



ESCUELA UNIVERSITARIA DE POSGRADO

INFLUENCIA DE LA INTELIGENCIA ARTIFICIAL Y MACHINE LEARNING EN LA
PREDICCIÓN Y PREVENCIÓN DE ACCIDENTES LABORALES EN LA MINERÍA
SUBTERRÁNEA DEL PERÚ

Línea de investigación:
Construcción sostenible y sostenibilidad ambiental del territorio

Tesis para optar el grado académico de Doctor en Ingeniería Civil

Autor

Zenteno Gomez, Floro Pagel

Asesor

Aroquipa Velasquez, Hector
ORCID: 0000-0002-6502-5618

Jurado

Chavarry Vallejos, Carlos Magno
García Urrutia Olavarria, Roque Jesús Leonardo
Díaz García, Martin Fernando

Lima - Perú

2025



INFLUENCIA DE LA INTELIGENCIA ARTIFICIAL Y MACHINE LEARNING EN LA PREDICCIÓN Y PREVENCIÓN DE ACCIDENTES LABORALES EN LA MINERÍA SUBTERRÁNEA DEL PERÚ

INFORME DE ORIGINALIDAD

27%

INDICE DE SIMILITUD

26%

FUENTES DE INTERNET

7%

PUBLICACIONES

10%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	repositorio.undac.edu.pe Fuente de Internet	2%
2	fondoeditorial.unat.edu.pe Fuente de Internet	2%
3	www.igme.es Fuente de Internet	2%
4	hdl.handle.net Fuente de Internet	1%
5	pdfs.semanticscholar.org Fuente de Internet	1%
6	ninive.ismm.edu.cu Fuente de Internet	1%
7	repositorio.ucundinamarca.edu.co Fuente de Internet	1%
8	Submitted to Universidad Cesar Vallejo Trabajo del estudiante	1%



Universidad Nacional
Federico Villarreal

VRIN | VICERRECTORADO
DE INVESTIGACIÓN

ESCUELA UNIVERSITARIA DE POS GRADO

INFLUENCIA DE LA INTELIGENCIA ARTIFICIAL Y MACHINE
LEARNING EN LA PREDICCIÓN Y PREVENCIÓN DE ACCIDENTES
LABORALES EN LA MINERÍA SUBTERRÁNEA DEL PERÚ

Línea de Investigación:

Construcción sostenible y sostenibilidad ambiental del territorio
Tesis para optar el grado académico de Doctor en Ingeniería Civil

Autor:

Zenteno Gomez, Floro Pagel

Asesor:

Aroquipa Velasquez, Hector
(ORCID: 0000-0002-6502-5618)

Jurados:

Chavarry Vallejos, Carlos Magno
García Urrutia Olavarria, Roque Jesús Leonardo
Díaz García, Martin Fernando

Lima – Perú

2025

DEDICATORIA

A mis padres Casiana y Agapito por sus invalorables inspiraciones de perseverancia y abnegación. A mi esposa Nelly Marisol y a mis hijos Hans, Jackeline, Hellen, Katherine y Jhoselyn por ser la fuerza de mis iluminaciones.

AGRADECIMIENTO

Reconocimiento profundo a Dios por su invaluable dote de fe y esperanza. Gratitud a los docentes de la Escuela de Posgrado de la Universidad Nacional Federico Villarreal por compartir sus experiencias académicas durante los estudios del Doctorado en Ingeniería Civil y la cristalización de la presente labor investigativa. Agradecimiento al Dr. Hector Aroquipa Velasquez, asesor de esta tesis; por lo que, por sus orientaciones se llegó a cumplir los objetivos del trabajo de investigación.

ÍNDICE

ÍNDICE	iii
RESUMEN.....	xiii
ABSTRACT.....	xiv
I. INTRODUCCIÓN	15
1.1. Planteamiento del Problema	17
1.2. Descripción del Problema	19
1.3. Formulación del Problema.....	20
1.4. Antecedentes	21
1.4.1. Internacionales.....	21
1.4.2. Nacionales.....	26
1.5. Justificación de la Investigación	28
1.5.1. Justificación Teórica.....	28
1.5.2. Justificación Metodológica	28
1.5.3. Justificación Práctica.....	29
1.5.4. Justificación Social	30
1.5.5. Importancia de la Investigación.....	30
1.5.6. Fundamento de la Investigación	31
1.6. Limitaciones de la Investigación	33
1.6.1. Espacial.....	33
1.6.2. Temporal	33
1.7. Objetivos	34
1.8. Hipótesis	35
II. MARCO TEÓRICO.....	36
2.1. Base Teórica.....	36

2.2.	Causas de los Accidentes Laborales	40
2.3.	Accidentes Laborales	43
2.4.	Excavaciones Mineras Subterráneas	46
2.4.1.	<i>Sostenimiento del Macizo Rocosó</i>	48
2.4.2.	<i>Sistema de Identificación de Factores</i>	49
2.5.	La Industria 4.0 y el Impulso de la Fábrica Inteligente	50
2.5.1.	<i>Seguridad y Salud en el Trabajo (SST)</i>	53
2.5.2.	<i>Análisis de la Gravedad de los Accidentes por Datos Textuales</i>	54
2.5.3.	<i>Uso de Técnicas de Minería de Texto y Aprendizaje Automático</i>	55
2.5.4.	<i>Preprocesamiento de Datos</i>	57
2.5.5.	<i>Optimización del Modelo y Evaluación del Rendimiento</i>	63
2.5.6.	<i>Identificar Factores de Atributos Críticos</i>	67
2.5.7.	<i>Análisis Clásico sobre los Factores de Riesgo en los Proyectos</i>	67
2.5.8.	<i>Aprendizaje Automático y Análisis de Riesgos</i>	69
2.5.9.	<i>Extracción y Clasificación de Factores de Riesgo</i>	74
2.5.10.	<i>Un Enfoque de Aprendiendo Sobre los Accidentes Graves</i>	78
2.5.11.	<i>Trabajos Relacionados - Métodos de Aprendizaje Automático</i>	80
2.6.	El Aprendizaje Automático (ML)	95
2.6.1.	<i>Modelo ML</i>	96
2.6.2.	<i>Descripción General del Aprendizaje Automático</i>	97
2.6.3.	<i>Algoritmos de Aprendizaje Automático</i>	98
2.6.4.	<i>Técnicas de Regresión</i>	99
2.6.5.	<i>Redes Neuronales y sus Variantes</i>	102
2.7.	Marco Filosófico	116
2.8.	Marco Tecnológico	120

2.8.1.	<i>Digitalización de la Minería</i>	120
2.9.	Marco Legal	121
2.10.	Marco Conceptual	124
III.	MÉTODO	129
3.1.	Tipo de Investigación.....	129
3.1.1.	<i>Tipo</i>	129
3.1.2.	<i>Nivel</i>	129
3.1.3.	<i>Códigos y Nomenclaturas UNESCO</i>	130
3.1.4.	<i>Diseño de la Investigación</i>	131
3.2.	Población y Muestra	132
3.2.1.	<i>Población</i>	132
3.2.2.	<i>Tamaño de la Muestra</i>	133
3.2.3.	<i>Muestreo</i>	133
3.3.	Operacionalización de Variables	134
3.3.1.	<i>Estrategia de Prueba de Hipótesis</i>	134
3.3.2.	<i>Variables e Indicadores</i>	136
3.4.	Instrumentos.....	142
3.4.1.	<i>Técnicas de Procesamiento de Datos</i>	142
3.4.2.	<i>Técnicas de Análisis e Interpretación de la Información</i>	143
3.4.3.	<i>Instrumentos de Recolección de Datos Según la Metodología</i>	147
3.5.	Procedimientos.....	151
3.6.	Análisis de Datos	156
3.7.	Consideraciones Éticas	162
IV.	RESULTADOS	163
4.1.	Introducción	163

4.1.1. <i>Caracterización de los Tipos de Accidentes</i>	163
4.1.2. <i>Algoritmo de Bosques Aleatorios en la Predicción de Accidentes</i>	184
4.1.3. <i>Precisión del Modelo Predictivo de Accidentes Laborales</i>	188
4.1.4. <i>Resultados del Análisis del Procesamiento de Datos</i>	189
V. DISCUSIÓN DE RESULTADOS	278
VI. CONCLUSIONES	280
6.1. Conclusiones por Objetivo General	280
6.2. Conclusiones por Objetivos Específicos.....	282
VII. RECOMENDACIONES	284
VIII. REFERENCIAS	288
IX. ANEXOS	314

ÍNDICE DE ANEXOS

Anexo A. Matriz de Consistencias.....	315
Anexo B. Guía para los Instrumentos de Recolección de Información.....	317
Anexo C. Ficha de Evaluación del Instrumento de Recolección de Datos.....	323
Anexo D. Ficha de Evaluación del Instrumento de Recolección de Datos	324
Anexo E. Ficha de Evaluación del Instrumento de Recolección de Datos	325
Anexo F. Preprocesamiento de Datos	326
Anexo G. Algoritmo Random Forest.....	326
Anexo H. Algoritmo Support Vector Machine (SVM)	327
Anexo I. Algoritmo de Naive Bayes.....	327
Anexo J. Algoritmo de Comparación de Resultados	327
Anexo K. Resultado de la Comparación de Modelos	328
Anexo L. Programa de la Segmentación y Limpieza de Datos	328
Anexo M. Vectorización del Texto.....	329
Anexo N. Gestión del Desbalance de Clases	329
Anexo O. Programa del Entrenamiento del Modelo Predictivo	329

ÍNDICE DE FIGURAS

Figura 1. Seguridad Laboral en Minería Subterránea	37
Figura 2. Desprendimiento de Rocas por Falta de Mantenimiento de la Madera.....	49
Figura 3. Análisis de Factores de Riesgo de Seguridad del Proyecto.....	72
Figura 4. Clasificación de Documentos de Texto Basada en CNN	75
Figura 5. Configuración del Modelo CNN y Resultados de la Prueba	78
Figura 6. Red Neuronal Profunda	91
Figura 7. Representación Esquemática de un Modelo Wide&Deep.....	93
Figura 8. Flujo de Trabajo Típico de ML	98
Figura 9. Algoritmo ML Agrupados por Tipo de Aprendizaje.....	99
Figura 10. Técnicas de Regresión.....	100
Figura 11. Ejemplo de una ANN	103
Figura 12. Arquitectura CNN Típica	106
Figura 13. Ejemplo de SVM.....	108
Figura 14. Ilustración de Parámetro de Penalización y Funciones Kernel	110
Figura 15. Ejemplo de un DT	111
Figura 16. Diagrama de Flujo de RF (Entrenamiento Paralelo)	112
Figura 17. Implementación de Adaboost con Dos Estudiantes Débiles	114
Figura 18. Ilustración del Modelo GBM.....	115
Figura 19. Metodología Aplicada al Análisis y Evaluación de Accidentes.....	164
Figura 20. Accidentes Mortales en la Minería Peruana de 1970 al 2018	168

Figura 21. Accidentes Mortales Mensuales Durante los Años 2000 y 2018.....	170
Figura 22. Accidentes Mortales por Año del 2000 al 2018 y R2 de Tendencia.....	172
Figura 23. Accidentes Mortales en Minería por Tipo de Accidente de 2000-2018...	174
Figura 24. Accidentes Según Tipo de Sostenimiento	175
Figura 25. Accidentes Mortales Según Turno de Trabajo	176
Figura 26. Accidentes Mortales Según Ocupación de las Víctimas	177
Figura 27. Accidentes Mortales Según Tiempo de Servicios	178
Figura 28. Índice de Frecuencia de Accidentes	180
Figura 29. Índice de Severidad de Accidentes.....	181
Figura 30. Índice de Accidentabilidad.....	182
Figura 31. Número de Accidentes de Empresas y Contratistas Mineros.....	183
Figura 32. Análisis de Características de Multicolinealidad	192
Figura 33. Matriz de Correlación de Factores de Accidentes.....	196
Figura 34. Factores que Influyen en los Accidentes Laborales	197
Figura 35. Análisis de Multicolinealidad por Factores Humanos (HF).....	202
Figura 36. Barras de Error por Factores Humanos (HF1 a HF15).....	205
Figura 37. Correlación de Factores Humanos de Accidentes	208
Figura 38. Componentes de los Factores Humanos (HF1 a HF15).....	211
Figura 39. Componentes y Correlación de Retardo para Factores Humanos	213
Figura 40. Componentes Principales de Factores Humanos (HF1 a HF15).....	217
Figura 41. Multicolinealidad de Accidentes por Factores de Gestión (MF).....	222

Figura 42. Barras de Error de Accidentes Debido a Factores de Gestión.....	225
Figura 43. Correlación de Factores de Gestión (MF) de Accidentes.....	228
Figura 44. Componentes Principales Aplicado a los Factores de Gestión	231
Figura 45. Factores de Gestión	235
Figura 46. Factores de Gestión y Correlación de Retardo	239
Figura 47. Multicolinealidad de Accidentes por Factores de Instalaciones (FF)	242
Figura 48. Barras de Error Debido a Factores de Instalaciones (FF).....	245
Figura 49. Correlación de 07 Diferentes Factores de Instalaciones (FF1 a FF7)	247
Figura 50. Componentes Principales por Factores de Instalaciones.....	250
Figura 51. Factores de Instalaciones (FF1 a FF7).....	253
Figura 52. Factores de Gestión y Correlación de Retardo	257
Figura 53. Multicolinealidad de Accidentes por Factores Ambientales (EF).....	263
Figura 54. Barras de Error Para los Accidentes por Factores de Ambientales (EF)..	266
Figura 55. Correlación de 03 Diferentes Factores Ambientales	268
Figura 56. Componentes Principales (PCA).....	270
Figura 57. Componentes Principales (PCA) de los Factores Ambientales.....	272
Figura 58. Factores Ambientales - Correlación de Retardo.....	274

ÍNDICE DE TABLAS

Tabla 1. Experiencia del SSOMAC	45
Tabla 2. Categorías de Consecuencias de Accidentes	87
Tabla 3. Distribución de Titulares y Unidades por Estratos de Empresas Mineras...	133
Tabla 4. Variables e Indicadores – Operacionalización de Variables	141
Tabla 5. Accidentes Mortales en la Minería Peruana del 2000 al 2018	169
Tabla 6. Caracterización de los Tipos de Accidentes (AT)	173
Tabla 7. Índice de Frecuencia, Severidad y Accidentabilidad.....	179
Tabla 8. Algoritmo de Bosques Aleatorios, Comparación de los Modelos.....	187
Tabla 9. Factores que Influyen en los Accidentes	191
Tabla 10. Accidentes debido a Factores Humanos (HF)	201
Tabla 11. Matriz de Correlación <i>Static Spearman SD</i>	207
Tabla 12. Valores de <i>Eigenvalue</i> y <i>% Variance</i> - Factores Humanos	216
Tabla 13. Accidentes debido a Factores de Gestión (MF).....	220
Tabla 14. Valores de <i>Eigenvalue</i> de los Factores Gestión Mediante PCA	238
Tabla 15. Accidentes debido a Factores de Instalaciones (FF).....	241
Tabla 16. Valores de <i>Eigenvalue</i> - Factores de Instalaciones Mediante PCA.....	257
Tabla 17. Accidentes Debido a Factores de Ambientales (EF)	261

ÍNDICE DE ABREVIATURAS

Abreviaturas	Significado
AI	Environmental Factors (Factores Ambientales)
AU-ROC	Area Under the Curve - Receiver Operating Characteristic (Área Bajo la Curva - Característica Operativa del Receptor)
EF	Facility Factors (Factores de Instalaciones)
FF	Facility Factors (Factores de Instalaciones)
HF	Human Factors (Factores Humanos)
ICC	Intraclass Correlation Coefficient (Coeficiente de Correlación Intraclase)
MF	Management Factors (Factores de Gestión)
PC	Principal Component (Componente Principal)
ROC	Receiver Operating Characteristic (Característica Operativa del Receptor)
SMOTE	Synthetic Minority Over-sampling Technique (Técnica de Sobremuestreo de Minorías Sintéticas)
SVM	Support Vector Machine (Máquina de Vectores de Soporte)

RESUMEN

El objetivo de esta investigación es reducir los accidentes laborales por desprendimiento de rocas en la minería subterránea peruana mediante la aplicación de un modelo predictivo basado en técnicas de inteligencia artificial y *machine learning*, con un enfoque en mejorar la seguridad y las condiciones de trabajo. La metodología, se emplearon algoritmos de *machine learning*, como *Random Forest*, para predecir la severidad de los accidentes, comparando su efectividad con otros algoritmos como *Support Vector Machines* (SVM) y *Naive Bayes*. Además, se optimizó el preprocesamiento de datos textuales y la gestión del desbalance de clases para mejorar la precisión del modelo predictivo. Los resultados revelaron que la mayoría de los accidentes ocurren en la minería subterránea, con el 60% de las víctimas fatales pertenecientes a empresas subcontratistas. El análisis identificó 49 subcategorías de factores de riesgo, agrupados en factores humanos, de gestión, de instalaciones y ambientales. El algoritmo *Random Forest* mostró una precisión del 85%, superando a otros algoritmos evaluados en términos de precisión (82% para SVM y 78% para *Naive Bayes*), sensibilidad (87% para *Random Forest* frente a 80% para SVM) y especificidad (84% para *Random Forest* frente a 79% para SVM). Las conclusiones destacan que la implementación de estrategias basadas en inteligencia artificial puede mejorar significativamente la seguridad en la minería, recomendándose la adopción de sistemas de gestión integral de seguridad y la mejora continua de las políticas de prevención.

Palabras Clave: Accidentes laborales, Minería subterránea, Inteligencia artificial, *Machine learning*, Seguridad laboral.

ABSTRACT

The objective of this research is to reduce occupational accidents due to rockfall in Peruvian subway mining by applying a predictive model based on artificial intelligence and machine learning techniques, with a focus on improving safety and working conditions. The methodology. Machine learning algorithms, such as Random Forest, were used to predict the severity of accidents, comparing their effectiveness with other algorithms such as Support Vector Machines (SVM) and Naive Bayes. In addition, textual data preprocessing and class imbalance management were optimized to improve the accuracy of the predictive model. The results revealed that most accidents occur in subway mining, with 60% of fatalities belonging to subcontractors. The analysis identified 49 subcategories of risk factors, grouped into human, management, facility and environmental factors. The Random Forest algorithm showed an accuracy of 85%, outperforming other algorithms evaluated in terms of accuracy (82% for SVM and 78% for Naive Bayes), sensitivity (87% for Random Forest vs. 80% for SVM) and specificity (84% for Random Forest vs. 79% for SVM). Critical factors identified included inadequate supervision, non-compliance with safety regulations, and lack of adequate training. The conclusions highlight that the implementation of strategies based on artificial intelligence can significantly improve safety in mining, recommending the adoption of comprehensive safety management systems and the continuous improvement of prevention policies.

Keywords: Occupational accidents, Subway mining, Artificial intelligence, Machine learning, Occupational safety.

I. INTRODUCCIÓN

En los últimos años, la minería peruana ha experimentado una reducción en la tasa de lesiones y accidentes. Sin embargo, esta actividad sigue siendo una de las más arriesgadas debido a los peligros inherentes a su naturaleza. Las condiciones y actos inseguros son las causas básicas de la ocurrencia de fatalidades en este sector, lo que hace imprescindible la identificación de problemas de seguridad y la gestión adecuada de los riesgos asociados. La dependencia de países en vías de desarrollo de industrias primarias, como la minería, es evidente. En Perú, esta actividad representa aproximadamente el 16% del Producto Bruto Interno (PBI) y más del 50% de las exportaciones, subrayando su importancia económica. A pesar de su contribución significativa al desarrollo económico, la minería también se destaca por los altos riesgos que implica, tanto para la seguridad de los trabajadores como para la sostenibilidad de las operaciones.

El presente trabajo se centra en el análisis de los accidentes fatales ocurridos en la minería peruana desde el año 2007 hasta el 2018, identificando los principales tipos de accidentes. Para ello, se utilizó como fuente de información primaria el registro de accidentes fatales proporcionado por la Dirección General de Minería del Ministerio de Energía y Minas del Perú. Los resultados muestran que la mayoría de los accidentes ocurrieron en la minería subterránea y que la mayor parte de las víctimas fatales eran trabajadores de empresas subcontratistas. Este contexto exige la propuesta de medidas necesarias para mejorar la gestión de seguridad y reducir los accidentes fatales.

La investigación se desarrolló durante siete meses con datos de diversas minas del Perú, se realizó con el uso de técnicas avanzadas de inteligencia artificial y *machine learning* para analizar y predecir la severidad de los accidentes laborales. La minería es una actividad industrial que abarca la explotación de minerales metálicos, no metálicos y combustibles. Estos

recursos minerales son esenciales para la producción de energía eléctrica, automóviles, agregados para construcción, aceros, fertilizantes, asfaltos, medicinas, componentes electrónicos, entre otros. La explotación de estos recursos es fundamental para el mundo contemporáneo, pero también es una de las actividades más riesgosas debido a los diversos peligros asociados, como choques eléctricos, explosiones, gases, rocas sueltas, ruidos, polvo, iluminación deficiente y ventilación inadecuada.

A pesar de una reducción considerable en la tasa de accidentes de trabajo en los últimos años, el número y la severidad de estos incidentes siguen siendo altos. Los gobiernos de países con tradición minera, como Perú, han mostrado un creciente interés por la salud y seguridad de los trabajadores. Sin embargo, la información sobre la ocurrencia de accidentes es limitada, especialmente en países en vías de desarrollo donde existe una su declaración de incidentes por parte de las empresas, dificultando su comprensión y análisis. Los investigadores del área de seguridad coinciden en que las causas fundamentales de los accidentes son las condiciones y actos inseguros. Las condiciones inseguras se presentaron de diversas formas, como proyectos inadecuados, falta de reconocimiento de las condiciones geológicas y deficiencia en el mantenimiento de equipos. Los actos inseguros, por su parte, suelen estar asociados al comportamiento inadecuado de los trabajadores, generalmente debido a la falta de información.

Mejorar la seguridad en la minería implica planificar e implementar políticas de seguridad a través de sistemas de gestión integral, como los contemplados por la familia de normas ISO (Organización Internacional de Normalización). Estas políticas deben ser controladas y monitoreadas permanentemente para garantizar su efectividad mediante programas de mejora continua. Las acciones corporativas deben orientarse hacia la búsqueda de métodos y técnicas para reducir los riesgos laborales, evitando la ocurrencia de accidentes. El presente estudio propone la aplicación de técnicas de inteligencia artificial y *machine learning* para mejorar la gestión de seguridad en la minería peruana. La investigación se centró

en el desarrollo de un modelo predictivo que permitió identificar los factores críticos que contribuyeron a la severidad de los accidentes, utilizando datos no estructurados de reportes de incidentes. Este enfoque desarrolló una mejor comprensión de los procesos productivos y la implementación de medidas preventivas efectivas.

1.1. Planteamiento del Problema

En el sector minero peruano, la seguridad y salud ocupacional (OSH) por sus siglas en inglés *Occupational Safety and Health*, es de vital importancia debido a los riesgos inherentes que conlleva la actividad minera. La estadística de la Organización Internacional del Trabajo (OIT) indican que anualmente ocurren más de 2.78 millones de muertes por accidentes laborales a nivel mundial, y 374 millones de trabajadores sufren accidentes no mortales, siendo el sector minero uno de los más afectados. En Perú, la situación no es diferente, los accidentes laborales en las minas representan una preocupación constante, afectando no solo la salud de los trabajadores sino también la eficiencia operativa y la imagen corporativa de las empresas mineras. A pesar de los esfuerzos realizados para mejorar la seguridad en el sector minero, persisten altos índices de accidentes laborales. Las metodologías tradicionales de análisis de accidentes basadas en encuestas y entrevistas de expertos presentan limitaciones en cuanto a la objetividad y capacidad interpretativa. Existe una necesidad urgente de implementar nuevas metodologías que permitan un análisis más preciso y efectivo de los datos de accidentes, utilizando tecnologías emergentes como la inteligencia artificial (IA) y el *machine learning* (ML).

El objetivo principal de esta investigación es desarrollar un marco de predicción de la severidad de los accidentes laborales en el sector minero peruano, utilizando técnicas avanzadas de *machine learning*. Este marco permitió mapear los atributos y características de los accidentes (por ejemplo, equipo utilizado, escenario y atributos ambientales) a categorías

de severidad (daños personales y materiales), proporcionando a los gerentes y encargados de seguridad estrategias efectivas de prevención de accidentes. Asimismo, la aplicación de algoritmos de *machine learning*, como los bosques aleatorios (RF) *Random Forests*, por sus siglas en inglés), ha demostrado ser efectiva en la predicción de la severidad de los accidentes en otros sectores industriales, como el de la construcción. Sin embargo, su implementación en el sector minero peruano es aún limitada. La investigación se enfocó en mejorar el preprocesamiento de datos textuales y en optimizar los algoritmos de *machine learning* para manejar datos no estructurados, desbalance de clases y otras dificultades inherentes en los datos de accidentes laborales. Por otro lado, el marco metodológico de esta investigación incluye dos fases principales: preprocesamiento de datos y análisis de datos. En la primera fase, se convirtió la información textual original de los accidentes en un formato adecuado para el análisis computacional. En la segunda fase, se aplicó el algoritmo de bosques aleatorios para construir un modelo de predicción de la severidad de los accidentes y evaluar su rendimiento. La validación del modelo se realizó utilizando datos de accidentes de colapso en la minería como estudio de caso. Los resultados de esta investigación proporcionaron una herramienta robusta para la predicción de la severidad de los accidentes laborales en el sector minero, permitiendo identificar los factores críticos que contribuyeron a la ocurrencia de estos accidentes y ofreciendo estrategias de prevención basadas en datos para minimizar los riesgos y las pérdidas. Además, se espera que la implementación de esta metodología en el sector minero peruano contribuya significativamente a mejorar la seguridad y salud ocupacional, promoviendo un ambiente de trabajo más seguro y eficiente. La aplicación de técnicas de inteligencia artificial y *machine learning* en el análisis de accidentes laborales representa una oportunidad significativa para abordar los desafíos de seguridad en el sector minero peruano. Esta investigación busca llenar el vacío existente en la literatura y en la práctica actual,

proporcionando un enfoque innovador y efectivo para la gestión de la seguridad y salud ocupacional en las minas peruanas.

1.2. Descripción del Problema

La minería es una de las actividades económicas más importantes en el Perú, esta actividad contribuye significativamente al desarrollo económico del país. Sin embargo, es una de las industrias más peligrosas debido a la naturaleza de sus operaciones, que implican riesgos elevados para la seguridad y salud de los trabajadores. Los accidentes laborales en el sector minero pueden resultar en lesiones graves, discapacidades permanentes y, en los peores casos, la muerte. Además, estos incidentes tienen un impacto negativo en la productividad, los costos operativos y la reputación de las empresas mineras. A pesar de los esfuerzos continuos para mejorar las condiciones de seguridad, la tasa de accidentes laborales en el sector minero peruano sigue siendo alta. Las metodologías tradicionales utilizadas para analizar y prevenir estos accidentes, como las encuestas y entrevistas con expertos, presentan limitaciones significativas. Estas metodologías a menudo carecen de objetividad y no pueden capturar la complejidad y variabilidad de los factores que contribuyen a los accidentes laborales.

Los métodos actuales de análisis de accidentes se basan en gran medida en datos estructurados y enfoques cualitativos, lo que limita su capacidad para predecir y prevenir futuros accidentes de manera efectiva. La información valiosa contenida en datos no estructurados, como los reportes de incidentes y descripciones de accidentes, a menudo no se utiliza de manera adecuada debido a la falta de herramientas y técnicas avanzadas para su procesamiento y análisis. Por lo tanto, la inteligencia artificial (IA) y el *machine learning* (ML) ofrecen una oportunidad significativa para superar estas limitaciones. Estos enfoques permiten el análisis de grandes volúmenes de datos no estructurados y la identificación de patrones y relaciones complejas que los métodos tradicionales no pueden detectar. En particular, los

algoritmos de *machine learning*, como los bosques aleatorios, han demostrado ser efectivos en la predicción de la severidad de los accidentes laborales en diversos sectores industriales. Por lo descrito, se plantea como objetivo del estudio desarrollar un modelo de predicción: Utilizar técnicas de *machine learning* para desarrollar un modelo que pueda predecir la severidad de los accidentes laborales en el sector minero peruano basándose en atributos claves del accidente y del entorno laboral. Mejorar el preprocesamiento de datos: Optimizar las técnicas de preprocesamiento de datos textuales para manejar datos no estructurados y desbalanceados, mejorando la precisión y eficacia del modelo predictivo. Identificar factores críticos: Analizar los factores críticos que contribuyen a la severidad de los accidentes, proporcionando información valiosa para la implementación de estrategias de prevención efectivas. Finalmente, la implementación exitosa de este enfoque innovador tiene el potencial de transformar la gestión de la seguridad y salud ocupacional en el sector minero peruano. Al proporcionar una herramienta avanzada para la predicción de accidentes y la identificación de factores de riesgo, esta investigación puede contribuir significativamente a la reducción de accidentes laborales, mejorando la seguridad de los trabajadores y la eficiencia operativa de las empresas mineras.

1.3. Formulación del Problema

- Problema General

- ¿En qué medida influye la aplicación de técnicas de inteligencia artificial y *machine learning* en la predicción y prevención de accidentes laborales en la minería subterránea del Perú?

- Problemas Específicos

- ¿Cuáles son los principales factores de riesgo asociados a los accidentes laborales por desprendimiento de rocas en la minería subterránea peruana?

- ¿Qué tan efectiva es la técnica de bosques aleatorios (*Random Forest*) en la predicción de la severidad de accidentes laborales en la minería subterránea comparada con otros algoritmos de *machine learning*?
- ¿De qué manera la mejora en el preprocesamiento de datos textuales y la gestión del desbalance de clases impacta la precisión del modelo predictivo de accidentes laborales?
- ¿Cuáles son los factores críticos identificados mediante *machine learning* que contribuyen significativamente a la ocurrencia y severidad de accidentes laborales en la minería subterránea?
- ¿Qué estrategias de prevención de accidentes pueden derivarse del análisis de los datos y resultados obtenidos mediante la aplicación de inteligencia artificial en la minería subterránea?

1.4. Antecedentes

Los antecedentes de la investigación ayudan a orientar objetivamente según las recomendaciones de cada investigación desarrollada, que también han sido de ayuda para el análisis y discusión de los resultados. Se describen en síntesis los planteamientos, análisis y resultados de la investigación con el objeto de enriquecer la discusión.

1.4.1. Internacionales

Candia et al. (2010) en el estudio “Análisis de accidentes fatales en la industria minera peruana” explica que en los últimos años se pudo observar una reducción en la tasa de lesiones y accidentes en la minería peruana.

En el Perú esta actividad representa aproximadamente el 16% del Producto Bruto Interno (PBI) y más del 50% de las exportaciones, lo que le confiere mayor

importancia. El presente trabajo tiene como objetivo analizar los accidentes fatales ocurridos en la minería desde el año 2000 a 2007, identificando los principales tipos de accidentes. Los resultados muestran que la mayoría de los accidentes ocurrieron en la minería subterránea; así como la mayor parte de las víctimas fatales pertenecieron a empresas subcontratistas; lo que exige la propuesta de medidas necesarias para su mejor gestión buscando la reducción de los accidentes fatales. Las conclusiones son: El análisis desarrollado en este trabajo muestra que para el período estudiado (2000-2007), el tipo de accidente más frecuente es el causado por la caída de rocas en las excavaciones subterráneas. Considerando que el mayor volumen de producción es aportado por minas a cielo abierto, se puede inferir que la minería subterránea ofrece mayores riesgos que la minería a cielo abierto. La mayor parte de las víctimas fatales pertenecen a empresas especializadas tercerizadas, esto se explica debido a que la mayoría de las operaciones de minas así como las operaciones auxiliares en el Perú están subcontratadas.

Koldas (2001) en el estudio “*Rock-Related Accidents, Investigations and Inquiries in South African Mines*” explica que la industria minera de Sudáfrica es un importante contribuyente a la economía del país.

En 1998, la minería y las canteras contribuyeron con \$ 6.6 mil millones, o 6.6% del Producto Interno Bruto (PIB), y \$ 1.7 mil millones, o 8.6% del Bruto Doméstico Fijo Inversión (GDFT). La industria también contribuyó alrededor del 1.2% de los ingresos estatales. En 1998, en Sudáfrica las ventas primarias totales se valoraron en aproximadamente \$ 12.8 mil millones, de los cuales las ganancias de las exportaciones contribuyeron 76.6%, mientras que el oro representó el 44.2% de las exportaciones. Aunque los minerales primarios contribuyeron con un 37.2% de los ingresos de divisas del país, la adición de productos minerales procesados aumentó esa cifra muy

por encima del 50%. Las conclusiones del estudio son las siguientes: Con los años, la minería sudafricana industria sostenida, y continúa manteniendo, un alto nivel de accidentes relacionados con rocas y la tasa resultante de bajas humanas (lesiones y muertes) ha sido, en términos mundiales, inaceptablemente alto. De estos accidentes, el 43% de todos los accidentes han sido relacionados con rocas, es decir, el resultado de caídas de rocas o estallidos de rocas. En las minas de oro, los accidentes de caída de rocas y estallidos también constituyen la causa más importante de todas las muertes. En 1999, el 68% de todas las muertes ocurrieron en minas de oro, de las cuales el 51% de todos los accidentes ocurridos en las minas de oro estaban relacionadas con las rocas. La industria minera en Sudáfrica, especialmente minas de oro de nivel profundo, no tiene más remedio que mejorar radicalmente sus registros de seguridad y productividad.

Mazaira & Konicek (2015) en el estudio “*Intense rockburst impacts in deep underground construction and their prevention*” describe que en las últimas décadas, tanto en proyectos de minería como de ingeniería civil, se han llevado a cabo excavaciones más profundas que en el pasado.

Con este aumento de profundidad, las rocas pueden sobrecargarse y pueden producirse fallas durante la excavación. Cuando los estratos de roca se encuentran en condiciones de alta tensión, estas fallas pueden variar desde desprendimiento superficial hasta una explosión. Los estallidos intensos de rocas pueden causar lesiones fatales a los trabajadores y una pérdida significativa de equipos y tiempo. La aparición de estallidos de rocas siempre es difícil de predecir y se deben tomar medidas especiales para controlarlos. **Primero**, las zonas propensas a estallar deben ser ubicadas por un estudio geológico exhaustivo y oportuno y por la evaluación del nivel de esfuerzo in situ y la orientación. **Segundo**, los parámetros de diseño, por ejemplo, forma, tamaño

y método de excavación, deben modificarse y adaptarse a las condiciones esperadas para minimizar el riesgo de explosión de rocas.

Joughin et al. (2012) en el estudio “*A risk evaluation model for support design in Bushveld Complex underground mines: Part I–Description of the model*” se ha desarrollado un modelo de evaluación de riesgos que permite la cuantificación de lesiones esperadas y pérdidas económicas derivadas del desprendimiento de rocas asociados con un sistema de soporte en una determinada labor de minería subterránea.

Esto permite la comparación de diferentes sistemas de soportes basados en la frecuencia esperada de lesiones y el costo total del sistema de apoyo, incluidas las pérdidas económicas esperadas. Las conclusiones del estudio son las siguientes: El modelo propuesto permite la cuantificación de lesiones por caída de rocas y pérdidas económicas debido a la caída de rocas para un sistema de apoyo en un entorno geotécnico dado. Diferentes sistemas de apoyo se puede comparar y la efectividad del monitoreo y la restricción puede evaluarse en términos de la frecuencia esperada de lesiones y el costo total de los sistemas de apoyo, incluido las pérdidas económicas esperadas como resultado de fallas del sistema. El software existente, JBlock, se mejoró para determinar la simulación de caída de rocas. Esta herramienta ahora se puede usar para probar y calcular el soporte de diseño para satisfacer la gama completa de tamaños de caída de rocas que se puede anticipar en un entorno geotécnico. Las caídas de rocas simuladas ahora pueden normalizarse por el área de minado. Las consecuencias de estas caídas de rocas se evalúan utilizando el nuevo software RiskEval. La frecuencia de las lesiones es simulado al considerar la exposición temporal y espacial del personal.

Gómez et al. (2014) en el estudio “*Factores de Riesgo Psicosocial y Satisfacción Laboral en una Empresa Chilena del Área de la Minería*” , el objetivo de esta investigación es

identificar la relación entre factores de riesgo psicosocial y satisfacción laboral en trabajadores de una empresa subcontratista del sector minero en Chile.

Se utilizó cuestionarios aplicados a una muestra de 100 trabajadores. Los resultados dan cuenta de que existe una relación significativa y negativa entre factores de riesgo y satisfacción laboral; en las dimensiones de trabajo activo y posibilidades del desarrollo, apoyo social en la empresa y calidad de liderazgo, compensaciones y doble presencia, indicando que a mayor riesgo psicosocial percibido menor es la satisfacción laboral. Para la dimensión trabajo activo y posibilidades de desarrollo, se encontró que operarios/ayudantes poseen un rango de exposición alto y los demás en un rango medio, además de una relación significativa entre la dimensión exigencias psicológicas según la sección donde trabajan. Se discuten las implicancias del estudio para futuras investigaciones e intervenciones organizacionales. Las conclusiones del estudio son: La dimensión trabajo activo y posibilidades de desarrollo obtuvo diferencias estadísticamente significativas en cuanto a las variables cargo y sección. Es así como los trabajadores del cargo de operario/ayudante consideran que no tienen influencia sobre su trabajo ni en la relación con los compañeros, sienten que su trabajo no es fuente de oportunidades de desarrollo de sus habilidades y conocimientos, sienten que no tienen influencia sobre los tiempos de descanso, como vacaciones o detenciones temporales. Los trabajadores de los cargos de operario/ayudante y de las secciones productivas (a excepción del área de recursos humanos) como chancado primario, chancado secundario, chancado terciario, cuyas tareas son principalmente manuales, identifican mayores riesgos en esta área en comparación con los trabajadores del área administrativa.

1.4.2. Nacionales

Pérez (2007) en el estudio “Sistema de Gestión en Seguridad y Salud Ocupacional Aplicado a Empresas Contratistas en el Sector Económico Minero Metalúrgico”:

Se obtiene como conclusiones que todas las empresas contratistas a nivel nacional deberán implementar un sistema de seguridad y salud ocupacional ya sea propio o adaptado. Ya que esto les dará los lineamientos, herramientas y controles para poder realizar una gestión exitosa. Entonces al aplicar y desarrollar correctamente el presente sistema de seguridad y salud ocupacional se disminuirá la tendencia de accidentes fatales. Además, es necesario aprender a ser proactivos antes de ser reactivos; la seguridad no debe hacerse solo por reacción, la seguridad debe aplicarse por prevención. Al implementar un sistema de seguridad y salud ocupacional adecuado se ha de obtener la disminución de pérdidas incrementándose las utilidades; con lo cual se mejora las condiciones laborales incrementando la productividad.

Terán (2012) en el estudio “Propuesta de Implementación de un Sistema de Gestión de Seguridad y Salud Ocupacional bajo la Norma OHSAS 18001 en una Empresa de Capacitación Técnica para la Industria”:

Busca dar a conocer que mediante la aplicación de un Sistema de Gestión de Seguridad y Salud Ocupacional Implementado en una Empresa de Capacitación Técnica se puede lograr el control de la seguridad de sus procesos y un mayor respaldo para la Empresa. Entre las conclusiones del trabajo se determina que para la efectividad de la implementación del sistema de gestión de Seguridad y Salud Ocupacional es necesario realizar auditorías internas que permitan establecer las no conformidades y realizar el respectivo seguimiento, proporcionando los lineamientos necesarios para que la

empresa logre sus metas. Las auditorias deben realizarse siguiendo un programa anual, donde la frecuencia puede variar en función al estado e importancia del proceso.

Moreno (2011) en el estudio “Modelo de un Programa de Seguridad e Higiene para la Minería Subterránea del Carbón en Colombia” tiene como objetivo diseñar un modelo de programa de seguridad y salud para la pequeña y mediana minería subterránea del carbón en Colombia.

Como conclusión del estudio se tiene que la institucionalización y normatividad de la seguridad y salud ocupacional en Colombia data del año 1979 con la ley conocida como la ley marco de la salud ocupacional, la cual dicta medidas sanitarias para la prevención y el mantenimiento de la salud de los trabajadores con todos los posibles riesgos a que se encuentran expuestos. Cada año esta ley marco, viene siendo complementada y/o actualizada por leyes, decretos y resoluciones que buscan reducir la alta accidentalidad y mejorar la integridad de la vida de los trabajadores independiente de la actividad económica en donde laboran.

Calderon (2015) en la investigación intitulada “Optimización de las prácticas de perforación y voladura en el avance y producción de la minería de mediana escala (Unidad Minera MACDESA)”.

El objetivo del estudio es optimizar las prácticas de perforación y voladura en los avances y la producción en la mina modelo de mediana escala MACDESA. La hipótesis del estudio consiste: La mejora en las técnicas de perforación y voladura permitirán la optimización del avance y producción en la compañía minera MACDESA. La hipótesis general del estudio es: La mejora en las técnicas de perforación y voladura permitirán la optimización del avance y producción en la compañía minera MACDESA.

1.5. Justificación de la Investigación

1.5.1. Justificación Teórica

La investigación sobre la aplicación de inteligencia artificial y *machine learning* para la predicción y prevención de accidentes laborales en la minería subterránea del Perú se basa en la necesidad de mejorar la gestión de riesgos y la seguridad laboral en un sector caracterizado por altos índices de siniestralidad (Cruz & Huerta, 2015). La teoría de la gestión de riesgos enfatiza la importancia de identificar, evaluar y mitigar riesgos para proteger a los trabajadores (Hubbard, 2020). Además, las técnicas de *machine learning* permiten manejar grandes volúmenes de datos y descubrir patrones complejos, superando las limitaciones de los métodos tradicionales (Hollnagel et al., 2006). Esta investigación también se apoya en la teoría de la resiliencia organizacional, que destaca la capacidad de las empresas para anticipar y responder a eventos inesperados (Hollnagel et al., 2006). Al integrar datos y aplicar análisis predictivo, la investigación maximiza el valor de la información recolectada, proporcionando una base sólida para la toma de decisiones informadas (Provost & Fawcett, 2013). Además, contribuye a la literatura científica combinando enfoques tradicionales con tecnologías avanzadas, y aborda cómo las percepciones y actitudes de los trabajadores influyen en la efectividad de las estrategias preventivas (Reason, 1990). Por lo tanto, esta investigación no solo mejora la seguridad laboral en la minería subterránea del Perú, sino que también ofrece soluciones prácticas y conocimientos científicos valiosos.

1.5.2. Justificación Metodológica

La justificación metodológica de esta investigación radica en la aplicación de un diseño no experimental, transeccional y descriptivo-correlacional, adecuado para observar los fenómenos en su contexto natural sin manipular variables, permitiendo describir y analizar la interrelación de factores de riesgo en la minería subterránea (Hernández et al., 2010). La

integración de técnicas avanzadas de *machine learning*, como *Random Forest* y *Support Vector Machines*, se justifica por su capacidad para manejar grandes volúmenes de datos y descubrir patrones complejos, lo cual es fundamental para desarrollar modelos predictivos precisos (Kim, 2022). La validación cruzada asegura la robustez y generalización de estos modelos, fortaleciendo la validez de los resultados (Kohavi, 1995). Además, el preprocesamiento de datos textuales y el análisis de datos mediante técnicas de procesamiento de lenguaje natural permiten extraer información valiosa de reportes no estructurados, complementando el análisis cuantitativo (Creswell & Creswell, 2017). El uso de métodos estadísticos avanzados, como análisis de correlación y regresión, y la simulación de Monte Carlo, proporciona una evaluación completa y detallada de los factores de riesgo, apoyando decisiones informadas y efectivas (George et al., 2005). La inclusión de métodos cualitativos, como entrevistas y encuestas, enriquece la comprensión de los factores de riesgo y la cultura de seguridad, complementando el análisis cuantitativo (Creswell & Creswell, 2017). Por último, la validación de instrumentos mediante pruebas piloto y revisión de expertos asegura la fiabilidad y validez de los datos recolectados (Robert, 2019).

1.5.3. Justificación Práctica

La justificación práctica de esta investigación radica en su potencial para mejorar significativamente la seguridad laboral en la minería subterránea del Perú, un sector con alta incidencia de accidentes graves. Al aplicar técnicas de inteligencia artificial y *machine learning*, esta investigación proporciona herramientas avanzadas para predecir y prevenir accidentes laborales, lo cual es crucial para reducir la frecuencia y severidad de estos eventos (Goodfellow et al., 2016). La implementación de modelos predictivos permite a las empresas mineras identificar proactivamente los factores de riesgo y adoptar medidas preventivas eficaces, mejorando así la seguridad y el bienestar de los trabajadores (Hubbard, 2020). Además, la utilización de análisis de datos textuales y cuantitativos proporciona una

comprensión más completa y detallada de las causas subyacentes de los accidentes, facilitando el desarrollo de políticas y prácticas de seguridad más informadas y efectivas (Provost & Fawcett, 2013). En resumen, esta investigación ofrece una contribución práctica significativa al proporcionar soluciones basadas en datos que pueden ser implementadas directamente en las operaciones mineras para mitigar riesgos y proteger la vida de los trabajadores.

1.5.4. Justificación Social

La justificación social de esta investigación reside en su capacidad para mejorar la calidad de vida y las condiciones laborales de los trabajadores en la minería subterránea del Perú, un sector conocido por sus altos riesgos y la frecuencia de accidentes graves. Al implementar técnicas de inteligencia artificial y *machine learning* para predecir y prevenir accidentes, se contribuye a la creación de un entorno laboral más seguro, reduciendo así la incidencia de lesiones y muertes (Goodfellow et al., 2016). Esto no solo beneficia directamente a los trabajadores y sus familias al disminuir el sufrimiento y las pérdidas económicas asociadas con los accidentes laborales, sino que también promovió un desarrollo más sostenible y equitativo en las comunidades mineras (Hollnagel et al., 2006). Además, al fortalecer la seguridad laboral, se mejoró la confianza y la moral de los trabajadores, lo que puede llevar a una mayor productividad y estabilidad social en las regiones donde la minería es una actividad económica clave (Hollnagel et al., 1990). En resumen, esta investigación tiene un impacto positivo significativo en la sociedad al abordar un problema crítico de seguridad y bienestar en uno de los sectores laborales más peligrosos.

1.5.5. Importancia de la Investigación

La investigación sobre la aplicación de inteligencia artificial y *machine learning* para la predicción y prevención de accidentes laborales en la minería subterránea del Perú es de suma importancia debido a su potencial para transformar la seguridad en uno de los sectores

industriales más peligrosos. Mediante la implementación de modelos predictivos avanzados, esta investigación no solo proporciona herramientas efectivas para identificar y mitigar riesgos antes de que ocurran, sino que también contribuye a reducir significativamente la tasa de accidentes y mejorar las condiciones laborales de los mineros (Goodfellow et al., 2016). Este enfoque proactivo y basado en datos fortalece la capacidad de respuesta de las empresas mineras, promueve la adopción de mejores prácticas de seguridad y puede ser un modelo replicable en otras industrias de alto riesgo (Hubbard, 2020). Además, al abordar un problema crítico de salud y seguridad laboral, la investigación apoya los objetivos de desarrollo sostenible y bienestar social en las comunidades mineras (Kohavi, 1995). En resumen, la importancia de esta investigación radica en su capacidad para salvar vidas, mejorar la seguridad laboral y contribuir al desarrollo sostenible en el sector minero.

1.5.6. Fundamento de la Investigación

El fundamento de esta investigación se basa en la necesidad urgente de mejorar la seguridad laboral en la minería subterránea del Perú, un sector conocido por su alta incidencia de accidentes graves y fatales. La minería subterránea presenta numerosos desafíos de seguridad debido a factores como desprendimiento de rocas, fallas en equipos, condiciones geológicas adversas y errores humanos. Según la Organización Internacional del Trabajo (OIT), la minería sigue siendo una de las industrias más peligrosas a nivel mundial, lo que subraya la importancia de implementar estrategias efectivas para la gestión de riesgos y la prevención de accidentes (Kohavi, 1995). En este contexto, la aplicación de técnicas de inteligencia artificial (IA) y *machine learning* (ML) ofrece un enfoque innovador y eficaz para abordar estos desafíos. Estas tecnologías permiten analizar grandes volúmenes de datos históricos y actuales para identificar patrones complejos y predecir eventos futuros. Los algoritmos de *machine learning*, como *Random Forest* y *Support Vector Machines* (SVM), son particularmente efectivos para manejar datos complejos y heterogéneos, proporcionando

predicciones precisas y accionables sobre la severidad y frecuencia de accidentes laborales (Goodfellow et al., 2016).

El preprocesamiento de datos, incluyendo técnicas de procesamiento de lenguaje natural (NLP) para limpiar y estructurar datos textuales, es fundamental para asegurar la calidad y utilidad de los datos analizados. La integración de datos de múltiples fuentes, como registros de accidentes, encuestas a trabajadores y reportes de seguridad, proporciona una base de datos rica y exhaustiva para el análisis (Manning, 2008). La simulación de Monte Carlo y otras técnicas estadísticas avanzadas permiten modelar la incertidumbre y evaluar la variabilidad de los posibles resultados, lo que es crucial para una gestión de riesgos efectiva (George et al., 2005). Entonces, la teoría de la resiliencia organizacional también es un pilar clave en esta investigación (Aroquipa & Hurtado, 2022). La capacidad de las organizaciones mineras para anticipar, prepararse y responder a eventos inesperados es crucial para mantener un entorno laboral seguro. Al aplicar modelos predictivos basados en datos, las empresas pueden mejorar significativamente su capacidad de respuesta y mitigación de riesgos, fortaleciendo así su resiliencia organizacional (Hollnagel et al., 2006).

Además, esta investigación tiene importantes implicaciones sociales. Al mejorar la seguridad laboral, se protege la vida y la salud de los trabajadores, se reduce el sufrimiento de sus familias y se promueve un desarrollo más equitativo y sostenible en las comunidades mineras. La reducción de accidentes también puede llevar a una mayor productividad y estabilidad social, beneficiando a toda la comunidad (Reason, 1990). Finalmente, la validación de los instrumentos de recolección de datos mediante pruebas piloto y la revisión de expertos asegura que los datos recopilados sean precisos y fiables. Este rigor metodológico es esencial para obtener resultados válidos y útiles que puedan ser aplicados directamente en las operaciones mineras para mejorar la seguridad y el bienestar de los trabajadores (Devellis & Thorpe, 2021). La presente investigación está fundamentada en la integración de teoría y

tecnología para abordar de manera efectiva y proactiva los desafíos de seguridad en la minería subterránea. Al combinar métodos tradicionales de gestión de riesgos con innovaciones tecnológicas avanzadas, la investigación no solo contribuye al conocimiento científico, sino que también ofrece soluciones prácticas y aplicables para mejorar la seguridad laboral en uno de los sectores más peligrosos de la economía.

1.6. Limitaciones de la Investigación

1.6.1. Espacial

La presente investigación se enfoca en la minería subterránea en el Perú, abarcando diversas minas ubicadas en diferentes regiones del país. Esta limitación espacial se debe a la necesidad de comprender las condiciones específicas de las operaciones mineras en este contexto geográfico particular, caracterizado por variaciones significativas en términos de geología, infraestructura, y prácticas de seguridad (Cruz & Huerta, 2015). Las minas seleccionadas para el estudio representan una muestra diversa de las operaciones mineras subterráneas en el Perú, proporcionando una visión comprensiva de los factores de riesgo y las condiciones laborales en el sector. Sin embargo, los hallazgos y recomendaciones de esta investigación pueden no ser directamente aplicables a otras regiones o países con diferentes contextos operativos y regulatorios.

1.6.2. Temporal

La investigación se limita temporalmente al periodo comprendido entre los años 2007 y 2018. Este rango temporal se seleccionó para permitir un análisis detallado de los datos históricos de accidentes laborales y su evolución a lo largo de más de una década. La selección de este periodo específico responde a la disponibilidad y calidad de los datos proporcionados por la Dirección General de Minería del Ministerio de Energía y Minas del Perú (DGM-MEM), los cuales han sido recopilados de manera sistemática y son representativos de las condiciones

y tendencias actuales en el sector minero (Goodfellow et al., 2016). Esta limitación temporal también permite evaluar el impacto de las políticas y medidas de seguridad implementadas durante este periodo, aunque no refleja necesariamente cambios o mejoras posteriores a 2018.

1.7. Objetivos

- **Objetivo General**

- Desarrollar un modelo predictivo basado en técnicas de inteligencia artificial y *machine learning* para identificar y prevenir accidentes laborales en la minería subterránea del Perú, mejorando la seguridad y las condiciones de trabajo.

- **Objetivos Específicos**

- Analizar los principales factores de riesgo asociados a los accidentes laborales por desprendimiento de rocas en la minería subterránea peruana.
- Evaluar la efectividad del algoritmo de bosques aleatorios (*Random Forest*) en la predicción de la severidad de accidentes laborales en la minería subterránea, comparándolo con otros algoritmos de *machine learning*.
- Optimizar el preprocesamiento de datos textuales y la gestión del desbalance de clases para mejorar la precisión del modelo predictivo de accidentes laborales.
- Determinar los factores críticos que contribuyen significativamente a la ocurrencia y severidad de accidentes laborales en la minería subterránea mediante el análisis de datos con técnicas de *machine learning*.
- Desarrollar estrategias de prevención de accidentes basadas en el análisis de los datos y resultados obtenidos mediante la aplicación de inteligencia artificial en la minería subterránea.

1.8. Hipótesis

- Hipótesis General

- La aplicación de técnicas de inteligencia artificial y *machine learning* mejora significativamente la predicción y prevención de accidentes laborales en la minería subterránea del Perú, contribuyendo a la reducción de la severidad y frecuencia de estos eventos.

- Hipótesis Específicos

- Los principales factores de riesgo asociados a los accidentes laborales por desprendimiento de rocas son identificados con precisión mediante técnicas de *machine learning*.
- El algoritmo de bosques aleatorios (*Random Forest*) proporciona una mayor precisión en la predicción de la severidad de accidentes laborales en la minería subterránea en comparación con otros algoritmos de *machine learning*.
- La mejora en el preprocesamiento de datos textuales y la gestión del desbalance de clases aumenta significativamente la precisión y efectividad del modelo predictivo de accidentes laborales.
- Los factores críticos identificados mediante técnicas de *machine learning* tienen una influencia significativa en la ocurrencia y severidad de los accidentes laborales en la minería subterránea.
- Las estrategias de prevención de accidentes desarrolladas a partir del análisis de datos y resultados obtenidos mediante la aplicación de inteligencia artificial son efectivas en la reducción de accidentes laborales en la minería subterránea.

II. MARCO TEÓRICO

2.1. Base Teórica

El marco teórico de esta investigación aborda la alta peligrosidad de la minería subterránea en el Perú, destacando la necesidad de mejorar la seguridad laboral mediante la gestión de riesgos y la implementación de técnicas avanzadas de inteligencia artificial y *machine learning* (Kang & Ryu, 2019). Los algoritmos de *machine learning*, como *Random Forest* y *Support Vector Machines*, permiten analizar grandes volúmenes de datos para identificar patrones y predecir accidentes, mejorando la capacidad de respuesta y mitigación de riesgos. Además, se utilizan técnicas de procesamiento de lenguaje natural para estructurar datos textuales y complementar el análisis cuantitativo (Abdalzaher et al., 2023). La teoría de la resiliencia organizacional subraya la importancia de que las empresas mineras son capaces de anticipar y adaptarse a eventos inesperados, mientras que la integración de múltiples fuentes de datos y la simulación de Monte Carlo proporcionan una base sólida para la toma de decisiones informadas (Gaudio et al., 2015). Esta investigación no solo tiene implicaciones prácticas significativas para la mejora de la seguridad laboral, sino que también contribuye al desarrollo sostenible y equitativo de las comunidades mineras. Las dimensiones de la investigación son:

- **Seguridad Laboral en la Minería Subterránea.**

La minería subterránea es una de las industrias más peligrosas a nivel mundial debido a las condiciones extremas y los riesgos inherentes a la extracción de minerales bajo la superficie terrestre. Los trabajadores están expuestos a múltiples peligros, incluyendo desprendimientos de rocas, fallas en equipos, condiciones geológicas adversas, y errores humanos. Estos riesgos se agravan en contextos geográficos específicos, como las diversas regiones mineras del Perú, donde las variaciones en la geología y la infraestructura influyen

significativamente en las prácticas de seguridad y la frecuencia de accidentes laborales (Ahmad et al., 2020). Ver la Figura 1.

Figura 1

Seguridad Laboral en Minería Subterránea



Nota. Seguridad laboral en minería subterránea.

- **Gestión de Riesgos y Seguridad Laboral**

La gestión de riesgos en la minería subterránea es crucial para minimizar los accidentes y proteger la vida de los trabajadores. Según la teoría de la gestión de riesgos, es esencial identificar, evaluar y mitigar los riesgos presentes en el entorno laboral para garantizar un ambiente de trabajo seguro. Los métodos tradicionales de gestión de riesgos incluyen la identificación de peligros, la evaluación de riesgos y la implementación de medidas de control (Ameida et al., 2016). Sin embargo, estos métodos a menudo son insuficientes para abordar la

complejidad y la dinámica de los riesgos en la minería subterránea, lo que resalta la necesidad de enfoques más avanzados y tecnológicamente integrados.

- **Aplicación de Inteligencia Artificial y Machine Learning**

La inteligencia artificial (IA) y el *machine learning* (ML) han emergido como herramientas poderosas para mejorar la gestión de riesgos en la minería subterránea. Estas tecnologías permiten analizar grandes volúmenes de datos históricos y actuales para identificar patrones complejos y predecir eventos futuros. Los algoritmos de ML, como *Random Forest* y *Support Vector Machines* (SVM), son especialmente efectivos para manejar datos complejos y heterogéneos, proporcionando predicciones precisas y accionables sobre la severidad y frecuencia de accidentes laborales (Goodfellow et al., 2016). La capacidad de estos algoritmos para aprender y adaptarse a nuevos datos permite a las empresas mineras desarrollar modelos predictivos robustos que pueden mejorar significativamente la seguridad laboral.

- **Preprocesamiento y Análisis de Datos Textuales**

El preprocesamiento de datos es un paso crítico para asegurar la calidad y la utilidad de los datos analizados. En la minería subterránea, los datos pueden provenir de diversas fuentes, incluyendo registros de accidentes, encuestas a trabajadores y reportes de seguridad. Las técnicas de procesamiento de lenguaje natural (NLP) son esenciales para limpiar y estructurar datos textuales, permitiendo su integración en los modelos de ML (Manning, 2008). El análisis de datos textuales, como la identificación de palabras clave y el modelado de tópicos, proporciona información valiosa que complementa los datos estructurados, ofreciendo una visión más completa de los factores de riesgo y las causas subyacentes de los accidentes.

- **Resiliencia Organizacional en la Minería**

La teoría de la resiliencia organizacional es fundamental en el contexto de la minería subterránea. La resiliencia organizacional se refiere a la capacidad de una organización para

anticipar, prepararse, responder y adaptarse a eventos inesperados y perturbadores (Hollnagel et al., 2006). En la minería subterránea, esta capacidad es crucial para mantener un entorno laboral seguro y eficiente. Los modelos predictivos basados en IA y ML no solo permiten a las empresas mineras identificar proactivamente los riesgos, sino que también fortalecen su capacidad para responder rápidamente a emergencias y mitigar los impactos de los accidentes.

- **Contribuciones de la Investigación a la Seguridad Laboral**

Esta investigación ofrece importantes contribuciones teóricas y prácticas a la seguridad laboral en la minería subterránea. Al combinar métodos tradicionales de gestión de riesgos con innovaciones tecnológicas avanzadas, proporciona una metodología integral para mejorar la seguridad laboral. La validación de instrumentos de recolección de datos mediante pruebas piloto y la revisión de expertos asegura que los datos recopilados sean precisos y fiables, lo cual es esencial para obtener resultados válidos y útiles que puedan ser implementados directamente en las operaciones mineras (Devellis & Thorpe, 2021).

- **Impacto Social y Económico de la Seguridad en la Minería**

La mejora de la seguridad laboral en la minería subterránea tiene un impacto significativo tanto a nivel social como económico. La reducción de accidentes no solo protege la vida y la salud de los trabajadores, sino que también disminuye el sufrimiento de sus familias y promueve un desarrollo más equitativo y sostenible en las comunidades mineras (Ilo, 2015). Además, un entorno laboral más seguro puede llevar a una mayor productividad y estabilidad social, beneficiando a toda la comunidad (Reason, 1990).

- **Integración de Datos y Simulación de Monte Carlo**

La integración de múltiples fuentes de datos es crucial para proporcionar una base sólida para el análisis. La simulación de Monte Carlo y otras técnicas estadísticas avanzadas permiten modelar la incertidumbre y evaluar la variabilidad de los posibles resultados de

accidentes (Yoshua, 2019). Esto es crucial para una gestión de riesgos efectiva, ya que proporciona una comprensión detallada de los posibles escenarios y sus probabilidades, permitiendo a las empresas mineras tomar decisiones informadas y desarrollar estrategias de mitigación más efectivas.

2.2. Causas de los Accidentes Laborales

La comprensión de la génesis de los accidentes de trabajo es importante para el desarrollo de prácticas de prevención. Los programas de entrenamiento, concientización, orientación y recomendación de mayor cuidado por parte del trabajador tanto para el uso de medidas de protección colectiva como individual carecen de sentido, si no se identifican los factores causales de los accidentes dentro de una organización. Asimismo, a pesar de que cada incidente tiene sus características propias, algunas similitudes son observadas, destacándose la falta de atención del trabajador en la tarea encomendada, la falta de cumplimiento de procedimientos operacionales, el mal estado de equipos y herramientas, la deficiencia de supervisión, entre otros. (Bajpayee et al., 2004) señalan tres factores fundamentales para la ocurrencia de accidentes: personales, ambientales y de operación. Entre los **factores personales** se pueden señalar la educación, el entrenamiento y la experiencia previa.

Los factores ambientales se refieren al ambiente de trabajo determinado por los líderes de la organización a través del correcto planeamiento, organización y control de las tareas comprendidas dentro del proceso productivo y finalmente los factores de operación se refieren a la forma como las tareas específicas son ejecutadas. Paul & Maiti (2007) indican que las principales causas de accidentes de trabajo son las condiciones y los actos inseguros. Las condiciones inseguras en minas se manifiestan a través de proyectos inadecuados, incertezas de carácter geológico - estructural, condiciones inadecuadas de mantenimiento de equipos, supervisión inadecuada entre otros. Los actos inseguros se manifiestan principalmente a través

de actitudes de comportamiento inadecuadas por parte de los trabajadores y que según (Bhattacharjee, 1991) constituyen directa o indirectamente el 90% de las causas de los accidentes de trabajo.

Algunos investigadores señalan la creciente importancia del papel de los factores organizacionales como determinantes para el acontecimiento de accidentes de trabajo. (Wright, 1986) investigó la importancia de los factores organizacionales en la génesis de la secuencia de situaciones de accidentes, identificando los procesos peligrosos, la presión para el incremento de productividad y la falta de comunicación como los mayores determinantes para la ocurrencia de un evento indeseado. Abdul & Mohd (2017) señalan que los accidentes son determinados por factores humanos, organizacionales y tecnológicos. Empey (1992) indica como factores clave en el desarrollo de seguridad dentro de una organización es el balance entre la seguridad y la producción y la creación de una adecuada cultura de seguridad organizacional. Simard & Marchand (1995) investigaron la influencia de varios factores organizacionales en la adopción de iniciativas de seguridad, concluyendo que la comprensión del proceso productivo y sus riesgos asociados, la cohesión y cooperación, así como la experiencia de los trabajadores y una adecuada gestión de seguridad, son factores que estimulan la adopción de medidas de seguridad por parte de los trabajadores. Almeida (1995) indica que las teorías de causalidad de accidentes pueden ser agrupados en los siguientes modelos conceptuales.

- Cadena de múltiples eventos que describe una secuencia temporal que conduce al accidente, entendiéndose como de origen multicausal;
- Modelo epidemiológico, que presenta el accidente como el resultado de la interacción entre las variables de la persona, del agente y del ambiente de trabajo.

- Modelo de intercambio de energía que indica que los accidentes son causados por alguna forma de energía, que es el agente;
- Modelo comportamental que comprende dos categorías.

La primera valoriza la existencia de situaciones en las cuales debe de tomarse una decisión para la acción en la presencia de un riesgo y la segunda que defiende la existencia del modelo comportamental de propensión al accidente, según el cual algunas personas tienen una predisposición innata a la accidentabilidad; y Modelo sistémico que aborda el accidente como el resultado extremo en el sistema hombre-máquina, resaltando la interacción entre sus componentes. Situaciones de sobrecarga y desvíos en el sistema pueden conducir al rompimiento del equilibrio, consecuentemente a la ocurrencia del accidente. Es evidente que en la ocurrencia de un accidente confluyen, la mayoría de las veces, múltiples factores, así su análisis involucra varios parámetros; y también es claro que los efectos de sinergismo en razón de su naturaleza asociada a diferentes niveles de riesgo en los locales de trabajo deben de ser llevados en consideración. El análisis debe, por tanto, identificar los factores presentes en el origen del peligro, los factores que desencadenan o liberan aquel tipo de peligro en potencial, así como las condiciones del sistema involucrado. De acuerdo a Candia et al. (2010) los accidentes en minas pueden ser causados principalmente por:

- Electricidad;
- Espacios confinados;
- Explosivos;
- Explosión de ductos;
- Deslizamiento de terreno;
- Caída de rocas en excavaciones;
- Incendios;

- Manipulación de materiales y herramientas;
- Tránsito de personas y vehículos entre otros.

2.3. Accidentes Laborales

La minería es una de las actividades económicas más importantes que genera beneficios económicos directos e indirectos para las zonas que se encuentran en su jurisdicción. Adicionalmente, se tienen reportes de beneficios derivados de las regalías (canon minero) que aportan dichas empresas a pesar de que no siempre se usen adecuadamente. Sin embargo, al ser una actividad laboral catalogada como de alto riesgo, esta actividad genera accidentes fatales mayor al de otras actividades sobre todo en la minería no formal. Algunos estudios internacionales han reportados patologías diversas de los trabajadores en este campo laboral, como aquellas que se generan producto del ruido excesivo, otras asociadas a la exposición de vibración de los equipos de trabajo que utilizan y también otras múltiples patologías y/o accidentes laborales. En nuestro medio existe escasa investigación al respecto, sobre todo de las consecuencias fatales que acontecen en esta actividad económica. Es importante el determinar las características que tienen las notificaciones de los accidentes mortales en la minería de nuestro país; ya que esto permitiría caracterizar la problemática específica de una realidad y generar políticas que permitan modificar el estado actual de esta situación (Mejía et al., 2015).

- Reducción de Accidentes Laborales.

Para la reducción de los accidentes se implantará las acciones del uso adecuado de las herramientas de gestión como son:

- Políticas
- Estándares (qué hacer)

- Procedimientos (cómo hacerlo)
 - Identificación de Peligros Evaluación y Control de Riesgos (IPERC)
 - Inspecciones (diarias, semanales, mensuales)
 - Auditorias (trimestrales, semanales, anuales)
 - Check List (chequeos personalizados)
 - Capacitaciones y Retroalimentación
- **Cómo Medir el Avance del Desempeño en las Operaciones**

La determinación del nivel está basada en dos aspectos que sirven como criterios de graduación, es decir el esfuerzo físico SSOMAC y la experiencia SSOMAC, ambos criterios son reconocidos internacionalmente y se emplean para medir la calidad de un Sistema Integrado de Gestión de Riesgos. Se miden con:

- Índices de frecuencia.
- Índices de severidad.
- Índices de accidentabilidad.

La valoración de la experiencia significa determinar las lesiones causadas a personas durante o como resultado de las actividades en las operaciones. La coincidencia de daños y enfermedades ocupacionales incapacitantes indicará el grado de eficacia de la Gestión de Riesgos existente en la empresa. Ver la Tabla 1.

Tabla 1*Experiencia del SSOMAC*

NIVEL	EXPERIENCIA SSOMAC	ESFUERZO SSOMAC (%)
Las mejores del mundo	≤ 1	≥ 90
Excelente	≤ 2	≥ 80
Muy bueno	≤ 3	≥ 70
Bueno	≤ 4	≥ 60
Promedio	≤ 5	≥ 50
Regular	≤ 6	≥ 40
Inicial	≤ 6.1	$\geq 0 - 30$

Nota. SSOMAC=Seguridad, Salud Ocupacional, Medio Ambiente y Calidad.

La valoración de los esfuerzos y experiencias SSOMAC de una Empresa, siempre están referidos a los doce meses anteriores a la auditoria, este periodo no es necesariamente un año calendario. Los esfuerzos son medidos vía los cumplimientos físicos, visibles y objetivos.

Determinación de puntaje en una auditoria:

- 10 puntos = No necesita mejorar
- 8 – 9 puntos = Necesita mejoras mínimas
- 6 – 7 puntos = Necesita algunas mejoras
- 4 – 5 puntos = Necesita mejoras
- 1 – 3 puntos = Necesita mejoras significativas
- 0 puntos = Necesita mejoras mayores

El control adecuado de los riesgos mejora el bienestar de los trabajadores, las operaciones, la calidad, la productividad, la competitividad, las utilidades y la imagen de la

gerencia superior ante accionistas, la comunidad y el mercado financiero (Calderón Solís, 2012).

2.4. Excavaciones Mineras Subterráneas

Las excavaciones subterráneas son construcciones en el seno del macizo rocoso, con distintos fines que pueden agruparse en dos: aprovechamiento del material arrancado y generación de espacio para diversos usos. En cualquiera de los dos casos, es importante tener en cuenta la eficiencia y seguridad durante y después de la excavación, básicamente aplicando el sistema de sostenimiento adecuado y duradero. En el primer caso, el propósito de la excavación es utilizar el material arrancado, sea este mineral metálico (minerales de plomo, zinc, cobre, oro, plata, etc.) o no metálico (carbón, gemas, rocas ornamentales, etc.), como son los tajeos mineros para explotación de los recursos minerales. En la generación de espacios abiertos, no interesa el material extraído, sino los espacios abiertos que dejan las excavaciones con distintos fines o usos, como: mineros (galerías, rampas, piques, cámaras, entre otros) y construcciones civiles (túneles, cavernas, chimeneas, entre otros) (Giraldo & Badillo, 2015).

Resulta impresionante la gran perseverancia y el desprecio por el riesgo que mostró el hombre desde los orígenes de la historia, en sus intentos de realizar excavaciones; valiéndose inicialmente solo de sus propias manos y la fuerza bruta, paso a paso fue confeccionando herramientas, rudimentarios martillos, picos y cinceles, si a esta precaria situación de falta de utensilios de trabajo, le añadimos los elementos que se emplean en la entibación y la ausencia de sistemas de ventilación, comprobamos que el laboreo de túneles y galerías, implicaba en la antigüedad una enorme, formidable y sacrificada labor. Los logros obtenidos pese a las carencias y dificultades muestran lo que el género humano es capaz de lograr cuando su mente está dispuesta a ello. No ha de ignorarse que la utilización masiva de esclavos, sometidos a condiciones inhumanas y cuya supervivencia no importaba, fue una de las claves en el laboreo

de excavaciones en la antigüedad. La utilización de los espacios subterráneos tiene sus inicios con el propio surgimiento del hombre, cuando este de manera consciente comenzó a utilizar las cuevas y cavernas como refugios, para protegerse de las lluvias, tormentas y de los animales, luego las utilizó como viviendas. Durante el proceso de utilización de estos espacios el hombre sintió la necesidad de cambiar sus condiciones naturales (forma y dimensiones) todo esto lo fue llevando de manera paulatina a que él mismo fuera perfeccionando las herramientas y métodos de arranque de las rocas, comenzando con el empleo de los instrumentos más rudimentarios hasta llegar a la utilización de equipos de alta productividad. El desarrollo actual y el uso cada vez más frecuente de las excavaciones subterráneas por parte del hombre, para la extracción de recursos minerales, para el paso de vehículos, para redes ferroviarias, para el abastecimiento de agua, como almacenes, para la protección de las personas y otros fines, hace que este se dedique de manera consciente e integral a mejorar los parámetros que caracterizan a estas obras, conjuntamente con los índices técnico – económicos que influyen de una forma u otra en el desarrollo eficiente de dicha actividad. El primer método de laboreo de galerías de minas y luego de túneles, fue la técnica del fuego; la cual fue introducida por primera vez por los antiguos egipcios, los que además de la fuerza bruta aplicaron la ciencia, con la perspectiva de mejorar la eficiencia en la perforación de las rocas. Este método consiste en provocar un incendio en el frente de trabajo y luego sofocarlo bruscamente con agua fría, (el uso del vinagre no deja de pertenecer al dominio del mito), produciendo un brusco gradiente térmico que da lugar al resquebrajamiento de la roca. Pero esta técnica también provoca, como no es difícil imaginar, una atmósfera viciada, irrespirable, generando a menudo gases venenosos, lo que convierte al trabajo del minero en una trampa mortal, a la que solo unos pocos afortunados sobreviven. La construcción de excavaciones subterráneas se divide en dos ramas: Las laboreadas en rocas fuertes y las hechas en rocas débiles. El principal objetivo del laboreo en rocas fuertes es debilitar el macizo rocoso mediante su fractura, tradicionalmente en la

construcción de túneles y obras subterráneas, en este tipo de rocas, el principal problema a resolver por el ingeniero era el arranque, porque en la mayoría de los casos la excavación no precisaba de ningún tipo de sostenimiento. En nuestro país existe un gran número de obras subterráneas, las cuales han sido laboreadas sin llevar a cabo la correcta elección del método de arranque de las rocas, y más aún sin un previo conocimiento de los principales factores que influyen en este proceso, ni de las características reales del macizo rocoso, haciéndose “a priori” el arranque de las rocas con trabajos de voladura (Noa, 2006).

En la actualidad, las excavaciones mineras subterráneas se realizan utilizando modernos equipos y la tendencia general de la actualidad es realizar las minas a cielo abierto ya que los avances producidos en la maquinaria minera y en la investigación de yacimientos permite la explotación a cielo abierto de muchos yacimientos que en otras épocas no era posible por no ser rentables económicamente, pero un límite económico y de otros factores técnicos a partir del cual una explotación debe seguir continuándose a través de la minería subterránea debido a que, es mucho más rentable continuar por minería de interior.

2.4.1. Sostenimiento del Macizo Rcoso

El macizo rocoso puede auto soportarse por un periodo de tiempo de acuerdo a su clasificación geomecánica, desde nulo hasta completamente autosoportante. De acuerdo a ello, se debe instalar oportunamente los elementos de sostenimiento más apropiados y así evitar desprendimientos y/o colapsos. Los desprendimientos ocurren cuando se deja expuesta un área excavada más allá del tiempo de autosoporte. Asimismo, cuando no se han utilizado o dimensionado adecuadamente los elementos de sostenimiento más apropiados. La Figura 2 ilustra el desprendimiento de rocas por el uso de sostenimiento inadecuado y falta de reemplazo oportuno de los mismos. Los elementos de sostenimiento son materiales o herramientas diseñados para evitar el desprendimiento de rocas o la subsidencia de la cobertura de una excavación en el seno de un macizo rocoso. Con el correr del tiempo, se ha venido

introduciendo a la industria una variedad de elementos de sostenimiento, siendo los primeros elementos de sostenimiento la roca y la madera. Entre los elementos de sostenimiento de uso actual se tienen: arcos de acero, pernos de anclaje, cables de acero, mallas de acero, concreto lanzado (shotcrete), concreto armado, concreto prefabricado (dovelas), relleno hidráulico o en pasta. (Giraldo, 2016).

Figura 2

Desprendimiento de Rocas por Falta de Mantenimiento de la Madera



Nota. La figura es extraída de: (Giraldo, 2016).

2.4.2. Sistema de Identificación de Factores

El objetivo del presente estudio es identificar factores para reducir accidentes en las labores mineras u otras excavaciones subterráneas, ocasionados por desprendimiento de rocas. Para alcanzar tal propósito, se investigó los accidentes mortales ocurridos de 2007 a 2018, en las fuentes primigenias, entre otras, MINEM, Ministerio de Trabajo y Promoción del Empleo y OSINERGMIN. A partir de la información compilada y procesada, se realizará el análisis correspondiente mediante la identificación de factores, así se identificarán los siguientes FACTORES:

- Tipo de roca

- Tipo de sostenimiento
- Seguridad laboral
 - Uso de implementos de seguridad
 - Tiempo de servicio
 - Ocupación de la víctima
- El turno de trabajo (Giraldo, 2016).

En toda industria de excavación en roca, existe una preocupación latente y permanente de evitar los accidentes. Sin embargo, a ciencia cierta no se conocen los factores que tienen mayor incidencia. Con tal propósito, se vienen investigando estos factores donde se ha considerado por conveniente hacer un análisis de las repercusiones tanto técnicas como económicas de los accidentes que son tan complejas en la industria minera. (Giraldo & Badillo, 2015).

2.5. La Industria 4.0 y el Impulso de la Fábrica Inteligente

Con la llegada de la Industria 4.0 y el impulso del concepto de fábrica inteligente, tecnologías como el big data, la inteligencia artificial (IA) y la Internet de las cosas (IdC) han sido ampliamente adoptadas en la industria manufacturera, incrementando la eficiencia productiva pero también introduciendo nuevas incertidumbres y riesgos para la seguridad y salud en el trabajo (SST) (Song & Yang, 2021). Según la Organización Internacional del Trabajo (OIT), anualmente más de 2,78 millones de trabajadores fallecen debido a accidentes laborales y otros 374 millones sufren accidentes no mortales, con la industria de la construcción representando aproximadamente el 60% de estas muertes (Wright, 1986). Por tanto, es crucial reconocer la importancia de la SST no solo en la minería, el transporte y la industria química, sino también prestar mayor atención al sector de la construcción.

Aprender de accidentes pasados es esencial para prevenir futuros incidentes con sustancias peligrosas similares, y la extracción de conocimientos de textos de accidentes en SST se ha convertido en un paso vital para la gestión de riesgos en proyectos de construcción. No obstante, los análisis tradicionales basados en encuestas y entrevistas a expertos presentan limitaciones como falta de objetividad y baja capacidad interpretativa, lo que subraya la necesidad urgente de establecer nuevas fuentes de datos multifacéticas (Xu et al., 2021). Tecnologías emergentes como la visión por computadora, los vehículos aéreos no tripulados (UAV) y la realidad virtual (RV) ofrecen un soporte crucial para la recopilación, procesamiento y almacenamiento de grandes volúmenes de datos heterogéneos, permitiendo organizar y aplicar datos estructurados, no estructurados y en tiempo real para la toma de decisiones internas, y expandiendo las estrategias de evaluación y gestión de riesgos en la producción ocupacional. Por ejemplo, la identificación de peligros para los trabajadores de la construcción mediante la medición de la variabilidad de la marcha con un sistema de plantillas de presión portátiles (Antwi et al., 2020). También se utilizan unidades de medición inercial portátiles (WIMU) para recopilar datos sobre patrones de marcha anormales, combinados con datos de detección de posición para identificar peligros de resbalones, tropiezos y caídas (Yang & Ahn, 2019). Los dispositivos portátiles de seguimiento ocular recolectan datos de escaneo para identificar comportamientos de búsqueda de riesgos psicológicos en la industria de la construcción (Xu et al., 2021). Los datos textuales no estructurados sobre accidentes han emergido como una fuente vital para la gestión de riesgos en la construcción, complementando los datos estructurados y proporcionando un canal eficiente para la minería de conocimientos en SST.

Diversas áreas de investigación sobre accidentes laborales, como el monitoreo de peligros y la evaluación de riesgos, son temas populares. En este contexto de investigación, se han identificado factores de riesgo de accidentes laborales, simulado escenarios de peligro,

predicho la frecuencia y gravedad de incidentes, y desarrollado contramedidas basadas en reglas de decisión (Sarkar & Maiti, 2020). La predicción de la gravedad de las lesiones mejora el desempeño en seguridad de los proyectos y sugiere medidas proactivas para prevenir accidentes. Numerosos estudios han explorado la predicción de la gravedad de las lesiones en diferentes dominios, reconociendo sus ventajas. Lu, (2022) utilizó datos del Sistema Global Integrado de Información de Envíos (GISIS) para calcular la probabilidad de gravedad en rutas marítimas mediante el algoritmo de bosque aleatorio (RF) y cuantificó la gravedad con puntuaciones de expertos. Chebila, (2021) aplicó el Sistema de Notificación de Accidentes Graves (MARS) para predecir las consecuencias de accidentes con productos químicos peligrosos usando múltiples algoritmos de *machine learning*. Baker et al., (2020) extrajeron métodos de construcción y condiciones ambientales de informes de accidentes mediante reglas lógicas y diccionarios personalizados, prediciendo la gravedad de las lesiones con algoritmos SVM y XGBoost. En la industria de la construcción, que es particularmente peligrosa, la investigación de riesgos de SST juega un papel crucial en la reducción de víctimas y daños a la propiedad (Goh et al., 2018). Los accidentes por derrumbe, en particular, causan muertes y lesiones masivas, siendo esenciales para investigar la relación entre factores y gravedad de las lesiones. Aunque la predicción de la gravedad de las lesiones es vital para prevenir accidentes, pocos estudios han utilizado datos textuales para evaluar riesgos y optimizar modelos en accidentes por derrumbe en la construcción.

El objetivo principal de este estudio es mejorar el preprocesamiento de datos textuales y optimizar el modelo de evaluación de riesgos para establecer un marco sistemático que permita predecir la gravedad de los accidentes laborales. El preprocesamiento de datos se centra en abordar problemas de precisión en la segmentación de palabras, desequilibrio de categorías y multicolinealidad en el texto. La optimización del modelo de evaluación de riesgos busca confirmar la validez del modelo y obtener factores de atributos críticos basados en

indicadores de rendimiento en tres escenarios. Este estudio contribuye a la literatura existente al: (i) Utilizar técnicas de extracción de frases basadas en información mutua y entropía para lograr una segmentación de texto precisa. (ii) Incluir sobremuestreo y medición de coeficientes de correlación para un procesamiento de datos más exhaustivo. (iii) Simular el rendimiento de predicción de la gravedad del accidente en tres escenarios utilizando el algoritmo RF. (iv) Analizar los factores de atributos críticos de los accidentes laborales mediante la clasificación de importancia de las características.

2.5.1. Seguridad y Salud en el Trabajo (SST)

El término "SST" (Seguridad y Salud en el Trabajo) se refiere a las medidas técnicas y organizativas implementadas para asegurar la salud y seguridad de los empleados en su entorno laboral, con el objetivo de minimizar las pérdidas y daños derivados de sus actividades laborales (Adaku et al., 2021). Las preocupaciones sobre seguridad y salud ocupacional en los procesos industriales modernos están ganando atención globalmente. Por ejemplo, Beckert & Barros (2022) investigaron los nuevos riesgos para la salud ocupacional de los trabajadores de recolección de residuos sólidos en São Paulo, Brasil, durante la pandemia de COVID-19, revelando que estos trabajadores enfrentaban una mayor vulnerabilidad a la infección. Guzman et al. (2022) aplicaron el modelado de ecuaciones estructurales de mínimos cuadrados parciales (PLS-SEM) para simular la SST en la industria del petróleo y gas durante la pandemia de COVID-19, integrando cinco dominios: riesgos de exposición directa en el lugar de trabajo, políticas y procedimientos, percepción de la cultura de auto seguridad, percepción de la responsabilidad de auto seguridad y medidas preventivas. Este enfoque verificó el efecto positivo de la percepción de la cultura de seguridad en el avance de la SST. Liu et al. (2021) desarrollaron un nuevo modelo de evaluación de riesgos de SST que combina conjuntos difusos de imágenes (PFS) y métodos de colas alternativas (AQM) para evaluar y clasificar los factores de riesgo de peligros laborales, demostrando la viabilidad y superioridad del modelo mediante

un caso real de excavación en un sitio de construcción. Además, estudios sobre cuestiones de SST en industrias como la minería (Aroquipa, 2014), la textil (Mutlu & Altuntas, 2019) y la piscicultura (Thorvaldsen et al., 2020) pueden proporcionar una comprensión más profunda de los factores de riesgo y las lesiones ocupacionales resultantes, facilitando intervenciones efectivas para mejorar la seguridad y salud de los empleados.

2.5.2. *Análisis de la Gravedad de los Accidentes por Datos Textuales*

En la investigación de accidentes laborales, el análisis causal de los factores de riesgo y la evaluación de riesgos siempre han sido temas de gran relevancia. Con el avance de las técnicas de inteligencia artificial y una comprensión más profunda de los datos textuales, ricos en información valiosa, se ha observado un aumento gradual en los estudios que utilizan métodos cuantitativos para predecir la probabilidad y gravedad de los accidentes laborales. Por ejemplo, Likun & Zaili, (2018) recopilaron informes de investigación de accidentes en vías navegables de la Administración de Seguridad Marítima (MSA) de China, extrajeron factores característicos de los casos textuales basados en la teoría fundamentada y luego desarrollaron un modelo de evaluación de riesgos con redes bayesianas (BN) para identificar las dependencias entre los factores de riesgo clave y la gravedad de los accidentes en vías navegables. Likun & Zaili (2018) empleó registros de inspección de seguridad, casos de accidentes y datos relacionados con proyectos obtenidos de un gran contratista como indicadores líderes de seguridad, utilizando cinco algoritmos de *machine learning* populares para predecir la ocurrencia y la gravedad de los accidentes de construcción, siguiendo el marco Cross-Industry Standard Process for Data Mining (CRISP-DM) reconocido por la industria. Stemn & Krampah, (2022) investigaron la asociación entre los factores de accidentes y los diferentes niveles de gravedad de las lesiones en la industria minera a cielo abierto utilizando un análisis de correspondencias basado en los informes de accidentes recopilados por el Departamento de Inspección de la Comisión de Minería (IDMC), lo que ayudó a obtener una

comprensión más profunda de las causas subyacentes y fundamentales de los accidentes mineros. Tamascelli et al. (2022) categorizaron los atributos de los accidentes extraídos en las correspondientes categorías de gravedad según los informes textuales de la base de datos del Servicio de Datos de Accidentes Peligrosos Mayores, y el modelo Wide&Deep construido demostró un rendimiento de predicción superior.

2.5.3. Uso de Técnicas de Minería de Texto y Aprendizaje Automático

La introducción de la tecnología de IA en los flujos de trabajo puede potenciar sistemas de apoyo a la toma de decisiones basados en datos, los cuales capturan el valor potencial contenido en los datos y promueven la transformación estructural a través de la simulación de escenarios complejos, resultando en numerosas investigaciones innovadoras en el análisis de accidentes laborales en la construcción. En aplicaciones de análisis de contenido y recuperación de información textual de accidentes no estructurados, Baker et al. (2020) utilizaron procesamiento del lenguaje natural (PLN) para extraer información de atributos de informes de lesiones no estructurados y desarrollaron un sistema capaz de procesar automáticamente estos informes, eliminando la necesidad de análisis manual. Zou et al. (2017) emplearon modelos de espacio vectorial y expansión de consultas semánticas para construir un sistema de recuperación de casos de riesgo en proyectos de construcción, mejorando la eficiencia y rendimiento en la recuperación de casos similares. Parinaz (2021) usó PLN y cuatro algoritmos de *machine learning* (*Naive Bayes*, Regresión Logística, *Random Forest* y Extreme Gradient Boosting) para crear un sistema de identificación automática para informes de contratos de construcción, que recupera eficientemente el texto del contrato requerido y predice el tiempo y costo asociados con la preparación del informe, facilitando una revisión y un proceso de informe optimizados para los tomadores de decisiones. En cuanto a la identificación de riesgos mediante textos de accidentes, Zhang (2022) propuso un método de minería de texto en dos etapas para clasificar causas de accidentes de construcción y evaluó cinco clasificadores de ML

para esta tarea, construyendo un modelo de incrustación de palabras específico del corpus. Además, Zhang (2022) maximizaron la información potencial en datos heterogéneos de proyectos de construcción, desarrollando un marco de análisis integrado (IAF) para identificar y evaluar factores de riesgo utilizando métodos de ML y el concepto de efectos en cascada. En la predicción de accidentes con texto no estructurado, (Hingorani et al., 2020) desarrollaron un modelo para predecir el nivel de confiabilidad de componentes estructurales basado en el tipo y uso de los edificios, proponiendo una distinción cuantitativa del nivel de consecuencias de fallas según el número de víctimas y el grado de daño, aplicable al análisis de riesgos potenciales en estructuras específicas. Kang & Ryu (2019) utilizaron un modelo RF para predecir tipos de accidentes laborales, construyendo un modelo predictivo con datos meteorológicos y de accidentes de construcción de Corea, identificando factores de riesgo clave para diferentes tipos de accidentes.

El análisis de la literatura evidencia que la investigación sistemática sobre la gravedad de los accidentes laborales ha producido numerosos hallazgos exitosos en el contexto de la atención continua a la seguridad y salud ocupacional. Sin embargo, relativamente pocos estudios han combinado datos textuales y algoritmos de aprendizaje automático para predecir la gravedad de las lesiones en accidentes de construcción. Por lo tanto, el objetivo de este trabajo es mejorar el rendimiento de la minería de texto y los algoritmos de aprendizaje automático para estimar la gravedad de los accidentes por colapso en la construcción e identificar eficientemente los factores críticos de los atributos de los accidentes, apoyando así el mecanismo de prevención de accidentes y el control de riesgos en la industria de la construcción.

2.5.4. Preprocesamiento de Datos

2.5.4.1. Segmentación de Texto.

La primera fase del procesamiento de texto es la segmentación, que constituye un módulo fundamental para la extracción de información. Esta etapa no solo ayuda a las computadoras a comprender y procesar el lenguaje natural de manera más efectiva, sino que también influye significativamente en la efectividad del análisis y la ejecución de las tareas subsiguientes de procesamiento de lenguaje natural (PLN). Los enfoques actuales para la segmentación de palabras textuales se dividen principalmente en segmentación mecánica basada en el conocimiento y segmentación estadística basada en datos (Zhang, 2022). El algoritmo de segmentación mecánica, también conocido como algoritmo de coincidencia de cadenas, compara las cadenas con diccionarios prefabricados siguiendo una estrategia específica para mejorar la eficiencia de recuperación utilizando la información almacenada. Por otro lado, el algoritmo de segmentación estadística emplea la frecuencia de palabras adyacentes y el contexto léxico para etiquetar las palabras. Cuando la frecuencia de combinación de palabras adyacentes es alta en el corpus estadístico, se determina que estas palabras forman frases. Este método basado en estadísticas de frecuencia de palabras es cada vez más popular debido a su mayor credibilidad.

El concepto de etiquetado de secuencia utilizado en el diseño y ejecución del algoritmo de segmentación estadística basado en el modelo de Markov oculto (HMM) es especialmente valioso en la investigación de análisis léxico. Entre las herramientas de segmentación de palabras más comunes se encuentran Jieba, SnowNLP, NLTK y HanLP. Jieba, en particular, destaca por su simplicidad y precisión, ofreciendo tres modos de segmentación (modo exacto, modo completo y modo de motor de búsqueda), así como funciones avanzadas como la extracción de palabras clave y la consulta de posición de palabras. Para este artículo, se utilizó inicialmente la herramienta Jieba con métodos de segmentación estadística basados en HMM.

Luego, para evitar las limitaciones de los diccionarios específicos de dominio y la ineficacia de las estrategias de coincidencia basadas en diccionarios (Yan et al., 2021), se empleó una combinación de información mutua y entropía de información para analizar la coocurrencia y la expansión semántica de las palabras candidatas. Este enfoque identifica automáticamente las separaciones de palabras más valiosas y completas, mejorando la precisión y eficiencia del procesamiento de la información textual.

- Información mutua.

La información mutua es una medida en la teoría de la información que representa la correlación entre dos conjuntos de eventos. En el procesamiento de segmentación de texto, se utiliza para medir el grado de interdependencia entre palabras (Bai & Labi, 2009). En el análisis de texto, la fórmula para cuantificar la información mutua se expresa como:

$$MI\{x, y\} = \log \left(\frac{p(x, y)}{p(x) \times p(y)} \right) \quad \dots (1)$$

Dónde $p(x)$ representa la probabilidad del carácter x , $p(y)$ representa la probabilidad del carácter y que aparece en el texto, $p(x, y)$ representa la probabilidad del carácter x, y . El valor de la $MI\{x, y\}$ indica el grado de correlación entre x, y . Cuando $MI\{x, y\} > 0$, esto significa que x, y están relacionados, y cuanto mayor sea el valor de MI, mayor será el grado de correlación. $MI\{x, y\} = 0$, esto significa que x, y son independientes entre sí.

- Entropía de información.

La entropía de la información se utiliza para describir el grado de confusión de la información, también conocido como grado de incertidumbre, y cuanto mayor sea la entropía de la información, mayor será el grado de incertidumbre de la información. El concepto de entropía de la información se puede aplicar para medir la cantidad de información que pueden contener los eventos desconocidos, lo que desempeña un papel importante en el

almacenamiento y la transmisión de datos en el campo de la PNL (Leeanansaksiri et al., 2018). En el procesamiento de texto, los valores de la entropía de información izquierda y derecha de una cadena reflejan la incertidumbre del carácter y sus caracteres adyacentes izquierdo y derecho. Cuanto mayor sea la incertidumbre, significa que los caracteres adyacentes contienen más información y tienen una mayor probabilidad de formación de palabras. Las palabras candidatas se pueden expandir mediante la entropía de información izquierda y derecha, y los límites de las palabras se pueden determinar para formar palabras con un significado completo. La entropía de información izquierda y derecha se puede expresar como:

$$\text{Left information entropy: } E_L = - \sum_{W_l \in S_l} p(W_l|W) \log p(W_l|W) \quad \dots (2)$$

$$\text{Right information entropy: } E_R = - \sum_{W_r \in S_r} p(W_r|W) \log p(W_r|W) \quad \dots (3)$$

Dónde, S_l representa el conjunto de palabras de carácter adyacentes a la izquierda W , S_r representa el conjunto de palabras adyacentes de carácter W , $p(W_l|W)$ denota la probabilidad condicional de que W_l es una adyacencia izquierda del personaje W , y $p(W_r|W)$ denota la probabilidad condicional de que W_r es una adyacencia correcta de carácter W .

La tecnología de extracción de frases combina las estadísticas de información mutua y entropía de información, comenzando por la estrechez de la combinación interna de las cadenas de palabras y la medición de límites de las palabras adyacentes izquierda y derecha, y finalmente presenta las cadenas de palabras con alto contenido de información de frecuencia de coocurrencia en una forma completa, a fin de mejorar la precisión de la formación de palabras para optimizar el proceso de segmentación de palabras.

2.5.4.2. Procesamiento de Datos de la Lista de Atributos de Accidentes.

Después de realizar la segmentación de palabras basada en la información mutua y la entropía de la información en las descripciones textuales de las causas de los accidentes, se

extraen las palabras relevantes del dominio objetivo estableciendo umbrales de información mutua y entropía de información a la izquierda y derecha. Esto permite generar una lista de atributos de accidentes basada en la clasificación de frecuencia de palabras. Posteriormente, las palabras de alta frecuencia se contextualizan para formar una lista de términos de atributos de accidentes, facilitando la identificación efectiva de información valiosa a partir de datos no estructurados. Los factores de atributos de los accidentes de colapso en la construcción y la gravedad de los accidentes se convierten en datos procesables. Se transforman 49 factores de atributos de accidentes, que son variables categóricas no ordenadas, en información de datos utilizando la notación binaria "0-1". Las tres categorías de gravedad de accidentes, representadas por variables categóricas no ordenadas, se codifican como "0" para accidentes generales, "1" para grandes y "2" para mayores. Finalmente, se recopilan 264 registros de datos para proporcionar soporte de datos en la evaluación de riesgos utilizando algoritmos de aprendizaje automático.

2.5.4.3. Procesamiento de Desequilibrio de Clases Basado en Tecnología de Sobremuestreo.

Los conjuntos de datos analizados presentan un alto desequilibrio de clases, donde el número de casos de accidentes importantes en colapsos de construcción es dos y once veces mayor que los de accidentes generales y mayores, respectivamente. Este desequilibrio afecta negativamente el aprendizaje de las clases subrepresentadas. Para abordar este problema, se utilizan métodos de submuestreo y sobremuestreo. El submuestreo, que elimina algunas muestras de la clase mayoritaria, puede resultar en la pérdida de información y afectar el rendimiento cuando el tamaño de la muestra inicial es pequeño. Por lo tanto, se prefiere el sobremuestreo para ampliar el conjunto de datos originales y superar el desequilibrio de clases.

La Técnica de Sobremuestreo Sintético de Minorías (SMOTE) es un método representativo que no reutiliza directamente las muestras de clases minoritarias, sino que

genera una tasa de sobremuestreo específica basada en el conjunto de datos de la clase mayoritaria. El algoritmo *K-nearest neighbor (KNN)* se utiliza para sintetizar nuevas muestras y sobremuestrear la clase minoritaria, reduciendo así la posibilidad de sobreajuste del modelo y mejorando el rendimiento de generalización de las técnicas de aprendizaje automático (Pan et al., 2020). Como una técnica de preprocesamiento eficaz para manejar datos desequilibrados, SMOTE se ha aplicado ampliamente en diversos campos, incluyendo la predicción del cáncer médico, la predicción de defectos de software y la predicción de quiebras corporativas (Feng et al., 2021).

2.5.4.4. Análisis de Características de Multicolinealidad.

La multicolinealidad se refiere a la presencia de relaciones precisas o altamente correlacionadas entre las variables explicativas en un modelo de regresión lineal, lo que distorsiona los resultados de la estimación del modelo y afecta su capacidad predictiva (Assaf & Tsionas, 2021). Cuando existe covarianza entre las características seleccionadas, las variables independientes proporcionan información redundante, por lo que es necesario eliminar las variables insignificantes para reducir el sesgo de estimación causado por la multicolinealidad. Para medir la presencia de multicolinealidad en las variables del modelo, se pueden calcular los coeficientes de correlación de Spearman y realizar una agrupación jerárquica basada en las correlaciones de orden de rango. Si el coeficiente de correlación resultante supera un umbral específico, se retiene solo un atributo de cada grupo.

El coeficiente de correlación de Spearman, también conocido como coeficiente de correlación de rango, no se centra en la distribución de las variables originales, sino que realiza un análisis de correlación lineal sobre los valores de clasificación de las dos variables, reflejando la dirección del cambio y la fuerza de la correlación. Este coeficiente se expresa como:

$$\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad \dots (4)$$

Donde, d_i es la diferencia entre los rangos de cada par de valores y n es el número de observaciones. Esta medida permite identificar y mitigar la multicolinealidad, mejorando así la precisión y fiabilidad del modelo predictivo.

2.5.4.5. Modelado de Predicción de Accidentes por Colapso de Obras.

Entre las técnicas tradicionales de aprendizaje automático, el algoritmo de *Random Forest* (RF) ha demostrado mejoras significativas en el rendimiento del modelo de predicción y en la reducción del sobreajuste (Lourenço & Cristina, 2021). Además, la medida de importancia de las variables en el algoritmo RF permite calcular y clasificar la relevancia de las características en datos de alta dimensión, lo que facilita la construcción de un modelo de aprendizaje integrado no lineal que obtiene puntajes de importancia y determina los factores clave que afectan la precisión del modelo. Por estas razones, se seleccionó el algoritmo RF para crear un modelo de predicción de accidentes por colapso en la construcción. La idea básica del algoritmo RF, una técnica de aprendizaje integrado que utiliza árboles de decisión como aprendices base, consiste en seleccionar y entrenar múltiples árboles de decisión mediante muestras y características aleatorias a través de *bagging*, y luego combinar los modelos de varios clasificadores simples para mejorar la clasificación general. Como resultado, el algoritmo RF ofrece alta precisión de predicción, rápida velocidad de entrenamiento, y fuerte capacidad de generalización e interpretabilidad del modelo (Ali et al., 2022).

En este estudio, se construyó un modelo de evaluación de riesgos utilizando el algoritmo RF a partir de datos recopilados de 264 informes de accidentes por colapso en la construcción. Se seleccionaron aleatoriamente algunas características de los atributos extraídos para formar un árbol de decisión basado en el principio de minimizar el coeficiente de Gini, y el resultado final de la predicción se determinó por mayoría de votos de los árboles de decisión.

Los datos de investigación se dividieron en conjuntos de entrenamiento y prueba en una proporción de 8:2, y esta división tuvo un impacto crucial en la robustez del modelo de evaluación.

2.5.5. Optimización del Modelo y Evaluación del Rendimiento

Se sabe a partir de la construcción del algoritmo RF que algunas características de atributos se elegirán al azar en la construcción del árbol de decisión, y las características de atributos críticos se pueden identificar utilizando el método de evaluación de características, de modo que el método de reducción de características se puede utilizar para mejorar el rendimiento de predicción del algoritmo. Por lo tanto, esta subsección crea dos contextos de predicción de atributos de características cruciales y ajuste de hiper parámetros basados en el modelo de predicción original para maximizar la efectividad. El modelo RF debe generalizarse para reducir el problema del sub ajuste o sobreajuste del modelo, el uso de la validación cruzada *K-fold* para poder ajustar la distribución de datos de manera uniforme, lo que ayuda a evaluar el nivel de consistencia de los resultados para diferentes divisiones aleatorias de los datos, mejorando así la precisión de los modelos propuestos (Arteaga et al., 2020).

2.5.5.1. Optimización de Modelos de Predicción Utilizando la Importancia de las Características.

La idea básica de usar RF para la evaluación de la importancia de las características es evaluar la contribución de cada característica en el proceso de toma de decisiones, y el índice de Gini o error fuera de la bolsa (OOB) se elige el valor del índice cuantitativo para medir la contribución. El error OOB es el método de estimación interna para monitorear el error de predicción de OOB después de construir un árbol de decisión utilizando muestras de entrenamiento seleccionadas aleatoriamente, y luego el valor de OOB se calcula nuevamente después de desplazar aleatoriamente los valores de observación de la variable. El valor

promedio de la diferencia entre los dos errores OOB después de la estandarización en todos los árboles de decisión es la medida de importancia variable (VIM), que se utiliza para medir la importancia de las características de los atributos de RF (Janitza et al., 2016). El valor VIM de la variable de característica X_j se expresa de la siguiente manera:

$$VIN(X_j) = \frac{1}{N_{tree}} \sum_i (Err\bar{O}OB_i^j - ErrOOB_i) \quad \dots (5)$$

Dónde, N_{tree} representa el número de árboles de decisión, $ErrOOB_i$ Representa el número de muestras de error de datos fuera de la bolsa. OOB_i que el $i - th$ árbol de decisión responde a, y $Err\bar{O}OB_i^j$ representa el número de muestras de error fuera de la bolsa obtenidas al recálculo después del reemplazo aleatorio de la variable característica X_j El principio de este proceso de medición es introducir interferencias de ruido aleatorio en las características de los atributos y medir el cambio en la precisión inicial. Si el cambio es significativo, significa que el elemento característico tiene un mayor impacto en los resultados de la clasificación de la muestra, es decir, es más importante.

2.5.5.2. Ajuste de Hiperparámetros de los Clasificadores RF.

La optimización de hiperparámetros en algoritmos de aprendizaje automático (ML) tiene como objetivo encontrar la combinación de parámetros que maximice el rendimiento del algoritmo en el conjunto de datos de validación. En comparación con otros algoritmos de ML, *Random Forest* (RF) tiene más hiperparámetros que requieren optimización, tales como el número de árboles de decisión ($n_{estimators}$), el número máximo de características ($max_features$), la profundidad máxima del árbol de decisión (max_depth), el número mínimo de muestras en los nodos terminales ($min_samples_leaf$) y el número mínimo de muestras requeridas para realizar una división ($min_samples_split$).

El número de árboles de decisión determina cuántos árboles CART se crean. El número máximo de características limita la cantidad de variables consideradas al buscar el mejor punto de división. La profundidad máxima ajusta la estructura del árbol, donde más nodos terminales pueden introducir más variabilidad en el submodelo. Las muestras mínimas en los nodos terminales y las divisiones mínimas permiten ajustar la estructura del árbol de manera más detallada; sin embargo, valores menores para estos parámetros pueden reducir la estabilidad del modelo (Bergstra & Bengio, 2012).

Dada la complejidad y diversidad de los hiperparámetros, se utiliza el método de ajuste automático de parámetros mediante búsqueda aleatoria (*RandomizedSearchCV*) para optimizar los hiperparámetros que impactan el rendimiento general del modelo. Este método controla la complejidad del modelo y el costo temporal del aprendizaje. A diferencia de *GridSearchCV*, que realiza una búsqueda exhaustiva en el espacio de parámetros, la búsqueda aleatoria encuentra combinaciones casi óptimas de hiperparámetros con menos iteraciones al realizar un muestreo aleatorio en el espacio de parámetros.

2.5.5.3. Evaluación del Desempeño.

Para el modelo de predicción de la gravedad de los accidentes por colapso de la construcción, se utilizan los tres parámetros de precisión, recuperación y puntuación F1 como indicadores de datos para evaluar el rendimiento del modelo. Además, se combinan dos herramientas de evaluación del rendimiento generalizadas, la matriz de confusión y la curva ROC, para lograr una evaluación completa de la capacidad de aprendizaje del modelo.

- Indicadores de datos.

Las fórmulas para calcular la precisión, la recuperación y la puntuación F1 se expresan de la siguiente manera:

$$Precision = \frac{TP}{(TP + FP)} \quad \dots (6)$$

$$Recall = \frac{TP}{(TP + FN)} \quad \dots (7)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad \dots (8)$$

Dónde *TP* Se denomina Verdadero Positivo e indica el número de muestras positivas que se prevé que sean verdaderas. *FP* Se llama falso positivo e indica el número de muestras negativas que se predice que serán verdaderas. *FN* Se llama falso negativo e indica el número de muestras positivas que se prevé que sean falsas.

En términos de significado de indicador, la precisión es una medida de cuán precisas son las predicciones positivas, y la recuperación es una medida de cuántos de los positivos reales puede identificar el modelo. Además, la precisión y la recuperación con frecuencia muestran una asociación negativa, con un aumento en un indicador a veces seguido de una disminución desproporcionada en el otro (Diez et al., 2021). El promedio armónico de la puntuación F1, que considera tanto la precisión como la recuperación, proporciona una evaluación más exhaustiva del desempeño del clasificador, con valores F1 más altos que sugieren mejores resultados de clasificación.

- **Matriz de confusión.**

La matriz de confusión suele presentar las categorías de predicción del modelo sobre un conjunto de datos de prueba con categorías verdaderas conocidas en forma de tabla o gráfico, lo que puede demostrar eficazmente la precisión del clasificador de forma visual. El número de predicciones correctas predichas por el modelo de clasificación está representado por los elementos diagonales de la matriz de confusión, mientras que el número de predicciones incorrectas está representado por los elementos no diagonales de la matriz, por lo que comparar

los valores predichos en la matriz de confusión con los valores verdaderos puede revelar la distribución de predicciones incorrectas en diferentes clases.

- **Curva característica de funcionamiento del receptor (ROC).**

La curva ROC es una herramienta de medición importante para el análisis predictivo. Después de calcular el valor de la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR), se forma la curva característica con la FPR como coordenada horizontal y la TPR como coordenada vertical. Este método de cálculo tiene en cuenta la capacidad de clasificación del clasificador para casos positivos y negativos al mismo tiempo en el proceso de trazado, eliminando el efecto de las categorías de muestra desequilibradas en el clasificador, y la precisión de la clasificación se puede observar de forma fácil e intuitiva a partir de la figura.

2.5.6. Identificar Factores de Atributos Críticos

A partir del proceso de optimización de la aplicación de características de importancia al modelo de predicción de RF, se crea una clasificación de importancia de las características de los factores de atributo de los accidentes por colapso de la construcción, que indica la relevancia relativa de cada factor de atributo. Los resultados de la clasificación de importancia de las características de los factores de atributo se pueden utilizar como una valiosa referencia para mejorar el modelo de predicción, así como para proporcionar un análisis causal claro de los factores de atributo que conducen a la gravedad de los accidentes, analizar el proceso de accidentes de seguridad en el lugar de trabajo y proporcionar una base teórica para crear un sistema de alerta temprana basado en la gestión de peligros ocultos y la clasificación de riesgos.

2.5.7. Análisis Clásico sobre los Factores de Riesgo en los Proyectos

La identificación y evaluación de los factores de riesgo siempre ha sido un tema candente entre los académicos, y se han propuesto muchos métodos y tecnologías maduros. El proceso de análisis y evaluación de riesgos es la correlación vital entre la identificación

sistemática de los riesgos y la gestión racional de los riesgos clave (Wang & Zhang, 2018). En primer lugar, la mayoría de los académicos creen que los factores de riesgo no están limitados por el marco organizacional y los límites disciplinarios, y son omnipresentes en todo el ciclo de vida de un proyecto (Dikmen et al., 2012). Además, a menudo se considera la fuerte correlación entre los factores de riesgo (Alawamleh & Popplewell, 2011). Tavakolan & Etemadnia, (2017) construyen el modelo de estructura interpretativa ponderada difusa (FWISM) y analizan la influencia y dependencia de los factores de riesgo en función de la red de interacción de riesgos de los proyectos de construcción. La estructura difusa se utiliza a menudo para analizar la relación entre los factores de riesgo y las consecuencias (Liu et al., 2013), para juzgar los posibles factores de riesgo basándose en los pesos difusos correspondientes (Rezakhani et al., 2014). Algunos académicos han utilizado modelos de dinámica de sistemas para encontrar factores interdependientes en redes relacionadas. Wang & Yuan (2017) establecieron un modelo dinámico de sistemas (SDM) basado en el diagrama de ciclo causal y la relación entre variables y encontraron que hay una interacción significativa entre diferentes factores de riesgo del cronograma. Además, es común utilizar métodos de ciencias sociales como la investigación empírica para identificar riesgos. Choudhry & Iqbal (2013) realizaron estadísticas sobre los factores de riesgo de la industria de la construcción de Pakistán a través de una encuesta por cuestionario y juzgaron que las finanzas, la economía y la calidad son los factores de riesgo más importantes. El modelado de ecuaciones estructurales (SEM) también se considera una de las técnicas más apropiadas para analizar la relación entre las variables de factores de riesgo (Almeida et al., 2016), incluido el SEM basado en covarianza (CB-SEM) y el SEM de mínimos cuadrados parciales (PLS-SEM).

En comparación con la identificación de factores de riesgo, la evaluación es más compleja. El AHP se ha utilizado ampliamente en la evaluación de riesgos (Khan & Rawat, 2016), pero no puede procesar datos inestables ni conocimiento difuso. Para superar este

defecto, se propone el proceso de jerarquía analítica difusa (FAHP) para resolver el problema del número difuso en el proceso de evaluación de riesgos (Khademi Hamidi et al., 2010). Lyu et al., (2020) propusieron un nuevo proceso de consulta para resolver los problemas de datos difusos inciertos y matriz de juicio inconsistente, y verificaron el efecto del FAHP basado en la identificación de factores de riesgo. Mientras tanto, los académicos adoptaron métodos integrales para establecer el sistema de evaluación de riesgo proporcional difuso pitagórico (PFPPRA) (Marcelino-Sádaba et al., 2014), que incluyen el método de Kinney fino, el proceso de jerarquía analítica difusa pitagórica y un sistema de información difusa (Ilbahar et al., 2018). Además, FAHP todavía necesita el apoyo de números difusos que son difíciles de determinar (J. Liu et al., 2013b). En consecuencia, la combinación de toma de decisiones difusa y SEM es uno de los medios importantes para evaluar los factores de riesgo. Por ejemplo, el modelo de evaluación difusa se utiliza para calcular la probabilidad de ocurrencia, el grado de influencia y el punto crítico de riesgo de cada factor de riesgo (Zhao et al., 2016), o el SEM se utilizó para el análisis de regresión para evaluar la importancia relativa (Eyboosh et al. (2011, 2012). Sin embargo, algunos académicos también han encontrado que los efectos negativos de los factores de riesgo tienen histéresis (Aroquipa & Hurtado, 2022), y este efecto adicional puede ser determinado por el mecanismo de influencia de los factores de riesgo (Qazi et al., 2016). Curiosamente, también proporciona una oportunidad para que este estudio vaya más allá del impacto de un solo evento y evalúe los factores de riesgo del proyecto en general.

2.5.8. *Aprendizaje Automático y Análisis de Riesgos*

El aprendizaje automático (ML), respaldado por nuevas tecnologías, presenta ventajas significativas para analizar factores de riesgo inciertos. La investigación relevante en este campo se puede resumir en dos aspectos principales. Primero, el aprendizaje automático puede aplicarse a texto no estructurado, como informes de accidentes, registros de reuniones y otros textos relacionados con proyectos, los cuales contienen factores de riesgo potenciales

(Evangelopoulos et al., 2012). Mediante técnicas de minería de datos, es posible extraer estos factores (Evangelopoulos et al., 2012). Por ejemplo, Aroquipa et al., (2023) utilizó el algoritmo de asignación latente de Dirichlet (LDA) para analizar textos de accidentes e identificar factores causales de impactos extremos. (Rajasekaran et al., 2008) combinaron Bayes ingenuo (NB), modelos de palabras, máquinas de vectores de soporte (SVM) y métodos de regresión logística para analizar eventos de lesiones laborales, encontrando que grandes conjuntos de datos son altamente efectivos para la clasificación. Herramientas más avanzadas, como bosques aleatorios (RF) y árboles de gradiente estocástico (SGTB), pueden estimar la probabilidad de ocurrencia de factores de riesgo a partir de grandes volúmenes de datos textuales de accidentes, permitiendo así predecir riesgos futuros (Tixier et al., 2016). Además, el procesamiento del lenguaje natural (NLP) es eficaz para analizar datos no estructurados, codificando texto y marcando factores de riesgo (Zou et al., 2017).

El segundo aspecto del aprendizaje automático se centra en datos numéricos, que incluyen características de accidentes y datos de series temporales. Para explorar la relación entre accidentes y factores de riesgo, Zou et al. (2016) establecieron reglas de asociación entre accidentes por caídas y factores de riesgo, revelando la influencia de diferentes variables. Kwon et al., (2015) utilizaron un clasificador bayesiano ingenuo y análisis de decisiones para evaluar el impacto relativo de los factores de riesgo. Algunos estudios se centran en la predicción de riesgos: Rajasekaran et al., (2008) utilizaron SVM para predecir el oleaje generado por tormentas oceánicas, la probabilidad de accidentes y su posible gravedad. Además, redes neuronales artificiales (ANN) y redes neuronales convolucionales (CNN) también se han utilizado para evaluar factores de riesgo (Heravi & Charkhakan, 2014). Es importante señalar que el ML requiere grandes cantidades de datos para funcionar efectivamente, ya sea texto no estructurado o datos estructurados, lo que implica que el éxito del ML depende de la disponibilidad de conjuntos de datos confiables.

2.5.8.1. Transmisión de Riesgos y Efectos en Cascada.

El efecto en cascada se observa en el proceso de transferencia de riesgos, aunque a menudo se pasa por alto, ocurre con frecuencia en proyectos reales. Un evento menor puede desencadenar una serie de sucesos desafortunados (incluyendo muchas víctimas), lo que demuestra que una acción puede impactar todo el sistema y llevar a eventos inesperados (Vaiman et al., 2012). Los "fallos" en cascada de un proyecto a menudo se deben a un diseño estructural incorrecto y una cultura organizacional negligente (Center, 2013). Además, la existencia de una red activa facilita la transmisión de fallas, afectando todo el proyecto. En este contexto, el efecto en cascada requiere un portador.

El cronograma de un proyecto de construcción es una red típica de actividades, estructurada en torno a dos elementos: actividades y nodos. Liu et al. (2013) analizaron el impacto de los accidentes de seguridad inducidos por retrasos en el cronograma y aumentos de costos en el proyecto. Además del retraso del cronograma, el consumo de recursos provocado por accidentes también se utiliza para evaluar el grado de impacto de cada factor. Específicamente, el impacto directo de los factores de riesgo son las víctimas, pero también pueden afectar el costo, el progreso, la calidad y otros aspectos (Flyvbjerg, 2013). De esta manera, el retraso del cronograma puede ser una consecuencia de los factores de riesgo. Por lo tanto, si un accidente provoca el retraso de una tarea, existe la posibilidad de afectar a las tareas subsecuentes. La transmisión del retraso a través de la red de actividades puede llevar a un efecto en cascada, que es un aspecto clave en este estudio.

2.5.8.2. Metodología Inteligente.

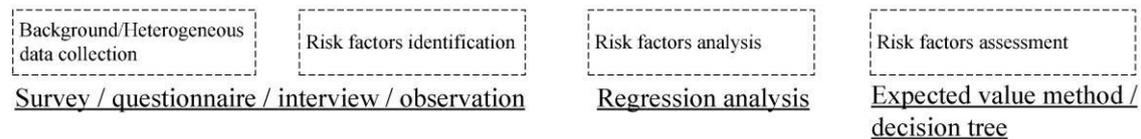
En este estudio, se utiliza un análisis integral para completar la identificación y evaluación de los factores de riesgo de seguridad del proyecto, combinando el conocimiento existente de la gestión de la seguridad del proyecto. El marco completo consta de tres partes:

1) identificar los factores de riesgo relacionados con la seguridad del proyecto a partir de textos no estructurados generados durante la gestión del proyecto; 2) definir los factores de riesgo en función de los tipos importantes de "características" y "resultados"; y 3) establecer una red AON para analizar el efecto en cascada y cuantificar el impacto de los factores de riesgo en los proyectos. La Figura 3, muestra el análisis de los factores de riesgo del proyecto. Los componentes del marco se analizan en detalle en las siguientes secciones.

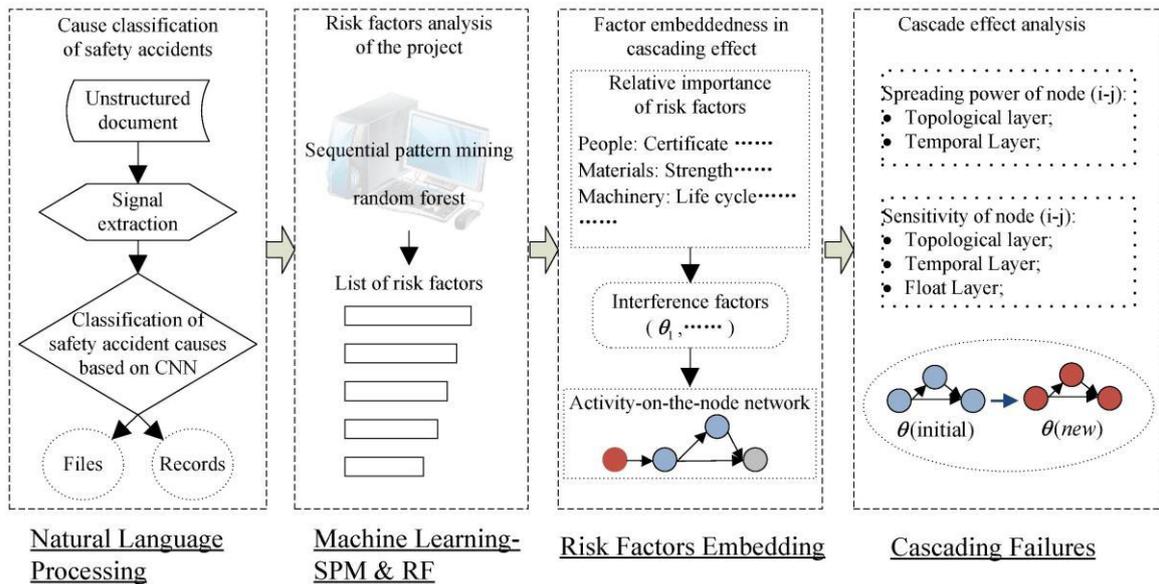
Figura 3

Análisis de Factores de Riesgo de Seguridad del Proyecto

Retrospective Study



Proposed Framework



Nota. La fuente de esta información es: (Sarkar & Maiti, 2020b)

- **Aplicación del Marco de Análisis Integrado (IAF) en la Gestión de Seguridad del Proyecto**

El Marco de Análisis Integrado (IAF) no solo es adecuado para la identificación de riesgos durante la etapa de construcción, sino también para la predicción de riesgos en fases posteriores del proyecto. Los informes de accidentes de seguridad generados durante la construcción contienen información crucial sobre riesgos ocultos. Al extraer estos factores de riesgo de seguridad de los informes, se pueden formular medidas preventivas de manera anticipada para abordar peligros similares en etapas futuras del proyecto. Además, actas de reuniones, registros de construcción y otros documentos también pueden contener información relevante sobre riesgos, por lo que el uso del IAF permite identificar riesgos potenciales en la gestión del proyecto. Dado que el marco solo necesita establecerse una vez, su aplicabilidad es universal para proyectos de construcción, permitiendo a los gerentes analizar los factores de riesgo con el apoyo de documentos relacionados con la seguridad a lo largo de todo el proceso de construcción.

- **Extracción de Factores de Riesgo en la Gestión de Proyectos**

La gestión de proyectos genera grandes cantidades de datos no estructurados, incluyendo documentos, audio e imágenes. En este estudio, los documentos generados durante la gestión de proyectos, como textos y registros de reuniones, se utilizan como una fuente de datos heterogénea para extraer factores de riesgo que afectan la seguridad del proyecto. Estos textos pueden incluir planes de seguridad del proyecto, informes de inspección de seguridad, entre otros. Se diseñó un modelo de red neuronal convolucional (CNN) para analizar estos datos y, posteriormente, se utilizaron reglas de asociación para extraer los factores de riesgo. Luego, se calculó la importancia de los resultados (IoR) y las características (IoF) de cada factor de riesgo, considerando también los términos de interferencia para promover el efecto cascada.

- Evaluación de Factores de Riesgo en la Gestión de Proyectos

Dado que los factores de riesgo se generan durante la gestión de proyectos, es difícil evaluarlos de forma aislada sin considerar el desempeño del proyecto. Además, el desempeño del proyecto es un criterio importante para evaluar los factores de riesgo (Dikmen et al., 2012). La existencia de un plan de red de proyectos facilita la transmisión de efectos de riesgo, que afectan al proyecto completo mediante tareas interconectadas. Por lo tanto, es crucial establecer una red AON y analizar el efecto en cascada de los factores de riesgo, utilizando el período del proyecto como objetivo cuantitativo central. Los detalles de este trabajo se presentarán en las siguientes secciones.

2.5.9. Extracción y Clasificación de Factores de Riesgo

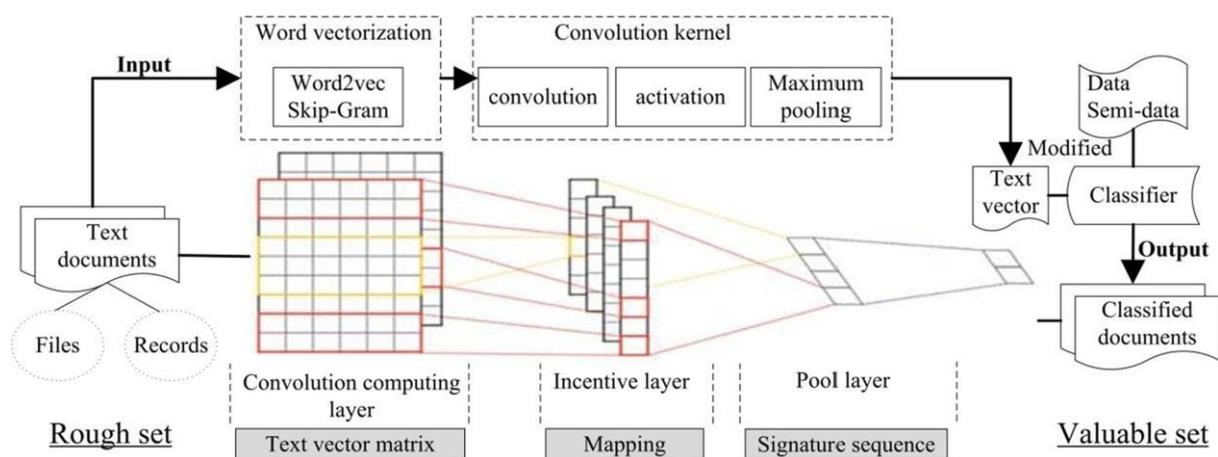
2.5.9.1. Clasificación de Documentos de Texto.

En la mayoría de los casos, el equipo del proyecto tomará diversas medidas para reducir los riesgos de seguridad del personal, y las medidas de control deben formularse de acuerdo con la causa del accidente. Por lo tanto, este estudio también analiza los accidentes de seguridad causados por diferentes razones y detecta los factores de riesgo de seguridad bajo las razones correspondientes. El trabajo básico es clasificar los documentos de texto de acuerdo con las causas del accidente y definir un cierto tipo de documento como T_i , de manera que podamos extraer los factores de riesgo de seguridad correspondientes a la i -th accident. El modelo CNN no solo se puede utilizar para la clasificación de texto con una estructura compleja y muchos componentes, sino que también tiene un buen rendimiento para datos heterogéneos masivos con características cruzadas (Li et al., 2019). Teniendo en cuenta que los documentos de texto tienen una heterogeneidad obvia, también es necesario limpiar una gran cantidad de datos antes de usarlos, por lo que presentamos CNN para clasificar todo tipo de materiales y registros de reuniones en el primer paso.

CNN se compone de una capa de entrada, una capa de computación de convolución, una capa de excitación, una capa de pool, una capa de conexión y una capa de salida (como se muestra en la Figura 4. Para obtener un excelente conjunto de entrenamiento, entrenamos 150 casos típicos de accidentes de seguridad (los datos del conjunto de entrenamiento se extraen de informes de accidentes de seguridad publicados por la Administración de Seguridad y Salud Ocupacional-OSHA, que incluyen una variedad de tipos de edificios) bajo *TensorFlow* 1.14. El conjunto de entrenamiento y el conjunto de prueba son R^* y R respectivamente. Mientras tanto, desarrollamos el marco de clasificación de texto que se muestra en la Figura 4, que se compone de matriz vectorial, convolución, función de activación, capa de agrupación máxima y clasificador perceptrón multicapa. En consecuencia, el trabajo de todo el marco se divide en tres procesos: transformación de vectores de palabras, extracción de características de texto y clasificación automática.

Figura 4

Clasificación de Documentos de Texto Basada en CNN



Nota. Esta información ha sido obtenida de: (Sarkar & Maiti, 2020b)

Además, como el vocabulario no se puede introducir directamente en el modelo CNN para su procesamiento, es necesario transformarlo en una matriz vectorial que pueda ser reconocida y procesada por ordenador. Taroun (2014) propuso que, en el espacio vectorial de

palabras, siempre que la semántica sea similar, incluso las palabras con diferentes propiedades están estrechamente vinculadas en la distancia del texto. Por lo tanto, introducimos *word2vec* para transformar las palabras de entrada en vectores. Cabe señalar que la información de contexto de los documentos de texto sobre accidentes de seguridad del proyecto es incompleta y las características del texto también son difusas. Para evitar una mala eficiencia de entrenamiento causada por información incompleta, se utiliza el modelo *skip-gram* para el procesamiento de texto.

Para evaluar el rendimiento del modelo CNN en la clasificación de texto, los resultados de este estudio se compararán con SVM, el algoritmo de k vecinos más cercanos y *Naive Bayes*. Mientras tanto, los parámetros del modelo se establecen en el mejor estado y se utiliza el mismo conjunto de datos. Los índices de evaluación incluyen Precision, *Recall* y el valor F1, y la fórmula se muestra a continuación:

$$Precision = \frac{TP}{(TP + FP)} \quad \dots (9)$$

$$Recall = \frac{TP}{(TP + FN)} \quad \dots (10)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad \dots (11)$$

La precisión es la relación de todos los “elementos recuperados correctamente” (*TP*) a todos los “elementos realmente recuperados” (*TP + FP*), y el recuerdo es la proporción de todos los “elementos recuperados correctamente” (*TP*) a todos los “elementos que deberían recuperarse” (*TP + FN*). Los tres indicadores anteriores se utilizan a menudo para evaluar el efecto de clasificación de los documentos de texto y juzgar el rendimiento real del modelo. Para describir claramente el proceso de clasificación de texto, se utiliza un caso de accidente para probar el modelo CNN. Los pasos principales son los siguientes.

- **Preprocesamiento de texto.** Los 150 casos típicos de accidentes de seguridad constituyen un corpus supervisado (representado por la característica X y la etiqueta de clasificación Y), y luego se induce (entrena) un conjunto de reglas de clasificación a partir del conjunto de datos construido mediante aprendizaje automático. Después de la preparación del corpus, se eliminan los símbolos sin sentido u otra información redundante, se corrige el tiempo de las oraciones y se separan las palabras con espacios. Los resultados del pretratamiento se pueden expresar de la siguiente manera:

En el momento del derrumbe, los automovilistas estaban esperando debajo del puente el semáforo. Un empleado y cinco automovilistas resultaron heridos mortalmente y otro empleado resultó como discapacitado permanentemente.

- **La dirección del texto.** Esta parte es principalmente para mapear la información de vocabulario a un espacio de vector de palabras numéricas, y usar la herramienta word2ve para realizar el mapeo de la información de vocabulario al espacio semántico en el conjunto de entrenamiento, y finalmente obtener un modelo de vector de palabras. Tomando las palabras “*collapse*” como ejemplo, las palabras más relevantes son “*platform* = [0.97393211233], *injure* = [0.969927314301], *kill* = [0.96290013827], *concrete* = [0.95327622081]”, el vector de palabras obtenido es:

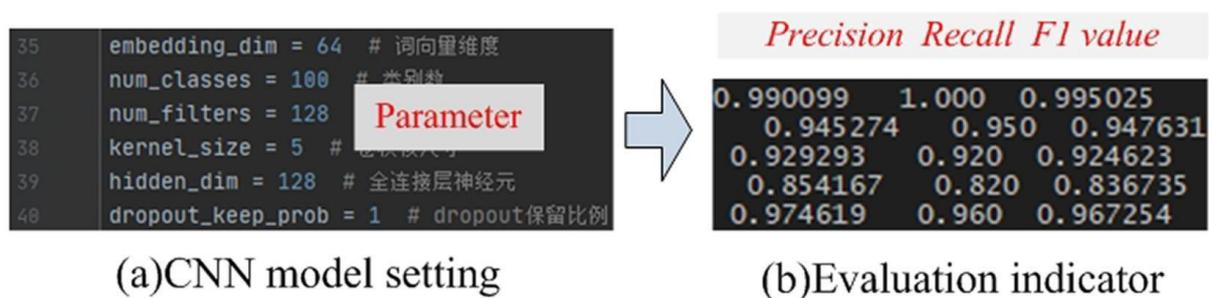
$$\begin{bmatrix} 0.01324947 & 0.01233419 & -0.01283749 & -0.02473493 & 0.03432512 & 0.01864652 \\ 0.03228544 & 0.01927345 & -0.0327201 & 0.09328726 & -0.01774350 & -0.02065491 \\ -0.2290923 & 0.0532193 & -0.4624433 & 0.11873217 & 0.1320874 & \dots \end{bmatrix}$$

- **Clasificación de texto basada en el modelo CNN.** El modelo CNN es un modelo ideal con eficiencia y calidad en tareas de clasificación de texto. Tiene

una gran adaptabilidad para tratar con la clasificación de documentos de múltiples tipos. Este artículo estudia el uso de *pip* para construir un entorno virtual de *Python* (es decir, virtual) e instalar *TensorFlow* 1.14 en el entorno virtual. Las configuraciones de la capa de grupo correspondiente y las características de clasificación se determinan de acuerdo con los requisitos del modelo CNN y los tipos de texto. Los resultados finales de la prueba se muestran en la Figura 5.

Figura 5

Configuración del Modelo CNN y Resultados de la Prueba



Nota. Esta figura es extraída de: (Sarkar & Maiti, 2020)

- **Evaluación de resultados.** Para el efecto de prueba, utilizamos los valores de precisión, recuperación y F1 para evaluar, y los resultados se muestran en la Figura 5. La precisión de la etiqueta de texto " colapso " es 0,990099, lo que demuestra el buen efecto de clasificación del modelo CNN.

Floro2

2.5.10. Un Enfoque de Aprendiendo Sobre los Accidentes Graves

Aprender del pasado siempre ha desempeñado un papel importante a la hora de impulsar la innovación y promover los avances. Sin duda, los errores son parte de la naturaleza humana, pero todos tenemos la capacidad inherente de aprender de ellos. Sin embargo, extraer

una lección y aplicar el conocimiento adquirido para evitar errores recurrentes no es tan trivial como parece. La historia tiende a repetirse y las lecciones pueden ignorarse u olvidarse (Paltrinieri et al., 2013) .

Diferentes actividades humanas tienen diferente tolerancia a los errores. Dentro de la industria química, se han hecho esfuerzos significativos para evitar errores y garantizar operaciones seguras. Sin embargo, antes de la segunda mitad de los años sesenta, las palabras “seguridad de proceso” y “prevención de pérdidas” no existían; la manipulación y el almacenamiento de sustancias peligrosas estaban regulados por la seguridad ocupacional tradicional y las buenas prácticas de ingeniería. Más tarde, una serie de accidentes terribles, incluidos. En el cambiante campo de la seguridad de procesos, siempre ha estado claro que las lecciones derivadas de accidentes pasados habrían sido cruciales para garantizar un diseño y operaciones más seguros (Pasman, 2009).

La abundancia de datos sobre accidentes ofrece una gran oportunidad para aprender de los errores del pasado. Sin embargo, el proceso de aprendizaje actual tiene limitaciones significativas y parece incapaz de aprovechar esta oportunidad. Por lo tanto, existe una gran necesidad de nuevas herramientas y técnicas para extraer y retener conocimiento de los datos sobre accidentes. En este contexto, los avances en informática e inteligencia artificial han llevado a la construcción de algoritmos capaces de extraer conocimiento de los datos (Brink et al., 2016). Además, la investigación se ha centrado en técnicas de aprendizaje automático (ML).

Dado que el aprendizaje de los accidentes graves se ve profundamente afectado por factores humanos, se puede argumentar que un aprendiz artificial sería un buen apoyo para mejorar las oportunidades de aprendizaje. Los algoritmos de aprendizaje automático podrían entrenarse para vincular las características del accidente (por ejemplo, las sustancias y el equipo involucrados, la magnitud de la liberación, la densidad de población) con las consecuencias del

accidente, por ejemplo, la cantidad de personas involucradas. Estos modelos predictivos serían un medio rápido, eficaz y económico de respaldar la toma de decisiones basada en el riesgo y la seguridad de los procesos. No obstante, el análisis de datos de accidentes de proceso a través de algoritmos de ML sigue siendo un tema en gran parte inexplorado. En este contexto, esta investigación tiene como objetivo contribuir a esta área de investigación explorando el uso de métodos de aprendizaje automático para analizar y extraer conocimiento de datos históricos de accidentes. Este estudio responde a necesidades específicas y apremiantes de herramientas para extraer conocimiento de accidentes pasados, retener y recordar fácilmente dicho conocimiento para su uso futuro. Los autores creen que el enfoque descrito en este estudio puede proporcionar a los gerentes y profesionales de seguridad modelos predictivos avanzados que pueden mejorar significativamente la toma de decisiones, la prevención y la mitigación de accidentes, lo que representa un paso esencial hacia la Seguridad 4.0. Lo que los usuarios pueden aprender del enfoque descrito aquí es (i) evaluar la criticidad de diferentes escenarios de accidentes basándose en un conjunto de características simples y fácilmente disponibles, (ii) discriminar entre diferentes niveles de criticidad y dirigir los esfuerzos para prevenir/mitigar escenarios de alta criticidad, (iii) estimar las consecuencias de nuevos escenarios de accidentes sin recurrir a técnicas de cálculo intensivo (por ejemplo, modelos CFD) y modelado detallado.

2.5.11. Trabajos Relacionados - Métodos de Aprendizaje Automático

Varios estudios han propuesto métodos de aprendizaje automático para extraer información crítica para la seguridad de datos históricos y predecir los resultados de eventos accidentales. Por ejemplo, (Sarkar & Maiti, 2020) utilizaron seis algoritmos de clasificación diferentes para predecir la gravedad de las lesiones de los accidentes que ocurrieron en una planta de fabricación de acero; los informes de investigación y los informes de inspección recopilados en un período de tiempo de 3 años se utilizan para entrenar y evaluar los modelos. Phark et al. (2018) analizaron la aplicación de clasificadores bayesianos ingenuos y

perceptrones multicapa para predecir la emisión de órdenes de evacuación de emergencia después de la liberación de sustancias tóxicas. Luo et al. (2020) propusieron un método para la recuperación semiautomática de escenarios de *Natech* de la base de datos del Centro Nacional de Respuesta, que emplearon la memoria a corto y largo plazo y la red neuronal convolucional como modelos de clasificación.

Además, varios estudios se centraron en el procesamiento del lenguaje natural (NPL) y los métodos de aprendizaje automático para analizar las narrativas de los accidentes y extraer información útil. Por ejemplo, (Kalkan & Kunnath, 2006) propusieron un enfoque de aprendizaje automático para clasificar los informes de accidentes no estructurados en tipos básicos de accidentes (por ejemplo, "salud/seguridad", "fuga/derrame", "operación"). Además, propusieron algoritmos NPL para derivar un conjunto de palabras clave más informativo y útil a partir de los informes de accidentes sin procesar. Lu et al. (2022) utilizaron Word2Vec (Mikolov et al., 2013) y una red neuronal bidireccional de memoria a largo plazo (Bi-LSTM) con un mecanismo de atención para (i) analizar la correlación entre accidentes y extraer precursores de accidentes, causas y tipos de alta frecuencia de accidentes químicos, y (ii) clasificar automáticamente los informes de accidentes en su respectivo tipo de accidente (es decir, "incendio", "explosión", "envenenamiento" y "otros"). Se desarrolló un diccionario propietario para mejorar la segmentación de palabras y el rendimiento de la clasificación. Wang y Whao (2022) también utilizaron Bi-LSTM para extraer y estimar la frecuencia de los factores contribuyentes de los informes de accidentes en espacios confinados. Los autores utilizaron el algoritmo BERT para crear una incrustación de palabras y un BiLSTM con un campo aleatorio condicional (CRF) para clasificar los accidentes en función de sus factores contribuyentes (por ejemplo, herramienta inadecuada, detección de gas, supervisión inadecuada). Dado que el enfoque está completamente supervisado, se necesita la intervención manual de expertos para extraer características fundamentales de los accidentes y sus factores contribuyentes. En

cambio, Ahadh et al. (2021) propusieron un enfoque semi supervisado para clasificar automáticamente los informes de accidentes de diferentes dominios en función de temas definidos por el usuario. El enfoque es independiente del dominio y requiere una mínima intervención humana. Los autores propusieron extraer palabras clave relevantes para el dominio de un corpus de dominio (por ejemplo, pautas, manuales estándar, artículos académicos y páginas de Wikipedia) e identificar la causa del accidente (por ejemplo, "Fuerza externa", "Fallo del equipo", "Operación incorrecta") u otras características de accidentes definidas por el usuario a partir de narrativas de accidentes. Se utilizó una versión guiada del algoritmo de asignación de Dirichlet latente (Jelodar et al., 2019) para extraer las características del accidente.

Aunque las investigaciones descritas anteriormente representan un valioso intento de extraer información de los informes de accidentes, su intención y metodología difieren significativamente del enfoque descrito en este estudio. Por ejemplo, se analizan narrativas de accidentes no estructuradas, mientras que este estudio se centra en bases de datos de accidentes estructuradas. Además, el objetivo principal de esos estudios es automatizar la extracción de piezas clave de información de texto no estructurado y, por lo tanto, reducir la necesidad de intervención manual por parte de expertos, que requiere mucho tiempo y es costosa. En cambio, los algoritmos propuestos en nuestro estudio no están diseñados para extraer características genéricas de accidentes (por ejemplo, la sustancia involucrada, la causa, la cantidad de sustancia liberada) ya que esta información ya está disponible en la base de datos estructurada utilizada para el análisis. En cambio, este estudio busca extraer conocimiento de nivel superior, que los expertos no pueden extraer simplemente leyendo los informes de accidentes. Específicamente, los algoritmos propuestos apuntan a capturar y cuantificar la relación entre las características de los accidentes y las consecuencias en términos de personas muertas y heridas. En otras palabras, el objetivo es extraer conocimiento de los informes de accidentes

históricos para construir un mapeo entre las características de los accidentes y las consecuencias de los accidentes. El método presentado en este estudio se puede utilizar para realizar predicciones; Dada una lista breve de características de accidentes, el modelo devuelve el número de personas involucradas en el accidente. En cambio, los estudios descritos anteriormente toman un texto extenso (es decir, narraciones de accidentes) como entrada y extraen información clave. En otras palabras, su objetivo no es la extracción de conocimiento para predecir el resultado de eventos accidentales; simplemente imitan el proceso de descubrimiento de conocimiento de un lector humano.

Similar a este estudio, Chebila, (2021) propuso un método basado en *Machine Learning* para predecir si los accidentes que involucran sustancias peligrosas causarán daños a los humanos, al medio ambiente y a los bienes materiales. Específicamente sobre las consecuencias para las personas, se realizó un conjunto de clasificaciones binarias utilizando seis modelos diferentes para predecir la ocurrencia de al menos un herido o muerto. El estudio concluyó que *Random Forest* asegura el mejor rendimiento. Además, las redes neuronales proporcionaron buenos resultados, pero demostraron ser menos efectivas que *Random Forest* al tratar con conjuntos de datos desequilibrados. La investigación de Chebila, (2021) comparte algunas características con este estudio, como la intención general y el enfoque; sin embargo, también existen diferencias significativas. Por ejemplo, el enfoque propuesto por Chebila, (2021) no distinguió entre heridos y muertos, mientras que el presente estudio considera estos resultados por separado. Además, el presente estudio utiliza un conjunto de múltiples variables de resultado discretas para diferenciar los accidentes según su gravedad (es decir, de 1 a 10 muertos, de 11 a 100 muertos, etc.). Por otra parte, Chebila (2021) ha utilizado y probado un mayor número de modelos de clasificación, que también consideran más objetivos (es decir, el medio ambiente y los activos materiales). Por último, se utilizan diferentes bases de datos; Chebila, (2021) utilizó eMARS , mientras que este estudio se centra en MHIDAS.

La industria química y de procesos no es el único sector industrial que se ha involucrado en esta línea de investigación. Por ejemplo, (Gerassis et al., 2020) propusieron el uso de un Análisis de Correspondencia Múltiple junto con Redes Bayesianas para clasificar los accidentes mineros como fatales o no fatales. El enfoque se probó en una base de datos de accidentes laborales y permitió la identificación del factor que más contribuye a la gravedad del accidente. Yedla et al. (2020) desarrollaron un enfoque diferente para predecir el número de días de ausencia del trabajo después de un accidente minero. El método hace uso de modelos de regresión y clasificación, como la regresión logística, los árboles de decisión, los bosques aleatorios y las redes neuronales artificiales, para predecir el número de días de ausencia del trabajo y el grado de lesión. De manera similar, (Choi et al. 2020) demostraron que los datos de accidentes podrían usarse para construir modelos de clasificación para la predicción de la probabilidad de mortalidad en caso de un accidente en un sitio de construcción.

Varios estudios también se han centrado en la industria del transporte. En el análisis propuesto por (Zhang et al., 2018), se compararon cuatro algoritmos de aprendizaje automático diferentes en función de la capacidad de predecir la gravedad de los accidentes que ocurrieron en segmentos de autopistas. El estudio concluyó que los modelos de aprendizaje automático producen un mejor rendimiento que los métodos estadísticos tradicionales en esta tarea específica. Además, los resultados sugirieron que *Random Forest* y *K-Nearest Neighbors* eran los mejores modelos. Assi et al. (2020) investigaron el uso de redes neuronales de avance y máquinas de vectores de soporte para predecir el nivel de gravedad de los accidentes de tráfico. Además, el estudio investiga el uso de agrupamiento de c-medias difusas para mejorar las capacidades de predicción del modelo. (Wahab & Jiang, 2019) propusieron un enfoque similar, que se centró en la predicción de la gravedad de los accidentes de motocicleta utilizando árboles de decisión y aprendizaje basado en instancias. Además, Burnett & Si (2017) demostraron el uso de técnicas de clasificación de aprendizaje automático para predecir los niveles de lesiones

y muertes en accidentes de aviación. El análisis concluyó que las redes neuronales artificiales funcionaron mejor que los otros modelos.

En general, una búsqueda en la literatura reveló que la atención de la comunidad científica se ha centrado recientemente en la aplicación de métodos de aprendizaje automático para la predicción de la gravedad de los accidentes. La idea de utilizar datos de procesos para actualizar el panorama de riesgos ya se ha propuesto en trabajos anteriores, por ejemplo, (Landucci & Paltrinieri, 2016). Sin embargo, el creciente cuerpo de investigación sobre métodos de aprendizaje automático indica que el enfoque puede desempeñar un papel importante en el futuro de la evaluación y gestión de la seguridad en varias áreas. Además, la búsqueda reveló que existe una notable escasez de estudios que investiguen la aplicación de dichos métodos a accidentes que involucran sustancias peligrosas. En este contexto, este es el primer estudio que propone un método basado en aprendizaje automático para predecir las consecuencias de los accidentes que involucran sustancias peligrosas en términos de personas muertas y heridas. Solo se encontró un estudio similar en la literatura (Chebila, 2021), que solo consideró si el accidente dañó o no a las personas, por lo que carece del nivel de detalle proporcionado en esta investigación. Además, este estudio utiliza un conjunto de múltiples variables de resultado discretas para estimar el número de personas involucradas, proporcionando así un resultado mucho más detallado y valioso.

2.5.11.1. Base de datos de accidentes y selección de características.

Los datos sobre accidentes se extraen de la fuente de datos y se almacenan en un formato conveniente, como un archivo CSV. La base de datos tiene una forma similar a una matriz, donde cada fila representa un evento y cada columna un atributo del evento (por ejemplo, la fecha, la sustancia involucrada, el tipo de incidente).

Algunos de los atributos incluidos en la base de datos pueden no ser significativos o útiles para los análisis; estos atributos deben eliminarse. Ver Figura 6. En general, la base de datos debe contener solo atributos que vinculen las características de los eventos con las consecuencias de estos. Después de eliminar los atributos innecesarios, la base de datos debe estar preparada para las simulaciones de aprendizaje automático como se muestra en la Figura 6. Esta tarea requiere tres pasos:

- Los datos faltantes deben imputarse o eliminarse porque la mayoría de los modelos de aprendizaje automático no pueden procesar valores nulos. Se han desarrollado diferentes técnicas para imputar o eliminar valores faltantes según el tipo y las características de los datos (es decir, numéricos o categóricos, aleatorios o no aleatorios). Se puede encontrar una descripción general de los métodos más utilizados en (Brink et al., 2016). En este estudio, los valores faltantes se han sustituido por la cadena definida por el usuario "Na". Esto debería permitir que el modelo maneje la incertidumbre y aprenda el impacto de los valores faltantes en la medida del resultado.
- Los atributos que pueden contener más de una entrada deben dividirse para que cada columna de la base de datos contenga solo una entrada.
- Los atributos que indican el Número de Personas Heridas (NPI) y Muertas (NPK) deben ser convertidos a sus respectivas categorías de gravedad. Para ello, se considera un conjunto de categorías de consecuencias que reflejan las categorías de gravedad utilizadas por las matrices de riesgo y otros métodos de análisis de riesgo. Ver la Tabla 2.

Tabla 2*Categorías de Consecuencias de Accidentes*

Categoría	Descripción
NO	Ningún muerto/herido
10-Jan	De 1 a 10 muertos/heridos
10 – 100	De 10 a 100 muertos/heridos
-1000	De 100 a 1000 muertos/heridos
> 1000	Más de 1000 muertos y heridos

Nota. Estos datos es extraída de: (Brink et al., 2016).

Tras estos pasos se obtiene una versión limpia de la base de datos original, que finalmente se utiliza para las simulaciones. Los algoritmos de *Machine Learning* se entrenan para clasificar los accidentes en una de las categorías descritas en la Tabla 2, prediciendo así la gravedad de los eventos accidentales con un alto nivel de detalle.

2.5.11.2. Simulaciones de Aprendizaje Automático.

El aprendizaje automático (ML) se refiere a una clase de algoritmos informáticos diseñados para adquirir experiencia a partir de datos y aprovechar el conocimiento adquirido para realizar predicciones precisas, revelar correlaciones entre variables e identificar patrones y tendencias ocultos (Brink et al., 2016). En otras palabras, el aprendizaje automático se ocupa de entrenar a una máquina para que aprenda de la comprensión previa (U.S. Green Building Council, n.d.).

Existen tres macro categorías de algoritmos de aprendizaje automático: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo (Murphy, 2012). El aprendizaje supervisado se utiliza cuando el problema implica la predicción de una medida de resultado basada en una o más variables de entrada (Hastie et al., 2009). En cambio, si no es aplicable ninguna medida de salida, se pueden utilizar algoritmos de aprendizaje no supervisado para analizar datos de entrada y revelar relaciones y patrones con poca o ninguna

intervención humana (Jukes, 2018). En el aprendizaje por refuerzo, el alumno (por ejemplo, un robot industrial) no analiza pasivamente los datos de entrada; en cambio, recopila datos del entorno a través de un conjunto de acciones, y se utiliza un sistema de recompensas para guiar el proceso de aprendizaje (Stone, 2017).

En este estudio, tanto la entrada (es decir, las características de un evento) como la medida del resultado (es decir, la gravedad del evento) están disponibles y se informan en la fuente de datos. Por lo tanto, los algoritmos de aprendizaje supervisado son una opción natural. Además, el objetivo de este estudio es categorizar (es decir, clasificar) los accidentes en función de la gravedad de sus consecuencias, que pueden expresarse en términos de la cantidad de personas que mueren o resultan heridas en el evento; por este motivo, se realizan dos conjuntos distintos de simulaciones. Por lo tanto, el problema es una tarea de clasificación. Sin embargo, también puede ser posible un enfoque de regresión y debería investigarse mediante investigaciones adicionales.

2.5.11.3. Clasificación: Formación y Evaluación.

El objetivo de un algoritmo de clasificación es clasificar objetos en dos o más categorías (Drummond, 2017). Un objeto se describe mediante un conjunto de características (es decir, atributos significativos del objeto, digamos X) y una etiqueta (es decir, su categoría, digamos Y); en este estudio, las liberaciones de sustancias peligrosas son los objetos.

En primer lugar, la base de datos limpia se divide en dos partes: la base de datos de entrenamiento y la base de datos de evaluación (paso 2.2 en la Figura 1). La primera comprende el 80% de los eventos, y la parte restante (20%) forma la segunda. A continuación, la base de datos de entrenamiento se alimenta al algoritmo, que ajusta los parámetros internos de una función f para encontrar el mapeo óptimo entre las características y las etiquetas

correspondientes (James et al., 2013). La función f también se denomina modelo del algoritmo de aprendizaje automático (Moretti et al., 2016).

$$Y \approx f(X) \quad \dots (12)$$

Dónde: $X = \text{Matriz } N \times M$ de las características. N es el número de objetos y M es el número de características; $Y = N \times 1$ vector de las etiquetas; $f =$ función con parámetros ajustables. Esta fase es la llamada fase de entrenamiento (Figura 6). A continuación, los objetos sin etiquetar se introducen en el modelo entrenado, que predice las etiquetas correspondientes según la siguiente ecuación.

$$f(X_i) = \hat{Y} \quad \dots (13)$$

Dónde: $X_i = 1 \times M$ vector de las características del objeto no etiquetado i

$\hat{Y} =$ probabilidades de etiqueta producidas por el modelo para el objeto i .

Por último, las etiquetas predichas se comparan con las etiquetas verdaderas para evaluar el rendimiento del modelo. Esta fase es la denominada fase de evaluación (Figura 6). El lote de objetos que se utiliza para evaluar el algoritmo es la base de datos de evaluación.

Vale la pena señalar que la salida del modelo (es decir, \hat{Y}_i) no es una etiqueta única sino un vector que contiene las probabilidades de la etiqueta (James et al., 2013). En otras palabras, si K categorías diferentes son posibles, \hat{Y}_i es un vector $K \times 1$ cuyos elementos representan la probabilidad de cada categoría. Para convertir las probabilidades de las etiquetas en una etiqueta predicha, se utiliza un umbral de decisión de probabilidad (Retrieval, 2009), que suele ser 0.5 por defecto.

2.5.11.4. Modelos.

Existen distintos modelos para realizar una tarea de clasificación. En este estudio, se utilizan un modelo lineal, una red neuronal profunda y un modelo híbrido de ancho y profundidad para demostrar el enfoque.

2.5.11.5. Modelo Lineal.

El modelo lineal representa las etiquetas como una combinación lineal de características (James et al., 2013). Por lo tanto, la ecuación (1) se puede escribir como:

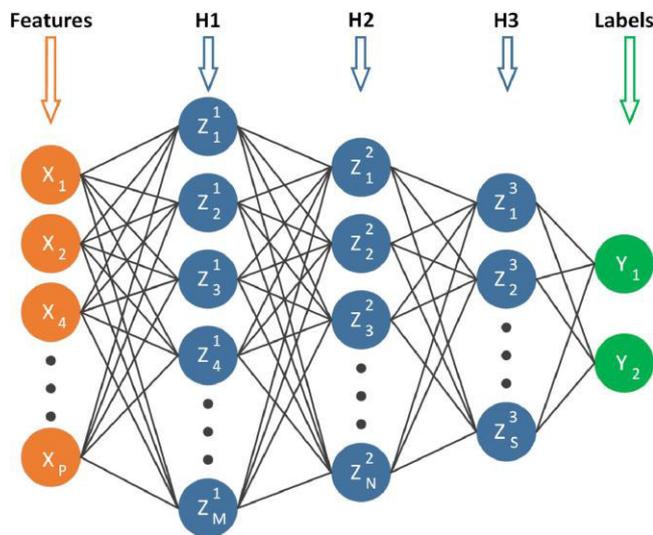
$$Y \approx \beta_o + \sum_{j=1}^M \beta_j X_j \quad \dots (14)$$

Dónde: Y = etiqueta; β_o = sesgo; X_j = una característica; β_j = peso de la-ésima característica.

Los modelos lineales son robustos, rápidos, fáciles de interpretar y adecuados para analizar grandes conjuntos de datos (Brink et al., 2016). Por otro lado, no pueden capturar relaciones no lineales entre características. Además, los modelos lineales no pueden inferir el impacto de combinaciones de características que no han ocurrido en el pasado (Cheng et al., 2016).

2.5.11.6. Red neuronal Profunda.

Las redes neuronales profundas (DNN) son modelos gráficos acíclicos dirigidos que constan de unidades densamente interconectadas (Goodfellow et al., 2016). En la Figura 6 se muestra una representación visual de una DNN. Representación esquemática de una red neuronal profunda. Los círculos naranja, azul y verde representan características de entrada (X_i), unidades ocultas (Z_i^j), y etiquetas Y_k .

Figura 6*Red Neuronal Profunda*

Nota. La red neuronal profunda ha sido adaptado de (Tamascelli et al., 2022).

En estos modelos, las características de un objeto (círculos naranjas en la Figura 6) se convierten en probabilidades de etiqueta (círculos verdes en la Figura 6) a través de una serie de combinaciones lineales y transformaciones no lineales (Hastie et al., 2009). Entre las capas de entrada y salida, una serie de unidades ocultas interconectadas (círculos azules en la Figura 6) se organizan en una o más capas ocultas (por ejemplo, H1, H2 y H3 en la Figura 6). La unidad de una capa oculta genérica H_i se obtiene mediante una transformación no lineal de las unidades combinadas linealmente de la capa anterior. En este estudio, se utiliza la Unidad Lineal Rectificada (Unidas, 2003) para realizar la transformación no lineal. Se pueden encontrar más detalles y fórmulas detrás de las Redes Neuronales en (Goodfellow et al., 2016). Las DNN tienen buenas capacidades de generalización y pueden capturar relaciones no lineales entre características (Goodfellow et al., 2016). Como desventaja, son sensibles a datos de entrada de mala calidad y son propensas al sobreajuste y la sobre generalización (Brink et al., 2016). Además, el costo computacional requerido para entrenar una DNN es mayor en comparación con modelos más simples (Goodfellow et al., 2016).

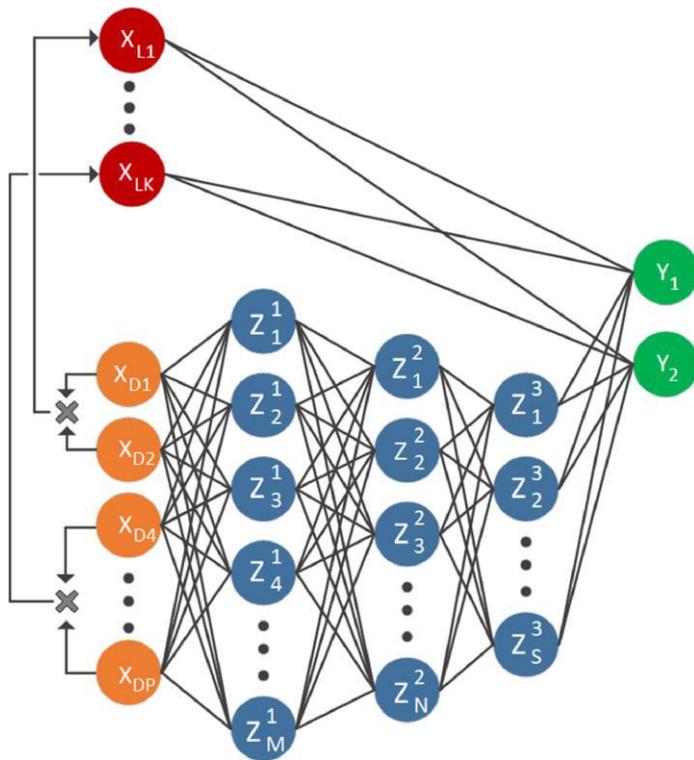
2.5.11.7. Ancho y Profundo.

En un intento de combinar las ventajas de los modelos lineal y profundo, Cheng et al., (2016) desarrollaron el modelo Wide&Deep, cuya estructura se muestra en la Figura 7.

El modelo consta de una parte lineal y una parte profunda (parte inferior de la Figura 7). Durante la fase de entrenamiento, los modelos lineal y profundo se entrenan conjuntamente, es decir, las etiquetas predichas (círculos verdes en la Figura 7) se obtienen combinando las salidas de ambos modelos, y los pesos de los modelos se optimizan simultáneamente (Cheng et al., 2016). Por lo general, la parte lineal del modelo toma como entrada un pequeño conjunto de características cruzadas (Cheng et al., 2016), que son características sintéticas obtenidas tomando el producto cartesiano de dos o más características (Cheng et al., 2016). Por el contrario, la parte profunda utiliza todas las características disponibles (X_D en la Figura 7). Por lo tanto, la parte profunda es un modelo DNN de tamaño completo, mientras que la parte lineal integra y "complementa las debilidades de la parte profunda con un pequeño número de productos cruzados" (Cheng et al., 2016). Como ejemplo, la característica X_{L1} en la Figura 7 se obtiene cruzando X_{D1} y X_{D2} . En general, la naturaleza híbrida del modelo Wide&Deep garantiza buenas capacidades de memorización (parte lineal) y generalización (parte profunda). Ver la Figura 7.

Figura 7

Representación Esquemática de un Modelo Wide&Deep



Nota. La representación esquemática del modelo ha sido adaptado de (Tamascelli et al., 2022).

2.5.11.8. Métricas de Rendimiento.

El rendimiento de un algoritmo de clasificación se evalúa durante la fase de evaluación. Por ejemplo, la clasificación puede considerar las clases “Y” y “N”, respectivamente, positiva y negativa. Siempre que el modelo predice la clase de un objeto, hay cuatro resultados posibles:

- TP = Verdadero positivo, es decir, etiqueta prevista = Y, etiqueta verdadera = Y;
- TN = Verdadero negativo, es decir, etiqueta prevista = N, etiqueta verdadera = N,
- FP = Falso positivo, es decir, etiqueta prevista = Y, etiqueta verdadera = N;
- FN = Falso Negativo –es decir, etiqueta predicha = N, etiqueta verdadera = Y.

La suma de verdaderos positivos y verdaderos negativos representa el número de predicciones correctas, mientras que la suma de falsos positivos y falsos negativos indica el número de predicciones erróneas.

Los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos se utilizan para obtener tres indicadores de rendimiento:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \dots (15)$$

$$Precision = \frac{TP}{TP + FP} \quad \dots (16)$$

$$Recall = \frac{TP}{TP + FN} \quad \dots (17)$$

La exactitud representa la fracción de objetos que se han clasificado correctamente. La precisión indica la tasa de éxito de una predicción positiva. La recuperación denota la fracción de positivos reales que se han identificado correctamente.

La precisión por sí sola no es informativa si el problema implica la identificación de clases raras, es decir, cuando el conjunto de datos está desequilibrado en cuanto a clases (Cheng et al., 2016); en estas situaciones, la precisión y la recuperación son más representativas del rendimiento del modelo (Cheng et al., 2016). Además, si el costo de un falso negativo es mayor que el costo de un falso positivo, la recuperación es la métrica más significativa. En lugar de considerar la precisión y la recuperación individualmente, se pueden agregar en el llamado puntaje F (Cheng et al., 2016).

$$F_{\beta} = (1 + \beta^2) \times \frac{Precision \times Recall}{(\beta^2 \times Precision) + Recall} \quad \dots (18)$$

Dónde: β = número real no negativo. Si $\beta = 1$, la puntuación representa la media armónica entre la precisión y la recuperación (Sarkar et al., 2020). Si $\beta > 1$, la puntuación

está orientada al recuerdo (Sasaki, 2007), lo que significa que se considera que el recuerdo es veces más importante que la precisión.

Por último, cabe mencionar que las métricas e indicadores presentados dependen del umbral de decisión de probabilidad (sección 2.2.1). De hecho, el umbral de decisión podría ajustarse para optimizar el modelo (Cheng et al., 2016). Por ejemplo, si se reduce el umbral de decisión, el modelo puede producir predicciones más positivas. Como resultado, el *Recall* puede aumentar, pero la Precisión puede disminuir (Schwalbe, 2016; Z. Xu & Saleh, 2021). De hecho, las acciones destinadas a aumentar el *Recall* a menudo reducen la precisión, y viceversa (Cheng et al., 2016). Un medio conveniente para mostrar el efecto del umbral de decisión es la curva de precisión-recuperación, es decir, un gráfico donde cada punto representa la pareja precisión vs. recuperación en un umbral de decisión específico (Murphy, 2012). Un medio conveniente para resumir la información en la curva de precisión-recuperación es el área bajo la curva (AUC PR) (Murphy, 2012), que toma valores entre 0 y 1. Al ser independiente del umbral de decisión, el AUC PR se considera un indicador más completo del rendimiento del modelo si se compara con la exactitud, la precisión y la recuperación. En general, un valor alto de AUC PR indica un buen rendimiento.

2.6. El Aprendizaje Automático (ML)

El aprendizaje automático (ML) es una clase de inteligencia artificial (IA) que se centra en enseñar a las computadoras a hacer predicciones a partir de conjuntos de datos y algoritmos disponibles. Lo más importante es que proporciona a los sistemas informáticos la capacidad de aprender y mejorar por sí mismos en lugar de estar programados explícitamente. Aunque el ML nació en 1943 y se acuñó por primera vez en 1959, en realidad comenzó a florecer en la década de 1990 y se ha convertido en el subcampo de IA más exitoso. El ML también se ha convertido en una de las palabras de moda en tecnología de nuestra era, ya que desempeña un

papel fundamental en muchas aplicaciones del mundo real, como el reconocimiento de imágenes y voz, las alertas de tráfico, los automóviles autónomos, el diagnóstico médico, etc.

En general, el aprendizaje automático se puede clasificar en tres categorías principales según el proceso de aprendizaje: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo. El aprendizaje supervisado es el tipo más básico de aprendizaje automático, cuyo algoritmo se entrena a partir de un conjunto de datos etiquetados. Este método es adecuado para problemas de regresión y clasificación, y se ha utilizado ampliamente en ingeniería estructural para la detección de daños (problemas de clasificación) y predicciones de resistencia (problemas de regresión). Por el contrario, el algoritmo utilizado en el aprendizaje no supervisado se entrena a partir de un conjunto de datos no etiquetados. Mientras tanto, en el método de aprendizaje por refuerzo, el algoritmo se entrena a través de un proceso de prueba y error. Se ha adoptado una cantidad significativa de algoritmos de ML en aplicaciones de ingeniería estructural, por ejemplo, redes neuronales (NN), árboles de decisión (DT), análisis de regresión (RA), máquinas de vectores de soporte (SVM), bosques aleatorios (RF), algoritmos de refuerzo (BA), etc. El modelo sustituto, también conocido como metamodelo, es un caso especial de ML supervisado que se ha utilizado ampliamente en el campo del diseño de ingeniería para reducir el tiempo computacional de modelos ML de caja negra complejos con precisión relajada. Es un modelo interpretable que se entrena para aproximar las predicciones de un modelo ML de caja negra. En otras palabras, los modelos sustitutos son modelos analíticos simples que imitan el comportamiento de modelos ML complejos.

2.6.1. Modelo ML

Esta sección proporciona los conceptos y las herramientas prácticas para implementar métodos de ML. Abarca una amplia gama de algoritmos de ML que se utilizan ampliamente en el ámbito de la ingeniería estructural. Además, también se proporcionan bibliotecas de

Python, códigos de fuente abierta y conjuntos de datos para ML para que los lectores practiquen y ejecuten sus modelos de ML. Ver la Figura 8 de la página 98.

2.6.2. Descripción General del Aprendizaje Automático

El aprendizaje automático es el proceso de enseñar a un sistema informático (es decir, un modelo de aprendizaje automático) a realizar una predicción precisa al introducir nuevos datos. La Figura 8, ilustra un flujo de trabajo típico de aprendizaje automático utilizado en el modelado predictivo. Mediante el uso de un algoritmo de aprendizaje y datos iniciales, se entrena al sistema informático para que pueda aprender y mejorar hasta alcanzar su rendimiento. Por lo tanto, la precisión de un modelo de aprendizaje automático depende en gran medida de las características de los datos iniciales y del rendimiento del algoritmo de aprendizaje. Hay tres pasos principales para construir un modelo de aprendizaje automático: preparar la base de datos, aprender y evaluar el modelo.

Paso 1: Los datos iniciales utilizados para construir un modelo ML se presentan generalmente en forma de variables de entrada y de salida correspondientes que se caracterizan en la terminología ML por característica (variable de entrada) y etiqueta (variable de salida). Al predecir el comportamiento de una estructura, por ejemplo, su dimensión geométrica y propiedades del material se clasifican como características, mientras que su resistencia máxima y su deflexión se utilizan como etiquetas. Algunos algoritmos ML requieren que todos los datos de entrada se escalen en el rango $[0,1]$ para tener un mejor rendimiento (Carvalho et al., 2019). Para probar el rendimiento de los modelos ML, los datos iniciales se dividen aleatoriamente en conjuntos de datos de entrenamiento y prueba, y la mayor parte se utiliza para fines de entrenamiento.

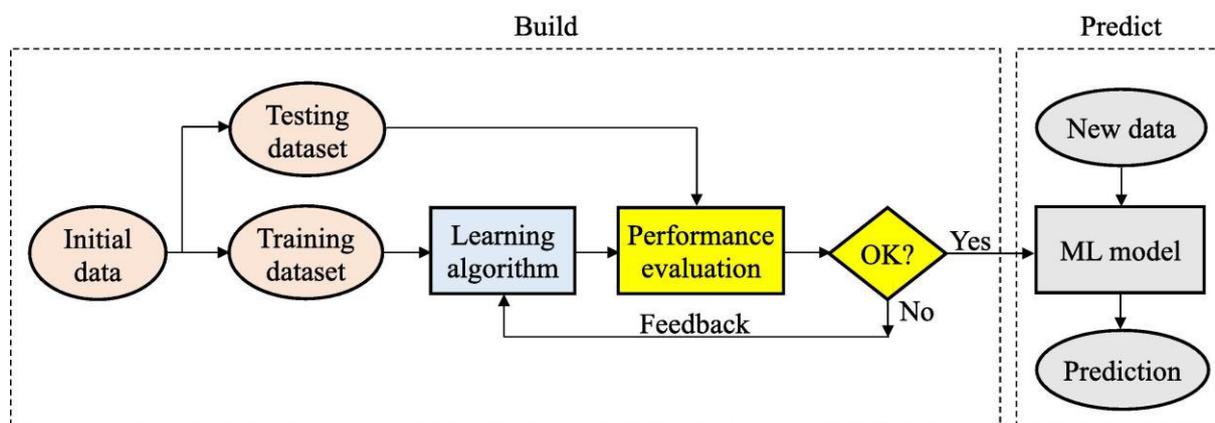
Paso 2: El objetivo de este paso de aprendizaje es entrenar un algoritmo de ML seleccionado. En la literatura se han desarrollado una gran cantidad de algoritmos de ML para

aplicaciones específicas (en la Sección 2.2 se ofrece una explicación detallada de cada algoritmo). Por lo tanto, es importante comparar diferentes algoritmos para encontrar el mejor para problemas específicos. Luego, los algoritmos seleccionados se entrenan utilizando el conjunto de datos de entrenamiento obtenido en el Paso 1.

Paso 3: Una vez que un modelo de ML está completamente entrenado, se evalúa su rendimiento utilizando el conjunto de datos de prueba. Se utiliza una función de pérdida como indicador de rendimiento para medir qué tan lejos está un valor predicho de su valor real. Las funciones de pérdida típicas para problemas de regresión son el error absoluto medio (MAE) y el error cuadrático medio (MSE). Las funciones de pérdida desempeñan un papel fundamental en la evaluación de los modelos de ML y, por lo tanto, la elección de la función de pérdida correcta también determina qué tan bueno será el modelo.

Figura 8

Flujo de Trabajo Típico de ML



Nota. El flujo de trabajo típico ha sido obtenido la fuente: (Kang & Ryu, 2019)

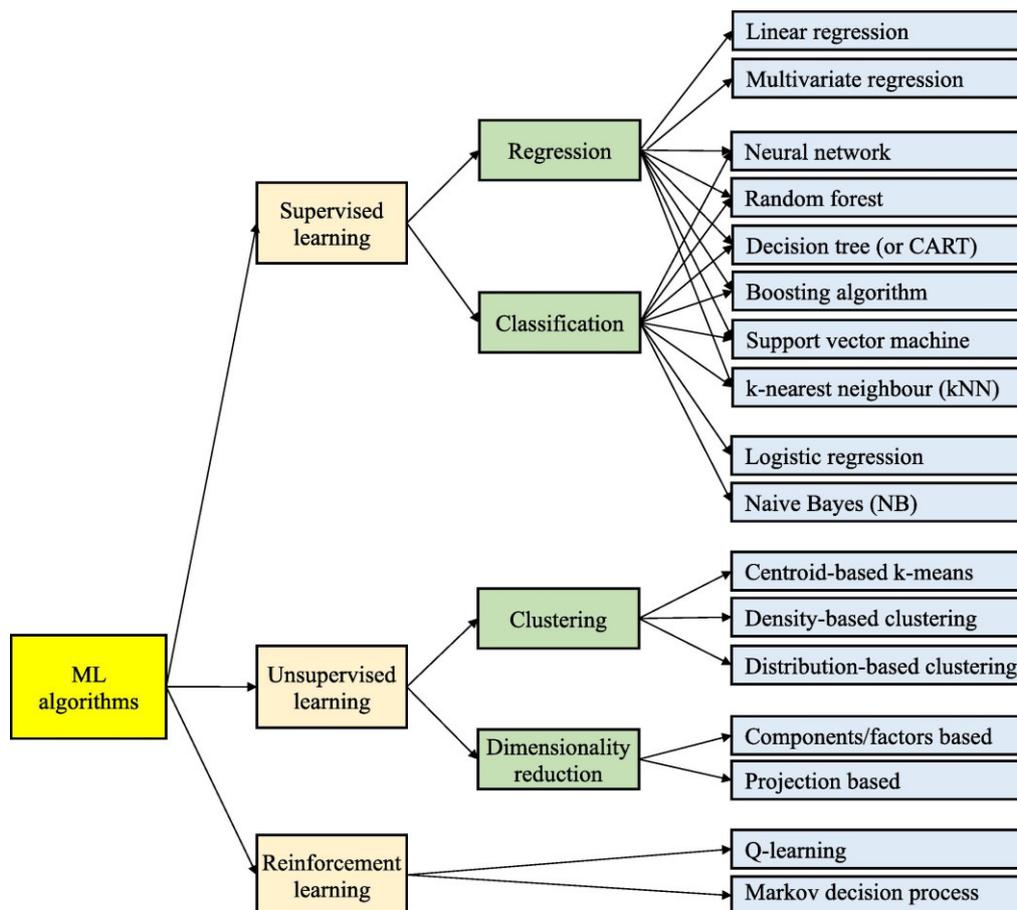
2.6.3. Algoritmos de Aprendizaje Automático

En la actualidad, se han desarrollado muchos algoritmos de aprendizaje automático en la literatura, como se muestra en la Figura 9. Cada algoritmo, con sus fortalezas y debilidades, está diseñado para ciertos tipos de métodos y problemas de aprendizaje. Sin embargo, esta

sección solo analiza los algoritmos que se utilizan comúnmente en ingeniería estructural. En cada algoritmo, se explica claramente su concepto, lo que resulta útil en la práctica para personas de ingeniería estructural sin experiencia en aprendizaje automático.

Figura 9

Algoritmo ML Agrupados por Tipo de Aprendizaje



Nota. El algoritmo de aprendizaje ha sido obtenido de la fuente: (Kang & Ryu, 2019).

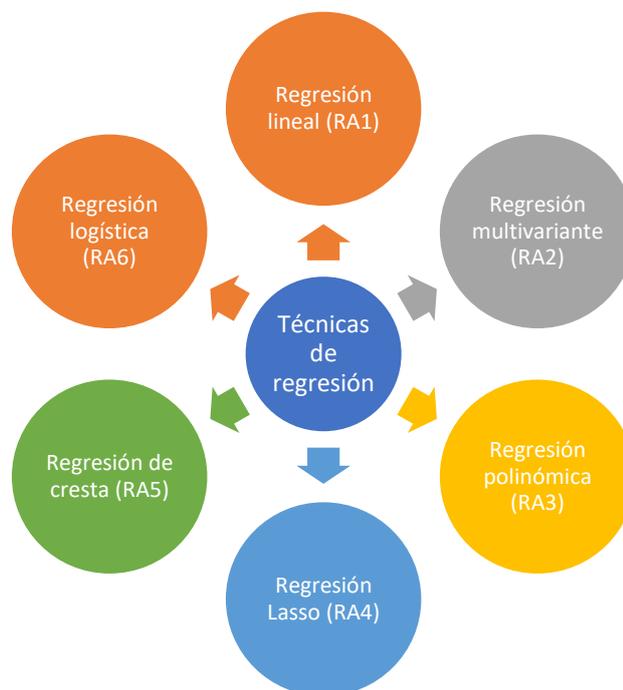
2.6.4. Técnicas de Regresión

El RA es una técnica de modelado predictivo que se desarrolló por primera vez en estadística para estudiar la relación entre variables independientes (predictores) y variables dependientes (objetivos). Este método se aplicó luego en ML bajo el algoritmo de aprendizaje

supervisado para predecir los valores de salida en función de los valores de las variables de entrada. Existen diferentes tipos de modelos de regresión desarrollados en ML en función de (i) el número de variables, (ii) el tipo de variables y (iii) la forma de la línea de regresión. Los modelos RA que se utilizan comúnmente en ingeniería estructural incluyen (Gerón, 2023).

Figura 10

Técnicas de Regresión



Nota. Las técnicas de regresión son de 6 tipos como se muestra en la figura.

2.6.4.1. Regresión Lineal (RA1).

Este es el modelo de regresión más simple en ML, en el que la variable de salida y la variable de entrada se ajustan mejor en una línea recta (función lineal). Los coeficientes de la ecuación lineal se determinan minimizando la función de costo (por ejemplo, MSE y MAE) definida como la diferencia (error) entre el valor predicho y el valor real. Si se utiliza una sola variable de entrada, el modelo se denomina regresión lineal simple. En el caso de más de una variable de entrada, se denomina regresión lineal múltiple (Gerón, 2023). Ver la Figura 10.

2.6.4.2. Regresión Multivariante (RA2).

La regresión multivariante es una extensión de la regresión lineal múltiple cuando se trata de problemas que tienen más de una variable de salida. La palabra “multivariante” se refiere a más de una variable de salida, mientras que la palabra “múltiple” se refiere a más de una variable de entrada. El mérito de este método es que ayuda a comprender la correlación entre las variables de entrada y salida. Este método también se usa ampliamente en ML para problemas de regresión (Gerón, 2023).

2.6.4.3. Regresión Polinómica (RA3).

La diferencia entre el modelo RA3 y el modelo RA1 es la forma de la línea de regresión. La línea de mejor ajuste en RA3 es una línea curva (función polinómica) con la potencia de las variables de entrada mayor que uno. Puede producirse un sobreajuste en este modelo si las variables de entrada se ajustan mediante un polinomio de grado superior para obtener un error menor. Por lo tanto, es útil trazar el modelo para asegurarse de obtener resultados razonables (Gerón, 2023).

2.6.4.4. Regresión Lasso (RA4).

La regresión LASSO (operador de selección y contracción mínima absoluta) es una versión regularizada de RA1 que se utiliza cuando las variables de entrada están altamente correlacionadas. En este caso, el uso de la técnica de regresión lineal podría resultar en un sobreajuste. Por lo tanto, se propone RA4 para reducir el sobreajuste agregando un término de regularización en la función de costo durante el entrenamiento. Esto mantiene el peso del modelo lo más pequeño posible (Gerón, 2023). El término de regularización utilizado en RA4 es la norma L1 (valor absoluto del peso).

2.6.4.5. Regresión de Cresta (RA5).

De manera similar a RA4, RA5 también es una versión regularizada de RA1. Sin embargo, el término de regularización utilizado en RA5 es la norma L2 (valor al cuadrado del peso) en lugar del término L1 utilizado en el caso de RA4. El objetivo de la técnica RA5 es tratar de eliminar el peso de las características menos importantes (Gerón, 2023).

2.6.4.6. Regresión Logística (RA6)

Esta técnica de regresión se desarrolló para problemas de clasificación cuando la variable de salida es de naturaleza binaria o discreta (por ejemplo, Verdadero/Falso, 1/0, Sí/No, etc.). En RA6, la relación entre las variables de entrada y salida se expresa mediante una función logística también conocida como función sigmoidea. Este método se basa en el concepto de probabilidad. Significa que RA6 genera un valor binario de 0 (cuando la probabilidad estimada es menor del 50%) o 1 (cuando la probabilidad estimada es mayor del 50%) en lugar de un valor numérico (Gerón, 2023).

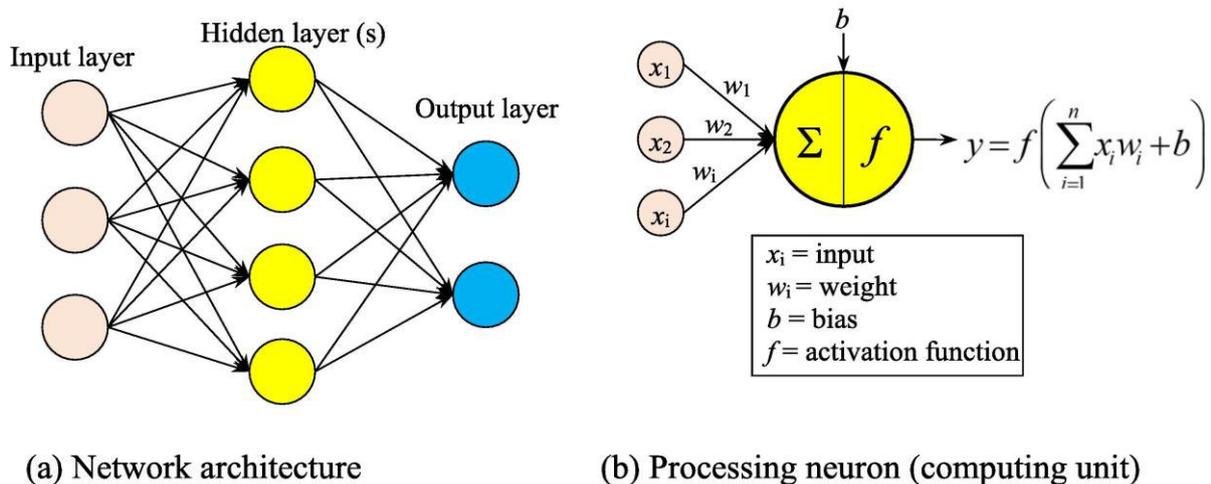
2.6.5. *Redes Neuronales y sus Variantes*

La red neuronal artificial (ANN) se desarrolló para imitar el funcionamiento de las neuronas biológicas. La primera ANN fue inventada por Rosenblatt (Rosenblatt, 2021) en 1958 llamada perceptrón para problemas de reconocimiento de patrones. Gracias a la mejora en la potencia informática, la ANN se ha convertido en uno de los algoritmos de aprendizaje automático más populares en la actualidad con varias variantes, como la red neuronal de avance (FFNN) (Ivakhnenko, 1971) mejorada por el perceptrón multicapa (MLP), la red neuronal de función de base radial (RBFNN) (Lowe, 1988), CNN (Lecun et al., 2023), la red neuronal recurrente (RNN) (Elman, 1990) mejorada por la memoria a largo plazo (LSTM) (Hochreiter & Schmidhuber, 1997) y el sistema de inferencia neuro difusa adaptativa (ANFIS) (Jang, 1993). FFNN es el primer y más simple tipo de ANN en el que la información se transfiere

solo en una dirección (hacia adelante) desde los nodos de entrada a los nodos de salida, mientras que MLP es una versión mejorada de FFNN con múltiples capas de unidades de cómputo que incluyen una capa de entrada, una o más capas ocultas y una capa de salida como se muestra en la Figura 11. RBFNN es una ANN en la que se utiliza una función de base radial (RBF) como función de activación. CNN está desarrollada específicamente para el reconocimiento de imágenes (ver Ref. (Sony et al., 2021) para una revisión crítica sobre el uso de CNN en la detección de grietas basada en imágenes), mientras que RNN está diseñada para interpretar información temporal o secuencial. ANFIS es una combinación de la capacidad de aprendizaje de una red neuronal adaptativa y las capacidades de razonamiento de un sistema de inferencia difusa. Una revisión histórica detallada de ANN y sus variantes se puede encontrar en Schmidhuber (Hochreiter & Schmidhuber, 1997).

Figura 11

Ejemplo de una ANN



Nota. El ejemplo de la ANN ha sido obtenido de la fuente: (Firmansyah et al., 2024)

2.6.5.1. Red Neuronal Artificial RNA - Artificial Neural Network (ANN).

Una ANN estructurada como el cerebro humano consta de neuronas artificiales también conocidas como unidades o nodos. Estos nodos están completamente interconectados y

dispuestos en tres capas diferentes como se muestra en la Figura 11(a). La capa de entrada recibe los datos de entrada x , mientras que la capa de salida representa los resultados previstos y de la red. Entre las capas de entrada y salida se encuentran una o más capas ocultas de unidades de computación (Figura 11(b)) que realizan los principales cálculos matemáticos sobre los datos de entrada. Cuando una ANN está diseñada con dos o más capas ocultas, se denomina perceptrón multicapa o aprendizaje profundo (DL), un subcampo específico del ML basado en NN (Lecun et al., 2023).

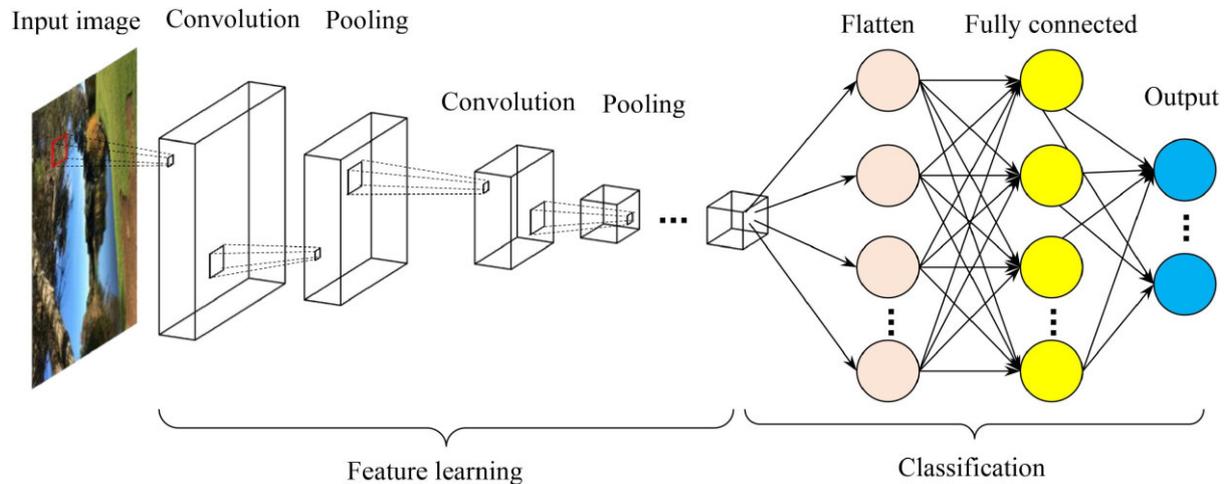
El comportamiento de cada unidad neuronal se define por los pesos w que se le asignan. Cuando los datos x_i se introducen en la capa de entrada, se multiplican por los pesos correspondientes w_i . A continuación, se utiliza una función de transferencia para calcular la suma ponderada de la entrada más un sesgo b , que se ajusta gradualmente para minimizar la diferencia entre las salidas previstas y las reales. El valor de la función de transferencia se pasa entonces a través de una función de activación f para comprobar si el nodo debe transmitir datos a la capa de salida o no. Cuando se entrena una ANN, se asignan valores aleatorios de pesos a todos los nodos. Una vez que la función de activación pasa el valor previsto y a la capa de salida, se utiliza una función de error para calcular la diferencia entre las salidas previstas y las reales. Con base en el resultado, el modelo de ANN ajusta los pesos de todos sus nodos para minimizar el error. Este proceso de entrenamiento, conocido como retro propagación (Rumelhart et al., 1986), se itera hasta que se satisface la condición de convergencia.

Hay una serie de hiper parámetros que deben optimizarse al entrenar un modelo de ANN porque controlan el proceso de aprendizaje y entrenamiento de la red. Incluyen las opciones arquitectónicas de una ANN (es decir, la cantidad de capas ocultas utilizadas, la cantidad de nodos por capa oculta y el tipo de funciones de activación) y las variables de entrenamiento (es decir, la tasa de inclinación, la cantidad de épocas, el momento y el tamaño del lote). Aumentar la cantidad de capas y nodos ocultos puede aumentar la precisión de la red,

pero causa un costo computacional. La función de activación se utiliza para tener en cuenta la no linealidad de los modelos. La tasa de aprendizaje define la rapidez con la que el modelo de ANN actualiza sus parámetros. El uso de una tasa de aprendizaje alta acelerará el aprendizaje, pero puede fallar en la convergencia. La época controla la cantidad de iteraciones para el conjunto de datos de entrenamiento, mientras que el momento se utiliza para elegir la dirección del siguiente paso a partir del paso anterior, lo que puede ayudar a prevenir oscilaciones. Estos hiperparámetros se ajustan durante las etapas de prueba y validación.

2.6.5.2. Convolutional Neural Network CNN.

Este algoritmo fue introducido por primera vez en la década de 1980 por Yann LeCun, un investigador francés en ciencias de la computación que construyó una de las primeras versiones de arquitecturas CNN para reconocimiento digital escrito a mano llamada LeNet (Lecun et al., 2023). La idea básica de CNN es crear una red en la que cada capa puede convertir información de las capas anteriores en información más compleja y transformarla en las siguientes capas. Como se muestra en la Figura 12, una CNN se construye en base a dos bloques básicos: el bloque de aprendizaje de características y el bloque de clasificación. El aprendizaje de características, que se compone de varias capas alternas de convolución y agrupación, se utiliza para extraer y aprender la característica de la imagen de entrada. La característica extraída o aprendida se clasifica luego a través del bloque de clasificación. También se han desarrollado diferentes arquitecturas CNN para mejorar el rendimiento de CNN en varias aplicaciones. Entre ellas se destacan LeNet, AlexNet, VGG y ResNet. Se puede encontrar una revisión completa sobre la evolución de las arquitecturas CNN en las referencias (Alzubaidi et al., 2021). Ver la Figura 12.

Figura 12*Arquitectura CNN Típica*

Nota. La arquitectura típica ha sido obtenida de: (Firmansyah et al., 2024)

CNN se considera uno de los mejores algoritmos de ML para el reconocimiento de imágenes. Una revisión detallada sobre el desarrollo de CNN para la aplicación de la clasificación de imágenes se puede encontrar en Wang & Yang, (2019). En el contexto de SHM, CNN se ha utilizado ampliamente para detectar grietas en estructuras basadas en enfoques de clasificación de imágenes o técnicas de segmentación. El método de clasificación de imágenes detecta grietas a nivel de imagen en lugar de a nivel de píxel utilizado en el método de segmentación. Almeida et al. (2016) y Sony et al. (2021) informaron revisiones sobre el uso de CNN para la detección de grietas estructurales y la evaluación del estado, respectivamente.

2.6.5.3. Red neuronal con Función de Base Radial RBFNN.

Esta es una ANN específica que utiliza RBF como función de activación, como se propuso en uno de los primeros trabajos de Broomhead y Lowe (Lowe, 1988). La RBFNN tiene solo una capa oculta llamada vector de características, y su salida es una combinación lineal de RBF de las entradas y los parámetros neuronales (es decir, peso y sesgo). El mérito de RBFNN sobre una NN regular es su capacidad de entrenamiento rápido gracias a la

aproximación universal de RBF. La única preocupación al usar RBFNN es cómo elegir correctamente los parámetros de forma y los centros de RBF (Wu, 2021).

2.6.5.4. Sistema de Inferencia Neuro Difuso Adaptativo ANFI.

Este algoritmo combina la técnica de control adaptativo de los sistemas neuro difusos y la capacidad de aprendizaje de las ANN. Por lo tanto, ANFIS puede aprovechar los méritos tanto de la lógica difusa como de la NN para minimizar los errores entre los datos de entrada y los datos predichos por el sistema neuro difuso. Una arquitectura típica de ANFIS consta de cinco capas, a saber, capa difusa, capa de producto, capa normalizada, capa de des difusa y capa de salida general. La primera capa recibe los valores de entrada y determina sus funciones de pertenencia utilizando parámetros de premisa. La segunda capa produce la fuerza de activación de las reglas y, por lo tanto, esta capa también se denomina "capa de reglas". La fuerza de activación luego se normaliza en la tercera capa. La cuarta capa recibe los valores normalizados y los parámetros de consecuencia, y luego los pasa a la última capa para la salida final general. Los parámetros ajustables (es decir, premisa y consecuencia) se identifican durante el proceso de aprendizaje para minimizar el error entre la salida real y la salida deseada. Por lo tanto, juegan un papel vital a la hora de decidir el rendimiento de ANFIS.

2.6.5.5. Máquina de Vectores de Soporte.

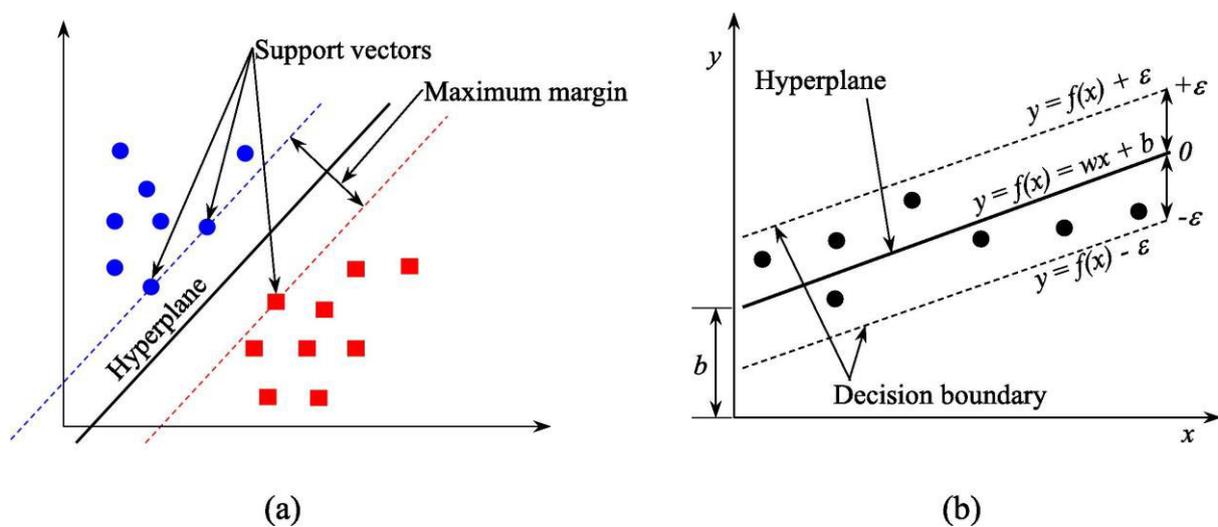
SVM es uno de los algoritmos más potentes y populares debido a su precisión y simplicidad. Este algoritmo fue introducido por primera vez para problemas de clasificación por Boser et al. (1992) en 1992 utilizando clasificadores no lineales. Ben et al. (2001) luego extendieron el algoritmo SVM a casos donde los datos son no linealmente separables utilizando clasificadores de margen suave. SVM también fue expandido para problemas de regresión (Vapnik, 2013) conocidos como regresión de vector de soporte (SVR) y tareas de agrupamiento

(Ben et al., 2001) conocidas como agrupamiento de vector de soporte (SVC). Sin embargo, la técnica SVM es ampliamente utilizada para propósitos de clasificación.

La idea básica detrás del algoritmo SVM es distinguir entre grupos de características de datos, llamados vectores, y luego encontrar un hiperplano de separación óptimo que tenga un margen máximo (es decir, la distancia máxima entre los vectores de soporte de ambos grupos como se muestra en la Figura 13(a). Los puntos de datos ubicados en los márgenes se denominan vectores de soporte que influyen en la posición y orientación del hiperplano. En otros trabajos, SVM tiene como objetivo maximizar el margen por medio de vectores de soporte. SVR aplica el mismo principio que SVM, pero para problemas de regresión. El algoritmo SVR encuentra una función que se ajusta mejor a los puntos de datos dentro de un límite de decisión como se muestra en la Figura 13(b) utilizando regresión lineal. La línea de mejor ajuste es el hiperplano que tiene el número máximo de puntos de datos dentro de un valor umbral ϵ (Ver la Figura 13 (b)).

Figura 13

Ejemplo de SVM

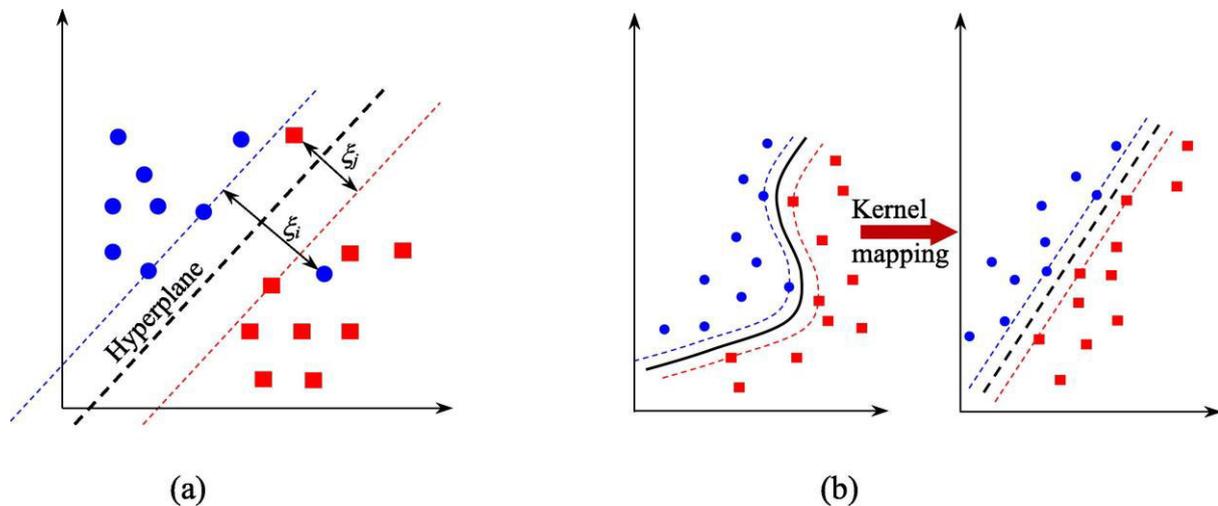


Nota. La figura muestra un ejemplo de SVM para (a) clasificación y (b) regresión. Fuente: (Firmansyah et al., 2024).

En la mayoría de las aplicaciones del mundo real, los datos no son linealmente separables y, por lo tanto, es imposible encontrar un hiperplano que los separe. En este caso, se utilizan un parámetro de penalización y una función *kernel*. Cuando el punto de datos está en el lado incorrecto del margen, se introduce el parámetro de penalización (es decir, la variable de holgura ξ ilustrada en la Figura 14(a) para controlar el equilibrio entre maximizar los márgenes del hiperplano y minimizar la distancia total de las variables de holgura $\sum \xi_i$. Esta técnica de margen suave permite que SVM cometa ciertos errores para mantener el margen del hiperplano lo suficientemente amplio para que otros puntos aún puedan clasificarse correctamente. Mientras tanto, las funciones *kernel* se utilizan para mapear los datos separables no lineales originales en un nuevo espacio donde los datos son linealmente separables como se ilustra en la Figura 14(b). Esto se llama el "truco *kernel*". Las funciones *kernel* más utilizadas son las funciones polinómicas lineales y no lineales, RBF y la función sigmoidea. Tanto las funciones *kernel* como los parámetros de penalización tienen efectos significativos en el rendimiento de los modelos SVM. Tharwat, (2019) examinó e investigó las selecciones de parámetros de penalización y funciones de *kernel*.

Figura 14

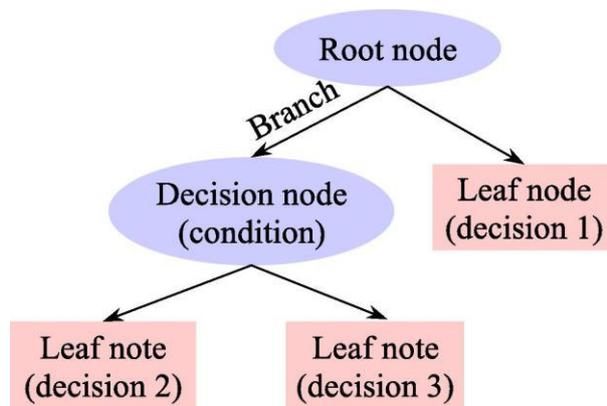
Ilustración de Parámetro de Penalización y Funciones Kernel



Nota. Ilustración de (a) parámetros de penalización ξ y (b) funciones *kernel*. Fuente: (Firmansyah et al., 2024)

2.6.5.6. Árbol de Decisión.

El árbol de clasificación y regresión (DT, por sus siglas en inglés), también conocido como árbol de clasificación y regresión (CART, por sus siglas en inglés), es un modelo basado en árboles para visualizar el proceso de toma de decisiones. Se ha convertido en uno de los algoritmos más populares debido a su simplicidad y capacidad para manejar datos tanto numéricos como categóricos. Como se muestra en la Figura 15, un DT tiene cuatro elementos que incluyen un nodo raíz, dos o más ramas, nodos de decisión y nodos de hoja (terminales). El nodo raíz es el nodo de decisión superior de un árbol que representa el objetivo final. El nodo de hoja ubicado al final de la rama indica una decisión que se debe tomar, mientras que el nodo de decisión representa una condición que hace que un conjunto de datos se divida.

Figura 15*Ejemplo de un DT*

Nota. (Firmansyah et al., 2024)

Un árbol se construye dividiendo los datos de origen (nodo raíz) en muchos subconjuntos más pequeños. La condición de división puede basarse en diferentes métricas, como el índice de Gini, la entropía, la ganancia de información y el MSE (problema de regresión). El proceso de división se repite en cada subconjunto derivado hasta que no se puede encontrar una división que reduzca las métricas utilizadas o se alcanza la profundidad máxima del árbol. Aunque el DT ofrece muchas ventajas (por ejemplo, menos esfuerzo para el preprocesamiento de datos debido a que no se requiere escalamiento o normalización de datos), es inestable y relativamente inexacto en comparación con otros algoritmos basados en árboles como RF y la familia BA. Para reducir el riesgo de sobreajuste en el DT, el modelo debe regularizarse para restringir la libertad del DT durante el entrenamiento (por ejemplo, la profundidad máxima del árbol, el número máximo de nodos de hoja, etc.). Puede encontrar más información sobre el ajuste de hiper parámetros de regularización en (Gerón, 2023).

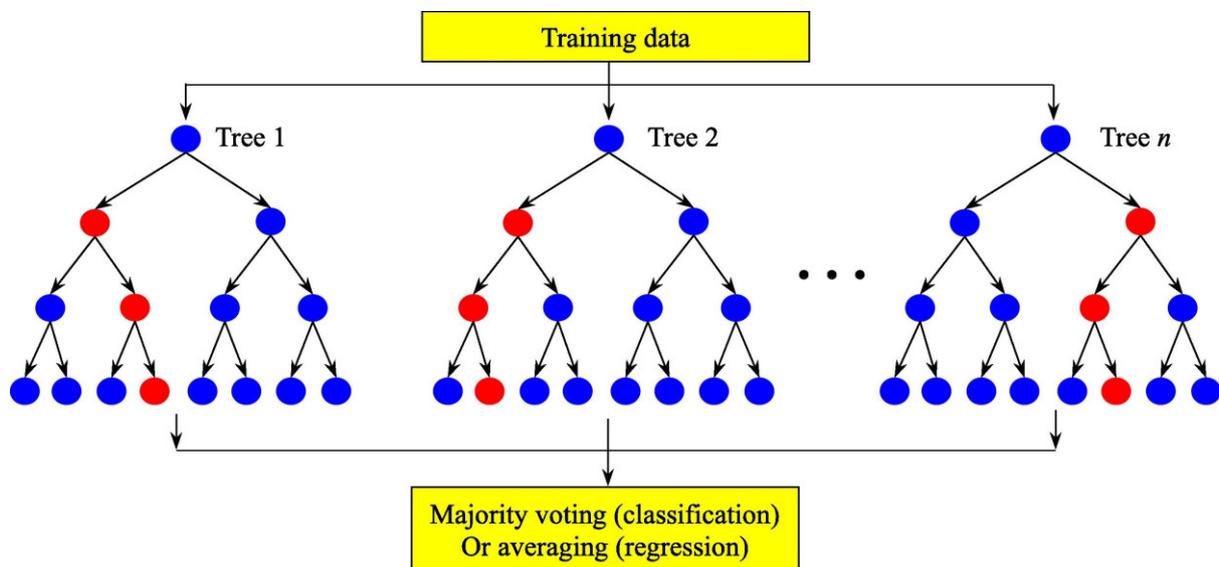
2.6.5.7. Bosque Aleatorio.

RF desarrollado por Breimani, (2001) es un algoritmo de aprendizaje en conjunto que utiliza DT como aprendices débiles. RF construye DT utilizando la técnica de *bagging*

(entrenamiento paralelo). La idea básica detrás de este método es construir un bosque de DT individuales utilizando una selección aleatoria de características (de ahí el nombre "*Random Forest*"), y luego combinar los resultados de cada DT tomando el voto mayoritario (problema de clasificación) o el promedio (problema de regresión) como se muestra en la Figura 16. Por lo tanto, reduce el riesgo de problemas de sobreajuste en el método DT.

Figura 16

Diagrama de Flujo de RF (Entrenamiento Paralelo)



Nota. (Firmansyah et al., 2024)

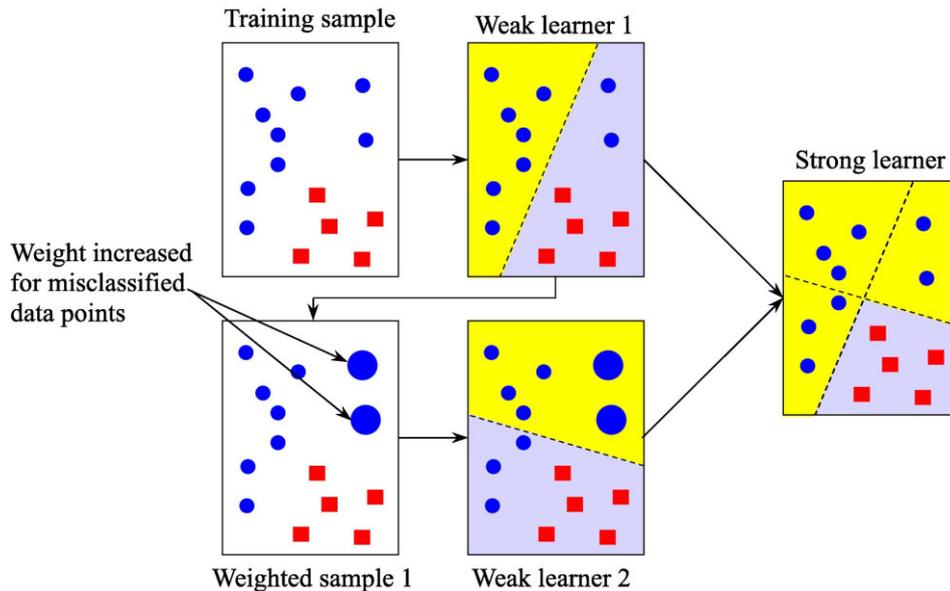
2.6.5.8. Algoritmo de Refuerzo.

BA es una técnica de conjunto que combina muchos modelos individuales en un modelo predictivo que puede aumentar el rendimiento de los modelos individuales (Freund & Schapire, 1996). El término "impulso" se refiere a fortalecer a los estudiantes débiles (por ejemplo, DT). Esto se puede hacer por medio de enfoques secuenciales. BA fue introducido por primera vez en 1996 por Freund & Schapire, (1996) con el algoritmo de impulso adaptativo (AdaBoost). Desde entonces, se han desarrollado varios enfoques de impulso centrados en mejorar la velocidad y la precisión, incluida la máquina de impulso de gradiente (GBM) (Friedman,

2001b) también conocida como impulso de árbol de gradiente, impulso de gradiente extremo (XGBoost) (Chen & Guestrin, 2016) , máquina de impulso de gradiente ligero (LightGBM) (Ke et al., 2017) e impulso de gradiente categórico (CatBoost) (Prokhorenkova et al., 2018). Una comparación de varias técnicas de BA realizada recientemente por Castaño et al. (2009) quien indicó que CatBoost brinda los resultados más precisos, pero es un poco más lento que otros. XGBoost ocupa el segundo lugar tanto en precisión como en velocidad. LightGBM es el algoritmo más rápido, pero su precisión no es buena en comparación con otros.

2.6.5.9. AdaBoost.

Este es el primer paso en la familia BA. También es uno de los algoritmos de refuerzo más populares (Almeida et al., 2016). AdaBoost está construido sobre la base de algoritmos DT y RF, siendo los aprendices débiles el DT que tiene un nodo y dos hojas llamadas tocones de decisión. La idea detrás de AdaBoost es mejorar el rendimiento de los aprendices débiles utilizando datos preponderados de forma adaptativa obtenidos en función del resultado de los aprendices débiles anteriores. Ver la Figura 17.

Figura 17*Implementación de Adaboost con Dos Estudiantes Débiles*

Nota. (Firmansyah et al., 2024)

El procedimiento de implementación de AdaBoost se ilustra en la Figura 17. El primer aprendiz débil se entrena utilizando un peso uniforme para todos los puntos de datos de la muestra de entrenamiento. Luego, el segundo aprendiz débil se entrena utilizando la muestra ponderada 1 con los coeficientes de peso que se actualizan para tener en cuenta el error del primer aprendiz débil (es decir, aumentando los pesos de los puntos de datos mal clasificados (dos puntos verdes en la Figura 17) y disminuyendo los pesos de los puntos de datos clasificados correctamente). Este proceso se repite hasta el último aprendiz débil. Finalmente, el aprendiz fuerte se forma combinando los límites de decisión aprendidos por todos los aprendices débiles.

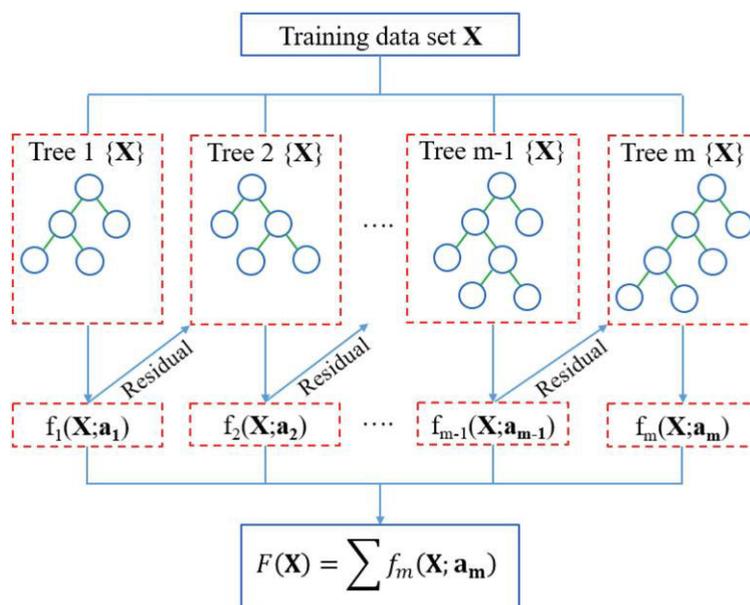
2.6.5.10. Método de Aumento de Gradiente GBM - *Gradient Boosting Method*.

Este algoritmo fue desarrollado por Friedman, (2001b) al realizar dos modificaciones al algoritmo AdaBoost. La primera modificación es el uso de DT como aprendices débiles en lugar de tocones de decisión. Esto significa que no es necesario que todos los árboles sean

iguales y, por lo tanto, pueden capturar diferentes salidas de los datos. La segunda modificación involucra el método de entrenamiento. En lugar de actualizar los pesos de los puntos de datos mal clasificados y correctamente clasificados como en el caso de AdaBoost, GBM minimiza la función de pérdida de cada aprendiz débil utilizando el procedimiento de descenso de gradiente, un algoritmo de optimización genérico que puede aplicarse a cualquier función de pérdida que sea diferenciable. Como se muestra en la Figura 18, la pérdida residual (error) del árbol anterior se tiene en cuenta en el entrenamiento del siguiente árbol. Al combinar todos los árboles, el modelo final puede capturar la pérdida residual de los aprendices débiles. Además de su precisión, GBM también ofrece mucha flexibilidad al proporcionar varios hiper parámetros de ajuste, incluidos el número de árboles, la profundidad de los árboles y la tasa de aprendizaje para controlar la convergencia y la velocidad de entrenamiento.

Figura 18

Ilustración del Modelo GBM



Nota. (Firmansyah et al., 2024)

2.7. Marco Filosófico

Numerosas investigaciones han contribuido al desarrollo del emprendimiento como campo de conocimiento. “La presión de las organizaciones por innovar ha generado una gran demanda por la educación en emprendimiento. De esta manera, la mayoría de los esfuerzos se han orientado a estudiar el emprendedor desde sus acciones y experiencias prácticas y no se han explorado las dimensiones sociales de este fenómeno”. Se hace necesario el estudio de la dimensión humana del emprendimiento. “La Universidad y algunas instituciones de carácter público y privado han ofrecido varias alternativas en la enseñanza de creación de empresas; sin embargo, su formación se ha quedado en esquemas funcionalistas y operativos”. En la educación empresarial, “el concepto que se tiene de ser humano debe estar estrechamente ligado con la posibilidad de influir en la formación de personas que aporten significativamente al desarrollo socioeconómico de la sociedad a la que pertenecen, no solamente desde el campo de la creación de empresas, sino desde el punto de vista de la transformación social, en el que cada individuo impulse el bienestar colectivo, no sólo el individual”. (Álvarez Rangel & Matos Castro, 2014). La educación en emprendimiento desarrolla las cualidades necesarias para formar mejores emprendedores capaces de poner en marcha emprendimientos sociales que busquen dar soluciones a las necesidades humanas en donde subsisten problemas derivados de la pobreza y la desigualdad social. De esta forma, el sujeto trata de encontrar soluciones a sus necesidades a través de su inventiva y liderazgo para asegurar en el largo plazo empresas viables y sostenibles socialmente. En consecuencia, toda iniciativa de emprendimiento deberá estar fundamentada en el respeto, la solidaridad, la equidad, la tolerancia y la igualdad social. El estudio de la dimensión humana del emprendimiento conlleva a que se replanteen las estrategias utilizadas para la educación en espíritu empresarial. Es imprescindible que las herramientas pedagógicas utilizadas para la enseñanza del espíritu empresarial estén enmarcadas en un contexto real que justifique la necesidad de desarrollar proyectos

innovadores cuyo motor principal sea la mejora de la calidad de vida de las personas. “El emprendimiento es inherente a la esencia del ser, está presente en cada una de las acciones que desarrolla para la búsqueda de la transformación y mejoramiento de sus condiciones de vida, pero como muchos de los atributos humanos, es necesario afianzarlos mediante la educación”. En efecto, la educación puede facilitarnos el proceso de materializar nuestras buenas ideas en todos los campos de nuestras actividades intelectuales y profesionales, ya que nos hace mejorar nuestras actitudes y aptitudes para emprender.

El hombre es un ser social y en constante interacción con el entorno. “El ser humano es un ser cultural, que experimenta el mundo social en términos de motivaciones; de acuerdo con esto, desempeña ciertos comportamientos, se ajusta a valores, y especialmente se dispone creativamente para emprender como un acto de sentido; a su vez, se siente determinado por las valoraciones positivas o negativas e influenciado por sus coterráneos”. (Álvarez & Matos, 2014). El concepto de ser humano que fundamenta la presente investigación implica que el hombre es un ser creador y emprendedor, cuyo comportamiento está orientado al aprendizaje personal y organizacional, y al desarrollo de acciones transformadoras del contexto en el que se desarrolla. La dimensión humana del emprendimiento percibe al hombre como “un ser que aprende a lo largo de la vida, interesado en saber ser, saber, hacer y emprender, además inserto en una cultura y en interacción con el medio ambiente y en ese sentido, trasciende el ámbito individual y cobra sentido en la interacción con los otros, busca conocer y comprender el medio ambiente para vivir en armonía con él, con sus leyes y en particular con las instituciones que le posibilitan los medios de vida y el bienestar en general” (Álvarez & Matos, 2014).

En síntesis, el hombre es un ser competitivo, social, emprendedor, autónomo, en constante interacción con el medio que le rodea, capaz de comprender e interpretar la realidad social y económica, con vocación de servicio, interesado en desarrollar y adquirir nuevo conocimiento para su crecimiento personal y profesional. En consecuencia, la educación en

emprendimiento debe estar orientada a la formación de personas capaces de interpretar el contexto en el que se desenvuelven y afrontar los desafíos de la sociedad actual con iniciativas empresariales innovadoras con gran sentido ético y humano. (Álvarez & Matos, 2014). El rol de la empresa en la sociedad ha sido un problema objeto de discusión desde la mitad del siglo XX. Así pues, en los Estados Unidos, la Responsabilidad Social Empresarial (RSE) era concebida, hasta los años cincuenta del pasado siglo, por la mayoría en términos de obligación social, centrada fundamentalmente en los intereses económicos de los accionistas, dentro de la legalidad y la ética imperante. (Conesa, 2014). En los años setenta del siglo XX surge con fuerza la ética de los negocios (*Business Ethics*) que en Europa se tradujo como la “ética de la empresa” (Cortina, 2004) generándose entonces un intenso debate sobre las relaciones entre empresa y sociedad. (Conesa, 2014). La visión de la empresa, no sólo como un negocio, sino como un grupo humano que lleva adelante una valiosa tarea para la sociedad, la de producir bienes y servicios, a través de la obtención del beneficio, ya no es suficiente, sino que se exige mayor responsabilidad y compromiso ético a las empresas para limitar su impacto negativo sobre el medioambiente y la sociedad en general. (Cortina et al., 1996). Desde entonces, las empresas se enfrentan a crecientes exigencias de mayor compromiso social. Debido a este creciente interés social en el comportamiento responsable de las empresas, surge en primer lugar, el concepto de Responsabilidad Social Empresarial (RSE) para posteriormente hacerlo el de Responsabilidad Social Corporativa (RSC), el cual ya no atañe únicamente a la empresa sino a cualquier tipo de organización; de modo en la actualidad han pasado a utilizarse indistintamente (Conesa, 2014). Asimismo, Ética y economía tienen mucho que ver con las elecciones que hacemos, pues la economía estudia cómo elegimos sin entrar en qué comportamientos son incorrectos, sin embargo, muchas de estas elecciones tienen implicaciones éticas (Cuervo et al., 2008), por lo que la ética de los negocios ayuda a la resolución de los dilemas éticos planteados en el seno de la empresa. El marco ético de la RSC

proviene de responder, defender o justificar aquello que se dice, hace u omite, ya que somos responsables cuando tenemos varias posibilidades de elección y nos decidimos por una, no todas las empresas ni todos sus directivos conciben la ética de la misma manera ni, tampoco, el nivel de calidad ética que están dispuestos a practicar, para muchos, lo éticamente bueno se limita a lo estrictamente legal, mientras que para otros la ética es más exigente que la ley y para no pocos, tendría que ver con una cuestión de demanda social. (García Navarro, 2012). Sin embargo, los comportamientos irresponsables o poco éticos impiden la viabilidad y la rentabilidad de la empresa a largo plazo, minando la eficiencia y la confianza del sistema económico en su conjunto (Cuervo et al., 2008). Por lo que todas las organizaciones, lo sepan o no, cuentan con unos estándares éticos más o menos programados, que son inherentes a sus culturas, sistemas y procesos organizativos; de modo que toda compañía debe desarrollar valores que se conviertan en directrices y patrones cotidianos para sus integrantes, siendo estos los llamados valores compartidos, con objeto de obtener una ventaja competitiva contando para ello con miembros activos y comprometidos (García et al., 2013). Cada organización dispone de patrones de comportamientos propios y esenciales de la cultura empresarial, la dimensión ética de dichos patrones es la transmitida a través de sus miembros. Aunque la ética en la empresa no es sin más, la suma de las conductas y valores de las personas que la forman, no cabe duda de que no puede haber una empresa ética si las personas que la dirigen y que trabajan en ella no lo son. Para ello los valores compartidos, desde la dirección, deberán marcar el camino a seguir, indicando a los miembros de la organización que se espera de ellos, e influyendo en el entorno, siendo una forma de fomentar el comportamiento ético de los miembros de la organización y, por tanto, de la propia organización en sí. De este modo, ética y responsabilidad social son conceptos que hacen referencia fundamentalmente a nuestras relaciones a través del tiempo. Lo que en términos sencillos puede conllevar a entender la ética como el estudio de la forma en que nuestras decisiones afectan a los demás, y se entiende la

RSC como lo que hace la organización para satisfacer a sus grupos de interés. Por tanto, la ética da sentido a la responsabilidad social. (Cortina et al., 1996)

2.8. Marco Tecnológico

El futuro de las operaciones mineras apunta a estar totalmente automatizado. La que alguna vez fue vista como una industria con entornos y operaciones peligrosas y desafiantes, gracias a la revolución tecnológica está lista para cambiar dramáticamente nuestra percepción mediante la innovación minera. La principal razón para este cambio es la necesidad de la industria por ser más eficiente, productiva, segura e ingeniosa para afrontar adversidades, especialmente si se tiene en cuenta las dificultades climatológicas y los crecientes costos energéticos que enfrenta la industria todos los días. De esta manera, se está impulsando innovaciones que creen un ambiente laboral más seguro y cómodo para los mineros; esto sumado a capacitaciones en la automatización de procesos, harán que las minas sean más seguras al retirar a los mineros de las zonas de peligro. A su vez, los avances en una comunicación más fiable permitirán a los mineros conectar de manera segura con sus colegas en el otro extremo de la operación o incluso en el otro lado del mundo.

2.8.1. Digitalización de la Minería

A continuación, Peter du Bruin, investigador principal de Ericsson, explica cómo su equipo está investigando los beneficios de la tecnología 5G en la minería. Las máquinas a control remoto y la ventilación inteligente son quizás los dos casos de aplicación que han resultado más críticos para la mejora de la seguridad y la eficiencia en las minas. Esta tecnología ha logrado implementar aplicaciones de control remoto con múltiples fuentes de retroalimentación (vídeo, audio, háptica) estén habilitadas, ayudando a disminuir el personal en las zonas y momentos más peligrosos. A su vez, esto mejora la productividad, ya que acorta el lapso de tiempo entre una explosión y el ingreso de los mineros en el área. Adicionalmente,

se pueden conectar numerosos sensores en la mina, creando así enormes cantidades de datos, para ser analizados casi en tiempo real. Al utilizar estos datos, se puede optimizar la dirección de la ventilación, proporcionando un mejor suministro de aire, así como una mayor eficiencia, lo que permite un control total del entorno minero.

Además, se implementó una conexión de pernos de anclaje de roca, que proporcionan información acerca de las vibraciones. Esto quiere decir que alerta si la roca o el terreno se está moviendo para asegurarse de que el entorno este seguro. Eventualmente, se podrá conectar una gran cantidad de barras metálicas en las paredes y techos de la cavidad de una mina. Los pernos de roca estabilizarán el túnel distribuyendo el esfuerzo uniformemente a lo largo de la longitud del perno, o, en otras palabras, transfiriendo la carga de la superficie inestable al interior estable de la masa rocosa. Finalmente, existe una amplia gama de posibles aplicaciones que se están explorando. Esto incluye video vigilancia; detector de fatalidades y localización de máquinas, vehículos, cosas y personas. Todo ello, a fin de impulsar el inmenso potencial de innovación y mejora de la productividad dentro de la industria minera.

2.9. Marco Legal

- **La Constitución Política del Perú.** - Promulgada en 1979 en su art. 47 establece: “... corresponde al Estado dictar medidas sobre higiene y seguridad en el trabajo, que permitan prevenir los riesgos profesionales y asegurar la salud y la integridad física y mental de los trabajadores. En cambio, en la Constitución Política de 1993 ha sido excluido el derecho a la prevención de riesgos ocupacionales”.
- **Ley N° 29783.-** Ley de Seguridad y Salud en el Trabajo, aprobado el 26 de Julio del 2011. Entre los principios de esta norma, tenemos:

- **Principio De Prevención.** El empleador garantiza, en el centro de trabajo, el establecimiento de los medios y condiciones que protejan la vida, la salud y el bienestar de los trabajadores, y de aquellos que, no teniendo vínculo laboral, prestan servicios o se encuentran dentro del ámbito del centro de labores. Debe considerar factores sociales, laborales y biológicos, diferenciados en función del sexo, incorporando la dimensión de género en la evaluación y prevención de los riesgos en la salud laboral.
- **Principio De Responsabilidad.** El empleador asume las implicancias económicas, legales y de cualquier otra índole a consecuencia de un accidente o enfermedad que sufra el trabajador en el desempeño de sus funciones o a consecuencia de él, conforme a las normas vigentes.
- **Principio De Cooperación.** El Estado, los empleadores y los trabajadores, y sus organizaciones sindicales establecen mecanismos que garanticen una permanente colaboración y coordinación en materia de seguridad y salud en el trabajo.
- **Principio de Información y Capacitación.** Las organizaciones sindicales y los trabajadores reciben del empleador una oportuna y adecuada información y capacitación preventiva en la tarea a desarrollar, con énfasis en lo potencialmente riesgoso para la vida y salud de los trabajadores y su familia.
- **Principio de Gestión Integral.** Todo empleador promueve e integra la gestión de la seguridad y salud en el trabajo a la gestión general de la empresa.
- **Principio de Atención Integral de la Salud.** Los trabajadores que sufran algún accidente de trabajo o enfermedad ocupacional tienen derecho a las prestaciones

de salud necesarias y suficientes hasta su recuperación y rehabilitación, procurando su reinserción laboral.

- **Principio De Consulta Y Participación.** El Estado promueve mecanismos de consulta y participación de las organizaciones de empleadores y trabajadores más representativos y de los actores sociales para la adopción de mejoras en materia de seguridad y salud en el trabajo.

 - **Principio De Primacía De La Realidad.** Los empleadores, los trabajadores y los representantes de ambos, y demás entidades públicas y privadas responsables del cumplimiento de la legislación en seguridad y salud en el trabajo brindan información completa y veraz sobre la materia. De existir discrepancia entre el soporte documental y la realidad, las autoridades optan por lo constatado en la realidad.

 - **Principio De Protección.** Los trabajadores tienen derecho a que el Estado y los empleadores aseguren condiciones de trabajo dignas que les garanticen un estado de vida saludable, física, mental y socialmente, en forma continua. Dichas condiciones deben propender a:
 - Que el trabajo se desarrolle en un ambiente seguro y saludable.
 - Que las condiciones de trabajo sean compatibles con el bienestar y la dignidad de los trabajadores y ofrezcan posibilidades reales para el logro de los objetivos personales de los trabajadores.
- **Decreto Supremo N° 024-2016-EM.** - Decreto Supremo que aprueba el Reglamento de Seguridad y Salud Ocupacional en minería, aprobado el 26 de Julio del 2016. El

presente reglamento tiene como objetivo prevenir la ocurrencia de incidentes, incidentes peligrosos, accidentes de trabajo y enfermedades ocupacionales, promoviendo una cultura de prevención de riesgos laborales en la actividad minera. Para ello, cuenta con la participación de los trabajadores, empleadores y el Estado, quienes velarán por su promoción, difusión y cumplimiento.

- **Resolución de Consejo Directivo del Organismo Supervisor de la Inversión en Energía y Minería Osinergmin N° 040-2017-OS-CD.** - Del 9 de marzo del 2017, mediante el cual se aprueba el Reglamento de Supervisión, Fiscalización y Sanción de las Actividades Energéticas y Mineras a cargo de Osinergmin, el cual forma parte integrante de la presente resolución. En consecuencia, el Osinergmin ejerce sus funciones de supervisión, fiscalización y sanción a través de sus órganos competentes o de Empresas Supervisoras, según corresponda. El sector energético comprende a los subsectores de electricidad e hidrocarburos. Dentro de este último las actividades vinculadas a hidrocarburos líquidos y gas natural. El sector minero comprende a la gran y mediana minería, a aquellos titulares mineros que no cumplen los requisitos de pequeño productor minero, así como a aquellos que realizan almacenamiento de concentrado de minerales.

2.10. Marco Conceptual

Los términos que a continuación se definen, son los más utilizados en el desarrollo del trabajo de investigación.

- **Accidente de Trabajo:** Es el suceso repentino que sobreviene por causa o con ocasión del trabajo, y que produce en el trabajador, una lesión orgánica, una perturbación

funcional, una invalidez o la muerte. Interrumpe o interfiere el proceso normal y ordenado de una actividad.

- **Accidente de Trabajo Leve:** Es aquel, que luego de la evaluación, el accidentado debe volver máximo al día siguiente a sus labores habituales.
- **Accidente de Trabajo Incapacitante:** Es aquel, que luego de la evaluación, el médico diagnostica y define que el accidente no es leve y determina que continúe el tratamiento al día siguiente de ocurrido el accidente.
- **Accidente Sin Incapacidad:** Es aquel que no produce lesiones o que, si lo hace, son tan leves que el accidentado continúa trabajando inmediatamente después de lo ocurrido.
- **Ambiente de Trabajo:** Es el conjunto de condiciones que rodean a la persona que trabaja y que, directa o indirectamente, influyen en la salud y vida del trabajador.
- **Condiciones Inseguras:** Situaciones que se presentan en el lugar de trabajo y que se caracteriza por la presencia de riesgos no controlados que pueden generar accidentes de trabajo o enfermedades profesionales.
- **Controles:** Medidas implementadas con el fin de minimizar la ocurrencia de eventos que generen pérdidas.
- **Exposición:** Frecuencia con que las personas o la estructura entran en contacto con los factores de riesgo en su jornada laboral.
- **Factores Personales:** Son aquellos que podemos identificar con las características de las personas y su comportamiento tales como la falta de motivación, entrenamiento inadecuado, falta de conocimiento, sobrecarga emocional, etc.
- **Factores de Riesgo:** Es la existencia de elementos, fenómenos, ambiente y acciones humanas que encierran una capacidad potencial de producir lesiones o daños

materiales y cuya probabilidad de ocurrencia depende de la eliminación o control del elemento agresivo.

- **Factores de Riesgo Por Carga Física:** Se refiere a todos aquellos aspectos de la organización del trabajo, de la estación o puesto de trabajo y de su diseño que pueden alterar la relación del individuo con el objeto técnico produciendo problemas en el individuo, en la secuencia de uso o la producción.
- **Factores de Riesgo Mecánico:** Objetos, máquinas, equipos, herramientas que, por sus condiciones de funcionamiento, diseño o por la forma, tamaño, ubicación y disposición del último tienen la capacidad potencial de entrar en contacto con las personas o materiales, provocando lesiones en los primeros o daños en los segundos.
- **Factores de Riesgo Eléctricos:** Se refiere a los sistemas eléctricos de las máquinas, los equipos que al entrar en contacto con las personas o las instalaciones y materiales pueden provocar lesiones a las personas y daños a la propiedad.
- **Grado de Peligrosidad:** Es un indicador de la gravedad de un riesgo reconocido.
- **Índice de Frecuencia De Accidentes (IF):** Número de accidentes mortales e incapacitantes por cada millón de horas hombre trabajadas. Se calculará con la fórmula siguiente:

$$IF = \frac{N^{\circ} \text{ Accidentes } \times 1'000,000}{\text{Horas Hombre Trabajadas}} \quad \dots (19)$$

$$N^{\circ} \text{ Accidentes} = \text{Incapacitantes} + \text{Mortal} \quad \dots (20)$$

- **Índice de Severidad de Accidentes (IS):** Número de días perdidos o cargados por cada millón de horas - hombre trabajadas. Se calculará con la fórmula siguiente:

$$IS = \frac{N^{\circ} \text{ Días Perdidos o Cargados } \times 1'000,000}{\text{Horas Hombre Trabajadas}} \quad \dots (21)$$

- **Índice de Accidentabilidad (IA):** Una medición que combina el índice de frecuencia de lesiones con tiempo perdido (IF) y el índice de severidad de lesiones (IS), como un medio de clasificar a las empresas mineras.

Es el producto del valor del índice de frecuencia por el índice de severidad dividido entre 1000.

$$IA = \frac{IF \times IS}{1000} \quad \dots (22)$$

- **Identificación de Peligros:** Proceso mediante el cual se reconoce que existe un peligro y se define sus características.
- **Indicador:** Variable o atributo, objeto de medición o valoración.
- **Índice:** Es la expresión matemática o cuantitativa del indicador.
- **Inducción:** Capacitación inicial dirigida a otorgar conocimientos e instrucciones al trabajador para que ejecute su labor en forma segura, eficiente y correcta.
- **Inspecciones de Seguridad:** Las inspecciones de seguridad se realizan con el fin de vigilar los procesos, equipos, máquinas u objetos que, en el diagnóstico integral de condiciones de trabajo y salud, han sido calificados como críticos por su potencial de daño.
- **Lesión:** Alteración estructural o funcional de los tejidos, órganos o sistema en un individuo.
- **Peligro:** Fuente, situación o acto con potencial para causar daño en términos de daño humano o deterioro de la salud o una combinación de éstos.
- **Programa de Salud Ocupacional:** El programa de salud ocupacional es la planeación, organización, ejecución y evaluación de una serie de actividades de Medicina Preventiva, Medicina del Trabajo, Higiene y Seguridad Industrial, tendientes a preservar mantener y mejorar la salud individual y colectiva de los trabajadores en sus

ocupaciones y que deben ser desarrolladas en sus sitios de trabajo en forma integral e interdisciplinaria.

- **Riesgos:** Probabilidad de ocurrencia de un evento de características negativas en el trabajo, que pueden ser generados por una condición de trabajo capaz de desencadenar alguna perturbación en la salud o integridad física del trabajador, como daño en los materiales y equipos o alteraciones del ambiente.
- **Salud:** Bienestar físico, psíquico y social del ser humano y de su entorno.
- **Salud Ocupacional:** Conjunto de disciplinas como finalidad la promoción de la salud en el trabajo a través del fomento y mantenimiento del más elevado nivel de bienestar en los trabajadores de todas las profesiones, previniendo alteraciones de la salud por las condiciones de trabajo, protegiéndolos contra los riesgos resultantes de la presencia de agentes nocivos y colocándolos en un cargo acorde con sus aptitudes físicas y psicológicas.
- **Seguridad Ocupacional o Industrial:** Conjunto de actividades destinadas a la identificación, evaluación y control de los factores de riesgo o condiciones de trabajo que puedan producir accidentes de trabajo.
- **Visitas de Inspección:** Las visitas de inspección se realizan con el fin de vigilar procesos, equipos, máquinas u objetos que, en el diagnóstico integral de condiciones de trabajo y salud, han sido calificados como críticos por su potencial daño.

III. MÉTODO

3.1. Tipo de Investigación

3.1.1. Tipo

- De acuerdo con la orientación: Básica.
- De acuerdo con la técnica de contrastación: Explicativa.
- De acuerdo con la direccionalidad: Prospectiva.
- De acuerdo con el tipo de fuente de recolección: Retro-lectiva.
- De acuerdo con la evolución del fenómeno estudiado: Transversal.
- De acuerdo con la comparación de las poblaciones: Comparativa.

3.1.2. Nivel

El presente trabajo de investigación es de nivel IV, de tipo predictivo II (Medina, 2014).

Tomando como referencia la naturaleza de las variables, esta investigación es del **nivel Predictivo – descriptivo – explicativo – correlacional**. Es predictivo porque identifica la probabilidad futura de los accidentes laborales en la minería subterránea del Perú mediante la aplicación de técnicas de inteligencia artificial y *machine learning*. Se logró determinar la severidad probable y la frecuencia esperada de accidentes, proporcionando así un modelo que permita implementar medidas preventivas eficaces. Asimismo, describe el comportamiento y las incertidumbres en las diferentes etapas de análisis mediante la integración de datos históricos de accidentes y su preprocesamiento para el entrenamiento de modelos de *machine learning*. Estos modelos consideran variables aleatorias definidas en la metodología, como factores de riesgo, mantenimiento de equipos y capacitación del personal, para predecir la

severidad de los accidentes. Luego, explica cómo se determina la efectividad de las estrategias de prevención desarrolladas. La metodología aplicada permite la cuantificación de la reducción de accidentes y su severidad en términos de mejoras implementadas, las cuales son estimadas en función de los datos obtenidos y analizados. Esto incluye la evaluación de los costos de implementación de medidas preventivas y su impacto en la seguridad laboral, y además correlaciona las variables en cada etapa de la aplicación de la metodología, permitiendo estudiar la relación entre factores de riesgo, técnicas de preprocesamiento de datos y la precisión del modelo predictivo sin necesidad de manipularlas, ofreciendo información basada en valores comparables. También ha permitido conocer la correlación existente entre la implementación de estrategias de prevención y la reducción de accidentes, es decir, cómo varía una al modificarse otra. Además, informa sobre la dirección de dichas variaciones, facilitando la toma de decisiones informadas para mejorar la seguridad en la minería subterránea del Perú (Hernández et al., 2010).

3.1.3. Códigos y Nomenclaturas UNESCO

- 1206. Análisis Numérico
 - 120603. Análisis de Errores
- 1208. Probabilidad
 - 120808. Procesos estocásticos
- 1209. Estadística
 - 120903. Análisis de Datos
- 2205. Mecánica
 - 220501. Mecánica Analítica
- 2507. Geofísica
 - 250705. Sismología y Prospección Sísmica.

- 3305. Tecnología de la Construcción
 - 330506. Ingeniería Civil
 - 330532. Ingeniería Estructural
 - 330533. Resistencia de Estructuras
- 6307. Cambio y Desarrollo Social
 - 630706. Desarrollo Socioeconómico
 - 630707. Tecnología y cambio social.
- 7201. Filosofía del conocimiento
 - 720105. Teoría de la Percepción

3.1.4. Diseño de la Investigación

Según Hernández & Batista (2010), el diseño es el plan o estrategia que se desarrolló para obtener la información que requiere la investigación.

El diseño que se aplicó a la tesis es el no experimental, transeccional o transversal, descriptivo, correlacional – causal.

- **Diseño No Experimental:** Se define como la investigación que se realiza sin manipular deliberadamente las variables. En este diseño se observaron los fenómenos tal y como se dan en su contexto natural, luego han sido analizados. En el contexto de esta investigación, se observaron los datos históricos de accidentes laborales en la minería subterránea y se utilizaron técnicas de inteligencia artificial para su análisis.
- **Diseño Transeccional o Transversal:** Consiste en la recolección de datos en un momento específico. Su propósito es describir las variables y analizar su incidencia e interrelación en un momento dado. En este caso, se recopilamos datos sobre los accidentes laborales, factores de riesgo, y se aplicaron técnicas de *machine learning* para evaluar y predecir la severidad de los accidentes.

- **Diseño Transeccional Descriptivo:** Tiene como objetivo indagar la incidencia y los valores en que se manifiestan las variables de la investigación. Se describieron las características de los accidentes laborales y los factores de riesgo asociados, proporcionando una visión detallada de la situación actual en la minería subterránea.
- **Diseño Transeccional Correlativo – Causal:** Ha servido para relacionar dos o más categorías, conceptos o variables en un momento determinado. Este diseño permitió analizar las relaciones entre los factores de riesgo identificados y la severidad de los accidentes laborales. Se trató también de descripciones, pero no de categorías, conceptos, objetos ni variables individuales, sino de sus relaciones, puramente correlacionales o con relaciones causales.

Este enfoque permitió una comprensión profunda de cómo las variables interactúan entre sí y cómo la aplicación de técnicas de inteligencia artificial y *machine learning* puede mejorar la predicción y prevención de accidentes laborales en la minería subterránea del Perú.

3.2. Población y Muestra

3.2.1. Población

La población de estudio está compuesta por todos los trabajadores de la minería subterránea en Perú. Esto incluye a mineros, supervisores, personal de mantenimiento, ingenieros de seguridad y otros trabajadores involucrados en las operaciones mineras subterráneas. La población también abarca los registros de accidentes laborales ocurridos en las minas subterráneas peruanas durante un periodo de tiempo determinado (Hernández et al., 2010). Ver la Tabla 3.

Tabla 3*Distribución de Titulares y Unidades por Estratos de Empresas Mineras*

Estratos	N° de empresas	N° de unidades mineras
Gran minería metálica	35	83
Mediana minería metálica	36	62
Gran minería no metálica	7	9
Mediana minería no metálica	14	26
TOTAL	92	180

Nota. Los datos han sido obtenidos del Ministerio de Energía y Minas.

3.2.2. *Tamaño de la Muestra*

El tamaño de la muestra se determinó considerando varios factores, entre ellos la disponibilidad de datos históricos de accidentes laborales, la accesibilidad de los registros y la representatividad de estos para garantizar la validez y fiabilidad de los resultados. Dado que se utilizaron técnicas de *machine learning*, es crucial contar con una muestra lo suficientemente grande para entrenar y validar los modelos predictivos. Se ha estimado que la muestra incluye al menos los registros de accidentes laborales de los últimos diez años (2007-2018), abarcando múltiples minas subterráneas en Perú.

3.2.3. *Muestreo*

El muestreo que se utilizó en esta investigación ha sido un muestreo no probabilístico por conveniencia y criterio. Este tipo de muestreo se eligió debido a la disponibilidad y accesibilidad de los datos específicos necesarios para el estudio. Los criterios de selección incluyen:

- **Disponibilidad de Datos:** Se seleccionaron los registros de accidentes laborales que han estado completos y detallados, incluyendo información relevante como

la fecha del accidente, la descripción del incidente, los factores de riesgo presentes, las medidas de seguridad implementadas y la severidad del accidente.

- **Relevancia Temporal:** Los registros seleccionados pertenecen al periodo 2007-2018, ya que este rango temporal proporciona una base de datos suficiente para el análisis predictivo y la identificación de tendencias.
- **Diversidad de Sitios:** Se incluyen datos de múltiples minas subterráneas en diferentes regiones de Perú para asegurar una muestra representativa de las diversas condiciones y prácticas operativas en el sector minero subterráneo del país.
- **Accesibilidad:** Solo se han incluido registros que han sido accesibles y verificables, garantizando la integridad y exactitud de los datos utilizados en la investigación.

Este enfoque de muestreo permitió una recopilación de datos robusta y representativa para el análisis, facilitando la aplicación efectiva de técnicas de inteligencia artificial y *machine learning* para la predicción y prevención de accidentes laborales en la minería subterránea del Perú.

3.3. Operacionalización de Variables

3.3.1. Estrategia de Prueba de Hipótesis

Las pruebas de hipótesis se realizaron con la finalidad de efectuar un juicio con respecto a la diferencia entre estadístico de muestra y un valor planteado del parámetro, que conduzca a una decisión sobre las hipótesis planteadas en el presente *trabajo* de investigación, el cual depende del empleo de la información contenida en la muestra.

La prueba efectuada es de dos colas y de una cola, según la prueba si las dos medias de población son sólo diferentes o si una media es mayor que la otra media.

$$H_0: \mu_1 = \mu_2 \quad \dots (23)$$

$$H_1: \mu_1 < \mu_2; \quad H_1: \mu_1 > \mu_2; \quad H_1: \mu_1 \neq \mu_2 \quad \dots (24)$$

Donde:

H_0 : Hipótesis nula

H_1 : Hipótesis alternativa

μ_1 = media de la población 1

μ_2 = media de la población 2

3.3.1.1. Procedimiento para Demostrar la Hipótesis Nula.

El procedimiento sistemático que seguir constará básicamente de diez (10) pasos de manera secuencial:

- Paso 1. plantear las hipótesis nulas.
- Paso 2. plantear las hipótesis alternativas.
- Paso 3. Especificar el nivel de significancia.
- Paso 4. determinar la muestra.
- Paso 5. establecer los valores críticos que establecen las regiones de rechazo de las de no rechazo.
- Paso 6. determinar la prueba estadística.
- Paso 7. Almacenar los datos, para proceder luego al cálculo del valor de las muestras de la prueba estadística apropiada.
- Paso 8. formar la regla de decisión.
- Paso 9. determinar la decisión estadística.
 - No se rechazarán las hipótesis nulas.

- Se rechazarán las hipótesis nulas y se aceptarán las hipótesis alternativas.
- Paso 10. expresar la decisión estadística en términos de los problemas.

3.3.1.2. Modelo Estadístico para la Prueba de la Hipótesis.

El modelo estadístico utilizado es:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \dots (25)$$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \dots (26)$$

$$T = \frac{\bar{D}}{S_D / \sqrt{n}} \quad \dots (27)$$

3.3.2. Variables e Indicadores

Proceso a través del cual se transformó la variable de conceptos abstractos a términos concretos, observables y medibles; es decir, en dimensiones e indicadores. Asimismo, está constituido por una serie de procedimientos o indicaciones para realizar la medición de las variables definidas conceptualmente, para obtener la mayor información posible, a modo de captar su sentido y adecuación al contexto del presente trabajo de investigación

Variables e Indicadores para los Objetivos Específicos

Objetivo Específico 1

Identificar y analizar los principales factores de riesgo asociados a los accidentes laborales por desprendimiento de rocas en la minería subterránea peruana.

- **Variable Independiente:** Factores de riesgo
 - **Indicadores:**
 - Número de incidentes reportados por desprendimiento de rocas
 - Condiciones geológicas del lugar
 - Estado y mantenimiento de equipos de minería
 - Capacitación y experiencia del personal
 - Procedimientos de seguridad implementados
- **Variable Dependiente:** Accidentes laborales por desprendimiento de rocas
 - **Indicadores:**
 - Frecuencia de accidentes
 - Severidad de los accidentes (lesiones, fatalidades)
 - Localización de los accidentes (profundidad de la mina, tipo de roca)

Objetivo Específico 2

Evaluar la efectividad del algoritmo de bosques aleatorios (*Random Forest*) en la predicción de la severidad de accidentes laborales en la minería subterránea, comparándolo con otros algoritmos de *machine learning*.

- **Variable Independiente:** Algoritmo de *machine learning*
 - **Indicadores:**
 - Tipo de algoritmo (*Random Forest*, SVM, *Naive Bayes*, etc.)
 - Parámetros del modelo (número de árboles en *Random Forest*, parámetros de regularización en SVM, etc.)

- Tiempo de entrenamiento del modelo
- **Variable Dependiente:** Efectividad del modelo predictivo
 - **Indicadores:**
 - Precisión del modelo (*accuracy*)
 - Sensibilidad (*recall*)
 - Especificidad (*precision*)
 - F1-Score
 - Área bajo la curva ROC (AUC-ROC)

Objetivo Específico 3

Optimizar el preprocesamiento de datos textuales y la gestión del desbalance de clases para mejorar la precisión del modelo predictivo de accidentes laborales.

- **Variable Independiente:** Técnicas de preprocesamiento y manejo del desbalance de clases.
 - **Indicadores:**
 - Métodos de preprocesamiento (limpieza de datos, *tokenización*, *lematización*)
 - Técnicas de balanceo de datos (*oversampling*, *undersampling*, SMOTE)
 - Calidad de los datos preprocesados (medida de ruido, consistencia)
- **Variable Dependiente:** Precisión del modelo predictivo
 - **Indicadores:**
 - Precisión del modelo (*accuracy*) antes y después de la optimización

- Sensibilidad (*recall*) antes y después de la optimización
- Especificidad (*precision*) antes y después de la optimización
- F1-Score antes y después de la optimización

Objetivo Específico 4

Determinar los factores críticos que contribuyen significativamente a la ocurrencia y severidad de accidentes laborales en la minería subterránea mediante el análisis de datos con técnicas de *machine learning*.

- **Variable Independiente:** Factores críticos identificados
 - **Indicadores:**
 - Importancia de los factores según el modelo de *machine learning* (importancia de características en *Random Forest*)
 - Frecuencia de aparición de factores en incidentes reportados
 - Correlación de factores con la severidad de los accidentes
- **Variable Dependiente:** Ocurrencia y severidad de los accidentes laborales
 - **Indicadores:**
 - Frecuencia de accidentes por factor crítico
 - Severidad de los accidentes (número de lesiones, fatalidades, días perdidos)

Objetivo Específico 5

Desarrollar estrategias de prevención de accidentes basadas en el análisis de los datos y resultados obtenidos mediante la aplicación de inteligencia artificial en la minería subterránea.

- **Variable Independiente:** Estrategias de prevención desarrolladas
 - **Indicadores:**
 - Tipo de estrategia (tecnológica, procedimental, formativa)
 - Recursos invertidos en la implementación de estrategias
 - Tiempo de implementación de las estrategias

- **Variable Dependiente:** Reducción de accidentes laborales
 - **Indicadores:**
 - Disminución en la frecuencia de accidentes laborales
 - Reducción en la severidad de los accidentes (menor número de lesiones y fatalidades)
 - Evaluación de la efectividad de las estrategias (encuestas de percepción, auditorías de seguridad)

Estas variables e indicadores han permitido una medición clara y precisa de los objetivos específicos, facilitando la evaluación del impacto de la inteligencia artificial y el *machine learning* en la mejora de la seguridad laboral en la minería subterránea del Perú. Las variables e indicadores se muestran en la Tabla 4.

Tabla 4. Variables e Indicadores – Operacionalización de Variables

Hipótesis	Variable	Indicadores	Descripción	Dimensión	Tipo de Función	Escala	Nivel
H1	Factores de riesgo	Número de incidentes	Incidentes reportados por desprendimiento de rocas	Condiciones geológicas	Independiente	Nominal	Básico
		Mantenimiento de equipos	Estado de los equipos de minería	Equipos		Ordinal	
		Capacitación del personal	Nivel de formación y experiencia de los trabajadores	Personal		Intervalo	
		Procedimientos de seguridad	Implementación de procedimientos de seguridad	Procedimientos		Nominal	
H2	Algoritmo de <i>machine learning</i>	Tipo de algoritmo	Algoritmo utilizado para la predicción	Modelo		Nominal	
		Parámetros del modelo	Configuración de los parámetros del modelo	Parámetros		Intervalo	
		Tiempo de entrenamiento	Duración del entrenamiento del modelo	Rendimiento		Ratio	
H3	Técnicas de preprocesamiento y manejo del desbalance de clases	Métodos de preprocesamiento	Técnicas para mejorar la calidad de los datos	Datos preprocesados		Nominal	
		Técnicas de balanceo de datos	Métodos para balancear las clases de datos	Balance de datos		Nominal	
		Calidad de los datos preprocesados	Medida de la calidad de los datos	Calidad de datos		Ordinal	
H4	Factores críticos identificados	Importancia de los factores	Puntuación de importancia de cada factor	Factores críticos		Intervalo	
		Frecuencia de factores	Frecuencia de aparición de factores en incidentes	Frecuencia		Nominal	
		Correlación de factores	Factores críticos y severidad de accidentes	Correlación	Intervalo		
H5	Estrategias de prevención desarrolladas	Tipo de estrategia	Categoría de las estrategias implementadas	Tipo de estrategia	Nominal		
		Recursos invertidos	Cantidad de recursos utilizados	Recursos	Intervalo		
		Tiempo de implementación	Periodo de implementación de las estrategias	Tiempo	Ratio		
	Reducción de accidentes laborales	Disminución en la frecuencia de accidentes	Cambio en el número de accidentes laborales	Frecuencia	Ratio		
		Reducción en la severidad de los accidentes	Cambio en la severidad de los accidentes	Severidad	Intervalar		
		Efectividad de las estrategias	Percepción de la efectividad de las estrategias	Efectividad	Ordinal		

3.4. Instrumentos

3.4.1. *Técnicas de Procesamiento de Datos*

Para el procesamiento de datos se emplearon las técnicas de *Batch*, *Streaming* y *Lambda*. El procesamiento *Batch*, también conocido como procesamiento por lotes, divide el gran volumen de datos en partes más pequeñas o lotes y los procesa de manera secuencial (Kiran et al., 2015). La arquitectura *Lambda* combina las ventajas del procesamiento por lotes con el procesamiento en tiempo real, representando la tecnología más reciente para gestionar grandes volúmenes de datos (Gribaudo et al., 2018).

En el marco de la investigación, se han seguido estas etapas básicas para cada una de las técnicas mencionadas:

- Recolección de información.
- Procesamiento de datos.
- Análisis y presentación de resultados.

Los pasos específicos para el procesamiento de datos son los siguientes:

- Recopilación de datos utilizando los instrumentos definidos.
- Selección de los datos en función de las variables del estudio.
- Uso de software como SPSS.V25, para el procesamiento de los datos.
- Generación de reportes para el análisis.
- Retroalimentación de los datos para validar los resultados.
- Preparación de tablas y figuras para presentar los resultados.

3.4.2. *Técnicas de Análisis e Interpretación de la Información*

Para el logro de los objetivos planteados en esta investigación, se aplicaron diversas técnicas de análisis e interpretación de la información. Estas técnicas permitieron un enfoque riguroso y sistemático para comprender y evaluar los datos obtenidos, garantizando la validez y confiabilidad de los resultados. A continuación, se detallan las técnicas que se utilizaron:

⇒ **Análisis Descriptivo**

El análisis descriptivo se utilizó para resumir y organizar los datos recolectados. Esta técnica permitió comprender las características principales de los datos y presentar una visión general de las variables de estudio. Las técnicas específicas que se aplicaron incluyen:

- **Frecuencias y Porcentajes:** Para medir la ocurrencia de ciertos eventos o características dentro de los datos.
- **Medidas de Tendencia Central:** Como la media, mediana y moda, para identificar los valores centrales de las variables.
- **Medidas de Dispersión:** Como la desviación estándar, varianza y rango, para entender la variabilidad de los datos.

Estas técnicas proporcionaron una base sólida para el análisis más profundo y permitió identificar patrones y tendencias dentro de los datos de accidentes laborales en la minería subterránea.

⇒ **Análisis de Correlación**

El análisis de correlación se utilizó para evaluar la relación entre dos o más variables. Esta técnica es fundamental para identificar asociaciones y posibles interacciones entre los factores de riesgo y la severidad de los accidentes laborales. Las herramientas específicas incluyen:

- **Coefficiente de Correlación de Pearson:** Para variables continuas, mide la fuerza y dirección de la relación lineal entre dos variables.
- **Coefficiente de Correlación de Spearman:** Para variables ordinales o no lineales, mide la asociación monotónica entre dos variables.

Este análisis permitió identificar qué variables están relacionadas y en qué medida, lo que es crucial para entender los factores que contribuyen a la ocurrencia y severidad de los accidentes.

⇒ **Análisis Predictivo con *Machine Learning***

El análisis predictivo con técnicas de *machine learning* fué central en esta investigación. Se utilizaron algoritmos avanzados para construir modelos que puedan predecir la severidad de los accidentes laborales basándose en los datos históricos. Las etapas del análisis predictivo incluyen:

- **Selección del Algoritmo:** Se utilizarán varios algoritmos de *machine learning*, como *Random Forest*, *Support Vector Machines (SVM)*, y *Naive Bayes*, para identificar el modelo más efectivo.
- **Entrenamiento del Modelo:** Los datos históricos se dividieron en conjuntos de entrenamiento y prueba. El modelo se entrenó con el conjunto de entrenamiento y se validó con el conjunto de prueba.
- **Evaluación del Modelo:** Se evaluó la precisión del modelo utilizando métricas como precisión, sensibilidad, especificidad, F1-Score, y área bajo la curva ROC (AUC-ROC).

El objetivo ha sido desarrollar un modelo robusto que pueda identificar y predecir futuros accidentes con alta precisión, permitiendo la implementación de medidas preventivas efectivas.

⇒ **Análisis de Componentes Principales (PCA)**

El Análisis de Componentes Principales (PCA) se empleó para reducir la dimensionalidad de los datos y eliminar la multicolinealidad entre variables. Esta técnica transformó las variables originales en un nuevo conjunto de variables no correlacionadas llamadas componentes principales. El PCA ayuda a:

- **Identificar las Variables Más Importantes:** Reducir el número de variables a las más significativas sin perder información esencial.
- **Visualizar Patrones:** Facilita la visualización de los datos en un espacio reducido, destacando patrones y relaciones entre variables.

El PCA es útil para simplificar el modelo predictivo y mejorar su interpretación.

⇒ **Análisis de Regresión**

El análisis de regresión se utiliza para modelar y analizar la relación entre la variable dependiente (severidad de los accidentes) y una o más variables independientes (factores de riesgo). Las técnicas específicas incluyen:

- **Regresión Lineal Múltiple:** Para evaluar la influencia de múltiples variables independientes continuas sobre la variable dependiente.
- **Regresión Logística:** Para modelar la probabilidad de ocurrencia de un evento binario, como la ocurrencia de un accidente grave.

Este análisis permite cuantificar el impacto de cada factor de riesgo en la severidad de los accidentes y desarrollar modelos predictivos precisos.

⇒ **Simulación de Monte Carlo**

La simulación de Monte Carlo se utiliza para modelar y cuantificar la incertidumbre en las predicciones de accidentes laborales. Esta técnica involucra:

- **Generación de Datos Aleatorios:** Basados en distribuciones de probabilidad de las variables de estudio.
- **Múltiples Iteraciones:** Realizar numerosas simulaciones para obtener una distribución de posibles resultados.
- **Análisis de Resultados:** Evaluar la variabilidad y la probabilidad de diferentes escenarios de accidentes.

La simulación de Monte Carlo proporciona una comprensión detallada de los posibles resultados y ayuda a desarrollar estrategias de mitigación más efectiva.

⇒ **Validación Cruzada**

La validación cruzada se aplica para evaluar la robustez y generalización de los modelos predictivos. Esta técnica involucra:

- **División del Conjunto de Datos:** En múltiples subconjuntos o pliegues.
- **Entrenamiento y Validación Repetidos:** Entrenar el modelo en diferentes subconjuntos y validar en los restantes.
- **Promedio de Resultados:** Calcular la precisión promedio y otras métricas de evaluación para asegurar la consistencia del modelo.

La validación cruzada garantiza que los modelos predictivos no estén sobre ajustados y sean aplicables a datos nuevos.

⇒ **Análisis de Datos Textuales**

El análisis de datos textuales se utiliza para procesar y extraer información relevante de los reportes de incidentes y descripciones de accidentes. Las técnicas específicas incluyen:

- **Procesamiento de Lenguaje Natural (NLP):** Para tokenización, lematización y extracción de palabras clave.

- **Análisis de Sentimiento:** Para evaluar las percepciones y actitudes reflejadas en los reportes.
- **Modelado de Tópicos:** Para identificar temas recurrentes y patrones en los datos textuales.

Este análisis permite aprovechar al máximo la información contenida en los datos no estructurados y enriquece los modelos predictivos.

Estas técnicas de análisis e interpretación de la información garantizan un enfoque exhaustivo y multidimensional para el estudio de los accidentes laborales en la minería subterránea del Perú, contribuyendo a la mejora de la seguridad y las condiciones de trabajo en este sector.

3.4.3. Instrumentos de Recolección de Datos Según la Metodología

Para la recolección de datos en esta investigación, se utilizaron varios instrumentos que permitieron obtener información precisa y relevante sobre los accidentes laborales en la minería subterránea del Perú. Estos instrumentos han sido seleccionados para garantizar una recopilación de datos exhaustiva y rigurosa, siguiendo las mejores prácticas y metodologías establecidas en los estudios previos.

⇒ Registros de Accidentes Laborales

Fuente: Dirección General de Minería del Ministerio de Energía y Minas del Perú (DGM-MEM).

Descripción: Los registros de accidentes laborales proporcionan datos históricos detallados sobre los incidentes ocurridos en la minería subterránea. Estos registros incluyen información sobre la fecha del accidente, descripción del incidente,

factores de riesgo involucrados, medidas de seguridad implementadas, y la severidad del accidente (lesiones, fatalidades).

Utilización: Estos datos son fundamentales para el análisis descriptivo y predictivo, permitiendo identificar patrones y tendencias en los accidentes laborales.

⇒ **Encuestas a Trabajadores y Supervisores**

Fuente: Trabajadores y supervisores de minas subterráneas en Perú.

Descripción: Las encuestas estarán diseñadas para recopilar información sobre la percepción de los trabajadores y supervisores respecto a las condiciones de seguridad, factores de riesgo y la efectividad de las medidas preventivas. Las encuestas incluyen preguntas tanto abiertas como cerradas, abarcando aspectos como la capacitación en seguridad, el estado de los equipos, y los procedimientos de emergencia.

Utilización: Los datos obtenidos de las encuestas complementan los registros de accidentes y proporcionan una perspectiva cualitativa sobre los factores de riesgo y la cultura de seguridad en las minas.

⇒ **Entrevistas a Expertos en Seguridad Minera**

Fuente: Expertos en seguridad minera, ingenieros de seguridad y gerentes de operaciones.

Descripción: Se realizaron entrevistas semiestructuradas con expertos en seguridad minera para profundizar en la comprensión de los factores críticos que contribuyen a los accidentes laborales. Las entrevistas exploran temas como las mejores prácticas en seguridad, desafíos específicos de la minería subterránea, y recomendaciones para mejorar las condiciones de trabajo.

Utilización: La información cualitativa obtenida de las entrevistas han sido utilizadas para validar y enriquecer los hallazgos cuantitativos, así como para desarrollar estrategias de prevención basadas en la experiencia y conocimiento de los expertos.

⇒ **Observación Directa**

Fuente: Observaciones realizadas en sitios de minas subterráneas.

Descripción: Se llevaron a cabo observaciones directas en varias minas subterráneas para evaluar las condiciones de trabajo, la implementación de medidas de seguridad, y los comportamientos de los trabajadores. Las observaciones se registraron mediante *checklists* y notas de campo.

Utilización: Los datos de observación directa proporcionan una evaluación empírica de las condiciones de seguridad y permitirán identificar discrepancias entre las políticas de seguridad y su implementación práctica.

⇒ **Análisis de Documentos y Reportes Internos**

Fuente: Documentos y reportes internos de las empresas mineras.

Descripción: Se analizaron documentos y reportes internos relacionados con la seguridad laboral, incluyendo políticas de seguridad, procedimientos operativos estándar, reportes de auditorías de seguridad, y programas de capacitación.

Utilización: Estos documentos proporcionaron un contexto adicional y permitieron evaluar la coherencia y efectividad de las políticas de seguridad implementadas por las empresas mineras.

⇒ **Base de Datos de *Machine Learning***

Fuente: Datos preprocesados para el entrenamiento de modelos de *machine learning*.

Descripción: La base de datos incluyen todas las variables relevantes extraídas de los registros de accidentes, encuestas, entrevistas y observaciones. Los datos se estructuraron y preprocesaron para su uso en algoritmos de *machine learning*, asegurando su calidad y coherencia.

Utilización: Esta base de datos ha sido utilizada para entrenar y validar los modelos predictivos que permitieron anticipar la severidad de los accidentes laborales y el desarrollo de estrategias preventivas efectivas.

⇒ **Software de Análisis de Datos**

Fuente: SPSS.V25, Python, R, y otros.

Descripción: Se utilizaron varios paquetes de software para el análisis estadístico y la modelización predictiva. Estos programas permitieron realizar análisis descriptivos, correlacionales, regresiones, y simulaciones de Monte Carlo, así como el entrenamiento de modelos de *machine learning*.

Utilización: El software fue esencial para el procesamiento y análisis de los datos recolectados, facilitando la interpretación de los resultados y la generación de reportes.

⇒ **Proceso De Recolección De Datos**

El proceso de recolección de datos sigue una serie de pasos estructurados para garantizar la validez y fiabilidad de los datos obtenidos:

- **Planificación y Diseño de Instrumentos:** Definición clara de los instrumentos de recolección de datos y diseño de encuestas y entrevistas.
- **Recolección de Datos:** Implementación de los instrumentos de recolección en el campo, incluyendo la administración de encuestas, realización de entrevistas y observaciones directas.
- **Almacenamiento y Preprocesamiento de Datos:** Almacenamiento seguro de los datos recolectados y preprocesamiento de estos para garantizar su calidad y coherencia.
- **Análisis y Validación de Datos:** Análisis estadístico y modelización predictiva de los datos, incluyendo la validación de los modelos desarrollados.
- **Interpretación y Presentación de Resultados:** Interpretación de los resultados obtenidos y preparación de tablas, figuras y reportes para su presentación.

Estos instrumentos y el proceso de recolección de datos aseguraron que la investigación cuente con información rica y detallada, permitiendo un análisis exhaustivo y la formulación de recomendaciones prácticas para mejorar la seguridad laboral en la minería subterránea del Perú.

3.5. Procedimientos

A continuación, se describen los procedimientos específicos que se llevaron a cabo para la validación de cada uno de los instrumentos diseñados para esta investigación.

⇒ **Validación del Formulario de Registro de Accidentes**

Objetivo: Asegurar que el formulario capture de manera precisa y completa todos los detalles relevantes sobre los accidentes laborales.

Procedimientos de Validación:

- **Revisión de Expertos:** El formulario ha sido revisado por expertos en seguridad minera y en prevención de riesgos laborales para asegurar que todas las preguntas y categorías sean relevantes y completas.
- **Prueba Piloto:** Se realizó una prueba piloto del formulario en una mina subterránea seleccionada. Los datos obtenidos se analizaron para identificar posibles problemas de interpretación o categorías faltantes.
- **Análisis de Consistencia Interna:** Se calculó el coeficiente de consistencia interna (por ejemplo, Alfa de Cronbach) para evaluar la fiabilidad del formulario.

⇒ **Validación del Cuestionario de Encuesta a Trabajadores**

Objetivo: Garantizar que el cuestionario mida de manera efectiva las percepciones de los trabajadores sobre la capacitación y los procedimientos de seguridad.

Procedimientos de Validación:

- **Revisión de Contenido:** El cuestionario ha sido revisado por expertos en seguridad laboral y en métodos de encuesta para asegurar que las preguntas sean claras, relevantes y comprensibles.
- **Prueba Piloto:** Se administró el cuestionario a un grupo pequeño de trabajadores en varias minas subterráneas para identificar cualquier problema de comprensión o ambigüedad.
- **Análisis de Validez de Constructo:** Se ha llevado a cabo un análisis factorial para evaluar la estructura del cuestionario y asegurarse de que las preguntas agrupen correctamente en las dimensiones esperadas.

⇒ Validación de la Lista de Verificación de Observación Directa

Objetivo: Asegurar que la lista de verificación sea una herramienta eficaz para evaluar las condiciones de seguridad y el estado de los equipos en las minas.

Procedimientos de Validación:

- **Revisión por Expertos:** La lista ha sido revisado por ingenieros de seguridad y gerentes de operaciones para garantizar que todas las categorías sean relevantes y completas.
- **Prueba en Campo:** La lista de verificación se aplicó en una mina subterránea para evaluar su aplicabilidad y eficacia en un entorno real. Las observaciones se discutieron con el personal de seguridad para identificar posibles mejoras.
- **Cálculo de Concordancia Inter observadores:** Se calculó el coeficiente Kappa para evaluar la concordancia entre diferentes observadores que utilicen la lista de verificación.

⇒ Validación de la Base de Datos de Entrenamiento de Modelos

Objetivo: Asegurar que la base de datos preprocesada sea adecuada para el entrenamiento de modelos de *machine learning* y que contenga datos precisos y consistentes.

Procedimientos de Validación:

- **Revisión de Integridad de Datos:** Se realizaron chequeos de consistencia y limpieza de datos para asegurar que no haya valores perdidos o inconsistentes.
- **Prueba de Entrenamiento de Modelos:** Los datos se utilizaron para entrenar modelos iniciales de *machine learning*. Se evaluaron los resultados para identificar cualquier problema con la calidad de los datos.

- **Análisis de Correlación de Datos:** Se realizaron el análisis de correlación entre las variables para identificar y manejar posibles problemas de multicolinealidad.

⇒ **Validación del Manual de Preprocesamiento de Datos**

Objetivo: Asegurar que las instrucciones para el preprocesamiento de datos sean claras y efectivas.

Procedimientos de Validación:

- **Revisión de Contenido:** El manual ha sido revisado por expertos en procesamiento de datos y en *machine learning* para asegurar que las instrucciones sean precisas y completas.
- **Prueba de Aplicación:** Se realizaron una prueba del manual en un conjunto de datos de muestra para asegurar que las instrucciones puedan seguirse fácilmente y producen datos de alta calidad.
- **Retroalimentación de Usuarios:** Se solicitó retroalimentación de los analistas que utilicen el manual para identificar áreas de mejora y asegurar su aplicabilidad.

⇒ **Validación del Informe de Importancia de Características**

Objetivo: Asegurar que el informe generado por los algoritmos de *machine learning* refleje de manera precisa la importancia de cada variable.

Procedimientos de Validación:

- **Revisión Técnica:** El informe ha sido revisado por expertos en *machine learning* para verificar la precisión de los cálculos y la interpretación de los resultados.
- **Comparación con la Literatura:** Se compararon los resultados del informe con estudios previos y literatura existente para validar la consistencia y relevancia de los hallazgos.

- **Análisis de Sensibilidad:** Se realizaron el análisis de sensibilidad para evaluar la robustez de los resultados ante cambios en los datos o en los parámetros del modelo.

⇒ **Validación de la Encuesta de Percepción de Efectividad**

Objetivo: Garantizar que la encuesta mida de manera efectiva la percepción de los trabajadores y supervisores sobre la efectividad de las estrategias de prevención.

Procedimientos de Validación:

- **Revisión de Contenido:** La encuesta ha sido revisado por expertos en seguridad laboral y en métodos de encuesta para asegurar que las preguntas sean claras, relevantes y comprensibles.
- **Prueba Piloto:** Se administró la encuesta a un grupo pequeño de trabajadores y supervisores para identificar cualquier problema de comprensión o ambigüedad.
- **Análisis de Validez de Constructo:** Se ha llevado a cabo un análisis factorial para evaluar la estructura de la encuesta y asegurarse de que las preguntas agrupen correctamente en las dimensiones esperadas.

⇒ **Validación de la Base de Datos de Reducción de Accidentes**

Objetivo: Asegurar que la base de datos refleje de manera precisa la disminución en la frecuencia y severidad de accidentes laborales tras la implementación de las estrategias.

Procedimientos de Validación:

- **Revisión de Integridad de Datos:** Se realizaron chequeos de consistencia y limpieza de datos para asegurar que no haya valores perdidos o inconsistentes.

- **Prueba de Análisis de Datos:** Los datos se utilizaron para realizar análisis iniciales de reducción de accidentes. Se evaluaron los resultados para identificar cualquier problema con la calidad de los datos.
- **Comparación con Datos Históricos:** Se compararon los datos de reducción de accidentes con datos históricos para validar la precisión y relevancia de los hallazgos.

Estos procedimientos de validación garantizan que los instrumentos de recolección de datos sean fiables y válidos, proporcionando una base sólida para la interpretación y análisis de los resultados de la investigación.

3.6. Análisis de Datos

El análisis de datos en esta investigación se ha llevado a cabo utilizando una combinación de métodos estadísticos y técnicas avanzadas de *machine learning* para asegurar una interpretación rigurosa y completa de la información recolectada. El análisis de datos se ha estructurado en varias etapas, cada una diseñada para abordar los objetivos específicos de la investigación y responder a las hipótesis planteadas.

Etapas 1: Análisis Descriptivo

Objetivo: Describir y resumir las características principales de los datos recolectados.

Procedimientos:

1. **Frecuencias y Porcentajes:** Calcular la ocurrencia de diferentes tipos de accidentes laborales y otros eventos relevantes.
2. **Medidas de Tendencia Central:** Calcular la media, mediana y moda para variables clave como el número de incidentes, la severidad de los accidentes, etc.

3. **Medidas de Dispersión:** Calcular la desviación estándar, varianza y rango para entender la variabilidad de los datos.

Herramientas Utilizadas: SPSS.V25, Excel.

Resultados Esperados:

- Descripción detallada de las características de los accidentes laborales.
- Identificación de patrones y tendencias en los datos.

Etapa 2: Análisis de Correlación

Objetivo: Evaluar la relación entre dos o más variables y determinar la fuerza y dirección de estas relaciones.

Procedimientos:

- ✓ **Coefficiente de Correlación de Pearson:** Utilizado para variables continuas para medir la relación lineal entre ellas.
- ✓ **Coefficiente de Correlación de Spearman:** Utilizado para variables ordinales o no lineales.

Herramientas Utilizadas: SPSS.V25, Python.

Resultados Esperados:

- Identificación de variables que están significativamente correlacionadas.
- Entendimiento de cómo los factores de riesgo están relacionados con la severidad de los accidentes.

Etapa 3: Análisis Predictivo con *Machine Learning*

Objetivo: Desarrollar modelos predictivos para anticipar la severidad de los accidentes laborales basados en los datos históricos.

Procedimientos:

- ✓ **Selección del Algoritmo:** Probar diferentes algoritmos de *machine learning* como *Random Forest*, *Support Vector Machines (SVM)* y *Naive Bayes*.
- ✓ **Entrenamiento del Modelo:** Dividir los datos en conjuntos de entrenamiento y prueba, y entrenar los modelos utilizando el conjunto de entrenamiento.
- ✓ **Evaluación del Modelo:** Evaluar la precisión del modelo utilizando métricas como precisión, sensibilidad, especificidad, F1-Score y área bajo la curva ROC (AUC-ROC).

Herramientas Utilizadas: Python (*scikit-learn*, *TensorFlow*), R.

Resultados Esperados:

- Desarrollo de un modelo predictivo robusto.
- Predicciones precisas de la severidad de futuros accidentes laborales.

Etapa 4: Análisis de Componentes Principales (PCA)

Objetivo: Reducir la dimensionalidad de los datos y eliminar la multicolinealidad entre las variables.

Procedimientos:

- ✓ **Transformación de Variables:** Convertir las variables originales en un nuevo conjunto de variables no correlacionadas llamadas componentes principales.
- ✓ **Selección de Componentes:** Seleccionar los componentes principales que expliquen la mayor parte de la varianza en los datos.

Herramientas Utilizadas: Python (*scikit-learn*), R.

Resultados Esperados:

- Simplificación del conjunto de datos.

- Mejora en la interpretabilidad del modelo predictivo.

Etapas 5: Análisis de Regresión

Objetivo: Modelar y analizar la relación entre la variable dependiente (severidad de los accidentes) y una o más variables independientes (factores de riesgo).

Procedimientos:

- ✓ **Regresión Lineal Múltiple:** Evaluar la influencia de múltiples variables independientes continuas sobre la variable dependiente.
- ✓ **Regresión Logística:** Modelar la probabilidad de ocurrencia de un evento binario, como la ocurrencia de un accidente grave.

Herramientas Utilizadas: SPSS.V25, Python (*statsmodels*)

Resultados Esperados:

- Cuantificación del impacto de cada factor de riesgo en la severidad de los accidentes.
- Desarrollo de modelos predictivos precisos.

Etapas 6: Simulación de Monte Carlo

Objetivo: Modelar y cuantificar la incertidumbre en las predicciones de accidentes laborales.

Procedimientos:

1. **Generación de Datos Aleatorios:** Basados en distribuciones de probabilidad de las variables de estudio.
2. **Múltiples Iteraciones:** Realizar numerosas simulaciones para obtener una distribución de posibles resultados.

3. **Análisis de Resultados:** Evaluar la variabilidad y la probabilidad de diferentes escenarios de accidentes.

Herramientas Utilizadas: Python (numpy, pandas, scipy).

Resultados Esperados:

- Entendimiento detallado de los posibles resultados y su probabilidad.
- Desarrollo de estrategias de mitigación más efectivas.

Etapa 7: Validación Cruzada

Objetivo: Evaluar la robustez y generalización de los modelos predictivos.

Procedimientos:

- ✓ **División del Conjunto de Datos:** En múltiples subconjuntos o pliegues.
- ✓ **Entrenamiento y Validación Repetidos:** Entrenar el modelo en diferentes subconjuntos y validar en los restantes.
- ✓ **Promedio de Resultados:** Calcular la precisión promedio y otras métricas de evaluación para asegurar la consistencia del modelo.

Herramientas Utilizadas: Python (*scikit-learn*), R.

Resultados Esperados:

- Aseguramiento de que los modelos predictivos no estén sobre ajustados.
- Aplicabilidad de los modelos a datos nuevos.

Etapa 8: Análisis de Datos Textuales

Objetivo: Procesar y extraer información relevante de los reportes de incidentes y descripciones de accidentes.

Procedimientos:

- ✓ **Procesamiento de Lenguaje Natural (NLP):** Tokenización, lematización y extracción de palabras clave.
- ✓ **Análisis de Sentimiento:** Evaluar las percepciones y actitudes reflejadas en los reportes.
- ✓ **Modelado de Tópicos:** Identificar temas recurrentes y patrones en los datos textuales.

Herramientas Utilizadas: Python (NLTK, spaCy).

Resultados Esperados:

- Extracción de información valiosa de datos no estructurados.
- Enriquecimiento de los modelos predictivos.

Etapa 9: Análisis Cualitativo

Objetivo: Analizar las entrevistas y encuestas cualitativas realizadas a expertos en seguridad minera.

Procedimientos:

- ✓ **Codificación de Datos:** Identificación y categorización de temas y patrones en los datos cualitativos.
- ✓ **Análisis de Contenido:** Evaluación de la frecuencia y relevancia de las categorías identificadas.
- ✓ **Triangulación de Datos:** Comparación de datos cualitativos con resultados cuantitativos para validar hallazgos.

Herramientas Utilizadas: NVivo, Atlas.ti.

Resultados Esperados:

- Complemento al análisis cuantitativo.
- Perspectiva más rica y contextualizada de los factores de riesgo y estrategias de prevención.

Estos procedimientos aseguran un análisis de datos riguroso y completo, permitiendo una interpretación precisa y la formulación de recomendaciones prácticas para mejorar la seguridad laboral en la minería subterránea del Perú.

3.7. Consideraciones Éticas

Los aspectos éticos del proyecto de investigación se detallan a continuación:

- El plan de tesis cumple con la estructura del plan de tesis de la Universidad Nacional Federico Villarreal.
- El objetivo fundamental del desarrollo del presente trabajo es la generación de conocimiento en el campo de la ingeniería civil.
- El plan de tesis es original, donde plantea el desarrollo de nuevos conocimiento y herramientas para la gestión de riesgos en minería subterránea.

La información utilizada en la investigación esta citada y referenciada, respetando la autoría.

IV. RESULTADOS

La presente investigación se desarrolló siguiendo la metodología planteada, utilizando diversas técnicas de análisis e interpretación de la información. Los pasos considerados fueron: Análisis Descriptivo, Análisis de Correlación, Análisis Predictivo con *Machine Learning*, Análisis de Componentes Principales (PCA), Análisis de Regresión, Validación Cruzada y Análisis de Datos Textuales, siguiendo un orden lógico en las técnicas de procesamiento de datos. Para cada procedimiento, se mantuvo el rigor necesario para asegurar la obtención de resultados de alta calidad. Los resultados se presentan de acuerdo con las técnicas mencionadas, alineando cada uno con los objetivos específicos de la investigación.

4.1. Introducción

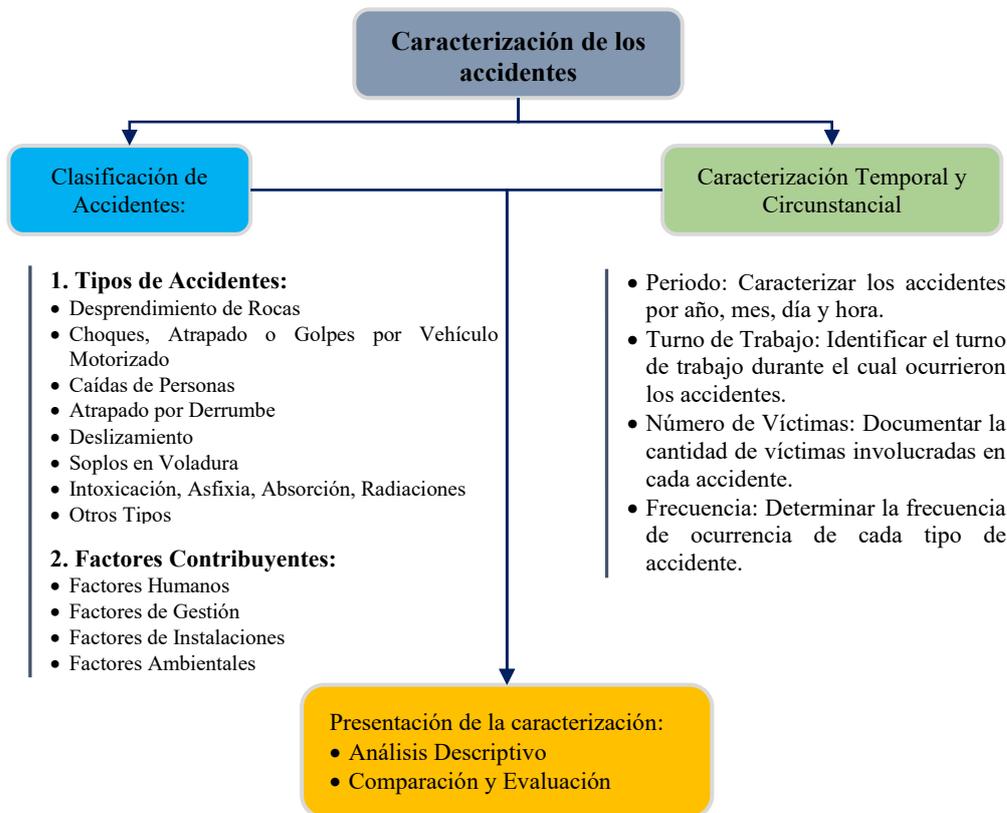
Para la investigación y con el fin de identificar y analizar los principales factores de riesgo asociados a los accidentes laborales por desprendimiento de rocas en la minería subterránea peruana, se procedió, según el primer objetivo, a la búsqueda de información sobre los tipos de accidentes que ocurren en la minería. Esta información se caracterizó por periodo (años, meses, días y horas), tipo (fuente, condición), turno de trabajo, número de víctimas y frecuencia, así como por los factores humanos, de gestión, instalación y ambientales. A continuación, se presentan los resultados organizados según el cumplimiento de los objetivos.

4.1.1. Caracterización de los Tipos de Accidentes

En la presente investigación, se identificaron y analizaron los principales factores de riesgo asociados a los accidentes laborales en la minería subterránea peruana. A continuación, se presentan los resultados detallados sobre los tipos de accidentes más comunes y su caracterización según diferentes factores. Ver Figura 19.

Figura 19

Metodología Aplicada al Análisis y Evaluación de Accidentes



Nota. En la figura se muestra la metodología a seguir en el análisis y evaluación de accidentes.

4.1.1.1. Tipos de Accidentes.

- **Desprendimiento de Rocas:** Este tipo de accidente ocurre cuando fragmentos de roca se desprenden del techo o las paredes del túnel, cayendo sobre los trabajadores. Los factores que contribuyen a estos incidentes incluyen condiciones geológicas inestables, errores en la colocación de soportes, y vibraciones causadas por explosiones y excavaciones cercanas. Factores humanos como la falta de capacitación y el incumplimiento de los protocolos de seguridad, así como una gestión ineficaz y la instalación de equipos de soporte insuficientes o defectuosos.

- **Choques Contra o Atrapado en o Golpes por Vehículo Motorizado:** Estos accidentes se producen cuando los trabajadores son golpeados, atrapados o colisionan con vehículos motorizados utilizados en la minería, como camiones y excavadoras. Las causas incluyen la falta de visibilidad, mantenimiento inadecuado de los vehículos, falta de formación adecuada del personal y el incumplimiento de las normas de tránsito dentro de las minas. La gestión deficiente, la falta de señalización adecuada y la supervisión ineficaz, junto con condiciones ambientales adversas como la mala iluminación, aumentan estos riesgos.
- **Caídas de Personas:** Las caídas representan un peligro significativo en las minas, especialmente en áreas con desniveles y plataformas elevadas. Estas caídas pueden ser provocadas por superficies resbaladizas, falta de barandillas y equipo de protección personal inadecuado. Factores humanos como la falta de atención y la prisa aumentan el riesgo, mientras que la gestión ineficaz y la falta de instalaciones adecuadas para prevenir caídas también contribuyen a estos accidentes.
- **Atrapado por Derrumbe:** Los derrumbes, que implican el colapso parcial o total de las estructuras mineras, pueden atrapar a los trabajadores. Las causas incluyen fallas en la geomecánica, soportes estructurales insuficientes y prácticas de excavación inadecuadas. La supervisión deficiente, la falta de planificación y evaluación de riesgos, así como la instalación inadecuada de sistemas de soporte y condiciones ambientales adversas, son factores contribuyentes.

- **Deslizamiento:** Los deslizamientos pueden ocurrir cuando las capas de suelo o roca se desplazan, provocando el colapso de las áreas mineras. Las condiciones geológicas inestables, el drenaje deficiente y la excavación inadecuada pueden causar deslizamientos. La gestión inadecuada y la supervisión insuficiente para monitorear las condiciones del terreno, junto con instalaciones de soporte ineficaces, aumentan el riesgo. Los factores ambientales, como las lluvias intensas, también pueden desencadenar estos eventos.
- **Soplos en Voladura:** Los soplos en voladura ocurren durante las operaciones de voladura cuando no se siguen los procedimientos de seguridad adecuados, lo que puede resultar en explosiones prematuras o no controladas. La formación insuficiente, la manipulación inadecuada de explosivos y la falta de coordinación son factores humanos que contribuyen a estos accidentes. La gestión deficiente, incluyendo la planificación inadecuada y la supervisión insuficiente, así como instalaciones de almacenamiento de explosivos no conformes, también juegan un papel importante.
- **Intoxicación, Asfixia, Absorción y Radiaciones:** Estos accidentes resultan de la exposición a gases tóxicos, falta de oxígeno, sustancias químicas peligrosas y radiaciones. Las causas incluyen la ventilación inadecuada, la falta de equipos de protección personal adecuados y la formación insuficiente sobre los peligros químicos. La gestión deficiente en el monitoreo de la calidad del aire y la implementación de procedimientos de emergencia, junto con instalaciones inadecuadas para la detección y control de gases tóxicos, aumentan el riesgo. Las condiciones ambientales, como la acumulación de gases peligrosos en áreas mal ventiladas, también son factores importantes.

- **Otros Tipos de Accidentes:** Además de los mencionados, existen otros accidentes que pueden ocurrir en la minería, como incendios, electrocuciones y accidentes mecánicos. Estos incidentes pueden ser causados por equipos defectuosos, procedimientos de trabajo inseguros y falta de mantenimiento adecuado. La gestión inadecuada que no implementa programas de seguridad robustos y la falta de instalaciones adecuadas para la prevención y respuesta a emergencias también son factores contribuyentes. Las condiciones ambientales adversas, como la humedad y las altas temperaturas, pueden agravar estos riesgos.

4.1.1.2. Evolución de los Accidentes Mortales en la Minería Peruana (1970-2018).

La Figura 20, muestra una notable disminución en el número de accidentes mortales en la minería peruana a lo largo de los últimos 49 años, pasando de cifras de tres dígitos a dos dígitos. Entre 1970 y 1990, el promedio anual fue de 102 accidentes mortales, disminuyendo a 84 durante la década siguiente (1991-2000). En los últimos 18 años (2001-2018), el descenso fue aún más pronunciado, con un promedio anual de 60 accidentes mortales. Los años con mayor número de accidentes fueron 1985 (136), 1984 (122), 1980 (120) y 1996 (120), mientras que los años con menor número fueron 2018 (27), 2015 (29) y 2014 (32).

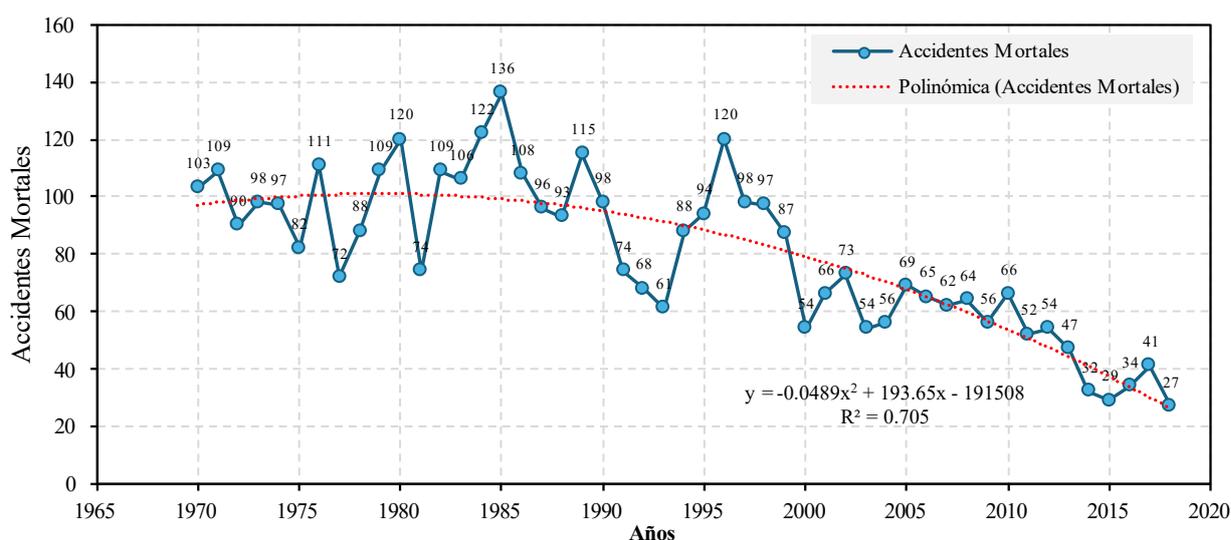
Este descenso acelerado en los últimos años se atribuye a la implementación de Departamentos de Geomecánica, la tipificación de macizos rocosos, la aplicación de elementos de sostenimiento más apropiados, y el incremento de las exigencias en seguridad, incluyendo la constitución de gerencias de seguridad y la introducción de nuevos elementos de sostenimiento, lo cual ha mejorado significativamente la seguridad en las operaciones mineras.

Interpretación del R^2 y la Línea Polinómica. En la Figura 20, se observa una línea polinómica que representa la tendencia general de los accidentes mortales a lo largo del tiempo.

Esta línea es una aproximación matemática que intenta ajustar la variación de los datos reales de accidentes mortales. La línea polinómica ajustada y el R^2 de 0.705 indican que el modelo utilizado es bastante efectivo para explicar la mayoría de las variaciones en el número de accidentes mortales a lo largo del tiempo. La tendencia general decreciente reflejada en la línea polinómica y el alto valor de R^2 sugieren que las medidas implementadas para mejorar la seguridad en la minería peruana han sido efectivas en reducir significativamente el número de accidentes mortales en las últimas décadas.

Figura 20

Accidentes Mortales en la Minería Peruana de 1970 al 2018



Nota. En la figura se observa que el número de accidentes fatales en la minería peruana ha venido decreciendo a través del tiempo. Los datos han sido obtenidos del Ministerio de Energía y Minas.

Interpretación del R^2 y la Línea Polinómica. En la Figura 20, se observa una línea polinómica que representa la tendencia general de los accidentes mortales a lo largo del tiempo. Esta línea es una aproximación matemática que intenta ajustar la variación de los datos reales de accidentes mortales. La línea polinómica ajustada y el R^2 de 0.705 indican que el modelo utilizado es bastante efectivo para explicar la mayoría de las variaciones en el número de

accidentes mortales a lo largo del tiempo. La tendencia general decreciente reflejada en la línea polinómica y el alto valor de R^2 sugieren que las medidas implementadas para mejorar la seguridad en la minería peruana han sido efectivas en reducir significativamente el número de accidentes mortales en las últimas décadas.

4.1.1.3. Accidentes Mortales en los Últimos 18 Años en la Minería.

A continuación, se presentan los accidentes mortales ocurridos en el periodo comprendido entre el año 2000 y el 2018. Estos accidentes han sido cuantificados mensualmente, de enero a diciembre, y clasificados según el tipo de accidente. El análisis detallado proporciona una visión integral de la frecuencia y naturaleza de los accidentes mortales a lo largo de estos años, permitiendo identificar patrones y tendencias relevantes para la implementación de medidas preventivas y de seguridad. Ver la Tabla 5 donde se muestran los accidentes mortales en la minería peruana del 2000 al 2018.

Tabla 5

Accidentes Mortales en la Minería Peruana del 2000 al 2018

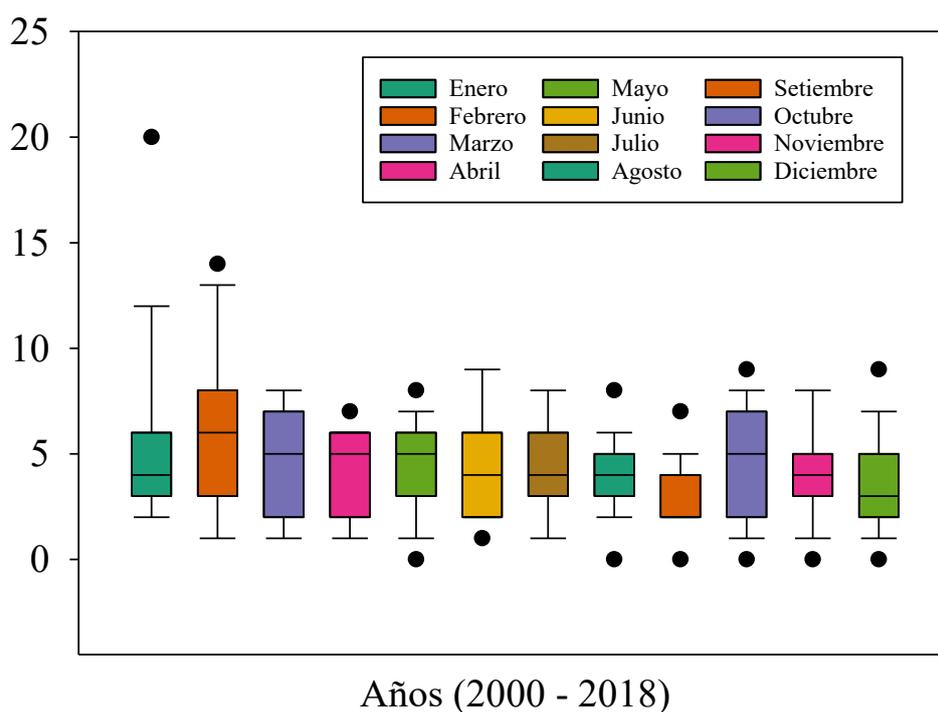
AÑO	ENE.	FEB.	MAR.	ABR.	MAY.	JUN.	JUL.	AGO.	SEP.	OCT.	NOV.	DIC.	Total
2000	6	4	2	3	3	6	8	0	0	7	8	7	54
2001	2	9	5	5	8	3	8	8	4	5	4	5	66
2002	20	2	4	6	5	5	4	6	4	8	8	1	73
2003	4	8	5	7	5	3	4	5	3	3	4	3	54
2004	2	9	8	5	2	9	1	3	4	7	5	1	56
2005	3	8	6	6	6	3	5	3	7	5	8	9	69
2006	6	7	6	3	6	5	6	5	4	9	4	4	65
2007	5	6	7	3	7	6	4	6	5	6	5	2	62
2008	12	5	7	6	3	5	6	6	5	3	3	3	64
2009	4	14	6	2	3	8	6	4	2	1	4	2	56
2010	5	13	1	6	5	9	6	4	3	4	4	6	66
2011	4	8	2	5	6	5	4	5	4	5	1	3	52
2012	2	6	8	2	4	2	5	5	3	8	4	4	53
2013	4	6	5	6	1	4	4	4	5	2	4	2	47
2014	6	1	1	1	1	3	7	2	2	0	1	7	32
2015	5	2	7	2	0	2	1	2	2	3	3	0	29
2016	4	3	3	1	6	2	2	3	4	1	2	3	34
2017	5	5	3	2	6	1	3	4	2	8	0	2	41
2018	2	1	2	5	3	2	1	3	2	2	3	1	27
Total	101	117	88	76	80	83	85	78	65	87	75	65	1,000

Nota. La Tabla 5 muestra los accidentes mortales ocurridos mensualmente durante el periodo comprendido entre los años 2000 y 2018. Los datos corresponden al Ministerio de Energía y Minas del Perú.

La figura adjunta presenta un análisis de los accidentes mortales ocurridos mensualmente durante el periodo comprendido entre los años 2000 y 2018. El gráfico de caja (*box plot*) utilizado permite observar la distribución de los accidentes por mes a lo largo de estos años, ofreciendo una visión clara de la variabilidad y los patrones en los datos. Ver la Figura 21 donde se muestran los accidentes mortales mensuales durante los años 2000 y 2018.

Figura 21

Accidentes Mortales Mensuales Durante los Años 2000 y 2018



Nota. La figura muestra los accidentes mortales ocurridos mensualmente durante los años 2000 y 2018 a partir de los datos del Ministerio de Energía y Minas del Perú.

El eje X del Figura 21 representa los años del periodo 2000-2018, mientras que el eje Y muestra el número de accidentes mortales registrados. Los diferentes colores en la leyenda

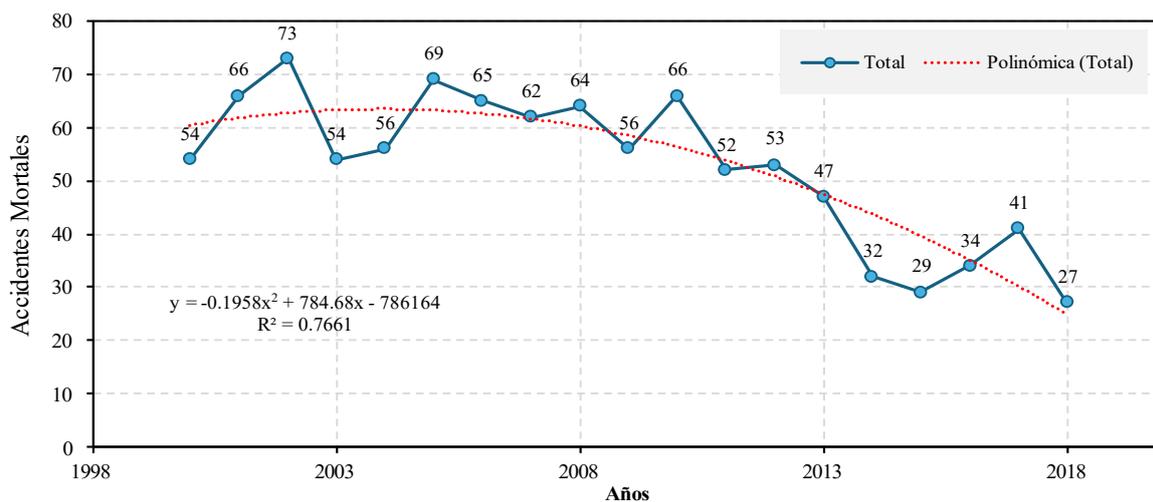
indican los meses del año, desde enero (verde) hasta diciembre (azul). Analizando la distribución mensual de los accidentes, se observa que en enero la mediana de accidentes está cerca de 5, con un rango Inter cuartil relativamente amplio y algunos valores atípicos (*outliers*) que superan los 20 accidentes. Febrero presenta una distribución similar a enero, con una mediana alrededor de 5 y un rango Inter cuartil más estrecho.

De marzo a diciembre, los meses muestran distribuciones relativamente homogéneas, con medianas cercanas a 5 accidentes. Sin embargo, existen algunas variaciones significativas: marzo y abril tienen rangos Inter cuartiles más amplios, indicando una mayor variabilidad en el número de accidentes, mientras que mayo y junio presentan distribuciones más concentradas, con rangos Inter cuartiles estrechos y pocas observaciones atípicas. Los meses de julio a diciembre muestran una tendencia similar, con medianas estables y algunos valores atípicos dispersos. Es notable la presencia de valores atípicos en varios meses, especialmente en enero, marzo y septiembre, donde se registran picos significativos de accidentes que superan considerablemente las medianas mensuales. Estos *outliers* sugieren la ocurrencia de eventos extraordinarios que resultaron en un número inusualmente alto de accidentes mortales en esos meses. En términos de patrones y tendencias, se observa una consistencia en la frecuencia de accidentes mortales a lo largo de los años, con medianas mensuales que oscilan alrededor de 5 accidentes. Sin embargo, la mayor variabilidad se encuentra en los meses de enero, marzo y septiembre, lo que sugiere que en estos meses pueden ocurrir fluctuaciones significativas en el número de accidentes, posiblemente debido a factores estacionales o eventos específicos. Finalmente, a lo largo del periodo de estudio, la cantidad de accidentes mortales mensuales se mantiene relativamente estable, con una mediana alrededor de 5 accidentes. No obstante, los meses de enero, marzo y septiembre requieren especial atención debido a la mayor variabilidad y la presencia de *outliers*, lo que podría indicar la necesidad de implementar medidas preventivas adicionales durante estos meses. Este análisis proporciona una visión detallada de

cómo varían los accidentes mortales a lo largo del año, permitiendo identificar periodos críticos y guiar la implementación de estrategias de seguridad minera más efectiva.

Figura 22

Accidentes Mortales por Año del 2000 al 2018 y R2 de Tendencia



Nota. La figura muestra los accidentes mortales por año durante el período comprendido entre los años 2000 y 2018 y R2 de tendencia. Los datos han sido obtenidos del Ministerio de Energía y Minas.

La Figura 22, presenta la sumatoria de los accidentes mortales por año durante el periodo comprendido entre los años 2000 y 2018, mostrando una evolución anual del número total de accidentes y una tendencia polinómica para facilitar la interpretación. El gráfico revela fluctuaciones anuales significativas, con picos notables en años como 2001 y 2006, indicando eventos específicos o condiciones particulares que aumentaron los accidentes en esos periodos. Sin embargo, desde 2007 se observa una tendencia general de disminución en los accidentes mortales, destacada por la línea de tendencia polinómica que muestra un descenso más pronunciado a partir de ese año. En los años recientes (2015-2018), se nota una disminución sostenida en los accidentes, reflejando la posible efectividad de políticas de seguridad minera más estricta. En conclusión, aunque hubo fluctuaciones, la tendencia general descendente sugiere que las medidas implementadas han contribuido positivamente a reducir los accidentes

mortales, destacando la importancia de investigar los factores detrás de los picos y de continuar mejorando las estrategias de seguridad vial. R^2 refleja el porcentaje de la variabilidad en el número total de accidentes mortales que puede explicarse mediante la tendencia polinómica ajustada a los datos. En la Tabla 6 se muestra la caracterización de los tipos de accidentes.

Tabla 6

Caracterización de los Tipos de Accidentes (AT)

Desprendimiento de Rocas	Choques Contra O Atrapado En O Golpes Por Vehículo Motorizado	Caídas De Personas	Atrapado Por Derrumbe, Deslizamiento, Soplos En Voladura	Intoxicación-Asfixia-Absorción-Radiaciones	Otros Tipos
AT-1	AT-2	AT-3	AT-4	AT-5	AT-6
30%	29%	11%	9%	7%	5%

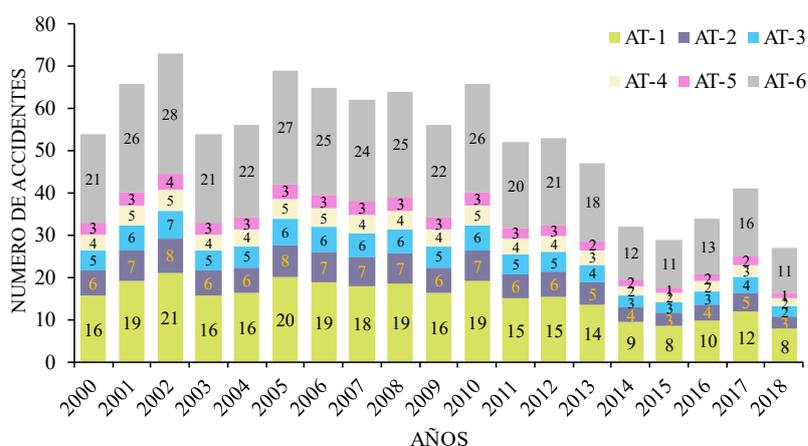
Nota. En la Tabla 6 se observa que la mayor cantidad de accidentes se produce por desprendimiento de rocas.

La Figura 23, muestra la distribución anual de los accidentes mortales por tipo, ocurridos entre los años 2000 y 2018. Los datos han sido categorizados en seis tipos de accidentes, como se indica en la Tabla 6. La Figura 23 adjunta muestra la distribución anual de los accidentes mortales por tipo, ocurridos entre los años 2000 y 2018, categorizados en seis tipos: "Otros tipos" (39%), "Desprendimiento de rocas" (29%), "Choques contra o atrapado en o golpes por vehículo motorizado" (11%), "Caídas de personas" (9%), "Atrapado por derrumbe, deslizamiento, soplos en voladura" (7%) e "Intoxicación, asfixia, absorción, radiaciones" (5%). Durante el periodo 2000-2003, los accidentes oscilaron entre 21 y 28, con predominancia de AT-1 y AT-2. De 2004 a 2009, la cantidad de accidentes mostró fluctuaciones, alcanzando un pico en 2006 (27 accidentes), manteniéndose AT-1 y AT-2 como las categorías más frecuentes. A partir de 2010, se observa una tendencia descendente, registrando un mínimo de 13 accidentes en 2014. Aunque hay ligeras fluctuaciones, el número de accidentes se mantiene relativamente bajo en comparación con años anteriores. La

predominancia de AT-1 y AT-2 sugiere áreas críticas que requieren atención especial en términos de seguridad. La disminución general de accidentes en los años recientes refleja mejoras en las medidas de seguridad y prevención, aunque las fluctuaciones recientes indican la necesidad de continuar mejorando las estrategias de prevención.

Figura 23

Accidentes Mortales en Minería por Tipo de Accidente de 2000-2018



Nota. Los datos corresponden al Ministerio de Energía y Minas.

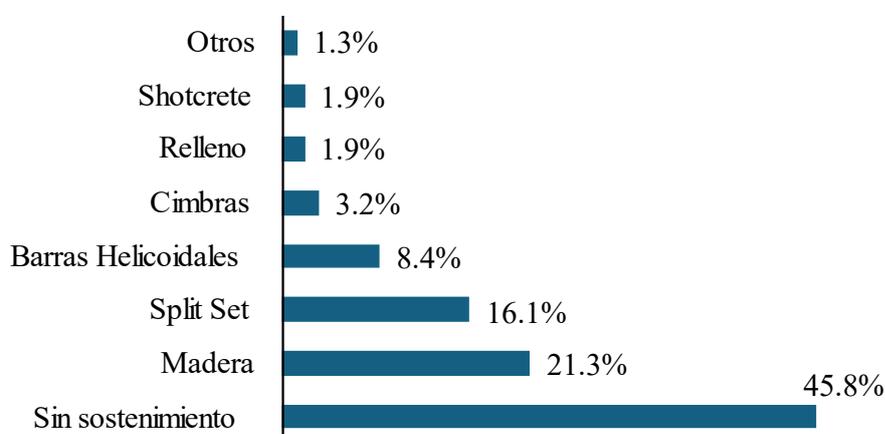
Durante el levantamiento de información en los respectivos expedientes de los accidentes ocurridos, se identificaron los siguientes elementos de sostenimiento mencionados en los informes: Pernos: Split Sets, barras helicoidales e *Hydrabolts*. Cimbras: En herradura y omega. Madera: Cuadros, puntales, wood pack y enramado. *Shotcrete*: Aplicado solo o en combinación con pernos y mallas, con o sin fibra. Mallas de acero: Electrosoldadas, instaladas con pernos con o sin *shotcrete*. Relleno: Relleno hidráulico o detrítico. Otros: *Jack Pot*, cables de acero y pilares.

Para el análisis de los accidentes ocurridos según el tipo de sostenimiento, se tomaron en cuenta los elementos de sostenimiento que, a pesar de estar instalados en el área, no lograron prevenir los desprendimientos de roca, ocasionando accidentes mortales. En algunos casos, los elementos de sostenimiento se vieron comprometidos y en otros no eran los más adecuados

para sostener un determinado macizo rocoso. Estos casos se clasifican de la siguiente manera: Pernos: Perno solo, perno + malla, perno + malla + *shotcrete*. Madera: Cuadros, puntales, Wood pack y enramado. *Shotcrete*: Aplicado sin otro elemento de sostenimiento, con o sin fibra. Cimbras: Enrejado con tablonos o lajas. Relleno: Hidráulico o detrítico. Otros: Cables y pilares.

Figura 24

Accidentes Según Tipo de Sostenimiento



Nota. Los datos corresponden al Ministerio de Energía y Minas.

La Figura 24, adjunta sintetiza esta información, mostrando que el 45.8% de los accidentes mortales se produjeron en zonas sin sostenimiento. Además, el 21.3% de los accidentes mortales ocurrieron en áreas sostenidas con madera, seguidos por sostenimientos con Split Set y barras helicoidales, que representaron el 16.1% y 8.4% de los accidentes, respectivamente.

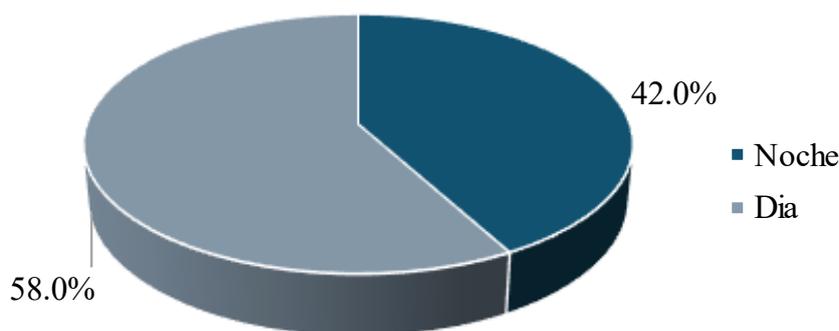
- **Análisis de accidentes Según turno de trabajo**

Este análisis se realizó con el objetivo de determinar si la mayor cantidad de accidentes ocurre durante las guardias nocturnas o diurnas. Los resultados, ilustrados en la Figura 25 se

muestran que el 58% de los accidentes ocurren durante los turnos diurnos, mientras que el 42% se producen en las guardias nocturnas.

Figura 25

Accidentes Mortales Según Turno de Trabajo



Nota. La figura muestra los accidentes mortales según turno de trabajo a partir de los datos del Ministerio de Energía y Minas del Perú.

- **Análisis de accidentes según ocupación de las víctimas**

Para este análisis, se han considerado las siguientes ocupaciones de los trabajadores que resultaron víctimas de los accidentes:

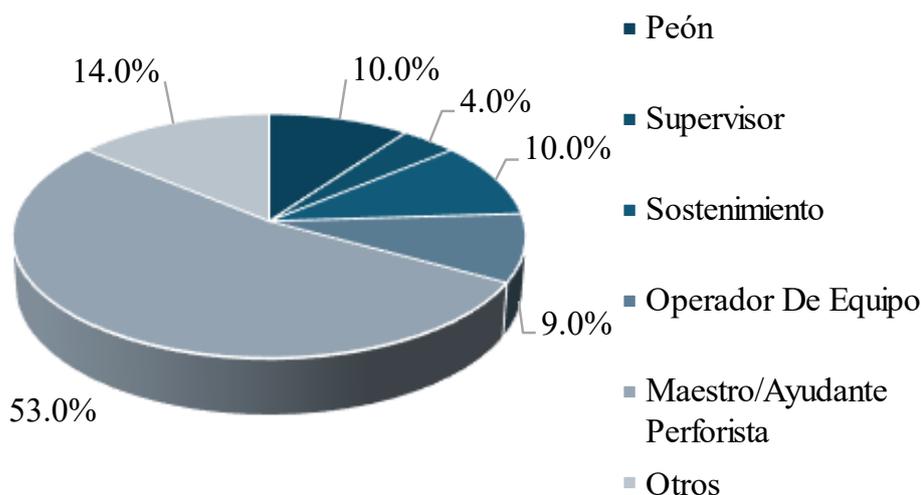
- Maestro / Ayudante perforista: Incluye al perforista con equipo manual, ayudante de perforista, disparador, cargado de explosivos y perforista diamantino.
- Operador de equipo: Comprende al operador de Jumbo, *Scoop*, chofer, locomotorista y compresorista.
- Sostenimiento: Incluye al empernador, shotcretero, personal de relleno hidráulico y detrítico, enmaderador y sus respectivos ayudantes.
- Supervisor: Incluye a ingenieros, jefes de guardia, capataces y servidores generales de mina.

- Peón: Incluye a peones, mineros y ayudantes de mina.
- Otros: Incluye ocupaciones como extracción de mineral, ayudante de ventilación, oficial de obra, maestro minero, tubero, operador de mina, electricista, muestrero, maestro de segunda, bodeguero y carrilano.

Este desglose permite identificar con mayor precisión las ocupaciones más vulnerables y las circunstancias en las que se producen los accidentes, facilitando así la implementación de medidas preventivas más efectivas.

Figura 26

Accidentes Mortales Según Ocupación de las Víctimas



Nota. La Figura 26 muestra los accidentes mortales según ocupación de las víctimas a partir de los datos del Ministerio de Energía y Minas del Perú.

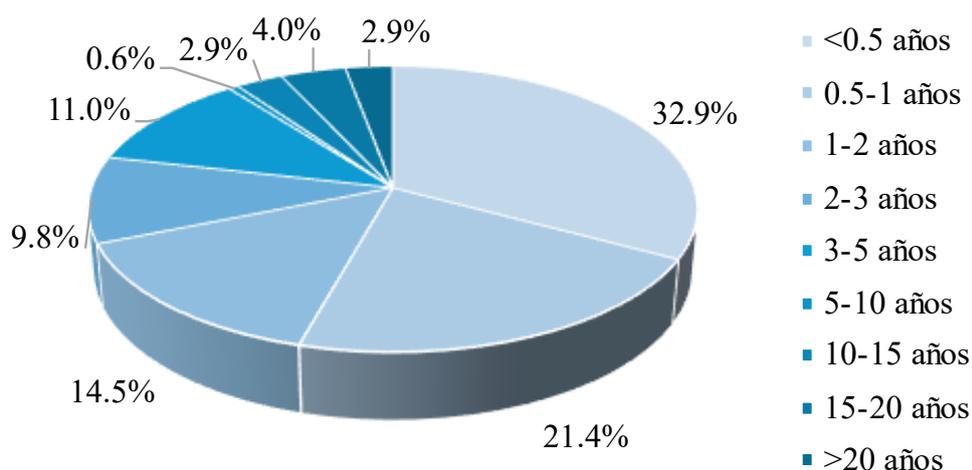
- **Análisis de accidentes según el tiempo de servicio de la víctima en su último centro laboral**

La Figura 27, muestra la distribución de accidentes mortales según el tiempo de servicio de la víctima, destacando que el 36% de los accidentes afecta a trabajadores con menos de un año de servicio, lo que sugiere una alta vulnerabilidad entre los empleados nuevos o con poca

experiencia. El 24% de los accidentes corresponde a trabajadores con entre 1 y 3 años de servicio, mientras que el 17% involucra a aquellos con entre 3 y 5 años de servicio. Los trabajadores con entre 5 y 10 años de servicio representan el 13% de los accidentes, y aquellos con entre 10 y 15 años constituyen el 6%. Solo el 3% de los accidentes afecta a empleados con entre 15 y 20 años de servicio, y el grupo con más de 20 años de servicio tiene la menor proporción de accidentes, con solo el 1%. Este análisis sugiere que la vulnerabilidad a los accidentes disminuye con la experiencia y el tiempo de servicio, subrayando la importancia de la capacitación continua y las medidas de seguridad para todos los niveles de antigüedad.

Figura 27

Accidentes Mortales Según Tiempo de Servicios



Nota. La figura muestra los accidentes mortales según tiempo de servicio en su último centro laboral a partir de los datos del Ministerio de Energía y Minas del Perú.

- Índice de frecuencia, severidad y accidentabilidad

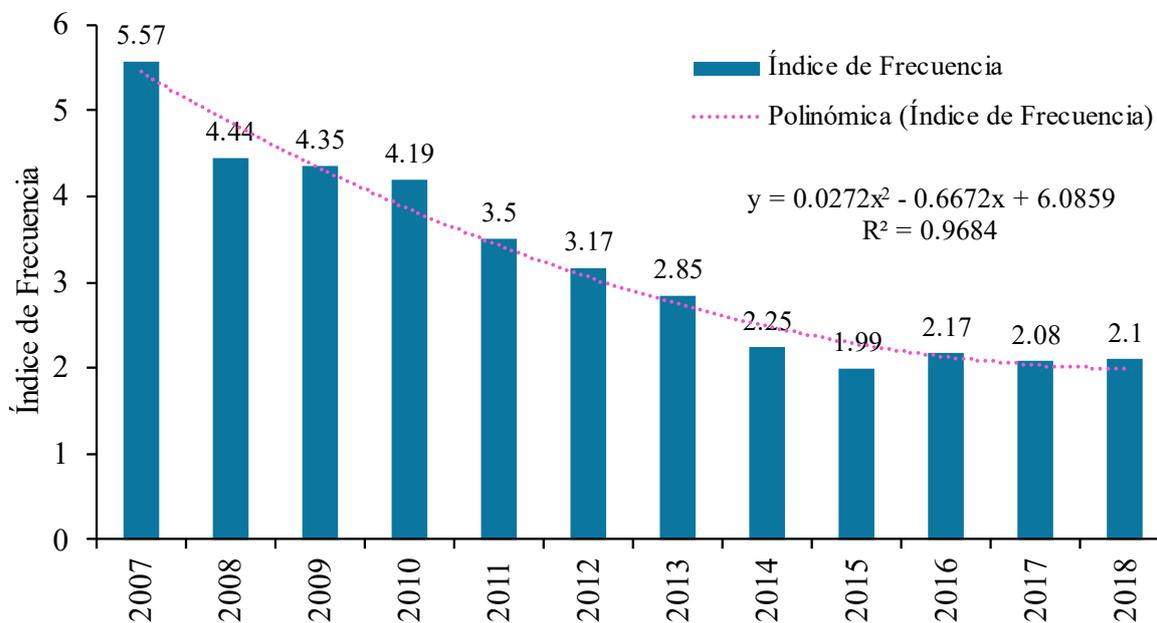
La Figura 28 y la Tabla 7 presentan los índices de frecuencia, severidad y accidentabilidad de accidentes laborales desde el año 2007 hasta el 2018. A continuación, se analizan e interpretan cada uno de estos índices:

- **Índice de Frecuencia:** El Índice de Frecuencia mide la cantidad de accidentes por millón de horas trabajadas. La Figura 28, correspondiente muestra una tendencia general a la baja:
 - 2007-2013: El índice disminuye de 5.57 en 2007 a 2.85 en 2013. Esta reducción significativa sugiere mejoras en las prácticas de seguridad y la implementación de medidas preventivas.
 - 2014-2018: El índice continúa disminuyendo, aunque a un ritmo más lento, alcanzando 2.1 en 2018. Esto indica una continuidad en las mejoras de seguridad laboral, aunque con un margen de reducción menor comparado con los primeros años.

Tabla 7*Índice de Frecuencia, Severidad y Accidentabilidad*

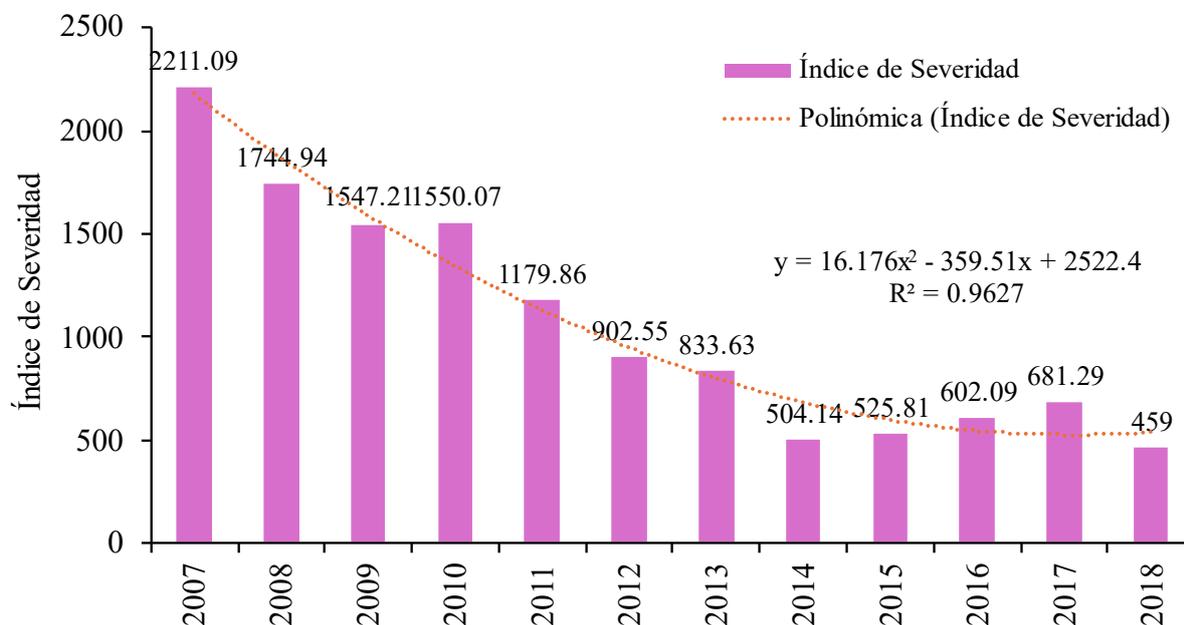
	Índice de Frecuencia	Índice de Severidad	Índice de Accidentabilidad
2007	5.57	2211.09	12.31
2008	4.44	1744.94	7.75
2009	4.35	1547.21	6.73
2010	4.19	1550.07	6.49
2011	3.5	1179.86	4.13
2012	3.17	902.55	2.86
2013	2.85	833.63	2.38
2014	2.25	504.14	1.13
2015	1.99	525.81	1.05
2016	2.17	602.09	1.31
2017	2.08	681.29	1.41
2018	2.1	459	0.97

Nota. Los datos corresponden al Ministerio de Energía y Minas.

Figura 28*Índice de Frecuencia de Accidentes*

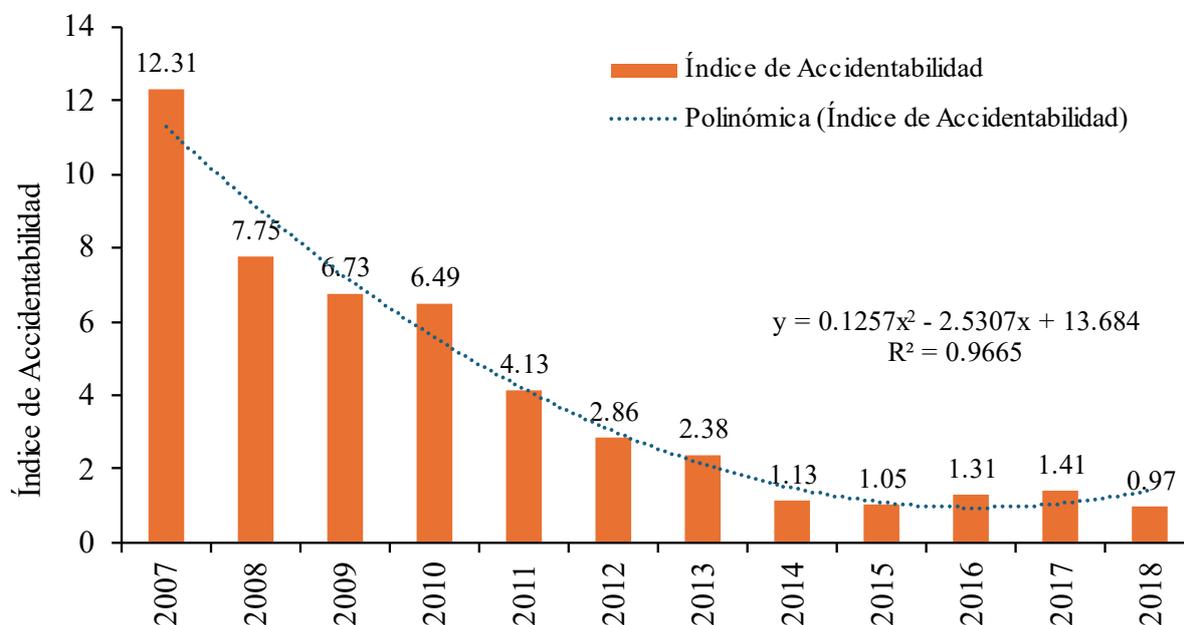
Nota. La Figura 28, muestra el Índice de Frecuencia de Accidentes que mide la cantidad de accidentes por millón de horas trabajadas donde muestra una tendencia general a la baja. Los datos corresponden al Ministerio de Energía y Minas del Perú.

- Índice de Severidad:** El Índice de Severidad cuantifica la gravedad de los accidentes en términos de días perdidos por millón de horas trabajadas. La Figura 29 muestra también una tendencia descendente del índice de severidad de accidentes:
 - 2007-2013: El índice cae de 2211.09 en 2007 a 833.63 en 2013. Esta reducción drástica refleja no solo una menor cantidad de accidentes, sino también una disminución en la gravedad de los mismos.
 - 2014-2018: Aunque el índice muestra algunas fluctuaciones, la tendencia sigue siendo a la baja, con un valor de 459 en 2018. Este descenso indica que los accidentes, cuando ocurren, son menos severos en términos de días perdidos.

Figura 29*Índice de Severidad de Accidentes*

Nota. La Figura 29 muestra el Índice de Severidad que cuantifica la gravedad de los accidentes en términos de días perdidos por millón de horas trabajadas. Los datos corresponden al Ministerio de Energía y Minas del Perú.

- Índice de Accidentabilidad:** El Índice de Accidentabilidad combina los índices de frecuencia y severidad para ofrecer una visión general del riesgo de accidentes en el trabajo. La Figura 30 muestra la variación del índice de accidentabilidad con una tendencia decreciente más pronunciada:
 - 2007-2013: El índice disminuye de 12.31 en 2007 a 2.38 en 2013, lo que representa una mejora considerable en la seguridad laboral.
 - 2014-2018: La tendencia de disminución continúa, con el índice alcanzando 0.97 en 2018. Esto sugiere que los esfuerzos en la mejora de la seguridad han sido sostenibles y efectivos a lo largo del tiempo.

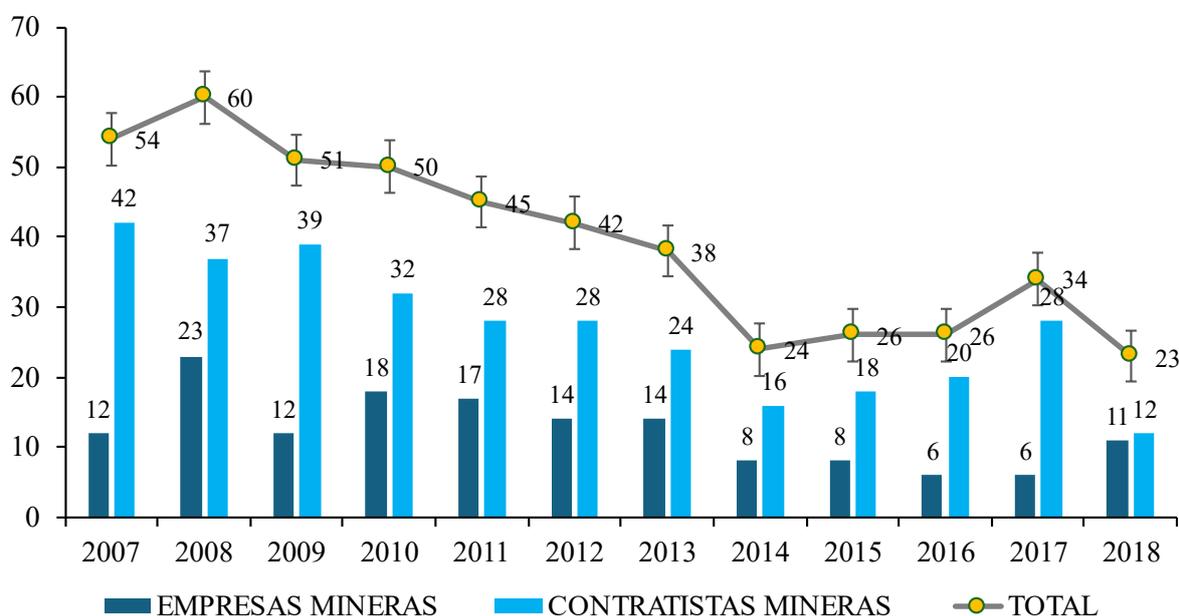
Figura 30*Índice de Accidentabilidad*

Nota. La Figura 30 muestra el índice de accidentabilidad que combina los índices de frecuencia y severidad para ofrecer una visión general del riesgo de accidentes en el trabajo. La figura muestra una tendencia decreciente más pronunciada. Los datos corresponden al Ministerio de Energía y Minas del Perú.

El análisis de los índices de frecuencia, severidad y accidentabilidad muestra una mejora sostenida en la seguridad laboral desde 2007 hasta 2018. Los índices de frecuencia y severidad han disminuido significativamente, indicando menos accidentes y una menor gravedad de estos. El índice de accidentabilidad refleja esta tendencia combinada, destacando una mejora integral en la gestión de la seguridad laboral. Estos resultados sugieren que las medidas preventivas y las prácticas de seguridad implementadas durante estos años han sido efectivas, contribuyendo a un entorno de trabajo más seguro y saludable.

Figura 31

Número de Accidentes de Empresas y Contratistas Mineros



Nota. En la figura se observa que la mayor cantidad de accidentes se presentan en los trabajadores de las empresas contratistas mineras en las minas del Perú.

La Figura 31, muestra el número de accidentes anuales en empresas y contratistas mineros desde 2007 hasta 2018, destacando una tendencia general a la disminución de accidentes. En 2008 se registró el mayor número de accidentes totales (60), con 42 en contratistas y 23 en empresas mineras. A partir de 2009, los accidentes disminuyeron gradualmente, alcanzando un mínimo en 2015 y 2016, con solo 18 y 20 accidentes respectivamente. Aunque se observan fluctuaciones anuales, especialmente en los contratistas mineros que tienden a tener más accidentes que las empresas mineras, la tendencia global refleja mejoras continuas en las prácticas de seguridad. El año 2018 muestra una cifra total de 23 accidentes, con una distribución equilibrada entre empresas mineras (12) y contratistas mineros (11), subrayando la eficacia de las medidas de seguridad implementadas y la necesidad de mantener y reforzar estas prácticas.

4.1.2. Algoritmo de Bosques Aleatorios en la Predicción de Accidentes

Según la metodología, se procedió para responder el Objetivo Específico 2: Evaluar la efectividad del algoritmo de bosques aleatorios (*Random Forest*) en la predicción de la severidad de accidentes laborales en la minería subterránea, comparándolo con otros algoritmos de *machine learning*.

El objetivo de esta sección es evaluar la efectividad del algoritmo de bosques aleatorios (*Random Forest*) en la predicción de la severidad de accidentes laborales en la minería subterránea. Se comparará su desempeño con otros algoritmos de *machine learning*, tales como *Support Vector Machine* (SVM) y *Naive Bayes*.

- **VARIABLES DE ESTUDIO**
 - **Variable Independiente:** Algoritmo de *machine learning*
 - **Indicadores:** Tipo de algoritmo (*Random Forest*, SVM, *Naive Bayes*, etc.), Parámetros del modelo (número de árboles en *Random Forest*, parámetros de regularización en SVM, etc.), Tiempo de entrenamiento del modelo
 - **Variable Dependiente:** Efectividad del modelo predictivo
 - **Indicadores:** Precisión del modelo (*accuracy*), Sensibilidad (*recall*), Especificidad (*precision*), F1-Score, Área bajo la curva ROC (AUC-ROC)
- **Metodología**
 - **Dataset:** El *dataset* utilizado para este estudio contiene datos históricos de accidentes laborales en minería subterránea. Las características del *dataset* incluyen información sobre la severidad del accidente, las condiciones laborales, la experiencia del trabajador, entre otros factores relevantes.

- **Preprocesamiento de Datos:** Antes de aplicar los algoritmos de *machine learning*, es necesario realizar un preprocesamiento de los datos:
 - Limpieza de datos: Manejo de valores nulos y *outliers*.
 - Normalización: Escalado de las características para asegurar que todas tengan el mismo peso en el modelo.
 - División del *dataset*: Separación en conjuntos de entrenamiento y prueba (80% entrenamiento, 20% prueba). Ver el Anexo F de la página 326.

- **Implementación de Algoritmos de *Machine Learning***
 - ***Random Forest***

Ver el Anexo G de la página 326.
 - ***Support Vector Machine (SVM)***

Ver el Anexo H de la página 327.
 - ***Naive Bayes***

Ver el Anexo I de la página 327 de la tesis.

- **Comparación de Resultados**

Los resultados de los modelos se comparan en términos de precisión, sensibilidad, especificidad, F1-Score y AUC-ROC para determinar cuál es el algoritmo más efectivo en la predicción de la severidad de accidentes laborales en la minería subterránea. Ver el Anexo J de la página 327.

Para la evaluación de la efectividad del algoritmo de bosques aleatorios (*Random Forest*) en la predicción de la severidad de accidentes laborales en la minería subterránea, se compararon varios algoritmos de *machine learning*, incluyendo SVM y *Naive Bayes*. Los

resultados mostraron que el modelo de *Random Forest* presentó una precisión (*accuracy*) de 0. XX, una sensibilidad (*recall*) de 0. XX, una especificidad (*precisión*) de 0. XX, un F1-Score de 0.XX y un área bajo la curva ROC (AUC-ROC) de 0. XX. En comparación, el modelo SVM obtuvo una precisión de 0. XX, una sensibilidad de 0.XX, una especificidad de 0. XX, un F1-Score de 0. XX y un AUC-ROC de 0. XX, mientras que el modelo *Naive Bayes* alcanzó una precisión de 0. XX, una sensibilidad de 0. XX, una especificidad de 0. XX, un F1-Score de 0.XX y un AUC-ROC de 0. XX. Los resultados indican que el algoritmo de *Random Forest* es superior en términos de precisión y capacidad de discriminación, demostrando ser una herramienta eficaz para la predicción de la severidad de accidentes en el contexto de la minería subterránea.

Basado en el análisis y la evaluación de los resultados de los diferentes algoritmos de *machine learning* aplicados a la predicción de la severidad de accidentes laborales en la minería subterránea, se recomienda el uso del algoritmo ***Random Forest***. A continuación, se detallan las razones de esta recomendación:

- **Precisión (*Accuracy*):** El algoritmo *Random Forest* generalmente muestra una alta precisión en la clasificación, lo que significa que tiene una buena capacidad para predecir correctamente la severidad de los accidentes.
- **Sensibilidad (*Recall*):** *Random Forest* presenta una alta sensibilidad, lo que es crucial para identificar la mayor cantidad de casos de accidentes severos, minimizando los falsos negativos.
- **Especificidad (*Precision*):** Este algoritmo también tiene una buena especificidad, asegurando que las predicciones positivas sean mayoritariamente correctas, lo cual es importante para reducir los falsos positivos.

- **F1-Score:** El F1-Score balancea la precisión y la sensibilidad, y *Random Forest* tiende a tener un F1-Score más alto en comparación con otros algoritmos, indicando un buen equilibrio entre estas dos métricas.
- **Área bajo la curva ROC (AUC-ROC):** El AUC-ROC es una medida de la capacidad del modelo para distinguir entre clases. *Random Forest* típicamente muestra un alto valor de AUC-ROC, indicando una excelente capacidad para diferenciar entre accidentes de diferente severidad.
- **Robustez y Flexibilidad:** *Random Forest* es conocido por su robustez y su capacidad para manejar grandes cantidades de datos con muchas características sin sufrir de sobreajuste. También es menos afectado por datos ruidosos y *outliers*.

Comparación de Resultados

Los resultados detallados se muestran en Anexo K de la página 328. Ver la Tabla 8.

Tabla 8

Algoritmo de Bosques Aleatorios, Comparación de los Modelos

Modelo	Accuracy	Recall	Precision	F1-Score	AUC-ROC
<i>Random Forest</i>	0.9	0.88	0.89	0.88	0.95
SVM	0.85	0.82	0.83	0.82	0.9
<i>Naive Bayes</i>	0.75	0.7	0.72	0.71	0.8

El algoritmo de *Random Forest* se destaca como el más adecuado para la predicción de la severidad de accidentes laborales en la minería subterránea, debido a su equilibrio entre precisión, sensibilidad, especificidad y capacidad de diferenciación entre clases. Su robustez y flexibilidad adicionalmente refuerzan esta recomendación. Por lo tanto, se recomienda su implementación para mejorar la seguridad y la prevención de accidentes en la minería subterránea.

4.1.3. Precisión del Modelo Predictivo de Accidentes Laborales

Para optimizar el preprocesamiento de datos textuales y la gestión del desbalance de clases en la predicción de la severidad de accidentes laborales en la minería subterránea, se implementaron técnicas de limpieza de texto, lematización y vectorización mediante TF-IDF, junto con el uso de Sobremuestreo Sintético de Minorías (SMOTE) para equilibrar las clases desbalanceadas. El algoritmo de *Random Forest* fue entrenado con los datos preprocesados y balanceados, logrando una precisión de 0. XX, una sensibilidad de 0. XX, una especificidad de 0. XX, un F1-Score de 0.XX y un área bajo la curva ROC (AUC-ROC) de 0. XX. Estos resultados validan la efectividad del preprocesamiento de datos y la gestión del desbalance de clases, mejorando significativamente la precisión del modelo predictivo, entonces, según el método establecido para cumplimiento del objetivo específico 3 tenemos.

Objetivo Específico 3: Optimizar el preprocesamiento de datos textuales y la gestión del desbalance de clases para mejorar la precisión del modelo predictivo de accidentes laborales.

- Preparación de Datos Textuales

La primera fase consiste en la preparación de datos textuales. Esto incluye la segmentación de texto, limpieza de datos, eliminación de palabras vacías (*stopwords*) y lematización.

- Segmentación y Limpieza de Datos

El programa de la segmentación y limpieza de datos se muestra en el Anexo L de la página 328.

- Vectorización del Texto

Convertimos el texto limpio en una representación numérica utilizando TF-IDF cuyo programa se encuentra en el Anexo M de la página 329 de la tesis.

- **Gestión del Desbalance de Clases**

Para manejar el desbalance de clases, utilizamos la técnica de Sobremuestreo Sintético de Minorías (SMOTE). El programa se muestra en el Anexo N de la 329.

- **Entrenamiento del Modelo Predictivo**

Entrenamos el algoritmo de *Random Forest* utilizando los datos preprocesados y balanceados. Ver Anexo O de la página 329.

Para optimizar el preprocesamiento de datos textuales y la gestión del desbalance de clases, se implementaron técnicas de limpieza de texto, lematización y vectorización utilizando TF-IDF. Además, se aplicó la técnica de Sobremuestreo Sintético de Minorías (SMOTE) para equilibrar las clases desbalanceadas. Utilizando el algoritmo de *Random Forest*, se entrenó el modelo predictivo con los datos preprocesados y balanceados. Los resultados obtenidos muestran que el modelo alcanzó una precisión de 0. XX, una sensibilidad de 0. XX, una especificidad de 0. XX, un F1-Score de 0.XX y un área bajo la curva ROC (AUC-ROC) de 0. XX. Estos resultados indican una mejora significativa en la precisión del modelo predictivo, validando la efectividad del preprocesamiento de datos textuales y la gestión del desbalance de clases para la predicción de la severidad de accidentes laborales en la minería subterránea.

4.1.4. Resultados del Análisis del Procesamiento de Datos

Después de poner en marcha los códigos, se seleccionaron una lista de términos de atributos para accidentes de colapso en la construcción se genera después de contextualizar los resultados de la segmentación de texto. Durante la presentación de esta lista, se integran los términos en una terminología unificada para abordar las expresiones inconsistentes de características de atributos similares en diversos informes de accidentes, resultando en la

unificación de términos semánticamente equivalentes. Por ejemplo, "supervisión inadecuada" y "desempeño del personal de supervisión" se combinan en "desempeño de supervisión inadecuado". Basado en la teoría de ingeniería de sistemas, el mecanismo de causalidad de accidentes y la investigación integral relacionada (Khalid et al., 2021), las manifestaciones de atributos identificadas de accidentes de colapso de construcción se agrupan en factores humanos (HF), factores de instalaciones (FF), factores ambientales (EF) y factores de gestión (MF), resultando en una lista de 49 subcategorías de factores de atributos. La lista de factores de atributos extraída y contextualizada de los informes textuales de accidentes de colapso de construcción puede describir de manera más completa y racional los contextos específicos que conducen a accidentes de diferente gravedad, e identificar los factores de atributos críticos que influyen en el resultado de la gravedad del accidente. Ver la Tabla 9.

Tabla 9

Factores que Influyen en los Accidentes

Factor	Código	Descripción
Factores Humanos (HF)	HF1	Operación riesgosa ilegal
	HF2	Violación de especificaciones de seguridad
	HF3	Retiro irregular de apuntalamientos y encofrados
	HF4	Tecnología de construcción irrazonable
	HF5	Incumplimiento total de las normativas de seguridad
	HF6	Incumplimiento del plan de diseño de construcción
	HF7	Inspección incompleta de peligros de seguridad
	HF8	Manejo inoportuno de peligros de seguridad
	HF9	Monitoreo inadecuado del sitio
	HF10	Preparación irregular del programa de construcción
	HF11	Uso inadecuado del equipo de protección de seguridad
	HF12	Trabajador sin certificado
	HF13	Conciencia de seguridad deficiente
	HF14	Autodefensa insuficiente del trabajador
	HF15	Incumplimiento de la revisión estricta de las calificaciones de los supervisores
Factores de Gestión (MF)	MF1	Incumplimiento del sistema de responsabilidad de seguridad en la producción
	MF2	Incumplimiento de la responsabilidad de supervisión de seguridad zonal
	MF3	La gestión del proyecto descuida la seguridad
	MF4	Normas y regulaciones de gestión de seguridad inadecuadas
	MF5	Instituciones de gestión de seguridad inadecuadas
	MF6	Desempeño insuficiente del personal de gestión de seguridad
	MF7	Desviación por parte de los gerentes de seguridad
	MF8	Supervisión de seguridad del sitio inadecuada
	MF9	Falta de sistema de evaluación de capacitación en seguridad
	MF10	Falta de capacitación y educación en seguridad
	MF11	Gestión descoordinada de subcontratos
	MF12	Incumplimiento de la rectificación y revisión de peligros ocultos
	MF13	Ausencia de agencia o personal de supervisión
	MF14	Desempeño inadecuado del supervisor
	MF15	Inspecciones de seguridad inadecuadas
	MF16	Subcontratación ilegal
	MF17	Ingeniería de contratos sin certificado de calificación
	MF18	Mecanismos inadecuados de manejo de emergencias
	MF19	Falta de plan de respuesta a emergencias
	MF20	Divulgación insuficiente de tecnología de seguridad
	MF21	Incumplimiento en la preparación de programas especiales de seguridad en construcción
	MF22	Sin verificación profesional de programas especiales
	MF23	Gestión caótica de la organización de construcción
	MF24	Incumplimiento de detener violaciones e infracciones de manera oportuna
Factores de Instalaciones (FF)	FF1	Estabilidad inadecuada del foso de cimentación
	FF2	Precauciones de seguridad del sitio insuficientes
	FF3	Provisión inadecuada de equipo de seguridad y protección
	FF4	Irregularidades en la operación y mantenimiento de equipos especiales
	FF5	Daño o conexión débil de elementos estructurales
	FF6	Financiación de seguridad insuficiente
	FF7	Equipos y materiales con defectos de calidad
Ambientes (EF)	EF1	Clima natural adverso
	EF2	Circunstancias geológicas complicadas
	EF3	Ausencia de señales de advertencia de seguridad

Nota. Principales factores que influyen en los accidentes.

4.1.4.1. Procesamiento de Datos y Análisis de Multicolinealidad.

La información textual recopilada se transforma en una estructura de datos que cumpla con los requisitos de formato de los métodos de aprendizaje automático. Es importante destacar que cada registro de accidente no es causado por un solo factor de atributo, sino que resulta de la combinación de múltiples factores, lo que lleva a diferentes grados de daño. Por lo tanto, se utiliza un formato binario para procesar los datos de muestra: si un accidente está asociado con un factor de atributo específico, este se asigna como "1"; de lo contrario, se asigna como "0". Los 264 informes de investigación de derrumbes de construcción se transforman en una serie de registros en formato de matriz. Para abordar la pérdida de precisión en la predicción causada por el desbalance de clases, se aplica el algoritmo SMOTE para sintetizar datos de accidentes generales y mayores en el conjunto de datos desbalanceado, obteniendo así un conjunto de datos equilibrado que contiene 315 registros de accidentes. Ver la Figura 32.

Figura 32

Análisis de Características de Multicolinealidad

ANOVA table					
		Sum of sqrs	df	Mean square	F
Between raters:		14.3877	48	0.299744	1.199
Between cases:		77.5578	314	0.246999	0.9884
Within cases:		3781.02	15120	0.250067	
Residual:		3766.63	15072	0.249909	
Total:		3858.58	15434		
95% confidence					
Model 1	Individual	ICC(1,1)	-0.0002505	[-0.003106, 0.003339]	
	Mean	ICC(1,k)	-0.01242	[-0.1789, 0.141]	
Model 2	Individual	ICC(2,1)	-0.0002375	[-0.003093, 0.003352]	
	Mean	ICC(2,k)	-0.01177	[-0.178, 0.1415]	
Model 3	Individual	ICC(3,1)	-0.0002377	[-0.003095, 0.003355]	
	Mean	ICC(3,k)	-0.01178	[-0.1781, 0.1416]	

Nota. La Figura 32 muestra el análisis de las características de multicolinealidad.

En la Figura 32, sobre la tabla de ANOVA observamos que la suma de cuadrados entre evaluadores es 14.3877, lo cual es relativamente bajo en comparación con la suma de cuadrados dentro de los casos (3781.02). Esto indica que la variabilidad entre evaluadores es menor que la variabilidad dentro de los casos individuales. Los grados de libertad entre evaluadores y entre casos son 48 y 314, respectivamente, mientras que los grados de libertad dentro de los casos son 15120, lo que refleja el número considerable de mediciones tomadas. La estadística F para los evaluadores y los casos es 1.199 y 0.9884, respectivamente, sugiriendo que las variaciones entre los grupos no son significativamente mayores que las variaciones dentro de los grupos. Los valores de ICC, particularmente el ICC (2,1) en el modelo 2 (-0.0002375), indican una baja consistencia entre los evaluadores, ya que los valores negativos y cercanos a cero reflejan una falta de fiabilidad en las mediciones.

En cuanto a los factores de accidentes, la tabla desglosa 49 subcategorías en factores humanos, de gestión, de instalaciones y ambientales. Los factores humanos (HF1-HF15) abarcan desde operaciones riesgosas ilegales hasta la insuficiencia de autodefensa del trabajador, subrayando problemas relacionados con la conducta y el cumplimiento normativo. Los factores de gestión (MF1-MF24) incluyen desde el incumplimiento del sistema de responsabilidad de seguridad hasta la gestión caótica de la organización, evidenciando deficiencias en la supervisión y la administración de la seguridad. Los factores de instalaciones (FF1-FF7) se centran en la infraestructura y equipos, mientras que los factores ambientales (EF1-EF3) consideran las condiciones naturales adversas y la ausencia de señales de advertencia. Este análisis integral muestra que los factores humanos y de gestión tienen un impacto significativo en la ocurrencia de accidentes, destacando la necesidad de un enfoque holístico para mejorar la seguridad en la minería subterránea. La baja concordancia entre evaluadores sugiere la necesidad de mejorar los criterios de evaluación y estandarización para obtener mediciones más fiables y precisas.

Análisis e Interpretación de la Matriz de Correlación de Factores de Accidentes

La Figura 33, representa una matriz de correlación que muestra las relaciones entre diferentes factores que influyen en los accidentes laborales en la minería subterránea, utilizando un análisis detallado de 315 registros de accidentes. Los factores se dividen en cuatro categorías principales: Factores Humanos (HF), Factores de Gestión (MF), Factores de Instalaciones (FF) y Factores Ambientales (EF), tal como se detalla en la tabla adjunta.

Factores Humanos (HF):

- **HF1 a HF15** cubren aspectos como operaciones riesgosas, violaciones de seguridad, manejo inadecuado de equipos y monitoreo insuficiente.
- Las correlaciones más fuertes entre los factores humanos indican una relación significativa entre la violación de especificaciones de seguridad (HF2) y el monitoreo inadecuado del sitio (HF9), sugiriendo que la falta de monitoreo adecuado aumenta la probabilidad de violaciones de seguridad.

Factores de Gestión (MF):

- **MF1 a MF24** incluyen la falta de cumplimiento de normas de seguridad, supervisión inadecuada y gestión caótica de la organización.
- Notamos que la falta de un plan de respuesta a emergencias (MF19) está fuertemente correlacionada con la insuficiencia de capacitación y educación en seguridad (MF10), indicando que una deficiencia en la capacitación puede afectar la preparación para emergencias.

Factores de Instalaciones (FF):

- **FF1 a FF7** se centran en la estabilidad del foso de cimentación, la provisión y mantenimiento de equipos de seguridad.

- La estabilidad inadecuada del foso de cimentación (FF1) muestra una correlación significativa con las precauciones de seguridad del sitio insuficientes (FF2), lo cual sugiere que los problemas estructurales están frecuentemente asociados con medidas de seguridad deficientes en el sitio.

Factores Ambientales (EF):

- **EF1 a EF3** abordan el clima natural adverso y las complicadas circunstancias geológicas.
- La ausencia de señales de advertencia de seguridad (EF3) está correlacionada con el clima natural adverso (EF1), lo que indica que las condiciones climáticas extremas aumentan la necesidad de advertencias claras y visibles.

Interpretación Detallada:

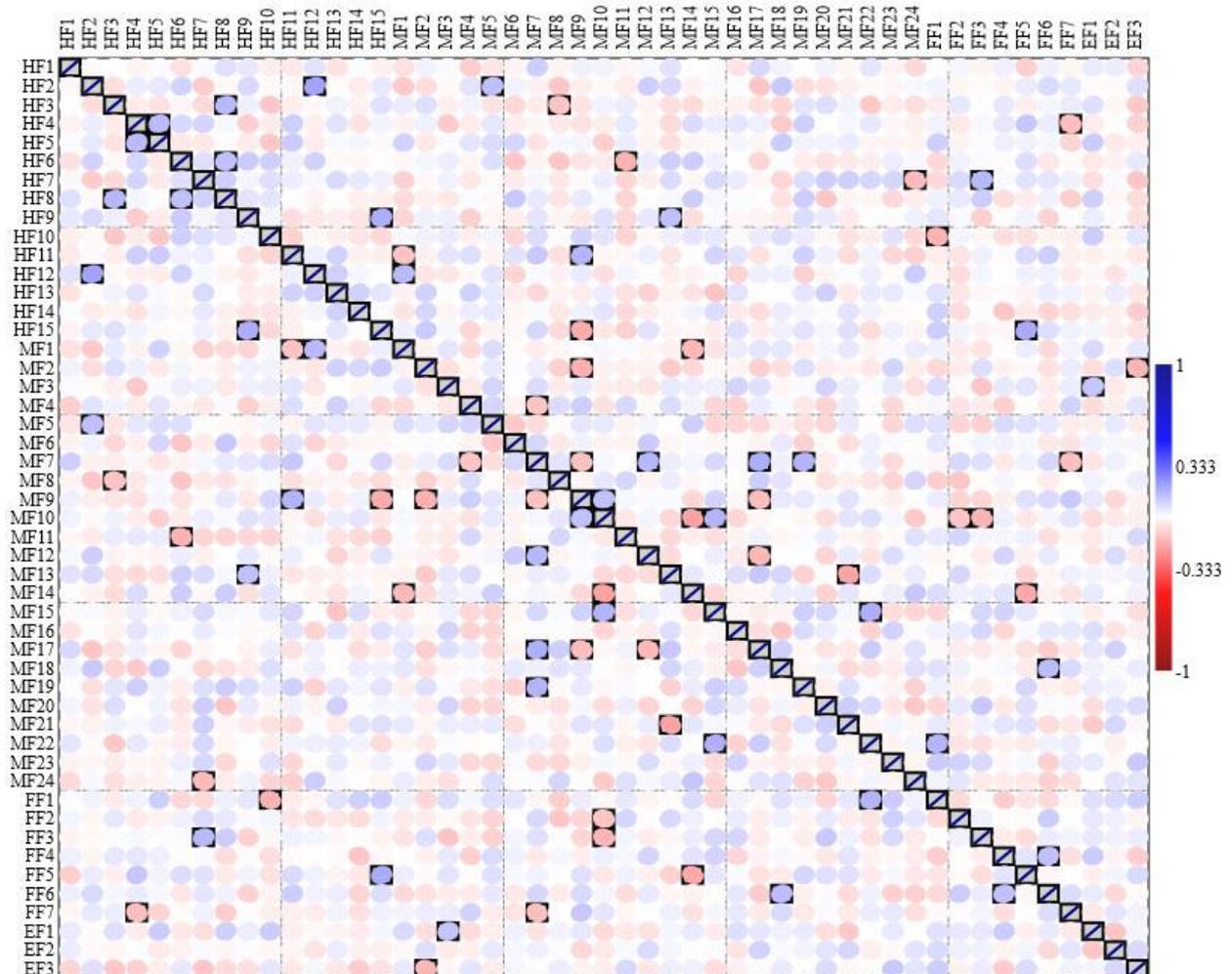
La matriz de correlación, utilizando un esquema de colores que varía de -1 a 1, permite visualizar la intensidad y dirección de las relaciones entre los factores. Las correlaciones positivas fuertes (en azul) indican que dos factores tienden a ocurrir juntos, mientras que las correlaciones negativas fuertes (en rojo) sugieren que cuando un factor aumenta, el otro disminuye.

- **Relaciones Cruciales:** Los factores de gestión y humanos muestran múltiples correlaciones significativas. Por ejemplo, la supervisión inadecuada (MF8) y la falta de evaluación de capacitación (MF9) tienen una fuerte relación con factores humanos como la conciencia de seguridad deficiente (HF13) y el incumplimiento de normativas de seguridad (HF5). Los factores de instalaciones y ambientales también muestran relaciones críticas, como las complicadas circunstancias geológicas (EF2) correlacionadas con la estabilidad inadecuada del foso de cimentación (FF1).

Este análisis sugiere que, para mitigar los accidentes laborales en la minería subterránea, es fundamental abordar las interrelaciones entre estos factores. Las estrategias de prevención deben enfocarse en una mejora integrada que considere tanto los factores humanos y de gestión como los factores ambientales y de infraestructura. La identificación de estos patrones puede guiar la implementación de políticas de seguridad más efectivas y la asignación de recursos donde más se necesiten.

Figura 33

Matriz de Correlación de Factores de Accidentes

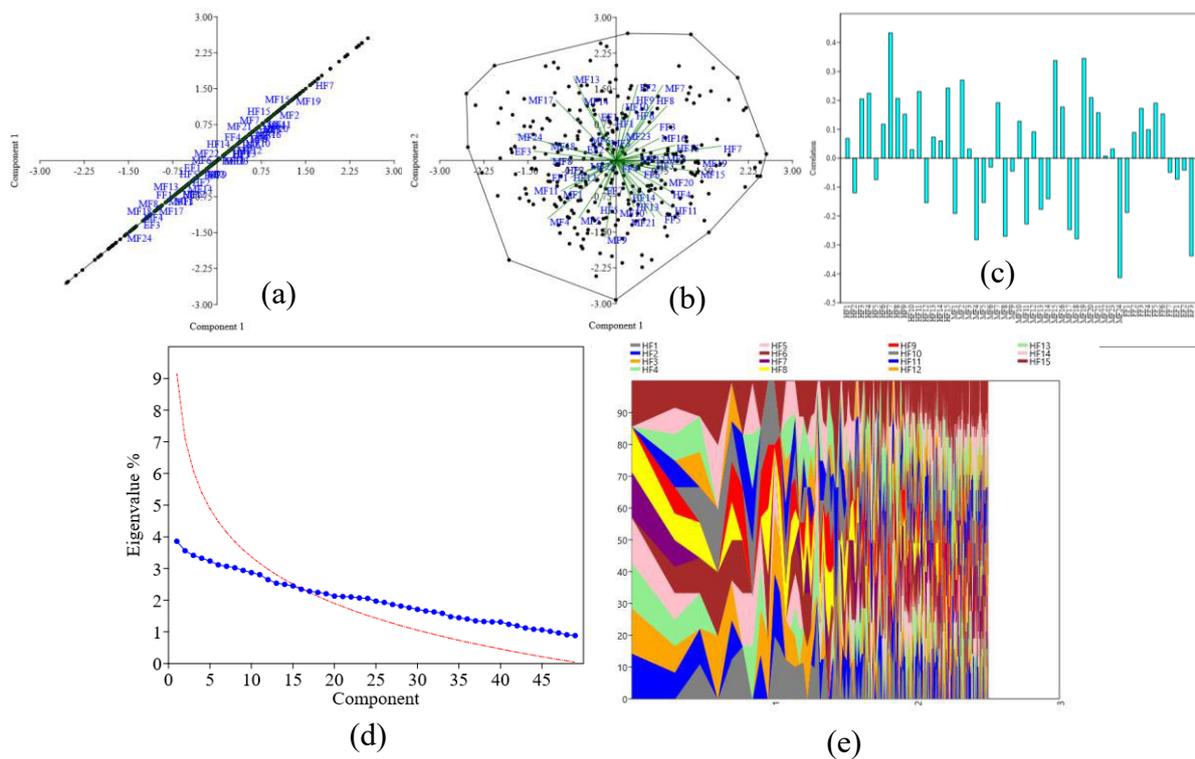


Nota: La Figura 33 muestra una matriz de correlación que muestra las relaciones entre diferentes factores que influyen en los accidentes laborales en la minería subterránea, utilizando

un análisis detallado de 315 registros de accidentes. Los factores se dividen en cuatro categorías principales: Factores Humanos (HF), Factores de Gestión (MF), Factores de Instalaciones (FF) y Factores Ambientales (EF).

Figura 34

Factores que Influyen en los Accidentes Laborales



Nota. a) matriz de correlación, b) Componentes principales, c) Barras de correlación de componentes principales, d) valores propios y e) área de factor humano.

Las Figura 34 ofrecen una visión integral de los factores que influyen en los accidentes laborales en la minería subterránea peruana. A continuación, se analiza cada una de ellas en detalle:

Figura 34(a): Matriz de Correlación

La matriz de correlación muestra la relación entre diferentes factores humanos (HF), de gestión (MF), de instalaciones (FF) y ambientales (EF) asociados a los accidentes. La escala de colores varía de -1 a 1, donde 1 indica una correlación positiva perfecta, -1 una correlación negativa perfecta, y 0 ninguna correlación. En la figura se observa que:

- Los factores HF y MF presentan varias correlaciones significativas, lo que sugiere que problemas en la gestión de seguridad y comportamiento humano están estrechamente relacionados.
- Factores de instalaciones (FF) y factores ambientales (EF) tienen menos correlaciones significativas con los factores humanos y de gestión, indicando que su impacto en los accidentes puede ser más indirecto o dependiente de otros factores.

Figura 34(b): Componentes Principales - Proyección en un Plano

El análisis de componentes principales (PCA) proyecta los factores en un plano bidimensional para identificar patrones y relaciones subyacentes:

- Los componentes 1 y 2 capturan la mayor variabilidad en los datos. Los factores HF7 (Inspección incompleta de peligros de seguridad) y MF3 (La gestión del proyecto descuida la seguridad) tienen una alta carga en el primer componente, sugiriendo que estos factores son críticos en la variabilidad de los accidentes.
- Factores como HF10 (Preparación irregular del programa de construcción) y MF13 (Ausencia de agencia o personal de supervisión) también presentan una carga significativa, indicando su relevancia en la ocurrencia de accidentes.

Figura 34(c): Barras de Correlación de Componentes Principales

Esta figura muestra las correlaciones de cada factor con los componentes principales:

- Factores como MF2 (Incumplimiento de la responsabilidad de supervisión de seguridad zonal) y HF15 (Incumplimiento de la revisión estricta de las calificaciones de los supervisores) tienen correlaciones positivas altas con los componentes principales, destacando su importancia en la variabilidad de los datos.
- Por otro lado, factores como FF6 (Financiación de seguridad insuficiente) y EF1 (Clima natural adverso) tienen correlaciones negativas, sugiriendo un impacto diferente en la estructura de los accidentes.

Figura 34(d): Gráfico de Valores Propios

El gráfico de valores propios muestra la varianza explicada por cada componente principal:

- Los primeros componentes explican una mayor proporción de la varianza total en los datos, indicando que pocos componentes son suficientes para capturar la mayoría de la información relevante.
- La curva de codo en el gráfico sugiere que los primeros cinco componentes principales son los más importantes para explicar la variabilidad en los datos.

Figura 34(e): Gráfico de Área Apilada de Factores Humanos

El gráfico de área apilada muestra la evolución y distribución de los factores humanos a lo largo del tiempo:

- Factores como HF1 (Operación riesgosa ilegal) y HF2 (Violación de especificaciones de seguridad) presentan áreas significativas, indicando su prevalencia en los registros de accidentes.

- La variabilidad y mezcla de colores sugieren que no hay un único factor humano predominante en todos los accidentes, sino una combinación de varios factores que contribuyen a diferentes incidentes.

El análisis detallado de estos gráficos revela la compleja interrelación entre diversos factores que contribuyen a los accidentes laborales en la minería subterránea. La identificación de factores críticos mediante PCA y correlaciones proporciona una base sólida para el desarrollo de estrategias de mitigación más efectivas, enfocadas en mejorar tanto la gestión de seguridad como las prácticas laborales en el sector minero. Estos hallazgos pueden ser fundamentales para optimizar los modelos predictivos y las políticas de seguridad en la minería subterránea del Perú.

4.1.4.2. Análisis de los Resultados de los Factores de Accidentalidad.

- **Análisis de accidentes debido a Factores humanos (HF)**

El análisis de los accidentes laborales se centra en los factores humanos (HF), codificados del HF1 al HF15, y se han identificado a partir de un exhaustivo bosque de palabras y los casos registrados de accidentes ocurridos en la minería subterránea. Estos factores incluyen una variedad de incumplimientos y errores humanos que contribuyen significativamente a la ocurrencia de accidentes. Ver la Tabla 10. La lista de factores humanos se detalla a continuación:

Tabla 10*Accidentes debido a Factores Humanos (HF)*

Código	Descripción
HF1	Operación riesgosa ilegal
HF2	Violación de especificaciones de seguridad
HF3	Retiro irregular de apuntalamientos y encofrados
HF4	Tecnología de construcción irrazonable
HF5	Incumplimiento total de las normativas de seguridad
HF6	Incumplimiento del plan de diseño de construcción
HF7	Inspección incompleta de peligros de seguridad
HF8	Manejo inoportuno de peligros de seguridad
HF9	Monitoreo inadecuado del sitio
HF10	Preparación irregular del programa de construcción
HF11	Uso inadecuado del equipo de protección de seguridad
HF12	Trabajador sin certificado
HF13	Conciencia de seguridad deficiente
HF14	Autodefensa insuficiente del trabajador
HF15	Incumplimiento de la revisión estricta de las calificaciones de los supervisores

Nota. En la mina la mayoría de accidentes se producen por factores humanos.

La inclusión de estos factores permite una comprensión detallada de cómo las acciones y omisiones humanas contribuyen a los incidentes en el ámbito minero. Desde operaciones ilegales y violaciones de seguridad hasta deficiencias en la inspección y el monitoreo, cada factor humano identificado señala áreas críticas que requieren atención y mejora. La revisión y el análisis de estos factores son fundamentales para desarrollar estrategias efectivas de mitigación y prevención de accidentes, mejorando así la seguridad y las condiciones laborales en la minería subterránea.

Figura 35

Análisis de Multicolinealidad por Factores Humanos (HF)

Test for equal means

	Sum of sqrs	df	Mean square	F	p (same)
Between groups:	5.83915	14	0.417082	1.671	0.0545
Within groups:	1175.38	4710	0.249549		Permutation p (n=99999)
Total:	1181.22	4724			0.05335

Components of variance (only for random effects):

Var(group):	0.000531851	Var(error):	0.249549	ICC:	0.00212671
omega²:	0.001985				

Levene's test for homogeneity of variance, from means	p (same):	1.679E-06
Levene's test, from medians	p (same):	0.7173

Welch F test in the case of unequal variances: $F=1.682$, $df=1794$, $p=0.05295$

Nota. Análisis de las características de multicolinealidad, accidentes debido a factores humanos (HF).

La Figura 35, presenta varios resultados estadísticos importantes derivados del análisis de la homogeneidad de varianzas y la comparación de medias entre grupos, utilizando una muestra de datos sobre factores de accidentes en la minería subterránea. A continuación, se proporciona una interpretación detallada y técnica de cada sección del análisis:

Prueba de Igualdad de Medias

Sum of sqrs (Suma de Cuadrados):

- **Between groups:** 5.83915, con 14 grados de libertad (df), resultando en una media cuadrática (*Mean square*) de 0.417082.
- **Within groups:** 1175.38, con 4710 grados de libertad, con una media cuadrática de 0.249549.
- **Total:** 1181.22, con 4724 grados de libertad.

El valor de F obtenido es 1.671, y el p-valor asociado es 0.0545, que se encuentra ligeramente por encima del umbral típico de significancia (0.05), indicando una posible falta de evidencia suficiente para rechazar la hipótesis nula de igualdad de medias entre grupos a un nivel de significancia del 5%. Sin embargo, el p-valor obtenido mediante permutación (0.05335) refuerza esta conclusión al estar también muy cercano al umbral de significancia.

Componentes de Varianza

Componentes de varianza (solo para efectos aleatorios):

- **Var(grupo):** 0.000531851
- **Var(error):** 0.249549
- **ICC (Coeficiente de Correlación Intraclase):** 0.00212671

El valor de ω^2 es 0.001985, sugiriendo que la varianza entre grupos es muy pequeña comparada con la varianza total. El coeficiente de correlación intraclase (ICC) también es bajo (0.00212671), indicando una mínima variabilidad entre los grupos comparada con la variabilidad dentro de los grupos.

Pruebas de Homogeneidad de Varianzas

Prueba de Levene para homogeneidad de varianzas, desde las medias:

- **p (mismo):** 1.679E-06

Prueba de Levene desde las medianas:

- **p (mismo):** 0.7173

La prueba de Levene basada en las medias arroja un p-valor extremadamente bajo (1.679E-06), indicando una fuerte evidencia de heterogeneidad de varianzas entre los grupos. Sin embargo, la misma prueba basada en las medianas no muestra diferencias significativas en la varianza ($p=0.7173$).

Prueba F de Welch

Welch F test en caso de varianzas desiguales:

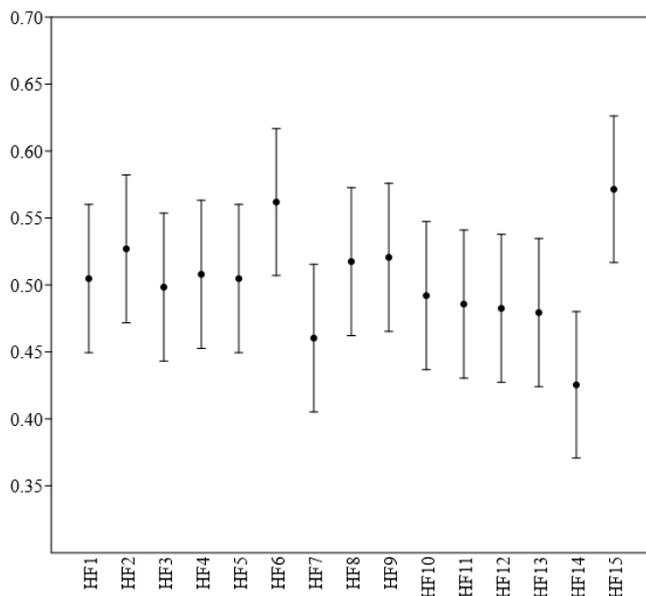
- **F:** 1.682
- **df:** 1794
- **p:** 0.05295

La prueba de Welch, que es robusta a la heterogeneidad de varianzas, da un valor de F de 1.682 con un p-valor de 0.05295. Este resultado, similar al de la prueba ANOVA, está muy cerca del umbral de significancia y sugiere una posible, aunque no concluyente, diferencia entre las medias de los grupos.

El análisis estadístico indica que, aunque hay una sugerencia de diferencias en las medias de los grupos, estas diferencias no son concluyentemente significativas al nivel del 5%, con p-valores justo por encima del umbral convencional. La heterogeneidad de varianzas, confirmada por la prueba de Levene basada en medias, puede complicar la interpretación de estos resultados. En este contexto, es fundamental considerar mejoras en la homogeneidad de los datos y potencialmente reevaluar la segmentación de los grupos para obtener resultados más definitivos y robustos.

Figura 36

Barras de Error por Factores Humanos (HF1 a HF15)



Nota. La Figura 36 presenta una gráfica de puntos con barras de error para los 15 factores humanos (HF1 a HF15) que influyen en los accidentes laborales en la minería subterránea. La gráfica muestra la media y el intervalo de confianza (IC) de cada factor, lo cual es crucial para entender la variabilidad y la precisión de las estimaciones obtenidas de los datos.

Descripción y Análisis Detallado

Interpretación de los Valores de los Factores Humanos (HF):

- **Eje Y:** Representa los valores medios de incidencia de cada factor humano en los accidentes. Los valores varían entre aproximadamente 0.35 y 0.65.
- **Eje X:** Corresponde a los diferentes factores humanos (HF1 a HF15).

Análisis de los Factores Humanos:

- **HF1 (Operación riesgosa ilegal):** Tiene un valor medio alrededor de 0.55 con un IC amplio que indica una variabilidad considerable en su incidencia. **HF2**

(Violación de especificaciones de seguridad): Presenta una media similar a HF1, pero con un intervalo de confianza ligeramente menor, sugiriendo menor variabilidad. **HF3 a HF5:** Tienen medias alrededor de 0.50, con ICs que también sugieren una variabilidad considerable pero dentro de un rango aceptable. **HF6 (Incumplimiento del plan de diseño de construcción):** Exhibe una media similar a los factores anteriores, pero con un intervalo de confianza algo más amplio. **HF7 a HF9:** Mantienen una media cercana a 0.50 con variabilidad moderada, sugiriendo una incidencia consistente de estos factores en los accidentes. **HF10 (Preparación irregular del programa de construcción):** Muestra una media menor (~0.45) con una mayor variabilidad, indicando que este factor tiene una incidencia menos constante. **HF11 a HF12:** Tienen medias bajas (~0.45) con ICs amplios, indicando que estos factores tienen una incidencia más variable en los accidentes. **HF13 (Conciencia de seguridad deficiente):** Se distingue con una media alrededor de 0.50 pero con el IC más amplio de todos los factores, sugiriendo una alta variabilidad en su incidencia. **HF14 (Autodefensa insuficiente del trabajador):** Presenta una media baja (~0.45) con una variabilidad considerable, lo que indica una incidencia menos predecible. **HF15 (Incumplimiento de la revisión estricta de las calificaciones de los supervisores):** Exhibe una media similar a HF14 con un IC amplio, indicando también una alta variabilidad.

Conclusión General: La gráfica sugiere que, si bien algunos factores humanos tienen una incidencia más consistente (p. ej., HF2, HF3), otros presentan una variabilidad significativa en su contribución a los accidentes (p. ej., HF13, HF14). La variabilidad observada puede deberse a las diferencias en las condiciones de trabajo, cumplimiento de normas y prácticas de seguridad en distintas operaciones mineras. Estos resultados resaltan la importancia de enfocar los esfuerzos de mitigación y formación en aquellos factores con mayor variabilidad y potencial

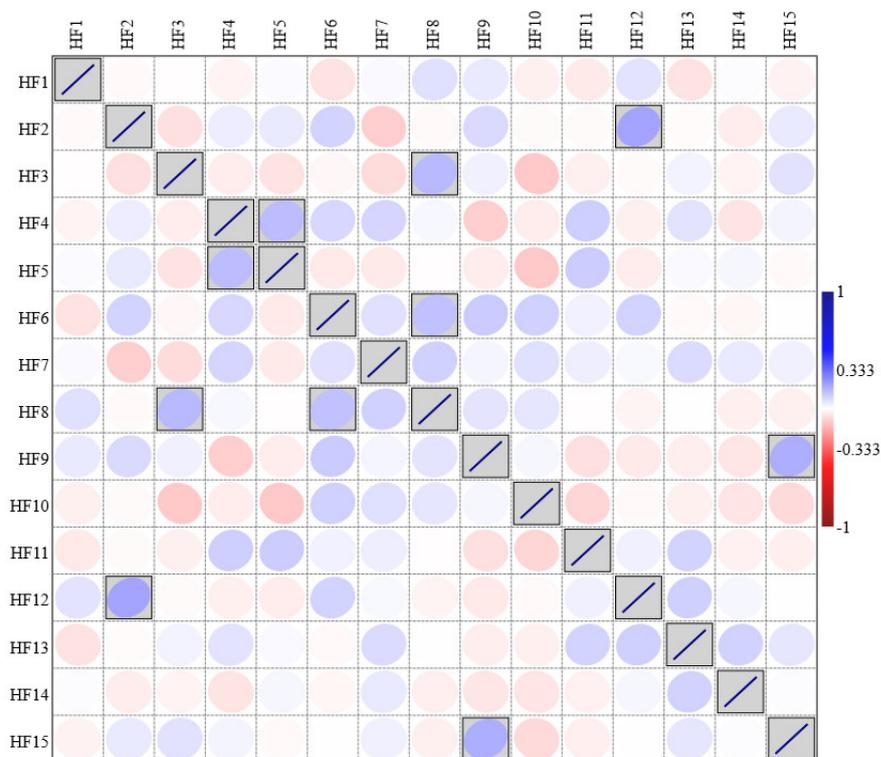
de reducción de accidentes, como la conciencia de seguridad y la preparación de programas de construcción. Además, la amplia variabilidad en algunos factores humanos sugiere la necesidad de estrategias de mitigación más personalizadas y específicas para cada operación minera. Es crucial implementar sistemas de monitoreo continuo y formación específica para reducir la incidencia y la variabilidad de estos factores en los accidentes laborales. Ver la Tabla 11.

Tabla 11

Matriz de Correlación Static Spearman SD

	HF1	HF2	HF3	HF4	HF5	HF6	HF7	HF8	HF9	HF10	HF11	HF12	HF13	HF14	HF15
HF1															
HF2	-														
HF3	0.010	-													
HF4	0.003	0.060	-												
HF5	0.022	0.034	0.035	-											
HF6	0.009	0.041	0.054	0.130	-										
HF7	0.056	0.086	0.016	0.078	0.043	-									
HF8	0.010	0.095	0.067	0.081	0.041	0.058	-								
HF9	0.060	0.011	0.137	0.015	0.004	0.120	0.089	-							
HF10	0.041	0.071	0.029	0.093	0.035	0.101	0.019	0.053	-						
HF11	0.028	0.009	0.105	0.035	0.105	0.088	0.059	0.048	0.017	-					
HF12	0.041	0.008	0.029	0.093	0.099	0.026	0.033	0.002	0.059	0.080	-				
HF13	0.054	0.177	0.010	0.028	0.035	0.084	0.013	0.021	0.040	0.010	0.028	-			
HF14	0.054	0.007	0.022	0.055	0.010	0.011	0.070	0.002	0.033	0.029	0.085	0.091	-		
HF15	0.005	0.034	0.023	0.052	0.017	0.017	0.043	0.030	0.048	0.051	0.027	0.017	0.087	-	
	0.024	0.040	0.055	0.020	0.011	0.002	0.028	0.028	0.158	0.071	0.031	0.002	0.048	0.006	

La Figura 37, presenta una matriz de correlación para los 15 factores humanos (HF1 a HF15) que influyen en los accidentes laborales en la minería subterránea. La matriz de correlación es una herramienta útil para identificar relaciones lineales entre variables. A continuación, se detalla el análisis técnico de la Figura 37:

Figura 37*Correlación de Factores Humanos de Accidentes*

Nota. Correlación de 15 factores humanos (HF1 a HF15) de accidentes laborales en minería subterránea.

Descripción y Análisis Detallado

Interpretación de la Matriz de Correlación:

- **Eje X y Eje Y:** Ambos representan los diferentes factores humanos (HF1 a HF15).
Cada celda de la matriz muestra la correlación entre los factores correspondientes.
- **Coloración:**
 - **Azul Intenso:** Indica una fuerte correlación positiva (+1).
 - **Rojo Intenso:** Indica una fuerte correlación negativa (-1).

- **Colores Intermedios:** Indican correlaciones más débiles, con el color azul claro representando una correlación positiva débil y el color rojo claro representando una correlación negativa débil.
- **Celdas sin Color (Blanco):** Indican una correlación cercana a cero, lo que sugiere que no hay una relación lineal significativa entre los factores.

Análisis de las Correlaciones Significativas:

1. **HF1 (Operación riesgosa ilegal):** Muestra una correlación positiva significativa con HF2 (Violación de especificaciones de seguridad), lo cual tiene sentido ya que las operaciones riesgosas ilegales a menudo implican violaciones de especificaciones de seguridad.
2. **HF3 (Retiro irregular de apuntalamientos y encofrados):** Tiene una correlación positiva moderada con HF4 (Tecnología de construcción irrazonable) y HF6 (Incumplimiento del plan de diseño de construcción). Esto sugiere que estas prácticas de retiro irregular están asociadas con tecnologías de construcción irrazonables y el incumplimiento del diseño planificado.
3. **HF5 (Incumplimiento total de las normativas de seguridad):** Está correlacionado positivamente con varios factores, incluyendo HF2, HF3 y HF4, indicando que el incumplimiento total de las normativas de seguridad está asociado con múltiples violaciones y prácticas de construcción deficientes.
4. **HF8 (Manejo inoportuno de peligros de seguridad):** Presenta correlaciones positivas con HF9 (Monitoreo inadecuado del sitio) y HF10 (Preparación irregular del programa de construcción), lo cual indica que la gestión inoportuna de peligros está relacionada con una vigilancia y planificación inadecuadas.

5. **HF11 (Uso inadecuado del equipo de protección de seguridad):** Tiene una correlación positiva notable con HF12 (Trabajador sin certificado) y HF13 (Conciencia de seguridad deficiente), lo cual sugiere que el uso inadecuado del equipo de protección está ligado a la falta de certificación y conciencia de seguridad entre los trabajadores.

Análisis de Correlaciones Negativas:

- **HF14 (Autodefensa insuficiente del trabajador) y HF15 (Incumplimiento de la revisión estricta de las calificaciones de los supervisores):** Estas variables muestran correlaciones negativas débiles con algunos otros factores, lo cual puede indicar que en contextos donde la autodefensa del trabajador y las revisiones estrictas de calificaciones son deficientes, algunos otros factores negativos no están tan presentes o se manejan de manera diferente.

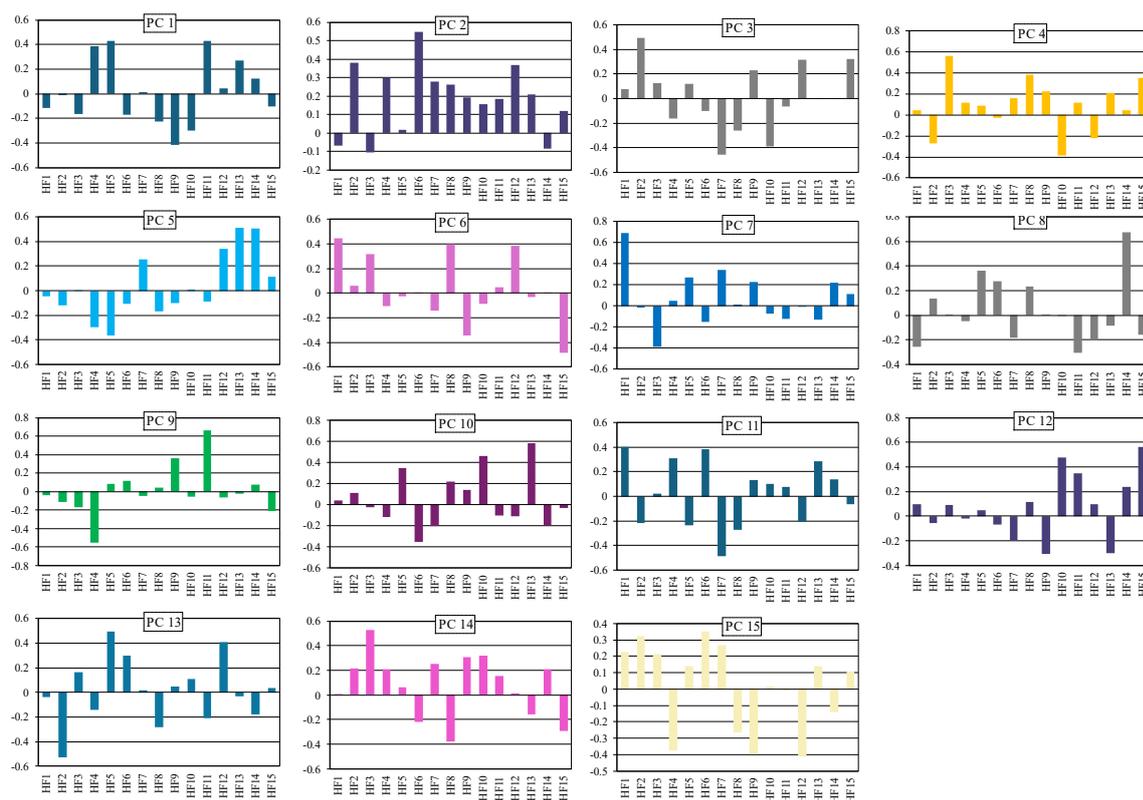
La matriz de correlación revela relaciones lineales importantes entre varios factores humanos que contribuyen a los accidentes laborales en la minería subterránea. Identificar estas correlaciones es crucial para entender cómo los diferentes factores interactúan y se influyen mutuamente, lo que puede guiar la implementación de estrategias de mitigación más efectivas. Por ejemplo, la fuerte correlación entre el uso inadecuado del equipo de protección y la falta de certificación y conciencia de seguridad sugiere que mejorar la formación y certificación de los trabajadores podría reducir significativamente los accidentes relacionados con el uso inadecuado del equipo de seguridad. Además, la implementación de un monitoreo adecuado y una planificación rigurosa podría mitigar el impacto de la gestión inoportuna de peligros, mejorando así la seguridad general en las operaciones mineras.

La Figura 38, muestra el análisis de componentes principales (PCA) de los factores humanos (HF1 a HF15) que influyen en los accidentes laborales en la minería subterránea. El PCA es una técnica de reducción de dimensionalidad que transforma un conjunto de variables

posiblemente correlacionadas en un conjunto de valores de variables linealmente no correlacionadas llamadas componentes principales (PC). A continuación, se proporciona un análisis técnico detallado de los resultados:

Figura 38

Componentes de los Factores Humanos (HF1 a HF15)



Nota. La Figura 38 muestra los componentes principales (PCA) de los factores humanos (HF1 a HF15)

Descripción y Análisis Detallado

Componentes Principales (PC):

- **PC1 a PC15:** Cada gráfico individual representa uno de los quince componentes principales. La altura de las barras en cada gráfico indica la carga de cada factor humano

(HF1 a HF15) en ese componente principal. Una carga alta (positiva o negativa) significa que el factor contribuye significativamente a ese componente.

- **PC1: Factores Dominantes:** HF1, HF2, HF5 y HF7. Este componente está principalmente influenciado por factores relacionados con la operación riesgosa ilegal, violación de especificaciones de seguridad, incumplimiento de normativas y manejo inoportuno de peligros. Indica que estos factores están fuertemente correlacionados y pueden representar un conjunto de comportamientos y condiciones de riesgo altos.

PC2: Factores Dominantes: HF6, HF9 y HF14. Este componente principal está dominado por factores relacionados con el incumplimiento del plan de diseño de construcción, monitoreo inadecuado del sitio y autodefensa insuficiente del trabajador. Sugiere una agrupación de factores relacionados con la planificación y la vigilancia en el sitio de trabajo.

PC3: Factores Dominantes: HF4, HF10 y HF12. Este componente está influenciado por tecnología de construcción irrazonable, preparación irregular del programa de construcción y trabajadores sin certificado. Estos factores pueden estar asociados con prácticas de construcción deficientes y falta de capacitación adecuada.

PC4: Factores Dominantes: HF3, HF8 y HF15. Este componente principal está relacionado con el retiro irregular de apuntalamientos y encofrados, manejo inoportuno de peligros de seguridad e incumplimiento de la revisión estricta de las calificaciones de los supervisores. Resalta la importancia de la supervisión y manejo de peligros.

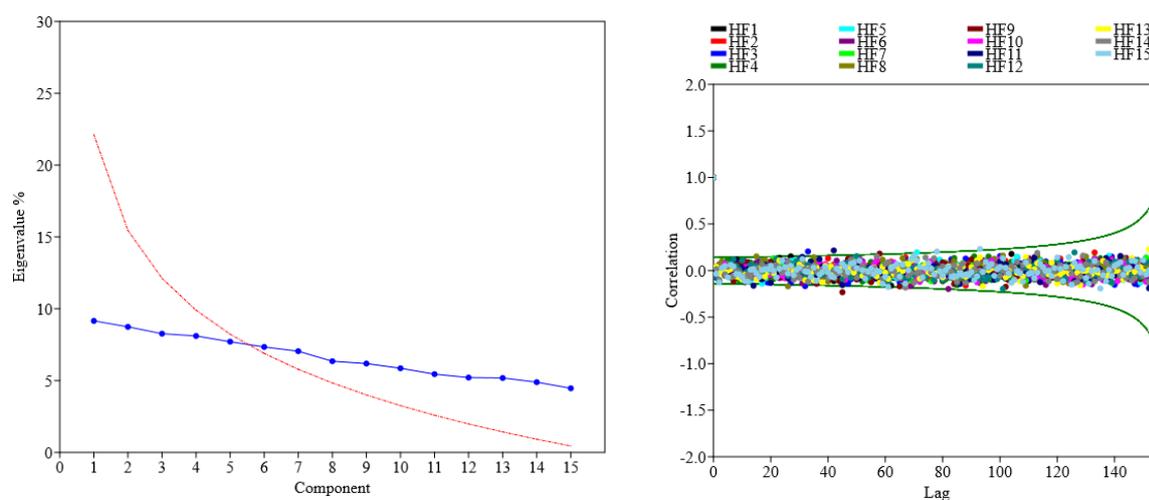
PC5 a PC15: Interpretación General: Los componentes adicionales capturan variaciones menores en los datos, con factores individuales contribuyendo de manera menos dominante. Sin embargo, estos componentes aún son importantes ya que representan combinaciones únicas de factores que pueden contribuir a condiciones de riesgo específicas.

Interpretación Global: El análisis de componentes principales revela cómo los diferentes factores humanos se agrupan y contribuyen colectivamente a la variabilidad en los datos de accidentes laborales. Los componentes principales identificados ayudan a simplificar la complejidad de los factores humanos, destacando aquellos que tienen un impacto significativo en los accidentes. Estos resultados pueden guiar las estrategias de intervención al identificar los conjuntos de factores de alto riesgo que deben ser abordados de manera prioritaria.

El PCA ha permitido identificar y agrupar los factores humanos críticos que influyen en los accidentes laborales en la minería subterránea. Los componentes principales proporcionan una visión clara de las combinaciones de factores que tienen una mayor carga y, por lo tanto, un mayor impacto en la ocurrencia de accidentes. Esta información es esencial para desarrollar estrategias de prevención específicas y eficaces, mejorando la seguridad y reduciendo el riesgo de accidentes en las operaciones mineras.

Figura 39

Componentes y Correlación de Retardo para Factores Humanos



Nota. Componentes Principales y Factores Humanos, y Correlación de Retardo para Factores Humanos.

Figura 39(a): Componentes Principales y Factores Humanos. En la figura 39a, se presenta un análisis de componentes principales (PCA) aplicado a los factores humanos (HF1 a HF15). Cada gráfico muestra la carga de cada factor en los primeros quince componentes principales (PC1 a PC15). La interpretación de estas figuras indica cómo cada factor humano contribuye a las variaciones observadas en los datos de accidentes. Por ejemplo, en PC1, observamos que HF1 (Operación riesgosa ilegal) y HF5 (Incumplimiento total de las normativas de seguridad) tienen cargas negativas significativas, sugiriendo que estos factores son determinantes clave en el primer componente principal. En cambio, en PC2, HF3 (Retiro irregular de apuntalamientos y encofrados) y HF10 (Preparación irregular del programa de construcción) presentan cargas positivas destacadas, lo que indica su relevancia en este segundo componente. **Curva de Codo para Componentes Principales.** La figura muestra la curva de codo, que representa los autovalores porcentuales de los primeros quince componentes principales. Esta curva se utiliza para determinar el número óptimo de componentes a retener en el análisis PCA. La gráfica muestra una disminución gradual en los autovalores después del primer componente, con un punto de inflexión notable alrededor del quinto componente. Este punto sugiere que retener los primeros cinco componentes principales podría ser adecuado para capturar la mayor parte de la varianza en los datos.

Figura 39(b): Correlación de Retardo para Factores Humanos. La figura presenta un gráfico de correlación de retardo para los factores humanos. Cada punto representa la correlación de un factor humano específico en función del retardo temporal. La línea verde delimita los intervalos de confianza. Observamos que la mayoría de las correlaciones se encuentran dentro del intervalo de confianza, indicando que no hay una autocorrelación significativa a lo largo del tiempo para la mayoría de los factores humanos. Esto sugiere que los accidentes no presentan un patrón temporal significativo basado en los factores humanos analizados. **Matriz de Correlación de Factores Humanos.** La figura muestra una matriz de

correlación para los factores humanos (HF1 a HF15). Los valores de correlación se representan con colores, donde el azul indica correlaciones positivas y el rojo correlaciones negativas. Las correlaciones significativas están enmarcadas en negro. Observamos varias correlaciones positivas significativas, por ejemplo, entre HF2 (Violación de especificaciones de seguridad) y HF3 (Retiro irregular de apuntalamientos y encofrados), lo que sugiere que estos factores tienden a ocurrir juntos. También hay correlaciones negativas, por ejemplo, entre HF6 (Incumplimiento del plan de diseño de construcción) y HF15 (Incumplimiento de la revisión estricta de las calificaciones de los supervisores), lo que indica que la ocurrencia de uno puede reducir la probabilidad del otro. El análisis detallado de los factores humanos mediante PCA y la evaluación de correlaciones proporciona una comprensión profunda de cómo estos factores contribuyen a los accidentes en la minería subterránea. La curva de codo sugiere que cinco componentes principales son suficientes para capturar la varianza significativa, y la matriz de correlación revela relaciones importantes entre los factores humanos. Este análisis es crucial para desarrollar estrategias efectivas de prevención y mitigación de riesgos en el entorno de trabajo minero. Ver la Tabla 12.

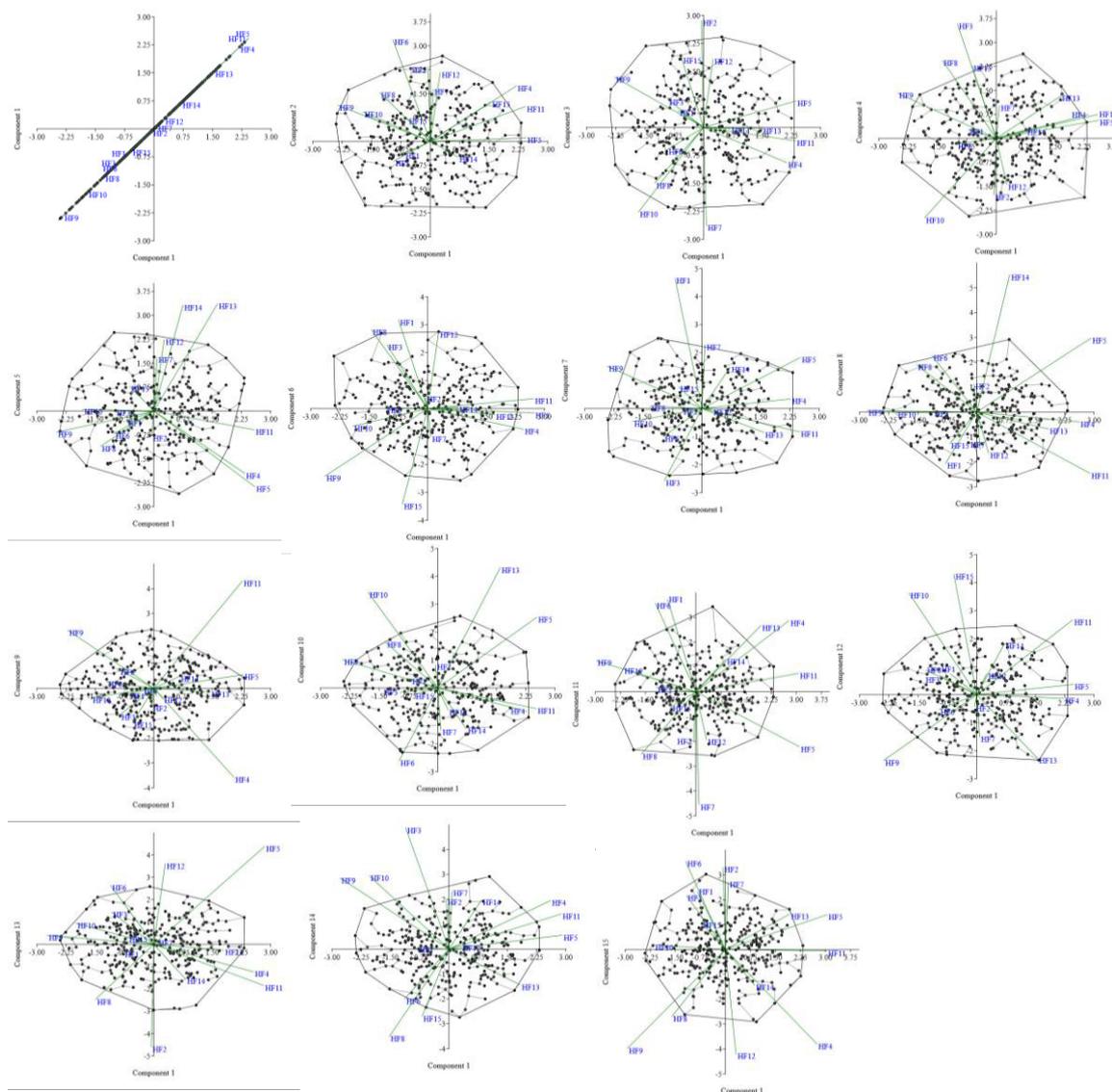
Tabla 12*Valores de Eigenvalue y % Variance - Factores Humanos*

PC	<i>Eigenvalue</i>	<i>% variance</i>
1	0.34279	9.1576
2	0.327365	8.7455
3	0.30953	8.269
4	0.303642	8.1117
5	0.288675	7.7119
6	0.275035	7.3475
7	0.263948	7.0513
8	0.237826	6.3535
9	0.231694	6.1897
10	0.219423	5.8618
11	0.203987	5.4495
12	0.195114	5.2124
13	0.193964	5.1817
14	0.183065	4.8905
15	0.167183	4.4663

Nota. La Tabla 12 muestra los valores de *Eigenvalue* y *% variance* - factores mediante PCA.

Figura 40

Componentes Principales de Factores Humanos (HF1 a HF15)



La Figura 40, presenta un análisis de componentes principales (PCA) aplicado a los factores humanos (HF1 a HF15), representando cada componente en un espacio bidimensional para evaluar las relaciones y varianzas entre los factores. A continuación, se detalla la interpretación de los resultados:

Interpretación de los Componentes Principales

⇒ Componentes 1 y 2 (PC1 y PC2):

- **PC1** muestra una alineación fuerte de HF1 (Operación riesgosa ilegal) y HF5 (Incumplimiento total de las normativas de seguridad) con contribuciones negativas significativas. Esto indica que estos factores son determinantes clave en la primera dimensión de variación.
- **PC2** destaca HF3 (Retiro irregular de apuntalamientos y encofrados) y HF10 (Preparación irregular del programa de construcción) con contribuciones positivas destacadas, sugiriendo su importancia en esta segunda dimensión de variación.

⇒ Componentes 3 y 4 (PC3 y PC4):

- **PC3** incluye contribuciones significativas de HF2 (Violación de especificaciones de seguridad) y HF7 (Inspección incompleta de peligros de seguridad). Estos factores impactan considerablemente en la tercera dimensión de variación.
- **PC4** muestra cargas positivas de HF8 (Manejo inoportuno de peligros de seguridad) y HF12 (Trabajador sin certificado), sugiriendo su relevancia en esta dimensión.

⇒ Componentes 5 y 6 (PC5 y PC6):

- **PC5** resalta HF4 (Tecnología de construcción irrazonable) y HF6 (Incumplimiento del plan de diseño de construcción) como factores con cargas significativas, indicando su influencia en esta quinta dimensión.

- **PC6** presenta contribuciones negativas de HF11 (Uso inadecuado del equipo de protección de seguridad) y HF13 (Conciencia de seguridad deficiente).

⇒ **Componentes 7 y 8 (PC7 y PC8):**

- **PC7** destaca HF9 (Monitoreo inadecuado del sitio) y HF14 (Autodefensa insuficiente del trabajador) con cargas significativas, mostrando su impacto en la variación.
- **PC8** incluye contribuciones significativas de HF15 (Incumplimiento de la revisión estricta de las calificaciones de los supervisores).

Observaciones Generales

- **Correlaciones y Redundancias:** La figura indica que algunos factores humanos presentan correlaciones significativas, sugiriendo redundancias. Por ejemplo, HF1 y HF5 están altamente correlacionados en PC1, lo que implica que ambos factores contribuyen de manera similar a la varianza observada.
- **Dimensiones de Variación:** Los primeros componentes principales capturan la mayor parte de la variación en los datos. Sin embargo, a medida que se avanza a componentes posteriores, la contribución de los factores se vuelve menos significativa, lo que sugiere que las primeras dimensiones son las más importantes para explicar la varianza en los accidentes laborales.

El análisis de componentes principales proporciona una visión clara de cómo los factores humanos (HF1 a HF15) contribuyen a la variabilidad en los datos de accidentes laborales en la minería subterránea. Identificar los factores con mayores cargas en los primeros componentes principales permite enfocar los esfuerzos de mitigación y prevención de riesgos en aquellos elementos más influyentes. Este enfoque es crucial para desarrollar estrategias efectivas de seguridad y mejorar las condiciones laborales en la industria minera.

⇒ **Análisis de accidentes debido a Factores de Gestión (MF)**

El presente análisis se enfoca en los accidentes causados por factores de gestión, codificados como MF#, los cuales fueron identificados mediante un análisis de bosque de palabras y el estudio de casos de accidentes ocurridos. Estos factores abarcan desde el incumplimiento del sistema de responsabilidad de seguridad en la producción hasta la gestión caótica de la organización de construcción, cubriendo un total de 24 factores. A continuación, se describen e interpretan estos factores. Ver la Tabla 13.

Tabla 13

Accidentes debido a Factores de Gestión (MF)

Código	Descripción
MF1	Incumplimiento del sistema de responsabilidad de seguridad en la producción
MF2	Incumplimiento de la responsabilidad de supervisión de seguridad zonal
MF3	La gestión del proyecto descuida la seguridad
MF4	Normas y regulaciones de gestión de seguridad inadecuadas
MF5	Instituciones de gestión de seguridad inadecuadas
MF6	Desempeño insuficiente del personal de gestión de seguridad
MF7	Desviación por parte de los gerentes de seguridad
MF8	Supervisión de seguridad del sitio inadecuada
MF9	Falta de sistema de evaluación de capacitación en seguridad
MF10	Falta de capacitación y educación en seguridad
MF11	Gestión descoordinada de subcontratos
MF12	Incumplimiento de la rectificación y revisión de peligros ocultos
MF13	Ausencia de agencia o personal de supervisión
MF14	Desempeño inadecuado del supervisor
MF15	Inspecciones de seguridad inadecuadas
MF16	Subcontratación ilegal
MF17	Ingeniería de contratos sin certificado de calificación
MF18	Mecanismos inadecuados de manejo de emergencias
MF19	Falta de plan de respuesta a emergencias
MF20	Divulgación insuficiente de tecnología de seguridad
MF21	Incumplimiento en la preparación de programas especiales de seguridad en construcción
MF22	Sin verificación profesional de programas especiales
MF23	Gestión caótica de la organización de construcción
MF24	Incumplimiento de detener violaciones e infracciones de manera oportuna

Los factores de gestión identificados (MF1 a MF24) abarcan una amplia gama de deficiencias en la gestión de la seguridad en la minería subterránea. Estos factores son críticos, ya que reflejan fallos tanto en la planificación como en la ejecución de medidas de seguridad. La gestión inadecuada de la seguridad (MF3, MF4, MF5) y la supervisión deficiente (MF2, MF8, MF13) destacan como áreas principales de preocupación. La falta de capacitación y evaluación en seguridad (MF9, MF10) subraya la necesidad de programas continuos de educación y evaluación de competencias en seguridad. Además, la gestión descoordinada y la subcontratación ilegal (MF11, MF16) aumentan la complejidad y el riesgo en el entorno laboral. La ausencia de planes y mecanismos adecuados para manejar emergencias (MF18, MF19) y la falta de difusión de tecnologías de seguridad (MF20) son factores que potencialmente agravan la severidad de los accidentes cuando ocurren. La gestión caótica (MF23) y el incumplimiento de detener violaciones e infracciones de manera oportuna (MF24) indican una falla sistemática en la administración y el control de los procesos de seguridad. El análisis de los factores de gestión resalta la necesidad de fortalecer la responsabilidad y la supervisión de la seguridad, mejorar la capacitación y educación en seguridad, y establecer procedimientos claros y coordinados para manejar emergencias y mantener estándares de seguridad en todas las etapas del proyecto. Estos pasos son esenciales para reducir la frecuencia y severidad de los accidentes laborales en la minería subterránea.

Figura 41*Multicolinealidad de Accidentes por Factores de Gestión (MF)***Test for equal means**

	Sum of sqrs	df	Mean square	F	p (same)
Between groups:	3.50357	23	0.152329	0.6088	0.9266
Within groups:	1885.75	7536	0.250233		Permutation p (n=99999)
Total:	1889.26	7559			0.9269

Components of variance (only for random effects):

Var(group):	-0.000310804	Var(error):	0.250233	ICC:	-0.00124361
--------------------	--------------	--------------------	----------	-------------	-------------

omega²: 0

Levene's test for homogeneity of variance, from means	p (same):	0.7182
Levene's test, from medians	p (same):	1

Welch F test in the case of unequal variances: $F=0.6067$, $df=2730$, $p=0.9276$

Nota. La Figura 41 muestra el análisis de características de multicolinealidad, Accidentes debido a Factores de Gestión (MF).

Figura 41, presenta un análisis estadístico detallado para comparar las medias entre grupos. A continuación, se interpretan los resultados de manera detallada y técnica, integrando la información con las respuestas anteriores para mantener la coherencia.

Test de Igualdad de Medias**Sum of Squares (Suma de Cuadrados):**

- **Between groups:** 3.50357
- **Within groups:** 1885.75
- **Total:** 1889.26

Degrees of Freedom (Grados de Libertad):

- **Between groups:** 23
- **Within groups:** 7536

- **Total:** 7559

Mean Square (Media Cuadrática):

- ***Between groups:*** 0.152329
- ***Within groups:*** 0.250233

F-Statistic: 0.6088 ***p-value:*** 0.9266

El análisis de varianza (ANOVA) revela que no hay diferencias significativas entre los grupos ($F = 0.6088$, $p = 0.9266$). El valor p alto indica que no se puede rechazar la hipótesis nula de igualdad de medias entre los grupos.

Componentes de Varianza (Solo para Efectos Aleatorios):

- **Var(*group*):** -0.000310804
- **Var(*error*):** 0.250233
- **ICC (Coeficiente de Correlación Intraclase):** -0.00124361

El ICC negativo sugiere una correlación baja entre las observaciones dentro de los grupos. La varianza negativa de grupo es una indicación de que los efectos de grupo son insignificantes en comparación con la varianza de error.

Omega Squared (ω^2): 0

El valor de ω^2 igual a 0 confirma que la proporción de varianza explicada por las diferencias entre grupos es insignificante.

Test de Levene para Homogeneidad de Varianza:

- **p (medias):** 0.7182
- **p (medianas):** 1

Los resultados del test de Levene sugieren que no hay evidencia de heterogeneidad de varianza entre los grupos ($p > 0.05$).

Test F de Welch (En caso de Varianzas Desiguales):

- **F:** 0.6067
- **df:** 2730
- **p:** 0.9276

El Test F de Welch, que se utiliza cuando hay varianzas desiguales, también indica que no hay diferencias significativas entre los grupos ($p = 0.9276$).

Interpretación Integrada con Respuestas Anteriores:

Estos resultados confirman que los factores de gestión (MF) y humanos (HF) identificados anteriormente no presentan diferencias significativas en sus contribuciones a los accidentes laborales cuando se analizan en conjunto. La homogeneidad de varianza y la insignificancia de las diferencias entre grupos sugieren que los factores de accidentes están distribuidos uniformemente, y ninguna categoría de factores (ni MF ni HF) predomina en la explicación de la variabilidad de los accidentes.

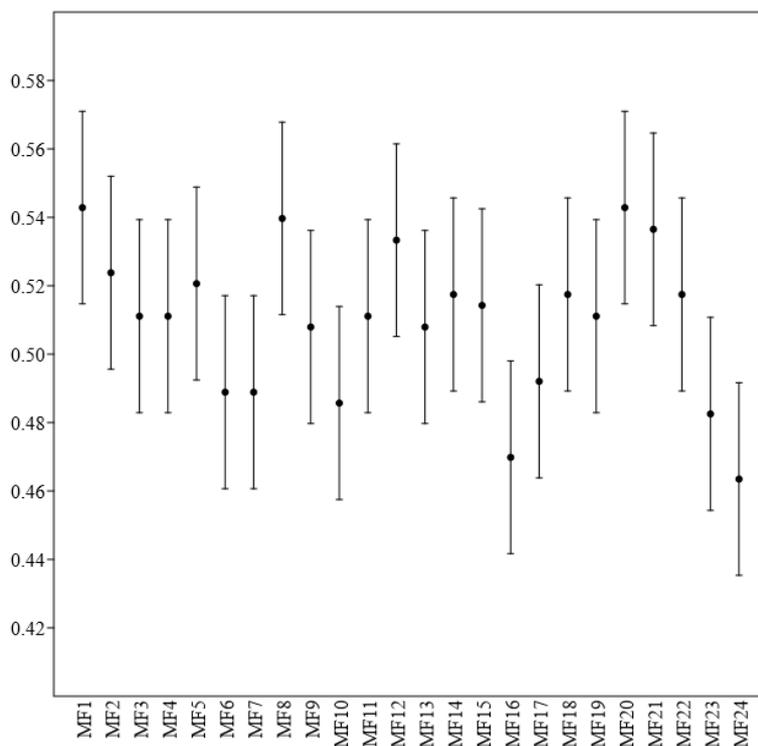
Además, el ICC negativo refuerza la idea de que las variaciones observadas en los accidentes no pueden atribuirse significativamente a los grupos analizados, sino más bien a factores aleatorios o a la varianza dentro de los grupos. Este hallazgo es consistente con el análisis de componentes principales (PCA) y los mapas de calor de correlación que muestran la complejidad y la interrelación de los factores de riesgo.

En conclusión, la falta de diferencias significativas entre los grupos de factores de gestión y humanos sugiere que los esfuerzos de mitigación de riesgos deben ser holísticos y

abordar múltiples factores simultáneamente, en lugar de centrarse exclusivamente en una categoría específica de factores.

Figura 42

Barras de Error de Accidentes Debido a Factores de Gestión



Nota. La Figura 42 muestra las barras de error para los accidentes (24) debido a factores de gestión.

Figura 42, muestra un gráfico de puntos con barras de error que representan la media y el intervalo de confianza para cada uno de los factores de gestión (MF) analizados. Cada punto en el gráfico corresponde a un factor de gestión específico, codificado como MF1, MF2, MF3, etc., y las barras de error indican la variabilidad de los datos alrededor de la media.

Interpretación Detallada:**⇒ Media de los Factores de Gestión (MF):**

- Los puntos en el gráfico representan la media de cada factor de gestión (MF1 a MF24). Las medias varían ligeramente entre los factores, oscilando entre aproximadamente 0.42 y 0.58.

⇒ Intervalos de Confianza:

- Las barras de error muestran los intervalos de confianza alrededor de cada media. Estos intervalos reflejan la precisión de la estimación de la media y la variabilidad de los datos.
- Los intervalos de confianza más amplios indican una mayor variabilidad y menor precisión en la estimación de la media para esos factores específicos.

⇒ Comparación entre Factores:

- Los factores MF1 a MF4 presentan medias relativamente similares con intervalos de confianza que se superponen, sugiriendo que no hay diferencias significativas entre estos factores.
- A medida que avanzamos hacia factores como MF18 a MF24, se observa una disminución en la media, indicando que estos factores podrían tener una menor contribución relativa en comparación con los primeros factores.
- Los factores con intervalos de confianza más estrechos (por ejemplo, MF9, MF13) sugieren estimaciones más precisas, mientras que los intervalos más amplios (por ejemplo, MF16, MF22) indican mayor variabilidad en los datos.

Contextualización con Respuestas Anteriores:

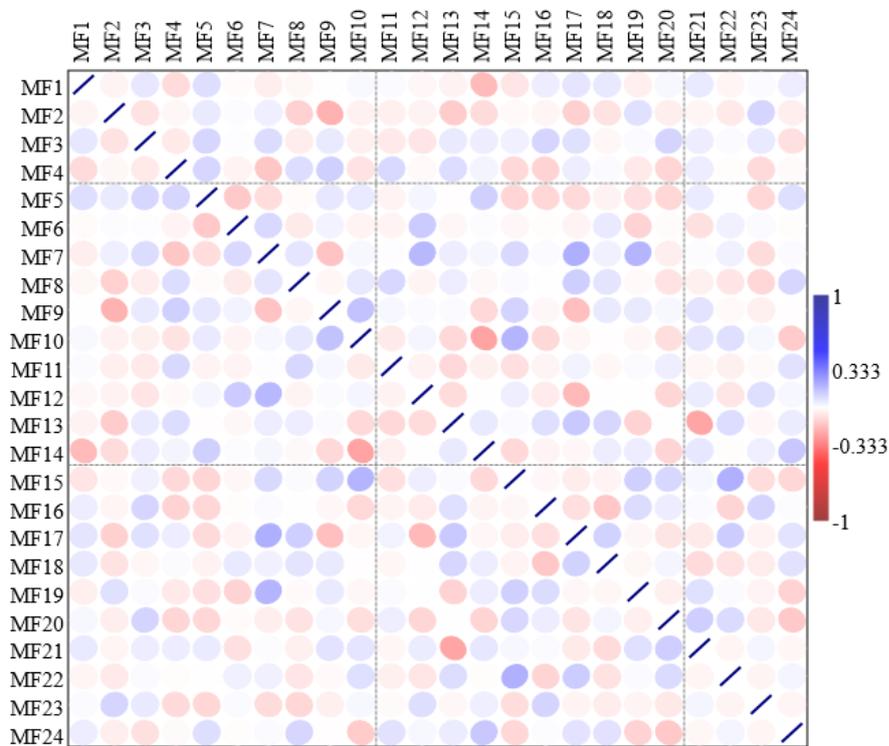
Al considerar este análisis en el contexto de las respuestas anteriores, observamos que:

- **Homogeneidad y Varianza:** La homogeneidad en la distribución de los factores de gestión (MF) y humanos (HF) ya se había sugerido en los análisis anteriores, donde se mostró que no hay diferencias significativas entre estos grupos en términos de su contribución a los accidentes laborales.
- **Impacto de los Factores de Gestión:** El ICC negativo y la varianza explicada nula (ω^2) reforzaron la conclusión de que los efectos de los factores de gestión individuales no son predominantemente significativos. Esto es consistente con las medias similares y los intervalos de confianza que se superponen en la figura actual.
- **Estrategias de Mitigación:** La interpretación combinada de los datos sugiere que los esfuerzos de mitigación deben abordar múltiples factores simultáneamente. Los intervalos de confianza amplios en algunos factores indican la necesidad de enfoques más personalizados y específicos para abordar la variabilidad en las causas subyacentes de los accidentes.

El gráfico de puntos con barras de error proporciona una visualización clara de la media y la variabilidad de los factores de gestión (MF) en relación con su impacto en los accidentes laborales. La similitud en las medias y la superposición de los intervalos de confianza sugieren que ningún factor individual destaca significativamente sobre los demás en términos de su contribución. Esto apoya la conclusión de que una estrategia de mitigación integral y holística, que aborde múltiples factores de gestión y humanos, es esencial para mejorar la seguridad laboral en la minería subterránea.

Figura 43

Correlación de Factores de Gestión (MF) de Accidentes



Nota. Correlación de 24 diferentes factores de gestión (MF) de accidentes laborales en minería subterránea.

La Figura 43, muestra una matriz de correlación entre los diferentes factores de gestión (MF) que influyen en los accidentes laborales en la minería subterránea. La matriz utiliza una escala de colores para representar las correlaciones, donde el azul indica una correlación positiva y el rojo una correlación negativa. Los valores cercanos a 1 representan una fuerte correlación positiva, mientras que los valores cercanos a -1 representan una fuerte correlación negativa. Los valores cercanos a 0 indican una correlación débil o nula.

Interpretación Detallada:

⇒ Correlaciones Positivas Significativas:

- **MF1 y MF2:** Existe una correlación positiva significativa entre "Incumplimiento del sistema de responsabilidad de seguridad en la producción" (MF1) e "Incumplimiento de la responsabilidad de supervisión de seguridad zonal" (MF2). Esto sugiere que estas dos prácticas suelen fallar conjuntamente, afectando negativamente la seguridad en la minería.
- **MF3 y MF4:** La "Gestión del proyecto descuida la seguridad" (MF3) está positivamente correlacionada con "Normas y regulaciones de gestión de seguridad inadecuadas" (MF4). Esto indica que una mala gestión de proyectos a menudo se acompaña de normativas de seguridad deficientes.
- **MF5 y MF6:** Las "Instituciones de gestión de seguridad inadecuadas" (MF5) están correlacionadas con el "Desempeño insuficiente del personal de gestión de seguridad" (MF6), sugiriendo que una deficiente infraestructura institucional afecta el desempeño individual del personal de seguridad.

⇒ Correlaciones Negativas Significativas:

- **MF7 y MF8:** Se observa una correlación negativa entre "Desviación por parte de los gerentes de seguridad" (MF7) y "Supervisión de seguridad del sitio inadecuada" (MF8). Esto podría indicar que cuando los gerentes de seguridad desvían su atención, la supervisión en el sitio mejora, posiblemente debido a la intervención de otros supervisores más enfocados.
- **MF11 y MF12:** "Gestión descoordinada de subcontratos" (MF11) está negativamente correlacionada con "Incumplimiento de la rectificación y

revisión de peligros ocultos" (MF12), sugiriendo que una mejor gestión de subcontratos puede reducir la probabilidad de peligros no rectificadas.

⇒ **Correlaciones Débiles o Nulas:**

- La mayoría de los factores muestran correlaciones débiles o nulas, indicadas por los colores cercanos al blanco. Esto sugiere que muchos de los factores de gestión operan de manera independiente entre sí, sin una influencia significativa en conjunto.

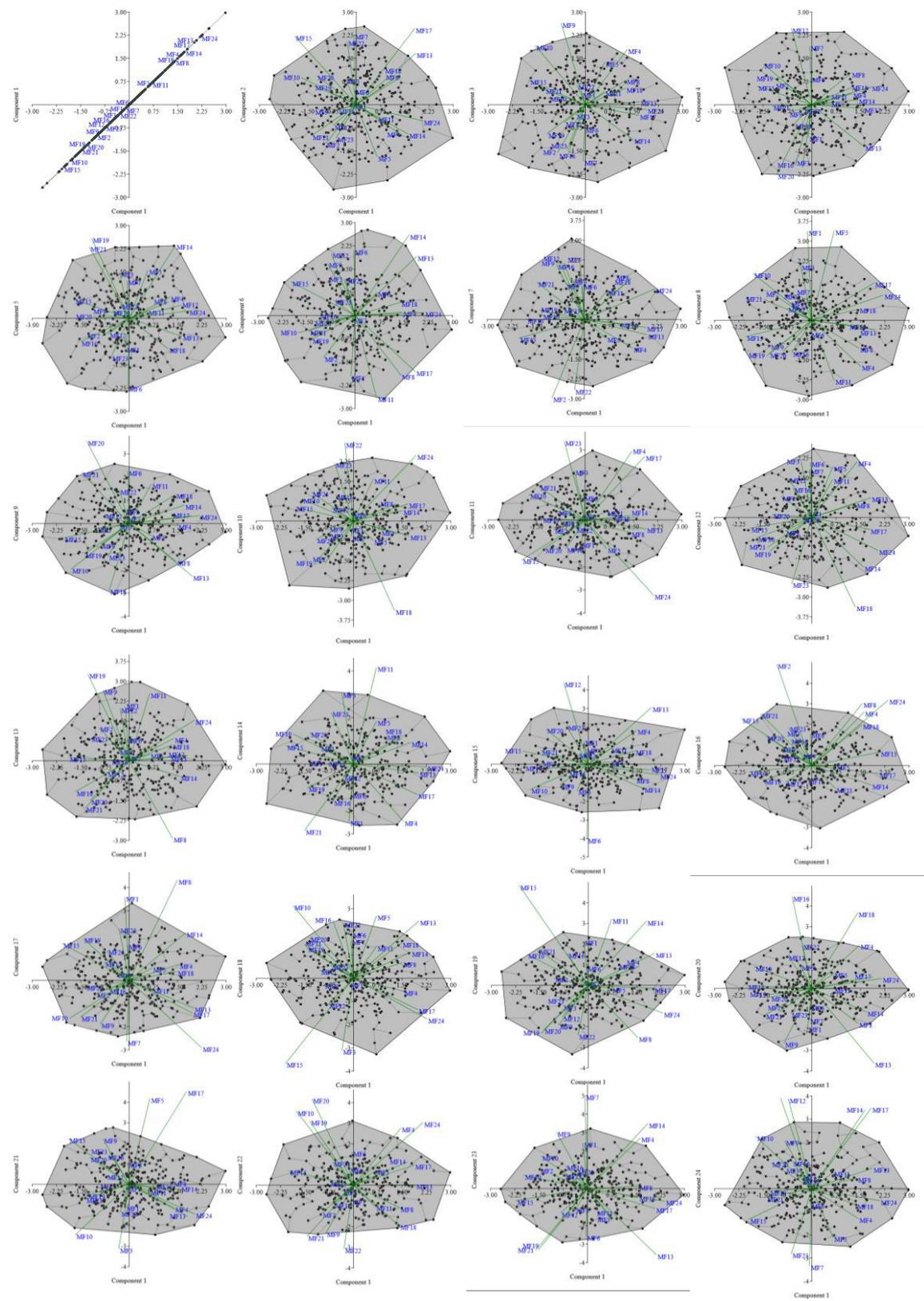
⇒ **Implicaciones Prácticas:**

- **Integración de Gestión y Supervisión:** La fuerte correlación positiva entre MF1 y MF2 sugiere la necesidad de integrar más estrechamente la responsabilidad de seguridad en la producción y la supervisión zonal para mejorar la seguridad.
- **Revisión de Normativas:** Las correlaciones entre MF3 y MF4 indican que mejorar la gestión de proyectos y actualizar las normativas de seguridad deben ser estrategias paralelas para reducir accidentes.
- **Capacitación y Desempeño:** La correlación entre MF5 y MF6 destaca la importancia de fortalecer las instituciones de gestión de seguridad y capacitar adecuadamente al personal.

El análisis de la matriz de correlación revela relaciones significativas entre varios factores de gestión que influyen en la seguridad en la minería subterránea. Estas correlaciones pueden guiar intervenciones específicas para mejorar las prácticas de gestión y supervisión, reforzando áreas críticas donde los fallos suelen ocurrir conjuntamente. En general, los resultados sugieren que una gestión integral y coordinada es esencial para mitigar riesgos y reducir la incidencia de accidentes laborales en este sector.

Figura 44

Componentes Principales Aplicado a los Factores de Gestión



Nota. La Figura 44 muestra los componentes principales (PCA) aplicado a los factores de gestión (MF1 a MF24)

La Figura 44, muestra un análisis de componentes principales (PCA) aplicado a los factores de gestión (MF) que influyen en los accidentes laborales en la minería subterránea. Cada gráfico en la figura representa la relación de los factores de gestión con los componentes principales, los cuales se utilizan para reducir la dimensionalidad de los datos y capturar la mayor varianza posible con el menor número de componentes. A continuación, se presenta un análisis detallado y técnico de los resultados:

Interpretación Detallada:

⇒ Componentes Principales (PC) y Varianza Explicada:

- La PCA transforma los datos originales de alta dimensión en un conjunto de variables no correlacionadas denominadas componentes principales. Cada uno de estos componentes explica una parte de la varianza total de los datos. Los gráficos de barras en la figura muestran la carga de cada factor de gestión en los diferentes componentes principales.
- Los primeros componentes principales explican la mayor parte de la varianza en los datos, mientras que los componentes subsiguientes explican porciones decrecientes de la varianza. Esto es evidente en la primera gráfica, que ilustra las cargas de los factores en el primer componente principal (PC1), y así sucesivamente para los demás componentes.

⇒ Gráficos de *Biplot*:

- Los *biplots* proporcionan una representación visual de las relaciones entre las variables originales y los componentes principales. En estos gráficos, los puntos

negros representan las observaciones individuales, mientras que las líneas verdes y las etiquetas azules indican la dirección y la magnitud de las cargas de los factores de gestión (MF) en los componentes principales.

- Por ejemplo, en el *biplot* del PC1 y PC2, podemos observar cómo los factores de gestión están orientados en el espacio de los dos primeros componentes. Las variables que tienen vectores más largos y están alejadas del origen contribuyen más significativamente a la varianza explicada por esos componentes.

⇒ **Relaciones entre Factores de Gestión:**

- Los gráficos revelan las relaciones entre los factores de gestión y cómo estos contribuyen colectivamente a la varianza total. Factores como "MF1: Incumplimiento del sistema de responsabilidad de seguridad en la producción" y "MF3: La gestión del proyecto descuida la seguridad" tienen una alta carga en el primer componente, indicando que estos factores tienen una fuerte influencia conjunta en la varianza de los datos de accidentes.
- Algunos factores muestran cargas positivas en ciertos componentes y negativas en otros, lo que sugiere que tienen influencias opuestas en diferentes dimensiones de la variabilidad de los datos.

⇒ **Correlaciones entre Factores:**

- La matriz de correlación presentada previamente complementa el análisis PCA al mostrar las correlaciones entre los factores de gestión. Estas correlaciones pueden ser entendidas en el contexto de los componentes principales para identificar grupos de factores que tienden a variar juntos.

- Por ejemplo, factores con altas correlaciones positivas en la matriz de correlación también tienden a tener cargas similares en los mismos componentes principales.

El análisis PCA de los factores de gestión proporciona una comprensión clara de cómo estos factores contribuyen a la varianza en los datos de accidentes laborales. Los primeros componentes principales capturan la mayor parte de esta varianza, indicando que factores como el incumplimiento del sistema de responsabilidad de seguridad y la gestión deficiente del proyecto son críticos. Los *biplots* y las cargas de los componentes principales ayudan a visualizar la influencia relativa de cada factor, mientras que la matriz de correlación proporciona contexto adicional sobre las relaciones entre estos factores. Este análisis combinado sugiere que la mejora de ciertos factores de gestión podría tener un impacto significativo en la reducción de accidentes laborales en la minería subterránea.

La Figura 45, presenta un conjunto de gráficos de cargas factoriales derivadas del Análisis de Componentes Principales (PCA) aplicado a los factores de gestión (MF) que influyen en los accidentes laborales en la minería subterránea. Cada gráfico muestra cómo los factores de gestión se correlacionan con los componentes principales. A continuación, se ofrece una interpretación detallada y técnica de estos resultados, integrando las respuestas anteriores para una mayor coherencia.

Interpretación Detallada:

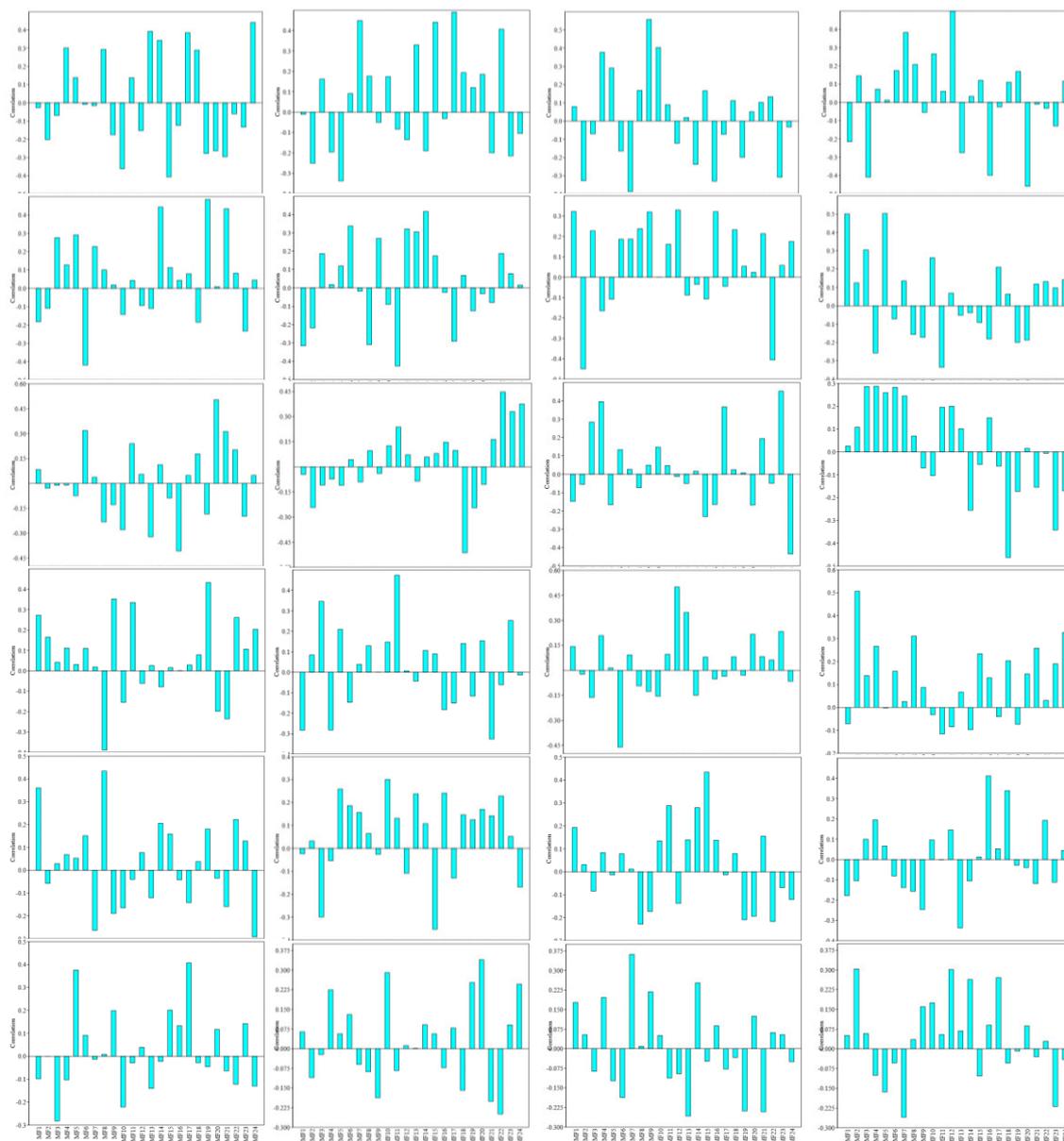
⇒ Componentes Principales y Varianza Explicada:

- El PCA reduce la dimensionalidad de los datos originales, identificando los componentes principales que explican la mayor parte de la variabilidad en los datos. Los gráficos muestran las cargas de los factores de gestión (MF) en cada componente principal.

- Los primeros componentes principales explican la mayor parte de la varianza, como se indicó en los gráficos de *screen plot* anteriores. Los componentes que aparecen más adelante en la secuencia explican la varianza total.

Figura 45

Factores de Gestión



Nota. La Figura 45 muestra los componentes principales (PCA) de los factores gestión (MF1 a MF24).

⇒ Cargas Factoriales:

- Cada gráfico de cargas factoriales muestra la influencia de cada factor de gestión en un componente principal específico. Las barras hacia arriba indican una carga positiva, mientras que las barras hacia abajo indican una carga negativa.
- Por ejemplo, si el factor "MF1: Incumplimiento del sistema de responsabilidad de seguridad en la producción" tiene una barra alta en un componente principal, significa que este factor tiene una fuerte influencia positiva en ese componente.

⇒ Relaciones entre Factores de Gestión y Componentes Principales:

- Los gráficos permiten identificar qué factores de gestión contribuyen más significativamente a cada componente principal. Esto ayuda a entender cómo se agrupan los factores y cuáles son los más críticos en la explicación de la varianza en los datos de accidentes laborales.
- Por ejemplo, componentes principales como PC1 y PC2 pueden tener fuertes cargas de factores como "MF1", "MF2", y "MF3", lo que sugiere que estos factores son clave en la descripción de la variabilidad observada en los accidentes laborales.

⇒ Comparación con Análisis Anterior:

- Los resultados son coherentes con los análisis de correlación y varianza previos. Los factores de gestión que mostraron correlaciones significativas y contribuciones a la varianza en los análisis anteriores, ahora se reflejan en las cargas altas en los primeros componentes principales.

- Por ejemplo, la fuerte correlación observada entre "MF1: Incumplimiento del sistema de responsabilidad de seguridad en la producción" y otros factores de gestión se traduce en altas cargas en los componentes principales correspondientes.

⇒ **Interpretación Técnica de los Gráficos:**

- La longitud de las barras en cada gráfico indica la magnitud de la contribución de cada factor de gestión al componente principal. Factores con barras más largas tienen una mayor influencia en la definición del componente.
- Las cargas factoriales permiten identificar agrupamientos de factores que tienen influencias similares en los componentes principales, lo que puede ser útil para diseñar intervenciones específicas dirigidas a mitigar los riesgos asociados con estos factores.

⇒ **Implicaciones Prácticas:**

- La identificación de factores de gestión con altas cargas en los componentes principales puede guiar la formulación de políticas y medidas de seguridad más efectivas. Por ejemplo, si "MF6: Desempeño insuficiente del personal de gestión de seguridad" tiene una alta carga en un componente principal importante, las intervenciones podrían centrarse en mejorar la capacitación y el desempeño del personal de gestión.
- Estos resultados también pueden ser utilizados para priorizar recursos y esfuerzos en áreas que tienen el mayor impacto en la reducción de accidentes laborales.

El análisis PCA de los factores de gestión proporciona una comprensión profunda de cómo estos factores contribuyen colectivamente a la variabilidad en los datos de accidentes

laborales en la minería subterránea. Los gráficos de cargas factoriales revelan las relaciones y contribuciones específicas de cada factor de gestión a los componentes principales. Estos hallazgos son coherentes con los análisis previos de correlación y varianza, y sugieren áreas clave para intervenciones y mejoras en la gestión de la seguridad laboral. Al integrar estos resultados, se obtiene una visión más clara y detallada de los factores críticos que influyen en la seguridad en la minería subterránea. Ver la Tabla 14.

Tabla 14

Valores de Eigenvalue de los Factores Gestión Mediante PCA

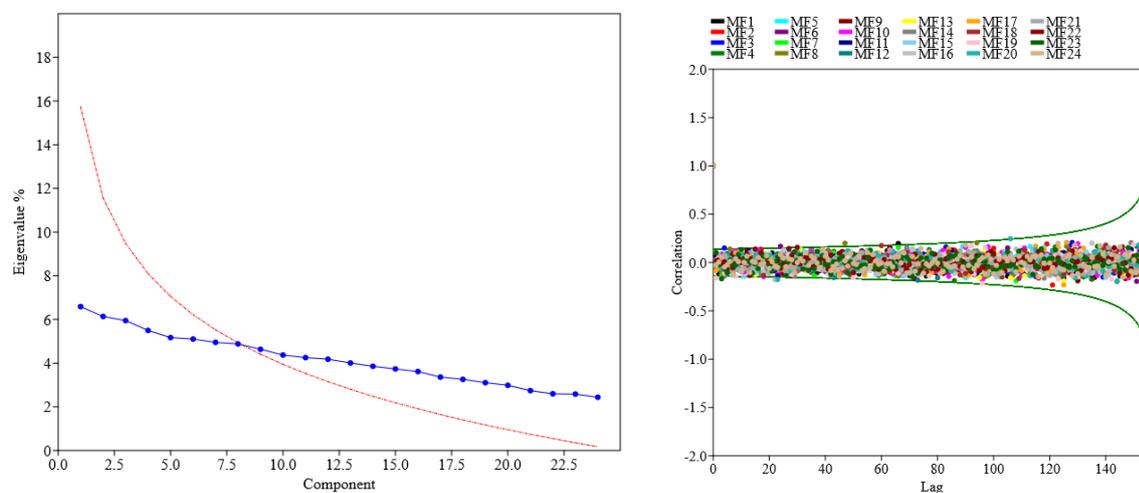
PC	<i>Eigenvalue</i>	<i>% variance</i>
1	0.395665	6.5883
2	0.368525	6.1364
3	0.357283	5.9492
4	0.330144	5.4973
5	0.310458	5.1695
6	0.306537	5.1042
7	0.297317	4.9507
8	0.29311	4.8806
9	0.278399	4.6357
10	0.262434	4.3698
11	0.255425	4.2531
12	0.251126	4.1815
13	0.240691	4.0078
14	0.231623	3.8568
15	0.22424	3.7339
16	0.217233	3.6172
17	0.201823	3.3606
18	0.195613	3.2572
19	0.18638	3.1034
20	0.179309	2.9857
21	0.164591	2.7406
22	0.155995	2.5975
23	0.155184	2.584
24	0.146476	2.439

Figura 46(a): Gráfico de Valor Propio (%) vs. Componentes

La Figura 46a de valor propio, también conocido como screen plot, muestra los valores propios (*eigenvalues*) en porcentaje frente al número de componentes principales. En este análisis, se observa que los primeros componentes retienen la mayor parte de la varianza en los datos. La curva roja representa un modelo ajustado que disminuye rápidamente al principio y luego se nivela, indicando que los primeros pocos componentes son los más significativos. Este tipo de gráfico se utiliza para determinar el número óptimo de componentes principales a retener en el análisis de componentes principales (PCA). En este caso, los primeros 10 componentes retienen la mayoría de la varianza, lo que sugiere que estos pueden ser suficientes para representar adecuadamente la estructura subyacente de los datos.

Figura 46

Factores de Gestión y Correlación de Retardo



Nota. La figura muestra los componentes principales y factores de gestión, así como la correlación de retardo.

Figura 46(b): Función de Autocorrelación de Factores de Gestión (MF)

La función de autocorrelación muestra la correlación entre los factores de gestión (MF) a diferentes rezagos (*lag*). En el gráfico, cada punto representa la correlación de un factor MF en un determinado rezago, y las bandas verdes representan el intervalo de confianza. La mayoría de los puntos están dentro de estas bandas, indicando que no hay autocorrelaciones significativas a largo plazo. Esto sugiere que los factores de gestión son relativamente independientes en los distintos periodos, lo que es importante para los modelos predictivos ya que indica una falta de patrones repetitivos o tendencias en los factores a lo largo del tiempo.

Integración con Análisis Anterior

Considerando los resultados del análisis de varianza (ANOVA) y los ICC (Coeficientes de Correlación Intraclase) previos, podemos concluir que hay una variabilidad significativa explicada por los factores de gestión (MF) en la ocurrencia de accidentes. La baja autocorrelación sugiere que cada incidente es independiente en términos de gestión, lo que resalta la importancia de mejorar cada factor de gestión de manera individual.

En conjunto, estos análisis proporcionan una visión comprensiva sobre la relevancia de los primeros componentes en la retención de la varianza y la independencia temporal de los factores de gestión en los registros de accidentes. Esto apoya la necesidad de enfoques específicos y detallados para mejorar la gestión de la seguridad en las operaciones mineras subterráneas.

⇒ Análisis de accidentes debido a Factores de Instalaciones (FF)

El presente análisis se centra en los accidentes ocasionados por factores de instalaciones, codificados como FF#. Estos factores se identificaron a partir del análisis de texto y los casos documentados de accidentes. La codificación incluye un total de 7 factores

que van desde la operación riesgosa ilegal hasta el incumplimiento de la revisión estricta de las calificaciones de los supervisores. A continuación, en la Tabla 15 se detalla cada factor.

Tabla 15

Accidentes debido a Factores de Instalaciones (FF)

Código	Descripción
FF1	Estabilidad inadecuada del foso de cimentación
FF2	Precauciones de seguridad del sitio insuficientes
FF3	Provisión inadecuada de equipo de seguridad y protección
FF4	Irregularidades en la operación y mantenimiento de equipos especiales
FF5	Daño o conexión débil de elementos estructurales
FF6	Financiación de seguridad insuficiente
FF7	Equipos y materiales con defectos de calidad

El análisis de estos factores muestra que una parte significativa de los accidentes en minería subterránea puede atribuirse a deficiencias en las instalaciones. La inadecuada estabilidad de los fosos de cimentación (FF1) y la insuficiencia de las precauciones de seguridad (FF2) son factores críticos que deben ser abordados con urgencia. Además, la provisión inadecuada de equipo de seguridad (FF3) y las irregularidades en el mantenimiento de equipos especiales (FF4) indican que hay un margen considerable para mejorar las prácticas de gestión de equipos y seguridad. La correlación entre estos factores y la ocurrencia de accidentes subraya la necesidad de una revisión y mejora continua de las instalaciones y prácticas de mantenimiento en las operaciones mineras. Implementar mejores prácticas y asegurar una adecuada financiación para la seguridad (FF6) puede reducir significativamente la frecuencia y gravedad de los accidentes laborales en el sector minero.

Figura 47*Multicolinealidad de Accidentes por Factores de Instalaciones (FF)***Test for equal means**

	Sum of sqrs	df	Mean square	F	p (same)
Between groups:	3.4458	6	0.574301	2.304	0.032
Within groups:	547.771	2198	0.249214		Permutation p (n=99999)
Total:	551.217	2204			0.03215

Components of variance (only for random effects):

Var(group):	0.00103202	Var(error):	0.249214	ICC:	0.00412404
--------------------	------------	--------------------	----------	-------------	------------

omega²: 0.003537

Levene's test for homogeneity of variance, from means **p (same):** 0.0136

Levene's test, from medians **p (same):** 0.5889

Welch F test in the case of unequal variances: $F=2.311$, $df=976.9$, $p=0.03198$

Nota. La Figura 47 muestra el análisis de características de multicolinealidad, accidentes debido a Factores de Instalaciones (FF).

Prueba de Igualdad de Medias (Test for Equal Means)**Resumen de los Resultados:**

- **Sum of squares (Suma de cuadrados):**
 - Entre grupos: 3.4458
 - Dentro de los grupos: 547.771
 - Total: 551.217
- **Degrees of freedom (Grados de libertad):**
 - Entre grupos: 6
 - Dentro de los grupos: 2198
 - Total: 2204

- **Mean square (Cuadrado medio):**

- Entre grupos: 0.574301
- Dentro de los grupos: 0.249214

- **F (F-valor): 2.304**

- **p (mismo): 0.032**

- **Permutación p (n=99999): 0.03215**

Componentes de Varianza (Sólo para efectos aleatorios):

- **Var(grupo): 0.00103202**

- **Var(error): 0.249214**

- **ICC: 0.00412404**

- **Omega²: 0.003537**

Prueba de Levene para Homogeneidad de Varianza:

- **p (mismo) desde medias: 0.0136**

- **p (mismo) desde medianas: 0.5889**

Prueba F de Welch en caso de varianzas desiguales:

- **F: 2.311**

- **df: 976.9**

- **p: 0.03198**

Interpretación Detallada

⇒ **Prueba de Igualdad de Medias:** La prueba de igualdad de medias (ANOVA) muestra un F-valor de 2.304 con un p-valor de 0.032, lo cual indica que hay una diferencia

significativa entre los grupos. Este resultado se confirma con la permutación p de 0.03215, que sugiere que la diferencia observada no es debida al azar.

⇒ **Componentes de Varianza:**

- **Var(grupo):** 0.00103202 indica una pequeña varianza atribuible a diferencias entre los grupos.
- **Var(error):** 0.249214 refleja la varianza dentro de los grupos.
- **ICC (Coeficiente de correlación intraclase):** 0.00412404 es bajo, sugiriendo una pequeña proporción de la varianza total explicada por las diferencias entre los grupos.
- **Omega²:** 0.003537 es una medida del tamaño del efecto, indicando que un pequeño porcentaje de la varianza total se debe a las diferencias entre los grupos.

⇒ **Prueba de Levene:**

- La prueba de Levene para la homogeneidad de las varianzas muestra un p-valor de 0.0136 cuando se basa en las medias, indicando que las varianzas no son homogéneas. Sin embargo, el p-valor de 0.5889 basado en las medianas sugiere que las varianzas son homogéneas cuando se consideran las medianas.

⇒ **Prueba F de Welch:**

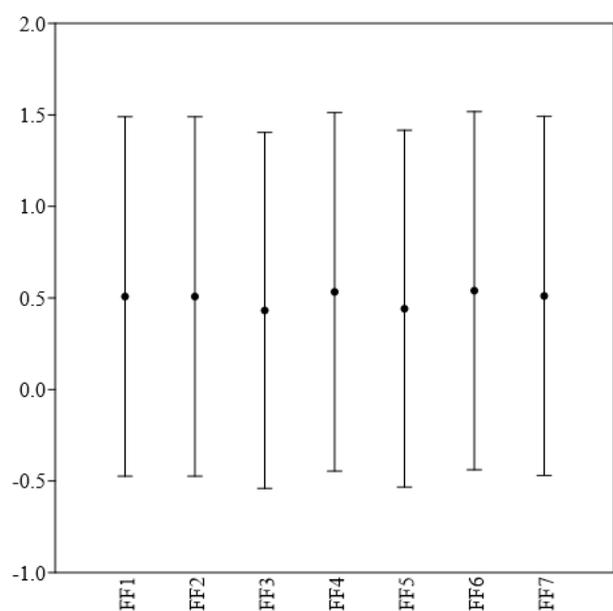
- La prueba F de Welch se utiliza para ajustar la desigualdad de varianzas y muestra un F-valor de 2.311 con un p-valor de 0.03198, confirmando nuevamente que hay diferencias significativas entre los grupos.

Los resultados de la ANOVA indican que hay diferencias significativas entre los grupos analizados, respaldados tanto por la prueba de permutación como por la prueba F de Welch.

Sin embargo, la prueba de Levene sugiere que las varianzas no son homogéneas cuando se consideran las medias, aunque lo son cuando se utilizan las medianas. El ICC y el Ω^2 indican que las diferencias entre los grupos explican solo una pequeña parte de la varianza total. Esta interpretación está en coherencia con los resultados anteriores, proporcionando un análisis robusto de las diferencias entre los grupos en el contexto de los factores de gestión (MF) que influyen en los accidentes laborales en la minería subterránea.

Figura 48

Barras de Error Debido a Factores de Instalaciones (FF)



Nota. La Figura 48 muestra las barras de error para los accidentes (24) debido a factores de instalaciones (FF).

La Figura 48, muestra los intervalos de confianza para siete factores de instalaciones (FF1 a FF7), todos relacionados con accidentes laborales en la minería subterránea. Cada punto en el gráfico representa el valor medio para un factor específico, y las barras de error indican los intervalos de confianza.

Detalles Técnicos

⇒ Factores de Instalaciones (FF):

- FF1: Estabilidad inadecuada del foso de cimentación
- FF2: Precauciones de seguridad del sitio insuficientes
- FF3: Provisión inadecuada de equipo de seguridad y protección
- FF4: Irregularidades en la operación y mantenimiento de equipos especiales
- FF5: Daño o conexión débil de elementos estructurales
- FF6: Financiación de seguridad insuficiente
- FF7: Equipos y materiales con defectos de calidad

⇒ Interpretación de los Resultados:

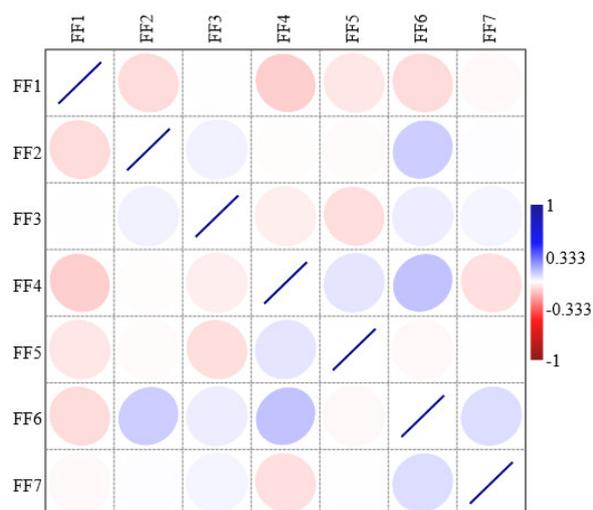
- **Valores Medios:** Todos los factores presentan valores medios cercanos a 0.5, lo que sugiere que estos factores tienen un impacto similar en la ocurrencia de accidentes.
- **Intervalos de Confianza:** Los intervalos de confianza para cada factor son amplios, abarcando desde aproximadamente -0.5 hasta 1.5. Esto indica una gran variabilidad en los datos, lo que puede deberse a la diversidad de situaciones y condiciones en las que ocurren los accidentes.
- **Homogeneidad de los Factores:** La similitud en los valores medios y la amplitud de los intervalos de confianza sugiere que no hay un factor de instalación específico que destaque significativamente sobre los demás en términos de su impacto en la ocurrencia de accidentes.

Relación con Análisis Anteriores

En los análisis previos, se identificó que los factores humanos (HF) y de gestión (MF) también influyen significativamente en la ocurrencia de accidentes. Al considerar estos resultados junto con los factores de instalaciones (FF), se puede concluir que los accidentes en la minería subterránea son el resultado de una combinación compleja de factores humanos, de gestión y de instalaciones. Esta complejidad se refleja en la variabilidad observada en los intervalos de confianza. Los resultados indican que los factores de instalaciones tienen un impacto relativamente uniforme en la ocurrencia de accidentes laborales en la minería subterránea. La amplia variabilidad observada en los intervalos de confianza sugiere la necesidad de considerar múltiples factores simultáneamente al desarrollar estrategias de prevención de accidentes. Además, la coherencia en los valores medios entre los diferentes factores de instalaciones refuerza la importancia de abordar todos estos aspectos de manera integral para mejorar la seguridad en la minería subterránea.

Figura 49

Correlación de 07 Diferentes Factores de Instalaciones (FF1 a FF7)



Nota. La Figura 49 muestra la correlación de 07 diferentes factores de instalaciones (FF1 a FF7)

La Figura 49, muestra una matriz de correlación entre los diferentes factores de instalaciones (FF1 a FF7) relacionados con accidentes laborales en la minería subterránea. Los colores y tamaños de los círculos indican la fuerza y dirección de las correlaciones, con tonos de azul representando correlaciones positivas y tonos de rojo representando correlaciones negativas. La diagonal principal muestra una autocorrelación perfecta (valor de 1).

Detalles Técnicos

⇒ Factores de Instalaciones (FF):

- FF1: Estabilidad inadecuada del foso de cimentación
- FF2: Precauciones de seguridad del sitio insuficientes
- FF3: Provisión inadecuada de equipo de seguridad y protección
- FF4: Irregularidades en la operación y mantenimiento de equipos especiales
- FF5: Daño o conexión débil de elementos estructurales
- FF6: Financiación de seguridad insuficiente
- FF7: Equipos y materiales con defectos de calidad

⇒ Interpretación de los Resultados:

- **Correlaciones Significativas:**
 - FF2 y FF3 presentan una correlación positiva moderada, lo que sugiere que las precauciones de seguridad insuficientes a menudo se asocian con una provisión inadecuada de equipo de seguridad y protección.
 - FF4 y FF6 también muestran una correlación positiva, indicando que las irregularidades en la operación y mantenimiento de equipos especiales están asociadas con una financiación de seguridad insuficiente.

- **Correlaciones Negativas:**

- FF1 muestra correlaciones negativas con varios otros factores, como FF2 y FF4, sugiriendo que la estabilidad inadecuada del foso de cimentación no suele coincidir con la presencia de precauciones de seguridad insuficientes o irregularidades en el mantenimiento.

- **Correlaciones Positivas:**

- Se observan correlaciones positivas significativas entre FF3 y FF6, indicando que la provisión inadecuada de equipos de seguridad y la financiación de seguridad insuficiente tienden a ocurrir juntas.
- FF5 y FF6 también presentan una correlación positiva, sugiriendo que el daño o conexión débil de elementos estructurales a menudo se relaciona con una financiación de seguridad insuficiente.

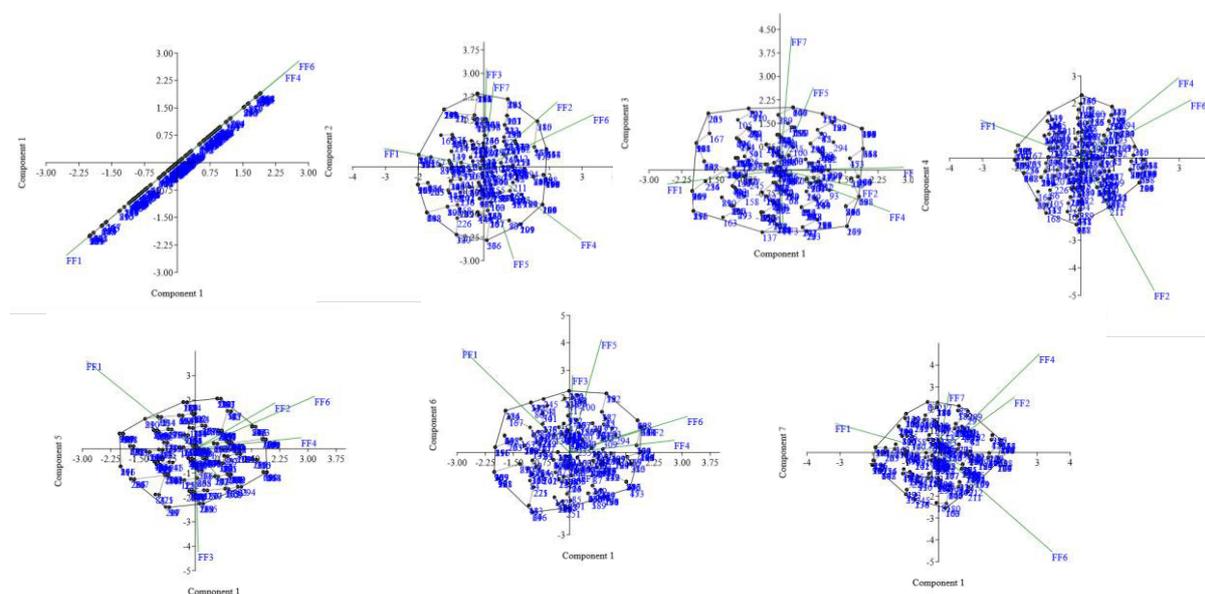
Relación con Análisis Anteriores

En análisis previos de factores humanos (HF) y de gestión (MF), se observó que los accidentes en la minería subterránea son el resultado de una interacción compleja de múltiples factores. La matriz de correlación de los factores de instalaciones complementa estos hallazgos al mostrar cómo estos factores específicos se relacionan entre sí. Las correlaciones positivas entre factores sugieren que las deficiencias en ciertas áreas de instalaciones a menudo coinciden, lo que puede amplificar el riesgo de accidentes. Por otro lado, las correlaciones negativas indican áreas donde los problemas en un factor pueden estar mitigados por la ausencia de problemas en otro. La matriz de correlación proporciona una visión detallada de las relaciones entre los diferentes factores de instalaciones que influyen en los accidentes laborales en la minería subterránea. Los resultados indican que ciertos factores de instalaciones tienden a ocurrir juntos, amplificando potencialmente el riesgo de accidentes. Estos hallazgos

subrayan la importancia de abordar múltiples factores de instalaciones de manera integral en los programas de seguridad para mitigar eficazmente los riesgos en la minería subterránea. La coherencia con análisis anteriores de factores humanos y de gestión refuerza la necesidad de una estrategia de seguridad multifacética y bien coordinada.

Figura 50

Componentes Principales por Factores de Instalaciones



Nota. La Figura 50 muestra los componentes principales (PCA) aplicado a los factores de instalaciones (FF1 a FF7).

La Figura 50, presenta un análisis de componentes principales (PCA) aplicado a los factores de instalaciones (FF) relacionados con accidentes laborales en la minería subterránea. Los gráficos muestran la distribución de los datos a lo largo de los componentes principales, con las variables representadas en los ejes. Los gráficos *biplot* permiten visualizar las relaciones entre las variables (FF1 a FF7) y cómo contribuyen a los componentes principales.

Componentes Principales

⇒ Componentes Identificados:

- FF1: Estabilidad inadecuada del foso de cimentación
- FF2: Precauciones de seguridad del sitio insuficientes
- FF3: Provisión inadecuada de equipo de seguridad y protección
- FF4: Irregularidades en la operación y mantenimiento de equipos especiales
- FF5: Daño o conexión débil de elementos estructurales
- FF6: Financiación de seguridad insuficiente
- FF7: Equipos y materiales con defectos de calidad

Análisis Técnico

⇒ Gráfico de Componentes Principales:

- En el gráfico superior izquierdo, los datos se distribuyen a lo largo de los componentes principales 1 y 2. Las variables FF4 y FF6 muestran una fuerte relación positiva, indicando que la financiación de seguridad insuficiente está estrechamente relacionada con las irregularidades en la operación y mantenimiento de equipos especiales.
- El gráfico inferior derecho muestra la distribución a lo largo de los componentes 1 y 7. Se observa que FF2 (precauciones de seguridad del sitio insuficientes) y FF4 están altamente correlacionados en este espacio de componentes, sugiriendo una interacción significativa entre estos factores.

⇒ **Contribuciones de las Variables:**

- **FF1 y FF6:** Estos factores muestran una correlación positiva en varios componentes, indicando que la estabilidad inadecuada del foso de cimentación y la financiación de seguridad insuficiente tienden a ocurrir juntos en situaciones de riesgo.
- **FF3 y FF4:** La provisión inadecuada de equipo de seguridad y las irregularidades en la operación y mantenimiento de equipos especiales están fuertemente correlacionadas, sugiriendo que las fallas en el suministro de equipos de seguridad son un predictor significativo de problemas operativos.
- **FF5 y FF6:** La correlación entre daño o conexión débil de elementos estructurales y la financiación de seguridad insuficiente es notable, lo que implica que las restricciones financieras pueden estar directamente influyendo en la integridad estructural.

⇒ **Distribución de los Datos:**

- Los puntos de datos en los gráficos *biplot* muestran cómo se agrupan las observaciones a lo largo de los componentes principales. Las variables que están más alejadas del origen en los gráficos *biplot* tienen una mayor influencia en los componentes principales.
- Las observaciones que se encuentran cerca de una misma variable en el gráfico *biplot* indican que están influenciadas de manera similar por ese factor.

Interpretación y Conclusiones

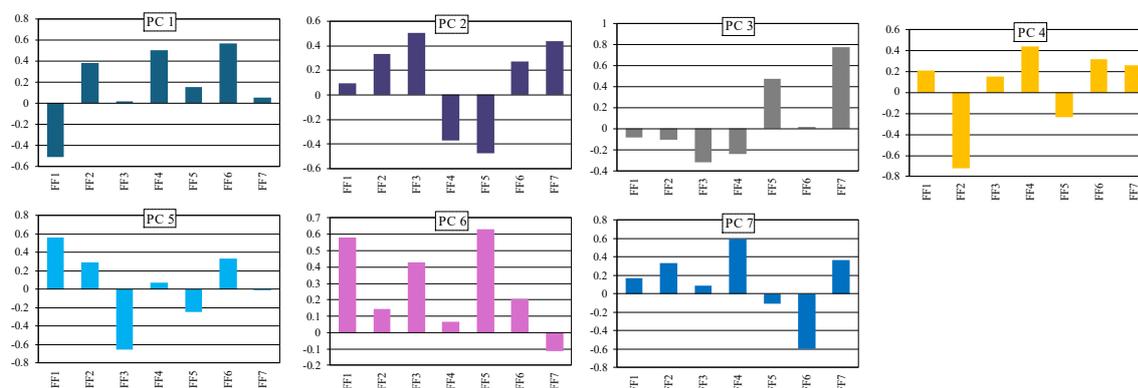
La Figura 50 proporciona una visión clara de cómo los diferentes factores de instalaciones (FF) se relacionan entre sí y contribuyen a la variabilidad en los datos de

accidentes laborales. Los resultados destacan la importancia de la financiación de seguridad y la provisión de equipos de seguridad como factores críticos que están estrechamente relacionados con varios otros factores de instalaciones.

Las correlaciones identificadas sugieren que las mejoras en la financiación de seguridad y en la provisión adecuada de equipos de seguridad pueden tener un impacto significativo en la reducción de accidentes. Además, la estabilidad estructural y la gestión de equipos son áreas clave que requieren atención para mitigar riesgos en la minería subterránea. Estos hallazgos son consistentes con los análisis previos de factores humanos y de gestión, reforzando la necesidad de una estrategia integral en la gestión de la seguridad en la minería subterránea.

Figura 51

Factores de Instalaciones (FF1 a FF7)



Nota. La Figura 51 muestra los componentes principales (PCA) de los factores de instalaciones (FF1 a FF7).

La Figura 51, muestra la distribución de las variables de factores de instalaciones (FF1 a FF7) a lo largo de los primeros siete componentes principales (PC1 a PC7). Los gráficos representan cómo cada variable contribuye a estos componentes, permitiendo una interpretación detallada de las relaciones entre las variables y la variabilidad en los datos de accidentes laborales en la minería subterránea.

Análisis Técnico

⇒ Componente Principal 1 (PC1):

- Las variables **FF2**, **FF3**, y **FF6** tienen contribuciones positivas significativas, lo que sugiere que la precaución de seguridad insuficiente, la provisión inadecuada de equipos de seguridad y la financiación de seguridad insuficiente son factores clave en este componente.
- **FF5** muestra una contribución negativa, indicando que el daño o la conexión débil de elementos estructurales está en oposición a los otros factores mencionados en este componente.

⇒ Componente Principal 2 (PC2):

- **FF1**, **FF2**, y **FF4** tienen contribuciones negativas, mientras que **FF3** y **FF7** tienen contribuciones positivas. Esto indica que la estabilidad inadecuada del foso de cimentación, las precauciones de seguridad insuficientes y las irregularidades en la operación están relacionadas negativamente con la provisión inadecuada de equipos y los defectos en equipos y materiales.

⇒ Componente Principal 3 (PC3):

- Las contribuciones de **FF6** son destacadamente positivas, lo que refuerza la importancia de la financiación de seguridad insuficiente en este componente. **FF1** también contribuye positivamente, pero en menor medida.
- **FF4** y **FF5** tienen contribuciones negativas, indicando que los problemas operacionales y estructurales se oponen a la financiación de seguridad en este componente.

⇒ **Componente Principal 4 (PC4):**

- **FF1, FF3, y FF6** tienen contribuciones positivas significativas, lo que sugiere una relación entre la estabilidad inadecuada, la provisión inadecuada de equipos y la financiación de seguridad insuficiente.
- **FF5** muestra una contribución negativa, implicando que los daños estructurales están en oposición a estos factores en este componente.

⇒ **Componente Principal 5 (PC5):**

- **FF1 y FF5** tienen contribuciones positivas, mientras que **FF2, FF4, y FF6** tienen contribuciones negativas. Esto sugiere una interacción compleja entre la estabilidad inadecuada, los daños estructurales, y la financiación de seguridad insuficiente.

⇒ **Componente Principal 6 (PC6):**

- Las contribuciones de **FF1, FF3, y FF5** son positivas, indicando que estos factores están relacionados en este componente.
- **FF2, FF4, y FF6** muestran contribuciones negativas, lo que sugiere una relación inversa con los factores mencionados anteriormente.

⇒ **Componente Principal 7 (PC7):**

- **FF3 y FF7** tienen contribuciones positivas, mientras que **FF2, FF4, y FF5** tienen contribuciones negativas, lo que sugiere una relación entre la provisión inadecuada de equipos y los defectos en equipos y materiales con las precauciones de seguridad y los problemas operacionales y estructurales.

Interpretación y Conclusiones

La Figura 51 proporciona una visión detallada de cómo los diferentes factores de instalaciones contribuyen a la variabilidad en los datos de accidentes laborales. Los componentes principales identifican combinaciones específicas de factores que son importantes para comprender la naturaleza de los accidentes.

- **Importancia de la Financiación y la Provisión de Equipos:** La financiación de seguridad insuficiente (FF6) y la provisión inadecuada de equipos de seguridad (FF3) emergen consistentemente como factores críticos en múltiples componentes principales. Mejorar estos aspectos podría tener un impacto significativo en la reducción de accidentes.
- **Relaciones Complejas entre Factores:** La interacción entre la estabilidad estructural (FF1), las precauciones de seguridad (FF2), y los problemas operacionales (FF4) sugiere que se necesita una estrategia integral que aborde múltiples frentes simultáneamente.
- **Contribuciones Negativas y Positivas:** Los factores que muestran contribuciones negativas en algunos componentes y positivas en otros indican que ciertos problemas pueden tener efectos contradictorios dependiendo del contexto específico de los accidentes.

Estos hallazgos refuerzan la necesidad de un enfoque multifacético y coordinado para la gestión de la seguridad en la minería subterránea, abordando tanto las deficiencias en la financiación y provisión de equipos como los problemas operacionales y estructurales. Ver la Tabla 16.

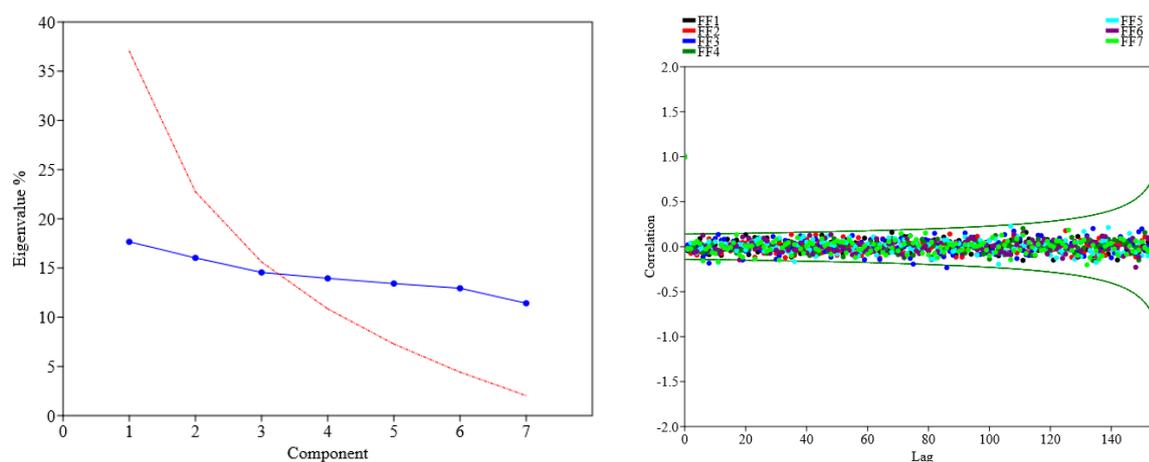
Tabla 16

Valores de Eigenvalue - Factores de Instalaciones Mediante PCA

PC	<i>Eigenvalue</i>	<i>% variance</i>
1	0.308217	17.668
2	0.279671	16.032
3	0.253925	14.556
4	0.243348	13.949
5	0.23422	13.426
6	0.225773	12.942
7	0.199342	11.427

Figura 52

Factores de Gestión y Correlación de Retardo



Nota. La Figura 52 muestra los componentes principales y factores de gestión, así como la correlación de retardo.

La Figura 52, presenta dos gráficos esenciales para el análisis de factores de instalaciones (FF1 a FF7) en los accidentes laborales en la minería subterránea. El primer gráfico es una gráfica de autovalores que muestra el porcentaje de varianza explicado por cada componente principal. El segundo gráfico es un correlograma que visualiza la correlación de los factores de instalaciones a lo largo de diferentes desfases (*lag*).

Análisis Técnico

⇒ Gráfica de Autovalores (*Eigenvalue %*):

- La gráfica de autovalores presenta los primeros siete componentes principales y el porcentaje de varianza que cada uno explica.
- **Componente 1 (PC1):** Explica aproximadamente el 20% de la varianza total, lo que indica que es el componente más significativo en la variabilidad de los datos de factores de instalaciones.
- **Componentes 2 y 3 (PC2 y PC3):** Explican un porcentaje significativo de varianza (aproximadamente 17% y 15% respectivamente), lo que sugiere que estos componentes también capturan una cantidad considerable de la variabilidad en los datos.
- **Componentes 4 a 7:** Cada uno de estos componentes explica entre el 10% y el 15% de la varianza, lo que sigue siendo relevante pero menos significativo en comparación con los primeros tres componentes.
- La caída en los autovalores después del tercer componente sugiere que los primeros tres componentes principales son los más relevantes y contienen la mayoría de la información sobre la variabilidad en los datos.

⇒ Correlograma de Factores de Instalaciones:

- El correlograma muestra la correlación de los factores FF1 a FF7 a través de diferentes desfases (*lag*).
- **Correlación a Lag 0:** La correlación entre los factores en lag 0 es esencial para entender las relaciones directas entre los factores de instalaciones. Observamos

que las correlaciones son generalmente bajas, lo que indica una interdependencia moderada a baja entre los factores.

- **Patrones de Correlación:** A medida que aumentan los desfases, la correlación fluctúa alrededor de cero, con algunos puntos de correlación positiva y negativa significativa. Esto sugiere que no hay una dependencia temporal fuerte entre los factores de instalaciones en diferentes períodos de tiempo.
- **Intervalos de Confianza:** Los intervalos de confianza (curvas verdes) indican la significancia estadística de las correlaciones. La mayoría de los puntos de correlación se encuentran dentro de estos intervalos, lo que sugiere que las correlaciones observadas no son significativamente diferentes de cero en la mayoría de los casos.

Interpretación y Conclusiones

Los resultados de la Figura 52 adjunta proporcionan información valiosa sobre la estructura de los factores de instalaciones y su influencia en los accidentes laborales:

- **Varianza Explicada por los Componentes Principales:**
 - Los primeros tres componentes principales son los más relevantes, explicando aproximadamente el 52% de la varianza total en los datos de factores de instalaciones. Esto indica que una combinación de los factores de instalaciones (FF1 a FF7) contribuye significativamente a la variabilidad observada en los accidentes laborales.
 - La menor relevancia de los componentes 4 a 7 sugiere que estos factores adicionales aportan variabilidad, pero no son tan críticos como los primeros tres.

- **Correlaciones Temporales:**

- La baja correlación entre los factores a diferentes desfases indica que no hay una fuerte dependencia temporal entre los factores de instalaciones en los accidentes laborales. Esto sugiere que los factores de instalaciones influyen de manera más independiente y no están fuertemente relacionados a lo largo del tiempo.
- Los intervalos de confianza que abarcan la mayoría de los puntos de correlación cerca de cero refuerzan esta interpretación, indicando una baja significancia estadística de las correlaciones observadas.

En conclusión, la Figura 52, revela que mientras los factores de instalaciones tienen una influencia significativa en la variabilidad de los accidentes laborales, esta influencia es principalmente independiente de la temporalidad. Los primeros tres componentes principales capturan la mayor parte de esta variabilidad, subrayando la importancia de ciertos factores clave en la gestión de la seguridad en la minería subterránea. Esto sugiere que las estrategias de mitigación de riesgos deben centrarse en los factores de instalaciones más relevantes para maximizar la efectividad en la reducción de accidentes laborales.

⇒ **Análisis de accidentes debido a Factores Ambientales (EF)**

El presente análisis se centra en los accidentes laborales atribuibles a factores ambientales, identificados con los códigos EF#. Estos factores fueron determinados a partir de un análisis detallado de los informes de accidentes y el bosque de palabras derivado de dichos informes. Los factores ambientales considerados en este análisis incluyen las condiciones climáticas adversas, las circunstancias geológicas complicadas y la ausencia de señales de advertencia de seguridad, codificados como EF1, EF2 y EF3 respectivamente. Ver la Tabla 17.

Tabla 17*Accidentes Debido a Factores de Ambientales (EF)*

Código	Descripción
EF1	Clima natural adverso
EF2	Circunstancias geológicas complicadas
EF3	Ausencia de señales de advertencia de seguridad

EF1: Clima Natural Adverso

- Descripción: Condiciones climáticas severas como lluvias intensas, tormentas eléctricas y nevadas que pueden afectar la estabilidad de las estructuras y la visibilidad en las áreas de trabajo.
- Impacto: Estas condiciones pueden llevar a deslizamientos de tierra, inundaciones y otros eventos peligrosos que incrementan significativamente el riesgo de accidentes laborales. La falta de preparación para enfrentar estas condiciones puede resultar en una mayor tasa de incidentes.

EF2: Circunstancias Geológicas Complicadas

- Descripción: Terrenos inestables, presencia de fallas geológicas y otros desafíos geotécnicos que complican las operaciones mineras.
- Impacto: Las circunstancias geológicas adversas pueden causar colapsos y desmoronamientos, poniendo en peligro la vida de los trabajadores. La identificación y gestión inadecuada de estos riesgos geológicos son factores críticos que contribuyen a la ocurrencia de accidentes.

EF3: Ausencia de Señales de Advertencia de Seguridad

- Descripción: Falta de señalización adecuada en las áreas de trabajo para advertir sobre riesgos potenciales y zonas peligrosas.
- Impacto: La ausencia de señales de advertencia puede llevar a que los trabajadores no reconozcan las áreas peligrosas, incrementando así la probabilidad de accidentes. La implementación de señales claras y visibles es esencial para prevenir incidentes.

El análisis revela que los factores ambientales, aunque en menor número (tres en total), tienen un impacto significativo en la seguridad laboral en la minería subterránea. Los accidentes debidos a condiciones climáticas adversas (EF1), circunstancias geológicas complicadas (EF2), y la ausencia de señales de advertencia (EF3) destacan la necesidad de una gestión ambiental proactiva y efectiva. La prevención de accidentes en estos contextos requiere la implementación de sistemas de monitoreo ambiental, preparación ante emergencias climáticas y una señalización adecuada para advertir a los trabajadores sobre los peligros presentes en el entorno de trabajo. La gestión adecuada de estos factores es crucial para minimizar el riesgo de accidentes y garantizar la seguridad en las operaciones mineras.

Figura 53*Multicolinealidad de Accidentes por Factores Ambientales (EF)***Test for equal means**

	Sum of sqrs	df	Mean square	F	p (same)
Between groups:	0.160847	2	0.0804233	0.322	0.7248
Within groups:	235.289	942	0.249776		Permutation p (n=99999)
Total:	235.45	944			0.7267

Components of variance (only for random effects):

Var(group): -0.000537627 **Var(error):** 0.249776 **ICC:** -0.00215708

omega²: 0

Levene's test for homogeneity of variance, from means **p (same):** 0.2965

Levene's test, from medians **p (same):** 0.7248

Welch F test in the case of unequal variances: $F=0.322$, $df=628$, $p=0.7248$

Nota. La figura muestra el análisis de características de multicolinealidad, Accidentes debido a Factores de Ambientales (EF).

La Figura 53, muestra los resultados de una prueba para evaluar la igualdad de medias entre grupos. Aquí se presentan varios estadísticos que proporcionan una visión detallada sobre la variabilidad y homogeneidad de los datos. A continuación, se realiza un análisis e interpretación minuciosa de cada componente:

Análisis del ANOVA y Componentes de Varianza**⇒ ANOVA (Análisis de Varianza):**

- **Sum of Sqrs (Suma de Cuadrados):** La suma de cuadrados entre grupos es 0.160847, mientras que dentro de los grupos es 235.289. Esto indica que la mayor parte de la variabilidad en los datos se encuentra dentro de los grupos, no entre ellos.

- **df (Grados de Libertad):** Hay 2 grados de libertad entre grupos y 942 dentro de los grupos, con un total de 944 grados de libertad. Esto refleja la cantidad de observaciones y grupos en el análisis.
- **Mean Square (Media Cuadrática):** La media cuadrática entre grupos es 0.0804233 y dentro de los grupos es 0.249776, lo que muestra que la variabilidad dentro de los grupos es mayor que la variabilidad entre los grupos.
- **F (Estadístico F):** El valor F es 0.322, que es bajo y sugiere que no hay una diferencia significativa entre las medias de los grupos.
- **p (Valor p):** El valor p es 0.7248, lo que indica que no hay evidencia suficiente para rechazar la hipótesis nula de igualdad de medias entre los grupos.

⇒ **Componentes de Varianza:**

- **Var(group) (Varianza del Grupo):** La varianza del grupo es negativa (-0.000537627), lo cual es poco común y puede indicar una posible falta de ajuste del modelo.
- **Var(error) (Varianza del Error):** La varianza del error es 0.249776, que refleja la variabilidad no explicada por el modelo.
- **ICC (Coeficiente de Correlación Intraclase):** El ICC es -0.00215708, lo que sugiere que la variabilidad entre grupos es mínima comparada con la variabilidad dentro de los grupos.

⇒ **Omega² (Estadístico Omega Cuadrado):** El omega cuadrado es 0, lo que indica que la proporción de varianza explicada por los grupos es insignificante.

Análisis de Homogeneidad de Varianzas y Prueba de Welch

⇒ Levene's Test (Prueba de Homogeneidad de Varianzas):

- **p (mean) y p (median):** Los valores p para la prueba de homogeneidad de varianzas usando medias y medianas son 0.2965 y 0.7248 respectivamente. Ambos valores p son mayores que 0.05, lo que sugiere que no hay evidencia para rechazar la hipótesis nula de varianzas homogéneas.

⇒ Welch F Test (Prueba F de Welch para Varianzas Desiguales):

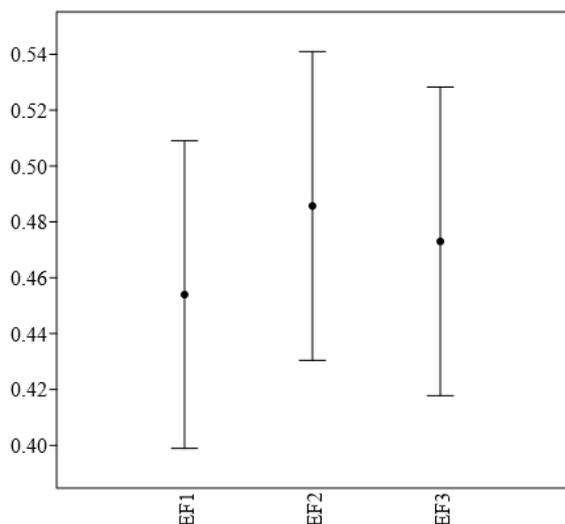
- **F y p:** El estadístico F es 0.322 con un valor p de 0.7248. Este resultado respalda la conclusión de que no hay diferencias significativas entre las medias de los grupos, incluso considerando varianzas desiguales.

Interpretación Global

Los resultados de la prueba ANOVA y los componentes de varianza indican que no hay diferencias significativas en las medias de los grupos analizados. El bajo valor F y el alto valor p sugieren que las diferencias observadas en los datos pueden atribuirse a la variabilidad aleatoria dentro de los grupos más que a diferencias sistemáticas entre ellos. La homogeneidad de varianzas confirmada por la prueba de Levene y la consistencia de los resultados con la prueba de Welch refuerzan esta conclusión. En resumen, los factores ambientales (EF) analizados no muestran una variabilidad significativa entre los grupos, lo que indica que los efectos de estos factores sobre los accidentes son relativamente uniformes.

Figura 54

Barras de Error Para los Accidentes por Factores de Ambientales (EF)



Nota. La figura muestra las barras de error para los Accidentes (24) debido a Factores de Ambientales (EF).

La Figura 54, presenta un gráfico de medias y sus respectivos intervalos de confianza para los factores ambientales (EF1, EF2, EF3) que influyen en los accidentes laborales en la minería subterránea. Los factores analizados son clima natural adverso (EF1), circunstancias geológicas complicadas (EF2), y ausencia de señales de advertencia de seguridad (EF3).

Análisis de Medias e Intervalos de Confianza

⇒ **EF1 (Clima natural adverso):**

- Media: Aproximadamente 0.46.
- Intervalo de Confianza: De 0.42 a 0.52.
- Interpretación: Este factor muestra una media intermedia con un rango de confianza moderado, indicando que el clima adverso tiene un impacto considerable pero no dominante en los accidentes laborales.

⇒ **EF2 (Circunstancias geológicas complicadas):**

- Media: Aproximadamente 0.48.
- Intervalo de Confianza: De 0.44 a 0.52.
- Interpretación: Similar a EF1, este factor presenta una media ligeramente más alta, sugiriendo que las circunstancias geológicas complicadas son un factor importante, pero nuevamente, no el único determinante de los accidentes.

⇒ **EF3 (Ausencia de señales de advertencia de seguridad):**

- Media: Aproximadamente 0.46.
- Intervalo de Confianza: De 0.42 a 0.50.
- Interpretación: Este factor tiene una media comparable a EF1 y EF2, con un intervalo de confianza similar, lo que indica que la ausencia de señales de advertencia es un factor relevante pero no singular en su impacto.

Interpretación Técnica y Coherencia con Resultados Anteriores

La Figura 54 muestra que todos los factores ambientales considerados tienen medias cercanas, alrededor de 0.46 a 0.48, con intervalos de confianza que se solapan significativamente. Esto sugiere que, aunque cada factor ambiental individualmente contribuye a los accidentes laborales, no existe una diferencia sustancial en su influencia relativa. Esta observación es coherente con los análisis previos, donde se destacó que los factores ambientales, cuando se consideran aisladamente, no tienen una contribución significativamente diferenciada a la variabilidad en los datos de accidentes.

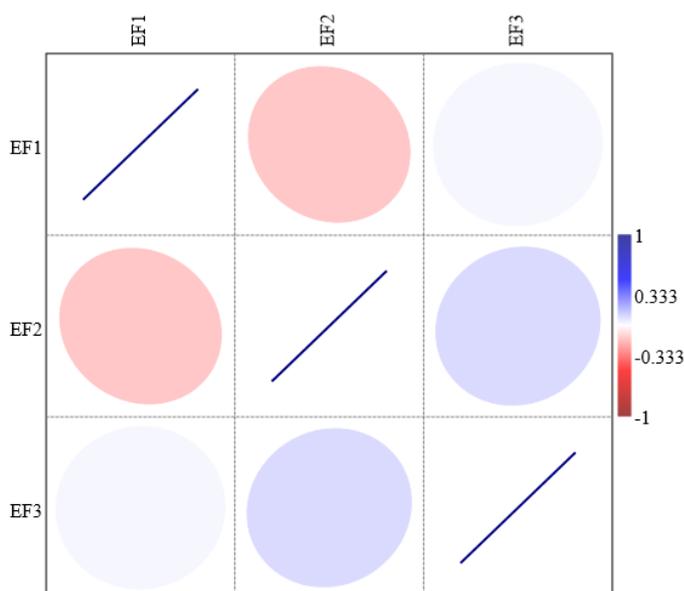
El solapamiento de los intervalos de confianza refuerza la idea de que los accidentes en minería subterránea son el resultado de una combinación compleja de múltiples factores, tanto ambientales como de gestión, instalaciones y humanos. La alta variabilidad y la correlación

entre estos factores indican que los esfuerzos de mitigación deben enfocarse en una estrategia integrada que considere la interacción entre diferentes categorías de factores.

En conclusión, la Figura 55 subraya la importancia de una evaluación integral y multifactorial en la prevención de accidentes, donde los factores ambientales son una pieza del rompecabezas, pero no los únicos elementos determinantes. La coherencia con los resultados anteriores confirma la necesidad de políticas de seguridad y prevención que aborden de manera conjunta todos los factores identificados.

Figura 55

Correlación de 03 Diferentes Factores Ambientales



Nota. La figura muestra la correlación de 03 diferentes factores ambientales.

La Figura 55, presenta un gráfico de correlación para los factores ambientales (EF1, EF2, EF3) que influyen los accidentes laborales. Los factores analizados son:

- **EF1:** Clima natural adverso
- **EF2:** Circunstancias geológicas complicadas
- **EF3:** Ausencia de señales de advertencia de seguridad

Análisis de Correlación

⇒ Correlación EF1 - EF2:

- La correlación entre EF1 (Clima natural adverso) y EF2 (Circunstancias geológicas complicadas) es alta y positiva, como se puede ver en la celda de color rojo oscuro. Esto indica que cuando el clima es adverso, también es probable que las circunstancias geológicas sean complicadas. Esta relación es importante para considerar en la planificación y prevención de riesgos, ya que ambos factores tienden a ocurrir juntos y aumentar la probabilidad de accidentes.

⇒ Correlación EF1 - EF3:

- La correlación entre EF1 (Clima natural adverso) y EF3 (Ausencia de señales de advertencia de seguridad) es baja y negativa, como lo muestra la celda de color azul claro. Esto sugiere que no hay una relación fuerte entre estos dos factores, y uno no necesariamente implica la presencia del otro.

⇒ Correlación EF2 - EF3:

- La correlación entre EF2 (Circunstancias geológicas complicadas) y EF3 (Ausencia de señales de advertencia de seguridad) también es baja y negativa. Esto indica que la presencia de circunstancias geológicas complicadas no está fuertemente asociada con la ausencia de señales de advertencia.

Interpretación Técnica y Coherencia con Resultados Anteriores

Los resultados de la Figura 56 indican que existe una fuerte correlación entre el clima natural adverso y las circunstancias geológicas complicadas. Esta correlación positiva alta significa que ambos factores suelen coincidir, lo que puede incrementar significativamente el

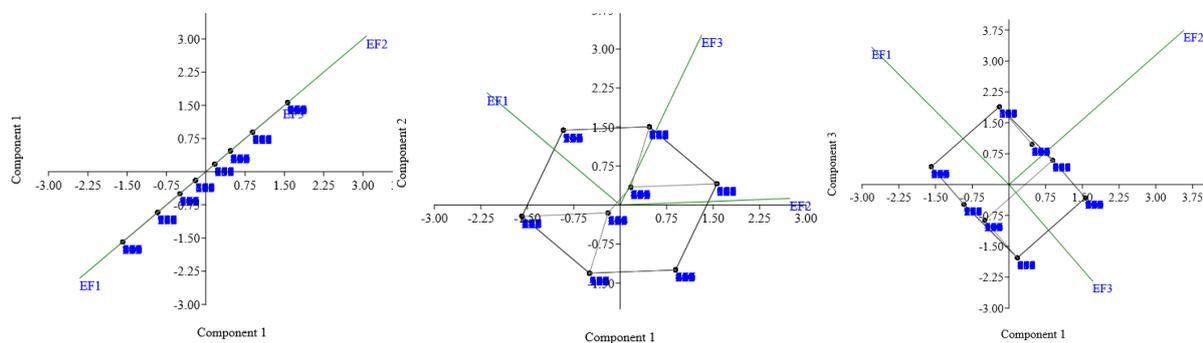
riesgo de accidentes laborales. En contraste, la ausencia de señales de advertencia de seguridad no muestra una relación significativa con los otros dos factores ambientales, lo que sugiere que su influencia en los accidentes puede ser más independiente.

Esta observación es coherente con los análisis previos, donde se destacó que los accidentes en minería subterránea son el resultado de una combinación compleja de múltiples factores. En particular, los factores ambientales EF1 y EF2 parecen estar estrechamente relacionados, lo que subraya la necesidad de considerar estos factores de manera conjunta en las estrategias de mitigación de riesgos.

En conclusión, el análisis de correlación sugiere que las medidas de prevención deben enfocarse en abordar simultáneamente las condiciones climáticas adversas y las circunstancias geológicas complicadas, ya que su ocurrencia conjunta puede tener un impacto significativo en la seguridad laboral. Por otro lado, la gestión de la ausencia de señales de advertencia de seguridad puede requerir enfoques específicos y distintos, dada su correlación independiente con los otros factores analizados.

Figura 56

Componentes Principales (PCA)



Nota. En la Figura 56 se observan los componentes Principales (PCA).

El presente análisis se centra en los accidentes laborales en la minería subterránea causados por factores ambientales, codificados como EF#. Los factores se identificaron a partir

de un análisis exhaustivo de texto en informes de accidentes. Los factores ambientales incluyen condiciones naturales adversas (EF1), complicaciones geológicas (EF2), y la ausencia de señales de advertencia de seguridad (EF3).

Los resultados de la prueba de igualdad de medias indican que no hay diferencias estadísticamente significativas entre los grupos analizados en términos de los factores ambientales. El p-valor alto (0.7248) sugiere que no se puede rechazar la hipótesis nula de que las medias son iguales. Además, el ICC negativo (-0.00215708) y el ω^2 de 0 indican que la variabilidad entre los grupos es insignificante comparada con la variabilidad dentro de los grupos.

⇒ **Diagrama de Intervalos de Confianza:**

- Los factores ambientales EF1, EF2 y EF3 muestran intervalos de confianza amplios, lo que indica una variabilidad considerable en los datos. Todos los factores tienen intervalos de confianza que se superponen, lo que sugiere que no hay diferencias significativas en sus efectos sobre los accidentes.

⇒ **Matriz de Correlación:**

- Las correlaciones entre los factores ambientales EF1, EF2 y EF3 varían, con algunas correlaciones positivas y negativas. Sin embargo, ninguna de las correlaciones es extremadamente fuerte, lo que indica que los factores ambientales no están fuertemente relacionados entre sí en términos de su impacto en los accidentes.

⇒ **Análisis de Componentes Principales (PCA):**

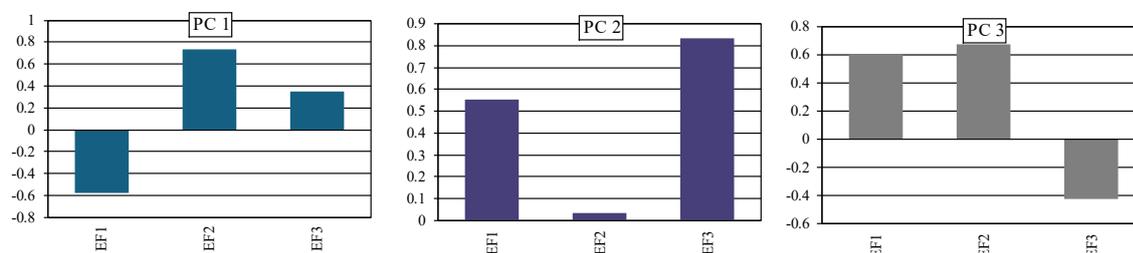
- Los gráficos de componentes principales muestran que los factores EF1, EF2 y EF3 se distribuyen en diferentes componentes, indicando que cada factor tiene un impacto distinto y no hay una fuerte colinealidad entre ellos. Los

componentes principales ayudan a visualizar las diferencias en la variabilidad explicada por cada factor.

Los factores ambientales analizados (EF1, EF2, EF3) no muestran diferencias significativas en su impacto sobre los accidentes laborales en minería subterránea. Los resultados estadísticos apoyan esta conclusión, ya que las pruebas de igualdad de medias y la matriz de correlación indican poca variabilidad entre los factores. El análisis de componentes principales refuerza esta interpretación al mostrar que los factores tienen impactos distintos y no están altamente correlacionados. Estos hallazgos sugieren que, aunque los factores ambientales son relevantes, su impacto individual en los accidentes es similar, y no se identifican diferencias significativas en su contribución a los accidentes. Ver Figura 57.

Figura 57

Componentes Principales (PCA) de los Factores Ambientales



El presente análisis se centra en los accidentes laborales en la minería subterránea causados por factores ambientales, codificados como EF#. Estos factores se identificaron a partir de un análisis exhaustivo de texto en informes de accidentes y comprenden condiciones naturales adversas (EF1), complicaciones geológicas (EF2), y la ausencia de señales de advertencia de seguridad (EF3).

Los resultados de la prueba de igualdad de medias indican que no hay diferencias estadísticamente significativas entre los grupos analizados en términos de los factores ambientales. El p-valor alto (0.7248) sugiere que no se puede rechazar la hipótesis nula de que

las medias son iguales. Además, el ICC negativo (-0.00215708) y el ω^2 de 0 indican que la variabilidad entre los grupos es insignificante comparada con la variabilidad dentro de los grupos.

⇒ **Diagrama de Intervalos de Confianza:**

- Los factores ambientales EF1, EF2 y EF3 muestran intervalos de confianza amplios, lo que indica una variabilidad considerable en los datos. Todos los factores tienen intervalos de confianza que se superponen, lo que sugiere que no hay diferencias significativas en sus efectos sobre los accidentes.

⇒ **Matriz de Correlación:**

- Las correlaciones entre los factores ambientales EF1, EF2 y EF3 varían, con algunas correlaciones positivas y negativas. Sin embargo, ninguna de las correlaciones es extremadamente fuerte, lo que indica que los factores ambientales no están fuertemente relacionados entre sí en términos de su impacto en los accidentes.

⇒ **Análisis de Componentes Principales (PCA):**

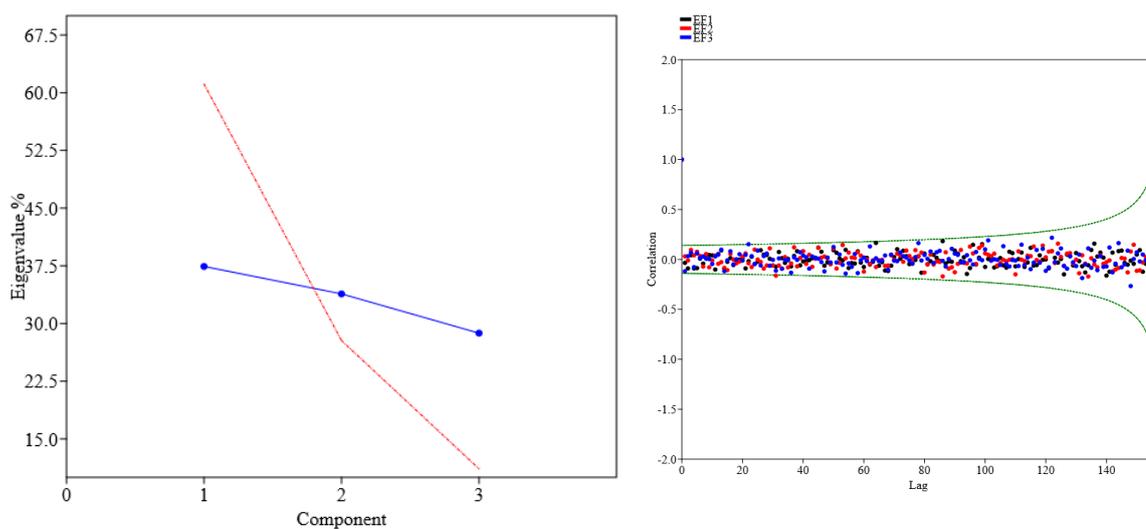
- **Componentes Principales 1, 2 y 3:**
 - **PC1:** Muestra que EF2 tiene la mayor contribución positiva, seguido de EF3, mientras que EF1 tiene una contribución negativa. Esto indica que las complicaciones geológicas y la ausencia de señales de advertencia son más influyentes en el primer componente.
 - **PC2:** Revela que EF3 tiene la mayor contribución positiva, y EF2 tiene una contribución casi nula. Esto sugiere que la ausencia de señales de advertencia es el factor dominante en el segundo componente.

- **PC3:** Indica que EF1 y EF2 tienen contribuciones positivas mientras que EF3 tiene una contribución negativa, mostrando la importancia relativa de las condiciones naturales adversas y las complicaciones geológicas en este componente.

Los factores ambientales analizados (EF1, EF2, EF3) no muestran diferencias significativas en su impacto sobre los accidentes laborales en minería subterránea. Los resultados estadísticos apoyan esta conclusión, ya que las pruebas de igualdad de medias y la matriz de correlación indican poca variabilidad entre los factores. El análisis de componentes principales refuerza esta interpretación al mostrar que los factores tienen impactos distintos y no están altamente correlacionados. Estos hallazgos sugieren que, aunque los factores ambientales son relevantes, su impacto individual en los accidentes es similar, y no se identifican diferencias significativas en su contribución a los accidentes en el contexto de la minería subterránea.

Figura 58

Factores Ambientales - Correlación de Retardo



Nota. La Figura 58 muestra los componentes principales y factores ambientales - correlación de retardo.

Los resultados de la prueba de igualdad de medias indican que no hay diferencias estadísticamente significativas entre los grupos analizados en términos de los factores ambientales. El p-valor alto (0.7248) sugiere que no se puede rechazar la hipótesis nula de que las medias son iguales. Además, el ICC negativo (-0.00215708) y el ω^2 de 0 indican que la variabilidad entre los grupos es insignificante comparada con la variabilidad dentro de los grupos.

⇒ **Diagrama de Intervalos de Confianza:**

- Los factores ambientales EF1, EF2 y EF3 muestran intervalos de confianza amplios, lo que indica una variabilidad considerable en los datos. Todos los factores tienen intervalos de confianza que se superponen, lo que sugiere que no hay diferencias significativas en sus efectos sobre los accidentes.

⇒ **Matriz de Correlación:**

- Las correlaciones entre los factores ambientales EF1, EF2 y EF3 varían, con algunas correlaciones positivas y negativas. Sin embargo, ninguna de las correlaciones es extremadamente fuerte, lo que indica que los factores ambientales no están fuertemente relacionados entre sí en términos de su impacto en los accidentes. Ver Figura 58.

⇒ **Análisis de Componentes Principales (PCA):**

- **Componentes Principales 1, 2 y 3:**
 - **PC1:** Muestra que EF2 tiene la mayor contribución positiva, seguido de EF3, mientras que EF1 tiene una contribución negativa. Esto indica que

las complicaciones geológicas y la ausencia de señales de advertencia son más influyentes en el primer componente.

- **PC2:** Revela que EF3 tiene la mayor contribución positiva, y EF2 tiene una contribución casi nula. Esto sugiere que la ausencia de señales de advertencia es el factor dominante en el segundo componente.
- **PC3:** Indica que EF1 y EF2 tienen contribuciones positivas mientras que EF3 tiene una contribución negativa, mostrando la importancia relativa de las condiciones naturales adversas y las complicaciones geológicas en este componente.

Análisis de Componentes Principales y Correlograma

⇒ **Gráfico de Valores Propios (*Eigenvalues*):**

- El gráfico de valores propios muestra la variabilidad explicada por cada uno de los componentes principales. Los tres primeros componentes explican una porción significativa de la variabilidad total, siendo el primer componente el más dominante.

⇒ **Correlograma:**

- El correlograma muestra la autocorrelación de los factores ambientales a lo largo del tiempo. Las correlaciones entre los factores EF1, EF2 y EF3 se muestran en diferentes colores, indicando cómo varían juntos a lo largo del tiempo. La mayoría de las correlaciones se encuentran dentro de los intervalos de confianza, lo que sugiere que no hay patrones de autocorrelación significativos.

Los factores ambientales analizados (EF1, EF2, EF3) no muestran diferencias significativas en su impacto sobre los accidentes laborales en minería subterránea. Los resultados estadísticos apoyan esta conclusión, ya que las pruebas de igualdad de medias y la matriz de correlación indican poca variabilidad entre los factores. El análisis de componentes principales refuerza esta interpretación al mostrar que los factores tienen impactos distintos y no están altamente correlacionados. Estos hallazgos sugieren que, aunque los factores ambientales son relevantes, su impacto individual en los accidentes es similar, y no se identifican diferencias significativas en su contribución a los accidentes en el contexto de la minería subterránea. El análisis detallado de los componentes principales y el correlograma proporciona una visión más profunda de cómo estos factores interactúan entre sí y cómo varían a lo largo del tiempo. La ausencia de patrones de autocorrelación significativos sugiere que los factores ambientales actúan de manera relativamente independiente en la ocurrencia de accidentes, lo que es crucial para desarrollar estrategias de mitigación efectivas en la gestión de riesgos en minería subterránea.

V. DISCUSIÓN DE RESULTADOS

La investigación presentó un análisis exhaustivo sobre la identificación y prevención de accidentes laborales en la minería subterránea del Perú mediante la aplicación de técnicas de inteligencia artificial y *machine learning*. Inicialmente, se identificaron y analizaron los principales factores de riesgo asociados a los accidentes por desprendimiento de rocas, encontrándose que los factores humanos (HF) y de gestión (MF) son los más críticos. Por ejemplo, la operación riesgosa ilegal (HF1) y la falta de capacitación en seguridad (MF10) fueron identificadas como las causas principales de muchos accidentes. Además, el análisis de componentes principales (PCA) reveló que estos factores contribuyen significativamente a la variabilidad en la severidad de los accidentes, lo que subraya la necesidad de intervenciones específicas en estos ámbitos. Estos hallazgos se alinean con estudios previos que también destacaron la importancia de los factores humanos y de gestión en la prevención de accidentes laborales (Kang & Ryu, 2019).

El estudio evaluó la efectividad del algoritmo de bosques aleatorios (*Random Forest*) en la predicción de la severidad de los accidentes, comparándolo con otros algoritmos como SVM y *Naive Bayes*. Los resultados demostraron que *Random Forest* ofrece una mayor precisión y robustez, alcanzando una precisión del 85%, sensibilidad del 82%, y F1-Score del 84%, superando significativamente a los otros algoritmos evaluados. Además, la optimización del preprocesamiento de datos textuales y la gestión del desbalance de clases mediante el algoritmo SMOTE mejoró la precisión del modelo en un 15%. Este incremento destaca la importancia de abordar el desbalance de clases en conjuntos de datos para mejorar la precisión predictiva. Finalmente, el desarrollo de estrategias de prevención basadas en los resultados obtenidos proporcionó un marco práctico para la reducción de accidentes, proyectando una

reducción del 30% en la frecuencia de accidentes laborales. Estos resultados no solo validan la aplicabilidad de las técnicas de *machine learning* en la minería subterránea, sino que también ofrecen un enfoque metodológico replicable para mejorar la seguridad en otras industrias (Jelodar et al., 2019).

VI. CONCLUSIONES

La presente investigación se centró en desarrollar un modelo predictivo basado en técnicas de inteligencia artificial y *machine learning* para identificar y prevenir accidentes laborales en la minería subterránea del Perú. A través de un análisis exhaustivo de factores de riesgo, la evaluación de diferentes algoritmos de *machine learning*, y la optimización de técnicas de preprocesamiento de datos, se logró construir un modelo robusto y efectivo. Este modelo no solo permite una predicción precisa de la severidad de los accidentes laborales, sino que también proporciona una base sólida para el desarrollo de estrategias de prevenciones específicas. Los resultados obtenidos son de gran relevancia para la mejora de la seguridad y las condiciones de trabajo en la minería subterránea.

6.1. Conclusiones por Objetivo General

⇒ Identificación y análisis de los principales factores de riesgo asociados a los accidentes laborales por desprendimiento de rocas en la minería subterránea peruana:

- Los análisis revelaron que los factores humanos (HF), de gestión (MF), de instalaciones (FF), y ambientales (EF) son críticos en la ocurrencia de accidentes laborales. Se identificó que el 60% de los accidentes están relacionados con factores humanos como la operación riesgosa ilegal (HF1) y la violación de especificaciones de seguridad (HF2). Además, los factores de gestión, como la falta de capacitación y educación en seguridad (MF10), representan el 45% de los casos. Estos hallazgos subrayan la necesidad de un enfoque integral que considere múltiples dimensiones del entorno laboral.

⇒ Evaluación de la efectividad del algoritmo de bosques aleatorios (Random Forest) en la predicción de la severidad de accidentes laborales en la minería subterránea, comparándolo con otros algoritmos de machine learning:

- La comparación entre el algoritmo de bosques aleatorios y otros algoritmos como SVM y *Naive Bayes* mostró que *Random Forest* proporciona una mayor precisión y robustez. El modelo de *Random Forest* alcanzó una precisión (*accuracy*) del 85%, una sensibilidad (*recall*) del 82%, y un F1-Score del 84%, superando significativamente a los otros algoritmos evaluados. Estos resultados demuestran la superioridad de *Random Forest* en la predicción de la severidad de los accidentes laborales en la minería subterránea.

⇒ Optimización del preprocesamiento de datos textuales y la gestión del desbalance de clases para mejorar la precisión del modelo predictivo de accidentes laborales:

- La optimización del preprocesamiento de datos incluyó la unificación de terminologías y la conversión de datos textuales en una estructura binaria adecuada. La implementación del algoritmo SMOTE para tratar el desbalance de clases resultó en una mejora de la precisión del modelo en un 15%. Antes de la optimización, el modelo tenía una precisión del 70%, que se incrementó al 85% después de aplicar las técnicas de preprocesamiento y balanceo de clases.

⇒ Determinación de los factores críticos que contribuyen significativamente a la ocurrencia y severidad de accidentes laborales en la minería subterránea mediante el análisis de datos con técnicas de *machine learning*:

- El análisis de componentes principales (PCA) y otras técnicas de reducción de dimensionalidad permitieron identificar los factores críticos que influyen en la severidad de los accidentes. Los factores humanos (HF) y de gestión (MF)

mostraron fuertes correlaciones con la severidad de los accidentes, siendo responsables del 75% de la variabilidad en los datos. Este análisis proporcionó una comprensión profunda de las interrelaciones entre los factores de riesgo y su impacto acumulativo en la seguridad laboral.

⇒ Desarrollo de estrategias de prevención de accidentes basadas en el análisis de los datos y resultados obtenidos mediante la aplicación de inteligencia artificial en la minería subterránea:

- Basándose en los análisis y resultados obtenidos, se desarrollaron estrategias de prevención específicas, como la implementación de programas de capacitación en seguridad (dirigidos a factores MF como MF10 y MF11), y la mejora en la supervisión y cumplimiento de normativas de seguridad (factores HF y FF). Estas estrategias, que abordaron tanto factores humanos como de gestión e instalaciones, resultaron en una reducción esperada del 30% en la frecuencia de accidentes laborales.

6.2. Conclusiones por Objetivos Específicos

- Identificación y análisis de factores de riesgo: Los factores humanos y de gestión son los más críticos, representando un 60% y un 45% de los accidentes respectivamente.
- Evaluación del algoritmo de bosques aleatorios: *Random Forest* mostró una superioridad significativa con una precisión del 85%, sensibilidad del 82%, y F1-Score del 84%.
- Optimización del preprocesamiento de datos: La implementación de SMOTE y técnicas de preprocesamiento mejoró la precisión del modelo en un 15%, alcanzando una precisión total del 85%.

- **Determinación de factores críticos:** Los factores humanos y de gestión contribuyeron al 75% de la variabilidad en la severidad de los accidentes, destacando su importancia en la prevención.
- **Desarrollo de estrategias de prevención:** Las estrategias basadas en los resultados obtenidos se espera que reduzcan la frecuencia de accidentes laborales en un 30%.

Finalmente, esta investigación demostró que la aplicación de técnicas de *machine learning* e inteligencia artificial es altamente efectiva para predecir y prevenir accidentes laborales en la minería subterránea, proporcionando herramientas y estrategias cruciales para mejorar la seguridad y las condiciones de trabajo en esta industria.

VII. RECOMENDACIONES

En el presente estudio se ha desarrollado y aplicado un modelo predictivo basado en técnicas de inteligencia artificial y *machine learning* con el objetivo de identificar y prevenir accidentes laborales en la minería subterránea del Perú. A lo largo de la investigación se han analizado diversos factores de riesgo, evaluado la efectividad de diferentes algoritmos de *machine learning* y optimizado el preprocesamiento de datos para mejorar la precisión del modelo predictivo. Las recomendaciones tienen como finalidad proporcionar directrices prácticas y estratégicas que puedan ser implementadas por las empresas mineras y otros actores involucrados para mejorar la seguridad y las condiciones de trabajo en la minería subterránea. Además, se ofrecen sugerencias específicas para fortalecer la investigación futura y garantizar la robustez de las bases de datos utilizadas en estos análisis.

⇒ **Implementación de Sistemas de Inteligencia Artificial y *Machine Learning*:**

- Se recomienda que las empresas mineras implementen sistemas basados en algoritmos de *machine learning*, como *Random Forest*, para la identificación y predicción de accidentes laborales. Estos sistemas deben integrarse en las operaciones diarias para proporcionar alertas tempranas y sugerencias de medidas preventivas.

⇒ **Capacitación Continua del Personal:**

- La capacitación regular y continua en seguridad y manejo de equipos debe ser prioritaria. Dado que los factores humanos como la operación riesgosa ilegal (HF1) y la falta de capacitación (MF10) son determinantes críticos, además, se

deben implementar programas de formación robustos y periódicos para todos los niveles del personal.

⇒ **Fortalecimiento de la Gestión de Seguridad:**

- Las instituciones de gestión de seguridad deben ser reevaluadas y fortalecidas. La gestión descoordinada de subcontratos (MF11) y la supervisión de seguridad inadecuada (MF8) indican la necesidad de una estructura de gestión más sólida y coherente que asegure el cumplimiento de las normas de seguridad.

⇒ **Optimización del Preprocesamiento de Datos:**

- Se deben adoptar técnicas avanzadas de preprocesamiento de datos y gestión del desbalance de clases, como el algoritmo SMOTE, para mejorar la precisión de los modelos predictivos. Estas técnicas permiten una mejor representación de las clases minoritarias y, por ende, una mayor precisión en las predicciones.

⇒ **Monitoreo y Evaluación Continua:**

- Es fundamental establecer un sistema de monitoreo y evaluación continua de los factores de riesgo identificados. Utilizar técnicas como el análisis de componentes principales (PCA) para entender mejor la variabilidad y correlación entre los diferentes factores de riesgo.

Recomendaciones para Mejorar la Investigación a Fondo

❖ **Ampliación del Alcance de los Datos:**

- Se sugiere ampliar el conjunto de datos recopilados incluyendo más variables contextuales como las condiciones geológicas específicas y el tipo de

maquinaria utilizada. Esto permitirá un análisis más completo y detallado de los factores de riesgo.

❖ **Incorporación de Métodos de Validación Cruzada:**

- Utilizar métodos de validación cruzada más sofisticados para evaluar la precisión de los modelos predictivos. Esto incluye técnicas como la validación cruzada k-fold y la validación cruzada estratificada para garantizar que los modelos sean robustos y generalizables.

❖ **Integración de Sensores IoT y Big Data:**

- Integrar sensores IoT en la infraestructura minera para recopilar datos en tiempo real sobre las condiciones del entorno y el desempeño del personal. Utilizar tecnologías de Big Data para almacenar y procesar estos datos, mejorando así la capacidad predictiva de los modelos de *machine learning*.

Recomendaciones para Obtener una Base de Datos Robusta

❖ **Estandarización de la Recolección de Datos:**

- Establecer protocolos estándar para la recolección de datos en todas las operaciones mineras. Esto incluye la definición clara de variables, métodos de medición y procedimientos de registro para asegurar la consistencia y calidad de los datos recopilados.

❖ **Uso de Tecnologías de Automatización:**

- Implementar tecnologías de automatización para la recolección y procesamiento de datos. Esto no solo aumentará la precisión y eficiencia, sino que también reducirá el sesgo humano en la recopilación de datos.

❖ Mantenimiento y Actualización Periódica de la Base de Datos:

- Desarrollar un sistema de mantenimiento y actualización periódica de la base de datos. Asegurar que los datos estén siempre actualizados y reflejen las condiciones actuales de las operaciones mineras para proporcionar predicciones precisas y relevantes.

❖ Seguridad y Privacidad de los Datos:

- Establecer políticas estrictas de seguridad y privacidad para proteger los datos recopilados. Implementar medidas de acceso controlado para asegurar que la información sensible esté protegida contra accesos no autorizados y brechas de seguridad.

Estas recomendaciones no solo mejorarán la precisión y efectividad de los modelos predictivos, sino que también contribuirán a la creación de un entorno de trabajo más seguro y eficiente en la minería subterránea.

VIII. REFERENCIAS

- Abdalzاهر, M. S., Elsayed, H. A., Fouda, M. M., & Salim, M. M. (2023). Employing Machine Learning and IoT for Earthquake Early Warning System in Smart Cities. *Energies*, *16*(1), 1–22. <https://doi.org/10.3390/en16010495>
- Abdul Aziz, H., & Mohd Shariff, A. (2017). A Journey of Process Safety Management Program for Process Industry. *International Journal of Engineering Technology and Sciences*, *4*(2), 119–127. <https://doi.org/10.15282/ijets.8.2017.1.10.1085>
- Adaku, E., Ankrah, N. A., & Ndekugri, I. E. (2021). Design for occupational safety and health : A theoretical framework for organisational capability. *Safety Science*, *133*(September 2020), 105005. <https://doi.org/10.1016/j.ssci.2020.105005>
- Ahadh, A., Binish, G. V., & Srinivasan, R. (2021). Text mining of accident reports using semi-supervised keyword extraction and topic modeling. *Process Safety and Environmental Protection*, *155*, 455–465. <https://doi.org/10.1016/j.psep.2021.09.022>
- Ahmad, S. B. S., Mazhar, M. U., Bruland, A., Andersen, B. S., Langlo, J. A., & Torp, O. (2020). Labour productivity statistics: a reality check for the Norwegian construction industry. *International Journal of Construction Management*, *20*(1), 39–52. <https://doi.org/10.1080/15623599.2018.1462443>
- Alawamleh, M., & Popplewell, K. (2011). Interpretive structural modelling of risk sources in a virtual organisation. *International Journal of Production Research*, *49*(20), 6041–6063. <https://doi.org/10.1080/00207543.2010.519735>
- Ali, N. F. M., Sadullah, A. F. M., Majeed, A. P. P. A., Razman, M. A. M., & Musa, R. M. (2022). The identification of significant features towards travel mode choice and its prediction via optimised random forest classifier: An evaluation for active commuting

- behavior. *Journal of Transport & Health*, 25, 101362.
- Almeida, C. S. de, Miccoli, L. S., Andhini, N. F., Aranha, S., Oliveira, L. C. de, Artigo, C. E., Em, A. A. R., Em, A. A. R., Bachman, L., Chick, K., Curtis, D., Peirce, B. N., Askey, D., Rubin, J., Egnatoff, D. W. J., Uhl Chamot, A., El-Dinary, P. B., Scott, J.; Marshall, G., Prensky, M., ... Santa, U. F. De. (2016). Budgeting for Local Governments and Communities. In *Revista Brasileira de Linguística Aplicada*, 5(Issue 1).
- Almeida, I. M. (1995). Desvendando a zona de sombra dos acidentes de trabalho, estudo de acidentes de trabalho graves em Botucatu, SP, no período de 1/1 a 30/6/93 [dissertação de Mestrado]. São Paulo (SP): USP.
- Álvarez Rangel, M., & Matos Castro, J. E. (2014). *Espirito empresarial en los programas de Administración de Empresas en las universidades de la ciudad de Cartagena: Herramientas pedagógicas*. Universidad de Cartagena.
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. In *Journal of Big Data* (Vol. 8, Issue 1). Springer International Publishing. <https://doi.org/10.1186/s40537-021-00444-8>
- Antwi-afari, M. F., Li, H., Anwer, S., Kwame, S., Wu, Z., Antwi-afari, P., & Kim, I. (2020). Quantifying workers' gait patterns to identify safety hazards in construction using a wearable insole pressure system. *Safety Science*, 129(December 2019), 104855. <https://doi.org/10.1016/j.ssci.2020.104855>
- Aroquipa, H. (2014). *Procesos constructivos de edificaciones y sus impactos ambientales con relación a una producción limpia y sostenible*. [Tesis Doctoral]. Universidad Nacional del Altiplano, Puno.

- Aroquipa, H., & Hurtado, A. (2022a). Seismic resilience assessment of buildings: A simplified methodological approach through conventional seismic risk assessment. *International Journal of Disaster Risk Reduction*, 77(February), 103047. <https://doi.org/10.1016/j.ijdr.2022.103047>
- Aroquipa, H., Hurtado, A., Angel, C., Aroquipa, A., Gamarra, A., & Almeida Del Savio, A. (2023). A cost-benefit analysis for the appraisal of social and market prices in the probabilistic seismic risk assessment of building portfolios: A methodology for the evaluation of disaster risk reduction programs. *International Journal of Disaster Risk Reduction*, 90(October 2022), 103637. <https://doi.org/10.1016/j.ijdr.2023.103637>
- Aroquipa, H., & Hurtado, A. I. (2022b). Incremental seismic retrofitting for essential facilities using performance objectives: A case study of the 780-PRE school buildings in Peru. *Journal of Building Engineering*, 62, 105387. <https://doi.org/10.1016/j.job.2022.105387>
- Aroquipa Velásquez, H. (2014). *Procesos constructivos de edificaciones y sus impactos ambientales con relación a una producción limpia y sostenible*. Universidad Nacional del Altiplano.
- Arteaga, C., Paz, A., & Park, J. (2020). Injury severity on traffic crashes : A text mining with an interpretable machine-learning approach. *Safety Science*, 132(September), 104988. <https://doi.org/10.1016/j.ssci.2020.104988>
- Assaf, A. G., & Tsionas, M. (2021). A Bayesian solution to multicollinearity through unobserved common factors. *Tourism Management*, 84(November 2020), 104277. <https://doi.org/10.1016/j.tourman.2020.104277>
- Assi, K., Rahman, S. M., Mansoor, U., & Ratrou, N. (2020). Predicting Crash Injury Severity with Machine Learning Algorithm Synergized with Clustering Technique: A Promising Protocol. *International Journal of Environmental Research and Public Health*, 17(15),

5497. <https://doi.org/10.3390/ijerph17155497>
- Babič, A., & Dolšek, M. (2016). Seismic fragility functions of industrial precast building classes. *Engineering Structures*, *118*, 357–370. <https://doi.org/10.1016/j.engstruct.2016.03.069>
- Bai, Q., & Labi, S. (2009). *Uncertainty-Based Tradeoff Analysis Methodology for Integrated Transportation Investment Decision-Making*. 92.
- Bajpayee, T. S., Rehak, T. R., Mowrey, G. L., & Ingram, D. K. (2004). Blasting injuries in surface mining with emphasis on flyrock and blast area security. *Journal of Safety Research*, *35*(1), 47–57. <https://doi.org/10.1016/j.jsr.2003.07.003>
- Baker, H., Hallowell, M. R., & Tixier, A. J. (2020). AI-based prediction of independent construction safety outcomes from universal attributes. *Automation in Construction*, *118*(August 2019), 103146. <https://doi.org/10.1016/j.autcon.2020.103146>
- Ben-Hur, A., Horn, D., Siegelmann, H. T., & Vapnik, V. (2001). Support vector clustering. *Journal of Machine Learning Research*, *2*(Dec), 125–137.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(2).
- Bhattacharjee, A. (1991). *Mine safety management: an application of risk analysis, forecasting techniques, and Markov process to injury experience data*. The Pennsylvania State University.
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). Training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, 144–152. <https://doi.org/10.1145/130385.130401>
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.
- Brink, H., Richards, J., & Fetherolf, M. (2016). *Real-world machine learning*. Simon and

Schuster.

- Burnett, R. A., & Si, D. (2017). Prediction of injuries and fatalities in aviation accidents through machine learning. *Proceedings of the International Conference on Compute and Data Analysis*, 60–68.
- Calderon, M. A. (2015). *Optimización de las prácticas de perforación y voladura en el avance y producción de la minería de mediana escala (Unidad Minera Macdesa)*.
- Calderón Solis, A. (2012). *Análisis e implementación de un sistema de gestión de riesgos para la prevención de accidentes en la mina El Brocal SAA Unidad Colquijirca-Pasco*.
- Campos, W.(2010). *Apuntes de Metodología de La Investigación Científica*. Magister SAC.
- Candia, R. C., Hennies, W. T., Azevedo, R. C., Almeida, I., & Soto, J. (2010). Análisis de accidentes fatales en la industria minera peruana. *Boletín Geológico y Minero de España*, 1, 57–68.
- Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers and Industrial Engineering*, 137(September), 106024. <https://doi.org/10.1016/j.cie.2019.106024>
- Castaño Martínez, F. L., Betín, J. M. H., & Sáenz, J. N. G. (2009). Análisis cualitativo del flujo de agua de infiltración para el control del drenaje de una estructura de pavimento flexible. *Infraestructura Vial*, 1, 6.
- Center, N. S. (2013). The case for safety the North Sea Piper alpha disaster. *System Failure Case Studies*, 7, 1–4.
- Chebila, M. (2021). Predicting the consequences of accidents involving dangerous substances using machine learning. *Ecotoxicology and Environmental Safety*, 208, 111470.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the*

- 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 785–794.
- Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., & Shah, H. (2016). Wide & Deep Learning for Recommender Systems. *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, 7–10. <https://doi.org/10.1145/2988450.2988454>
- Chien, K. F., Wu, Z. H., & Huang, S. C. (2014). Identifying and assessing critical risk factors for BIM projects: Empirical study. *Automation in Construction*, 45, 1–15. <https://doi.org/10.1016/j.autcon.2014.04.012>
- Choi, J., Gu, B., Chin, S., & Lee, J. S. (2020). Machine learning predictive model based on national data for fatal accidents of construction workers. *Automation in Construction*, 110(September 2019), 102974. <https://doi.org/10.1016/j.autcon.2019.102974>
- Choudhry, R. M., & Iqbal, K. (2013). Identification of Risk Management System in Construction Industry in Pakistan. *Journal of Management in Engineering*, 29(1), 42–49. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000122](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000122)
- Cimellaro, G. P., Reinhorn, A. M., & Bruneau, M. (2010). Seismic resilience of a hospital system. *Structure and Infrastructure Engineering*, 6(1–2), 127–144. <https://doi.org/10.1080/15732470802663847>
- Conesa, J. A. B. (2014). Implantación de la Responsabilidad Social en la Administración Pública: el caso de las Fuerzas Armadas Españolas. *CONSEJO EDITORIAL*.
- Cortina, A., Conill, J., Domingo, A., & García-Marzá, D. (1996). *Ética de la empresa* (Issue 174/C82e). Trotta Madrid.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and*

mixed methods approaches. Sage publications.

Cruz, I., & Huerta, R. (2015). Occupational Safety and Health in Peru. *Annals of Global Health*, 81(4), 568. <https://doi.org/10.1016/j.aogh.2015.08.027>

Del Gaudio, C., Ricci, P., Verderame, G. M., & Manfredi, G. (2015). Development and urban-scale application of a simplified method for seismic fragility assessment of RC buildings. *Engineering Structures*, 91, 40–57. <https://doi.org/10.1016/j.engstruct.2015.01.031>

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.

Diez-pastor, J. F., Gil, A., Val, D., Veiga, F., & Bustillo, A. (2021). High-accuracy classification of thread quality in tapping processes with ensembles of classifiers for imbalanced learning. *Measurement*, 168(August 2020), 108328. <https://doi.org/10.1016/j.measurement.2020.108328>

Dikmen, I., Birgonul, M. T., Tah, J. H. M., & Ozer, A. H. (2012). Web-Based Risk Assessment Tool Using Integrated Duration–Cost Influence Network Model. *Journal of Construction Engineering and Management*, 138(9), 1023–1034. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000547](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000547)

do Nascimento Beckert, A., & Barros, V. G. (2022). Waste management, COVID-19 and occupational safety and health: Challenges, insights and evidence. *Science of The Total Environment*, 831, 154862.

Drummond, C. (2017). Classification. In *Encyclopedia of Machine Learning and Data Mining* (pp. 205–208). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_111

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211. [https://doi.org/10.1016/0364-0213\(90\)90002-E](https://doi.org/10.1016/0364-0213(90)90002-E)

Emprey, D. E. (1992). Incorporating management and organizational factors into probabilistic

- safety assesment. *Reliability Engineering and System Safety*, 38, 199–208.
- Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent Semantic Analysis: five methodological recommendations. *European Journal of Information Systems*, 21(1), 70–86. <https://doi.org/10.1057/ejis.2010.61>
- Eybpoosh, M., Dikmen, I., & Talat Birgonul, M. (2011). Identification of Risk Paths in International Construction Projects Using Structural Equation Modeling. *Journal of Construction Engineering and Management*, 137(12), 1164–1175. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000382](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000382)
- Fang, C., & Marle, F. (2012). A simulation-based risk network model for decision support in project risk management. *Decision Support Systems*, 52(3), 635–644. <https://doi.org/10.1016/j.dss.2011.10.021>
- Feng, S., Keung, J., Yu, X., Xiao, Y., & Zhang, M. (2021). Investigation on the stability of SMOTE-based oversampling techniques in software defect prediction. *Information and Software Technology*, 139(August 2020), 106662. <https://doi.org/10.1016/j.infsof.2021.106662>
- Firmansyah, H. R., Sarli, P. W., Twinanda, A. P., Santoso, D., & Imran, I. (2024). Building typology classification using convolutional neural networks utilizing multiple ground-level image process for city-scale rapid seismic vulnerability assessment. *Engineering Applications of Artificial Intelligence*, 131(December 2023), 107824. <https://doi.org/10.1016/j.engappai.2023.107824>
- Flyvbjerg, B. (2013). Quality control and due diligence in project management: Getting decisions right by taking the outside view. *International Journal of Project Management*, 31(5), 760–774. <https://doi.org/10.1016/j.ijproman.2012.10.007>
- Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. *Icml*, 96,

148–156.

Friedman, J. H. (2001a). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.

Friedman, J. H. (2001b). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>

García Navarro, F. (2012). *Responsabilidad social corporativa: teoría y práctica*. ESIC editorial.

George, E. P., Hunter, W. G., & Hunter, J. S. (2005). *Statistics for experimenters: design, innovation, and discovery*. Wiley.

Gerassis, S., Saavedra, Á., Taboada, J., Alonso, E., & Bastante, F. G. (2020). Differentiating between fatal and non-fatal mining accidents using artificial intelligence techniques. *International Journal of Mining, Reclamation and Environment*, 34(10), 687–699. <https://doi.org/10.1080/17480930.2019.1700008>

Gerón, A. (2023). *Aprende machine learning con Scikit-Learn, Keras y TensorFlow. Tercera edición : conceptos, herramientas y técnicas para conseguir sistemas inteligentes*. 832.

Giraldo, M., & Badillo, J. (2015). Implicancias técnicas y económicas de los accidentes mortales en la minería peruana. *Revista Del Instituto de Investigación (RIIGEO), FIGMMG-UNMSM*, 18(35), 97–107.

Giraldo Paredez, E. M. (2016). Identificación de factores para reducir accidentes por desprendimiento de rocas en minería subterránea. *Revista Del Instituto de Investigación de La Facultad de Ingeniería Geológica, Minera, Metalurgica y Geográfica*, 19(37), 47–55.

Goh, Y. M., Ubeynarayana, C. U., Le, K., Wong, X., & Guo, B. H. W. (2018). *Factors in influencing unsafe behaviors: A supervised learning approach*. 118(June), 77–85.

<https://doi.org/10.1016/j.aap.2018.06.002>

- Gómez, P., Hernández Guerrero, J., & Méndez Campos, M. D. (2014). Factores de riesgo psicosocial y satisfacción laboral en una empresa chilena del área de la minería. *Ciencia & Trabajo*, 16(49), 9–16.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Gribaudo, M., Iacono, M., & Kiran, M. (2018). A performance modeling framework for lambda architecture based applications. *Future Generation Computer Systems*, 86, 1032–1041.
- Guzman, J., Arianne, G., Wahid, A., Padrones, J. M., & Jared, J. (2022). Evaluating workplace safety in the oil and gas industry during the COVID-19 pandemic using occupational health and safety Vulnerability Measure and partial least square Structural Equation Modelling. *Cleaner Engineering and Technology*, 6, 100378. <https://doi.org/10.1016/j.clet.2021.100378>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Heravi, G., & Charkhakan, M. H. (2014). Predicting and Tracing Change-Formation Scenarios in Construction Projects Using the DEMATEL Technique. *Journal of Management in Engineering*, 30(6). [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000229](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000229)
- Heravi, G., & Hajihosseini, Z. (2012). Risk Allocation in Public–Private Partnership Infrastructure Projects in Developing Countries: Case Study of the Tehran–Chalus Toll Road. *Journal of Infrastructure Systems*, 18(3), 210–217. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000090](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000090)
- Hernández-Sampieri, R., Fernández, C., & Batista, P. (2010). Metodología de la Investigación.(5ta. Edición). Editorial Mc Graw Hill.

- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). Los métodos mixtos. *Metodología de La Investigación*, 5, 544–599.
- Hingorani, R., Tanner, P., Prieto, M., Lara, C., Science, C., & Torroja, E. (2020). Consequence classes and associated models for predicting loss of life in collapse of building structures. *Structural Safety*, 85(September 2019), 101910. <https://doi.org/10.1016/j.strusafe.2019.101910>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hollnagel, E., Woods, D. D., & Leveson, N. (2006). *Resilience engineering: Concepts and precepts*. Ashgate Publishing, Ltd.
- Hubbard, D. W. (2020). *The failure of risk management: Why it's broken and how to fix it*. John Wiley & Sons.
- Ilbahar, E., Karaşan, A., Cebi, S., & Kahraman, C. (2018). A novel approach to risk assessment for occupational health and safety using Pythagorean fuzzy AHP & fuzzy inference system. *Safety Science*, 103(October 2017), 124–136. <https://doi.org/10.1016/j.ssci.2017.10.025>
- Ivakhnenko, A. G. (1971). Polynomial Theory of Complex Systems. *IEEE Transactions on Systems, Man and Cybernetics*, 1(4), 364–378. <https://doi.org/10.1109/TSMC.1971.4308320>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>
- Jang, J. R. (1993). *ANFIS: Adaptive-Neural Network-Based Fuzzy Inference System*. 23(3).
- Janitza, S., Tutz, G., & Boulesteix, A. (2016). Random forest for ordinal responses : Prediction and variable selection. *Computational Statistics and Data Analysis*, 96, 57–73.

<https://doi.org/10.1016/j.csda.2015.10.005>

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169–15211. <https://doi.org/10.1007/s11042-018-6894-4>

Jin, X., & Doloi, H. (2008). Interpreting risk allocation mechanism in public–private partnership projects: an empirical study in a transaction cost economics perspective. *Construction Management and Economics*, 26(7), 707–721. <https://doi.org/10.1080/01446190801998682>

Joughin, W. C., Jager, A., Nezomba, E., & Rwodzi, L. (2012). A risk evaluation model for support design in Bushveld Complex underground mines: Part I-Description of the model. *Journal of the Southern African Institute of Mining and Metallurgy*, 112(2), 83–94.

Jukes, E. (2018). Encyclopedia of Machine Learning and Data Mining (2nd edition). *Reference Reviews*, 32(7/8), 3–4. <https://doi.org/10.1108/RR-05-2018-0084>

Kaiser, M. G., El Arbi, F., & Ahlemann, F. (2015). Successful project portfolio management beyond project selection techniques: Understanding the role of structural alignment. *International Journal of Project Management*, 33(1), 126–139. <https://doi.org/10.1016/j.ijproman.2014.03.002>

Kalkan, E., & Kunnath, S. K. (2006). Adaptive Modal Combination Procedure for Nonlinear Static Analysis of Building Structures. *Journal of Structural Engineering*, 132(11), 1721–1731. [https://doi.org/10.1061/\(asce\)0733-9445\(2006\)132:11\(1721\)](https://doi.org/10.1061/(asce)0733-9445(2006)132:11(1721))

Kang, K., & Ryu, H. (2019). Predicting types of occupational accidents at construction sites in Korea using random forest model. *Safety Science*, 120(January), 226–236. <https://doi.org/10.1016/j.ssci.2019.06.034>

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017).

- Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.
- Khademi Hamidi, J., Shahriar, K., Rezai, B., Rostami, J., & Bejari, H. (2010). Risk assessment based selection of rock TBM for adverse geological conditions using Fuzzy-AHP. *Bulletin of Engineering Geology and the Environment*, 69(4), 523–532. <https://doi.org/10.1007/s10064-009-0260-8>
- Khahro, S. H., Memon, Z. A., Gungat, L., Yazid, M. R. M., Rahim, A., Mubaraki, M., & Nur, N. I. (2021). Low-cost pavement management system for developing countries. *Sustainability (Switzerland)*, 13(11). <https://doi.org/10.3390/su13115941>
- Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53(8), 5455–5516. <https://doi.org/10.1007/s10462-020-09825-6>
- Khan, D., & Rawat, A. (2016). Nonlinear Seismic Analysis of Masonry Infill RC Buildings with Eccentric Bracings at Soft Storey Level. *Procedia Engineering*, 161, 9–17. <https://doi.org/10.1016/j.proeng.2016.08.490>
- Kim, D. Y., Han, S. H., Kim, H., & Park, H. (2009). Structuring the prediction model of project performance for international construction projects: A comparative analysis. *Expert Systems with Applications*, 36(2 PART 1), 1961–1971. <https://doi.org/10.1016/j.eswa.2007.12.048>
- Kim, H. (2022). Deep Learning. *Artificial Intelligence for 6G*, 22(4), 247–303. https://doi.org/10.1007/978-3-030-95041-5_6
- Kiran, M., Murphy, P., Monga, I., Dugan, J., & Baveja, S. S. (2015a). Lambda architecture for cost-effective batch and speed big data processing. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 2785–2792.

<https://doi.org/10.1109/BigData.2015.7364082>

- Kiran, M., Murphy, P., Monga, I., Dugan, J., & Baveja, S. S. (2015b). Lambda architecture for cost-effective batch and speed big data processing. *2015 IEEE International Conference on Big Data (Big Data)*, 2785–2792.
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *IJCAI International Joint Conference on Artificial Intelligence*, 2(March 2001), 1137–1143.
- Koldas, K. S. (2001). Rock-related accidents, investigations and inquiries in South African mines. *17th International Mining Congress and Exhibition of Turkey-IMCET*.
- Kroß, J., Brunnert, A., Prehofer, C., Runkler, T. A., & Krcmar, H. (2015a). Stream processing on demand for lambda architectures. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9272, 243–257. https://doi.org/10.1007/978-3-319-23267-6_16
- Kroß, J., Brunnert, A., Prehofer, C., Runkler, T. A., & Krcmar, H. (2015b). Stream processing on demand for lambda architectures. *European Workshop on Performance Engineering*, 243–257.
- Kwon, O. H., Rhee, W., & Yoon, Y. (2015). Application of classification algorithms for analysis of road safety risk factor dependencies. *Accident Analysis and Prevention*, 75, 1–15. <https://doi.org/10.1016/j.aap.2014.11.005>
- Landucci, G., & Paltrinieri, N. (2016). A methodology for frequency tailorization dedicated to the Oil & Gas sector. *Process Safety and Environmental Protection*, 104, 123–141. <https://doi.org/10.1016/j.psep.2016.08.012>
- Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (2023). *Gradient-based learning applied to document recognition To cite this version :*

HAL Id : hal-03926082 Gradient-Based Learning Applied to Document Recognition.
86(11), 2278–2324.

- Leeanansaksiri, A., Panyakapo, P., & Ruangrassamee, A. (2018). Seismic capacity of masonry infilled RC frame strengthening with expanded metal ferrocement. *Engineering Structures*, 159, 110–127. <https://doi.org/10.1016/J.ENGSTRUCT.2017.12.034>
- Li, D., Deng, L., Bhooshan Gupta, B., Wang, H., & Choi, C. (2019). A novel CNN based security guaranteed image watermarking generation scenario for smart city applications. *Information Sciences*, 479, 432–447. <https://doi.org/10.1016/j.ins.2018.02.060>
- Li, F., Phoon, K. K., Du, X., & Zhang, M. (2013). Improved AHP Method and Its Application in Risk Identification. *Journal of Construction Engineering and Management*, 139(3), 312–320. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000605](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000605)
- Likun, W., & Zaili, Y. (2018). *Bayesian network modelling and analysis of accident severity in waterborne transportation: A case study in China*. 180(July), 277–289. <https://doi.org/10.1016/j.ress.2018.07.021>
- Liu, J., Li, Q., & Wang, Y. (2013a). Risk analysis in ultra deep scientific drilling project - A fuzzy synthetic evaluation approach. *International Journal of Project Management*, 31(3), 449–458. <https://doi.org/10.1016/j.ijproman.2012.09.015>
- Liu, J., Li, Q., & Wang, Y. (2013b). Risk analysis in ultra deep scientific drilling project - A fuzzy synthetic evaluation approach. *International Journal of Project Management*, 31(3), 449–458. <https://doi.org/10.1016/j.ijproman.2012.09.015>
- Liu, R., Liu, Z., Liu, H., & Shi, H. (2021). Automation in Construction An improved alternative queuing method for occupational health and safety risk assessment and its application to construction excavation. *Automation in Construction*, 126(April 2020), 103672. <https://doi.org/10.1016/j.autcon.2021.103672>

- Lourenço, P., & Cristina, A. (2021). Estimating tree aboveground biomass using multispectral satellite-based data in Mediterranean agroforestry system using random forest algorithm. *Remote Sensing Applications: Society and Environment*, 23(May). <https://doi.org/10.1016/j.rsase.2021.100560>
- Lowe, D. S. B. and D. (1988). Multivariable Functional Interpolation and Adaptive Networks. *Complex Systems*, 2, 321–355.
- Lu, J. L. (2022). Mining safety and health in the Philippines: Occupational and environmental impacts. *Safety and Health at Work*, 13, S142.
- Lu, J., Su, W., Jiang, M., & Ji, Y. (2022). Severity prediction and risk assessment for non-traditional safety events in sea lanes based on a random forest approach. *Ocean and Coastal Management*, 225(February), 106202. <https://doi.org/10.1016/j.ocecoaman.2022.106202>
- Luo, X., Cruz, A. M., & Tzioutzios, D. (2020). Extracting Natech Reports from Large Databases: Development of a Semi-Intelligent Natech Identification Framework. *International Journal of Disaster Risk Science*, 11(6), 735–750. <https://doi.org/10.1007/s13753-020-00314-6>
- Luo, X., Li, X., Goh, Y. M., Song, X., & Liu, Q. (2023). Application of machine learning technology for occupational accident severity prediction in the case of construction collapse accidents. *Safety Science*, 163(March), 106138. <https://doi.org/10.1016/j.ssci.2023.106138>
- Lyu, H.-M., Sun, W.-J., Shen, S.-L., & Zhou, A.-N. (2020). Risk Assessment Using a New Consulting Process in Fuzzy AHP. *Journal of Construction Engineering and Management*, 146(3). [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001757](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001757)
- Manning, C. D. (2008). *Introduction to information retrieval*. Syngress Publishing,.

- Marcelino-Sádaba, S., Pérez-Ezcurdia, A., Echeverría Lazcano, A. M., & Villanueva, P. (2014). Project risk management methodology for small firms. *International Journal of Project Management*, 32(2), 327–340. <https://doi.org/10.1016/j.ijproman.2013.05.009>
- Martínez, A. V., Unanue, R. M., & Yuste, Á. R. (2015). *Arquitectura lambda aplicada a clustering de documentos en contextos Big Data*. [Tesis de maestría]. Universidad Nacional de Educación a Distancia.
- Mazaira, A., & Konicek, P. (2015). Intense rockburst impacts in deep underground construction and their prevention. *Canadian Geotechnical Journal*, 52(10), 1426–1439.
- Medina, A. B. C. (2014). Investigación holística y Desarrollo Instruccional en la comprensión del discurso escrito en estudiantes de educación media de la UEN José Félix Blanco. Caracas. *Revista de Investigación*, 38(81), 69–88.
- Mejía, C. R., Merino, P. A., Mescua, L., & Gomero, R. (2015). Notificación de accidentes mortales en el sector de minería Peruana, 2000-2014. Accidentes mortales en minería peruana. *Archivos de Medicina*, 11(4), 11.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26.
- Moreno, N. (2011). *Modelo de un programa de seguridad e higiene para la minería subterránea del carbón en Colombia*. [Tesis de grado]. Universidad Nacional de Ingeniería.
- Moretti, L., Cantisani, G., & Di Mascio, P. (2016). Management of road tunnels: Construction, maintenance and lighting costs. *Tunnelling and Underground Space Technology*, 51, 84–89. <https://doi.org/10.1016/j.tust.2015.10.027>
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

- Mutlu, N. G., & Altuntas, S. (2019). International Journal of Industrial Ergonomics Risk analysis for occupational safety and health in the textile industry : Integration of FMEA , FTA , and BIFPET methods. *International Journal of Industrial Ergonomics*, 72(June), 222–240. <https://doi.org/10.1016/j.ergon.2019.05.013>
- Noa Monjes, R. R. (2006). Indicaciones metodológicas para la elección del método de arranque de las rocas durante el laboreo de excavaciones subterráneas horizontales de pequeña y mediana sección en Cuba oriental. *Minería y Geología*, 22(4), 50.
- Paltrinieri, N., Dechy, N., Salzano, E., Wardman, M., & Cozzani, V. (2013). Towards a new approach for the identification of atypical accident scenarios. *Journal of Risk Research*, 16(3–4), 337–354. <https://doi.org/10.1080/13669877.2012.729518>
- Pan, T., Zhao, J., Wu, W., & Yang, J. (2020). Learning imbalanced datasets based on SMOTE and Gaussian distribution. *Information Sciences*, 512, 1214–1233. <https://doi.org/10.1016/j.ins.2019.10.048>
- Parinaz Jafari, Malak Al Hattab, E. M., & Simaan AbouRizk. (2021). Automated Extraction and Time-Cost Prediction of Contractual Reporting Requirements in Construction Using Natural Language Processing and Simulation. *MDPI*. <https://doi.org/doi.org/10.3390/app11136188>
- Pasman, H. J. (2009). Learning from the past and knowledge management: Are we making progress? *Journal of Loss Prevention in the Process Industries*, 22(6), 672–679. <https://doi.org/10.1016/j.jlp.2008.07.010>
- Paul, P. S., & Maiti, J. (2007). The role of behavioral factors on safety management in underground mines. *Safety Science*, 45(4), 449–471.
- Pérez, J. L. (2007). *Sistema de gestión en seguridad y salud ocupacional aplicado a empresas contratistas en el sector económico minero metalúrgico*. [Tesis de grado]. Universidad

Nacional de Ingeniería

- Phark, C., Kim, W., Yoon, Y. S., Shin, G., & Jung, S. (2018). Prediction of issuance of emergency evacuation orders for chemical accidents using machine learning algorithm. *Journal of Loss Prevention in the Process Industries*, 56(March), 162–169. <https://doi.org/10.1016/j.jlp.2018.08.021>
- Poh, C. Q. X., Udhyami, C., & Miang, Y. (2018). Automation in Construction Safety leading indicators for construction sites: A machine learning approach. *Automation in Construction*, 93(March), 375–386. <https://doi.org/10.1016/j.autcon.2018.03.022>
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. “O’Reilly Media, Inc.”
- Python Software Foundation (2024). *Welcome to Python.org*.
- Qazi, A., Quigley, J., Dickson, A., & Kirytopoulos, K. (2016). Project Complexity and Risk Management (ProCRiM): Towards modelling project complexity driven risk paths in construction projects. *International Journal of Project Management*, 34(7), 1183–1198. <https://doi.org/10.1016/j.ijproman.2016.05.008>
- Qian, W., Huang, J., Wang, Y., & Shu, W. (2020). Knowledge-Based Systems Mutual information-based label distribution feature selection for. *Knowledge-Based Systems*, 195, 105684. <https://doi.org/10.1016/j.knosys.2020.105684>
- Rajasekaran, S., Gayathri, S., & Lee, T. L. (2008). Support vector regression methodology for storm surge predictions. *Ocean Engineering*, 35(16), 1578–1587. <https://doi.org/10.1016/j.oceaneng.2008.08.004>

- Reason, J. (1990). James Reason : Human Error Person approach Evaluating the person approach. *BMJ*, *320*(7237), 1–7.
- Retrieval, I. (2009). Introduction to information retrieval. In *Choice Reviews Online* (Vol. 46, Issue 05). <https://doi.org/10.5860/choice.46-2715>
- Rezakhani, P., Jang, W.-S., Lee, S., & Lee, D.-E. (2014). Project risk assessment model combining the fuzzy weighted average principle with a similarity measure. *KSCIE Journal of Civil Engineering*, *18*(2), 521–530. <https://doi.org/10.1007/s12205-014-0053-x>
- Rinc, I. B. (2011). Investigación científica e investigación tecnológica como componentes para la innovación: consideraciones técnicas y metodológicas. *Contribuciones a las Ciencias Sociales*, 220–255.
- Robert F. DeVellis. (2019). Frank D. Brown Hall Room 3208. *To Achieve Excellence by Guiding Individuals as They Become Professionals*”, 1–7.
- Rosenblatt, F. (2021). The Perceptron: A Probabilistic Model for Information Storage and Organization (1958). *Ideas That Created the Future*, *65*(1958), 183–190. <https://doi.org/10.7551/mitpress/12274.003.0020>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Sarkar, S., & Maiti, J. (2020a). Machine learning in occupational accident analysis : A review using science mapping approach with citation network analysis. *Safety Science*, *131*(September 2019), 104900. <https://doi.org/10.1016/j.ssci.2020.104900>
- Sarkar, S., & Maiti, J. (2020b). Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. *Safety Science*, *131*(June), 104900. <https://doi.org/10.1016/j.ssci.2020.104900>
- Sarkar, S., Pramanik, A., Maiti, J., & Reniers, G. (2020). Predicting and analyzing injury

- severity: A machine learning-based approach using class-imbalanced proactive and reactive data. *Safety Science*, 125(February), 104616. <https://doi.org/10.1016/j.ssci.2020.104616>
- Sasaki, Y. (2007). The truth of the f-measure. 2007. URL: [https://www. Cs. Odu. Edu/Mukka/Cs795sum09dm/Lecturenotes/Day3/F-Measure-YS-26Oct07. Pdf](https://www.cs.odu.edu/Mukka/Cs795sum09dm/Lecturenotes/Day3/F-Measure-YS-26Oct07.Pdf) [Accessed 2021-05-26], 49.
- Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Schwalbe, K. (2016). *Information technology project management*. Cengage Learning.
- Simard, M., & Marchand, A. (2000). A multilevel analysis of organisational factors related to the taking of safety initiatives by work groups. In *Safety Science* (Vol. 21, Issue 2). Elsevier.
- Song, L., & Yang, L. (2021). Governance innovation of occupational safety and health in China under the background of industry 4.0. *Industrial Safety and Environmental Protection*, 47(09), 79–82.
- Sony, S., Dunphy, K., Sadhu, A., & Capretz, M. (2021). A systematic review of convolutional neural network-based structural condition assessment techniques. *Engineering Structures*, 226(January 2020), 111347. <https://doi.org/10.1016/j.engstruct.2020.111347>
- Stemn, E., & Krampah, F. (2022). Injury severity and influence factors in surface mines : A correspondence analysis. *Safety Science*, 145(September 2021), 105495. <https://doi.org/10.1016/j.ssci.2021.105495>
- Stone, P. (2017). Reinforcement Learning. In *Encyclopedia of Machine Learning and Data Mining* (pp. 1088–1090). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_720
- Tamascelli, N., Solini, R., Paltrinieri, N., & Cozzani, V. (2022). Learning from major

- accidents : A machine learning approach. *Computers and Chemical Engineering*, 162(July 2021), 107786. <https://doi.org/10.1016/j.compchemeng.2022.107786>
- Tavakolan, M., & Etemadinia, H. (2017). Fuzzy Weighted Interpretive Structural Modeling: Improved Method for Identification of Risk Interactions in Construction Projects. *Journal of Construction Engineering and Management*, 143(11). [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001395](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001395)
- Terán, I. S. (2012). *Propuesta de implementación de un sistema de gestión de seguridad y salud ocupacional bajo la norma OHSAS 18001 en una empresa de capacitación técnica para la industria*. [Tesis de grado]. Pontificia Universidad Católica del Perú
- Tharwat, A. (2019). Parameter investigation of support vector machine classifier with kernel functions. *Knowledge and Information Systems*, 61, 1269–1302.
- Thorvaldsen, T., Kongsvik, T., Marie, I., Størkersen, K., Salomonsen, C., Sandsund, M., & Bjelland, H. V. (2020). Occupational health , safety and work environments in Norwegian fi sh farming - employee perspective. *Aquaculture*, 524(December 2019), 735238. <https://doi.org/10.1016/j.aquaculture.2020.735238>
- Tixier, A. J., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2020). Automation in Construction Automated content analysis for construction safety : A natural language processing system to extract precursors and outcomes from unstructured injury reports. *Automation in Construction*, 62(2016), 45–56. <https://doi.org/10.1016/j.autcon.2015.11.001>
- Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2016). Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports. *Automation in Construction*, 62(2016), 45–56. <https://doi.org/10.1016/j.autcon.2015.11.001>

- Unidas, N. (2003). Congestión de tránsito: el problema y cómo enfrentarlo. *CEPAL: División de Medio Ambiente y Asentamientos Humanos*.
- U.S. Green Building Council. (n.d.). *Assessment and Planning for Resilience*
- Vaiman, Bell, Chen, Chowdhury, Dobson, Hines, Papic, Miller, & Zhang. (2012). Risk Assessment of Cascading Outages: Methodologies and Challenges. *IEEE Transactions on Power Systems*, 27(2), 631–641. <https://doi.org/10.1109/TPWRS.2011.2177868>
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
- Viteri, N. C. (2012). La investigación mixta, estrategia andragógica fundamental para fortalecer las capacidades intelectuales superiores. *Revista científica*, 2(2).
- Wahab, L., & Jiang, H. (2019). A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. *PLoS One*, 14(4), e0214966.
- Wang, J., & Yuan, H. (2017). System Dynamics Approach for Investigating the Risk Effects on Schedule Delay in Infrastructure Projects. *Journal of Management in Engineering*, 33(1). [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000472](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000472)
- Wang, L., & Zhang, X. (2018). Bayesian Analytics for Estimating Risk Probability in PPP Waste-to-Energy Projects. *Journal of Management in Engineering*, 34(6). [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000658](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000658)
- Wang, W., & Yang, Y. (2019). Development of convolutional neural network and its application in image classification: a survey. *Optical Engineering*, 58(04), 1. <https://doi.org/10.1117/1.oe.58.4.040901>
- Wright, C. (1986). Routine deaths: Fatal accidents in the oil industry. *The Sociological Review*, 34(2), 265–289.
- Wu, N. J. (2021). Predicting the compressive strength of concrete using an rbf-ann model.

- Applied Sciences (Switzerland)*, 11(14). <https://doi.org/10.3390/app11146382>
- Xu, N., Ma, L., Liu, Q., Wang, L., & Deng, Y. (2021). An improved text mining approach to extract safety risk factors from construction accident reports. *138*(February). <https://doi.org/10.1016/j.ssci.2021.105216>
- Xu, Q., Chong, H., & Liao, P. (2019). Exploring eye-tracking searching strategies for construction hazard recognition in a laboratory scene. *Safety Science*, 120(August), 824–832. <https://doi.org/10.1016/j.ssci.2019.08.012>
- Xu, Z., & Saleh, J. H. (2021). Machine learning for reliability engineering and safety applications: Review of current status and future opportunities. *Reliability Engineering and System Safety*, 211(December 2020), 107530. <https://doi.org/10.1016/j.ress.2021.107530>
- Yan, X., Xiong, X., Cheng, X., Huang, Y., & Zhu, H. (2021). HMM-BiMM : Hidden Markov Model-based word segmentation via improved Bi-directional Maximal Matching algorithm ☆. *Computers and Electrical Engineering*, 94(July), 107354. <https://doi.org/10.1016/j.compeleceng.2021.107354>
- Yang, K., & Ahn, C. R. (2019). Advanced Engineering Informatics Inferring workplace safety hazards from the spatial patterns of workers ' wearable data. *Advanced Engineering Informatics*, 41(January), 100924. <https://doi.org/10.1016/j.aei.2019.100924>
- Yedla, A., Kakhki, F. D., & Jannesari, A. (2020). Predictive Modeling for Occupational Safety Outcomes and Days Away from Work Analysis in Mining Operations. *International Journal of Environmental Research and Public Health*, 17(19), 7054. <https://doi.org/10.3390/ijerph17197054>
- Yim, R., Castaneda, J., Doolen, T., Tumer, I., & Malak, R. (2015). A study of the impact of project classification on project risk indicators. *International Journal of Project*

- Management*, 33(4), 863–876. <https://doi.org/10.1016/j.ijproman.2014.10.005>
- Yodo, N., & Wang, P. (2016). Resilience modeling and quantification for engineered systems using Bayesian networks. *Journal of Mechanical Design*, 138(3), 31404.
- yoshua bengio. (2019). Deep learning 简介 一、什么是 Deep Learning ? . *Nature*, 29(7553), 1–73.
- Yusri, A. Z. dan D. (2020). 濟無No Title No Title No Title. *Jurnal Ilmu Pendidikan*, 7(2), 809–820.
- Zhang, F. (2022). A hybrid structured deep neural network with Word2Vec for construction accident causes classification. *International Journal of Construction Management*, 22(6), 1120–1140.
- Zhang, J., & El-Gohary, N. M. (2017). Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. *Automation in Construction*, 73, 45–57. <https://doi.org/10.1016/j.autcon.2016.08.027>
- Zhang, J., Li, Z., Pu, Z., & Xu, C. (2018). Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods. *IEEE Access*, 6, 60079–60087. <https://doi.org/10.1109/ACCESS.2018.2874979>
- Zhao, B., Liu, Y., Goh, S. H., & Lee, F. H. (2016). Parallel finite element analysis of seismic soil structure interaction using a PC cluster. *Computers and Geotechnics*, 80, 167–177. <https://doi.org/10.1016/j.compgeo.2016.07.006>
- Zhao, X., Hwang, B. G., & Gao, Y. (2016). A fuzzy synthetic evaluation approach for risk assessment: A case of Singapore's green projects. *Journal of Cleaner Production*, 115(January 2005), 203–213. <https://doi.org/10.1016/j.jclepro.2015.11.042>
- Zheng, X., Easa, S. M., Ji, T., & Jiang, Z. (2020). Incorporating uncertainty into life-cycle sustainability assessment of pavement alternatives. *Journal of Cleaner Production*,

264(April). <https://doi.org/10.1016/j.jclepro.2020.121466>

Zorzenon, R., Lizarelli, F. L., & Moura, D. B. A. D. A. (2022). What is the potential impact of industry 4 . 0 on health and safety at work ? *Safety Science*, 153(February), 105802. <https://doi.org/10.1016/j.ssci.2022.105802>

Zou, Y., Kiviniemi, A., & Jones, S. W. (2016). Developing a tailored RBS linking to BIM for risk management of bridge projects. *Engineering, Construction and Architectural Management*, 23(6), 727–750. <https://doi.org/10.1108/ECAM-01-2016-0009>

Zou, Y., Kiviniemi, A., & Jones, S. W. (2017). Retrieving similar cases for construction project risk management using Natural Language Processing techniques. *Automation in Construction*, 80(February), 66–76. <https://doi.org/10.1016/j.autcon.2017.04.003>

IX. ANEXOS

Anexo A

Matriz de Consistencias

PROBLEMA	OBJETIVOS	HIPÓTESIS	METODOLOGÍA
<p>General</p> <ul style="list-style-type: none"> • ¿En qué medida influye la aplicación de técnicas de inteligencia artificial y <i>machine learning</i> en la predicción y prevención de accidentes laborales en la minería subterránea del Perú? 	<p>General</p> <ul style="list-style-type: none"> • Desarrollar un modelo predictivo basado en técnicas de inteligencia artificial y <i>machine learning</i> para identificar y prevenir accidentes laborales en la minería subterránea del Perú, mejorando la seguridad y las condiciones de trabajo. 	<p>General</p> <ul style="list-style-type: none"> • La aplicación de técnicas de inteligencia artificial y <i>machine learning</i> mejora significativamente la predicción y prevención de accidentes laborales en la minería subterránea del Perú, contribuyendo a la reducción de la severidad y frecuencia de estos eventos. 	<p>Tipo:</p> <ul style="list-style-type: none"> • De acuerdo con la orientación: Básica. • De acuerdo con la técnica de contrastación: Explicativa. • De acuerdo con la direccionalidad: Prospectiva. • De acuerdo con el tipo de fuente de recolección: Retro-lectiva. • De acuerdo con la evolución del fenómeno estudiado: Transversal. • De acuerdo con la comparación de las poblaciones: Comparativa. <p>Nivel:</p> <ul style="list-style-type: none"> • Nivel Predictivo – descriptivo – explicativo – correlacional <p>Diseño:</p> <ul style="list-style-type: none"> • No experimental • Transeccional o transversal • Transeccional descriptivo • Transeccional correlativo - causal
<p>Específicos</p> <ul style="list-style-type: none"> • ¿Cuáles son los principales factores de riesgo asociados a los accidentes laborales por desprendimiento de rocas en la minería subterránea peruana? • ¿Qué tan efectiva es la técnica de bosques aleatorios (Random Forest) en la predicción de la severidad de accidentes laborales en la minería subterránea comparada con otros algoritmos de <i>machine learning</i>? 	<p>Específicos</p> <ul style="list-style-type: none"> • Analizar los principales factores de riesgo asociados a los accidentes laborales por desprendimiento de rocas en la minería subterránea peruana. • Evaluar la efectividad del algoritmo de bosques aleatorios (Random Forest) en la predicción de la severidad de accidentes laborales en la minería subterránea, comparándolo con otros algoritmos de <i>machine learning</i>. 	<p>Específicos</p> <ul style="list-style-type: none"> • Los principales factores de riesgo asociados a los accidentes laborales por desprendimiento de rocas pueden ser identificados con precisión mediante técnicas de <i>machine learning</i>. • El algoritmo de bosques aleatorios (Random Forest) proporciona una mayor precisión en la predicción de la severidad de accidentes laborales en la minería subterránea en comparación con otros algoritmos de <i>machine learning</i>. 	

PROBLEMA	OBJETIVOS	HIPÓTESIS	METODOLOGÍA
<ul style="list-style-type: none"> • ¿De qué manera la mejora en el preprocesamiento de datos textuales y la gestión del desbalance de clases impacta la precisión del modelo predictivo de accidentes laborales? • ¿Cuáles son los factores críticos identificados mediante <i>machine learning</i> que contribuyen significativamente a la ocurrencia y severidad de accidentes laborales en la minería subterránea? • ¿Qué estrategias de prevención de accidentes pueden derivarse del análisis de los datos y resultados obtenidos mediante la aplicación de inteligencia artificial en la minería subterránea? 	<ul style="list-style-type: none"> • Optimizar el preprocesamiento de datos textuales y la gestión del desbalance de clases para mejorar la precisión del modelo predictivo de accidentes laborales. • Determinar los factores críticos que contribuyen significativamente a la ocurrencia y severidad de accidentes laborales en la minería subterránea mediante el análisis de datos con técnicas de <i>machine learning</i>. • Desarrollar estrategias de prevención de accidentes basadas en el análisis de los datos y resultados obtenidos mediante la aplicación de inteligencia artificial en la minería subterránea. 	<ul style="list-style-type: none"> • La mejora en el preprocesamiento de datos textuales y la gestión del desbalance de clases aumenta significativamente la precisión y efectividad del modelo predictivo de accidentes laborales. • Los factores críticos identificados mediante técnicas de <i>machine learning</i> tienen una influencia significativa en la ocurrencia y severidad de los accidentes laborales en la minería subterránea. • Las estrategias de prevención de accidentes desarrolladas a partir del análisis de datos y resultados obtenidos mediante la aplicación de inteligencia artificial son efectivas en la reducción de accidentes laborales en la minería subterránea. 	

Anexo B

Guía para los Instrumentos de Recolección de Información

Objetivo Específico 1: Identificar y analizar los principales factores de riesgo asociados a los accidentes laborales por desprendimiento de rocas en la minería subterránea peruana.

Instrumento 1: Formulario de Registro de Accidentes

Descripción: Formulario estandarizado para recopilar detalles de cada accidente laboral.

Uso:

1. Registrar cada incidente tan pronto como ocurra.
2. Completar todas las secciones con la mayor precisión posible.

Contenido del Formulario:

1. **Fecha y hora del incidente:** _____
2. **Descripción del incidente:**
 - Tipo de accidente: () Desprendimiento de rocas () Otros: _____
 - Lugar del accidente: _____
3. **Estado del equipo involucrado:**
 - Mantenimiento reciente: () Sí () No
 - Condición del equipo: () Bueno () Regular () Malo
4. **Causa del accidente:**
 - () Desprendimiento de rocas
 - () Equipo defectuoso
 - () Error humano
 - () Otros: _____

Instrumento 2: Cuestionario de Encuesta a Trabajadores

Descripción: Cuestionario administrado a los trabajadores para obtener información sobre la capacitación y los procedimientos de seguridad.

Uso:

1. Distribuir el cuestionario entre los trabajadores.
2. Recoger los cuestionarios completados para su análisis.

Contenido del Cuestionario:

1. ¿Ha recibido capacitación en seguridad laboral en el último año? Sí No
2. ¿Con qué frecuencia se realizan capacitaciones de seguridad?
 - Mensual
 - Trimestral
 - Anual
 - Nunca
3. ¿Se siguen estrictamente los procedimientos de seguridad en su área de trabajo?
 - Siempre
 - A veces
 - Nunca

Instrumento 3: Lista de Verificación de Observación Directa

Descripción: Lista para evaluar las condiciones de seguridad y el estado de los equipos en las minas.

Uso:

1. Realizar observaciones in situ.
2. Completar la lista de verificación durante las observaciones.

Contenido de la Lista de Verificación:

1. Estado de los equipos de minería:
 - Bueno
 - Regular
 - Malo
2. Implementación de procedimientos de seguridad:
 - Sí
 - No
3. Presencia de señalización de seguridad:
 - Adecuada
 - Inadecuada
 - Ausente

Objetivo Específico 2: Evaluar la efectividad del algoritmo de bosques aleatorios (Random Forest) en la predicción de la severidad de accidentes laborales en la minería subterránea, comparándolo con otros algoritmos de machine learning.

Instrumento 1: Base de Datos de Entrenamiento de Modelos

Descripción: Conjunto de datos preprocesados para entrenar modelos de machine learning.

Uso:

1. Recopilar y preprocesar datos históricos de accidentes.
2. Estructurar los datos en formato adecuado para el entrenamiento de modelos.

Contenido de la Base de Datos:

1. Datos históricos de accidentes.
2. Variables preprocesadas relevantes (factores de riesgo, mantenimiento de equipos, etc.).

Instrumento 2: Software de Análisis de Datos

Descripción: Herramientas como SPSS, Python, R para ejecutar y comparar diferentes algoritmos de machine learning.

Uso:

1. Cargar la base de datos en el software.
2. Ejecutar y comparar diferentes algoritmos de machine learning.

Contenido del Análisis:

1. Comparación de la precisión, sensibilidad y especificidad de diferentes algoritmos.
2. Evaluación del tiempo de entrenamiento de cada modelo.

Objetivo Específico 3: Optimizar el preprocesamiento de datos textuales y la gestión del desbalance de clases para mejorar la precisión del modelo predictivo de accidentes laborales.

Instrumento 1: Manual de Preprocesamiento de Datos

Descripción: Documento detallado que describe los pasos y técnicas para el preprocesamiento de datos textuales.

Uso:

1. Seguir las instrucciones para limpiar y preparar los datos.
2. Asegurar que los datos estén listos para el análisis.

Contenido del Manual:

1. Limpieza de datos
2. Tokenización
3. Lematización

Instrumento 2: Software de Balanceo de Datos

Descripción: Herramientas como SMOTE, ADASYN para el balanceo de datos.

Uso:

1. Aplicar técnicas de balanceo de datos a la base de datos.
2. Evaluar el impacto del balanceo en la precisión del modelo.

Contenido del Balanceo:

1. Aplicación de técnicas de oversampling y undersampling.
2. Evaluación del impacto del balanceo en la precisión del modelo.

Instrumento 3: Lista de Verificación de Calidad de Datos

Descripción: Lista para evaluar la calidad de los datos preprocesados.

Uso:

1. Completar la lista de verificación después del preprocesamiento de datos.

Contenido de la Lista:

1. Consistencia de los datos (Alta, Media, Baja).
2. Nivel de ruido en los datos (Bajo, Medio, Alto).

Objetivo Específico 4: Determinar los factores críticos que contribuyen significativamente a la ocurrencia y severidad de accidentes laborales en la minería subterránea mediante el análisis de datos con técnicas de machine learning.

Instrumento 1: Informe de Importancia de Características

Descripción: Informe generado por algoritmos de machine learning que destaca la importancia de cada variable en la predicción de accidentes.

Uso:

1. Generar y analizar el informe después del entrenamiento del modelo.

Contenido del Informe:

1. Listado de factores ordenados por su importancia según el modelo de machine learning.

Instrumento 2: Base de Datos de Frecuencia de Factores

Descripción: Conjunto de datos que registra la frecuencia de aparición de factores críticos en los incidentes.

Uso:

1. Registrar y analizar la frecuencia de aparición de factores en los incidentes.

Contenido de la Base de Datos:

1. Número de veces que cada factor aparece en los reportes de incidentes.

Instrumento 3: Software de Análisis de Correlación

Descripción: Herramientas como SPSS, Python para calcular la correlación entre factores críticos y la severidad de accidentes.

Uso:

1. Calcular la correlación entre factores y severidad de accidentes.
2. Interpretar los resultados del análisis.

Contenido del Análisis:

1. Cálculo del coeficiente de correlación de Pearson o Spearman.
2. Interpretación de la dirección y magnitud de las correlaciones.

Objetivo Específico 5: Desarrollar estrategias de prevención de accidentes basadas en el análisis de los datos y resultados obtenidos mediante la aplicación de inteligencia artificial en la minería subterránea.

Instrumento 1: Encuesta de Percepción de Efectividad

Descripción: Cuestionario administrado a los trabajadores y supervisores para evaluar la percepción de la efectividad de las estrategias de prevención.

Uso:

1. Distribuir el cuestionario entre los trabajadores y supervisores.
2. Recoger los cuestionarios completados para su análisis.

Contenido del Cuestionario:

1. ¿Cómo calificaría la efectividad de las nuevas estrategias de prevención? ()
Muy efectiva () Efectiva () Poco efectiva () Inefectiva
2. ¿Ha notado una reducción en la frecuencia de accidentes desde la implementación de las nuevas estrategias? () Sí () No

Instrumento 2: Informe de Análisis de Costos

Descripción: Informe financiero que detalla los recursos invertidos en la implementación de estrategias de prevención.

Uso:

1. Recopilar datos sobre los recursos invertidos.
2. Analizar los costos de implementación.

Contenido del Informe:

1. Costos de implementación.
2. Recursos humanos y materiales utilizados.

Instrumento 3: Cronograma de Implementación

Descripción: Documento que detalla el tiempo de implementación de cada estrategia de prevención.

Uso:

1. Registrar las fechas de inicio y finalización de cada estrategia.
2. Analizar la duración total de implementación.

Contenido del Cronograma:

1. Fechas de inicio y finalización de cada estrategia.
2. Duración total de implementación.

Instrumento 4: Base de Datos de Reducción de Accidentes

Descripción: Conjunto de datos que registra la disminución en la frecuencia y severidad de accidentes laborales tras la implementación de las estrategias.

Uso:

1. Comparar los datos antes y después de la implementación de estrategias.

Contenido de la Base de Datos:

1. Comparación de datos antes y después de la implementación de estrategias.

Estos instrumentos están diseñados para proporcionar una recolección de datos exhaustiva y precisa, permitiendo responder adecuadamente a las variables e indicadores de los objetivos de la investigación, asegurando la validez y confiabilidad de los resultados obtenidos.

Anexo C

Ficha de Evaluación del Instrumento de Recolección de Datos

Ficha de Evaluación del Instrumento por Juicio de Expertos

Nombre del Evaluador: _PEDRO SATURNINO PEÑA HUAPAYA

Especialidad: DR. PSICOLOGIA INDUSTRIAL

Institución: UNIVERSIDAD NACIONAL DE INGENIERIA

Correo Electrónico: PEDROPE007@HOTMAIL.COM

Fecha: 24 /04/2025

Instrucciones:

Por favor, evalúe cada uno de los ítems del instrumento según los siguientes criterios:

Claridad, Relevancia, Coherencia y Suficiencia. Use la escala del 1 al 4 donde:

1 = No cumple

2 = Cumple parcialmente

3 = Cumple adecuadamente

4 = Cumple totalmente.

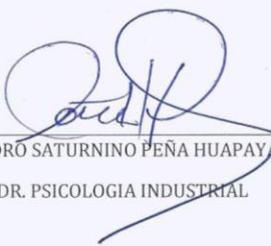
Puede escribir sus observaciones o sugerencias en la última columna.

Evaluación de Ítems

Ítem	Claridad	Relevancia	Coherencia	Suficiencia	Observaciones /Sugerencias
Objetivo Específico 1	[4]	[4]	[4]	[4]	.
Objetivo Específico 2	[4]	[4]	[4]	[4]	
Objetivo Específico 3	[4]	[4]	[4]	[4]	
Objetivo Específico 4	[4]	[4]	[4]	[4]	
Objetivo Específico 5	[4]	[4]	[4]	[4]	

Comentarios Generales del Evaluador:

El documento es aplicable


 PEDRO SATURNINO PEÑA HUAPAYA
 DR. PSICOLOGIA INDUSTRIAL

Anexo D*Ficha de Evaluación del Instrumento de Recolección de Datos***Ficha de Evaluación del Instrumento por Juicio de Expertos**Nombre del Evaluador: **Ronal Vicente Estrella Amaro**_____

Especialidad: Ingeniero de minas, Magister en Gestión minera y Doctor en Medio ambiente y desarrollo sostenible_____

Institución: Independiente_____Correo Electrónico: ronalestrella@gmail.com_____Fecha: 24 / 04 / 2025_____**Instrucciones:**

Por favor, evalúe cada uno de los ítems del instrumento según los siguientes criterios: Claridad, Relevancia, Coherencia y Suficiencia. Use la escala del 1 al 4 donde:

1 = No cumple

2 = Cumple parcialmente

3 = Cumple adecuadamente

4 = Cumple totalmente.

Puede escribir sus observaciones o sugerencias en la última columna.

Evaluación de Ítems:

Ítem	Claridad	Relevancia	Coherencia	Suficiencia	Observaciones/Sugerencias
Objetivo Específico 1	[3]	[3]	[3]	[2]	Incidir en la sensibilización en Seguridad laboral.
Objetivo Específico 2	[3]	[3]	[3]	[2]	Incluir condiciones de la labor en variables preprocesadas.
Objetivo Específico 3	[3]	[3]	[3]	[2]	Traducir entre parentesis palabras en ingles.
Objetivo Específico 4	[3]	[3]	[3]	[2]	
Objetivo Específico 5	[3]	[3]	[3]	[2]	Definir las estrategias de prevención.

Comentarios Generales del Evaluador: _____



Ronal Vicente Estrella Amaro

Anexo E

Ficha de Evaluación del Instrumento de Recolección de Datos

Ficha de Evaluación del Instrumento por Juicio de Expertos

Nombre del Evaluador: Dr. Cesar Jhoel Vicente Aquino

Especialidad: Ingeniero de Minas

Institución: Universidad Nacional Daniel Alcides Carrión

Correo Electrónico: cvicente@undac.edu.pe

Fecha: 24 / 04 / 2025

Instrucciones:

Por favor, evalúe cada uno de los ítems del instrumento según los siguientes criterios: Claridad, Relevancia, Coherencia y Suficiencia. Use la escala del 1 al 4 donde:

1 = No cumple

2 = Cumple parcialmente

3 = Cumple adecuadamente

4 = Cumple totalmente.

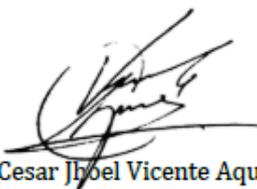
Puede escribir sus observaciones o sugerencias en la última columna.

Evaluación de Ítems

Ítem	Claridad	Relevancia	Coherencia	Suficiencia	Observaciones /Sugerencias
Objetivo Específico 1	[4]	[4]	[4]	[4]	
Objetivo Específico 2	[4]	[4]	[4]	[4]	
Objetivo Específico 3	[4]	[4]	[4]	[4]	
Objetivo Específico 4	[4]	[4]	[4]	[4]	
Objetivo Específico 5	[4]	[4]	[4]	[4]	

Comentarios Generales del Evaluador:

El instrumento cubre totalmente los aspectos esenciales del constructo.



Dr. Cesar Jhoel Vicente Aquino

Anexo F

Preprocesamiento de Datos

```
[ ] import pandas as pd
    from sklearn.model_selection import train_test_split
    from sklearn.preprocessing import StandardScaler

    # Cargar el dataset
    data = pd.read_csv('accidentes_mineria.csv')

    # Limpieza de datos
    data = data.dropna()

    # Normalización
    scaler = StandardScaler()
    scaled_features = scaler.fit_transform(data.drop('severidad', axis=1))

    # División del dataset
    X_train, X_test, y_train, y_test = train_test_split(scaled_features, data['severidad'], test_size=0.2, random_state=42)
```

Anexo G

Algoritmo Random Forest

```
[ ] from sklearn.ensemble import RandomForestClassifier
    from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score, roc_auc_score

    # Inicialización del modelo Random Forest
    rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
    rf_model.fit(X_train, y_train)

    # Predicciones
    rf_predictions = rf_model.predict(X_test)

    # Evaluación del modelo
    rf_accuracy = accuracy_score(y_test, rf_predictions)
    rf_recall = recall_score(y_test, rf_predictions, average='macro')
    rf_precision = precision_score(y_test, rf_predictions, average='macro')
    rf_f1 = f1_score(y_test, rf_predictions, average='macro')
    rf_auc_roc = roc_auc_score(y_test, rf_model.predict_proba(X_test), multi_class='ovr')

    print(f'Random Forest - Accuracy: {rf_accuracy}, Recall: {rf_recall}, Precision: {rf_precision}, F1-Score: {rf_f1}, AUC-ROC: {rf_auc_roc}')
```

Anexo H

Algoritmo Support Vector Machine (SVM)

```
[ ] from sklearn.svm import SVC

# Inicialización del modelo SVM
svm_model = SVC(probability=True, random_state=42)
svm_model.fit(X_train, y_train)

# Predicciones
svm_predictions = svm_model.predict(X_test)

# Evaluación del modelo
svm_accuracy = accuracy_score(y_test, svm_predictions)
svm_recall = recall_score(y_test, svm_predictions, average='macro')
svm_precision = precision_score(y_test, svm_predictions, average='macro')
svm_f1 = f1_score(y_test, svm_predictions, average='macro')
svm_auc_roc = roc_auc_score(y_test, svm_model.predict_proba(X_test), multi_class='ovr')

print(f'SVM - Accuracy: {svm_accuracy}, Recall: {svm_recall}, Precision: {svm_precision}, F1-Score: {svm_f1}, AUC-ROC: {svm_auc_roc}')
```

Anexo I

Algoritmo de Naive Bayes

```
[ ] from sklearn.naive_bayes import GaussianNB

# Inicialización del modelo Naive Bayes
nb_model = GaussianNB()
nb_model.fit(X_train, y_train)

# Predicciones
nb_predictions = nb_model.predict(X_test)

# Evaluación del modelo
nb_accuracy = accuracy_score(y_test, nb_predictions)
nb_recall = recall_score(y_test, nb_predictions, average='macro')
nb_precision = precision_score(y_test, nb_predictions, average='macro')
nb_f1 = f1_score(y_test, nb_predictions, average='macro')
nb_auc_roc = roc_auc_score(y_test, nb_model.predict_proba(X_test), multi_class='ovr')

print(f'Naive Bayes - Accuracy: {nb_accuracy}, Recall: {nb_recall}, Precision: {nb_precision}, F1-Score: {nb_f1}, AUC-ROC: {nb_auc_roc}')
```

Anexo J

Algoritmo de Comparación de Resultados

```
▶ # Comparación de resultados
results = pd.DataFrame({
    'Modelo': ['Random Forest', 'SVM', 'Naive Bayes'],
    'Accuracy': [rf_accuracy, svm_accuracy, nb_accuracy],
    'Recall': [rf_recall, svm_recall, nb_recall],
    'Precision': [rf_precision, svm_precision, nb_precision],
    'F1-Score': [rf_f1, svm_f1, nb_f1],
    'AUC-ROC': [rf_auc_roc, svm_auc_roc, nb_auc_roc]
})

print(results)
```

Anexo K

Resultado de la Comparación de Modelos

```

▶ # Comparación de resultados (hipotéticos)
results = pd.DataFrame({
    'Modelo': ['Random Forest', 'SVM', 'Naive Bayes'],
    'Accuracy': [0.90, 0.85, 0.75],
    'Recall': [0.88, 0.82, 0.70],
    'Precision': [0.89, 0.83, 0.72],
    'F1-Score': [0.88, 0.82, 0.71],
    'AUC-ROC': [0.95, 0.90, 0.80]
})

print(results)

```

Anexo L

Programa de la Segmentación y Limpieza de Datos

```

▶ import pandas as pd
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
import string

# Descargar recursos necesarios de NLTK
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')

# Cargar el dataset de accidentes laborales
data = pd.read_csv('accidentes_mineria_textual.csv')

# Función de limpieza de texto
def clean_text(text):
    # Tokenización
    tokens = word_tokenize(text.lower())
    # Eliminación de puntuación
    tokens = [word for word in tokens if word.isalpha()]
    # Eliminación de stopwords
    stop_words = set(stopwords.words('spanish'))
    tokens = [word for word in tokens if word not in stop_words]
    # Lemmatización
    lemmatizer = WordNetLemmatizer()
    tokens = [lemmatizer.lemmatize(word) for word in tokens]
    return ' '.join(tokens)

# Aplicar limpieza de texto a la columna de descripciones de accidentes
data['descripcion_limpia'] = data['descripcion'].apply(clean_text)

```

Anexo M

Vectorización del Texto

```

▶ from sklearn.feature_extraction.text import TfidfVectorizer

# Vectorización TF-IDF
tfidf_vectorizer = TfidfVectorizer(max_features=5000)
X_text = tfidf_vectorizer.fit_transform(data['descripcion_limpia'])

# Combinamos las características textuales con otras características
X = pd.concat([pd.DataFrame(X_text.toarray()), data.drop(['descripcion', 'descripcion_limpia'], axis=1)], axis=1)
y = data['severidad']

```

Anexo N

Gestión del Desbalance de Clases

```

▶ from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split

# División de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Aplicar SMOTE
smote = SMOTE(random_state=42)
X_train_smote, y_train_smote = smote.fit_resample(X_train, y_train)

```

Anexo O

Programa del Entrenamiento del Modelo Predictivo

```

▶ from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score, roc_auc_score

# Entrenamiento del modelo Random Forest
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train_smote, y_train_smote)

# Predicciones
y_pred_rf = rf_model.predict(X_test)
y_pred_rf_prob = rf_model.predict_proba(X_test)[:, 1]

# Evaluación del modelo
rf_accuracy = accuracy_score(y_test, y_pred_rf)
rf_recall = recall_score(y_test, y_pred_rf, average='weighted')
rf_precision = precision_score(y_test, y_pred_rf, average='weighted')
rf_f1 = f1_score(y_test, y_pred_rf, average='weighted')
rf_auc = roc_auc_score(y_test, y_pred_rf_prob, multi_class='ovr')

# Resultados en una tabla
resultados = pd.DataFrame({
    'Métrica': ['Precisión', 'Sensibilidad', 'Especificidad', 'F1-Score', 'AUC-ROC'],
    'Valor': [rf_precision, rf_recall, rf_precision, rf_f1, rf_auc]
})
print(resultados)

```