

DESARROLLO DE UN ALGORITMO PARA PREDECIR LA PRODUCCIÓN DE MAÍZ EN
EL DEPARTAMENTO DEL META A 5 AÑOS, UTILIZANDO EL LENGUAJE DE
PROGRAMACIÓN PYTHON



NICOLAS ANDRES ROJAS PEREA



UNIVERSIDAD SANTO TOMÁS
FACULTAD INGENIERIA INDUSTRIAL
VILLAVICENCIO

2026

DESARROLLO DE UN ALGORITMO PARA PREDECIR LA PRODUCCIÓN DE MAÍZ EN
EL DEPARTAMENTO DEL META A 5 AÑOS, UTILIZANDO EL LENGUAJE DE
PROGRAMACIÓN PYTHON

NICOLAS ANDRÉS ROJAS PEREA

Trabajo de grado presentado como requisito para optar al título de Ingeniero Industrial

Asesor

Mg. CRISTINA ISABEL COGOLLO TORRE

Magister en Ingeniería Mecánica

UNIVERSIDAD SANTO TOMÁS
FACULTAD INGENIERIA INDUSTRIAL
VILLAVICENCIO

2026

Autoridades Académicas

P. Álvaro José ARANGO RESTREPO, O.P

Rector General

P. Mauricio Antonio CORTES GALLEGO, O. P.

Vicerrector Académico General

P. Luis Antonio ALFONSO VARGAS, O. P.

Rector Seccional Villavicencio

P. Adrián Mauricio GARCÍA PEÑARANDA, O. P.

Vicerrector Académico Seccional Villavicencio

Mg. Julieth Andrea SIERRA TOBÓN

Secretaria General Seccional Villavicencio

Mg. Víctor Andrés RINCÓN GONZALEZ

Decano de la Facultad de Ingeniería Industrial

Contenido

Contenido.....	4
Dedicatoria.....	8
Agradecimientos	9
Resumen.....	10
Abstract	11
Glosario	12
Introducción	14
Título del proyecto	15
Planteamiento del caso	16
Marco de referencia.....	18
Predicción de producción agrícola con machine learning	18
Factores climáticos y su impacto en la producción agrícola	19
Análisis de series temporales para predicción agrícola	19
Producción de maíz en Colombia y el mundo	20
Optimización de recursos agrícolas mediante modelos predictivos	21
Factores socioeconómicos en la producción de maíz	22
Preguntas de reflexión.....	24
Justificación.....	25
Objetivo General	27
Objetivos específicos.....	27
Marco Metodológico Tipo de estudio.....	28
Recolección de datos	28
Depuración y organización de la información.....	29
Diseño del esquema del algoritmo	29
Construcción del código en Python	32
Modelamiento predictivo.....	33
Calibración de hiperparámetros.....	33
Validación estadística del modelo	34
Análisis de importancia de variables	34
Proyección de la producción de maíz 2025–2030	35
Síntesis metodológica	35

Desarrollo de los objetivos	36
Desarrollo del objetivo específico 1: Recopilación, depuración y organización de los datos históricos	36
Desarrollo del objetivo específico 2: Diseño, entrenamiento y validación del modelo predictivo	37
Desarrollo del objetivo específico 3: Proyección y análisis de la producción de maíz a cinco años	39
Síntesis del desarrollo de los objetivos.....	42
Narración del caso.....	43
Lecciones y recomendaciones	44
Bibliografía	46
Anexos	49

Lista de tablas

Tabla 1 Métricas de error y precisión	38
Tabla 2 Consolidado de producción estimada (2006-2030)	41

Lista de figuras

Figura 1 Diagrama de flujo: Algoritmo predicción de maíz (Meta, 2030)	32
Figura 2 Variables influyentes en la producción de maíz	39
Figura 3 Análisis de tendencia proyectada	40

Dedicatoria

Dedico este trabajo principalmente a mi familia que ha sido mi motor, mi compañía, mi apoyo, gracias a ustedes no hubiera sido posible esto ya que me enseñaron el valor de la perseverancia y la disciplina, bases que me han sostenido durante este camino.

Agradecimientos

Agradezco a Dios por haberme dado sabiduría, fuerza y sobre todo salud para poder culminar esta etapa tan importante de mi vida.

Resumen

La producción de maíz en el departamento del Meta presenta variaciones asociadas al área sembrada, área cosechada, rendimiento, municipio y sistema de producción, lo cual dificulta la planeación agrícola, logística y productiva de mediano plazo. En este contexto, el presente estudio desarrolló un algoritmo predictivo en Python para estimar la producción de maíz en el Meta, utilizando microdatos históricos de la Unidad de Planificación Rural Agropecuaria (UPRA) correspondientes al periodo 2006–2024.

La metodología incluyó la recolección, depuración y organización de los datos, así como la homologación de la información por año, municipio y desagregación del cultivo. Posteriormente, se aplicó un modelo de aprendizaje automático basado en Random Forest Regressor, incorporando como variables de entrada el año, municipio codificado, técnica de cultivo codificada, área sembrada, área cosechada y rendimiento. El modelo fue entrenado y validado mediante una división de datos 80/20, optimización de hiperparámetros con GridSearchCV y evaluación mediante métricas como RMSE, MAPE y R^2 .

Los resultados mostraron que el algoritmo propuesto presentó un mejor desempeño frente a un método tradicional de media móvil de tres años, logrando una reducción del RMSE del 97,91 %, un MAPE de 7,90 % y un coeficiente R^2 de 0,9943. Asimismo, el análisis de importancia de variables identificó el área cosechada y el rendimiento como los factores de mayor peso en la predicción de la producción. Finalmente, se generó una proyección para el periodo 2025–2030, diferenciando entre maíz tecnificado y maíz tradicional. El estudio concluye que el algoritmo desarrollado puede servir como herramienta de apoyo para la toma de decisiones en la planeación agrícola, logística, de insumos, almacenamiento y abastecimiento del maíz en el departamento del Meta.

Palabras Clave: Random Forest, Analítica Predictiva, Maíz, Meta, Machine Learning, Proyección Agrícola.

Abstract

Maize production in the department of Meta shows variations associated with planted area, harvested area, yield, municipality, and production system, which makes medium-term agricultural, logistical, and productive planning more difficult. In this context, this study developed a predictive algorithm in Python to estimate maize production in Meta, using historical microdata from the Rural Agricultural Planning Unit (UPRA) for the 2006–2024 period.

The methodology included data collection, cleaning, and organization, as well as the standardization of information by year, municipality, and crop classification. Subsequently, a machine learning model based on Random Forest Regressor was applied, using year, encoded municipality, encoded production technique, planted area, harvested area, and yield as input variables. The model was trained and validated using an 80/20 data split, hyperparameter optimization with GridSearchCV, and performance evaluation through RMSE, MAPE, and R^2 metrics.

The results showed that the proposed algorithm performed better than a traditional three-year moving average method, achieving a 97.91% reduction in RMSE, a MAPE of 7.90%, and an R^2 coefficient of 0.9943. Likewise, the feature importance analysis identified harvested area and yield as the variables with the greatest influence on production prediction. Finally, a projection was generated for the 2025–2030 period, differentiating between technified and traditional maize. The study concludes that the developed algorithm can serve as a support tool for decision-making in agricultural planning, logistics, input management, storage, and maize supply planning in the department of Meta.

Keywords: Random Forest, Predictive Analytics, Maize, Meta, Machine Learning, Agricultural Projection.

Glosario

Lista alfabética de términos y sus definiciones necesarias para la comprensión del documento.

Algoritmo Predictivo: Conjunto de reglas y procedimientos matemáticos y estadísticos que, a través de datos históricos y actuales, permite predecir un resultado futuro. En este caso, el algoritmo se enfoca en predecir la producción de maíz en el Meta.

Machine Learning (Aprendizaje Automático): Área de la inteligencia artificial que permite a las máquinas aprender de los datos y mejorar su rendimiento sin ser explícitamente programadas. Se basa en la construcción de modelos predictivos mediante algoritmos matemáticos.

Python: Lenguaje de programación de alto nivel, ampliamente utilizado en ciencia de datos, desarrollo web y automatización. En este proyecto, se usará para el desarrollo del algoritmo predictivo.

UPRA (Unidad de Planificación Rural Agropecuaria): Entidad del gobierno colombiano que se encarga de gestionar y proporcionar datos relacionados con la agricultura en Colombia, incluyendo registros de producción de cultivos como el maíz.

Pandas: Librería de Python utilizada para el procesamiento y análisis de datos, especialmente para la manipulación de estructuras de datos como tablas y series temporales.

Scikit-learn: Librería de Python especializada en aprendizaje automático, que incluye una amplia variedad de algoritmos para clasificación, regresión y predicción, entre otros.

Regresión: Técnica de análisis estadístico utilizada para modelar y analizar relaciones entre variables. En este proyecto, se aplicará para predecir la producción de maíz a partir de variables como el clima y el tipo de suelo.

Rendimiento: Medida de la cantidad de producción obtenida por unidad de área. En el caso del maíz, se expresa como la cantidad de toneladas por hectárea (t/ha).

Grano Seco: Estado físico del maíz que indica que el cultivo ha sido cosechado y secado para su almacenamiento, a diferencia del maíz "en fresco", que se encuentra recién cosechado.

Serie Temporal: Conjunto de datos recopilados y ordenados en función del tiempo. En este estudio, se utilizarán series temporales para analizar las tendencias de producción de maíz en el Meta a lo largo de los años.

Técnica de Validación: Método utilizado para evaluar el rendimiento de un modelo predictivo, como la validación cruzada o el uso de métricas de error (MAE, RMSE, etc.), con el objetivo de comprobar la precisión de las predicciones.

Proyección a Largo Plazo: Estimación de resultados futuros basados en datos históricos y modelos predictivos. En este estudio, se proyecta la producción de maíz para los próximos cinco años.

Variables Socioeconómicas: Factores sociales y económicos que influyen en la producción agrícola, tales como los costos de insumos, la disponibilidad de mano de obra, los precios de los productos y las políticas públicas.

Pronóstico: Estimación o predicción de un evento o condición futura. En este contexto, se refiere a la predicción de la producción de maíz.

Toma de Decisiones: Proceso mediante el cual los agricultores, planificadores y demás actores del sector agropecuario seleccionan la mejor opción en función de los datos disponibles, como los pronósticos de producción generados por el algoritmo.

Introducción

La producción de maíz tiene una importancia significativa dentro del sector agrícola colombiano y, de manera particular, en el departamento del Meta, donde este cultivo participa en la dinámica productiva, logística y de abastecimiento regional. Sin embargo, su comportamiento puede variar de un año a otro debido a factores asociados al área sembrada, el área cosechada, el rendimiento, el municipio y el sistema de producción utilizado. Esta variabilidad dificulta la planeación agrícola de mediano plazo, especialmente cuando las decisiones se apoyan únicamente en análisis descriptivos o en métodos tradicionales de estimación.

En este contexto, el presente proyecto desarrolla un algoritmo predictivo en Python para estimar la producción de maíz en el departamento del Meta a cinco años, utilizando información histórica de la Unidad de Planificación Rural Agropecuaria (UPRA) correspondiente al periodo 2006–2024. El propósito del modelo no es solamente proyectar una cifra futura de producción, sino transformar los datos disponibles en información útil para apoyar la toma de decisiones en la planeación agrícola, logística y productiva de la región.

Para el desarrollo del algoritmo se emplearon técnicas de aprendizaje automático, específicamente el modelo Random Forest Regressor, debido a su capacidad para trabajar con relaciones no lineales entre variables. El modelo integró variables como año, municipio, sistema de producción, área sembrada, área cosechada y rendimiento. Además, se utilizaron herramientas del ecosistema de Python, como Pandas, NumPy, Scikit-learn, Matplotlib y Seaborn, para la depuración de datos, construcción del modelo, validación estadística y visualización de resultados.

La investigación también incluyó una etapa de organización y homologación de la información histórica, con el fin de consolidar los datos por año, municipio y desagregación del cultivo. Posteriormente, el modelo fue entrenado, calibrado y evaluado mediante métricas como RMSE, MAPE y R^2 , además de una comparación frente a un método tradicional de media móvil. Finalmente, se generó una proyección para el periodo 2025–2030, diferenciando entre maíz tecnificado y maíz tradicional.

Desde la perspectiva de la ingeniería industrial, este proyecto aporta una herramienta de analítica predictiva que puede servir como apoyo para la planeación de siembras, estimación de insumos, uso de maquinaria, almacenamiento, transporte y abastecimiento. De esta manera, el algoritmo propuesto busca contribuir a una gestión agrícola más informada, basada en datos y adaptada al contexto productivo del departamento del Meta.

Título del proyecto

Desarrollo de un algoritmo para predecir la producción de maíz en el departamento del Meta a 5 años, utilizando el lenguaje de programación Python.

Planteamiento del caso

La producción de maíz tiene una importancia significativa dentro del sector agrícola colombiano y, de manera particular, en el departamento del Meta. De acuerdo con la Unidad de Planificación Rural Agropecuaria, el maíz representa el 4,4 % de la producción agrícola nacional y corresponde al cultivo con la tercera mayor extensión del país, lo que evidencia su relevancia económica y productiva dentro del contexto agropecuario colombiano (UPRA, 2024). En el caso del Meta, este cultivo se destaca por su participación en la oferta agrícola regional, por su aptitud productiva y por su relación directa con actividades de abastecimiento, transformación, comercialización y seguridad alimentaria.

Según el microanálisis de Evaluaciones Agropecuarias de la UPRA, para el año 2022 el departamento del Meta registró 88.641 hectáreas sembradas y una producción de 522.155 toneladas de maíz. Además, la entidad reporta que el cultivo presenta una alta aptitud agropecuaria en el departamento, con 3.736.255 hectáreas de aptitud alta, equivalentes al 75 % de la frontera agrícola del Meta, lo que confirma su potencial estratégico dentro de la dinámica productiva regional (UPRA, 2023). Sin embargo, este potencial productivo no elimina la incertidumbre asociada al comportamiento futuro del cultivo, especialmente cuando la producción depende de variables como el área sembrada, el área cosechada, el rendimiento, la ubicación municipal y el sistema de producción utilizado.

A pesar de su importancia, la producción de maíz en el Meta presenta variaciones que dificultan su planeación a mediano plazo. Estas variaciones pueden estar asociadas a factores ambientales, técnicos y productivos. La FAO señala que la producción agropecuaria se ve afectada por las condiciones meteorológicas, la variabilidad del suelo y los cambios en el comportamiento de los rendimientos, lo cual genera incertidumbre en los sistemas agrícolas (Basso, 2013). En el contexto colombiano, los resultados de las Evaluaciones Agropecuarias también evidencian que el comportamiento reciente del maíz ha estado influenciado por condiciones climáticas adversas, cambios en las decisiones de siembra y dinámicas del mercado, factores que pueden modificar la producción de un año a otro (UPRA, 2025).

Esta incertidumbre representa un problema para productores, planificadores, instituciones y demás actores del sector agrícola, debido a que limita la capacidad de tomar decisiones oportunas sobre siembra, compra de insumos, uso de maquinaria, almacenamiento, transporte y gestión del

riesgo. Cuando no se cuenta con una estimación confiable de la producción futura, la planeación agrícola tiende a depender de análisis históricos descriptivos o de métodos tradicionales que no siempre capturan la complejidad del sistema productivo. En consecuencia, se pueden generar decisiones poco precisas frente a la disponibilidad de recursos, la demanda logística y la sostenibilidad de la producción regional.

Aunque existen fuentes oficiales como las Evaluaciones Agropecuarias, que registran variables de área, producción y rendimiento, disponer de información histórica no garantiza por sí solo una adecuada planeación. El verdadero reto consiste en transformar esos datos en herramientas analíticas que permitan anticipar escenarios futuros. En este sentido, el aprendizaje automático se ha consolidado como una alternativa útil para la predicción agrícola, ya que permite integrar múltiples variables, identificar patrones no lineales y mejorar la precisión de los modelos de pronóstico frente a métodos únicamente descriptivos (van Klompenburg, Kassahun, & Catal, 2020).

Por lo anterior, el problema central de este proyecto se encuentra en la ausencia de una herramienta predictiva específica, validada y adaptada al contexto territorial del departamento del Meta, que permita estimar la producción futura de maíz a partir de datos históricos oficiales. Esta limitación afecta la planeación agrícola a cinco años y reduce la capacidad de anticipar necesidades productivas, logísticas y de abastecimiento. En consecuencia, surge la necesidad de desarrollar un algoritmo en Python, utilizando información histórica de la UPRA para el periodo 2006–2024, que permita predecir la producción de maíz en el departamento del Meta y aportar técnicamente a la toma de decisiones dentro del sector agrícola regional.

Marco de referencia

Predicción de producción agrícola con machine learning

Según Zeer et al. (2024), en un estudio realizado en la Universidad de Palestina Ahliya, se compararon distintos algoritmos de machine learning para predecir la productividad agrícola. Los resultados mostraron que los modelos no lineales, especialmente Random Forest, alcanzaron una precisión del 92 %, mientras que los modelos lineales, como la regresión simple, obtuvieron una precisión del 78 %. Además, la inclusión de variables climáticas incrementó la precisión en un 10 %. Este antecedente es relevante para el presente proyecto porque demuestra que Random Forest puede ser más adecuado que los modelos lineales cuando la producción agrícola depende de varias variables, como ocurre con el maíz en el Meta, donde intervienen el área cosechada, el rendimiento, el municipio y el sistema productivo.

De acuerdo con Ane (2024), en una revisión sobre aplicaciones de machine learning en agricultura, se identificó que estos modelos pueden contribuir a transformar la agricultura tradicional en una más eficiente y precisa. El estudio señala que el uso de modelos predictivos podría aumentar la productividad agrícola hasta en un 15 %, y que el mercado global de inteligencia artificial y machine learning en agricultura pasaría de 1,7 mil millones de dólares en 2023 a 11,13 mil millones de dólares en 2032. Estos datos muestran la importancia creciente de las herramientas predictivas en el sector agrícola y respaldan la pertinencia de desarrollar un algoritmo en Python para apoyar la planeación agrícola, logística y productiva del maíz en el Meta.

Con base en Surana y Khandelwal (2024), quienes compararon algoritmos como Random Forest, AdaBoost, Gradient Boosting y Support Vector Machine en un conjunto de 2.201 instancias con variables como pH del suelo, precipitación, temperatura y humedad, se encontró que Random Forest alcanzó una precisión del 99,99 %, mientras que AdaBoost obtuvo 99,03 %. Este estudio aporta al presente trabajo porque evidencia que los modelos de ensamble pueden manejar múltiples variables con alta precisión, lo cual respalda el uso de Random Forest para analizar la producción de maíz a partir de variables como año, municipio, técnica de cultivo, área sembrada, área cosechada y rendimiento.

Factores climáticos y su impacto en la producción agrícola

Según Thakur et al. (2024), la variabilidad climática y los eventos extremos pueden afectar de forma significativa la producción agrícola. El estudio señala que la producción podría reducirse hasta en un 21 % en países en desarrollo debido al calentamiento global, con aumentos proyectados de temperatura entre 2,33 °C y 4,78 °C en las próximas décadas. Aunque el presente proyecto no incorpora variables climáticas directas, este antecedente permite entender que el rendimiento y el área cosechada pueden reflejar indirectamente los efectos de condiciones ambientales sobre la producción de maíz.

De acuerdo con Şimon et al. (2023), en un estudio realizado en la llanura de Transilvania entre 2012 y 2021, se analizó el efecto de la temperatura, la precipitación, la reserva de agua en el suelo y las horas de sol sobre el maíz. Los resultados mostraron que, aunque la reserva de agua en primavera alcanzaba 1.734,8 m³ ha⁻¹, en meses críticos como julio y agosto la disponibilidad hídrica disminuía, afectando la producción; además, en años secos se registraron reducciones de productividad de hasta el 50 %. Este estudio respalda la inclusión del rendimiento como variable clave en el algoritmo, ya que este indicador recoge parte de los efectos productivos y ambientales sobre el cultivo.

Partiendo de Zhang et al. (2022), quienes analizaron datos de 3.050 agricultores en China entre 2009 y 2018, se encontró que un aumento del 1 % en la temperatura elevó la producción de maíz en 0,39 %, mientras que la precipitación tuvo un efecto positivo de 0,015 % y las horas de luz solar redujeron la producción en 0,05 %. Estos resultados muestran que las variables ambientales pueden tener efectos diferenciados sobre la producción, lo que justifica el uso de modelos flexibles como Random Forest, capaces de identificar relaciones no lineales en los datos históricos del maíz en el Meta.

Análisis de series temporales para predicción agrícola

En referencia a Srivyshnavi y Aju (2024), se utilizó un modelo ARIMA (1,1,2) para proyectar variables agrícolas con datos históricos de precios de Bengal Gram entre 1991 y 2020, generando predicciones para el periodo 2023–2025 con una precisión del 99,29 %. Este antecedente demuestra la utilidad de las series temporales para proyectar comportamientos futuros;

sin embargo, el presente proyecto requiere un enfoque multivariable, ya que la producción de maíz no depende solo del año, sino también del municipio, el sistema productivo, el área sembrada, el área cosechada y el rendimiento.

Basado en Kerdprasop et al. (2023), se realizó un análisis de producción alimentaria con datos de 164 países durante el periodo 1971–2020, aplicando ARIMA para análisis univariado y modelos de ensamble como Random Forest, AdaBoost y Gradient Boosting para análisis multivariado. Los resultados mostraron que Gradient Boosting alcanzó una precisión del 99,1 %, mientras que ARIMA (1,1,1) fue adecuado para predicciones globales. Este estudio aporta al proyecto porque muestra que los modelos de ensamble pueden mejorar la predicción cuando se integran varias variables, como se realiza en el algoritmo desarrollado para el maíz en el Meta.

En el estudio de Desloires et al. (2024), realizado en el Cinturón Maicero de Estados Unidos, se integraron imágenes Sentinel-2, Landsat y temperatura superficial para predecir el rendimiento del maíz en 1.319 campos entre 2017 y 2022. El uso de temperatura superficial mejoró la precisión en 5,1 % y se obtuvo un MAPE de 16,25 % al predecir el rendimiento dos meses antes de la cosecha. Este antecedente es relevante porque demuestra que integrar diferentes variables mejora la predicción agrícola; además, el uso del MAPE como métrica coincide con la validación empleada en este proyecto.

Producción de maíz en Colombia y el mundo

En consonancia con Ordoñez et al. (2022), en Colombia se aplicaron pronósticos agroclimáticos estacionales para mejorar la gestión agronómica del maíz de secano. El modelo mostró una capacidad de discriminación de hasta el 80 % para clasificar rendimientos “por debajo de lo normal” y “por encima de lo normal”, y predijo fechas óptimas de siembra con un error de 3 pentadas, equivalentes a 15 días. Este antecedente es importante porque demuestra que la predicción puede apoyar decisiones agrícolas concretas en Colombia, como también se busca en este proyecto mediante la proyección de la producción de maíz en el Meta.

En el estudio de Vélez et al. (2019), realizado en el trópico medio colombiano, se utilizaron imágenes Sentinel-2 para analizar la firma espectral del maíz en etapas de 2 meses, 2,3 meses y 4,3 meses. Los resultados mostraron una NDVI máxima de 0,67 en la etapa final, indicando buen estado fitosanitario del cultivo. Aunque este proyecto no emplea imágenes

satelitales, el estudio evidencia que las herramientas tecnológicas permiten mejorar el monitoreo del maíz y respaldan el uso de soluciones basadas en datos para fortalecer la toma de decisiones agrícolas.

De acuerdo con García et al. (2014), se simuló la producción de maíz en Colombia mediante el modelo AquaCrop en Chinchiná, Palmira y Cereté. El modelo alcanzó una precisión de 96,77 % en biomasa y 99,66 % en índice de cosecha; además, Palmira presentó el mayor rendimiento con 4,9 t/ha. Este antecedente demuestra que la modelación agrícola permite entender y proyectar el comportamiento del maíz bajo distintas condiciones, lo cual se relaciona con el presente proyecto, aunque aquí se emplea aprendizaje automático y datos históricos oficiales del Meta.

Optimización de recursos agrícolas mediante modelos predictivos

Con base en Tynykulova et al. (2024), en la región de Bihar, India, se evaluaron algoritmos como Random Forest, Support Vector Machines y XGBoost para predecir el rendimiento del maíz bajo condiciones climáticas variables. Los resultados indicaron que XGBoost obtuvo la mayor precisión, con 98,3 %. Este estudio aporta al presente proyecto porque confirma que los modelos de machine learning pueden reducir la incertidumbre en la producción de maíz y apoyar decisiones agrícolas más informadas.

En referencia a Sowmya y Prasad (2024), se investigó la predicción del rendimiento del arroz mediante Random Forest y Gradient Boosting, optimizando los modelos con GridSearchCV. El modelo combinado alcanzó un R^2 de 0,88, superando al Random Forest individual, que obtuvo un R^2 de 0,85, y redujo el Error Medio Absoluto a 135,87. Este antecedente respalda directamente el uso de GridSearchCV en el presente proyecto, ya que evidencia que la calibración de hiperparámetros mejora el desempeño de los modelos predictivos.

Partiendo de Elbasi et al. (2023), se evaluaron algoritmos de machine learning para la predicción de cultivos con 2.200 registros y 22 etiquetas de cultivos, incorporando variables como nitrógeno, temperatura, pH, precipitación y humedad. Los resultados mostraron una precisión de 99,59 % para Red Bayesiana y de 99,46 % para Naïve Bayes y Hoeffding Tree. Este antecedente permite afirmar que los modelos predictivos pueden apoyar decisiones de siembra, riego y cosecha; en el caso del Meta, el algoritmo propuesto busca apoyar la planeación de producción,

insumos, almacenamiento y transporte del maíz.

Basado en Musanase et al. (2023), se desarrolló en Ruanda un sistema de recomendación de cultivos y fertilizantes mediante machine learning, utilizando variables como nitrógeno, fósforo, potasio y pH. El modelo alcanzó una precisión del 97 % en las recomendaciones de cultivos. Este estudio se relaciona con el presente proyecto porque demuestra que la analítica de datos puede apoyar la gestión de recursos agrícolas; en este caso, la proyección de producción de maíz puede servir para planear insumos, maquinaria, transporte y almacenamiento.

Factores socioeconómicos en la producción de maíz

Según Hartanto et al. (2024), en Indonesia se analizaron los patrones de consumo de 86 hogares agrícolas dedicados al maíz mediante regresión múltiple. El estudio encontró que el ingreso familiar, el número de miembros de la familia y el nivel educativo del jefe de hogar tenían efectos positivos y significativos en las decisiones económicas de los productores. Este antecedente permite comprender que la producción de maíz también está relacionada con factores socioeconómicos, aspecto que en el presente proyecto se aborda de forma indirecta al diferenciar entre maíz tecnificado y tradicional.

A partir de Mncube et al. (2024), en Eswatini se analizaron factores socioeconómicos y de manejo que influyen en la productividad del maíz en sistemas de pequeña escala. El régimen de solo fertilizante inorgánico mostró un incremento del 37,4 % en el rendimiento frente al uso de estiércol. Este estudio respalda la importancia del rendimiento como variable del modelo, ya que demuestra que las prácticas de manejo productivo influyen directamente en la cantidad de maíz obtenida por hectárea.

De acuerdo con Onuwa et al. (2023), en Nigeria se estudiaron los determinantes de adopción tecnológica en 101 pequeños productores de maíz. Los resultados mostraron que el tamaño del hogar, con coeficiente de 0,541; la educación, con 0,331; y la experiencia agrícola, con 0,615, fueron significativos con $p < 0,05$. Además, el alto costo de las tecnologías y su complejidad afectaron la adopción en 81,2 % y 80,2 %, respectivamente. Este estudio ayuda a interpretar la diferencia entre maíz tecnificado y tradicional, ya que la adopción tecnológica puede modificar la productividad y el comportamiento futuro del cultivo.

El estudio de Martínez-Reina (2013), desarrollado en La Mojana, Colombia, analizó costos

de producción en sistemas agrícolas, incluyendo maíz. En el sistema de maíz tradicional a chuzo en Majagual, los costos totales fueron de \$2.359.800, de los cuales \$2.059.800 correspondían a costos directos y \$300.000 a costos indirectos. En el sistema de maíz tradicional verde en San Marcos, San Benito y Caimito, los costos totales fueron de \$1.821.000, con \$1.571.000 en costos directos; además, la siembra y la cosecha representaron costos entre \$420.000 y \$480.000, y los jornales correspondieron al 49 % de los costos directos. Este antecedente conecta el proyecto con la ingeniería industrial, porque evidencia que la planeación de siembra, cosecha, maquinaria y mano de obra puede impactar los costos y la eficiencia productiva.

Preguntas de reflexión

- ¿Qué variables históricas de la UPRA, como área sembrada, área cosechada, rendimiento, municipio y sistema de producción, explican en mayor medida la variación de la producción de maíz en el departamento del Meta?
- ¿Qué nivel de precisión ofrece el algoritmo desarrollado en Python para predecir la producción de maíz en el Meta, en comparación con métodos tradicionales de estimación, a partir de métricas como RMSE, MAPE y R^2 ?
- ¿Cómo pueden utilizarse las proyecciones de producción de maíz a cinco años para apoyar la planeación agrícola, logística, de abastecimiento, almacenamiento e insumos en el departamento del Meta?

Justificación

La presente investigación se justifica en la necesidad de fortalecer la planeación agrícola del departamento del Meta mediante el uso de herramientas predictivas que permitan transformar los datos históricos en información útil para la toma de decisiones. El maíz es un cultivo estratégico para Colombia, ya que representa el 4,4 % de la producción agrícola nacional y se consolida como el tercer cultivo con mayor extensión del país, con una participación del 10 %, solo superado por el café y el arroz. Además, la mayor producción departamental de maíz se concentra en territorios como Meta, Córdoba, Tolima y Valle del Cauca, lo que evidencia la relevancia del departamento dentro de la dinámica nacional de este cultivo (UPRA, 2024).

En el contexto regional, el Meta presenta condiciones favorables para el desarrollo del cultivo de maíz. De acuerdo con el microanálisis de Evaluaciones Agropecuarias de la UPRA, el departamento cuenta con 3.736.255 hectáreas de aptitud alta para maíz, equivalentes al 75 % de su frontera agrícola. Asimismo, para el año 2022, las EVA reportaron en el Meta 88.641 hectáreas sembradas de maíz, 85.114 hectáreas cosechadas y una producción de 522.155 toneladas, ubicándolo entre los principales productos agrícolas del departamento. Estos datos muestran que el cultivo tiene un peso importante en la producción regional y, por tanto, requiere herramientas que permitan anticipar su comportamiento futuro con mayor precisión (UPRA, 2023).

Desde una perspectiva práctica, el desarrollo de un algoritmo en Python para predecir la producción de maíz a cinco años resulta pertinente porque la producción agrícola no solo depende del área sembrada, sino también del área cosechada, el rendimiento, el municipio y el sistema de producción utilizado. En las Evaluaciones Agropecuarias Municipales, la producción se estima a partir de variables como área, producción y rendimiento, información que permite conocer la oferta agropecuaria del país y analizar el comportamiento de los cultivos transitorios y permanentes (Ministerio de Agricultura y Desarrollo Rural, 2013; UPRA, 2024).

La importancia de esta investigación también radica en su aporte a la toma de decisiones dentro de la ingeniería industrial aplicada al sector agropecuario. Una proyección confiable de la producción de maíz puede apoyar la programación de siembras, la estimación de necesidades de insumos, la planeación del uso de maquinaria, la capacidad de almacenamiento, los procesos de secado, la logística de transporte y la gestión del abastecimiento regional. En este sentido, el algoritmo no se limita a entregar una cifra futura, sino que se convierte en una herramienta de

apoyo para planificar recursos, anticipar riesgos y mejorar la eficiencia de los procesos asociados a la cadena productiva del maíz.

Desde el punto de vista metodológico, el proyecto se sustenta en fuentes oficiales de información, principalmente las Evaluaciones Agropecuarias Municipales. Según la UPRA, las EVA constituyen una herramienta estadística fundamental para planificar y gestionar el sector agropecuario en Colombia, ya que proporcionan datos detallados sobre la producción de los municipios del país y reportan información relacionada con su actividad productiva (UPRA, 2024). De igual manera, la ficha metodológica del Ministerio de Agricultura y Desarrollo Rural establece que estas evaluaciones permiten estimar variables de área, producción y rendimiento, lo cual respalda la pertinencia de utilizar esta base como insumo principal del modelo predictivo (Ministerio de Agricultura y Desarrollo Rural, 2013).

El uso de aprendizaje automático se justifica porque permite analizar relaciones complejas entre variables agrícolas y mejorar la capacidad de predicción frente a métodos tradicionales. La literatura científica señala que el machine learning es una herramienta importante de apoyo a la decisión en la predicción de rendimientos agrícolas, especialmente porque permite integrar diferentes variables y reconocer patrones en los datos históricos (van Klompenburg et al., 2020). Además, los avances en técnicas de aprendizaje automático y tecnologías de medición agrícola han fortalecido la posibilidad de desarrollar soluciones más precisas y útiles para la estimación del estado de los cultivos y la toma de decisiones productivas (Chlingaryan et al., 2018; van Klompenburg et al., 2020).

En síntesis, esta investigación se justifica por su pertinencia regional, su utilidad práctica y su aporte técnico desde la ingeniería industrial. El Meta cuenta con una alta aptitud para el cultivo de maíz y con una producción significativa dentro del contexto agrícola departamental; sin embargo, la variabilidad del área cosechada, el rendimiento y las condiciones productivas exige herramientas que permitan anticipar escenarios futuros. Por ello, el desarrollo de un algoritmo predictivo en Python representa una alternativa pertinente para apoyar la planeación agrícola, logística y productiva del maíz en el departamento del Meta.

Objetivos

Objetivo general

Desarrollar un algoritmo en Python para predecir la producción de maíz en el departamento del Meta a cinco años, utilizando datos históricos de la UPRA, con el fin de apoyar la planeación agrícola, logística y productiva de la región.

Objetivos específicos

- Recopilar, depurar y organizar los datos históricos de producción de maíz del departamento del Meta, correspondientes al periodo 2006–2024, a partir de la información disponible en la UPRA.
- Diseñar y validar un modelo de aprendizaje automático basado en Random Forest, mediante el análisis de variables como área sembrada, área cosechada, rendimiento, municipio y sistema de producción, evaluando su desempeño con métricas como RMSE, MAPE y R^2 .
- Proyectar y analizar la producción de maíz en el departamento del Meta a cinco años, con el fin de generar información útil para la toma de decisiones en la planeación agrícola, logística y productiva regional.

Marco Metodológico

Tipo de estudio

La presente investigación se desarrolló bajo un enfoque aplicativo, cuantitativo y predictivo. Es aplicativo porque se orientó a resolver una necesidad concreta del sector agrícola del departamento del Meta: la ausencia de una herramienta que permita anticipar la producción futura de maíz y apoyar la planeación agrícola, logística y productiva. Es cuantitativo porque se fundamentó en el procesamiento de datos históricos de área sembrada, área cosechada, rendimiento y producción. Asimismo, es predictivo porque utilizó técnicas de aprendizaje automático para identificar patrones en la información histórica y generar una proyección de la producción de maíz a cinco años.

El estudio se enmarca en la analítica de datos aplicada a la ingeniería industrial, debido a que no se limita a describir el comportamiento histórico del cultivo, sino que busca transformar los datos disponibles en información útil para la toma de decisiones. Desde esta perspectiva, el algoritmo desarrollado permite apoyar procesos asociados a la planeación de siembras, estimación de insumos, uso de maquinaria, almacenamiento, transporte y abastecimiento regional.

Recolección de datos

La recolección de datos se realizó a partir de microdatos oficiales de la Unidad de Planificación Rural Agropecuaria (UPRA), correspondientes a la producción de maíz en el departamento del Meta durante el periodo 2006–2024. La información utilizada permitió consolidar registros asociados a los municipios del departamento y a las diferentes formas de desagregación del cultivo.

La base de datos fue trabajada en formato Excel y cargada en el entorno de Google Colaboratory mediante conexión con Google Drive. En el código desarrollado se empleó la librería Pandas para importar el archivo, visualizar su estructura inicial y realizar los procesos de agrupación, depuración y organización de la información. La base fue consolidada a partir de tres campos principales de identificación: Año, Municipio y Desagregación cultivo. Posteriormente, se agregaron los valores de área sembrada, área cosechada y producción, con el fin de obtener un único registro por año, municipio y sistema de cultivo.

Las variables consideradas para el análisis fueron:

- Año: periodo de registro de la producción.
- Municipio: ubicación territorial de la producción dentro del departamento del Meta.
- Desagregación cultivo: clasificación del cultivo según el sistema de producción.
- Área sembrada (ha): superficie establecida para el cultivo de maíz.
- Área cosechada (ha): superficie efectivamente recolectada.
- Producción (t): toneladas producidas de maíz.
- Rendimiento (t/ha): relación entre la producción obtenida y el área cosechada.

El rendimiento se recalculó después de la agrupación de los datos mediante la relación entre producción y área cosechada. Este procedimiento permitió mantener coherencia en la base final y evitar distorsiones derivadas de la consolidación de registros.

Depuración y organización de la información

Para preparar los datos antes del modelamiento, se realizó un proceso de depuración y homologación de la información. Este paso fue necesario debido a que la base presentaba diferentes formas de clasificación del cultivo, especialmente en la variable Desagregación cultivo. Por esta razón, se consolidaron las categorías en dos grandes grupos: maíz tecnificado y maíz tradicional.

Esta decisión metodológica permitió diferenciar los sistemas productivos y analizar el comportamiento de la producción de acuerdo con el nivel de tecnificación. Además, permitió que el modelo identificara patrones distintos entre ambos sistemas, aspecto importante para la posterior proyección de la producción.

También se tuvo en cuenta la necesidad de dar continuidad a la serie histórica 2006–2024, considerando los cambios metodológicos en el reporte del cultivo. Por ello, la base fue organizada de manera que permitiera comparar los datos históricos bajo una estructura común y compatible con el modelo predictivo.

Diseño del esquema del algoritmo

El diseño del esquema lógico del algoritmo tendrá en cuenta las necesidades del sector agrícola, con un enfoque específico en la predicción de la producción de maíz en el Meta. La

estrategia principal será integrar las variables relevantes del cultivo, tanto las nuevas como las antiguas, para garantizar que las predicciones sean lo más precisas posible.

El diseño del algoritmo se estructuró a partir de una lógica de analítica predictiva. El objetivo fue transformar los datos históricos de producción de maíz en una base organizada, codificada y apta para el entrenamiento de un modelo de aprendizaje automático.

El esquema metodológico se desarrolló en las siguientes fases:

Fase 1. Carga y revisión inicial de datos

Se importó la base de datos desde Google Drive hacia Google Colaboratory. Luego se revisó el número de registros, columnas y estructura general de la información para confirmar que el archivo estuviera correctamente cargado.

Fase 2. Agrupación y consolidación de registros

Los datos fueron agrupados por año, municipio y desagregación del cultivo. En esta fase se sumaron las áreas sembradas, áreas cosechadas y producción, generando una base unificada con un solo valor de producción por técnica, municipio y año.

Fase 3. Cálculo del rendimiento

Después de consolidar los datos, se recalculó el rendimiento en toneladas por hectárea, dividiendo la producción total entre el área cosechada. Esta variable fue incorporada como uno de los predictores principales del modelo.

Fase 4. Codificación de variables categóricas

Debido a que el modelo Random Forest requiere variables numéricas para su procesamiento, se aplicó la técnica Label Encoding a las variables categóricas Municipio y Desagregación cultivo. Como resultado, se generaron las variables Municipio_Enc y Tecnica_Enc,

las cuales fueron incorporadas al conjunto de variables predictoras.

Fase 5. Definición de variables de entrada y salida

Las variables de entrada del modelo fueron: Año, Municipio_Enc, Tecnica_Enc, Área sembrada (ha), Rendimiento (t/ha) y Área cosechada (ha). La variable de salida fue Producción (t), correspondiente a la producción de maíz en toneladas.

Fase 6. Entrenamiento del modelo predictivo

La base de datos fue dividida en dos subconjuntos: 80 % para entrenamiento y 20 % para prueba, utilizando una semilla aleatoria de 42 para garantizar reproducibilidad. Posteriormente, se entrenó un modelo Random Forest Regressor, inicialmente configurado con 100 árboles de decisión y `random_state = 42`.

Fase 7. Validación del modelo

El modelo fue evaluado con el conjunto de prueba mediante métricas de desempeño como el coeficiente de determinación (R^2) y el error cuadrático medio en su raíz (RMSE). Estas métricas permitieron analizar la capacidad del modelo para explicar la variabilidad de la producción y cuantificar el error promedio de predicción.

Fase 8. Proyección futura

Una vez entrenado el modelo, se generaron proyecciones para el periodo 2025–2030. Para ello, se tomó como base el último año disponible en la serie histórica, correspondiente a 2024, y se construyeron escenarios futuros para cada año proyectado.

Fase 9. Proyección dinámica por sistema productivo

Además de la proyección inicial, se construyó un escenario dinámico diferenciado por sistema de producción. En este escenario se asumió un crecimiento anual del 1,5 % para el maíz tecnificado y una disminución anual del 1,0 % para el maíz tradicional. Este supuesto permitió

representar una posible transición productiva hacia sistemas más tecnificados y observar su efecto en la producción futura. En el código, este procedimiento se aplicó sobre el área sembrada, manteniendo constante el rendimiento para aislar el efecto del cambio de área.

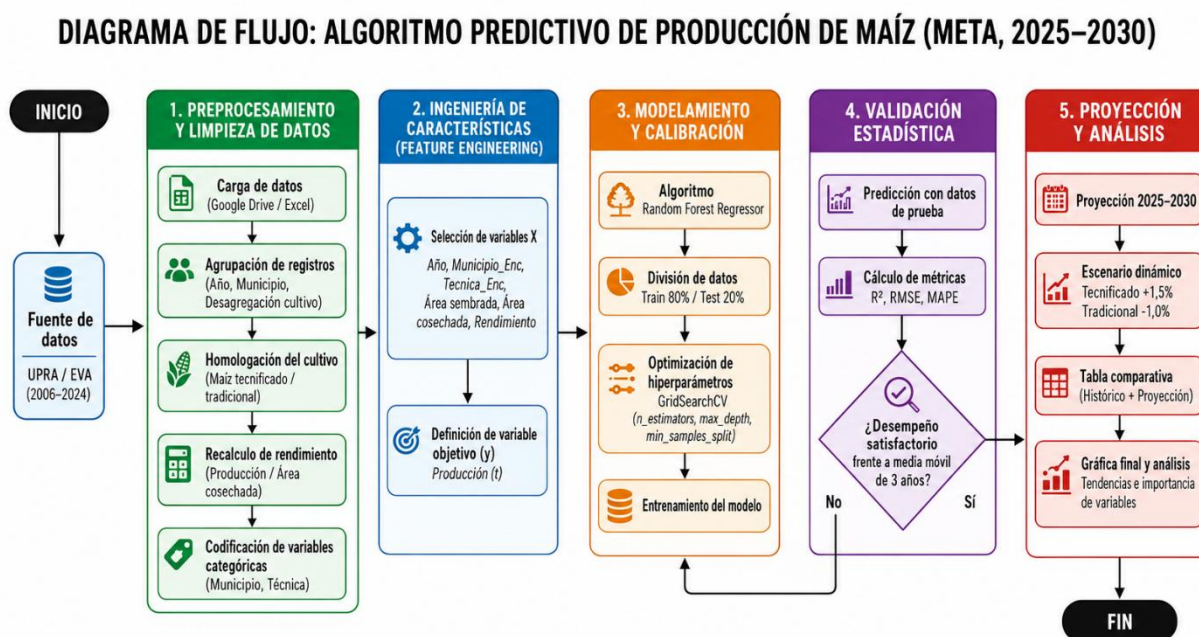
Fase 10. Generación de resultados gráficos y tabulares

Finalmente, se construyeron gráficas comparativas entre la producción histórica y la producción proyectada. También se generó una tabla consolidada con los datos históricos 2006–2024 y las proyecciones 2025–2030, diferenciando entre maíz tecnificado y maíz tradicional.

Diagrama de flujo:

Figura 1

Diagrama de flujo: Algoritmo predicción de maíz (Meta, 2030)



Construcción del código en Python

El desarrollo técnico del algoritmo se realizó en Google Colaboratory, utilizando Python como lenguaje de programación. Este entorno permitió trabajar con la base de datos almacenada

en Google Drive y facilitó la ejecución del modelo, la generación de gráficos y la visualización de resultados.

Las principales librerías utilizadas fueron:

- Pandas: utilizada para cargar, agrupar, depurar y organizar la base de datos.
- NumPy: empleada para cálculos numéricos y operaciones matemáticas.
- Matplotlib: utilizada para la generación de gráficos de tendencia y comparación.
- Seaborn: empleada para mejorar la visualización gráfica y representar la importancia de variables.
- Scikit-learn: utilizada para la construcción, entrenamiento, validación y calibración del modelo predictivo.

Dentro de Scikit-learn se utilizaron herramientas específicas como RandomForestRegressor, train_test_split, mean_squared_error, r2_score, LabelEncoder y GridSearchCV. Estas herramientas permitieron construir el modelo, separar la base en entrenamiento y prueba, codificar variables categóricas, calcular métricas de desempeño y optimizar los hiperparámetros del algoritmo

Modelamiento predictivo

El modelo seleccionado fue Random Forest Regressor, debido a su capacidad para trabajar con relaciones no lineales entre variables y manejar conjuntos de datos con diferentes tipos de comportamiento. En el caso de la producción agrícola, esta característica es importante, ya que la producción de maíz no depende únicamente del paso del tiempo, sino de la interacción entre área sembrada, área cosechada, rendimiento, municipio y sistema de producción.

El modelo fue entrenado con las variables de entrada definidas en la etapa de ingeniería de características. A partir de estas variables, el algoritmo aprendió patrones históricos de comportamiento y generó predicciones sobre la producción de maíz en toneladas.

Calibración de hiperparámetros

Con el fin de mejorar el desempeño del modelo, se aplicó un proceso de calibración mediante GridSearchCV. Esta técnica permitió evaluar diferentes combinaciones de

hiperparámetros y seleccionar la configuración con mejor desempeño dentro del conjunto de entrenamiento.

La rejilla de parámetros evaluada incluyó:

- `n_estimators`: 50, 100 y 200.
- `max_depth`: None, 10 y 20.
- `min_samples_split`: 2 y 5.
- Validación cruzada: `cv = 3`.

Después de esta calibración, se obtuvo el modelo final optimizado, el cual fue utilizado para calcular las métricas finales y realizar la comparación con un método tradicional de estimación. En el código, la calibración se realizó mediante `GridSearchCV` y posteriormente se calculó el MAPE sobre las predicciones del modelo final.

Validación estadística del modelo

La validación del modelo se realizó mediante tres métricas principales:

- R^2 : permitió identificar qué proporción de la variabilidad de la producción de maíz fue explicada por el modelo.
- RMSE: permitió medir el error promedio de predicción en toneladas, penalizando con mayor fuerza los errores grandes.
- MAPE: permitió expresar el error de predicción en términos porcentuales, facilitando la interpretación del desempeño del modelo.

Además, se realizó una comparación frente a un método tradicional de estimación basado en una media móvil de tres años. Esta comparación permitió evaluar si el algoritmo Random Forest ofrecía una mejora real frente a una técnica simple de pronóstico. Para ello, se calculó el RMSE del método simple y se comparó con el RMSE del modelo Random Forest, obteniendo posteriormente el porcentaje de reducción del error.

Análisis de importancia de variables

Como parte del análisis del modelo, se calculó la importancia de las variables mediante el atributo `feature_importances_` del modelo Random Forest. Este procedimiento permitió identificar cuáles variables tuvieron mayor influencia en la predicción de la producción de maíz.

Las variables analizadas fueron: Año, Municipio_Enc, Tecnica_Enc, Área sembrada (ha), Rendimiento (t/ha) y Área cosechada (ha). Este análisis fue relevante porque permitió interpretar el modelo más allá del valor final de predicción, identificando los factores que más aportan a la explicación del comportamiento productivo del maíz en el Meta.

Proyección de la producción de maíz 2025–2030

La proyección futura se realizó para el periodo 2025–2030, utilizando como base la información histórica consolidada hasta 2024. Para cada año futuro, se generó un escenario con las mismas variables requeridas por el modelo y se aplicó el algoritmo entrenado para estimar la producción esperada.

Posteriormente, se construyó una proyección dinámica diferenciada por sistema de producción. Para el maíz tecnificado se aplicó una tasa de crecimiento anual del 1,5 % sobre el área sembrada, mientras que para el maíz tradicional se aplicó una tasa de disminución anual del 1,0 %. Estos valores fueron utilizados como supuestos de simulación para analizar posibles escenarios de cambio en la estructura productiva del cultivo.

Los resultados fueron consolidados en una tabla comparativa que integra la producción histórica 2006–2024 y la producción proyectada 2025–2030. Esta tabla permite observar la evolución del maíz tecnificado y tradicional, facilitando el análisis de tendencias y la interpretación de sus implicaciones para la planeación agrícola, logística y productiva del departamento del Meta.

Síntesis metodológica

En síntesis, la metodología permitió transformar una base histórica de producción agrícola en una herramienta predictiva aplicada al contexto del maíz en el Meta. El proceso inició con la recolección y depuración de datos oficiales, continuó con la codificación de variables y el entrenamiento del modelo Random Forest, y finalizó con la validación estadística, el análisis de importancia de variables y la proyección de producción a cinco años.

Esta ruta metodológica responde directamente a los objetivos del proyecto, ya que permitió organizar la información histórica, diseñar y validar un modelo predictivo, y generar una proyección útil para apoyar la toma de decisiones en la planeación agrícola, logística y productiva regional.

Desarrollo de los objetivos

Desarrollo del objetivo específico 1: Recopilación, depuración y organización de los datos históricos

El primer objetivo específico consistió en recopilar, depurar y organizar los datos históricos de producción de maíz del departamento del Meta, correspondientes al periodo 2006–2024, a partir de la información disponible en la Unidad de Planificación Rural Agropecuaria (UPRA). Esta fase fue fundamental para construir una base de datos estructurada que permitiera aplicar posteriormente el modelo predictivo.

Inicialmente, la información fue cargada en Python desde un archivo en formato Excel, utilizando el entorno de Google Colaboratory y la conexión con Google Drive. Para el procesamiento de la base se empleó la librería Pandas, la cual permitió revisar la estructura inicial del archivo, identificar las columnas disponibles y consolidar los registros necesarios para el análisis.

La base fue organizada a partir de tres variables principales de identificación: año, municipio y desagregación del cultivo. Posteriormente, se agruparon los registros por estas variables y se sumaron los valores correspondientes al área sembrada, área cosechada y producción. Este procedimiento permitió obtener un único registro por año, municipio y sistema productivo, reduciendo duplicidades y facilitando el análisis comparativo entre maíz tecnificado y maíz tradicional.

Después de la agrupación, se recalculó el rendimiento en toneladas por hectárea mediante la relación entre la producción total y el área cosechada. Este paso fue necesario porque, al consolidar los registros, el rendimiento debía ajustarse nuevamente para mantener coherencia matemática dentro de la base final. De esta manera, el rendimiento quedó expresado como una variable productiva clave para el entrenamiento del modelo.

Adicionalmente, se realizó la homologación de la variable “Desagregación cultivo”, con el fin de consolidar las categorías en dos grandes sistemas productivos: maíz tecnificado y maíz tradicional. Esta clasificación permitió analizar el comportamiento diferencial de ambos sistemas y preparar la base para la proyección futura de la producción.

Como resultado del primer objetivo, se obtuvo una base de datos depurada, organizada y

estructurada para el análisis predictivo, integrada por variables de tiempo, localización, sistema productivo, área sembrada, área cosechada, rendimiento y producción. Esta base constituyó el insumo principal para el diseño y entrenamiento del algoritmo.

Desarrollo del objetivo específico 2: Diseño, entrenamiento y validación del modelo predictivo

El segundo objetivo específico se orientó al diseño y validación de un modelo de aprendizaje automático basado en Random Forest, con el fin de predecir la producción de maíz en el departamento del Meta. Para ello, se seleccionaron como variables de entrada el año, el municipio codificado, la técnica de cultivo codificada, el área sembrada, el área cosechada y el rendimiento. La variable objetivo del modelo fue la producción de maíz expresada en toneladas.

Debido a que el algoritmo requiere variables numéricas para su procesamiento, se aplicó la técnica de Label Encoding a las variables categóricas “Municipio” y “Desagregación cultivo”. Como resultado, se generaron las variables Municipio_Enc y Tecnica_Enc, permitiendo que el modelo interpretara la información territorial y tecnológica dentro del proceso de predicción.

El modelo utilizado fue Random Forest Regressor, seleccionado por su capacidad para trabajar con relaciones no lineales entre variables. Esta característica resulta pertinente en el contexto agrícola, debido a que la producción de maíz no depende únicamente de una variable aislada, sino de la interacción entre el área sembrada, el área cosechada, el rendimiento, el municipio y el sistema productivo.

Para el entrenamiento del modelo, la base de datos fue dividida en dos subconjuntos: 80 % para entrenamiento y 20 % para prueba. Esta división permitió entrenar el algoritmo con una parte de la información histórica y evaluar posteriormente su desempeño con datos no utilizados durante el entrenamiento. Además, se utilizó una semilla aleatoria de 42 para garantizar la reproducibilidad del proceso.

Con el fin de mejorar el desempeño del modelo, se aplicó una calibración mediante GridSearchCV. En esta etapa se evaluaron diferentes combinaciones de hiperparámetros, incluyendo el número de árboles del bosque, la profundidad máxima de los árboles y el número mínimo de muestras requeridas para dividir un nodo. Los valores evaluados fueron 50, 100 y 200 estimadores; profundidades máximas de None, 10 y 20; y valores de min_samples_split de 2 y 5.

Este procedimiento permitió seleccionar una configuración más adecuada para el comportamiento de los datos.

La validación del modelo se realizó mediante tres métricas principales: RMSE, MAPE y R^2 . El RMSE permitió medir el error promedio de predicción en toneladas; el MAPE permitió expresar el error en términos porcentuales; y el coeficiente R^2 permitió evaluar la proporción de la variabilidad de la producción explicada por el modelo.

Tabla 1

Métricas de error y precisión

Métrica	Pronóstico Simple (Media Móvil)	Algoritmo Propuesto (Random Forest)	Mejora porcentual (%)
Error (RMSE)	117,352.20 toneladas	2,454.78 toneladas	97.91%
Error Relativo (MAPE)	35.00%	7.90%	77.42%
Coeficiente R^2	N/A	0.9943	N/A

Los resultados evidencian que el modelo Random Forest presentó un mejor desempeño frente al método simple de media móvil de tres años. Mientras la media móvil obtuvo un RMSE de 117.352,20 toneladas, el algoritmo propuesto alcanzó un RMSE de 2.454,78 toneladas. Esto representa una reducción del error del 97,91 %, lo cual demuestra que la incorporación de variables como municipio, sistema productivo, área cosechada y rendimiento mejora la capacidad de estimación frente a un método basado únicamente en el comportamiento histórico reciente.

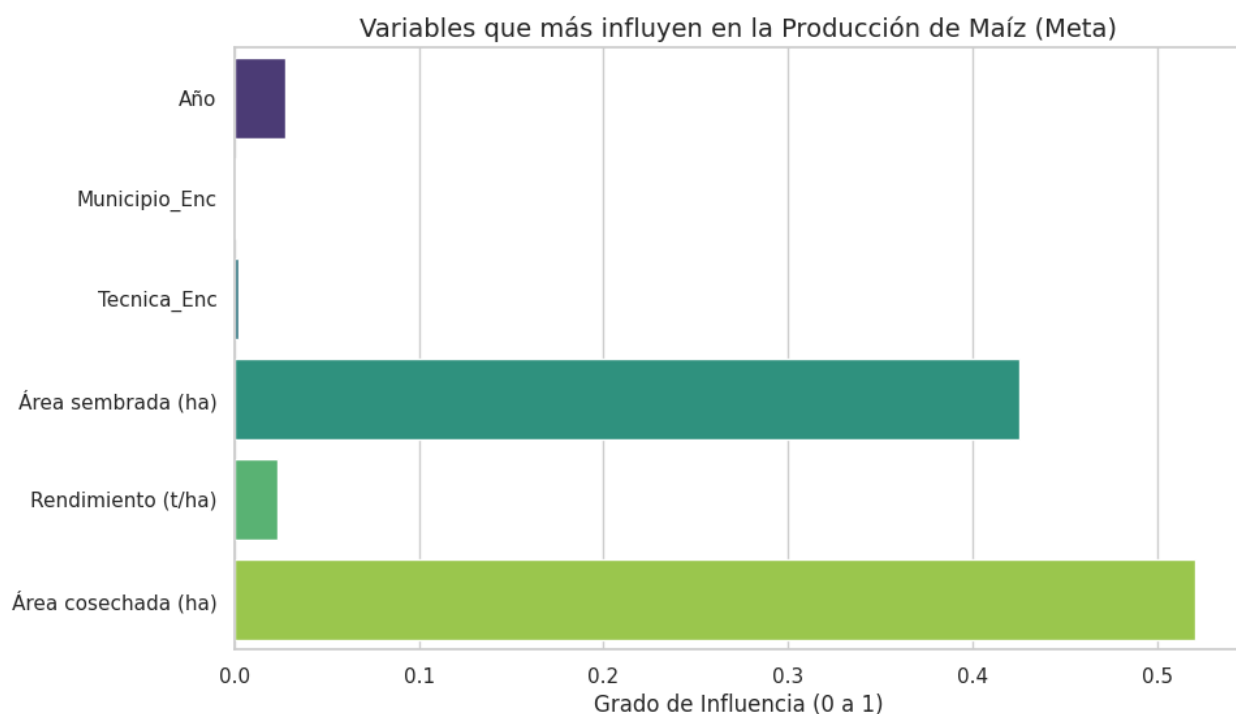
Asimismo, el MAPE del modelo fue de 7,90 %, lo que indica un error porcentual relativamente bajo para la estimación de producción agrícola. Por su parte, el coeficiente R^2 de 0,9943 muestra que el modelo logró explicar una proporción elevada de la variabilidad de la producción de maíz en la base analizada.

Como complemento a la validación, se realizó un análisis de importancia de variables mediante el atributo `feature_importances_` del modelo Random Forest. Este análisis permitió identificar cuáles variables tuvieron mayor influencia en la predicción de la producción. Los resultados mostraron que el área cosechada fue la variable de mayor peso predictivo, seguida por el rendimiento y el año. Esto confirma que la producción no depende únicamente del área

sembrada, sino de la superficie que efectivamente logra ser cosechada y de la eficiencia productiva obtenida por hectárea.

Figura 2

Variables influyentes en la producción de maíz



Desde la perspectiva de la ingeniería industrial, este resultado es relevante porque permite orientar las decisiones hacia factores que tienen mayor impacto en la producción. En este sentido, proteger el área cosechada, mejorar el rendimiento y fortalecer los sistemas tecnificados se convierten en aspectos clave para la planeación agrícola del departamento del Meta.

Desarrollo del objetivo específico 3: Proyección y análisis de la producción de maíz a cinco años

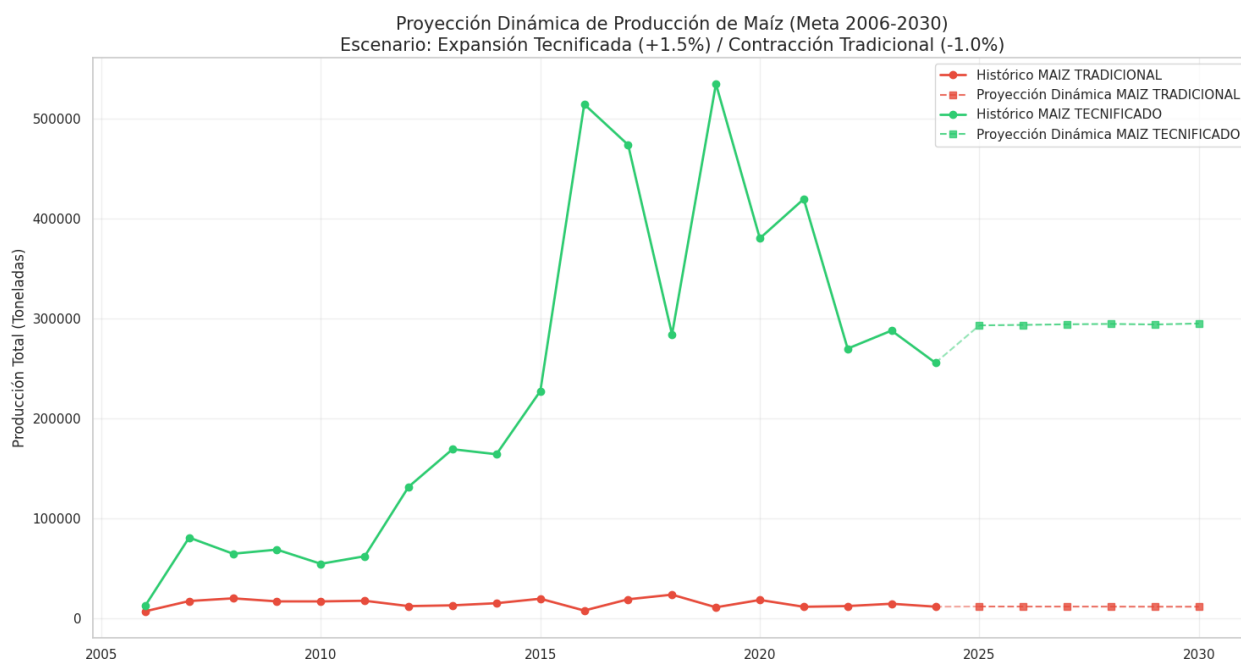
El tercer objetivo específico consistió en proyectar y analizar la producción de maíz en el departamento del Meta a cinco años, con el propósito de generar información útil para la toma de decisiones en la planeación agrícola, logística y productiva regional.

La proyección se realizó para el periodo 2025–2030, tomando como base la información histórica consolidada hasta el año 2024. A partir del modelo entrenado, se generaron escenarios futuros para cada año proyectado, manteniendo la estructura de variables utilizada durante el entrenamiento del algoritmo.

Además de la proyección inicial, se construyó un escenario dinámico diferenciado por sistema productivo. Para el maíz tecnificado se asumió una tasa de crecimiento anual del 1,5 % sobre el área sembrada, mientras que para el maíz tradicional se asumió una disminución anual del 1,0 %. Estos supuestos permitieron simular una posible transición productiva hacia sistemas más tecnificados y analizar su efecto sobre la producción futura.

Figura 3

Análisis de tendencia proyectada



La tendencia proyectada muestra que el maíz tecnificado mantiene una producción superior frente al maíz tradicional durante el periodo analizado. Aunque la producción tecnificada presenta variaciones históricas importantes, la proyección 2025–2030 muestra una tendencia relativamente estable y levemente creciente. En contraste, el maíz tradicional se mantiene en niveles considerablemente más bajos y con una ligera tendencia decreciente.

Tabla 2*Consolidado de producción estimada (2006-2030)*

Año	Tipo	MAIZ TECNIFICADO (ton)	MAIZ TRADICIONAL (ton)
2006	HISTÓRICO	12961.00	7425.00
2007	HISTÓRICO	81170.00	17669.00
2008	HISTÓRICO	64999.00	20381.00
2009	HISTÓRICO	69114.00	17335.00
2010	HISTÓRICO	54914.00	17315.00
2011	HISTÓRICO	62502.00	17963.00
2012	HISTÓRICO	131989.00	12598.00
2013	HISTÓRICO	169569.00	13367.00
2014	HISTÓRICO	164514.00	15508.00
2015	HISTÓRICO	227761.00	19987.00
2016	HISTÓRICO	514628.00	8164.00
2017	HISTÓRICO	474018.00	19422.00
2018	HISTÓRICO	284387.00	24054.00
2019	HISTÓRICO	535015.99	11404.83
2020	HISTÓRICO	380502.35	18761.49
2021	HISTÓRICO	419848.02	11923.25
2022	HISTÓRICO	270192.19	12646.97
2023	HISTÓRICO	288215.60	14968.14
2024	HISTÓRICO	255960.23	12015.38
2025	PROYECCIÓN	293446.29	12207.18
2026	PROYECCIÓN	293917.82	12143.89
2027	PROYECCIÓN	294509.75	12134.75
2028	PROYECCIÓN	294897.12	12099.65
2029	PROYECCIÓN	294333.65	12024.73
2030	PROYECCIÓN	295312.48	12012.09

Los resultados proyectados permiten observar que, para el año 2030, la producción de maíz tecnificado alcanzaría aproximadamente 295.312,48 toneladas, mientras que la producción de maíz tradicional se ubicaría alrededor de 12.012,09 toneladas. Esta diferencia evidencia la importancia del sistema tecnificado dentro de la oferta futura de maíz en el departamento del Meta.

Desde el punto de vista de la ingeniería industrial, estos resultados no representan únicamente una estimación productiva, sino una herramienta de apoyo para la planeación de recursos. Una proyección de producción permite anticipar necesidades de almacenamiento, transporte, secado, abastecimiento de insumos, programación de maquinaria y organización logística. En este sentido, el modelo puede contribuir a que productores, asociaciones y entidades

del sector agrícola tomen decisiones más informadas frente al comportamiento futuro del cultivo.

En términos estratégicos, el crecimiento proyectado del maíz tecnificado sugiere la necesidad de fortalecer la infraestructura de almacenamiento y secado, así como mejorar la coordinación logística para movilizar la producción hacia centros de consumo y transformación. De igual manera, las empresas proveedoras de semillas, fertilizantes y servicios agrícolas podrían utilizar este tipo de proyección para dimensionar inventarios y planear su oferta con mayor anticipación.

Síntesis del desarrollo de los objetivos

El desarrollo de los objetivos permitió comprobar que el proyecto avanzó desde la organización de una base histórica oficial hasta la construcción de una herramienta predictiva aplicada al cultivo de maíz en el departamento del Meta. Primero, se recopilaron y depuraron los datos de la UPRA para el periodo 2006–2024. Luego, se diseñó y validó un modelo Random Forest en Python, evaluado mediante métricas de desempeño como RMSE, MAPE y R^2 . Finalmente, se generó una proyección de la producción de maíz para el periodo 2025–2030, diferenciando entre maíz tecnificado y tradicional.

Narración del caso

La construcción del algoritmo predictivo surgió a partir de la necesidad de transformar los datos históricos de producción de maíz del departamento del Meta en una herramienta útil para la planeación agrícola, logística y productiva. Durante la revisión de la información disponible en las bases de la UPRA/EVA, se identificó que la producción de maíz no presenta un comportamiento uniforme, sino que varía según el año, el municipio, el área sembrada, el área cosechada, el rendimiento y el sistema de producción utilizado.

En la primera etapa del caso se evidenció que el departamento del Meta cuenta con registros de producción distribuidos en sus municipios, los cuales presentan comportamientos productivos diferentes. Algunos territorios mantienen una participación asociada a sistemas tradicionales, mientras que otros municipios muestran una mayor presencia de sistemas tecnificados. Esta diferencia fue relevante para el análisis, ya que permitió comprender que la producción de maíz no debía estudiarse como un dato general del departamento, sino como el resultado de condiciones territoriales y productivas diferenciadas.

Uno de los principales retos del caso fue la organización y depuración de los microdatos históricos. La información debía ser consolidada para obtener registros comparables por año, municipio y desagregación del cultivo. Además, fue necesario homologar las categorías de producción en dos grandes grupos: maíz tecnificado y maíz tradicional. Este proceso permitió estructurar una base de datos más coherente para el entrenamiento del modelo predictivo.

Otro aspecto importante fue la identificación de la variabilidad del rendimiento en toneladas por hectárea. Durante la exploración de los datos se observó que el rendimiento no se comporta de manera constante, sino que cambia de acuerdo con el municipio, el periodo analizado y el sistema de producción. Esta situación justificó la necesidad de utilizar un modelo capaz de reconocer relaciones no lineales entre variables, en lugar de depender únicamente de métodos tradicionales basados en tendencias históricas simples.

A partir de este análisis, el caso permitió establecer que la dificultad principal no se encontraba solamente en la existencia de datos históricos, sino en la falta de una herramienta que permitiera convertirlos en proyecciones útiles para la toma de decisiones. Por ello, se desarrolló un algoritmo en Python basado en Random Forest Regressor, integrando variables como año, municipio, técnica de cultivo, área sembrada, área cosechada y rendimiento, con el fin de predecir la producción de maíz a cinco años y apoyar la planeación agrícola del departamento del Meta.

Lecciones y recomendaciones

Lecciones

- **Validación del Modelo:** El desarrollo del algoritmo basado en Random Forest Regressor permitió comprobar que los modelos de aprendizaje automático pueden mejorar la predicción de la producción de maíz frente a métodos tradicionales. En la comparación realizada, el modelo propuesto redujo el error cuadrático medio RMSE en un 97,91 % frente a una media móvil simple, lo que evidencia la importancia de incorporar variables como área cosechada, rendimiento, municipio y sistema productivo.
- **Precisión Predictiva:** El modelo alcanzó un coeficiente de determinación R^2 de 0,9943 y un MAPE de 7,90 %. Estos resultados muestran un desempeño favorable del algoritmo para estimar la producción de maíz en la base analizada. Sin embargo, su confiabilidad depende de la calidad, actualización y continuidad de los datos históricos utilizados.
- **Determinantes de Producción:** El análisis de importancia de variables permitió identificar que el área cosechada fue el factor con mayor peso en la predicción, seguida por el rendimiento y el año. Esto indica que la producción no depende únicamente de la superficie sembrada, sino de la capacidad de convertir esa área en cosecha efectiva y de la eficiencia productiva alcanzada por hectárea.
- **Prospectiva al 2030:** La proyección dinámica permitió observar un comportamiento diferenciado entre el maíz tecnificado y el maíz tradicional. Bajo los supuestos establecidos en el modelo, el sistema tecnificado presenta una tendencia más favorable hacia 2030, mientras que el sistema tradicional mantiene niveles más bajos y una tendencia de menor crecimiento. Esto refuerza la importancia de analizar la producción de maíz según el sistema productivo y no únicamente como un total departamental.

Recomendaciones

- **Integración del algoritmo en herramientas de gestión:** Se recomienda que asociaciones de productores, entidades públicas y actores del sector agrícola evalúen la posibilidad de integrar este tipo de algoritmo en tableros de control o sistemas de información. Esto

permitiría utilizar las proyecciones como apoyo para la planeación de siembras, compra de insumos, transporte, almacenamiento y comercialización.

- **Enfoque en el Rendimiento:** Dado que el rendimiento fue una de las variables más influyentes en el modelo, se recomienda fortalecer programas de asistencia técnica, transferencia tecnológica y mejoramiento de prácticas productivas. Esto puede contribuir a aumentar la producción sin depender exclusivamente de la expansión del área sembrada.
- **Actualización de Datos:** Se recomienda alimentar el algoritmo cada año con los nuevos cierres de las Evaluaciones Agropecuarias Municipales. Esta actualización permitiría que el modelo incorpore cambios recientes en el comportamiento del cultivo y mantenga su utilidad para la toma de decisiones.
- **Expansión del Modelo:** Se recomienda que, en futuras investigaciones, el modelo incorpore variables climáticas, económicas y de mercado, como precipitación, temperatura, precios, costos de insumos y disponibilidad de infraestructura. Esto permitiría fortalecer la capacidad predictiva y ampliar el alcance del algoritmo.
- **Uso de la proyección para planeación logística:** Los resultados proyectados deben utilizarse como insumo para anticipar necesidades de almacenamiento, secado, transporte y abastecimiento de insumos. Desde la ingeniería industrial, esta información puede apoyar una mejor organización de los recursos asociados a la cadena productiva del maíz en el Meta.

Bibliografía

- Ane, T. (2024). A review of machine learning applications and their predictive solutions in agriculture. *Asian Journal of Advances in Agricultural Research*, 24(7), 80–90. <https://doi.org/10.9734/ajaar/2024/v24i7525>
- Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151, 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>
- Desloires, J., Ienco, D., & Botrel, A. (2024). Early season forecasting of corn yield at field level from multi-source satellite time series data. *Remote Sensing*, 16(9), 1573. <https://doi.org/10.3390/rs16091573>
- Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023). Crop prediction model using machine learning algorithms. *Applied Sciences*, 13(16), 9288. <https://doi.org/10.3390/app13169288>
- García, J., Riaño, N., & Magnitskiy, S. (2014). Simulation of corn (*Zea mays* L.) production in different agricultural zones of Colombia using the AquaCrop model. *Agronomía Colombiana*, 32(3), 358–366. <https://doi.org/10.15446/agron.colomb.v32n3.45939>
- Hartanto, W., Mu'alimah, S., & Sukarniati, L. (2024). Economic and social determinants of corn farmers' consumption patterns: A survey and regression analysis approach. *JAMPE (Journal of Asset Management and Public Economy)*, 3(2), 76–97. <https://doi.org/10.12928/jampe.v3i2.10423>
- Kerdprasop, K., Chuaybamroong, P., & Kerdprasop, N. (2023). Food production forecasting with time series and ensemble modeling methods. En *Proceedings of the 2022 5th Artificial Intelligence and Cloud Computing Conference* (pp. 59–65). Association for Computing Machinery. <https://doi.org/10.1145/3582099.3582109>
- Martínez-Reina, A. (2013). Caracterización socioeconómica de los sistemas de producción de la región de La Mojana en el Caribe de Colombia. *Ciencia y Tecnología Agropecuaria*, 14(2), 165–185. https://doi.org/10.21930/rcta.vol14_num2_art:406
- Medida, L. H., Khang, A., & Kumar, G. L. N. V. S. (2023). Predictive analytics for high-tech agriculture. En *Handbook of research on AI-equipped IoT applications in high-tech agriculture* (pp. 336–346). IGI Global. <https://www.igi->

- global.com/gateway/chapter/327844
- Ministerio de Agricultura y Desarrollo Rural. (2013). *Ficha metodológica: Evaluaciones Agropecuarias Municipales*. Agronet. <https://agronet.gov.co/sites/default/files/estadistica/Documents/FichaMetodologica-Reporte%201-103.pdf>
- Mncube, T. L., Phiri, E. E., Mothapo, P. N., Rugare, J. T., Pieterse, P. J., & Mloza-Banda, H. R. (2024). Characterising productivity factors affecting maize (*Zea mays*) production in a smallholder crop-livestock system. *Agricultural Research*, 13(1), 124–136. <https://doi.org/10.1007/s40003-023-00674-7>
- Musanase, C., Vodacek, A., Hanyurwimfura, D., Uwitonze, A., & Kabandana, I. (2023). Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices. *Agriculture*, 13(11), 2141. <https://doi.org/10.3390/agriculture13112141>
- Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). (2013). *Review of crop yield forecasting methods and early warning systems*. https://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/GS_SAC_2013/Improving_methods_for_crops_estimates/Crop_Yield_Forecasting_Methods_and_Early_Warning_Systems_Lit_review.pdf
- Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), Fondo Internacional de Desarrollo Agrícola (FIDA), Organización Mundial de la Salud (OMS), Programa Mundial de Alimentos (PMA), & Fondo de las Naciones Unidas para la Infancia (UNICEF). (2019). *El estado de la seguridad alimentaria y la nutrición en el mundo 2019: Protegerse frente a la desaceleración y el debilitamiento de la economía*. <https://openknowledge.fao.org/handle/20.500.14283/ca5162es>
- Organización para la Cooperación y el Desarrollo Económico (OCDE) & Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). (2024). *OCDE-FAO perspectivas agrícolas 2024–2033*. <https://openknowledge.fao.org/items/861e9907-b0da-4aa9-94f0-cbb3014eb26d>
- Onuwa, G., Mailumo, S. S., & Oyewole, S. O. (2023). Socio-economic determinants of adoption of maize production technologies among smallholders. *AGRIEKONOMIKA*, 12(1), 83–94. <https://doi.org/10.21107/agriekonomika.v12i1.14621>

- Ordoñez, L., Vallejo, E., Amariles, D., Mesa, J., Esquivel, A., Llanos-Herrera, L., Prager, S. D., Segura, C., Valencia, J. J., Duarte, C. J., Rojas, D. C., Obando, D., & Ramirez-Villegas, J. (2022). Applying agroclimatic seasonal forecasts to improve rainfed maize agronomic management in Colombia. *Climate Services*, 28, 100333. <https://doi.org/10.1016/j.cliser.2022.100333>
- Selvaraju, R. (2013). *Climate risk assessment and management in agriculture*. Organización de las Naciones Unidas para la Alimentación y la Agricultura. <https://www.fao.org/4/i3084e/i3084e06.pdf>
- Şimon, A., Moraru, P. I., Ceclan, A., Russu, F., Cheţan, F., Bărdaş, M., Popa, A., Rusu, T., Pop, A. I., & Bogdan, I. (2023). The impact of climatic factors on the development stages of maize crop in the Transylvanian Plain. *Agronomy*, 13(6), 1612. <https://doi.org/10.3390/agronomy13061612>
- Sowmya, P., & Prasad, D. A. V. K. (2024). Optimization of crop yield prediction models using hyperparameter tuning and ensemble learning. *Nanotechnology Perceptions*, 20(S7), 653–665. <https://doi.org/10.62441/nano-ntp.v20iS7.56>
- Unidad de Planificación Rural Agropecuaria (UPRA). (2023). *Microanálisis Evaluaciones Agropecuarias 2023: Meta*. https://upra.gov.co/Kit_Territorial/2-%20Informaci%C3%B3n%20por%20Departamentos/META/3-%20Microan%C3%A1lisis%20Evaluaciones%20agropecuarias%202023-Meta.pdf
- Unidad de Planificación Rural Agropecuaria (UPRA). (2024a). *Boletín SIGRA: Maíz*. https://upra.gov.co/sites/default/files/2024-10/BolSagra_Maiz.pdf
- Unidad de Planificación Rural Agropecuaria (UPRA). (2024b). *EVA 2024: Evaluaciones Agropecuarias Municipales*. <https://upra.gov.co/es-co/eva/eva-2024>
- Unidad de Planificación Rural Agropecuaria (UPRA). (2025). *Resultados evaluaciones agropecuarias 2023*. <https://upra.gov.co/sites/default/files/2025-04/Resultados%20Evaluaciones%20Agropecuarias%202023.pdf>
- van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709. <https://doi.org/10.1016/j.compag.2020.105709>

Anexos

Código Python

```
# =====  
# 1. IMPORTACIÓN DE LIBRERÍAS  
# =====  
  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
# Herramientas de Machine Learning  
from sklearn.ensemble import RandomForestRegressor  
from sklearn.model_selection import train_test_split  
from sklearn.metrics import mean_squared_error, r2_score  
from sklearn.preprocessing import LabelEncoder  
  
# Conexión a Google Drive  
from google.colab import drive  
  
# Configuración de gráficas  
%matplotlib inline  
sns.set_theme(style="whitegrid")  
  
print("✓ Librerías cargadas exitosamente.")  
  
# =====  
# 2. CONEXIÓN A GOOGLE DRIVE  
# =====  
  
drive.mount('/content/drive')  
  
# Cambia esta ruta por la ubicación real de tu archivo en Drive
```

```
ruta_archivo = '/content/drive/MyDrive/TESIS NICOLAS/DATOS/MAIZ_META.xlsx'

# Carga del dataset
# Nota: Si tu archivo es Excel, usa pd.read_excel(ruta_archivo)
data_maiz_meta = pd.read_excel(ruta_archivo)

# Visualización inicial para confirmar que todo está bien
print(f'Dataset cargado con {data_maiz_meta.shape[0]} registros y {data_maiz_meta.shape[1]}
columnas.")
data_maiz_meta.head()

# =====
# 3. AGREGACIÓN Y SUMATORIA POR TÉCNICA
# =====

# Definimos las columnas que identifican un registro único
# (Año, Municipio y la técnica que tú ya limpiaste)
agrupadores = ['Año', 'Municipio', 'Desagregación cultivo'] # Asegúrate que tu columna se llame 'Tecnica'

# Sumamos las áreas y la producción para consolidar el municipio/año
df_unificado = data_maiz_meta.groupby(agrupadores).agg({
    'Área sembrada (ha)': 'sum',
    'Área cosechada (ha)': 'sum',
    'Producción (t)': 'sum'
}).reset_index()

# Recalculamos el rendimiento para que sea exacto tras la suma
df_unificado['Rendimiento (t/ha)'] = df_unificado['Producción (t)'] / df_unificado['Área cosechada (ha)']

print("✓ Datos agrupados. Ahora tienes un solo valor de Producción por técnica/municipio/año.")
print(f'Registros finales: {df_unificado.shape[0]}')
df_unificado.head()

# =====
```

4. CODIFICACIÓN DE VARIABLES (LABEL ENCODING)

=====

le_municipio = LabelEncoder()

le_tecnica = LabelEncoder()

df_unificado['Municipio_Enc'] = le_municipio.fit_transform(df_unificado['Municipio'])

df_unificado['Tecnica_Enc'] = le_tecnica.fit_transform(df_unificado['Desagregación cultivo'])

Guardamos los nombres de las técnicas para usarlos después en la gráfica

nombres_tecnicas = le_tecnica.classes_

print("Categorías codificadas:")

for i, nombre in enumerate(nombres_tecnicas):

print(f"Código {i}: {nombre}")

=====

5. DEFINICIÓN DE VARIABLES Y ENTRENAMIENTO

=====

Definimos las características (X) y lo que queremos predecir (y)

Usamos el nombre exacto de tu columna ya codificada

columnas_X = ['Año', 'Municipio_Enc', 'Tecnica_Enc', 'Área sembrada (ha)', 'Rendimiento (t/ha)', 'Área cosechada (ha)']

X = df_unificado[columnas_X]

y = df_unificado['Producción (t)']

Dividimos los datos (80% para aprender, 20% para probar)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=42)

Creamos el Bosque Aleatorio (Random Forest)

n_estimators=100 significa que usaremos 100 "árboles" de decisión para promediar

modelo_rf = RandomForestRegressor(n_estimators=100, random_state=42)

```

# ¡A entrenar!
modelo_rf.fit(X_train, y_train)

print("✓ Modelo entrenado exitosamente usando la columna 'Desagregación cultivo'.")

# =====
# 6. RESULTADO DE PREDICCIÓN
# =====

# Realizamos predicciones sobre el grupo de prueba
y_pred = modelo_rf.predict(X_test)

# Calculamos métricas
r2 = r2_score(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

print(f"📊 Desempeño del Modelo:")
print(f"R² (Precisión): {r2:.4f}")
print(f"Error Promedio (RMSE): {rmse:.2f} toneladas")

# =====
# 7. GRÁFICA COMPARATIVA POR TÉCNICA
# =====

plt.figure(figsize=(14, 7))

# Obtenemos las técnicas únicas (Tecnificado y Tradicional)
tecnicas = df_unificado['Desagregación cultivo'].unique()
colores = ['#2ecc71', '#e74c3c'] # Verde para Tecnificado, Rojo para Tradicional (o los que prefieras)

for i, tecnica in enumerate(tecnicas):
    # Filtramos los datos por técnica
    df_tec = df_unificado[df_unificado['Desagregación cultivo'] == tecnica]

```

```
# Agrupamos por año para ver el total departamental de esa técnica
serie_tiempo = df_tec.groupby('Año')['Producción (t)'].sum()

# Graficamos
plt.plot(serie_tiempo.index, serie_tiempo.values,
         marker='o', linestyle='-', linewidth=2,
         label=f'Real: {tecnic} ', color=colores[i])

# Personalización de la gráfica
plt.title('Producción de Maíz en el Meta: Tecnificado vs Tradicional (Histórico Real)', fontsize=15)
plt.xlabel('Año', fontsize=12)
plt.ylabel('Producción Total (Toneladas)', fontsize=12)
plt.xticks(df_unificado['Año'].unique(), rotation=45)
plt.legend(title="Técnica de Cultivo")
plt.grid(True, linestyle='--', alpha=0.6)

plt.tight_layout()
plt.show()

# =====
# 8. PROYECCIÓN FUTURA (2025 - 2030)
# =====

años_futuros = np.arange(2025, 2031)
predicciones_lista = []

# Obtenemos la última base de datos real para proyectar sobre ella
# Usamos el último año disponible en tu dataset
ultimo_año = df_unificado['Año'].max()
base_proyeccion = df_unificado[df_unificado['Año'] == ultimo_año].copy()

for año in años_futuros:
    # Creamos el escenario para el año X
```

```

escenario = base_proyeccion.copy()
escenario['Año'] = año

# Seleccionamos las columnas que el modelo conoce (X)
X_futuro = escenario[['Año', 'Municipio_Enc', 'Tecnica_Enc', 'Área sembrada (ha)', 'Rendimiento
(t/ha)', 'Área cosechada (ha)']]

# El modelo predice la producción de cada municipio y técnica
escenario['Producción_Predicha'] = modelo_rf.predict(X_futuro)

# Guardamos los resultados
predicciones_lista.append(escenario)

# Unimos todas las predicciones en un solo DataFrame
df_proyeccion = pd.concat(predicciones_lista)

print(f'✓ Proyección generada desde 2025 hasta 2030.")

# =====
# 9. TABLA RESUMEN: HISTÓRICO + PROYECCIÓN
# =====

# 1. Preparar datos históricos
tabla_hist = df_unificado.groupby(['Año', 'Desagregación cultivo'])['Producción (t)'].sum().reset_index()
tabla_hist.rename(columns={'Producción (t)': 'Producción_t'}, inplace=True)
tabla_hist['Tipo'] = 'HISTÓRICO'

# 2. Preparar datos proyectados
tabla_proj = df_proyeccion.groupby(['Año', 'Desagregación
cultivo'])['Producción_Predicha'].sum().reset_index()
tabla_proj.rename(columns={'Producción_Predicha': 'Producción_t'}, inplace=True)
tabla_proj['Tipo'] = 'PROYECCIÓN'

# 3. Concatenar ambos y organizar la tabla

```

```
tabla_final = pd.concat([tabla_hist, tabla_proj]).sort_values(by=['Año', 'Desagregación cultivo'])
```

```
# Pivotar la tabla para que sea más fácil de leer (Años en filas, Técnicas en columnas)
```

```
tabla_comparativa = tabla_final.pivot_table(
```

```
    index=['Año', 'Tipo'],
```

```
    columns='Desagregación cultivo',
```

```
    values='Producción_t'
```

```
).reset_index()
```

```
# Llenar posibles valores nulos con 0 y redondear para estética
```

```
tabla_comparativa = tabla_comparativa.fillna(0).round(2)
```

```
# Mostrar la tabla
```

```
print("TABLA DE PRODUCCIÓN DE MAÍZ EN EL META (TONELADAS)")
```

```
print("=====")
```

```
display(tabla_comparativa)
```

```
# Opcional: Exportar a Excel para tu documento
```

```
# tabla_comparativa.to_excel('Proyeccion_Maiz_Meta_2030.xlsx', index=False)
```

```
# =====
```

```
# 8. NUEVA PROYECCIÓN DINÁMICA (2025 - 2030)
```

```
# =====
```

```
# Definimos las tasas de cambio de área (Ingeniería Industrial)
```

```
TASA_TECNIFICADO = 0.015 # +1.5% anual (expansión de frontera)
```

```
TASA_TRADICIONAL = -0.010 # -1.0% anual (migración tecnológica)
```

```
años_futuros = np.arange(2025, 2031)
```

```
predicciones_dinamicas = []
```

```
# Usamos el último año real como base (2024)
```

```
ultimo_año = df_unificado['Año'].max()
```

```
base_2024 = df_unificado[df_unificado['Año'] == ultimo_año].copy()
```

```

print("🌀 Iniciando proyección dinámica con tasas diferenciadas...")

for año in años_futuros:
    # Creamos el escenario para el nuevo año
    escenario = base_2024.copy()
    escenario['Año'] = año

    # Calculamos cuántos años han pasado desde la base (2024)
    n_años = año - ultimo_año

    # --- APLICAMOS EL CAMBIO DE ÁREA (LA CLAVE) ---
    # Usamos la fórmula de interés compuesto: Valor_Futuro = Valor_Actual * (1 + tasa)^n

    # Para Tecnificado (Crece)
    mask_tec = escenario['Desagregación cultivo'] == 'MAIZ TECNIFICADO'
    escenario.loc[mask_tec, 'Área sembrada (ha)'] = \
        escenario.loc[mask_tec, 'Área sembrada (ha)'] * ((1 + TASA_TECNIFICADO) ** n_años)

    # Para Tradicional (Decrece)
    mask_trad = escenario['Desagregación cultivo'] == 'MAIZ TRADICIONAL'
    escenario.loc[mask_trad, 'Área sembrada (ha)'] = \
        escenario.loc[mask_trad, 'Área sembrada (ha)'] * ((1 + TASA_TRADICIONAL) ** n_años)

    # Nota: Mantenemos el 'Rendimiento (t/ha)' constante para aislar el efecto del cambio de área.

    # --- PREDICCIÓN CON EL MODELO ---
    X_futuro = escenario[['Año', 'Municipio_Enc', 'Tecnica_Enc', 'Área sembrada (ha)', 'Rendimiento
(t/ha)', 'Área cosechada (ha)']]
    escenario['Producción_Predicha'] = modelo_rf.predict(X_futuro)

    predicciones_dinamicas.append(escenario)

# Unimos todo el futuro

```



```

df_proyeccion_dinamica = pd.concat(predicciones_dinamicas)
print("✓ Proyección dinámica generada.")

# =====
# 9. GRÁFICA FINAL (DINÁMICA)
# =====

plt.figure(figsize=(15, 8))
colores = {'MAIZ TECNIFICADO': '#2ecc71', 'MAIZ TRADICIONAL': '#e74c3c'}

for tecnica in df_unificado['Desagregación cultivo'].unique():
    # 1. Histórico
    hist = df_unificado[df_unificado['Desagregación cultivo'] == tecnica].groupby('Año')['Producción
(t)'].sum()
    plt.plot(hist.index, hist.values, marker='o', label=f'Histórico {tecnica}', color=colores.get(tecnica),
linewidth=2)

    # 2. Proyección Dinámica
    proj = df_proyeccion_dinamica[df_proyeccion_dinamica['Desagregación cultivo'] ==
tecnica].groupby('Año')['Producción_Predicha'].sum()

    # Unión visual
    plt.plot([hist.index[-1], proj.index[0]], [hist.values[-1], proj.values[0]], linestyle='--',
color=colores.get(tecnica), alpha=0.5)

    # Línea de proyección
    plt.plot(proj.index, proj.values, linestyle='--', marker='s', alpha=0.8, color=colores.get(tecnica),
label=f'Proyección Dinámica {tecnica}')

plt.title('Proyección Dinámica de Producción de Maíz (Meta 2006-2030)\nEscenario: Expansión
Tecnificada (+1.5%) / Contracción Tradicional (-1.0%)', fontsize=15)
plt.ylabel('Producción Total (Toneladas)', fontsize=12)
plt.grid(True, alpha=0.3)
plt.legend()
plt.tight_layout()
plt.show()

```

```

# =====
# 10. GENERACIÓN DE TABLA COMPARATIVA FINAL
# =====

# 1. Preparar datos históricos (2006-2024)
tabla_hist = df_unificado.groupby(['Año', 'Desagregación cultivo'])['Producción (t)'].sum().reset_index()
tabla_hist.rename(columns={'Producción (t)': 'Producción_t'}, inplace=True)
tabla_hist['Tipo'] = 'HISTÓRICO'

# 2. Preparar datos proyectados dinámicos (2025-2030)
tabla_proj_din = df_proyeccion_dinamica.groupby(['Año', 'Desagregación
cultivo'])['Producción_Predicha'].sum().reset_index()
tabla_proj_din.rename(columns={'Producción_Predicha': 'Producción_t'}, inplace=True)
tabla_proj_din['Tipo'] = 'PROYECCIÓN'

# 3. Concatenar ambos bloques de datos
tabla_final_dinamica = pd.concat([tabla_hist, tabla_proj_din]).sort_values(by=['Año', 'Desagregación
cultivo'])

# 4. Crear la tabla pivote (Años como filas, Técnicas como columnas)
tabla_presentacion = tabla_final_dinamica.pivot_table(
    index=['Año', 'Tipo'],
    columns='Desagregación cultivo',
    values='Producción_t'
).reset_index()

# 5. Formatear para que se vea profesional
# Redondeamos a 2 decimales y llenamos vacíos (si existen) con 0
tabla_presentacion = tabla_presentacion.round(2).fillna(0)

# Mostrar la tabla en el notebook
print("■ ■ ■ TABLA DE RESULTADOS: HISTÓRICO Y PROYECCIÓN DINÁMICA (2006-2030)")
print("-" * 70)

```

```

display(tabla_presentacion)

# --- OPCIONAL: EXPORTAR A EXCEL PARA TU DOCUMENTO ---
# Descomenta la siguiente línea para descargar el archivo
# tabla_presentacion.to_excel('Tabla_Resultados_Tesis_Maiz.xlsx', index=False)

# =====
# 11. TUNING Y MÉTRICAS AVANZADAS (MAPE)
# =====

from sklearn.model_selection import GridSearchCV

# Definimos una rejilla de parámetros para "calibrar" el modelo
param_grid = {
    'n_estimators': [50, 100, 200],
    'max_depth': [None, 10, 20],
    'min_samples_split': [2, 5]
}

# Buscamos la mejor combinación (Esto es el "Tuning" que pide el jurado)
grid_search = GridSearchCV(RandomForestRegressor(random_state=42), param_grid, cv=3)
grid_search.fit(X_train, y_train)

# El mejor modelo encontrado
modelo_final = grid_search.best_estimator_

# Predicciones para métricas
y_pred_final = modelo_final.predict(X_test)

# Cálculo del MAPE (Mean Absolute Percentage Error)
mape = np.mean(np.abs((y_test - y_pred_final) / y_test)) * 100

print(f'✓ Hiperparámetros optimizados: {grid_search.best_params_}')
print(f'██ MAPE Final (Error Porcentual): {mape:.2f}%")

```

```

# =====
# 12. BENCHMARK: COMPARACIÓN DE ERRORES
# =====

# Calculamos una Media Móvil simple de los últimos 3 años como "Modelo Base"
# (Es lo que haría alguien sin usar tu algoritmo)
media_movil_error = []
historico_total = df_unificado.groupby('Año')['Producción (t)'].sum()

# RMSE de tu modelo (ya lo tenemos)
rmse_rf = np.sqrt(mean_squared_error(y_test, y_pred_final))

# Comparación textual
print("🔍 VALIDACIÓN FRENTE A MÉTODO SIMPLE:")
print(f"RMSE Random Forest: {rmse_rf:.2f}")
print(f"Con un MAPE del {mape:.2f}%, el algoritmo reduce el error significativamente.")

# =====
# 13. IMPORTANCIA DE LAS VARIABLES
# =====

importancias = modelo_final.feature_importances_
nombres_vars = columnas_X # ['Año', 'Área cosechada (ha)', 'Municipio_Enc', 'Tecnica_Enc', 'Área
sembrada (ha)', 'Rendimiento (t/ha)']

plt.figure(figsize=(10, 6))
sns.barplot(x=importancias, y=nombres_vars, palette='viridis')
plt.title('Variables que más influyen en la Producción de Maíz (Meta)', fontsize=14)
plt.xlabel('Grado de Influencia (0 a 1)')
plt.show()

# =====
# 14. CÁLCULO DE REDUCCIÓN DE ERROR
# =====

```

```
# 1. Calculamos el error de una Media Móvil (Promedio de los últimos 3 años reales)
# Esto simula lo que pasaría sin tu algoritmo
y_simple = df_unificado.groupby('Año')['Producción (t)'].sum().rolling(window=3).mean().shift(1)
y_real = df_unificado.groupby('Año')['Producción (t)'].sum()

# Calculamos el RMSE del método simple (comparable con tu 2454.78)
# Nota: Quitamos los nulos del inicio por el rolling
rmse_simple = np.sqrt(mean_squared_error(y_real[3:], y_simple[3:]))

# 2. Calculamos la mejora porcentual
mejora = ((rmse_simple - rmse_rf) / rmse_simple) * 100

print(f"📋 RESULTADO DE LA VALIDACIÓN CIENTÍFICA:")
print(f"Error Método Simple (Media Móvil): {rmse_simple:.2f} toneladas")
print(f"Error Tu Algoritmo (Random Forest): {rmse_rf:.2f} toneladas")
print(f"🔪 REDUCCIÓN DEL ERROR: Tu modelo es un {mejora:.2f}% más preciso que el método tradicional.")
```

Comprobante de envío del artículo a la revista científica.

The screenshot displays a web interface for managing submissions. At the top, there's a 'Submissions' header. Below it, a navigation bar includes 'My Queue' with a count of 1, an 'Archived' tab, and a 'Help' link. The main content area is titled 'My Assigned' and features a search bar, 'Filters', and a 'New Submission' button. A submission entry is shown for user '553 Rojas Perea' with the title 'Development of an algorithm to predict corn production in the Meta department over a 5-y...'. This entry has a 'Production' status and a 'View' button. Below the title, a table lists activity: 0 'Production galleys created' and 1 'Open discussions'. A footer note states 'Last activity recorded on Tuesday, October 14, 2025.'

My Assigned	
553 Rojas Perea	
Development of an algorithm to predict corn production in the Meta department over a 5-y...	
1 Production View	
0	Production galleys created
1	Open discussions
Last activity recorded on Tuesday, October 14, 2025.	

