

Maestría en

CIENCIA DE DATOS Y MÁQUINAS DE APRENDIZAJE CON MENCIÓN EN INTELIGENCIA ARTIFICIAL

Trabajo previo a la obtención de título de
Magister en Ciencia de Datos y Máquinas de
Aprendizaje

AUTOR/ES:

Samuel José Peña Bautista
Lourdes Valeria Andino Célleri
Ana Gabriela Morocho Quingalagua
José Eduardo Alarcón López
Luis Patricio Calo Catota

TUTOR/ES:

Karla Estefanía Mora Cajas
Fernanda Paulina Vizcaíno Imacaña

TEMA: “Estimación de la producción de Banano en el litoral
ecuatoriano mediante modelos de Machine Learning e integración
de variables agroclimáticas (2012-2024)”


Certificación de autoría

Nosotros, **Samuel José Peña Bautista, Ana Gabriela Morocho Quingalagua, José Eduardo Alarcón López, Lourdes Valeria Andino Céleri, Luis Patricio Calo Catota**, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.



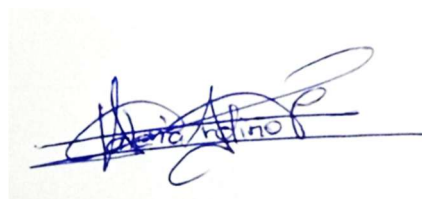
Firma
Samuel José Peña Bautista



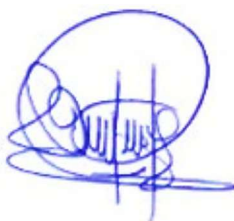
Firma
Ana Gabriela Morocho Quingalagua



Firma
José Eduardo Alarcón López



Firma
Lourdes Valeria Andino Céleri



Firma
Luis Patricio Calo Catota

Autorización de Derechos de Propiedad Intelectual

Nosotros, **Samuel José Peña Bautista, Ana Gabriela Morocho Quingalagua, José Eduardo Alarcón López, Lourdes Valeria Andino Céleri, Luis Patricio Calo Catota**, en calidad de autores del trabajo de investigación titulado ***Estimación de la producción de Banano en el litoral ecuatoriano mediante modelos de Machine Learning e integración de variables agroclimáticas (2012-2024)***, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, 19 diciembre 2025



Firma
Samuel José Peña Bautista



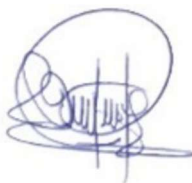
Firma
Ana Gabriela Morocho Quingalagua



Firma
José Eduardo Alarcón López



Firma
Lourdes Valeria Andino Céleri

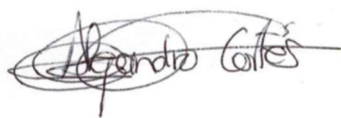


Firma
Luis Patricio Calo Catota

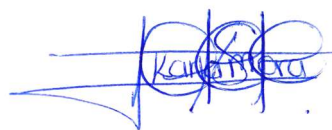
Aprobación de dirección y coordinación del programa

Nosotros, **Alejandro Cortes López Director EIG y Karla Estefanía Mora Cajas**

Coordinadora UIDE, declaramos que: Samuel José Peña Bautista, Ana Gabriela Morocho Quingalagua, José Eduardo Alarcón López, Lourdes Valeria Andino Célleri, Luis Patricio Calo Catota son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortes López
Director/a de la
Maestría en Ciencia de Datos y Máquinas
de Aprendizaje



Karla Estefanía Mora Cajas
Coordinador/a de la
Maestría en Ciencia de Datos y Máquinas
de Aprendizaje

DEDICATORIA

Dedicamos el presente trabajo a nuestras familias, por el apoyo constante, la comprensión y la motivación brindada durante todo el proceso de desarrollo de esta investigación. Su acompañamiento ha sido fundamental para superar los desafíos académicos y alcanzar los objetivos propuestos.

De igual manera, dedicamos este trabajo a todas las personas que, de forma directa o indirecta, contribuyeron a nuestra formación académica y profesional, y cuyo apoyo y ejemplo hicieron posible la culminación de esta etapa de formación.

AGRADECIMIENTOS

En primer lugar, agradecemos a Dios por guiar cada paso de este camino, por darnos fortaleza en los momentos de dificultad y por ser la luz que orientó nuestras decisiones a lo largo de este proceso académico.

A quienes, a lo largo de este camino, nos inspiraron a crecer profesionalmente y a ser mejores personas; a aquellos cuya sola presencia se convierte en un impulso constante, en un momento de alegría, que celebran sinceramente nuestros logros y permanecen a nuestro lado con una palabra de aliento cuando sentimos que no podemos continuar.

Un agradecimiento especial a los docentes y miembros de la maestría, por compartir generosamente sus conocimientos y experiencias, dejando en nosotros no solo aprendizajes académicos, sino también parte de su vocación, esfuerzo y compromiso profesional.

Resumen

El presente trabajo desarrolla un modelo predictivo mediante técnicas de aprendizaje automático para estimar la producción anual de banano en la región Costa del Ecuador, utilizando información levantada en el periodo 2012–2024. El modelo integra datos agro-productivos oficiales de la Encuesta de Superficie y Producción Agropecuaria Continua ESPAC y variables climáticas satelitales de NASA POWER.

Con el objetivo de superar las limitaciones de modelos estadísticos tradicionales para representar la complejidad del sistema agrícola, se evaluaron múltiples arquitecturas de modelo, incluyendo regresión lineal, máquinas de vectores de soportes (SVM) y métodos de ensamble. Estas arquitecturas fueron analizadas tanto en términos de desempeño predictivo como de interpretabilidad, apoyándose en técnicas de explicabilidad del modelo (SHAP), las cuales permitieron identificar a la superficie cosechada y a variables estructurales asociadas como las predictoras dominantes del rendimiento.

Los resultados obtenidos evidencian que el algoritmo Bagged Trees (Ensamble de árboles) presenta el desempeño más robusto entre los modelos evaluados, alcanzando un coeficiente de determinación (R^2) superior al 0.95 durante la fase de entrenamiento. De igual forma, la validación temporal con datos correspondientes al 2024 confirmó su alta capacidad de predicción, registrando un error porcentual medio (MAPE) de 1.69% para la provincia de Los Ríos y 5.57% para Guayas. Estos resultados consolidan al modelo propuesto como una herramienta sólida y confiable para la planificación y evaluación de escenarios en el sector bananero del Ecuador.

Palabras clave: Producción de banano, Machine Learning, Modelos de ensamble, Análisis agroclimático, Costa ecuatoriana.

Abstract

This research develops a predictive model using machine learning techniques to estimate annual banana production in the Ecuadorian Coastal region, utilizing data collected over the period 2012–2024. The model integrates official agro-productive data from the Continuous Agricultural Surface and Production Survey (ESPAC) with satellite-derived climatic variables from NASA POWER.

To overcome the limitations of traditional statistical models in capturing the complexity of agricultural systems, multiple modeling architectures were evaluated, including linear regression, Support Vector Machines (SVM), and ensemble-based methods. These architectures were assessed both in terms of predictive performance and interpretability, supported by model explainability techniques (SHAP), which identified harvested areas and associated structural variables as the dominant predictors of production variability.

The results indicate that the Bagged Trees algorithm (ensemble of decision trees) exhibits the most robust performance among the evaluated models, achieving a coefficient of determination (R^2) greater than 0.95 during the training phase. Likewise, temporal validation using data from 2024 confirmed its high predictive capability, registering a Mean Absolute Percentage Error (MAPE) of 1.69% for Los Ríos province and 5.57% for Guayas. These findings consolidate the proposed model as a reliable and effective tool for scenario evaluation and strategic planning within Ecuador's banana production sector.

Keywords: Banana production; Machine Learning; Ensemble models; Agroclimatic analysis; Ecuadorian Coast.

TABLA DE CONTENIDOS (Índice)

Capítulo 1	1
Definición del proyecto	1
Justificación e importancia del trabajo de investigación	2
Alcance	3
Objetivos.....	5
Capítulo 2.....	6
Revisión de Literatura	6
Conceptos fundamentales	6
Estado de Arte	7
Marco Teórico	14
Capítulo 3.....	18
Desarrollo	18
Metodología.....	18
Fuentes primarias de información	21
Obtención de datos	22
Depuración, análisis preliminar y tratamiento de valores atípicos	25
Interpretación y análisis de importancia de variables.....	35
Evaluación y validación del modelo.....	40
Análisis de escenarios.....	48
Arquitectura del sistema web	49
Backend	50

Frontend.....	50
Visualización Geográfica y 3D.....	52
Integración Frontend – Backend.....	52
Despliegue en la nube	53
Seguridad, rendimiento y mantenibilidad.....	53
Estrategia de pruebas y verificación del sistema web.....	54
Capítulo 4.....	56
Análisis de Resultados.....	56
Evaluación del rendimiento de modelos entrenados (2012-2023)	56
Análisis de Resultados	59
Capítulo 5.....	63
Conclusiones.....	63
Recomendaciones	64
Referencias Bibliográficas	66

LISTA DE TABLAS (Índice de tablas)

Tabla 1 <i>Comparación de artículos científicos</i>	11
Tabla 2 <i>Coordenadas geográficas utilizadas por provincia para la obtención de variables climáticas mediante la API NASA POWER</i>	23
Tabla 3 <i>Configuración temporal del filtrado de datos por provincia y muestras finales</i>	32
Tabla 4 <i>Importancia y correlación de las variables de entrada según los métodos Random Forest, F-Test, MRMR y Spearman</i>	38
Tabla 5 <i>Características climáticas seleccionadas para el modelo</i>	39
Tabla 6 <i>Resultados de la evaluación del Modelo 1 y sus variantes sobre el conjunto de prueba</i>	41
Tabla 7 <i>Resultados de la evaluación del modelo Support Vector Regression (SVR)</i>	44
Tabla 8 <i>Valores optimizados de los hiperparámetros obtenidos para el modelo Support Vector Regression (SVR)</i>	44
Tabla 9 <i>Hiperparámetros optimizados para el modelo Random Forest Regressor</i>	45
Tabla 10 <i>Resultados de la evaluación del modelo Random Forest Regressor en su configuración base y optimizada</i>	46
Tabla 11 <i>Componentes principales del Frontend</i>	51
Tabla 12 <i>Comparación del desempeño de los algoritmos de regresión mediante R^2, RMSE y MAE</i>	57

LISTA DE FIGURAS (Índice de figuras)

Figura 1 <i>Descripción del proceso metodológico</i>	18
Figura 2 <i>Producción total anual de banano en la región Costa del Ecuador durante el periodo 2012–2024</i>	26
Figura 3 <i>Producción anual de banano por provincia en el Ecuador (2012–2024)</i>	30
Figura 4 <i>Volumen total de producción del dataset de entrenamiento considerando años y provincias con datos válidos</i>	34
Figura 5 <i>Producción anual por provincia en el dataset final de entrenamiento excluyendo años atípicos</i>	35
Figura 6 <i>Comparación de métodos de interés con variables normalizadas</i>	37
Figura 7 <i>Impacto direccional de variables (SHAP)</i>	38
Figura 8 <i>Evolución de la producción total en contraste con la superficie plantada y cosechada</i>	42
Figura 9 <i>Impacto de variables en la producción (SHAP Beeswarm)</i>	43
Figura 10 <i>Evaluación anual de variables meteorológicas</i>	47
Figura 11 <i>Estabilidad interanual de variables durante el periodo 2012-2024</i>	47
Figura 12 <i>Interfaz general del sistema de estimación de producción de banano</i>	50
Figura 13 <i>Visualización 3D (GLB) sobre provincia seleccionada.</i>	52
Figura 14 <i>Verificación de datos presentados en la interfaz web</i>	54
Figura 15 <i>Comparación entre valores reales y predichos del modelo</i>	59
Figura 16 <i>Comparación entre la producción real y la producción predicha en 2024 por provincia utilizando el modelo Bagged Trees.</i>	60

Figura 17 <i>Análisis de sensibilidad del impacto de la superficie cultivada sobre la producción estimada de banano utilizando el modelo Bagged Trees</i>	61
--	----

Capítulo 1

Definición del proyecto

El presente proyecto desarrolla un modelo predictivo de aprendizaje automático orientado a estimar la producción anual del banano, uno de los cultivos permanentes más relevantes de la región Costa del Ecuador, utilizando registros históricos del periodo (2012-2024). La investigación integra bases oficiales de producción agropecuaria de la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC), los cuales son registrados a escala anual, con variables agroclimáticas derivadas de fuentes satelitales como NASA POWER, aplicando metodologías que permitan agregar y seleccionar los meses climáticamente más influyentes para el ciclo productivo del banano. Esta integración permite modelar relaciones no lineales entre variabilidad climática, condiciones ambientales y rendimiento agrícola, superando las limitaciones inherentes a los modelos estadísticos tradicionales.

Diversas investigaciones demuestran que los factores climáticos influyen de manera no lineal en la productividad agrícola y que los enfoques convencionales basados en regresiones lineales presentan limitaciones para capturar estos efectos. En este escenario, el uso de algoritmos avanzados de Machine Learning (ML) permite descubrir patrones ocultos, mejorar la precisión predictiva y generar métricas útiles para la planificación agrícola, gestión del riesgo climático y optimización de recursos.

El enfoque planteado en este proyecto contribuye al desarrollo de herramientas analíticas robustas orientadas a comprender el impacto del clima sobre un cultivo estratégico como el banano, evaluar posibles escenarios futuros y apoyar la toma de decisiones basada en datos en el sector agro productivo ecuatoriano. Además, establece un marco replicable para estudios

posteriores centrados en otros cultivos o que incorporen modelos de predicción multiescalar basados en datos espaciales o series temporales de mayor granularidad.

Justificación e importancia del trabajo de investigación

El banano es uno de los principales cultivos del Ecuador y de numerosos países tropicales, no solo por su alta demanda internacional, sino por su impacto económico y social que genera a través de su cadena de valor. La producción de banano sostiene actividades agrícolas, logísticas y comerciales que contribuyen a la generación de divisas, empleo y estabilidad económica de distintos sectores agro-productivos. En este contexto, el poder estimar la producción futura del banano con base en datos históricos y variables climáticas resulta fundamental para la planificación agrícola, optimización de recursos y reducción de incertidumbre en un entorno influenciado por factores ambientales complejos.

La relación entre variables meteorológicas y rendimiento es muy marcada en cultivos tropicales como el banano, los cuales responden de manera sensible a variaciones de temperatura, precipitación, humedad del suelo y radiación solar. Esta relación requiere de un enfoque que permita analizar datos no lineales y patrones que los modelos tradicionales no siempre logran describir adecuadamente.

Las tendencias actuales en ciencia de datos y agricultura de precisión demuestran que técnicas como regresión estadística, modelos de aprendizaje automático y algoritmos de predicción multivariable representan una herramienta de alto potencial para anticipar el comportamiento productivo de cultivos agrícolas. Sin embargo, en el caso del banano, existe una brecha importante en la literatura regional: los estudios predictivos disponibles son escasos o se

centran en agrupaciones de cultivos, dejando un vacío respecto al análisis específico de este cultivo emblemático.

Por ello, la presente investigación propone un enfoque metodológico robusto para analizar la producción bananera e implementar modelos predictivos que permitan estimar el rendimiento futuro del cultivo en la región costa del Ecuador. Esta investigación no solo busca aportar un conocimiento técnico valioso sobre la relación entre variables climáticas y productivas, sino que también ofrece un conjunto de herramientas que puedan ser utilizadas por productores, técnicos agrícolas y exportadores para la toma de decisiones estratégicas.

De igual forma, el uso de modelos estadísticos y técnicas de aprendizaje automático contribuyen a la modernización del análisis agrícola, alineándose con prácticas internacionales de agricultura y ciencia de datos aplicada. La construcción de modelos predictivos confiables permite avanzar hacia sistemas de alerta temprana, optimización de ciclos de producción y mejores en la eficiencia de recursos. De este modo, el trabajo se justifica en términos científicos, económicos y prácticos, consolidándose como un aporte para el análisis de cultivos en el Ecuador.

Alcance

El alcance de este proyecto comprende el desarrollo, validación y comparación de modelos predictivos de aprendizaje automático orientados a estimar la producción anual de banano en la región Costa del Ecuador, para el periodo 2012-2024. El estudio se limita a las provincias para las cuales existe información anual consolidada en la ESPAC: El Oro, Esmeraldas, Guayas, Los Ríos, Santa Elena y Manabí.

Dado que la producción agrícola disponible es anual, mientras que las variables climáticas satelitales presentan resolución mensual, se aplica metodologías de agregación y selección de meses relevantes que permitan generar predictores climáticos consistentes con los ciclos de producción del banano. Se consideran variables climáticas como temperatura, precipitación, humedad, radiación solar y condiciones atmosféricas que influyen de manera directa en el rendimiento del cultivo.

El alcance metodológico incluye la comparación de algoritmos de ML como arboles de decisión, Random Forest, XGBoost y máquinas de vectores de soporte, así como procesos de selección de características mediante técnicas estadísticas y basadas en modelos. Los algoritmos se evaluarán mediante métricas como MAE, RMSE y R^2 para seleccionar el modelo con mejor desempeño.

El proyecto incorpora el desarrollo de un dashboard interactivo y una visualización web que permita comunicar los resultados del modelo y facilitar el análisis por parte de usuarios técnicos. Quedan fuera del alcance aspectos relacionados con manejo agronómico, presencia de plagas, fertilización, logística productiva o modelación espacial.

Objetivos

Objetivo general

Desarrollar un modelo predictivo de aprendizaje automático para estimar la producción anual de banano en la región Costa del Ecuador durante el periodo 2012-2024, integrando información agroproductiva oficial con variables climáticas satelitales, a fin de identificar patrones entre clima y rendimiento.

Objetivo específico

- Recolectar y procesar datos de producción anual de banano provenientes de la ESPAC, así como variables agroclimáticas de NASA POWER, correspondientes a las provincias de El Oro, Manabí, Esmeraldas, Guayas, Santa Elena y Los Ríos para el periodo 2012-2024.
- Depurar e integrar el conjunto de datos consolidado mediante procesos de limpieza, validación, tratamiento de valores atípicos y estandarización, con el fin de armonizar adecuadamente la información.
- Analizar de manera exploratoria y multivariable el comportamiento histórico de la producción anual de banano, identificando tendencias, variaciones interanuales y años atípicos.
- Entrenar y evaluar modelos de algoritmos de aprendizaje supervisado para determinar el modelo con mejor desempeño predictivo según métricas de validación (MAE, RMSE, R^2).
- Implementar un dashboard interactivo que permita comunicar los resultados del modelo, analizar escenarios y facilitar la interpretación por parte de usuarios.

Capítulo 2

Revisión de Literatura

Conceptos fundamentales

Este capítulo presenta una síntesis de los avances recientes en la modelización de rendimientos agrícolas frente a cambios climáticos. Se aborda la problemática desde una perspectiva que abarca variables agronómicas, climatológicas y de ciencia de datos. El análisis se centra en el estado actual de técnicas de machine Learning (ML) y su capacidad de capturar relaciones no lineales entre la variabilidad climática y la productividad de cultivos. En este contexto de investigación, se define el rendimiento agrícola como la producción de un cultivo por unidad de superficie.

- **Rendimiento agrícola y complejidad de su predicción:** El rendimiento en la producción de cultivos se ve afectado por diversas variables; pronosticar su variación requiere de modelos capaces de representar relaciones no lineales y dependencias temporales. Estimar el rendimiento con anticipación permite planificar la producción y comercialización de los productos, específicamente en regiones con alta variabilidad climática.
- **Variables climáticas y agroclimáticas:** Las variaciones meteorológicas como precipitación, temperatura máxima/mínima, humedad relativa y evaporación son determinantes para el análisis del rendimiento. Los índices agroclimáticos combinan todas estas variables para entrenar modelos de predicción.
- **Aprendizaje automático (ML) y aprendizaje profundo (DL):** Modelos de ML supervisado se centran en predecir resultados a partir de datos y se diferencian de la estadística clásica al priorizar la predicción predictiva sobre la inferencia causal. Los modelos de ML pueden manejar relaciones complejas y datos de distintos formatos; entre ellos se destacan

métodos basados en árboles, máquinas de soporte vectorial, redes neuronales multicapa y sistemas neuro difusos. Modelos de DL, como LSTM y Transformers, capturan dependencias temporales y permiten integrar grandes volúmenes de datos.

- **Inteligencia artificial explicable (XAI):** El uso de XAI permite interpretar como los modelos complejos llegan a una predicción. Técnicas como SHAP o LIME cuantifican la contribución de cada variable y facilitan la adopción de modelos para agricultores. La interpretabilidad es esencial cuando se busca influenciar en prácticas de manejo o producción.

Estado de Arte

En los últimos años, la aplicación de la inteligencia artificial y el Machine Learning en la agricultura ha cobrado fuerza debido a su capacidad para mejorar la precisión en la predicción de rendimientos y ser un apoyo en la toma de decisiones. La literatura reciente muestra una transición clara desde métodos estadísticos tradicionales hacia enfoques computacionales más robustos, capaces de capturar relaciones no lineales y patrones complejos en sistemas agro-productivos.

Modelos estadísticos clásicos. Por muchos años la relación clima-rendimiento se modeló mediante regresiones lineales (LR) y modelos econométricos. Estos modelos suelen asumir relaciones lineales o polinomiales. Estudios recientes como el de Sidhu et al. (2023), comparan LR con boosted regression trees (BRT) para estimar el impacto del clima en arroz, mostrando que LR subestima la complejidad de las interacciones y BRT mejora considerablemente la precisión del modelo. Sin embargo, los autores advierten que modelos demasiado flexibles pueden confundir efectos temporales con variables climáticas. Crane-Droesch (2018) demostró que las redes neuronales semiparamétrica superan en precisión a modelos estadísticos y neural networks

no paramétricas, además de mostrar proyecciones menos pesimistas de impactos del cambio climático. Este trabajo marca una transición hacia modelos de ML.

Adopción de modelos de aprendizaje. A partir del aumento de datos climáticos y computacionales, la década de 2020 vio un auge en modelos de árboles de decisión, bosques aleatorios, gradient boosting y métodos ensemble. Estos modelos son capaces de capturar relaciones no lineales. La investigación de Clercq y Mahdi (2022) sobre arroz en la india entrenó diecinueve modelos de ML, incluyendo CatBoost, LightGBM y Extra Trees, utilizando datos climáticos y satelitales. Los mejores modelos alcanzaron coeficientes de determinación de hasta 0,82 y errores medios absolutos por debajo de 0,29; estos resultados superan a estudios previos demostrando la viabilidad de aplicar modelos de ML. En Mohan et al. (2025) integran modelos de árboles (Decision Tree Regressor, Random Forest Regressor, LightGBM Regressor) y análisis exploratorio de datos para pronosticar rendimientos en un entorno de agricultura de precisión. Los modelos lograron un $R^2=0,92$ con errores cuadráticos medio (MSE) de 0,02 y errores absolutos promedios de 0,015.

Aprendizaje profundo y series de tiempo. Para cultivos como arroz, Thirumal y Latha (2023) emplearon un modelo multiplicativo LSTM integrado con simulaciones DSSAT para evaluar impactos de variaciones climáticas en arroz, mostrando que las series temporales permiten representar anomalías de temperatura y precipitación. La formalización de la predicción como problemas de series temporales fue abordada por Tamayo-Vera et al. (2024) para el caso de la papa en Prince Edward Island. El estudio demostró que incluir dependencias temporales mejora sustancialmente la precisión y permite proyectar reducciones de rendimiento de hasta 70% bajo escenarios de altas emisiones SSP5-8.5. Estos estudios resaltan que los modelos

tradicionales tienen a tratar variables climáticas como entradas estáticas y omiten la variabilidad del rendimiento.

Otro aporte en DL es el de Vignesh et al. (2023), quienes proponen un modelo híbrido de Deep Belief Network (DBN) con una red VGG optimizado mediante un algoritmo Chicken Swarm Optimization (CSO) para clasificar y predecir la producción agrícola. El modelo procesa parámetros ambientales y de manejo en múltiples capas, alcanzando una precisión de 97% superando modelos existentes. Este tipo de arquitectura evidencia como las redes neuronales profundas pueden extraer características complejas, mejorando la predicción del rendimiento.

Modelos híbridos y metamodelos. Además de modelos ML y DL, se han desarrollado modelos híbridos que acoplan simuladores biogeofísicos con algoritmos de aprendizaje, Xiao et al. (2022) utilizaron un modelo de simulación APSIM, y entrenaron emuladores de ML que replicaron las predicciones de APSIM sobre rendimiento y dinámica de carbono del suelo con mucha menos información y costo computacional. Los emuladores les permitieron explorar rápidamente miles de escenarios de manejo de riego y determinar combinaciones óptimas para maximizar rendimientos. Este enfoque demuestra como los metamodelos pueden ampliar la aplicabilidad en modelos de cultivos.

Avances recientes, interpretabilidad y sostenibilidad. En los últimos años la atención se ha centrado hacia la explicabilidad (XAI), integración de nuevas tecnologías y sostenibilidad. El trabajo de Mohan et al. (2025) utiliza XAI para interpretar los factores que más influyen en los modelos de regresión basados en árboles, demostrando que la temperatura y la precipitación son los principales determinantes del rendimiento de un cultivo.

Paralelamente, las revisiones de documentación indican que en una base de 125 artículos de rendimiento entre 2016 y 2025, la mayoría se enfocan en DL (42 trabajos) y ML (36),

mientras que 24 se enfocan en metamodelos. Esta revisión enfatiza que las variables más influyentes son: variables meteorológicas, propiedades del suelo e índices vegetativos de satélite.

Una de las conclusiones más importantes es que los modelos boosting muestran un mejor equilibrio entre precisión, estabilidad y capacidad para manejar datos heterogéneos. Esto es útil para regiones donde la producción agrícola presenta grandes variaciones entre años, como ocurre en las provincias de la Costa ecuatoriana en periodos afectados por fenómenos externos como el terremoto de 2016 o la pandemia COVID-19.

Turpo Mamani (2024) revisa 15 estudios y concluye que el uso combinado de modelos predictivos, sensores IoT y algoritmos de optimización son esenciales para la gestión de recursos y acceso a información frente a cambios climáticos. El artículo destaca tecnologías emergentes como blockchain o big data para mejorar la resiliencia agrícola.

Brechas de estudio. A pesar de estos avances, persisten brechas significativas:

- Falta de estudios en cultivos tropicales perennes: la mayoría de las investigaciones se centran en cultivos como arroz, trigo y papa. Hay escasez de modelos aplicados a cultivos de la región andino-amazónica como el banano.
- Escala regional y contexto local: pocos trabajos utilizan datos de América Latina o integran estadísticas oficiales con productos climáticos de reanálisis. Existe un vacío en estudios que combinen datos de estaciones locales, bases satelitales (NASA POWER) y registros de rendimientos (ESPAC/INEC) para el banano en la Costa Ecuatoriana.
- Exploración de escenarios de cambio climático: aunque estudios como el de Tamayo-Vera et al. (2024) utilizan escenarios SSP para la papa, no existen proyecciones equivalentes para el banano en la región costa ecuatoriana.

- Exploración y análisis de escenarios socioeconómicos históricos que hayan

afectado al rendimiento del banano en el periodo (2012-2024).

Estas brechas justifican el presente proyecto: aplicar y comparar modelos de ML, integrando datos climáticos regionales y rendimientos de banano, bajo diferentes escenarios de cambio climático. Además, de incorporar técnicas de explicabilidad para que agricultores y técnicos puedan comprender la influencia de cada variable y resultados obtenidos para la toma de decisiones.

En la **Tabla 1** se resumen los principales artículos consolidados, mostrando de forma sintética el título, autores, cultivos, modelos utilizados y su aporte principal.

Tabla 1

Comparación de artículos científicos

Título del artículo	Autores (año)	Cultivo y Región	Método/Algoritmo	Hallazgos
Machine learning methods for crop yield prediction and climate change impact assessment in agriculture	Crane-Droesch A. (2018)	Multicultivo	SNN, DNN, regresión, ML	Propone modelos semiparamétricos y regresiones para mejorar la predicción de rendimiento
Machine Learning Algorithms for Modelling Agro-Climatic Indices for Crop Yield Forecasting, Monitoring and Management	Satapathy S.C. et al. (2021)	Multicultivo	Regresión, ML	Introduce índices agroclimáticos como variables de entrada para ML
Recent Trends in Machine Learning, Deep Learning, Ensemble Learning, and Explainable Artificial Intelligence Techniques for Evaluating Crop Yields Under Abnormal Climate Condition	Múltiples autores (2024)	Multicultivo	ML, DL, XAI	Revisión amplia que sintetiza tendencias de ML, DL y XAI en la predicción de rendimiento
Next-gen agriculture: integrating AI and XAI for precision crop yield prediction	Mohan R.N.V.J. et al. (2025)	Multicultivo	Decision Tree, Random Forest, LightGBM + XAI	Demuestra que combinar modelos avanzados de regresión con técnicas XAI (SHAP, LIME) mejora la precisión ($R^2 \approx 0,92$)
AI-Driven Predictive Analytics in Precision Agriculture	Diversos autores (2023)	Multicultivo / Agricultura de precisión	ML + DL	Presenta flujos de analítica predictiva integrando sensores y datos climáticos

Predictive Analysis and Optimization in Sustainable Agriculture Facing Climate Change with Emerging Technological Approaches	Diversos autores (2023)	Multicultivo	ML, optimización	Revisa tecnologías emergentes aplicadas a la optimización de sistemas agrícolas
Adaptive neuro-fuzzy inference system based multicrop yield prediction in the semi-arid region of India	Borse K. et al. (2025)	Arroz, maíz, maní, caña	ANFIS, ANN y árboles	Muestra que ANFIS supera a ANN en precisión y permite integrar múltiples variables climáticas
Machine Learning-Based Predictive Assessments of Impacts of Influential Climatic Conditions for the Sustainable Productivity of Paddy Crops	Thirumal S., Latha R. (2023)	Paddy (India)	mLSTM + DSSAT	Combina mLSTM con DSSAT para evaluar paddy bajo RCP4.5 y RCP8.5, identificando variables y periodos críticos.
Feasibility of Machine Learning-Based Rice Yield Prediction in India at the District Level Using Climate Reanalysis Data	Diversos autores (2023)	Arroz (India)	Random Forest, SVM, ANN, regresión	Demuestra la viabilidad de predecir el rendimiento de arroz a nivel distrital con datos climáticos de reanálisis y modelos ML.
Forecasting the Impact of Climate Change on Rice Crop Yields under RCP4.5 and RCP8.5 in Central Luzon, Philippines, Using Machine Learning Algorithms	Baltazar R.G. (2024)	Arroz (Filipinas)	Random Forest, Gradient Boosting, ANN, regresión	Integra escenarios RCP4.5/8.5 con ML para proyectar rendimientos de arroz y cuantificar pérdidas climáticas.
How can machine learning help in understanding the impact of climate change on crop yields?	Sidhu B.S. et al. (2022)	Arroz, trigo, mijo perla	Boosted Regression Trees, regresión	Compara BRT con regresiones lineales y muestra mejoras en precisión pero advierte sobre la necesidad de cautela para inferencia causal.
Projecting future changes in potato yield using machine learning techniques: a case study for Prince Edward Island, Canada	Tamayo-Vera D. et al. (2024)	Papa (Canadá)	Ensembles, series temporales, OMP	Predice rendimientos de papa usando series temporales y proyecta fuertes caídas bajo escenarios SSP1/2/5-8.5, evidenciando sensibilidad al clima.
Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability	Diversos autores (2023)	Papa	Random Forest, ensambles, análisis de riesgo	Cuantifica el riesgo de pérdidas de rendimiento de papa bajo escenarios climáticos mediante ML y análisis probabilístico.
Climate-driven projections of sunflower yield across the United States	Majumder S., Mason C.M. (2025)	Girasol (EE. UU.)	Árboles + XAI	Identifica temperatura máxima y precipitación estival como variables

using machine-learning and explainable artificial intelligence				críticas y proyecta descensos de rendimiento; utiliza XAI para explicar umbrales.
Machine Learning Integrated Climate-Agriculture Forecasting: A Transformer-Based Approach to Predict Precipitation and Wheat Production Amidst Climate Change	Diversos autores (2023)	Trigo	Transformers, DL	Integra modelos tipo transformer para predecir precipitación y rendimiento de trigo, demostrando capacidad para procesar múltiples series climáticas.
A Eurasian Case Study on Climate Change and Agricultural Land Suitability Using Interpretable Machine Learning	Ramesh B. et al. (2025)	Aptitud de tierras (Eurasia)	Clasificadores ML + XAI	Usa datos CMIP6 y modelos interpretables para evaluar la aptitud de tierras bajo SSP1-2.6/2-4.5/5-8.5 y señala regiones vulnerables.
Mitigating climate change impacts on agriculture through AI-driven crop improvement	Diversos autores (2023)	Mejoramiento genético	ML/DL	Revisa el uso de IA en mejoramiento de cultivos y su potencial para incrementar tolerancia al estrés climático.
Optimized Deep Learning Methods for Crop Yield Prediction	Diversos autores (2022)	Multicultivo	Redes neuronales profundas optimizadas	Presenta arquitecturas de DL optimizadas para mejorar la predicción de rendimiento con grandes volúmenes de datos.
Optimizing Agricultural Yields with Artificial Intelligence	Diversos autores (2024)	Multicultivo	ML/DL + optimización	Explora la integración de modelos predictivos con esquemas de optimización de manejo para maximizar rendimientos.
Coupling agricultural system models with machine learning to facilitate regional predictions of management practices and crop production	Xiao L. et al. (2022)	Maíz (China)	APSIM + ML (emuladores)	Acopla modelos de procesos (APSIM) con ML, mostrando que los emuladores reproducen con precisión las predicciones.

Nota. Elaboración propia a partir de la revisión de literatura científica.

Marco Teórico

Bases teóricas de la producción de banano. El banano constituye un eje transversal en la economía ecuatoriana. Ecuador se ha consolidado como uno de los mayores exportadores de dicho cultivo, siendo una fuente económica vital para el país. La producción se concentra mayormente en la región Costa, donde las condiciones climáticas son óptimas para su desarrollo. El rendimiento de banano es altamente sensible a condiciones ambientales, se requiere rangos específicos de temperatura, precipitación y radiación solar para optimizar su cultivo. Entre las variables más influyentes tenemos:

- **Temperatura:** Según la literatura, una de las variables mas influyentes para su producción. Variaciones de temperatura pueden retrasar la maduración del cultivo.
- **Precipitación:** Determina la disponibilidad de agua. Excesos o déficits pueden afectar directamente el rendimiento del cultivo.
- **Radiación solar:** Es la energía que permite la fotosíntesis. Sin la suficiente radiación, el cultivo no se produce adecuadamente.
- **Humedad:** Ayuda a regular el intercambio de agua y nutrientes.

Datos y fuentes de información agrícola. El estudio emplea datos oficiales provenientes de la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC). La elección de una periodicidad anual (2012-2024) obedece a criterios metodológicos que buscan:

- Observar tendencias estructurales a largo plazo
- Reducir el ruido o la estacionalidad de corto plazo propia de los ciclos de cosecha mensuales
- Alinearse con la disponibilidad de datos oficiales del país

Ante la falta de información nacional específica sobre las condiciones meteorológicas terrestres en ciertas zonas rurales, la teledetección se presente como una solución robusta. Los datos meteorológicos históricos son adquiridos de NASA POWER (Prediction of Worldwide Energy Resources). Fuente que proporciona registros continuos de variables agroclimáticas como temperatura, precipitación, humedad, entre otras, permitiendo modelar patrones en zonas de interés.

Aprendizaje supervisado y algoritmos de regresión. El aprendizaje automático supervisado consiste en construir un modelo a partir de datos de entrada y salida que permiten predecir valores de variables continuas o discretas. Los algoritmos de RL fueron los primeros en emplearse para analizar relaciones entre clima y rendimiento, pero su limitante es la necesidad de asumir una forma funcional específica o lineal. Para evitar esta limitante, se adoptan métodos más flexibles como los árboles de decisión, bosques aleatorios o métodos de boosting. Estos modelos permiten generar estructuras que segmenten los atributos de manera recursiva, permitiendo identificar interacciones más complejas y no lineales.

Algoritmos de predicción relevantes. Basado en el estado del arte, los algoritmos más eficientes para el rendimiento de cultivos son:

- **Árboles de decisión y random forest:** Métodos de ensamblaje que construyen múltiples arboles de decisión durante su entrenamiento. Son muy robustos frente el overfitting y manejan bien datos no lineales como lo son los datos climáticos.
- **Gradient Boosting (XGBoost, LightGBM, CatBoost):** Son modelos secuenciales, donde cada nuevo modelo busca corregir los errores del anterior. Tienen un alto desempeño en estudios agrícolas debido a su equilibrio entre precisión y velocidad.

- **Support Vector Regression (SVR):** Modelos útiles para conjuntos de datos más pequeños, que buscan un hiperplano que mejor se ajuste a los datos de entrada.

Análisis de predicción, validación y métricas. Un elemento clave para el aprendizaje automático es la validación cruzada, utilizada para evaluar el desempeño de los modelos. Las métricas como error cuadrático medio (MSE), error absoluto medio (MAE) y coeficientes de determinación (R^2) son utilizados para cuantificar la precisión de los modelos de predicción.

Análisis de escenarios y modelado climático. Para evaluar el impacto del cambio climático en la producción agrícola se utilizan escenarios climáticos generados por modelos de circulación general (GCM) y resumidos en trayectorias de concentraciones representativas (RCP) y sus versiones socioeconómicas (SSP). Estos escenarios representan posibles variaciones climáticas hasta finales de siglo. Por ejemplo, como se trabajó en el estudio de Tamayo-Vera et al. (2024), se proyectó rendimientos bajo escenarios SSP1 (bajas emisiones), SSP2 (intermedio) y SSP5-8.2 (altas emisiones). El uso de distintos escenarios permite cuantificar la incertidumbre asociada a las proyecciones.

Anomalías y eventos atípicos. Un modelo predictivo robusto debe considerar no solo variables climáticas, sino también situaciones que alteren la capacidad productiva. Es importante contemplar la influencia de eventos atípicos en los periodos de estudio:

- Eventos geofísicos (Terremoto 2016): desastres naturales que afectan la infraestructura logística y productiva de zonas de producción.
- Crisis sanitarias globales (COVID-19): Factores que impactan la producción debido a interrupciones de la cadena de suministros, mano de obra y restricciones de logística.
- Transición y calentamiento (El niño 2023): Cambios abruptos que generan estrés térmico y un aumento en la nubosidad y precipitaciones. El exceso de humedad incrementa

exponencialmente la presión de enfermedades fúngicas, las cuales impactan el rendimiento final del cultivo.

Aplicación al caso del banano ecuatoriano. El banano es un cultivo fundamental para la economía ecuatoriana y requiere de un análisis específico de sus factores de producción y respuesta al clima. El trabajo de investigación se apoyará en los modelos ML para desarrollar predicciones de rendimiento de banano a escala provincial. Se consideran variables climáticas (temperatura máxima/mínima, precipitación, humedad relativa). Además, se utilizarán escenarios de validación para evaluar el rendimiento del banano.

Capítulo 3

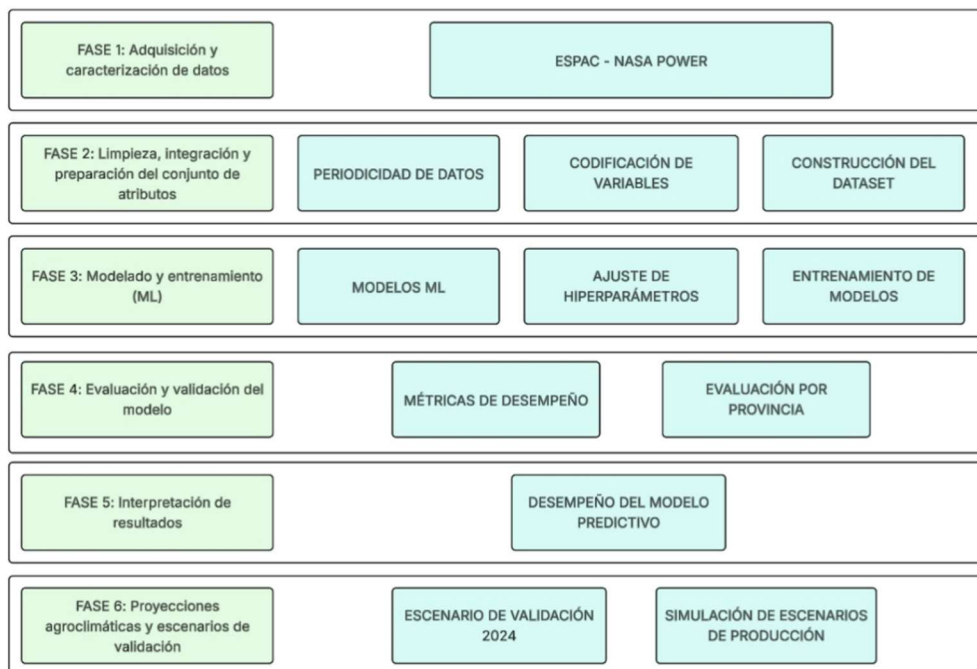
Desarrollo

Metodología

El estudio adopta un enfoque cuantitativo y correlacional, estructurado bajo un flujo de trabajo secuencial (pipeline) orientado a la construcción de modelos predictivos de rendimiento agrícola. Cada fase está diseñada para asegurar coherencia temporal, integridad del conjunto de datos y entrenamiento adecuado de los modelos. En la **Figura 1** se presenta un diagrama de flujo general de la metodología, donde se visualizan las etapas principales del proceso.

Figura 1

Descripción del proceso metodológico



Nota. Elaboración propia

El proceso se articula en las siguientes fases:

Fase 1: Adquisición y caracterización del conjunto de datos. Esta etapa establece la base empírica del estudio mediante la extracción de datos de fuentes primarias y oficiales.

- **Componente Agronómico:** Se recopilan las estadísticas anuales y superficie (plantada y cosechada) del cultivo a partir de la Encuesta de Superficie y Producción Agropecuaria Continua (Instituto Nacional de Estadística y Censos, s.f.).
- **Componente Climático:** Se obtienen series temporales de variables meteorológicas mediante API de NASA POWER (National Aeronautics and Space Administration [NASA, s.f.]), asegurando cobertura espacial continua sobre las zonas de estudio.

Fase 2: Limpieza, integración y preparación del conjunto de atributos. Esta etapa corresponde al procesamiento de la base de datos, donde se incluye:

- Identificación de valores atípicos, inconsistentes y diferencias de formato entre fuentes.
- Estandarización de unidades, tipos de datos y nomenclaturas.
- Evaluación de anomalías en la serie temporal de producción para identificación de valores atípicos que no correspondan a variabilidad climática.
- Integración de valores anuales a mensuales utilizando método proporcional basado en los meses productivos del cultivo.

Fase 3: Modelado y entrenamiento del sistema predictivo. En esta fase se implementan los modelos de aprendizaje automático orientados a estimar el rendimiento equivalente del cultivo.

Selección de algoritmos:

- Selección de modelos robustos y no lineales ampliamente reconocidos y ejecutados por su precisión en estudios agroclimáticos.

División del dataset:

- División del conjunto de datos en subconjuntos de entrenamiento y prueba.
- Validación cruzada para evitar sobreajuste y evaluar estabilidad del modelo.

Entrenamiento del modelo:

- Ajuste de hiperparámetros.
- Entrenamiento de múltiples configuraciones para comparación y selección de alternativa con mejor desempeño.

Fase 4: Evaluación y validación del modelo. Una vez entrenados los modelos, se evalúan mediante métricas estadísticas estándar.

Métricas utilizadas:

- Error cuadrático medio (RMSE).
- Coeficiente de determinación (R^2).

Proceso de validación:

- Evaluación de modelos sobre el conjunto de prueba.
- Comparación de resultados para selección de mejor modelo.

Fase 5: Interpretación de resultados y análisis de importancia de variables. Una vez seleccionado el modelo, se realizará:

- Evaluación de la contribución de las variables climáticas.
- Análisis técnico de la relación entre clima y productividad agrícola.

Fase 6: Proyecciones agroclimáticas y escenarios. El modelo final será utilizado para:

- Generar predicciones bajo condiciones climáticas históricas.
- Simular variaciones climáticas y evaluar posibles impactos en la producción.
- Explorar escenarios agroclimáticos orientados a toma de decisiones.

Síntesis final, conclusiones y recomendaciones. Se consolidarán los resultados obtenidos en el estudio y se formularán recomendaciones técnicas para investigaciones futuras y aplicaciones en gestión agrícola.

Fuentes primarias de información

Instituto Nacional de Estadística y Censos (INEC) y la ESPAC. El Instituto Nacional de Estadística y Censos (INEC) es el órgano oficial encargado de levantar y difundir estadísticas económicas, demográficas y agropecuarias del Ecuador. Correspondiente el área agrícola, la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC) se considera la principal fuente de información oficial del sector a nivel nacional. La ESPAC sigue un marco de muestreo diseñado conforme a recomendaciones de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO); en este marco se recopila anualmente información como superficie sembrada y cosechada, producción y ventas de 52 productos agrícolas, además de información sobre ganado, producción regional y empleo agrícola.

La relevancia de la ESPAC radica en que es una fuente censal anual que garantiza consistencia y representatividad estadística de la producción agropecuaria del país. Sin embargo, es importante recalcar que la información proporcionada en los tabulados de la ESPAC mantiene un formato de informe consistente a partir del año 2018, en años previos los informes contienen menos información.

Datos climáticos NASA POWER. Los datos climáticos se obtuvieron de NASA POWER (Prediction of Worldwide Energy Resources), un servicio brindado por la NASA que provee de parámetros meteorológicos y de radiación derivados de observaciones satelitales y de modelos de reanálisis. Según la página oficial, NASA POWER ofrece “conjuntos de datos de

libre acceso, de baja latencia y alta precisión”, permiten una libre elección de variables, unidades y formatos personalizables. Estos datos están orientados a usuarios del sector energético, de infraestructura y agro climatología.

Los datos meteorológicos provienen del modelo de asimilación atmosférica MERRA-2 del Goddard Space Flight Center con una resolución de $0,5^\circ \times 0,625^\circ$. Los datos son procesados diariamente y su disponibilidad es casi en tiempo real. La combinación de mediciones y modelos permite obtener estimaciones de datos continuas de variables como temperatura, precipitación, humedad relativa, radiación solar y velocidad del viento, las cuales resultan de gran utilidad para modelos de rendimiento agrícola.

Obtención de datos

Extracción de datos de la ESPAC. La ESPAC publica sus informes de resultados anuales en archivos Excel titulados “Tabulados de la Encuesta de Superficie y Producción Agropecuaria Continua”. A partir del año 2018 los tabulados presentan una estructura homogénea, con múltiples hojas (pestañas) de cada cultivo, presentando una tabla consolidada con variables de superficie, producción y ventas del cultivo, de manera regional y provincial.

Para el periodo 2018-2024, se desarrolló un script en Python que automatiza la extracción de la información relevante. El algoritmo navega por las pestañas de interés, identificando la fila y columna correspondiente a la información del banano y seleccionando valores de variables como:

- Superficie plantada (Has.)
- Superficie cosechada (Has.)
- Producción (Tm.)
- Ventas (Tm.)

Únicamente se extrajo los registros anuales pertenecientes a las provincias de la región Costa (El Oro, Esmeraldas, Guayas, Los Ríos, Santa Elena y Manabí), ya que son las zonas productoras evaluadas en este proyecto. Los datos fueron consolidados en un único archivo, facilitando su unión posterior con la información climática.

Para los años 2012-2017 la ESPAC no proporciona tabulados en formato Excel sino reportes en formato PDF, con tablas incrustadas en un único documento. A diferencia de los tabulados del periodo 2018-2024, estos informes no cuentan con información consistente de diversos cultivos, únicamente se centran en los más representativos a nivel nacional: banano, palma africana, caña de azúcar. Dada la estructura de estos archivos y la ausencia de un formato estructurado, se optó por extraer manualmente las cifras de producción, superficie plantada y cosechada correspondientes a cada provincia de interés. Estos valores se digitalizaron y estandarizaron para mantener una consistencia histórica.

Adquisición de datos NASA POWER. La adquisición de variables climáticas se realizó mediante la API de NASA POWER, la cual permite definir coordenadas geográficas, periodos de interés, parámetros a recuperar y formatos de exportación de datos. Para representar las condiciones climáticas de cada zona, se seleccionó un punto geográfico central por provincia. En la **Tabla 2** se resumen las coordenadas empleadas:

Tabla 2

Coordenadas geográficas utilizadas por provincia para la obtención de variables climáticas mediante la API NASA POWER

Provincia (Costa)	Latitud	Longitud
El Oro	-3.5278	-79.8383
Esmeraldas	0.8718	-79.4098
Guayas	-2.0391	-80.0433

Santa Elena	-2.1609	-80.5874
Manabí	-0.6415	-80.0671
Los Ríos	-1.4891	-79.6478

Nota. Las coordenadas geográficas de los puntos de muestreo fueron seleccionadas y extraídas mediante la plataforma NASA POWER.

La selección de variables climáticas se fundamentó en la literatura científica reciente, la cual identifica los factores que más influyen en el desarrollo de cultivos tropicales. Las variables extraídas son:

- Temperatura máxima (°C)
- Temperatura mínima (°C)
- Temperatura media (°C)
- Precipitación (mm)
- Presión superficial (kPa)
- Humedad del suelo (adimensional)
- Humedad capa superficial (adimensional)
- Humedad específica (g/kg)
- Radiación (kW/m²)
- Velocidad del viento (m/s)

NASA POWER permite obtener datos en frecuencias diarias, mensuales o anuales. En este caso se empleó la frecuencia mensual, ya que:

- Facilita la integración con datos productivos anuales.
- Reduce la presencia de ruido y variabilidad.
- Es compatible con modelos ML que requieren series suavizadas.

Integración de bases productivas y climáticas. La integración de datos requirió un proceso de unificación temporal y estructural, con la finalidad de generar un conjunto de datos consistente. Los registros de producción del INEC están disponibles únicamente en formato anual, mientras que las variables climáticas extraídas de NASA POWER tienen una resolución mensual. Para compatibilizar ambas fuentes se aplicó una metodología de armonización temporal. En el caso del banano, cuya cosecha es continua durante todo el año, la producción se dividió de manera uniforme entre los doce meses. Esta aproximación permite mantener el valor total reportado por el INEC y evita introducir variaciones estacionales que no correspondan al comportamiento real del cultivo. Metodologías similares se utilizan en estudios agroclimáticos cuando se deben combinar datos con diferentes granularidades temporales y es necesario mantener coherencia en las series.

El proceso de integración permitió construir una matriz de análisis donde cada registro mensual combina las condiciones climáticas del periodo con el rendimiento equivalente mensual del cultivo, facilitando el uso de técnicas de aprendizaje automático que requieren series homogéneas y alineadas temporalmente. Como resultado, se obtuvo un registro con 780 filas y 19 columnas, dando un total de 14820 datos en el dataset final. Es importante recalcar que el algoritmo de extracción de datos nos permitió adquirir la información correspondiente a banano, palma africana y caña de azúcar (cultivos más representativos de la costa ecuatoriana), dando un total de 27588 datos, útiles para futuros trabajos o investigaciones.

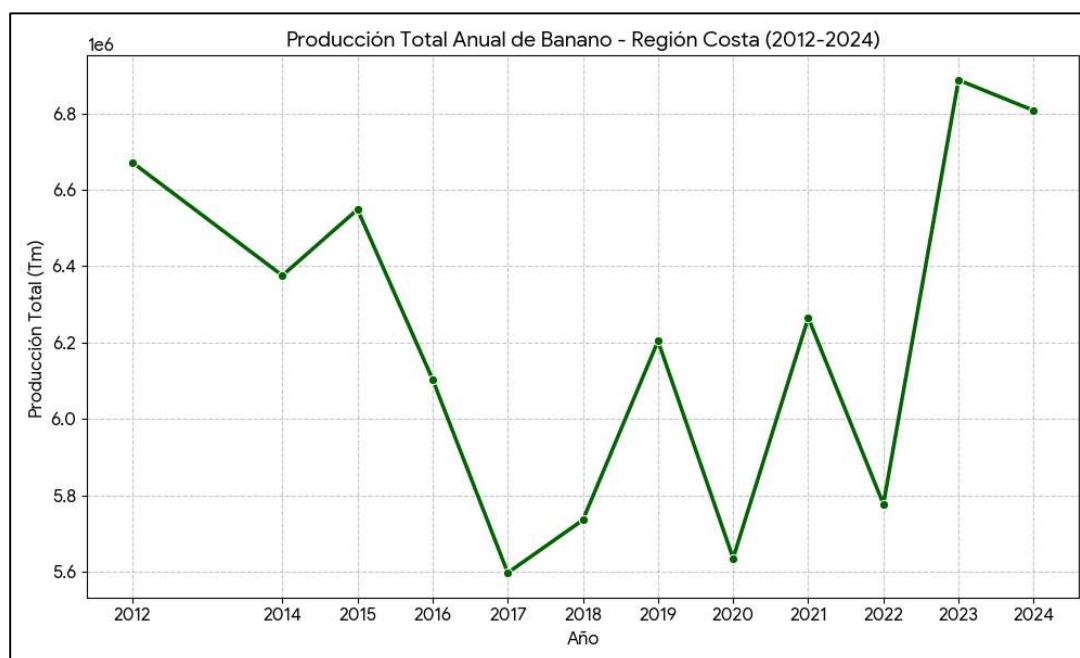
Depuración, análisis preliminar y tratamiento de valores atípicos

Análisis preliminar de la serie de producción. Como parte del proceso de depuración se realizó un análisis exploratorio inicial utilizando la serie anual de producción total del banano

para el periodo 2012-2024. En la **Figura 2** se muestra la producción de la región costa, la cual constituye la gran mayoría de la producción nacional de banano en Ecuador.

Figura 2

Producción total anual de banano en la región Costa del Ecuador durante el periodo 2012–2024



Nota. Elaboración propia con datos provenientes de la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC).

El comportamiento histórico evidencia una dinámica no lineal con fluctuaciones marcadas por eventos atípicos. En términos generales, el volumen productivo se ha mantenido en un rango entre los 5.0 y 6.9 millones de toneladas métricas anuales. Sin embargo, la serie temporal revela un punto de inflexión crítica en el periodo 2016-2017, donde se registra la caída más severa. Tras este evento, se observa una tendencia de recuperación sostenida que experimenta variaciones durante los años 2020 y 2022. Para el cierre del periodo analizado (2023-2024), la producción consolidada de la Costa muestra una estabilización al alza, superando

los 6.8 millones de toneladas, lo que muestra la capacidad de resiliencia del sector frente a anomalías registradas en la última década.

Anomalías y eventos estructurales en la serie temporal anual (2012-2024). Para el desarrollo de un modelo predictivo robusto es necesario identificar y caracterizar eventos atípicos que generan ruido en los datos históricos. Durante el periodo de estudio, la producción de banano no solo respondió a la variabilidad climática estacional, sino que fue impactada por situaciones atípicas de diversa naturaleza. Dichas caídas de producción se identificaron como valores atípicos dentro de la serie anual.

Aunque en menor magnitud que afectaciones posteriores, el año 2014 registró una leve disminución del volumen de producción, descendiendo a 6.38 millones de toneladas frente a los 6.67 millones reportados en 2012. Esta variación negativa representa una fluctuación cíclica propia de la dinámica de mercado y variabilidad climática estándar, la cual fue revertida rápidamente con la recuperación productiva evidenciada en 2015

El periodo 2016-2017 constituye el punto de inflexión más complejo de la serie temporal, caracterizado por la combinación de eventos climáticos, económicos y geofísicos.

Climáticamente, el ciclo 2015-2016 estuvo influenciado por el fenómeno de El Niño, cuyas lluvias torrenciales en la región Costa generaron preocupaciones por estrés hídrico y proliferación de enfermedades en cultivos. Las proyecciones agronómicas de la época estimaban que el exceso de humedad podría reducir la producción bananera entre un 20% y 30% (Swissinfo.ch, 2016, párr. 16). Paralelamente, los principales mercados de exportación presentaban frenos, destacándose la devaluación del rublo en Rusia y su reducción de poder de compra (Swissinfo.ch, 2016). En este contexto, el terremoto del 16 de abril actuó como catalizador de crisis. El sismo provocó una ruptura severa en la actividad agrícola costera,

combinado con choques de oferta y demanda. Las consecuencias económicas fueron devastadoras, con pérdidas estimadas de USD 525 millones y destrucción de aproximadamente 21.823 empleos (Banco Mundial, 2021, párr. 4), afectando directamente ingresos de exportación agrícola. Los efectos fueron particularmente en las provincias de Manabí y Esmeraldas, territorios donde más del 60% de la población depende económicamente de actividades agropecuarias (Pacheco Gil, 2017). Evaluaciones post-desastre indican que el 80% de los productores rurales en Manabí sufrieron daños, afectando a cultivos clave como cacao (24%), palma africana (23%), maíz (17%) y plátano (7%) (Pacheco Gil, 2017). Si bien estos eventos ocurrieron en el año 2016, los datos de producción relevan que la caída más profunda se materializó en 2017, comportamiento que responde al efecto de rezago agronómico.

A diferencia de la ruptura estructural causada por el terremoto de 2016, el año 2020 presentó un escenario de estabilidad climática que contrasta con una pandemia global. Los registros meteorológicos indicaron que las condiciones se mantuvieron dentro de patrones estacionales habituales, y a pesar de que hubo reportes de inundaciones, no existieron riesgos de alto impacto que justificaran agronómicamente una caída del rendimiento del banano (Pesantes, 2020). La pandemia COVID-19 provocó una recesión global que afectó cadenas de suministro y demanda de productos. Sin embargo, es importante mencionar que la agricultura costera se fortaleció en exportaciones de Banano, a pesar de tener una caída del 6% con respecto al 2019. El banano, siendo el principal producto no petróleo de país, alcanzó un récord histórico de exportación: en total se embarcaron 380 millones de cajas de banano, 7% en volumen y 15% en valor respecto al 2019 (Cámara Marítima del Ecuador, 2021, párr. 2).

El caso del 2020 en la costa ecuatoriana ilustra perfectamente la complejidad multidimensional de la data agrícola. Mientras que las variables meteorológicas y operativas

(como las restricciones por pandemia) pueden explicar una caída del 6% en el volumen físico cosechado, las variables comerciales responden a dinámicas globales de demanda y logística que pueden moverse en dirección opuesta, generando récords de exportación.

Esto demuestra que el análisis agrícola no puede ser lineal; requiere un análisis que cruce dos escenarios distintos:

- Agronómico: Donde variables meteorológicas, condiciones de suelo y estacionalidad dictan el rendimiento del cultivo.
- Económico: Donde el stock, precio y demanda internacional dictan el éxito financiero.

El año 2022 representa un caso de estudio caracterizado por anomalías meteorológicas de larga duración y una alta inestabilidad sociopolítica. Este año estuvo marcado por el prolongado fenómeno climático de La Niña (2020-2023), el más largo en 66 años, el cual generó anomalías climáticas en zonas costeras. La Niña ocasiono un déficit hídrico y un descenso de temperatura, creando escenarios opuestos a los requerimientos fisiológicos del banano para un crecimiento óptimo. El impacto fue determinante, según la Asociación de Comercialización y Exportación de Banano (ACORBANEC), cerca del 80% de la caída del volumen de exportaciones bananeras de 2022 se debió al impacto de La Niña (González, 2023, párr. 15). A la presión climática se sumaron factores internacionales, el estallido del conflicto bélico entre Rusia y Ucrania alteró uno de los mercados más importantes para la fruta ecuatoriana, generando problemas logísticos y de pago. El desempeño agrícola se vio finalmente fracturado por la inestabilidad política interna. El “Paro Nacional” de junio de 2022 paralizó las actividades productivas del país durante 18 días. Según diversos medios de prensa, los bloqueos estratégicos en vías de la Costa y Sierra rompieron la cadena de suministros, impidiendo la cosecha, empaquetado y transporte de la fruta

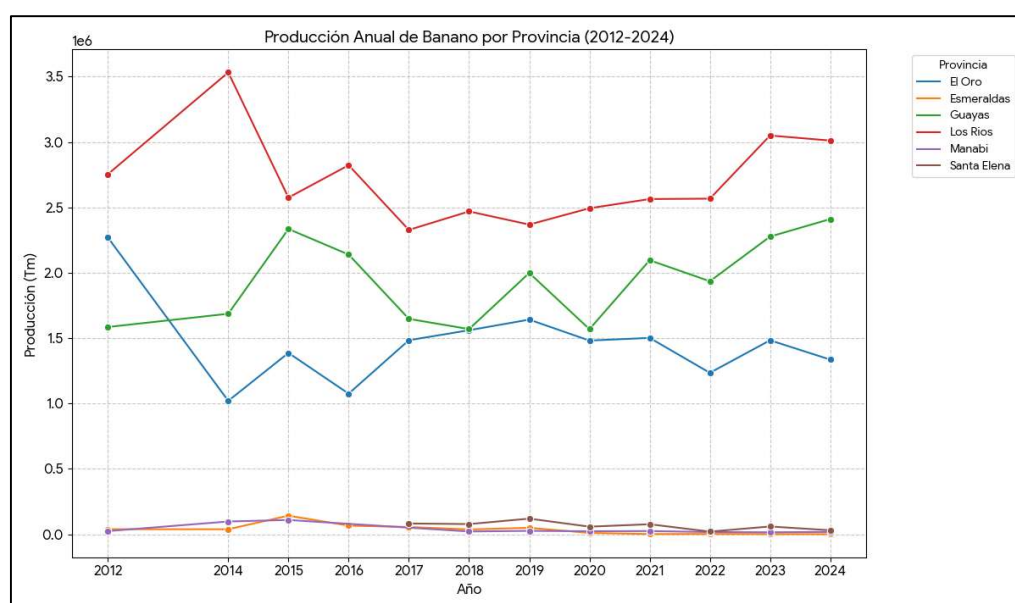
(Roa Chejín, 2022). La Asociación de Exportadores de Banano del Ecuador (AEBE) estimó pérdidas sectoriales cercanas a los USD 2 millones derivadas de esta combinación de factores: guerra, condiciones climáticas y conflictividad interna (EFEAgro, 2023).

Análisis provincial: Identificación de datos atípicos.

La **Figura 3** muestra la producción anual del banano en las distintas provincias más representativas de la región costa.

Figura 3

Producción anual de banano por provincia en el Ecuador (2012–2024)



Nota. Elaboración propia con base en los registros provinciales de producción de la ESPAC.

El análisis consolidado para el periodo 2012-2024 indica cierta resiliencia ante desastres por parte de ciertas provincias, teniendo una capacidad de recuperación tras eventos críticos, mientras que otras son más vulnerables a periodos de crisis.

Los datos permiten identificar ciertos roles:

- La provincia de Los Ríos muestra una tendencia creciente y sostenida, actuando como amortiguador productivo. Su ubicación geográfica le ha permitido mitigar impactos directos a eventos atípicos.
- La provincia de Guayas tiene variaciones más pronunciadas, particularmente en el 2020.
- Esmeraldas muestra un comportamiento estable con una anomalía evidente en 2016, caída asociada al terremoto y factores presentados anteriormente.
- Santa Elena y Manabí son las provincias con menor participación, mantienen niveles de producción reducidos, pero reflejan patrones coherentes. Sus fluctuaciones no ejercen mayor impacto sobre valores totales, pero aportan información útil para la caracterización del cultivo.

La identificación de valores atípicos en este estudio se realizó en base al conocimiento e investigación del contexto agro-productivo y socioeconómico, y no exclusivamente mediante criterios estadísticos automáticos como el rango intercuartílico (IQR) o medidas de dispersión. Esta decisión se fundamenta en que las anomalías detectadas corresponden a eventos extraordinarios ampliamente documentados (como el terremoto de 2016 y la pandemia de COVID-19 en 2020) que no reflejan la relación agroclimática natural entre variables meteorológicas y producción de banano

Selección y depuración del conjunto de datos. Para garantizar estabilidad y convergencia de los modelos predictivos, se aplica una limpieza y depuración de datos para minimizar el ruido generado por inconsistencias en la serie temporal. A partir del análisis exploratorio sobre las seis provincias costeras, se determinó acotar el alcance del modelamiento a

las tres provincias más significativas de producción de banano ecuatoriano: Los Ríos, Guayas y El Oro. Esta selección se sustenta bajo los siguientes criterios:

- **Representatividad:** En conjunto, estas tres provincias concentran el 98.2% de la producción histórica registrada en el periodo de análisis.
- **Continuidad de la serie:** Presentan registros completos en el periodo de análisis (2012-2024), a diferencia de provincias como Manabí o Santa Elena que exhiben vacíos de información o series truncadas.
- **Estabilidad volumétrica:** Se excluye a la provincia de Esmeraldas debido a su colapso estructural (caída del 90% de su producción), comportamiento que representa una anomalía irreversible y no un patrón climático ajustable.

Filtrado de datos. Con el objetivo de integrar la relación Clima-Producción, se procedió a excluir del set de entrenamiento aquellos periodos afectados por interrupciones físicas o logísticas. Es importante aclarar que la reducción de datos no busca maximizar métricas, sino validar la naturaleza estructural del sistema productivo. La configuración temporal de información se detalla en la **Tabla 3**.

Tabla 3

Configuración temporal del filtrado de datos por provincia y muestras finales

Provincia	Años seleccionados	Periodos excluidos	Justificación técnica de la exclusión	Nº de muestras finales (Meses)
Los Ríos	2012-2015, 2018-2019, 2021-2024	2016, 2017, 2020	2016: Sismo 2017: Rezago agronómico 2020: Pandemia	108

Guayas	2012-2015, 2018-2019, 2021-2024	2016, 2017, 2020	2016: Sismo 2017: Rezago agronómico 2020: Pandemia	108
El Oro	2012-2015, 2017, 2018- 2019, 2021- 2024	2016, 2020	2016: Sismo 2020: Pandemia (Se mantiene 2017 por evidenciar recuperación normal)	120

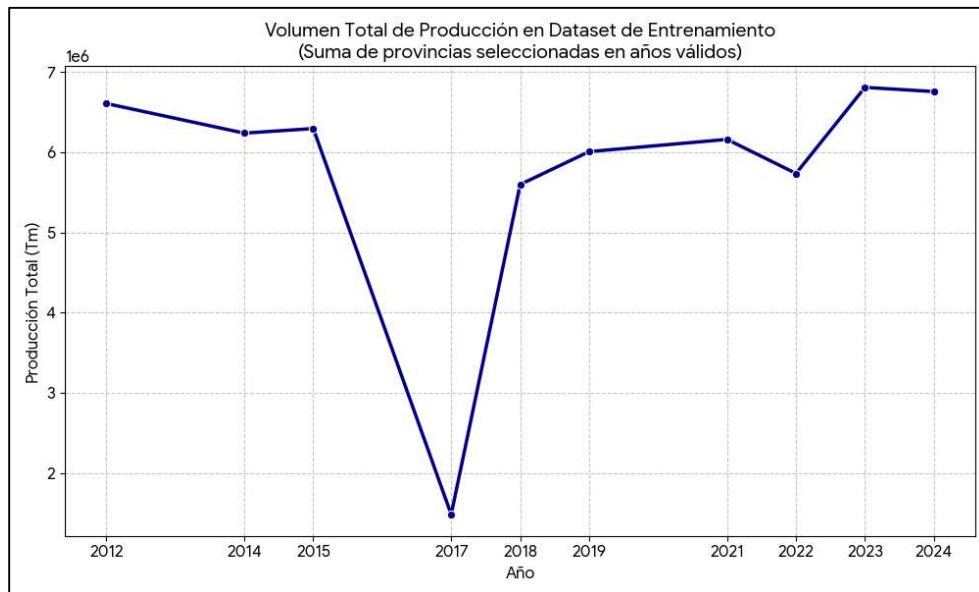
Nota. Se excluyeron años con valores atípicos atribuidos a factores no climáticos documentados, garantizando así la robustez estadística del análisis.

El dataset de entrenamiento quedó conformado por 336 registros mensuales. Este volumen de datos garantiza un balance adecuado entre cantidad de muestras necesarias para convergencia y pureza de la influencia climática.

La **Figura 4** muestra una caída pronunciada en el año 2017, por lo cual es importante aclarar que eso no representa una caída en la producción, sino una reducción de la disponibilidad de muestras válidas para el modelo durante dicho año. Esto es debido a la disponibilidad de datos y acceso a información.

Figura 4

Volumen total de producción del dataset de entrenamiento considerando años y provincias con datos válidos

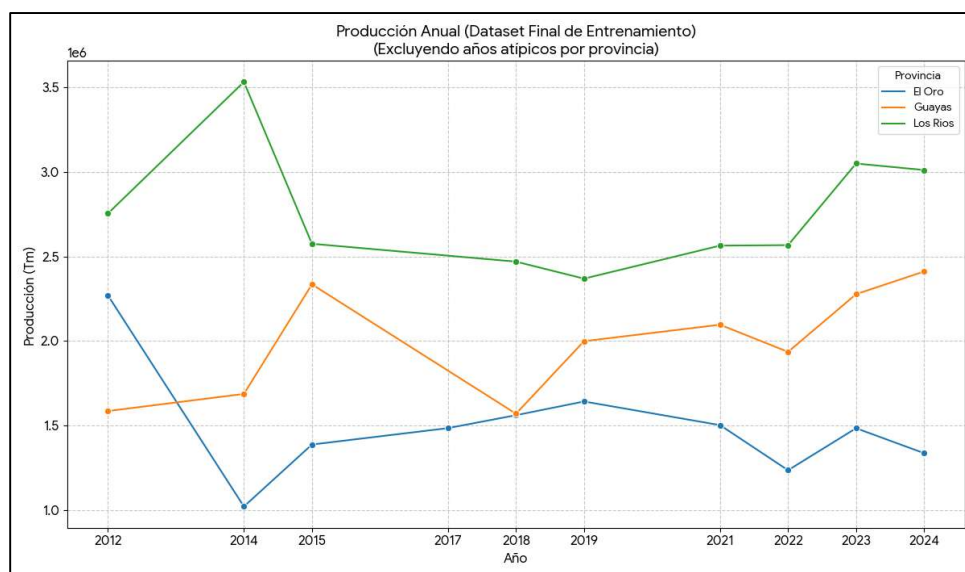


Nota. Elaboración propia

En la **Figura 5** notamos la distribución de la producción mensual en las tres principales provincias utilizando el dataset depurado. Se evidencia una clara jerarquía productiva donde Los Ríos presenta la mediana más alta, consolidándose como la zona de mayor volumen, seguida por Guayas, que exhibe un rendimiento intermedio con una dispersión considerable, similar a la primera, lo que refleja su sensibilidad a las variaciones estacionales. En el caso de El Oro se exhibe una distribución más compacta, lo que denota una producción más estable y menos propensa a oscilaciones drásticas mes a mes.

Figura 5

Producción anual por provincia en el dataset final de entrenamiento



Nota: Elaboración propia

Interpretación y análisis de importancia de variables

Estrategias de selección. En esta fase, el objetivo es no se limita a la capacidad predictiva de los modelos, sino a la interpretación agronómica de los resultados, identificando las variables climáticas que explican la variabilidad productiva en las provincias de Los Ríos, Guayas y El Oro.

Con el fin de reducir el sesgo asociado a un único método de selección, se adopta una estrategia de análisis que contraste tres perspectivas:

- **Análisis Estadístico Univariante:** Se aplica la prueba F-test para evaluar la dependencia lineal individual de cada variable climática respecto a la producción, complementando con coeficientes de correlación de Spearman.
- **Optimización de Redundancia (MRMR):** Se utiliza el algoritmo de Mínima Redundancia y Máxima Relevante, el cual selecciona el subconjunto óptimo de predictores que

maximiza la correlación con la variable objetivo mientras penaliza la colinealidad entre variables explicativas.

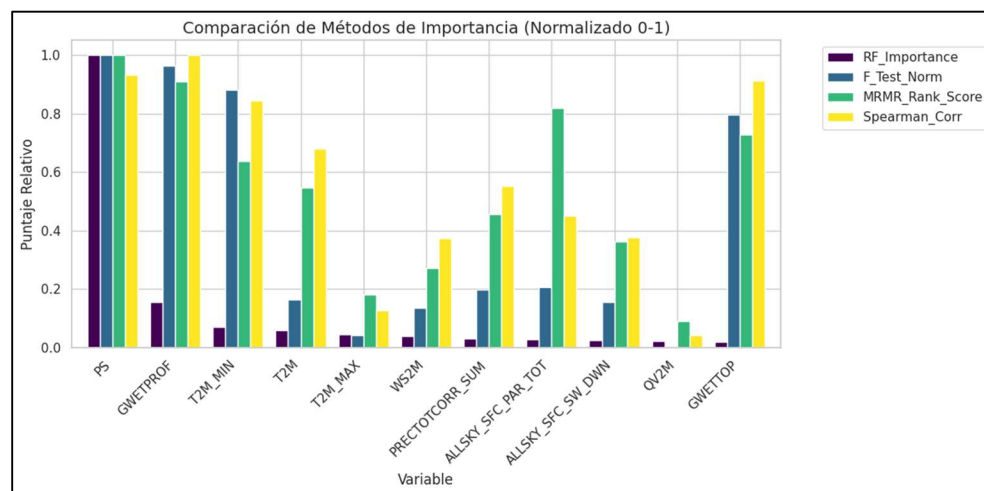
- **Importancia no lineal y Explicabilidad (SHAP):** Se examina la importancia de características basada en la reducción de impureza mediante Random Forest, y se integra el análisis de valores SHAP (Shapley Additive exPlanations), el cual permite cuantificar la contribución individual de cada variable.

Justificación del enfoque SHAP. A diferencia de métodos tradicionales, el análisis SHAP permite descomponer la predicción para entender la direccionalidad del impacto de cada variable. Esto permite identificar no únicamente la magnitud de importancia, sino también la dirección del impacto (favorecer o afectar), permitiendo discernir si valores altos de determinada variable actúan como estimulantes o limitantes para el rendimiento.

Métricas de evaluación climáticas. En la **Figura 6**, se observa la convergencia entre los distintos algoritmos evaluados. Las barras representan la relevancia de cada variable climática escalada entre 0 y 1. A pesar de las diferencias matemáticas de cada algoritmo, existe cierta similitud entre las variables dominantes. Las variables con valores cercanos a 1 indican una influencia fuerte en el rendimiento del cultivo.

Figura 6

Comparación de métodos de interés con variables normalizadas

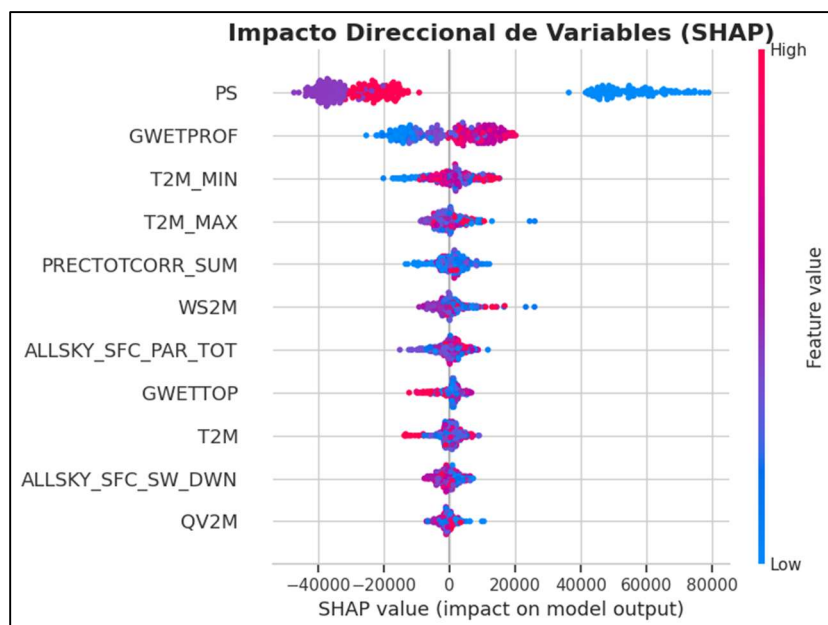


Nota. Comparación de la relevancia de las variables climáticas según los métodos Random Forest, F-Test, MRMR y correlación de Spearman, con valores normalizados entre 0 y 1.

La gráfica SHAP debe interpretarse bajo tres dimensiones:

- Jerarquía (eje Y): Las variables se ordenan de mayor a menor importancia global.
- Magnitud de impacto (eje X): La dispersión horizontal indica que tanto altera la variable la predicción del rendimiento del cultivo. Los puntos alejados del centro representan un fuerte impacto (De forma positiva a la derecha, y negativa a la izquierda).
- Relación de valor: El gradiente de color revela el tipo de relación.

La distribución de puntos evidencia la no linealidad del sistema. Esto sugiere que, para ciertas variables, los valores pueden tener un efecto desproporcionado en el rendimiento del cultivo. El análisis de la **Figura 7** revela que la PS (Presión Superficial) tiene el mayor rango de dispersión en el eje horizontal, definiéndola como la variable predominante en la varianza del modelo.

Figura 7*Impacto direccional de variables (SHAP)*

Nota. Representación del impacto y dirección de influencia de variables climáticas.

Los valores bajos de presión (puntos azules) se asocian con contribuciones positivas al rendimiento. Adicionalmente, variables como Humedad y Temperatura muestran una relación directa; valores elevados (puntos rojos) de estas variables impulsan el alza en producción.

Complementando los resultados visuales, la **Tabla 4** detalla los valores numéricos obtenidos por cada métrica de evaluación. Esta tabulación permite una comparación directa y objetiva.

Tabla 4

Importancia y correlación de las variables de entrada según los métodos Random Forest, F-Test, MRMR y Spearman

Variable	Random Forest Importancia	F_Test	MRMR Rank Score	Spearman Correlación
PS	0.670714	1.000000	11	0.467508

GWETPROF	0.103818	0.962792	10	0.501107
T2M_MIN	0.047069	0.880904	7	0.423509
T2M	0.039052	0.163075	6	0.340780
T2M_MAX	0.029974	0.042618	2	0.064018
WS2M	0.025497	0.136548	3	0.187926
PRECTOTCORR_SUM	0.020507	0.197292	5	0.276823
ALLSKY_SFC_PAR	0.019677	0.207583	9	0.225169
ALLSKY_SFC_SW	0.016319	0.154446	4	0.188561
QV2M	0.014688	0.001607	1	0.021633
GWETTOP	0.012685	0.797147	8	0.457457

Nota. Elaboración propia

Selección del vector de características. Una vez aplicados los métodos de importancia y el análisis de su influencia en el rendimiento del cultivo, se procede a realizar una depuración del conjunto de datos. Se descartan aquellas variables que presentaron una contribución cercana a cero o cierta redundancia. El objetivo es reducir el riesgo de sobreajuste y entrenar el algoritmo únicamente con señales climáticas influyentes.

El vector de características final, vinculando variables de ESPAC y climáticas, será la entrada para el entrenamiento de los modelos, conformado por las variables en la **Tabla 5**:

Tabla 5

Características climáticas seleccionadas para el modelo

Variable	Nombre Descriptivo	Tipo de dato	Justificación de selección
PS	Presión superficial	Valor continuo	Variable más influyente. Indicador principal de estacionalidad en la costa ecuatoriana.

T2M	Temperatura a 2 metros	Valor continuo	Esencial para modelar caracterizar fisiología del cultivo.
QV2M	Humedad Específica	Valor continuo	Alta correlación. Crucial para representar estrés hídrico, muy influyente para el cultivo.
PRECTOTCORR	Precipitación corregida	Valor continuo	Correlación con humedad específica.
WS2M	Velocidad del viento	Valor continuo	Define el riesgo físico por viento
T2M_MIN	Temperatura Mínima	Valor continuo	Permite al algoritmo detectar frentes fríos que pausan el crecimiento del cultivo.
GWETPROF	Humedad del perfil del suelo	Valor continuo	Disponibilidad hídrica real.
Superficie plantada	N/A	Valor continuo	Alta relación con variable objetivo
Superficie cosechada	N/A	Valor continuo	Alta relación con variable objetivo

Nota. Elaboración propia

Evaluación y validación del modelo

La evaluación y validación de modelos se realizó con el objetivo de cuantificar su capacidad de verificar la estabilidad de las estimaciones frente a datos no utilizados durante el entrenamiento. Para ello, se empleó un esquema de validación basado en separar la información temporal, utilizando la data histórica hasta el año 2023 para el entrenamiento y la data del 2024 como conjunto de prueba independiente.

El desempeño de los modelos se evaluó mediante métricas como Error Cuadrático Medio (RMSE), Error Absoluto Medio (MAE) y el Coeficiente de Determinación (R^2). Estas métricas proporcionan una evaluación complementaria del modelo, donde el RMSE penaliza errores grandes, el MAE ofrece una medida robusta de error promedio y el R^2 cuantifica la proporción de varianza.

Evaluación del modelo 1: Regresión lineal. El primer conjunto de modelos corresponde a regresión lineal, utilizados como referencia debido a su interpretabilidad e influencia proporcional de las variables productivas incluidas en el estudio. Se implementó un modelo de regresión línea simple, y variantes: Ridge, Lasso y ElasticNet.

En la **Tabla 6** se presentan los resultados obtenidos sobre el conjunto de prueba correspondiente al año 2024.

Tabla 6

Resultados de la evaluación del Modelo 1 y sus variantes sobre el conjunto de prueba

Modelo	RMSE	MAE	R^2
Regresión Lineal	13 887.19	11 551.58	0.9422
Ridge Regression	13 624.99	11 778.07	0.9443
Lasso Regression	13 831.09	11 508.22	0.9427
ElasticNet	13 585.24	11 549.05	0.9447

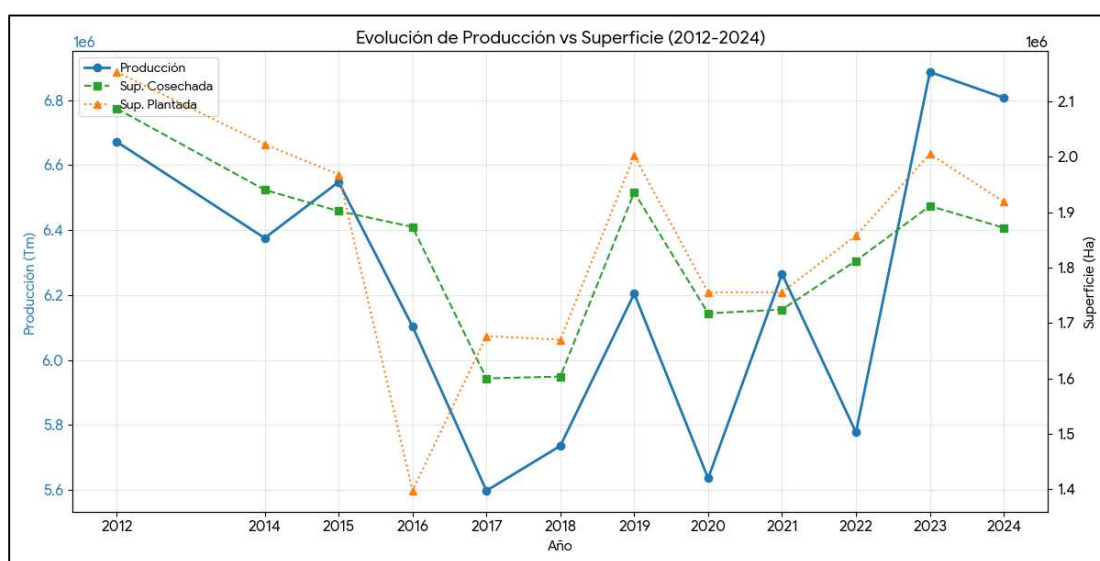
Nota. Entrenamiento de modelos realizados en Google Colab.

Los resultados muestran que las variantes de regresión lineal tienen un alto nivel de ajuste, con valores de R^2 superiores a 0.94. El modelo ElasticNet presenta el mejor desempeño global, reflejando la ventaja de combinar penalizaciones L1 y L2 en un contexto caracterizado por alta colinealidad entre predictores estructurales, como lo son la superficie plantada y cosechada.

Como se observa en la **Figura 8**, se evidencia una relación estrecha y consistente entre producción y variables estructurales del sistema productivo (dataset completo). A lo largo de la serie temporal, se observa que las variaciones en la producción siguen de manera cercana las fluctuaciones de ambas superficies, manteniendo un comportamiento proporcional incluso en escenarios de recuperación o expansión del área cultivada.

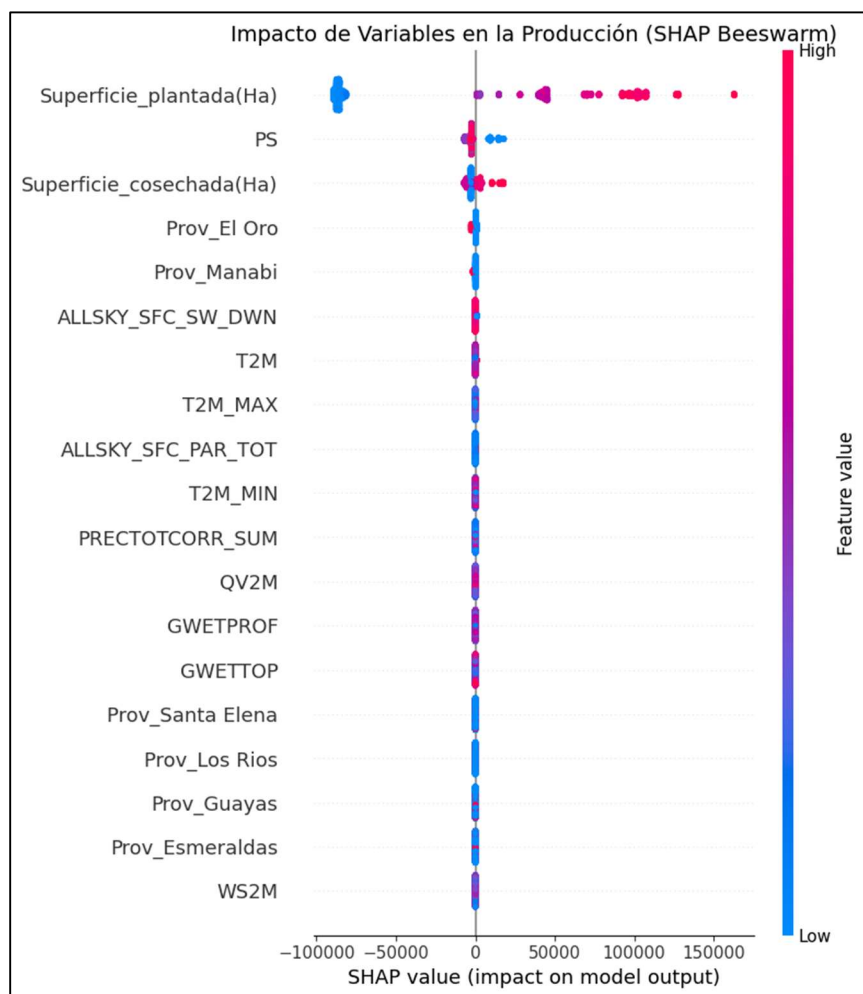
Figura 8

Evolución de la producción total en contraste con la superficie plantada y cosechada



Nota. Comparación temporal entre la producción total del banano y variables estructurales, provenientes de datos históricos de la ESPAC.

Con el fin de analizar la contribución individual de cada variable, se aplica el método de interpretabilidad SHAP con variables estructurales y climáticas. Como se muestra en la **Figura 9**, la variable de superficie plantada presenta el mayor valor de impacto, seguida de PS y superficie cosechada, confirmando su influencia dominante en la predicción del modelo.

Figura 9*Impacto de variables en la producción (SHAP Beeswarm)*

Nota. Distribución de los valores SHAP entre variables climáticas y estructural.

Evaluación del modelo 2: Support Vector Regression (SVM). El modelo de Support Vector Regression se implementó con un enfoque no lineal con el objetivo de evaluar la existencia de relaciones no complejas entre las variables explicativas y el rendimiento del banano. En una primera aproximación el modelo se desempeñó de forma deficiente, reflejando un valor de coeficiente de determinación negativo. Aplicando un proceso de optimización de hiperparámetros, mediante validación cruzada, se mejoró el desempeño del modelo alcanzando

un R^2 superior al 0.90 en el conjunto de prueba como se puede evidenciar en los valores obtenidas en la **Tabla 7**.

Tabla 7

Resultados de la evaluación del modelo Support Vector Regression (SVR)

Modelo	RMSE	MAE	R^2
SVR (baseline)	61 471	57 953	-0.13
SVR (tuned)	17 833	14 508	0.905

Nota. Modelos entrenados en Matlab 2025b

En la **Tabla 8** se presentan las métricas de desempeño obtenidas para el modelo SVR tras el proceso de optimización.

Tabla 8

Valores optimizados de los hiperparámetros obtenidos para el modelo Support Vector Regression (SVR)

Parámetro	Tipo de parámetro	Valor optimizado
C	Controla el equilibrio entre precisión del ajuste y la suavidad del modelo	100
Épsilon	Define la amplitud de la región alrededor de la función de regresión. Que tan sensible responde el modelo a variaciones	0.01
Kernel	Determina la forma de la relación entre las variables de entrada y objetivo. Lineal o RBF (no lineal)	lineal

Nota. Valores obtenidos en base a resultados de las iteraciones de entrenamiento de modelos.

En conjunto, los parámetros optimizados indican que el modelo SVR converge hacia una solución cercana a una regresión lineal regularizada, reforzando la evidencia de que el sistema productivo analizado responde principalmente a relaciones lineales.

Evaluación del modelo 3: Random Forest Regressor. Se implementó un modelo Random Forest Regressor basado en ensamble de árboles, con el fin de contrastar su capacidad predictiva frente a modelos lineales y SVR. Random Forest trabaja directamente con variables numéricas y captura interacciones no lineales mediante particiones sucesivas. La optimización de hiperparámetros se realizó mediante validación cruzada de 5 particiones, evaluando el desempeño con RMSE en el conjunto de entrenamiento. Los parámetros optimizados se muestran en la **Tabla 9**.

Tabla 9

Hiperparámetros optimizados para el modelo Random Forest Regressor

Estimators	Min Samples Split	Min Samples leaf	Max features	Bootstrap
800	5	1	0.8	True

Nota. Valores obtenidos en base a resultados de las iteraciones de entrenamiento de modelos.

En la **Tabla 10** se muestra el desempeño del modelo en su configuración base y optimizada. Se evidencia una reducción de error y un incremento del coeficiente de determinación en el modelo ajustado.

Tabla 10

Resultados de la evaluación del modelo Random Forest Regressor en su configuración base y optimizada

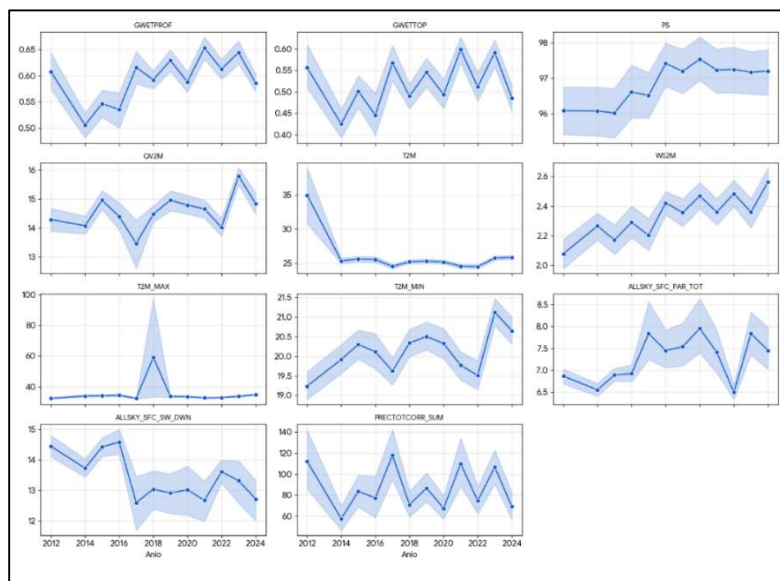
Modelo	RMSE	MAE	R2
RandomForest (tuned)	16 304	13 311	0.9203
RandomForest (baseline)	16 507	13 432	0.9183

Nota. Entrenamientos realizados y testeados en Matlab 2025b.

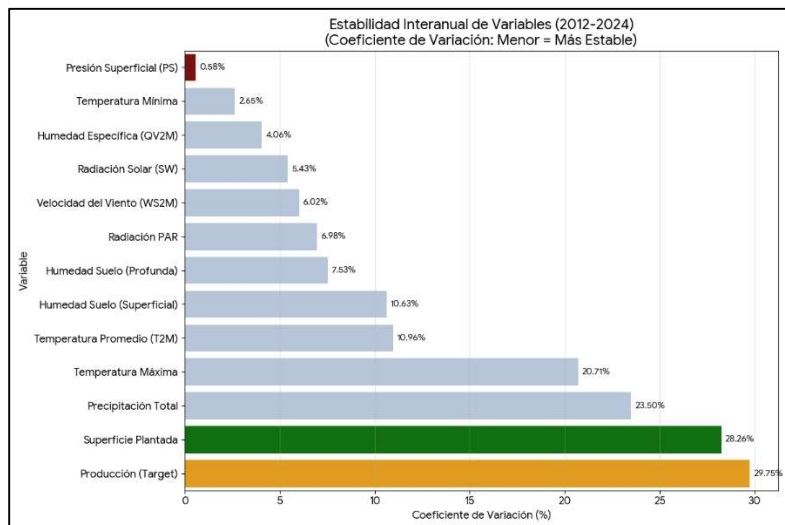
Reducción de dimensionalidad y validación de modelo estructural en Matlab. Para reforzar los resultados obtenidos y verificar la consistencia de los modelos entrenados, se realizó un procedimiento de ingeniería de características orientado a una nueva reducción de dimensionalidad.

Se evaluó la dinámica temporal de las variables de entrada climáticas y productivas mediante el cálculo del Coeficiente de Variación Interanual (CV) y los Intervalos de Confianza del 95% (IC) para el periodo de análisis. Como se observa en la **Figura 10** y **Figura 11**, los resultados revelan:

- **Variables Estáticas (CV < 10%):** Variables como Humedad Especifica, Radiación solar, velocidad del viento y presión superficial, mostraron un comportamiento altamente estable, con bandas de confianza estrechas.
- **Variables Dinámicas (CV > 25%):** Variables como superficie plantada y cosechada.

Figura 10*Evaluación anual de variables meteorológicas*

Nota. La figura muestra el comportamiento interanual de las variables meteorológicas obtenidas.

Figura 11*Estabilidad interanual de variables durante el periodo 2012-2024*

Nota. La figura indica la estabilidad de la variable PS, siendo la más representativa de todas.

Dada la evidencia de que las variables climáticas, en su mayoría, actúan como “ruido de fondo” constante y no como influencia en el rendimiento anual, se procedió a validar un modelo simplificado. Utilizando la herramienta Regression Learner de Matlab, se entrenaron nuevos algoritmos predictivos eliminando las variables climáticas secundarias y conservando:

1. Superficie Plantada
2. Provincia (Aplicando One-hot Encoding)
3. Presión Superficial (PS)

Análisis de escenarios

Diseño experimental y definición de escenarios de validación. Inicialmente, se consideró la exclusión de los periodos 2016 (Terremoto) y 2020 (Pandemia) bajo la premisa de que actuaban como Outliers en un modelo climático. Sin embargo, al reorientar el modelo hacia un enfoque basado en variables estructurales, justificado por el análisis previo, estos eventos dejan de ser ruido y se convierten en puntos de validación crítica.

Una vez definida la arquitectura del modelo y seleccionadas las variables de mayor relevancia (Superficie, Provincia y PS), se estableció un protocolo de pruebas para evaluar la precisión del modelo y su utilidad práctica como herramienta de planificación agrícola.

1. **Escenario de validación temporal (Realidad 2024):** Se utiliza el conjunto de datos correspondientes al año 2024 para comparar las predicciones del modelo frente a los datos reales reportados por organismos oficiales.
2. **Escenario de Sensibilidad Estructural:** Dado que la variable de superficie fue identificada como el predictor dominante, se simulan escenarios hipotéticos de expansión (+10% de superficie) y contracción (-10% de superficie) manteniendo constantes el resto de las variables.

Arquitectura del sistema web

El proyecto implementa un sistema web para apoyar el análisis y visualización de información agrícola, con énfasis en la producción de banano. El sistema se compone de un backend en Python (FastAPI) que expone servicios REST para consulta y predicción, y un frontend en React + Vite orientado a la visualización interactiva sobre mapa (MapLibre + deck.gl) y elementos 3D (GLB). El despliegue en producción se realizó separando responsabilidades: el backend se hospeda en Railway para controlar el runtime de Python y dependencias científicas, mientras que el frontend se hospeda en Vercel, plataforma optimizada para aplicaciones con Content Delivery Network (CDN) global y despliegue continuo desde GitHub.

Los componentes principales de la arquitectura se describen a continuación:

- **Cliente Web:** Desarrollado en React con Vite, responsable de la interfaz de usuario, consumo de la API REST y el renderizado de mapas y modelos 3D.
- **API REST:** Implementada en FastAPI, encargada de validar las solicitudes, prepara las variables, ejecutar el modelo de ML y retornar los valores de predicción.
- **Modelo ML:** Un pipeline de la librería scikit-learn serializado en formato joblib, entrenado con el conjunto de datos histórico.
- **Datos Históricos:** Archivos CSV utilizados para listar las opciones de superficie y presión superficial por provincia disponibles para el usuario.

La interfaz del sistema en producción se muestra en la **Figura 12**.

Figura 12

Interfaz general del sistema de estimación de producción de banano



Nota. La figura indica la captura general de la aplicación web en producción

Backend

El Backend está implementado con FastAPI por su rendimiento, tipado y generación automática de documentación (OpenAPI/Swagger). El servicio carga inicialmente el modelo serializado y el dataset histórico para alimentar las opciones del formulario.

Frontend

El Frontend se construyó con React para componer la interfaz en componentes reutilizables con Vite como bundler por su rapidez en desarrollo y optimización en build. La lógica de la aplicación se distribuye en componentes clave descritos en la **Tabla 11**.

Tabla 11*Componentes principales del Frontend*

Componente	Responsabilidad	Detalle técnico
api.js	Cliente HTTP del backend.	Gestiona base URL.
usePrediction.jsx	Hook personalizado que maneja el estado, la carga de datos y la ejecución de la predicción.	Carga provincias y opciones.
PredictionPanel.jsx	Interfaz de usuario para la selección de parámetros y visualización del resultado de la variable objetivo.	Botones y muestra de información de predicción.
EcuadorCropsNoTerrain.jsx	Componente encargado de la visualización del mapa y la capa 3D.	MapLibre, deck.gl, ScenegraphLayer
DropdownMenu.jsx – ProfileCard.jsx	Navegación y perfiles	Menu accesible y tarjetas con información de integrantes del equipo.
MobileMenu.jsx	Navegación responsive	Alternativa para pantallas pequeñas (md:hidden).

Nota. Elaboración propia

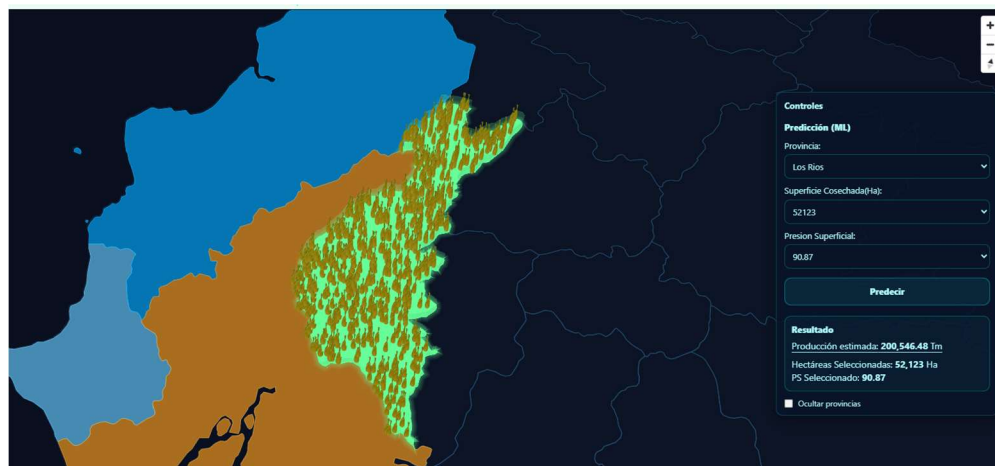
Visualización Geográfica y 3D

Uno de los componentes centrales del sistema es la visualización geoespacial, implementada mediante la integración de MapLibre (como mapa base) y deck.gl para la superposición de capas avanzadas.

El sistema carga un archivo GeoJSON con los límites provinciales de la región de estudio, aplicando estilos dinámicos de relleno y extrusión. Para la representación tridimensional de los cultivos, se utiliza la capa *ScenographLayer* de deck.gl, la cual renderiza modelos en formato GLB distribuidos de manera aleatoria dentro de las regiones provinciales. La **Figura 13** muestra el detalle de la visualización tridimensional sobre una provincia seleccionada.

Figura 13

Visualización 3D (GLB) sobre provincia seleccionada.



Nota. La figura muestra la selección de la provincia Los Ríos, en la cual se muestran los objetos 3D dependiendo del resultado de producción estimada.

Integración Frontend – Backend

La comunicación se realiza mediante peticiones HTTP al backend. El Frontend define un cliente en `api.js` que centraliza `fetch`, parsea JSON y normaliza errores. La URL base se obtiene desde la variable `VITE_API_BASE` (entorno) con fallback a la URL de Railway, permitiendo

separar la configuración por ambiente (local, preview, producción). Para garantizar una integración fluida y segura, el sistema usa variables de entorno (como `VITE_API_BASE`) que definen la dirección de la API. Adicionalmente, se implementan políticas de seguridad CORS en el backend para restringir el acceso a los recursos, autorizando explícitamente solo las peticiones que provienen del dominio de producción en Vercel o del entorno local, lo que previene el consumo no autorizado de los servicios.

Despliegue en la nube

El despliegue de la infraestructura en la nube se realizó utilizando contenedores Docker en la plataforma Railway para el backend, garantizando la reproducibilidad del entorno y la compatibilidad binaria de las dependencias (Numpy, Pandas, Scipy) mediante el uso de una imagen base *Python:3.11-slim* ejecutada con un servidor Uvicorn gestionado por Gunicorn. El frontend se despliega como un sitio estático construido por Vite. Vercel publica el directorio y distribuye el contenido por CDN. Esto reduce latencia y simplifica la operación, manteniendo el backend como servicio independiente.

Seguridad, rendimiento y mantenibilidad

En términos de seguridad, el sistema web desarrollado adopta prácticas alineadas con arquitecturas modernas orientadas a servicios. Se implementaron políticas de control de acceso mediante restricción de orígenes permitidos (CORS), junto con esquemas de validación de datos de entrada que garantizan la integridad de solicitudes y un manejo controlado de errores HTTP. La arquitectura aprovecha las capacidades de FastAPI para operaciones de entrada y salida eficientes, cargando el modelo de ML una única vez durante la inicialización del servicio.

Estrategia de pruebas y verificación del sistema web

La estrategia de verificación se organizó en tres niveles. En el primer nivel se implementaron pruebas funcionales sobre la interfaz de programación de aplicaciones (API), validando el correcto funcionamiento de los endpoints mediante herramientas de documentación interactiva. Estas pruebas permitieron comprobar que las solicitudes de predicción se procesan adecuadamente cuando se ingresan parámetros válidos.

En el segundo nivel, se ejecutaron pruebas de integración orientadas a verificar la comunicación efectiva entre el backend y el frontend del sistema. En esta etapa se confirmó la recepción de datos en la interfaz web. Este nivel de verificación permitió asegurar que los datos generados por el modelo predictivo sean correctamente interpretados y presentados al usuario final como se muestra en la **Figura 14**.

Figura 14

Verificación de datos presentados en la interfaz web

Controles

Predicción (ML)

Provincia:
Los Rios

Superficie Cosechada(Ha):
53566.06575

Presion Superficial:
91.04

Predecir

Resultado

Producción estimada: **204.545,59 Tm**

Hectáreas Seleccionadas: **53.566,066 Ha**

PS Seleccionado: **91,04**

☐ Ocultar provincias

Nota. Captura de pantalla de los resultados obtenidos en la interfaz web con datos históricos.

Finalmente, se llevaron a cabo pruebas de verificación en un entorno de producción, enfocadas en evaluar la operatividad del sistema una vez desplegada la infraestructura en la nube. Estas pruebas incluyeron validación de las direcciones de acceso a los servicios, la revisión de registros de ejecución y el análisis de los tiempos de respuesta, con el objetivo de identificar posibles fallos, errores de comunicación o problemas asociados a restricciones de seguridad y control de acceso.

Capítulo 4

Análisis de Resultados

En este capítulo se presentan los resultados obtenidos tras el entrenamiento, validación y prueba de los modelos predictivos para estimar el rendimiento de la producción de banano. El análisis se estructura en tres fases: primero, la validación métrica de los algoritmos entrenados con la serie histórica completa (2012-2023); segundo, la validación del modelo seleccionado frente a datos históricos del 2024; tercero, la simulación de escenarios de sensibilidad para medir la respuesta del modelo ante cambios estructurales en la superficie cultivada.

Evaluación del rendimiento de modelos entrenados (2012-2023)

Se utilizó la herramienta Regression Learner de Matlab, empleando un enfoque de validación cruzada (k-Fold Cross Validation, $k=5$) para garantizar que los resultados no dependen de un sesgo en la partición de datos. Las variables predictoras utilizadas fueron:

- Superficie cosechada (Ha)
- Provincia (One-Hot Encoding)
- Presión superficial (Ps)

Comparativa de algoritmos y métricas de error. Se sometieron distintas familias de algoritmos para evaluar su capacidad de ajuste. Los resultados obtenidos se resumen en la **Tabla 12**, donde se contrastan el coeficiente de determinación, error cuadrático medio y el error absoluto medio.

Tabla 12

Comparación del desempeño de los algoritmos de regresión mediante R^2 , RMSE y MAE

Algoritmo	Tipo	R^2 (Ajuste)	RMSE (Tm)	MAE (Tm)
Linear Regression	Linear	0.91284	13353	9261.9
Linear Regression	Interactions Linear	0.91531	13162	9401.2
Linear Regression	Robust Linear	0.90305	14082	9748.1
Stepwise Linear Regression	Stepwise Linear	0.91181	13431	9513
SVM	Linear SVM	0.90541	13910	9879.3
SVM	Quadratic SVM	0.90824	13700	10047
SVM	Cubic SVM	0.166	41304	31525
SVM	Fine Gaussian SVM	0.94997	10117	7984.4
SVM	Medium Gaussian SVM	0.89575	14603	10232
SVM	Coarse Gaussian SVM	0.90365	14039	9976
Ensemble	Boosted Trees	0.96581	8363.4	7488.2
Ensemble	Bagged Trees	0.99067	4369	2850.8

Nota. Los entrenamientos y resultados obtenidos en la tabla fueron realizados en Matlab, utilizando el módulo de Regression Learner.

Los resultados muestran diferencias claras en el desempeño predictivo entre los distintos algoritmos evaluados.

Modelos Lineales

Las variantes de regresión lineal (lineal, interacciones y stepwise) presentan un desempeño estable, con valores de R^2 entre 0.91 y 0.92, y errores RMSE del orden de 13 000 a 14 000. Esto confirma que una gran parte de la variabilidad de la producción puede ser explicada mediante relaciones aproximadamente lineales, asociadas a variables estructurales.

Modelos SVM

El desempeño de los modelos SVM es altamente dependiente del kernel utilizado.

Mientras que el Fine Gaussian SVM alcanza un buen ajuste superior al 0.90, otros kernels como el cúbico presentan comportamientos deficientes, evidenciando problemas de sobreajuste o poca interpretación de los datos.

Modelos de ensamble

Los mejores resultados se obtienen con los modelos de ensamble. Particularmente con Boosted Trees y Bagged Trees, los cuales superan al resto de modelos entrenados. Esto indica una alta capacidad para capturar relaciones lineales.

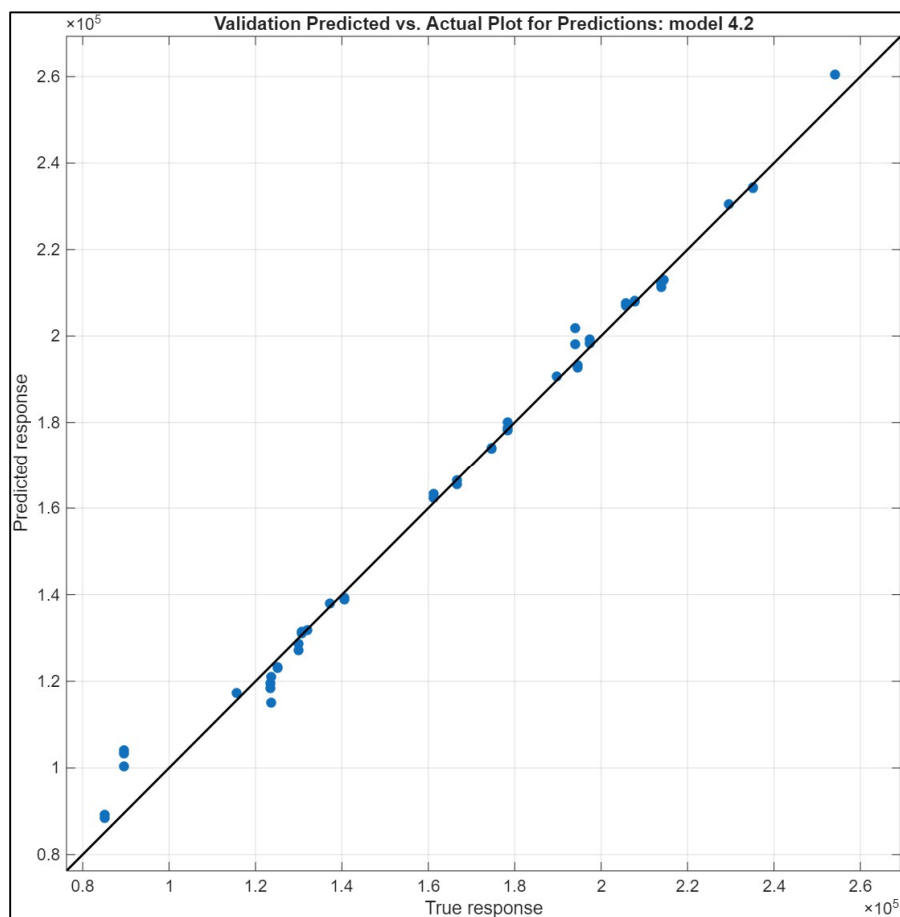
Selección del modelo. Los modelos de ensamble demostraron un comportamiento superior al resto de arquitecturas. Bagged Trees tiene una capacidad consistente para reproducir el rendimiento. Esta estabilidad resulta relevante para la fase de escenarios, en la cual se evalúan modificaciones estructurales.

Los modelos lineales se mantienen como referencia interpretativa del sistema productivo, pero no se seleccionan como base para escenarios debido a su menor capacidad para capturar variaciones locales. De la misma forma, modelos SVM se descartan debido a su elevada dependencia de la configuración e hiperparámetros para mejorar su estabilidad.

En consecuencia, se selecciona Bagged Trees como modelo principal para la simulación de escenarios, al presentar el mejor equilibrio entre precisión predictiva, estabilidad y capacidad de generalización, estableciendo una base sólida y confiable para el análisis del rendimiento del banano. La **Figura 15** presenta la comparación entre los valores reales y predichos del modelo, evidenciando una fuerte concordancia. La dispersión reducida indica que el modelo mantiene un comportamiento estable tanto para niveles bajos como altos de producción.

Figura 15

Comparación entre valores reales y predichos del modelo



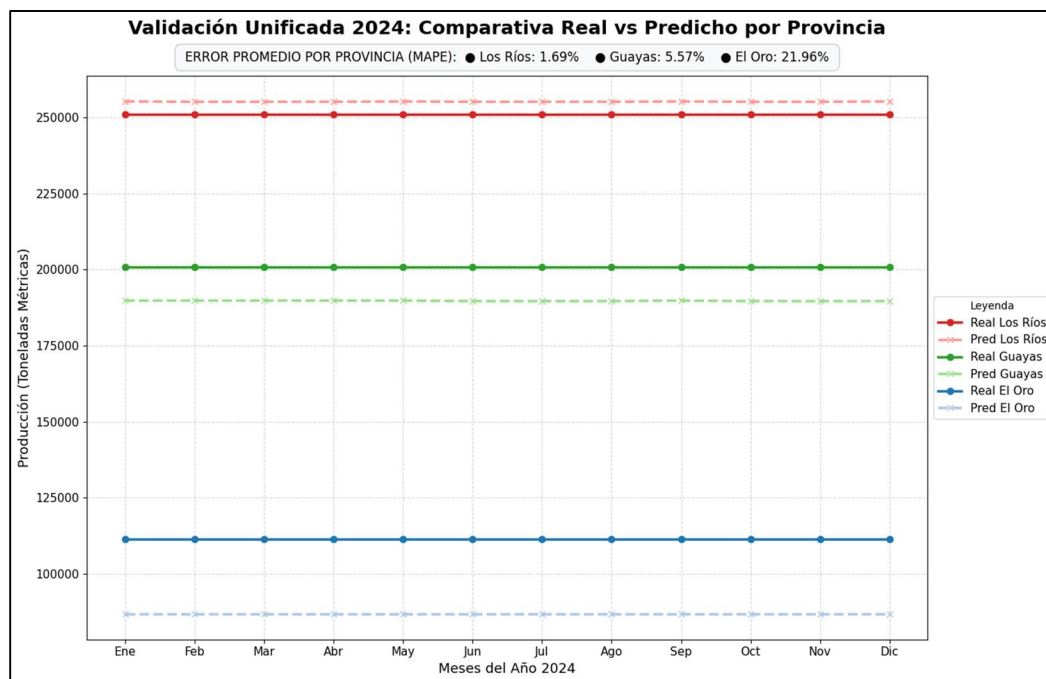
Nota. La Figura muestra el resultado obtenido con el mejor modelo entrenado en Matlab.

Análisis de Resultados

Escenario de validación temporal (Realidad 2024). La **Figura 16** presenta la comparación entre la producción real y la producción predicha para el año 2024, separada por provincia, utilizando el modelo Bagged Trees. Los resultados muestran un comportamiento diferenciado entre provincias, asociado a su estructura productiva y representatividad histórica de los datos utilizados en el entrenamiento.

Figura 16

Comparación entre la producción real y la producción predicha en 2024 por provincia utilizando el modelo Bagged Trees.



Nota. Se seleccionaron los datos del año 2024 para realizar el testeo del modelo, tomando en cuenta las provincias más relevantes del cultivo en la región.

En el caso de Los Ríos, el modelo reproduce con alta precisión el nivel de producción, con un error porcentual promedio (MAPE) de 1.69%. Este resultado confirma que el modelo captura adecuadamente la dinámica productiva de la provincia.

Para Guayas, la predicción presenta una desviación moderada respecto a los valores reales, con un MAPE de 5.57%. Este resultado es aceptable desde el punto de vista predictivo y puede atribuirse a una mayor variabilidad histórica en la producción provincial.

En contraste, El Oro presenta la mayor discrepancia entre valores reales y predichos, con un MAPE de 21.96%. Este comportamiento muestra una menor capacidad de predicción del

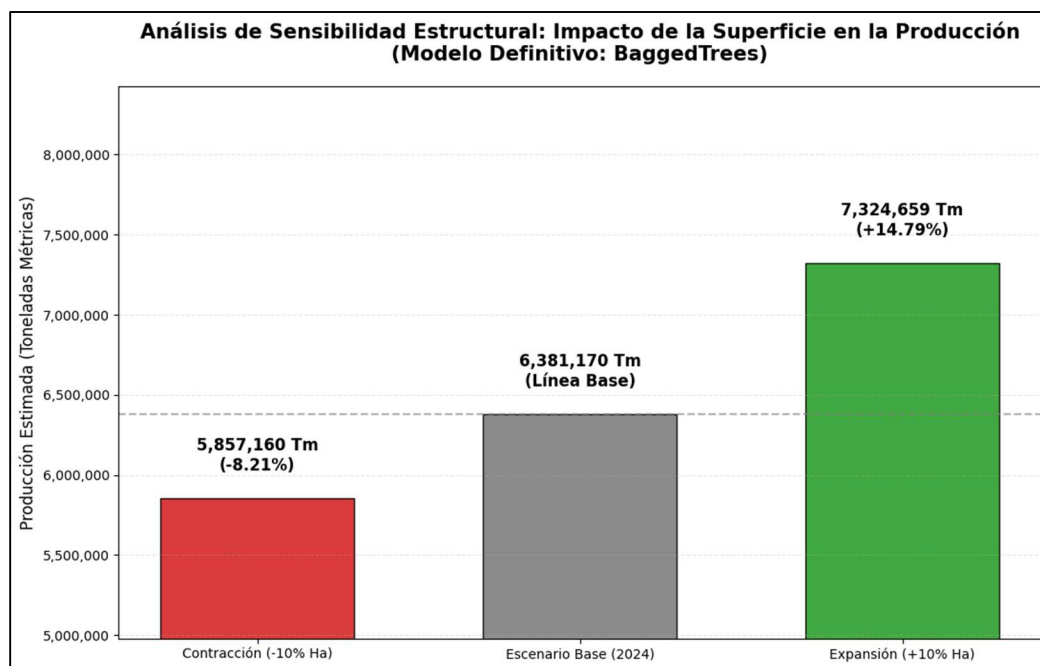
modelo, lo cual es consistente con su menor peso relativo en el conjunto de entrenamiento y una estructura productiva distinta.

De manera general, la validación confirma que el modelo Bagged Trees mantiene una alta capacidad predictiva a nivel provincia, especialmente en zonas con mayor estabilidad estructural y mayor representatividad histórica.

Escenario de Sensibilidad Estructural. La **Figura 17** presenta el escenario de expansión y contracción de la superficie cultivada anualmente, evaluado mediante el modelo Bagged Trees. En este análisis se simulan variaciones estructurales en el área productiva, manteniendo constantes el resto de las variables del sistema, con el objetivo de cuantificar su impacto directo sobre la producción estimada del banano.

Figura 17

Análisis de sensibilidad del impacto de la superficie cultivada sobre la producción estimada de banano utilizando el modelo Bagged Trees



Nota. Elaboración propia

Los resultados evidencian una alta sensibilidad ante cambios de la superficie. En el escenario de contracción, la producción estimada se reduce a 5 857 160 toneladas métricas, lo que representa una reducción del -8.21%, respecto al escenario base del año 2024. Por el contrario, el escenario de expansión alcanza un valor de 7 324 659 toneladas métricas, correspondientes a un incremento de +14.79%.

Los resultados indican que los escenarios de reducción y aumento de superficie no presentan una relación estrictamente proporcional, sino que tiene efectos no lineales asociados a la estructura productiva y a la interacción de sus variables de entrada. Sin embargo, la tendencia se mantiene consistente con la lógica agronómica: el aumento en la superficie se traduce a un crecimiento significativo de la producción.

Capítulo 5

Conclusiones

- El proceso de adquisición e integración de datos constituyó un componente crítico del estudio, al requerir la unificación de fuentes productivas oficiales y datos climáticos de distinta periodicidad. La información obtenida de la ESPAC y NASA POWER consistió en registros anuales y mensuales consistentes y estructuralmente homogéneos para el periodo de análisis 2012-2024.
- El conjunto final de datos utilizados para el entrenamiento del modelo se construyó a partir de los años con información completa y estructuralmente consistente, lo que dispuso de un volumen suficiente de observaciones para el periodo 2012-2023. El año 2024 fue reservado exclusivamente para la validación temporal, fortaleciendo la evaluación del modelo.
- La producción bananera en la Costa ecuatoriana presenta una dinámica estructural a escala anual, donde las variables de superficie plantada y cosechada explican la variabilidad observada entre el periodo 2012-2024. Este comportamiento fue consistente a lo largo de todos los modelos evaluados.
- Las variables climáticas actúan principalmente como factores moduladores secundarios, mostrando baja variabilidad interanual y contribuciones en la predicción del rendimiento anual. Su influencia es más relevante a escalas temporales menores o en contextos de eventos extremos.
- Los modelos de ensamble superaron sistemáticamente a los modelos lineales y SVM, destacándose Bagged Trees como el algoritmo con mejor equilibrio entre

precisión, estabilidad y capacidad de generalización, alcanzando valores de $R^2 > 0.99$ y los menores errores RMSE y MAE.

- La validación temporal con datos reales del año 2024 confirmó la robustez del modelo seleccionado, obteniendo altos niveles de precisión en provincias con mayor representatividad histórica (Los Ríos y Guayas), y evidenciando limitaciones en provincias con menor peso relativo como El Oro.
- En conjunto, el estudio demuestra que la combinación de datos oficiales, una metodología rigurosa de integración temporal y modelos de ensamble robustos permite generar predicciones confiables del rendimiento bananero, aportando una base sólida para simulaciones de escenarios y la toma de decisiones en el sector agrícola.
- El sistema web demuestra la viabilidad de integrar un modelo entrenado de ML con una interfaz web rica en visualización para apoyar análisis agrícolas. La separación de despliegues (Railway/Vercel) mejora la operabilidad y permite escalar componentes de forma independiente.

Recomendaciones

- Se recomienda que futuras investigaciones incorporen predictores a nivel provincial y cantonal de manera sistemática. Aunque el modelo anual mostró un desempeño estable, la variabilidad climática y productiva no es homogénea entre provincias de la región Costa. Una mayor desagregación permitiría capturar mejor los efectos locales.

- Se recomienda incorporar en futuros estudios variables que expliquen la dinámica de la superficie agrícola. Esto permitiría modelar de manera más integral las decisiones de expansión y contracción para escenarios de validación.
- Se recomienda estructurar futuros modelos en dos etapas: un primer modelo orientado a la evolución de la superficie, y un segundo enfocado en la estimación del rendimiento a partir de dicha superficie. Esta separación permitiría reducir el sesgo estructural y mejorar la interpretación de datos.
- Se recomienda en futuros trabajos extender la cobertura del sistema web incorporando información correspondiente a un mayor número de provincias y cultivos. De la misma forma, la integración de múltiples modelos predictivos para realizar comparaciones técnicas y mejorar la capacidad de análisis.
- Se recomienda implementar flujos automatizados de pruebas y validación continua que permitan detectar errores de forma temprana y asegurar estabilidad del sistema ante futuras modificaciones de la interfaz web.

Referencias Bibliográficas

- Baltazar, R. G. (2024). Forecasting the impact of climate change on rice crop yields under RCP4.5 and RCP8.5 scenarios in Central Luzon, Philippines, using machine learning algorithms. *Ciencia e Investigación Agraria: Revista Latinoamericana de Ciencias de la Agricultura*, 51(1), 10–26. <https://doi.org/10.7764/ijanr.v51i1.2494>
- Instituto Nacional de Estadística y Censos. (s.f.). *Estadísticas agropecuarias*. Recuperado el 10 de septiembre de 2025, de <https://www.ecuadorencifras.gob.ec/estadisticas-agropecuarias-2/>
- National Aeronautics and Space Administration. (s.f.). *POWER Daa Access Viewer*. Recuperado el 10 de septiembre de 2025, de <https://power.larc.nasa.gov/data-access-viewer/>
- Borse, K., Agnihotri, P. G., Jothimani, M., & Jena, R. K. (2025). Adaptive neuro fuzzy inference system based multicrop yield prediction in the semi-arid region of India. *Scientific Reports*, 15(1), 18787. <https://doi.org/10.1038/s41598-025-03334-8>
- Choi, J. W., Hidayat, M. S., Cho, S. B., Hwang, W. H., Lee, H., Cho, B. K., & Kim, G. (2025). Recent trends in machine learning, deep learning, ensemble learning, and explainable artificial intelligence techniques for evaluating crop yields under abnormal climate conditions. *Plants*, 14(18), 2841. <https://doi.org/10.3390/plants14182841>
- Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11), 114003. <https://doi.org/10.1088/1748-9326/aae159>
- De Clercq, D., & Mahdi, A. (2024). Feasibility of machine learning-based rice yield prediction in India at the district level using climate reanalysis and remote sensing data. *Agricultural Systems*, 220, 104099. <https://doi.org/10.1016/j.agry.2024.104099>

- Edwin Prem Kumar, G., & Lydia, M. (2021). Machine learning algorithms for modelling agro-climatic indices: A review. *In Smart computing techniques and applications: Proceedings of the fourth international conference on smart computing and informatics* (Vol. 2, pp. 15–23). Springer. https://doi.org/10.1007/978-981-16-1502-3_3
- SWI swissinfo.ch. (2016, 22 de enero). *El 2016 será un año complicado para el banano de Ecuador*. <https://www.swissinfo.ch/spa/el-2016-ser%C3%A1-un-a%C3%B1o-complicado-para-el-banano-de-ecuador/41913254>
- Banco Mundial. (2021, 27 de abril). *A cinco años del terremoto, Ecuador sigue trabajando en su resiliencia frente a desastres*. <https://www.bancomundial.org/es/news/feature/2021/04/27/a-cinco-a-os-del-terremoto-ecuador-sigue-trabajando-en-su-resiliencia-frente-a-desastres>
- Pesantes, K. (2020, 6 de marzo). *En seis ciudades de Ecuador cayeron más de 90 litros de lluvia por metro cuadrado. Primicias*. <https://www.primicias.ec/noticias/sociedad/seis-ciudades-costa-inundaciones/>
- Cámara Marítima del Ecuador. (2021, 22 de enero). *Exportación bananera alcanza nuevo récord al comercializar 380 millones de cajas en el 2020*. <https://www.camae.org/exportaciones/exportacion-bananera-alcanza-nuevo-record-al-comercializar-380-millones-de-cajas-en-el-2020/>
- González, P. (2023, 7 de marzo). *Ecuador enfrenta el fenómeno de La Niña más largo en 120 años. Primicias*. <https://www.primicias.ec/noticias/economia/fenomeno-nina-costos-banano-pesca/>

- Farooq, H., Mazhar, R., & Jan, S. A. (2025). Mitigating climate change impacts on agriculture through AI-driven crop improvement. *MOJ Biology and Medicine*, 10(1), 29–34.
<https://doi.org/10.15406/mojbm.2025.10.00237>
- Jabed, M. A., & Murad, M. A. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*, 10(24).
<https://doi.org/10.1016/j.heliyon.2024.e40836>
- Majumder, S., & Mason, C. M. (2025). Sunflower yield modeling with XAI: Historical weather impacts and forecasting. *bioRxiv*. <https://doi.org/10.1101/2025.02.27.640573>
- Mamani, L. M. T. (2024). *Predictive analysis and optimization in sustainable agriculture facing climate change with emerging technological approaches*.
<https://doi.org/10.20944/preprints202412.1416.v1>
- Mohan, R. J., Rayanoothala, P. S., & Sree, R. P. (2025). Next-gen agriculture: Integrating AI and XAI for precision crop yield predictions. *Frontiers in Plant Science*, 15, 1451607.
<https://doi.org/10.3389/fpls.2024.1451607>
- Pacheco Gil, H. (2017). Efectos del sismo del 16 de abril de 2016 en el sector productivo agropecuario de Manabí. *La Técnica: Revista de las Agrociencias*, (17).
https://doi.org/10.33936/la_tecnica.v0i17.692
- Ramesh, B., & Veerasha, M. (2025). A Eurasian case study on climate change and agricultural land suitability using interpretable machine learning. *International Journal of Engineering Research and Science & Technology*, 21(2), 1115–1124.
<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1115-1124>

- Renuka, A. (2025). AI-driven predictive analytics in precision agriculture. *Scientific Journal of Artificial Intelligence and Blockchain Technologies*, 2(3), 9–17.
<https://doi.org/10.63345/sjaibt.v2.i3.102>
- Roa Chejín, S. (2022, 24 de junio). *Así ha afectado el paro nacional al sector bananero de Ecuador*. GK City.
<https://gk.city/2022/06/24/asi-ha-afectado-paro-nacional-sector-bananero-ecuador/>
- EFEAgro. (2023, 31 de enero). *La exportación de banano de Ecuador cae 6,57 % en 2022 con 354 millones de cajas*.
<https://efeagro.com/exportacion-banano-ecuador/>
- Sghir, A., & Iroshan, A. (2025). Machine learning integrated climate–agriculture forecasting: A transformer-based approach to predict precipitation and wheat production amidst climate change. *SSRN*. <https://doi.org/10.2139/ssrn.5137941>
- Sidhu, B. S., Mehrabi, Z., Ramankutty, N., & Kandlikar, M. (2023). How can machine learning help in understanding the impact of climate change on crop yields? *Environmental Research Letters*, 18(2), 024008. <https://doi.org/10.1088/1748-9326/acb164>
- Tamayo-Vera, D., Liu, K., Bolufé-Röhler, A., & Wang, X. (2024). Projecting future changes in potato yield using machine learning techniques: A case study for Prince Edward Island, Canada. *Environmental Research Communications*, 6(10), 105025.
<https://doi.org/10.1088/2515-7620/ad85c5>
- Thirumal, S., & Latha, R. (2023). Machine learning-based predictive assessments of impacts of influential climatic conditions for the sustainable productivity of paddy crops. *SSRG International Journal of Electronics and Communication Engineering*, 10(5), 116–127.
<https://doi.org/10.14445/23488549/IJECE-V10I5P111>

Vignesh, K., Askarunisa, A., & Abirami, A. M. (2023). Optimized deep learning methods for crop yield prediction. *Computer Systems Science & Engineering*, 44(2).

<https://doi.org/10.32604/csse.2023.024475>

Xiao, L., Wang, G., Zhou, H., Jin, X., & Luo, Z. (2022). Coupling agricultural system models with machine learning to facilitate regional predictions of management practices and crop production. *Environmental Research Letters*, 17(11), 114027.

<https://doi.org/10.1088/1748-9326/ac9c71>

Zidan, F., & Febriyanti, D. E. (2024). Optimizing agricultural yields with artificial intelligence-based climate adaptation strategies. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, 5(2), 136–147. <https://doi.org/10.34306/itsdi.v5i2.663>