

Modelo para estimar la producción de aceite de palma en Colombia

Robert Sebastián Castellanos Rodríguez¹

¹Analista del Sistema de información estadística del sector palmero (Sispa). Autor para correspondencia: rcastellanos@fedepalma.org
Federación Nacional de Cultivadores de Palma de Aceite, Fedepalma.

Introducción

La demanda del mercado global de aceite de palma se ha intensificado en las últimas décadas, posicionando a Colombia como un productor líder en América Latina. Tradicionalmente, las estimaciones de producción se han basado en modelos estáticos que utilizan datos de censos nacionales desactualizados y relaciones básicas de área-rendimiento. Tales métodos no tienen en cuenta la interacción dinámica de factores climáticos y prácticas agrícolas en evolución. En contraste, las técnicas modernas de aprendizaje automático (ML) ofrecen la capacidad de analizar grandes conjuntos de datos multifactoriales y capturar relaciones no lineales complejas que los modelos tradicionales pasan por alto.

El estudio actual aborda estas limitaciones proponiendo un modelo predictivo basado en la Máquina de Aprendizaje Extremo (ELM), un tipo de red neuronal conocida por su rápido entrenamiento y alta precisión (Huang *et al.*, 2012). Al integrar diversas fuentes de datos, incluidos registros meteorológicos y tasas de extracción, el enfoque propuesto busca proporcionar una previsión más oportuna de la producción de palma de aceite. Este modelo no solo está dirigido a mejorar la precisión, sino también a apoyar la toma de decisiones estratégicas para productores, gerentes de la cadena de suministro y formuladores de políticas en el sector agroindustrial.

Además, el creciente impacto del cambio climático, marcado por sequías más frecuentes, patrones de lluvia irregulares y fluctuaciones de temperatura, exige modelos predictivos que sean adaptativos y robustos. La integración del aprendizaje automático en la estimación de la producción de aceite de palma crudo representa un avance significativo en la agricultura de precisión, ofreciendo beneficios potenciales como la optimización de recursos, la mejora de la gestión de riesgos y la estabilidad del rendimiento (Mohd Nain. F.N *et al.*, 2022).

Objetivos

- Implementar un modelo de red neuronal que permita analizar grandes conjuntos de datos multifactoriales, integrando registros meteorológicos y prácticas agronómicas regionales, para prever la producción de palma de aceite con mayor precisión y rapidez.
- Desarrollar un modelo robusto y adaptativo que responda a las variaciones climáticas, como sequías, patrones de lluvia irregulares y fluctuaciones de temperatura, mejorando la gestión de riesgos y la estabilidad del rendimiento del modelo.
- Estimar la producción de aceite de palma crudo en Colombia para el año 2024 con el modelo que quede seleccionado luego de analizar las métricas de rendimiento.

Metodología

Los datos utilizados en este estudio fueron obtenidos de Fedepalma, abarcando registros mensuales de producción de 2015 a 2023. El conjunto de datos incluye las siguientes variables: volumen de producción (toneladas) de aceite de palma crudo (APC); datos climáticos como temperatura, precipitación y humedad relativa y tasas de extracción y rendimientos. El preprocesamiento de datos incluyó: 1) Manejo de valores atípicos mediante el método de Tukey. 2) Normalización de variables para mejorar la estabilidad del modelo. 3) Segmentación del conjunto de datos en entrenamiento (80 %) y prueba (20 %) para la validación del modelo.

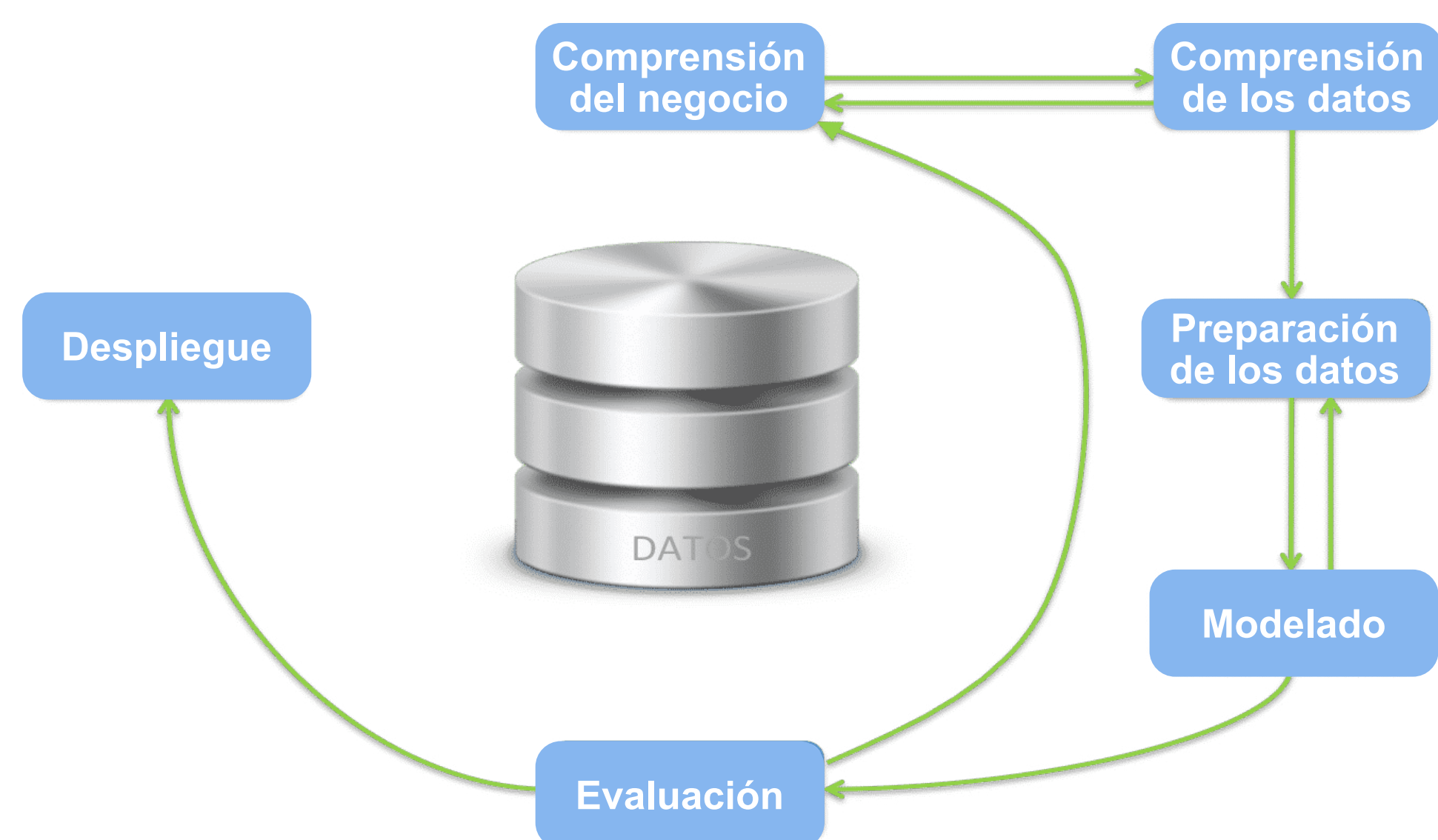


Figura 1. Metodología CRISP-DM.

KNN: número de vecinos más próximos: 3. Distancia: euclidiana.
ANN: número de neuronas en la capa oculta: 10. Tipo de función de activación: TANH (función tangente hiperbólica).
RF: número de árboles: 150.
SVM: tipo de función de núcleo: función de base radial.
ELM: tipo de función de activación: Sigmoide, Número de neuronas en la capa oculta: 11.

Para medir la precisión de los modelos, se utilizaron las siguientes métricas de evaluación:

Tiempo de ejecución: determina la eficiencia computacional del modelo, midiendo el tiempo requerido para completar la predicción.
Sensibilidad: mide la capacidad del modelo para identificar correctamente los casos positivos.
Especificidad: evalúa la capacidad del modelo para identificar los casos negativos.
Error Porcentual Absoluto Medio (MAPE): indica el porcentaje de error en la predicción.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \times 100$$

Si el MAPE es menor a 10, el pronóstico es altamente preciso; si está entre 10-20, el pronóstico es bueno; si está entre 20-50, el pronóstico es razonable y si es mayor a 50, el pronóstico es impreciso (Lewis, C.D. 1982).

Resultados

Los modelos de aprendizaje automático (ML) fueron evaluados en términos de su capacidad predictiva sobre la producción de aceite de palma crudo del 2023. Los resultados obtenidos se presentan en la Figura 2, donde se comparan las métricas de error de cada modelo.

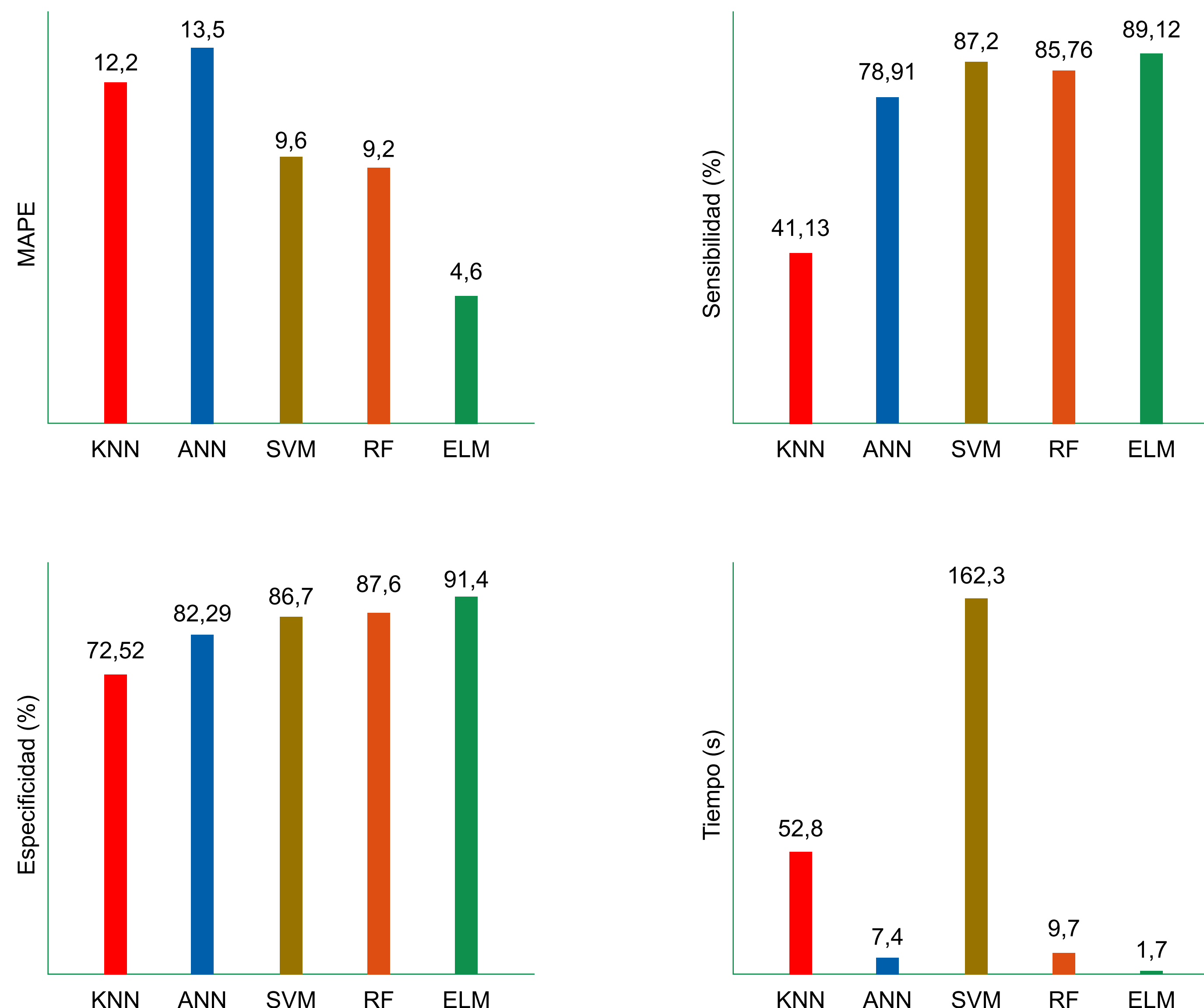


Figura 2. Comparación de Modelos.

Después de la evaluación de diferentes técnicas de aprendizaje automático, se escogió el método de *Extreme Learning Machine* (ELM) para hacer el análisis predictivo de la producción de palma. Al tener el MAPE más bajo (4,6), menor tiempo de tratamiento (1.7 segundos).

Una vez seleccionado el modelo (ELM), se procedió a ejecutarlo para estimar la producción de aceite de palma en 2024. El modelo se corrió en marzo de 2024, arrojando una predicción de 1.710.000 toneladas. Al comparar este valor con la producción real reportada en la plataforma Sisplus, al cierre de febrero de 2025, se constató que la producción real fue de 1.720.000 toneladas a nivel nacional, lo que implica un MAPE de aproximadamente 0,58 % según la fórmula. Este MAPE tan bajo indica una precisión excepcional del modelo, ya que la diferencia relativa entre la predicción y el valor real es mínima, demostrando la alta eficacia y confiabilidad del enfoque ELM para estimar la producción de aceite de palma.

Conclusiones

Este estudio confirma que la aplicación de técnicas avanzadas de aprendizaje automático, en particular la Máquina de Aprendizaje Extremo (ELM), permite realizar estimaciones altamente precisas de la producción de aceite de palma en Colombia. La comparación con modelos tradicionales y otras metodologías de aprendizaje supervisado demuestra que el ELM ofrece la mejor combinación de precisión (MAPE menor a 5), rapidez en el procesamiento (1.7 segundos) y robustez frente a variaciones en los datos.

A futuro, la metodología desarrollada podría expandirse a otras regiones y cultivos, permitiendo mejorar la eficiencia productiva en diferentes contextos agrícolas. Asimismo, la integración de fuentes de datos adicionales, como imágenes satelitales y sensores IoT, podría potenciar aún más la capacidad predictiva del modelo. Con estos avances, se refuerza el papel del aprendizaje automático en la agricultura de precisión, promoviendo la sostenibilidad y la competitividad del sector palmero en Colombia y América Latina.

Referencias bibliográficas

- Huang, G., Zhou, H., Ding, X., Zhang, R. (2012). Extreme Learning Machine for Regression and Classification. IEEE Transactions on Systems, Man, and Cybernetics.
- Lewis, C.D. (1982). Industrial and business forecasting methods. London: Butterworths.
- Mohd Nain. F.N., Ahamed Hassain Malim. N.H., Abdullah. R., Abdul Rahim. M.F., Ahmad Mokhtar, M.A., Mohamad Fauzi. N.S. (2022). A Review of an Artificial Intelligence Framework for Identifying the Most Effective Palm Oil Prediction. 15, 218.

Agradecimientos

Los autores ofrecen sus agradecimientos al Fondo de Fomento Palmero, administrado por Fedepalma, por la financiación.