



Universidad Nacional
Federico Villarreal

VRIN | VICERRECTORADO
DE INVESTIGACIÓN

ESCUELA UNIVERSITARIA DE POSGRADO

DESARROLLO DE UN MODELO PARA LA MEDICIÓN DE RADIACIÓN
UV-B UTILIZANDO TÉCNICAS DE INTELIGENCIA ARTIFICIAL

Línea de Investigación:
Sistemas inteligentes, robótica, domótica

Tesis para optar el grado académico de
Doctor en Ingeniería de Sistemas

Autor:
Sanchez Atuncar, Giancarlo

Asesor:
Hilario Falcon Francisco
(ORCID: 0000-0003-3153-9343)

Jurado:
Manrique Suarez Luis Humberto
Flores Vidal Higinio Exequiel
Vales Carrillo Jorge Alberto

Lima – Perú
2022

Índice

I.	INTRODUCCIÓN.....	11
	1.1. Planteamiento del problema	13
	1.2. Descripción del problema.....	17
	1.3. Formulación del problema.....	19
	1.3.1. Problema general	19
	1.3.2. Problemas específicos	19
	1.4. Antecedentes	19
	1.5. Justificación de la investigación.....	30
	1.6. Limitaciones de la investigación	32
	1.7. Objetivos	32
	1.7.1. Objetivo general.....	32
	1.7.2. Objetivos específicos	32
	1.8. Hipótesis.....	33
	1.8.1. Hipótesis general.....	33
	1.8.2. Hipótesis específicas	33
II.	MARCO TEÓRICO	34
	2.1. Marco conceptual	34
III.	MÉTODO	46
	3.1. Tipo de investigación	46
	3.2. Población y muestra	47

3.3. Operacionalización de variables.....	48
3.4. Instrumentos	49
3.5. Procedimientos	49
3.6. Análisis de datos.....	50
3.7. Consideraciones éticas	50
IV. RESULTADOS	51
V. DISCUSIÓN DE RESULTADOS	80
VI. CONCLUSIONES.....	83
VII. RECOMENDACIONES	84
VIII. REFERENCIAS	84
IX. ANEXOS.....	95
Anexo 1: Matriz de consistencia.....	959
Anexo 2: Etapas del modelado de desarrollo.....	102
Anexo 3: Validacion de instrumentoe	114

Índice de tablas

Tabla 1	Aportes de los artículos del estado del arte.....	30
Tabla 2	Limitaciones de la investigación.....	33
Tabla 3	Categorías del índice UV	43
Tabla 4	Matriz de confusión – árbol de decisión	53
Tabla 5	Matriz de observación – árbol de decisión.....	54
Tabla 6	Métricas de evaluación – árbol de decisión	55
Tabla 7	Matriz de confusión – KNN.....	56
Tabla 8	Matriz de observación – KNN	56
Tabla 9	Métricas de evaluación – KNN	57
Tabla 10	Matriz de confusión – Regresión logística.....	58
Tabla 11	Matriz de observación – Regresión logística	58
Tabla 12	Métricas de evaluación – Regresión logística.....	59
Tabla 13	Matriz de confusión – Redes bayesianas	60
Tabla 14	Matriz de observación – Redes bayesianas.....	60
Tabla 15	Métricas de evaluación – Redes bayesianas	61
Tabla 16	Matriz de confusión – Redes neuronales	62
Tabla 17	Matriz de observación – Redes neuronales.....	62
Tabla 18	Métricas de evaluación – Redes neuronales	63
Tabla 19	Matriz de confusión – SVM.....	64
Tabla 20	Matriz de observación – SVM	64
Tabla 21	Métricas de evaluación – SVM.....	65
Tabla 22	Métrica de evaluación H1– Precisión árbol de decisión.....	66
Tabla 23	Métrica de evaluación H1– Precisión KNN.....	66
Tabla 24	Métrica de evaluación H1– Precisión Regresión logística.....	67

Tabla 25	Métrica de evaluación H1– Precisión Redes bayesianas	67
Tabla 26	Métrica de evaluación H1– Precisión Redes neuronales	68
Tabla 27	Métrica de evaluación H1 – Precisión SVM.....	68
Tabla 28	Cuadro comparativo de resultados según el indicador precisión	69
Tabla 29	Métrica de evaluación H2– Sensibilidad árbol de decisión	70
Tabla 30	Métrica de evaluación H2– Sensibilidad Precisión KNN.....	70
Tabla 31	Métrica de evaluación H2– Sensibilidad Regresión logística.....	71
Tabla 32	Métrica de evaluación H2– Sensibilidad Redes bayesianas	71
Tabla 33	Métrica de evaluación H2– Sensibilidad Redes neuronales	71
Tabla 34	Métrica de evaluación H2 – Sensibilidad Precisión SVM.....	72
Tabla 35	Cuadro comparativo de resultados según el indicador sensibilidad	72
Tabla 36	Métrica de evaluación H3– Especificidad árbol de decisión	73
Tabla 37	Métrica de evaluación H3– Especificidad KNN.....	73
Tabla 38	Métrica de evaluación H3– Especificidad Regresión logística.....	74
Tabla 39	Métrica de evaluación H3– Especificidad Redes bayesianas	74
Tabla 40	Métrica de evaluación H3– Especificidad Redes neuronales	74
Tabla 41	Métrica de evaluación H3 – Especificidad SVM.....	75
Tabla 42	Cuadro comparativo de resultados según el indicador especificidad	75
Tabla 43	Métrica de evaluación H4– Recall árbol de decisión.....	76
Tabla 44	Métrica de evaluación H4– Recall KNN	76
Tabla 45	Métrica de evaluación H4– Recall Regresión logística	77
Tabla 46	Métrica de evaluación H4– Recall Redes bayesianas.....	77
Tabla 47	Métrica de evaluación H4– Recall Redes neuronales.....	77
Tabla 48	Métrica de evaluación H4 – Recall SVM	78
Tabla 49	Cuadro comparativo de resultados según el indicador recall.....	78

Tabla 50	Métrica de evaluación H5– F1 Score árbol de decisión.....	79
Tabla 51	Métrica de evaluación H5– F1 Score KNN	79
Tabla 52	Métrica de evaluación H5– F1 Score Regresión logística	80
Tabla 53	Métrica de evaluación H5– F1 Score Redes bayesianas.....	80
Tabla 54	Métrica de evaluación H5– F1 Score Redes neuronales.....	80
Tabla 55	Métrica de evaluación H5 – F1 Score SVM	81
Tabla 56	Cuadro comparativo de resultados según el indicador F1 Score	81
Tabla 57	Matriz de consistencia.....	97

Índice de figuras

Figura 1	Estación: Von Humboldt.....	13
Figura 2	Índice de Radiación Ultravioleta según horario (2016-2021)	18
Figura 3	Ecuación de radiación	36
Figura 4	Metodología Kdd	100
Figura 5	Análisis exploratorio de datos.....	101
Figura 6	Visualización de datos extraídos.....	101
Figura 7	Limpieza de datos y conversión de datos.....	102
Figura 8	Convirtiendo de Celsius a Kelvin	102
Figura 9	Modelo de estimación de radiación solar.....	103
Figura 10	Datos de radiación solar	103
Figura 11	Conversión a Mw/m ²	104
Figura 12	Índice de radiación Ultravioleta (IUV)	104
Figura 13	Índice de radiación IUV	105
Figura 14	Descargando datos del modelo	105
Figura 15	Modelamiento de árbol de decisión	106
Figura 16	Análisis del modelo de radiación del árbol de decisión.....	106
Figura 17	Matriz de confusión de modelo de radiación de árbol de decisión	107
Figura 18	Modelamiento de Regresión logística.....	107
Figura 19	Análisis del modelo de radiación de la logística.....	108
Figura 20	Matriz de confusión de modelo de radiación logística	108
Figura 21	Modelamiento de KNN	109
Figura 22	Análisis del modelo de radiación del KNN	109
Figura 23	Matriz de confusión de modelo de radiación de KNN.....	110
Figura 24	Modelamiento de Redes bayesianas	110

Figura 25	Análisis del modelo de radiación de redes bayesianas	111
Figura 26	Matriz de confusión de modelo de radiación de red bayesiana	111
Figura 27	Modelamiento de Redes neuronales	112
Figura 28	Precisión del modelo de radiación de la red neuronal de decisión	112
Figura 29	Matriz de confusión de modelo de radiación de red neuronal	112
Figura 30	Modelamiento de SVM	113
Figura 31	Análisis del modelo de radiación de SVM de decisión	113
Figura 32	Matriz de confusión de modelo de radiación del SVM	113

Resumen

El propósito de la investigación es desarrollar un modelo mediante técnicas de inteligencia artificial para medir la radiación UV-B en el Distrito de La Molina, cuyo tipo de investigación es aplicada y diseño experimental de tipo pre-experimental con enfoque cuantitativo. La metodología utilizada para el desarrollo fue la metodología KDD (Knowledge Discovery in Database) la cual consta de las siguientes etapas: Selección de datos, procesamiento de datos, transformación de datos, minería de datos e interpretación. Para el estudio se utilizaron 43425 mil datos meteorológicos del SENAMHI entre 2016 y 2021, de los cuales el 80% y 20% de los datos se utilizaron para entrenamiento y validación de los respectivos modelos. Los resultados obtenidos muestran que de los seis algoritmos utilizados, entre los que se encuentra el árbol de decisión, Vecinos Cercanos, Regresión logística, Redes Bayesianas, Redes Neuronales y SVM, se pudo determinar que con respecto al Árbol de Decisión: precisión = 100%, sensibilidad = 100 %, especificidad = 100 %, recuperación = 100 % y puntuación F1 = 1,98; KNN: precisión = 99,24 %, sensibilidad = 98,12 %, especificidad = 99,52 %, Recall = 98 % y puntaje F1 = 1,94; Regresión logística: precisión = 99,77 %, sensibilidad = 99,44 %, especificidad = 99,86 %, recuperación = 99 % y puntaje F1 = 1,96; Redes bayesianas: precisión = 80,44 %, sensibilidad = 62,36 %, especificidad = 86,80 %, recuperación = 62 % y puntuación F1 = 0,6; Redes neuronales: precisión = 90,24 %, sensibilidad = 98,09 %, especificidad = 99,52 %, recuperación = 98 % y puntaje F1 = 1,94; SVM: precisión = 99,39 %, sensibilidad = 98,16 %, especificidad = 24,92 %, Recall = 98 % y puntuación F1 = 1,94. Concluyendo que el árbol de decisión permite predecir con alta precisión, sensibilidad, especificidad, Recall y puntuación F1 de la radiación UV-B .

Palabras clave: algoritmos, radiación solar, árbol de decisión, vecinos cercanos, regresión logística, redes bayesianas, redes neuronales y SVM

Abstract

The purpose of the research is to develop several machine learning models to predict UV-B radiation and determine the best algorithm, whose type of research is applied and experimental design of pre-experimental type with a quantitative approach. The methodology used for the development was the KDD (Knowledge Discovery in Database) methodology which consists of the following stages: Data selection, data processing, data transformation, data mining and interpretation. For the study, 43425 thousand meteorological data from SENAMHI between 2016 and 2021 were used, of which 80% and 20% of the data were used for training and validation of the respective models. The results obtained show that of the six algorithms used, among which is the decision tree, Close Neighbors, Logistic regression, Bayesian networks, neural networks and SVM, it was possible to determine that with respect to the Decision Tree: accuracy = 100%, sensitivity = 100%, specificity = 100%, Recall = 100% and F1 score = 1.98; KNN: accuracy = 99.24 %, sensitivity = 98.12%, specificity = 99.52%, Recall = 98% and F1 score= 1.94; Logistic Regression: accuracy = 99.77%, sensitivity = 99.44%, specificity = 99.86%, Recall = 99% and F1 score = 1.96; Bayesian networks: accuracy = 80.44%, sensitivity = 62.36%, specificity = 86.80 %, Recall = 62% and F1 score = 0.6; Neural networks: accuracy = 90.24%, sensitivity = 98.09%, specificity = 99.52 %, Recall = 98% and F1 score = 1.94; SVM: accuracy = 99.39%, sensitivity = 98.16%, specificity = 24.92 %, Recall = 98% and F1 score = 1.94. Concluding that the decision tree allows to predict with high precision, sensitivity, specificity, Recall and F1 score of UV radiation.

Keywords: algorithms, solar radiation, decision tree, Close Neighbors, Logistic regression, Bayesian networks, neural networks and SVM

I. INTRODUCCIÓN

El presente trabajo de investigación da a conocer sobre la importancia de tener radiación solar en lima metropolitana. La energía liberada por el Sol, que se transmite al exterior mediante radiaciones electromagnéticas, se conoce como la radiación solar. Esta energía es una mezcla de radiaciones de longitudes de onda entre 200 nm y 4000 nm, que se denominan radiación ultravioleta, luz visible y radiación infrarroja. A su vez, la magnitud que mide la radiación solar que llega a la Tierra es la irradiancia, que se define como la potencia por unidad de superficie, y se mide en vatios/metro cuadrado (W/m^2). De la misma manera, la irradiancia que llega a la atmósfera superior de la Tierra es definida como constante solar, teniendo por valor, el establecido por la NASA de $1353 W/m^2$ y de $1373 W/m^2$ según la WMO (Organización Mundial de Meteorología).

La atmósfera no es calentada directamente por la radiación solar, sino de manera indirecta a través de la reflexión de dicha radiación en el suelo y en la superficie de mares y océanos.

Según Kolmakov et al. (2020) el agotamiento en la estratosfera de la capa de ozono, ha llevado a incrementar la radiación solar UV en la superficie de la tierra. Existen efectos positivos asociados con la radiación UV que incluye síntesis cutánea de vitamina D y prevención de raquitismo, osteomalacia y osteoporosis (Schmalwieser et al., 2005). Otros efectos positivos también son la reducción del riesgo de algunas enfermedades tales como tuberculosis o propagación de enfermedades por aire como: influenza, posibles beneficios para soriasis, hipertensión, enfermedad isquémica del corazón, diabetes y riesgo reducido de cáncer y mama.

Los efectos negativos de la radiación UV son agudas y lesiones crónicas en la piel humana (quemaduras, lesiones solares crónicas y fotodermatosis), ojos (fotoqueratitis aguda, fotoconjuntivitis, cáncer de córnea y conjuntiva, cataratas y degeneración macular), disminución

del sistema inmunológico y también algún tipo de cáncer en la piel (Epstein et al., 1961; Garssen et al., 1998).

En el ser humano, se conoce que una exposición prolongada a la radiación UV solar puede producir efectos agudos y crónicos en la salud de la piel, los ojos y el sistema inmunitario. Las quemaduras solares y el bronceado son los efectos agudos más conocidos de la exposición excesiva a la radiación UV; a largo plazo, se produce un envejecimiento prematuro de la piel como consecuencia de la degeneración de las células, del tejido fibroso y de los vasos sanguíneos inducida por la radiación UV. La radiación UV puede producir también reacciones oculares de tipo inflamatorio, como la queratitis actínica.

La piel es insensible a la incidencia de la radiación UV, por esta razón, notamos los efectos de la exposición a los rayos UV en la piel muy tarde, siendo estos: oscurecimiento, bronceado y ardor en la superficie de la piel. Desafortunadamente, siempre es demasiado tarde aplicarse la protección cuando la sobreexposición ya hizo efectos.

Por lo tanto, es conveniente contar con un sistema de advertencia y monitoreo en tiempo real de los rayos UV para que los usuarios conozcan antes de exponerse al sol. Los métodos adecuados de evaluación de la radiación UV incidente sobre un individuo deben ser capaces de reconocer el índice de radiación UV, tiempo de exposición y tipo de piel, así como brindar recomendaciones acerca de la protección solar a utilizar.

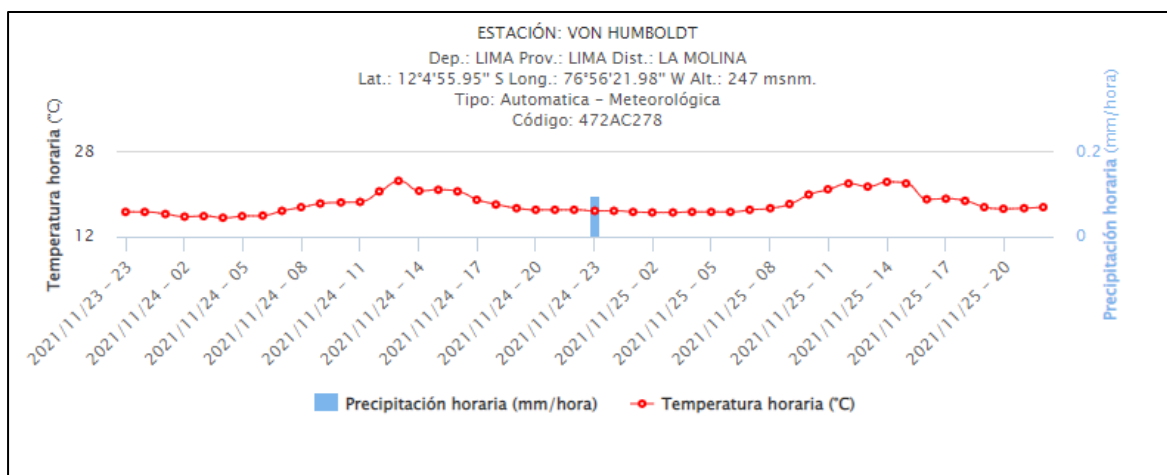
1.1. Planteamiento del problema

El tema a investigar está relacionado en el ámbito meteorológico y en la actualidad es muy complicado el monitoreo de la radiación solar debido a que se necesita múltiples componentes, aparatos para la implementación de dichas estaciones meteorológicas y además suele ser un proceso en el manejo y monitoreo, debido a eso se puede utilizar la inteligencia artificial para darle una solución a este problema y convertir este proceso más eficiente y más rápido con una optimización inteligente.

El proceso de cálculo de radiación solar requiere estaciones meteorológicas con capacidad de medición radiación solar que requiere de paneles solares posicionados además de especialistas del campo que son técnicos e ingenieros meteorólogos especializados en el área para calcular parámetros donde les permitirán obtener valores de radiación solar. En la actualidad el Senamhi proporciona datos a los ciudadanos con libre acceso, pero de solo datos de (Temperatura(°C) precipitación (mn/hora) humedad (%), dirección de viento y velocidad del viento) en diferentes estaciones a nivel nacional, sin radiación solar.

Figura 1

Estación: Von Humboldt



1.1.1. Internacional

Existen a nivel internacional estaciones que toman datos tanto meteorológicas como hidrológicos en sus complejos hidroeléctricos que miden parámetros como son: lluvia, temperatura ambiental, humedad relativa, presión atmosférica, velocidad del viento entre otros una de las más conocidas es el instituto nacional meteorología e Hidrología (INAMHI) que es la encargada del monitoreo del clima (Palaguachi, 2018).

Desde hace muchos años atrás, se ha detectado en todo el mundo un alto crecimiento de anomalías de cánceres de piel, es importante e indispensable dar a conocer a detalle los efectos nocivos sobre la radiación UV, a su vez impulsar cambios en los estilos de vida que reduzcan los continuos casos de cánceres de piel.

Según Turner et al. (2020) afirma que los efectos de la radiación ultra violeta (UV) sobre la tierra han sido continuamente objeto de investigación. La sobre exposición a la radiación ultravioleta es dañina, pero se requieren pequeñas cantidades de exposición para buena salud. Por lo tanto, es importante que los seres humanos optimicen su propia exposición a los rayos UV y no excedan niveles de UV suficientes biológicas esenciales.

La organización mundial de la salud (OMS) menciona que las radiaciones ultravioletas (UV) son radiaciones electromagnéticas con longitudes de onda entre 100 y 400nm. En cantidades pequeñas las radiaciones ultravioletas son beneficiosas para la salud y desempeñan una función esencial en la producción de vitamina D, a su vez, sin embargo, la exposición excesiva a ellas se relaciona con diferentes tipos de cáncer cutáneo.

Según como menciona Assas et al. (2016) en su investigación que presenta un conjunto de modelos de redes neuronales artificiales (ANN) para estimar la radiación solar global diaria utilizando variables como por ejemplo (intensidad media diaria de radiación solar, las horas del

sol máxima posibles, humedad relativa diaria, temperatura media diaria máxima del aire, presión atmosférica media diaria y velocidad del viento.) para la ciudad de Djelfa en Argelia.

Por lo general una red neuronal artificial proporciona un enfoque computacionalmente eficiente para determinar una relación empírica probablemente no lineal entre un numero de entrada y de salidas, en la base teórica se han propuesto varios modelos basados en el enfoque ANN con el fin de predecir los datos de radiación solar global.

En el ámbito de la investigación se ha vuelto cada vez más interesante debido a sus amplias disciplinas y a sus diferentes características relacionadas con el mundo real y a sus metodologías emergentes así lo afirma (Méndez y Morales, 2007) que estamos a puerta a una revolución tecnológica sin precedentes, los avances permiten a un programa de software aprender sobre el mejor conocimiento disponible alrededor de diferentes temas.

La principal dificultad en la producción de energía solar es la volatilidad en interrupción de la generación de energía del sistema fotovoltaico debido principalmente a las condiciones climáticas. Según lo menciona Lee et al. (2020) que principalmente la irradiación solar está muy relacionada con la captación de energía solar y que su predicción puede ser un buen indicador de producción de energía. Y por lo tanto la predicción precisa de la irradiancia solar se está volviendo vital para reducir el impacto de la incertidumbre y los costos de energía que permita la integración adecuada de los sistemas fotovoltaicos en una red inteligente.

Cabe mencionar que se han realizado muchos estudios de modelos y algoritmos para predecir la irradiancia solar basándose en varios factores meteorológicos que se miden de forma rutinaria, como la temperatura o la humedad, dado que la predicción precisa de la irradiancia solar se convierte en la columna vertebral de redes inteligentes por el aumento de la instalación de sistemas fotovoltaicos.

La energía solar es ampliamente reconocida como la energía renovable más limpia y abundante, fuente que es disponible para los seres humanos. El futuro de la tecnología es sin duda muy comprometedor, la tecnología fotovoltaica (PV), la concentración de esta energía solar puede ayudar a aprovechar para una variedad de usos, para uso doméstico, fuente de calefacción, agua caliente, etc (Pang et al., 2020).

Wang et al. (2020) afirma que, en los últimos años, los modelos de predicción solar basados en inteligencia artificial han atraído la atención de muchos investigadores. El aprendizaje profundo como técnica emergente de aprendizaje automático se ha aplicado en la predicción solar basado en el aprendizaje por conjuntos, los resultados demuestran que este predictor es capaz de extraer características de alta precisión de pronóstico.

Las características dependientes del tiempo de la irradiancia solar basado en el aprendizaje automático se aplicaron en la predicción solar.

1.1.2. Nacional

En el ámbito nacional podemos dar a conocer que la Municipalidad de Lima (2020) trabaja actualmente en un proyecto que impulsa una innovadora plataforma digital para monitorear la calidad del aire en el cercado como también el comportamiento de las variables meteorológicas de temperatura, radiación ultravioleta, presión atmosférica, humedad relativa, esta plataforma corresponde a una iniciativa del proyecto de investigación multidisciplinario “Desarrollo de un sistema de calidad de aire en zonas urbanas con módulos de medición de bajo costo en tiempo real y técnicas de inteligencia artificial.

1.1.3. Local

En el ámbito a nivel local podemos apreciar según Senamhi (2016) indica que el pasado martes 14 de abril, en la estación meteorológica campo de marte que está ubicada en el distrito de Jesús maría reporto un 28.7°C, mientras que la estación ubicada en el aeropuerto Jorge

Chávez registro 28.3°C y la estación Von Humboldt en la molina alcanzo 29.5°C. se estima que en lo posterior estas temperaturas incrementen y alcancen valores de hasta 31°C en Lima Este. Los especialistas del Senamhi indicaron que este incremento de la temperatura está influenciado por el ingreso de viento del norte y la presencia de cielo con escasa nubosidad.

1.2. Descripción del problema

Florez et al. (2018) menciona que debido al cambio climático y la exposición solar sin los debidos cuidados, el cáncer de piel ha aumentado considerablemente, es por ello que ahora muchos países vienen implementando campañas educativas que inculcan hábitos para protegerse de los daños solares severos como por ejemplo el autoexamen y la consulta oportuna al dermatólogo. El melómano y el cáncer de piel, muestran una tasa de incidencia cada vez más mayor en todo el mundo.

La piel es insensible a la incidencia de la radiación UV, por esta razón, notamos los efectos de la exposición a los rayos UV en la piel muy tarde, siendo estos: oscurecimiento, bronceado y ardor en la superficie de la piel. Es por ello que en la actualidad las persona comúnmente suelen exponerse al sol sin medir sus consecuencias.

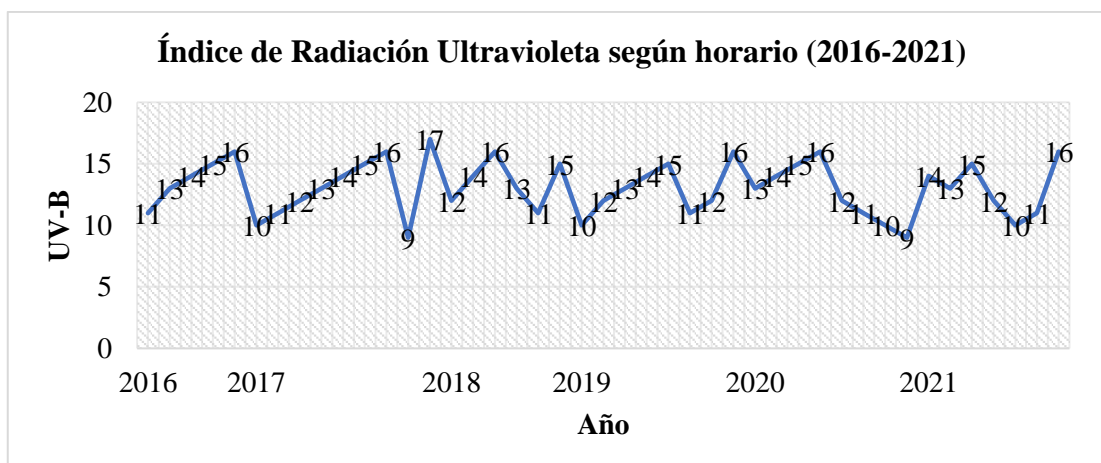
Según Senamhi (2022), con respecto al pronóstico de radiación UV-B . , informa que en la región andina los niveles de radiación ultravioleta son considerados muy altos debido a que su presencia solar supera los 30 °C. ante esta situación recomiendan a las personas considerar las siguientes medidas de prevención es decir minimizar la exposición en las horas de 10:00 am a 16:00 pm es muy importante si salimos a la calle debemos de utilizar sombreros, gorros de ala ancha, sombrillas y lentes de sol cuyos cristales absorban la radiación UV -B.

Como se sabe la exposición excesiva de la radiación UV-B causa efectos negativos a corto como a largo plazo en la salud de las personas, principalmente en la vista y en la piel, desde quemaduras solares, queratitis y complicaciones inmunológicas hasta más graves como cataratas o tumores malignos, resultando por ende los casos más común hoy en día como es el

cáncer de piel, En base a la gran cantidad de información acerca de la radiación UV-B de datos analizados, se ha podido determinar que en el Distrito de la Molina existe una gran exposición solar en determinadas horas del día durante los últimos 5 años, evidenciándose en el alto índice de radiación UV-B, el distrito de la molina se ubica geográficamente en un área limite, entre lo que se constituye la parte alta de la región chala y la parte baja de la región yunga a una altitud que va de 350 a 900 m.s.n.m tal como se muestra en la figura 2, siendo entre los horarios de 11:00 am hasta las 4:00 pm cuando es la hora pico de mayor radiación solar, en otras palabras es “Muy alta” teniendo esto riesgos para la salud de las personas como se detalló con anterioridad .

Figura 2

Índice de Radiación Ultravioleta según horario (2016-2021)



Nota. La figura muestra el índice de radiación ultravioleta de los últimos 5 años del distrito de La Molina extraídos de la base de datos del Perú. Fuente: Senamhi (2022)

Este estudio será de utilidad para las personas en general, ya que brindará un pronóstico a corto plazo sobre el índice de radiación ultravioleta, para alertar y tomar medidas de prevención evitando algunos problemas de salud en las personas que viven en del Distrito de la Molina que en un futuro es probable que presenten alguna de estas enfermedades.

1.3. Formulación del problema

1.3.1. Problema general

¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir la radiación UV-B en el Distrito de La Molina?

1.3.2. Problemas específicos

¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir la precisión de la radiación UV-B en el Distrito de La Molina?

¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir la sensibilidad de la radiación UV-B en el Distrito de La Molina?

¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir la especificidad de la radiación UV-B en el Distrito de La Molina?

¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir el recall de la radiación UV-B en el Distrito de La Molina?

¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir el F1 score de la radiación UV-B en el Distrito de La Molina?

1.4. Antecedentes

Existen investigaciones a nivel nacional acerca del desarrollo de un modelo para la medición de radiación UV-B utilizando técnicas de inteligencia:

Mamani (2021), hizo una investigación sobre un modelo para pronosticar basado en minería de datos, tomado en cuenta ejemplos de la investigación en radiación UV-B comparando 3 modelos, el estudio tuvo como propósito la propuesta de un modelo de precisión alta en cifras de radiación UV en Arequipa para el año 2017, la metodología fue utilizar programas como Octave y Python, para después dividirlo en: entrenamiento en los años 2010-2016 y los ensayos en la realización en dicho años, siendo 10% de información, obteniendo

como resultados que el modelo de redes neuronales con Memoria de Largo y Corto Plazo (LSTM) obtuvo un mayor coeficiente de correlación de $R= 0.72$ y un % RMSE de 28 %, en comparación del modelo de Autoregresivo de Medias Móviles (ARMA) con $R= 0.022$ y el modelo Razonamiento Basado en Casos (RBC) se obtuvo un coeficiente de correlación de $R = 0.695$ y % RMSE de 32 %, concluyendo que el mejor modelo es el de redes neuronales para poder realizar una buena predicción del modelo y la información estudiada, cumpliendo con la predicción por un periodo establecido para el nivel de aire.

Sánchez (2018), ejecutó su investigación sobre el diseño y construcción de un medidor de radiación con seguimiento solar y registrador de datos, donde tuvo como propósito diseñar un medidor de radiación solar a partir de la captación de rayos solares por una celda fotovoltaica, su transformación a niveles de radiación, la simulación del diseño y construcción del mismo, asimismo en la investigación se utilizó un sistema para rastrear en altura , a través del manejo de un motor con microcontrolador, recibiendo ondas de la radiación en un panel y convirtiéndolo en señalizaciones digitales en contorno de controlar información en transcurso del tiempo y procesándolo en aplicaciones, teniendo como resultados que con el Solar Power Meter Mod SM206, se logró una linealidad moderadamente, con variaciones de 200 W/M² en radiaciones globales de ± 13 w , pudiéndose notar solo un cambio bajo en la detección de +10 y concluyendo para lograr la medición adecuada basado en parámetros del panel, en voltios y suministro de energía, se debe tomar en cuenta la ruta del sol y cuando es necesario adquirir un microcontrolador.

Mujaico (2020), en su estudio diseñó e implementó un sistema para medir y pronosticar la radiación UV usando internet y machine learning, para esto se usó la metodología del Machine learning y IoT(MATLAB-Cloud Computing ThingSpeak), con un sensor de medición de rayos UV, emitiendo señales en intensidad de un microcontrolador con comunicación a una plataforma, analizados con algoritmo de predicción, obteniendo como resultado que la prueba

de medición tiene un $R= 25$ minutos y una IUUV promedio de 1.84, mínimo de 1.02 y máximo de 3.85, como también el comportamiento de la línea de pronóstico de $R= 0.37$ (rango máximo de separación), Concluyendo que su aplicación de la regresión lineal es en muchos ámbitos para lograr predecir de forma eficaz, así pues las pruebas de medición UV se realizaron con éxito , y el valor de error se redujo cada vez que está acercándose al horario.

Guillen (2021), implemento en su estudio un modelo para monitorear y medir las condiciones ambientales usando tecnología de Arduino y Raspberry, cuyo objetivo fue proponer la implementación de software para registro en tiempo real de los valores de las condiciones ambientales de temperatura, humedad, calidad del aire e índice de radiación ultra violeta, usando metodologías como el uso de software Raspberry, Arduino y tecnología del Internet de las Cosas (IoT), con Software de Proceso Unificado de Racional (RUP), teniendo como resultado que con Raspberry y Arduino tuvo 2 muestras de: 0.69 y 1.34 respectivamente, cambios por condiciones del ambiente, no habiendo resultados para Account Weather, Tutiempo.net, Weather Online y Clima.com , siendo más seguro el primer método, ya que muestra datos reales y confiables, diferencias del 18%, como conclusión que la implementación de la propuesta es buena, porque toma en cuenta parámetros de pronósticos reales y sensores de monitoreo, mostrando una aceptación por parte de los interesados.

Existen investigaciones a nivel internacional acerca del desarrollo de un modelo para la medición de radiación UV-B utilizando técnicas de inteligencia:

Koo et al. (2019), realizó un estudio acerca de la estimación del potencial de la radiación solar con complejos patrones espaciales vía utilizando técnicas de machine learning, donde el objetivo fue el desarrollo de estimación en promedio de cada mes en la radiación solar MADSR, utilizando la metodología de Clustering (k-means) y razonamiento en casos, es así que se recolectó datos de 97 lugares en China durante un lapso determinado de tiempo, teniendo como resultados que con ese método se obtuvo un 98.23% de exactitud de aproximación en la

radiación, concluyendo así que este tipo de tecnología ayuda a desarrollar una nueva aproximación en la estimación y tiene grandes ventajas.

Gil (2017), ejecutó un modelo para pronosticar la radiación UV ajustando con el clima de Barranquilla para dimensionar los sistemas fotovoltaicos, teniendo como objetivo seleccionar un modelo de predicción de irradiación solar para el dimensionamiento de sistemas fotovoltaicos, que se ajuste a las condiciones y disponibilidad de variables climatológicas de la ciudad de Barranquilla, así pues, se aplicó como metodología Box Jenkins con modelo ARIMA (1,0,0) y redes neuronales, teniendo como resultado que con las mejores características es el modelo de redes neuronales, estableciendo indicadores $MBE=0,4950 \text{ KWh/m}^2 \text{ /día}$; $RMSE=1,2779 \text{ KWh/m}^2 \text{ /día}$; $MAPE=27,6202 \%$; $MPE=14,1858\%$ y $MABE= 1,0858 \text{ KWh/m}^2 \text{ /día}$ para el año 2013, concluyendo que con este modelo se alcanzó tener una herramienta sofisticada, correcta para utilizar y estimar la irradiación en Bogotá

Qin (2020), realizó una revisión acerca de un modelo eficaz de banda ancha para la derivación de los rayos ultravioleta UV diaria, el objetivo de la investigación fue el desarrollo de una banda ancha en base en modelos físicos y estimar la radiación UV con rapidez, exactitud y continuidad, teniendo como metodología el utilizar una banda ancha FASTUV y re-análisis MERRA, como también el desarrollo de un método IA, evaluando mediciones en estaciones del CERN, teniendo como resultado que el método ANN está relacionado con mediciones UV con nivel bajo $RMSE (9.092 \text{ Wm}^{-2}-15.190 \text{ Wm}^{-2})$, bajo $MAE (6.909 \text{ Wm}^{-2}-12,217 \text{ Wm}^{-2})$, $RMSE$ bajo $(8,004\%-25,769\%)$, $MAER$ bajo $(6,263\%-20,726\%)$ y R alto $(0,934-0,980)$, concluyendo que el modelo ayuda en medios solares e investigaciones, siendo eficiente y buenas precisiones.

Larrosa y Cauritongo (2019), en su estudio desarrolló una aplicación usando sensores ambientales de hardware para su lectura en tiempo real de los rayos UV-B, en lugares de Guayaquil para su adecuada verificación de calidad y la información, así pues el propósito de

dicha investigación fue la generación de datos en tiempo real de variables meteorológicas usando hardware, utilizando la metodología de SCRUM, estableciendo roles para cada uno, ensamblando y conectando piezas Arduino Uno R3, teniendo como resultados que el 75% de los expertos consideran las predicciones meteorológicas de forma positiva y muestran satisfacción en la página, en otras palabras el 93,8 % lo toma como una fortaleza, así mismo su calificación del 94,38% de los expertos califican positivamente los procedimientos, finalmente se concluye que al utilizar la herramienta Arduino y sensores electrónicos brindan grandes oportunidades en la programación de sistemas y a las personas que lo usan para tener conocimiento de datos precisos en tiempo real.

Marín et al. (2020), realizó un estudio acerca de una estación de medición de rayos UV, siendo fuente de energía un sistema fotovoltaico, cuyo objetivo fue desarrollar un sistema electrónico que permita la visualización del IUV en la Universidad del Quindío y presente información preventiva a la comunidad usando el concepto IoT, siguiendo con la metodología de la implementación de una estación para medir la accesibilidad de las personas dentro de la universidad, usando 3 métodos: emitir mensajes de precaución, el Top Dow y el Timer off , teniendo como resultado que el Top Dow es el más adecuado para medir el clima y humedad, siendo indispensables para saber los cambios de rayos UV, así como con el Time Off, el consumo de la energía baja y llega a ser de 5-7 horas, concluyendo que la estación de medición de radiación ultravioleta por sistema fotovoltaico efectúa el objetivo de sensibilización a las personas, y el método Time Off es fuente para suministrar energía solar, siendo un excelente sistema de medición.

La investigación se realizará a partir de artículos relacionados al tema de estudio de investigación, tratando, temas relacionados a medición de radiación (UV) utilizando técnicas de inteligencia artificial, para ello se realizaron búsquedas para proporcionar una bibliografía

exhaustiva de la investigación. En dichas búsquedas se utilizó el método de revisión sistemática utilizado por Kitchenham (2004).

El autor Assas et al. (2016) realiza un estudio sobre el uso de la red neuronal artificial y datos meteorológicos para predecir la radiación solar global diarias. Enfoca su estado del arte en investigar sobre la energía solar, trata de obtener información precisa sobre los datos y aclara que no es fácil de adquirir cierta información debido a las secuencias medidas de valores de radiación esta disponibles solo para unas pocas ubicaciones o regiones de cada país. Para ello realiza estudios sobre: Clasificar modelos de predicción solar en dos clases:

- Modelos estocásticos como AR (Auto regresivo)-(Promedio móvil) Y ARIMA (AR Integrado MA) estos modelos se basan en esencialmente en la estimación de probabilidad.
- Modelos de redes Neuronales (ANN) que se basan para conocer el comportamiento de un sistema, por lo tanto, predecir datos invisibles.

El autor Fahrni et al. (2011) enfoca su estado del arte en investigar sobre la importancia de la salud, debido a la exposición excesiva a la luz solar que conlleva importantes riesgos, que van desde quemaduras solares hasta cáncer de piel. A pesar que las personas tienen conocimiento del riesgo, los daños a la piel con la luz solar, han aumentado en las últimas décadas. Para abordar más en el tema se realizó lo siguiente:

- Se implemento Droide solar, Sundroid que mide la radiación ultravioleta que se comunica de forma inalámbrica con el teléfono inteligente del usuario. El teléfono actúa como una interfaz de usuario para presentar los datos medidos de una manera intuitiva e notificar al usuario la cantidad critica de luz solar.
- Cuentan con un sistema que será calibrado utilizando dos trazos de seis horas registrados en el centro mundial de la radiación: señala que el sensor es capaz de medir con precisión la radiación absorbida por la piel.

- Implementación de un sistema prototipo portátil que está preparado para: el rastreo de exposición a la luz solar del usuario en tiempo real.

El autor Jiang et al. (2017) realizó un estudio acerca de la predicción diaria de la radiación solar cada dos horas con un enfoque de cambio de Markov, cuyo objetivo fue proponer el enfoque de inferencia bayesiana para llevar a cabo las predicciones puntuales e interválicas de la energía solar cada dos horas, utilizando un modelo de Markov (MS) con un enfoque de inferencia bayesiana para realizar las predicciones de intervalo y puntos diarios de la radiación solar cada dos horas, como modelo de persistencia, modelo autorregresivo, modelo de regresión de proceso gaussiano, modelo de red neuronal y métricas de evaluación de desempeño, obteniendo como resultado, los resultados muestran que el modelo de Markov es el método más adecuado, manteniéndose en 350 W/m^2 y 450 W/m^2 , es decir que brinda pronósticos de rayos UV-B más claros que los otros modelos que muestran retraso en predicciones con 25 W/m^2 y 45 W/m^2 y no reaccionan a la luz, concluyendo que el método Markov es el más adecuado en predicciones y múltiples distribuciones con más cobertura reales, para saber las condiciones de rayos UV-B en el futuro.

El autor Heo et al. (2018) realizó un estudio acerca de dosímetros inalámbricos, sin batería, flexibles y miniaturizados monitorean la exposición a la radiación solar y a la luz para fototerapia, donde su objetivo fue diseñar, fabricar y analizar dosímetros inalámbricos, flexibles y sin batería, utilizándose la metodología de la construcción de dispositivos de cobre flexible electrónico con polímero transparente a radiación solar, probados en personas sanas al aire libre y bebés (luz azul), teniendo como resultado que con la utilización de los dispositivos se mostró una exposición solar y correlación lineal de $[R^2 = 0,94, \text{pendiente: } 8,8 \pm 0,2 \text{ mV (mJ/cm}^2\text{)}^{-1}]$, con dosímetros mm-NFC UVA en el RBH con lo medido con los dosímetros comerciales en la muñeca durante las caminatas en la mañana ($n = 13$), vespertinas ($n = 9$) y en las noches ($n = 11$), es decir en la mañana, tarde y noche muestran correlaciones

lineales ($R^2 = 0,80$, $R^2 = 0,67$ y $R^2 = 0,88$), concluyendo que estos métodos nos ayudan a medir la radiación ultravioleta, ya que esto en cantidades controladas tiene un uso beneficioso para tratar afecciones inflamatorias de la piel.

Los autores Park et al. (2019) En el estudio que realiza sobre el desarrollo de un dispositivo de medición portátil basado en un sensor de índice UV con la relación EUVB de luz natural, mencionando el objetivo principal fue brindar asesoramiento sobre las actividades al aire libre para reducir los efectos negativos de la exposición a los rayos UV, haciendo uso de un aparato UV portátil con sensores TOCON E2 para la medición de los rayos UV, aplicando la actividad de eritema, así como el patrón Bluetooth HM-10, comunicándose inalámbricamente y minimizar emplear mucha energía, y el Arduino Nano con AT mega328 para la MCU, obteniendo como resultado que el error medio absoluto encontrado en el espectrómetro y el aparato fue de $0,0034 \text{ W/m}^2$, usando ecuaciones de regresión lineal para su corrección se logró minimizar el error al más de 50%, mostrando un error de $0,0017 \text{ W/m}^2$, teniendo como conclusión que con el dispositivo se logra datos más precisos y seguros, así pues ponerlo en práctica sería muy buena opción.

Los autores Husein y Chung (2019) realizó un estudio sobre el pronóstico diario de los rayos solares para microrredes usando una red neuronal recurrente de memoria a corto plazo, siendo su objetivo proponer un modelo de previsión de irradiación solar por hora y día, utilizando solo datos meteorológicos ampliamente disponibles, a saber, temperatura de bulbo seco, temperatura del punto de rocío y humedad relativa. El modelo se desarrolló utilizando una red neuronal recurrente de memoria profunda a largo plazo a corto plazo (LSTM-RNN), donde se comparó bajo el enfoque de esta red neuronal feedforward 19 (FFNN) para el pronóstico de la irradiación solar, los resultados demuestra que el error se redujo de 30% a 20% a través de la red neuronal pre-alimentada (FFNN), con el modelo LSTM se alcanzó una alta precisión y su RMSE fue de $60,31 \text{ W/m}^2$, es así que el error absoluto medio para LSTM =

36,90 W/m² y FFNN=72,45 W/m², mostrándose los puntos más altos para el modelo LSTM sobre el modelo FFNN, en conclusión el usar la tecnología más adecuada garantiza un pronóstico ideal y se logra ahorrar mucho más, aumentando el ahorro al año de suministro eléctrico en un 2%.

Los autores Tan et al. (2018) hizo una investigación acerca de la distribución de la dosis Eritemal ultravioleta solar en malasia península utilizando un instrumento de monitoreo de ozono, utilizando de metodología el uso del Ozone Monitoring Instrument (OMI) que se encuentran a bordo del satélite AURA, con esto se obtuvo la información, mapeando la distribución de tasa diaria de eritema UV, teniendo como resultados que usando el satélite AURA se pudo pronosticar que el pico más alto de rayos UV es de 7711,43 J/m² el 10 de abril, mientras que la más baja es el 10 de diciembre con 5518,13 J/m², estos datos coincidieron el tiempo de monzones y lluvias en Malasia que solo se sabía que se divide en tres bandas espectrales: (*UV-A con rango de longitud de onda 315 a 400nm, UV-C absorbida por el oxígeno O₂ y O₃, UV-B absorbido de manera eficiente pero no completamente por el O₃ estratosférico), concluyendo que haciendo uso de la tecnología se puede obtener un pronóstico más acertado y conocer las variaciones tanto espaciales como temporales de las dosis eritemal UV.

Salao (2021), desarrolló un modelo con técnica de inteligencia artificial para la medición de rayos UV en Riobamba, siendo su objetivo desarrollar un algoritmo utilizando técnicas de inteligencia artificial para medir la eficiencia de un colector solar de tubos al vacío para el calentamiento de agua de uso residencial bajo la Irradiancia de la ciudad de Riobamba, usando como metodología el uso del modelo perceptrón multicapa (MLP), sistema de neuro-difuso (ANFIS) y el uso de criterios de cuadrado de error, el error cuadrático medio de la raíz y el coeficiente de determinación y algoritmos de entrenamiento : backpropagation, Levenberg-Marquardt (LM), mostrando como resultados que el MLP, es el que tiene más precisión con la

información real $R^2=0.99215$, $RMSE=0.0101$ y $MSE=0.00019$, a diferencia de ANFIS con los datos que evidencian $R^2=0.7742$, $RMSE=0.0702$ y $MSE=0.00492$, concluyendo que al tener una herramienta adecuada, ayuda a conocer de manera exacta y medir de manera eficiente el colector solar UV-B, siendo muy eficaz.

Raksasat et al. (2021), realizó un estudio acerca de la medición de pronóstico de los rayos UV aplicando tecnología con redes neuronales, como metodología consistió en trabajar con 2 partes: decodificadores (salida) y codificadores(entrada), constando de LSTM multicapas de red neuronal, teniendo como resultados que el mejor rendimiento fue el SurfUVNet (redes neuronales), pronosticando eficientemente en errores porcentuales absolutos medios (MAPE) de 10,41 y 10,51, a diferencia del modelo Seq2Seq que obtuvo 10,60, siendo el primero seguro, ya que si cambiamos de 7-21 días tiene la tendencia de aumentar el MAPE de LSTM-SG a 13,87–17,74, logrando un pronóstico todo el día, concluyendo que el SurfUVNet es una buena metodología para medir la radiación solar, imitando el modelo de campana UV diaria en validarlo, y es una buena opción para utilizarlo por su eficiencia.

Ravinesh (2019), ejecutó su estudio acerca de máquinas de aprendizaje integrados con satélites MODIS para el pronóstico de los rayos UV, usando como metodología la utilización de diferentes modelos: redes neuronales artificial (ANN), regresión lineal múltiple (MLR) y del promedio móvil integrado autorregresivo (ARIMA), y es así que los resultados arrojaron que la ANN prevaleció en los otros 2 modelos, mostrando el 39% de error en previsión acumulado, a diferencia de MLR con 15% , y ARIMA 25%., concluyendo que utilizar un buen predictor para medir los rayos Uv, es un ayuda útil para tomar decisiones , y es que el árbol de decisión es considerado el peor medidor.

Jamroen et al. (2021), realizó una investigación sobre un sistema basado en sensores UV para medir la radiación solar, con el objetivo de desarrollar nuevos diseños o diferentes estrategias de control de seguimiento, cuya metodología se usó fue sistemas fotovoltaicos (PV)

con espectro ultravioleta (UV) con sensores UV, mostrando como resultados que el sistema propuesto aumentó el suministro de energía en 19,97% a diferencia del LDR 11,00%, por otra parte, el producir este método fue 1,31% más alto, es decir fotovoltaicas sobre el astronómico, concluyendo que el método es factible y generaría más ingresos.

Bamisile (2020) realizó un estudio acerca de los algoritmos del machine learning y el deep learning para medir la radiación solar, como metodología se utilizó diferentes modelos de aprendizaje profundo como: red neuronal artificial, red neuronal convolucional y red neuronal recurrente. Teniendo como resultados que el rendimiento del machine learning tuvo un puntaje $r = 0.9476$, $RMSE = 88.15 \text{ W/m}^2$, $NMBE = 6.67\%$ y $MAE = 42.63 \text{ W/m}^2$, así como el Deep learning obtuvo una calificación de RNN ($r = 0.8979$, $MAE = 20 \text{ W/m}^2$, $RMSE = 49.94 \text{ W/m}^2$, $NMBE = 1.72\%$), concluyendo que el modelo de machine learning (ANN, CNN y RNN), obtuvo una mayor utilidad a diferencia del deep learning, en otras palabras, es el más apropiado para medir los rayos UV.

Voyant (2017), ejecutó una investigación sobre métodos de machine learning para predecir los rayos UV-B, así pues, se utilizó métodos como: redes neuronales, regresión de vectores de soporte, árbol de regresión, bosque aleatorio, aumento de gradiente, obteniendo un error de predicción (MAPE) del 18,5% en verano y del 21,8% en invierno, $ANN = 28.47$, errores por horarios (día, mes y año) de 5% y el 15%, y el 79% de IA, basado en un enfoque conexionista (ANN), teniendo como conclusión que la metodología ANN y ARIMA son semejantes cuando se trata de calidad de predecir en condiciones variables, aunque la flexibilidad de ANN como cercanía no lineal obteniendo más preferencia que ARIMA habitual.

1.5. Justificación de la investigación

Después de una búsqueda sistematizada de artículos y tesis se procedió a realizar un análisis riguroso para el estado de arte ubicados en la Tabla N°1 que presenta los aportes de artículos relacionados a la tesis y agrupados también como competidores.

Tabla 1

Aportes de los artículos del estado del arte

Artículo	Aporte	Tipo
Assas, O., Bouzgou, H., Fetah, S., Salmi, M. y Boursas, A. 2016	En este estudio se realizó el uso de redes neuronales artificiales (Perceptrón multicapa) para modelar y predecir la radiación solar global diaria en una superficie horizontal, utilizando parámetros meteorológicos (Intensidad media diaria de radiación solar extraterrestre G_0 , las horas de sol máximas posibles, entonces media diaria humedad relativa, temperatura media diaria humedad relativa, temperatura media diaria del aire T , presión atmosférica media diaria P y velocidad del viento $V(x)$ dicha información fue proporcionada por la oficina meteorológica de Argelia.	Competidor
Heo, S., Kim, J., Gutruf, P., Banks, A., Wei, P., Pielak, R 2018	Actualmente existe una necesidad crítica de tecnología que puedan medir con precisión y promover la exposición segura a los rayos UV. Se Implementará dispositivos portátiles en forma de insignia O Muñequeras, la electrónica convencional que funciona con baterías en estos sistemas facilita el funcionamiento inalámbrico. Para medir la dosis de exposición a los rayos UVA es un modo de acumulación continua, doblado entre el pulgar y el índice.	Competidor
Husein, M. y Chung, I. 2019	En este estudio se realizó un modelo para predecir la irradiancia solar usando solo entradas exógenas; esto significa que el modelo puede generar irradiancia solar para una ubicación determinada sin una medición previa	Competidor

de la irradiancia solar en la ubicación. Se demostrará que al utilizar un enfoque de aprendizaje profundo LST-RNN pude lograr una mayor precisión que otros modelos, Se realizará un aprendizaje profundo de LSTM-RNN Y FFNN.

De la tabla N° 1 se observa que estos trabajos como de Assas et al., 2016, Heo et al., 2018, Tan et al., 2018, Husein y Chung., 2019. Son trabajos competidores que están estudiando sobre la radiación ultra violeta UV-B, indican que tan probable e importante es la exposición al solar esto se debe a que nuestro sol emite diferentes tipos de radiación electromagnética, asimismo los rayos infrarrojos son aquellos que nos proporciona calor y además los rayos ultravioletas (UV-B) que son peligrosos para nuestra salud.

La tabla N°1 sobre los aportes de los artículos del estado de arte fueron realizados con diferentes tecnologías que evalúan la radiación (UV-B) a diferencia de Husein y Chung., 2019 que desarrollo un modelo para predecir la irradiancia solar mediante inteligencia artificial se podrá obtener un aprendizaje profundo llamado LSTM-RNN Y FNN. Debido a los aportes encontrados se realizará la tesis dando a conocer el nuevo aporte innovador.

Si bien es cierto existen trabajos realizados de desarrollo de equipos wearable portátiles y aplicaciones para el teléfono celular que midan la radiación UV-B en el mundo. Nuestro trabajo será el diseño de un modelo que mida la incidencia de radiación UV-B en diversas zonas de Lima-Perú. El principal aporte de esta investigación es que en base a mi estudio se puede implementar un modelo predictivo de radiación solar obteniendo similares resultados que con bases meteorológicas el cual permite reducir costos y además permita tomar mejores resultados y mejores decisiones a consecuencias de la radiación solar: “problemas de salud cáncer a la piel” además, con el modelo obtenido se podrán realizar predicciones acerca de la radiación solar UV-B en la zona de estudio.

1.6. Limitaciones de la investigación

En el presente estudio del desarrollo de un modelo para la medición de radiación UV-B utilizando técnicas de inteligencia artificial, se pudieron identificar tres aspectos como limitantes de la investigación, y cada aspecto cumple cierto impacto que se representa a continuación:

Tabla 2

Limitaciones de la investigación

N°	Limitación Identificada	Impacto del proyecto
1	Permanecer mucho tiempo en el sol puede traer consecuencias graves.	Alto
2	Los equipos que servirán para la recopilación de datos están algo escasos.	Medio
3	Unos de los posibles obstáculos podrían ser el clima, se debe de ubicar en una zona cálida para realizar la recopilación de datos.	Alto

1.7. Objetivos

1.7.1. *Objetivo general*

Desarrollar un modelo mediante técnicas de inteligencia artificial para medir la radiación UV-B en el Distrito de La Molina.

1.7.2. *Objetivos específicos*

Determinar en qué porcentaje las técnicas de inteligencia artificial permite medir la precisión de la radiación UV-B en el Distrito de La Molina.

Determinar en qué porcentaje las técnicas de inteligencia artificial permite medir la sensibilidad de la radiación UV-B en el Distrito de La Molina.

Determinar en qué porcentaje las técnicas de inteligencia artificial permite medir la especificidad de la radiación UV-B en el Distrito de La Molina.

Determinar en qué porcentaje las técnicas de inteligencia artificial permite medir el recall de la radiación UV-B en el Distrito de La Molina.

Determinar en qué porcentaje las técnicas de inteligencia artificial permite medir el F1 score de la radiación UV-B en el Distrito de La Molina.

1.8. Hipótesis

1.8.1. Hipótesis general

El desarrollo de un modelo mediante técnicas de inteligencia artificial permite medir la radiación UV-B en el Distrito de La Molina.

1.8.2. Hipótesis específicas

El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir la precisión de la radiación UV-B en el Distrito de La Molina.

El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir la sensibilidad de la radiación UV-B en el Distrito de La Molina.

El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir la especificidad de la radiación UV-B en el Distrito de La Molina.

El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir el recall de la radiación UV-B en el Distrito de La Molina.

El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir el F1 score la radiación UV-B en el Distrito de La Molina.

II. MARCO TEÓRICO

2.1. Marco conceptual

2.1.1. *Inteligencia artificial*

Según Luis y Rodríguez (2020) define la inteligencia Artificial (IA) como:

Unos de los avances tan importantes en los últimos años que está siendo experimentado en el campo de la inteligencia artificial (IA) que son presentados por los medios de comunicación como un fenómeno de esperanza para la resolución de muchos problemas que enfrenta el ser humano. Por lo tanto, este gran conocimiento que la humanidad ha creado, la tecnología, es decir la técnica basada en la ciencia que pasa a tener implicaciones que comprometen la vida y el futuro y exigen ser enfrentadas desde la actitud crítica y responsable con el fin de que el sueño de la maquina inteligente no se convierta en una pesadilla que escape a nuestras manos.

Según De la Fuente Chacón (2019) afirma que la IA trabaja en diferentes campos como la representación del conocimiento, en la búsqueda heurística, procesamiento de lenguaje natural, aprendizaje automático y en sus comienzos en el desarrollo de sistemas expertos. Los sistemas expertos pueden ser considerados como unos de los campos que más se ha trabajado en la IA. Un sistema experto resuelve problemas complejos del mundo real que requieren experiencia mediante un computador, cabe mencionar que los sistemas bien diseñados imitan el proceso de razonamiento que los expertos utilizan para resolver problemas específicos.

Según Büyükçolpan y Tol (2019) define la IA como:

Técnicas aplicadas al diagnóstico de enfermedades que han sido utilizadas en estudios de problemas complejos, alcanzando un aceptado grado de certeza en los resultados obtenidos con respecto a la identificación de un tipo específico de enfermedad. Estas aplicaciones son ventajosas debido a que facilitan la construcción y estudio de sistemas capaces de aprender a partir de un conjunto de datos y mejorar procesos de clasificación y predicción.

2.1.2. *El sol*

Según Rodríguez (2021) el Sol es una estrella con tiene estructura en capas que consta de un núcleo, una región de radiación, una región de convección, una fotosfera, una cromosfera, una región de transición y una corona. Las manchas solares son los primeros rasgos visibles en el Sol a primera vista. Asimismo, pueden ayudar a describir la actividad magnética del sol. Llegando afectar el ambiente espacial de la tierra. Por su parte, Arancibia y best (2010) muestran que el sol da comienzo a la energía que conserva vivo al planeta Tierra, emitiendo constantemente 62.600 kilovatios (o kilovatios) de energía por cada metro cuadrado de su superficie.

2.1.3. *Redes neuronales*

Según Mateus et al. (2014) define la red neuronal como:

Las redes neuronales realizan increíbles aportes a la medicina, nuevas investigaciones sobre el cerebro a menudo estimulan la creación de nuevos patrones en las redes neuronales. Un nuevo enfoque está utilizando conexiones que se extienden mucho más allá y capas de procesamiento de enlace en lugar de estar siempre localizado en las neuronales adyacentes.

Obando-Paredes et al., (2017) Los autores definen la red neuronal también:

Para determinar con precisión la serie temporal de datos climáticos. De acuerdo con el error asociado en la predicción oscile entre 5% y un 15%, alrededor del 79 % de las técnicas usadas en predicción de potencia se basan en un enfoque de conexiones con red neuronal. Aproximadamente el 20% restante se utilizan técnicas derivadas de redes neuronales tales como lógica difusa, wavelength RBFNN, a su vez existe otra técnica denominada K-nearest Neighbors (k-NN) está basada en un algoritmo de reconocimiento de patrones el cual compara el estado original con las muestras entrenadas.

2.1.4. Radiación solar

(Wang et al., 2020) Los autores lo definen también como aporte en:

La inteligencia artificial (IA) han traído una atención considerable. Los métodos típicos de IA incluyen la máquina de vectores de soporte (SVM) máquina de aprendizaje externa, transformada wavelet, aprendizaje profundo y aprendizaje en conjunto. Las características dependientes del tiempo de irradiancia solar se descubren y se proporciona un modelo de predicción de la irradiancia solar basado en el aprendizaje por conjuntos. El aprendizaje profundo, como técnica emergente de aprendizaje automático se ha aplicado en la predicción solar, cabe destacar que los resultados muestran que este predictor es capaz de extraer características con una alta precisión de pronóstico.

Voyant et al. (2017) Afirma que tengamos en cuenta que los métodos de predicción solar en la literatura van a más allá de las predicciones puntuales. Los pronósticos probabilísticos también se utilizan con frecuencia y a menudo son soluciones más prácticas para las necesidades de energía solar. Las redes neuronales se han estudiado en muchas partes del mundo y los investigadores han demostrado la capacidad de estas técnicas para predecir con precisión la serie temporal de datos meteorológicos.

Según Garnacho et al. (2020), los efectos de radiación solar sobre las personas tienen efectos demostrativos sobre su salud, donde las quemaduras solares, la fotosensibilidad, la foto pigmentación, la inmunosupresión, las imágenes y la carcinogénesis son uno de los sucesos más desfavorables.

La radiación solar y la temperatura ambiente son recursos fundamentales para la dispersión a gran escala de tecnologías solares. El aprovechamiento de la energía solar en sistemas fotovoltaicos y solares térmicos necesita el análisis de datos de radiación solar y los modelos manejados para evaluar el rendimiento del sistema (Clemente y Alcántara, 2020).

2.1.5. Clasificación de la radiación solar

2.1.5.1. Radiación extraterrestre. Según Olejua y Navarro (2020) lo definen como radiación solar diaria recogida sobre un área horizontal ubicada en el límite superior del espacio. Además, está en función solar continua. (G_s), la latitud del sitio (l), el factor de corrección de excentricidad de la órbita terrestre (E_0), la declinación solar (δ) y el ángulo medio de la hora de salida del sol (ω_s); H_0 se puede expresar mediante la siguiente ecuación:

Figura 3

Ecuación de radiación

$$H_0 = \frac{24 \times 3600 G_s}{\pi} E_0 \left[\cos l \cos \delta \sin \omega_s + \left(\frac{\pi \omega_s}{180} \right) \sin l \cos \delta \right] \quad (2)$$

2.1.5.2. Radiación sobre la superficie terrestre. La radiación solar se define como la energía del Sol y se mide en la superficie terrestre. Esta energía se mide mediante la irradiancia solar global, intensidad que encarna en número de energía electromagnética que transgrede sobre un área por unidad de tiempo, por unidad de área, y cuya magnitud es continua en toda la superficie más allá de la atmósfera terrestre, con un valor cercano de 1368 W m^{-2} (Contreras et al., 2018).

Según Noruega et al. (2017) hay dos tipos de radiación: la radiación directa, se conoce como la radiación que se produce en línea recta, y la radiación difusa, que proviene de diferentes direcciones debido a las nubes y la turbidez atmosférica, que contribuyen al aumento de radiación emitida en comparación con la radiación directa, y la suma de estos dos producen radiación global.

2.1.5.3. Radiación difusa. Según Vanegas et al. (2015) esta energía puede representar alrededor del 15% de la radiación global en días soleados, mientras que en nublados es muy frágil, esto representa una proporción mayor. En cambio, las superficies

horizontales reciben la radiación más difusa, porque "ven" todo el hemisferio celeste, mientras que las áreas verticales recogen menos al ver la mitad del asteroide.

2.1.5.4. Radiación directa. Según Noruega et al (2017) muestra que, bajo la radiación directa, los objetos pueden mostrar una sombra, que se mueve a medida que cambia el ángulo del sol, lo que no ocurre con la radiación difusa.

2.1.5.5. Radiación reflejada (Albedo). Es la radiación que irradia desde el área terrestre, también conocida como albedo. Las áreas horizontales no recogen emisión reflejada, las áreas verticales reciben la radiación más reflejada. De manera similar, el porcentaje de radiación UVB irradiada se altera según el color y la textura, la reflectividad del césped o del suelo es inferior al 5 %, las áreas de nieve superan el 80 %, la reflectancia del agua es de aproximadamente el 20 %, el suelo arenoso, el cemento del 7 al 18 %., etc. (López Costales, 2017).

2.1.5.6. Radiación global. Es la emisión solar recogida en un ángulo consistente de 2π estereorradián sobre un área horizontal. Asimismo, contiene la emisión recibida rectamente del disco solar y la radiación celeste esparcida a medida que atraviesa la atmósfera (Camayo et al., 2019).

2.1.6. Radiación ultravioleta

Según Polo y Miranda (2020) la radiación ultravioleta es una representación energética resplandeciente que resguarda el rango de longitud de onda de 100 a 400 nanómetros conveniente a una energía de aproximadamente 12 eV, el ozono en la atmósfera reduce totalmente la radiación de 100 a 280 nm, y para el rango de 280 a 315, la atmósfera que reduce casi el 90 de esta radiación se llama UV-B. Además, el UV-B se puede calcular como irradiación indicada en W/m², o como radiación de exposición, a través de un período, expresada en J/m².

Gutiérrez et al. (2017) define la radiación ultravioleta como:

La superficie terrestre está compuesta por un 95% de radiación UVA (320-400 nm) y un 4% de radiación UVB (280-320 nm). Los UV-B son la causa predominante de eritema o quemaduras solares y daño del ADN. Por otro lado, los rayos UVA están más relacionados con el bronceado y la foto envejecimiento, que también causa daño al ADN indirectamente a través de la formación de especies reactivas de oxígeno. La radiación UCV (200-280nm) comúnmente no está presente en la superficie terrestre, excepto en regiones de gran altitud.

Hacker et al. (2018) define la radiación ultra violeta como:

La exposición a la luz solar es el principal factor de riesgo ambiental para los cánceres de piel de melánoma y queratinocitos. El melánoma es el cáncer más común entre las personas de 15 a 39 años en Australia, como también en los estados unidos y varios países europeos reportan niveles más altos de quemaduras solares en comparación con los adultos mayores. A pesar de estas campañas de salud públicas, los adultos y jóvenes continúan reportando altos niveles de exposición, los jóvenes ahora se están alejando cada vez más de los medios tradicionales, como periódicos y televisión, favoreciendo la transmisión basada en la web. Los avances tecnológicos han permitido la implementación de intervenciones de salud relevantes para el tiempo y el contexto.

Park et al. (2019) define la radiación ultra violeta como:

Los rayos ultravioletas (UV-B) son ondas electromagnéticas que representan aproximadamente el 5% de la luz solar, cuando se sobreexponen, presentan efectos malévolos sobre la piel y la salud humana, sin embargo, también se sabe que los rayos UV-B desempeñan un papel positivo, incluido el apoyo a la síntesis de vitamina D en el cuerpo humano, sin embargo, el cálculo de EUVB, que se puede realizar aplicando la función ponderada

eritematosa sobre el valor de intensidad en longitud de onda, requiere dispositivos de medición óptica especializados, a los que el público en general no puede acceder fácilmente.

Piekarski et al. (2020) los autores afirman que:

La región infrarroja a través del visible ultravioleta hasta los rayos X, da como resultado una luz de sincrotrón con alto brillo donde la función principal es que el sincrotrón solari es proporcionar un haz de electrones estables de alta calidad. Dado que existe una gran demanda de un sistema de detección de anomalías preciso y confiable, proponemos implementar las últimas arquitecturas de redes neuronales convolucionales, además con aprendizaje de transferencia para realizar una detección precisa de fallas, los modelos de redes neuronales profundas requieren no solo recursos computacionales, sino también una gran cantidad de datos de entrenamiento, validación y prueba para ajustar bien todos los parámetros, que es la principal limitación para construir redes profundas desde cero.

2.1.7. Índice de radiación ultravioleta

El índice de irradiación ultravioleta mide la fuerza de radiación ultravioleta solar en el área terrestre, llegando así a formar parte del espectro electromagnético de 100 a 00 (nm) nanómetros (nm), incluye tres subcategorías distintas ("A", "B" y "C") según el rango espectral en el que se encuentren. En el rango de 280 nm a 315 nm, los rayos UVB son más peligrosos que los rayos UVA, considerados una de las causas transcendentales de cáncer de piel, quemaduras solares, cataratas en los ojos (Dedios, 2016).

Tabla 3*Categorías del índice UV*

Categoría	Código de Color	UV	Eery (Mw/m ²)
Bajo	Verde	0 a 2	0 a 75
Moderado	Amarrillo	3 a 5	75 a 150
Alto	Naranja	6 a 7	150 a 200
Muy alto	Rojo	8 a 10	200 a 275
Extremo	Morado	11 +	275+

2.1.8. Influencia de radiación UV sobre la salud humana

Según Garnacho et al. (2020) las quemaduras solares, la inmunosupresión, las imágenes y la carcinogénesis son algunos de los efectos secundarios más significativos de la radiación solar en los humanos. Los rayos UVB pueden perjudicar la piel humana y causar enrojecimiento leve (eritema) o ardor; incluso con el tiempo, pueden causar molestias severas, lunares, espinillas, incluso cáncer de piel, la radiación UV-B en dosis normales tiene un efecto beneficioso, ya que ayuda en la producción de ciertas vitaminas en el cuerpo, lo que previene enfermedades como el raquitismo.

2.1.9. Machine learning

Según Romero et al. (2021) modelar a partir de estos datos es una tarea complicada para los humanos. Los algoritmos de Machine Learning son relevantes porque toman automáticamente estos patrones. Por lo tanto, Maaken (2020) señala que ML es una de los materiales digitales que se pueden usar en la extracción de datos para conseguir ganancias, desarrollo y éxito a largo plazo, llegando a aprender a través de algún tipo de conocimiento previo, haciéndolas más inteligentes y capaces de entregar efectos cercanos a la inteligencia humana.

2.1.10. Paradigma del machine learning

Según Méndez et al. (2020) muestran que entre los modelos manipulados para adiestrar algoritmos se encuentran modelos lógicos, modelos geométricos y modelos probabilísticos. Además, algunos modelos de aprendizaje incluyen:

- Aprendizaje supervisado. Basado en el pasado, el modelo construye una función de entrelazar las variables predictoras con las variables de clase. Esta función se utiliza luego para hacer predicciones (clasificaciones).
- Aprendizaje no supervisado. Su objetivo es obtener grupos, de forma que dentro de cada grupo existan peticiones parecidas, mientras que los grupos son diversos entre ellos. Durante este aprendizaje, el propio modelo se encarga de ejecutar sus propios cortes.
- Aprendizaje por refuerzo. El algoritmo aprende, con su interacción con el mundo rodeado, por tanto, se origina una retroalimentación que va cambiando y purificando su conducta.

2.1.11. Metodologías del machine learning

2.1.11.7. SEMMA. Esta metodología, tiene como objetivo ayudar la exploración estadística, los métodos de visualización, la selección y transformación de las variables más importantes en la predicción, el modelado transforma variable para predecir la salida y posteriormente ratifica la precisión del modelo. Al evaluar el resultado de cada paso en este proceso, puede establecer cómo modelar preguntas planteadas por resultados pasados. Como resultado, puede pasar a la fase de exploración para refinar aún más los datos. Sus fases son: muestreo, exploración, manipulación, modelado y evaluación (Rodríguez y García, 2016).

2.1.11.8. CRISP DM. Según Espinoza (2020) es un modelo de proceso de minería de datos que describe cómo abordar un problema. Conjuntamente, compone tareas requeridas en proyectos, desde la fase de comprensión del problema hasta la implementación

de sistemas automatizados de análisis, predicción y/o futuro, se puede utilizar a una variedad de problemas, como investigar registros de clientes fraudulentos, valorar la probabilidad de que los clientes de una empresa cambien a la competencia, o identificar patrones de compra y luego encomendar bienes de interés.

Entre sus fases se encuentran:

- a. Comprensión del problema o negocio: Este es el paso fundamental, porque si no se entiende el problema correctamente, los siguientes pasos serán innecesarios.
- b. Comprensión de datos: se reduce la cantidad de datos que pueden ser procesados para encarnar una rotunda información.
- c. Preparación de datos: Por lo general, este es el paso más largo del proyecto y donde los datos se transforman en función de los resultados del paso anterior seleccionado para usar en el paso de modelado.
- d. Modelado: En este paso, el modelo de minería de datos se obtiene correctamente.
- e. Evaluación del modelo: En este punto, la calidad del modelo se determina por el análisis de alguna medida estadística del modelo, comparando resultados actuales con anteriores, o examinando resultados con ayuda de expertos en el área del problema.
- f. Implementación del modelo: Este paso descarga, a través de acciones específicas, el conocimiento obtenido a través del modelo.

2.1.11.9. KDD. Según Calvache et al. (2018) El método KDD es esencialmente un proceso automatizado donde se mezclan el hallazgo y el análisis. El proceso radica en extraer modelos en forma de reglas o funciones de los datos para que los usuarios los analicen. Sus etapas son las siguientes:

- a. Pre - etapa de identificación de objetivos: Incluye conocer y definir los objetivos de KDD desde el punto de vista del cliente. Este paso es importante porque primero se conocen

los objetivos, procesos y actividades de la organización en la que quieres inmiscuirse, pues tener conocimientos previos consentirá dirigir el proceso.

- b. Etapa de selección: El objetivo de este paso es establecer la fuente de datos y el tipo de investigación a manejar, este es el paso donde se extraen los datos selectos para el análisis de las fuentes.
- c. Etapa de pre-procesamiento: Incluye un análisis en profundidad del vínculo de datos selectos en el paso anterior, donde se sitúan en práctica las operaciones y métodos necesarios para excluir el ruido, las inconsistencias o la redundancia que puedan tener los datos.
- d. Etapa de transformación: Es el paso que suministra una vista procesable, es decir, los datos están listos para ser desarrollados por algoritmos de minería.
- e. Etapa de minería de datos: Busca genera conocimiento, es decir, hallar vínculos entre datos no detectados, utilizados por clientes u organizaciones para respaldar sus decisiones. Algoritmos del machine learning.

2.1.11.10. Algoritmos de machine learning.

A. Árbol de decisión. Un árbol de decisión es un esquema que abarca posibles resultados de diversas decisiones vinculadas, lo cual a su vez permite al sujeto comparar posibles acciones en relación con costos, probabilidades y beneficios. Asimismo, la organización de un árbol de decisión inicia con un nodo raíz, donde se originan todas las ramas que tienen forma de una cadena de nodos donde se toman decisiones sobre el valor del carácter condicional, permitiendo el progreso a través del árbol. Un proceso se interrumpe en un nodo hoja, donde se toma una decisión de clasificación. Basándose en la regla de asociación (Sulla et al., 2018).

B. Regresión lineal. La regresión lineal se utiliza para pronosticar la conducta de una variable en relación con otra variable. Implicando supuestos como la linealidad de la

relación, la normalidad, la aleatoriedad de la muestra y la uniformidad de varianzas, donde cada una asiste de manera aditiva y es continua para todos los fenómenos observados (Ríos et al., 2020).

C. Redes neuronales. Según Campos y Trujillo (2021) Una red neuronal artificial imita el funcionamiento básico del cerebro. Llegando así la información fluir entre las neuronas y, según la estructura y la ganancia de las conexiones neuronales, actúa de manera diferente. En estas redes, cada neurona está acoplada con otra por un factor de corrección o peso.

D. Máquina de vectores. La máquina de vectores de soporte (SVM) es una técnica poderosa del aprendizaje automático, aunque es simple, ha confirmado ser un algoritmo poderoso que trasciende bien a los problemas reales. Se considera como una alternativa efectiva considerando las limitaciones de ANN frente al tamaño y ruido en los datos. Varios estudios han demostrado que SVM es mejor que los modelos de promedio móvil de recuperación automática, las redes neuronales artificiales y los sistemas de inferencia difusos neuronales adaptativos (Cuevas et al., 2019).

E. Bayesianos. La estadística bayesiana refiere la contrapartida de las llamadas estadísticas clásicas o "regulares", con técnicas similares a todas las siguientes y quizás con un mayor número de aplicaciones; sin embargo, el método bayesiano puede verse como una extensión del método clásico (Quero, 2020).

III. MÉTODO

3.1. Tipo de investigación

La investigación fue de tipo aplicada, el cual según Hernández et al. (2014), El tipo de investigación aplicada construye, transforma y se utiliza en una práctica determinada, además, busca dar soluciones rápidas y prácticas a un problema previamente tratado, asimismo, procura abordar un problema que afecta lo estudiado previamente conocido por el investigador.

El nivel de investigación será explicativo. Según Sánchez et al. (2018), lo define como aquel estudio donde el investigador hace la formulación de preguntas sobre las causas de los fenómenos que se está estudiando, identificando la relación de causalidad; en otras palabras, su finalidad es entender el fenómeno que se está haciendo referencia Hernández Sampieri et al. (2014)

El diseño de investigación fue experimental de tipo pre experimental. Según Ortiz et al. (2018) lo definen como una serie de pruebas en las que se realizan cambios intencionales en las variables de entrada de un proceso o sistema con el fin de observar y determinar las razones de los cambios, el cambio se puede observar en la respuesta de salida.

“En el presente estudio se consideró la cantidad de 43403 datos que fueron obtenidos del Senamhi para luego ser ingresado para diferentes modelos de predicción que nos permitió obtener diferentes métricas de precisión y determinar cuál fue el más adecuado y de esta manera hacer una predicción exacta” X representa el estímulo

$$G \text{ — } X \text{ — } O$$

G= 43403 datos del Senamhi

X= Modelos de predicción

O= Observación de la métrica

3.1.1. Enfoque de la investigación

La tesis fue de enfoque cuantitativo debido a que los datos fueron expresados numéricamente al momento de aplicar los instrumentos de medición y recolectar los datos.

3.1.2. Alcance de investigación

El alcance del estudio fue del tipo experimental.

3.2. Población y muestra

La población de esta investigación estuvo compuesta por 43403 mil datos obtenidos de Senamhi. Según Ventura León (2017) una población es un conjunto de elementos que sujetan ciertos rasgos para fines de investigación.

Mientras tanto la muestra estuvo constituida por la totalidad de la población, es decir por 43403,000 mil datos obtenidos de Senamhi. Por su parte, Ventura León (2017) lo define como un subconjunto extraído de la población conformado por unidades de análisis.

Asimismo, el muestreo empleado fue no probabilístico. Según Otzen y Manterola (2017) esta técnica de muestreo no probabilístico, la selección de los sujetos de investigación dependerá de determinadas características, criterios, etc que los investigadores están revisando en este momento; por lo tanto, pueden ser inválidos y confiables o reproducibles; debido a que este tipo de muestra no corresponde a una base probabilística.

3.3. Operacionalización de variables

3.3.1. Variable independiente

Técnicas de inteligencia artificial

3.3.1.1. Definición conceptual

Técnicas aplicadas al diagnóstico de enfermedades que han sido utilizadas en estudios de problemas complejos, alcanzando un aceptado grado de certeza en los resultados obtenidos con respecto a la identificación de un tipo específico de enfermedad. Estas aplicaciones son ventajosas debido a que facilitan la construcción y estudio de sistemas capaces de aprender a partir de un conjunto de datos y mejorar procesos de clasificación y predicción (Büyükçolpan y Tol, 2019).

3.3.2. Variable dependiente

Medición de radiación UV-B utilizando técnicas de Inteligencia

3.3.2.1. Definición conceptual

Las redes neuronales y el aprendizaje automático se utilizan ampliamente en aplicaciones de radiación solar, evaluando el rendimiento de los perceptrones multicapa (MLP) y los árboles de decisión de energía combinándolos con la regresión lineal para estimar la energía solar (Ordoñez et al., 2020).

3.3.2.2. Definición operacional

Para medir la radiación UV-B se usaron métricas de precisión, la cual se obtuvieron a través de una herramienta de Machine Learning

3.3.2.3. Indicadores

Como única dimensión se tuvo a Métricas de precisión, y sus indicadores son los siguientes: $\text{Accuracy} = (\text{TP} + \text{TN} / \text{TP} + \text{TN} + \text{FP} + \text{FN}) * 100$; $\text{Sensibilidad} = (\text{TP} / \text{TP} + \text{FN}) * 100$; $\text{Especificidad} = (\text{TN} / \text{TN} + \text{FP}) * 100$; $\text{recall} = \text{TP} / \text{TP} + \text{FN}$; $\text{F1 score} = 2 * (\text{Recall} * \text{Precisión}) / (\text{Recall} + \text{Precisión})$

3.3.2.4.Escala:

La escala fue de razón. La matriz de operacionalización de variables se encuentra en el anexo N° 2.

3.4. Instrumentos

Para la presente investigación se utilizó técnicas de observación para recopilar los datos necesarios. Según Gómez (2020) esta técnica está dirigida a observar el objeto de investigación en una situación particular, es por ello que se realiza sin interferir ni modificar el entorno en el que se despliega el objeto. Las observaciones de los participantes se pueden utilizar para capturar el "significado" de una acción social, o la observación sistemática para cuantificar este comportamiento.

Al mismo tiempo, se utilizó como instrumento la ficha de observación, este instrumento fue utilizado en relación con la recopilación y compilación de datos. Como menciona Fernández (2014) la herramienta de recolección de datos (fichas de observación) permite recolectar registros procesales de las pruebas previas y posteriores al juicio, y luego se realizan las comparaciones necesarias para verificar la conformidad propuesta.

La validación de las herramientas se realizó en opinión de expertos, bajo el denominado método juicio de expertos. Basándose en la recolección de opiniones proporcionadas por expertos profesionales en el campo de la encuesta y para evaluar los indicadores que se estudian.

3.5. Procedimientos

Para desarrollar el estudio, se siguió una sucesión de pasos, iniciando con la problemática con relación a la radiación solar, en donde la recolección de información se dio en Senamhi , para ello se procedió a plantear una solución al estudio acerca del desarrollo de un modelo para la medición de radiación UV-B, donde se utilizó técnicas de inteligencia

artificial, así pues, se procedió a buscar investigaciones similares para sustentar y fortalecer el marco teórico, realizando la revisión de la metodología, para después pasar a la aplicación de los instrumentos con la debida evaluación, y fueron procesados en el programa Python y SPSS Modeler, finalizando con vaciar la información, analizar los resultados de dicho estudio y establecer las conclusiones.

3.6. Análisis de datos

Para el análisis de datos obtenidos, se utilizaron dos métodos: La estadística descriptiva y la estadística predictiva.

En este estudio se utilizó variables de entrada y una variable de salida respecto al desarrollo de un modelo para la medición de radiación UV-B utilizando técnicas de inteligencia artificial, que fue enfocado a diferentes algoritmos de predicción. A partir de ello, los datos fueron vaciados al Programa Microsoft Excel 2019, en donde se usó la estadística descriptiva para después analizar las variables de entrada y salida en tablas y figuras.

En la estadística predictiva, se realizó el análisis luego de haber procesado los datos en el Programa Python, SPSS Modeler, se pudo hacer contraste de las hipótesis de investigación mediante las estadísticas predictivas haciendo uso de la matriz de confusión. Por otra parte, para el desarrollo de un modelo para la medición de radiación UV-B utilizando técnicas de inteligencia artificial, se usaron diferentes algoritmos de predicción, a través de los resultados obtenidos de los modelos se pudo verificar si hubo o no un rechazo respecto a las determinadas hipótesis planteadas, así mismo se logró identificar el mejor algoritmo para el caso planteado.

3.7. Consideraciones éticas

Desde el punto de vista ético, este estudio se realizó de acuerdo con las normas y lineamientos establecidos por la Universidad de Villareal. Para la recolección de datos se utilizó el instrumento de investigación.

IV. RESULTADOS

En este capítulo se detalló los resultados obtenidos de la encuesta, en base a los indicadores de precisión, sensibilidad y especificidad, recall, F1 score comparados entre los 6 algoritmos de aprendizaje para determinar el mejor.

4.1. Matriz de confusión de algoritmos

A continuación, se muestran los resultados de la matriz de confusión obtenida a partir de la herramienta SPSS Modeler para los siguientes algoritmos:

4.1.1. *Árbol de decisión*

Tabla 4

Matriz de confusión – árbol de decisión

Radiación	Alto	Bajo	Extremo	Muy alto	Moderado
Alto	195	0	0	0	0
Bajo	0	7527	0	0	0
Extremo	0	0	15	0	0
Muy alto	0	0	0	114	0
Moderado	0	0	0	0	83
SUBTOTAL	195	7527	15	114	83
TOTAL			7934		

Tabla 5*Matriz de observación – árbol de decisión*

Clase	Medidas			
	TP	TN	FP	FN
Alto	195	7739	0	0
Bajo	7527	407	0	0
Extremo	15	7919	0	0
Muy alto	114	7820	0	0
Moderado	83	7851	0	0
TOTALES	7934	31736	0	0

Para un total de 7934 registros, se identificó a 195 lugares con radiación alto de manera correcta y 0 de manera incorrecta, 7527 lugares con radiación de manera correcta y 0 incorrectos, seguidamente de 15 espacios con radiación extremo y 0 incorrectos, así también 114 lugares tuvieron una radiación muy alto y 0 predicciones incorrectas. Finalmente 83 lugares tuvieron una radiación moderada con 0 incorrectos.

Asimismo, en la siguiente tabla N° 06 se muestra las métricas de evaluación:

Tabla 6*Métricas de evaluación – árbol de decisión*

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	$Accuracy = (TP+TN/TP+TN+FP+FN) * 100$	100 %
2	Especificidad	Razón	$Especificidad = (TN/TN+FP) * 100$	100%
3	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	100 %
4	Recall		$recall=TP/TP+FN$	100 %
5	F 1 SCORE	Razón	$2 * (Recall * Precisión) / (Recall + Precisión)$	100 %

Interpretación: Como se observa en la Tabla 6, utilizando el algoritmo árbol de decisión se obtuvo una precisión igual a 100%, asimismo una especificidad con 100%, de igual manera se obtuvo una sensibilidad de 100%, Recall con un valor de 100% y F1 score con 100%.

4.1.2. KNN vecinos cercanos

Tabla 7*Matriz de confusión – KNN*

Radiación	Alto	Bajo	Extremo	Muy alto	Moderado
Alto	152	0	0	6	37
Bajo	0	7500	0	0	27
Extremo	0	0	4	11	0
Muy alto	21	0	0	93	0
Moderado	12	49	0	0	772
SUBTOTAL	185	7549	4	110	836
TOTAL			8684		

Tabla 8*Matriz de observación – KNN*

Clase	Medidas			
	TP	TN	FP	FN
Alto	152	8456	33	43
Bajo	7500	1108	49	27
Extremo	4	8669	0	11
Muy alto	93	8553	17	21
Moderado	772	7787	64	61
TOTALES	8521	34573	163	163

Para un total de 8521 registros, se identificó a 152 lugares con radiación alto de manera correcta y 43 de manera incorrecta, 7500 lugares con radiación de baja y 27 incorrectos; seguidamente de 4 espacios con radiación extremo y 11 incorrectos; así también 93 lugares tuvieron una radiación muy alto y 21 predicciones incorrectas. Finalmente 772 lugares tuvieron una radiación moderada con 61 incorrectos.

De igual manera, en la siguiente tabla se muestra las métricas de evaluación:

Tabla 9

Métricas de evaluación – KNN

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	$Accuracy = (TP+TN/TP+TN+FP+FN) * 100$	98 %
2	Especificidad	Razón	$Especificidad = (TN/TN+FP) * 100$	99 %
3	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	76 %
4	Recall		$recall=TP/TP+FN$	76 %
5	F 1 SCORE	Razón	$2 * (Recall * Precisión) / (Recall + Precisión)$	79 %

Interpretación: Como se observa en la Tabla 9, utilizando el algoritmo KNN se obtuvo una precisión igual a 98%, asimismo una especificidad de 99%, de igual manera para sensibilidad se obtuvo un valor de 76%, seguido de Recall con 76% y finalmente F1 score con 79%.

4.1.3. Regresión logística

Tabla 10

Matriz de confusión – Regresión logística

Radiación	Alto	Bajo	Extremo	Muy alto	Moderado
Alto	152	0	0	0	0
Bajo	0	7527	0	0	0
Extremo	0	0	15	0	0
Muy alto	0	0	0	114	0
Moderado	0	49	0	0	883
SUBTOTAL	152	7576	15	114	883
TOTAL			8740		

Tabla 11

Matriz de observación – Regresión logística

Clase	Medidas			
	TP	TN	FP	FN
Alto	152	8588	0	0
Bajo	7527	1164	49	0
Extremo	15	8725	0	0
Muy alto	114	8626	0	0
Moderado	883	7808	0	49
TOTALES	8691	34911	49	49

Para un total de 8691 registros, se identificó a 152 lugares con radiación alto de manera correcta y 0 de manera incorrecta, 7527 lugares con radiación baja y 0 incorrectos; seguidamente de 15 espacios con radiación extremo y 0 incorrectos; así también 114 lugares tuvieron una radiación muy alto y 0 predicciones incorrectas. Finalmente 883 lugares tuvieron una radiación moderada con 49 incorrectos.

Asimismo, en la siguiente tabla se muestra las métricas de evaluación:

Tabla 12

Métricas de evaluación – Regresión logística

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	Accuracy = $(TP+TN/TP+TN+FP+FN) * 100$	99 %
2	Especificidad	Razón	Especificidad = $(TN/TN+FP) * 100$	99 %
3	Sensibilidad	Razón	Sensibilidad = $(TP/TP+FN) * 100$	99 %
4	Recall		recall= $TP/TP+FN$	99 %
5	F 1 SCORE	Razón	$2 * (Recall * Precisión) / (Recall +$ Precisión)	99 %

Interpretación: Como se observa en la Tabla 12, utilizando el algoritmo Regresión logística se obtuvo una precisión igual a 99%, asimismo una especificidad de 99%, de igual manera para sensibilidad se obtuvo un valor de 99%, seguido de Recall con 99% y finalmente F1 score con 99%.

4.1.4. Redes Bayesianas

Tabla 13

Matriz de confusión – Redes bayesianas

Radiación	Alto	Bajo	Extremo	Muy alto	Moderado
Alto	168	0	0	13	14
Bajo	0	7331	0	0	195
Extremo	10	0	0	5	0
Muy alto	84	0	0	30	0
Moderado	0	124	0	0	709
SUBTOTAL	262	7455	0	48	918
TOTAL			8683		

Tabla 14

Matriz de observación – Redes bayesianas

Clase	Medidas			
	TP	TN	FP	FN
Alto	168	8394	94	27
Bajo	7331	1033	124	195
Extremo	0	8668	0	15
Muy alto	30	8551	18	84
Moderado	709	7641	209	124
TOTALES	8238	34287	445	445

Para un total de 8238 registros, se identificó a 168 lugares con radiación alto de manera correcta y 27 de manera incorrecta, 7331 lugares con radiación baja y 195 incorrectos; seguidamente de 0 espacios con radiación extremo y 15 incorrectos; así también 30 lugares tuvieron una radiación muy alto y 84 predicciones incorrectas. Finalmente 709 lugares tuvieron una radiación moderada con 124 incorrectos.

Asimismo, en la siguiente tabla se muestra las métricas de evaluación:

Tabla 15

Métricas de evaluación – Redes bayesianas

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	Accuracy = $(TP+TN/TP+TN+FP+FN) * 100$	95 %
2	Especificidad	Razón	Especificidad = $(TN/TN+FP) * 100$	98 %
3	Sensibilidad	Razón	Sensibilidad = $(TP/TP+FN) * 100$	59 %
4	Recall		recall= $TP/TP+FN$	59 %
5	F 1 SCORE	Razón	$2 * (Recall * Precisión) / (Recall +$ Precisión)	58 %

Interpretación: Como se observa en la Tabla 15, utilizando el algoritmo Redes bayesianas se obtuvo una precisión igual a 95%, asimismo una especificidad de 98%, de igual manera para sensibilidad se obtuvo un valor de 59%, seguido de Recall con 59% y finalmente F1 score con 58%.

4.1.5. Redes Neuronales

Tabla 16

Matriz de confusión – Redes neuronales

Radiación	Alto	Bajo	Extremo	Muy alto	Moderado
Alto	165	0	0	0	30
Bajo	0	7524	0	0	3
Extremo	9	0	0	6	0
Muy alto	114	0	0	0	0
Moderado	0	4	0	0	829
SUBTOTAL	288	7528	0	6	862
TOTAL			8684		

Tabla 17

Matriz de observación – Redes neuronales

Clase	Medidas			
	TP	TN	FP	FN
Alto	165	8366	123	30
Bajo	7524	1153	4	3
Extremo	0	8669	0	15
Muy alto	0	8564	6	114
Moderado	829	7818	33	4
TOTALES	8518	34570	166	166

Para un total de 8684 registros, se identificó a 165 lugares con radiación alto de manera correcta y 30 de manera incorrecta, 7524 lugares con radiación baja y 3 incorrectos; seguidamente de 0 espacios con radiación extremo y 15 incorrectos; así también 0 lugares tuvieron una radiación muy alto y 114 predicciones incorrectas. Finalmente 824 lugares tuvieron una radiación moderada con 4 incorrectos.

Asimismo, en la siguiente tabla se muestra las métricas de evaluación:

Tabla 18

Métricas de evaluación – Redes neuronales

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	Accuracy = $(TP+TN/TP+TN+FP+FN) * 100$	98 %
2	Especificidad	Razón	Especificidad = $(TN/TN+FP) * 100$	100 %
3	Sensibilidad	Razón	Sensibilidad = $(TP/TP+FN) * 100$	57 %
4	Recall		recall= $TP/TP+FN$	57 %
5	F 1 SCORE	Razón	$2 * (Recall * Precisión) / (Recall +$ Precisión)	53 %

Interpretación: Como se observa en la Tabla 18, utilizando el algoritmo Redes neuronales se obtuvo una precisión igual a 98%, asimismo una especificidad de 100%, de igual manera para sensibilidad se obtuvo un valor de 57%, seguido de Recall con 57% y finalmente F1 score con 53%.

4.1.6. SVM

Tabla 19

Matriz de confusión - SVM

Radiación	Alto	Bajo	Extremo	Muy alto	Moderado
Alto	175	0	0	8	12
Bajo	0	7516	0	0	11
Extremo	0	0	10	5	0
Muy alto	12	0	0	102	0
Moderado	7	49	0	0	816
SUBTOTAL	194	7565	10	115	839
TOTAL			8723		

Tabla 20

Matriz de observación - SVM

Clase	Medidas			
	TP	TN	FP	FN
Alto	175	8509	19	20
Bajo	7516	1147	49	11
Extremo	10	8708	0	5
Muy alto	102	8596	13	12
Moderado	816	7828	23	56
TOTALES	8619	34788	104	104

Para un total de 8723 registros, se identificó a 175 lugares con radiación alto de manera correcta y 20 de manera incorrecta, 7516 lugares con radiación baja y 11 incorrectos; seguidamente de 10 espacios con radiación extremo y 5 incorrectos; así también 102 lugares tuvieron una radiación muy alto y 12 predicciones incorrectas. Finalmente 816 lugares tuvieron una radiación moderada con 56 incorrectos.

Asimismo, en la siguiente tabla se muestra las métricas de evaluación:

Tabla 21

Métricas de evaluación - SVM

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	Accuracy = $(TP+TN/TP+TN+FP+FN) * 100$	99 %
2	Especificidad	Razón	Especificidad = $(TN/TN+FP) * 100$	99 %
3	Sensibilidad	Razón	Sensibilidad = $(TP/TP+FN) * 100$	88 %
4	Recall		recall= $TP/TP+FN$	88 %
5	F 1 SCORE	Razón	$2 * (Recall * Precisión) / (Recall +$ Precisión)	75 %

Interpretación: Como se observa en la Tabla 21, utilizando el algoritmo SVM se obtuvo una precisión igual a 99%, asimismo una especificidad de 99%, de igual manera para sensibilidad se obtuvo un valor de 88%, seguido de Recall con 88% y finalmente F1 score con 75%.

A continuación, se muestran los resultados obtenidos para las hipótesis específicas planteadas en esta investigación, a partir de la herramienta SPSS Statistics.

4.2.Hipótesis 1:

H1: ¿El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir con precisión la radiación UV-B en el distrito de la molina?

4.2.1. *Árbol de decisión*

Tabla 22

Métrica de evaluación H1– Precisión árbol de decisión

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	Accuracy = $(TP+TN/TP+TN+FP+FN) * 100$	100 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir con precisión la radiación UV-B con una precisión igual a 100 % utilizando el algoritmo de árbol de decisión.

4.2.2. *KNN – vecinos cercanos*

Tabla 23

Métrica de evaluación H1– Precisión KNN

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	Accuracy = $(TP+TN/TP+TN+FP+FN) * 100$	98 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir con precisión la radiación UV-B con una precisión igual a 98 % utilizando el algoritmo KNN – vecinos cercanos.

4.2.3. Regresión logística

Tabla 24

Métrica de evaluación H1 – Precisión regresión logística

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	Accuracy = $(TP+TN/TP+TN+FP+FN) * 100$	99 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir con precisión la radiación UV-B con una precisión igual a 99 % utilizando el algoritmo de regresión logística.

4.2.4. Redes bayesianas

Tabla 25

Métrica de evaluación H1 – Precisión redes bayesianas

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	Accuracy = $(TP+TN/TP+TN+FP+FN) * 100$	95 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir con precisión la radiación UV-B con una precisión igual a 95 % utilizando el algoritmo de redes bayesianas.

4.2.5. Redes neuronales

Tabla 26

Métrica de evaluación H1 – Precisión redes neuronales

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	Accuracy = $(TP+TN/TP+TN+FP+FN) * 100$	98 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir con precisión la radiación UV-B con una precisión igual a 98 % utilizando el algoritmo de redes neuronales.

4.2.6. SVM

Tabla 27

Métrica de evaluación H1 – Precisión SVM

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precisión (exactitud)	Razón	Accuracy = $(TP+TN/TP+TN+FP+FN) * 100$	99 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir con precisión la radiación UV-B con una precisión igual a 99 % utilizando el algoritmo SVM.

4.2.7. Cuadro comparativo de resultados según el indicador precisión

Tabla 28

Cuadro comparativo de resultados según el indicador precisión

Algoritmo	Resultado (%)
Árbol de decisión	100 %
K-NN	98 %
Regresión logística	99 %
Redes bayesianas	95 %
Redes neuronales	98 %
SVM	99 %

Nota. En la tabla 28 se evidenció que el algoritmo con mejor resultado en cuanto al indicador precisión que evaluó el porcentaje de radiación UV-B correctamente fue el “Árbol de decisión” con 100%, seguido de “Regresión logística” y “SVM” con 99%, asimismo el algoritmo “K-NN” y “Redes neuronales” con un resultado igual a 98 %. Finalmente “Redes bayesianas” con un resultado igual a 95%.

4.3.Hipótesis 2:

HE2: ¿El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir con sensibilidad la radiación UV-B en el distrito de la molina?

4.3.1. *Árbol de decisión*

Tabla 29

Métrica de evaluación H2 – Sensibilidad árbol de decisión

Ítem	Indicador	Medida	Fórmula	Sensibilidad
2	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	100 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir la radiación UV-B con una sensibilidad igual a 100 % utilizando el algoritmo de árbol de decisión.

4.3.2. *KNN – vecinos cercanos*

Tabla 30

Métrica de evaluación H2 – Sensibilidad KNN

Ítem	Indicador	Medida	Fórmula	Sensibilidad
2	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	76 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir la radiación UV-B con una sensibilidad igual a 76 % utilizando el algoritmo de KNN.

4.3.3. Regresión logística

Tabla 31

Métrica de evaluación H2 – Sensibilidad regresión logística

Ítem	Indicador	Medida	Fórmula	Sensibilidad
2	Sensibilidad	Razón	$\text{Sensibilidad} = (\text{TP}/\text{TP}+\text{FN}) * 100$	99 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir la radiación UV-B con una sensibilidad igual a 99 % utilizando el algoritmo de regresión logística.

4.3.4. Redes bayesianas

Tabla 32

Métrica de evaluación H2 – Sensibilidad redes bayesianas

Ítem	Indicador	Medida	Fórmula	Sensibilidad
2	Sensibilidad	Razón	$\text{Sensibilidad} = (\text{TP}/\text{TP}+\text{FN}) * 100$	59 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir la radiación UV-B con una sensibilidad igual a 59 % utilizando el algoritmo de redes bayesianas.

4.3.5. Redes neuronales

Tabla 33

Métrica de evaluación H2 – Sensibilidad redes neuronales

Ítem	Indicador	Medida	Fórmula	Sensibilidad
2	Sensibilidad	Razón	$\text{Sensibilidad} = (\text{TP}/\text{TP}+\text{FN}) * 100$	57 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir la radiación UV-B con una sensibilidad igual a 57 % utilizando el algoritmo de redes neuronales.

4.3.6. SVM

Tabla 34

Métrica de evaluación H2 – Sensibilidad SVM

Ítem	Indicador	Medida	Fórmula	Sensibilidad
2	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	88 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia artificial permitió medir la radiación UV-B con una sensibilidad igual a 88 % utilizando el algoritmo de SVM.

4.3.7. Cuadro comparativo de resultados según el indicador sensibilidad

Tabla 35

Cuadro comparativo de resultados según el indicador sensibilidad

Algoritmo	Resultado (%)
Árbol de decisión	100 %
K-NN	76 %
Regresión logística	99 %
Redes bayesianas	59 %
Redes neuronales	57 %
SVM	88 %

Nota. En la tabla 35 se evidenció que el algoritmo con mejor resultado en cuanto al indicador sensibilidad que evaluó el porcentaje de radiación UV-B correctamente fue el “Árbol de decisión” con un 100%, seguido del algoritmo “Regresión logística” con un resultado igual a 99 %; de igual manera “SVM” con 88%, asimismo “K-NN” con 76 %; “Reyes bayesianas” con 59%; y finalmente “Redes neuronales” con 57 %.

4.4.Hipótesis 3

HE3: ¿El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir con especificidad la radiación UV-B en el distrito de la Molina?

4.4.1. *Árbol de decisiones*

Tabla 36

Métrica de evaluación H3 – Especificidad árbol de decisión

Ítem	Indicador	Medida	Fórmula	Especificidad
3	Especificidad	Razón	$\text{Especificidad} = (\text{TN}/\text{TN}+\text{FP}) * 100$	100 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una especificidad igual a 100 % utilizando el algoritmo árbol de decisión.

4.4.2. *KNN – vecinos cercanos*

Tabla 37

Métrica de evaluación H3 – Especificidad KNN

Ítem	Indicador	Medida	Fórmula	Especificidad
3	Especificidad	Razón	$\text{Especificidad} = (\text{TN}/\text{TN}+\text{FP}) * 100$	99 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una especificidad igual a 99 % utilizando el algoritmo KNN.

4.4.3. Regresión logística

Tabla 38

Métrica de evaluación H3– Especificidad regresión logística

Ítem	Indicador	Medida	Fórmula	Especificidad
3	Especificidad	Razón	$\text{Especificidad} = (\text{TN}/\text{TN}+\text{FP}) * 100$	99 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una especificidad igual a 99 % utilizando el algoritmo de regresión logística.

4.4.4. Redes bayesianas

Tabla 39

Métrica de evaluación H3 – Especificidad redes bayesianas

Ítem	Indicador	Medida	Fórmula	Especificidad
3	Especificidad	Razón	$\text{Especificidad} = (\text{TN}/\text{TN}+\text{FP}) * 100$	98 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una especificidad igual a 98 % utilizando el algoritmo de redes bayesianas.

4.4.5. Redes neuronales

Tabla 40

Métrica de evaluación H3 – Especificidad redes neuronales

Ítem	Indicador	Medida	Fórmula	Especificidad
3	Especificidad	Razón	$\text{Especificidad} = (\text{TN}/\text{TN}+\text{FP}) * 100$	100 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una especificidad igual a 100 % utilizando el algoritmo de redes neuronales.

4.4.6. SVM

Tabla 41

Métrica de evaluación H3 – Especificidad SVM

Ítem	Indicador	Medida	Fórmula	Especificidad
3	Especificidad	Razón	$\text{Especificidad} = (\text{TN}/\text{TN}+\text{FP}) * 100$	99 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una especificidad igual a 99 % utilizando el algoritmo de SVM.

4.4.7. Cuadro comparativo de resultados según el indicador especificidad

Tabla 42

Cuadro comparativo de resultados según el indicador especificidad

Algoritmo	Resultado (%)
Árbol de decisión	100 %
K-NN	99 %
Regresión logística	99 %
Redes bayesianas	98 %
Redes neuronales	100 %
SVM	99 %

Nota. En la tabla 42 se evidenció que los algoritmos con mejor resultado en cuanto al indicador especificidad que evaluó el porcentaje de radiación UV-B correctamente fueron “Árbol de decisión” y “Redes neuronales” con 100 %, seguido de “K-NN”, “Regresión logística” y “SVM” con un 99%. Y finalmente “Redes bayesianas” con 98%.

4.5.Hipótesis 4

HE4: ¿El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir con recall la radiación UV-B en el distrito de la molina?

4.5.1. *Árbol de decisiones*

Tabla 43

Métrica de evaluación H4– Recall árbol de decisión

Ítem	Indicador	Medida	Fórmula	Recall
4	Recall	Razón	$\text{recall}=\text{TP}/\text{TP}+\text{FN}$	1

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una recall igual a 100 % utilizando el algoritmo árbol de decisión.

4.5.2. *KNN – vecinos cercanos*

Tabla 44

Métrica de evaluación H4– Recall KNN

Ítem	Indicador	Medida	Fórmula	Recall
4	Recall	Razón	$\text{recall}=\text{TP}/\text{TP}+\text{FN}$	0.76

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una Recall igual a 0.76 utilizando el algoritmo KNN.

4.5.3. Regresión logística

Tabla 45

Métrica de evaluación H4 – Recall regresión logística

Ítem	Indicador	Medida	Fórmula	Recall
4	Recall	Razón	$\text{recall} = \text{TP} / (\text{TP} + \text{FN})$	0.99

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una recall igual a 0.99 % utilizando el algoritmo de regresión logística.

4.5.4. Redes bayesianas

Tabla 46

Métrica de evaluación H4 – Recall redes bayesianas

Ítem	Indicador	Medida	Fórmula	Recall
4	Recall	Razón	$\text{recall} = \text{TP} / (\text{TP} + \text{FN})$	0.59

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una recall igual a 0.59 utilizando el algoritmo de redes bayesianas.

4.5.5. Redes neuronales

Tabla 47

Métrica de evaluación H4 – Recall redes neuronales

Ítem	Indicador	Medida	Fórmula	Recall
4	Recall	Razón	$\text{recall} = \text{TP} / (\text{TP} + \text{FN})$	0.57

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una recall igual a 0.57 utilizando el algoritmo de redes neuronales.

4.5.6. SVM

Tabla 48

Métrica de evaluación H4– Recall SVM

Ítem	Indicador	Medida	Fórmula	Recall
4	Recall	Razón	$\text{recall} = \text{TP}/\text{TP}+\text{FN}$	0.88

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una recall igual a 0.88 utilizando el algoritmo de SVM.

4.5.7. Cuadro comparativo de resultados según el indicador recall

Tabla 49

Cuadro comparativo de resultados según el indicador recall

Algoritmo	Resultado (%)
Árbol de decisión	100 %
K-NN	76 %
Regresión logística	99 %
Redes bayesianas	59 %
Redes neuronales	57 %
SVM	88 %

Nota. En la tabla 49 se evidenció que el algoritmo con mejor resultado en cuanto al indicador recall que evaluó el porcentaje de radiación UV-B correctamente fue el “Árbol de decisión” con un 100%, seguido del algoritmo “Regresión logística” con un resultado igual a 99 %; además “SVM” con 88%, asimismo “K-NN” con 76 %; “Reyes bayesianas” con 59%; y finalmente “Redes neuronales” con 57 %.

4.6.Hipótesis 5

HE5: ¿El desarrollo de un modelo utilizando técnicas de inteligencia permite medir con F1 score la radiación UV-B en el distrito de la molina?

4.6.1. *Árbol de decisiones*

Tabla 50

Métrica de evaluación H5– F1 Score árbol de decisión

Ítem	Indicador	Medida	Fórmula	F1 Score
5	F1 Score	Razón	$2 * (\text{Recall} * \text{Precisión}) / (\text{Recall} + \text{Precisión})$	100 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una F1 Score igual a 100 % utilizando el algoritmo árbol de decisión.

4.6.2. *KNN – vecinos cercanos*

Tabla 51

Métrica de evaluación H5 – F1 Score KNN

Ítem	Indicador	Medida	Fórmula	F1 Score
5	F1 Score	Razón	$2 * (\text{Recall} * \text{Precisión}) / (\text{Recall} + \text{Precisión})$	79 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una F1 Score igual a 79 % utilizando el algoritmo KNN.

4.6.3. Regresión logística

Tabla 52

Métrica de evaluación H5 – F1 Score regresión logística

Ítem	Indicador	Medida	Fórmula	F1 Score
5	F1 Score	Razón	$2 * (\text{Recall} * \text{Precisión}) / (\text{Recall} + \text{Precisión})$	99 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una F1 Score igual a 99 % utilizando el algoritmo de regresión logística.

4.6.4. Redes bayesianas

Tabla 53

Métrica de evaluación H5 – F1 Score redes bayesianas

Ítem	Indicador	Medida	Fórmula	F1 Score
5	F1 Score	Razón	$2 * (\text{Recall} * \text{Precisión}) / (\text{Recall} + \text{Precisión})$	58 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una F1 Score igual a 58 % utilizando el algoritmo de redes bayesianas.

4.6.5. Redes neuronales

Tabla 54

Métrica de evaluación H5 – F1 Score redes neuronales

Ítem	Indicador	Medida	Fórmula	F1 Score
5	F1 Score	Razón	$2 * (\text{Recall} * \text{Precisión}) / (\text{Recall} + \text{Precisión})$	53 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una F1 Score igual a 53 % utilizando el algoritmo de redes neuronales.

4.6.6. SVM

Tabla 55

Métrica de evaluación H5 – F1 Score SVM

Ítem	Indicador	Medida	Fórmula	F1 Score
5	F1 Score	Razón	$2 * (\text{Recall} * \text{Precisión}) / (\text{Recall} + \text{Precisión})$	75 %

Nota. El desarrollo de un modelo utilizando técnicas de inteligencia permitió medir la radiación UV-B con una *F1 Score* igual a 75 % utilizando el algoritmo de SVM.

4.6.7. Cuadro comparativo de resultados según el indicador F1 score

Tabla 56

Cuadro comparativo de resultados según el indicador F1 score

Algoritmo	Resultado (%)
Árbol de decisión	100 %
K-NN	79 %
Regresión logística	99 %
Redes bayesianas	58 %
Redes neuronales	53 %
SVM	75 %

Nota. En la tabla 56 se evidenció que el algoritmo con mejor resultado en cuanto al indicador F1 score que evaluó el porcentaje de radiación UV-B correctamente fue el “Árbol de decisión” con 100 %, seguido del algoritmo “Regresión logística” con 99%, asimismo “K-NN” con un resultado igual a 79%; “SVM” con 75 %; de igual manera “Reyes bayesianas” con 58 %. Y finalmente “Redes neuronales” con 53%.

V. DISCUSIÓN DE RESULTADOS

A continuación, se muestran las discusiones elaboradas en base a los resultados que se dieron durante la investigación.

Para este estudio se utilizó la metodología KDD, un método que te indica que hacer mas no el cómo hacerlo, por sus características es mucho más ágil por ende toma menor tiempo la implementación. Además, para desarrollar el modelo predictivo se utilizó como herramienta de modelado el SPSS Modeler por su interfaz amigable e interactivo, además por la documentación y comunidades de ayuda, esto permite realizar el desarrollo con mayor facilidad y rapidez, mientras que autores como Gil (2017) utilizó como metodología Box Jenkins con modelo ARIMA (1,0,0) y redes neuronales. Por su parte, Larrosa y Cauritongo (2019) utilizaron la metodología scrum.

Después de comparar entre los algoritmos: árbol de decisión, SVM, K-NN, regresión logística, redes bayesianas, redes neuronales, encontramos que el mejor modelo es generado por árbol de decisión, porque la sensibilidad, especificidad, precisión, recall y f1 score son del 100%, ya que es el más adecuado para predecir los resultados de radiación UV-B. Otros autores como Jiang et al. (2017) utilizaron un modelo de Markov (MS) con un enfoque de inferencia bayesiana para realizar las predicciones de intervalo y puntos diarios de la radiación solar cada dos horas. Husein & Chung (2019) realizó un estudio sobre el pronóstico diario de los rayos solares para microrredes usando una red neuronal recurrente de memoria a corto plazo.

En cuanto a los indicadores de este estudio se obtuvieron resultados positivos. En esta investigación se pudo observar una mejor precisión al utilizar “Árbol de decisión” con un resultado de 100% para predecir la radiación solar UV-B. Sin embargo, Koo et al (2019) estimó en promedio de cada mes en la radiación solar MADSR, utilizando la metodología de Clustering (k-means) y razonamiento en casos, teniendo como resultados que con ese método se obtuvo un 98.23% de exactitud de aproximación en la radiación. Mamani (2021), hizo una investigación sobre un modelo para pronosticar basado en minería de datos, tomado en cuenta ejemplos de la

investigación en radiación UV-B comparando 3 modelos y llegando a concluir que el mejor modelo es el de redes neuronales para poder realizar una buena predicción del modelo y la información estudiada, cumpliendo con la predicción por un periodo establecido para el nivel de aire. Salao (2021) dice que, al tener una herramienta adecuada, ayuda a conocer de manera exacta y medir de manera eficiente el colector solar UV-B, siendo muy eficaz.

También para el indicador sensibilidad para medir el modelo Machine Learning se obtuvo un porcentaje de 100% utilizando el algoritmo árbol de decisión, otros autores como Ravinesh (2019), ejecutó su estudio acerca de máquinas de aprendizaje integrados con satélites MODIS para el pronóstico de los rayos UV, usando como metodología la utilización de diferentes modelos: redes neuronales artificial (ANN), regresión lineal múltiple (MLR) y del promedio móvil integrado autorregresivo (ARIMA), y es así que los resultados arrojaron que la ANN prevaleció en los otros 2 modelos, mostrando el 39% de error en previsión acumulado, a diferencia de MLR con 15% , y ARIMA 25%.

Para el indicador especificidad en el estudio se alcanzó un porcentaje de 100% utilizando “Árbol de decisión” y “Redes neuronales”. Sin embargo, Park et al (2019) en sus estudios han obtenido resultado que el error medio absoluto encontrado en el espectrómetro y el aparato fue de 0,0034 W/m², usando ecuaciones de regresión lineal para su corrección se logró minimizar el error al más de 50%, mostrando un error de 0,0017 W/m², teniendo como conclusión que con el dispositivo se logra datos más precisos y seguros. De igual manera, Ravinesh (2019), utilizó redes neuronales artificial (ANN), regresión lineal múltiple (MLR) y del promedio móvil integrado autorregresivo (ARIMA), y es así que los resultados arrojaron que la ANN prevaleció en los otros 2 modelos, mostrando el 39% de error en previsión acumulado, a diferencia de MLR con 15%, y ARIMA 25%.

Asimismo, para el indicador recall se alcanzó un porcentaje de 100% utilizando el algoritmo de árboles de decisión. Por su parte, Mamani (2021) indica que siendo 10% de información, obteniendo como resultados que el modelo de redes neuronales con Memoria de

Largo y Corto Plazo (LSTM) obtuvo un mayor coeficiente de correlación de $R= 0.72$ y un % RMSE de 28 %, concluyendo que el mejor modelo es el de redes neuronales para poder realizar una buena predicción del modelo y la información estudiada, cumpliendo con la predicción por un periodo establecido para el nivel de aire.

Finalmente, para el indicador f1 score se consiguió una precisión del 100 % utilizando el algoritmo de árboles de decisión. Por su lado Pereira, Donandon & Oliveira (2021), dicen que esta métrica es la media armónica entre precisión y revocación, cuyos valores representan respectivamente los parámetros P y R. El parámetro β se puede utilizar para asignar diferentes pesos a las medidas utilizadas en la ecuación. En la evaluación de desempeño utilizada en este proyecto, β recibió un valor igual a 1; por lo tanto, la precisión y la renovación tienen la misma importancia.

Por lo tanto, cabe destacar la igual importancia de precisión y revocación en la ecuación de puntuación f1 - score. Además de ser preciso, el algoritmo debe detectar la mayor cantidad de radiaciones UV-B en una escena dada, expresada representada por la medida de rendimiento de la puntuación f1 obtenida en este proyecto a través de un equilibrio entre precisión y reversibilidad.

VI. CONCLUSIONES

- Se puede concluir que se utilizaron 6 algoritmos para predecir la radiación UV-B, siendo estos: Árbol de decisión, Regresión logística, SVM, K-NN, Redes neuronales y Redes bayesianas, el cual árbol de decisión fue el que permitió predecir con mayor porcentaje de precisión (100%), sensibilidad (100%), especificidad, Recall (100%) y F1 score (100%) de la radiación UV-B en el distrito de la molina.
- Se llegó a la conclusión que luego de la medición de las técnicas de inteligencia artificial con respecto a la precisión, obteniéndose los siguientes valores: árbol de decisión un porcentaje de 100%, regresión logística un porcentaje de 99%, SVM un porcentaje de 99%, K-NN un porcentaje de 98%, redes neuronales 98% y finalmente redes bayesianas en un 95% de la radiación UV-B en el distrito de la molina.
- Se concluyó después de la medición que las técnicas de inteligencia artificial sobre sensibilidad: árbol de decisión se obtuvo 100%, regresión logística con un porcentaje de 99%, SVM con un porcentaje de 88%, K-NN con un porcentaje de 76%, redes bayesianas 59% y finalmente redes neuronales en un 57% de la radiación UV-B en el distrito de la molina.
- Se pudo concluir, luego de medir las técnicas de inteligencia artificial, acerca de la especificidad, los siguientes valores: 100% utilizando árbol de decisión, redes neuronales un 100%, regresión logística con un porcentaje de 99%, SVM con un porcentaje de 99%, K-NN con un porcentaje de 99% y finalmente redes bayesianas en un 98% de la radiación UV-B en el distrito de la molina.
- Se concluye después de la medición que las técnicas de inteligencia artificial sobre la Recall, se obtuvieron los siguientes valores: árbol de decisión 100%, regresión logística un porcentaje de 99%, SVM un porcentaje de 88%, K-NN un porcentaje de 76%, redes bayesianas 59% y finalmente redes neuronales en un 57% de la radiación UV-B en el distrito de la molina.
- Se concluye después de la medición que las técnicas de inteligencia artificial sobre la F1 Score, los siguientes valores: árbol de decisión un porcentaje de 100%, regresión logística un porcentaje de 99%, K-NN un porcentaje de 79%, SVM un porcentaje de 75%, redes bayesianas 58% y finalmente redes neuronales en un 53% de la radiación UV-B en el distrito de la molina.

VII. RECOMENDACIONES

- Se recomienda usar la técnica de árbol de decisión ampliando el estudio en otros distritos con características similares dado que se obtuvo en la presente investigación mayor precisión, sensibilidad, especificidad, recall, f1 score.
- Crear modelos predictivos basados en las probabilidades obtenidas y de esta manera mejorar los resultados obtenidos con los modelos usados de manera independiente.
- Se aconseja crear un modelo predictivo empleando mas factores relacionados a la medición de radiación UV-B, como por ejemplo posición del sol, altitud, latitud cubrimiento de las nubes, cantidad de ozono en la atmosfera y reflexión terrestre.
- Se sugiere hacer estudios sobre la utilización de algoritmos combinados diferentes más factores relacionados a la medición de radiación UV-B, como por ejemplo posición del Sol, altitud, latitud, cubrimiento de las nubes, cantidad de ozono en la atmósfera y reflexión, como por ejemplo cascading, stacking, con la finalidad de mejorar el modelo creado en base a solo un algoritmo.
- Seguir haciendo uso de diferentes herramientas de machine learning para poder determinar de manera más detallada cuál sería el más adecuado para con respecto a la predicción de la radiación.
- Se sugiere hacer pruebas con el modelo realizado en diferentes países del mundo, para poder determinar la precisión en cada uno de ellos y ver su comportamiento, y así de generalizar los resultados.
- Se aconseja capacitar al personal encargado con respecto al modelo construido, consiguiendo el compromiso de todos para fomentar el liderazgo y las habilidades del personal para la prevención de enfermedades de piel a través de los datos correctos de radiación UV-B.

- Cuevas-Soto, V., Alvares-Iriarte, S., Azcona-Romero, M. y Rodríguez-Rogert, A. (2019). Capacidad predictiva de las Máquinas de Soporte Vectorial. Una aplicación en la planificación financiera. *Revista Cubana de Ciencias Informáticas*, 13(3), pp. 59-75. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992019000300059&lng=es&tlng=es
- Dedios, N. (2016). Radiación Ultravioleta. Análisis de su comportamiento estacional en diferentes sectores de la región Piura. Costa Norte del Perú. *Revista Ingenierías USB Med*, 7(1), pp. 26-30.
- De la Fuente, J. (2019). La inteligencia artificial y su aplicación en el mundo militar. *Inteligencia Artificial Aplicado a La Defensa*, pp. 69–98. Ministerio de Defensa. <http://publicacionesoficiales.boe.es/>
- Epstein, J., Sullivan, F., y Epstein, W. (1961). The effect of ultraviolet light on chemical carcinogenesis. *The Journal of Investigative Dermatology*, 36(2), pp. 73–77. <https://doi.org/10.1038/jid.1961.15>
- Espinoza, J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería, investigación y tecnología*, 21(1). pp. 1-13. <https://doi.org/10.22201/fi.25940732e.2020.21n1.008>
- Fahrni, T., Kuhn, M., Sommer, P., Wattenhofer, R. y Welten, S. (2011). Sundroid: Solar Radiation Awareness with Smartphones. *Computer Engineering and Networks Laboratory ETH Zurich, Switzerland*, pp. 365–374. <https://doi.org/10.1145/2030112.2030162>
- Fernández, C. (2014). Instrumentos de Investigación. <https://tiposdeinvestigacion.org/instrumentos-de-investigacion/>
- Florez-Morales, I., Bertel-Rodríguez, D., Correa-Londoño, L. y Velásquez-Lopera, M. (2018). Reporte de la experiencia de la Sección de Dermatología de la Universidad de Antioquia en la quinta jornada de detección de cáncer de piel, Asocolderma 2016. *IATREIA*, 31(4), pp. 362–370. <https://doi.org/10.17533/udea.iatreia.v31n4a03>

- Garnacho, G., Salido, R. y Moreno, J. (2020). Effects of solar radiation and an update on photoprotection. *Anales de pediatría*, 92(6), pp. 377.e1-377.e9. <https://doi.org/10.1016/j.anpedi.2020.04.014>
- Garsen, J., Norval, M., El-Ghorr, A., Gibbs, N. K., Jones, C. D., Cerimele, D., De Simone, C., Caffieri, S., Dall'Acqua, F., De Grujil, F. R., Sontag, Y. y Van Loveren, H. (1998). Estimation of the effect of increasing UVB exposure on the human immune system and related resistance to infectious diseases and tumours. *Journal of Photochemistry and Photobiology B: Biology*, 42(3), pp. 167–179. [https://doi.org/10.1016/S1011-1344\(97\)00122-X](https://doi.org/10.1016/S1011-1344(97)00122-X)
- Gil, C. (2017). *Modelo de predicción de irradiación solar ajustado al clima de Barranquilla, para el dimensionamiento de sistemas fotovoltaicos*. [Tesis de maestría, Universidad de la Costa]. Repositorio CUC. <https://repositorio.cuc.edu.co/bitstream/handle/11323/191/72200547.pdf?sequence=1&isAllowed=y>
- Gómez, G. (2020). Métodos y técnicas de investigación utilizados en los estudios sobre comunicación en España. *Mediterránea de comunicación*, 12(1), pp. 115-127. <https://doi.org/10.14198/MEDCOM000018>
- Gutiérrez-Mesías, L., Romero-Qwisgaard, A., Chávez-Untiveros, G., Palomino-Kobayashi, L., Moromisato-Shimabukuro, L. y Kitazono-Sugahara, A. (2017). Comparison of the Photoprotective Effects of Sunscreens Using Spectrophotometric Measurements or the Survivability of Yeast Cells Exposed to UV Radiation. *Rev. Soc. Quím Perú*, 83(3), pp. 294–307.
- Guillen, I. (2021). *Modelo de implementación para el monitoreo y control de condiciones ambientales basados con tecnologías de arduino y raspberry*. [Tesis de maestría, Universidad Nacional de San Agustín de Arequipa]. Repositorio UNSA. <http://190.119.145.154/bitstream/handle/20.500.12773/12677/UPgugoi%281%29.pdf?sequence=3&isAllowed=y>

- Hacker, E., Horsham, C., Vagenas, D., Jones, L., Lowe, J. y Janda, M. (2018). A mobile technology intervention with ultraviolet radiation dosimeters and smartphone apps for skin cancer prevention in young adults: Randomized controlled trial. *JMIR MHealth and UHealth*, 6(11). <https://doi.org/10.2196/mhealth.9854>
- Heo, S. Y., Kim, J., Gutruf, P., Banks, A., Wei, P., Pielak, R., Balooch, G., Shi, Y., Araki, H., Rollo, D., Gaede, C., Patel, M., Kwak, J. W., Peña-alcántara, A. E., Lee, K. y Yun, Y. (2018). Monitor exposure to solar radiation and to light for phototherapy. *Science Translational Medicine*, 10(470). <https://doi.org/10.1126/scitranslmed.aau1643>
- Hernández, R., Fernández, C. y Baptista, P. (2014). *Metodología de la investigación*. (6ta. ed.). McGraw Hill. <https://www.esup.edu.pe/wpcontent/uploads/2020/12/2.%20Hernandez.%20Fernandez%20y%20Baptista-Metodolog%C3%ADa%20Investigacion%20Cientifica%206ta%20ed.pdf>
- Husein, M. y Chung, I. Y. (2019). Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach. *Energies*, 12(10). <https://doi.org/10.3390/en12101856>
- Jiang, Y., Long, H., Zhang, Z. y Song, Z. (2017). Day-Ahead Prediction of Bihourly Solar Radiance with a Markov Switch Approach. *IEEE Transactions on Sustainable Energy*, 8(4), pp. 1536–1547. <https://doi.org/10.1109/TSTE.2017.2694551>
- Kolmakov, K., Winter, F. R. y Sednev, M. V. (2020). factors in their STED microscopy performance. *Photochemical & Photobiological Sciences*, 12, pp. 1615-1794. <https://doi.org/10.1039/d0pp00304b>
- Koo, C., Li, W., Hyun S. y Zhang, S. (2019). A novel estimation approach for the solar radiation potential with its complex spatial pattern via machine-learning techniques. *Renewable Energy*, 133, pp. 575–592. <https://doi.org/10.1016/j.renene.2018.10.066>.
- Larrosa-Carrera, A., Cauritongo-Valdiviezo, M. (2019). *Desarrollo de un aplicativo web utilizando sensores ambientales de hardware libre para la lectura de variables meteorológicas en tiempo real, en dos áreas metropolitanas diferentes de la ciudad de Guayaquil, localizadas en el Sur en las afueras de la aduana ecuatoriana y en el Norte en*

las afueras de la urbanización mucho lote 2, verificando la calidad de los datos con la estación meteorológica automatizada del INAMHI ubicada en la Facultad de CCNN.

[Tesis de pregrado, Universidad de Guayaquil]. Repositorio Institucional de la Universidad de Guayaquil. <http://repositorio.ug.edu.ec/handle/redug/39484>

Lee, J., Wang, W., Harrou, F. y Sun, Y. (2020). Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. *Energy Conversion and Management*, 208, 112582. <https://doi.org/10.1016/j.enconman.2020.112582>

López Costales, J. A. (26 de junio de 2017). Prevención y protección de la radiación solar. <http://www.ccooaytomadrid.es/documentos/general/saludlaboral/Solar.pdf>. Madrid: CCOO salud laboral.

Mamani Yanapa, W. (2021). Modelo de predicción a partir de la minería de datos basado en casos, vinculados al estudio de la radiación ultravioleta UV-B. Arequipa 2017 <http://www.repositorio.unjbg.edu.pe/handle/UNJBG/4323>

Marín-García, E., Alzate-Plaza, S. y Serna-Ruiz, A. (2020). Estación de medición de Rayos Ultravioleta energizado por un Sistema Fotovoltaico. *INGE CUC*, 16(2), pp. 163–179, 2020. <http://doi.org/10.17981/ingecuc.16.2.2020.12>

Mateus, S. P., González, N. y Branch, J. W. (2014). Aplicación de redes neuronales artificiales en entornos virtuales inteligentes. *Informacion Tecnologica*, 25(5), pp. 103–112. <https://doi.org/10.4067/S0718-07642014000500015>

Méndez, I., Ramírez, A., y Mora, R. (2020). Aprendizaje automático aplicado en física: Una revisión de la literatura científica. *Research in Computing Science* 149(8), pp. 803–816.

Méndez Palma, J. y Morales Marín, R. (2007). Inteligencia Artificial. *Inteligencia Artificial*, 2(6), pp. 1–33. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Inteligencia+Artificial:+un+enfoco+moderno#0>

Mujaico, A. (2020). *Diseño e implementación de un sistema de medición y pronóstico de radiación ultravioleta utilizando Internet de las cosas y Machine Learning*. [Tesis de maestría,

Universidad de Ciencias y Humanidades]. Repositorio UCH.
https://repositorio.uch.edu.pe/bitstream/handle/20.500.12872/521/Mujaico_AA_tesis_ingeneria_electronica_2020.pdf?sequence=1&isAllowed=y

Obando-Paredes, E., Carvajal-quintero, S. X. y Pineda, J. (2017). Comparación metodológica para pronosticar capacidad de generación de energía fotovoltaica basado en datos climáticos Methodological comparison to predict photovoltaic generation capacity based on climatic data. *Editorial Universidad Nacional*, 9, pp. 1–8.

Noruega, O., Pinto, R., y Villareal, J. (2017). La eficiencia de los nuevos materiales fotosensibles usados en la fabricación de paneles solares. *ITECKNE*, 15(1), pp. 7-16.

Olejua, C., y Navarro, A. (2020). Estimación de la Radiación Solar Mediante Métodos Empíricos en Bucaramanga, Colombia. *RINDERESU*, 5(2), pp. 724-733.

Ordoñez-Palacios, L., León-Vargas M. Sc, D., Bucheli-Guerrero Ph. D, V., y Ordoñez-Eraso Ph. D.H. (2020). Predicción de radiación solar en sistemas fotovoltaicos utilizando técnicas de aprendizaje automático. *Revista Facultad de Ingeniería*, 29 (54), e11751.
<https://doi.org/10.19053/01211129.v29.n54.2020.11751>

Ortiz Aguilar, W., Enrique Hevia, F., y García La Rosa, J. (2018). Pre-experimento para constatar cambios al aplicar un resultado científico. Ejemplo práctico de una metodología para mejorar el desarrollo de habilidades del pensamiento Geométrico Espacial. *Opuntia Brava*, 10(3), 329-346. Recuperado a partir de <http://opuntiabrava.ult.edu.cu/index.php/opuntiabrava/article/view/563>

Otzen, T. y Manterola, C. (2017). Técnicas de Muestreo sobre una Población a Estudio. *Int. J. Morphol*, 35(1), pp. 227-232.

Palaguachi, S. (2018). Diseño, Desarrollo E Implementación De Una Estación Meteorológica Basada En Una Red Jerárquica De Sensores, Software Libre Y Sistemas Embebidos Para La Empresa Elecaastro En La Minicentral Gualaceo Utilizando Comunicación Mqtt Y Modbus. Universidad Politécnica Salesiana Sede Cuenca, 148.
<https://dspace.ups.edu.ec/bitstream/123456789/5264/1/UPS-CT002757.pdf>

- Pang, Z., Niu, F., y O'Neill, Z. (2020). Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renewable Energy*, 156, pp. 279–289. <https://doi.org/10.1016/j.renene.2020.04.042>
- Park, D. H., Oh, S. T. y Lim, J. H. (2019). Development of a UV index sensor-based portable measurement device with the EUVB ratio of natural light. *Sensors (Switzerland)*, 19(4), pp. 1–12. <https://doi.org/10.3390/s19040754>
- Piekarski, M., Jaworek-Korjakowska, J., Wawrzyniak, A. y Gorgon M. (2020). Convolutional neural network architecture for beam instabilities identification in Synchrotron Radiation Systems as an anomaly detection problema. *Elsevier Ltd*, 165(0), pp. 1-13. <https://doi.org/10.1016/j.measurement.2020.108116>
- Polo Bravo, C., y Miranda-Espinoza, G. (2020). Cuantificación y caracterización de la radiación ultravioleta-b en la ciudad de Tacna, periodo: 2012-2014. *Tecnia*, 30(1), pp. 43-52. <https://dx.doi.org/10.21754/tecnica.v30i1.854>
- Quero, M. (2020). El bayesianismo: de lo casuístico a lo histórico O Bayesianismo: Do casuístico ao histórico. *Revista de Ciências Sociais, Humanas e Engenharias*, 6(1), 124-130. <https://www.redalyc.org/journal/5727/572763612011/html/>
- Wenmin Qin, Lunche Wang , Jing Wei, Bo Hu, Xun Liang. (s.f.). A novel efficient broadband model to derive daily surface solar Ultraviolet radiation (0.280–0.400 μm) <https://www.sciencedirect.com/science/article/abs/pii/S0048969720330308>
- Rodríguez T. y Terrones Luis (2020). Inteligencia artificial, responsabilidad y compromiso cívico y democrático * Inteligência artificial, responsabilidade e compromisso cívico e democrático Artificial Intelligence, Responsibility and Civic and Democratic Commitment. *Revista CTS*, 44(44), pp. 253–276. <http://ojs.revistacts.net/index.php/CTS/article/view/166>
- R. Raksasat, P. Sri iesaranusorn , J. Pemcharoen⁴, P. Laiwarin , S. Buntoung , S. Janjai , E. Boontaveeyuwat , P.Asawanonda , S. Sriswasdi y E. Chuangsuwanich. (s.f.). Accurate

surface ultraviolet radiation forecasting for clinical applications with deep neural network

<https://www.nature.com/articles/s41598-021-84396-2.pdf>

Ravinesh C., Mehmet, Ş. Adamowskic, J. and Jianchun, M. (s.f.) deployable extreme learning machines integrated with remotely sensed MODIS satellite predictors over Australia to forecast global solar radiation: A new approach

<https://www.sciencedirect.com/science/article/abs/pii/S1364032119300048>

Ríos, V., Montalvo, E., Tairo, R., Farfán, M., Lavilla, W. y Aedo, E. (2020). Modelo de regresión lineal múltiple del peso vivo de ovejas de raza asblack en función de variables zoométricas. *Revista El Ceprosimad*, 8(2), pp. 13-29.

Rodríguez, J. (2021). La influencia del Sol en la Tierra y otros planetas: Clima espacial. *Rev. Bras. Ensino Fis*, 43.

<https://www.scielo.br/j/rbef/a/YBVbzHZb4FSy5r9RsTxp9zM/?format=pdf&lang=es>

Rodríguez León, C. y García Lorenzo, M. (2016). Adecuación a metodología de minería de datos para aplicar a problemas no supervisados tipo atributo-valor. *Revista Universidad y Sociedad*, 8(4), pp. 43-53. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202016000400005&lng=es&tlng=es.

Romero, G., Macgluf, A., Rodríguez, L., Espinoza, J. y Suárez, A. (2021). Aplicación de Machine Learning en la industria 4.0 en tiempos de pandemia. *Interconectados saberes*, 6(11), pp. 1-7. <https://doi.org/10.25009/is.v0i11.2692>

Salao Bravo, J. (2021). *Desarrollo de un modelo con técnicas de inteligencia artificial para medir la eficiencia de un colector solar de tubos al vacío bajo la irradiancia de la ciudad de Riobamba*. [Tesis de maestría, Escuela Superior Politécnica de Chimborazo]. Repositorio ESPOCH. <http://dspace.esPOCH.edu.ec/handle/123456789/14627>

Sánchez Poma, R. (2018). *Diseño y construcción de medidor de radiación con seguimiento solar y registrador de datos*. [Tesis de maestría, Universidad Nacional del Centro del Perú]. Repositorio UNCP.

<https://repositorio.uncp.edu.pe/bitstream/handle/20.500.12894/4869/S%C3%A1nchez%20Poma.pdf?sequence=1>

Sánchez, H., Reyes, C. y Mejía, K. (2018). Manual de términos en investigación científica, tecnológica y humanística (1era. ed.). Universidad Ricardo Palma. <https://www.urp.edu.pe/pdf/id/13350/n/libro-manual-de-terminos-en-investigacion.pdf>

Schmalwieser, A. W., Schauburger, G., Janouch, M., Nunez, M., Koskela, T., Berger, D. y Karamanian, G. (2005). Global Forecast Model to Predict the Daily Dose of the Solar Erythemally Effective UV Radiation. *Photochemistry and Photobiology*, 81(1), pp. 154. <https://doi.org/10.1562/2003-12-03-ra-019.1>

Servicio Nacional de Meteorología e Hidrología del Perú [Senamhi]. (2022). *Pronóstico de radiación UV máximo (cielo despejado y mediodía solar) a nivel nacional*. <https://www.senamhi.gob.pe/?p=radiacion-uv>

Servicio Nacional de Meteorología e Hidrología del Perú [Senamhi]. (2022). *Datos Hidrometeorológicos a Nivel Nacional*. <https://www.senamhi.gob.pe/servicios/?p=estaciones>

Sulla-Torres, J., Gómez-Campos, R. y Cossio-Bolaños, M.A. (2018). Aplicación de un árbol de decisión difusa con clasificación de ambigüedad para determinar el exceso de peso en escolares. *Revista mexicana de ingeniería biomédica*, 39(2), pp. 128-143. <https://doi.org/10.17488/rmib.39.2.1>

Tan, K. C., Lim, H. S. y Mat Jafri, M. Z. (2018). Study on solar ultraviolet erythemal dose distribution over Peninsular Malaysia using Ozone Monitoring Instrument. *Egyptian Journal of Remote Sensing and Space Science*, 21(1), 105–110. <https://doi.org/10.1016/j.ejrs.2017.01.001>

Turner, J., Igoe, D., Parisi, A. V., McGonigle, A. J., Amar, A. y Wainwright, L. (2020). A review on the ability of smartphones to detect ultraviolet (UV) radiation and their potential to be

used in UV research and for public education purposes. *Science of the Total Environment*, 706, 135873. <https://doi.org/10.1016/j.scitotenv.2019.135873>

Vanegas, M., Villicaña, E. y Arrieta, L. (2015). Quantification and characterization of solar radiation at the department of La Guajira-Colombia by calculating atmospheric transmissivity. *Prospect*, 13(2), pp. 54-63.

Ventura-León, J. (2017) ¿Población o muestra?: Una diferencia necesaria. *Revista Cubana de Salud Pública*, 43(4).

Wang, H., Cai, R., Zhou, B., Aziz, S., Qin, B., Voropai, N., Gan, L. y Barakhtenko, E. (2020). Solar irradiance forecasting based on direct explainable neural network. *Energy Conversion and Management*, 226. <https://doi.org/10.1016/j.enconman.2020.113487>

IX. ANEXOS

Anexo 1: Matriz de consistencia

Tabla 57

Matriz de consistencia

Título	Problema	Objetivos	Hipótesis	Variable	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Escala de medición
	General	General	General	Independiente	Técnicas aplicadas al diagnóstico de enfermedades que han sido utilizadas en estudios de problemas complejos, alcanzando un aceptado grado de certeza en los resultados obtenidos con respecto a la identificación de un tipo específico de enfermedad. Estas aplicaciones son ventajosas debido a que facilitan la construcción y estudio de sistemas capaces de aprender a partir de un conjunto de datos y mejorar procesos de clasificación y predicción (Büyükçolpan y Tol, 2019).				
	¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir la radiación UV-B en el Distrito de La Molina?	Desarrollar un modelo mediante técnicas de inteligencia artificial para medir la radiación UV-B en el Distrito de La Molina.	El desarrollo de un modelo mediante técnicas de inteligencia artificial permite medir la radiación UV-B en el Distrito de La Molina.	Técnicas de inteligencia artificial					
	Específicos	Específicos	Específicos	Dependiente	Las redes neuronales y el aprendizaje automático se utilizan ampliamente en aplicaciones de radiación solar, evaluando el rendimiento de los perceptrones multicapa (MLP) y los árboles de decisión de energía combinándolos con la regresión lineal para estimar la energía solar (Ordoñez, León, Bucheli y Ordoñez, 2020).	Para medir la radiación UV-B se ha hecho uso de las métricas de precisión, la cual será obtenida a través de técnicas de inteligencia artificial en el Distrito de La Molina.	Métricas de precisión	Accuracy = $(TP+TN/TP+TN+F P+FN) * 100$	Razón
	¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir la precisión de la radiación UV-B en el Distrito de La Molina?	Determinar en qué porcentaje las técnicas de inteligencia artificial permite medir la precisión de la radiación UV-B en el Distrito de La Molina.	El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir la precisión de la radiación UV-B en el Distrito de La Molina.	Medición de radiación UV-B					
	¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir la sensibilidad de la radiación UV-B	Determinar en qué porcentaje las técnicas de inteligencia artificial permite medir la sensibilidad de la radiación UV-B	El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir la sensibilidad de la radiación UV-B					Sensibilidad = $(TP/TP+FN) * 100$	

<p>en el Distrito de La Molina?</p> <p>¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir la especificidad de la radiación UV-B en el Distrito de La Molina?</p>	<p>en el Distrito de La Molina.</p> <p>Determinar en qué porcentaje las técnicas de inteligencia artificial permite medir la especificidad de la radiación UV-B en el Distrito de La Molina.</p>	<p>en el Distrito de La Molina.</p> <p>El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir la especificidad de la radiación UV-B en el Distrito de La Molina.</p>				<p>Especificidad = $(TN / (TN + FP)) * 100$</p>	
<p>¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir el recall de la radiación UV-B en el Distrito de La Molina?</p>	<p>Determinar en qué porcentaje las técnicas de inteligencia artificial permite medir el recall de la radiación UV-B en el Distrito de La Molina.</p>	<p>El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir el recall de la radiación UV-B en el Distrito de La Molina.</p>				<p>Recall = $TP / (TP + FN)$</p>	
<p>¿Cómo el desarrollo de un modelo mediante técnicas de inteligencia artificial ayudará a medir el F1 score de la radiación UV-B en el Distrito de La Molina?</p>	<p>Determinar en qué porcentaje las técnicas de inteligencia artificial permite medir el F1 score de la radiación UV-B en el Distrito de La Molina.</p>	<p>El desarrollo de un modelo utilizando técnicas de inteligencia artificial permite medir el F1 score de la radiación UV-B en el Distrito de La Molina.</p>				<p>$2 * (Recall * Precisión) / (Recall + Precisión)$</p>	

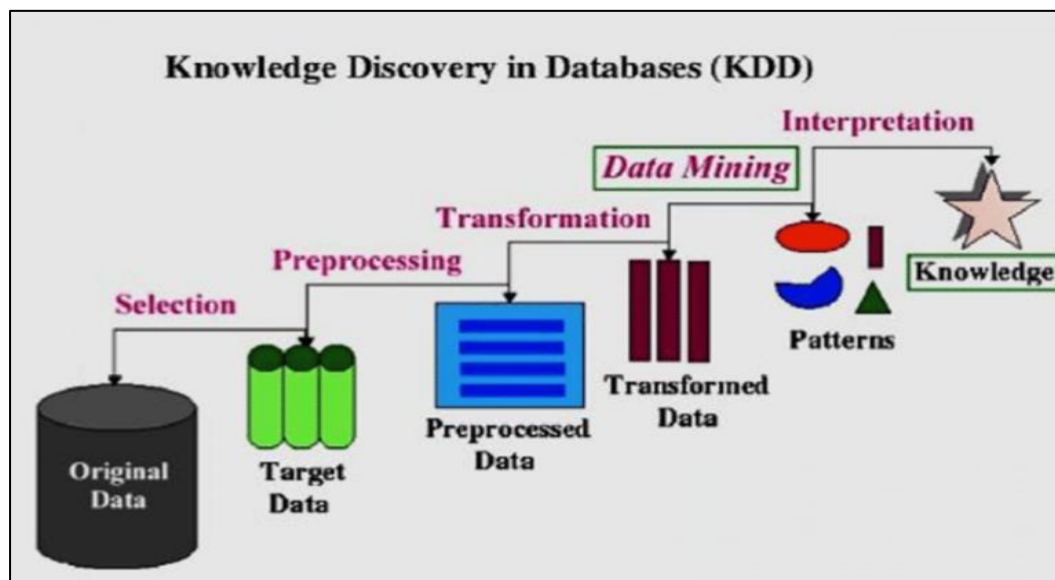
Nota. La tabla muestra la Matriz de consistencia del estudio “Desarrollo de un modelo para la medición de radiación UV-B utilizando técnicas de inteligencia artificial”

Anexo 2: Etapas del modelo desarrollado

Implementación aplicando Knowledge Discovery in Databases (KDD):

Figura 4:

Metodología Kdd



1.1. Pasos para la implementación del Sistema

Los pasos para la implementación son:

- Selección
- Procesamiento
- Transformación
- Data mining
- Interpretación

Detallados a continuación:

Selección:

Del conjunto de datos recolectados y ya definidos los objetivos por alcanzar, se deben elegir datos disponibles para realizar el estudio e integrarlos en uno solo que pueda favorecer a llegar a alcanzar a los objetivos del análisis. Se están usando más de 43 mil datos de SENAMHI con

datos meteorológicos diarios desde 2016 hasta el 2021 para optimizar el entrenamiento de la red neuronal y alcanzar una mejor Precisión.

Figura 5

Análisis exploratorio de datos

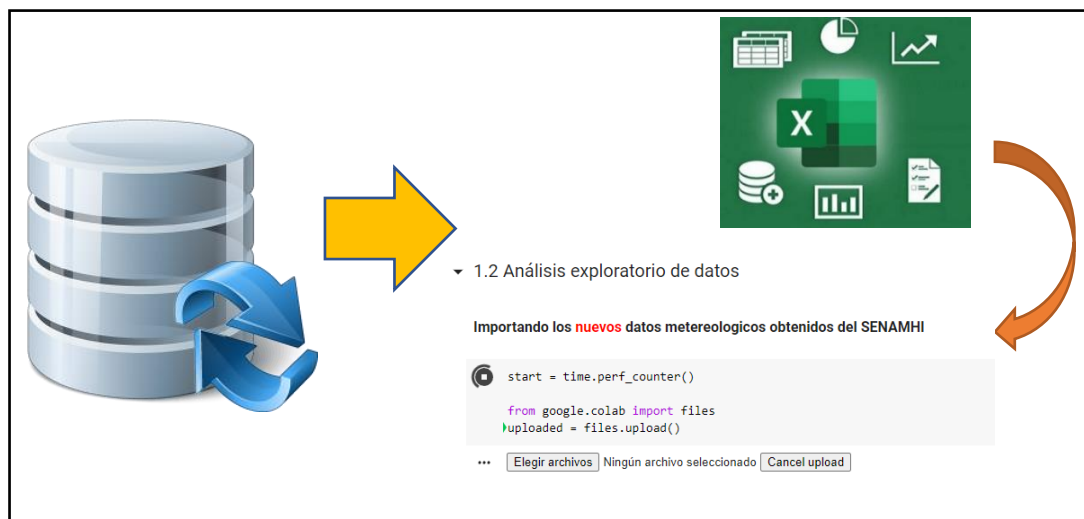


Figura 6

Visualización de datos extraídos

```
[ ] df.columns
Index(['date', 'time', 'prec_mmph', 'wind_dir', 'wind_speed', 'humidity',
      'temp_c'],
      dtype='object')

[ ] # Visualización de datos extraídos
df.head()

   date    time  prec_mmph  wind_dir  wind_speed  humidity  temp_c
0 2016-04-01 00:00:00     0.0     309         0.0         74     23.9
1 2016-04-01 01:00:00     0.0     264         1.7         77     23.2
2 2016-04-01 02:00:00     0.0     303         0.7         79     22.5
3 2016-04-01 03:00:00     0.0     241         0.0         79     22.4
4 2016-04-01 04:00:00     0.0     287         0.0         80     22.2

[ ] # mas de 43 mil datos (filas)
df.shape

(43425, 7)
```

Procesamiento:

Se hace la limpieza de datos (tratamiento de datos perdidos o remover valores atípicos). Esto implica eliminar variables o atributos con datos faltantes o eliminar información no útil para este tipo de tareas como el texto. Convirtiendo a las columnas numéricas de tipo String a tipo float para hacer operaciones matemáticas.

Figura 7

Limpieza de datos y conversión de datos

2.1 Limpieza de Datos

. Conjunto de datos de SENAMHI contiene "datos perdidos" en la forma de "S/D", toda fila que tenga "S/D" debe ser reemplazada por np.nan

```
[ ] df.replace('S/D', np.nan, inplace=True)
```

. Despues de que los datos perdidos sean reemplazados por np.nan, se puede hacer una limpieza de esas filas, lo cual ayuda a que el modelo de redes neuronales sea optimizado despues

```
[ ] df.dropna(inplace=True)
```

2.2 Conversión de Datos

. Convirtiendo a las columnas numéricas de tipo String a tipo float para hacer operaciones matemáticas

Ahora se esta incluyendo Precipitación

```
[ ] df['wind_dir'] = df['wind_dir'].astype(float)
df['wind_speed'] = df['wind_speed'].astype(float)
df['humidity'] = df['humidity'].astype(float)
df['temp_c'] = df['temp_c'].astype(float)
df['prec_mmph'] = df['prec_mmph'].astype(float)
```

Ahora se esta incluyendo las horas del día

. Convirtiendo la variable Categorica "time" de un formato horas a número entero






Figura 8

Convirtiendo de Celsius a Kelvin

FORMULA

°K = °C + 273.15

°K = Grados Kelvin

°C = Grados Centigrados

```
[ ] # Convirtiendo de Celsius a Kelvin
df["temp_k"] = df.temp_c + 273.15
```

```
[ ] df.head()
```

	date	time	prec_mmph	wind_dir	wind_speed	humidity	temp_c	temp_k
0	2016-04-01	0	0.0	309.0	0.0	74.0	23.9	297.05
1	2016-04-01	1	0.0	264.0	1.7	77.0	23.2	296.35
2	2016-04-01	2	0.0	303.0	0.7	79.0	22.5	295.65
3	2016-04-01	3	0.0	241.0	0.0	79.0	22.4	295.55
4	2016-04-01	4	0.0	287.0	0.0	80.0	22.2	295.35

Figura 9

Modelo de estimación de radiación solar.

```

2.3 Modelo de Estimación de Radiación Solar


$$R = 2.88204 (10^{-26}) e^{0.211622755T}$$


De acuerdo a la investigación hecha por Usha Bajpai y Kalpana Singh en su investigación "Estimation of instant solar radiation by using of instant temperature" (Singh y Bajpai, 2009, página 4) Hacer click aqui para abrir el pdf. La ecuación obtenida como modelo de radiación solar vs temperatura ambiental es:

R = E * e^{0.2117T}

Donde:
R = Radiación solar en [Watts/m²]
E = Constante Calculada por Investigación = 2.88204 x 10^{-26}
T = Temperatura Ambiental en grados Kelvin

# Creando las variables necesarias para implementar el modelo
E = 2.88204 * 10**(-26)
c = 0.211622755
df["temp_k"] = df.temp_k + 11
T = np.array(df.temp_k)

# Aplicando el modelo: R = E * e^{(0.2117 * T)}
# Guardandolo en nueva columna: "solar_radiation"
df["solar_radiation"] = E * np.exp(c * T)

```

Figura 10

Datos de radiación solar

	date	time	prec_mmph	wind_dir	wind_speed	humidity	temp_c	temp_k	solar_radiation
0	2016-04-01	0	0.0	309.0	0.0	74.0	23.9	308.05	590.917041
1	2016-04-01	1	0.0	264.0	1.7	77.0	23.2	307.35	509.555974
2	2016-04-01	2	0.0	303.0	0.7	79.0	22.5	306.65	439.397196
3	2016-04-01	3	0.0	241.0	0.0	79.0	22.4	306.55	430.196252
4	2016-04-01	4	0.0	287.0	0.0	80.0	22.2	306.35	412.368329
...
43420	2021-03-26	18	0.0	301.0	3.8	66.0	22.0	306.15	395.279220
43421	2021-03-26	19	0.0	288.0	2.6	73.0	20.6	304.75	293.923799
43422	2021-03-26	20	0.0	296.0	3.1	75.0	20.2	304.35	270.067369
43423	2021-03-26	21	0.0	288.0	1.4	78.0	19.3	303.45	223.231860
43424	2021-03-26	22	0.0	305.0	2.5	81.0	19.0	303.15	209.500083

43403 rows x 9 columns

Transformación

Dividiendo conjunto de datos, seleccionando las variables de entrenamiento Seleccionando la nueva variable Creada por nuestro modelo "solar_radiation" como variable objetivo a Predecir.

Figura 11

Conversión a Mw/m2

3.1 Conversión a mW/m2

```
[ ] df_mw = df.copy()
df_mw["solar_radiation_mw/m2"] = df_mw.solar_radiation / 10
df_mw
```

	date	time	prec_mmph	wind_dir	wind_speed	humidity	temp_c	temp_k	solar_radiation	solar_radiation_mw/m2
0	2016-04-01	0	0.0	309.0	0.0	74.0	23.9	308.05	590.917041	59.091704
1	2016-04-01	1	0.0	264.0	1.7	77.0	23.2	307.35	509.555974	50.955597
2	2016-04-01	2	0.0	303.0	0.7	79.0	22.5	306.65	439.397196	43.939720
3	2016-04-01	3	0.0	241.0	0.0	79.0	22.4	306.55	430.196252	43.019625
4	2016-04-01	4	0.0	287.0	0.0	80.0	22.2	306.35	412.368329	41.236833
...
43420	2021-03-26	18	0.0	301.0	3.8	66.0	22.0	306.15	395.279220	39.527922
43421	2021-03-26	19	0.0	288.0	2.6	73.0	20.6	304.75	293.923799	29.392380
43422	2021-03-26	20	0.0	296.0	3.1	75.0	20.2	304.35	270.067369	27.006737
43423	2021-03-26	21	0.0	288.0	1.4	78.0	19.3	303.45	223.231860	22.323186
43424	2021-03-26	22	0.0	305.0	2.5	81.0	19.0	303.15	209.500083	20.950008

43403 rows x 10 columns

Figura 12

Índice de radiación Ultravioleta (IUV)

3.2 Conversión de mW/m2 a "Índice de Radiación Ultravioleta" (IUV)

Conversión a UVI de acuerdo a "Solar radiation and the UV index: An application of numerical integration, trigonometric functions, online education and the modelling process" según Downs et al. (2016) [Hacer click aquí para abrir el PDF](#).

```
d = {range(0, 25) : 0,
range(25, 50) : 1,
range(50, 75) : 2,
range(75, 100) : 3,
range(100, 125) : 4,
range(125, 150) : 5,
range(150, 175) : 6,
range(175, 200) : 7,
range(200, 225) : 8,
range(225, 250) : 9,
range(250, 275) : 10,
range(275, 300) : 11,
range(300, 325) : 12,
range(325, 350) : 13,
range(350, 375) : 14,
range(375, 400) : 15,
range(400, 425) : 16,}

df_mw["UVI"] = df_mw['solar_radiation_mw/m2'].apply(lambda x: next((v for k, v in d.items() if int(x) in k), x))
```

Figura 13*Índice de radiación IUV*

	date	time	prec_mmph	wind_dir	wind_speed	humidity	temp_c	temp_k	solar_radiation	solar_radiation_mw/m2	UVI
0	2016-04-01	0	0.0	309.0	0.0	74.0	23.9	308.05	590.917041	59.091704	2
1	2016-04-01	1	0.0	264.0	1.7	77.0	23.2	307.35	509.555974	50.955597	2
2	2016-04-01	2	0.0	303.0	0.7	79.0	22.5	306.65	439.397196	43.939720	1
3	2016-04-01	3	0.0	241.0	0.0	79.0	22.4	306.55	430.196252	43.019625	1
4	2016-04-01	4	0.0	287.0	0.0	80.0	22.2	306.35	412.368329	41.236833	1
5	2016-04-01	5	0.0	295.0	0.0	82.0	21.9	306.05	387.002103	38.700210	1
6	2016-04-01	6	0.0	270.0	0.0	83.0	21.7	305.85	370.964205	37.096421	1
7	2016-04-01	7	0.0	263.0	0.0	84.0	22.0	306.15	395.279220	39.527922	1
8	2016-04-01	8	0.0	252.0	0.7	71.0	25.4	309.55	811.682918	81.168292	3
9	2016-04-01	9	0.0	252.0	4.1	62.0	27.2	311.35	1188.005023	118.800502	4
10	2016-04-01	10	0.0	261.0	5.1	59.0	28.2	312.35	1467.995998	146.799600	5
11	2016-04-01	11	0.0	253.0	5.7	52.0	29.9	314.05	2103.614147	210.361415	8
12	2016-04-01	12	0.0	237.0	5.0	53.0	29.9	314.05	2103.614147	210.361415	8
13	2016-04-01	13	0.0	249.0	6.6	48.0	30.5	314.65	2388.416372	238.841637	9
14	2016-04-01	14	0.0	235.0	7.3	46.0	30.7	314.85	2491.674792	249.167479	9

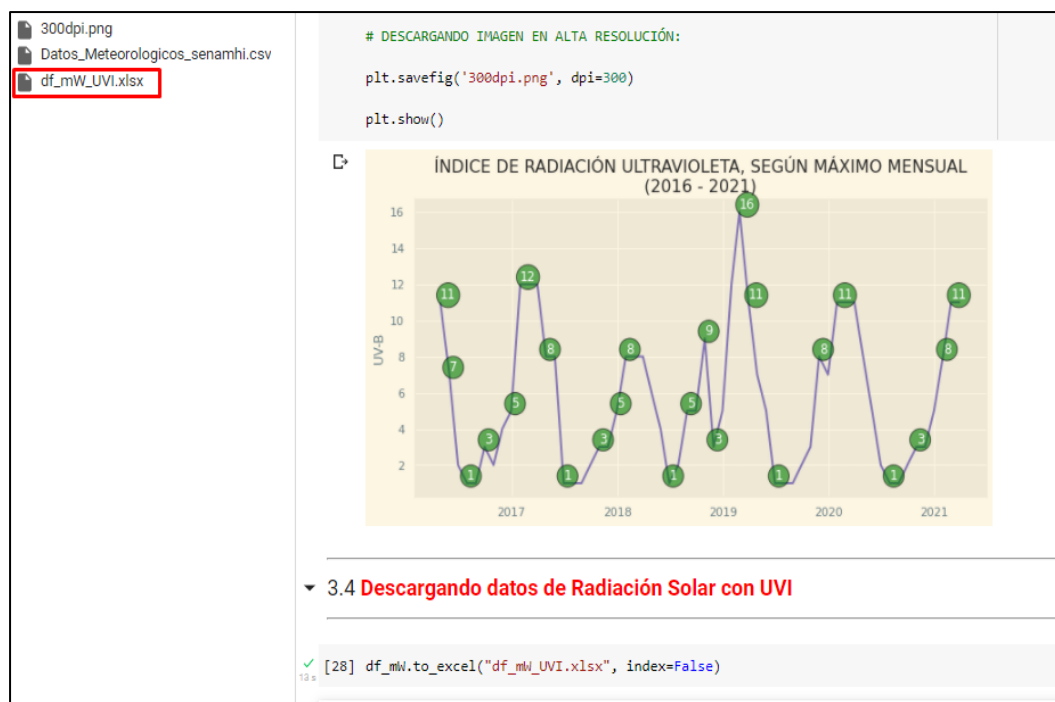
Figura 14*Descargando datos del modelo*

Figura 15
Modelamiento de árbol de decisión

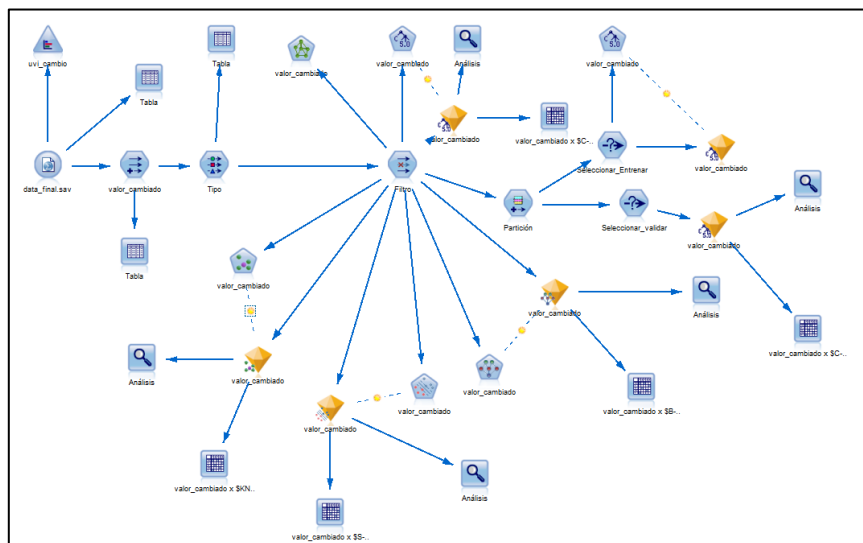


Figura 15
Análisis del modelo de radiación del árbol de decisión

Análisis de [valor_cambiado]

Archivo Editar

Análisis Anotaciones

Plegar todo Desplegar todo

Resultados para el campo de resultado valor_cambiado

Comparando \$C-valor_cambiado con valor_cambiado

'Partición'	2_Validar		
Correctos	8.684	100%	
Erróneos	0	0%	
Total	8.684		

Figura 17
Matriz de confusión de modelo de radiación de árbol de decisión

Matriz de valor_cambiado por \$C-valor_cambiado

Archivo Editar Generar

Matriz Aspecto Anotaciones

valor_cambiado	\$C-valor_cambiado				
	ALTO	BAJO	EXTREMO	MALTO	MODERADO
ALTO	195	0	0	0	0
BAJO	0	7527	0	0	0
EXTREMO	0	0	15	0	0
MALTO	0	0	0	114	0
MODERADO	0	0	0	0	833

Las casillas contienen: tabulación cruzada de campos (incluidos valores perdidos)

Chi-cuadrado = 34.736, gl = 16, probabilidad = 0

Figura 18
Modelamiento de Regresión logística

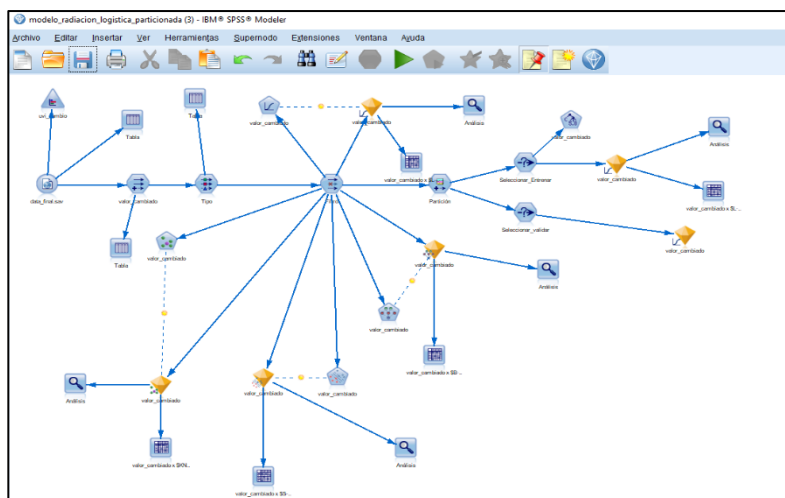


Figura 19
Análisis del modelo de radiación de la logística

The screenshot shows the 'Análisis de [valor_cambiado]' window. It displays a confusion matrix for the 'valor_cambiado' field, comparing the predicted values against the actual values. The matrix shows a high percentage of correct classifications (95.2%) and a low percentage of errors (4.8%).

Resultados para el campo de resultado valor_cambiado			
Comparando \$L-valor_cambiado con valor_cambiado			
Correctos	41.318		95,2%
Erróneos	2.085		4,8%
Total	43.403		

Figura 20
Matriz de confusión de modelo de radiación logística

The screenshot shows the 'Matriz de valor_cambiado por \$L-valor_cambiado' window. It displays a confusion matrix for the 'valor_cambiado' field, comparing the predicted values against the actual values. The matrix shows a high percentage of correct classifications (95.2%) and a low percentage of errors (4.8%).

\$L-valor_cambiado					
valor_cambiado	ALTO	BAJO	EXTREMO	MALTO	MODERADO
ALTO	131	0	0	45	800
BAJO	0	37025	0	554	97
EXTREMO	48	0	1	8	0
MALTO	414	0	0	82	5
MODERADO	0	105	0	9	4079

Las casillas contienen: tabulación cruzada de campos (incluidos valores perdidos)
Chi-cuadrado = 71.013,568, gl = 16, probabilidad = 0

Figura 21
Modelamiento de KNN

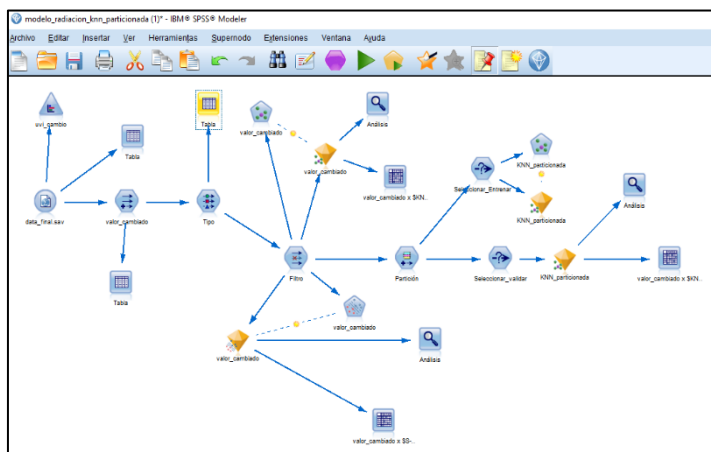


Figura 22
Análisis del modelo de radiación del KNN

Análisis de [valor_cambiado] #2

Archivo Editar

Análisis Anotaciones

Plegar todo Desplegar todo

Resultados para el campo de resultado valor_cambiado

Comparando \$KNN-valor_cambiado con valor_cambiado

'Partición'	2_Validar		
Correctos	8.521	98,12%	
Erróneos	163	1,88%	
Total	8.684		

Figura 16
Matriz de confusión de modelo de radiación de KNN

Matriz de valor_cambiado por \$KNN-valor_cambiado

Archivo Editar Generar

Matriz Aspecto Anotaciones

\$KNN-valor_cambiado

valor_cambiado	ALTO	BAJO	EXTREMO	MALTO	MODERADO
ALTO	152	0	0	6	37
BAJO	0	7500	0	0	27
EXTREMO	0	0	4	11	0
MALTO	21	0	0	93	0
MODERADO	12	49	0	0	772

Las casillas contienen: tabulación cruzada de campos (incluidos valores perdidos)

Chi-cuadrado = 22.129,802, gl = 16, probabilidad = 0

Aceptar

Figura 24
Modelamiento de Redes Bayesianas

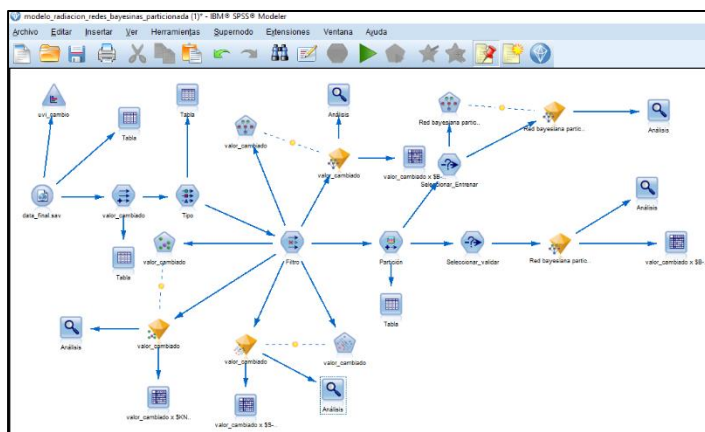


Figura 25
Análisis del modelo de radiación de redes bayesianas

The screenshot shows a window titled 'Análisis de [valor_cambiado] #4'. It has a menu bar (Archivo, Editar) and buttons for 'Análisis' and 'Anotaciones'. Below the menu, there are buttons for 'Plegar todo' and 'Desplegar todo'. The main content area displays the following text and table:

Resultados para el campo de resultado valor_cambiado
Comparando \$B-valor_cambiado con valor_cambiado

'Partición'	2_Validar	
Correctos	8.238	94,86%
Erróneos	446	5,14%
Total	8.684	

An 'Aceptar' button is located at the bottom right of the window.

Figura 26
Matriz de confusión de modelo de radiación de red Bayesiana

The screenshot shows a window titled 'Matriz de valor_cambiado por \$B-valor_cambiado'. It has a menu bar (Archivo, Editar, Generar) and buttons for 'Matriz', 'Aspecto', and 'Anotaciones'. The main content area displays a confusion matrix table:

		\$B-valor_cambiado			
valor_cambiado	\$null\$	ALTO	BAJO	MALTO	MODERADO
ALTO	0	168	0	13	14
BAJO	1	0	7331	0	195
EXTREMO	0	10	0	5	0
MALTO	0	84	0	30	0
MODERADO	0	0	124	0	709

Below the table, the text reads: 'Las casillas contienen: tabulación cruzada de campos (incluidos valores perdidos)'. At the bottom, it states: 'Chi-cuadrado = 14.378,121, gl = 16, probabilidad = 0'. An 'Aceptar' button is at the bottom right.

Figura 27
Modelamiento de Redes Neuronales

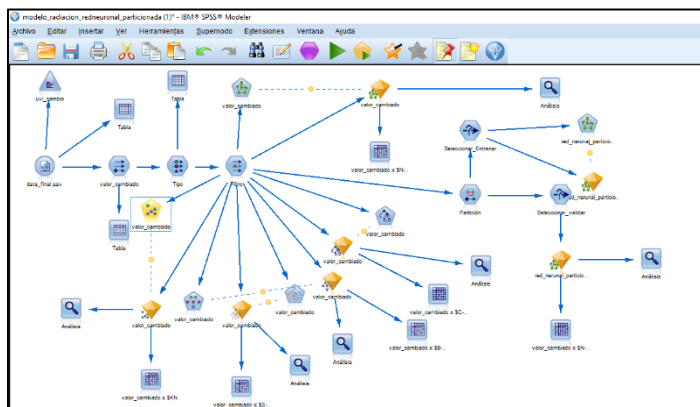


Figura 28
Precisión del modelo de radiación de la red neuronal de decisión

Análisis de [valor_cambiado] #5

Archivo Editar

Análisis Anotaciones

Plegar todo Desplegar todo

Resultados para el campo de resultado valor_cambiado

Comparando \$N-valor_cambiado con valor_cambiado

'Partición'	2_Validar	
Correctos	8.518	98,09%
Erróneos	166	1,91%
Total	8.684	

Aceptar

Figura 29
Matriz de confusión de modelo de radiación de red neuronal

Matriz de valor_cambiado por \$N-valor_cambiado

Archivo Editar Generar

Matriz Aspecto Anotaciones

\$N-valor_cambiado				
valor_cambiado	ALTO	BAJO	MALTO	MODERADO
ALTO	165	0	0	30
BAJO	0	7524	0	3
EXTREMO	9	0	6	0
MALTO	114	0	0	0
MODERADO	0	4	0	829

Las casillas contienen: tabulación cruzada de campos (incluidos valores perdidos)

Chi-cuadrado = 19.633,537, gl = 12, probabilidad = 0

Aceptar

Figura 30
Modelamiento de SVM

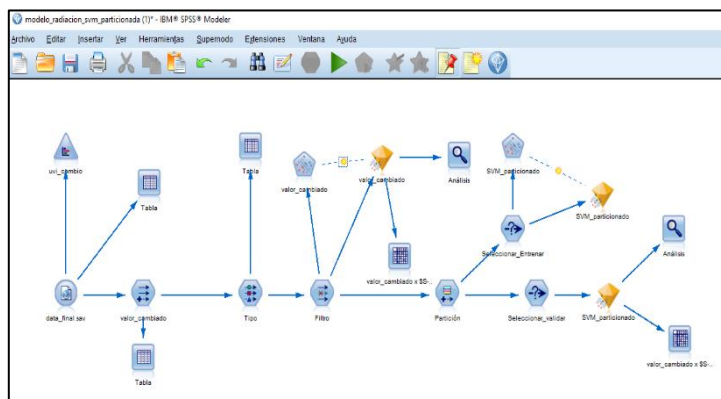


Figura 31
Análisis del modelo de radiación de SVM de decisión

Análisis de [valor_cambiado] #6

Archivo Editar

Análisis Anotaciones

Plegar todo Desplegar todo

Resultados para el campo de resultado valor_cambiado

Comparando \$S-valor_cambiado con valor_cambiado

'Partición'	2_Validar		
Correctos	8.619	99,25%	
Erróneos	65	0,75%	
Total	8.684		

Aceptar

Figura 32
Matriz de confusión de modelo de radiación del SVM

Matriz de valor_cambiado por \$S-valor_cambiado

Archivo Editar Generar

Matriz Aspecto Anotaciones

\$S-valor_cambiado

valor_cambiado	ALTO	BAJO	EXTREMO	MALTO	MODERADO
ALTO	175	0	0	8	12
BAJO	0	7516	0	0	11
EXTREMO	0	0	10	5	0
MALTO	12	0	0	102	0
MODERADO	7	10	0	0	816

Las casillas contienen: tabulación cruzada de campos (incluidos valores perdidos)

Chi-cuadrado = 28.178,085, gl = 16, probabilidad = 0

Aceptar

ANEXO 3 Validación de Instrumentos

EXPERTO 1

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Precision
Experto 1	Dr. Alfredo Daza Vergaray
DNI:	40466240

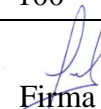
Ítem	Indicador	Medida	Fórmula	Precisión
1	Precision	Razón	$Accuracy = (TP+TN/TP+TN+FP+FN) * 100$	



Firma

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Sensibilidad
Experto 1	Dr. Alfredo Daza Vergaray
DNI:	40466240

Ítem	Indicador	Medida	Fórmula	Sensibilidad
2	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	



Firma

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Especificidad
Experto 1	Dr. Alfredo Daza Vergaray
DNI:	40466240

Ítem	Indicador	Medida	Fórmula	Sensibilidad
3	Especificidad	Razón	$Especificidad = (TN/TN+FP) * 100$	



Firma

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Recall
Experto 1	Dr. Alfredo Daza Vergaray
DNI:	40466240

Ítem	Indicador	Medida	Fórmula	Recall
4	Recall	Razón	$\text{recall} = \text{TP} / (\text{TP} + \text{FN})$	

Firma


Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	F1 Score
Experto 1	Dr. Alfredo Daza Vergaray
DNI:	40466240

Ítem	Indicador	Medida	Fórmula	F1 Score
5	F1 Score	Razón	$2 * (\text{Recall} * \text{Precisión}) / (\text{Recall} + \text{Precisión})$	

Firma


EXPERTO 2

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Precision
Experto 1	Phd. Victor Manuel Cabrejos Yalan
DNI:	42052260

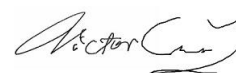
Ítem	Indicador	Medida	Fórmula	Precisión
1	Precision	Razón	$Accuracy = (TP+TN/TP+TN+FP+FN) * 100$	



Firma

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Sensibilidad
Experto 1	Phd. Victor Manuel Cabrejos Yalan
DNI:	42052260

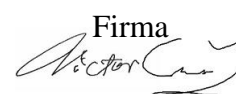
Ítem	Indicador	Medida	Fórmula	Sensibilidad
2	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	



Firma

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Especificidad
Experto 1	Phd. Victor Manuel Cabrejos Yalan
DNI:	42052260

Ítem	Indicador	Medida	Fórmula	Sensibilidad
3	Especificidad	Razón	$Especificidad = (TN/TN+FP) * 100$	



Firma

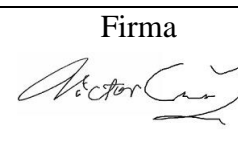
Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Recall
Experto 1	Phd. Victor Manuel Cabrejos Yalan
DNI:	42052260

Ítem	Indicador	Medida	Fórmula	Recall
4	Recall	Razón	$\text{recall} = \text{TP} / (\text{TP} + \text{FN})$	

Firma


Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	F1 Score
Experto 1	Phd. Victor Manuel Cabrejos Yalan
DNI:	42052260

Ítem	Indicador	Medida	Fórmula	F1 Score
5	F1 Score	Razón	$2 * (\text{Recall} * \text{Precisión}) / (\text{Recall} + \text{Precisión})$	

Firma


EXPERTO 3

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Precision
Experto 1	Dr. Ivan carlo Petrlik azabache
DNI:	10140461

Ítem	Indicador	Medida	Fórmula	Precisión
1	Precision	Razón	$Accuracy = (TP+TN/TP+TN+FP+FN) * 100$	




Firma

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Sensibilidad
Experto 1	Dr. Ivan carlo Petrlik azabache
DNI:	10140461

Ítem	Indicador	Medida	Fórmula	Sensibilidad
2	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	




Firma

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Especificidad
Experto 1	Dr. Ivan carlo Petrlik azabache
DNI:	10140461

Ítem	Indicador	Medida	Fórmula	Sensibilidad
3	Especificidad	Razón	$Especificidad = (TN/TN+FP) * 100$	




Firma

Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	Recall
Experto 1	Dr. Ivan carlo Petrlik azabache
DNI:	10140461

Ítem	Indicador	Medida	Fórmula	Recall
4	Recall	Razón	$\text{recall} = \text{TP} / (\text{TP} + \text{FN})$	



Título del Proyecto	Desarrollo de un modelo para la medición de radiación UV-B Utilizando técnicas de inteligencia artificial
Responsable de investigación	Giancarlo Sanchez Atuncar
Indicador	F1 Score
Experto 1	Dr. Ivan carlo Petrlik azabache
DNI:	10140461

Ítem	Indicador	Medida	Fórmula	F1 Score
5	F1 Score	Razón	$2 * (\text{Recall} * \text{Precisión}) / (\text{Recall} + \text{Precisión})$	

