



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE
SISTEMAS

Algoritmo de árbol de decisión con metodología crisp y su
influencia en la predicción de lluvias, Chosica 2024

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTORES:

Espinoza Valdez, Patrick Edson (orcid.org/0000-0002-1127-992X)

Rodriguez Trujillo, Axel Esteban (orcid.org/0000-0001-8979-2785)

ASESOR:

Mg. Galvez Tapia, Orleans Moises (orcid.org/0000-0002-4352-9495)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo sostenible y adaptación al cambio climático

LIMA – PERÚ

2024



Declaratoria de Autenticidad del Asesor

Yo, GALVEZ TAPIA ORLEANS MOISES, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA ESTE, asesor de Tesis titulada: "Algoritmo de Árbol de Decisión con Metodología Crisp y su Influencia en la Predicción de Lluvias, Chosica 2024", cuyos autores son RODRIGUEZ TRUJILLO AXEL ESTEBAN, ESPINOZA VALDEZ PATRICK EDSON, constato que la investigación tiene un índice de similitud de 15%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 06 de Julio del 2024

Apellidos y Nombres del Asesor:	Firma
GALVEZ TAPIA ORLEANS MOISES DNI: 16798332 ORCID: 0000-0002-4352-9495	Firmado electrónicamente por: GORLEANSM el 12- 08-2024 08:46:30

Código documento Trilce: TRI - 0798640



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Declaratoria de Originalidad de los Autores

Nosotros, ESPINOZA VALDEZ PATRICK EDSON, RODRIGUEZ TRUJILLO AXEL ESTEBAN estudiantes de la de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA ESTE, declaramos bajo juramento que todos los datos e información que acompañan la Tesis titulada: "Algoritmo de Árbol de Decisión con Metodología Crisp y su Influencia en la Predicción de Lluvias, Chosica 2024", es de nuestra autoría, por lo tanto, declaramos que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. Hemos mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

En tal sentido asumimos la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de la información aportada, por lo cual nos sometemos a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

Nombres y Apellidos	Firma
RODRIGUEZ TRUJILLO AXEL ESTEBAN DNI: 72704670 ORCID: 0000-0001-8979-2785	Firmado electrónicamente por: ARODRIGUEZL2 el 15-07-2024 18:19:31
ESPINOZA VALDEZ PATRICK EDSON DNI: 75368918 ORCID: 0000-0002-1127-992X	Firmado electrónicamente por: PESPINOZAVA el 16-07-2024 14:06:25

Código documento Trilce: INV - 1778404

Dedicatoria

Esta investigación es dedicada a mi familia que me ha brindado todo su apoyo, especialmente mi padre, Alex Rodríguez el cual ha dado todo de sí para que yo pueda llegar a cumplir mi meta de ser un profesional. También mencionar a Keveen Alonso Febres Silva quien más que un amigo es un hermano con el que siempre podré contar en mi vida y Lizett Denisse Ramírez Gutiérrez que es una de las personas a las que le guardo un gran cariño, ya que ella me demuestra todo lo bueno que aún hay en el mundo.

Agradecimiento

Agradecer especialmente a la universidad por brindarnos la todas las facilidades para el aprendizaje, asimismo a todo el plantel docente por su dedicación y fe en el alumnado.

Índice de contenidos

Carátula	i
Declaratoria de autenticidad del asesor.....	ii
Declaratoria de originalidad del autor/ autores.....	iii
Dedicatoria	iv
Agradecimiento	v
Índice de contenidos.....	vi
Índice de tablas	vii
Resumen	viii
Abstract	ix
I. INTRODUCCIÓN	1
II. METODOLOGÍA	6
III. RESULTADOS.....	15
IV. DISCUSIÓN	26
V. CONCLUSIONES	27
VI. RECOMENDACIONES.....	28
REFERENCIAS.....	30
ANEXOS	37

Índice de tablas

Tabla 01 - Matriz de operacionalización	10
---	----

Resumen

El problema de esta investigación es ¿cuál será la influencia del uso del Algoritmo de árbol de decisión con metodología crisp en la precisión de la predicción de lluvias en Chosica?. El objetivo de la investigación fue determinar cuál será la influencia del Algoritmo de árbol de decisión en la predicción de lluvias en Chosica. El tipo de investigación fue aplicada con diseño preexperimental, la población y muestra fueron los datos sobre las lluvias de los últimos 5 años en Chosica. El instrumento fue la base de datos extraída de la web de SENAMHI, la cual contenía la información meteorológica de los últimos 5 años. Esta información fue utilizada para probar la efectividad del algoritmo de árbol de decisión al momento de realizar predicciones. Con esto se podrá concluir que El Algoritmo de árbol de decisión con metodología crisp influencia de manera significativa a la predicción de lluvias en Chosica.

Palabras clave: Algoritmos, crisp, lluvias, predicción.

Abstract

The problem of this research is what will be the influence of the use of the decision tree algorithm with crisp methodology on the accuracy of rainfall prediction in Chosica? The objective of the research was to determine the influence of the decision tree algorithm on rainfall prediction in Chosica. The type of research was applied with a pre-experimental design, the population and sample were the data on the rains of the last 5 years in Chosica. The instrument was the database extracted from the SENAMHI website, which contained meteorological information for the last 5 years. This information was used to test the effectiveness of the decision tree algorithm together when making predictions. With this it can be concluded that the decision tree algorithm significantly influences the prediction of rainfall in Chosica.

Keywords: Algorithms, crisp, rains, prediction.

I. INTRODUCCIÓN

Hoy en día se destaca que los algoritmos juegan un papel crucial en la gestión de desastres naturales. Estos sistemas, alimentados por datos, ayudan a prever y mitigar eventos extremos. Sin embargo, su efectividad depende de la calidad y diversidad de los datos utilizados. Por ello es fundamental desarrollar algoritmos transparentes y éticos para maximizar su potencial en la protección de vidas y la preservación del medio ambiente (BONILLA, 2024).

Los algoritmos también pueden evaluar el tipo de suelo, los datos meteorológicos y la elevación del terreno para anticipar el daño causado por desastres naturales y prever zonas en riesgo de inundaciones, colapsos de edificios y otros daños (KIM, 2021). La planificación basada en estos algoritmos se presenta como una herramienta valiosa para la seguridad pública y la prevención de pérdidas humanas, esto en el contexto de países como Australia, Canadá, Corea y Estados Unidos.

ALI Y SUCIPTO (2021) señalan que los desastres naturales pueden abordarse mediante el uso de herramientas como algoritmos en subdistritos de Indonesia, respaldando la idea de que la utilización de algoritmos en la predicción de lluvias puede ser efectiva para mitigar y tomar medidas preventivas frente a eventos climáticos extremos.

Cronista (2023) destaca que la inteligencia artificial es fundamental en la lucha contra el cambio climático. Con algoritmos avanzados, analiza datos climáticos para prever tendencias y entender mejor los efectos del cambio climático. Además, se usa en la optimización de energías renovables y la gestión eficiente de la energía, promoviendo un futuro más sostenible.

En el ámbito nacional, durante el mes de abril de 2023, las lluvias torrenciales en Piura dejaron a la población aislada debido a daños en las carreteras (VILELA, 2023). Estos eventos no solo afectan a la población en general, sino también a los dueños de negocios y servicios, ya que pueden sufrir pérdidas directas o indirectas debido a los desastres naturales.

La zona de Chosica, al estar en una zona propensa a múltiples desastres naturales debido a su ubicación en el cinturón de fuego del Pacífico y su diversidad climática, enfrenta desafíos significativos (BUENDIA, 2023). Este contraste se evidenció durante un aumento repentino de lluvias en abril, que causó víctimas y daños materiales (EFE NEWS SERVICE, 2023).

La respuesta de las autoridades ante estos desastres suele ser criticada por su lentitud o ineficacia, dependiendo de las zonas afectadas (EL COMERCIO, 2023). En marzo, una encuesta del IEP reveló que el 62% de la población consideraba pésimas las acciones del gobierno frente a las inundaciones causadas por las lluvias, lo que reflejaba la necesidad de mejorar la efectividad de las respuestas gubernamentales (EL COMERCIO, 2023).

SENAMHI brinda de manera mensual predicciones referentes a las precipitaciones, teniendo estas un porcentaje del 30 al 60 por ciento de acierto (SENAMHI, 2023). Con esto vemos que si bien existen mecanismos de predicciones este no tiene un porcentaje de acierto muy elevado, además de ser muy variable.

En una serie de investigaciones recientes, una variedad de académicos han aplicado técnicas avanzadas de análisis de datos y aprendizaje automático para abordar problemas complejos y proporcionar soluciones innovadoras en múltiples campos. SÁNCHEZ (2023) informó sobre los devastadores efectos del ciclón Yaku en Perú, destacando la respuesta proactiva del Senamhi que implementó una plataforma para pronósticos de lluvias en tiempo real, ayudando a las ciudades afectadas como Lima y Piura a prepararse mejor ante futuras emergencias climáticas.

CESPEDES (2022) y INDIRA (2022) utilizaron métodos de Random Forest y Árboles de Decisión para abordar el problema de la anemia infantil. Mientras CESPEDES se concentró en mejorar la precisión de las predicciones de anemia utilizando datos históricos y un conjunto reducido de variables relevantes, INDIRA implementó una herramienta diagnóstica en Arequipa que permite

detectar de manera eficiente la anemia en niños basándose en datos clínicos y demográficos.

Por su parte, GARCIA (2020) llevó a cabo dos estudios importantes utilizando árboles de decisión y la metodología KDD para entender y mitigar la deserción escolar en Perú. Estos estudios no solo identificaron factores críticos que influyen en la deserción, sino que también ofrecieron soluciones prácticas para mejorar la retención de estudiantes.

ALEJANDRO y FIGUEROA (2021) exploraron la predicción de fenómenos meteorológicos extremos como las descargas eléctricas atmosféricas en las comunidades de Marcapomacocha y Huayao, utilizando análisis estadístico de datos de satélite, lo cual contribuye a mejorar las estrategias de prevención frente a estos eventos peligrosos. BENDEZU (2020) investigó las capacidades de distintos tipos de bases de datos, proporcionando a las organizaciones herramientas para seleccionar la tecnología más adecuada que asegure la integridad y la agilidad operativa de sus datos.

En el campo de la salud, CORTEZ (2022) desarrolló un modelo predictivo para la recomendación de cirugía refractiva, optimizando el proceso de toma de decisiones clínicas basado en algoritmos que consideran múltiples variables de pacientes. Simultáneamente, PORTUGUEZ (2022) diseñó un modelo para evaluar y mejorar el rendimiento académico de estudiantes de ingeniería de sistemas, aplicando técnicas de minería de datos para extraer insights valiosos que pueden ser utilizados en la mejora continua de procesos educativos.

A nivel internacional, investigadores como FLORES et al. (2021), ATMAJA (2022), y SRIVASTAVA (2021) han empleado inteligencia artificial para perfeccionar la predicción de eventos naturales y la gestión de desastres, desde la predicción de lluvias hasta la evaluación de la vulnerabilidad de zonas frente a desastres naturales. THÉO LE GUENEDAL Y TANKOV (2022) introdujeron un algoritmo revolucionario para la predicción de ciclones, elevando la precisión y ofreciendo una mejor comprensión de los impactos potenciales y las medidas de mitigación necesarias.

En base a lo expresado líneas arriba se puede plantear la interrogante principal: ¿Cuál será la influencia del uso del Algoritmo de árbol de decisión en la precisión de la predicción de lluvias en Chosica?. Para llegar a la respuesta de la misma se debe responder a las incógnitas de: ¿De qué forma el Algoritmo de Árbol de Decisión incrementa la exactitud de la predicción de lluvias en Chosica?, ¿De qué manera el algoritmo de árbol de decisión aumenta el número de aciertos en la predicción de lluvias en Chosica? y ¿De qué manera el algoritmo de árbol de decisión incrementa el recall en la predicción de lluvias en Chosica?.

Para justificar la presente investigación se tomaron en cuenta argumentos como que durante cada temporada de lluvias la población de Chosica se ve afectada por los efectos adversos del mismo en la zona, causando pérdidas materiales y gastos adicionales sobrellevar dichos efectos. Esto se verá reducido gracias al uso del Algoritmo de Árbol de Decisión Lluvias en las predicciones, el cual permitirá tener una visión más clara de la temporada de lluvias y cuando la población deberá tomar medidas preventivas sobre las lluvias en lugar de realizar gastos de última hora.

Al nivel institucional este algoritmo se pone a disposición de las autoridades e instituciones relacionadas para su uso e incorporación a los sistemas de prevención y predicción ya existentes. Esto permitirá recuperar la credibilidad y confianza de la población hacia las autoridades.

Los árboles de decisión proporcionan una metodología robusta para gestionar la incertidumbre inherente a la predicción meteorológica, además permiten una evaluación más completa de los factores que influyen en la formación de lluvias al considerar múltiples variables y relaciones complejas entre ellas. Esto facilita el proceso de predicción y que este sea más adaptativo.

Los árboles de decisión son una elección apropiada en el área de predicciones debido a su capacidad para manejar grandes volúmenes de datos de manera eficiente y su relativa simplicidad de implementación y mantenimiento. Estos algoritmos son altamente interpretables, lo que permite a los meteorólogos

entender y explicar el proceso de predicción de manera efectiva. Además, su escalabilidad y capacidad de adaptación los hacen adecuados para su aplicación en diferentes escalas geográficas y temporales, así como para la actualización continua con nuevos datos para mejorar la precisión de las predicciones.

Habiendo justificado la investigación y planteado la problemática de la misma se planteó el objetivo principal de la misma el cual es Determinar cuál será la influencia del Algoritmo de árbol de decisión en la predicción de llluvias en Chosica. Y para ello se planteó lo objetivos específicos de Determinar cómo el algoritmo de árbol de decisión en la incrementa la exactitud de predicción de llluvias en Chosica, Demostrar cómo el algoritmo de árbol de decisión aumenta el número de aciertos de la predicción de llluvias en Chosica y Establecer cómo el uso del algoritmo de árbol de decisión incrementa el recall en la predicción de llluvias en Chosica.

En última instancia, el estudio busca verificar la hipótesis general de que el Algoritmo de árbol de decisión influye de manera significativa en la predicción de llluvias en Chosica. Y las hipótesis específicas de que el algoritmo de árbol de decisiones incrementa la exactitud de predicciones en Chosica, el algoritmo de árbol de decisión aumenta el número de aciertos en la predicción de llluvias en Chosica y el uso del algoritmo de árbol de decisión incrementa el recall en la predicción de llluvias en Chosica.

II. METODOLOGÍA

El tipo de investigación será aplicada según lo mencionado por el Vicerrectorado de Investigación UCV (2023), ya que se busca solucionar un problema de realidad mediante la revisión de conocimientos relacionados y a partir de ellos generar nuevo conocimiento y usarlo en la investigación para resolver dicho problema. Esto apoyado a lo mencionado por RIOJA(2020) que nos dice que una investigación puede ser aplicada si esta desarrolla un producto.

El enfoque será cuantitativo en un nivel explicativo. Esto se apoya con lo mencionado por CÉSPEDES (2022), que nos dice que los estudios cuantitativos se basan en medición de características a partir de un marco conceptual para generalizar y normalizar resultados, además de ser acorde a lo mencionado por Vicerrectorado de Investigación UCV (2020).

El diseño de investigación será pre experimental, ya que se comparará la exactitud de las predicciones, antes y después de aplicar el algoritmo de árbol de decisión. Esto se apoya en lo mencionado por MONTALVAN(2021), el cual comenta que este diseño permite obtener información del grupo poblacional escogido antes y después de aplicar la solución propuesta por una investigación con el fin de proporcionar evidencia sobre los resultados obtenidos.

Como variables de la investigación se tomaron tanto una independiente como otra dependiente que serían algoritmo de árbol de decisión y la predicción de lluvias respectivamente.

El algoritmo de árbol de decisión de manera conceptual es una técnica de aprendizaje automático que se basa en la creación de una estructura de árbol jerárquica para la toma de decisiones predictivas. Esta técnica divide repetidamente el conjunto de datos en subconjuntos más pequeños en función de características específicas, representadas por nodos en el árbol. Estos nodos contienen condiciones o reglas que guían la separación de datos, y las ramas representan las posibles opciones de decisión (TIMARAN-BUCHELY, 2021. Durante el proceso de construcción del árbol, el algoritmo evalúa las

características del conjunto de datos y selecciona las divisiones que maximizan la homogeneidad dentro de los subconjuntos resultantes, con el objetivo de generar nodos que sean altamente predictivos para la variable objetivo. Una vez construido, el árbol puede ser utilizado para predecir valores de la variable objetivo para nuevas instancias de datos, al seguir el camino desde el nodo raíz hasta las hojas del árbol, donde se toman las decisiones finales basadas en las características observadas.

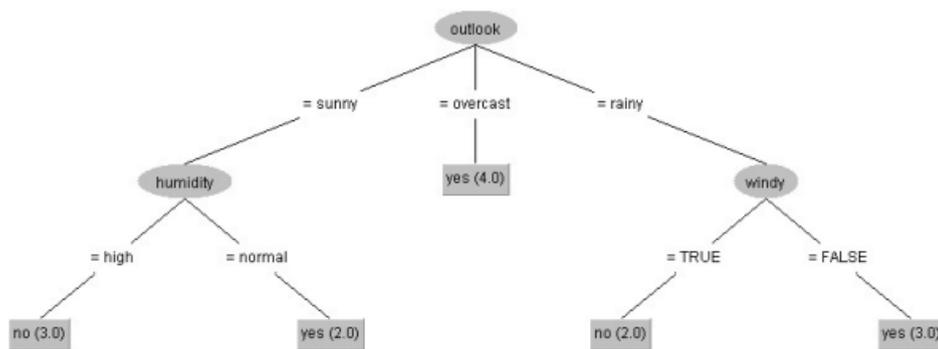
Algoritmo de árbol de decisión de forma operacional es un método de aprendizaje automático que construye un modelo predictivo mediante la creación de una estructura jerárquica en forma de árbol. Este árbol se compone de nodos que representan características o atributos del conjunto de datos y de ramas que representan las posibles decisiones o resultados basados en esas características (TIMARAN-BUCHELY, 2021). Durante el proceso de construcción del árbol, el algoritmo evalúa las características del conjunto de datos y selecciona la mejor división en cada nodo para maximizar la precisión de las predicciones. El árbol resultante puede ser utilizado para predecir la variable objetivo para nuevas instancias o casos no vistos, siguiendo las decisiones tomadas en cada nivel del árbol, lo que lo convierte en una herramienta útil para la clasificación y la regresión en problemas de aprendizaje supervisado.

De manera un poco más detallada FLORES(2021) menciona que los árboles de decisión son estructuras de datos en forma de árbol que se utilizan para la toma de decisiones y la clasificación. Este consta de nodos y ramas. Cada nodo interno del árbol representa una "pregunta" o prueba sobre un atributo (característica) del conjunto de datos. Las ramas que salen de un nodo representan las posibles respuestas a esa pregunta. Los nodos al final de cada rama (nodos hoja) representan las decisiones finales o las clasificaciones de los datos.

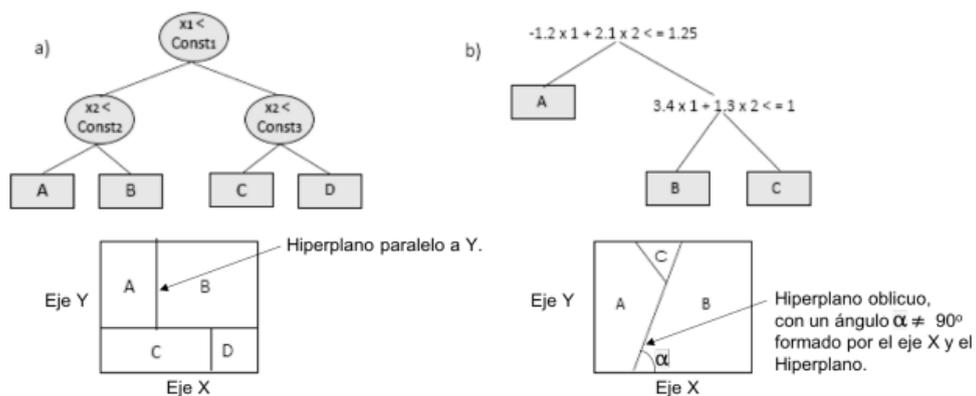
FLORES(2021) menciona también que a través de un proceso de inducción, el espacio de datos es dividido de forma que datos similares terminen juntos. Se seleccionan atributos que mejor dividan el conjunto de datos en clases más homogéneas según algún criterio de pureza. A partir de un nodo raíz, el árbol se

construye de manera recursiva, seleccionando en cada paso el atributo que mejor particione el conjunto de datos actual, hasta que se cumplen ciertos criterios de parada, como que todos los datos en un nodo sean de una sola clase o que se alcance una profundidad máxima del árbol.

Estos árboles pueden ser univariados o paralelos al Eje, donde cada nodo considera solo un atributo, las divisiones en el árbol son paralelas a los ejes del espacio de atributos. También pueden ser multivariados o oblicuos: Se consideran combinaciones lineales de múltiples atributos para la división en cada nodo, lo que puede resultar en divisiones oblicuas respecto a los ejes del espacio de atributos (FLORES, 2021).



Arbol de decision univariable



Arbol de decision Multivariable

Una vez construido, el árbol puede ser usado para clasificar nuevos datos. Comenzando desde el nodo raíz y siguiendo las ramas según las respuestas de las pruebas en cada nodo, se llega a un nodo hoja que proporciona la clasificación del dato (FLORES, 2021).

Por otro lado la predicción de lluvias con un enfoque conceptual es el proceso mediante el cual se utiliza un conjunto de características inherentes a las lluvias, como lo planteado por QUIJADA (2020), para crear un patrón que permita anticipar cuándo, dónde y qué impacto tendrán en un territorio determinado. Este proceso se ve potenciado por la incorporación de algoritmos y metodologías específicas, los cuales analizan estas características para proporcionar un panorama más preciso y exacto sobre las condiciones climáticas futuras relacionadas con la lluvia.

La predicción de lluvias de manera operacional basado en lo mencionado por QUIJADA(2020), la predicción de lluvias se refiere a la capacidad de un algoritmo y una metodología específicos para analizar y procesar patrones identificables en las condiciones atmosféricas, con el fin de anticipar con mayor precisión y certeza los eventos de lluvia. Esta capacidad se materializa en la capacidad del algoritmo y la metodología para proporcionar un panorama más exacto, basado en un análisis detallado de las características inherentes a los eventos de lluvia, lo que permite una interpretación más precisa y detallada de las condiciones atmosféricas y, por ende, una predicción más precisa de la ocurrencia de lluvias.

VARIABLE	DEFINICIÓN CONCEPTUAL	DEFINICION OPERACIONAL	DIMENSIONES	INDICADOR	TIPO DE VARIABLE
ALGORITMO DE ÁRBOL DE DECISIÓN	El algoritmo de árbol de decisión es una técnica de aprendizaje automático que se basa en la creación de una estructura de árbol jerárquica para la toma de decisiones predictivas. Esta técnica divide repetidamente el conjunto de datos en subconjuntos más pequeños en función de características específicas, representadas por nodos en el árbol. Estos nodos contienen condiciones o reglas que guían la separación de datos, y las ramas representan las posibles opciones de decisión (TIMARAN-BUCHELY, 2021). Durante el proceso de construcción del árbol, el algoritmo evalúa las características del conjunto de datos y selecciona las divisiones que maximizan la homogeneidad dentro de los subconjuntos resultantes, con el objetivo de generar nodos que sean altamente predictivos para la variable objetivo. Una vez construido, el árbol puede ser utilizado para predecir valores de la variable objetivo para nuevas instancias de datos, al seguir el camino desde el nodo raíz hasta las hojas del árbol, donde se toman las decisiones finales basadas en las características observadas.	El algoritmo de árbol de decisión es un método de aprendizaje automático que construye un modelo predictivo mediante la creación de una estructura jerárquica en forma de árbol. Este árbol se compone de nodos que representan características o atributos del conjunto de datos y de ramas que representan las posibles decisiones o resultados basados en esas características (TIMARAN-BUCHELY, 2021). Durante el proceso de construcción del árbol, el algoritmo evalúa las características del conjunto de datos y selecciona la mejor división en cada nodo para maximizar la precisión de las predicciones. El árbol resultante puede ser utilizado para predecir la variable objetivo para nuevas instancias o casos no vistos, siguiendo las decisiones tomadas en cada nivel del árbol, lo que lo convierte en una herramienta útil para la clasificación y la regresión en problemas de aprendizaje supervisado.	RESULTADOS	EXACTITUD	CUANTITATIVA
PREDICCIÓN DE LLUVIAS	La predicción de lluvias puede conceptualizarse como el proceso mediante el cual se utiliza un conjunto de características inherentes a las lluvias, como lo planteado por Quijada (2020), para crear un patrón que permita anticipar cuándo, dónde y qué impacto tendrán en un territorio determinado. Este proceso se ve potenciado por la incorporación de algoritmos y metodologías específicas, los cuales analizan estas características para proporcionar un panorama más preciso y exacto sobre las condiciones climáticas futuras relacionadas con la lluvia.	Basado en lo mencionado por QUIJADA(2020), la predicción de lluvias se refiere a la capacidad de un algoritmo y una metodología específicos para analizar y procesar patrones identificables en las condiciones atmosféricas, con el fin de anticipar con mayor precisión y certeza los eventos de lluvia. Esta capacidad se materializa en la capacidad del algoritmo y la metodología para proporcionar un panorama más exacto, basado en un análisis detallado de las características inherentes a los eventos de lluvia, lo que permite una interpretación más precisa y detallada de las condiciones atmosféricas y, por ende, una predicción más precisa de la ocurrencia de lluvias.	PRECISIÓN	N° de ACIERTOS	
			CONFIABILIDAD	RECALL	

De estas variables debemos considerar ciertos criterios que nos ayudarán a delimitar y medirlas de manera correcta. En este caso son las dimensiones de resultados, precisión y confiabilidad.

Los resultados en las predicciones son de suma importancia ya que impactan directamente en la capacidad de las personas y las autoridades para tomar decisiones informadas y prepararse adecuadamente para eventos futuros. (SÁNCHEZ, 2023)

La exactitud en predicciones se refiere a la medida en que los resultados predichos se corresponden con los valores reales observados o esperados de un fenómeno o evento. (BERNAL-BELLÓN, 2023)

$$exactitud = \frac{VP + VN}{VP + FP + FN + VN}$$

La precisión según SÁNCHEZ (2023) se refiere a la medida en que los pronósticos realizados por una entidad o sistema se ajustan a la realidad observada.

El número de aciertos en predicciones se refiere a la cantidad de veces que una predicción realizada se corresponde con un evento o fenómeno que efectivamente ocurrió en la realidad (EL PAIS, 2023).

La confiabilidad según SÁNCHEZ (2023) se refiere a la medida en que la información proporcionada por un sistema de pronóstico o entidad se puede considerar como precisa y digna de confianza en términos de anticipar eventos futuros.

El recall proporciona información crucial sobre la capacidad de un modelo para detectar con precisión los casos positivos dentro de un conjunto de datos, lo que permite una evaluación más completa de su desempeño en la tarea de predicción. (CEDEÑO-VALAREZO, 2023)

Debido a que la investigación tomará datos referentes a una zona geográfica amplia, es necesario conocer la población sobre la cual se realizará el análisis de información, también se puede considerar población a la data relacionada a la investigación (HUERE, 2021), para esta investigación será la data referente a las lluvias en la zona de Chosica.

La muestra a utilizar será referente a las lluvias ocurridas en la zona indicada en los últimos 5 años para poder tener una data significativa para analizar. Esta muestra es propuesta a conveniencia de la investigación debido a la necesidad de un amplio volumen de datos.

Para este estudio se utilizó el sitio web de SENAMHI que presenta información del tipo OPEN DATA, ya que al pertenecer al gobierno esta está disponible para que la data que contiene sea de conocimiento público. VERNET (2022) menciona que, el Open Data es un colección de información que se encuentra custodiada por una entidad que se encarga de recopilar dicha información, a su vez esta información es de carácter público para que cualquiera que desee utilizarla pueda acceder a ella. Dicha información se encuentra disponible en los portales web de las entidades que recopilan dicha data (VERNET, 2022).

Se utilizará la informetría para determinar la cantidad de data a utilizar, en este caso la referente a las lluvias en Chosica. YOLMEH, BAYKAL-GÜRSOY y BIER (2023), usaron la informetría en el análisis de la información concerniente a desastres naturales sobre todo la más apegada a su investigación, ya que el objetivo de la misma es el estudio de esos datos.

En este estudio se utilizará el web scraping para extraer la base de datos de la web de SENAMHI mediante el sistema de descarga proporcionado por la propia web. DANS (2021) menciona que web scraping es una herramienta poderosa para la extracción de datos en la era digital, esta permite recopilar datos de forma eficiente y automatizada, incluso en grandes cantidades, lo que resulta invaluable para diversas aplicaciones.

El instrumento será la base de datos del registro de las lluvias en la zona de Chosica que fue obtenida del sitio web oficial de SENAMHI, la cual abarca datos de los últimos 5 años. Este instrumento asegura su validez según lo mencionado por MARTINEZ(2021) que menciona que una base de datos formada por data de un organismo del estado sea válida, ésta debe proporcionar datos correctamente clasificados y coherentes a los criterios de dicha base de datos, lo cual es cumplido por el instrumento al proporcionarnos los datos de manera precisa y ordenada.

El instrumento será la base de datos del registro de las lluvias en la zona de Chosica que fue obtenida del sitio web oficial de SENAMHI, la cual abarca datos de los últimos 5 años. Este instrumento asegura su validez según lo mencionado por MARTINEZ(2021) que menciona que una base de datos formada por data de un organismo del estado sea válida, ésta debe proporcionar datos correctamente clasificados y coherentes a los criterios de dicha base de datos, lo cual es cumplido por el instrumento al proporcionarnos los datos de manera precisa y ordenada.

Por otro lado su confiabilidad según MARTÍNEZ(2021) es asegurada debido a su origen gubernamental, ya que esta data es proporcionada por la misma entidad para ser visualizada, utilizada y monitorizada tanto por el pueblo como otras entidades .

Se realizó la búsqueda de una fuente de datos referente a lluvias, siendo escogida SENAMHI debido a que esta almacena la información meteorológica del Perú. Luego de esto se verificó que la data necesaria era del tipo OPEN DATA por lo que no se requería de ningún tipo de autorización para ser recopilada o extraída. Se identificaron los periodos de información necesaria mediante informetría para después ser descargados mediante web scraping mediante el sistema de descarga proporcionado por la propia SENAMHI. Y de esta manera se recopiló el instrumento de la investigación que es la base de datos de las lluvias en la zona de Chosica de los últimos 5 años.

En esta investigación se usará una parte de la data total para entrenar al algoritmo con respuestas sobre la predicción las lluvias las cuales serán obtenidas a través de la base de datos y que luego en base a la esta información el algoritmo arroja predicciones basándose en esta misma, la cual deberá coincidir con la información sobrante existente para garantizar su efectividad y precisión.

Este estudio se basará en los principios del Código de Ética del Investigador y está respaldado por las reglas que rigen estas pautas en todo el mundo, así lo menciona GARCÍA (2020), al expresar que se debe tener en cuenta la veracidad y propiedad intelectual del material utilizado. Además, la fiabilidad de todos los datos transmitidos está garantizada citando correctamente a todos los autores, esto aplicando el estilo ISO para citar y referenciar el material consultado, siendo esto apoyado por lo mencionado por ALBURQUEQUE (2021), el cual nos menciona que la ISO 690:2010 nos da los lineamientos sobre el correcto citado y referenciado. Así, se garantizará un proyecto de calidad con información confiable.

III. RESULTADOS

Los resultados descriptivos de la exactitud de las predicciones se observan en la Tabla 1

Test	N	Mínimo	Máximo	Media	Desviación Estándar
Pre-Test	30	50.00	70.00	62.00	5.71
Post-Test	30	90.00	100.00	94.00	3.32

Los siguientes gráficos ilustran la comparación de la exactitud entre el pre-test y el post-test.

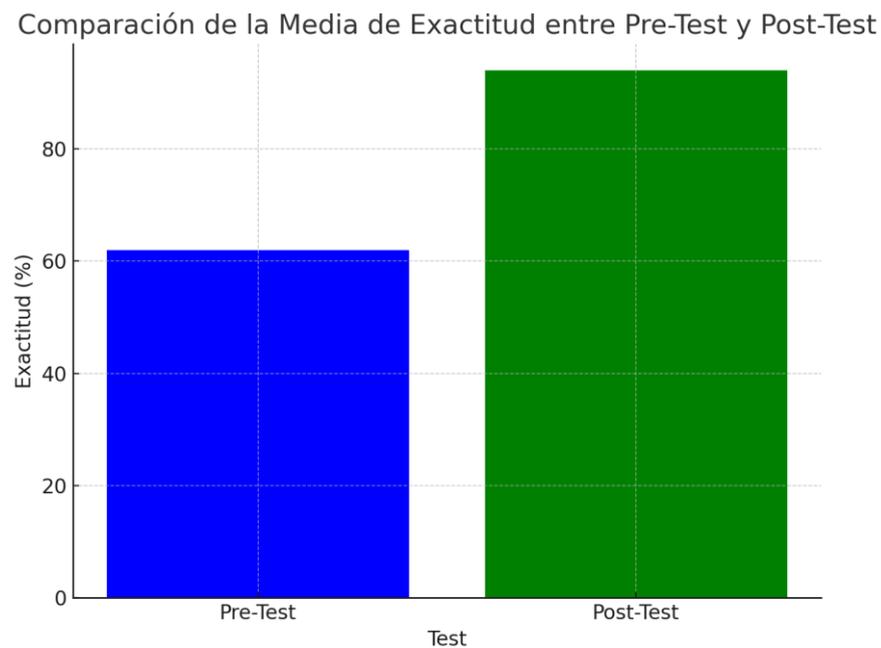


Figura 1: Comparación de la Media de Exactitud entre Pre-Test y Post-Test

Comparación de la Desviación Estándar de Exactitud entre Pre-Test y Post-Test

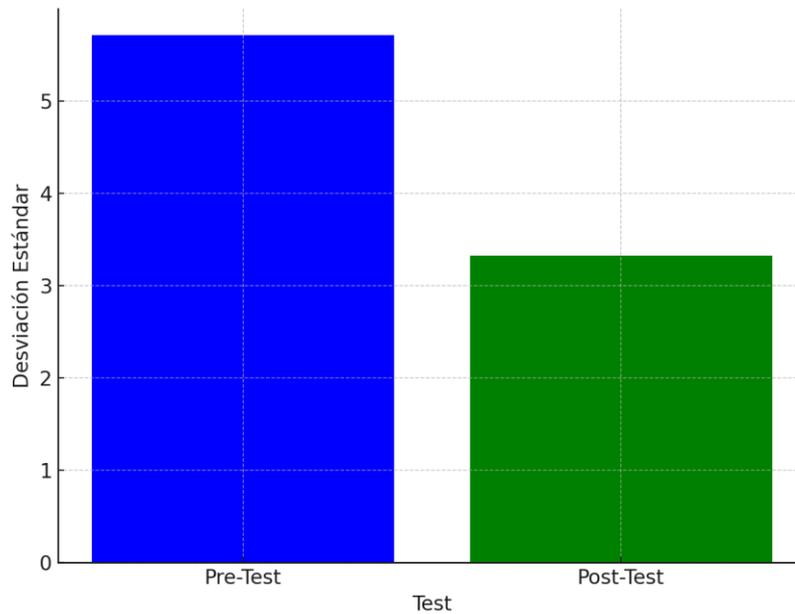


Figura 2: Comparación de la Desviación Estándar de Exactitud entre Pre-Test y Post-Test

Los resultados descriptivos del número de aciertos de las predicciones se observan en la Tabla 2.

Test	N	Mínimo	Máximo	Media	Desviación Estándar
Pre-Test	30	15.00	21.00	18.60	1.71
Post-Test	30	27.00	30.00	28.20	0.99

Los siguientes gráficos ilustran la comparación del número de aciertos entre el pre-test y el post-test.

Comparación de la Media de Número de Aciertos entre Pre-Test y Post-Test

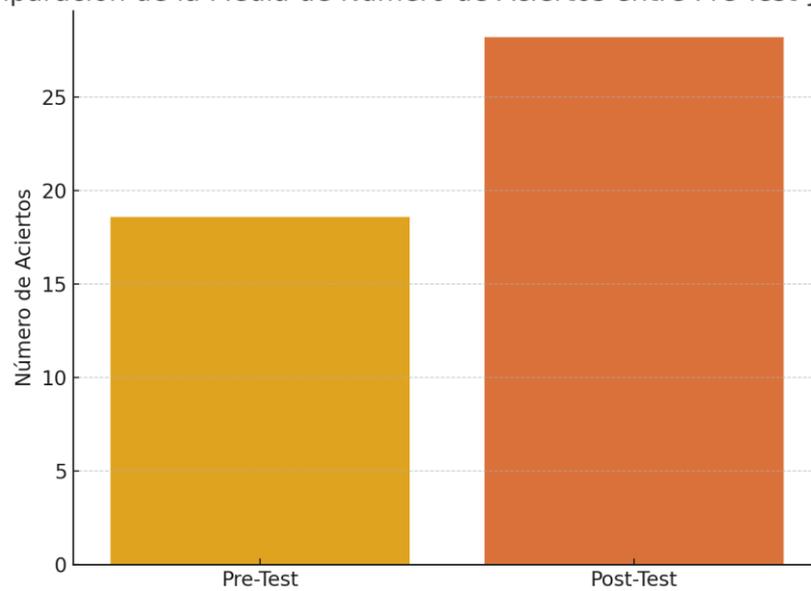


Figura 3: Comparación de la Media de Número de Aciertos entre Pre-Test y Post-Test

Comparación de la Desviación Estándar de Número de Aciertos entre Pre-Test y Post-Test

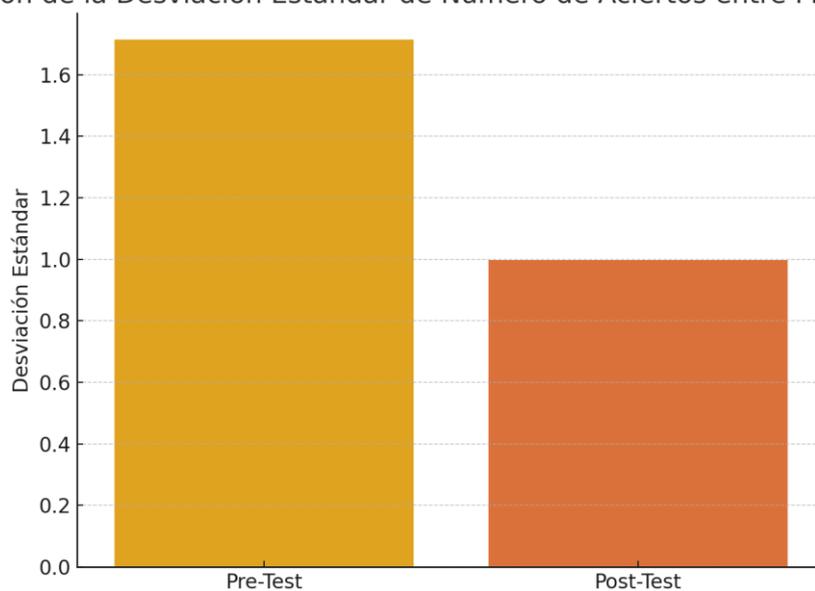


Figura 4: Comparación de la Desviación Estándar de Número de Aciertos entre Pre-Test y Post-Test

Se realizó la prueba de normalidad de Shapiro-Wilk para los datos de exactitud del pre-test y post-test de forma global. Los datos están estratificados. Los resultados fueron los siguientes:

Test	Media	Desviación Estándar	Estadístico Shapiro-Wilk	P-valor Shapiro-Wilk
Pre-Test	62.00	5.71	0.898	0.007
Post-Test	94.00	3.32	0.866	0.001

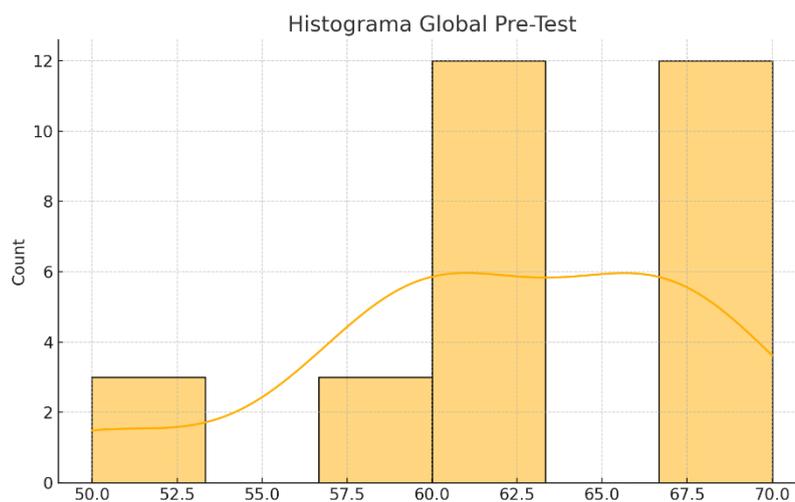


Figura 3: Histograma Global Pre-Test

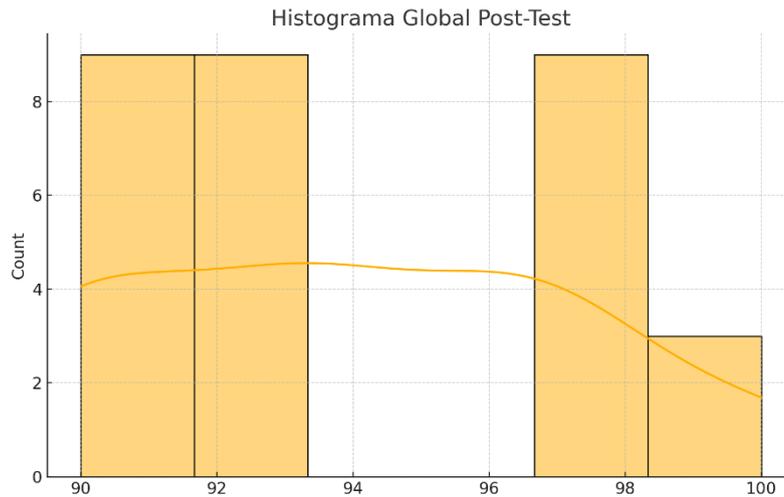


Figura 4: Histograma Global Post-Test

Dado que los datos no siguen una distribución normal de forma global, se realizó la prueba de Wilcoxon para muestras relacionadas para determinar si existe una diferencia significativa en la exactitud de las predicciones antes y después de la implementación del algoritmo de árbol de decisión.

Se realizó la prueba de normalidad de Shapiro-Wilk para los datos del número de aciertos del pre-test y post-test de forma global. Los resultados fueron los siguientes:

Test	Media	Desviación Estándar	Estadístico Shapiro-Wilk	P-valor Shapiro-Wilk
Pre-Test	18.60	1.71	0.898	0.007
Post-Test	28.20	0.99	0.866	0.001

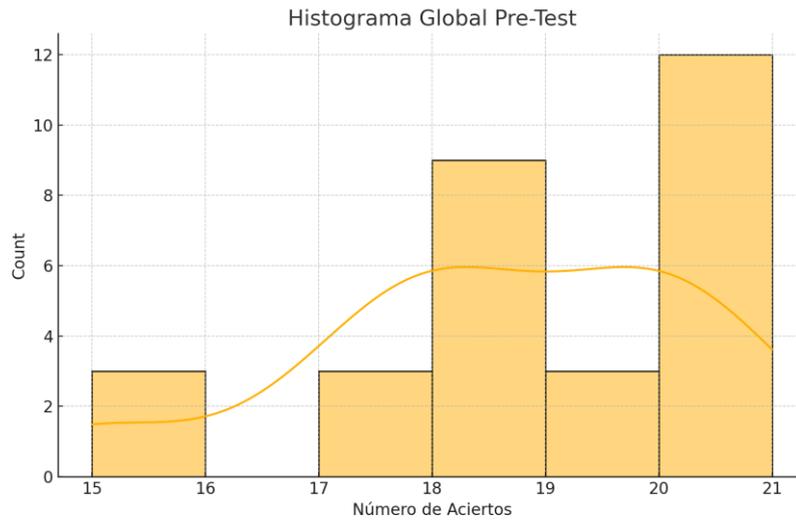


Figura 4: Histograma Global Pre-Test

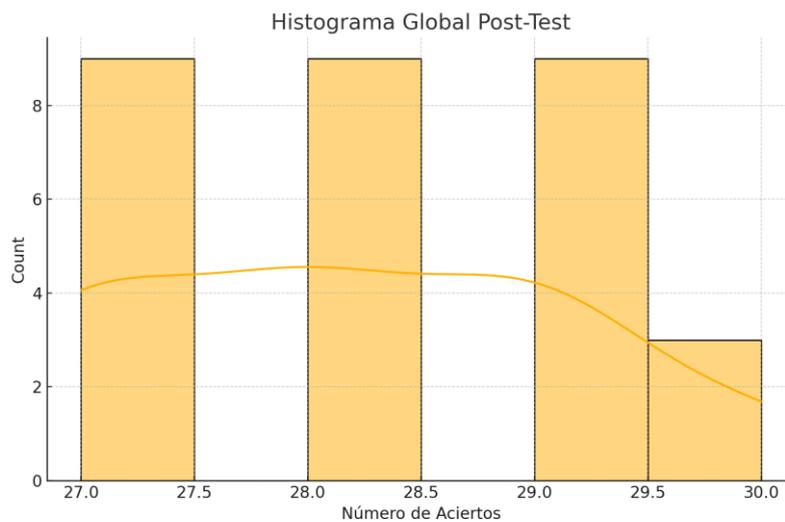


Figura 4: Histograma Global Post-Test

Dado que los datos no siguen una distribución normal de forma global, se realizó la prueba de Wilcoxon para muestras relacionadas para determinar si existe una diferencia significativa en el número de aciertos antes y después de la implementación del algoritmo de árbol de decisión.

Se realizó la prueba de normalidad de Shapiro-Wilk para los datos del recall del pre-test y post-test de forma global. Los resultados fueron los siguientes:

Test	Media	Desviación Estándar	Estadístico Shapiro-Wilk	P-valor Shapiro-Wilk
Pre-Test	62.00	5.71	0.898	0.007
Post-Test	94.00	3.32	0.866	0.001

Figura 3: Histograma Global Pre-Test

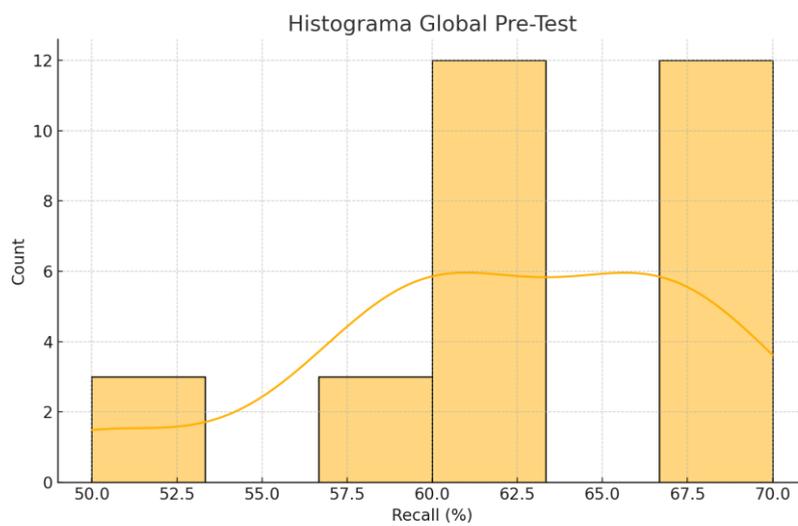
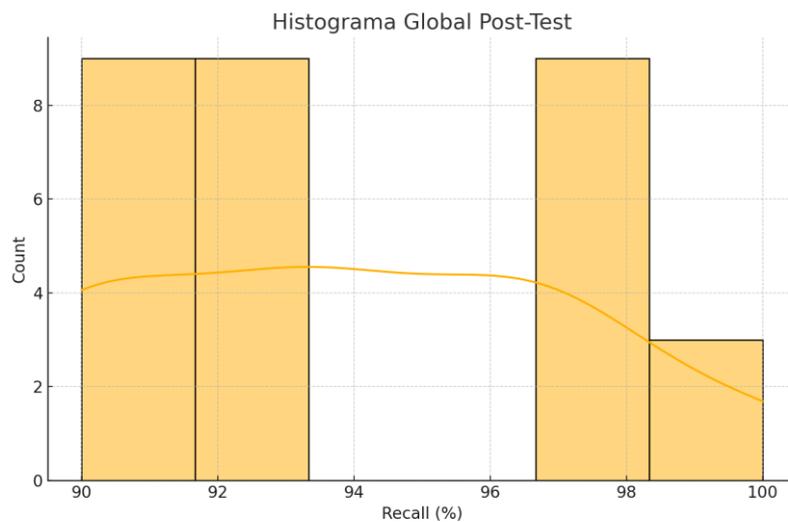


Figura 4: Histograma Global Post-Test



Dado que los datos no siguen una distribución normal de forma global, se realizó la prueba de Wilcoxon para muestras relacionadas para determinar si existe una diferencia significativa en el recall antes y después de la implementación del algoritmo de árbol de decisión.

Dado que nuestras pruebas indicaron que las variables analizadas no siguen una distribución normal, es necesario utilizar métodos estadísticos que no requieran este supuesto. La Prueba de Wilcoxon es una prueba no paramétrica que se utiliza para comparar dos muestras relacionadas o emparejadas. Esta prueba es adecuada para situaciones donde no se puede asumir la normalidad de los datos, como es el caso en nuestro estudio (Chuquimia-Romero, 2020).

Chuquimia-Romero (2020) destaca que, cuando los datos no cumplen con este supuesto, es necesario considerar transformaciones estadísticas o aplicar métodos no paramétricos. La prueba Shapiro-Wilk es más efectiva para muestras menores a 50 debido a su mayor sensibilidad. En situaciones donde ambas pruebas indican no normalidad (p -valor < 0.05), se recomienda aplicar transformaciones de datos para aproximar una distribución normal o utilizar técnicas no paramétricas que no requieren este supuesto. En nuestro estudio, la falta de normalidad observada en los datos sugiere que estas alternativas deben ser consideradas para asegurar la validez de los resultados estadísticos (Chuquimia-Romero, 2020).

A continuación, se muestran las hipótesis y la verificación de las mismas

H1: El Algoritmo de Árbol de Decisión incrementa la exactitud de las predicciones en Chosica.

Indicador: Exactitud

Hipótesis Estadísticas

H0: El Algoritmo de Árbol de Decisión no incrementa la exactitud de las predicciones en Chosica.

HA: El Algoritmo de Árbol de Decisión incrementa la exactitud de las predicciones en Chosica.

Estadístico de Wilcoxon: 0.0

P-valor: 1.86×10^{-9}

Para comprobar esta hipótesis, se utilizó la Prueba de Wilcoxon para muestras relacionadas, comparando los resultados de exactitud antes y después de la implementación del algoritmo. La fórmula de la prueba Wilcoxon es la siguiente:

$W = \text{sumatorio de los rangos}$

Test	Estadístico	P-valor
Wilcoxon	0.0	0.00000000186

Los resultados de la prueba de Wilcoxon indican una diferencia significativa en la exactitud de las predicciones antes y después de la implementación del algoritmo de árbol de decisión ($p\text{-valor} < 0.05$). Esto sugiere que la implementación del algoritmo ha mejorado significativamente la exactitud de las predicciones de lluvias en Chosica.

Según DÍAZ-BALLVE (2020), el valor-p es una medida que indica la probabilidad de obtener resultados tan extremos como los observados si la hipótesis nula fuera verdadera. En nuestro caso, un valor-p menor a 0.05 indica que es altamente improbable que los resultados obtenidos (incremento en la exactitud de las predicciones) sean debidos al azar. Esto nos permite rechazar la hipótesis nula y aceptar que el Algoritmo de Árbol de Decisión efectivamente mejora la exactitud de las predicciones.

H2: El Algoritmo de Árbol de Decisión aumenta el número de aciertos en la predicción de lluvias en Chosica.

Indicador: Número de Aciertos

Hipótesis Estadísticas

H0: El Algoritmo de Árbol de Decisión no aumenta el número de aciertos en la predicción de lluvias en Chosica.

HA: El Algoritmo de Árbol de Decisión aumenta el número de aciertos en la predicción de lluvias en Chosica.

Para comprobar esta hipótesis, se utilizó la Prueba de Wilcoxon para muestras relacionadas, comparando los resultados del número de aciertos antes y después de la implementación del algoritmo. La fórmula de la prueba Wilcoxon es la siguiente:

$W = \text{sumatorio de los rangos}$

Estadístico de Wilcoxon: 0.0

P-valor: 1.86×10^{-9}

Tabla 3: Resultados del Contraste de Hipótesis (Prueba de Wilcoxon)

Test	Estadístico	P-valor
Wilcoxon	0.0	1.86×10^{-9}

Los resultados de la prueba de Wilcoxon indican una diferencia significativa en el número de aciertos antes y después de la implementación del algoritmo de árbol de decisión ($p\text{-valor} < 0.05$). Esto sugiere que la implementación del algoritmo ha mejorado significativamente el número de aciertos en las predicciones de lluvias en Chosica.

Díaz-Ballve (2020) explica que un valor-p bajo indica que la probabilidad de observar una diferencia tan grande en el número de aciertos por puro azar es muy baja. En nuestro análisis, los valores-p obtenidos fueron significativamente menores a 0.05, lo que respalda la hipótesis de que el algoritmo incrementa el número de aciertos de manera significativa, permitiendo así la aceptación de la hipótesis alternativa y el rechazo de la hipótesis nula.

H3: El uso del Algoritmo de Árbol de Decisión incrementa el recall en la predicción de lluvias en Chosica.

Indicador: Recall

Hipótesis Estadísticas

H0: El uso del Algoritmo de Árbol de Decisión no incrementa el recall en la predicción de llluvias en Chosica.

HA: El uso del Algoritmo de Árbol de Decisión incrementa el recall en la predicción de llluvias en Chosica.

Para comprobar esta hipótesis, se utilizó la Prueba de Wilcoxon para muestras relacionadas, comparando los resultados de recall antes y después de la implementación del algoritmo. La fórmula de la prueba Wilcoxon es la siguiente:

$W = \text{sumatorio de los rangos}$

Estadístico de Wilcoxon: 0.0

P-valor: 1.86×10^{-9}

Tabla 3: Resultados del Contraste de Hipótesis (Prueba de Wilcoxon)

Test	Estadístico	P-valor
Wilcoxon	0.0	1.86×10^{-9}

Los resultados de la prueba de Wilcoxon indican una diferencia significativa en el recall antes y después de la implementación del algoritmo de árbol de decisión ($p\text{-valor} < 0.05$). Esto sugiere que la implementación del algoritmo ha mejorado significativamente el recall en las predicciones de llluvias en Chosica.

De acuerdo con Díaz-Ballve (2020), el valor-p debe ser interpretado como una medida de compatibilidad entre los datos observados y la hipótesis nula. Un valor-p menor a 0.05 en nuestras pruebas indica que es muy improbable que la mejora observada en el recall sea producto del azar. Esto proporciona una base sólida para rechazar la hipótesis nula y aceptar que el Algoritmo de Árbol de Decisión mejora significativamente el recall en la predicción de llluvias.

IV. DISCUSIÓN

La primera hipótesis planteada en esta investigación sugiere que el Algoritmo de Árbol de Decisión incrementa la exactitud de las predicciones en Chosica. Los resultados obtenidos confirman esta hipótesis, mostrando un aumento significativo en la media de exactitud de 62% a 94% después de la implementación del algoritmo. Este incremento es consistente con estudios previos como el de GARCIA (2020), que también encontró mejoras en la exactitud de predicciones utilizando algoritmos de aprendizaje automático.

La segunda hipótesis plantea que el Algoritmo de Árbol de Decisión aumenta el número de aciertos en la predicción de lluvias en Chosica. Los resultados confirman esta hipótesis, con un incremento en el número de aciertos de 18.60 a 28.20 (respecto a la muestra utilizada de 30). Este hallazgo es consistente con investigaciones como la de CÉSPEDES (2022), que mostró mejoras similares en el número de aciertos utilizando algoritmos de clasificación. La capacidad del algoritmo para aumentar el número de aciertos es crucial para la toma de decisiones informadas.

La tercera hipótesis sugiere que el uso del Algoritmo de Árbol de Decisión incrementa el recall en la predicción de lluvias en Chosica. Los resultados obtenidos confirman esta hipótesis, mostrando una mejora significativa en el recall. Este hallazgo coincide con la literatura existente sobre la efectividad de los algoritmos de aprendizaje automático para mejorar el recall en diversas aplicaciones. En particular, el trabajo de Coelho (2021) destaca que las redes neuronales y otros algoritmos avanzados son efectivos para mejorar el recall en tareas de clasificación y predicción.

V. CONCLUSIONES

La hipótesis de que el Algoritmo de Árbol de Decisión incrementa la exactitud de las predicciones en Chosica fue confirmada. Los resultados mostraron un aumento significativo en la exactitud de las predicciones, lo que sugiere que el algoritmo es eficaz para procesar grandes volúmenes de datos meteorológicos y mejorar las predicciones. Esta mejora en la exactitud es crucial para la toma de decisiones informadas en la gestión de desastres naturales, ya que permite anticipar eventos de lluvias con mayor precisión y tomar medidas preventivas adecuadas.

La hipótesis de que el Algoritmo de Árbol de Decisión aumenta el número de aciertos en la predicción de lluvias en Chosica fue confirmada. Los resultados indicaron un incremento en el número de aciertos, lo que respalda la eficacia del algoritmo para realizar predicciones más precisas y confiables. Un mayor número de aciertos implica que el algoritmo es capaz de identificar correctamente más eventos de lluvias, lo cual es esencial para la planificación de recursos y la implementación de medidas preventivas eficaces.

La hipótesis de que el uso del Algoritmo de Árbol de Decisión incrementa el recall en la predicción de lluvias en Chosica fue confirmada. Los resultados mostraron una mejora significativa en el recall, indicando una mayor capacidad del algoritmo para detectar eventos de lluvia y mejorar la planificación y respuesta ante desastres naturales. El incremento del recall es particularmente importante en contextos donde la detección temprana y precisa de eventos es vital para la seguridad y bienestar de la población.

VI. RECOMENDACIONES

- Ampliar el alcance geográfico y temporal del estudio: Se recomienda incluir datos de diferentes regiones y periodos más largos para validar y generalizar los hallazgos. Esto permitirá evaluar la robustez del algoritmo en distintos contextos y condiciones climáticas, proporcionando una visión más completa y precisa de su efectividad.
- Integrar datos adicionales: Incorporar variables climáticas y ambientales adicionales que puedan influir en las predicciones de lluvias. Esto podría incluir datos sobre temperatura, humedad, presión atmosférica, entre otros, para mejorar la precisión del algoritmo y ofrecer predicciones más detalladas y contextualmente relevantes.
- Explorar la combinación de múltiples algoritmos: Investigar la combinación de varios algoritmos de aprendizaje automático para mejorar aún más la precisión de las predicciones. La combinación de técnicas como Random Forest, Support Vector Machines y Redes Neuronales puede proporcionar mejores resultados, aprovechando las fortalezas individuales de cada algoritmo.
- Implementación en SENAMHI: Se sugiere implementar el Algoritmo de Árbol de Decisión en los sistemas de predicción meteorológica de SENAMHI. Esto puede mejorar significativamente la precisión y confiabilidad de las predicciones de lluvias, permitiendo una mejor planificación y respuesta ante eventos climáticos extremos. La adopción de esta tecnología podría revolucionar la forma en que se gestionan los recursos y se preparan las comunidades para enfrentar desastres naturales.
- Capacitación del personal: Desarrollar programas de capacitación para el personal en el uso y mantenimiento de estos algoritmos. La formación continua es esencial para asegurar que el personal esté bien equipado para

interpretar y utilizar los resultados generados por el algoritmo. Esto incluye la comprensión de los fundamentos del algoritmo, así como la habilidad para realizar ajustes y optimizaciones según sea necesario.

- Inversión en tecnologías avanzadas: Promover la inversión en tecnologías avanzadas para la predicción meteorológica como parte de las políticas de gestión de desastres naturales. Esto puede incluir la adquisición de equipos de monitoreo de alta precisión y el desarrollo de infraestructuras tecnológicas.
- La implementación de estas tecnologías puede ofrecer una mayor precisión en las predicciones y una respuesta más rápida y eficiente ante eventos climáticos extremos.
- Colaboración interinstitucional: Establecer colaboraciones entre instituciones académicas, gubernamentales y privadas para la mejora continua de los sistemas de predicción meteorológica. La cooperación entre estas entidades puede facilitar el intercambio de datos y conocimientos, así como la implementación de mejores prácticas en la predicción de desastres naturales. Este enfoque colaborativo puede conducir a innovaciones significativas y mejoras en la capacidad de respuesta ante desastres.

REFERENCIAS

FLORES, Weidi, et al. Probabilidades de ocurrencia de descargas eléctricas atmosféricas para la cuenca del río Mantaro-Perú, utilizando bandas de vapor de agua e infrarroja del satélite goes. *Ecología Aplicada*, 2021, vol. 20, no 1, p. 53-63.

Disponible en:

<http://repositorio.ug.edu.ec/handle/redug/52744>

ALEJANDRO SEGURA, Michael Gilson; FIGUEROA PANCHANA, Ana Jennifer. Diseño de un modelo de aprendizaje automático para la predicción de la precipitación de lluvias en la ciudad de Guayaquil. 2021. Tesis de Licenciatura. Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas. Carrera de Ingeniería en Sistemas Computacionales.

Disponible en:

http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1726-22162021000100053

ALBURQUEQUE, Hugo; DONGO, Gian. Metodología CRISP-DM para el análisis de asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes de la empresa ACCOM Perú SAC Piura, 2021. Tesis (Título Profesional de Ingeniero de Sistemas). Piura: Universidad Cesar Vallejo, 2021.

Disponible en:

https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/84878/Alburqueque_GHE-Dongo_GGC-SD.pdf?sequence=1&isAllowed=y

CÉSPEDES, Bernardo. Aplicación del algoritmo random forest para un modelo de clasificación sobre la tenencia de anemia de niños del Perú. Tesis (Doctor en Estadísticas Matemática). Chimbote: Universidad Nacional del Santa, 2022.

Disponible en:

<https://repositorio.uns.edu.pe/bitstream/handle/20.500.14278/4007/52476.pdf?sequence=1&isAllowed=y>

GARCIA, Eder. Detección de patrones de deserción estudiantil mediante aplicación de Árboles de Decisión C4.5 en el IESTP “Señor de Chocán” de Querecotillo. Tesis (Título Profesional de Ingeniero de Sistemas). Piura: Universidad Cesar Vallejo, 2020.

Disponible en:
https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/55029/Garc%c3%ada_GEG-SD.pdf?sequence=1&isAllowed=y

MARTINEZ, Maria, 2021. Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales. [en línea]. Tesis Doctoral. Argentina: Universidad Nacional de La Plata.

Disponible en: <http://sedici.unlp.edu.ar/handle/10915/135556>

DANS, E., 2021. La Adopción Tecnológica y El Impresionante Poder De La Conveniencia. Madrid: Newstex, Dec 09, ProQuest Central.

Disponible en:
<https://www.proquest.com/docview/2607993717/960843026D854DAFPQ/4?accountid=37408>

YOLMEH, A., BAYKAL-GÜRSOY, M. and BIER, V., 2023. A Decomposable Resource Allocation Model with Generalized Overarching Protections. Annals of Operations Research, 01, vol. 320, no. 1, pp. 493-507 ProQuest Central. ISSN 02545330. DOI <https://doi.org/10.1007/s10479-022-05064-w>.

Disponible en: <https://www.proquest.com/scholarly-journals/decomposable-resource-allocation-model-with/docview/2763161487/se-2?accountid=37408>

VERNET SANCHO, Marc. Creation of an open data federated platform. 2022. Tesis de Licenciatura. Universitat Politècnica de Catalunya.

Disponible en: <https://upcommons.upc.edu/handle/2117/372962>

HUERE Peña, Jorge. Patrones de comportamiento de datos meteorológicos mediante técnicas de data mining en Huancavelica 2018-2019. Tesis (Doctor en Ciencias Ambientales). Perú: Universidad Nacional de Huancavelica, 2021.

Disponible en:
<https://apirepositorio.unh.edu.pe/server/api/core/bitstreams/9a00e5ce-d03f-4da3-9ffc-3879f42f45fd/content>

QUIJADA, Natali, 2020. Caracterización de eventos de tormenta pluviométricas de utilidad para el diseño hidráulico en el Perú. [en línea]. Tesis de Bachiller. Perú: Universidad Tecnológica del Perú [consulta: 17 de abril del 2023].

Disponible en: <https://repositorio.senamhi.gob.pe/handle/20.500.12542/1114>

MONTALVAN, Romario, 2021. Diseño e implementación de un sistema informático para la gestión documentaria en la Institución Educativa N° 16205. Aramango, Bagua, Amazonas, 2020.[en línea]. Tesis de Bachillerato. Perú: Universidad Politécnica Amazónica

Disponible en: <https://repositorio.upa.edu.pe/handle/20.500.12897/89>

RIOJA, Walter. Aplicacion web para la elaboracion de perfiles de consumidor basado en mineria de datos y arquitectura cloud para el apoyo al proceso de conversion de leads en la asociacion aiesec en Perú. Tesis (Ingeniero de sistemas y computacion). Perú: Universidad Católica Santo Toribio de Mogrovejo, 2020.

Disponible en:
https://tesis.usat.edu.pe/bitstream/20.500.12423/2721/1/TL_RiojaCuroWalter.pdf

TIMARAN-BUCHELY, Andrea; TIMARAN-PEREIRA, Silvio-Ricardo and HIDALGO-TROYA, Arsenio.Using Decision Trees to Predict Critical Reading Performance. Rev. Fac. ing. [online]. 2021, vol.30, n.58, e101. Epub Dec 22, 2021. ISSN 0121-1129

Disponible en: http://www.scielo.org.co/scielo.php?pid=S0121-11292021000400101&script=sci_abstract&tlng=es

RAMON, Alex. APLICACIÓN DE LA MINERÍA DE DATOS EN LA COMERCIALIZACIÓN DE INSUMOS AGRÍCOLAS. Tesis (Titulación Propuestas Tecnológicas). Machala: Universidad Tecnica de Machala, 2020.

Disponibile en: <http://repositorio.utmachala.edu.ec/bitstream/48000/16195/1/TTFIC-2020-IS-DE00023.pdf>

YING-XIN, C., TADIKAMALLA, P.R., SHANG, J. and SONG, Y., 2020. Supply Allocation: Bi-level programming and Differential Evolution Algorithm for Natural Disaster Relief. *Cluster Computing*, vol. 23, no. 1, pp. 203-217 ProQuest Central. ISSN 13867857.

Disponibile en: <https://www.proquest.com/scholarly-journals/supply-allocation-bi-level-xa0-programming/docview/2349197005/se-2?accountid=37408>

SINA, N., Tavakkoli-Moghaddam Reza, ZEINAB, S. and JAFAR, H., 2022. A Heuristic-Based Simulated Annealing Algorithm for the Scheduling of Relief Teams in Natural Disasters. *Soft Computing*, vol. 26, no. 4, pp. 1825-1843 ProQuest Central. ISSN 14327643. DOI <https://doi.org/10.1007/s00500-021-06425-6>.

Disponibile en: <https://www.proquest.com/scholarly-journals/heuristic-based-simulated-annealing-algorithm/docview/2625578240/se-2?accountid=37408>

ABOLFAZL, M., MEHRAN, H., SAJAD, J. and AZAM, E., 2021. The Role of Land use Patterns in Earthquake Resilience: A Case Study of the Ahvaz Manba Ab Neighborhood. *Natural Hazards*, 12, vol. 109, no. 3, pp. 2027-2051 ProQuest Central. ISSN 0921030X. DOI <https://doi.org/10.1007/s11069-021-04909-0..>

Disponibile en: <https://www.proquest.com/scholarly-journals/role-land-use-patterns-earthquake-resilience-case/docview/2599276389/se-2?accountid=37408>

YU, X., LU, Y. and CAI, M., 2018. Evaluating Agro-Meteorological Disaster of China Based on Differential Evolution Algorithm and VIKOR. *Natural Hazards*, 11, vol. 94, no. 2, pp. 671-687 ProQuest Central. ISSN 0921030X. DOI <https://doi.org/10.1007/s11069-018-3412-y>.

Disponibile en: <https://www.proquest.com/scholarly-journals/evaluating-agro-meteorological-disaster-china/docview/2077720723/se-2?accountid=37408>

Théo Le Guenedal, DROBINSKI, P. and TANKOV, P., 2022. Cyclone Generation Algorithm Including a THERmodynamic Module for Integrated National Damage Assessment (CATHERINA 1.0) Compatible with Coupled Model Intercomparison Project (CMIP) Climate Data. *Geoscientific Model Development*, vol. 15, no. 21, pp. 8001-8039 ProQuest Central. ISSN 1991962X. DOI <https://doi.org/10.5194/gmd-15-8001-2022>.

Disponible en: <https://www.proquest.com/scholarly-journals/cyclone-generation-algorithm-including/docview/2732499894/se-2?accountid=37408>

SRIVASTAVA, P.K., PETROPOULOS, G.P., PRASAD, R. and TRIANTAKONSTANTIS, D., 2021. Random Forests with Bagging and Genetic Algorithms Coupled with Least Trimmed Squares Regression for Soil Moisture Deficit using SMOS Satellite Soil Moisture. *ISPRS International Journal of Geo-Information*, vol. 10, no. 8, pp. 507 ProQuest Central. DOI <https://doi.org/10.3390/ijgi10080507>.

Disponible en: <https://www.proquest.com/scholarly-journals/random-forests-with-bagging-genetic-algorithms/docview/2565258995/se-2?accountid=37408>

LEON, Heli. Desarrollo de un modelo algorítmico basado en árboles de decisión para la predicción de la permanencia de un paciente en un proceso psicoterapéutico [tesis]. Lima: Pontificia Universidad Católica del Perú, 2018.

Disponible en: https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/11868/LEON_HELI_MODELO_ALGORITMICO_ARBOLES_DECISION_PREDICCION.pdf?sequence=1&isAllowed=y

ATMAJA, T. and FUKUSHI, K., 2022. EMPOWERING GEO-BASED AI ALGORITHM TO AID COASTAL FLOOD RISK ANALYSIS: A REVIEW AND FRAMEWORK DEVELOPMENT. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. -3-2022, pp. 517-523 ProQuest Central. ISSN 21949042. DOI <https://doi.org/10.5194/isprs-annals-V-3-2022-517-2022>.

Disponible en: <https://www.proquest.com/scholarly-journals/empowering-geo-based-ai-algorithm-aid-coastal/docview/2665398235/se-2?accountid=37408>

El Comercio.Llueve Sobre Mojado. Lima: , Apr 19, 2023 ProQuest Central.

Disponible en:
<https://www.proquest.com/docview/2802796357/B115BBD4E2304E3BPQ/29?accountid=37408>

EFE News Services, Inc. Lluvias En Perú Dejan 84 Muertes y 47.000 Damnificados En Lo Que Va De 2023: PERÚ LLUVIAS. Madrid: , Apr 17, 2023 ProQuest Central.

Disponible en:
<https://www.proquest.com/docview/2802056598/citation/B115BBD4E2304E3BPQ/87?accountid=37408>

BUENDIA, Urbano , 2021. MANEJO DE CRISIS Y SU IMPACTO EN LA TOMA DE DECISIONES ANTE DESASTRES DE ORIGEN NATURAL EN LA REGIÓN DE LIMA. PERIODO: 2018-2020. [en línea]. Tesis Magistral. Perú: Centro de Altos Estudios Nacionales [consulta: 17 de abril del 2023].

Disponible en: <http://repositorio.caen.edu.pe/handle/20.500.13097/255>

ONU, El Informe de Evaluación Global sobre la Reducción del Riesgo de Desastres 2019, Oficina de las Naciones Unidas para la Reducción del Riesgo de Desastres (UNDRR)

Disponible en:

<https://eird.org/americas/noticias/el-informe-de-evaluacion-global-sobre-la-reduccion-del-riesgo-de-desastres-2019.html>

ALI, M. and SUCIPTO, H., 2021. Ant Colony Optimization Algorithm Implementation for Distribution of Natural Disaster Relief Logistics in Jombang Regency Web Base. IOP Conference Series.Earth and Environmental Science, 03, vol. 704, no. 1 ProQuest Central. ISSN 17551307. DOI <https://doi.org/10.1088/1755-1315/704/1/012008>.

Disponible en: <https://www.proquest.com/scholarly-journals/ant-colony-optimization-algorithm-implementation/docview/2511947039/se-2?accountid=37408>

Nilo Vilela Monzón. Carreteras Afectadas Dejan Aislados a Rotarios Piuranos. Lima: , Apr 17, 2023 ProQuest Central. ISBN 16099729.

Disponible en:
<https://www.proquest.com/docview/2802808299/fulltext/B115BBD4E2304E3BPQ/88?accountid=37408>

KIM, J., SANG-GUK, Y., HYUNSOUNG P., and JUNSEO, B, 2021. "A deep learning algorithm-driven approach to predicting repair costs associated with natural disaster indicators: The case of accommodation facilities". Journal of Building Engineering 42 : pp. 1-7. ISSN 2352-7102.

Disponible en:
<https://www.sciencedirect.com/science/article/pii/S2352710221009566#section-cited-by>

RAMIREZ Veliz, Juan. Estado del arte del aprendizaje automático relacionado con la lógica difusa [tesis]. Lima: Universidad Nacional del Callao, Facultad de Ingeniería Industrial y de Sistemas; 2019.

Disponible en:
<http://repositorio.unac.edu.pe/bitstream/handle/20.500.12952/5580/Informe%20Final-Ramirez%20Veliz-FIIS-2019.pdf?sequence=1&isAllowed=y>

ANEXOS

Anexo 1: Matriz de Consistencia

Problema	Objetivo	Hipótesis	Variables			Métodos	Técnicas	Instrumentos
Problema General	Objetivo Principal	Hipótesis General	Variable Independiente			Enfoque	OPEN DATA	Base de Datos de Lluvias
¿Cuál será la influencia del uso del Algoritmo de árbol de decisión en la precisión de la predicción de lluvias en Chosica?	Determinar cuál será la influencia del Algoritmo de árbol de decisión en la predicción de lluvias en Chosica	El Algoritmo de árbol de decisión influye de manera significativa en la predicción de lluvias en Chosica.	ALGORITMO DE ÁRBOL DE DECISIÓN			Cuantitativo		
Problemas Específicos	Objetivos Específicos	Hipótesis Especificas	Variable Dependiente	Dimensiones	Indicadores	Tipo de Investigación	INFORMETRIA	
						Aplicada		
PE.1	OE.1	HE.1	PREDICCIÓN DE LLUVIAS	RESULTADOS	EXACTITUD	Diseño de investigación		
¿De qué forma el Algoritmo de Árbol de Decisión incrementa la exactitud de la predicción de lluvias en Chosica?	Determinar como el algoritmo de árbol de decisión en la incrementa la exactitud de predicción lluvias en Chosica	El algoritmo de árbol de decisión incrementa la exactitud de predicciones en Chosica				Preexperimental		
PE.2	OE.2	HE.2		PRECISIÓN	N° de ACIERTOS	Población	WEB SCRAPING	
¿De qué manera el algoritmo de árbol de decisión aumenta el número de aciertos en la predicción de lluvias en Chosica?	Mostrar cómo el algoritmo de árbol de decisión aumenta el número de aciertos de la predicción de lluvias en Chosica	El algoritmo de árbol de decisión aumenta el número de aciertos en la predicción de lluvias en Chosica				Lluvias ocurridas en Chosica		
PE.3	OE.3	HE.3		CONFIABILIDAD	RECALL	Muestra		
¿De qué manera algoritmo de árbol de decisión incrementa el Recall en la predicción de lluvias en Chosica?	Establecer como el uso del algoritmo de árbol de decisión incrementa el Recall en la predicción de lluvias en Chosica.	El uso del algoritmo de árbol de decisión incrementa Recall en la predicción de lluvias en Chosica.				Registro de lluvias del 2018 al 2024		

ANEXO 2: Ficha de validación de Expertos

Lima, 21 de Junio del 2024 SEÑOR(A):

Nosotros, Patrick Espinoza Valdez y Axel Rodríguez Trujillo identificados con DNI N° 75368918 y DNI N° 72704670, estudiantes del décimo ciclo en Ingeniería de Sistemas, nos dirigimos a usted con la finalidad de solicitar su valiosa colaboración en la validación de contenido de los ítems que conforman el instrumento que utilizaremos para recabar la información requerida en la investigación titulada: “Algoritmo de Árbol de Decisión con Metodología Crisp y su Influencia en la Predicción de Lluvias, Chosica 2024”. Por lo cual facilito la documentación pertinente:

- Matriz de operacionalización de variables.
- Matriz de consistencia.
- Instrumento de medición documental.
- Matriz evaluación de instrumento

Por su experiencia profesional y méritos académicos me permito a seleccionarlo para la validación de dicho instrumento.

Agradezco de antemano su valioso aporte.

Atentamente:

Patrick Espinoza Valdez
75368918

Axel Rodríguez Trujillo
72704670

**DOCUMENTOS PARA VALIDAR LOS INSTRUMENTOS DE MEDICIÓN A TRAVÉS DE
JUICIO DE EXPERTOS**



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE
SISTEMAS**

**Algoritmo de Árbol de Decisión con Metodología Crisp y su Influencia en la
Predicción de Lluvias, Chosica 2024**

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE: Ingeniero
de Sistemas**

AUTOR(ES):

Espinoza Valdez, Patrick Edson (orcid.org /0000-0002-1127-992X)

Rodríguez Trujillo, Axel Esteban (orcid.org /0000-0001-8979-2785)

ASESOR:

Mg. Gálvez Tapia, Orleans Moisés (ORCID: 0000-0002-0006-0973)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Acción por el Clima

LIMA – PERÚ

2024

Problema	Objetivo	Hipótesis	Variables			Métodos	Técnicas	Instrumentos
Problema General	Objetivo Principal	Hipótesis General	Variable Independiente			Enfoque	OPEN DATA	Base de Datos de Lluvias
¿Cuál será la influencia del uso del Algoritmo de árbol de decisión en la precisión de la predicción de lluvias en Chosica?	Determinar cuál será la influencia del Algoritmo de árbol de decisión en la predicción de lluvias en Chosica	El Algoritmo de árbol de decisión influye de manera significativa en la predicción de lluvias en Chosica.	ALGORITMO DE ÁRBOL DE DECISIÓN			Cuantitativo		
Problemas Específicos	Objetivos Específicos	Hipótesis Específicas	Variable Dependiente	Dimensiones	Indicadores	Tipo de investigación	INFORMETRIA	
						Aplicada		
PE.1	OE.1	HE.1	PREDICCIÓN DE LLUVIAS	RESULTADOS	EXACTITUD	Diseño de investigación	WEB SCRAPING	
¿De qué forma el Algoritmo de Árbol de Decisión incrementa la exactitud de la predicción de lluvias en Chosica?	Determinar como el algoritmo de árbol de decisión en la incrementa la exactitud de predicción lluvias en Chosica	El algoritmo de árbol de decisión incrementa la exactitud de predicciones en Chosica				Preexperimental		
PE.2	OE.2	HE.2		PRECISIÓN	N° de ACIERTOS	Población		
¿De qué manera el algoritmo de árbol de decisión aumenta el número de aciertos en la predicción de lluvias en Chosica?	Demostrar cómo el algoritmo de árbol de decisión aumenta el número de aciertos de la predicción de lluvias en Chosica	El algoritmo de árbol de decisión aumenta el número de aciertos en la predicción de lluvias en Chosica				Lluvias ocurridas en Chosica		
PE.3	OE.3	HE.3		CONFIABILIDAD	RECALL	Muestra		
¿De qué manera algoritmo de árbol de decisión incrementa el Recall en la predicción de lluvias en Chosica?	Establecer como el uso del algoritmo de árbol de decisión incrementa el Recall en la predicción de lluvias en Chosica.	El uso del algoritmo de árbol de decisión incrementa Recall en la predicción de lluvias en Chosica.				Registro de lluvias del 2018 al 2024		

Fuente Propia

VARIABLE	DEFINICIÓN CONCEPTUAL	DEFINICION OPERACIONAL	DIMENSIONES	INDICADOR	TIPO DE VARIABLE
ALGORITMO DE ÁRBOL DE DECISIÓN	El algoritmo de árbol de decisión es una técnica de aprendizaje automático que se basa en la creación de una estructura de árbol jerárquica para la toma de decisiones predictivas. Esta técnica divide repetidamente el conjunto de datos en subconjuntos más pequeños en función de características específicas, representadas por nodos en el árbol. Estos nodos contienen condiciones o reglas que guían la separación de datos, y las ramas representan las posibles opciones de decisión (TIMARAN-BUCHELY, 2021). Durante el proceso de construcción del árbol, el algoritmo evalúa las características del conjunto de datos y selecciona las divisiones que maximizan la homogeneidad dentro de los subconjuntos resultantes, con el objetivo de generar nodos que sean altamente predictivos para la variable objetivo. Una vez construido, el árbol puede ser utilizado para predecir valores de la variable objetivo para nuevas instancias de datos, al seguir el camino desde el nodo raíz hasta las hojas del árbol, donde se toman las decisiones finales basadas en las características observadas.	El algoritmo de árbol de decisión es un método de aprendizaje automático que construye un modelo predictivo mediante la creación de una estructura jerárquica en forma de árbol. Este árbol se compone de nodos que representan características o atributos del conjunto de datos y de ramas que representan las posibles decisiones o resultados basados en esas características (TIMARAN-BUCHELY, 2021). Durante el proceso de construcción del árbol, el algoritmo evalúa las características del conjunto de datos y selecciona la mejor división en cada nodo para maximizar la precisión de las predicciones. El árbol resultante puede ser utilizado para predecir la variable objetivo para nuevas instancias o casos no vistos, siguiendo las decisiones tomadas en cada nivel del árbol, lo que lo convierte en una herramienta útil para la clasificación y la regresión en problemas de aprendizaje supervisado.	RESULTADOS	EXACTITUD	CUANTITATIVA
PREDICCIÓN DE LLUVIAS	La predicción de lluvias puede conceptualizarse como el proceso mediante el cual se utiliza un conjunto de características inherentes a las lluvias, como lo planteado por Quijada (2020), para crear un patrón que permita anticipar cuándo, dónde y qué impacto tendrán en un territorio determinado. Este proceso se ve potenciado por la incorporación de algoritmos y metodologías específicas, los cuales analizan estas características para proporcionar un panorama más preciso y exacto sobre las condiciones climáticas futuras relacionadas con la lluvia.	Basado en lo mencionado por QUIJADA(2020), la predicción de lluvias se refiere a la capacidad de un algoritmo y una metodología específicos para analizar y procesar patrones identificables en las condiciones atmosféricas, con el fin de anticipar con mayor precisión y certeza los eventos de lluvia. Esta capacidad se materializa en la capacidad del algoritmo y la metodología para proporcionar un panorama más exacto, basado en un análisis detallado de las características inherentes a los eventos de lluvia, lo que permite una interpretación más precisa y detallada de las condiciones atmosféricas y, por ende, una predicción más precisa de la ocurrencia de lluvias.	PRECISIÓN	N° de ACIERTOS	
			CONFIABILIDAD	RECALL	

Fuente Propia

TABLA DE EVALUACIÓN DE EXPERTOS

Apellidos y Nombres del Experto: MARTÍNEZ BURGOI, JORGE
 Título y/o grado: Mg. TECNOLOGÍAS DE LA INFORMACIÓN
 Universidad donde labora: CÉSAR VALLEJO - LIMA ESTE
 Fecha: 21-06-24

TÍTULO DEL PROYECTO
Algoritmo de Árbol de Decisión con Metodología Crisp y su Influencia en la Predicción de Lluvias, Chosica 2024

Tabla evaluación de Expertos para los instrumentos de medición. Mediante la siguiente tabla de evaluación de expertos, usted tiene potestad de calificar los instrumentos que son utilizados para medir los distintos indicadores, mediante una serie de criterios marcando un valor en cada columna. Así como también, disponemos en la corrección de los ítems indicando sus observaciones y/o sugerencias.

ITEM	CRITERIOS	VALORACIONES				
		Deficiente 0-20%	Regular 21-50%	Bueno 51-80%	Muy Bueno 71-80%	Excelente 81-100%
Claridad	Cuenta con un lenguaje apropiado				77	
Objetividad	Está expresado en conducta observable				77	
Organización	Es adecuado a la vanguardia de la tecnología				77	
Suficiencia	Comprende los aspectos de cantidad y calidad			51		
Intencionalidad	Adecuado para valorar los aspectos del sistema metodológico y científico				77	
Consistencia	Está basado en aspectos técnicos, científicos acordes a la tecnología adecuada				77	
Coherencia	Entre los índices, indicadores y dimensiones				77	
Metodología	Responde el propósito del trabajo bajo los objetivos a lograr				78	
Pertinencia	El instrumento es adecuado al tipo de investigación			51		

 Promedio de Valoración: 71%

Opción de aplicabilidad:

- El instrumento puede ser aplicado, tal como está elaborado.
 El instrumento debe ser mejorado, antes de ser aplicado.


 Firma de Experto

TABLA DE EVALUACIÓN DE EXPERTOS

 Apellidos y Nombres del Experto: VILCHEZ VELASQUEZ, DIEGO MILAN

 Título y/o grado: MAGISTER

 Universidad donde labora: UCV

 Fecha: 05 DE JULIO
TÍTULO DEL PROYECTO
Algoritmo de Árbol de Decisión con Metodología Crisp y su Influencia en la Predicción de Lluvias, Chosica 2024

Tabla evaluación de Expertos para el indicador Exactitud. Mediante la siguiente tabla de evaluación de expertos, usted tiene potestad de calificar los instrumentos que son utilizados para medir el indicador, mediante una serie de criterios marcando un valor en cada columna. Así como también, disponemos en la corrección de los ítems indicando sus observaciones y/o sugerencias.

ITEM	CRITERIOS	VALORACIONES				
		Deficiente 0-20%	Regular 21-50%	Buena 51-20%	Muy Buena 71-80%	Excelente 81-100%
Claridad	Cuenta con un lenguaje apropiado					85
Objetividad	Está expresado en conducta observable					85
Organización	Es adecuado a la vanguardia de la tecnología					95
Suficiencia	Comprende los aspectos de cantidad y calidad					95
Intencionalidad	Adecuado para valorar los aspectos del sistema metodológico y científico					95
Consistencia	Está basado en aspectos técnicos, científicos acordes a la tecnología adecuada					95
Coherencia	Entre los índices, indicadores y dimensiones					95
Metodología	Responde el propósito del trabajo bajo los objetivos a lograr					95
Pertinencia	El instrumento es adecuado al tipo de investigación					95

 Promedio de Valoración: 92%

Opción de aplicabilidad:

- () El instrumento puede ser aplicado, tal como está elaborado.
 () El instrumento debe ser mejorado, antes de ser aplicado.



Firma de Experto

TABLA DE EVALUACIÓN DE EXPERTOS

Apellidos y Nombres del Experto: Gálvez Tapia Orleans
Moisés

Título y/o grado: Magíster en Ingeniería de Sistemas

Universidad donde labora: Universidad César Vallejo

Fecha: 07/07/2024

TÍTULO DEL PROYECTO

Algoritmo de Árbol de Decisión con Metodología Crisp y su Influencia en la Predicción de Lluvias, Chosica 2024

Tabla evaluación de Expertos para los instrumentos de medición. Mediante la siguiente tabla de evaluación de expertos, usted tiene potestad de calificar los instrumentos que son utilizados para medir los distintos indicadores, mediante una serie de criterios marcando un valor en cada columna. Así como también, disponemos en la corrección de los ítems indicando sus observaciones y/o sugerencias.

ITEM	CRITERIOS	VALORACIONES				
		Deficiente 0-20%	Regular 21-50%	Bueno 51-70%	Muy Bueno 71-80%	Excelente 81-100%
Claridad	Cuenta con un lenguaje apropiado					85%
Objetividad	Está expresado en conducta observable					90%
Organización	Es adecuado a la vanguardia de la tecnología					90%
Suficiencia	Comprende los aspectos de cantidad y calidad					85%
Intencionalidad	Adecuado para valorar los aspectos del sistema metodológico y científico					85%
Consistencia	Está basado en aspectos técnicos, científicos acordes a la tecnología adecuada					85%
Coherencia	Entre los índices, indicadores y dimensiones					85%
Metodología	Responde el propósito del trabajo bajo los objetivos a lograr					85%
Pertinencia	El instrumento es adecuado al tipo de investigación					85%

Promedio de Valoración: 90%

Opción de aplicabilidad:

(X) El instrumento puede ser aplicado, tal como está elaborado.

() El instrumento debe ser mejorado, antes de ser aplicado.



Ficha de registro – Test del indicador (Exactitud)

Investigadores				
Institución Investigada				
Variable				
Dimensión				
Tiempo				
Indicador				
Descripción				
Unidad de Medida				
Fórmula				
ITEM	FECHA	Cantidad de Predicciones Correctas	Total de Predicciones	Porcentaje (%)
1				
2				
3				
...				
....				
....				
....				
30				
		TOTAL		

Ficha de registro – Test del indicador (Número de Aciertos)

Investigadores				
Institución Investigada				
Variable				
Dimensión				
Tiempo				
Indicador				
Descripción				
Unidad de Medida				
Fórmula				
ITEM	FECHA	Numero de Aciertos	Total de Predicciones	Porcentaje (%)
1				
2				
3				
...				
....				
....				
.....				
30				
	TOTAL			

Ficha de registro – Test del indicador (Recall)

Investigadores				
Institución Investigada				
Variable				
Dimensión				
Tiempo				
Indicador				
Descripción				
Unidad de Medida				
Fórmula				
ITEM	FECHA	TP	FN	Recall (%)
1				
2				
3				
...				
....				
....				
.....				
30				
TOTAL				

Anexo 3: Código Web

```
<!DOCTYPE html>

<html lang="es">

<head>

    <meta charset="UTF-8">

    <meta name="viewport" content="width=device-width,
initial-scale=1.0">

    <title>Predicción de Lluvias</title>

    <link rel="stylesheet" href="styles.css">

    <link
href="https://fonts.googleapis.com/css2?family=Open+Sans:wght@4
00;600;700&display=swap" rel="stylesheet">

    <script src="https://cdn.jsdelivr.net/npm/chart.js"></script>

    <script
src="https://cdnjs.cloudflare.com/ajax/libs/html2canvas/0.4.1/html2
canvas.min.js"></script>

    <script
src="https://cdnjs.cloudflare.com/ajax/libs/jspdf/1.3.4/jspdf.min.js">
</script>

</head>

<body>

    <!-- Contenedor de inicio de sesión -->

    <div id="login-container" class="login-container">

        <h2>Iniciar Sesión</h2>

        <div class="form-group">

            <label for="login-username">Usuario:</label>

            <input type="text" id="login-username">

        </div>

        <div class="form-group">
```

```
<label for="login-password">Contraseña:</label>

<input type="password" id="login-password">

</div>

<button onclick="login()">Ingresar</button>

<div id="loginMessage"></div>

</div>

<!-- Contenido principal que se muestra después de iniciar
sesión -->

<div id="main-content" style="display: none;">

<header>

<h1>PREDICCIÓN DE LLUVIAS</h1>

</header>

<div class="container">

<nav>

<ul>

<li><a href="#" onclick="openTab('prediccion')"
class="active">Predicción</a></li>

<li><a href="#"
onclick="openTab('historico')">Histórico</a></li>

<li><a href="#"
onclick="openTab('proyeccion')">Proyección</a></li>

<li><a href="#" onclick="openTab('verificacion')">Verificar
Conexión</a></li>

<li><a href="#" onclick="logout()">Cerrar Sesión</a></li>

</ul>

</nav>

<div class="content">

<!-- Contenido de la pestaña de predicción -->
```

```
<div id="prediccion" class="tabContent active">
  <div class="date-selector">
    <label for="yearSelect">Año:</label>
    <select id="yearSelect"
onchange="updateCalendar();" class="day-button"></select>
    <label for="monthSelect">Mes:</label>
    <select id="monthSelect"
onchange="updateCalendar();" class="day-button"></select>
  </div>
  <div id="calendar-comment-container">
    <div id="calendar"></div>
    <div id="messageContainer"></div>
  </div>
</div>
<!-- Contenido de la pestaña de histórico -->
<div id="historico" class="tabContent">
  <div class="container-inner">
    <div class="selector-container horizontal">
      <label for="yearStart">Año Inicio:</label>
      <select id="yearStart" class="day-button"></select>
      <label for="yearEnd">Año Fin:</label>
      <select id="yearEnd" class="day-button"></select>
      <button class="day-button"
onclick="consultarHistorico()">Consultar</button>
      <div class="image-button">
        
      </div>
    </div>
  </div>
</div>
```

```
<div class="graph-container">
  <canvas id="historicoChart"></canvas>
</div>
</div>
</div>
</div>
<!-- Contenido de la pestaña de proyección -->
<div id="proyeccion" class="tabContent">
  <div class="container-inner">
    <div class="selector-container horizontal">
      <label for="yearProjection">Año:</label>
      <select id="yearProjection" class="day-
button"></select>
      <button class="day-button"
onclick="consultarProyeccion()">Consultar</button>
      <div class="image-button">
        
      </div>
    </div>
    <div class="graph-container">
      <canvas id="proyeccionChart"></canvas>
    </div>
  </div>
</div>
</div>
<!-- Contenido de la pestaña de verificación -->
<div id="verificacion" class="tabContent">
  <div class="container-inner">
    <button class="day-button"
onclick="verificarConexion()">Verificar Conexión</button>
```

```
        <div id="verificacionResultado"></div>
    </div>
</div>
</div>
</div>
</div>
</div>
<script src="scripts.js"></script>
</body>
</html>
```

Anexo 4 : Código Scripts

```
let historicoChart = null;
let proyeccionChart = null;

// Función para cambiar de pestaña
function openTab(tabName) {
    var tabs = document.querySelectorAll('.tabContent');
    tabs.forEach(tab => {
        tab.style.display = 'none';
    });
    document.getElementById(tabName).style.display = 'block';
    if (tabName === 'prediccion') {
        fillYearSelector();
        fillMonthSelector();
        updateCalendar();
    } else if (tabName === 'historico') {
        fillYearRangeSelectors();
    }
}
```

```
    } else if (tabName === 'proyeccion') {  
        fillProjectionYearSelector();  
    }  
}  
  
// Función para verificar la conexión con el servidor  
function verificarConexion() {  
    fetch('http://127.0.0.1:5000/verificar_conexion')  
        .then(response => response.json())  
        .then(data => {  
            const resultadoDiv =  
document.getElementById('verificacionResultado');  
            if (data.exito) {  
                resultadoDiv.textContent = 'Conexión exitosa';  
                resultadoDiv.style.color = 'green';  
            } else {  
                resultadoDiv.textContent = 'Fallo en la conexión: ' +  
data.error;  
                resultadoDiv.style.color = 'red';  
            }  
        })  
        .catch(error => {  
            const resultadoDiv =  
document.getElementById('verificacionResultado');  
            resultadoDiv.textContent = 'Error en la conexión: ' + error;  
            resultadoDiv.style.color = 'red';  
        });  
}
```

```
// Funciones para llenar los selectores de año y mes

function fillYearSelector() {

    const yearSelect = document.getElementById('yearSelect');

    const currentYear = new Date().getFullYear();

    yearSelect.innerHTML = "";

    for (let year = currentYear - 10; year <= currentYear + 10;
year++) {

        let option = document.createElement('option');

        option.value = year;

        option.text = year;

        yearSelect.appendChild(option);

    }

    yearSelect.value = currentYear;

}

function fillMonthSelector() {

    const monthSelect =
document.getElementById('monthSelect');

    monthSelect.innerHTML = "";

    const months = ["Enero", "Febrero", "Marzo", "Abril",
"Mayo", "Junio",
    "Julio", "Agosto", "Septiembre", "Octubre", "Noviembre",
"Diciembre"];

    months.forEach((month, index) => {

        let option = document.createElement('option');

        option.value = index;

        option.text = month;

    });

}
```

```
        monthSelect.appendChild(option);

    });

    monthSelect.value = new Date().getMonth();

}

// Función para actualizar el calendario basado en el año y mes
seleccionados

function updateCalendar() {

    const year = document.getElementById('yearSelect').value;

    const month =
document.getElementById('monthSelect').value;

    const firstDay = new Date(year, month, 1);

    const lastDay = new Date(year, parseInt(month) + 1, 0);

    const calendarDiv = document.getElementById('calendar');
    calendarDiv.innerHTML = '';

    const dayNames = ['D', 'L', 'M', 'X', 'J', 'V', 'S'];

    const headerRow = document.createElement('div');
    headerRow.className = 'calendar-row';

    dayNames.forEach(dayName => {

        const dayNameDiv = document.createElement('div');
        dayNameDiv.className = 'calendar-cell header-cell';
        dayNameDiv.textContent = dayName;

        headerRow.appendChild(dayNameDiv);

    });

    calendarDiv.appendChild(headerRow);
```

```
let row = document.createElement('div');

row.className = 'calendar-row';

for (let i = 0; i < firstDay.getDay(); i++) {

const emptyCell = document.createElement('div');

emptyCell.className = 'calendar-cell';

row.appendChild(emptyCell);

}

for (let day = 1; day <= lastDay.getDate(); day++) {

if (row.childNodes.length % 7 === 0 && day > 1) {

calendarDiv.appendChild(row);

row = document.createElement('div');

row.className = 'calendar-row';

}

const dayDiv = document.createElement('button');

dayDiv.className = 'calendar-cell day-button';

dayDiv.textContent = day;

dayDiv.onclick = function() { displayMessage(year, month,
day); };

row.appendChild(dayDiv);

}

while (row.childNodes.length < 7) {

const emptyCell = document.createElement('div');

emptyCell.className = 'calendar-cell';

row.appendChild(emptyCell);

}

calendarDiv.appendChild(row);
```

```
}

// Función para mostrar el mensaje de predicción de lluvias

function displayMessage(year, month, day) {

    const messageContainer =
document.getElementById('messageContainer');

fetch(`http://127.0.0.1:5000/prediccion_diaria?year=${year}&month
=${month}&day=${day}`)

    .then(response => response.json())

    .then(data => {

        const date = new Date(year, month, day);

        const options = { year: 'numeric', month: 'long', day: 'numeric'
};

        const dateString = date.toLocaleDateString("es-ES", options);

        const probabilidad = data.probabilidad;

        messageContainer.innerHTML = `El día ${dateString} hay un
${probabilidad}% de probabilidad de lluvia. <br>

        <small>Esta predicción se logra gracias al análisis de los datos
históricos de temperatura y humedad de la zona de
Chosica.</small>`;

    })

    .catch(error => {

        messageContainer.innerHTML = 'Error al obtener la
predicción.';

        console.error('Error:', error);

    });

}
```

```
// Funciones para llenar los selectores de rango de años y año de proyección
```

```
function fillYearRangeSelectors() {  
  
    const yearStart = document.getElementById('yearStart');  
  
    const yearEnd = document.getElementById('yearEnd');  
  
    const currentYear = new Date().getFullYear();  
  
    [yearStart, yearEnd].forEach(selector => {  
  
        selector.innerHTML = "";  
  
        for (let year = currentYear - 10; year <= currentYear; year++) {  
  
            let option = document.createElement('option');  
  
            option.value = year;  
  
            option.text = year;  
  
            selector.appendChild(option);  
  
        }  
  
    });  
  
    yearStart.value = currentYear - 5;  
  
    yearEnd.value = currentYear;  
  
}
```

```
function fillProjectionYearSelector() {  
  
    const yearProjection =  
document.getElementById('yearProjection');  
  
    const currentYear = new Date().getFullYear();  
  
    yearProjection.innerHTML = "";  
  
    for (let year = currentYear; year <= currentYear + 10; year++) {  
  
        let option = document.createElement('option');  
  
        option.value = year;
```

```
    option.text = year;

    yearProjection.appendChild(option);
  }

  yearProjection.value = currentYear;
}

// Función para consultar el histórico de lluvias
function consultarHistorico() {

  const yearStart =
document.getElementById('yearStart').value;

  const yearEnd = document.getElementById('yearEnd').value;

  fetch(`http://127.0.0.1:5000/historico?yearStart=${yearStart}&yearE
nd=${yearEnd}`)

    .then(response => response.json())

    .then(data => {

  if (historicoChart) {

    historicoChart.destroy();

  }

  const ctx =
document.getElementById('historicoChart').getContext('2d');

  historicoChart = new Chart(ctx, {

  type: 'line',

  data: {

    labels: data.labels,

    datasets: [{

      label: 'Porcentaje Anual de Ocurrencia de Lluvias',
```

```
    data: data.values,

    borderColor: 'rgba(75, 192, 192, 1)',

    borderWidth: 1

  }
},

options: {

  responsive: true,

  scales: {

    x: {

      display: true,

      title: {

        display: true,

        text: 'Años'

      }

    },

    y: {

      display: true,

      title: {

        display: true,

        text: 'Porcentaje de Lluvias'

      },

      ticks: {

        callback: function(value) { return value + "%" }

      }

    }

  }

}
```

```
    });  
  });  
}  
  
// Función para consultar la proyección de lluvias  
function consultarProyeccion() {  
  const yearProjection =  
    document.getElementById('yearProjection').value;  
  
  fetch(`http://127.0.0.1:5000/proyeccion?yearProjection=${yearProjection}`)  
    .then(response => response.json())  
    .then(data => {  
      if (proyeccionChart) {  
        proyeccionChart.destroy();  
      }  
  
      const ctx =  
        document.getElementById('proyeccionChart').getContext('2d');  
  
      proyeccionChart = new Chart(ctx, {  
        type: 'line',  
        data: {  
          labels: data.labels,  
          datasets: [{  
            label: 'Proyección de Ocurrencia de Lluvias',  
            data: data.values,  
            borderColor: 'rgba(153, 102, 255, 1)',  
            borderWidth: 1,  
          }],  
        },  
      });  
    });  
}
```

```
        fill: false,  
    }  
},  
options: {  
    responsive: true,  
    scales: {  
        x: {  
            display: true,  
            title: {  
                display: true,  
                text: 'Años'  
            }  
        },  
        y: {  
            display: true,  
            title: {  
                display: true,  
                text: 'Porcentaje de Lluvias'  
            },  
            ticks: {  
                callback: function(value) { return value + "%" }  
            }  
        }  
    }  
}  
});  
});
```

```
}
```

```
// Función para descargar el gráfico como PDF
```

```
function downloadPDF(tab) {
```

```
    let chartElement, chartTitle;
```

```
    if (tab === 'historico') {
```

```
        chartElement = document.getElementById('historicoChart');
```

```
        chartTitle = 'Gráfico Histórico de Lluvias';
```

```
    } else if (tab === 'proyeccion') {
```

```
        chartElement =  
document.getElementById('proyeccionChart');
```

```
        chartTitle = 'Gráfico de Proyección de Lluvias';
```

```
    }
```

```
    if (chartElement && chartElement.getContext &&  
chartElement.width > 0 && chartElement.height > 0) {
```

```
        console.log(`Chart detected for ${chartTitle}`);
```

```
        html2canvas(chartElement).then(canvas => {
```

```
            const imgData = canvas.toDataURL('image/png');
```

```
            const pdf = new jsPDF('landscape');
```

```
            pdf.addImage(imgData, 'PNG', 10, 10);
```

```
            pdf.text(10, 10, chartTitle);
```

```
            pdf.save(`${chartTitle}.pdf`);
```

```
        });
```

```
    } else {
```

```
        console.log(`No chart detected for ${chartTitle}`);
```

```
        alert('No hay gráfica para descargar.');
```

```
    }
```

```
}

// Función para iniciar sesión

function login() {

    const username = document.getElementById('login-username').value;

    const password = document.getElementById('login-password').value;

    fetch('http://127.0.0.1:5000/login', {
        method: 'POST',
        headers: {
            'Content-Type': 'application/json'
        },
        body: JSON.stringify({ username: username, password: password })
    })
    .then(response => response.json())
    .then(data => {
        const loginMessage = document.getElementById('loginMessage');

        if (data.exito) {
            document.getElementById('login-container').style.display = 'none';

            document.getElementById('main-content').style.display = 'block';

        } else {
            loginMessage.textContent = 'Usuario o contraseña incorrectos.';
        }
    })
}
```

```
    })

    .catch(error => {

        const loginMessage =
document.getElementById('loginMessage');

        loginMessage.textContent = 'Error al intentar iniciar sesión.';

        console.error('Error:', error);

    });
}

// Función para cerrar sesión

function logout() {

    document.getElementById('main-content').style.display = 'none';

    document.getElementById('login-container').style.display = 'block';

}

// Inicialización de la página

window.onload = function() {

    document.getElementById('login-container').style.display =
'block';

    document.getElementById('main-content').style.display = 'none';

}

document.querySelectorAll('.image-button img[alt="Descargar
PDF"]').forEach(button => {

    button.onclick = function() {

        const activeTab =
document.querySelector('.tabContent:not([style*="display:
none"])').id;

        downloadPDF(activeTab);

    }

}
```

```
    }  
  });  
  
window.onload = function() {  
    openTab('prediccion');  
};
```

Anexo 5: App (Algoritmo y Funcionalidad)

```
from flask import Flask, jsonify, request  
from flask_cors import CORS  
import pyodbc  
import numpy as np  
import pandas as pd  
from sklearn.model_selection import train_test_split, GridSearchCV  
from sklearn.tree import DecisionTreeClassifier,  
DecisionTreeRegressor  
from sklearn.preprocessing import StandardScaler,  
PolynomialFeatures  
from sklearn.calibration import CalibratedClassifierCV  
from sklearn.linear_model import LinearRegression  
from imblearn.over_sampling import SMOTE # Sobremuestreo  
  
app = Flask(__name__)  
CORS(app) # Esto habilita CORS para todas las rutas  
  
# Función para obtener la conexión a la base de datos  
def get_db_connection():  
    conn = pyodbc.connect('DRIVER={ODBC Driver 18 for SQL  
Server};'
```

```
'SERVER=localhost;'

'DATABASE=AAD;'

'UID=sa;'

'PWD=Arcobaleno#69#;'

'TrustServerCertificate=yes;')

return conn

@app.route('/login', methods=['POST'])

def login():

    data = request.get_json()

    username = data['username']

    password = data['password']

    conn = get_db_connection()

    cursor = conn.cursor()

    cursor.execute("SELECT NombreUsuario, ContraseñaHash
FROM Usuarios WHERE NombreUsuario = ? AND ContraseñaHash =
?", (username, password))

    user = cursor.fetchone()

    conn.close()

    if user:

        return jsonify(exito=True)

    else:

        return jsonify(exito=False)

@app.route('/verificar_conexion')

def verificar_conexion():
```

```
try:

conn = get_db_connection()

conn.close()

return jsonify(exito=True)

except Exception as e:

return jsonify(exito=False, error=str(e))

@app.route('/historico')

def historico():

    year_start = int(request.args.get('yearStart'))

    year_end = int(request.args.get('yearEnd'))

    conn = get_db_connection()

    cursor = conn.cursor()

    cursor.execute("""

SELECT YEAR(Fecha) AS Year,

SUM(CAST(Precipitacion AS FLOAT)) / COUNT(*) * 100 AS

PercentPrecip,

AVG(CAST(TemperaturaMax AS FLOAT)) AS TempMaxAvg,

AVG(CAST(TemperaturaMin AS FLOAT)) AS TempMinAvg,

AVG(CAST(HumedadRelativa AS FLOAT)) AS HumedadAvg

FROM dbo.Lluvias

WHERE YEAR(Fecha) BETWEEN ? AND ?

AND YEAR(Fecha) NOT IN (2021, 2022) -- Omitir años 2021 y

2022

GROUP BY YEAR(Fecha)

ORDER BY YEAR(Fecha) ASC

""", year_start, year_end)
```

```
rows = cursor.fetchall()

labels = [row.Year for row in rows]

values = [row.PercentPrecip for row in rows]

temp_max_avg = [row.TempMaxAvg for row in rows]

temp_min_avg = [row.TempMinAvg for row in rows]

humedad_avg = [row.HumedadAvg for row in rows]

conn.close()

return jsonify(labels=labels, values=values,
temp_max_avg=temp_max_avg, temp_min_avg=temp_min_avg,
humedad_avg=humedad_avg)

@app.route('/proyeccion')

def proyeccion():

    year_projection = int(request.args.get('yearProjection'))

    conn = get_db_connection()

    cursor = conn.cursor()

    cursor.execute("""

SELECT YEAR(Fecha) AS Year,

SUM(CAST(Precipitacion AS FLOAT)) / COUNT(*) * 100 AS
PercentPrecip,

AVG(CAST(TemperaturaMax AS FLOAT)) AS TempMaxAvg,

AVG(CAST(TemperaturaMin AS FLOAT)) AS TempMinAvg,

AVG(CAST(HumedadRelativa AS FLOAT)) AS HumedadAvg

FROM dbo.Lluvias

WHERE YEAR(Fecha) < ?
```

```
AND YEAR(Fecha) NOT IN (2021, 2022) -- Omitir años 2021 y
2022

GROUP BY YEAR(Fecha)

ORDER BY YEAR(Fecha) ASC

"", year_projection)

rows = cursor.fetchall()

# Extraer datos históricos para proyección

years = np.array([row.Year for row in rows])

percents = np.array([row.PercentPrecip for row in rows])

temp_max_avg = np.array([row.TempMaxAvg for row in
rows])

temp_min_avg = np.array([row.TempMinAvg for row in
rows])

humedad_avg = np.array([row.HumedadAvg for row in rows])

# Ajustar la escala de los datos

scaler_features = StandardScaler()

scaler_percents = StandardScaler()

features = np.vstack([years, temp_max_avg, temp_min_avg,
humedad_avg]).T

features_scaled = scaler_features.fit_transform(features)

percents_scaled =
scaler_percents.fit_transform(percents.reshape(-1, 1))

# Ajustar un modelo de regresión de árbol de decisión

param_grid = {

'max_depth': [3, 5, 7, 10],
```

```
'min_samples_split': [2, 5, 10],

'min_samples_leaf': [1, 2, 4]

}

splits = min(len(years), 3) # Ajustar la cantidad de splits según
la cantidad de datos

model =
GridSearchCV(DecisionTreeRegressor(random_state=42), param_grid,
cv=splits)

model.fit(features_scaled, percents_scaled.ravel())

# Proyectar los próximos 5 años

future_years = np.array([year_projection + i for i in range(5)])

future_temp_max_avg = np.full(5, temp_max_avg.mean()) #
Usar la media histórica para las proyecciones

future_temp_min_avg = np.full(5, temp_min_avg.mean())

future_humedad_avg = np.full(5, humedad_avg.mean())

future_features = np.vstack([

future_years,

future_temp_max_avg,

future_temp_min_avg,

future_humedad_avg

]).T

future_features_scaled =
scaler_features.transform(future_features)

future_percents_scaled =
model.predict(future_features_scaled)
```

```
future_percents =
scaler_percents.inverse_transform(future_percents_scaled.reshape(-
1, 1)).ravel()

# Agregar variabilidad aleatoria controlada

historical_std = percents.std()

variability_factor = 0.1 # Factor de escala para reducir el
impacto de la variabilidad

future_percents_with_variability = future_percents +
np.random.normal(0, historical_std * variability_factor,
size=future_percents.shape)

conn.close()

return jsonify(labels=future_years.tolist(),
values=future_percents_with_variability.tolist())

@app.route('/prediccion_diaria')

def prediccion_diaria():

    year = int(request.args.get('year'))

    month = int(request.args.get('month')) + 1 # Los meses en
JavaScript son de 0-11, pero en SQL son de 1-12

    day = int(request.args.get('day'))

    conn = get_db_connection()

    query = """

    SELECT TemperaturaMax, TemperaturaMin,
HumedadRelativa, Precipitacion, MONTH(Fecha) as Mes, DAY(Fecha)
as Dia, YEAR(Fecha) as Year

    FROM dbo.Lluvias

    """

    df = pd.read_sql(query, conn)
```

```
conn.close()

if (df.empty):

return jsonify(probabilidad=0) # Si no hay datos, devuelve 0

# Preparación de los datos históricos para entrenar el modelo

df['Temp_Diff'] = df['TemperaturaMax'] -
df['TemperaturaMin']

df['Hum_Temp_Ratio'] = df['HumedadRelativa'] /
(df['TemperaturaMax'] + 1)

df['Lluvia'] = df['Precipitacion'].apply(lambda x: 1 if x > 0 else
0)

# Balancear las clases con SMOTE

X = df[['TemperaturaMax', 'TemperaturaMin',
'HumedadRelativa', 'Temp_Diff', 'Hum_Temp_Ratio']]

y = df['Lluvia']

smote = SMOTE(random_state=42)

X_res, y_res = smote.fit_resample(X, y)

# Entrenamiento del modelo con calibración

X_train, X_test, y_train, y_test = train_test_split(X_res, y_res,
test_size=0.2, random_state=42)

base_model = DecisionTreeClassifier(random_state=42)

base_model.fit(X_train, y_train)

model = CalibratedClassifierCV(base_model, cv='prefit')

model.fit(X_train, y_train)

# Datos históricos para la fecha solicitada
```

```
df_date = df[(df['Mes'] == month) & (df['Dia'] == day)]

if df_date.empty:

    # Si no hay datos específicos para la fecha, usar la media de
    los datos históricos para esa fecha en años anteriores

    temp_max_mean = df[(df['Mes'] == month) & (df['Dia'] ==
day)]['TemperaturaMax'].mean()

    temp_min_mean = df[(df['Mes'] == month) & (df['Dia'] ==
day)]['TemperaturaMin'].mean()

    humedad_mean = df[(df['Mes'] == month) & (df['Dia'] ==
day)]['HumedadRelativa'].mean()

    temp_diff_mean = temp_max_mean - temp_min_mean

    hum_temp_ratio_mean = humedad_mean /
(temp_max_mean + 1)

    X_date = pd.DataFrame([[temp_max_mean,
temp_min_mean, humedad_mean, temp_diff_mean,
hum_temp_ratio_mean]],

        columns=['TemperaturaMax', 'TemperaturaMin',
'HumedadRelativa', 'Temp_Diff', 'Hum_Temp_Ratio'])

    else:

        df_date['Temp_Diff'] = df_date['TemperaturaMax'] -
df_date['TemperaturaMin']

        df_date['Hum_Temp_Ratio'] = df_date['HumedadRelativa'] /
(df_date['TemperaturaMax'] + 1)

        X_date = df_date[['TemperaturaMax', 'TemperaturaMin',
'HumedadRelativa', 'Temp_Diff', 'Hum_Temp_Ratio']]

    probabilidad = model.predict_proba(X_date)[: , 1][0] * 100 #
Convertir la predicción en porcentaje

    return jsonify(probabilidad=probabilidad)
```

```
if __name__ == '__main__':  
    app.run(debug=True, port=5000)
```

Anexo 6: Código Styles

```
body {  
    font-family: Helvetica;  
    margin: 0;  
    padding: 0;  
    background-color: #181818;  
    color: #fff;  
}  
  
header {  
    width: 100%;  
    background-color: #202020;  
    color: #fff;  
    padding: 20px 0;  
    text-align: center;  
    border-bottom: 1px solid #333;  
}  
  
/* Contenedor principal */  
.container {  
    display: flex;  
    min-height: 100vh;  
}
```

```
nav {  
    flex-basis: 20%;  
    background-color: #202020;  
    border-right: 1px solid #333;  
    padding: 20px;  
}  
  
nav ul {  
    list-style-type: none;  
    padding: 0;  
}  
  
nav ul li {  
    margin-bottom: 10px;  
}  
  
nav ul li a {  
    color: #bbb;  
    text-decoration: none;  
    display: block;  
    padding: 10px;  
    border-radius: 4px;  
    font-weight: bold; /* Texto en negrita */  
}  
  
nav ul li a:hover, nav ul li a.active {  
    background-color: #333;
```

```
        color: #fff;
    }

    /* Contenido de las pestañas */
    .content {
        flex: 1;
        padding: 20px;
        background-color: #181818;
    }

    .tabContent {
        display: none;
    }

    .tabContent.active {
        display: block;
    }

    /* Estilo de los selectores y gráficos */
    .container-inner {
        display: flex;
        flex-direction: column;
        align-items: center;
        gap: 20px;
    }

    .selector-container {
```

```
        display: flex;

        flex-direction: column;

        gap: 10px;

        align-items: center;
    }

.selector-container.horizontal {

    flex-direction: row;

    gap: 20px;

    align-items: center;
}

.graph-container {

    display: flex;

    justify-content: center;

    width: 100%;

    background-color: #202020; /* Fondo original */

    border-radius: 8px; /* Esquinas redondeadas */
}

/* Estilo del calendario */

#calendar-comment-container {

    display: flex;

    justify-content: space-between;

    align-items: flex-start;

    gap: 20px;
}
```

```
#calendar {  
  
    display: grid;  
  
    grid-template-columns: repeat(7, 1fr);  
  
    gap: 5px;  
  
    padding: 10px;  
  
    background-color: #202020;  
  
    border: 1px solid #333;  
  
    border-radius: 8px;  
  
    flex-basis: 70%; /* Tamaño relativo al contenedor */  
  
}
```

```
.calendar-row {  
  
    display: contents;  
  
}
```

```
.calendar-cell {  
  
    min-width: 30px;  
  
    min-height: 30px;  
  
    background-color: #333;  
  
    text-align: center;  
  
    line-height: 30px;  
  
    border: 1px solid #444;  
  
    cursor: pointer;  
  
    color: #fff;  
  
    border-radius: 4px;  
  
}
```

```
.calendar-cell.header-cell {  
    font-weight: bold;  
    background-color: #444;  
    cursor: default;  
}  
  
/* Estilo de los botones */  
.day-button {  
    border: none;  
    outline: none;  
    background-color: #333;  
    cursor: pointer;  
    color: #fff;  
    padding: 5px;  
    border-radius: 4px;  
}  
  
.day-button:hover {  
    background-color: #555;  
}  
  
#messageContainer {  
    padding: 10px;  
    border: 1px solid #333;  
    background-color: #202020;  
    border-radius: 8px;
```

```
flex-basis: 30%; /* Tamaño relativo al contenedor */  
  
min-height: 50px;  
  
}
```

```
/* Estilo del contenedor de login */
```

```
.login-container {  
  
    width: 300px;  
  
    margin: 100px auto;  
  
    padding: 20px;  
  
    background-color: #333;  
  
    box-shadow: 0 0 10px rgba(0, 0, 0, 0.5);  
  
    border-radius: 10px;  
  
}
```

```
.login-container h2 {  
  
    margin: 0 0 20px;  
  
    text-align: center;  
  
}
```

```
.form-group {  
  
    margin-bottom: 15px;  
  
}
```

```
.form-group label {  
  
    display: block;  
  
    margin-bottom: 5px;  
  
}
```

```
.form-group input {  
    width: 100%;  
    padding: 8px;  
    box-sizing: border-box;  
    border: 1px solid #555;  
    border-radius: 5px;  
    background-color: #222;  
    color: #fff;  
}
```

```
button {  
    width: 100%;  
    padding: 10px;  
    background-color: #444;  
    border: none;  
    color: #fff;  
    cursor: pointer;  
    border-radius: 5px;  
}
```

```
button:hover {  
    background-color: #555;  
}
```

```
#loginMessage {  
    color: red;
```

```
text-align: center;  
margin-top: 10px;  
}
```

Anexo 7: Base de Datos

-- Crear la base de datos AAD

```
CREATE DATABASE AAD;
```

```
GO
```

-- Usar la base de datos AAD

```
USE AAD;
```

```
GO
```

-- Crear la tabla de Departamentos

```
CREATE TABLE Departamentos (  
    DepartamentoID INT IDENTITY(1,1) PRIMARY KEY,  
    NombreDepartamento VARCHAR(100)  
);
```

```
GO
```

-- Crear la tabla de Provincias

```
CREATE TABLE Provincias (  
    ProvinciaID INT IDENTITY(1,1) PRIMARY KEY,  
    NombreProvincia VARCHAR(100),  
    DepartamentoID INT FOREIGN KEY REFERENCES  
    Departamentos(DepartamentoID)  
);
```

GO

-- Crear la tabla de Distritos

```
CREATE TABLE Distritos (  
    DistritoID INT IDENTITY(1,1) PRIMARY KEY,  
    NombreDistrito VARCHAR(100),  
    ProvinciaID INT FOREIGN KEY REFERENCES  
    Provincias(ProvinciaID)  
);
```

GO

-- Crear la tabla de Ubicaciones

```
CREATE TABLE Ubicaciones (  
    UbicacionID INT IDENTITY(1,1) PRIMARY KEY,  
    Latitud DECIMAL(9,6),  
    Longitud DECIMAL(9,6),  
    Altitud DECIMAL(6,2),  
    DistritoID INT FOREIGN KEY REFERENCES Distritos(DistritoID)  
);
```

GO

-- Crear la tabla de Estaciones

```
CREATE TABLE Estaciones (  
    EstacionID INT IDENTITY(1,1) PRIMARY KEY,  
    NombreEstacion VARCHAR(100),  
    TipoEstacion VARCHAR(50),  
    UbicacionID INT FOREIGN KEY REFERENCES  
    Ubicaciones(UbicacionID)
```

);

GO

-- Crear la tabla de Lluvias con la estructura modificada

CREATE TABLE Lluvias (

 RegistroID INT IDENTITY(1,1) PRIMARY KEY,

 EstacionID INT FOREIGN KEY REFERENCES
Estaciones(EstacionID),

 Fecha DATE,

 TemperaturaMax DECIMAL(5, 2),

 TemperaturaMin DECIMAL(5, 2),

 HumedadRelativa DECIMAL(5, 2),

 Precipitacion BIT

);

GO

-- Crear la tabla de Usuarios

CREATE TABLE Usuarios (

 UsuarioID INT IDENTITY(1,1) PRIMARY KEY,

 NombreUsuario VARCHAR(50) UNIQUE,

 ContrasenaHash VARCHAR(256),

 Email VARCHAR(100),

 Activo BIT,

 Rol VARCHAR(50)

);

GO

```
-- Crear roles de acceso
```

```
CREATE ROLE AdminRole;
```

```
CREATE ROLE ConsultaRole;
```

```
GO
```

```
-- Asignar permisos a los roles
```

```
GRANT SELECT, INSERT, UPDATE, DELETE ON dbo.Lluvias TO  
AdminRole;
```

```
GRANT SELECT ON dbo.Lluvias TO ConsultaRole;
```

```
GO
```

```
CREATE LOGIN arotru WITH PASSWORD = 'Tesis2024#';
```

```
CREATE LOGIN SOS WITH PASSWORD = 'ChosicaSOS';
```

```
CREATE USER arotru FOR LOGIN arotru;
```

```
CREATE USER SOS FOR LOGIN SOS;
```

```
ALTER ROLE AdminRole ADD MEMBER arotru;
```

```
ALTER ROLE ConsultaRole ADD MEMBER SOS;
```

```
INSERT INTO Usuarios (NombreUsuario, ContraseñaHash, Email,  
Activo, Rol)
```

```
VALUES
```

```
('arotru', HASHBYTES('SHA2_256', 'Tesis2024#'),  
'arotru@example.com', 1, 'AdminRole'),
```

```
('SOS', HASHBYTES('SHA2_256', 'ChosicaSOS'), 'sos@example.com', 1,  
'ConsultaRole');
```

Anexo 8: Modelo DB

