

Recibido: 20-08-2024 Aceptado: 1-10-2024

Desarrollo de un modelo de *machine learning* para determinar el comportamiento de heladas en la Sabana de Bogotá¹

Development of a machine learning model to determine the behavior of frost in the Bogotá Savannah.

Robinson Castillo Mendez²

Julian Andres Camacho C.³

Cómo citar:

Mendez, R. & Camacho, J. (2024). Desarrollo de un modelo de machine learning para determinar el comportamiento de heladas en la Sabana de Bogotá. *Vía Innova*, 11 (1), 150-164. <https://doi.org/10.23850/2422068X.6819>

¹ Proyecto Implementación de un modelo de Machine Learning para determinar el comportamiento de heladas y variables meteorológicas fundamentales en la agricultura

² Robinson Castillo Méndez Ingeniero Electrónico Máster en Microelectrónica, Centro de Electricidad, Electrónica y Telecomunicaciones CEET (Servicio Nacional de Aprendizaje - SENA), rcastillom@sena.edu.co, Bogotá D.C.

³ Julián Andrés Camacho Ingeniero Electrónico Magister en Gestión Sostenible de la Energía, Centro de Electricidad, Electrónica y Telecomunicaciones CEET (Servicio Nacional de Aprendizaje - SENA), jacamachoc@sena.edu.co, Bogotá D.C"

RESUMEN

Se presenta el avance realizado en el desarrollo e implementación de un modelo de *machine learning* para determinar el comportamiento de las condiciones climatológicas fundamentales en la agricultura, y de la ocurrencia de una helada en la Sabana de Bogotá. Inicia con la creación de una base de información relacionada a estas variables, la caracterización, diseño, simulación, implementación y validación del sistema para llegar a la obtención de dicho modelo. Se incorpora un modelo de regresión múltiple y árboles de decisión para su implementación. Se establece una base de información actual e histórica acerca del comportamiento de las variables meteorológicas de Temperatura, Pluviosidad y Humedad Relativa, así como de la eventualidad de ocurrencia de una helada en la Sabana de Bogotá, logrando así la creación de la base de aprendizaje. Se ha caracterizado el sistema y se han definido los requerimientos de diseño, los cuales se encuentran en etapa de desarrollo e implementación. Se espera obtener un alto nivel de asertividad en el pronóstico realizado para aplicarlo en la agricultura de la Sabana de Bogotá inicialmente, así como mejorar las capacidades para realizar con mayor precisión cualquier actividad relacionada con el cultivo a partir de la recolección de información y el aprendizaje.

Una importante herramienta para determinar el comportamiento que tendrá un cultivo o una cosecha es la predicción. Con *machine learning* la agricultura está evolucionando en ese sentido, permitiendo al agricultor tener mayores elementos de decisión y de procesamiento de información sobre su cultivo.

Palabras clave: Aprendizaje Automático; Modelo de Predicción de Heladas; Sabana de Bogotá; Variables Climatológicas.

ABSTRACT

This work presents the progress made in the development and implementation of a machine learning model to determine the behavior of fundamental weather conditions in agriculture and the occurrence of frost in the Sabana de Bogotá. The project started with the establishment of an information base pertaining to the aforementioned variables. Then by the characterization, design, simulation, implementation, and validation of the system, with the objective of obtaining the model. A multiple regression model and decision trees have been incorporated for implementation purposes. A database has been constructed comprising current and historical data on the behaviour of meteorological variables: temperature, rainfall and relative humidity. Also the database includes information on the probability of frost occurrence in the Sabana de Bogotá. The learning base has been created and the system has been characterised and designed. The latter is currently in the development and implementation phase. It is expected to the forecasting will be highly accurate and will initially be applied to agriculture in the Sabana de Bogotá. Furthermore, it is expected that the precision of activities related to crop management, from data collection to learning.

A significant instrument for forecasting the behavior of a crop or harvest is the application of machine learning. The field of agriculture is making progress, enabling farmers take in more comprehensive decision-making and information processing regarding their crops.

Palabras clave: Machine learning; Frost prediction model; Sabana de Bogotá; Climatological variables.

1. INTRODUCCIÓN

El sector agrícola tiene una alta importancia dentro de la economía del país, pero tiene también un bajo nivel de aprovechamiento de los recursos tecnológicos disponibles, bien sea por brechas de conocimiento, altos costos, o por ser un sector poco flexible frente a cambios tecnológicos. Adicionalmente, el cambio climático tiene un papel relevante dentro de la dinámica de la actividad agrícola en la región y en el país; las variables meteorológicas y las heladas afectan considerablemente los cultivos y su rendimiento.

La agricultura en Colombia se ha visto poco beneficiada por los avances tecnológicos en materia de 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 la actividad agrícola 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 su tejido y reduciendo su rendimiento y calidad.

La transformación digital de la agricultura ha hecho evolucionar varios aspectos de la gestión hacia sistemas inteligentes artificiales en aras de la puesta en valor de los datos cada vez más numerosos, originados en diversas fuentes. El aprendizaje automático tiene un considerable potencial para atender numerosos desafíos en el establecimiento de sistemas agrícolas a partir del conocimiento. La inteligencia artificial y el Machine Learning (ML) aplicados en la agricultura, mejoran las capacidades para realizar con mayor precisión cualquier actividad relacionada con el cultivo, basados en la recolección de información y el aprendizaje.

A partir de información histórica de variables meteorológicas 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, implementando modelos de predicción basados en aprendizaje de máquina. El trabajo planteado tendrá impacto en el sector agrícola, por lo que se propone desarrollar e implementar un modelo de *machine learning* que permita determinar el comportamiento de las variables meteorológicas: temperatura, pluviosidad y humedad relativa, así como de las heladas en la sabana de Bogotá, con el fin de brindar nuevas herramientas y modelos de apoyo para el desarrollo de las actividades en el sector de la agricultura tradicional de la región. Se trata de un trabajo aún en curso, por lo que este documento presenta

el avance realizado en el desarrollo e implementación de este modelo, aplicado inicialmente a la región de la Sabana de Bogotá.

1.1 Generalidades

La agricultura tradicional se ve afectada por factores ambientales externos, los cuales reducen el rendimiento y la calidad de los cultivos. El cambio climático es uno de los principales factores que incide en este campo (Ramírez Gómez, 2020). 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. La agricultura en Colombia no se ha visto muy beneficiada por los avances tecnológicos en materia de *machine learning* y ciencia de datos, estas técnicas ayudan a entender los patrones y dinámicas que sigue la naturaleza. Trabajos como los desarrollados por García Cañón (2019), González Botero (2018) o Marqués Gozalbo (2020) abordan técnicas de aprendizaje automático y su aplicación en la agricultura en relación a variables meteorológicas, del suelo y de heladas.

Así mismo, los trabajos presentados por autores como Cunha et al. (2018) o Sharma et al. (2021) son evidencia de que el *machine learning* cada vez tiene mayor relevancia en el sector agrícola, teniendo en cuenta que la agricultura desempeña un papel fundamental en el crecimiento económico de cualquier país y también, que con el aumento de la población, los frecuentes cambios en las condiciones climáticas y los recursos limitados, satisfacer las necesidades alimentarias de la población actual se convierte en una tarea difícil.

154

1.2 Acerca de las heladas: nociones básicas

De acuerdo con la Organización de las Naciones Unidas para la Alimentación y Agricultura - FAO, una *helada* se considera como la baja temperatura que empieza a causar daños en los tejidos de las plantas, evidenciándose en sus aspectos fisiológicos. Se considera que en cualquier región del mundo es posible la ocurrencia de temperaturas por debajo de 0°, exceptuando las zonas de clima exclusivamente tropical; lo anterior significa que una helada puede ocurrir también en países de clima tropical en altitudes elevadas (Snyder et al., 2010).

El término “Helada” hace referencia a la formación de cristales de hielo sobre las superficies, tanto por congelación del rocío como por un cambio de fase de vapor de agua a hielo. En general, una helada corresponde a un evento meteorológico cuando los cultivos y otras plantas experimentan daño por congelación (Corporación Autónoma Regional de Cundinamarca (CAR), 2020). El Instituto de Hidrología Meteorología y Estudios Ambientales - IDEAM, define la helada como la ocurrencia de una temperatura igual o menor de 0°C a una altura de dos metros sobre el nivel del suelo.

Desde el punto de vista meteorológico, se alcanza el grado de helada cuando la temperatura del aire, a dos metros de la superficie del suelo, es igual o inferior a cero grados. Así mismo, el término “helada” puede definirse en el campo agrometeorológico como la temperatura a la cual los tejidos de la planta comienzan a sufrir daño, considerando aspectos fisiológicos como la resistencia susceptible del cultivo a bajas temperaturas en sus diferentes estados de desarrollo, altura de la planta sobre el nivel del suelo y la temperatura de la hoja (Gonzalez Botero, 2018). Cuando la temperatura alcanza los cero grados centígrados, debido a la posibilidad de nucleación del hielo, aumenta la posibilidad de éxtasis y desintegración en las células vegetales, y los productos agrícolas sensibles pueden resultar dañados, lo que tiene efectos importantes en la producción (Barooni et al., 2023).

155

La radiación solar, la circulación del viento, el vapor de agua en la atmósfera, la altitud, latitud y otras características topográficas de la zona, determinan la temperatura del aire y, consecuentemente, la posible ocurrencia de una helada. Las principales variables atmosféricas que inciden en las heladas son: Humedad Relativa, Velocidad del Viento, Temperatura, Precipitación y Radiación Solar. Según la Corporación Autónoma Regional de Cundinamarca (2020), de acuerdo a las causas, las heladas se clasifican en:

Heladas por Radiación: la pérdida de calor que se presenta en las plantas y el suelo al ceder a la atmósfera durante la noche por medio del proceso de radiación, es originada por las heladas de radiación. Este tipo de helada es producto de la inversión de la temperatura (las temperaturas aumentan con la altura respecto a la superficie del suelo). Se presentan dos subcategorías de heladas por radiación: una denominada “blanca”, la cual se caracteriza porque el vapor de agua se deposita sobre la superficie y forma una capa blanca de hielo que se denomina normalmente “escarcha”; el otro tipo es la helada “negra” que ocurre cuando la temperatura cae por debajo de 0°C y no se forma hielo debido a que la humedad es suficientemente baja. Normalmente, la helada blanca

produce menos daño que la helada negra (Corporación Autónoma Regional de Cundinamarca (CAR), 2020).

Heladas por Advección: ocurren debido a la invasión de masas de aire frío originario de las regiones polares y se presentan de manera continua por varios días. Son más características en latitudes medias y altas (Corporación Autónoma Regional de Cundinamarca (CAR), 2020).

Heladas por Evaporación: se presentan cuando, después de una precipitación, desciende la humedad relativa del aire, la vegetación intercepta parte de la precipitación reteniendo el agua en el follaje y troncos, se produce una intensa evaporación, el calor de evaporación que el agua necesita para pasar del estado líquido al gaseoso lo toma de las plantas y en consecuencia, la temperatura de algunos órganos vegetales desciende a límites que ocasionan daños (Corporación Autónoma Regional de Cundinamarca (CAR), 2020).

1.3 Aprendizaje Automático (Machine Learning o Aprendizaje de Máquina)

El Aprendizaje Automático puede definirse como un tipo de Inteligencia Artificial que desarrolla técnicas que permiten a máquinas tomar decisiones; también como un proceso automatizado que extrae patrones de los datos. Para construir los modelos utilizados en las aplicaciones de análisis predictivo de datos, se utiliza el aprendizaje automático supervisado (Kelleher et al., 2015). Los algoritmos de *machine learning* funcionan buscando a través de un conjunto de posibles modelos de predicción, el modelo que captura mejor la relación entre las características descriptivas y la característica de destino en un conjunto de datos.

A medida que los algoritmos incorporan datos de entrenamiento, es posible producir modelos más precisos basados en dichos datos. En general, en Aprendizaje de Máquina e Inteligencia Artificial, cuando se encuentran en un entorno controlado, reciben una gran cantidad de datos y sus respectivos resultados con los que se podrá crear relaciones entre ideas (Aprendizaje Supervisado). El aprendizaje supervisado generalmente comienza con un conjunto establecido de datos y cierta comprensión de cómo se clasifican, por lo que está destinado a encontrar patrones en los datos que se pueden aplicar a un proceso de análisis. Estos datos tienen características etiquetadas que definen su significado (Hurwitz & Kirsch, 2018).

El aprendizaje supervisado usa datos etiquetados por humanos y se usa comúnmente cuando los datos pueden predecir eventos probables. En otras palabras, es una entrada cuando se conoce la salida deseada. El algoritmo aprende un conjunto de entradas junto con las salidas correctas

correspondientes y aprende comparando su salida real con las salidas correctas para encontrar errores. Una vez que encuentra los errores, puede modificar el modelo en consecuencia.

2. METODOLOGÍA

De manera general, para el alcance planteado en este trabajo, el cual llega hasta la descripción del diseño de alto nivel de la solución propuesta, se ha abordado la siguiente metodología:

Creación de base de información: se establece una base de información actual e histórica acerca del comportamiento de las variables meteorológicas de interés y de la ocurrencia de heladas en la sabana de Bogotá, logrando así la creación de la base de aprendizaje e información para el modelo a desarrollar.

Figura 1. Metodología base de información.



157

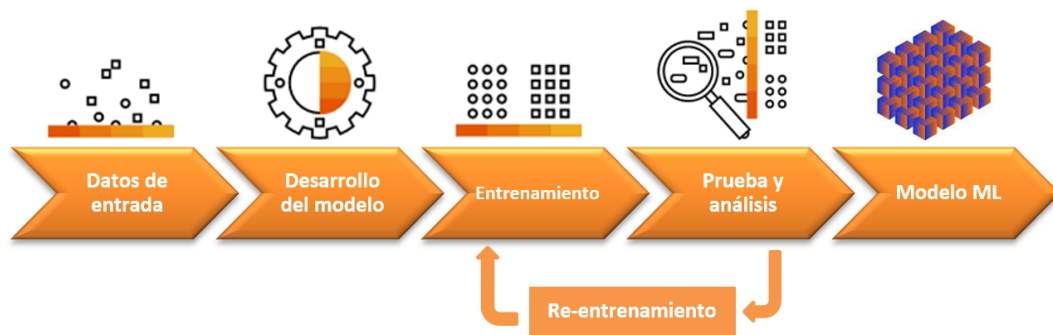
Fuente: (Elaboración propia, 2024)

La recolección de datos se realizó de las estaciones meteorológicas del Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM pertenecientes a la red de heladas y ubicadas en la región de la Sabana de Bogotá. Para este caso, las variables de interés fueron: Humedad Relativa, Precipitación, Velocidad del Viento, Temperatura: mínima y máxima diarias, al suelo (10 cm). La depuración y análisis de la información se ha realizado mediante las herramientas Python y MATLAB®, eliminando vacíos y acondicionando los datos para su procesamiento (modelo de regresión múltiple).

Caracterización del modelo: se caracteriza el sistema y se definen los requerimientos de diseño. En esta etapa se plantea también determinar la herramienta de desarrollo en la cual se creará y bajo la cual el modelo *machine learning* determinará el comportamiento de las variables meteorológicas definidas y de las heladas en la Sabana de Bogotá.

El sistema se basa en un modelo de aprendizaje supervisado (Figura 2), incorporando árboles de decisión (estructura donde en cada nodo se hace una pregunta y, en función de la respuesta, se sigue hacia un nodo u otro) como base del modelo. Una vez caracterizados y definidos los parámetros, se inician las etapas de desarrollo y simulación.

Figura 2. Modelo de Machine Learning.

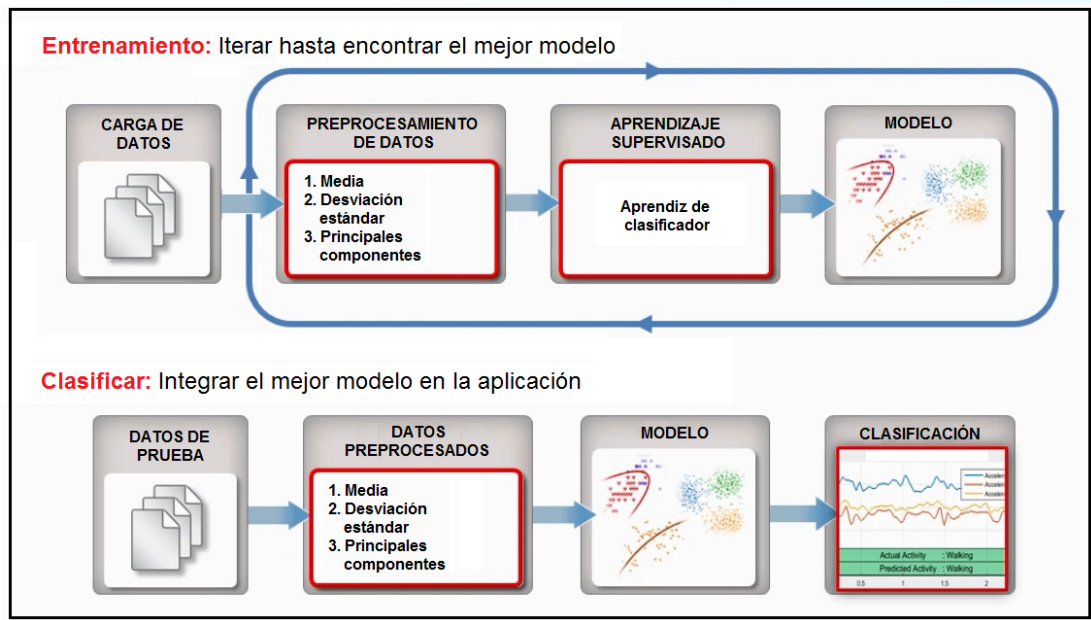


Fuente: Castillo Méndez (2023)

Desarrollo y simulación del sistema: una vez especificado el sistema en función de los requerimientos de diseño y de los parámetros definidos, de acuerdo a la herramienta de desarrollo y/o programación seleccionada, se realiza el diseño y simulación del modelo Machine Learning propuesto, empleando las herramientas Python y MATLAB®. El diseño y desarrollo del sistema consta de dos grandes etapas: entrenamiento basado en iteraciones hasta identificar el modelo ML de mejor desempeño; y la clasificación, la cual integra el modelo determinado en la solución, como lo muestra la Figura 3.

A un alto nivel, el diseño del sistema consta de los siguientes bloques básicos: el primer bloque del sistema se encarga de acceder a la información. A continuación, se realiza la extracción de características y una vez conformado el vector de entrada, durante la etapa de desarrollo, esta información se usará para el entrenamiento (aprendizaje supervisado, Figura 2). Eso permitirá determinar las capas y pesos requeridos para la red neuronal, la cual generará un vector de salida que se corresponderá con alguna de las salidas esperadas. Desde el punto de vista de interfaz de usuario, se desplegará la información correspondiente a la salida del sistema, lo cual equivale a indicar el estado de las variables climatológicas de interés y probabilidad de ocurrencia de la helada.

Figura 3. Metodología para desarrollo del modelo ML.



Fuente: Elaboración propia basada en MathWorks®, 2024.

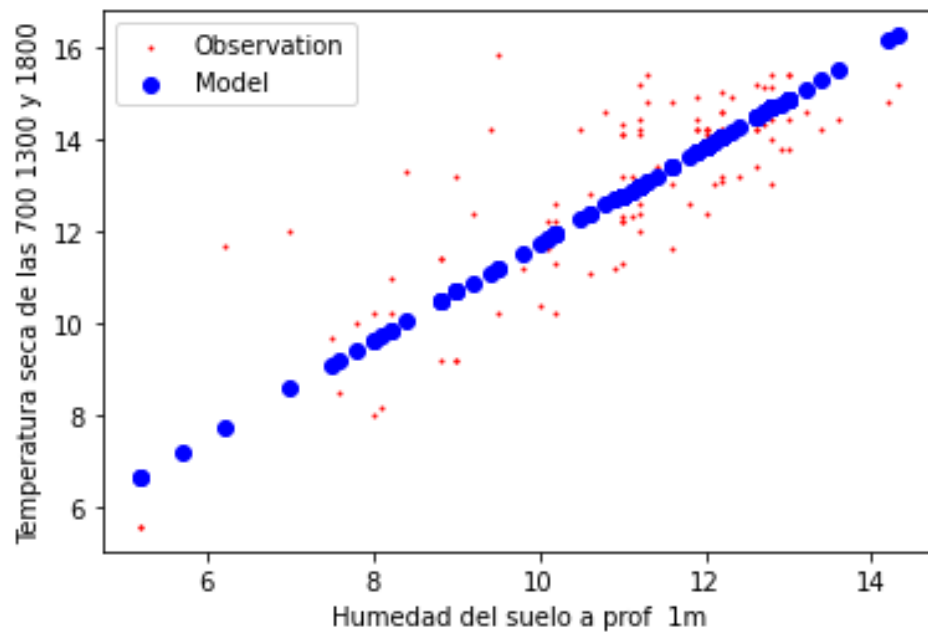
Implementación del sistema: una vez simulado el sistema, se adecuará el ambiente de operación y pruebas técnicas y funcionales a nivel de laboratorio. (Etapa en desarrollo).

Validación del sistema: mediante una prueba de desempeño, se aplicarán pruebas funcionales, verificando el desempeño del modelo. Una vez realizado el análisis de los resultados, se determinará si se realiza un ciclo de ajuste y nuevas pruebas.

3. RESULTADOS

Se evaluó parcialmente el comportamiento del modelo frente a las variables climatológicas de interés. Como se aprecia en la Figura 4, el comportamiento del modelo se acerca a lo esperado de acuerdo a los datos de las observaciones realizadas:

Figura 4. Comportamiento del modelo.



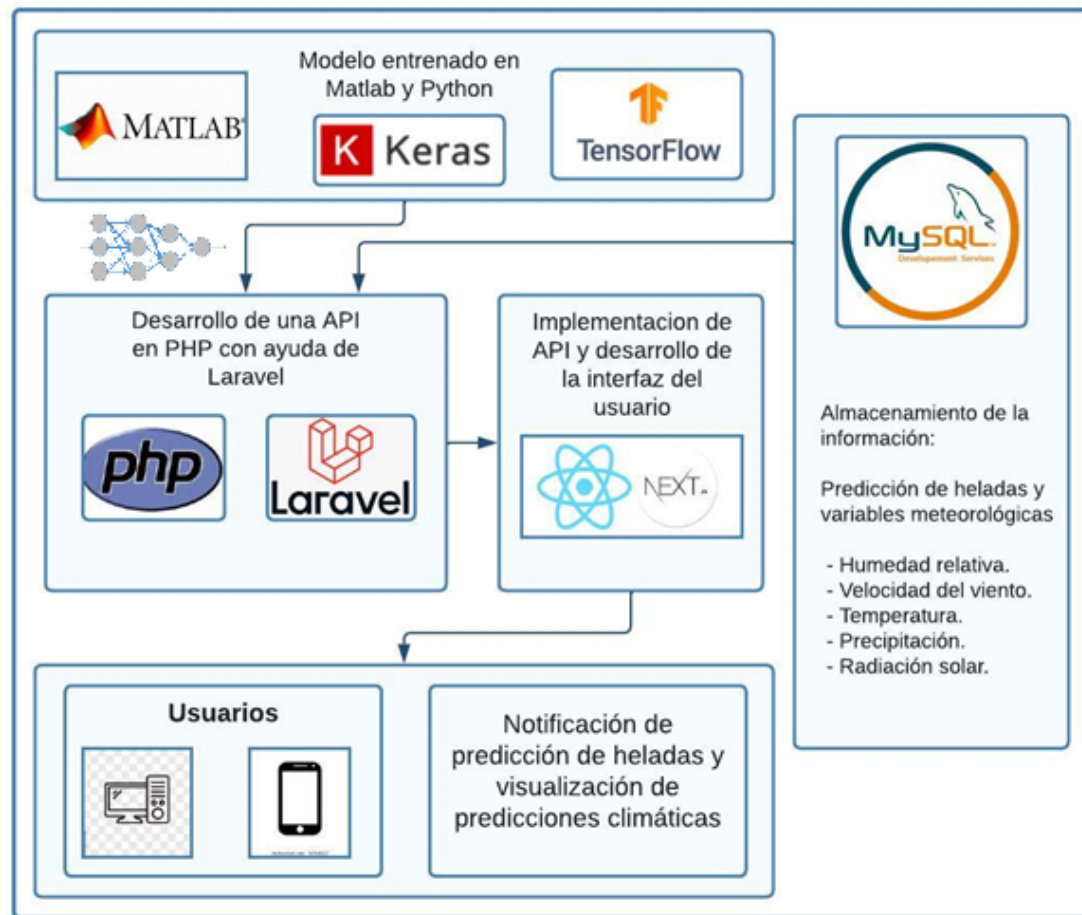
Fuente: Elaboración propia.

160

Es importante mencionar que, dentro del alcance dado a partir de los requerimientos del sistema, además de determinar el comportamiento de las variables meteorológicas y la probabilidad de evento de helada, a nivel de interfaz de usuario, la información deberá mostrarse en un aplicativo web de tal manera que las predicciones realizadas puedan ser consultadas por cualquier usuario.

La arquitectura y herramientas del sistema han tenido las siguientes consideraciones: para el acondicionamiento y análisis de datos, desarrollo e implementación del sistema, se ha realizado el diseño y entrenamiento del modelo combinando las herramientas Python y MATLAB®, y se ha estructurado la incorporación de una Interfaz de Programación de Aplicaciones (API) soportando el almacenamiento de la información en una base de datos que permita la provisión requerida para administrar el flujo de datos requerido (ver Figura 5).

Figura 5. Arquitectura de alto nivel y herramientas del sistema.



Fuente: elaboración propia.

El algoritmo para el modelo aprende un conjunto de entradas junto con las salidas correctas correspondientes y lo hace comparando su salida real con las salidas correctas para encontrar errores, establecer los parámetros y realizar estimaciones y predicciones. A medida que va identificando errores, puede ajustar los parámetros del modelo en consecuencia (pesos de la red).

A nivel de interfaz de usuario, las figuras 6 y 7, muestran la vista de la aplicación web y su estructura, respectivamente. Un estado de variables climatológicas y predicción de probabilidad de helada son mostrados inicialmente, adicionalmente, se ha contemplado que el usuario pueda generar reportes de predicciones sobre la región de interés (permitido a partir del registro del usuario).

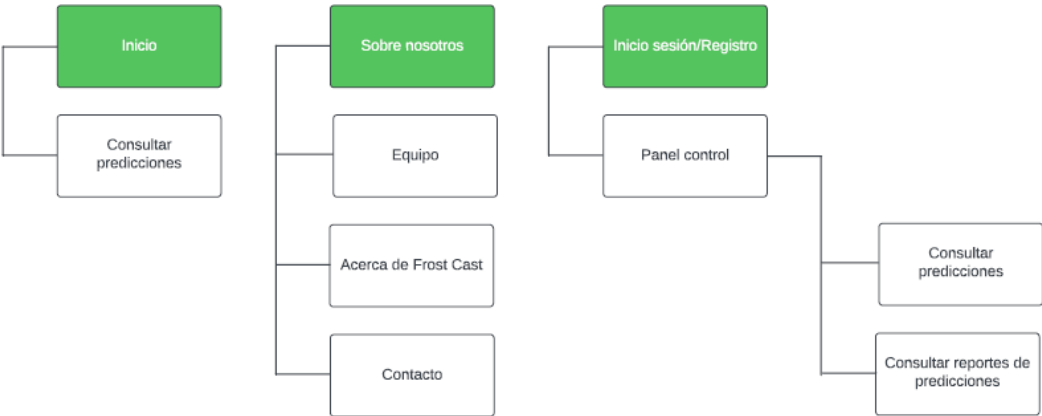
El proceso de implementación aún se encuentra en ejecución. La validación y pruebas funcionales para verificar el desempeño del modelo y de todo el sistema, permitirá realizar un análisis de los resultados y determinar si se realiza un ciclo de ajuste y nuevas pruebas.

Figura 6. Landing aplicación web para interfaz de usuario.



Fuente: (Elaboración propia, 2024)

Figura 7. Estructura aplicación web para interfaz de usuario.



Fuente: Elaboración propia.

4. CONCLUSIONES

Se han presentado de manera general los avances realizados en el desarrollo de un modelo de aprendizaje automático para determinar heladas y el comportamiento de variables climatológicas relevantes en agricultura en la Sabana de Bogotá. Dentro de la definición de requerimientos, se ha establecido que el despliegue hacia el usuario se realizará mediante un aplicativo web. Los principales avances en este aspecto también se abordan en este documento.

El aprendizaje de máquina (Machine Learning) tiene amplia aplicación en diferentes procesos y en diferentes áreas; se define como un proceso automatizado que extrae patrones de los datos y es un subcampo importante de la Inteligencia Artificial, tecnología relevante dentro de la Industria 4.0, por esto es importante para el Grupo de Investigación GICS generar capacidades internas que le permitan cerrar brechas de conocimiento en estas tecnologías, a partir del desarrollo de procesos de investigación aplicada y desarrollo, como el presentado en este trabajo.

En relación a las heladas, el enfoque del desarrollo de este trabajo es abordar las que se ocasionan por radiación. Cuanto más variables se consideren, existen mayores probabilidades de obtener un modelo con un nivel de asertividad más alto. Para efectos del presente trabajo se han considerado las variables climáticas de Humedad Relativa, Precipitación, Velocidad del Viento, Temperatura: mínima y máxima diarias, al suelo (10 cm), lo anterior por disponibilidad de información histórica en la región.

Es recomendable dar continuidad al proceso de desarrollo del sistema para poder materializar una solución que apoye el proceso de predicción de variables climatológicas y eventos de heladas en la Sabana de Bogotá, y fortalecer las herramientas tecnológicas disponibles para el sector agrícola de la región, problemática inicial que originó el planteamiento y ejecución del proyecto descrito. Así mismo, es importante fortalecer la base de conocimiento y realizar futuras iteraciones de entrenamiento y re-entrenamiento para robustecer el sistema que será implementado **más adelante**.

Como trabajo futuro, de manera inmediata se encuentra finalizar las fases de implementación y validación. Así mismo, se espera que al finalizar el proyecto, se puedan entregar resultados concluyentes en cuanto a la posibilidad de generar un modelo de predicción temprana de eventos de heladas basado en información agrometeorológica histórica. Es importante mencionar que esto corresponde a un trabajo inicial en dicha temática y resulta necesario validar la metodología presentada y sistema a obtener con registros asociados a diferentes estaciones meteorológicas ubica-

das en distintas zonas geográficas del país para, de esta manera, lograr establecer una generalidad de mayor aplicación al modelo propuesto.

5. REFERENCIAS

- Barooni, M., Ziarati, K., & Barooni, A. (2023). Frost Prediction Using Machine Learning Methods in Fars Province. *2023 28th International Computer Conference, Computer Society of Iran, CSICC 2023*. <https://doi.org/10.1109/CSICC58665.2023.10105391>
- Castillo Méndez, R. (2023). Clasificación automática de defectos de paneles solares aplicando aprendizaje de máquina. *Revista Teinnova*, 7, 89–97. <https://doi.org/10.23850/25007211.5750>
- Corporación Autónoma Regional de Cundinamarca (CAR). (2020). *Informe del registro e impactos de heladas en el territorio CAR durante Enero de 2020*
- Cunha, R. L. F., Silva, B., & Netto, M. A. S. (2018). A scalable machine learning system for pre-season agriculture yield forecast. *Proceedings - IEEE 14th International Conference on EScience, e-Science 2018*, 423–430. <https://doi.org/10.1109/eScience.2018.00131>
- García Cañón, H. S. (2019). Implementación de técnicas de machine learning para la predicción de variables meteorológicas y del suelo que afectan la agricultura. In *Implementación de técnicas de machine learning para la predicción de variables meteorológicas y del suelo que afectan la agricultura*. Universidad de los Andes.
- Gonzalez Botero, D. F. (2018). *Planteamiento de un modelo de predicción de heladas en cultivos de rosa en la sabana de Bogotá*. Universidad Militar Nueva Granada.
- Hurwitz, J., & Kirsch, D. (2018). Machine Learning. In *Journal of the American Society for Information Science* (Vol. 35, Issue 5). John Wiley & Sons, Inc.
- Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics* (T. M. Press, Ed.). The MIT Press.
- Marqués Gozalbo, M. Á. (2020). *Modelos predictivos de producción agroindustrial con Machine Learning a partir de fuentes de información pública*. Universidad de Córdoba.
- Ramírez Gómez, C. A. (2020). Aplicación del Machine Learning en agricultura de precisión. *Revista Cintex*, 25(2), 14–27.
- Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine Learning applications for precision agriculture: A comprehensive review. *IEEE Access*, 9, 4843–4873. <https://doi.org/10.1109/ACCESS.2020.3048415>
- Snyder, R. L., de Melo-Abreu, J. Paulo., & Matulich, Scott. (2010). *Protección contra las heladas: fundamentos, práctica y economía* (2nd ed.). Organización de las Naciones Unidas para la Agricultura y la Alimentación.