HEURISTIC APPROACH TO IMPROVE EXPECTED TIME OF ARRIVAL PREDICTION FOR GLOBAL INTERMODAL TRANSPORT NETWORKS

by

Arun Singh, B. Tech., GMBA

DISSERTATION

Presented to the Swiss School of Business and Management Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

JUNE, 2024

HEURISTIC APPROACH TO IMPROVE EXPECTED TIME OF ARRIVAL PREDICTION FOR GLOBAL INTERMODAL TRANSPORT NETWORKS

by

Arun Singh

Supervised by

Dr. Anna Provodnikova

APPROVED BY

Nihar Behera.
____
Dissertation chair

RECEIVED/APPROVED BY:

____
Admissions Director

Dedication

This dissertation is dedicated to everyone who buys and sells goods across the globe and dreams to have Visibility on his/her shipment's current location and predict its accurate time to arrival at the destination.

"The line between disorder and order lies in logistics…"

– Sun Tzu

Acknowledgements

Throughout my journey, I've encountered numerous individuals who have entered my life, some fleeting, while others have remained steadfast through the highs and lows. The completion of this dissertation owes a debt of gratitude to these remarkable individuals who have guided and supported me tirelessly over the years.

Foremost among them is my family, the cornerstone of my strength. My wife, Roohi, and my son, Arjun, provide me with unwavering motivation to continually strive for excellence.

My parents, who bestowed upon me the freedom to forge my own path early on, allowing me to grapple with challenges and their consequences, yet consistently offering their unwavering support when needed.

My in-laws, too, have imparted invaluable lessons on humility and pride in one's achievements.

I am fortunate to have siblings who, despite being younger, have often assumed the role of elders. My sister, Shweta, and my brother, Pulkit, have enriched my life with their continuous support.

My boss, Yogesh, has been a trusted friend, mentor, and a grounding force, providing invaluable insights and reality checks when needed.

My colleagues, Praveesh and Amit, deserve special mention for their wealth of knowledge in technology and business, offering unwavering support and candid feedback over the years.

My dear friends, Ravi, Ankit, and Gaurav (include my initials and we become R.A.G.A.), whose unwavering support has been a constant source of strength through thick and thin.

And finally, I extend my heartfelt appreciation to my guide and DBA mentors, Dr. Lidija Preglej and Dr. Anna Provodnikova, for their invaluable support and guidance throughout my journey, from the inception of my thesis concept to its final submission.

I am deeply grateful to all those who have played a role in shaping my perspective on life, reminding me of its transient nature and the importance of embracing both its joys and sorrows.

In conclusion, I acknowledge that this achievement would not have been possible without the unwavering support and guidance of each, and every individual mentioned here. Thank you, from the bottom of my heart, for being a part of this journey and contributing to its success.

ABSTRACT

HEURISTIC APPROACH TO IMPROVE EXPECTED TIME OF ARRIVAL PREDICTION FOR GLOBAL INTERMODAL TRANSPORT NETWORKS

Arun Singh
2024

Dissertation Chair: <Chair’s Name>
Co-Chair: <If applicable. Co-Chair’s Name>

This research endeavours to delve into the intricate dynamics of managing Estimated Time of Arrivals (ETA) of shipments arriving to the destination in the global supply chain ecosystem. In today's interconnected world, where buyers and suppliers are scattered across continents, the accuracy of estimated departure and arrival predictions holds immense significance. These forecasts serve as critical benchmarks around which various operational decisions are made, ranging from transport planning to inventory management and sales forecasting.

At the heart of this research lies a central question: What are the accuracies of current ETA predictions of intermodal shipments, and how can they be improved? After literature review and analysis of various reports, it was found that the ETA prediction accuracy is least when it comes to Ocean moves. When main move of transport is Air or Rail, the accuracy of the ETA prediction is on a higher side. But in a move that contains Dray-Ocean-Dray, the accuracy predictions drop drastically. To address this challenge

comprehensively, the study aims to analyse the existing practices employed by Ocean carriers and Logistics service providers in estimating departure and arrival times.

A multitude of factors, including weather conditions, port congestion, vessel schedules, and unforeseen disruptions, influence the accuracy of the ETA predictions. Against this backdrop, the research aims to unravel the operational realities faced by stakeholders involved in managing supply chain logistics across the globe.

Armed with a deep understanding of existing practices and operational challenges, the research endeavours to develop a more robust approach to estimate predictions. By leveraging insights gleaned from data analysis and industry expertise, the study seeks to propose innovative solutions that enhance prediction accuracy and reliability.

The research methodology encompasses a multifaceted approach, combining qualitative and quantitative analysis to gain a comprehensive understanding of the research subject. Data will be collected from various sources, including Ocean carriers, Logistics service providers, industry reports, and academic literature.

The findings of this research are expected to have significant implications for stakeholders across the global supply chain ecosystem. By shedding light on the accuracies of current ETA predictions and proposing strategies for improvement, the study aims to empower buyers, suppliers, carriers, and logistics service providers with the knowledge and tools needed to optimize supply chain operations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ETA	Expected Time of Arrival
3PLs	Third-Party Logistics
4PLs	Fourth Party Logistics
MtCO2	Million Tons of Carbon Dioxide
IMO	International Maritime Organization
TMS	Transportation Management Systems
IoT	Internet of Things
AIS	Automatic Identification System
VTS	Vessel Traffic Services
ETD	Estimated Time of Departure
DTT	Dynamic Travel Time
DDT	Dynamic Delay Time Model
R	Radius
PCI	Port Congestion index
O-D pair	Origin-Destination pair
VTR	Vessel Travel Time
VWT	Vessel Wait Time (VWT)
VDT	Vessel Delay Time
CN	China
CO	Cambodia

HK	Hong Kong
ID	Indonesia
IN	India
JP	Japan
KR	Korea
LK	Sri Lanka
SG	Singapore
TH	Thailand
TW	Taiwan
VN	Vietnam
US	United States of America

CHAPTER I:

INTRODUCTION

The globalization of trade and commerce has ushered in an era of unprecedented connectivity, facilitating the seamless exchange of goods across vast distances and diverse transportation modes. From the bustling manufacturing hubs of Asia to the consumer markets of North America and Europe, the intricate network of global intermodal transport plays a pivotal role in sustaining the modern economy. At the heart of this complex web of logistics lies the critical task of predicting the Expected Time of Arrival (ETA) for shipments, a fundamental determinant of supply chain efficiency, customer satisfaction, and overall business success.

In the bustling metropolis of Chicago, a customer eagerly awaits the arrival of a shipment containing the latest iPhones manufactured in China. Little does the customer know of the intricate journey these iPhones undertake, traversing continents and crossing oceans before reaching their final destination. From the factory floor to the nearest railyard, from the port of departure to the port of entry, and from the warehouse to the retailer's doorstep, each leg of the shipment journey is fraught with uncertainties and challenges that can disrupt the carefully choreographed flow of goods.

The importance of accurate ETA prediction cannot be overstated, as delays along the supply chain can have far-reaching consequences for businesses and consumers alike. A delay in the importation of raw materials can cascade into production bottlenecks, missed delivery deadlines, and unhappy customers. Moreover, delays incur additional costs, including detention and demurrage charges at ports and terminals, further eroding profit margins and straining cash flow. In an era where just-in-time inventory management and fast delivery have become the norm, the ability to predict ETA with precision is a competitive advantage that cannot be ignored.

1.1 Background

Intermodal transport is a system of transporting goods using multiple modes of transportation, seamlessly coordinated to optimize efficiency, cost-effectiveness, and sustainability. It involves the movement of cargo using a combination of methods such as trucks, trains, ships, and airplanes, with containers or other standardized units that can easily be transferred between different modes of transport without the need to handle the contents.

Here's a breakdown of the key components and benefits of intermodal transport:

1.1.1 Modes of Transportation

Intermodal transport integrates various modes of transportation, including road (trucks), rail (trains), water (ships, barges), and air (airplanes). Each mode is utilized for its strengths and efficiencies in different parts of the transportation journey (European Commission, 2023).

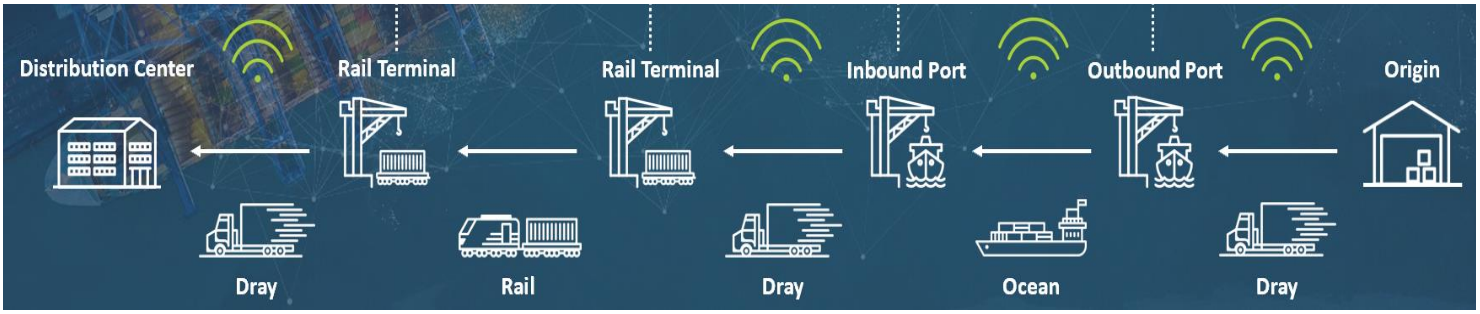


Figure 1: Modes of Intermodal Transport
(Source: Created by Learner)

- **Road (Trucks):** Trucks provide flexibility and accessibility for door-to-door delivery. They are ideal for short to medium-distance transport and for reaching destinations not easily accessible by other modes (European Commission, 2023).
- **Rail (Trains):** Rail transport is highly efficient for long-distance haulage, especially for bulk goods. Trains can carry large volumes of cargo over land at relatively low costs and with lower environmental impact compared to trucks (European Commission, 2023).
- **Water (Ships, Barges):** Water transport, primarily via ships and barges, is optimal for transporting goods over long distances, especially for bulk cargo. It's cost-effective, energy-efficient, and reduces congestion on roads and highways (European Commission, 2023).
- **Air (Airplanes):** Air transport offers unparalleled speed for time-sensitive cargo and goods requiring rapid delivery across long distances. While more

expensive than other modes, air transport is essential for certain high-value or perishable goods (European Commission, 2023).

1.1.2 Containerization

A central feature of intermodal transport is containerization. Goods are typically packed into standardized containers that can be easily transferred between different modes of transport without the need to unpack and repack the cargo. Standardized containers ensure compatibility across various modes and simplify handling and transfer processes (TCIL, 2018).

Figure 2: Containerization in Intermodal Logistics
(Source: TCIL, 2018)

- Standardized containers, typically made of steel, enable efficient handling and transfer of goods between different modes of transport.
- Containers come in various sizes, with the most common being 20-foot and 40-foot lengths, accommodating different types of cargo.
- Containerization simplifies loading and unloading processes, reduces the risk of damage to goods, and enhances security during transit.

1.1.3 Efficiency and Cost-effectiveness

Intermodal transport offers increased efficiency and cost-effectiveness compared to single-mode transportation. By utilizing multiple modes, companies can optimize routes, reduce transit times, and lower overall transportation costs. For example, long-haul transport by rail or water may be more economical than using trucks for the entire journey (Dyrda, 2016).

- Intermodal transport optimizes transportation routes and modes to minimize costs and maximize efficiency.
- By combining multiple modes, companies can leverage the strengths of each mode, such as the cost-effectiveness of rail for long-haul transport and the flexibility of trucks for last-mile delivery.
- Intermodal networks streamline operations, reduce empty miles, and improve asset utilization, ultimately leading to cost savings for businesses.

1.1.4 Flexibility and Reliability

Intermodal transport provides greater flexibility and reliability in supply chain management. It allows companies to adapt to changing market conditions, transportation capacity constraints, and unforeseen disruptions by leveraging alternative modes of transport or routes (Dyrda, 2016).

a. Intermodal transport provides flexibility to adapt to changing market conditions, customer demands, and disruptions in the transportation network.
b. Companies can switch between modes or routes based on factors such as capacity constraints, weather conditions, or infrastructure maintenance.
c. Intermodal networks often offer redundancy and alternative routes, enhancing reliability and minimizing the impact of disruptions on supply chains.

1.1.5 Environmental Sustainability

Intermodal transport is often more environmentally sustainable compared to single-mode transportation. By shifting freight from road to rail or waterways, companies can reduce carbon emissions and alleviate traffic congestion on highways. Additionally, newer intermodal facilities and equipment are designed with sustainability in mind, incorporating technologies for energy efficiency and emissions reduction (Dyrda, 2016).

a. Shifting freight from road to rail or waterways reduces carbon emissions and alleviates congestion on highways, contributing to environmental sustainability.

b. Intermodal transport promotes modal shift initiatives, incentivizing businesses to choose more environmentally friendly transportation options.

c. Investments in green technologies, such as electrified rail systems and low-emission trucks, further enhance the sustainability of intermodal transport.

1.1.6 Global Connectivity

Intermodal transport facilitates global trade by seamlessly connecting different regions and countries through various transportation networks. Ports, rail hubs, and intermodal terminals serve as key nodes in the global supply chain, enabling the efficient movement of goods between different modes of transport and international markets (Bogusz, 2016).

a. Intermodal transport connects regions and countries through integrated transportation networks spanning land, sea, and air.

b. Ports, rail hubs, and intermodal terminals serve as critical nodes in the global supply chain, facilitating the seamless movement of goods between different modes and international markets.

c. Intermodal corridors, such as the Trans-European Transport Network (TEN-T) or the North American Free Trade Agreement (NAFTA) corridors, enhance connectivity and trade flows across continents.

By considering these aspects, businesses can harness the full potential of intermodal transport to optimize their supply chain operations, improve efficiency, and achieve their strategic goals.

Overall, intermodal transport plays a crucial role in modern supply chain management, offering businesses a strategic approach to logistics that emphasizes efficiency, flexibility, and sustainability in the movement of goods.

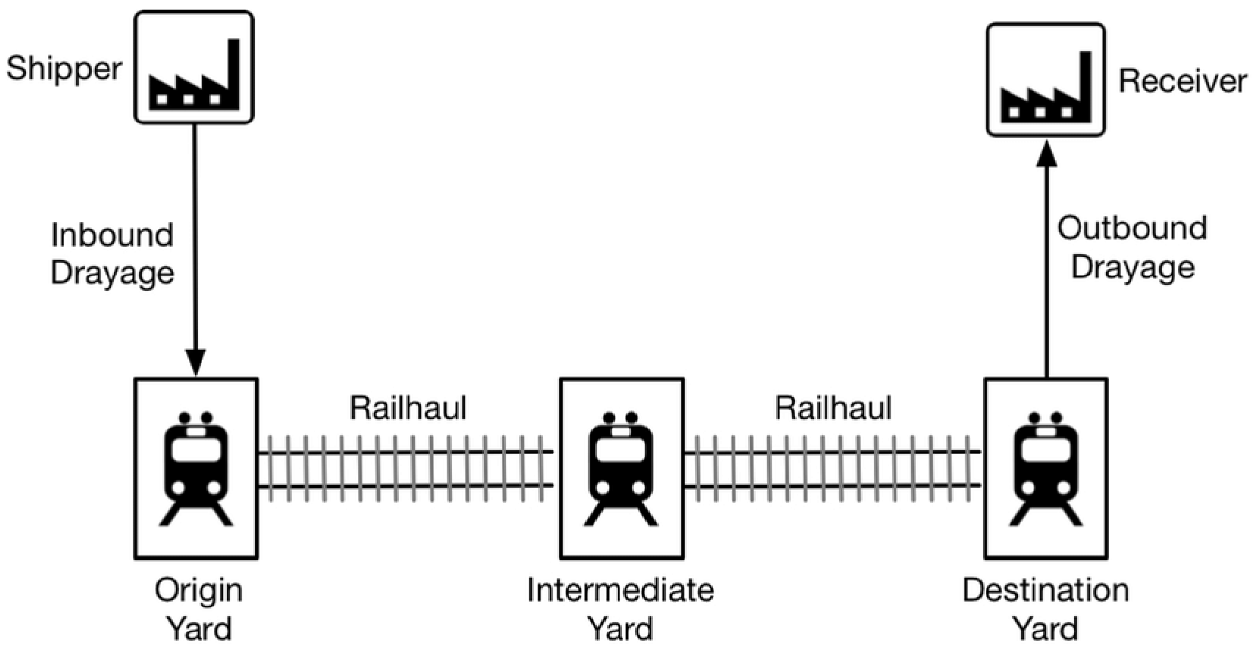


Figure 3: Example of Intermodal Transport
(Source: Created by Learner)

As shown in the above example there is a Dray – Rail – Rail – Dray movement while transporting a shipment from Shipper's factory to the Origin Rail yard on a Truck then from Origin Rail yard to Intermediate Rail yard on a Train then from Intermediate Rail yard Destination Rail yard on a train again and finally from the Destination Rail yard to the Receiver's warehouse on a Truck. There can be multiple other examples of Intermodal transport where we have Ocean and Air involvements as well and the shipment journey could be Dray – Ocean – Dray, Dray – Air – Dray, Dray – Rail – Ocean – Rail – Dray, Dray – Rail – Air – Rail – Dray and so many other combinations depending upon the Origin and Destination for that shipment.

The concept of intermodal transport, which involves the seamless transfer of goods between different modes of transportation, has gained prominence in recent decades as businesses seek to optimize supply chain efficiency and reduce transportation costs. Intermodal transport networks encompass a wide array of transportation modes, including maritime shipping, rail transport, road transport, and air freight, each with its unique advantages and challenges (Bogusz, 2016).

Globalization has further fuelled the growth of intermodal transport, as businesses increasingly rely on international trade to access new markets, source raw materials, and expand their customer base. The rise of e-commerce has further accelerated this trend, as consumers demand fast and reliable delivery of goods purchased online, regardless of their geographic location.

In this interconnected world, the accurate prediction of ETA for shipments has emerged as a critical success factor for businesses engaged in global trade. Timely delivery of goods is not only essential for meeting customer expectations but also for optimizing inventory levels, reducing transportation costs, and maximizing operational efficiency. However, achieving accurate ETA prediction in the context of global intermodal transport networks is a formidable challenge, given the inherent complexities and uncertainties involved (Bogusz, 2016).

1.2 Challenges

While intermodal transport offers numerous benefits, it also presents several challenges that companies and transportation providers must address. Here are some key challenges:

1.2.1 Infrastructure Misalignment

Infrastructure for different modes of transport may not always be well-aligned, causing inefficiencies in intermodal operations. This misalignment can include disparities in gauge standards for railways, lack of intermodal terminals or ports equipped to handle containerized cargo, and inadequate road connections between transport hubs (Zografos, 2004).

1.2.2 Gauge Incompatibility

Railways often have different gauge standards across regions and countries, complicating seamless cross-border transportation. Gauge incompatibility requires transshipment or the use of specialized equipment for cargo transfer (Zografos, 2004).

1.2.3 Terminal Capacity

Inadequate infrastructure at ports, rail yards, and intermodal terminals can lead to congestion and delays in cargo handling. Expansion and modernization of terminals are necessary to accommodate growing volumes of containerized cargo (Zografos, 2004).

1.2.4 Road Connectivity

Poor road connections between transport hubs and industrial centers can hinder efficient intermodal operations. Improving road infrastructure and addressing last-mile connectivity issues are essential for seamless transport (Zografos, 2004).

1.2.5 Interoperability Issues

Ensuring seamless interoperability between various modes of transport is essential for efficient intermodal operations. However, differences in equipment standards, technology systems, and operational practices among transportation providers can hinder interoperability and increase complexity in coordination and communication (Zografos, 2004).

- **Technology Integration:** Different transportation modes may use disparate technology systems for tracking, communication, and documentation. Achieving seamless integration and data exchange between these systems is crucial for effective coordination and visibility throughout the supply chain.
- **Standardization:** Lack of standardized processes and equipment specifications among transportation providers can hinder interoperability. Adopting common standards for equipment, operations, and documentation can facilitate smoother intermodal transactions.

1.2.6 Transshipment Delays

Transshipment points where cargo transfers between different modes of transport are susceptible to delays, especially if there are inefficiencies in handling processes or congestion at terminals. Delays at transshipment points can disrupt supply chains, increase transit times, and impact overall service reliability (Zografos, 2004).

- **Terminal Congestion:** High volumes of cargo, limited terminal capacity, and inefficient handling processes can result in congestion at transshipment points. Implementing efficient terminal operations, optimized yard layouts, and advanced cargo handling technologies can alleviate congestion and reduce transshipment delays.

- **Scheduling Challenges:** Coordinating arrivals and departures of different modes of transport at transshipment hubs can be complex. Improved scheduling algorithms, real-time tracking, and predictive analytics can enhance coordination and minimize dwell times at terminals.

1.2.7 Regulatory and Administrative Hurdles

Intermodal transport involves crossing multiple jurisdictions, each with its own regulatory requirements, customs procedures, and administrative processes. Navigating these regulatory hurdles can be challenging, leading to delays, paperwork burdens, and compliance costs for companies involved in cross-border transport (Zografos, 2004).

- **Customs Clearance:** Cross-border shipments require compliance with customs regulations, import/export documentation, and inspection procedures. Streamlining customs processes, implementing electronic data interchange (EDI), and harmonizing trade policies can expedite clearance and reduce administrative burdens.
- **Regulatory Compliance:** Intermodal transport operations must adhere to diverse regulatory requirements, including safety standards, environmental regulations, and labour laws. Ensuring compliance with regulations across jurisdictions requires robust regulatory oversight and adherence to industry best practices.

1.2.8 Equipment Imbalance

Balancing equipment availability, such as containers and chassis, across different modes of transport can be a challenge. Imbalances in equipment supply and demand can lead to inefficiencies, increased costs, and disruptions in intermodal operations, particularly during peak shipping seasons or in regions with asymmetrical trade flows (Zografos, 2004).

- **Container Repositioning:** Balancing equipment supply and demand for containers, chassis, and other intermodal assets is challenging. Empty container repositioning, equipment sharing agreements, and dynamic pricing mechanisms can help mitigate imbalances and optimize asset utilization.
- **Peak Season Demand:** During peak shipping seasons or fluctuations in trade volumes, equipment shortages or surpluses can occur. Flexibility in equipment procurement, leasing options, and strategic deployment strategies are essential for managing seasonal demand variations.

1.2.9 Last-Mile Connectivity

While intermodal transport excels in long-distance haulage, the last-mile segment of the supply chain, involving delivery to final destinations, often poses challenges. Limited access to intermodal terminals, congestion in urban areas, and restrictions on trucking operations can impede last-mile connectivity and increase costs for shippers (Zografos, 2004).

- **Urban Congestion:** Delivering cargo from intermodal terminals to final destinations in urban areas can be hindered by congestion, traffic restrictions, and limited access to distribution centres. Implementing efficient urban logistics solutions, such as consolidation centres, micro-distribution hubs, and alternative delivery modes, can improve last-mile connectivity.

- **Infrastructure Investment:** Investing in road infrastructure upgrades, last-mile delivery facilities, and sustainable transportation solutions is crucial for enhancing last-mile connectivity and reducing delivery lead times.

1.2.10 Security Risks

Intermodal transport involves multiple handovers and transfer points, increasing the risk of cargo theft, tampering, or damage. Maintaining security throughout the supply chain, especially during transshipment and storage, requires robust security protocols, tracking systems, and collaboration between stakeholders (Regan, 2004).

- **Cargo Theft:** Intermodal transport involves multiple handovers and transfer points, increasing the risk of cargo theft or tampering. Implementing security protocols, surveillance systems, and cargo tracking technologies can mitigate security risks and enhance cargo visibility throughout the supply chain.
- **Supply Chain Transparency:** Enhancing transparency and traceability of shipments through digitalization and blockchain technologies can improve supply chain security and facilitate rapid response to security incidents or breaches.

1.2.11 Environmental Impact

While intermodal transport offers environmental benefits compared to single-mode transportation, challenges remain in reducing the carbon footprint of intermodal operations. Addressing emissions from trucks, trains, ships, and other transportation modes, as well as minimizing the environmental impact of infrastructure development, requires investments in green technologies and sustainable practices (Zografos, 2004).

- **Emissions Reduction:** Despite its environmental advantages, intermodal transport still contributes to carbon emissions and pollution. Investing in low-emission technologies, alternative fuels, and energy-efficient infrastructure can mitigate the environmental impact of intermodal operations.
- **Sustainable Practices:** Promoting sustainable practices, such as modal shift initiatives, eco-friendly packaging, and route optimization, can further reduce the environmental footprint of intermodal transport and support climate action goals.

Addressing these challenges requires collaboration among stakeholders, innovative solutions, and strategic investments in infrastructure, technology, and sustainable practices to build resilient and efficient intermodal transport networks. Idea behind this research is not to solve all the problems related to Intermodal Transport. Rather, this research aims to focus upon the right predictions for Estimated Time of Arrival so that the alternate plans can be made in case of delays and a lot of these problems can be resolved (Regan, 2004).

1.3 Research Problem

The challenge of predicting ETA for global intermodal transport networks is multifaceted, encompassing a range of factors, including weather conditions, traffic congestion, port delays, and unforeseen disruptions. These factors introduce variability and uncertainty into the shipment journey, making it difficult to forecast arrival times with precision. As a result, businesses face the constant risk of missed delivery deadlines, increased costs, and customer dissatisfaction.

The impact of shipment delays is not limited to individual businesses but extends across entire industries and economies. The British Toy and Hobby Association reported significant losses in sales due to shipment delays, highlighting the far-reaching consequences of disruptions in global supply chains. Similar challenges are faced by businesses across diverse sectors, from electronics and automotive to pharmaceuticals and consumer goods, underscoring the urgent need for solutions to improve ETA prediction accuracy.

1.4 Purpose of Research

The purpose of the research is to investigate the complexities involved in managing Estimated Time of Arrivals (ETA) of shipments within the global supply chain. Specifically, the study aims to understand the accuracies of current ETA predictions, particularly focusing on intermodal shipments, and to identify factors contributing to inaccuracies, such as those prevalent in Dray-Ocean-Dray moves. By analyzing existing practices employed by Ocean carriers and Logistics service providers, the research seeks to develop innovative solutions to enhance the accuracy and reliability of ETA predictions. Through a comprehensive research methodology combining qualitative and quantitative analysis, the study aims to provide insights that empower stakeholders across the supply chain ecosystem to optimize their operations (Monios, 2014).

1.5 Objective of the Study

The primary objective of this research is to develop and evaluate heuristic approaches aimed at improving ETA prediction for global intermodal transport networks. Specifically, the study aims to:

- To identify the main reasons for shipment delays in global intermodal transport networks.
- To investigate strategies for improving the visibility of planned and active shipments to enhance ETA prediction accuracy.
- To analyse existing methodologies employed by shipment parties to calculate ETAs and identify opportunities for improvement.

To develop heuristic algorithms leveraging data intelligence, carrier inputs, and external factors to enhance the accuracy of ETA predictions.

1.6 Research Questions

Against this backdrop, this research endeavour seeks to address the following questions:

Central Question

- How can we improve ETA prediction for Global Intermodal Transport Networks?

Sub-Questions

- What are the primary reasons for shipment delays in global intermodal transport networks?
- How can the visibility of planned and active shipments be improved to enhance ETA prediction accuracy?
- What methodologies are currently employed by shipment parties to calculate ETAs, and what are their limitations?
- How can we leverage data intelligence, carrier inputs, and external factors such as weather and traffic conditions to improve the accuracy of ETA predictions?

1.7 Research Hypotheses

This research is focused on testing following research hypotheses:

H1: The accuracy of Ocean ETA predictions is significantly improved by incorporating factors like vessel position, vessel speed and port congestions.

H01: The accuracy of Ocean ETA predictions is not significantly improved by incorporating factors like vessel position, vessel speed and port congestions.

H2: The accuracy of Ocean ETA predictions is significantly improved by considering carrier predictions also in the approach and choosing the best of both worlds in a hybrid approach.

H02: The accuracy of Ocean ETA predictions is not significantly improved by considering carrier predictions also in the approach and choosing the best of both worlds in a hybrid approach.

1.8 Significance of the Study

The findings of this research are expected to have significant implications for businesses engaged in global trade, logistics service providers, policymakers, and academia. By improving the accuracy of ETA predictions for global intermodal transport networks, businesses can optimize supply chain operations, reduce costs, and enhance customer satisfaction. Logistics service providers can leverage the insights gained from this research to develop innovative solutions and differentiate themselves in a competitive market landscape. Policymakers can use the findings to inform regulatory frameworks and infrastructure investments aimed at improving the efficiency and resilience of global intermodal transport networks. Finally, academia can benefit from the development of new methodologies and approaches for ETA prediction in complex and dynamic transportation environments (Monios, 2014).

1.9 Scope and Limitations

It is essential to delineate the scope and limitations of this study to provide clarity and context for the research findings. The scope of this study encompasses the development and evaluation of heuristic approaches for improving ETA prediction in global intermodal transport networks. The research will focus on identifying the main reasons for shipment delays, investigating strategies for improving shipment visibility, analysing existing methodologies for calculating ETAs, and developing heuristic algorithms to enhance prediction accuracy.

However, it is important to acknowledge the limitations of this study, including constraints related to data availability, access to proprietary information, and the inherent uncertainties of real-world transportation systems. Additionally, while the research aims to develop generalizable insights and methodologies, the applicability of the findings may be influenced by factors such as geographic location, industry sector, and specific operational contexts.

1.10 Structure of the Thesis

This thesis is structured into five main chapters, each addressing a specific aspect of the research topic:

- Chapter 1: Introduction
- Chapter 2: Literature Review
- Chapter 3: Methodology
- Chapter 4: Data Analysis and Findings
- Chapter 5: Discussion and Implications and Recommendations

Each chapter will delve into the respective topic area in detail, drawing upon relevant literature, theoretical frameworks, and empirical evidence to provide a comprehensive understanding of the research subject.

1.11 Summary

In summary, this introduction provides an overview of the research topic, problem statement, research questions, objectives, significance, scope, and structure of the thesis. The subsequent chapters will delve deeper into each aspect of the research, drawing upon a diverse array of literature, methodologies, and empirical evidence to address the research questions and objectives outlined in this introduction. Through this comprehensive inquiry, the study aims to contribute valuable insights and practical solutions to the challenge of improving ETA prediction for global intermodal transport networks.

CHAPTER II:

REVIEW OF LITERATURE

2.1 Introduction

ETA predictions can be influenced by various factors such as weather conditions, port congestion, customs clearance, and unforeseen events, making accurate predictions challenging.

Ocean carriers use a combination of data, technology, and predictive analytics to estimate the arrival times of their vessels. Here are some key factors and methods involved in the process of generating Estimated Time of Arrival (ETA) predictions:

- Vessel Tracking Systems: Carriers utilize advanced vessel tracking systems that often include GPS technology. These systems provide real-time information on the location and movement of the vessels (Altinkaya, 2013).
- Voyage Planning Software: Carriers often use sophisticated voyage planning software that considers various parameters such as vessel speed, fuel consumption, weather patterns, and optimal routing. These tools help optimize the voyage for efficiency and accuracy in ETA predictions (Altinkaya, 2013).
- Historical Data and Performance Metrics: Carriers analyse historical data related to their vessels' previous voyages, considering factors like average speed, route efficiency, and port turnaround times. Performance metrics help in understanding the typical behaviour of a vessel (Altinkaya, 2013).

While these factors help in giving them a high-level forecast of estimated departure and arrival dates however, these predictions are not always precise due to the dynamic nature of maritime logistics.

2.1.1 Current state of Container Shipping and Ports Industry

Globalization, deregulation, the integration of logistics, and the widespread use of containerization have transformed the landscape of the port and shipping industry. Port and maritime enterprises face the task of redefining their functional roles within the value chain to generate customer value and secure the sustainability and expansion of their operations. Rather than maintaining the status quo, companies are actively engaged in disrupting existing paradigms. Terminal operators and shipping lines often pursue distinct paths, seeking higher profit margins and enhanced customer satisfaction, and they find themselves compelled to alter their course at various points. Although the Atlantic Rim served as the birthplace of containerization, the economically vibrant East Asia has emerged as the predominant global container region. Over the years, Asia's contribution to worldwide container port throughput has surged from 25 percent in 1980 to approximately 46 percent today, while Europe has experienced a decline from 32 percent to 23 percent. The ascent of global containerization can be attributed to the complex interplay of macroeconomic, microeconomic, and policy-related factors (Notteboom, 2004). Below is a table that depicts Mergers and Acquisitions and strategic alliances amongst the top global carriers -

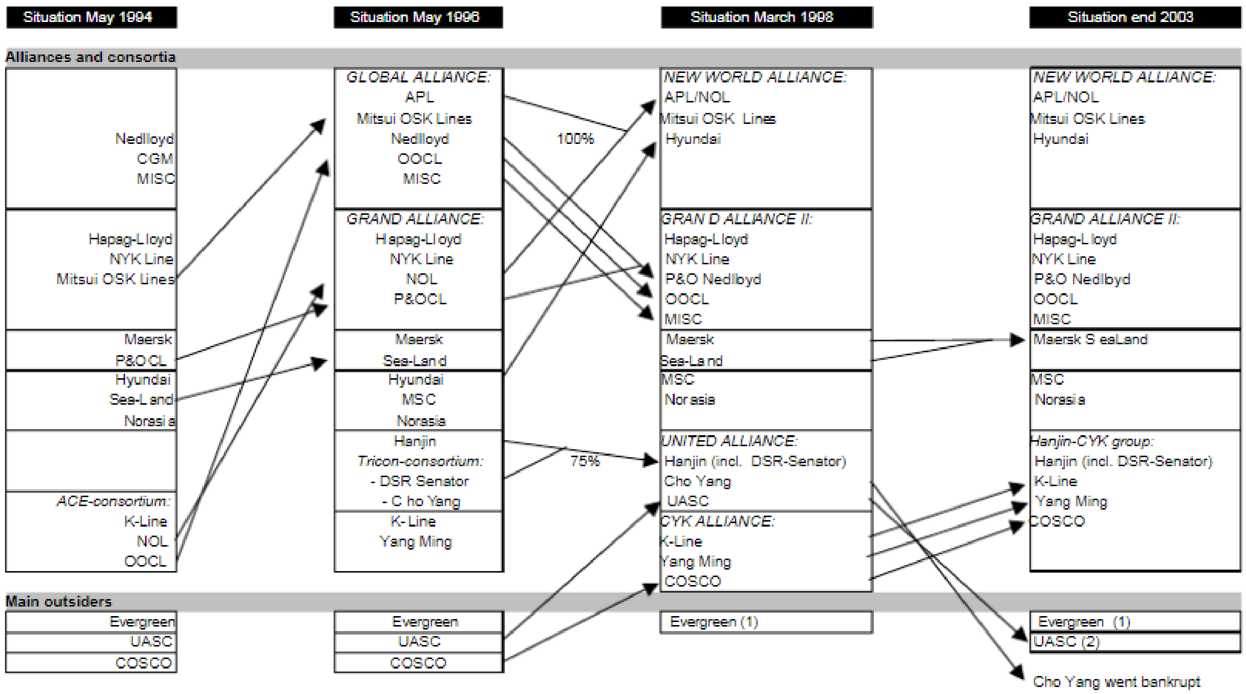


Figure 4: Mergers, Acquisitions, Strategic Alliances.
(Source: Notteboom, 2004)

The implementation of market liberalization has been shown to foster the advancement of logistics on a global scale. The intricacies of international supply chains have heightened, and logistic frameworks undergo continual evolution due to various influences such as globalization, expansion into new markets, the shift towards mass customization in response to product and market segmentation, the adoption of lean manufacturing practices, and associated changes in costs. The demand from customers for a more extensive range of global services, along with the desire for truly integrated services and capabilities (including design, build, and operate functions), has prompted the adoption of integrated logistics strategies. This shift has moved away from transportation centric Third-Party Logistics (3PLs) towards providers specializing in warehousing and distribution. Simultaneously, it has paved the way for innovative forms of non-asset-related logistics service provision, notably Fourth Party Logistics (4PL).

The escalating competition on the supply side has exerted pressures on cost management and profit margins. These transformations in supply chains and logistics models compel container ports and shipping lines to reassess their roles within the broader logistics process (Notteboom, 2004).

2.2 Theoretical Framework

The global logistics market size was accounted at USD 7.98 trillion in 2022 and it is expected to reach around USD 18.23 trillion by 2030. More than 90% of this global trade happens over sea. Annually, ships transport a substantial 11 billion tons of goods worldwide, averaging an impressive 1.5 tons per person given the current global population. The crucial role of shipping in efficiently transferring goods from production to consumption areas is fundamental to modern life (International Chambers of Shipping, 2023).

In economic regions like the European Union, shipping plays a pivotal role, constituting 80% of total imports and exports by volume and approximately 50% by value. By 2019, the overall value of the annual global shipping trade had surpassed an impressive 14 trillion US Dollars (International Chambers of Shipping, 2023).

To evaluate the impact of the delays in the logistics industry, research was conducted in Chinese Shipping industry and ports and the results of the report are astonishing. During the survey the import / export managers were requested to delineate the nature of expenses linked to supply chain interruptions and delays in transportation, with lost sales being a prevalent response. In every instance, the ramifications of delays were unmistakably tied to substantial harm to a company's image, standing, and customer connections. Exporters pointed out that delays might compel them to extend price discounts, offer rebates, or incur penalty payments. Prolonged delays could potentially lead to order cancellations, cargo returns, or the auctioning of shipments at foreign ports. Exporters also expressed heightened concerns regarding increased transport costs due to rerouting or changes in transportation modes. The contractual and payment terms had a noteworthy impact on exporters' cash flows when delays occurred. Yet, when questioned about the specific costs arising from these delays, the study found that the Increase Administration Workload and Costs was up to 42.9%, Increase Transport Costs was up to 37.1%, Increase in Inventory Costs was up to 28.6% and up to 35.7% affect was seen on their Sales and Promotion plans (Zhang, 2010).

This is where the need of knowing a potential delay comes into play. If there is a potential delay in the Estimated departure of a vessel, the exporter might plan a different route / mode all together. If there is a potential delay in the Estimated arrival of a vessel, the importer might plan for a faster service for last mile or a substitute goods all together.

2.1.1 Minimizing carbon dioxide emissions in the Maritime Industry

International shipping contributed approximately 843 million tons of carbon dioxide (MtCO2) annually, constituting around 3% of the total anthropogenic carbon emissions (IMO, 2007). To provide context, if classified as an emissions source, "international shipping" would rank just below Germany and ahead of the UK. Only six nations surpass international shipping in greenhouse gas production. The emissions from shipping are reported to have doubled since 1990, and without regulatory interventions, they are anticipated to increase by a factor of 2.4 to 3 by 2050 (Balster, 2020).

Addressing shipping emissions is intricate due to the industry's global nature. Ocean-going vessels procure fuel globally and emit carbon dioxide during journeys between countries, making it challenging to attribute these emissions to specific nations. Consequently, international shipping was not included in the emission reduction targets outlined in the Kyoto Protocol. The responsibility to address shipping emissions was delegated to developed economies through the Kyoto Protocol, mandating collaboration with the International Maritime Organization (IMO). However, as of now, there has been no consensus within the IMO on a framework to cap global shipping emissions (House of commons, 2009). Now all this emission cannot be reduced to zero. However, it can certainly be reduced if the delays during the transit or on the ports can be avoided. As the saying goes – “A lean supply chain is a clean supply chain”. Therefore, it becomes ever critical to have an accurate forecast of the departures and arrivals of the vessels so that the busier routes and ports and sometimes modes can be avoided for a quicker and greener route, port or mode respectively.

2.3 Theory of Reasoned Action

In recent years, ship routing has garnered significant attention from both academia and industry. Forecasting ship routing involves predicting the most optimal route for a vessel based on various factors such as weather conditions, sea currents, traffic, fuel efficiency, and safety considerations. Common objectives include minimizing operating costs, reducing fuel consumption, and avoiding delays associated with passage risks or adverse conditions. Challenges within this field involve determining the most efficient route and optimal sailing speed for a given voyage, considering environmental factors such as wind and waves.

Companies have started the integration of multimodal transportation management systems for optimizing ship routing. There are various challenges and opportunities in incorporating multiple modes of transportation such as ships, trucks, and trains into a unified TMS framework. Algorithms and optimization techniques are being explored to improve the efficiency and reliability of ship routing within multimodal logistics networks (Zhang, et al., 2019). Here are some key aspects and methods involved in forecasting ship routing such as:

Weather Conditions - Analyzing meteorological data to anticipate weather patterns and their impact on sea conditions. This includes considerations for wind speed, wave height, and storm predictions (Zontul, et al. 2013).

Sea Currents and Tides - Understanding the influence of ocean currents and tides on ship movements. Incorporating information on favorable currents can significantly impact fuel efficiency and voyage duration (Zontul, et al. 2013).

Traffic and Navigation - Monitoring maritime traffic to avoid congested routes and potential collisions. Utilizing real-time data and historical traffic patterns helps in planning efficient and safe routes (Zontul, et al. 2013).

Fuel Efficiency - Implementing algorithms that consider fuel consumption and emissions to determine the most fuel-efficient route. This involves optimizing speed, route deviations, and engine efficiency (Zontul, et al. 2013).

Safety Considerations - Factoring in safety parameters such as proximity to hazardous areas, piracy-prone zones, and navigational obstacles. Ensuring compliance with international maritime regulations (Zontul, et al. 2013).

Vessel Characteristics - Considering the specific attributes of the vessel, including its size, draft, speed, and cargo capacity. Tailoring the route based on the vessel's capabilities enhances overall performance (Zontul, et al. 2013).

2.3.1 Current forecasting methods in Multimodal TMS

Examining past voyages and their outcomes to identify patterns and trends. Historical data helps in understanding common routes, weather challenges, and optimal strategies. Also, utilizing machine learning models to process vast amounts of data and predict optimal routes. These models can continuously learn and adapt to changing conditions, improving accuracy over time **(**Smith, et. al., 2020).

There are multiple new aged forecasting techniques used for ship routing within multimodal Transportation Management Systems (TMS), such as time series analysis, machine learning algorithms, and hybrid approaches. Each of these have their own strengths and limitations. But after a lot of hit and trial, many advancements have been made in ship routing forecasting, though they come with their own set of challenges. Similarly, incorporating data from onboard sensors, satellite systems, and other Internet of Things (IoT) devices are helping in Estimated arrival times for Multimodal TMS voyages. Real-time information enhances the accuracy of routing forecasts **(**Smith, et. al., 2020).

Weather also plays a big impact on Forecasting Models of Multimodal TMS for Ship Routing Optimization. Wind speed, tides, and disruptions like tsunamis can turn all forecasting relied upon historical data, upside down. Therefore, it is important to know the weather data in advance and the alternate routes that the vessel can take to navigate through the bad weather. The algorithms should consider these moves and update the forecasted ETAs accordingly to mitigate these challenges to improve ship routing decisions (Johnson, et. al., 2017).

Forecasting ship routing is an evolving field that benefits from advancements in technology, data analytics, and the integration of smart systems. The goal is to enhance the efficiency, safety, and sustainability of maritime transportation.

2.3.2 Challenges with current forecasting methods in Multimodal TMS

Multimodal Transportation Management Systems (TMS) play a critical role in optimizing logistics operations by facilitating the efficient movement of goods across various modes of transportation such as road, rail, sea, and air. Forecasting methods are essential components of TMS, as they enable logistics managers to anticipate demand, allocate resources effectively, and optimize transportation routes. However, despite advancements in technology and methodologies, challenges persist in accurately forecasting demand and optimizing multimodal transportation (Johnson, et. al., 2017). This literature review aims to explore the key challenges associated with current forecasting methods in multimodal TMS.

2.3.3 Data Integration and Accuracy

One of the primary challenges in forecasting for multimodal TMS is the integration of disparate data sources from different modes of transportation. Transportation data often comes from various sources, including carriers, shippers, and third-party providers, each with its own format and level of granularity. Ensuring the accuracy and consistency of data across these sources poses significant challenges. Moreover, data quality issues such as incomplete, inconsistent, or outdated information can lead to inaccurate forecasts, resulting in suboptimal transportation planning and execution (Giannopoulos et al., 2017).

2.3.4 Mode Selection and Coordination

Multimodal transportation involves selecting the most cost-effective combination of transportation modes to move goods from origin to destination. However, determining the optimal mode selection and coordination poses challenges due to the complexity of factors involved, including transportation costs, transit times, capacity constraints, and service reliability. Forecasting methods must account for these factors and dynamically adjust mode selections based on changing market conditions and operational constraints (Higgins et al., 2019).

2.3.5 Uncertainty and Variability

Forecasting for multimodal TMS is inherently challenging due to the high degree of uncertainty and variability in transportation operations. External factors such as weather conditions, traffic congestion, labor strikes, and geopolitical events can significantly impact transportation networks, leading to delays, disruptions, and fluctuations in demand. Traditional forecasting methods often struggle to account for these uncertainties and may produce inaccurate predictions, leading to inefficiencies and increased costs (Savelsbergh et al., 2018).

2.3.6 Dynamic Demand Patterns

The demand for transportation services exhibits dynamic and non-linear patterns, making it challenging to forecast accurately. Seasonal fluctuations, market trends, promotional activities, and unforeseen events can cause demand to vary significantly over time and across different transportation modes. Forecasting methods must be able to capture and adapt to these dynamic demand patterns to ensure optimal resource allocation and service levels (Yang et al., 2020).

2.3.7 Scalability and Performance

As transportation networks continue to expand and become more complex, forecasting methods must be scalable and capable of handling large volumes of data in real time. Traditional forecasting techniques may struggle to cope with the scalability requirements of multimodal TMS, leading to performance issues such as slow processing times and reduced accuracy. Advanced forecasting models leveraging machine learning and artificial intelligence show promise in addressing these scalability challenges and improving forecasting accuracy (Huang et al., 2021).

Forecasting for multimodal Transportation Management Systems (TMS) presents various challenges, including data integration, mode selection, uncertainty, dynamic demand patterns, and scalability. Addressing these challenges requires the development of advanced forecasting methodologies that can effectively integrate heterogeneous data sources, adapt to dynamic operating conditions, and provide accurate predictions in real time.

2.3.8 Mitigation tactics to overcome the current challenges

Transportation disruption, a common source of business interruptions, can cause significant economic loss to a lean supply chain (Paul, 2019). To have the real-time Visibility of the shipments some companies have started using AIS data to predict the route and position of a Shipment, predict the ETA of the shipment to a destination well ahead of time and reduce the uncertainty of the arrival times (Meijer, 2017). The Automatic Identification System (AIS) is an automatic tracking system that uses transceivers on ships and is used by Vessel Traffic Services (VTS).

A similar study shows that adapting to the latest machine learning capabilities, some companies are also trying to make use of their historical shipment journey data and predict how the future shipment journey on the same Origin-Destination pairs may look like, including the delays, ETAs, and the exceptions on the way (Balster, 2020). Companies are adapting innovative techniques to have better visibility and accurate ETA predictions as having uncertainty leads to disaster (Urciuoli, 2018). To emphasize this, researchers modelled the impacts of not having accurate shipment visibility.

While these are some great strides towards ETA accuracy and shipment visibility, there is still a lot to do. Not enough research has been done to explore how we can combine a few of the ML capabilities with AIS data or Carrier inputs with weather and traffic data and come up with a hybrid approach to improving the ETA accuracy. That is exactly where my research aims at.

2.4 Literature Gap

The literature on Estimated Time of Arrival (ETA) predictions for ocean transport is extensive as described above. However, there are still several gaps and areas for further research. Here are some of the key literature gaps in this field:

2.4.1 Data Integration and Fusion

Research in this area could explore methods for integrating diverse data sources, such as AIS data, satellite imagery, oceanographic data, and historical voyage data. Techniques for data fusion, including statistical fusion methods, machine learning approaches, and Bayesian inference, could be investigated to combine information from multiple sources and improve prediction accuracy (Yoon et al., 2023). Additionally, studies could focus on addressing data quality issues, such as missing or erroneous data, through data cleaning, imputation techniques, and quality control measures.

2.4.2 Machine Learning Models

There is a need for comparative studies evaluating the performance of different machine learning models for ETA predictions in ocean transport. Research could investigate the strengths and limitations of various models under different operating conditions, voyage scenarios, and data characteristics (Yoon et al., 2023). Furthermore, studies could explore ensemble learning techniques that combine predictions from multiple models to improve overall accuracy and robustness.

2.4.3 Uncertainty Quantification

Research in this area could develop probabilistic forecasting models that provide not only point estimates of arrival times but also probabilistic distributions capturing the uncertainty associated with predictions. Techniques such as Monte Carlo simulation, ensemble forecasting, and Bayesian inference could be utilized to quantify and propagate uncertainty (Bodunov et al., 2018). Studies could also focus on developing decision-making frameworks that account for uncertainty in ETA predictions, considering risk tolerance levels, cost-benefit trade-offs, and decision sensitivity to uncertainty.

2.4.4 Dynamic Model Adaptation

Research could explore adaptive learning algorithms that continuously update prediction models based on incoming data and changing environmental conditions. Techniques such as online learning, reinforcement learning, and adaptive filtering could be investigated to dynamically adjust model parameters and improve prediction accuracy over time (Fancello et al., 2011). Additionally, studies could examine methods for detecting and responding to sudden changes or anomalies in data streams, such as vessel behavior deviations, adverse weather events, or unexpected port disruptions.

2.4.5 Port Operations and Congestion

There is a need for research that integrates port-related factors into ETA prediction models, such as berth availability, terminal congestion, and port infrastructure constraints. Studies could analyze historical port data to identify patterns and trends that impact vessel turnaround times and arrival schedules (Bodunov et al., 2018). Furthermore, research could explore simulation modeling techniques to assess the impact of port operations on vessel schedules and develop predictive models that account for port-specific dynamics and constraints.

2.4.6 End-to-End Supply Chain Integration

Studies could investigate the integration of ETA predictions with broader supply chain processes, such as inventory management, production planning, and distribution logistics. Research could develop decision support systems that leverage ETA predictions to optimize supply chain operations, minimize inventory holding costs, and enhance overall efficiency (Fancello et al., 2011). Additionally, collaborative research involving stakeholders across the supply chain could identify opportunities for coordination and synchronization of activities based on ETA information, leading to improved resource allocation and enhanced customer service levels.

2.4.7 Environmental Impact Assessment

Research in this area could assess the environmental implications of different ETA prediction strategies, considering factors such as fuel consumption, emissions, and marine pollution risks. Studies could develop models to quantify the environmental impact of vessel routing decisions, speed adjustments, and operational practices based on ETA predictions (Fancello et al., 2011). Furthermore, research could explore the development of green navigation strategies that leverage ETA predictions to optimize vessel routes, minimize fuel consumption, and reduce greenhouse gas emissions in maritime transportation.

2.4.8 User-Centric Decision Support

There is a need for research that focuses on understanding the information needs and decision-making processes of stakeholders involved in ocean transport, including shippers, carriers, port operators, and logistics providers. Studies could employ user-centered design methodologies, such as stakeholder interviews, surveys, and usability testing, to elicit user requirements and preferences. Additionally, research could develop decision support tools that provide tailored ETA predictions, visualization tools, and scenario analysis capabilities to support informed decision-making by different user groups. User-centric interfaces, alerts, and notifications could be designed to communicate ETA predictions and associated uncertainties effectively to stakeholders and facilitate timely decision-making (Hasheminia et al., 2017).

By addressing these literature gaps, researchers can contribute to advancing the state-of-the-art in ETA predictions for ocean transport and developing innovative solutions that improve operational efficiency, enhance environmental sustainability, and support informed decision-making across the maritime supply chain.

2.5 Summary

For centuries, the transportation of goods via ocean routes has been a common practice. However, the precise tracking of these goods and the vessels transporting them only began a few decades ago, along with the ability to predict their expected arrival times at their destinations. This presents a distinct opportunity to enhance the accuracy of estimated arrival predictions. Although organizations are endeavoring to refine these predictions through data engineering and other innovative techniques, there remains ample room for improvement (Bodunov et al., 2018).

As described in the Literature Gap that there are various aspects that can be explored to improve the accuracy of the estimate predictions. This research aims to utilize the aspects such as diverse data sources, such as AIS data, satellite imagery, oceanographic data, and historical voyage data and develop an inventive hybrid approach by incorporating additional external factors such as -

- Weather Forecasting: Weather conditions significantly impact maritime travel. Carriers incorporate weather forecasts into their ETA predictions to account for potential delays or adjustments in the voyage route (Wartsila Publication, 2022).

Figure 5: Impact of bad weather
(Source: Wartsila publication, 2022)

- Port Congestion and Delays: Port congestion and delays in port operations can affect arrival times. Carriers monitor the status of ports along the route and consider historical data on port efficiency and delays (Wartsila Publication, 2022).
- Communication with Ports and Authorities: Carriers maintain communication with ports, terminal operators, and relevant authorities to receive updates on berth availability, customs clearance procedures, and any other factors that could impact the vessel's schedule (Wartsila Publication, 2022).
- Machine Learning and Predictive Analytics: Some carriers employ machine learning algorithms and predictive analytics to enhance the accuracy of ETA

predictions. These systems can continuously learn from new data and improve their forecasting capabilities over time (Wartsila Publication, 2022).

Better predictions and Visibility into when the inventory are arriving will help the buyers in reduction of Inventory Carrying cost, reduce stock outs and lost sales, reduce safety stock and reduce price mark downs as they can plan better in terms what, when and how much to buy.

Having said that, there is also a call out here that despite all these efforts, the unexpected events, such as emergencies, geopolitical issues, or equipment failures, can still lead to deviations from the initially estimated arrival times. ETA predictions are, therefore, considered as estimates and subject to change based on real-time circumstances.

CHAPTER III:

METHODOLOGY

3.1 Overview of the Research Problem

This chapter of the research is focused towards explaining the methods that are used to conduct study on "Heuristic Approach to Improve Expected Time of Arrival Prediction for Global Intermodal Transport Networks". This section delves into the intricate details of the research methodology, elucidating the guiding principles, and the underlying methods that are used to drive the study forward. It explains the strategies and approaches selected to gather, sort, process, and analyze the data so that effective research outcome can be produced. Additionally, justification is provided behind selecting particular research method for the investigation purpose as compared to others.

There is special concern given towards conducting this chapter systematically so that the chances of losing any point can be avoided. Additionally, conducting research methodology in a step-by-step manner is crucial as it allows ensuring clarity, consistency, and validity in the research process (Panneerselvam, 2014). Structured approach allows to systematically address complex questions while minimizing biases and errors and enhancing the reliability and relevancy of the research outcome. It has also supported to represent the overall chapter to the readers in a manner that it could be easy to understand. Saunder’s Research Onion model is used to lead this chapter as it offers systematic approach to conduct research study effectively (Sinha, et al., 2018).

Additionally, this chapter also explains the ethical considerations that are maintained to ensure respect for the participants, confidentiality, and anonymity while keeping in mind about plagiarism, copyright, and patent aspect throughout the study to eliminate any kind of hindrance.

3.2 Research Purpose and Questions

This research endeavour seeks to address the following questions:

Central Question

- How can we improve ETA prediction for Global Intermodal Transport Networks?

Sub-Questions

- What are the primary reasons for shipment delays in global intermodal transport networks?
- How can the visibility of planned and active shipments be improved to enhance ETA prediction accuracy?
- What methodologies are currently employed by shipment parties to calculate ETAs, and what are their limitations?
- How can we leverage data intelligence, carrier inputs, and external factors such as weather and traffic conditions to improve the accuracy of ETA predictions?

3.3 Operationalization of Theoretical Constructs

Saunder's Research Onion is a methodological framework that is designed by Mark Saunders and his colleagues to guide the development of a research strategy and to select appropriate methods for conducting different kind of research (Saunders and Tosey, 2012). Its systematic approach allows conducting research with the help of the framework depicted as layers of an onion and each layer represents crucial aspect of the research process. Under this model, researchers require to peel away the outer layers before reaching the core components of methodology that allows ensuring thorough consideration of each aspect (Iovino and Tsitsianis, 2020).

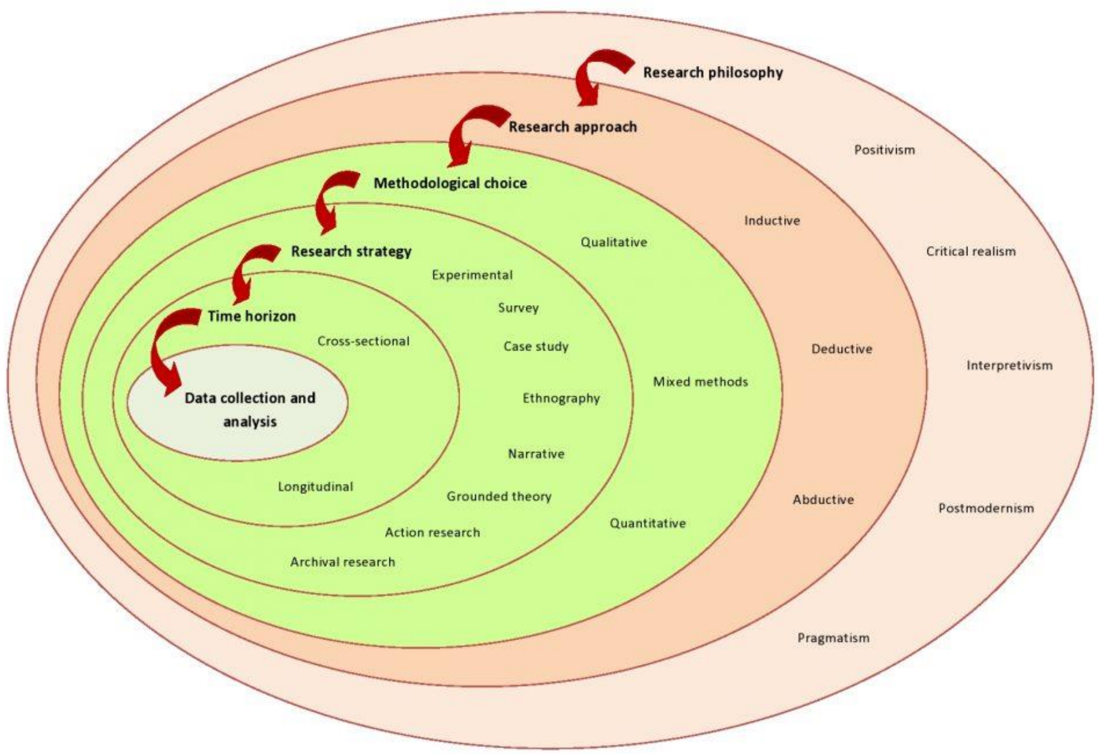


Figure 6: Saunder's Research Onion
(Source: Saunders, and Tosey, 2012)

Each step builds on the previous one which further helps in ensuring that the research question is thoroughly explored through appropriate data collection, analysis, and interpretation methods. This methodical approach facilitates clear communication of the research process and findings that contributes to the reliability and credibility of the research outcomes within the scientific community. The structured breakdown of this model is as follow:

3.4 Research Philosophy

Research philosophy can be stated as a set of guidelines and ideas that offer direction to the researchers for taking decisions related to planning, execution, and interpretation of the research studies (Bhattacharyya, 2006). Research philosophy allows the researcher to underpin the nature and development of the knowledge about the research topic. It dictates the strategy and methods that a researcher required to use to lead the research process effectively. Researchers majorly select among three research philosophies, they are: positivism, interpretivism, and realism (Rivas, 2010). Every philosophy has its own assumptions about the situations and the way in which research is required to be continued. Each philosophy interprets the study and the associated issue in different manner while maintaining consistency between the views and the selected research process. Hence, research philosophy plays significant role in research methodology and allows leading the overall study in the right direction. It makes it essential to select appropriate research philosophy so that appropriate research outcome can be produced. It also explains the identification and classification of data which further helps in result interpretation.

Positivism research philosophy majorly emphasizes on scientific knowledge that are derived from logical and mathematical treatments and remain concern towards reporting sensory experience is of facts (Ryan, 2018). This philosophy remains relevant for the researchers where there is a requirement analyzing research outcome on the basis of quantifiable data and statistical analysis. This philosophy allows leading systematic study with the help of objectivity and observable components. According to this philosophy, reliable knowledge can be gained through sensory experience and can be interpreted through reasoning, logic and scientific inquiry as it allows advocating objectivity, clarity, and predictability (Alharahsheh and Pius, 2020). Positivism relies on empirical evidence as they are measurable in nature and helps to explore natural and social phenomena.

Similarly, according to interpretivism philosophy, reality is multi-layered and complex which makes it important to understand single phenomenon with the help of multiple interpretations (Junjie and Yingxin, 2022). This philosophy helps to interpret the research outcome in a subjective manner which make it relevant for qualitative research. This philosophy remains appropriate for the researches that has multidimensional and socially created reality as it requires grasping the research issue in a subjective manner so that human experiences and social contexts can be represented effectively (Ryan, 2018). This philosophy helps to develop understanding about the experiences in subjective manner.

On the other hand, realism philosophy remains concern towards believing that reality is independent from human thoughts and beliefs, and it can be understood through observation (Pather and Remenyi, 2005). This philosophy states that senses and scientific instruments can support to give accurate representations of the world as it allows uncovering the universal laws and objective truths (Rivas, 2010).

In the context of undertaken research, positivism philosophy is selected as it has supported to induce the knowledge about the selected research topic ‘heuristic approach to improve expected time of arrival prediction for global intermodal transport networks’ and has supported to achieve the research objectives of identifying the main reasons for shipment delays in global intermodal transport networks, investigating the strategies for improving the visibility of planned and active shipments to enhance ETA prediction accuracy, analyzing the existing methodologies employed by shipment parties to calculate ETAs and identify opportunities for improvement and developing heuristic algorithms leveraging data intelligence, carrier inputs, and external factors to enhance the accuracy of ETA predictions. Positivism research philosophy allows involving observation, experiment, and logical analysis that supports giving importance to observable phenomena as well as to the quantifiable facts that supports to eliminate metaphysical speculation (Junjie and Yingxin, 2022). Use of positivism philosophy has supported in this research to represent the data statistically to offer factual knowledge. It has supported to develop effective research outcome empirically with the help of scientific methods.

3.5 Research Approach

Research approach can be defined as the logical sequence and methodology that is required to employ in the research to address the research question (Davidavičienė, 2018). It allows explaining broad assumptions concerning to the data collection and supports to offer proper explanations. It helps to address the research issue appropriately while offering comprehensive framework which further supports to acquire, examine, and evaluate the data so that research can reach to the effective conclusion.

There are majorly two research approaches, they are: inductive and deductive (Soiferman, 2010). Selection of the research approach takes place on the basis of data collection method, organized, and interpreted so that thorough investigation can take place effectively for the chosen research problem. Selection of appropriate research approach allows representing the research design that influences data collection, analysis, and conclusions.

Deductive approach initiates with designing hypotheses and further moves with collecting data and analyzing them to test those hypotheses (Woiceshyn and Daellenbach, 2018). This approach utilizes top-down method that initiates with generalizations and further moves towards observations to evaluate pre-designed hypotheses (Newman, 2000). On the other hand, inductive approach initiates with data collection which further moves towards identifying patterns and relationships so that formulation of new theories can take place. This approach focuses towards exploratory and utilizes bottom-up method.

To lead this research, deductive research approach is chosen as it remains appropriate for quantitative research. It helped to design hypotheses (The accuracy of Ocean ETA predictions is significantly improved or not by incorporating factors like vessel position, vessel speed and port congestions and the accuracy of Ocean ETA predictions is significantly improved by considering carrier predictions or not in the approach and choosing the best of both worlds in a hybrid approach) and allowed surrounding the study towards testing those hypotheses. It has supported to test the relationship among the research study. It has allowed initiating the study with designing expected patterns and further supported to test those hypotheses against observations. This approach supported to explain the causal association between concepts and variables so that hypotheses can be tested.

3.6 Methodological Choices

Methodological choices can be referred as the decisions that researchers make for the selection of particular methods and procedures that is required to be used to collect and analyze the data (Goundar, 2012). Within academia, two predominant methodological choices emerge: quantitative methodological choice and qualitative methodological choice, each offering unique methodologies and approaches (Sinha and Dhiman, 2002).

Quantitative research design primarily focuses on discerning relationships between variables. This method entails collecting data from samples and subjecting them to statistical analysis. It aligns seamlessly with the deductive approach, wherein hypotheses are tested against empirical evidence (Chawla and Sodhi, 2011). Conversely, qualitative research design complements the inductive and abductive approaches. It involves immersive techniques such as interviewing individuals, posing probing inquiries, and extracting nuanced insights from their responses (Flick, 2015). While quantitative research design probes the interplay between variables, qualitative research design delves into the relational dynamics between entities (Flick, 2015). Quantitative strategies are appropriate for the studies that involve mathematical, statistical, and fact-based approaches, while qualitative strategies are applied to studies that involve the life experiences of subjects. Hence, this research has utilized quantitative approach through the means of historical data and other key external factors.

To answer the research questions proposed in the previous section, this researcher has applied quantitative research design. This researcher has used historical data, and the external factors on top of it had affected the final-outcome and has assisted to improve the predictions. The historical data that is collected is further analyzed with the help of descriptive data analysis method to lead cleansing, model selection, model training, model testing and then final prediction. On the basis of result, prediction have taken place which is used to formulate the most likely scenario or explanation for the phenomenon.

3.7 Research Design

Research strategy can be defined as a comprehensive plan that remains concern towards outlining that how a researcher can answer the research questions or hypotheses that is set out in the study (Mukherjee, 2019). It represents the selected methods, techniques, and procedures that is required to be used for the collection, analysis, and interpretation of the data. According to Saunder's research onion, some of the common research strategies are: experimental, survey, observation, case study, ethnography, narrative, grounded theory, action research and archival research (Melnikovas, 2018). Each research strategy has its own **strengths,** and it is chosen on the basis of the nature of the research question, discipline, available resources, and desired level of detail and depth of understanding (Rajasekar and Verma, 2013).

In the context of undertaken research, action research strategy is used as it has supported to involve a participatory research strategy for the collection of the data and has supported to offer practical problem solving so that effective change can be led (Walker, 1997). This strategy involves a cyclical process of planning, acting, observing, and reflecting, to achieve the goal of improving management practices and addressing the issues within a specific context. This strategy has allowed collaborating with the research subjects to generate knowledge and solutions that are directly applicable. It has supported to ensuring that the research findings are relevant and also beneficial to those involved. Use of action research remains supportive to lead promote continuous improvement and adaptive learning.

In the context of undertaken research, using third-party providers for data acquisition offered several advantages. Firstly, it ensured that the data collected was relevant and tailored to the specific needs of the research. These providers typically offer customizable data packages, allowing researchers to access precisely the information required for their analysis.

Secondly, the data obtained from third-party providers was valuable in terms of its accuracy and comprehensiveness. These providers employ advanced technologies and methodologies to collect and process data from diverse sources, including satellite tracking, AIS (Automatic Identification System), and other maritime monitoring systems. As a result, the data provided by these sources is often of high quality and covers a wide range of maritime activities.

Additionally, relying on third-party providers for data collection saved time and resources that would otherwise be spent on manually gathering and verifying information from multiple sources. By leveraging existing data infrastructure and expertise, the research was able to focus on analysis and interpretation rather than data collection logistics.

Overall, the utilization of third-party providers for sailing data ensured that the research had access to relevant, valuable, and reliable data sets essential for conducting meaningful analysis and drawing insightful conclusions.

3.8 Time Horizon

Time horizon can be defined as the period of time over which the research is required to be conducted and data will be collected **(**Patel and Patel, 2019). Determination of the time horizon takes place on the basis of research objectives and type of investigation. Moreover, researcher interest of studying the population may vary at a certain point in time or over a period. While focusing towards all these aspects, time horizon is required to be selected. Time horizon can be majorly divided into two types, they are: cross-sectional time horizon and longitudinal time horizon.

Under cross-sectional time horizon, data id collected at a single point of time as it remains concern towards analyzing a particular phenomenon within a specific time frame so that snapshot of the variables can be provided at that moment (Cummings, 2018). Cross-sectional time horizon remains highly efficient to quickly collect the data from large sample size. However, as data is captured at a particular point, hence, cross-sectional studies do not become able to assess change or develop over time which limits the ability to infer causality.

In the contrary, longitudinal time horizon collects the data over extended period of time that allows researchers to observe changes and developments over time. This type of research helps to study trends, long-term effects, and causal relationships, as it tracks the same subjects repeatedly (Kothari, 2004). However, this method remains more costly, time-consuming, and subject to challenges like participant dropout. Both time horizons have their pros and cons and both of them provide valuable insights however, they serve different purposes based on the research objectives.

Cross-sectional designs offer a momentary glance, suitable for descriptive analysis, hence, to conduct this research, cross-sectional time horizon is selected over longitudinal. Selection of this time horizon has supported to accomplish the overall research within short term time frame and has supported to accomplish the study within deadline.

3.9 Data Collection

Data collection method can be defined as a systematic approach for the collection of the data in a systematic manner so that research questions can be answered in a manner that effective research outcome can be produced (Baker, 2000). In regard to this, researchers select between primary and secondary data collection method according to the research need.

Primary data collection involves gathering new data that does not already exist. This method helps to collect first-hand data directly from the population who are facing the research problem. It involves survey questionnaire, interview, observation, experiment and focus group methods (Mazhar, et al., 2021). Selection of the primary data collection method depends on the requirement of data, sample size, availability of time and effort required.

On the other hand, secondary data collection refers to the use of existing data, collected for some purpose other than the current research project (Nayak and Singh, 2021). It remains appropriate for the research where there are already various past studies conducted related to the selected research topic (Goddard and Melville, 2004). It involves case study analysis, systematic literature review, case study, etc. (Goddard and Melville, 2004). Secondary data collection technique is relatively less time, effort and money consuming as compared to primary.

Both the techniques have their own benefits and drawbacks; hence the selection of the method usually takes place based on complexity of the study problem, data type and availability of time and financial constraints for the researchers (Daniel and Sam, 2011).

For this research, secondary data collection method is used which is collected with the help of third part. Obtaining the necessary data required a specialized approach due to its niche nature. Unlike data readily available on social platforms or easily gathered through questionnaires, the specific sailing data needed for the research was not widely accessible. Consequently, the primary instrument utilized was third-party providers who regularly publish sailing data. These third-party providers specialize in collecting, processing, and disseminating maritime data, including information on vessel movements, routes, speeds, and other relevant parameters. By relying on these providers, the research ensured access to comprehensive and reliable data sets that were essential for the analysis.

3.9.1 Sampling

To conduct the research, it is not possible to include all the relevant population in the study hence, sampling is used. Sampling technique allows selecting a particular set from larger population for conducting research where the selected sample size showcases the whole population (Mangal and Mangal, 2013). Random sampling, non-probability sampling, stratified sampling, systematic sampling, etc. are the types of sampling techniques (Vehovar, et al., 2016).

For this study, non-probability sampling method is used, and different dataset is collected from various groups who are involved in shipping stuff by sea. For this purpose, ocean shipping companies, freight companies, ports, and places where ships dock are contacted. All this info helped to make good guesses about when ships will arrive at their destinations.

From the ocean carriers, we got the details like when ships are supposed to leave and arrive at different ports. Specifics about the ship's journey, such as - where it's stopping along the way, how long it'll be there, etc. are also considered. Additionally, it also checks out data about the routes ships take and the lanes they travel in the ocean.

Additionally, it is also focused to keep an eye on where the ships are in real-time and how fast they're going. This info comes from tracking systems used by ships. Knowing this helped to predict when they'll reach their destination more accurately. Additionally, proper gauge maintained on how busy ports are, which helped to understand if there are any big traffic jams or problems.

In essence, this study represents a concerted effort to harness the power of data-driven insights, transforming the landscape of maritime logistics with unprecedented precision and foresight. By aggregating, analyzing, and operationalizing data from a diverse array of stakeholders, it seeks to empower decision-makers across the shipment ecosystem, fostering a culture of adaptability, efficiency, and resilience in the face of dynamic logistical challenges.

3.10 Data Analysis

Data analysis can be defined as a method that is used in the research to analyze the raw data that is collected from the data collection method to interpret the research outcome effectively (Taylor, et al., 2006). Data analysis helps to apply statistical and logical techniques in a systematic manner so that data can be defined, abbreviated, summarized, interpreted and evaluated effectively to develop effective understanding about patterns and relationships.

In the context of undertaken research, descriptive data analysis method is selected as it has allowed gaining insights for effective decision-making and hypothesis testing which has further assisted to draw conclusions that are relevant to the research objectives (Opoku, et al., 2016). Descriptive data analysis can be defined as a statistical method that allows describing the basic features of the data while providing simple summaries about the sample and the measures (Singh, 2006). Utilization of this method in this study has supported to use graphical representations to organize and simplify large amounts of data and make it easier to understand at a glance. It has supported to provide an overview of the data characteristics which has further supported to test the hypotheses and has allowed conducting inferential analysis.

This research under scrutiny adopts descriptive study framework to scrutinizing various parameters aimed at enhancing the accuracy of Estimated Time of Arrival (ETA) predictions. These enhancements are poised to revolutionize industry practices, offering tangible benefits such as time savings, cost reduction, and mitigation of carbon emissions. Through meticulous analysis and innovative methodologies, this research endeavors to carve out new pathways for optimizing ETA predictions, thereby contributing to both academic scholarship and practical industry applications.

The investigation undertaken for this study will be of descriptive nature. This research has attempted to explore and explain the themes and provide additional insights into the Global logistics specially through Ocean moves, through the creation of a theoretical framework.

The research methodology employed in this study revolves around a comprehensive analysis of current sailing data, involving several key stages to extract meaningful insights and develop predictive models with high accuracy –

Observation of Sailing Data: This research starts with the observation of sailing data obtained from third-party providers. This data encompasses a wide range of variables, including vessel positions, speeds, routes, weather conditions, and other relevant parameters. By closely examining this data, the researcher aims to gain an understanding of the dynamics and patterns inherent in maritime transportation.

Understanding Nuances and Patterns: Through meticulous analysis, this researcher identifies nuanced patterns within the sailing data. This involves detecting trends, correlations, and anomalies that may impact vessel movements and arrival times. Understanding these nuances is crucial for developing accurate predictive models that account for various factors influencing maritime operations.

Outlier Detection and Cleansing: Outliers, or data points that deviate significantly from the norm, can distort analysis and compromise the accuracy of predictive models. To address this, outlier detection techniques are employed to identify and cleanse the data of anomalies. This process ensures that the dataset used for model development is reliable and representative of typical sailing conditions.

Model Building: Once the data has been cleansed, predictive models built based on the least error method. This involves selecting an appropriate modeling approach, such as regression analysis, machine learning algorithms, or time series forecasting methods. The goal is to develop a model that minimizes prediction errors and accurately estimates vessel arrival times based on relevant input variables.

Training and Testing the Model: The developed model is then trained using a subset of the data, with the remaining portion reserved for testing. During the training phase, the model learns from historical data patterns and relationships between variables. Subsequently, the model's performance is evaluated using the test dataset to assess its predictive accuracy and generalization capabilities.

Final Predictions and Validation: With a trained and validated model in hand, final predictions of vessel arrival times is produced with confidence. These predictions is validated against real-world observations to ensure their accuracy and reliability. Any discrepancies or inaccuracies is addressed through iterative refinement of the model or adjustments to input variables.

This methodology enables the researcher to uncover valuable insights, and improved accuracy for the ETA predictions.

3.11 Research Ethics

Research ethics govern the code of conduct to lead the research ethically. It enables the researchers to understand the importance of adhering fundamental moral and ethical principles as it allows ensuring integrity in the overall research process (Thomas, 2021). It supports to protect the rights, and dignity of the participants involved and further allows promoting accuracy while eliminating the chances of deception (Dooly, et al., 2017). Following research ethics in the study helps to foster trust, accountability, and reliability of the findings and also support to eliminate legal obstacles while submitting the study.

While conducting the literature review, it is focused to write the collected content in own words that were referenced from the past studies to eliminate the chances of plagiarism. Additionally, proper in-text citation and referencing is done in the study to give respect to the researchers for their valuable findings to eliminate the chances of copyright and patent (Thomas, 2021).

The robustness of a research study hinges upon its ability to withstand scrutiny in terms of reliability and validity. As highlighted by Shoaib and Mujtaba in their 2016 work, ensuring dependability, transferability, credibility, and conformability is crucial for establishing the quality of research (Shoaib & Mujtaba, 2016).

In the context of this quantitative analysis, reliability and validity are paramount objectives. Achieving these aims entails ensuring the trustworthiness and cleanliness of the data, as well as selecting the most appropriate model for the given dataset. To this end, several methodologies are employed, including transcript analysis, triangulation, and verification.

Transcript analysis involves meticulously examining and interpreting the data transcripts to identify patterns, trends, and discrepancies. This process helps ensure the accuracy and consistency of the data under examination. Triangulation, on the other hand, involves cross-referencing multiple sources of data or methodologies to validate findings and enhance the credibility of the research outcomes (Hasheminia et al., 2017). By corroborating information from different sources or employing diverse analytical approaches, the researcher guards against biases and strengthens the reliability of the results.

In evaluating the effectiveness of the proposed approach for predicting dynamic ETAs, the researcher conducts comparative analyses against ETAs predicted by carriers and actual arrival times. By comparing the predicted ETAs generated through the proposed model with those provided by carriers and benchmarking them against actual arrival times, the researcher assesses the predictive accuracy and reliability of the model.

The evaluation aims to quantify the improvement in prediction accuracies achieved with the proposed model compared to existing methods (Hasheminia et al., 2017). By demonstrating the superiority of the proposed approach in predicting ETAs with greater precision and reliability, the research establishes its credibility and contributes to advancing knowledge and practices in the field of maritime logistics.

Overall, by adhering to principles of reliability and validity and employing rigorous methodologies for data analysis and model selection, the research endeavors to produce robust and credible findings that withstand scrutiny and contribute to the advancement of knowledge in the domain of ETA prediction for vessels.

3.12 Research Design Limitations

While this research focuses on enhancing the accuracy of Estimated Time of Arrival (ETA) predictions for vessels and the shipments they carry, a significant challenge lies in accurately predicting the Estimated Time of Departure (ETD) from ports. This involves forecasting the turnaround time of a vessel after it arrives at a port and estimating the duration until its departure. However, numerous complexities and uncertainties make accurate ETD predictions a formidable challenge. Several factors contribute to the unpredictability of ETDs such as:

Container Positioning: The arrangement and handling of containers on board vessels can vary, impacting the efficiency of unloading and reloading operations. The precise positioning of containers within the vessel affects the time required for cargo handling processes (Yoon et al., 2023).

Customs Procedures: Custom holds and releases can significantly delay vessel departures. Compliance with customs regulations, inspections, and documentation processes adds uncertainty to the turnaround time at ports (Yoon et al., 2023).

Terminal Congestion: Congestion at port terminals can lead to delays in vessel operations. Limited berth availability, inefficiencies in cargo handling, and high volumes of inbound and outbound traffic contribute to terminal congestion, affecting ETDs (Yoon et al., 2023).

Yard Congestion: Congestion in container yards, where containers are stored and transferred between vessels and other transportation modes, further complicates ETD predictions. Limited space, inefficient logistics operations, and challenges in container handling contribute to yard congestion (Yoon et al., 2023).

Other Delays: Various unforeseen factors, such as mechanical issues, weather disruptions, labor strikes, or regulatory changes, can cause delays in vessel departures.

Given the multitude of unknowns and uncertainties inherent in port operations, accurately predicting vessel departure times is a daunting task. As a result, the ETD prediction aspect falls beyond the scope of this research focused on refining ETA predictions.

Addressing the challenges associated with ETD predictions would require comprehensive data integration, advanced predictive modeling techniques, and collaboration among stakeholders across the maritime supply chain. While this remains a critical area for improvement in maritime logistics, it presents a distinct and complex research problem that warrants dedicated attention and resources.

3.13 Conclusion

In this chapter, the research method used to conduct allover study is discussed while providing proper justification behind selecting particular method. From the above discussion, it can be summarized that to conduct this research, positivism philosophy, deductive research approach, action research strategy, cross-sectional time horizon, non-probability sampling method and descriptive data analysis method. Quantitative research design is used to comprehensively address the research objectives (Bhattacharya et al., 2006). Quantitative research design proves beneficial when the phenomenon under investigation pertains to observable data readily available in secondary sources. The research instruments employed in this study primarily involved gathering sailing schedules from carrier portals and compiling actual/predicted data for specified trade lanes.

Once the requisite data was gathered and cleansed to ensure its reliability and integrity, the researcher proceeded to develop two distinct models. These models were designed to calculate both travel time and wait time experienced by vessels at terminals along the specified trade lanes. Subsequently, these models were leveraged to predict Estimated Time of Arrivals (ETAs) for selected lanes.

The accuracy of the ETA predictions generated through the developed models was then systematically evaluated and compared against existing methods (Bhattacharya et al., 2006). Specifically, the researcher compared the predicted ETAs with those provided by carriers to ascertain the efficacy and reliability of the proposed approach. By conducting this comparative analysis, the researcher aimed to validate the predictive capabilities of the models and identify any potential areas for improvement.

In the final stages of the study, the researcher meticulously articulated the limitations inherent in the research methodology and findings.

By openly acknowledging these limitations, the researcher ensured transparency and integrity in the research process, thereby enhancing the credibility and validity of the study's conclusions.

Through systematic data collection, model development, prediction analysis, and reflection on limitations, the study contributes to advancing understanding and knowledge within the domain of maritime logistics and ETA prediction methodologies.

CHAPTER IV:

RESULTS

4.1 Introduction

This chapter of the research is focused towards producing results from the raw data and because of the results, it is focused towards discussing the findings. In regard to this, initially, this chapter explains the prediction models (Travel Time Prediction Model and Delay Prediction Model) that are used to calculate the ETA. Further, additional data layers (Weather Layer and Traffic Layer) are discussed that are used to improve the accuracy of predictions of the data. After that, calculation of dynamic ETA and Static ETA has taken place with the help of third-party data for (October, November and December 2023). Further, study is moved forward towards predicting the accuracy by evaluating the data that are collected from different sources and accordingly personal intelligence is integrated along with the historical predictions.

From the above research finding, it can be concluded that it remained assistive to achieve the research objectives at the greater extent and to answer the research questions effectively. It can be further discussed below:

4.2 Research Question One

What are the primary reasons for shipment delays in global intermodal transport networks?

From the research, it can be concluded that there are multiple reasons behind the shipment delays in global intermodal transport networks. In regard to this, it is identified that weather and traffic are major reasons behind the delay. Focusing on external factors such as weather conditions, port congestions, and traffic can remain highly assistive to improve the accuracy of Ocean ETA predictions (Bodunov et al., 2018). The reason behind it is that weather at sea, including factors like wind speed, wave height, and storm patterns, can directly influence the speed and efficiency of maritime transport. Adverse weather conditions may slow down vessels or even necessitate route changes, affecting their ETA predictions unless these conditions are accounted for. Similarly, congestion at ports can cause delays in loading and unloading cargo, as well as in berthing and departure times for vessels. Factors contributing to port congestion include high cargo volumes, limited infrastructure, labor strikes, and administrative procedures. Incorporating port congestion data into ETA predictions can account for potential delays in vessel operations.

Additionally, traffic congestion along shipping routes, such as busy shipping lanes or congested waterways near ports, can impact vessel transit times. Vessels may experience slowdowns or detours to avoid congested areas, affecting their ETA. Moreover, traffic conditions at ports, including vessel queues and navigational restrictions, can influence the arrival and departure times of ships (Fetzer et al., 2018). By considering these external factors it can be suggested that more accurate and reliable estimates of arrival times can be achieved. This would be particularly beneficial for various stakeholders involved in maritime logistics, including shipping companies, port authorities, cargo owners, and logistics providers, as it would allow for better planning and optimization of operations.

On the other hand, this research has also supported to conclude that there are some situations where external factor such as weather conditions, port congestions, and traffic may have minimal impact on Ocean ETA predictions. It posits that the variability in arrival times primarily stems from factors internal to the shipping operations, such as vessel speed, route planning, and operational efficiency, rather than external conditions. The complexity introduced by incorporating external factors into ETA prediction models may not result in proportionate improvements in accuracy. It suggests that the additional data and computational resources required to account for weather conditions, port congestions, and traffic may not yield significant enhancements in prediction performance compared to simpler models that rely solely on historical voyage data and basic scheduling algorithms (Klimberg et al., 2010). Consideration of data quality and uncertainty associated with external factors. It suggests that the available data on weather conditions, port congestions, and traffic may be insufficient or unreliable for accurate prediction modelling. Inaccuracies in data collection, measurement errors, and uncertainties in forecasting future conditions could undermine the effectiveness of incorporating these external factors into ETA predictions. Despite the intuitive appeal of considering external factors in Ocean ETA predictions, there may be limitations and challenges that prevent meaningful improvements in prediction accuracy (Klimberg et al., 2010). For this purpose, rigorous statistical analysis has taken place while comparing the prediction models with and without the inclusion of external factors to assess whether the observed differences in accuracy are statistically significant or simply due to random chance.

4.3 Research Question Two

How can the visibility of planned and active shipments be improved to enhance ETA prediction accuracy?

This research has supported to identify the difference between planned shipments and active shipments and how delay takes place and how time of the shipment delivery get increased due to various factors. It is discussed with the help of travel-time model and delay model. For this purpose, AIS data is used to track the shipment in real time and also to observe that the shipment is following the planned route and schedule or not or it is getting off track in terms of location as well as time.

Improving the visibility of planned and active shipments is crucial for enhancing Estimated Time of Arrival (ETA) prediction accuracy in global supply chains. Implementing real-time tracking technologies such as GPS, RFID, or IoT sensors to monitor the location, status, and condition of shipments throughout their journey help in the cause. These technologies provide continuous visibility into the movement of goods and enable stakeholders to track shipments in transit accurately. Freight Forwarders should also integrate data from various sources, including transportation carriers, logistics providers, suppliers, and customers, into a centralized platform or system. By consolidating shipment data in one place, stakeholders can access real-time information about planned and active shipments, enabling better coordination and decision-making (Fancello et al., 2011). They should also establish collaborative information-sharing platforms or networks that allow stakeholders across the supply chain to share relevant data and updates about planned and active shipments. By sharing information about shipment schedules, delays, or disruptions, stakeholders can work together to address issues proactively and minimize the impact on ETA predictions.

Leveraging predictive analytics and machine learning algorithms to analyze historical shipment data, identify patterns, and predict potential delays or disruptions in transit help improving the prediction accuracy for both Planned and Active shipments. By analyzing factors such as weather conditions, traffic patterns, carrier performance, and historical transit times, these algorithms can generate more accurate ETA predictions and provide early warnings about potential issues (Fancello et al., 2011). On top of it, they can implement dynamic routing and scheduling algorithms that can adjust shipment routes and schedules in real-time based on changing conditions or priorities. By dynamically rerouting shipments to avoid congestion, optimize transit times, and minimize delays, these algorithms can improve ETA prediction accuracy and ensure on-time delivery.

Implementing exception management systems that automatically detect deviations from planned shipment schedules or predefined performance thresholds and trigger alerts to relevant stakeholders helps the Freight Forwarder to focus only on those shipments where there are issues (Hasheminia et al., 2017). By proactively identifying potential delays or disruptions, these systems enable stakeholders to take corrective actions promptly and mitigate the impact on ETA predictions. Leading players in this market foster collaboration and communication among supply chain partners to improve forecasting accuracy and reduce uncertainties in shipment planning. By sharing demand forecasts, production schedules, and inventory levels, stakeholders can better anticipate future shipment volumes and proactively adjust transportation capacity to meet demand.

4.4 Research Question Three

What methodologies are currently employed by shipment parties to calculate ETAs, and what are their limitations?

From the research finding, it can be concluded that the data collected from the third-party (SPIRE Global) is used to calculate ETA for 3 months (October, November, December – 2023) which is further used with actual ETA to predict error mean values with the help of Err_Mean_Model. It has not only allowed to evaluate the error mean values of the prediction of ocean model but has also supported to compare it with carrier model. It has supported to identify that in most of the cases, ocean model has shown more accuracy as compared to carrier model however, there were some situations where carrier model is found performing better as compared to ocean model. It has further supported to come up with hybrid approach which is the novelty that is added by the researcher in this field.

The calculation of Estimated Time of Arrival (ETA) in shipment logistics involves a range of methodologies, each leveraging different technologies and data sources to provide accurate predictions. The main methodologies currently employed, along with their limitations and explored in this research are as follows:

4.4.1 Historical Data Analysis

This involves analyzing past shipment data to predict future delivery times. Historical data on routes, weather patterns, traffic conditions, and delivery times are used to create predictive models. The limitations being as follows:

Static Nature - Historical data can become outdated quickly, especially with changing traffic patterns, new regulations, or infrastructure changes.

Lack of Real-Time Adjustments - This method does not account for real-time variables that can affect delivery times.

4.4.2 GPS and Real-Time Tracking

GPS technology allows for real-time tracking of shipment vehicles, providing up-to-the-minute location data. This data is used to calculate ETAs based on current position, speed, and known route information. The limitations being as follows:

Signal Interference - GPS signals can be disrupted by urban environments, weather conditions, or other factors, leading to inaccuracies.

Dynamic Variables - Real-time tracking needs to be coupled with dynamic data (like real-time traffic updates) to be fully effective, which can be complex and resource intensive.

4.4.3 Machine Learning and AI

Advanced algorithms and AI models analyze a wide array of data points, including historical data, real-time tracking information, weather conditions, and traffic patterns, to predict ETAs (Klimberg et al., 2010). These systems can learn and improve over time. The limitations being as follows:

Data Quality and Volume - Machine learning models require vast amounts of high-quality data to train effectively. Inaccurate or insufficient data can lead to poor predictions.

Complexity - Implementing and maintaining AI systems can be complex and costly, requiring specialized knowledge and continuous monitoring.

4.4.4 Traffic and Weather Data Integration

Integrating real-time traffic data and weather forecasts into ETA calculations helps in adjusting predictions based on current road conditions and weather patterns (Klimberg et al., 2010). The limitations being as follows:

Data Availability - Access to accurate and up-to-date traffic and weather data can be inconsistent, particularly in remote or less-developed regions.

Integration Challenges - Combining multiple data sources in real-time and ensuring their compatibility and reliability is technically challenging.

4.4.5 Carrier and Route-Specific Algorithms

Customized algorithms are developed based on the specific carrier's operational characteristics and commonly used routes. These take into account factors like typical loading/unloading times, stop frequencies, and driver behavior (Klimberg et al., 2010). The limitations being as follows:

Scalability - Custom algorithms may not be easily scalable or applicable to different carriers or routes.

Flexibility - These models can lack flexibility if significant changes occur in the logistics network or operational practices.

4.4.6 Crowdsourced Data

Utilizing data from crowdsourced platforms (e.g., Waze, Google Maps) to get real-time traffic and route information from other users on the road. The limitations being as follows:

Data Reliability - Crowdsourced data can vary in accuracy and reliability depending on the user base and participation level.

Coverage - Effectiveness depends on the volume of users in a given area; rural or less-traveled areas may lack sufficient data points.

4.4.7 Summary of Limitations

- Data Quality and Availability - All methods rely on accurate, up-to-date data. Poor data quality can significantly affect ETA accuracy.
- Integration Complexity - Combining various data sources and technologies can be technically challenging and resource intensive.
- Adaptability - Methods that do not adapt in real-time or learn from new data can become outdated quickly.

Overall, while each methodology has its strengths, the limitations highlight the need for continuous improvement in data collection, integration, and analysis to enhance the accuracy of ETA predictions in shipment logistics.

4.5 Research Question Four

How can we leverage data intelligence, carrier inputs, and external factors such as weather and traffic conditions to improve the accuracy of ETA predictions?

From the research finding, it is identified that there were some situations such as Yokohama Port in Japan to the Port of Shanghai in China and Yantian Port in China to Ningbo Port in China, where the ETA calculated by ocean model have shown better outcomes by representing lower error rate as compared to carrier model and there was a situation (Port Yantian to Port of Los Angeles) where carrier model has shown less error as compared to ocean model. However, it is also identified that there was a situation where in some days, ocean and in some days carrier model was showcasing less error. In that kind of situation, it can become confusing that use of which model can remain relevant. Hence, this research has supported to conclude that to deal with such kind of situation, utilization of hybrid approach can remain relevant as it offers the potential to increase the relevancy of the research outcome.

Improving the accuracy of ETA predictions in shipment logistics involves the integration of various data sources and leveraging advanced data intelligence techniques. Here's how data intelligence, carrier inputs, and external factors like weather and traffic conditions can be combined effectively:

4.5.1 Data Intelligence

Machine Learning Algorithms - Implemented machine learning models that can process vast amounts of data, identify patterns, and make predictions. These models were trained on historical data and continuously updated with new data to improve accuracy.

Big Data Analytics - Usage of big data platforms to aggregate and analyze data from various sources, including historical shipment data, real-time tracking data, and external data sources. This helped identify trends and anomalies that affect ETAs.

Predictive Analytics - Developed predictive analytics models that consider multiple variables simultaneously, allowing for more accurate forecasting. These models predict potential delays and suggest alternative routes or schedules.

4.5.2 Carrier Inputs

Operational Data - Collected detailed operational data from carriers, including loading and unloading times, driver schedules, stop frequencies, and typical route deviations. This data provided insights into the carrier-specific factors that affect ETAs.

Driver / Captain Feedback - Incorporated real-time feedback from drivers regarding route conditions, delays, and other on-the-ground / in-the-sea factors. Driver and Captain apps used to report and update conditions in real-time.

Fleet Management Systems - Integrated data from fleet management systems that track vehicle health, fuel consumption, and driver behavior. This helped in anticipating delays due to vehicle issues or inefficiencies.

4.5.3 External Factors

Weather Data Integration - Used weather APIs to integrate real-time weather forecasts and historical weather data into the ETA calculation models. This helped predict delays due to adverse weather conditions.

Traffic Data - Integrated real-time traffic data from sources like GPS systems, traffic monitoring services (e.g., Google Maps, Waze), and local traffic authorities. Traffic congestion, accidents, and roadworks significantly impact delivery times.

Ocean Conditions - Included data on ocean conditions, such as construction zones, canal closures, and detours. This data is sourced from government databases and real-time reporting tools.

4.5.4 Data Collection and Integration

Data Aggregation Platforms - Developed platforms that aggregate data from various sources, ensuring they are standardized and interoperable. This enabled seamless integration and analysis.

APIs and IoT Devices - Used APIs to pull data from external sources (e.g., weather, traffic) and IoT devices for real-time tracking and monitoring of shipments and vehicle conditions.

4.5.5 Advanced Analytics and AI

Real-Time Data Processing - Implemented systems capable of processing data in real-time to provide up-to-date ETAs. Stream processing frameworks like Apache Kafka and Apache Flink were found useful.

AI-Driven Adjustments - Used AI to make real-time adjustments to ETAs based on changing conditions. For instance, if a port congestion is detected, the system can reroute the shipment and update the ETA accordingly.

4.5.6 Feedback Loops

Continuous Learning - Implemented feedback loops where the system learns from actual delivery times versus predicted ETAs, adjusting models to improve future predictions.

Driver and Customer Feedback - Collected feedback from drivers, captains and customers regarding delivery times and experiences to refine prediction models.

4.5.7 Visualization and Communication

Dashboards - Dashboards can be created for logistics managers that visualize real-time data and predictions, highlighting potential delays and suggesting proactive measures.

Automated Notifications - Similarly, automated notifications can be set up to inform stakeholders (drivers, customers, logistics managers) of any changes in ETA or potential delays.

4.5.8 Challenges and Solutions

Data Quality and Reliability - Ensured data accuracy and completeness by validating and cross-referencing data from multiple sources. Implemented data cleaning and preprocessing steps to handle inconsistencies.

Scalability - Developed scalable systems that can handle increasing data volumes and complexity. Cloud-based solutions offer the necessary scalability and flexibility.

Integration Complexity - Used middleware and data integration tools to streamline the integration process and ensure seamless data flow between different systems.

By leveraging data intelligence, carrier inputs, and external factors in a cohesive and integrated manner, the researcher could significantly improve the accuracy of estimated predictions of Vessel Arrival, leading to more efficient operations and higher customer satisfaction.

4.6 Summary of Findings

Based on collected data, comparative data analysis has taken place between the model predicted ETA and carrier ETA. This comparative analysis of the results has supported to evaluate the error mean value of the ETA calculated to predict effective way for the arrival of the vessel which has allowed discussing that utilization of which method is appropriate for predicting ETA. Additionally, it has also supported to discuss that in which situation utilization of hybrid approach is effective. Based on overall results, findings are discussed. Findings are further correlated with the literature review to increase the reliability of the result outcome.

4.7 Conclusion

Above research helps to conclude that to conduct this research, quantitative data collection method is used which has supported to represent the carrier model data and ocean model data effectively. Utilization of Saunder's research onion model has supported to conduct the overall research in a step-by-step manner. Giving concern towards ethical aspect has supported to lead this overall research in ethical manner to eliminate the chances of any kind of hindrance during research submission. Additionally, it has supported to conduct comparative analysis of error in ETA in both the models. Further, it has allowed identifying the model that is required to be implemented in different situations. Use of effective tables and graphs have supported to represent the collected data in a statistical manner which has supported to understand the research outcome in a glance.

CHAPTER V:

DISCUSSION

5.1 Discussion of Results

It is important to touch upon a typical Logistics End to End Workflow to understand at which stage this research comes into play (Port of Antwerp, 2020):

- Demand and supply plans are made for upcoming seasons.
- Buyer places a Purchase Order to buy inventory.
- Purchase Order is received by the shipper. Shipper responds with a Sales Order and confirms the Order for the same inventory.
- Shipper allocates the Raw materials / parts from various Warehouses and Sources them to its Assembly Plant.
- Shipper creates a Work Order to assemble the parts and create the Finished Goods. These finished goods are packed and picked.
- A Shipment is created to fulfil the initial Purchase Order. This shipment is accepted by **a** Logistics Service Provider that hires a first mile Truck Load or Less than Truck Load move (from shipper DC to nearest port), an Ocean move (from the origin Port to a Port in buyer's country) and a last mile Truck Load or Less than Truck Load move (from the destination Port to the buyer's DC).
- When Shipment is planned the Shipment parties can see its estimated time of arrival and make sure that it is within the Promised Delivery Date.
- When Shipment is on the move (active shipment) it is tracked and in case of any delays, the ETAs get updated, and the Alerts get generated. This is where this research

comes into play to predict the estimated time of arrivals. Focus of this research will be only on the Ocean moves which will help the end users to know the changes in final ETA at the destination as well.

5.2 Discussion of Research Question One

The first question that the researcher started on this journey was – "What are the primary reasons for shipment delays in global intermodal transport networks?". Literature clearly says that one of the major reasons for shipment delays is Port Congestions. Ports are critical nodes in the global supply chain, where goods are transferred between different modes of transportation. Congestion at ports can occur due to factors such as high cargo volumes, labor shortages, inefficient operations, or delays in customs clearance. As a result, ships may have to wait in line to unload or load cargo, leading to delays in transit times (Yoon et al., 2023).

The other reason often seen for shipment delays is Vessel Schedule Disruption. Changes in vessel schedules due to mechanical failures, maintenance issues, labor disputes, or unexpected events like accidents can disrupt the planned transportation routes and timings (Yoon et al., 2023). Even minor delays in departure or arrival times can have ripple effects across the supply chain, impacting the timely delivery of goods to their destination.

Then there are Customs Clearance Delays. International shipments often require customs clearance procedures to ensure compliance with import/export regulations, tariffs, and duties. Delays in customs clearance can occur due to documentation errors, incomplete paperwork, discrepancies in declared values, or regulatory changes (Port of Antwerp, 2020). These delays can hold up shipments at border crossings or ports of entry until customs procedures are completed. Predicting custom clearances are next to impossible as there are many parameters in play. Custom clearances can sometimes happen a day before arrival of the shipment or sometimes 2 weeks after the shipment arrival at the international ports.

Supply Chain Disruptions also cause a lot of delays and hinder accurate ETA predictions. Disruptions in the upstream or downstream supply chain, such as production delays, supplier issues, distribution bottlenecks, or inventory shortages, can impact the availability and timely delivery of goods. These disruptions can cascade through the supply chain, leading to delays in shipments as companies struggle to overcome challenges and restore normal operations.

5.3 Discussion of Research Question Two

The second question that the researcher wanted to explore in this journey was – "How can the visibility of planned and active shipments be improved to enhance ETA prediction accuracy?"

5.3.1 Planned Shipments:

Before the shipment starts its Ocean journey on a Vessel, it is just a Planned shipment. Based upon the historical data of the shipment journey between an origin and destination, the researcher derived –

1. What route would a shipment take in future and how many legs will it have?
    a. Are there any planned Transit ports?

b. Are there any unplanned Transit ports?

c. The entire Ocean journey is being carried out on the same Vessel or there are some Feeder Vessels and Mother Vessels are involved as well?

5.3.1.1 Feeder Vessel: A feeder vessel is a smaller ship designed to transport containers between smaller ports or terminals to larger hub ports where they can be transferred onto larger vessels for long-distance transport. Feeder vessels typically have shallower drafts and lower capacities compared to mother vessels. They serve to collect containers from various smaller ports within a region and consolidate them at a larger hub port for onward shipment to distant destinations. Feeder vessels play a crucial role in facilitating the distribution of cargo from regional ports to global trade routes, enhancing connectivity and efficiency in the supply chain (Port of Antwerp, 2020).

5.3.1.2 Mother Vessel: A mother vessel, also known as a mainline vessel or ocean-going vessel, is a large container ship designed to carry vast quantities of cargo over long distances between major ports around the world. These vessels are typically massive in size, with high container capacity and deep drafts, enabling them to navigate deep-sea routes and traverse oceans efficiently. Mother vessels serve as the backbone of international maritime trade, transporting containers between major hubs and facilitating the global movement of goods on a large scale. They often call at key ports to load and unload containers, acting as pivotal nodes in the global container shipping network (Port of Antwerp, 2020).

2. What would be the time taken to complete those legs and the milestones (planned time)?
    a. Historical data analysis and cleansing.
    b. Model selection based upon least error method.
    c. Model training.
    d. Model testing.
    e. Final prediction using the chosen model.

5.3.2 Active Shipments

When the shipment journey starts, the researcher used the AIS (Automatic Identification System) data for the container that is carrying this shipment. Using the AIS data that is readily available for free via various opensource platforms such as Open-AIS, SPIRE global, Kpler.com etc., the researcher tracks the shipment in real time and observe if the shipment is following the planned route and schedule or is getting off track in terms of its location and time.

Researcher built two types of models here - one purely for travel time that is once the Vessel departs from Origin port and it is on the way, how much more time it will take to arrive to the anchorage of its destination port. And second for the delay that happens from anchorage to the berthing at the terminal.

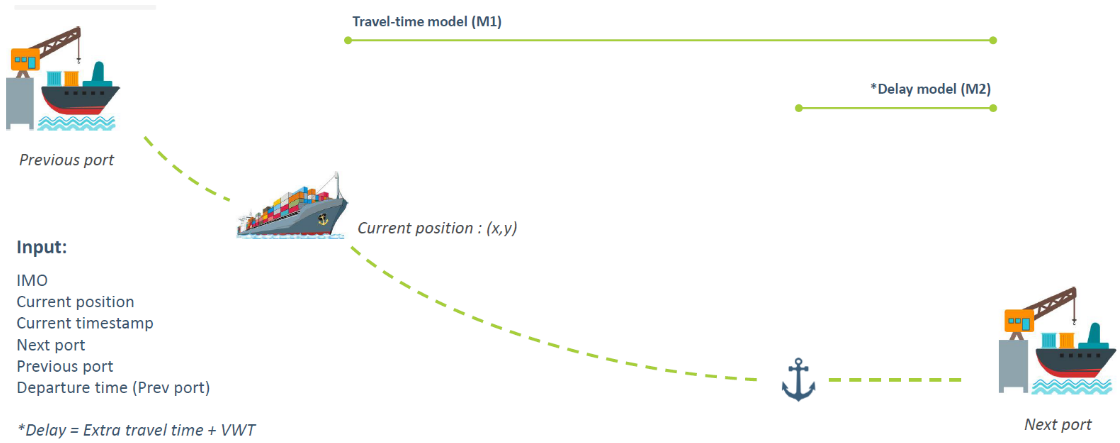


Figure 7: Travel Time and Delay Time Model
(Source: Created by Learner)

5.4 Discussion of Research Question Three

The third question that the researcher wanted to answer through this research was – "What methodologies are currently employed by shipment parties to calculate ETAs, and what are their limitations?". For this purpose, the researcher came up with some of his own Prediction models which adds novelty to this area. Prediction models for ETA in global intermodal transport networks are sophisticated tools which are designed to forecast the arrival times of goods moving through transportation mode (Abdi and Amrit, 2024). For this purpose, algorithms and machine learning techniques are required to utilize to analyze historical and real-time data which helps to enhance accuracy in predicting delays or early arrivals. Effective prediction is crucial as it helps to optimize supply chain efficiency, reduce costs, and improves service reliability in the interconnected and dynamic environment of global intermodal transport. In regards to this, it is identified that there are majorly two models that are essential to consider to get the accurate predictions as they mostly work in conjunction. Above finding has shown similarity with the findings of Zhang, et al. (2019) that contemporary organizations have increased concern towards integrating the multimodal transportation management systems to optimize the ship routing as it can assist in exploring various opportunities and can assist in creating unified TMS framework.

Regarding this, following models are built by the researcher -

5.4.1 Travel Time Prediction Model

Travel time prediction model is used to estimate the ETA in the global intermodal transport networks of a vehicle to travel from a particular point to another while considering current and historical traffic data (Hardij, 2018). Analysis of patterns and real-time inputs help to lead accurate forecasts that are essential for route planning, navigation systems, and traffic management. Use of this model helps to enhance the efficiency of transportation systems and to improve the commuting experience.

Regarding undertaken research, when the Vessel is on the move, it has already completed part of its journey and due to this reason, actuals are already known till that point which is the current position (latitude, longitude being x and y). However, it is required to calculate how much time will it take to reach the anchorage of the next port. For this purpose, following parameters are used:

5.4.1.1 Inputs

1. IMO Number: Works as a Ship identification number provided by International Maritime Organisation.

2. Current position: Tells what the Vessel's current coordinates (latitude, longitude) in the sea are.
3. Current time stamp: Tells when the Vessel was at the above coordinates.
4. Previous port: Tells where the Vessel started its journey and more importantly which was the last port of call.
5. Departure time from the previous port: Tells when the Vessel left the last Port of call.
6. Next port: Tells what is the next Port that the Vessel is going to and how far is that from the current position of the Vessel.

All these parameters are available in the AIS inputs that is sourced from the third-party open sources like Spire Global.

5.4.2 Delay Prediction Model

Once the Travel time is calculated, i.e. how much of the Ocean journey is complete, how much more is left and how much time the Vessel will take to arrive to the destination port, it is important to know how much time it will take from arrival at anchorage (berthing) to actual arrival at the terminal for loading / unloading of the containers. This gets calculated by the second model that is the Delay Prediction Model. Utilization of this model helps to analyze the historical data on workflows, schedules, and disruptions so that patterns and potential bottlenecks can be identified that has the potential to cause future delays (Zhang, et al., 2024). Incorporation of real-time data with delay prediction model helps to enhance the predictive accuracy. It helps to anticipate delays so that organizations can proactively manage resources, adjust schedules, and communicate more effectively with the stakeholders while minimizing the impact of delays on operations and improving overall efficiency.

To calculate that, the primary input that goes into the model is ‘Port congestion’. Using the data from the terminals and the Vessel turnaround times for the previous vessels on the same terminal in last 24 hours will provide an average wait time / delay time at the terminal, using which in combination of the Travel time, the exact prediction of estimated arrival at the destination port can be calculated.

On top of the above two models (Travel Time Prediction Model and Delay Prediction Model), there have been few additional data layers too that are essential to consider increasing the accuracy of predictions. For this purpose, weather layer and traffic layer are also considered. Under each of these layer, multiple aspects are considered. They are described as follow:

5.4.2.1 Weather Layer

The researcher would also use the weather of the lanes and locations where the shipment is passing through and predict its effect on the shipment timeliness. Adding a weather layer to estimated arrival predictions for vessels can significantly improve the accuracy of these predictions by considering environmental factors that can affect vessel speed and performance. Views from literature review have also shown similarity with the above discussion that algorithms are required to be considered as there is a requirement of forecasting ETAs in a manner that it can assist to deal with the climatic challenges so that ship routing decisions can be improved (Johnson et al., 2017).

Analyzing meteorological data to anticipate weather patterns and their impact on sea conditions. This includes considerations for wind speed, wave height, and storm predictions.

- **Wind and Currents:** Weather data provides information on wind patterns and ocean currents, which have a significant impact on vessel speed and fuel consumption. By factoring in wind direction and strength, as well as current speed and direction, predictive models can more accurately estimate how these elements affect a vessel's progress along its route (Johnson et al., 2017).
- **Sea State:** Weather data includes information on sea state, such as wave height and direction. Rough seas can slow down vessels and increase fuel consumption, while favourable sea conditions can improve speed and efficiency. Incorporating sea state data allows predictive models to adjust arrival estimates accordingly (Johnson et al., 2017).
- **Visibility:** Poor visibility due to fog, rain, or other weather conditions can affect navigation and require vessels to slow down or alter their course for safety reasons. By considering visibility conditions along the vessel's route, arrival predictions can account for potential delays caused by reduced speed or route deviations (Johnson et al., 2017).
- **Temperature and Humidity:** Extreme temperatures and humidity levels can impact engine performance and may require adjustments to operating conditions. Including temperature and humidity data in arrival predictions helps account for

these effects and provides a more accurate assessment of the vessel's expected arrival time (Johnson et al., 2017).

This weather data received from SPIRE Global which are further integrated into the Travel time model using machine learning algorithms to analyze historical weather patterns and their impact on vessel performance. By training models on past data, they can learn to make more accurate predictions about how future weather conditions will affect vessel arrivals.

Overall, incorporating a weather layer into estimated arrival predictions for vessels enables operators to make more informed decisions, optimize voyage planning, and better manage resources such as fuel and crew time (Kim, et al., 2023). Involvement of all these factors in ETA prediction can support to enhance cost savings, improved efficiency, and enhanced safety for maritime operations.

5.4.2.2 Traffic Layer

The researcher also used the traffic data of the origin, destination, and other port calls where the shipment is passing through and predict its effect on the shipment timeliness. Transportation disruption is one of the common sources of business interruptions that has the potential to cause significant economic loss to a lean supply chain (Paul, et al., 2019).

Integrating a traffic layer into estimated arrival predictions for vessels can offer several benefits such as:

- **Route Optimization:** Traffic data can help identify congested areas along the vessel's route, allowing for route adjustments to avoid delays. Just as drivers use real-time traffic information to choose the fastest route, vessels can navigate around heavy traffic areas to maintain their schedule (Paul, et al., 2019).

- **Port Congestion:** Traffic data can highlight port congestion, enabling operators to anticipate delays in docking or unloading/loading cargo. By factoring in port congestion, arrival predictions can be adjusted accordingly, providing more accurate estimates of arrival times (Paul, et al., 2019).
- **Navigational Safety:** Understanding vessel traffic in busy waterways is crucial for navigational safety. By incorporating traffic data, vessel operators can better plan their route to avoid collisions and navigate safely through areas with heavy maritime traffic (Paul, et al., 2019).
- **Berth Availability:** Traffic data can provide information on berth availability at ports, helping vessel operators coordinate arrival times with berth availability to minimize waiting times and optimize port operations (Paul, et al., 2019).
- **Dynamic Routing:** Real-time traffic data allows for dynamic routing, where vessels can adjust their course based on current traffic conditions to optimize their journey. This flexibility enables vessels to adapt to changing traffic patterns and minimize delays (Paul, et al., 2019).
- **Port Queuing:** Traffic data can reveal queues forming at ports, indicating potential delays in the docking process. By accounting for port queuing in arrival predictions, operators can better manage their schedules and resources (Paul, et al., 2019).

Incorporating a traffic layer into estimated arrival predictions enhances the prediction accuracy and help the vessel operators make more informed decisions to optimize their routes, minimize disruptions, and ensure timely arrivals at their destinations.

Both the weather and traffic layers are not yet used in this model. However, the researcher wants to convey that there is a good chance that if we use those, the prediction accuracy will improve further. Also mentioned the reasons why these could be a big factor in the ETA prediction.

5.4.2.3 ETA prediction Accuracy by Third-Party in the Last 3 Months

Mentioned data in the Appendixes 1, 2, and 3 are collected with the help of third-party (SPIRE Global) for October, November and December, 2023. It has supported to represent the average of accuracy of different ports. Further, accuracy of these ports is evaluated on T-2, T-5, T-10, T-20, T-30 and T-40 parameters.

On the basis of above evaluation, calculation of average accuracy & average error. has taken place and forecasting of Delta-T threshold has taken place on weekly basis.

calculated month	days bucket	Average of acc	Average of err(hours)
October	0-7 days	0.85847451	15.85415569
	7-14 days	0.790010601	27.26669622
	14-21 days	0.841490889	31.81409498
	21-28 days	0.873948408	33.2656356
	>28 days	0.831445594	35.55445908
November	0-7 days	0.859123002	15.58801334
	7-14 days	0.795650768	26.58394838
	14-21 days	0.845603169	30.91876799
	21-28 days	0.874422442	33.51794499
	>28 days	0.836858202	38.1636299
December	0-7 days	0.859767386	15.65108621
	7-14 days	0.788327744	27.59038864
	14-21 days	0.831048454	33.66940797
	21-28 days	0.858310627	36.12141235
	>28 days	0.848604651	35.62714729

Forecasting days before arrival at port	Delta-T thresholds	Points
0 to 7 days	Delta-T <= 12h	95%
	12h< Delta-T <= 24h	85%
	24h< Delta-T <= 48h	75%
	Delta-T > 48h	30%
7 to 14 days	Delta-T<= 16h	95%
	16h< Delta-T <= 24h	85%
	24h< Delta-T <= 48h	75%
	Delta-T > 48h	50%
14 to 21 days	Delta-T<= 24h	95%
	24h< Delta-T <= 48h	85%
	48h< Delta-T <= 72h	75%
	Delta-T > 72h	50%
21 to 28 days	Delta-T<= 48h	95%
	48h< Delta-T <= 72h	85%
	72h< Delta-T <= 96h	75%
	Delta-T > 96h	50%
28+ days	Delta-T<= 48h	95%
	48h< Delta-T <= 72h	85%
	72h< Delta-T <= 96h	75%
	Delta-T > 96h	50%

Table 1: Weekly Average Accuracy and Error
(Source: Created by Learner)

The detail of value calculation in step-by-step manner mentioned below:

- Days buckets are identified initially, dividing the timeframe into segments. In this case, the day buckets are: 0-7 days, 7-14 days, 14-21 days, and 21-28 days.
- For each day within a bucket, the error (err) is calculated as the difference between my model's ETA and the actual ETA. The average error is then computed for each day within the bucket. For Example, within the 0-7 days bucket, the error for each day from day 1 to day 7 is calculated, and then an average error is derived for the entire 0-7 days period.
- Further, the average accuracy of my model is determined:

- The average error calculated for each day within the bucket is compared to a predefined forecasting table. For example, if the average error for day 7 is less than or equal to 12 hours, it is assigned a score of 0.95.
- This process is repeated for each day within the bucket, and then the average accuracy score is computed for the entire bucket.

From the provided data, it is deduced that my model's ETA exhibits an accuracy of 85% for the 1-7 days, 14-21 days, and 21-28 days buckets. However, my model's ETA demonstrates an accuracy of less than 80% for the 7-14 days bucket.

5.4.3 Use of Travel Time Prediction Model and Delay Prediction Model to Predict Arrival Time Prediction

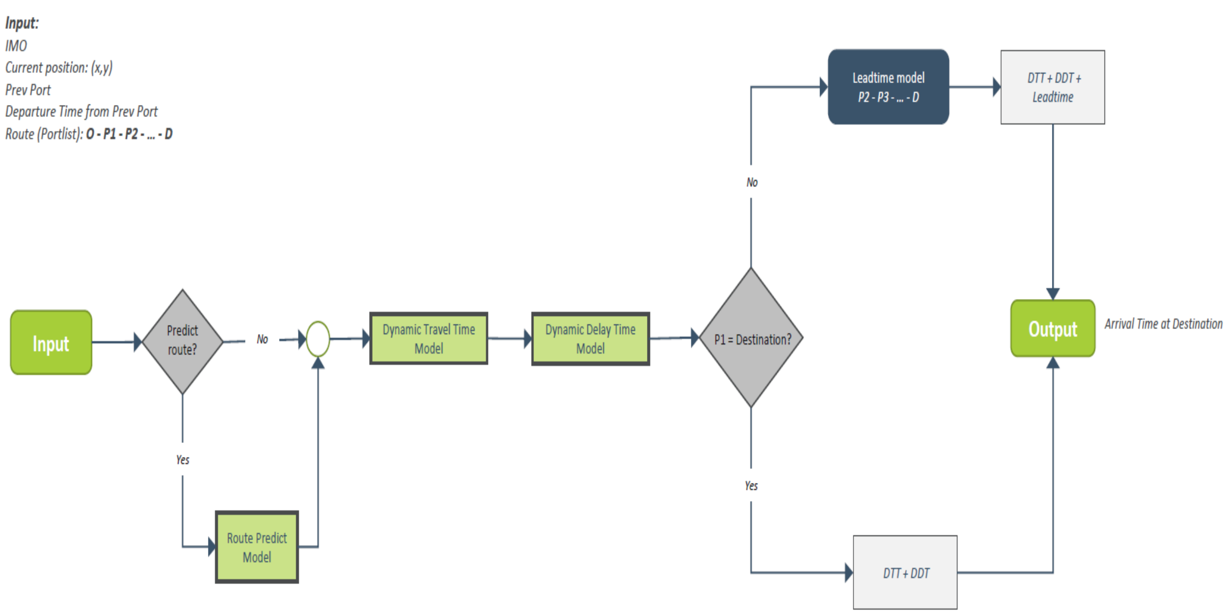


Figure 8: Calculating Appropriate Arrival Time
(Source: Created by learner)

Above flow chart represents how travel prediction model and delay prediction model is used to calculate appropriate arrival time. From this flow chart, it can be analyzed that to predict the arrival time at destination, there is a requirement of predicting the route on the basis of input. It reflects that do we know the route, if yes then there is no requirement of predicting the route. However, if it is no, then there is requirement of predicting the route which creates the need of utilizing route predicting model. Then, further, dynamic travel time (DTT) model (to calculate travel time from origin to destination) and dynamic delay time model (DDT) (to calculate the delay happened due to various obstacles such as – traffic, congested terminal) is required to use. With the help of these models, time calculation will take place.

- If P1 is final destination, then output will be calculated with the formula: DTT + DDT.
- If P1 is not the final destination, then there is a requirement of utilizing Leadtime Model and output will be calculated with the formula: DTT + DDT + Leadtime.

5.4.3.1 Port congestion Indicators

Objective of this step is to Derive metrices that gives quantitative congestion measure at port. For this purpose, I used the AIS historical data for last two years and Port centroid information provided by the secondary sources such as Spire Global.

5.4.3.2 Features used:

1. max_port_distance/pci_distance : Radius which cover the whole port.
2. displacement_R : Displacement from a lat/long to port centroid.
3. lat_mod,lon_mod : rounded values of lat/long with range +- 1.

5.4.3.3 Congestion calculation

Find the following congestion metrics inside the pci_distance for each port:

1. Convex Hull Area: Spread of vessels
2. Geolocation Area: Density of vessels
3. Average Vessel Proximity: Closeness of vessels.

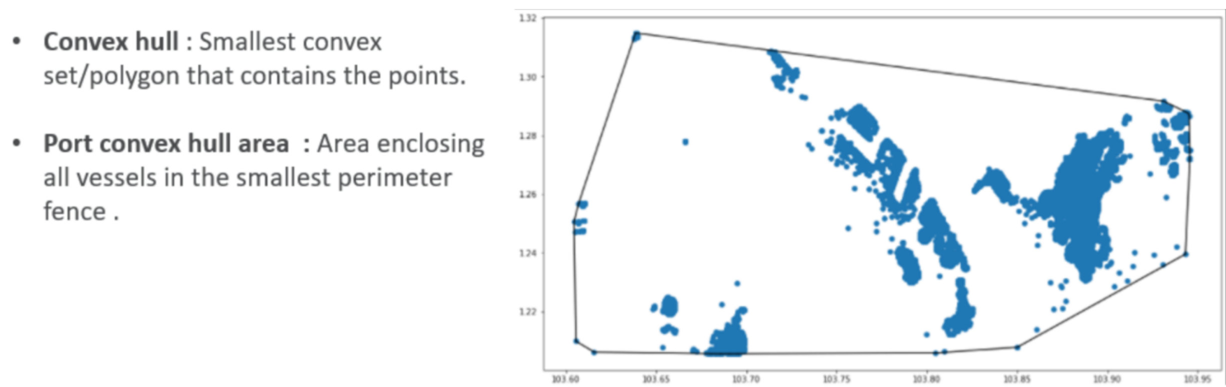


Figure 9: Congestion Calculation
(Source: Created by Learner)

5.4.3.4 Data Preparation

1. Convert the data into 12-hour interval data, where each vessel is represented by one point for every 12-hour period, with the data being the last event recorded within that 12-hour window.
2. Merge the data with port centroid data based on rounded values of latitude and longitude (lat_mod, lon_mod). Each MMSI will be associated with multiple lat_mod, lon_mod values, representing multiple nearby ports.

3. Filter the data for the points which are within any port's boundary.

5.4.3.5 Port Congestion Indicators (PCI) Calculation

1. Convex Hull Area:
    a. Calculate Convex Hull
    b. Calculate Area of Convex Hull in Earth Miles.
2. Geolocation Area:
    a. Calculate Geohash for each latitude and longitude.
    b. Calculate the total area using precision 7 within a timestamp and UNLOCODE (United Nations Code for Trade and Transport Locations).
3. Average Vessel Proximity (AVP):
    a. Calculate haversine distance between each latitude-longitude pair within a timestamp and UNLOCODE (United Nations Code for Trade and Transport Locations).
    b. Remove distances which are 0 or greater than 95th quantile.
4. If there are Nan values:
    a. Average Vessel Proximity: Fill it with minimum Average Vessel Proximity (To denote vessels are far apart).
    b. Geolocation Area: Fille it with minimum Geolocation Area (To denote less density of vessels).

5.4.3.6 Features Used in model based on Port Congestion Indicators

1. Spatial Concentration: Spatial distribution of ships within the convex area.
    a. Formula = Average Vessel Proximity /Convex Area
2. Spatial Density: Area used by the vessels.
    a. Formula = Geolocation Area / Convex Area

5.4.4 Calculation of Dynamic ETA and Static ETA

The Ocean ETA model (produced by learner) has been built using ML algorithms that are being trained using vessel's AIS historical data, which involves combination of features such as distance, vessel properties and ports/ terminals past behavior. All the mentioned features are derived using complex statistical methods. In short, the ETA (predictive analytics) relies on past data to predict what might happen in the future.

However, there is still a lot of other factors which might add error to the ETA predictions such as weather, ocean currents, vessel breakdowns, short come of labor at the destination for loading / unloading of the containers. etc. While these erratic factors contribute a lot to the Ocean ETA, this information aren't always available while calculating ETA from the model predicted method.

Ocean Carriers schedules are powerful data source. The carriers have proforma schedule that they need to stick with; they have agreements with terminals and are aware of the back-office negotiations with terminals which affect operational schedules - something outsiders have no visibility into (Fancello et al., 2011). Hence, utilization of hybrid intelligence model remains relevant in the situation where instead of ignoring all the useful information available from the carriers both Model Predicted ETAs and Carrier ETAs are used as it helps to enhance the accuracy level of the ETA prediction.

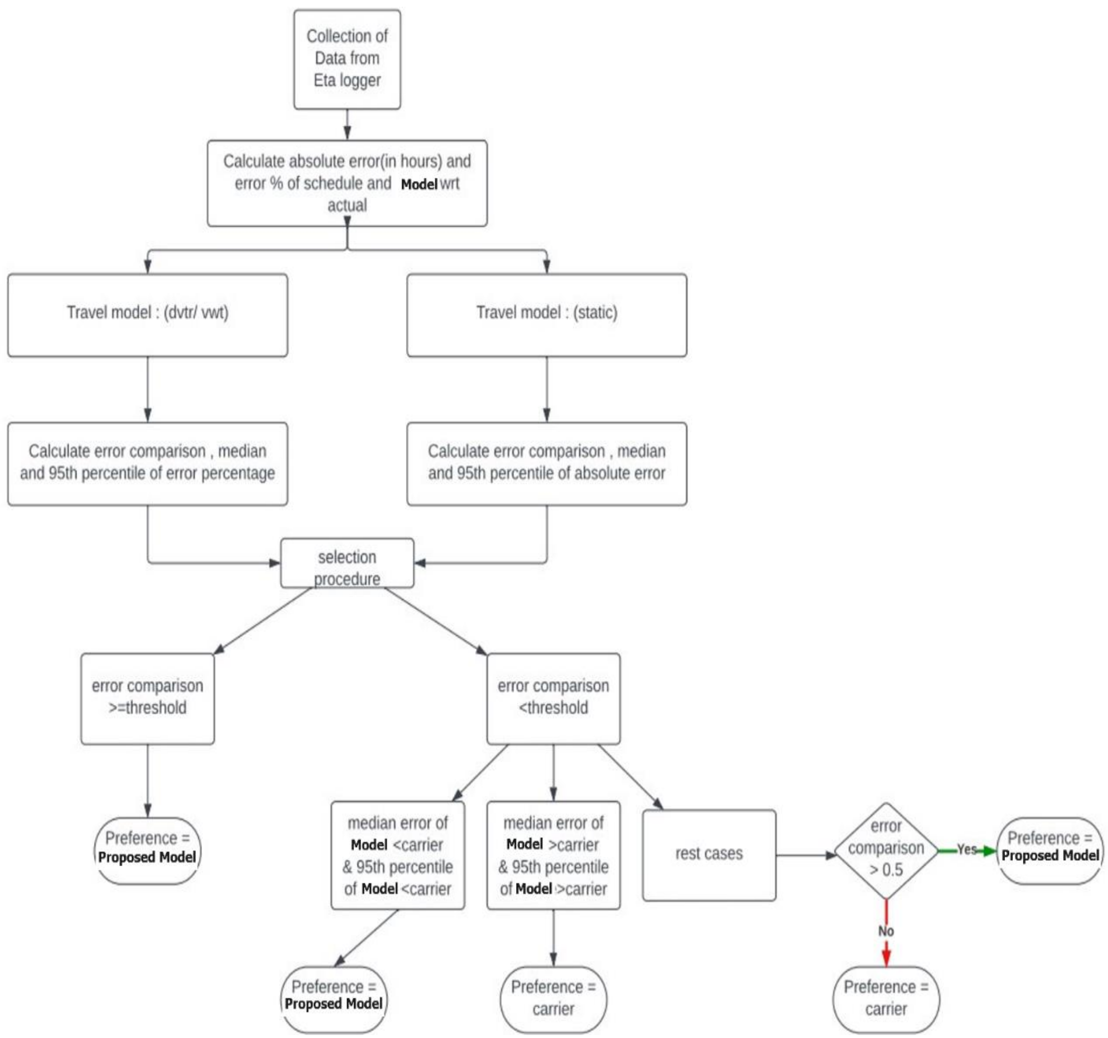


Figure 10: Selecting ETA Calculation Method
(Source: Created by Learner)

Initially, the data will be collected from ETA logger (third party). This collected data will be used to calculate absolute error (in hours) and in error % of schedule and model with respect to the actuals. It is done with the help of DVTR and static model and both of them assisted to calculate error comparison, median and 95th percentile of error %. Both the output is further combined to lead selection procedure. Then, error comparison takes place where selection between Ocean model and Carrier model takes place based on error%. If error percentage is higher more than or equal to threshold, then proposed model (produced by learner) is required to take. In the contrary, if error percentage is lower than the threshold, then there can be 3 situations. They are: if median error of model is lower than carrier & 95^{th} percentile of model then proposed model (produced by learner) will be used. In the contrary, if median error of model is higher than carrier & 95^{th} percentile of model then carrier model will be used. In the case, where both are equal, again error comparison will take place. If error comparison is higher than 0.5 then proposed model (produced by learner) will be used whereas if error comparison is lower than 0.5 then carrier model will be used.

5.4.4.1 Dynamic ETA

Dynamic ETA comes to play only when a vessel has left its last stop, and currently on its way to the next stop (A). Depending upon the vessel's current live position there are two different scenarios for which dynamic ETA is calculated at port A:

Initially, radius (R) is calculated for each ocean port using AIS historical data. R is the radius where it is observed that most of the vessels waiting due to congestion of the port at anchorage area. Travel time to reach at port A is calculated based on the vessel's current location whether it currently lies outside R or in the inside R. Further, Port Congestion index (PCI) plays role when vessel is inside the R radius of the port A. In that case, vessels currently waiting is considered at anchorage and at berthing state at port A. In the situation, where vessel is outside R distance of destination port A, ETA is predicted of the vessel from the data science models that are built for vessel specific for particular O-D pair (Origin-Destination pair).

In cases when there is no historical occurrence of the current vessel travelling to the O-D pair, then speed of the vessel is predicted using distance / displacement left to destination D from origin O. Hence calculate the average ETA.

For the overall VTR model, vessel's current live speed is also considered to predict travel time in outside R case.

Therefore, if vessel current live location is outside R –

- ETA at port A = Travel time to reach at port A (VTR)

If vessel current live location is inside R –

- ETA at port A = Vessel Wait Time (VWT)

The data sources used for the calculation of Dynamic ETA are as follows -

1. AIS (Automatic Identification System):
    a. Basic Information: MMSI, timestamp
    b. Positional Information: latitude, longitude
    c. Speed & Course Information: speed, course, heading, rot, status, maneuver.
2. Vessel:
    a. Static Vessel Information: length, width, name, ship and cargo type, IMO.
3. Port Information:

a. Static Port & terminal Information: latitude, longitude (geofence)

Data transformations & Feature Engineering steps:

1. Converted AIS data with 1 sec frequency into 1 hr frequency.
2. Mapping of Stop and Travel Events and Arrival and Departure Events.
3. Data Filtered for Last 2 years.
4. Data filtered for more frequent and recent vessels and Ports.
5. Data Includes data from entering the anchorage area (determined by radius R) till destination port.
6. Target Variable: Arrival Date at Destination port – Current timestamp of the vessel.

The features used for the calculation of Dynamic ETA are as follows -

1. Percentage of space occupied on a terminal.
2. Time taken by a vessel to reach point A (current position) from the origin port.
3. Length: length of the vessel.
4. Speed: current speed of the vessel.
5. Heading: heading/direction of the vessel
6. Status: status of the vessel.
7. Displacement left to be cover from point A (current position) to the destination port.
8. Bearing: angular slope between point A (current position) to the destination port.
9. Complexity: Average vessel proximity between Vessel and Port.
10. Density: Total Area taken by anchored / moored vessels at port.

The technique used for the calculation of Dynamic ETA are as follows -

1. Trained model for each terminal separately.
2. Removed Outliers using Interquartile range method method based on distance left to be covered.
3. Used XGBoost Regressor as a training model while scaling the numerical features and encoding the categorical features.
4. For terminals with not enough count an average model is used:
   a. Displacement left to be covered divided by minimum of (2, current speed).

The techniques used for the performance validation are as follows -

- Evaluated accuracy metrics like MAE, MAPE to determine the best-performing model.
- Used test set for validation.
- Compare predicted values with actual predictions.

5.4.4.2 Static ETA

For all future stops after stop A, I used STATIC ETA models for ETA calculation. Estimated Time of Arrival and Estimated Time of Departure at port B for a given vessel (when PORT B comes after port A, as explained in the previous Scenario):

ETA at port B = VTR + VDT

Here:

VTR - (Vessel Travel Time) Travel time reach port B from its last stop.

VDT – (Vessel Delay Time) gives any possible additional time that the vessel could take to travel between an OD pair. To predict Arrival of the vessel we are going to comparison of three different sources, they are:

- Model predicted ETA.
- Carrier ETA.
- Captain ETA.

Whichever source has latest updated time (between Carrier vs Captain) calling it Carrier ETA, I used that to compare with my ETA (Model predicted).

If my ETA < Carrier ETA, model chooses Carrier ETA handling the early Arrival case (with this there will be no Early arrival at Point of Load).

Tested the model on the Ocean transport data received from various third party open source publishers such as https://ecomm.one-line.com/one-ecom/schedule/long-range-schedule and https://www.track-trace.com/container.

Similar results were found when the results are tested for different trade lane from Yokohama Port in Japan to the Port of Shanghai in China. The Mean error was significantly less in my model predicted ETAs and Carrier predicted ETAs when compared those with Actual events over a period of time.

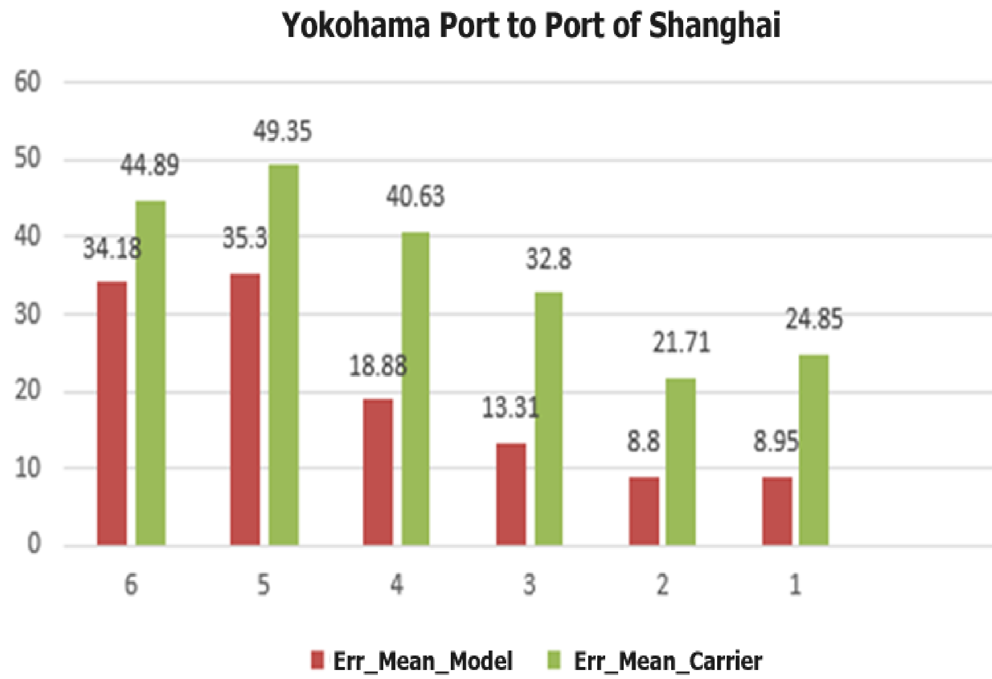


Figure 11: Yokohama Port to Shanghai Port
(Source: Created by Learner)

In the above graph,

- x-axis denotes - Days away from destination.
- y-axis denotes - Scale of error in hours.
- Err_Mean_Model – In Red (Produced by the learner) in hours.
- Err_Mean_Carrier - In Green (Produced by the carrier) in hours.

The above graph compares the Err mean values between my model's ETA and Carrier's ETA. The x-axis shows the number of days away from the destination, and the y-axis shows the error in hours. The error is calculated by subtracting the actual ETA from Ocean model (produced by learner) from carrier ETA.

The error mean, represented by Err_Mean_Model and Err_Mean_Carrier, is calculated by subtracting the actual ETA from the corresponding estimated ETA (Model or Carrier) for the route JPYOK (Port of Yokohama) – CNSHA (Port of Shanghai). Then, the average of these differences is taken for a specific number of days away from the destination (represented on the x-axis).

Above graph helps to evaluate that in regard to 1 day accuracy; the error mean value of ocean model (produced by learner) was 8.95 hours whereas the error mean value of Carrier's ETA was 24.85 hrs. Similarly, in regards to 2 day accuracy, the error mean value of ocean model was 8.8 hours whereas Carrier's ETA was 21.71 hours; for 3 day, the error mean value of ocean model was 13.31 hours whereas Carrier's ETA was 32.8 hours; for 4 day, it was 18.88 hours for ocean model whereas 40.63 hours for Carrier's ETA; for 5 day, it was 35.3 hours for ocean model whereas it reached to 49.35 hours for Carrier's ETA. Similarly, for 6 days, it was 34.18 hours for ocean model and 44.89 hours for Carrier's ETA.

For example, on day 5, the average error (Err_Mean_Model) represents the typical difference between the model ETA and the actual ETA for all routes arriving on day 5. This average difference is plotted in hours on the y-axis.

5.4.4.3 Error Calculation

- To calculate the error value for a shipment, we need the Model Estimate Day, Carrier Estimate Date, and Actual Date.

  - For example, if the Ocean Model Estimate Day is February 24, 2024, the Carrier Estimate Date is March 5, 2024, and the Actual Date is January 24, 2024:
  - Model Err value = Actual Date - Model Estimate Date = 31 days
  - Carrier Err value = Actual Date - Carrier Estimate Date = 41 days
- Average Error Calculation for a Day:
- Repeat the error calculation process for multiple vessels within the same 6-day period.
- Once the Model/Carrier err values are obtained for all vessels, calculate the average for that day.
  - For instance, if there are 20 vessels in the bucket period and the sum of Model err values equals 684 days:
  - Average Model Err = Sum of Model err values / Number of vessels = 684 / 20 = 34.2 days
- Similarly, calculate the average for Carrier.
  - For example, if there are 20 vessels in the bucket period and the sum of Carrier err values equals 898 days:
  - Average Carrier Err = Sum of Carrier err values / Number of vessels

Overall, the graph suggests that Ocean model (produced by learner) provides more accurate ETAs than carrier.

Here are some additional insights that can be gleaned from the graph:

It reflects that the average error for Carrier is higher than the average error for ocean model (produced by learner) for the route JPYOK - CNSHA. This suggests that Ocean model's ETAs are more accurate than Carriers ETAs. The error for both model and Carrier appear to be more variable for days further away from the destination. Overall, it can be evaluated from the graph that Ocean model provides more accurate ETAs than carrier.

To evaluate the prediction, first of all, the example of one of the trade lanes from Yantian Port in China to Ningbo Port in China is selected to test the accuracy of the predictions of Expected time of arrivals predicted by my model and the Carrier. It has supported to lead comparative analysis of both the aspects with the help of Actuals over a period of time. This comparison was done using the Mean error method and there was a clear improvement in terms of accuracy of the predictions.

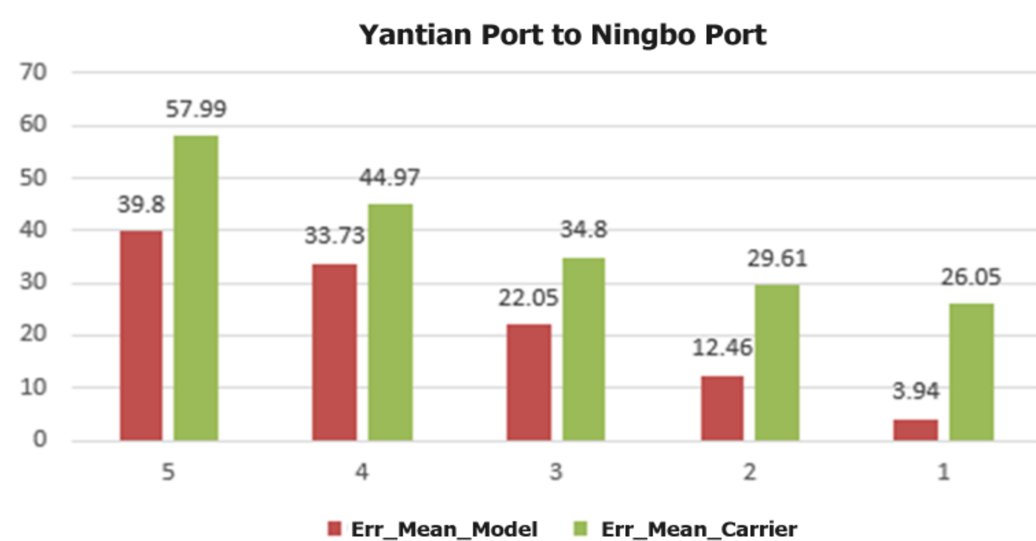


Figure 12: Yantian Port to Ningbo Port
(Source: Created by Learner)

In the above graph,

- x-axis denotes - Days away from destination.
- y-axis denotes - Scale of error in hours.
- Err_Mean_Model – In Red (Produced by the learner) in hours.
- Err_Mean_Carrier - In Green (Produced by the carrier) in hours.

The graph illustrates the comparison between the Err mean Values between my model and Carrier for the route CNYTN (Port of Yantian) – CNNBG (Port of Ningbo). The x-axis denotes the number of days away from the destination, while the y-axis displays the error in hours, calculated by subtracting the actual ETA from the estimated ETA by either my model or Carrier. From the above graph, it can be evaluated that in regard to 1 hour accuracy; the error mean value of my model was 3.94 whereas Carrier's ETA was 26.05. Similarly, in regards to 2 hour accuracy, the error mean value of my model was 12.46 whereas Carrier's ETA was 29.61; for 3 hour, the error mean value of my model was 22.05 whereas Carrier's ETA was 34.8; for 4 hour, the error mean value of my model was 33.73 whereas Carrier's ETA was 44.97 and for 5 hour, it is 39.8 for my model whereas 57.99 for Carrier's ETA.

By analyzing the data, it is found that my model's average error (Err_Mean_Model) is consistently lower than Carrier's across different days leading up to the destination. This indicates that my model tends to provide more accurate ETAs for this route compared to Carrier.

Moreover, both my model and Carrier exhibit higher variability in errors as the days from the destination increase. This suggests that predicting ETAs becomes more challenging as the timeframe extends, potentially due to factors such as changing weather conditions or logistical complexities.

In summary, the graph suggests that the model suggested by me (the learner) offers superior accuracy in ETA as compared to Carrier for the CNYTN (Port of Yantian) – CNNBG (Port of Ningbo) route.

Hence, it is analyzed that in both the above cases the preference is given to the Ocean model's ETA as the mean error value is lesser than the Carrier ETA.

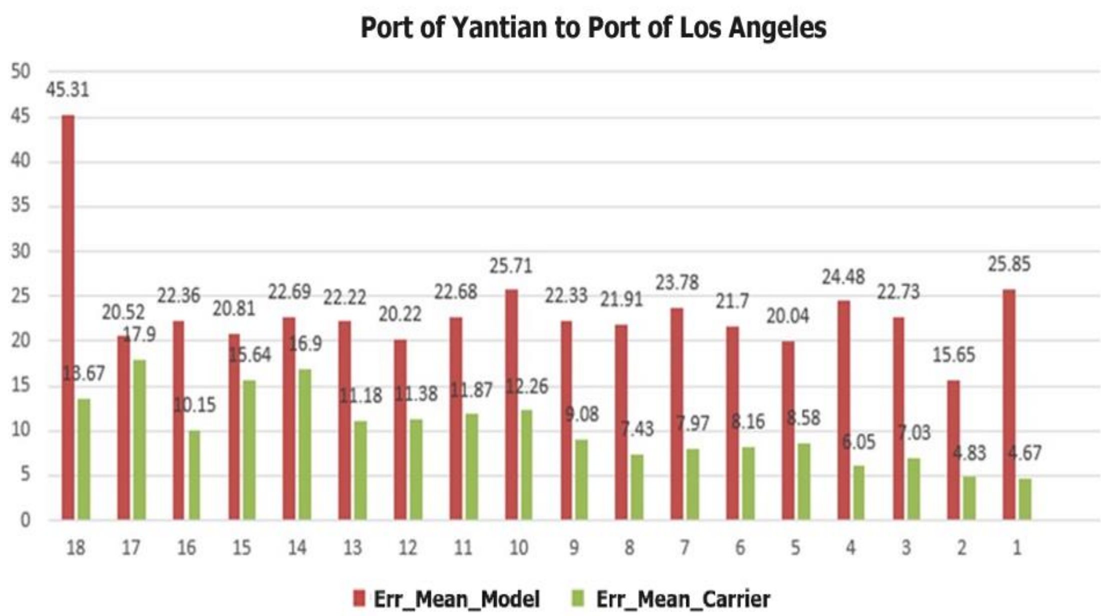


Figure 13: Port Yantian to Port of Los Angeles
(Source: Created by Learner)

The graph shows the comparison between the Err mean Values between my model and Carrier for the route CNYTN (Port of Yantian) – USLAX (Port of Los Angeles). The x-axis denotes the number of days away from the destination, while the y-axis displays the error in hours, calculated by subtracting the actual ETA from the estimated ETA by either my model or Carrier.

By analyzing the data, we find that Carrier's average error is consistently lower than that of my model across different days leading up to the destination. This indicates that Carrier tends to provide more accurate ETAs for this route compared to my model.

In summary, the graph suggests that Carrier offers more accuracy in ETAs compared to my model for the CNYTN (Port of Yantian) – USLAX (Port of Los Angeles) route.

In the above cases the preference is given to the Carrier ETA over my model's ETA as the mean error value is lesser. This is the situation where the Carrier ETA is performing better than my model ETA.

There were some longer trade lanes such as from Busan New Port in Korea to Long Beach Port in the US where the results were mixed i.e. 10 plus days away from the destination, Carrier predictions were better but as we come closer to the destination, Ocean model (produced by learner) predictions were seemingly better.

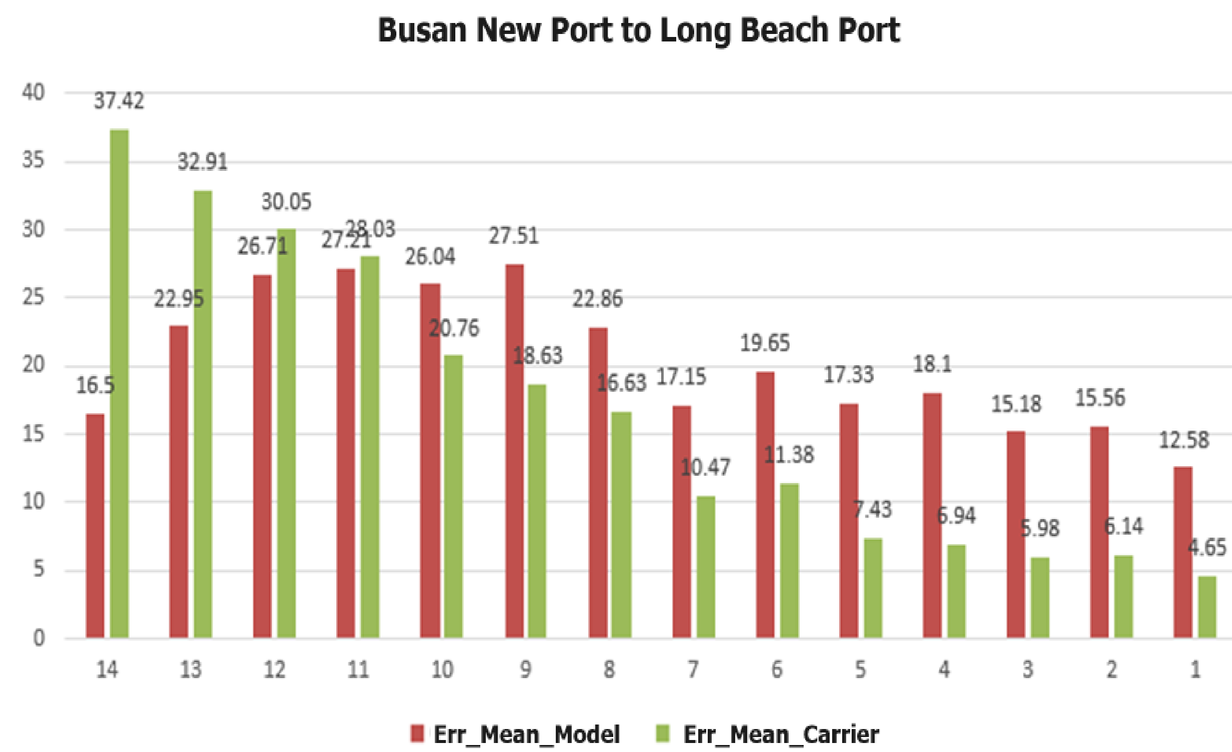


Figure 14: Busan New Port to Long Beach Port
(Source: Created by Learner)

In the above graph,

- x-axis denotes - Days away from destination.
- y-axis denotes - Scale of error in hours.
- Err_Mean_Model – In Red (Produced by the learner) in hours.
- Err_Mean_Carrier - In Green (Produced by the carrier) in hours.

Err Value Calculation Process – Busan New Port to Long Beach Port:

- To calculate the error value for a shipment in the 11th day bucket, we need the Model Estimate date, Carrier Estimate Date, and Actual Date.

- For example, if the Model Estimate date is February 20, 2024, the Carrier Estimate Date is February 11, 2024, and the Actual Date is January 21, 2024:
    - Model Err value = Actual Date - Model Estimate Date = 27 days
    - Carrier Err value = Actual Date - Carrier Estimate Date = 28 days

Average Error for a Day was calculated as follows -

- Repeat the error calculation process for multiple vessels within the same 11th day period.
- Once the Model/Carrier err values are obtained for all vessels, calculate the average for that day.
- For instance, if there are 20 vessels in the bucket period and the sum of Model err values equals 561 days:
    - Average Model Err = Sum of Model err values / Number of vessels = 544 / 20 = 27.2 days
- Similarly, I calculate the average for Carrier.
- For example, if there are 20 vessels in the bucket period and the sum of Carrier err values equals 561 days:
    - Average Carrier Err = Sum of Carrier err values / Number of vessels = 561 / 20 = 28.05 days

Finally, the above process was repeated for each day to obtain average error values for Model and Carrier estimates.

From the above graph, it can be analyzed that the model proposed by Ocean model (produced by learner) is performing better closer to the destination (day 11 to day 14). However, it is not performing well, from day 1 to day 10. It helps to represent that in some cases, Ocean model (produced by learner) ETA is performing better, while in some cases Carrier ETA is performing better.

From the above graph, it can be illustrated that the comparison between the Err mean Values between my model and Carrier for the route KRBNP (Port of Busan) – USLGB (Port of Long Beach). The x-axis denotes the number of days away from the destination, while the y-axis displays the error in hours, calculated by subtracting the actual ETA from the estimated ETA by either my model or Carrier.

By analyzing the data, we find that my model's average error (Err_Mean_Model) is lower at certain days and the carrier's error (Err_Mean_Carrier) is lower at certain days. This is the case where the Hybrid approach comes into play.

5.5 Discussion of Research Question Four

The fourth and the final question that the researcher wanted to explore through this was – "How can we leverage data intelligence, carrier inputs, and external factors such as weather and traffic conditions to improve the accuracy of ETA predictions?". For this the learner came with the following pointers and gave a Hybrid approach which is best of both worlds' predictions –

- Using Captain ETA for immediate next port, Captain ETA is manually inputted by the captain of the ship, who can increase / decrease the speed of the vessel to reach the next port early or late as per the planned.
- Using Live average speed of the vessel to compare the travel time predicted from historical data and using aggregate travel time.
- Using my Ocean ETA logger (a collection containing all information such as Blume ETA, Schedule ETA, and actual ETA for every possible journey), I analysed all possible scenarios in which my model performed better and in which Carrier prediction performed better.
- Built maximum possible and minimum possible travel time threshold for each vessel and its OD pairs (Origin – Destination pairs) to validate prediction at each level.
- Below is the High-Level Design flow for the Hybrid Approach for Ocean ETA:

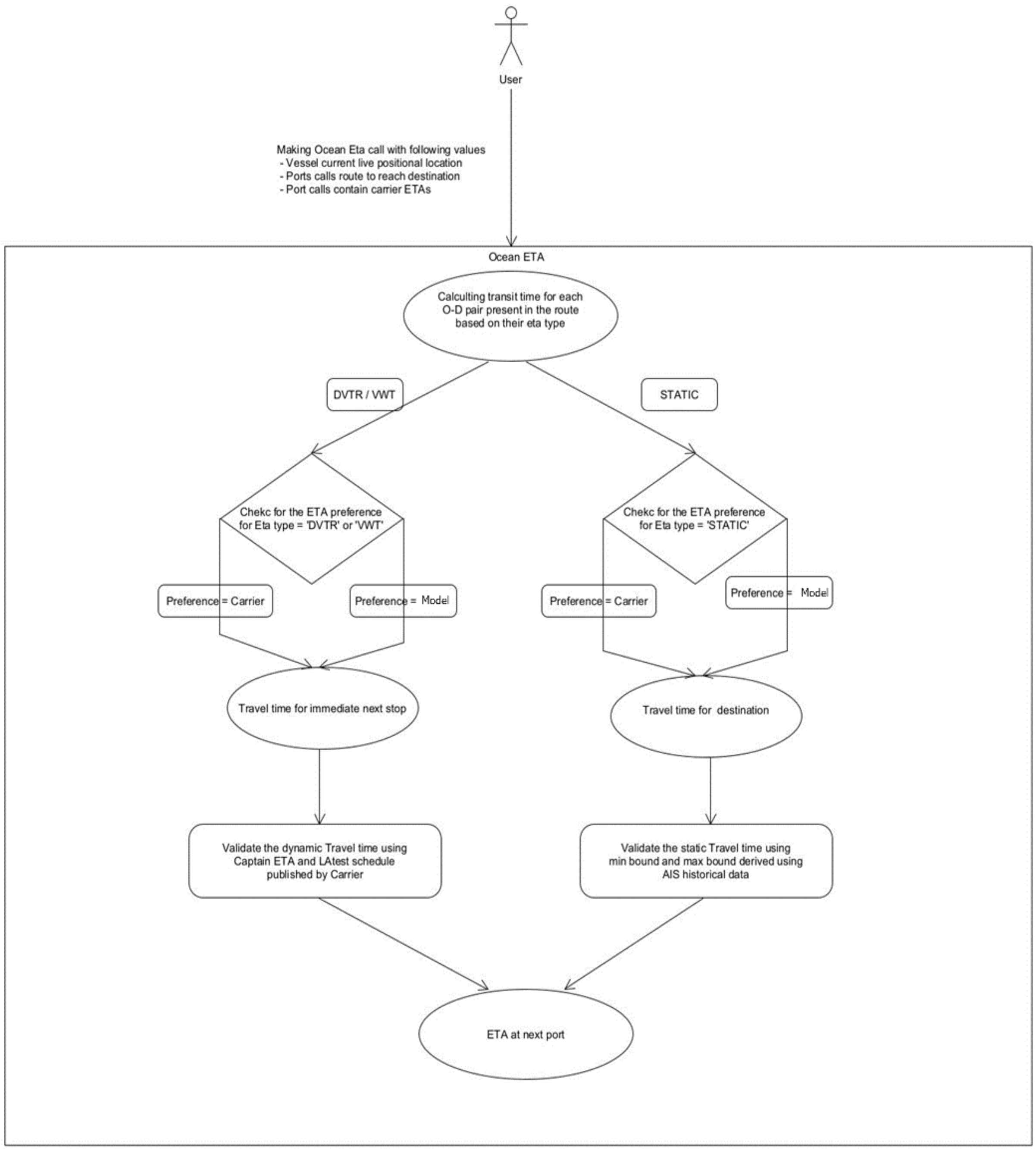


Figure 15: Intelligence with Historical Predictions
(Source: Created by Learner)

This flowchart represents the way arrival times of vessels is managed and forecasted while considering both dynamic conditions and static schedules. It represents that Ocean ETA call is made with the help of vessel's current live position, ports calls route to reach destination and port calls contain carrier ETAs. Further, it represents that to calculate transit time for each O-D pair present in the route based on their ETA type. DVTR and static model help to check for the ETA preference for ETA type. According to the ETA preference is required to given to Carrier or Ocean ETA in both the situation. On the basis of preference, DVTR helps to calculate travel time for immediate next step and static model helps to calculate travel time for destination. Further, validation of dynamic travel time takes place on the basis of Carrier ETA and latest schedule published by carrier. In the contrary, static static travel time is validated by utilizing minimum bound and maximum bound derived using AIS historical data.

Above finding has shown similarity with the research finding of Balster (2020) that adaptation of the latest machine learning capabilities with historical shipment journey data helps to improve the prediction of the future shipment journey which can be further use on the same Origin-Destination pairs may look like while including the delays, ETAs, as well as the exceptions on the way.

On top of the above model, few catch all rules are applied as below -

5.5.1 For Immediate Next Port Destination

Vessel journey is divided into 2 parts Transit Time (DVTR), Vessel Wait Time (VWT) and Turnaround Time (VTT). DVTR and VWT is used to predict ETA at POL (Point of Load), where VTT is used to predict ETD at POL.

- DVTR - is the transit time before the vessel starts waiting for the anchorage, from the historical eta error analysis my DVTR has performed quite well.
- VWT - is the vessel wait time by vessel before it gets berthed, my VWT has performed sufficiently good but however in new journeys by vessel or exceptional cases VWT increase / decrease varies quite a lot from the historical VWT.

Hence, for VWT, over the hybrid approach if my VWT is within the VWT Threshold (24 hours for now), the model uses my VWT and predict ETA else it uses Carrier VWT and predicts ETA for POL.

- VTT - If my VTT is within the VTT Threshold (difference between carrier VTT and my VTT, 24 hours for now), the model uses my VTT to predict ETD else it uses Carrier VTT and predict ETD at POL.

5.5.2 Hybrid Approach

From the above results, it is found that the Ocean ETA approach has supported to represent the culmination of cutting-edge technological advancements in data science and maritime logistics. The reason found behind it is that its core lies a sophisticated ensemble of machine learning (ML) models meticulously trained on a vast repository of historical data sourced from vessels' Automatic Identification System (AIS) transmissions. This treasure trove of information encompasses a myriad of vital parameters, including but not limited to:

- **Distance:** A comprehensive understanding of the geographical expanse covered by vessels, factoring in diverse routes and destinations, forms the bedrock of our predictive analytics framework.
- **Vessel properties:** Each vessel encapsulates a unique set of characteristics ranging from its physical dimensions to its propulsion system and operational capabilities. Our models delve deep into these nuances to glean invaluable insights into their performance trajectories.
- **Ports/Terminals past behaviour:** Ports and terminals serve as pivotal nodes in the intricate web of global maritime logistics. By scrutinizing historical data pertaining to vessel arrivals and departures at these crucial junctures, we unravel intricate patterns and trends that underpin our predictive ETA algorithms.

The amalgamation of these features, painstakingly extracted through the application of advanced statistical methodologies, furnishes with a robust foundation for forecasting the Estimated Time of Arrival (ETA) of vessels traversing the high seas.

Yet, the maritime domain is replete with myriad uncertainties, many of which elude the purview of conventional predictive models. The capricious whims of Mother Nature, manifested in the form of erratic weather patterns and unpredictable ocean currents, pose formidable challenges to our endeavor. Moreover, the specter of unforeseen contingencies such as vessel breakdowns and labor shortages at destination ports looms large, casting a shadow of doubt over the reliability of our ETA predictions.

Acknowledging the inherent limitations of relying solely on historical data, it has pioneered a paradigm shift in maritime ETA forecasting through the conceptualization of a hybrid intelligence model. At its essence lies a symbiotic fusion of predictive analytics and real-time data gleaned from ocean carriers' schedules.

Ocean carriers, endowed with intimate knowledge of their vessels' operational cadence and fortified by established agreements with terminals, wield an unparalleled vantage point in navigating the labyrinthine corridors of maritime logistics. Leveraging this wealth of firsthand insights, synergized with their provided ETAs with our predictive models, culminating in a holistic and nuanced understanding of vessel arrival dynamics.

This convergence of methodologies affords the ability to adapt swiftly to evolving circumstances, mitigating the impact of unforeseen disruptions on our ETA predictions. Augmented by prescriptive analytics, our hybrid approach transcends the realm of mere prognostication, empowering stakeholders with actionable recommendations to optimize their logistical strategies.

To validate and refine the efficacy of our hybrid model, it has meticulously curated a comprehensive dataset comprising predictive ETA, schedule ETA, and actual ETA data spanning multiple vessels since August 2023. Through exhaustive analysis and iterative refinement, we continue to chart new frontiers in maritime ETA forecasting, propelling the industry towards unparalleled levels of efficiency and reliability.

With above Hybrid approach few more Rules are applied -

5.5.3 For Immediate Next Port Destination

Vessel journey is divided into 2 parts Transit Time (DVTR), Vessel Wait Time (VWT) and Turnaround Time (VTT). DVTR and VWT is used to predict ETA at POL, where VTT is used to predict ETD at POL.

DVTR - is the transit time before the vessel starts waiting for the anchorage, from the historical eta error analysis my model DVTR has performed quite well.

VWT - is the vessel wait time by vessel before its gets berthed, model VWT has performed sufficiently good but however in new journeys by vessel or exceptional cases VWT increase / decrease varies quite a lot from the historical VWT. So, for VWT, over the hybrid approach if model VWT is within the VWT Threshold (24 hours), I used my model VWT and predict ETA else I used Carrier VWT and predict ETA for Point of load.

VTT - If my model VTT is within the VTT Threshold (24 hours), I used my model VTT to predict ETD else I used Carrier VTT and predict ETD for Point of load.

Further To predict Arrival of the vessel we are going to compare 2 different sources -

- My model ETA.
- Carrier ETA (that includes Captain ETA as well).

If my model ETA < Carrier ETA, I choose Carrier ETA to handle the early Arrival case (with this there will be no Early arrival at POL).

Hence, from the research outcome, it is found that utilization of hybrid method will remain appropriate as it will help to utilize mix of Carrier's ETA and the Model ETA (introduced by the learner). This concept has taken place from the Prescriptive Analytics as it involves making specific and actionable recommendations that are based on these forecasts. In the case of the selected research, carrier provided ETA will be utilized. The reason behind it is that carriers are providing ETAs from a long time, hence, it is not appropriate to completely neglect them. Hence, to enhance the effectiveness of ETA calculation. It would be relevant to utilize the method that can offer reduced error result. To be able to achieve the results, model's ETA (produced by learner) was logged in (Schedule ETA and actual ETAs of multiple vessels) from August 2023. It has supported to analyze the results to find a method for the hybrid approach.

CHAPTER VI:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

The researcher explored two Research Hypothesis in this thesis and two Null Hypothesis against it. It was concluded that at the end of the research two hypothesis were selected and two were rejected. These are as follows –

	Hypotheses	**Selection / Rejection**
H1	The accuracy of Ocean ETA predictions is significantly improved by incorporating factors like vessel position, vessel speed and port congestions.	**Selected**
H01	The accuracy of Ocean ETA predictions is not significantly improved by incorporating factors like vessel position, vessel speed and port congestions.	**Rejected**
H2	The accuracy of Ocean ETA predictions is significantly improved by considering carrier predictions also in the approach and choosing the best of both worlds in a hybrid approach.	**Selected**
H02	The accuracy of Ocean ETA predictions is not significantly improved by considering carrier predictions also in the approach and choosing the best of both worlds in a hybrid approach.	**Rejected**

Table 2: Hypotheses Testing
(Source: Created by Learner)

Regarding hypotheses 1, from the research finding, it can be discussed that accurate estimation of vessel arrival times (ETA) is a cornerstone of efficient maritime logistics operations. Despite advancements in prediction models, traditional approaches often overlook dynamic factors such as real-time vessel position updates, speed variations, and port congestion metrics. This study proposes an innovative framework that integrates these dynamic parameters into Ocean ETA prediction algorithms. Researcher hypothesizes that by incorporating dynamic factors such as real-time vessel position updates, speed variations, and port congestion metrics into Ocean ETA prediction algorithms, researcher anticipates a marked enhancement in predictive accuracy. This improvement will manifest through a reduction in ETA deviation from actual arrival times, as measured by mean absolute error metrics. Researcher posits that the integration of these multifaceted parameters will enable the model to capture the intricate interplay between maritime traffic dynamics and port operations, thus yielding more nuanced and contextually relevant predictions. Consequently, stakeholders across the maritime supply chain, including shipping companies, port authorities, and cargo owners, stand to benefit from more reliable scheduling, optimized resource allocation, and reduced operational uncertainties. The inclusion of vessel position, speed, and port congestion factors into Ocean ETA prediction algorithms will not yield a statistically significant improvement in predictive accuracy when compared to conventional models that rely solely on historical data and basic route calculations. This suggests that the additional complexity introduced by integrating dynamic variables does not contribute meaningfully to reducing the discrepancy between predicted and actual arrival times. Consequently, any observed differences in prediction accuracy between the enhanced model and the baseline model are likely due to random variation or noise within the data rather than the effectiveness of the incorporated factors.

In regard to hypotheses 2, from the research finding, it can be discussed that the accuracy of Ocean ETA (Estimated Time of Arrival) predictions can be markedly enhanced through the integration of carrier predictions alongside historical data within a hybrid prediction approach. This hybridization aims to capitalize on the complementary strengths of both methodologies, leveraging the robustness of historical data and the real-time insights provided by carrier predictions. By synthesizing these diverse sources of information, the hybrid model is anticipated to achieve a synergistic improvement in predictive accuracy, offering more dependable and finely tuned ETA estimations for maritime logistics operations. In this research hypothesis, the researcher proposes that by considering carrier predictions in conjunction with historical data, the hybrid approach will demonstrate a discernible reduction in ETA deviation from actual arrival times. By quantifying and analyzing key performance metrics such as mean absolute percentage error (MAPE), researcher anticipates observing a statistically significant improvement in prediction accuracy compared to traditional models relying solely on historical data.

Furthermore, the integration of carrier predictions into the hybrid model is expected to enable a more adaptive and responsive forecasting framework. By dynamically incorporating real-time insights from carriers regarding vessel schedules, routing adjustments, and potential disruptions, the model can better capture the evolving dynamics of maritime transportation networks. This adaptability is crucial for mitigating the impact of unforeseen events such as weather disturbances, port congestion, or vessel delays, which can profoundly influence arrival times. Through the hybridization of prediction methodologies, researcher's hypothesis posits that the model will effectively navigate the trade-off between historical data stability and carrier prediction timeliness, striking a balance that optimizes predictive accuracy. Consequently, stakeholders across the maritime supply chain, including shipping companies, port authorities, and cargo owners, stand to benefit from more informed decision-making, enhanced operational efficiency, and reduced uncertainties in scheduling and resource allocation. The validation of this hypothesis through empirical analysis and comparative evaluation will contribute valuable insights into the efficacy of hybrid prediction approaches in maritime logistics. Moreover, the findings of this research have the potential to inform the development of more advanced and adaptable predictive models, facilitating continuous improvement in the efficiency and resilience of global maritime transportation systems.

The incorporation of carrier predictions alongside historical data in a hybrid approach to Ocean ETA predictions will not yield a statistically significant improvement in predictive accuracy compared to models relying solely on historical data or carrier predictions alone. This hypothesis suggests that the hybridization of prediction methodologies does not confer any meaningful advantage over the individual approaches in terms of reducing the discrepancy between predicted and actual arrival times. Instead, any observed disparities in prediction accuracy between the hybrid model and the individual methodologies are likely attributable to chance variation or noise within the dataset rather than the efficacy of the hybridization approach.

In further detail, the null hypothesis posits that the hybrid model, which integrates both historical data and carrier predictions, will not exhibit a discernible reduction in ETA deviation from actual arrival times compared to models utilizing only historical data or carrier predictions independently. The null hypothesis implies that the additional complexity introduced by the hybrid approach does not result in a statistically significant enhancement of predictive accuracy. Therefore, any differences observed in prediction performance metrics, such as mean absolute percentage error (MAPE), between the hybrid model and the individual methodologies are likely to be within the realm of random chance. Furthermore, the null hypothesis suggests that the hybrid model's purported ability to adaptively leverage both historical data stability and real-time insights from carrier predictions does not confer any substantive advantage over models relying solely on one source of information. Any observed improvements in predictive accuracy attributed to the hybrid approach are considered spurious and not indicative of a genuine enhancement in model effectiveness.

6.2 Implications

This research has multiple implications as lots of people will benefit from the findings of this study. Ship owners and the folks who own the cargo on those ships will find it really helpful. With the accurate predictions, they can plan their shipping schedules better, manage their goods more efficiently, and deal with any unexpected delays or issues.

However, it's not just them who benefit as other important people in the shipping world, like freight companies, port workers, and even customs officials, can use our predictions to plan ahead too. They can adjust their schedules and resources based on what we expect to happen with the ships (Port of Antwerp, 2020).

At the same time, the findings of this study can also be implemented for academic implications. Students from logistics, supply chain management, transportation, and international business can utilize the finding of this research to develop theoretical, practical, and professional knowledge about the importance of ETA prediction and its accuracy in calculation.

This research is based on limited ports; however, the findings of this research can be implemented on almost every port that deals in global intermodal transport networks.

From this, it can be evaluated that this research has multiple implications. On behalf of this, it is not wrong to say that this research is valuable for the global intermodal transport network system.

6.3 Limitations

To conduct this research, quantitative data collection method is used however, if qualitative data collection method is also used then it could support to enhance the relevancy of the research finding while involving subjective approach too. At the same time, it is also analyzed that this research has utilized secondary data collection method which has limited the research outcome as it remains dependent on existing studies. However, if primary data collection method would also be used then it could support to collect data from the current population who are actually facing the problem which could further support to increase the relevancy of the finding. It could also support to collect the data that can assist to increase the accuracy level of ETA. It creates opportunity for the future researchers to involve mixed data collection method while including both primary and secondary data collection method.

Moreover, this research has not involved external weather conditions. Hence there is also opportunity for the future researchers to involve these aspects too to increase the relevancy of the model. There is also opportunity for the future researchers to additionally explore how the vessel's characteristics can have some effect upon the overall travel time in the data science models.

6.4 Recommendations for Future Research

Future research on the topic of managing Estimated Time of Arrivals (ETA) in the global supply chain could explore several avenues to further advance understanding and improve practices –

6.4.1 Integration of Advanced Technologies

Investigate the potential of emerging technologies such as artificial intelligence, machine learning, and blockchain in enhancing ETA prediction accuracy. Explore how these technologies can be integrated into existing systems and processes to mitigate uncertainties and disruptions in supply chain operations (Fadda et al., 2011).

6.4.2 Real-Time Data Analytics

Conduct research on real-time data analytics capabilities to continuously monitor and analyze various factors affecting ETA predictions, such as weather conditions, port congestion, and vessel schedules. Explore how predictive analytics models can be refined and updated dynamically to improve accuracy and reliability (Fadda et al., 2011).

6.4.3 Collaborative Forecasting and Information Sharing

Investigate the benefits and challenges of collaborative forecasting and information sharing among stakeholders in the supply chain ecosystem. Explore strategies to facilitate seamless communication and data sharing to enhance the accuracy of ETA predictions and mitigate risks associated with information asymmetry (Fadda et al., 2011).

6.4.4 Environmental Sustainability Considerations

Examine the impact of environmental factors, such as carbon emissions and fuel consumption, on supply chain logistics and ETA predictions. Explore sustainable transportation alternatives and green logistics practices that can improve efficiency while reducing environmental footprint (Bodunov et al., 2018).

6.4.5 Supply Chain Resilience and Risk Management

Investigate strategies for enhancing supply chain resilience and risk management in the face of unforeseen disruptions and uncertainties. Explore how proactive risk mitigation measures and contingency planning can improve the robustness of ETA predictions and minimize the impact of disruptions on supply chain operations (Bodunov et al., 2018).

6.4.6 Regulatory Compliance and Trade Policies

Analyze the implications of regulatory compliance requirements and trade policies on supply chain logistics and ETA predictions, particularly in international trade scenarios. Explore how changes in regulations and trade agreements affect transportation routes, lead times, and ETA accuracy, and identify strategies to adapt to regulatory changes effectively (Fetzer et al., 2018).

6.4.7 Customer-Centric Approaches

Explore customer-centric approaches to ETA management, focusing on improving transparency, communication, and service levels to meet customer expectations. Investigate how personalized delivery options, real-time tracking capabilities, and proactive notification systems can enhance the overall customer experience while improving ETA prediction accuracy (Fetzer et al., 2018).

By addressing these areas of research, future studies can contribute to advancing the understanding and practices related to managing Estimated Time of Arrivals in the global supply chain, ultimately leading to more efficient, reliable, and sustainable logistics operations.

6.5 Conclusion

From the above research, it can be recommended that while calculating the accuracy of Ocean ETA predictions, it is essential to incorporate factors like vessel position, vessel speed and port congestions as it helps to improve the accuracy of Ocean ETA predictions.

It also helps to recommend that the accuracy of Ocean ETA predictions can be significantly improved by considering carrier predictions also in the approach as it helps to consider hybrid approach which further help to choose the best of both worlds and allows to increase the relevancy of predictions while eliminating the limitations of both the aspects.

The research outcome also helps to recommend that weather aspect is also required to involve in the calculation of Ocean ETA prediction to increase the relevancy of the findings.

Utilization of above recommendations in global intermodal transport networks can remain highly assistive to increase the chances of predicting the ETA accurately.

REFERENCES

Abdi, A. and Amrit, C. (2024) 'Enhancing vessel arrival time prediction: A fusion-based deep learning approach', *Expert Systems with Applications*, 252 [online]. Available at: https://www.sciencedirect.com/science/article/pii/S0957417424008546/pdfft?md5=80ea8c6a8d82b5388f7391fc8ed6b9f8&pid=1-s2.0-S0957417424008546-main.pdf (Accessed: 22 April 2023).

Ahlstrom, D., Bruton, G.D. and Yeh, K.S. (2007) 'Venture capital in China: Past, present, and future', *Asia Pacific Journal of Management*, 24(3) [online]. Available at: https://doi.org/10.1007/s10490-006-9032-1 (Accessed: 12 March 2023).

Bhattacharyya, D.K. (2006) '*Research methodology'*. Excel Books India.

Baker, M.J. (2000) 'Selecting a research methodology', *The Marketing Review*, 1(3) [online]. Available at: https://doi.org/10.1362/1469347002530736 (Accessed: 01 June 2022).

Balster, A. (2020) 'An Eta prediction model for Intermodal Transport Networks based on machine learning', *Business & Information Systems Engineering*, 62(5) [online]. Available at: https://api.semanticscholar.org/CorpusID:219404826 (Accessed: 02 January 2024).

Bodunov, O., Schmidt, F., Martin, A., Brito, A., and Fetzer, C. (2018) 'Real-time destination and ETA prediction for maritime traffic', *Proceedings of the 12th ACM* *international conference on distributed and event-based systems*, 198 [online]. Available at: https://dl.acm.org/doi/10.1145/3210284.3220502 (Accessed: 02 March 2024).

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, 3(2) [online]. Available at: http://eprints.uwe.ac.uk/11735 (Accessed: 15 November 2022)

Creswell, J.W. (2014) 'Writing Strategies and Ethical Considerations' in John W.C. (ed.) *Research design: Qualitative, quantitative, and mixed methods approach*. Sage Publications, pp. 134-189.

Ernie, T. (2008) 'This month in Physics history', *APS News*, 17(9) [online]. Available at: https://www.aps.org/publications/apsnews/200810/physicshistory.cfm (Accessed: 23 November 2021).

Fancello, G., Pani, C., Pisano, M., Serra, P., Zuddas, P. and Fadda. P. (2011) 'Prediction of arrival times and human resources allocation for container terminal', *Maritime Economics & Logistics* 13 (2) [online]. Available at: http://www.palgrave-journals.com/mel/journal/v13/n2/full/mel20113a.html (Accessed: 01 May 2022).

Flick, U. (2015) 'Steps in the Research Process', *Introducing research methodology: A beginner's guide to doing a research project.* Sage Publications, pp. 73-89.

Giannopoulos, G.A. and Pardalos, P.M. (2017) 'Data-driven optimization and modeling for transportation systems', *Handbook of Optimization in Complex Networks*, Springer Science & Business Media, pp. 255-275.

Goddard, W. and Melville, S. (2004) '*Research methodology: An introduction*'. Juta and Company Ltd.

Goundar, S. (2012) 'Research methodology and research method' in Goundar, S. (ed.) *Cloud Computing*. Research Gate Publications, pp. 1-47.

Hamari, J. and Järvinen, A. (2011) 'Building customer relationship through game mechanics in social games' in Cruz-Cunha, M., Carvalho, V., and Tavares, P. (eds.) *Business, Technological and Social Dimensions of Computer Games: Multidisciplinary Developments*. Sage Publications, pp. 348-365.

Hasheminia, H., and Jiang, C. (2017) 'Strategic trade-off between vessel delay and schedule recovery: An empirical analysis of container liner shipping', *Maritime Policy & Management* 44(4): [online]. Available at: https://ideas.repec.org/s/taf/marpmg.html (Accessed: 06 March 2023).

Higgins, A., Chow, P.S. and Jeihani, M. (2019) 'Transportation management systems and multimodal transportation', in *Supply Chain Management and Logistics: Innovative Strategies and Practical Solutions*. Springer Science & Business Media, pp. 187-200.

House of Commons (2009) 'Reducing CO2 and other emissions from shipping', *Parliament Publications UK*, [online]. Available at: https://publications.parliament.uk/pa/cm200809/cmselect/cmenvaud/528/528.pdf (Accessed: 3 January 2023).

Huang, T., Qiu, R.Z., Zhang, X. and Liu, Z. (2021) 'Scalable deep learning for multimodal transportation demand forecasting: Challenges and opportunities', in *Transactions on Intelligent Transportation Systems*. Springer Science & Business Media, pp. 87-99.

ICS (2023) 'Shipping and world trade', *Driving Prosperity,* [online]. Available at: https://www.ics-shipping.org/shipping-fact/shipping-and-world-trade-driving-prosperity/ (Accessed: 8 January 2023).

Johnson, M. and Wang, J. (2017) 'Integration of Weather Forecasting Models into Multimodal TMS for Ship Routing Optimization', *Journal of Transport Geography*, 62(9) [online]. Available at: https://www.sciencedirect.com/journal/journal-of-transport-geography/vol/62/suppl/C (Accessed: 1 March 2023).

Kerr, W.R. and Nanda, R. (2009) 'Financing constraints and entrepreneurship', *National Bureau of Economic Research,* 152(8) [online]. Available at: https://www.nber.org/papers/w15498 (Accessed: 12 March 2023).

Kim, J.S., Park, H.J., Shin, W., Park, D. and Han, S.W. (2023) 'WAY: Estimation of Vessel Destination in Worldwide AIS Trajectory', in *Transactions on Aerospace and Electronic Systems*. Springer Science & Business Media, pp. 109-121.

Klimberg, R.K., Sillup, G.P., Boyle, K.J. and Tavva, V. (2010) 'Forecasting performance measures - What are their practical meaning?'. *Advances in Business and Management Forecasting,* 7 [online]. Available at: https://www.emerald.com/insight/content/doi/10.1108/S1477-4070(2010)0000007012/full/html (Accessed: 13 March 2022).

Kwesi-Buor, J. (2019) 'Scenario Analysis and Disaster Preparedness for Port and Maritime Logistics Risk Management', *Accident Analysis & Prevention,* 123(2) [online]. Available at: https://doi.org/10.1016/j.aap.2016.07.013 (Accessed: 3 March 2023).

Makridakis, S. (1993) ‘Accuracy measures: Theoretical and practical concerns’, *International Journal of Forecasting* 9 (4) [online]. Available at: https://flora.insead.edu/fichiersti_wp/Inseadwp1993/93-53.pdf (Accessed: 12 April 2023).

Maskooki, A., Virjonen, P., Kallio, M., (2020) 'Assessing the Prediction Uncertainty in a Route Optimization Model for Autonomous Maritime Logistics', *International Transactions in Operational Research*, 28(4) [online]. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/itor.12882 (Accessed: 12 April 2023).

Meijer, R. (2017) 'Predicting the ETA of a container vessel based on route identification using AIS data', *MOT TU Delft ETA prediction*, 44(4) [online]. Available at: https://repository.tudelft.nl/islandora/object/uuid%3Acba0ef59-dd23-49aa-91d5-bed239e27395 (Accessed: 18 February 2022).

Monios, J. (2014). ‘Institutional Challenges to Intermodal Transport and Logistics’, *Governance in Port Regionalization and Hinterland Integration.* [online]. Available at: https://doi.org/10.4324/9781315588834 (Accessed: 28 August 2022).

Notteboom, T.E. (2004) 'Container Shipping and Ports: An Overview', *Review of Network Economics*, 3(2) [online]. Available at: https://doi.org/10.2202/1446-9022.1045 (Accessed: 12 April 2023).

Nunis, B.T. and V. (2020) 'I’ve lost £20,000 in sales because of shipping delays', *BBC News* [online]. Available at: https://www.bbc.com/news/business-55329573 (Accessed: 19 January 2024).

Paul, S.K., Arisian, S. and Goh, M. (2019) 'Managing sudden transportation disruptions in supply chains under delivery delay and quantity loss', *Annals of Operations Research* [online]. Available at: https://doi.org/10.1007/s10479-017-2684-z (Accessed: 19 January 2024).

Press Gazette. (2021) 'Bharatpe starts giving BMW bikes to IT professionals as joining bonus, extends offer', *Livemint*, 29 July 2021 [online]. Available at: https://www.livemint.com/companies/news/bharatpe-starts-giving-bmw-bikes-other-perks-to-it-professionals-extends-offer-11627559450874.html (Accessed: 24 January 2022).

Savelsbergh, M.W., Song, J. and Wang, S. (2018) 'Modeling and optimization for multimodal freight transportation systems: Challenges and opportunities', *Transportation Research Part C: Emerging Technologies*, 90 [online]. Available at: https://www.sciencedirect.com/journal/transportation-research-part-c-emerging-technologies (Accessed: 21 September 2022).

Smith, J. and Urumsah, D. (2020) 'A Review of Forecasting Techniques for Ship Routing in Multimodal Transportation Management Systems', *International Journal of Logistics Management*, 4 [online]. Available at: https://www.emerald.com/insight/publication/issn/0957-4093/vol/31/iss/4 (Accessed: 23 February 2024).

Urciuoli, L. (2018*)* 'An algorithm for improved ETAs estimations and potential impacts on supply chain decision making', *Procedia Manufacturing (Volume 25),* [online]. Available at: https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6979 (Accessed: 5 January 2024).

Wei, Z. (2019) 'Integration of Multimodal Transportation Management Systems for Efficient Ship Routing', *IEEE Transactions on Intelligent Transportation Systems 62(9) [online]*. Available at: https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6979 (Accessed: 5 January 2024).

Yang, D., Banerjee, A. and Subramanian, N. (2020) 'Multimodal transportation demand forecasting using deep learning', *Transportation Research Part C: Emerging Technologies*, 120 [online]. Available at: https://www.sciencedirect.com/journal/transportation-research-part-c-emerging-technologies (Accessed: 22 September 2022).

Yoon, J.-H., Kim, D-H., Yun, S-W. and Kim, S. (2023) 'Enhancing Container Vessel Arrival Time Prediction through Past Voyage Route Modeling: A Case Study of Busan New Port', *Journal of Marine Science and Engineering*, 11(6) [online]. Available at: https://doi.org/10.3390/jmse11061234 (Accessed: 22 January 2024).

Zhang, Z. and Figliozzi, M.A. (2010) 'A Survey of China's Logistics Industry and the Impacts of Transport Delays on Importers and Exporters', *Transport Reviews*, 30(2) [online]. Available at: https://doi.org/10.1080/01441640902843232 (Accessed: 5 December 2023).

Zhao, B. (2019) 'Data-Driven next Destination Prediction and ETA Improvement for Urban Delivery Fleets', *IET Intelligent Transport Systems*, 13(11) [online]. Available at: https://doi.org/10.1049/iet-its.2019.0148 (Accessed: 5 December 2023).

Zhang, X., Fu, X., Xiao, Z., Xu, H., Wei, X., Koh, J., Ogawa, D. and Qin, Z. (2024) 'Prediction of Vessel Arrival Time to Pilotage Area Using Multi-Data Fusion and Deep Learning', *Journal of Business Venturing*, 29(1) [online]. Available at: https://arxiv.org/abs/2403.09969 (Accessed: 03 April 2024).

Ziebland, S. and McPherson, A. (2006) 'Making sense of qualitative data analysis: an introduction with illustrations from DIPEx', *Medical Education*, 40 [online]. Available at: https://asmepublications.onlinelibrary.wiley.com/doi/10.1111/j.1365-2929.2006.02467.x (Accessed: 03 April 2024).

Zis, T.P.V., Nikolaides, A.V. and Scarlatos, P.D. (2020) 'Ship Weather Routing: A Taxonomy and Survey', *Ocean Engineering*, 213 [online]. Available at: https://doi.org/10.1016/j.oceaneng.2020.107697 (Accessed: 11 December 2023).

Zografos, K. G., and Regan, A. C. (2004) ‘Current Challenges for Intermodal Freight Transport and Logistics in Europe and the United States’, *Transportation Research Record*, 1873(1) [online]. Available at: https://doi.org/10.1016/j.oceaneng.2020.107697 (Accessed: 28 October 2022).

Zontul, M. and Altinkaya, M. (2013) ‘Urban Bus Arrival Time Prediction: A Review of Computational Models’, *International Journal of Recent Technology and Engineering (IJRTE)*, 2(4) [online]. Available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=34e59070faa4472fb94e3d3413c897f5ad409b11 (Accessed: 8 July 2022).

APPENDIX A:

OCTOBER ACCURACY

October Accuracy						
Average of Accuracy	Column Labels					
Row Labels	T-2 Accuracy	T-5 Accuracy	T-10 Accuracy	T-20 Accuracy	T-30 Accuracy	T-40 Accuracy
CN			81.36	95.00	93.75	95.00
HK	89.07	84.61	90.81	79.64	93.89	94.63
ID	90.50	80.79	81.80	87.39	91.34	93.36
IN	94.18	89.63	85.33	86.46	90.73	92.39
JP	88.68	84.67	84.17	81.45	87.42	93.45
KR				89.07	90.00	
SG	90.48	77.90	82.15	78.63	93.35	
TH	85.47	91.25	81.77	76.70	84.52	89.66
TW			85.00	68.13		
US	94.07	88.04	84.83	84.77	89.86	92.25
VN		95.00	95.00	66.67		
Grand Total	89.03	86.55	83.26	82.41	88.44	92.51

APPENDIX B:

NOVEMBER-DECEMBER ACCURACY

Nov-Dec Accuracy						
Average of Accuracy	Column Labels					
Row Labels	T-2 Accuracy	T-5 Accuracy	T-10 Accuracy	T-20 Accuracy	T-30 Accuracy	T-40 Accuracy
CN	94.09	95.00	95.00	89.83	85.00	
CO	86.69	86.54	85.22	93.37		
HK	92.37	82.62	86.02	93.51	84.56	88.92
ID	93.93	89.59	83.60	80.99	89.10	91.58
IN	93.91	86.16	84.81	83.73	86.85	91.39
JP	92.41	87.57	83.93	85.60	89.18	93.07
KR		95.00	87.35	85.00		
LK	93.75	89.00	74.74	61.67		
SG	91.02	81.92	89.53	86.10	91.70	85.00
TH	91.29	89.41	86.62	83.37	88.68	91.16
TW	95.00	87.25	82.81	54.37	78.75	
US	93.20	89.37	86.43	85.03	89.95	91.33
Grand Total	91.97	88.82	85.83	84.19	88.96	91.54

APPENDIX C:

OVERALL ACCURACY OF PAST THREE MONTHS

Overall Accuracy of past 3 months						
Average of Accuracy	**Column Labels**					
Row Labels	**T-2 Accuracy**	**T-5 Accuracy**	**T-10 Accuracy**	**T-20 Accuracy**	**T-30 Accuracy**	**T-40 Accuracy**
CN	**94.09**	**95.00**	**88.18**	**89.92**	**91.15**	**95.00**
CO	**86.69**	**86.54**	**85.22**	**93.37**		
HK	**91.33**	**82.85**	**86.85**	**83.92**	**88.06**	**89.90**
ID	**92.12**	**86.95**	**82.96**	**82.42**	**89.55**	**92.09**
IN	**94.04**	**87.47**	**85.01**	**84.43**	**87.63**	**91.55**
JP	**90.78**	**86.50**	**84.01**	**84.43**	**88.61**	**93.24**
KR		**95.00**	**87.35**	**87.97**	**90.00**	
LK	**93.75**	**89.00**	**74.74**	**61.67**		
SG	**90.77**	**79.97**	**87.15**	**85.10**	**92.26**	**85.00**
TH	**89.99**	**89.64**	**85.97**	**82.60**	**88.39**	**91.00**
TW	**95.00**	**87.25**	**82.95**	**55.92**	**78.75**	
US	**93.47**	**88.96**	**85.87**	**84.95**	**89.93**	**91.69**
VN		**95.00**	**95.00**	**66.67**		
Grand Total	**91.02**	**88.32**	**85.24**	**83.78**	**88.85**	**91.84**

APPENDIX D:

ANCHORAGE AREA RADIUS CALCULATION

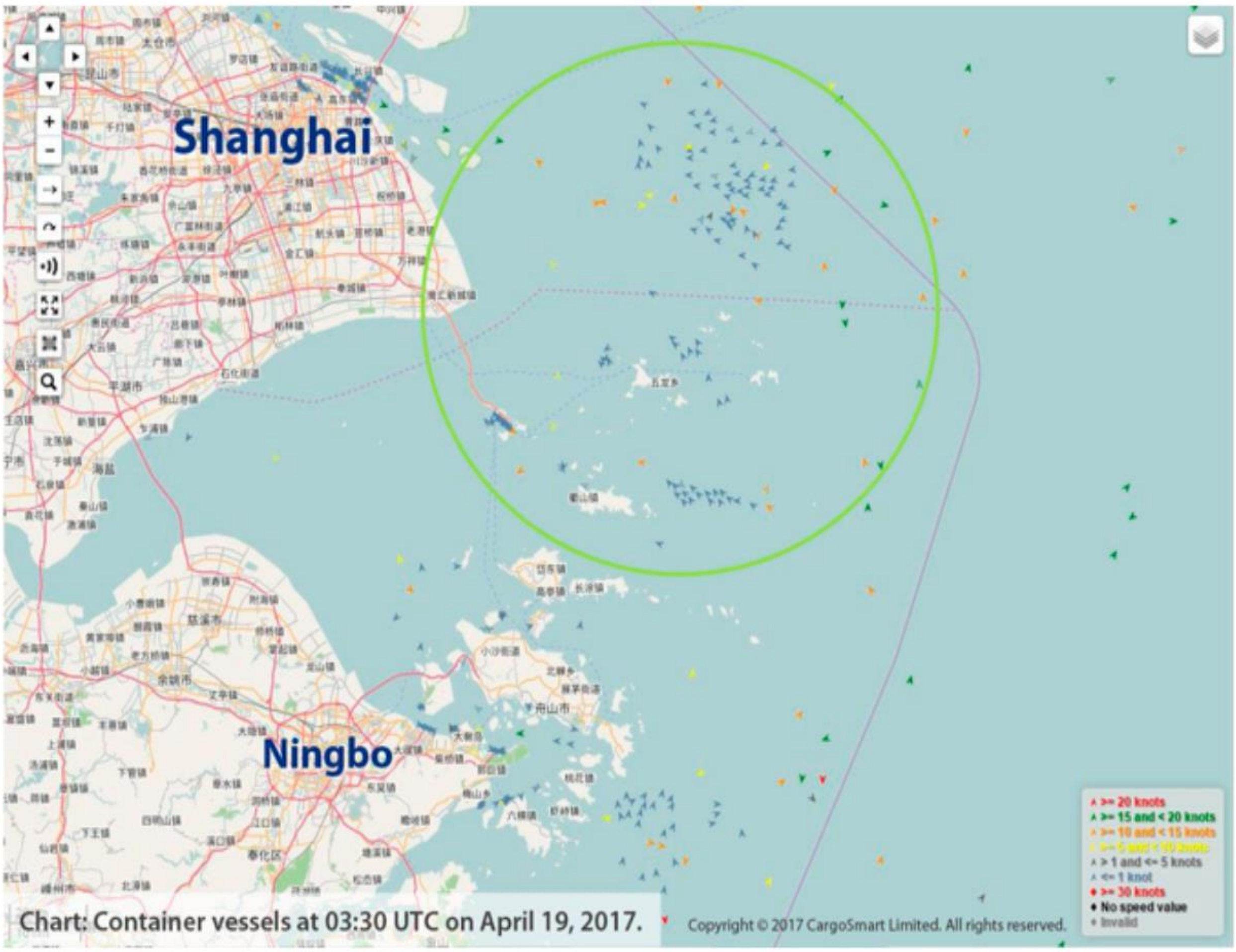
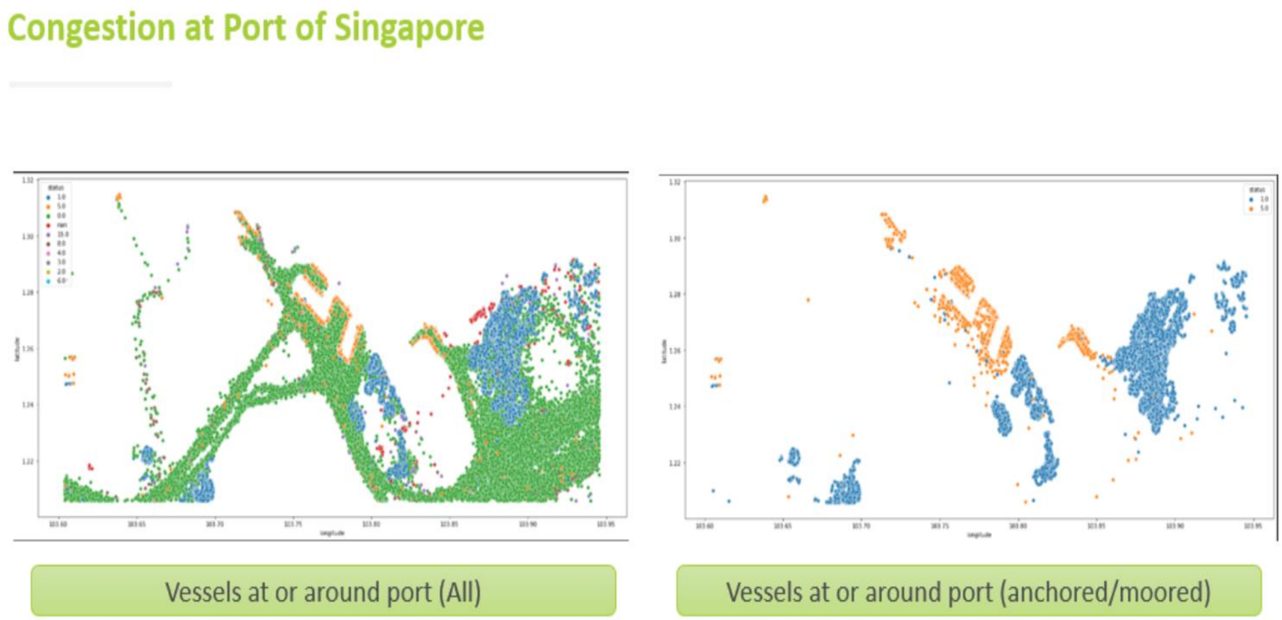


Chart: Container vessels at 03:30 UTC on April 19, 2017.

APPENDIX E:

PORT CONGESTION INDICATORS (PCI)

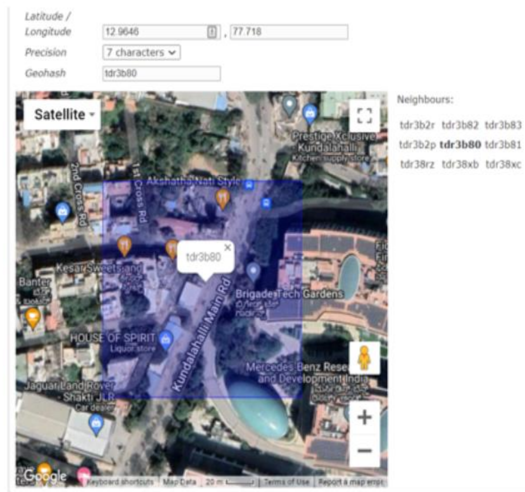


APPENDIX F:

GEOLOCATION AREA

Unique identifier of a specific region on the Earth.

- Geohash algorithm converts its latitude and longitude into a string.
- Points within close geographical proximity will have the same Geohash.
- More precision means a longer geohash and smaller individual regions.



APPENDIX G:

COMPARISON OF PREDICTIONS AND ACCURACIES

days_away	count_timestamp	median_err_dvtr	median_err_stp	mean_err_dvtr	mean_err_stp	accuracy_dvtr	accuracy_stp	p_accuracy_dvtr	p_accuracy_stp	p_acc_median_dvtr	p_acc_median_stp
0	23349	3.300001	7.000000	5.784314	10.951334	NaN	NaN	NaN	NaN	NaN	NaN
1	16845	5.500000	8.599998	8.312457	13.264110	65.364762	44.732877	75.978569	60.945736	83.593742	75.409836
2	10640	7.000000	10.200001	11.322104	13.176791	76.412282	72.548352	80.671669	77.377716	88.225952	82.601181
3	5536	9.000002	12.000000	14.003133	16.018948	80.551204	77.751462	83.031898	80.599701	88.952293	85.173935
4	2994	12.000000	14.599998	18.593306	20.684870	80.631973	78.453261	82.652565	80.650444	88.869713	86.373932
5	1699	18.699997	19.500000	24.734764	27.389349	79.387697	77.175543	81.064987	79.085060	85.943192	85.426353
6	973	19.900009	23.000000	23.598644	30.230009	83.612053	79.006938	84.739815	80.440834	86.928101	85.429558
7	556	21.349991	25.400009	23.651440	31.892813	85.921762	81.016183	86.739967	82.101990	87.899223	85.435760
8	224	35.550003	43.949997	34.879459	47.425457	81.833615	75.299241	82.773605	76.650986	81.888725	78.310822
9	143	27.500000	51.500000	27.065035	52.795105	87.469891	75.557822	88.037750	76.752678	88.090080	77.029434
10	110	21.250008	68.000000	18.819996	56.815453	92.158335	76.326895	92.515015	77.415512	91.584366	72.781693
11	106	20.749985	67.199997	18.775473	58.357536	92.888079	77.894873	93.193237	78.833168	92.410339	75.149811
12	89	10.799988	37.599976	13.315730	51.928085	95.376483	81.969415	95.551208	82.664124	96.467133	87.654724
13	68	13.400009	70.950012	12.189704	71.414711	96.093044	77.110670	96.232666	77.935776	95.811577	78.237572
14	33	10.500000	32.200012	11.209095	52.000000	96.663960	84.523810	96.745354	84.829964	97.013649	90.533562
15	21	17.799988	25.899994	17.961901	24.400002	95.010583	93.222222	95.200974	93.438133	95.363380	93.253456
16	21	18.100006	27.200012	18.376186	26.790478	95.214535	93.023313	95.366150	93.256279	95.415550	93.195419
17	8	18.700012	28.399994	18.287510	28.724995	95.517767	92.959560	95.572205	93.039543	95.483292	93.139877

APPENDIX H:

MODELS GENERATED WITH LESS DATA POINTS

UNLOCODE	PREV_PORT	journeys
JPHIC	JPKWS	9
CNNBG	MYBKI	9
FRDKK	GHTEM	9
JPYAT	JPNAH	9
MYTPP	INPAV	8
PAROD	PEPIO	9
SIKOP	EGALY	8

APPENDIX I:

UNIFORM THRESHOLD IMPOSITION ACROSS O-D PAIRS

		nunique										...	count									
		count	mean	std	min	25%	50%	65%	75%	80%	90%	...	mean	std	min	25%	50%	65%	75%	80%	90%	max
dispBin	LB																					
(0.507, 255.023]	0.507	5322.0	6.219842	6.821740	1.0	1.0	3.0	6.0	9.0	12.0	19.0	...	13.970688	51.181785	1.0	1.0	3.0	6.0	10.0	14.0	29.0	1585.0
(255.023, 468.821]	255.023	3941.0	4.829231	5.901563	1.0	1.0	2.0	3.0	6.0	8.0	15.0	...	8.321238	25.803825	1.0	1.0	2.0	4.0	7.0	9.0	18.0	833.0
(468.821, 674.23]	468.821	3232.0	4.524134	5.733321	1.0	1.0	2.0	3.0	5.0	7.0	14.0	...	7.240099	17.328455	1.0	1.0	2.0	3.0	6.0	8.0	18.0	265.0
(674.23, 924.338]	674.230	2860.0	4.234615	5.438633	1.0	1.0	2.0	3.0	5.0	7.0	13.0	...	6.473776	16.375758	1.0	1.0	2.0	3.0	5.0	7.0	15.0	324.0
(924.338, 1220.644]	924.338	2419.0	3.907813	5.134681	1.0	1.0	1.0	2.0	4.0	6.0	12.0	...	5.347664	11.551457	1.0	1.0	1.0	2.0	4.0	6.0	14.0	200.0
(1220.644, 1634.773]	1220.644	1941.0	3.495621	4.961668	1.0	1.0	1.0	2.0	3.0	4.0	11.0	...	4.981453	12.043892	1.0	1.0	1.0	2.0	3.0	4.0	12.0	156.0
(1634.773, 2332.252]	1634.773	1680.0	3.066071	4.235170	1.0	1.0	1.0	2.0	3.0	4.0	9.0	...	3.926190	9.075648	1.0	1.0	1.0	2.0	3.0	4.0	10.0	158.0
(2332.252, 3804.678]	2332.252	1427.0	2.577435	3.557206	1.0	1.0	1.0	1.0	2.0	3.0	7.0	...	3.146461	7.353448	1.0	1.0	1.0	1.0	2.0	3.0	7.0	158.0
(3804.678, 5568.095]	3804.678	1103.0	2.358114	2.916805	1.0	1.0	1.0	1.0	2.0	3.0	6.0	...	2.632820	4.229605	1.0	1.0	1.0	1.0	2.0	3.0	6.0	60.0
(5568.095, 12423.645]	5568.095	891.0	2.377104	2.898914	1.0	1.0	1.0	1.0	2.0	3.0	7.0	...	2.746352	4.432601	1.0	1.0	1.0	1.0	2.0	3.0	7.0	43.0

APPENDIX I:

PERSONAL EXPERIENCE AND GROWTH FROM THIS STUDY

Embarking on the journey towards a Doctorate in Business Administration (DBA) felt surreal to me. Having completed my postgraduate studies nearly 12 years prior, I found myself completely removed from the habits of academia. Learning something new outside the professional realm posed a monumental challenge.

Yet, with the invaluable guidance of my mentor and the unwavering support of well-wishers, I swiftly regained my academic stride far sooner than anticipated. Soon, I found myself delivering sections of my thesis with unexpected efficiency, which served as a significant morale boost.

As I delved deeper into the research process, much of my initial apprehension dissipated. Initially, "research" seemed an inscrutable and daunting endeavor, accessible only to the intellectually gifted few. However, as I gained clarity on the methodology, it transformed into an achievable and worthy challenge.

My reading and writing habits underwent a transformation. Prior to embarking on this journey, my reading primarily consisted of blogs, newsletters, industry reports, and occasional books to stay abreast of trends. I also dabbled in writing blogs on Supply Chain updates and, occasionally, Hindi poetry. The DBA journey introduced me to scholarly articles, periodicals, journals, and other theses, expanding my avenues of knowledge acquisition.

Consistently writing concisely, a skill I previously struggled to cultivate and was often criticized for, became a habit. Expressing oneself solely through words, devoid of the nuances of tone and body language - which I had always considered integral - proved to be a fascinating challenge. It required frequent revisiting of my writing to ensure the essence of my message was effectively conveyed.

While there were numerous other lessons and experiences throughout this journey, many are difficult to articulate or comprehend without firsthand experience. Nevertheless, for an endeavor that most undertake only once in their lifetime - myself included - this period will undoubtedly stand out as one of the most memorable chapters of my life!!