

ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE

<https://dx.doi.org/10.14482/inde.43.01.155.454>

Predicción de heladas y variables meteorológicas relevantes en agricultura en la Sabana de Bogotá usando machine learning

Frost and relevant meteorological variables forecast in agriculture in the Sabana de Bogotá using machine learning

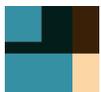
ROBINSON CASTILLO MÉNDEZ *
JULIÁN ANDRÉS CAMACHO CASTRO **

* Investigador Grupo de Investigación del CEET SENA (GICS). Centro de Electricidad, Electrónica y Telecomunicaciones (CEET). Servicio Nacional de Aprendizaje (SENA). Magíster en Microelectrónica: Diseño y Aplicaciones de Sistemas Micro/Nanométricos. Orcid-ID: <https://orcid.org/0000-0002-7412-6306>. rcastillom@sena.edu.co

** Investigador Grupo de Investigación del CEET SENA (GICS). Centro de Electricidad, Electrónica y Telecomunicaciones (CEET). Servicio Nacional de Aprendizaje (SENA). Magíster en Gestión Sostenible de la Energía. Orcid-ID: <https://orcid.org/0009-0004-3545-5539>. jacamachoc@sena.edu.co

Correspondencia: Robinson Castillo. Dirección: Av. Cra. 30 n.º 17-91 Sur. Bogotá, D.C., Colombia. Celular: +573004779236.

Subvenciones: Proyecto ejecutado por la línea programática de investigación aplicada y semilleros, en el Centro de Electricidad, Electrónica y Telecomunicaciones (CEET), en el marco de la estrategia SENNOVA, del Servicio Nacional de Aprendizaje (SENA). Código: SGPS-10692-2023. Proyecto: “Implementación de un modelo de machine learning para determinar el comportamiento de heladas y variables meteorológicas fundamentales en la agricultura”.



Resumen

A partir de información histórica de variables climatológicas y de las heladas, es posible mejorar las decisiones tomadas en las actividades de la agricultura, buscando determinar patrones que garanticen mayor rendimiento y calidad de los cultivos e implementando modelos de predicción basados en *machine learning* (ML). Se propone desarrollar un modelo ML que permita determinar el comportamiento de las variables meteorológicas temperatura, pluviosidad y humedad relativa, así como de las heladas en la Sabana de Bogotá. Se ha partido de la conformación de una base de datos históricos de estas variables desde 2010 hasta abril de 2023, considerando información de diez estaciones meteorológicas diferentes de la región. Ha sido necesario implementar técnicas de imputación de datos en los vacíos de información. Para determinar el modelo con la respuesta más cercana a la realidad, se desarrolló un modelo basado en regresión lineal múltiple y otro en redes neuronales artificiales. De acuerdo con los resultados obtenidos y el nivel de error absoluto, el segundo modelo aproxima sus predicciones más cerca a los datos reales. El trabajo desarrollado puede ser una herramienta esencial para generar un sistema de alerta temprana que ayude a los agricultores de la Sabana de Bogotá.

Palabras clave: aprendizaje automático, predicción de heladas, regresión lineal múltiple, red neuronal de avance, variables meteorológicas.

Abstract

Taking into account historical information on climatological and frost variables, it is possible to improve decisions made in agricultural activities, seeking to determine patterns that guarantee greater yield and quality of crops and implementing forecast models based on machine learning (ML). This work presents the development of a ML model that allows determining the behavior of the meteorological variables, temperature, rainfall, and relative humidity, as well as frost, in the Sabana de Bogotá. The starting point was the creation of a historical database of these variables from 2010 to April 2023, considering information from ten different meteorological stations in the region. It has been necessary to implement data imputation techniques in information gaps. To determine the model with the response closest to reality, a model based on multiple linear regression and another on artificial neural networks were developed. According to the results and the level of absolute error, the second model approximates its forecasts closer to the real data. The work developed can be an essential tool to generate an early warning system that helps farmers in the Sabana de Bogotá.

Keywords: feedforward neural network, frost forecasting, machine learning, meteorological conditions, multiple linear regression.

INTRODUCCIÓN

La agricultura se enfrenta a crecientes desafíos se estima que la población mundial humana llegará a los 9 mil millones de personas hacia 2050, incrementando aproximadamente un 70 % la demanda de alimentos. Para satisfacer esta creciente demanda alimentaria, el consumo de agua en la agricultura debería aumentar aproximadamente un 41 %. Adicionalmente, el mundo y la agricultura se enfrentan a problemas como el cambio climático y las apropiaciones de recursos naturales superiores en un 30 % a la capacidad de la naturaleza para regenerarse [1]. Organizaciones como la FAO (Organización de las Naciones Unidas para la Agricultura y la Alimentación) indican, como ocurre en el informe “El estado mundial de la agricultura y la alimentación 2021”, que la estabilidad y seguridad de los sistemas agroalimentarios están cada vez más en mayor riesgo debido a factores como el cambio climático, la deforestación, la degradación de los recursos naturales y otras crisis prolongadas [2].

En relación con el cambio climático, una de las problemáticas más comunes en la meteorología y en la agricultura son la ocurrencia de heladas, el frío o el congelamiento. Este fenómeno ocurre cuando la temperatura ambiente mínima alcanza y disminuye respecto a cierto valor, ocasionando daños a los diferentes cultivos. Abordar y superar esta situación requiere el lograr predecir la temperatura mínima, y con un tiempo suficiente para implementar las medidas necesarias. La FAO ha proporcionado algunos métodos empíricos que pueden predecir la temperatura mínima, pero no con suficiente tiempo. Alternativamente, se pueden implementar soluciones que incorporen métodos de aprendizaje automático para modelar la temperatura mínima [3] y otras variables climáticas relevantes.

En Colombia, el sector agrícola no se ha visto muy beneficiado por los avances tecnológicos en materia de *machine learning* (también conocido como “aprendizaje automático”) y ciencia de datos; estas técnicas ayudan a entender los patrones y las dinámicas que sigue la naturaleza. La temperatura, la pluviosidad y la humedad relativa son variables climatológicas fundamentales en la agricultura; así mismo, las heladas son eventos climáticos de gran preocupación en dicha actividad, debido al potencial de pérdidas sociales y económicas que de ellas se derivan. Los daños en los cultivos por heladas tienen un efecto negativo para la planta, afectando el tejido y reduciendo su rendimiento y calidad. El aprendizaje automático tiene un considerable potencial para manejar numerosos desafíos en el establecimiento de sistemas agrícolas basados en el conocimiento. La inteligencia artificial y el *machine learning* aplicados en la agricultura mejoran las capacidades para realizar con mayor precisión cualquier actividad relacionada con el cultivo, a partir de la recolección de información y el aprendizaje.

El desarrollo de un modelo basado en ML que integre variables meteorológicas fundamentales (Temperatura, Pluviosidad y Humedad relativa) permitirá predecir con alta precisión la ocurrencia de heladas en la Sabana de Bogotá. Esta herramienta reducirá significativamente la incertidumbre asociada al comportamiento climático, facilitando la implementación de prácticas agrícolas más eficientes y la adopción de nuevos modelos especializados en el sector agrícola tradicional de la región.

ANTECEDENTES

La calidad y el rendimiento de los cultivos en la agricultura tradicional son afectados por factores ambientales externos. El cambio climático es uno de los principales que inciden en esta situación [4]. Una importante herramienta para determinar el comportamiento que tendrá un cultivo o una cosecha es la predicción con aprendizaje automático; la agricultura está evolucionando en ese sentido, permitiendo al agricultor tener mayores elementos de decisión y de procesamiento de información sobre su cultivo.

Trabajos presentados por autores como Cunha [5], quien describe un sistema que incorpora conjuntos de datos de precipitación y propiedades del suelo derivados de satélites, datos de pronóstico climático estacional de modelos físicos y otras fuentes para producir una predicción previa a la temporada del rendimiento de un cultivo de soja/maíz; o como Sharma [6], quien realiza una revisión del *machine learning* y su aplicación en la agricultura, siendo sus áreas de interés la predicción de parámetros del suelo (por ejemplo, carbono orgánico o contenido de humedad), la detección de enfermedades, especies y malas hierbas en los cultivos, así como la predicción de su rendimiento, o como Elghamrawy [7], quien propone un modelo de predicción de los efectos del cambio climático en la producción de cultivos (PMCCCE) basado en tecnologías de inteligencia artificial (IA) e internet de las cosas (IoT) y emplea una red neuronal convolucional (CNN), son evidencia de que el *machine learning* cada vez tiene mayor relevancia en el sector agrícola.

Como se mencionó anteriormente, la agricultura en Colombia no se ha visto muy beneficiada por los avances tecnológicos en materia de *machine learning* y ciencia de datos, pero otros trabajos, como los desarrollados por García Cañón [8], quien muestra una aproximación a las aplicaciones que tiene el ML a través de un modelo predictivo en variables del suelo, su humedad y temperatura a diferentes profundidades, o como González Botero [9], quien sugiere una opción basada en índices meteorológicos que posibilitan calcular un rango de probabilidad de ocurrencia de una helada en cultivos de Rosa en la Sabana de Bogotá, o Marqués Gozalbo [10], quien presenta ejemplos para la predicción de productividad agrícola partiendo de datos públicos, y de esta manera implementar modelos predictivos con algoritmos de inteligencia ar-

tificial, son ejemplos de desarrollos que abordan técnicas de aprendizaje automático y su aplicación en la agricultura en relación con variables meteorológicas, del suelo y de heladas.

Los algoritmos de *machine learning* elaboran un modelo a partir de un conjunto de datos de entrenamiento, con el fin de hacer predicciones o tomar decisiones sin estar programados de manera específica para ello. Se han realizado diferentes esfuerzos para utilizar la inteligencia artificial y el aprendizaje automático en la predicción de heladas. Se han introducido enfoques basados en una máquina de vectores de soporte (SVM) para predecir eventos de heladas, empleado elementos ambientales, como valores de temperatura, humedad y radiación, para predecir la probabilidad futura de heladas. Otros autores han investigado enfoques de ML para el pronóstico de heladas, especialmente durante la primavera. Se han diseñado modelos basados en redes neuronales artificiales (RNA) para clasificaciones de heladas. Otros autores han utilizado enfoques anteriores basados en ML mientras consideraban condiciones termodinámicas y asumieron que estas condiciones pueden ser informativas para la predicción [11].

Entre otros, lo anterior evidencia la necesidad de abordar una problemática que se presenta en el sector agrícola de Bogotá-Región: la baja certeza de los cultivadores acerca del comportamiento de las principales variables meteorológicas en la agricultura y de las heladas, así como de nuevas herramientas y modelos especializados para determinar su comportamiento.

METODOLOGÍA

La figura 1 muestra las fases generales consideradas para el desarrollo e implementación de la solución propuesta:



Fuente: elaboración propia.

FIGURA 1. BASE METODOLÓGICA

El objetivo principal de este trabajo es determinar el comportamiento de las variables meteorológicas Temperatura, Pluviosidad y Humedad relativa, así como su relación con la ocurrencia de heladas en la Sabana de Bogotá. Se busca implementar un

modelo basado en ML para reducir la incertidumbre en la predicción de heladas en la región. Se trata de un trabajo principalmente observacional y predictivo, utilizando datos históricos de variables meteorológicas y registros de eventos de heladas en la Sabana de Bogotá.

Conformación de base de datos

Consiste en la recopilación de información histórica y actual de las variables temperatura (T °C), humedad relativa (H.R. %) y pluviosidad (P mm); lo anterior con la colaboración del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). Se consideraron diez (10) diferentes estaciones meteorológicas pertenecientes a la Sabana de Bogotá; seleccionando el registro de mediciones diario correspondiente a las 7: 00 a. m., 1: 00 p.m. y 6: 00 p.m. para cada variable desde marzo de 2010 hasta abril de 2023.

Análisis de la información

Se observó que para las tres variables hacían falta mediciones, presentando vacíos en la base de información conformada, para lograr tener la totalidad de los datos para el periodo establecido en las tres horas de medición de cada día, se combinaron técnicas de imputación, empleando las técnicas de imputación mediante media y regresión, así como K-Nearest Neighbor [12], [13].

Para la información de cada una de las variables, la imputación mediante media se empleó en los vacíos de uno o dos datos consecutivos. En los vacíos de tres a cinco datos seguidos, la imputación empleada fue la de regresión. Finalmente, para vacíos mayores a cinco datos consecutivos, el algoritmo K-Nearest Neighbor, siendo K impar y de valor más cercano a la cantidad de datos faltantes identificados.

Posteriormente, se identificaron y corrigieron valores atípicos en el conjunto de datos de entrada imputado y se agruparon por ubicación geográfica de cada estación. Para su posterior uso en el desarrollo de los modelos, el almacenamiento del conjunto de datos se realizó en el formato de archivos Feather (siendo más compatible para leer y escribir datos en lenguajes como Python y R); lo anterior empleando el apoyo de herramientas *software* y plataformas de programación y cálculo numérico para ingeniería y ciencia de datos.

Desarrollo del modelo

En primera instancia se desarrolló un modelo basado en regresión lineal múltiple [14]:

$$y_i = b_0 + b_1 \cdot x_{i1} + b_2 \cdot x_{i2} + \dots + b_k \cdot x_{ik} + u_i, \text{ para } i = 1, 2, 3, \dots, n \quad (1)$$

Siendo:

y_i = valor de la variable dependiente para el i -ésimo caso

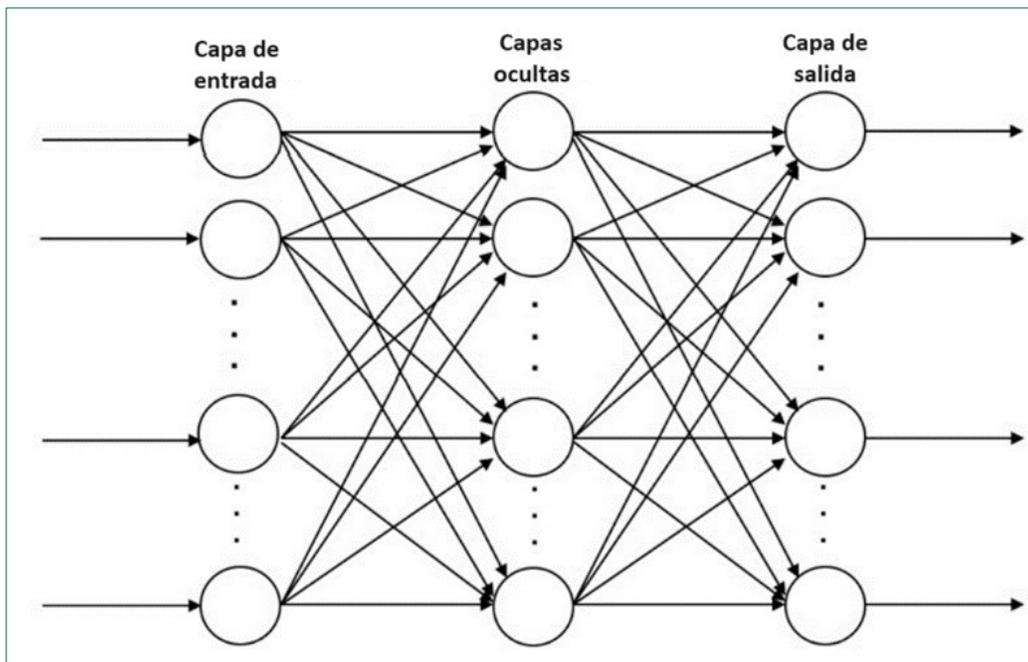
x_{ik} = valor de la k -ésima variable independiente para el i -ésimo caso

b_0 = es el Y-intercepto de la superficie de regresión

u_i = componente de error aleatorio para el i -ésimo caso

Cada b_j , donde $j = 1, 2, \dots, k$, representa la pendiente de la superficie de regresión respecto a la variable x_j . En la ecuación (1) hay n observaciones y k predictores (donde $n > k + 1$). El modelo estableció una relación entre las variables temperatura (T °C), humedad relativa (H.R. %) y pluviosidad (P mm), desde 2018.

Posteriormente, se desarrolló un segundo modelo basado en redes neuronales artificiales, como lo muestra la figura 2; de base se ha empleado una red de alimentación hacia adelante (*Feed-forward*), transmitiendo los datos desde la entrada en una sola dirección:



Fuente: elaboración propia.

FIGURA 2. RED NEURONAL DE ALIMENTACIÓN HACIA ADELANTE, BASE PARA EL SEGUNDO MODELO

La red empleada en este segundo modelo empleó once (11) neuronas en la capa de entrada (para las variables temperatura (T °C), humedad relativa (H.R. %) y pluviosidad (P mm), setenta y cinco (75) neuronas en la capa oculta y once neuronas en la salida. Para determinar el número de neuronas en la capa oculta, se realizó un proceso iterativo, considerando el tamaño del conjunto de datos y la capacidad computacional, ajustando el número de capas hasta llegar al tamaño empleado.

Se empleó la retropropagación de regularización bayesiana como algoritmo de entrenamiento, actualizando los valores de peso y sesgo según la optimización de Levenberg-Marquardt. La retropropagación se utiliza para calcular la matriz jacobiana jX del rendimiento $pref$ con respecto a las variables de peso y sesgo X [15]:

$$jj = jX * jX \quad (2)$$

$$je = jX * e \quad (3)$$

$$dX = -(jj + I * mu) \setminus je \quad (4)$$

donde corresponde a errores, I es la matriz identidad y mu parámetro de adaptación de Marquardt. La función de activación en la capa oculta, basada en la tangente sigmoidea, está dada por:

$$f(X) = \frac{1}{1 + e^{-x}} \quad (5)$$

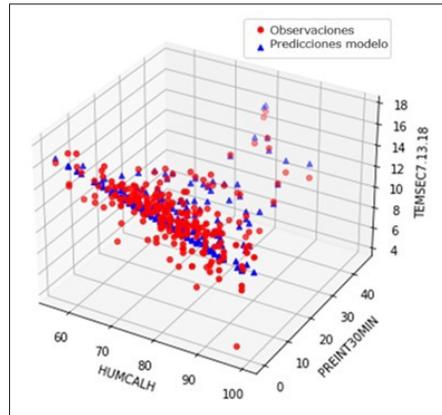
El modelo estableció una relación entre las variables temperatura (T °C), humedad relativa (H.R. %) y pluviosidad (P mm), desde 2010.

Validación

Se tomaron mediciones de las variables temperatura (T °C), humedad relativa (H.R. %) y pluviosidad (P mm), así como de la ocurrencia de helada de treinta (30) días posteriores a abril de 2023 en la Sabana de Bogotá; este conjunto de datos se empleó para probar la repuesta del modelo basado en regresión y para el desarrollado con la red neuronal para contrastar las respuestas para predicción de las variables climáticas consideradas y la probabilidad de ocurrencia de helada en esta región. Se realizó una comparación de las respuestas de cada modelo frente a los datos reales, y se obtuvo métricas de asertividad, nivel de error y predicción por cada variable para el periodo de validación considerado.

RESULTADOS Y DISCUSIONES

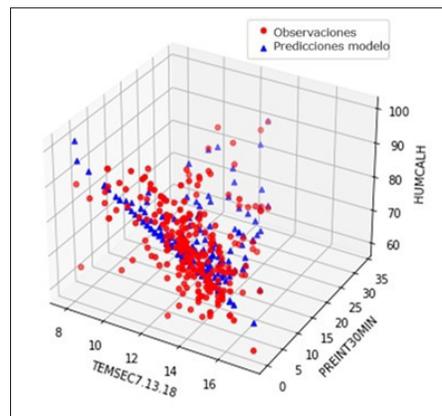
Las figuras 3, 4 y 5 muestran la respuesta del modelo basado en regresión lineal múltiple durante su etapa de desarrollo con la base de datos históricos desde 2010 hasta abril de 2023:



Fuente: elaboración propia.

FIGURA 3. RESPUESTA DEL MODELO RESPECTO A LA VARIABLE $T^{\circ}\text{C}$

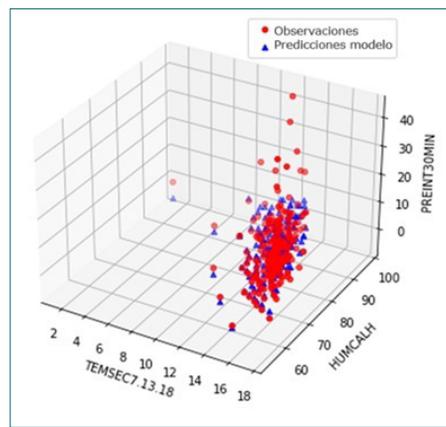
La figura 3 representa la relación entre la variable temperatura (en $^{\circ}\text{C}$) y los valores predichos por el modelo. La nube de puntos rojos corresponde a las observaciones, representados en los valores históricos para esta variable, mientras que la nube de puntos azules, las predicciones del modelo. Se encuentra que los puntos se encuentran cerca de la línea de regresión, indicado un buen ajuste del modelo a los datos. La pendiente observada indica que a medida que aumenta el valor de la variable independiente, también aumenta el valor de la variable dependiente.



Fuente: elaboración propia.

FIGURA 4. RESPUESTA DEL MODELO RESPECTO A LA VARIABLE H.R. %

La figura 4 representa la relación entre la variable humedad relativa (en %) y los valores predichos por el modelo. De manera similar a la figura anterior, la nube de puntos rojos corresponde a las observaciones, representados en los valores históricos para esta variable, mientras que la nube de puntos azules; las predicciones del modelo. Se observa que los puntos azules, aunque en su mayoría se encuentran cerca de la línea de regresión, la concentración de algunos no lo está, indicado un ajuste aceptable del modelo a los datos. La pendiente observada indica que a medida que aumenta el valor de la variable independiente, también aumenta el valor de la variable dependiente.

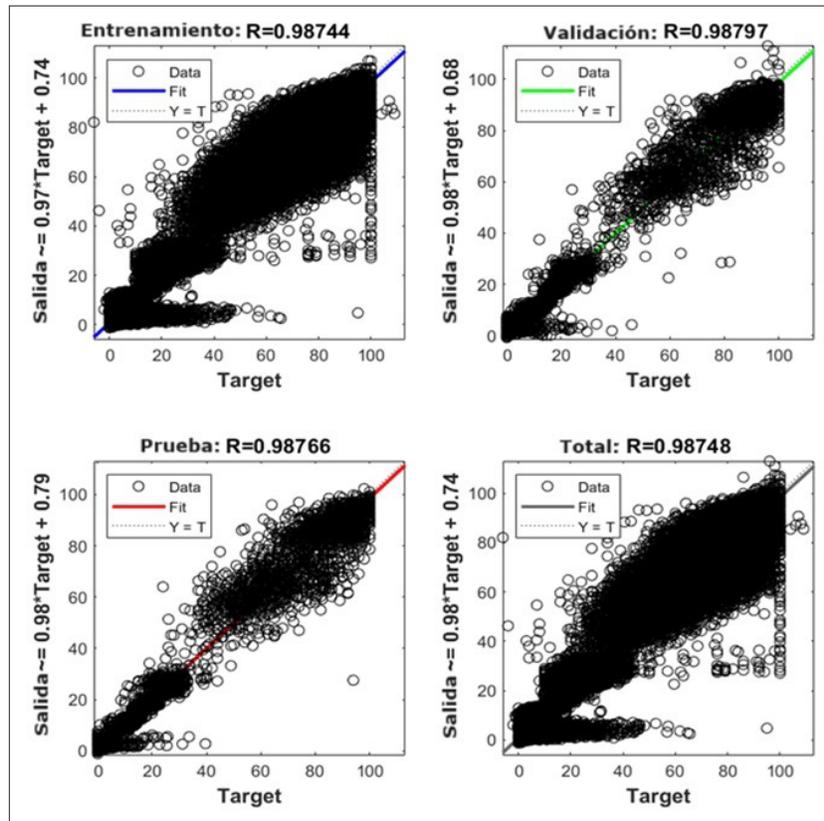


Fuente: elaboración propia.

FIGURA 5. RESPUESTA DEL MODELO RESPECTO A LA VARIABLE P MM

La figura 5 representa la relación entre la variable pluviosidad (en mm) y los valores predichos por el modelo. En este caso también la nube de puntos rojos corresponde a las observaciones, representados en los valores históricos para esta variable, mientras que la nube de puntos azules, las predicciones del modelo. De manera similar al caso anterior, se observa que los puntos azules, aunque en su mayoría se encuentran cerca de la línea de regresión, la concentración de algunos no lo está, indicado un ajuste aceptable del modelo a los datos. La pendiente observada indica que a medida que aumenta el valor de la variable independiente, también aumenta el valor de la variable dependiente.

Para el modelo basado en red neuronal se empleó un 90 % de los datos para el aprendizaje, un 5 % como base de validación y otro 5 % como prueba. La figura 6 muestra la respuesta de este modelo durante su etapa de desarrollo con la base de datos históricos desde 2010 hasta abril de 2023:

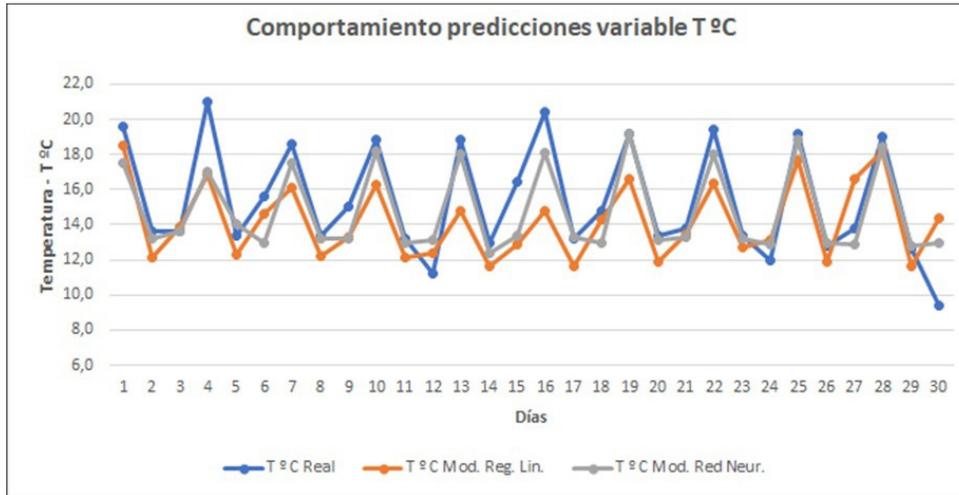


Fuente: elaboración propia.

FIGURA 6. RESPUESTA DEL MODELO BASADO EN RED NEURONAL PARA TODAS LAS VARIABLES CONSIDERADAS

Como puede observarse, en función de la precisión del modelo, este modelo logró ajustar los datos de entrenamiento al 98,74 %, así mismo que la red generaliza nuevos datos al 98,79 %, y se estima que el rendimiento de la red en un escenario real puede aproximarse al 98,76 %, indicado un ajuste bueno del modelo a los datos.

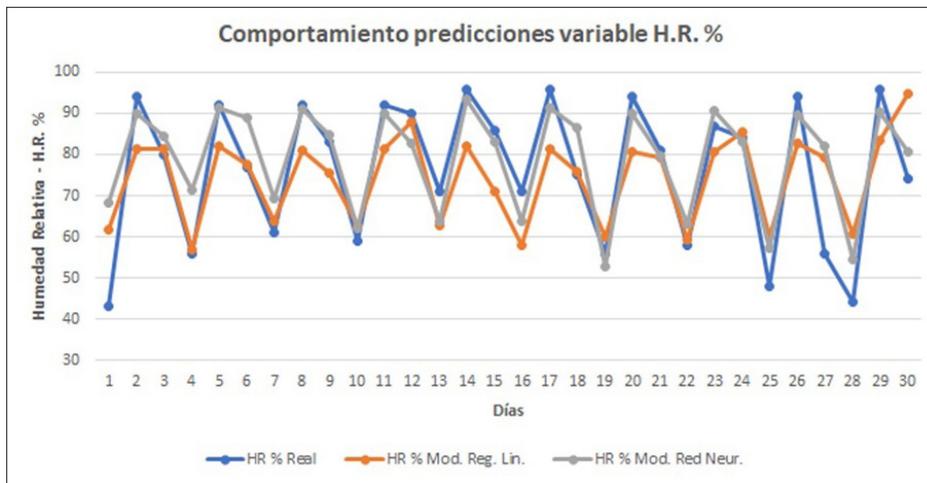
Considerando las respuestas mostradas en las figuras anteriores, se evaluó el comportamiento de los dos modelos frente a la predicción de las variables climatológicas de interés y la probabilidad de ocurrencia de helada en la Sabana de Bogotá durante treinta (30) días posteriores a abril de 2023, es decir, posteriores al conjunto de datos de desarrollo y entrenamiento de los modelos. Como se observa en las figuras 7, 8, 9 y 10, los dos modelos presentan predicciones cercanas al comportamiento real (variables climáticas medidas por el IDEAM), siendo la red neuronal la que en promedio presenta un menor nivel de error:



Fuente: elaboración propia.

FIGURA 7. TEMPERATURA REAL VS. PREDICCIONES DE LOS DOS MODELOS

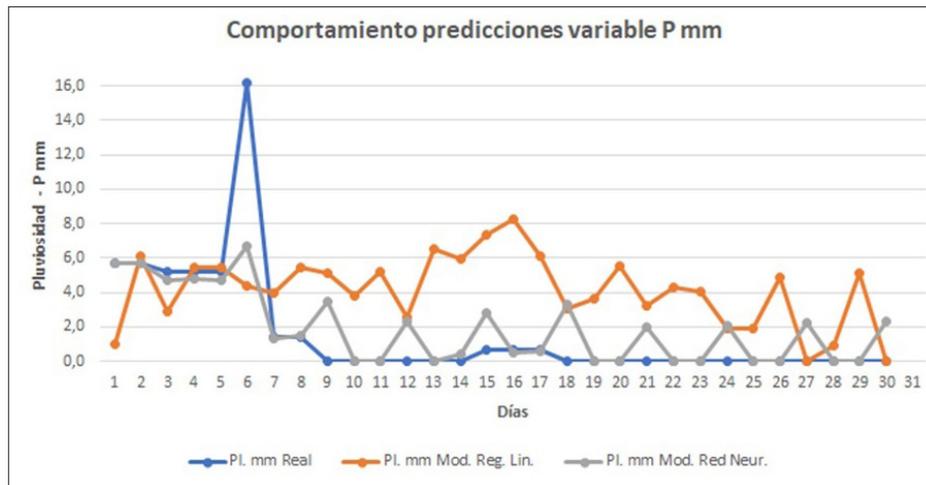
La figura 7 presenta una comparación entre los valores reales de la variable temperatura y las predicciones realizadas por los dos modelos: regresión lineal múltiple y el de la red neuronal. Como tendencia general, valores reales y predicciones siguen una tendencia similar, indicando que ambos modelos capturan la variabilidad general de la temperatura a lo largo del tiempo. La línea de la regresión lineal múltiple presenta una tendencia más suave y menos detallada que los datos reales, mientras que la red neuronal se ajusta más a los picos y valles de los datos reales, indicando que captura patrones más complejos y ofrece una predicción más precisa de la variable.



Fuente: elaboración propia.

FIGURA 8. HUMEDAD RELATIVA VS. PREDICCIONES DE LOS DOS MODELOS

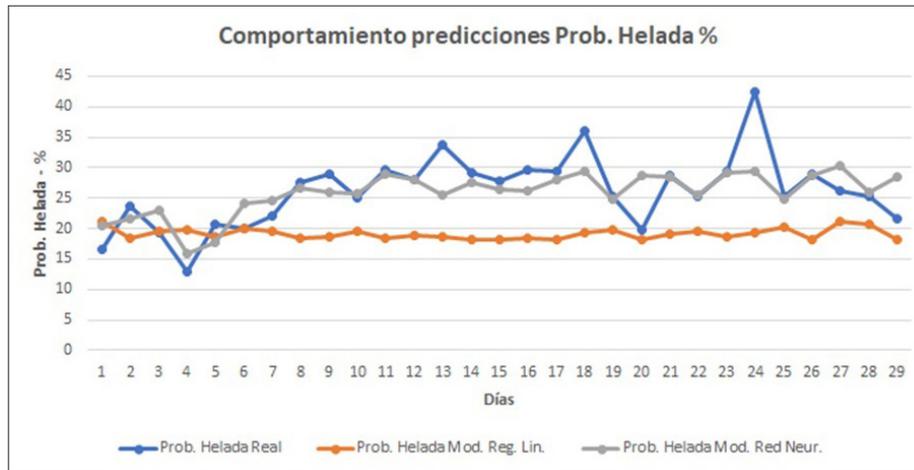
La figura 8 presenta una comparación entre los valores reales de la variable humedad relativa y las predicciones realizadas por los dos modelos: regresión lineal múltiple y el de la red neuronal. Como tendencia general, y de manera similar al caso de la variable temperatura, valores reales y predicciones siguen una tendencia similar, indicando que ambos modelos capturan la variabilidad general de la humedad relativa a lo largo del tiempo. La línea de la regresión lineal múltiple en este caso también presenta una tendencia más suave y menos detallada que los datos reales, mientras que la red neuronal se ajusta más a los picos y valles de los datos reales, indicando que captura patrones más complejos y ofrece una predicción más precisa de la variable.



Fuente: elaboración propia.

FIGURA 9. PLUVIOSIDAD VS. PREDICCIONES DE LOS DOS MODELOS

La figura 9 presenta una comparación entre los valores reales de la variable pluviosidad y las predicciones realizadas por los dos modelos: regresión lineal múltiple y el de la red neuronal. Como tendencia general, en este caso también los valores reales y predicciones siguen una tendencia similar, indicando que ambos modelos capturan la variabilidad general de la pluviosidad a lo largo del tiempo, aunque en menor medida para el primer modelo. La línea de la regresión lineal múltiple en este caso también presenta una tendencia más suave y un poco menos detallada que los datos reales, mientras que la red neuronal se ajusta más a los picos y valles de los datos reales, indicando que captura patrones más complejos y ofrece una predicción más precisa de la variable. Adicionalmente se observa un punto donde las líneas de predicción se alejan significativamente de los datos reales, representando un error significativo en la predicción.

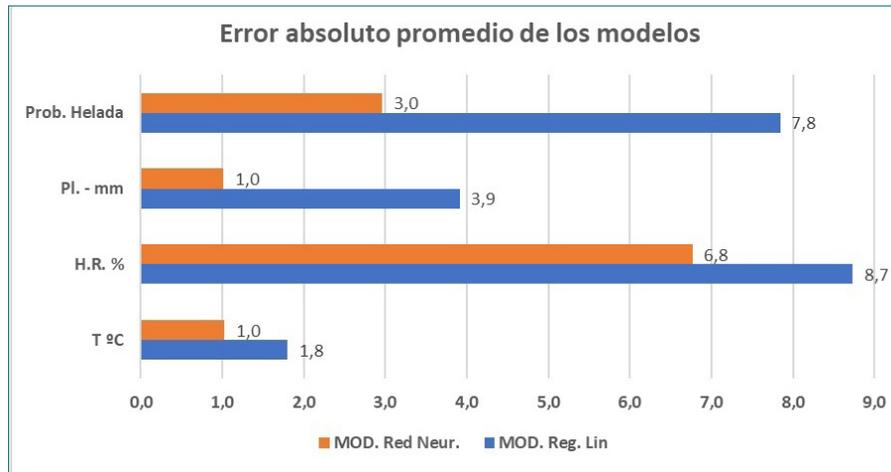


Fuente: elaboración propia.

FIGURA 10. PROBABILIDAD DE HELADA VS. PREDICCIONES DE LOS DOS MODELOS

La figura 10 presenta una comparación entre los valores reales de la probabilidad de ocurrencia de helada en la Sabana de Bogotá y las predicciones realizadas por los dos modelos: regresión lineal múltiple y el de la red neuronal. Como tendencia general, se observa que ambos modelos capturan la variabilidad general de la probabilidad de helada a lo largo del tiempo. Se nota que la línea de la regresión lineal múltiple presenta una tendencia más suave y menos detallada que los datos reales, mientras que la red neuronal se ajusta más a los picos y valles de los datos reales, indicando que captura patrones más complejos y ofrece una predicción más precisa de la variable. En general, el primer modelo se encuentra más alejado en su predicción y se observan dos puntos donde las líneas de predicción se alejan significativamente de los datos reales, representando un error significativo en la predicción.

La figura 11 muestra el error absoluto de las predicciones realizadas por los dos modelos; en el caso del modelo de regresión lineal, la variable de menor nivel de error ha sido la humedad relativa, y la más acertada, la temperatura. Un comportamiento similar se presenta en el modelo basado en la red neuronal, aunque en este segundo caso el nivel de error absoluto es menor.



Fuente: elaboración propia.

FIGURA 11. ERROR ABSOLUTO PROMEDIO EN LAS PREDICCIONES DE LOS DOS MODELOS

Modelar y predecir las variables climatológicas y la ocurrencia de heladas corresponde con un sistema no lineal, el cual requiere del análisis de un importante número de variables, con funciones y comportamientos desconocidos. El desarrollo de este trabajo ha considerado las variables más influyentes para la agricultura en la Sabana de Bogotá: la temperatura, la humedad relativa y la pluviosidad, y con estas variables, especialmente con la primera, estimar la probabilidad de ocurrencia de heladas en esta región (se consideraron estas tres variables también por la cantidad de datos disponibles para otras variables en las estaciones de interés para este estudio y desarrollo de los modelos).

Al tratarse de un sistema predominantemente caótico, se verificó el comportamiento basado en la aplicación de un modelo de regresión lineal múltiple que correlaciona las tres variables consideradas, pero también la posibilidad de aplicar modelos no lineales hizo factible implementar un modelo basado en redes neuronales artificiales.

Frente los resultados obtenidos en la predicción de las variables de interés y la probabilidad de ocurrencia de heladas en la Sabana de Bogotá, ambos modelos presentaron un comportamiento muy aceptable; sin embargo, la asertividad del modelo basado en redes neuronales fue mayor, teniendo un margen de error absoluto de 3.3 puntos en promedio por debajo del sistema basado en regresión lineal considerando todas las variables (el desempeño ha sido mucho más cercano a la realidad en las variables temperatura y pluviosidad).

Dado que los modelos se basan en tres mediciones diarias de cada variable, y la predicción se realiza considerando el siguiente momento de medición, estos modelos

pueden llegar a contribuir en el incremento de la certeza de los agricultores sobre el comportamiento de las principales variables meteorológicas y de las heladas, favoreciendo su actividad, todo esto desde el desarrollo de aplicaciones de *machine learning*.

CONCLUSIONES

Este trabajo propone dos modelos para la predicción de temperatura, humedad relativa, pluviosidad y ocurrencia de heladas en la Sabana de Bogotá: el primero, basado en regresión lineal múltiple, y el segundo, en redes neuronales artificiales. Se emplearon datos de diez (10) estaciones meteorológicas diferentes de la región, con información de estas variables desde 2010 hasta marzo de 2023, con mediciones diarias a las 7: 00 a.m., 1: 00 p.m. y 6: 00 p.m. Fue necesario emplear técnicas de imputación de datos para completar los vacíos en la información.

Los resultados indicaron un rendimiento muy alto, especialmente en el enfoque propuesto mediante red neuronal artificial. El enfoque propuesto puede considerarse una alternativa o un complemento a los métodos tradicionales de predicción de heladas, y puede ser una herramienta esencial para generar un sistema de alerta temprana que ayude a los agricultores de la Sabana de Bogotá.

Haciendo referencia a las heladas, es importante tener en cuenta que el enfoque del desarrollo de este trabajo es abordar las que se ocasionan por radiación en la Sabana de Bogotá. Cuanto más variables se consideren, existen mayores probabilidades de obtener un modelo con un nivel de asertividad mucho mayor.

Como trabajo futuro se propone desarrollar un modelo basado en redes neuronales más robusto, validar la metodología presentada y sistema que se va a obtener con registros asociados a diferentes estaciones meteorológicas ubicadas en distintas zonas geográficas del país, y de esta manera lograr establecer una generalidad de mayor aplicación al modelo propuesto.

REFERENCIAS

- [1] M. K. Sott et al., “Precision Techniques and Agriculture 4.0 Technologies to Promote Sustainability in the Coffee Sector: State of the Art, Challenges and Future Trends”, *IEEE Access*, vol. 8, pp. 149854-149867, 2020. doi: 10.1109/ACCESS.2020.3016325.
- [2] J. P. Tovar-Soto, M. O. Gonzalez y J. A. S. Sanchez, “Digital agriculture for urban crops: design of an IoT platform for monitoring variables”, en *2022 IEEE International Conference on Automation/25th Congress of the Chilean Association of Automatic Control: For the Development of Sustainable Agricultural Systems*, ICA-ACCA 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICA-ACCA56767.2022.10006036.

- [3] M. Barooni, K. Ziarati y A. Barooni, “Frost Prediction Using Machine Learning Methods in Fars Province”, en *2023 28th International Computer Conference, Computer Society of Iran, CSICC 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/CSICC58665.2023.10105391.
- [4] C. A. Ramírez Gómez, “Aplicación del Machine Learning en agricultura de precisión”, *Revista Cintex*, vol. 25, n°. 2, pp. 14-27, 2020.
- [5] R. L. F. Cunha, B. Silva y M. A. S. Netto, “A scalable machine learning system for pre-season agriculture yield forecast”, *Proceedings-IEEE 14th International Conference on eScience, e-Science 2018*, pp. 423-430, 2018, doi: 10.1109/eScience.2018.00131.
- [6] A. Sharma, A. Jain, P. Gupta y V. Chowdary, “Machine Learning applications for precision agriculture: A comprehensive review”, *IEEE Access*, vol. 9, pp. 4843-4873, 2021. doi: 10.1109/ACCESS.2020.3048415.
- [7] S. Elghamrawy, “An AI-Based Prediction Model for Climate Change Effects on Crop production using IoT”, en *2023 International Telecommunications Conference, ITC-Egypt 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 497-503. doi: 10.1109/ITC-Egypt58155.2023.10206201.
- [8] H. S. García Cañón, “Implementación de técnicas de machine learning para la predicción de variables meteorológicas y del suelo que afectan la agricultura”, Universidad de los Andes, 2019.
- [9] D. F. González Botero, “Planteamiento de un modelo de predicción de heladas en cultivos de rosa en la sabana de Bogotá”, Universidad Militar Nueva Granada, Bogotá D.C., 2018.
- [10] M. Á. Marqués Gozalbo, “Modelos predictivos de producción agroindustrial con Machine Learning a partir de fuentes de información pública”, Universidad de Córdoba, 2020.
- [11] M. G. Herabad y N. P. Afshar, “Fuzzy-based Deep Reinforcement Learning for Frost Forecasting in IoT Edge-enabled Agriculture”, en *Proceedings - 2022 8th International Iranian Conference on Signal Processing and Intelligent Systems, ICSPIS 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICSPIS56952.2022.10044063.
- [12] W. Lv, H. Huang, W. Tang y T. Chen, “Research and Application of Intersection Similarity Algorithm Based on KNN Classification Model”, en *Proceedings - 2021 International Conference on Artificial Intelligence, Big Data and Algorithms, CAIBDA 2021*, Institute of Electrical and Electronics Engineers Inc., mayo 2021, pp. 141-144. doi: 10.1109/CAIBDA53561.2021.00037.
- [13] M. Akhiladevi, K. Anitha, K. Amrutha, M. Amrutha y K. R. Chandanashree, “Accident Prediction Using KNN Algorithm”, en *4th International Conference on Emerging Research in Electronics, Computer Science and Technology, ICERECT 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICERECT56837.2022.10059746.

- [14] M. R. Putri, I. G. P. S. Wijaya, F. P. A. Praja, A. Hadi y F. Hamami, “The Comparison Study of Regression Models (Multiple Linear Regression, Ridge, Lasso, Random Forest, and Polynomial Regression) for House Price Prediction in West Nusa Tenggara”, en *ICADEIS 2023 - International Conference on Advancement in Data Science, E-Learning and Information Systems: Data, Intelligent Systems, and the Applications for Human Life, Proceeding*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICADEIS58666.2023.10270916.
- [15] MathWorks®, “Bayesian regularization backpropagation” [En línea], Bayesian regularization backpropagation. Disponible en: <https://la.mathworks.com/help/deeplearning/ref/trainbr.html?lang=en> [Accedido: 25 nov. 2023].