Augmenting Organizational Efficiency by Predicting Shipment Arrival Using Statistics and Machine Learning

Chethan Manjunath, Keerthana Nemili, Soham Patil, Sesha Sreeja Sadhu, Gauri Vaidya, Matthew A. Lanham

Purdue University, Krannert School of Management

cmanjuna@purdue.edu; knemili@purdue.edu; patil124@purdue.edu; sadhu0@purdue.edu; vaidyag@purdue.edu; lanhamm@purdue.edu

PURDUE UNIVERSITY® | Mitchell E. Daniels, Jr. School of Business

ABSTRACT

In today's global economy, organizations commonly procure goods and materials from multiple vendors across the globe through sea-going vessels. Although carriers in charge of shipments provide their own estimated time of arrivals (ETA), they are often inaccurate and provide an ETA with larger variance. Moreover, multiple ETAs from multiple carriers degrades the decision-making of organizations causing disruptions to subsequent activities in the chain. In partnership with an award winning container shipment tracking aggregator, we designed and tested several approaches to effectively use these noisy carrier predictions in an ensembled fashion to generate better vessel ETA predictions. Among several experiments we found using machine learning predictions in conjunction with a custom linear programming solution to assign optimal weights led to more accurate and reliable ETA predictions. The impact of this work will not only improve the value the partner provides their clients, but will help their clients plan better which will positively impact P&L and the planet.

BUSINESS PROBLEM

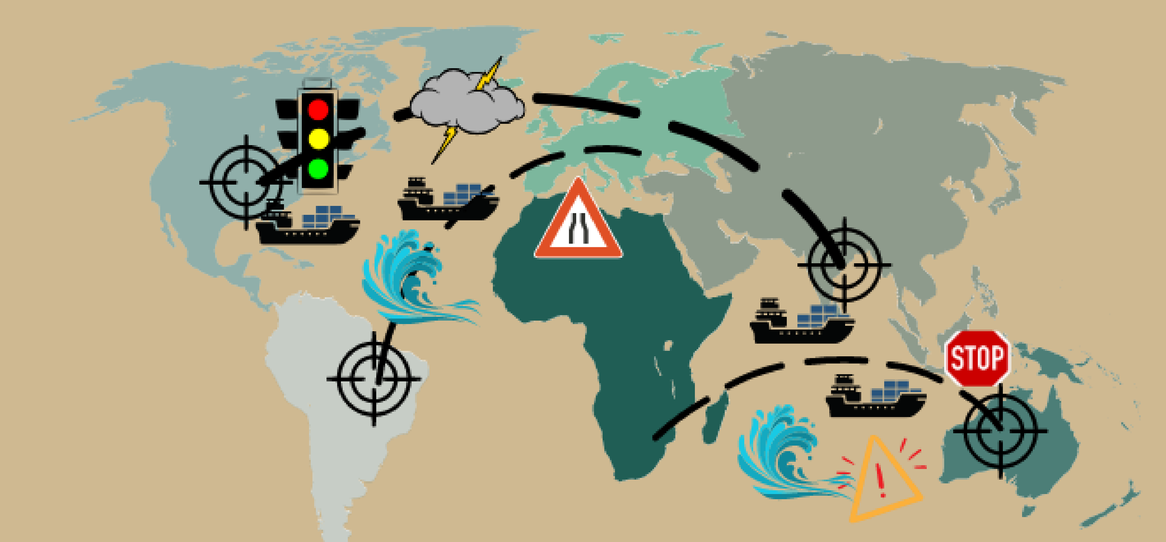

Fig 1. Supply chain complexity

Supply chain is a kind of industry where a minor hiccup at any stage can trigger a domino effect creating disruptions throughout the chain. At the head of this issue lies the understanding and obtaining an accurate arrival times of shipments, so that subsequent activities in the supply chain can be planned and executed. The main objective of our research is to improve the accuracy of shipment arrival forecasts, which will enhance consumer confidence. This improvement in projected estimated time of arrival (ETA) will enable businesses to prepare and deliver products on time, optimize resource allocation, reduce expenses, increase profitability, and contribute to a greener environment by reducing the carbon footprint of the supply chain.

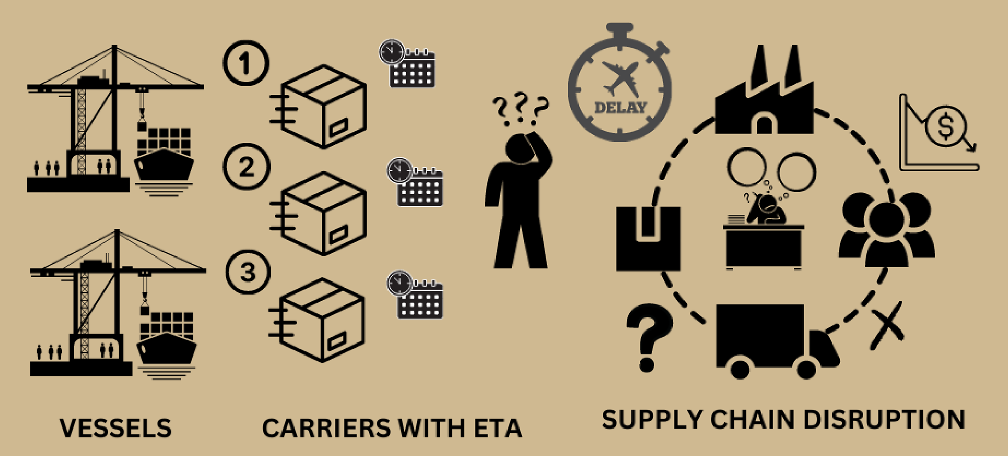

Fig 2. Carrier ETA prediction and possible supply chain disruption

RESEARCH QUESTIONS

- What is the best way to aggregate many carrier predictions?
- Are there certain carrier predictions that we can easily rule out?
- Are there regional dependencies?
- What is the best forecast or prediction for when the containers will arrive at the port?

ANALYTICS PROBLEM

Multiple carriers can provide multiple predictions for a given vessel – port combination (Fig 3). At a given point in time which carrier should the customers rely on? To tackle this, we split the prediction into time buckets (T-1,T-2, T-10 etc.) and in each time bucket, error in number of days and accuracy within 24hrs will be measured. Based on this, the idea is to assign weights to carriers and rank them according to the success metrics. Further we use this information in regression-based machine learning models and linear programming to generate optimized predictions.

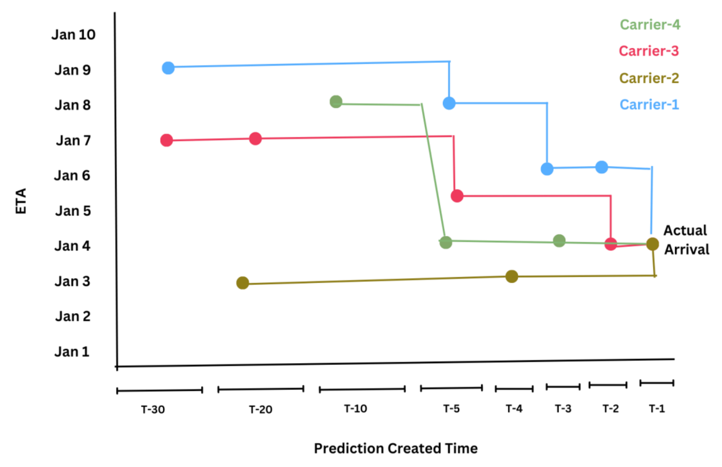

Fig 3. Carrier prediction at various point in time

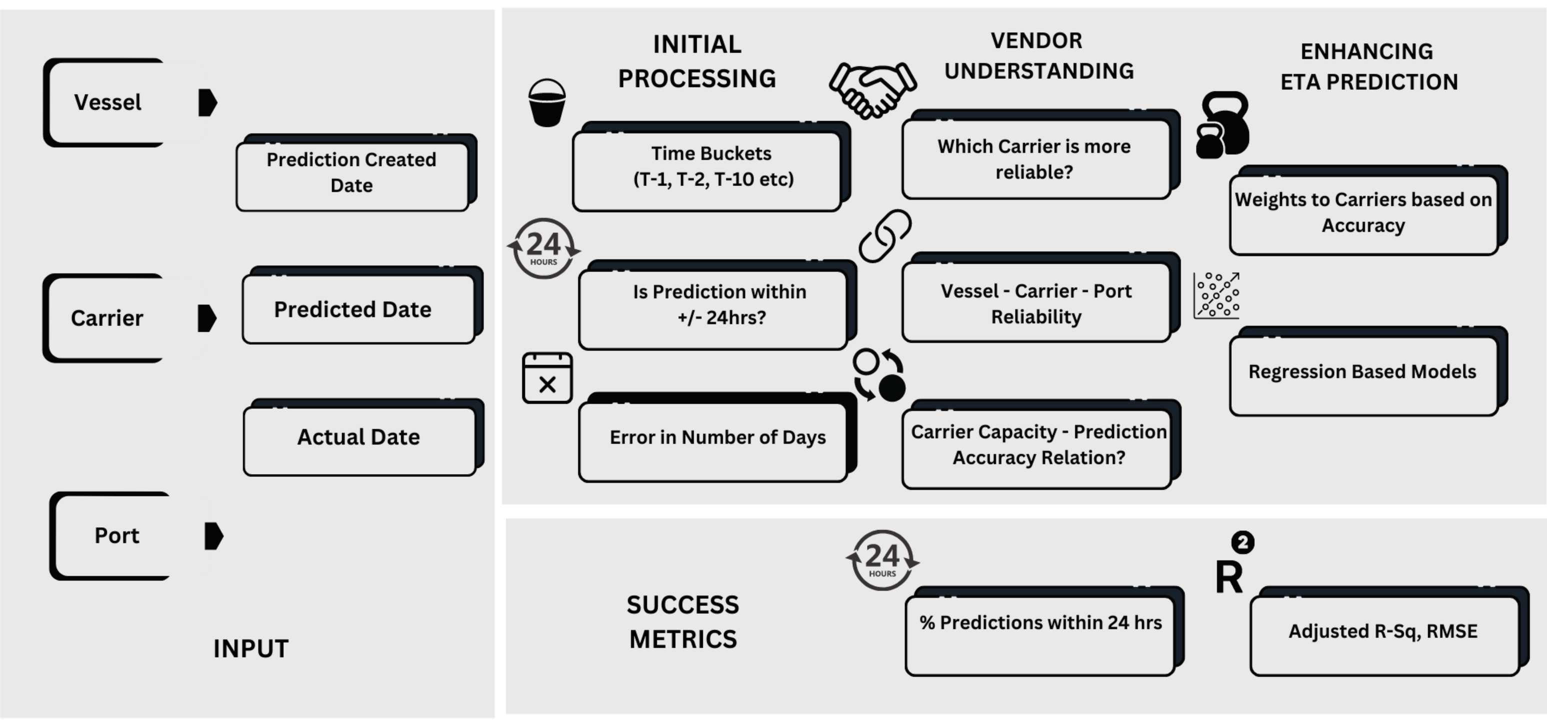

Fig 4. Stages of analytics problem framing

DATA

When a vessel arrives at a port, it is recorded as PortCall ID. One port can have multiple vessels and each vessel can have multiple carriers in it. A vessel will be a fixed capacity. Carries provide an ETA of the vessel to their customers at different point in time. The ETA can change as the vessel moves closer to the arrival and tends to become more accurate. For training the model and to identify patterns, actual arrival time for each of the predicted instance is given. The source of the data is from our collaborator's aggregated platform. Data from multiple vessel – carrier instances, gives a chance to improve the prediction and give a one stop solution for customers to track.

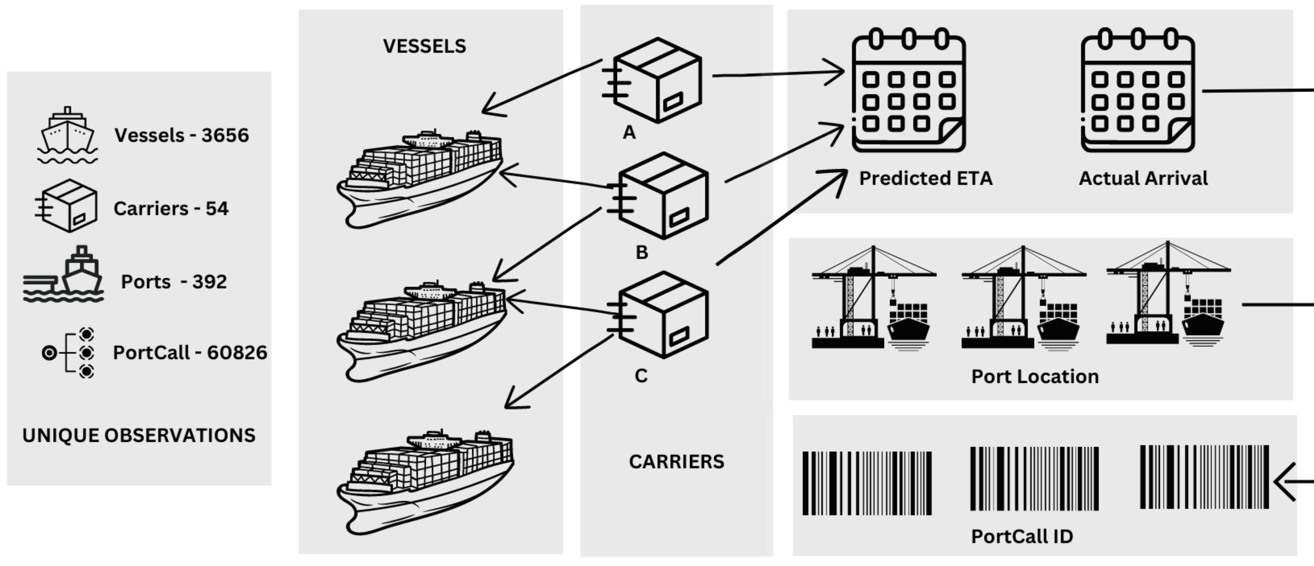

Fig 5. Data relationship

METHODOLOGY

As a first step, we explored the data to establish clear relationships and answer questions that were created initially. We explored multiple methods like liner programming, aggregate average, weighted moving average to assign weights to rank carriers. In parallel, insights and patterns related to carrier, vessel routes were created. Using these attributes, regression based models were trained. By analyzing results we identified scope for improvements and tweaked parameters to achieve better results.

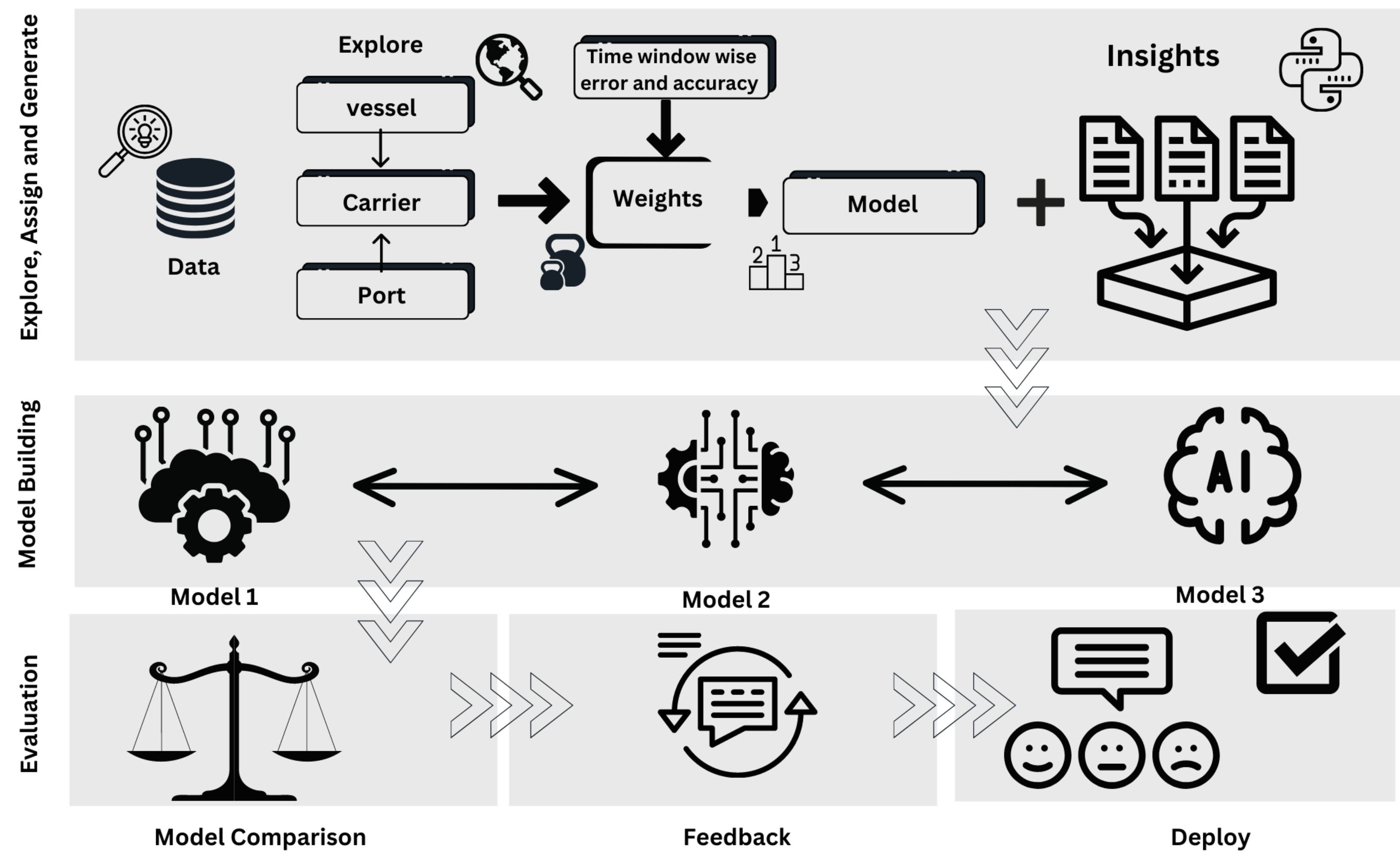

Fig 6. Methodology diagram

MODEL BUILDING AND RESULTS

Data analysis outcomes: The error distribution was highly left skewed with skewness of –1.2 therefore, median error was chosen as the input to the model rather than mean error. Another observation was the accuracy of predictions improves as the ETA approaches therefore, time window along with the carrier is considered as other input.

Final Model:

- **Inputs :** Carrier predictions, past actual arrival dates, carrier accuracies, median errors, days between carrier prediction and the prediction created on date.
- **Model:**

  New Days = (%accuracy * days) + (1 - %accuracy) * (Days + median error)

  New ETA = Pred_created_on + New Days
- Current carrier prediction accuracy is 46%. The proposed model gave us an improvement of 3%. Even though the percentage improvement is only by a small margin, the act of aggregation gives confidence i.e., improved accuracy and a single date for organizations to work with and plan subsequent activities.

Results

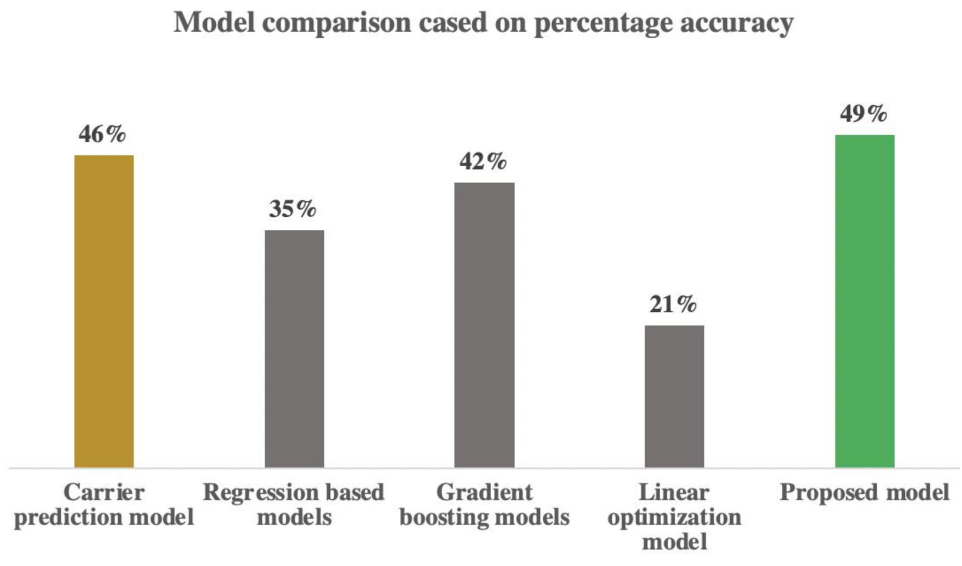

Fig 7. Model comparison of various methods used

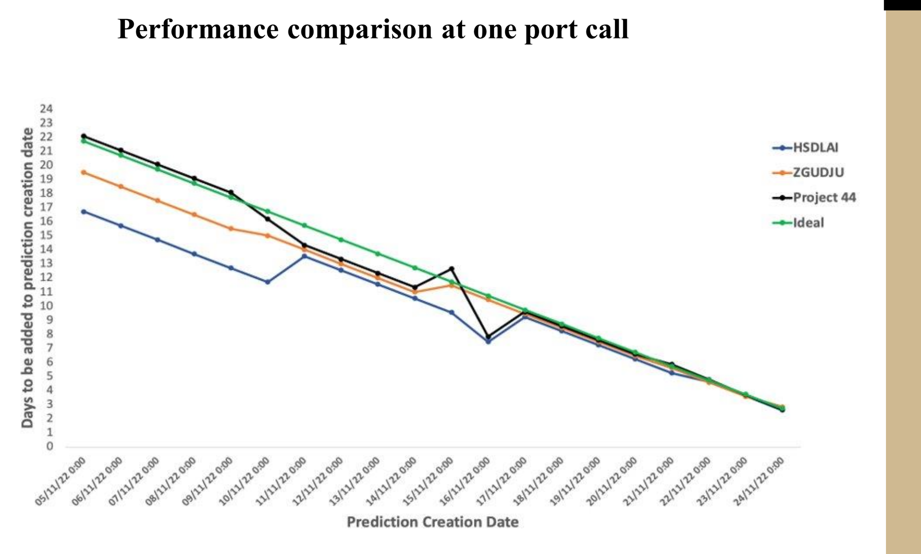

Fig 8. Model performance comparison of various carrier vs proposed model for one of the portcall.

DEPLOYMENT & LIFE CYCLE MANAGEMENT

- Documentation – A standard operating procedure document is created detailing steps to input carrier predictions, weights and error factors to get the vessel arrival prediction.
- Code – The excel and R-files shared with ready, fully functional code to generate required parameters.
- Integration with client system – using the Data Transfer Tools (DTL) the solution is deployed making it user-friendly for the client to get the predictions via just one click.

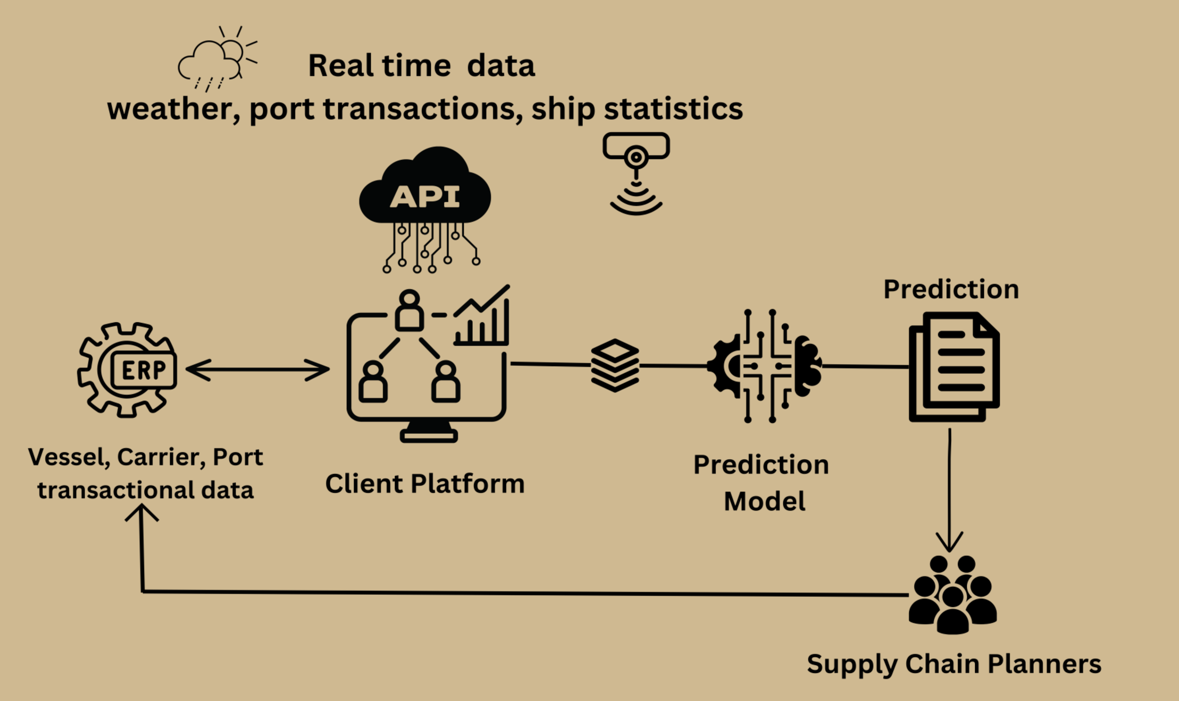

Fig 9. Deployment methodology and user workflow

CONCLUSION

Operational Efficiency

Increased operational efficiency and profits as better ETA predictions enable more effective planning, reduced idle times, and improved resource utilization, leading to cost savings and increased revenues.

Customer Satisfaction

High levels of customer satisfaction by ensuring timely delivery of items based on improved ETA's, which enhances the organization's reputation and customer loyalty.

Greener Earth

Positively contributes to a greener earth by reducing carbon footprints by approximately 30% for a major organization, as it facilitates better planning of transportation and reduces unnecessary emissions.

ACKNOWLEDGEMENTS

We would like to thank our industry partner for their guidance and support on this project as well as the Purdue MS BAIM program for partially funding this work.

informs BUSINESS ANALYTICS CONFERENCE