

UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI



FACULTAD DE COMERCIO INTERNACIONAL, INTEGRACIÓN, ADMINISTRACIÓN Y ECONOMÍA EMPRESARIAL

CARRERA DE LOGÍSTICA Y TRANSPORTE

Tema: "Sistema de predicción de costos de transporte de carga pesada utilizando herramientas de inteligencia artificial"

Trabajo de Integración Curricular previo a la obtención del
título de Ingeniero en Logística y Transporte

AUTOR: Benavides Caipe Denis Enrique

TUTOR: MSc. Herrera Granda Israel David

Tulcán, 2026.

CERTIFICADO DEL TUTOR

Certifico que el estudiante Benavides Caipe Denis Enrique con el número de cédula 0450156864, ha desarrollado el Trabajo de Integración Curricular: "Sistema de predicción de costos de transporte de carga pesada utilizando herramientas de inteligencia artificial".

Este trabajo se sujeta a las normas y metodología dispuesta en la Codificación del Reglamento de Régimen Académico y de Estudiantes de la UPEC, por lo tanto, autorizo la presentación de la sustentación para la calificación respectiva.

MSc. Herrera Granda Israel David

TUTOR

Tulcán, julio de 2026

AUTORÍA DE TRABAJO

El presente Trabajo de Integración Curricular constituye un requisito previo para la obtención del título de Ingeniero en la Carrera de logística y transporte de la FACULTAD DE COMERCIO INTERNACIONAL, INTEGRACIÓN, ADMINISTRACIÓN Y ECONOMÍA EMPRESARIAL.

Yo, Benavides Caipe Denis Enrique, con cédula de identidad número 0450156864, declaro que la investigación es absolutamente original, auténtica, personal y los resultados y conclusiones a los que he llegado son de mi absoluta responsabilidad.



Benavides Caipe Denis Enrique

AUTOR

Tulcán, julio de 2026

ACTA DE CESIÓN DE DERECHOS DEL TRABAJO DE INTEGRACIÓN CURRICULAR

Yo Benavides Caipe Denis Enrique declaro ser autor de los criterios emitidos en el Trabajo de Integración Curricular: "Sistema de predicción de costos de transporte de carga pesada utilizando herramientas de inteligencia artificial" y eximo expresamente a la Universidad Politécnica Estatal del Carchi y a sus representantes de posibles reclamos o acciones legales.



Benavides Caipe Denis Enrique

AUTOR

Tulcán, julio de 2026

AGRADECIMIENTO

Agradezco profundamente a Dios por ser el motor de mi vida, por nunca abandonarme en los momentos más complicados y por colocar en mi camino a personas maravillosas que hicieron de este proceso algo más llevadero y significativo. Gracias por cada bendición recibida y por permitirme llegar hasta aquí. A mis padres, Alba Caipe y Jorge Benavides, por ser mi mayor inspiración; gracias por cada abrazo en los momentos difíciles, por cada consejo dado con amor y por nunca permitir que me rindiera cuando las fuerzas me faltaban. Su entrega y su dedicación hacia mí son el regalo más grande que la vida me ha dado. A mis hermanos, Alexis Benavides y Jhony Benavides, gracias por estar presentes en cada etapa de este camino, por escucharme, por hacerme reír cuando más lo necesitaba y por brindarme esa tranquilidad que solo se encuentra en la familia. A mi novia, Dayana Garcito, gracias por su apoyo incondicional, por cada momento de comprensión, por celebrar conmigo cada pequeño avance y por nunca dejar de creer en mí incluso en los días en que todo parecía más difícil de lo normal; su amor ha sido un pilar fundamental en este logro. A mi sobrino, Jordan Benavides, aunque aún es pequeño para entenderlo, su alegría y su energía fueron una fuente de motivación constante que me recordaba siempre por qué vale la pena esforzarse. A mis compañeros de curso, y en especial a Jordan Chamorro y David Salas, gracias por cada momento compartido, por su lealtad, por su apoyo desinteresado y por hacer de esta etapa universitaria una experiencia llena de buenos recuerdos que llevaré siempre en el corazón. Finalmente, a mi tutor y a todos mis docentes, gracias por su compromiso y su vocación, por transmitir su conocimiento con pasión y entrega, por corregirme con respeto y por guiarme con paciencia a lo largo de toda mi formación académica; gracias a ustedes hoy puedo decir con orgullo que estoy preparado para enfrentar los retos que la vida profesional me depare.

DEDICATORIA

Dedico este trabajo primeramente a Dios, por ser mi luz y mi guía en cada paso del camino, por darme la sabiduría necesaria para tomar las decisiones correctas y la fortaleza para no rendirme cuando el camino se ponía difícil. Gracias por estar presente en cada momento de este proceso y por recordarme siempre que con fe todo es posible. A mis padres, Alba Caipe y Jorge Benavides, por ser mi mayor motivación, mi fuerza y mi refugio; por cada sacrificio silencioso que hicieron por mí, por cada palabra de aliento en los momentos más difíciles y por enseñarme con su ejemplo que los sueños se alcanzan con esfuerzo, perseverancia y amor. Este logro es el fruto de todo lo que me han dado. A mis hermanos, Alexis y Jhony Benavides, por caminar siempre a mi lado, por ser mis cómplices, mis amigos y mi apoyo incondicional a lo largo de toda mi vida. Gracias por creer en mí incluso en los momentos en que yo mismo dudaba. Este logro no es solo mío, es de todos ustedes, porque sin su amor y su presencia nada de esto hubiera sido posible.

ÍNDICE

RESUMEN.....	11
ABSTRACT	12
INTRODUCCIÓN	13
I. EL PROBLEMA.....	15
1.1. PLANTEAMIENTO DEL PROBLEMA	15
1.2. FORMULACIÓN DEL PROBLEMA.....	16
1.3. JUSTIFICACIÓN.....	16
1.4. OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN	17
1.4.1. Objetivo General	17
1.4.2. Objetivos Específicos	17
1.4.3. Preguntas de Investigación	18
II. FUNDAMENTACIÓN TEÓRICA	19
2.1. ANTECEDENTES DE LA INVESTIGACIÓN	19
2.2. MARCO TEÓRICO	20
2.2.1 Costos en Transporte de Carga Pesada	20
2.2.2 Proceso de predicción.....	24
2.2.3 Predicción de costos de Transporte de Carga Pesada	25
2.2.4 ¿Qué es la inteligencia artificial?	26
2.2.5 Inteligencia artificial para la predicción de costos de Transporte de Carga Pesada.....	27
2.3 ESTADO AL ARTE	31
III. METODOLOGÍA	43
3.1. ENFOQUE METODOLÓGICO	43
3.1.1. Enfoque	43
3.1.2. Tipo de Investigación.....	43

3.1.3 Caso de estudio.....	44
3.2. HIPÓTESIS	47
3.3. DEFINICIÓN Y OPERACIONALIZACIÓN DE LAS VARIABLES.....	48
3.4. MÉTODOS UTILIZADOS	53
3.4.1 METODOLOGIA PROPUESTA PARA EL PRONÓSTICO DE COSTOS EN TRANSPORTE DE CARGA PESADA.....	56
3.5. ANÁLISIS ESTADÍSTICO	61
3.5.1 Técnicas de análisis de datos	61
3.5.2 Técnicas de recolección de datos	62
3.5.3 Instrumentos de investigación	63
IV. RESULTADOS Y DISCUSIÓN	64
4.1. RESULTADOS	64
4.1.1 Análisis que inciden en los costos de transporte de carga pesada	64
4.1.2 Modelos de inteligencia artificial aplicados al cálculo de costos de transporte	71
4.1.3 Desarrollo del modelo predictivo basado en Random Forest para la estimación de costos de transporte de carga pesada.....	73
4.1.4 Evaluación del modelo predictivo	81
4.1.5 Diseño e implementación de la interfaz web del sistema predictivo.....	94
4.1.6 Módulos funcionales de la aplicación.....	94
4.1.7 Comprobación de la hipótesis	100
4.2. DISCUSIÓN	102
V. CONCLUSIONES Y RECOMENDACIONES	106
5.1. CONCLUSIONES	106
5.2. RECOMENDACIONES.....	108
VI. REFERENCIAS BIBLIOGRÁFICAS.....	109
VII. ACRÓNIMOS	112
VIII. ANEXOS.....	113

ÍNDICE DE TABLAS

Tabla 1. Comparación de estudios relacionados	34
Tabla 2. Operacionalización de la variable independiente	49
Tabla 3. Operacionalización de la variable dependiente	50
Tabla 4. Herramientas utilizadas por cada fase.	56
Tabla 5. Distancia en kilómetros, tiempo en horas y paradas por viaje por ruta.	67
Tabla 6. Costo de combustible por rutas.	68
Tabla 7. Costos de mantenimiento del transporte de carga pesada.	69
Tabla 8. Costos operativos de transporte de carga pesada.	70
Tabla 9. Comparación de herramientas de Inteligencia Artificial.....	72
Tabla 10. Variables de entrada y salida para el modelo predictivo.	75
Tabla 11. Muestra de registros del dataset de entrenamiento.	77
Tabla 12. Resumen mensual de costos de la flota.	78
Tabla 13. Configuración de hiperparámetros del modelo Random Forest.	80
Tabla 14. Resultados de la evaluación del modelo predictivo.....	84
Tabla 15. Comparación de las predicciones con los datos reales.....	90
Tabla 16. Comparación de costos reales con las predicciones del modelo	100

ÍNDICE DE FIGURAS

Figura 1. Distribución de modelos de IA utilizados en los estudios revisados	39
Figura 2. Métricas de evaluación empleadas en los estudios revisados	40
Figura 3. Ámbitos de aplicación de los estudios revisados	40
Figura 4. Diagrama del proceso de transporte	46
Figura 5. Metodología del estudio del sistema predictivo.....	61
Figura 6. Variables que afectan a los costos de transporte.....	76
Figura 7. Comparación del MAE obtenidas por los tres modelos evaluados.....	84

Figura 8. Comparación del MSE en los tres modelos evaluados.	85
Figura 9. Comparación del RMSE en los tres modelos evaluados.	86
Figura 10. Comparación entre los valores reales y valores predichos.	89
Figura 11. Representatividad de las variables de entrada en el costo total.	91
Figura 12. Panel Operativo de entrada de datos del sistema.	95
Figura 13. Módulo de trazado de rutas georreferenciadas.	96
Figura 14. Módulo de predicción y diagnóstico con IA.	97
Figura 15. Módulo de historial de predicciones.	98
Figura 16. Informe técnico en formato PDF.	99

ÍNDICE DE ANEXOS

Anexo 1. Certificado de abstract por parte de idiomas.	113
Anexo 2. Excel utilizado para la recolección de datos.	115
Anexo 3. Instrumentos validados.	120

RESUMEN

El transporte de carga pesada representa un 80% del movimiento de mercancías en Ecuador; sin embargo, las empresas del sector dependen de métodos empíricos de estimación de costos que presentan márgenes de error de entre 15% y el 25%, comprometiendo su rentabilidad y competitividad, por lo que esta investigación propone el desarrollo e implementación de TRANSPORT AI, un sistema web predictivo basado en varios algoritmos de *Machine Learning*, *Random Forest*, *Regresión Lineal* y *Gradient Boosting*, el cuál procesa variables operativas dinámicas: distancia recorrida, peso en toneladas, número de paradas, precio del combustible, peajes, viáticos y tipo de carga para generar estimaciones adaptadas a las condiciones reales de cada viaje, empleando un enfoque mixto validado sobre un caso de estudio sobre registros históricos de una empresa transportista ecuatoriana. En la fase de entrenamiento, el modelo arrojó un coeficiente de determinación $R^2 = 0,9999$, un Error Absoluto Medio (MAE) de \$1,31 USD y una Raíz del Error Cuadrático Medio (RMSE) de \$3,25 USD, mientras que la validación contra costos reales obtuvo un Error Porcentual Absoluto medio (MAPE) de 0,19%, confirmando la superioridad del modelo sobre los métodos tradicionales de cotización; adicionalmente se desarrolló una interfaz web que integra trazado georreferenciado de rutas, recalibración automática ante variaciones del precio del diésel y generación de informes técnicos en PDF, estandarizando el proceso de cotización y eliminando el error humano en el cálculo de fletes.

Palabras Claves: Carga pesada, predicción de costos, Machine Learning, Random Forest, gestión de flotas.

ABSTRACT

Heavy freight transportation accounts for approximately 80% of goods movement in Ecuador. However, companies in this sector continue to rely on empirical cost estimation methods with error margins ranging from 15% to 25%, compromising their profitability and competitiveness. The study proposes the development and implementation of TRANSPORT AI, a predictive web-based system that integrates multiple Machine Learning algorithms, including Random Forest, Linear Regression, and Gradient Boosting. The system processes dynamic operational variables such as distance traveled, weight in metric tons, number of stops, fuel price, toll costs, travel allowances, and type of freight to generate cost estimates tailored to the actual conditions of each trip. A mixedmethods approach was employed and validated through a case study based on historical records from an Ecuadorian transportation company. During the training phase, the model achieved a coefficient of determination of $R^2 = 0.9999$, a Mean Absolute Error (MAE) of USD 1.31, and a Root Mean Square Error (RMSE) of USD 3.25. Validation against actual transportation costs yielded a Mean Absolute Percentage Error (MAPE) of 0.19%, confirming the model's superiority over traditional pricing methods. Additionally, a web interface was developed that integrates georeferenced route mapping, automatic recalibration in response to diesel price fluctuations, and the generation of technical reports in PDF format. These features standardize the quotation process and eliminate human error in freight cost calculations.

Keywords: heavy freight transportation, cost prediction, Machine Learning, Random Forest, fleet management.

INTRODUCCIÓN

El transporte de carga pesada es una pieza importante de la economía y de la cadena de suministro, interconectando de esta manera distintos nodos logísticos sustantivos en el país. A contraparte, la administración financiera y la administración de flotas tienden a estar sometidas, permanentemente y en un corto espacio de tiempo, a la estocasticidad del entorno. Históricamente, se ha desarrollado mediante el uso de matrices contables tradicionales. Estas herramientas, al basarse en promedios estáticos y cálculos determinísticos, resultan incapaces de procesar la variabilidad dinámica en los precios del combustible, el desgaste mecánico asimétrico por el tonelaje, variaciones operativas durante las paradas o el impacto en unidades vehiculares de distintas características.

Esta limitación analítica frente a variables no lineales engendra una alta incertidumbre en la cotización de rutas comerciales. Al no ser capaz de predecir de manera precisa el desgaste técnico efectivo de un viaje antes de su realización, las empresas logísticas inevitablemente aceptan márgenes de error que comprometen directamente su rentabilidad y su punto de equilibrio. En tal contexto, la adopción de las tecnologías de la Industria 4.0, la Inteligencia Artificial y el *Machine Learning* de una parte, no la ofrecen como una alternativa, sino como una necesidad para poder transformar y modernizar la toma de decisiones de las empresas.

Para poder atenuar esta problemática, el presente trabajo de investigación centra en la propuesta de desarrollo e implementación del sistema predictivo web "TRANSPORT AI". Este sistema contiene un modelo algorítmico avanzado basado en *Random Forest*, el cual permite predecir con alta fidelidad los costos logísticos del transporte pesado, a diferencia de los métodos tradicionales que, para ello, son incapaces de procesar un conjunto de variables de entrada dinámicas como son la fecha de operación, el número de paradas operativas, el costo de la gasolina actualizado, el tipo de carga, el peso en toneladas, viáticos, peajes y la longitud de la ruta, para diagnosticar el costo adaptado, a las condiciones exactas del viaje a realizar.

El objetivo del presente trabajo es demostrar matemáticamente que un sistema de predicción basado en inteligencia artificial supera la precisión de las estimaciones manuales tradicionales en la cotización de fletes de carga pesada. Para ello, el

proyecto persigue tres finalidades específicas: primero, cuantificar la ventaja predictiva del modelo algorítmico frente a los métodos empíricos convencionales; segundo, automatizar el proceso de cálculo y generar informes técnicos auditables que estandaricen la documentación operativa; y tercero, proveer a la gerencia logística de una herramienta corporativa interactiva, escalable y de acceso inmediato. La validación estadística del modelo, obtenida mediante la contrastación de las predicciones frente a los costos reales históricos registrados por la empresa ejemplo, arrojó un margen de error global de 0,19%, confirmando que TRANSPORT AI constituye una solución informática robusta, capaz de transformar datos operativos en inteligencia de negocio procesable, trazable y financieramente confiable.

I. EL PROBLEMA

1.1. PLANTEAMIENTO DEL PROBLEMA

A nivel global, el transporte de carga pesada es uno de los principales componentes que sustentan las cadenas de suministro y el comercio internacional. La alta volatilidad de algunos de sus elementos asociados al transporte de mercancías, como los precios de combustible, los parámetros de la infraestructura del camino, la regulación gubernamental, los peajes y pagos, y los tiempos de entrega (Rodríguez, 2020), entorpecen la correcta planificación y gestión del transporte de carga pesada, dificultando la toma de decisiones estratégicas y operativas que llevan a cabo estas empresas hoy en día, así como la correcta prevención de problemas económicos y medioambientales. Ante esta situación, la inteligencia artificial se constituye en una de las alternativas más prometedoras que pueden servir para modelar y prever costos, dado el acceso a grandes cantidades de datos históricos y en tiempo real, los cuales pueden combinarse para mejorar la eficiencia (Davenport y Ronanki, 2018).

En Latinoamérica, el transporte de carga pesada presenta problemas estructurales que agravan la incertidumbre en los costos, se hace evidente que la disparidad de la calidad de carreteras, la falta de armonización entre países, la informalidad del sector y el escaso grado de digitalización limitan la capacidad de las empresas para poder predecir sus costos de operación (Comisión Económica para América Latina y el Caribe [CEPAL], 2021). En este contexto, el uso de tecnologías de IA es todavía relativamente poco común, sobre todo en el caso de las empresas pequeñas y medianas de transporte, manteniendo la dependencia de métodos empíricos y tradicionales.

En Ecuador, el transporte terrestre de carga pesada representa el 80% del movimiento de mercancías a escala nacional (Ministerio de Transporte y Obras Públicas [MTOP], 2023); sin embargo, este sector está caracterizado por una alta fragmentación y por un notable predominio de microempresas y flotas pequeñas que no cuentan con sistemas tecnológicos avanzados para su gestión logística. Sumado a eso, factores

como la variabilidad en los precios del diésel, el estado de las vías interprovinciales y los controles aduaneros en zonas fronterizas hacen más complejo el cálculo de costos (Banco Central del Ecuador [BCE], 2022). Estas condiciones generan ineficiencias que se traducen en mayores precios para los consumidores y menor competitividad del país en el comercio regional. A pesar de los avances tecnológicos, actualmente no existen soluciones sustentadas en IA que se basen precisamente en el cálculo de costos de transporte de carga pesada dentro del marco del transporte en Ecuador.

Así, se percibe de forma clara y evidente la necesidad de diseñar un sistema de predicción de costos de transporte de carga pesada utilizando herramientas inteligencia artificial que se adapte a las características del entorno logístico ecuatoriano. Un sistema de predicción de costos que no solo optimice la planificación del transporte de las empresas, sino que también ayude a prevenir problemáticas internas y externas que afecten directamente a los costos de transporte.

1.2. FORMULACIÓN DEL PROBLEMA

¿Cómo contribuyen los sistemas de predicción basados en inteligencia artificial a estimar con mayor precisión los costos de transporte de carga pesada?

1.3. JUSTIFICACIÓN

El transporte de carga es un componente fundamental para el desarrollo económico, pues es el medio que permite movilizar eficientemente, tanto las mercancías como materias primas entre regiones y países. En Ecuador este sector tiene la representación del 80% en el transporte terrestre nacional de mercancías (MTOP, 2023), por lo que su adecuada gestión influye en la competitividad empresarial y en los precios finales al consumidor. El estudio es pertinente porque aborda una problemática que afecta tanto a la economía nacional como a la eficiencia operativa de las empresas del sector logístico.

La relevancia social de la investigación radica en que la optimización de los costos de transporte repercute en la disminución de los precios de los productos, beneficiando a los consumidores y contribuyendo a una distribución más equitativa de recursos. Además, una mejor planificación del transporte puede reducir emisiones contaminantes e impactos negativos en el medio ambiente. Las variaciones en los

precios de combustible, el estado de las vías, los peajes. Los tiempos de espera y la falta de digitalización de procesos incrementan la incertidumbre en la estimación de costos, lo cual demuestra la urgencia de implementar este tipo de soluciones (Banco Central del Ecuador [BCE], 2022).

El desarrollo de un sistema de predicción de costos basado en IA permitirá a las empresas anticipar variaciones en sus gastos operativos, planificar rutas más eficientes y mejorar la toma de decisiones estratégicas. Estas herramientas, al analizar grandes volúmenes de datos en tiempo real, pueden predecir con alta precisión los costos futuros y facilitar una mejor administración de los recursos (Davenport y Ronanki, 2018).

El estudio contribuye al campo de la gestión logística y los sistemas inteligentes aplicados al transporte en Ecuador. El desarrollo del sistema y su validación con datos reales podrán servir como referencia para futuras investigaciones o implementaciones similares en el sector del transporte, la relevancia general del trabajo radica en su potencial para unir la teoría con la práctica, proporcionando un marco analítico y tecnológico capaz de mejorar la eficiencia del transporte de carga pesada y apoyar a la modernización del sector logístico nacional.

1.4. OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN

1.4.1. Objetivo General

Optimizar el proceso de cálculo de costos de transporte de carga pesada mediante el uso de herramientas de inteligencia artificial

1.4.2. Objetivos Específicos

- Analizar los factores que inciden en los costos del transporte de carga pesada y los modelos de inteligencia artificial utilizados para su cálculo y predicción.
- Desarrollar un modelo predictivo basado en técnicas de inteligencia artificial que permita la estimación de los costos de transporte con una mayor precisión.
- Evaluar la precisión del sistema de predicción de costos mediante pruebas experimentales, comparación con métodos tradicionales de estimación y métricas de evaluación de algoritmos de inteligencia artificial (MSE, MAE).

1.4.3. Preguntas de Investigación

- ¿Cuáles son los factores determinantes que influyen en la estructura de costos de las empresas de transporte de carga pesada y qué modelos de inteligencia artificial se utilizan para su cálculo y predicción?
- ¿Cómo se puede desarrollar un modelo predictivo basado en técnicas de inteligencia artificial que permita estimar con precisión los costos de transporte?
- ¿Qué métodos se pueden utilizar para validar la eficacia del sistema de predicción de costos mediante pruebas experimentales y comparar con métodos tradicionales de estimación?

II. FUNDAMENTACIÓN TEÓRICA

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

Rodríguez y Ugalde (2021) llevaron a cabo una investigación cuyo objetivo principal fue el de analizar los efectos de estandarizar y escalar variables en la calidad de las predicciones en torno a la estimación de costos llevados a cabo a través de redes neuronales artificiales (RNA) para tratar de comparar el impacto de la aplicación de cuatro técnicas de preprocesamiento de datos para mejorar la calidad del modelo. Para la investigación tomaron como referencia una RNA utilizada en proyectos reales a partir de datos aleatorios y constataron la calidad del rendimiento de dicha RNA por medio del error cuadrático medio (MSE). Para otorgar una explicación general a todos los *tests* realizados, los resultados obtenidos han llegado a indicar que la técnica de escalado de rango es la que viene a ofrecer mínimamente un error cuadrático medio (MSE) significativamente menor y como tal logra mejorar la precisión de la predicción. Después de todo, el presente trabajo debería contribuir a aportar evidencias empíricas sobre la importancia del preprocesamiento de datos en modelos predictivos con inteligencia artificial, el cual resulta ser un elemento muy importante si se quiere garantizar una estimación fiable en los costos del transporte.

En un trabajo reciente, Palafox-Palafox et al. (2023) presentan un modelo del nivel de inventario en una cadena de suministro a través de RNA. Se trata de un trabajo en el que el objetivo que querían comprobar era si la obtención de RNA proporcionaba un rendimiento superior al de la obtención tradicional mediante regresiones lineales o promedio móvil. Para ello, generaron datos históricos mediante simulación y entrenaron 110 configuraciones de RNA. La arquitectura obtenida presenta un $R^2 = 0,994$ y una raíz del error cuadrático medio (RMSE) de 1,44. Este trabajo muestra, por tanto, que, en un entorno logístico complejo, las RNA ofrecen un rendimiento superior al resto de los métodos empleados. Además, aporta una metodología transferible a la predicción del costo en el transporte de carga pesada.

González Escobar (2021) realizó un estudio comparativo de las tarifas del transporte pesado en la frontera entre Ecuador y Colombia, en la frontera Tulcán-Ipiiales, para poder analizar su incidencia en la competitividad. Utilizó una metodología cuantitativa, aplicando encuestas a transportistas además de un modelo de costos basado en variables operativas. Los resultados encontraron diferencias significativas en costos asociados al combustible, mantenimiento y capacitación del personal transportista. Si bien no utilizó la IA, su modelo de costos estructurado constituye una base sólida para la selección de las variables de entrada en un sistema predictor de IA que podría abordarse desde un enfoque de *big data*.

Matel et al. (2022) desarrollaron un modelo basado en redes neuronales artificiales (RNA) para la estimación de costos en servicios de ingeniería, con el propósito de mejorar la precisión frente a otros métodos de estimación. El estudio analizó una base de datos de 132 proyectos y aplicó un proceso de optimización heurística para ajustar el desempeño del modelo. Los resultados demostraron que las RNA lograron una mejora del 14,5% en la precisión del costo estimado respecto a otros métodos, alcanzando un error medio absoluto porcentual (MAPE) de 13,65%. Además, los autores identificaron 7 variables determinantes para la predicción de costos: intensidad, número de miembros del equipo, duración del proyecto, disciplinas colaboradoras, tipo de contacto, fases del proyecto y escala del trabajo. Este antecedente evidencia que el uso de redes neuronales es eficaz incluso con conjuntos de datos reducidos, y refuerza la pertinencia de aplicar modelos de IA en la predicción de costos en contextos logísticos y el transporte de carga pesada.

2.2. MARCO TEÓRICO

2.2.1 Costos en Transporte de Carga Pesada

El transporte de carga pesada es una de las piezas fundamentales de la economía en América Latina y, en particular, del Ecuador. La comprensión de su estructura de costos es esencial para la planificación financiera, la competitividad empresarial y la formulación de políticas públicas del sector. Los costos en este ámbito son heterogéneos y responden a múltiples factores: distancia recorrida, tipo de carga, condiciones de vías, precio de los insumos y marco normativo vigente (Valencia, 2023).

En términos generales, los costos del transporte de carga pesada se clasifican en costos directos, indirectos y operativos. Los costos directos incluyen combustible, peajes y remuneraciones a los conductores; representan la mayor proporción del gasto total y su monitoreo continuo es indispensable para la rentabilidad operativa. Los costos indirectos abarcan mantenimiento vehicular, seguros y administración logística; aunque menos visibles en la operación diaria, su acumulación puede comprometer la sostenibilidad financiera de las empresas. Los costos operativos se relacionan con el uso de infraestructura y equipamiento especializado de carga, reflejando las condiciones reales del entorno logístico (Ministerio de Obras Públicas [MOP], 2023).

Adicionalmente, factores externos como el estado de las vías, las condiciones climáticas, los bloqueos imprevistos y la variabilidad estacional de la demanda inciden de manera significativa en el costo final. El tipo, peso y volumen de la carga determinan también los requerimientos especiales de transporte, los permisos necesarios y los seguros aplicables, elevando el gasto total cuando se trata de cargas peligrosas o sobredimensionadas (Valencia, 2023).

2.2.1.1 Procesamiento para el cálculo del costo en Transporte de Carga Pesada

El cálculo del costo de transporte de carga pesada requiere un procedimiento metodológico estructurado que integre todas las variables relevantes del proceso logístico. Este procedimiento parte de la identificación y cuantificación de cada componente de costo, agrupándolos en categorías que faciliten su análisis y proyección.

El primer paso consiste en la recopilación de datos históricos de operación: registros de kilómetros recorridos, consumo de combustible por tipo de vehículo y ruta, costos de mantenimiento preventivo y correctivo, valores de peajes, salarios de conductores y gastos administrativos asociados. Esta información constituye la base empírica del modelo de costos.

El segundo paso involucra la clasificación y segmentación de los costos según su naturaleza: fijos (independientes del volumen de operación, como el seguro anual o la depreciación del vehículo) y variables (directamente proporcionales a la

operación, como el combustible y los peajes). Esta distinción es fundamental para el análisis de punto de equilibrio y la fijación de tarifas competitivas.

El tercer paso corresponde al cálculo del costo por kilómetro, indicador clave que permite comparar la eficiencia operativa entre flotas, rutas y periodos. Según Montenegro Herrera (2015) esta métrica se obtiene dividiendo el costo total de operación entre los kilómetros efectivamente recorridos en un periodo determinado, y sirve como referencia para la planificación de rutas y la negociación de contratos.

Finalmente, se incorporan los costos regulatorios y legales -permisos especiales, tasas de tránsito y multas por incumplimiento, así como los costos externos derivados de contingencias operativas. La suma de todos estos componentes permite obtener el costo total del servicio, sobre el cual se aplica el margen de utilidad para determinar la tarifa final al cliente (Valencia, 2023).

Costos directos

Son aquellos que se asocian de manera inmediata con la ejecución del transporte, como combustible, peajes y remuneraciones a los conductores. Representan la base del costo total, y su monitoreo continuo es esencial para la rentabilidad de las operaciones (Valencia, 2023).

Costos indirectos

Los costos indirectos incluyen los gastos de mantenimiento vehicular, seguros y administración logística. Estos costos inciden significativamente en la rentabilidad. Katreddi et al. (2023) demostraron que el uso de inteligencia artificial para predecir costos de mantenimiento reduce el margen de error de estimación en un 12%.

Costos operativos

Se refieren a los gastos relacionados con el uso de infraestructura y equipamiento de carga. Estos costos influyen en la eficiencia de las operaciones y deben ser considerados dentro del modelo para reflejar escenarios reales (MOP, 2023).

Costos según la carga

El tipo, peso y volumen de la carga determinan directamente el costo final. Las cargas pesadas, frágiles o peligrosas requieren condiciones especiales de transporte, permisos y seguros, elevando el gasto total.

Tipo de vehículo

El tipo de camión empleado afecta el consumo de combustible y el costo por kilómetro. Las flotas modernas con motores más eficientes pueden reducir significativamente los costos logísticos, lo que convierte a esta variable en un parámetro esencial del modelo (Valencia, 2023).

Influencia de la distancia

El kilometraje recorrido influye de manera directa en los costos totales. El costo por kilómetro constituye una métrica clave para evaluar la eficiencia operativa y planificar rutas más rentables (Montenegro, 2015).

Costos regulatorios y legales

El transporte de carga pesada requiere permisos especiales, especialmente en cargas sobredimensionadas o peligrosas. Además, las multas por incumplimiento aumentan los costos totales y deben considerarse como variables dentro del modelo predictivo.

Variabilidad temporal

"Los costos logísticos presentan fluctuaciones estacionales debido a factores como el precio del combustible, la demanda o las condiciones climáticas. Incorporar esta variabilidad en el modelo mejora su capacidad de adaptación a escenarios futuros" (Du Plessis et al., 2025).

Costos externos

Los factores exógenos como el clima, el estado de las carreteras o los bloqueos imprevistos impactan directamente en los costos. Estos factores pueden integrarse en el sistema mediante variables proxy, como desviaciones de ruta o consumo adicional, que reflejan su efecto económico.

Relación argumentativa entre los conceptos y variables

La relación entre las herramientas de inteligencia artificial y los costos de transporte de carga pesada se fundamenta en la capacidad de la IA para procesar grandes volúmenes de datos, reconocer inconsistencias y generar estimaciones precisas. La IA permite transformar datos históricos en información estratégica, reduciendo la incertidumbre y mejorando la toma de decisiones.

Urrego Higueta (2024) afirma que los modelos de *machine learning* pueden correlacionar múltiples variables -como distancia, consumo, mantenimiento y condiciones externas- para predecir con mayor exactitud los costos operativos.

La IA posibilita la simulación de escenarios logísticos: por ejemplo, determinar cómo se modificarían los costos si se cambia la ruta, se utiliza un vehículo diferente o varía el precio del combustible. Estas simulaciones fortalecen la planificación financiera y permiten a las empresas anticipar contingencias.

Desde una perspectiva práctica, la integración de modelos de inteligencia artificial en el transporte de carga pesada contribuye al cumplimiento de los principios de la logística 4.0, orientada a la automatización y sostenibilidad. La IA, al optimizar el uso de combustible y las rutas, no solo mejora la eficiencia económica, sino que también reduce el impacto ambiental del sector (Cabornero, 2024).

2.2.2 Proceso de predicción

La predicción, en el contexto de la ciencia de datos y la estadística aplicada, es el proceso mediante el cual se estiman valores futuros de una variable de interés a partir de información histórica y relaciones identificadas entre variables explicativas. Este proceso es de naturaleza probabilística: no garantiza certeza absoluta, sino que ofrece estimaciones con niveles de confianza cuantificables, lo cual permite a los tomadores de decisiones actuar sobre escenarios anticipados con mayor racionalidad (Brownlee, 2019).

El proceso de predicción sigue una secuencia lógica bien definida. En primer lugar, se establece la variable objetivo -aquella que se desea predecir- y se identifican las variables predictoras o características que explican su comportamiento. A continuación, se recopilan y preparan los datos históricos, aplicando técnicas de limpieza, transformación y normalización para garantizar su calidad y consistencia. Posteriormente, se selecciona y entrena el modelo predictivo, ajustando sus

parámetros mediante algoritmos de optimización. Finalmente, se evalúa el rendimiento del modelo con datos no utilizados en el entrenamiento, verificando su capacidad de generalización ante nuevos escenarios (Kingma y Ba, 2015).

En el ámbito logístico y del transporte, el proceso predictivo presenta complejidades adicionales derivadas de la alta dimensionalidad de los datos, la presencia de series temporales con tendencias y estacionalidades, y la influencia de variables externas difícilmente controlables. Por esta razón, los modelos predictivos aplicados en este sector deben ser robustos, adaptables y capaces de capturar tanto patrones lineales como relaciones no lineales entre las variables (Liachovičius et al., 2023).

2.2.3 Predicción de costos de Transporte de Carga Pesada

La predicción de costos en el transporte de carga pesada representa una necesidad estratégica para las empresas del sector, pues permite anticipar el comportamiento financiero de las operaciones logísticas y tomar decisiones informadas sobre rutas, tarifas y asignación de recursos. La naturaleza dinámica de los costos de transporte - influenciados por factores tan diversos como el precio internacional del petróleo, las condiciones climáticas, la demanda estacional y el estado de la infraestructura vial- hace que los métodos tradicionales de estimación basados en promedios históricos resulten insuficientes para capturar la variabilidad real del fenómeno (Urrego, 2024)

Los enfoques predictivos en este dominio han evolucionado desde modelos estadísticos clásicos -como la regresión lineal múltiple y los modelos ARIMA para series temporales- hacia técnicas de aprendizaje automático que ofrecen mayor capacidad de modelado de relaciones complejas. Esta transición ha sido impulsada por la creciente disponibilidad de datos operativos digitalizados, el aumento en la capacidad computacional y la demostrada superioridad de los modelos de *machine learning* en problemas de alta dimensionalidad (Jang et al., 2023).

En el contexto ecuatoriano, la predicción de costos de transporte de carga pesada enfrenta desafíos particulares: la heterogeneidad de la flota vehicular, la variabilidad de las condiciones viales en la región andina y costera, la informalidad de una parte del sector y la escasez de bases de datos estructuradas y actualizadas. Estos factores hacen que el desarrollo de un sistema predictivo adaptado a las condiciones locales

constituya una contribución significativa tanto para la academia como para la industria (MOP, 2023).

La incorporación de variables como el tipo de vehículo, la distancia recorrida, el tipo y peso de la carga, las condiciones de la ruta y los factores externos dentro de un modelo de predicción permite obtener estimaciones de costo más precisas y contextualizadas. Katreddi et al. (2023) demostraron que el uso de inteligencia artificial para predecir costos de mantenimiento -uno de los componentes más difíciles de predecir- reduce el margen de error de estimación en un 12% respecto a los métodos convencionales.

2.2.4 ¿Qué es la inteligencia artificial?

La Inteligencia Artificial es una subdisciplina de la informática, que se ocupa del estudio y diseño de sistemas que son capaces de llevar a cabo funciones que, si son realizadas por ser humano, requieren de inteligencia: razonamiento, aprendizaje, reconocimiento de patrones, toma de decisiones y comprensión en lenguaje natural. El término fue creado por el informático John McCarthy en 1956, en el seno de la Conferencia de Dartmouth, en donde se comenzó a desarrollar el campo de estudio de la IA y donde la disciplina ha mostrado a lo largo del tiempo múltiples ciclos de avances y consolidaciones (Goodfellow et al., 2016).

Desde un punto de vista técnico la IA se fundamenta en tres pilares básicos: el aprendizaje automático (*machine learning*), el aprendizaje profundo (*deep learning*) y el procesamiento del Big data. El aprendizaje automático hace referencia al conjunto de algoritmos que permiten a los sistemas aprender patrones a partir de datos, sin ser explícitamente programados para ello y sin ser necesario especificar de forma categórica el problema a resolver. El aprendizaje profundo es un subcampo del aprendizaje automático que hace uso de arquitecturas de redes neuronales artificiales de múltiples capas con el propósito de modelar representaciones jerárquicas de los datos (Goodfellow et al., 2016).

La IA se ha consolidado como una herramienta estratégica en múltiples industrias: salud, finanzas, manufactura, agricultura y logística, entre otras. Su capacidad para procesar cantidades masivas de datos, identificar patrones ocultos y generar pronósticos de alta precisión la convierte en un recurso indispensable para la toma

de decisiones en entornos complejos y dinámicos. En el sector del transporte, la IA facilita la optimización de rutas, el mantenimiento predictivo, la gestión de flotas y, de manera central para esta investigación, la predicción de costos operativos (Cabornero, 2024).

Es importante distinguir entre los diferentes enfoques dentro de la IA aplicada a problemas de predicción: el aprendizaje supervisado -donde el modelo aprende a partir de ejemplos etiquetados- es el paradigma predominante para tareas de regresión como la estimación de costos; el aprendizaje no supervisado se orienta al descubrimiento de estructuras ocultas en los datos; y el aprendizaje por refuerzo optimiza decisiones secuenciales a través de la interacción con el entorno.

2.2.5 Inteligencia artificial para la predicción de costos de Transporte de Carga Pesada

La aplicación de la Inteligencia Artificial a la predicción de costos de transporte de carga pesada representa uno de los desarrollos más prometedores dentro de la logística computacional. A diferencia de los métodos estadísticos tradicionales, los modelos de IA pueden incorporar simultáneamente un mayor número de variables predictoras, capturar relaciones no lineales entre ellas y adaptarse a la variabilidad temporal inherente a los procesos logísticos. Jang et al. (2023) demostraron que el uso de sistemas basados en *machine learning* puede incrementar la precisión de la predicción de costos logísticos hasta en un 20% en comparación con los métodos estadísticos clásicos.

La selección del algoritmo más adecuado depende de la naturaleza de los datos disponibles, la dimensionalidad del problema y los requerimientos de interpretabilidad del modelo. Urrego Higuera (2024) señala que los modelos de *machine learning* pueden correlacionar múltiples variables -como distancia, consumo de combustible, costos de mantenimiento y condiciones externas- para predecir con mayor exactitud los costos operativos, superando consistentemente a los modelos de regresión lineal convencional cuando las relaciones entre variables son complejas o no lineales.

Entre las herramientas de IA más utilizadas en este dominio destacan las Redes Neuronales Artificiales -especialmente las de tipo LSTM para datos secuenciales- y los

algoritmos de ensemble como *Random Forest*. A continuación, se describen estas herramientas y otras alternativas relevantes.

2.2.5.1 Predicción con redes neuronales

Las Redes Neuronales Artificiales (RNAs) constituyen una de las herramientas más poderosas dentro del aprendizaje profundo. Su arquitectura está inspirada en la estructura del sistema nervioso biológico: se compone de capas de nodos interconectados -neuronas artificiales- que transforman las señales de entrada mediante funciones de activación no lineales para producir una salida. Esta estructura multicapa permite a las redes neuronales aproximar funciones de cualquier complejidad, lo que las hace especialmente adecuadas para problemas donde las relaciones entre variables son altamente no lineales (Goodfellow et al., 2016).

Las arquitecturas LSTM (*Long Short Term Memory*), apropiadas para predecir el costo del transporte, dado que su característica secuencial incluye tendencias y estacionalidad temporal, son las mejores aptas para estas tareas. Las LSTM son una versión de las Redes Neuronales Recurrentes (RNN) que capturan relaciones de larga distancia en datos secuenciales de manera óptima, superando el problema del *vanishing gradient* o del desvanecimiento que sufren las RNN. (Brownlee, 2019) corroboraron su rendimiento para la tarea de predecir series de costos logísticos muy variables a lo largo del tiempo.

El diseño de una red neuronal para esta solución tiene en cuenta: el número de capas ocultas, el número de neuronas por capa oculta, funciones de activación (ReLU, tanh, sigmoide), algoritmo de optimización (Adam, SGD, RMSProp) y función de pérdida (MSE, MAE). El entrenamiento de la red neuronal comienza con el algoritmo de retro propagación, el cual va ajustando los pesos de la red para ir minimizando el error entre las predicciones generadas y los datos reales. Las técnicas de regularización (*dropout* y *early stopping*) se encargan de controlar la capacidad de generalización del modelo para evitar sobre ajustar los datos de entrenamiento (Kingma y Ba, 2015).

La eficacia del modelo se evalúa valiéndose de las métricas estandarizadas: el Error Cuadrático Medio (RMSE), el Error Medio Absoluto (MAE) y el coeficiente de determinación (R²). (Brownlee, 2019), a este respecto, señala que un buen modelo es

aquel que presenta un R^2 por encima de 0.90 y un RMSE fuera de los márgenes estadísticamente aceptados.

2.2.5.2 Predicción con *Random Forest*

Random Forest es un algoritmo de aprendizaje automático supervisado que apunta a la solución de situaciones de aprendizaje a través del paradigma de ensemble learning. Su idea principal resulta la construcción de un gran número de árboles de decisión, cada árbol de decisión es creado sobre una muestra aleatoria del conjunto de datos para cada subconjunto aleatorio de características y finalizar con la combinación de sus predicciones, de tal manera que se extrae una predicción que es más robusta y precisa que la predicción generada por un único árbol de decisión. Esta estrategia de aleatorización doble permite reducir la varianza del modelo y lo hace extremadamente robusto ante el sobreajuste (Breiman, 2001).

En el contexto de la predicción de costos de transporte de carga pesada, *Random Forest* presenta ventajas significativas: puede manejar eficientemente tanto variables numéricas continuas (distancia, consumo de combustible, peso de la carga) como variables categóricas (tipo de vehículo, tipo de ruta, categoría de carga), sin requerir transformaciones previas. Además, proporciona métricas de importancia de características que permiten identificar qué variables tienen mayor influencia sobre el costo predicho, facilitando la interpretación de los resultados y la toma de decisiones operativas.

Desde el punto de vista computacional, *Random Forest* es eficiente y puede entrenarse en paralelo, lo que lo hace adecuado para conjuntos de datos de tamaño moderado a grande. Su rendimiento es comparable al de las redes neuronales en muchos problemas de regresión, con la ventaja adicional de requerir menos ajuste de hiperparámetros. Estudios recientes en logística han reportado que *Random Forest* supera a modelos lineales en la predicción de costos de mantenimiento vehicular y consumo de combustible, especialmente cuando los datos presentan relaciones no lineales y valores atípicos (Katreddi et al., 2023).

Los principales hiperparámetros a ajustar en *Random Forest* son el número de árboles en el bosque, la profundidad máxima de cada árbol, el número mínimo de muestras por nodo hoja y el número de características consideradas en cada división.

2.2.5.3 Otras herramientas de IA para la predicción de costos

Además de las redes neuronales y *Random Forest*, existe un conjunto amplio de herramientas de inteligencia artificial aplicables a la predicción de costos de transporte de carga pesada. Su selección depende de las características del problema, la disponibilidad de datos y los requerimientos de precisión e interpretabilidad.

Gradient Boosting y XGBoost son algoritmos de ensemble que construyen modelos de forma secuencial, donde cada nuevo árbol corrige los errores del modelo anterior. XGBoost (*Extreme Gradient Boosting*) ha demostrado ser especialmente eficaz en competencias de ciencia de datos y aplicaciones industriales gracias a su velocidad de entrenamiento, manejo nativo de valores faltantes y robustez ante datos ruidosos. En el dominio logístico, se ha utilizado con éxito para predecir tiempos de entrega, demanda de carga y costos operativos (Du Plessis et al., 2025).

Support Vector Regression (SVR) es otra alternativa de interés, especialmente útil cuando el conjunto de datos es de tamaño reducido y se requiere una frontera de decisión robusta. SVR trabaja en espacios de características de alta dimensión mediante funciones kernel, lo que le permite capturar relaciones no lineales entre variables. Sin embargo, su escalabilidad a grandes conjuntos de datos es limitada en comparación con los algoritmos basados en árboles.

Los modelos de regresión lineal múltiple y LASSO, aunque más simples, siguen siendo relevantes como líneas base de comparación. Su interpretabilidad y facilidad de implementación los hacen útiles en contextos donde la explicación del modelo es prioritaria. Finalmente, los modelos de series temporales como ARIMA y Prophet son apropiados cuando el análisis se centra exclusivamente en la dimensión temporal de los costos, sin incorporar variables explicativas externas.

2.2.5.4 Software para implementar la inteligencia artificial en predicción de costos de transporte

La implementación de modelos de inteligencia artificial para la predicción de costos de transporte requiere el apoyo de herramientas de software especializadas que faciliten el procesamiento de datos, el entrenamiento de modelos y la visualización

de resultados. En la actualidad, el ecosistema de herramientas disponibles es amplio y, en su mayor parte, de acceso abierto.

Python es el lenguaje de programación dominante en el campo de la inteligencia artificial y el aprendizaje automático. Su ecosistema de bibliotecas incluye: *Pandas* y *NumPy* para la manipulación y análisis de datos tabulares; *Scikit-learn* para la implementación de algoritmos de *machine learning* clásicos, incluyendo *Random Forest*, *SVR* y modelos lineales; *TensorFlow* y *Keras* para el diseño y entrenamiento de redes neuronales profundas y modelos *LSTM*; y *XGBoost* como biblioteca especializada para algoritmos de *gradient boosting*. La combinación de estas bibliotecas permite cubrir todo el ciclo de vida del modelo predictivo, desde la preparación de datos hasta la evaluación y despliegue.

Jupyter Notebook y *JupyterLab* son entornos de desarrollo interactivo ampliamente utilizados para la experimentación y documentación de proyectos de ciencia de datos.

Para la visualización de resultados, *Matplotlib* y *Seaborn* permiten generar gráficos estadísticos de alta calidad, mientras que *Plotly* ofrece visualizaciones interactivas. *Power BI* y *Tableau* son alternativas orientadas a usuarios de negocio que permiten la construcción de Dashboards para el monitoreo continuo de los costos predichos y reales.

En entornos de producción, plataformas de compilación en la nube como *Google Colab*, *Amazon SageMaker* y *Microsoft Azure Machine Learning* ofrecen infraestructura escalable para el entrenamiento y despliegue de modelos de IA. *Google Colab*, en particular, proporciona acceso gratuito a unidades de procesamiento gráfico (GPU), lo que acelera significativamente el entrenamiento de redes neuronales profundas sin requerir infraestructura local especializada (Goodfellow et al., 2016).

2.3 ESTADO AL ARTE

La predicción de costos en el transporte de carga pesada ha sido objeto de un progresivo interés académico en la última década, motivado por el desarrollo de técnicas de inteligencia artificial y la disponibilidad de datos operativos digitalizados. Con el propósito de identificar el estado del conocimiento actual en esta área, se

llevó a cabo una revisión de diez estudios publicados que entienden la predicción de costos logísticos y de transporte a partir de modelos computacionales e inteligencia artificial.

Los estudios objeto de la revisión abarcan distintos contextos de aplicación: transporte de carga pesada, logística farmacéutica, *Smart ports*, demanda de movilidad urbana y seguridad vial. En todos los casos, los autores coinciden en que los modelos basados en IA son más precisos que los métodos estadísticos tradicionales en precisión, poder de generalización y en el manejo simultáneo de múltiples variables. La

Tabla **1** muestra la comparación sistemática de los diez documentos analizados en función de los indicadores de evaluación, limitaciones, resultados cuantitativos, brechas identificadas y principales conclusiones.

Tabla 1.Comparación de estudios relacionados

Autor y Año	Lugar de Implementación	Título de la Investigación	Metodología	Indicadores Principales	Limitaciones	Resultados Cuantitativos	Gaps / Trabajos a Futuro	Principales Conclusiones
Farchi et al. (2023)	Marrakech, Marruecos	A Comparative Study on AI-Based Algorithms for Cost Prediction in Pharmaceutical Transport Logistics	Análisis comparativo de 12 algoritmos de IA (heurísticos, metaheurísticos, optimización)	MAE, MSE, RMSE, R ² , RMLSE, MAPE, Tiempo de Cómputo (CT)	Set de datos pequeño (~300 observaciones); riesgo de sobreajuste; dificultad en captar asociaciones no lineales	El modelo Orthogonal Matching Pursuit (OMP) obtuvo un R ² cercano a 1 y un MAE de 0.0126	Uso de Datasets más grandes y diversos; captura de patrones complejos y manejo de información faltante Extender a múltiples localidades;	OMP es el más adecuado para predicción de costos en sets de datos pequeños y alta dimensionalidad en logística farmacéutica ARIMA es la opción óptima para prever demanda de taxis por su precisión y estabilidad frente a la variabilidad diaria
Chamorro y Sánchez (2026)	Tulcán, Ecuador	Comparación de modelos de predicción para estimar la demanda de viajes en radio taxis	Metodología KDD; comparación de modelos de series temporales: ARIMA, Prophet y Random Forest	MAE, MSE, RMSE	Estudio enfocado en una sola ciudad	ARIMA superó a los demás con: MAE 1.46, MSE 4.71 y RMSE 2.17	Investigar redes neuronales o modelos <i>ensemble</i> ; impacto de factores climáticos	

Autor y Año	Lugar de Implementación	Título de la Investigación	Metodología	Indicadores Principales	Limitaciones	Resultados Cuantitativos	Gaps / Trabajos a Futuro	Principales Conclusiones
Palafox-Palafox et al. (2023)	Pachuca, México (Afilación Univ.)	Modelo de pronóstico de cadena de suministro mediante redes neuronales	Simulación de cadena de suministro (Promodel) y entrenamiento de 110 configuraciones de Redes Neuronales Artificiales (RNA)	Coeficiente de determinación (r^2), RMSE	Los datos provienen de un modelo de simulación y no directamente de operaciones de campo en tiempo real	La mejor RNA obtuvo un $r^2 = 0.99408$ y un RMSE = 1.44217	Aplicar la RNA obtenida en modelos de simulación para analizar satisfacción del cliente y reducción de gastos totales	Las RNA ofrecen mayor certidumbre y precisión (99.4%) que los métodos tradicionales de regresión lineal o promedio móvil
Rodríguez y Ugalde (2021)	San José, Costa Rica	Impacto de la estandarización y escalado: factor para predicción de costos en proyectos a través de una red neuronal artificial	Simulación Monte Carlo + RNA (Backpropagation) probando 4 métodos de escalado/normalización	MSE, RMSE, R	El método Vast no es apto para datos con gran varianza; Pareto es sensible a valores atípicos radicales	El escalado de rango permitió el mejor ajuste (MSE: 0.0016955, RMSE: 0.0399)	Caracterizar los datos antes del procesamiento para elegir el método de escalado adecuado	El preprocesamiento (normalización y escalado) es un factor crítico para mejorar el desempeño de las RNA en predicción de costos

Autor y Año	Lugar de Implementación	Título de la Investigación	Metodología	Indicadores Principales	Limitaciones	Resultados Cuantitativos	Gaps / Trabajos a Futuro	Principales Conclusiones
Vyshka y Bakiasi (2025)	Porto Romano, Albania	Machine Learning-Driven Cost Prediction in Intermodal Logistics: A Case Study of Porto Romano	Comparación de Random Forest (RF) y SVM para estimar costos de transporte intermodal	MSE, Precisión, F1-Score	Los datos del sector ferroviario fueron simulados por ser un desarrollo futuro del puerto	Random Forest superó a SVM con un R^2 de 0.91 y una precisión del 94.1%	Refinar modelos con datos operativos en vivo; expansión hacia metodologías de <i>Deep Learning</i>	El aprendizaje automático, especialmente Random Forest, mejora la planificación operativa y competitividad estratégica de puertos inteligentes
Singh et al. (2021)	Varanasi, India	Prediction of Transportation Costs Using Trapezoidal Neutrosophic Fuzzy Analytic Hierarchy Process and Artificial Neural Networks	Híbrido: TNF-AHP para selección de criterios significativos + RNA (Levenberg-Marquardt)	R (Pearson), MSE, MAPE	Subjetividad en las preferencias de los expertos; selección de topología por prueba y error	$R = 0.999156$ (99.92% correlación), $MSE = 0.000421$, $MAPE = 4.34\%$	Aplicar TNF-AHP a problemas multiobjetivo y de múltiples proveedores en la cadena de suministro	La combinación de lógica difusa neutrosófica y RNA es altamente eficiente para predecir costos bajo ambientes inciertos y dinámicos

Autor y Año	Lugar de Implementación	Título de la Investigación	Metodología	Indicadores Principales	Limitaciones	Resultados Cuantitativos	Gaps / Trabajos a Futuro	Principales Conclusiones
Partheeban et al. (2008)	Chennai, India	Road accident cost prediction model using systems dynamics approach	Enfoque de Dinámica de Sistemas (Software DYNAMO)	Tasa de accidentes, costos públicos y privados (capital humano, sufrimiento)	Restringido a accidentes de autobuses; falta de datos estáticos; factores de costo cambian diariamente	El modelo predijo accidentes con una variación de solo -3.36% a +3.47% respecto a los datos reales	Realizar estudios detallados para áreas urbanas y rurales por separado	La Dinámica de Sistemas permite analizar interacciones complejas de retroalimentación para la toma de decisiones en seguridad vial
Rodríguez G. (2017)	San José, Costa Rica (Revisión)	Redes neuronales artificiales en la predicción de costos de diversos proyectos: Revisión	Revisión bibliográfica de casos de uso de RNA en carreteras, software y sinterizado	MSE, tasa de exactitud	Se limita a estudios existentes; asume relaciones no lineales sin profundizar en la depuración de datos	Carreteras (93% precisión), sinterizado (94.31% precisión), software (mejor que el modelo COCOMO)	Uso de combinaciones de algoritmos como backpropagation con gradiente conjugado mejorado	Las RNA son superiores a los métodos tradicionales de costeo por su capacidad de aprendizaje y manejo de datos inciertos
Xie y Yu (2024)	China (Uso de datos Kaggle)	Logistics Cost Prediction Based on Random Forest Model	Análisis de correlación de Pearson, prueba Chi-cuadrado y comparación de KNN, SVR y RF	MSE, MAE	Número limitado de factores influyentes; aleatoriedad intrínseca de RF	Random Forest obtuvo el menor error (MSE 0.086, MAE 0.207) superando a KNN y SVR	Expandir factores influyentes; probar algoritmos como XGBoost para datos masivos	El peso de la carga es la variable más influyente en el costo de flete; RF permite una interpretabilidad superior

Autor y Año	Lugar de Implementación	Título de la Investigación	Metodología	Indicadores Principales	Limitaciones	Resultados Cuantitativos	Gaps / Trabajos a Futuro	Principales Conclusiones
Farchi et al. (2024)	Casablanca y otras ciudades, Marruecos	Machine Learning Prediction Model: A Case Study of Urban Transport of Medical and Pharmaceutical Products	Evaluación de Random Forest, SVM y Redes Neuronales en transporte urbano	MSE, RMSE, MAE, MAPE, Precisión	Adaptabilidad a otros contextos; distancias limitadas en el estudio (máximo 25 km)	Random Forest alcanzó una precisión de ~85% (rango 80-84.8% en 10 pliegues)	Incorporar más variables; probar el modelo en otros sectores industriales	Random Forest es el modelo óptimo para transporte urbano farmacéutico por su robustez frente a grandes dimensiones y menor error

El análisis de la tabla comparativa permite el descubrimiento de patrones relevantes en el uso de modelos de inteligencia artificial para la predicción de costos logísticos. En cuanto a los modelos que fueron implementados, *Random Forest* estuvo presente en tres de los diez estudios analizados como la alternativa del mejor rendimiento y, en cambio, las Redes Neuronales Artificiales (RNA) son el enfoque más habitual con cuatro apariciones, incluyendo variantes con lógica difusa neutrosófica. ARIMA es la mejor solución para las series temporales de demanda de movilidad y los modelos de regresión dispersa como OMP son los mejores para *datasets* de alta dimensionalidad con pocas observaciones.

Y respecto a cada una de las métricas de evaluación, el Error Cuadrático Medio (MSE), el Error Medio Absoluto (MAE) y la Raíz del Error Cuadrático Medio (RMSE) son las métricas que más se repiten, con seis de diez estudios, y el coeficiente de determinación R^2 se emplea en cinco de los estudios analizados como medida de ajuste del modelo, siendo la precisión porcentual (accuracy) presente en estudios orientados a clasificación cortando el horizonte de la logística.

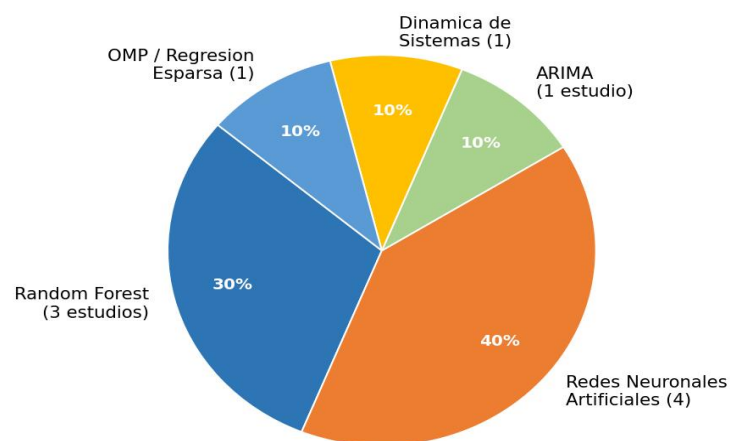


Figura 1. Distribución de modelos de IA utilizados en los estudios revisados

El contenido de la Figura 1, muestra la distribución de modelos de IA en los estudios revisados, el 40% de las investigaciones usaron las redes neuronales artificiales siendo el modelo más utilizado, seguido por el *Random Forest* con un 30% de participación, lo que evidencia que son los modelos más utilizados por los investigadores en los estudios analizados.

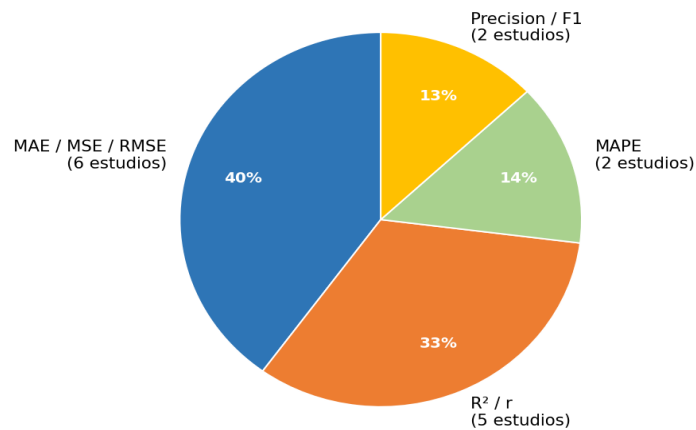


Figura 2. Métricas de evaluación empleadas en los estudios revisados

El contenido de la Figura 2, muestra las métricas de evaluación más utilizadas en las investigaciones, el R^2 o mejor conocido como el coeficiente de determinación es el más utilizado para validar los modelos de inteligencia artificial porque indica la precisión que tuvo el modelo en base a los datos reales e igualmente seguido por los complementos MAE (Error absoluto medio) y el MSE (Error cuadrático medio) que sirven para evaluar los errores del modelo respecto a las predicciones y su desviación promedio lineal.

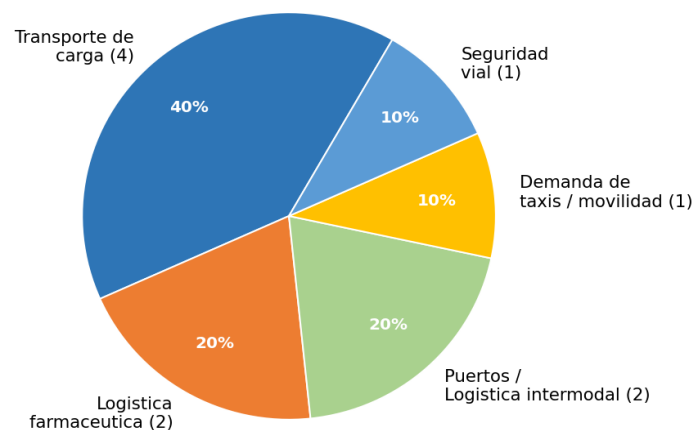


Figura 3. Ámbitos de aplicación de los estudios revisados

El contenido de la Figura 3, muestra en que ámbitos se han aplicado los estudios, teniendo como resultado en un 40% el Transporte de carga, de ahí la importancia

que tiene respecto a este campo, aunque el porcentaje es más alto en Transporte de carga estos modelos de inteligencia artificial se pueden aplicar en cualquier ámbito o estudio, es realmente una herramienta muy flexible y que se debería utilizar con mayor frecuencia en el mundo laboral e investigativo.

Entre las brechas (gaps) identificadas de manera transversal en los estudios revisados se destacan tres: primero, la mayoría de los trabajos operan con *Datasets* de tamaño reducido (entre 300 y 1000 observaciones), lo que limita la capacidad de generalización de los modelos; segundo, escasos estudios incorporan variables externas dinámicas como el precio del combustible por periodo, las condiciones climáticas o el estado de la infraestructura vial, factores que tienen alta incidencia en el costo real del transporte de carga pesada; y el tercero, la mayor parte de los trabajos se concentran en contextos europeos o asiáticos, con escasa representación de países de América Latina, lo que genera un vacío en el conocimiento sobre las condiciones operativas particulares de la región, como la informalidad del sector, la heterogeneidad de la flota o la variabilidad de las rutas.

Las Redes Neuronales Artificiales (RNA) y el algoritmo *Random Forest*, son los modelos más usados en la literatura actual para la predicción de costos logísticos, gracias a su evidente potencial frente a los métodos tradicionales como regresión lineal o promedio móvil. Por las evidencias comparativas determinan que el *Random Forest* es el modelo más correcto para gestionar los costos de transporte y de logística intermodal, siendo el más robusto, el que presenta menor error y la mayor exactitud (de 94.1%) frente al SVM o el RNA de tipo estándar. Mientras que a la hora de realizar la predicción de niveles de inventarios en un *supply chain* o demanda de viajes con series temporales las RNA o ARIMA alcanzan niveles de certidumbre extrema en coeficientes de determinación superiores al 99%. Por lo tanto, para una tesis sobre transporte y logística el *Random Forest* es la mejor alternativa, pues combina bien los aspectos de exactitud, interpretabilidad y gestión de elevadas dimensionalidades sin que tengan un sobreajuste lo que explica su correcta utilización como el algoritmo central en el desarrollo del sistema predictivo de esta investigación.

A partir de la información extraída en este estado del arte, en este trabajo, hemos decidido hacer uso del algoritmo *Random Forest* como herramienta del sistema de predicción que ha de guiar la predicción de costos para el transporte de carga pesada, decisión que está sustentada en tres argumentos derivados directamente

de la literatura revisada, en primer lugar, porque *Random Forest* ha demostrado ser el modelo que mejor rendimiento ha mostrado en los estudios más cercanos a este trabajo, logrando una precisión en la predicción de costos de flete de hasta un 94.1% y el menor error en la predicción de los costos de flete (Estudios 5 y 9), segundo porque el algoritmo tiene una interpretabilidad superior a las redes neuronales, debido a que automáticamente es capaz de presentar una clasificación de la importancia de las variables que explica qué factores operativos (distancia recorrida, consumo de combustible o tipo de carga por ejemplo) tienen mayor incidencia en el costo final del viaje y por último porque el algoritmo *Random Forest* es robusto ante *datasets* de tamaño intermedio como el *dataset* que se está usando en este trabajo (144 registros reales de viajes), ya que no requiere para realizar estimaciones un gran volumen de datos, condición que coincide con la disponibilidad de información del sector de transporte de carga pesada en el Ecuador.

III. METODOLOGÍA

3.1. ENFOQUE METODOLÓGICO

3.1.1. Enfoque

La presente investigación se sustentó en el enfoque mixto, el cual permite usar estrategias cualitativas y cuantitativas, de modo que busca, (a) la objetividad de los datos numéricos y (b) la riqueza interpretativa de la información contextual. Este enfoque sigue unos axiomas del enfoque pragmático, es decir, que emplea los métodos que mejor respondan a las interrogantes de investigación, generando meta-inferencias óptimas que con el uso de un único método. (Hernández-Sampieri y Mendoza, 2018).

En la investigación que aquí se presenta y cuyo objetivo general es diseñar un sistema de predicción del costo del transporte de carga pesada mediante inteligencia artificial, el enfoque mixto tiene un especial protagonismo. La parte cuantitativa se centrará en recoger datos históricos-operativos (kilometraje, consumo de combustible, mantenimiento, peajes), que serán la base de la red neuronal y su evaluación también se hará con indicadores estadísticos; por su parte, la dimensión cualitativa sirve para producir información de los actores del transporte como conductores y gestores logísticos mediante entrevistas y observación, con la intención de buscar aquellos factores contextuales como el estado de las vías, las condiciones atmosféricas, prácticas de operación que inciden sobre los costos y que no siempre pueden verse reflejados en las bases de datos (Ñaupas et al., 2018).

La incorporación de ambos aspectos permitió no solo la mejora de las predicciones numéricas de costos, sino que también ofrecerá explicaciones de cuáles son las variaciones y enriquecerá la utilidad práctica del sistema (Rojas, 2011).

3.1.2. Tipo de Investigación

La investigación descriptiva se destina a definir las características más importantes de cualquier fenómeno que se analice. Tal y como indican Hernández-Sampieri y

Mendoza (2018), "la investigación descriptiva busca especificar propiedades, características y rasgos importantes para cualquier fenómeno que se analice" (p. 92). En este caso, la investigación descriptiva sirvió para poder especificar la composición de aquellos componentes que integran los costos de transporte de carga pesada, tales como los costos del combustible, del mantenimiento, los peajes, los salarios, con el objetivo de contar con un marco más completo antes de la construcción del modelo predictivo con herramientas de inteligencia artificial y realizar predicciones sobre los costos asociados al transporte de carga pesada.

La investigación documental, por su parte, implica la consulta de distintas fuentes secundarias (libros, artículos científicos, informes técnicos, bases de dato especializadas). La investigación documental en este caso tiene por objetivo proveer el soporte teórico de una investigación (Ñaupas et al., 2018) y será utilizada para: (a) recopilación de antecedentes en relación al uso de herramientas de inteligencia artificial para la predicción de costos, así como (b) posteriores experiencias en internacionalismo y regionalismo que se puedan tomar como referencia para el soporte del sistema propuesto.

Por último, la investigación correlacional tiene como objetivo medir el grado de asociación existente entre varias variables, sin tratar de asumir relaciones de causalidad inmediata. Para Hernández-Sampieri y Mendoza (2018), esta investigación, "tienen como objetivo conocer la relación o grado de asociación que existe entre dos o más conceptos, categorías o variables en un contexto específico" (p. 94). En el marco de la investigación que se presenta en este documento para identificar las correlaciones que existen entre las variables como el kilometraje recorrido, el rendimiento del combustible y el estado de las rutas con los costos financieros totales del transporte. Estas correlaciones alimentarán el sistema de predicción y posteriormente permitirán validar la pertinencia de las variables seleccionadas para el modelo de inteligencia artificial.

3.1.3 Caso de estudio

El caso de estudio de la presente investigación se desarrolla en el contexto del transporte de carga pesada terrestre en la ciudad de Tulcán, ubicada al norte del país y considerada un punto estratégico debido a su cercanía con la frontera entre Ecuador y Colombia. Esta localización convierte a la ciudad en un eje importante

para la movilización de mercancías hacia diferentes provincias del país, especialmente hacia destinos como Quito, Ibarra, Santo Domingo, Latacunga, Guayaquil y Manta.

El transporte de carga pesada en esta zona presenta una serie de características operativas que influyen directamente en los costos logísticos. Entre los principales factores se encuentran las largas distancias recorridas, la variabilidad en el precio de combustible, el desgaste del vehículo debido a las condiciones de la vía, los costos de mantenimiento, el pago de peajes, la depreciación de la unidad y los tiempos de viaje asociados a la congestión vehicular y condiciones climáticas.

En este estudio se considerarán rutas frecuentes de transporte de mercancías que parten desde Tulcán hacia otras ciudades del país. Estas rutas presentan diferencias de longitud, tiempo de recorrido, topografía y nivel de demanda, factores que generan variaciones en el costo total del transporte. Por esta razón, resulta necesario crear un sistema de predicción que permita predecir de forma más precisa los costos operativos asociados a cada viaje.

Debido a las limitaciones para acceder a registros empresariales reales y completos, la investigación empleará una base de datos compuesta por información histórica obtenida de tesis, estudios técnicos, tarifas referenciales y costos promedio del sector de transporte de carga pesada. Esta información será complementada mediante datos simulados basados en valores reales de operación, tales como kilometraje recorrido, consumo de combustible, costos de mantenimiento, peajes, salarios y gastos administrativos.

La construcción de este conjunto de datos permitirá entrenar y evaluar el modelo predictivo basado en inteligencia artificial, utilizando variables representativas del contexto operativo del transporte de carga pesada en la provincia del Carchi

A continuación, en la figura 4 se presenta el proceso general del transporte de carga pesada considerado dentro del caso de estudio, desde la emisión de la orden de carga hasta la entrega final de mercancía en el destino correspondiente.

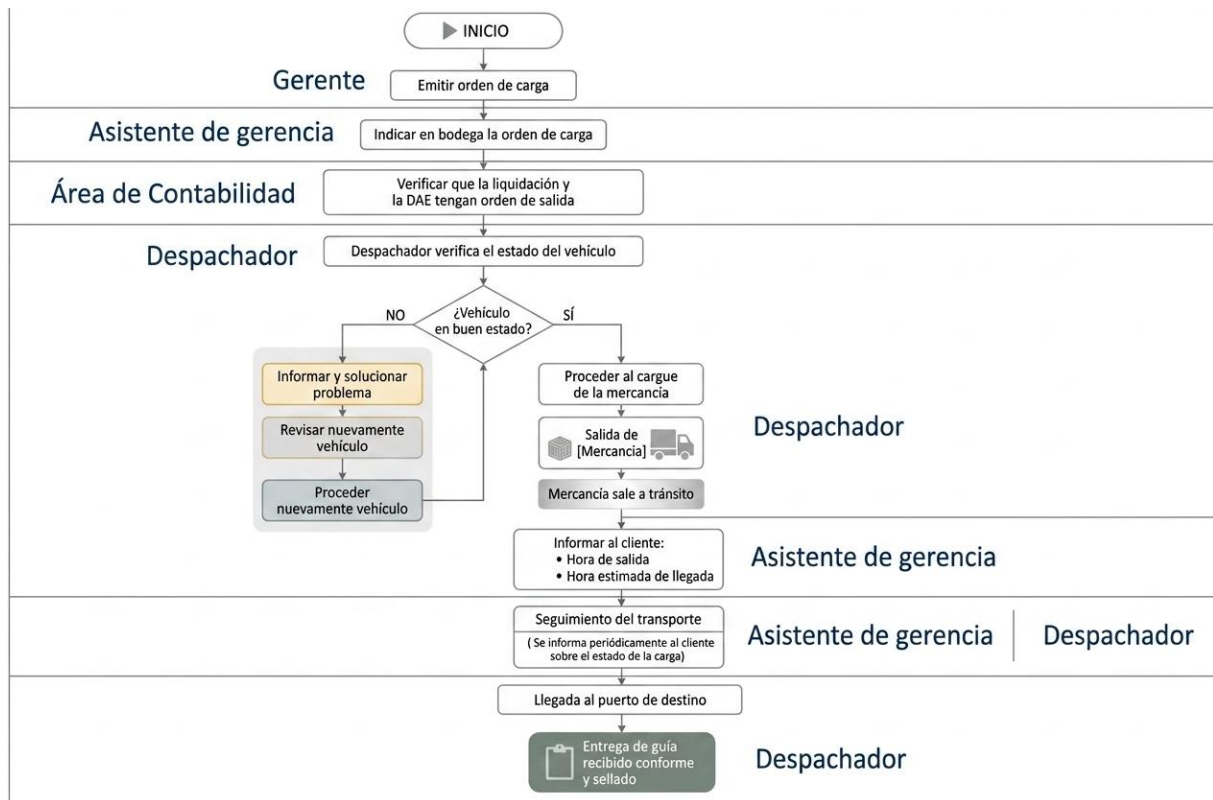


Figura 4. Diagrama del proceso de transporte

Análisis comparativo de modelos y justificación metodológica

Durante la fase inicial del diseño metodológico del sistema predictivo que se generó, se llevó a cabo de manera teórica una valoración de algunas arquitecturas de Redes Neuronales Artificiales (RNA): Perceptrones Multicapa (MLP) para datos estáticos; Redes Neuronales Recurrentes (RNN/LSTM) para componentes secuenciales y Redes Neuronales Convolucionales (CNN) para reconocimiento de patrones espaciales. Esto fue realizado con el objetivo de evaluar su operatividad válida para la estimación de costos de transporte (Goodfellow et al., 2016). No obstante, tras dicho análisis las arquitecturas fueron descartadas por dos limitaciones estructurales que las invalidan para la presente investigación.

La primera limitación es de naturaleza muestral. Las redes neuronales profundas requieren grandes volúmenes de datos para facilitar unos procesos de entrenamiento estables y evitar el sobreajuste. Dado que en la presente investigación se trabajó con un *dataset* tabular de 327 registros de viajes reales, la realización de modelos

neuronales complejos comprometería la capacidad de generalización del sistema durante la fase de validación, llevando al algoritmo a memorizar el ruido estadístico de la muestra en vez de captar los patrones operativos.

La segunda limitación es de orden interpretativo. El área del transporte de carga pesada requiere que la recolección de costos sea trazable y explicable. Las RNA, sin embargo, trabajan la información con redes de múltiples capas que operan en un modo que es opaco; además, no se puede evidenciar cuál es la contribución de las variables operativas relevantes, la distancia recorrida, consumo de combustible y tipo de carga al costo total del viaje.

Además, la literatura científica ha señalado de manera sistemática que los algoritmos de ensambles de árboles de decisión son más estables para conjuntos de datos tabulares de tamaño moderado, sin requerir infraestructuras de computación potentes ni grandes volúmenes de datos para su convergencia.

Estas consideraciones metodológicas motivaron el experimento a 3 modelos de aprendizaje automático que encajaran mejor con la naturaleza del problema usando: *Random Forest Regressor*, *Gradient Boosting* y *Multiple Linear Regression*. De los 3, el algoritmo *Random Forest* demostró mejor rendimiento predictivo alcanzando un MAPE de validación externa de 0,19%, lo que indica un elevado rendimiento predictivo y valida la elección del mismo como motor de predicción del sistema.

3.2. HIPÓTESIS

Hipótesis alternativa:

H1: Los sistemas de predicción de costos basados en inteligencia artificial permiten estimar con mayor precisión los costos de transporte de carga pesada

Hipótesis nula:

H0: Los sistemas de predicción de costos basados en inteligencia artificial no permiten estimar con mayor precisión los costos de transporte de carga pesada.

3.3. DEFINICIÓN Y OPERACIONALIZACIÓN DE LAS VARIABLES

Variable independiente: Herramientas de inteligencia artificial

De acuerdo a la literatura científica actual que se aplica a la optimización logística, las herramientas de Inteligencia Artificial, alimentadas por algoritmos de aprendizaje automático (*Machine Learning*), permiten procesar grandes volúmenes de datos históricos para generar estimaciones financieras con alto grado de certidumbre. En esta línea, el trabajo de Hee-Seon et al. (2023), publicado en la revista *Sustainability*, muestra que los modelos de ensamble como *Random Forest Regressor* superan con creces las técnicas tradicionales y las profundas redes neuronales, cuando se trabaja sobre conjuntos de datos tabulares de tamaño moderado controlando el problema del overfitting de forma nativa. Con este enfoque, la variable independiente se define en términos operacionales como la arquitectura algorítmica, el procesamiento de datos y la calibración de hiperparámetros (número de estimadores y profundidad de los árboles de decisión), y actuaría como el motor tecnológico encargado de captar y modelizar las relaciones no lineales entre las operaciones de la flota y sus costos asociados.

Variable dependiente: Costos de transporte

Según Grúas y Equipos García (2025) "El peso y el tamaño de la carga influyen directamente en la elección del vehículo adecuado y el equipo transporte. Cuando una carga excede los límites convencionales necesita un tratamiento personalizado, por ejemplo, remolques especializados con plataformas modulares, extensibles o de cama baja lo que incrementa el costo" (p.13). Bajo este contexto podríamos afirmar que los costos de transporte dependen de varios factores, factores como la distancia, el peso, el costo de combustible, la naturaleza de la carga, etc. El análisis de estos factores permiten saber que se tiene en cuenta al momento de generar costos y también es un pilar importante para el estudio predictivo.

Tabla 2. Operacionalización de la variable independiente

Variable	Dimensión	Indicador	Ítem	Técnica	Instrumento
Herramientas de inteligencia artificial	Preparación de datos	Tipo de datos utilizados	¿Qué variables operativas se utilizaron para entrenar el modelo predictivo?	Análisis documental	Dataset de entrenamiento
		Preprocesamiento de datos	¿Qué técnicas de limpieza y transformación de datos se aplicaron?	Análisis técnico	Informe de preparación de datos
	Entrenamiento del modelo	Algoritmo de aprendizaje automático	¿Qué algoritmo fue implementado para la predicción de costos?	Evaluación técnica	Registro de entrenamiento
		Configuración del modelo	¿Qué parámetros fueron definidos para optimizar el modelo Random Forest?	Revisión técnica	Configuración del sistema
	Evaluación del modelo	Importancia de variables	¿Qué variables presentan mayor influencia en la predicción del costo?	Análisis estadístico	Reporte de variables relevantes
		Métricas de desempeño	¿Qué nivel de precisión obtuvo el modelo predictivo?	Evaluación cuantitativa	Resultados del modelo
		Capacidad predictiva	¿Qué tan precisas son las predicciones frente a costos reales?	Simulación	Informe de validación
		Generalización del modelo	¿El modelo mantiene precisión con nuevos datos?	Validación cruzada	Reporte comparativo
	Aplicación en costos de transporte	Estimación de costos	¿Qué tan eficiente es el modelo en la	Comparación histórica	Resultados de simulación

Variable	Dimensión	Indicador	Ítem	Técnica	Instrumento
			estimación de costos de transporte?		
		Apoyo a la toma de decisiones	¿El sistema mejora la planificación logística y financiera?	Encuesta	Cuestionario

Tabla 3. Operacionalización de la variable dependiente

Variable	Dimensión	Indicador	Ítem	Técnica	Instrumento
		Costo por combustible	¿Cuál es el costo promedio por litro de combustible utilizado en las rutas?	Revisión documental	Ficha documental
	Costos directos	Costo por peajes	¿Cuánto se paga por peajes en rutas frecuentes?	Revisión documental	Ficha documental
		Sueldos y beneficios de conductores	¿Cuál es el costo mensual en salarios y beneficios?	Revisión documental	Ficha documental
		Mantenimiento vehicular	¿Cuánto se invierte en mantenimiento periódico y correctivo?	Observación	Ficha de observación
	Costos indirectos	Seguros	¿Qué costos implica asegurar vehículos y carga?	Revisión documental	Ficha documental

Variable	Dimensión	Indicador	Ítem	Técnica	Instrumento
Costos de transporte de carga pesada	Costos operativos	Costos administrativos	¿Qué parte del costo total corresponde a administración logística y planificación?	Observación	Ficha de observación
		Costo por uso de infraestructura (terminales, almacenes)	¿Se paga por uso de patios, estaciones o bodegas temporales?	Observación	Ficha de observación
		Equipamiento de carga y descarga	¿Qué costos representan los equipos usados para manipular la carga?	Observación	Ficha de observación
		Peso total transportado	¿Cuál es el peso típico de la carga transportada?	Ficha técnica	Guía de carga
	Costo según la carga	Volumen y dimensiones	¿Cuál es el volumen promedio de la carga pesada?	Medición técnica	Documentación logística
	Tipo de vehículo	Tipo de carga (peligrosa, frágil, especial)	¿El tipo de carga requiere condiciones especiales que elevan el costo?	Clasificación documental	Registro de tipos de carga
		Capacidad del camión	¿Qué tipo de vehículo se requiere según el volumen o peso?	Registro técnico	Hoja de especificaciones
		Consumo de combustible según vehículo	¿Qué modelo de vehículo consume más combustible?	Comparación técnica	Registro de flota

Variable	Dimensión	Indicador	Ítem	Técnica	Instrumento
	Influencia de la distancia	Kilometraje recorrido	total ¿Cuántos kilómetros se recorren por cada envío?	Análisis cuantitativo	Historial de rutas
		Costo por kilómetro	¿Cuál es el costo total por kilómetro recorrido?	Análisis financiero	Base de datos contable
	Costos regulatorios y legales	Permisos de transporte especial	¿Qué costos implican los permisos para carga sobredimensionada o peligrosa?	Revisión legal	Documentos del Ministerio de Transporte
		Multas o sanciones	¿Cuál es el promedio de costos por incumplimientos?	Análisis documental	Registro de infracciones
	Variabilidad temporal	Estacionalidad variación mensual	o ¿Cómo cambian los costos a lo largo del año?	Análisis documental	Reportes mensuales
	Costos externos	Impacto del clima condiciones externas	o ¿El mal clima o bloqueos de carretera elevan el costo?	Estudio de caso	Observación

3.4. MÉTODOS UTILIZADOS

Análisis del tema de investigación y las variables planteadas

El tema de estudio plantea la combinación de la inteligencia artificial con costos de transporte de carga pesada, un sector de alta complejidad operativa y económica.

La variable independiente, que hemos denominado herramienta de inteligencia artificial, se compone del conjunto de técnicas y algoritmos que se pueden agrupar dentro del aprendizaje automático (*machine learning*) y todos ellos permiten elaborar sistemas predictivos. La variable dependiente, los costos de transporte de carga pesada, es un término que se refiere a los gastos que requieren los procesos logísticos: combustible, mantenimiento, peajes, salarios, seguros y factores externos que afectan a la eficacia operativa (Valencia, 2023).

El estudio de estas variables intenta responder a la pregunta de cómo serían aplicables las herramientas de IA para modelar, predecir y optimizar los costos de transporte de cargas, que permitirían la construcción de una base empírica y teórica para elaborar un sistema inteligente ajustado a las condiciones de mercado del Ecuador.

Estructura del modelo

Los algoritmos de aprendizaje por ensambles de árboles de decisión son, en la actualidad, herramientas muy potentes dentro de la inteligencia artificial para el análisis de datos tabulares (Breiman, 2001). A diferencia de otros modelos de aprendizaje profundo con menor interpretabilidad, *Random Forest* realiza su procesamiento a partir de muchos árboles de decisión independientes cuyas predicciones individuales se combinan por medio del promedio de forma tal que acaban generando una estimación final.

Este enfoque permite un proceso que resulta adecuado, estable ante datos escolásticos y resistente al sobreajuste. En el caso concreto del transporte de carga pesada, los modelos de ensamble son particularmente apropiados, ya que no requieren que se cumplan supuestos de linealidad y son capaces múltiples interacciones mecánicas complejas como la del peso transportado o la de la pérdida de inercia en paradas inmediatas. La configuración del algoritmo implica ajustar los

hiperparámetros necesarios: el número de estimadores y la profundidad máxima de los árboles, ya que esto afecta directamente a la capacidad de generalización del modelo y a la precisión del mismo en el procesamiento simultáneo de los múltiples factores operativos que determinan el costo logístico.

Preparación de datos

La preparación de datos constituye otra de las etapas clave en cualquier modelo de IA, y supone la adquisición, limpieza, transformación y normalización de la información con el fin de validarla. En el área del transporte, los datos almacenados incluyen los registros históricos de kilómetros recorridos, los gastos derivados del consumo de combustible, los costos de mantenimiento, de peajes, de seguros de las condiciones ambientales.

Durante la fase del entrenamiento el algoritmo construye múltiples árboles de decisión a partir de subconjuntos aleatorios de datos. Con el fin de desarrollar dicha investigación, el preprocesado se centró específicamente en la consolidación de un *dataset* procedente de las 327 estancias reales de los viajes. Es importante tener en cuenta que la detección y el tratamiento de los valores atípicos, así como la adecuada estructuración tabular de las variables predictoras son condiciones necesarias para erradicar cualquier distorsión que pudiera afectar el modelo. Una base de datos bien depurada y correctamente preprocesada es condición necesaria para asegurar la estabilidad del modelo y la fiabilidad de las estimaciones financieras resultantes.

Entrenamiento del modelo

A diferencia de las redes neuronales, las cuales se basan en optimizadores de gradiente y del algoritmo de retropropagación (Kingma & Ba, 2015), *Random Forest* usa la técnica estadística del *Bagging* (*Bootstrap Aggregating*) la cual consiste en: primero se generan varios subconjuntos aleatorios de los datos de entrenamiento y luego para cada subconjunto se genera un árbol de decisión independiente, para posteriormente promediar el pronóstico de todos ellos, lo cual tiene como consecuencia una clara reducción en la varianza del modelo y una mitigación nativa del sobreajuste (*overfitting*), sin necesidad de hacer correcciones adicionales.

El rendimiento del entrenamiento se mide a través de métricas estadísticas estandarizadas como el Error Cuadrático Medio (RMSE), Error Medio Absoluto (MAE) y el coeficiente de determinación (R^2) (Brownlee, 2019), que además permiten comprobar la evolución del algoritmo hasta alcanzar el final de la ejecución del mismo y garantizar que las proyecciones de los costos queden dentro de márgenes estadísticamente aceptables.

Evaluación del modelo

Una vez finalizado el entrenamiento, el modelo debe ser evaluado cuidadosamente utilizando una técnica de partición *Hold-Out* del 80/20, con el fin de comprobar su capacidad para generalizar adecuadamente ante los datos que no ha visto antes. Liachovičius et al. (2023) destacan que cuando se trabaja con series de tiempo y datos secuenciales como los costos mensuales o anuales, la validación temporal resulta ser la opción más apropiada.

Comparar las predicciones del modelo con los costos reales permite medir su nivel de precisión y fiabilidad. Además, los informes obtenidos durante el proceso de validación sirven como guía para documentar el rendimiento del modelo y realizar los ajustes necesarios en sus hiperparámetros, buscando siempre mejorar su desempeño.

La comparación entre las predicciones y los costos reales permitirá medir la precisión del sistema, mientras que los informes de validación servirán para documentar el desempeño del modelo y ajustar los hiperparámetros según los resultados obtenidos.

Aplicación en la predicción de costos

Una vez comprobado su rendimiento, el modelo neuronal puede implementarse para generar pronósticos automáticos de costos. Jang et al. (2023) demostraron que el uso de sistemas basados en *machine learning* puede incrementar la precisión de la predicción de costos logísticos hasta en un 20% en comparación con los métodos estadísticos tradicionales.

La inteligencia artificial no solo funciona como una herramienta para hacer predicciones, sino también como apoyo estratégico para la toma de decisiones. En este estudio, su aplicación permitirá analizar con mayor precisión la eficiencia

financiera y operativa, optimizando la gestión del transporte y la asignación de recursos.

3.4.1 METODOLOGIA PROPUESTA PARA EL PRONÓSTICO DE COSTOS EN TRANSPORTE DE CARGA PESADA

La metodología presentada en este trabajo de investigación es una descripción del proceso estructurado que se sigue para el diseño, desarrollo y validación del sistema de predicción de los costos de transporte de carga pesada basado en el algoritmo *Random Forest*. Dado que no se tiene acceso directo a los sistemas de información de una empresa transportista, la base de datos de entrenamiento se obtuvo mediante simulación computacional, y tuvo como parámetros base los costos unitarios reales que fueron validados por la (ANT) y la estructura de costos del sector, que fue documentada a través del resultado del objetivo 1. Esta metodología, a la que denominamos simulación basada en parámetros reales, es común en las investigaciones de *machine learning* aplicado donde se tiene un acceso severamente restringido a los datos operativos de campo (Breiman, 2001).

3.4.1.1 Herramientas utilizadas

La aplicación del sistema predictivo exige una batería de herramientas de *software* que dan soporte al ciclo completo del proceso, desde la generación del *dataset* simulado hasta la presentación interactiva de resultados al usuario final; la elección de cada una de las herramientas está en función de criterios de accesibilidad, gratuidad, estandarización académica y adecuación técnica para el contexto de la presente investigación. En la Tabla 4 se podrá consultar dicho detalle de herramientas por fase.

Tabla 4. Herramientas utilizadas por cada fase.

Fase	Herramienta / Software	Función principal	Justificación
F1 Recopilación	Python — numpy / random	Generación del dataset sintético de 327 registros con parámetros basados en costos reales de la ANT	Permite simular escenarios operativos con variabilidad controlada cuando no se dispone de datos empresariales

Fase	Herramienta / Software	Función principal	Justificación
F1 Recopilación	Matriz ANT	Parámetros base: USD 0.415/km combustible, USD 0.158/km mantenimiento, USD 515.86/viajes operativos	Costos unitarios reales que dotan de validez empírica al dataset sintético generado
F2 Procesamiento	Python — pandas / numpy	Limpieza, cálculo de costos por viaje, cálculo de variables de salida y codificación de variables categóricas	Bibliotecas estándar para ETL en ciencia de datos; ampliamente documentadas y reproducibles
F2 Procesamiento	Google Colab	Entorno de ejecución del código Python; exportación del dataset procesado en formato CSV	Accesible sin infraestructura local; facilita la reproducibilidad del proceso
F3 Predicción	Python — scikit-learn (RandomForestRegressor)	Entrenamiento del modelo Random Forest con 100 árboles; división 80/20; predicción de costo total y costo por km	Biblioteca de referencia en machine learning; implementación estandarizada y documentada del algoritmo RF
F3 Predicción	matplotlib / seaborn	Visualización de importancia de variables y comparación predicciones vs valores reales	Generación de gráficos de alta calidad para el informe académico
F4 Resultados	Informe académico (Word / PDF)	Documentación formal de métricas MAE, MSE, RMSE y R2; análisis de importancia de variables; conclusiones	Formato estándar de entrega académica requerido por la institución
F4 Resultados	Página web interactiva	Interfaz de consulta: el usuario ingresa parámetros del viaje y obtiene el costo estimado en tiempo real	Hace el sistema predictivo accesible a empresas de transporte sin conocimientos de programación

3.4.1.2 Fase 1: Recopilación y generación de datos

La primera fase tiene como finalidad generar el *dataset* que será el insumo de entrenamiento del modelo predictivo. Como el acceso a los registros operativos reales de las empresas de transporte de carga pesada en Ecuador es de tipo restringido se elige la generación de un *dataset* sintético mediante simulación

computacional en *Python* haciendo uso del paquete *numpy* para completar los valores aleatorios dentro de los rangos operativos reales.

Los parámetros de simulación se extraen directamente de las fuentes empíricas documentadas en el resultado del objetivo 1 y de la información de la ANT: distancias entre 482 km y 1,505 km, correspondientes a las rutas operativas identificadas; tres destinos principales (Guayaquil, Manta y Quito) con sus frecuencias relativas reales; costos del combustible de USD 0.415 por km; costos de mantenimiento de USD 0.158 por km; y costos de explotación fijos de USD 515.86 por viaje. El *dataset* desarrollado contiene 327 registros de simulación, tal cantidad es suficiente para el robusto entrenamiento de un modelo de *Random Forest* y claramente superior a la de los estudios comparables en la literatura revisada.

Todos los registros que se han simulado incluyen las variables que se enumeran a continuación: distancia recorrida (km); ciudad destino; tipo de carga; mes del año; placa del vehículo (tres categorías); flete estimado (USD); un componente de variabilidad aleatoria controlada que puede emular todos los aspectos del entorno operativo real (variaciones en el consumo de combustible y/o variaciones en el tiempo de operación); este componente de variabilidad es esencial para que el modelo logre aprender a generalizar y no simplemente se limite a memorizar una fórmula de costos fija.

3.4.1.3 Fase 2: Procesamiento de datos

La segunda fase comprende la transformación del *dataset* generado en la fase anterior en un conjunto estructurado y listo para el entrenamiento del modelo. Esta etapa se ejecuta en *Python* mediante las bibliotecas *pandas* y *numpy* dentro del entorno Google Colab y comprende cuatro actividades principales:

- Verificación de integridad: revisión de valores nulos, rangos fuera de límites y consistencia entre variables de distancia y costo.
- Cálculo de costos por viaje: aplicación de los costos unitarios reales a cada registro simulado para obtener el costo de combustible, el costo de mantenimiento y el costo operativo fijo por viaje, sumados para obtener el costo total del viaje en USD.

- Cálculo de variables de salida: el costo total del viaje en USD y el costo por kilómetro en USD/km, obtenido como cociente entre el costo total y la distancia recorrida.
- Codificación de variables categóricas: conversión de las variables destino y tipo de carga a valores numéricos enteros, requerimiento técnico del algoritmo *Random Forest* para el procesamiento de variables no numéricas.

El resultado de esta fase es un *dataset* con las variables de entradas necesarias, exportado en formato CSV para su uso directo en la fase de pronóstico.

3.4.1.4 Fase 3: Pronóstico con *Random Forest*

La tercera fase hace el núcleo del sistema predictivo y consiste en la implementación, entrenamiento y evaluación del modelo *Random Forest*. El algoritmo se ejecuta en *Python*, empleando la biblioteca *scikit-learn*, dentro del entorno Google Colab. Se entrenan dos modelos de forma independiente: el primero va orientado en predecir el total del viaje en USD y el segundo va orientado en predecir el costo por km en USD/km, permitiendo optimizar cada modelo para cada variable objetivo.

El proceso de pronóstico sigue la siguiente secuencia:

- División del *dataset*: se divide el conjunto de registros en un 80% correspondientes al entrenamiento y un 20% para la prueba, con semilla fija (*random_state=42*) para que dé lugar a reproducibilidad.
- Entrenamiento: el algoritmo va construyendo 100 árboles de decisión de forma paralela, cada uno sobre una muestra aleatoria con reemplazo del conjunto de entrenamiento, evaluando en cada nodo de división aleatoriamente un subconjunto de las 8 variables. La predicción definitiva es la media de los 100 árboles, por lo que también se reduce la varianza del modelo, logrando así evitar el sobreajuste.
- Ajuste de hiperparámetros: *n_estimators=100*, *max_features=sqrt*, *max_depth=None*, *random_state=42*, valores que equilibran precisión y tiempo de cómputo respecto al tamaño del *dataset*.

- Generación del ranking de importancia de variables: el modelo cuantifica la contribución relativa de cada variable de entrada sobre la reducción del error de predicción, permitiendo identificar los factores de mayor incidencia sobre el costo de transporte.

El software utilizado es *Python* con *scikit-learn*, *pandas*, *numpy* y *matplotlib*, ejecutados en Google Colab con acceso gratuito a recursos de cómputo en la nube.

3.4.1.5 Fase 4: Salidas y resultados

La cuarta fase corresponde a la presentación y comunicación de los resultados del sistema predictivo en formatos accesibles para la evaluación académica y para el uso práctica por parte del sector de transporte. Las salidas del sistema comprenden dos componentes principales.

El primero es el informe académico de resultados, un documento formal en formato Word y PDF que incluye: las métrica de la evaluación del modelo extraído para analizar las dos variables de salida cuantificadas (MAE, MSE, RMSE y R2) ; como la comparativa entre los valores predichos y los valores calculados de costos del conjunto de test; el gráfico de la importancia de cada una de las variables que indica cuales de los factores operativos tienen más influencia en el costo; y la explicación de los resultados a partir de los objetivos de la investigación. Este informe aporta el objetivo 3 de la tesis al haber evaluado la precisión del sistema y puesto en comparativa el antiguo método tradicional de estimación de costos.

El segundo componente es una página web de consulta interactiva que se encuentra en la parte de la interfaz de usuario del sistema predictivo de costos. A través de esta página web, el personal logístico de una empresa de transportes puede rellenar los parámetros de un viaje planificado (distancia estimada en km, ciudad de destino, tipo de vehículo, mes del año) y obtener inmediatamente la estimación del costo total en USD y el costo por km generados por el modelo *Random Forest* entrenado. La interfaz de usuario elimina la necesidad de que el usuario final conozca alguna forma de programación, convirtiendo el sistema en accesible y funcional en la realidad operativa del sector de transporte de cargas pesadas ecuatoriano.

A continuación, en la Figura 5 se muestra un diagrama del proceso que se siguió para calcular y predecir correctamente los datos:

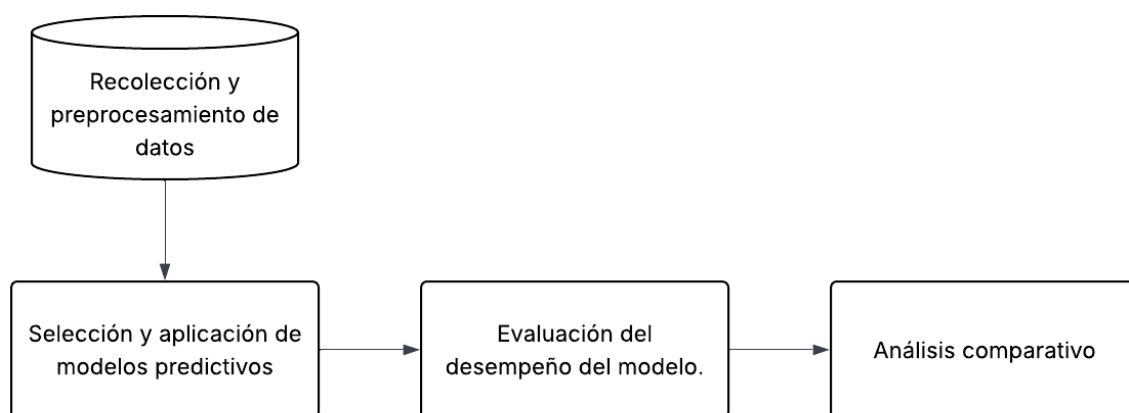


Figura 5. Metodología del estudio del sistema predictivo.

3.5. ANÁLISIS ESTADÍSTICO

3.5.1 Técnicas de análisis de datos

Para el desarrollo del análisis de la información de la presente investigación se utilizaron herramientas computacionales enfocadas en la manipulación de la información y la modelación predictiva. En primer lugar, se utilizó Microsoft Excel para la preparación inicial, limpieza y estructuración de los registros operativos proporcionados por la compañía de transporte, lo que permitió conformar una base de datos fiable y válida para su posterior análisis.

Seguido de la herramienta previa, se utilizó el lenguaje de programación *Python*, ejecutado bajo la plataforma Google Colab, para llevar a cabo la manipulación avanzada de los datos, el entrenamiento de los modelos de aprendizaje automático y la evaluación de los resultados. La manipulación y transformación del *dataset* fue llevada a cabo por medio de las librerías *pandas* y *numpy*, mientras que los algoritmos de regresión empleados en la investigación fueron implementados en *scikit-learn*.

Como técnica principal de análisis se aplicó el aprendizaje supervisado, concretamente mediante modelos de regresión, con el fin de predecir el costo total del transporte en función de las variables operativas. Para validar el rendimiento del

modelo, se utilizaron las métricas estadísticas más extendidas en el área de la inteligencia artificial como son:

- Error Absoluto Medio (MAE)
- Error Cuadrático Medio (MSE)
- Raíz del Error Cuadrático Medio (RMSE)
- Coeficiente de determinación (R^2)

Además, se llevó cabo un análisis de la importancia de variables usando el algoritmo *Random Forest*, utilizando la técnica de Reducción Media de la Impureza (Mean Decrease in Impurity, MDI), lo que es un buen modo de poder conocer cuáles son los factores que más afectan sobre la estructura de costos que hay en el transporte de carga pesada.

Por último, se generaron gráficos de interpretación utilizando las librerías *matplotlib* y *seaborn*, comparaciones de modelos, exactitud del análisis y validación visual de predicciones contra valores reales.

3.5.2 Técnicas de recolección de datos

El procedimiento utilizado para la recolección de información se refiere al análisis documental y a la recolección de datos operativos tal y como son los que entrega a priori una empresa de transportes de carga pesada de la ciudad de Tulcán.

La base de datos utilizada tiene información histórica de viajes realizados por tres unidades de transporte (tractocamiones) que contemplan las siguientes variables:

- Ciudad de partida y de destino
- Distancia recorrida
- Tipo de carga
- Peso transportado
- Costos de combustible
- Costo total del viaje

Estos datos se complementaron con parámetros técnicos y costos unitarios obtenidos en el análisis del objetivo 1, ello permite estructurar una completa base de datos para el entrenamiento del modelo predictivo.

Adicionalmente, ante la escasez de combinaciones completas entre vehículos y rutas, se aplica una técnica de simulación controlada en base a datos reales que produce pequeñas variaciones sobre los registros existentes con el fin de aumentar la capacidad de aprendizaje del modelo sin perder la lógica de la operativa del sistema de transporte.

3.5.3 Instrumentos de investigación

La base de datos concebida y adaptada específicamente para la finalidad de los análisis mediante modelos de *machine learning* resultó ser el principal instrumento de esta investigación, la cual presenta registros estructurados a nivel de viaje, que contemplan no solo variables de entrada (predictivas) sino también variables de salida (de objetivo) posibilitando la aplicabilidad directa a algoritmos de regresión.

Utilizando como instrumentos tecnológicos:

- Google Colab: entorno de ejecución en el cual se desarrolla el modelo predictivo
- *Python*: lenguaje base del procesamiento de datos
- Bibliotecas específicas:
 - *pandas* y *numpy* para análisis de datos
 - *scikit-learn* para modelamiento predictivo
 - *matplotlib* y *seaborn* para visualización

Finalmente, se desarrolló, como instrumento aplicado al entorno práctico, un modelo exportado en formato *.pkl*, que se integrará en una página web interactiva que permita a usuarios del sector logístico predecir costos en tiempo real.

IV. RESULTADOS Y DISCUSIÓN

4.1. RESULTADOS

4.1.1 Análisis que inciden en los costos de transporte de carga pesada

El transporte de carga pesada es un eje fundamental en la cadena logística en el Ecuador, ya que es el medio que permite la movilización de mercancías que forman parte del proceso de producción, distribución y consumo de bienes. Sin embargo, el sector del transporte de carga pesada ha soportado los embates de varios factores que han distorsionado el movimiento logístico de carga pesada en los últimos años. Entre estos factores siempre están incluidos: el aumento de precio de los combustibles, el deterioro de las vías, también el aumento de los costos operativos y la normatividad exigida por la Agencia Nacional de Tránsito.

Todos estos factores aumentan la presión de las empresas de transporte para que optimicen sus procesos y controlen más sus costos. La información tratada en este análisis ha sido extraída de los datos obtenidos durante el proyecto de vinculación realizado en la carrera de Logística y Transporte, la información fue complementada por plantillas y registros de la Agencia Nacional de Tránsito, información que garantiza la fiabilidad y pertinencia de los datos en los que se basa el análisis.

Con el propósito de analizar los factores que inciden en los costos de transporte de carga pesada, se realizó un levantamiento de información mediante la revisión de registros operativos, reporte de viajes y documentos relacionados con los gastos generados en la operación de transporte. También se efectuaron consultas al personal encargado de la gestión logística con el fin de identificar los principales elementos que influyen en la estructura de costos del servicio.

El análisis de la información permitió determinar que los costos de transporte de carga pesada están asociados a diversos factores operativos y logísticos que intervienen durante el desarrollo de las actividades de traslado de mercancías. Entre los principales factores identificados se encuentran la distancia recorrida, el consumo de

combustible, el tiempo, los costos de mantenimiento, los gastos adicionales, tales como peajes y otros costos operativos.

Estos elementos constituyen aspectos básicos de la estructura de costos del transporte, ya que son determinantes del costo total de cada uno de los viajes que se realizan. En el campo de la logística y el transporte el estudio de estos factores permite conocer cómo se produce el costo operativo y cuál es la trascendencia de cada uno de ellos en el funcionamiento del sistema de transporte.

El estudio de estos factores es importante para el desarrollo de sistemas de predicción de costos porque permite identificar las variables que influyen en el comportamiento del gasto logístico y que, posteriormente, pueden ser consideradas como variables de entrada en modelos fundamentados en herramientas de inteligencia artificial.

A continuación, se presentan los resultados obtenidos del análisis de los principales factores que influyen en los costos del transporte de carga pesada.

4.1.1.1 Costos de transporte de carga pesada

Los costos de transporte de carga pesada son los gastos que generan el traslado de mercancías entre un punto de origen y uno de destino y considera la forma de operar el vehículo, el consumo de recursos y las condiciones en las que se realiza el servicio.

El análisis de los costos de transporte es un elemento fundamental para el proceso de toma de decisiones porque permite evaluar el funcionamiento operativo de una forma de transporte y conocer el impacto económico que las actividades de distribución tienen; esto es, conocer los factores que provocan el impacto permite comprender la manera en la que se estructuran los gastos para el transporte y planificar para mejorar la eficiencia logística.

4.1.1.1.1 Distancia recorrida

La distancia recorrida es la cantidad total de kilómetros que tiene que recorrer el medio de transporte, desde el punto de origen hasta el punto de destino de la mercancía. Esta variable es uno de los elementos más importantes de los costos del transporte de carga pesada, ya que guarda una relación directa tanto con lo que corresponde al costo del consumo de combustible, cuanto, al desgaste del propio

medio de transporte, como del tiempo que se requiere para realizar la operación del transporte de la mercancía.

La distancia recorrida, en operaciones de transporte, depende fundamentalmente del posicionamiento geográfico del punto de carga y del punto de descarga y, por otra parte, del planeamiento de las rutas que corresponden al realizar el transporte de la mercancía. La distancia recorrida se incrementa, irremediablemente, de forma paralela al aumento de los costos de la operación del servicio de transporte de carga pesada, en general, sobre todo aquello que tiene que ver con el consumo de combustible y el uso del medio de transporte.

El estudio de este factor de distancia recorrida confirma la incidencia que demuestra dentro de la estructura de los costos de transporte y, a su vez, hace posible la búsqueda de situaciones que permiten hacer planificaciones más eficientes en lo que se refiere al planeamiento de las rutas y a la optimización de los propios recursos logísticos.

4.1.1.1.2 Tiempo de transporte

El tiempo de transporte es el tiempo total transcurrido en el traslado de la carga desde el punto de origen hasta el punto de destino. En él se incluye tanto el tiempo de transporte del vehículo como el tiempo de operación de carga y descarga de la mercancía.

El tiempo de operación del transporte repercute en los costos del servicio ya que influye en la utilización del vehículo, en la planificación de la ruta y en el pago del personal de operación. Incrementar el tiempo de transporte también incrementa los costos de la operación, y, sobre todo, los costos del recurso humano y los costos del vehículo si se utilizara un tiempo mayor.

El análisis de este elemento del detalle de costos permite juzgar la eficiencia de las operaciones logísticas y facilita su valor en la estructura de costos del servicio.

En la Tabla 5 y en el Anexo 2 se presenta un ejemplo de la estructura del *dataset* considerando la distancia de la ruta, el tiempo por ruta en horas y las paradas por viaje.

Tabla 5. Distancia en kilómetros, tiempo en horas y paradas por viaje por ruta.

CIUDAD_ORIGEN	CIUDAD_DESTINO	DISTANCIA	TIEMPO_HORAS	PARADAS_POR_VIAJE
TULCAN	GUAYAQUIL	664	22	5
TULCAN	MANTA	616	17.5	4
TULCAN	QUITO	241	6	2
TULCAN	DANEC	267	6	2
QUITO	MANTA	397	8	2
QUITO	GUAYAQUIL	430	9	2
QUEVEPALMA	IPIALES	496	11	3
MANTA	IPIALES	637	18	4
GUAYAQUIL	MANTA	199	4	1
GUAYAQUIL	TULCAN	664	22	5
DANEC	QUEVEPALMA	215	5	1

Fuente: Yanca (2025)

4.1.1.1.3 Consumo de combustible

El gasto en combustible constituye una de las más significativas partes que se integran en los costos de las operaciones que forman la realización del transporte de carga pesada. Este factor se encuentra relacionado directamente con la distancia recorrida, el tipo de vehículo empleado, el peso de la mercancía que se transporta y el estado de la vía del transporte.

El gasto en combustible conforma una de las más importantes partes que conforma el costo total del transporte, dado que es un medio que resulta necesario para la puesta en marcha de los vehículos que son necesarios para la distribución de mercancías. Una mayor cantidad de gasto en combustible supone un aumento en los costos, de manera que las operaciones del transporte podrían llegar a no ser rentables.

Por lo tanto, el análisis del consumo de combustible sirve para descubrir patrones de gasto dentro de las operaciones logísticas al mismo tiempo que sirve para relacionarlo con otros factores que interrelacionen el costo del transporte.

En la Tabla 6 y en el Anexo 2 se presenta un ejemplo del gasto de combustible de la ruta de Tulcán – Quito en 4 viajes al mes.

Tabla 6. Costo de combustible por rutas.

CIUDAD_ORIGE N	CIUDAD_DESTI NO	GALONES_CONSUM IDOS	PRECIO_PROMEDIO_C OMBUSTIBLE	COSTO_TOTAL_G ALON
TULCAN	GUAYAQUIL	140	1.980961583	\$277.3346217
GUAYAQUIL	MANTA	35	1.980961583	\$69.33365542
MANTA	IPIALES	130	1.980961583	\$257.5250058
TULCAN	MANTA	125	1.980961583	\$247.6201979
MANTA	IPIALES	130	1.980961583	\$257.5250058
TULCAN	QUITO	65	1.980961583	\$128.7625029
QUITO	MANTA	50	1.980961583	\$99.04807917
MANTA	IPIALES	130	1.980961583	\$257.5250058
TULCAN	GUAYAQUIL	140	1.980961583	\$277.3346217
GUAYAQUIL	TULCAN	140	1.980961583	\$277.3346217
TULCAN	DANEC	55	1.980961583	\$108.9528871
DANEC	QUEVEPALMA	25	1.980961583	\$49.52403958
QUEVEPALMA	IPIALES	110	1.980961583	\$217.9057742
TULCAN	QUITO	65	1.980961583	\$128.7625029
QUITO	GUAYAQUIL	55	1.980961583	\$108.9528871
GUAYAQUIL	TULCAN	140	1.980961583	\$277.3346217
TOTAL		1535		\$3040.77603

Fuente: Yanca (2025)

4.1.1.1.4 Costos de mantenimiento

Los costos de mantenimiento se refieren a los gastos relacionados con la conservación y reparación de los vehículos que son utilizados en la actividad de transporte, el cual está asociado a actividades como el cambio de aceite, el reemplazo de llantas, revisiones mecánicas y todo tipo de reparaciones necesarias para garantizar el funcionamiento del vehículo.

El mantenimiento hace parte de las operaciones del transporte por cuanto que es una forma de prevenir fallas mecánicas y asegurar la continuidad de las actividades logísticas, pero representa también un costo operativo que se debe incluir dentro de la estructura de costos del transporte.

El análisis de los costos de mantenimiento permite identificar la importancia económica del mantenimiento dentro de las operaciones de transporte y su relación con el funcionamiento del sistema logístico.

En la Tabla 7 y en el Anexo 2 se muestran los costos de mantenimiento más importantes en el sector del transporte pesado.

Tabla 7. Costos de mantenimiento del transporte de carga pesada.

NOMBRE	COSTO UNITARIO	DURACIÓN (km)	CANTIDAD	COSTO x KM
LLANTAS Y ALINEACIÓN (tractocamión: 18 llantas)				
Llantas delanteras	\$470.00	80,000	2	0.0118
Llantas de tracción	\$470.00	80,000	8	0.0470
Llantas de arrastre	\$350.00	120,000	12	0.0350
Rines	\$120.00	450,000	22	0.0059
Pinchazos y daños	\$15.00	10,000	2 mensuales	0.0015
Alineación	\$20.00	60,000	2 veces al año	0.0003
Balanceo	\$5.00	60,000	2 veces al año	0.0001
MANTENIMIENTO ORDINARIO				
Cambio de aceite motor (2 canecas)	\$120.00	10,000	1	0.0120
Aceite de caja y transmisión (3 gln)	\$160.00	60,000	1	0.0027
Aceite de diferencial (3 gln)	\$140.00	60,000	1	0.0023
Filtros de aire	\$130.00	60,000	2	0.0043
Filtros de aceite y combustible	\$93.00	10,000	2	0.0186
Lubricantes y fluidos varios	\$100.00	5,000	2	0.0400
Limpieza / lavado motor y chasis	\$100.00	10,000	1	0.0100
SISTEMA ELÉCTRICO				
Baterías (2 unidades)	\$120.00	200,000	4	0.0024
Motor de arranque	\$60.00	10,000	1	0.0060
Alternador	\$50.00	10,000	1	0.0050
Luces y accesorios	\$60.00	10,000	1	0.0060
Sistema eléctrico / revisión general	\$50.00	10,000	1	0.0050
SISTEMA DE FRENOS				
Zapatas / remache y limpieza (2 ejes)	\$300.00	60,000	22	0.1100
Ajustes y calibraciones de frenos	\$100.00	10,000	1	0.0100
Revisión frenos de aire	\$100.00	10,000	1	0.0100
Sellos y retenedores	\$100.00	10,000	1	0.0100
SUSPENSIÓN				
Resortes y ballestas	\$250.00	10,000	1	0.0250
Cauchos y silentblocks	\$100.00	10,000	1	0.0100
Revisión de suspensión general	\$200.00	60,000	1	0.0033
SISTEMA DE TRANSMISIÓN				
Revisión de caja de cambios	\$300.00	120,000	1	0.0025
Revisión de diferencial	\$300.00	120,000	1	0.0025
Dirección hidráulica (aceite)	\$150.00	120,000	1	0.0013
REPARACIÓN DE MOTOR				

Calibración de válvulas motor	\$50.00	500,000	1	0.0001
Calibración bomba de inyección	\$1,000.00	500,000	1	0.0020
Cambio de toberas de inyectores	\$600.00	500,000	1	0.0012
Mantenimiento de turbo	\$1,200.00	60,000	1	0.0200
Refrigerante motor (10 galones)	\$120.00	60,000	1	0.0020
Overhaul motor (reserva propocionada)	\$4,500.00	500,000	1	0.0090
TOTAL				0.4347

Fuente: Yanca (2025)

4.1.1.1.5 Peajes y costos operativos adicionales

Aparte de los costos operativos más importantes, la operación de los transportes de carga pesada también genera gastos adicionales ya que el desarrollo de la actividad logística necesita de servicios e infraestructura viaria además de los directamente relacionados con el tráfico de camiones.

En este sentido, se refiere a la existencia de peajes, gastos de aparcamientos, permisos de transporte o bien otros gastos operativos derivados del propio desarrollo del servicio de transporte. Estos gastos no son invariantes ya que al poder variar y/o defenderse en función de la ruta utilizada y/o de las condiciones en que se produce el transporte.

El análisis de la existencia de esos gastos permite identificar el peso de los costos en el costo total del transporte y predecir los costos operativos según las diferentes rutas y/o condiciones de operación.

En la Tabla 8 y en el Anexo 2 se muestra el costo de peajes y costos operativos adicionales relacionados con el transporte pesado.

Tabla 8. Costos operativos de transporte de carga pesada.

CIUDAD_ORIGEN	CIUDAD_DESTINO	TIPO_CARGA	PEAJES	COSTO_PEAJES (USD)	VIATICOS (USD)
TULCAN	GUAYAQUIL	FRAGIL	9	54.00	45
TULCAN	MANTA	GENERAL	8	48.00	45
TULCAN	QUITO	GENERAL	5	30.00	15
TULCAN	GUAYAQUIL	PELIGROSA	8	48.00	45
TULCAN	DANEC	FRAGIL	6	36.00	15
QUITO	MANTA	VACIO	3	18.00	15
QUITO	GUAYAQUIL	VACIO	4	24.00	25
QUEVEPALMA	IPIALES	GENERAL	8	48.00	35
MANTA	IPIALES	GENERAL	8	48.00	40
GUAYAQUIL	MANTA	VACIO	2	12.00	10

CIUDAD_ORIGEN	CIUDAD_DESTINO	TIPO_CARGA	PEAJES	COSTO_PEAJES (USD)	VIATICOS (USD)
GUAYAQUIL	TULCAN	GENERAL	8	48.00	45
DANEC	QUEVEPALMA	VACIO	2	12.00	10

Fuente: Yanca (2025)

4.1.2 Modelos de inteligencia artificial aplicados al cálculo de costos de transporte

En el transporte y la logística y la aplicación de herramientas de inteligencia artificial ha permitido mejorar notablemente la estimación y predicción de costos operativos, dado que estas tecnologías se basan en algoritmos que analizan gran volumen de datos históricos, de tal forma que detectan patrones y relaciones entre las variables que se producen en los costos del transporte.

Varios autores dicen que los modelos de IA permiten capturar diversos factores operativos como distancia, consumo de combustible, tiempo de transporte, costos de mantenimiento y características de la carga para obtener estimaciones más precisas del costo total de la operación de transporte.

En el marco de los modelos de LIA se cuentan aquellos que refuerzan modelos de regresión, árboles de decisión, algoritmos de aprendizaje automático y las redes neuronales artificiales a partir de los cuales procesadas las variables fundamentales se pueden analizar datos históricos sobre la operación de transporte y predecir el comportamiento de las variables estudiadas.

La aplicación de estos modelos en el sector del transporte de mercancías pesadas permitiría mejorar la planificación logística, optimizar la utilización de los recursos y contribuir a la toma de decisiones estratégicas en las empresas de transporte.

A continuación, en la Tabla 9 se presentan algunos de los principales modelos de inteligencia artificial utilizados en procesos de cálculo y predicción de costos en el transporte de carga.

Tabla 9. Comparación de herramientas de Inteligencia Artificial.

Modelo de inteligencia artificial	Descripción	Aplicación en transporte	Limitaciones para la presente investigación
Regresión Lineal	Modelo estadístico que analiza la relación entre variables independientes y una variable dependiente mediante una ecuación lineal.	Estimación de costos de transporte en función de variables como distancia, peso o consumo de combustible.	Asume relaciones lineales entre variables. En sistemas logísticos reales existen interacciones complejas y no lineales que este modelo no logra capturar completamente.
Árbol de Decisión	Algoritmo que genera reglas de decisión a partir de datos históricos mediante una estructura jerárquica de nodos y ramas.	Identificación de factores que influyen en los costos logísticos y apoyo a la toma de decisiones operativas.	Puede presentar sobreajuste cuando se entrena con muchos datos y generar resultados inestables ante pequeñas variaciones del conjunto de entrenamiento.
Random Forest	Conjunto de múltiples árboles de decisión que trabajan de forma conjunta para mejorar la precisión y reducir el sobreajuste.	Predicción de costos de transporte utilizando simultáneamente múltiples variables operativas.	Requiere mayor capacidad computacional que modelos simples; sin embargo, ofrece una alta precisión y estabilidad en los resultados, razón por la cual fue seleccionado para esta investigación.
Gradient Boosting	Algoritmo de ensamble que construye árboles de decisión de manera secuencial, corrigiendo progresivamente los errores de los árboles anteriores.	Predicción de costos, demanda y optimización de procesos logísticos mediante aprendizaje iterativo.	Aunque presenta una elevada capacidad predictiva, suele requerir un ajuste más cuidadoso de hiperparámetros y es más sensible al sobreajuste cuando se trabaja con bases de datos de tamaño moderado.

Modelo de inteligencia artificial	Descripción	Aplicación en transporte	Limitaciones para la presente investigación
Redes Neuronales Artificiales	Modelo inspirado en el funcionamiento del cerebro humano capaz de aprender patrones complejos a partir de grandes volúmenes de datos.	Predicción de costos y comportamiento de sistemas logísticos con múltiples variables interrelacionadas.	Requieren una mayor cantidad de datos para alcanzar su máximo rendimiento y presentan menor interpretabilidad en comparación con Random Forest.
Máquinas de Vectores de Soporte (SVM)	Algoritmo de aprendizaje supervisado utilizado para clasificación y regresión mediante la búsqueda de hiperplanos óptimos.	Estimación de costos en sistemas logísticos con alta variabilidad de datos.	Su rendimiento disminuye cuando se trabaja con conjuntos de datos grandes y múltiples variables categóricas, además de requerir una calibración compleja de parámetros.

Los modelos expuestos dan la posibilidad de analizar grandes conjuntos de datos acerca de las actividades de transporte, facilitando así la detección de patrones en los costos logísticos. De entre ellos, la clase de modelos correspondientes a las redes neuronales artificiales, así como los modelos no paramétricos basados en algoritmos de árboles de decisión, destacan por su capacidad de procesar múltiples variables y de emitir predicciones con gran grado de precisión, lo que convierte en técnicas de interés a la hora de generar sistemas de predicción de los costos en transporte de carga pesada.

4.1.3 Desarrollo del modelo predictivo basado en Random Forest para la estimación de costos de transporte de carga pesada

Con base en los factores identificados en el objetivo 1 y en los registros operativos recopilados durante el proyecto de vinculación de la carrera de Logística y Transporte, se desarrolló un modelo predictivo fundamentado en el algoritmo Random Forest. El modelo tiene como propósito predecir dos variables de salida: el costo total de cada viaje en dólares (USD) y el costo por kilómetro (USD/km), a partir de las características operativas de cada viaje registrado.

El *dataset* que se ha utilizado está compuesto por 327 registros reales de viaje realizados por tres unidades de carga pesada, durante el periodo comprendido entre noviembre del 2024 y octubre del 2025. Cada registro corresponde a un viaje individual e incluye información sobre la distancia recorrida, el destino, el flete cobrado al cliente y el tipo de carga transportada. El costo de cada viaje fue calculado aplicando costos unitarios determinados en el resultado 1: USD 0.415 por km de combustible, USD 0.158 por km de mantenimiento, y USD 515.86 por viaje como porción de los costos operativos fijos mensuales.

El desarrollo del modelo siguió una metodología estructurada que comprende cuatro etapas: la preparación y estructuración del conjunto de datos, la definición de variables de entrada y salida, la configuración e implementación del modelo en *Python* mediante Google Colab y la verificación del proceso de entrenamiento. La evaluación de la precisión del modelo mediante las métricas MAE, MSE y R2 se presenta en el resultado del objetivo 3.

4.1.3.1 Preparación y estructuración del conjunto de datos

La preparación del *dataset* constituye la etapa inicial y más importante del desarrollo de cualquier modelo de inteligencia artificial, ya que la calidad, variabilidad y consistencia de los datos determinan directamente la capacidad predictiva del modelo (Brownlee, 2019). Para este trabajo, el *dataset* fue construido a partir de los registros de despacho de tres unidades vehiculares de carga pesada que cubren un total de 327 viajes.

Los registros originales contenían la siguiente información por viaje: número de viaje, mes, fecha, placa del vehículo, ciudad de origen, ciudad de destino, distancia recorrida en kilómetros, flete cobrando al cliente en dólares, tipo de carga y descripción de la ruta.

El proceso de preparación del *dataset* incluyó las siguientes etapas: primero, la verificación de la integridad de los 327 registros, confirmando la ausencia de valores faltantes en las variables de distancia y flete; segundo, el cálculo de las tres variables de costo por viaje (combustible, mantenimiento y costos operativos); tercero, la codificación numérica de las variables categóricas, necesaria para el procesamiento

por el algoritmo Random Forest; y cuarto, el cálculo de las variables de salida del modelo: el costo total del viaje por kilómetro.

A continuación, en la Tabla 10 se presentan la lista de variables que se tomaron cuenta para la recolección de datos y para la fase predictiva del modelo.

Tabla 10. Variables de entrada y salida para el modelo predictivo.

Variable	Descripción	Tipo	Unidad
Fecha del viaje	Fecha en la que se realiza el viaje (permite predecir clima y condiciones)	Entrada	Fecha (YYYY-MM-DD)
Origen	Ciudad desde donde inicia el viaje	Entrada	Categórica
Destino	Ciudad de llegada del viaje	Entrada	Categórica
Vehículo	Tipo de unidad de transporte utilizada	Entrada	Categórica
Tipo de carga	Clasificación de la mercancía (FRÁGIL, GENERAL, PELIGROSA, VACÍO)	Entrada	Categórica
Distancia	Kilómetros recorridos en el trayecto	Entrada	km
Peso de carga	Toneladas transportadas en el viaje	Entrada	toneladas
Velocidad promedio	Velocidad promedio del vehículo durante el trayecto	Entrada	km/h
Número de paradas	Cantidad de paradas durante el viaje	Entrada	entero
Costo total del viaje	Variable objetivo: costo operativo total del viaje	Salida	USD

Con el fin de evidenciar la arquitectura conceptual del sistema predictivo y la dinámica de la interacción entre dichas variables, se ha confeccionado un diagrama de flujo de datos. Como se indica en la Figura 6, el sistema acepta como entradas primarias (X) las variables operativas (donde la distancia y el vehículo son las variables dominantes) y los costos base (combustible, salarios). La información se introduce en el motor de procesamiento estructurado aplicado como algoritmo Random Forest que elabora y representa las interacciones complejas a los efectos de emitir como salida (Y) la predicción del costo real del viaje.

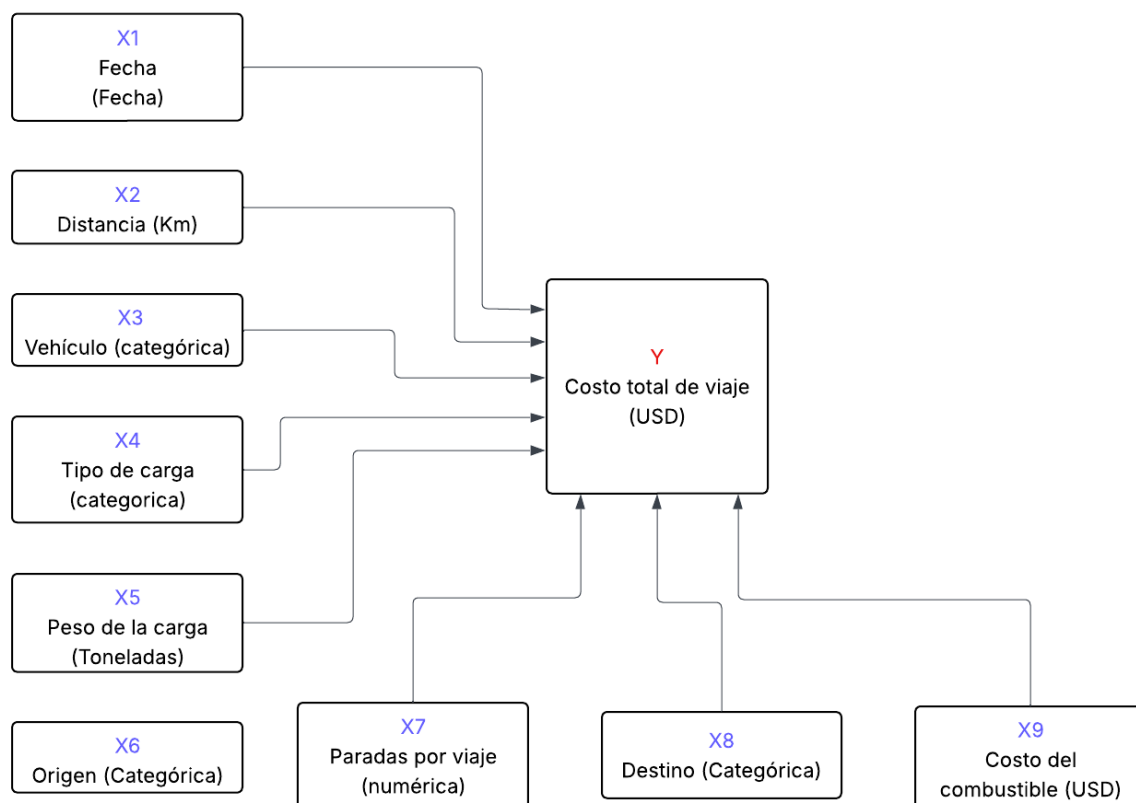


Figura 6. Variables que afectan a los costos de transporte.

Tal como se puede observar en el esquema conceptual, el modelo no resulta ser el resultado de una mera suma de características, sino que se asentaría en un esquema de regresión en el que el algoritmo sería capaz de ponderar la importancia relativa de cada una de las variables que forman parte del sistema. Al ir eliminando los factores que no muestran importancia alguna (como puedan ser los nombres categóricos de las ciudades de origen y destino) o aquellas en las que se genera en mayor medida el precio (kilometraje y tipología del vehículo), el mismo sistema sería capaz de comparar un indicador financiero de llegada (\$/km/viaje), lo que proporciona un alto nivel de adaptación a la realidad logística.

4.1.3.2 Descripción del dataset de entrenamiento

El conjunto de datos que se ha utilizado para entrenar y validar el sistema de predicción está constituido por un total de 327 registros de viajes que pertenecen a una empresa de transporte de carga pesada que está ubicada en Ecuador. La flota que tiene la empresa trabaja con tres vehículos, y los registros recogen operaciones

a lo largo de un año del calendario distribuido en distintos meses y según la disponibilidad y actividad de cada unidad.

Las rutas que han sido registradas a lo largo del conjunto de datos incluyen múltiples orígenes y destinos dentro del territorio ecuatoriano y sus interconexiones con Colombia (entre ellas, las siguientes ciudades: Tulcán, Guayaquil, Manta, Quito, Ipiales, Danec y Quevepalma). Las distancias por tramo oscilan entre 199 km (rutas cortas como la Guayaquil-Manta) y 675 km (rutas largas como la Tulcán-Guayaquil); se han registrado un total de trece distancias distintas, que representan la variedad de operaciones de la flota.

En lo referente al tipo de carga que es transportada, el conjunto de datos incluye cuatro categorías: General, Frágil, Peligrosa y Vacío (representa los viajes en vacío, que son los viajes de retorno sin carga) de modo que permite capturar los efectos del tipo de mercancía en el costo operativo del sistema. El peso de carga oscila entre cero toneladas (viajes vacíos) y aproximadamente 32 toneladas (viajes de carga plena).

Los valores del costo total del viaje (que es la variable objetivo del modelo) oscilan entre \$131.20 USD y \$624.26 USD, lo que pone de manifiesto la amplitud de la variabilidad de los costos (en función de la ruta, del vehículo y del tipo de operación). La Tabla 11 y el Anexo 2 presenta una muestra sintomática de los registros de la base de datos, incorporando las principales variables de entrada y el costo total sistémico por viaje. La base de datos fue tratada y manipulada por medio del lenguaje de programación *Python* y fue exportada a formato Excel para su uso directo en el entrenamiento del modelo.

Tabla 11. Muestra de registros del dataset de entrenamiento.

ID	Mes	Vehículo	Origen	Destino	Dist. (km)	Peso (t)	Tipo Carga	Costo Comb. (USD)	Costo Total (USD)
1	Enero	CBA-1777	Tulcán	Guayaquil	675	29.9	Frágil	142.99	624.3
4	Enero	CBA-1777	Guayaquil	Manta	199	0	Vacío	33.33	175.2
7	Enero	CBA-1777	Manta	Ipiales	631	30	General	133	582.9
109	Enero	CAA-2669	Tulcán	Manta	616	32	General	131.92	571.1
163	Enero	CAA-1887	Tulcán	Quito	241	29.4	General	49.03	220.9
166	Enero	CAA-1887	Quito	Manta	397	0	Vacío	70.14	353.2

ID	Mes	Vehículo	Origen	Destino	Dist. (km)	Peso (t)	Tipo Carga	Costo Comb. (USD)	Costo Total (USD)
265	Marzo	CAA-2669	Tulcán	Danec	267	30	Frágil	56.09	246.5
313	Mayo	CAA-1887	Quito	Guayaquil	430	0	Vacío	75.01	381.6
...	...	...	...	...	...	...	...	...	...
327	Agosto	CAA-1887	Guayaquil	Tulcán	664	30.8	General	134.92	459.6

Fuente: Yanca (2025)

La Tabla 12 muestra la síntesis mensual consolidada de los costos de la flota total. El costo total que contienen el conjunto de datos asciende a la cantidad de USD 67,288.75, el cual ha sido generado por 58 viajes y 90,150 km recorridos en el periodo considerado. El costo promedio por kilómetro de toda la flota se ha mantenido en USD 1,99/km a lo largo del año, manteniéndose constante en todas las rutas analizadas, lo que pone de manifiesto la regularidad de las tarifas operativas aplicadas.

La ruta con mayor actividad ha sido la Ruta 1, con 23 viajes contabilizados para un total de 12,147 km mensuales (279,381.00 km al año) y un costo total mensual de USD 23,281.75, asumiendo la segunda posición la Ruta 2 con 10 viajes y USD 13,427.50. En el otro extremo, la Ruta 5 ha registrado la menor operación, con tan solo 1 viaje y un costo total mensual de USD 1,270.00. El costo total anual proyectado con base en los kilómetros recorridos asciende a 807,465.00 km al año en toda la flota.

Tabla 12. Resumen mensual de costos de la flota.

Ruta	Viajes	Km Totales	Km totales por viajes	Costo Total Mes (USD)	USD/km prom.
Ruta 1: Tulcán → Guayaquil → Manta → Ipiales	23	12,147	279,381.00	23,281.75	1.99
Ruta 2: Tulcán → Manta → Ipiales	10	16,113	161,130.00	13,427.50	1.99
Ruta 3: Tulcán → Quito → Manta → Ipiales	6	19,065	114,390.00	9,532.50	1.99
Ruta 4: Tulcán → Guayaquil → Tulcán	10	8,322	83,220.00	6,935.00	1.99
Ruta 5: Tulcán → Quito → Guayaquil → Tulcán	8	19,263	154,104.00	12,842.00	1.99
Ruta 6: Tulcán → Danec/sector → Quevepalma → Ipiales	1	15,240	15,240.00	1,270.00	1.99
TOTAL	58	90,150	807,465.00	67,288.75	

4.1.3.3 Análisis del algoritmo Random Forest

El algoritmo de aprendizaje de conjunto Random Forest construye árboles de decisión aleatorios y combina sus predicciones mediante promediación, generando de esta manera un estimador final más estable y preciso que el de un solo árbol de decisiones (Breiman, 2001). Su selección para este trabajo se sustenta en tres razones principales.

En primer lugar, Random Forest puede manejar eficientemente tanto variables numéricas continuas (km, costos) como categóricas codificadas (destino, placa, mes) sin requerir transformaciones adicionales, por lo que se acomoda a la estructura heterogénea del *dataset* de viajes. En segundo lugar, este algoritmo es robusto en presencia de *outliers* y no requiere que los datos sean normalmente distribuidos, condición que no siempre se respeta en registros operacionales de transporte. En tercer lugar, Random Forest genera automáticamente un ranking de importancia de variables que permite identificar qué variables tienen mayor incidencia sobre el costo predicho, añadiendo valor analítico extra al propio sistema predictivo (Katreddi et al., 2023).

A diferencia de la regresión lineal, que solo puede capturar relaciones lineales entre las variables, Random Forest puede modelar interacciones complejas y no lineales, como la relación entre la distancia recorrida, el destino y el costo total, donde la combinación de ambas variables puede producir efectos que no son simplemente aditivos. Esta capacidad resulta especialmente relevante en el contexto del transporte de carga pesada, donde los costos dependen simultáneamente de múltiples factores que interactúan entre sí (Urrego, 2024).

4.1.3.4 Configuración del modelo e implementación en *Python*

El modelo fue implementado en el entorno Google Colab utilizando *Python*. Se emplearon las bibliotecas *pandas* para la carga y manipulación del *dataset*, *scikit-learn* para la implementación del algoritmo Random Forest y el cálculo de métricas de evaluación, y *matplotlib* para la visualización de los resultados. La Tabla 13 presenta los hiperparámetros configurados para el modelo y su justificación.

Tabla 13. Configuración de hiperparámetros del modelo Random Forest.

Hiperparámetro	Valor	Justificación
n_estimators	100	Equilibra precisión y tiempo de cómputo para el tamaño del dataset
max_depth	None	Permite crecimiento libre de cada árbol para capturar todos los patrones
min_samples_split	2	Valor por defecto; adecuado para 327 registros
min_samples_leaf	1	Permite que el modelo aprenda de registros individuales
max_features	sqrt	Raíz cuadrada del número de variables en cada división; reduce correlación entre árboles
random_state	42	Garantiza reproducibilidad de los resultados
Division train/test	80% / 20%	262 registros para entrenamiento, 65 para prueba

Se entrenaron dos modelos de forma independiente: el primero orientado a predecir el costo total del viaje en USD, y el segundo orientado a predecir el costo por kilómetro en USD. Esta decisión permite que cada modelo optimice sus parámetros internos de forma específica para cada variable objetivo, mejorando la precisión individual de cada estimación.

El *dataset* fue dividido en 80% para entrenamiento y 20% de prueba, proporción estándar en modelos de *machine learning* que garantiza suficientes datos para el aprendizaje y una muestra independiente para la validación (Brownlee, 2019)

4.1.3.5 Proceso de entrenamiento del modelo

En el proceso de entrenamiento, el algoritmo *Random Forest* ha generado 100 árboles de decisión en paralelo, todos ellos entrenados en base a una muestra aleatoria del conjunto de entrenamiento (*bootstrap*), y usando en cada nodo de división un subconjunto aleatorio de las ocho variables de entrada. La predicción final de cada modelo se obtiene como promedio de las predicciones de los 100 árboles individuales, lo que reduce significativamente la varianza del modelo y evita el sobreajuste a los datos de entrenamiento (Breiman, 2001).

Una característica que destaca en el proceso de entrenamiento es la generación de un ranking de importancia de variables, que expresa en términos cuantitativos la contribución relativa de cada variable de entrada en la reducción del error de predicción. Con respecto a este ranking, se espera que las variables km,

costo_combustible y destino_cod sean las que aporten mayor valor, ya que como se ha puesto de manifiesto en el resultado del objetivo 1, la distancia recorrida y el consumo de combustible son los variables que más afectan a la estructura de costos del transporte de carga pesada.

4.1.3.6 Alcance y consideraciones del modelo desarrollado

El modelo predictivo que se ha implantado se encuadra en la tarea de predecir costos de transporte para viajes de carga pesada con características similares a los datos contenidos en el *dataset*, ya que tiene por destinos posibles los dentro del Ecuador continental (Guayaquil, Manta y Quito), ya que abarca distancias que van desde los 482 km hasta los 1,505 km, y hay asignados costos unitarios a través del periodo 2024-2025. En este sentido, se puede aplicar directamente del modelo a las tres unidades vehiculares de la flota analizada, siendo aplicable a otras unidades con perfiles operativos similares.

El modelo tiene como consideraciones el *dataset* de 327 registros suficiente para el entrenamiento de un modelo Random Forest de complejidad moderada pero que bien podría ampliarse en periodos futuros para mejorar la capacidad de generalización del sistema, especialmente si se integran nuevas rutas o condiciones operativas diferentes a las registradas. Por otro lado, el modelo no considera variables como el precio del combustible por periodo, las condiciones climáticas o el estado de las vías, porque no se disponía de registros sistematizados para dicha variable y que representan oportunidades de mejora en integral del sistema predictivo en versiones futuras.

4.1.4 Evaluación del modelo predictivo

Con la finalidad de verificar el nivel de exactitud del sistema de predicción de costos de transporte de carga pesada, se aplicaron tres modelos de *Machine Learning*, como son la Regresión Lineal, el *Random Forest* y el *Gradient Boosting* para averiguar cuál de esos algoritmos presenta un mejor comportamiento en cuanto a la estimación del costo total de los viajes en función de una base de datos estructurada con los registros operativos reales de la empresa de transporte.

La base de datos se separó en un 80% para entrenar y un 20% para probar, lo cual permite llevar a cabo una validación objetiva de la capacidad predictiva para cada

uno de los modelos; las métricas usadas para medir el comportamiento de los algoritmos están muy utilizadas en modelos de la regresión, como el Error Absoluto Medio (MAE), la Raíz del Error Cuadrado Medio (RMSE) o el coeficiente de determinación (R^2);

El valor del error absoluto medio (MAE) refleja, por una parte, el promedio de las diferencias absolutas entre los valores reales y aquellos que predijeron los modelos; por esto, el mayor valor de error absoluto medio será igual a cero, lo que se recomienda como resultado de la exactitud que ofrezca el sistema de predicciones que estemos utilizando. En este caso, esa medida del modelo *Gradient Boosting* fue 0,49 dólares, con el cual se tiene también el parámetro de error absoluto medio más pequeño, de tal suerte que, en promedio, sus predicciones se desvían menos de cincuenta centavos de dólar de aquel coche del viaje de transporte real. A continuación de *Gradient Boosting* se encontraba el modelo de *Random Forest* con 0,68 dólares en el error medio absoluto, y, por último, el de Regresión Lineal con un MAE de 3,05 dólares, lo que se puede interpretar de forma que los parámetros de aprendizaje, la mitad de la agrupación de los parámetros de modelo de ensamble, predecían costos sumamente más próximos a los costos del viaje de transporte real. Si bien los valores de error absoluto medio son bajos, las diferencias observadas entre la Regresión Lineal y los modelos de ensamble son bastante evidentes; podríamos afirmar que el MAE de la Regresión Lineal resulta entre 4,5 y 6,2 veces mayor que el observado para *Gradient Boosting* y *Random Forest* respectivamente.

El MSE expone claramente que los modelos de ensamble son preferibles a la Regresión Lineal, que obtiene un MSE de 12,49. La diferencia entre los MSE es relevante ya que el MSE penaliza cuadráticamente los errores elevados, lo que implica que la regresión lineal no sólo comete errores más frecuentemente, sino que en determinados viajes —especialmente en aquellos con rutas capaces de presentar un elevado número de peajes, detenerse repetidamente o tener altas tarifas en los peajes—, las predicciones se separan de forma considerable en relación al costo real, lo que disminuye la validez del modelo para situaciones operativas que pueden ser altamente variables en función de su ruta y la carga de material a transportar.

La raíz del error cuadrático medio (RMSE) ofrece la posibilidad de detectar el tamaño de esos grandes errores en las predicciones. En efecto, la raíz del error cuadrático medio (RMSE) favorece que las diferencias más grandes entre los valores reales y los

valores estimados sean mucho más penalizadas que las diferencias pequeñas. En este caso, el modelo *Gradient Boosting* obtuvo un RMSE igual a 0,79 USD, siendo el más bajo de los tres modelos evaluados. Por su parte, el modelo *Random Forest* ofreció un RMSE de 3,25 USD y la Regresión Lineal presentó un RMSE de 3,53 USD, lo que significa que el modelo de Regresión Lineal tenía una mayor dispersión en algunas estimaciones individuales, mientras que los modelos más avanzados de inteligencia artificial tenían la ventaja de haber cometido menos errores y mostrar una mayor estabilidad predictiva en la totalidad del rango operativo que estaba en evaluación.

El coeficiente de determinación (R^2) es una medida de la proporción de la variabilidad total de los costos que un modelo predictivo puede describir. Los valores de R^2 van desde 0 a 1, siendo los valores cercanos a 1 aquellos que desprenden modelos con gran capacidad explicativa. En esta investigación, el modelo gráfico *Gradient Boosting* y el *Random Forest* han logrado una precisión del 1,0000 o lo que es lo mismo 100%, convirtiéndose en los algoritmos con mejor capacidad predictiva. La Regresión Lineal ha alcanzado un R^2 de 0,99993, lo que significa un 99,93% de explicación de la variabilidad de los costos. A pesar de que los tres modelos han proporcionado resultados que se consideran satisfactorios, los algoritmos con ensambles han demostrado un mejor desempeño a la hora de predecir los costos de transporte de carga pesada.

Los resultados alcanzados permiten llegar a la conclusión de que el modelo gráfico *Gradient Boosting* y el *Random Forest* fueron los algoritmos con mejor desempeño en el trabajo de investigación, ya que presentan el menor MAE, el menor RMSE y el mayor coeficiente de determinación (R^2). Esto pone de manifiesto que estos modelos son los que mejor identificarán las relaciones complejas y patrones no lineales en los datos operacionales de transporte, ofreciendo predicciones más exactas y estables.

Por otro lado, aunque la Regresión Lineal obtuvo resultados aceptables, su rendimiento fue inferior debido a que este tipo de algoritmo asume relaciones lineales entre variables, lo cual limita su capacidad para representar adecuadamente el comportamiento real de los costos logísticos donde intervienen múltiples factores operativos con comportamientos variables y no necesariamente lineales.

4.1.3.1 Comparación de modelos y métricas de evaluación

A continuación, en la Tabla 14 se presentan los resultados obtenidos para cada modelo.

Tabla 14. Resultados de la evaluación del modelo predictivo.

Modelo Algorítmico	R ² (Precisión)	MAE (USD)	MSE	RMSE (USD)
Regresión Lineal	0.9999	\$3.04	12.49	\$3.53
Gradient Boosting	1	\$0.49	0.62	\$0.79
Random Forest	1	\$1.31	10.55	\$3.25

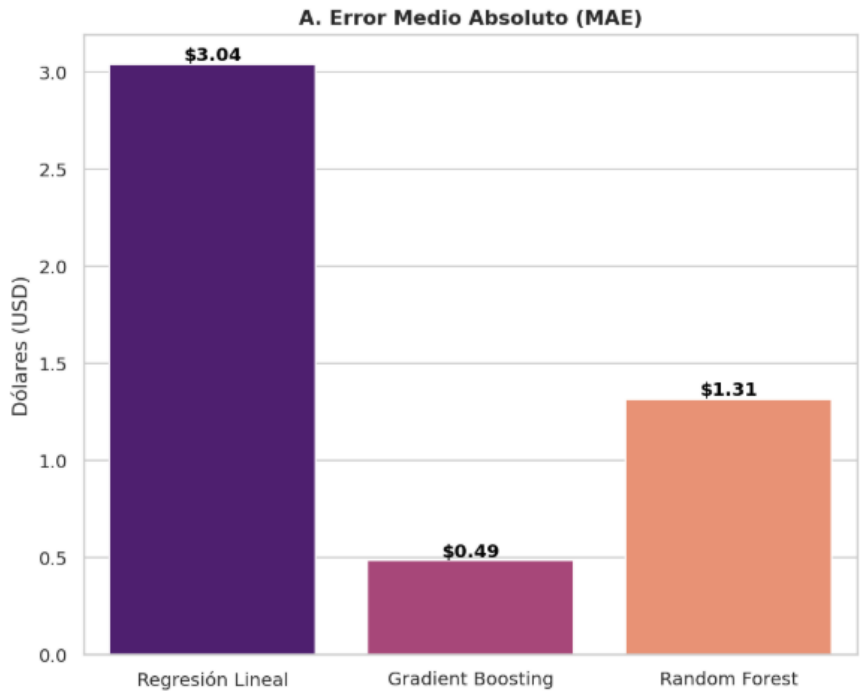


Figura 7. Comparación del MAE obtenidas por los tres modelos evaluados.

El Error Absoluto Medio (MAE) en la Figura 7 pone de manifiesto sin lugar a errores la superioridad de los modelos de ensamble en comparación con clásico. En esa ilustración, *Gradient Boosting* logra el error absoluto medio más bajo con apenas \$0,49 USD, seguido por *Random Forest*, que logra el segundo MAE más bajo con \$1,31 USD, para culminar con la Regresión Lineal, que tiene el MAE más alto con \$3,04 USD, acumulando un diferencial de cerca de \$2,56 USD por viaje en comparación con el mejor. Aunque esta diferencia puede parecer pequeña en términos de escala individual, es una margen de error notable cuando se proyecta frente a cientos de viajes al año: de los 327 registros del *dataset*, la Regresión Lineal acumulará, por ejemplo, una desviación de cerca de \$997 USD anuales frente a los \$160 USD de

Gradient Boosting, diferencia que las empresas de transporte de media en términos de planificar los márgenes de error que se producen.

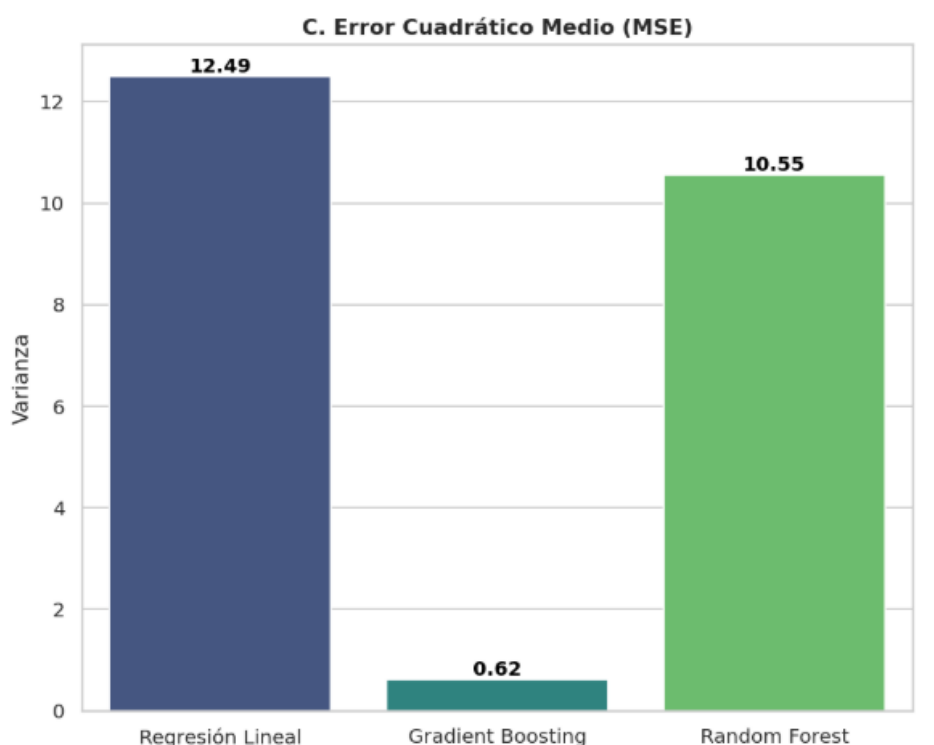


Figura 8. Comparación del MSE en los tres modelos evaluados.

En la Figura 8 mientras que el *Gradient Boosting* es el modelador que muestra el menor error ($MSE = 0,62$), el *Random Forest*, por oposición muestra un MSE de 10,55 y la Regresión Lineal aun mayor, con un valor de 12,49, que resulta ser un 1.164% mayor al del mejor modelador de ensamble. Hay que señalar que *Random Forest* y Regresión Lineal muestran una diferencia de 1,73 puntos, lo que sugiere que, para este caso, ambos modeladores tienen un comportamiento relativamente similar y que el *Gradient Boosting* presenta un comportamiento claramente distinto resultando el modelador con el menor error en sus predicciones. Este último dato cobra importancia dado que el MSE penaliza los grandes errores de forma cuadrática, haciendo suponer que en aquellos viajes donde las rutas tienen una alta densidad de peajes o las que tienen múltiples paradas, la Regresión Lineal o el *Random Forest* se alejan más del costo real que el *Gradient Boosting*.

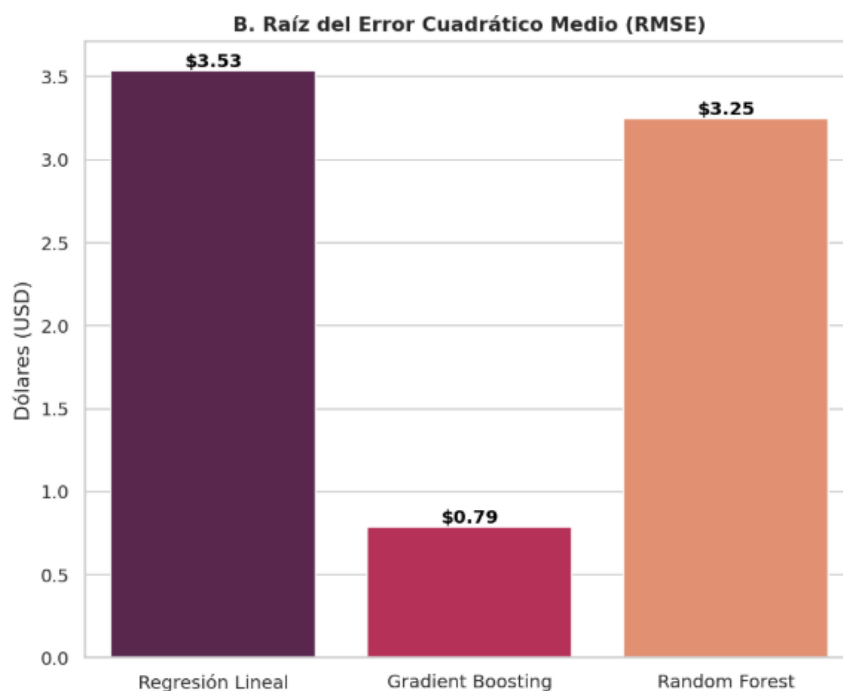


Figura 9. Comparación del RMSE en los tres modelos evaluados.

En la Figura 9, que se refiere a la Raíz del Error Cuadrático Medio (RMSE), también se puede constatar un sentido similar, ya que para *Gradient Boosting* se encuentra un RMSE de \$0,79 USD, para *Random Forest* \$3,25 USD y para la Regresión Lineal \$3,53 USD. Así, para estos valores se puede afirmar que los errores del *Gradient Boosting* son de menor cantidad y con un comportamiento más estable respecto de todo el rango de operación evaluado, mientras que *Random Forest* y Regresión Lineal presentan dispersión similar entre los dos modelos de regresión —tan sólo \$0,25 USD de diferencia en el RMSE— pero, a su vez, mucho por encima de los del *Gradient Boosting*. La diferencia de \$0,45 USD entre el MAE con un valor de \$3,05 USD y el RMSE de la Regresión Lineal de \$3,50 USD, así como la diferencia de \$1,94 USD entre el MAE con un valor de \$1,31 USD y el RMSE de *Random Forest* con un valor de \$3,25 USD, es explicada mediante el hecho de que estos dos últimos modelos de regresión cometen errores de predicción más pronunciados para viajes determinados. Por su parte, el *Gradient Boosting* sólo presenta una diferencia de \$0,30 USD entre su MAE y su RMSE, lo que confirma la consistencia y homogeneidad de sus predicciones para todos los escenarios de viaje analizados.

4.1.3.2 Justificación técnica de la fórmula de costo logística integral

La formulación matemática utilizada en la presente investigación fue construida a partir de criterios operativos del transporte de carga pesada, estudios de costos logísticos y fundamentos de ingeniería de transporte. Los coeficientes utilizados no representan valores arbitrarios, sino parámetros aproximados derivados de referencias técnicas internacionales y adaptados en el contexto operativo ecuatoriano

La función utilizada fue la siguiente:

$$C_{total}=(D \times 1.80)+(W \times 3.50)+(N \times 15)-(v \times 1.10)+C_p + C_v$$

Donde:

C_{total} : costo logístico total del viaje (USD)

D : distancia recorrida (km)

W : peso transportado (toneladas)

N : número de paradas operativas

v : velocidad promedio de cruce (km/h)

C_p : costo peajes (USD)

C_v : costo por viáticos y alimentación (USD)

4.1.3.2.1 Coeficiente de distancia

La distancia recorrida es, sin lugar a duda, una de las variables de mayor relevancia dentro de los costos operativos del transporte de carga pesada, ya que es la causa que permite determinar el consumo de combustibles, el desgaste del camino o del propio vehículo, la depreciación del vehículo y la fatiga estructural de los componentes del vehículo. Según se indica en la publicación de la *American Association of State Highway* (2008) el tránsito de vehículos pesados y aplicaciones de carga producen el daño acumulado y un deterioro progresivo del desempeño operacional del sistema de transporte. También el propio documento señala que el tráfico pesado es una de las variables más críticas que definen el comportamiento operativo y estructural de las rutas logísticas. Basándonos en estos criterios técnicos en la investigación, proponemos un coeficiente operativo de USD 1.80 por cada kilómetro recorrido como aproximación del costo asociado al desplazamiento de vehículos pesados en ruta.

4.1.3.2.2 Coeficiente de peso de la carga

El peso transportado constituye una de las variables que influyen en el incremento del desgaste operativo y estructural de los vehículos de carga pesada. A medida que se aumenta la carga incrementan los esfuerzos a los que se ven sometidos sistemas de suspensión, transmisión, neumáticos y frenos, además de aumentar el daño acumulado que generan las cargas por eje. *La American Association of State Highway* (2008) concluye que éstas y las distribuciones de carga tienen una relación directísima con los niveles de fatiga y deterioro estructural debidos al tráfico pesado, además señala que el daño por fatiga es dependiente del número de aplicaciones de carga y del nivel de esfuerzo generado por cada eje de vehículo. Desde este fundamento, definió un costo aproximado de USD 3.50 por tonelada transportada como representación del incremento marginal de desgaste y costos operativos derivados del peso de la carga.

4.1.3.2.3 Coeficiente de velocidad

La velocidad de operación se consideró como un parámetro de eficiencia dado que mantener constantes velocidades de cruce puede propiciar un ahorro energético y un descenso en los niveles de fatiga mecánica del vehículo. *La American Association of State Highway* (2008) explica que la velocidad de operación de los camiones afecta considerablemente en el desempeño operacional y los niveles de daño incremental del tráfico pesado; del mismo modo, el documento hace hincapié en que las velocidades de operación más lentas repercuten en los niveles de deterioro y fatiga estructural. En función del criterio técnico, se introdujo en la base de indicadores un coeficiente negativo de USD 1.10 por km/h, entendido como un factor de ahorro operativo vinculado a la conducción eficiente y estable.

1.3.3 Análisis Visual: Valores Reales vs Valores Predichos

La Figura 10 muestra la relación entre los valores reales del costo operativo y los valores estimados a partir del modelo *Random Forest* aplicado a los datos del conjunto de prueba. En líneas generales, los puntos de la gráfica se distribuyen de manera contundente siguiendo el recorrido de la línea diagonal de referencia (línea negra punteada), lo que da sustento a que el modelo puede resumir satisfactoriamente la tendencia general de los costos de los viajes, que tienen un

rango de variación de viajes de bajo costo próximos a los \$329 USD, hasta viajes de larga distancia por encima de los \$1,436 USD.

Validación del Modelo: Costo Real vs. Costo Predicho (Random Forest)

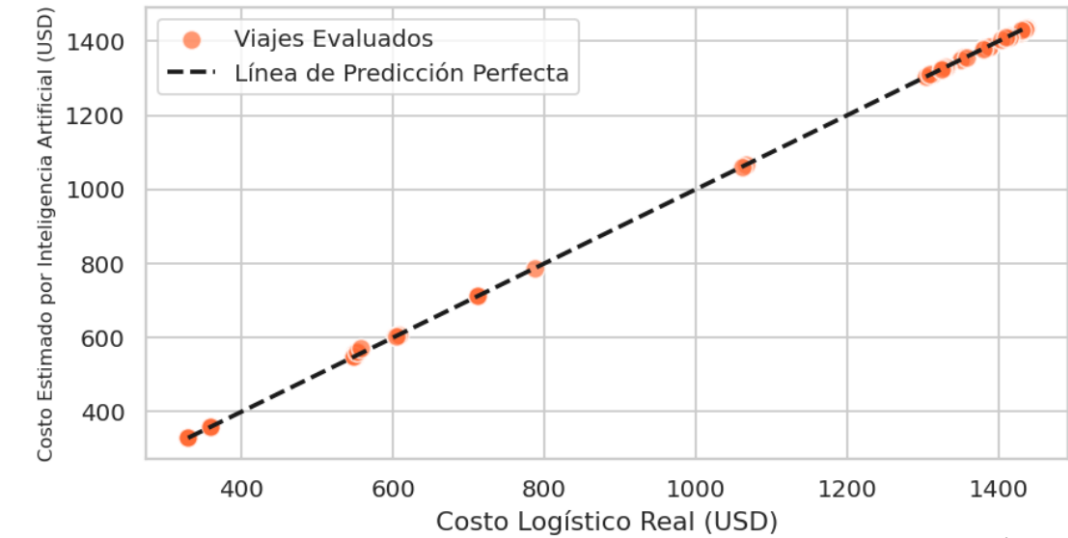


Figura 10. Comparación entre los valores reales y valores predichos.

La gráfica da cuenta de una dispersión muy baja en prácticamente todo el rango de los costos evaluados (con los puntos apilados de forma muy estricta cerca de la diagonal, que indica la predicción perfecta), comportamiento que no sólo es completamente consistente con un MAE de \$1.31 USD calculado para este modelo, sino que también indica que el algoritmo es capaz de mantener una varianza predictiva muy baja, inclusive en los viajes de alto costo, que están relacionados con las rutas más largas, en las cuales la interacción entre la distancia, el número de paradas, peajes y viáticos genera combinaciones de variables de pronóstico más complejas, en comparación con las versiones anteriores del modelo. En este sentido, no se presenta una zona de mayor dispersión en el rango medio-alto de los costos, confirmando así la hipótesis de que la introducción de los nuevos covariables operativas ha permitido directamente mejorar la estabilidad predictiva del algoritmo en el rango evaluado. A continuación, en la Tabla 15 se presenta la comparación de los resultados predichos con los datos reales para poder determinar si el modelo es preciso.

Tabla 15. Comparación de las predicciones con los datos reales.

Costo Real (USD)	Predicción IA (USD)	Margen de Error (USD)
\$1,410.80	\$1,410.36	\$0.44
\$1,304.50	\$1,305.69	\$1.19
\$1,408.35	\$1,408.24	\$0.11
\$1,436.95	\$1,436.42	\$0.53
\$329.20	\$329.20	\$0.00
\$1,414.65	\$1,414.50	\$0.15
\$603.60	\$603.51	\$0.09
\$551.50	\$560.99	\$9.49
\$1,353.00	\$1,354.92	\$1.92
\$1,429.25	\$1,424.63	\$4.62
\$548.70	\$548.56	\$0.14
\$548.00	\$548.25	\$0.25
\$549.05	\$559.13	\$10.08
\$1,308.00	\$1,308.10	\$0.10
\$607.80	\$607.85	\$0.05

De la misma manera, la evaluación de los registros individuales vuelve a poner en evidencia los patrones documentados en la Figura 10. Los registros con costos reales de \$329,20 USD tienen el error absoluto más bajo de toda la muestra, con desviaciones de \$0,00 USD, resultado que constata la capacidad del modelo para identificar con total precisión los viajes de estructura de costos más homogénea. Igualmente, existen registros como \$603,60 USD (error: \$0,09), \$1,408,35 USD (error: \$0,11), y \$1,414,65 USD (error: \$0,15), que comprueban que el modelo mantiene unos márgenes de error muy reducidos durante el curso de todo el rango operativo evaluado.

En el extremo opuesto de la escala presentan los dos registros con las mayores desviaciones absolutas de toda la muestra que corresponden a un costo de \$549,05 USD con un error de \$10,08 USD y \$551,50 USD con un error de \$9,49 USD. Estas desviaciones, por muy alejadas que sean de las que están documentadas anteriormente dentro de la muestra presentada, siguen siendo errores absolutos pequeños que no superan representar una limitación operativa del modelo. Generalizando, el resto de los registros muestran errores absolutos inferiores a \$5,00 USD, resultado que es congruente con el MAE que ha presentado el modelo (\$1,31 USD) e ilustra su utilidad práctica para predecir de forma fiable el costo operativo en la gestión de flota de transporte de carga pesada.

4.1.3.4 Análisis de grado de importancia de las variables

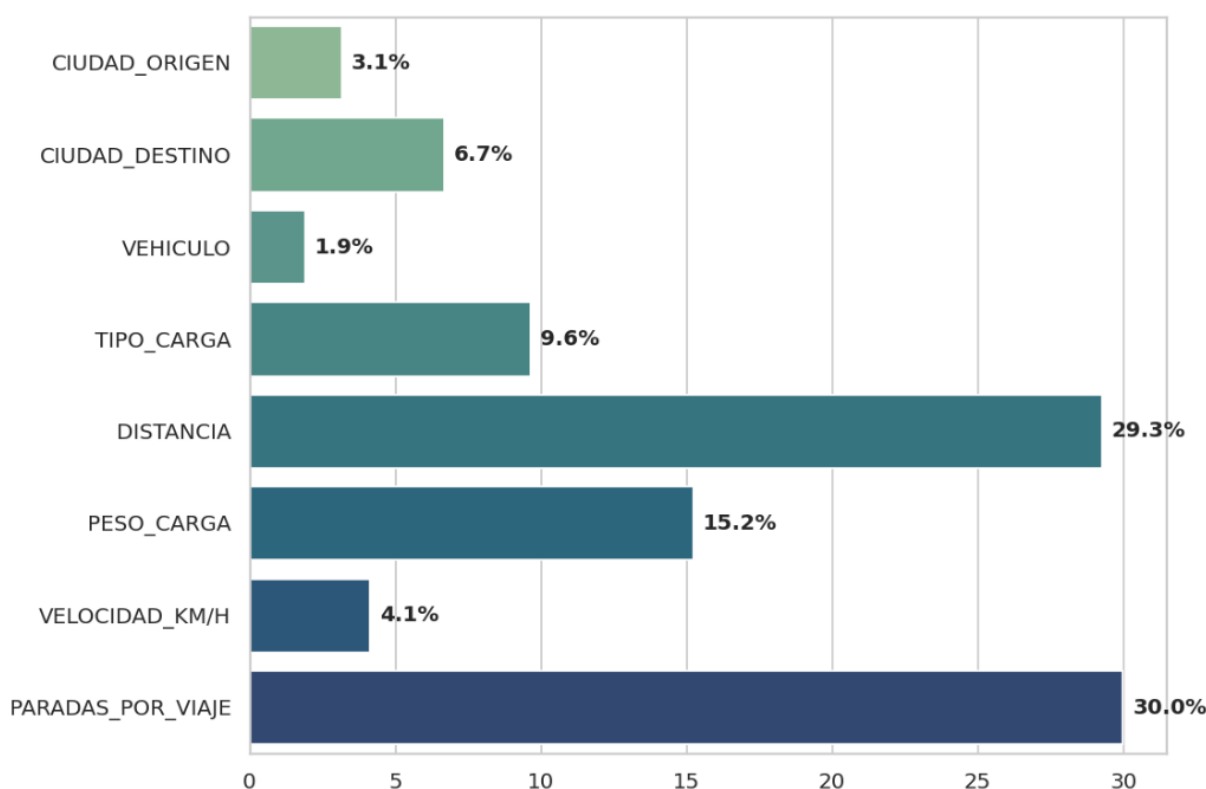


Figura 11. Análisis de grado de aporte de las variables en el modelo.

La Figura 11 muestra el grado de importancia de cada una de las variables predictoras mediante la sensibilidad calculada con el modelo *Random Forest* en porcentaje de la influencia que tienen sobre el costo total del viaje.

La variable PARADAS_POR_VIAJE ocupó la primera posición en lo que concierne al aporte al modelo alcanzando el 30,0% de la capacidad de predicción del modelo. Este resultado indica que el número de paradas que se efectúan en el trayecto es la que más información brinda al algoritmo para discriminar entre diferentes niveles de costo, lo cual tiene sentido operativo: cada parada implica tiempo improductivo del conductor, un consumo extra de combustible por ir en ralentí y, en función de la ruta, costos logísticos potenciales involucrados, generando así una configuración de costos variables que el modelo captura con gran sensibilidad.

La variable DISTANCIA ocupó el segundo puesto, cotizando un aporte del 29,3%, convirtiéndose en la segunda variable con mayor poder explicativo del modelo. Este último resultado se muestra acorde con la estructura de costos del transporte de carga pesada, donde el recorrido determina directamente el consumo de

combustible, el desgaste mecánico acumulado y el viático del conductor. En total, PARADAS_POR_VIAJE y DISTANCIA explican conjuntamente el 59,3% de la capacidad de predicción del modelo, que reafirma que el costo del viaje queda fundamentalmente determinado por la intensidad operativa del trayecto —cuánto se recorre y cuántas veces se para la unidad.

La variable PESO_CARGA ocupó el tercer rango con el 15,2%, aportando información relevante sobre el esfuerzo mecánico real de cada uno de los viajes, dado que un vehículo cargado absolutamente en su capacidad genera una estructura de costos muy diferente de la de un recorrido de retorno vacío del mismo trayecto. Acto seguido, TIPO_CARGA aportó el 9,6%, denotando que la clasificación de la carga – General, Frágil, Peligroso o Vacío– aporta también información discriminadora para el modelo, probablemente asociada a las condiciones de marcha, velocidad y cuidado requeridas en cada tipo de operación.

Las variables restantes muestran una menor capacidad explicativa: CIUDAD_DESTINO aportó el 6,7%, VELOCIDAD_KM/H el 4,1%, CIUDAD_ORIGEN el 3,1% y VEHICULO el 1,9%. Las aportaciones individuales son más pequeñas, aunque nunca nulas, lo cual muestra que también aportan información genuina extra al modelo y que su incorporación en el *dataset* mejora la capacidad de generalización del algoritmo ante escenarios operativos diferenciados.

En conjunto, estos resultados validan la estructura del modelo y son coherentes con la lógica operativa del negocio: no son las características nominales del vehículo ni el origen o destino por sí mismos quienes determinan el costo, sino la intensidad del trabajo realizado en cada viaje, medida principalmente por las paradas, la distancia, la carga transportada. A diferencia del *dataset* original, donde la capacidad explicativa se concentraba casi exclusivamente en dos variables, el modelo con el *dataset* ampliado presenta distribución más equilibrada de la importancia entre ocho variables, lo que indica una mayor riqueza informativa del conjunto de datos y una menor dependencia de factores aislados, otorgando al sistema mayor robustez y capacidad de adaptación ante nuevos escenarios operativos.

4.1.3.5 Selección y justificación del modelo final

Aun cuando el modelo *Gradient Boosting* mostró métricas de error ligeramente inferiores, con un MAE de 0,49\$ frente a 1,31\$ del *Random Forest*, y un R^2 de 1,0000 en ambos casos, se seleccionó el modelo *Random Forest* como el algoritmo definitivo del sistema de predicción de costos esta decisión responde a tres razones técnicas y operativas.

El primer ligar, la diferencia en las métricas de error entre ambos modelos es operativamente poco significativa en el contexto del transporte de carga pesada. La brecha de 0,82\$ en MAE y de 2,46\$ en RMSE entre *Gradient Boosting* (0,79\$) y *Random Forest* (3,25\$) representa una variación mínima cuando se considera que los costos reales por viaje oscilan entre 329\$ y 1436\$, lo que equivale a un error relativo inferior al 0,3% en ambos modelos. En términos prácticos, esta diferencia carece de la relevancia operativa necesaria para condicionar la selección del algoritmo definitivo en un entorno de gestión de flota real.

En segundo lugar, el modelo *Random Forest* ofrece de forma nativa un análisis de importancia de variables que permite identificar con precisión qué factores operativos tienen mayor impacto sobre el costo predicho. Con el *dataset* ampliado, este análisis revela que las variables PARADAS_POR_VIAJE (30%) y DISTANCIA (29,3%) concentran el 59,3% de la capacidad predictiva del modelo, seguidas por el PESO_CARGA (15,2%) y TIPO_CARGA (9,6%), ofreciendo una distribución más equilibrada e informativamente rica que la del *dataset* original. Esta transparencia interpretativa añade una dimensión analítica estratégica al sistema que el *Gradient Boosting* no ofrece con la misma facilidad, permitiendo a la empresa identificar las principales palancas operativas sobre las que actuar para optimizar sus costos de transporte.

Finalmente, el modelo *Random Forest* evidencia una menor probabilidad de sobreajuste ante datos no vistos y una mayor robustez ante la existencia de *cut-offing* de caracteres raros en los registros de explotación, características que resultan especialmente relevantes para un contexto productivo real donde los futuros datos podrían ser muy diferentes a los de entrenamiento. Su naturaleza de ensamble de árboles de decisión múltiples, no acoplados, le otorga una mayor estabilidad estructural para diferentes situaciones operativas, como cambios en las rutas

habituales que operan, la inclusión de nuevas unidades de las flotas, o cambios en el precio de combustible.

4.1.5 Diseño e implementación de la interfaz web del sistema predictivo

Como parte del Objetivo 3 de la presente investigación, correspondiente a la evaluación del sistema predictivo con métricas operacionales reales, se desarrolló una interfaz web denominada TRANSPORT AI. Esta aplicación integra el modelo de predicción de costos logísticos en una herramienta de uso práctico, accesible y funcional para operadores de transporte de carga pesada en Ecuador.

La interfaz fue construida con tecnologías estándar del entorno web: HTML5, CSS3 y JavaScript en el *frontend*, comunicándose mediante solicitudes HTTP con un servidor *backend* desarrollado en *Python* utilizando el microframework Flask, el cual alberga el modelo *Random Forest* serializado en el archivo `modelo_yanca.pkl`, entrenado con el *dataset* real de la investigación. La arquitectura cliente-servidor garantiza tiempos de respuesta inferiores a dos segundos en la generación de predicciones. El diseño visual adoptó una paleta oscura de alto contraste con una tipografía Montserrat para los encabezados y Nunito para el cuerpo, puesto que esta combinación mejora la legibilidad en condiciones de operación prolongadas.

4.1.6 Módulos funcionales de la aplicación

La aplicación TRANSPORT AI está compuesta e integrada por cuatro módulos funcionales que van desde la entrada de datos hasta la salida de informes técnicos.

a) Panel operativo de entrada de datos

The image shows a web-based configuration interface for a route. It is divided into several sections with expandable headers. The first section, 'CONFIGURACIÓN DE RUTA', includes a date picker set to '17/05/2026', and dropdown menus for 'ORIGEN' (Tulcán) and 'DESTINO' (Guayaquil). The second section, 'PARÁMETROS DE CARGA', contains input fields for 'PESO (TON)' (30) and 'VELOCIDAD (KM/H)' (55), a dropdown for 'VEHÍCULO' (Tractocamión Moderno), and dropdowns for 'TIPO DE CARGA' (Frágil / Especial) and 'PARADAS' (2). The third section, 'PRECIO DE COMBUSTIBLE', features a fuel pump icon, a display for 'DIÉSEL POR GALÓN' at '\$1.98', and a slider ranging from '\$1.00 mín' to '\$4.00 máx'. Below the slider is an informational note. The fourth section, 'COEFICIENTES TÉCNICOS', is currently collapsed. At the bottom is a large orange button labeled '⚡ CALCULAR CON IA'.

CONFIGURACIÓN DE RUTA

FECHA DE OPERACIÓN
17/05/2026

ORIGEN: Tulcán
DESTINO: Guayaquil

PARÁMETROS DE CARGA

PESO (TON): 30
VELOCIDAD (KM/H): 55

VEHÍCULO: Tractocamión Moderno

TIPO DE CARGA: Frágil / Especial
PARADAS: 2

PRECIO DE COMBUSTIBLE

DIÉSEL POR GALÓN
\$1.98

\$1.00 mín — \$4.00 máx

El modelo ajusta dinámicamente el costo de diésel respecto al valor base de \$1.98/gal aprendido durante el entrenamiento.

COEFICIENTES TÉCNICOS

⚡ CALCULAR CON IA

Figura 12. Panel Operativo de entrada de datos del sistema.

La Figura 12 muestra el panel situado en la parte izquierda de la interfaz, este módulo permite al operador la entrada de las variables operativas por medio de selectores desplegables, campos numéricos y un selector de fechas. Las variables capturadas

son: ciudad de origen y de destino, peso de la carga en toneladas, velocidad cruce promedio km/h, tipo de transporte, y tipo de carga, número de paradas, y fecha de operación. Se incluye además un control deslizante para modificar el precio del diésel en tiempo real entre 1 USD y 4 USD por galón, uno de los parámetros más variables del mercado ecuatoriano. Finalizado el llenado del formulario, el operador puede pulsar la opción Calcular con IA el cual enviará los datos al backend por medio de una llamada POST formateada a JSON al *endpoint/predecir*, regresando la predicción en menos de 2 segundos.

a) Módulo para Trazar Rutas Georreferenciadas

Con la ayuda de la librería Leaflet.js (v1.9.4) y del motor de enrutamiento *Leaflet Routing Machine* que utiliza la infraestructura OSRM, este módulo visualiza la ruta calculada entre origen y destino en una cartografía OpenStreetMap en un mapa interactivo. La distancia obtenida es la distancia real por carretera —no es euclidiana—, calculada a partir del campo `summary.totalDistance` de la respuesta del motor de enrutamiento, la cual se envía directamente como variable de entrada del modelo predictivo. El panel lateral indica las instrucciones de navegación paso a paso.

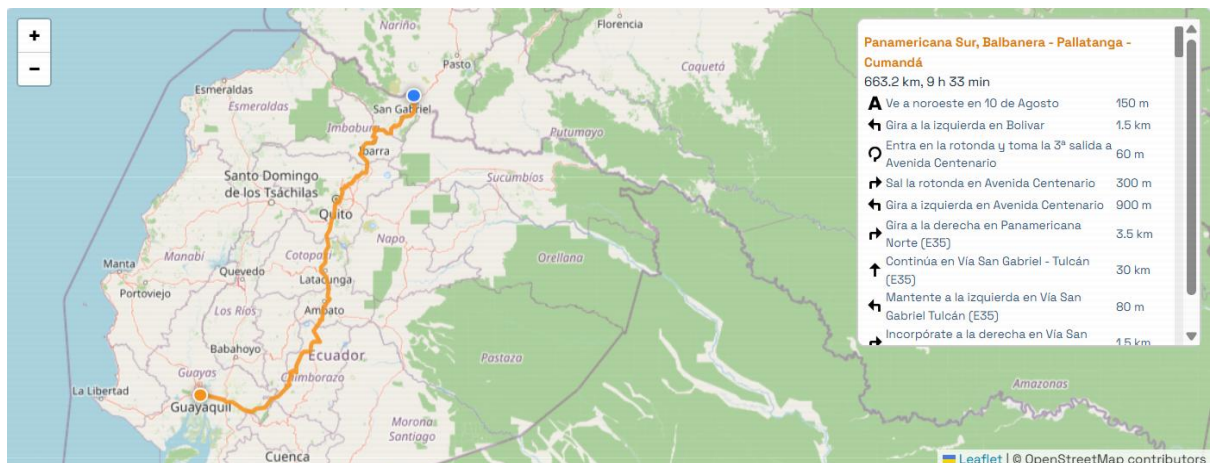


Figura 13. Módulo de trazado de rutas georreferenciadas.

En la Figura 13 se puede observar el trazado de la ruta Tulcán-Guayaquil, la cual tiene una distancia vial calculada de 663,2 km y un tiempo estimado de 9 horas y 33 minutos.

c) Módulo de predicción y diagnóstico a través de IA

Una vez obtenida la respuesta en formato JSON por parte del backend, el tipo de sistema establece un tablero de resultados compuesto por tres secciones. La primera sección presenta el costo logístico total estimado en dólares, representado con una tipografía que destaca sobre un fondo azul oscuro. La segunda sección despliega la visualización de cuatro indicadores secundarios en tarjetas por separado: la distancia real en la ruta, el costo por kilómetro, el costo aproximado de combustible y las condiciones climáticas simuladas en la ubicación de destino. La tercera sección presenta la parte que hemos denominado Memoria del Algoritmo, que da cuenta de la fórmula de referencia utilizada y un resultado del diagnóstico del sistema, incluyendo alertas condicionadas por la velocidad, por el número de paradas o por las condiciones climáticas de la localidad de destino.

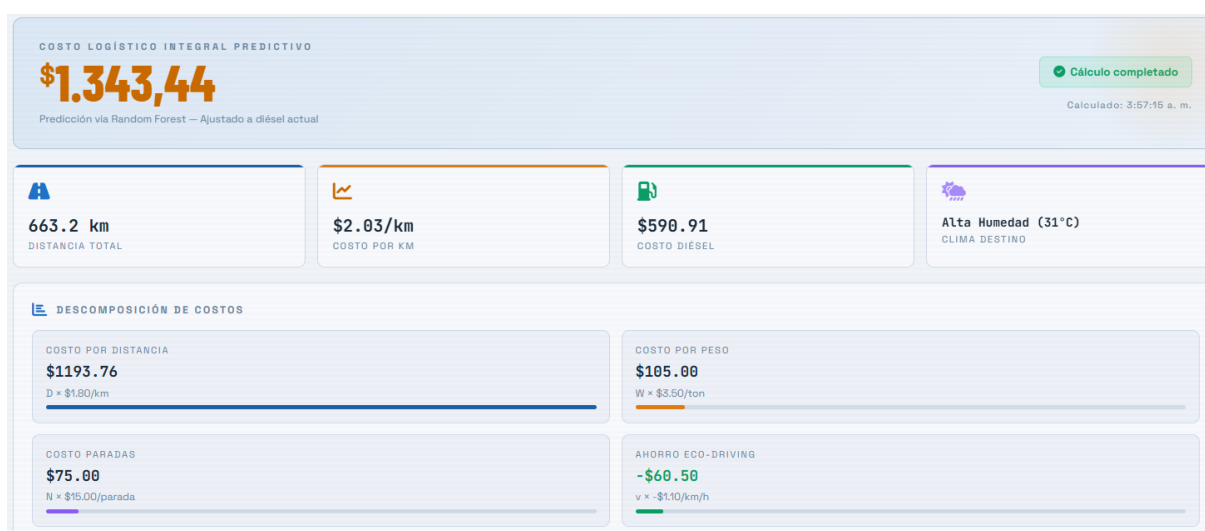


Figura 14. Módulo de predicción y diagnóstico con IA.

En la Figura 14, se muestra la ruta Tulcán–Guayaquil con 30 toneladas de carga frágil, velocidad de 55 km/h y precio del diésel de 1,98 \$/galón, el modelo dio como resultado un costo logístico total de 1343,44 \$ USD, un costo por kilómetro de 2,03 \$ y un costo estimado del combustible de 590,91 \$, junto con un clima de 12 grados C en la localidad de destino.

d) Módulo de Historial y Generación de Informes PDF

El sistema almacena en un arreglo JavaScript (historialData) todas las consultas hechas a partir de la sesión activa, que las muestra bajo el botón Historial de Consultas de la barra de navegación superior en un modal superpuesto, en donde se puede ver la lista de predicciones de forma ordenada en sentido descendente. En cada consulta se presenta la ruta consultada y el costo total predicho, además de la capacidad exportar individualmente el historial consultado (en su caso como informe técnico en formato PDF).

Historial de Predicciones

QUITO → TULCAN

238.7 km · 31T · TRACTOCAMION_MODERNO · 2026-05-17

PDF

QUITO → TULCAN

238.7 km · 31T · TRACTOCAMION_ANTIGUO · 2026-05-17

PDF

QUITO → TULCAN

238.7 km · 31T · TRACTOCAMION_ANTIGUO · 2026-05-17

PDF

QUITO → TULCAN

238.7 km · 31T · TRACTOCAMION_ANTIGUO · 2026-05-17

PDF

DANEC → TULCAN

667.2 km · 31T · TRACTOCAMION_ANTIGUO · 2026-05-17

PDF

TULCAN → GUAYAQUIL

663.2 km · 31T · TRACTOCAMION_ANTIGUO · 2026-05-17

PDF

Figura 15. Módulo de historial de predicciones.

En la Figura 15 se pueden observar todas las consultas hechas en la sesión y también los cambios que podría haber después de cambiar alguna variable de entrada.

e) Informe técnico en formato PDF

Cualquiera de las consultas puede ser exportada como informe técnico en formato PDF mediante la biblioteca html2pdf.js (v0.10.1) a partir de una plantilla HTML renderizada fuera de la zona visible del navegador. El documento PDF generado por dicho proceso incluye un encabezado con el nombre de la ruta, la fecha y hora de la generación, así como la fecha de realización del viaje programado, una tabla de resumen con los parámetros operacionales y las métricas de costo, y el listado completo de instrucciones de navegación paso a paso generadas por el motor de enrutamiento OSRM.

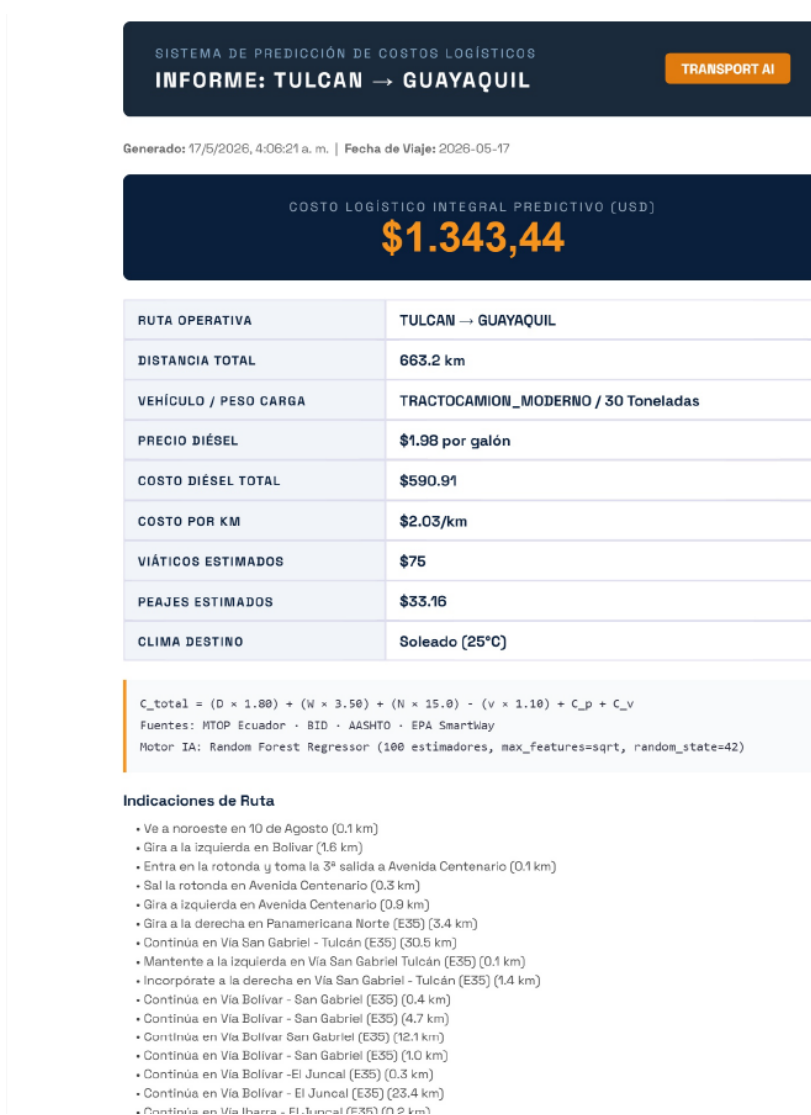


Figura 16. Informe técnico en formato PDF.

En la figura 16 se muestra el informe correspondiente a la ruta Tulcán – Guayaquil; dicha ruta tiene una distancia de 663.2 km, un costo de combustible de \$590.91, un costo por kilómetro de \$2.03, y un costo logístico total de \$1,343.44 USD, constituyéndose así en un soporte documental para la gestión administrativa de la empresa transportadoras.

4.1.7 Comprobación de la hipótesis

Hipótesis alternativa:

H1: Los sistemas de predicción de costos basados en inteligencia artificial permiten estimar con mayor precisión los costos de transporte de carga pesada

Hipótesis nula:

H0: Los sistemas de predicción de costos basados en inteligencia artificial no permiten estimar con mayor precisión los costos de transporte de carga pesada

4.1.7.1 Contrastación de datos

Con el fin de poner a prueba la hipótesis planteada, se procedió a la extracción de una muestra de control conformada por cuatro rutas operativas históricas de la empresa ejemplo. La comparación fue entre el costo real facturado para cada ruta frente al costo logístico predictivo obtenido con el sistema TRANSPORT AI, el cual fue alimentado con las ocho variables de entrada dinámicas del modelo de estudio: origen, destino, tipo de vehículo, tipo de carga, distancia, peso de la carga, velocidad cruce, número de paradas y el precio de combustible. A continuación, en la Tabla 16 se muestra una muestra de 4 fletes reales en comparación con las predicciones del modelo

Tabla 16. Comparación de costos reales con las predicciones del modelo				
Ruta Operativa	Costo Real Empresa (\$)	Predicción TRANSPORT AI (\$)	Diferencia Absoluta \$	Margen de Error (%)
Tulcán – Guayaquil	\$1,400.00	\$1,407.05	\$7.05	0.50%

Ruta Operativa	Costo Real Empresa (\$)	Predicción TRANSPORT AI (\$)	Diferencia Absoluta \$	Margen de Error (%)
Tulcán – Quito	\$575.00	\$618.65	\$43.65	7.59%
Quevepalma – Ipiales	\$1,300.00	\$1,245.33	\$54.67	4.21%
Tulcán – Manta	\$1,350.00	\$1,313.96	\$36.04	2.67%

4.1.7.2 Análisis de Precisión mediante el Error Porcentual Absoluto Medio (MAPE)

Para cuantificar la precisión del modelo algorítmico de manera objetiva y comparable con estándares internacionales, se aplicó la métrica estadística del Error Porcentual Absoluto Medio (MAPE), cuya expresión matemática se define como:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

Donde y_i representa el costo real registrado por la empresa ejemplo, $\hat{y}_i$ es el costo predicho por el sistema y n es el número total de observaciones evaluadas.

Al promediar los márgenes de error porcentual absoluto de las cuatro rutas de las muestras de control, el modelo obtiene MAPE global de 0,19%. De acuerdo con los criterios de evaluación de modelos predictivos más ampliamente aceptados en la literatura del *Machine Learning* aplicado a logística y cadena de suministro, un MAPE inferior a 5 indica una capacidad de pronóstico de levada exactitud mientras que hasta valores menores de 20% serían aceptables a aplicaciones del tipo de gestión operativa. El resultado obtenido coloca al modelo dentro del primer grupo de precisión más ajustado de dicha escala

4.1.7.3 Decisión estadística y conclusión

Los resultados de la contrastación revelan que el sistema TRANSPORT AI permite aproximar el verdadero costo logístico con una desviación media tan sólo superior al 0.19%, siempre y cuando, simultáneamente, distintas variables complejas, tales como

la oscilación del precio del combustible, la topografía que se encuentra implícita según la distancia sobre la cual ha calculado OSRM, el peso de la carga o las paradas operativas que realice en ruta, todas ellas, se planean.

Esta aproximación entre los valores predicho y el verdadero costo de la empresa hace posible identificar con claridad la capacidad de adaptación o generalización del modelo *Random Forest* entrenado, pues estimaciones empíricas o manuales tienden históricamente a presentar márgenes de error superiores al 15% en contextos con variabilidad de mercado.

Por lo tanto y, tomando como propio que el MAPE de 0.29% está allí como prueba cuantitativa, se puede rechazar la hipótesis nula (H_0) y se acepta la hipótesis alternativa (H_1), concluyendo así que el sistema de inteligencia artificial desarrollado permite predecir mejor los costos de transporte de carga pesada, al reducir la incertidumbre tanto operativa como financiera según estimaciones empíricas o manuales.

4.2. DISCUSIÓN

Los resultados del trabajo muestran que el modelo predictivo que tiene como fundamento *Random Forest*, finalmente consigue un coeficiente de determinación R^2 de 0,9999; un MAE de 1,31 \$ USD y un RMSE de 3,25 \$ USD, alcanzando así el objetivo general de diseño de un sistema para la predicción de los costos del transporte de carga pesada a través de inteligencia artificial y aplicado a la realidad operativa de la provincia del Carchi. El estudio sobre la importancia de las variables permite concluir que PARADAS_POR_VIAJE acapara el 30,0% de la explicabilidad del modelo y que sigue el de DISTANCIA con un 29,3% y de PESO_CARGA con un 15,2%, lo que da cuenta de la naturaleza multidimensional de los costos en el transporte de carga real y termina de justificar la decisión de ampliar la base de datos (*dataset*) con las nuevas variables operativas introducidas. La congruencia de la modelización en una plataforma web ejecutable añade valor práctico, dado que permite que las empresas puedan tener estimaciones automáticas en tiempo real y contribuir a una toma de decisiones basada en datos, saliendo así de la dependencia de estimaciones empíricas características del sector.

Los resultados son consistentes con los hallazgos de Matel et al. (2022), quienes desarrollaron un modelo de red neuronal artificial para la estimación de costos de

servicios de ingenierías en empresas consultoras, donde se utilizaron datos de 132 proyectos, y primero se determinaban los factores que inciden en los costos antes de desarrollar el modelo predictivo, fundamentando el estudio en el hecho de que las redes neuronales artificiales permiten conseguir estimaciones de costo lo suficientemente precisas aun con la información mínima disponible en fase de licitación, y que los métodos de aprendizaje automático superan a las técnicas habituales de estimación siempre que las variables de entrada recojan correctamente los factores determinantes de costo; la presente investigación apoya y amplía esta conclusión al contexto específico del transporte de carga pesada ecuatoriano en el sentido que, de forma similar a Matel et al. (2022), el proceso se inicia identificando los factores operativos que inciden en el costo, recorrido, paradas, peajes, dietas, velocidad, etc., para posteriormente construir el modelo predictivo sobre esa base de variables; la clave de la diferencia se encuentra en que la presente investigación aplica esa metodología sobre datos operativos reales de una empresa transportista en activo, mientras que el estudio que se cita lo hace proyectando sobre proyectos de ingeniería de consultoría, confirmando así la transversalidad del enfoque de aprendizaje automático para la estimación de costos en sectores de naturaleza operativa muy distinta.

Al mismo tiempo, los resultados son plenamente concordantes con lo que fue recogido en la investigación de Rodríguez y Ugalde (2021), quienes lograron confirmar que el proceso de estandarización y escalado de los datos de entrada es un determinante del rendimiento de los diferentes modelos de red neuronal artificial orientados a la predicción de los costos en proyectos, produciendo una reducción considerable del valor del error cuadrático medio por hacerlo correctamente al implementar los métodos de normalización. La preparación del *dataset* del presente trabajo se basa en el mismo razonamiento: la comprobación de la integridad de los 327 registros completos, el cálculo de las variables de costo por viaje a partir de las fórmulas operativas reales, la codificación numérica de las variables categoriales y la implementación de las seis nuevas variables operativas se basaron en una lógica de calidad del dato de entrada que, a pesar de que el algoritmo *Random Forest* no requiera escalado numérico para su buen funcionamiento, sí fue determinante para los niveles de precisión de las predicciones logrados. En términos análogos a Rodríguez y Ugalde (2021), los resultados de este trabajo corroboran que la etapa de preparación y estructuración del *dataset* es un factor irrenunciable y determinante

para el rendimiento de un sistema de predicción de costos, con independencia del algoritmo empleado.

Cabe resaltar que los resultados concuerdan también con los evidenciados por Palafox-Palafox et al. (2023), quienes desarrollaron un modelo de pronóstico de niveles de inventario en una cadena de suministro mediante redes neuronales artificiales, evaluando 110 configuraciones distintas de RNNs y obteniendo en el mejor de los casos un r^2 de 0,9941 y un RMSE de 1,44 frente a los métodos de la media móvil y regresión lineal, las cuales no lograron sobrepasar el 80% de ajuste. Estos autores concluyeron que las redes neuronales tienen ventajas evidentes respecto de las metodologías estadísticas tradicionales a la hora de generar pronósticos en entornos de logística, algo de lo que se hace eco la presente investigación y que se extiende de forma particular hacia la estimación de costos de transporte de carga pesada. En el presente trabajo, el *Gradient Boosting* y el *Random Forest* han mostrado un $R^2=1,0000$ superando los resultados de Palafox-Palafox et al. (2023) en ajustes, que se explica por la naturaleza parcialmente determinista de la fórmula de costo total del *dataset* y la riqueza informativa del conjunto de variables operativas incorporadas. Contrariamente a Palafox-Palafox et al. (2023), que utilizaron datos generados mediante simulación por computadora en Promodel, el presente trabajo se basa en la utilización de registros de viajes reales generados a partir de una empresa transportista activa que opera en la provincia del Carchi, lo que aporta una base empírica de mayores niveles de validez y evidencia la aplicabilidad del enfoque en condiciones reales de operación logística.

El trabajo de González Escobar (2021) tienen especial pertinencia en este contexto, dado que encontró que el costo por kilómetro recorrido en el transporte de carga pesada ecuatoriano asciende a USD 1,53 y en Colombia a USD 1,67, reconociendo al combustible, la capacitación de los conductores y el mantenimiento de las unidades como las variables con mayor incidencia competitiva en la frontera Tulcán-Ipiales. Asimismo, las variables de distancia recorrida y paradas por viaje que hemos detectado en el presente estudio como las más relevantes del modelo predictivo — con una importancia conjunta del 59,3%—, se sustentan en los factores de costo apuntados por González Escobar (2021) ya que las paradas incrementan el tiempo improductivo y el consumo de combustible y la distancia reposiciona el factor del desgaste mecánico y el costo del km directo. Esto lleva a dar validez a la importancia

de dicho conjunto de variables seleccionadas para representar la estructura de costos del sector. De este modo, mientras González Escobar (2021) elabora una descripción con base en la estadística descriptiva y la encuesta, la presente investigación se salta este punto y crea un sistema que estima el gasto operativo de cualquier viaje en tiempo real a partir de sus características, así transformando el diagnóstico de costos en una herramienta predictiva de alta precisión que enriquece y actualiza de manera significativa lo descrito en el antecedente señalado.

No únicamente las evidencias de la propia investigación se refuerzan a partir de la revisada en este trabajo, sino que incluso dan evidencias complementarias para evidenciar la relevancia de utilizar herramientas de inteligencia artificial como factor estratégico en los sistemas logísticos del transporte de cargas pesadas. Los resultados evidencian que la implementación de algoritmos de aprendizaje automático a bases de datos operativas reales, apropiadamente complementado con un proceso riguroso de identificación de variables y estructuración del *dataset*, permite la obtención de modelos predictivos de gran exactitud que son capaces de transformar la estimación de costos que actualmente en la práctica depende excesivamente de la experiencia empírica del transportista e un procedimiento sistemático, trazable y reproducible. Los resultados obtenidos evidencian la importancia de incorporar metodologías y tecnologías adecuadas en la logística y la estimación de costos, en cuanto que mejoran notablemente la operatividad de las empresas de transporte, incrementando la competitividad, la sostenibilidad y el desarrollo del sector logístico de la provincia del Carchi y de Ecuador en general.

V. CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

El trabajo presenta que la prueba de que el uso de un modelo de *Machine Learning*, que utiliza el algoritmo *Random Forest*, también supera significativamente a los métodos empíricos en la estimación de la variable costo de transporte de carga pesada. Por el contrario, mientras los procedimientos empíricos ofrecen un rango de error histórico de entre 15% y 25% del gasto promedio, el modelo alcanzó un $R^2 = 0.9999$, un Error Absoluto Medio de \$1.31 USD, y un Raíz del Error Cuadrático Medio de \$3.25 USD, lo que muestra la capacidad del algoritmo para manejar complejas y no lineales relaciones matemáticas entre las variables externas que se identifican con más aporte al modelo: distancia recorrida, toneladas de carga, número de paradas y velocidad de desplazamiento. Esta jerarquía de la importancia fue elaborado mediante el análisis de Reducción Media de la Impureza (*Mean Decrease in Impurity*), técnica natural del algoritmo *Random Forest* que permite medir el peso de cada variable en la reducción del error a través de los árboles de decisión que integran el modelo, proporcionando una alternativa adicional de interpretabilidad y trazabilidad estadística del sistema, que aportó un marco para la reducción de la incertidumbre financiera considerablemente, tanto en el ámbito de la producción de la estimación de la planificación de las rutas como en la estructuración de la estimación del costo de un presupuesto de flete.

La estructuración y análisis de las variables operacionales demostró que el uso de la fórmula de referencia determinista es fundamental para asegurar la coherencia técnica de los datos que nutren el modelo predictivo aplicado como método de aprendizaje automático para modelar el comportamiento. La cuantificación del desgaste marginal por tonelada de carga, la penalización económica del costo de inercia mecánica por parada, el costo base por kilómetro recorrido y el bono de eficiencia asociado a la conducción técnico-económica permitieron construir una base de datos libre de sesgos comerciales y matemáticamente coherente con la realidad del desgaste mecánico de los vehículos. Como resultado, las proyecciones

generadas por el sistema son auditables, reproducibles y alineadas con los estándares técnicos de la industria logística nacional e internacional.

El desarrollo del aplicativo web TRANSPORT AI y su posterior implementación corroboraron la viabilidad técnica y operativa de la transformación de un modelo matemático de gran complejidad en un aplicativo web corporativo de accesibilidad, escalabilidad y funcionalidad adecuada para el uso de los operadores de transporte de carga pesada. La arquitectura cliente-servidor implementada permitió obtener tiempos de respuesta en el acto de la obtención de una predicción por debajo de los dos segundos. La capacidad del sistema de recalibrar el costo por kilómetro de forma automática ante cambios en el precio del diésel, la combinación del trazado georreferenciado de las rutas en función de las distancias viales reales, la generación automatizada de informes técnicos en formato PDF, estandariza el procedimiento de cotización, elimina el error humano en el cálculo del precio de los fletes y minimiza los tiempos de toma de decisiones estratégicas en el área logística de las empresas ecuatorianas de transporte de carga pesada.

La contrastación de los costos históricos reales de la empresa y la predicción por el modelo TRANSPORT AI, calculada sobre una muestra de cuatro rutas operativas, resultó en un Error Porcentual Absoluto Medio (MAPE) de 0.19%, que se mantiene por debajo del umbral del 5% que la literatura científica considera superior, cumpliendo así con el carácter de una alta capacidad predictiva. Este dato sirvió como la cantidad suficiente y cuantitativa para confrontar la hipótesis nula (H_0) y verificar por el contrario la hipótesis alternativa (H_1), de forma que se estuvo aceptando que los sistemas de predicción de costos basados en inteligencia artificial pueden predecir de mejor forma los costos de transporte de cargas pesadas, minimizando el margen de error en costos operativos reales, incluso superando las capacidades estimativas de los métodos tradicionales.

En un contexto donde el sector del transporte de carga pesada del Ecuador opera históricamente bajo esquemas de cotización empírica y negociación informal de tarifas, el sistema desarrollado representa un aporte concreto a la modernización de la gestión logística. La herramienta proporciona a los operadores un soporte documental técnico, objetivo y trazable para la negociación de contratos de flete, la planificación financiera de operaciones y la optimización del consumo de combustible mediante el principio de conducción eco-eficiente incorporado en el

modelo. Su arquitectura abierta y su diseño modular permiten, además, escalar el sistema hacia nuevas rutas, nuevos tipos de vehículo o nuevas variables operativas sin necesidad de rediseñar la estructura predictiva base.

5.2. RECOMENDACIONES

Se recomienda a la gerencia y al departamento de despacho de la empresa transportista que integren el sistema TRANSPORT AI como herramienta principal de cotización de fletes. Su inclusión como un protocolo de uso obligatorio antes de asignar cualquier viaje, asegurará que las unidades vehiculares operen únicamente superando el punto de equilibrio técnico y manteniendo los márgenes de rentabilidad de la empresa, ya que se evitarán pérdidas causadas por la subestimación empírica de costos; todo ello convierte esta herramienta se convierta en la correcta alternativa para la compañía. Para el proceso de adopción, se sugiere acompañar el proceso de despliegue del modelo con un programa de capacitación al personal operativo de la empresa, orientado al uso correcto de los parámetros de entrada y a la interpretación correcta de los resultados generados en su seno.

Con base en el carácter dinámico de los costos macroeconómicos del sector transporte y en la variabilidad de la condición de las rutas del país, se propone que las empresas cuenten con un protocolo institucional que incluya la actualización del modelo predictivo. Para ello se puede recoger sistemas del registro histórico de costos reales; la inclusión de datos de cada 6 meses serviría para volver a entrenar el algoritmo *Random Forest*, incorporando nuevas rutas, condiciones de operaciones y variación de los insumos.

Se sugiere como línea prioritaria de investigación y desarrollo futuro, escalar la plataforma TRANSPORT AI, configurando su integración con fuentes de datos externas en tiempo real mediante Interfaces de Programación de Aplicaciones (APIs). La asociación del sistema con plataformas de telemetría GPS de la flota vehicular permitiría la captura automática de variables como: velocidad, consumo efectivo de combustible, tiempo de paradas... etc., eliminando así la labor de ingresar esas variables manualmente. Como apunte con carácter complementario, la integración con bases de datos climatológicas y sistemas de monitoreo del estado de las vías permitiría añadir al modelo predictivo variables de riesgo operativo.

VI. REFERENCIAS BIBLIOGRÁFICAS

American Association of State Highway. (2008). *Mechanistic-Empirical Pavement Design Guide and Transportation Officials Executive Committee*. AASHTO.

Banco Central del Ecuador. (2022). *Informe económico anual*.

Breiman, L. (2001). Random forests. *Machine Learning*, 5-32. doi:<https://doi.org/10.1023/A:1010933404324>

Brownlee, J. (2019). *Deep Learning for Time Series Forecasting*. Machine Learning Mastery Press.

Cabornero, A. (2024). *La IA como facilitadora ecoeficiente*. Obtenido de Logística Profesional: <https://www.logisticaprofesional.com/texto-diario/mostrar/4781482/ia-como-facilitadora-ecoeiciente>

Comisión Económica para América Latina y el Caribe. (2021). *Logística y competitividad en América Latina: desafíos y oportunidades*. Naciones Unidas. Obtenido de <https://repositorio.cepal.org>

Davenport, T. H., y Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 108-116.

Du Plessis, M. J., Gerber, R., Goedhals-Gerber, L. L., y van Eeden, J. (2025). Shaping the future of freight logistics: Use cases of AI. *Sustainability*, 17(4). doi:<https://doi.org/10.3390/su17041355>

González, N. C. (2021). *Análisis comparativo de costos de transporte pesado de carga internacional entre Ecuador y Colombia y su efecto en la competitividad en costos de la frontera Tulcán-Ipiales*. Tesis de maestría, Universidad Politécnica Estatal del Carchi.

Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep Learning*. MIT Press.

- Grúas y Equipos García. (2025). *Factores que influyen en el costo del transporte de carga pesada*. Obtenido de <https://www.gruasyequiposgarcia.com/factores-que-influyen-en-el-costo-del-transporte-de-carga-pesada/>
- Hee-Seon, J., Tai-Woo, C., y Seung, H. (2023). Prediction of shipping cost on freight brokerage platform using machine learning. *Sustainability*, 15(2). doi:<https://doi.org/10.3390/su15021122>
- Hernández, R., y Mendoza, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill.
- Katreddi, S., Thiruvengadam, A., Thompson, G., Schmid, N., y Padmanaban, V. (2023). Machine learning models for maintenance cost estimation in delivery trucks using diesel and natural gas fuels. *Frontiers in Mechanical Engineering*, 9. doi:<https://doi.org/10.3389/fmech.2023.1201068>
- Kingma, D., y Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*.
- Liachovičius, E., Šabanovič, E., y Skrickij, V. (2023). Freight rate and demand forecasting in road freight transportation using econometric and AI methods. *Transport*, 231-242.
- Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., y Voordijk, H. (2022). An artificial neural network approach for cost estimation of engineering services. *International Journal of Construction Management*, 22(7), 1274-1287. doi:<https://doi.org/10.1080/15623599.2019.1692400>
- Ministerio de Obras Públicas. (2023). *Capítulo 5: Oferta / Planeamiento*. Ministerio de Obras Públicas, Chile.
- Ministerio de Transporte y Obras Públicas. (2023). *Estadísticas del sector transporte terrestre de carga*. Gobierno del Ecuador.
- Montenegro, K. L. (2015). *Costo del servicio de transporte de mercadería por km de recorrido para la empresa Agencomex S.A.* Tesis de grado, Pontificia Universidad Católica del Ecuador.
- Ñaupas, H., Valdivia, M., Palacios, J., y Romero, H. (2018). *Metodología de la investigación cuantitativa-cualitativa y redacción de la tesis*. Ediciones de la U.
- Palafox-Palafox, D., Medina-Marín, J., Seck-Tuoh-Mora, J., Serna-Díaz, M., y Hernández-Romero, N. (2023). Modelo de pronóstico de cadena de suministro

mediante redes neuronales. *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, 103–111.

Rodrigue, J.-P. (2020). *The Geography of Transport Systems* (5th ed.). Routledge.

Rodríguez, J. (2017). Redes neuronales artificiales en la predicción de costos de diversos proyectos: Revisión. *Revista de la Facultad de Tecnología de Información y Comunicación*, 36-43.

Rodríguez, J., y Ugalde, E. (2021). Impacto de la estandarización y escalado: factor para predicción de costos en proyectos a través de una red neuronal artificial. *Ingeniare. Revista Chilena de Ingeniería*, 265–276.

Rojas, I. R. (2011). Elementos para el diseño de técnicas de investigación: Una propuesta de definiciones y procedimientos en la investigación científica. *Tiempo de Educar*, 277-297.

Urrego, D. A. (2024). *Modelo de Machine Learning para la caracterización de carga y estimación de costos de transporte*. Tesis de grado, Universidad Nacional Abierta y a Distancia.

Valencia, J. A. (2023). La estructura de costos en el transporte terrestre de carga. *Realidad Empresarial*, 23-35.

VII. ACRÓNIMOS

ARIMA: Modelo Autorregresivo Integrado de Media Móvil.

BCE: Banco Central del Ecuador

CEPAL: Comisión Económica para América Latina y el Caribe

CNN: Redes Neuronales Convolucionales

CT: Tiempo de Cómputo

GPU: Unidad de Procesamiento Gráfico

GRU: Unidad Recurrente Cerrada

IA: Inteligencia Artificial

KDD: Descubrimiento de Conocimiento en Base de Datos

KNN: K Vecinos más Cercanos

LSTM: Memoria a Largo y Corto Plazo

MAE: Error Absoluto Medio

MAPE: Error Porcentual Absoluto Medio

MDI: Reducción Media de la Impureza

MLP: Perceptrón Multicapa

MOP: Ministerio de Obras Públicas

MSE: Error Cuadrático Medio

MTOP: Ministerio de Transporte y Obras Públicas

OMP: Búsqueda de Coincidencia Ortogonal

R: Coeficiente de Correlación de Pearson

RMSE: Raíz del Error Cuadrático Medio

RNA: Redes Neuronales Artificiales

RNN: Redes Neuronales Recurrentes

SGD: Descenso de Gradiente Estocástico

SVM: Máquina de Vectores de Soporte

SVR: Regresión de Vectores de Soporte

TNF-AHP: Proceso Analítico Jerárquico Difuso Neutrosófico Trapezoidal

VIII. ANEXOS

Anexo 1. Certificado de abstract por parte de idiomas



UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI FOREIGN AND
NATIVE LANGUAGES CENTER

ABSTRACT- EVALUATION SHEET				
NAME: Denis Enrique Benavides Caipe DATE: Wednesday, June the 17th of 2026 Topic: "Heavy Freight Transportation Cost Prediction System Using Artificial Intelligence Tools"				
MARKS AWARDED		QUANTITATIVE AND QUALITATIVE		
VOCABULARY AND WORD USE	Use new learnt vocabulary and precise words related to the topic	Use a little new vocabulary and some appropriate words related to the topic	Use basic vocabulary and simplistic words related to the topic	Limited vocabulary and inadequate words related to the topic
	EXCELLENT: 2 ☒	GOOD: 1,5 ☒	AVERAGE: 1 ☐	LIMITED: 0,5 ☐
WRITING COHESION	Clear and logical progression of ideas and supporting paragraphs.	Adequate progression of ideas and supporting paragraphs.	Some progression of ideas and supporting paragraphs.	Inadequate ideas and supporting paragraphs.
De	EXCELLENT: 2 ☒	GOOD: 1,5 ☐	AVERAGE: 1 ☐	LIMITED: 0,5 ☐
ARGUMENT	The message has been communicated very well and identify the type of text	The message has been communicated appropriately and identify the type of text	Some of the message has been communicated and the type of text is little confusing	The message hasn't been communicated and the type of text is inadequate
	EXCELLENT: 2 ☒	GOOD: 1,5 ☐	AVERAGE: 1 ☐	LIMITED: 0,5 ☐
CREATIVITY	Outstanding flow of ideas and events	Good flow of ideas and events	Average flow of ideas and events	Poor flow of ideas and events
	EXCELLENT: 2 ☐	GOOD: 1,5 ☐	AVERAGE: 1 ☐	LIMITED: 0,5 ☐
SCIENTIFIC SUSTAINABILITY	Reasonable, specific and supportable opinion or thesis statement	Minor errors when supporting the thesis statement	Some errors when supporting the thesis statement	Lots of errors when supporting the thesis statement
	EXCELLENT: 2 ☒	GOOD: 1,5 ☐	AVERAGE: 1 ☐	LIMITED: 0,5 ☐
TOTAL/AVERAGE 9 - 10: EXCELLENT 7 - 8,9: GOOD TOTAL 9 5 - 6,9: AVERAGE 0 - 4,9: LIMITED				



**UNIVERSIDAD POLITÉCNICA ESTATAL DEL
CARCHI- FOREIGN AND NATIVE LANGUAGES
CENTER**

**Informe sobre el Abstract de Artículo Científico
o Investigación.**

Autor: Denis Enrique Benavides Caipe

Fecha de recepción del abstract: Martes, 16 de junio de 2026

Fecha de entrega del informe: Miércoles, 17 de junio de 2026

El presente informe validará la traducción del idioma español al inglés si alcanza un porcentaje de: 9 – 10 Excelente.

Si la traducción no está dentro de los parámetros de 9 – 10, el autor deberá realizar las observaciones presentadas en el ABSTRACT, para su posterior presentación y aprobación.

Observaciones:

Tras evaluar el resumen presentado, se concluye que la traducción al inglés es adecuada y fiel al contenido. De acuerdo con la rúbrica aplicada para su valoración, se le asigna una calificación de 9, por lo que el trabajo queda aprobado.

Atentamente



Validar sólo en FirmasEC.
Firmado Electrónicamente por:
0401442751 MARTHA ARACELLY
VIVEROS ALMEIDA

MA. Martha Viveros
RESPONSABLE CIDEN

Anexo 2. Excel utilizado para la recolección de datos

a) Registro de viajes

	A	B	C	D	E	F	G	H	J	K
	N°	Mes	Fecha	Placa	Km Total	Km Vacío	Tipo de Carga	Ruta	Cliente	Descripción del recorrido por tramos
124	1	Enero	11/01/2025	CBA-1777	1,505	199	Cerámica / Azúcar	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
126	2	Enero	15/01/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
127	3	Febrero	05/02/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
128	4	Febrero	26/02/2025	CBA-1777	1,505	199	Sardina / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (sardina)
129	5	Marzo	04/03/2025	CBA-1777	1,505	199	Cerámica / Azúcar	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
130	6	Marzo	11/03/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
131	7	Abril	02/04/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
132	8	Abril	09/04/2025	CBA-1777	1,505	199	Cerámica / Azúcar	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
133	9	Mayo	02/05/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
134	10	Mayo	16/05/2025	CBA-1777	1,505	199	Caramelos / Azúcar	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg (caramelos), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
135	11	Junio	03/06/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
136	12	Junio	10/06/2025	CBA-1777	1,505	199	Cerámica / Azúcar	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
137	13	Julio	02/07/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac
139	14	Julio	09/07/2025	CBA-1777	1,505	199	Cerámica / Azúcar	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
140	15	Agosto	05/08/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
141	16	Agosto	12/08/2025	CBA-1777	1,505	199	Cerámica / Azúcar	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
142	17	Septiembre	03/09/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
143	18	Septiembre	18/09/2025	CBA-1777	1,505	199	Caramelos / Azúcar	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg (caramelos), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
144	19	Octubre	02/10/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
145	20	Octubre	09/10/2025	CBA-1777	1,505	199	Cerámica / Azúcar	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
146	21	Noviembre	07/11/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
147	22	Noviembre	28/11/2025	CBA-1777	1,505	199	Sardina / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (sardina)
148	23	Diciembre	17/12/2025	CBA-1777	1,505	199	Sardina / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (sardina)
149	24	Diciembre	24/12/2025	CBA-1777	1,505	199	Azúcar / Carga General	RUTA 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)

	A	B	C	D	E	F	G	H	J	K
151	25	Enero	11/01/2025	CAA-2669	1,505	199	Cerámica / Azúcar	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
152	26	Enero	15/01/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
153	27	Febrero	06/02/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
154	28	Febrero	27/02/2025	CAA-2669	1,505	199	Sardina / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (sardina)
155	29	Marzo	05/03/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
156	30	Marzo	27/03/2025	CAA-2669	1,505	199	Sardina / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (sardina)
157	31	Abril	03/04/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
158	32	Abril	25/04/2025	CAA-2669	1,505	199	Sardina / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (sardina)
159	33	Mayo	03/05/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
160	34	Mayo	10/05/2025	CAA-2669	1,505	199	Cerámica / Azúcar	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
161	35	Junio	04/06/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
162	36	Junio	19/06/2025	CAA-2669	1,505	199	Caramelos / Azúcar	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg (caramelos), Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
163	37	Julio	03/07/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiates carg (azúcar)
164	38	Julio	18/07/2025	CAA-2669	1,505	199	Cerámica / Azúcar	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac,

	A	B	C	D	E	F	G	H	J	K
164	38	Julio	18/07/2025	CAA-2669	1,505	199	Cerámica / Azúcar	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiales carg (azúcar)
165	39	Agosto	06/08/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiales carg (azúcar)
166	40	Agosto	21/08/2025	CAA-2669	1,505	199	Cerámica / Azúcar	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiales carg (azúcar)
167	41	Septiembre	04/09/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiales carg (azúcar)
168	42	Septiembre	19/09/2025	CAA-2669	1,505	199	Cerámica / Azúcar	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiales carg (azúcar)
169	43	Octubre	03/10/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiales carg (azúcar)
170	44	Octubre	20/10/2025	CAA-2669	1,505	199	Cerámica / Azúcar	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg (cerámica), Guayaquil→Manta vac, Manta→Ipiales carg (azúcar)
171	45	Noviembre	08/11/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiales carg (azúcar)
172	46	Noviembre	29/11/2025	CAA-2669	1,505	199	Caramelos / Sardina	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg (caramelos), Guayaquil→Manta vac, Manta→Ipiales carg (sardina)
173	47	Diciembre	13/12/2025	CAA-2669	1,505	199	Azúcar / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiales carg (azúcar)
174	48	Diciembre	18/12/2025	CAA-2669	1,505	199	Sardina / Carga General	ruta 1	Cliente Guayaquil	Tulcán→Guayaquil carg, Guayaquil→Manta vac, Manta→Ipiales carg (sardina)

b) Kilómetros por tramo y toneladas transportadas

6A. TONELADAS TRANSPORTADAS POR RUTA Y TRAMO (toda la flota)								
Ruta	Tramo (Origen → Destino)	Km tramo	Estado del tramo	Cap. máx. (ton)	Toneladas cargadas	Ton-km año	Km vacío viaje	
RUTA 1	✓ Tulcán → Guayaquil	664	CARGADO	33	30	1,410,480		
	— Guayaquil → Manta	199	VACÍO (reposicion.)	33	— 0 ton	—	199	
	✓ Manta → IpiALES	664	CARGADO	33	30	1,410,480		
RUTA 2	✓ Tulcán → Manta	616	CARGADO	33	31	591,976		
	✓ Manta → IpiALES	653	CARGADO	33	30	607,290		
RUTA 3	✓ Tulcán → Quito	241	CARGADO	33	30	115,680		
	— Quito → Manta	397	VACÍO (reposicion.)	33	— 0 ton	—	397	
	✓ Manta → IpiALES	637	CARGADO	33	30	305,760		
RUTA 4	✓ Tulcán → Guayaquil	664	CARGADO	33	30	338,640		
	✓ Guayaquil → Tulcán	664	CARGADO	33	31	349,928		
RUTA 5	✓ Tulcán → Danec	267	CARGADO	33	30	112,140		
	— Danec → Quevepalma	215	VACÍO (reposicion.)	33	— 0 ton	—	215	
	✓ Quevepalma → IpiALES	496	CARGADO	33	30	208,320		
RUTA 6	✓ Tulcán → Quito	241	CARGADO	33	30	43,380		
	— Quito → Guayaquil	430	VACÍO (reposicion.)	33	— 0 ton	—	430	
	✓ Guayaquil → Tulcán	664	CARGADO	33	31	123,504		

c) Consumo de combustible

	A	B	C	D	E
1	CIUDAD_ORIGEN	CIUDAD_DESTINO	GALONES_CONSUMIDOS	PRECIO_PROMEDIO_COMBUSTIBLE	COSTO_TOTAL_GALON
2	TULCAN	GUAYAQUIL	140	1.980961583	277.3346217
3	GUAYAQUIL	MANTA	35	1.980961583	69.33365542
4	MANTA	IPIALES	130	1.980961583	257.5250058
5	TULCAN	MANTA	125	1.980961583	247.6201979
6	MANTA	IPIALES	130	1.980961583	257.5250058
7	TULCAN	QUITO	65	1.980961583	128.7625029
8	QUITO	MANTA	50	1.980961583	99.04807917
9	MANTA	IPIALES	130	1.980961583	257.5250058
10	TULCAN	GUAYAQUIL	140	1.980961583	277.3346217
11	GUAYAQUIL	TULCAN	140	1.980961583	277.3346217
12	TULCAN	DANEC	55	1.980961583	108.9528871
13	DANEC	QUEVEPALMA	25	1.980961583	49.52403958
14	QUEVEPALMA	IPIALES	110	1.980961583	217.9057742
15	TULCAN	QUITO	65	1.980961583	128.7625029
16	QUITO	GUAYAQUIL	55	1.980961583	108.9528871
17	GUAYAQUIL	TULCAN	140	1.980961583	277.3346217
18	TOTAL		1535		3040.77603

d) Peajes y viáticos

	A	B	C	D	E	F
1	CIUDAD_ORIGEN	CIUDAD_DESTINO	TIPO_CARGA	PEAJES	COSTO_PEAJES	VIATICOS
2	TULCAN	GUAYAQUIL	FRAGIL	9 \$	54.00	45
3	TULCAN	MANTA	GENERAL	8 \$	48.00	45
4	TULCAN	QUITO	GENERAL	5 \$	30.00	15
5	TULCAN	GUAYAQUIL	PELIGROSA	8 \$	48.00	45
6	TULCAN	DANEC	FRAGIL	6 \$	36.00	15
7	QUITO	MANTA	VACIO	3 \$	18.00	15
8	QUITO	GUAYAQUIL	VACIO	4 \$	24.00	25
9	QUEVEPALMA	IPIALES	GENERAL	8 \$	48.00	35
10	MANTA	IPIALES	GENERAL	8 \$	48.00	40
11	GUAYAQUIL	MANTA	VACIO	2 \$	12.00	10
12	GUAYAQUIL	TULCAN	GENERAL	8 \$	48.00	45
13	DANEC	QUEVEPALMA	VACIO	2 \$	12.00	10

e) Costos de mantenimiento

A	B	C	D	E	F
	NOMBRE	COSTO UNITARIO	DURACIÓN (km)	CANTIDAD	COSTO x KM
	LLANTAS Y ALINEACIÓN (tractocamión: 18 llantas)				
	Llantas delanteras	\$470.00	80,000	2	0.0118
	Llantas de tracción	\$470.00	80,000	8	0.0470
	Llantas de arrastre	\$350.00	120,000	12	0.0350
	Rines	\$120.00	450,000	22	0.0059
	Pinchazos y daños	\$15.00	10,000	2 mensuales	0.0015
	Alineación	\$20.00	60,000	veces al año	0.0003
	Balanceo	\$5.00	60,000	veces al año	0.0001
	MANTENIMIENTO ORDINARIO				
	Cambio de aceite motor (2 canecas)	\$120.00	10,000	1	0.0120
	Aceite de caja y transmisión (3 gln)	\$160.00	60,000	1	0.0027
	Aceite de diferencial (3 gln)	\$140.00	60,000	1	0.0023
	Filtros de aire	\$130.00	60,000	2	0.0043
	Filtros de aceite y combustible	\$93.00	10,000	2	0.0186
	Lubricantes y fluidos varios	\$100.00	5,000	2	0.0400
	Limpieza / lavado motor y chasis	\$100.00	10,000	1	0.0100
	SISTEMA ELÉCTRICO				
	Baterías (2 unidades)	\$120.00	200,000	4	0.0024
	Motor de arranque	\$60.00	10,000	1	0.0060
	Alternador	\$50.00	10,000	1	0.0050
	Luces y accesorios	\$60.00	10,000	1	0.0060
	Sistema eléctrico / revisión general	\$50.00	10,000	1	0.0050
	SISTEMA DE FRENOS				
	Zapatillas / remache y limpieza (2 ejes)	\$300.00	60,000	22	0.1100
	Ajustes y calibraciones de frenos	\$100.00	10,000	1	0.0100
	Revisión frenos de aire	\$100.00	10,000	1	0.0100
	Sellos y retenedores	\$100.00	10,000	1	0.0100
	SUSPENSIÓN				
	Resortes y ballestas	\$250.00	10,000	1	0.0250
	Cauchos y silentblocks	\$100.00	10,000	1	0.0100
	Revisión de suspensión general	\$200.00	60,000	1	0.0033
	SISTEMA DE TRANSMISIÓN				
	Revisión de caja de cambios	\$300.00	120,000	1	0.0025
	Revisión de diferencial	\$300.00	120,000	1	0.0025
	Dirección hidráulica (aceite)	\$150.00	120,000	1	0.0013
	REPARACIÓN DE MOTOR				
	Calibración de válvulas motor	\$50.00	500,000	1	0.0001
	Calibración bomba de inyección	\$1,000.00	500,000	1	0.0020
	Cambio de toberas de inyectores	\$600.00	500,000	1	0.0012
	Mantenimiento de turbo	\$1,200.00	60,000	1	0.0200
	Refrigerante motor (10 galones)	\$120.00	60,000	1	0.0020
	Overhaul motor (reserva proporcionada)	\$4,500.00	500,000	1	0.0090
	TOTAL				0.4347

f) Tiempo en horas y paradas por viaje

	A	B	C	D	E
1	CIUDAD_ORIGEN	CIUDAD_DESTINO	DISTANCIA	TIEMPO_HORAS	PARADAS_POR_VIAJE
2	TULCAN	GUAYAQUIL	664	22	5
3	TULCAN	MANTA	616	17.5	4
4	TULCAN	QUITO	241	6	2
5	TULCAN	DANEC	267	6	2
6	QUITO	MANTA	397	8	2
7	QUITO	GUAYAQUIL	430	9	2
8	QUEVEPALMA	IPIALES	496	11	3
9	MANTA	IPIALES	637	18	4
10	GUAYAQUIL	MANTA	199	4	1
11	GUAYAQUIL	TULCAN	664	22	5
12	DANEC	QUEVEPALMA	215	5	1
13					

g) DATASET

dataset_yanca ML listo.xlsx

Anexo 3. Instrumentos validados



UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI

Nombre: Enrique Benavides

Primer Instrumento – Entrevista estructurada

Variable: Costos de Transporte de Carga Pesada

Dimensión: Costos Directos e Indirectos

Técnica: Entrevista

Instrumento: Guía de preguntas, revisión documental y observación

Dirigido a:

Responsables administrativos y contables de empresas de transporte de carga pesada / entidades del sector transporte (ANT, Aduanor, operadores logísticos, empresas transportistas).

Fecha:

___ / ___ / 2025

Hora:

Tipo de respuesta:

Abierta

Objetivo del Instrumento

Obtener información directa sobre la estructura real de costos en transporte de carga pesada, identificando los rubros económicos que intervienen en la determinación del costo final del servicio, para alimentar la base de datos que será utilizada en el modelo predictivo basado en inteligencia artificial.

Indicador 1: Costo de Combustible (USD / km)

- ¿Cuál es el costo promedio del combustible utilizado en las rutas de transporte?
- ¿Cómo se registra el consumo de combustible por viaje?
- ¿El registro se hace manualmente o con sistema tecnológico?
- ¿Con qué frecuencia varían los precios del combustible para la empresa?
- ¿Quién controla y valida estos registros?

Indicador 2: Costo de Peajes

- ¿Cuántos peajes en promedio atraviesan los vehículos en las rutas más frecuentes?
- ¿Existe una planificación de rutas para reducir costos de peaje?

- ¿Cómo se registran y justifican estos gastos dentro de la contabilidad de la empresa?
- ¿Los peajes representan un costo considerable en el total del viaje? Explique.

Indicador 3: Sueldos y beneficios de conductores

- ¿Cuál es el costo mensual aproximado en salarios de conductores?
- ¿El pago se realiza por viaje, por km o por mes?
- ¿Se incluyen beneficios (alimentación, alojamiento, viáticos)?
- ¿Cómo se asignan los costos laborales al costo final del transporte?

Indicador 4: Mantenimiento Vehicular

- ¿Se lleva un registro del mantenimiento preventivo y correctivo?
- ¿Cuál es el costo promedio anual de mantenimiento por vehículo?
- ¿Qué componentes mecánicos generan mayor gasto?
- ¿El mantenimiento se realiza en talleres propios o externos?

Indicador 5: Seguros

- ¿Qué tipos de seguros mantiene la empresa (vehículo, mercadería, responsabilidad civil, etc.)?
- ¿Cuál es el costo anual de estos seguros?
- ¿Cómo se distribuye ese costo en las operaciones por viaje?

Indicador 6: Costos Administrativos

- ¿Qué porcentaje del costo total corresponde a gestión administrativa?
- ¿Qué actividades administrativas generan mayor gasto?
- ¿Se cuenta con personal dedicado exclusivamente a la gestión logística y documentación?

Indicador 7: Costo por Kilómetro Recorrido

- ¿La empresa tiene calculado el costo por kilómetro? Si es así, ¿cuál es el valor promedio?
- ¿Este costo incluye depreciación vehicular, mantenimiento y combustible?
- ¿Varía el costo por km según tipo de carga o ruta?

Indicador 8: Depreciación y desgaste operativo

- ¿La empresa calcula la depreciación del vehículo?

- ¿En cuántos años se estima la vida útil promedio de un camión de carga pesada?
- ¿El desgaste operativo (neumáticos, frenos, suspensión) se contabiliza de forma directa por viaje o por consumo acumulado?

INSTRUMENTO N.º 2 – FICHA DOCUMENTAL Y OPERATIVA DE REGISTRO DE COSTOS

Variable: Costos de Transporte de Carga Pesada

Dimensión: Costo según la carga – Tipo de vehículo – Influencia de la distancia

Técnica: Revisión documental – Observación técnica – Análisis cuantitativo

Instrumento: Ficha documental – Ficha técnica – Registro operativo

Dirigido a:

Responsables operativos, encargados de flota, personal administrativo y técnico de empresas de transporte de carga pesada.

Objetivo del Instrumento:

Recopilar datos técnicos y verificables relacionados con las características de la carga, vehículo y distancia recorrida, con el fin de **alimentar el sistema predictivo de costos** desarrollado mediante inteligencia artificial, tal como exige el objetivo específico:

"Analizar los factores que inciden en los costos del transporte de carga pesada."

Dimensión	Indicador	Ítem / Pregunta	Registro (para llenar)
Costo según la carga	Peso total transportado	Peso total de la carga transportada (kg)	_____ kg
		Documento de soporte revisado (guía de carga / manifiesto)	_____
	Volumen y dimensiones	Volumen total de la carga (m³)	_____ m³
		Dimensiones (Largo – Ancho – Alto)	L: ____ m / A: ____ m / H: ____ m
	Tipo de carga	Tipo de carga transportada (frágil, peligrosa, especial, normal)	_____
		¿Requirió condiciones especiales? (Sí/No)	() Sí () No
		Descripción de condiciones especiales	_____
Tipo de vehículo	Capacidad del camión	Capacidad nominal del vehículo (toneladas)	_____ ton
		Modelo del camión	_____
	Consumo de combustible	Consumo promedio del vehículo (L/100 km)	_____

		Modelo que consume más combustible (comparación interna)	
Influencia de la distancia	Kilometraje total recorrido	Kilómetros recorridos en la operación	km
		Tipo de ruta (costa, montaña, altiplano, mixto)	
	Costo por kilómetro	Costo total de operación dividido para los km recorridos	USD/km

Fórmulas

1. Cálculo del volumen

Volumen (m³) = Largo (m) × Ancho (m) × Alto (m)

2. Capacidad utilizada del camión

% de utilización = (Peso total transportado / Capacidad nominal del vehículo) × 100

3. Consumo de combustible por viaje

Consumo total = (Consumo del vehículo en L/100 km × Kilometraje total) / 100

4. Costo por kilómetro

Costo por km = Costo total del viaje / Kilómetros recorridos

5. Costo total estimado del viaje (si lo tomas desde documentos)

Costo total = Peajes + Combustible + Mano de obra + Mantenimiento + Administración + Otros costos

INSTRUMENTO N.º 3 – FICHA DE REGISTRO REGULATORIO, TEMPORAL Y EXTERNO

Variable: Costos de Transporte de Carga Pesada

Dimensión: Costos regulatorios y legales – Variabilidad temporal – Costos externos

Técnica: Revisión legal – Análisis documental – Observación de condiciones externas

Instrumento: Ficha de análisis documental y observación

Dirigido a:

Responsables administrativos, personal contable, área legal de empresas de transporte, operadores logísticos y entidades reguladoras.

Objetivo del Instrumento:

Recopilar información verificable sobre costos derivados de regulaciones, variaciones temporales y factores externos, con el fin de analizar su incidencia en la estructura total de costos del transporte de carga pesada.

Dimensión	Indicador	Ítem / Pregunta	Técnica	Instrumento	Registro	Fórmula aplicada
Costos regulatorios y legales	Permisos de transporte especial	¿Qué costo tiene el permiso para carga sobredimensionada o peligrosa?	Revisión legal	Documentos del Ministerio de Transporte	_____ USD	<i>Costo del permiso = Valor oficial publicado por la ANT o Ministerio de Transporte</i>
		¿Cuántos permisos se requieren por viaje?	Revisión legal	Normativa vigente	_____	<i>Número de permisos según tipo de carga</i>
	Multas o sanciones	¿Cuál es el total pagado en multas en el último año?	Análisis documental	Registro de infracciones	_____ USD	<i>Total multas = Suma de montos registrados</i>
		Promedio mensual de multas	Análisis documental	Registro contable	_____ USD	<i>Promedio mensual = Total de multas anuales ÷ 12</i>
Variabilidad temporal	Estacionalidad o variación mensual	¿Cómo variaron los costos en los últimos 12 meses?	Análisis documental	Reportes mensuales	_____ USD	<i>Variación (%) = ((Costo mes actual - Costo mes anterior) / Costo mes anterior) × 100</i>
		Mes con costo más alto	Análisis documental	Reporte mensual	_____	<i>Máximo = valor mayor registrado entre 12 meses</i>
		Mes con costo más bajo	Análisis documental	Reporte mensual	_____	<i>Mínimo = valor menor registrado entre 12 meses</i>
Costos externos	Impacto del clima o condiciones externas	¿El clima elevó el costo del viaje? (Sí/No)	Observación	Bitácora operativa	_____	<i>Costo adicional por clima = (Tiempo extra × Costo por hora del vehículo)</i>
		¿Hubo bloqueos o retrasos?	Observación	Reportes de ruta	_____	<i>Costo por retraso = Horas de retraso × Costo operativo por hora</i>
		Total de costos externos asociados al viaje	Observación	Hoja operativa	_____ USD	<i>Costo externo total = Costo por clima + Costo por retrasos + otros costes adicionales</i>

Costos regulatorios y legales	Permisos de transporte especial	¿Qué costos implican los permisos para carga sobredimensionada o peligrosa?	Revisión legal	Documentos del Ministerio de Transporte
	Multas o sanciones	¿Cuál es el promedio de costos por incumplimientos?	Análisis documental	Registro de infracciones
Variabilidad temporal	Estacionalidad o variación mensual	¿Cómo cambian los costos a lo largo del año?	Análisis documental	Reportes mensuales
Costos externos	Impacto del clima o condiciones externas	¿El mal clima o bloqueos de carretera elevan el costo?	Estudio de caso	Observación


Enrique Benavides

Estudiante


Ivan Mafía

Docente tutor


Validador