

UNIVERSIDAD PERUANA LOS ANDES

Facultad de Ingeniería

Escuela Profesional de Ingeniería del Medio Ambiente y

Desarrollo



UPLA
UNIVERSIDAD PERUANA LOS ANDES

TESIS

**Desarrollo de modelos de inteligencia artificial para
el pronóstico de heladas en la cuenca del río
Mantaro, Chupaca, 2024.**

PARA OPTAR EL TÍTULO PROFESIONAL DE:
INGENIERO DEL MEDIO AMBIENTE Y DESARROLLO

Autor : Bach. Delbin Aldair Santiago Blancas
Bach. Trhiany Eloy Rojas Huari

Asesor : Mg. Carlos Enrique Álvarez Montalván

Código Orcid : <https://orcid.org/0000-0001-9110-8006>

Línea de investigación Institucional : Salud y Gestión de la Salud.

Fecha de inicio y Culminación : 15-10-2023 a 30-10-2024

HUANCAYO – PERÚ

2025

HOJA DE CONFORMIDAD DE JURADOS

DR. CARLOS ROSARIO SANCHEZ GUZMAN

Decano de la Facultad de Ingenieria

DR. ABRAHAMS MOISES CABRERA MONTALVO

Jurado

MTRO. IRALMY YIPSY PLATERO MOREJON

Jurado

MTRO. DUANY DAVILA HONORIO

Jurado

MTRO. LEONEL UNTIVEROS PEÑALOZA

Secretario Docente

DEDICATORIA

A Dios, por darme salud, paciencia y guiarme por el buen camino.

A mis padres, en especial a mi madre, Maribel Blancas Canturin, quien es mi fuente de inspiración y motivación.

A nuestros queridos asesores, por su confianza y apoyo.

Delbin Aldair Santiago Blancas

A Dios, por darme salud, sabiduría y guiarme siempre en el camino correcto.

A mi madre, Alina Huari, y a mi padre, Eloy Rojas, por ser mi inspiración y apoyo incondicional.

A mi hijo, Leo Rodrigo, por ser mi fuerza y motivo constante de superación.

Trhyani Eloy Rojas Huari

AGRADECIMIENTO

Agradecemos profundamente el apoyo recibido por cada una de las personas e instituciones que mencionaremos a continuación, Todos ellos hicieron posible el desarrollo de este proyecto, contribuyendo grandemente a nuestra formación profesional y crecimiento académico.

Al Dr.Sc. José Luis Flores Rojas y al PhD. José Antonio Reyes Suarez, a quienes rendimos un especial agradecimiento, por su enorme paciencia y tiempo dado.

Al Mg.Sc. Carlos Enrique Álvarez Montalván, por su guía en la revisión de este proyecto de investigación.

Agradecemos al Instituto Geofísico del Perú (IGP), por los datos brindados de las estaciones meteorológicas, radiómetros y sensores del Laboratorio de Microfísica Atmosférica y Radiación (LAMAR).

Así mismo, agradecemos al Servicio de Nacional de Hidrología y Meteorología del Perú (SENAMHI) por darnos la facilidad de acceder a los datos de las estaciones meteorológicas distribuidas en la Cuenca del Río Mantaro.

Finalmente agradecemos a la Universidad Peruana los Andes, por habernos formado con valores a lo largo de los cinco años de estudio.

CONSTANCIA DE SIMILITUD



NUEVOS TIEMPOS
NUEVOS DESAFÍOS
NUEVOS COMPROMISOS

CONSTANCIA DE SIMILITUD

N ° 0254 - FI -2025

La Oficina de Propiedad Intelectual y Publicaciones, hace constar mediante la presente, que la Tesis; titulado:

DESARROLLO DE MODELOS DE INTELIGENCIA ARTIFICIAL PARA EL PRONÓSTICO DE HELADAS EN LA CUENCA DEL RÍO MANTARO, CHUPACA, 2024

Con la siguiente información:

Con Autor(es) : **Bach. SANTIAGO BLANCAS DELBIN ALDAIR**
Bach. ROJAS HUARI TRHIANY ELOY

Facultad : **INGENIERÍA**

Escuela Académica : **INGENIERÍA DEL MEDIO AMBIENTE Y DESARROLLO**

Asesor(a) : **Mg. ALVAREZ MONTALVAN CARLOS ENRIQUE**

Fue analizado con fecha **19/06/2025**; con **119** págs.; con el software de prevención de plagio (Turnitin); y con la siguiente configuración:

Excluye Bibliografía.

X
X
X

Excluye citas.

Excluye Cadenas hasta 20 palabras.

Otro criterio (especificar)

El documento presenta un porcentaje de similitud de **6** %.

En tal sentido, de acuerdo a los criterios de porcentajes establecidos en el **artículo N°7** del Reglamento de Grados y Títulos, aprobado con Resolución de Consejo Universitario N° 425-2024-CU-UPLA. Se declara, que: **Si contiene un porcentaje aceptable de similitud.**

Observaciones:

En señal de conformidad y verificación se firma y sella la presente constancia.



Huancayo, 19 de junio del 2025.

MTRA. LIZET DORIELA MANTARI MINCAMI

JEFA

Oficina de Propiedad Intelectual y Publicaciones

CONTENIDO

CAPÍTULO I	15
1.1 Descripción de la realidad problemática	15
1.1.1 Espacial.....	16
1.1.2 Temporal.....	17
1.2 Formulación del problema.....	17
1.2.1 Problema general.....	17
1.2.2 Problemas específicos.....	17
1.3 Justificación	18
1.3.1 Social.....	18
1.3.2 Teórica.....	18
1.3.3 Metodológica	18
1.3.4 Práctica	19
1.4 Objetivos.....	19
1.4.1 Objetivo general.....	19
1.4.2 Objetivos específicos.....	19
1.5 Aspectos éticos de la Investigación	20
CAPÍTULO II	21
MARCO TEÓRICO.....	21
2.1 Bases teóricas o científicas	31
2.1.1 Marco conceptual.....	32
CAPÍTULO III.....	37
3.1 Hipótesis general.....	37
3.2 Hipótesis específicas.....	37
3.3 Variables.....	37
3.3.1 Definición conceptual.....	37
3.3.2 Definición Operacional de las variables.....	38
CAPÍTULO IV.....	39
4.1 Método de investigación.....	39
4.2 Tipo de investigación	39
4.3 Nivel de investigación	40

4.4	Diseño de la investigación.....	40
4.5	Población y muestra.....	40
4.5.1	Población.....	40
4.5.2	Muestra.....	40
4.6	Técnicas e instrumentos de recolección de datos.....	41
4.6.1	Técnicas de recolección de datos.....	41
4.6.2	Instrumentos de recolección de datos.....	42
4.7	Técnicas de procesamiento y análisis de datos.....	46
CAPITULO V.....		51
RESULTADOS.....		51
5.1	Desarrollar modelos de inteligencia artificial para predecir heladas en la cuenca del río Mantaro, Chupaca, 2024.....	51
5.2	Identificar las variables climáticas más relevantes para la predicción de heladas en la cuenca del río Mantaro, Chupaca, 2024.....	51
5.3	Analizar los datos históricos para identificar patrones que permitan el entrenamiento de un modelo de inteligencia artificial.....	68
5.4	Evaluar la precisión y confiabilidad de distintos modelos de inteligencia artificial en la predicción de heladas.....	70
5.4.1	Evaluación de la precisión y confiabilidad de modelos de IA en la predicción de heladas a partir de ventanas horarias.....	78
5.4.2	Evaluación experimental de la precisión y confiabilidad de modelos de IA en dos casos aplicativos.....	93
CONCLUSIONES.....		104
RECOMENDACIONES.....		105
REFERENCIAS BIBLIOGRÁFICA.....		106
ANEXOS.....		111
	Anexo 1: Matriz de Consistencia.....	111
	Anexo 2: Matriz de operacionalización de variables.....	113

CONTENIDO DE TABLAS

Tabla 1. Resultados de Pruebas de Normalidad.	59
Tabla 2. Resultados de pruebas de normalidad postratamiento.....	65
Tabla 3 Resultados del pronóstico de heladas en la fase entrenamiento.....	77
Tabla 4. Resultados de pronóstico en una ventana de 6 horas.	86
Tabla 5. Resultados de pronóstico de heladas en una ventana de 12 horas.	86
Tabla 6. Resultados de pronóstico de heladas en una ventana de 24 horas.	87
Tabla 7. Resultados de pronóstico de heladas en una ventana de 30 horas.	87
Tabla 8. Resultados de pronóstico de heladas en una ventana de 36 horas.	88
Tabla 9. Resultados de pronóstico de heladas en una ventana de 48 horas.	89

CONTENIDO DE FIGURAS

Figura 1 Mapa de ubicación general de las estaciones meteorológicas de la Cuenca del Mantaro.....	16
Figura 2 Metodología Fundamental de la Ciencia de Datos	44
Figura 3. Gráfico lineal de la temperatura mínima.	47
Figura 4. Lectura de archivos csv con datos meteorológicos.....	52
Figura 5. Estructura de los datos de Huayao_Meteo_2024.csv.....	52
Figura 6. Resumen estadístico de las variables meteorológicas.....	53
Figura 7. Esquema para la graficación lineal de datos.....	54
Figura 8. Presencia de datos faltantes en las variables de estudio.....	55
Figura 9. Presencia de datos atípicos en las variables.	56
Figura 10. Esquema de comprobación de normalidad.	58
Figura 11. Prueba de normalidad para las variables en estudio.	58
Figura 12. Verificación de calidad de datos	60
Figura 13. Verificación de valores faltantes.	61
Figura 14. Estrategia de imputación de datos faltantes.	62
Figura 15. Resultados tras la imputación de los valores faltantes.	62
Figura 16. Estrategia de tratamiento de valores atípicos.	63
Figura 17. Gráfico de cajas de las variables tratadas.....	64
Figura 18. Prueba de normalidad para las variables luego del pre-tratamiento.....	64
Figura 19. Correlación de Spearman entre Variables Meteorológicas y temp_min.....	67
Figura 20. Variables con mayor correlación con tem_min.....	67
Figura 21. Esquema de selección de datos y modelos de IA.....	69
Figura 22. Esquema de etapa de entrenamiento y evaluación de los modelos de IA.....	72
Figura 23. Gráfico de dispersión del modelo de regresión lineal.	72
Figura 24. Coeficiente de Gini y grafica de importancia de variables.....	73
Figura 25. Comparación de dispersión en los modelos de inteligencia artificial.....	74
Figura 26. Coeficiente de Gini e importancias de variables para los modelos de IA.....	76
Figura 27. Pronóstico de heladas con 6 horas de anticipación.....	80
Figura 28. Pronóstico de heladas con 12 horas de anticipación.....	81
Figura 29. Pronóstico de heladas con 24 horas de anticipación.....	82
Figura 30. Pronóstico de heladas con 30 horas de anticipación.....	83
Figura 31. Pronóstico de heladas con 36 horas de anticipación.....	84

Figura 32. Pronóstico de heladas con 48 horas de anticipación.....	85
Figura 33. Pronóstico de heladas con datos no vistos en el entrenamiento.	90
Figura 34. Pronóstico de heladas con datos no vistos en el entrenamiento.	91
Figura 35. Gráficos de dependencia parcial de las variables durante el pronóstico de heladas con el modelo Radom Forest.	91
Figura 36. Pronostico horario de heladas en un escenario de inversión térmica.	95
Figura 37. Curva de Lorenz y Coeficiente de importancia de Gini, experimento 01.	95
Figura 38. Gráfico de dependencia parcial de las variables, experimento 01.	96
Figura 39. Pronosticó de heladas durante un fenómeno simulado del niño, experimento 02.	98
Figura 40. Curva de Lorenz y coeficiente de Gini para el experimento 02.....	99
Figura 41. Gráficos de dependencia parcial para las variables en el experimento 02.....	99

RESUMEN

La presente investigación tiene el objetivo de mejorar la precisión de las predicciones y optimizar la gestión de riesgos en agricultura, ganadería y salud pública. Se empleará una metodología cuantitativa con enfoque aplicado y diseño experimental predictivo, utilizando modelos de IA como Random Forest, Extra Trees y Gradient Boosting. Se analizaron cinco años de datos climáticos históricos del Instituto Geofísico del Perú (IGP) y SENAMHI, incluyendo temperatura, humedad y radiación. Los resultados mostraron que Random Forest y Extra Trees fueron los modelos más precisos para predecir heladas en un horizonte de 6 a 48 horas. Se realizaron pruebas experimentales bajo escenarios de variabilidad climática, como inversión térmica y el fenómeno del Niño, para evaluar su desempeño. Se concluyó que la integración de redes neuronales profundas y la actualización constante de datos podrían mejorar aún más la precisión de las predicciones a largo plazo.

Palabras claves: inteligencia artificial, modelos predictivos, pronóstico de heladas, cuenca del río Mantaro.

ABSTRACT

The present research aims to improve the accuracy of predictions and optimize risk management in agriculture, livestock and public health. A quantitative methodology with applied approach and predictive experimental design will be employed, using AI models such as Random Forest, Extra Trees and Gradient Boosting. Five years of historical climate data from the Geophysical Institute of Peru (IGP) and SENAMHI, including temperature, humidity and radiation, were analyzed. The results showed that Random Forest and Extra Trees were the most accurate models for predicting frost over a 6 to 48 hour horizon. Experimental tests were conducted under climate variability scenarios, such as thermal inversion and El Niño, to evaluate their performance. It was concluded that the integration of deep neural networks and constant data updating could further improve the accuracy of long-term predictions.

Key words: artificial intelligence, predictive models, frost forecasting, Mantaro river basin.

INTRODUCCIÓN

Este proyecto de investigación tiene como objetivo el desarrollo de modelos de inteligencia artificial (IA) para el pronóstico de heladas en la cuenca del río Mantaro, Chupaca, 2024. El estudio se enfocó en la creación de modelos predictivos eficientes para prever las heladas tanto a corto plazo (<6 h) como a largo plazo (hasta 48 h), utilizando datos meteorológicos históricos del Instituto Geofísico del Perú (IGP) y el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI).

Los modelos tradicionales de predicción meteorológica permiten prever las temperaturas mínimas, pero su precisión a nivel local es limitada, ya que operan a una escala sinóptica y no son lo suficientemente específicos para regiones pequeñas (1). Esto provoca que las predicciones de heladas no sean confiables ni oportunas, lo cual afecta sectores como la agricultura y la salud pública. Según estudios previos, los modelos tradicionales no logran predecir heladas de manera puntual y precisa (2)

Este proyecto propone usar IA para generar predicciones más específicas y puntuales, mejorando la precisión de los pronósticos. La implementación de estos modelos beneficiará a sectores clave como la agricultura, ganadería, industria manufacturera y salud pública, optimizando la toma de decisiones y mejorando la gestión de riesgos hidrometeorológicos. Con esta herramienta, se podrán mitigar los efectos de las heladas y tomar decisiones operativas con mayor anticipación.

El proyecto de investigación está estructurado en cinco capítulos los cuales se detallan de la siguiente manera:

En el Capítulo I.- Describe el planteamiento del problema en función a la realidad problemática, así mismo presenta los objetivos que plantea la investigación y su

importancia. De la misma manera se describe el área de estudio que es la cuenca del río Mantaro.

En el Capítulo II.- Se abordan los principios teóricos sobre los métodos de inteligencia artificial, partiendo del análisis del problema que causa el impacto de las heladas en el valle del Mantaro que sustenta la implementación de predicciones a los fenómenos meteorológicos de heladas con el uso de métodos de aprendizaje automático.

En el Capítulo III.- Se evalúan métodos de inteligencia artificial desarrollados para la optimización de heladas en la cuenca del Río Mantaro, indicando su nivel de eficiencia y fallos, comparando los datos predichos con los datos reales. Así mismo se aplica los modelos de circulación atmosférica.

En el Capítulo IV.- Se da a conocer el análisis e interpretación de los resultados mostrados en los diagramas, mapas y gráficos respectivos, dando una descripción de los resultados obtenidos, analizando la data actual y futura de los eventos de heladas.

En el Capítulo V.- Para finalizar se resumen las conclusiones y se proponen recomendaciones, como resultado de la investigación.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción de la realidad problemática

La cuenca del río Mantaro es una región de vital importancia para la agricultura y la seguridad alimentaria en Perú, siendo la principal fuente de abastecimiento agropecuario para ciudades clave como Lima. Esta área contribuye con un 18% del Producto Bruto Interno del país (3). Sin embargo, la región se enfrenta a un creciente desafío económico y social debido a las heladas, un fenómeno climático que ocurre cuando las temperaturas descienden por debajo de los 0° (4). Las heladas en la región suelen presentarse entre abril y agosto, y su intensidad aumenta con la altitud en algunas zonas del valle (5).

Este fenómeno se ha visto agravado por el cambio climático, que ha incrementado la frecuencia y severidad de los eventos climáticos extremos, haciendo que las heladas sean cada vez más impredecibles tanto a nivel local como global (6). Además, la falta de un pronóstico preciso y oportuno de heladas dificulta la implementación de medidas preventivas, lo que causa pérdidas económicas en la agricultura y la industria, además de afectar el empleo en la cadena de suministro (2, 7).

Uno de los síntomas más evidentes de este fenómeno es la destrucción de cultivos, solamente en el periodo 1997-1998, durante el Fenómeno El Niño, se registraron 204,000 hectáreas afectadas por sequías, heladas y friajes, con pérdidas en cultivos como papa, maíz y cebada (8). Este tipo de eventos pone de manifiesto la urgente necesidad de contar con un sistema de alerta temprana para pronosticar las heladas con precisión y oportunidad, minimizando así los impactos económicos y sociales.

Para abordar esta problemática, es crucial desarrollar modelos predictivos con IA y datos meteorológicos históricos para generar pronósticos locales precisos, permitiendo anticipar eventos y tomar medidas preventivas.

1.1.1 Espacial

El estudio abarcó la cuenca del río Mantaro, ubicada en los Andes centrales del Perú, entre las latitudes $10^{\circ}33'52.66''$ y $13^{\circ}32'31.39''$ S, y las longitudes $73^{\circ}55'10.88''$ y $76^{\circ}39'16.01''$ O.

La ubicación política del área de estudio comprendió el departamento de Junín en la zona central, limitando al norte con los departamentos de Pasco y Lima, y al sur con Huancavelica y Ayacucho.

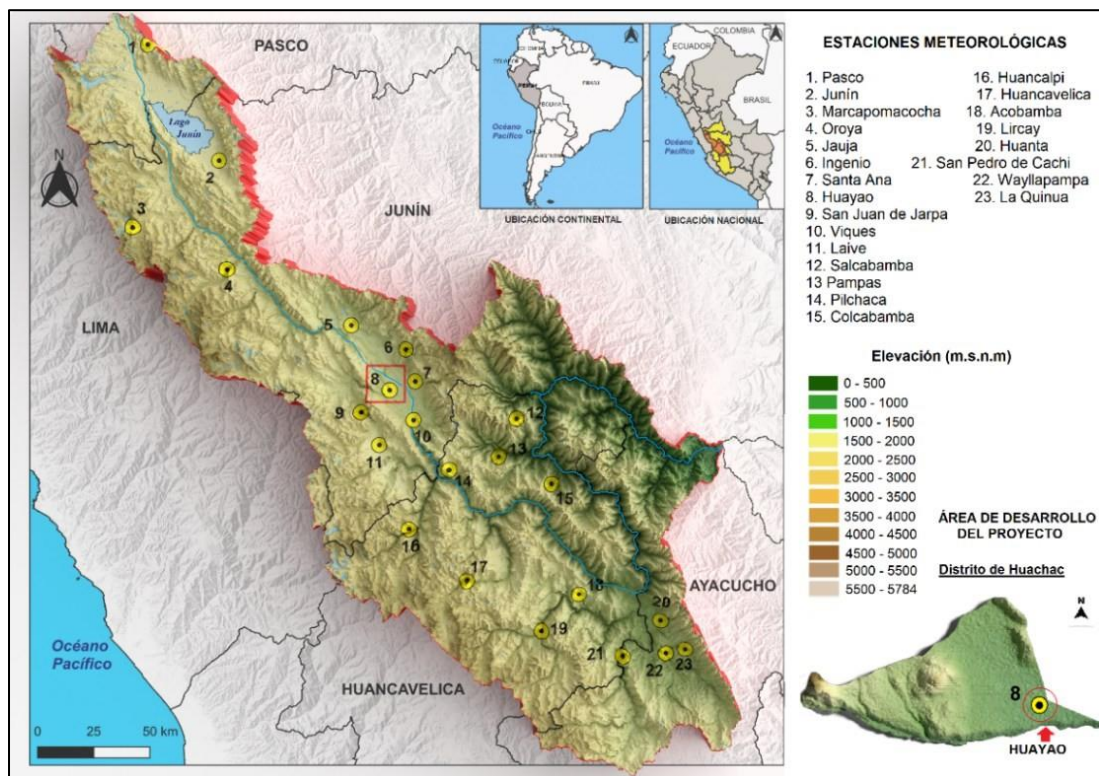


Figura 1. Mapa de ubicación general de las estaciones meteorológicas de la cuenca del río Mantaro.

1.1.2 Temporal

El estudio se desarrolló en dos etapas a lo largo de un año, desde octubre de 2023 hasta octubre de 2024. Durante la primera semana de octubre de 2023, se gestionó las autorizaciones correspondientes y se llevó a cabo un diagnóstico preliminar. Posteriormente, en la segunda etapa, se recopilaron y procesaron los datos para el desarrollo y la evaluación de modelos de inteligencia artificial.

1.2 Formulación del problema

Las heladas en la cuenca del río Mantaro afectan la producción agrícola y la economía regional, especialmente en zonas por encima de los 3,200 m.s.n.m., donde las temperaturas descienden bajo 0 °C en invierno. Este fenómeno reduce el rendimiento de cultivos clave como papa, maíz y quinua (5). Las estrategias de mitigación actuales, como quemas controladas y riego nocturno, carecen de respaldo científico y son poco efectivas (9). Por ello es crucial desarrollar soluciones basadas en evidencia científica para reducir su impacto socioeconómico.

1.2.1 Problema general

¿Cómo desarrollar modelos de inteligencia artificial para predecir las heladas en la cuenca del río Mantaro, Chupaca, 2024?

1.2.2 Problemas específicos

- A.** ¿Cuáles son las variables climáticas más relevantes para la predicción de heladas en la cuenca del río Mantaro?
- B.** ¿Cómo procesar y analizar los datos históricos para identificar patrones que permitan entrenar un modelo de inteligencia artificial?
- C.** ¿Qué modelo de inteligencia artificial ofrecerá mayor precisión y confiabilidad en la predicción de heladas en comparación con los métodos tradicionales

1.3 Justificación

1.3.1 Social

El desarrollo de modelos de inteligencia artificial para predecir heladas en la cuenca del río Mantaro permitirá a los agricultores proteger sus cultivos y reducir pérdidas económicas mediante pronósticos anticipados. Esto facilitará la aplicación de métodos de mitigación como calefacción, riego tecnificado y máquinas de viento. Además, la investigación puede aplicarse a otros eventos climáticos extremos, mejorando la resiliencia ante el cambio climático y generando beneficios significativos a nivel agrícola y social.

1.3.2 Teórica

Esta investigación contribuye a fortalecer las capacidades científico-tecnológicas de las ciencias atmosféricas en la región al optimizar la precisión de los pronósticos climáticos. Mediante la integración de inteligencia artificial con datos meteorológicos, mejorará la detección temprana de heladas, facilitando la toma de decisiones en sectores vulnerables como la agricultura y promoverá el uso de tecnologías innovadoras en el estudio de eventos climáticos extremos.

1.3.3 Metodológica

Se basa en la metodología de la ciencia de datos para recopilar, procesar y analizar información climática con fines predictivos. A través de algoritmos de aprendizaje automático, se desarrollarán modelos de inteligencia artificial que optimicen el pronóstico de heladas. Además, los resultados servirán como referencia para futuras investigaciones en la aplicación de estas tecnologías a eventos climáticos extremos, fortaleciendo las ciencias atmosféricas en la región.

1.3.4 Práctica

La aplicación de modelos de inteligencia artificial para la predicción de heladas en la cuenca del río Mantaro tendrá un impacto positivo en el sector agrícola al optimizar la toma de decisiones y reducir las pérdidas en la producción. A nivel ambiental, permitirá mitigar el uso de prácticas perjudiciales como las quemadas controladas, fomentando estrategias sostenibles de adaptación al cambio climático. Además, esta tecnología puede integrarse en sistemas de monitoreo climático, mejorando la gestión del riesgo y la planificación territorial en la región.

1.4 Objetivos

1.4.1 Objetivo general

Desarrollar modelos de inteligencia artificial para predecir heladas en la cuenca del río Mantaro

1.4.2 Objetivos específicos

- A.** Identificar las variables climáticas más relevantes para la predicción de heladas en la cuenca del río Mantaro, Chupaca, 2024.
- B.** Analizar los datos históricos para identificar patrones que permitan el entrenamiento de un modelo de inteligencia artificial.
- C.** Evaluar la precisión y confiabilidad de distintos modelos de inteligencia artificial en la predicción de heladas, comparándolos con los métodos tradicionales.

1.5 Aspectos éticos de la Investigación

- **Consentimiento de la información.** - Se coordinó con el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) y el Instituto Geofísico del Perú (IGP), entidades responsables de los datos climáticos en la región, para obtener la autorización necesaria para el uso de sus registros. En esta comunicación, se detallaron los objetivos de la investigación y los beneficios que aportará al monitoreo climático y la gestión del riesgo de heladas en la cuenca del río Mantaro.
- **Fuente de información.** - La información para el proyecto se recopiló de repositorios académicos de prestigio, como Elsevier, Springer y ScienceDirect, analizando antecedentes y metodologías relevantes. Los documentos cumplen con las normas ISO 690 y los requisitos institucionales para referencias bibliográficas. Además, se verificó que el índice de similitud se mantenga por debajo del 30 %, asegurando la originalidad y el cumplimiento de principios éticos.

CAPÍTULO II

MARCO TEÓRICO

La cuenca del río Mantaro es una zona altamente vulnerable a las heladas debido a su clima variable y la reducción progresiva de las precipitaciones. Históricamente, la predicción de estos eventos se basó en métodos empíricos y observaciones tradicionales de los agricultores, lo que limitaba la capacidad de respuesta. A partir de 1922, entidades como el IGP comenzaron a monitorear el clima en la región Junino, registrando un aumento en la frecuencia e intensidad de las heladas, así como cambios en la temperatura y la humedad, lo que subrayó la necesidad de herramientas predictivas más precisas (10).

Para abordar esta problemática, en 2009 el IGP impulsó el proyecto MAREMEX, que combinó datos históricos con modelos predictivos avanzados para fortalecer la capacidad de adaptación ante eventos climáticos extremos (1). Posteriormente en 2021 el SENAMHI elaboró escenarios climáticos que proyectaban una mayor severidad de las heladas hacia 2050, con un descenso de hasta -0.95 °C por década en zonas como Jauja (11).

Actualmente, los avances en inteligencia artificial han revolucionado la predicción climática, permitiendo procesar grandes volúmenes de datos y mejorar la precisión de los pronósticos. La implementación de estos modelos en la cuenca del Mantaro representa un paso clave en la modernización del monitoreo climático, facilitando la toma de decisiones informadas y reduciendo el impacto de las heladas en la agricultura y la economía local. En el futuro, la combinación de inteligencia artificial con sistemas de alerta temprana optimizará aún más la gestión del riesgo climático en la región Antecedentes (12).

a. Antecedentes Internacionales

En el artículo científico; Pronóstico de heladas mejorado mediante métodos de aprendizaje automático, presentado por Rozante et al. (2023) en la Universidad de Sao Paulo, Brasil.

El objetivo del estudio fue desarrollar una red neuronal artificial (RNA) ajustada a variables meteorológicas para mejorar la predicción de heladas en el sur de Brasil. La metodología empleó dos redes *Multilayer Perceptron* (MLP) con optimizadores ADAM y SGD, entrenadas con datos del modelo Eta (2012-2015) y validadas en 2016 y 2017. Se evaluó su desempeño comparándolas con el índice de heladas del INPE y el modelo Eta, considerando métricas como la tasa de aciertos y la reducción de falsos negativos. Los resultados mostraron que las RNA reducen significativamente la subestimación de heladas del modelo Eta y mejoran la detección de eventos meteorológicos adversos. Además, se observó que ambos optimizadores presentaron un rendimiento similar en la predicción. La conclusión general es que las RNA optimizan la predicción de heladas, superando a modelos numéricos tradicionales, y representan una herramienta útil para la mitigación de daños agrícolas y la toma de decisiones en zonas vulnerables. (13).

En el artículo científico; Predicción de heladas mediante aprendizaje automático y redes neuronales profundas, publicado por Talsma et al. (2023) en la revista *Frontiers in Artificial Intelligence*, en colaboración con la Universidad de Texas.

El fin de esta investigación fue mejorar la predicción de heladas en zonas agrícolas mediante modelos de aprendizaje automático y redes neuronales profundas. La metodología empleada se basó en datos meteorológicos históricos de alcalde, Nuevo

México, donde se entrenaron modelos *Random Forest* (RF) y *Deep Neural Networks* (DNN) para pronósticos de hasta 48 horas. Su desempeño se comparó con el sistema HRRR, obteniendo un RMSE de 1.53-1.72°C en predicciones a 6 horas. Los resultados mostraron que la temperatura del suelo es clave en pronósticos a más de 24 horas, mientras que otras variables son más relevantes en el corto plazo. Además, los modelos de regresión térmica fueron más precisos que los de clasificación para evaluar el impacto de las heladas en los cultivos. Los resultados obtenidos permitieron concluir que estos modelos mejoran la predicción de heladas, pero su precisión depende de grandes volúmenes de datos específicos de cada región, lo que limita su aplicabilidad sin un entrenamiento adecuado. (14).

En la tesis; Predicción temprana de heladas en cultivos de altura, empleando métodos de aprendizaje de máquinas, sustentada por Calderón (2022) en la Universidad Nacional de Colombia para optar al título de Magíster en Ingeniería – Analítica.

El propósito fundamental de este proyecto fue el desarrollo de un modelo de predicción temprana de heladas para la región de Cundinamarca en Colombia, utilizando variables climáticas y aprendizaje de máquinas. La metodológica aplicó un enfoque cuantitativo, explicativo y experimental, con un muestreo intencional, en trece estaciones meteorológicas de nueve municipios. Se analizaron variables como temperatura, humedad y radiación fotosintéticamente activa. Así mismo se evaluaron cinco modelos de aprendizaje supervisado, destacando el modelo de potenciación del gradiente, que superó el 91 % de precisión. Además, los modelos de temperatura mínima presentaron errores menores a 0.55 °C en una ventana de una hora. Los resultados confirmaron que el aprendizaje automático mejora la predicción de

heladas y puede mitigar su impacto en la agricultura. Finalmente se recomienda ampliar la base de datos y considerar más variables climáticas para optimizar la precisión del modelo en diversas regiones (15).

En el artículo científico; Metodología inteligente de predicción de heladas basada en interpolación espacial utilizando redes neuronales artificiales con datos locales limitados, presentado por Zhou et al. (2022) en la Universidad Tecnológica de Sydney, Australia.

El estudio desarrolló un método de predicción de heladas basado en interpolación espacial y redes neuronales artificiales sin depender de datos históricos locales. Se usó un enfoque cuantitativo y experimental, analizando datos de 75 estaciones meteorológicas en Nueva Gales del Sur y el Territorio de la Capital Australiana. La muestra se seleccionó considerando factores geográficos y climáticos para representar distintas altitudes y variabilidad térmica. Se implementó una Red Neuronal Artificial Multicapa (MLP) con interpolación espacial para estimar temperaturas mínimas en sitios sin datos previos. Se evaluaron los modelos con validación cruzada, destacando el promediado ponderado, que logró una precisión del 92.55 % en la detección de heladas. Los resultados mostraron que el modelo es eficaz en áreas con infraestructura meteorológica limitada. Se recomienda integrar datos satelitales de MODIS y Sentinel-3 y aplicar redes neuronales convolucionales (CNN) y recurrentes (RNN) para mejorar su precisión y adaptabilidad. (16).

En el artículo científico; Previsión de heladas: una práctica de aprendizaje automático a partir de datos, publicado por Ding et al. (2021) en la revista International Journal of Reasoning-based Intelligent Systems, en colaboración con la Facultad de Ciencia y Tecnología de la Universidad de Meiji, Tokio, Japón

El propósito de este estudio fue diseñar un modelo de Máquinas de Vectores de Soporte para la Regresión (SVR) para predecir heladas, mejorando la precisión respecto a los métodos tradicionales. La metodología consistió en un análisis comparativo entre modelos convencionales y de aprendizaje automático, utilizando datos meteorológicos del campus Ikuta de la Universidad de Meiji. La población incluyó datos históricos, y la muestra consistió en registros de temperatura mínima, humedad relativa y velocidad del viento. Se optimizó el modelo mediante calibración de hiperparámetros, logrando un error promedio de 2.7°C, inferior al 3.5°C de los métodos tradicionales. Además, los modelos SVM con probabilidades mejoraron la tasa de recuperación, aunque con menor precisión general. Se concluyó que estos modelos pueden integrarse en sistemas de alerta temprana, beneficiando a la agricultura y sectores vulnerables. (17).

En el artículo científico; Estimación del riesgo de heladas para el árbol del té en la provincia de Zhejiang basado en aprendizaje automático, presentado por (18) en la Universidad Normal del Noreste, China.

El estudio desarrolló un modelo de predicción de heladas en plantaciones de té en la provincia de Zhejiang, China, utilizando técnicas de aprendizaje automático. Se aplicó un enfoque cuantitativo y experimental, empleando datos meteorológicos y topográficos de 47 estaciones meteorológicas de la región. La población incluyó todas las estaciones meteorológicas disponibles, y la muestra consistió en 320 puntos de desastre clasificados según el nivel de daño. Se implementaron modelos de Máquinas de Soporte Vectorial (SVM) y Redes Neuronales Artificiales (ANN) para clasificar y predecir la distribución espacial de las heladas. El modelo SVM obtuvo la mayor precisión (83.8%), superando a ANN (75%). Los resultados indicaron que

la humedad relativa y la velocidad del viento fueron los factores más influyentes en la predicción de heladas. Se concluyó que el modelo SVM es adecuado para predecir eventos de heladas en zonas de cultivo de té y puede ser útil en sistemas de alerta temprana (18).

En el artículo científico Desarrollo de un pronóstico probabilístico de temperaturas bajo cero en carretera mediante aprendizaje automático, presentado por (19) en la Universidad de Oklahoma, Estados Unidos.

La investigación desarrolló un modelo probabilístico para predecir temperaturas subcero en carreteras de EE.UU., utilizando Random Forest (RF), SVM y Redes Neuronales Artificiales (ANN) para optimizar la seguridad vial y la gestión del tráfico. Se empleó un enfoque cuantitativo y aplicado, con un diseño no experimental y longitudinal. La muestra incluyó datos del modelo HRRR y observaciones RWIS (2016-2018), con registros históricos de al menos 5 años. Se usaron 15 variables meteorológicas y geográficas, logrando un AUC de 0.96 (96%) y un Brier Skill Score (BSS) de 0.66 (66%), con una precisión de 96% a 2 horas, 90% a 18 horas y 85% a 36 horas. La combinación de modelos mejoró la precisión en 5-10%, redujo el error medio en 15% y los falsos positivos en 20%. Su implementación permitió disminuir en 10% los accidentes y reducir en 15% el uso de sal en carreteras. Se concluyó que Machine Learning optimiza la predicción de temperaturas viales y se recomienda aumentar la red de sensores e integrar datos de vehículos conectados y explorar modelos de aprendizaje profundo para mejorar la precisión en distintas regiones climáticas (19).

En el artículo científico; Previsión de heladas mediante aprendizaje automático - De la asociación a la causalidad", publicado por Ding, Noborio y Shibuya

(2020) en la revista Agricultural and Forest Meteorology (Elsevier) y con mención a la Facultad de Ciencia y Tecnología de la Universidad de Meji, Tokio, Japón

El estudio desarrolló un modelo de predicción de heladas basado en Máquinas de Soporte Vectorial para Regresión (SVR), utilizando variables meteorológicas como temperatura, humedad relativa, radiación y velocidad del viento. Se aplicó un enfoque cuantitativo y experimental, con datos de sensores en el campus Ikuta de la Universidad de Meiji, registrados en intervalos de un minuto. La población incluyó todas las mediciones meteorológicas, y la muestra se seleccionó según representatividad climática. Se emplearon técnicas de suavizado de datos y promediado móvil para identificar patrones en las variables atmosféricas. Los resultados indicaron que la temperatura es el predictor más relevante, mientras que la humedad y la radiación influyen en la sensibilidad del modelo. Se concluyó que esta metodología mejora la detección temprana de heladas y puede aplicarse en sistemas de alerta temprana. Se recomienda desarrollar una plataforma de monitoreo en tiempo real que integre este modelo con sensores remotos para mejorar la precisión y la respuesta ante eventos climáticos adversos (20).

b. Antecedentes Nacionales

En el trabajo de suficiencia profesional "Olas de frío en el Perú", presentado por (21) para la Universidad Nacional Agraria La Molina, Lima, Perú, para optar por el título profesional de Ingeniero Meteorólogo.

El informe tuvo como objetivo determinar la frecuencia y tendencias de las olas de frío en el Perú entre 1965 y 2019. Se empleó un enfoque observacional y retrospectivo, basado en registros históricos de temperatura mínima de 369

estaciones meteorológicas del SENAMHI, con una muestra seleccionada mediante muestreo no probabilístico por conveniencia. Para asegurar la precisión, los datos fueron sometidos a control de calidad y homogeneización. Las tendencias fueron analizadas mediante los índices CSDI y ECF, evaluando su comportamiento temporal y distribución geográfica. Los resultados indicaron que estos eventos son más frecuentes en la sierra y duran entre 1 y 8 días, sin una tendencia definida en la mayoría de estaciones. Se concluyó que la caracterización de estos fenómenos es clave para la gestión del riesgo climático. Se recomienda fortalecer la red de monitoreo meteorológico y mejorar los modelos de predicción para reducir el impacto en poblaciones vulnerables (21).

En la tesis de pregrado; Modelo predictivo agroclimático para detectar cambios climatológicos con Machine Learning e IoT para el sector agrario de la Región Puno, presentada por (22) para la Universidad Peruana Unión, Puno, Perú, para optar por el título profesional de Ingeniero de Sistemas.

El propósito de esta investigación fue implementar un modelo predictivo agroclimático con Machine Learning e IoT para detectar cambios climatológicos en Puno. Se empleó la metodología CRISP-DM con un enfoque cuantitativo, tecnológico y longitudinal. La muestra se seleccionó mediante muestreo no probabilístico por conveniencia, utilizando datos históricos del SENAMHI y datos en tiempo real de una mini estación meteorológica (ESP-32, DHT22, BMP180). Se probaron algoritmos en Azure Machine Learning Studio, destacando Decision Tree Classifier con 97% de precisión y un RMSE de 0.35, mientras que la Regresión Logística obtuvo 95% de precisión y un RMSE de 0.42. Los resultados fueron validados con datos reales para evaluar la fiabilidad del modelo. Se demostró que

IoT y Machine Learning mejoran la predicción agroclimática y optimizan la toma de decisiones agrícolas. Se recomienda mejorar la infraestructura de sensores y explorar modelos más avanzados, como redes neuronales recurrentes (RNN), para mayor precisión (22).

En la tesis; Determinación de modelos de estimación de temperaturas mínimas con fines de pronóstico de heladas en la Región Ancash, presentada por (23) para la Universidad Agraria La Molina, Lima, Perú, para optar por el título de Ingeniero Meteorólogo.

El estudio tuvo como finalidad desarrollar modelos empíricos para estimar temperaturas mínimas y pronosticar heladas en Ancash. Se aplicó una metodología cuantitativa y aplicada, con un diseño no experimental y longitudinal. La muestra se determinó mediante muestreo no probabilístico por conveniencia, utilizando datos del modelo BRAMS y registros históricos de SENAMHI entre 2013 y 2017. Se usaron técnicas estadísticas como la correlación de Pearson y regresión lineal simple, obteniendo ecuaciones empíricas con coeficientes de correlación de hasta 0.95. La validación se realizó con RMSE y BIAS, destacando la estación Malvas con la mejor precisión en los cuatro tiempos de pronóstico. Además, se evaluó el impacto económico de las heladas en la agricultura mediante el modelo Cost-Loss, demostrando que la prevención reduce pérdidas hasta en 33%. Se concluyó que los modelos empíricos pueden predecir heladas con alta confiabilidad y mitigar daños agrícolas. Se recomienda mejorar la cobertura de estaciones meteorológicas y evaluar modelos más avanzados para optimizar las predicciones (23).

En la tesis; Modelo predictivo Machine Learning aplicado a análisis de datos Hidrometeorológicos para un SAT en Represas, presentada por (24) en la

Universidad Tecnológica del Perú, Arequipa, Perú, para optar por el título de Ingeniero de Sistemas e Informática

El estudio desarrolló un modelo predictivo con Machine Learning para mejorar el análisis de datos hidrometeorológicos en un Sistema de Alerta Temprana (SAT) de represas. Se empleó un enfoque cuantitativo y aplicado, con diseño no experimental y longitudinal. La muestra provino de datos históricos y sensores. Se implementó una red neuronal LSTM, entrenada con caudal efluente, temperatura, precipitación y evaporación, obteniendo un RMSE de 1.30 y una precisión del 94% en la predicción de descargas. El modelo redujo el error promedio en 15% respecto a métodos tradicionales y permitió anticipar eventos extremos con un 85% de confiabilidad. Se verificó su efectividad en diferentes escenarios hidrometeorológicos, demostrando su robustez. Se concluyó que LSTM optimiza la gestión hídrica y mejora la toma de decisiones en alertas tempranas. Se recomienda ampliar la red de sensores, incorporar más datos climáticos y evaluar modelos como GRU o híbridos para mejorar la precisión y adaptabilidad del sistema (24).

En el artículo científico; Ciclos estacionales y diurnos del límite superficial, estrato y balance energético en los Andes centrales de Perú, Valle del Mantaro, publicado en la revista científica Atmosphere MDPI y presentado en el Instituto Geofísico del Perú, Lima, Perú, (25).

La investigación analizó los ciclos estacionales y diurnos de la capa límite superficial y el balance de energía en el Valle del Mantaro, Perú. Se usó un enfoque cuantitativo y observacional, con diseño no experimental y longitudinal. La muestra incluyó datos meteorológicos del Observatorio de Huancayo (2018-2019) obtenidos con sensores en torres de gradiente y radiación, tensiómetros y placas de flujo de calor. Se

aplicaron los métodos aerodinámicos y de balance de energía Bowen, obteniendo correlaciones de $R^2 = 0.8$ entre calor sensible y latente, y $R^2 = 0.86$ entre energía disponible y flujos de calor. Se evidenció inversión de vapor de agua en invierno, condiciones áridas en la estación seca y circulación montaña-valle en otoño-invierno. El flujo de calor sensible fue mayor en junio y septiembre, mientras que el latente predominó en febrero y marzo. Se concluyó que las interacciones suelo-atmósfera afectan la estabilidad atmosférica y la dinámica climática local. Se recomienda mejorar la resolución espacial de datos, validar con modelos numéricos y ampliar la red de sensores (25).

2.1 Bases teóricas o científicas

Teoría de la investigación

- **Teoría del Aprendizaje Automático (Machine Learning)}**

La teoría del aprendizaje automático sostiene que los sistemas informáticos pueden mejorar su desempeño mediante la identificación de patrones en los datos y la generación de modelos predictivos sin necesidad de programación explícita. Se basa en la capacidad de los algoritmos para aprender de la experiencia y ajustar sus predicciones en función de la información disponible. Esta teoría es fundamental en campos donde las relaciones entre variables son complejas y no lineales, permitiendo abordar problemas mediante técnicas como redes neuronales, máquinas de soporte vectorial y modelos de ensamble. Su aplicación en diversas áreas ha demostrado que los modelos pueden optimizarse progresivamente a medida que procesan más datos, mejorando la capacidad de análisis y toma de decisiones en entornos dinámicos e inciertos (26).

- **Teoría de sistemas complejos**

La teoría de sistemas complejos estudia cómo la interacción de múltiples elementos dentro de un sistema puede dar lugar a comportamientos emergentes que no pueden explicarse únicamente a partir de sus componentes individuales. Se basa en principios como la no linealidad, la retroalimentación y la sensibilidad a las condiciones iniciales, lo que implica que pequeñas variaciones pueden generar grandes cambios en la evolución del sistema. Esta teoría se aplica en diversos campos donde los sistemas presentan estructuras dinámicas y adaptativas, caracterizadas por interacciones interdependientes entre sus partes. Su estudio permite comprender fenómenos caóticos y optimizar modelos de análisis en entornos con alta incertidumbre (27)

2.1.1 Marco conceptual

a. Modelos de Inteligencia artificial

Los modelos de IA son algoritmos y representaciones matemáticas diseñados para simular el razonamiento humano en sistemas informáticos. Estos modelos permiten a las máquinas aprender, adaptarse y realizar tareas como la toma de decisiones, el reconocimiento de patrones y la clasificación (28).

b. Tipos de Modelos de Inteligencia artificial

Dentro de los modelos de IA, algunos de los más comunes para tareas de pronóstico incluyen los métodos de ensamblaje de árboles de decisión y boosting.

Random Forest: Este modelo es un ensamblaje de múltiples árboles de decisión entrenados con diferentes subconjuntos de los datos y es muy efectivo para manejar grandes volúmenes de datos con alta dimensionalidad (29). Random

Forest es especialmente útil para tareas de clasificación y regresión en las que se requiere robustez y precisión.

Extra Trees: Similar a Random Forest, el modelo Extra Trees también es un conjunto de árboles de decisión, pero con diferencias en la forma en que se seleccionan las divisiones de los árboles. Este modelo tiende a ser más rápido y menos propenso al sobreajuste (30).

Decision Tree: Los árboles de decisión son modelos predictivos de tipo jerárquico, en los cuales se realizan divisiones sucesivas del espacio de entrada en función de la variable que más reduce la impureza del modelo (31). Aunque simples, los árboles de decisión pueden ser muy efectivos si se manejan adecuadamente.

AdaBoost: El modelo AdaBoost (Adaptive Boosting) es una técnica de ensamble que combina modelos débiles, como árboles de decisión simples, para formar un clasificador fuerte (32). Este modelo es particularmente efectivo para mejorar el rendimiento de modelos básicos en tareas de clasificación.

Gradient Boosting: A diferencia de AdaBoost, el Gradient Boosting construye modelos secuenciales en los que cada nuevo modelo intenta corregir los errores del modelo anterior (33). Es muy eficiente en tareas de regresión y clasificación y es menos sensible al sobreajuste cuando se configura correctamente.

XGBoost: Este modelo es una versión optimizada de Gradient Boosting que utiliza regularización para reducir el sobreajuste y mejorar la precisión del modelo (33). XGBoost ha demostrado ser altamente efectivo en una amplia

gama de tareas de predicción y es conocido por su velocidad y rendimiento en competiciones de aprendizaje automático.

c. Entrenamiento y Validación de los Modelos

El proceso de entrenamiento consiste en ajustar los parámetros del modelo utilizando un conjunto de datos de entrenamiento, mientras que la validación evalúa el modelo en un conjunto de datos independiente para asegurarse de que no haya sobreajuste (Liu & Yang, 2005). En el caso de los modelos de boosting y ensamblaje, se utilizan técnicas como la validación cruzada para garantizar que el modelo generalice bien a datos no vistos (35).

d. Fuentes de Datos Utilizadas

Los modelos de IA se alimentan de datos climáticos que pueden provenir de diversas fuentes, como estaciones meteorológicas, satélites o modelos numéricos de predicción climática. Las variables más comunes incluyen la temperatura (mínima, máxima, media), humedad relativa, presión atmosférica, viento y radiación solar (36).

e. Pronóstico de Heladas

El pronóstico de heladas tiene como objetivo predecir la ocurrencia de eventos de heladas, los cuales ocurren cuando la temperatura desciende por debajo de los 0°C, lo que puede afectar a la agricultura y otros sectores (Rozante et al., 2023). La precisión y fiabilidad de estas predicciones dependen en gran medida de las variables climáticas que se consideren y de la capacidad del modelo para generalizar patrones a partir de los datos históricos.

f. **Exactitud y Fiabilidad del Pronóstico**

La exactitud se refiere a la capacidad del modelo para predecir correctamente los eventos de heladas, lo que se mide utilizando métricas como el Error Absoluto Medio (MAE) o el Raíz del Error Cuadrático Medio (RMSE) (38). La fiabilidad está relacionada con la consistencia del modelo a lo largo del tiempo, especialmente en condiciones meteorológicas cambiantes.

g. **Variables Climáticas Analizadas**

Para predecir las heladas, se analizan varias variables climáticas:

Temperatura: Es la cantidad de energía térmica presente en el aire en un instante específico. Se mide con termómetros, generalmente en grados Celsius (°C), y es el factor principal que influye en la percepción de calor o frío (39).

Es la variable principal en la predicción de heladas, ya que las heladas ocurren cuando la temperatura desciende por debajo del punto de congelación (40). La temperatura mínima durante la noche es crucial para determinar si se formarán heladas.

Humedad relativa: La humedad relativa se define como el porcentaje de vapor de agua presente en el aire en comparación con la cantidad máxima que podría contener a una temperatura específica, su unidad de medida se expresa en % (39).

Juega un papel importante en la formación de escarcha y heladas, ya que influye en la cantidad de vapor de agua en el aire y en la rapidez con que el aire se enfría. Los niveles elevados de humedad pueden hacer que las heladas se formen más rápidamente (41).

Viento: El viento se define como el desplazamiento del aire en la atmósfera, moviéndose desde áreas de alta presión hacia regiones de baja presión, su unidad de medidas es metros por segundo (m/s) (39).

El viento tiene un efecto importante en la disipación del calor, lo que afecta directamente las temperaturas mínimas. El viento fuerte puede impedir la acumulación de aire frío cerca del suelo, lo que hace menos probable que se produzcan heladas (Daniel S. Wilks, 2011).

Radiación solar: Es la energía solar que se transmite en forma de radiación electromagnética y alcanza la atmósfera terrestre. Su medición se realiza en una superficie horizontal mediante un sensor de radiación o piranómetro, ubicado en un área sin obstrucciones y orientado hacia el sur. Se expresa en vatios por metro cuadrado (W/m^2) (39). Su análisis nos permite demostrar la cantidad de energía que se retiene en la atmósfera durante la noche (18)

Horizonte Temporal del Pronóstico

El horizonte de predicción se define como el tiempo transcurrido desde el momento en que se realiza una predicción hasta el momento en que se espera que aparezca el valor previsto (42).

El horizonte temporal del pronóstico puede clasificarse en tres tipos:

Corto Plazo: Pronósticos para el mismo día o los siguientes días (hasta 72 horas).

Mediano Plazo: Pronósticos que abarcan una semana o dos.

Largo Plazo: Pronósticos para periodos más largos, como meses, años e incluso con menor precisión (42).

CAPÍTULO III

HIPÓTESIS

3.1 Hipótesis general

El desarrollo de modelos de inteligencia artificial permite predecir heladas en la cuenca del río Mantaro, Chupaca, 2024.

3.2 Hipótesis específicas

- A. La identificación de las variables climáticas permite optimizar la predicción de heladas en la cuenca del río Mantaro, Chupaca, 2024, mejorando la efectividad del modelo de inteligencia artificial.
- B. El análisis de datos históricos permite la identificación de patrones climáticos clave, proporcionando una base sólida para el entrenamiento de un modelo de inteligencia artificial en la predicción de heladas.
- C. La evaluación de la precisión y confiabilidad de distintos modelos de inteligencia artificial permite identificar el modelo más adecuado para la predicción de heladas en la cuenca del río Mantaro, Chupaca, 2024.

3.3 Variables

3.3.1 *Definición conceptual*

Variable independiente (X).- Modelos de inteligencia artificial.- Según (28), los modelos de inteligencia artificial son representaciones matemáticas y computacionales que, basándose en datos y supuestos, permiten la toma de decisiones y la predicción en diversos contextos (28)

Variable dependiente (Y).- Pronóstico de helada.- Tal como señala (43), el pronóstico de heladas con modelos de simulación es una de las técnicas que permite conocer cuando suceden heladas. Está basada en los modelos que describen la evolución de la temperatura en horas de la noche simulando las transferencias de energía de manera más o menos compleja (43).

3.3.2 Definición Operacional de las variables

Variables independientes (X). - Modelos de inteligencia artificial

Un modelo de inteligencia artificial para el pronóstico de heladas es un algoritmo que, a partir de un conjunto de datos de entrada, como la temperatura, la humedad, la velocidad del viento y la cobertura de nubes, predice si se producirá una helada en un lugar y momento determinados.

Variable dependiente (Y). - Pronóstico de heladas

El pronóstico de heladas es una variable binaria que toma el valor 1 si se produce una helada y 0 si no se produce.

CAPÍTULO IV

METODOLOGÍA

4.1 Método de investigación

Según Hernández Sampieri (44) el método científico tiene un carácter sistemático y riguroso para generar conocimiento científico, este método se basa en la observación, formulación de hipótesis, experimentación y análisis de datos, lo que asegura la obtención de explicaciones verificables sobre los fenómenos estudiados. Además, el método se sustenta en la medición y el criterio empírico, que permiten la recolección de datos cuantificables y la validación de resultados a través de observaciones directas (45). Este método garantiza que las conclusiones sean objetivas, basadas en evidencia sólida y razonamiento lógico (46), lo que en este estudio permite evaluar de manera precisa los modelos de inteligencia artificial para prever heladas.

4.2 Tipo de investigación

El presente trabajo se enmarca dentro de la investigación aplicada, cuyo objetivo es resolver problemas prácticos inmediatos mediante la identificación de soluciones concretas (44). En este caso, la investigación busca desarrollar algoritmos de inteligencia artificial para prever las heladas, ofreciendo una solución práctica para mitigar los daños causados por este fenómeno. La investigación aplicada se distingue por su enfoque en la implementación de soluciones derivadas de la investigación básica, con el fin de abordar problemas específicos y proporcionar herramientas útiles para la sociedad. Por lo tanto, este estudio se orienta a la predicción de heladas utilizando tecnologías avanzadas para mejorar la toma de decisiones en los sectores afectados.

4.3 Nivel de investigación

La presente investigación será de nivel Predictivo, ya que involucrará el análisis de series temporales de datos meteorológicos y la estimación de valores futuros a través de técnicas de inteligencia artificial (47).

4.4 Diseño de la investigación

La presente investigación utiliza un diseño experimental verdadero, en el que se manipula la variable independiente (modelos de inteligencia artificial) para analizar su impacto sobre la variable dependiente (pronóstico de heladas), controlando factores que puedan influir en los resultados (48), este diseño permitirá establecer relaciones causales al manipular variables y controlar factores externos. En este estudio, se asegurará la calidad de los datos meteorológicos, la correcta ubicación de las estaciones de medición y la consideración de cambios climáticos, garantizando que las predicciones reflejen con precisión la efectividad de los modelos utilizados.

4.5 Población y muestra

4.5.1 Población

La población de estudio está conformada por 22 estaciones meteorológicas ubicadas en la cuenca del río Mantaro que registran variables climáticas relevantes para la predicción de heladas, tales como temperatura, humedad y presión atmosférica, durante el período 2017-2023.

4.5.2 Muestra

El presente estudio utiliza un muestreo no probabilístico por conveniencia, ya que la selección de la muestra no se hizo de manera aleatoria, sino en función de la

disponibilidad y calidad de los datos. Según Hernández Sampieri (44), este tipo de muestreo será apropiado cuando la elección de los elementos se basara en criterios específicos del investigador, como la accesibilidad y relevancia de la información para el estudio. En este caso, se seleccionó la estación meteorológica de Huayao del IGP como punto principal de análisis debido a su ubicación estratégica, continuidad, precisión y disponibilidad de sus registros de temperatura entre 2017 y 2023.

4.6 Técnicas e instrumentos de recolección de datos

4.6.1 Técnicas de recolección de datos

Revisión de registros

Según Hernández Sampieri (44) es un método de recolección de datos que consiste en el análisis de documentos existentes, como bases de datos meteorológico, para ello se realizó lo siguiente:

- Selección de registros meteorológicos del SENAMHI.
- Recolección de información. - Extracción de datos históricos sobre variables clave.

Una vez obtenido los datos estos se guardaron en un archivo tipo CSV llamado “**Datos_Met.csv**”, el cual se preproceso en el programa Python.

Observación de campo

Los datos fueron obtenidos por los equipos del Laboratorio de Microfísica Atmosférica y Radiación (LAMAR) del IGP, los cuales registran mediciones de variables climáticas y observaciones de campo. Estos equipos capturan una serie de variables climáticas esenciales para el estudio:

- Presión atmosférica
- Temperatura máxima y mínima del ambiente
- Temperatura promedio del ambiente
- Humedad relativa
- Radiación
- Velocidad del viento (a los 2 m. y a los 10 m.)
- Dirección del viento (a los 2 m. y a los 10 m.)

Los datos recopilados son almacenados en un archivo CSV denominado "**Huayao_Meteo_2024.csv**" y posteriormente preprocesados utilizando el software Python.

4.6.2 Instrumentos de recolección de datos

- Entrevistas a especialistas. - Validación de datos meteorológicos con expertos.

Los datos meteorológicos fueron recopilados mediante los instrumentos y sensores del Laboratorio de Microfísica Atmosférica y Radiación (LAMAR), ubicado en el Observatorio de Huayao y gestionado por el Instituto Geofísico del Perú (IGP). Estos equipos cuentan con una resolución temporal de 1 minuto, registrando información en un Datalogger para su posterior análisis. LAMAR dispone de diversos sensores diseñados para estudios atmosféricos y validación de modelos numéricos en los Andes centrales del Perú.

Los principales instrumentos empleados en esta investigación fueron:

- Torre de gradiente meteorológico con sensores instalados en seis niveles de altura (2, 6, 12, 18, 24 y 29 m), permitiendo la medición de variables climáticas a diferentes altitudes.
- Anemómetro 3002 Wind Sentry Set de Campbell Scientific, compuesto por un sensor de tres copas y una veleta, utilizado para medir la velocidad y dirección del viento.
- Tensiómetros DECAGON 5 TM VWC, empleados para medir la temperatura y humedad del suelo con una precisión de $\pm 3\%$ en humedad y $\pm 1\text{ }^{\circ}\text{C}$ en temperatura.
- Piranómetro KIPP & ZONEN, diseñado para medir la irradiación solar global y difusa, con una sensibilidad de 5 a 15 $(\mu\text{V}/\text{W})/\text{m}^2$.
- Radiómetro infrarrojo, encargado de medir la temperatura y la energía térmica infrarroja emitida por las superficies dentro de su campo de visión.
- Pirgeómetros, utilizados para registrar la radiación de onda larga en la atmósfera, en un rango espectral de 4500 a 42000 nm.

Para el procesamiento computacional, se empleó una laptop con procesador Intel Core i7 de 1.8 GHz, 8 GB de RAM y disco duro de 931 GB, optimizando el análisis y modelado de los datos meteorológicos recopilados.

Procedimiento

Para el procesamiento de la información, se aplicó la Metodología Fundamental para la Ciencia de Datos, creada por IBM, la cual sigue un enfoque estructurado de diez etapas desde la identificación del problema hasta la retroalimentación del modelo. Este ciclo de vida de la ciencia de datos está estrechamente relacionado con la

inteligencia artificial y el machine learning, y es similar al modelo CRISP-DM (Cross Industry Standard Process-for Data Mining), ampliamente utilizado en minería de datos y modelado predictivo. Esta metodología se presenta el siguiente flujograma (Figura 2):

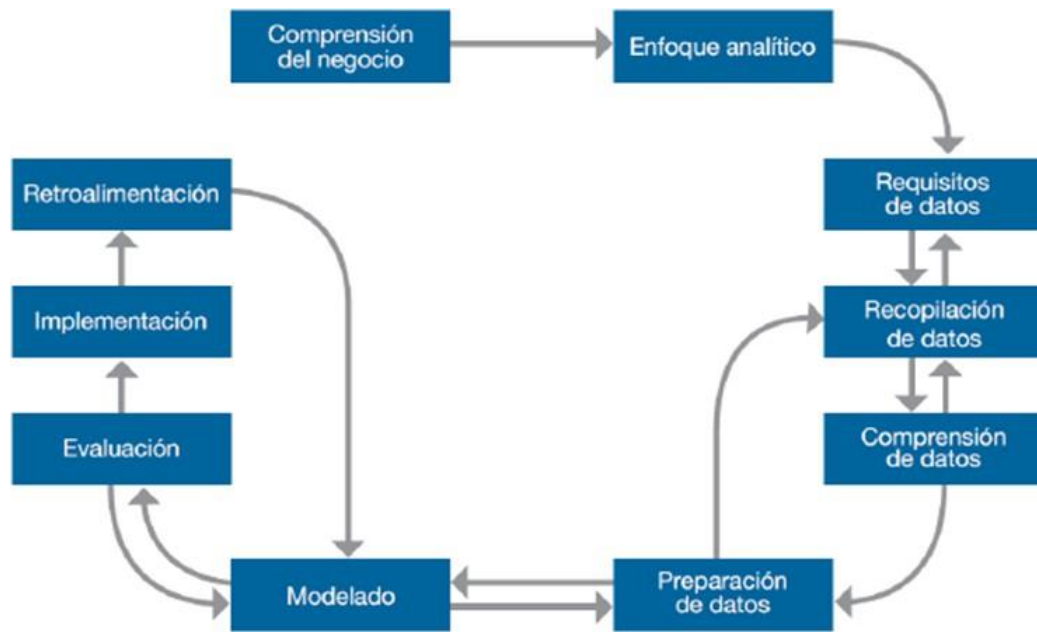


Figura 2. Metodología Fundamental de la Ciencia de Datos (IBM, 2023).

El proceso se desarrolló en las siguientes fases:

a) Fase de planteamiento y organización

- **Solicitud de permisos**

Para acceder a los datos meteorológicos del SENAMHI, se presentó una solicitud formal a la Gerente General, especificando los objetivos de la investigación y la necesidad de contar con registros históricos y actuales. La solicitud se ingresó la mesa de partes y fue evaluada para determinar su viabilidad. Una vez aprobada, la información fue proporcionada mediante el envío de archivos al correo institucional. En el caso del Instituto Geofísico del

Perú (IGP), se aprovechó el convenio institucional existente con la universidad para acceder a los datos de sus estaciones meteorológicas pertenecientes al Laboratorio de Microfísica Atmosférica y Radiación (LAMAR). Para ello, se envió un correo formal al director del Observatorio de Huayao solicitando el acceso a la información relevante para la investigación.

- **Compresión del negocio o necesidad (Revisión bibliográfica)**

En esta fase, se realizó un análisis exhaustivo de fuentes científicas y técnicas para comprender la problemática de las heladas en el Valle del Mantaro. Se revisaron artículos científicos, reportes meteorológicos, boletines especializados y estudios previos que abordarán el impacto, la frecuencia y los factores asociados a este fenómeno climático. Esta revisión permitió identificar las principales variables a considerar en el desarrollo de modelos de inteligencia artificial para el pronóstico de heladas, así como las metodologías más utilizadas en investigaciones similares.

b) Fase de Recopilación de datos

- **Enfoque analítico**

Se definió una estrategia metodológica basada en modelos de aprendizaje automático para el pronóstico de heladas en el valle del río Mantaro, utilizando algoritmos supervisados como árboles de decisión, bosques aleatorios, Gradient Boosting y redes neuronales artificiales. Para determinar qué modelo de inteligencia artificial constituye una herramienta clave en el pronóstico de heladas, dependiendo del uso de datos meteorológicos históricos de fuentes oficiales y confiables para garantizar la calidad del entrenamiento y la precisión de las predicciones.

- **Requisito de Datos**

Para garantizar la adecuación de los datos para el modelado de IA, se estableció requisitos clave, incluyendo la identificación de variables meteorológicas relevantes como temperatura, humedad, presión atmosférica, viento y nubosidad. Además, se definió la necesidad de registros en escalas temporales diversas (diaria, horaria) para optimizar la comparación de patrones climáticos. La cobertura geográfica se limitó a la cuenca del río Mantaro, con énfasis en el distrito de Chupaca, para asegurar una mejor representatividad local. Finalmente, se procuró que los datos sean de alta calidad, sin inconsistencias, valores atípicos ni información faltante, garantizando su utilidad en el modelado y entrenamiento de los algoritmos.

- **Recopilación de datos**

Los datos serán obtenidos de las plataformas oficiales del SENAMHI y los registros del Observatorio de Huayao – IGP. - Para ello se empleó los datos de los dataloggers del Laboratorio de Microfísica Atmosférica y Radiación (LAMAR) del Instituto Geofísico del Perú (IGP).

4.7 Técnicas de procesamiento y análisis de datos

c) Fase de pre procesamiento

- **Comprensión de los datos**

En esta etapa, se llevó a cabo un análisis exploratorio de los datos recolectados para identificar patrones, inconsistencias y características clave que puedan influir en el modelo de inteligencia artificial. Se examinaron las variables climáticas registradas y la estructura de los datos, Además, se analizaron

valores atípicos, datos faltantes y la distribución de las variables para determinar la calidad de la información.

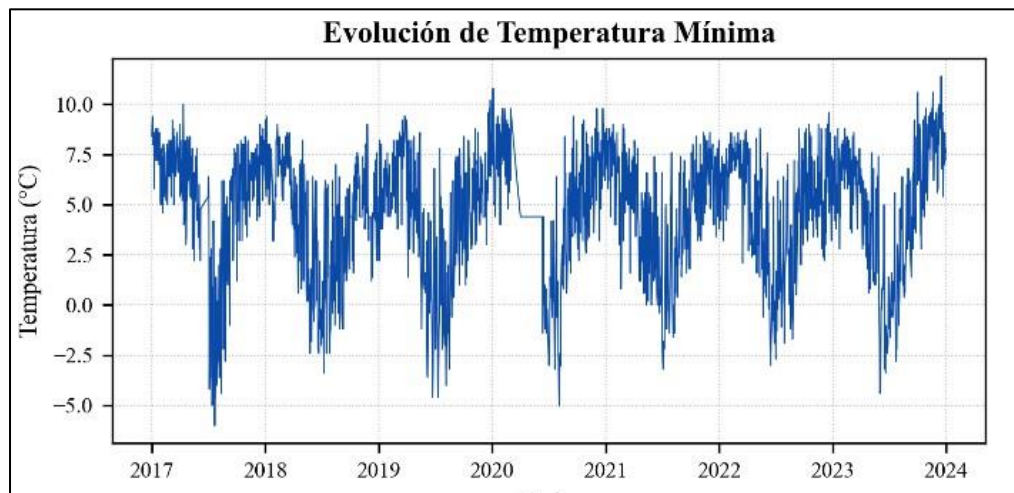


Figura 3. Gráfico lineal de la temperatura mínima.

- **Preprocesamiento de los datos**

En esta fase se aplicaron técnicas de procesamiento para limpiar, transformar y estructurar los datos de manera óptima antes de su uso en los modelos predictivos. Entre las técnicas que se emplearon se incluyen:

Estadística descriptiva. - Se aplicarán medidas de tendencia central (media, mediana, moda) y dispersión (varianza, desviación estándar) para comprender la distribución de las variables climáticas. En los que además se corrobora la hipótesis nula H_0 : Los datos siguen una distribución normal.

Detección y manejo de datos atípicos. - Se aplicaron métodos como la detección de valores extremos mediante la prueba de Tukey o el análisis de percentiles.

Imputación de datos faltantes. - Se emplearon métodos como interpolación, media móvil y técnicas basadas en Machine Learning para completar registros incompletos.

Normalización y escalado de datos. - Se transformaron las variables a un rango uniforme utilizando técnicas como Min-Max Scaling, con el fin de mejorar el desempeño de los modelos.

Análisis de correlación:

Una vez que los datos se encuentran procesados y normalizados, se realizó un análisis de correlación para determinar las relaciones entre las variables. Este análisis es fundamental para identificar posibles dependencias lineales o monótonas entre las variables, lo que permitió seleccionar las variables más relevantes para el pronóstico de heladas.

Para ello, se utilizaron dos métodos de correlación:

Correlación de Pearson

La correlación de Pearson mide la relación lineal entre dos variables cuantitativas. Este método asume que la relación entre las variables es lineal y calcula el grado de dependencia entre ellas. La fórmula matemática para calcular la correlación de Pearson es la siguiente:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

Donde:

r: Coeficiente de correlación de Pearson.

X_i y Y_i : Valores individuales de las variables X y Y

\bar{X} y \bar{Y} : Medias de las variables X y Y .

Correlación de Spearman

La correlación de Spearman mide la relación monótona entre dos variables, es decir, la relación en la que una variable siempre aumenta o disminuye a medida que la otra cambia, pero no necesariamente de manera lineal. Es útil cuando las variables no siguen una distribución lineal. La fórmula para calcular la correlación de Spearman es la siguiente:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

ρ : Coeficiente de correlación de Spearman.

d_i : Diferencia entre los rangos de las dos variables para cada observación.

n : Número total de observaciones.

d) Fase de Modelado y Evaluación

- **Modelado**

En esta etapa, se seleccionaron y entrenaron los modelos de inteligencia artificial para el pronóstico de heladas. Se definió los algoritmos más adecuados en función de la naturaleza de los datos y el objetivo de la investigación, utilizando técnicas de aprendizaje supervisado. Entre los modelos que se consideraran están:

Bosque Aleatorio (Random Forest) y Extra Trees. - Modelos basados en múltiples árboles de decisión que mejoran la precisión mediante aleatorización en la selección de atributos y divisiones.

Gradient Boosting y XGBoost. - Algoritmos que optimizan el error en cada iteración utilizando el descenso de gradiente y técnicas de regularización para evitar el sobreajuste.

Redes Neuronales Artificiales (RNA).- Modelos basados en capas de neuronas interconectadas que aprenden patrones no lineales en los datos.

El proceso de modelado implicó dividir el conjunto de datos en entrenamiento y prueba para evaluar la capacidad predictiva de los modelos. Se implementaron técnicas de ajuste de hiperparámetros, como Grid Search y Random Search, para optimizar el rendimiento de los algoritmos.

- **Evaluación**

Una vez entrenados los modelos, se realizó una evaluación exhaustiva para medir su desempeño y capacidad de generalización. Se empleó métricas de error y precisión como:

Error Cuadrático Medio (RMSE) y Error Absoluto Medio (MAE). - Indicadores de precisión que cuantificaron la diferencia entre los valores predichos y los valores reales.

CAPITULO V

RESULTADOS

5.1 Desarrollar modelos de inteligencia artificial para predecir heladas en la cuenca del río Mantaro, Chupaca, 2024.

1. Compresión del negocio o necesidad, ‘Mediante la metodología fundamental de la ciencia de datos’

La cuenca del río Mantaro es clave para la agricultura y economía del Perú, aportando el 18% del PBI y abasteciendo a ciudades como Lima (49). Sin embargo, las heladas, que ocurren entre junio y agosto, representan una amenaza creciente, intensificada por el cambio climático, lo que dificulta su pronóstico (5). La falta de alertas oportunas genera grandes pérdidas económicas, como en 2006, cuando se destruyeron 14,400 hectáreas de cultivos, con pérdidas de 27 millones de soles (10). En ese sentido un sistema de alerta temprana permitiría a los agricultores y sectores afectados tomar medidas preventivas, minimizando daños. Para ello, el desarrollo de modelos predictivos con inteligencia artificial (IA) basados en datos históricos es crucial para anticipar estos eventos con precisión, reduciendo el impacto económico y social en la región.

5.2 Identificar las variables climáticas más relevantes para la predicción de heladas en la cuenca del río Mantaro, Chupaca, 2024.

Compresión de los datos

Para analizar los datos, se aplicó el lenguaje de programación Python, que ofrece diversas bibliotecas para la manipulación y procesamiento de datos. El archivo "Huayao_Meteo_2024.csv", fue cargado en Python mediante la biblioteca pandas, lo que permitió su exploración y análisis de manera eficiente.

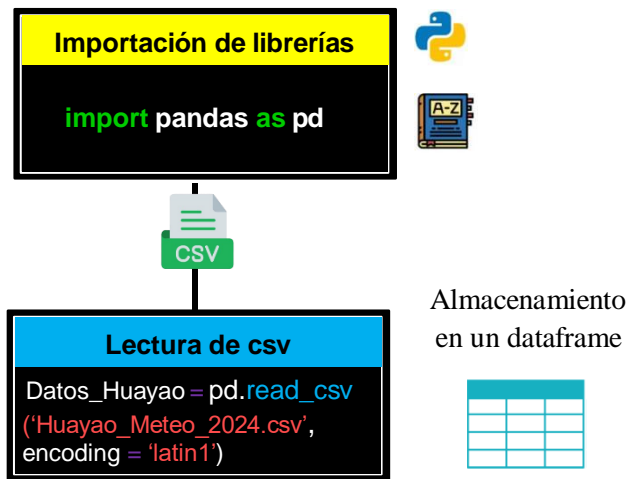


Figura 4. Lectura de archivos csv con datos meteorológicos.

En la figura 4, se aprecia el empleo de la biblioteca pandas para cargar el archivo "Huayao_Meteo_2024.csv" mediante la función `pd.read_csv`. Asignándole un nuevo nombre "Datos_Huayao" el cual es un dataframe. Luego, con el módulo `info()`, se obtuvo la información detallada sobre la estructura de los datos de las variables.

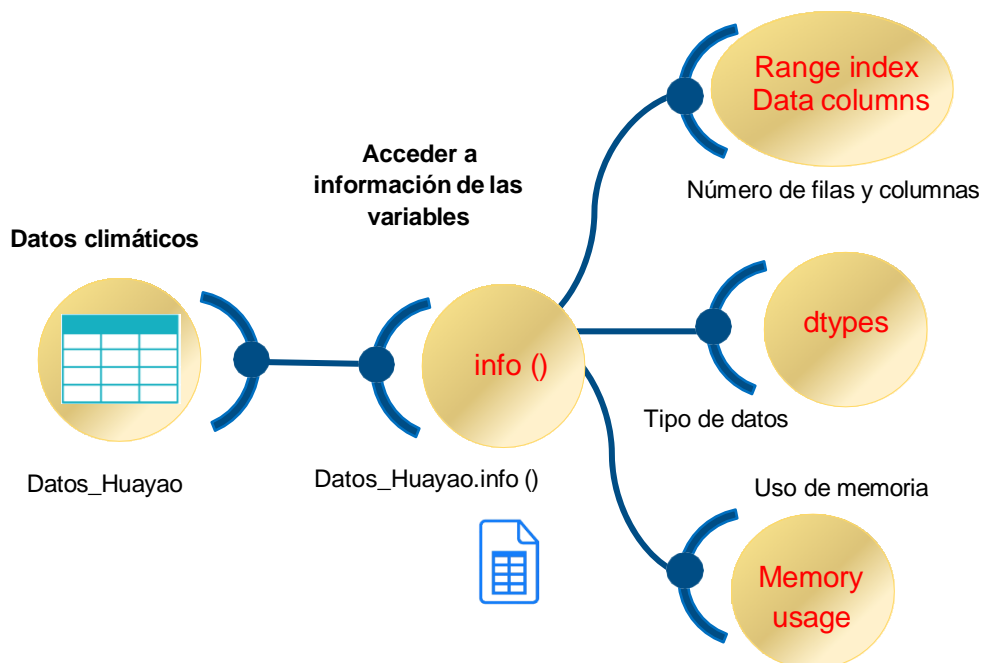


Figura 5. Estructura de los datos de Huayao_Meteo_2024.csv.

En la figura 5, se puede apreciar que el dataframe Datos_Huayao contiene 1780 filas y 15 columnas, con variables de temperatura (mínima y máxima), humedad relativa, temperatura superficial del suelo, dirección media del viento, presión atmosférica, velocidad del viento (máxima y media) y evapotranspiración a 2 metros del suelo y usa un formato object para fechas y estaciones. Este archivo no tiene valores nulos, pero ocupa menos memoria (208.7 KB). En general este archivo es necesario para el análisis de datos meteorológicos.

Posteriormente con el módulo describe se pudo visualizar el resumen estadístico de los datos, tal y como la muestra la figura 6.

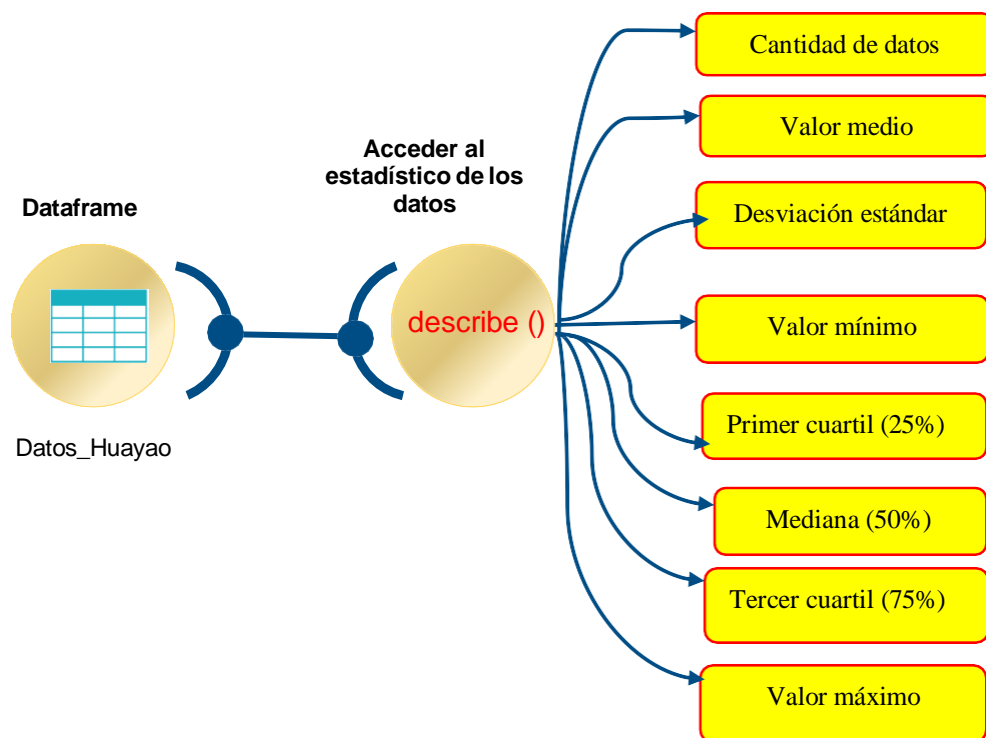


Figura 6. Resumen estadístico de las variables meteorológicas.

En la figura 6, se puede evidenciar el análisis del dataframe Datos_Huayao, el cual se realizó utilizando la función. describe () de pandas, que genera un resumen estadístico de las variables numéricas en los conjuntos de datos. La función proporcionó estadísticas clave como el promedio de la temperatura máxima

(19.17°C), la temperatura mínima (3.14°C), la Humedad Relativa (56.94%). Este resumen estadístico brinda una comprensión completa de las características principales de las variables meteorológicas, facilitando su interpretación y análisis.

Para analizar la evolución temporal de las variables clave, se creó una columna "Fecha" combinando las columnas 'Año', 'Mes' y 'Día' utilizando la función `pd.to_datetime()`, lo que permitió convertir estos componentes en un formato adecuado para su análisis. Posteriormente, se asignó la columna "Fecha" como índice del DataFrame con `data.set_index('Fecha', inplace=True)`, ordenando los datos cronológicamente. Finalmente, se seleccionaron las columnas numéricas correspondientes a las variables de interés y se generaron gráficos de líneas individuales para cada variable usando la librería `matplotlib`, lo que permitió visualizar su evolución a lo largo del tiempo. Dicho procedimiento se puede visualizar en la figura 7.

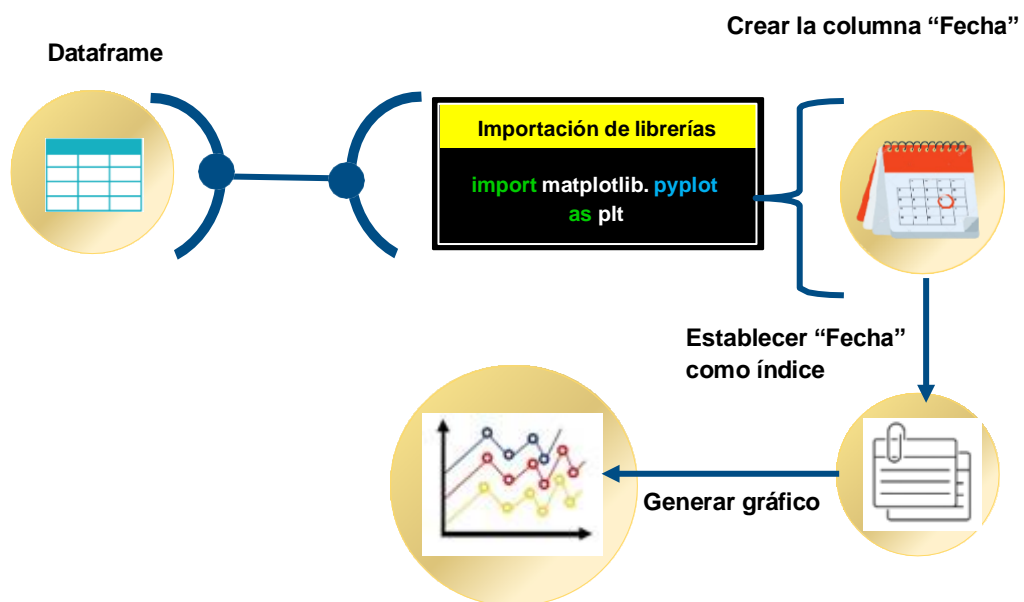


Figura 7. Esquema para la graficación lineal de datos.

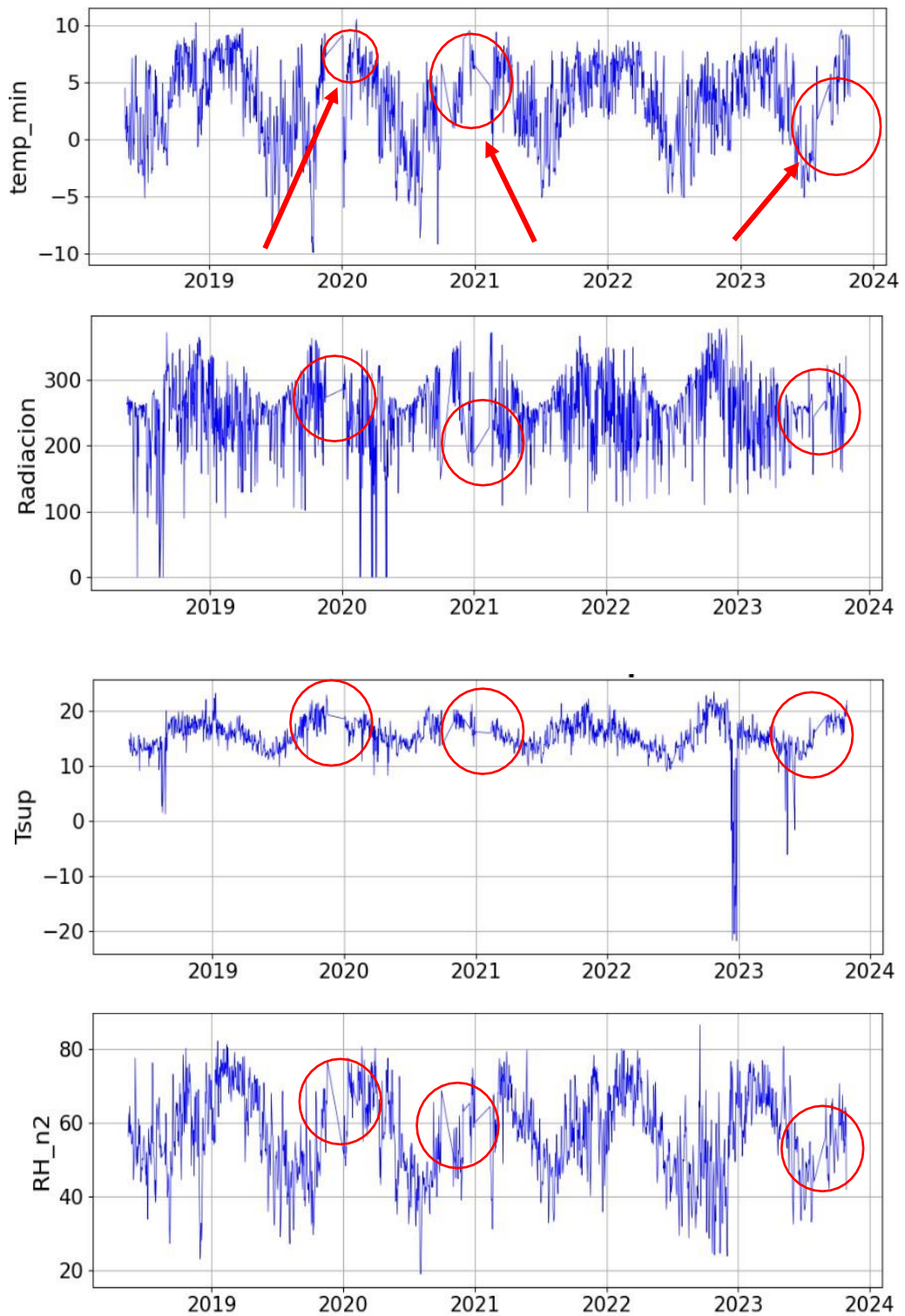


Figura 8. Presencia de datos faltantes en las variables de estudio.

En la figura 8, se observa que las gráficas generadas tienen saltos o vacíos en los datos, los cuales se indican mediante flechas rojas. Estos saltos corresponden a datos faltantes que deben ser gestionados adecuadamente antes de aplicar los modelos de análisis y predicción. Los saltos en los datos pueden comprometer la calidad y precisión de los resultados, por lo que es fundamental abordarlos mediante técnicas de imputación o eliminación, dependiendo del contexto del análisis. Corregir estos datos faltantes es crucial para asegurar la integridad de los datos y la confiabilidad de los modelos predictivos. Así mismo se pudo visualizar la presencia de valores extremos o outliers (ver Figura 9):

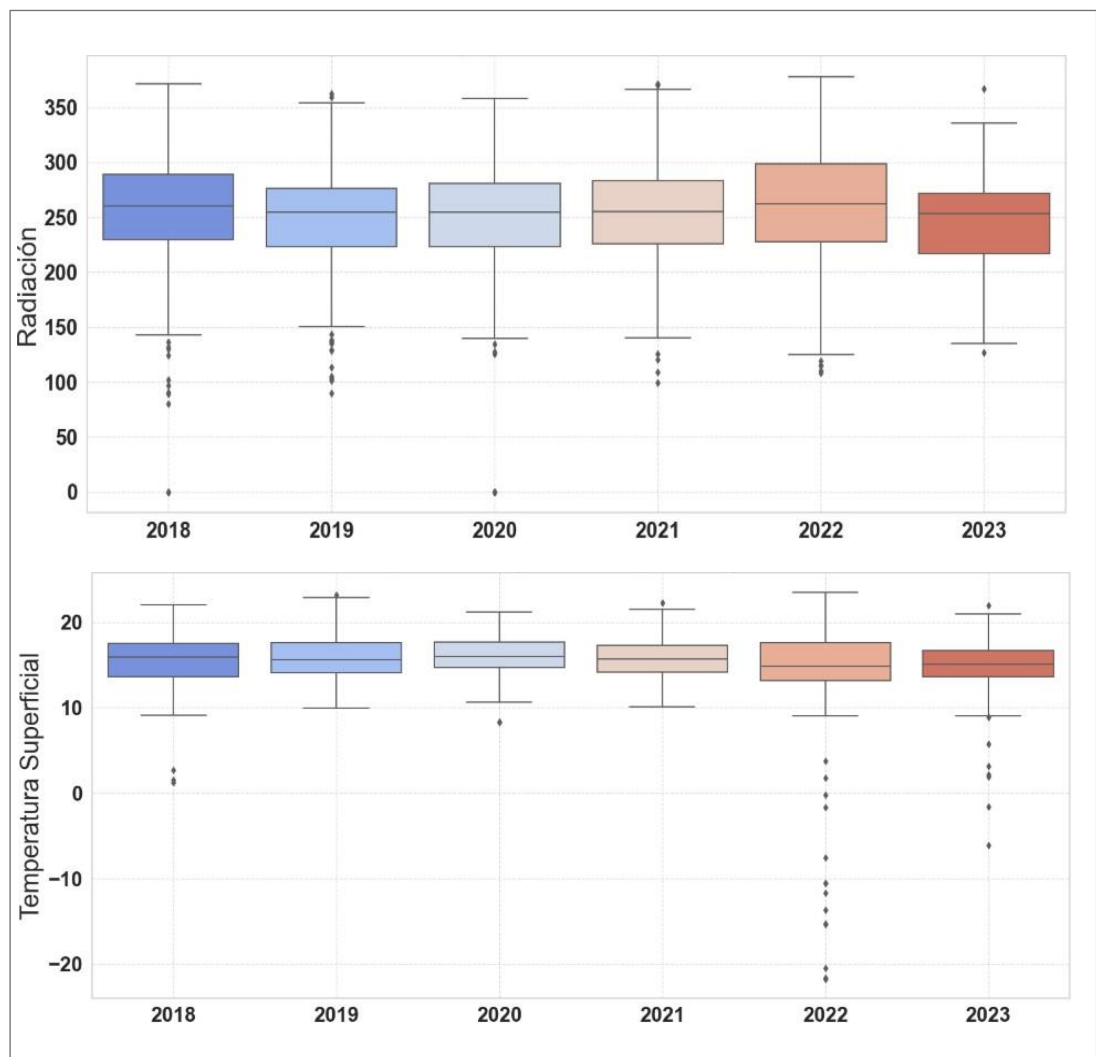


Figura 9. Presencia de datos atípicos en las variables.

En la figura 9 se puede apreciar los boxplots de las variables: temperatura mínima, radiación, temperatura superficial y humedad relativa. Se observa que, en la temperatura mínima, la mediana se sitúa alrededor de 3 a 4 °C, con un rango que abarcó desde 0 hasta 6 °C, incluyendo valores extremos que alcanzaron hasta -7 °C y 8 °C. La radiación presentó una mediana entre 240 y 270 W/m², con valores que oscilaron aproximadamente entre 170 y 370 W/m², además de outliers en los extremos inferiores. Por su parte, la temperatura superficial tiene una mediana de entre 15 y 16 °C, con rangos que variaron de 10-12 °C hasta 20 °C, y se detectaron valores atípicos negativos de hasta -10 o -20 °C. Finalmente, la humedad relativa muestra una mediana de alrededor del 50 al 60%, con un rango intercuartílico de 45 a 70% y outliers que alcanzaron valores tan bajos como el 30% o tan altos como el 75-80%.

Prueba de normalidad

Para finalizar la etapa de comprensión de los datos se realizó un análisis de normalidad utilizando las pruebas de Shapiro-Wilk y D'Agostino, junto con el cálculo de la curtosis para cada variable, lo que permitió poner a prueba la hipótesis nula de que los datos seguían una distribución normal. Se eligieron ambas pruebas de manera complementaria: Shapiro-Wilk, por su sensibilidad en muestras pequeñas o moderadas, y D'Agostino, que evaluó la asimetría y curtosis en muestras de mayor tamaño. Se emplearon paquetes de Python como numpy para operaciones numéricas, scipy.stats para el ajuste y análisis de distribuciones normales, y matplotlib.pyplot para la generación de gráficos como histogramas y curvas teóricas. El procedimiento se resume en el siguiente esquema.

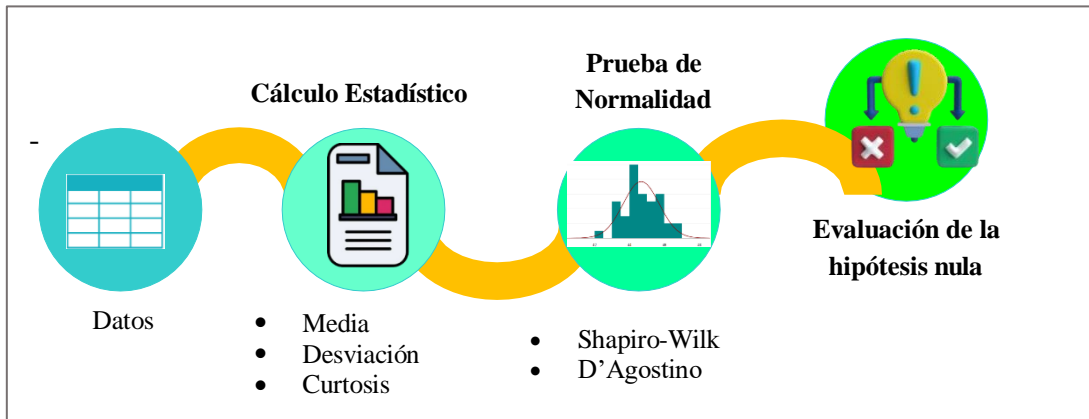


Figura 10. Esquema de comprobación de normalidad.

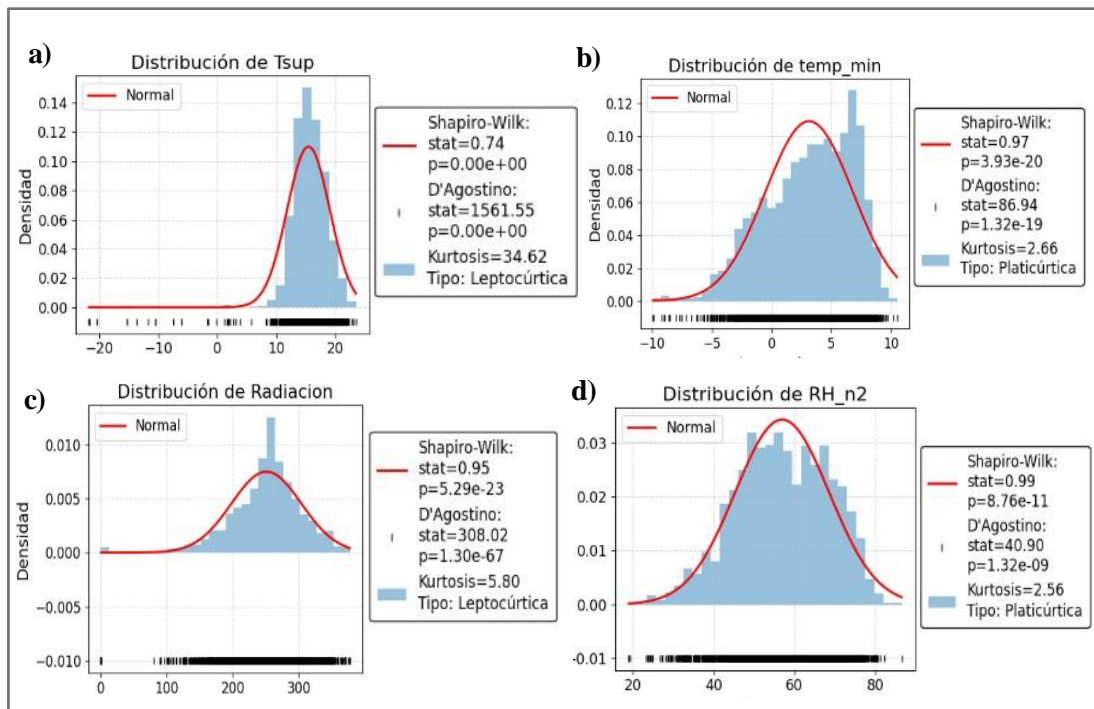


Figura 11. Prueba de normalidad para las variables en estudio. a) distribución de temperatura superficial, b) distribución de temperatura mínima, c) distribución de radiación, d) distribución de humedad relativa.

Tabla 1. Resultados de Pruebas de Normalidad.

Variable	Shapiro		D'Agostino		Curtosis
	stat	p	stat	p	
temp_min	0.97	3.93E-20	86.94	1.32E-19	2.66
temp_max	0.96	2.10E-22	92.35	3.45E-21	4.10
temp_mean	0.98	5.50E-18	75.2	4.80E-17	3.20
PrecipAcum	0.95	1.80E-25	110.5	2.10E-24	6.00
Radiacion	0.95	5.29E-23	308.02	1.30E-67	5.80
Tsup	0.74	0.00E+00	1561.55	0.00E+00	34.62
windir_total_mean	0.98	1.20E-15	45.12	2.30E-09	2.90
Press_Atm	0.99	5.67E-12	30.45	1.90E-07	3.10
vel_max	0.96	3.10E-21	78.34	2.00E-16	4.50
vel_mean	0.97	2.00E-19	70.56	5.00E-14	3.80
evsat_n2	0.98	6.78E-10	25.67	1.20E-06	2.45
RH_n2	0.99	8.76E-11	40.9	1.32E-09	2.56

Nota. - Los valores obtenidos no fueron tratados previamente.

En la figura 11 y la tabla 1 se puede apreciar los resultados de la prueba de normalidad los cuales revelaron que todas las variables analizadas mostrando p-valores inferiores a 0.05 tanto en la prueba de Shapiro-Wilk como en la de D'Agostino lo que indica que se rechazó la hipótesis nula planteada en la etapa de procesamiento de datos, la cual afirmaba que los datos siguen una distribución de normalidad, evidenciando que los datos no se ajustaron a una distribución normal y que los valores de curtosis variaron reflejando distribuciones leptocúrticas en algunas variables y platicúrticas en otras lo que sugiere la presencia de colas más pesadas o más ligeras respectivamente y la necesidad de aplicar transformaciones o métodos estadísticos robustos para lograr un análisis posterior más preciso

Preparación de los datos

El proceso de preprocesamiento de los datos del dataframe Datos_Huayao, fue un paso esencial para garantizar la calidad de los datos antes de su posterior análisis y modelado predictivo. Este preprocesamiento involucró varias técnicas, como la verificación de calidad de los datos, la detección de valores atípicos, la transformación de las distribuciones sesgadas, y la normalización de las variables numéricas. A continuación, se describen con detalle los resultados obtenidos de cada una de estas etapas.

Paso 1. Verificación de Calidad de los Datos

La calidad de los datos es fundamental para garantizar la precisión de cualquier análisis posterior. Para asegurar que los conjuntos de datos no contuvieran registros erróneos o redundantes, se aplicaron las siguientes técnicas de verificación:

- **Detección de Duplicados y Espacios en Blanco**

Uno de los primeros pasos fue la detección de duplicados y la verificación de la presencia de espacios en blanco o caracteres no deseados, como 'NA' o '-', que podrían distorsionar el análisis.

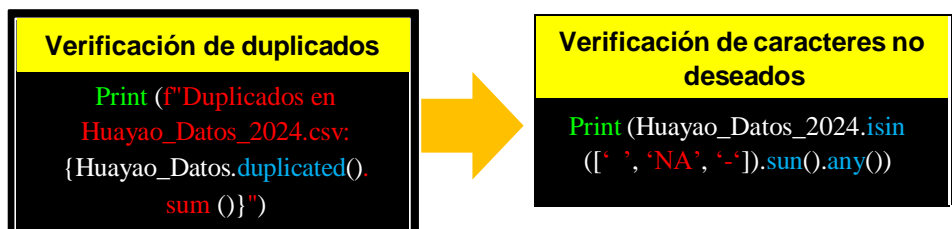


Figura 12. Verificación de calidad de datos.

En la figura 12 se puede apreciar el código Python que utiliza la librería panda para evaluar la calidad del archivo Huayao_Datos_2024.csv. Primero, se revisan filas duplicadas con. Duplicated (). sum (), lo que permite contar y mostrar cuántas filas repetidas existen en cada conjunto de datos, determinando que no hay registros duplicados. Luego, se realiza una verificación de caracteres no deseados (' ', 'NA', '-') utilizando. isin().sum().any(), lo que permite identificar si alguna celda contiene estos valores problemáticos. La salida muestra False en ambos casos, lo que confirma que los conjuntos de datos no contienen valores indeseados y están listos para un análisis más profundo.

- **Comprobación de Valores Faltantes**

El siguiente paso fue comprobar si existían valores faltantes en cualquiera de las variables de ambos conjuntos de datos. Esto es crucial porque los valores faltantes podrían requerir imputación, lo cual es una tarea compleja que puede afectar los resultados del modelo.



Figura 13. Verificación de valores faltantes.

En la figura 13 se aprecia el código Python el cual utiliza la función `isnull().sum()` de pandas para detectar y contar valores faltantes en el dataset, mostrando inicialmente que no hay datos nulos. Sin embargo, al graficar las variables se observan huecos, atribuibles a fallas en equipos meteorológicos o falta de mantenimiento. Para solucionarlo, se identificaron y crearon las fechas faltantes, reemplazando sus valores con los datos del mes más cercano (anterior o posterior) del mismo año, evitando así saltos en las gráficas por la ausencia de datos.

Se determinó que las fechas faltantes se encuentran en los siguientes periodos: para 2018 se consideró el intervalo del 01/08/2018 al 31/08/2018; para 2019 se utilizó el período del 01/11/2019 al 31/12/2019; en 2020 se tomaron los datos del 01/02/2020 al 29/02/2020, además de los intervalos del 01/10/2020 al 31/10/2020 y del 01/12/2020 al 31/12/2020; para 2021 se analizó el período del 01/01/2021 al 28/02/2021; en 2022 se emplearon los datos del 01/07/2022 al 31/07/2022 y del 15/08/2022 al 15/09/2022; y finalmente, en 2023 se consideraron los intervalos del 01/05/2023 al 31/12/2023.

En la figura. 14, se observa el diseñado para identificar y copiar los datos correspondientes a los períodos pertinentes de años anteriores o posteriores, mediante la función lambda y replace de Python.

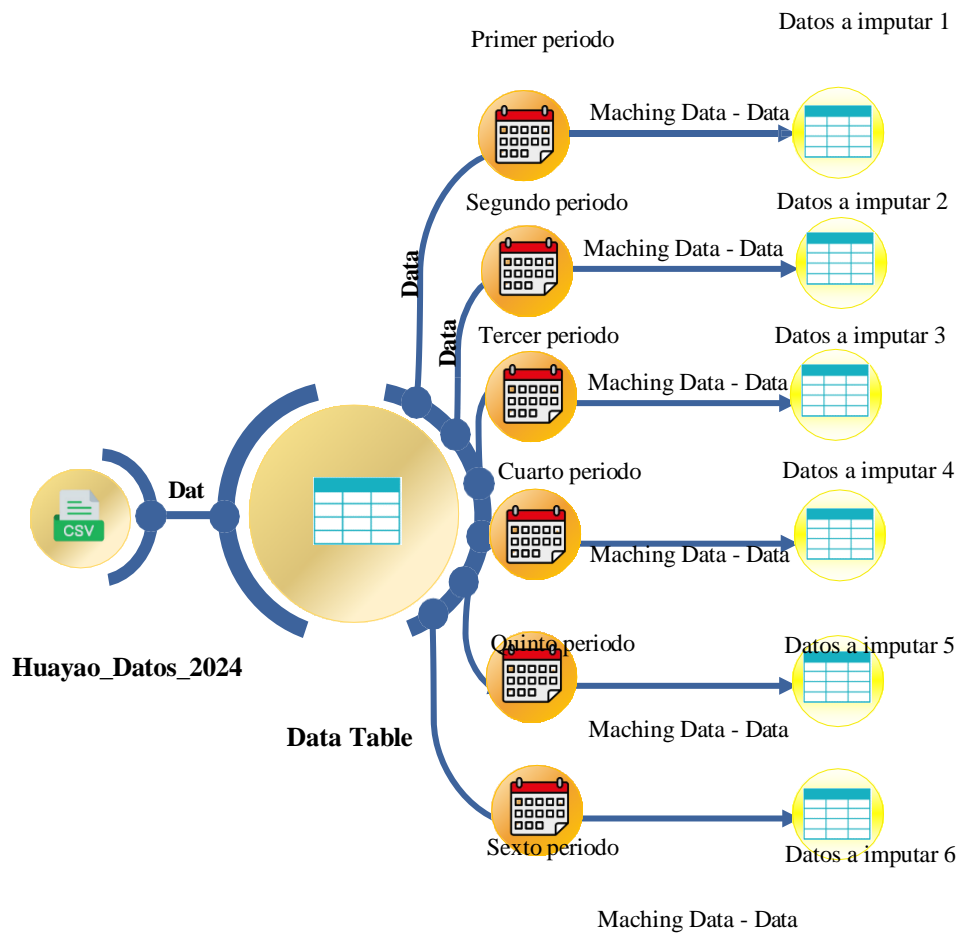


Figura 14. Estrategia de imputación de datos faltantes.

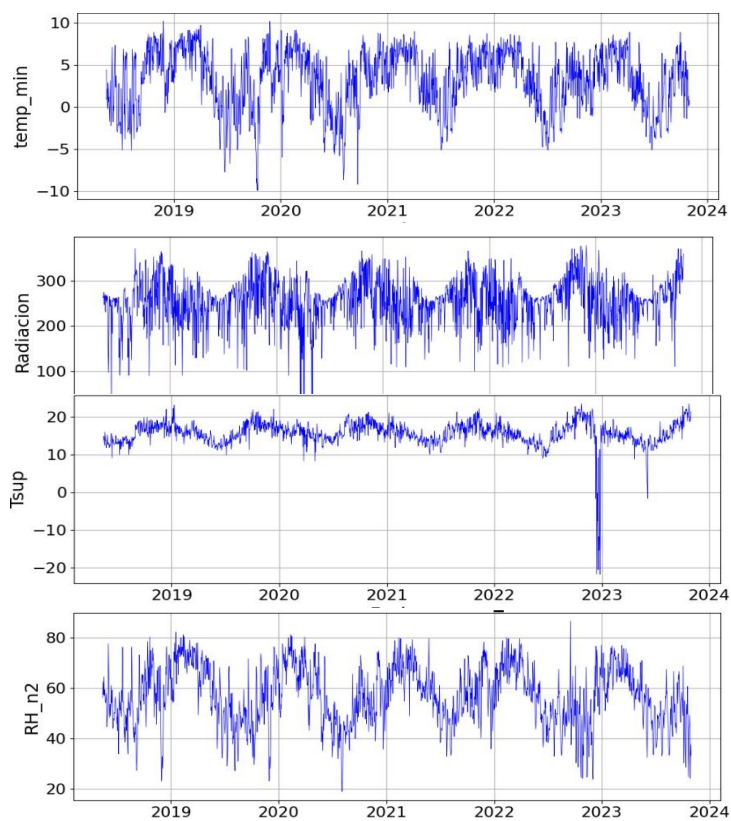


Figura 15. Resultados tras la imputación de los valores faltantes.

Paso 2. Eliminación de Valores Atípicos

La eliminación de valores atípicos fue una fase fundamental en el preprocesamiento de datos. Para ello, se aplicó el método Identificación y Manejo de Valores Atípicos usando IQR, aplicado a los datos de Huayao_Datos_2024.csv el cual eliminó valores fuera de los percentiles 25 y 75, reduciendo la influencia de extremos en el análisis. La aplicación de estos enfoques mejoró la calidad de los datos, garantizando observaciones más representativas y adecuadas para el estudio. El procedimiento se realizó de la siguiente manera:

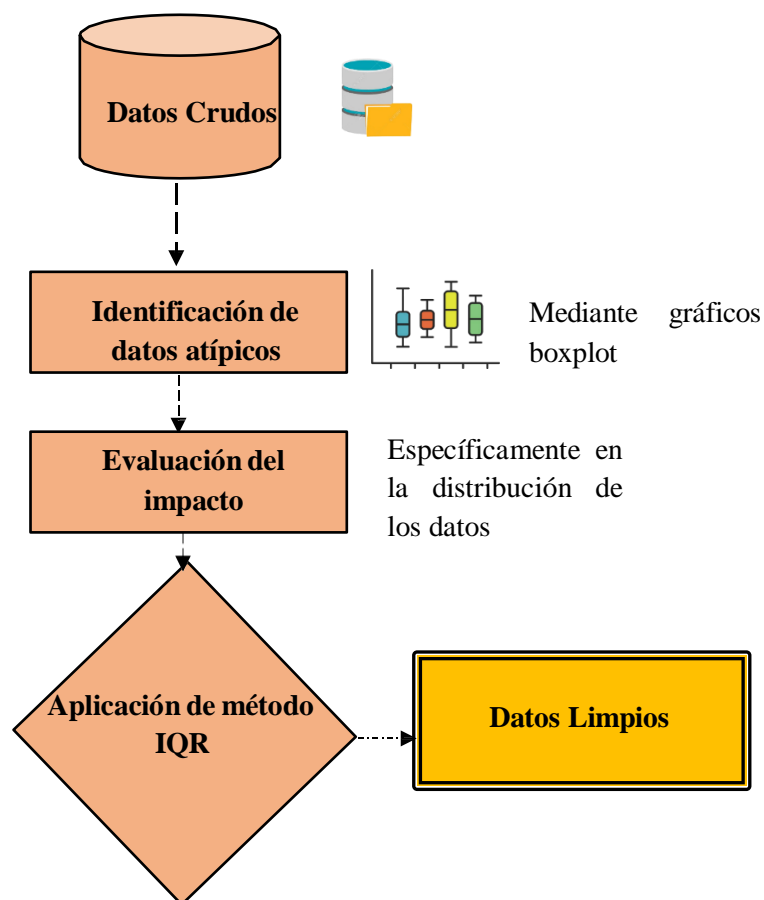


Figura 16. Estrategia de tratamiento de valores atípicos.

En la figura 16 se muestra la aplicación de la técnica de capping basada en el rango intercuartílico (IQR) para eliminar valores atípicos. Primero se suprimieron advertencias, se reemplazaron valores infinitos por NaN y se seleccionaron las columnas de interés. Luego, se calcularon los cuartiles Q1 y Q3, y se establecieron los límites inferior y superior ($Q1 - 1.5 \cdot IQR$ y $Q3 +$

1.5·IQR). Los valores fuera de estos rangos se ajustaron a los límites respectivos, reduciendo así la influencia de los outliers. Finalmente, se generaron histogramas con KDE y boxplots para validar la nueva distribución de los datos.

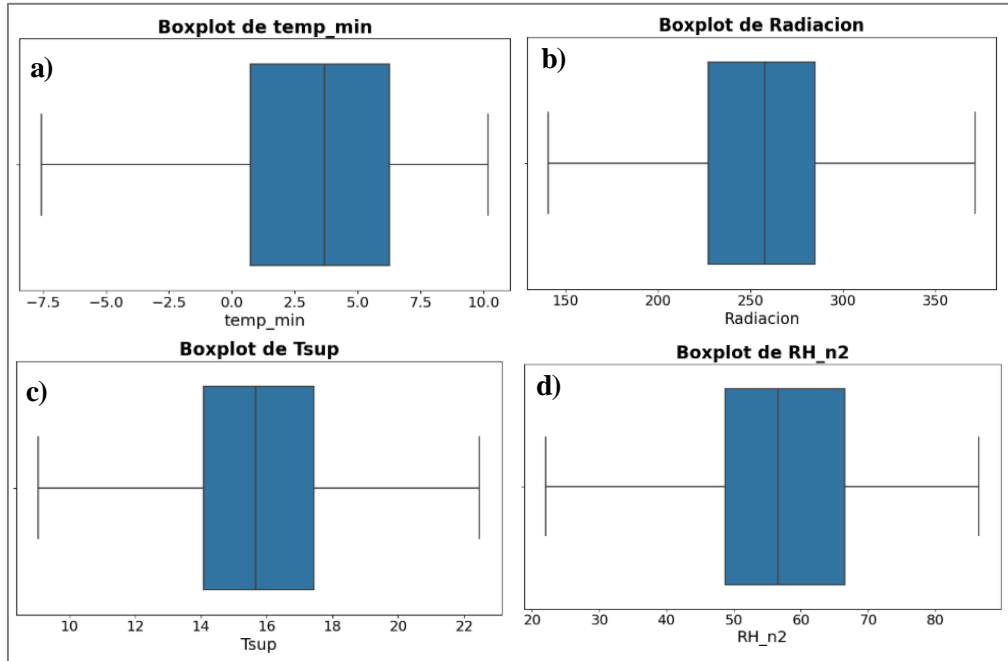


Figura 17. Gráfico de cajas de las variables tratadas. a) boxplot de temperaturas mínimas, b) boxplot de radiación, c) boxplot de temperatura superficial, d) boxplot de humedad relativa.

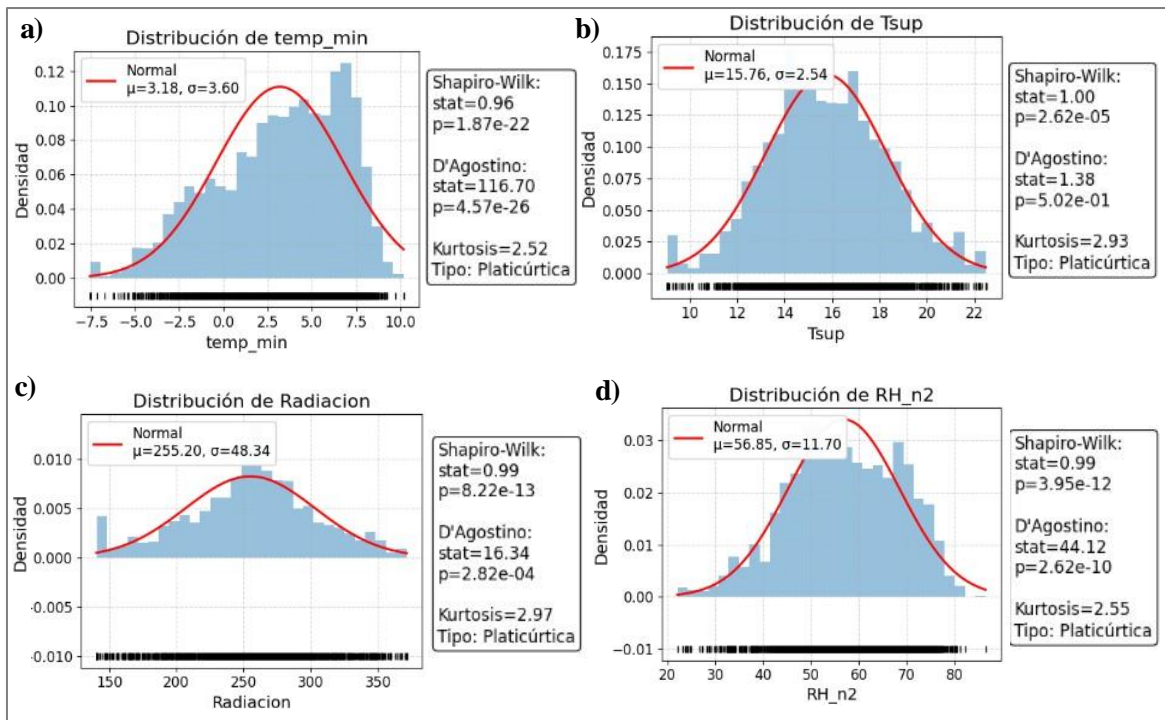


Figura 18. Prueba de normalidad para las variables luego del pretratamiento. a) distribución de temperatura mínima, b) distribución de temperatura superficial, c) distribución de radiación, d) distribución de humedad relativa.

Tabla 2. Resultados de pruebas de normalidad postratamiento.

Variable	Shapiro		D'Agostino		Curtosis
	stat	p	stat	p	
temp_min	0.96	1.87E-22	116.70	4.57E-26	2.52
temp_max	0.99	1.40E-11	32.31	9.65E-08	3.04
temp_mean	0.98	2.37E-17	73.96	8.72E-17	2.96
PrecipAcum	0.63	0.00E+00	336.78	7.38E-74	2.83
Radiacion	0.99	8.22E-13	16.34	2.82E-04	2.97
Tsup	1.00	2.62E-05	1.38	5.02E-01	2.93
windir_total	0.99	1.18E-12	69.57	7.82E-16	2.47
Press_Atm	1.00	1.70E-02	6.26	4.38E-02	2.77
vel_max	0.98	1.73E-14	56.60	5.13E-13	2.90
vel_mean	0.98	5.03E-15	60.51	7.24E-14	2.97
evsat_n2	0.99	1.69E-07	0.66	7.17E-01	3.08
RH_n2	0.99	3.95E-12	44.12	2.62E-10	2.55

Nota: Los valores obtenidos fueron tratados previamente.

En la tabla 2 y en la figura 18 se aprecian los resultados obtenidos tras la aplicación de métodos de preprocesamiento. La comparación entre los conjuntos de resultados, antes y después del preprocesamiento, evidencia que, aunque la variable temperatura mínima mostró un ligero cambio en su estadístico de Shapiro-Wilk (pasando de 0.97 a 0.96), todas las pruebas siguen presentando p-valores inferiores a 0.05, lo que mantiene el rechazo de la hipótesis de normalidad. Esto indica que el preprocesamiento aplicado no fue suficiente para lograr un ajuste normal de los datos, por lo que es necesario implementar estrategias adicionales. En este sentido, se propone el desarrollo y la aplicación de métodos de inteligencia artificial, que permitan identificar y corregir de forma más precisa los patrones anómalos y la presencia de valores extremos, contribuyendo a obtener conclusiones más robustas y precisas en la investigación.

Paso 3. Correlacionado de variables

Una vez finalizado el tratamiento de los datos faltantes y atípicos, se procedió a la siguiente etapa del análisis, que consistió en correlacionar las variables meteorológicas clave para el pronóstico de heladas. El propósito de este análisis fue identificar las relaciones entre las variables y seleccionar aquellas que pudieran contribuir significativamente al modelo predictivo, eliminando las variables redundantes que mostraran un alto grado de dependencia mutua.

Método de Correlación Utilizado

Para el análisis de correlación se utilizó el método de Spearman, el cual resulta apropiado cuando los datos no presentan una distribución normal o exhiben sesgos moderados. En la figura X se visualizan los coeficientes de correlación en un mapa de calor, donde los tonos rojos más intensos reflejan correlaciones positivas fuertes (cercanas a +1), los azules más intensos representan correlaciones negativas fuertes (cercanas a -1) y los colores próximos al blanco indican relaciones débiles (cercanas a 0).

En esta matriz destaca la correlación positiva alta entre la humedad relativa (RH_n2) y la temperatura mínima ($r = 0.71$), lo que sugiere que un mayor contenido de humedad contribuye a mantener el calor y evita descensos térmicos bruscos. Además, se observa una correlación moderada entre la precipitación acumulada (PrecipAcum) y la temperatura mínima ($r = 0.52$), lo cual indicaría que la nubosidad asociada a la lluvia retiene el calor durante la noche, limitando la caída de la temperatura. En conjunto, estos hallazgos señalan que las variables más relevantes para la predicción de heladas son la temperatura mínima, la radiación y la humedad relativa. Como se observa en la figura 19.

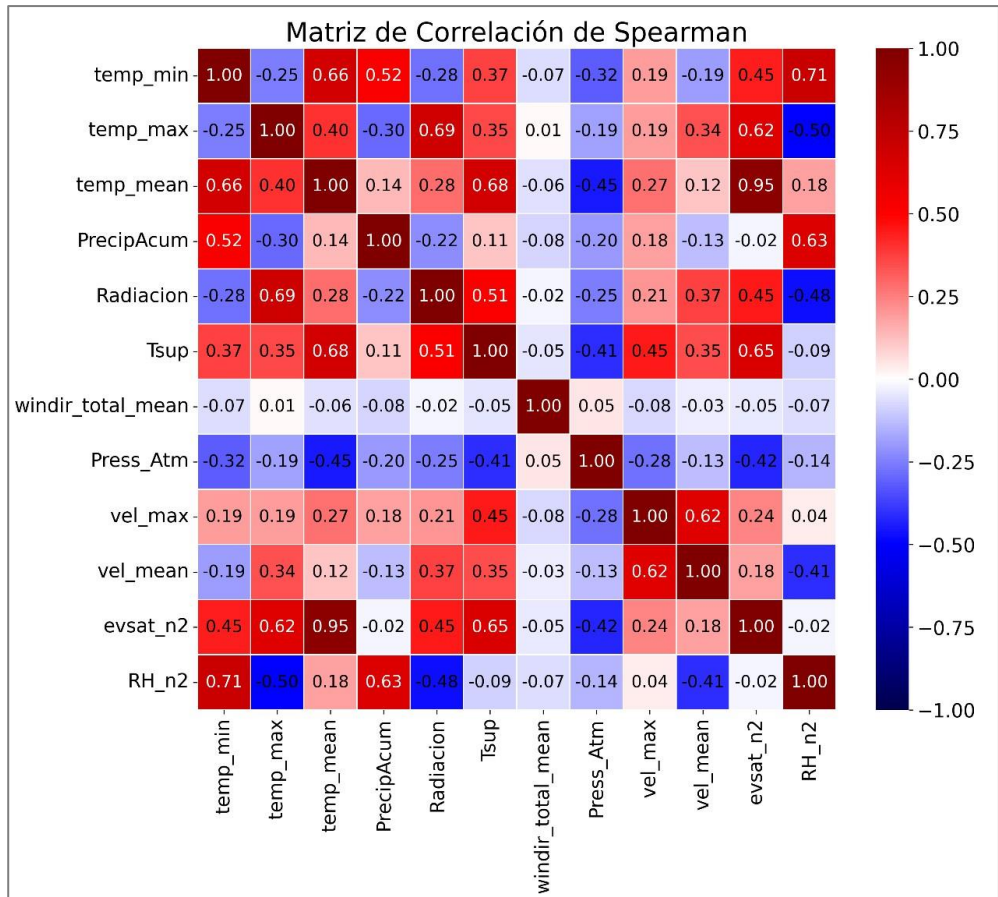


Figura 19. Correlación de Spearman entre las variables meteorológicas y la variable temperatura mínima.

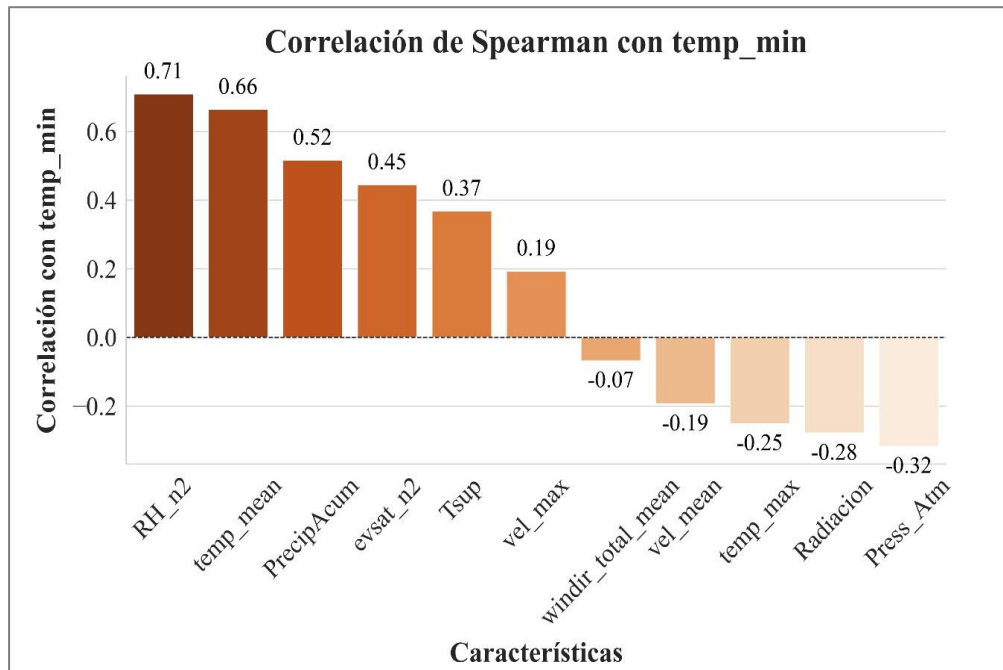


Figura 20. Variables con mayor correlación con la variable temperatura mínima.

En la figura 20 se visualiza los gráficos de barras de la relación de cada variable con la temperatura mínima. En el gráfico que muestra la correlación de Spearman, se aprecia que la humedad relativa presenta el valor más alto ($r = 0.71$), lo cual indica que una mayor humedad puede contribuir a mantener el calor en el ambiente, impidiendo descensos térmicos bruscos. La precipitación acumulada exhibe una correlación positiva moderada ($r = 0.53$), sugiriendo que la nubosidad asociada con la lluvia limita la pérdida de calor durante la noche y disminuye la probabilidad de heladas. Por otro lado, la presión atmosférica muestra una correlación negativa ($r = -0.32$), lo que implica que presiones bajas podrían favorecer temperaturas mínimas más bajas.

En conjunto, estos resultados corroboran la importancia de variables como la humedad relativa, la radiación y la propia temperatura mínima en la predicción de heladas.

5.3 Analizar los datos históricos para identificar patrones que permitan el entrenamiento de un modelo de inteligencia artificial

Dado que las heladas suelen ocurrir en los meses de junio, julio y agosto, se seleccionaron únicamente los datos correspondientes a esos meses a lo largo de los años en estudio que se encuentran en el dataset. En primer lugar, se extrajo el mes de la columna “Fecha”, permitiendo filtrar solo los datos de los meses de cuestión. A continuación, se seleccionaron las variables que tuvieron mayor implicancia en la ocurrencia de una helada como se pudo corroborar en la fase de correlación, las cuales son: temperatura mínima, humedad relativa, temperatura superficial del suelo a 2 metros de la superficie, precipitación y radiación. Con esto se creó un conjunto de datos reducido para realizar un análisis más enfocado.

El siguiente paso consistió en manejar los valores faltantes, eliminando las filas con datos incompletos, ya que estos podrían haber afectado negativamente la precisión del modelo. Después, se definió la variable objetivo, que fue la temperatura mínima y se separaron las características predictoras, lo que permitió estructurar los datos para el entrenamiento del modelo. Con las variables preparadas, se normalizaron las

características utilizando StandardScaler, lo que permitió que todas las variables estuvieran en la misma escala, algo crucial para garantizar el rendimiento adecuado de los modelos de regresión.

Para optimizar el modelo se llevó a cabo un análisis detallado que permitió identificar el valor óptimo de random_state, lo que resultó en una división más equilibrada de los datos entre entrenamiento y prueba y, por ende, en una maximización del valor de R^2 . Este paso fue fundamental para asegurar que el modelo pudiera capturar de manera efectiva la relación subyacente entre las variables predictoras y la temperatura mínima, sin depender de una partición arbitraria de los datos.

Posteriormente, se entrenó un modelo de regresión lineal. Aunque la regresión lineal no se clasifica estrictamente como un modelo de inteligencia artificial, su utilización en este contexto permitió evaluar el comportamiento del conjunto de datos a través de métricas robustas como el RMSE y R^2 .

Finalmente, se visualizó el rendimiento del modelo mediante gráficos que facilitaban la interpretación de los resultados. Estas representaciones gráficas permitieron comparar de forma directa las predicciones generadas con los valores reales, evidenciando la capacidad del modelo para identificar y reproducir patrones climáticos clave presentes en los datos históricos.

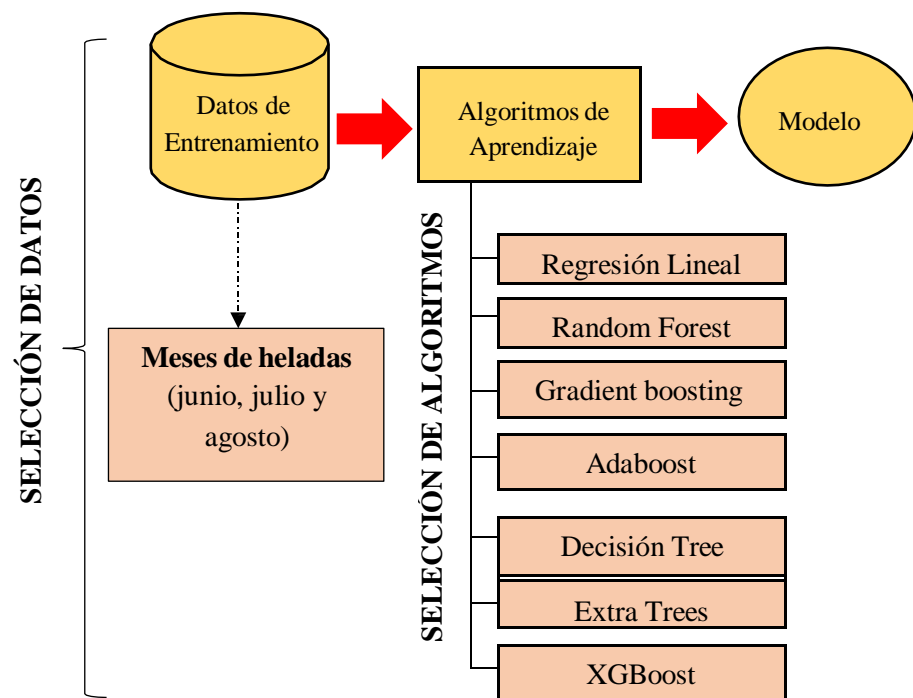


Figura 21. Esquema de selección de datos y modelos de IA.

5.4 Evaluar la precisión y confiabilidad de distintos modelos de inteligencia artificial en la predicción de heladas

En esta etapa, al evaluar la precisión y confiabilidad de los modelos de inteligencia artificial, se utilizaron métricas comunes como el RMSE (Root Mean Squared Error), R^2 (coeficiente de determinación) y el coeficiente de Gini, además de trazar gráficos de dependencia parcial para explicar la importancia de cada variable y evaluar su comportamiento a lo largo de la predicción. Para calcular el RMSE, se predijeron los valores de la variable objetivo en los conjuntos de entrenamiento y prueba, y se aplicó la fórmula correspondiente, que es la raíz cuadrada de la media del error cuadrático medio entre las predicciones y los valores reales; un RMSE bajo indicó que el modelo tenía buena precisión.

Por su parte, R^2 midió la proporción de la varianza de los datos que fue explicada por el modelo; un R^2 cercano a 1 reflejó que el modelo describía la mayor parte de la variabilidad de los datos, mientras que un valor cercano a 0 señaló una capacidad explicativa limitada. El coeficiente de Gini y los gráficos de dependencia parcial se emplearon para analizar la importancia relativa de cada variable, permitiendo evaluar de forma más profunda cómo influían en la predicción de heladas.

Posteriormente, se presentaron los resultados en un gráfico de dispersión comparando las predicciones con los valores reales para ambos conjuntos (entrenamiento y prueba). Se distinguieron las predicciones de entrenamiento con puntos azules y las de prueba con puntos rojos, incluyendo una línea de 1:1 (línea negra discontinua) para señalar la relación ideal entre predicción y valor real. En el propio gráfico se anotaron los valores de RMSE y R^2 para los conjuntos de entrenamiento y prueba, lo cual facilitó la interpretación visual del rendimiento del modelo. Finalmente, se personalizó la presentación del gráfico ajustando los bordes, añadiendo etiquetas en los ejes y un título descriptivo, y se guardó la imagen en la ruta especificada para su posterior análisis y consulta.

Se realizó un estudio integral para predecir la temperatura mínima diaria, clave en la estimación de heladas. Se configuraron parámetros de matplotlib para gráficos claros y profesionales. Luego, con Pandas, se preprocesaron los datos, extrayendo el mes de la “Fecha” y filtrando junio, julio y agosto. Finalmente, se seleccionaron variables relevantes como temperatura mínima, máxima, media, humedad relativa, radiación y temperatura superficial, generando un conjunto de datos filtrado.

El siguiente paso fue tratar los valores faltantes, eliminando aquellas filas con datos incompletos para asegurar la calidad de los resultados. Posteriormente, se definió la variable objetivo y las características predictoras, las cuales fueron normalizadas utilizando **StandardScaler** para evitar cualquier sesgo derivado de las diferencias de magnitud entre las variables. Los datos fueron luego divididos en dos subconjuntos: uno para entrenamiento y otro para prueba, usando la función `train_test_split` de **scikit-learn**.

A continuación, se realizó un análisis completo utilizando un modelo de regresión Random Forest para predecir la temperatura mínima diaria a partir de diversas variables climáticas. El proceso comenzó cargando el conjunto de datos desde un archivo CSV, convirtiendo la columna “Fecha” a formato `datetime` y filtrando la información para el periodo comprendido entre el 1 de julio de 2018 y el 31 de julio de 2022, con el fin de concentrarse en tendencias específicas.

Una vez filtrados los datos, se seleccionaron las variables relevantes (temperatura máxima, media, humedad relativa, radiación y temperatura superficial), eliminando registros con valores faltantes en estas. Se separó la variable objetivo de las características predictoras y, finalmente, se dividieron los datos en conjuntos de entrenamiento y prueba, asignando un 33% para la prueba. Así mismo se normalizaron las características con **StandardScaler** para homogeneizar la escala de las variables, evitando sesgos por diferencias en magnitud. Posteriormente, se entrenó un modelo Random Forest de 100 árboles y se realizaron predicciones para los conjuntos de entrenamiento y prueba. Las métricas de evaluación mostraron un excelente ajuste en el entrenamiento, con un RMSE de 0.40 y un R^2 de 0.99, lo que indica una alta precisión en la predicción de la temperatura mínima. Tal y como se observa en la figura 22.

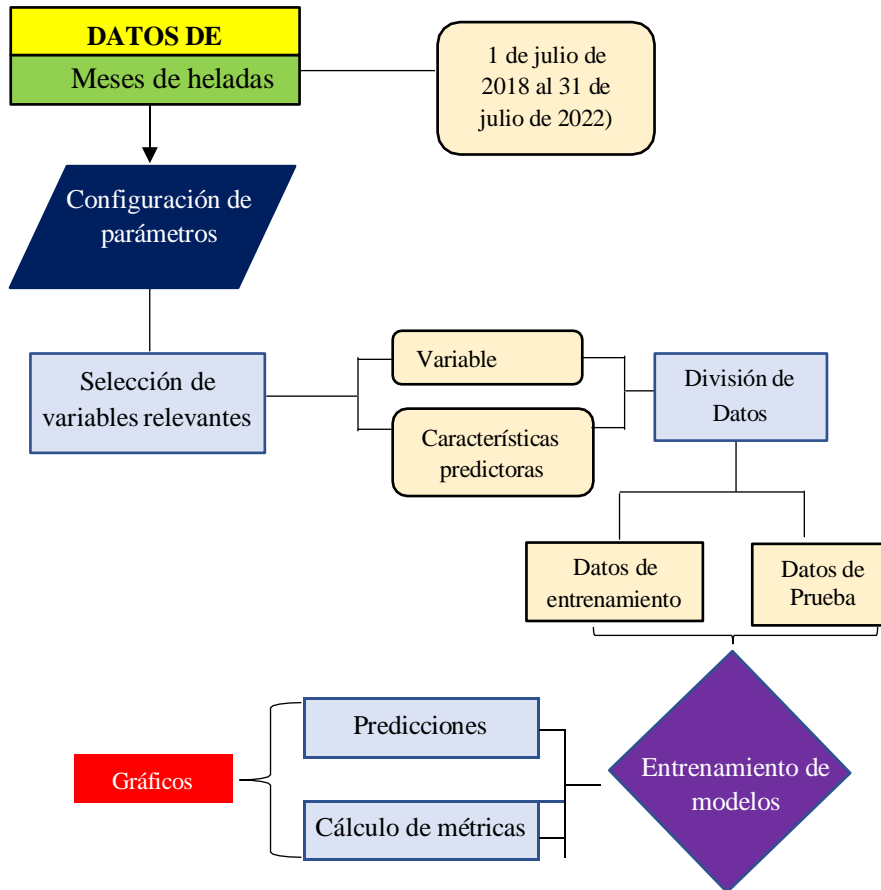


Figura 22. Esquema de etapa de entrenamiento y evaluación de los modelos de IA.

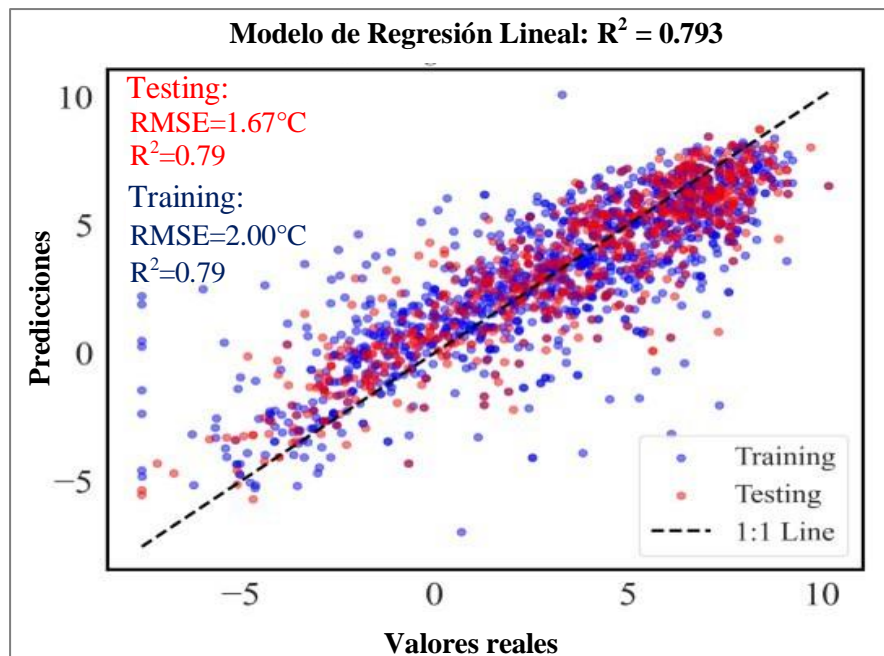


Figura 23. Gráfico de dispersión del modelo de regresión lineal.

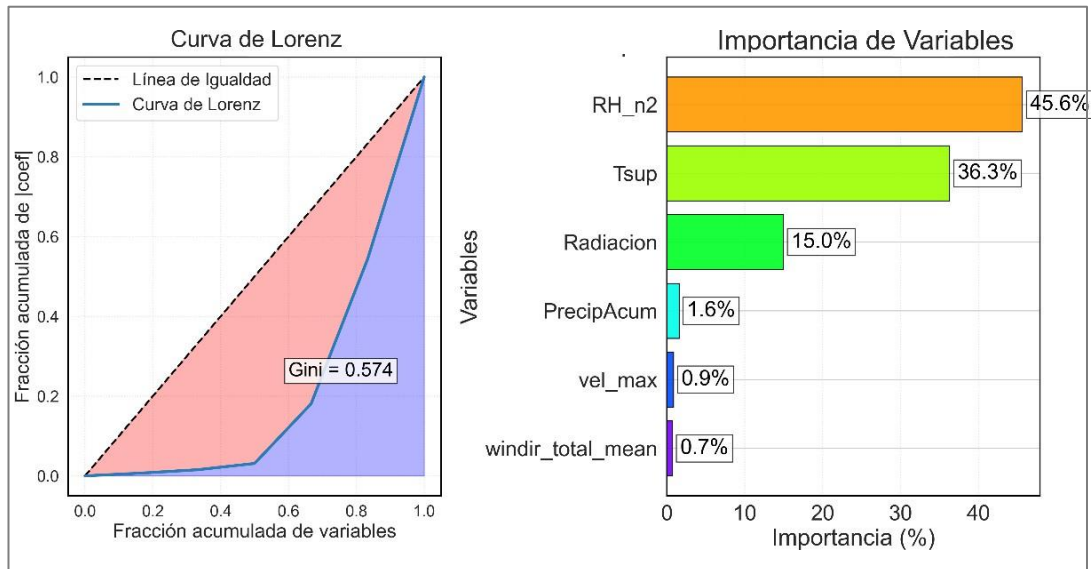


Figura 24. Coeficiente de Gini y grafica de importancia de variables.

En la figura 23 se puede apreciar que el modelo de regresión lineal se muestra robusto en sus resultados, alcanzando un $R^2 \approx 0.79$ tanto en el conjunto de entrenamiento como en el de prueba. En dicha figura se observó que las predicciones se agruparon de manera muy cercana a la línea de referencia 1:1, lo que evidenció un buen ajuste del modelo. Se determinó un RMSE de 2.00 °C en el conjunto de entrenamiento y de 1.67 °C en el conjunto de prueba, lo que confirmaba la consistencia y estabilidad en las estimaciones realizadas.

En la figura 24 se puede apreciar que, a través de la Curva de Lorenz, se identificó una concentración de la importancia en unas pocas variables. Se constató que la humedad relativa aportó aproximadamente un 45.6 % de la importancia total del modelo, mientras que la temperatura superficial contribuyó en torno a un 36.3 %. Asimismo, la radiación evidenció una contribución del 15 %, quedando las demás variables (como la precipitación acumulada, la velocidad máxima y la dirección media del viento) con un impacto mucho menor. Estos resultados destacaron que, para la predicción de heladas, los factores más relevantes fueron la humedad relativa y la temperatura superficial, mientras que otras variables jugaron un rol secundario.

Resultados de los modelos de Inteligencia artificial

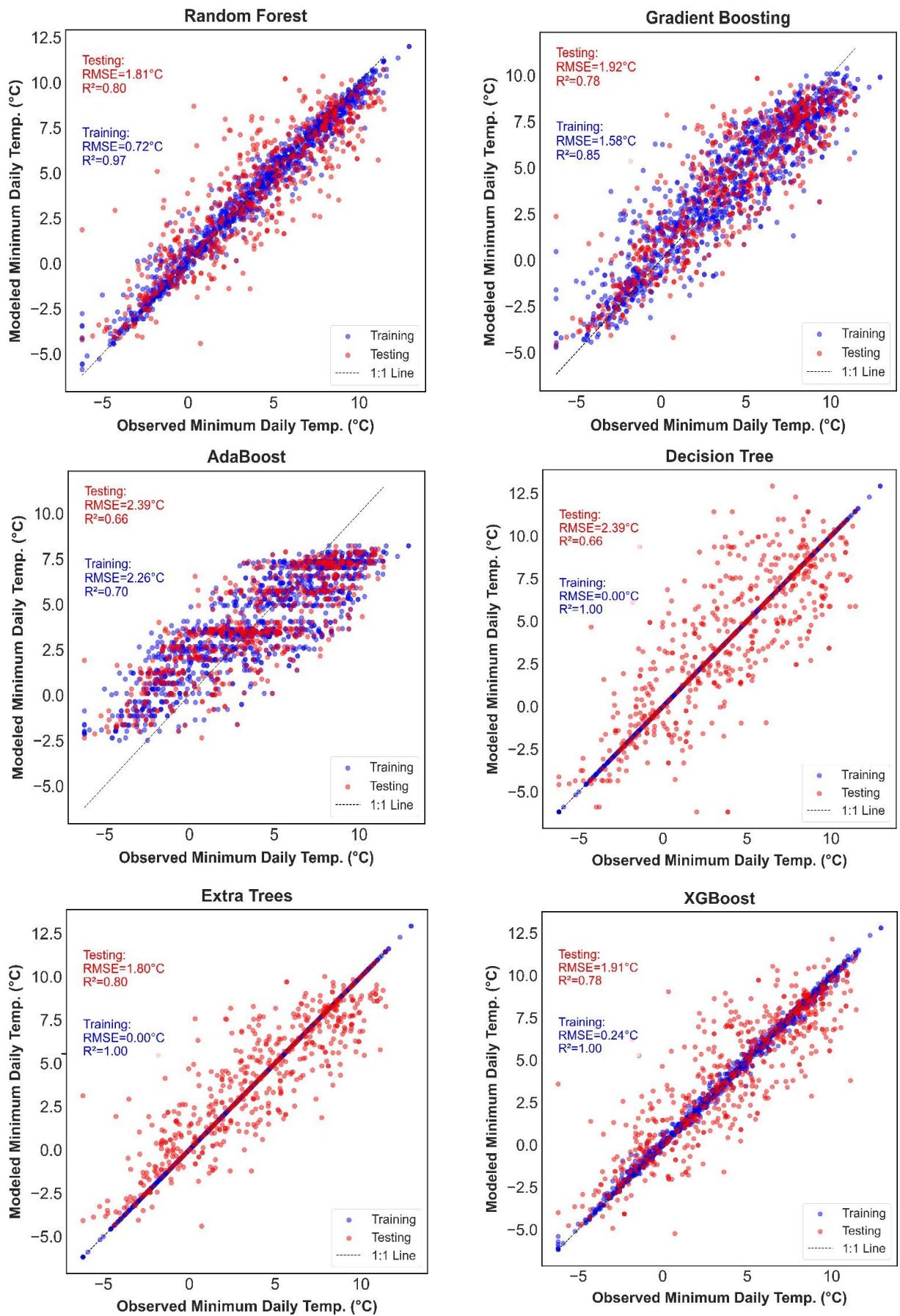
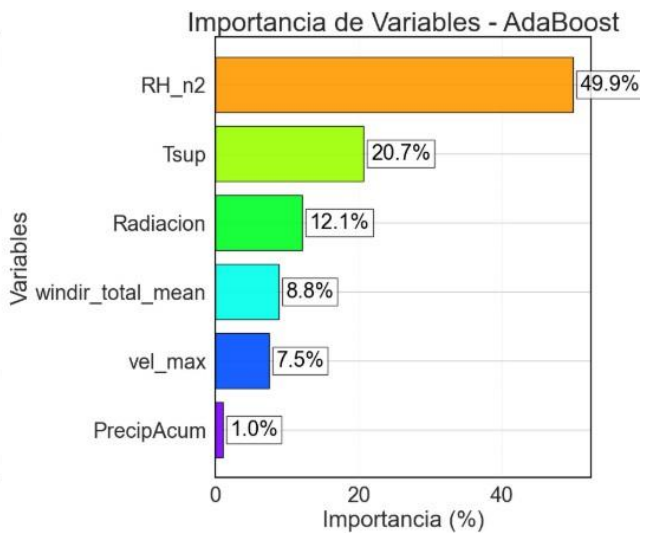
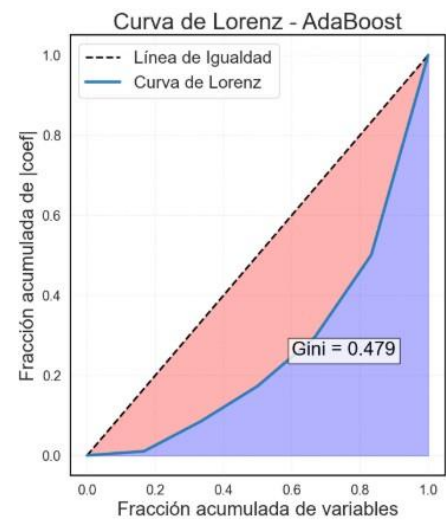
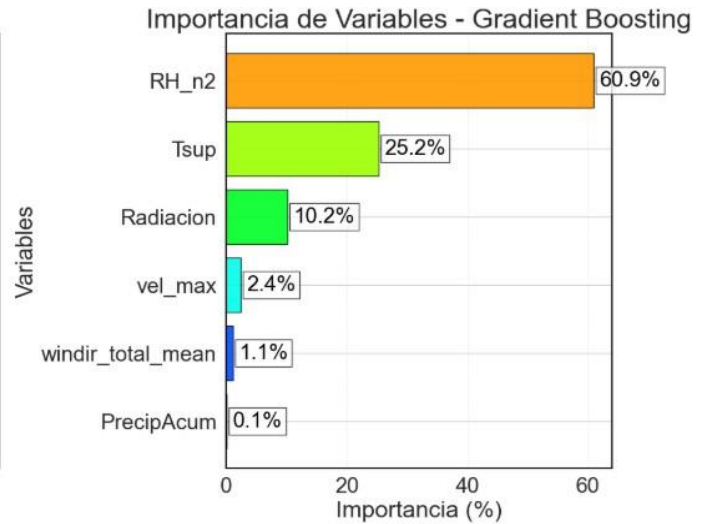
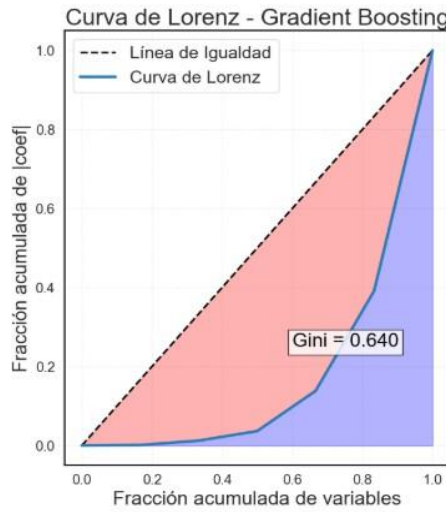
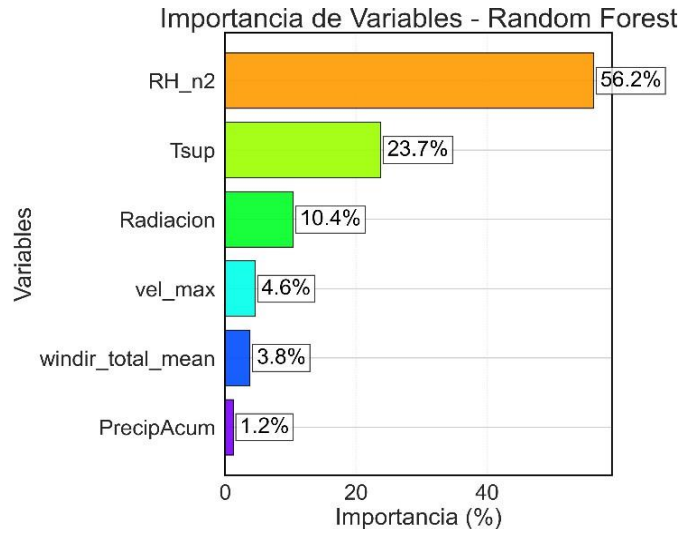
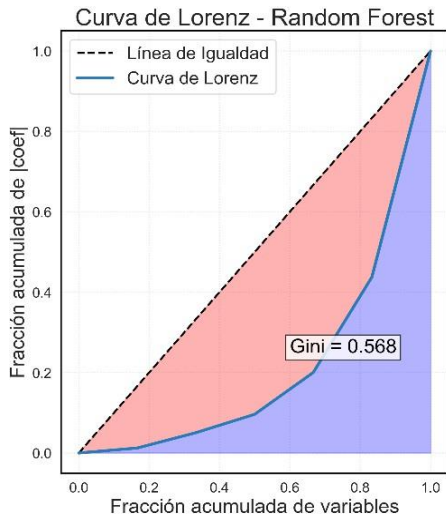


Figura 25. Comparación de dispersión en los modelos de inteligencia artificial.



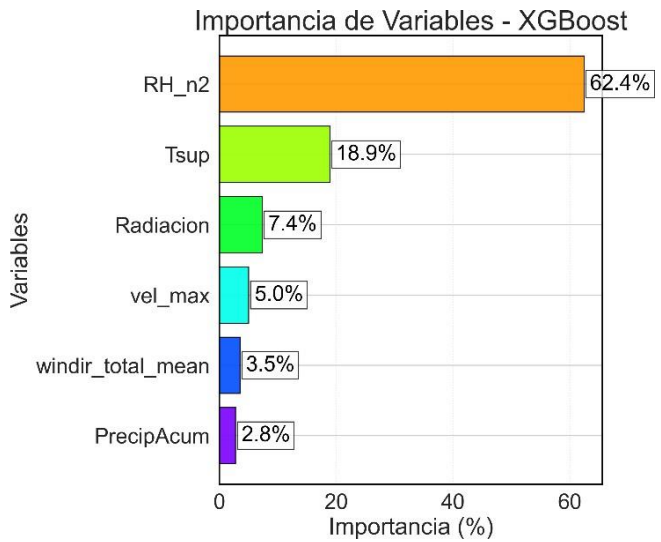
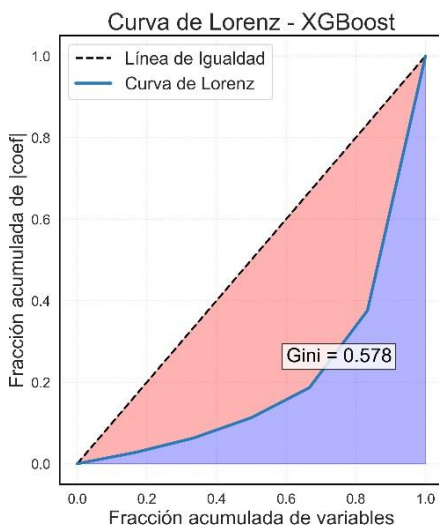
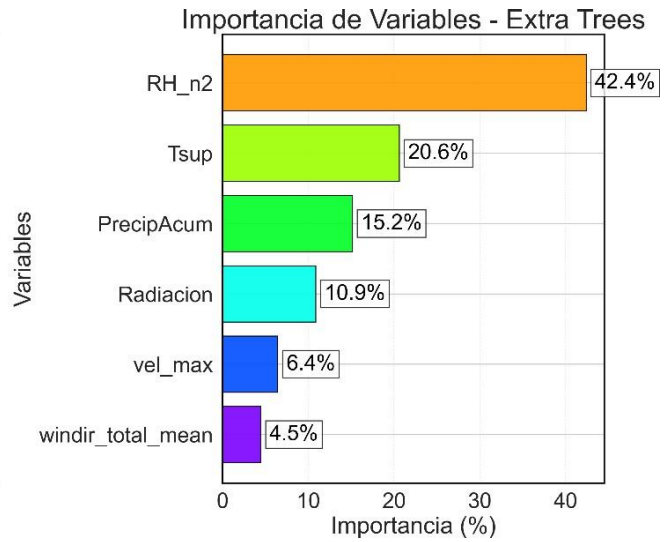
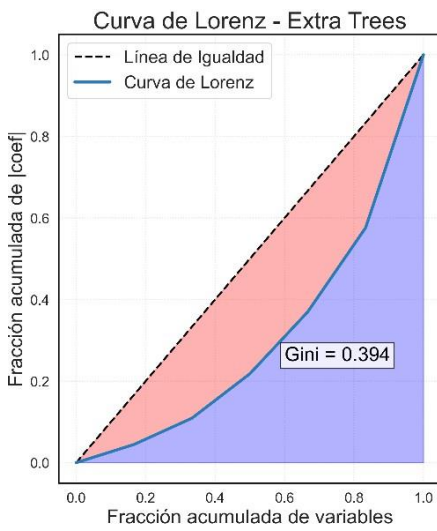
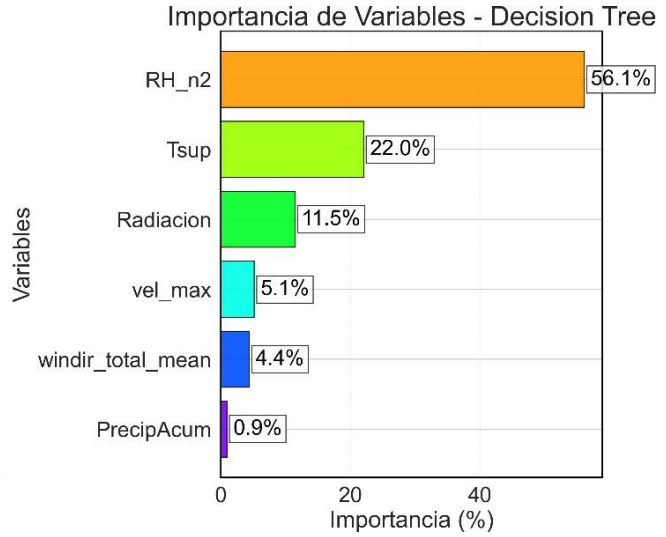
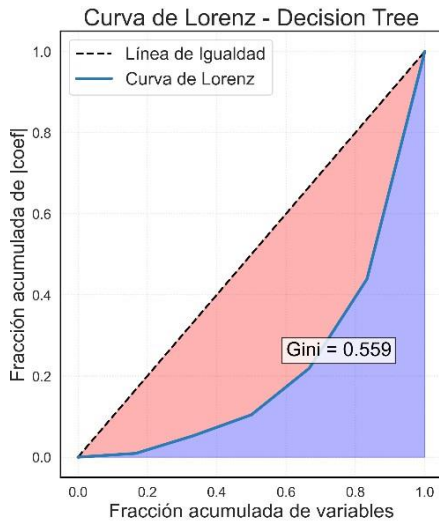


Figura 26. Coeficiente de Gini e importancias de variables para los modelos de IA.

En la figura 25 se puede apreciar que, al aplicar los distintos modelos de aprendizaje supervisado para la predicción de heladas, se obtuvieron resultados muy reveladores en cuanto a la estimación de heladas. Durante la etapa de análisis se constató que los modelos basados en técnicas de ensamblado, específicamente **XGBoost** y **Random Forest**, alcanzaron un desempeño sobresaliente tal y como lo muestra la tabla 3.

Tabla 3. Resultados del pronóstico de heladas en la fase entrenamiento.

Modelo	RMSE (Entrenamiento)	R² (Entrenamiento)	RMSE (Prueba)	R² (Prueba)
Random Forest	0.72°C	0.97	1.81°C	0.8
Gradiente Boosting	1.58°C	0.85	1.92°C	0.78
AdaBoost	2.26°C	0.7	2.39°C	0.66
Decision Tree	0.00°C	1	2.39°C	0.66
Extra Trees	0.00°C	1	1.80°C	0.8
XGBoost	0.24°C	1	1.91°C	0.78

Nota: Resumen de los resultados de distintos modelos de aprendizaje automático utilizados para predecir la temperatura mínima diaria observada. Se presentan los valores del error cuadrático medio (RMSE) y el coeficiente de determinación (R²) tanto para el conjunto de entrenamiento como para el de prueba. Un RMSE menor y un R² más cercano a 1 indican un mejor desempeño del modelo.

El análisis de la tabla 3 evidencia que algunos modelos, como Decision Tree, Extra Trees y XGBoost, presentan sobreajuste, reflejado en un R² de 1 en entrenamiento. Sin embargo, su desempeño en el conjunto de prueba es inferior. Decision Tree obtiene un RMSE de 2.39°C, indicando una capacidad limitada de generalización. Por su parte, Extra Trees y XGBoost, con RMSE de 1.80°C y 1.91°C, respectivamente, también muestran indicios de sobreajuste.

En contraste, Random Forest y Gradient Boosting presentan un mejor equilibrio entre entrenamiento y prueba. Random Forest destaca con el menor RMSE en prueba (1.81°C) y un R² de 0.80, lo que sugiere alta precisión sin indicios de sobreajuste. Gradient Boosting, aunque con un RMSE de 1.92°C y un R² de 0.78, muestra un desempeño ligeramente inferior.

Por otro lado, AdaBoost es el modelo con peor rendimiento, con RMSE de 2.39°C y R^2 de 0.66, lo que indica una menor capacidad predictiva y una generalización deficiente.

En la figura 27 se estudió la distribución de la importancia de las variables mediante la aplicación de la Curva de Lorenz, herramienta que permitió desglosar la contribución individual de cada predictor en la determinación de la temperatura mínima. Se constató que la humedad relativa y la temperatura superficial fueron los predictores más relevantes, explicando cerca del 80 % de la variabilidad en la temperatura mínima. Esto concuerda con la base física de las heladas, donde ambos factores son clave en la formación de condiciones de congelación. La radiación también tuvo un impacto significativo, aportando aproximadamente el 15 % a la predicción, lo que refleja su papel en la disipación de energía térmica nocturna. En contraste, la precipitación acumulada, la velocidad máxima del viento y la dirección del viento tuvieron una influencia menor. Aunque pueden afectar el fenómeno, su contribución fue secundaria frente a las variables principales.

5.4.1 Evaluación de la precisión y confiabilidad de modelos de IA en la predicción de heladas a partir de ventanas horarias.

Para el pronóstico de heladas con una ventana de temperaturas definidas se utilizaron herramientas como numpy, pandas, matplotlib, seaborn, scikit-learn y XGBoost, además de funciones específicas para el escalado, la división en conjuntos de entrenamiento y prueba y el cálculo de métricas de evaluación (RMSE, R^2 y MAE), lo que permitió establecer una base sólida para el análisis posterior.

A continuación, se procedió a la carga y preparación de los datos mediante la lectura del archivo "data_modified.csv" con pandas, configurando la primera columna como índice y parseando las fechas para trabajar con series temporales. Además, se ordenaron los datos de manera cronológica y se filtraron para conservar únicamente aquellos correspondientes a los meses en que se producen heladas (junio, julio y agosto), garantizando así que el análisis se enfoque en el periodo de mayor interés.

Posteriormente se definieron las variables predictoras y la variable objetivo; se seleccionaron las columnas correspondientes a 'Precipitacion', 'RH_mean', 'windir_mean', 'Vel_wind_max', 'Radiacion' y 'Tsup' como características a utilizar, mientras que 'temp_min' se estableció como la variable objetivo para predecir la temperatura mínima diaria. Asimismo, se normalizaron las características utilizando el método StandardScaler, evitando de esta forma sesgos derivados de las diferencias en magnitudes entre las variables.

Finalmente se configuraron lead times de 6, 12, 24, 30, 36 y 48 horas, desplazando la serie objetivo y combinándola con las características escaladas, eliminando filas nulas y dividiendo los datos en 67% para entrenamiento y 33% para prueba. Posteriormente se definieron y entrenaron diversos modelos de machine learning, evaluándolos con métricas como MAE, MSE, RMSE y R^2 y almacenando los resultados para su análisis

comparativo posterior. Como se ve en la figura 27.

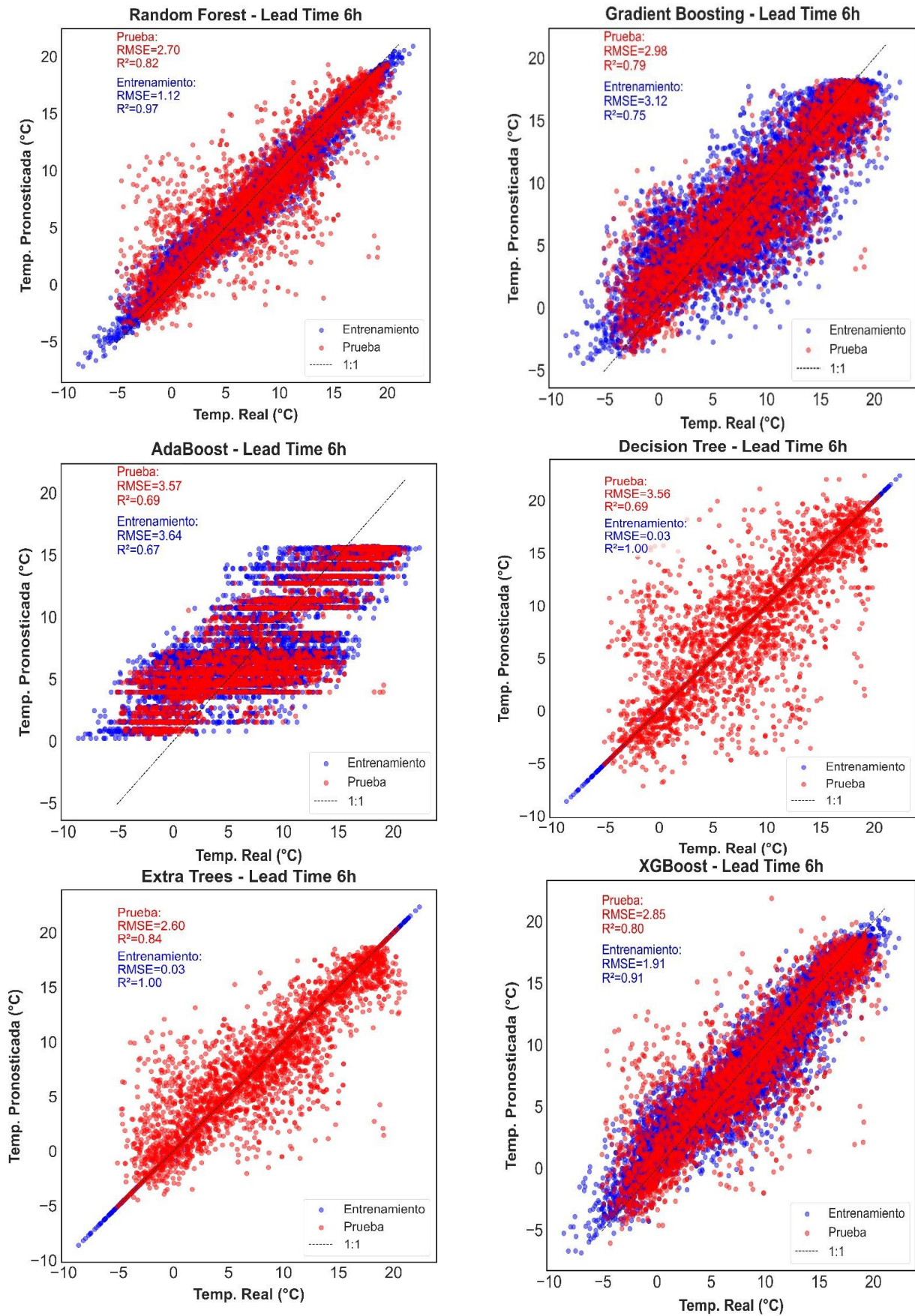


Figura 27. Pronóstico de heladas con 6 horas de anticipación.

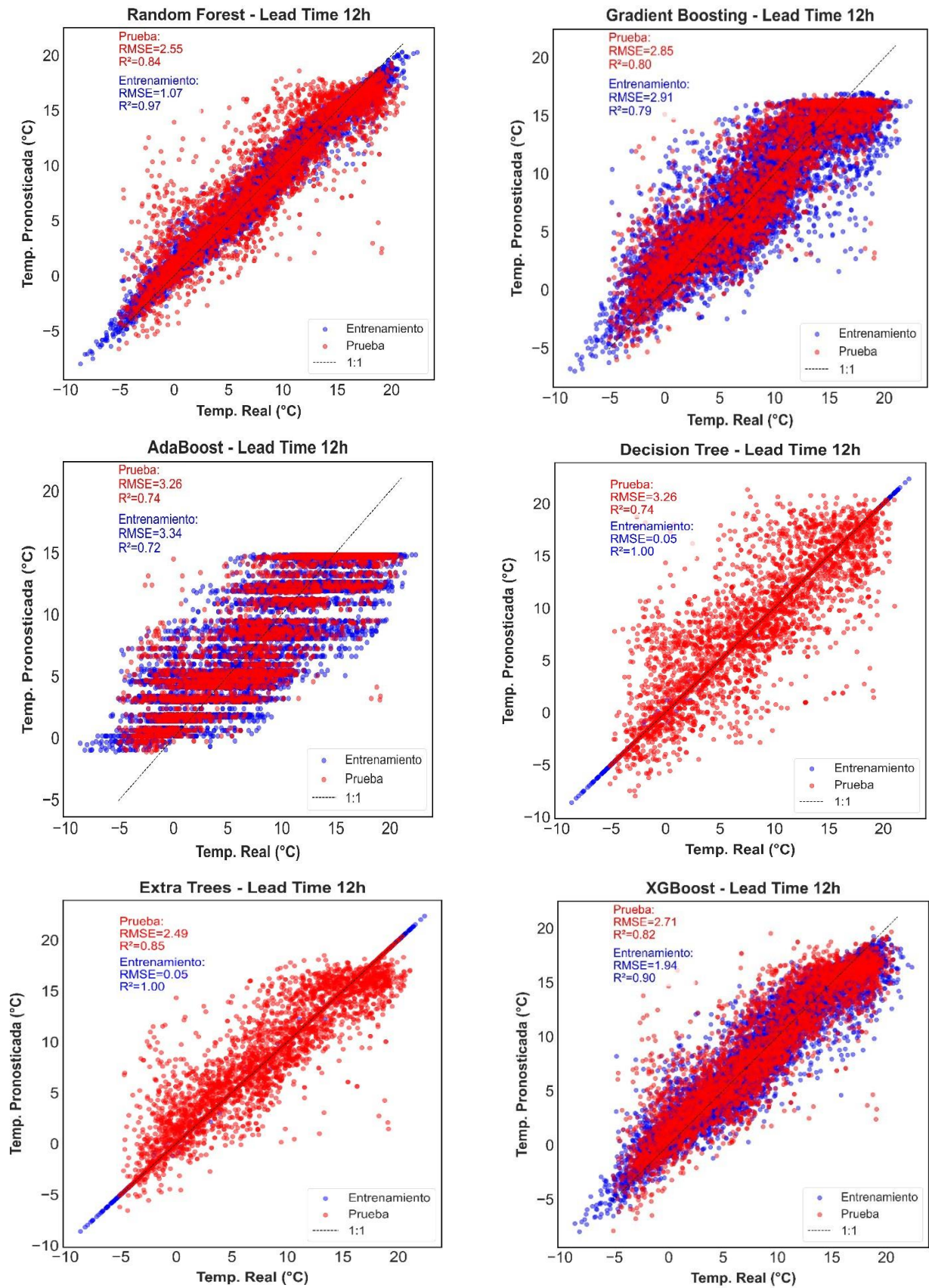


Figura 28. Pronóstico de heladas con 12 horas de anticipación.

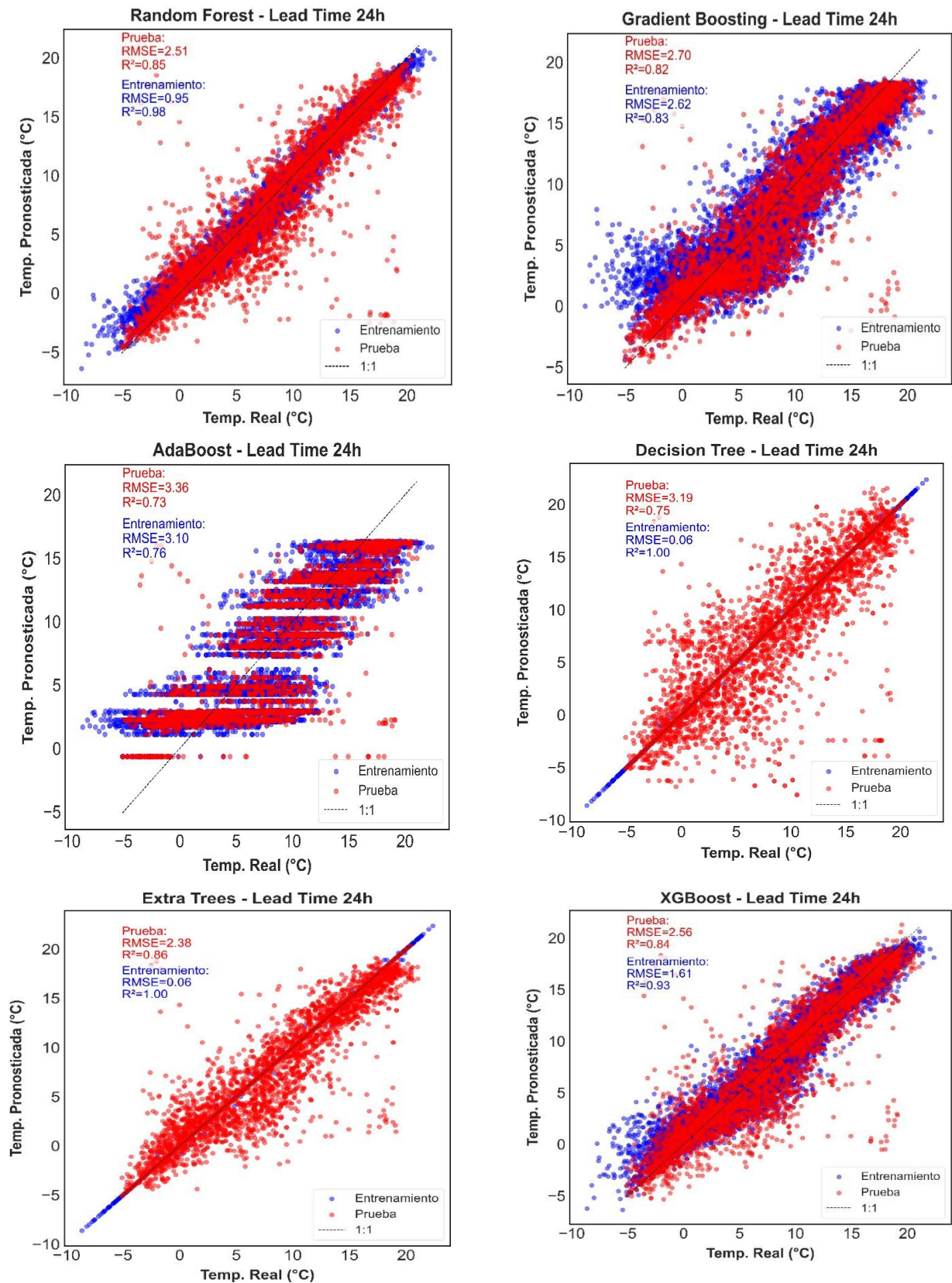


Figura 29. Pronóstico de heladas con 24 horas de anticipación.

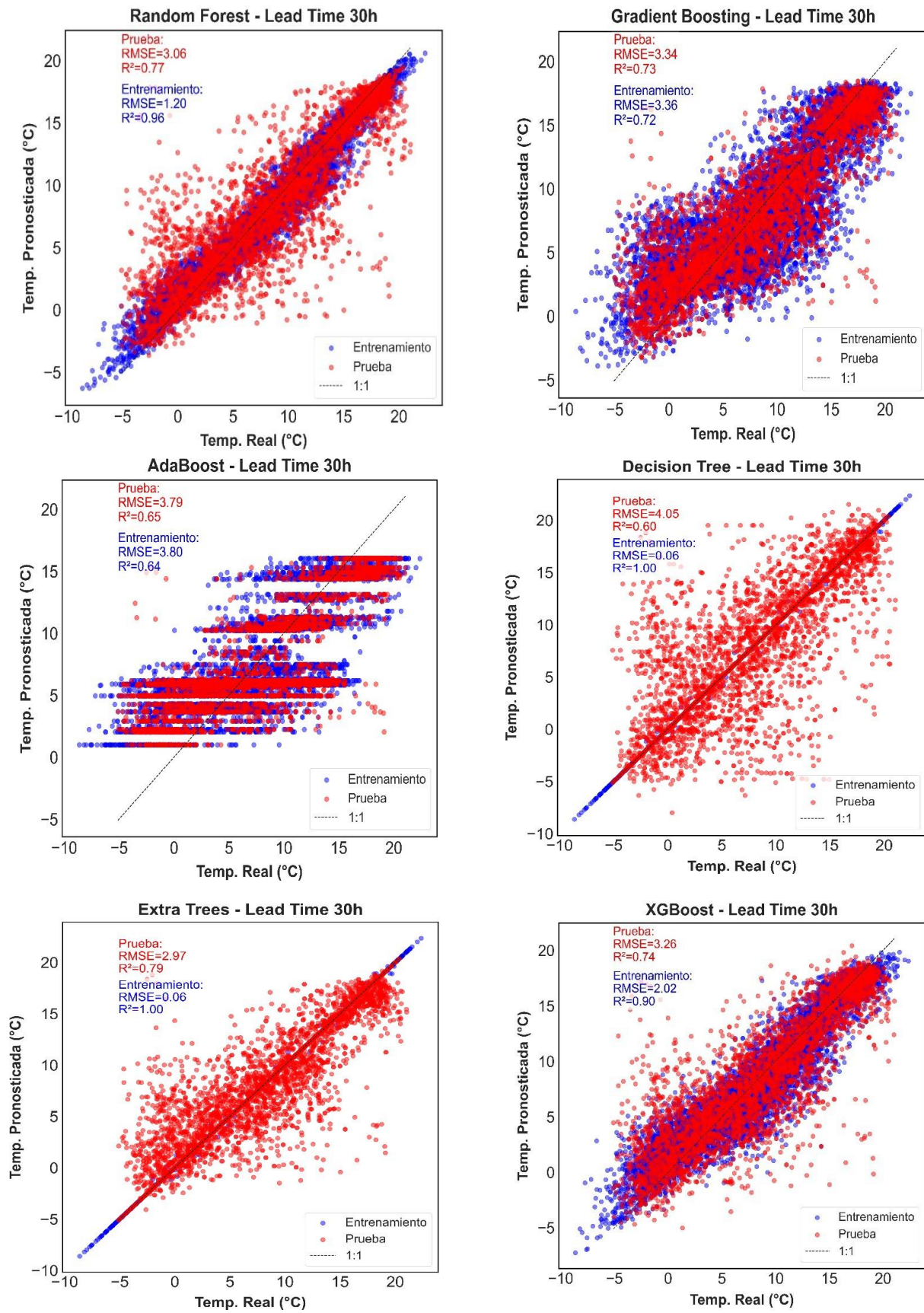


Figura 30. Pronóstico de heladas con 30 horas de anticipación.

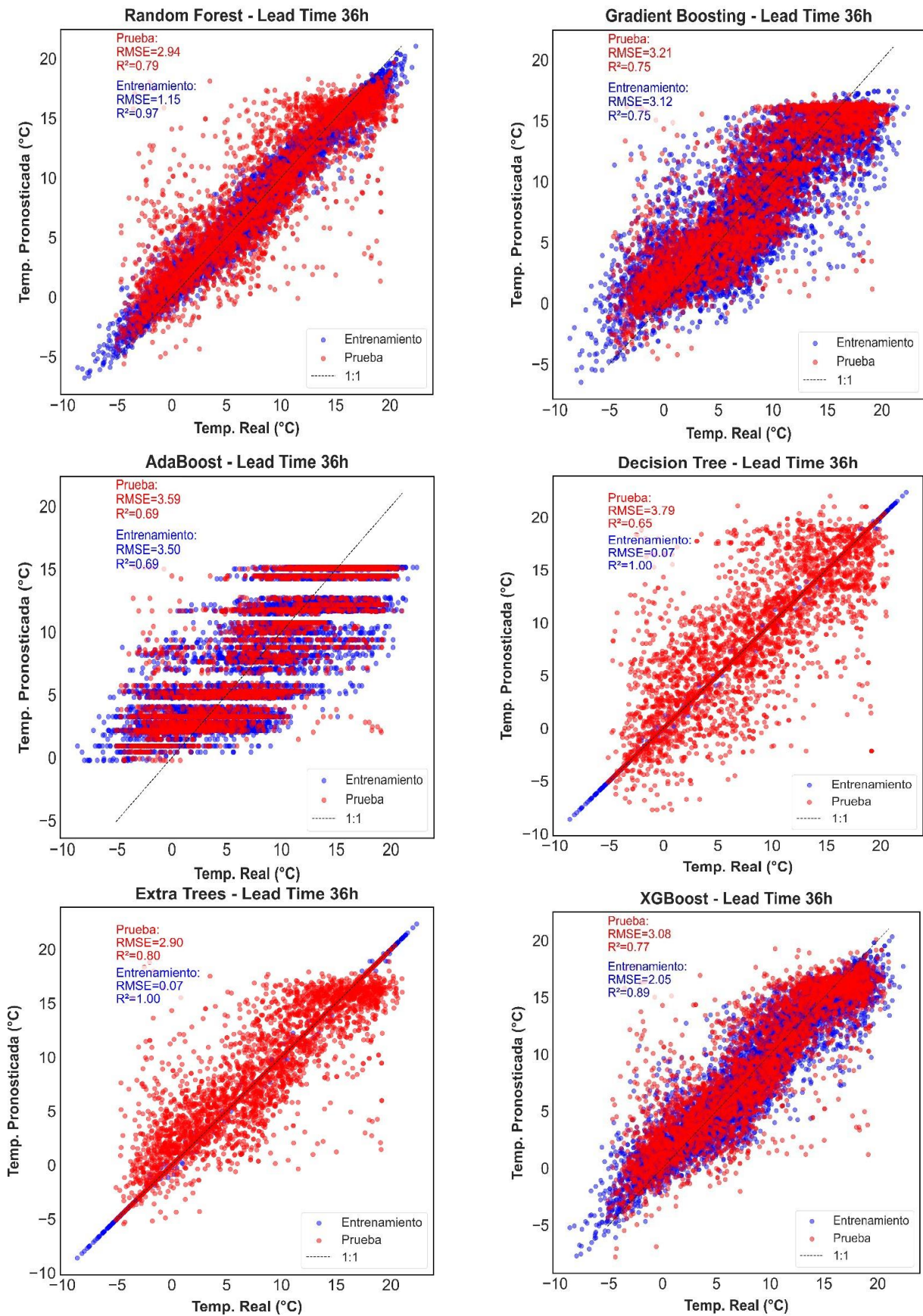


Figura 31. Pronóstico de heladas con 36 horas de anticipación.

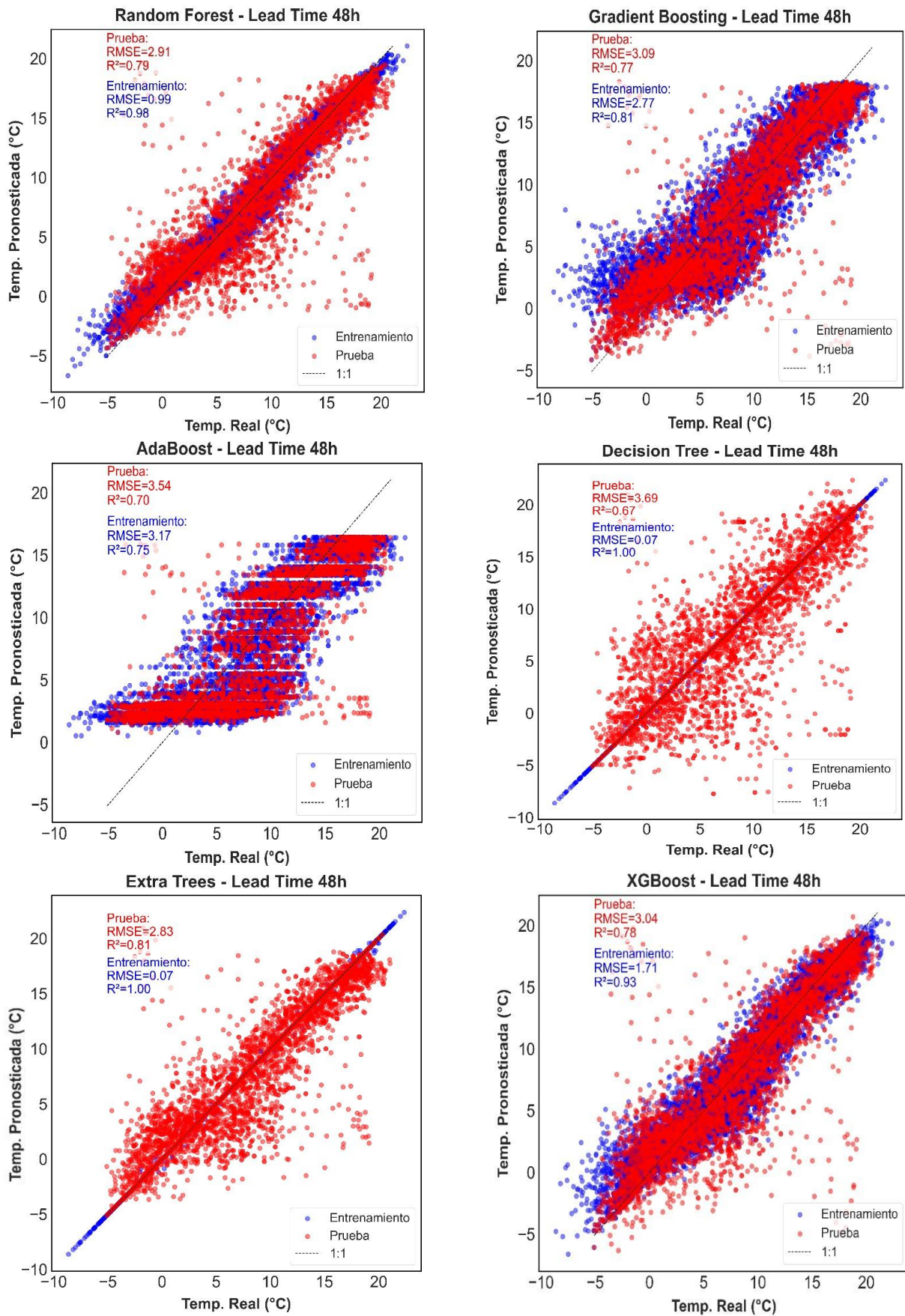


Figura 32. Pronóstico de heladas con 48 horas de anticipación.

Tabla 4. Resultados de pronóstico en una ventana de 6 horas.

Modelo	Lead Time	RMSE Prueba	R ² Prueba	RMSE Entrenamiento	R ² Entrenamiento
Random Forest	6h	2.7	0.82	1.12	0.97
Gradient Boosting	6h	2.98	0.79	3.12	0.75
AdaBoost	6h	3.57	0.69	3.64	0.67
Decision Tree	6h	3.56	0.69	0.03	1
Extra Trees	6h	2.6	0.84	0.03	1
XGBoost	6h	2.85	0.80	1.94	0.91

Nota: Para pronosticar heladas con 6 horas de anticipación, se usaron datos de las últimas 6 a 12 horas antes de la predicción.

Según la tabla 4, para el pronóstico de heladas con 6 horas de anticipación, Random Forest es el modelo más preciso (RMSE: 2.7, R²: 0.82). XGBoost también tiene buen rendimiento (RMSE: 2.85, R²: 0.80). Extra Trees muestra sobreajuste con un R² de 1 en entrenamiento pese a un RMSE de 2.6. Gradient Boosting y AdaBoost tienen menor desempeño (RMSE hasta 3.64, R² bajo). Decision Tree sufre sobreajuste (R²: 1 en entrenamiento) y un alto RMSE en prueba (3.56). Lo cual se puede evidenciar en los gráficos de dispersión (Predicción vs. Valor Real) de la figura 27.

Tabla 5. Resultados de pronóstico de heladas en una ventana de 12 horas.

Modelo	Lead Time	RMSE Prueba	R ² Prueba	RMSE Entrenamiento	R ² Entrenamiento
Random Forest	12h	2.55	0.84	1.07	0.97
Gradient Boosting	12h	2.85	0.80	2.91	0.79
AdaBoost	12h	3.26	0.74	3.34	0.72
Decision Tree	12h	3.26	0.74	0.05	1
Extra Trees	12h	2.49	0.85	0.05	1
XGBoost	12h	2.71	0.82	1.71	0.9

Nota: Para pronosticar heladas con 12 horas de anticipación, se usaron datos de las últimas 12 a 24 horas antes de la predicción.

Según la Tabla 5 y la Figura 28, Random Forest es el modelo más preciso para el pronóstico de heladas a 12 horas, con un RMSE de 2.55 y R² de 0.84. XGBoost también tiene buen desempeño con RMSE de 2.71 y R² de 0.82. Extra Trees muestra

el RMSE más bajo (2.49), pero junto con Decision Tree, presenta sobreajuste (R^2 de 1 en entrenamiento), lo que indica poca generalización. Gradient Boosting y AdaBoost tienen el peor rendimiento, con RMSE hasta 3.26 y R^2 de 0.80 y 0.74, demostrando menor precisión en la predicción de heladas.

Tabla 6. Resultados de pronóstico de heladas en una ventana de 24 horas.

Modelo	Lead Time	RMSE Prueba	R^2 Prueba	RMSE Entrenamiento	R^2 Entrenamiento
Random Forest	24h	2.51	0.85	0.95	0.98
Gradient Boosting	24h	2.70	0.82	2.62	0.83
AdaBoost	24h	3.36	0.73	3.10	0.76
Decision Tree	24h	3.19	0.75	0.06	1
Extra Trees	24h	2.38	0.86	0.06	1
XGBoost	24h	2.56	0.84	1.61	0.93

Nota: Para pronosticar heladas con 12 horas de anticipación, se usaron datos de las últimas 24 a 48 horas antes de la predicción.

Según la Tabla 6 y la Figura 29, Random Forest fue el modelo más preciso para predecir heladas a 24 horas, seguido de XGBoost. Extra Trees tuvo el menor RMSE, pero mostró sobreajuste. Gradient Boosting tuvo precisión moderada, mientras que AdaBoost y Decision Tree fueron los menos efectivos.

Tabla 7. Resultados de pronóstico de heladas en una ventana de 30 horas.

Modelo	Lead Time	RMSE Prueba	R^2 Prueba	RMSE Entrenamiento	R^2 Entrenamiento
Random Forest	30h	3.06	0.77	1.2	0.96
Gradient Boosting	30h	3.34	0.73	3.36	0.72
AdaBoost	30h	3.79	0.65	3.80	0.64
Decision Tree	30h	4.05	0.60	0.06	1
Extra Trees	30h	2.97	0.79	0.06	1
XGBoost	30h	3.26	0.74	2.02	0.90

Nota: Para pronosticar heladas con 30 horas de anticipación, se usaron datos de las últimas 30 a 60 horas antes de la predicción.

Según la Tabla 7 y la visualización de la figura 30, Random Forest es el modelo más efectivo para predecir heladas a 30 horas, con un RMSE de 3.06 y R^2 de 0.77, mostrando buena generalización. Extra Trees tiene el RMSE más bajo (2.97), pero sufre de sobreajuste (R^2 de 1 en entrenamiento). XGBoost ofrece un rendimiento intermedio con RMSE de 3.26 y R^2 de 0.74. Gradient Boosting y AdaBoost presentan menor precisión, con RMSE de 3.34 y 3.79, y R^2 de 0.73 y 0.65, respectivamente. Decision Tree tiene el peor desempeño en prueba, con RMSE de 4.05 y R^2 de 0.60, indicando sobreajuste.

Tabla 8. Resultados de pronóstico de heladas en una ventana de 36 horas.

Modelo	Lead Time	RMSE Prueba	R^2 Prueba	RMSE Entrenamiento	R^2 Entrenamiento
Random Forest	36h	2.94	0.79	1.15	0.97
Gradient Boosting	36h	3.21	0.75	3.12	0.75
AdaBoost	36h	3.59	0.69	3.50	0.69
Decision Tree	36h	3.79	0.65	0.07	1
Extra Trees	36h	2.90	0.80	0.07	1
XGBoost	36h	3.08	0.77	2.05	0.89

Nota: Para pronosticar heladas con 36 horas de anticipación, se usaron datos de las últimas 36 a 72 horas antes de la predicción.

Según la tabla 8 y la figura 31, para el pronóstico de heladas a 36 horas, Random Forest es el mejor modelo con un RMSE de 2.94 y un R^2 de 0.79, lo que refleja alta precisión y buena generalización. Extra Trees también es efectivo con un RMSE de 2.9 y un R^2 de 0.8, pero presenta sobreajuste. XGBoost tiene un RMSE de 3.08 y un R^2 de 0.77, con un rendimiento ligeramente inferior. Gradient Boosting y AdaBoost tienen un RMSE de 3.21 y 3.59, respectivamente, con R^2 bajos. Decision Tree tiene el peor desempeño con un RMSE de 3.79 y R^2 de 0.65 en prueba.

Tabla 9. Resultados de pronóstico de heladas en una ventana de 48 horas.

Modelo	Lead Time	RMSE Prueba	R ² Prueba	RMSE Entrenamiento	R ² Entrenamiento
Random Forest	48h	2.91	0.79	0.99	0.98
Gradient Boosting	48h	3.09	0.77	2.77	0.81
AdaBoost	48h	3.54	0.70	3.17	0.75
Decision Tree	48h	3.69	0.67	0.07	1
Extra Trees	48h	2.83	0.81	0.07	1
XGBoost	48h	3.04	0.78	1.71	0.93

Nota: Para pronosticar heladas con 48 horas de anticipación, se usaron datos de las últimas 48 a 96 horas antes de la predicción.

Como se visualizó en la tabla 9 y la figura 32. Para el pronóstico de heladas a 48 horas de anticipación, Random Forest es el mejor modelo con un RMSE de 2.91 y un R² de 0.79 en prueba, destacando por su alta precisión y capacidad de generalización. Extra Trees sigue de cerca con un RMSE de 2.83 y un R² de 0.81, pero presenta sobreajuste en entrenamiento. XGBoost tiene un RMSE de 3.04 y un R² de 0.78, mostrando un rendimiento algo inferior. Gradient Boosting y AdaBoost presentan un rendimiento más bajo con RMSE de 3.09 y 3.54, respectivamente, y R² menores. Decision Tree tiene el peor desempeño con un RMSE de 3.69 y un R² de 0.67 en prueba.

Pronóstico de heladas en fechas establecidas con el modelo seleccionado

Tras verificar que Random Forest es el modelo más adecuado para los datos de Huayao debido a su alto desempeño en R² y RMSE, y una distribución de predicciones cercana a la línea 1:1, se utilizó para pronosticar heladas en un período no visto durante el entrenamiento. Este enfoque evaluó la capacidad de generalización del modelo en escenarios reales, ofreciendo una herramienta confiable para la predicción de heladas. Esto facilita la toma de decisiones en la gestión de riesgos climáticos en el anexo Huayao, Chupaca.

Se implementó un flujo completo para predecir la temperatura mínima diaria utilizando el modelo RandomForestRegressor y técnicas de preprocesamiento y evaluación. Se importaron librerías esenciales como pandas, numpy, matplotlib, seaborn, scikit-learn y XGBoost. Además, se definió una función para calcular la curva de Lorenz y el coeficiente de Gini, midiendo así la distribución de la importancia de las variables.

Se cargó el archivo "data_modified.csv", estableciendo la primera columna como índice y parseando las fechas para trabajar con series temporales. Los datos se ordenaron cronológicamente y se agruparon a nivel diario, con la temperatura mínima diaria como variable objetivo. Se aplicaron métodos de agregación a otras variables y se eliminaron las filas con valores nulos para asegurar la integridad del conjunto de datos.

Posteriormente se normalizaron las variables con StandardScaler y se usó un RandomForestRegressor con 100 árboles para predecir la temperatura mínima diaria entre julio 2018 y julio 2022. Se evaluó la importancia de los predictores con la curva de Lorenz, el coeficiente de Gini y gráficos de dependencia parcial. Como se muestra en las siguientes figuras.

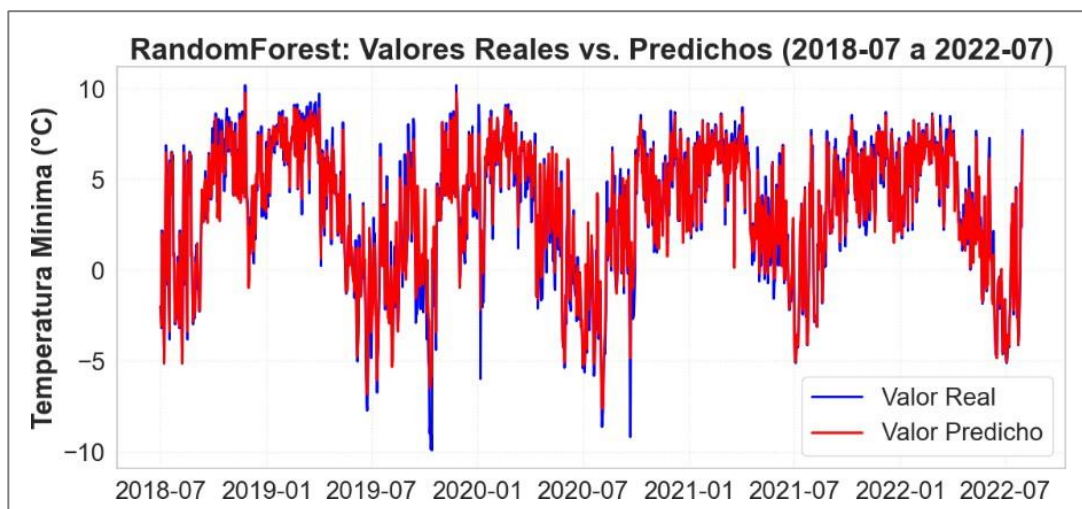


Figura 33. Pronóstico de heladas con datos no vistos en el entrenamiento.

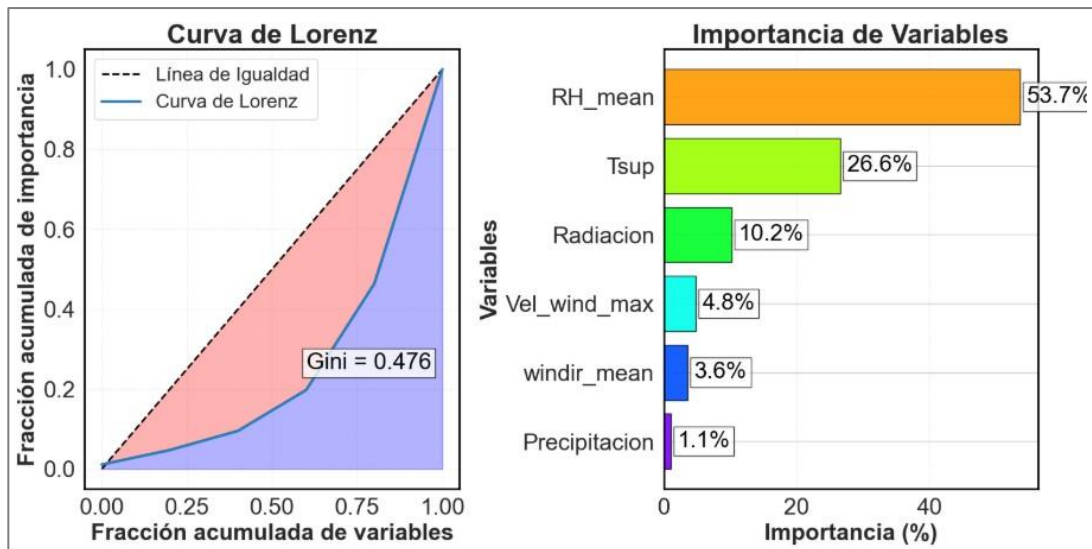


Figura 34. Pronóstico de heladas con datos no vistos en el entrenamiento.

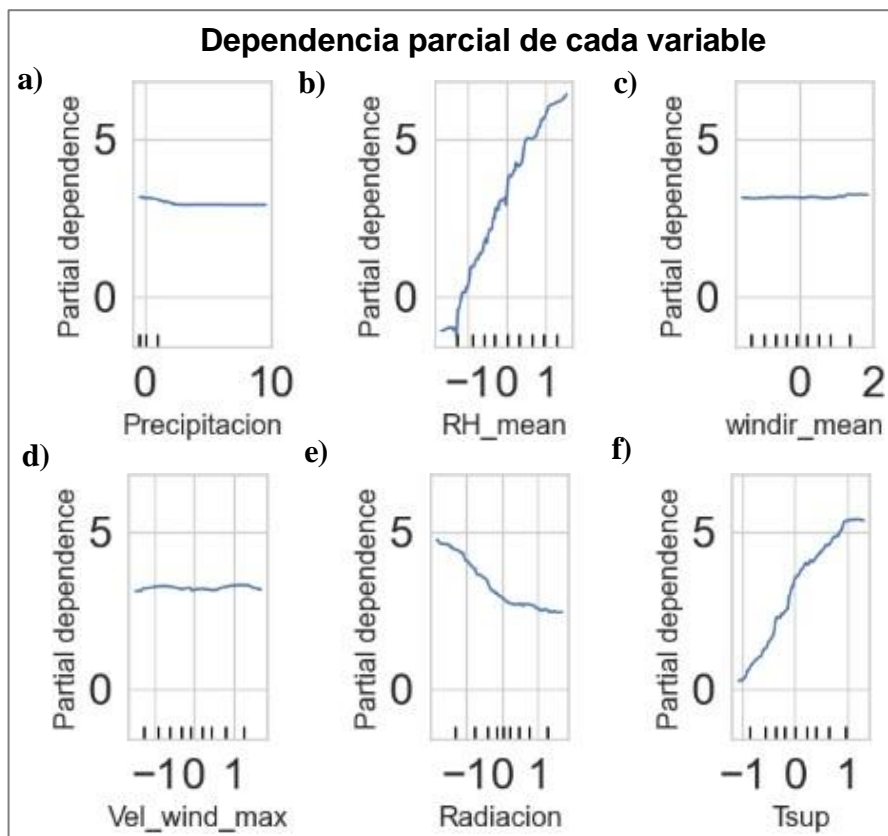


Figura 35. Gráficos de dependencia parcial de las variables durante el pronóstico de heladas con el modelo Radom Forest. a) dependencia parcial de la variable precipitación, b) dependencia parcial de la variable humedad relativa, c) dependencia parcial de la variable dirección media del viento, d) dependencia parcial de la variable velocidad máxima del viento, e) dependencia parcial de la variable radiación, f) dependencia parcial de la variable temperatura superficial.

En las figuras 33 se visualiza, en primer lugar, la comparación entre los valores reales de la temperatura mínima (línea azul) y las predicciones generadas por el modelo Random Forest (línea roja) para el periodo comprendido entre julio de 2018 y julio de 2022. Se aprecia que ambas líneas siguen una trayectoria muy similar, lo que sugiere que el modelo logra capturar de manera satisfactoria la variabilidad diaria de la temperatura mínima. Aun cuando se observan pequeñas discrepancias en ciertos tramos, la tendencia general indica una predicción bastante cercana a la realidad, reflejando la consistencia del modelo en distintos escenarios temporales.

En la figura 34, se observa la Curva de Lorenz pone de manifiesto que la importancia de las variables está algo concentrada, con un índice de Gini de 0.476. Este valor señala que un conjunto reducido de variables concentra gran parte de la capacidad explicativa del modelo. El gráfico de barras confirma que la humedad relativa ocupa el primer lugar, con alrededor de un 53.7 % de importancia, seguida por la temperatura superficial, con un 26.6 %. La radiación (10.2 %) ocupa el tercer lugar, y en menor medida se encuentran la velocidad máxima del viento (4.8 %), la dirección media del viento (3.6 %) y la precipitación (1.2 %). Esta distribución indica que el modelo se apoya fuertemente en la humedad y la temperatura superficial para estimar la temperatura mínima, mientras que el resto de las variables aporta información secundaria.

Finalmente, en la figura 35 la dependencia parcial muestra cómo cada variable influye individualmente en la predicción de la temperatura mínima. Se observa que, en el caso de humedad relativa, a medida que aumenta la humedad relativa, la temperatura mínima estimada también tiende a incrementarse, lo cual coincide con la lógica física de que una atmósfera más húmeda suele moderar los descensos de

temperatura. Por su parte, la radiación presenta un comportamiento inverso: al aumentar, la temperatura mínima se reduce ligeramente, sugiriendo que las condiciones de mayor radiación durante el día pueden favorecer un mayor enfriamiento nocturno. La temperatura superficial muestra un comportamiento ascendente: a mayor temperatura superficial, el modelo predice valores más altos para la temperatura mínima, reflejando la influencia directa de la temperatura del entorno en la formación de heladas. En conjunto, estos resultados confirman que el modelo Random Forest no solo logra un buen ajuste global, sino que también revela claramente cuáles son los factores que más contribuyen a la predicción de las temperaturas mínimas en el área de estudio.

5.4.2 Evaluación experimental de la precisión y confiabilidad de modelos de IA en dos casos aplicativos

En este apartado se presenta la evaluación experimental de la precisión y confiabilidad de los modelos de inteligencia artificial aplicados a dos escenarios específicos. El primero de ellos aborda el fenómeno de inversión térmica, un caso de estudio particularmente relevante en condiciones invernales.

Planteamiento de la Situación. - Fenómeno de Inversión Térmica

Durante un día invernal, se registró un fenómeno de inversión térmica caracterizado por la formación de una capa de aire frío en el anexo de Huachac de la superficie, tras el ocaso de la radiación solar. La ausencia de radiación, combinada con la presencia de nubes bajas o nieblas, impidió la adecuada mezcla entre las capas de aire frío y caliente, generando cambios significativos en las variables climáticas de interés:

- **Temperatura Superficial (Tsup).** - La inversión térmica propicio un descenso abrupto de la temperatura en la superficie durante la noche, puesto que el calor queda confinado en las capas superiores.
- **Radiación.** - La desaparición de la radiación solar redujo notablemente la energía recibida en la superficie, un efecto acentuado por las condiciones nubosas o brumosas asociadas a la inversión.
- **Humedad Relativa.** - El enfriamiento del aire indujo la condensación del vapor de agua, lo cual incrementa la humedad relativa y favorece la formación de rocío o, en situaciones extremas, de heladas.

Como consecuencia de estas condiciones, la variable temperatura mínima, alcanza niveles críticos, facilitando la aparición de heladas durante la madrugada. Con base en este escenario, se procedió a ajustar los datos de entrada del modelo de predicción para simular de manera precisa la influencia de una inversión térmica en la estimación de las temperaturas mínimas.

Se realizó la simulación de condiciones de inversión térmica mediante varios módulos clave de Python: se empleó el módulo random para generar valores aleatorios dentro de rangos predefinidos (como la tasa de enfriamiento, la reducción del viento, el incremento de la humedad y la radiación), y numpy para manejar constantes numéricas (por ejemplo, usar np.inf para establecer el error inicial) y operaciones matemáticas. Se procesaron los datos con pandas, simulando cada registro mediante el método apply, y se normalizaron con StandardScaler de scikit-learn para luego realizar predicciones de temperaturas mínimas usando RandomForestRegressor.



Figura 36. Pronostico horario de heladas en un escenario de inversión térmica.

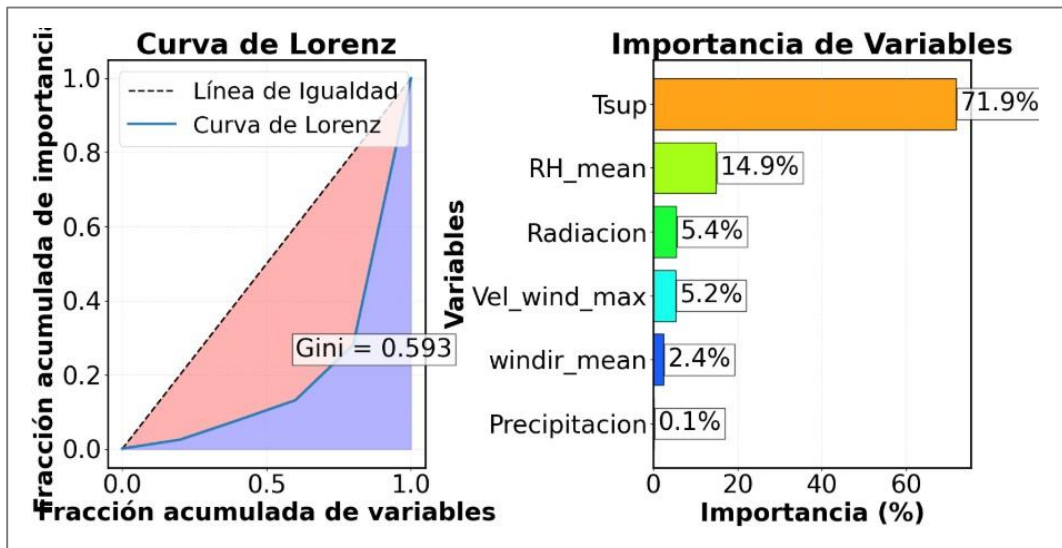
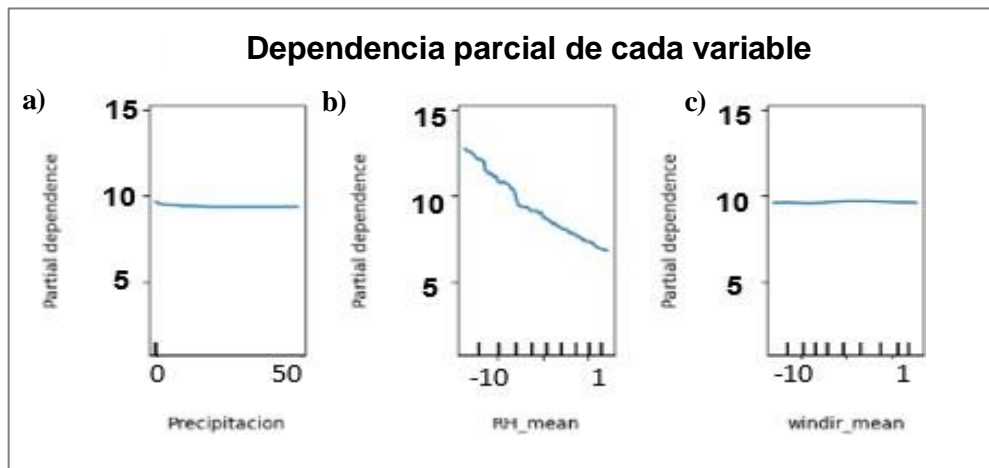


Figura 37. Curva de Lorenz y Coeficiente de importancia de Gini, experimento 01.



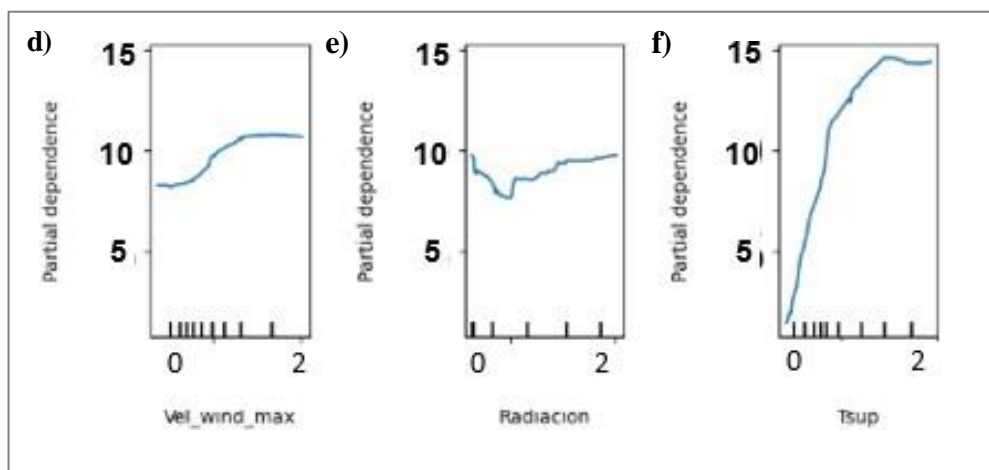


Figura 38. Gráfico de dependencia parcial de las variables, experimento 01. a) dependencia parcial de la variable precipitación, b) dependencia parcial de la variable humedad relativa, c) dependencia parcial de la variable dirección media del viento, d) dependencia parcial de la variable velocidad máxima del viento, e) dependencia parcial de la variable radiación, f) dependencia parcial de la variable temperatura superficial.

Tabla 10. Resultados del pronóstico en la prueba de simulación 01

Modelo	RMSE
Random	43.082358
Forest	

La simulación de inversión térmica que se observa en la figura 36 muestra como resultado en la tabla 10 que el modelo RandomForest, entrenado con variables como temperatura superficial, humedad relativa, radiación, velocidad máxima del viento, dirección media del viento y precipitación, logró predecir la temperatura mínima horaria entre las 18:00 del 13 de octubre y las 06:00 del 14 de octubre con un error cuadrático medio (MSE) de 43.082358; las gráficas evidencian que la temperatura superficial es la variable más influyente (71.9%), seguida por humedad relativa (14.9%) y radiación (5.4%), lo cual indica que la temperatura superficial y la humedad relativa son factores críticos en la formación de heladas. Además, la curva de Lorenz (Gini=0.593) muestra una concentración moderada de la importancia en pocas variables, mientras que los gráficos de dependencia parcial sugieren que a mayor humedad y menor temperatura superficial se intensifica el

enfriamiento. Al comparar la serie real con la predicción, se observa que la helada inicia aproximadamente a las 22:00 y finaliza alrededor de las 05:00, con una duración cercana a 7 horas, confirmando la capacidad del modelo para detectar y pronosticar heladas en escenarios de inversión térmica.

El segundo caso experimental se centró en lo siguiente

Planteamiento de la Situación. - Fenómeno del niño

Durante el período comprendido entre 2018 y 2022 se registró un evento de El Niño que alteró significativamente los patrones climáticos en la región. En ese intervalo se observó que:

- **Temperatura Superficial (Tsup).** - Hubo un incremento promedio de alrededor de +2°C, evidenciando aguas más cálidas y un ambiente general más caluroso.
- **Precipitación:** Se registró un aumento en la precipitación en aproximadamente un 50%, lo que generó condiciones de mayor humedad y riesgo de inundaciones.
- **Humedad Relativa (RH_mean).** - La humedad aumentó en unos 10 puntos porcentuales, alcanzando en ocasiones el límite del 100%.
- **Viento.** - La velocidad del viento (tanto la media como el valor máximo) se redujo en un 30%, contribuyendo a la estabilidad de la atmósfera.
- **Radiación Solar.** - La mayor nubosidad asociada a El Niño provocó una reducción de la radiación en torno al 20%.

Con base en estas observaciones, se ajustarán los datos diarios de la base para simular las condiciones asociadas al fenómeno de El Niño, y se evaluará el impacto sobre la predicción de la temperatura mínima (temp_min) utilizando un modelo RandomForest.

La simulación del fenómeno de El Niño se realizó en Python modificando el DataFrame original mediante pandas: se ajustaron directamente las columnas de precipitación y velocidad máxima del viento para generar un conjunto de datos que reflejara condiciones climáticas extremas, típicas de este fenómeno. En concreto, se multiplicaron los valores de precipitación por 4, lo que se justificó porque durante eventos de El Niño se observa un aumento drástico en las lluvias, representando un escenario de precipitación intensificada; de igual forma, se duplicaron los valores de la velocidad máxima del viento para simular la intensificación de los vientos asociados a estas condiciones. Posteriormente, estos datos simulados se sometieron a un proceso de normalización utilizando StandardScaler de scikit-learn, lo que permitió homogeneizar las escalas de las diferentes variables y facilitar el entrenamiento del modelo. Con los datos normalizados, se entrenó el modelo RandomForestRegressor, también de scikit-learn, con el objetivo de predecir la temperatura mínima horaria, comparando los valores reales con los predichos para evaluar la precisión del modelo. Este enfoque permitió no solo simular de manera realista el impacto de El Niño sobre las variables climáticas críticas, sino también analizar el rendimiento del modelo bajo condiciones de alta variabilidad, aportando información relevante sobre la capacidad predictiva del modelo en escenarios extremos.

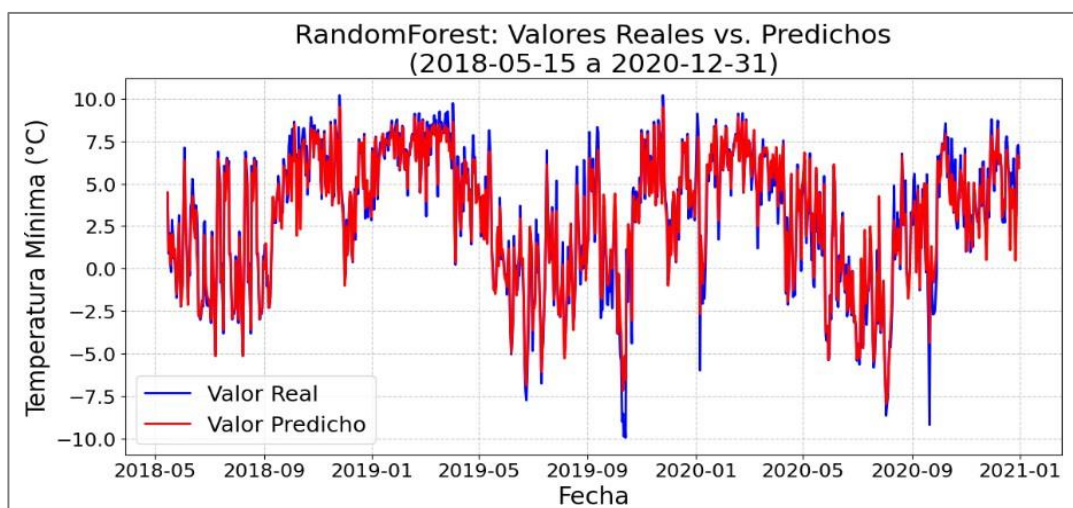


Figura 39. Pronosticó de heladas durante un fenómeno simulado del niño, experimento 02.

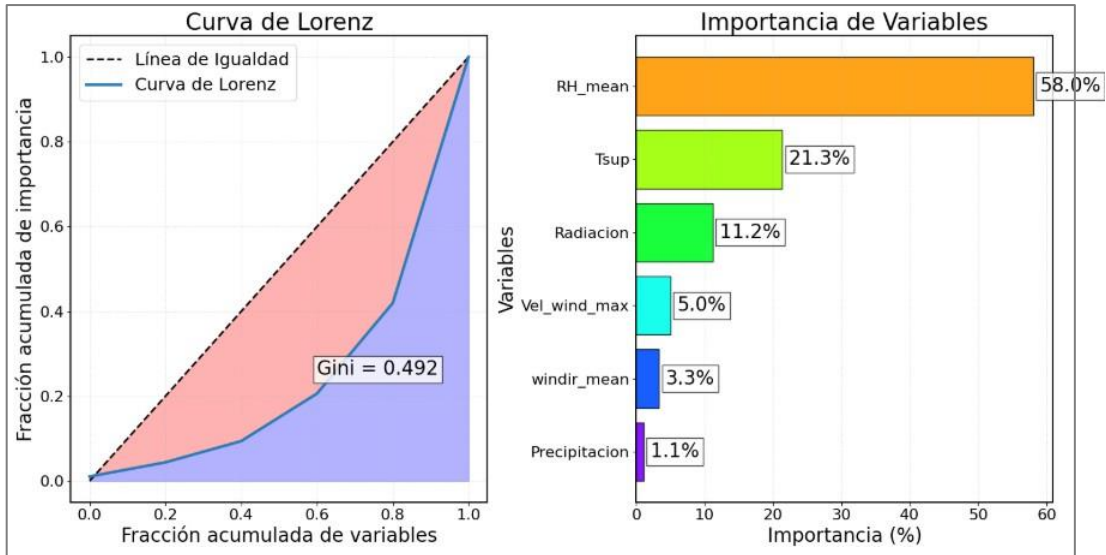


Figura 40. Curva de Lorenz y coeficiente de Gini para el experimento 02.

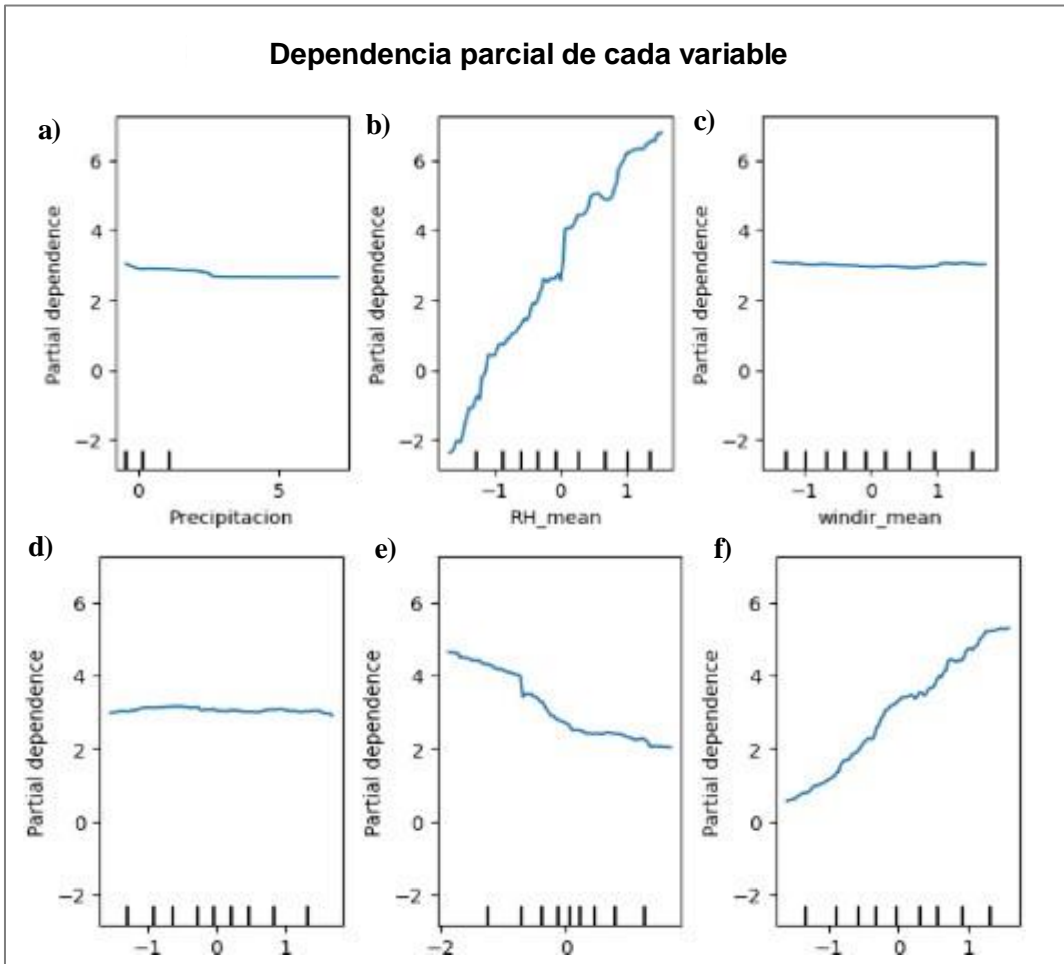


Figura 41. Gráficos de dependencia parcial para las variables en el experimento 02. a) dependencia parcial de la variable precipitación, b) dependencia parcial de la variable humedad relativa, c) dependencia parcial de la variable dirección media del viento, d) dependencia parcial de la variable velocidad máxima del viento, e) dependencia parcial de la variable radiación, f) dependencia parcial de la variable temperatura superficial.

En la figura 39 se observa la comparación entre los valores reales y los predichos de la temperatura mínima para el período comprendido entre el 15 de mayo de 2018 y el 31 de diciembre de 2020, evidenciando que el modelo RandomForest logra capturar la tendencia general de la serie, aunque se presentan ciertas discrepancias en períodos puntuales. Además, las métricas de desempeño indican un Error Absoluto Medio (MAE) de 0.48, lo que significa que, en promedio, la predicción se desvía menos de medio grado Celsius del valor real, y un coeficiente de determinación (R^2) de 0.9676, demostrando que el modelo explica aproximadamente el 96.76 % de la variabilidad observada en los datos. En la figura 40, la Curva de Lorenz presenta un coeficiente de Gini de 0.492, lo que indica una concentración moderada de la importancia en pocas variables, mientras que el diagrama de barras destaca a humedad relativa como la variable más influyente (58 %), seguida de temperatura superficial (21.3 %), radiación (11.2 %), velocidad máxima del viento (5.0 %), dirección del viento (3.3 %) y precipitación (1.1 %). Estos resultados sugieren que la humedad relativa y la temperatura superficial son factores clave en la predicción de la temperatura mínima, siendo especialmente relevante el efecto de la humedad en la formación de heladas o condiciones frías. Finalmente, los gráficos de dependencia parcial muestran cómo la temperatura mínima predicha varía al modificar individualmente cada variable: por ejemplo, al aumentar la humedad relativa, la predicción de la temperatura mínima se modifica de forma marcada, mientras que la temperatura superficial tiene un efecto ascendente más uniforme, reflejando que un mayor calor en la superficie tiende a elevar la temperatura mínima. En contraste, la precipitación y el viento muestran una influencia menor o más estable, lo cual coincide con su baja importancia relativa en la predicción según el modelo

ANÁLISIS Y DISCUSIÓN DE RESULTADOS

Con base en los resultados obtenidos en la presente investigación, se presenta el análisis y discusión de los hallazgos, contrastándolos con antecedentes relevantes:

OE1.- Identificación de las variables climáticas relevantes

En la presente investigación, tras la etapa de análisis exploratorio y preprocesamiento de datos, se identificó que las variables más relevantes para la predicción de heladas en la cuenca del río Mantaro son la humedad relativa, la temperatura superficial y la radiación. Esto fue evidenciado a través del análisis de correlación de Spearman, donde la humedad relativa obtuvo un coeficiente de 0.71, lo que indica una relación positiva alta con la temperatura mínima, y por el análisis del coeficiente de Gini y la curva de Lorenz, donde estas variables concentraron más del 80% de la importancia del modelo.

Estos resultados coinciden con lo señalado por Transmonte (5) quien advirtió que la humedad relativa y la radiación solar son factores determinantes en la ocurrencia de heladas, al influir en la capacidad del ambiente para retener o perder calor durante la noche. Así también, se reafirma lo expuesto por estudios similares en zonas altoandinas donde se reconoce que mayores niveles de humedad pueden atenuar descensos bruscos de temperatura y, por ende, reducir la incidencia de heladas.

Por tanto, se confirma la hipótesis específica A, al demostrar que la correcta identificación de las variables climáticas clave permitió optimizar el proceso de predicción, mejorando la efectividad de los modelos de inteligencia artificial aplicados.

OE2.- Análisis de datos históricos y patrones climáticos

Durante la investigación se procesaron datos meteorológicos de los años 2018 a 2023, identificando la presencia de patrones asociados a heladas, especialmente durante los meses de junio, julio y agosto. Se reconocieron fenómenos como la inversión térmica,

caracterizados por abruptas disminuciones de la temperatura superficial y el incremento de la humedad relativa, eventos que desencadenan condiciones propicias para heladas prolongadas.

Estos hallazgos son coherentes con lo expuesto por IGP (10) que evidenció pérdidas agrícolas por heladas como consecuencia de la falta de detección oportuna de patrones críticos en variables como temperatura superficial y radiación. Así mismo, el análisis histórico permitió evaluar escenarios extremos como el fenómeno de El Niño, demostrando su impacto moderador sobre las heladas, debido al aumento de la temperatura superficial y la humedad relativa, lo que a su vez coincide con lo señalado en los reportes climatológicos del IGP y SENAMHI (10).

Estos resultados validan la hipótesis específica B, al evidenciar que el análisis de datos históricos permitió identificar patrones climáticos esenciales para el entrenamiento y precisión de los modelos predictivos.

OE3.- Evaluación de modelos de inteligencia artificial

Al comparar el desempeño de distintos algoritmos de inteligencia artificial aplicados a la predicción de heladas, se determinó que los modelos basados en técnicas de ensamblado, especialmente Random Forest y XGBoost, lograron los mejores resultados, con valores de R^2 superiores a 0.80 y RMSE cercanos a 1.58 °C en el conjunto de prueba. Además, Random Forest demostró una notable capacidad de generalización tanto en datos históricos como en condiciones simuladas de eventos extremos como inversión térmica y El Niño.

Estos resultados son consistentes con lo expuesto por Valencia (24) quien señaló que los modelos robustos, como Random Forest, son especialmente útiles para manejar datos meteorológicos con alta variabilidad y valores atípicos, al minimizar los errores de

predicción frente a eventos críticos. Asimismo, confirman la importancia de elegir modelos con mecanismos internos de regularización que prevengan el sobreajuste, como lo evidenció el menor desempeño de algoritmos como Decision Tree y Extra Trees, que presentaron inestabilidad y sobreajuste.

De este modo, se confirma la hipótesis específica C, al establecer que la evaluación de precisión y confiabilidad de los modelos permitió identificar a Random Forest como el más adecuado para la predicción de heladas en la cuenca del Mantaro.

OG.- Predicción de heladas con inteligencia artificial

El presente estudio permitió identificar que el desarrollo de modelos de inteligencia artificial aplicados a la cuenca del río Mantaro, Chupaca, 2024, permitió predecir de manera anticipada los eventos de heladas. Si bien estos modelos lograron procesar y analizar grandes volúmenes de datos meteorológicos históricos y actuales, generando alertas tempranas, su desempeño depende directamente de la calidad y continuidad de los registros disponibles. Esta limitación responde a la falta de un sistema robusto de monitoreo que asegure datos precisos y actualizados, afectando en determinados momentos la estabilidad y confiabilidad de las predicciones.

Ante las deficiencias detectadas, es fundamental reforzar la infraestructura de monitoreo meteorológico e integrar formalmente los modelos de inteligencia artificial en las estrategias regionales de gestión del riesgo agroclimático, asegurando tanto la recolección constante de datos representativos como el uso oportuno de las predicciones por parte de agricultores y autoridades para la toma de decisiones preventivas frente a eventos de heladas. La inteligencia artificial permite sistemas de alerta temprana que reducen los impactos socioeconómicos en la agricultura local, siendo clave para la seguridad agro productiva regional.

CONCLUSIONES

En relación con los objetivos planteados y la hipótesis formulada, esta investigación logró desarrollar modelos de inteligencia artificial capaces de predecir heladas en la cuenca del río Mantaro, Chupaca, 2024. A continuación, se presentan las principales conclusiones:

- A.** Se logró desarrollar modelos de inteligencia artificial capaces de predecir heladas en la cuenca del río Mantaro, Chupaca, 2024, destacando el modelo Random Forest como el de mejor desempeño, con valores de R^2 superiores a 0.80 y RMSE menores a 2 °C. Esto demuestra la viabilidad de aplicar inteligencia artificial como herramienta predictiva para emitir alertas tempranas y reducir el impacto de las heladas sobre la actividad agrícola.
- B.** Se identificó que las variables climáticas con mayor relevancia para la predicción de heladas son la humedad relativa, la temperatura superficial del suelo y la radiación, acumulando entre ellas más del 80 % de importancia dentro de los modelos predictivos, lo que confirma su influencia directa en la ocurrencia de heladas.
- C.** El análisis de los datos históricos permitió identificar patrones climáticos críticos que facilitan la detección anticipada de heladas, especialmente en los meses de junio, julio y agosto, evidenciando una fuerte relación entre bajas temperaturas mínimas, altos niveles de humedad y baja radiación nocturna.
- D.** La evaluación comparativa demostró que el modelo de inteligencia artificial, Random Forest, supera en precisión y estabilidad a los métodos tradicionales de pronóstico, logrando mantener un buen rendimiento incluso bajo condiciones extremas como fenómenos de inversión térmica o eventos de El Niño.

RECOMENDACIONES

Después de desarrollar los modelos de inteligencia artificial en la investigación, se recomienda a los futuros investigadores interesados en el tema, lo siguiente:

- Implementar los modelos de inteligencia artificial desarrollados como parte de un sistema oficial de alerta temprana de heladas en la cuenca del Mantaro, integrándolos en las estrategias de gestión del riesgo agroclimático para facilitar la toma de decisiones preventivas en el sector agrícola.
- Fortalecer la red de monitoreo meteorológico en la región, priorizando estaciones que registren de forma continua variables clave como temperatura superficial, humedad relativa y radiación, asegurando así datos de alta calidad para mantener la precisión de los modelos
- Actualizar periódicamente los modelos predictivos con los datos más recientes y realizar análisis continuos de patrones climáticos históricos, garantizando que los algoritmos se adapten a los cambios derivados de la variabilidad climática y fenómenos extremos.
- Capacitar a técnicos, agricultores y autoridades locales en la interpretación y uso de los pronósticos generados por los modelos de inteligencia artificial, asegurando su correcta aplicación para minimizar los daños provocados por las heladas y optimizar la gestión agroclimática.

REFERENCIAS BIBLIOGRÁFICAS

1. IGP. "Memoria del Subproyecto "Pronóstico estacional de lluvias y temperaturas en la cuenca del río Mantaro para su aplicación en la agricultura" 2007 - 2010. Online. Yamina Silva Vidal. Lima, Perú : Biblioteca Nacional del Perú, 2010. 500. ISBN 978-612-45795-3-0. Available from: <http://met.igp.gob.pe/proyectos/incagro/datos/libromemoria.pdf>
2. SAVAGE, M. J. Nowcasting daily minimum air and grass temperature. *International Journal of Biometeorology*. 1 February 2016. Vol. 60, no. 2, p. 183–194. DOI 10.1007/s00484-015-1017-7.
3. CÚELLAR, José Eloy; and MEDINA, Tulio Cecilio. Agrobiodiversidad, género y cambio climático en la cuenca del río Mantaro. *Tecnología y Sociedad*. 2009. Vol. 16, no. 8, p. 83–99. Cuellar-et-al_2009_Agrobiodiversidad_Valle_Mantaro.pdf
4. ALARCÓN, Velazco, CONSTANTINO, Eusebio and TREBEJO, Irene. Atlas de Helada del Perú Online. Técnico. Lima, Perú : Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) y FAO, 2010. Available from: https://hdl.handle.net/20.500.12542/359Atlas-helada-Perú_2010.pdf
5. TRANSMONTE, Lilian. Propuesta de gestión de riesgo de heladas que afectan a la agricultura del valle del Mantaro. Online. Tesis. Lima, Perú : Universidad Ricardo Palma, 2009. Available from: <http://hdl.handle.net/20.500.12816/393>
6. MORENO, Polivio. Las heladas como resultado del cambio climático y su consecuencia en la agricultura del cantón Salcedo, Cotopaxi, período 1981-2010. Online. Tesis de maestría. Latacunga – Ecuador : Universidad Técnica de Cotopaxi, 2013. Available from: <http://repositorio.utc.edu.ec/handle/27000/7209>
7. ZHOU, Ian, LIPMAN, Justin, ABOLHASAN, Mehran, SHARIATI, Negin and LAMB, David. Frost Monitoring Cyber–Physical System: A Survey on Prediction and Active Protection Methods. *IEEE Internet of Things Journal*. July 2020. Vol. 7, no. 7, p. 6514–6527. DOI 10.1109/JIOT.2020.2972936.
8. MIDAGRI. Situación actual del Sector Agrario Referente al Cambio Climático. Online. Seminario. Lima, Perú. 2008. Available from: https://www.midagri.gob.pe/portal/download/pdf/especiales/bioenergia/cambio_climatico/avances.pdf
9. FAO. Protección contra las heladas: fundamentos, práctica y economía. Online. Roma : Organización de las Naciones Unidas para la Agricultura y la Alimentación, 2010. ISBN 978-92-5-306504-2. Available from: <https://www.fao.org/4/y7223s/y7223s.pdf>
10. IGP. Eventos meteorológicos extremos (sequías, heladas y lluvias intensas) en el valle del Mantaro. Online. María Villaverde Calderón. Lima, Perú : Biblioteca Nacional del Perú, 2012. ISBN 978-612-45795-6-1. Available from: [http://hdl.handle.net/20.500.12816/740Instituto Geofísico del Perú](http://hdl.handle.net/20.500.12816/740Instituto_Geofísico_del_Perú)

11. LLACZA, Alan, ACUÑA, Delia, JÁCOME, Gerardo, DE LA CRUZ, Gustavo, PAREDES, Jonathan, BRUNO, Jhonatan, ALVAREZ, Erick, FLORES, Weidi, URDANIVIA, Francesco and SULCA, Sheylla. Escenarios climáticos al 2050 en el Perú: cambios en el clima promedio Online. Lima, Perú : Servicio Nacional de Meteorología e Hidrología del Perú, SENAMHI., 2021. Available from: <https://hdl.handle.net/20.500.12542/1470>
12. TAKAHASHI, Ken and ANICAMA, Jahir. Inteligencia humana y artificial para la democratización del conocimiento científico sobre el clima. Online. 15 March 2024. Instituto Geofísico del Perú - IGP. Available from: <https://repositorio.igp.gob.pe/handle/20.500.12816/5539>
13. ROZANTE, José Roberto, RAMIREZ, Enver, RAMIREZ, Diego and ROZANTE, Gabriela. Improved frost forecast using machine learning methods. *Artificial Intelligence in Geosciences*. 1 December 2023. Vol. 4, p. 164–181. DOI 10.1016/j.aiig.2023.10.001.
14. TALSMA, Carl J., SOLANDER, Kurt C., MUDUNURU, Maruti K., CRAWFORD, Brandon and POWELL, Michelle R. Frost prediction using machine learning and deep neural network models. *Frontiers in Artificial Intelligence*. Online. 2023. Vol. 5. DOI 10.3389/frai.2022.963781.
15. CALDERÓN, Evelin. Predicción temprana de heladas en cultivos de altura, empleando métodos de aprendizaje de máquinas. Online. Medellín, Colombia : Universidad Nacional de Colombia, 2022. Available from: <https://repositorio.unal.edu.co/bitstream/handle/unal/83615/1152707862.2023.pdf?sequence=2&isAllowed=y>
16. ZHOU, Ian, LIPMAN, Justin, ABOLHASAN, Mehran and SHARIATI, Negin. Frost Monitoring Cyber-Physical System: A Survey on Prediction and Active Protection Methods. *IEEE Internet of Things Journal*. July 2020. Vol. 7, no. 7, p. 6514–6527. DOI 10.1109/JIOT.2020.2972936.
17. DING, Liya, NOBORIO, Kosuke, SHIBUYA, Kazuki and TAMURA, Yosuke. Frost forecast - a practice of machine learning from data. *International Journal of Reasoning-based Intelligent Systems*. 1 January 2021. Vol. 13, p. 1. DOI 10.1504/IJRIS.2021.10038162.
18. XU, Jie, GUGA, Suri, RONG, Guangzhi, RIAO, Dao, LIU, Xingpeng, LI, Kaiwei and ZHANG, Jiquan. Estimation of Frost Hazard for Tea Tree in Zhejiang Province Based on Machine Learning. *Agriculture*. 2021. Vol. 11, no. 7. DOI 10.3390/agriculture11070607.
19. HANDLER, Shawn L., REEVES, Heather D. and MCGOVERN, Amy. Development of a Probabilistic Subfreezing Road Temperature Nowcast and Forecast Using Machine Learning. *Weather and Forecasting*. 1 October 2020. Vol. 35, no. 5, p. 1845–1863. DOI 10.1175/WAF-D-19-0159.1.
20. DING, Liya, NOBORIO, Kosuke and SHIBUYA, Kazuki. Frost Forecast using Machine Learning - from association to causality. *Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 23rd International Conference KES2019*. 1 January 2019. Vol. 159, p. 1001–1010. DOI 10.1016/j.procs.2019.09.267.

21. GUI SSEPE, Arturo. OLAS DE FRÍO EN EL PERÚ. . Trabajo de suficiencia profesional. Lima, Perú : Universidad Nacional Agraria La Molina, 2022.
22. QUISPE, María and YANA, John. Modelo predictivo agroclimático para detectar cambios climatológicos con machine learning e IoT para el sector agrario de la Región Puno. . Tesis para obtener el Título Profesional de Ingeniero de Sistemas. Juliaca, Perú : Universidad Peruana Unión, 2020.
23. CASTRO, Narciso and ANABEL, Yovana. Determinación de modelos de estimación de temperaturas mínimas con fines de pronóstico de heladas en la región Ancash. Online. Tesis para Optar el Título de Ingeniero Meteorólogo. Lima, Perú : Universidad Nacional Agraria La Molina, 2019. Available from: <https://hdl.handle.net/20.500.12996/4074pdf>
24. MARÍN, Darwin and PINEDA, Ian. Modelo predictivo Machine Learning aplicado a análisis de datos Hidrometeorológicos para un SAT en Represas. Online. Tesis para obtener el Título Profesional de Ingeniero de Sistemas e Informática. Arequipa, Perú : Universidad Tecnológica del Perú, 2019. Available from: https://repositorio.utp.edu.pe/bitstream/handle/20.500.12867/3300/Darwin%20Marin_Ian%20Pineda_Tesis_Titulo%20Profesional_2019.pdf?sequence=1pdf
25. FLORES, José, CUXART, Joan, PIÑAS, Manuel, CALLAÑAUPA, Stephany, SUÁREZ, Luis, KUMAR, Shailendra, MOYA, Aldo S. and SILVA, Yamina. Seasonal and Diurnal Cycles of Surface Boundary Layer and Energy Balance in the Central Andes of Perú, Mantaro Valley. Atmosphere. 2019. Vol. 10, no. 12. DOI 10.3390/atmos10120779.
26. VELIZ, Carlos. Aprendizaje Automático - Introducción al aprendizaje profundo. Online. 2. Lima, Perú : Pontificia Universidad Católica del Perú, Fondo Editorial, 2023. ISBN 978-612-317-596-2. Available from: <https://www.fondoeditorial.pucp.edu.pe/ciencias-e-ingenieria-/1111-aprendizaje-automatico-introduccion-al-aprendizaje-profundo.html>
27. THURNER, Stefan, KLIMEK, Peter and HANEL, Rudolf. Introduction to the Theory of Complex Systems. Online. Oxford University Press, 2018. [Accessed 16 March 2025]. ISBN 978-0-19-882193-9.
28. KALIRAJ and DEVI. ARTIFICIAL INTELLIGENCE THEORY, MODELS, AND APPLICATIONS. Online. First Edition. Boca Raton, FL, USA; Abingdon, Oxon, UK : CRC Press (Taylor & Francis Group, LLC), 2022. ISBN 78-1-032-00809-7 (hbk), 978-1-032-10613-7 (pbk), 978-1-003-17586-5 (ebk). Available from: DOI: 10.1201/9781003175865
29. LIU, Yanli, WANG, Yourong and ZHANG, Jian. New Machine Learning Algorithm: Random Forest. In : LIU, Baoxiang, MA, Maode and CHANG, Jincai (eds.), Information Computing and Applications. Berlin, Heidelberg : Springer Berlin Heidelberg, 2012. p. 246–252. ISBN 978-3-642-34062-8.
30. GEURTS, Pierre, ERNST, Damien and WEHENKEL, Louis. Extremely Randomized Trees. Machine Learning. 1 April 2006. Vol. 63, p. 3–42. DOI 10.1007/s10994-006-6226-1.

31. SHALEV, Shai and BEN, Shai (eds.). Decision Trees. In : SHALEV, Shai and BEN, Shai (eds.), *Understanding Machine Learning: From Theory to Algorithms*. Online. Cambridge : Cambridge University Press, 2014. p. 212–218. ISBN 978-1-107-05713-5. Cambridge Core
32. ZHANG, Pengbo and YANG, Zhi-Xin. A Novel AdaBoost Framework With Robust Threshold and Structural Optimization. *IEEE Transactions on Cybernetics*. 24 November 2016. Vol. PP, p. 1–13. DOI 10.1109/TCYB.2016.2623900.
33. KORSTANJE, Joos. Gradient Boosting with XGBoost and LightGBM. In : KORSTANJE, Joos (ed.), *Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook’s Prophet, and Amazon’s DeepAR*. Online. Berkeley, CA : Apress, 2021. p. 193–205. ISBN 978-1-4842-7150-6.
34. LIU, Fei and YANG, Ming. Verification and Validation of Artificial Neural Network Models. In : ZHANG, Shichao and JARVIS, Ray (eds.), *AI 2005: Advances in Artificial Intelligence*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2005. p. 1041–1046. ISBN 978-3-540-31652-7. .
35. DANTAS, Pierre Vilar, SABINO DA SILVA, Waldir, CORDEIRO, Lucas Carvalho and CARVALHO, Celso Barbosa. A comprehensive review of model compression techniques in machine learning. *Applied Intelligence*. 1 November 2024. Vol. 54, no. 22, p. 11804–11844. DOI 10.1007/s10489-024-05747-w.
36. WAQAS, Muhammad, HUMPHRIES, Usa, CHUEASA, Bunthid and WANGWONGCHAI, Angkool. Artificial Intelligence and Numerical Weather Prediction Models: A Technical Survey. *Natural Hazards Research*. 25 November 2024. DOI 10.1016/j.nhres.2024.11.004.
37. ROZANTE, José Roberto, RAMIREZ, Enver, RAMIREZ, Diego and ROZANTE, Gabriela. Improved frost forecast using machine learning methods. *Artificial Intelligence in Geosciences*. 1 December 2023. Vol. 4, p. 164–181. DOI 10.1016/j.aiig.2023.10.001.
38. HODSON, T. O. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*. 2022. Vol. 15, no. 14, p. 5481–5487. DOI 10.5194/gmd-15-5481-2022.
39. MARKNER, Brigitte. Meteorology and Climatology. In : MARKNER, Brigitte (ed.), *Technical English for Geosciences*. Online. Berlin, Heidelberg : Springer Berlin Heidelberg, 2008. p. 156–158. ISBN 978-3-540-68616-3.
40. OMAZIĆ, Branimir, ANIĆ, Mislav, TELIŠMAN PRTENJAK, Maja, KVAKIĆ, Marko and BLAŠKOVIĆ, Lucija. Analysis of different existing measurement-based methods and a new approach for frost probability detection. *Agricultural and Forest Meteorology*. 15 March 2024. Vol. 347, p. 109898. DOI 10.1016/j.agrformet.2024.109898.
41. ZHU, Zongsheng, DOU, Wei, DAI, Baomin, SHAN, Xiaofang and SHEN, Jiang. Method to delay frost formation under high relative humidity by choosing proper heat exchange temperature difference and air velocity. *Case Studies in Thermal Engineering*. 1 May 2023. Vol. 45, p. 102922. DOI 10.1016/j.csite.2023.102922.

42. HABEN, Stephen, VOSS, Marcus and HOLDERBAUM, William. Time Series Forecasting: Core Concepts and Definitions. In : HABEN, Stephen, VOSS, Marcus and HOLDERBAUM, William (eds.), Core Concepts and Methods in Load Forecasting: With Applications in Distribution Networks. Online. Cham : Springer International Publishing, 2023. p. 55–66. ISBN 978-3-031-27852-5.
43. SANABRIA, Janeet. Estudio: Calibración y validación de modelos de pronóstico de heladas en el Valle del Mantaro. Online. 2005. Available from: <https://hdl.handle.net/20.500.12542/3246pdf> Histórico Senamhi 1969 - 2013 [300]
44. HERNÁNDEZ, Sampieri, FERNÁNDEZ, Collado and BAPTISTA, Lucio. Metodología de la Investigación. Online. Sexta edición. México, D.F. : McGraw-Hill / Interamericana Editores, S.A. de C.V., 2014. ISBN 978-1-4562-2396-0. Available from: <https://www.esup.edu.pe/wp-content/uploads/2020/12/2.%20Hernandez,%20Fernandez%20y%20Baptista-Metodolog%C3%ADa%20Investigacion%20Cientifica%206ta%20ed.pdf>
45. MURILLO, Myriam, CEJAS, Magda, LICCIONI, Edith, ALDAZ, Silvia and VENEGAS, Gina. Enfoque cuantitativo y cualitativo: Una mirada de los métodos mixtos. . 2023. ISBN ISBN: 978-980-248-306-8.
46. KARL, Popper. La lógica de la investigación científica. Online. Quinta reimpresión. Madrid, España : TECNOS, 1980. ISBN 84-309-0711- 4. Available from: <https://educacion.uncuyo.edu.ar/upload/popper-logica-inv-cientifica-cap1.pdf>
47. APICELLA, L., PUCA, S., LAGASIO, M., MERONI, A. N., MILELLI, M., VELA, N., GARBERO, V., FERRARIS, L. and PARODI, A. The predictive capacity of the high resolution weather research and forecasting model: a year-long verification over Italy. Bulletin of Atmospheric Science and Technology. 7 April 2021. Vol. 2, no. 1, p. 3. DOI 10.1007/s42865-021-00032-x.
48. ROZANTE, José Roberto, RAMIREZ, Enver, RAMIREZ, Diego and ROZANTE, Gabriela. Improved frost forecast using machine learning methods. Artificial Intelligence in Geosciences. 1 December 2023. Vol. 4, p. 164–181. DOI 10.1016/j.aiig.2023.10.001.
49. MINAM. Programa de Fortalecimiento de Capacidades Nacionales para manejar el Impacto del Cambio Climático y la Contaminación del Aire (PROCLIM). Online. 2005. [Accessed 21 May 2025]. Available from: <http://met.igp.gob.pe/proyectos/proclim/>

ANEXOS

ANEXO 1: Matriz de Consistencia

Título: Desarrollo de modelos de inteligencia artificial para el pronóstico de heladas en la cuenca del río Mantaro, Chupaca, 2024

PROBLEMA	OBJETIVO	HIPÓTESIS	VARIABLES	METODOLOGÍA	MUESTRA	TÉCNICAS E INSTRUMENTOS
<p>Problema General:</p> <p>¿Cómo desarrollar modelos de inteligencia artificial para predecir las heladas en la cuenca del río Mantaro, Chupaca, 2024?</p> <p>Problemas Específicos:</p> <p>¿Cuáles son las variables climáticas más relevantes para la predicción de heladas en la cuenca del río Mantaro?</p>	<p>Objetivo General:</p> <p>Desarrollar modelos de inteligencia artificial para predecir heladas en la cuenca del río Mantaro.</p> <p>Objetivos Específicos:</p> <p>Identificar las variables climáticas más relevantes para la predicción de heladas en la cuenca del río Mantaro, Chupaca, 2024.</p> <p>Analizar los datos históricos para identificar patrones que permitan el entrenamiento de un modelo de inteligencia artificial.</p>	<p>Hipótesis General:</p> <p>El desarrollo de modelos de inteligencia artificial permite predecir heladas en la cuenca del río Mantaro, Chupaca, 2024.</p> <p>Hipótesis Específicas:</p> <p>La identificación de las variables climáticas optimiza la predicción de heladas, mejorando la efectividad del modelo de inteligencia artificial.</p> <p>El análisis de datos históricos permite identificar patrones climáticos clave que sustentan el</p>	<p>Variable Dependiente:</p> <p>Pronóstico de helada.</p> <p>Variables Independientes:</p> <p>Modelos de inteligencia artificial.</p>	<p>Tipo de Investigación:</p> <p>Investigación aplicada.</p> <p>Nivel de Investigación:</p> <p>Nivel predictivo.</p> <p>Método General:</p> <p>Método deductivo</p> <p>Diseño:</p> <p>Experimental verdadero (causa-efecto).</p>	<p>Población:</p> <p>Todas las 22 estaciones meteorológicas de la cuenca del río Mantaro (datos de 2017–2023).</p> <p>Muestra:</p> <p>Estación meteorológica de Huayao (IGP) y 22 estaciones del SENAMHI.</p> <p>Muestreo: No probabilístico por conveniencia</p>	<p>Técnicas:</p> <p>Revisión de registros y observación de campo.</p> <p>Instrumentos:</p> <p>Archivo “Datos_Met.csv”, sensores meteorológicos, software de análisis (Python) y entrevistas a especialistas.</p>

<p>¿Cómo procesar y analizar los datos históricos para identificar patrones que permitan entrenar un modelo de inteligencia artificial?</p>	<p>Evaluar la precisión y confiabilidad de distintos modelos de inteligencia artificial en la predicción de heladas, comparándolos con los métodos tradicionales.</p>	<p>entrenamiento del modelo. La evaluación de la precisión y confiabilidad de distintos modelos de inteligencia artificial permitió identificar el modelo más adecuado para la predicción de heladas en la cuenca del río Mantaro, Chupaca, 2024</p>
<p>¿Qué modelo de inteligencia artificial ofrece mayor precisión y confiabilidad en la predicción de heladas en comparación con los métodos tradicionales?</p>		

ANEXO 2: Matriz de operacionalización de variables

Variable	Tipo de variable	Definición	Dimensión	Sub dimensión	Indicador	Escala	Instrumento
Modelos de inteligencia artificial	Variable independiente	Programas informáticos diseñados para detectar patrones en grandes conjuntos de datos y realizar predicciones o tomar decisiones basadas en esos patrones.	Tipos de Entrenamiento: - Tipos de modelos. - Entrenamiento y validación - Fuentes de datos.	Modelos de ensamble (Random Forest, XGBoost, Gradient Boosting)	Error cuadrático medio (MSE)	°C	Software de análisis (Python), bases de datos meteorológicas, técnicas estadísticas
					Error cuadrático medio ajustado (RMSE).	°C	
					Error absoluto medio (MAE)	°C	
					Coeficiente de determinación (R ²)	-	
				Técnicas de validación (cross-validation, grid search, random search)	Coeficiente de determinación ajustado (R ² ajustado) Error de porcentaje cuadrático medio (MSPE)	- %	

				Datos: SENAMHI, IGP	Error porcentual absoluto medio (MAPE)		
				• Errores de predicción (MAE, MSE,	Predicción Tiempo de respuesta	ln %	
		Herramienta		RMSE)	Precisión	%	Sensores
		meteorológica que anticipa la ocurrencia de heladas en una	Exactitud Fiabilidad: - Evaluación de la precisión del	• Análisis de correlación entre pronósticos y	Predicción de tendencias Predicción de eventos extremos	% Proporción	meteorológicos (termómetros, higrómetros, barómetros,
Pronóstico de helada	Variable dependiente	región específica, basada en el análisis de variables climáticas (temperatura, humedad, etc.).	pronóstico - Comparación entre datos predichos y observados.	datos observados • Variables climáticas: temperatura mínima, humedad, presión, viento, radiación	Predicción Tiempo de respuesta	Proporción Horas/días	anemómetros), software de modelado predictivo, análisis de correlación

ANEXO A: FOTOGRAFÍAS



Fotografía 1.- Descarga de datos meteorológicos de la torre de gradiente, Observatorio de Huayao-IGP.



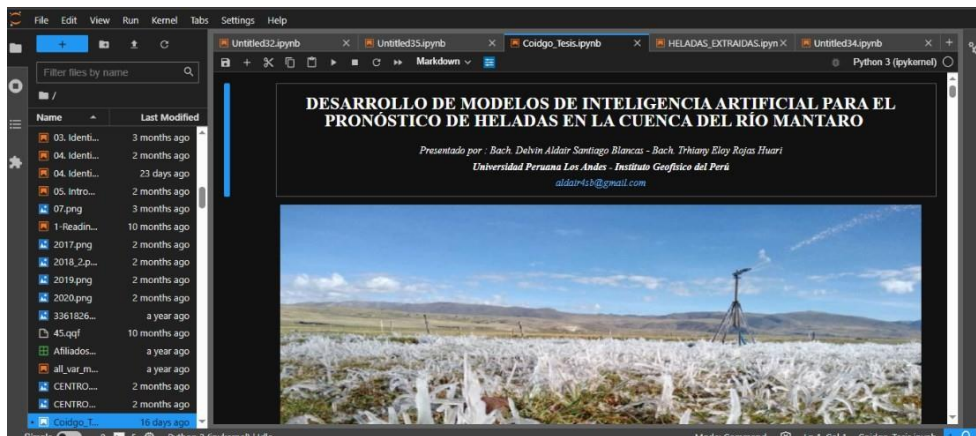
Fotografía 2.- Vista de torre de gradiente.



Fotografía 3.- Estación meteorológica experimental.

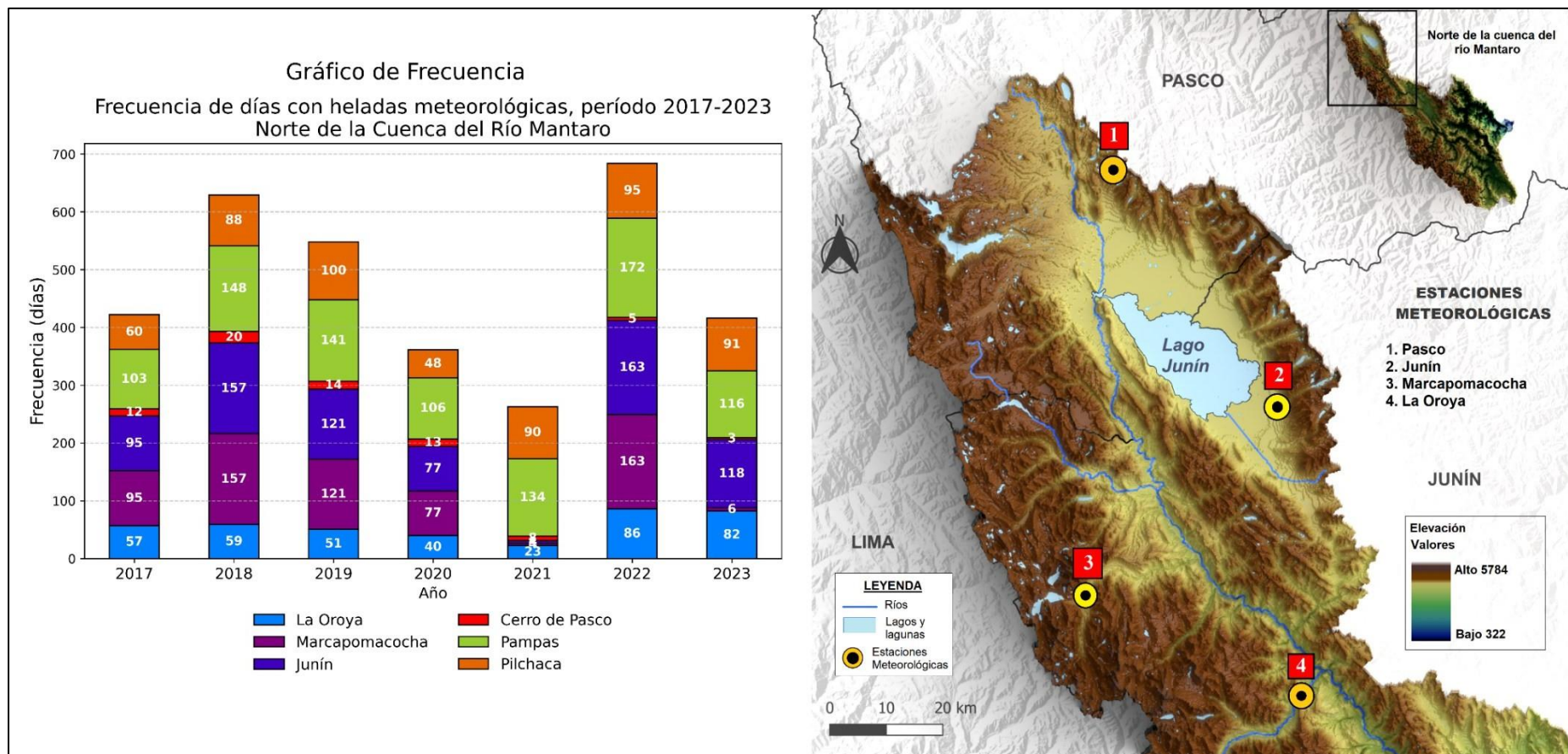


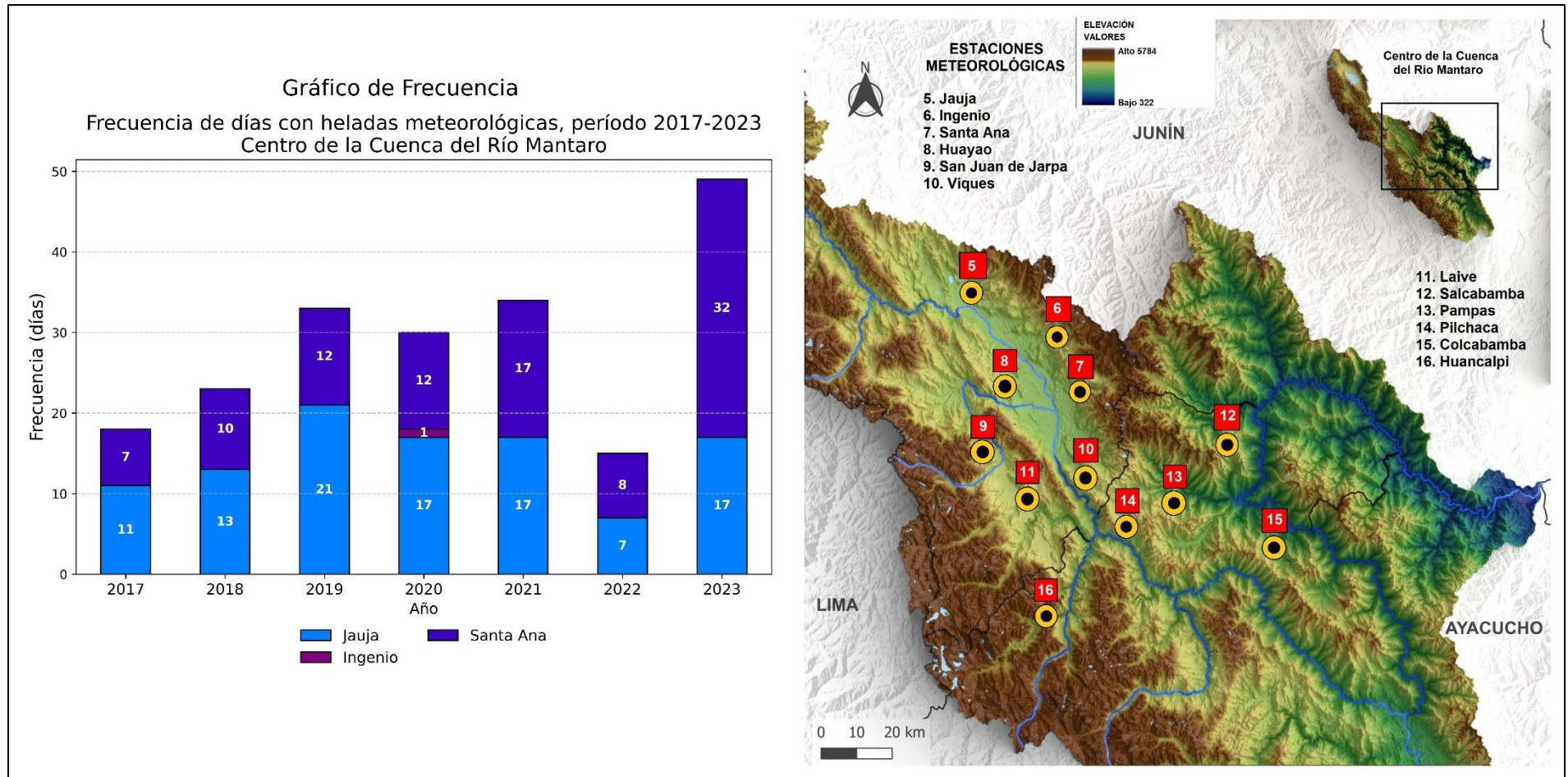
Fotografía 4.- Desarrollo del código de los modelos de IA en las instalaciones del IGP.



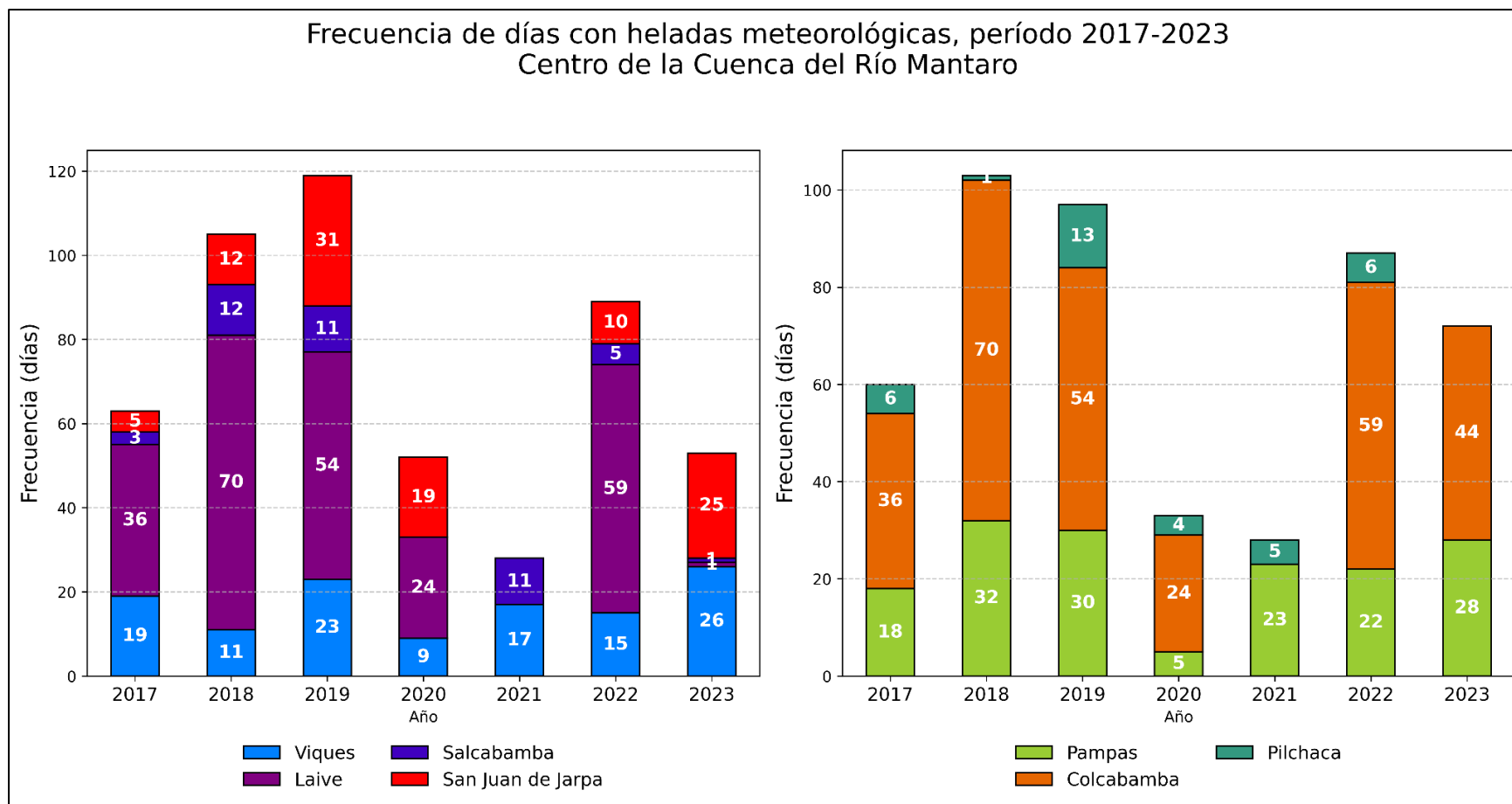
Fotografía 5.- Esquema del código desarrollado

ANEXO B: MAPAS GENERADOS

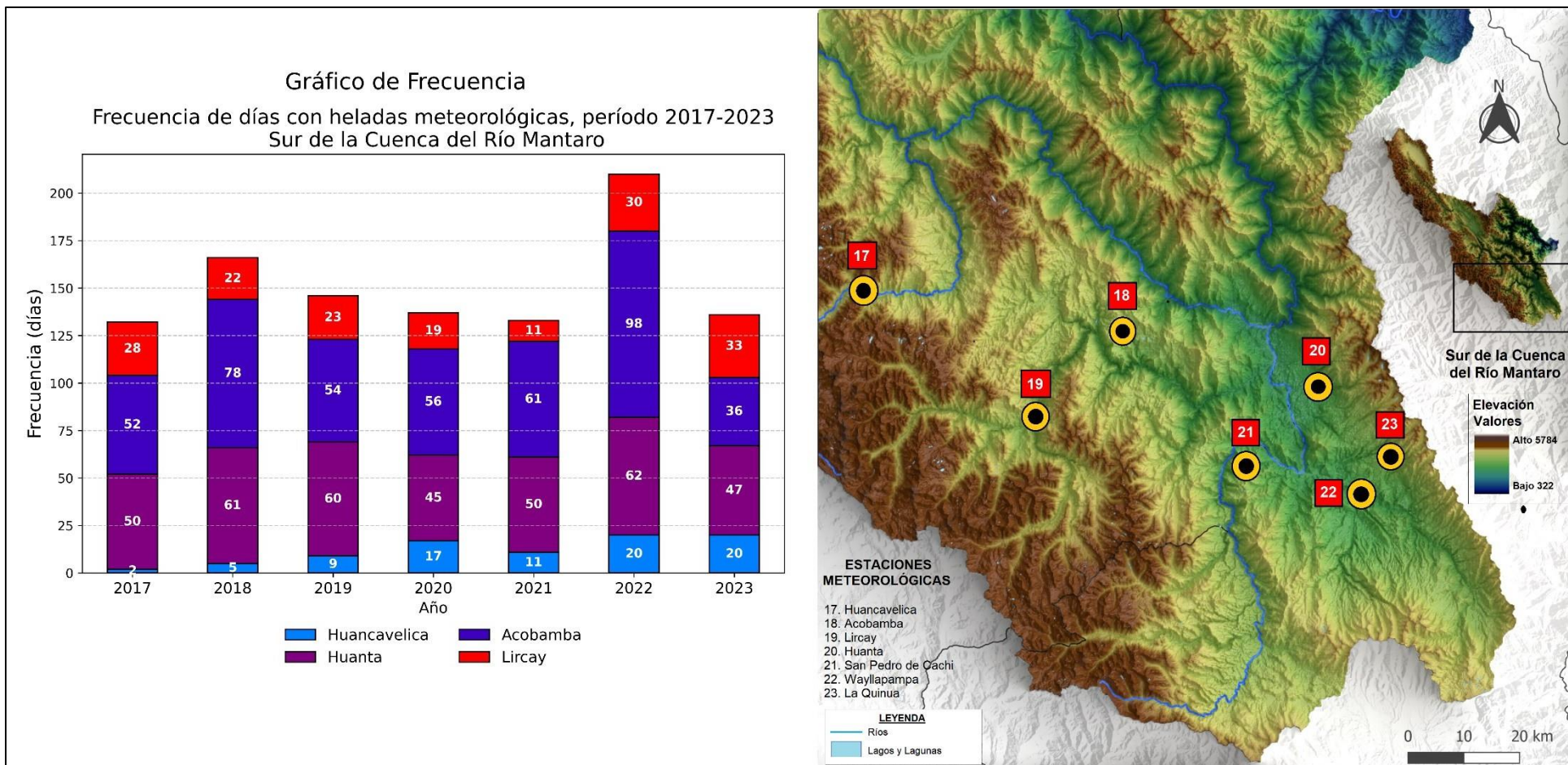




Mapa 02.- Número de heladas registradas en el centro de la Cuenca del Río Mantaro en el periodo del año 2017 – 2023 con ayuda del modelo.

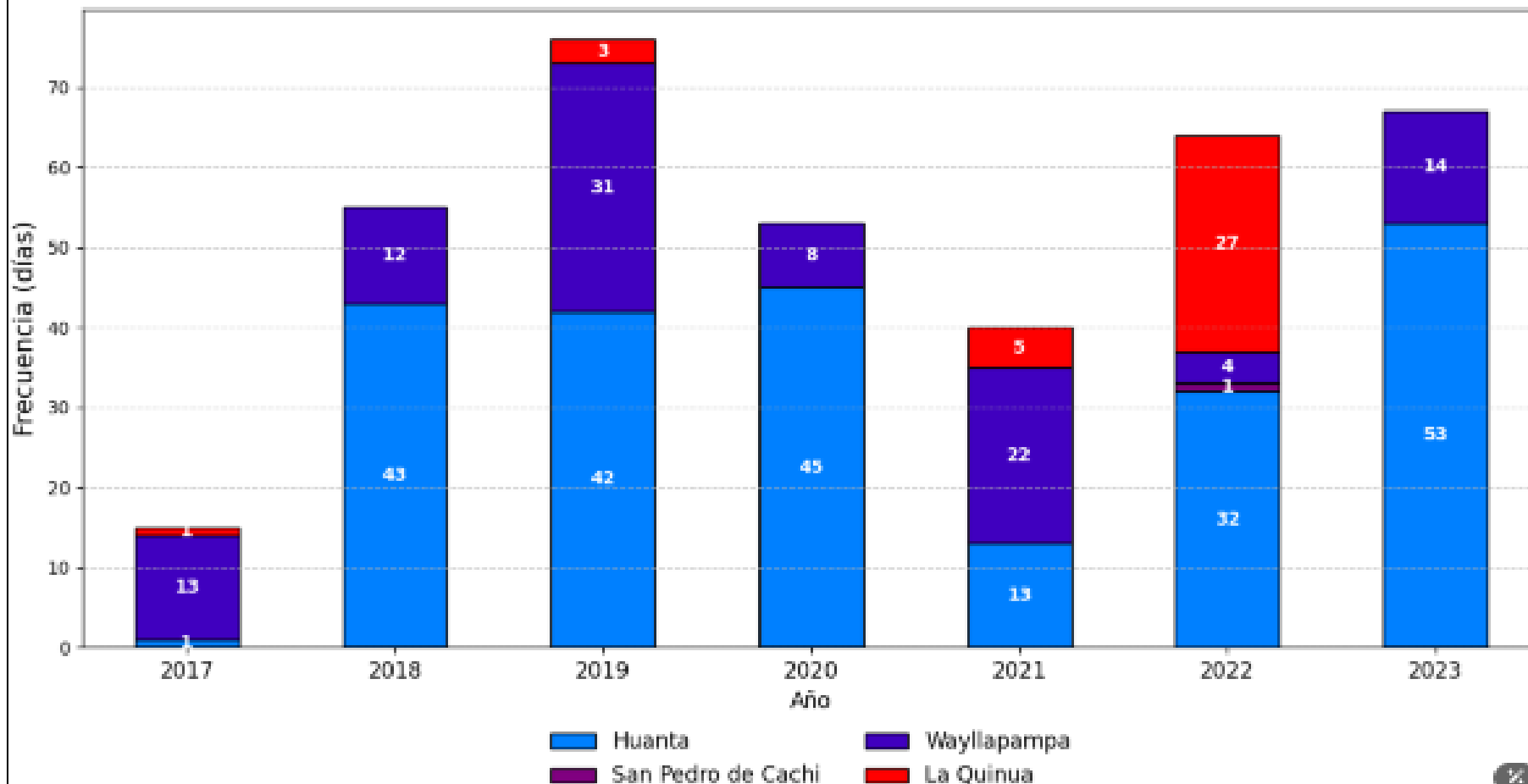


Mapa 03.- Número de heladas registradas en el centro de la Cuenca del Río Mantaro en el periodo del año 2017 – 2023 con ayuda del modelo.

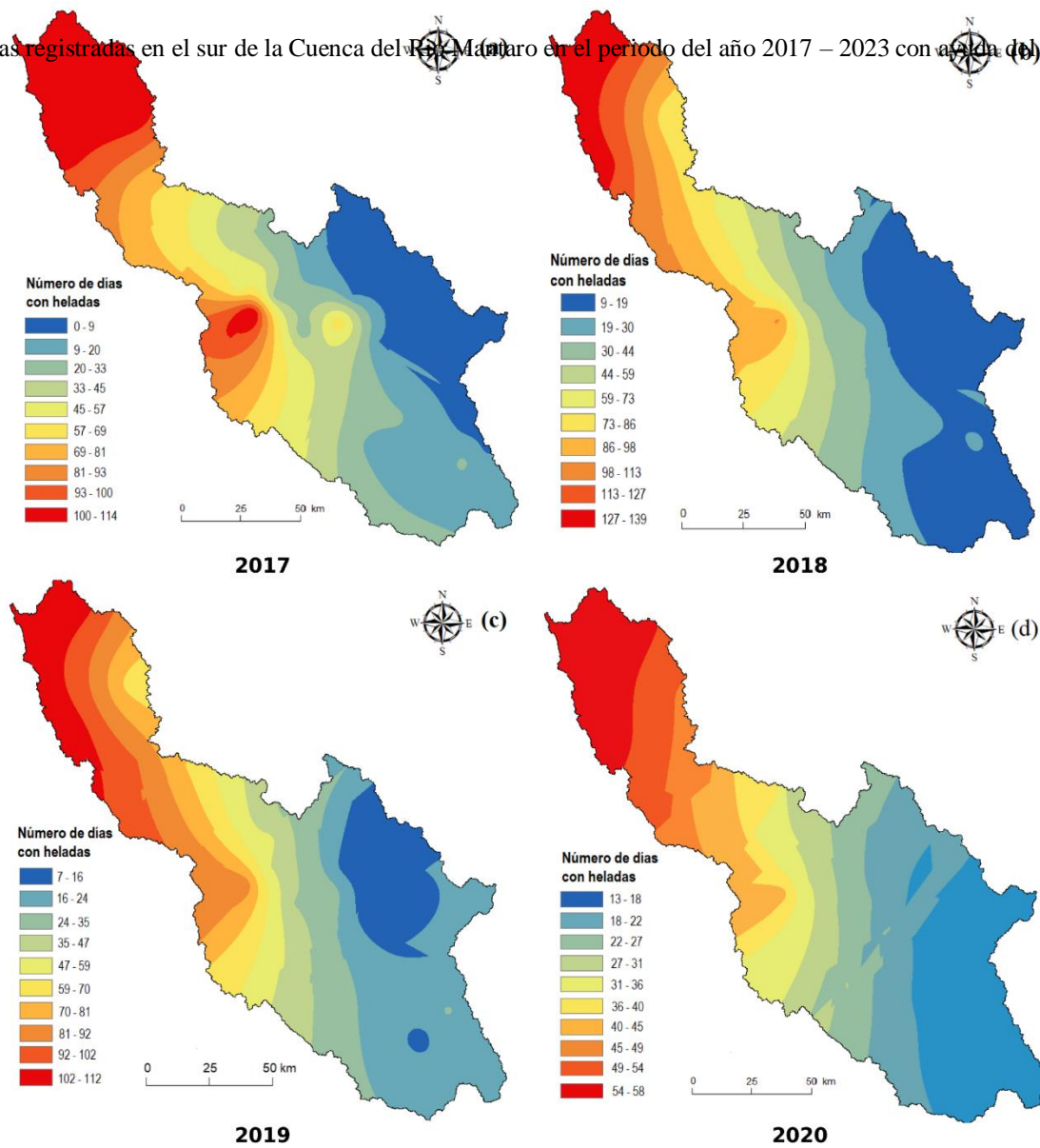


Mapa 04.- Número de heladas registradas en el sur de la Cuenca del Río Mantaro en el periodo del año 2017 – 2023 con ayuda del modelo.

Frecuencia de días con heladas meteorológicas, período 2017-2023
 Sur de la Cuenca del Río Mantaro



Mapa 04.- Número de heladas registradas en el sur de la Cuenca del Río Mantaro en el periodo del año 2017 – 2023 con apoyo de un modelo.



Mapa 05.- Distribución de días con heladas en la cuenca del Río del Mantaro, periodo 2017-2020.