



UNIVERSIDAD TÉCNICA DEL NORTE

Facultad de Ingeniería en Ciencias Aplicadas

Carrera de Ingeniería en Mecatrónica

**SISTEMA INTELIGENTE PARA PREDICCIÓN DE VARIABLES CLIMÁTICAS
LOCALES**

Trabajo de grado previo a la obtención del título de Ingeniero en Mecatrónica

Autor:

Robert Michael Alvarez Carasayo

Director:

Carlos Xavier Rosero Chandi

Ibarra - Ecuador

2025



Universidad Técnica del Norte
Biblioteca Universitaria
Autorización de uso y publicación a favor de la
Universidad Técnica del Norte
Identificación de la Obra

En cumplimiento del Art. 144 de la Ley de Educación Superior, hago la entrega del presente trabajo a la Universidad Técnica del Norte para que sea publicado en el Repositorio Digital Institucional, para lo cual pongo a disposición la siguiente información:

Datos del Autor			
Cédula de identidad	1752863629		
Apellidos y Nombres	Alvarez Carasayo Robert Michael		
Dirección	Panamericana Norte E35 y Lucila Molineros, Cayambe		
Email	rmalvarezc@utn.edu.ec		
Teléfono fijo	022164153	Teléfono móvil	0997965505
Datos de la Obra			
Tema	SISTEMA INTELIGENTE PARA PREDICCIÓN DE VARIABLES CLIMÁTICAS LOCALES		
Apellidos y Nombres	Alvarez Carasayo Robert Michael		
Fecha	27/02/2025		
Programa	Pregrado		
Título por el que opta	Ingeniero en Mecatrónica		
Director	Carlos Xavier Rosero Chandi		



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS
CONSTANCIA

El autor manifiesta que la obra objeto de la presente autorización es original y se la desarrolló sin violar derechos de autor de terceros, por lo tanto la obra es original, y que es el titular de los derechos patrimoniales, por lo que asume la responsabilidad sobre el contenido de la misma y saldrá en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, a los 27 días del mes de febrero de 2025

A handwritten signature in blue ink, appearing to read "Robert Michael Alvarez Carasayo".

Robert Michael Alvarez Carasayo

C.I.: 1752863629



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS
CERTIFICACIÓN

En calidad de director del trabajo de grado “ SISTEMA INTELIGENTE PARA PREDICCIÓN DE VARIABLES CLIMÁTICAS LOCALES ”, presentado por el egresado Robert Michael Alvarez Carasayo, que opta por el título de Ingeniero en Mecatrónica, certifico que el mencionado proyecto fue realizado bajo mi dirección.

Ibarra, 28 de Febrero de 2025


Carlos Xavier Rosero Chandi
DIRECTOR DE TESIS

Agradecimientos

Quiero expresar mi más profundo agradecimiento a mis padres, Rocío y Roberto, quienes fueron un apoyo gigantesco a lo largo de todo este camino universitario. Gracias a su esfuerzo, sacrificio y amor incondicional, hoy puedo alcanzar el título de ingeniero. Este logro también es suyo, pues sin su respaldo nada de esto habría sido posible. Extiendo también mi gratitud a toda mi familia, cuyo apoyo ha sido fundamental en mi vida.

A mi tutor, Ph.D. Xavier Rosero, una persona excepcional, formidable y de una inteligencia admirable, le agradezco profundamente por haberme guiado tanto como profesor como tutor. Su orientación fue clave en mi formación académica y profesional. Asimismo, expreso mi gratitud a mi asesor de tesis, cuya amabilidad y disposición fueron fundamentales para ayudarme y guiarme en este proceso.

A mis amistades y conocidos, gracias por llenarme de alegría durante todo este tiempo. En especial, a aquellos que no solo fueron amigos, sino hermanos: mi grupo de siempre, los Ratos”. Con ustedes viví experiencias inolvidables que me forjaron como persona y enriquecieron mi vida de innumerables maneras. Su compañía fue invaluable, especialmente cuando estuve lejos de casa, y siempre tendrán un espacio en mi corazón.

No puedo dejar de mencionar a la señora de los .Agachaditos”, cuyo cariño y sazón llenaban de alegría mis días y mi estómago, convirtiéndose en un pequeño pero significativo refugio en esta etapa de mi vida.

Por último, pero no menos importante, quiero agradecer a mi novia, quien con su gran apoyo estuvo a mi lado en todo momento. Su paciencia, comprensión y motivación fueron clave para ayudarme a alcanzar este gran objetivo.

A todos ustedes, gracias de corazón.

Dedicatoria

Este trabajo, fruto de incontables horas de esfuerzo y dedicación, lo dedico con el corazón lleno de gratitud a mis padres, Rocío y Roberto, quienes han sido mi faro en la oscuridad y mi refugio en la tormenta. Su amor inquebrantable, sus sacrificios y su guía han forjado en mí la determinación para alcanzar este logro. Sin su apoyo, este camino habría sido imposible.

A mi familia, cuyo aliento y confianza en mí han sido un sostén invaluable en cada etapa de este recorrido. Su presencia ha sido el impulso que me ha mantenido firme ante los desafíos.

A Dios, quien me ha acompañado en cada paso, dándome fuerza en los momentos más difíciles y sosteniéndome cuando sentí que podía caer. Su luz ha iluminado mi camino y me ha permitido perseverar.

A mi mentor, Xavier, un ejemplo de inteligencia y sabiduría, cuyas enseñanzas y orientación han dejado una huella imborrable en mi formación.

Al Ing. David, un ser excepcional cuya guía y confianza en mis capacidades fueron fundamentales en este proceso. Su apoyo significó mucho más de lo que las palabras pueden expresar.

A toda mi familia, amigos y conocidos que, con su cariño y apoyo, me ayudaron a crecer, a superar obstáculos y a convertirme en la persona que soy hoy. Cada palabra de aliento, cada gesto de apoyo, ha sido un ladrillo en la construcción de este sueño.

Y, finalmente, me dedico este logro a mí mismo, por no rendirme, por cada noche de desvelo, por cada reto superado y por la pasión que puse en este camino. Porque este esfuerzo ha sido grande, pero el aprendizaje y la satisfacción de haber llegado hasta aquí son aún mayores.

Gracias a todos, y gracias a mí por haberlo logrado.

Resumen

Esta tesis diseña y analiza un sistema para la predicción de variables climáticas, identificando los factores esenciales para el pronóstico del clima y la influencia de los microclimas en actividades económicas dependientes de las condiciones meteorológicas. Se implementa un sistema basado en la comparación de sensores, microcontroladores y microcomputadoras, priorizando la optimización de costos para garantizar su accesibilidad. Además, se desarrolla una base de datos robusta mediante la integración de información histórica y datos obtenidos tras la implementación del sistema, destacando la importancia del equilibrio, preprocesamiento y depuración para mejorar la confiabilidad de la información. Se entrenan y comparan cinco modelos de aprendizaje automático para clasificación, evaluándolos con bases de datos de distintas dimensiones y etiquetas para seleccionar el algoritmo más eficiente. Finalmente, se diseña una interfaz gráfica que visualiza los valores recopilados por los sensores, las predicciones generadas y la probabilidad asociada a cada pronóstico.

Abstract

This thesis designs and analyzes a system for predicting climatic variables, identifying key factors for weather forecasting and the influence of microclimates on economic activities dependent on meteorological conditions. The system is implemented based on a comparison of sensors, microcontrollers, and microcomputers, prioritizing cost optimization to ensure accessibility. Additionally, a robust database is developed by integrating historical information with data collected after system implementation, emphasizing the importance of balancing, preprocessing, and cleaning data to enhance reliability. Five machine learning classification models are trained and compared using databases of different dimensions and labels to identify the most efficient algorithm. Finally, a graphical interface is designed to display sensor-collected values, generated predictions, and the probability associated with each forecast.

Keywords: Weather forecasting, Microclimates, Machine Learning, Climatic variables, Data preprocessing, Open-source hardware.

Índice general

Autorización de uso y publicación a favor de la Universidad Técnica del Norte	II
Agradecimientos	V
Dedicatoria	VI
Resumen	VII
Abstract	VIII
1 Introducción	1
1.1 Planteamiento del problema	1
1.2 Objetivos	2
1.2.1 Objetivo general	2
1.2.2 Objetivos específicos	2
1.3 Alcance	2
1.4 Justificación	3
1.5 Contexto	4

2	Análisis literario	6
2.1	Estado del arte	6
2.2	Marco teórico	8
2.2.1	Meteorología	8
2.2.2	Estaciones meteorológicas	8
2.2.3	Efecto invernadero	8
2.2.4	Gases de efecto invernadero	9
2.2.5	Cambio climático en Ecuador y su capital	10
2.2.6	Predicción de variables climáticas	11
2.2.7	Modelos climáticos	11
2.2.7.1	Modelo ECHAM4.5	12
2.2.7.2	Modelo CCM3.6	13
2.2.7.3	Modelo de balance de energía (EBMs)	13
2.2.7.4	Modelos radiativos-convectivos (RCMs)	13
2.2.7.5	Modelo de restricciones dimensionales	14
2.2.8	Declaración de variables atmosféricas	14
2.2.9	Firebase	14
2.2.10	Aprendizaje automático	15
2.2.10.1	Aprendizaje supervisado	15
2.2.10.2	Aprendizaje no supervisado	16
2.2.11	Técnicas de aprendizaje utilizadas para predecir el clima.	16

3	Desarrollo	18
3.1	Metodología	18
3.2	Diseño del sistema de predicción del clima	19
3.2.1	Selección de los sensores	19
3.2.2	Selección de microcontrolador	25
3.2.3	Selección del microprocesador	27
3.2.4	Selección de la base de datos en tiempo real	29
3.3	Implementación del sistema de predicción del clima	31
3.3.1	Partes usadas para la construcción del hardware del sistema	31
3.3.2	Construcción y optimización del hardware y software del sistema	32
3.3.3	Implementación del código para el funcionamiento del circuito	33
3.3.4	Comunicación del puerto serial con el microcomputador	35
3.3.5	Comunicación entre el microcomputador y la base de datos en tiempo real	36
3.3.6	Creación de la base de datos	39
3.4	Descripción de los modelos de machine learning utilizados.	44
3.4.1	Árbol de Decisión	44
3.4.2	K-Nearest Neighbors	44
3.4.3	Máquinas de Vectores de Soporte	45
3.4.4	Bosque Aleatorio	46
3.4.5	Interfaz gráfica para visualización de los resultados de la predicción. . .	46
4	Pruebas de funcionamiento	48

4.1	Descripción de las pruebas	48
4.1.1	Métricas de validación	49
4.2	Análisis de resultados	49
4.2.1	Análisis de los modelos con 3 dimensiones y 4 etiquetas	49
4.2.2	Análisis de los modelos con 2 dimensiones y 4 etiquetas	51
4.2.3	Análisis de los modelos con 3 dimensiones y 2 etiquetas	52
4.3	Análisis comparativo de algoritmos para la selección del modelo final	53
4.3.1	Resultados y análisis	54
4.3.1.1	Comparación de resultados con 3 dimensiones y 4 etiquetas .	54
4.3.1.2	Comparación de resultados con 2 dimensiones y 4 etiquetas .	54
4.3.1.3	Comparación de resultados con 2 dimensiones y 2 etiquetas .	55
4.3.2	Propuesta de un modelo optimizado para mayor eficiencia	56
5	Conclusiones y trabajo futuro	57
5.1	Conclusiones	57
5.2	Recomendaciones	58
5.3	Trabajo futuro	59
	Anexos	69
A	Repositorio del código de Arduino	70
B	Repositorio de código Python para conexión serial y almacenamiento en tiempo real.	71

C	Conjunto de base de datos disponibles	72
D	Algoritmos utilizados para el entrenamiento de modelos de Machine Learning.	73
E	Interfaz gráfica para la visualización de resultados.	74

Índice de tablas

3.1	Comparativa de sensores de temperatura y humedad	20
3.2	Ventajas y desventajas de los sensores de temperatura y humedad	21
3.3	Comparativa de sensores de presión atmosférica	22
3.4	Ventajas de los sensores de presión atmosférica	23
3.5	Comparativa de sensores de lluvia	24
3.6	Comparativa de microcontroladores	26
3.7	Ventajas comparativas de microcontroladores	26
3.8	Comparativa de microprocesadores	28
3.9	Ventajas comparativas de microprocesadores	28
3.10	Comparativa de bases de datos en tiempo real	30
3.11	Clasificación de la lluvia según su intensidad	40
3.12	Etiquetas para la base de datos creada	41
4.1	Métricas para evaluar los modelos.	49
4.2	Exactitud de los diferentes modelos 3d 4e	50
4.3	Precisión de los diferentes modelos 3d 4e	50
4.4	Sensibilidad de los diferentes modelos 3d 4e	50

4.5	Exactitud de los diferentes modelos 2d 4e	51
4.6	Precisión de los diferentes modelos 2d 4e	51
4.7	Sensibilidad de los diferentes modelos 2d 4e	52
4.8	Exactitud de los diferentes modelos 3d 2e	52
4.9	Precisión de los diferentes modelos 3d 2e	53
4.10	Sensibilidad de los diferentes modelos 3d 2e	53
4.11	Exactitud de los diferentes modelos 3d 4e resultados	54
4.12	Exactitud de los diferentes modelos 2d 4e resultados	55
4.13	Exactitud de los diferentes modelos 3d 2e resultados	55
4.14	Precisión de los diferentes modelos 3d 2e resultados	56
4.15	Sensibilidad de los diferentes modelos 3d 2e resultados	56

Índice de figuras

3.1	Modelo en Cascada.	19
3.2	Sensor DHT22	21
3.3	Sensor BMP180	23
3.4	Sensor DSK raindrops	25
3.5	Microcontrolador ESP32	27
3.6	Miprocador Raspberri Pi 3b+	29
3.7	Secuencia de funcionamiento del sistema	32
3.8	Circuito para la recepción de datos	33
3.9	Diagrama de bloques del sistema de recepción de datos	34
3.10	Conexión entre la Raspberri Pi y la ESP32 mediante SSH	36
3.11	Diagrama de comunicación de la Raspberry Pi con Firebase	38
3.12	Datos añadidos a la base de datos creada	39
3.13	Total de datos recopilados y almacenados en la base de datos local.	41
3.14	Base de datos con datos balanceados	42
3.15	Base de datos con datos balanceados	43
3.16	Interfaz para visualización de los resultados de la predicción	47

Capítulo 1

Introducción

1.1. Planteamiento del problema

El cambio en el clima se atribuye de forma directa e indirecta a la actividad humana ya que altera la composición del ambiente. Este cambio conlleva a que se excedan los rangos de variabilidad de la temperatura y de las precipitaciones. Cabe mencionar que el clima de un determinado territorio es el resultado de varios factores e interacciones [1].

Hoy en día existen aplicaciones móviles y web que proporcionan la predicción del clima. Sin embargo, estas predicciones se realizan en función de datos obtenidos a través de radares meteorológicos que cubren grandes áreas. La predicción a nivel local en países con microclimas es poco fiable [2].

Mediante el sensado de variables atmosféricas basado en hardware abierto y de bajo costo y a través de un sistema inteligente de gestión y análisis de esta información, se podría predecir el tiempo [2]. Así, se pretende desarrollar una aplicación de pronóstico del tiempo confiable, eficiente y de bajo costo utilizando el concepto de aprendizaje automático en software libre y hardware abierto.

A través del desarrollo de esta herramienta se podría realizar la predicción de variables

climáticas en un entorno local, con microclimas, servicio no proporcionado por las aplicaciones existentes.

1.2. Objetivos

1.2.1. Objetivo general

Diseñar un sistema inteligente para la predicción de variables climáticas locales.

1.2.2. Objetivos específicos

Diseñar el modelo para la predicción de variables climáticas considerando las características.

Implementar un sistema prototipo inteligente que considere la interacción entre los dispositivos finales y las aplicaciones a través de una plataforma de internet de las cosas.

Integrar el modelo sobre el sistema para la evaluación de varios algoritmos de aprendizaje de máquina.

1.3. Alcance

El presente proyecto consiste en un sistema inteligente que reciba la información de variables atmosféricas y la usa para predecir variables climáticas locales.

Se trabajará dentro de una plataforma de internet de las cosas, tomando información desde

sensores y almacenándola en una base de datos local. Esta información estará disponible para que pueda ser accedida local o remotamente.

Se desarrollará hardware de prueba de concepto para implementar el modelo de sensado y almacenamiento de datos diseñado. Además, usando esta información, se predecirá por lo menos una variable climática que podría ser la precipitación, a través de aprendizaje de máquina supervisado, la predicción será presentada mediante un programa tipo ventana en el ordenador, se usará hardware abierto y software matemático libre para el desarrollo de todo el proyecto.

1.4. Justificación

Actualmente, con el avance de la tecnología se han creado diferentes sistemas para la predicción de variables climáticas que no se enfocan en sectores específicos, sino en sectores amplios ciudades y cantones.

El sistema inteligente de predicción del clima en función de variables locales tiene la finalidad de brindar una solución barata y sectorial que puede escalada en función de diversas necesidades. Un valor agregado de la propuesta es que la información guardada en la base de datos será accesible gratuitamente para aplicaciones ingenieriles y académicas.

La inteligencia artificial permite analizar datos automáticamente sin importar el tamaño de la muestra. Se pueden incluso utilizar algoritmos capaces de aprender a predecir el clima basándose en datos históricos de una zona [3]. Este es un valor agregado del sistema, convertirse en un dispositivo de monitoreo y almacenamiento que permita guardar grandes volúmenes de información climática para su posterior análisis.

1.5. Contexto

Los factores que influyen de manera significativa en este procedimiento son los meteorológicos, ya que son usados para determinar el estado del clima. Entre ellos se encuentra la temperatura, precipitación, velocidad del viento, nubosidad, niebla y humedad relativa. Cada uno de ellos aporta información importante para la predicción de variables [4].

Las variables de precipitación y temperatura resaltan cambios relevantes en los datos de variables climatológicas. Sin embargo, existe una falta de registro en meses, incluso hasta años, de esta información [5].

El registro histórico de los cambios climáticos se realiza a través de instrumentos de medida que han evolucionado hasta la aplicación de técnicas actuales no convencionales. De allí nacen los dispositivos informativos que traen consigo nuevas formas de recopilar información climática. Entre estas destacan los drones, nuevas tecnologías de sensores, radares, entre otros [6].

Los sensores de temperatura del aire, luminosidad, velocidad del viento y humedad, generalmente presentan resultados inconsistentes en el momento de recolectar información. Es por ello que se deben realizar mantenimientos mensuales a los instrumentos de medida ya que esto servirá para recolectar información de manera eficiente [7].

Por otra parte, el internet de las cosas (IoT, Internet of Things) consiste en objetos de uso cotidiano que se conectan a internet de manera automática (sin la interacción humana) (Pizarro, 2020). Es así como la internet permite la interconexión entre los objetos del mundo físico, equipados con sensores, actuadores y tecnología de comunicación [8].

Diseñan una solución portátil y de bajo costo para la predicción del tiempo. Desarrollan un sistema de pronóstico del tiempo que puede usarse en áreas remotas. El método de análisis de datos y aprendizaje automático llamado “random forest” se utiliza para predecir las condiciones meteorológicas [2].

Exploran tres modelos de aprendizaje automático para la predicción meteorológica: Support Vector Machine (SVM), Artificial Neural Network (ANN) y Time Series Recurrent Neural Network (RNN) [2]. También discuten los pasos seguidos para lograr los resultados. Para el pronóstico del tiempo usan Pandas, NumPy, Keras, Git, Matplotlib, TensorFlow, Anaconda y Google Cloud Services. Observan que RNN basado en series temporales hace el mejor trabajo de predicción del tiempo.

Capítulo 2

Análisis literario

Este capítulo presenta estudios relacionados con la predicción de variables climáticas. El estado del arte proporciona información actualizada en áreas como la ciencia de datos y la agricultura de precisión, destacando el uso de hardware libre y plataformas gratuitas, especialmente bases de datos en tiempo real como Firebase.

2.1. Estado del arte

El estudio Wireless Wine: Estimación de rendimiento y ubicación de sensores para la predicción de heladas en los viñedos destaca la influencia del cambio climático en la ocurrencia de heladas, afectando la producción vitivinícola en Mendoza, donde en 2016 se registró una reducción del 40 % respecto a 2015 [9]. Se propone el uso de sensores IoT para optimizar la previsión del rendimiento agrícola mediante la recopilación de datos meteorológicos en los viñedos, complementados con información histórica y en tiempo real, lo que permite anticipar condiciones

climáticas adversas y facilitar la toma de decisiones.

En 2023 se presenta el estudio “Análisis y Predicción de Radiación en Sistemas Fotovoltaicos Haciendo Uso de Machine Learning” [10], centrado en el desarrollo de un modelo para la predicción de la radiación solar. Este enfoque permite anticipar los niveles de radiación en un área específica, facilitando la planificación y ejecución de proyectos fotovoltaicos para la integración a la red eléctrica nacional. El estudio enfatiza el procesamiento y limpieza de datos, así como la optimización de redes neuronales LSTM para la predicción de variables climáticas esenciales en la toma de decisiones.

“El Estudio de factibilidad del uso de Machine Learning con múltiples fuentes de datos en el pronóstico del tiempo” [11] analiza el impacto de los eventos climáticos extremos en la economía, como huracanes e inundaciones en zonas turísticas. Propone un modelo de recopilación de datos mediante una miniestación meteorológica basada en Arduino, que combina mediciones en tiempo real con datos históricos. Esta integración facilita la aplicación de técnicas y algoritmos de Machine Learning para la predicción climática, empleando variables como humedad, presión y temperatura.

En 2023 se presenta el estudio “Cuantificación y Predicción de Lluvia a partir de Información Obtenida de Radioenlaces de Microondas y Comunicaciones Satelitales Aplicando Machine Learning” [12], el cual analiza los riesgos asociados a eventos meteorológicos que afectan a la población. Destaca la recopilación de datos mediante sensores y la implementación de un sistema de detección de precipitaciones atmosféricas, logrando una mejora del 5,88 % en métricas de calidad respecto a otros modelos. Además, se entrenan redes neuronales en MATLAB

utilizando la función `trainNetwork` con parámetros como `sequence`, `response`, `layers` y `options`. El estudio subraya la diversidad de enfoques para la predicción climática, considerando tanto software con licencia como herramientas de código abierto.

2.2. Marco teórico

2.2.1. Meteorología

La meteorología estudia los fenómenos atmosféricos y sus procesos mecánicos con el objetivo de realizar predicciones. El término proviene del griego *meteoros* (alto) y *logos* (tratado). Esta disciplina clasifica los fenómenos atmosféricos en categorías como aéreos, acuosos, granizo, auroras polares, arcoíris y rayos [13].

2.2.2. Estaciones meteorológicas

Las estaciones meteorológicas son puntos de observación ubicados en sitios estratégicos para la recopilación de datos atmosféricos con fines de análisis en tiempo real y diferido. Estos espacios deben cumplir con estándares y técnicas normalizadas que garanticen la precisión y confiabilidad de las mediciones [14].

2.2.3. Efecto invernadero

El efecto invernadero es un fenómeno natural causado por la presencia de ciertos gases en la atmósfera, los cuales retienen parte de la radiación solar que intenta escapar al espacio. Este

proceso devuelve la radiación a la litósfera, generando un aumento en la temperatura ambiental. Se trata de un mecanismo esencial para la regulación térmica de cualquier planeta con atmósfera [15].

2.2.4. Gases de efecto invernadero

Los gases acumulados en la atmósfera por causas naturales o antropogénicas se denominan gases de efecto invernadero. En los últimos años, las actividades humanas han incrementado significativamente su concentración, intensificando el efecto invernadero y elevando la temperatura global [15].

Las principales consecuencias del cambio climático acelerado incluyen:

- Daños en la biodiversidad.
- Incendios.
- Deshielo.
- Subida del nivel del mar.
- Desertización.
- Aumento de la temperatura.
- Fenómenos meteorológicos extremos [15].

2.2.5. Cambio climático en Ecuador y su capital

La evidencia del cambio climático y su impacto global es contundente, reflejada en transformaciones visibles como el aumento significativo del nivel del mar.

El derretimiento de glaciares y capas polares, junto con el aumento de la temperatura global, provoca fenómenos extremos como olas de calor y frío. Estos cambios constituyen una evidencia clara del cambio climático.

La segunda comunicación nacional sobre el cambio climático identifica varias influencias probables en Ecuador, entre ellas:

1. Intensificación de eventos climáticos extremos, como los asociados a la Oscilación del Sur-El Niño;
2. Retroceso de glaciares;
3. Reducción del drenaje anual;
4. Aumento en la transmisión de enfermedades tropicales;
5. Expansión de especies invasoras en Galápagos y otros ecosistemas sensibles; y
6. Pérdida de biodiversidad.

Este último aspecto es especialmente crítico, dado el alto valor del patrimonio genético del país, cuya conservación es fundamental no solo a nivel nacional, sino también global. Estos análisis subrayan la necesidad de adoptar medidas alineadas con las disposiciones constitucionales del Ecuador [16].

2.2.6. Predicción de variables climáticas

La predicción climática se basa en una amplia variedad de modelos desarrollados por distintos grupos e instituciones, los cuales emplean variables atmosféricas como temperatura, presión, humedad relativa y precipitación para su entrenamiento. La implementación de estos métodos requiere herramientas adecuadas que optimicen el uso del modelado numérico, con el objetivo de ofrecer soluciones teórico-prácticas para futuras proyecciones climáticas [17].

El estudio de las características climáticas de una región es fundamental para anticipar fenómenos inesperados y mitigar sus efectos. Comprender y predecir estas variaciones permite tomar decisiones informadas a nivel nacional, optimizando el uso de conocimientos, técnicas y herramientas para la gestión climática.

2.2.7. Modelos climáticos

Para comprender los fenómenos físicos de la atmósfera, se han desarrollado modelos que estudian las variables climáticas y su comportamiento. Los primeros intentos se remontan al siglo III a. C., cuando Eratóstenes relacionó el clima con la posición del sol. Posteriormente, Edmond Halley formuló la primera teoría climática, publicada después de la Segunda Guerra Mundial, basada en datos de la corriente en chorro [17].

En la década de 1950, Norman Phillips desarrolló el primer modelo computarizado de circulación global. Aunque limitado por la capacidad informática de la época, empleó placas cilíndricas para simplificar la geometría. A pesar de estas restricciones, el modelo logró simular el movimiento atmosférico a escala mensual y a largo plazo, sentando las bases para el desarrollo de

los modelos de circulación general [17].

A partir de la década de 1960, surgieron nuevos modelos climáticos más avanzados. En 1969 se describió el modelo de balance energético (EBM), seguido en 1970 por el modelo estadístico dinámico (SDM) y, posteriormente, el modelo de complejidad intermedia (EMIC). Estos últimos han experimentado importantes avances en las últimas décadas.

Los modelos climáticos simulan los procesos atmosféricos de la Tierra mediante ecuaciones diferenciales basadas en la física, la mecánica de fluidos y la química. Desde la década de 1970, el aumento de la capacidad informática ha permitido incorporar un mayor número de variables, agrupadas en procesos radiativos, cinéticos, superficiales y químicos.

El componente radiativo abarca la entrada de radiación solar, su absorción en la atmósfera y los océanos, y la emisión de radiación infrarroja. En cuanto a la dinámica atmosférica, la energía se transporta a través del viento y las corrientes oceánicas. Los procesos superficiales involucran cambios en los océanos, el hielo, la vegetación y la humedad [17].

2.2.7.1. Modelo ECHAM4.5

El modelo de circulación general *ECHAM4.5* es el cuarto de una serie desarrollada en el *Max Planck Institut für Meteorologie* (MPIM) y deriva del modelo del *European Centre for Medium-Range Weather Forecasts* (ECMWF).

Con base en los trabajos de Phillips y Lohman & Benmartz, este modelo emplea como variables la vorticidad, la divergencia, el logaritmo de la presión en la superficie, la temperatura y las razones de mezcla del vapor de agua, el agua líquida en las nubes y el hielo. Las ecuaciones

se resuelven en 19 niveles verticales mediante un sistema de coordenadas híbridas, utilizando el método espectral con truncación triangular en el número de onda 42.

2.2.7.2. Modelo CCM3.6

Es un modelo general desarrollado por el *National Center for Atmospheric Research* (NCAR). Siguiendo a Rasch & Kristjánsson, el *CCM3.6* se caracteriza por el uso de ecuaciones de evolución para el calor y el momento, empleando una representación espectral para el tratamiento horizontal y diferencias finitas de primer orden para gradientes verticales [18].

2.2.7.3. Modelo de balance de energía (EBMs)

Este modelo evalúa el cambio en el sistema climático a partir del ingreso de energía al planeta. Se centra en la energía termodinámica, ya que no altera directamente el sistema climático. No proporciona una dimensión espacial concreta, sino que refleja un promedio global. Generalmente, se emplea para el análisis de fenómenos externos y la interpretación de resultados complejos [18].

2.2.7.4. Modelos radiativos-convectivos (RCMs)

Estos modelos pueden representarse en una o dos dimensiones. Los unidimensionales simulan la transferencia de energía y las variaciones de radiación a lo largo de la altitud atmosférica, mientras que los bidimensionales incorporan tanto la dimensión vertical como la horizontal. Finalmente, los *RCMs* determinan los efectos de las variaciones en los gases de efecto invernadero sobre la radiación terrestre (emisividad), influyendo en la temperatura [18].

2.2.7.5. Modelo de restricciones dimensionales

Estos modelos incorporan dos o más dimensiones latitudinales. Un conjunto de estadísticas resume la velocidad y dirección de los vientos, mientras que un coeficiente de difusión regula el transporte de energía. Denominados *Estadístico Dinámico (SDMs)*, se emplean para analizar las corrientes horizontales y los procesos que las afectan [18].

2.2.8. Declaración de variables atmosféricas

Gracias al modelo *ARIMA* propuesto en “Predicción de variables meteorológicas por medio de modelos *ARIMA*”, se identifican variables atmosféricas esenciales para el monitoreo y la predicción climática. Este modelo emplea datos almacenados a nivel horario en estaciones meteorológicas, incluyendo fecha y hora, velocidad del viento, precipitación, radiación solar, temperatura media y humedad relativa [19].

El proceso de predicción del clima se define como pronóstico del tiempo. Mediante la adquisición de datos en tiempo real, es posible prever la lluvia utilizando variables atmosféricas como humedad, presión y temperatura, las cuales se obtienen a través de sensores específicos [2].

2.2.9. Firebase

Es una base de datos en tiempo real, utilizada en aplicaciones móviles y en la nube para gestionar información de manera eficiente. Su velocidad supera a *HTTP*, ya que opera mediante una única conexión *socket*. Funciona como una base de datos de *back-end* estructurada en tiem-

po real, proporcionando a los desarrolladores una API para sincronizar y almacenar datos en la nube [20].

2.2.10. Aprendizaje automático

Se define como una rama de la inteligencia artificial enfocada en replicar las capacidades de aprendizaje humano mediante algoritmos. Un algoritmo se considera inteligente si aprende y se adapta a los cambios del entorno, así como a nuevas situaciones de forma automática. El proceso de aprendizaje optimiza los parámetros del modelo al minimizar un criterio de error [21].

También conocido como *machine learning*, es una rama de la inteligencia artificial que permite a las máquinas aprender mediante algoritmos que procesan muestras de datos sin requerir un modelo matemático preprogramado [22].

El *machine learning* es una herramienta robusta aplicable a diversos ámbitos. Sin embargo, la variedad de modelos desarrollados requiere una selección adecuada según las necesidades específicas y los datos disponibles para cada aplicación.

Existen dos tipos de *machine learning*: supervisado y no supervisado.

2.2.10.1. Aprendizaje supervisado

En este tipo de algoritmo, la intervención humana es fundamental, ya que el aprendizaje depende de datos introducidos, clasificados y etiquetados por el usuario. Estos datos pueden emplearse para tareas de clasificación, donde los objetos se asignan a distintas categorías, o de

regresión, cuando se busca predecir un valor numérico [22].

2.2.10.2. Aprendizaje no supervisado

En este tipo de algoritmo, no se requiere intervención humana, ya que el aprendizaje se basa en la identificación de patrones a partir de datos no etiquetados. Existen dos enfoques principales: *clustering*, donde se agrupan los datos según similitudes, y *asociación*, que permite identificar reglas y relaciones dentro del conjunto de datos [22].

2.2.11. Técnicas de aprendizaje utilizadas para predecir el clima.

En la predicción climática, diversas variables pueden analizarse, como la velocidad del viento, cuya estimación resulta compleja debido a la variedad de modelos disponibles. Para su predicción, se emplean diferentes enfoques de *machine learning*, como redes neuronales, vecinos cercanos, regresión lineal y árboles aleatorios. Además, es posible combinar estos modelos para mejorar la precisión de los resultados [23].

Uno de los métodos más relevantes es *k-nearest neighbors (K-NN)*, un algoritmo de clasificación y predicción sencillo pero eficaz. Su enfoque no se basa en determinar un número fijo de vecinos, sino en considerar aquellos más cercanos según el valor de K seleccionado. La distancia euclidiana es la métrica más utilizada para esta aproximación, aunque también pueden emplearse otras como Chebyshev, Manhattan y Mahalanobis. Estas características hacen de *K-NN* un algoritmo idóneo para la predicción climática [23].

Algoritmos como la regresión lineal no son adecuados para la predicción climática, ya que

su enfoque se basa en modelar la relación entre una variable de interés y un conjunto de variables predictoras mediante una ecuación lineal. Sin embargo, el clima es altamente variable y no sigue un comportamiento lineal, lo que limita la precisión de este modelo en dicho contexto.

Capítulo 3

Desarrollo

Este capítulo describe el proceso de selección de variables, la implementación del sistema o servidor local para la recopilación de datos ambientales, el almacenamiento en la nube y la gestión de la información. Finalmente, se presenta el diseño de un sistema capaz de predecir, a corto plazo, al menos una variable climática mediante un algoritmo de *machine learning*.

3.1. Metodología

El modelo en cascada, también conocido como secuencial, se emplea en proyectos experimentales como el sistema de predicción climática. Este enfoque divide el desarrollo en fases estructuradas, estableciendo objetivos específicos en cada etapa y permitiendo un avance ordenado del proyecto.

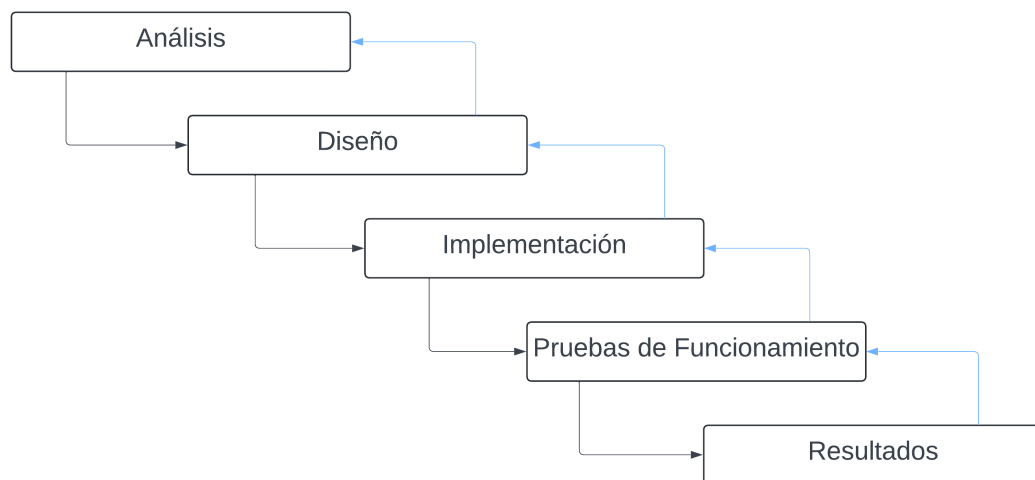


Figura 3.1: Modelo en Cascada.

- **Análisis de variables atmosféricas:** Se seleccionaron tres variables atmosféricas dependientes y una independiente. El estudio se basó en un artículo que considera la humedad relativa, la temperatura ambiental y la presión atmosférica como variables clave para el análisis y la predicción climática. Estas variables son fundamentales para la precisión del modelo. Como variable a predecir se eligió la lluvia, dado que puede ser etiquetada según su ocurrencia [2].

3.2. Diseño del sistema de predicción del clima

3.2.1. Selección de los sensores

En la selección de sensores se consideraron la precisión, el costo, el rango de medición y otras características diferenciadoras. Se priorizó la optimización de gastos y la versatilidad del sistema para su aplicación en diversos campos.

Las primeras variables seleccionadas para el sensor son la temperatura y la humedad relativa, las cuales pueden medirse con distintos dispositivos. En la Tabla 3.1 se presentan las diferencias entre ellos para facilitar su selección.

Tabla 3.1: Comparativa de sensores de temperatura y humedad

Sensor	Rango de temp.	Precisión de temp.	Rango de hum.	Precisión de hum.
DHT11	0°C a 50°C	$\pm 2^{\circ}\text{C}$	20 % a 80 % RH	$\pm 5\%$ RH
DHT22	-40°C a 125°C	$\pm 0.5^{\circ}\text{C}$	0 % a 100 % RH	$\pm 2-5\%$ RH
BME280	-40°C a 85°C	$\pm 1^{\circ}\text{C}$	0 % a 100 % RH	$\pm 3\%$ RH
AHT20	-40°C a 85°C	$\pm 0.3^{\circ}\text{C}$	0 % a 100 % RH	$\pm 2\%$ RH

La información se extrae de [24, 25], donde se realiza una comparativa de distintos sensores. Se destaca que el sensor DHT22 es uno de los más confiables para este sistema debido a su amplio rango de temperatura, lo que permite su adaptación a diversos entornos sin inconvenientes. Además, presenta una alta precisión en la medición y una tolerancia mínima en comparación con otros sensores analizados. La Tabla 3.2 detalla sus ventajas y desventajas.

Tabla 3.2: Ventajas y desventajas de los sensores de temperatura y humedad

Sensor	Ventajas	Desventajas
DHT11	Económico, fácil de usar	Menor precisión, rango de medición limitado
DHT22	Alta precisión, amplio rango de medición	Más caro que el DHT11, comunicación más lenta
BME280	Alta precisión, mide presión atmosférica, interfaz I2C	Más caro, más complejo de integrar
AHT20	Alta precisión, interfaz I2C, buena estabilidad en condiciones adversas	Relativamente nuevo, menos documentación disponible

En la Figura 3.2 se puede observar el sensor usado [26].

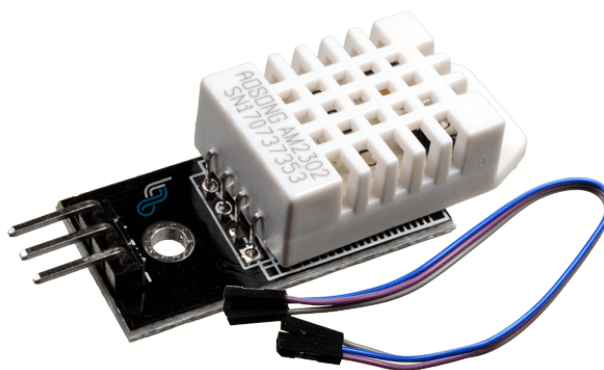


Figura 3.2: Sensor DHT22

La siguiente variable considerada para la selección del sensor fue la presión atmosférica.

Los criterios de elección incluyeron precisión, amplitud de medición y fiabilidad. En la Tabla 3.3 se presenta una comparativa de los sensores evaluados.

Tabla 3.3: Comparativa de sensores de presión atmosférica

Sensor	Rango de Presión	Precisión	Resolución
BMP180	300 a 1100 hPa	± 0.12 hPa	0.01 hPa
BMP280	300 a 1100 hPa	± 1 hPa	0.16 hPa
BME280	300 a 1100 hPa	± 1 hPa	0.18 hPa
MS5611	10 a 1200 hPa	± 0.03 hPa	0.01 hPa
LPS22HB	260 a 1260 hPa	± 0.1 hPa	0.01 hPa

La información se extrae de [27, 28, 29, 30, 31], donde se comparan distintas características de los sensores evaluados. Se selecciona el sensor BMP180 por su alta precisión y bajo consumo de energía, características ideales para la predicción climática en diversas condiciones. La Tabla 3.4 presenta sus ventajas.

Tabla 3.4: Ventajas de los sensores de presión atmosférica

Sensor	Ventajas
BMP180	Alta precisión, bajo consumo de energía, compacto.
BMP280	Mejor precisión que el BMP180, también mide temperatura y humedad.
BME280	Alta precisión, mide temperatura y humedad, interfaz I2C y SPI.
MS5611	Muy alta precisión, amplio rango de presión.
LPS22HB	Alta precisión, bajo consumo de energía, compacto.

El sensor seleccionado se observa en la Figura 3.3 [32].

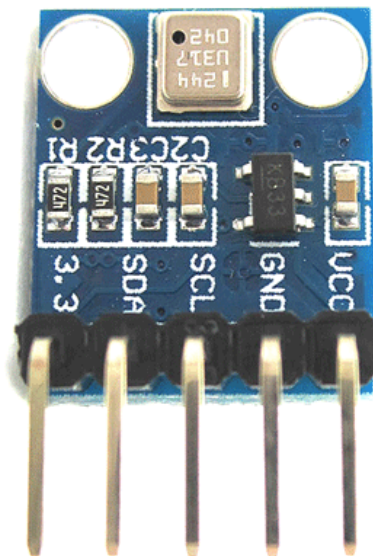


Figura 3.3: Sensor BMP180

Finalmente, se selecciona un sensor para detectar la presencia de lluvia y validar su ocu-

rencia en el sistema. La Tabla 3.5 presenta una comparativa de los sensores evaluados para esta función, con el objetivo de garantizar una recopilación de datos precisa y confiable.

Tabla 3.5: Comparativa de sensores de lluvia

Sensor	Protocolo de Comunicación	Voltaje de Suministro	Ventajas
DSK Raindrops	Analógico, Digital	3.3V a 5V	Alta sensibilidad, fácil de usar, resistente a la corrosión
YL-83	Analógico, Digital	3.3V a 5V	Alta sensibilidad, incluye módulo comparador
FC-37	Analógico, Digital	3.3V a 5V	Alta sensibilidad, resistente a la corrosión

En esta sección no se identifica una diferencia significativa entre los sensores evaluados [33, 34], ya que presentan un desempeño similar en términos de sensibilidad, durabilidad y precisión. Por esta razón, cualquiera de ellos resulta adecuado para la detección de lluvia. En este caso, se selecciona el sensor DSK Raindrops, representado en la Figura 3.4 [35].

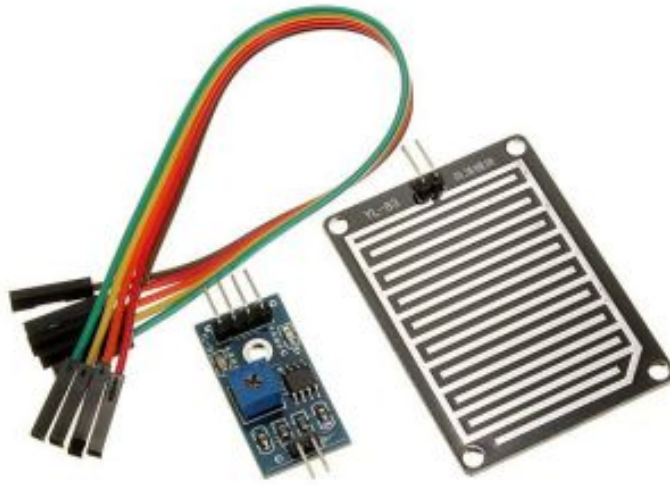


Figura 3.4: Sensor DSK raindrops

3.2.2. Selección de microcontrolador

El uso de un microcontrolador se considera debido a sus múltiples ventajas, entre ellas su eficiencia, escalabilidad y estabilidad [36]. A diferencia de un microprocesador, que cuenta con un hardware más robusto y una mayor cantidad de funcionalidades, un microcontrolador resulta más adecuado para este tipo de aplicaciones, ya que evita posibles limitaciones derivadas del exceso de características innecesarias. La Tabla 3.6 y la Tabla 3.7 presentan las diferencias clave para la selección del microcontrolador más adecuado.

Tabla 3.6: Comparativa de microcontroladores

Microcontrolador	Procesador	Frecuencia	Mem Flash	RAM	Conect.	GPIO
ESP32	Tensilica	Hasta	Hasta 16 MB	520 KB	Wi-Fi,	34
	Xtensa	240 MHz			Bluetooth	
	LX6 (dual-core)					
Arduino Uno	ATmega328P	16 MHz	32 KB	2 KB	Ninguna	14
STM32F103C8T6	ARM Cortex-M3	72 MHz	64 KB	20 KB	Ninguna	37

Tabla 3.7: Ventajas comparativas de microcontroladores

Microcontrolador	Ventajas
ESP32	Alta potencia de procesamiento, conectividad Wi-Fi y Bluetooth integrada, bajo costo
Arduino Uno	Fácil de usar, gran comunidad de soporte, ideal para principiantes
STM32F103C8T6	Alto rendimiento, bajo costo, soporte para periféricos avanzados

La información se extrae de [37, 38]. Como se observa en la comparativa, los tres microcontroladores evaluados resultan adecuados para este sistema. Sin embargo, la ESP32 es la opción

más indicada debido a su alto rendimiento, mayor cantidad de entradas y amplias funcionalidades, lo que permite la integración de más sensores y mejora la precisión del sistema de predicción. La Figura 3.5 presenta este microcontrolador [39].

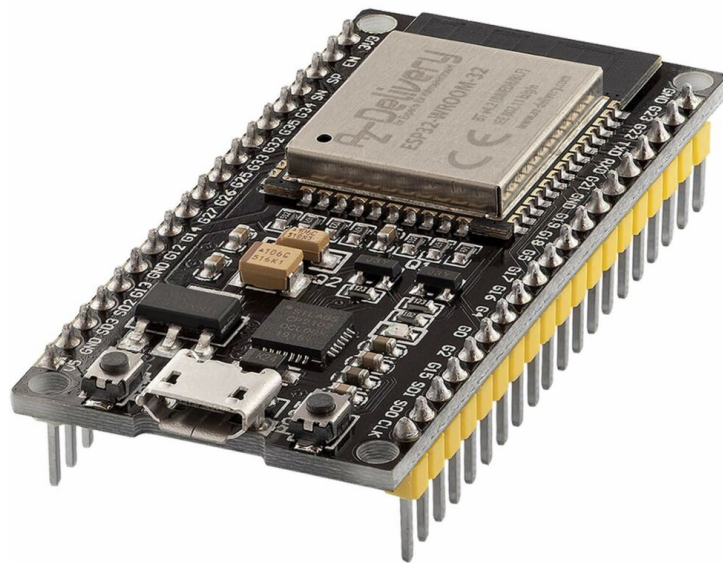


Figura 3.5: Microcontrolador ESP32

3.2.3. Selección del microprocesador

Un microprocesador incorpora una *CPU*, memoria y periféricos de entrada y salida en un solo chip, permitiendo el control eficiente de dispositivos electrónicos y sensores [40].

Esta sección analiza las características de diversos microcontroladores con funciones similares a las de un ordenador. Su propósito es seleccionar el más adecuado en términos de simplicidad y eficiencia. La Tabla 3.6 y la Tabla 3.9 presentan su comparación.

Tabla 3.8: Comparativa de microprocesadores

Microprocesador	Procesador	Frecuencia	Mem RAM	Conectividad	GPIO
Raspberry Pi 3B+	Broadcom BCM2837B0 (quad-core)	1.4 GHz	1 GB	Wi-Fi, Bluetooth, Ethernet	40
Intel Celeron N3350	Intel Celeron N3350 (dual- core)	1.1 GHz (2.4 GHz turbo)	2 GB	Wi-Fi, Ethernet	N/A
ODROID-C4	Amlogic S905X3 (quad-core)	2.0 GHz	4 GB	Wi-Fi (con adap- tador), Ethernet	40

Tabla 3.9: Ventajas comparativas de microprocesadores

Microprocesador	Ventajas
Raspberry Pi 3B+	Bajo costo, fácil de usar, gran comunidad de soporte, conectividad integrada
Intel Celeron N3350	Mayor rendimiento en tareas de procesamiento, soporte para gráficos integrados
ODROID-C4	Alto rendimiento, mayor memoria RAM, soporte para 4K, comunidad activa

La información se extrae de [41, 42]. Tras analizar las características de estos microprocesa-

dores y procesadores, se selecciona el más adecuado para esta aplicación, priorizando simplicidad y eficiencia. Dado que el sistema operativo a emplear es Linux, se aprovechan sus ventajas sobre sistemas más robustos, permitiendo el uso de aplicaciones específicas como la conectividad del equipo, la comunicación con el microcontrolador y el funcionamiento de la distribución seleccionada. Estas tareas requieren un uso mínimo de los recursos del dispositivo, por lo que se elige la *Raspberry Pi 3B+*, representada en la Figura 3.6 [43].

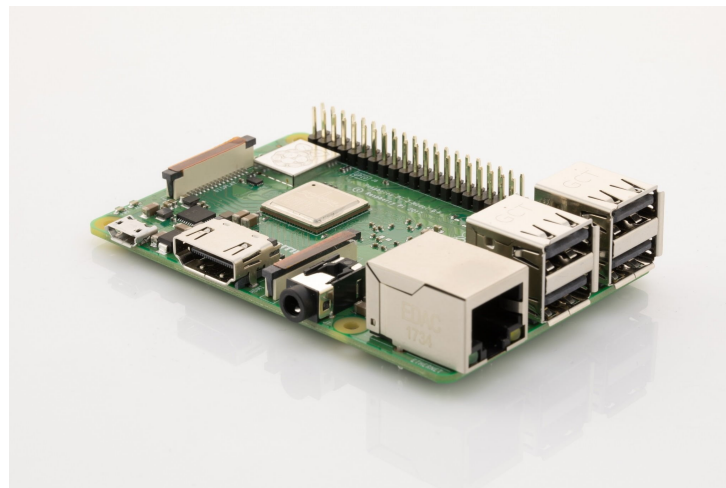


Figura 3.6: Miprosesador Raspberri Pi 3b+

3.2.4. Selección de la base de datos en tiempo real

Las bases de datos en tiempo real son sistemas de gestión de datos diseñados para entornos donde el tiempo es un factor crítico. A diferencia de las bases de datos tradicionales, están optimizadas para garantizar la ejecución de operaciones dentro de plazos estrictos [44].

El diseño de estas bases de datos incorpora técnicas como la planificación de tareas en tiempo real, la priorización de consultas y la gestión eficiente de recursos, asegurando la ejecución

de operaciones dentro de los límites de tiempo requeridos. Además, deben procesar grandes volúmenes de datos sin afectar el rendimiento, lo que las convierte en herramientas esenciales en entornos donde la velocidad y la precisión son fundamentales [44].

A continuación, la Tabla 3.10 presenta una comparación de distintas bases de datos en tiempo real para determinar la más adecuada para el sistema.

Tabla 3.10: Comparativa de bases de datos en tiempo real

Característica	Firebase RTDT	MongoDB	Amazon DynamoDB
Modelo de datos	NoSQL (JSON)	NoSQL (Document-based)	NoSQL (Key-Value, Document-based)
Sincronización en tiempo real	Sí	No (requiere configuración adicional)	Sí
Escalabilidad	Alta (con múltiples instancias)	Alta (sharding)	Alta (autoescalado)
Seguridad	Reglas de seguridad basadas en expresiones	Reglas de acceso basadas en roles	IAM (Identity and Access Management)
Facilidad de uso	Alta (Múltiples plataformas)	Media (requiere configuración)	Alta (gestión automática)

La información utilizada en este estudio proviene de las fuentes citadas en [45], [46] y [47]. El análisis comparativo de las características técnicas de diversas bases de datos en tiempo real determina que Firebase es la opción más adecuada para el sistema propuesto. Su capacidad de sincronización en tiempo real permite actualizar los datos instantáneamente en todos los dispo-

sitivos conectados. Además, su funcionalidad offline almacena la información localmente y la sincroniza automáticamente al restablecer la conexión a Internet. Por último, su sistema de seguridad proporciona un control granular sobre el acceso a los datos, garantizando su protección y gestión eficiente. Estas características consolidan a Firebase como la mejor alternativa para el desarrollo del sistema.

3.3. Implementación del sistema de predicción del clima

3.3.1. Partes usadas para la construcción del hardware del sistema

El desarrollo del sistema se basa en un flujo metodológico estructurado, desde la adquisición de datos hasta la generación de pronósticos climáticos. Requiere una selección rigurosa de componentes tecnológicos y variables meteorológicas, integrando sensores especializados y dispositivos de procesamiento cuya interoperabilidad optimiza la eficiencia del sistema.

Se implementa Firebase como plataforma de gestión en la nube, permitiendo el monitoreo en tiempo real y la visualización de las variables atmosféricas. Además, facilita el almacenamiento local de datos, esencial para su preprocesamiento. Una vez organizada la información, se entrenan los modelos predictivos y se evalúa su desempeño para validar la precisión de las proyecciones climáticas.

La Figura 3.7 muestra la secuencia operativa del sistema, resaltando la interconexión lógica entre cada etapa y su contribución a la generación de predicciones basadas en análisis científico-técnico. Esta representación gráfica facilita la comprensión del mecanismo propuesto, eviden-

ciando su coherencia metodológica y su potencial replicabilidad en contextos similares. El circuito se diseña utilizando el software Fritzing.

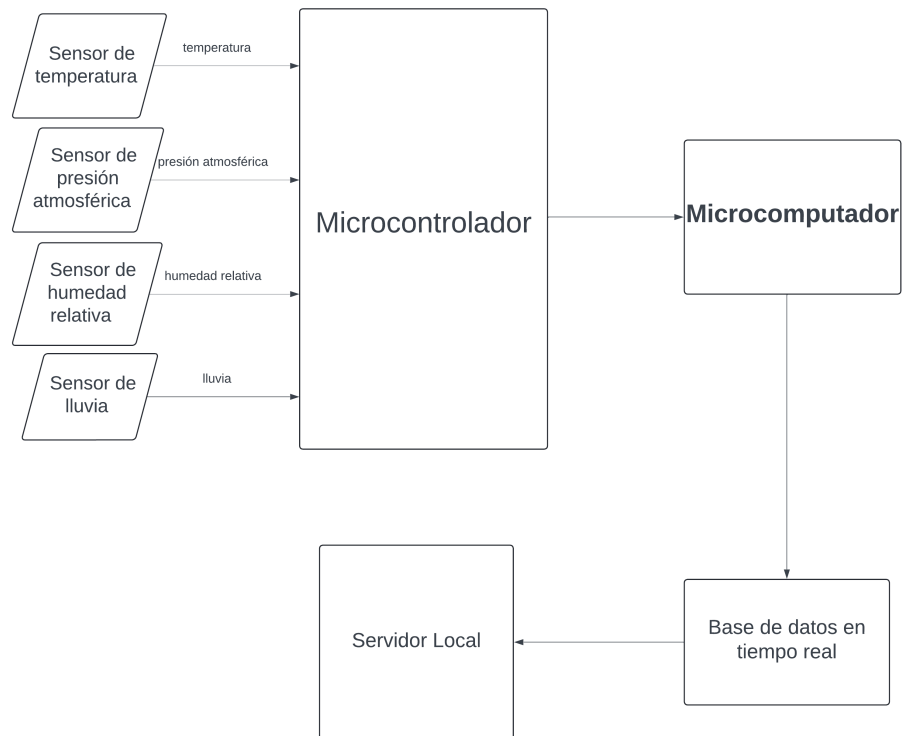


Figura 3.7: Secuencia de funcionamiento del sistema

3.3.2. Construcción y optimización del hardware y software del sistema

Los componentes empleados en la construcción del sistema incluyen:

- Sensor de humedad DHT22.
- Sensor de temperatura DHT22.
- Sensor de presión atmosférica BMP180.
- Sensor de lluvia DSK raindrops.

- Microcontrolador ESP32.
- Microcomputador Raspberry Pi 3B+.

Esta sección describe el circuito de conexión del sistema y sus componentes. La Figura 3.7 muestra el diseño, donde la integración de estos elementos asegura un funcionamiento óptimo y eficiente.

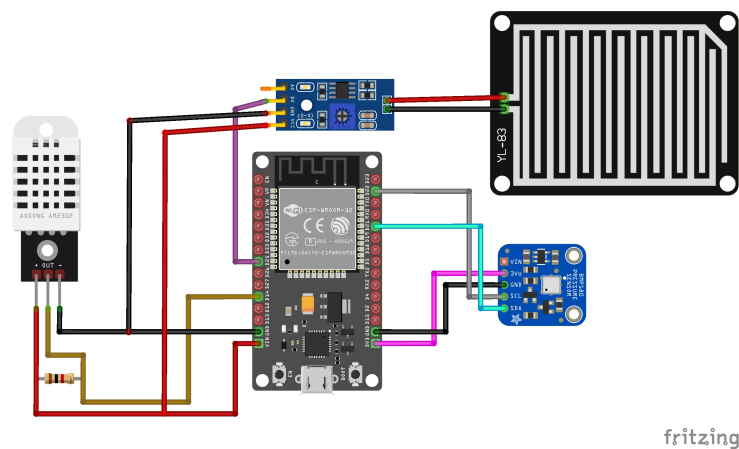


Figura 3.8: Circuito para la recepción de datos

3.3.3. Implementación del código para el funcionamiento del circuito

El código¹ inicia con la importación de las librerías necesarias para el correcto funcionamiento de los sensores. Luego, se definen las variables y arreglos requeridos durante la ejecución del programa. Posteriormente, se inicializa el puerto serial y se configura un contador de tiempo basado en el reloj interno de la *ESP32*, evitando pausas o retardos y garantizando una ejecución eficiente y continua de las funciones.

¹Consulte el código en “Código de Arduino”.

Al iniciar el sistema, se verifica el correcto funcionamiento de los sensores. Confirmada su operatividad, se realizan las mediciones empleando las bibliotecas específicas. Los datos recopilados se almacenan en un arreglo para su procesamiento. Finalmente, la función print transmite la información a través del puerto serial de forma estructurada y precisa. La Figura 3.9 muestra el diagrama de bloques del sistema.

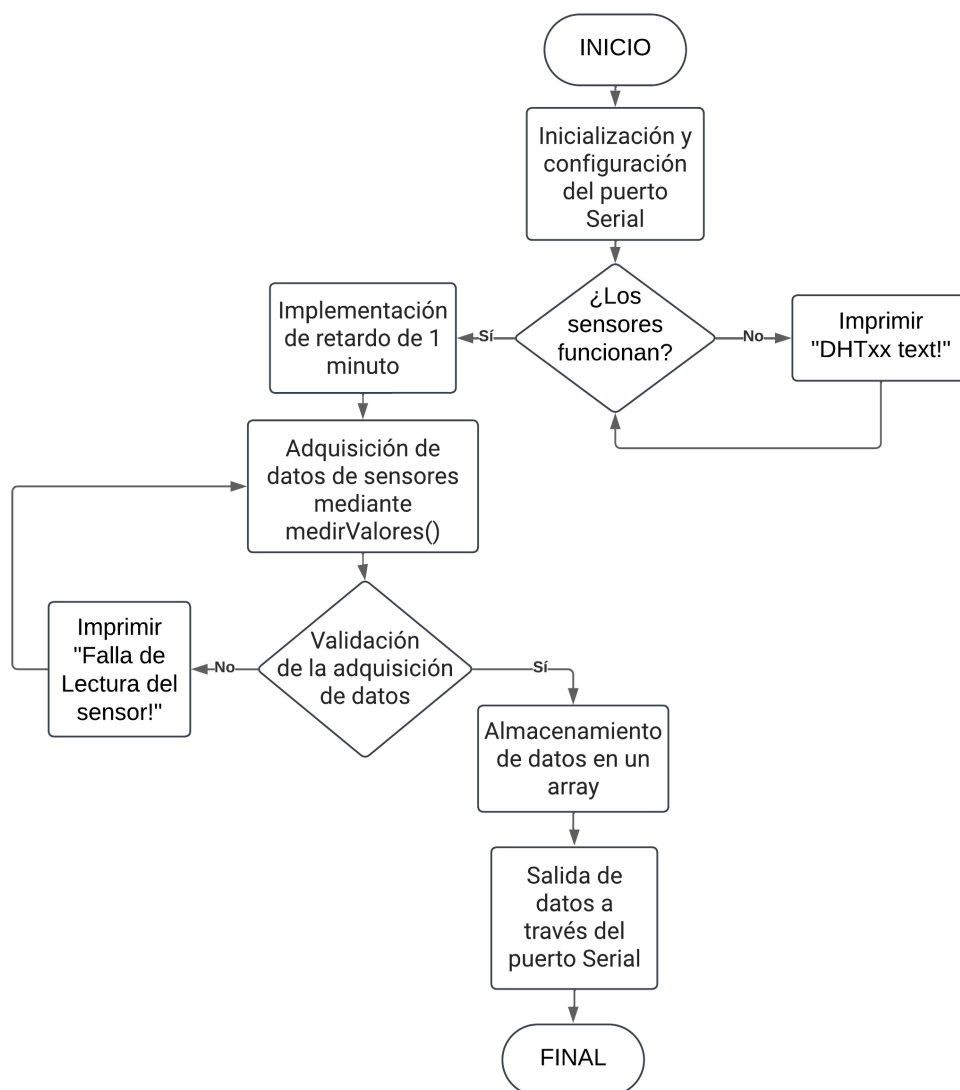


Figura 3.9: Diagrama de bloques del sistema de recepción de datos

3.3.4. Comunicación del puerto serial con el microcomputador

Se instala el sistema operativo *DietPi* en la *Raspberry Pi 3B+*, una distribución ligera y optimizada basada en Debian, diseñada para dispositivos de bajo rendimiento. Su elección responde a la necesidad de minimizar el consumo de recursos durante la recepción y comunicación de datos entre el microcontrolador *ESP32* y el microcomputador, garantizando así un funcionamiento eficiente.

Ante la ausencia de una pantalla en el microcomputador, se prescinde de interfaces gráficas para evitar el consumo innecesario de recursos y posibles ralentizaciones. En su lugar, se configura la *Raspberry Pi* para operar mediante *SSH (Secure Shell)*, un protocolo que permite la comunicación y transferencia de datos entre dispositivos a través de una red WiFi. Esta conexión, establecida dentro de la misma red local, facilita la administración remota y el monitoreo del sistema sin requerir hardware adicional.

Se establece la comunicación SSH mediante MobaXterm, configurando la dirección IP y la clave de acceso. Una vez conectados los dispositivos, la Figura 3.10 muestra la comunicación del puerto serial entre el microcontrolador y el microcomputador, garantizando una transmisión de datos precisa y eficiente.

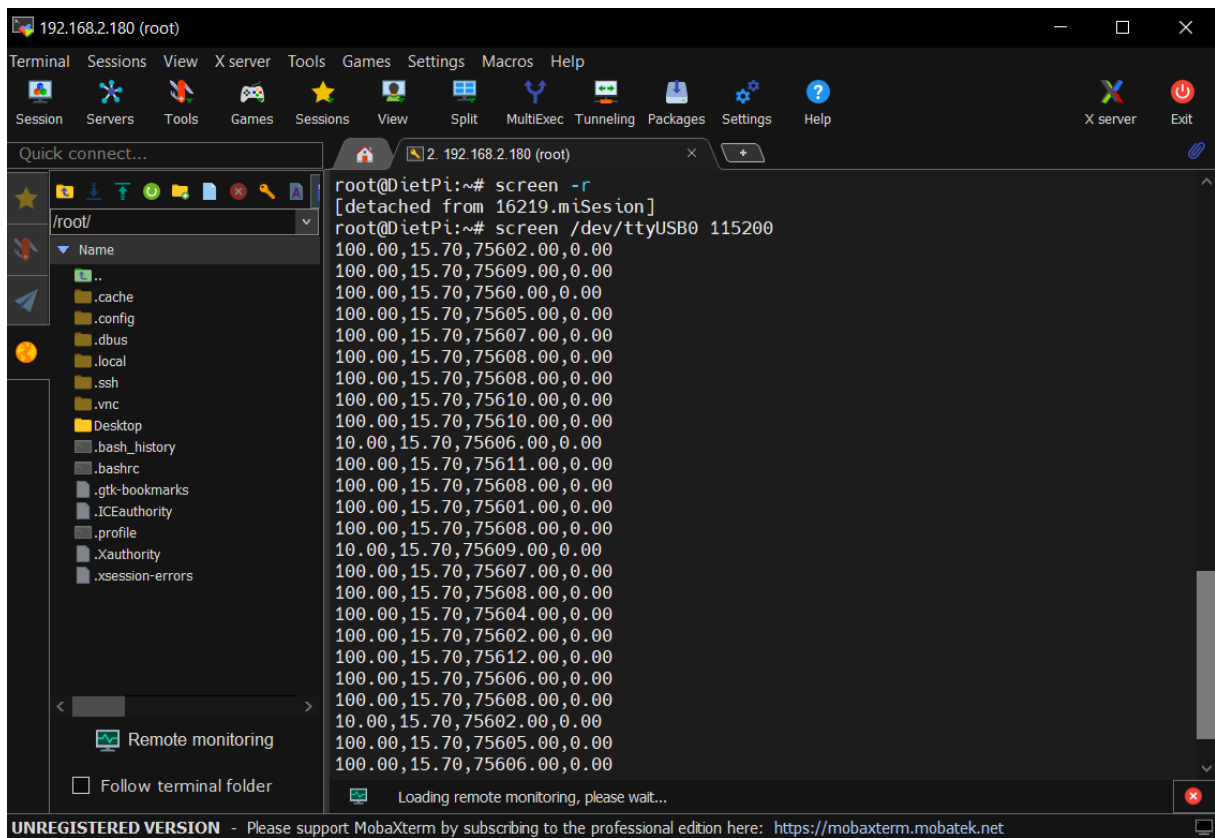


Figura 3.10: Conexión entre la Raspberri Pi y la ESP32 mediante SSH

3.3.5. Comunicación entre el microcomputador y la base de datos en tiempo real

En esta etapa, se configura *Firebase* como base de datos en tiempo real. El proceso incluye la definición de reglas de seguridad, la obtención de la *URL* de acceso a los datos y la descarga de un archivo *JSON* con las credenciales de autenticación. Finalmente, un *script* inicializa la conexión y establece la comunicación con la base de datos.

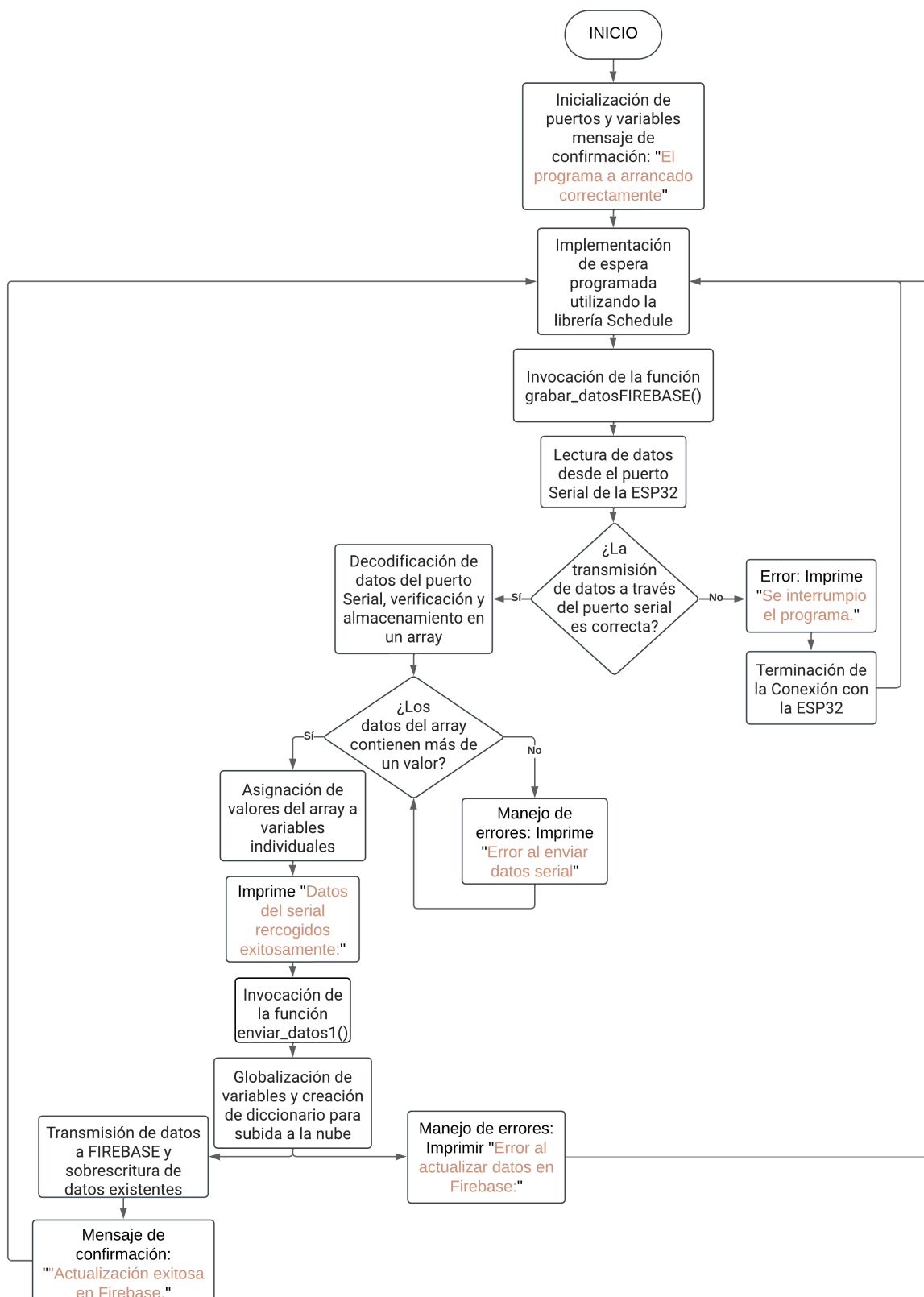
Para garantizar la comunicación efectiva entre los datos del puerto *serial* y la base de datos

en tiempo real, se implementa un código desarrollado en *Python* que emplea librerías especializadas en gestión de comunicación serial. Estas herramientas permiten la captura y almacenamiento temporal de los datos en variables estructuradas para su posterior transmisión. El sistema incorpora funciones condicionales que gestionan excepciones, previenen interrupciones del proceso y aseguran la continuidad operativa ante eventuales fallos mediante iteraciones automáticas.

Entre las librerías más importantes utilizadas se encuentran *firebase-admin*, que permite leer las credenciales de Firebase y gestionar la conexión con la base de datos; *logging*, un módulo incorporado en Python que genera registros de los eventos del programa, facilita la depuración de errores y permite definir filtros, formatos y destinos para los registros; *JSON*, utilizada para codificar y decodificar datos en formato *JSON*, fundamental en este script; y *schedule*, que ayuda a programar tareas específicas en intervalos de tiempo determinados.

Estas librerías, junto con otras empleadas en el desarrollo y el código². Además, en la Figura 3.11 se presenta un diagrama de bloques que resume la lógica del código implementado.

²Consulte el código en “Código de Python”.



3.3.6. Creación de la base de datos

En esta sección, se crea una base de datos mediante un algoritmo³desarrollado en *Jupyter Notebook*, una herramienta web que permite crear, compartir documentos y ejecutar *scripts* de manera interactiva.

En este *script*, primero se establece una conexión con Firebase utilizando sus credenciales para acceder a la base de datos en tiempo real. El objetivo es recibir los datos sensados al instante desde *Firebase*, almacenarlos en variables y luego utilizarlos en la base de datos creada. Para gestionar esta base de datos, se emplea la librería *sqlite3* en *Python*, que permite añadir los datos sensados por el sistema de recepción en intervalos de tiempo configurados. En la Figura 3.12, se observa cómo se agrega cada dato a la base de datos local.

```
2025-01-21 20:14:34,542 - INFO - Datos subidos exitosamente 24
{'Humedad (0-100)% RH': 100.0, 'Presion Atmosferica Pa': 75.66, 'Rain': 0, 'Temperatura C': 14.4, 'timestamp': 1737508408.0211568}
Data actualizada exitosamente
(1737506021.2683346, 14.8, 100.0, 75.6, 0)
(1737506617.7362952, 14.5, 100.0, 75.63, 0)
(1737507214.2063231, 14.1, 100.0, 75.64, 0)
(1737507810.0540345, 14.0, 100.0, 75.65, 0)
(1737508408.0211568, 14.4, 100.0, 75.66, 0)

2025-01-21 20:24:35,150 - INFO - Datos subidos exitosamente 25
{'Humedad (0-100)% RH': 100.0, 'Presion Atmosferica Pa': 75.67, 'Rain': 0, 'Temperatura C': 14.2, 'timestamp': 1737509004.0559192}
Data actualizada exitosamente
(1737506617.7362952, 14.5, 100.0, 75.63, 0)
(1737507214.2063231, 14.1, 100.0, 75.64, 0)
(1737507810.0540345, 14.0, 100.0, 75.65, 0)
(1737508408.0211568, 14.4, 100.0, 75.66, 0)
(1737509004.0559192, 14.2, 100.0, 75.67, 0)

2025-01-21 20:34:35,784 - INFO - Datos subidos exitosamente 26
{'Humedad (0-100)% RH': 100.0, 'Presion Atmosferica Pa': 75.67, 'Rain': 0, 'Temperatura C': 14.4, 'timestamp': 1737509605.1523612}
Data actualizada exitosamente
(1737507214.2063231, 14.1, 100.0, 75.64, 0)
(1737507810.0540345, 14.0, 100.0, 75.65, 0)
(1737508408.0211568, 14.4, 100.0, 75.66, 0)
(1737509004.0559192, 14.2, 100.0, 75.67, 0)
(1737509605.1523612, 14.4, 100.0, 75.67, 0)
```

Figura 3.12: Datos añadidos a la base de datos creada

³Consulte el código en “Código en Jupyter Notebook”.

También se emplean datos históricos del mismo sector sensado, obtenidos a través de *NASA POWER*, una página web que permite acceder a registros meteorológicos antiguos. Estos datos mejoran la base de datos mediante un proceso de preprocesamiento y depuración, eliminando información que podría confundir al modelo predictivo [48].

Finalmente, la base de datos creada incluye tanto datos sensados sobre la probabilidad de lluvia, obtenidos con el hardware del sistema, como datos históricos proporcionados por *NASA POWER*. Cabe mencionar que la herramienta *NASA POWER* proporciona datos en *mm/hora*, por lo que clasifica y etiqueta la información según la cantidad de *mm/hora*, de acuerdo con la Tabla 3.11.

Tabla 3.11: Clasificación de la lluvia según su intensidad

Intensidad	mm/h	mm/6h	mm/15 días
Ligera	>2.5	>15	>225
Moderada	2.5-7.5	15-45	225-675
Fuerte	>7.5	>45	>675

La información utilizada en este estudio proviene de [49] y facilita la construcción de la base de datos⁴. Inicialmente, se lleva a cabo el preprocesamiento de los registros obtenidos mediante la herramienta *NASA POWER*.

Una vez estructurada la base de datos, procede a etiquetar los registros de manera adecuada, utilizando como referencia la Tabla 3.11. Este proceso clasifica la intensidad de las lluvias según los criterios establecidos. Los resultados de este etiquetado se observan en la Tabla 3.12, donde se detallan las categorías asignadas a cada registro.

Al unir la base de datos local con registros antiguos extraídos y registros actuales sensados

⁴Consulte la información en la sección “Base de datos *NASA POWER*”.

Tabla 3.12: Etiquetas para la base de datos creada

Estado de la lluvia	mm/h	Etiqueta
Sin lluvia	0	0
Lluvia ligera	0-2.5	1
Lluvia moderada	2.5-7.5	2
Lluvia fuerte	>7.5	3

con el sistema, se obtiene una base de datos⁵ con 15,112 datos almacenados como se ilustra en la Figura 3.13.

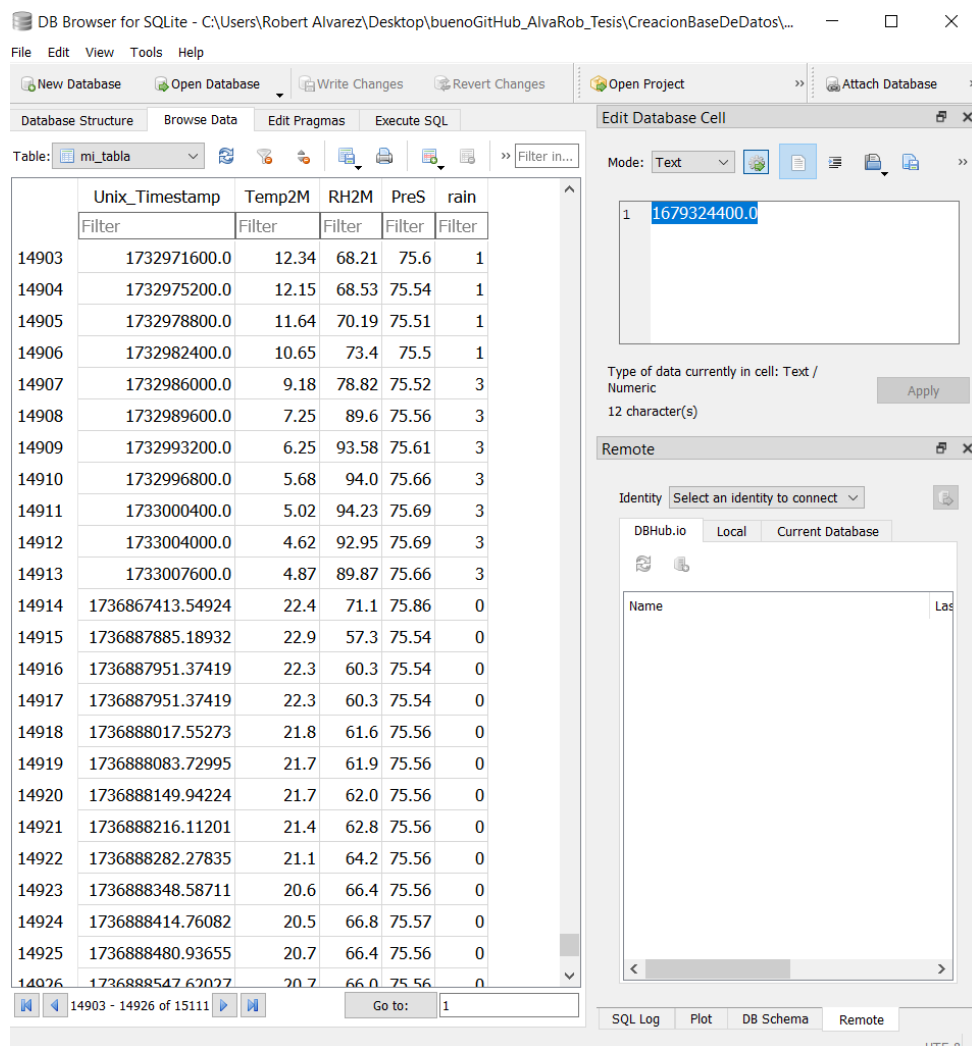
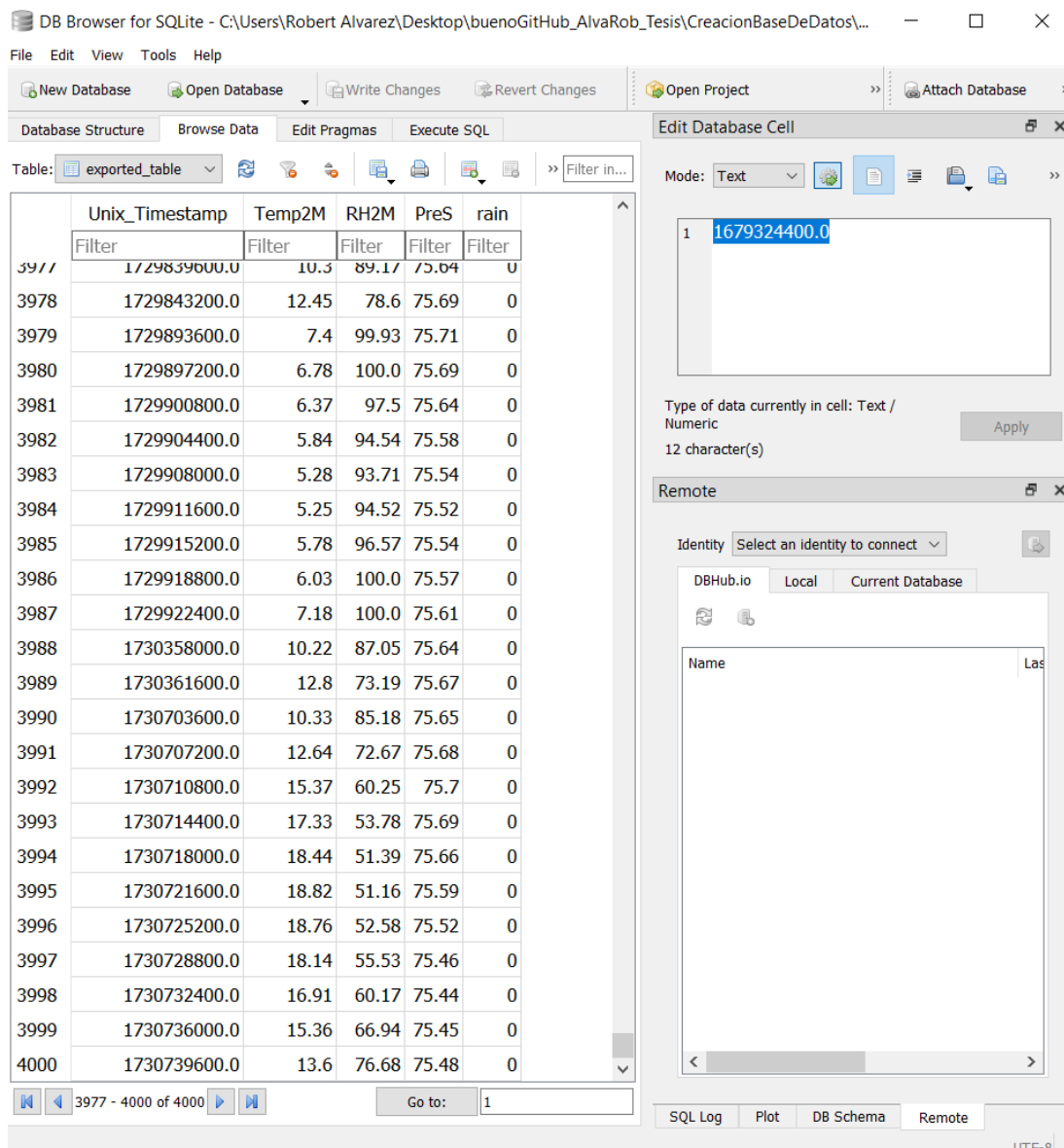


Figura 3.13: Total de datos recopilados y almacenados en la base de datos local.

⁵Consulte la información en la sección “Base de datos unificada”.

Mediante el balanceo de datos, se reduce significativamente el volumen de información y se depuran etiquetas ruidosas que podrían generar confusión durante el entrenamiento del modelo. Como resultado⁶, se obtiene un conjunto de 4000 datos limpios y organizados, los cuales se pueden observar en la Figura 3.14.



	Unix_Timestamp	Temp2M	RH2M	PreS	rain
3977	1729839600.0	10.3	89.17	75.64	0
3978	1729843200.0	12.45	78.6	75.69	0
3979	1729893600.0	7.4	99.93	75.71	0
3980	1729897200.0	6.78	100.0	75.69	0
3981	1729900800.0	6.37	97.5	75.64	0
3982	1729904400.0	5.84	94.54	75.58	0
3983	1729908000.0	5.28	93.71	75.54	0
3984	1729911600.0	5.25	94.52	75.52	0
3985	1729915200.0	5.78	96.57	75.54	0
3986	1729918800.0	6.03	100.0	75.57	0
3987	1729922400.0	7.18	100.0	75.61	0
3988	1730358000.0	10.22	87.05	75.64	0
3989	1730361600.0	12.8	73.19	75.67	0
3990	1730703600.0	10.33	85.18	75.65	0
3991	1730707200.0	12.64	72.67	75.68	0
3992	1730710800.0	15.37	60.25	75.7	0
3993	1730714400.0	17.33	53.78	75.69	0
3994	1730718000.0	18.44	51.39	75.66	0
3995	1730721600.0	18.82	51.16	75.59	0
3996	1730725200.0	18.76	52.58	75.52	0
3997	1730728800.0	18.14	55.53	75.46	0
3998	1730732400.0	16.91	60.17	75.44	0
3999	1730736000.0	15.36	66.94	75.45	0
4000	1730739600.0	13.6	76.68	75.48	0

Figura 3.14: Base de datos con datos balanceados

⁶Consulte la información en la sección “Base de datos balanceada”.

Finalmente, se obtiene una base de datos mejorada con solo dos etiquetas, lo que ofrece mejores resultados para entrenar el modelo, como se explicará más adelante. En esta base de datos final⁷, se cuenta con un total de 2000 datos, obtenidos después de preprocesar y balancear la información, como se muestra en la Figura 3.15.

	Unix_Timestamp	Temp2M	RH2M	PreS	rain
	Filter	Filter	Filter	Filter	Filter
766	1686013200.0	5.8	99.07	75.69	1
767	1686016800.0	5.47	98.84	75.66	1
768	1686020400.0	5.32	98.36	75.65	1
769	1686024000.0	5.33	97.81	75.66	1
770	1686027600.0	5.52	97.29	75.68	1
771	1686031200.0	6.25	96.52	75.7	1
772	1686034800.0	8.18	92.74	75.74	1
773	1686038400.0	9.82	89.96	75.78	1
774	1686042000.0	11.35	83.91	75.8	1
775	1686045600.0	12.54	78.65	75.79	1
776	1686070800.0	10.95	85.56	75.58	1
777	1686074400.0	8.99	96.16	75.61	1
778	1686078000.0	8.21	99.39	75.65	1
779	1686081600.0	7.79	100.0	75.71	1
780	1686085200.0	7.39	100.0	75.76	1
781	1686088800.0	6.85	100.0	75.78	1
782	1686092400.0	6.33	99.82	75.77	1

Figura 3.15: Base de datos con datos balanceados

⁷Consulte la información en la sección “Base de datos optimizada para el modelo de *Machine Learning*”.

3.4. Descripción de los modelos de machine learning utilizados.

A continuación, se presenta una breve descripción de los modelos de machine learning utilizados para predecir el clima, junto con algunos detalles específicos:

3.4.1. Árbol de Decisión

El algoritmo de árbol de decisión opera como un árbol jerárquico, donde cada nodo representa una característica de los datos y cada rama una decisión basada en esa característica. Emplea la estrategia de “*divide y vencerás*” para identificar los mejores puntos de división, evaluando todas las posibles divisiones y seleccionando la más adecuada. Los hiperparámetros, como la profundidad del árbol, se optimizan mediante una búsqueda en cuadrícula, probando diferentes combinaciones y evaluando el rendimiento del modelo. Este enfoque permite crear un modelo predictivo eficaz y fácil de interpretar, basado en las características más relevantes de los datos [50].

3.4.2. K-Nearest Neighbors

El algoritmo *K-Nearest Neighbors* (*KNN*) clasifica nuevas instancias basándose en la proximidad con los datos de entrenamiento más cercanos. Para implementarlo en Python, primero prepara los datos, dividiéndolos en conjuntos de entrenamiento y validación con la función *train_test_split* de *scikit-learn*. Luego, escala las características utilizando técnicas como Min-

MaxScaler para garantizar uniformidad [50].

El modelo se entrena utilizando la clase *KNeighborsClassifier*, donde se calculan las distancias entre los puntos de datos. Para optimizar el valor de '*k*', se emplean métodos como validación cruzada o búsqueda en cuadrícula, maximizando la precisión en datos no vistos. Finalmente, el modelo se evalúa con métricas como precisión, sensibilidad y *F1-score*, y se analiza su eficiencia computacional, incluyendo tiempos de entrenamiento y predicción [50].

3.4.3. Máquinas de Vectores de Soporte

Las Máquinas de Vectores de Soporte (*SVM*) constituyen un algoritmo de aprendizaje supervisado, empleado principalmente para clasificación. Este algoritmo maximiza el margen entre clases para identificar el hiperplano de separación óptimo. En casos donde los datos no son linealmente separables, *SVM* aplica el truco del kernel para proyectarlos a un espacio de mayor dimensión.

En la práctica, los datos se dividen en conjuntos de entrenamiento y validación. El clasificador *SVM* se entrena con el conjunto de entrenamiento y se evalúa mediante validación cruzada, generando métricas como precisión, sensibilidad y *F1-score*. Las predicciones se realizan en el conjunto de prueba, construyendo una matriz de confusión y calculando el tiempo total de eficiencia computacional. La selección de hiperparámetros, como el tipo de *kernel* y el parámetro de regularización, se realiza para optimizar el rendimiento del modelo [50].

3.4.4. Bosque Aleatorio

El algoritmo de bosques aleatorio es conocido por su capacidad para construir múltiples árboles de decisión y combinar sus predicciones, ofreciendo resultados precisos y robustos. En su implementación, se cargan y dividen los datos, y se define una cuadrícula de hiperparámetros para optimizar el modelo mediante *GridSearchCV* de *scikit-learn*. Se ajustan parámetros como el número de árboles y la profundidad máxima de cada árbol. Luego, se utilizan los datos de prueba para hacer predicciones y se generan métricas de rendimiento detalladas. También se construye y visualiza una matriz de confusión para evaluar el modelo de manera intuitiva. Finalmente, se calcula el tiempo total de eficiencia computacional, incluyendo el tiempo de entrenamiento, predicción y visualización [50].

3.4.5. Interfaz gráfica para visualización de los resultados de la predicción.

Primero, para acceder a la interfaz final de predicción, es necesario utilizar el modelo previamente entrenado y mejor puntuado, que se explica en la siguiente sección. Luego, se conecta a la base de datos en tiempo real e inicializa algunas variables y arrays necesarios. Para que esta interfaz sea más amigable e intuitiva, se presentan los valores sensados mediante *frames* usando la interfaz de *Tkinter* en *Python*. Dentro de estos *frames*, se utiliza el módulo *matplotlib* para graficar cada valor sensado de manera consecutiva.

Finalmente, al presionar el botón “*adquirir*”, los datos se actualizan mediante la base de datos en tiempo real y se grafican los datos actuales. En otro *frame*, se observan los datos de

salida de la lluvia, tanto los sensados como los predichos. Además, se muestra la probabilidad de lluvia⁸, que varía de 0 a 1, según el modelo entrenado. El resultado final de la interfaz se observa en la Figura 3.16.

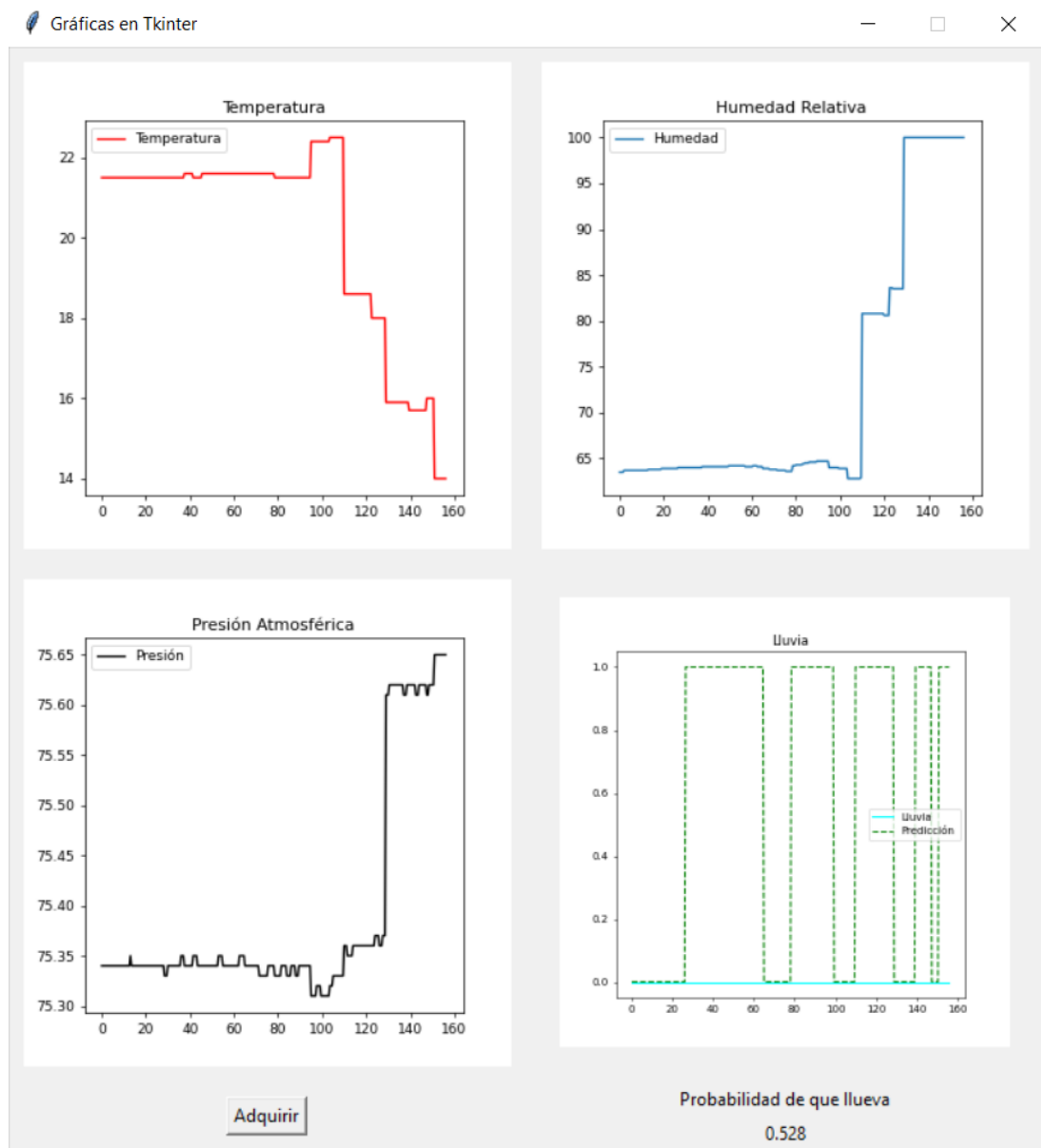


Figura 3.16: Interfaz para visualización de los resultados de la predicción

⁸Consulte el resultado en "Interfaz gráfica."

Capítulo 4

Pruebas de funcionamiento

En este capítulo, se analizan y evalúan diferentes parámetros para determinar cuál es el mejor modelo entrenado que resuelve el problema de predicción del clima. El objetivo es identificar el modelo que ofrezca las predicciones más precisas y confiables, basándose en los datos recopilados y procesados.

4.1. Descripción de las pruebas

A continuación, se presentan los resultados obtenidos mediante diversas métricas para evaluar los algoritmos de predicción del clima. Se describen los criterios de evaluación y las métricas utilizadas para comparar los algoritmos seleccionados.

Para este análisis, se utiliza una división del conjunto de datos en un 75 % para entrenamiento y un 25 % para prueba. Esta proporción permite validar las mejores valoraciones en cuanto a las métricas, garantizando así la selección del algoritmo más robusto y eficaz.

4.1.1. Métricas de validación

Existen diversas métricas para evaluar los mejores modelos. Los principios de evaluación aplican diferentes métricas para probar la eficacia y agilidad del algoritmo. Por ello, se seleccionan las siguientes métricas que se muestran en la Tabla 4.1.

Tabla 4.1: Métricas para evaluar los modelos.

Métrica	Significado
Exactitud (Accuracy)	Proporción de instancias clasificadas correctamente por el modelo.
Sensibilidad (Recall)	Proporción de valores positivos correctamente clasificados.
Precisión (Precision)	Proporción de predicciones positivas correctas.

4.2. Análisis de resultados

Se analiza cada uno de los algoritmos seleccionados para abordar este problema, evaluando su rendimiento y eficacia. Para ello, se toman en consideración las métricas presentadas en la Tabla 4.1, comparando el desempeño de cada algoritmo en función de las bases de datos disponibles, incluyendo aquellas balanceadas con 4, 3 y 2 etiquetas, y también reduciendo una dimensión.

4.2.1. Análisis de los modelos con 3 dimensiones y 4 etiquetas

En esta sección, se analizan los resultados del entrenamiento de los modelos seleccionados, utilizando las diferentes métricas presentadas en la Tabla 4.1.

Al observar la Tabla 4.2, se aprecia que el modelo *KNN* presenta la mejor exactitud. Sin embargo, su precisión es demasiado baja para considerarlo como el modelo final.

Tabla 4.2: Exactitud de los diferentes modelos 3d 4e

Modelo	Exactitud
KNN	0.46
SVM (LINEAR)	0.37
SVM (RBF)	0.40
Random Forest	0.41
Árbol de decisión	0.38

En la Tabla 4.3, se observa que la precisión de los modelos presenta dificultades significativas en las etiquetas 1, 2 y 3, lo cual afecta a todos los algoritmos evaluados. Sin embargo, entre ellos, el modelo *KNN* destaca como el de mejor rendimiento en esta prueba.

Tabla 4.3: Precisión de los diferentes modelos 3d 4e

Modelo	Etiqueta 0	Etiqueta 1	Etiqueta 2	Etiqueta 3
KNN	0.61	0.39	0.44	0.41
SVM (LINEAR)	0.45	0.32	0.34	0.39
SVM (RBF)	0.56	0.35	0.29	0.39
Random Forest	0.48	0.48	0.27	0.3
Árbol de decisión	0.48	0.37	0.33	0.34

En la Tabla 4.4, se observa que todos los modelos¹ tienen dificultades con las etiquetas 2 y 3, lo cual representa un problema para su uso. Sin embargo, el modelo *KNN* destaca al obtener una de las mejores puntuaciones entre los demás modelos.

Tabla 4.4: Sensibilidad de los diferentes modelos 3d 4e

Modelo	Etiqueta 0	Etiqueta 1	Etiqueta 2	Etiqueta 3
KNN	0.53	0.47	0.42	0.41
SVM (LINEAR)	0.58	0.51	0.16	0.22
SVM (RBF)	0.5	0.41	0.21	0.45
Random Forest	0.34	0.63	0.18	0.25
Árbol de decisión	0.48	0.34	0.37	0.34

¹Consulte los resultados de los modelos entrenados en la sección “Modelos de *Machine Learning* en 4d 3e.”.

4.2.2. Análisis de los modelos con 2 dimensiones y 4 etiquetas

En esta sección, se analizan los modelos de *machine learning* seleccionados para la predicción y se observan los resultados de las diferentes métricas elegidas, como se muestra en la Tabla 4.1.

Primero, al observar la Tabla 4.5, se nota que prácticamente no hay diferencias en comparación con el análisis anterior. En este caso, los resultados son peores al eliminar una dimensión.

Tabla 4.5: Exactitud de los diferentes modelos 2d 4e

Modelo	Exactitud
KNN	0.41
SVM (LINEAR)	0.37
SVM (RBF)	0.40
Random Forest	0.36
Árbol de decisión	0.32

En la Tabla 4.6, se observa una caída significativa en los puntajes de precisión para las etiquetas 1, 2 y 3. Esto resalta que el único modelo que sobresale es el de *KNN*.

Tabla 4.6: Precisión de los diferentes modelos 2d 4e

Modelo	Etiqueta 0	Etiqueta 1	Etiqueta 2	Etiqueta 3
KNN	0.59	0.30	0.33	0.41
SVM (LINEAR)	0.44	0.31	0.37	0.37
SVM (RBF)	0.58	0.34	0.33	0.38
Random Forest	0.54	0.32	0.28	0.3
Árbol de decisión	0.48	0.29	0.27	0.29

En la Tabla 4.7, se observa una ligera mejora en la sensibilidad, pero no es suficientemente eficiente. De hecho, las etiquetas 2 y 3 presentan problemas más críticos en algunos algoritmos².

Si se utilizaran estos modelos, serían muy ineficientes y cometerían muchos errores.

²Consulte los resultados de los modelos entrenados en la sección “Modelos de *Machine Learning* en 2d 4e.”.

Tabla 4.7: Sensibilidad de los diferentes modelos 2d 4e

Modelo	Etiqueta 0	Etiqueta 1	Etiqueta 2	Etiqueta 3
KNN	0.59	0.27	0.43	0.34
SVM (LINEAR)	0.58	0.49	0.08	0.37
SVM (RBF)	0.49	0.38	0.3	0.44
Random Forest	0.5	0.3	0.29	0.33
Árbol de decisión	0.5	0.26	0.29	0.3

4.2.3. Análisis de los modelos con 3 dimensiones y 2 etiquetas

En esta sección, se analizan los modelos de *machine learning* utilizados para la predicción y se comparan los resultados de las diferentes métricas seleccionadas, como se muestra en la Tabla 4.1.

En la Tabla 4.8, se observa una mejora significativa en comparación con otros algoritmos. Dado que la predicción del clima es una problemática compleja, se asume que esta configuración de variables es una de las más adecuadas para entrenar el modelo con cualquiera de estos algoritmos.

Tabla 4.8: Exactitud de los diferentes modelos 3d 2e

Modelo	Exactitud
KNN	0.72
SVM (LINEAR)	0.7
SVM (RBF)	0.7
Random Forest	0.71
Árbol de decisión	0.66

En la Tabla 4.9, se observa una precisión muy alta para la etiqueta 0. Sin embargo, no se debe subestimar la precisión de la etiqueta 1, ya que para este problema representa una condición favorable del sistema. Este parámetro es crucial para entrenar el sistema de manera efectiva.

En la Tabla 4.10, se observa una mejora notable en comparación con las secciones anteriores. Por ello, es altamente recomendable seleccionar el mejor algoritmo³utilizando esta configura-

ción de etiquetas y dimensiones.

Tabla 4.9: Precisión de los diferentes modelos 3d 2e

Modelo	Etiqueta 0	Etiqueta 1
KNN	0.77	0.69
SVM (LINEAR)	0.71	0.69
SVM (RBF)	0.76	0.67
Random Forest	0.71	0.69
Árbol de decisión	0.65	0.67

Tabla 4.10: Sensibilidad de los diferentes modelos 3d 2e

Modelo	Etiqueta 0	Etiqueta 1
KNN	0.63	0.81
SVM (LINEAR)	0.67	0.73
SVM (RBF)	0.59	0.81
Random Forest	0.65	0.75
Árbol de decisión	0.68	0.64

4.3. Análisis comparativo de algoritmos para la selección del modelo final

En las subsecciones anteriores, se destacan los dos modelos con las mejores puntuaciones en las métricas comparadas: *K-Vecinos Cercanos* y bosque aleatorio. En esta sección, se comparan estos modelos para determinar cuál es el más adecuado para predecir el clima. La comparación se realiza utilizando configuraciones de 3 dimensiones y 4 etiquetas, 2 dimensiones y 4 etiquetas, y 3 dimensiones y 2 etiquetas, siguiendo el mismo enfoque de la comparación anterior.

³Consulte los resultados de los modelos entrenados en la sección “Modelos de *Machine Learning* en 3d 2e.”.

4.3.1. Resultados y análisis

4.3.1.1. Comparación de resultados con 3 dimensiones y 4 etiquetas

El análisis comparativo revela que la presencia de datos no relacionados adecuadamente o con valores significativamente similares en la base de datos provoca numerosos errores en el modelo. Esto se refleja en una exactitud muy baja, como se muestra en la Tabla 4.11. Por esta razón, no es necesario comparar precisión y sensibilidad. Este formato de base de datos contiene muchos datos falsos o ruidosos que confunden al algoritmo, y la presencia de datos con desviaciones provoca predicciones erróneas.

Tabla 4.11: Exactitud de los diferentes modelos 3d 4e resultados

Modelo	Exactitud
KNN	0.46
Random Forest	0.41

4.3.1.2. Comparación de resultados con 2 dimensiones y 4 etiquetas

En esta sección, se analiza con una dimensión menos, tomando la exactitud como la métrica principal a evaluar. Al comparar estas dos secciones, se observa que la de 2 dimensiones presenta peores resultados, como se muestra en la Tabla 4.12. Por ello, no es necesario analizar la precisión y la sensibilidad debido a las predicciones erróneas, lo que resultaría en un modelo inexacto. La causa principal de estos errores son los datos ruidosos y falsos presentes en la base de datos de 4000 registros.

Tabla 4.12: Exactitud de los diferentes modelos 2d 4e resultados

Modelo	Exactitud
KNN	0.41
Random Forest	0.36

4.3.1.3. Comparación de resultados con 2 dimensiones y 2 etiquetas

Es relevante detallar esta sección, ya que en ella los dos algoritmos obtienen las mejores puntuaciones según las métricas establecidas en la Tabla 4.1. A continuación, se observan y comparan estos dos algoritmos con las mejores puntuaciones.

- Exactitud: En la Tabla 4.13, se observa una notable mejoría en la puntuación de la exactitud. Esto permite continuar analizando el sistema utilizando las otras dos métricas para determinar cuál de los dos algoritmos es el más adecuado. En esta instancia, el *KNN* resulta ser más exacto.

Tabla 4.13: Exactitud de los diferentes modelos 3d 2e resultados

Modelo	Exactitud
KNN	0.72
Random Forest	0.71

- Precisión: En la Tabla 4.14, se observa que al comparar la clase 0, el algoritmo *K-Vecinos Cercanos* muestra una mayor precisión. Sin embargo, es importante tener en cuenta que este algoritmo tiene una duración de ejecución más lenta en comparación con el bosque aleatorio. Por otro lado, al comparar la clase 1, ambos algoritmos presentan puntuaciones idénticas. Dado que la ejecución del *K-Vecinos Cercanos* es más lenta, se concluye que el bosque aleatorio es más eficiente y eficaz en esta sección.

Tabla 4.14: Precisión de los diferentes modelos 3d 2e resultados

Modelo	Etiqueta 0	Etiqueta 1
KNN	0.77	0.69
Random Forest	0.71	0.69

- Sensibilidad: Como se observa en la Tabla 4.15, la sensibilidad varía considerablemente.

En la clase 1, el algoritmo *K-Vecinos Cercanos* obtiene una puntuación superior, mientras que en la clase 0, el bosque aleatorio presenta una mejor puntuación. Sin embargo, dado que el *K-Vecinos Cercanos* no es tan eficiente, su implementación en un software podría hacerlo más lento.

Tabla 4.15: Sensibilidad de los diferentes modelos 3d 2e resultados

Modelo	Etiqueta 0	Etiqueta 1
KNN	0.63	0.81
Random Forest	0.65	0.75

4.3.2. Propuesta de un modelo optimizado para mayor eficiencia

Tras evaluar las métricas y puntuaciones de cada algoritmo, se selecciona el bosque aleatorio por su ligereza y capacidad para evitar el sobreajuste (*overfitting*). A diferencia de *K-Vecinos Cercanos*, ofrece predicciones rápidas, además de ser el modelo más equilibrado y eficiente.

En resumen, los resultados indican que el bosque aleatorio es uno de los modelos más eficaces para entrenar con la base de datos disponible en la predicción climática. La precisión del modelo depende de una base de datos sólida, depurada y libre de información ruidosa, falsificada o corrupta.

Capítulo 5

Conclusiones y trabajo futuro

5.1. Conclusiones

- Se diseñó un sistema inteligente para la predicción de variables climáticas locales, considerando la interacción entre distintos factores atmosféricos. Este diseño proporciona una herramienta eficiente y accesible que contribuye a la toma de decisiones en sectores dependientes del clima, optimizando recursos y mejorando la planificación en actividades como la agricultura.
- Se desarrolló un modelo de predicción climática basado en la identificación y análisis de variables relevantes, garantizando su aplicabilidad en entornos con microclimas específicos. La construcción de una base de datos balanceada y preprocesada permitió mejorar la calidad de las predicciones y reducir la influencia de datos atípicos.
- Se implementó un sistema prototipo que integra sensores, dispositivos de procesamiento

y una plataforma de internet de las cosas (IoT). La comunicación en tiempo real entre estos elementos asegura la recolección, almacenamiento y análisis de datos climáticos, lo que permite generar predicciones con una infraestructura de bajo costo y fácil acceso.

- Se evaluaron y compararon diversos algoritmos de aprendizaje automático para seleccionar el modelo más eficiente en la predicción climática. El algoritmo de bosque aleatorio mostró el mejor desempeño en términos de precisión y optimización de recursos, superando otras alternativas como K-Vecinos Cercanos, que requería mayor capacidad computacional. Esta selección garantiza un equilibrio entre exactitud y viabilidad operativa en el sistema propuesto.

5.2. Recomendaciones

- Se recomienda utilizar el algoritmo de bosque aleatorio para la predicción climática, debido a su versatilidad y eficiencia. Este modelo es adecuado para entrenar sistemas con un número reducido de variables de entrada, permitiendo obtener predicciones rápidas y precisas.
- Para mejorar la precisión del sistema de predicción en microclimas, es fundamental contar con una base de datos sólida y un registro histórico de las variables climáticas del sector. Esto facilita el balanceo y la corrección de valores sensados, reduciendo la dispersión de los datos y minimizando errores en las predicciones.
- Se sugiere incrementar el número de variables de entrada en el modelo. En este estudio

se emplearon únicamente temperatura, humedad relativa y presión atmosférica. La inclusión de más variables permitirá mejorar la precisión y confiabilidad de las predicciones, aumentando la exactitud del modelo sin riesgo de sobreajuste.

5.3. Trabajo futuro

- El sistema actual muestra una mejora significativa en las métricas analizadas. No obstante, la inclusión de una variable adicional, como la velocidad del viento, permitiría optimizar su desempeño. Esto facilitaría la evaluación del modelo con distintos algoritmos de clasificación, verificando su viabilidad. Además, la infraestructura de hardware admite la incorporación de sensores adicionales para medir estas variables y reducir errores en las predicciones.
- El cambio climático impacta directamente en sectores como la agricultura, donde la incertidumbre climática representa un desafío constante. Un modelo de predicción más preciso y eficiente beneficiaría a quienes dependen de la producción agrícola, minimizando pérdidas. En Ecuador, uno de los principales exportadores de rosas y otros productos, un sistema avanzado de pronóstico climático contribuiría a la estabilidad del sector. Se recomienda, a futuro, mejorar el diseño del sistema mediante la incorporación de más variables climáticas, la comparación con métricas adicionales y la exploración de técnicas de aprendizaje profundo para optimizar las predicciones a largo plazo.

Bibliografía

- [1] C. A. Sánchez Zavaleta, “Evolución del concepto de cambio climático y su impacto en la salud pública del Perú,” *Revista peruana de medicina experimental y salud pública*, vol. 33, pp. 128–138, 2016.
- [2] N. Singh, S. Chaturvedi, and S. Akhter, “Weather forecasting using machine learning algorithm,” in *2019 International Conference on Signal Processing and Communication (ICSC)*. IEEE, 2019, pp. 171–174.
- [3] C. A. Mosquera Barrionuevo, “Aplicación de la tecnología iot (internet of things) para la medición de variables hidrometeorológicas en la agricultura sostenible: optimización del recurso hídrico mediante la tecnología iot a través del sensor de nivel de agua, esto con el fin de evitar un sobre exceso o un déficit de este recurso durante el riego de los cultivos de ciclo corto aplicado hacia una agricultura sostenible,” B.S. thesis, Quito: EPN, 2022., 2022.
- [4] J. F. Mendoza Sánchez, E. ADAME VALENZUELA, O. MARCOS PALOMARES, and L. GRADILLA HERNANDEZ, “Sistema de informacion metereologica y climatica para

- el diseno y operacion de las carreteras en mexico,” *Publicación Técnica*, no. 575, 2020.
- [5] A. D. Leal Quintero and S. T. Portes Rodriguez, “Análisis de las variables climatológicas de precipitación y temperatura de los municipios de tocaima y jerusalén,” Ph.D. dissertation, 2020.
- [6] S. A. GUILLÉN MEZA, “Diseño, construcción y aplicación de un dispositivo electrónico basado en arduino para monitoreo en tiempo real de parámetros ambientales,” B.S. thesis, Jipijapa-Unesum, 2022.
- [7] B. Guerrero, W. Daniel, M. J. Mora Villacís, D. V. Carrera Villacrés, and E. R. M. Guerrón Varela, “Instalación de sensores climáticos, generador atmosférico iot y análisis lineal y no lineal de las variables generadas.”
- [8] I. Bonilla-Fabela, A. Tavizon-Salazar, M. Morales-Escobar, L. T. Guajardo-Muñoz, and C. I. Laines-Alamina, “Iot, el internet de las cosas y la innovación de sus aplicaciones,” *Vinculatégica efan*, vol. 2, no. 1, pp. 2313–2340, 2016.
- [9] D. Dujovne, T. Watteyne, G. Mercado, A. Diedrichs, C. Taffernaberry, and J. P. Peña, “Wireless wine: estimación de rendimiento y ubicación de sensores para la predicción de heladas en los viñedos,” in *Workshop de Investigadores en Ciencias de la Computación (WICC)*, 2020.
- [10] A. Herrera Jiménez, “Análisis y predicción de radiación en sistemas fotovoltaicos haciendo uso de machine learning,” 2023.

- [11] N. Gnoza Tansini and M. E. Barberena Allietti, “Estudio de factibilidad del uso de machine learning con múltiples fuentes de datos en el pronóstico del tiempo,” 2018.
- [12] G. Ojeda Suárez, “Cuantificación y predicción de lluvia a partir de información obtenida de radioenlaces de microondas y comunicaciones satelitales aplicando machine learning,” Master’s thesis, 2023.
- [13] M. Rivas Navarro, *Procesos cognitivos y aprendizaje significativo*. Comunidad de Madrid. Consejería de Educación. Viceconsejería de Organización . . . , 2008.
- [14] M. Ángel Herminda Núñez, “Las condiciones meteorológicas en la ruta la coruña-dublín y las maniobras en la estación trinquete del crucero internacional 2012; propuesta de empleo del velamen en las distintas condiciones de maniobra.” 2013. [Online]. Available: <http://repositorio.espe.edu.ec/handle/21000/8425>
- [15] I. G. Romero González, “Evaluación del efecto del cambio climático y la perdida de agrobiodiversidad en la comunidad de santa lucia, parroquia antonio josé holguín, cantón salcedo,” Master’s thesis, 2020.
- [16] J. G. Venegas Zapata *et al.*, “Análisis estadístico de datos meteorológicos mensuales y diarios en el periodo 2006-2018 para la determinación de variabilidad climática y cambio climático en el distrito metropolitano de quito,” Master’s thesis, Quito, EC: Universidad Andina Simón Bolívar, Sede Ecuador, 2020.

- [17] J. J. Madrigal Barrera, “Análisis de pronósticos climáticos para la previsión de sequías meteorológicas e hidrológicas,” Ph.D. dissertation, Universitat Politècnica de València, 2020.
- [18] E. R. Rivera and J. A. Amador, “Predicción estacional del clima en centroamérica mediante la reducción de escala dinámica. parte ii: aplicación del modelo mm5v3,” *Revista de Matemática: Teoría y Aplicaciones*, vol. 16, no. 1, pp. 76–104, 2009.
- [19] G. J. Aguado-Rodríguez, A. Quevedo-Nolasco, M. Castro-Popoca, R. Arteaga-Ramírez, M. A. Vázquez-Peña, and B. P. Zamora-Morales, “Predicción de variables meteorológicas por medio de modelos arima,” *Agrociencia*, vol. 50, pp. 1–13, 2016. [Online]. Available: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-31952016000100001&lng=es&nrm=iso
- [20] D. Villalón Pardo, “Crear y desarrollar una aplicación de alto rendimiento con bajo coste utilizando flutter y firebase,” Ph.D. dissertation, Universitat Politècnica de València, 2021.
- [21] J. J. C. Fernández, “Visualización, predicción y análisis de variables climáticas del océano atlántico mediante técnicas de aprendizaje automático,” Ph.D. dissertation, Universitat de València, 2020.
- [22] G. A. Albornoz Cabello, “Aplicación del aprendizaje automático supervisado en el mantenimiento predictivo de los motores eléctricos de inducción en las empresas mineras del Perú,” 2021.

- [23] G. M. Montelongo, “Pronóstico de la velocidad del viento aplicando un esquema de votación,” Master’s thesis, Instituto Tecnológico de Culiacán, Culiacán, Sinaloa, México, agosto 2018. [Online]. Available: <https://www.culiacan.tecnm.mx/wp-content/uploads/2020/07/Tesis-Giovanni-Manjarrez-Montelongo.pdf>
- [24] I. Components, “Dht11 vs. dht22: A comprehensive comparison of temperature and humidity sensors,” <https://www.ic-components.es/blog/dht11-vs-dht22-a-comprehensive-comparison-of-temperature-and-humidity-sensors.jsp>, accessed: 2025-01-05.
- [25] H. Libre, “Adafruit aht20 vs. dht22 vs. dht11: Comparación de sensores,” <https://www.hwlibre.com/adafruit-aht20-vs-dht22-vs-dht11-comparacion-de-sensores/>, accessed: 2025-01-05.
- [26] Nanoparuro, “Sensor de humedad y temperatura asair dht22,” <https://nanoparuro.com/shop/sensor-de-humedad-y-temperatura-asair-dht22-826>, accessed: 2025-01-05.
- [27] Components101, “Bmp180 atmospheric pressure sensor,” <https://components101.com/sensors/bmp180-atmospheric-pressure-sensor>, accessed: 2025-01-05.
- [28] B. Sensortec, “Humidity sensor bme280,” <https://www.bosch-sensortec.com/products/environmental-sensors/humidity-sensors-bme280/>, accessed: 2025-01-05.
- [29] —, “Pressure sensor bmp280,” <https://www.bosch-sensortec.com/products/environmental-sensors/pressure-sensors/bmp280/>, accessed: 2025-01-05.

- [30] T. Connectivity, “Ms5611-01ba03 barometric pressure sensor, with stainless steel cap,” https://www.te.com/commerce/DocumentDelivery/DDEController?Action=showdoc&DocId=Data+Sheet%7FMS5611-01BA03%7FB3%7Fpdf%7FEnglish%7FENG_DS_MS5611-01BA03_B3.pdf%7FCAT-BLPS0036, accessed: 2025-01-05.
- [31] STMicroelectronics, “Mems nano pressure sensor: 260-1260 hpa absolute digital output barometer lps22hb,” <https://www.st.com/resource/en/datasheet/lps22hb.pdf>, accessed: 2025-01-05.
- [32] Components101, “Bmp180 - atmospheric pressure sensor,” 2018, accessed: 2025-01-05. [Online]. Available: <https://components101.com/sensors/bmp180-atmospheric-pressure-sensor>
- [33] Alvesoaj, “Yl-83 + lm393 (rain sensor) with arduino,” 2013, accessed: 2025-01-05. [Online]. Available: <https://blog.alvesoaj.com/2013/01/yl-83-lm393-rain-sensor-with-arduino.html>
- [34] R. N. Tutorials, “Guide for rain sensor fc-37 or yl-83 with arduino,” 2018, accessed: 2025-01-05. [Online]. Available: <https://randomnerdtutorials.com/guide-for-rain-sensor-fc-37-or-yl-83-with-arduino/>
- [35] Ardumotive, “How to use the raindrops sensor module with arduino,” 2018, accessed: 2025-01-05. [Online]. Available: <https://www.ardumotive.com/how-to-use-the-raindrops-sensor-moduleen.html>

- [36] J. D. y A. Smith, “Comparative analysis of esp32 and raspberry pi for iot sensor data acquisition,” *International Journal of Embedded Systems*, vol. 12, no. 4, pp. 123–134, 2021, discusión sobre ventajas de microcontroladores para tareas en tiempo real.
- [37] Sekcompsci, “Comparison chips (socs) table for esp8266/esp32/esp32-s2/esp32-s3/esp32-c3/esp32-c6,” 2023, accessed: 2025-01-05. [Online]. Available: <https://gist.github.com/sekcompsci/2bf39e715d5fe47579fa184fa819f421>
- [38] S. Marsden, “Esp32 comparison chart,” 2023, accessed: 2025-01-05. [Online]. Available: <https://stevemarsden.com/electronics/esp32-comparison-chart/>
- [39] AranaCorp, “Descripción general del microcontrolador nodemcu esp32,” 2024, accessed: 2025-01-05. [Online]. Available: <https://www.aranacorp.com/es/descripcion-general-del-microcontrolador-nodemcu-esp32/>
- [40] K. L. B. Tapia and M. F. Q. Tomas, “Diseño y construcción de una tarjeta entrenadora basado en microcontroladores microchip 16fxxx y 18fxxx para optimizar el tiempo de práctica en los estudiantes de ingeniería de la universidad continental,” Master’s thesis, Universidad Continental, Huancayo, Perú, 2022. [Online]. Available: https://repositorio.continental.edu.pe/bitstream/20.500.12394/12309/2/IV_FIN_113_TE_Balbin_Quispe_2022.pdf
- [41] SocialCompare, “Raspberrypi models comparison,” 2024, accessed: 2025-01-05. [Online]. Available: <https://socialcompare.com/en/comparison/raspberrypi-models-comparison>

- [42] S. Wazir, A. A. Ikram, H. A. Imran, H. Ullah, A. J. Ikram, and M. Ehsan, "Performance comparison of mpich and mpi4py on raspberry pi-3b beowulf cluster," 2019, accessed: 2025-01-05. [Online]. Available: <https://arxiv.org/abs/1911.03709>
- [43] R. P. DK, "Raspberry pi 3 model b+," 2025, accessed: 2025-01-05. [Online]. Available: <https://raspberrypi.dk/en/product/raspberry-pi-3-model-b-plus-2/>
- [44] J. L. Gómez and R. Martínez, "Sistemas de bases de datos en tiempo real: Fundamentos y aplicaciones en entornos críticos," Master's thesis, Universidad Tecnológica de Innovación, Ciudad Futura, 2020, tesis de Maestría.
- [45] Google, "Firebase realtime database," 2025, accedido: 2025-02-03. [Online]. Available: <https://firebase.google.com/docs/database/>
- [46] I. MongoDB, "Mongodb features & key characteristics," 2025, accedido: 2025-02-03. [Online]. Available: <https://www.mongodb.com/resources/products/capabilities/features>
- [47] A. W. Services, "Amazon dynamodb features," 2025, accedido: 2025-02-03. [Online]. Available: <https://aws.amazon.com/dynamodb/features/>
- [48] NASA Langley Research Center (LaRC), "Nasa power data access viewer," 2025, accedido: 2024-09-03. [Online]. Available: <https://power.larc.nasa.gov/data-access-viewer/>
- [49] G. J. M. Murillo and J. A. M. Espinoza, "Implementación de un sistema de alerta temprana agroclimático ante la ocurrencia de enfermedades (moniliasis y mazorca negra) en las plantaciones de cacao en la provincia de los ríos - ecuador. fases (1) conocimiento

del riesgo, (2) y monitoreo y servicios de alerta,” Proyecto Integrador, Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador, 2021. [Online]. Available: <https://www.dspace.espol.edu.ec/bitstream/123456789/50615/1/T-109764%20Mendoza-Moran.pdf>

- [50] S. B. S. Santander, “Modelo de riego inteligente determinado a través de aprendizaje de máquina,” Master’s thesis, Universidad Técnica del Norte, Ibarra, Ecuador, 2024. [Online]. Available: <https://repositorio.utn.edu.ec/handle/123456789/15988>

Anexos

Anexo A

Repositorio del código de Arduino

Arduino código.

Anexo B

Repositorio de código Python para conexión serial y almacenamiento en tiempo real.

Código Python.

Anexo C

Conjunto de base de datos disponibles

Repositorio de bases de datos.

Anexo D

Algoritmos utilizados para el entrenamiento de modelos de Machine Learning.

Repositorio de modelos entrenados con “3 dimensiones y 4 etiquetas”.

Repositorio de modelos entrenados con “2 dimensiones y 4 etiquetas”.

Repositorio de modelos entrenados con “3 dimensiones y 2 etiquetas”.

Anexo E

Interfaz gráfica para la visualización de resultados.

Código de la interfaz en Python.
