Optimización de la productividad agrícola en Boyacá mediante técnicas de Machine Learning

Samuel Mantilla Velásquez

Asesor

Jorge Eliecer Ospino Portillo

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de datos y Analítica

2025

Dedicatoria

Dedico este trabajo a mi madre, hermana y a mi pareja, por ser mi motivación y por su apoyo incondicional. Su amor, confianza y compañía han sido fundamentales en cada paso de este camino, recordándome siempre que los sueños se construyen con esfuerzo y se alcanzan con el apoyo de quienes verdaderamente creen en nosotros.

Resumen

En este proyecto de investigación pretende la implementación de modelos predictivos de machine learning ML, para optimizar las decisiones agrícolas en el departamento de Boyacá. Por medio de la integración de datos locales (fuentes de datos libres y departamentales) sobre características del suelo, condiciones climáticas y prácticas agrícolas, con el fin de desarrollar un sistema que pueda predecir rendimientos, identificar deficiencias en el suelo para poder realizar recomendaciones y así mejorar la producción agrícola. Investigaciones previas han demostrado el potencial de estas tecnologías para optimizar la agricultura, donde la mayoría de estos estudios se centran en contextos internacionales y carecen de aplicaciones locales en Colombia. Este trabajo busca llenar ese vacío al adaptar las soluciones tecnológicas al contexto específico de Boyacá.

Por medio de la recopilación y el análisis de datos, en este proyecto busca mejorar la eficiencia en la toma de decisiones enfocado en cultivos, rotación y manejo de recursos. Se espera que los resultados proporcionen a los agricultores herramientas para optimizar sus procesos ahora basados en datos y, a su vez, promuevan la sostenibilidad y mejoras en la eficiencia agrícola en la región de Boyacá. Además, los resultados podrían influir en la adopción de tecnologías avanzadas en políticas públicas agrícolas. (Sabogal García, 2021).

***Palabras claves*:** Agrícola, Boyacá, ML, Modelo de Predicción, Productividad

Abstract

This research project aims to implement predictive machine learning (ML) models to optimize agricultural decision-making in the department of Boyacá. By integrating local data (open and departmental data sources) on soil characteristics, climatic conditions, and agricultural practices, the project seeks to develop a system capable of predicting yields, identifying soil deficiencies, and providing recommendations to improve agricultural production. Previous studies have demonstrated the potential of these technologies to enhance agriculture, but most focus on international contexts and lack local applications in Colombia. This work aims to bridge that gap by adapting technological solutions to the specific context of Boyacá.

Through data collection and analysis, this project seeks to improve decision-making efficiency, focusing on crops, rotation, and resource management. The expected results aim to provide farmers with data-driven tools to optimize their processes while promoting sustainability and efficiency improvements in agriculture in Boyacá. Additionally, the findings could influence the adoption of advanced technologies in agricultural public policies (Sabogal García, 2021).

Keywords: Agriculture, Boyacá, ML, Predictive Model, Productivity

Tabla de Contenido

Lista de Tablas

Lista de Figuras

Introducción

La agricultura es un pilar fundamental en la economía de Boyacá, donde una gran parte de la población depende de esta actividad para su sustento. Sin embargo, los productores enfrentan desafíos como el cambio climático, la variabilidad en la calidad del suelo y la falta de acceso a tecnologías avanzadas, lo que limita el rendimiento de los cultivos y afecta su competitividad en el mercado.

En este contexto, la implementación de herramientas basadas en el análisis de datos y machine learning se presenta como una alternativa prometedora para optimizar la producción agrícola. Estas tecnologías permiten analizar factores agroclimáticos, predecir rendimientos y mejorar la toma de decisiones en el campo, contribuyendo así a una producción más eficiente y sostenible.

Este proyecto busca integrar estos enfoques innovadores en la agricultura de Boyacá, con el propósito de mejorar la productividad, reducir pérdidas y fomentar el desarrollo sostenible en la región.

Planteamiento del Problema

La agricultura en Boyacá, uno de los principales departamentos agrícolas de Colombia, se caracteriza por su estabilidad y diversidad en la producción de cultivos como la papa en sus distintas variedades, el maíz, las hortalizas y frutas, (Ministerio de Agricultura y Desarrollo Rural, 2021). Este sector es muy competente y eficiente en relación con otros departamentos y su productividad, existe un gran potencial para optimizar la productividad mediante nuevas tecnologías basadas en datos. (Gobernación de Boyacá, 2023). Los agricultores de la región utilizan en su mayoría prácticas tradicionales, que, si bien han garantizado rendimientos aceptables, no en todos los casos se ha aprovechado al máximo las condiciones tales como suelo, el clima y factores adicionales ejemplo las características de los tipos de cultivo. Según el Instituto Geográfico Agustín Codazzi (IGAC, 2020), más del 40% de los suelos agrícolas en Boyacá presentan características físicas o químicas que requieren manejos específicos para mejorar la capacidad productiva y se rendimiento.

Uno de los principales retos para alcanzar mayores niveles de eficiencia (productiva) es la baja adopción de tecnologías que permitan una toma de decisiones basada en datos. La integración de herramientas como el machine learning (ML) puede ofrecer recomendaciones en funciones de parámetros para mejorar la gestión de cultivos y optimizar el uso de insumos (Parra-Peña et al., 2021). Aunque la agricultura de precisión está en crecimiento en Colombia, su implementación en regiones como Boyacá es limitada, lo que deja margen para mejorar la productividad mediante el uso de modelos predictivos y análisis de datos locales.

Estudios internacionales, como los realizados por (Rambauth Ibarra, 2022) han demostrado que el uso de machine learning en la agricultura permite optimizar procesos como el manejo de insumos, la predicción de rendimientos y la detección temprana de deficiencias

nutricionales. Sin embargo, gran parte de estas investigaciones se han llevado a cabo en contextos internacionales, dejando un vacío importante en cuanto a su implementación en Colombia, y particularmente siendo Boyacá un departamento destacado en aspectos agrícolas. Esta falta de referencias locales presenta una oportunidad para aplicar estas tecnologías en el contexto específico de la región, adaptándolas a las condiciones agroclimáticas y socioeconómicas de los agricultores boyacenses.

El impacto de este problema es significativo, ya que la falta de herramientas tecnológicas limita la capacidad de los agricultores para mejorar sus rendimientos y sostenibilidad. La implementación de modelos predictivos que integren datos sobre el suelo, el clima y las prácticas agrícolas podría ser la clave para maximizar la eficiencia en la producción. Investigaciones previas sugieren que la combinación de datos agroclimáticos con algoritmos de machine learning puede predecir rendimientos y ofrecer recomendaciones personalizadas, lo que a su vez incrementaría la productividad agrícola (Gupta et al., s.f.).

Este proyecto se propone abordar estas limitaciones al desarrollar modelos predictivos basados en machine learning que se adapten a las condiciones específicas de los agricultores en Boyacá.

Justificación

La agricultura en Boyacá, como una de las actividades económicas más relevantes del departamento, tiene un gran potencial para mejorar su productividad mediante la incorporación de tecnologías avanzadas como el ML. A pesar de su estabilidad y diversidad, el sector agrícola en la región no ha aprovechado plenamente las oportunidades que brindan las herramientas tecnológicas modernas, lo que representa una oportunidad estratégica para maximizar los rendimientos y optimizar el uso de los recursos naturales (Gobernación de Boyacá, 2023).

El uso de machine learning en la agricultura se ha consolidado como una herramienta clave para transformar los procesos productivos a nivel mundial ya sea para la optimización de procesos, mejoras en rendimiento o implementación medidas de prevención de plagas (García & Eisenhower, 2021). Según (Kamilaris & Prenafeta-Boldú, 2018) esta tecnología permite analizar grandes volúmenes de datos, identificar patrones complejos y generar predicciones que optimizan la toma de decisiones agrícolas. Su aplicación puede impactar positivamente en la productividad al predecir rendimientos, ajustar insumos como fertilizantes y agua, y mejorar la sostenibilidad general de los cultivos. En el caso de Boyacá, donde la agricultura depende de suelos que presentan limitaciones físicas o químicas, estas herramientas podrían proporcionar un enfoque más eficiente y personalizado para el manejo de los cultivos (Maquera-Callo et al., 2023).

Además, este proyecto busca llenar un vacío en la implementación local de tecnologías avanzadas, adaptando las soluciones tecnológicas a las necesidades específicas de los agricultores boyacenses. Mientras que la literatura internacional demuestra el impacto positivo de estas tecnologías en diversos contextos, en Colombia, y especialmente en Boyacá, su

adopción ha sido limitada. Este trabajo propone no solo mejorar los rendimientos agrícolas, sino también ofrecer un modelo replicable que pueda ser aplicado en otras regiones del país.

Los beneficios de esta investigación se extienden más allá de los agricultores de Boyacá. Por un lado, el desarrollo de modelos predictivos contribuirá al avance del conocimiento en ciencia de datos aplicada a la agricultura (Ramírez Gómez, 2020). Por otro, las herramientas generadas tendrán el potencial de influir en la modernización tecnológica del sector agrícola en Colombia para la gestión de recomendaciones. Además, como señala el (Ministerio de Agricultura y Desarrollo Rural, 2021), iniciativas que integren tecnologías avanzadas pueden generar impactos económicos positivos al incrementar los ingresos netos de los productores y garantizar prácticas agrícolas más sostenibles

En conclusión, este proyecto no solo busca optimizar la productividad agrícola en Boyacá, sino también demostrar cómo la integración de machine learning puede transformar la agricultura en un sector más eficiente, rentable y sostenible. La investigación servirá como un puente entre la tecnología y la práctica agrícola local, siendo una base para una agricultura moderna y adaptada a las necesidades del futuro.

Objetivos

Objetivo General

Evaluar la efectividad de las técnicas de machine learning (ML) en la mejora de la productividad agrícola en Boyacá.

Objetivos Específicos

Implementar modelos de *machine learning* para predecir la productividad agrícola y detectar factores limitantes, como problemas de suelo o riesgos climáticos, considerando condiciones agroclimáticas de Boyacá.

Interpretar los resultados del sistema de optimización y ajustar las recomendaciones basadas en las condiciones específicas de cada zona agrícola de Boyacá.

Analizar el impacto de las recomendaciones generadas por los modelos en la toma de decisiones de los agricultores y su efectividad en la mejora de la productividad agrícola.

Marco de Referencia

Estado del Arte

El machine learning (ML) se ha consolidado como una herramienta transformadora en el sector agrícola, ofreciendo soluciones innovadoras para optimizar diversas etapas del ciclo productivo. Su capacidad para analizar grandes volúmenes de datos y extraer patrones significativos ha impulsado avances notables, desde el análisis del suelo hasta la optimización de la cosecha. El ML permite la toma de decisiones informadas en diversas fases de la producción agrícola, promoviendo la precisión, la eficiencia y la sostenibilidad en las prácticas agrícolas. Se emplean técnicas de ML, incluyendo el aprendizaje supervisado, no supervisado y por refuerzo, para mejorar la exactitud de las predicciones y decisiones en la agricultura. La integración del ML con tecnologías como el *IoT*, drones e imágenes satelitales apoya aún más las iniciativas de agricultura inteligente (Ećim et al., 2024; Reddy et al., 2025; Jhajharia y Mathur, 2022).

Aplicaciones Clave del Machine Learning en la Agricultura

El ML ha permeado múltiples aspectos de la agricultura moderna como:

Gestión de Cultivos y Predicción de Rendimientos: Se emplean algoritmos de ML como árboles de decisión y redes neuronales para la predicción del rendimiento de los cultivos, ayudando a los agricultores a optimizar las estrategias de siembra y la asignación de recursos (Reddy et al., 2025; Jhajharia y Mathur, 2022). Además, algoritmos de ML predicen las condiciones ideales para el crecimiento de los cultivos, considerando factores ambientales clave como la temperatura, la humedad y las precipitaciones (Vinothkumar et al., 2024). Las Redes Neuronales Convolucionales (CNN) son particularmente efectivas en la clasificación de especies de plantas y la detección de enfermedades, lo que facilita una gestión precisa de los cultivos (Sharma y Abrol, 2024)

Gestión del Suelo y del Agua: Los modelos de ML asisten en la evaluación de la calidad del suelo y la optimización de los recursos hídricos, lo cual es crucial para prácticas agrícolas sostenibles (Reddy et al., 2025; Jhajharia y Mathur, 2022). Estas tecnologías contribuyen a un riego y una fertilización eficientes, reduciendo el desperdicio y el impacto ambiental (Sharma y Abrol, 2024; Yadav et al., 2024).

Control de Plagas y Enfermedades: El ML permite la detección temprana y la gestión de enfermedades y plagas en las plantas, minimizando las pérdidas de cosechas y mejorando la seguridad alimentaria (Ećim et al., 2024; Jhajharia y Mathur, 2022). Los modelos predictivos pueden anticipar brotes de enfermedades, permitiendo intervenciones oportunas (Yadav et al., 202). Las CNN han demostrado ser particularmente efectivas en la detección de enfermedades en cultivos específicos como el trigo y la soja, alcanzando altos niveles de precisión (Sharma y Abrol, 2024; Benos et al., 2021)

Análisis de Suelo: Se emplean Redes Neuronales Artificiales (ANN) para analizar las propiedades del suelo, como el contenido de materia orgánica y la fertilidad, permitiendo decisiones más informadas sobre la selección de cultivos y la aplicación de recursos (Biswas y Banik, 2024).

Optimización del Riego: El aprendizaje supervisado se utiliza para analizar datos de sensores remotos y modelos climáticos, prediciendo las necesidades hídricas de los cultivos y minimizando el desperdicio de este recurso vital (Reddy et al., 2025; Yadav etal., 2024).

Gestión de Malezas: Los sistemas de recomendación basados en ML ayudan a identificar y cuantificar las especies de malezas, facilitando la aplicación selectiva de herbicidas y reduciendo el impacto ambiental (Jhajharia & Mathur, 2022; Biswas y Banik, 2024).

Técnicas y Algoritmos Predominantes

La diversidad de problemas agrícolas ha llevado a la aplicación de una amplia gama de técnicas de ML.

Las redes neuronales artificiales (ANN) y las redes neuronales convolucionales (CNN) son fundamentales para el análisis de imágenes y datos complejos, destacando su eficacia en la clasificación de especies de plantas y la detección de enfermedades (Ećim et al., 2024; Sharma y Abrol, 2024).

El aprendizaje supervisado se utiliza cuando los datos están etiquetados, como en la clasificación de especies de plantas, mientras que el aprendizaje no supervisado se aplica en la segmentación de imágenes de suelo y cultivos para identificar patrones inherentes en los datos (Ećim et al., 2024; Reddy et al., 2025).

Las técnicas de aumento de gradiente, como el *gradient boosting*, han probado ser robustas en tareas de predicción y clasificación, demostrando una alta precisión en la predicción de rendimientos y la clasificación de la salud de los cultivos (Biswas y Banik, 2024).

La transferencia de aprendizaje emerge como una técnica valiosa al permitir la reutilización de modelos pre-entrenados para tareas específicas en agricultura. Esto reduce significativamente los tiempos de entrenamiento y mejora la precisión de los modelos en aplicaciones agrícolas particulares (Sharma y Abrol, 2024).

Los sistemas de recomendación basados en ML son utilizados para la gestión de plagas y malezas, sugiriendo tratamientos selectivos y optimizando el uso de pesticidas (Biswas y Banik, 2024).

Desafíos y Perspectivas Futuras

A pesar del prometedor panorama, la adopción generalizada del ML en la agricultura enfrenta desafíos significativos. La escasez de datos de calidad y etiquetados, especialmente en

ciertas regiones, limita la eficacia de los modelos de ML (Reddy et al., 2025; Benos et al., 2021). Los costos de implementación asociados con la adquisición de sensores, drones y otros dispositivos necesarios para la recopilación de datos pueden ser una barrera considerable para los pequeños agricultores (Reddy et al., 2025; Biswas y Banik, 2024). La falta de expertise técnico entre muchos agricultores para implementar y utilizar modelos de ML requiere iniciativas de capacitación y asistencia para facilitar su adopción (Reddy et al., 2025; Benos et al., 2021). Finalmente, la interpretabilidad de algunos modelos de ML, como las CNN, puede ser limitada debido a su complejidad, lo que dificulta la comprensión de los resultados por parte de los usuarios finales (Reddy et al., 2025; Benos et al., 2021).

El futuro del ML en la agricultura se vislumbra en la integración con el Internet de las Cosas *IoT*, lo que permitirá una monitorización en tiempo real de los cultivos y suelos, facilitando la toma de decisiones más precisas y oportunas (Jhajharia y Mathur, 2022; Biswas y Banik, 2024). Se espera un mayor desarrollo de modelos más interpretativos que proporcionen explicaciones claras de sus predicciones, aumentando así la confianza entre los agricultores (Reddy et al., 2025; Benos et al., 2021). La expansión de datos abiertos y el fomento de la colaboración entre investigadores, agricultores y entidades gubernamentales serán cruciales para superar los desafíos actuales e impulsar la adopción del ML en la agricultura a nivel global (Yadav et al., 2024). Abordar estas barreras a través de la colaboración entre investigadores, agricultores y formuladores de políticas es esencial para la adopción generalizada del ML en la agricultura (Yadav et al., 2024).

Marco Teórico

La agricultura en Boyacá, al igual que en muchas regiones agrícolas del mundo, enfrenta el desafío de optimizar sus procesos para incrementar la productividad y la sostenibilidad.

Tecnologías emergentes como el machine learning (ML) Ofrecen herramientas poderosas para abordar estas necesidades al analizar grandes volúmenes de datos y proporcionar recomendaciones específicas y basadas en evidencia, es decir basadas en datos.

Productividad Agrícola

La agricultura en Boyacá es un sector clave para su economía regional. Según el informe de la (Gobernación de Boyacá, 2023), el departamento ocupa un lugar importante en la producción agrícola nacional gracias a su clima favorable y diversidad de cultivos como papa, maíz, frutas y hortalizas. Sin embargo, aún existe un potencial significativo para mejorar la eficiencia productiva mediante la modernización de las prácticas agrícolas

La productividad agrícola es un indicador clave que refleja la eficiencia de un sistema agrícola en la generación de productos en relación con los insumos utilizados, como tierra, agua, fertilizantes y mano de obra. (FAO, 2019) señala que la mejora en la productividad es fundamental para garantizar la seguridad alimentaria, especialmente en un contexto de creciente demanda global de alimentos debido al aumento de la población y cambios en los patrones de consumo. Además, la productividad agrícola está influenciada por factores como la tecnología, las prácticas de manejo agronómico y las condiciones ambientales.

Machine Learning en Agricultura

El machine learning (ML) se ha convertido en una herramienta poderosa en la agricultura moderna, permitiendo la automatización de procesos y la toma de decisiones basadas en datos. Las técnicas de ML, como el aprendizaje supervisado y no supervisado, se utilizan para analizar grandes volúmenes de datos provenientes de sensores, imágenes satelitales y registros históricos (Azeem Ayaz Mirani, 2021). Esto permite predecir el rendimiento de los cultivos, identificar enfermedades y plagas, y optimizar el uso de recursos La implementación de estas tecnologías

puede llevar a una agricultura de precisión, donde las decisiones se toman a nivel de parcela, mejorando la eficiencia y sostenibilidad.

Por otra parte, el uso de ML puede promover la sostenibilidad al reducir el impacto ambiental de la agricultura. Por ejemplo, se puede evitar el uso excesivo de fertilizantes al ajustar las dosis según las necesidades específicas de cada cultivo, lo cual es especialmente relevante en una región como Boyacá, donde el manejo eficiente de recursos es una prioridad.

Análisis de Datos y Toma de Decisiones

El análisis de datos en la agricultura implica la recopilación y procesamiento de información relevante para la toma de decisiones informadas. Las tecnologías de *big data* y *análisis predictivo* permiten a los agricultores identificar patrones y tendencias que pueden influir en la producción. Por ejemplo, el análisis de datos climáticos puede ayudar a anticipar sequías o inundaciones, permitiendo a los agricultores ajustar sus prácticas de cultivo en consecuencia (Zhang, 2019). La capacidad de tomar decisiones basadas en datos puede mejorar significativamente la resiliencia de los sistemas agrícolas ante cambios ambientales y económicos.

Sostenibilidad Agrícola

La sostenibilidad en la agricultura se refiere a prácticas que buscan equilibrar la producción de alimentos con la conservación de los recursos naturales y la protección del medio ambiente. Según (Tilman, 2011), la agricultura sostenible no solo se centra en aumentar la producción, sino también en minimizar el impacto ambiental, promoviendo la biodiversidad y el uso eficiente de los recursos. La integración de tecnologías como el machine learning puede facilitar la adopción de prácticas sostenibles, al proporcionar información precisa sobre el uso de insumos y la gestión de cultivos.

Contexto Agroclimático de Boyacá

Boyacá es un departamento de Colombia caracterizado por su diversidad agroclimática, lo que influye en la producción agrícola. Las condiciones climáticas, como la temperatura, la precipitación y la altitud, varían considerablemente dentro de la región, afectando la viabilidad de diferentes cultivos. Comprender estas particularidades es esencial para aplicar técnicas de machine learning que se adapten a las condiciones locales (IDEAM, 2020). La recopilación de datos agroclimáticos específicos permitirá desarrollar modelos predictivos que consideren las variaciones regionales y ayuden a optimizar la producción agrícola.

Desafíos y Oportunidades

A pesar del potencial del machine learning para transformar la agricultura, existen desafíos que deben abordarse, como la disponibilidad de datos de calidad, la capacitación de los agricultores en el uso de tecnologías, y la integración de sistemas de información. Sin embargo, las oportunidades son significativas. La implementación de modelos de ML puede llevar a una mayor eficiencia en la producción, reducción de costos y una mejor gestión de los recursos, contribuyendo a la sostenibilidad y resiliencia del sector agrícola en Boyacá.

Metodología

Este proyecto se sigue el marco metodológico *CRISP-DM* Cross (Industry Standard Process for Data Mining), el cual es un enfoque ampliamente utilizado para la minería de datos y el aprendizaje automático en proyectos de ciencia de datos. Este marco es especialmente adecuado para nuestro proyecto, que busca aplicar técnicas de Machine Learning (ML) para optimizar la productividad agrícola en Boyacá a través de datos agrícolas libres o públicos. Esta metodología fue establecida a finales de 1990 en donde se establece las siguientes etapas.

Etapas

1. Comprensión comercial: identificación de los objetivos y requisitos del proyecto.
2. Comprensión de los datos: recopilación y exploración de datos para descubrir información.
3. Preparación de datos: limpieza y transformación de datos para el modelado.
4. Modelado: selección y aplicación de varias técnicas de modelado.
5. Evaluación: evaluar la eficacia del modelo en comparación con los objetivos empresariales.
6. Implementación: implementación del modelo en un entorno de producción.

Esta metodología permite la flexibilidad de fuentes de información para la primera fase, se recopilan datos de fuentes públicas, como *IDEAM* (Instituto de Hidrología, Meteorología y Estudios Ambientales) y Ministerio de Agricultura, los cuales fueron procesados mediante técnicas de minería de datos para garantizar su calidad. Posteriormente, se entrenarán modelos de Machine Learning para identificar patrones en la productividad agrícola.

Comprensión Comercial

Este proyecto busca evaluar el impacto de técnicas de Machine Learning (ML) en la optimización de la productividad agrícola en el departamento de Boyacá, esta es región caracterizada por su diversidad agroclimática y desafíos asociados a la variabilidad del suelo y el clima. A pesar de la creciente disponibilidad de datos agrícolas, muchos productores (campesinos / Empresas) aún dependen de métodos tradicionales en la toma de decisiones, lo que limita su capacidad para adaptarse a condiciones cambiantes y mejorar su rendimiento.

El objetivo general del proyecto es evaluar la efectividad de los modelos de ML en la predicción y en la optimización de la productividad agrícola. Para esto, se definió los siguientes objetivos:

Implementar modelos de machine learning para predecir la productividad agrícola y detectar factores limitantes, como problemas de suelo o riesgos climáticos, considerando condiciones agroclimáticas de Boyacá.

Interpretar los resultados del sistema de optimización y ajustar las recomendaciones basadas en las condiciones específicas de cada zona agrícola de Boyacá.

Analizar el impacto de las recomendaciones generadas por los modelos en la toma de decisiones de los agricultores y su efectividad en la mejora de la productividad agrícola.

Comprensión de los Datos

En el proceso de recopilación de datos, luego de revisar fuentes bibliográficas se optó por trabajar con datos asociados a datos Agriculturas y Desarrollo Rural, datos de redes Meteorológicas, como temperatura mínima del aire, Velocidad el Viento y precipitaciones. Cuyas fuentes fueron IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales)

mediante Datos abiertos Colombia y Ministerio de Agricultura dependencia Agronet, continuación se presenta los metadatos de los datos crudos empleados en este proyecto.

Tabla 1

Metadatos Evaluaciones Agropecuarias

Atributo	Descripción
Descripción	Base histórica de los años 2019 a 2023, relacionada con la producción agrícola nacional. Además, brinda a los diversos actores, con especial énfasis en los productores, información agrícola regional y nacional que fortalece procesos productivos y de comercialización.
Información General	
Fecha de Actualización	28 de junio de 2024
Última Actualización de los Datos	24 de noviembre de 2019
Última Actualización de Metadatos	26 de marzo de 2024
Fecha de Creación	24 de noviembre de 2019
Vistas	106K
Descargas	24.8K
Información de la Entidad	
Fuente de Datos	Ministerio de Agricultura y Desarrollo Rural
Propietario del Conjunto de Datos	Ministerio de Agricultura
Nombre de la Entidad	Ministerio de Agricultura y Desarrollo Rural
Orden	Nacional
Área o Dependencia	Agricultura y Desarrollo Social Agronet
Información de Datos	
Idioma	Español
Cobertura Geográfica	Nacional
Frecuencia de Actualización	Anual

Atributo	Descripción
Fecha de Emisión	25 de noviembre de 2019
Categoría	Agricultura y Desarrollo Rural
URL del Conjunto de Datas	Evaluaciones Agropecuarias Municipales EVA

Tabla 2

Metadatos Evaluaciones Agropecuarias

Atributo	Descripción
Descripción	Temperatura mínima del aire a 2 metros. Datos crudos provenientes de sensores de estaciones automáticas del IDEAM y terceros. Se ofrecen como datos abiertos para gestión de riesgos, sin validación oficial, posibles errores e inconsistencias. Su uso e interpretación son responsabilidad del usuario, sin valor jurídico ni justificación posterior del IDEAM.
Información General	
Fecha de Actualización	2 de abril de 2025
Última actualización de los datos	2 de abril de 2025
Última actualización de metadatos	5 de febrero de 2025
Fecha de creación	27 de agosto de 2019
Vistas	2.580
Descargas	1.127
Información de la Entidad	
Suministró los datos	Instituto de Hidrología, Meteorología y Estudios Ambientales
Propietario del conjunto de datos	Oficina de Informática IDEAM
Departamento	Bogotá D.C.
Municipio	Bogotá D.C.

Atributo	Descripción
Nombre de la entidad	Instituto de Hidrología, Meteorología y Estudios Ambientales
Orden	Nacional
Sector	Agricultura y Desarrollo Social
Área o dependencia	GESTIÓN DE DATOS Y RED METEOROLÓGICA
Información de Datos	
Idioma	Español
Cobertura geográfica	Nacional
Frecuencia de actualización	Mensual
Fecha de emisión	27 de agosto do 2019
Categoría	Ambiente y Desarrollo Sostenible
Etiquetas	Este conjunto de datos no tiene ninguna etiqueta
Licencia y atribución	
Licencia	Public Domain
Enlace de la fuente	IDEAM

Tabla 3

Metadatos de la Velocidad del Viento

Atributo	Descripción
Descripción	Velocidad del viento cada 10 minutos. Datos crudos de sensores en estaciones automáticas del IDEAM y terceros. Velocidad del viento cada 10 minutos. Datos crudos de sensores en estaciones automáticas del IDEAM y terceros. Se publican como datos abiertos para gestión de riesgos, sin validación oficial, con posibles errores e inconsistencias. Su uso e interpretación son responsabilidad del usuario, sin valor jurídico ni justificación posterior del IDEAM.
Información general	
Fecha de actualización	2 de abril de 2025

Atributo	Descripción
Última actualización de los datos	2 de abril de 2025
Última Actualización de metadatos	5 de febrero de 2025
Fecha de creación	27 de agosto de 2019
Vistas	14.600
Descargas	3.582
Información de la entidad	
Suministró los datos	Instituto de Hidrología, Meteorología y Estudios Ambientales
Propietario del conjunto de datos	Oficina de Informática IDEAM
Departamento	Bogotá D.C.
Municipio	Bogotá D.C.
Nombre de la entidad	Instituto de Hidrología, Meteorología y Estudios Ambientales
Orden	Nacional
Sector	Agricultura y Desarrollo Social
Área o dependencia	GESTIÓN DE DATOS Y RED METEOROLÓGICA
Información de datos	
Idioma	Español
Cobertura geográfica	Nacional
Frecuencia de actualización	Mensual
Fecha de emisión	27 de agosto do 2019
Categoría	Ambiente y Desarrollo Sostenible
Etiquetas	Este conjunto de datos no tiene ninguna etiqueta
Licencia y atribución	
Licencia	Public Domain
Enlace de la fuente	IDEAM

Tabla 4

Metadatos del Conjunto Agrícola y Fertilizantes

Atributo	Descripción
Descripción	Productos agroquímicos
Información general	
Fecha de actualización	23 de marzo de 2024
Última actualización de los datos	17 de julio de 2023
Última actualización de Metadatos	26 de marzo de 2024
Fecha de creación	20 de diciembre de 2016
Vistas	2.224
Descargas	406
Información de la entidad	
Suministró los datos	VECOL S.A.
Propietario del conjunto de datos	VECOL
Departamento	Bogotá D.C.
Municipio	Bogotá D.C.
Nombre de la entidad	Empresa Colombiana de Productos Veterinarios VECOL S.A
Orden	Nacional
Sector	Agricultura y Desarrollo Social
Área o dependencia	Gerencia Comercial
Información de datos	
Idioma	Español
Cobertura geográfica	Nacional
Frecuencia de actualización	Mensual
Fecha de emisión	27 de agosto do 2019
Categoría	Ambiente y Desarrollo Sostenible
Etiquetas	Este conjunto de datos no tiene ninguna etiqueta
Licencia y atribución	

Atributo	Descripción
Licencia	Public Domain

Tabla 5

Metadatos del Conjunto de Datos sobre Precipitación

Atributo	Descripción
Descripción	Metadatos del Conjunto de Datos sobre Precipitación
Información general	
Fecha de Actualización	2 de abril de 2025
Última actualización de los datos	2 de abril de 2025
Última actualización de metadatos	5 de febrero de 2025
Fecha de creación	27 de agosto de 2019
Vistas	22,8k
Descargas	7.799
Información de la entidad	
Suministró los datos	Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM)
Propietario del conjunto de datos	Oficina de Informática IDEAM
Departamento	Bogotá D.C.
Municipio	Bogotá D.C.
Nombre de la entidad	Instituto de Hidrología, Meteorología y Estudios Ambientales
Orden	Nacional
Sector	Ambiente y Desarrollo Sostenible
Área o dependencia	Gestión de Datos y Red Meteorológica
Información de datos	
Idioma	Español
Cobertura geográfica	Nacional
Frecuencia de actualización	Mensual

Atributo	Descripción
Fecha de emisión	27 de agosto do 2019
Categoría	Ambiente y Desarrollo Sostenible
Etiquetas	Estación, Precipitación, IDEAM
Licencia y atribución	
Licencia	Public Domain
Enlace de la fuente	IDEAM

La base principal en la que se centró este proyecto corresponde a información contenida de la página https://agronet.gov.co de datos libre de la entidad pública *AgroNet*, la cual cuenta con herramientas estadísticas de sectores como: Agrícola, Pecuario, Precios Comercio, Créditos, insumos entre otros. Para el caso de estudio se decidió trabajar con datos Agrícolas correspondientes a Reportes de Evaluaciones Agropecuarias - EVA y Anuario Estadístico del Sector Agropecuario, ver figura 1. La base de datos seleccionada fue el reporte más actualizado a la fecha '06-2024'.

Figura 1

Evaluaciones Agropecuarias - EVA y Anuario Estadístico del Sector Agropecuario

Nota. Fuente ArgoNet página web

Dado que nuestra población de interés corresponde a los municipios del departamento de Boyacá, se decidió filtrar la información por departamento. Asimismo, considerando los temas de interés, como la agricultura y la productividad, el método de escogencia del conjunto de datos se fundamentó en la fiabilidad de la entidad proveedora en este caso el (Ministerio de Agricultura y Desarrollo Rural) y en que el tiempo de actualización de los datos no superara los dos años.

Tras la implementación de los métodos de minería de datos pertinentes, se logró consolidar la base de datos resultante, que posteriormente sirvió como referencia para elaborar el diccionario de datos. Ver tabla 7.

Durante el proceso de limpieza de dato, transformaciones e integraciones, y la selección de variables relevantes en la investigación se documentó su proceso en un repositorio de *git* para su visualización y control de versiones Gihub. Luego de revisar fuentes bibliográficas de evaluaciones Agricultura y Desarrollo Rural, se integró información proveniente de redes meteorológicas, tales como temperatura mínima del aire, velocidad del viento y precipitaciones. estas fuentes fueron consultadas en la base del *IDEAM* Instituto de Hidrología, Meteorología y Estudios Ambientales)

Tomando los conjuntos de datos de precipitaciones y temperatura mínima, se consolidaron ambas medidas en una misma base mediante una llave por municipio y día, con el fin de unificar la información a nivel diario. Esto debido a que los registros meteorológicos estaban distribuidos a lo largo del día en distintos momentos.

Se planteó entonces calcular el promedio diario de temperatura mínima y la suma de las precipitaciones diarias, generando como resultado un dataset con las siguientes variables: ‘Fecha’, ‘Municipio’, ‘Precipitacion_Total’ y ‘Temp_Min_Promedio’. ver tabla 6

Tabla 6

Diccionario de datos - Consolidado Meteorológico Diario

Titulo	Consolidado Diario de Precipitaciones y Temperaturas Mínimas por Municipio - Boyacá		
Url	IDEAM	Datos Abiertos Colombia	
Nombre del datase	*clima_diario.csv*	Separador: ‘,‘	
Filas, Columnas	(35376, 6)	Cantidad de registros: 35376	
Documentación			
Atributo	Ejemplo	Tipo	Descripción
Fecha	2025-04-02	Date	Fecha de consolidación diaria de datos meteorológicos.
Municipio	PAIPA	String	Nombre del municipio donde se tomaron las mediciones.
Precipitacion_Tota1	2.3	Float	Suma total de las precipitaciones (en mm) registradas durante el día en el municipio
Temp_Min_Promedio	14.84	Float	Promedio de las temperaturas mínimas registradas durante el día en el municipio (°C).

Nota. Diccionario de datos fuentes agroclimáticas.

Tabla 7

Diccionario de Datos de Evaluaciones Agropecuarias

Titulo	Evaluaciones Agropecuarias Municipales - 2019 A 2023		
Url	agronet.gov.co	Boyacá	
Nombre del datase	*EvaluacionAgro.csv*	Separador: ‘,‘	
Filas, Columnas	(12265, 18)	Cantidad de registros: 12265	
Documentación			
Atributo	Ejemplo	Tipo	Descripción
codDaneDpto	15	String	Muestra el código del Departamento
Dpto	Boyacá	Int	Muestra el nombre del Departamento

Titulo	Evaluaciones Agropecuarias Municipales - 2019 A 2023		
codDaneMunicipio	15022	String	Muestra el código del Municipio
Municipio	Almeida	String	Muestra el nombre del Municipio
desagregacionCultivo	Tomate Invernadero	String	Se refiere a la subdivisión o separación de los datos agrícolas en categorías más específicas o detalladas.
Cultivo	Tomate	String	Muestra el nombre del cultivo
cicloDelCu1tivo	Transitorio	String	Indica el periodo de tiempo desde la siembra hasta la cosecha de un cultivo.
grupoCultivo	Hortalizas	String	Muestra categorías más amplias o generales de cultivos que comparten características comunes.
SubGrupo	Hortalizas de Fruto	String	Muestra categorías más específicas o detalladas dentro de un grupo de cultivos más amplio.
Anio	2019	Int	Muestra el año del cultivo
periodo	2019A	String	Indica el periodo del cultivo el cual se encuentran años que finalizan con la letra A o la letra B, donde los años que finalizan con la letra A se refieren al primer semestre y los años que finalizan con la letra B se refieren al segundo semestre.
areaSembradaHa	2	Float	Representa la extensión de tierra (en hectáreas) en la que se ha sembrado un cultivo específico durante un período de tiempo determinado.
areaCosechadaHa	2	Float	

Titulo	Evaluaciones Agropecuarias Municipales - 2019 A 2023		
produccionTon	136	Float	Representa la extensión de tierra (en hectáreas) que ha sido efectivamente cosechada para recolectar el cultivo.
rendimientoTonHa	68	Float	Indica la cantidad total de producto agrícola (en toneladas) que se ha obtenido de la cosecha. Representa la medida de la eficiencia de la producción agrícola y se expresa como la cantidad de producto agrícola (en toneladas) obtenida por unidad de área de tierra sembrada (hectáreas).
codCultivo	1052902	Int	Es decir, es la producción por hectárea de tierra sembrada.
nombreCientifico	Lycopersicon esculentum	String	Muestra el código del cultivo
Cultivo	En fresco	String	Muestra el nombre científico del cultivo
estadoFisicoCultivo			Describe el estado físico del cultivo en el momento de la medición o recolección.

Nota. Diccionario de datos de Evaluaciones agropecuarias de los departamentos de Boyacá, Fecha de actualización: 28-06-2024

Por otro lado, en el proceso de limpieza de datos y creación de variables útiles, se utilizó la base de datos de Evaluaciones Agropecuarias. Uno de los primeros pasos fue el replanteamiento de la variable periodo, ya que, dada la naturaleza temporal del análisis, era necesario interpretar adecuadamente los registros. de la siguiente manera:

A → Primer semestre

B → Segundo semestre

C → Año completo

Este cambio permitió identificar con mayor claridad cuáles fueron los periodos más productivos en cada año.

Posteriormente, al realizar diagramas de distribución para las variables numéricas (ver Figura 3. Se observó que la mayoría de los datos estaban sesgados hacia la izquierda, con algunos valores extremos alejados de la media. Ante esta situación, se procedió a validar la coherencia lógica entre variables clave.

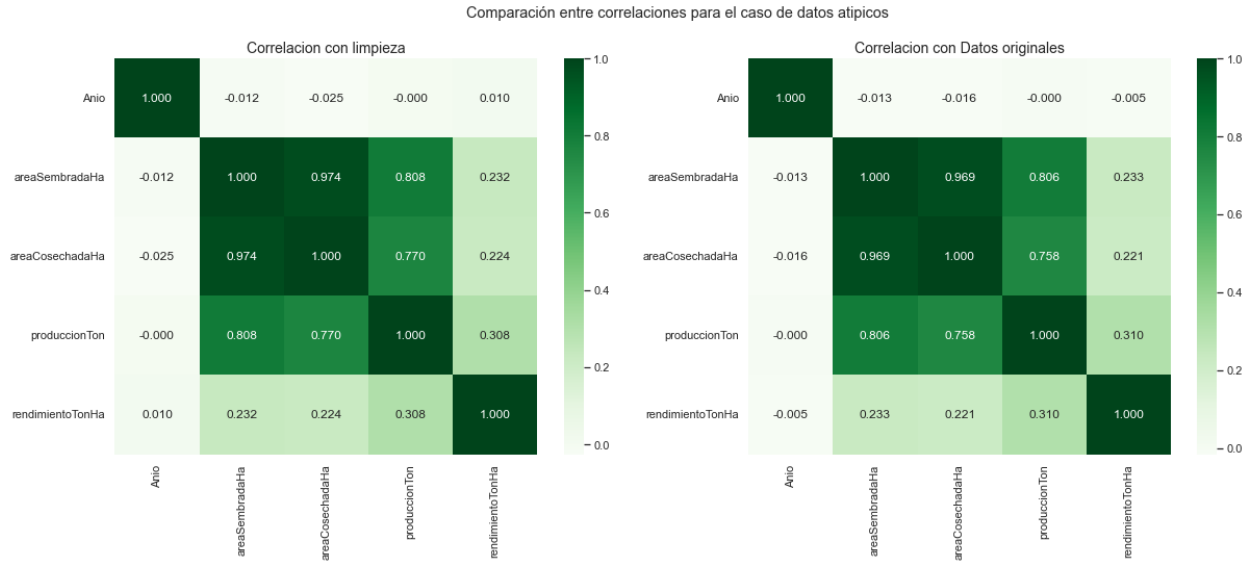
Se encontró una relación esperada entre las variables *areaSembradaHa* y *areaCosechadaHa*. Según sus definiciones, el área sembrada representa la superficie total dedicada a un cultivo, mientras que el área cosechada indica la parte efectivamente productiva. Por tanto, se debe cumplir que:

$$areaSembrada \geq areaCosechada$$

Se eliminaron los registros que no cumplían esta condición, lo que redujo el tamaño del conjunto de datos de 12,265 a 11,511 registros, lo que equivale a una disminución del 6.15\% aproximadamente. Esta depuración mejoró la consistencia interna del dataset y fortaleció las correlaciones entre variables, como puede observarse en la matriz de correlación, ver Figura 2.

Figura 2

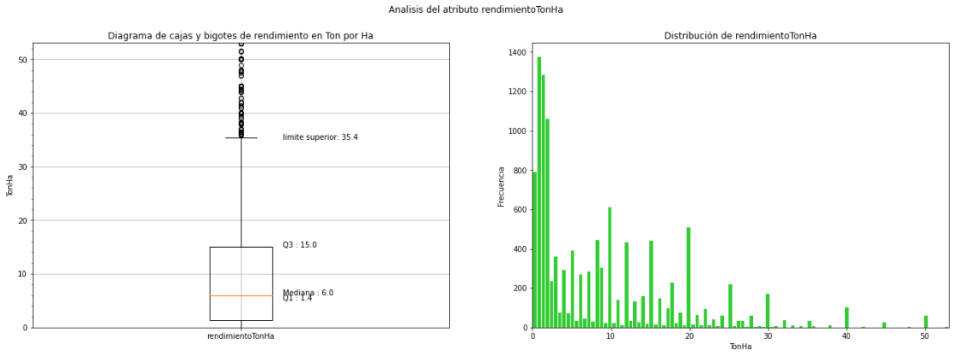
Distribución de la Variable produccionTon



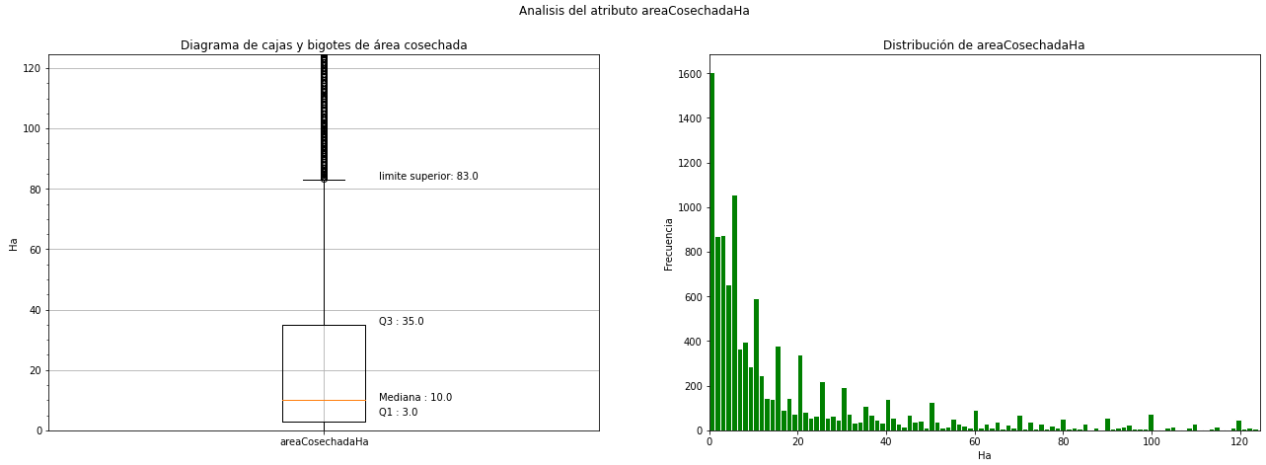
Nota. A la izquierda se muestran las correlaciones luego de aplicar filtros para asegurar consistencia entre las variables areaSembradaHa y areaCosechadaHa, eliminando valores atípicos. A la derecha se presentan las correlaciones originales, donde se observan pequeñas diferencias especialmente en la relación entre produccionTon y las variables de área

Figura 3

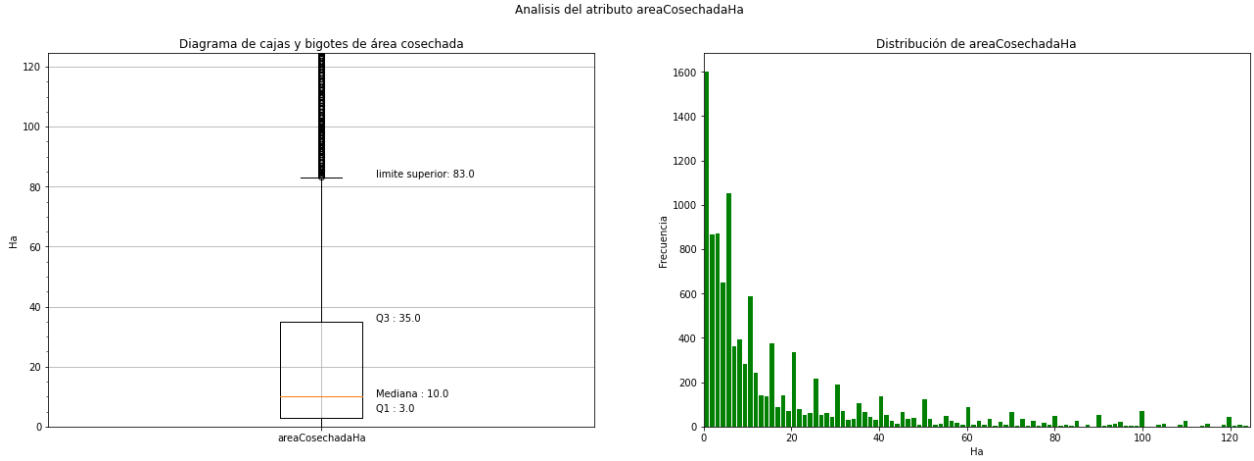
Distribuciones de Variables Numéricas Relevantes para el Análisis de Producción Agrícola



(a) Distribución de la variable rendimientoTonHa.



(b) Distribución de la variable areaCosechadaHa.



(c) Distribución de la variable produccionTon.

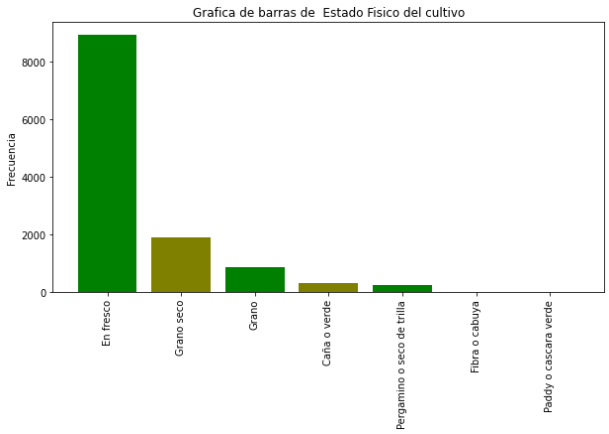
Modelado

En la elaboración del modelo se optó por implementar la técnica de codificación *One-Hot Encoding* con el fin de integrar variables categóricas consideradas útiles para el modelo. La selección de estas variables se basó en el análisis exploratorio de datos (EDA), el cual permitió identificar aquellas con una distribución clara y pocos valores únicos, lo que evita una alta dimensionalidad en el conjunto de datos resultante.

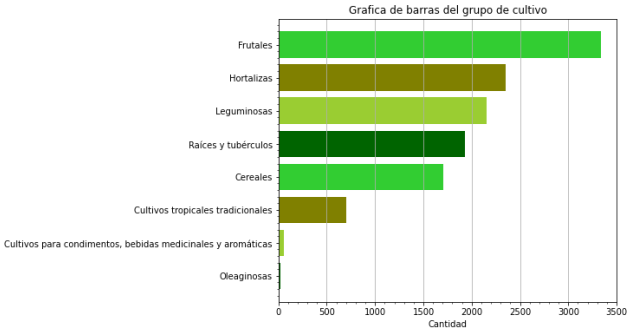
Las variables seleccionadas fueron: *Periodo*, *cicloDelCultivo*, *grupoCultivo* y *estadoFisicoCultivo*. A continuación, se presentan los diagramas de barras correspondientes a cada una de estas variables:

Figura 4

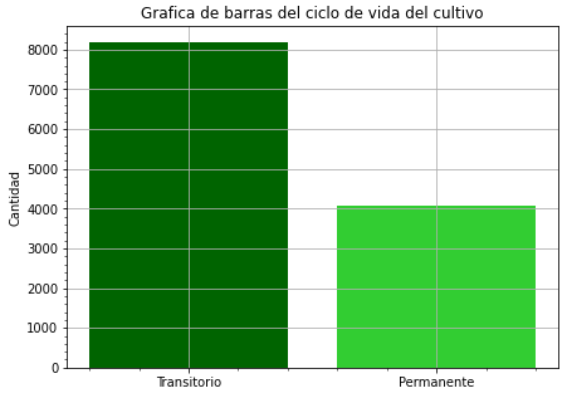
Distribución de las Variables Categóricas



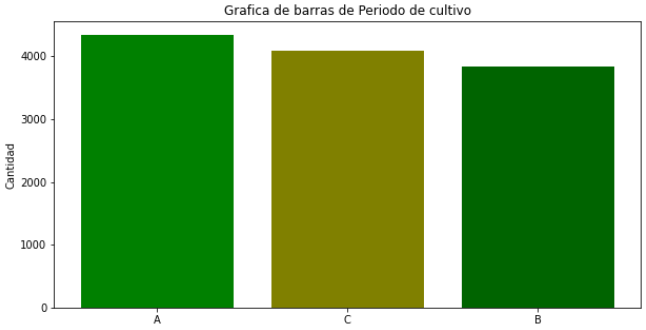
(a) Distribución de estadoFisicoCultivo



(b) Distribución de grupoCultivo.



(c) Distribución de cicloDelCultivo.



(d) Distribución de Periodo.

Nota. Distribución de las variables categóricas utilizadas en el modelo para la técnica de codificación One-Hot.

A las variables categóricas seleccionadas se les aplicó la función *get_dummies* de la biblioteca pandas, con el objetivo de convertirlas en variables numéricas mediante codificación

One-Hot. Esto permitió descomponer cada variable en sus respectivas subcategorías, generando un conjunto de aproximadamente 25 variables. Posteriormente, se aplicó un proceso de estandarización utilizando la función *StandardScaler* () de sklearn, la cual transforma los datos para que tengan una media aproximada de cero ($\mu \approx 0$) y una desviación estándar igual a uno ($\sigma = 1$), facilitando así la comparación entre variables con distintas escalas.

Dado el aumento en la dimensionalidad del conjunto de datos, se implementó una técnica de reducción de dimensiones mediante Análisis de Componentes Principales (PCA). Esta técnica permitió reducir el número de variables a 11 componentes principales, los cuales conservan el 90% de la varianza explicada del conjunto original, asegurando así una representación adecuada de la información sin comprometer el rendimiento del modelo.

División del Conjunto de Datos y Evaluación del Modelo

Para la etapa de modelado se trabajó con algoritmos de regresión, específicamente árboles de decisión, máquinas de vectores de soporte (*SVM*) y bosques aleatorios (*Random Forest*). El objetivo principal fue predecir el rendimiento agrícola, definido como la razón entre la producción total y el área cosechada. Esta variable fue elegida como variable objetivo debido a que representa de manera más eficiente el concepto de productividad agrícola, alineándose con el propósito de este estudio: implementar modelos de machine learning para predecir la productividad agrícola y detectar factores limitantes.

El conjunto de datos fue dividido en datos de entrenamiento y prueba utilizando validación cruzada. Para la evaluación del desempeño de los modelos se utilizaron las siguientes métricas:

Coeficiente de determinación (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}\left(y_{test} - \hat{y}_{pred}\right)^2}{\sum_{i=1}(y_{test} - \bar{y}_{test})^2}$$

Error Cuadrático Medio (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{test} - \hat{y}_{pred}\right)^2$$

Raíz del Error Cuadrático Medio (RMSE):

$$MSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{test} - \hat{y}_{pred}\right)^2}$$

Estas métricas permiten cuantificar la calidad de las predicciones, siendo deseable un valor de R^2 cercano a 1 (lo que indica una alta capacidad explicativa del modelo) y un valor de *RMSE* lo más bajo posible, que refleje un menor error promedio. Además, se consideró como intervalo de confianza de predicción el rango.

$$y = \hat{y}_{pred} \pm RMSE$$

lo que permite interpretar si las predicciones se ajustan adecuadamente a los valores reales observados.

Evaluación

Para el proceso de evaluación de los modelos de regresión Árboles de Decisión, Máquinas de Vectores de Soporte (SVM) y Bosques Aleatorios (*Random Forest*) se aplicaron técnicas de validación cruzada con cinco particiones (cv=5) para estimar la estabilidad y generalización de cada modelo. Adicionalmente, se realizó un ajuste de hiperparámetros mediante búsqueda en malla (*GridSearchCV*) con el objetivo de encontrar la configuración óptima de cada modelo, específicamente para el caso de *Random Forest*.

El modelo que presentó mejor desempeño fue el *Random Forest*, cuya configuración óptima se resume a continuación:

```
from sklearn.model_selection import GridSearchCV
```

```
# Definir los parámetros a ajustar
parametros_regresion = {
    'max_depth': (32, 16, 8),
    'n_estimators': (100, 50, 20)
}

# Instanciar el modelo base
modelo_Forest_Regressor = RandomForestRegressor(random_state=42)

# Configurar GridSearchCV
busqueda = GridSearchCV(
    estimator=modelo_Forest_Regressor,
    param_grid=parametros_regresion,
    cv=5,
    scoring='neg_mean_squared_error',
    n_jobs=-1
)

# Ajustar el modelo
busqueda.fit(X_train, y_train)

print ('Mejores parámetros encontrados:', busqueda.best_params_)

# Evaluar el modelo ajustado
mejor_modelo = busqueda.best_estimator_
y_pred_grid = mejor_modelo.predict(X_test)
r2_grid = r2_score(y_test, y_pred_grid)
mse_grid = mean_squared_error(y_test, y_pred_grid)
rmse_grid = math.sqrt(mse_grid)

print(f'Random Forest (ajustado) R^2: {r2_grid:.3f}')
print(f'Random Forest (ajustado) MSE: {mse_grid:.3f}')
print(f'Random Forest (ajustado) RMSE: {rmse_grid:.3f}')
```

En la Tabla 8 se resumen las métricas promedio obtenidas en la validación cruzada para cada modelo:

Tabla 8

Resumen de Evaluación de Modelos con Validación Cruzada

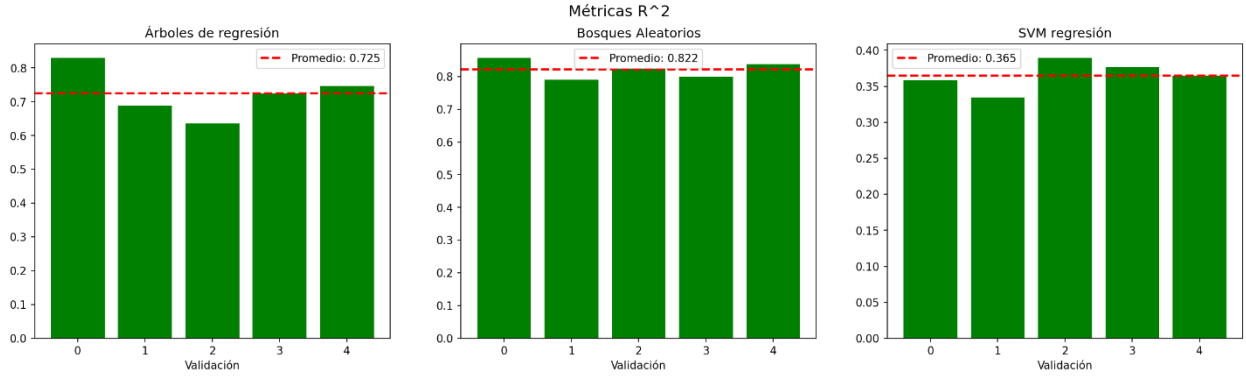
Modelo	R^2 Promedio	MSE Promedio	RMSE Promedio
Árbol de Regresión	0.725	79.49	8.81

Modelo	R^2 Promedio	MSE Promedio	RMSE Promedio
Bosque Aleatorio	0.822	50.81	7.11
SVM	0.365	181.78	13.46

Además, la Figura 5 ilustra comparativamente el desempeño de los modelos con respecto al coeficiente de determinación (R^2).

Figura 5

Comparación de Modelos Según la Métrica R^2



Nota. El modelo de Random Forest presenta un comportamiento más consistente y estable. Según los resultados, este modelo logra explicar aproximadamente el 82% de la variabilidad de los datos.

Implementación

En esta etapa, se preparó el modelo predictivo para su uso práctico por parte de los agricultores o entidades interesadas. Para facilitar su adopción, se desarrolló un script *uso_modelo_agricola.py* y se generó una documentación detallada que guía paso a paso el proceso de predicción con datos reales.

Flujo

Este flujo de implementación incluye, Archivos requeridos para la predicción:

1. modelo_random_forest.joblib: Modelo entrenado de RandomForestRegressor.
2. scaler.joblib: Objeto StandardScaler para estandarizar los datos.
3. pca.joblib: Objeto PCA para reducción de dimensionalidad.
4. uso_modelo_agricola.py: Script de ejemplo para cargar el modelo y hacer predicciones.

Todos estos archivos deben estar en la misma carpeta, o se deben ajustar las rutas en el script.

Parámetros de Entrada Requeridos

El modelo espera los siguientes parámetros (columnas) en el mismo orden y con los mismos nombres

Tabla 9

Variables Requeridas Implementación Modelo

Columnas	
Anio	GpOleaginosas
areaSembradaHa	grupoCultivo_Raíces y tubérculos
areaCosechadaHa	Boyaca
produccionTon	EdoCañaVerde
CicloPermanente	Edofresco
CicloTransitorio	EdoFibraCabuya
GpCereales	EdoGrano
grupoCultivo_Cultivos para condimentos, bebidas medicinales y aromáticas	EdoPaddyCascaraVerde
grupoCultivo_Cultivos tropicales tradicionales	EdoSecoTrilla
GpFrutales	Periodo A
GpHorta1izas	Periodo B

Columnas	
GpLeguminosas	Periodo C

Nota. Si falta alguna de las columnas requeridas o si sus nombres no coinciden exactamente con los esperados, el modelo no podrá ejecutarse correctamente. A partir de la columna CicloPermanente, las variables se implementan como indicadores binarios: se utiliza 1 si aplica y 0 si no aplica. El detalle completo puede consultarse en el anexo 1 correspondiente.

Resultados

Modelos

En la Tabla 10 se muestran los valores obtenidos de R^2, MSE y RMSE para cada partición y cada modelo. A partir de estos resultados se calcularon los promedios generales presentados en la Figura 6, lo que permite comparar de forma clara el desempeño de cada enfoque. Estos valores resumen el comportamiento de los modelos en términos de precisión y consistencia, permitiendo identificar cuál ofrece mejores predicciones sobre los datos evaluados.

Tabla 10

Resultados por Partición (Fold) para cada Modelo

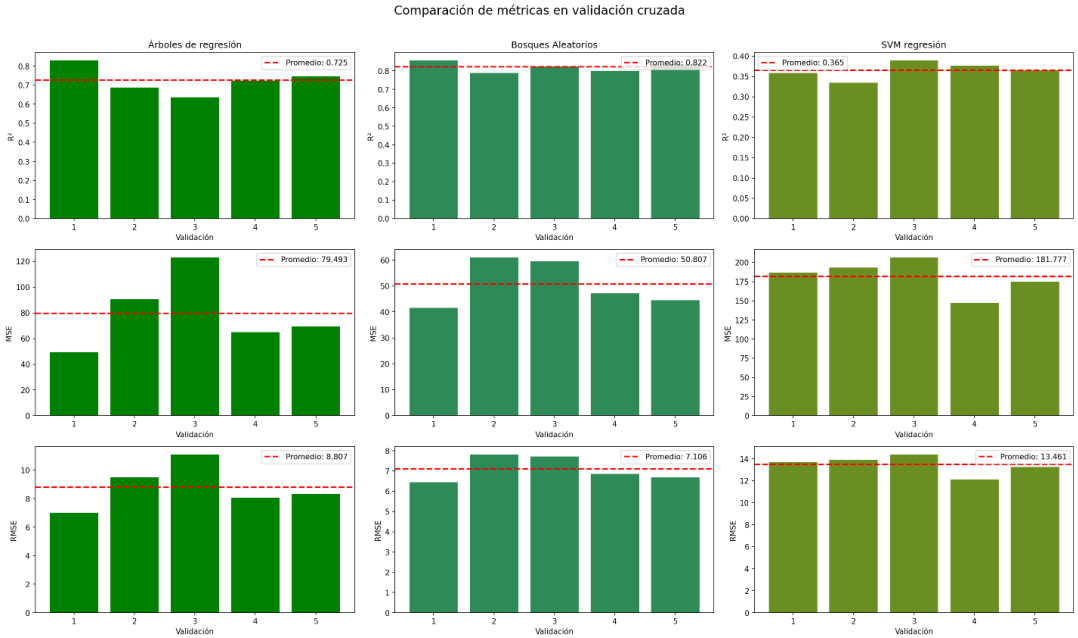
Fold	Modelo	R^2	MSE	RMSE
1	Árbol de Regresión	0.830	49.41	7.03
2	Árbol de Regresión	0.688	90.52	9.51
3	Árbol de Regresión	0.636	123.05	11.09
4	Árbol de Regresión	0.725	64.88	8.05
5	Árbol de Regresión	0.747	69.61	8.34
1	Bosque Aleatorio	0.857	41.64	6.45
2	Bosque Aleatorio	0.790	61.04	7.81
3	Bosque Aleatorio	0.824	59.69	7.73
4	Bosque Aleatorio	0.800	47.12	6.86
5	Bosque Aleatorio	0.838	44.54	6.67
1	SVM	0.358	186.84	13.67
2	SVM	O. 334	193.16	13.90
3	SVM	0.390	206.62	14.37
4	SVM	0.376	147.15	12.13
5	SVM	0.365	175.12	13.23

Los resultados muestran que el modelo de Bosque Aleatorio obtuvo el mejor rendimiento general entre los modelos evaluados. Su valor promedio de R^2 fue de 0.822, superior al del Árbol de Regresión (0.725) y al de SVM (0.365), lo que indica una mayor capacidad para explicar la variabilidad de la variable objetivo. Además, este modelo mantuvo una mayor estabilidad entre las particiones, reflejada en una menor dispersión de sus métricas.

En cuanto al error cuadrático medio (MSE), el Bosque Aleatorio también presentó el menor valor promedio (50.8), lo que sugiere predicciones más cercanas a los valores reales. Esta tendencia se confirma al observar la raíz del error cuadrático medio (RMSE), donde nuevamente obtiene el menor promedio (7.106), lo que refuerza su superioridad en términos de precisión frente a los otros enfoques

Figura 6

Desempeño por Partición en Validación Cruzada con Promedios



Nota. Cada barra representa el valor obtenido de cada métrica para cada partición (fold) durante la validación cruzada; la línea horizontal indica el promedio del modelo respectivo

Perfil Agrícola

Con el objetivo de identificar patrones temporales y caracterizar el comportamiento de los cultivos en el departamento de Boyacá, se llevó a cabo un análisis descriptivo utilizando mapas de calor y tablas comparativas. Este análisis se centró en observar la evolución anual de variables clave como la producción (en toneladas), el área sembrada y cosechada (en hectáreas), y el rendimiento (toneladas por hectárea), segmentadas por grupo de cultivo.

La visualización tipo heatmap permitió detectar fácilmente las tendencias a lo largo de los años, resaltando cuáles cultivos han mantenido una producción constante, cuáles han presentado un incremento o disminución significativa, y cómo se ha comportado su rendimiento agrícola. Esto facilita la identificación de los cultivos más productivos, los que presentan mejor desempeño por unidad de área, y aquellos con mayor superficie sembrada o cosechada.

Figura 7

Evolución de la Producción Agrícola por Grupo de Cultivo en Boyacá

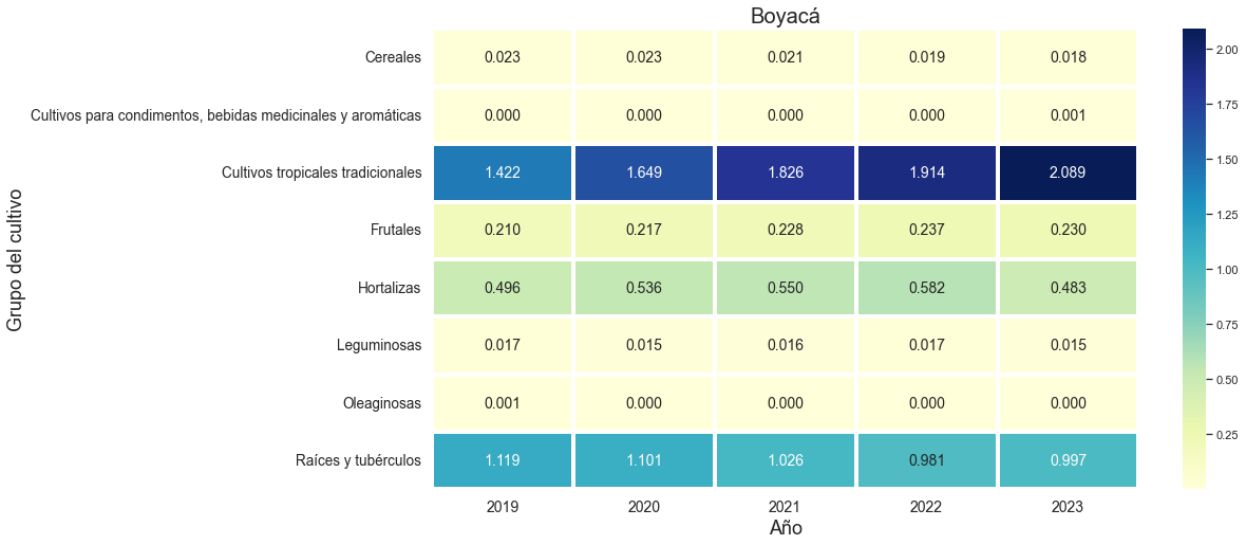


Figura 8

Evolución del Área Sembrada por Grupo de Cultivo en Boyacá

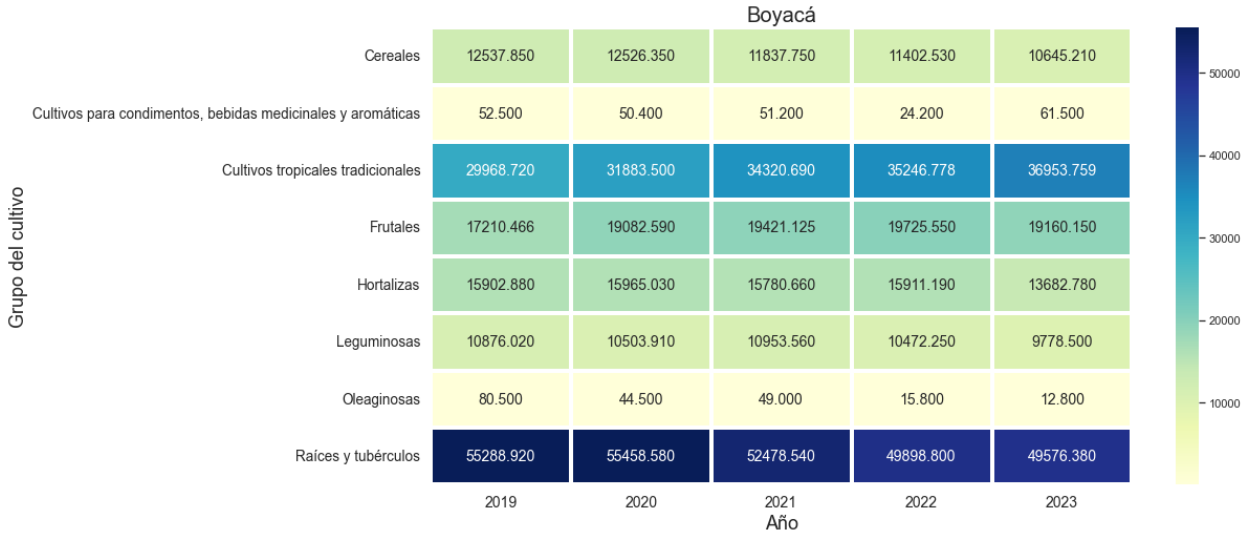


Figura 9

Evolución del Área Cosechada por Grupo de Cultivo en Boyacá

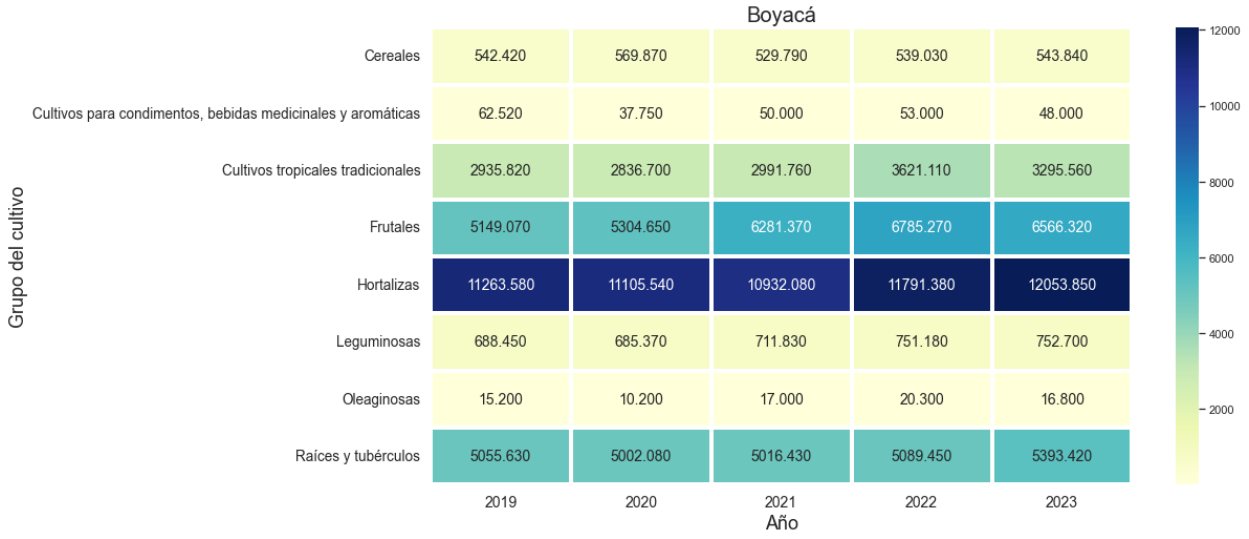
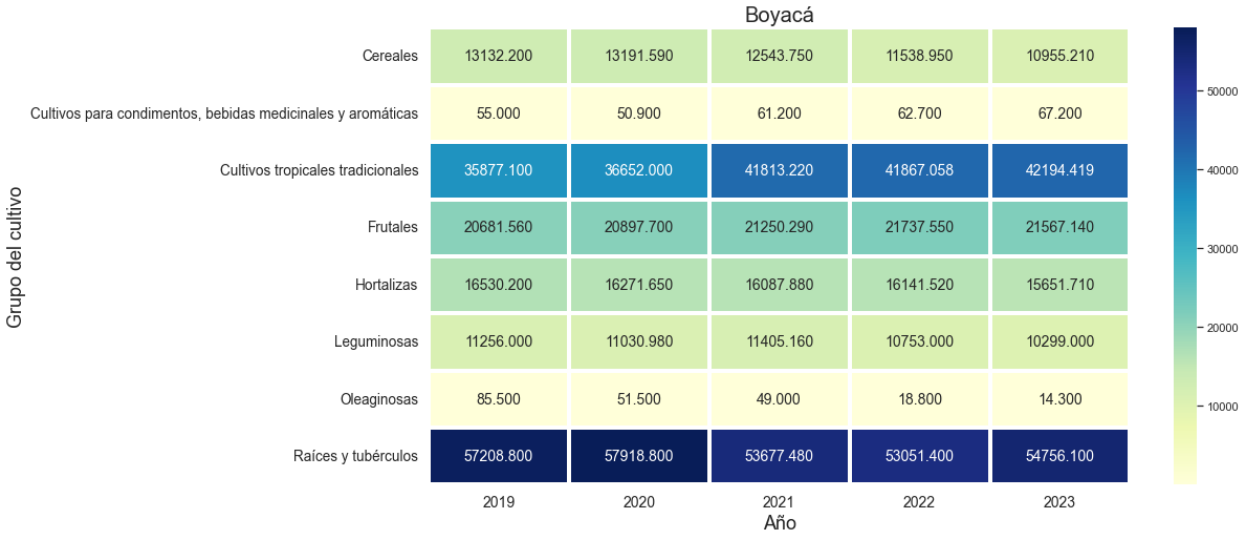


Figura 10

Evolución del Rendimiento Agrícola (ton/ha) por Grupo de Cultivo en Boyacá



A partir del análisis exploratorio de los datos de producción agrícola en el departamento de Boyacá durante el período 2019–2023, se logró identificar una serie de perfiles productivos que permiten caracterizar el comportamiento de los distintos grupos de cultivos. Esta clasificación se fundamenta en el análisis conjunto de variables clave como la producción total, el área sembrada, el área cosechada y el rendimiento (toneladas por hectárea).

Con el propósito de facilitar la interpretación de los datos y apoyar la toma de decisiones estratégicas en el sector agrícola, se definieron perfiles que agrupan los cultivos según su desempeño relativo. Entre estos se incluyen: expansivo y productivo, productivo, pero poco extendido, y limitado y de bajo rendimiento. Esta tipología permite resaltar aquellos cultivos con alto potencial para consolidar o expandir su impacto productivo, y también permite identificar aquellos que enfrentan desafíos importantes en términos de eficiencia o cobertura territorial.

Tabla 11

Perfiles Productivos Agrícolas

Perfil	Descripción
Expansivo y productivo	Cultivo ampliamente sembrado y con alto rendimiento. Es eficiente y tiene buena cobertura en el territorio.
Productivo pero poco extendido	Tiene alto rendimiento, pero poca área sembrada. Produce bien, pero no se siembra en gran escala.
Moderadamente productivo	Tiene un rendimiento intermedio y un área sembrada relevante. Cumple un rol estable, pero sin sobresalir.
Limitado y de bajo rendimiento	Se siembra poco y su rendimiento también es bajo. Requiere atención si se quiere mejorar o potenciar.
Estancado o en retroceso	Su producción o rendimiento tiende a disminuir a lo largo del tiempo. Puede indicar desinversión o problemas estructurales.
Emergente con potencial	Área o rendimiento en crecimiento. Aunque aún no es dominante, muestra señales de avance. Ideal para inversión futura.

Tabla 12

Clasificación Cualitativa de Grupos de Cultivo según Desempeño Agrícola

Grupo de cultivo	Área sembrada	Producción	Rendimiento	Perfil productivo
Hortalizas	Alta y estable	Alta	Muy alto	Expansivo y productivo
Frutales	Media	Alta	Muy alto	Productivo pero poco extendido
Cereales	Alta	Media	Bajo	Estancado o en retroceso
Cultivos tropicales tradicionales	Media	Media	Medio-alto	Moderadamente productivo
Leguminosas	Baja	Baja	Medio	

Grupo de cultivo	Área sembrada	Producción	Rendimiento	Perfil productivo
Oleaginosas	Muy baja	Muy baja	Muy bajo	Emergente con potencial
Raíces y tubérculos	Media-alta	Alta	Medio-alto	Limitado y de bajo rendimiento
Condimentos, bebidas medicinales y aromáticas	Muy baja	Muy baja	Inestable y bajo	Moderadamente productivo Limitado y de bajo rendimiento

Nota. Esta clasificación se basa en los promedios de rendimiento y tendencias observadas en los datos de 2019 a 2023.

Creación del Reporte Interactivo

Como resultado complementario al modelo predictivo, se desarrolló un reporte interactivo utilizando Power BI, el cual permite a los usuarios visualizar la información de manera clara y dinámica. Este reporte facilita la exploración de los principales cultivos, la comparación de rendimientos por municipio y la identificación de patrones temporales en la producción agrícola.

El reporte está diseñado para ser una herramienta de apoyo en la toma de decisiones, brindando a los agricultores, empresas locales o/u otros actores del sector agropecuario una forma accesible de interpretar los resultados del modelo y aplicar estrategias basadas en datos.

La construcción de este reporte consta de tres hojas: Análisis Agroclimático y Productividad por Municipio, Análisis Climático y Productividad Agrícola por Año, y Correlación Clima - Rendimiento. Todo el proceso fue automatizado mediante la ingesta de datos desde una API OData, la cual es consumida a través de un dataflow Gen2 en el servicio Microsoft Fabric. El modelo semántico del reporte conecta con los datos precargados en dicho

dataflow. El informe está disponible en el siguiente enlace: Rendimiento Agrícola, y se encuentra compartido a nivel organizacional.

Figura 11

Análisis Climático y Productividad Agrícola por Año

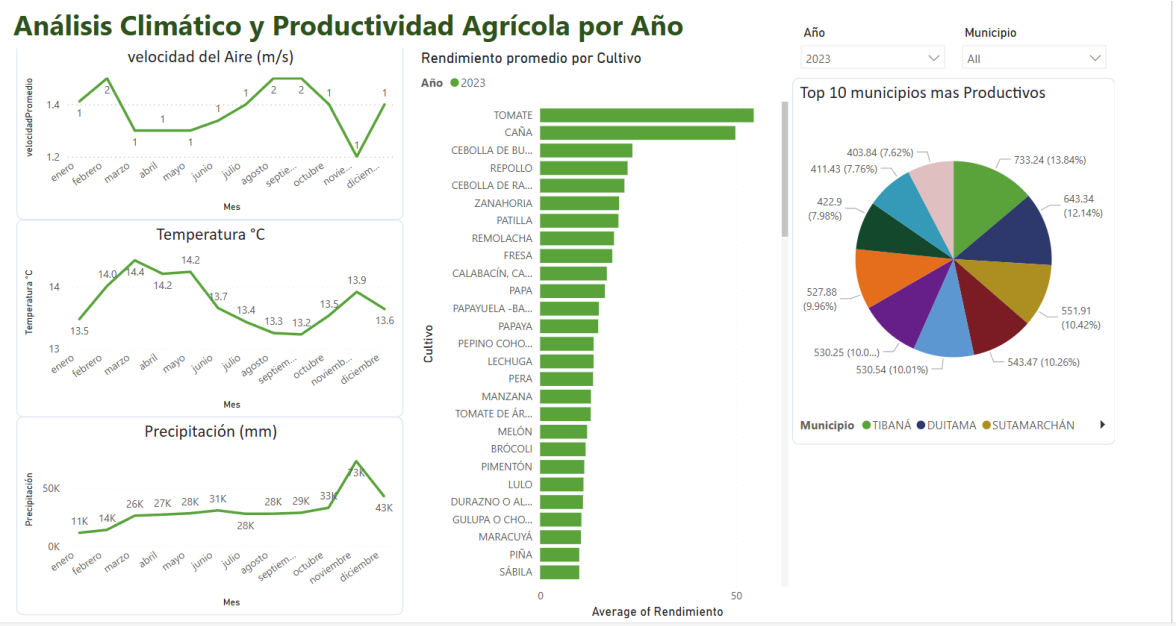
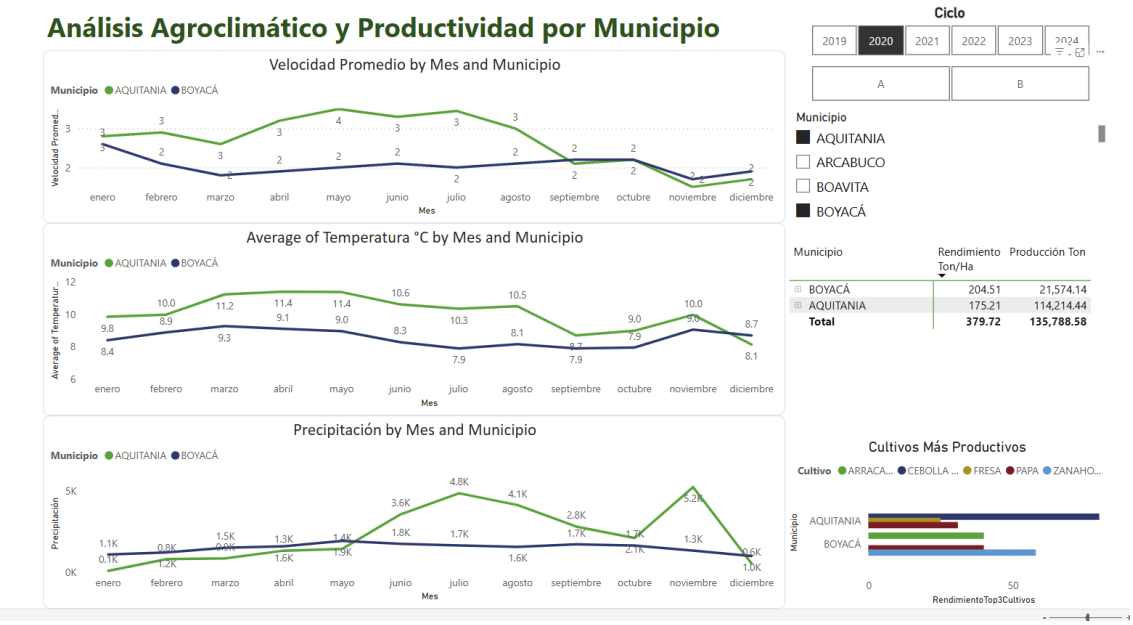


Figura 12

Análisis Agroclimático y Productividad por Municipio

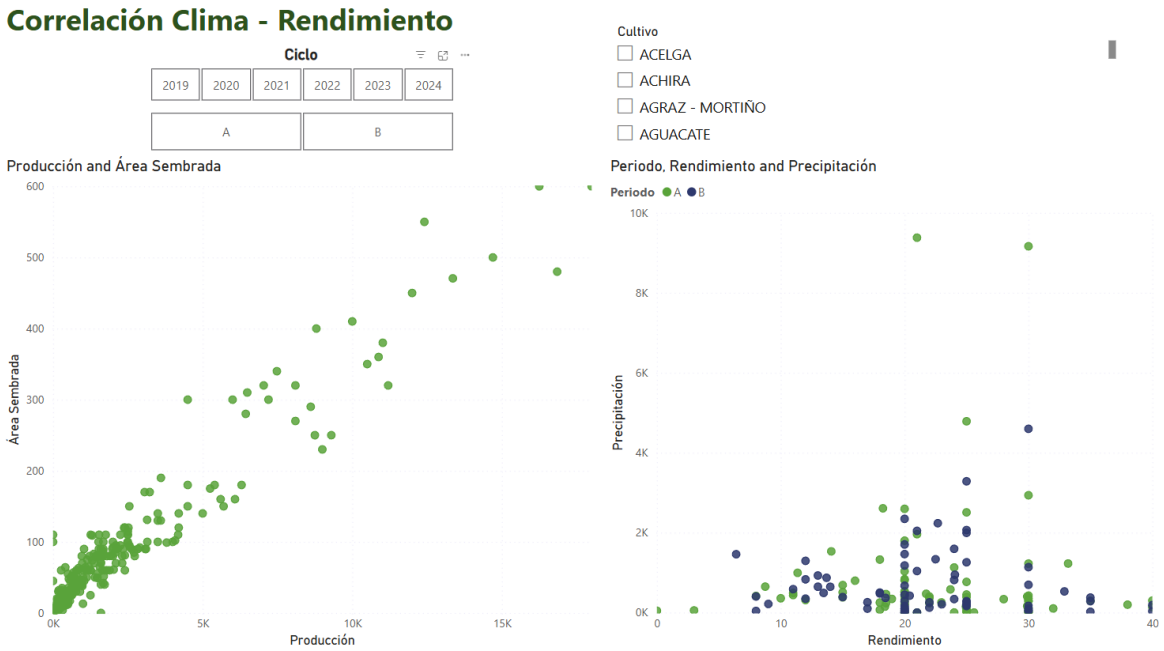


La visual Análisis Climático y Productividad Agrícola por Año, permite analizar cómo las condiciones climáticas anuales se relacionan con el rendimiento promedio de los cultivos y la productividad por municipio. Su propósito es facilitar la identificación de tendencias generales que pueden influir en la producción agrícola y reconocer los municipios con mayor aporte productivo en el periodo seleccionado.

Por otro lado, la visual Análisis Agroclimático y Productividad por Municipio tiene como finalidad facilitar la comprensión integrada entre las condiciones climáticas y el comportamiento productivo de los principales municipios agrícolas del departamento. Las visualizaciones están diseñadas para brindar un panorama comparativo mensual sobre variables clave como temperatura, precipitación y velocidad del viento, permitiendo identificar posibles patrones o tendencias que inciden en el rendimiento de los cultivos.

Figura 13

Correlación Clima - Rendimiento



Por último, la visual Correlación Clima - Rendimiento, permite explorar la relación entre la producción, el área sembrada, el rendimiento y la precipitación según el ciclo productivo. Su objetivo es facilitar la identificación de patrones entre variables climáticas y agrícolas, y así aportar insumos visuales para el análisis de posibles correlaciones entre el clima y el rendimiento de los cultivos en distintos periodos.

Conclusiones

El desarrollo de este proyecto permitió evidenciar el potencial del uso de técnicas de machine learning aplicadas al análisis de datos agrícolas en el departamento de Boyacá, generando un sistema predictivo y visual que contribuye a la toma de decisiones en torno a la productividad de los cultivos.

En primer lugar, se logró implementar un modelo de predicción utilizando Random Forest Regressor, que presentó un desempeño superior frente a otros modelos evaluados, alcanzando un coeficiente de determinación promedio de $R^2 = 0{,}822$ y un bajo error ($RMSE \approx 7.1$). Este resultado refleja una buena capacidad del modelo para predecir el rendimiento agrícola a partir de variables temporales, climáticas, tipo de cultivo y características productivas como el área sembrada, cosechada y la producción registrada.

En segundo lugar, a través del análisis exploratorio de datos se logró identificar perfiles productivos agrícolas, clasificando los cultivos según su comportamiento en términos de rendimiento y cobertura en el tiempo. Esta caracterización permite reconocer tanto los cultivos con mayor potencial productivo como aquellos que presentan limitaciones, lo que aporta información estratégica para orientar políticas agrícolas o decisiones de planificación.

Adicionalmente, se diseñó un reporte interactivo automatizado en Power BI, el cual integra visualizaciones dinámicas que permiten consultar la evolución de variables climáticas, los cultivos más productivos por municipio y la relación entre clima y rendimiento. Esta herramienta fue construida mediante flujos de datos conectados a fuentes públicas (API OData) y se encuentra disponible para su consulta en línea, con posibilidad de ser actualizado periódicamente.

Por último, se documentó el proceso de uso del modelo predictivo, permitiendo que otros usuarios -técnicos, instituciones o productores- puedan replicar el flujo de análisis con sus propios datos. Esto sienta las bases para una adopción progresiva de herramientas de analítica agrícola en el contexto local, y abre la puerta a futuras mejoras.

los resultados alcanzados dan cumplimiento a los objetivos propuestos y aportan una base sólida para continuar desarrollando herramientas basadas en datos que mejoren la toma de decisiones en el sector agropecuario. Esta iniciativa puede extenderse y adaptarse a otros territorios con necesidades similares, promoviendo prácticas agrícolas más informadas y eficientes.

Insumos finales, este proyecto fue desarrollado y publicado en GitHub y se encuentra público, allí se puede encontrar el desarrollo en sus diferentes etapas, así como sus productos finales (implementación). GiHub, por otro lado, se encuentra el reporte final público a nivel organizacional (personal/Estudiantes de la UNAD). Reporte Rendimiento Agrícola

Referencias Bibliográficas

Azeem Ayaz Mirani, M. S. (2021). Machine Learning In Agriculture: *A Review. International Journal of Scientific and Technological Research* (IJSTR), 10.

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21* (11), 3758.

Biswas, A., & Banik, R. (2024). Machine Learning Integration in Agriculture Domain: Concepts and Applications. *Fog Computing for Intelligent Cloud IoT Systems*, 71-97.

Ećim-Đurić, O., Miodragović, R., Rajković, A., Milanović, M., Mileusnić, Z., & Dragičević, A. (2024). Primena mašinskog učenja u poljoprivredi. *Poljoprivredna tehnika-Agricultural engineering*, *49*(4), 108-125.

FAO. (2019). The state of food and agriculture: Moving forward on food loss and waste reduction. *The State of the World*.

García, S., & Eisenhower, J. (2021). *Aplicación del machine learning y Big Data en el análisis de la huella de asfalto en cultivos de maíz* [Tesis de grado, Universidad Cooperativa de Colombia]. https://repositorio.ucm.edu.co/handle/10839/3254

Gobernación de Boyacá. (2023). *Minagricultura, proyectos agrícolas y potencial tecnológico* [Informe]. www.minagricultura.gov.co/ministerio/direcciones/Paginas/PDEA/Boyaca.pdf

Gupta, S., Kumar, D., & Mishra, S. Big Data Analytics in Agriculture: Harnessing IoT-generated Insights. In *Agriculture 4.0* (pp. 185-201). CRC Press.

IDEAM. (2020). *Estudio de caracterización agroclimática de Boyacá*. https://www.ideam.gov.co/

IGAC. (2020). *Informe Nacional sobre Suelos y Agricultura en Colombia*. https://www.igac.gov.co

Jhajharia, K., & Mathur, P. (2022). A comprehensive review on machine learning in agriculture domain. *IAES International Journal of Artificial Intelligence*, *11*(2), 753.

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science*, *156*(3), 312-322.

Maquera-Callo, E., Pino-Vargas, E., Choque, G., Huayna, G., Fernández-Cutire, O., & Ramos-Fernández, L. (2023). Machine Learning para clasificar el uso indiscriminado del suelo con fines agrícolas y su relación con el cambio climático, cabecera Desierto Atacama. *Idesia (Arica)*, *41*(4), 101-113.

Ministerio de Agricultura y Desarrollo Rural. (2021). *Boletín Técnico: Análisis del sector agrícola en Colombia.* https://www.minagricultura.gov.co

Parra-Peña, R. I., R., P., & Yepes, F. (2021). *Análisis de la productividad del sector agropecuario en Colombia y su impacto en temas como: encadenamientos productivos, sostenibilidad e internacionalización, en el marco del programa Colombia más competitiva* (inf. téc.). Fedesarrollo. http://hdl.handle.net/11445/4092

Rambauth Ibarra, G. E. (2022). Agricultura de Precisión: La integración de las TIC en la producción Agrícola. *CESTA*, *3* (1), 34-38. https://doi.org/10.17981/cesta.03.01.2022.04

Ramírez Gómez, C. A. (2020). Aplicación del Machine Learning en agricultura de precisión. *Revista CINTEX*, *25* (2), 14-27. https://doi.org/10.33131/24222208.356

Reddy, G. S., Reddy, M., Joshi, A., Chaitanya, K., et al. (2025). Leveraging Machine Learning Techniques in Agriculture: Applications, Challenges, and Opportunities. *Journal of Experimental Agriculture International*, *47* (1), 43-57.

Sabogal García, J. E. (2021). Aplicación del machine learning y Big Data en el análisis de la huella de asfalto en cultivos de maíz

Sharma, P., & Abrol, P. (2024). Agricultural Advancements through Machine Learning Technologies. *Environment and Ecology*, *42* (2B), 775-779.

Tantalaki, N., Souravlas, S., & Roumeliotis, M. (2019). Data-Driven Decision Making in Precision Agriculture: The Rise of Big Data in Agricultural Systems. *Journal of Agricultural & Food Information*, *20*(4), 344–380. https://doi.org/10.1080/10496505.2019.1638264

Tilman, D. B. (2011). Global Food Demand and the Sustainable Intensification of Agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 20260-20264. https://doi.org/10.1073/pnas.1116437108

Vinothkumar, S., Varadhaganapathy, S., Shanthakumari, R., Dhivya, E., Jayaharitha, K., & Livithasri, J. (2024). Crop Prediction Based on Factors of the Agricultural Environment Using Machine Learning. *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1-6.

Yadav, S., Rajendaran, M., Nagpal, A., Karthik, A., Aravinda, K., & Geetha, B. (2024). Exploring the Potential of Machine Learning in Enhancing Agricultural Practices and Food Production. *2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)*, 663-669.