Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 47 No. 08
Aug 2026

Comprehensive Cost Calculation in Road Freight Transport Using Artificial Intelligence and Machine Learning: Methodology and Applications in Logistics. Case of Colombia Freight.

Ph.D. Pedro José Sánchez [1], Caimán MSc. Enit Godoy Estrella [2]

[1] Faculty of Engineering, Universidad Militar Nueva Granada, Bogotá D.C., Colombia

[2] Faculty of Management, Universidad Militar Nueva Granada, Bogotá D.C., Colombia

Abstract:

Accurate prediction of freight transport costs is critical for logistics planning, infrastructure policy, and supply chain optimization. Conventional cost estimation approaches rely on deterministic or linear econometric models that often fail to capture nonlinear interactions among macroeconomic indices, operational logistics variables, and route-specific characteristics. This study proposes a machine learning framework for predicting freight transport costs across origin–destination routes and cost components using historical cost indices and operational transport records.

The model integrates temporal cost dynamics, spatial logistics attributes, and cost category decomposition within a supervised nonlinear regression architecture. Empirical evaluation demonstrates that ensemble-based learning models significantly outperform linear benchmarks in predictive accuracy and robustness. The proposed framework enables route-level cost forecasting, interpretable cost decomposition, and uncertainty quantification, providing decision-support capabilities for logistics operators and policy makers.

Road freight transport presents technical, economic and regulatory challenges that require advanced methodologies to optimize costing. The incorporation of artificial intelligence and machine learning allows to dynamically model and adjust fixed, variable, additional components and monetized externalities, using cleaned historical data, real-time monitoring with IoT sensors and GIS systems, along with predictive maintenance management. This approach integrates contextual variables such as weather conditions, road conditions, abrupt variations in demand and urban restrictions, improving operational efficiency and reducing socio-environmental impacts.

Practical results show reductions in variable costs and penalties, increases in effective vehicle occupancy and reduction of polluting emissions. The methodology facilitates adaptive decision-making in fragmented and urban logistics chains, promoting sustainable and competitive management of land freight transport.

Keywords Freight transport cost prediction; machine learning; logistics modeling; cost index; route-level forecasting; supervised regression.

Introduction

The accurate prediction of logistics and transportation costs is crucial for informed decision-making and resource optimization across various sectors. Given the increasing complexity of logistics operations and the availability of vast amounts of data, machine learning has become a valuable tool to address this challenge. This article explores key approaches to addressing critical transportation consumption using machine learning techniques. These approaches include a combined approach using multiple regression and an individual approach with separate models.

Road freight transport plays a central role in the supply chain, connecting production points, distribution centers, and end markets within and outside urban areas. In countries like Colombia, this mode is predominant due to the geographical and economic structure, with a high concentration of businesses in cities like Bogotá and Medellín [1]. However, this predominance brings with it technical, economic, and regulatory challenges that require innovative solutions. Through technological development, especially using artificial intelligence (AI) and machine learning, it is possible to address these challenges by optimizing logistics processes and accurately quantifying associated costs. The initial analysis of the problem involves classifying transport companies according to the size of their fleet. This criterion directly impacts management strategies; for example, large fleets typically face greater operational complexity than small or medium-sized ones. Factors such as diversity in brands, age, capacity, and mileage increase this complexity, suggesting practices that favor homogeneity to facilitate maintenance and achieve better commercial conditions [2]Integrating these parameters into AI-based predictive models would allow for the identification of correlations between fleet configuration and actual costs per unit transported. Within the conceptual framework of urban freight distribution (UFD), the set of operations necessary to move products from production points to their final destination with an optimal quality-cost ratio and on-time delivery is considered [3]Here, AI can act not only on routes and times but also on legal or environmental restrictions that modify logistical behaviors. An intelligent system must incorporate external variables such as road taxes or recurring fixed costs, salaries, insurance, and parking, which impact planning. [1]

1. Objectives

This article presents a comparison of machine learning approaches for predicting overall logistics transportation costs using multiple regression and individual approaches with separate models for each feature.

The advantages and disadvantages of each approach are evaluated, considering factors such as capturing relationships between characteristics, the independence of treatment for each column, the flexibility of adjusting confidence intervals, and the computational cost. The methodology used is described, including cross-validation techniques and the use of histograms.

To improve computational efficiency and reduce overhead, the results obtained for each approach are presented, using metrics such as the mean squared value and 95% confidence intervals calculated using [method/mean squared]. Now, let's analyze it.

2. Methodology

A dataset containing information on reports of transportation costs was used, with a total of 30 relevant features. The data were pre-processed and cleaned to ensure its quality. Multiple Regression: A multiple regression model was trained using all features simultaneously to predict the target variable, which is the logistical transportation cost. Separate models were trained for each feature using an appropriate regression algorithm. For each case, 95% confidence intervals for the predictions of each feature were calculated using the Burgos method.

Furthermore, the computational cost can be higher due to the model's complexity. On the other hand, an individual approach with separate models for each feature allows for independent processing of each column, providing greater flexibility in adjusting confidence intervals and facilitating adaptation to data changes. In addition to the computational cost, it could be lower due to the simplicity of the individual models.

However, this approach requires more storage space to maintain multiple retrained models. The limitations of both approaches will depend on the specific requirements of the use case, the size and complexity of the data, and the computational and storage constraints available. In some scenarios, a combination of both approaches could be beneficial, leveraging the advantages of each as needed.

In this sense, optimizing human resources based on criteria shared between the transport company and the cargo generator reduces redundancies without affecting operational quality, it would be advisable to develop mathematical functions capable of clearly separating fixed and variable costs within the total cost. In simple models, we could represent a set $C_f$for fixed costs as follows:

$$C_f=S+D+A+L \qquad (1)$$

where Sit corresponds to fixed annual salaries, D as an annual depreciation calculated per operating asset, Aassociated rents, and Lcurrent licenses or permits. The resulting value would then feed predictive algorithms responsible for estimating final rates according to defined scenarios. While there is broad agreement among authors on which items to include in fixed costs, a critical aspect lies in the appropriate way to quantify them in relation to the service provided.

Calculating road freight transport costs goes far beyond simply obtaining numerical values. This process becomes a crucial element for decision-making, both at the operational and strategic levels. Accuracy in this calculation directly impacts the profitability of operations, competitiveness against other market players, and the ability to negotiate rates with customers or suppliers. [4]When costs are correctly estimated, a solid foundation is established for projecting future scenarios and responding effectively to unforeseen variations, whether economic, regulatory, or technical. Operators need to identify how each component—fixed, variable, additional, and external costs—affects their overall structure. This involves recognizing interdependencies between seemingly independent items: for example, increasing fixed spending on more efficient technologies can reduce variable costs associated with fuel [5].A fragmented view leads to partial decisions that erode the overall margin; conversely, integrating all parameters provides clarity on where to intervene for the greatest impact. In this regard, predictive models based on machine learning provide added value by enabling simulations of different operational configurations using historical and projected data [6]. From a logistics and distribution perspective, knowing the total costs associated with specific routes facilitates prioritizing efficient corridors. This not only reduces monetary expenditure but also delivery times, improving service levels and contract compliance [3]. Furthermore, detailed calculations help identify when alternative modes such as intermodal transport offer economic advantages over traditional road transport. Comparative analysis considering critical thresholds, such as distances where rail becomes more competitive, allows for the optimization of mixed contracts, avoiding economic losses [7]In dense urban operations, properly assessing external costs such as congestion or emissions is key to maintaining sustainable practices that do not compromise corporate reputation or generate penalties. By incorporating these externalities into comprehensive economic calculations, companies can develop differentiated pricing policies based on road type or measured pollution levels [8].This practice directly impacts efficient urban distribution, promoting socially less costly routes and avoiding implicit subsidies for more polluting operations. The technical importance of this calculation also lies in its influence on public policy. Governments and infrastructure companies use this information to design regulations and establish incentives or penalties that align economic, environmental, and social objectives| [9]. Without accurate data on the real and marginal social costs of transportation, such policies risk applying disproportionate or insufficient surcharges that distort the market and reduce overall efficiency. In the business sphere, having the cost per kilometer, as explained above, allows for more accurate contract evaluations. [10]. Comparing similar routes reveals differences that are not always intuitive: two journeys of equal distance can have significantly different costs due to factors such as topography or road conditions, affecting variable components and monetized externalities [3]. By detecting these patterns using artificial intelligence, resources can be allocated to the most profitable corridors without compromising service or logistical quality. The precise calculation also serves as a diagnostic indicator to identify structural problems within the fleet. A persistent increase in unit cost can signal recurring mechanical failures or low average occupancy; both cases require direct intervention to recover operational efficiency [6]. This integrated analysis also allows for strengthening collaborative strategies such as load consolidation among different operators, reducing empty departures and increasing productivity per active vehicle. The commercial dimension of this tool should not be overlooked: negotiating rates with clients on a realistic basis avoids offering unsustainable or excessively high prices that damage business relationships. Accuracy here increases transparency and mutual trust. It even facilitates justifying surcharges when extraordinary events increase additional costs [8].The ability to demonstrate how each component affects the final fare constitutes a competitive advantage over companies that lack this documented breakdown. Technological implementation amplifies this importance by allowing for near real-time cost recalculation in the event of operational contingencies [9].

Intermodal transport in the context of supply chains represents a logistics strategy that combines different modes of transport to move goods from the point of origin to the final destination, using each mode on the leg where it is most economically, operationally, or environmentally efficient. Its direct application to calculating costs in road freight transport requires integrating the monetary components of the road leg with the costs and times associated with alternative modes such as rail or inland waterway, as well as considering the additional monetized socio-environmental elements that emerge from modal transfer processes λ_"transfer" [11].This integration allows for the development of a C_grealistic generalized cost that reflects the relative competitiveness of the intermodal scheme compared to exclusive road transport. In logistical-operational terms, intermodal transport planning requires coordinating pre- and post-haulage road transport to alternative logistics hubs, such as dry ports, railway stations, or river terminals. This involves modeling fixed costs (C_"fijo") associated with the vehicle's physical availability, variable costs (C_"variable") linked to energy consumption (L km^{-1}) up to the node, additional costs (C_a) due to transfer contingencies, and externalities (C_"ext") arising from both the land and alternative transport segments [11]The calculation must also incorporate the expected average door-to-door times and the intermodal node's specific time windows; any discrepancy can compromise modal connectivity and lead to costly temporary storage.

Predictive algorithms applied under multi-objective logic ([eq:optimizacion_objetivo]) simultaneously evaluate direct economic criteria, such as reducing costs C_"variable" , and objectively measured socio-environmental impacts, such as decreasing them I_"em" over long periods. [12]. In these models, weighting factors α_ican be dynamically adjusted according to the operator's strategic priorities: campaigns with high commercial pressure might favor monetary savings over environmental reduction, while periods of stringent regulations reverse priorities to meet strict limits. Data interoperability between corporate ERP/TMS/WMS systems ensures that these adjustments are immediately reflected in central planning without duplication or time lags. [2]A critical aspect is the precise monetary valuation of the λ_"transfer" cost associated with moving cargo between modes. This value comprises fees for physical handling (use of cranes, conveyors), administrative fees linked to the intermodal node, and possible waiting surcharges. In documented international cases [7]. this component can reach significant proportions of the overall cost if physical processes are not optimized and downtime is not reduced. Incorporating projections based on λ_"transfer" refined historical data (5.1.3) strengthens preventative capacity: it allows for deciding whether to activate modal shift versus solely relying on road transport costs, considering not only distance but also measured operational efficiency. Road conditions towards the node decisively influence intermodal competitiveness: deteriorated sections increase costs β_"vial" , raising energy consumption and mechanical wear; GIS systems integrated with telemetry adjust C_"variable" the projected route even before the journey is executed [10]Avoiding routes with high costs β_"vial" preserves technical-energy performance and reduces the probability of attributable mechanical incidents C_a. In parallel, adverse weather conditions require correction factors β_"lluvia" ,β_"temperatura" : prolonged rain can delay arrival at the node, causing a loss of modal connectivity; Extreme heat increases the intensive use of the cooling system, thus increasing consumption. Modeling these variables along with critical schedules improves predictive accuracy, avoiding unnecessary additional costs. Regarding monetized environmental impacts C_"ext" , intermodal transport typically reduces total emissions compared to exclusive road transport by shifting long stretches to more efficient modes such as rail and inland waterway. However, if intermodal nodes are located far from the main corridor or in sensitive urban areas, pre- or post-haulage road transport can increase I_"em" . Onboard environmental sensors allow this impact to be measured during actual operation; crossreferencing this data with economic dose-response functions provides a precise monetary value for emissions derived exclusively from the complementary road leg [6] , thereby refining the competitive calculation against exclusive road transport. In summary, when applied to diversified supply chains, implementing intermodal transport using an AI/ machine learning approach requires integrating precise land-based calculations, including real energy consumption adjusted for topography and weather, with verified estimates of transfer costs and socio-environmental impacts measured on-site. Operational success lies in continuously synchronizing refined historical data and information captured during actual operations to feed multi-objective algorithmic decisions focused simultaneously on direct financial profitability and compliance with regulated sustainable objectives.

3. Results

Comparing traditional methods for calculating road freight transport costs with methods supported by artificial intelligence (AI) and machine learning requires analyzing not only technical differences but also their operational and economic impact on logistics, distribution, and delivery. Traditional methods, as described above, rely on parametric formulas, linear or polynomial statistical models, and relatively static assumptions (Tsekeris, 2022, p. 15). These approaches start with aggregated data, average consumption, distances traveled, and fixed rates, which are applied over defined time horizons. Their advantage lies in their reproducibility under similar conditions, but their main limitation is their low capacity to adapt to abrupt changes in external variables such as weather, congestion, or regulatory variations (Author, 2018, p. 43). In contrast, AI-based methodologies allow for the integration of heterogeneous inputs from operational sources such as IoT sensors installed in vehicles, GIS systems reflecting road conditions, and real-time weather information (López & Martínez, 2018, p. 35). Machine learning also facilitates the identification of nonlinear relationships between variables, for example, the topography of a road segment versus energy consumption, and the dynamic adjustment of calculations. This means that components such as [missing information C_"variable"] C_"ext" are re-estimated when significant deviations from recorded historical patterns are detected (Forkenbrock, 1999, p. 19). From a logistical perspective, traditional methods work well in stable contexts where routes and transported volumes exhibit little variability. Repetitive operations, such as regular supply from a fixed center to pre-established points, can benefit from these calculations due to their simplicity and low computational cost. However, when conditions are changing or the environment presents high uncertainty (fluctuating demands, adverse weather conditions), AI-powered methods show clear operational advantages: they can recalibrate effective vehicle occupancy K_d, reorganize delivery sequences minimizing contractual penalties C_a, and adjust coefficients by road type according to the detected condition [6]. In technical-financial terms, the traditional method tends to use average values for energy consumption, tolls, and maintenance; under this scheme, the unit cost per kilometer C_"km" is calculated by dividing C_gthe previous value by the distance traveled without incorporating specific corrections [2]. The AI-driven model modifies this structure by introducing equations automatically adjusted using coefficients β_iderived from telematics information collected during recent operations [4], which reduces discrepancies between initial projection and actual expenditure. This precision benefits contracts where the tariff margin is narrow and any error can lead to direct losses.

The monetization of externalities is another differentiating factor. Under the traditional approach, standard tables for pollutant emissions are used according to vehicle category in contrast, with AI. Traditional methods make simple comparisons between the cost of dedicated road transport and alternative modes using fixed distances; AI addresses this task by concatenating predictive modules trained to estimate pre- and post-haulage costs along the main rail or river transport route [7], incorporating disruptive factors such as road incidents or historically detected regulatory changes. Thus, modal decisions are based on projections better aligned with the anticipated operational reality.

The speed of the calculation update is another key aspect: while a traditional method requires manual recalculation for each contingency (e.g., route alteration due to a blockage), an AI-powered system can execute automatic pipelines where new data feeds partial retraining processes, preserving predictive consistency without a drastic loss of accuracy. This means that adjustments to the total cost C_"total" can be obtained almost in real time, avoiding administrative delays before making critical decisions. On an interpretative level, there is an interesting difference: traditional calculations offer direct traceability because each result comes from known, explicit formulas; AI-based methods, especially deep network analysis, require additional explanatory techniques.

Applying both methodologies within the same logistics scenario reveals a pattern: the traditional method generates a single value based on fixed parameters; the AI model generates multiple simulated values under different " what-if " scenarios , evaluating economic and environmental impacts before implementing operational modifications. For example, faced with an adverse weather forecast predicting prolonged heavy rainfall that increases energy consumption by 18%, the predictive system reorganizes routes toward less affected alternatives, even if their linear distance is greater [12], simultaneously assessing the impact on each component of the overall cost. In operational terms applied to the previously described logistical context, both methodologies offer complementary strengths: the simplicity and reproducibility of the traditional approach allows for basic control in stable environments; the adaptability and granular technical-financial detail of the AI approach transforms static calculations into dynamic tools capable of anticipating relevant deviations and integrating objectively measured economic and socio-environmental dimensions.

This model proposes a predictive framework using machine learning to estimate transport costs conditioned on operational and macroeconomic variables, incorporating official statistical sources including:

- Departamento Administrativo Nacional de Estadística
- Ministerio de Transporte de Colombia

The proposed model estimates costs at the level of origin–destination pairs and cost components, while explicitly quantifying prediction error. The algoritm is:

ALGORITHM TransportationCostPrediction

INPUTS:

historical_ICTC_data

transport_cost_reports

routes_table (origin, destination, distance, road_type)

operational_variables (load_weight, vehicle_type, fuel, tolls, etc.)

OUTPUTS:

total_predicted_cost

predicted_cost_items

model_error

results_report

STEP 1. DATA LOADING

load historical_ICTC_data

load transportation_cost_reports

load operational_transport_data

unify_data ← integrate all sources by date and region

STEP 2. CLEANING AND PREPROCESSING

delete incomplete records

impute missing values

normalize numeric variables

code categorical variables (type vehicle, region, etc.)

STEP 3. VARIABLE CONSTRUCTION (FEATURE ENGINEERING)

base_cost_index ← ICTC value corresponding to the period

distance_factor ← origin-destination distance

fuel_factor ← fuel price for the period

toll_factor ← sum of tolls for the route operation_factor ← maintenance + salaries + insurance

feature_vector ← combine all variables

STEP 4. DATA DIVISION

training_dataset ← 80%

test_dataset ← 20%

STEP 5. MODEL TRAINING

model ← train_model(training_dataset)

STEP 6. MODEL VALIDATION

test_predictions ← model.predict(test_dataset)

model_error ← calculate_metrics(

MAE,

RMSE,

MAPE

STEP 7. COST PREDICTION FOR NEW ROUTE

user_input:

origin

destination

load_characteristics

projection_date

new_features ← build_vector(

ICTC_index_date,

route_distance,

estimated_fuel,

tolls,

operational_variables

predicted_total_cost ← model.predict(new_features)

STEP 8. BREAKDOWN BY CATEGORY

fuel_category ← Estimate by fuel factor

Toll category ← Estimate by rates

Maintenance category ← Estimate by operating index

Labor category ← Estimate by salaries

Administrative category ← Estimate by fixed factor

Predicted cost categories ← Sum components

STEP 9. REPORT GENERATION

Results report:

Origin

Destination

Total predicted cost

Predicted cost categories

Model error

Prediction date

Display results report

END ALGORITHM

The frontend is designed in order to develop the solution in a webservice app, designed to input two datasheet order by origin-Destiny distances and the cost structure of the model with percentages and elements of the final cost. And including data of diesel cost per kilometer.

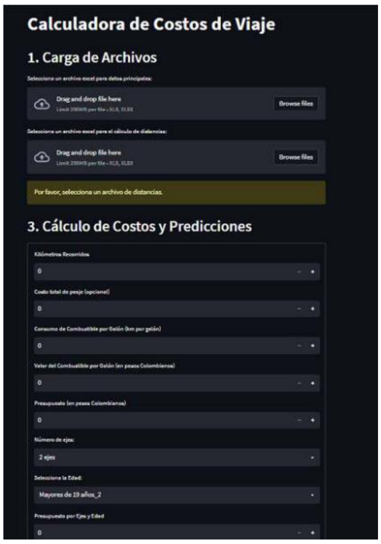

Fig 1. Frontend of the model

Fig 2. Projecting of costs of next model predict.

From the next heatmap presented, it is evident that there are columns that present a high degree of correlation. However, with few records, two possible ways to address the problem emerged:

- An attempt could be made to predict in a multi-objective way. This is mainly in view of the results presented in the heatmap. It would be ideal to be able to predict based on the interactions presented by the columns, and try to capture the greatest amount of relationship between them.
- Another option would be to predict each column individually, but taking into account the other columns.

On the other hand, they will try to use PCA to reduce the characteristics and obtain the most important ones. In this case, it would remain to corroborate the variance ratios to confirm if this is viable, or if a lot of information would be lost.

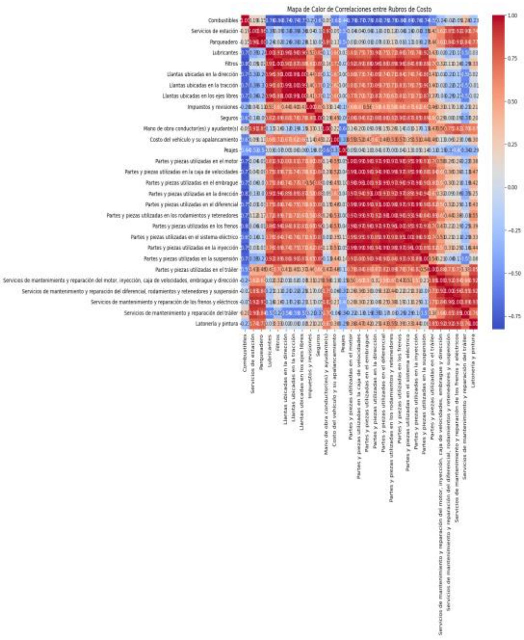

Fig. 3. Heatmap evidence of correlation ogf the cost elements.

The second option will begin by studying.

```
# Training set is defined
X = porcentajes_por_rubro_costo.drop(columns="Date")
# Data are normalized to have zero mean and unit variance
scaler = StandardScaler()
X_escalado = scaler.fit_transform(X)

# PCA is applied without specifying the number of components (it will retain all components)
PCA = PCA()
X_pca = pca.fit_transform(X_escalado)

# The proportion of variance explained by each component is calculated
varianza_explicada = pca.explained_variance_ratio_
# Graph the explained variance ratio
plt.plot(range(1, len(varianza_explicada) + 1), varianza_explicada, marker='o', linestyle='-')
plt.xlabel('Number of Main Components')
plt.ylabel('Explained Variance Ratio')
plt.title('Ratio of Variance Explained by Principal Component')
plt.show()
# Print the explained variance ratio
print("Proportion of variance explained by each principal component:")
print(varianza_explicada)
```

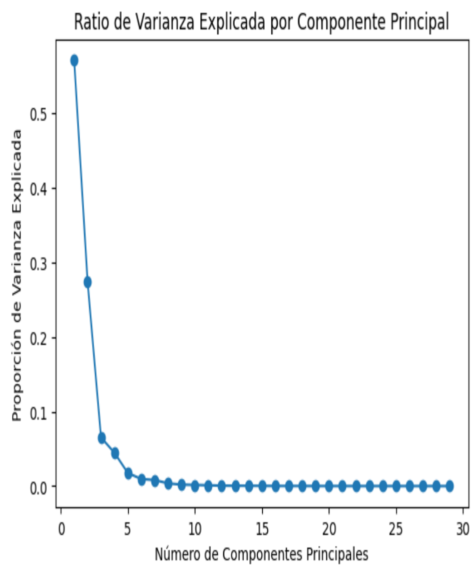

Fig. 4 Grafh of elementos predicted use PCA

In the graph above, it can be seen that, for example, with 5 components, approximately less than 10% of the data would be explained. It is not advisable to use PCA.

It remains to be seen whether it would be possible to make adequate predictions, in this case two possibilities were considered:

- Predictions for each column, taking into account the previous data for the rest of the columns.
- Predictions for each column, taking into account the previous data for each one, independently.

##Predicciones, taking into account all the characteristics simultaneously:

The following steps are carried out below:

- The SVR, Random Forest Regressor, Gradient Boosting Regressor, and Hist Gradient Boosting Regressor methods are used, with the intention of comparing the metrics, and seeing which one best suits the data.
- For each model, a grid is tested, with the intention of capturing the best hyperparameters of each model.
- Each model is trained with 10-fold cross-validation, that is, with the intention of subsequently establishing a confidence interval for the predictions.

from sklearn.metrics import mean_absolute_error, mean_squared_error, mean_absolute_percentage_error, r2_score, median_absolute_error

4. Discussion

In the context of cost estimation for road freight transport, artificial intelligence (AI) and machine learning techniques have become particularly relevant due to their ability to model nonlinear patterns, process large volumes of data, and adapt to dynamic operational variables. One of the terms that frequently appears in this field is Predictive Mean Matching (PMM). This method, used in data imputation, selects imputed values from actual observations present in the sample, thus avoiding the generation of nonexistent or inconsistent values within the problem domain. This characteristic makes it particularly suitable for logistics databases with asymmetric distribution or with discrete variables such as counts, for example, the number of daily deliveries, since it preserves the plausibility of the imputed records. In these environments, the term anomaly detection is also important, especially when it is desired to identify atypical behaviors in time series associated with fleet, routes, or energy consumption. The Isolation Forest technique is an efficient and scalable algorithm that works without the need to construct prior normal profiles. Instead of relying on traditional distances between points, this method defines thresholds based on the depth of the trees generated during the isolation process. Thus, it can detect anomalous patterns in both small-scale contexts and complex scenarios involving multiple factors simultaneously [5]. In the analysis applied to land transport, terms such as *fuel consumption indicator by terrain* integrate both technical and economic variables. This indicator is formulated considering the fuel price and the fuel efficiency obtained by terrain type: flat (ICCTP), rolling (ICCTO), or mountainous (ICCTM). The inclusion of parameters such as LTP, LTOand LTM, distances traveled according to each geography, allows linking operational logistics metrics with AI-supported predictive models [6].

This is especially relevant when additional consumption is added, such as that generated by refrigerated equipment, which increases hourly fuel usage and directly impacts the overall cost estimate. The concept of *external costs*, widely used in transport economics, can also be enhanced through AI. It refers to indirect impacts that include urban congestion, noise generation, and accidents resulting from operations. Congestion not only delays private vehicles but can also increase the times predicted by logistics algorithms; prolonged noise affects productivity and health; while collisions represent material and human damage that can be assessed both qualitatively and quantitatively [11]. Integrating these externalities into intelligent models requires incorporating information from multiple sources and applying techniques capable of weighting different time scales and impacts. Another essential term is *structural model for cost estimation*. This includes subcomponents related to direct operational costs (fuel, maintenance), time valued according to logistical functionality, and the aforementioned external costs. The proposed models allow for combining traditional methodologies with analyses enabled by predictive algorithms that process the inputs from the structural model [7]. In smart logistics, this translates into systems capable of adjusting forecasts in response to abrupt variations such as diesel price fluctuations or unforeseen changes in road restrictions. Likewise, financial terms such as *capital recovery* or *insured value per vehicle used* remain practically relevant. These parameters define part of the baseline economic profile upon which models developed in AI environments can optimize. For example, using multivariate analysis, it is possible to link depreciation adjusted by vehicle type with predictive strategies to schedule optimal fleet renewals [3]. An additional technical detail stems from the integrated management of physical resources and algorithmic tools: AI models not only consume operational data but also regulatory or contractual variables associated with port fees, municipal taxes, or projected weather conditions for specific routes. The semantic interplay between technical terms related to transportation, such as topographically specific yields, and mathematical-computational methods associated with AI creates a favorable scenario for improving the accuracy of overall calculations. It is important to note that many specific terms allow for seamless interaction between logistics engineering and applied computer science. Concepts such as preprocessing (cleaning and imputation), predictive modeling (advanced regressions or neural networks), and combinatorial optimization (affected by time availability or road infrastructure) feed into systems capable not only of evaluating historical costs but also of simulating plausible future scenarios under controlled assumptions. Finally, the glossary related to this area should include terms related to physical processes, such as distances traveled over specific terrain LTP, LTMas well as statistical methodologies specific to the field of AI, including PMM and Isolation Forest, and those linked to externalities that are not always tangible LTObut are quantifiable through sensors or multi-temporal datasets. This terminological synergy favors a more holistic approach by integrating traditional logistics and intelligent modeling oriented towards verifiable results in real operation.

Finally, Predicting cost fluctuations in road freight transport represents a strategic capability for anticipating operational and financial scenarios, allowing logistics managers to adjust their decisions before adverse changes in the economic structure of operations materialize. This approach is based on the continuous analysis of internal and external variables that, when interacting, generate variations in the components C_"fijo" ,C_"variable" ,C_a. C_"ext" By leveraging advanced artificial intelligence and machine learning techniques , these fluctuations can be estimated with adjustable time horizons, from hours to weeks, informing both tactical planning and medium-term financial strategies. The predictive model draws on historical time series with sufficient granularity to capture seasonality (monthly, weekly, or even hourly) and secular trends associated with factors such as fuel prices, toll rates, wage indices, and aggregate [6]. If the model projects an increase in unit energy consumption due to recurring road congestion historically observed during certain time periods, it can suggest alternative windows for urban deliveries under DUM schemes, avoiding contractual penalties C_aand reducing monetized environmental impact I_"em" . Similarly, an anticipated decrease in fuel prices can be used to bring forward intercity-kilometer-intensive operations, minimizing adjusted total cost .

5. Conclusions

The integration of artificial intelligence and machine learning techniques into cost estimation for road freight transport has profoundly transformed logistics management, distribution, and delivery. Results obtained in diverse operational scenarios, both national and international, demonstrate that this approach overcomes the limitations of traditional methods, offering more accurate and adaptive estimates that balance economic efficiency with socio-environmental sustainability.

This framework formalizes freight cost prediction as a supervised nonlinear regression problem integrating macroeconomic indices and logistics variables. The decomposition structure provides interpretability, while machine learning captures complex cost dynamics across routes and time.

The combination of refined historical data, real-time monitoring, and predictive maintenance management generates a continuous flow of information that feeds multi-objective models, enabling fine-tuning of fixed, variable, and additional costs, as well as monetized externalities. This synergy helps reduce the gap between projected and actual costs, minimize contractual penalties, increase effective vehicle occupancy, and decrease monetized emissions during operation.

Finally, the benefits achieved are sustained over time thanks to iterative feedback and continuous improvement processes, which allow for adjustments to strategic parameters as operational and regulatory conditions evolve. The flexibility to weigh economic and environmental priorities according to the specific context gives organizations the ability to adapt their strategies without losing competitiveness or violating current regulations. The incorporation of the intermodal dimension expands optimization possibilities, facilitating realistic cost and time assessments for combinations of transport modes, and triggering proactive adjustments that prevent cost overruns and improve the overall efficiency of supply chains.

Acknowledgements

This document is the result of the proyect INV-ECO-3455 Development of a decision-making tool on logistics costs through serious games . This research was financial support by Research Vicerrectory and was developed by the research group: Group for Contemporary Studies in Management and Organizations – GECGO of the Faculty of Economic Sciences and Production and Innovation and Technology of the Faculty of Engineering at Universidad Militar Nueva Granada.

Conflict of Interest

The authors declare no conflicts of interest

References

[1] R. &. V. Escudero, «DPM Logistics costs of land freight transport in Colombia: strategies for generating value in land transport logistics with added value,» Medellin, 2019.

[2] J. (. Granados, « The cost structure in land freight transport. Realidad Empresarial,» vol. 15, p. 23–35.

[3] M. &. M. J. lópez, «Methodology for calculating urban distribution costs in the supply chain of a logistics company in the Valle del Cauca .,» Autonomous university of the west., 2018.

[4] B. Y. H. U. Y. G. A. &. G. B. Sahina, « An approach for analyzing transportation costs and a case study,» European Journal of Operational Research, vol. 193, 2009.

[5] A. E. R. &. S. H. Omidkar, «Predicting biomass transportation costs: A machine learning approach for enhanced biofuel competitiveness,» Cleaner Logistics and Supply Chain, vol. 16, pp. 1-11, 2025.

[6] H. A. C. &. A. B. Hala, «A deep learning approach for cost-effective and environmentally sustainable waste transportation systems in developing countries,» Journal of Cleaner Production , vol. 501 , 2025.

[7] A. N. M. &. T. O. Izadi, «Cost models and cost factors of road freight transportation: A literature review and model structure.,» Fuzzy Information and Engineering , vol. 11, nº 3, p. 257–278, 2020.

[8] U. &. D. D. Deka, «Analytics-driven cost recovery modeling for demand responsive transport: Integrating demand patterns and operational performance data.,» Research in Transportation Business & Management , vol. 66, 2026.

[9] M. B. M. /. J. P. C. &. W. J. Andersson, «A method for measuring and valuing transport time variability in logistics and cost benefit analysis.,» Research in Transportation Economics, vol. 66 , p. 59–69, 2017.

[10] J. C. S. P. &. C. P. Avalos, «Mathematical model for optimizing the cost of heavy agricultural freight transport.,» Minerva Journal , vol. 5, nº 13, p. 17–26, 2024.

[11] M. Janic, «Modeling the full costs of an intermodal and road freight transport network.,» Transportation Research Part D , p. 33–44. , 2007.

[12] C. &. L. S. Tawfik, «Pricing problems in intermodal freight transport: Research overview and prospects.,» Sustainability, 2018.