POAC'25
St. John's, Newfoundland and Labrador, Canada

Proceedings of the 28th International Conference on Port and Ocean Engineering under Arctic Conditions
Jul 13-17, 2025
St. John's, Newfoundland and Labrador Canada

A Unified Data Platform for Maritime Analytics and Predictive Modeling in Arctic Operations

Balsher Singh[1], Joshua Barnes[1], Allison Kennedy[1], Matthew Hamilton[2], Samarasimha Reddy Chittamuru[1]

[1] Ocean, Coastal, and River Engineering, National Research Council Canada, St. John's, Canada

[2] Dept. of Computer Science Memorial University of Newfoundland St. John's, Canada

ABSTRACT

As the shipping industry embraces digitalization, marine operations generate vast and diverse data streams, offering significant opportunities for advanced analytics. This work presents an advanced software platform powered by a high-performance Online Analytical Processing (OLAP) database, centralizing data from multiple Canadian Coast Guard (CCG) vessels, including the Larsen, Laurier, Cygnus, Tully, and Tanu. The system integrates operational data (e.g., fuel consumption, navigation) and environmental datasets (e.g., ERA5 weather, ocean currents, ice charts) while harmonizing sensor nomenclature for consistency. For example, vessel-specific labels such as "ENG_RPM" and "ENGINE_SPEED" are standardized to a unified "Engine RPM" field using a middle vocabulary layer between database and , enabling consistent cross-vessel comparisons.

The platform's visualization tool enables quality assurance (QA), exploratory data analysis (EDA), and comparative analysis across vessels using scatter plots, trajectory maps, and fuel consumption summaries. Its emissions module provides detailed hourly and trip-based emissions calculations, supporting regulatory compliance and environmental assessments. Benchmarking features include univariate and bivariate plots, time series boxplots, and heatmaps for cross-vessel comparisons under varying conditions.

A key innovation is its accessible machine learning (ML) tool, allowing non-expert users to train predictive models on custom time intervals for forecasting metrics like fuel consumption, speed, and emissions. A dedicated ML Model Explainer dashboard enhances interpretability with regression statistics, feature importance rankings, and dependency charts.

Additionally, the platform supports strategic planning through historical route mapping and predictive analytics, enabling users to test models under Arctic conditions. By integrating visualization, benchmarking, and ML-driven forecasting, this platform enhances maritime decision-making, optimizing operational efficiency and environmental sustainability across challenging marine environments.

KEY WORDS: Maritime Analytics; Machine Learning; Emissions Tracking; Data Visualization; Operational Efficiency.

INTRODUCTION

Background

The maritime industry is undergoing a significant transformation, driven by digitalization, data analytics, and the need for sustainable operations. As environmental regulations tighten and the demand for efficiency grows, fleet managers are increasingly turning to data-driven solutions to optimize performance, reduce costs, and minimize emissions. The International Maritime Organization (IMO) has set ambitious goals, including a 40% reduction in carbon intensity by 2030, making real-time monitoring and predictive analytics crucial for compliance and operational improvement.

To address these challenges, this work presents the development of a comprehensive maritime analytics platform that integrates high-performance data processing, visualization, emissions tracking, and machine learning. This platform centralizes data from multiple Canadian Coast Guard vessels, enabling data-driven decision-making through enhanced benchmarking and predictive modeling.

While previous work has explored isolated aspects of marine data analytics (Piercey et al., 2024), this work presents unified systems with a suite of analytics capability together with integrated machine learning through a simplified interface, all driven by an optimized database implementation.

We present such a system, discussing its design and demonstrating its useful capabilities in the context of a user case designed to support environmental sustainability analysis, emissions tracking, operational efficiency and cost reduction.

Motivation

The environmental and operational challenges faced by the maritime industry necessitate the collection of vessel performance data and the application of advanced data analytics solutions. There are two primary motivations behind this work:

1. **Environmental Sustainability and Emissions Tracking-**

The shipping industry is a significant contributor to global greenhouse gas (GHG) emissions. The IMO's decarbonization targets require ship operators to closely monitor fuel consumption and emissions, ensuring compliance with evolving regulations. Traditional manual methods for emissions tracking are inefficient and prone to errors. This platform automates emissions calculations, offering granular insights into pollutants such as NOx, CO, and CO_2, helping fleet managers track and mitigate environmental impact.

2. **Operational Efficiency and Cost Reduction-**

In addition to sustainability, optimizing operational performance is a key industry priority. Fuel is one of the largest expenses in maritime operations, and inefficiencies can lead to excessive consumption. By integrating advanced visualization tools and benchmarking capabilities, the platform enables fleet managers to compare vessel performance, identify inefficiencies, and optimize routes for better fuel economy.

Key Design Objectives

This work aims to develop a high-performance, data-driven maritime analytics platform that supports predictive modeling, performance baselining, and emissions tracking, while providing the framework necessary for the implementation of real-time predictions. The key design objectives include:

- **Enabling real-time analytics from a wide variety of data sources related to operational efficiency of vessels**- Complex analytics requires a user to analyze multiple sources of data in the context of complex operational requirements. This means accessing multiple datasets simultaneously without significant delay to support the user's analysis process in a timely manner.
- **Developing Advanced Visualization and Benchmarking Tools**-
  The platform provides comprehensive visualization capabilities, enabling detailed maritime analytics through various plots and statistics. These tools empower fleet managers with actionable insights, supporting data-driven decision-making and operational optimization.
- **Implementing an Automated Emissions Tracking Module**-
  The emissions tracking system calculates pollutant levels based on fuel consumption and operational modes, using emission factors indicated by the Fourth IMO GHG Study (IMO, 2020). This module has the capability to provides accurate, real-time emissions data, eliminating the need for manual calculations and supporting regulatory compliance.
- **Integrating Machine Learning for Predictive Analytics**-
  A built-in machine learning (ML) module allows non-expert users to train models for forecasting key metrics such as fuel consumption, power, emissions, and other performance metrics. The ML Model Explainer Dashboard enhances transparency, providing: feature importance rankings, dependency charts & regression statistics. The explainer dashboard ensures that even users without data science expertise can interpret model results and make informed adjustments.

SYSTEM DESIGN AND IMPLEMENTATION

Implementing ClickHouse database for High-Performance Analytics

ClickHouse is an open-source columnar database management system designed for high-performance analytical workloads. Unlike traditional row-oriented databases, ClickHouse stores data in columns, making it particularly efficient for read-heavy operations and aggregations, ideal for time-series data that underpins maritime operational analytics. Key features of ClickHouse that enhance maritime data processing include columnar storage, partitioning and sharding and materialized views. The primary data channels stored include vessel location (latitude, longitude), speed, heading, fuel flow, and engine parameters (RPM, torque, power). Data is stored at 1Hz frequency for ship system. On average, the platform ingests approximately 2.6 million records per vessel per month. The migration from PostgreSQL to ClickHouse significantly improved the platform's efficiency. Aggregating data in the aggregate_df.csv process which is used to plot monthly and annual emission trends and other baselining trends, previously took two days to complete and now can be completed in 90 minutes, marking a 96% reduction in processing time. The comparison was conducted using a Linux server equipped with 64 GB RAM and a 6-core Intel Xeon W-2133 processor. No distributed cloud deployment was required for current operational volumes. The development of this processing framework enables real-time analytics, faster visualizations, and improved decision-making across multiple vessels.

The Need for an Intuitive Visualization Tool

As maritime operations generate vast datasets, fleet managers need a clear and interactive way to explore ship performance, fuel usage, and emissions trends. A raw dataset, no matter how detailed, remains difficult to interpret without proper visualization. To bridge this gap, our platform provides dynamic, interactive, and customizable visualizations, offering insights into fleet behavior in real time. The visualization interface is primarily developed using HoloViews and Plotly, integrated into a custom frontend and backend built with Panel, which leverages Bokeh as its rendering engine. Data queries and analytics for the plots are handled through an API layer that applies a standardized vocabulary, performs data cleaning, and interfaces with the database to retrieve relevant data.

Integrating environmental datasets such as ERA5 weather reanalysis, Copernicus Ocean currents, and Canadian ice charts presented challenges due to varied access mechanisms, including cloud storage, FTP servers, and public APIs. These datasets were pre-collected and stored in a centralized big-data storage array, from which the platform automatically synchronizes relevant files. While no formal license-checking system was implemented, all sources used were publicly available and appropriately cited. ERA5 data were accessed from the Copernicus Climate Data Store [ECMWF, 2023], ocean current data were sourced from the Copernicus Marine Service's Global Ocean Physics Reanalysis model [CMEMS, 2023], and Canadian sea ice charts were obtained from the Canadian Ice Service under Environment and Climate Change Canada [CIS, 2023].

The system allows users to:

- Select time ranges and ships to analyze specific operational periods.
- Explore ship movement on maps to study past voyages and patterns.
- Compare ship performance across multiple parameters, from fuel efficiency to emissions.
- Train and explain machine learning models, making predictive analytics accessible to all users.

With these functionalities, our tool transforms raw maritime data into actionable insights.

Foundational UI Elements

These components streamline data access, reliability, and usability, ensuring a smooth and efficient visualization experience.

- Date-Time Range Picker - Allows users to select a specific time range for analysis, with auto-restricted limits based on available data in the database, ensuring valid selections.

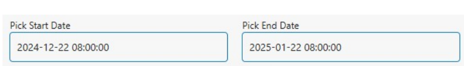


Figure 1. Datetime Range Picker

- Data Availability Info Pop-up - Displays availability of data for each vessel over different periods on click of button, helping users quickly identify accessible data before performing analyses. This is auto populated from the database.

Figure 2. Data Availability Pop-up

- Connection/Disconnection Alerts - Banners notify users of real-time connection status, alerting them to server disconnections or data retrieval failures, ensuring uninterrupted workflow.

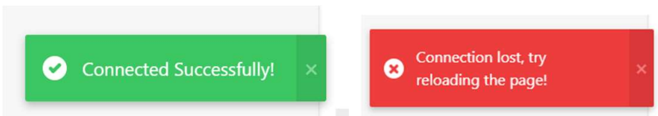


Figure 3. Connection Alerts

- Data Unavailability Pop-up - If a user selects a time range or dataset with no available data, a pop-up notification informs them immediately, preventing confusion and improving user experience.

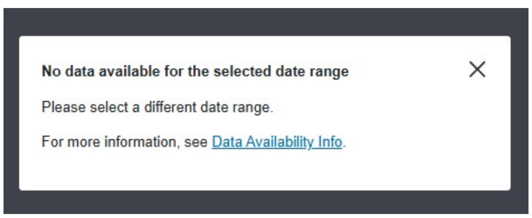


Figure 4. No Data Pop-up

Mapping Ship Trajectories: Visualizing Movement Over Time

The map plot feature allows users to visualize ship trajectories, offering insights into:

- **Route optimization**: Identify deviations, inefficiencies, and patterns in ship movement.
- **Fuel efficiency by route**: Compare fuel consumption across different paths.
- **Environmental factors**: Overlaying ice charts, wind conditions, and wave heights to understand navigational challenges.

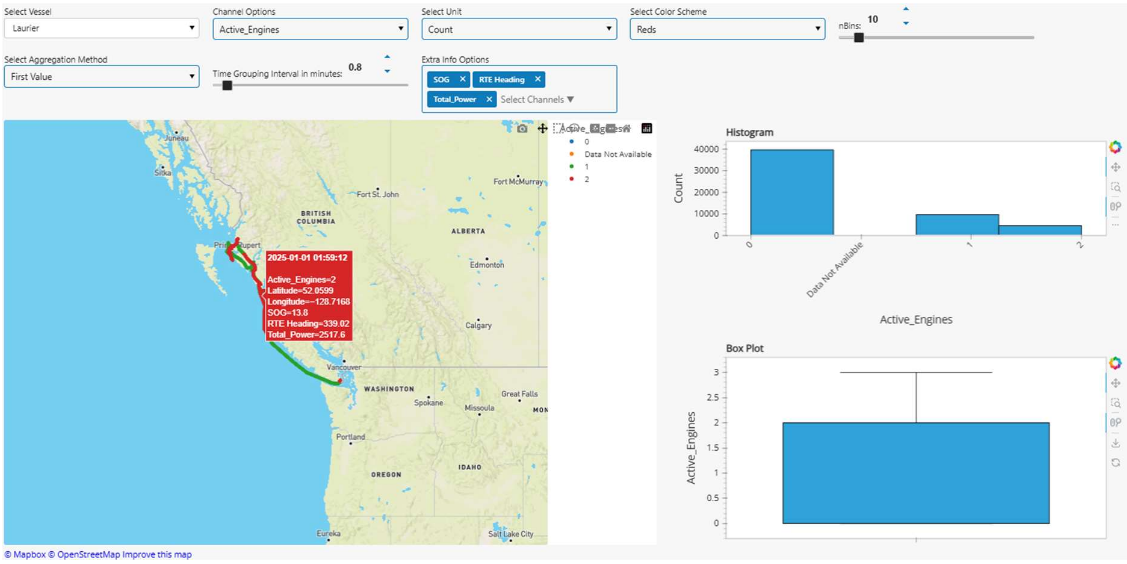


Figure 5. GPS plot for Laurier ship showing Active Engines

Scatter and Time-Series Plots for Performance Insights

- Scatter plots reveal correlations (e.g., fuel consumption vs. speed, engine power vs. emissions).
- Multi-ship scatter plots allow fleet-wide performance comparisons, by enabling a user to plot scatter plots relating to two or more vessels on the same graph.

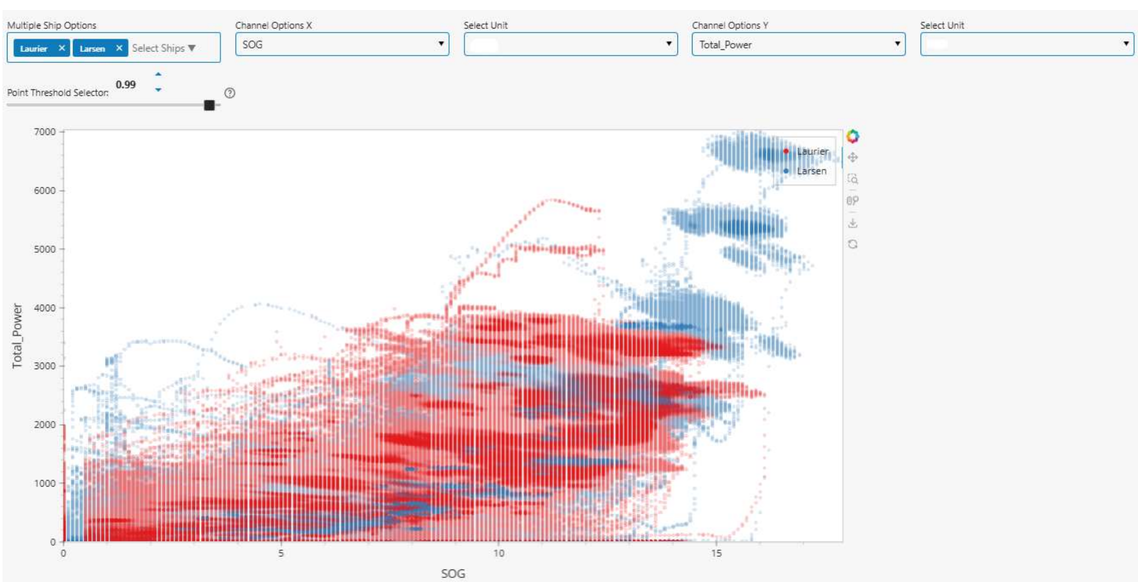


Figure 6. Performance Scatter plot for two ships

- Time-series plots track trends over time, detecting anomalies and seasonal variations. The time-series plot option allows the user to plot any parameter versus date and time, for the period of interest, e.g., for a single season or for a single week.

Benchmarking and Baselining Tools

Establishing performance baselines is key to operational efficiency, thus is a key user task that our system must support. In the context of this report performance baselining means the development of plots that demonstrate the vessel requirements (e.g. power, fuel, emissions) relating to operations at different speeds and in different environmental conditions. Our tool includes:

- **Boxplots**: Identify distribution, outliers, and variability in power, fuel consumption and emissions.
- **PDF & CDF**: Define "normal" operating conditions and detect threshold exceedances.
- **Polar Plots**: Analyze fuel use and emissions relative to wind, waves, and heading.
- **Time-Series Boxplots & Heatmaps**: Summarize variations across months, seasons, or routes.

These benchmarking tools **help fleet managers optimize vessel performance** through comparative analysis.

Fuel Consumption and Emissions Tracking

A tabular summary provides engine-specific fuel usage, breaking down emissions by operational mode (cruising, maneuvering, hotelling). The emissions module:

- Compares emissions across voyages to track sustainability improvements.
- Provides information for comparing against environmental regulations.
- Calculates pollutants (CO_2, NO_x, PM10, NMVOCs) in real time.

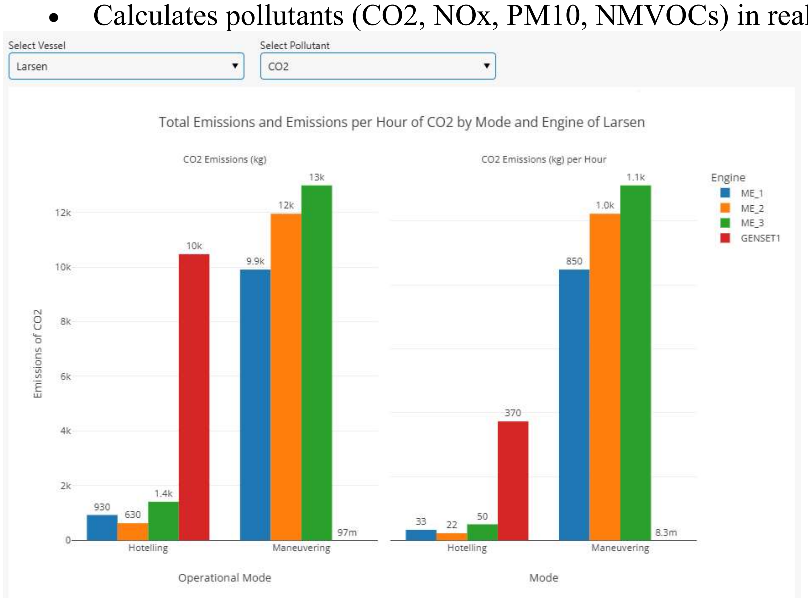


Figure 7. Total CO2 Emissions (Kg) and CO2 Emissions per hour (kg) for a given vessel

This feature supports the development of data-driven fuel efficiency and emissions reduction strategies. For example, by quantifying the amount of emissions from a given vessel while hoteling (primarily at dock), operators could use this information to help build justification cases to support the development of shore power options when possible.

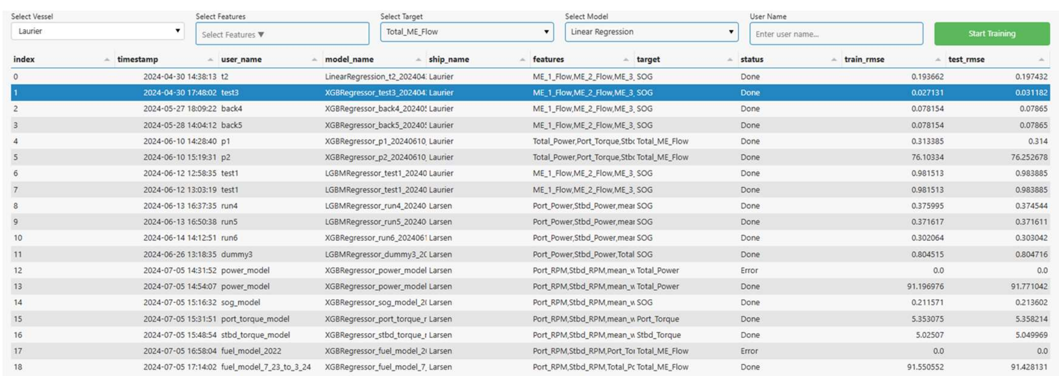
MACHINE LEARNING CAPABILITIES FOR PREDICTIVE ANALYTICS

At the core of the platform's innovation is its integrated machine learning (ML) module, enabling users—regardless of technical expertise—to leverage predictive analytics for maritime operations. This module provides a structured, interactive workflow for selecting input features, defining target variables, and training models on custom time intervals where the user can select a time range from date pickers and choose data only within that interval as training data. By incorporating fuel consumption forecasting, emissions prediction, speed over ground (SOG) estimation, and engine performance monitoring, the ML feature supports data-driven decision-making for optimizing vessel efficiency and sustainability. Users can predict fuel consumption for past transits and compare predictions to actual values in ML model Error tab, or estimate fuel consumption for future transits based on assumed parameters in ML Explainer tab. This capability aids in fuel cost estimation and voyage planning, providing valuable insights for operational efficiency.

Customizable ML Workflow

The machine learning module is designed to be user-friendly and adaptable to various analytical needs:

- Users can select input features (e.g., engine RPM, torque, weather conditions) and specify a target variable (e.g., fuel flow, emissions, speed over ground).
- The intuitive interface allows defining time intervals for model training and selecting algorithms from a list of supported regression and classification models.
- Performance metrics such as Root Mean Square Error (RMSE) for regression models and classification accuracy for categorization tasks provide immediate feedback on model effectiveness.



index	timestamp	user_name	model_name	ship_name	features	target	status	train_rmse	test_rmse
0	2024-04-30 14:38:13	t2	LinearRegression_t2_202404	Laurier	ME_1_Flow,ME_2_Flow,ME_3	SOG	Done	0.193662	0.197432
1	2024-04-30 17:48:02	test3	XGBRegressor_test3_202404	Laurier	ME_1_Flow,ME_2_Flow,ME_3	SOG	Done	0.027131	0.031182
2	2024-05-27 18:09:22	back4	XGBRegressor_back4_20240	Laurier	ME_1_Flow,ME_2_Flow,ME_3	SOG	Done	0.078154	0.07865
3	2024-05-28 14:04:12	back5	XGBRegressor_back5_20240	Laurier	ME_1_Flow,ME_2_Flow,ME_3	SOG	Done	0.078154	0.07865
4	2024-06-10 14:28:40	p1	XGBRegressor_p1_20240610	Laurier	Total_Power,Port_Torque,Stb	Total_ME_Flow	Done	0.313385	0.314
5	2024-06-10 15:19:31	p2	XGBRegressor_p2_20240610	Laurier	Total_Power,Port_Torque,Stb	Total_ME_Flow	Done	76.10334	76.252678
6	2024-06-12 12:58:35	test1	LGBMRegressor_test1_20240	Laurier	ME_1_Flow,ME_2_Flow,ME_3	SOG	Done	0.981513	0.983885
7	2024-06-12 13:03:19	test1	LGBMRegressor_test1_20240	Laurier	ME_1_Flow,ME_2_Flow,ME_3	SOG	Done	0.981513	0.983885
8	2024-06-13 16:37:35	run4	LGBMRegressor_run4_20240	Larsen	Port_Power,Stbd_Power,mear	SOG	Done	0.375995	0.374544
9	2024-06-13 16:50:38	run5	LGBMRegressor_run5_20240	Larsen	Port_Power,Stbd_Power,mear	SOG	Done	0.371617	0.371611
10	2024-06-14 14:12:51	run6	XGBRegressor_run6_2024061	Larsen	Port_Power,Stbd_Power,mear	SOG	Done	0.302064	0.303042
11	2024-06-26 13:18:35	dummy3	LGBMRegressor_dummy3_2(	Larsen	Port_Power,Stbd_Power,Total	SOG	Done	0.804515	0.804716
12	2024-07-05 14:31:52	power_model	XGBRegressor_power_model	Larsen	Port_RPM,Stbd_RPM,mean_w	Total_Power	Error	0.0	0.0
13	2024-07-05 14:54:07	power_model	XGBRegressor_power_model	Larsen	Port_RPM,Stbd_RPM,mean_w	Total_Power	Done	91.196976	91.771042
14	2024-07-05 15:16:32	sog_model	XGBRegressor_sog_model_2(	Larsen	Port_RPM,Stbd_RPM,mean_w	SOG	Done	0.211571	0.213602
15	2024-07-05 15:31:51	port_torque_model	XGBRegressor_port_torque_r	Larsen	Port_RPM,Stbd_RPM,mean_w	Port_Torque	Done	5.353075	5.358214
16	2024-07-05 15:48:54	stbd_torque_model	XGBRegressor_stbd_torque_r	Larsen	Port_RPM,Stbd_RPM,mean_w	Stbd_Torque	Done	5.02507	5.049969
17	2024-07-05 16:58:04	fuel_model_2022	XGBRegressor_fuel_model_2(	Larsen	Port_RPM,Stbd_RPM,Port_Tor	Total_ME_Flow	Error	0.0	0.0
18	2024-07-05 17:14:02	fuel_model_7_23_to_3_24	XGBRegressor_fuel_model_7	Larsen	Port_RPM,Stbd_RPM,Total_Pc	Total_ME_Flow	Done	91.550552	91.428131

Figure 8. Machine learning Trainer Tab

This workflow ensures that users without prior machine learning experience can build and deploy predictive models with minimal effort while maintaining analytical rigor.

ML Model Explainer Dashboard

To enhance model interpretability, the platform features an ML Model Explainer Dashboard, providing comprehensive insights into model performance and prediction reliability.

- Regression Statistics - The dashboard includes summary metrics such as RMSE, R-squared, and Mean Absolute Error to evaluate the accuracy of regression models.

- Feature Importance Rankings - A visual ranking of features highlights the most influential variables driving the model's predictions, helping users understand which factors contribute most to operational outcomes.
- Dependency Charts - Dependency plots illustrate how changes in specific input variables impact the target variable, enabling scenario-based analysis for decision-making.

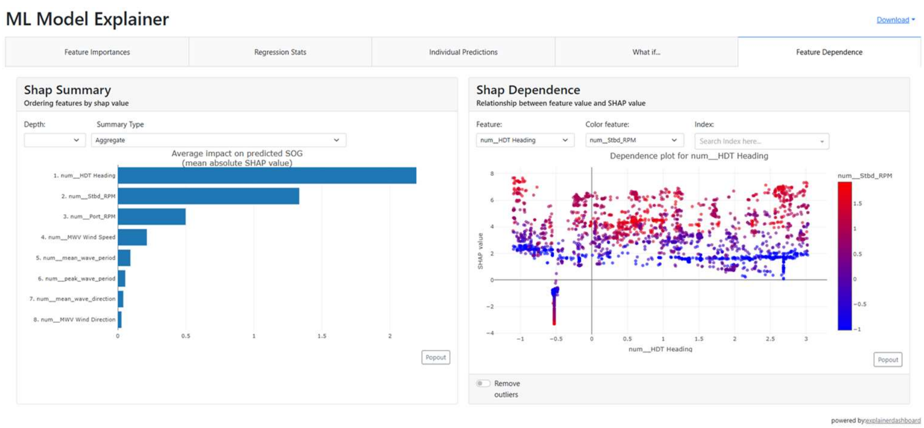


Figure 9. Machine learning Model Explainer dashboard

To enhance transparency and foster user trust in model outputs, the platform includes interactive tools that expose the inner workings of predictive models. Summary regression metrics such as Root Mean Square Error (RMSE), R-squared, and Mean Absolute Error offer users a high-level assessment of model accuracy. More importantly, feature importance rankings visually highlight which variables most strongly influence model predictions enabling users to interpret results and validate whether key operational drivers are being captured.

Complementing this, dependency charts illustrate how changes in individual input variables affect predicted outcomes, supporting "what-if" style analyses for decision-making. These capabilities align with emerging best practices in explainable artificial intelligence (XAI) , as explored by Molnar (2022) and further demonstrated in widely adopted methods such as SHAP by Lundberg and Lee (2017), particularly in high-stakes domains such as maritime operations, where model transparency is essential. By surfacing interpretable relationships and providing accessible visual cues, the system promotes broader adoption of machine learning among operational teams.

By providing transparency into model logic, these tools empower users to refine models, identify key operational factors, and enhance decision-making accuracy.

PREDICTIVE MODELING IN ACTION: SPEED OVER GROUND (SOG) FORECASTING

To demonstrate the practical application of the platform's machine learning capabilities, a Speed Over Ground (SOG) prediction model was developed using historical ship data. This model helps fleet managers anticipate vessel performance under varying operational and environmental conditions, optimizing route planning and fuel efficiency. Figure 10 presents the error between the predicted and actual SOG along a historical vessel route, visualized on a map. This figure is the ML Model Error Tab, which provides users with a detailed assessment of model accuracy, highlighting areas where predictions deviate from real-world measurements. User can select any model trained in ML Trainer tab and plot its error on a historical route on a map. Such insights enable iterative improvements to model performance and support better-informed decision-making for maritime operations.

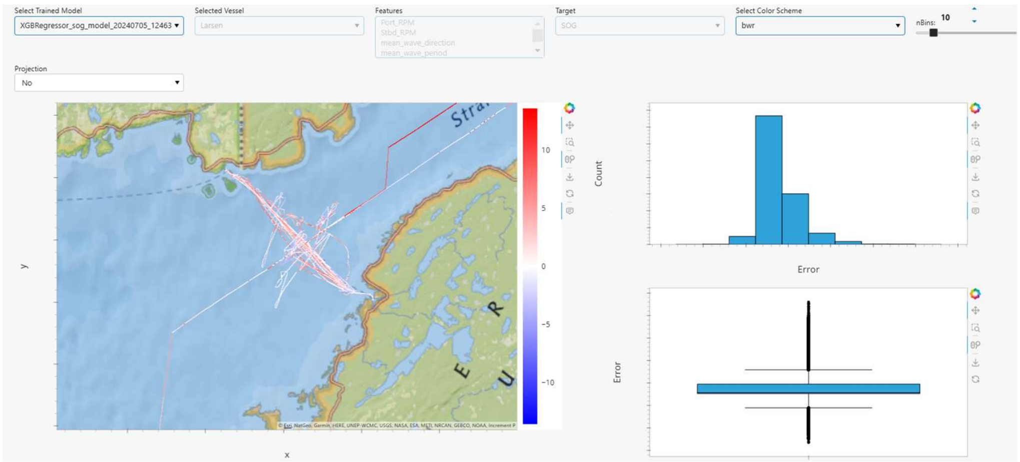


Figure 10. Error of Predicted vs. Actual Speed Over Ground (SOG) along a Historical Vessel Route

Model Development and Key Features

The SOG prediction model was trained using:

- Engine Parameters: RPM, power output, and fuel consumption.
- Weather Conditions: Wind resistance, wave height, and ice coverage.
- Historical Speed Data: Past SOG values across different voyages and environments.

By incorporating these inputs, the model can predict future vessel speed under different operational scenarios.

CONCLUSIONS

The work performed in this study allowed for the development of a data analytics framework and platform that is tailored to process, prepare and evaluate key data collected within the CCG fleet. The tool enables the CCG to visualize and calculate key features and metrics that will inform their decision making relating to optimized fuel consumption and performance. The development of this tool demonstrates further the value of collecting vessel performance data, towards providing evidence-based information to guide future operations.

Although the system was developed with Arctic maritime operations in mind, its modular design enables easy extension to fleets operating globally. Data ingestion workflows can be adapted for other sources through customizable pipelines, and new algorithms can be incorporated easily within the code.

The platform enables real-time analytics by integrating multiple data sources, allowing users to analyze vessel performance efficiently. Advanced visualization and benchmarking tools provide actionable insights through intuitive plots and statistical comparisons, supporting informed decision-making.

The automated emissions tracking module accurately calculates pollutant levels based on fuel consumption, aiding regulatory compliance and sustainability efforts. Additionally, the machine learning module allows users to forecast fuel consumption, emissions, and vessel performance, with the ML Model Explainer Dashboard ensuring transparency and accessibility.

By bridging maritime operations with data science, this platform enhances efficiency, safety, and sustainability, supporting the industry's transition toward greener and more efficient fleet management.

ACKNOWLEDGEMENTS

We extend our sincere gratitude to the Canadian Coast Guard for their invaluable support and collaboration in this project. Their provision of operational data, domain expertise, and insights into maritime challenges have been instrumental in developing this data analytics and visualization platform. Their commitment to advancing data-driven decision-making in maritime operations has significantly contributed to the success of this work.

We appreciate their efforts in facilitating access to vessel data and their guidance in understanding key operational and environmental factors. This project would not have been possible without their continued support and dedication to innovation in maritime analytics and sustainability.

REFERENCES

C. Piercey, M. Hamilton, B. Veitch, J. Barnes & X. Jiang, 2024. Ship Data Clustering for Improved Fuel Efficiency Optimization Decision Support Systems. *OCEANS 2024 - Halifax*, pp. 1-6.

S.R. Chittamuru, M. Hamilton, A. Akinturk, A. Kennedy, J. Barnes & B. Singh, 2024. Machine learning-based power prediction for the icebreaker Henry Larsen using vessel motions and environmental data in open water conditions. *Proceedings of the 33rd Annual Newfoundland Electrical and Computer Engineering Conference - St. John's*

Agand, P., Kennedy, A., Harris, T., Bae, C., Chen, M., & Park, E.J., 2023. Fuel consumption prediction for a passenger ferry using machine learning and in-service data: A comparative study. *Ocean Engineering*, **284**, p.115271.

Murrant, K., Kennedy, A., Pallard, R., & Montrose, M., 2019. Effects of hull and propeller cleaning on propulsion efficiency of an offshore patrol vessel. In: *Proceedings of the 4th Hull Performance & Insight Conference*, 8 May 2019, pp. 272-291.

International Maritime Organization (IMO), 2020. *Fourth IMO Greenhouse Gas Study 2020.* London: IMO

Molnar, C., 2022. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Leanpub.

Lundberg, S.M. & Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30.

Dijk, O., O. Sam, R. Bell, L. Simon-Free, B. Serna, R. Gupt, et al., 2023. ExplainerDashboard 0.4.2: Dtreeviz v2 Compatibility. Zenodo. https://doi.org/10.5281/zenodo.7633294.

ECMWF, 2023. ERA5 Reanalysis. Copernicus Climate Data Store. Available at: https://cds.climate.copernicus.eu/

CMEMS, 2023. Global Ocean Physics Reanalysis. Copernicus Marine Service. Available at: https://marine.copernicus.eu/

CIS, 2023. Canadian Ice Service. Environment and Climate Change Canada. Available at: https://ice-glaces.ec.gc.ca/