

Predicting Container Delivery Dates Using Machine Learning Techniques: A Regression Approach

Godwin Nwachukwu NKEM^{1*}, Adedeji Daniel GBADEBO²

1. Operation Department, Lekki Freeport Terminal, Lagos, Nigeria

2. Department of Accounting Science, Walter Sisulu University, Mthatha, South Africa.

*e-mail : nkem.godwinnwachukwu@gmail.com

email: agbadebo@wsu.ac.za

Article Info

Keywords:

Seaport;

Container Terminal;

Inland Delivery;

Machine Learning;

Random Forest Model

Received: 2025-06-08; **Reviewed:** 2025-09-15;

Revised: 2025-10-28; **Accepted:** 2025-10-29;

Published: 2025-11-29

Abstract

Inland container delivery constitutes a critical component of the global maritime logistics chain, acting as the final phase that connects international ports to inland destinations. Accurate prediction of inland container delivery times is crucial for enhancing operational efficiency, minimizing demurrage and detention costs, and improving customer satisfaction across global supply chains.

Purpose – This study leverages historical container movement data across key international ports to develop a robust machine learning model for predicting inland container delivery timelines.

Methodology – Using a Random Forest Regressor, the model was trained to forecast the total inland delivery time based on features such as container size, type, shipping line, dispatch weekday, and temporal patterns.

Findings – The findings have practical implications for shipping lines, freight forwarders, port authorities, and inland terminal operators seeking to optimize logistics planning, reduce uncertainty, and improve supply chain. Evaluation of the model's performance yielded a Mean Absolute Error of 4.59 days, a Root Mean Squared Error of 10.55 days, and a coefficient of determination of 0.68, indicating moderate predictive accuracy. Supporting visualizations - including learning curves, gain curves, feature importance plots, residual distributions, and prediction bands - illustrate the model's strengths and areas for further refinement.

Originality – The study contributes to the growing field of intelligent logistics and maritime informatics by providing a data-driven framework for improving inland delivery predictability.

INTRODUCTION

Globalization is not just a trend but a force that extends local supply chains across boundaries, catering to customers' demands worldwide and changing the landscape of trade and commerce. With increasing globalization, many products and operations are outsourced and moved across countries (Yu,

M., & Qi, X. 2013). The impact of transportation on the supply chain has always been crucial and is ever-expanding (Viellechner, A., & Spinler, S. 2020). In international logistics, the timely delivery of containers is essential for maintaining supply chain efficiency. Delays in shipment can lead to significant financial losses and operational disruptions. Inland container delivery represents a critical segment in the end-to-end supply chain, acting as the final leg in maritime container transportation after discharge at international ports (Khiari, J., & Olaverri-Monreal, C. 2020). Once containers arrive at major global maritime gateways such as the Port of Singapore, Port Klang (Malaysia), Lekki Deep Seaport (Nigeria), or the Port of Rotterdam (Netherlands), they are typically moved inland to final destinations like bonded terminals, inland container depots, warehouses, or customer premises (Giallombardo et al., 2023). These movements occur via various transport modes, including road, rail, and barge, depending on the regional infrastructure and logistics planning (Kavirathna et al., 2018).

This inland leg is often subject to significant variability due to port congestion, customs clearance delays, road conditions, and terminal handling inefficiencies. As such, timely and reliable inland container delivery is essential to ensure predictability in supply chain operations, minimize additional costs (e.g., demurrage and detention), and maintain customer satisfaction (Hanedar et al., 2024). Effective inland container tracking hinges on robust data integration from multiple touchpoints, including port systems, shipping lines, inland depots, and transportation providers. This data integration is not just a technical requirement, but a crucial aspect of supply chain management (Padi, J. N., & Setty, V. 2024). However, in many cases, real-time visibility is limited, and processes remain fragmented. This lack of synchronization creates uncertainty and operational inefficiencies, highlighting the need for comprehensive data integration (Padi, J. N., & Setty, V. 2024).

Traditional methods for estimating delivery times rely on historical averages and manual adjustments, which lack precision. However, recent advancements in Machine Learning (ML) offer the potential to revolutionize this process. ML can outperform conventional approaches by creating dynamic, data-driven predictive models, leading to more accurate and reliable delivery time estimates (Notteboom, T., & Rodrigue, J. P., 2008; Darendeli et al., 2021). This paper explores the application of regression techniques to predict the actual delivery dates of containers, demonstrating the potential impact of ML on the supply chain.

The dataset used in this research captures detailed information about the inland movement process, recording the progression of each container from the Place of Dispatch through the Port of Loading, Port of Discharge, and the Post-Port of Discharge phase until it reaches its final Delivered Date. Time intervals such as "Dispatch to Loading," "Loading to Discharge," and "Post-Discharge to Delivery" provide crucial insight into operational bottlenecks, regional performance disparities, and transit time variability.

While applying ML in predicting inland delivery times offers significant benefits, it also comes with challenges. These challenges can range from data quality issues to the need for specialized skills and resources. ML can offer a robust solution by addressing these challenges (Jonquais, A., & Krempl, F. 2019). By leveraging historical data and incorporating factors such as container size, type, shipping line, and dispatch timing, these models can improve decision-making, enable proactive scheduling, and reduce the overall cycle time from port to consignee (de Araujo, A. C., & Etemad, A. 2021).

With major ports such as Singapore, Klang, Lekki Deep Seaport, and Rotterdam functioning as key transshipment and entry points for global cargo flows, the relevance of efficient inland logistics becomes even more pronounced (Yu, M., & Qi, X. 2013). Enhancing container delivery prediction and tracking in such corridors contributes to the performance of individual supply chains and the competitiveness of broader trade facilitation and logistics (Hathikal et al., 2020).

In the literature, extensive research exists for improving the operations of inland delivery efficiency from port of discharge in logistics from various perspectives via ML techniques that use extensive data. Therefore, ML is considered one of the most essential elements of Big Data Analytics (Jonquais, A., & Krempl, F., 2019).

The foundational work of Notteboom & Rodrigue (2008) examines the functional and spatial development of inland terminals and container ports. The effectiveness of container flow, according to the author, depends on inland logistics networks, especially dry ports and intermodal links. They created a conceptual model that highlights elements such as port congestion, transportation bottlenecks, and

regulatory impediments to represent the increasing complexity of hinterland logistics. Their results lend credence to the notion that inland variables, such as train schedules, terminal dwell durations, and the quality of regional infrastructure, must be included in prediction models for inland container delivery, in addition to sea data. This study establishes the foundation for determining important operational variables that ought to be fed into ML models, even if it does not employ ML itself.

Kavirathna et al. (2018) introduce a multi-criteria decision-making framework for evaluating container terminal performance based on operational KPIs such as berth occupancy, crane productivity, truck turnaround time, and gate throughput. While the authors used quantitative analysis and simulation techniques rather than ML, the amount of operational data they examined is extremely useful for delivery prediction. Their analysis demonstrates that internal port processes, such as poor stacking and delayed customs procedures, have a significant impact on final delivery timeframes. Integrating such insights into ML models, particularly Random Forests or Gradient Boosting, which handle structured tabular data well, can significantly improve forecast accuracy.

Khiari & Olaverri (2020) investigated various ML models for predicting delivery times in postal services in their paper "Boosting Algorithms for Delivery Time Prediction in Transportation Logistics". The study compared linear regression, bagging, and boosting algorithms, including LightGBM and CatBoost. The results showed that boosting algorithms beat others in terms of accuracy and runtime economy. LightGBM and CatBoost, in particular, demonstrated improved performance, suggesting that they are well-suited for predicting real-time delivery times in logistics operations.

Theofanis & Boile (2009) analysed inland container logistics from a policy and infrastructural standpoint. The evaluation of performance measures for inland terminals highlighted the necessity for public-private collaboration to enhance container flow efficiency. Their studies indicated that insufficient data sharing among stakeholders—terminal operators, transporters, and customs—generates considerable difficulties in delivery estimation. Although they did not employ ML methods, their policy analysis guides the development of intelligent transport systems (ITS) and digital logistics platforms that may ultimately provide the requisite real-time data inputs for ML models. They contend that predictive systems should be engineered to interact with human decision-makers, including logistics planners and port officials.

Giallombardo et al. (2023) examine predictive analytics in maritime logistics, with a focus on terminal operations and the integration of hinterland transportation. The authors demonstrated, through simulation and predictive modeling, utilizing neural networks and decision trees, that dwell time at the port is a critical predictor of inland delivery delays. They also evaluated models that included external variables, such as vessel schedule reliability and carrier-specific behaviors. A significant addition is the stratified approach to modeling - initially forecasting port performance and subsequently integrating that output into inland prediction models. This complex structure reflects the principles of Random Forests or deep learning ensembles, highlighting the sequential nature of delay propagation. Their findings support the use of modular predictive systems, each trained on a distinct transport layer.

The effectiveness of ML models, particularly ensemble techniques such as Random Forest and boosting algorithms, in forecasting delivery times and maritime logistics expenses is demonstrated by these studies collectively. Prediction accuracy is further enhanced by incorporating real-time contextual data, such as traffic and weather conditions. Other models, such as LightGBM and deep learning architectures, may perform better based on the application and data properties, even if Random Forest offers resilience and interpretability.

RESEARCH METHOD

The paper implements ML methods to predict the inland destinations covering shipment activities in the United States. We use real-world operational records for maritime container shipments to model inland delivery times. One of the primary challenges encountered during the project was the limited availability and transparency of data within the maritime shipping industry. Despite the global importance of container transport, public datasets remain scarce, outdated, or incomplete. As documented in the Review of Maritime Transport 2016 by UNCTAD, both port authorities and private

shipping lines often withhold data due to commercial confidentiality and strategic interests (Romeo, B. 2016).

Nonetheless, for this project, relevant shipment records covering the years 2021 and 2022 were obtained. The data tracks containers from major Asian ports, including Yantian (China), Shanghai (China), and Vung Tau (Vietnam), to inland destinations in the United States, such as Dallas and Gainesville, with Long Beach, California, as the primary U.S. port of discharge. Each container's journey was captured through key timestamps: dispatch, port loading, port discharge, post-discharge, predicted delivery, and actual delivery dates. From these timestamps, intermediate durations and total transit time were derived. This aggregated duration served as the target variable in the regression modelling.

Figure 1 describes the data handling processes. The dataset was pre-processed to clean missing and transform raw timestamps into proper time intervals representing each container journey segment. Feature engineering, transforming raw data into a more valuable and efficient format for ML, was employed to capture temporal dynamics and geographic transitions. This allowed the model to account for complex relationships such as port congestion, inland travel variability, and multimodal handling times. The dataset was kept intact by handling missing values, such as NaN, Blank values, null, or Continuous zeroes, by substituting them with the corresponding feature mean.

To enhance model performance and lower computational complexity, redundant or unnecessary features were eliminated from the dataset. A Standard Scaler was used in data cleaning to achieve a standardized distribution with a mean of zero and a standard deviation of one. It standardizes features by subtracting the mean value from the feature and dividing the result by the feature's standard deviation. Alternatively, these missing values can be handled by using techniques such as imputation (e.g., replacing missing values with the mean, median, or mode of the feature), deleting rows or columns with missing values, or advanced imputation methods, such as K-nearest neighbours (KNN) imputation.

Moreso, a pre-processing procedure called "categorical encoding" was used to translate categorical data into a numerical representation that ML algorithms can understand. This categorical encoding involves converting categorical variables into a numerical format that ML algorithms can use. Standard encoding techniques include one-hot encoding, label encoding, and target encoding. This step is essential because most ML algorithms require numerical input data. Categorical variables must be encoded for our ML models to properly use numerical input, which is how these models are trained. We use method of one-hot encoding to encode categorical values. The method set up binary dummy variables inside the categorical variable for every category. A binary column with a value of 1 denoting the presence of a category and a value of 0 denoting its absence is used to represent each category.

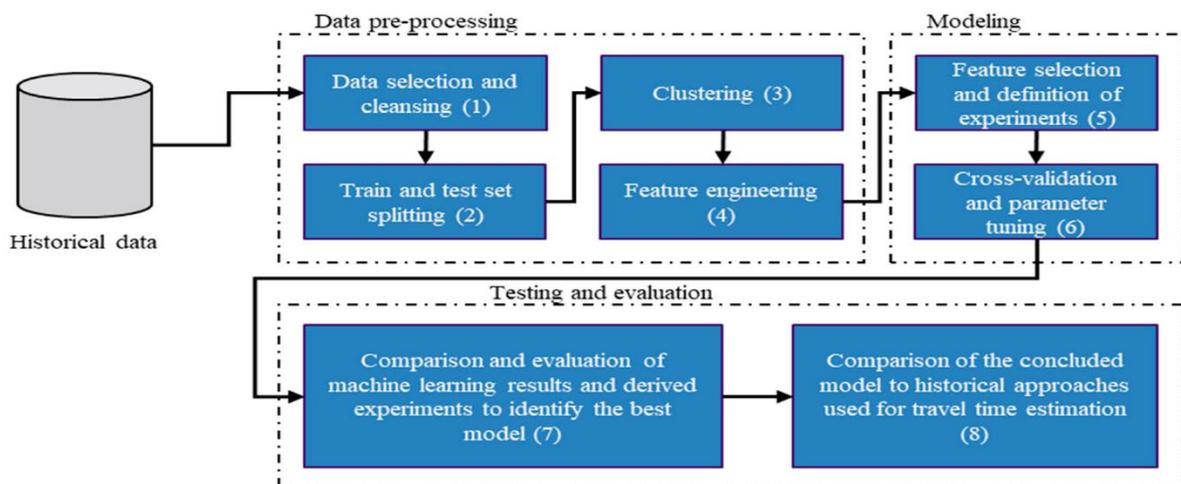


Figure 1: Data Handling Process

Prediction Models

Predictive modelling was conducted using several ML algorithms. Logistic regression, decision trees, support vector machines, Naïve Bayes, and k-nearest neighbours were initially evaluated for suitability (Hathikal et al., 2020). Logistic Regression, although widely used for classification, was not ideal for this regression-based task. It assumes a linear relationship between the log-odds of the outcome and the independent variables. While logistic regression performs well on categorical outputs, it is limited when dealing with non-linear and continuous target variables such as delivery time [11]. The logistic function is mathematically represented as:

$$p(X) = \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{(\beta_0 + \beta_1 X)}} \quad (1)$$

$$\frac{p(X)}{(1-p(X))} = e^{(\beta_0 + \beta_1 X)} \quad (2)$$

$$\log\left(\frac{p(X)}{(1-p(X))}\right) = \beta_0 + \beta_1 X \quad (3)$$

To better model complex, non-linear interactions present in the data, Random Forest Regression was selected for the final implementation. Random Forest is an ensemble learning method that constructs multiple decision trees and averages their outputs to improve accuracy and reduce overfitting. It efficiently handles high-dimensional, noisy, and heterogeneous data, and can model intricate feature interactions without requiring extensive preprocessing.

The mathematical representation of the Random Forest regression model is:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B h_b(X) \quad (4)$$

Where: \hat{y} is the predicted value, B is the total number of decision trees in the forest, and $h_b(X)$ is the prediction from the b -th decision tree for input X . Each tree h_b is trained using a different bootstrap sample and a random subset of features, allowing the best to learn diverse patterns and reducing the likelihood of overfitting.

Evaluation Metrics

Three statistical measures (Mean Absolute Error, Root Mean Square Error and Coefficient of Determination) were used to evaluate prediction accuracy, assuming actual values $a_1, a_2, a_3, \dots, a_n$ and predicted values $p_1, p_2, p_3, \dots, p_n$. These metrics are: *Mean Absolute Error (MAE)*: Measures the average magnitude of errors in a set of predictions, without considering their direction.

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - a_i| \quad (5)$$

Root Mean Square Error (RMSE): Penalizes larger errors more than MAE and gives a better sense of prediction reliability.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(p_i - a_i)^2}{n}} \quad (6)$$

Coefficient of Determination (R^2): Represents the proportion of variance in the dependent variable that is predictable from the independent variables.

$$R^2 = 1 - \frac{SS_E}{SS_T} \tag{7}$$

Where: SS_E is sum of squared errors; SS_T is total sum of squares; and R^2 is coefficient of determination.

RESULTS AND DISCUSSIONS

The outcome shows that the evaluation of the Random Forest regression model's performance yielded a Mean Absolute Error (MAE) of 4.59 days, a Root Mean Squared Error (RMSE) of 10.55 days, and a coefficient of determination of 0.68, indicating moderate predictive accuracy.

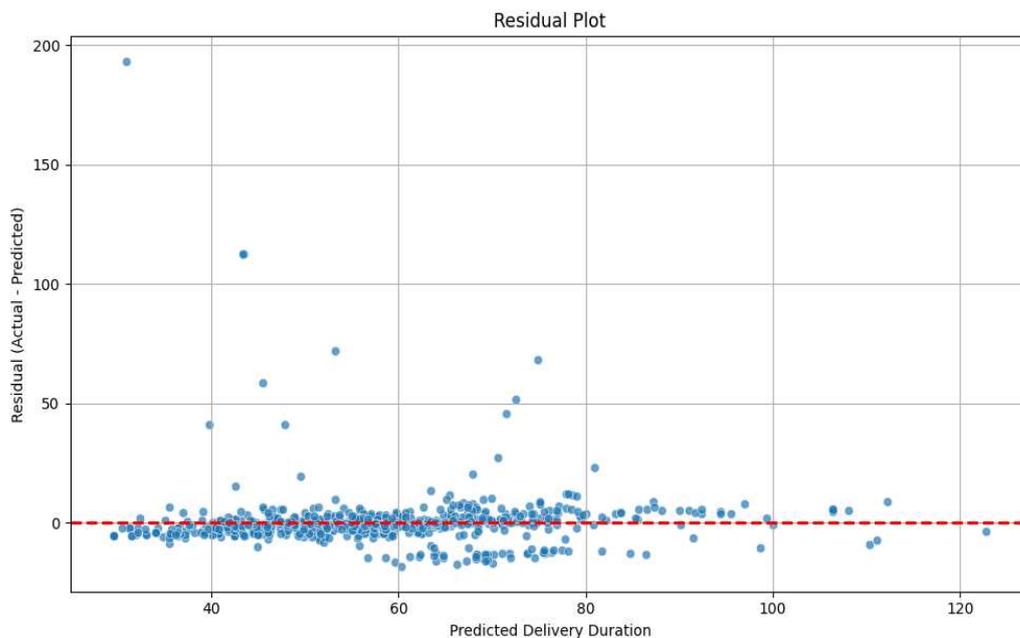


Figure 2: The residual of the model performance

Source: Authors (2025)

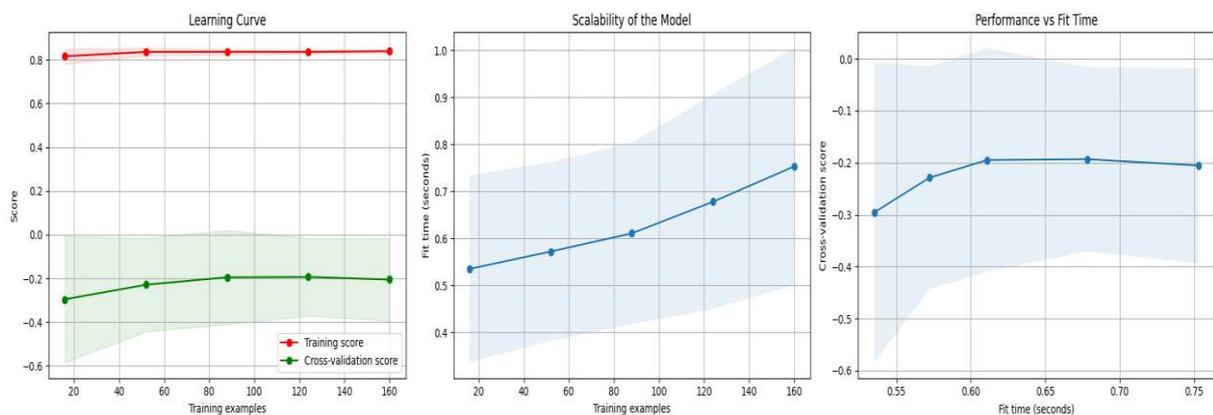


Figure 3: The model performance

Source: Authors (2025)

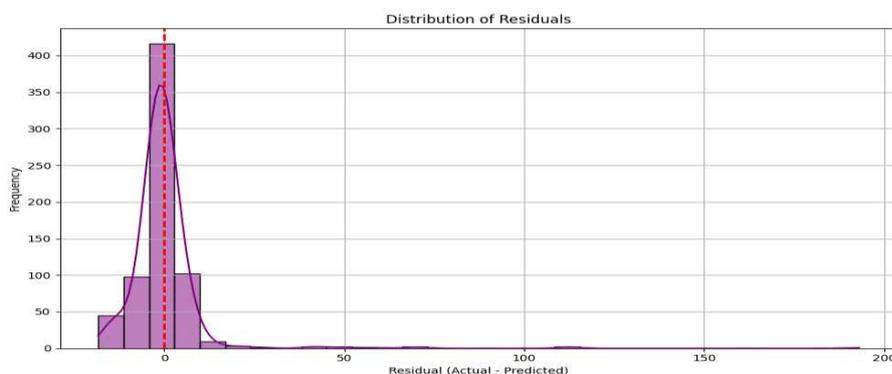


Figure 4: The residuals distribution of the model performance

Source: Authors (2025)

The Random Forest regression model is employed to forecast the duration of inland container deliveries. The performance and limitations of the model are illustrated in Figures 2, figure 3, and figure 4. In general, the residual of the model performance plot, which compares predicted values to residuals, demonstrates that many predictions are tightly clustered around zero, suggesting that the model is performing well. In certain instances, residuals are substantially positive, indicating that the model may be underestimating delivery times. These outliers likely reflect exceptional delays due to rare or unmodeled operational disruptions, highlighting the need for caution when interpreting predictions that fall near the model's performance boundaries. These outliers likely indicate significant delays resulting from infrequent or unaccounted operational disruptions, underscoring the necessity for prudence when interpreting predictions that approach the model's performance limits.

The model performance graphs underscore the model's constraints. Despite high and steady training scores, the cross-validation scores are low and relatively flat, suggesting that the model is overfitting: it excels on the training data but fails to generalize to unseen data. With more training examples the fit time increases as expected, but the gain in model performance plateaus early, showing limited benefit from additional training data under the current feature setup. This raises concerns about the model's robustness in operational settings, especially under variable or evolving conditions.

Lastly, the residuals distribution of the model performance plot demonstrates a steep peak near zero with a long right tail, suggesting that, despite the majority of predictions being in close agreement with the actual values, a small but significant subset of cases involves significant underestimation. This skewness suggests a potential operational risk, as delays may be systematically underestimated in specific scenarios. The Random Forest model is reasonably practical for general delivery duration estimation, as indicated by these plots. However, its operational deployment should be supplemented with cautionary mechanisms, such as the integration of additional context-specific features to improve generalization or the handling of exceptions for high-risk predictions.

Figure 5 compares the observed delivery times and the values predicted by a Random Forest regression model. Each point on the plot corresponds to a container shipment, with the x-axis representing the actual number of days taken for delivery and the y-axis representing the model's predicted delivery days. A dashed diagonal line indicates perfect predictions, where the predicted values match the actual ones. Visually, the actual vs predicted delivery days plot demonstrates that most points are closely clustered around the diagonal, suggesting that the model performs well in estimating delivery durations across most shipments. This alignment indicates a strong predictive ability of the Random Forest model, particularly in capturing the core dynamics influencing container delivery time. However, there are a few noticeable deviations from the diagonal, representing instances where the model either overestimated or underestimated delivery duration. These outliers may be attributed to unanticipated

real-world disruptions such as port congestion, customs clearance delays, weather-related issues, and familiar sources of variability in maritime logistics

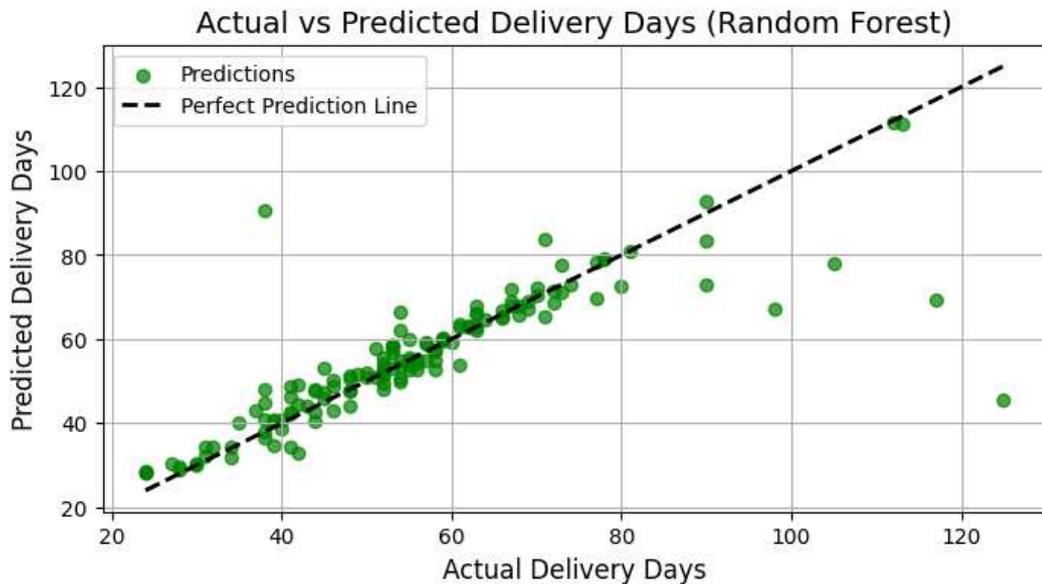


Figure 5: Actual vs Predicted Delivery Days (Random Forest)

Source: Authors (2025)

In practical terms, this plot provides strong evidence that the model is suitable for generating estimated times of arrival (ETAs) that stakeholders in the maritime supply chain, such as shipping lines, port authorities, and freight forwarders, can rely on for planning purposes. By improving the accuracy of ETA predictions, organizations can better allocate port resources, manage container yard operations, and provide more accurate delivery forecasts to customers, thereby improving overall supply chain visibility and efficiency.

The Cumulative Gain Curve for Long Delivery Prediction in Figure 6 provides insight into the model's ability to identify containers that are likely to experience high delivery delays. On the x-axis, the graph plots the proportion of containers sorted by predicted delivery delay. At the same time, the y-axis reflects the cumulative proportion of actual high-delay containers captured by the model. The blue curve represents the performance of the trained model, whereas the diagonal grey line represents a baseline random model, which assumes no learning or predictive capability.

This visualization shows that the model significantly outperforms random selection. For example, by examining just 20% of the containers predicted to have the highest delivery times, the model successfully captures around 70%–80% of all containers that experienced long delivery delays. This steep early rise in the gain curve indicates that the model is highly effective in prioritizing the riskiest shipments, those most likely to be delayed.

From a maritime logistics perspective, this is especially valuable. Port operators, logistics planners, or freight forwarders can focus their monitoring and contingency planning on a smaller subset of containers the model flags as high risk. This proactive focus enables better resource allocation, enhances customer communication, and facilitates more reliable scheduling. The ability to preemptively identify problematic deliveries before they occur can significantly reduce bottlenecks and improve supply chain resilience.

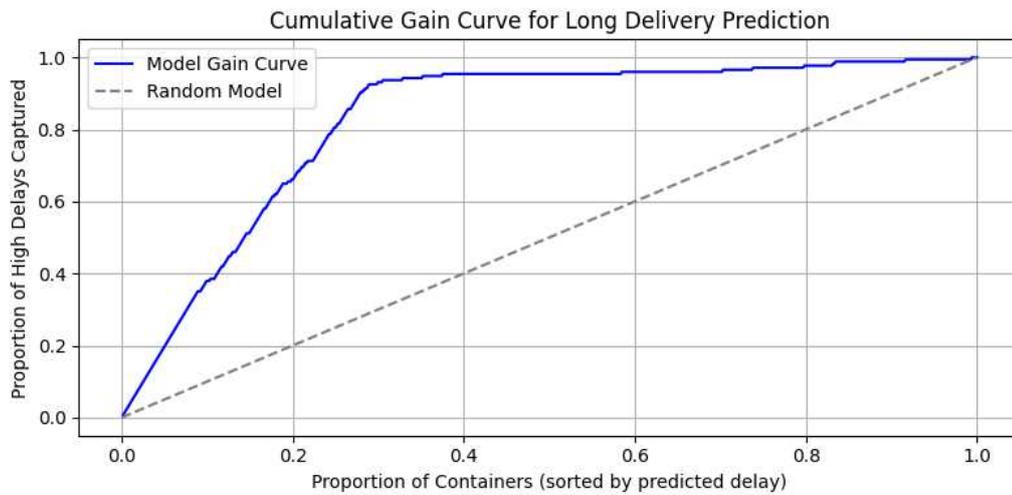


Figure 6: Cumulative gain curve for long delivery prediction

Source: Authors (2025)

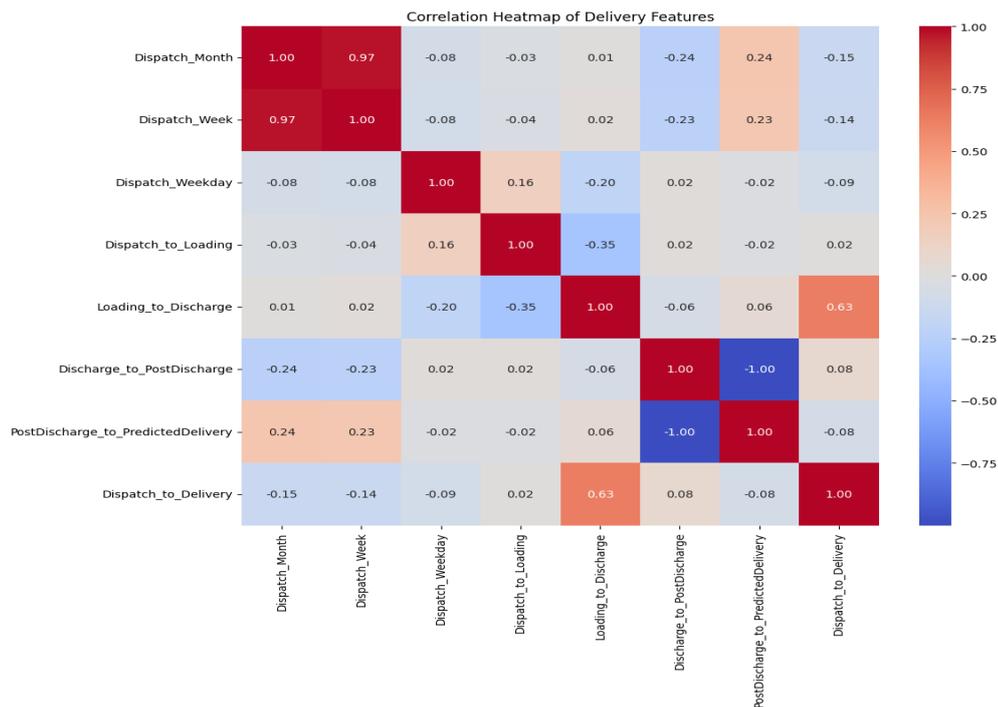


Figure 7. Delivery correlation heatmap

Source: Authors (2025)

Figure 7 is the delivery correlation heatmap. It provides an intuitive overview of how different stages and attributes in the maritime logistics process relate, particularly in terms of total delivery time (Dispatch to Delivery). Each cell in the heatmap represents the Pearson correlation coefficient between two variables, ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation), with color coding making it easier to spot strong relationships. One notable insight from the heatmap is the strong positive correlation (0.63) between Loading to Discharge and Dispatch to Delivery, indicating that delays in the main sea leg of the journey (from port of loading to port of discharge) significantly

contribute to the total delivery delay. This finding is particularly relevant for maritime operations, underscoring the importance of maintaining reliable ocean transit and adhering to schedule integrity.

Conversely, as expected, Dispatch Month and Dispatch_Week exhibit a robust correlation with each other ($r = 0.97$), since they are derived from the same timestamp. However, these time-based features have weak correlations with delivery delays, suggesting that seasonality alone is not a dominant driver of delivery performance in this dataset. Interestingly, there is a perfect negative correlation (-1.00) between Discharge_to_PostDischarge and PostDischarge_to_PredictedDelivery, which may indicate a data or modeling artifact—possibly due to how predicted and actual durations are computed or imputed. This warrants a deeper review to ensure no inverse relationships were introduced through feature engineering.

The delivery correlation heatmap helps stakeholders identify which stages of the delivery pipeline most influence overall performance, allowing targeted interventions such as reducing discharge-port dwell time or improving vessel scheduling accuracy to improve overall supply chain efficiency.

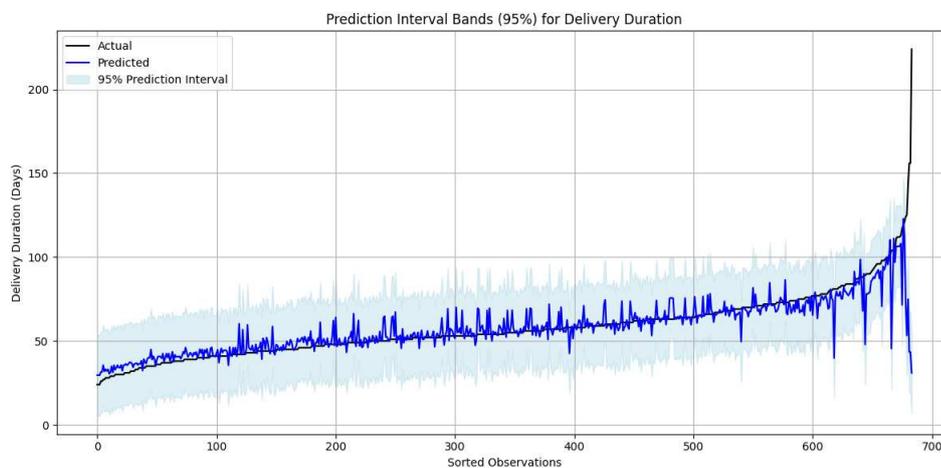


Figure 8. Interval bands for the delivery duration

Source: Authors (2025)

The prediction interval bands for the delivery duration in Figure 8 provide a compelling view into the uncertainty associated with the model's predictions across all delivery duration estimates. In this plot, the black line represents the actual delivery durations, sorted in ascending order, while the blue line denotes the model's predicted durations. The shaded area represents the 95% confidence interval, indicating the range within which the model predicts the actual delivery duration will fall with high certainty.

This visualization is especially valuable in maritime logistics, where unpredictable delays due to port congestion, weather events, labour issues, or customs procedures can significantly affect delivery performance. The wider bands seen toward the higher end of the sorted deliveries reflect the increased uncertainty for extreme delays, which is both expected and practically applicable. It suggests that while the model performs reliably for most typical deliveries, it appropriately acknowledges and accounts for variability in long-tail delivery times.

In operations, such prediction intervals enable risk-aware planning and decision-making. For example, supply chain managers can use these intervals to proactively buffer inventory levels or adjust customer commitments when the upper bounds of the prediction suggest potential severe delays. The reasonably tight band around most of the delivery range also reflects confidence in the model's generalization, reinforcing its practical value in real-time forecasting systems for container delivery.

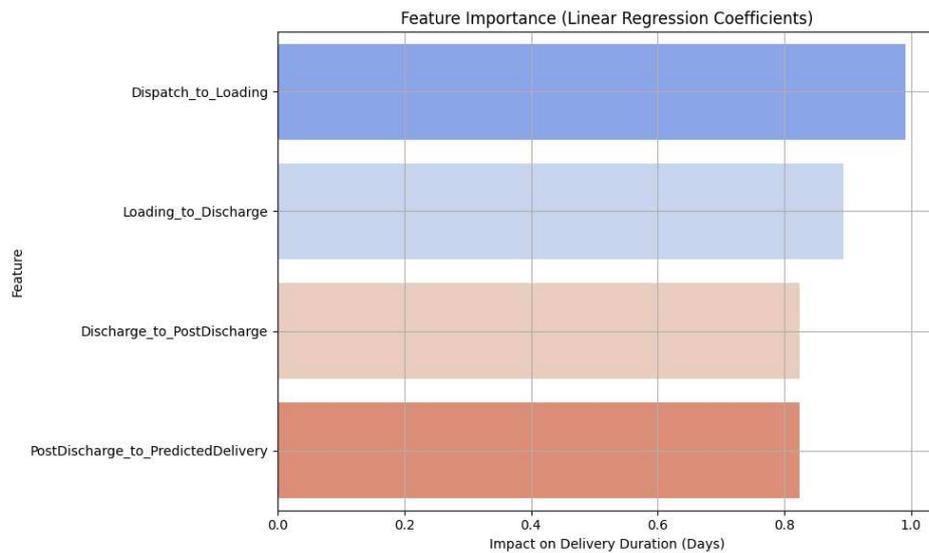


Figure 9: Feature importance (linear regression coefficients)

Source: Authors (2025)

The feature importance (linear regression coefficients) chart, depicted in Figure 9, is a crucial tool for understanding the relative influence of key delivery stages on the overall duration of container delivery. The regression coefficients, which are normalized for interpretability, indicate the degree to which each feature contributes to predicting total delivery time. Longer bars indicate a higher impact, providing valuable insights for maritime logistics planners. The most influential segment is Dispatch to Loading, which exhibits the highest coefficient, approaching 1.0. This underscores the urgent need to optimize pre-port processes, as delays at the very start of the logistics chain, often linked to inland haulage, documentation, or congestion at origin terminals, can significantly extend the overall delivery duration. It is a call to action to reduce end-to-end delays.

The Loading to Discharge feature encapsulates the maritime transit phase and is the next most impactful. This aligns with expectations in the marine context, where variations in voyage times due to vessel scheduling, weather disruptions, or port omissions can directly and often substantially affect delivery accuracy. It underscores the importance of stability in the maritime transit phase. Post-discharge to Predicted Delivery and Discharge to Post-Discharge show slightly lower but notable impacts. These stages include container handling, customs clearance, and last-mile coordination at the destination port. Though these delays may be shorter in absolute terms, they can become critical bottlenecks, especially in ports with high throughput or inefficiencies. Overall, this visualization provides actionable insight for maritime logistics planners: early-stage process efficiency and maritime leg stability are key levers for improving delivery time predictions and reliability. The potential benefits of targeted interventions at these stages can yield outsized benefits in reducing delivery variability and enhancing customer satisfaction, offering a promising future for the industry.

CONCLUSIONS

This study evaluated the predictive ability of a Random Forest regression model for forecasting inland container delivery times. The model had a Mean Absolute Error (MAE) of 4.59 days, a Root Mean Square Error (RMSE) of 10.55 days, and an R^2 score of 0.68, showing a good fit and ability to identify patterns in the data. Despite these positive results, residual analysis revealed heteroskedasticity and a significant underestimate in longer delivery durations, implying model bias and considerable operational risk. The learning curve also suggested overfitting, with high training scores (about 0.84) but significantly lower cross-validation scores (around -0.2), indicating a generalization gap caused by data sparsity or feature restrictions. To enhance the model's reliability and operational applicability,

future studies should consider incorporating more detailed and diverse features, such as real-time traffic data, weather conditions, and port congestion rates. Furthermore, balancing the dataset and looking at various ensemble or deep learning methodologies may increase predicted accuracy, particularly in outlier cases. These enhancements may help ensure the model's practical utility for logistical decision-making.

ACKNOWLEDGEMENTS

The authors would like to use this opportunity to acknowledge the following people for their support, suggestions and feedback during the period of this research: Augustine Nkem and Nkem Stephen.

REFERENCES

- Notteboom, T., & Rodrigue, J. P. (2008). Containerisation, box logistics and global supply chains: The integration of ports and liner shipping networks. *Maritime economics & logistics*, 10, 152-174.
- Kavirathna, C., Kawasaki, T., Hanaoka, S., & Matsuda, T. (2018). Transshipment hub port selection criteria by shipping lines: the case of hub ports around the bay of Bengal. *Journal of Shipping and Trade*, 3, 1-25.
- Khiari, J., & Olaverri-Monreal, C. (2020). Boosting algorithms for delivery time prediction in transportation logistics. In *2020 International Conference on Data Mining Workshops (ICDMW)* (pp. 251-258). IEEE.
- Theofanis, S., & Boile, M. (2009). Empty marine container logistics: facts, issues and management strategies. *GeoJournal*, 74, 51-65.
- Giallombardo, G., Mirabelli, G., & Solina, V. (2023). Optimal storage and shipment management of perishable agri-products with a hybrid priority policy: a case study. In *Sustainable Logistics Systems Using AI-based Meta-Heuristics Approaches* (pp. 157-168). Routledge.
- Darendeli, A., Alparslan, A., Erdoğan, M. S., & Kabadurmuş, Ö. (2021). Container demand forecasting using machine learning methods: a real case study from Turkey. In *Digital Conversion on the Way to Industry 4.0: Selected Papers from ISPR2020, September 24-26, 2020 Online-Turkey* (pp. 842-852). Springer International Publishing.
- Lee, Y., & Chao, S. L. (2009). A neighborhood search heuristic for pre-marshalling export containers. *European Journal of Operational Research*, 196(2), 468-475.
- Viellechner, A., & Spinler, S. (2020). Novel data analytics meets conventional container shipping: predicting delays by comparing various machine learning algorithms.
- Hathikal, S., Chung, S. H., & Karczewski, M. (2020). Prediction of ocean import shipment lead time using machine learning methods. *SN Applied Sciences*, 2(7), 1272.
- Jonquais, A., & Krempf, F. (2019). Predicting shipping time with machine learning.
- Padi, J. N., & Setty, V. (2024). Machine Learning Approaches to Predicting Sea Freight Carbon Emissions: A Comparative Evaluation of Decision Trees, Support Vector Machines, and Random Forests.
- Zhang, C., Bin, J., Wang, W., Peng, X., Wang, R., Halldearn, R., & Liu, Z. (2020). AIS data driven general vessel destination prediction: A random forest - based approach. *Transportation Research Part C: Emerging Technologies*, 118, 102729.
- Yu, M., & Qi, X. (2013). Storage space allocation models for inbound containers in an automatic container terminal. *European journal of operational research*, 226(1), 32-45.
- Hanedar, S., Ulus, C., & Akay, M. F. Development of Cargo Delivery Time Prediction Models. *Cukurova University Journal of Natural and Applied Sciences*, 3(1), 31-35.
- de Araujo, A. C., & Etemad, A. (2021). End-to-end prediction of parcel delivery time with deep learning for smart-city applications. *IEEE Internet of Things Journal*, 8(23), 17043-17056.
- Romeo, B. (2016). International Relations Development Strategies from the Perspective of Maritime Freight Transport Development. *Annals (Constanța Maritime University)*, 17(25).