

Proyecto:
Modelo Predictivo de Exportaciones Agrícolas
en Colombia hasta 2026

Ricardo Andres Sanchez Sanchez
Alexander González Troncoso

5 de abril de 2025

1. Introducción

El comercio de productos agrícolas en Colombia ha sido un pilar fundamental de la economía nacional a lo largo de la historia. Gracias a su diversidad climática y riqueza en recursos naturales, el país ha logrado posicionarse como un exportador clave de productos como café, banano, flores, cacao, palma de aceite y una amplia variedad de frutas exóticas. Además, Colombia también exporta animales vivos, resaltando la importancia de la biodiversidad en su comercio exterior. Estas exportaciones han contribuido significativamente al desarrollo económico, generando empleo y fortaleciendo las relaciones comerciales con diversos países.

Sin embargo, el comportamiento de las exportaciones agrícolas está sujeto a una serie de factores que dificultan su predicción, tales como la volatilidad del mercado, la variabilidad climática, la política comercial y las tendencias de consumo global. En este contexto, la aplicación de modelos predictivos basados en ciencia de datos y aprendizaje automático se convierte en una herramienta clave para la toma de decisiones en el sector.

Este proyecto desarrolla un modelo de predicción de exportaciones agrícolas en Colombia hasta 2026, utilizando técnicas avanzadas de análisis de datos. Se implementó un proceso ETL (Extract, Transform, Load) para limpiar y estructurar la información, seguido de un análisis exploratorio para identificar patrones y tendencias. Se compararon diferentes enfoques de modelado, incluyendo regresión lineal múltiple, algoritmos de machine learning como Random Forest, XGBoost y redes neuronales. Además, se integraron dashboards interactivos en Power BI para facilitar la visualización de los resultados y su interpretación por parte de los actores del sector agroexportador.

Con este trabajo, se busca proporcionar una herramienta robusta y precisa que permita optimizar la gestión de las exportaciones agrícolas en Colombia, ayudando a mitigar riesgos, identificar oportunidades de crecimiento y fortalecer la competitividad del país en el mercado global.

2. Identificación del problema

La exportación de productos agrícolas colombianos enfrenta diversos desafíos que afectan su estabilidad y crecimiento. La falta de herramientas predictivas adecuadas impide anticipar las fluctuaciones del mercado y tomar decisiones informadas sobre la producción y distribución. Se identifican las siguientes problemáticas principales:

- Variabilidad de los mercados internacionales y su impacto en la demanda de productos colombianos.
- Efectos del cambio climático en la producción agrícola y disponibilidad de productos para la exportación.
- Falta de acceso a modelos predictivos que permitan planificar de manera estratégica las exportaciones.
- Necesidad de optimizar la logística y distribución de productos para minimizar pérdidas y maximizar ganancias.

3. Objetivos

3.1. Objetivo General

Desarrollar un modelo predictivo para la gestión de las exportaciones agrícolas tradicionales y no tradicionales en Colombia, utilizando técnicas avanzadas de análisis de datos para optimizar la toma de decisiones en el sector.

3.2. Objetivos Específicos

- Implementar un proceso ETL con SQL y Python para limpiar, transformar y estructurar la información de exportaciones agrícolas colombianas.
- Visualizar tendencias y patrones de exportación mediante dashboards interactivos en Power BI.
- Construir un modelo predictivo basado en regresión logística, series temporales (ARIMA, Prophet) y técnicas de machine learning.

- Optimizar el modelo predictivo mediante ajuste de hiperparámetros y validación con datos recientes.
- Integrar el modelo en un sistema de soporte a la toma de decisiones con herramientas como Streamlit o Looker.
- Analizar el impacto del modelo en la planificación y gestión del comercio agrícola.

4. Análisis descriptivo

El conjunto de datos utilizado en este estudio proporciona información detallada sobre las exportaciones del sector agrícola en Colombia, medidas en dólares FOB y toneladas. Las exportaciones incluidas abarcan tanto productos agrícolas regulados por la Organización Mundial del Comercio (OMC) como productos del sector pesquero.

El presente conjunto de datos proporciona información sobre las exportaciones agrícolas tradicionales y no tradicionales de Colombia, medidas en dólares FOB y toneladas. Estas exportaciones abarcan el ámbito de la Organización Mundial del Comercio (OMC) y el sector pesquero.

- **Última actualización:** 11 de septiembre de 2024.
- **Fuente de los datos:** Oficina de Asuntos Internacionales, Ministerio de Agricultura y Desarrollo Rural.
- **Propietario del conjunto de datos:** Ministerio de Agricultura.
- **Cobertura geográfica:** Nacional.
- **Frecuencia de actualización:** Mensual.
- **Número de filas:** 476,000.
- **Número de columnas:** 9.
- **Cada fila representa:** Una exportación.
- **Licencia:** Creative Commons Attribution — NoDerivatives 4.0 International License.

- Enlace de la fuente: <https://www.minagricultura.gov.co/>.

Columnas (9)

Nombre de la columna	Descripción	Nombre del campo API	Tipo de Dato
# Partida	Subpartida arancelaria a 10 dígitos	partida	Número
Tr Descripción Partida10 Dig	Descripción de la subpartida arancelaria a 10 dígitos	descripcion_partida10_dig	Texto
Tr Tradición productos	Tipo de tradicionalidad del producto	tradici_n_productos	Texto
Tr Departamento	Departamento de origen	departamento	Texto
Tr País	País de destino de la exportación	pais	Texto
# Año	Año	a_o	Número
Tr Mes	Mes	mes	Texto
# Exportaciones en valor (Miles USD FOB)	Valor de la exportación en miles de dólares FOB	exportaciones_en_valor_miles	Número
# Exportaciones en volumen (Toneladas)	Cantidad de toneladas de la exportación	exportaciones_en_volumen	Número

Figura 1: Descripción de las columnas del DataFrame utilizado en el análisis.

Los datos utilizados para complementar este análisis provienen del portal de datos abiertos del gobierno colombiano. En particular, se ha trabajado con el conjunto de datos disponible en:

https://www.datos.gov.co/Agricultura-y-Desarrollo-Rural/Exportaciones-agr-colas-no-tradicionales-y-tradici/5fct-ib9u/data_preview

Este conjunto de datos permite analizar la evolución de las exportaciones agrícolas en el país y establecer patrones de comportamiento que son fundamentales para la construcción del modelo predictivo desarrollado en este estudio.

	Partida	Descripción Partida10 Díg	Tradicón productos	Departamento	País	Año	Mes	Exportaciones en valor (Miles USD FOB)	Exportaciones en volumen (Toneladas)
12	101210000	Caballos reproductores de raza pura, vivos	No tradicional	Antioquia	Estados Unidos	2019	Abril	30	5
14	101210000	Caballos reproductores de raza pura, vivos	No tradicional	Antioquia	Estados Unidos	2019	Junio	119	2
15	101210000	Caballos reproductores de raza pura, vivos	No tradicional	Antioquia	Estados Unidos	2019	Julio	3	1
16	101210000	Caballos reproductores de raza pura, vivos	No tradicional	Antioquia	Estados Unidos	2019	Septiembre	16	3
23	101210000	Caballos reproductores de raza pura, vivos	No tradicional	Antioquia	Estados Unidos	2020	Octubre	18	3

Figura 2: Muestra del Dataset en Jupyter.

Proceso ETL (Extracción, Transformación y Carga)

El proceso de ETL (Extract, Transform, Load) es fundamental en cualquier proyecto de análisis de datos, ya que permite estructurar la información de manera adecuada para su posterior exploración y modelado. En este proyecto, el conjunto de datos utilizado contiene información detallada sobre las exportaciones agrícolas en Colombia, incluyendo variables como el tipo de producto, el país de destino, el departamento de origen, el valor en dólares FOB y el volumen exportado en toneladas.

Importancia de la limpieza y estructuración del dataset

Tener un dataset limpio y bien estructurado es esencial por varias razones:

- **Precisión de los modelos predictivos:** Datos inconsistentes, duplicados o con valores faltantes pueden afectar la calidad de las predicciones.
- **Interpretabilidad de los resultados:** Un dataset bien organizado facilita la identificación de patrones y tendencias en las exportaciones.
- **Eficiencia computacional:** Datos sin errores y correctamente tipificados permiten un procesamiento más rápido y eficiente.
- **Toma de decisiones basada en evidencia:** Los datos confiables aseguran que las estrategias de exportación se fundamenten en información veraz y actualizada.

Resumen de la estructura del dataset

El dataset contiene **475,785 registros** y **9 columnas**, sin valores nulos, lo que indica una buena integridad en los datos iniciales. Sin embargo, se realizó un análisis exploratorio para entender la distribución de las variables clave.

Distribución de las variables principales

- **Año de exportación:** Los datos abarcan desde **2019 hasta 2024**, con una media de **2021.37**, lo que permite analizar tendencias recientes en el comercio agrícola.
- **Exportaciones en valor (Miles USD FOB):**
 - El valor máximo registrado en una exportación es de **92.600.000 USD**, mientras que el mínimo es **0 USD**.
 - Un **50 % de las exportaciones tienen valores inferiores a 1.000 USD**, lo que sugiere una gran presencia de exportaciones de bajo volumen económico.
- **Exportaciones en volumen (Toneladas):**
 - La media es de **64.71 toneladas por exportación**, con un máximo de **60.979 toneladas**, lo que indica una amplia variabilidad en la cantidad exportada.

En esta fase del proceso ETL, se implementaron técnicas para manejar datos faltantes, estandarizar los nombres de las columnas y transformar variables cuando fue necesario. En la siguiente etapa, se analizarán patrones y tendencias con técnicas de visualización de datos para comprender mejor el comportamiento de las exportaciones agrícolas en Colombia.

4.1. Transformación del Dataset

El proceso de transformación del conjunto de datos es una etapa crítica en el análisis de datos, ya que permite preparar la información en un formato adecuado para su posterior procesamiento y modelado. A continuación, se describen los pasos aplicados para estructurar el dataset de exportaciones agrícolas, enfatizando su justificación y relevancia para el análisis.

4.1.1. Unificación de las columnas de fecha

Originalmente, la información temporal del dataset se encontraba dividida en dos columnas separadas: **Año** y **Mes**. Este formato representa una dificultad técnica, ya que muchos algoritmos de análisis de series temporales requieren un campo de fecha unificado. Para solucionar este problema, se realizó la conversión del nombre del mes (en español) a su valor numérico utilizando el siguiente diccionario:

```
meses_dict = {
    "Enero": 1, "Febrero": 2, "Marzo": 3, "Abril": 4,
    "Mayo": 5, "Junio": 6, "Julio": 7, "Agosto": 8,
    "Septiembre": 9, "Octubre": 10, "Noviembre": 11, "
    Diciembre": 12
}
df["Mes"] = df["Mes"].map(meses_dict)
```

Posteriormente, se unificaron las columnas **Año** y **Mes** en una sola columna de tipo **datetime**, llamada **Fecha**:

```
df["Fecha"] = pd.to_datetime(df["Año"].astype(str) + "-" +
    df["Mes"].astype(str), format="%Y-%m")
```

Esta transformación permitió trabajar de forma más eficiente con operaciones de agrupación, visualización y predicción temporal.

4.1.2. Conversión de moneda y normalización de unidades

El dataset original contenía los valores de exportación en la columna **Exportaciones en valor (Miles USD FOB)**, expresados en miles de dólares estadounidenses. Para facilitar la interpretación local del análisis, se decidió convertir dichos valores a pesos colombianos (COP). Esta conversión se realizó en tiempo real utilizando una API de tipo REST de tasas de cambio actualizadas, tal como se muestra a continuación:

```
url = "https://v6.exchangerate-api.com/v6/MI_API_KEY/latest/USD"
response = requests.get(url)
data = response.json()
tasa_cambio = data["conversion_rates"]["COP"]
```

Luego, se realizaron las conversiones correspondientes:

- Multiplicación por 1,000 para pasar de miles de USD a USD reales.

- Multiplicación por la tasa de cambio para convertir USD a COP.
- División por 1 billón (1,000,000) para expresar los valores en millones de pesos colombianos y facilitar la visualización y comparación entre datos.

```
df["Exportaciones en USD"] = df["Exportaciones en valor (
    Miles USD FOB)"] * 1000
df["Exportaciones en COP"] = df["Exportaciones en USD"] *
    tasa_cambio
df["Exportaciones en COP"] = df["Exportaciones en COP"] / 1
    _000_000
```

4.1.3. Exportación del dataset transformado

Finalmente, el conjunto de datos limpio y transformado fue exportado a un archivo CSV, el cual servirá como base para las siguientes etapas del análisis y modelado predictivo:

```
df.to_csv("Exportaciones_limpias.csv", index=False)
```

Importancia de estas transformaciones

Estas transformaciones son esenciales para garantizar la calidad del análisis. Unificar fechas permite aplicar modelos de predicción con mayor precisión, mientras que la conversión de moneda local facilita la interpretación económica del fenómeno estudiado. Adicionalmente, normalizar unidades y escalar valores garantiza una mayor robustez en la visualización, facilita la convergencia de los modelos de machine learning y mejora la experiencia del lector al presentar resultados claros y consistentes.

Análisis Descriptivo de las Exportaciones

A continuación, se presentan los resultados del análisis descriptivo de las exportaciones agrícolas de Colombia, con especial atención a su evolución en el tiempo y los principales productos exportados.

Tendencia de Exportaciones por Año

En la Figura 3, se observa una tendencia al alza en las exportaciones desde 2019 hasta 2022, alcanzando un pico en ese último año con aproximadamente 48 billones de COP.

Sin embargo, en 2023 se evidencia una leve disminución y en 2024 hay una caída más pronunciada. Esto podría estar relacionado con factores económicos globales o nacionales, cambios en la demanda de productos exportados o variaciones en el tipo de cambio.

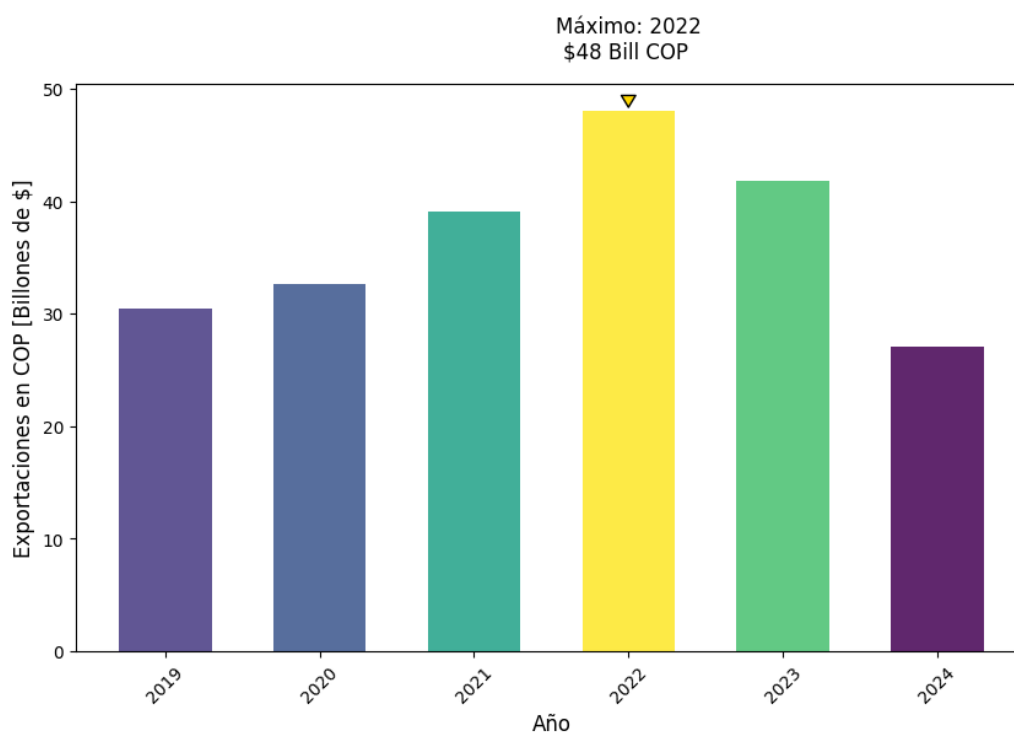


Figura 3: Evolución de las exportaciones en COP de 2019 a 2024.

Principales Productos Exportados

En la Figura 4, se muestra el ranking de los 10 productos más exportados en términos de valor en COP.

El café sin tostar es, con diferencia, el producto más exportado, generando 67 billones de COP, muy por encima del segundo producto más exportado: bananas tipo Cavendish Valery, con 21 billones de COP.

También destacan las exportaciones de flores y productos agrícolas como aceite de palma y azúcar. Esto reafirma la importancia del sector agroindustrial en la economía exportadora del país.

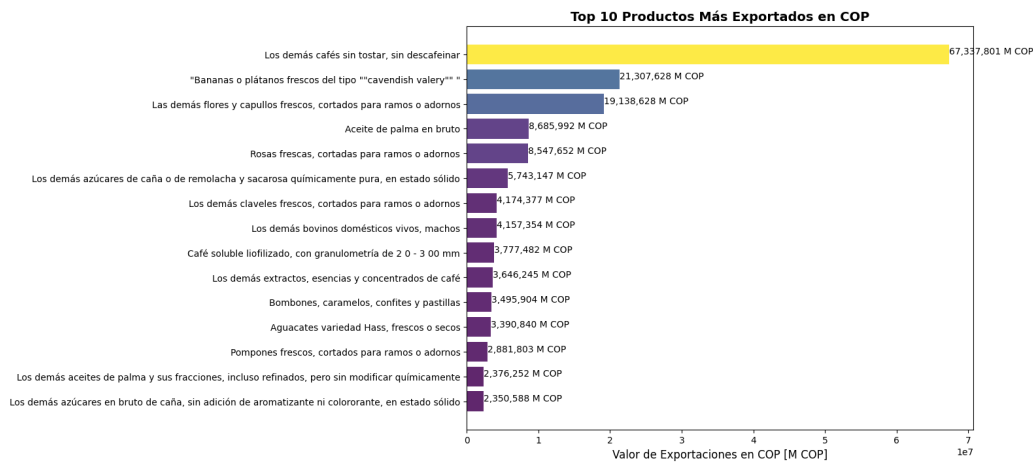


Figura 4: Top 10 productos más exportados en COP.

Distribución de Exportaciones según la Tradición de los Productos

La Figura 5 muestra la proporción de las exportaciones según su clasificación en productos tradicionales y no tradicionales. Se observa que los productos tradicionales representan el 62,2% del total exportado, mientras que los no tradicionales abarcan el 37,8%.

Esto sugiere que, aunque hay una diversificación en los productos exportados, los sectores históricos como el café, el banano y las flores siguen dominando el mercado exterior. Este predominio indica que la economía exportadora del país aún depende en gran medida de estos productos agrícolas, reflejando su importancia histórica y estructural.

Distribución de Exportaciones según Tradición de Productos

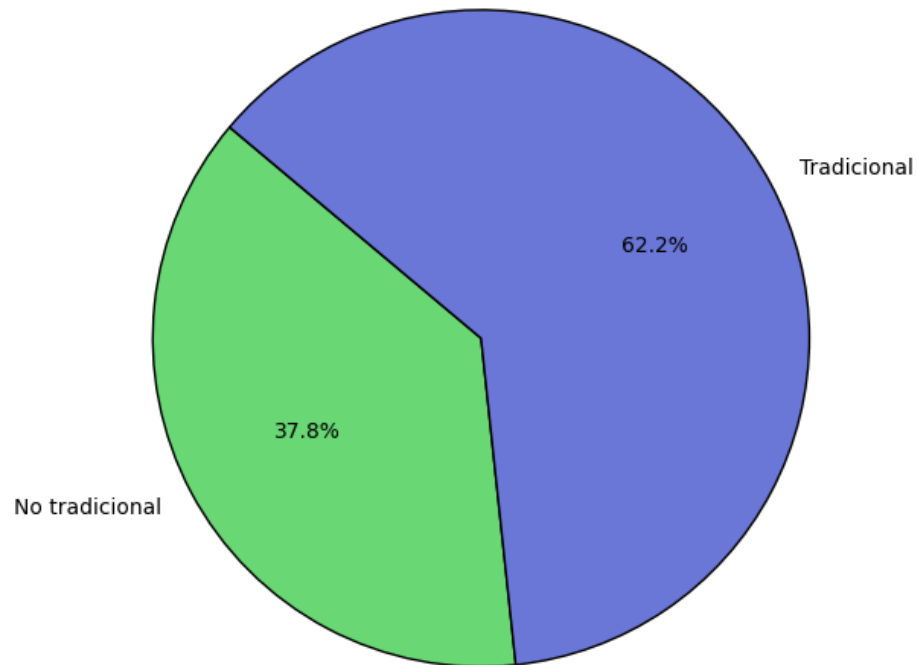


Figura 5: Distribución de las exportaciones según tradición de los productos.

Exportaciones por País de Destino

En la Figura 6, se presentan las exportaciones colombianas según el país de destino, diferenciando entre productos tradicionales y no tradicionales. Se observa que **Estados Unidos** es el principal receptor de exportaciones colombianas, seguido por **Países Bajos, Bélgica, Alemania, Canadá y Japón**.

Observaciones:

- Estados Unidos es el principal socio comercial de Colombia.
- Europa, en especial Países Bajos y Alemania, tiene una participación clave en las exportaciones colombianas.

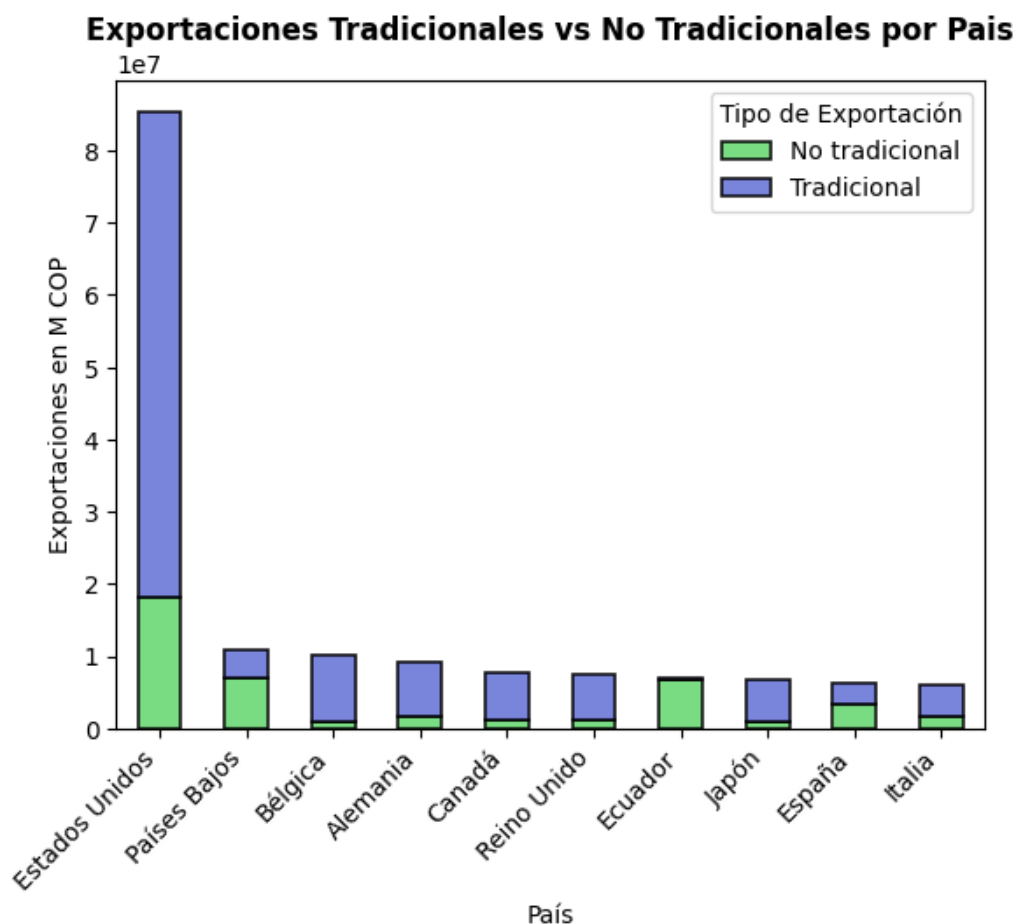


Figura 6: Exportaciones tradicionales y no tradicionales por país de destino.

- Japón y Canadá también reciben una parte importante de los productos colombianos.

Exportaciones por Departamento de Origen

La Figura 7 muestra la distribución de exportaciones por departamento de origen. Se destaca que **Antioquia y Bogotá** lideran las exportaciones nacionales, con Antioquia superando los 50 millones de COP en exportaciones.

Exportaciones Tradicionales vs No Tradicionales por Departamento

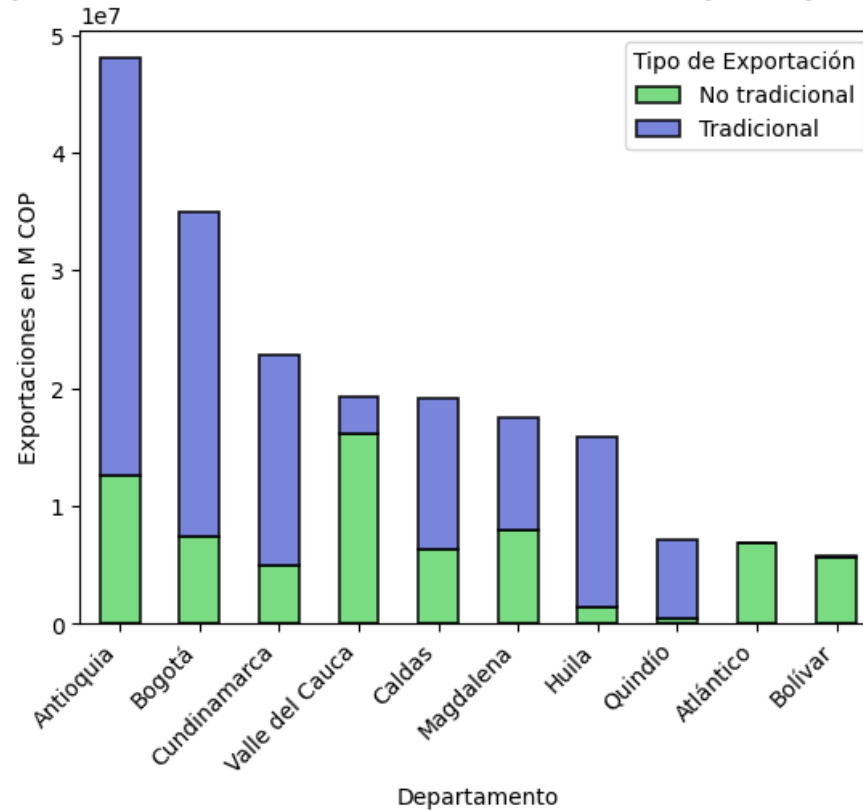


Figura 7: Exportaciones tradicionales y no tradicionales por departamento.

Observaciones:

- Antioquia y Bogotá son los departamentos con mayor volumen de exportaciones.
- Departamentos como Cundinamarca, Valle del Cauca, Caldas y Huila tienen una participación significativa, principalmente en productos agrícolas.
- La distribución muestra una diversificación territorial en las exportaciones.

Evolución de las Exportaciones

La Figura 9 muestra la evolución de las exportaciones colombianas desde 2019 hasta 2024. Se observa un crecimiento constante hasta 2022, seguido por una caída en 2023 y 2024.

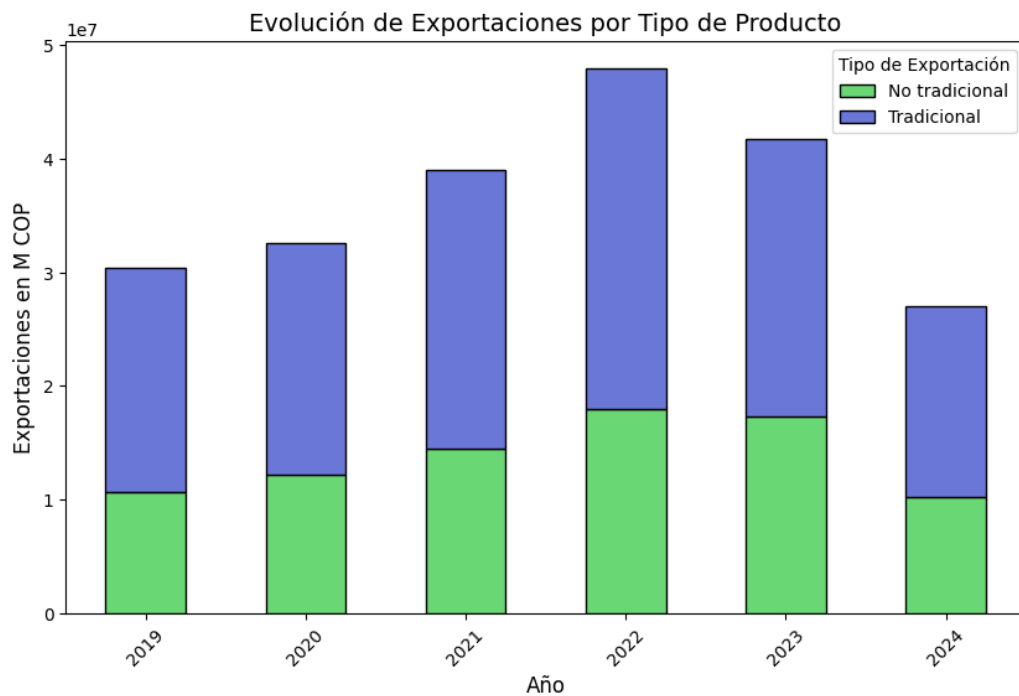


Figura 8: Evolución de las exportaciones colombianas por tipo de producto.

Observaciones:

- Las exportaciones crecieron de 2019 a 2022, alcanzando un pico en 2022.
- A partir de 2023, se observa una reducción en las exportaciones, posiblemente debido a factores macroeconómicos globales.
- Los productos tradicionales siguen siendo la base de las exportaciones colombianas, pero los no tradicionales han ganado participación con el tiempo.

5. Modelo Predictivo

Descripción del flujo de modelado predictivo y comparación de modelos

En esta sección se describe de forma detallada el procedimiento de análisis, modelado y evaluación de distintos algoritmos de predicción aplicados a una base de datos con registros de exportaciones agrícolas de Colombia entre los años 2019 y 2024. El objetivo principal es comparar el rendimiento de diferentes modelos de regresión y seleccionar aquel que proporcione las mejores proyecciones para el período 2025–2026.

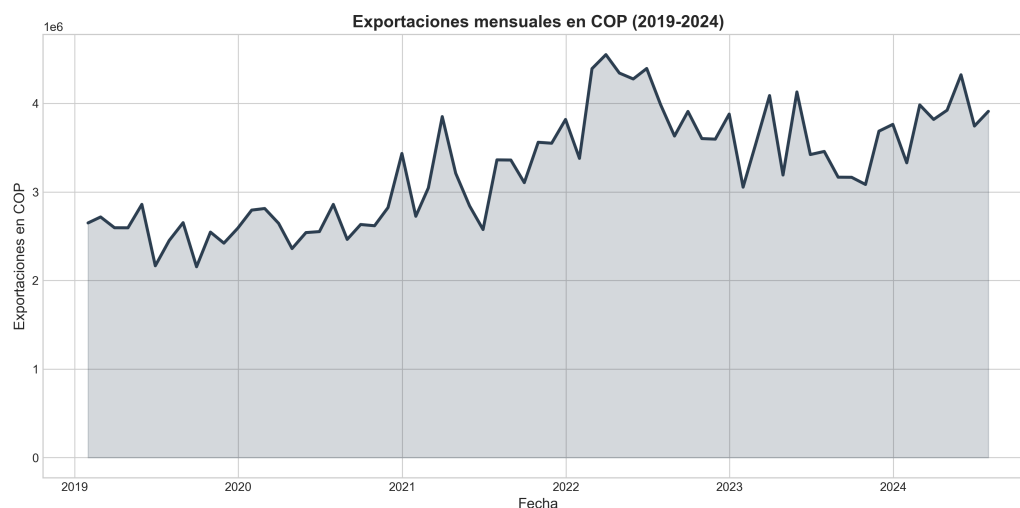


Figura 9: En este gráfico vemos el comportamiento de las exportaciones en M COP, punto de partida para establecer los modelos.

Preprocesamiento del conjunto de datos

Se parte de una base de datos con registros mensuales de exportaciones, clasificados según tipo de producto (tradicional o no tradicional), el valor monetario exportado y la fecha. Las transformaciones iniciales incluyen:

- Conversión de la columna de fechas al tipo `datetime`.

- Extracción de variables temporales como Año y Mes.
- Conversión de la variable categórica Tradicional a binaria.
- Agrupación por mes, año y tipo de producto para conformar una serie temporal consolidada.

Estas transformaciones permiten preparar el conjunto de datos para su uso en modelos de predicción supervisada.

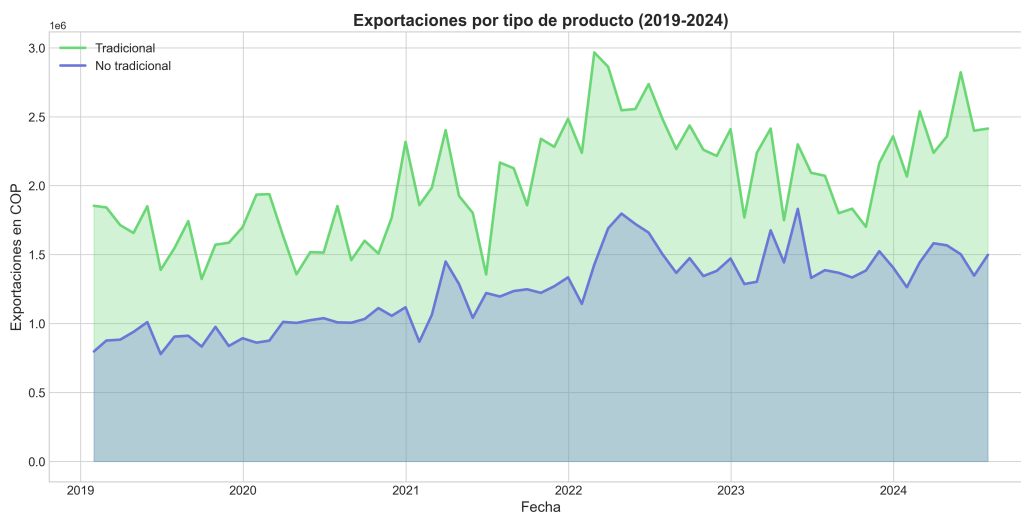


Figura 10: En este grafico vemos el comportamiento de las exportaciones en M COP, punto de partida para establecer los modelos.

Selección de variables

Se definieron como variables explicativas las siguientes:

- Año: permite capturar tendencias a largo plazo.
- Mes: útil para modelar estacionalidades.
- Tradicional: variable categórica binaria que indica si el producto es tradicional.

La variable objetivo fue:

- Exportaciones en millones COP.

División de datos

El conjunto de datos fue dividido en conjuntos de entrenamiento (80 %) y prueba (20 %) utilizando la función `train_test_split`. Esta división es fundamental para validar la capacidad de generalización de los modelos.

Modelos entrenados

Se entrenaron cinco modelos de regresión, cada uno con características y fundamentos matemáticos distintos:

Regresión Lineal Múltiple

Modelo estadístico clásico que asume una relación lineal entre las variables predictoras y la respuesta. La forma general del modelo es:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

Este modelo es útil por su interpretabilidad, aunque limitado en presencia de relaciones no lineales.

Random Forest Regressor

Algoritmo de ensamblado basado en árboles de decisión. Construye múltiples árboles sobre subconjuntos aleatorios de los datos y promedia sus resultados. Su fórmula se expresa como:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x)$$

donde $h_t(x)$ representa la predicción del árbol t . Es robusto ante el sobreajuste y captura relaciones no lineales.

XGBoost Regressor

Modelo de *gradient boosting* optimizado, que agrega árboles secuencialmente para corregir los errores de las predicciones anteriores. Su formulación incluye regularización explícita, lo que favorece la generalización. El modelo se actualiza según:

$$\hat{y}_t = \hat{y}_{t-1} + \eta \cdot f_t(x)$$

donde η es la tasa de aprendizaje y f_t representa el árbol ajustado en el paso t .

Gradient Boosting Regressor

Modelo similar a XGBoost, pero implementado en `scikit-learn`. También construye árboles secuenciales, pero sin la eficiencia ni la regularización incorporada de XGBoost. A pesar de ello, mantiene una buena capacidad para capturar relaciones complejas.

Red Neuronal (MLPRegressor)

Modelo de red neuronal multicapa (Multilayer Perceptron) con una arquitectura de tipo *feedforward*. La red contiene una capa oculta con 50 neuronas y función de activación ReLU. Se entrena utilizando el optimizador Adam y la función de pérdida cuadrática media. Este modelo es capaz de modelar relaciones altamente no lineales, aunque con menor interpretabilidad y mayor necesidad de ajuste de hiperparámetros.

- Exportaciones en millones COP.

Los resultados de los modelos establecidos en este caso se muestran en la Figura 12, en este se muestran los resultados de los modelos para las fechas establecidas. Haciendo el comparativo entre todos, en donde es evidente que algunos se ajustan mucho mejor que otros.

Comparación de Modelos Predictivos

Para evaluar el desempeño de los diferentes modelos predictivos, se compararon las métricas de error cuadrático medio (MSE), raíz del error cuadrático medio (RMSE) y coeficiente de determinación (R^2). Los resultados obtenidos son los siguientes:

De acuerdo con la tabla 1, el modelo que obtuvo el mejor desempeño en términos del coeficiente de determinación (R^2) fue **Random Forest**, con un valor de 0,6096. Esto indica que este modelo es el que mejor logra explicar la variabilidad de los datos.

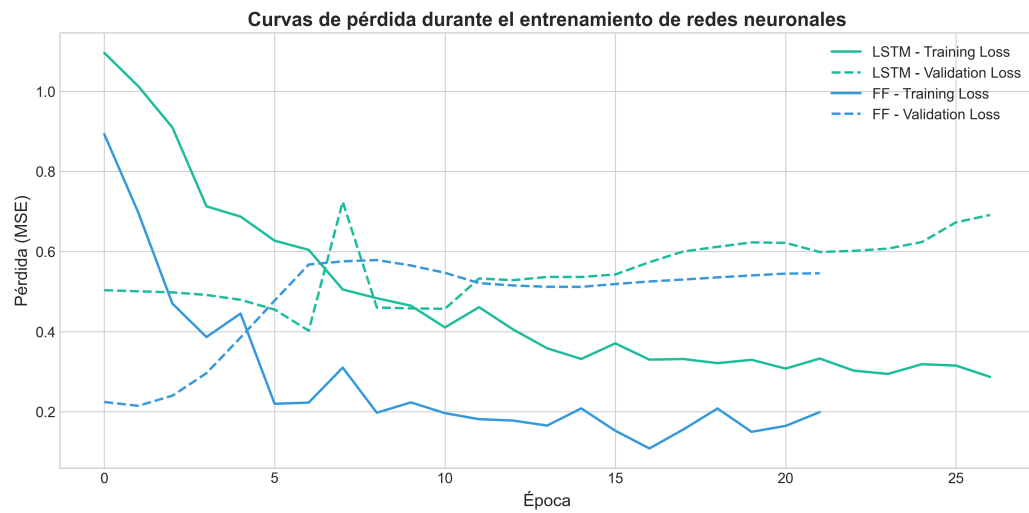


Figura 11: En este grafico se muestra como se comporta las curvas de pérdida durante el entrenamiento de redes neuronales.

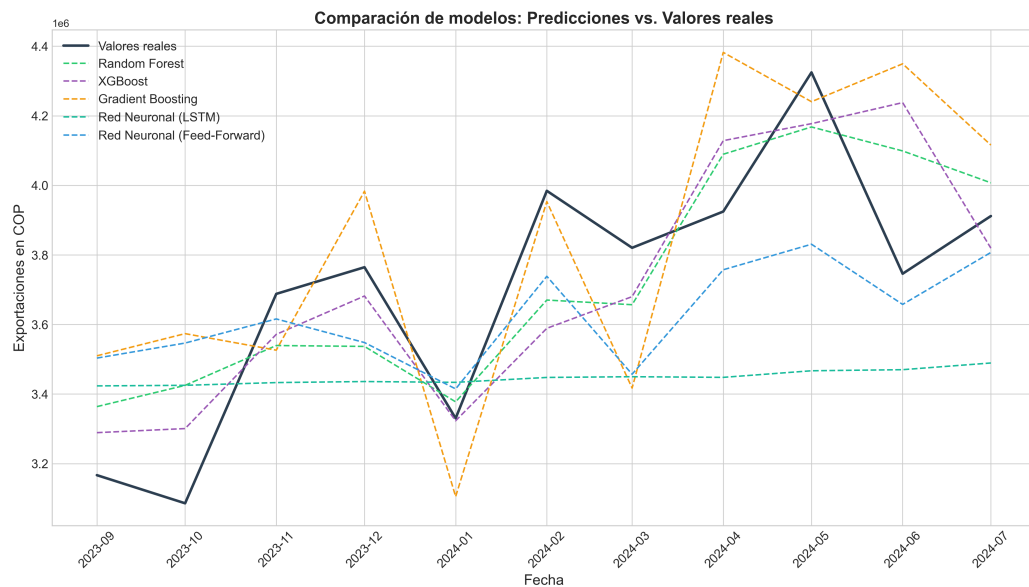


Figura 12: En este grafico vemos el comportamiento de las exportaciones en M COP, punto de partida para establecer los modelos.

Modelo	MSE	RMSE	R^2
Random Forest	$4,92 \times 10^{10}$	221,715,52	0,6096
XGBoost	$5,19 \times 10^{10}$	227,875,83	0,5876
Gradient Boosting	$1,15 \times 10^{11}$	339,270,94	0,0859
Red Neuronal (LSTM)	$1,82 \times 10^{11}$	427,197,73	-0,4493
Red Neuronal (Feed-Forward)	$7,89 \times 10^{10}$	280,895,24	0,3734

Cuadro 1: Resultados de los modelos predictivos

Análisis de los resultados

- **Random Forest:** Presenta el mejor R^2 , lo que sugiere que la combinación de múltiples árboles de decisión le permite capturar bien las relaciones en los datos. Además, su capacidad para manejar datos no lineales y mitigar el sobreajuste mediante la agregación de árboles ha sido clave para su éxito en esta tarea.

- **XGBoost:** Aunque suele superar a Random Forest en muchas aplicaciones, en este caso su desempeño es ligeramente inferior. Posibles explicaciones incluyen que los hiperparámetros aún pueden ser optimizados o que la estructura del dataset favorece la aleatorización de Random Forest sobre el boosting secuencial de XGBoost.

- **Gradient Boosting:** Su desempeño es considerablemente peor que Random Forest y XGBoost, con un R^2 de apenas 0,0859. Esto sugiere que el modelo puede estar subajustado o no haber capturado correctamente las tendencias de los datos.

- **Redes Neuronales (LSTM y Feed-Forward):** En general, estos modelos muestran el peor desempeño, con valores de R^2 negativos en el caso de la LSTM. Esto indica que la red no logró generalizar bien a los datos de prueba. Es posible que la arquitectura o el tamaño del conjunto de datos no hayan sido suficientes para aprovechar el potencial de las redes neuronales.

Efecto del uso del año 2024 completo en las predicciones

Durante la experimentación, se observó un comportamiento interesante: en algunas ejecuciones, cuando se incluía el año 2024 completo en el conjunto de predicción, los modelos de redes neuronales mejoraban notablemente su desempeño, incluso superando a Random Forest.

Este fenómeno puede explicarse por varias razones:

1. **Mayor cantidad de datos recientes:** Las redes neuronales tienden a beneficiarse de patrones temporales recientes. Si los datos del 2024 siguen tendencias similares a los años anteriores y se incluyen de manera completa, el modelo puede ajustarse mejor a las dinámicas más recientes del mercado de exportaciones.

2. **Sensibilidad a la distribución de datos:** A diferencia de los modelos basados en árboles, que pueden manejar mejor distribuciones no estacionarias, las redes neuronales pueden sufrir cuando hay cambios en la distribución de los datos entre el entrenamiento y la prueba. Si la división de los datos de entrenamiento y prueba es desfavorable para la red neuronal, su rendimiento puede verse afectado.

3. **Normalización y escalado:** Si la normalización de los datos incluye información del 2024 de manera distinta en distintas pruebas, podría afectar el desempeño de la red. En modelos basados en redes neuronales, una escala diferente en los datos de entrada puede impactar directamente la capacidad de aprendizaje.

4. **Secuencialidad en LSTM:** Las redes LSTM dependen en gran medida de la secuencialidad de los datos. Si la estructura de la entrada cambia al incluir un año adicional completo, el modelo puede encontrar mejores correlaciones en los patrones estacionales.

Observaciones

Si bien los modelos basados en redes neuronales pueden mejorar en condiciones específicas (como incluir el 2024 completo en la predicción), la comparación general de modelos sugiere que **Random Forest es el modelo más robusto y consistente en este contexto**. Su capacidad para manejar datos tabulares, combinada con su flexibilidad para capturar relaciones complejas sin requerir ajustes excesivos, lo convierte en la mejor opción para predecir las exportaciones agrícolas de Colombia.

Para futuras mejoras, se recomienda optimizar el preprocesamiento de datos para las redes neuronales y realizar una validación más rigurosa con esquemas de validación temporal para evaluar la estabilidad de los modelos en distintos periodos de predicción.

Proyección de exportaciones a futuro

Una vez identificado el modelo con mejor desempeño, se generaron predicciones para el periodo 2025–2026. Para ello, se construyó un nuevo conjunto de datos con las combinaciones futuras de año, mes y tipo de producto, y se aplicó el modelo entrenado para obtener las proyecciones correspondientes del valor de exportaciones.

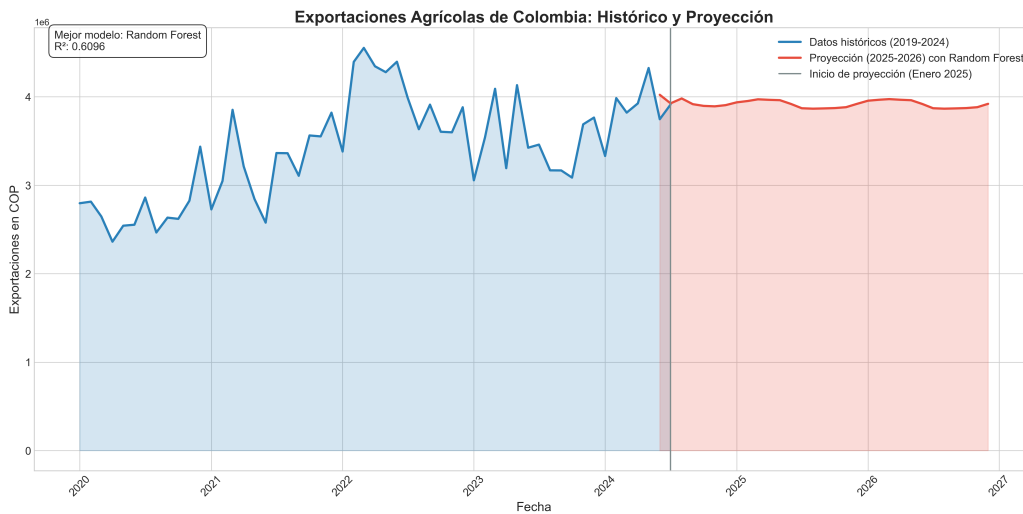


Figura 13: En este grafico vemos el comportamiento de las exportaciones en M COP, punto de partida para establecer los modelos.

Observaciones

El flujo implementado integra preprocesamiento, selección de variables, entrenamiento de modelos, evaluación comparativa y proyección a futuro. Este enfoque permite identificar el modelo con mejor desempeño y utilizarlo para realizar predicciones fundamentadas sobre la evolución esperada de las exportaciones agrícolas colombianas.

Al observar los gráficos a futuro, es fundamental observar que se tiene una predicción que parece acoplarse a la tendencia que se espera a futuro, estos resultados muestran un solapamiento entre datos del 2024 para lograr un mejor resultado, porque los resultados dependen de la continuidad de los datos que se tengan, por otro lado se logra ver las tendencias e incrementos

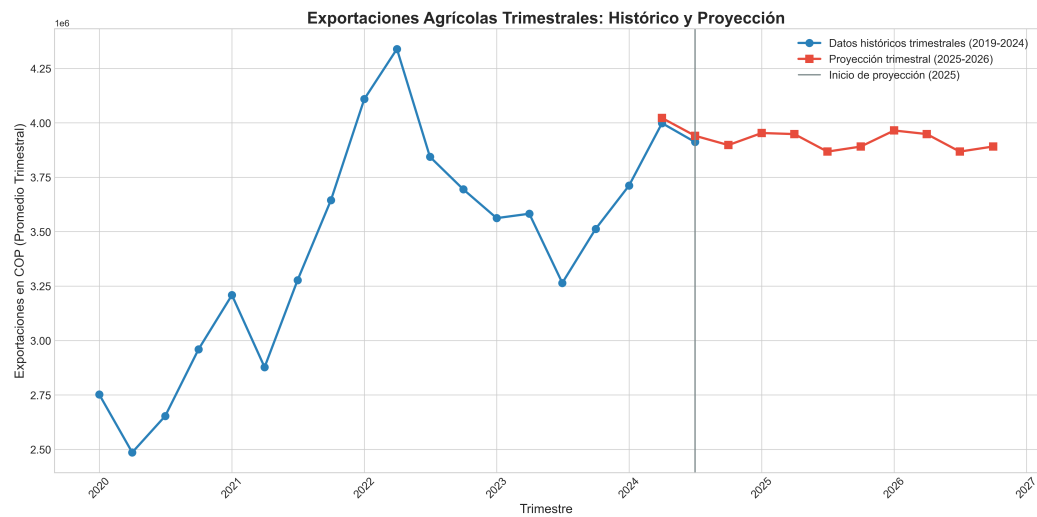


Figura 14: En este grafico vemos el comportamiento de las exportaciones en M COP, punto de partida para establecer los modelos.

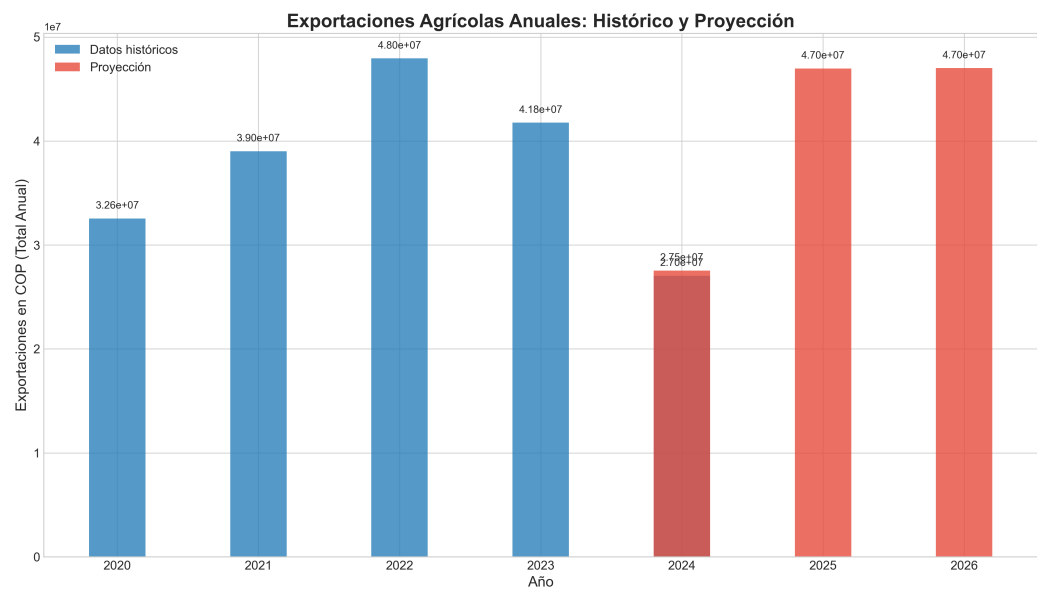


Figura 15: En este grafico vemos el comportamiento de las exportaciones en M COP, punto de partida para establecer los modelos.

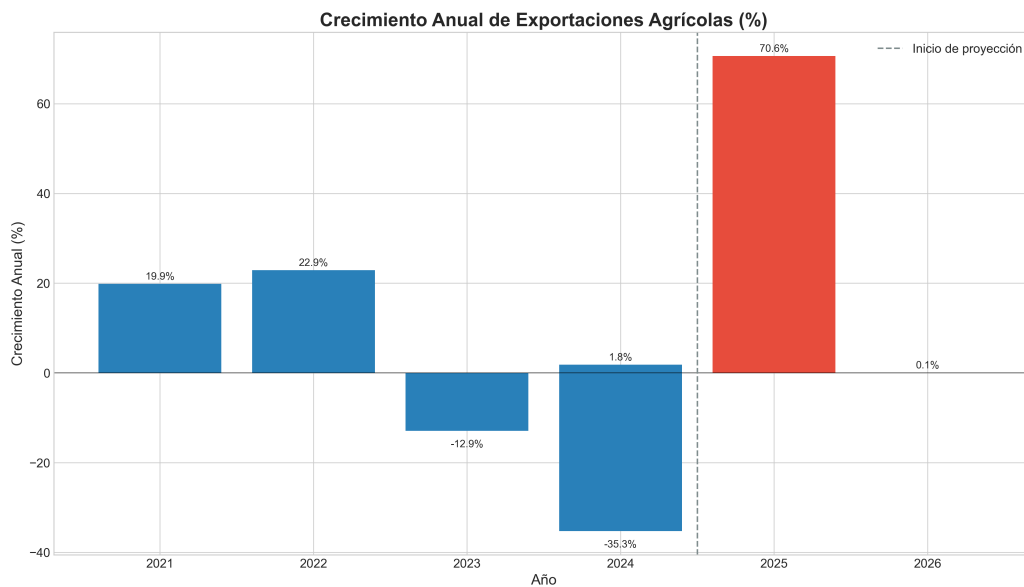


Figura 16: En este grafico vemos el comportamiento de las exportaciones en M COP, punto de partida para establecer los modelos.

en los meses extremos, es decir, Diciembre y Enero muestran unos picos elebados respecto al resto de año y está muy asociado con el comportamiento del mercado en general durante estas épocas específicas.

6. Dashboard

El dashboard desarrollado en Power BI permite visualizar y analizar en tiempo real el comportamiento de las exportaciones agrícolas colombianas, brindando herramientas clave para la toma de decisiones informadas por parte de expertos del sector.

Objetivos principales del dashboard:

- Monitorear el estado diario de las exportaciones por departamento.
- Analizar tendencias por tipo de producto (tradicional y no tradicional).
- Identificar mercados de destino y sus variaciones en el tiempo.

- Facilitar el seguimiento del valor y volumen exportado en millones de COP.
- Esta herramienta permite una gestión más eficiente y estratégica de las exportaciones agrícolas, promoviendo el desarrollo comercial del sector.

Accede al dashboard aquí: Dashboard Power BI – Exportaciones Agrícolas

<https://app.powerbi.com/groups/me/reports/0f101035-b65e-469f-bf09-177863d7b49e/4ac62c9c0050d0c001e8?ctid=99e1e721-7184-498e-8aff-b2ad4e53c1c2&experience=power-bi>

7. Conclusiones y Recomendaciones

¿Por qué Random Forest fue el modelo más efectivo en este trabajo?

1. Capacidad de generalización con pocas variables independientes. El modelo Random Forest se destaca especialmente cuando:
 - El número de variables independientes (features) no es muy alto, pero estas variables sí contienen información relevante y con capacidad explicativa.
 - No se requiere una compleja ingeniería de características (feature engineering) para capturar relaciones útiles, ya que el modelo construye automáticamente divisiones que optimizan la reducción de varianza.

En nuestro caso, las variables predictoras eran limitadas en número, lo cual tiende a restringir la capacidad de aprendizaje de modelos más complejos como redes neuronales profundas, que requieren alta dimensionalidad y volumen de datos para aprovechar su estructura.

2. Robustez frente a relaciones no lineales Las relaciones entre los factores temporales (como el año, mes) y las exportaciones probablemente no sean estrictamente lineales ni perfectamente secuenciales. Random Forest:

- Modela relaciones no lineales de forma natural, sin necesidad de transformar las variables.
- Aprende interacciones entre variables sin requerir que estas sean explícitamente incluidas en la matriz de entrada (como ocurre en regresión lineal).

En contraste, modelos como regresión múltiple o incluso algunas redes neuronales podrían no capturar adecuadamente esas no linealidades sin una transformación cuidadosa de los datos.

3. Estabilidad ante ruido y sobreajuste Random Forest, al ser un modelo de ensamble que promedia múltiples árboles entrenados sobre diferentes subconjuntos de datos y variables, tiene ventajas notables:

- Es menos propenso al sobreajuste (overfitting), a diferencia del Gradient Boosting o las redes neuronales, que pueden sobreajustarse fácilmente si no se regularizan con precisión.
- Tiene una alta tolerancia al ruido, es decir, puede ignorar datos atípicos sin que afecten significativamente el rendimiento global. Este punto es importante porque los datos de exportaciones agrícolas pueden tener fluctuaciones por causas externas (clima, conflictos, mercados internacionales), y Random Forest logra mantener un buen desempeño bajo estas condiciones.

4. No requiere fuerte preprocesamiento En este proyecto:

- Se escalaron los datos para que pudieran ser usados por redes neuronales y métodos de boosting.
- Sin embargo, Random Forest no requiere normalización ni escalado, lo que lo hace más directo y menos sensible a problemas de escala entre variables.

Esto hace que, en términos de implementación práctica, Random Forest sea más robusto en entornos con datasets reales, como el de exportaciones, que puede contener variaciones de magnitudes entre variables.

¿Por qué otros modelos no funcionaron tan bien? ¿Influyó la baja cantidad de variables?

La baja dimensionalidad (pocas variables predictoras distintas a la variable objetivo) tuvo un impacto directo en el rendimiento de modelos más complejos:

- Redes Neuronales: Necesitan muchas variables y datos para generalizar. Con pocos features, tienden a sobreajustar y no logran capturar relaciones complejas si no están bien representadas en los datos.
- Boosting: Aunque potente, es más sensible al ajuste de hiperparámetros y puede fallar si no se identifican las combinaciones adecuadas, especialmente en datasets más pequeños o con estructuras simples.
- Regresión Lineal: Suponiendo relaciones lineales entre las variables, se vuelve insuficiente cuando existen patrones no lineales o interacciones implícitas no representadas en la forma matemática del modelo.

Conclusión técnica: Random Forest fue el modelo más efectivo porque: Tolerar bien datasets con baja dimensionalidad pero relaciones complejas. Tiene flexibilidad estructural sin necesidad de gran configuración. Es robusto ante ruido, sobreajuste y datos no transformados. Requiere poca intervención manual para funcionar bien. En conjunto, estas propiedades hacen que Random Forest sea especialmente eficaz cuando se trabaja con un número moderado de datos históricos y un conjunto limitado de variables predictoras, como en este estudio de exportaciones agrícolas.

Referencias

- [1] Sala de Prensa, Exportaciones agropecuarias de Colombia alcanzan USD 962 millones en febrero, un aumento del 12,5 . (s. f). <https://upra.gov.co/es-co/saladeprensa/Paginas/Exportaciones-agropecuarias-de-Colombia-alcanzan-USD-962-millones-en-febrero,-un-aumento-del-12,5—.aspx>