**A Framework for Maritime Emission Prediction and Optimal Berth**

**Scheduling to Reduce Environmental Impact**

**1 Gollangi Ojaswini Sree, 2 Samuroutu Niveditha, 3 Kurmapu Hemanth Sai, 4 Aakifah**

1,2,3,4 Department of Information Technology, Anil NeeruKonda Institute of Technology and Sciences,

Visakhapatnam, Andhra Pradesh.

Email: 1ojaswinisree@gmail.com, 2 nivedithasamuroutu@gmail.com,3 hemanthsai8003@gmail.com,

4 aakifah968@gmail.com

**Abstract:** Despite being essential to international trade, the maritime sector greatly contributes to environmental pollution through carbon emissions and ineffective port operations. Emission forecasting and effective berth scheduling are two significant maritime logistics issues that are addressed in this study using an integrated approach that combines machine learning and optimization algorithms. The goal of the project is to create a predictive model that uses real-time data from environmental variables, fuel consumption, and vessel movements to accurately estimate maritime emissions. To improve the accuracy of emission forecasts, machine learning techniques will be applied to the analysis of both historical and current data. Furthermore, this study provides a framework for berth scheduling optimization that can meet different shipping demands while ensuring lower fuel consumption and fewer delays. Port operations will be streamlined using modern optimization algorithms that maintain both environmental sustainability and economic viability. By integrating these two components, the proposed approach enhances decision-making for maritime stakeholders and makes it possible to create informed scheduling and emission control plans. The results will reduce the carbon footprint of maritime logistics while improving operational effectiveness. This study is a significant step toward sustainable and intelligent maritime transportation, which will benefit the global economy and environment.

***Index******terms:***Maritime Industry, Port Operations, Berth Assigning, Emission, Sustainability.

**1. Introduction**

One essential element of international trade is maritime transportation. Maritime transportation is a vital component of international trade and logistics because it facilitates the movement of goods across borders. However, concerns regarding environmental sustainability are raised by the industry's significant contribution to greenhouse gas emissions, air pollution, and climate change. As ports and shipping operations expand, there is an increasing need for more efficient and sustainable maritime solutions [1].

One of the biggest environmental problems facing maritime operations is emissions from ships while they are berthing in ports and transiting through them. Ineffective berth scheduling increases carbon emissions by causing long wait times and excessive fuel use [2]. To address this issue, a comprehensive system for forecasting maritime emissions and scheduling berths is required. By combining state-of-the-art data analytics and optimization techniques, this system aims to increase operational efficiency while lowering the environmental impact of maritime logistics [3].

The economic and environmental advantages of this research make it significant. In addition to lowering congestion and increasing port efficiency, optimized berth scheduling also lowers operating expenses and fuel consumption [4]. Furthermore, precise emissions forecasting makes it possible to adhere to environmental standards, which lessens the negative climate effects of shipping operations.

Therefore, this research focuses on emission forecasting and berth scheduling optimization to improve maritime operations. A Random Forest-based machine learning model is developed to predict shipping emissions, helping ports make proactive decisions for sustainability. Additionally, a Mixed Integer Linear Programming (MILP) model is designed to optimize berth allocation, ensuring efficient port operations while minimizing environmental impact. The rest of this paper is organized as follows: Section 2 provides a review of existing literature, identifying key insights. Section 3 presents the system architecture and our strategic approach. Section 4 assesses the effectiveness of the proposed methodologies. Finally, the study concludes with key findings.

**2. Literature Review**

In recent years, there has been a growing interest in applying machine learning techniques to various challenges in maritime operations, particularly in reducing emissions, optimizing port logistics, and improving air quality. Several studies have explored how machine learning models can be leveraged to enhance environmental sustainability by forecasting and mitigating shipping emissions. Fletcher et al. developed predictive models using machine learning to create a comprehensive shipping emission inventory [5]. Similarly, Paternina-Arboleda et al. assessed sulfur dioxide emissions at maritime facilities, contributing to the development of greener port initiatives [6]. Expanding on this, Peng et al. introduced a machine learning model that incorporated green port activities, focusing on predicting energy consumption at ports, thus demonstrating the potential of data-driven sustainability approaches [7].

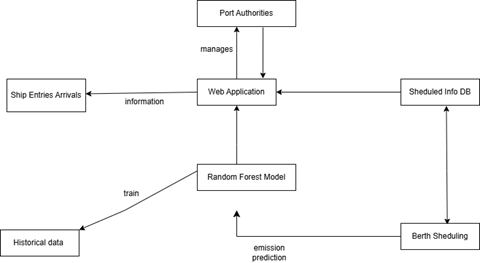
Various studies have focused on using machine learning to streamline port operations and optimize scheduling strategies. Guo et al introduced a multi-period coordinated optimization model for yard and berth assignment in container terminals, demonstrating how machine learning can enhance port logistics [8]. Beyond emissions reduction and air quality management, machine learning has also been applied to maritime infrastructure optimization, ship scheduling, and terminal management. Christiansen and Fagerholt emphasized the importance of robust ship scheduling strategies across different time frames [9].

These studies demonstrate the role of machine learning in improving maritime logistics and environmental outcomes by integrating predictive analytics, real-time data, and optimization techniques.

**3. Methodology:**

**3.1 Architecture:**

The proposed architecture integrates machine learning for emission forecasting with MILP-based berth scheduling through a centralized web application. As shown in Figure 1, our system collects real-time ship data, such as vessel type, size, fuel usage, arrival time, and cargo details. A web application that updates and retrieves berth schedules from a database process this data. We employ the Random Forest model, which helps port authorities make better decisions by analyzing historical and current data [10]



**Figure 1:** Architectural Flow of the Maritime Emission Prediction and Scheduling System

Simultaneously, berth scheduling is optimized through the use of Mixed Integer Linear Programming (MILP), which lowers fuel consumption and ship waiting times [10]. Lastly, the web application allows for real-time schedule monitoring and modification, improving the efficiency of port operations. The key components as shown in Figure 1 are:

**Ship Arrivals:** Ships offer up-to-date information on cargo details, estimated arrival time, fuel consumption, emission levels, vessel type, size, and weight. The web application processes this data to maintain the emissions models and berth scheduling current. Quick reactions to modifications in ship traffic and port conditions are made possible by the system's assurance of a smooth and accurate information flow. The system can improve berth assignments and forecast emissions by incorporating this real-time data, which will ultimately increase port efficiency and lessen its environmental impact [11].

**Web Application:** By processing real-time ship data and enabling communication between the berth scheduling engine, the emission prediction model, and the ship arrival system, the web application acts as the system's main control center. It serves as the port authorities' interface, enabling them to keep an eye on and modify schedules as necessary. Port operators can react swiftly to changes in ship arrivals and environmental conditions thanks to the web application's processing of real-time data and generation of actionable insights [12].

**Scheduled Data (DB):** In order to maintain consistency between ship arrival times and berth availability, the database keeps track of ship data and berth schedules. Berth allocations are continuously optimized thanks to its support for dynamic updates based on current ship status and port traffic.

Efficient information retrieval and storage are made possible by the database, which offers a solid basis for scheduling and emission prediction. This guarantees that port operations continue to be effective and flexible in the face of unforeseen delays or changes in ship arrivals.

**Random Forest Model (Emission Prediction):**

The system uses a Random Forest model to predict emissions based on historical data and real-time ship movement information. The model considers vessel size, speed, fuel type, and environmental factors to generate accurate emission forecasts. The output from this model is displayed through the web application, allowing port authorities to monitor and manage emissions more effectively. This predictive capability ensures compliance with environmental regulations and helps reduce the ecological impact of port operations [13].

**Berth Scheduling (MILP):** Mixed Integer Linear Programming (MILP) is used to optimize berth scheduling in order to minimize ship waiting times, improve berth utilization, and lower emissions and fuel consumption. The model enhances efficiency and sustainability by dynamically modifying schedules in response to port traffic and real-time ship arrivals [14].

**Port Authorities:** The web application is used by port authorities to control emissions and ship schedules. With the use of real-time data and forecasts, they can modify ship priorities and berth assignments. This aids in their decision-making to increase efficiency and lower fuel and emission levels [15].

**3.2 Implementation:**

**3.2.1 Software Requirements:**

The model was implemented using Python with Scikit-learn for training and deploying the Random Forest model for emission forecasting and Flask for creating a lightweight API to handle model predictions and scheduling updates.

**3.2.2 Dataset Description:**

The dataset used for training and testing consists of 1,000,000 entries with 9 features. It includes details such as ship type, size, fuel type, vessel age, fuel consumption, engine type, emission control technology, load factor, and emission scores.

**3.2.2 Emission Forecasting Using Machine Learning:**

We developed a machine learning-based predictive model to estimate ship emissions using Random Forest, analyzing both historical and real-time data for accurate predictions.

**3.2.2.1 Data Preparation:**

To ensure accuracy and reliability we have normalized numerical features, removed outliers, and handled missing values to ensure data accuracy. Additionally, Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance class distribution, improving the model’s fairness and predictive performance [16], where SMOTE is a method used to address class imbalance in datasets.

**3.2.2.2 Model Training:**

The Random Forest model was trained using an 80-20 data split, with 80% of the data for training and 20% for validation. The input data included ship type, fuel type, vessel age, engine type, fuel consumption, load factor, and emission control technology. The model was designed to predict ship emissions, categorizing them as Low, Medium, or High based on these factors. It follows a bootstrapping technique, creating multiple decision trees using different random subsets of data to improve generalization. At each split, a random subset of features was chosen to reduce overfitting. Hyperparameter tuning was done by adjusting the number of trees (50), maximum depth (10), and minimum samples per split (7) to ensure stability and reliable decision-making. During prediction, each tree provides an output, and the final result is obtained through majority voting for classification or averaging for regression, making the model more accurate and robust.

**3.2.2.3 Mathematical Approach Of Random Forest Model:**

Random Forest handles data by creating bootstrapped subsets, where the dataset is randomly sampled with replacement to generate multiple training sets. At each tree node, a random subset of features is selected for splitting, ensuring diversity among trees. The best feature is chosen based on Gini Impurity(G) for classification, which measures the likelihood of an incorrect classification if a random sample were classified according to the distribution of classes in a node.

It is calculated as:

(1)

From equation (1) Pi represents the probability of a class in the node, and i represents each class, c represents number of classes. A lower Gini Impurity value indicates a purer node with fewer mixed classes.

For regression tasks, the Mean Squared Error (MSE) is minimized:

(2)

From Equation (2) whereyi the actual value, and ​ is the predicted value, n is number of samples and I is index of each data point..

**3.2.2.4 Prediction and Evaluation:**

Emission levels were predicted using fuel consumption, and vessel size. Emissions were categorized by the model as High, Moderate, and Low orders. Evaluation was performed using delicacy, perfection, recall, and F1- score, with the model achieving 98% delicacy and all other criteria exceeding 96%, indicating high prophetic trustability and balanced performance.

**3.2.3 Berth Scheduling Optimization Using MILP:**

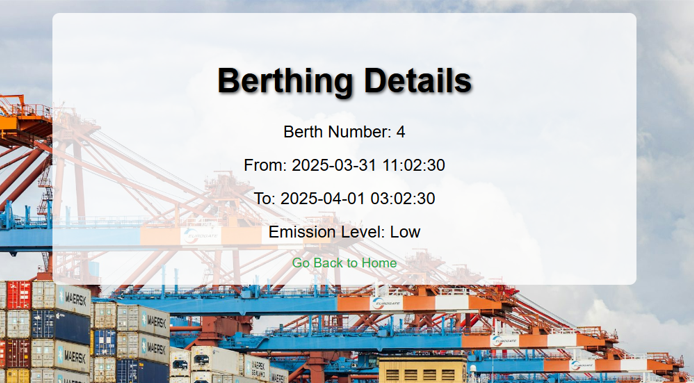
Berth scheduling can be optimized using Mixed Integer Linear Programming (MILP) by minimizing waiting times and fuel consumption while ensuring fair berth allocation. Optimizing berth can greatly increase port efficiency and lower emission [17].

**3.2.3.1 Scheduling and Optimization:**

The MILP model assigns ships to berths using decision variables such as berth assignment, start time, and waiting time. The objective function minimizes total ship waiting time and fuel consumption, balancing operational efficiency and sustainability. Constraints ensure berth capacity limits, prioritize high-ranking ships, and favor vessels with lower emissions.

Ships are categorized based on size, priority level, fuel consumption, and emissions. A priority score is assigned, favouring emergency cases and eco-friendly vessels. Ships are allocated berths immediately if available; otherwise, they enter a waiting queue, ordered by priority and emission index [18].

As shown in figure 3 it confirms that that a ship has been successfully assigned to a berth.



**Figure 3:** Berthing Details

**4. Results and Discussions:**

**4.1. Performance Metrics and Evaluation:**

After training, the machine learning model for ship emission forecasting was evaluated using multiple quantitative metrics. The testing confirmed that the model generalizes well to new data and accurately predicts ship emissions under varying port conditions.

**4.2 Evaluation Metrics:**

To assess the model’s performance, the following metrics were applied:

* **Mean Average Precision (mAP):**
  + mAP@50: Measures the model's accuracy at an Intersection over Union (IoU) threshold of **0.5**, confirming that the model accurately predicts ship emissions and berth schedules when moderate overlap is allowed.
  + mAP@50-95: Evaluates the model’s precision across a range of IoU thresholds (0.50 to 0.95), providing a comprehensive assessment of its detection and prediction capabilities.
* **Intersection over Union (IoU):** IoU measures the overlap between predicted and actual ship emissions and berth assignments. The model was tuned to achieve an IoU threshold of **0.7**, ensuring high alignment between predicted and actual outcomes.
* **Accuracy:** The model achieved an overall accuracy of 98%, indicating high consistency in predicting ship emissions and berth schedules based on real-time data and historical patterns. High accuracy reflects the model's ability to minimize both false positives and false negatives.
* **Precision and Recall:**
  + Precision: Measures the ratio of rightly predicted ship emissions and berth assignments to total predictions. The model demonstrated a high precision of 1.00, 0.98, and 0.96 across different classes, suggesting minimal false positive rates.
  + Recall: Measures the ratio of correctly predicted positive cases to total actual positive cases. The recall scores of 0.98, 0.99, and 0.99 indicate that the model successfully captures utmost true positive cases, ensuring that few positive cases are missed.
* **F1-Score:** The F1-score is the harmonious mean of precision and recall, giving a balanced measure of the model’s accuracy. The model achieved F1-scores of 0.99, 0.99, and 0.97 across different classes, showing that it maintains a good balance between precision and recall.
* **Confusion Matrix Analysis:** The confusion matrix illustrates the model's classification performance across different ship emission levels and berth categories. High true positive rates and low misclassification rates confirm the model’s ability to differentiate between different ship and berth classes accurately. The normalized confusion matrix provides insight into class-wise accuracy, helping to identify potential misclassification patterns for further optimization.

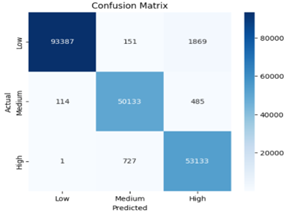


Figure 2: Confusion Matrix of the Random Forest Model for Emission Classification.

As shown in figure 2, The confusion matrix illustrates the model's accuracy in classifying ship emissions into High, Moderate, and Low categories. Most predictions align with the diagonal, reflecting high classification accuracy and minimal misclassification.

**4.3 Comparison with Existing Models:**

To evaluate the efficiency of the Random Forest model, we conducted experiments using a proprietary dataset collected from maritime operations. The dataset consists of 100,000 samples with 9 features, which was split into 80% for training and 20% for testing. We applied Random Forest, Logistic Regression, and Gradient Boosting to predict the emission levels and fuel usage patterns. The outcomes of this comparison are outlined in Table 1.

**Table 1:** Comparison of Machine Learning Models for Emission Forecasting.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Precisi-on | Recall | F1-Score | Accur-acy | Inferen-ce Time-(ms) |
| Logistic Regression | 69.2% | 70.1% | 69.2% | 70.1% | 25 |
| Gradient Boosting | 92.1% | 91.8% | 91.9% | 91.7% | 55 |
| Random Forest | 98.0% | 98.1% | 98.0% | 98.0% | 45 |

From the Table 1, Our approach utilizes the Random Forest model, which outperformed better than other logistic regression [19] and gradient boosting [20], achieving 98% accuracy with a precision of 98% and a recall of 98.1%. Its fast inference time of 45 ms makes it suitable for real-time prediction of ship emissions and berth scheduling.

**5. Conclusion:**

This study focused on the integration of machine learning and optimization techniques to enhance maritime operations through emission forecasting and berth scheduling optimization. A Random Forest-based predictive model was developed to estimate shipping emissions, enabling ports to make informed decisions that contribute to environmental sustainability. Additionally, a Mixed Integer Linear Programming (MILP) model was employed to optimize berth scheduling, reducing ship waiting times, fuel consumption, and overall emissions.

The experimental results demonstrated that the Random Forest model outperformed traditional methods such as Logistic Regression and Gradient Boosting in emission prediction, achieving an accuracy of 98% with a precision of 98% and a recall of 98.1%. The berth scheduling system, optimized

through MILP, efficiently allocated berths based on ship priority, size, and emission levels. This approach not only improved port efficiency but also contributed to minimizing environmental impact by reducing idle times and unnecessary fuel consumption. The integration of real-time data through a web application provided a dynamic decision-making framework for port authorities, enhancing operational flexibility and responsiveness.

This research demonstrates that machine learning and optimization can improve maritime logistics by making port operations more efficient and eco-friendly. By combining predictive analytics with intelligent scheduling, the system reduces delays, fuel consumption, and emissions. Future improvements could include deep learning for better predictions and real-time satellite data for more accurate berth allocations. Overall, this approach supports sustainable and data-driven maritime operations while ensuring compliance with global regulations.

**6. References:**

[1] M. Mansoursamaei, M. Moradi, G. González-Ramírez, & & E. Lalla-Ruiz. (2023), “Machine learning for promoting environmental sustainability in ports”, Artificial Intelligence Approaches for Green Transportation Planning.

[2]Mingyuan Yue, Yubing Wang, Siqing Guo, Lei Dai, & Hao Hu (2024), “A multi-objective optimization study of berth scheduling considering shore side electricity supply”, Transportation Research Part E: Logistics and Transportation Review.

[3] Lorenz Kolley, Nicolas Rückert, Marvin Kastner, Carlos Jahn, & Kathrin Fischer (2022), “Robust berth scheduling using machine learning for vessel arrival time prediction”, Flexible Services and Manufacturing Journal.

[4] Yuquan Du, Qiushuang Chen, Xiongwen Quan, Lei Long, & Richard Y.K. Fung (2011), “Berth allocation considering fuel consumption and vessel emissions”, Transportation Research Part E: Logistics and Transportation Review.

[5] T. Fletcher, Vikram Garaniya, Shuhong Chai, Rouzbeh Abbassi, Hongyang Yu, Thuy Chu Van, Richard J. Brown, & Faisal Khan (2018), “An application of machine learning to shipping emission inventory”, The International Journal of Maritime Engineering.

[6] Carlos D. Paternina-Arboleda, Dayana Agudelo-Castañeda, Stefan Voß, & Shubhendu Das, “Towards Cleaner Ports: Predictive Modeling of Sulfur Dioxide Shipping Emissions in Maritime Facilities Using Machine Learning”, Sustainability in Logistics and Supply Chain Management.

[7] Yun Peng, Huakun Liu, Xiangda Li, Jian Huang, & Wenyuan Wang (2020), "Machine learning method for energy consumption prediction of ships in port considering green ports", Energy Reports.

[8] Wei Guo, Minghao Ji, & Hai Zhu (2021), “Multi-Period Coordinated Optimization on Berth Allocation and Yard Assignment in Container Terminals Based on Truck Route”, IEEE Access.

[9] Fabregat, A., García-Martínez, A., Rizza, U., & Millán, & M. M. (2021), “Impact of port and ship traffic on urban air quality: A case study”, Atmospheric Environment.

[10] Wenxin Xie, Yong Li, Yang Yang, Peng Wang, Zhishan Wang, Zhaoxuan Li, Qiang Mei, & Yaqi Sun, (2023), “Maritime greenhouse gas emission estimation and forecasting through AIS data analytics: a case study of Tianjin port in the context of sustainable development”, Frontiers in Marine Science.

[11] Houjun Lu, & Xiao Lu (2025), “Joint Optimization of Berths and Quay Cranes Considering Carbon Emissions: A Case Study of a Container Terminal in China”, Journal of Marine Science and Engineering.

[12] Zhihui Hu, Yongxin Jing, Qinyou Hu, Sukanta Sen, Tianrui Zhou, & Mohd Tarmizi Osman (2019), "Prediction of Fuel Consumption for Enroute Ship based on Machine Learning", IEEE Access.

[13] Yinchen Lin, & Chuanxu Wang (2025), “Prediction of ship CO2 emissions and fuel consumption using Voting-BRL model”, Sustainability.

[14] Haolin Li, Jiajing Gao, Lu Zhen, & Xueting He (2024), “Berth and yard scheduling optimization for a port with a diagonal yard layout”, Flexible Services and Manufacturing Journal.

[15] Yonggai Dai, Zongchen Li, & Boyu Wang (2023), “Optimizing Berth Allocation in Maritime Transportation with Quay Crane Setup Times Using Reinforcement Learning”, ResearchGate.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, & W. P. Kegelmeyer (2002), “SMOTE: Synthetic minority over-sampling technique”, Journal of Artificial Intelligence Research

[17] H.L. Ma, S.H. Chung., H.K. Chan, & Li Cui (2017),” An integrated model for berth and yard planning in container terminals with multi-continuous berth layout”, Annals of Operations Research.

[18] Xing Jiang, Ming Zhang, Jiahui Shi, & Weifeng Li (2024), “Optimization of integrated scheduling of restricted channels, berths, and yards in bulk cargo ports considering carbon emissions”, ResearchGate.

[19] Pietukhov, R., Ahtamad, M., Faraji-Niri, M., & El-Said, T. (2023), “A hybrid forecasting model with logistic regression and neural networks for improving key performance indicators in supply chains”, Supply Chain Analytics.

[20] Mohamed M. Ahmed, & Mohamed Abdel-Aty (2013), “Application of stochastic gradient boosting technique to enhance reliability of real-time risk assessment: Use of automatic vehicle identification and remote traffic microwave sensor data.”, Transportation Research Record: Journal of the Transportation Research Board.