

# Modelo predictivo de transporte marítimo portuario de carga movilizada en cabotaje y transbordo\*

Predictive model of maritime port transport of cargo mobilized in cabotage and transshipment

Andrés Felipe Yanes Quintero\*\*

Anly del Rosario Lafont Badel\*\*\*

Jairo Jamith Palacios Rozo\*\*\*\*

Fecha de recepción: 18 de julio de 2025

Fecha de aprobación: 23 de octubre de 2025

Fecha de publicación: 30 de diciembre de 2025

DOI:<https://doi.org/10.18041/1900-0642/criteriolibre.2025v23n43.13296>

## Resumen

Aunque existen algunos avances en el transporte marítimo y en el movimiento portuario, resulta necesario conocer el sector transporte vinculado al movimiento de carga a través de los diferentes puertos autorizados en Colombia, dado que este constituye un factor estratégico para el comercio internacional. A pesar de que el seguimiento se realiza mediante los reportes de las sociedades portuarias de servicio público y privado por medio del sistema de información de Supervisión al Transporte (VIGIA), persisten vacíos importantes tanto en el conocimiento del transporte marítimo de cabotaje, como en el registro de las toneladas de carga movilizadas según zona portuaria, tráfico portuario, tipo de tráfico, tipo de carga y sociedad portuaria. Estos aspectos son fundamentales para la toma de decisiones estratégicas. El propósito de este estudio fue analizar el tipo de carga portuaria movilizada en operaciones de cabotaje y transbordo en los puertos de Colombia. Para ello, se aplicó la metodología

**Citar como:** Yanes, A. F., Lafont, A. R. y Palacios, J. J. (2025). Modelo predictivo de transporte marítimo portuario de carga movilizada en cabotaje y transbordo, 24 (43), 1-13. <https://doi.org/10.18041/1900-0642/criteriolibre.2025v23n43.13296>

Esta obra está bajo una licencia internacional [Creative Commons Atribución-NoComercial-SinDerivadas 4.0](#)



\*Proyecto de grado Maestría en Administración de Negocios Internacionales. Transporte marítimo portuario en Colombia. Universidad Autónoma de Colombia.

\*\*Estudiante de la Universidad Autónoma de Colombia. Bogotá, Colombia. ORCID: <https://orcid.org/0000-0002-8699-5066>  
Correo electrónico: ayanes@universidadmayor.edu.co

\*\*\*Estudiante de la Universidad Autónoma de Colombia. Bogotá, Colombia. ORCID: <https://orcid.org/0009-0002-7976-4769>  
Correo electrónico: anly.lafont@fuac.edu.co

\*\*\*\*Docente de la Universidad Colegio Mayor de Cundinamarca, Doctorando en Administración y Desarrollo de la Universidad de Celaya México. ORCID: <https://orcid.org/0000-0002-1437-9838> CvLAC: [https://scienti.minciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod\\_rh=0000548197](https://scienti.minciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0000548197) Correo electrónico: jjpalacios@universidadmayor.edu.co

CART (Classification and Regression Trees) un algoritmo de aprendizaje supervisado automático. La información secundaria utilizada proviene del conjunto de datos aportado por las sociedades portuarias de servicio público y privado, con un total de 7.139 registros correspondientes a zonas portuarias públicas y privadas del país entre los años 2018 y marzo de 2024. Los resultados presentan un modelo predictivo basado en machine learning para la clasificación de tipos de carga, utilizando un árbol de decisión con una variable objetivo (target) categórica que distingue entre carga de carbón a granel, contenedores, carga general, granel líquido y granel sólido distinto del carbón. Las variables predictorias (features) consideradas fueron cabotaje y transbordo. Se concluye que el modelo de árbol de decisión, en su configuración actual, presenta un rendimiento limitado en la clasificación del tipo de carga utilizando únicamente dichas características.

A diferencia de otros estudios realizados en Colombia o en la región, los cuales se centran principalmente en la optimización logística, la gestión de terminales o el análisis de la eficiencia portuaria mediante indicadores cuantitativos, esta investigación incorpora un enfoque basado en inteligencia artificial y aprendizaje automático. El uso de la metodología CART constituye una innovación metodológica, al permitir la identificación de patrones de clasificación de carga portuaria a partir de variables categóricas y supervisadas, brindando una perspectiva más predictiva que descriptiva sobre el comportamiento del tráfico marítimo nacional.

**Palabras Clave:** transporte marítimo portuario, cabotaje, transbordo, modelo predictivo.

## Abstract

Although there have been some advances in maritime transport and port operations, it remains necessary to understand the transport sector related to cargo movement through the various authorized ports in Colombia, as this constitutes a strategic factor for international trade. Despite the fact that monitoring is carried out through reports from public and private port companies via the Transport Supervision Information System (VIGIA), significant gaps persist both in the understanding of coastal maritime transport (cabotage) and in the recording of cargo tonnage moved by port zone, port traffic, type of traffic, type of cargo, and port company. These aspects are essential for strategic decision making. The purpose of this study was to analyze the types of port cargo moved in cabotage and transshipment operations Colombian ports. To this end, the CART (Classification and Regression Trees) methodology—an automatic supervised learning algorithm—was applied. The secondary data used came from datasets provided by public and private port companies, totaling 7,139 records corresponding to public and private port zones in the country between 2018 and March 2024. The results present a predictive model based on machine learning for cargo type classification, using a decision tree with a categorical

target variable that distinguishes between bulk coal cargo, containers, general cargo, liquid bulk, and solid bulk other than coal. The predictive variables (features) considered were cabotage and transshipment. It is concluded that the decision tree model, in its current configuration, shows limited performance in classifying cargo type when using only these features.

Unlike other studies conducted in Colombia or the region—which focus mainly on logistics optimization, terminal management, or port efficiency analysis through quantitative indicators—this research incorporates an artificial intelligence and machine learning approach. The use of the CART methodology represents a methodological innovation, as it enables the identification of port cargo classification patterns based on categorical and supervised variables, offering a more predictive rather than descriptive perspective on the behavior of national maritime traffic.

**Keywords:** maritime transport, port cargo, cabotage, transshipment, predictive model.

## 1. Introducción

Para mantener su relevancia en un entorno cada vez más competitivo, los puertos deben desarrollar estrategias efectivas que optimicen sus operaciones y mejoren su eficiencia. Este escrito explora diversas estrategias implementadas en los puertos marítimos, su impacto en la eficiencia y la competitividad, así como su aplicación práctica en el contexto actual. La innovación portuaria abarca una amplia gama de áreas, desde la infraestructura física hasta la gestión operativa y la adopción de tecnologías emergentes.

En cuanto a la infraestructura, los puertos están incorporando diseños más eficientes y sostenibles, que integran tecnologías verdes y resistentes al cambio climático (Smith, 2020). Además, la digitalización está transformando la gestión portuaria, permitiendo la optimización en tiempo real de las operaciones y la reducción de los tiempos

de espera (Jones & Wang, 2019). Entre las tendencias actuales en innovación portuaria, destacan la digitalización, la sostenibilidad y la mejora de la conectividad intermodal.

La digitalización, impulsada por tecnologías como el Internet de las cosas (IoT) y la inteligencia artificial (IA), está permitiendo la creación de puertos inteligentes, capaces de monitorear y gestionar eficientemente el flujo de mercancías (García & López, 2021; Gesé Bordils et. al, 2021). En este sentido, resulta esencial que los países de América Latina y el Caribe (ALC) desarrollen puertos modernos y eficientes para mejorar sus relaciones comerciales y contribuir al desarrollo de la región, igualmente es importante que se planteen reformas y mejoras en el sector portuario que generen oportunidades en el transporte marítimo. El crecimiento del tráfico de contenedores ha dado lugar

al surgimiento de nuevas rutas y a la construcción de buques y puertos de mayor escala (Fancello et al., 2022).

En términos de sostenibilidad y eficiencia, los puertos están adoptando prácticas ecoamigables y tecnologías de reducción de emisiones para mitigar su impacto ambiental (Chen et al., 2022). La implementación de estrategias sostenibles y tecnologías ecoamigables contribuye a la reducción de la huella de carbono de los puertos y promueve un desarrollo más sostenible a largo plazo (Wu et al., 2021). Por otro lado, la mejora de la conectividad intermodal está optimizando la eficiencia logística y reduciendo la congestión en las áreas portuarias (Tan et al., 2023). La innovación en los puertos marítimos tiene un impacto significativo en la eficiencia y la sostenibilidad de las operaciones portuarias, por lo que la digitalización y la mejora de la conectividad intermodal permiten una gestión más ágil y eficiente del flujo de mercancías, reduciendo los tiempos de espera y los costos operativos (Zhang, 2020).

Las estrategias en puertos marítimos abarcan una amplia gama de áreas, desde la infraestructura y la gestión operativa hasta la implementación de tecnologías avanzadas y la promoción de la sostenibilidad. Estas estrategias están diseñadas para abordar desafíos específicos y aprovechar oportunidades emergentes en el mercado global.

Una estrategia fundamental para mejorar la eficiencia en los puertos marítimos es la optimización de la infraestructura portuaria. Esto implica la modernización y expansión de instalaciones clave, como muelles, terminales de contenedores y

sistemas de almacenamiento, para aumentar la capacidad de manipulación de carga y reducir los tiempos de espera (Chen & Zhang, 2019). De igual manera, la mejorar de la conectividad terrestre, —mediante la construcción de ferrocarriles y carreteras de acceso eficientes, facilita el movimiento rápido y fluido de mercancías dentro y fuera del puerto (Jones et al., 2020).

Otra estrategia importante es la implementación de tecnologías avanzadas para optimizar las operaciones portuarias. Esto incluye el uso de sistemas de gestión de terminal (TOS) basados en la nube, que permiten una planificación y coordinación más eficientes de las actividades portuarias (García & López, 2022). Las funciones de apoyo a la carga —como la manipulación, el almacenamiento, la inspección, la consolidación y la gestión de la información— constituyen servicios esenciales para el adecuado funcionamiento de un puerto (Tapía et al., 2014).

Además, la automatización de procesos, como la carga y descarga de contenedores, mediante el uso de robots y sistemas de transporte autónomos, puede aumentar la productividad y reducir los costos operativos (Smith & Wang, 2018). La sostenibilidad es un aspecto cada vez más importante en la gestión portuaria, y las estrategias para promover prácticas sostenibles pueden mejorar la competitividad a largo plazo de un puerto. Incluye la adopción de tecnologías y prácticas ecoamigables, como la electrificación de equipos de manejo de carga y la implementación de energías renovables, que reducen la huella de carbono del puerto y cumplen con regulaciones ambientales cada vez más

estrictas (Tan et al., 2021), donde la gestión eficiente de residuos y la conservación de recursos naturales son componentes clave de una estrategia de sostenibilidad integral en puertos marítimos (Li & Wu, 2023).

La promoción de la sostenibilidad no solo reduce los costos a largo plazo asociados con el cumplimiento de regulaciones ambientales, sino que también mejora la reputación del puerto y su atractivo para clientes y socios comerciales (Tan et al., 2021). En conclusión, la aplicación efectiva de estas estrategias tiene un impacto directo en la eficiencia y la competitividad de un puerto marítimo.

## 2. Materiales y Métodos

Esta investigación se basa en información secundaria proveniente de datos abiertos aportados por las sociedades portuarias de servicio público y privado de Colombia. El conjunto de datos (dataset) está compuesto por 7.139 registros correspondientes a las diferentes zonas portuarias públicas y privadas del país, abarcando el periodo comprendido entre 2018 y marzo de 2024.

Para el análisis, se aplicó la metodología CART (Classification and Regression Trees), también conocida como árbol de clasificación, un algoritmo supervisado de aprendizaje automático que permite realizar un análisis riguroso de los patrones y factores que influyen en la dinámica portuaria. Esta metodología facilita la identificación de relaciones entre variables y la generación de modelos predictivos, con el propósito de ofrecer información que contribuya a la toma de decisiones estratégicas orientadas a mejorar la competitividad del sector.

La recopilación de datos cubre un periodo de siete años y tres meses, lo cual posibilita evaluar las tendencias y variaciones a lo largo del tiempo. De acuerdo con Sarduy (2007), la naturaleza de la información recolectada para dar respuesta al problema de investigación se enmarca dentro del paradigma cuantitativo, dado que implica la medición y el análisis de variables numéricas y categóricas.

El modelo se diseñó para comparar las respuestas categóricas de la variable objetivo —el tipo de carga, clasificada en carbón a granel, contenedores, carga general, granel líquido y granel sólido distinto del carbón— utilizando como características predictoras (features) las variables de cabotaje y transbordo. Asimismo, se emplearon métricas de evaluación para determinar el desempeño del modelo, incluyendo la precisión, la tasa de verdaderos positivos, la tasa de falsos positivos y el valor predictivo negativo, con el fin de cuantificar la capacidad del árbol de decisión para clasificar correctamente los diferentes tipos de carga portuaria.

## 3. Resultados

En este apartado se presentan los resultados obtenidos a partir del análisis realizado mediante técnicas de machine learning, específicamente con el algoritmo de clasificación CART (árbol de decisión), donde se mostró que la salida del target es la variable tipo de carga la cual es de forma nominal y las entradas es decir features las variables cabotaje y transbordo.

**Tabla 1.** Clasificación del árbol de decisiones

Splits	n(Train)	n(Test)	Test Accuracy
11	5712	1427	0.162

Posteriormente, se interpreta el fenómeno identificado a partir del análisis de Machine Learning con el modelo de clasificación basado en un árbol de decisión, el cual permitió identificar diversas características estructurales del modelo. Entre los elementos más relevantes se encuentran los siguientes:

- Splits: el modelo generó once divisiones (11) o puntos de partición, para clasificar los datos en función de las características cabotaje y transbordo. Estas divisiones determinan las reglas de decisión que guían la asignación de cada registro a una categoría específica del tipo de carga.
- n(Train): corresponde al número de datos de entrenamiento utilizados para construir el modelo, con un total de 5.712, lo que representa aproximadamente el 80 % del conjunto total de datos.
- n(Test): hace referencia al número de datos empleados para la fase de prueba del modelo, con un total de 1.427, equivalentes al 20 % del conjunto total.
- Test Accuracy: el modelo alcanzó una precisión de 0.162, es decir 16.2%, lo que indica un rendimiento bajo en la clasificación del

tipo de carga a partir de las variables cabotaje y transbordo. Este resultado sugiere que las características empleadas no aportan suficiente información discriminante para predecir de manera efectiva las categorías de carga portuaria.

La división de datos (data split) permite entrenar el modelo en un subconjunto de datos y después es evaluado en un conjunto de datos independiente el cual proporciona la estimación de su rendimiento en los datos no vistos. Para el siguiente caso aplicado se vislumbra un total de 7.139 observaciones divididas en 5.712 datos para entrenamiento (Train) y 1427 datos para prueba (Test).

Esta separación se emplea para validar el respectivo rendimiento del modelo. Posteriormente presenta un análisis del cómo interpretar, evaluar y distribuir los datos, se divide los datos de entrenamiento con los datos para prueba, es decir el train sobre el test (5.712/1.427) el cual representa aproximadamente un (80%/20%) en términos de proporción entre el conjunto de entrenamiento y el conjunto de prueba. El conjunto de datos permite que el modelo aprenda de las características y relaciones presentes en los datos, siendo importante para los modelos en este caso los árboles de decisión para capturar y determinar los patrones complejos.

**Figura 1.** Data Split

**Tabla 2.** Matriz de confusión

		Previsto	
		contenedores	general
Observación General	carbón al granel	0	0
	Contenedores	0.06	0.05
	granel liquido	0	0.08
	granel sólido diferente de carbón	0	0.03
	granel sólido diferente de carbón	0	0.18

En la interpretación de las predicciones contiene las predicciones correctas de como el modelo está prediciendo con éxito la clase granel líquido en una probabilidad de (28%) cuando el observado es también granel líquido. La clase general tiene una predicción correcta en la columna correspondiente con una probabilidad de (8%), los errores de clasificación cuando la clase observada es Carbón a granel, el modelo predice Granel líquido con una probabilidad de (12%), este es un error importante ya que la clase observada es muy diferente a la predicha.

La clase contenedores se clasifica erróneamente en granel líquido y general, con probabilidades pequeñas de (6%) y (5%) lo que indica que el modelo tiene dificultades para distinguir esta clase de otras. Para las observaciones de granel sólido diferente de carbón, se predice incorrectamente como granel líquido (18%) y como general (1%).

Este es otro caso importante de confusión entre clases, lo que refleja que el modelo tiene problemas para diferenciar estas clases.

A partir de la matriz de confusión, se calcularon las siguientes métricas importantes para evaluar el rendimiento del modelo, precisión (accuracy), donde la precisión es el porcentaje total de predicciones correctas. Para su cálculo se usa la siguiente fórmula.

$$\text{Precision} = \frac{\text{Predicciones Correctas}}{\text{Total de Observaciones}}$$

El cálculo de precisión y recall para la clase carbón al granel, observa la no predicción correcta en esa clase (Predicted: Carbón al granel = 0). Esto afecta negativamente tanto la precisión como el recall para esta clase. La matriz de confusión proporciona un desglose de las predicciones correctas e incorrectas para cada clase, útil para evaluar cómo se comporta el modelo con cada clase individual.

**Tabla 3.** Class Proportions

	Data Set	Training Set	Test Set
Carbón al granel	0.122	0.122	0.123
Contenedores	0.166	0.165	0.171
General	0.217	0.220	0.205
Granel liquido	0.302	0.301	0.306
Granel sólido diferente de carbón	0.193	0.192	0.196

El análisis de la Tabla 1 – Proporciones de clases (*Class Proportions*) presenta la distribución de las categorías en tres subconjuntos de datos: conjunto completo de datos (data set), conjunto de entrenamiento (training set) y conjunto de prueba (test set).

A continuación, se desglosa e interpreta la distribución de las clases en cada uno de los subconjuntos: **1) Distribución general de clases:** las clases que se analizan son: carbón a granel; contenedores; general; granel líquido; granel sólido diferente de carbón. Las proporciones de cada clase en el conjunto de datos total, entrenamiento y prueba son las siguientes: **2) Análisis de la distribución entre el conjunto de entrenamiento y el conjunto de prueba:** la distribución de clases en los conjuntos de entrenamiento y prueba parece estar bastante equilibrada, lo que es positivo porque significa que ambas particiones reflejan de manera representativa la distribución global de las clases.

Aquí algunos puntos clave: **2.1) Proporciones similares entre los conjuntos:** las proporciones de cada clase en los conjuntos de entrenamiento y prueba son muy cercanas, lo que sugiere que no hay sesgo significativo en cómo se distribuyen las clases entre estos subconjuntos. Es decir, el modelo de entrenamiento tendrá una representación adecuada de cada clase durante el proceso de aprendizaje y de igual manera el conjunto de prueba tendrá una distribución proporcionalmente similar para la evaluación. **2.2) Proporciones de las clases en detalle:** carbón al granel: la clase de

carbón al granel tiene una proporción en torno al (12%) en todos los conjuntos, lo que implica que está equilibrada entre el conjunto de entrenamiento y el conjunto de prueba, lo que garantiza que el modelo podrá aprender y generalizar bien para esta clase; contenedores: la clase de contenedores tiene una proporción algo mayor en el conjunto de prueba (17.1%) que en el conjunto de entrenamiento (16.5%), lo que podría indicar que el conjunto de prueba tiene ligeramente más ejemplos de esta clase, esto no debería ser un problema grave, ya que la diferencia es pequeña, pero podría ser importante si las diferencias fueran mayores; general: la clase general tiene una proporción ligeramente mayor en el conjunto de entrenamiento (22%) que en el conjunto de prueba (20.5%).

Esto sugiere que el modelo tiene más ejemplos de esta clase durante el entrenamiento, lo que podría llevar a que el modelo aprenda mejor esta clase en particular, la ligera disminución de la proporción en el conjunto de prueba no debería afectar demasiado la capacidad de generalización; granel líquido: la clase granel líquido tiene una proporción bastante consistente, con (30.2%) en el conjunto completo de datos, (30.1%) en el conjunto de entrenamiento y (30.6%) en el conjunto de prueba, la distribución es equilibrada, lo que significa que el modelo podrá aprender a clasificar correctamente los ejemplos de esta clase; granel sólido diferente de carbón: la clase granel sólido diferente de carbón tiene una proporción muy similar en todos los conjuntos: (19.3%) en el conjunto completo, (19.2%) en el conjunto de

entrenamiento y (19.6%) en el conjunto de prueba, esta estabilidad en las proporciones también es positiva, ya que el modelo verá esta clase representada de manera consistente tanto en el entrenamiento como en la prueba.

En la tabla 1. Análisis del Splits in Tree donde se presenta los puntos de split (división) en el árbol de decisiones, junto con el número de observaciones en cada división (Obs. in Split), el valor del punto de división (Split Point) y la mejora (Improvement) en la desviación asociada con esa división. El análisis de esta tabla ayudará a entender cómo el árbol de decisiones está tomando decisiones en cuanto a las divisiones en función de las características de Transbordo y Cabotaje, y qué tan efectivo es cada split para mejorar el modelo.

La interpretación de las Columnas:- Obs. en Split: Esta columna indica el número de observaciones (instancias de datos) que están siendo evaluadas en ese punto de división del árbol. Por ejemplo, en el primer split de Transbordo, hay (5.712) observaciones involucradas.- Split Point: El punto de división indica el valor específico de la característica en el que el árbol toma

la decisión de dividir. Un valor negativo indica que el árbol está dividiendo según valores menores que este punto. Por ejemplo, para el primer split de Transbordo, el árbol divide en (-0.138).

- Improvement: El mejoramiento en la devianza es una métrica que indica cuán efectiva es esa división para mejorar la precisión del modelo. La devianza mide la falta de ajuste del modelo,

por lo que un aumento en el mejoramiento significa que el modelo se ajusta mejor a los datos tras realizar la división. Cuanto mayor sea el valor del Improvement, mejor es la partición. 3.

### **Análisis de los Splits en el Árbol: 3.1 Primer Split en “Transbordo”:**

Observaciones: (5712); Punto de Split: (-0.138); Mejora: (305.081). Este primer split en Transbordo parece ser extremadamente efectivo en la mejora del modelo, ya que tiene una mejora de (305.081), que es significativamente alta. Esto sugiere que el valor de (-0.138) para la variable Transbordo es un punto clave para la clasificación en este árbol de decisiones, lo que indica que este split es de gran importancia para diferenciar las clases en el modelo. 3.2 Segundo Split en “Transbordo”: Observaciones: (665); Punto de Split: (-0.103); Mejora: (84.899); este split en Transbordo también

**Tabla 4. Splits in Tree**

	<b>Obs. in Split</b>	<b>Split Point</b>	<b>Improvement</b>
Transbordo	5712	-0.138	305.081
Transbordo	665	-0.103	84.899
Cabotaje	392	-0.073	26.060
Cabotaje	5047	-0.006	108.664
Cabotaje	4546	-0.073	107.122

Note. For each level of the tree, only the split with the highest improvement in deviance is shown.

tiene un punto de división similar (-0.103) y una mejora considerable, aunque menor que el primer split. A pesar de que solo involucra (665) observaciones, la mejora sigue siendo sustancial con (84.899), lo que indica que esta división es útil para mejorar la precisión del árbol.

**3.3 Primer Split en “Cabotaje”:** Observaciones: (392); Punto de Split: (-0.073); Mejora: (26.060). Este split en Cabotaje tiene una mejora mucho más baja (26.060) en comparación con los anteriores. Además, involucra un número relativamente pequeño de observaciones (392). Esto sugiere que esta división tiene un impacto menos importante en la mejora del rendimiento del modelo en comparación con los splits previos.

**3.4 Segundo Split en “Cabotaje”:** Observaciones: (5047); Punto de Split: (-0.006); Mejora: (108.664). Este split en Cabotaje involucra (5047) observaciones, lo que indica que es una división con una representación más grande del conjunto de datos. Aunque el punto de división es cercano a (0), la mejora de (108.664) sugiere que esta división sigue siendo relevante para mejorar el modelo, aunque no tan significativa como los primeros splits de Transbordo.

**3.5 Tercer Split en “Cabotaje”:** Observaciones: (4546); Punto de Split: (-0.073); Mejora: (107.122). Este split en Cabotaje tiene una mejora de (107.122), que es casi tan alta como la del segundo split de Cabotaje (108.664), lo que indica que esta partición también contribuye de manera significativa a la mejora en la desviación. Sin embargo, al ser un valor de split similar al anterior (-0.073), es posible que el árbol haya dividido en un punto muy parecido para mejorar aún más la clasificación de las observaciones.

## 5. Conclusiones

La innovación constituye un eje fundamental para el futuro de los puertos marítimos, al impulsar la eficiencia, la sostenibilidad y la competitividad en el comercio global. Las tendencias actuales —la digitalización, la sostenibilidad y la mejora de la conectividad intermodal—están transformando la industria portuaria a nivel mundial. Sin embargo, para aprovechar plenamente los beneficios de la innovación es necesario un compromiso continuo con la investigación, el desarrollo y la implementación de nuevas tecnologías y prácticas. Solo así se puede garantizar que los puertos marítimos sigan siendo motores de crecimiento económico y desarrollo sostenible en el siglo XXI.

Desde la perspectiva técnica, el análisis de machine learning basado en el modelo de clasificación CART (árbol de decisión) evidencia la salida del target (carga) de forma nominal y las entradas (cabotaje y transbordo) donde el modelo del splits vislumbra la clasificación de datos en función de las características cabotaje y transbordo. Otro elemento el n(Train) el número de datos de entrenamiento usados para construir el modelo y n(Test) mostrando el número de datos utilizados para probar el modelo. Y por último el Test Accuracy analizó la precisión del modelo en los datos de prueba de las predicciones hechas por el árbol de decisiones. El procedimiento de data split analizó el modelo entrenado en un subconjunto de datos y evaluó un conjunto de datos independiente el cual proporciona la estimación de su rendimiento en los datos no vistos. Para el

caso aplicado, el cual representa aproximadamente los términos de proporción entre el conjunto de entrenamiento y el conjunto de prueba. El conjunto de datos permite que el modelo aprenda de las características y relaciones presentes en los datos, siendo importante para los modelos en este caso los árboles de decisión para capturar y determinar los patrones complejos.

Las predicciones correctas como el modelo está prediciendo con éxito la clase granel líquido en una probabilidad cuando el observado es también granel líquido. La clase general tiene una predicción correcta en la columna correspondiente con la probabilidad, los errores de clasificación cuando la clase observada es carbón al granel, el modelo predice granel líquido con una probabilidad, este es un error importante ya que la clase observada es muy diferente a la predicha. La clase contenedores se clasificó erróneamente en granel líquido y general, con probabilidades pequeñas lo que indica que el modelo tiene dificultades para distinguir esta clase de otras. Para las observaciones de granel sólido diferente de carbón, se predice incorrectamente como granel líquido y como general. Este es otro caso importante de confusión entre clases, lo que refleja que el modelo tiene problemas para diferenciar estas clases. Métricas para la matriz de confusión para evaluar el rendimiento del modelo donde la precisión donde el cálculo de precisión y recall para la clase carbón al granel, se observa la no predicción correcta en esa clase. Afectando negativamente la precisión como el recall para esta clase.

La matriz de confusión proporciona un desglose de las predicciones correctas e incorrectas para cada clase, útil para evaluar cómo se comporta el modelo con cada clase individual. La distribución de clases en los conjuntos de entrenamiento y prueba parece estar bastante equilibrada, lo que es positivo porque significa que ambas particiones reflejan de manera representativa la distribución global de las clases. Las proporciones de cada clase en los conjuntos de entrenamiento y prueba son muy cercanas, lo que sugiere que no hay sesgo significativo en cómo se distribuyen las clases entre estos subconjuntos. Es decir, el modelo de entrenamiento tendrá una representación adecuada de cada clase durante el proceso de aprendizaje y de igual manera el conjunto de prueba tendrá una distribución proporcionalmente similar para la evaluación.

La clase de carbón al granel tiene una proporción en torno a todos los conjuntos, lo que implica que está equilibrada entre el conjunto de entrenamiento y el conjunto de prueba, lo que garantiza que el modelo podrá aprender y generalizar bien para esta clase. La clase de contenedores tiene una proporción algo mayor en el conjunto de prueba que en el conjunto de entrenamiento, lo que podría indicar que el conjunto de prueba tiene ligeramente más ejemplos de esta clase. Esto no debería ser un problema grave, ya que la diferencia es pequeña, pero podría. La clase general tiene una proporción ligeramente mayor en el conjunto de entrenamiento que en el conjunto de prueba esto sugiere que el modelo tiene más

ejemplos de esta clase durante el entrenamiento, lo que podría llevar a que el modelo aprenda mejor esta clase en particular. La ligera disminución de la proporción en el conjunto de prueba no debería afectar demasiado la capacidad de generalización. La clase granel líquido tiene una proporción bastante consistente con el conjunto completo de datos en el conjunto de entrenamiento y en el conjunto de prueba. La distribución es equilibrada lo que significa que el modelo podrá aprender a clasificar correctamente los ejemplos de esta clase. La clase granel sólido diferente de carbón tiene una proporción muy similar en todos los conjuntos, el conjunto completo, el conjunto de entrenamiento y el conjunto de prueba. Esta estabilidad en las proporciones también es positiva ya que el modelo verá esta clase representada de manera consistente tanto en el entrenamiento como en la prueba.

Los splits de Transbordo son los que tienen el mayor impacto en la mejora del modelo con mejoras significativamente altas indicando

que esta variable tiene un gran poder discriminatorio y es muy relevante para el modelo. Donde los splits de Cabotaje también son importantes aunque su impacto es menor en términos de mejora de desviación. Sin embargo, debido al número de observaciones involucradas estos splits aún tienen un efecto importante en el ajuste general del modelo. Los splits en Cabotaje tienen una mejora relativamente menor en comparación con los de Transbordo lo que podría sugerir que Transbordo es una variable más influyente o relevante para la clasificación en el modelo.

En general, los splits más efectivos tienden a ser aquellos que involucran variables con un mayor número de observaciones y aquellos en los que el modelo logra una mejora significativa en la desviación. El análisis de los splits del árbol de decisiones muestra cómo el modelo utiliza diferentes umbrales de las características para mejorar la precisión en cada etapa de la clasificación con un enfoque en transbordo y cabotaje como las principales características discriminatorias.

## Referencias Bibliográficas

- Chen, Y., et al. (2022). *Sustainable practices in port management: A review*. *Journal of Sustainable Development*, 15(3), 45–58.
- Chen, Y., & Zhang, J. (2019). *Optimization of port infrastructure: A case study of major ports in Asia*. *Journal of Transportation Engineering*, 25(3), 198–212.
- Fancello, G., Vitiello, D. M., & Serra, P. (2022). *The impact of globalization on Mediterranean container terminals*. *International Maritime Transport and Logistics Conference*, 11(1), 86–95.
- García, A., & López, M. (2021). *Digital transformation in port management: Opportunities and challenges*. *International Journal of Digital Innovation in Port Management*, 8(2), 120–135.

- García, A., & López, M. (2022). *Cloud-based terminal operating systems: A review of current trends*. *International Journal of Port Management*, 9(1), 45–58.
- Gesé Bordils, M., González-Cancelas, N., & Molina Serrano, B. (2021). *Indicadores clave de rendimiento en terminales de contenedores y su relación con la sostenibilidad ambiental: Aplicación al sistema portuario español*. *Ingeniare. Revista Chilena de Ingeniería*, 29(4), 647–660. <https://doi.org/10.4067/S0718-33052021000400647>
- Jones, R., & Wang, Q. (2019). *The role of digitalization in port operations: A case study of major ports in Asia*. *Journal of Maritime Studies*, 25(4), 321–335.
- Jones, R., et al. (2020). *Improving land connectivity in port development: Lessons from case studies*. *Maritime Policy & Management*, 37(2), 120–135.
- Li, H., & Wu, X. (2023). *Sustainable practices in port management: A comparative analysis*. *Journal of Sustainable Development*, 16(2), 89–102.
- Sarduy Domínguez, Y. (2007). *El análisis de información y las investigaciones cuantitativa y cualitativa*. *Revista Cubana de Salud Pública*, 33(3).
- Smith, P., & Wang, Q. (2018). *Automation in port operations: Challenges and opportunities*. *Journal of Automation in Transportation*, 12(4), 321–335.
- Smith, P. (2020). *Green infrastructure in port development: A comparative analysis*. *Journal of Environmental Planning and Management*, 28(1), 56–68.
- Tan, L., et al. (2021). *Promoting sustainable port development through renewable energy: A case study of major ports in Europe*. *Journal of Environmental Planning and Management*, 30(3), 212–225.
- Tan, L., et al. (2023). *Intermodal connectivity in port development: Lessons from case studies*. *Maritime Policy & Management*, 40(3), 212–225.
- Wu, X., et al. (2021). *Sustainable port development: Best practices and future directions*. *Journal of Sustainable Transportation*, 12(4), 89–102.
- Zhang, L. (2020). *Empirical study on improving international dry port competitiveness based on logistics supply chain integration: Evidence from China*. *International Journal of Logistics Management*, 33(3), 1040–1068. <https://doi.org/10.1108/IJLM-06-2020-0256>