



MODELO PREDICTIVO

DE LAS EXPORTACIONES COLOMBIANAS

Ejecutores

Margarita María Orozco
Natalia Betancur Herrera
Frank Yesid Zapata Castaño
Camila Díaz

Participantes

José Fernando García
Rubén Darío García
Gerardo D'La Vid



| Tabla de contenido | Página |
|-------------------------------------|--------|
| 1. Introducción | 3 |
| 2. Planteamiento del problema | 3 |
| 3. Objetivos | 4 |
| 3.1. Objetivo General | 4 |
| 3.2. Objetivos específicos | 4 |
| 4. Justificación | 5 |
| 5. Alcance | 5 |
| 6. Presupuesto | 6 |
| 6.1. Costos Directos | 6 |
| 6.1.1. Recursos Humanos | 6 |
| 6.1.2. Costos de contingencia | 7 |
| 6.2. Resumen de costos | 7 |
| 6.3. Descripción del recurso humano | 7 |
| 7. Metodología utilizada | 8 |
| 8. Modelado de datos | 10 |
| 9. Tipos de variables | 10 |
| 10. Bases de datos y variables | 10 |
| 10.1. Análisis de las variables | 13 |
| 11. Modelo utilizado y resultados | 13 |
| 11.1. Análisis estadísticos | 17 |
| 12. Anexos | 22 |
| 13. Bibliografía | 22 |

1. Introducción

El año 2020 fue para el mundo del siglo XXI un punto de quiebre significativo al enfrentar una pandemia que, además de cuestionar el vivir y producir actual, nos llevó a situaciones de gran preocupación, dada la dificultad que ante nuestros ojos y los de los científicos y autoridades mundiales se presentaba: LA SUPERVIVENCIA.

No era solamente tratar de salir vivos de ella, sino aprender y trascender, reinventando procesos, y dando un trato mejor y más adecuado al planeta, a sus campos y moradores, de tal manera que, entre todos pudiéramos acogernos, protegernos y mediante alianzas, transformar al mundo en un entorno saludable, próspero y seguro para los que quedaran vivos y para las generaciones venideras.

En esta carrera por la vida, todas las cosas o al menos la mayoría de la vida en las ciudades quedó paralizada, mientras que los campos y cultivos continuaron sus actividades, logrando así, abastecer y mitigar las necesidades básicas de quienes estábamos acuartelados.

Como resultado de esta pandemia, muchas actividades, personas, productos y procesos cambiaron o desaparecieron, dando paso a otras iniciativas, a otras necesidades y a otras economías que redefinieron su prioridad, garantizando que el mundo siguiera su rumbo como si nunca hubiera pasado nada.

En un mundo cada vez más globalizado, las exportaciones juegan un papel fundamental en la economía de Colombia, ya que representan una fuente clave de divisas y crecimiento; sin embargo, la fluctuación de los mercados internacionales, los cambios en las políticas comerciales y los factores externos, como la pandemia de COVID-19, cambios de gobierno, o las crisis económicas globales, han llevado a que las predicciones sean complejas y desafiantes.

Este proyecto se enfoca en el desarrollo de un modelo de predicción para las exportaciones colombianas, utilizando datos históricos de los años 2019 a 2024, con los que se buscará identificar patrones y generar proyecciones que ayuden a las empresas y entes gubernamentales a anticiparse a los cambios y tomar decisiones informadas, mejorando la competitividad y la capacidad de adaptación de las exportaciones Colombianas.

2. Planteamiento del Problema

Colombia es un país agrícola por excelencia con una gran diversidad de cultivos, climas y productos de primer orden que ha ido ganando aceptación en los mercados internacionales, gracias al trabajo de los productores para mejorar la calidad de estos y cumplir con los estándares exigidos por países con los que Colombia ha suscrito tratados de libre comercio.

Colombia con su economía abierta, depende significativamente de las exportaciones para su desarrollo económico. Los sectores clave de exportación incluyen productos agrícolas, pecuarios, petróleo, minería y manufacturas. Sin embargo, la dinámica de las exportaciones se ve influenciada por una variedad de factores tanto internos como externos, incluyendo políticas económicas, acuerdos comerciales, fluctuaciones de precios internacionales, tipo de cambio, y condiciones globales como crisis económicas o desastres naturales.

La incertidumbre generada por estos factores hace que la previsión de exportaciones sea una tarea compleja y fundamentalmente necesaria para optimizar los recursos, establecer estrategias de mercado y generar políticas públicas acertadas que fomenten especialmente las exportaciones de productos no minerales, que históricamente han liderado la balanza exportadora de nuestro país.

Los métodos tradicionales de análisis no siempre logran captar la dinámica completa del gran número de factores que intervienen en el mercado exportador, lo que dificulta la planificación a largo plazo, revelando la ausencia de herramientas predictivas robustas que permitan anticipar con precisión las fluctuaciones en las exportaciones. Esta falta de previsión puede llevar a decisiones inadecuadas en términos de inversión, producción y comercio exterior, lo cual afecta la competitividad y estabilidad del sector exportador colombiano.

Mediante el análisis de la información histórica recolectada del DANE y la DIAN por los años 2019 a 2024, se plantea analizar el comportamiento de las exportaciones colombianas considerando las variaciones presentadas especialmente en el periodo de pandemia, y centrando la atención en aquellos productos y mercados que sean más representativos, proporcionando dicha información a modelos predictivos basados en Inteligencia Artificial, que puedan proyectar tendencias futuras de exportaciones en determinados mercados internacionales.

3. Objetivos

3.1. Objetivo General

Construir un modelo predictivo que estime los valores de las exportaciones colombianas a corto y mediano plazo, basado en datos históricos y variables clave.

3.2. Objetivos específicos

3.2.1. Analizar los datos históricos de exportaciones de Colombia entre 2019 y 2024 para identificar patrones y comportamientos significativos.

3.2.2. Implementar y evaluar diferentes modelos de predicción, como series temporales y algoritmos de aprendizaje automático, para determinar cuál ofrece el pronóstico más preciso.

3.2.3. Desarrollar, mediante modelos predictivos de Inteligencia Artificial, unos algoritmos que permitan analizar y sacar provecho de momentos, tendencias y otras variables que nos ayuden a proyectar las actividades del mercado exportador.

4. Justificación

Nuestro proyecto de MODELO PREDICTIVO DE LAS EXPORTACIONES COLOMBIANAS nos acerca al correcto aprovechamiento de los recursos, enfocando esfuerzos en dinamizar el mercado exportador colombiano.

El avance tecnológico ha transformado la manera en que las industrias analizan y procesan datos. En el contexto de la predicción de exportaciones, los avances en inteligencia artificial (IA), machine learning (aprendizaje automático) y análisis de grandes volúmenes de datos (big data) ofrecen oportunidades sin precedentes para realizar pronósticos mucho más precisos y complejos que los métodos tradicionales.

Mientras que las técnicas clásicas de predicción (como los modelos estadísticos univariantes) se basan en aproximaciones simplificadas de los datos históricos, las nuevas tecnologías permiten integrar múltiples fuentes de información y considerar interacciones complejas entre diversas variables. Por ejemplo, mediante el uso de redes neuronales y algoritmos de aprendizaje profundo, es posible capturar patrones no lineales y adaptarse a cambios repentinos en las condiciones del mercado, algo que los modelos tradicionales no pueden hacer con eficacia.

Este enfoque tecnológico no solo optimiza la precisión de las predicciones, sino que también permite una respuesta más rápida ante las dinámicas cambiantes del comercio internacional. De esta forma, la aplicación de estos avances tecnológicos es clave para mejorar la competitividad y la resiliencia del sector exportador colombiano, posicionando al país en un entorno global más dinámico y competitivo.

Con estas herramientas de apoyo, las empresas y en general los productores, prestarán más atención a las necesidades reales de los mercados, diversificarán cadenas productivas, aprovecharán los tiempos, los suelos, las líneas de producción, los recursos humanos y demás recursos, con un enfoque más eficiente.

5. Alcance

Nuestro proyecto se enfocará en mostrar el comportamiento histórico de las exportaciones Colombianas hacia los mercados internacionales, en etapas de Pre Pandemia, Pandemia y pos Pandemia, entendiendo que, las actividades propias de exportaciones debieron invertir

en: educación, capacitación, motivación e interacción con los consumidores finales llevando productos mejorados, competitivos en precio y calidad, garantizándoles buena aceptación y buenas utilidades, además de su permanencia como productores y exportadores de bienes y servicios teniendo en cuenta que la gran mayoría de hábitos de consumo se transformaron, cambiaron o se reinventaron, así mismo la oferta debe acomodarse a la nueva normalidad.

El alcance del proyecto abarca el análisis de los datos históricos de exportaciones de Colombia desde el año 2019 hasta el 2024, los cuales serán utilizados para entrenar y evaluar el modelo predictivo. Este análisis se realizará considerando los principales sectores exportadores como minería, petróleo, agricultura, y manufacturas; además, el modelo incluirá variables externas clave, como los cambios en los precios internacionales, políticas comerciales y factores macroeconómicos globales que afectan directamente las exportaciones de Colombia.

El proyecto también buscará proporcionar un marco de referencia útil para futuros estudios, facilitando la creación de modelos más detallados o sectoriales que puedan ajustarse mejor a las particularidades de cada industria.

El proyecto se implementará en varias fases, con la primera enfocada en la recolección y limpieza de los datos históricos, seguida por el desarrollo del modelo predictivo, su validación y finalmente la elaboración de recomendaciones basadas en los resultados obtenidos.

Este enfoque contribuirá a la creación de un sistema robusto que pueda ser utilizado de forma práctica tanto por las empresas como por las autoridades gubernamentales para la toma de decisiones estratégicas.

6. Presupuesto

6.1. Costos Directos

Corresponde a todos los costos en que se incurra para el desarrollo del proyecto y está segregado en los siguientes conceptos:

6.1.1. Recursos Humanos.

Hace referencia a los costos por salarios u honorarios del personal que estará involucrado directamente en la ejecución del proyecto.

Por ser una tarea de poca duración, se trabajará en modo outsourcing, es decir, cada persona contratada deberá poseer su computador con plan de datos para capturar, procesar la información y entregarla en los tiempos adecuados, siendo el centro de acopio de datos, la oficina del Ingeniero Informático, al cual le rendirán los informes respectivos en los tiempos asignados.

| CARGO | # EMPL | COSTO MES | CANT MESES | C TOTAL |
|--|--------|-------------|------------|---------------------|
| Analista de Datos | 1 | \$5.000.000 | 4 | \$20.000.000 |
| Trabajador de Campo | 2 | \$3.000.000 | 4 | \$24.000.000 |
| Auxiliar Informático | 1 | \$2.000.000 | 2 | \$ 4.000.000 |
| Ingeniero Informático | 1 | \$7.000.000 | 2 | \$14.000.000 |
| Ingeniero Informático de retroalimentación | 1 | \$7.000.000 | 3 | \$21.000.000 |
| Contingencias | | | | \$8.300.000 |
| TOTAL | | | | \$91.300.000 |

6.1.2. Se destina un 10% adicional para Contingencias relacionadas con el desarrollo del proyecto por valor de \$8.300.000

Costo total del proyecto \$91.300.000

6.2. Descripción del Recurso Humano

6.2.1. Analista de Datos: Refiere la persona que se encargará de encontrar las bases de datos de carácter público, las depurará y seleccionará las variables que aportarán la información más relevante para el proyecto, dimensionará los períodos de tiempo funcionales por períodos significativos y además por el peso en tamaño digital que será procesado por los equipos de cómputo sin que estos colapsen.

6.2.2. Trabajador de Campo: Hace referencia a las personas que se encargan de ubicar a las empresas exportadores actuales, pasados y aquellas que están convirtiéndose en exportadores y que tienen líneas de productos calificados y avalados por los mercados internacionales. Entre sus funciones deberá recopilar datos relevantes como: - Nombre de la compañía, - Ubicación geográfica, - Línea de producción a exportar, - Destinos y cantidades.

6.2.3. Auxiliar Informático: Es el encargado de ingresar la información a las bases de datos recopilada por los Trabajadores de Campo y con cuyos datos se aplicarán las herramientas informáticas y algoritmos que arrojarán los resultados esperados.

6.2.4. Ingeniero Informático: Es la persona que se encargará de analizar, entender, y diagramar los algoritmos para luego generar el código en lenguaje

de programación y trabajar así con las bases de datos previamente ingresadas, a las cuales se les aplicará los modelos predictivos de Inteligencia Artificial que ayudarán a desarrollar las tareas trazadas.

6.2.5. Ingeniero informático de Retroalimentación: Es la persona que dará soporte al ingeniero informático y será el puente entre éste y la empresa, articulando sus labores y haciendo las correcciones oportunamente, garantizando que se alcancen los objetivos trazados y que cada recurso destinado logre la eficiencia y eficacia para lo que fue asignado.

7. Metodología utilizada

Para analizar y entender el proceso adelantado, nos hemos apoyado en la metodología CRISP-DM, que rige el ciclo de vida para proyectos informáticos, el cual consta de varias etapas descritas a continuación:

7.1. Comprensión del negocio: En nuestro caso, comprendemos el proceso de exportación general de productos, los desafíos que enfrenta el sector exportador, y los objetivos específicos del modelo de exportación Colombiano.

7.2. Consecución de los datos: Para esta etapa, se requirió identificar repositorios y/o fuentes de información confiables que proveyeran datos útiles, confiables y que permitieran su manipulación y análisis.

7.3. Comprensión de los datos: Aquí identificamos todos los datos relevantes relacionados con las exportaciones (por ejemplo, demanda histórica de productos en diferentes mercados, estadísticas de exportación, tendencias de consumo, costos de transporte, etc.).

7.4. Preparación de los datos: Procesamiento y depuración de los datos. Este proceso involucra la normalización de datos, eliminación de valores atípicos, manejo de datos faltantes y eliminación de variables innecesarias.

7.5. Modelado: Como hemos mencionado, por el tipo de datos disponibles, que son etiquetados, se utilizará un modelo de **aprendizaje supervisado** que se ajuste a los objetivos de predicción de la demanda de nuestros productos en diferentes mercados y nos apoyaremos en **algoritmos de clasificación** que nos ayuden a identificar los países con más necesidades y los productos más demandados a nivel internacional.

7.6. Evaluación del modelo: Necesariamente nuestro modelo entra en una fase de monitoreo para saber si cumple con los objetivos trazados. No es suficiente con que arroje algunos resultados sino validar si éstos se ajustan y/o son creíbles y si nos está mostrando patrones confiables.

La eficiencia de nuestro modelo es fundamental, por lo tanto hemos de complementar con herramientas como: precisión (Accuracy), recall, F1-score y RMSE, que son totalmente aplicables y se utilizan para evaluar el rendimiento de este, buscando aproximarnos a los resultados más fiables, y para nuestro caso en particular, hemos utilizado el **Recall** aplicándolo de manera general, pero centrando el análisis en los 20 productos más representativos; los primeros 10 por su cantidad de exportaciones y los siguientes 10 por el valor económico de estas.

7.6.1. Precisión (Accuracy): Es la proporción de predicciones correctas sobre el total de predicciones. Se usa cuando las clases están balanceadas.

Accuracy=Total de Predicciones/Predicciones Correctas

7.6.2. Recall (Sensibilidad o Exhaustividad): Mide la capacidad del modelo para encontrar todos los elementos relevantes de una clase (es decir, las instancias positivas en una clasificación binaria).

Recall=Predicciones positivas correctas/Total de instancias positivas reales

Se utiliza principalmente en clasificación cuando las clases están equilibradas, es decir, no hay muchas más instancias de una clase que de otras.

7.6.3. Otra herramienta que podría ser utilizada sería el **F1-score siendo un balance entre Accuracy y Recall.**

7.6.4. Despliegue: Una vez superadas las etapas anteriores, llevaremos el modelo a un entorno real para predicción y apoyando la gestión de exportaciones donde se pueda utilizar para mejorar las decisiones productivas, comerciales y logísticas.

8. Modelado de los datos.

8.1. Características de los datos

Los datos trabajados corresponden a la información histórica de exportaciones de 6 años, que fue extraída de los Documentos de Exportación diligenciada por los usuarios

aduaneros, clasificados en un archivo por cada año, en las cuales hay variedad de información con características comunes como fecha, valores, códigos, nombres, sin valores duplicados, y el total de los datos es la sumatoria de al menos 6 bases de datos consolidadas, con valores únicos, normalizados, sin inconsistencias en los mismos y con un aproximado de 2.800.000 registros sin valores atípicos y con variables que se relacionan entre sí, para suministrarlos diferentes niveles de información a ser considerada en diferentes momentos y de la cual podemos obtener mucha utilidad.

La cantidad de variables disponibles al interior de la base de datos es de 28 con funciones únicas, pero se utilizarán 8 como las más relevantes para el análisis de la información.

9. Tipos de variables

9.1. Categóricas: Existe tres tipos de variables categóricas:

9.2. Fecha: Las que almacenan los datos que contienen fechas y que podrían dar lugar a otras variables

9.3. Ordinales: Las usadas para denotar una sucesión de orden y prelación ej. Días de la semana, Meses del año, Primero, segundo, tercero, etc.

9.4. Nominale: Las que almacenan información general ej. nombre, ciudad, país, producto, departamento, etc.

9.5. Numéricas: Corresponde a las variables que almacenan valores numéricos con los cuales se pueden realizar cálculos matemáticos.

10. Bases de datos

A partir del repositorio datos.gov.co, se obtuvo información estadística de exportaciones recopilada por el DANE desde los años 2011 a 2024. Se decidió por parte del grupo investigador, seleccionar las bases de 2019 a 2024, de forma que se incluyera estadística de las ventas de años prepandemia y años postpandemia hasta la actualidad 2024.

Estas bases de datos 2019 a 2024 cuenta con 2.866.602 registros detallados por año así:

2019: 492.610

2020: 440.319

2021: 498.945

2022: 495.318

2023: 459.332

2024: 480.078

Cada una de las bases de datos contiene 28 variables que se repiten.

Las variables incluidas en las bases de datos contienen la siguiente información y características:

| Variable | Tipo | Descripción |
|----------|---------|--|
| FECH | Fechas | Fecha de la declaración de exportación |
| ADUA | Nominal | Código de la aduana por donde se realiza la exportación |
| PAIS | Nominal | Código del país destino de la mercancía exportada |
| COD_PA14 | Nominal | Código alfabético del país destino de la mercancía exportada |
| COD_SAL1 | Nominal | Código lugar de salida de la mercancía exportada |
| COD_SAL | Nominal | Código lugar de salida de la mercancía exportada |
| DPTO2 | Nominal | Código de procedencia de la mercancía |
| VIA | Nominal | Código del modo de transporte utilizado |
| BANDERA | Nominal | Bandera del medio de transporte |
| REGIM | Nominal | Régimen aduanero al que se acoge la exportación |
| MODAD | Nominal | Modalidad de importación que la precede |
| FINALID | Nominal | Modalidad de la exportación |
| CER_ORI1 | Nominal | VER TIPOS DE CERTIFICADOS DE ORIGEN |
| SISESP | Nominal | Sistemas especiales. Indica si la exportación tiene sistemas especiales o no |
| POSAR | Ordinal | Subpartida arancelaria del producto exportado |
| PROD | Ordinal | Designación del producto exportado |

| | | |
|---------|----------|--|
| DEPTO1 | Nominal | Código de departamento de procedencia de la mercancía |
| UNID | Nominal | Código de la unidad comercial de la mercancía |
| CODUNI2 | Nominal | Código de la unidad física de la mercancía |
| CANTI | Numérica | Cantidad exportada del producto en la unidad de medida correspondiente |
| PBK | Numérica | Peso bruto en Kg |
| PNK | Numérica | Peso neto en Kg |
| FOBDOL | Numérica | Valor FOB (free on board) total en dólares Estadounidenses |
| FOBPES | Numérica | Valor FOB (free on board) total en pesos Colombianos |
| AGRENA | Numérica | Valor agregado nacional |
| FLETES | Numérica | Costo de los fletes asociados al transporte de la mercancía |
| SEGURO | Numérica | Costo del seguro de la mercancía durante el transporte |
| OTROSG | Numérica | Otros gastos asociados a la exportación |
| CIFDOL | Numérica | Valor CIF (Cost. insurance freight) total en dólares estadounidenses |
| CIFPES | Numérica | Valor CIF (Cost. insurance freight) total en pesos Colombianos |

Todos los valores están presentes dentro de las bases de datos y con sus respectivos contenidos, es decir, no hay valores faltantes, lo que garantiza una continuidad y unos resultados basados en el 100% de los mismos valores dentro de ellas y así, cada cálculo será elaborado sobre valores existentes, no sobre promedios o porcentajes o información sintética.

El proyecto Proyección de exportaciones basado en el desarrollo de un modelo predictivo de exportaciones está soportado en los siguientes elementos:

10.1. Análisis de las variables

Del total de 28 variables contenidas en las bases de datos las más representativas utilizadas para el análisis de datos son:

10.1.1. Código del producto: Representa el código del producto y está determinado por la posición arancelaria que es un código único que individualiza el producto que es objeto de exportación

10.1.2. Cantidad de exportaciones realizadas: Representa la cantidad o número de veces que se realizó exportaciones del producto en el rango de tiempo evaluado y por cada uno de los países a los que se ha exportado.

10.1.3. Valor fob en dólares: Representa la cantidad de dólares cobrados a nuestro cliente por la exportación y por cada una de las operaciones. Información declarada ante las autoridades aduaneras que realizan el control aduanero y cambiario.

10.1.4. País destino de la exportación: Representa el código y nombre del país destino de la mercancía exportada.

10.1.5. Cantidad de exportación: Corresponde a la cantidad de exportación en unidades de medida (Metros cúbicos, kilos, unidades, etc), realizadas.

10.1.6. Departamento de origen: Corresponde al Departamento de Colombia en donde se produjo o generó la carga a exportar.

11. Modelo Utilizado

Dado que la información (Datos) utilizados contienen etiquetas con las que se puede hacer agrupaciones, mediciones y operaciones matemáticas, requiere la utilización de un Modelo Supervisado; en nuestro caso se utilizó el modelo Light GBM.

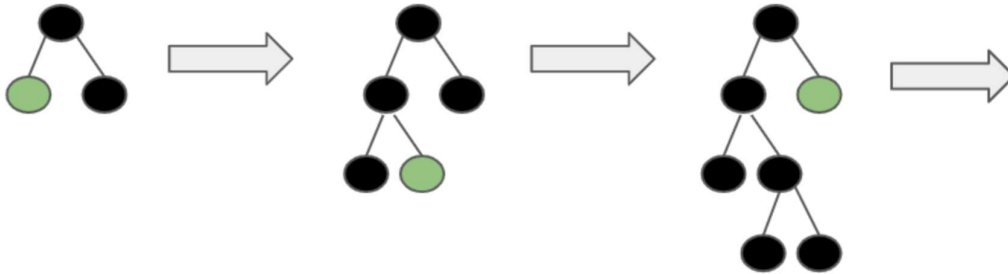
El LIGHT GBM o GBM ligero es una estructura rápida, apropiada e impulso de gradiente, que depende del cálculo del árbol de elección, y que se utiliza para el posicionamiento, la caracterización y otras numerosas asignaciones de la IA.

Dado que depende de los cálculos del árbol, divide la hoja del árbol con el mejor ajuste, mientras que otros cálculos de impulso dividen la profundidad del árbol de manera astuta o nivelada en oposición a la hoja. Así que cuando se desarrolla en una hoja similar en Light GBM, el cálculo por hojas puede generar más desaciertos que el cálculo por niveles y, por lo tanto, produce una precisión mucho mayor que puede ser lograda de vez en cuando por cualquiera de los cálculos de impulso actuales. De la misma manera, es sorprendentemente rápido, de ahí la palabra 'Ligero'.

Utiliza algoritmos de árboles de decisión para la clasificación, el ranking y otras tareas de aprendizaje automático. LGBMClassifier utiliza una técnica novedosa de muestreo unilateral

basado en gradientes (GOSS) y agrupación de características exclusivas (EFB) para manejar datos a gran escala con precisión, lo que los hace más rápidos y reduce el uso de memoria.

LightGBM leaf-wise



Resultados del Modelo

Una vez aplicado el modelo a los datos disponibles se logra los siguientes análisis.

11.1. Análisis gráfico

En el análisis preliminar del año 2024 se pudo analizar unos datos relevantes que se muestran a continuación:

- El top 10 de los productos con mayor valor de exportaciones en dólares y que representan el 69.2% de las exportaciones son:

| Descrip | FOB DOL | Partic. % | Acum % % |
|-----------------|-----------|-----------|----------|
| Petroleo | 12.062,00 | 24,36% | 24,36% |
| Hullas | 6.013,00 | 12,14% | 36,50% |
| Oro | 4.104,00 | 8,29% | 44,79% |
| Café | 3.545,00 | 7,16% | 51,95% |
| Aceites petrol | 2.733,00 | 5,52% | 57,46% |
| Flores | 2.348,00 | 4,74% | 62,21% |
| Bananas | 1.225,00 | 2,47% | 64,68% |
| Coques | 1.094,00 | 2,21% | 66,89% |
| Construcciones | 605,00 | 1,22% | 68,11% |
| Ferroaleaciones | 542,00 | 1,09% | 69,20% |

- Los 20 productos con mayor valor de exportaciones no minerales (sin petróleo o sus derivados), que representan el 56.81%, son:

| Descrip | FOBDO | Partic. % | Acum % |
|--------------------------------|----------|-----------|--------|
| Café | 3.545,00 | 15,08% | 15,08% |
| Flores | 2.348,00 | 9,99% | 25,06% |
| Bananas | 1.225,00 | 5,21% | 30,27% |
| Construcciones | 605,00 | 2,57% | 32,84% |
| Ferroaleaciones | 542,00 | 2,30% | 35,15% |
| Papel, textiles fotográficos | 507,00 | 2,16% | 37,30% |
| Plantas. | 442,00 | 1,88% | 39,18% |
| Margarina. | 417,00 | 1,77% | 40,96% |
| Máquinas de clasificar. | 395,00 | 1,68% | 42,64% |
| Manufacturas de fundición | 387,00 | 1,65% | 44,28% |
| Hormonas. | 342,00 | 1,45% | 45,74% |
| Extractos de café | 339,00 | 1,44% | 47,18% |
| Resinas aminicas | 328,00 | 1,39% | 48,57% |
| Dátiles, higos, aguacates | 327,00 | 1,39% | 49,96% |
| Aceites crudos de petróleo | 318,00 | 1,35% | 51,32% |
| Polímeros de propileno | 290,00 | 1,23% | 52,55% |
| Artículos de confitería sin ca | 283,00 | 1,20% | 53,75% |
| Preparaciones de belleza | 281,00 | 1,19% | 54,95% |
| Tapas plásticas | 222,00 | 0,94% | 55,89% |
| Aceites crudos de petróleo o | 215,00 | 0,91% | 56,81% |

Basados en la información analizada, por los años 2019 a 2024 se observaron 2 variables de mucha relevancia a saber:

- Productos exportados (PROD), nos da el producto y el número de veces que éste se exportó.
- Valor de las exportaciones en dólares (FOBDO).

Con base en estas 2 variables se determinaron los 20 productos más representativos en cuanto a valor de exportaciones.

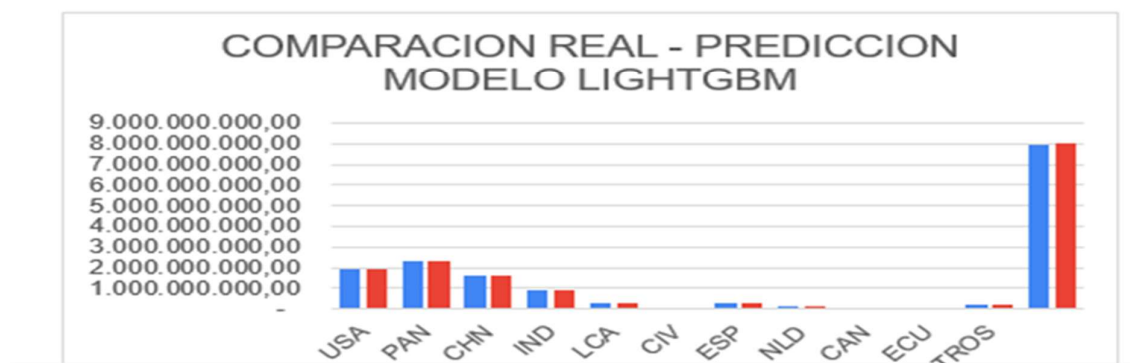
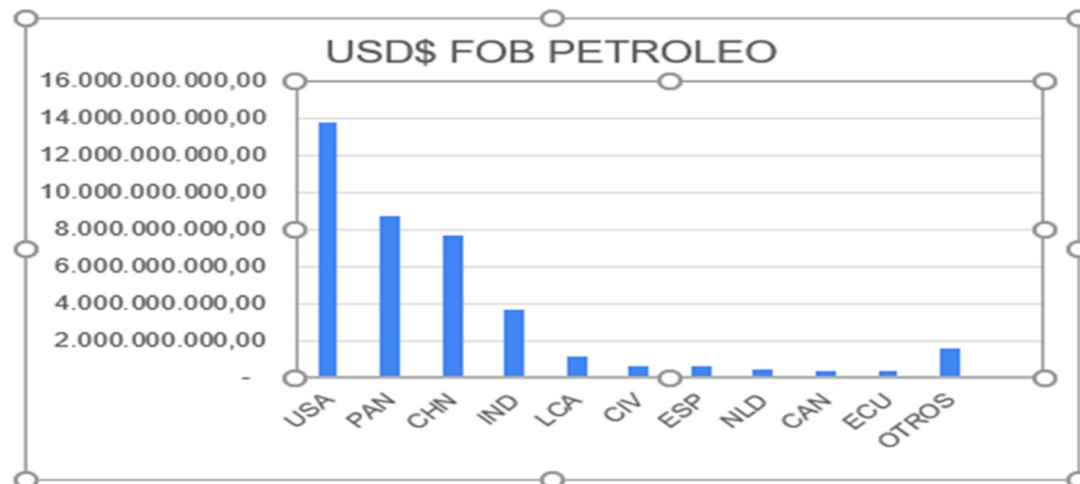
Para las predicciones realizadas del valor de las exportaciones se entregó al modelo el 80% de la base de datos para entrenamiento y el 20% para pruebas y validación.

Modelo de predicción Producto: Petróleo

Se muestra a continuación el modelo sobre el análisis del producto Petróleo y el comparativo del valor real Vs el valor predicho.

Resultado del modelo: La gráfica del resultado del modelo muestra una precisión del R^2 del 99.61%, lo cual indica que el modelo de predicción es altamente confiable, veamos:

Exportaciones de Petróleo



| COD_PA4 | USD\$ FOB | Real USD\$ FOB | Proyeccion USD\$ FOB |
|---------|-------------------|------------------|----------------------|
| USA | 13.776.269.732,08 | 1.962.044.051,09 | 1.959.670.734,19 |
| PAN | 8.702.514.620,45 | 2.327.957.022,02 | 2.337.543.480,27 |
| CHN | 7.640.475.837,59 | 1.629.790.128,76 | 1.655.866.797,09 |
| IND | 3.703.153.341,53 | 954.583.718,47 | 941.783.911,77 |
| LCA | 1.112.322.774,57 | 307.516.010,79 | 307.953.456,11 |
| CIV | 603.392.524,41 | 38.395.268,00 | 42.457.015,40 |
| ESP | 580.387.100,42 | 263.452.633,20 | 269.937.898,21 |
| NLD | 390.115.849,27 | 117.391.984,00 | 113.689.005,23 |
| CAN | 379.064.999,15 | 42.971.974,68 | 43.513.308,30 |
| ECU | 370.542.475,38 | 56.825.692,79 | 54.874.097,29 |
| OTROS | 1.536.884.254,92 | 244.896.550,85 | 255.266.832,72 |
| | 38.795.123.509,77 | 7.945.825.034,65 | 7.982.556.536,58 |

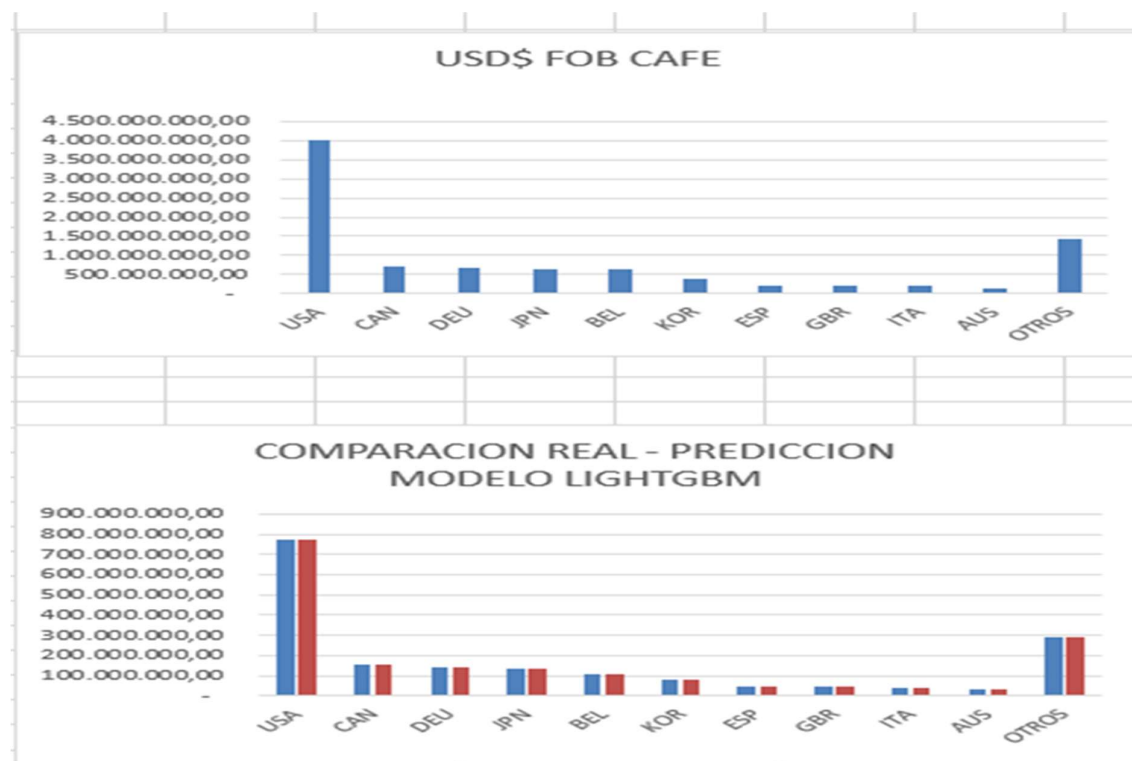
Modelo de predicción Producto: Cafe

Se muestra a continuación el modelo sobre el análisis del producto Café y el comparativo del valor real Vs el valor predicho.

Para la predicción del valor de las exportaciones igual se entregó al modelo el 80% de la información de la base de datos para entrenamiento y el 20% para pruebas y validación.

Resultado del modelo: La gráfica del resultado muestra una precisión del R^2 del 99.52%, lo cual indica que el modelo de predicción es altamente confiable, veamos:

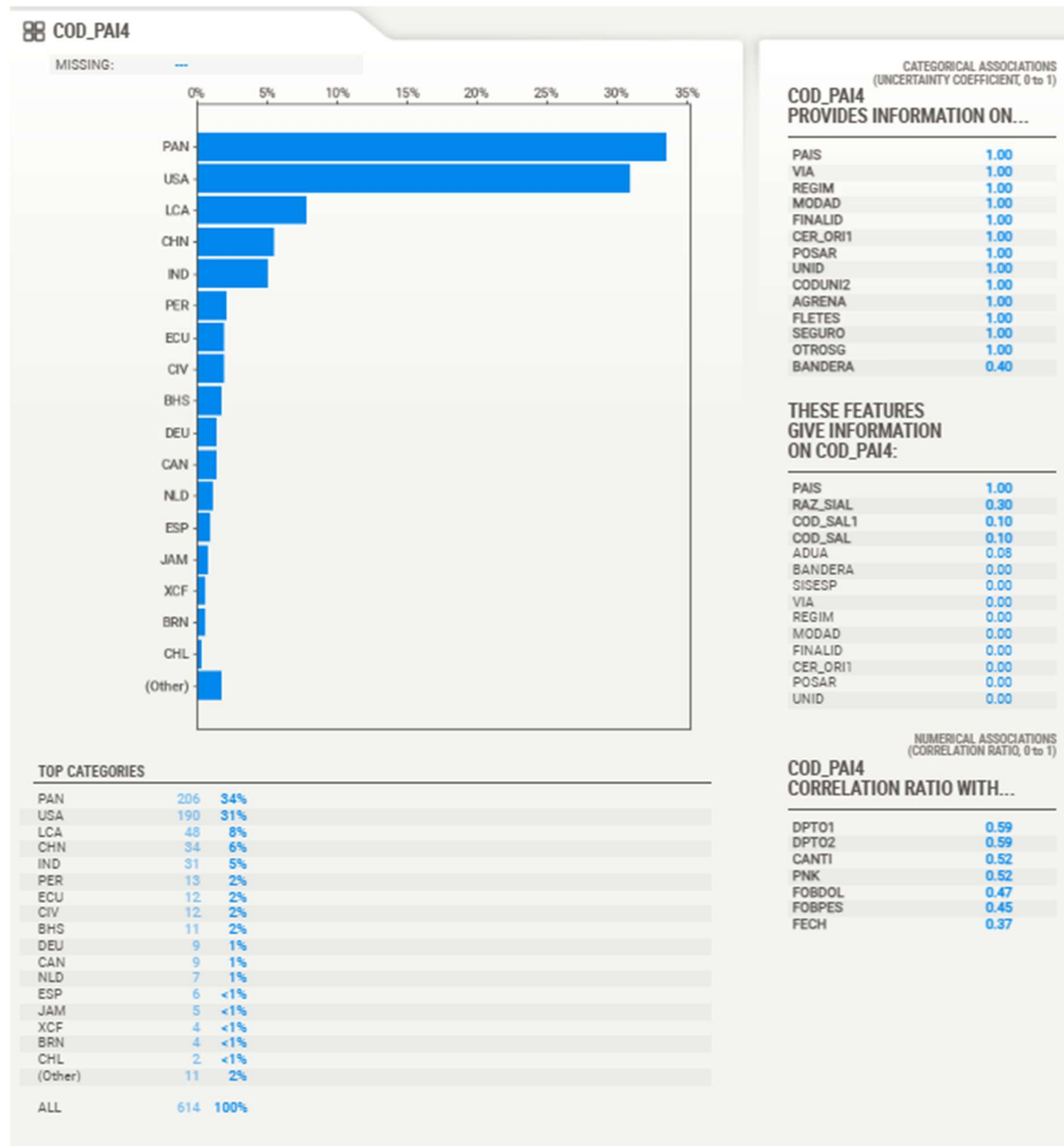
| COD_PAIS | USD\$ FOB | Real USD\$ FOB | Proyeccion USD\$ FOB |
|----------|------------------|------------------|----------------------|
| USA | 4.010.908.419,42 | 771.620.887,17 | 770.952.754,88 |
| CAN | 707.308.143,10 | 152.698.708,74 | 153.054.103,28 |
| DEU | 692.903.586,78 | 140.753.251,50 | 140.801.939,72 |
| JPN | 656.435.043,50 | 136.877.837,50 | 136.996.446,51 |
| BEL | 648.543.759,71 | 108.730.183,10 | 109.177.370,38 |
| KOR | 385.054.113,15 | 78.061.033,94 | 78.207.869,60 |
| ESP | 214.066.196,48 | 46.556.794,61 | 46.192.979,72 |
| GBR | 211.204.207,78 | 45.285.563,47 | 45.240.246,28 |
| ITA | 194.093.426,66 | 43.382.867,41 | 43.489.050,49 |
| AUS | 155.890.237,06 | 30.406.426,71 | 30.422.199,32 |
| OTROS | 1.436.982.656,09 | 294.503.588,38 | 294.590.139,56 |
| | 9.313.389.789,73 | 1.848.877.142,53 | 1.849.125.099,74 |



11.1 Análisis Estadístico de Datos (EDA)

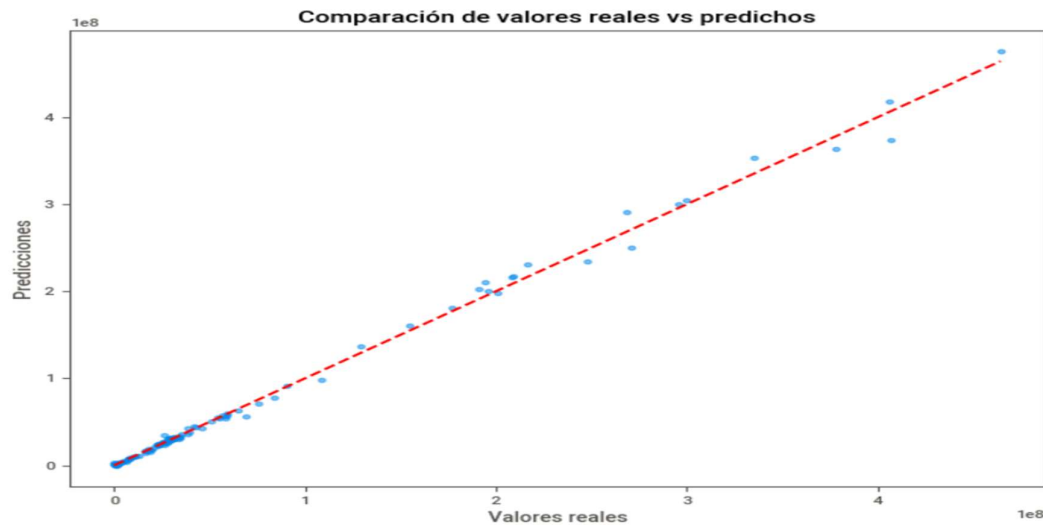
De análisis de las variables más relevantes (Valor FOB exportaciones y Cantidad de Exportaciones), se analizó las más sobresalientes de 18 productos modelados.

El análisis estadístico de datos para el producto **2709 petróleo** es el siguiente:



Los países a los que se ha efectuado mayor número de exportaciones de petróleo en el tiempo analizado son Panamá, Estados Unidos, Santa Lucía, China e India.

También se visualiza que el mayor exportador de petróleo en Colombia es Ecopetrol, con el 30% de las operaciones realizadas.



El modelo aplicado nos arrojó una precisión del 99,61% que se observa en la imagen anterior.

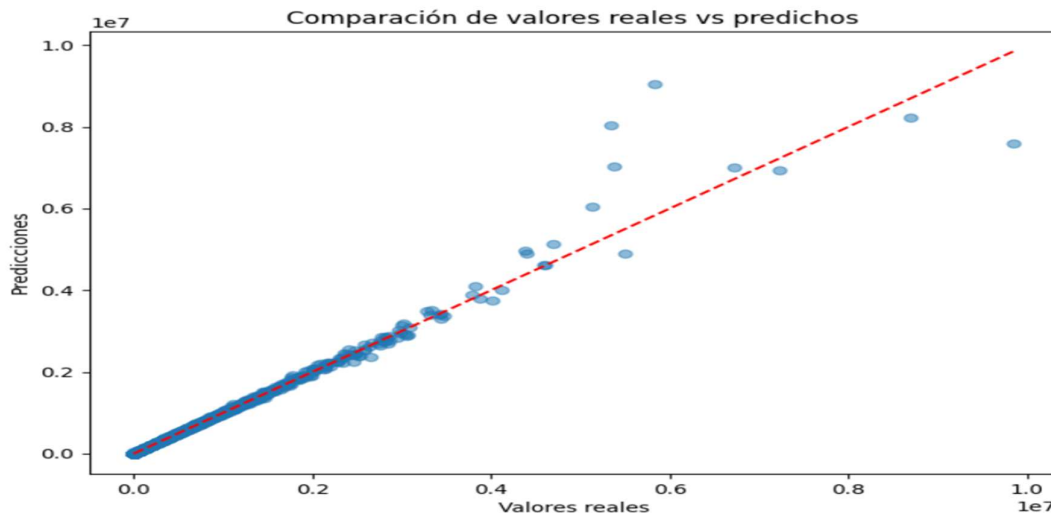
Para el caso del producto 0901 Café, podemos observar:

| COD_PA14 | | | |
|-----------|---------------|--------|---------|
| VALUES: | 36,084 (100%) | 10,895 | 30% |
| MISSING: | --- | 2,583 | 7% |
| | | 2,345 | 6% |
| DISTINCT: | 102 (<1%) | 1,981 | 5% |
| | | 1,675 | 5% |
| | | 1,577 | 4% |
| | | 1,403 | 4% |
| | | 13,625 | 38% |
| | | | (Other) |
| | | | USA |
| | | | CAN |
| | | | DEU |
| | | | JPN |
| | | | KOR |
| | | | BEL |
| | | | GBR |

Colombia efectuó exportaciones de café a 102 países en 36.084 operaciones de exportación, procedente de 24 departamentos.

El principal comprador del café Colombiano es Estados Unidos, con el 30% de las exportaciones realizadas y el principal exportador es la Federación Nacional de Cafeteros de Colombia con el 11% de las operaciones.

El modelo aplicado al café nos arrojó una precisión del 99,52% que se observa en la siguiente imagen.



Para el producto **0603 Flores:**

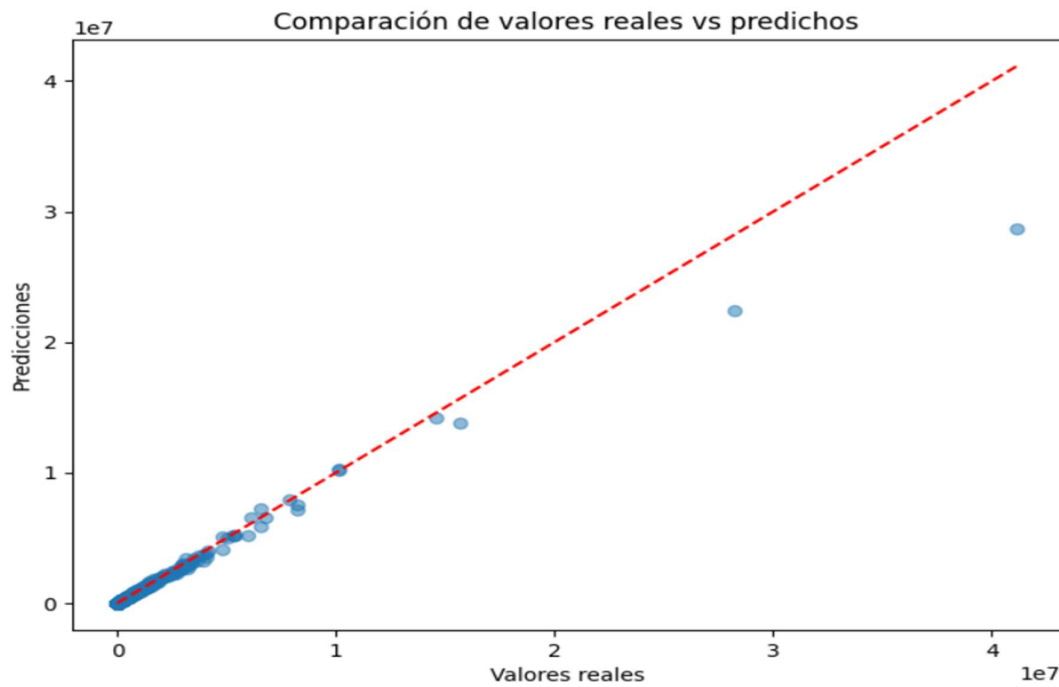
Se realizaron 205.012 operaciones de exportación a 113 países.

Destinos más representativos:

| COD_PAIA | | | |
|-----------|----------------|--------|---------|
| VALUES: | 205,012 (100%) | 75,246 | 37% |
| MISSING: | --- | 20,176 | 10% |
| | | 12,047 | 6% |
| DISTINCT: | 113 (<1%) | 11,048 | 5% |
| | | 9,388 | 5% |
| | | 8,556 | 4% |
| | | 7,509 | 4% |
| | | 61,042 | 30% |
| | | | (Other) |

Los departamentos más exportadores de flores son Cundinamarca, Bogotá y Antioquia.

| MOST FREQUENT VALUES | | |
|----------------------|--------|-------|
| 25 | 68,886 | 33.6% |
| 11 | 67,474 | 32.9% |
| 5 | 65,875 | 32.1% |
| 8 | 894 | 0.4% |
| 47 | 874 | 0.4% |
| 15 | 412 | 0.2% |
| 13 | 348 | 0.2% |
| 76 | 86 | <0.1% |
| 17 | 77 | <0.1% |
| 66 | 66 | <0.1% |
| 63 | 17 | <0.1% |
| 23 | 1 | <0.1% |
| 81 | 1 | <0.1% |
| 94 | 1 | <0.1% |



El modelo aplicado nos arrojó una precisión del 99,66% que se observa en la imagen anterior.

12. Anexos

Al presente informe se anexa:

- 12.1. Cuadernos de código
- 12.2. Análisis EDA de 18 productos
- 12.3. Tablas de los principales productos de exportación, con el resultado de la predicción y sus respectivas métricas.
- 12.4. 18 Gráficos comparativos de real Vs Predicción.
- 12.5. 18 Gráficos con las variables más relevantes por producto.

13. Bibliografía

Bibliografía (Dane, Dirección de impuestos y aduanas nacionales, Repositorios ministerio de Comercio exterior, Hacienda, BanRepública (TRM), SUIN-Juriscol Minjusticia.

Departamento Administrativo Nacional de Estadística (DANE). (2016, mayo). *Cobertura de las estadísticas de exportaciones.*

Departamento Administrativo Nacional de Estadística (DANE). (2016, septiembre). *Ficha metodológica de exportaciones - EXPO.*

Departamento Administrativo Nacional de Estadística (DANE). (2017, abril). *Glosario de exportaciones.*

Naciones Unidas. (2010). *Estadísticas del comercio internacional de mercancías: Conceptos y definiciones.* Naciones Unidas.

Comunidad Andina de Naciones. (2001). *Elaboración de las estadísticas del comercio exterior de bienes de la Comunidad Andina y de sus Países Miembros: Decisión 511.* Comunidad Andina de Naciones.

Dirección de Impuestos y Aduanas Nacionales – DIAN. (2016). *Declaración de Importación, Andina del Valor en Aduanas y Exportación.* Dirección de Impuestos y Aduanas Nacionales.

Decreto 1881 de 2021: Arancel de Aduanas. Ministerio de Hacienda, Industria y Turismo. (2021).

Gobierno de Colombia. *Página www.datos.gov.co.* Recuperado de <https://www.datos.gov.co>

Gobierno de Colombia. *Página www.dian.gov.co.* DIAN. *Aplicativo - Prevalidador Declaración de Exportación.*

Decreto 1165 de 2019. Ministerio de Hacienda y Crédito Público. (2019). *Sistema Único de Información Normativa - SUIN.*

Decreto Ley 2685 de 1999: Legislación Aduanera de Colombia. Presidencia de la República. (1999).

Resolución 4240 de 2000: Sistematización de los procedimientos aduaneros. Dirección de Impuestos y Aduanas Nacionales. DIAN. (2000).