NUANSA INFORMATIKA
Volume 19 Number 2 July 2025

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

Predicting Basic Shipping Tariff Using Machine Learning

Nisa Hanum Harani[1], M. Yusril Helmi Setyawan[2], Dani Ferdinan*[3]
[1,2,] Universitas Logistik dan Bisnis Internasional, Bandung
E-mail: [1]nisa@ulbi.ac.id,[2]yusrilhelmi@ulbi.ac.id, *[3]**daniferdinandall@gmail.com**

Abstract

This study explores the application of machine learning algorithms in predicting the Basic Shipping Tariff for logistics, focusing on variables such as Item Price, Shipment Weight, and Distance (KM). Random Forest Regressor and Linear Regression models were used as comparison methods. Experimental results show that the Random Forest Regressor outperforms Linear Regression, achieving an R^2 value of 0.915 and RMSE of 0.154, while Linear Regression reached an R^2 value of 0.706 and RMSE of 0.113. Additionally, the Random Forest model achieved lower error values with MSE of 0.000 and MAE of 0.003, compared to Linear Regression with MSE of 0.001 and MAE of 0.007. These error metrics further highlight the superiority of the Random Forest model. In-depth analysis reveals significant relationships between these variables and the Basic Shipping Tariff, showcasing the model's potential application in dynamic pricing strategies within the Indonesian logistics industry. This study aims to contribute to operational efficiency and improve pricing accuracy in the logistics business in Indonesia.

Keywords— *Machine Learning, Random Forest Regressor, Linear Regression, Basic Shipping Tariff, Logistics*

Submission: May 07, 2025 *Accepted: June 14, 2025* *Published: July 10, 2025*
DOI Article : https://doi.org/10.25134/ilkom.v19i2.388

1. INTRODUCTION

The logistics industry in Indonesia plays a vital role in supporting national economic growth. In recent years, this sector has undergone significant transformation, particularly due to the increasing volume of e-commerce transactions, which has driven the demand for fast and efficient delivery services. However, complex challenges such as the efficient management of operational costs and the determination of optimal shipping rates have become increasingly urgent. In this context, logistics companies are faced with the need to respond to the ever-evolving market dynamics without compromising operational cost efficiency [1].

One of the main issues faced by the logistics industry is the method of determining shipping rates, which is still often conducted manually or based on static policies. The Theory of Pricing in Logistics refers to the theoretical and practical framework used by logistics companies to determine the cost of their services. This approach involves a comprehensive analysis of operational costs, market demand, and competitive dynamics to ensure that the prices set not only cover all expenses but also maximize profit margins and market competitiveness [2].

Pricing models in logistics also take into account external factors, such as government regulations and economic fluctuations, which can significantly influence pricing structures. In a study conducted by [3], pricing was simulated using a game theory approach, which serves as a crucial element in shaping operational decisions. The game theory approach involves a thorough analysis of market demand and competitive dynamics to ensure that the prices set maximize profit margins [4].

Challenges in logistics cost management have also been highlighted, wherein efforts to reduce costs in one area may lead to increased costs in another [5]. Therefore, logistics companies need to adopt pricing strategies that are adaptive and responsive to market fluctuations.

This approach is not fully responsive to changes in demand or other variables that affect costs, such as item weight, item value, and delivery distance. Inefficiencies in rate setting can result in significant negative impacts, including a decline in the competitiveness of logistics companies in an increasingly competitive market.

In addressing these challenges, machine learning technology offers a promising solution [6]. Machine learning is a branch of artificial intelligence that enables systems to automatically learn from data and improve from experience without being explicitly programmed [7]. It focuses on building predictive models that identify patterns and relationships between variables, allowing future outcomes to be estimated based on past data [8]. Predictive modeling, as a core application of machine

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

learning, involves training algorithms on historical datasets to forecast values such as shipping costs, demand, or delivery times [9]. By leveraging predictive algorithms, companies can optimally analyze and utilize data to identify patterns and relationships among complex variables. Previous studies, such as those conducted on Alibaba's Cainiao system, have demonstrated that the implementation of machine learning can significantly enhance pricing efficiency. While such international implementations have proven effective, empirical machine learning-based models using real-world Indonesian logistics data remain underrepresented. Most existing research still focuses on global platforms like Alibaba [10], leaving a gap in localized, data-driven pricing models tailored to the operational context of Indonesian logistics companies. This highlights the substantial potential of this technology in overcoming the issues faced by the logistics industry.

This study focuses on the application of the Random Forest Regressor and Linear Regression algorithms to develop a model for predicting the base rate of logistics shipping. Random Forest is an ensemble algorithm composed of multiple decision trees that collectively work to produce more accurate predictions. Its main advantage lies in its ability to capture non-linear relationships within the data, making it particularly suitable for handling complex datasets [11]. The method employs a bagging technique, where multiple trees are trained on different subsets of the data to reduce the risk of overfitting. The average prediction from all trees is then used to generate a more stable and accurate final result.

On the other hand, Linear Regression is a simpler statistical method that aims to represent the relationship between one or more independent variables and a dependent variable through a linear equation [12]. Although this method is easier to interpret, its performance tends to decline when dealing with non-linear relationships or data involving complex interactions among variables. Therefore, it is important to compare these two methods in the context of shipping rate determination.

In evaluating model performance, the key metrics used include R^2 (R-squared), MSE (Mean Squared Error), RMSE (Root Mean Squared Error), and MAE (Mean Absolute Error). R^2 measures the extent to which the model explains the variance in the data, while MSE and RMSE assess the level of prediction error, with RMSE offering a more intuitive interpretation in the original units of the data. MAE, on the other hand, indicates the average absolute error [13], providing a direct measure of model accuracy.

By utilizing historical data from PT Pos Indonesia, this study is expected to offer a more adaptive approach to improving pricing prediction accuracy, while also providing relevant insights for the Indonesian logistics sector in addressing the challenges of global competition.

2. RESEARCH METHODS

This study adopts a quantitative approach grounded in positivist principles, aiming for objective and empirically testable results through numerical data analysis [14]. The process follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, which includes six standardized phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [15]. Developed collaboratively by Daimler Chrysler, SPSS, and NCR, CRISP-DM has been widely used since its official release in 1999 as a robust framework for solving data-driven problems across industries.

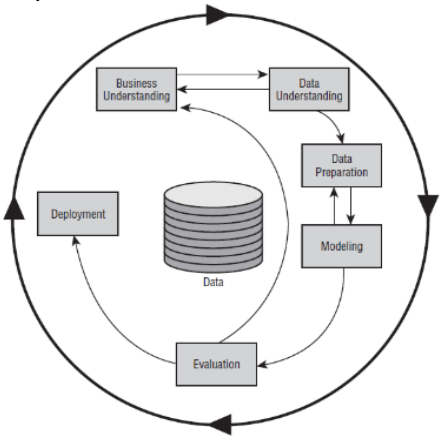

Figure 1. Research Methods

2.1. Bussiness Understanding

The objective of this study is to improve the accuracy of base rate predictions in logistics by utilizing historical data analysis from PT Pos Indonesia. It is expected that the findings of this research will support the company in automating its pricing strategy.

2.2. Data Understanding

The PT Pos shipping dataset also includes several additional columns that provide detailed information related to the shipping process. The "Nomor Resi" serves as a unique

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

identifier for each package shipped. The "ExtID" record an additional identification extension for company-specific purposes. "Kota Penerima" indicates the destination location of the recipient, while "Kantor Asal" and "Kantor Tujuan" record the origin and final destination offices of the shipment. "Bea Dasar" refers to the total shipping cost calculated based on various factors. "Harga Barang" reflects the declared value of the item being shipped. "Total" refers to the sum of the Bea Dasar (base shipping cost) and the Harga Barang (price of goods), and "Berat Kiriman" records the total weight of the package. The "Produk" column categorizes the type of service or shipping product used. "Tanggal Kirim" notes the date the shipment was dispatched, and "Isi Kiriman" provides a description of the contents or type of goods being delivered.

2.3. Data Preparation

Data Preparation is a crucial initial stage in this study to ensure that the data to be analyzed is clean, standardized, and ready for use [16], [17]. The initial dataset consists of 3,446 entries containing information on Harga Barang, Berat Kiriman, and Bea Dasar. However, several entries were incomplete due to missing values. Records with missing data in critical variables such as Harga Barang or Berat Kiriman were removed to ensure the quality of the dataset. After this cleaning process, a total of 3,433 valid entries remained and were deemed suitable for modeling.

The next step involves calculating the distance between the sender and recipient cities. Coordinates for each city were obtained by aligning the city names in the dataset with coordinate data retrieved from a public repository available at https://github.com/yusufsyaifudin/wilayah-indonesia. This repository provides geographic coordinates (latitude and longitude) for various regions in Indonesia, enabling accurate mapping of each city in the dataset to its corresponding coordinates.

Once the coordinates were obtained, the distance between the sender and recipient cities was calculated using the Python library Geopy, a tool designed for working with geographical data. The calculated distance values were then added to the dataset under the column named "Jarak (KM)".

2.4. Model Development

The implementation phase of the prediction algorithm was carried out using two machine learning models: Random Forest Regressor and Linear Regression. These algorithms were selected due to their distinct characteristics, offering valuable insights into the effectiveness of ensemble-based approaches versus simple linear regression in predicting the Bea Dasar variable.

The implementation process began with splitting the data into two groups: training data (80%) and testing data (20%). This data partitioning was performed using the train_test_split function from the Scikit-learn library, which randomly divides the dataset into two subsets according to the predefined proportion. This process aims to ensure that the model is trained on a sufficiently large subset to prevent overfitting, while also maintaining a representative testing set for performance evaluation.

2.4.1. Linear Regression

Linear regression is a fundamental technique in statistical modeling used to predict quantitative values based on one or more predictor variables. This approach is based on the assumption that there is a linear association between the dependent variable and the independent variables. Simple linear regression involves a single predictor variable, whereas multiple linear regression includes more than one. Recent studies [18] have shown that linear regression models can be enhanced by integrating them with modern computational techniques, thereby improving their predictive performance and expanding their applications in various fields such as economics, healthcare, and more. These advancements have also led to more robust error estimation and model assessment in real-world applications. In statistics, linear regression analysis is a method for identifying causal relationships between variables. It is a widely used approach across multiple disciplines. Both regression and correlation are employed to measure and understand relationships between variables. In regression analysis, estimation equations are developed to model the patterns or functions that describe the relationships among variables.

Linear regression is divided into two types: Simple Linear Regression and Multiple Linear Regression. Simple Linear Regression is used to examine the linear relationship between two variables, namely the independent variable (X), which can be controlled, and the dependent variable (Y), which reflects the response to the independent variable.

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

The equation of Y in relation to X in linear regression is presented in Equation (1).

$$\hat{Y} = a + bX \quad (1)$$

In linear regression, $\hat{Y}$ represents the predicted value of the dependent variable based on the changes or influence of the independent variable. **a** is the value of Y when X is zero; it is the constant term, also known as the intercept, indicating the point where the regression line crosses the Y-axis. This reflects the initial value of Y in the absence of any influence from X. **b** is the slope of the regression line, or the regression coefficient, which indicates the magnitude of change in the dependent variable in response to changes in the independent variable. A positive (+) coefficient results in an upward-sloping regression line, while a negative (–) coefficient produces a downward slope. **X** refers to a specific value of the independent variable, which can be controlled or set in order to observe its effect on the dependent variable.

Technically, the value of **b** (or the regression coefficient) is the tangent of the angle of the regression line's slope. It represents the ratio or rate of change in the dependent variable relative to the change in the independent variable, once the regression equation has been established [19].

The next type of linear regression is Multiple Linear Regression, which involves a linear relationship between the dependent variable (Y) while maintaining the property of linearity. Researchers use this method to predict changes in the dependent variable (criterion) based on two or more independent variables that serve as predictor factors. Multiple regression analysis is applied when there are at least two independent variables involved [20].

The general form of the multiple linear regression equation is presented in Equation (2).

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n + e \quad (2)$$

In this context, **Y** represents the response or dependent variable, **a** is the constant (intercept), **e** denotes the error term (disturbance), $\mathbf{b_1, b_2, ..., b_n}$ are the regression coefficients for the independent variables, and $\mathbf{X_1, X_2, ..., X_n}$ are the predictor variables (independent variables).

If the dependent variable is associated with two independent variables, the multiple linear regression equation can be expressed as shown in Equation (3).

$$Y = a + b_1X_1 + b_2X_2 \quad (3)$$

Let **Y** be the dependent variable, while $\mathbf{X_1}$ and $\mathbf{X_2}$ are the independent variables. The parameter **a** represents the value of Y when both X_1 and X_2 are equal to zero. The coefficients $\mathbf{b_1}$ and $\mathbf{b_2}$ are parameters in the multiple linear regression model, where $\mathbf{b_1}$ indicates the change in Y, in specific units, when X_1 increases or decreases by one unit while X_2 is held constant. Similarly, $\mathbf{b_2}$ represents the change in Y, in specific units, when X_2 increases or decreases by one unit, assuming X_1 remains constant.

2.4.2. Random Forest Regressor

Forest Regression is an ensemble-based machine learning method used to predict numerical values of a target variable. This method is designed to address the limitations of a single decision tree model, such as overfitting, by constructing an ensemble of decision trees [21]. In this approach, the model operates by training multiple trees on randomly selected subsets of the data using a technique known as bagging (bootstrap aggregating) [22]. This process not only helps reduce model variance but also enhances generalization, allowing the model to make more accurate predictions on previously unseen data.

Each tree in the forest generates a prediction based on the given input, and the final prediction is obtained by averaging all the predictions made by the individual trees in the model. This method helps improve the overall stability and accuracy of the model [23]. In addition, Random Forest has the capability to capture non-linear relationships within the data, making it highly effective for handling complex datasets with numerous variables.

One of the key advantages of Random Forest is its ability to provide estimates of feature importance, allowing users to identify which variables have the most influence on the predictions. This is particularly valuable in the context of data analysis, where understanding the factors that affect outcomes can offer critical insights for informed decision-making [24].

This model was implemented using the Random Forest Regressor class from Scikit-learn, a widely used Python library for machine learning. Scikit-learn provides a variety of tools and functions that facilitate the building, training, and evaluation of Random Forest models. Through this library, users can easily configure model parameters such as the number of trees in the forest (n_estimators), the maximum depth of the trees (max_depth), and the splitting criterion

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

(criterion) to optimize model performance according to their specific needs.

Overall, Random Forest Regression is a powerful and flexible method for data analysis, particularly in the context of predicting numerical variables. With its ability to overcome overfitting, capture non-linear relationships, and provide insights into feature importance, this method is an excellent choice for various applications, including in the logistics industry, where accurate shipping rate predictions are essential to improving operational efficiency and enhancing competitive advantage.

2.5. Evaluation

The evaluation method is a critical step in this study to assess the performance of the prediction model on previously unseen data, namely the testing data [25]. It not only measures the model's ability to predict values on the training data but also evaluates its capacity to deliver accurate and reliable results when applied to new, unseen data.

The evaluation was conducted using several model performance metrics, namely R^2 (R-squared), MSE (Mean Squared Error), RMSE (Root Mean Squared Error), and MAE (Mean Absolute Error). These metrics were chosen because each provides a different perspective on how well the model predicts the value of basic shipping.

R^2 (R-squared): This metric measures the extent to which the model can explain the variance present in the data. R^2 values range from 0 to 1, where a higher value indicates that the model explains a greater proportion of the data variability. The closer the value is to 1, the better the model's ability to represent the existing variation.

RMSE (Root Mean Squared Error): RMSE calculates the average magnitude of prediction errors, expressed in the same unit as the target variable. It reflects how far, on average, the model's predictions deviate from the actual values. A lower RMSE indicates better model performance in predicting the value of Bea Dasar.

MSE (Mean Squared Error): MSE is the average of the squared differences between predicted and actual values. While it also provides insight into the magnitude of error, it is more sensitive to outliers compared to RMSE. A lower MSE suggests higher accuracy in the model's predictions.

MAE (Mean Absolute Error): MAE measures the average absolute difference between predicted and actual values. This metric offers a straightforward interpretation of model accuracy and is generally easier to understand compared to MSE and RMSE. A lower MAE indicates that the model has a smaller prediction error on average.

By employing this combination of metrics, the study provides a comprehensive analysis of the model's performance. This is essential to ensure that the resulting model is not only highly accurate but also stable and reliable when applied in real-world scenarios, such as shipping rate determination in the logistics industry. Such thorough evaluation is expected to offer deeper insights into the model's effectiveness and support more informed decision-making in the future.

2.6. Deployment

The analysis of results provides recommendations for integrating the model into the company's automated pricing system. This integration aims to support more accurate, data-driven tariff decisions, thereby enhancing operational efficiency and competitiveness within the logistics industry.

3. RESULTS AND DISCUSSION

3.1. Data Preprocessing

In the modeling process, distance data is required. This distance data was obtained by first retrieving the coordinates of the destination city, then calculating the distance from the origin city based on these coordinates. Once the distance data was obtained, data preprocessing was carried out.

To obtain the distance data, the coordinates were first identified by matching the city names in the dataset with a public coordinate dataset available on GitHub, accessible via the following link: https://github.com/yusufsyaifudin/wilayah-indonesia.

Subsequently, the distance between the origin and destination cities was calculated using the Python library Geopy, based on their geographic coordinates. Once the distance data was generated, data preprocessing could be initiated.

Following an analysis of variable relationships, the researcher selected Jarak (KM), Harga Barang, and Berat Kiriman as independent variables that influence the dependent variable, Bea Dasar. After selecting the variables to be used, missing values in the

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

three selected variables were removed. Out of the total 3,446 data entries, 3,433 remained free of missing values and were used in the modeling process.

3.2. Model Development

Once the dataset was cleaned, the modeling process could be conducted. In this phase, two models were employed: Linear Regression and Random Forest Regression.

The first step involved separating the independent features, namely Berat Kiriman, Jarak (KM), and Harga Barang. The data was then split into training and testing sets using the `train_test_split` method, with 80% of the data allocated for training and the remaining 20% for testing.

Subsequently, both the Random Forest and Linear Regression models were constructed. The Random Forest model was trained using the training data and consisted of 100 decision trees. Once the model was trained, predictions were made on the testing data to generate base rate (Bea Dasar) predictions.

A simple Linear Regression model was also built and trained on the same data as the Random Forest model. This model aimed to identify a linear relationship between the features and the target variable. After training, predictions were similarly performed on the test data.

3.3. Model Evaluation

The performance of both models was evaluated by calculating various evaluation metrics, namely R^2 (R-squared), MSE (Mean Squared Error), RMSE (Root Mean Squared Error), and MAE (Mean Absolute Error) for the two models: Random Forest Regressor and Linear Regression. The calculated results for each model are presented in Figure 2:

The evaluation results indicate that the Random Forest Regressor outperforms the Linear Regression model. A lower MSE value in the Random Forest model suggests a smaller prediction error. Additionally, the lower MAE further confirms the model's ability to produce more accurate predictions. However, the slightly higher RMSE observed in the Random Forest Regressor indicates a greater average prediction error, particularly when considered in the original scale of the target variable.

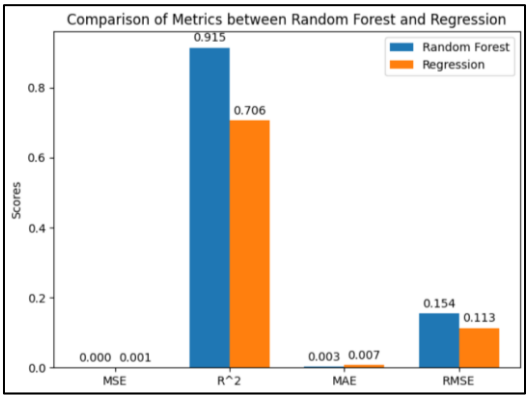

Figure 2. Comparison of Model Evaluation Results

The R^2 value for Random Forest reached 0.915, meaning the model was able to explain approximately 91.5% of the variance in the data, whereas Linear Regression only explained about 70.6% of the variance, with an R^2 of 0.706.

Based on these results, it can be concluded that the Random Forest Regressor demonstrates superior performance compared to Linear Regression, both in terms of predictive accuracy and the model's ability to explain data variability.

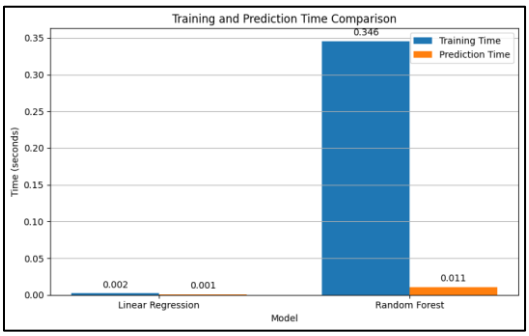

Figure 3. Comparison of Model Evaluation Results

Figure 3 illustrates the computational time comparison between the Linear Regression and Random Forest models during the training and prediction processes. Linear Regression demonstrates significantly faster computational performance, with a training time of 0.002 seconds and a prediction time of 0.001 seconds. In contrast, the Random Forest model requires 0.346 seconds for training and 0.011 seconds for prediction. These findings indicate that although Random Forest generally provides higher accuracy, it incurs a higher computational cost compared to the more efficient Linear Regression model.

3.4. Advantages of Random Forest

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

Random Forest offers significant advantages in handling complex relationships between variables. This model is capable of capturing non-linearity within the data, something that cannot be adequately modeled by linear approaches such as Linear Regression. By aggregating multiple decision trees, Random Forest can account for interactions among variables, resulting in more accurate predictions, especially in scenarios where the relationships between variables are not strictly linear.

In contrast, Linear Regression is limited in its ability to handle high-complexity data, as it assumes a purely linear relationship between predictor variables and the target. This limitation leads to lower performance when dealing with data that exhibits more intricate patterns, as reflected in the evaluation results.

3.5. Relationships Between Variables

A positive correlation was found between Harga Barang, Berat Kiriman, and Jarak (KM) with Bea Dasar. Among these variables, Jarak (KM) exhibited the strongest correlation, which aligns with the increasing operational costs as delivery distance grows. Berat Kiriman also showed a strong influence on Bea Dasar, reflecting the additional costs associated with the mass of the shipment. Meanwhile, Harga Barang demonstrated a more moderate but still significant correlation, indicating that the value of goods also contributes to the base rate calculation.

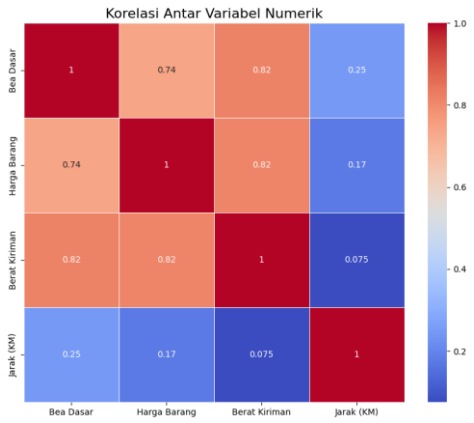

Figure 3. Feature Correlation

4. CONCLUSION

The Random Forest Regressor demonstrated superior performance in predicting Basic Shipping Tariffs (Bea Dasar) compared to Linear Regression. It achieved a higher coefficient of determination (R^2 = 0.915) and lower error metrics, including MSE = 0.000 and MAE = 0.003, outperforming Linear Regression which recorded R^2 = 0.706, MSE = 0.001, and MAE = 0.007. These results confirm the robustness of Random Forest in capturing the nonlinear relationships among key shipping variables such as Item Price, Shipment Weight, and Distance. However, it is worth noting that Random Forest also showed a slightly higher RMSE (0.154) compared to Linear Regression (0.113), indicating greater variance in certain prediction instances, which may need to be addressed for real-world deployment.

This study confirms the potential of machine learning as a strategic tool for enhancing operational efficiency and pricing accuracy in the logistics industry. For future research, it is recommended to enrich the dataset by including additional variables such as commodity type, packaging costs, and insurance value, and to extend the historical data range to capture seasonal effects. Additionally, developing an interactive web-based system that integrates the trained model could facilitate automated pricing, improve user engagement, and support decision-making in real-time logistics operations.

5. SUGGESTION

Based on the findings of this study on base rate prediction, several recommendations can be considered for further development.

5.1. Enriching the Shipping Dataset

To improve the accuracy of the predictive model, it is recommended to expand the dataset by incorporating additional relevant variables, such as commodity type, country of origin, insurance value, packaging costs, and previous shipment history. Furthermore, utilizing historical data over a longer time span may help the model capture seasonal patterns and long-term trends more effectively.

5.2. Web Interface Development

To ensure that the research outcomes are more accessible and useful to end-users, a more interactive and responsive web interface should be developed. The website design should prioritize user-friendliness and include features such as prediction visualizations, manual or file-based data input options, and informative summaries of prediction results.

REFERENCES

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

[1] K. Deorah, "Why Logistics Technology Matters," *Forbes Technology Council*, Dec. 2021.

[2] A. Zhahra Lubis, L. Lastrian Nahulae, N. Marliana Anggraini, R. Adawiyah, and U. Islam Negeri Sumatera Utara, "Analisis Faktor-Faktor yang Memengaruhi Penetapan Harga," *Jurnal Masharif al-Syariah: Jurnal Ekonomi dan Perbankan Syariah*, vol. 9, no. 1, pp. 25–28, 2024, doi: 10.30651/jms.v9i1.21412.

[3] J. Kong, Z. Chen, and X. Liu, "A Review of Logistics Pricing Research Based on Game Theory," *Sustainability (Switzerland)*, vol. 14, no. 17, 2022, doi: 10.3390/su141710520.

[4] A.-F. Raka Praditya and S. Yulianto Joko Prasetyo, "Game Theory Dalam Penentuan Strategi Pemasaran Optimal Dalam (Studi Kasus Persaingan E-Commerce Shopee dan TokoPedia)," *TIN: Terapan Informatika Nusantara*, vol. 2, no. 2, 2021, Accessed: Dec. 13, 2024. [Online]. Available: https://ejurnal.seminar-id.com/index.php/tin/article/view/799

[5] R. Muha, "An overview of the problematic issues in logistics cost management," *Pomorstvo*, vol. 33, no. 1, pp. 102–109, 2019, doi: 10.31217/p.33.1.11.

[6] N. T. Nafisah, F. Maria, M. R. Amanatullah, and T. Sutabri, "Penggunaan Teknologi Artificial Intelligence Untuk Peningkatan Pembelajaran Pada SMA Nurul Iman Palembang Menggunakan ITIL V3," vol. 18, no. 1, pp. 34–40, 2024, [Online]. Available: https://journal.fkom.uniku.ac.id/ilkom

[7] William F. Schneider and Hua Guo, "Machine Learning," *Journal of Physical Chemistry Letters*, vol. 8, no. 12, pp. 2689–2694, Jun. 2017, doi: 10.1021/acs.jpclett.7b01072.

[8] D. N. Argade, S. D. Pawar, V. V. Thitme, and A. D. Shelkar, "Machine Learning: Review," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 251–256, Jul. 2021, doi: 10.48175/ijarsct-1719.

[9] E. Akyuz, K. Cicek, and M. Celik, "A Comparative Research of Machine Learning Impact to Future of Maritime Transportation," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 275–280. doi: 10.1016/j.procs.2019.09.052.

[10] X. Li, Y. Zheng, Z. Zhou, and Z. Zheng, "Demand Prediction, Predictive Shipping, and Product Allocation for Large-scale E-commerce," *Social Science Research Network (SSRN)*, 2018, [Online]. Available: https://ssrn.com/abstract=3277125

[11] I. Maulita and A. Mu'amar Wahid, "Prediksi Magnitudo Gempa Menggunakan Random Forest, Support Vector Regression, XGBoost, LightGBM, dan Multi-Layer Perceptron Berdasarkan Data Kedalaman dan Geolokasi Predicting Earthquake Magnitude Using Random Forest, Support Vector Regression, XGBoost, LightGBM, and Multi-Layer Perceptron Based on Depth and Geolocation Data," *Jurnal Pendidikan dan Teknologi Indonesia (JPTI).jpti*, vol. 470, no. 5, pp. 221–232, 2024, doi: 10.52436/1.jpti.470.

[12] A. Maurice, C. Refiyana, and E. A. Vefiadytria, "Uji Asumsi Klasik dalam Regresi Linier pada Perhitungan Menggunakan Laporan Keuangan di Sektor Telekomunikasi Bursa Efek Indonesia (BEI)," *Jurnal Ilmiah Manajemen Ekonomi Dan Akuntansi*, vol. 1, no. Februari, pp. 107–118, 2024, doi: 10.62017/jimea.

[13] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput Sci*, vol. 7, no. 3, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[14] A. Shabur, M. Amadi, and N. Anwar, "Perbandingan Metodologi Studi Islam Tradisional dan Modern di Indonesia," *Jurnal Pendidikan Tambusai*, vol. 7, no. 3, pp. 22519–22526, 2023, doi: https://doi.org/10.31004/jptam.v7i3.10134.

[15] M. S. Murugaiyan and S. Balaji, *Succeeding with Agile software development*. 2012.

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

[16] R. Rianti, R. Andarsyah, and R. M. Awangga, "Penerapan PCA dan Algoritma Clustering untuk Analisis Mutu Perguruan Tinggi di LLDIKTI Wilayah IV," *NUANSA INFORMATIKA*, vol. 18, no. 2, pp. 67–77, 2024, [Online]. Available: https://journal.fkom.uniku.ac.id/ilkom

[17] E. Setiana and V. Retreva Danestiara, "Analisis Sentimen Pelaksanaan Kuliah Online Menggunakan Algoritma Support Vector Machine," *NUANSA INFORMATIKA*, vol. 17, no. 2, pp. 66–70, 2023, [Online]. Available: https://journal.fkom.uniku.ac.id/ilkom

[18] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 140–147, 2020, doi: 10.38094/jastt1457.

[19] P. D. Sugiono, *Statistik Untuk Penelitian*. 2007.

[20] M. I. Hasan, *Pokok-pokok materi statistik 1 (statistik deskriptif)*. 2003.

[21] M. Aditya Pratama, M. Munawaroh, W. Joko Pranoto, P. Studi Teknik Informatika, F. Sains dan Teknologi, and U. Muhammadiyah Kalimantan Timur, "Perbandingan Performa Algoritma Linear Regresi dan Random Forest untuk Prediksi Harga Bawang Merah di Kota Samarinda," *Jurnal Ilmu Teknik*, vol. 1, no. 2, pp. 172–182, 2024, doi: 10.62017/tektonik.

[22] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, "Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 187–192, Feb. 2021, doi: 10.29207/resti.v5i1.2813.

[23] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)*, vol. 4, no. 1, pp. 2723–1453, 2023, doi: 10.52158/jacost.491.

[24] C. Clara Afrisca, H. N. Rofiq, D. D. Atmoko, K. Keuangan, and R. Indonesia, "Penilaian Properti: Penggunaan Machine learning untuk Prediksi Nilai Sewa," *Jurnal Manajemen Keuangan Publik*, vol. 8, no. 2, pp. 138–155, 2024, doi: https://doi.org/10.31092/jmkp.v8i2.2922.

[25] K. Faizin, U. Islam, N. Sunan, and A. Surabaya, "Analisis Penggunaan Metode Penelitian Evaluasi Pada Penelitian Bahasa Arab Model Pengembangan," *Tabyin: Jurnal Pendidikan Islam*, vol. 03, no. 01, pp. 39–53, 2020, [Online]. Available: http://e-journal.stai-iu.ac.id/index.php/tabyin