

# **RETOS DE INNOVACIÓN TECNOLÓGICA:**

## **HACKING PREDICTIVO DEL MERCADO**

Autores

María Úsuga,

Astrid Agudelo M,

Santiago Ceballos y

Mauricio Alvarado

Institución

Universidad de Antioquia,

Universidad de Caldas y

Ubicua Technology.

Talento Tech

Docente

Feibert Guzmán

Noviembre 28 de 2025



## Tabla de Contenido

RETOS DE INNOVACIÓN TECNOLÓGICA:.....	1
HACKING PREDICTIVO DEL MERCADO .....	1
1. Introducción.....	4
2. Información del Equipo .....	5
3. Reto Seleccionado.....	5
3.1. Objetivos SMART .....	6
4. Requerimientos del Sistema.....	7
5. Arquitectura Propuesta.....	7
5.1 Descripción del Proyecto.....	7
5.2 Estructura del Proyecto .....	8
5.3 Instalación y Configuración.....	9
5.4 Dataset .....	10
5.5 Modelos de Machine Learning .....	11
5.6 Resultados y Comparación .....	11
6. Aplicación del Ciclo de Vida del Machine Learning. ....	12
7. Plan de Pruebas .....	16
7.1 Modelos de Machine Learning .....	16
7.2 Resultados esperados y observados.....	18
8. Evidencias del Desarrollo .....	19
9. Conclusiones .....	27
10. Bibliografía.....	27

## Tabla de Figuras

<b>Figura 1.</b> Captura de pantalla interfaz Export AI Dashboard .....	13
<b>Figura 2.</b> Captura de pantalla interfaz ExporAI Predicción ML .....	13
<b>Figura 3.</b> Captura de pantalla interfaz ExporAI Logística .....	14
<b>Figura 4.</b> Captura de pantalla interfaz ExporAI Logros .....	14
<b>Figura 5.</b> Captura de pantalla interfaz ExporAI Configuración .....	15
<b>Figura 6.</b> Captura de pantalla del Data proyecto .....	20
<b>Figura 7.</b> Captura de pantalla archivos en Google Drive .....	21
<b>Figura 8.</b> Captura de pantalla los Notebook .....	21
<b>Figura 9.</b> Captura de pantalla del Modelo de Árboles de Decisión .....	22
<b>Figura 10.</b> Top 10 características más importantes del modelo de Árboles de decisión .....	22
<b>Figura 11.</b> Métricas de evaluación del modelo de Árboles de Decisión .....	23
<b>Figura 12.</b> Captura de pantalla del Modelo de KNN .....	23
<b>Figura 13.</b> Distribución de la Variable Target .....	24
<b>Figura 14.</b> Métricas de evaluación optimizada del modelo KNN.....	24
<b>Figura 15.</b> Captura de pantalla del Modelo de regresión Lineal.....	25
<b>Figura 16.</b> Estadística de la Variable dependiendo FOB USD en Modelo Regresión Lineal.....	25
<b>Figura 17.</b> Captura de pantalla del Modelo de Regresión Logística .....	26
<b>Figura 18.</b> Métricas de evaluación del Modelo de Regresión Logística.....	26



## 1. Introducción

La correcta identificación de la demanda internacional de clientes potenciales en el mercado global de productos colombianos es crucial para el dinamismo de las empresas exportadoras y, consecuentemente, para la economía nacional. Las últimas tendencias en el marco de los retos empresariales de Bootcamp se han centrado en el diseño, desarrollo y validación de una solución tecnológica innovadora dirigida a la identificación de clientes potenciales en el mercado internacional.

Este proyecto toma como base la información histórica de exportaciones de Colombia, clasificada por continente, cliente, país, precio unitario Free on Board (FOB) mercancía puesta en puerto de origen, sobre el buque en United States Dollar (USD), dólar estadounidense, mercancía o producto, entre otras. El análisis de estos datos permitirá focalizar el alcance comercial hacia aquellos mercados y compradores que han mostrado mayor dinamismo en la adquisición de productos colombianos. Esta estrategia, apoyada en el análisis de datos de comercio internacional, busca optimizar los procesos de internacionalización y generar un impacto real. Por ello, se ha aceptado el desafío de crear una solución tecnológica que ayude a dinamizar la presencia internacional de las empresas colombianas, propiciando oportunidades de expansión y alianzas estratégicas.

Este proyecto no solo busca aplicar conocimientos técnicos, sino también aportar valor tangible al proceso de internacionalización de la economía colombiana.



## 2. Información del Equipo

Nombre del Integrante	Rol (Scrum Master / Dev / QA / UI/UX /ML)	Correo Institucional
Astrid Agudelo M.	ML	<a href="mailto:iabasicoudea@gmail.com">iabasicoudea@gmail.com</a>
Santiago Ceballos	Dev	<a href="mailto:ceballosaguadelesantiago9@gmail.com">ceballosaguadelesantiago9@gmail.com</a>
Mauricio Alvarado	Scrum Master	<a href="mailto:mfaconsulting77@gmail.com">mfaconsulting77@gmail.com</a>
María Úsuga	UI/UX	<a href="mailto:musugag0@gmail.com">musugag0@gmail.com</a>

## 3. Reto Seleccionado

Nombre del reto: HACKING PREDICTIVO DEL MERCADO

Empresa retadora: GRUPO IA UNIDAD PREDICTIVA DEI COMERCIO EXTERIOR (UPCE)

Descripción breve del problema:

La identificación de la demanda en mercados internacionales para productos de origen nacional es un factor crítico en países en vías de desarrollo y en naciones con economías emergentes. El análisis del perfil y las necesidades de los clientes internacionales resulta fundamental para determinar las oportunidades de exportación, lo cual, consecuentemente, tiene un impacto directo en el crecimiento económico. Específicamente, este proceso aporta al aumento del Producto Interno Bruto (PIB) como a la mejora del PIB per cápita, variables que reflejan la prosperidad económica general y la distribución de la riqueza por habitante, respectivamente.

Pregunta

¿Cuál es el modelo predictivo de Machine Learning (clasificación o regresión) óptimo, entrenado con datos históricos de exportaciones colombianas (Variable Independiente: producto, país de destino, valor y volumen exportado, frecuencia de transacción, entre otras), que permite predecir la demanda internacional (Variable Dependiente: potencial de adquisición o dinamismo de compra, cuantificado o



categorizado, representa Valor FOB (USD): Valor de la exportación en dólares estadounidenses) de clientes internacionales específicos (Unidad de Análisis), utilizando sus características históricas y demográficas (Variable Propia: identificador de cliente, sector de actividad, ubicación geográfica) a partir de los registros de comercio internacional (Espacio: mercados globales, exceptuando la variabilidad macroeconómica global como Variable Extraña) recopilados en el periodo más reciente disponible (Unidad Temporal y Tiempo), siendo este periodo 2024 hasta primer semestre 2025, con el objetivo de focalizar las estrategias comerciales de las empresas exportadoras colombianas?

1. Unidad de análisis: clientes internacionales específicos.
2. Variable dependiente: Valor FOB (USD).
3. Variables Independientes principales: país de destino, departamento origen, cantidad, peso en kilos netos y brutos, vía de transporte, número de artículos, mercancía, entre.
4. Variables Extrañas: Continente.
5. Variables propias de los individuos si aplica.
6. Unidad Temporal y tiempo: periodo 2024 hasta primer semestre 2025.
7. Espacio: los países importadores.

### 3.1. Objetivos SMART

Limpiar el conjunto de datos históricos de exportaciones colombianas obtenido del período 2024 a primer semestre 2025, asegurando la calidad de los datos y el tratamiento de valores atípicos y faltantes.

1. Entrenar al menos cuatro modelos de Machine Learning (regresión lineal, regresión logística, K-Nearest Neighbors (KNN) y Árboles de Decisión para la clasificación/predicción del potencial de la demanda internacional.
2. Validar el modelo seleccionado, revisando las métricas.

3. Implementar modelos que visualice la predicción del potencial de demanda de clientes internacionales, permitiendo la priorización de los 10 clientes y/o mercados más dinámicos.

#### 4. Requerimientos del Sistema

1. Requerimientos funcionales para el modelo de ML:

Base de datos de la Dirección Nacional de Impuestos y Aduanas Nacionales (DIAN) con datos históricos de exportaciones colombianas obtenido del período 2024 a primer semestre 2025.

2. Requerimientos no funcionales para el modelo de ML:

Modelos que visualice la predicción del potencial de demanda de clientes internacionales, permitiendo la priorización de los 10 clientes y/o mercados más dinámicos.

Cuatro modelos estadísticos ML: regresión lineal, regresión logística, K-Nearest Neighbors (KNN) y Árboles de Decisión.

#### 5. Arquitectura Propuesta

##### 5.1 Descripción del Proyecto

Sistema integral de Machine Learning para análisis y predicción de exportaciones colombianas, que incluye:

- Análisis Exploratorio (EDA) y Transformación de Datos (ETL): Proceso de ciencia de datos y análisis, cargar, extraer y transformar los datos.
- Cuatro Modelos de ML supervisados (regresión lineal, regresión logística, K-Nearest Neighbors (KNN) y Árboles de Decisión).
- Aplicación FullStack que nos permite comparar los modelos y saber cuál es el mejor, de acuerdo a nuestra base de datos.



- Notebooks Jupyter: entornos interactivos basados en la web que permiten combinar código (Python, R, etc.), texto explicativo (en formato Markdown), ecuaciones, visualizaciones y salidas de forma secuencial en un único documento.

## 5.2 Estructura del Proyecto

Se describe la estructura teniendo en cuenta sólo los cuatro modelos practicados a elegir, según consultas previas y recomendaciones de la IA, sin embargo, sólo dos fueron con mejores métricas.

### *Data.*

DATAPROYECTO.xlsx: es la Dataset original (80,000 registros)

data\_clean.csv: datos limpios

data\_processed.csv: datos procesados para ML

data\_numeric.csv: datos numéricos

### *Notebooks.*

01\_EDA\_ETL.ipynb: Análisis exploratorio y limpieza

02\_Regresion\_Lineal.ipynb: modelo de regresión.

03\_Regresion\_Logistica.ipynb: clasificación multiclas.

04\_KNN.ipynb: Clasificar basándose en vecinos más cercanos.

06\_Arboles\_Decision.ipynb: clasificar mediante reglas jerárquicas.

### *Modelos entrenados.*

model\_regresion\_lineal.pkl

model\_regresion\_logistica.pkl

model\_knn.pkl

model\_decision\_tree.pkl



TIC

TALENTO  
TECH

*Aplicación fullstack.*

### *Documentación.*

README.md: archivo con la estructura

INSTALLATION.md: guía de instalación

API\_DOCUMENTATION.md: Documentación de API

USER\_GUIDE.md: guía de usuario

### *Utilidades.*

generate\_all\_notebooks.py: Script generador

## 5.3 Instalación y Configuración

### *Requisitos Previos.*

Python 3.8 o superior

pip (gestor de paquetes de Python)

8GB RAM mínimo

2GB espacio en disco

Paso 1: clonar/Descargar el Proyecto

Paso 2: crear Entorno Virtual

Paso 3: instalar Dependencias

Dependencias principales: pandas, numpy, scikit-learn, matplotlib, seaborn, flask, flask-cors, openpyxl, jupyter, scipy

Paso 4: ejecutar Notebooks

Paso 5: iniciar Aplicación Web: <http://localhost:5000>



## 5.4 Dataset

### *Información General.*

Nombre: DATAPROYECTO.xlsx

Registros: 80,000 exportaciones

Variables: 30 columnas

Período: datos de exportaciones colombianas

Tamaño: ~17 MB

### *Variables Principales.*

#### *Variables Numéricas.*

Valor FOB (USD): valor de la exportación

Peso en kilos netos: peso neto del producto

Peso en kilos brutos: peso bruto incluyendo embalaje

Cantidad(es): cantidad de unidades

Número de artículos: número de artículos en la exportación

Precio Unitario FOB (USD) Peso Neto: precio por unidad de peso

#### *Variables Categóricas.*

País de Destino: país al que se exporta

Continente Destino: continente de destino

Departamento Origen: departamento colombiano de origen

Vía de transporte: medio de transporte utilizado

Moneda de negociación: moneda utilizada

Forma de pago: método de pago



## 5.5 Modelos de Machine Learning

De los cuales se ampliará explicación en el numeral 8.1.

- Regresión Lineal (Supervisado - Regresión).
- Regresión Logística (Supervisado - Clasificación).
- K-Nearest Neighbors (KNN) (Supervisado - Clasificación).
- Árboles de Decisión (Supervisado - Clasificación).

## 5.6 Resultados y Comparación

El notebook 09\_Comparativo\_Integrador.ipynb` realiza una comparación exhaustiva:

Criterios de Evaluación

Para Clasificación:

Accuracy (30% peso): precisión general

F1-Score (25% peso): balance precision-recall

Precision (20% peso): predicciones correctas

Recall (15% peso): cobertura de clases

CV Accuracy (10% peso): estabilidad

Para Regresión:

R<sup>2</sup> Score: varianza explicada

RMSE: error cuadrático medio

MAE: error absoluto medio

Visualizaciones Incluidas.

Aplicación FullStack

Archivo



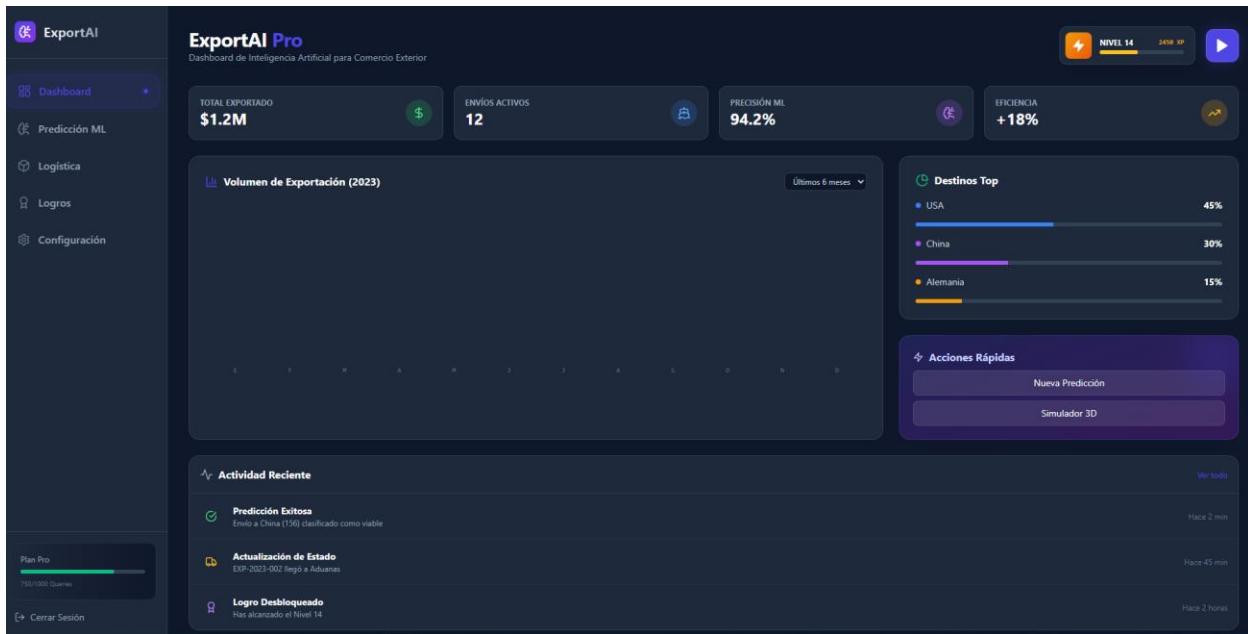
## 6. Aplicación del Ciclo de Vida del Machine Learning.

A continuación, se observa algunas capturas de pantalla como figuras de los principales flujos de navegación en aplicación del ciclo de vida de ML.

**Figura 1.** *El usuario ingresa a un dashboard claro y moderno donde ve de inmediato las métricas clave del sistema: exportaciones, envíos activos, precisión del modelo y eficiencia. El diseño oscuro con elementos resaltados facilita la lectura. El gráfico central muestra la actividad anual y, a la derecha, los países destino más relevantes. En la parte inferior, la actividad reciente resume las acciones del sistema, mientras que los botones de acciones rápidas permiten acceder fácilmente a nuevas predicciones y al simulador.* **Figura 2.** *El usuario accede al módulo de Predicción ML donde puede elegir entre diferentes modelos (clasificación, regresión y clustering). La interfaz organiza claramente los parámetros de carga, datos comerciales y logística, permitiendo ingresar información de forma rápida y ordenada. A la derecha, las métricas en vivo muestran el rendimiento del modelo en tiempo real. Finalmente, un botón destacado permite ejecutar la predicción de manera inmediata, haciendo el proceso intuitivo y guiado.*

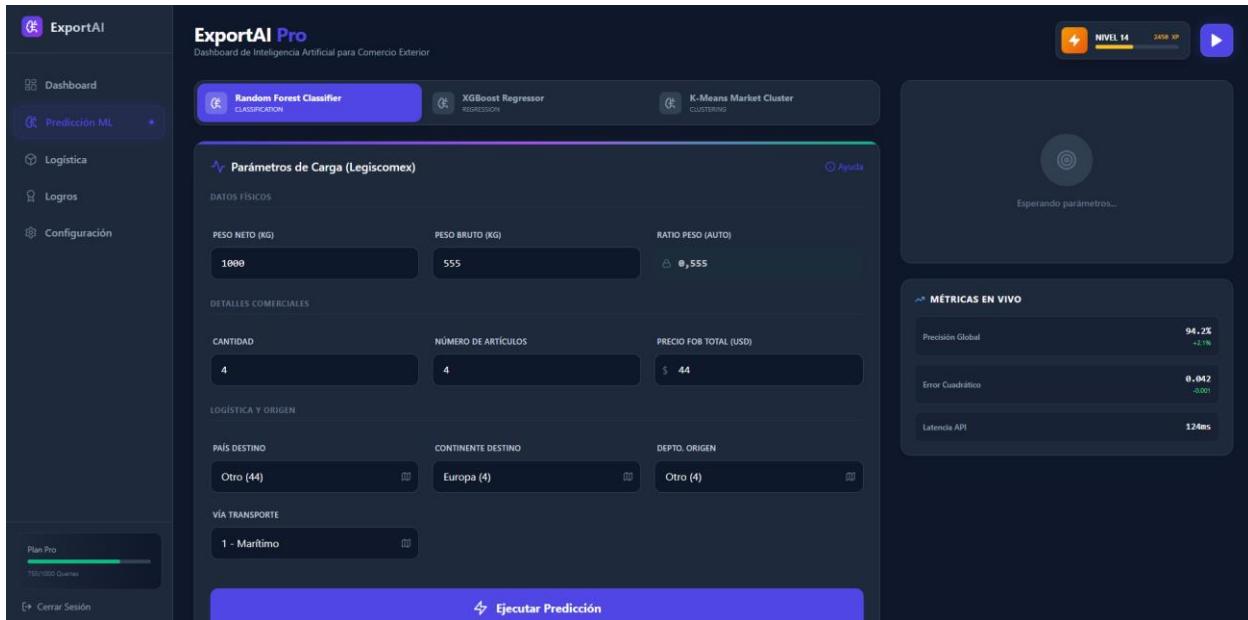
**Figura 3.** *Monitorear en tiempo real los despachos de exportación, mostrando: Cuántos están en tránsito, en aduanas o completados, el detalle de cada envío (destino, transporte, valor, fecha), y el estado actual de cada despacho (In Transit, Customs, Delivered, Pending).* **Figura 4.** *Este módulo muestra el nivel de experiencia del usuario dentro de la plataforma y las medallas o logros obtenidos según sus acciones en el sistema.* **Figura 5.** *Es la sección donde el usuario ajusta su perfil, activa o desactiva notificaciones y administra su seguridad y claves API.*

**Figura 1.** Captura de pantalla interfaz Export AI Dashboard



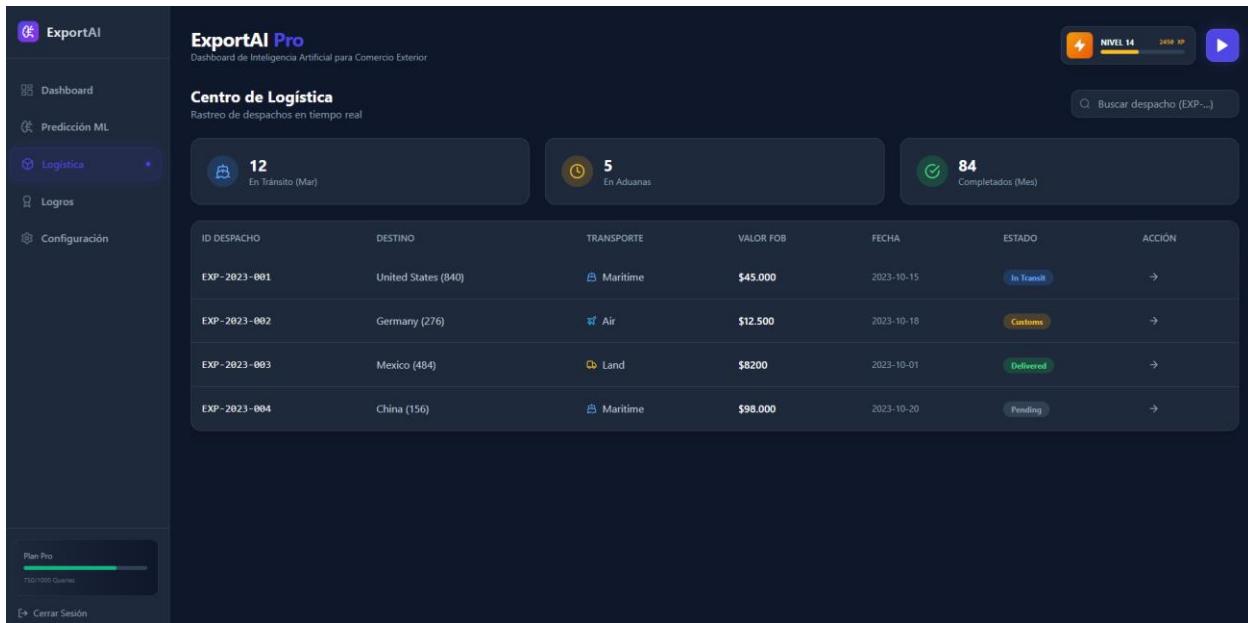
**Nota.** Interfaz Dashboard. Elaboración propia.

**Figura 2.** Captura de pantalla interfaz ExporAI Predicción ML



**Nota.** Interfaz parámetros de carga. Elaboración propia.

**Figura 3.** Captura de pantalla interfaz ExporAI Logística



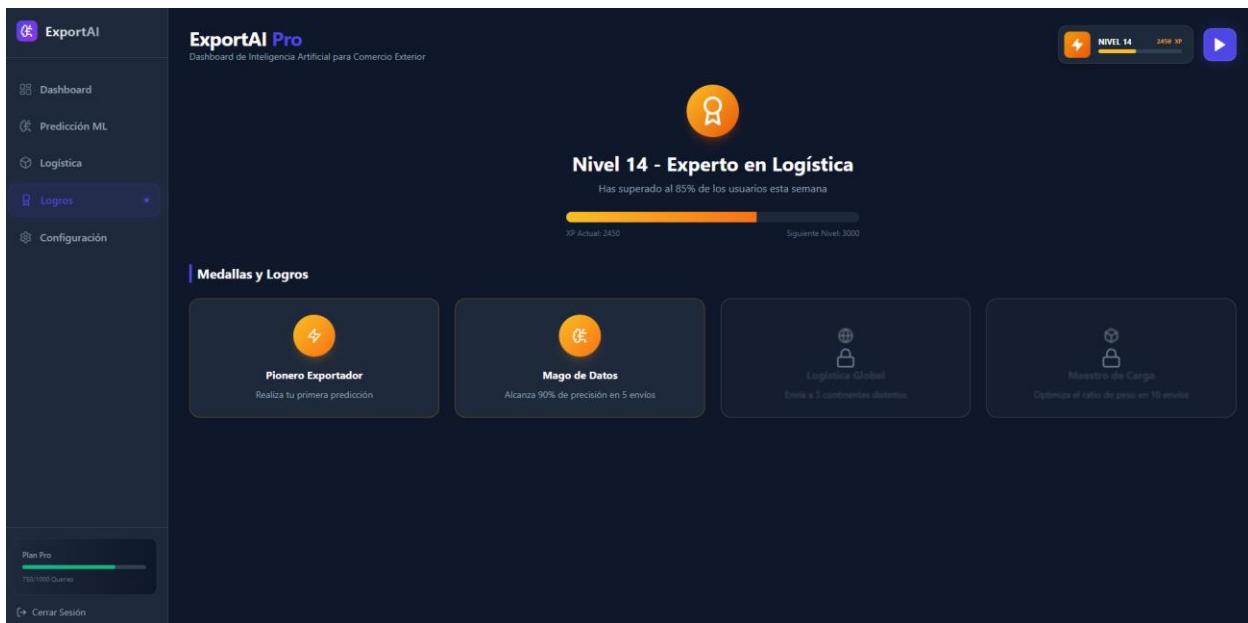
The screenshot shows the ExportAI Pro dashboard for Logística. On the left, a sidebar menu includes: Dashboard, Predicción ML, Logística (selected), Logros, and Configuración. A progress bar at the bottom indicates "Plan Pro" with 750/1000 Querries completed. The main area is titled "Centro de Logística" and displays "Rastreo de despachos en tiempo real". It features three summary cards: "12 En Tránsito (Mar)" (blue icon), "5 En Aduanas" (orange icon), and "84 Completados (Mes)" (green icon). Below these are four rows of shipping details:

ID DESPACHO	DESTINO	TRANSPORTE	VALOR FOB	FECHA	ESTADO	ACCIÓN
EXP-2023-001	United States (840)	Maritime	\$45.000	2023-10-15	In transit	→
EXP-2023-002	Germany (276)	Air	\$12.500	2023-10-18	Customs	→
EXP-2023-003	Mexico (484)	Land	\$8200	2023-10-01	Delivered	→
EXP-2023-004	China (156)	Maritime	\$98.000	2023-10-20	Pending	→

At the top right, there's a "NIVEL 14" badge with 2458 XP and a play button icon. A search bar says "Buscar despacho (EXP...)".

**Nota.** Interfaz rastreo de despachos en tiempo real. Elaboración propia.

**Figura 4.** Captura de pantalla interfaz ExporAI Logros



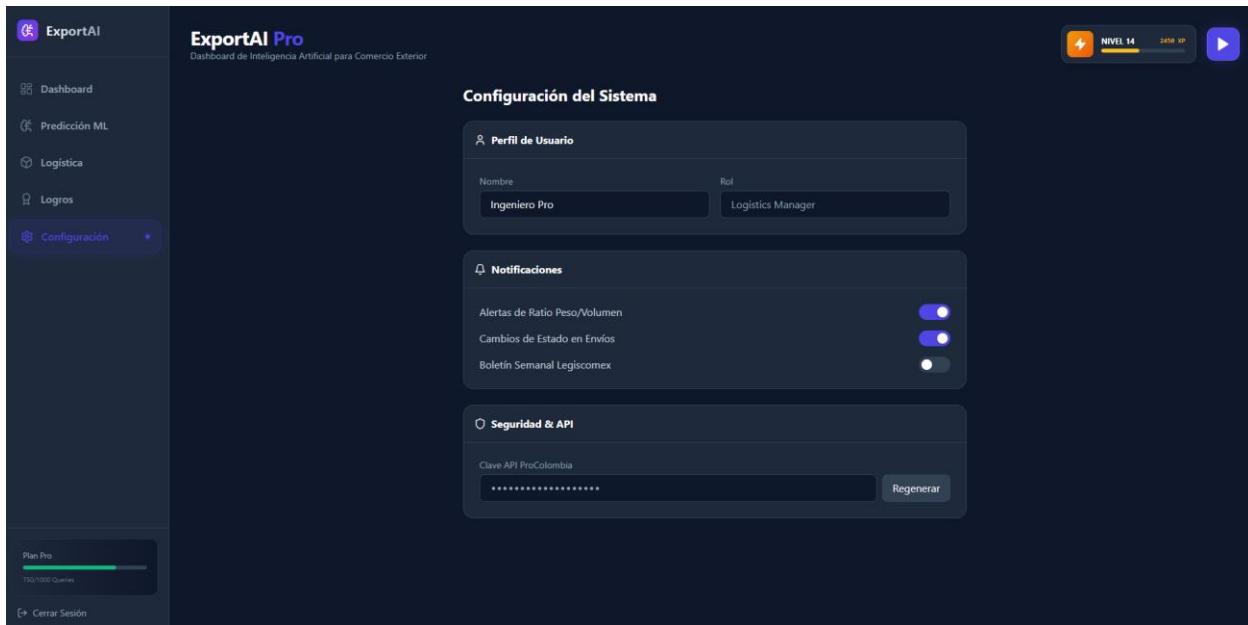
The screenshot shows the ExportAI Pro dashboard for Logros. The sidebar menu is identical to Figure 3. The main area features a large orange circular icon with a gear symbol. Below it, the text "Nivel 14 - Experto en Logística" is displayed, with a note "Has superado al 85% de los usuarios esta semana". A progress bar shows "XP Actual: 2450" and "Siguiente Nivel: 3000". A section titled "Medallas y Logros" lists four achievements:

- Pionero Exportador**: Realiza tu primera predicción
- Mago de Datos**: Alcanza 90% de precisión en 5 envíos
- Logística Global**: Envía a 3 continentes distintos
- Maestro de Carga**: Optimiza el ratio de peso en 10 envíos

At the bottom, there's a progress bar for "Plan Pro" with 750/1000 Querries completed, and a "Cerrar Sesión" button.

**Nota.** Interfaz medallas y logros. Elaboración propia.

**Figura 5.** Captura de pantalla interfaz ExporAI Configuración



**Nota.** Interfaz configuración del sistema. Elaboración propia.

Respecto a la experiencia como usuarios, se obtuvo la base de datos de la dirección: Base de datos de la Dirección Nacional de Impuestos y Aduanas Nacionales (DIAN). A partir de ella, se formuló un prompt con el problema de nuestro proyecto y se anexó la base de datos en Genspark AI, para realizar el EDA, el proceso ETL y algunos modelos de ML, con el fin de desarrollar un prototipo funcional.

Mediante trabajo en equipo se migró la carga de los resultados al drive, se visualizó los gráficos e interacciones de las variables, se observó los gráficos de barras, matrices, métricas, entre otros, de cuatro modelos utilizados, revisando el Full Stack y se concluyó sobre dos modelos finales, los cuales fueron el Árbol de Decisión y KNN.

Para el desarrollo del proyecto se utilizaron diversas herramientas de IA y plataformas que apoyaron cada etapa del proceso, algunas fueron: Manus AI, Google Drive, Netlify, Invideo, Google Gemini y Github principalmente, luego, se cargó al repositorio Github siguiendo indicaciones para su entrega.



## 7. Plan de Pruebas

Modelos aplicados de Machine Learning Supervisados como los siguientes:

### 7.1 Modelos de Machine Learning

#### *Regresión Lineal (Supervisado - Regresión).*

Objetivo: predecir el Valor FOB (USD) de las exportaciones

Variables:

Dependiente: valor FOB (USD)

Independientes: peso, cantidad, precio unitario, país,  
departamento, etc.

Métricas de Evaluación:

R<sup>2</sup> Score

RMSE (Root Mean Squared Error)

MAE (Mean Absolute Error)

MAPE (Mean Absolute Percentage Error)

Archivo: 02\_Regresion\_Lineal.ipynb, model\_regresion\_lineal.pkl

#### *Regresión Logística (Supervisado - Clasificación).*

Objetivo: clasificar exportaciones en categorías de valor

(Bajo/Medio/Alto/Muy Alto)

Variables:

Dependiente: categoria\_Valor (4 clases)

Independientes: variables numéricas y codificadas

Métricas de Evaluación:

Accuracy (Precisión)



Precision

Recall

F1-Score

Matriz de Confusión

Archivo: 03\_Regresion\_Logistica.ipynb, model\_regresion\_logistica.pkl

### **K-Nearest Neighbors (KNN) (Supervisado - Clasificación).**

Objetivo: clasificar exportaciones basándose en vecinos más cercanos

Parámetros:

n\_neighbors = 5

weights = distance

metric = euclidean

Características:

No asume distribución de datos

Adapta a datos no lineales

! Computacionalmente costoso

! Sensible a escala de features

Archivo: 04\_KNN.ipynb, model\_knn.pkl

### **Árboles de Decisión (Supervisado - Clasificación).**

Objetivo: Clasificar mediante reglas jerárquicas

Parámetros:

max\_depth = 10

min\_samples\_split = 20

min\_samples\_leaf = 10

## Características:

Altamente interpretable

Maneja datos no lineales

No requiere normalización

 Propenso a overfitting

 Inestable con pequeños cambios en datos

Archivo: 06\_Arboles\_Decision.ipynb, model\_decision\_tree.pkl

## 7.2 Resultados esperados y observados

Se esperaba observar principalmente, entre otras variables, qué continentes y países son los más importadores, las mercancías o productos con mayor movimiento en las exportaciones y los valores comerciales de estos flujos. Por lo cual, basados en las métricas de rendimiento, gráficos y demás, el modelo más adecuado para predecir el Continente Destino de las exportaciones es el K-Nearest Neighbors (KNN).

En dichos resultados, se obtuvo información de los países mayores importadores, y la participación que tiene cada continente en dichos desplazamientos de productos, las mercancías que más circulan, la clasificación del valor de exportación por categorías bajo, medio, alto, muy alto, entre otra información relevante.

Aunque el Árbol de Decisión y KNN tienen un rendimiento muy similar, el Árbol de Decisión muestra una ligera ventaja en precisión y F1-Score, mientras, el modelo KNN presenta la mayor exactitud (0.9335).

- KNN (K-Nearest Neighbors): muestra la mejor exactitud (0.9335) y un excelente balance en las demás métricas (F1-Score de 0.9306). Este modelo es robusto y funciona bien en espacios de

características bien definidos, lo que sugiere que las exportaciones a un continente tienden a agruparse en el espacio de características transformado.

- Árbol de Decisión: ofrece el mejor F1-Score (0.9333), lo que indica un excelente equilibrio entre precisión y exhaustividad. Es una alternativa muy fuerte, especialmente si se requiere interpretabilidad.

Recomendación Final: Se recomienda el modelo K-Nearest Neighbors (KNN) por su rendimiento superior en la métrica de exactitud, que es a menudo la métrica principal para problemas de clasificación balanceados. Sin embargo, el Árbol de Decisión es una alternativa casi idéntica en rendimiento y ofrece la ventaja de ser más interpretable.

## 8. Evidencias del Desarrollo

De acuerdo al desarrollo del prototipo, se observa en entorno de desarrollo, a continuación,

**Figura 6.** *Captura de pantalla del Data Proyecto* utilizado, en la que se observan celdas importantes como país destino, mercancía, entre otros. Adicional en **Figura 7.** *Captura de pantalla archivos en Google Drive*, encontramos las carpetas de la estructura del Proyecto cargadas en el drive. Igualmente,

**Figura 8.** *Captura de pantalla los Notebook*, nos muestra los modelos del sistema, de los cuales practicamos cuatro mencionados anteriormente en este trabajo. En la **Figura 9.** *Captura de pantalla del Modelo de Árboles de Decisión*, se muestra capturas del entorno en desarrollo para dicho modelo.

**Figura 10.** *Top 10 características más importantes del modelo de Árboles de decisión*, el cual muestra la mayor predicción de las variables. **Figura 11.** Métricas de evaluación del modelo de Árboles de Decisión, las cuales son muy confiables y altas para nuestros datos, por encima de 0.90. Se observa el objetivo en línea del Modelo K-Nearest Neighbors en la **Figura 12.** *Captura de pantalla del Modelo de KNN*. En la

**Figura 13.** *Distribución de la Variable Target*, se muestra las Categorías del valor alto, najo, medio, muy alto. **Figura 14.** *Métricas de evaluación optimizada del modelo KNN*, se observa el valor de las Métricas calculadas para el modelo aún optimizado, siendo por encima de 0.70. **Figura 15.** *Captura de pantalla*

del Modelo de Regresión Lineal, se observa en línea el entorno y el objetivo de dicho modelo. **Figura 16.**

Estadística de la Variable dependiendo FOB USD en Modelo Regresión Lineal, nos muestra datos de la variable objetivo. **Figura 17.** Captura de pantalla del Modelo de Regresión Logística, donde se muestra el objetivo de dicho modelo. **Figura 18.** Métricas de evaluación del Modelo de Regresión Logística, se observa Reporte de las métricas, los cuales son valores bajos.

Enlace al repositorio GitHub: [GitHub - mundoexport21-afk/Proyecto-Final-: Proyecto final de Machine Learning Talento Tech 2](https://github.com/mundoexport21-afk/Proyecto-Final-)

**Figura 6.** Captura de pantalla del Data proyecto

A	B	C	D	E	F	G	H	I	J	K	L	
1	Razón social del importador	País de Destino	Mes de la	Capítulo Del Arancel	Código	Descripción de la	Descripción de la	Fecha de	Número de la	Aduana De	Código	Agente(s)
2	CONTINENTAL FARF	ESTADOS UNIDOS	01	Plantas vivas y product	0603129000	Los demás claveles		20250101	6007748194791	BOGOTÁ	830074208	AGENCIA
3	CONTINENTAL FARF	ESTADOS UNIDOS	01	Plantas vivas y product	0603199090	Los demás flores y		20250101	6007748194791	BOGOTÁ	830074208	AGENCIA
4	CONTINENTAL FARF	ESTADOS UNIDOS	01	Plantas vivas y product	0603199090	Los demás flores y		20250101	6007748194791	BOGOTÁ	830074208	AGENCIA
5	CONTINENTAL FARF	ESTADOS UNIDOS	01	Plantas vivas y product	0603193000	Alstromerias frescas,	Especie:ALSTROEMERI	20250101	6007748194791	BOGOTÁ	830074208	AGENCIA
6	BANDY RANCH FILO	ESTADOS UNIDOS	01	Plantas vivas y product	0603141000	Pompones frescos,	Especie:POMPON ,	20250101	6007748230406	MEDELLÍN	830023585	AGENCIA
7	AVIANCA INC	ESTADOS UNIDOS	01	Plástico y sus manufac	3917329000	Los demás tubos sin	Nombre	20250101	6007748220773	MEDELLÍN	800251957	SIACO SIA
8	AVIANCA INC	ESTADOS UNIDOS	01	Manufacturas de fundi	73181159000	Los demás tornillos y		20250101	6007748210201	BOGOTÁ	800251957	SIACO SIA
9	AVIANCA INC	ESTADOS UNIDOS	01	Manufacturas de fundi	7307990000	Los demás accesorios		20250101	6007748210201	BOGOTÁ	800251957	SIACO SIA
10	AVIANCA INC	ESTADOS UNIDOS	01	Plástico y sus manufac	3926903000	Tornillos, pernos,		20250101	6007748210201	BOGOTÁ	800251957	SIACO SIA
11	AVIANCA INC	ESTADOS UNIDOS	01	Plástico y sus manufac	3926903000	Tornillos, pernos,		20250101	6007748210201	BOGOTÁ	800251957	SIACO SIA
12	AVIANCA INC	ESTADOS UNIDOS	01	Aeronaves, vehículos e	8802400000	Aviones y demás	Nombre	20250101	6007748210201	BOGOTÁ	800251957	SIACO SIA
13	AVIANCA INC	ESTADOS UNIDOS	01	Reactores nucleares, c:	8412290000	Los demás motores		20250101	6007748210201	BOGOTÁ	800251957	SIACO SIA
14	AVIANCA INC	ESTADOS UNIDOS	01	Reactores nucleares, c:	8471800000	Las demás unidades de	Nombre	20250101	6007748210184	BOGOTÁ	800251957	SIACO SIA
15	AVIANCA INC	ESTADOS UNIDOS	01	Reactores nucleares, c:	8471800000	Las demás unidades de	Nombre	20250101	6007748210184	BOGOTÁ	800251957	SIACO SIA
16	AVIANCA INC	ESTADOS UNIDOS	01	Aluminio y sus manufa	7616100000	Puntas, clavos, grapas		20250101	6007748131315	BOGOTÁ	800251957	SIACO SIA
17	AVIANCA INC	ESTADOS UNIDOS	01	Herramientas y útiles, i	8204120000	Llaves de ajustes de	Nombre	20250101	6007748131315	BOGOTÁ	800251957	SIACO SIA
18	AVIANCA INC	ESTADOS UNIDOS	01	Instrumentos y aparato	9014900000	Partes y accesorios de		20250101	6007748131331	BOGOTÁ	800251957	SIACO SIA
19	AVIANCA INC	ESTADOS UNIDOS	01	Cauchó y sus manufact	4008211000	Placas, hojas y tiras, de		20250101	6007748131331	BOGOTÁ	800251957	SIACO SIA
20	AVIANCA INC	ESTADOS UNIDOS	01	Cauchó y sus manufact	4016991000	Otros artículos para		20250101	6007748131331	BOGOTÁ	800251957	SIACO SIA
21	ZOELLNER E HIJOS C CHILE	ESTADOS UNIDOS	01	Manufacturas diversas	9602001000	Cápsulas de gelatina	Nombre	20250101	6007748194435	CARTAGENA	890105424	AGENCIA
22	AVIANCA INC	ESTADOS UNIDOS	01	Cauchó y sus manufact	4016930000	Juntas o		20250101	6007747869425	BOGOTÁ	800251957	SIACO SIA

**Nota.** Base de datos empleada. Elaboración propia.

**Figura 7.** Captura de pantalla archivos en Google Drive

Compartidos con... > ... > Proyecto Final ML\_F...

Tipo Personas Modificado Fuente

Nombre ↑

- 📁 DOCUMENTACIÓN
- 📁 NOTEBOOKS
- 📁 DATA
- 📁 exportai-pro
- 📁 FULLSTACK
- 📁 exportai-pro.zip ↗

*Nota.* Despliegue del prototipo en drive. Elaboración propia.

**Figura 8.** Captura de pantalla los Notebook

Compartidos connigo > ... > PROTOTIPO

Tipo Personas Modificado Fuente

Nombre ↑

- 📄 01\_EDA\_ETL.ipynb ↗
- 📄 02\_Regresión\_Lineal.ipynb ↗
- 📄 03\_Regresión\_Logística.ipynb ↗
- 📄 04\_KNN.ipynb ↗
- 📄 05\_SVM.ipynb ↗
- 📄 06\_Arboles\_Decisión.ipynb ↗
- 📄 07\_Naive\_Bayes.ipynb ↗
- 📄 08\_KMeans.ipynb ↗

*Nota.* Despliegue de los Notebooks. Elaboración propia.

**Figura 9.** Captura de pantalla del Modelo de Árboles de Decisión

### Modelo 5: Árboles de Decisión

#### Clasificación Basada en Reglas de Decisión

##### Objetivo

Clasificar exportaciones colombianas en categorías de valor usando el algoritmo de Árboles de Decisión.

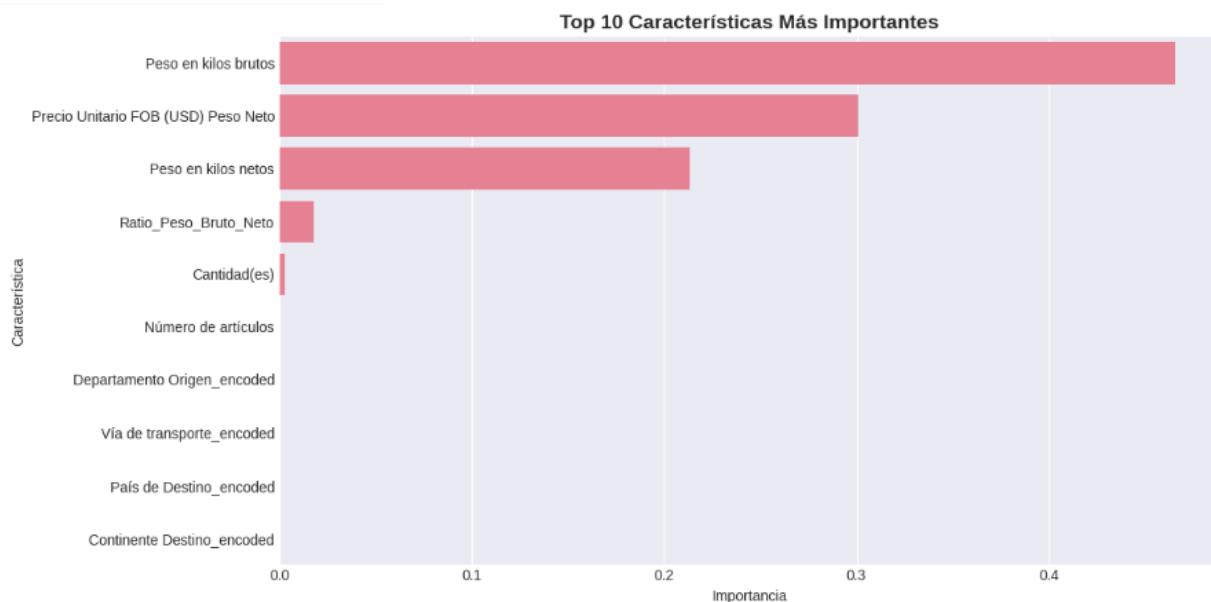
Los árboles de decisión son modelos predictivos que mapean observaciones sobre un ítem a conclusiones sobre el valor objetivo del ítem. Son fáciles de interpretar y pueden manejar tanto datos numéricos como categóricos.

##### Variables

- **Target (Variable objetivo):** Categoría\_Valor - Clasificación del valor FOB en categorías (Bajo/Medio/Alto/Muy Alto)
- **Features (Características):** Variables numéricas y categóricas codificadas relacionadas con las exportaciones

**Nota.** Objetivo del Modelo Árbol de decisión. Elaboración propia.

**Figura 10.** Top 10 características más importantes del modelo de Árboles de decisión.



- Interpretación:**
- Mayor importancia = mayor poder predictivo
  - Características con importancia 0 no se usan en el árbol
  - Características utilizadas: 8 de 11

**Nota.** Mayor importancia del poder predictivo de variables del Modelo. Elaboración propia.

**Figura 11.** Métricas de evaluación del modelo de Árboles de Decisión.

```

 MÉTRICAS DE EVALUACIÓN:
=====
Accuracy Train: 0.9753
Accuracy Test: 0.9669
Precision: 0.9670
Recall: 0.9669
F1-Score: 0.9669

 Rendimiento equilibrado

```

**Nota.** Rendimiento equilibrado del modelo. Elaboración propia.

**Figura 12.** Captura de pantalla del Modelo de KNN

 **Modelo 3: K-Nearest Neighbors (KNN)**

Clasificación por Vecinos Más Cercanos

---

 **Objetivo**

Clasificar exportaciones colombianas en categorías de valor usando el algoritmo K-Nearest Neighbors (KNN).

El KNN es un algoritmo de aprendizaje supervisado que clasifica nuevos ejemplos basándose en la similitud con los k ejemplos más cercanos en el espacio de características.

 **Variables**

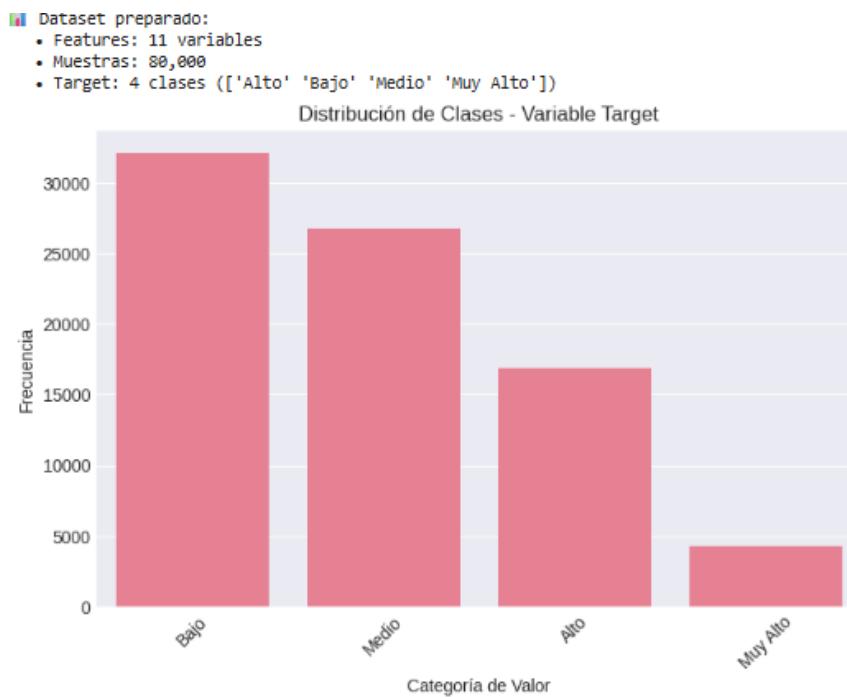
- **Target (Variable objetivo):** `Categoría_Valor` - Clasificación del valor FOB en categorías (Bajo/Medio/Alto/Muy Alto)
- **Features (Características):** Variables numéricas y categóricas codificadas relacionadas con las exportaciones

 **Contenido**

1. Importación de librerías
2. Carga y exploración de datos
3. Preprocesamiento de datos
4. Preparación de features y target
5. Entrenamiento del modelo KNN
6. Evaluación del modelo
7. Validación cruzada
8. Guardado del modelo

**Nota.** Objetivo del Modelo K-Nearest Neighbors. Elaboración propia.

**Figura 13.** Distribución de la Variable Target



**Nota.** Categorías del valor alto, bajo, medio, muy alto. Elaboración propia.

**Figura 14.** Métricas de evaluación optimizada del modelo KNN

■ MÉTRICAS DE EVALUACIÓN DEL MODELO OPTIMIZADO:  
=====

Accuracy Test (Optimizado):	0.7057
Precision (Optimizado):	0.7081
Recall (Optimizado):	0.7057
F1-Score (Optimizado):	0.7045

✓ Métricas calculadas para el modelo KNN optimizado.

**Nota.** Métricas calculadas para el modelo KNN optimizado. Elaboración propia.

**Figura 15.** Captura de pantalla del Modelo de regresión Lineal

**Modelo 1: Regresión Lineal**

Predicción del Valor FOB de Exportaciones

*Nota.*

**Objetivo del Modelo**

**Predictir el valor FOB (USD) de las exportaciones** basándose en características como peso, cantidad, departamento de origen y país de destino.

*Objetivo*

**Tipo de Problema**

**Regresión** - Predicción de una variable continua (valor monetario)

*del*

**Variables del Modelo**

Variable Dependiente (Target):

*Modelo*

- **Valor FOB (USD):** Valor de la exportación en dólares estadounidenses

**Regresión Lineal.** Elaboración propia.

*de*

**Figura 16.** Estadística de la Variable dependiendo FOB USD en Modelo Regresión Lineal



**Nota.** Datos de la variable objetivo en estadísticas. Elaboración propia.



**Figura 17.** Captura de pantalla del Modelo de Regresión Logística

Modelo 2: Regresión Logística

Clasificación de Categorías de Valor de Exportación

Objetivo del Modelo

Clasificar las exportaciones en categorías de valor: Bajo, Medio, Alto, Muy Alto

Tipo de Problema

Clasificación Multiclas - Predecir categorías discretas

Variables del Modelo

Variable Dependiente (Target):

- Categoría\_Value: Clasificación del valor FOB (Bajo/Medio/Alto/Muy Alto)

**Nota.** Objetivo del Modelo de Regresión Logística. Elaboración propia.

**Figura 18.** Métricas de evaluación del Modelo de Regresión Logística.

MÉTRICAS DE EVALUACIÓN

=====

Accuracy:

Train: 0.4316

Test: 0.4369

Métricas (Test - Weighted):

Precision: 0.4166

Recall: 0.4369

F1-Score: 0.4079

CLASSIFICATION REPORT:

	precision	recall	f1-score	support
0	0.34	0.18	0.24	4000
1	0.50	0.64	0.57	4125
2	0.38	0.24	0.29	3875
3	0.43	0.67	0.53	4000
accuracy			0.44	16000
macro avg	0.42	0.43	0.41	16000
weighted avg	0.42	0.44	0.41	16000

**Nota.** Reporte de las métricas del modelo. Elaboración propia.



## 9. Conclusiones

Dentro de los principales aprendizajes, se tienen el uso de diferentes aplicaciones muy prácticas para realizar diferentes objetivos planteados. Cómo dificultad se encontró algunos momentos en la programación de Python, ya que arrojaba algunos errores y la capacidad de uso gratis de algunas aplicaciones y el tiempo prolongado de trabajo del sistema. Sin embargo, se logró elaborar diferentes objetos virtuales, prototipos, análisis estadísticos, entre otros, como este proyecto; que de acuerdo a su desarrollo y basado en los modelos y las métricas arrojadas por la IA, se observó que los modelos ML: KNN Y Árbol de decisiones fueron los más adaptados a nuestra base de datos.

Este proyecto fue desarrollado como parte de un trabajo académico y profesional de ML, por los autores Santiago Ceballos, Mauricio Alvarado Cardona, María Úsuga y Astrid Agudelo M. Fecha: 2025.

Versión: 1.0.0.

## 10. Bibliografía

Dirección Nacional de Impuestos y Aduanas Nacionales (DIAN). (2025). Base de datos 2024 – 2025.

ExportAI Pro (<https://curious-gingersnap-e15d43.netlify.app/>). (2025). Dashboard de Inteligencia Artificial para Comercio Exterior.

Flask Documentation (<https://flask.palletsprojects.com/>). (2025).

Genspark AI (<https://www.genspark.ai/>). (2025).

Github (<https://github.com>). (2025).

Google Gemini (<https://gemini.google.com/app?hl=es>). (2025).

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. (2025).

James, G., et al. (2013). An Introduction to Statistical Learning.

Manus AI (<https://manus.im>). (2025).



Neflity (<https://www.netlify.com/>). (2025).

Pandas Documentation (<https://pandas.pydata.org/docs/>). (2025).

Scikit-learn Documentation (<https://scikit-learn.org/stable/>). (2025).