

PREDICCIÓN DE BROTES DE *Astaena pygidialis* EN EL CULTIVO DE AGUACATE HASS UTILIZANDO MÉTODOS DE APRENDIZAJE AUTOMÁTICO.

Oscar Pulgarin¹, Edna Z Rincon²

ABSTRACT

Pests are currently one of the major problems in food production worldwide, having a significant impact on agricultural productivity and therefore on the food security of populations, so early detection of outbreaks is key to reduce productivity losses, the economic impact they cause and the reduction of competitiveness in international markets. Machine learning (ML) methods have shown great promise in predicting the presence of pests in agriculture, offering an efficient and accurate alternative to traditional methods of pest monitoring, so the objective of this work is to predict outbreaks of the brown cucarron insect based on an open dataset, in which the climatic variables were fundamental for the prediction. Three Logistic Regression classification algorithms were evaluated, which showed an outstanding performance in the test set with an AUC of 1.00; however, overfitting was evidenced and corroborated by cross-validation. The KNN model obtained an AUC of 0.94 with a more consistent behavior, indicating a better generalization capacity in unseen contexts. Finally, the Random Forest model achieved a balance of precision (0.87) and sensitivity (0.91), which makes it a robust alternative for pest detection. The confusion matrix and the F1-score, which positions it as a reliable option for early warning systems or support in integrated pest management decisions, validated the good performance of the model.

Keywords: Pest prediction, machine learning, classification, regression, marceño beetle

RESUMEN

Las plagas en la actualidad constituyen uno de los mayores problemas en la producción de alimentos a nivel mundial, teniendo un impacto significativo en la productividad agrícola y por ende en la seguridad alimentaria de las poblaciones, por lo cual la detección temprana de brotes es clave para reducir la pérdida de producción, las afectaciones económicas y la reducción de competitividad en los mercados internacionales. Los métodos de aprendizaje automático (AA) han demostrado ser muy prometedores para predecir la presencia de plagas en la agricultura, ofreciendo una alternativa eficiente y precisa a los métodos tradicionales de monitoreo de plagas, por lo cual el objetivo de este trabajo es realizar la predicción de brotes del insecto cucarron marceño basando en datos abiertos, en el cual las variables climáticas fueron fundamentales para la predicción. En este contexto se evaluaron tres algoritmos de clasificación, Regresión Logística que mostró un rendimiento sobresaliente en el conjunto de prueba con un AUC de 1.00; sin embargo, se evidenció sobreajuste (overfitting) corroborado por validación cruzada. El modelo KNN el cual obtuvo una AUC de 0.94 con un comportamiento más consistente, lo que indica una mejor capacidad de generalización en contextos no vistos. Y por último el modelo Random Forest el cual logró un equilibrio de precisión (0.87) y sensibilidad (0.91), lo que lo convierte en una alternativa robusta para la detección de la plaga, el buen desempeño del modelo fue validado por la matriz de confusión y el F1-score, lo posiciona como una opción confiable para sistemas de alerta temprana o apoyo en decisiones de manejo integrado de plagas.

Palabras clave: Predicción plagas, aprendizaje automático, clasificación, Regresión, cucarrón marceño

¹Ingeniero Eléctrico, Universidad Autónoma de Occidente, Santiago de Cali. Especialización en Telemática, Universidad Autónoma & Universidad del Cauca, Country. Email: oscar_f.pulgarin@uao.edu.co

²Administradora Ambiental, Universidad Autónoma de Occidente, Santiago de Cali; Colombia. Email: Edna_zul.rincon@uao.edu.co

Introducción

En la actualidad el aguacate *Persea americana* cv Hass está tomando gran relevancia en los mercados internacionales, la producción mundial de aguacate mostró una tendencia de crecimiento equivalente al 6,4 % anual. Para Colombia las exportaciones de la fruta han mantenido un crecimiento promedio del 126 %; panorama que explica la ampliación de áreas sembradas y la participación de nuevos actores en la cadena productiva nacional siendo considerado como uno de los principales frutales en Colombia. [1] sin embargo, la producción y calidad se está viendo afectado por plagas de importancia económica como lo son el complejo *Melolonthidae* (*Astaena pygidialis* Kirsch) y la chinche *Monalonion* (*Monalonion veleangeli* Carvalho y Costa), que comprometen la producción y por ende el posicionamiento en mercados internacionales. [2]

Astaena pygidialis Kirsch, conocido comúnmente como cucarrón marceño es una de las plagas de mayor importancia económica para la producción de aguacate hass en Colombia, los daños ocasionados se generan cuando el insecto perfora el material vegetal y succiona la savia dejando lesiones en los frutos, lo cual deteriora su calidad ocasionando secamiento y afectando el crecimiento del mismo. [3]

Es una de las especies más común en el oriente antioqueño, su aparición se presenta entre los meses de marzo y junio, son denominados comúnmente “marceños” o “cuaresmeros” debido a su frecuente aparición en el mes de marzo o durante el tiempo de cuaresma, que coincide con la llegada de las lluvias; este evento climático facilita la salida de los adultos del suelo. Los cucarrones marceños atraviesan por cuatro estados de desarrollo, huevo, larva, pupa y adulto; están presentes en varias regiones de Colombia y su diversidad e importancia varían de una región. [4]

Según datos de producción, los agricultores del Oriente antioqueño pierden hasta el 30 % de la fruta debido al daño causado por los marceños. Esta plaga puede atacar entre el 40 % y el 60 % de la fruta formada, además de afectar hojas y

flores jóvenes, las cuales quedan rasgadas o esqueletizadas. Por esta razón, se considera una plaga de importancia económica, ya que deteriora la calidad de los frutos y sus larvas pueden dañar las raíces al alimentarse de ellas, lo que podría generar pérdidas aún mayores si no se implementa un seguimiento y control adecuados. [5]

La variabilidad de las plagas en relación con los factores climáticos puede ser simulada mediante modelos agro-meteorológicos. Para ello, se emplean técnicas de aprendizaje automático, las cuales se basan en el análisis de datos y buscan automatizar la creación de modelos predictivo, este enfoque también se utiliza para identificar patrones mediante la teoría del aprendizaje computacional y la inteligencia artificial. Estos algoritmos son herramientas prometedoras para realizar análisis a gran escala de forma eficiente y precisa. Entre los modelos más utilizados se encuentran el K-vecino más cercano (KNN), la regresión lineal múltiple (RLM), las redes neuronales artificiales y el bosque aleatorio (RF), entre otros. [6]

Estas plagas han causado reducción en la calidad y rendimiento de la producción del aguacate, las estrategias más comunes para el control de estas en su mayoría es la aplicación de agroquímicos para su control, estrategia que genera contaminación al medio ambiente y a la salud de los agricultores además de incrementar los costos de producción, por lo tanto, el uso de algoritmos de aprendizaje automático puede predecir la aparición de plagas para generar estrategias de control eficientes y ambientalmente amigables. [7]

En ese contexto el objetivo de este artículo es correlacionar la presencia de Marceño con las variables climáticas para predecir la incidencia de la plaga en el cultivo de aguacate utilizando algoritmos de aprendizaje automático como lo son Regresor de K vecinos (KNN); Regresor de bosque aleatorio (RFT) y Regresor Logístico (RL).

Trabajos Relacionados

La predicción de brotes de plagas mediante métodos de aprendizaje automático cuenta con precedentes. En los últimos años, se ha incrementado la investigación sobre estas técnicas en la agricultura, dado su potencial para generar alertas tempranas, sin embargo, persisten brechas considerables en la disponibilidad de datos con información relevante, siendo escasos para realizar predicciones precisas. Por ello, gran parte de los trabajos recientes se han centrado en el uso de imágenes para el reconocimiento e identificación de plagas, mas no en predicción de las mismas, mas adelante se detallan algunos trabajos relacionados con predicción de plagas.

Autores como *F. Rodríguez et al., 2024*, realizaron pronósticos de clasificación y regresión basados en técnicas de aprendizaje automático para la predicción de plagas en cultivos de olivares donde los modelos mostraron una alta precisión en la predicción de brotes de mosca del olivo, considerando variables como la velocidad del viento, y la radiación solar, determinantes para la propagación de la plaga en el cultivo.[8]

Oliveria et al., 2020, utilizaron algoritmos de aprendizaje automático para predecir la incidencia de la roya del café, cercospora, el minador y la broca del café, con datos meteorológicos obtenidos de campo en plantaciones de café en la región de Cerrado Mineiro, en Brasil, los algoritmos probados para la predicción fueron Regresión Múltiple (RLM), Regresor de K vecinos (KNN); Regresor de bosque aleatorio (RFT) y Redes neuronales artificiales (MLP), Las variables climáticas que mostraron correlaciones significativas con la enfermedad de la roya del café fueron la temperatura máxima del aire, el número de días con humedad relativa y la humedad relativa. Y el RFT fue el que mostró mayor precisión en las predicciones.[6]

De igual manera *Domingues, T. Brandão, y J. C. Ferreira, 2022* realizaron una revisión exhaustiva sobre la aplicación de técnicas de aprendizaje automático en la detección y predicción de enfermedades y plagas en cultivos, destacando diversos modelos de aprendizaje automático utili-

zados para la predicción de plagas en cultivos. Entre ellos, las Máquinas de Soporte Vectorial (SVM) las cuales son valoradas por su capacidad para clasificar datos no lineales con alta precisión, las Redes Neuronales Artificiales (ANN) y las Redes Neuronales Convolucionales (CNN) que se emplean ampliamente en el reconocimiento de patrones complejos, especialmente en el análisis de imágenes, los Árboles de Decisión y los Bosques Aleatorios (Random Forests) los cuales son apreciados por su robustez y facilidad de interpretación. Además mencionan modelos como K-Vecinos más Cercanos (KNN), útiles en tareas de clasificación simple, y modelos de ensemble como XGBoost, que combinan algoritmos para mejorar la precisión y finalmente, los modelos de series temporales, como ARIMA y LSTM, se utilizan para predecir la aparición de plagas en función del tiempo y condiciones climáticas. [7]

En ese contexto los métodos de aprendizaje automático ampliamente utilizados son los árboles de decisión que se destacan por su eficiencia y precisión en la predicción de plagas cuando los datos son sesgados y escasos además los bosques aleatorios también tienen un alto rendimiento en tareas de clasificación y predicción superando otros métodos[9], existiendo factores claves que influyen en la precisión como son la selección de variables, que esta relacionada con los factores climáticos y que tienen incidencia directa en la precisión de los modelos, las características de los datos es otro factor fundamental porque de ello depende el rendimiento de los modelos de aprendizaje automático según el tamaño y calidad de los mismos. [10]

METODOLOGIA

1. Descripción del Conjunto de Datos

Este conjunto de datos fue tomado de Mendeley disponible como recurso abierto, la información fue recolectada en una finca productora de aguacate Hass ubicada en la zona rural del municipio de Rionegro, en el oriente antioqueño de Colombia, durante los años 2022 y 2023. Contiene información sobre variables agroclimáticas rele-

vantes, como temperatura, precipitación, humedad relativa, radiación solar y velocidad del viento. [11]

Además, incluye registros sobre los daños causados por dos plagas identificadas durante el monitoreo del cultivo: el complejo *Melolonthidae* (*Astaena pygidialis* Kirsch) y la chinche *Monalonion* (*Monalonion velezeangeli* Carvalho y Costa). La información se obtuvo a partir del seguimiento de treinta árboles de aguacate Hass seleccionados aleatoriamente y distribuidos en diferentes lotes de la finca.

El conjunto también incorpora datos de rendimiento correspondientes a dos cosechas, junto con detalles específicos de los árboles evaluados. Esta base de datos permite realizar análisis predictivos del cultivo, facilitando herramientas para que los agricultores de la región gestionen los riesgos asociados a eventos fitosanitarios y condiciones agroclimáticas, con el objetivo de optimizar la producción.

En resumen, este conjunto de datos proporciona información valiosa sobre las variaciones agroclimáticas, el impacto de plagas y los niveles de producción del aguacate Hass en la zona de estudio, constituyendo un insumo clave para la toma de decisiones y la gestión de riesgos en el contexto agrícola. En las siguientes tablas se muestran las variables de los datos asociados a variables climáticas, datos de árboles y datos de la presencia de la plaga, no se muestra datos de la producción porque no se tuvo en cuenta para esta investigación.

Tabla 1. Datos Variables Climáticas

rawClimatePart1.csv; rawClimatePart2.csv		
Variable	Nomenclature	Unit
Record timestamp	timestamp	YYYY-MM-DD HH:MM
Relative humidity	relativeHumidity	Percent (%)
Air temperature	airTemp	Degrees Celsius (°C)
Rainfall	Rainfall	Millimeters of rainfall (mm)

Wind speed	speedWind	Meters per second (m/s)
Solar radiation	radiationSolar	Watts per square meter (W/m ²)
Raw milivolts value	rawSolarValue	Milivolts (mV)
Solar total flux density	fluxSolar	Megajoules per square meter (MJ/m ²)

Tabla 2. Datos Arboles

dataTrees.csv		
Variable	Nomenclature	Unit
Tree identifier	idTree	
Planting date	datePlanting	YYYY-MM-DD
Latitude	latitude	Decimal degrees (°)
Longitude	longitude	Decimal degrees (°)
Altitude	altitude	Meters above sea level (masl)
Slope orientation	orientSlope	Cardinal and ordinal directions
Slope direction	dirSlope	Degrees (°)
Type of plantation land	typePlantation	
Drainage distance	distDrainage	Meters (m)
Drainage depth	depthDrainage	Meters (m)

Tabla 3. Datos Plaga Marceño

pestMarceno.csv		
Variable	Nomenclature	Unit
Tree identifier	idTree	
Observation date	dateObservation	YYYY-MM-DD
Pest presence	presenceMarceno	Binary (0-1)
Damage to old leaves	oldLeavesMarceno	NA = No damage BA = Low damage ME = Medium damage
Damage to new leaves	newLeavesMarceno	NA = No damage BA = Low damage

		ME = Medium damage
Damage to fruits	fruitDamage-Marceno	NA = No damage
		BA = Low damage
		ME = Medium damage

2. Observación y limpieza de Datos

El dataset analizado reporta datos para dos plagas (cucarrón marceño y *Monalonion spp.*); sin embargo, este estudio se enfocó exclusivamente en la predicción de incidencia del cucarrón marceño (*Astaena pygidialis Kirsch*). Para ello, se implementaron y compararon múltiples metodologías de aprendizaje automático.

En primer lugar, se utilizó un dataset denominado **“Variables agroclimáticas, daños causados por dos insectos plaga y su impacto en la producción de un cultivo de aguacate Hass Colombiano”**. Este conjunto de datos incluye registros individuales por árbol, datos de cosecha, dos bases de datos climáticas (2022 y 2023) con variables meteorológicas y dos bases de datos sobre información sobre la incidencia de dos plagas, lo cual se detalló anteriormente, estos registros permiten correlacionar las condiciones ambientales con los daños observados en el cultivo.

Una vez cargados los datos en Python, se procedió a realizar la observación, limpieza y filtrado de los mismos. Inicialmente, se utilizó la función `head()` para visualizar las primeras cinco filas de cada dataframe. Posteriormente, se llevó a cabo la conversión de fechas, asignándoles un formato reconocible para la librería pandas.

Además, se realizó la verificación de valores nulos y su debido proceso de limpieza. Durante este análisis, se identificó que en el dataframe **pest_marceno** las variables `oldleavesMarceno`, `NewLeavesMarceno` y `fruitDamageMarceno` presentaban una cantidad significativa de valores incompletos.

Asimismo, se observó que los datos climáticos de las partes 1 y 2 presentaban diferencias en sus columnas. En particular, la parte 2 incluía variables adicionales como radiación solar (W/m^2), milivoltios (mV) y densidad de flujo solar (MJ/m^2). Para garantizar consistencia tempo-

ral en el análisis, se seleccionaron únicamente las columnas presentes en ambos conjuntos de datos, obteniendo así un dataset homogéneo para todo el período de estudio. Posteriormente, se realizaron las siguientes acciones:

- Unificación de datos climáticos: Se combinaron las partes 1 y 2 (verificando las dimensiones con `raw_climate.shape`).
- Integración de datos de plagas y árboles: Se unió la información de plagas con las características de los árboles (analizada previamente mediante `df_marceno_tree.describe()`).
- Limpieza de datos climáticos: Dada la presencia de valores nulos mencionada anteriormente, se aplicó un promedio de valores para imputar los datos faltantes.
- Organización de datos de plagas: Se estructuraron los datos por árbol y fecha de observación.
- Creación del DataFrame final: Se construyó un dataset consolidado que relaciona las observaciones de plagas (por árbol y fecha) con los promedios climáticos calculados desde la última observación.

Finalmente, se generó una matriz de correlación para analizar las relaciones entre las variables numéricas del conjunto de datos integrado (`raw_climate_tree_pest`).

3. Metodos de Aprendizaje Automatico

Los métodos de pronóstico fueron los siguientes: Regresión Logística (RL); Regresor de K vecinos (KNN); Regresor de bosque aleatorio (RFT).

Se realizó la separación de variables donde la presencia de marceño es la variable dependiente, y los datos climáticos y características de los árboles, las variables independientes de los modelos.

En todas las metodologías, el 80% de los datos se utilizó para el entrenamiento y el 20% para la calibración de los modelos. Esta separación se

realizó mediante la biblioteca de Python (`sklearn.model_selection.train_test_split`).

Utilizamos el método ridge de RLM. Este método evita el mal condicionamiento de la matriz de las variables regresoras, controlando la inflación y la inestabilidad general encontrada en los estimadores de mínimos cuadrados. Ridge evita el problema de multicolinealidad sin tener que excluir variables regresoras, por lo que no tiene pérdida de información. El algoritmo KNN es una técnica simple y de fácil implementación y es muy flexible. En el KNN, se identificaron los tres vecinos más cercanos, y la métrica utilizada para calcular las distancias fue la distancia euclidiana. En el RFT, se creó aleatoriamente un bosque con un conjunto de 100 árboles de decisión para predecir plagas y enfermedades en función del clima. La red neuronal artificial utilizada fue el Perceptrón Multicapa (MLP), con tres capas de neuronas. En cada una de estas capas, se utilizaron diez neuronas (`hidden_layer_sizes = 10, 10, 10`). El entrenamiento de MLP se realizó mediante retropropagación.

Para evaluar el rendimiento de los modelos de clasificación se utilizaron diversas métricas estándar en aprendizaje automático:

- Accuracy (Precisión global): indica el porcentaje de predicciones correctas sobre el total de casos. Es útil cuando las clases están balanceadas.
- Recall (Sensibilidad o Tasa de Verdaderos Positivos): mide la capacidad del modelo para identificar correctamente los casos positivos (en este caso, la presencia de Marceño). Es especialmente importante cuando los falsos negativos tienen un alto costo.
- F1-score: es la media armónica entre la precisión y el recall, y ofrece un balance entre ambos. Es útil cuando se desea evaluar el rendimiento considerando tanto los errores por omisión como por comisión.
- AUC-ROC (Área bajo la curva ROC): representa la capacidad del modelo para

discriminar entre clases (presencia o ausencia de plaga). Un valor cercano a 1 indica un excelente poder discriminativo, mientras que un valor cercano a 0.5 indica un modelo sin capacidad predictiva.

RESULTADOS Y DISCUSIÓN

Correlación entre el clima y las plagas

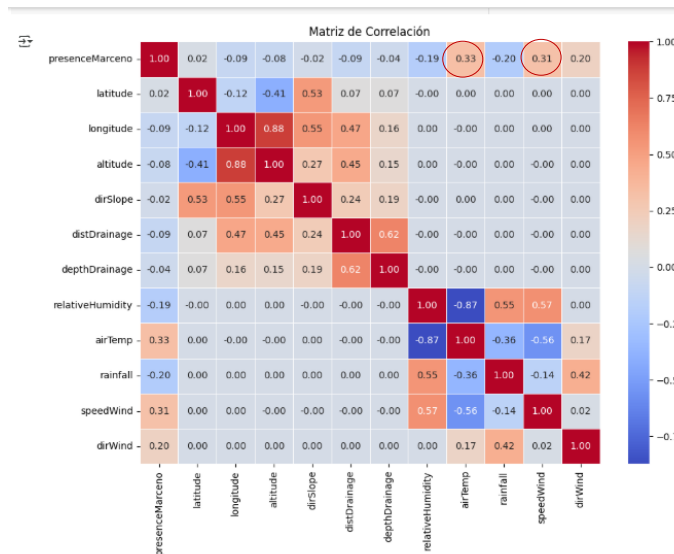
Comprender la relación que tiene los factores climáticos y la distribución de plagas es fundamental para el establecimiento de estrategias de control y manejo integrado de brotes en cultivos, en esta evaluación se realizó la correlación de variables climáticas y la presencia de cucarrón marceño en el cultivo de aguacate hass. (Ver grafica 1)

Dentro de las variables climáticas consideradas, la temperatura del aire presenta una correlación positiva moderada con la presencia del marceño con un valor correspondiente de 0.33, lo que indica que las condiciones cálidas pueden favorecer la aparición de esta plaga, de igual manera se observó correlación positiva en la velocidad del viento con un valor de 0.31 y la dirección del viento con 0.20, lo que puede indicar que el viento cumple un rol en la dispersión espacial del insecto pero en una magnitud menor.

Respecto a la humedad relativa con un valor de -0.19 y la precipitación de -0.20, las correlaciones fueron negativas, este hallazgo puede contradecir algunos reportes previos de la literatura que señalan la presencia del insecto durante los meses lluviosos, sin embargo la matriz de correlación representa tendencias globales en el espacio y la literatura hace referencia a patrones estacionales, es posible que las lluvias sean el inicio de la actividad biológica de la plaga, mientras su establecimiento y condiciones climáticas sean favorables para su desarrollo.

Estos resultados resaltan la necesidad de unir datos espaciales y temporales en futuros modelos de predicción, así como de validar los hallazgos en campo con monitoreos en diferentes periodos climáticos, para tener resultados más orientadores.

Grafica 1. Matriz correlación

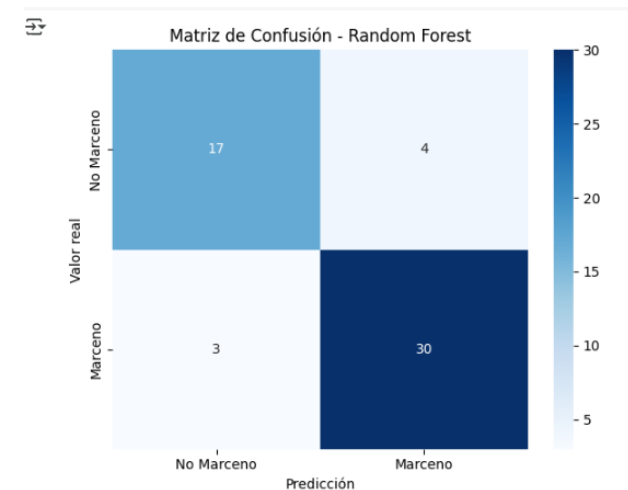


Evaluación de algoritmos

Según los modelos aplicados a partir de variables ambientales se entrenaron y evaluaron tres algoritmos de clasificación que se presentan a continuación, las métricas de evaluación empleadas fueron accuracy, recall, f1-score y AUC, ROC, matriz de confusión y validación cruzada.

El modelo **Random Forest** presento un buen desempeño con una precisión (accuracy) de 0,87, la información del reporte de clasificación también obtuvo un recall de 0.909 y un f1-score de 0.896 para la clase positiva es decir presencia del insecto, lo que indica una alta capacidad para realizar la identificación de casos positivos como se indica en la matriz de confusión (ver grafica 2) donde los resultados son 30 verdaderos positivos versus 3 falsos negativos.

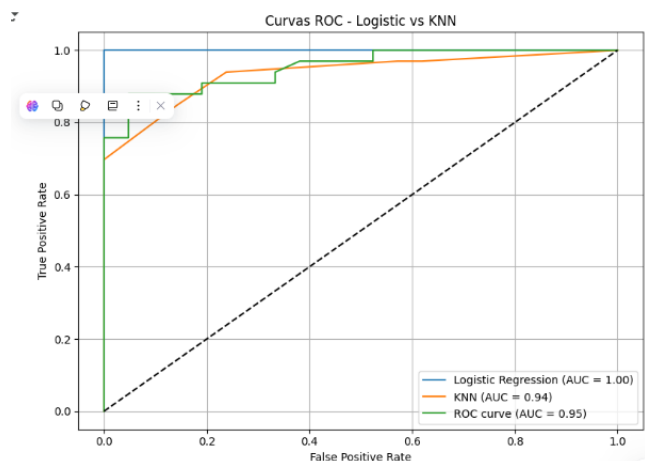
Grafica 2. Matriz de Confusión RFT



El modelo **KNN** presento un buen rendimiento con una precisión del 0.85 y un f1-score de 0.87, el área bajo de curva ROC fue de 0.935 lo que refleja una capacidad alta discriminativa, sin embargo su rendimiento fue menos al Random Forest con un recall de 0.848.

Por ultimo el modelo de **Regresión Logística** en una observación inicial las métricas se muestran aparentemente buenas con un accuracy de 1.0 y demás métricas con valores iguales iguales, sin embargo cuando se realizó la validación cruzada se logró evidenciar un comportamiento típico de overfitting es decir sobre ajuste con un AUC de 0.804 y un f1-score de 0.79.

Grafica 3. Curvas ROC



CONCLUSIONES

- En el análisis de correlación, la velocidad del viento se identifica como una variable climática relevante que puede influir en el diseño agroforestal del cultivo. Esta información resulta útil para la implementación de estrategias como la instalación de barreras naturales o árboles en los bordes del cultivo, que actúan como protección frente a condiciones adversas asociadas al viento y propagación de la plaga.
- De los algoritmos evaluados el modelo Random Forest fue el que logró un mejor equilibrio, una alta precisión y una alternativa robusta para la detección de la plaga, comparado con los demás modelos que uno de ellos presentó sobre ajuste, esto se valida y sustenta con la literatura mencionada en apartados anteriores.
- Estos enfoques permiten desarrollar sistemas predictivos más eficientes y sostenibles en el manejo agrícola, estas tecnologías pueden reducir la dependencia de pesticidas, minimizando así los impactos ambientales, los costos de producción que facilitan decisiones agrícolas más sostenibles y eficientes.

DISPONIBILIDAD DE DATOS

<https://data.mendeley.com/datasets/69nhkcxhp9/3> [11]

REFERENCIAS

[1] L. Díaz Ramírez, A. M. Bravo Parra, M. J. García Botina, M. Jäger, y J. J. Hurtado, Modelo empresarial de competitividad e inclusión de productores de aguacate Hass de pequeña escala en el Valle del Cauca. Bogotá, Colombia / Palmira, Colombia: Editorial Universidad Nacional de Colombia / Universidad Nacional de Colombia, 2021, 82 p.

[2] F. Torres Jaimes et al., "Relationship of Monalonion veezangeli Carvalho & Costa (Hemiptera: Miridae) with the phenology of avocado (Persea americana Mill., cv. Hass)," Rev. Fac.

Nac. Agron. Medellín, vol. 65, no. 2, pp. 6659–6665, 2012. [Online]. Available: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0304-28472012000200010&lng=en&tlng=en

[3] A. Carabalí Muñoz, A. M. Caicedo Vallejo, y C. M. Holguín, Guía para el reconocimiento y manejo de las principales plagas de aguacate cv. Hass en Colombia. Corporación Colombiana de Investigación Agropecuaria – AGROSAVIA; ICA, 2021. [Online]. Available: <https://doi.org/10.21930/agrosavia.nbook.7404913>

[4] J. A. Bernal y C. A. Díaz, Comps., Actualización tecnológica y buenas prácticas agrícolas (BPA) en el cultivo de aguacate, 2.^a ed. Corporación Colombiana de Investigación Agropecuaria – AGROSAVIA, 2020.

[5] T. Kondo et al., Insectos y ácaros. [Online]. Available: <http://hdl.handle.net/20.500.12324/36879>

[6] L. de Oliveira Aparecido, G. de Souza Rolim, J. R. da Silva Cabral De Moraes et al., "Machine learning algorithms for forecasting the incidence of Coffea arabica pests and diseases," Int. J. Biometeorol., vol. 64, pp. 671–688, 2020. [Online]. Available: <https://doi.org/10.1007/s00484-019-01856-1>

[7] Domingues, T. Brandão, y J. C. Ferreira, "Machine learning for detection and prediction of crop diseases and pests: A comprehensive survey," Agriculture, vol. 12, no. 9, art. 1350, 2022. [Online]. Available: <https://doi.org/10.3390/agriculture12091350>

[8] F. Rodríguez-Díaz, A. M. Chacón-Maldonado, A. R. Troncoso-García, y G. Asencio-Cortés, "Explainable olive grove and grapevine pest forecasting through machine learning-based classification and regression," Data Science & Big Data Lab, Universidad Pablo de Olavide, Seville, Spain, Oct. 11, 2024. [Online]. Available: <https://doi.org/10.1016/j.simpat.2024.102163>

[9] Kishi S, Sun J, Kawaguchi A, Ochi S, Yoshida M, Yamanaka T. Characteristic features of statistical models and machine learning methods derived from pest and disease monitoring datasets. *R Soc Open Sci.* 2023 Jun 28;10(6):230079. doi: 10.1098/rsos.230079

[10] E. C. Sumido, L. S. Feliscuzo y C. J. G. Aliac, “Clasificación y predicción de plagas: análisis del impacto del clima en la aparición de plagas me-

dante aprendizaje automático,” *Revista de Ciencia y Tecnología de Ingeniería*, 2023.

[11] Ramirez-Guerrero, Tomas; Hernandez-Perez, Maria Isabel; Tabares, Marta S.; Villanueva, Eduart; Peña, Alejandro (2024), “Agroclimatic variables, damage caused by two insect pests and their impact on the production of a Colombian Hass avocado crop”, *Mendeley Data*, V3, doi: 10.17632/69nhkcxhp9.3