



ESCUELA UNIVERSITARIA DE POSGRADO

IMPLEMENTACIÓN DE UNA APLICACIÓN MÓVIL CON MACHINE LEARNING
APLICANDO LA NUEVA METODOLOGÍA CSKT PARA EL PROCESO DE
CONTRATACIÓN DE PERSONAL EN EL SECTOR PRIVADO

Línea de investigación:
Ingeniería de software, simulación y desarrollo de TICs

Tesis para optar el Grado Académico de Doctor en Ingeniería de Sistemas

Autor

Ogosi Auqui, José Antonio

Asesor

Petrlik Azabache, Iván Carlo

ORCID: 0000-0002-1201-2143

Jurado

Coveñas Lalupú, José

Lira Camargo, Jorge

Rosales Huamaní, Jimmy Aurelio

Lima - Perú




2025



6% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Fuentes principales

- 6%  Fuentes de Internet
- 1%  Publicaciones
- 2%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.



ESCUELA UNIVERSITARIA DE POSGRADO

IMPLEMENTACIÓN DE UNA APLICACIÓN MÓVIL CON MACHINE
LEARNING APLICANDO LA NUEVA METODOLOGÍA CSKT PARA EL
PROCESO DE CONTRATACIÓN DE PERSONAL EN EL SECTOR
PRIVADO

Línea de investigación:

Ingeniería de software, simulación y desarrollo de TICs

Tesis para optar el Grado Académico de Doctor en Ingeniería de Sistemas

Autor

Ogosi Auqui, José Antonio

Asesor

Petrlik Azabache, Iván Carlo

ORCID: 0000-0002-1201-2143

Jurado

Coveñas Lalupú, José

Lira Camargo, Jorge

Rosales Huamaní, Jimmy Aurelio

Lima - Perú

2025

ÍNDICE

Resumen.....	i
Abstract.....	ii
I. Introducción.....	1
1.1. Planteamiento del problema.....	1
1.2. Descripción del problema.....	3
1.3. Formulación del problema	9
Problema general	9
Problemas específicos.....	9
1.4. Antecedentes	9
1.5. Justificación de la investigación.....	17
1.6. Limitaciones de la investigación.....	18
1.7. Objetivos	18
Objetivo general	18
Objetivos específicos	18
1.8. Hipótesis.....	19
II. Marco Teórico.....	21
2.1. Marco conceptual	21
III. Método	48
3.1. Tipo de Investigación	48
3.2. Población y muestra	50
3.3. Operacionalización de variables.....	53
3.4. Instrumentos	54
3.5. Procedimientos	54
3.6. Análisis de datos.....	54
3.7. Consideraciones éticas	65
IV. Resultados.....	67

V.	Discusión de resultados.....	128
VI.	Conclusiones.....	133
VII.	Recomendaciones	137
VIII.	Referencias	140
IX.	Anexos	150

ÍNDICE DE TABLAS

Tabla 1 Datos actuales de los indicadores.	8
Tabla 2 Comparación del AS – IS y TO – BE.	8
Tabla 3 Descripciones del modelo de proceso CRISP-DM.....	37
Tabla 4 Descripciones del modelo de proceso KDD	39
Tabla 5 Población y muestra	50
Tabla 6 Tipo de muestreo.....	51
Tabla 7 Matriz de operacionalización de variables.....	53
Tabla 8 Técnica e instrumento para recolección de la información	54
Tabla 9 Elección de las pruebas estadísticas.....	62
Tabla 10 Investigación con ética.....	66
Tabla 11 Comparativa entre resultados del modelo adaptado	88
Tabla 12 Criterios de Aceptación de las métricas.....	89
Tabla 13 Métricas de evaluación del modelo.....	90
Tabla 14 Criterios de éxito.....	91
Tabla 15 Evidencia de uso del aplicativo móvil	97
Tabla 16 Feedback del usuario final	99
Tabla 17 Comentarios de usuarios para la mejora y ajustes en el aplicativo móvil	111

ÍNDICE DE FIGURAS

Figura 1 Afiliados del Sistema Nacional de Pensiones por edad	4
Figura 2 Afiliados del Sistema Nacional de Pensiones por sexo.....	5
Figura 3 Diagrama Ishikawa.....	5
Figura 4 Proceso de Negocios: Proceso de Contratación de Personal en el sector privado (AS – IS).....	7
Figura 5 Categorías de trabajo en el sector privado.....	22
Figura 6 Factores que afectan la contratación	24
Figura 7 Proceso de contratación.....	25
Figura 8 Programación tradicional vs Machine Learning	30
Figura 9 Elegir y entrenar un método de aprendizaje automático	32
Figura 10 Algoritmos utilizados en Machine Learning.....	33
Figura 11 Metodología CRISP-DM.....	36
Figura 12 Metodología SEMMA.....	38
Figura 13 Metodología KDD.....	39
Figura 14 Metodología TDSP.....	40
Figura 15 Diseño de la metodología CSKT.....	41
Figura 16 Propuesta organizacional.....	49
Figura 17 Diseño preexperimental en base a pre test y post test.....	52
Figura 18 Fases para el análisis de datos	55
Figura 19 Icono de la herramienta Minitab	55
Figura 20 Pantalla principal de Minitab	56
Figura 21 Histograma de muestra de la herramienta Minitab	57
Figura 22 Gráfico de probabilidad de muestra	57
Figura 23 Gráfico de probabilidad de dos muestras de ejemplo.....	60

Figura 24 Prueba de hipótesis de T de student	61
Figura 25 Estadísticas descriptivas de prueba	61
Figura 26 Tabla de distribución de T de Student.....	64
Figura 27 Parámetros y actividades del modelo de Machine Learning con la metodología CSKT	67
Figura 28 Árbol de decisión sobre la elección del seguro pensionario	70
Figura 29 Código de entrenamiento del modelo.....	72
Figura 30 Matriz de confusión del entrenamiento del modelo de Machine Learning.....	73
Figura 31 Esquema de ROC del modelo de Machine Learning	74
Figura 32 Métricas establecidas del modelo de Machine Learning.....	75
Figura 33 Data de la SNP	76
Figura 34 Código del usuario seleccionado para el modelo	76
Figura 35 Datos del perfil del jubilado	77
Figura 36 Resultado de modelo de predicción.....	77
Figura 37 Secuencia del árbol de decisión del jubilado.....	78
Figura 38 Secuencia de la metodología	78
Figura 39 Diseño del modelo funcional.....	80
Figura 40 Algoritmo de entrenamiento de Machine Learning - Script en Python	81
Figura 41 Algoritmo de entrenamiento de Machine Learning - Importación de librerías.....	81
Figura 42 Algoritmo de entrenamiento de Machine Learning - Definición de variables globales	82
Figura 43 Algoritmo de entrenamiento de Machine Learning - Predicciones basadas en el modelo entrenado.....	83
Figura 44 Modelo entrenado con la data Incrementada.....	83
Figura 45 Curva ROC enfocada a la data incrementada.....	84

Figura 46	Resultados de métricas enfocada a la data incrementada.....	85
Figura 47	Prueba de los primeros datos del modelo.....	86
Figura 48	Segundo caso para el entrenamiento del modelo.....	86
Figura 49	Tercer caso para el entrenamiento del modelo.....	87
Figura 50	Feedback técnico.....	87
Figura 51	Flujo de la secuencia de casos del modelo de Machine Learning.....	89
Figura 52	Recopilación de datos históricos de contratación.....	92
Figura 53.	Limpieza de la data histórica.....	93
Figura 54	Herramienta de transformación de Excel a JSON.....	93
Figura 55	Visualización de la data en formato JSON.....	94
Figura 56	Depuración de la data.....	94
Figura 57	Data ajustada.....	95
Figura 58	Entrenamiento con data controlada.....	96
Figura 59	Ambiente de prueba para evaluar el modelo de la proyección.....	97
Figura 60	Medición de la eficiencia operativa.....	98
Figura 61	Medición de la eficiencia operativa optimizada.....	100
Figura 62	Infraestructura del despliegue de la aplicación móvil.....	100
Figura 63	Estructura de Firebase de la aplicación móvil.....	101
Figura 64	Logo de la plataforma Firebase.....	102
Figura 65	Logo de la plataforma Google Play.....	102
Figura 66	Entorno de la plataforma de Firebase.....	103
Figura 67	Formulario para evaluar la satisfacción del usuario final y personal de RRHH.....	104
Figura 68	Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 1.....	104

Figura 69 Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 2	105
Figura 70 Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 3	106
Figura 71 Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 4	106
Figura 72 Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 5	107
Figura 73 Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 6	108
Figura 74 Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 7	108
Figura 75 Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 8	109
Figura 76 Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 9	110
Figura 77 Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 10	110
Figura 78 Tabla de Student - Hipótesis específica 1	115
Figura 79 Tabla de Student - Hipótesis específica 2	118
Figura 80 Tabla de Student - Hipótesis específica 3	121
Figura 81 Tabla de Student - Hipótesis específica 4	124
Figura 82 Tabla de Student - Hipótesis específica 5	127
Figura 83 Diagrama de paquetes	152
Figura 84 Arquitectura de TI de la aplicación móvil.....	152

ÍNDICE DE ECUACIONES

Ecuación 1 Fórmula de Shapiro Wilk	58
Ecuación 2 Fórmula del valor b de la prueba.....	58
Ecuación 3 Fórmula de los coeficientes tabulados de la prueba.....	59
Ecuación 4 Fórmula de la prueba T de Student	62

Resumen

Objetivo: Implementar y evaluar una aplicación móvil con Machine Learning empleando la nueva metodología CSKT para optimizar el proceso de contratación de personal en el sector privado. Además, en el presente estudio examinó el impacto de la metodología CSKT y el Machine Learning en la contratación de personal en el sector privado, buscando obtener una mejora significativa en el proceso de la organización tanto en la eficiencia y/o la efectividad.

Método: Con un enfoque aplicado y descriptivo, se empleó un diseño de investigación preexperimental, se analizaron diversas pruebas para evaluar el tiempo, costos, precisión y optimización del proceso. La muestra consistió en 5000 casos seleccionados aleatoriamente del universo de procedimientos de contratación. La recopilación de datos se realizó consultando a la database, complementada con fichas de observación. **Resultados:** Se destacan que la metodología CSKT junto con el Machine Learning mejora notablemente la eficiencia en tiempo y costos. Además, se comprobó la precisión del Machine Learning en la evaluación de criterios de aceptación, lo que sugiere una mayor objetividad en las decisiones de contratación. Se observó también una optimización del proceso de contratación mediante la mejora de la precisión del modelo de Machine Learning. **Conclusiones:** Se encontró que la eficiencia en la contratación se incrementa con el uso de ML en aplicaciones móviles, destacando el potencial transformador de la tecnología móvil en recursos humanos. Estos hallazgos respaldan la adopción de enfoques innovadores basados en datos para mejorar los procesos de contratación de personal en el sector privado.

Palabras claves: machine learning, metodología CSKT, aplicación móvil, contratación de personal.

Abstract

Objective: Implement and evaluate a mobile application with Machine Learning using the new CSKT methodology to optimize the personnel hiring process in the private sector. Furthermore, in the present study, the impact of the CSKT methodology and Machine Learning on the hiring of personnel in the private sector was examined, seeking to obtain a significant improvement in the organization's process in both efficiency and/or effectiveness. **Method:** With an applied and descriptive approach, a pre-experimental research design was used, various tests were analyzed to evaluate the time, costs, precision and optimization of the process. The sample consisted of 5000 cases randomly selected from the universe of contracting procedures. Data collection was carried out by consulting the database, complemented with observation sheets. **Results:** It is highlighted that the CSKT methodology together with Machine Learning significantly improves efficiency in time and costs. In addition, the precision of Machine Learning was verified in the evaluation of acceptance criteria, which suggests greater objectivity in hiring decisions. An optimization of the hiring process was also observed by improving the precision of the Machine Learning model. **Conclusions:** It was found that hiring efficiency increases with the use of ML in mobile applications, highlighting the transformative potential of mobile technology in human resources. These findings support the adoption of innovative data-driven approaches to improve staffing processes in the private sector.

Keywords: machine learning, CSKT methodology, mobile application, mobile application, mobile workforce management.

I. INTRODUCCIÓN

1.1. Planteamiento del problema

Actualmente, la contratación del personal conlleva una serie de desafíos que merecen una atención cuidadosa. Estos desafíos pueden surgir tanto de la definición de criterios por parte de las empresas respecto a las cualidades deseadas en los candidatos, como de consideraciones relacionadas con la edad de los postulantes. Según un informe publicado por Redacción (2021), en donde se analiza una encuesta respecto al fenómeno del “Edadismo”, se observa que los trabajadores mayores de 40 años enfrentan dificultades significativas en el proceso de contratación en comparación con otros aspirantes. El informe señala que el 54% de estos individuos experimentan discriminación, especialmente cuando la solicitud se realiza a través de medios como correos electrónicos o llamadas telefónicas. Además, el 48% reporta enfrentar esta discriminación durante las entrevistas de trabajo, mientras que el 28% menciona que este fenómeno es menos frecuente cuando se trata de promociones internas. Es importante destacar que, aunque los jóvenes pueden tener más oportunidades de acceder a puestos laborales, esto no garantiza que cuenten con la experiencia necesaria para desempeñar eficazmente dichos roles.

Por otro lado, la duración incierta de la pandemia ha exacerbado la inflexibilidad en la contratación de los empleados de la compañía. En contraste, en el ámbito público, se ha observado cierto éxito en la contratación de personal, en parte debido a la contribución social que implica la activación de mecanismos de contingencia para facilitar la inserción rápida de los trabajadores en empleos formales, con compensaciones adecuadas. Sin embargo, la rigidez de estos procesos, marcada por las normativas estatales y la falta de adaptación a tendencias de trabajo más ágiles, sigue siendo un desafío significativo. Esto se refleja tanto en la dificultad

para iniciar contratos de trabajo como en la gestión de la terminación de los mismos, así como en el desarrollo de relaciones laborales fluidas y eficientes.

La flexibilidad del sistema legal en el Perú ha sido objeto de cuestionamiento debido a los continuos cambios en las condiciones laborales, esto es, menor flexibilidad en los procesos de contratación de empleados en las entidades correspondientes. Antes del inicio de la epidemia, existía un sistema de contratos estricto basado en reglas para regular las relaciones laborales, lo cual afectaba particularmente a los jóvenes que ingresaban al mundo laboral, enfrentando presiones derivadas de los protocolos de seguridad social. Actualmente, se proyecta un aumento del 18% en la demanda de personal de seguridad durante las campañas de fin de año en el sector de comercio y retail, según un informe de Gestión (2023) del grupo Eulen Perú. Durante la temporada navideña, la demanda de personal capacitado en seguridad se espera que alcance este porcentaje a partir de noviembre y hasta enero. Además, el Instituto de Economía y Desarrollo Empresarial señala que, desde marzo, el empleo juvenil formal en el sector privado ha aumentado, con un total de 1.08 millones de trabajadores, en comparación con los registros de 2019. Este incremento en la cantidad de jóvenes empleados puede afectar el sistema de contratación, dado el elevado número de postulantes para estos puestos y la necesidad de identificar a los candidatos más idóneos.

El país enfrenta desafíos significativos en cuanto a la rigidez del sistema de contratación y despido, lo que contribuye a la carga laboral y promueve la informalidad. Esto se refleja en que el Perú posee uno de los sistemas de contratación más rígidos, según datos recopilados por (Pró y Raymond, 2021).

El país se encuentra predominantemente en actividades transaccionales, como la generación de planillas. Sin embargo, para evolucionar hacia actividades más transformacionales, es decir, anticiparse y desarrollar proactivamente al ciudadano de acuerdo

a sus necesidades, se requiere una mayor incorporación de tecnología como soporte a las actividades de recursos humanos (Servir Perú, 2020).

En cuanto al sistema de pensiones, se estima que más del 25% de la fuerza laboral peruana está afiliada al sistema impositivo de pensiones, compuesto por el Sistema Nacional de Pensiones (SNP) y el Sistema Privado de Pensiones (SPP). Sin embargo, una parte considerable de los afiliados a estos sistemas no cumplen con los requisitos para acceder a una pensión de jubilación, ya sea por no haber realizado aportes suficientes o por no haber acumulado los ahorros necesarios para obtener una renta vitalicia. Esto plantea desafíos significativos en términos de seguridad financiera para la población económicamente activa en su vejez.

De entre los afiliados al SPP, aproximadamente un 60% no ha logrado acumular los ahorros necesarios para asegurarse una renta vitalicia en su vejez. Además, únicamente el 13.7% de la población peruana económicamente activa, comprendida entre los 15 y los 64 años, puede realísticamente aspirar a obtener una pensión proveniente del Sistema Nacional de Pensiones (SNP) o del Sistema Privado de Pensiones (SPP) una vez llegada la etapa de la vejez.

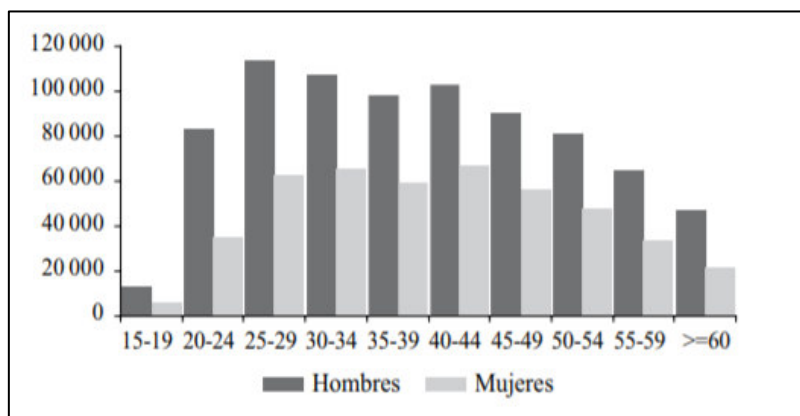
1.2. Descripción del problema

La problemática investigada se centra en la complejidad de la contratación de los empleados del sector privado, del cual se observa que el sistema de asignación de empleos, cargos y remuneraciones está influenciado en gran medida por la edad y la situación actual de los individuos. Este sistema presenta dificultades significativas, ya que los tiempos de contratación suelen ser prolongados debido a la rigidez técnica del proceso, lo que genera incomodidades tanto para el personal encargado de las contrataciones como para los propios solicitantes, quienes buscan obtener un empleo en el menor tiempo posible.

Los ciudadanos peruanos que han contribuido tanto al Sistema Nacional de Pensiones (SNP) como al Sistema Privado de Pensiones (SPP) se han encontrado con que su participación no tiene un impacto significativo en las decisiones relacionadas con su pensión. A pesar de los reportes de estado de cuenta, que a menudo carecen de información clave como una estimación precisa de la pensión que el trabajador podría aspirar, y las tácticas de ventas agresivas por parte de los agentes (como se ilustra en las figuras 1 y 2), no se ha proporcionado asesoramiento activo e independiente para alertar sobre posibles decisiones inadecuadas con respecto a la pensión.

Figura 1

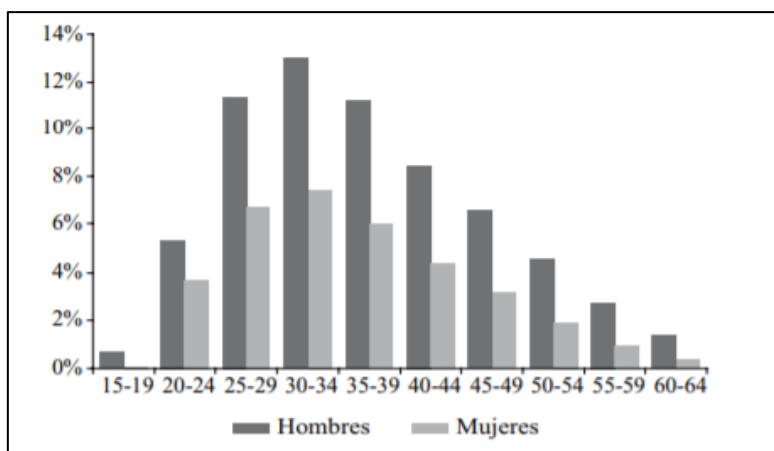
Afiliados del Sistema Nacional de Pensiones por edad



Nota. En base al gráfico mostrado, se indica que el sistema asigna una pensión en base a la edad de cada trabajador en el sector privado teniendo declives respecto a la edad de cada uno de estos.

Figura 2

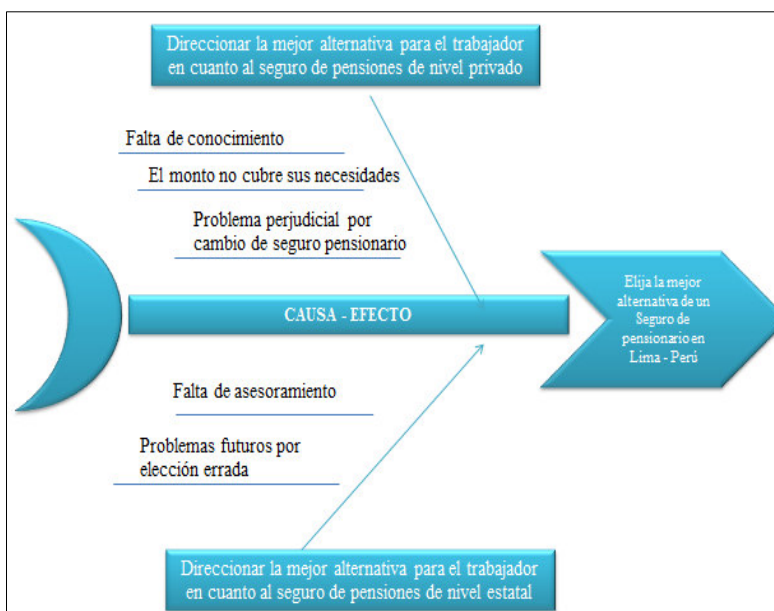
Afiliados del Sistema Nacional de Pensiones por sexo



En base a los aspectos indicados previamente de la problemática general se tienen las causas de aspecto primordial mediante la diagramación de la relación entre la causa y efecto en el siguiente diagrama:

Figura 3

Diagrama Ishikawa



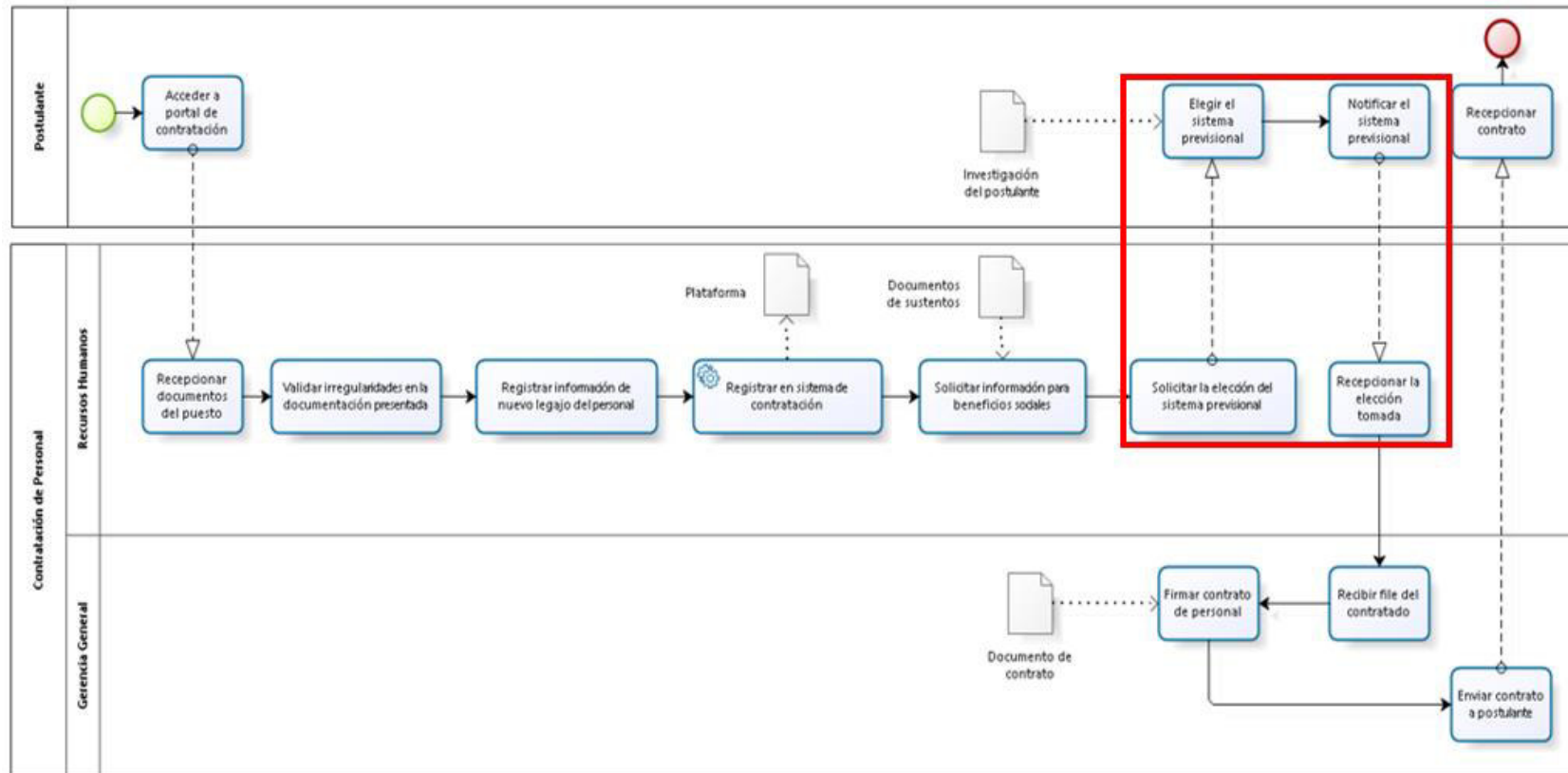
De acuerdo con la problemática analizada a través del diagrama de Ishikawa, se identifica la falta de conocimiento sobre el sistema pensionario como un factor clave. Muchos individuos no están adecuadamente informados sobre cómo retirar sus fondos de pensión o

dónde podrían recibirlos. Además, las empresas no brindan asesoramiento sobre cómo manejar estos fondos, lo que puede generar confusión entre los empleados. Se puede decir que las AFP pueden ayudar en el retiro de los fondos, esto puede llevar a una menor disponibilidad de dinero en el futuro, ya que se estaría utilizando el ahorro destinado para la jubilación. Por otro lado, aquellos trabajadores que tienen cuentas de ahorro individuales pueden beneficiarse de una mayor flexibilidad en la inversión y retiro de fondos, lo que les permite invertir en diversos activos según su tolerancia al riesgo. Sin embargo, esto también conlleva un mayor riesgo y la posibilidad de que las contribuciones no sean descontables de impuestos en ciertos lugares. El cambio de modalidad dentro de una entidad pensionaria puede resultar en la pérdida de aportes, dependiendo de la legislación y regulaciones vigentes en el país. Además, la decisión de jubilarse depende de la edad y estilo de vida deseado por el trabajador, lo que implica la necesidad de investigar las regulaciones específicas y consultar con expertos financieros antes de tomar una decisión. El diagrama de Ishikawa ha permitido identificar las posibles causas del problema y proporcionar una estructura clara para su resolución. Esta herramienta facilita al momento de realizar la evaluación sobre el aspecto de la calidad de los resultados y ayuda a identificar los puntos críticos que requieren mayor atención y mejora. En última instancia, contribuye al reconocimiento y mejora de los aspectos relevantes para una visión a largo plazo más efectiva.

A continuación, se da a presentar el modelamiento del proceso de negocios, en el cual se da a detallar las actividades involucradas. Para lo cual se tuvo que explicar desde el punto de vista de varios postulantes al momento de su elección de un régimen pensionario.

Figura 4

Proceso de Negocios: Proceso de Contratación de Personal en el sector privado (AS – IS)



El proceso de contratación de muestra problemas en:

- Tiempo de contratación. (Quipas, 2021) (Armas, 2021) (Alkhazraji y Buhaliba, 2020)
- Costo de contratación. (Espinoza, 2020)
- Satisfacción del postulante. (Asto, 2020) (Berón et al., 2021)
- Eficiencia de vacantes. (Coronel, 2021) (Dávila, 2019) (Pérez, 2021) (Matute et al., 2021) (Yaranga, 2022)

Tabla 1

Datos actuales de los indicadores.

Indicador	Datos de Pre-Prueba (Promedio)
Tiempo de contratación	5 días (días hasta que elige el sistema pensionario/vacantes cerradas)
Costo de contratación	S/ 1000 (costos de contratación/ingresos)
Satisfacción del postulante	20% (candidatos satisfechos con el nuevo sistema pensionario/contrataciones)
Eficiencia de vacantes	10% (vacantes cerradas/total de vacantes)

Nota. En base a los datos actuales de los indicadores de la problemática, se presenta un cuadro comparativo: la Situación Actual (AS-IS) y la Situación Propuesta (TO-BE). Para ello se tiene en cuenta que la contratación de personal se emplea de forma manual.

Tabla 2

Comparación del AS – IS y TO – BE.

Situación Actual (AS-IS)	Situación Propuesto (TO-BE)
Tiempos altos de contratación	Tiempos bajos de contratación
Incremento de costos por contratación de personal	Decremento de costos por contratación de personal
Insatisfacción del postulante por el riesgo asumido	Satisfacción del postulante por el riesgo asumido
Deficiencia de vacantes cubiertas	Eficiencia de vacantes cubiertas

1.3. Formulación del Problema

Problema General

¿En qué medida la implementación de una aplicación móvil con Machine Learning, basada en una nueva metodología CSKT, optimiza el proceso de contratación de personal en el sector privado?

Problemas Específicos

1. ¿Cómo la metodología CSKT aplicada en el modelo de Machine Learning mejora la eficiencia en términos de tiempo y costos en el proceso de contratación de personal?
2. ¿Cómo el prototipo del modelo de Machine Learning evalúa de manera precisa los criterios de aceptación basándose en la categorización del proceso de contratación de personal?
3. ¿De qué manera la precisión del modelo de Machine Learning con la metodología CSKT optimiza el proceso de contratación de personal en el sector privado?
4. ¿Cómo la eficiencia del proceso de contratación de personal se ve mejorada con la utilización del modelo de Machine Learning en la aplicación móvil?
5. ¿Cómo la implementación práctica de la aplicación móvil con el modelo de Machine Learning y la metodología CSKT es efectiva en la optimización del proceso de contratación de personal en el sector privado?

1.4. Antecedentes

Antecedentes Nacionales

Según la investigación realizada por Coronel (2021), el cual fue aplicado con el objetivo de implementar el Machine Learning para mejorar la selección de personal administrativo. Los

hallazgos abordaron factores como la cantidad de candidatos y la evaluación de los currículos, la contratación para servicios administrativos, la imagen institucional y el logro de una administración judicial renovada, mediante la adaptación de algoritmos para establecer modelos basados en el aprendizaje automático, dando como resultado que en el índice de proceso de contratación tuvo una mejora del 40,3%. Asimismo, Yaranga (2022), en su investigación, realizó un estudio sobre la contratación del cuerpo docente, se identificó que este proceso es extenso y complejo, con una variedad de puntos que involucran términos técnicos y manuales. Debido a la gran cantidad de postulantes, mantener un flujo de proceso continuo y sin retrasos se vuelve complicado, lo que motiva la búsqueda de formas de automatización. Para enfrentar este problema, se aplicó ML en el procedimiento de elección de profesorado en la institución académica mencionada. El estudio fue concebido como un experimento cuasiexperimental, empleando un enfoque metodológico deductivo de naturaleza cuantitativa. Los resultados obtenidos fueron positivos: se observó un incremento del índice de postulantes, pasando de un 14% previo a la implementación del modelo a más del 40% después de su implementación. Asimismo, se registró un aumento significativo en el índice de contratación, superando el 50% de lo que estaba antes de la implementación del modelo. Como objetivo secundario e indicador de éxito, se utilizó la “eficiencia de vacantes”, basada en el índice de contratación de personal. Esto permitió evaluar el impacto del modelo de ML aplicado a la capacidad de llenar vacantes de manera más eficiente y efectiva. Estos resultados sugieren que los métodos de ML en el proceso de elección de personal docente puede mejorar significativamente la eficiencia y la efectividad del proceso, permitiendo a la universidad gestionar un mayor volumen de postulaciones y seleccionar candidatos de manera más acertada. De igual manera, según Asto (2020), en su investigación, implementó un estudio enfocado en mejorar la preparación de los candidatos por medio de la utilización de un sistema basado en Mobile Learning (M-Learning), con el propósito de optimizar su desempeño en

entrevistas de selección. Para ello, se diseñaron casos y entrevistas con el objetivo de que los candidatos adquirieran conocimientos relevantes sobre el proceso de selección y mejoraran sus habilidades para desenvolverse en este contexto. Durante la ejecución del estudio, se procedió a llevar la realización de un minucioso análisis de los métodos y técnicas utilizados en la selección de personal y los procedimientos parte del reclutamiento, junto con las herramientas de inteligencia artificial disponibles para mejorar este proceso. Se utilizó el sistema M-Learning como medio para proporcionar material educativo e interactuar con los postulantes.

Posteriormente, tras la implementación del modelo de preparación y entrevista, se procedió a validar y evaluar su efectividad. Los resultados obtenidos indicaron una mejora significativa en la preparación de los candidatos y un mayor apoyo en la del selección de personal y en procedimientos parte del reclutamiento, asimismo, en el índice de facilidad de operación tuvo una mejora del 83% y en el índice de eficiencia de ejecución tuvo una mejora del 76%. En conclusión, se determinó que la propuesta de modelo era efectiva y recomendable, especialmente al adaptarse al estilo de aprendizaje de cada usuario. En base a las conclusiones del autor, se consideró válido el uso de recomendaciones de M-Learning para mejorar tanto la contratación de personal como la satisfacción del postulante. Este enfoque proporciona una herramienta eficaz para preparar a los candidatos de manera más efectiva y mejorar su desempeño en las entrevistas de selección, lo que a su vez puede aumentar las probabilidades de éxito en la contratación de personal adecuado para las necesidades de la organización.

También, según Armas (2021), en su investigación, se determinó cómo las organizaciones peruanas han ido adoptando la Inteligencia Artificial, dado que, los cambios que involucra su utilización proporcionan un impacto significativo sobre otras organizaciones, y cómo se relaciona esto con la normativa laboral vigente. La investigación se centró en analizar si las regulaciones laborales en Perú promueven el avance tecnológico y una

administración empresarial adecuada, haciendo hincapié en el uso de la Inteligencia Artificial. Para llevar a cabo esta investigación, se recopilaron y analizaron 203 opiniones a través de una técnica de encuesta. Los hallazgos indicaron que la investigación identifica una conexión directa entre la eficiencia y el beneficio económico de las compañías y la preocupación por perder empleo a causa de la introducción de la Inteligencia Artificial, dando como resultado que en el índice de eficiencia de la inteligencia artificial tuvo una mejora del 31%. Del mismo modo, según Espinoza (2020), en su investigación, analizó la problemática del desconocimiento financiero personal en los fondos privados de pensiones, caracterizado por un acceso limitado, información oficial poco confiable e ininteligible para el ciudadano promedio.

Para abordar esto, un modelo analítico de ML fue propuesto. El diseño del programa tiene como objetivo anticipar el capital acumulado de un adherente contemplando su venidera jubilación. Este modelo se fundamentó en factores de crecimiento y se desarrollaron técnicas, metodologías y herramientas de valoración, junto con una aplicación móvil, todas basadas en Machine Learning. Durante el proceso, se validó y elaboró el modelo de predicción de estimación y proyección. Se ha demostrado mediante lo hallado que, instaurar este modelo mejoró significativamente el conocimiento sobre los aportes mensuales y generó un mejor entendimiento de los afiliados respecto a los parámetros utilizados por la SBS, dando como recomendación que en el índice de regresión lineal tuvo una mejora del 0.006535, en el índice de regresión de redes neuronales tuvo una mejora del 0.006752, en el índice de regresión de árbol de decisión tuvo una mejora del 0.000088, en el índice de regresión lineal bayesiana tuvo una mejora del 0.006535 y en el índice de regresión de bosques aleatorios tuvo una mejora del 0.000109. Como recomendación final, se sugirió que los individuos consulten con especialistas financieros para adquirir una comprensión más exhaustiva y holística de sus finanzas y futuras pensiones.

Antecedentes Internacionales

Según Dávila (2019), en su investigación, se enfocó en los procedimientos administrativos y lo relacionado a la gestión de la planilla de trabajadores, con especial atención en los procedimientos llevados a cabo por el área de RR.HH. Se estableció de manera precisa el proceso de contratación, abordando aspectos como el perfil requerido por los empleadores, la rotación de personal y la gestión de puestos de trabajo. Una herramienta clave utilizada en este estudio fue la data Mining, este último ha servido como aquel recurso de apoyo al procesamiento de candidatos. El objetivo primordial consistió en diseñar un método eficiente, tomando en cuenta los perfiles de puesto para diferenciar las reglas de decisión proporcionadas a los candidatos, con la finalidad de acelerar la decisión making dentro del ámbito de competencia, dando como resultado que el índice de análisis de reglas ideales tuvo una mejora del 80%. El enfoque basado en la Data Mining permitió identificar a los mejores candidatos a través del análisis de vacantes, ofertas de empleo y reglas de decisión establecidas. Además, se implementaron etapas de entrenamiento basadas en Data Mining en pro de la optimización del proceso de selección. También, según, Pérez (2021), en su investigación, indicó que el Data Mining emerge como un campo que, si bien ofrece oportunidades valiosas, enfrenta desafíos considerables que deben ser superados. Es crucial desarrollar criterios reproducibles que faciliten la aplicación efectiva de esta técnica y que contribuyan al avance investigativo del área. En la contextualización de la elección de las decisiones discrecionales en el procedimiento de reclutamiento de recursos humanos, se destaca la importancia de elegir y evaluar candidatos de manera efectiva. Este proceso implica la identificación de las características más apropiadas para cada posición, así como la valoración de aptitudes y competencias pertinentes. Para interactuar eficazmente con la tecnología, es fundamental replantear el liderazgo transformacional. Esto implica proporcionar un pensamiento crítico que fomente la adaptación a los avances tecnológicos y promueva una cultura organizacional que valore la innovación y

el aprendizaje continuo. Una propuesta valiosa sería desarrollar minería de datos textual que genere valor a través de evidencia reproducible. Esto permitiría una evaluación y selección más efectiva de candidatos, optimizando así el proceso de contratación y elección de empleados.

En este contexto, es esencial estudiar los conceptos básicos de selección de personal, organización y discursos psicológicos. Estos conocimientos ofrecen un fundamento robusto para comprender los aspectos clave del proceso de selección de personal y facilitan la implementación de estrategias eficaces en el ámbito de la administración de recursos humanos. A esto se le suma, el estudio realizado por Lu et al. (2024), en el cual se ha planteado el objetivo medular de mapear sistemáticamente, describir y resumir la utilización actual de inteligencia artificial en el proceso de reclutamiento y retención de profesionales clínicos. Se llevo a cabo una revisión exhaustiva del alcance, que abarca tanto literatura revisada por pares como literatura gris y explorar la utilización de herramientas basadas en IA. La revisión seguirá la metodología marco proporcionado por el Instituto Joanna Briggs (JBI) para asegurar la transparencia y la conformidad de la presentación de informes, estándares, la extensión.

Asimismo, según, Matute et al. (2021), en su artículo, se realizó una exhaustiva recolección de datos pertinentes, que incluyó desde las estrategias más eficientes en la administración de capital humano hasta los enfoques más innovadores en materia de inclusión laboral. Se logró identificar y comprender los factores de éxito específicos que pueden influir en la promoción de una cultura organizacional inclusiva. Además, se hizo hincapié en la capacidad de gestionar y realizar cambios dentro de la organización. Esto implicó no solo identificar los factores clave para la inclusión laboral, sino también desarrollar estrategias efectivas para implementar estos cambios y fomentar una cultura de diversidad y equidad en el lugar de trabajo. Un aspecto importante de la investigación fue el uso del sector público como base del marco referencial. Se exploraron las prácticas y políticas exitosas implementadas en

el sector público, que podrían adaptarse y aplicarse en entornos empresariales para mejorar la inclusión laboral. Además, se exploraron las posibilidades de aplicar técnicas de aprendizaje automático con el fin de optimizar el proceso de contratación de empleados, lo que podría contribuir significativamente a garantizar una contratación más inclusiva y equitativa, dando como resultado que el índice de precisión de función lineal tuvo una mejora del 0.805, el índice de precisión de función cuadrática tuvo una mejora del 0.815 y el índice de precisión de función cúbica tuvo una mejora del 0.818. También, según, Berón et al. (2021), en la investigación de su autoría, se basó en teorías fundamentadas en el empirismo y enfoques de técnicas de Data Mining a fin de capturar la base de datos de enfermedades y utilizarla para predecir el motivo de ausentismo en el trabajo. El propósito medular fue prever el comportamiento del motivo de ausentismo, lo que permitió identificar las posibles causas que podrían intervenir en este fenómeno. Para lograr este propósito, se desarrollaron especificaciones de metodología reproducible que proporcionaron las BD confiables para la predicción del ausentismo. Esto implicó la recopilación y análisis de datos relacionados con enfermedades, ausentismo laboral y otros factores relevantes que podrían influir en la asistencia de los empleados al trabajo. Se hace uso de técnicas complejas de Data Mining para explorar y analizar los datos recopilados, identificando patrones, tendencias y relaciones que podrían ayudar a predecir el motivo de ausentismo. Estas técnicas permitieron la construcción de modelos predictivos precisos y confiables que podrían utilizarse para anticipar el ausentismo laboral y tomar medidas preventivas para abordar las causas subyacentes, dando como resultado que el índice de efectividad tuvo una mejora del 94,72%. Por otro lado, Fraij y László (2021), en su estudio, cuyo objetivo medular fue revisar la implementación de IA en procedimientos relacionados con la contratación de la Gestión de RR.HH. Se adoptó una revisión sistemática en la que se revisaron artículos académicos, artículos de revistas y sitios web altamente calificados en campos relacionados. Sin embargo, la metodología de investigación utilizada parece razonable

y aceptable ya que cubre un buen número de artículos relacionados con el área temática principal. Los resultados y las conclusiones fueron casi claros: el uso de la IA tiene ventajas en el ámbito de la contratación, ya que la tecnología puede ser la mejor opción en este ámbito. De igual forma, según Alkhazraji y Buhaliba (2020), en su trabajo de investigación, mencionó que el aprendizaje para el reclutamiento en las organizaciones ha demostrado ser una herramienta invaluable que no solo ha ahorrado tiempo a los reclutadores, sino que también ha mejorado significativamente la calidad de las contrataciones. Este avance no solo es relevante para el presente, sino que también ofrece una base robusta para investigaciones posteriores en el área de administración de los empleados. El uso de aprendizaje automático ha transformado las funcionalidades al gestionar al área de RR.HH., permitiendo una selección más precisa y eficiente de candidatos. Para aprovechar al máximo esta tecnología, es crucial replantear los procedimientos relacionados al reclutamiento con el fin de agilizar la selección de candidatos y adaptarse a los avances tecnológicos. Una propuesta valiosa sería determinar la importancia del uso de M-Learning, que genera valor mediante una evidencia de adopción sólida. Esto permitiría comprender mejor cómo esta tecnología puede mejorar eficazmente el procedimiento de contratación y elección de empleados en las organizaciones. Además, sería beneficioso desarrollar minería de datos textual para generar valor mediante una evidencia reproducible. Esta técnica podría facilitar la evaluación y selección de candidatos de manera más efectiva, optimizando así el proceso de reclutamiento, dando como resultado que el índice de precisión tuvo una mejora del 90%. Para llevar a cabo estos avances, es esencial estudiar los conceptos básicos de selección de personal, organización y discursos psicológicos. Esta comprensión proporciona una base sólida para abordar los desafíos y oportunidades asociados con la implementación de tecnologías avanzadas en el procedimiento de contratación y elección de empleados.

1.5. Justificación de la investigación

Conveniencia

El estado de progreso de esta investigación actual sirve para mejorar la contratación del personal en el sector privado, por ende, se logra disminuir el tiempo de contratación, se reduce el costo por reparación social, aumenta la satisfacción del postulante e incrementa el grado de contento de los individuos involucrados en el proceso para la prestación servicios para el desarrollo de la eficiencia de la cultura previsional para los postulantes del sector privado.

Relevancia social

Mediante la aplicación de la metodología CSKT se podrá obtener una ventaja mediante el perfeccionamiento del procedimiento de Contratación de personal en el sector privado utilizando Machine Learning para reducir el desconocimiento que tienen muchos postulantes al momento de su elección de un régimen pensionario. Esto a su vez mejora la calidad de estrategias ejecutadas de la oficina de recursos humanos, con la utilización de un software en óptimas condiciones y correctamente calibrado.

Implicaciones Prácticas

Por medio del estudio actual se pretende manifestar una solución frente a los problemas presentados en los procedimientos relacionados a la Contratación de los empleados en el sector privado, consintiendo que la obtención de información para la decisión de un régimen pensionario sea más rápida, y por ende los postulantes eleven su satisfacción en asumir los riesgos de su elección; asimismo, permite al postulante obtener una cultura previsional tempranamente de las ventajas y desventajas que estos ofrecen.

Valor teórico

La metodología CSKT que se obtenga de la comparación de las metodologías ya existentes, a posteriori, ya implementada en el proceso de Contratación de personal, se podrá volver a aplicar a los demás procesos que impliquen un material de selección brindados por el sector privado.

Aportes

Con la presente investigación se logra mejorar el proceso de Contratación de personal del sector privado, logrando disponer de una herramienta móvil que facilite decidir por una elección de un seguro pensionario confiable.

1.6. Limitaciones de la investigación

- a) Pedir información de las planillas de la SBS a un costo de uso para acceder a dichas planillas no solo se limita a los investigadores.
- b) El investigador solo tiene acceso a las planillas de la SBS por edad y sexo.
- c) Para el desarrollo del trabajo de investigación el ingreso al puesto de trabajo está restringido; asimismo se desconoce la forma de análisis de las plantillas de personal que le ofrecen al postulante.

1.7. Objetivos

Objetivo general

Implementar y evaluar una aplicación móvil con Machine Learning empleando la nueva metodología CSKT para optimizar el proceso de contratación de personal en el sector privado.

Objetivos específicos

1. Verificar la eficacia de la metodología CSKT aplicada en el modelo de Machine Learning mejora la eficiencia en términos de tiempo y costos en el proceso de contratación de personal.

2. Desarrollar y analizar el prototipo de Machine Learning que evalúe los criterios de aceptación en el proceso de contratación de personal.
3. Medir la precisión del modelo de Machine Learning con la metodología CSKT y su impacto en la optimización del proceso de contratación de personal en el sector privado.
4. Evaluar la eficiencia operativa del modelo de Machine Learning integrado en la aplicación móvil.
5. Implementar la aplicación móvil con Machine Learning y la metodología CSKT y estudiar su efectividad en el proceso de contratación de personal.

1.8. Hipótesis

Hipótesis General

La implementación de una aplicación móvil con Machine Learning y la metodología CSKT optimiza significativa el proceso de contratación de personal en el sector privado.

Hipótesis Específicas

1. La metodología CSKT aplicada en el modelo de Machine Learning mejora la eficiencia en términos de tiempo y los costos en el proceso de contratación de personal.
2. El prototipo de Machine Learning evalúa de manera precisa los criterios de aceptación basándose en la categorización del proceso de contratación de personal.
3. La tasa de precisión del modelo de Machine Learning con la metodología CSKT optimiza significativamente el proceso de contratación en el sector privado.
4. La eficiencia del proceso de contratación de personal mejora con la utilización del modelo de Machine Learning en la aplicación móvil.

5. La implementación práctica de la aplicación móvil con Machine Learning y la metodología CSKT es efectiva en la optimización del proceso de contratación de personal en el sector privado.

II. MARCO TEÓRICO

2.1. Marco conceptual

2.1.1. *Contratación de personal*

A continuación, para la descripción de la variable dependiente que es **Contratación de personal**, se consideró los siguientes conceptos:

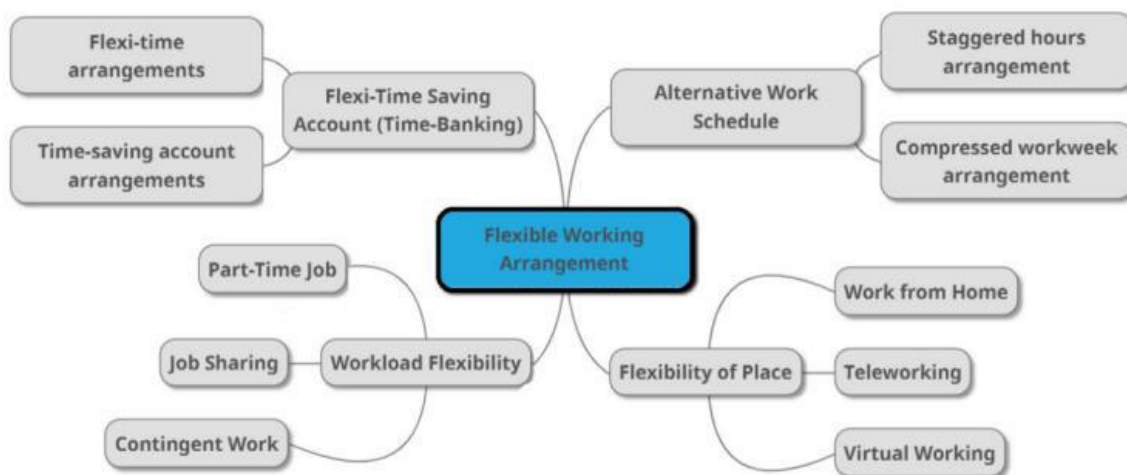
Según Borisova et al. (2020), la principal definición con este concepto es el estrecho vínculo estricto con el estado “Buscando empleo”. Este proceso implica la búsqueda exhaustiva en todos los lugares donde una persona deja huellas de información en línea. La búsqueda avanzada tiene el potencial de ampliar significativamente el flujo de entrada de candidatos, lo cual es una gran ventaja si se busca implementar estrategias de reclutamiento o contratación de profesionales. La lista de ubicaciones digitales candidatas potencialmente adecuadas sirve como base para establecer parámetros destinados a la entrega de información a los grupos objetivo, así como para seleccionar el método de contacto con los solicitantes. El conjunto de herramientas disponibles incluye una amplia variedad de opciones, como anuncios dirigidos, contenido personalizado, banners emergentes y publicidad en redes sociales. La elección de una herramienta en particular su dependencia es por múltiples factores, tales como: la ubicación de los potenciales candidatos y sus respectivas preferencias en cuanto a la recepción de comunicaciones y oportunidades laborales. Es esencial considerar estas variables al decidir qué herramientas utilizar para maximizar la efectividad del proceso de reclutamiento y contratación.

Según Shaari y Amiru (2020), los procedimientos relacionados a la contratación de personal están experimentando una evolución significativa, pasando de los métodos convencionales a una etapa más flexible que beneficia tanto a los empleadores como a los empleados. Esta evolución conlleva una mayor eficiencia en la retención del talento y un mayor

compromiso por parte de los empleados. Los empleados ahora tienen la ventaja de poder equilibrar mejor sus responsabilidades laborales y familiares, así como tener un mayor control sobre su trabajo y la oportunidad de buscar su crecimiento personal. Esta flexibilidad es especialmente común en los países desarrollados, donde se ha vuelto una práctica estándar en la esfera empresarial con la meta de fomentar o impulsar la reactivación económica. Para muchas grandes empresas, esta flexibilidad se ha convertido en una parte integral de sus planes estratégicos para el desarrollo socioeconómico del país. Los términos de duración y horarios de trabajo ofrecidos durante el proceso de contratación reflejan esta tendencia hacia una mayor flexibilidad y adaptabilidad en el entorno laboral actual. La siguiente figura muestra las categorías de trabajo en el sector privado:

Figura 5

Categorías de trabajo en el sector privado



Factores que afectan la contratación de persona

Según Anayochukwu y Anayochukwu (2021), el proceso de implementar un proceso válido en el reclutamiento y selección de servidores públicos debe cumplir rigurosamente con los principios de desempeño del servicio público, utilizando el mérito como criterio fundamental para establecer reglas equitativas. Esto implica la especificación detallada de

todos los atributos relevantes de los solicitantes de empleo, que van desde aspectos sociológicos y físicos hasta educativos y psicológicos, garantizando así un proceso completo y justo. Por consiguiente, la efectividad del proceso de selección tiene un impacto crucial en la capacidad durante un tiempo prologado de la corporación para conseguir sus objetivos. La introducción de empleados debe basarse en un debido proceso de contratación y selección que tenga en cuenta diversos factores, como los sociológicos e institucionales. Los factores sociológicos, que incluyen consideraciones políticas y de patrocinio, a menudo se puede percibir su influencia dentro del proceso de selección, lo que puede llevar a decisiones que no se basan únicamente en las calificaciones y el mérito de los candidatos. Por otro lado, los factores institucionales, como las pruebas de gestión y desempeño, las habilidades, la experiencia y las calificaciones, deben ser los principales criterios para evaluar a los candidatos.

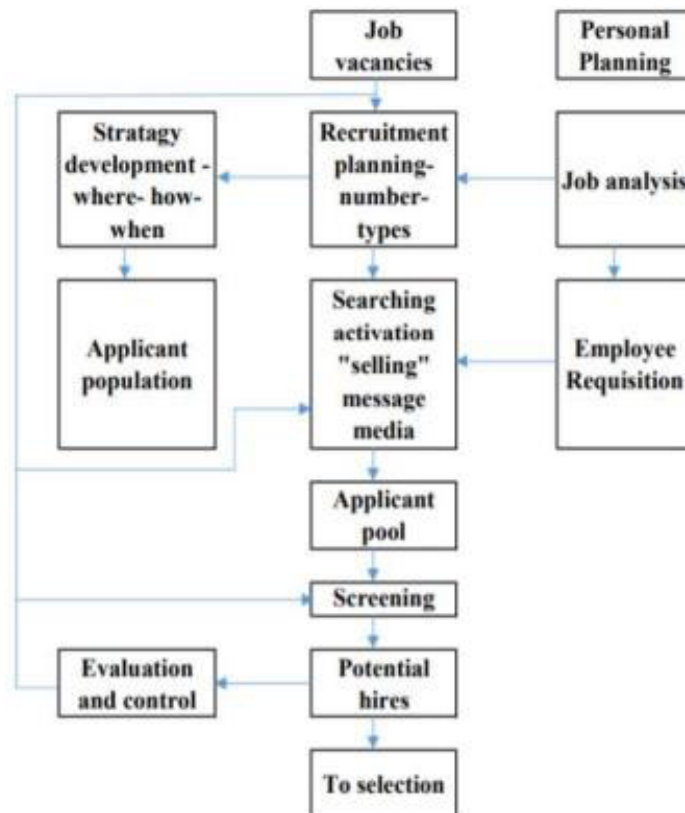
Además, según el autor Lu et al. (2024), los problemas de reclutamiento y retención afectan a diferentes aspectos de los servicios de salud y atención; se considera a la rotación y la escasez de personal variables que impactan en la excelencia del cuidado y el paciente.

Adicionalmente, Alkan et al. (2024) hace mención de otro factor muy involucrado es la rotación de personal que indica problemas de retención en el sector sanitario personal. De lo anteriormente mencionado, existen lagunas en la comprensión de la raíz a causa de las deficiencias de la fuerza laboral de una organización individual y cómo los factores del entorno de práctica interactúan para impactar en la captación y conservación del personal. También hace mención del aumento de la carga de trabajo y el agotamiento son factores, los cuales están relacionados con la COVID19 u otras enfermedades exacerbaban la salud del personal.

Es esencial evitar la influencia política y local en la selección de personal para garantizar la idoneidad y la eficacia del proceso de contratación en el ámbito público. La siguiente figura muestra los factores que inciden en la contratación:

Figura 6*Factores que afectan la contratación***Proceso de contratación de personal**

Según Dhabuwala y Pitroda (2021), la selección del personal es un paso primordial en la administración de RR.HH. en cualquier entidad. Su meta primordial consiste en identificar y elegir individuos calificados para desempeñar roles específicos dentro de la compañía. Este se inicia en la exploración activa de nuevos empleados y concluye con la incorporación de los candidatos seleccionados en los puestos vacantes. El ciclo de contratación comprende diversas etapas, desde la identificación de las necesidades de personal hasta la presentación de las solicitudes de empleo por parte de los candidatos. Durante este proceso, se evalúan las habilidades, competencias y experiencias de los postulantes, con el fin de encontrar el mejor ajuste para los requerimientos del puesto.

Figura 7*Proceso de contratación*

Con relación al flujo anterior se puede mencionar que:

a. Vacante de empleo: Esto se debe a que el reclutador está tomando medidas para encontrar candidatos adecuados para problemas recién creados o definidos por la organización, definido como un puesto renumerado. Para ello, el candidato debe tener la intención de completar los requisitos para un puesto de trabajo y de esa manera puede ser localizado por el reclutador de manera inmediata o en un lapso temporal definido.

b. Planificación de la contratación: Los planes de contratación se refieren a las estrategias que utiliza la empresa para contratar empleados. Esto sirve como base de planificación para que las empresas encuentren candidatos elegibles sin incurrir en tiempo de inactividad.

El plan de contratación identifica los siguientes pasos para un puesto en particular:

1. Identificar la oferta de trabajo: Una oferta de trabajo es cuando el candidato es invitado formalmente a trabajar para la empresa.
2. Decidir cómo llenar la vacante: Esto incluye la búsqueda de empleo y los anuncios de la bolsa de trabajo, así como también el propio sitio web del candidato para poder llenar la vacante.
3. Identificación del grupo objetivo: Consiste en seleccionar a quienes están más interesados por la empresa, posición de trabajo o puesto remunerado.
4. Notificar al grupo objetivo: Significa que el grupo objetivo es llamado para recibir una información específica sobre su proceso de reclutamiento.
5. Reunión con los candidatos: Es una parte importante del proceso de reclutamiento, donde los candidatos que completen más requisitos por la empresa al participar en una reunión con encargados de RR.HH. y los gerentes de reclutamiento, quienes necesitan saber cómo llevar a cabo la contratación de personal con los candidatos que quedan.

Indicadores de la contratación de personal

En el estudio se han tomado en cuenta cuatro (4) medidas, que son: Duración del proceso de contratación, Gastos asociados al proceso de contratación, Nivel de satisfacción del candidato y Efectividad en el llenado de vacantes. A continuación, se explica los conceptos y fórmulas correspondientes:

Según Jaramillo (2020), el Tiempo de contratación permite profundizar en los tiempos como el comienzo de la actividad y la fecha límite de presentación de postulantes, teniendo en cuenta el tiempo de asistencia requerida por la prioridad y volumen de los procesos considerando el (Acuerdo de Nivel de Servicio) establecido, que tiene como fórmula:

Tiempo de contratación = (días hasta que elige el sistema pensionario / vacantes cerradas)

De acuerdo con el concepto antes mencionado, se especifica entonces que es la división de días hasta que elige el sistema pensionario entre el número de vacantes cerradas con un resultado en días.

Según Guamán et. al. (2020), el Costo de contratación es desarrollar efectivamente un proceso de selección pensando en los costos que se requieren para llevarlo a cabo. Entre estos, se pueden mencionar algunas cantidades: vacantes, horas de trabajo, comidas y, en algunos casos, alojamiento, que tiene como fórmula:

$$\text{Costo de contratación} = (\text{costos de contratación} / \text{ingresos})$$

De acuerdo con el concepto antes mencionado, se logra especificar entonces que es la división de costos de contratación entre el número de ingresos con un resultado en moneda (soles).

Según Sánchez (2020), la Satisfacción del postulante es lograr que la dedicación y la fidelidad de los empleados no solo se limitan al bienestar laboral personal, sino estén ligados a su desempeño y niveles de eficiencia, potenciando la buena reputación de una empresa en particular, que tiene como fórmula:

$$\text{Satisfacción del postulante} = (\text{candidatos satisfechos con el nuevo sistema pensionario} / \text{contrataciones})$$

De acuerdo al concepto antes mencionado, se especifica entonces que es la división de candidatos satisfechos con el nuevo sistema pensionario entre el número de contrataciones con un resultado en porcentaje (%).

De acuerdo con el estudio realizado por Hanze en 2020, la finalidad de la Eficiencia de Vacantes es acelerar a la exploración de candidatos por parte de las empresas que utilizan este sistema, facilitando de esta manera el proceso de búsqueda y contratación de nuevos talentos para cubrir las vacantes disponibles, que tiene como fórmula:

Eficiencia de vacantes = (vacantes cerradas / total de vacantes)

De acuerdo con el concepto antes mencionado, se logra especificar entonces que es la división del número de vacantes cerradas entre el total de vacantes con un resultado en porcentaje (%).

2.1.2. Aplicación móvil

En el presente punto se explica lo que es una aplicación móvil en base a los siguientes conceptos:

Según Mendoza et al. (2020), una app consiste en ser un software que está especialmente diseñado para que pueda ser ejecutado en dispositivos móviles, ofreciendo mayor practicidad y eficacia en comparación con los sistemas web o de escritorio. Estas aplicaciones pueden ser utilizadas de manera efectiva en dispositivos móviles sin depender de una conexión constante a una fuente de energía cercana, lo que las hace prácticas y eficientes para los usuarios.

Al estar conectada a una base de datos mediante un servicio de plataforma como Firebase, las aplicaciones móviles pueden ofrecer soluciones que aborden diversas necesidades de los usuarios, como el proceso de contratación de personal. Estas soluciones pueden incluir un software que facilite el control y la verificación de las vacantes disponibles en una empresa, así como la evaluación de qué candidatos se ajustan mejor al perfil requerido. Esto ayuda a evitar la contratación de personas que no posean los conocimientos necesarios para

desempeñarse eficazmente en el área específica dentro de la organización (Ortiz y Valadez, 2019).

2.1.3. Machine Learning

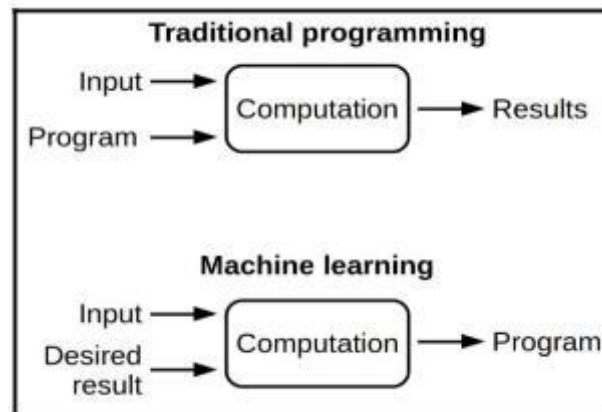
A continuación, se ha tenido que considerar ciertos conceptos para poder realizar la descripción de la variable independiente: **Machine Learning**. Como se detalla a continuación:

Según Campesonato (2020), el aprendizaje automático, como subconjunto de la inteligencia artificial, aborda tareas complejas que son difíciles de resolver con métodos de programación tradicionales. Los filtros de spam de correo electrónico son un ejemplo temprano de aplicación de aprendizaje automático. Este enfoque a menudo mejora la precisión de los algoritmos anteriores. Aunque existen varios algoritmos de ML, la calidad de los datos es fundamental y puede presentar desafíos como la insuficiencia, baja calidad, errores o datos irrelevantes. En este capítulo, se explora formas de abordar estos problemas. En términos simples, hace referencia a una matriz de datos estructurada que representan características específicas para cada punto de datos. Cada fila puede representar, por ejemplo, un cliente, y cada columna sería una característica específica de ese cliente, como edad, género, ingresos, etc. Los conjuntos de datos suelen estar en formatos como archivos CSV o hojas de cálculo, facilitando su manipulación y análisis. (p. 24). En este mismo, coincide Raß-Kettler y Lehnervp (2019), quienes consideran a Machine Learning, en pocas palabras, consiste en el proceso donde la máquina tiene la capacidad de aprendizaje por sí sola sin la dependencia de estar programado para realizar de una determinada forma aprendiendo de sus errores. Por ejemplo, ellos también hacen mención a AlphaGo, un programa de aprendizaje automático de Google diseñado para jugar GO, antes de convertirte en un maestro y vencer a los jugadores humanos, perdió varios juegos, lo que le ayudó a aprender y desarrollarse nuevas estrategias ganadoras (p. 3).

A posteriori, se observa la derivación automática de acciones a partir de datos sin intervención manual, como diferenciador importante entre ML y la programación tradicional:

Figura 8

Programación tradicional vs Machine Learning



Nota. Este gráfico fue extraído de Chinnamgari (2019)

En ese sentido, según Chinnamgari (2019), otro diferenciador importante entre ML y la programación tradicional es que el conocimiento adquirido a través de ML va más allá de la generalización de ejemplos de entrenamiento. Los programas codificados en la programación tradicional solo pueden ejecutar las respuestas contenidas como parte del código, pero el algoritmo interpreta datos que nunca se habían visto.

Por su lado, desde otro punto de vista sobre la figura, según Lee (2019), si ha escrito un programa, debe estar familiarizado con el diagrama que se muestra. Para la programación tradicional, puede crear programas que realicen contabilidad, como listas de inventario. El programa calcula las pérdidas en base a datos y ganancias o registros de ventas. Tal vez también pueda crear algunos gráficos agradables y extravagantes para mostrar su rendimiento de ventas. En este caso, el resultado es un estado de resultados y otros gráficos.

En otros casos, el aprendizaje automático ha transformado el paradigma de programación tradicional de la Figura 5 en el nuevo paradigma que se muestra en la sección Aprendizaje automático.

En lugar de alimentar los datos al programa, los datos recopilados y la salida ahora se usan para derivar el programa (también conocido como el modelo). Usando el mismo ejemplo de contabilidad en el paradigma de aprendizaje automático, se toma en cuenta registros de ventas diarios (tanto datos como resultados) y se utiliza para derivar un conjunto de reglas y hacer predicciones. Puede usar este modelo para predecir los artículos más populares o menos populares para vender el próximo año.

Según Hart et al. (2021), además de simplemente predecir la presencia de un compuesto, el aprendizaje automático se puede entrenar para predecir directamente otras propiedades microscópicas observables, como la dureza, la ductilidad, la tenacidad y la temperatura de Curie. En 2009, demostraron que PCA se puede utilizar para reproducir mapas estructurales que representan la relación entre estructura y propiedades (propiedades electrónicas y parámetros de estructura cristalina) con Data Mining. En 2018, utilizaron un algoritmo de bosque aleatorio (RF) para estudiar las características, la capacidad calorífica, la entropía de vibración y la energía libre 59 del espectro de fonones para mejorar la predicción de la estabilidad termodinámica de temperatura finita. El modelo puede incluir condiciones de aleación como parte del vector de características para predecir el procesamiento.

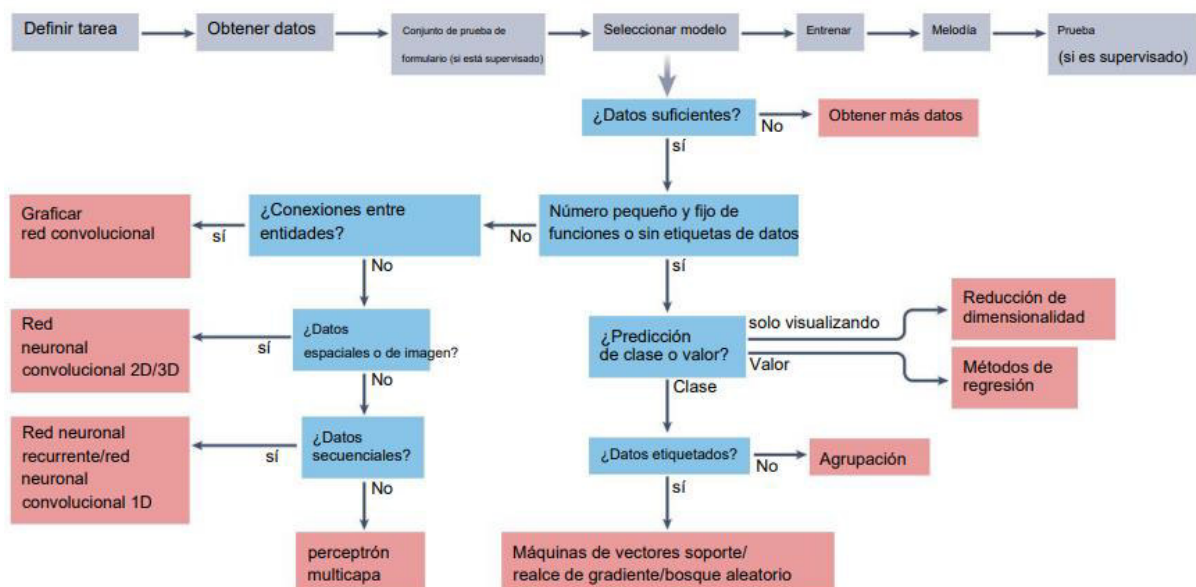
Según Janiesch et al. (2021), en general, ML implica que el desempeño de un programa de computadora se perfecciona mediante la adquisición de experiencia en la medición de una clase particular de trabajo y rendimiento. Entonces, la meta es automatizar la creación de modelos analíticos para llevar a cabo actividades de cognición, a manera de ejemplo, la identificación de objetos o el hecho de traducir del lenguaje humano. Esto se logra mediante la

aplicación de algoritmos que aprenden iterativamente datos de entrenamiento específicos de la tarea, lo que permite que las computadoras encuentren información oculta y patrones complejos sin programarlos explícitamente. Especialmente en tareas que involucran datos multidimensionales como clasificación, regresión y agrupación, Machine Learning evidencia una buena adaptabilidad.

Según Greener et al. (2022), “Machine Learning” se refiere en términos generales al proceso de adaptar modelos predictivos a los datos e identificar grupos de información dentro de los datos. En el campo del aprendizaje automático, básicamente se utilizan cálculos para intentar estimar o imitar la capacidad humana para reconocer patrones, aunque sea de manera objetiva. El aprendizaje automático es especialmente útil cuando el conjunto de datos que se analiza es demasiado grande (muchos puntos de datos individuales) o complejo (incluidas muchas características) para el análisis humano y/o cuando se automatiza el proceso de análisis de datos. Cree una canalización reproducible y eficiente en el tiempo. El procedimiento para entrenar técnicas de Machine Learning se muestra a continuación:

Figura 9

Elegir y entrenar un método de aprendizaje automático

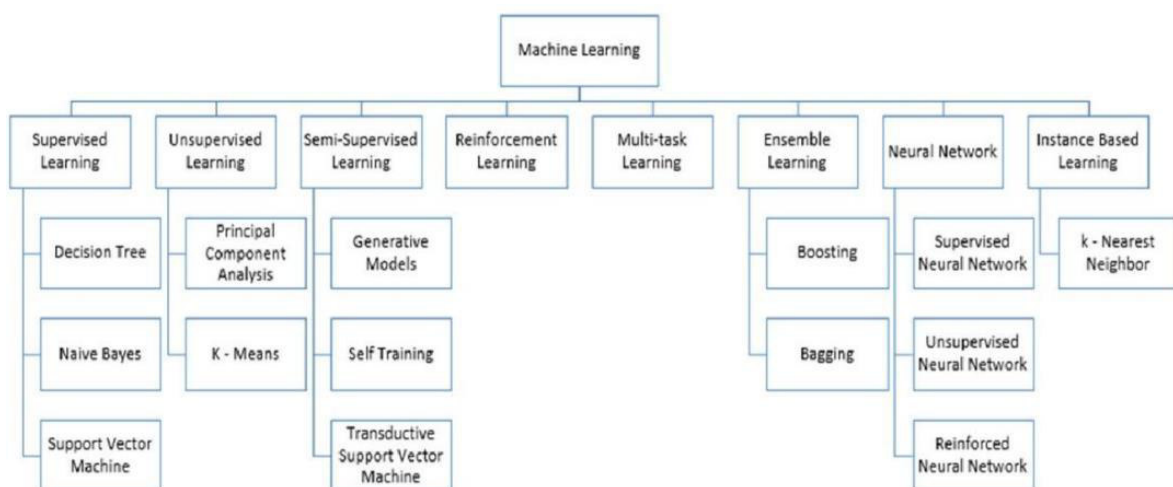


Algoritmos empleados en Machine Learning

Según Mahesh (2020), ML conviene cuando se busca instruir a las máquinas sobre cómo procesar datos de manera más eficaz. Cuando los datos son complejos y difíciles de interpretar directamente, se recurre al aprendizaje automático. Dado el vasto conjunto de datos disponibles, la demanda de ML está en aumento, siendo utilizado por diversas industrias para extraer información relevante. El propósito medular del ML consiste en la asimilación de conocimientos a partir de los datos, se han efectuado muchos estudios con el propósito dar la posibilidad que las máquinas desarrollen competencias autónomamente, prescindiendo de instrucciones de programación explícitas. Matemáticos y programadores emplean variadas estrategias para enfrentar este desafío en una data extensa. La selección del algoritmo adecuado depende del problema en cuestión, el número de variables, el modelo óptimo, entre otros factores. Machine Learning sustentada dada su diversidad de algoritmos para finiquitar cuestiones relacionadas con la información, y aquellos destacados, según los científicos, se visualizan en la siguiente representación:

Figura 10

Algoritmos utilizados en Machine Learning



Indicadores de ML

Para la evaluación del modelo de ML empleado en el proyecto, se tiene en cuenta las siguientes métricas:

Tasa de precisión del modelo de ML: Hace referencia al índice o porcentaje de todos los valores verdaderos y positivos en base a todas las predicciones que el modelo de aprendizaje automático realiza durante una operación (Mon, 2022).

Eficiencia del modelo de ML: Se entiende por el alcance que tiene el modelo en obtener excelentes niveles de rendimiento por medio de los recursos que tenga a su disposición y los tiempos reducidos de su entrenamiento logrando obtener niveles de eficiencia deseados (Mirjalili y Raschka, 2020).

Evaluación del prototipo de ML: Enfocado en el desarrollo completo del modelo de la aplicación móvil con tecnología de aprendizaje automático considerando que cumpla cada requerimiento establecido durante las pruebas de su funcionamiento (Ortiz y Andres, 2021).

Los siguientes indicadores son a considerar en base al objetivo de validación del funcionamiento del modelo de aprendizaje automático:

Accuracy: El presente indicador hace alusión al índice total de las observaciones realizadas en base a las respuestas acertadas que el modelo de aprendizaje automático, es decir, las mediciones y actividades que dicho modelo realiza se hacen tomando en cuenta la cantidad de las respuestas verdaderas positivas o negativas que este modelo obtiene al momento de realizar las pruebas (Bozzi, 2023).

$$\text{Accuracy} = \frac{\text{Verdaderos positivos} + \text{Verdaderos negativos}}{\text{Verdaderos positivos} + \text{Verdaderos negativos} + \text{Falsos positivos} + \text{Falsos negativos}}$$

Precisión: En el caso del presente indicador se menciona el porcentaje de los valores verdaderos positivos los cuales fueron predichos de forma correcta (Pérez, 2021).

$$\text{Precisión} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos positivos}}$$

Recall: Dentro del presente indicador para el modelo de machine learning se consideran todas las predicciones positivas que fueron dadas de manera correcta (Pérez, 2021).

$$\text{Recall} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos negativos}}$$

F1 – Score: Para el presente indicador se toma en cuenta los indicadores previos los cuales son la Precisión y el Recall, haciendo que se pueda predecir mejor los resultados positivos a diferencia de la métrica de Accuracy (Borja et al., 2020).

$$\text{F1 – Score} = (\text{Precisión} * \text{Recall}) / (\text{Precisión} + \text{Recall})$$

Metodologías de desarrollo de Machine learning

En la investigación se han analizado cuatro (4) tipos de enfoques metodológicos, que son: CRISP-DM, SEMMA, KDD y TDSP. Posteriormente, se explica la terminología de cada uno de ellos:

METODOLOGÍA CRISP-DM

De acuerdo con Schröer et al. (2021), el modelo CRISP-DM es una metodología estándar y ampliamente adoptada en la industria para llevar a cabo proyectos de Data Mining de manera eficiente. Este modelo consta de seis fases principales: (a) Comprensión del negocio, donde se busca entender los objetivos y requisitos empresariales para el proyecto de Data Mining. (b) Data understanding, en la que se recopilan datos pertinentes y se da la realización de una investigación preliminar para la comprensión de su estructura y calidad. (c) data preparation, donde se realiza: *data cleaning*, *integration*, y *transformation*; con el propósito de prepararlos para el modelado. (d) Modeling, donde se eligen y emplean técnicas de modelado apropiadas en el análisis de la data y el logro del alcance de los objetivos planteados del proyecto. (e) Evaluation, donde se evalúan los modelos creados para determinar su precisión y efectividad resolutive frente a los problemas identificados. (f) Deployment: Los modelos y los resultados se implementan en el entorno operativo del negocio para su uso continuo y monitoreo. La sgte. figura evidencias etapas pertenecientes a la metódica CRISP-DM y seguido se evidencia las descripciones del modelo de proceso CRISP-DM:

Figura 11

Metodología CRISP-DM

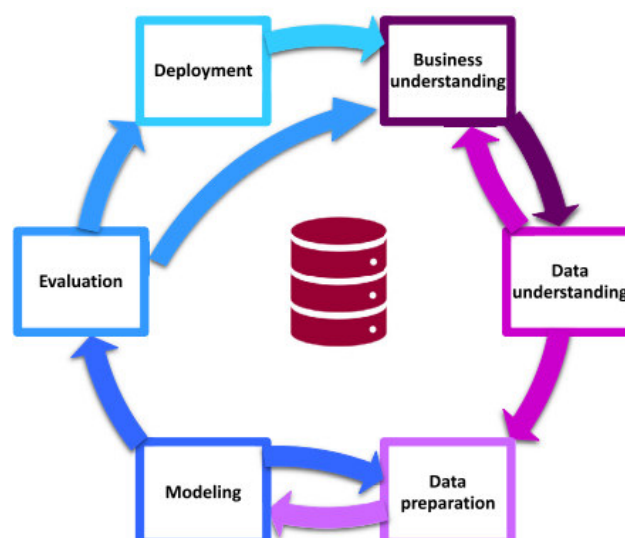


Tabla 3*Descripciones del modelo de proceso CRISP-DM*

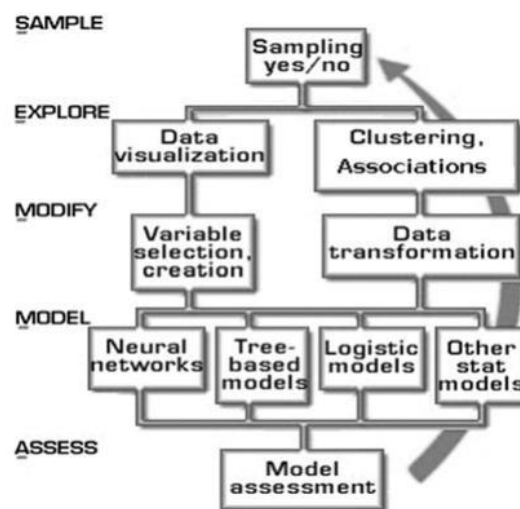
Fase	Breve descripción
Comprensión de la empresa	Conseguir una perspectiva completa de la disponibilidad de los medios imprescindibles y requeridos, es crucial su evaluación y/o análisis de la situación de la organización empresarial. Ello supone el establecimiento con la finalidad de la recolección de la data, el cual es considerado uno de los aspectos primordiales de la etapa. Durante este procedimiento, es imprescindible explicar las diversas modalidades de extracción de datos (como categorización, entre otras) y los estándares de éxito vinculados a esta tarea (como exactitud, entre otros), seguido por la creación de un plan de proyecto minucioso y obligatorio.
Data understanding	Recolectar información de diversas fuentes, examinarla exhaustivamente y valorar su calidad es una labor esencial en esta fase. El manual de instrucciones detalla la tarea de explicar la información mediante análisis estadísticos, lo que lleva a la identificación de sus características y conexiones.
Data preparation	Resulta esencial efectuar una selección de datos estableciendo filtros de admisión y rechazo durante este periodo. La disminución de la excelencia de la data puede abordarse por medio del procedimiento de la depuración de la data, de acuerdo con el modelo que ha sido utilizado (previamente identificado en la fase inicial). En consecuencia, es crucial definir una característica derivado de todos estos procedimientos para poder hacer frente a posibles variaciones según el modelo.
Modeling	La etapa de implementación usualmente se detalla en la guía de usuario, la cual puede ser tanto el informe final o una parte del módulo perteneciente al software. En dicho manual se describe que la fase correspondiente a la implementación engloba la ejecución, control y planificación del mantenimiento (Solano et al., 2022).
Evaluation	Para llevar a cabo una valoración del modelo, se aconseja inspeccionarlo mediante los parámetros de evaluación disponibles para identificar cuál es el más pertinente. Por lo tanto, resulta relevante establecer un enfoque distinto para analizar los resultados, además de revisar el proceso desde el inicio para asegurar la consecución de los resultados óptimos.
Deployment	La etapa de implementación comúnmente se especifica en la manual de usuario, el cual puede ser tanto el informe definitivo como el módulo de software. La guía aclara que la etapa de despliegue incluye la ejecución, control y programación de la gestión (Solano et al., 2022).

METODOLOGÍA SEMMA

Según Butka et al. (2020), SEMMA es una lista de pasos que guían a los usuarios en la ejecución de un proyecto de Data Mining. Aunque SEMMA proporciona todavía una visión bastante general del PDC, los autores afirman que es una organización más lógica de sus herramientas para cubrir las tareas principales de Data Mining (conocidas como SAS Enterprise Miner). Cabe resaltar que SEMMA cuenta con las siguientes fases: (a) Muestreo, (b) Explorar, (c) Modificar, (d) Modelar y (e) Evaluar. Se puede decir entonces que proporciona la descripción del proceso para seguir la herramienta de Data Mining. La siguiente figura muestra el proceso de la metodología SEMMA:

Figura 12

Metodología SEMMA



METODOLOGÍA KDD

Según Murnawan y Nugraha (2021), KDD se utiliza para referirse al proceso de descubrir e identificar patrones en un conjunto de datos. Cabe resaltar que KDD cuenta con las siguientes fases: (a) Selección, (b) Reprocesamiento o limpieza, (c) Transformación y (d) Data Mining. De acuerdo con el concepto antes mencionado, KDD es una metodología que permite

la interpretación o predicción de eventos futuros. La siguiente figura muestra el proceso de la metodología KDD y seguido se muestra la descripción del modelo de proceso de KDD:

Figura 13

Metodología KDD

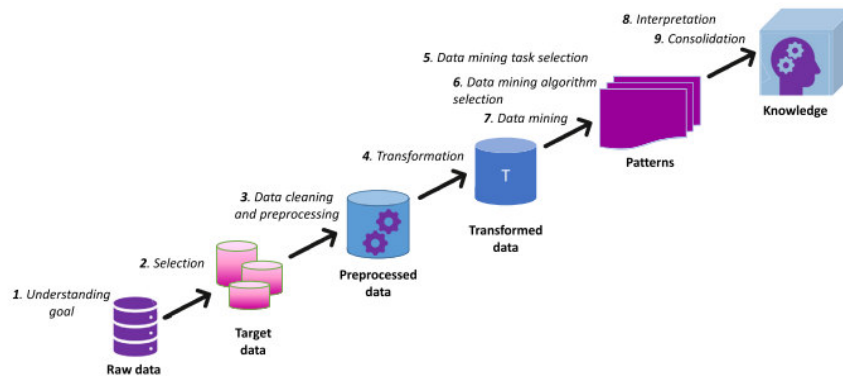


Tabla 4

Descripciones del modelo de proceso KDD

Fase	Breve descripción
Limpieza e integración	El objetivo de data cleaning es que la data procesados por Data Mining sean limpios y de alta calidad.
Selección de datos	A partir del entrenamiento de datos, después de que se hayan realizado las fases de limpieza e integración, la selección de datos ya ingresados se realiza en la fuente de datos ya limpios.
Transformación de datos	Consiste en aquel proceso de conversión de la data en el formato requerido para el siguiente proceso de extracción de datos. Por lo tanto, en esta etapa, se realizan algunos cambios en los atributos de la forma que aún no son adecuados para la excavación.
Data Mining	Es la fase en la que se utiliza el proceso de búsqueda de patrones o información de interés.

METODOLOGÍA TDSP

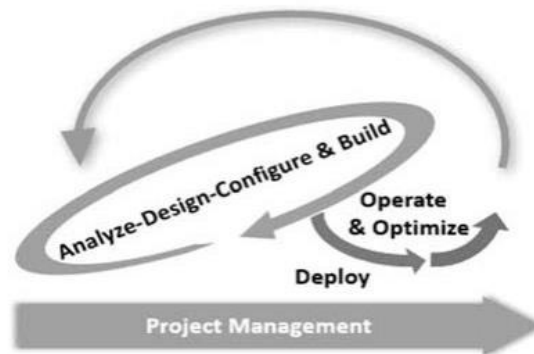
Según Costa y Aparicio (2020), TDSP es una metodología de Data Science rápida e iterativa. El TDSP también sugiere cómo los roles de equipo funcionan mejor juntos. Cabe resaltar que TDSP cuenta con las siguientes fases: (a) Analyze, Design, Configure, and Build, (b) Deploy, and (c) Operate and Optimize. De acuerdo al concepto antes mencionado, TDSP

es una metodología que tiene por objetivo el mejorar la cooperación y el aprendizaje de un equipo de analista de datos al seguir los procesos, de principio a fin, cuando se llevan a cabo.

La siguiente figura muestra el proceso de la metodología TDSP:

Figura 14

Metodología TDSP

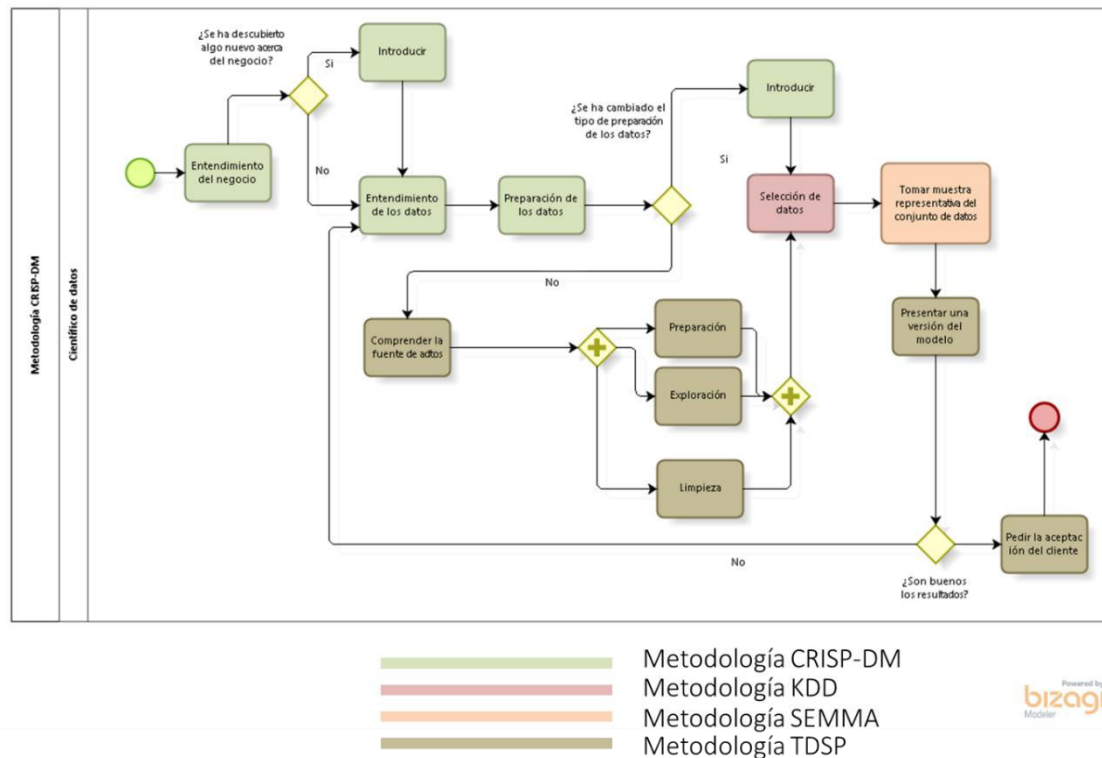


Metodología CSKT

En base a las metodologías explicadas dentro del presente capítulo es que se realiza el diseño de la metodología de la metodología CSKT en donde se consideran a algunos de los factores importantes de las metodologías CRISP-DM, KDD, SEMMA y TDSP en donde, mediante el siguiente diagrama se muestra la siguiente estructura.

Figura 15

Diseño de la metodología CSKT



En base al gráfico mostrado se tiene en cuenta que dentro de la metodología inicia con el entendimiento del negocio en donde se realiza una toma de decisión en el que, si se descubrió algo nuevo para el negocio es que se procede a introducir estos puntos nuevos dentro del proyecto para luego proceder al entendimiento de los datos, de ser el caso contrario se procede al entendimiento de los datos directamente, luego, se preparan los datos donde luego surge una condicional en donde sí se realizaron cambios en el tipo de preparación de los datos, de ser así, se introducen dichos cambios en el proyecto como registros y luego seleccionar estos datos se toma una muestra significativa dentro de dicho conjunto de datos, en caso contrario, se procede a realizar una comprensión de la fuente de los datos en los que, se realiza una preparación, exploración y limpieza con tal de obtener una clara y filtrada fuente de datos en donde luego se procede a la selección de estos nuevos datos volviendo a la toma de la muestra significativa de dicha data, luego de ello se procede a presentar la versión del modelo en donde si estos

resultados son factibles, se solicita la aceptación del cliente cerrando así el proceso de dicha metodología, de ser en caso contrario se retorna al entendimiento de los datos hasta que se cumplan los requisitos previos, asimismo, el diseño de la metodología fue aceptada por la Universidad Internacional de la Rioja en Perú (Ogosi, 2022).

2.2. Glosario de términos

CSKT: Metodología la cual toma referencia de otras metodologías enfocadas en la Data Mining las cuales son CRISP-DM, SEMMA, KDD, TDSP.

Entrenamiento: Término el cual hace referencia instruir al modelo de aprendizaje automático por medio de pruebas que ayuden con la mejora del mismo.

Sistema pensionario: Es la formación de normas, regulaciones y mecanismos los cuales rigen el modo de pago y financiamiento de las pensiones a trabajadores de un país o región.

Pensión de jubilación: Se le considera un privilegio monetario a personas quienes hayan alcanzado la edad de jubilación o que hayan cumplido ciertos criterios con el sistema nacional de pensión de un país.

Android: Sistema operativo de dispositivos móviles el cual es un intermediario para el acceso a aplicaciones móviles que muchos usuarios emplean para múltiples propósitos.

APP: Es un abreviado para el termino aplicación o application, es usualmente usado para hacer referencia a las aplicaciones informáticas desarrolladas para su aplicación en dispositivos electrónicos.

Database: Es aquella colección de la data archivada cuya información está estructurada y relacionada a un fin o temática la cual puede ser administrada de manera eficiente.

2.3. Estado del arte

El empleo de T.I. en el ámbito empresarial responde a las demandas y competencias del mercado laboral y productivo. Se busca mejorar diversos aspectos operativos con el fin de aumentar la eficiencia y el progreso de las empresas. Contextualmente, las TIC cumplen un rol imprescindible al influir en múltiples aspectos organizacionales. La contratación de personal cualificado y adecuado es crucial para el éxito de estas empresas, y cobra un papel relevante para el desempeño y desarrollo continuo. La selección y contratación efectivas de individuos idóneos, capacitados y comprometidos contribuyen significativamente a optimizar las operaciones empresariales. Es importante destacar que, una vez identificados y aceptados, estos profesionales pueden ocupar diversos roles en el entorno de las PYMES, generando mejoras progresivas en su funcionamiento y competitividad (Osorio, 2021). Por consiguiente, la relevancia de aplicativo web en el proceso organizacional que forma parte del objeto de estudio tiende a variar según los requerimientos específicos proporcionado por los usuarios, considerando además el crecimiento de la empresa y su necesidad de adoptar tecnologías que faciliten tanto la contratación de personal como la elección de aspirantes adecuados para las posiciones vacantes. En este sentido, dicha tecnología debe ser capaz de identificar y evaluar tanto a las habilidades y/o competencias necesarias para cada puesto, permitiendo así una contratación más efectiva y acorde al cargo. Por lo tanto, existen varias empresas que requieren encontrar soluciones que simplifiquen y agilicen los procedimientos relacionados a la contratación, asegurando así la selección de la persona adecuada para cada puesto disponible (Miranda, 2020), por otro lado, la aplicación de metodologías ágiles conlleva una serie de ventajas que facilitan la estructuración y administración de proyectos de software. Estas metodologías ofrecen un enfoque flexible que permite adaptarse a los cambios en el desarrollo del proyecto, ya que proporcionan herramientas para la elaboración progresiva y paso a paso de los diferentes componentes del software deseado. Además, permiten la distribución eficiente

de las tareas, asignando responsabilidades específicas a cada miembro del equipo de desarrollo. Esto facilita la comprensión y el seguimiento de las actividades totalmente requeridas para llevar a cabo el proyecto de manera efectiva. (Martín, 2020).

En base al desarrollo de software en donde se debe de aplicar posibles técnicas de innovación y manejo de nuevas tecnologías; Gil y Seguro (2022), destacan que la utilización de Machine Learning a fin de evaluar el rendimiento de proyectos de desarrollo de software. Su enfoque proporciona valiosas perspectivas destinadas a potenciar la eficiencia y calidad en este sector específico logrando determinar si las técnicas de machine learning resultan ser factibles o no para el ámbito en el cual se está enfocando, por otra parte, Forero y Bannasar (2024), en base a su investigación quienes han realizado un análisis estructurado que examina las técnicas y aplicaciones de ML e IA en la educación, enfatizando el impacto transformador de estas tecnologías en la educación remota en donde se determina que el uso de esta tecnología se destaca con el manejo de recursos necesarios para optimizar este método de educación sin presentar inconveniente alguno.

La aplicación de machine learning resulta una herramienta altamente aplicable para lo que son el desarrollo de herramientas como el reconocimiento de voz como de imágenes, la creación de aplicaciones de NLP y modelos de análisis predictivo, así como es altamente recomendable su aplicación para el desarrollo de herramientas enfocadas al análisis financiero o estrategias de marketing, sin dejar de tomar en cuenta su constante evolución y el potencial de descubrir nuevas aplicaciones para esta tecnología (Sharifani y Amini, 2023). Un ejemplo de ello se destaca en una investigación en donde se hizo una revisión sobre investigaciones en donde aplicaban machine learning en el tópico del COVID-19 enfocado a diversos objetivos y diferentes tipos de algoritmos, en donde de las revisión de 16 artículos 14 aplicaron el aprendizaje supervisado, uno de ellos aplico el aprendizaje no supervisado y el ultimo aplico

una combinación de ambos métodos, de las 14 investigaciones 5 aplicaron el algoritmo de regresión logística mostrando resultados prometedores en el ámbito de atención y aplicación de la salud en el tema del COVID-19, tres de ellos aplicaron redes neuronales artificiales los cuales también brindaron notables resultados, los demás 14 artículos que aplicaron una variedad de algoritmos tanto supervisados como no supervisados mostraron también resultados precisos, específicos y con alta sensibilidad (Kwekha et al., 2023).

En lo que respecta al campo médico, el uso de esta tecnología resulta ser relevante ante casos en los que se desea conocer con mayor precisión el estado de los pacientes, así como la búsqueda de historia médicas y determinar el diagnóstico médico mediante registros relacionándolos ante casos similares o existentes hasta el momento. Pineda (2022), resalta la utilización de modelos predictivos fundamentados en Machine Learning (ML), centrando su atención en cómo estos modelos pueden ser aplicados para potenciar los resultados en la prestación de servicios médicos cumpliendo con las expectativas tanto de los pacientes como del propio personal en lo que respecta al uso de esta tecnología, del mismo modo, Valdivieso et al. (2019), investigan las posibilidades de combinar mHealth, big data y ML como herramientas tecnológicas de apoyo en la salud humana colombiana, indicando el impacto beneficioso que esto tiene en el nivel de prestación de servicios y en la supervisión de la salud.

El uso de la tecnología de ML se basa en la recolección de los datos, un dato de vital importancia cuando se trata de labores y situaciones que involucran los procesos que se dan dentro de una empresa en el que, mediante una recolección de datos tomando en cuenta los criterios que la empresa solicita en un empleado, así como la información de los potenciales postulantes al cargo de trabajo, una vez dentro del puesto contratado por la entidad, se tiene un contrato bajo un salario dado de manera adecuada para que el empleado contratado se le pueda pagar en base a los datos recabados en su experiencia siendo esta la facultad de entrenamiento

de un ML la cual facilita el asunto de la contratación de personal en compañías que requieran de esta (Fajardo, 2023). Asimismo, en el tema de las pensiones se tiene en cuenta la utilización de los modelos de ML para el proceso de discernimiento de pagos de trabajadores en empresas en donde se tiene en cuenta el desempeño de los trabajadores que se dan durante los meses siendo este modelo el cual por medio del entrenamiento que se da en el proceso es que se asigna un sueldo en específico al trabajador cerciorándose de que estos reciban una pensión justa dependiendo de su rendimiento en lo laboral (Gil, 2022), en base al tema de la educación, se presentan casos en los que se desea emplear técnicas para optimizar el rendimiento académico mediante investigaciones o casos en los que ayuden con la educación de los estudiantes aplicando las técnicas de aprendizaje automático siendo el caso de Cruz et al. (2022), resaltan la aplicación de métodos de ML y destacan los aspectos en base a la recolección de información relacionada ante un tema o curso de educación, qué posibles métodos se emplearían dentro de una aula de clase y qué recursos se estarían empleando destacando libros que contengan los temas que se pueden enseñar durante el año escolar para los estudiantes, además, en un enfoque similar. Estrada et al. (2020), proponen un modelo selectivo del talento humano basado en ML, destacando la utilidad práctica de las RNA en los procedimientos de selección.

Asimismo, dentro de las soluciones que puede brindar esta tecnología, se debe tener en cuenta el modo en que se utiliza dado que para obtener un buen desempeño dentro de la solución ante muchos problemas, se toma como criterios la tasa de precisión en el cual se requiere determinar la cantidad de aciertos que se da durante el entrenamiento de esta inteligencia puesto que, antes de lanzar esta solución a la implementación, se requiere de una capacitación continua de la tecnología de aprendizaje automático ante el problema al cual se incorpora para cumplir con su propósito y no generar mayores dificultades en el proceso (Mon, 2022), dentro del tema de rendimiento, se considera que, para una mejor calidad de servicio de esta tecnología frente a un proceso el cual se está incorporando, se debe considerar el tema de

la eficiencia del modelo de Machine Learning ya que, al momento de establecer las funciones y criterios dentro de dicho modelo, se debe de realizar un constante entrenamiento poniéndolo a prueba dentro del proceso el cual se está integrando con la finalidad de que, al momento de emplear todos los recursos que estén a su disposición, no deba consumir demasiado espacio ni recursos innecesarios debido a la cantidad de fallos que pueda obtener durante su funcionamiento (Mirjalili y Raschka, 2020). No obstante, para realizar y proceder con las pruebas iniciales del aplicativo con la tecnología de aprendizaje automático, es necesario emplear un prototipo con la finalidad de evaluar su desempeño en lo que respecta al proceso a practicar puesto que, esto no solo brinda la ventaja de que los propios desarrolladores puedan visualizar qué aspectos puede mejorar dicho software sino que, además, los propios usuarios quienes manejen la aplicación puedan tener una primera perspectiva de cómo es la aplicación una vez se tenga el producto final, cosa que puede implicar en la mejora de algunas funciones o en las modificaciones de otras funciones (Ortiz y Andres, 2021). También, en base a la evaluación del software, se tienen en cuenta las métricas de validación de ML con la finalidad de que esta al momento de realizar constantes entrenamientos se pueda determinar si el desempeño que esta revela puede beneficiar a la misma organización que requiera de dicha tecnología, para ello se tiene en cuenta emplear la métrica de F1 – Score puesto que dentro de dicho indicador se conforma con las métricas del tema de accuracy y recall con el detalle de que en este punto ayuda a que en la aplicación se logre determinar la cantidad de predicciones acertadas durante el entrenamiento si en caso esta haya obtenido una mejora en la dichas predicciones o hubo un deceso, esto se tiene en cuenta si se desea validar completamente el funcionamiento y efectividad de la herramienta en cuestión para la organización a la cual empieza con su labor (Borja et al., 2020).

III. MÉTODO

3.1. Tipo de Investigación

Básica

La siguiente metodología aumenta probabilidades en las organizaciones de implementarla en el desarrollo móvil con machine learning para el proceso de objeto de estudio que se necesita para el siguiente proyecto.

Se efectúa la revisión de la utilización de diferentes metodologías, lo que permite utilizar el pensamiento sistémico para integrar las ventajas y aportes acorde a la elaboración de una nueva metodología de bajos costos de implementación para el sector privado.

Se entiende por investigación básica dado que se busca una investigación a base de hechos encontrados en otras investigaciones, es decir, se basa en la teoría recolectada de otros documentos relacionados con el hecho o fenómeno a investigar (Esteban, 2018).

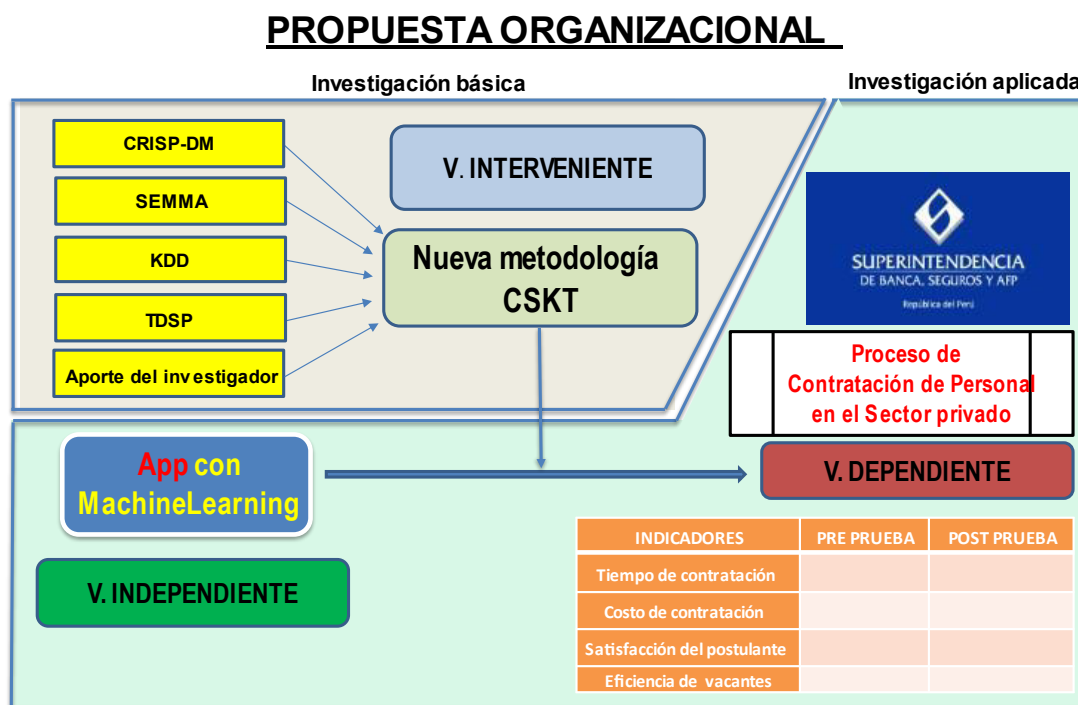
Aplicada

La siguiente metodología realiza un cambio positivo a la problemática que genera la contratación de personal que debe ser aplicada para resolver los problemas en el sector privado con la solución de una App con Machine Learning.

Se entiende por investigación aplicada a dichos hechos o soluciones que se realizan a partir de la teoría rebuscada en la investigación de tipo básica siendo que esta solución influye dentro del objeto o zona de estudio donde se está realizando dicho experimento (Esteban, 2018).

Figura 16

Propuesta organizacional



En base a la figura 16 sobre la propuesta organizacional se hace el estudio de los componentes de la investigación básica y aplicada en donde se muestra la variable interviniente que es la nueva metodología CSKT conformada por metodologías de Data Mining, resaltando los aspectos ágiles para el respectivo proyecto, por otro lado (en base a la investigación aplicada), se muestran las dos variables a realizar la investigación cuantitativa en donde el objeto de estudio a evaluar es la aplicación con machine learning, después se muestra la variable dependiente: “proceso de contratación” complementado con los indicadores (identificados explicados en el marco teórico).

3.1.1. Nivel de Investigación

○ Descriptiva

En este estudio se detalla las consecuencias de asumir los riesgos en la elección de un sistema pensionario dentro del proceso de Contratación laboral.

Asimismo, se describe el comportamiento asumido por falta de la cultura previsional de los postulantes.

Dentro de este nivel se busca definir lo que se mide dentro de la investigación, asignando a los responsables para realizar y proceder con la recolección de la data imprescindible (Esteban, 2018).

○ **Explicativa**

Esta propuesta de una metodología diferente que asuma la tarea de traer una cultura previsional que pueda ser adaptada en la provincia de Lima por los altos tiempos de contratación de los postulantes; asimismo, esto podrá ser aplicado sin problemas en todas las provincias del Perú.

Esta investigación se diferencia de la descriptiva dado que la magnitud del estudio es más profundo y complejo siendo el principal uso de las hipótesis planteadas su principal propósito en dicho nivel en donde se busca descubrir nuevas teorías que ayuden con la investigación en cuestión (Esteban, 2018).

3.2. Población y muestra

Tabla 5

Población y muestra

Unidad muestral	Proceso de Contratación de personal Restricciones: <ul style="list-style-type: none"> • Empresas formalizadas con gestión de planillas. • Empresas del sector privado. • Nivel nacional.
Universo	Todos los procedimientos relacionados con la selección del personal en organizaciones formalizadas del sector privado con gestión de planillas a nivel nacional. Debido a que el número de procedimientos no se puede determinar, se tiene: N = indeterminado

Muestra	Procesos de Contratación del personal de la compañía privada con gestión de planillas a nivel nacional. n = 5000
---------	---

Nota. Considerando que la presente investigación no tiene la capacidad para abordar a todos las actividades de Selección de recursos humanos en la industria del sector privado a nivel nacional. Sería ideal trabajar siempre con una población, pero esto no suele ser posible. En conclusión, se tiene una muestra según Zambrana et al (2020), que realmente puede representar a la población mediante la cantidad de 5000 transacciones.

3.2.1. Tipo de muestreo

Tabla 6

Tipo de muestreo

Unidad muestral	Proceso de Contratación de personal Restricciones: <ul style="list-style-type: none"> • Empresas formalizadas con gestión de planillas. • Empresas del sector privado. • Nivel nacional.
Universo	Todos los procedimientos relacionados con la selección del personal en organizaciones formalizadas del sector privado con gestión de planillas a nivel nacional. Debido a que el número de procedimientos no se puede determinar, se tiene: N = indeterminado
Muestra	Procesos de Contratación de personal en una empresa formalizada del sector privado formalizadas con gestión de planillas a nivel nacional. n = 5000
Tipo de muestreo	Aleatorio

Nota. El tipo de muestreo a realizarse es aleatorio según (Zambrana et al., 2020).

Con el objetivo de que los procedimientos de Reclutamiento de recursos humanos en organizaciones formalizadas del sector privado con gestión de planillas a nivel nacional sean parte de una muestra aleatoria se haría una selección por sorteo de los 5000 procesos mediante la base de datos y posteriormente se evaluaría individualmente cada una de las transacciones.

3.2.2. *Diseño de la investigación*

Diseño de Post-prueba con grupo de control:

El diseño preexperimental consiste en emplear pruebas pre y post en base a la muestra seleccionada en el estudio con la finalidad de determinar los efectos de una solución propuesta por los investigadores obteniendo resultados ya sean positivos o negativos (Hernández y Mendoza, 2018).

Figura 17

Diseño preexperimental en base a pre test y post test



En este caso, para la investigación se considera emplear el diseño pre experimental en base a aplicar pruebas pre y post para una sola muestra para determinar los posibles efectos que se obtienen luego implementar la app con la tecnología de aprendizaje automático.

3.3. Operacionalización de variables

Tabla 7

Matriz de operacionalización de variables

Tipo	Variable	Dimensión	Indicadores	Descripción
Independiente	Machine Learning	Manejo de recursos	Tasa de precisión del modelo	Índice o porcentaje de todos los valores verdaderos y positivos en base a todas las predicciones que el modelo de aprendizaje automático realiza durante una operación (Bianco & Philip, 2022). Alcance que tiene el modelo en obtener excelentes niveles de rendimiento por medio de los recursos que tenga a su disposición y los tiempos reducidos de su entrenamiento logrando obtener niveles de eficiencia deseados (Mirjalili y Raschka, 2020). Desarrollo completo del modelo de la aplicación móvil con tecnología de aprendizaje automático en el que, se considera que esta cumpla cada uno de los requerimientos ya establecidos en el transcurso del desarrollo de las pruebas de su funcionamiento (Ortiz & Andres, 2021). Se toma en cuenta los indicadores previos los cuales son la Precisión y el Recall, haciendo que se pueda predecir mejor los resultados positivos a diferencia de la métrica de Accuracy (Borja et al., 2020).
			Eficiencia del modelo	
		Funcionalidades	Evaluación del prototipo	
		Adaptabilidad	F1 – Score	
Interviniente	Metodología CSKT en el modelo de machine learning	Metodología	Aplicabilidad	Implica la adecuación y pertinencia de la metodología CSKT para enfrentar y mejorar el procedimiento de contratación de personal en el ámbito privado. Este aspecto implica evaluar la capacidad de la metodología para ajustarse y demostrar eficacia en un contexto particular.
		Aplicación móvil	Usabilidad	Implica la facilidad con la cual los usuarios, especialmente los profesionales responsables del proceso de contratación pueden emplear y relacionarse con la metodología CSKT. Se centra en la evaluación de la eficacia, eficiencia y grado de satisfacción de los usuarios.
Dependiente	Proceso contratación de personal	Elección de personal	Tiempo de contratación	Según Jaramillo (2020), permite profundizar en los tiempos como el comienzo de la fecha del proceso y la fecha límite de entrega de postulantes, teniendo en cuenta el t. de atención requerido por la prioridad y volumen de los procesos considerando el Acuerdo de Nivel de Servicio establecido.
			Costo de contratación	Según Guamán (2020), desarrolla efectivamente un proceso de selección pensando en los costos que se requieren para llevarlo a cabo. Entre estos, se pueden mencionar algunas cantidades: vacantes, horas de trabajo, comidas y, en algunos casos, alojamiento.
		Satisfacción	Satisfacción del postulante	Según Sánchez (2020), logra que el compromiso y la lealtad de los empleados no se limiten a la satisfacción laboral individual, sino estén ligados a su desempeño y niveles de eficiencia, potenciando la buena reputación de una empresa en particular.
		Manejo de vacantes	Eficiencia de vacantes	Según Hanze (2020), tiene por objetivo agilizar el proceso exploratorio de los postulantes en las compañías que utilicen este servicio y, asimismo, el facilitar la exploración y contratación de nuevos empleados que necesiten dar respuesta a las vacantes.

3.4. Instrumentos

Se procede a emplear los instrumentos de fichas de observación enfocados a los procesos dentro de la variable dependiente para luego utilizar los datos recabados para los futuros resultados de posteriores investigaciones.

Tabla 8

Técnica e instrumento para recolección de la información

Técnica	Instrumento
La Observación Indirecta: <ul style="list-style-type: none"> • Consulta a BD 	<ul style="list-style-type: none"> • Fichas de observación

3.5. Procedimientos

En esta investigación, se utilizan fichas de observación como instrumentos, centrándose en los procesos asociados con la variable dependiente. Posteriormente, se emplea los datos recopilados para obtener los resultados futuros.

3.6. Análisis de datos

Las etapas para el análisis de resultados cuantitativos y cualitativos resultan efectuadas tras las siguientes fases:

Figura 18

Fases para el análisis de datos

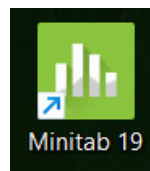


Fase 1:

Se selecciona el programa estadístico Minitab para el análisis de datos.

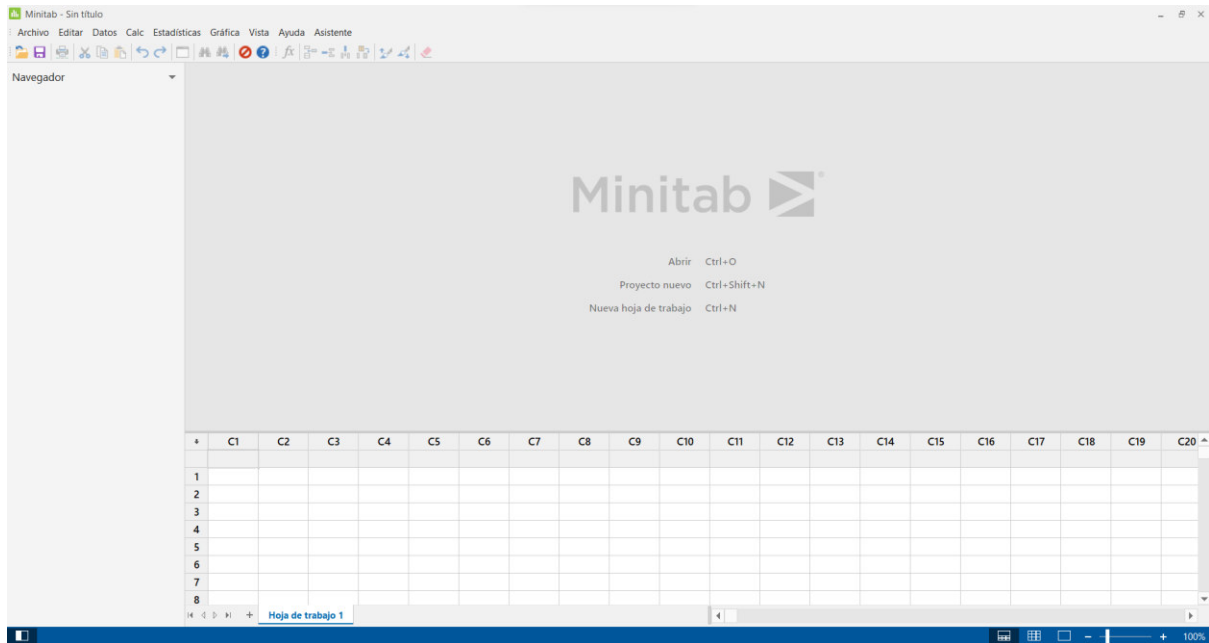
Figura 19

Icono de la herramienta Minitab



Fase 2:

Se ejecuta el programa Minitab para el análisis de datos.

Figura 20*Pantalla principal de Minitab***Fase 3:**

Se procede con el mapeo de la data, lo que conlleva el análisis y la representación visual de cada variable de investigación.:

- **Distribución de frecuencias:** Histogramas, Tipo Pastel, Tabla de frecuencias, Los polígonos de frecuencias.
- **Medida de tendencia central:** moda, mediana, media.
- **Medidas de variabilidad:** El rango, la desv. est. o característica.
- **Otras estadísticas descriptivas:** La asimetría, la curtosis.

Figura 21

Histograma de muestra de la herramienta Minitab

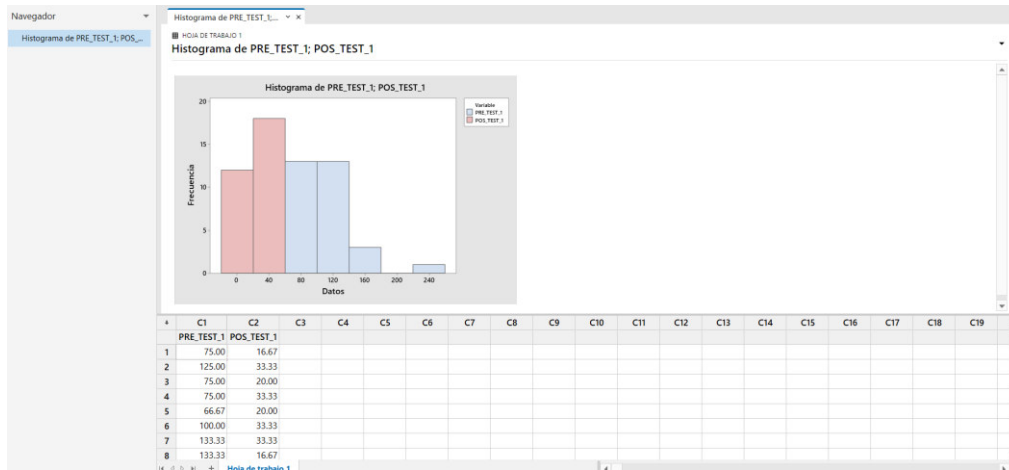
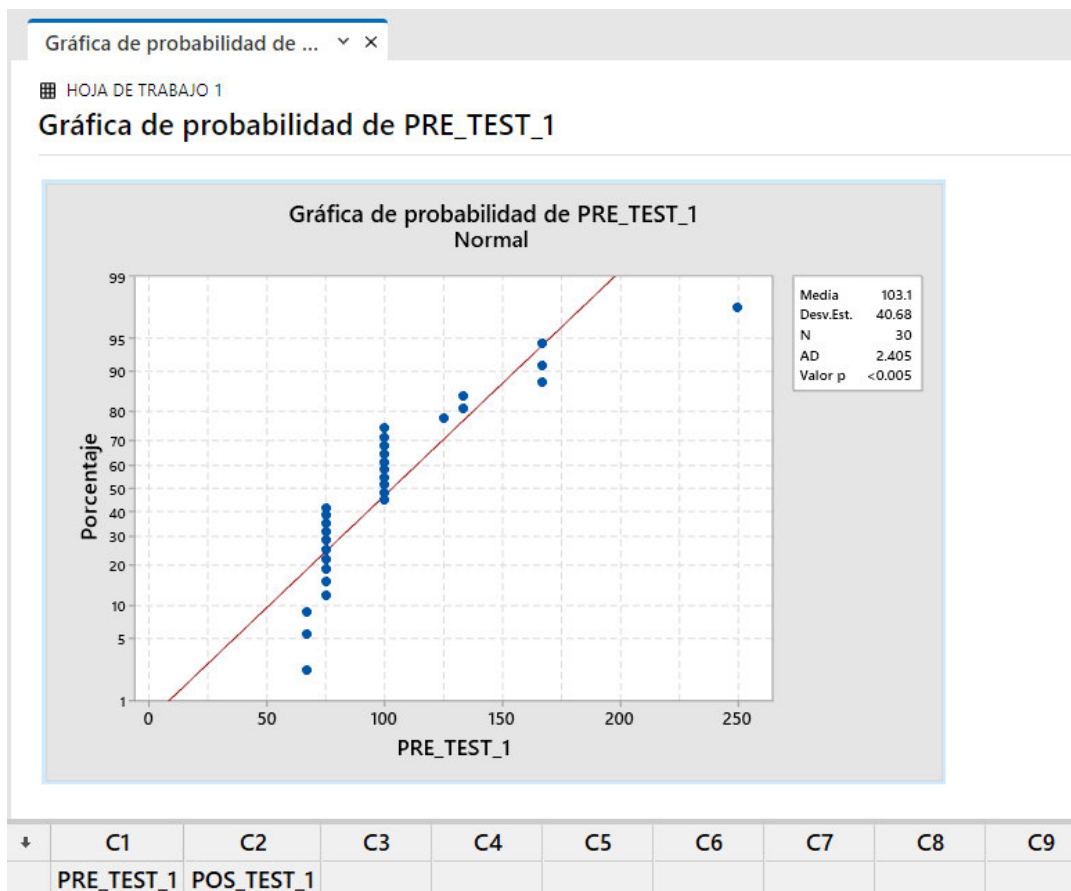


Figura 22

Gráfico de probabilidad de muestra



En base a las pruebas de normalidad se procede a emplear la prueba de Shapiro Wilk siendo que la muestra seleccionada es menor a 50 en donde se evalúa si la significancia es menor o mayor al 5% del error estimado determinando así, si el indicador presenta un comportamiento normal o no (Flores y Flores, 2021). S continuación, se presentan las ecuaciones para la prueba de normalidad a través de las siguientes figuras (Flores et al., 2019):

Ecuación 1

Fórmula de Shapiro Wilk

El cálculo de la fórmula de Shapiro Wilk se utiliza en estadística para probar la normalidad de una muestra. Un valor alto de W sugiere que la muestra no proviene de una distribución normal, tal como se muestra en la ecuación (1).

$$W = \frac{b^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

Nota. Ecuación de la prueba de normalidad mediante Shapiro Wilk de Flores et al., 2019, *Perfiles*, 1(21)

Ecuación 2

Fórmula del valor b de la prueba

Esencialmente, b mide cómo se relacionan los datos observados con lo que se esperaría en una distribución normal. Si los datos se ajustan bien a una distribución normal, b se acerca a 1. Sin embargo, si los datos muestran desviaciones significativas de la normalidad, b será menor, lo que sugiere que la muestra no proviene de una distribución normal. La fórmula se muestra en la ecuación (2).

$$b = \sum_{i=1}^n a_i [X_{(n-i-1)} - x_i] \quad (2)$$

Nota. Fórmula del valor b de la prueba de normalidad de Shapiro Wilk de Flores et al., 2019, *Perfiles*, 1(21)

Ecuación 3

Fórmula de los coeficientes tabulados de la prueba

Esta fórmula proporciona los valores críticos o de referencia para determinar si los resultados de una prueba son estadísticamente significativos. Estos coeficientes son utilizados para comparar el valor calculado de una estadística de prueba con un valor crítico específico, basado en un nivel de significancia predefinido y los grados de libertad asociados con la prueba. Como se aprecia en la ecuación (3).

$$a_i = m'V^{-1}(m'V^{-1}m)^{-1/2} \quad (3)$$

Nota. Fórmula de los coeficientes tabulados de la prueba de normalidad de Shapiro Wilk de Flores et al., 2019, *Perfiles*, 1(21)

Donde:

$[X_{(n-i-1)} - X_1]$: Es la resta que se da entre el primer y último valor de la muestra seguido del segundo con el penúltimo de forma sucesiva

α_1 : Se le considera como uno o más coeficientes tabulados dentro de la tabla de Shapiro.

m : Se le considera como el valor medio del estadístico ordenado ya sea creciente o decreciente de toda la muestra seleccionada.

V: Es la matriz de covarianzas del estadístico, es decir, de la muestra seleccionada de forma organizada.

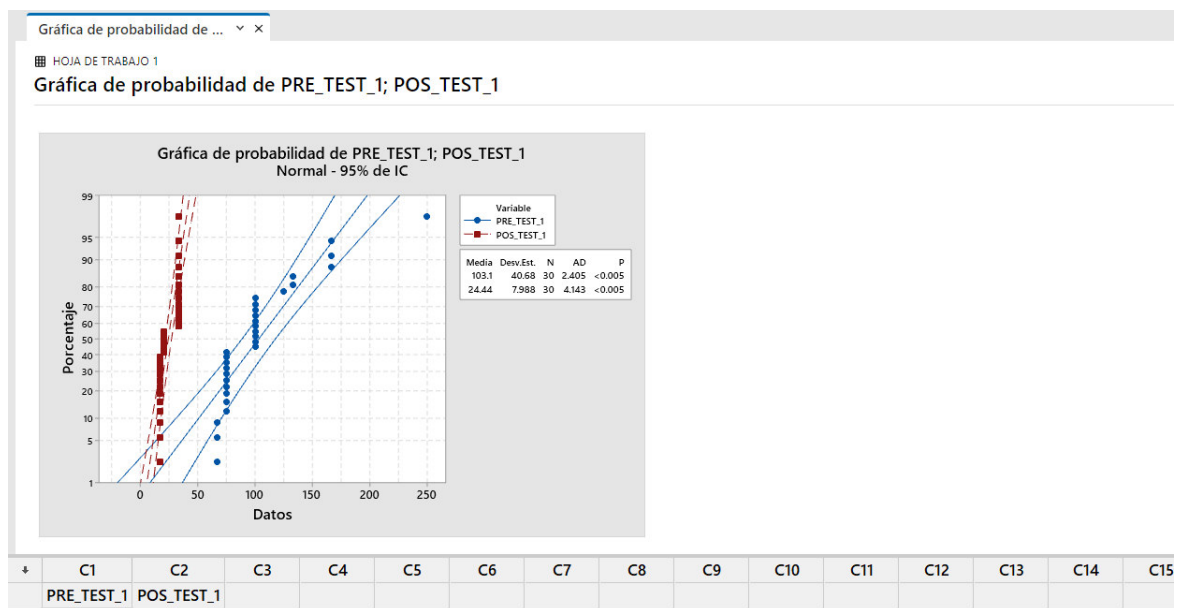
Fase 4:

Se realiza el análisis estadístico inferencial para probar las hipótesis planteadas y estimar los parámetros:

- **Nivel de significancia (α):** Es equivalente a 0.05.
- **Prueba de hipótesis:** Análisis paramétrico, prueba t de student, prueba de Mann-Whitney.

Figura 23

Gráfico de probabilidad de dos muestras de ejemplo

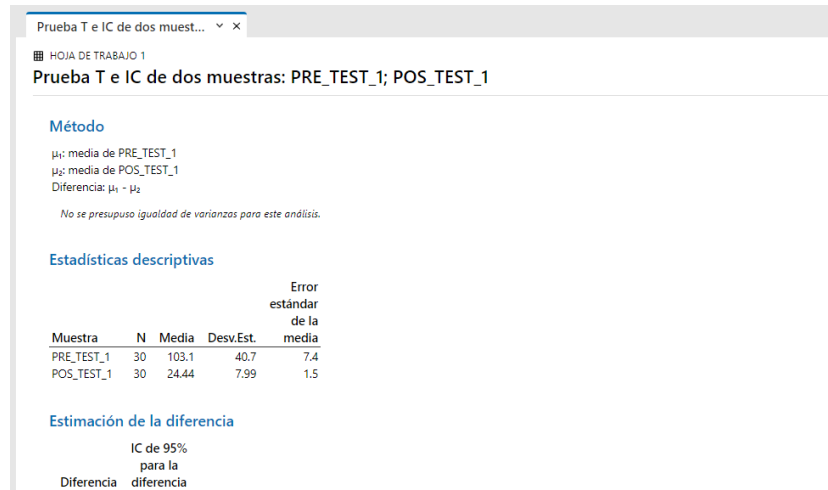


Fase 5:

Se efectúa análisis adicionales como la prueba t de 2 muestras.

Figura 24

Prueba de hipótesis de T de student



Fase 6:

Se preparan los resultados para presentarlos mediante tablas, gráficas, figura, cuadros.

Figura 25

Estadísticas descriptivas de prueba

Estadísticas descriptivas

Muestra	N	Media	Desv.Est.	Error estándar de la media
PRE_TEST_1	30	103.1	40.7	7.4
POS_TEST_1	30	24.44	7.99	1.5

Estimación de la diferencia

Diferencia	IC de 95% para la diferencia
78.61	(63.18; 94.05)

Prueba

Hipótesis nula	$H_0: \mu_1 - \mu_2 = 0$	
Hipótesis alterna	$H_1: \mu_1 - \mu_2 \neq 0$	
Valor T	GL	Valor p
10.39	31	0.000

Tabla 9*Elección de las pruebas estadísticas*

Pregunta de investigación Diferencia de grupos	Prueba
Dos indicadores o grupos:	
<ul style="list-style-type: none"> • Muestras independientes • Datos ordinales 	<ul style="list-style-type: none"> • Prueba de Mann-Whitney U • t de Student

Nota. Después de completar la evaluación de normalidad, para, posteriormente, hacer uso de la prueba de T de Student, esta considera: si el valor p dentro de la prueba se considera si el valor de T se encuentra dentro de la zona de aceptación en base al grado de libertad (gl) y a la significancia estimada del 5% (Molinas et al., 2020), representándose por medio de la presente figura y la siguiente tabla:

Ecuación 4*Fórmula de la prueba T de Student*

La estadística de prueba t se compara con un valor crítico de la distribución t de Student para determinar si hay una diferencia significativa entre las medias de las dos muestras. Esta prueba se utiliza comúnmente para comparar los promedios de dos grupos y determinar si la diferencia observada es estadísticamente significativa. Tal como se aprecia en la ecuación (4).

$$T = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad (4)$$

Nota. Fórmula de la prueba de hipótesis de T de Student la cual se enfoca en la investigación, Cascante y Villacís, 2022

Donde:

T: Se le considera el punto crítico

x: La media o promedio de la población seleccionada

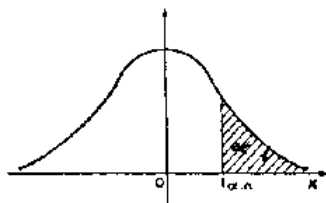
μ : La media de la muestra seleccionada

σ : Desviación estándar de la muestra

D: el tamaño de la muestra

Figura 26

Tabla de distribución de T de Student



$\alpha/2$	0,40	0,30	0,20	0,10	0,050	0,025	0,010	0,005	0,001	0,0005
1	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22	12,94
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032	5,893	6,859
6	0,263	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,535	0,863	1,337	1,746	2,120	2,583	2,921	3,686	4,013
17	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,611	3,922
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,256	0,531	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,255	0,529	0,851	1,303	1,648	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,849	1,298	1,676	2,009	2,403	2,678	3,262	3,495
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	0,254	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,415
100	0,254	0,526	0,845	1,290	1,660	1,984	2,365	2,626	3,174	3,389
200	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,339
500	0,253	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,106	3,310
∞	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Nota. Gráfico de distribución de la prueba T de Student considerando el grado de libertad y el error significativo, Fistierra, 2001, <https://www.fisterra.com/formacion/metodologia-investigacion/metodos-parametricos-para-comparacion-dos-medias.-t-student/#sec0>

El caso es el siguiente respecto a la figura 28, si el grado de libertad y el error significativo seleccionan por ejemplo en el caso de la investigación el valor 1.725, este se distribuiría en un rango entre el valor negativo y valor positivo, si el valor p se encuentra dentro del rango especificado, en consecuencia, se deniega la H_0 y, en tanto que, se afirma H_a .

3.7. Consideraciones éticas

Una consideración ética crucial es la adecuada atribución de fuentes externas. Esto implica citar cualquier información, ya sea proveniente de literatura científica, código fuente o contenido multimedia como videos de YouTube, que haya contribuido al desarrollo del trabajo de investigación. La omisión de esta atribución puede conducir a problemas de plagio y falta de reconocimiento a los autores originales. Es esencial que los investigadores se comprometan a realizar citas exhaustivas y detalladas, incluso en casos donde la información citada sea discontinua, para mantener la integridad académica y respetar el trabajo intelectual de otros.

Además, la consideración ética en la investigación también abarca el tratamiento justo y equitativo de los datos y participantes involucrados en la investigación, esto conlleva a solicitar a los encuestados, el consentimiento informado; así como de salvaguardar la privacidad y confidencialidad de los datos recopilados, y garantizar que cualquier uso de la información recopilada sea éticamente justificado y beneficioso para la sociedad. Los investigadores deben estar comprometidos con la práctica de la investigación responsable, asegurándose de que sus acciones no solo cumplan con los estándares éticos establecidos, sino que también contribuyan positivamente al avance del conocimiento científico y al bienestar de la comunidad en general. Toda información que no pertenece al autor de la presente investigación fue citada, siendo posible realizar un ejemplo, aunque discontinuado de la citación a nivel detallado:

Tabla 10*Investigación con ética*

Palabras clave			
Del título	Relacionados al título	Tipo de diseño de la investigación	Técnica estadística
Implementación	CRISP-DM	Tipo de diseño de la investigación	regresión lineal múltiple
App	SEMMA	Estudio experimental	regresión logística binaria
Machine	KDD	Estudio descriptivo	regresión logística multinomial
Nueva	Estudio de tendencias	Estudio explicativo	análisis factorial exploratorio
Metodología	T. de contratación	análisis de componentes principales	análisis factorial confirmatorio
CSKT	Inversión en reclutamiento	Clasificación	análisis de conglomerados
Proceso	Complacencia del candidato	Modelo	modelamiento de ecuaciones
Contratación	Rendimiento de las vacantes	Tipología	estructurales
Personal	selección de personal	Taxonomía	Promedio móvil
Sector	Revisión bibliográfica	Revisión de literatura	
privado	Reclutamiento	Holt-Winter	
	reclutamiento y selección	Revisión sistemática	
	candidatos a puestos de trabajo	Evolución histórica	
	política activa de mercado de trabajo		
Nuevas palabras encontradas			
Palabras	Referencia directa		
selección de personal	Nabeeh, N. A., Smarandache, F., Abdel-Basset, M., El-Ghareeb, H. A., & Aboelfetouh, A. (undefined 2019). An integrated neutrosophic-TOPSIS approach and its application to personnel selection: A new trend in brain processing and analysis. IEEE access: practical innovations, open solutions, 7, 29734–29744. https://doi.org/10.1109/access.2019.2899841		
modelos de aprendizaje	Solano, J. A., Lancheros Cuesta, D. J., Umaña Ibáñez, S. F., & Coronado-Hernández, J. R. (2022). Predictive models assessment based on CRISP-DM methodology for students' performance in Colombia - Saber 11 Test. <i>Procedia Computer Science</i> , 198, 512–517. https://doi.org/10.1016/j.procs.2021.12.278		
reclutamiento	Pessach, D., Singer, G., Avrahami, D., Chalutz Ben-Gal, H., Shmueli, E., & Ben-Gal, I. (2020). Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. <i>Decision Support Systems</i> , 134(113290), 113290. https://doi.org/10.1016/j.dss.2020.113290		
reclutamiento y selección	Hmoud, B. (s/f). Will Artificial Intelligence take over humanresources recruitment and selection? <i>Seaopenresearch.eu</i> . Recuperado el 4 de febrero de 2022, de https://seaopenresearch.eu/Journals/articles/NIS_13_3.pdf		
candidatos a puestos de trabajo	Liem, C. C. S., Langer, M., Demetriou, A., Hiemstra, A. M. F., Sukma Wicaksana, A., Born, M. P., & König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. En <i>The Springer Series on Challenges in Machine Learning</i> (pp. 197–253). Springer International Publishing.		
Estructura tentativa bajo ética			
Dimension 1 (autor1,año;autor2,año)			
item D1.1 (autor, año, p. N; autor4, año, p. N)			
item D1.2			
item D1.3			
item D1.4			
Dimension 2			
item D2.1			
item D2.2			
item D2.3			
item D2.4			

Nota. Adaptado del taller de elaboración de artículos científicos por Emigdio Antonio Alfaro Paredes (2022). Este método consiste en establecer cruce de palabras o combinaciones para descubrir nueva información. Luego de descubrir nuevas palabras claves, se ha citación y se propone una estructura para su posterior consulta.

IV. RESULTADOS



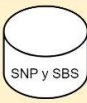



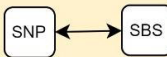

4.1. Verificación de la eficacia de la metodología CSKT aplicada en el modelo de Machine Learning

4.1.1. Diseño del Modelo

4.1.1.1. Definición de los parámetros y actividades considerados en el modelo de Machine Learning bajo la metodología CSKT

Figura 27

Parámetros y actividades del modelo de Machine Learning con la metodología CSKT

ML \ CSKT	Modelo Empresarial	Preparación	Exploración	Visualización
Obtención de la DATA	Entendimiento del negocio	Entendimiento de los datos	Selección de datos	
ETL		Preparación de los datos	Preparación - Exploración - Limpieza	
Entrenamiento del modelo		Toma muestra de DATASETS	Comprender fuentes de datos	 Modelo preparado
Prueba de la DATA	 Usuario		 Visualización de datos	Presentar versión del modelo
Optimización				Aceptación del cliente

Parámetros del modelo de ML:

- **Obtención de DATA:** se obtiene la data de la fuente del SNP y de la SBS.
- **Extracción, Transformación y Carga:** Se realiza el filtro de la data del SNP y de la SBS.
- **Entrenamiento de modelo:** Se procede con el entrenamiento del modelo de ML dentro del sistema nacional de pensiones.
- **Prueba de DATA:** Se realizan las respectivas pruebas de la data seleccionada de la fuente del sistema nacional de pensiones y de la SBS.

- **Optimización:** Se concluye con los respectivos cambios para su futura entrega del cliente.

Actividades de la metodología CSKT:

- **Entendimiento del negocio:** El proceso de entendimiento del negocio implica no solo comprender la naturaleza y los objetivos del negocio en general, sino también enfocarse específicamente en cómo el modelo de aprendizaje automático (ML) puede contribuir a esos objetivos. Se realiza un detallado análisis de los requerimientos comerciales, identificando áreas donde el ML puede generar valor agregado, como la optimización de procesos, la personalización del servicio al cliente o la predicción de tendencias del mercado. Este enfoque garantiza que el modelo ML esté alineado estratégicamente con las metas y objetivos comerciales a largo plazo.
- **Entendimiento de la data:** Para comprender plenamente la data disponible, se realizó un procedimiento analítico detallado y completo sobre la información extraída sobre las fuentes del sistema nacional de pensiones y de la SBS. Esto implica identificar el tipo de datos disponibles, como datos demográficos, financieros o históricos, y evaluar su calidad y relevancia para el desarrollo del modelo ML. Además, se estudia cómo estos datos pueden ser utilizados para abordar los desafíos específicos del negocio y cómo pueden ser integrados de manera efectiva la modelización.
- **Selección de datos:** Después de identificar los datos relevantes, se procede a simplificarlos mediante la selección de una muestra representativa. Esto implica identificar las variables clave que son necesarias para el análisis y el desarrollo del modelo ML, descartando aquellos datos que no contribuyen significativamente a los objetivos comerciales establecidos. La selección cuidadosa de datos garantiza que el

modelo se entrene y se evalúe utilizando la información más relevante y significativa, resultando con precisión y utilidad.

- **Preparación de los datos:** esta etapa resulta crucial, debido a que, implica una serie de acciones como depurar datos, modificar variables y fusionar conjuntos de datos, con el objetivo de asegurar que estos estén aptos. Durante esta fase, se detectan y corrigen valores atípicos, se codifican variables categóricas y se normalizan los datos para mantener su coherencia y calidad. Una adecuada preparación de datos es fundamental para garantizar la precisión y eficacia del modelo de aprendizaje automático generado.
- **Preparación – Exploración – Limpieza:** Tras la selección inicial de datos, se procede con una fase de preparación, exploración y limpieza más detallada. Durante esta etapa, se clasifican y ordenan los datos para comprender mejor su estructura y distribución. Se llevan a cabo análisis estadísticos y visualizaciones de datos para identificar posibles problemas, como datos faltantes o valores atípicos, y se aplican técnicas de limpieza para abordar estos problemas y garantizar la integridad de los datos. Esta fase es crucial para obtener una muestra de datos más clara y definida, lista para su uso en el proceso de modelización.
- **Toma muestra de DATASETs:** se procede a tomar muestras de los DATASETs resultantes, luego de finalizar el proceso de ETL. Estas muestras representativas proporcionan una visión general de la data final obtenida procedente del SNP y de la SBS. Se utilizan herramientas de visualización de datos para explorar y analizar la estructura y las características de los datos, ofreciendo una comprensión más profunda previa a continuar con el análisis y la modelización.
- **Comprender fuentes de datos:** Una vez que se ha filtrado la data, se realiza un análisis adicional para confirmar si es la adecuada para presentar el modelo o si se requiere repetir el proceso. Este análisis implica realizar una evaluación respecto a la calidad, la

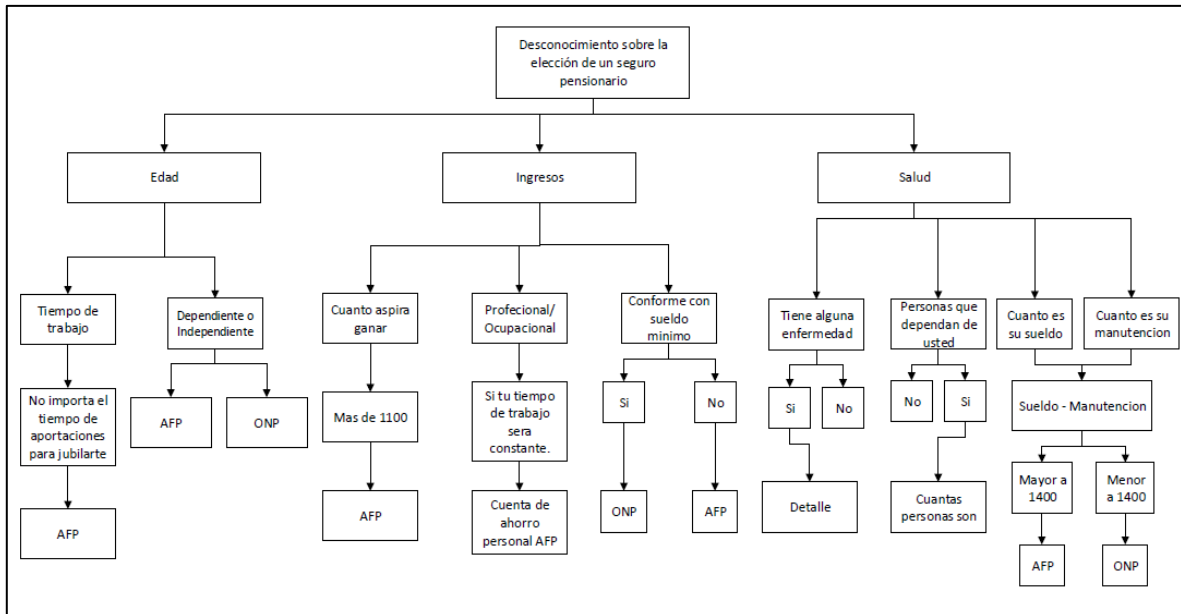
integridad y la preponderancia de la data seleccionada para garantizar que sean apropiados para el desarrollo del modelo ML. Se identifican posibles sesgos o limitaciones en los datos y se toman medidas para mitigar su impacto en los resultados del modelo. Si es necesario, se repite el proceso de selección y preparación de datos hasta que se obtenga una muestra adecuada y representativa para el análisis y la modelización.

- **Presentar versión del modelo:** Se presenta la primera versión del modelo de ML, utilizando los datos ingresados y seleccionados previamente. Esta versión inicial del modelo se utiliza para realizar pruebas preliminares y evaluar su rendimiento y precisión. Se realizan ajustes y refinamientos según sea necesario antes de proceder con la validación final y la implementación del modelo.
- **Aceptación del cliente:** Una vez que se cumplen con los criterios establecidos y se ha validado internamente el modelo de ML, se procede a entregarlo al cliente para su validación final. Se proporciona al cliente la oportunidad de revisar el modelo y proporcionar retroalimentación adicional antes de su implementación final. Se realizan ajustes según las sugerencias del cliente para garantizar que el modelo cumpla con sus expectativas y requisitos específicos.

4.1.1.2. Diseño del esquema preliminar del modelo ML

Figura 28

Árbol de decisión sobre la elección del seguro pensionario



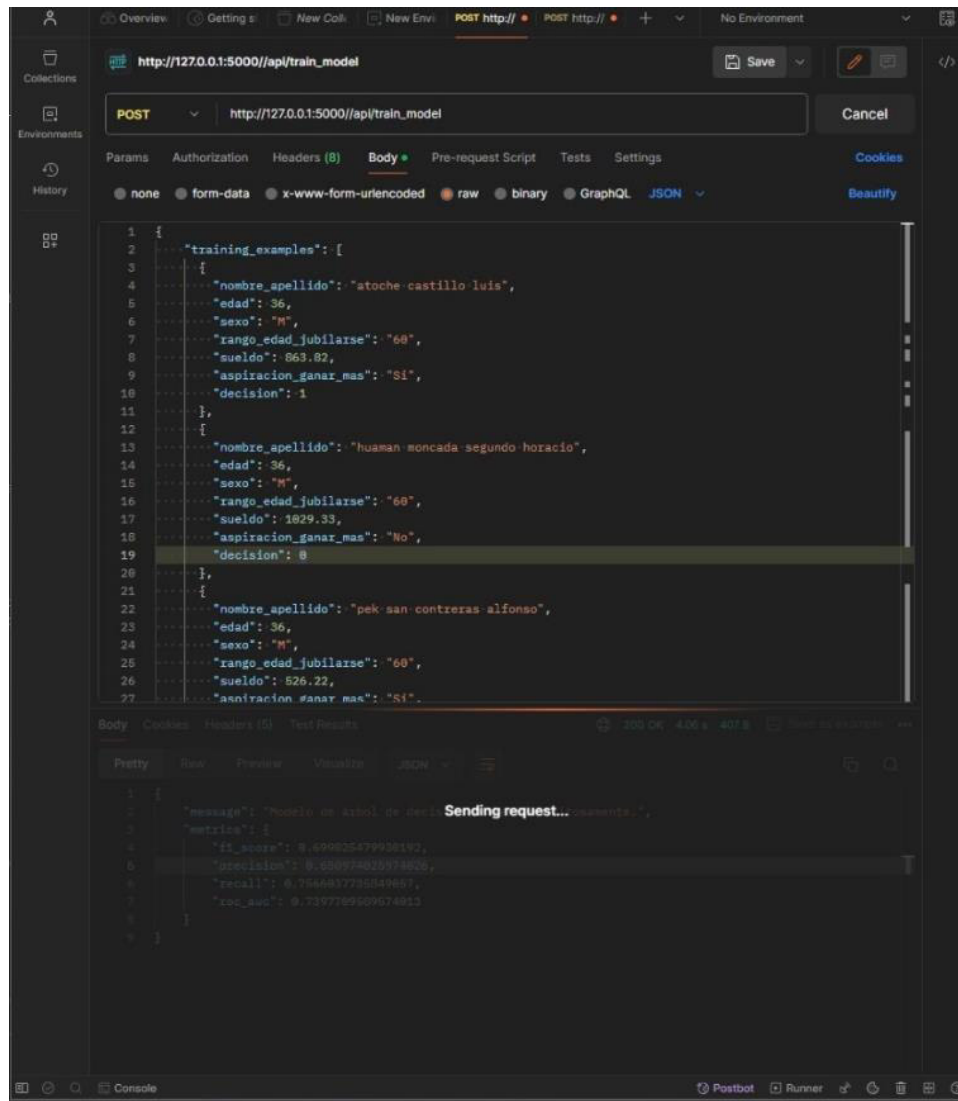
En el árbol de decisión se presentan las categorías de Edad, Ingresos y Salud, en la categoría de salud se presentan dos opciones que son “Tiempo de trabajo” y “Dependiente o Independiente”, en “Tiempo de trabajo” se muestra que no importa el tiempo de aportaciones para jubilarse, por lo que se procede a una AFP, mientras que, por otro lado, se puede seleccionar AFP u ONP, en la categoría de Ingresos se tienen tres opciones, los cuales son “la cantidad que aspira ganar el usuario”, “si es profesional u Ocupacional” y “si es conforme con el sueldo mínimo”; respecto a la primera opción, sigue una secuencia donde el usuario coloca 1100 soles y es AFP, en cuanto al segundo caso se tiene una condición que, si el tiempo de trabajo del usuario es constante, entonces va a contar con una cuenta de ahorro personal AFP, por último, también posee una condicional, en la cual, si el usuario está conforme con el sueldo mínimo, entonces se le asigna ONP, de ser el caso contrario se le otorga AFP, por último, en la categoría de Salud tiene cuatro opciones, primero, “si el usuario presenta una enfermedad”, debe detallarse, luego, “si este tiene personas que dependen de usted los cuales son su familia” entonces debe incluir la cantidad de personas en total, y los dos últimos puntos son referentes a “cuanto es su sueldo” y “cuanto es su manutención”, en donde si ambos son mayores a 1400, se le asigna AFP, caso contrario, se le asigna ONP.

4.1.2. Simulación y Análisis

4.1.2.1. Realizar simulaciones preliminares para evaluar la eficacia de la metodología con el modelo ML.

Figura 29

Código de entrenamiento del modelo



```

1  {
2    "training_examples": [
3      {
4        "nombre_apellido": "atoche castillo luis",
5        "edad": 36,
6        "sexo": "M",
7        "rango_edad_jubilarse": "60",
8        "sueldo": 863.82,
9        "aspiracion_ganar_mas": "Si",
10       "decision": 1
11     },
12     {
13       "nombre_apellido": "huaman moncada segundo horacio",
14       "edad": 36,
15       "sexo": "M",
16       "rango_edad_jubilarse": "60",
17       "sueldo": 1829.33,
18       "aspiracion_ganar_mas": "No",
19       "decision": 0
20     },
21     {
22       "nombre_apellido": "pek san contreras alfonso",
23       "edad": 36,
24       "sexo": "M",
25       "rango_edad_jubilarse": "60",
26       "sueldo": 526.22,
27       "aspiracion_ganar_mas": "Si"
28     }
29   ]
30 }

```

```

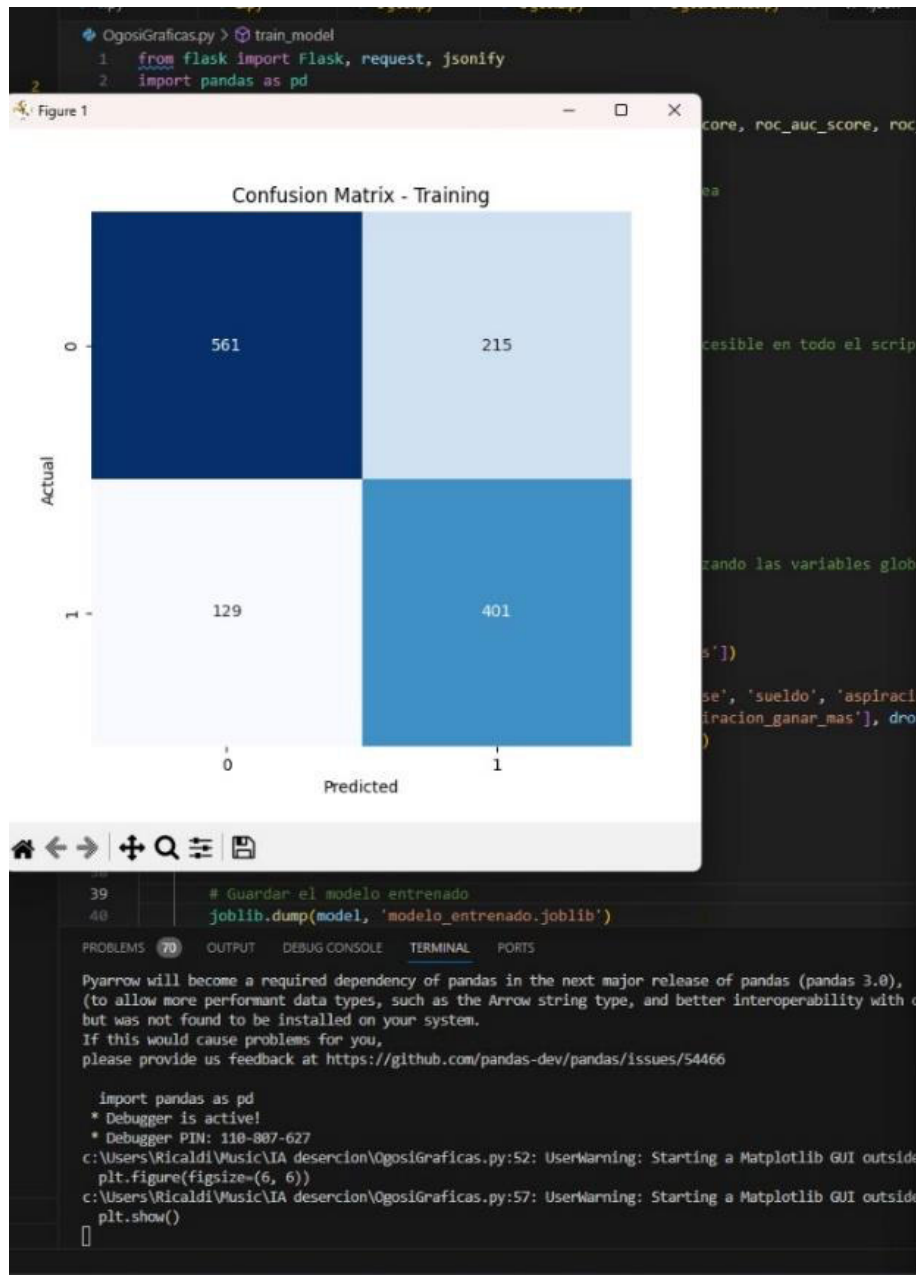
1  {
2    "message": "Modelo de árbol de decisión entrenado.",
3    "matriz": {
4      "f1_score": 0.69825479938192,
5      "precision": 0.85897482074826,
6      "recall": 0.754463773284967,
7      "roc_auc": 0.7397189509074812
8    }
9  }

```

Se muestra el código del entrenamiento del modelo, en el cual, al momento de su ejecución, se muestra los verdaderos positivos y negativos, así como los falsos positivos y negativos, siendo las respuestas que se obtendrán al momento de comenzar la simulación, dando como resultado la siguiente matriz.

Figura 30

Matriz de confusión del entrenamiento del modelo de Machine Learning



Se presentan dos valores de 1 y 0, donde 1 representa ONP y 0 representa AFP; en referencia a la asignación del seguro de pensiones, los cuales al usuario se le asigna ONP o AFP en donde se muestran los datos en escala actual, en el cual presenta una cantidad total de 129 falsos negativos y 215 falsos positivos, asimismo, presenta 561 verdaderos positivos y 401 verdaderos negativos al momento de entrenar el presente modelo.

La próxima secuencia de código, presentada en formato JSON, muestra cómo se crea la curva de falsos positivos en función a los datos proporcionados del modelo actual. Esta representación gráfica tiene como objetivo analizar la eficacia del modelo y determinar el valor de la Característica Operativa del Receptor (ROC). Una vez que el código se ejecuta, se genera el esquema mostrado en la Figura 35, ofreciendo una representación visual de la relación entre falsos positivos y verdaderos positivos durante la evaluación del modelo.

Figura 31

Esquema de ROC del modelo de Machine Learning



Después de la ejecución del código (utilizando los datos suministrados), se visualiza el diagrama de la curva ROC del modelo actual, el cual está siendo capacitado con la información respectiva. Esta representación gráfica exhibe la proyección de falsos positivos a lo largo de la curva, presentando debajo de la curva un área (AUC) equivalente a 0,74.

Figura 32

Métricas establecidas del modelo de Machine Learning

```
1  {  
2    "message": "Modelo de árbol de decisión entrenado exitosamente.",  
3    "metrics": {  
4      "f1_score": 0.699825479930192,  
5      "precision": 0.650974025974026,  
6      "recall": 0.7566037735849057,  
7      "roc_auc": 0.7397709589574013  
8    }  
9  }
```

Después de examinar las métricas durante la fase inicial de entrenamiento del modelo de árbol de decisión, se resalta que el F1-Score presenta un valor específico de 0.699825, indicando un equilibrio entre precisión y recall. La precisión individual del modelo se sitúa en 0.650974, destacando la habilidad para prever positivos con precisión. En cuanto al recall, se observa un valor de 0.756603, mostrando la proporción de instancias positivas identificadas correctamente por el modelo. Por último, la curva ROC exhibe un valor de 0.739770, proporcionando una medida integral de la capacidad de discriminación del modelo en diferentes puntos de corte. Estos los indicadores proporcionan una perspectiva o punto de vista íntegramente del rendimiento inicial del modelo de árbol de decisión.

4.1.2.2. Análisis de los resultados para el ajuste del diseño del modelo ML.

Inicialmente tuvo en cuenta como punto de gesta a la BD *DATA_SNP_2010_2022_data.xls*, información que fue acumulada y enviada por la Superintendencia de Banca y Seguros, dicha data tiene omisiones de algunos registros debido a que los sistemas transaccionales de la ONP y AFP podrían presentar incidencias en las migraciones de información, así como atenciones en horarios alternos de trabajo.

Figura 33

Data de la SNP

1	DETALLE_OCUPACION	U	V	W	X	Y	Z	AA	AB	AC
2	2021	BARBA	CALSINO	MARINO MARCELINO	13/02/1984	LIMA	BARANCOCHA	PARAMONGA	AV RAMON CASTILLA MZ LIT A URB 1 DE JUNIO	
2022	OBRERO	BARBA	SOLÍS	MANUEL LUIS	13/02/1984	LIMA	BARANCOCHA	PARAMONGA	AV RAMON CASTILLA MZ LIT A URB 1 DE JUNIO	
7025	CONTINUACION FACULTATIVA	BARBA	FAMREZ DE CALDERON	MARTHA CECILIA	29/07/1945	LIMA	LIMA	JR WASHINGTON N° 3645		
5870	OBRERO	BARBA	SANTAN	MANUEL	17/09/1940	LIMA	LIMA	JR LAS PALMARIAS 1014		
34320	OBRERO	BARBA	TRINIDAD	HENRO WALTER	1/08/1944	LIMA	BARANCOCHA	CALLE LIBERTAD N° 76 URB. MIGUEL GRAU		
34321	SERVICIOS DOMESTICOS	BARBA	USALCE	LAURO MARCELA	23/08/1971	LIMA	JESUS MARIA	EDIF. LOS HOGALES BLOCK 50 DPTO 101 RES. SAN FELIX		
3444	SERVICIOS DOMESTICOS	BARBA	TRINIDAD	JOSUE MANUEL	29/05/1940	LIMA	LIMA	VILLA MARIA DEL TRUJILLO		
34793	EMPLAZADO	BARBA	QUILCATE	SALU LEOPOLDO	14/03/1940	LAMBAYEQUE	CHICLAYO	CALLE TERESA JANNINI N° 654 P. JOSÉ OLAYTA		
34827	EMPLAZADO	BARBA	HEREDIA	ZOLA JUANITA	31/12/1947	LIMA	MULILRA	URB HUACHO MZ C LT 18		
44838	EMPLAZADO	BARBA	HEREDIA	ZOLA JUANITA	31/12/1947	LIMA	MULILRA	URB HUACHO MZ C LT 18		
44839	EMPLAZADO	BARBA	HEREDIA	ZOLA JUANITA	31/12/1947	LIMA	MULILRA	URB HUACHO MZ C LT 18		
50478	FACULTATIVO INDEPENDIENTE	BARBA	FRUENDENCO	EULADIA ANITA	30/12/1950	LA LIBERTAD	TRUJILLO	CLAS MORENAS N° 447 DPTO. 301-B URB. CALIFORNIA		
58602	OBRERO	BARBA	COSTILLA	ORIBEN SEBASTIAN	20/01/1994	LIMA	HERTA	CALLE GENERAL FRANCISCO VIGAL N° 990		
70321	EMPLAZADO	BARBA	SOLÍS	JESUS MARCELO	20/05/1955	LIMA	LIMA	AV SANTA MZ LIT 1		
76248	EMPLAZADO	BARBA	GUANILLO	MARTHA ISABEL	14/02/1944	LAMBAYEQUE	CHICLAYO	CALLEON BARANDARANAN 185 INT 202 URB. STA VICTORIA		
89977	EMPLAZADO	BARBA	DE BELLIS	MARIA YAREL	24/05/1949	LIMA	BARANCOCHA	CALLE BOLIVAR 495		
93363	FACULTATIVO INDEPENDIENTE	BARBA	MARCOS	SABINO	11/20/1944	LIMA	COMAS	JR ALEJANDRO HIGUERA N° 341 AÑO NUEVO		
97862	EMPLAZADO	BARBA	DE CARDOZAS	ROSALBA	24/12/1944	LAMBAYEQUE	CHICLAYO	CALLE LOS OLIVOS N° 299 URB.ERENACION FERRON V		
103933	OTROS	BARBA	MISTA DE JOD	MARIA ELFINA	17/01/1958	LIMA	LA VICTORIA	JIRON LOS RUISES N° 972 INT. D. URBANIZACION		
110776	EMPLAZADO	BARBA	OLAYA	BERTHA NOHI	23/12/1961	PUNTA	SULLANA	CALLE SAN JUAN BOSCO N° 208 URB. SANTA ROSA		
114668	EMPLAZADO	BARBA	QUISPE	EUSEBIA	8/02/1951	LIMA	LIMA	AV PINA DEL 900 2006		
115132	OBRERO	BARBA	RAYON	MERARDO OCTAVIO	8/06/1949	LIMA	LIMA	SAN MARTIN DE PORRES		
120570	OBRERO	BARBA	SANCHEZ	SEBASTIAN ANTONIO	20/05/1960	TIARRES	ZARUMILLA	ADOLFO VERDES		
134300	EMPLAZADO	BARBA	CASOLA	OLGA HORTENCIA	20/06/1951	LIMA	LIMA	PARQUE LA PUENTE Y CORTEZ 137		
138811	EMPLAZADO	BARBA	DE PALMIRINO	MARCELO LUDGANDA	28/06/1951	LIMA	BARANCOCHA	CALLE INDEPENDENCIA N° 206 INT. DPTO. 803		
146119	FACULTATIVO INDEPENDIENTE	BARBA	PIERRES DE RODAS	MARIA	28/07/1947	LIMA	LIMA	VILLA MARIA DEL TRUJILLO		
154222	EMPLAZADO	BARBA	MISTA	JOSE ESPERIDIO	27/12/1951	LAMBAYEQUE	FERRERÑE	CALLE 100 N° 750		
157752	EMPLAZADO	BARBA	CORRALDO	JOSE ALFONSO	14/02/1949	LIMA	LIMA	JR CONDELO BORJA 303 INT. 4		
169338	OTROS SECTORES	BARBA	SOTO	WALFEMAR FAIDO	28/06/1954	LIMA	ATE	ASOC. HUCOS DE ARIPIMAC MZ C LT 17 ETAPA URB SAN		
171729	EMPLAZADO	BARBA	MALAMANI	MARIA AURORA	7/02/1948	LAMBAYEQUE	FERRERÑE	CALLE TRES MARQUES 407		
217059	OTROS SECTORES	BARBA	RIOS	CEGAR ALVARDO	18/08/1953	CAMARCA	SANTA CRUZ	SECTOR EL SEG MZ. LT. 23 - ESPINAL CARRETERA A		
175740	SERVICIOS DOMESTICOS	BARBA	MALAMANI	DOLORES EMPERATRIZ	6/04/1954	LA LIBERTAD	TRUJILLO	JR CHICO DE ALMORAN N° 284		
219705	OTROS SECTORES	BARBA	BLATTES	HUGO ALBERTO	21/05/1951	LA LIBERTAD	CHICLAYO	CALLE TRUJILLO 376		
248990	OTROS SECTORES	BARBA	OTANGOPUN	TORRES LUIS	13/07/1950	LIMA	LIMA	AV ELMER FAUSETT 371 BLOCK A-301		
249353	EMPLAZADO	BARBA	SANCHEZ	GABRIEL MIGUEL	1/09/1950	LIMA	LIMA	JR HERRERA N° 1028 DPTO. 3-B		
260626	EMPLAZADO	BARBA	MARCIA	JOSE MANUEL	11/07/1951	LAMBAYEQUE	CHICLAYO	JOSÉ EDUARDO ORTIZ		
261480	CENTRO DE PROD. MINERA, METALURGICA Y SIDERURGICA	BARBA	ASCANCA	LUIS	21/06/1954	LIMA	LIMA	JR HUALLAGA 180 INT 317 TERCEP PISO		
211683	CENTRO DE PROD. MINERA, METALURGICA Y SIDERURGICA	BARBA	ASCANCA	LUIS	21/06/1954	LIMA	LIMA	JR HUALLAGA 180 INT 317 TERCEP PISO		
234290	OBRERO	BARBA	MEMONZA	JUAN LEOPOLDO	18/09/1958	PROV CONST CALLAO	CALLAO	JR ARQUITA NORTE 822		

Para el presente ajuste se ha tomado como referencia el caso de BARBA CALSINO MARINO MARCELINO el cual se cuenta con la siguiente información en código JSON.

Figura 34

Código del usuario seleccionado para el modelo

```

POST http://127.0.0.1:5000/api/prediccion

{
  "perfil": {
    "nombre_apellido": "barba calsino marino marcelino",
    "edad": 38,
    "sexo": "M",
    "tempo_estad_subjarse": "68",
    "sueldo": 1236.64,
    "aspiracion_ganar_mas": "si"
  }
}

{
  "decision": "Predicci3n para barba calsino marino marcelino: Se elige AFP."
}
    
```

El código hace referencia al usuario seleccionado para el entrenamiento de dicho modelo donde se ingresan los respectivos datos requeridos para la elección del seguro pensionario, donde al final se muestran los resultados de que al jubilado se le asigna una AFP como sistema de pensiones.

Figura 35

Datos del perfil del jubilado

```
{
  "perfiles": {
    "nombre_apellido": "barba calsino marino marcelino",
    "edad": 20,
    "sexo": "M",
    "rango_edad_jubilarse": "60",
    "sueldo": 1236.54,
    "aspiracion_ganar_mas": "si"
  }
}
```

Se muestra los datos extraídos de la tabla del Excel convertidos en formato JSON donde solo se toma en cuenta la información de entrada para la evaluación del modelo Machine Learning adaptado con la metodología CSKT, se ha tomado en cuenta el nombre y apellido concatenado (por un tema de orden en la información), el campo edad, sexo y sueldo; es de vital importancia los datos tomados en cuenta para la lógica del árbol de decisión. El rango de edad de jubilación es un campo tomado en cuenta de la base de datos y que hace referencia estrictamente a la ley de SNP. El campo de “aspiración a ganar más” fue tomando en cuenta con el fin de considerar la percepción de crecimiento profesional y económico de cada trabajador o persona que va a iniciar su vida laboral.

Figura 36

Resultado de modelo de predicción

```

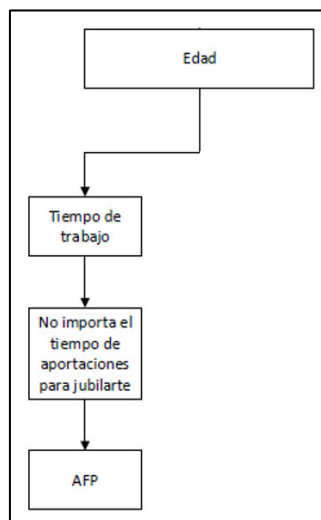
Body Cookies Headers (5) Test Results
Pretty Raw Preview Visualize JSON
1 {
2   "decision": "Predicción para barba calsino marino marcelino: Se elige AFP."
3 }

```

A continuación, el resultado del caso señalado se dedujo que le conviene AFP, esto debido a que entró por el criterio de edad, tiempo de trabajo ingresado y el salario mostrado; esto se corrobora más por medio del árbol de decisión en donde se tomó la siguiente secuencia en base a los criterios y la data ingresada para el presente usuario.

Figura 37

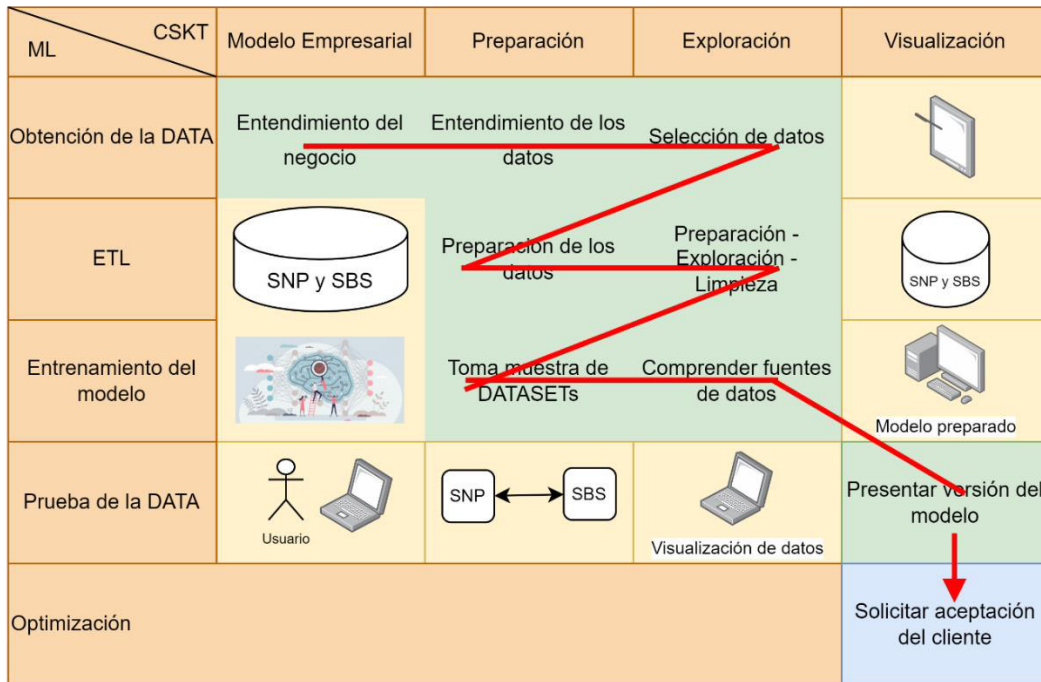
Secuencia del árbol de decisión del jubilado



El modelo sigue el camino de la metodología CSKT logrando la aceptación de la decisión evaluada por el caso del señor BARBA CALSINO MARINO MARCELINO donde primero se determinó la cantidad de tiempo de trabajo estimado, seguido de ello se procede a lo siguiente; que es independiente al tiempo de aportaciones, para que el señor pueda jubilarse dando como producto la asignación de una AFP.

Figura 38

Secuencia de la metodología



En base a la secuencia mostrada, la metodología sigue una secuencia detallada la cual comienza con la comprensión del negocio, donde se busca obtener un conocimiento profundo de las necesidades y objetivos de la entidad. Luego, se dirige hacia la comprensión de los datos, recopilando información relevante. A continuación, se lleva a cabo una cuidadosa selección de una proporción específica de la totalidad de los datos, seguida de la preparación de los mismos. Durante la etapa de preparación, se aplican filtros para descartar posibles puntos de datos incompletos o repeticiones de usuarios.

Después de la fase de filtrado, se presenta la muestra ya depurada para una revisión adicional, si resulta el caso, se lleva a cabo una nueva selección de muestras con el fin de asegurar la excelencia de la calidad de la data. Una vez que la información está lista, se somete al modelo de Machine Learning para su evaluación; los resultados obtenidos se utilizan para presentar la primera versión del modelo al cliente. En esta etapa, se busca la aprobación del cliente con respecto al modelo desarrollado. Este enfoque secuencial asegura una implementación efectiva y una colaboración continua con el cliente a lo largo del proceso.

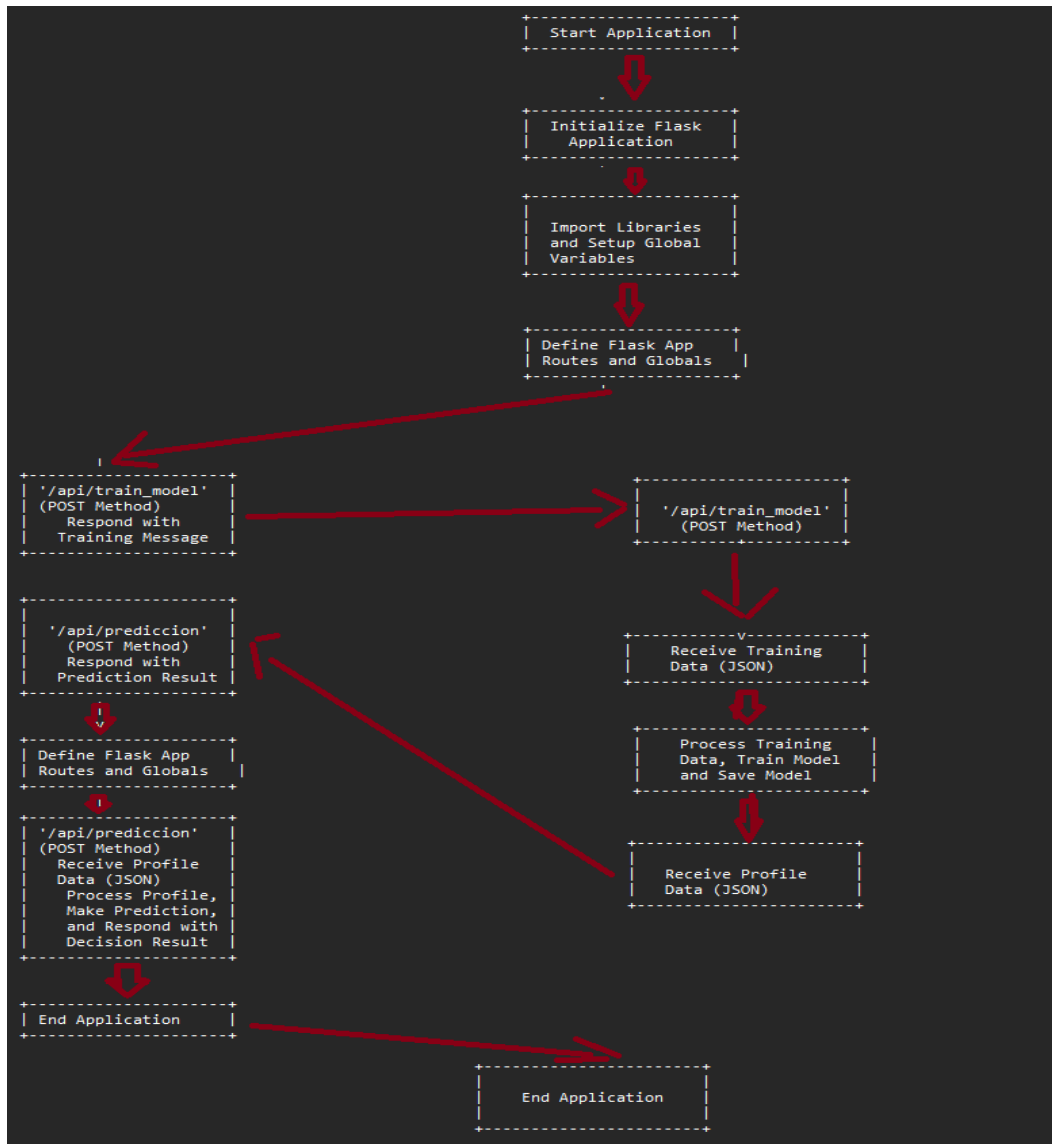
4.2. Desarrollo y análisis del prototipo de Machine Learning

4.2.1. Desarrollo del Prototipo

4.2.1.1. Implementar el diseño del modelo en un prototipo funcional.

Figura 39

Diseño del modelo funcional



Se muestra la secuencia de inicio del modelo de Machine Learning que va desde la ejecución de la aplicación, hasta el proceso que va desde la importación de librerías, las rutas que el modelo debe de seguir hasta la ejecución de los resultados que el modelo expresa mediante los datos que el usuario ingresa culminando con el cierre del aplicativo.

4.2.1.2. Programar los algoritmos de Machine Learning con la metodología CSKT.

Figura 40

Algoritmo de entrenamiento de Machine Learning - Script en Python

```

from flask import Flask, request, jsonify
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import precision_score, recall_score, f1_score, roc_auc_score
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
import matplotlib.pyplot as plt
import joblib

app = Flask(__name__)

# Definir X_train y X_val_scaled a nivel global para que sea accesible en todo el script
X_train = None
y_train = None
model = None

# Crear un clasificador de árboles de decisión
model = DecisionTreeClassifier(random_state=42)

@app.route('/api/train_model', methods=['POST'])
def train_model():
    global X_train, y_train, model # Indicar que estamos utilizando las variables globales
    try:
        # Resto del código de entrenamiento aquí...
        training_data = request.json
        df_train = pd.DataFrame(training_data['training_examples'])

        X_train = df_train[['edad', 'sexo', 'rango_edad_jubilarse', 'sueldo', 'aspiracion_ganar_mas']]
        X_train = pd.get_dummies(X_train, columns=['sexo', 'aspiracion_ganar_mas'], drop_first=True)
        X_train_scaled = StandardScaler().fit_transform(X_train)

```

Este script en Python está diseñado para implementar la metodología CSKT (Capture, Select, Kreate, Train), específicamente un clasificador de árboles de decisión. La metodología CSKT se centra en capturar datos, seleccionar características relevantes, crear el modelo y, finalmente, entrenarlo.

Figura 41

Algoritmo de entrenamiento de Machine Learning - Importación de librerías

```

y_train = df_train["decision"].astype(float)

# Entrenar el modelo de árbol de decisión
model.fit(X_train_scaled, y_train)

# Guardar el modelo entrenado
joblib.dump(model, 'modelo_entrenado.joblib')

# Calcular métricas después del entrenamiento
y_pred_train = model.predict(X_train_scaled)
precision_train = precision_score(y_train, y_pred_train)
recall_train = recall_score(y_train, y_pred_train)
f1_train = f1_score(y_train, y_pred_train)
roc_auc_train = roc_auc_score(y_train, y_pred_train)

# Devolver métricas como parte de la respuesta JSON
metrics_response = {
    'precision': precision_train,
    'recall': recall_train,
    'f1_score': f1_train,
    'roc_auc': roc_auc_train
}

return jsonify({'message': 'Modelo de árbol de decisión entrenado exitosamente.', 'metrics': metrics_response})

except Exception as e:
    return jsonify({'error': str(e)}), 400

@app.route('/api/prediccion', methods=['POST'])
def predecir_perfil():

```

El código comienza importando las bibliotecas necesarias, como Flask para la creación de una API web, Pandas para la manipulación de datos y Scikit-learn para el modelo de árbol de decisión y métricas de rendimiento.

Figura 42

Algoritmo de entrenamiento de Machine Learning - Definición de variables globales

```

@app.route('/api/prediccion', methods=['POST'])
def predecir_perfil():
    try:
        # Recibir datos en formato JSON desde la aplicación
        data = request.json

        # Acceder al perfil directamente
        perfil = data['perfiles']

        # Crear un DataFrame a partir de los datos recibidos
        df = pd.DataFrame([perfil])

        # Asegurarse de tener todas las columnas necesarias
        X = df[['edad', 'sexo', 'rango_edad_jubilarse', 'sueldo', 'aspiracion_ganar_mas']]

        # Asegurarse de que X tenga las mismas columnas que se utilizaron durante el entrenamiento
        X = X.reindex(columns=X_train.columns, fill_value=0)

        # Transformar las características usando el StandardScaler ajustado durante el entrenamiento
        X_scaled = StandardScaler().fit(X_train).transform(X)

        y_pred = model.predict(X_scaled)

        # Tomar decisión ONP o AFP basada en la salida del modelo
        decision = "Predicción para {}:".format(perfil['nombre_apellido'])
        decision += 'Se elige AFP.' if y_pred[0] > 0.5 else 'Se elige ONP.'

        # Responder con la decisión
        return jsonify({'decision': decision})

    except Exception as e:

```

Se definen variables globales para almacenar datos de entrenamiento (X_{train} , y_{train}) y el modelo entrenado. Luego, se crea un clasificador de árboles de decisión. El código incluye dos endpoints principales. El primero ('/api/train_model') se utiliza para recibir datos de

entrenamiento, procesarlos, escalar las características y entrenar el modelo. También se almacenan métricas de rendimiento y se guarda el modelo entrenado en un archivo.

Figura 43

Algoritmo de entrenamiento de Machine Learning - Predicciones basadas en el modelo entrenado

```

data = request.json

# Acceder al perfil directamente
perfil = data['perfiles']

# Crear un DataFrame a partir de los datos recibidos
df = pd.DataFrame([perfil])

# Asegurarse de tener todas las columnas necesarias
X = df[['edad', 'sexo', 'rango_edad_jubilarse', 'sueldo', 'aspiracion_ganar_mas']]

# Asegurarse de que X tenga las mismas columnas que se utilizaron durante el entrenamiento
X = X.reindex(columns=X_train.columns, fill_value=0)

# Transformar las características usando el StandardScaler ajustado durante el entrenamiento
X_scaled = StandardScaler().fit(X_train).transform(X)

y_pred = model.predict(X_scaled)

# Tomar decisión ONP o AFP basada en la salida del modelo
decision = "Predicción para {}: ".format(perfil['nombre_apellido'])
decision += 'Se elige AFP.' if y_pred[0] > 0.5 else 'Se elige ONP.'

# Responder con la decisión
return jsonify({'decision': decision})

except Exception as e:
    return jsonify({'error': str(e)}), 400

if __name__ == '__main__':
    app.run(debug=True)

```

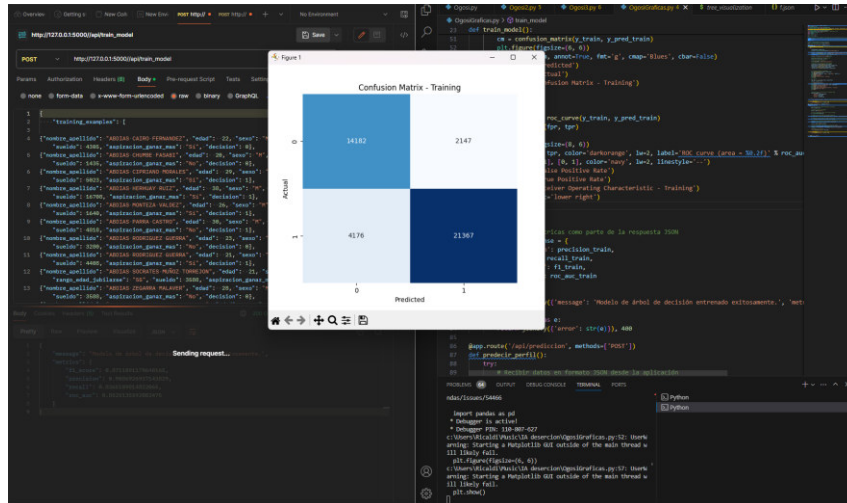
El segundo endpoint ('/api/prediccion') realiza predicciones basadas en el modelo entrenado. Procesa los datos del perfil, realiza las transformaciones necesarias y hace la predicción. La decisión final (AFP o ONP) se determina según la salida del modelo. La aplicación Flask se inicia en modo de depuración para permitir la ejecución y prueba del código. Este código Python implementa la metodología CSKT para entrenar un clasificador de árboles de decisión a través de una API web, permitiendo tanto el entrenamiento del modelo como la realización de predicciones basadas en perfiles específicos.

4.2.2. Pruebas de Prototipo

4.2.2.1. Realizar pruebas del prototipo en escenarios controlados.

Figura 44

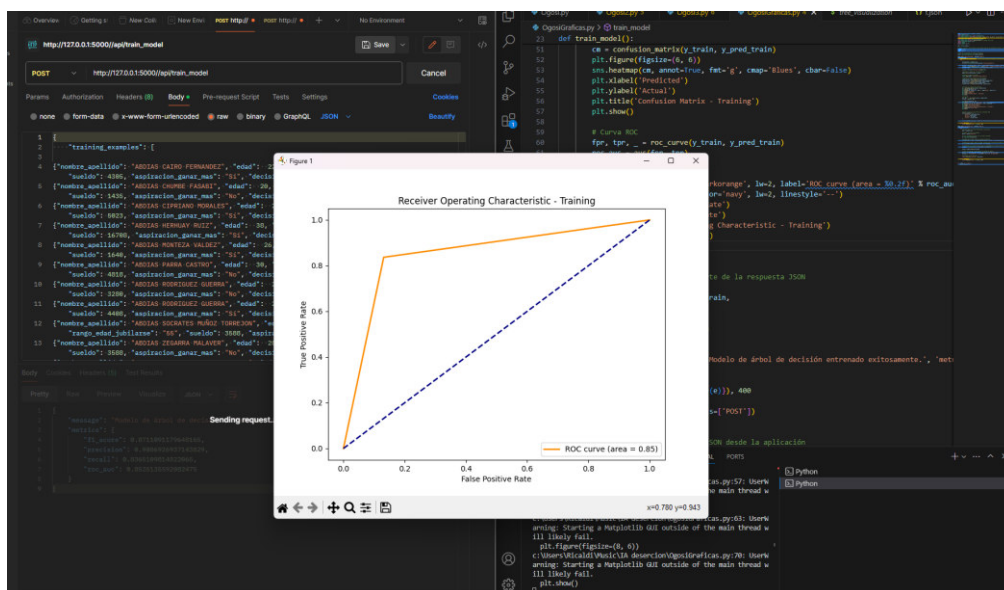
Modelo entrenado con la data Incrementada



En base a la matriz de confusión del entrenamiento del modelo de Machine Learning adaptado con la data de la SBS incrementada, se presentan dos valores de 1 y 0; donde 1 representa a ONP y 0 representa a AFP, en referencia a la asignación del seguro de pensiones los cuales al usuario se le asigna ONP o AFP mostrando los datos en escala actual que presenta una cantidad total de 4176 como falsos negativos y 2147 como falsos positivos, mientras que en los verdaderos positivos presenta una cantidad de 14182 y en los verdaderos negativos presenta una cantidad de 21367 al momento de entrenar el presente modelo.

Figura 45

Curva ROC enfocada a la data incrementada



Tras la ejecución del código utilizando los datos suministrados para el modelo adaptado con la data de la SBS incrementada, se visualiza el diagrama de la curva ROC del modelo actual, el cual está siendo capacitado con la mencionada información. Esta representación gráfica exhibe la proyección de falsos positivos a lo largo de la curva, presentando debajo de la curva un área (AUC) equivalente de 0,85.

Figura 46

Resultados de métricas enfocada a la data incrementada

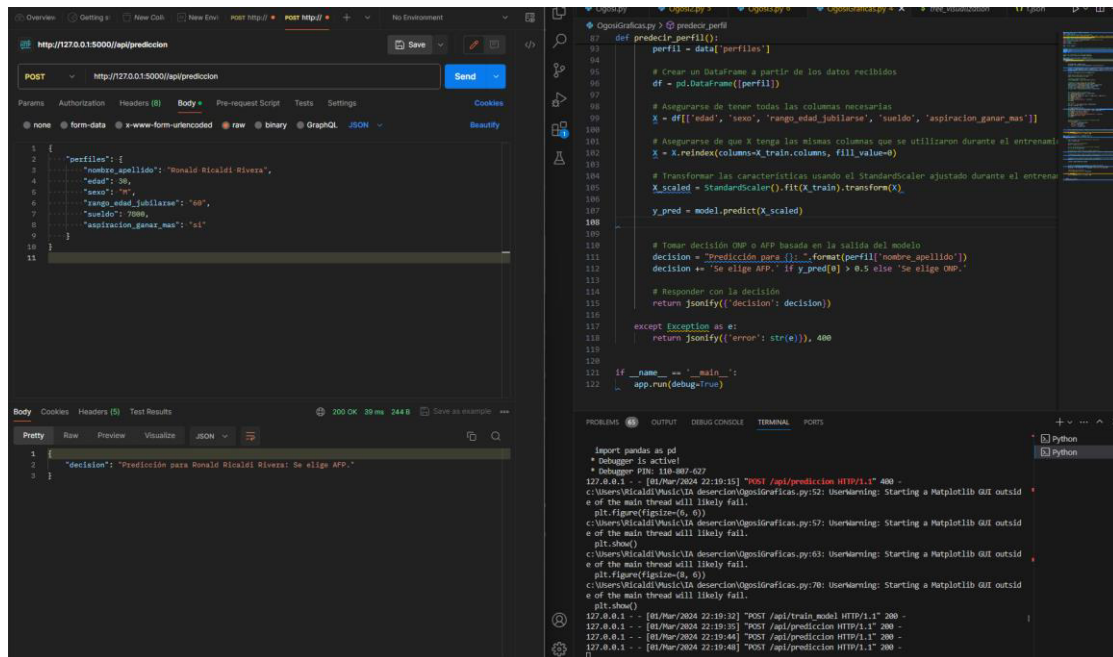
```

1 POST http://127.0.0.1:5000/api/train_model
2
3 "training_examples": [
4 ["nombre_apellido": "ABDIAS CAIRO FERNANDEZ", "edad": 22, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 4366, "aspiracion_ganar_mas": "Si", "decision": 0},
5 ["nombre_apellido": "ABDIAS CHURRI FASAZ", "edad": 20, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 1435, "aspiracion_ganar_mas": "No", "decision": 0},
6 ["nombre_apellido": "ABDIAS CIRIACO ROMALES", "edad": 29, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 5923, "aspiracion_ganar_mas": "Si", "decision": 1},
7 ["nombre_apellido": "ABDIAS HERIBAY RIZZO", "edad": 38, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 10700, "aspiracion_ganar_mas": "Si", "decision": 1},
8 ["nombre_apellido": "ABDIAS MONTEA VALDEZ", "edad": 26, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 1648, "aspiracion_ganar_mas": "Si", "decision": 1},
9 ["nombre_apellido": "ABDIAS PARRA CASTRO", "edad": 38, "sexo": "M", "rango_edad_jubilarse": "65", "sueldo": 4318, "aspiracion_ganar_mas": "No", "decision": 1},
10 ["nombre_apellido": "ABDIAS RODRIGUEZ QUEROA", "edad": 23, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 3288, "aspiracion_ganar_mas": "No", "decision": 0},
11 ["nombre_apellido": "ABDIAS RODRIGUEZ QUEROA", "edad": 24, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 4468, "aspiracion_ganar_mas": "Si", "decision": 1},
12 ["nombre_apellido": "ABDIAS SOCRATES NUÑO TORIBIO", "edad": 21, "sexo": "M", "rango_edad_jubilarse": "65", "sueldo": 3585, "aspiracion_ganar_mas": "No", "decision": 0},
13 ["nombre_apellido": "ABDIAS ZEGARRA MALAVEY", "edad": 23, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 3588, "aspiracion_ganar_mas": "No", "decision": 0},
14 ["nombre_apellido": "ABDITO LEONIDAS GUERRON SALAS", "edad": 37, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 1438, "aspiracion_ganar_mas": "Si", "decision": 1},
15 ["nombre_apellido": "ABDON ALEJANDRO ACEVEDO FLORES", "edad": 21, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 2889, "aspiracion_ganar_mas": "Si", "decision": 1},
16 ["nombre_apellido": "ABDON ALEJANDRO ACEVEDO FLORES", "edad": 32, "sexo": "M", "rango_edad_jubilarse": "65", "sueldo": 16893, "aspiracion_ganar_mas": "No", "decision": 0},
17 ["nombre_apellido": "ABDON ALEJANDRO BALLEARES ACHICARRA", "edad": 36, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 9636, "aspiracion_ganar_mas": "Si", "decision": 1},
18 ["nombre_apellido": "ABDON ALFONSO GALENES PINTO", "edad": 21, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 3895, "aspiracion_ganar_mas": "No", "decision": 1},
19 ["nombre_apellido": "ABDON ASCENCIO QUENEA", "edad": 29, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 3383, "aspiracion_ganar_mas": "Si", "decision": 0},
20 ["nombre_apellido": "ABDON AZAÑA GARBILLO", "edad": 23, "sexo": "M", "rango_edad_jubilarse": "65", "sueldo": 3383, "aspiracion_ganar_mas": "Si", "decision": 1},
21 ["nombre_apellido": "ABDON CABELLO EMPEROR", "edad": 22, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 3998, "aspiracion_ganar_mas": "No", "decision": 1},
22 ["nombre_apellido": "ABDON CALLE HUACCA", "edad": 37, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 1013, "aspiracion_ganar_mas": "No", "decision": 1},
23 ["nombre_apellido": "ABDON CANDIOTTI SARMIENTO", "edad": 20, "sexo": "M", "rango_edad_jubilarse": "60", "sueldo": 1546, "aspiracion_ganar_mas": "Si", "decision": 1},
24 ]
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329

```

Figura 47

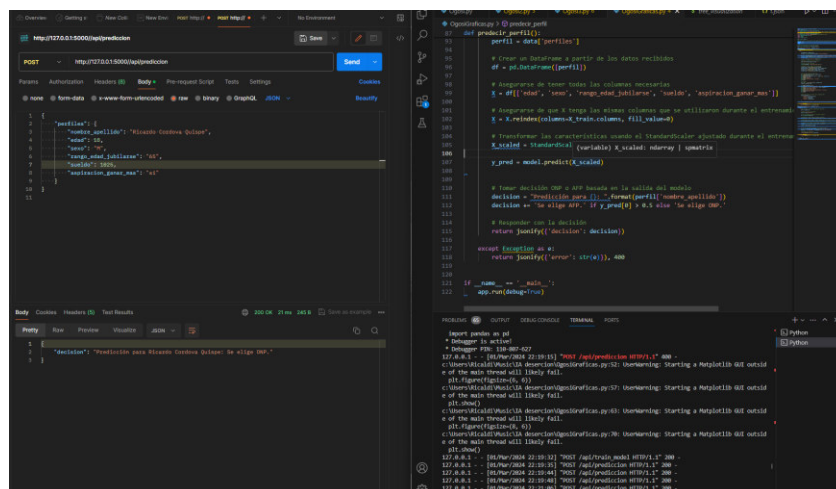
Prueba de los primeros datos del modelo



Como pruebas de las predicciones del modelo ML, se puso a prueba con el registro de uno de los trabajadores como primer dato, el cual se pudo obtener que marc3 como correcto la selecci3n del AFP.

Figura 48

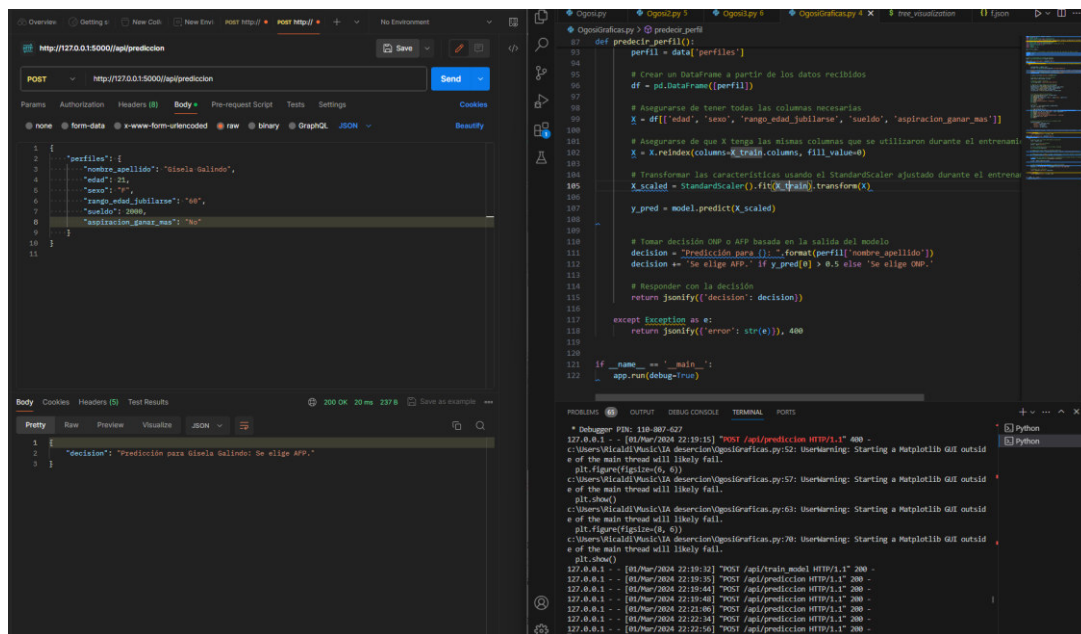
Segundo caso para el entrenamiento del modelo



Respecto a este segundo caso, se tiene en cuenta los datos de un joven de 18 años que tiene el puesto de Operario en donde, mediante sus datos ingresados en el modelo, se pudo obtener como resultado que eligió ONP en base a los mismos campos ingresados por el primer caso.

Figura 49

Tercer caso para el entrenamiento del modelo

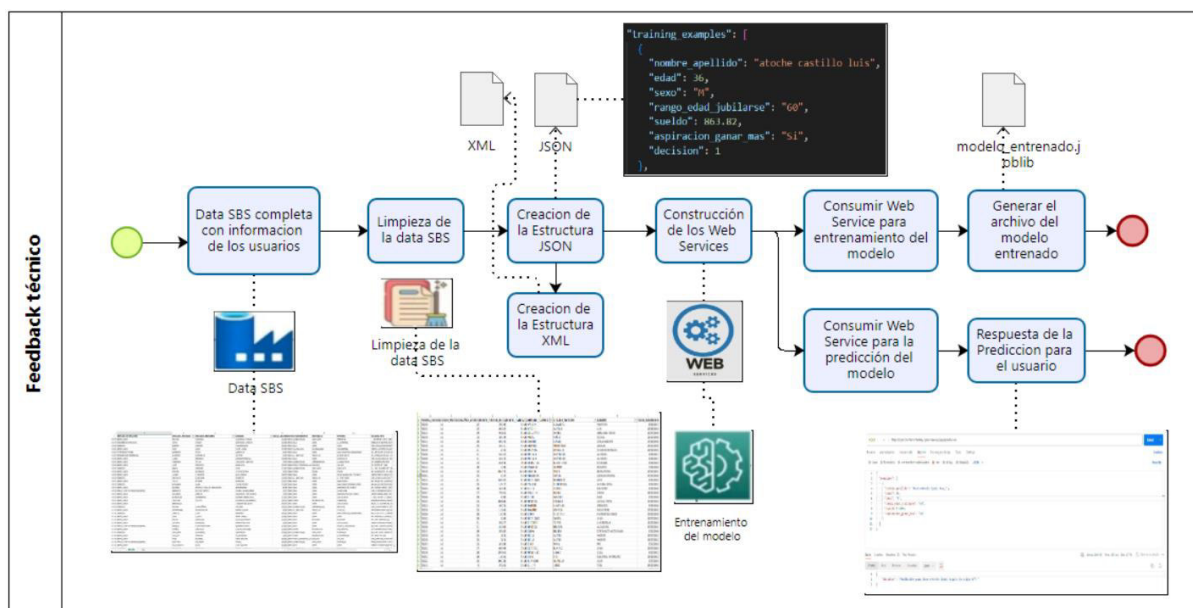


Respecto al tercer caso, se tiene los datos de una programadora Jr., quien, al momento de elegir las opciones dadas mediante sus datos ingresados, el modelo mediante dicha data ingresada fue quien predijo que su mejor opci3n para ella era AFP.

4.2.2.2. Recoger feedback y realizar ajustes t3cnicos.

Figura 50

Feedback t3cnico



En base al feedback técnico mostrado en la figura 50, se determina que se debe realizar unos ajustes propios según lo definido para el entrenamiento del modelo basado en ML, para este caso la data de la SBS debe seguirse limpiando antes de pasar al escenario de producción, debido a posibles errores que se presenten durante el procesamiento de las épocas. Otro ajuste a realizarse es el consumo de Web Services, y esto se debe al incremento de la data en formato JSON el cual va a consistir en aumentar la información para tener una mayor precisión en el modelo entrenado.

Tabla 11

Comparativa entre resultados del modelo adaptado

Métrica	Valor alcanzado en la adaptación del modelo con la metodología CSKT	Valor alcanzado del modelo adaptado con la data de la SBS incrementada
Precisión	0.69982547	0.90869269
Recall	0.65097402	0.83651098
F1 – Score	0.75660377	0.87110911
AUC - ROC	0.73977095	0.85251355

4.2.3. Análisis de Resultados

4.2.3.1. Evaluar el desempeño del prototipo según los criterios de aceptación definidos.

Tabla 12

Criterios de Aceptación de las métricas

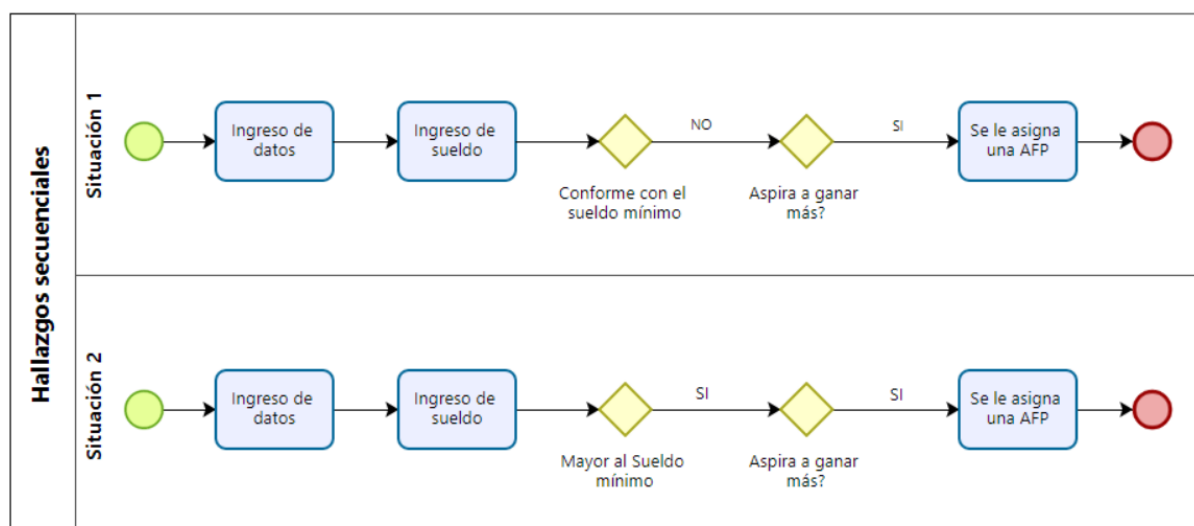
Métrica	Rango de Aceptación	Comentarios
Precisión	Mayor al 0.7 o 0.8	Busca minimizar los falsos positivos.
Recall	Mayor al 0.7 o 0.8	Busca minimizar los falsos negativos.
F1 – Score	Mayor al 0.7 o 0.8	Media armónica de precisión y recall.
AUC - ROC	Mayor al 0.7 o 0.8	Busca distinguir clases, un valor de 0.5 señala aleatorias predicciones.

4.2.3.2. Documentar los hallazgos y las áreas de mejora.

En base a las pruebas realizadas en el prototipo se tomaron en cuenta la secuencia de dos casos de trabajadores que cuentan con datos sumamente distintos por lo que se tiene el siguiente esquema.

Figura 51

Flujo de la secuencia de casos del modelo de Machine Learning



En base a los casos explicados anteriormente, se plantaron esta secuencia en base a las decisiones tomadas al momento de ingresar sus datos, en el cual uno de los casos indica que su sueldo es menor al mínimo pero que el aspira a ganar una mayor capital, mientras que el otro caso indica que su sueldo si es mayor al sueldo mínimo pero este desea ganar una mayor capital dando como resultado para ambos casos puestos en el modelo que se les asigna AFP, por ende, se da a entender que sin importar el monto que ganan, si estos desean ganar más aportes, entonces el modelo le asigna AFP en base a los datos obtenidos de dichos registros.

4.3. Medición de la precisión del modelo de Machine Learning

4.3.1. Definición de las Métricas de Precisión

4.3.1.1. Establecer métricas claras para evaluar la precisión del modelo.

Tabla 13

Métricas de evaluación del modelo

Métrica	Descripción
F1 – Score	En relación con este indicador específico, se incorporan las métricas anteriores, precisión y recall, para brindar una evaluación más completa de la capacidad del modelo para prever de manera eficiente los resultados positivos, en contraste con la métrica de Accuracy (Borja et al., 2020). El F1-Score se formula considerando tanto la precisión como el recall, buscando un equilibrio entre la capacidad de identificar correctamente los resultados positivos y minimizar los falsos positivos y negativos. Esta métrica proporciona una medida más equitativa del rendimiento del modelo en comparación con la precisión o el recall por separado.
Precisión	En lo que respecta a este indicador en particular, se hace referencia al porcentaje de valores verdaderos positivos que fueron correctamente anticipados por el modelo (Pérez, 2021). En otras palabras, la precisión de este indicador refleja la proporción de predicciones acertadas en relación con los casos reales positivos, proporcionando una medida de la exactitud del modelo en identificar correctamente los resultados positivos.

Recall	<p>En el contexto de este indicador específico para el modelo de machine learning, se incluyen todas las predicciones positivas que fueron acertadas (Pérez, 2021). Esto implica que el indicador de recall evalúa si el modelo es capaz de identificar y/o capturar de manera correcta la totalidad de los casos positivos existentes en el conjunto de datos, proporcionando así una medida de la exhaustividad y eficacia del modelo en la detección de resultados positivos.</p>
ROC - AUC	<p>De acuerdo con Arias et al. (2020), la Curva ROC se convierte en un recurso esencial al momento de validar la capacidad predictiva de modelos, tales como los Grupos de Morbilidad Ajustada (GMA), al ser contrastados con otras herramientas destinadas a la estratificación de la población. Esta representación gráfica no solo ofrece una visión global del desempeño del modelo, sino que también suministra detalles específicos sobre su rendimiento en distintos contextos. En consecuencia, la Curva ROC desempeña un papel crucial al evaluar la eficacia predictiva del modelo en diversos escenarios, contribuyendo significativamente al proceso de evaluación de su desempeño general.</p>

4.3.1.2. Determinar los criterios de éxito.

Tabla 14

Criterios de éxito

Métricas Utilizadas	Criterios de Aceptación	Valor Alcanzado	Criterios de Éxito
F1 – Score	Mayor al 0.7 o 0.8	0.87110911	SI
Precisión	Mayor al 0.7 o 0.8	0.90869269	SI
Recall	Mayor al 0.7 o 0.8	0.83651098	SI
ROC - AUC	Mayor al 0.7 o 0.8	0.85251355	SI

Nota. En base a la tabla de criterios mostrado, se logró determinar que para el caso del F1 – Score presenta un valor de 0.87110911 siendo mayor al criterio de aceptación propuesto, luego,

para la métrica de precisión, esta presenta un valor de 0.90869269 lo cual se acepta referente a los criterios de éxito, para el caso del recall, presenta un valor de 0.83651098 lo cual determina como un criterio de éxito considerado luego de obtener los datos respectivos sobre el modelo ML, por último, dada la métrica de ROC se determinó que presenta el valor de 0.85251355 siendo mayor al criterio de aceptación lo cual se pudo demostrar que presenta buenos estándares el presente modelo.

4.3.2. Recolección y Preparación de Datos

4.3.2.1. Recopilar datos históricos de contratación para pruebas.

Figura 52

Recopilación de datos históricos de contratación

1	DETALLE_Ocupacion	APELLIDO_PATERNO	APELLIDO_MATERNO	NOMBRE	FECHA_NACIMIENTO	DEPARTAMENTO	PROVINCIA	DISTRITO	DI_BENE_TITU
2051	OBERO	BARBA	CALEMO	MARINO MARCELINO	12/02/1949	LIMA	BARRANCA	PARAMONGA	AV RAMON CASTILLA
2062	OBERO	BARBA	ROBLES	MANUEL LUJAN	18/02/1948	LIMA	BARRANCA	PARAMONGA	AV RAMON CASTILLA
7005	CONTINUACION FACULTATIVA	BARBA	RAMIREZ DE CALDERON	MARTHA CECILIA	23/07/1940	LIMA	LIMA	LIMA	JR. WASHINGTON N° 2
8087	OBERO	BARBA	IBALTAN	MANUEL	13/06/1947	LIMA	LIMA	LOS OLIVOS	AV LAS PALMERAS 340
8431	OBERO	BARBA	TRINIDAD	PEDRO WALTER	1/08/1944	LIMA	BARRANCA	PARAMONGA	CALLE LIBERTAD N° 70
8670	AMAS,SERV DOMESTICO	BARBA	UGALDE	LAURA MARCELA	18/06/1947	LIMA	LIMA	JESUS MARIA	EDIF. LOS NOGALES BL
8844	SERVICIOS DOMESTICOS	BARBA	TRINIDAD	JOSUE MANUEL	29/05/1940	LIMA	LIMA	VILLA MARIA DEL TRIUNFO	AV. PEDRO TEJADA N°
8970	EMPLEADO	BARBA	QUIRIGATE	SALVA LEOPOLDO	14/02/1945	LAMBAYEQUE	CHICLAYO	CHICLAYO	CALLE TERESA FRANK
8667	EMPLEADO	BARBA	HERRERA	ZOLA AURORA	3/11/1947	LIMA	HUACRA	HUACRA	URB HUACHO MZ. C.11
8688	EMPLEADO	BARBA	HERRERA	ZOLA AURORA	3/11/1947	LIMA	HUACRA	HUACRA	URB HUACHO MZ. C.11
8690	EMPLEADO	BARBA	HERRERA	ZOLA AURORA	3/11/1947	LIMA	HUACRA	HUACRA	URB HUACHO MZ. C.11
9047	FACULTATIVO INDEPENDIENTE	BARBA	PRUDENCIO	ZULIMA AMINA	10/12/1950	LA LIBERTAD	TRUJILLO	TRUJILLO	CL LAS MORENAS N° 4
8980	OBERO	BARBA	COSTILLA	ONTERE SEBASTIAN	20/01/1950	LIMA	BREÑA	BREÑA	CALLE GENERAL FRAN
7033	EMPLEADO	BARBA	GONZALES	JESUS SAMUEL	20/03/1955	LIMA	LIMA	SAN MARTIN DE PORRES	AV CANITA MCL 17 1
7644	EMPLEADO	BARBA	GUANILO	MARTHA YSABEL	16/02/1948	LAMBAYEQUE	CHICLAYO	CHICLAYO	CL LLON BARRANCA
8097	EMPLEADO	BARBA	DE URBILUS	MARIA YSABEL	24/03/1940	LIMA	BARRANCA	SUPE	CALLE BOLIVAR 499 -
9366	FACULTATIVO INDEPENDIENTE	BARBA	MARCOS	SABINO	12/10/1944	LIMA	LIMA	COMAS	JR. ALLANORO VILESI
9702	EMPLEADO	BARBA	DE CARDENAS	ROSA MARY	14/12/1948	LAMBAYEQUE	CHICLAYO	CHICLAYO	CALLE LOS OLIVOS N°
10530	OTROS	BARBA	MESTA DE JOO	MARIA ELENA	13/01/1958	LIMA	LIMA	LA VICTORIA	JIRON LOS RUBES N°
11078	EMPLEADO	BARBA	OLAYA	BERTHA NORI	23/12/1943	PUNO	SULLANA	SULLANA	CALLE SAN JUAN BOS
11468	EMPLEADO	BARBA	QUISPE	EVGENIA	8/02/1951	LIMA	LIMA	SAN MARTIN DE PORRES	AV. PINAR DEL RIO 16
11517	OBERO	BARBA	BAYLON	MERARDO OCTAVIO	8/06/1949	LIMA	LIMA	SAN MARTIN DE PORRES	JR. LAS TUNAS 171 UR
12207	OBERO	BARBA	SANCHEZ	SEGUNDO ANTONIO	29/05/1950	TUMBES	ZARUMILLA	AGUAS VERDES	JR. EL SALVADOR N°2
12400	EMPLEADO	BARBA	CACEDA	OLGA HOSTERIA	20/06/1951	LIMA	LIMA	PUBLICO LIBRE	PARRQUE LA PUERTI Y
13811	EMPLEADO	BARBA	DE PALOMINO	NAIDA LUZGARDA	28/06/1951	LIMA	LIMA	MIRAFLORES	CALLE INDEPENDENCI
14510	FACULTATIVO INDEPENDIENTE	BARBA	FLORES DE RODAS	MARIA	28/07/1947	LIMA	LIMA	VILLA MARIA DEL TRIUNFO	JR. ALBINO TORRES N°
14623	EMPLEADO	BARBA	MESTA	JOSE ESPERDINO	27/12/1951	LAMBAYEQUE	FERREÑAFE	FERREÑAFE	CALLE LIO N° 70
15735	EMPLEADO	BARBA	CABALLERO	JOSE ALFONSO	19/10/1949	LIMA	LIMA	LIMA	JR. CORNELIO BORDA
18938	OTROS SECTORES	BARBA	SOTO	WALDEMAR ITALO	28/06/1954	LIMA	LIMA	ATE	ASOC HIJOS DE APURI
17372	EMPLEADO	BARBA	MANQUE	MARIA AURORA	7/02/1946	LAMBAYEQUE	FERREÑAFE	FERREÑAFE	CALLE TRES MARIAS 4
17359	OTROS SECTORES	BARBA	RIOS	CESAR ALVARO	18/06/1953	CAJAMARCA	SANTA CRUZ	CATACHE	SECTOR EL SEIS MZ. A
17340	SERVICIOS DOMESTICOS	BARBA	HUAMAN	DOLORES EMPERARIZ	6/04/1954	LA LIBERTAD	TRUJILLO	TRUJILLO	JR. DIOSO DE ALMAG
17970	OTROS SECTORES	BARBA	ALBIRIN	LUIS ALBERTO	21/09/1952	LA LIBERTAD	CHEPEN	CHEPEN	CALLE TRIUNFO 376
18800	OTROS SECTORES	BARBA	OTANGUREN	TOSKA ELBA	11/07/1950	LIMA	LIMA	SAN MIGUEL	AV. ELMIR FAUCETT 31

En base al gráfico mostrado, se presenta la *database* que ha sido empleada para el proyecto, esta hoja de cálculo se aprecia en la nube por medio de este enlace de OneDrive:

<https://unfype->

my.sharepoint.com/:x:/g/personal/2021008037_unfy_edu_pe/EU4QEKLl1s1Pup3k9y2IdHQB

7ZIW9u4O0CNQNY8UZNVBMw?e=tEyEhh

4.3.2.2. Preparar y limpiar los datos para su uso en el modelo.

Figura 53.

Limpieza de la data histórica

ID	PRIMERA EMISION	ESTADO	PRESTACION AÑOS	APORTES MONTO	PENSION MONTO	HABILITADO	APELLIDO_PATERNO	APELLIDO_MATERNO	NOMBRE	FECHA_NACIMIENTO
2001	201011	AC	33	129.42	893.00	BARBA	CALDINO		MARINO MARCELINO	13/02/1949
2002	201011	AC	33	129.42	893.00	BARBA	ROBLES		MANUEL LUIS	13/02/1949
7005	201103	AC	20	229.34	500.00	BARBA	BAMREZ DE CALDERON		MARTHA CECILIA	23/07/1945
30870	201210	AC	35	352.70	500.00	BARBA	GAITAN		MANUEL	17/06/1947
14330	201303	AC	48	228.47	893.00	BARBA	TRINIDAD		PEDRO WALTER	3/02/1944
34707	201303	AC	23	300.90	500.00	BARBA	UGALDE		LAURA MARCELA	13/06/1947
34804	201308	AC	21	313.80	500.00	BARBA	TRINIDAD		JOSUE MANUEL	29/05/1940
30791	201307	AC	41	624.58	686.00	BARBA	QUILGATE		SALIL LEONOLDO	14/03/1943
46637	201403	AC	33	608.98	673.00	BARBA	HERRERA		ZOLA AURORA	3/11/1947
46638	201403	AC	33	608.98	673.00	BARBA	HERRERA		ZOLA AURORA	3/11/1947
46639	201403	AC	33	608.98	673.00	BARBA	HERRERA		ZOLA AURORA	3/11/1947
50478	201407	AC	25	334.85	500.00	BARBA	PRUDENCIO		ZULEMA AMINA	10/12/1950
69000	201509	AC	20	218.47	500.00	BARBA	COSTILLA		ONTERE SEBASTIAN	20/01/1936
70323	201807	AC	31	258.79	500.00	BARBA	GONZALES		JESUS SAMUEL	20/02/1955
76148	201508	AC	20	352.31	500.00	BARBA	GUANILO		MARTHA YSABEL	16/02/1949
80997	201510	AC	25	376.75	500.00	BARBA	DE URRILLUS		MARIA YSABEL	24/03/1949
93046	201606	AC	11	172.55	500.00	BARBA	MARCOS		SABINO	12/10/1944
97801	201407	AC	25	424.73	509.00	BARBA	DE CARDENAS		ROSA MARY	14/12/1948
105391	201705	AC	33	336.51	500.00	BARBA	MESTA DE JOD		MARIA ELENA	17/01/1956
110770	201807	AC	38	393.98	500.00	BARBA	OLAYA		BETHYAN NORI	23/12/1961
114664	201810	AC	29	370.86	500.00	BARBA	OLUSE		EUGENIA	8/02/1951
115117	201806	AC	42	172.27	893.00	BARBA	BAYLON		MERARDO OCTAVIO	8/06/1949
123574	201609	AC	32	563.72	632.00	BARBA	SANCHEZ		SEGUNDO ANTONIO	29/02/1950
134304	201712	AC	24	384.40	500.00	BARBA	CACEDA		OLGA HORTENSIA	20/06/1952
138111	201703	AC	35	541.80	613.00	BARBA	DE PALOMINO		NAIDA LUIGARDA	28/06/1951
140133	201302	AC	20	97.65	500.00	BARBA	FLORES DE ROBAS		MARIA	28/07/1947
154221	201303	AC	20	425.00	500.00	BARBA	MESTA		JOSE ESPOSORIO	27/12/1952
157720	202003	AC	29	1366.27	893.00	BARBA	CABALLERO		JOSE ALFONSO	19/10/1949
169919	201911	AC	39	640.64	772.00	BARBA	SOTO		BRADENBAR TIALDO	24/02/1954
171720	201808	AC	44	576.44	644.00	BARBA	MAIQUE		MARIA AURORA	7/02/1948
173509	201811	AC	35	841.13	878.00	BARBA	RIOS		CESAR ALADINO	18/06/1953
177948	202004	AC	12	172.22	500.00	BARBA	HUAMAN		DOLORES EMPERATRIZ	6/04/1954
179750	201808	AC	36	696.17	769.00	BARBA	ALBITRES		LUIS ALBERTO	21/09/1952
188906	202010	DATA	44	303.52	500.00	BARBA	OVANGUREN		TOSSA ELBA	13/07/1950

Tal como se muestra en esta figura, se procede a realizar la limpieza de datos, es decir, se retira las columnas que no presentan relevancia y solo se considera los datos respectivos para el entrenamiento del modelo de Machine Learning.

Figura 54

Herramienta de transformación de Excel a JSON

EXCEL a JSON
Convertir EXCEL a JSON en línea gratis

- 1: Elija varios archivos EXCEL locales o
- 2: Ingrese la URL del archivo EXCEL en línea.
- 3: Elija "JSON" como formato de destino y configure las opciones.
- 4: Haga clic en el botón "INICIAR CONVERSIÓN" para convertir EXCEL a JSON en línea.

Elija archivos o ingrese la URL del archivo remoto

FORMATO DE DESTINO: Establecer opciones de conversión (opcional)

Acepto términos y privacidad

ARCHIVOS DE SALIDA

#	NOMBRE DEL ARCHIVO	TAMAÑO DE ARCHIVO	DESCARGAR

Luego de la limpieza de los datos respectivo para el modelo, mediante una herramienta de formatos en línea, se procede con la transformación de la data desde el formato Excel al formato JSON.

Figura 55

Visualización de la data en formato JSON

```

1 [
2   {
3     "PRIMERA_EMISION": "201010", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "22", "MONTO_PENSION_INICIAL": "284.2", "MONTO_HABILITADO": "500", "APELLI":
4   },
5   {
6     "PRIMERA_EMISION": "201011", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "32", "MONTO_PENSION_INICIAL": "863.82", "MONTO_HABILITADO": "893", "APELLI":
7   },
8   {
9     "PRIMERA_EMISION": "201111", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "29", "MONTO_PENSION_INICIAL": "911.83", "MONTO_HABILITADO": "893", "APELLI":
10  },
11  {
12   "PRIMERA_EMISION": "201010", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "13", "MONTO_PENSION_INICIAL": "305.19", "MONTO_HABILITADO": "500", "APELLI":
13  },
14  {
15   "PRIMERA_EMISION": "201111", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "20", "MONTO_PENSION_INICIAL": "686.33", "MONTO_HABILITADO": "670", "APELLI":
16  },
17  {
18   "PRIMERA_EMISION": "201010", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "25", "MONTO_PENSION_INICIAL": "323.11", "MONTO_HABILITADO": "500", "APELLI":
19  },
20  {
21   "PRIMERA_EMISION": "201010", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "21", "MONTO_PENSION_INICIAL": "29.33", "MONTO_HABILITADO": "500", "APELLI":
22  },
23  {
24   "PRIMERA_EMISION": "201206", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "31", "MONTO_PENSION_INICIAL": "526.22", "MONTO_HABILITADO": "640", "APELLI":
25  },
26  {
27   "PRIMERA_EMISION": "201206", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "31", "MONTO_PENSION_INICIAL": "526.22", "MONTO_HABILITADO": "640", "APELLI":
28  },
29  {
30   "PRIMERA_EMISION": "201010", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "24", "MONTO_PENSION_INICIAL": "631.76", "MONTO_HABILITADO": "710", "APELLI":
31  },
32  {
33   "PRIMERA_EMISION": "201112", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "20", "MONTO_PENSION_INICIAL": "14.36", "MONTO_HABILITADO": "500", "APELLI":
34  },
35  {
36   "PRIMERA_EMISION": "201010", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "34", "MONTO_PENSION_INICIAL": "684.75", "MONTO_HABILITADO": "4400", "APELLI":
37  },
38  {
39   "PRIMERA_EMISION": "201301", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "25", "MONTO_PENSION_INICIAL": "25.22", "MONTO_HABILITADO": "500", "APELLI":
40  },
41  {
42   "PRIMERA_EMISION": "201111", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "41", "MONTO_PENSION_INICIAL": "1367.92", "MONTO_HABILITADO": "893", "APELLI":
43  },
44  {
45   "PRIMERA_EMISION": "201010", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "29", "MONTO_PENSION_INICIAL": "109.17", "MONTO_HABILITADO": "500", "APELLI":
46  },
47  {
48   "PRIMERA_EMISION": "201204", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "30", "MONTO_PENSION_INICIAL": "334.28", "MONTO_HABILITADO": "527", "APELLI":
49  },
50  {
51   "PRIMERA_EMISION": "201007", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "27", "MONTO_PENSION_INICIAL": "784.03", "MONTO_HABILITADO": "628", "APELLI":
52  },
53  {
54   "PRIMERA_EMISION": "201206", "ESTADO_PRESTACION": "ACT", "AÑOS_APORTES": "20", "MONTO_PENSION_INICIAL": "46.88", "MONTO_HABILITADO": "500", "APELLI":
55  },
56  }
57 ]

```

Luego de proceder con la transformación del formato de los datos, se muestra en esta figura toda la data visualizada a través de la herramienta Visual Studio Code el archivo en formato JSON.

4.3.3. Evaluación y Ajustes

4.3.3.1. Aplicar el modelo a los datos y medir su precisión.

Figura 56

Depuración de la data

NOMBRE APELLIDO_PATERNO APELLIDO MATERNO	E	edad	E	sexo	E	rango_edad_jubilarse	RANGO DE JUBILACION	E	sueldo
'Nombre_apellido' ABDIAS CHAMO FERNANDEZ		22	seco	M		' rango_edad_jubilarse'	60	'sueldo'	
'Nombre_apellido' ABDIAS CHAMBE FASABI		20	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDIAS CERRANO MORALES		29	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDIAS HERILAY RUIZ		38	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDIAS MONTEZA VALDEZ		26	seco	M		' rango_edad_jubilarse'	65	'sueldo'	
'Nombre_apellido' ABDIAS FABRA CASTRO		30	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDIAS RODRIGUEZ GUERRA		23	seco	M		' rango_edad_jubilarse'	45	'sueldo'	
'Nombre_apellido' ABDIAS RODRIGUEZ GUERRA		21	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDIAS SOCRATES MUÑOZ TORREÓN		21	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDIAS ZEGARRA MALAVEY		28	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON LEONIDAS GARCERAN SALAS		37	seco	M		' rango_edad_jubilarse'	65	'sueldo'	
'Nombre_apellido' ABDON ALEJANDRO AZEVEDO FLORES		21	seco	M		' rango_edad_jubilarse'	60	'sueldo'	
'Nombre_apellido' ABDON ALEJANDRO AZEVEDO FLORES		32	seco	M		' rango_edad_jubilarse'	65	'sueldo'	
'Nombre_apellido' ABDON ALEJANDRO BALLEGAARD SCHICHTANA		36	seco	M		' rango_edad_jubilarse'	60	'sueldo'	
'Nombre_apellido' ABDON ALFONSO CALENES PINTO		21	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON ANDRÉS CUENCA		29	seco	M		' rango_edad_jubilarse'	60	'sueldo'	
'Nombre_apellido' ABDON AZARÁ CABRILLO		23	seco	M		' rango_edad_jubilarse'	65	'sueldo'	
'Nombre_apellido' ABDON CABALLERO ENRIQUEZ		22	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON CALLE RIVERA		37	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDON CANDOTTI SARMENTO		20	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON CLAUDIO SOGA MONTAÑA		20	seco	M		' rango_edad_jubilarse'	65	'sueldo'	
'Nombre_apellido' ABDON COSAGLIA ROMAN		18	seco	M		' rango_edad_jubilarse'	65	'sueldo'	
'Nombre_apellido' ABDON ESPINOZA QUISEP		19	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDON FERNANDEZ PALOMINO		36	seco	M		' rango_edad_jubilarse'	60	'sueldo'	
'Nombre_apellido' ABDON FERNANDEZ PALOMINO		31	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDON FERNANDEZ PALOMINO		32	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON FIGUEROA LEIVA		35	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON FLORES CALATO		37	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON FRIASCHIO CAPIA		36	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDON FRIASCHIO CAPIA		37	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON GAMBORA NAUPAS		20	seco	M		' rango_edad_jubilarse'	60	'sueldo'	
'Nombre_apellido' ABDON GAMBORA NAUPAS		31	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON GARCIA SEGUEBROS		28	seco	M		' rango_edad_jubilarse'	65	'sueldo'	
'Nombre_apellido' ABDON HILANCA QUISEP		21	seco	M		' rango_edad_jubilarse'	65	'sueldo'	
'Nombre_apellido' ABDON HUGO CORDOVA CHAVEZ		32	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDON ILIARDO LAZO		37	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDON ISIDRO POSTIGO RAMOS		38	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON ISRAEL MORENO CASTILLO		28	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON MANUEL CAJAMA		18	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON MANUEL VALERO GUZMAN		36	seco	M		' rango_edad_jubilarse'	65	'sueldo'	
'Nombre_apellido' ABDON MONDRAGON MORETO		23	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDON NAVARRETE PACHECO		37	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDON PALLI YANA		33	seco	M		' rango_edad_jubilarse'	45	'sueldo'	
'Nombre_apellido' ABDON PARRAMAZA CHOQUEHUANCA		36	seco	M		' rango_edad_jubilarse'	55	'sueldo'	
'Nombre_apellido' ABDON QUISEP PALCAZ		33	seco	M		' rango_edad_jubilarse'	50	'sueldo'	
'Nombre_apellido' ABDON RAMIREZ CORREA		27	seco	M		' rango_edad_jubilarse'	65	'sueldo'	
'Nombre_apellido' ABDON RAMIREZ CORREA		31	seco	M		' rango_edad_jubilarse'	65	'sueldo'	

Se realiza la depuración de datos para garantizar que el conjunto de datos sea preciso, asimismo, confiable. Esto implica corregir errores de entrada, eliminar datos duplicados o inconsistentes, y estandarizar formatos para evitar confusiones. Al depurar datos, se eliminan barreras que podrían distorsionar los resultados o conducir a conclusiones erróneas. Con la data depurada y precisa, se confía en la información presentada para respaldar sus acciones, esto es crucial para la planificación estratégica y la gestión efectiva de los resultados esperados para el presente trabajo.

4.3.3.2. Realizar ajustes basados en los resultados y repetir la evaluación hasta alcanzar los criterios de éxito.

Figura 57
Data ajustada

NOMBRE_EMPLEADO	PRESTACION	AÑO	APORTES	MONTO_PENSION	MONTO_MANTENIMIENTO	APELLIDO_PATERNO	APELLIDO MATERNO	NOMBRE	FECHA_NACIMIENTO
1100251102	AC	30	124.24	300.00	300.00	MENDOZA	CADANILAS	LUZ ANSELICA	20/07/1938
1100251102	AC	30	124.24	300.00	300.00	CANALAS	RAMAYO	GUERRA NEGRERA	28/07/1944
1100251104	AC	34	483.90	337.00	337.00	DE LA CRUZ	CABALLERO	JUAN CESAR	19/06/1939
1100251102	AC	34	764.47	807.00	807.00	CRUZ	GARCIA	CARLOS	04/02/1946
1100251102	AC	23	384.23	300.00	300.00	VOC	MIRANDA	ALFREDO	12/02/1940
1100251102	AC	29	30.98	300.00	300.00	CHACABANCAO	SEMPERNA	TERESA ALDAMERA	01/11/1931
1100251102	AC	19	133.78	810.00	810.00	RAM	MAMANI	LUIS	04/10/1947
1100251102	AC	20	6.08	300.00	300.00	MARINOSO	LUNA	LORENZO ORMANET	07/06/1942
1100251103	AC	28	204.04	300.00	300.00	MERINO	PIERA	DENIS	07/09/1948
1100251102	AC	27	95.88	770.00	770.00	MALDONADO	SANCHEZ	FRANCISCO	01/09/1946
1100251103	AC	28	95.88	770.00	770.00	MALDONADO	SANCHEZ	FRANCISCO	03/03/1945
1100251102	AC	28	6.05	300.00	300.00	PANANDEIA	RODRIGUEZ	VICENTE EDUARDO	29/10/1940
1100251102	AC	33	379.82	300.00	300.00	RAUZ	FRIEZ	CARLOS	07/06/1945
1100251102	AC	33	352.90	300.00	300.00	SORIA	VILLALBERTO	SANTOS DANIEL	10/07/1933
1100251102	AC	30	232.25	300.00	300.00	ESPINOSA	CLARE	JUAN ANTONIO	05/03/1943
1100251102	AC	20	461.17	300.00	300.00	PAREDES	SANCHEZ	JOSE CLEMENTE	07/09/1940
1100251102	AC	42	1341.76	300.00	300.00	MARQUE	SHIMONAKA	ANA	04/02/1943
1100251102	AC	43	1598.44	810.00	810.00	AVILES	VEGA	EDUARDO	01/11/1941
1100251102	AC	33	381.80	300.00	300.00	DE TRUJILLO	SANABANDEIA	DORISNIA ENRIQUETA	04/02/1934
1100251102	AC	33	48.30	300.00	300.00	VILLALVEJA	CORREA	TEODORO OSAMAR	14/11/1943
1100251102	AC	27	498.27	300.00	300.00	SODONA	TORRES	VICTOR FAL	22/02/1942
1100251102	AC	42	1124.84	491.00	491.00	DAVILA	FERNANDEZ	RICARDO	10/04/1949
1100251102	AC	36	1722.08	810.00	810.00	SOTOMAYOR	PAZ	DAVID RETHALDO	03/07/1944
1100251102	AC	33	376.46	300.00	300.00	ACOSTA	CRUZ	JOSÉ	07/06/1945
1100251102	AC	36	486.40	551.00	551.00	FRANQUE	ROQUE DE FRETO	MARIA FELICIA	12/05/1935
1100251102	AC	23	316.45	300.00	300.00	MARABANDA	MUÑOZ	JOSE	12/06/1940
1100251102	AC	26	383.86	810.00	810.00	PINEDA	PALOMINO	MARGARITA	12/06/1940
1100251102	AC	36	397.70	450.00	450.00	ESPINOSA	CACABANCAO	ALVARO AGUSTIN	20/06/1946
1100251102	AC	36	461.90	720.00	720.00	MARTINEZ	LAY	EDUARDO NARCIZO	12/06/1940
1100251104	AC	25	361.30	300.00	300.00	SANCHEZ	ROMERO DE BARRÉ	JULIA FABRILA	7/04/1945
1100251104	AC	25	401.90	300.00	300.00	VEP	GALLERES	LIDIA MURTA	01/11/1934

En base a la data recolectada, se presentó un total de 50000 de los datos, en el cual se tuvo que simplificar a un total de 5000 registros, debido a que se descartaron registros incompletos que presentaban redundancia o también que los datos no eran correctos, por lo cual se hizo las modificaciones en la hoja de cálculo.

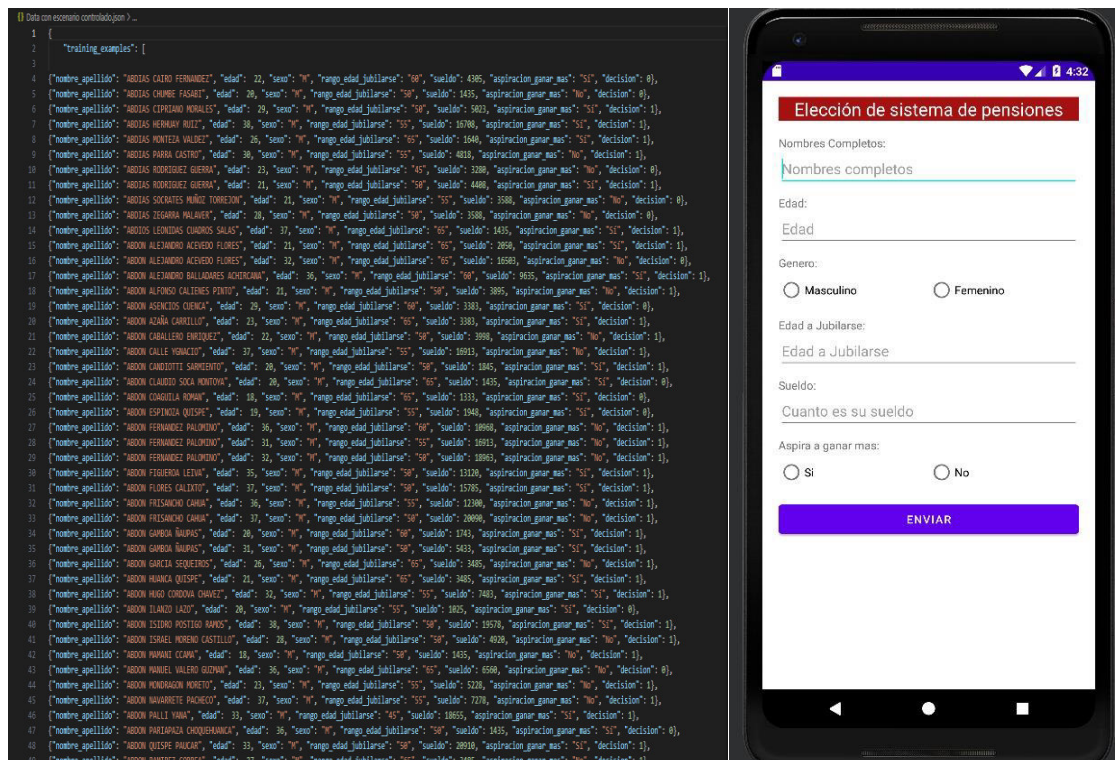
4.4. Evaluación de la eficiencia operativa del modelo

4.4.1. Implementación en Ambiente de Prueba

4.4.1.1. Integrar el modelo en una versión de prueba de la aplicación móvil.

Figura 58

Entrenamiento con data controlada

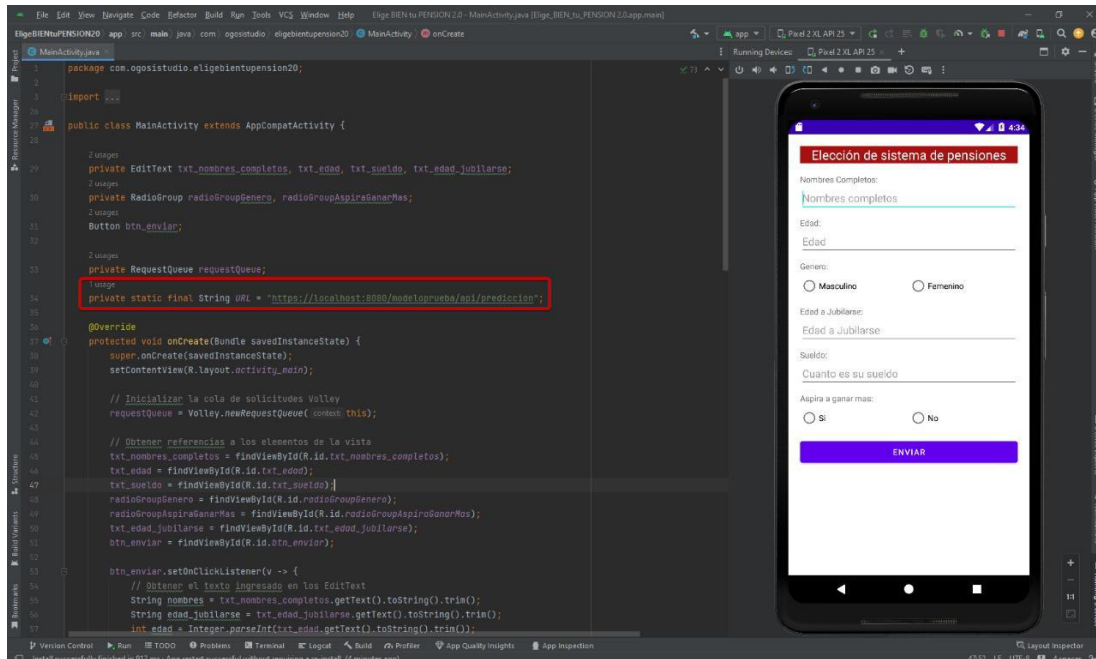


Se ejecuta la data empleando la información controlada, la cual está organizada en formato JSON, con la finalidad de derivar los índices de exactitud mediante el uso de la aplicación móvil.

4.4.1.2. Configurar un ambiente de prueba para la evaluación.

Figura 59

Ambiente de prueba para evaluar el modelo de la proyección



En el entorno de pruebas de Android Studio, se ha establecido la configuración con el propósito de efectuar la evaluación del modelo de proyección y en base a ello poder determinar el valor de la precisión.

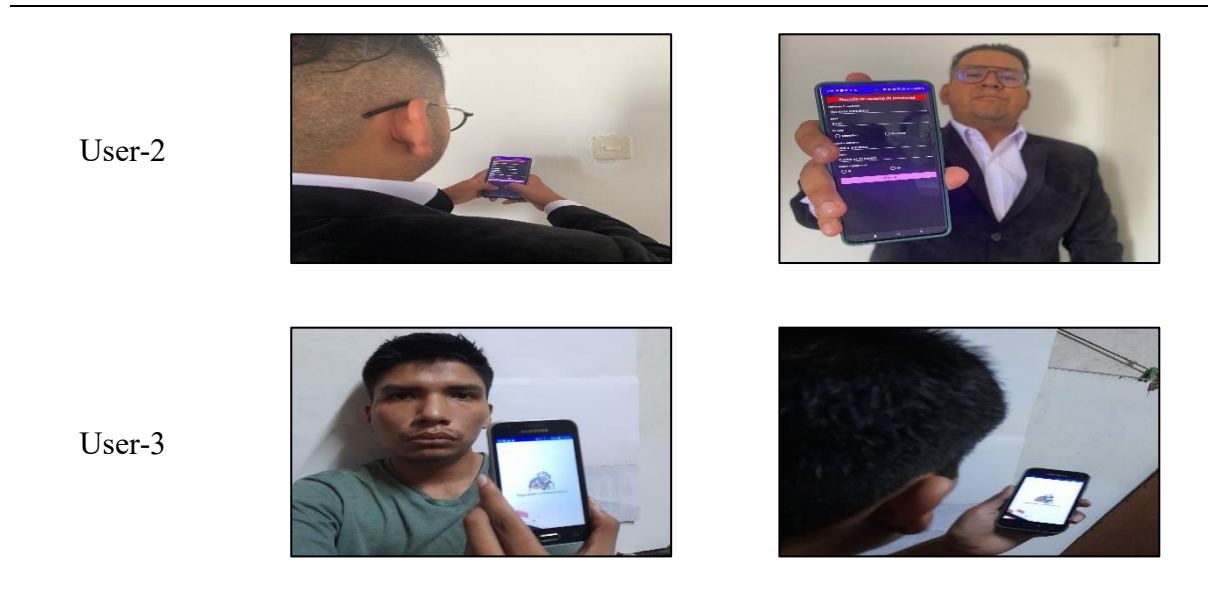
4.4.2. Pruebas de Usabilidad y Eficiencia

4.4.2.1. Realizar pruebas de usabilidad con usuarios finales.

Tabla 15

Evidencia de uso del aplicativo móvil

Usuario	Evidencias
User-1	



4.4.2.2. Medir la eficiencia operativa en términos de tiempo de respuesta y recursos utilizados.

Figura 60

Medición de la eficiencia operativa



Se exhibe el envío de la información de un usuario para medir la eficiencia operativa, el cual muestra un mensaje de confirmación, el cual indica que se ha realizado la recepción de la data, y en base a ello, se observa que la precisión es de 0,7129.

4.4.3. Análisis de Feedback y Optimización

4.4.3.1. Analizar el feedback de los usuarios y los datos de eficiencia.

Tabla 16

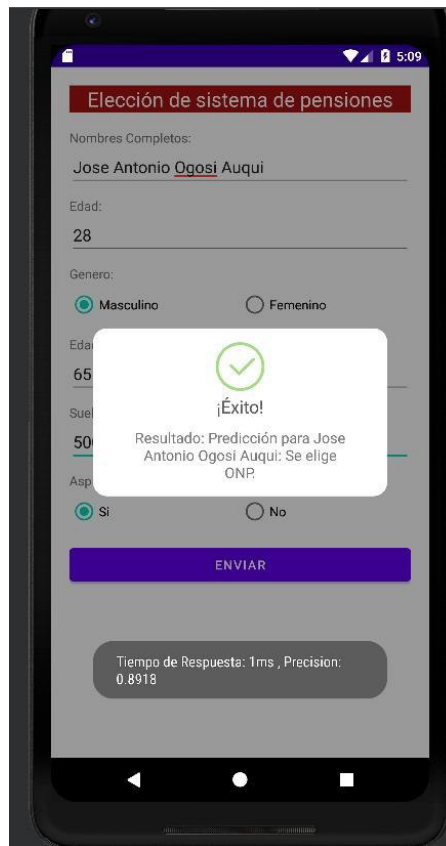
Feedback del usuario final

Usuario	Comentario final
User-1	“¡Increíble aplicación! Siempre he estado confundido sobre si elegir AFP u ONP, pero esta app lo hace fácil. Solo ingresé mis datos y recibí una predicción clara y detallada sobre cuál sería la mejor opción para mi jubilación. ¡Altamente recomendado!”
User-2	“Me sorprendió lo precisa que fue esta aplicación al prever mi situación pensionaria. Antes de probarla, estaba indeciso sobre qué sistema elegir, pero ahora tengo una dirección clara. Además, la interfaz presenta la característica relevante de ser intuitiva y de fácil uso. Definitivamente la utilizaré para hacer mis decisiones financieras en el futuro.”
User-3	“La aplicación móvil del sistema pensionario es una herramienta indispensable para cualquier persona preocupada por su jubilación. Me encantó cómo considera una amplia gama de factores para proporcionar recomendaciones personalizadas. Es como tener a un asesor financiero en mi bolsillo. ¡No puedo dejar de recomendarla a mis amigos y familiares!”

4.4.3.2. Optimizar el modelo y la aplicación para mejorar la eficiencia operativa.

Figura 61

Medición de la eficiencia operativa optimizada



Se representa la transmisión de los datos de un usuario para medir la eficiencia operativa optimizada, el cual muestra un mensaje de confirmación, el cual indica que se ha realizado la recepción de la data, y en base a ello, se observa que la precisión es de 0,8918.

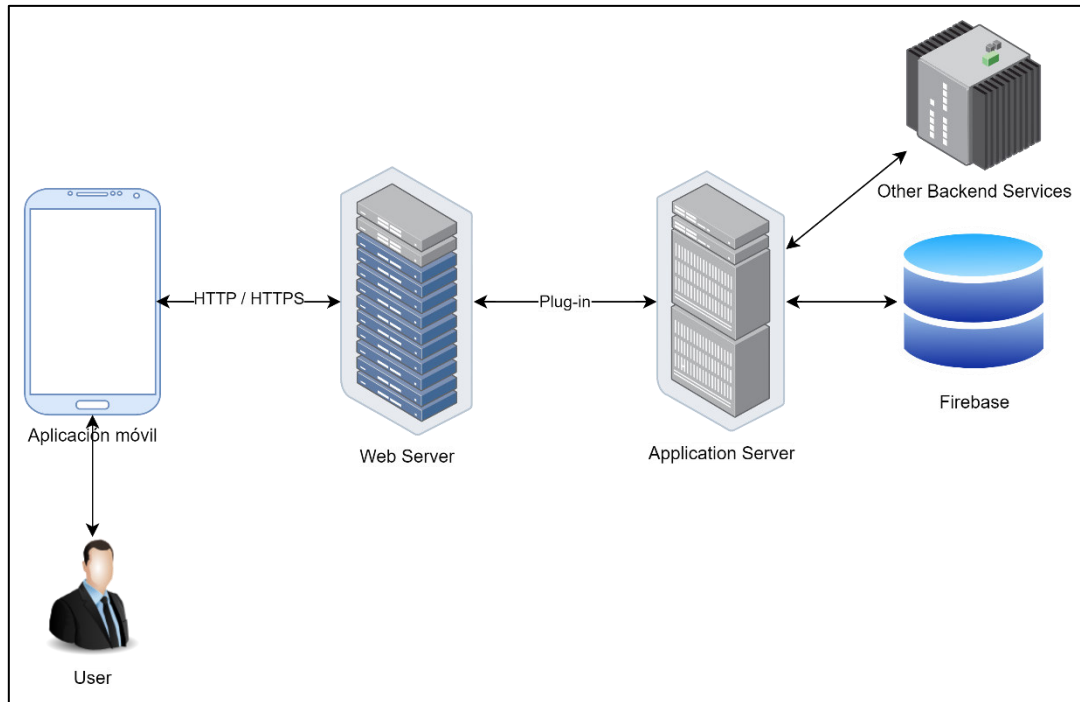
4.5. Implementación de la aplicación móvil y su efectividad

4.5.1. Despliegue de la Aplicación

4.5.1.1. Preparar la infraestructura necesaria para el despliegue de la aplicación móvil.

Figura 62

Infraestructura del despliegue de la aplicación móvil



En base a la infraestructura mostrada, el usuario comienza con la interacción de la aplicación móvil en el cual envía instrucciones al servidor web sumado al servidor de la aplicación la cual esta está conectada a otros servicios de backend y a la database proveniente del servicio de Firestore de Firebase en donde estas dos últimas envían la respuesta solicitada del servidor de la aplicación, seguido de la respuesta del servidor Web para luego ser proyectada en la aplicación móvil en vista del usuario quien fue que envió dicha solicitud.

Figura 63

Estructura de Firebase de la aplicación móvil



Ahora bien, la plataforma de Firebase cuenta con la siguiente estructura en donde esta transmite toda la data que se solicita por medio de la aplicación móvil la cual esta está conectada al servicio de Firestore; responsable de contener toda la data que se almacena para dicha aplicación.

Figura 64

Logo de la plataforma Firebase



Esta plataforma de desarrollo de apps mobile y web se considera, debido a su amplia gama de servicios integrados, que incluye los sgtes. servicios: user authentication , database realtime y cloud storage. Su relevancia reside en su capacidad para simplificar el proceso de desarrollo al brindar herramientas poderosas y escalables que posibilitan a la gesta de apps de alta calidad rápida y eficientemente. Además, ofrece una fiabilidad y seguridad en su infraestructura, lo que significa un escalamiento del sistema de forma fácil a medida que aumentan en usuarios y funciones, convirtiéndola en una elección popular tanto para empresas como para desarrolladores independientes.

4.5.1.2. Desplegar la aplicación en el entorno de producción.

Figura 65

Logo de la plataforma Google Play



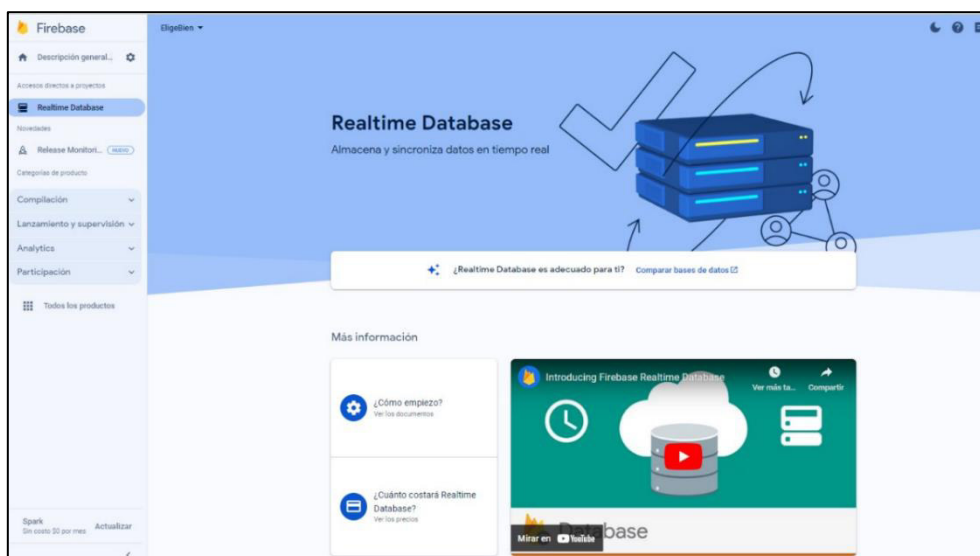
La aplicación móvil al momento de concretarse en su totalidad, esto es registrado para luego ser lanzada a la plataforma de Google Play en un lapso de 2 a 3 meses.

4.5.2. *Monitoreo y Recolección de Datos*

4.5.2.1. Monitorear la aplicación en uso real y recoger datos sobre su efectividad.

Figura 66

Entorno de la plataforma de Firebase



En base a la figura mostrada, se tiene conocimiento constante de la data que se registra dentro de la BD de la app en donde, gracias al servicio de Firestore, se puede ver en tiempo real, la base de datos donde, como se había mencionado, todos los datos que se ingresan dentro del modelo de Machine Learning.

4.5.2.2. Realizar encuestas y entrevistas con usuarios finales y personal de RRHH.

Figura 67

Formulario para evaluar la satisfacción del usuario final y personal de RRHH

**ENCUESTA DE SATISFACCIÓN
SOBRE LA IMPLEMENTACIÓN DE
UNA APLICACIÓN MÓVIL CON
MACHINE LEARNING APLICANDO LA
NUEVA METODOLOGÍA CSKT PARA
EL PROCESO DE CONTRATACIÓN DE
PERSONAL EN EL SECTOR PRIVADO**

¡Saludos a todos nuestros valiosos participantes! Queremos invitarlos cordialmente a formar parte de nuestra encuesta. Sus opiniones son esenciales para nosotros, y hemos estructurado cuidadosamente cada pregunta para capturar de manera eficiente sus valiosas perspectivas. Además, queremos enfatizar que la confidencialidad es nuestra prioridad, ya que todas las respuestas serán tratadas de forma anónima, garantizando un ambiente seguro para compartir sus ideas. Les agradecemos de antemano por su tiempo y colaboración en este proceso, ¡esperamos con entusiasmo conocer sus respuestas y enriquecer nuestro entendimiento con sus valiosas contribuciones!

[Iniciar sesión en Google](#) para guardar lo que llevas hecho. [Más información](#)

* Indica que la pregunta es obligatoria

Se realiza un formulario en Google Forms para poder medir la satisfacción de los usuarios en base al aplicativo móvil con machine learning aplicando la metodología CSKT, cabe resaltar que se respeta la confiabilidad, intimidad y confidencialidad de la información brindada por los usuarios, lo mismo que la seguridad de los datos a brindarse. El formulario realizado está albergado en:

https://docs.google.com/forms/d/e/1FAIpQLScg45fSCdgGR3xsqHcM6cLLoXS0H7jJFz6iQG_DWBGGbB2b29Q/viewform

4.5.3. Evaluación de Impacto y Ajustes

4.5.3.1. Evaluar el impacto de la aplicación en el proceso de contratación de personal.

Figura 68

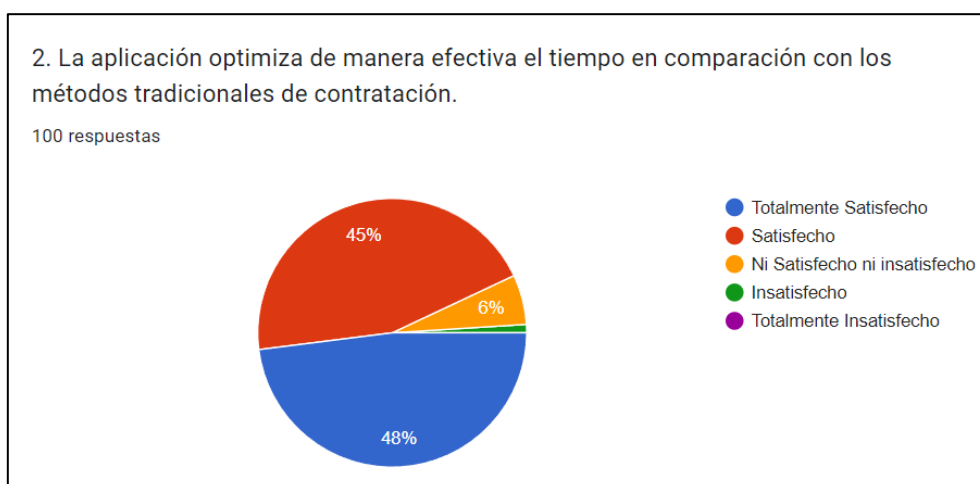
Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 1



Las respuestas recopiladas de los encuestados revelan un nivel significativo de satisfacción entre los encuestados. La mayoría de los encuestados, representada por el 67%, indicó estar “Totalmente Satisfecho”, subrayando una percepción altamente positiva en cuanto a la eficiencia de la interfaz para facilitar la gestión de recursos y mejorar la eficiencia del proceso de contratación. Adicionalmente, el 29% expresó estar “Satisfecho”, respaldando la conclusión general de que la interfaz cumple con las expectativas y resulta satisfactoria para una parte considerable de los usuarios. Sugiriendo que el 96% de los participantes experimentan una satisfacción respecto a la interfaz del aplicativo. Estos resultados indican que, se ha logrado impactar positivamente, según la perspectiva de los encuestados.

Figura 69

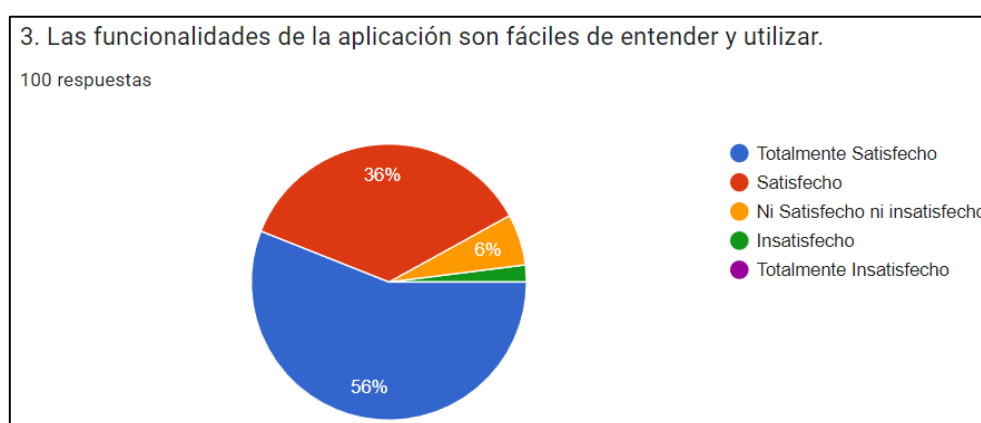
Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 2



En relación con la eficiencia temporal comparada con métodos tradicionales, un 67% de los encuestados se manifestó “Totalmente Satisfecho”, respaldando la afirmación de que la aplicación optimiza efectivamente el tiempo. Además, un 48% reflejan estar totalmente satisfechos y el otro 45% de encuestados indicaron que están satisfechos. Se infiere que el 93% del total de encuestados están satisfechos. Por ello, se puede decir que, la aplicación optimiza de manera efectiva el tiempo en comparación con los métodos tradicionales de contratación.

Figura 70

Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 3



Las respuestas recopiladas a la pregunta revelan una tendencia generalmente positiva por parte de los usuarios. Con un 56% de los encuestados indican “Totalmente Satisfecho” y un 36% declarando “Satisfecho”, la mayoría muestra una percepción favorable hacia la accesibilidad y usabilidad de la aplicación móvil. Además, se obtiene que el 92% de los participantes encuestados y encuentra satisfechos con las funcionalidades del aplicativo dada la facilidad de entendimiento y utilización. Esta distribución refleja una aceptación mayoritaria de las funcionalidades de la aplicación, destacando su efectividad en términos de comprensión y utilización, aunque aún se debe considerar la retroalimentación de aquellos usuarios que no se sintieron completamente satisfechos para posibles mejoras futuras.

Figura 71

Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 4



Las respuestas recopiladas, reflejan una impresión mayoritariamente favorable de parte de los usuarios. Un significativo 45% de los encuestados expresaron estar “Totalmente Satisfechos”, mientras que otro 45% manifestó estar “Satisfecho”, indicando así un consenso general en cuanto a la eficacia de la aplicación en proporcionar las herramientas necesarias para un proceso de contratación completo. Se obtiene que el 90% que representa al total de encuestados estás satisfechos con las herramientas que proporciona la app para el proceso de contratación. Estos resultados respaldan la efectividad de la aplicación en satisfacer a las necesidades en términos de herramientas proporcionadas para la contratación de personal.

Figura 72

Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 5

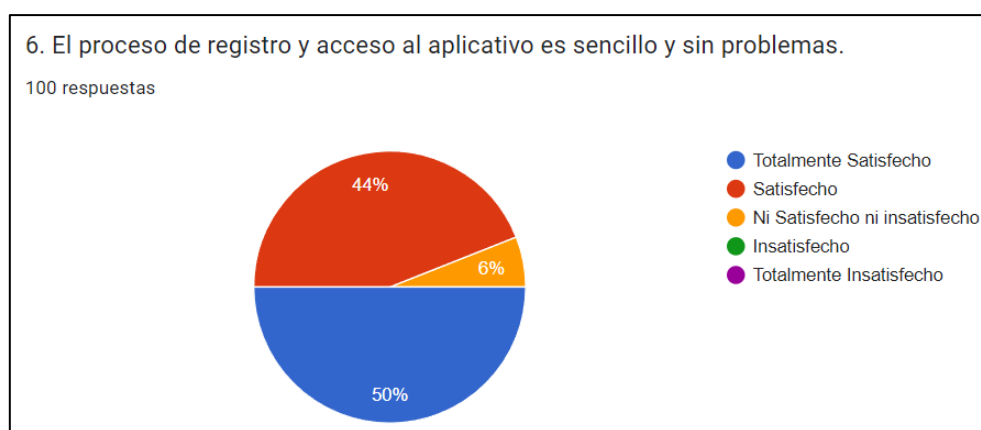


Las respuestas recabadas con relación a la pregunta se revelan: con un notable 53% de los encuestados indicando estar “Totalmente Satisfechos” y un sólido 42% expresando estar

“Satisfechos”, dado lo mostrado visualmente se obtiene que la mayor parte de los usuarios encuentra la interfaz del aplicativo fácil de entender y navegar. El 95% del total de encuestados presentan una satisfacción relacionada al entendimiento y navegación de la app. En conjunto, estos resultados reflejan una experiencia positiva y eficiente en la interacción con la interfaz del aplicativo, lo cual es esencial para una experiencia de usuario exitosa.

Figura 73

Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 6



Las respuestas recopiladas, reflejan una satisfacción significativa por parte de los usuarios. Con un notable 50% de los encuestados manifestando estar “Totalmente Satisfechos” y un sólido 44% indicando estar “Satisfechos”, dado lo mostrado visualmente se obtiene que gran parte del conjunto de usuarios experimenta un proceso de registro y acceso sin complicaciones. Se obtiene que el 94% de encuestados se encuentran satisfechos sobre la sencillez que brinda el aplicativo en el proceso de registro y su acceso en sí. En conjunto, estas respuestas indican una experiencia generalmente positiva y exitosa en el manejo del proceso de registro y acceso al aplicativo, lo cual contribuye significativamente a una experiencia de usuario positiva y sin contratiempos.

Figura 74

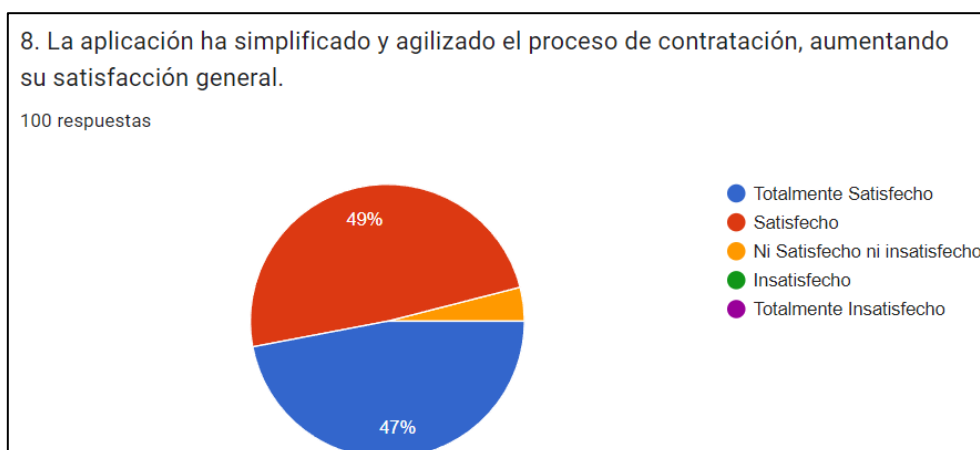
Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 7



Las respuestas obtenidas en relación con la pregunta reflejan una evaluación positiva por parte de los usuarios. Con un sólido 52% de los encuestados indicando estar “Totalmente Satisfechos” y un significativo 36% expresando estar “Satisfechos”, la mayoría percibe que el aplicativo destaca de manera efectiva las habilidades y la experiencia clave de los candidatos. Se infiere que el porcentaje que se encuentra satisfecho sobre la efectividad de las habilidades claves, así como, la experiencia de los candidatos corresponde al valor de 88%. En conjunto, estos resultados sugieren que el aplicativo ha logrado, en su mayoría, resaltar de manera exitosa las habilidades y la experiencia esenciales de los candidatos, contribuyendo así a una evaluación positiva por parte de los usuarios.

Figura 75

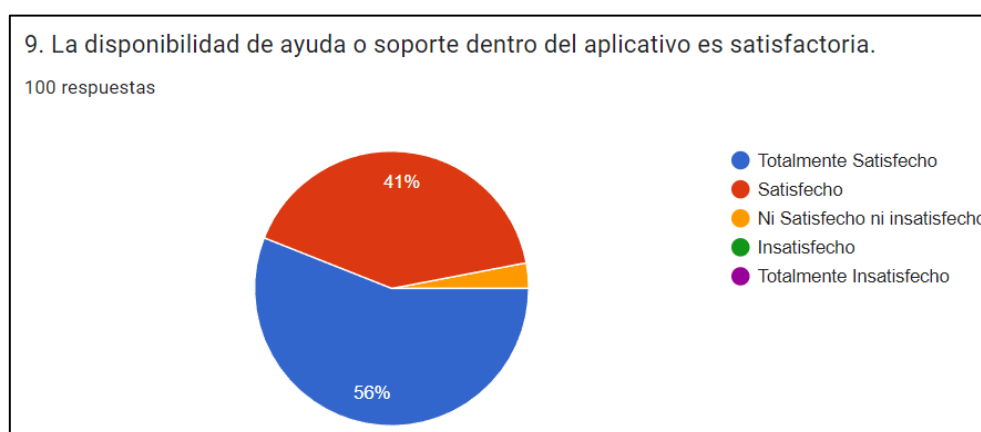
Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 8



Las respuestas recabadas en torno a la pregunta reflejan una evaluación positiva y optimista por parte de los usuarios. Con un destacado 47% de los participantes indicando estar “Totalmente Satisfechos” y un sólido 49% manifestando estar “Satisfechos”, la gran mayoría percibe que la aplicación ha logrado simplificar y agilizar eficazmente el proceso de contratación. Los usuarios encuestados que se encuentran satisfechos sobre la simplificación y agilización que brinda la app corresponde al 96% del total de encuestados. En conjunto, estos resultados reflejan un impacto positivo significativo de la aplicación al mejorar la eficiencia y la satisfacción general en el proceso de contratación, indicando una experiencia exitosa y beneficiosa para los usuarios.

Figura 76

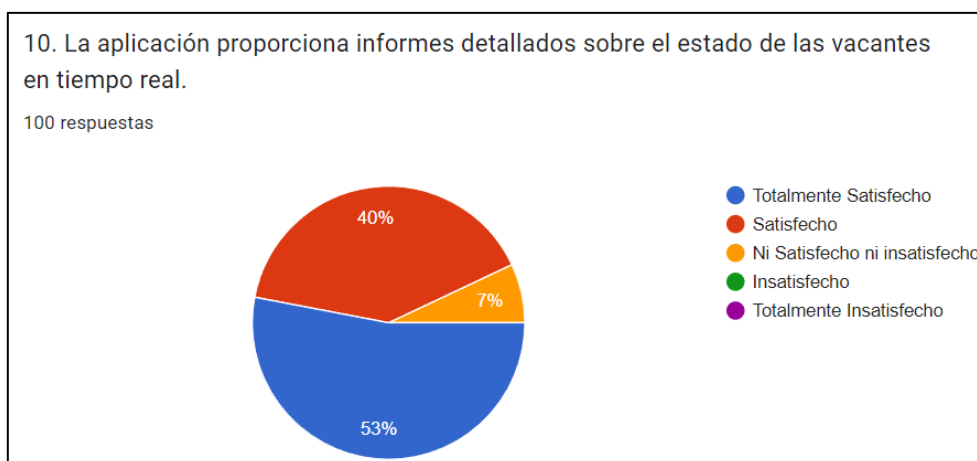
Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 9



Las respuestas obtenidas respecto a la pregunta reflejan: un destacado 56% de los encuestados indicando estar “Totalmente Satisfechos” y un significativo 41% expresando estar “Satisfechos”, la mayoría equivalente al 97% de los usuarios reporta un nivel de satisfacción considerable con el soporte disponible dentro de la aplicación. En conjunto, estos resultados sugieren que la aplicación ha logrado proporcionar un nivel satisfactorio de ayuda y soporte a los usuarios, contribuyendo así a una experiencia general positiva y eficiente.

Figura 77

Respuesta del usuario final para evaluar la satisfacción del aplicativo móvil - Pregunta 10



Las respuestas recabadas, da a mostrar una consideración positiva por parte de los usuarios. Con un notorio 53% de los encuestados indicando estar “Totalmente Satisfechos” y un sólido 40% expresando estar “Satisfechos”, la mayoría de los usuarios percibe que la aplicación ha cumplido efectivamente con proporcionar informes detallados en tiempo real sobre el estado de las vacantes. El 93% del total de usuarios encuestados reflejan una satisfacción sobre el estado de las vacantes proporcionadas por el aplicativo. En resumen, estos resultados sugieren que la aplicación satisface a los usuarios al proporcionar informes detallados en tiempo real sobre las vacantes, contribuyendo así a una experiencia de usuario positiva y eficiente.

4.5.3.2. Identificar áreas de mejora y realizar ajustes en la aplicación y/o el modelo de Machine Learning.

Tabla 17

Comentarios de usuarios para la mejora y ajustes en el aplicativo móvil

Usuario	Comentario final
1-User	Sería genial tener más opciones de personalización en la aplicación.
2-User	Agregar un chat en vivo para resolver consultas al momento sería útil.
3-User	Podrían mejorar la velocidad de carga de algunas funciones.
4-User	Incluir tutoriales cortos para usuarios nuevos sería beneficioso.

5-User	Sería útil tener un calendario de eventos relacionados con la jubilación.
6-User	Agregar una sección de preguntas frecuentes podría reducir las consultas.
7-User	Mejorar la precisión de las simulaciones de jubilación sería excelente.
8-User	Incluir una opción de retroalimentación para usuarios ayudaría a mejorar la aplicación.
9-User	Agregar la posibilidad de gestionar documentos directamente desde la aplicación sería conveniente.
10-User	Sería beneficioso tener un sistema de recompensas por el uso continuo de la aplicación.
11-User	Incorporar un sistema de notificaciones más personalizable sería útil.
12-User	Ampliar la compatibilidad con diferentes dispositivos mejoraría la accesibilidad.
13-User	Agregar un centro de recursos con información adicional sobre pensiones sería valioso.
14-User	Implementar un sistema de seguimiento de trámites para mayor transparencia.
15-User	Ofrecer herramientas de planificación financiera más avanzadas sería beneficioso.
16-User	Mejorar la integración con otros servicios financieros facilitaría la gestión global.
17-User	Incluir un acceso directo a asesores expertos para consultas específicas.
18-User	Ofrecer opciones de idioma adicionales para usuarios internacionales.
19-User	Desarrollar una versión lite para dispositivos con recursos limitados.

20-User	Incorporar un sistema de retroalimentación <i>realtime</i> en pro de la experiencia del usuario.
---------	--

Prueba de HE1

HE1: La metodología CSKT aplicada en el modelo de Machine Learning mejora la eficiencia en términos de tiempo y los costos en el proceso de contratación de personal.

H_0 = La metodología CSKT aplicada en el modelo de Machine Learning **no mejora significativamente** la eficiencia en términos de tiempo y los costos en el proceso de contratación de personal.

H_1 = La metodología CSKT aplicada en el modelo de Machine Learning **mejora significativamente** la eficiencia en términos de tiempo y los costos en el proceso de contratación de personal.

Procedimiento de la prueba de hipótesis:

Se requiere:

- Una medida justa para evaluar el rendimiento de la metodología CSKT aplicada en el modelo de Machine Learning en la eficiencia en términos de tiempo y los costos en el proceso de contratación de personal, como el Recall.
- Un **dataset relevante** con y sin la adaptación del modelo para entrenar y probar la data con la metodología CSKT y la data de la SBS incrementada.
- Debido a que se debe comparar el rendimiento de la metodología CSKT aplicada en el modelo de Machine Learning, se puede utilizar:
 - La **prueba t** si se comparan las **medias de la métrica de evaluación** entre los métodos.

- La prueba de hipótesis para proporciones si estamos comparando tasas de eficiencia en término de tiempo y costos.
- A posteriori, se selecciona el **nivel sig.**, generalmente 0.05.
- Utilizar la **data recopilada** para calcular la **estadística de prueba** (diferencia de medias o diferencia de proporciones) y determinar si es estadísticamente significativa.
- Si $p < \text{umbral sig. establecido}$, entonces se invalida la H_0 y se deduce la existencia de evidencia adecuada para sustentar la hipótesis alternativa, H_1 .
- Las limitaciones del estudio de:
 - **sesgos en los datos,**
 - **tamaño del conjunto de datos,**
 - **representatividad de las muestras, etc.**

Prueba de hipótesis:

H_1 = La metodología CSKT aplicada en el modelo de Machine Learning **mejora significativamente** la eficiencia en términos de tiempo y los costos en el proceso de contratación de personal.

- Valor alcanzado del modelo adaptado con la data de la SBS incrementada: recall = 0,84
- Valor alcanzado en la adaptación del modelo con la metodología CSKT: recall = 0,65
- **Tamaño de la muestra:** 5000 Procesos de Contratación de personal
- **Prueba t** para comparar el recall de los dos valores.

Medias de las muestras \bar{x}_1, \bar{x}_2

Desviaciones estándar de las muestras S_1, S_2

Tamaño de las muestras n_1, n_2

- $\bar{x}_1=0.84$

- $\bar{x}_2=0.65$
- $S_1=S_2=\sqrt{\frac{0.84-(1-0.84)}{5000}}=\sqrt{\frac{0.68}{5000}}=0.0117$

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

$$t=\frac{0.84-0.65}{\sqrt{\frac{0.0117^2}{5000}+\frac{0.0117^2}{5000}}}$$

$$t=\frac{0.19}{\sqrt{0.000000027378+0.000000027378}}$$

$$t=\frac{0.19}{0.000234}$$

$$t=811,97$$

- El grado de libertad se calculan con n-1 para cada grupo.

$$gl = 5000 - 1 = 4999$$

Se hace uso de la tabla t de Student o programas estadísticos para determinar el valor crítico correspondiente al $\alpha = 0.05$ con 4999 gl.

Figura 78

Tabla de Student - Hipótesis específica 1

80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9852	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2903	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

$$t = 1.65$$

Comparando el valor de $t(811,97)$ con el valor crítico (1.65), $t >$ el valor crítico. Por lo tanto, se rechaza la H_0 .

Se concluye afirmando que la metodología CSKT aplicada en el modelo de ML **mejora significativamente** la eficiencia en términos de tiempo y los costos en el proceso de contratación de personal.

Prueba de HE2

HE2: El prototipo de Machine Learning evalúa de manera precisa los criterios de aceptación basándose en la categorización del proceso de contratación de personal.

H_0 = El prototipo de Machine Learning **no evalúa de manera precisa** los criterios de aceptación basándose en la categorización del proceso de contratación de personal.

H_1 = El prototipo de Machine Learning **evalúa de manera precisa** los criterios de aceptación basándose en la categorización del proceso de contratación de personal.

Procedimiento de la prueba de hipótesis:

Se requiere:

- Una medida justa para evaluar de manera precisa los criterios de aceptación basándose en la categorización del proceso de contratación de personal, como la Precisión.
- Un **dataset relevante** con y sin la adaptación del modelo para entrenar y probar la data con la metodología CSKT y la data de la SBS incrementada.
- Debido a que se debe comparar el rendimiento de la metodología CSKT aplicada en el modelo de Machine Learning, se puede utilizar:
 - La **prueba t** si se comparan las **medias de la métrica de evaluación** entre los métodos.

- La prueba de hipótesis para proporciones si estamos comparando criterios de aceptación basados en la categorización del proceso de contratación.
- Seleccionar el **nivel sig.**, generalmente 0.05.
- Utilizar los **datos recopilados** para calcular la **estadística de prueba** (diferencia de medias o diferencia de proporciones) y determinar si es estadísticamente significativa.
- Si el **valor p** es inferior al **nivel sig.**, rechazamos la **H_a** y se concluye que hay evidencia suficiente para respaldar la **hipótesis H₁**.
- Las limitaciones del estudio de:
 - **sesgos en los datos,**
 - **tamaño del conjunto de datos,**
 - **representatividad de las muestras, etc.**

Prueba de hipótesis:

H₁ = El prototipo de Machine Learning **evalúa de manera precisa** los criterios de aceptación basándose en la categorización del proceso de contratación de personal.

- Valor alcanzado en la adaptación del modelo con la metodología CSKT: Precisión=0,70
- Valor alcanzado del modelo adaptado con la data de la SBS incrementada: Precisión=0,91
- **Tamaño de la muestra:** 5000 Procesos de Contratación de personal
- **Prueba t** para comparar la precisión de los dos valores.

Medias de las muestras \bar{x}_1, \bar{x}_2

Desviaciones estándar de las muestras S_1, S_2

Tamaño de las muestras n_1, n_2

- $\bar{x}_1=0.91$
- $\bar{x}_2=0.70$

$$\bullet S_1=S_2=\sqrt{\frac{0.91-(1-0.91)}{5000}}=\sqrt{\frac{0.82}{5000}}=0.0128$$

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

$$t=\frac{0.91-0.70}{\sqrt{\frac{0.0128^2}{5000}+\frac{0.0128^2}{5000}}}$$

$$t=\frac{0.21}{\sqrt{0.000000032768+0.000000032768}}$$

$$t=\frac{0.21}{0.000256}$$

$$t=820.31$$

- El grado de libertad se calculan con n-1 para cada grupo.

$$gl = 5000 - 1 = 4999$$

Se consulta la tabla t de Student o software estadístico para encontrar el valor crítico respecto al α de 0.05 y 4999 gl.

Figura 79

Tabla de Student - Hipótesis específica 2

80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9852	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2903	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

$$t = 1.65$$

Comparando el valor de $t(820.31)$ con el valor crítico (1.65), $t >$ el valor crítico. Por lo tanto, se rechaza la H_0 .

Se afirmando que el prototipo de Machine Learning **evalúa de manera precisa** los criterios de aceptación basándose en la categorización del proceso de contratación de personal.

Prueba de HE3

HE3: La tasa de precisión del modelo de Machine Learning con la metodología CSKT optimiza significativamente el proceso de contratación en el sector privado.

H_0 = La tasa de precisión del modelo de Machine Learning con la metodología CSKT **no optimiza significativamente** el proceso de contratación en el sector privado.

H_1 = La tasa de precisión del modelo de Machine Learning con la metodología CSKT **optimiza significativamente** el proceso de contratación en el sector privado.

Procedimiento de la prueba de hipótesis:

Se requiere:

- Una medida justa para evaluar el rendimiento de la metodología CSKT aplicada en el modelo de Machine Learning en la eficiencia de la tasa de precisión en el proceso de contratación de personal, como el F1-Score.
- Un **dataset relevante** con y sin la adaptación del modelo para entrenar y probar la data con la metodología CSKT y la data de la SBS incrementada.
- Debido a que se debe comparar el rendimiento de la metodología CSKT aplicada en el modelo de Machine Learning, se puede utilizar:
 - La **prueba t** si se comparan las **medias de la métrica de evaluación** entre los métodos.
 - La prueba de hipótesis para proporciones si estamos comparando tasas de precisión del modelo de Machine Learning con la metodología CSKT.

- Seleccionar el **nivel sig.**, generalmente 0.05.
- Utilizar los **datos recopilados** para calcular la **estadística de prueba** (diferencia de medias o diferencia de proporciones) y determinar si es estadísticamente significativa.
- Sea el caso de **p** inferior que el **nivel de sig.**, rechazamos la H_0 y se concluye respaldando la **hipótesis H1**.
- Las limitaciones del estudio de:
 - **sesgos en los datos,**
 - **tamaño del conjunto de datos,**
 - **representatividad de las muestras, etc.**

Prueba de hipótesis:

H_1 = La tasa de precisión del modelo de Machine Learning con la metodología CSKT **optimiza significativamente** el proceso de contratación en el sector privado.

- Valor alcanzado del modelo adaptado con la data de la SBS incrementada: F1-Score=0,87
- Valor alcanzado en la adaptación del modelo con la metodología CSKT: F1-Score=0,76
- **Tamaño de la muestra:** 5000 Procesos de Contratación de personal
- **Prueba t** para comparar el F1-Score de los dos valores.

Medias de las muestras \bar{x}_1, \bar{x}_2

Desviaciones estándar de las muestras S_1, S_2

Tamaño de las muestras n_1, n_2

- $\bar{x}_1=0.87$
- $\bar{x}_2=0.76$
- $S_1=S_2=\sqrt{\frac{0.87-(1-0.87)}{5000}} = \sqrt{\frac{0.74}{5000}} = 0.0122$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$t = \frac{0.87 - 0.76}{\sqrt{\frac{0.0122^2}{5000} + \frac{0.0122^2}{5000}}}$$

$$t = \frac{0.11}{\sqrt{0.000000029768 + 0.000000029768}}$$

$$t = \frac{0.11}{0.000234}$$

$$t = 470,09$$

- El grado de libertad se calculan con n-1 para cada grupo.

$$gl = 5000 - 1 = 4999$$

Se consulta la tabla t de Student o software estadístico para encontrar el valor crítico respecto al nivel de sig. de 0.05 y 4999 gl.

Figura 80

Tabla de Student - Hipótesis específica 3

80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9852	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2903	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

$$t = 1.65$$

Comparando el valor de t(470,09) con el valor crítico (1.65), $t >$ el valor crítico. Por lo tanto, se rechaza la H_0 .

Se concluye afirmando que la tasa de precisión del modelo de Machine Learning con la metodología CSKT **optimiza significativamente** el proceso de contratación en el sector privado.

Prueba de la HE4

HE4: La eficiencia del proceso de contratación de personal mejora con la utilización del modelo de Machine Learning en la aplicación móvil.

H_0 = La eficiencia del proceso de contratación de personal **no mejora significativamente** con la utilización del modelo de Machine Learning en la aplicación móvil.

H_1 = La eficiencia del proceso de contratación de personal **mejora significativamente** con la utilización del modelo de Machine Learning en la aplicación móvil.

Procedimiento de la prueba de hipótesis:

Se requiere:

- Una medida justa para evaluar el rendimiento de la metodología CSKT aplicada en el modelo de ML en la eficiencia de la utilización del modelo de Machine Learning en la aplicación móvil, como el Recall.
- Un **dataset relevante** con y sin la adaptación del modelo para entrenar y probar la data con la metodología CSKT y la data de la SBS incrementada.
- Debido a que se debe comparar el rendimiento de la metodología CSKT aplicada en el modelo de Machine Learning, se puede utilizar:
 - La **prueba t** si se comparan las **medias de la métrica de evaluación** entre los métodos.
 - La prueba de hipótesis para proporciones si estamos comparando tasas de la eficiencia del proceso de contratación de personal.
- Seleccionar el **nivel sig.**, generalmente 0.05.

- Utilizar los **datos recopilados** para calcular la **estadística de prueba** (diferencia de medias o diferencia de proporciones) y determinar si es estadísticamente significativa.
- Sea el caso de **p** inferior L **nivel de sig.**, rechazamos la H_0 y se concluye respaldando a la **hipótesis H_1** .
- Las limitaciones del estudio de:
 - **sesgos en los datos,**
 - **tamaño del conjunto de datos,**
 - **representatividad de las muestras, etc.**

Prueba de hipótesis:

H_1 = La eficiencia del proceso de contratación de personal **mejora significativamente** con la utilización del modelo de Machine Learning en la aplicación móvil.

- Valor alcanzado del modelo adaptado con la data de la SBS incrementada: recall = 0,84
- Valor alcanzado en la adaptación del modelo con la metodología CSKT: recall = 0,65
- **Tamaño de la muestra:** 5000 Procesos de Contratación de personal
- **Prueba t** para comparar el recall de los dos valores.

Medias de las muestras \bar{x}_1, \bar{x}_2

Desviaciones estándar de las muestras S_1, S_2

Tamaño de las muestras n_1, n_2

- $\bar{x}_1=0.84$
- $\bar{x}_2=0.65$
- $S_1=S_2=\sqrt{\frac{0.84-(1-0.84)}{5000}} = \sqrt{\frac{0.68}{5000}} = 0.0117$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$t = \frac{0.84 - 0.65}{\sqrt{\frac{0.0117^2}{5000} + \frac{0.0117^2}{5000}}}$$

$$t = \frac{0.19}{\sqrt{0.000000027378 + 0.000000027378}}$$

$$t = \frac{0.19}{0.000234}$$

$$t = 811,97$$

- El grado de libertad se calculan con n-1 para cada grupo.

$$gl = 5000 - 1 = 4999$$

Se consulta la tabla t de Student o software estadístico para encontrar el valor crítico respecto al nivel sig. de 0.05 y 4999 gl.

Figura 81

Tabla de Student - Hipótesis específica 4

80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9852	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2903	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

$$t = 1.65$$

Comparando el valor de t(811,97) con el valor crítico (1.65), $t >$ el valor crítico. Por lo tanto, se rechaza la H_0 .

Se concluye afirmando sobre la eficiencia en la contratación de personal **mejora significativamente** con la utilización del modelo de Machine Learning en la aplicación móvil.

Prueba de la HE5

HE5: La implementación práctica de la aplicación móvil con Machine Learning y la metodología CSKT es efectiva en la optimización del proceso de contratación de personal en el sector privado.

H_0 = La implementación práctica de la aplicación móvil con Machine Learning y la metodología CSKT **no es efectiva** en la optimización del proceso de contratación de personal en el sector privado.

H_1 = La implementación práctica de la aplicación móvil con Machine Learning y la metodología CSKT **es efectiva** en la optimización del proceso de contratación de personal en el sector privado.

Procedimiento de la prueba de hipótesis:

Se requiere:

- Una medida justa para evaluar el rendimiento de la metodología CSKT aplicada en el modelo de Machine Learning en la eficiencia de la implementación práctica de la aplicación móvil en el proceso de contratación de personal, como el roc_auc.
- Un **dataset relevante** con y sin la adaptación del modelo para entrenar y probar la data con la metodología CSKT y la data de la SBS incrementada.
- Debido a que se debe comparar el rendimiento de la metodología CSKT aplicada en el modelo de Machine Learning, se puede utilizar:
 - La **prueba t** si se comparan las **medias de la métrica de evaluación** entre los métodos.
 - La prueba de hipótesis para proporciones si estamos comparando tasas de eficiencia en la optimización del proceso de contratación de personal.
- Seleccionar el **nivel sig.**, generalmente 0.05.

- Utilizar los **datos recopilados** para calcular la **estadística de prueba** (diferencia de medias o diferencia de proporciones) y determinar si es estadísticamente significativa.
- Sea el caso de p inferior L nivel de sig., rechazamos la H_0 y se concluye respaldando a la hipótesis H_1 .
- Las limitaciones del estudio de:
 - **sesgos en los datos,**
 - **tamaño del conjunto de datos,**
 - **representatividad de las muestras, etc.**

Prueba de hipótesis:

H_1 = La implementación práctica de la aplicación móvil con Machine Learning y la metodología CSKT es **efectiva** en la optimización del proceso de contratación de personal en el sector privado.

- Valor alcanzado del modelo adaptado con la data de la SBS incrementada: $roc_auc = 0,85$
- Valor alcanzado en la adaptación del modelo con la metodología CSKT: $roc_auc = 0,74$
- **Tamaño de la muestra:** 5000 Procesos de Contratación de personal
- **Prueba t** para comparar el roc_auc de los dos valores.

Medias de las muestras \bar{x}_1, \bar{x}_2

Desviaciones estándar de las muestras S_1, S_2

Tamaño de las muestras n_1, n_2

- $\bar{x}_1 = 0.85$
- $\bar{x}_2 = 0.74$
- $S_1 = S_2 = \sqrt{\frac{0.85 - (1 - 0.85)}{5000}} = \sqrt{\frac{0.70}{5000}} = 0.0118$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$t = \frac{0.85 - 0.74}{\sqrt{\frac{0.0118^2}{5000} + \frac{0.0118^2}{5000}}}$$

$$t = \frac{0.11}{\sqrt{0.000000027848 + 0.000000027848}}$$

$$t = \frac{0.19}{0.000236}$$

$$t = 805,08$$

- El grado de libertad se calculan con n-1 para cada grupo.

$$gl = 5000 - 1 = 4999$$

Se consulta la tabla t de Student o software estadístico para encontrar el valor crítico respecto al nivel de sig. de 0.05 y 4999 gl.

Figura 82

Tabla de Student - Hipótesis específica 5

80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9852	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2903	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

$$t = 1.65$$

Comparando el valor de t(805,08) con el valor crítico (1.65), $t >$ el valor crítico. Por lo tanto, se rechaza la H_0 .

Se concluye afirmando que la implementación práctica de la aplicación móvil con Machine Learning y la metodología CSKT es efectiva en la optimización del proceso de contratación de personal en el sector privado.

V. DISCUSIÓN DE RESULTADOS

La implementación de la metodología CSKT en el modelo de Machine Learning presenta una mejora notable en la eficiencia del proceso de contratación de personal en base a la gestión efectiva de los recursos con una mejora del 96%. Esta cifra es relevante al ser comparada con diversos estudios, siendo superior a los resultados de Coronel (2021), el cual destaca cómo el uso de Machine Learning en el proceso de selección de personal administrativo ha generado mejoras significativas, su enfoque se centra en la adaptación de algoritmos para establecer modelos basados en el autoaprendizaje, lo que ha resultado un aumento del índice de postulantes y una mejora en la evaluación curricular y contratación de un 40,3%. Además, es superior a los resultados de Yaranga (2022) el cual se enfoca en el proceso de selección de personal docente, donde el uso de Machine Learning ocasiona un incremento sustancial del 26% en el índice de postulados y 50% en el índice de contratación. Estas cifras muestran cómo el uso de Machine Learning ha generado mejoras notables en el proceso de selección de personal, de manera similar a lo encontrado en este estudio. Asimismo, es superior a los resultados de Dávila (2019) el cual se destaca por su enfoque en la Data Mining para mejorar los procesos de reclutamiento de personal utilizándolo como herramienta de apoyo, además, logró determinar mejores candidatos a través del análisis de perfiles de puesto y reglas de decisión con una mejora del 80%, la cual respalda la eficiencia del proceso de contratación. Realizando la comparación, la metodología CSKT sobresale en su contribución al proceso de contratación de personal, logrando una mejora del 96% en la eficiencia, superando notablemente los resultados de los autores citados en términos de aumento del índice de postulantes y contratación. A diferencia del enfoque centrado en autoaprendizaje de Coronel y el énfasis en la selección docente de Yaranga, la metodología CSKT demuestra una gestión más efectiva de recursos. Además, frente al enfoque de Dávila en Data Mining, esta metodología ofrece una mejora significativa sin comprometer la precisión en la identificación

de candidatos, consolidando así su papel destacado en la optimización de procesos de contratación a nivel nacional e internacional.

La implementación del prototipo de Machine Learning evalúa de manera precisa los criterios de aceptación basándose en la categorización del proceso de contratación de personal con un incremento del 29,85% indicando un hallazgo significativo en la evaluación de los criterios de aceptación que se puede comparar con otros antecedentes para evaluar su impacto y eficacia en el contexto en la contratación de personal. Esta cifra puede ser comparada con otros antecedentes para evaluar su impacto y eficacia en el contexto de la contratación de personal, siendo superior a los resultados de Espinoza (2020) el cual resalta cómo la aplicación de métodos y recursos de aprendizaje automático ha posibilitado un pronóstico predictivo para estimar la valoración del capital acumulado del miembro en el sistema de jubilación privado con una mejora del 0,65%. Esta investigación muestra cómo el ML puede aplicarse de manera efectiva para resolver problemas complejos y proporcionar predicciones precisas, lo que sugiere que también podría ser eficaz en la evaluación de criterios de aceptación en procesos de contratación de personal. Por otro lado, es inferior a los resultados de Alkhazraji y Buhaliba (2020) el cual se centra en el uso de software de aprendizaje automático en la contratación de recursos humanos para los candidatos de la Academia de Policía de Dubai con una mejora del índice de precisión del 90%, ello resalta la capacidad de Machine Learning para mejorar la calidad de la contratación y ahorrar tiempo a los reclutadores. Luego de la comparativa, se obtiene que la metodología CSKT se diferencia al centrarse en la evaluación precisa de los criterios de aceptación en el proceso de contratación de personal, logrando un incremento del 29,85%. A diferencia de los enfoques de los autores citados, la metodología CSKT ofrece una herramienta específica y eficaz para optimizar la evaluación de candidatos en el contexto laboral.

El resultado de la tasa de precisión del modelo de Machine Learning aplicando la metodología CSKT optimiza significativamente el proceso de contratación en el sector privado con una mejora del 90,86%; ello es un hallazgo importante que se puede comparar con otros antecedentes para evaluar su influencia en el impacto y eficacia en la optimización del proceso de contratación, siendo superior a los resultados de Armas (2021) el cual destaca cómo la inteligencia artificial está siendo adoptada por empresas peruanas para mejorar la productividad y rentabilidad, el cual tuvo una mejora del índice de facilidad de operación del 83%, y en el índice de eficiencia de ejecución del 76%, resultando que el modelo es más preciso en sus predicciones, y que las operaciones sean más fluidas y eficientes. Por otra parte, es inferior a los resultados de Berón et al. (2021) los cuales se enfocan en las principales causas de ausentismo laboral proponiendo el uso de técnicas de Data Mining para predecir el motivo de ausentismo el cual tuvo una mejora del 94,72% en el índice de efectividad. Posterior a la comparación, se afirma que la metodología CSKT sobresale al lograr una mejora del 90,86% en la tasa de precisión del modelo de ML, optimizando significativamente el proceso de contratación en el sector privado. En contraste con los enfoques de los autores mencionados, la metodología CSKT se destaca por su precisión en las predicciones específicas relacionadas con la contratación, lo que permite una evaluación más eficaz de los candidatos y una optimización del proceso de selección.

El modelo de Machine Learning en la aplicación móvil optimiza la eficiencia del proceso de contratación de personal con una mejora del 25,09%, siendo un hallazgo crucial que puede ser comparado con otros antecedentes para evaluar su relevancia y eficacia en la mejora de procesos similares, siendo inferior a los resultados de Asto (2020) el cual destaca cómo la implementación de un modelo de sistemas ML para el reclutamiento y selección de talento humano ha mejorado la preparación de los candidatos y ha brindado un buen soporte para el proceso de contratación de personal con una mejora del 83% en el índice de facilidad

de operación y una mejora del 76% en el índice de eficiencia de ejecución. Aunque el enfoque no es exactamente el mismo, este estudio proporciona evidencia de que el uso de tecnologías móviles y Machine Learning puede mejorar la eficiencia en el ámbito del reclutamiento y la selección de personal. Asimismo, es inferior a los resultados de Dávila (2019) el cual se enfoca en la extracción de datos para mejorar el proceso de selección de empleados con una mejora del 80%. Comparando los resultados, se observa que todos comparten el objetivo de reforzar la eficiencia en el proceso que forma parte del objeto de estudio mediante la utilización de tecnologías innovadoras. No obstante, dada la cifra porcentual de mejora del modelo de la investigación presente es inferior a lo hallado por los autores, la contribución de la metodología CSKT radica en su enfoque específico en la optimización del proceso de contratación a través de la aplicación móvil y el ML, lo que proporciona una herramienta valiosa para mejorar la eficiencia en este ámbito laboral.

La implementación de la aplicación móvil con Machine Learning y la metodología CSKT es efectiva en la optimización del proceso de contratación de personal en el sector privado, con una mejora del 93%, es un hallazgo importante que puede ser comparado con otros antecedentes para evaluar su relevancia y eficacia en la mejora de procesos similares, siendo superior a los resultados de Coronel (2021) el cual destaca cómo el uso de Machine Learning en el proceso de selección de personal administrativo ha generado mejoras significativas, su enfoque se centra en la adaptación de algoritmos para establecer modelos basados en el autoaprendizaje, lo que ha resultado en un aumento del índice de postulantes del 40,3%, el cual resalta la efectividad del Machine Learning en mejorar la eficiencia del proceso de contratación en el sector privado. Además, es superior a los resultados de Matute et al. (2021) los cuales se enfocan en el uso de software de aprendizaje automático en el proceso de contratación de recursos humanos con una mejora en la precisión del 80,5%. Comparando, resulta que la metodología CSKT contribuye significativamente con una mejora 93% en la

optimización del proceso de contratación de personal; la distinción de CSKT radica en su enfoque específico y eficaz, lo que respalda su relevancia y eficacia en la mejora de procesos similares en el ámbito laboral.

VI. CONCLUSIONES

- 6.1. La eficacia de la metodología CSKT aplicada en modelos de Machine Learning posibilita la mejora de la eficiencia en el proceso de contratación de personal, asimismo, otros procesos: monitoreo y/o mantenimiento de los modelos de ML, seguimiento del rendimiento del modelo en producción, evaluaciones de impacto comercial, generación y/ o análisis de informes, análisis de riesgos, entre otros. Asimismo, para lograr una mayor reducción significativa en los errores de asignación en pro de la efectividad de los modelos, es necesario seleccionar la data representativa y de alta calidad; realización de un preprocesamiento exhaustivo de la data; utilización de técnicas de validación para optimizar el rendimiento del modelo de ML, por ejemplo: la validación cruzada; evaluación del modelo mediante métricas sólidas como precisión, recall, F1-Score y AUC-ROC; análisis y ajustes del modelo; utilización de técnicas de validación cruzada. La implementación exitosa de aplicaciones móviles es aplicable sobre la infraestructura de Firebase, Kubernetes, Google Cloud Platform, Amazon Web Services, IBM Watson, Microsoft Azure, entre otros. Además, la metodología permite simplificar el desarrollo y mejorar la eficiencia otras actividades en la gestión de recursos y procesos de contratación, tales como: filtrado automático de currículums, predicción de la retención de empleados, automatización de tareas administrativas y el análisis de perfiles de candidatos en redes sociales. Asimismo, posibilita la mejora de los procesos empresariales, tales como: gestión de inventarios o cadena de suministro, análisis de marketing o ventas, personalización de productos o servicios y la toma de decisiones estratégicas.
- 6.2. La eficacia de la metodología CSKT aplicada en el modelo de Machine Learning mejora la eficiencia en términos de tiempo y costos en el proceso de contratación de personal, además, otros procesos beneficiados por la metodología resultan:

automatización de tareas repetitivas, análisis predictivos y detección de anomalías. Demostrando una reducción en los errores de asignación del seguro de pensiones, respaldando la efectividad del modelo con una cantidad total de 129 falsos negativos y 215 falsos positivos, asimismo, presenta 561 verdaderos positivos y 401 verdaderos negativos al momento de su entrenamiento, cabe recalcar que, la efectividad mejora si se realiza una optimización continua de los algoritmos de ML mediante el ajuste de parámetros; integración con sistemas existentes en la organización previa evaluación, capacitación del personal. Además, la aplicación de la metodología CSKT permitió determinar la asignación más adecuada del seguro de pensiones, asimismo, otros de los procesos organizacionales que se ven beneficiados son: gestión financiera, gestión de la cadena de suministro, servicio al cliente y la planificación de actividades.

- 6.3. Asimismo, el estudio ha demostrado la efectividad del modelo de Machine Learning basado en la metodología CSKT para clasificar seguros de pensiones. Los resultados revelan una precisión del 69.98%, un recall del 65.10%, un F1-Score del 75.66%, y un AUC-ROC del 73.98%, estos resultados se pueden mejorar con el procesamiento de los datos, selección de características relevantes, optimización de hiperparámetros, aumento del tamaño del conjunto de entrenamiento, técnicas de validación cruzada. Estas métricas respaldan la capacidad del modelo para tomar decisiones precisas en la asignación de seguros de pensiones, independientemente del salario inicial de los individuos. Además, se identificaron áreas de mejora en la limpieza de datos y el consumo de Web Services para optimizar la precisión del modelo. En conjunto, estos hallazgos indican la importancia de los datos de entrada en la toma de decisiones del modelo y su impacto en la planificación financiera personal, brindando así una herramienta valiosa para la gestión de pensiones.

- 6.4. El estudio ha validado la eficacia del modelo de ML propuesto, evidenciado por métricas como F1-Score, precisión, recall y ROC que superan los criterios de aceptación establecidos. La limpieza de datos se ha revelado como un paso crucial para garantizar la fiabilidad y precisión de la información utilizada en el modelo, mediante la corrección de errores, eliminación de duplicados y estandarización de formatos. Aunque se haya reducido el conjunto de datos de 50,000 a 5,000 registros para mantener la calidad y coherencia de los datos, esta acción ha demostrado ser esencial para optimizar el rendimiento del modelo, dado que, al disminuir la cantidad de registros ocasionó las siguientes ventajas: menor complejidad computacional para el modelo, reducción de realizar varios sobreajustes y la mejora general de los datos utilizados para el entrenamiento. Asimismo, la gestión eficaz de datos desempeña un papel fundamental en el éxito de los proyectos de ML tales como: proyectos de clasificación con una data histórica o en proyectos de predicción especializados en áreas específicas de la organización; respaldando la toma de decisiones estratégicas y la obtención de resultados precisos.
- 6.5. Por otro lado, el estudio utilizó datos organizados en formato JSON y una aplicación móvil para evaluar la precisión de un modelo de proyección. En el entorno de pruebas de Android Studio, se configuró la evaluación del modelo, y se envió información de usuarios para medir la eficiencia operativa. Los resultados mostraron una precisión inicial de 0.7129, que mejoró a 0.8918 después de implementar las medidas de optimización: aumento del tamaño del conjunto de datos para el entrenamiento, optimización de hiperparámetros, el refinamiento del preprocesamiento de datos y la optimización en tiempo de ejecución. Estos hallazgos destacan la capacidad de mejorar la precisión mediante ajustes en la eficiencia operativa, lo que subraya la importancia de la evaluación continua y la optimización en el desarrollo de modelos

predictivos, asimismo, es aplicable a otros procesos, tales como: gestión de activos físicos, optimización de precios, detección de fraudes, entre otros.

- 6.6. No obstante, se presenta una aplicación móvil desarrollada sobre la infraestructura de Firebase, destacando su capacidad para simplificar el desarrollo y proporcionar una variedad de servicios integrados. La implementación exitosa de la aplicación, demostrada por el lanzamiento en Google Play en un corto período de tiempo, refleja su eficacia en la gestión de recursos y el proceso de contratación. Las encuestas realizadas muestran una alta satisfacción de los usuarios en términos de eficiencia operativa, experiencia de usuario y calidad del soporte proporcionado. Estos hallazgos respaldan la conclusión de que la aplicación ha cumplido con éxito su objetivo de proporcionar una solución efectiva y satisfactoria para las necesidades de contratación de personal.

VII. RECOMENDACIONES

- 7.1. Explorar técnicas avanzadas de preprocesamiento de datos dado que la limpieza de datos se ha identificado como un paso crucial para garantizar la fiabilidad y precisión de los modelos de ML, por ende, se sugiere investigar y aplicar técnicas más avanzadas de limpieza: eliminación de duplicados, corrección tipográfica, gestión de valores atípicos, imputación de data faltante, normalización o estandarización, codificación de variables categóricas, entre otros; y preprocesamiento de datos: reducción de la dimensionalidad de variables, discretización de variables, balanceo de datos mediante la generación de data sintéticas en la capacidad del modelo para lograr la predicción de clases minoritarias y la transformación de características, esto podría incluir la identificación y corrección automática de errores a través de la reducción de registros logrando una menor cantidad de reajustes ahorrando tiempo, así como la detección y manejo de datos faltantes de manera más efectiva mediante la normalización u otras técnicas de limpieza y preprocesamiento como anteriormente se ha mencionado.
- 7.2. Investigar modelos de Machine Learning más avanzadas, dado que, aunque los modelos basados en la metodología CSKT han demostrado ser efectivos, aún se debe explorar y comparar con otros modelos más avanzados, tales como: 3D CNN, Siamese Networks with Attention, Transformer Networks with Memory, Deep Bayesian Neural Networks, entre otros. Estos modelos podrían ofrecer una mayor capacidad de generalización y precisión en la asignación de seguros de pensiones.
- 7.3. Optimizar el consumo de web services, dado que se identificaron áreas de mejora en el consumo de servicios web, por ello, se sugiere investigar y aplicar técnicas de optimización para mejorar la eficiencia y velocidad de acceso a estos servicios, esto podría incluir el uso de técnicas de almacenamiento en caché (caché distribuida o

de fragmentos), paralelización de solicitudes (indexación) y optimización de consultas.

- 7.4. Implementar técnicas de evaluación continua y optimización para mejorar la precisión de los modelos de Machine Learning en el tiempo, asimismo se sugiere implementar técnicas de evaluación continua y optimización, ya que esto podría incluir la monitorización regular del rendimiento del modelo en producción, la identificación de posibles áreas de mejora, la actualización periódica del modelo con nuevos datos y técnicas, tales como: regularización L1 y Ensemble Learning.
- 7.5. Explorar nuevas tecnologías, plataformas de desarrollo, lenguajes de programación como: Python, Java, Dart, Kotlin, entre otros; frameworks o librerías: TensorFlow, PyTorch, Android SDK, Flutter, entre otros; plataformas en la nube: Amazon Web Services, Google Cloud Platform o Microsoft Azure. Puesto que, aunque la infraestructura de Firebase ha demostrado ser efectiva, se debe explorar y evaluar otras tecnologías y plataformas de desarrollo para aplicaciones móviles, entre ellos se encuentran: Android Studio, React Native, Flutter e Ionic; en los cuales se podrían ofrecer una mayor flexibilidad y rendimiento durante el proceso de desarrollo.
- 7.6. Realizar un estudio más exhaustivo sobre la satisfacción del usuario, aunque las encuestas realizadas mostraron una alta satisfacción de los usuarios, se puede realizar un estudio más exhaustivo para obtener mejores resultados, para ello, se sugiere la utilización de otras herramientas, tales como: análisis de comentarios, análisis de informes internos, grupos de discusión o entrevistas, análisis de uso de datos, Net Promoter Score, Customer Satisfaction Score, Consistency Score, Customer Loyalty Index, First Contact Resolution, etc; incluyendo una mayor diversidad de usuarios y escenarios de uso: asignación de recursos financieros,

planificación financiera personal, evaluación de riesgos financieros, asesoramiento financiero automatizado, prevención del fraude financiero, entre otros; esto podría proporcionar información adicional sobre las necesidades y expectativas de los usuarios, así como áreas específicas de mejora para la aplicación. Asimismo, se sugiere también realizar análisis comparativos con otras metodologías similares, tales como CRISP-DM (Cross-Industry Standard Process for Data Mining), SEMMA (Sample, Explore, Modify, Model, Assess), TDSP (Team Data Science Process), KDD (Knowledge Discovery in Databases), etc; y evaluar el impacto de la metodología en los diferentes sectores industriales anteriormente mencionados. Estos enfoques podrían ampliar el conocimiento sobre la efectividad y aplicabilidad de la metodología CSKT.

VIII. REFERENCIAS

- Alkan, E., Cushen-Brewster, N. and Anyanwu, P. (2024). Investigating Healthcare Workforce Recruitment and Retention: A Mixed-Methods Study Protocol. *BMJ Open*, 14, (2). <https://doi.org/10.1136/bmjopen-2023-074100>.
- Alkhozraji, I., & Buhaliba, A. (2020). *Using machine learning software in the human resource recruiting process for candidates from Dubai Police Academy*. [Tesis de doctorado publicada] Rochester Institute of Technology.
- Anayochukwu, A., & Anayochukwu, G. (2021). Aplicación móvillicación of due process in recruitment and selection of personnel into the federal civil service. Part 1. *RUDN Journal of Public Administration*, 8(4), 375–385.
- Arias, C., Val, M. P. R., Fernández, L., Sánchez, L., Díaz, A. & Ramírez, M. (2020). Validación del poder predictivo de los Grupos de Morbilidad Ajustada (GMA) respecto de otras herramientas de estratificación de la población. *Revista española de salud pública*, (94), 75.
- Armas, C. (2021). La Inteligencia Artificial en empresas peruanas e impactos laborales en los trabajadores. *Iberoamerican business journal*, 5(1), 83–105. <https://doi.org/10.22451/5817.ibj2021.vol5.1.11053>
- Asto, P. (2020). *Modelo de sistemas M-Learning para el reclutamiento y selección de talento humano / caso: entrevistas*. [Tesis de pregrado publicada] Universidad Nacional de San Agustín de Arequipa.
- Berón, E., Mejía, D., & Castrillón, O. (2021). Principales causas de ausentismo laboral: una aplicación desde la minería de datos. *CIT Informacion Tecnologica*, 32(2), 11–18. <https://doi.org/10.4067/s0718-07642021000200011>

- Borisova, A., Rakhimberdinova, M., Madiyarova, E., Riazantseva, I., & Mikidenko, N. (2020). Staffing search and recruitment of personnel on the basis of artificial intelligence technologies. *Journal of Entrepreneurship and Sustainability Issues*, 7(3), 2456–2469. [https://doi.org/10.9770/jesi.2020.7.3\(66\)](https://doi.org/10.9770/jesi.2020.7.3(66))
- Borja-Robalino, R., Monleon-Getino, A., & Rodellar, J. (2020). Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E30), 184-196.
- Bozzi, M. (2023). *Modelos de machine learning aplicados a la estimación de la probabilidad de default de entidades bancarias del sistema bancario argentino*. Universidad de San Andrés. Escuela de Negocios. <http://hdl.handle.net/10908/23089>
- Butka, P., Bednár, P., & Ivančáková, J. (2020). Methodologies for knowledge discovery processes in context of AstroGeoInformatics. *En Knowledge Discovery in Big Data from Astronomy and Earth Observation* (pp. 1–20). Elsevier.
- Campesonato, O. (2020). *Artificial intelligence, machine learning, and deep learning*. Mercury Learning & Information.
- Chinnamgari, S. (2019). *R Machine Learning Projects: Implement Supervised, Unsupervised, and Reinforcement Learning Techniques Using R 3. 5*. 1st edition. Birmingham: Packt Publishing Ltd, 2019. Print.
- Coronel, E. (2021). *Machine learning en la mejora del proceso de selección del personal administrativo de la Corte Superior de Justicia de Lima, 2020*. [Tesis de maestría publicada] Universidad César Vallejo. <https://hdl.handle.net/20.500.12692/61903>
- Costa, C., & Aparicio, J. (2020). POST-DS: A methodology to boost data science. *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6.

- Cruz, E., González, M., & Rangel, J. C. (2022). Técnicas de machine learning aplicadas a la evaluación del rendimiento y a la predicción de la deserción de estudiantes universitarios, una revisión. *Prisma Tecnológico*, 13(1), 77–87. <https://doi.org/10.33412/pri.v13.1.3039>
- Dávila, C. (2019). *Minería de datos a partir de la WEB para la mejora del proceso de reclutamiento de personal*. [Tesis de pregrado publicada]. Benemérita Universidad Autónoma de Puebla. <https://hdl.handle.net/20.500.12371/15362>
- Dhabuwala, P., & Pitroda, J. (2021). Recruitment, Selection and Training of Human Resource in Construction: A Review. *RT&A e-Journal*, 60, 11–120.
- Espinoza, J. (2020). *Análisis predictivo para el cálculo de la valoración del fondo acumulado del afiliado en el Sistema Privado de Pensiones usando técnicas y herramientas de machine Learning*. [Tesis de pregrado publicada]. Universidad Peruana de Ciencias Aplicadas. <http://hdl.handle.net/10757/652592>
- Esteban, N. (2018). *Tipos de investigación*. Repositorio Institucional UDSG. <http://repositorio.usdg.edu.pe/handle/USDG/34>
- Estrada, M. L., Pérez, N. V. R., Ramírez, N. N. R., Crespo, E. R. O., & Rodríguez, J. A. A. (2020). Propuesta de un modelo de selección de capital humano utilizando aprendizaje automático a través de rna (proposal of a human capital selection model using automatic learning through rna). *Pistas Educativas*, 42(137). <https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/2312>
- Fajardo, J. (2023). Inteligencia artificial aplicada al proceso de selección de personal. *Polo del Conocimiento: Revista científico-profesional*, 8(9), 726-740. <https://dialnet.unirioja.es/servlet/articulo?codigo=9152551>

- Flores, P., Muñoz, L., & Sánchez, T. (2019). Estudio de potencia de pruebas de normalidad usando distribuciones desconocidas con distintos niveles de no normalidad. *Perfiles*, 1(21), 4–11. <https://doi.org/10.47187/perf.v1i21.42>
- Flores Tapia, C. E., & Flores Cevallos, K. L. (2021). Pruebas para comprobar la normalidad de datos en procesos productivos: anderson-darling, ryan-joiner, shapiro-wilk y kolmogórov-SMIRNOV. *Societas*, 23(2), 83–106. <https://matriculapre.up.ac.pa/index.php/societas/article/view/2302>
- Forero-Corba, W., & Bennasar, F. N. (2024). Técnicas y aplicaciones del Machine Learning e Inteligencia Artificial en educación: una revisión sistemática. *RIED Revista Iberoamericana de Educación a Distancia*, 27(1). <https://doi.org/10.5944/ried.27.1.37491>
- FraiJ, J., & László, V. (2021). A Literature Review: Artificial Intelligence Impact on the Recruitment Process. *International Journal of Engineering and Management Sciences*, 6(1), 108-119. <https://doi.org/10.21791/IJEMS.2021.1.10>
- Gestión. (2023, noviembre 3). Demanda de personal de seguridad aumentará 18% por campaña de fin de año. *Gestión*. <https://gestion.pe/economia/empresas/demanda-de-personal-de-seguridad-aumentara-18-por-campana-de-fin-de-ano-noticia/>
- Gil Rubio, R. (2022). *Modelos de machine learning para clasificar la cartera en un fondo de pensiones*. Universidad Santo Tomás. <http://hdl.handle.net/11634/47294>
- Gil-Vera, V. D., & Seguro-Gallego, C. (2022). Machine learning aplicado al análisis del rendimiento de desarrollos de software. *Revista Politécnica*, 18(35), 128–139. <https://doi.org/10.33571/rpolitec.v18n35a9>

- Greener, J., Kandathil, S., Moffat, L., & Jones, D. (2022). A guide to machine learning for biologists. *Nature Reviews. Molecular Cell Biology*, 23(1), 40–55. <https://doi.org/10.1038/s41580-021-00407-0>
- Guamán Chacha, K. A., Hernández Ramos, E. L., & Lloay Sánchez, S. I. (2020). El positivismo y el positivismo jurídico. *Revista Universidad y Sociedad*, 12(4), 265-269.
- Guamán, D. (2020). *Una revisión conceptual a la Marca Empleadora como generador estratégico de captación y retención de talentos*. [Tesis de maestría publicada]. Universidad Andina Simón Bolívar. <http://hdl.handle.net/10644/7765>
- Hanze, E. (2020). *El reclutamiento a través de redes sociales como nueva tendencia en la selección de personal y su aplicación en la empresa Tata Consultancy Services Ecuador*. [Tesis de maestría publicada]. Universidad Andina Simón Bolívar. <http://hdl.handle.net/10644/7873>
- Hart, G., Mueller, T., Toher, C., & Curtarolo, S. (2021). Machine learning for alloys. *Nature Reviews. Materials*, 6(8), 730–755. <https://doi.org/10.1038/s41578-021-00340-w>
- Hernández, R., & Mendoza, C. P. (2018). *Metodología de la investigación: las rutas: cuantitativa, cualitativa y mixta*. McGraw-Hill Educación. <http://repositorio.uasb.edu.bo:8080/handle/54000/1292>
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685–695. <https://doi.org/10.1007/s12525-021-00475-2>
- Jaramillo, M. (2020). *Diseño de un modelo de perfiles de competencias para la gestión de selección de personal en la empresa Química Suiza Industrial del Ecuador*. [Tesis de

maestría publicada]. Universidad Andina Simón Bolívar.
<http://hdl.handle.net/10644/7467>

Kwekha, A. S., Abduljabbar, H. N., & Alhayani, B. (2023). Coronavirus disease (COVID-19) cases analysis using machine-learning applications. *Applied Nanoscience*, *13*(3), 2013–2025. <https://doi.org/10.1007/s13204-021-01868-7>

Lu, X. (2024). Artificial Intelligence Tools for Optimising Recruitment and Retention in Clinical Trials: A Scoping Review Protocol. *BMJ Open*, *14*(3). <https://doi.org/10.1136/bmjopen-2023-080032>.

Mahesh, B. (2020). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, *9*, 381–386.

Martín Gómez, S. (2020). Aplicación de las Metodologías Ágiles al proceso de enseñanza-aprendizaje universitario. *Revista d'Innovació Docent Universitària*, 62–73. <https://doi.org/10.1344/ridu2020.12.7>

Matute, K., & Bojorque, R. (2021). Apoyo a los subsistemas de talento humano, selección y reclutamiento a partir de un sistema experto. Caso de estudio. *Ingenius*, *26*, 41–48. <https://doi.org/10.17163/ings.n26.2021.04>

Mendoza Romero, G. R., Quinde Obando, P. M., & Director, S. I. P. (2020). *Plan de Negocios de aplicación móvil con geolocalización de personal cualificado para el cuidado de niños*. ESPOL. FCSH. <https://www.dspace.espol.edu.ec/handle/123456789/53143>

Miranda De Freitas, K. D. (2020). *Desarrollo de una aplicación Web configurable para agilizar los procesos de selección y contratación de personal en las organizaciones*. [Tesis de pregrado]. Repositorio Universidad Central de Venezuela. <http://hdl.handle.net/10872/20494>

- Mirjalili, V., & Raschka, S. (2020). *Python Machine Learning*. Marcombo.
- Mon Bianco, J. P. (2022). *Mejora de la eficiencia computacional de técnicas machine learning para la detección temprana de defectos superficiales en piezas sometidas a procesos de fundición y estampado de chapa*. [Tesis de Maestría]. Universidad de Oviedo. <http://hdl.handle.net/10651/61970>
- Murnawan & Nugraha, U. (2021). Classify Event Participants in Universities and Industries Using Knowledge Discovery in Databases. *Review of International Geographical Education (RIGEO)*, 11(1), 526-542. doi: 10.48047/rigeo.11.1.36
- Ogosi Auqui, J. A. (2022). *Nueva metodología CSKT para mejorar los proyectos de implementación de Machine Learning en Ingeniería Industrial*. [Tesis de maestría]. Universidad Internacional de La Rioja. <https://renati.sunedu.gob.pe/handle/sunedu/3542110>
- Ortiz de la Torre, L., & Valadez Najera, C. (2019). *Aplicación móvil para la difusión de oferta de empleos y gestión de información para el reclutamiento de personal de maquiladoras de Ciudad Juárez*. [Tesis de pregrado]. Repositorio Universidad Autónoma de Ciudad Juárez. <http://erecursos.uacj.mx/handle/20.500.11961/5785>
- Ortiz, G., & Andres, O. (2021). Aplicación del sistema de Machine Learning para aumentar la eficiencia de las organizaciones. *Administración de Empresas*.
- Osorio Perdomo, R. (2021). Aplicación de las redes sociales en el proceso de reclutamiento y selección del personal en el sector pyme de Villavicencio, Colombia. <http://dx.doi.org/10.19137/perspectivas-2021-v11n1a12>

- Pérez González, G. A. (2021). *Detección de transacciones fraudulentas en tarjetas de crédito mediante el uso de modelos de Machine Learning*. Universidad de los Andes. <http://hdl.handle.net/1992/53571>
- Pérez, J. (2021). *Minería de Datos/Texto–Estructural– Multicriterio como recurso estratégico en la selección de personal*. [Tesis de doctorado publicada] Universidad Nacional de Colombia. <https://repositorio.unal.edu.co/handle/unal/80145>
- Pineda, J.M. (2022). Modelos predictivos en salud basados en aprendizaje de maquina (machine learning). *Revista médica Clínica Las Condes*, 33(6), 583–590. <https://doi.org/10.1016/j.rmclc.2022.11.002>
- Pró, G. & Raymond, W. (2021). *Laborem. SPDTSS*. Obtenida el 18 de febrero de 2022, de <https://www.spdtss.org.pe/wp-content/uploads/2021/09/Laborem17.pdf>
- Quipas, M. (2021). *Machine Learning en la mejora de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021*. [Tesis de Maestría publicada]. Universidad César Vallejo.
- Rąb-Kettler, K. & Lehnervp, B. (2019). Recruitment in the times of machine learning. *Management Systems in Production Engineering*, 27(2), 105-109.
- Redacción. (20 diciembre 2021). *Edadismo: la mitad de los mayores de 40 se ha sentido en desventaja por su edad en un proceso de contratación*. RRHH Digital. <https://www.rrhhdigital.com/secciones/seleccion/150135/Edadismo-la-mitad-de-los-mayores-de-40-se-ha-sentido-en-desventaja-por-su-edad-en-un-proceso-de-contratacion/>
- Salas, D. (2020). *Manual de Epistemología para la Investigación en Negocios*. @CENTRUM PUCP. <http://repositorio.pucp.edu.pe/index/handle/123456789/174767>

- Sánchez, I. (2020). *Diseño de un sistema de reclutamiento y selección por competencias para disminuir la rotación de personal en Sepronac Cía. Ltda. Quito*. [Tesis de maestría publicada]. Universidad Andina Simón Bolívar.
- Schröer, C., Kruse, F., & Gómez, J. (2021). A systematic literature review on aplicación móvillying CRISP-DM process model. *Procedia Computer Science*, *181*, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Servir Perú. (2020). *Técnicas modernas de selección virtual de personal*. Obtenida el 22 de enero de 2022, de https://www.facebook.com/watch/live/?ref=watch_permalink&v=172462984187547
- Shaari, S., & Amiru, S. (2020). Regulating flexible working arrangements (FWAs) in Malaysian private sector: Are we there yet? *International Journal of Business, Economics and Law*, *21*(4), 13–22
- Sharifani, K., & Amini, M. (2023). *Machine learning and deep learning: A review of methods and applications*. <https://papers.ssrn.com/abstract=4458723>
- Solano, J., Lancheros, D., Umaña Ibáñez, S., & Coronado, J. (2022). Predictive models assessment based on CRISP-DM methodology for students' performance in Colombia - Saber 11 Test. *Procedia Computer Science*, *198*, 512–517. <https://doi.org/10.1016/j.procs.2021.12.278>
- Valdivieso, A., Díaz, C., & Sarmiento, J. (2019). Mhealth con bigdata y machine learning como soporte tecnológico para la salud en Colombia. *LOGINN Investigación Científica y Tecnológica*, *3*(2), 51–69. <https://doi.org/10.23850/25907441.2635>

Yaranga Vite, I. P. (2022). Machine Learning en la mejora del proceso de selección del personal docente en una universidad nacional, Lima 2021. [Trabajo de grado, Universidad César Vallejo]. <https://hdl.handle.net/20.500.12692/85185>

Zambrana Vera, A. M., Avilés Ribera, S. A., Gumucio Zabalaga, F., Luizaga López, M., Pineda Gamarra, P., & Illanes Velarde, D. (2020). Muestreo aleatorio de base espacial y su utilidad en la investigación epidemiológica. *Gaceta Médica Boliviana*, 43(1), 74-79.

IX. ANEXOS

Anexo A. Matriz de Consistencia

TÍTULO: Implementación de una Aplicación móvil con Machine Learning aplicando la nueva metodología CSKT para el Proceso de contratación de personal en el sector privado

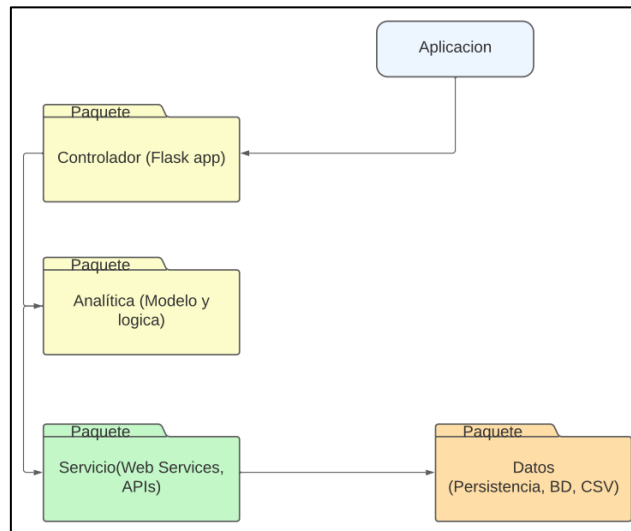
PROBLEMA GENERAL	OBJETIVO GENERAL	HIPÓTESIS GENERAL	VARIABLES	DIMENSIONES	INDICADOR(ES)	TIPO DE INVESTIGACIÓN
¿En qué medida la implementación de una aplicación móvil con Machine Learning, basada en una nueva metodología CSKT, optimiza el proceso de contratación de personal en el sector privado?	Implementar y evaluar una aplicación móvil con Machine Learning empleando la nueva metodología CSKT para optimizar el proceso de contratación de personal en el sector privado.	La implementación de una aplicación móvil con ML y la metodología CSKT optimiza significativa el proceso de contratación de personal en el sector privado.	Variable Independiente	Manejo de recursos	Tasa de precisión del modelo	<ul style="list-style-type: none"> • Básica • Aplicada
			Aplicación móvil con Machine Learning.	Funcionalidades	Eficiencia del modelo	NIVEL DE INVESTIGACIÓN
				Adaptabilidad	Evaluación del prototipo	<ul style="list-style-type: none"> • Explicativa • Descriptiva • Predictiva
					F1 – Score	DISEÑO DE INVESTIGACIÓN
						<ul style="list-style-type: none"> • Preexperimental
						UNIVERSO
						N = Procedimientos del proceso de contratación de personal
						MUESTRA:
						n = 5000
						MUESTREO:
						Aleatorio
						TÉCNICA:
						Consulta a la base de datos
						INSTRUMENTO:
						Ficha de observación
PROBLEMAS ESPECÍFICOS	OBJETIVOS ESPECÍFICOS	HIPÓTESIS ESPECÍFICAS	Variable Dependiente			
¿Cómo la metodología CSKT aplicada en el modelo de Machine Learning mejora la eficiencia en términos de tiempo y costos en el proceso de contratación de personal?	Verificar la eficacia de la metodología CSKT aplicada en el modelo de Machine Learning mejora la eficiencia en términos de tiempo y costos en el proceso de contratación de personal.	La metodología CSKT aplicada en el modelo de Machine Learning mejora la eficiencia en términos de tiempo y los costos en el proceso de contratación de personal.		Elección de personal	Tiempo de contratación.	
¿Cómo el prototipo del modelo de Machine Learning evalúa de manera precisa los criterios de aceptación basándose en la categorización del proceso de contratación de personal?	Desarrollar y analizar el prototipo de Machine Learning que evalúe los criterios de aceptación en el proceso de contratación de personal.	El prototipo de Machine Learning evalúa de manera precisa los criterios de aceptación basándose en la categorización del proceso de contratación de personal.		Satisfacción	Costo de contratación.	
¿De qué manera la precisión del modelo de Machine Learning con la metodología CSKT optimiza el proceso de contratación de personal en el sector privado?	Medir la precisión del modelo de Machine Learning con la metodología CSKT y su impacto en la optimización del proceso de contratación de personal en el sector privado.	La tasa de precisión del modelo de Machine Learning con la metodología CSKT optimiza significativamente el proceso de contratación en el sector privado.	Proceso de Contratación de personal.	Manejo de vacantes	Satisfacción del postulante.	
¿Cómo la eficiencia del proceso de contratación de personal se ve mejorada con la utilización del modelo de		La eficiencia del proceso de contratación de personal mejora con la utilización			Eficiencia de vacantes.	

Machine Learning en la aplicación móvil? ¿Como la implementación practica de la aplicación móvil con el modelo de Machine Learning y la metodología CSKT es efectiva en la optimización del proceso de contratación de personal en el sector privado?	Evaluar la eficiencia operativa del modelo de Machine Learning integrado en la aplicación móvil. Implementar la aplicación móvil con Machine Learning y la metodología CSKT y estudiar su efectividad en el proceso de contratación de personal.	del modelo de Machine Learning en la aplicación móvil. La implementación practica de la aplicación móvil con Machine Learning y la metodología CSKT es efectiva en la optimización del proceso de contratación de personal en el sector privado.
--	---	---

Anexo B. Diagrama de paquetes

Figura 83

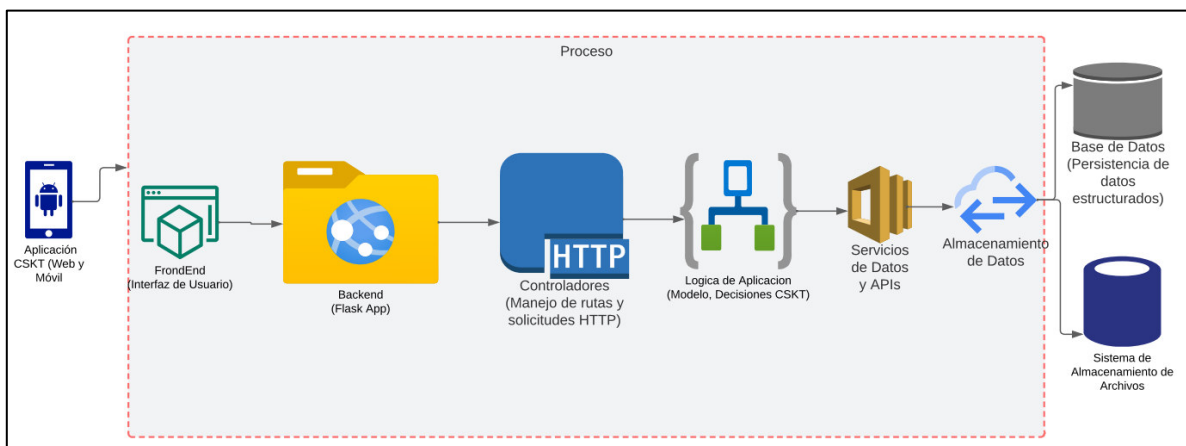
Diagrama de paquetes



Anexo C. Arquitectura de TI de la aplicación móvil

Figura 84

Arquitectura de TI de la aplicación móvil



Anexo D. Depuración de la data histórica en Excel

nombre_apellido	edad	sexo	RANGO DE FUMILACION
ABDIAS CAIRO FERNANDEZ	22	seño	M
ABDIAS CHUMBE FASADI	20	seño	M
ABDIAS CIRIACO MORALES	29	seño	M
ABDIAS HERLUAY RUIZ	58	seño	M
ABDIAS MONTEZA VALDEZ	26	seño	M
ABDIAS PARRA CASTRO	30	seño	M
ABDIAS RODRIGUEZ GUERRA	23	seño	M
ABDIAS RODRIGUEZ GUERRA	21	seño	M
ABDIAS SOKATES NUÑOZ TORREJON	21	seño	M
ABDIAS ZEGARRA MALAVER	28	seño	M
ABDIOS LEONIDAS GUARDOS SALAS	37	seño	M
ABDON ALEJANDRO ACEVEDO FLORES	21	seño	M
ABDON ALEJANDRO BALACABRE KICHICANA	32	seño	M
ABDON ALFONSO CALENES FINTO	21	seño	M
ABDON ASENCIOS GUENCA	29	seño	M
ABDON AZÁLA CABRILLO	23	seño	M
ABDON CABALLERO ENRIQUEZ	22	seño	M
ABDON GALE YONACIO	37	seño	M
ABDON CANDIOTTI SARMIENTO	20	seño	M
ABDON CLAUDIO SOCA MONTOYA	20	seño	M
ABDON CONGILIA ROMAN	18	seño	M
ABDON ESPINOZA GUISPE	19	seño	M
ABDON FERNANDEZ PALOMINO	36	seño	M
ABDON FERNANDEZ PALOMINO	31	seño	M
ABDON FERNANDEZ PALOMINO	32	seño	M
ABDON FLORES CALIXTO	37	seño	M
ABDON FRISANCHO CAPIUA	36	seño	M
ABDON FIGUEROA LEIVA	37	seño	M
ABDON GAMBOA RAUPAS	20	seño	M
ABDON GAMBOA RAUPAS	31	seño	M
ABDON GARCIA BOLESIOS	26	seño	M
ABDON HUANCA GUISPE	21	seño	M
ABDON HUGO GONZALEZ CHAVEZ	32	seño	M
ABDON LLANZO LAZO	20	seño	M
ABDON ISIDRO POSTIGO RAMOS	38	seño	M
ABDON ISRAEL MORENO CASTILLO	28	seño	M
ABDON MAMANI CCAMA	18	seño	M
ABDON MANUEL VALERO GUEZMAN	36	seño	M
ABDON MONSERRAT MORETO	23	seño	M
ABDON NAVARRETE FACHEDO	37	seño	M
ABDON NALLI RAMA	33	seño	M
ABDON FARIAPAZA CHOLQUEHUANCA	36	seño	M
ABDON GUISPE FAUCAR	33	seño	M
ABDON SANKREZ CORREA	27	seño	M
ABDON BIVIS PUGA	37	seño	M

Fuente:

[https://unfvpe-](https://unfvpe-my.sharepoint.com/:x/g/personal/2021008037_unfv_edu_pe/EQ1SVxqja79Cvg6lTbAoIUUByHLwCCD_6TeEUJT6r-71FQ?e=qD9bE8)

[my.sharepoint.com/:x/g/personal/2021008037_unfv_edu_pe/EQ1SVxqja79Cvg6lTbAoIUU](https://unfvpe-my.sharepoint.com/:x/g/personal/2021008037_unfv_edu_pe/EQ1SVxqja79Cvg6lTbAoIUUByHLwCCD_6TeEUJT6r-71FQ?e=qD9bE8)

[ByHLwCCD_6TeEUJT6r-71FQ?e=qD9bE8](https://unfvpe-my.sharepoint.com/:x/g/personal/2021008037_unfv_edu_pe/EQ1SVxqja79Cvg6lTbAoIUUByHLwCCD_6TeEUJT6r-71FQ?e=qD9bE8)

Anexo E. Algoritmo de Machine Learning con la metodología CSKT

```

1 from flask import Flask, request, jsonify
2 import pandas as pd
3 from sklearn.preprocessing import StandardScaler
4 from sklearn.metrics import precision_score, recall_score, f1_score, roc_auc_score
5 from sklearn.model_selection import train_test_split
6 from sklearn.tree import DecisionTreeClassifier
7 import matplotlib.pyplot as plt
8 import joblib
9
10 app = Flask(__name__)
11 # Definir X_train y X_val_scaled a nivel global para que sea accesible en todo el script
12 X_train = None
13 y_train = None
14 model = None
15
16 # Crear un clasificador de árboles de decisión
17 model = DecisionTreeClassifier(random_state=42)
18
19 @app.route('/api/train_model', methods=['POST'])
20 def train_model():
21     global X_train, y_train, model # Indicar que estamos utilizando las variables globales
22     try:
23         # Resto del código de entrenamiento aquí...
24         training_data = request.json
25         df_train = pd.DataFrame(training_data['training_examples'])
26

```

```

26
27 X_train = df_train[['edad', 'sexo', 'rango_edad_jubilarse', 'sueldo', 'aspiracion_ganar_mas']]
28 X_train = pd.get_dummies(X_train, columns=['sexo', 'aspiracion_ganar_mas'], drop_first=True)
29 X_train_scaled = StandardScaler().fit_transform(X_train)
30 y_train = df_train['decision'].astype(float)
31 # Entrenar el modelo de árbol de decisión
32 model.fit(X_train_scaled, y_train)
33
34 # Guardar el modelo entrenado
35 joblib.dump(model, 'modelo_entrenado.joblib')
36
37 # Calcular métricas después del entrenamiento
38 y_pred_train = model.predict(X_train_scaled)
39 precision_train = precision_score(y_train, y_pred_train)
40 recall_train = recall_score(y_train, y_pred_train)
41 f1_train = f1_score(y_train, y_pred_train)
42 roc_auc_train = roc_auc_score(y_train, y_pred_train)
43
44 # Devolver métricas como parte de la respuesta JSON
45 metrics_response = {
46     'precision': precision_train,
47     'recall': recall_train,
48     'f1_score': f1_train,
49     'roc_auc': roc_auc_train
50 }
51

```

```

51
52 return jsonify({'message': 'Modelo de árbol de decisión entrenado exitosamente.', 'metrics': metrics_response})
53 except Exception as e:
54     return jsonify({'error': str(e)}), 400
55 @app.route('/api/prediccion', methods=['POST'])
56 def predecir_perfil():
57     try:
58         # Recibir datos en formato JSON desde la aplicación
59         data = request.json
60
61         # Acceder al perfil directamente
62         perfil = data['perfiles']
63
64         # Crear un DataFrame a partir de los datos recibidos
65         df = pd.DataFrame(perfil)
66
67         # Asegurarse de tener todas las columnas necesarias
68         X = df[['edad', 'sexo', 'rango_edad_jubilarse', 'sueldo', 'aspiracion_ganar_mas']]
69
70         # Asegurarse de que X tenga las mismas columnas que se utilizaron durante el entrenamiento
71         X = X.reindex(columns=X_train.columns, fill_value=0)
72         # Transformar las características usando el StandardScaler ajustado durante el entrenamiento
73         X_scaled = StandardScaler().fit(X_train).transform(X)
74         y_pred = model.predict(X_scaled)
75         # Tomar decisión ONP o AFP basada en la salida del modelo decision = "Predicción para {}:".format(perfil['nombre_ape
76         # Responder con la decisión
77         return jsonify({'decision': decision})
78
79     except Exception as e:
80         return jsonify({'error': str(e)}), 400
81     if __name__ == '__main__':
82         app.run(debug=True)
83

```

Help us improve our support for JavaScript

Take Short Survey