Journal of Research in Social Science And Humanities (2025); 5(4):
DOI: http://dx.doi.org/10.47679/jrssh.v5i4.497
ISSN 2809-3356 (print), ISSN 2807-3916 (online)

RESEARCH ARTICLE

# Pricing System For Indonesia's Freight Forwarding Industry: A Case Study Of PT Adhi Surya Amanah

Shofi Rafiiola[1]

Published online: 15 December 2025

**Abstract**

PT Adhi Surya Amanah (ASA Trans), a domestic B2B freight forwarding company in Indonesia, faces persistent inefficiencies in its pricing process that hinder operational speed, accuracy, and competitiveness. The company relies heavily on manual quotation preparation using fragmented information spread across spreadsheets, emails, and handwritten records. This results in inconsistent pricing decisions, long quote response times, and a heavy reliance on staff interpretation, limiting the company's ability to respond efficiently to customer demand in an increasingly competitive logistics environment. These inefficiencies pose both scientific and managerial challenges, as accurate and timely pricing is a key factor in determining profitability and market performance in freight forwarding. Therefore, this study aims to develop, test, and implement a data-driven price forecasting infrastructure that can increase efficiency, reduce human error, and improve strategic decision-making at ASA Trans. The research begins by identifying the root causes of inefficiency through a detailed assessment of ASA Trans's operational workflow. This evaluation identified four key challenges: unstructured and fragmented pricing information, manual verification, lack of access to real-time cost and market data, and over-reliance on subjective judgment. All of these issues lead to decreased operational reliability and limit the company's ability to grow. To address these challenges, this paper proposes a forecasting model to be developed based on the premise that ASA Trans' historical records of freight, cost, and pricing information are sufficient to provide the operational trends necessary for accurate forecasting. The study also assumes that market factors such as fuel and toll prices are subject to specific, predictable trends. The objectives of this study include determining the data requirements for price prediction, comparing forecasting techniques, creating a practical prototype of a price forecasting system, and integrating the system into ASA Trans' workflow. To achieve these objectives, the researchers employed a multi-stage methodology. Primary data was collected using internal freight data, operational cost records, and historical prices, while secondary data was collected in the form of freight benchmarks and competitor rate cards. Analysis Phase: The machine learning process involved is structured and includes data cleaning, standardization, exploratory data analysis, model development, and validation. Supervised learning technology was applied to three models: Multiple Linear Regression, Random Forest, and LightGBM to determine which approach provided the most accurate and operational predictions. Model evaluation is conducted using MAE, RMSE, and MAPE, supported by business-oriented validation criteria that measure quotation speed, usability, and alignment with financial objectives. The findings indicated that the best model was the Random Forest model, which had the lowest error and the most consistent results across various shipping situations. The fact that this model can capture nonlinear relationships makes it a favorable solution for the multifaceted and cost-based structure found in the Indonesian domestic freight market. The implementation of the Random Forest model was realized in a web-based Price Forecasting System created at ASA Trans. The implementation phase transformed the model into a working tool that can input estimated shipping costs, estimated operational costs, and estimated lead times with minimal user input. Test results showed that bid completion time was reduced from several hours to minutes, indicating significant improvements in operational efficiency. The fact that this system improves consistency, reduces administrative workload, and enhances the accuracy of pricing decisions was confirmed by user feedback during the training and socialization phases. This system is also capable of providing structured and continuous cost details, which increases transparency and enhances customer trust. This research contributes to scientific knowledge by demonstrating the practical application of machine learning in pricing optimization for domestic freight forwarding. It also establishes an empirical foundation for integrating predictive analytics into pricing workflows in small to mid-scale logistics companies in emerging markets. In the case of ASA Trans, the forecasting system is a strategic move towards digitalization, which will have long-term advantages with better pricing management, uniform decision-making, and preparedness to undertake an automation initiative in the future, like a more comprehensive Transportation Management System (TMS).

**Keyword:** Pricing Forecasting, Pricing System, Freight Forwarder

Institut Teknologi Bandung

*) corresponding author

Correspondin Author name

Email: shofi.rafiola@gmail.com

## Introduction

Indonesia's logistics sector plays a vital role in the country's economy and is currently experiencing a major transformation. This shift is driven by the rapid growth of e-commerce and an increasing demand for fast and reliable delivery services. By 2024, the domestic logistics market is expected to reach approximately US$39 billion, growing steadily at a rate of 5–7% per year (ALI, 2023). This growth is further fueled by the government's strong investment in transportation infrastructure and the ongoing digitalization of supply chains, as noted in recent market reports (Indonesia Transport and Logistics Market Forecast Group, 2024; Mordor Information, 2025). The transformation of Indonesia's logistics sector aligns with the global digitalization trend in freight and container transportation. According to a BCG analysis (2023), seven digital trends, including e-platforms, advanced analytics, AI, blockchain, and automation, are projected to transform logistics operations by facilitating online ordering, dynamic pricing, predictive maintenance, and automated workflows (BCG, 2023). These innovations significantly reduce operational inefficiencies by replacing manual verification with data-driven systems, shortening quote turnaround times from hours to minutes, and increasing overall supply chain transparency. Global companies such as DHL, DB Schenker, and Kuehne + Nagel have integrated these digital solutions to offer faster and more transparent services (Michel and Siegfried, 2020). Meanwhile, local companies such as JNE Logistics, Wahana Selamat, and RPX Logistics are stepping up their game by improving their B2B services and investing in price automation and unique service options (Muthutantrige, Hasan, and Shah, 2023). Because of this competition, customers now expect faster service, clear transparency, and consistent pricing; these factors have become key to winning and keeping their business. In today's fast-paced logistics market, getting prices right and responding quickly are key to staying competitive. But many shipping companies in Indonesia still use old-fashioned, manual pricing methods. This usually means checking fixed-rate cards, confirming payments with third parties by phone or email, and waiting for several management approvals. This slow manual process leads to long wait times for quotes, inconsistent pricing, and a greater likelihood of errors, which can damage customer trust and reduce profits.

## Method

### Data Analysis Method

Data analysis for this study followed a structured machine learning (ML) process, which combined statistical preprocessing, supervised regression modeling, and rigorous validation. This process consisted of the following steps:

### Data Preprocessing

The pre-processing phase ensures that historical shipping, operational, and pricing data is cleaned, organized, and converted into a machine learning-ready format. This step is critical because ASA Trans' raw logistics data is fragmented across various spreadsheets, handwritten notes, and vendor emails. The following activities are implemented:

### Data Cleaning

In the cleaning phase, the dataset was first reviewed for duplicates, as some shipment records (e.g., Jakarta-Surabaya quotes) were recorded multiple times; only one was retained to avoid bias. Missing values were handled systematically: continuous variables (distance, volume, fuel price) were filled by imputing the average of similar shipments, while categorical variables (transport mode, cargo type) were imputed using the most frequently occurring category on comparable routes. For example, missing weights in Jakarta-Surabaya "General Cargo" records were replaced with the average of similar shipments, and missing modes for Jakarta-Medan were imputed as "Truck" if they were used in 80% of the previous cases. This process ensured the integrity, consistency, and reliability of the data for model training.

### Data Standardization

Data standardization ensures that all numerical variables in the dataset use consistent units. Distances are standardized to kilometers (km), volumes are converted to cubic meters ($m^3$), and all monetary amounts are expressed in Indonesian Rupiah (IDR). This process is crucial to avoid discrepancies caused by differing units and to ensure that machine learning models can interpret the variables accurately. By using consistent units across all datasets, predictive models can process inputs more reliably and make more accurate predictions.

### Model Development (Machine Learning)

The development of the forecasting model in this study is based on a structured dataset derived from operational, historical, and reference information from ASA Trans. After preprocessing and standardization, the refined dataset provides a reliable basis for testing various forecasting approaches. Three statistical and computational methods were selected: multiple linear regression (MLR), random forest (RF), and LightGBM, as they offer complementary advantages in terms of interpretability, predictive accuracy, and computational efficiency.

## Results and Discussion

### Data Input

The dataset used in this study was compiled from a combination of internal records and external industry references, resulting in a comprehensive picture of ASA Trans's daily operations and pricing realities. Rather than relying on a single source, the study combined shipping details, operational cost components, and historical pricing data that reflect how ASA Trans transports goods throughout Indonesia. These internal records were then supplemented with secondary data, such as port charges, ferry rates, and competitor rates, to fill in information gaps, particularly for areas not covered in ASA Trans's original rate cards. By combining these diverse sources, the dataset is rich, realistic, and representative of the complex logistics environment in which the company operates.

**Table 1: Summary of Data Sources Used in the Machine LearningDataset**

| Data Type | Data Source | Key Variables Included | Important Variables in the Dataset |
|---|---|---|---|
| **Shipment Data** | Primary | Province, Destination City, Price/m³, Volume (m³) | Provides the basic characteristics of the shipment and the price references that ASA Trans uses in its quotation procedures. |
| **Operational Data** | Primary | Mode of Transport, Distance (km), Fuel Price, Handling Fee | Reflects the actual costs incurred during the shipment and is a key component for cost estimation. |
| **Historical Shipment Records (2024–2025)** | Primary | Distance, Mode, Volume, Fuel, Handling, Actual Cost, Quoted Cost, Lead Time | This data contains 588 actual shipments and trains a machine learning model to identify cost patterns over time. |
| **Freight Rate Benchmarks** | Secondary | Port charges, Ferry charges | It is used to correct for or evaluate handling costs, so the model reflects industry standards. |
| **Competitor Pricing Models** | Secondary | Tariffs for uncovered areas (e.g., Aceh Barat, Kalimantan Barat) | Completes the lack of geographic coverage in ASA Trans own internal rate card and adds to the accuracy of the model. |

By combining these data sources, the resulting dataset is more than just numbers: it reflects the real-world operational challenges ASA Trans faces every day, from fuel price fluctuations to the complexities of regional logistics. This integrated dataset enables machine learning models to learn from a wide range of shipping scenarios, enhancing their ability to generate accurate, consistent, and actionable pricing predictions. Ultimately, this database enables ASA Trans to move toward faster, more transparent, and more reliable pricing decisions across its operations.

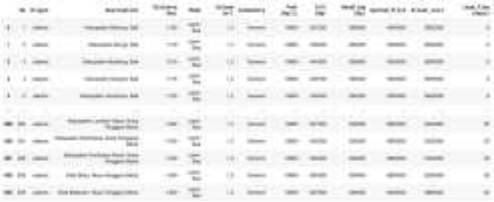

**Figure.1: Dataset ASA Trans**
Source: Author (2025)

**Data Cleaning**

Data Structuring for Fuel (Rp/L)

The data structuring process for fuel costs aims to standardize and correct inconsistencies in recorded fuel price values across different modes of transport. Initially, the dataset showed a uniform price of 10,800 IDR per liter for all modes (land, sea, and combined), as shown in Figure 4.3. This uniformity did not accurately reflect actual operating conditions, where fuel costs vary depending on the type of transport and the source.

**Table 2: Data Format for Fuel Before Data Structuring**

| Mode | Fuel (Rp/L) |
|---|---|
| **Land-Sea** | 10800 |
| **Sea** | 10800 |
| **Land** | 10800 |

During the data structuring phase, adjustments were made based on the latest operational references and recent fuel consumption records. As illustrated in Figure 4.4, the corrected dataset differentiates fuel prices by transportation mode: Rp15,197/L for land transportation, Rp1,300/L for sea transportation, and Rp10,800/L for multimodal routes (land-sea). These improvements ensure that the dataset more accurately represents actual variations in logistics costs, thereby improving the reliability of modeling and subsequent analysis in machine learning-based price prediction systems.

**Table 3: Data Format for Fuel After Data Structuring**

| Mode | Fuel (Rp/L) |
|---|---|
| **Land-Sea** | 10800 |
| **Sea** | 1300 |
| **Land** | 15197 |
| **Sea-Land** | 1300 |

Data Structuring for Destination (Province)

Initially, the destination data combined city or district and province names into a single column, as shown in Figure 4.5. This format limited the flexibility of data analysis, making it difficult to categorize shipments by geographic level or aggregate by province. To increase data granularity and facilitate spatial analysis, the column was split into two separate columns: Destination_City and Destination_Province, as illustrated in Figure 4.6.

**Table 4: Data Format for Destination Before Data Structuring**

| Origin | Destination |
|---|---|
| **Jakarta** | Kabupaten Badung, Bali |
| **Jakarta** | Kabupaten Bangli, Bali |
| **Jakarta** | Kabupaten Buleleng, Bali |

This restructuring allows machine learning models to more effectively differentiate between intercity and interprovincial shipments, facilitating a deeper analysis of variations in costs, delivery times, and operational performance by region. The improved format also optimizes data readability, standardization, and consistency, essential aspects for accurate model training and regional cost predictions.

**Table 5: Data Format for Destination After Data Structuring**

| Origin | Destination_City | Destination_Provice |
|---|---|---|
| **Jakarta** | Kabupaten Badung | Bali |
| **Jakarta** | Kabupaten Bangli | Bali |
| **Jakarta** | Kabupaten Buleleng | Bali |

**Data Distribution**

Destination Province

The distribution of shipping destinations by province, as shown in Figure 4.7, illustrates the geographic concentration of ASA Trans's logistics operations. The data indicate that most shipments during the 2024-2025 period were destined for East Java, Central Java, and West Java, reflecting the company's strong operational presence in the Java corridor, Indonesia's main economic and industrial hub.

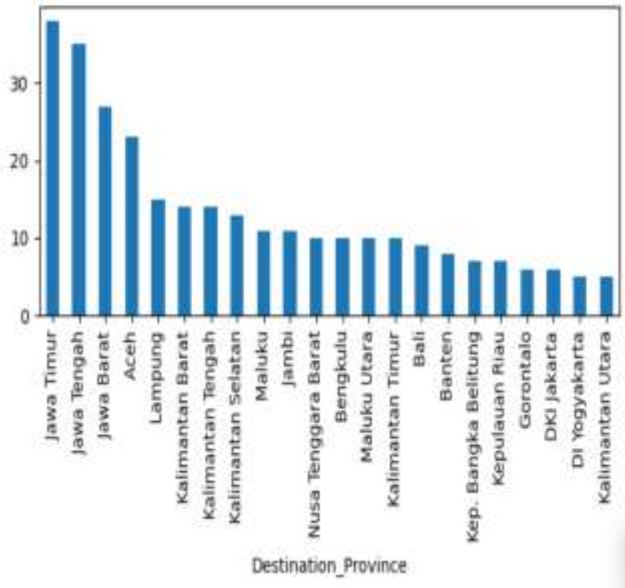


**Figure .2: Destination Province Distribution Data**

Other provinces, such as Aceh, Lampung, West Kalimantan, and Central Kalimantan, show moderate shipping activity, while Gorontalo, Jakarta, and North Kalimantan recorded lower frequencies. This pattern highlights ASA Trans's focus on high-demand routes within Java, while maintaining moderate activity in Sumatra and Kalimantan.

**Transportation Mode**

The distribution of transport modes used by ASA Trans, as shown in Figure 4.8, indicates that land transport predominates in the company's logistics operations, followed by multimodal land and sea routes. This pattern reflects the company's focus on a national land distribution network, particularly in Java and on inter-island connections that require a combination of land and sea transport.

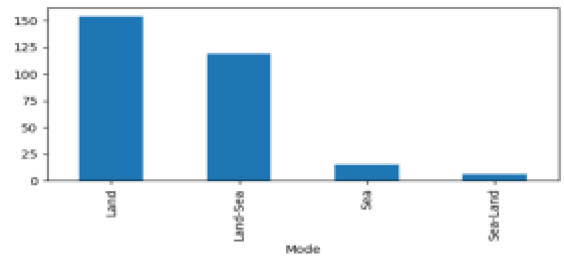


**Figure 3: Transportation Mode Distribution Data**

On the other hand, sea and land-sea modes represent a smaller proportion of total shipments, indicating activity limited to purely maritime routes or reverse multimodal routes. These finding highlights ASA Trans's operational dependence on land transport infrastructure and underscores the importance of optimizing the efficiency of land logistics in the company's pricing and forecasting models.

**Distance (km)**

The histogram in Figure 4.9 depicts the distribution of delivery distances in ASA Trans' logistics operations. The majority of deliveries fall within the 0 to 1,000 km range, indicating that most deliveries are short- and medium-haul, primarily within the islands of Java and Sumatra. A small, but still significant, number of deliveries exceeds 1,500 km, representing long-haul routes to eastern regions such as Sulawesi and Kalimantan.

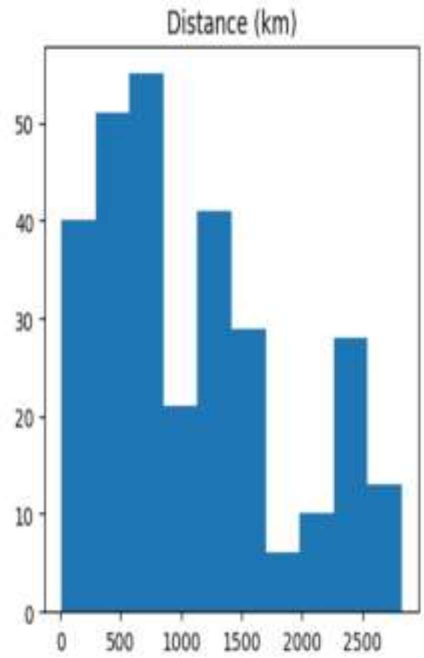


**Figure 4: Distance (km) Distribution Data**

This distribution indicates that ASA Trans' operational network primarily focuses on short intra- and inter-island routes, while long-haul deliveries account for a smaller proportion of total deliveries. Understanding this pattern is crucial for optimizing pricing strategies and improving the accuracy of distance-based cost forecasting in machine learning models.

**Lead Time (Days)**

The histogram in Figure 4.10 shows the distribution of lead times used to determine shipping prices in ASA Trans' operational data. Most shipments are completed within 1–5 days, indicating a high concentration of shipments with short lead times, particularly for land routes in Java and surrounding islands. A smaller proportion of shipments fall within the 7–15-day range, typically representing multimodal routes combining land and sea transportation.

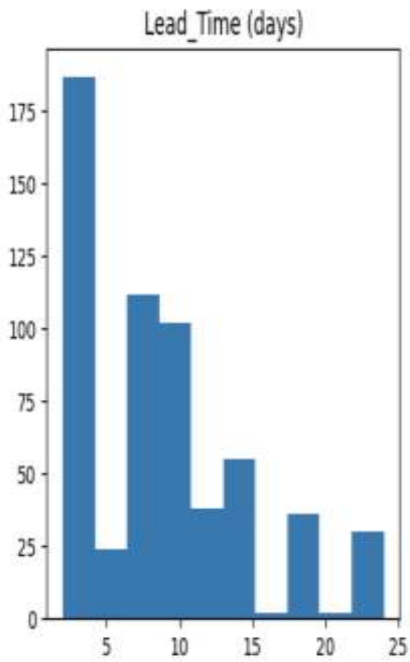


**Figure 5: Lead Time (days) Distribution Data**

Ships exceeding 15 days are relatively rare and generally relate to long-haul or inter-island shipments to eastern Indonesia. This distribution confirms that ASA Trans' core operations prioritize short- to medium-haul freight shipments, reflecting its strategic focus on fast and reliable domestic freight delivery.

**Data Analysis**

Correlation Analysis

The correlation analysis in Figure 4.X identifies key relationships among ASA Trans's operating variables. The results show that distance (km) has a strong positive correlation with tolls (r = 1.00), fuel (r = 0.64), quoted price (r = 0.64), and actual cost (r = 0.64), indicating that distance is a key cost determinant. Management costs also show a high correlation with both quoted and actual costs (r = 0.70), reflecting their significant impact on pricing. On the other hand, waiting time shows a moderate correlation with distance (r = 0.59) and fuel (r = 0.45), indicating that longer routes require more time and fuel. Variables with missing or inconsistent data, such as weight and volume, were excluded to maintain data integrity in the model.

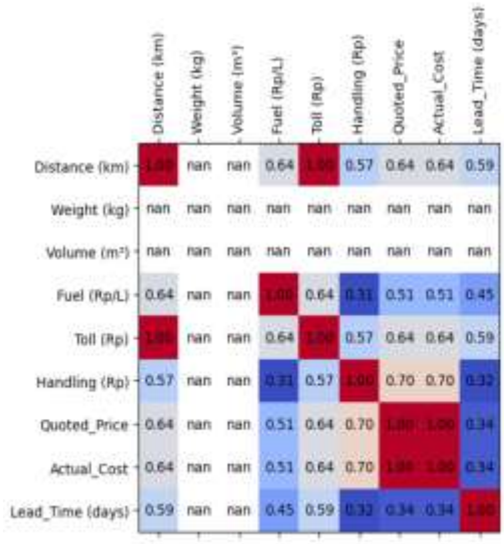


**Figure 6: Correlation Analysis ASA Trans Data**

**Delete Column**

During the data preprocessing phase, several columns were removed to ensure the machine learning model only learned meaningful and informative variables. Exploratory Data Analysis (EDA) revealed that several columns had no variance, meaning they had constant values across all data points. These columns also did not provide predictive information for the model, as they lacked a clear pattern. Therefore, removing these columns helped simplify the dataset, reduce noise, and improve model performance. Two of these columns, Commodity and Origin, were removed because they had the same value in all rows and were irrelevant for cost prediction.

**Table 6: Column Deleted in the Dataset**

| Column Name | Destination_City | Destination_Provice |
|---|---|---|
| **Commodity** | Constant value (no variance) | All commodity shipments are of the same type providing no differentiation or predictability (General) |
| **Origin** | Constant value (no variance) | The column has no effect on the cost variation; all the shipments are made at one location. (Jakarta) |

**Duplicate Value Analysis**

Duplicate entry analysis was performed to ensure that only unique shipment records were included in the dataset. In logistics data, shipments can occur multiple times, for example, when quotes are changed or when there are conflicts between data sources. If these duplicates are not removed, some routes or cost patterns may appear more frequently than expected, which can bias the model and reduce its prediction accuracy. To address this, identical rows across all columns were removed from the dataset. No duplicates were allowed, so each record corresponded to a single shipment. This cleaning process helped the model learn to use accurate and balanced data, thereby increasing the reliability of its predictions.

**Data Pre-Processing**

Feature engineering in data pre-processing is the process of creating new variables from existing data to improve the performance and interpretability of machine learning models. In this study, several additional features were developed to better understand ASA Trans's operational and cost patterns. The new features include:

- Toll per kilometer: Represents the toll cost per kilometer, helping to measure the cost efficiency of routes.
- Average delivery time by province: Shows the average delivery time by province, reflecting regional logistical conditions.
- Shipment scope: Classifies routes as intra-island or inter-island shipments, depending on whether the shipment remains within Java or crosses other islands.

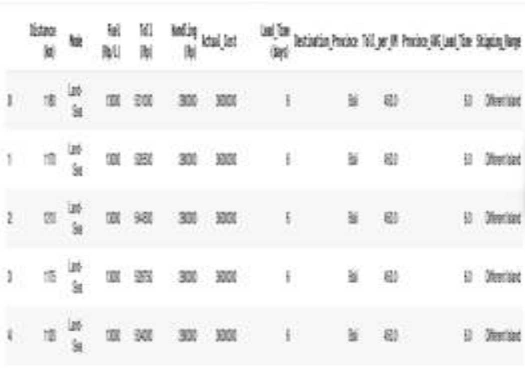

**Figure 7: Feature Engineering Additional Column**

These engineered features strengthen the dataset by highlighting geographic, cost, and performance variations, enabling the machine learning model to generate more accurate and contextualized price predictions.

**Machine Learning**

**Data Modelling**

The dataset was divided into two subsets, training data and testing data, to prepare for machine learning model development. The independent variables (X) include all operational and cost-related predictors, while the dependent variable (y) is Actual_Cost, representing the target output for prediction.

```
X = df_augmented.drop('Actual_Cost', axis=1)
y = df_augmented['Actual_Cost']

# Perform train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

print("Shape of X_train:", X_train.shape)
print("Shape of X_test:", X_test.shape)
print("Shape of y_train:", y_train.shape)
print("Shape of y_test:", y_test.shape)

Shape of X_train: (470, 11)
Shape of X_test: (118, 11)
Shape of y_train: (470,)
Shape of y_test: (118,)
```

**Figure 8: Data Modeling Shape**

Using the train_test_split function, the dataset was split with an 80:20 ratio, where 470 records (80%) were used for model training and 118 records (20%) for testing. The random_state=42 parameter was applied to ensure reproducibility and consistent results across model iterations. This approach allows the model to learn underlying patterns from the training data while the testing data is reserved for performance evaluation, ensuring that the predictive model can generalize effectively to unseen data and avoid overfitting.

**Implementation Multiple Linear Regression Model**

A multiple linear regression (MLR) model was implemented to analyze the relationship between several independent variables, such as distance, fuel, tolls, handling charges, and waiting time, and the dependent variable, the actual cost. This model aims to generate a linear equation that predicts shipping costs based on these operational factors. As shown in Figure 4.15, the comparison between actual and predicted costs indicates that the MLR model successfully captures the overall cost trend, with predicted values (red dots) closely aligning with actual values (blue dots). However, slight deviations are observed at certain data points, particularly for shipments with higher cost variance, which could be due to unconsidered external factors such as weather, port delays, or fluctuations in supplier prices.

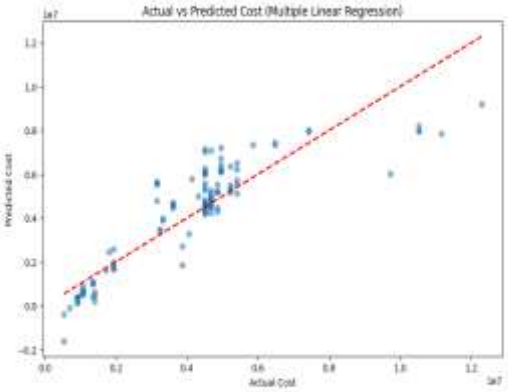


**Figure 9: Multiple Linear Regression Comparison Cost**

The scatter plot in Figure 4.16 further confirms this relationship, with most data points clustered near the regression line, indicating a positive linear correlation between the actual and predicted values. This suggests that the MLR model provides a solid foundation for cost estimation, although further refinement using nonlinear models (e.g., Random Forest or LightGBM) could improve prediction accuracy.

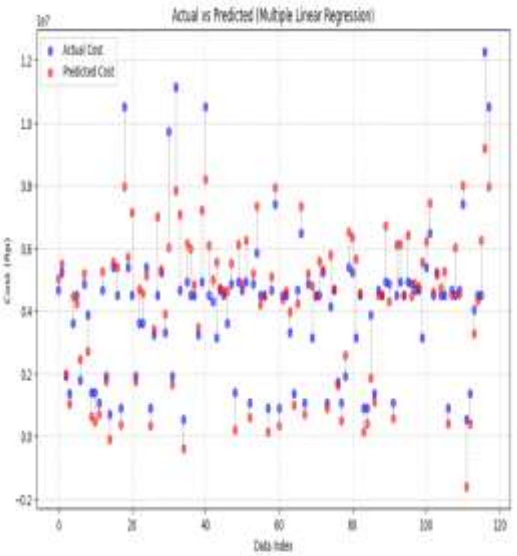


**Figure 10: Scatter Plot Multiple Linear Regression**

**Implementation Random Forest Model**

A Random Forest regression model was implemented to improve the predictive accuracy of linear models by capturing the nonlinear relationship between operational variables and costs. Unlike multiple linear regression, which assumes a linear relationship, the Random Forest model builds multiple decision trees and averages the results, reducing variance and improving consistency. The example decision tree shown in Figure 4.17 illustrates how the model segments the dataset based on the most influential variables: Average delivery time by province, handling fees, tolls, and fuel prices. These characteristics are identified as key determinants of total shipping costs. The decision structure shows that shorter average delivery times and lower handling fees or tolls tend to result in lower overall costs, while routes with higher fuel and handling fees significantly increase total cost estimates.

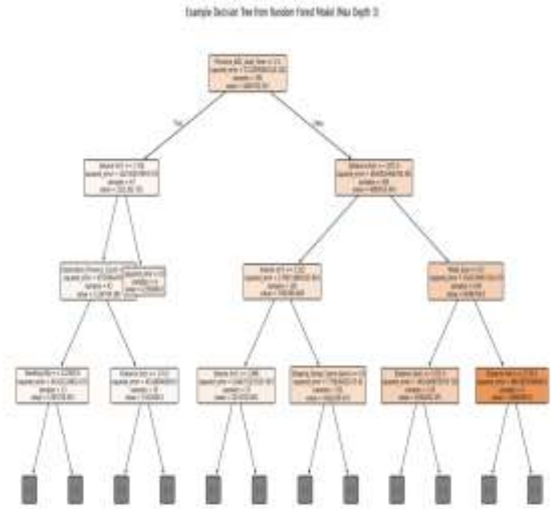

**Figure.11: Decision Tree Random Forest Model**

As visualized in Figure 4.18, the Random Forest model demonstrates a strong fit between actual costs (blue) and predicted costs (red). The points almost completely overlap, indicating high predictive accuracy and lower error variance compared to the linear model.

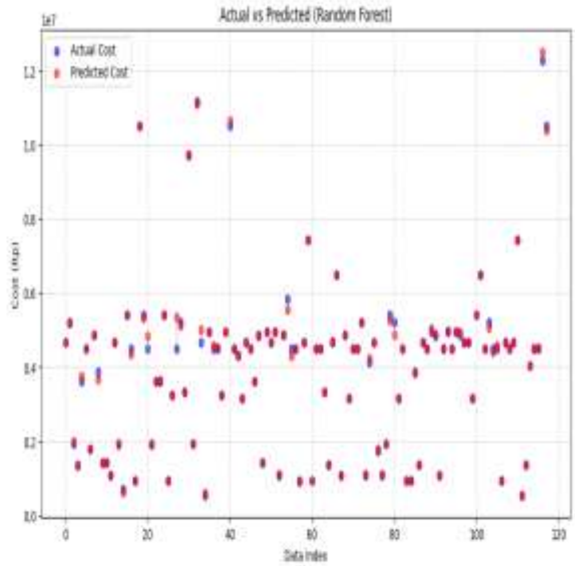


**Figure 12: Scatter Plot Random Forest**

This confirms that Random Forest effectively captures the complex cost interactions in ASA Trans' logistics operations, making it a more reliable tool for dynamic, data-driven price forecasting.

**Implementation LightGBM Model**

The LightGBM (Light Gradient Boosting Machine) model was implemented to improve predictive performance and computational efficiency compared to previous models. LightGBM uses a gradient boosting framework with a leaf-tree growth strategy, enabling it to more effectively capture the complex nonlinear relationships between cost variables. As shown in Figure 4.13, the decision tree generated by the model illustrates how LightGBM partitions the data using key variables such as Handling (Rp), Distance (km), Average Provincial Delivery Time (Province), and Delivery Time (days). This partitioning demonstrates that handling costs and delivery distance fundamentally influence total shipping costs, while delivery time and average provincial delivery time provide additional adjustments for regional variations.


**Figure 13: Decision Tree from Light GBM Model**

In Figure 4.20, the comparison of actual and predicted costs shows consistent alignment, with most data points following a similar pattern. This demonstrates that LightGBM effectively learns from operational and cost variables to estimate shipping prices with relatively small prediction errors, reflecting the model's ability to more accurately represent real-world cost behavior.

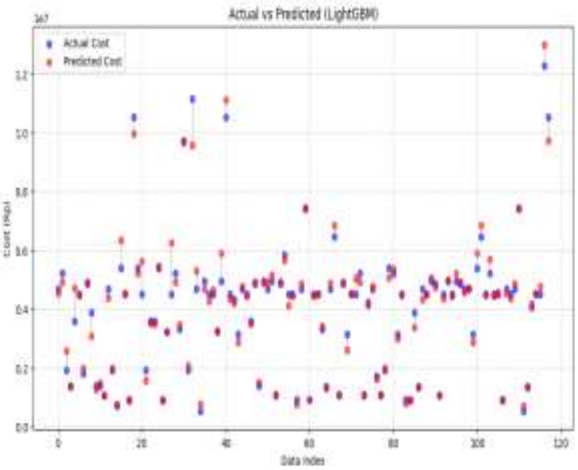


**Figure14: Scatter Plot LightGBM Model**

**Machine Learning Evaluation**

Model performance was evaluated using1-line standard regression metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). These metrics measure how accurately each model predicts shipping costs compared to the actual values in the dataset.

| | MAE | RMSE | MAPE |
|---|---|---|---|
| Multiple Linear Regression | 971444.295978 | 1.270411e+06 | 36.923044 |
| Random Forest | 206101.490205 | 6.808286e+05 | 4.626843 |
| LightGBM | 413965.265702 | 1.091906e+06 | 10.298935 |

**Figure .15: Machine Learning Evaluation**

As shown in Figure 4.21, the Multiple Linear Regression model produced the highest error values (MAE = 971,444.30; RMSE = 1,270,411.00; MAPE = 36.92%). The Random Forest model showed substantial improvement, reducing the average prediction error to MAE = 206,101.49 and achieving the lowest MAPE of 4.63%, demonstrating strong accuracy and robustness in learning complex patterns from the data. Meanwhile, the LightGBM model achieved comparable results with MAE = 413,965.27, RMSE = 1,091,906.00, and MAPE = 10.30%. Although its error rate was slightly higher than Random Forest, it maintained computational efficiency and stability across a larger dataset. Overall, these evaluation metrics confirm that both ensemble-based models (Random Forest and LightGBM) outperformed traditional linear regression in predicting logistics costs, reflecting their ability to handle nonlinearities and feature interactions in ASA Trans operational data.

**KFold Cross Validation**

To ensure that the model evaluation results were reliable and unaffected by random data splitting, K-Fold Cross-Validation was performed. This technique divides the dataset into K equal subsets (in this case, K=3) and iteratively trains the model on two folds while testing it on the remaining folds. This process is repeated three times, and the average Root Mean Squared Error (RMSE) is calculated to assess the consistency of the overall model performance.

| | RMSE |
|---|---|
| Multiple Linear Regression | 1.674210e+06 |
| Random Forest | 4.858347e+05 |
| LightGBM | 7.852066e+05 |

**Figure 16: KFold Cross Validation Root Mean Squared Error (RMSE)**

As shown in Figure 4.16, the Random Forest model achieved the lowest average RMSE value (485,834.69), followed by LightGBM (785,206.56) and Multiple Linear Regression (1,674,210.89). These results indicate that Random Forest consistently produced the most stable and accurate predictions across all validation folds. These validation processes demonstrated that the Random Forest model performed consistently across multiple data splits, demonstrating good generalization ability to previously unseen data. The consistent results indicate that the model effectively captures key price patterns in the ASA Trans operational data sets without significant bias.

**Conclusions and Recommendations**

The study concluded that the main challenges and inefficiencies in ASA Trans's current manual pricing process stem from its heavy reliance on fragmented data, manual verification, and sequential approvals. The lack of a centralized system forces administrators to consolidate information from spreadsheets, emails, and supplier messages, resulting in long bid turnaround times, inconsistent pricing, and frequent calculation errors. These inefficiencies not only reduce operational speed but also undermine competitiveness and customer trust, especially in Indonesia's rapidly evolving freight market, where clients demand fast and accurate. **What are the main challenges and inefficiencies in ASA Trans's current manual pricing process?** The main challenges in ASA Trans's manual pricing process stem from its reliance on fragmented data, sequential verification steps, and human judgment. Data is available in various calculation times, electronic mail, and messaging data, requiring administrative staff to manually consolidate and verify information. This results in long bid completion times, often exceeding three hours for inter-island shipments, and increases the risk of miscalculations and inconsistencies. The lack of a standardized pricing logic also leads to variations in bids among staff, reducing price transparency and overall customer trust. As a result, ASA Trans faces operational inefficiencies that limit its responsiveness in Indonesia's fast-moving freight market. What types of data are needed to build a reliable price forecasting framework for domestic freight transportation in Indonesia? To develop an accurate and reliable price forecasting framework for domestic freight transportation, this study identified three key data categories that were also applied in ASA Trans's modeling process. First, operational data: including shipment origin and destination, distance (km), transportation mode (land, sea, or multimodal), and dwell time (days): serve as key predictors affecting total shipping costs. Second, cost data components, such as fuel price (Rp/L), toll fees (Rp), and handling fees (Rp), represent direct operational costs used as key input variables in the Random Forest and LightGBM models. The integration of these variables resulted in a structured and representative dataset that reflects real-world operational conditions, enabling the forecasting model to generate precise and contextually relevant price predictions for ASA Trans's domestic freight operations. **What advanced statistical and computational methods demonstrate the strongest forecasting performance for freight transportation pricing?** A comparison of the Multiple Linear Regression, Random Forest, and LightGBM models shows that the ensemble-based method provides the most accurate and reliable forecasting results for transportation pricing. The Multiple Linear Regression model remains useful for understanding how each variable affects costs, but it cannot fully capture the complex nonlinear relationships found in the ASA Trans pricing data. In contrast, the Random Forest model significantly improves prediction accuracy by processing a diverse dataset and reducing errors through its decision tree ensemble. LightGBM performs similarly well, offering faster computation and high precision when handling large and complex data. Based on the model evaluation results, the Random Forest model achieves the lowest error rate and the most consistent predictions, making it the most reliable and suitable model for the ASA Trans pricing forecasting system. **How can a forecasting system be integrated into ASA Trans's business processes to reduce administrative workload and improve decision-making?** Integrating a forecasting system into ASA Trans's workflow has transformed the pricing process from manual to automated. The system allows transport managers to enter shipment details, such as route, weight, and handling type, into a centralized web platform that instantly generates accurate price estimates. These results can be aggregated and converted into official quotes and invoices through direct integration with the finance module. This automation reduces repetitive administrative tasks, minimizes human error, and ensures price consistency across all divisions. Furthermore, real-time data updates and model retraining enable the system to continuously evolve, adapting to current carrier and supply dynamics to support faster, data-driven decision-making. **What strategic benefits can ASA Trans achieve by adopting a data-driven forecasting system compared to existing manual methods?** Implementing a data-driven forecasting system offers significant strategic advantages for ASA Trans. The system reduces quote response time from hours to minutes, improves pricing accuracy, and strengthens profit margins by mitigating overpricing and underpricing. Furthermore, it promotes transparency and standardization, fostering greater customer trust. Internally, the system facilitates interdepartmental collaboration by integrating Finance, Operations, and IT functions into a unified digital workflow. Beyond efficiency, this transformation positions ASA Trans for long-term competitiveness, as the forecasting system enables the development of data-driven strategies, continuous improvement, and preparedness for future digital innovations in the Indonesian freight transport sector.

## References


ALI. (2023). Indonesia Logistics Market Outlook 2024. Indonesian Logistics Association.

Bappenas, & World Bank. (2023). Indonesia Logistics Cost Study.

BCG. (2018). The digital imperative in container shipping. https://www.bcg.com/publications/2018/digital-imperative-container-shipping

Cherepanov, V., & Popov, A. (2024). Design and design thinking in digital transformation. Journal of Business Research, 169, 113–126.

Christopher, M. (2016). Logistics & supply chain management (5th ed.). Pearson Education.

Gujarati, D. N., & Porter, D. C. (2009). Basic econometrics (5th ed.). McGraw-Hill/Irwin.

Harani, N. H., Setyawan, M. Y. H., & Ferdinan, D. (2025). Predicting basic shipping tariff using machine learning. Nuansa Informatika, 19(2), 58–66. https://doi.org/10.25134/ilkom.v19i2.388

IBM. (n.d.). Data-driven decision making. https://www.ibm.com/think/topics/data-driven-decision-making

Jang, H.-S., Chang, T.-W., & Kim, S.-H. (2023). Prediction of shipping cost on freight brokerage platform using machine learning. Sustainability, 15(2), 1122. ttps://doi.org/10.3390/su15021122

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.

Liang, Y., Chen, Z., & He, X. (2024). Intelligent pricing decision support system for the logistics industry based on big data platform. International Journal of Logistics Research and Applications, 27(4), 523–540.

Michel, S., & Siegfried, P. (2020). Digital freight forwarders in food logistics. Transportation Research Procedia, 46, 65–72.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis (6th ed.). Wiley.

Muha, R. (2019). Problematic issues in logistics cost management. Pomorstvo, 33(1), 102–109. https://doi.org/10.31217/p.33.1.11

Muthutantrige, D., Hasan, R., & Shah, B. (2023). Project management framework for digitalization of logistics processes. IEEE Access, 11, 45321–45335.

Oliveira, P., Zancul, E., & Salerno, M. (2024). Capability building for digital transformation through design thinking. Journal of Service Design, 15(1), 35–50.

Prasetyo, H., & Sutopo, W. (2018). The application of multiple linear regression for logistics cost estimation in freight forwarding. IOP Conference Series: Materials Science and Engineering, 337(1), 012046. https://doi.org/10.1088/1757-899X/337/1/012046

Shi, H., Guo, Y., & Wang, J. (2019). Application of LightGBM in demand forecasting for logistics. Procedia Computer Science, 162, 646–653. https://doi.org/10.1016/j.procs.2019.12.037

Wang, S., Meng, Q., & Sun, Z. (2018). Forecasting container transport demand with a multiple linear regression model. Maritime Policy & Management, 45(7), 985–1000. https://doi.org/10.1080/03088839.2018.1446105