Østfold University College
Department of Engineering

Insights into Maritime Shipping's Carbon Footprint

Assessing the State of Emissions in Maritime Shipping Through Big Data Analysis

Freddy Rameshchandra Thobhani

Masters in Green Energy Technology (Smart Energy)

Department of Engineering

Lucian Mihet-Popa

August 9, 2023

Insights into Maritime Shipping's Carbon Footprint

Assessing the State of Emissions in Maritime Shipping Through Big Data Analysis

Freddy Rameshchandra Thobhani

Abstract

This is the abstract.

Acknowledgements

I would like to thank...

Contents

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

In the 21st century, climate change is the biggest challenge faced by humanity. It poses a substantial danger to the survival of the inhabitants of our planet. Human activities such as deforestation, extraction and burning of fossil fuels have led to a rise in global temperatures. The consequences of such activities are an increase in sea levels, extreme weather events, and loss of biodiversity. There is an urgent and undeniable need to reduce greenhouse gas emissions and transition to a sustainable and low-carbon future.

Maritime shipping is essential to the global economy. It accounts for transporting 90% of the world's goods by volume. It is also a major source of greenhouse gas emissions, with the International Maritime Organization (IMO) estimating that maritime shipping accounts for 3% of global carbon dioxide emissions. While 3% may seem small, it is important to note that this is a rapidly growing sector. Without action, maritime shipping contribution to carbon emissions can increase by up to 10-13% in the next few decades. Due to this fact, there is a growing global effort to reduce emissions from this sector. [10].

The European Green Deal is a significant initiative by the European Union to make Europe the world's first climate-neutral continent by 2050 . It aims to transform various sectors, including shipping, to reduce environmental impacts. Through meticulous analysis of total carbon emissions within the maritime sector using advanced data techniques, this research can significantly contributes to the goals of the European Green Deal. It can aid in formulating policies, monitoring progress, and promoting sustainable practices, aligning with the Green Deal's emphasis on innovation, global cooperation, and eco-conscious industrial transformation [20]. This study's alignment with the European Green Deal underscores its relevance and importance within the broader context of sustainability-focused endeavors.

In accordance with Sustainable Development Goal 13, in 2018, the initial strategy was adopted by IMO's Environmental Protection Committee (MEPC), during its 72nd session at IMO Headquarters in London, United Kingdom. According to this strategy, the IMO will work towards reducing the total annual greenhouse gas emissions from international shipping by at least 50% by 2050 compared to 2008 [27]. In the 76th session of MEPC in 2021, serval mandatory measures were adopted to reduce greenhouse gas emissions from international shipping, which will help in achieving the goal of reducing emissions by 50% by 2050 [28]. One of the important measures is the Carbon Intensity Indicator (CII).

Maritime shipping is a complex and highly volatile system, generating very large data sets. Big data analytics can be used to understand the complex system and make informed decisions. It can facilitate operations such as monitoring of emissions and predictive analysis of vessel performance. This can help in reducing emissions and improving the efficiency of the maritime sector [31].

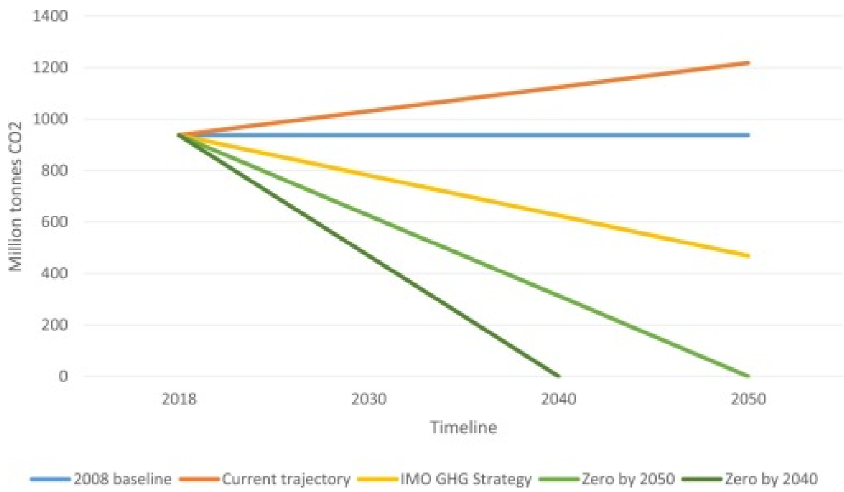

Figure 1.1: Emission trajectories for different levels of ambition for emission reduction targets

1.2 Big Data Analysis

Big data analytics is where advanced analytic techniques operate on big data sets. Hence, big data analytics is really about two things — *big data* and *analytics*.

1.2.1 Big Data

As the name suggests, big data is a large amount of data. There are other important attributes of big data. These are: data variety and data velocity.

Thus we can define big data using 3 V's: *volume*, *variety*, and *velocity* as showin in figure 1.2.

Beyond these three V's, Big Data is also about how complicated the computing problem is. Given the number of variables and number of data points for analysing the maritime shipping data. It is a very complicated problem. Thus, in addition to the three V's identified by IBM, it would also be necessary to take complexity into account as shown in figure 1.3 [13].

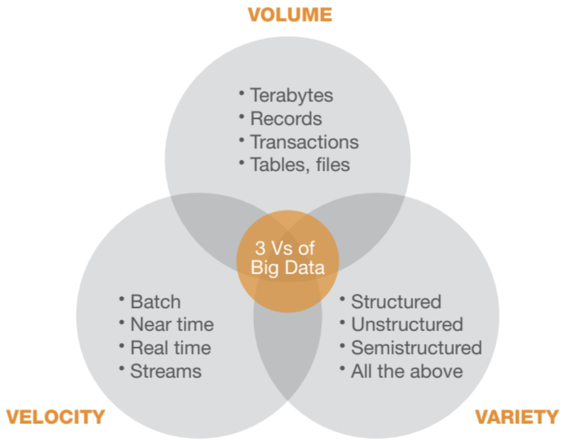

Figure 1.2: Big Data: 3 V's [11]

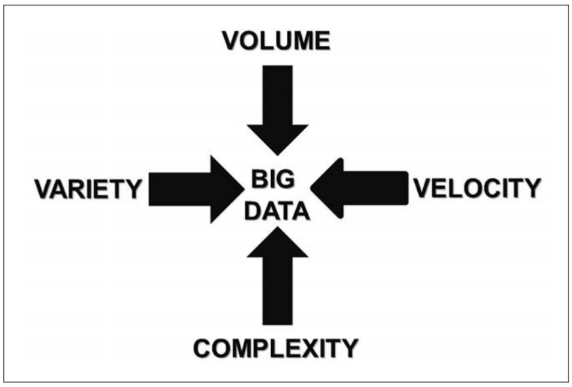

Figure 1.3: Big Data: Beyong 3 V's - volume, velocity, variety, and complexity

1.2.2 What is Big Data Analytics?

Big data analytics is the process of examining large and varied data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful information that can help organizations make more-informed business decisions.

Thus, Data analytics revolves around deriving valuable knowledge and meaningful insights from extensive sets of data. This process involves crafting hypotheses, often rooted in gathered experiences and uncovering correlations between variables, sometimes even through serendipitous discoveries. Data analytics can be classified into four distinct types [16]:

1. Descriptive Analytics

Descriptive analytics focuses on explaining past events and presenting them in a comprehensible manner. The collected data is structured into visual aids like bar charts, graphs, pie charts, maps, and scatter diagrams, facilitating easy interpretation that offers insights into the data's implications. This mode of data representation is often termed a dashboard, reminiscent of a car's dashboard that provides details such as speed, engine status, fuel levels, and distance traveled. A classic instance of descriptive analytics involves displaying population census data, which categorizes a nation's population by gender, age brackets, education, income, population density, and similar criteria [16].

2. Predictive Analytics

Predictive analytics extends beyond existing data to forecast forthcoming events. It anticipates what is likely to occur in the immediate future. Techniques like time series analysis utilizing statistical methods, neural networks, and machine learning algorithms are employed for this extrapolation. A significant application of predictive analytics is seen in marketing, where it understands customer preferences and needs. For instance, when purchasing shoes online, an advertisement for socks may appear. Another prevalent application is in orchestrating election campaigns. This involves gathering diverse data, such as the demographics of voters in different areas and their perceived needs like infrastructure and local concerns [16].

3. Prescriptive Analytics

This process detects opportunities for enhancing existing solutions by analyzing collected data. Essentially, it guides us on the actions to undertake in order to accomplish a particular objective. An illustrative instance is observed in the aviation industry where airlines determine seat pricing through analysis of historical travel patterns, popular travel origins and destinations, significant events, holidays, and more. This approach is employed to optimize profit generation [16].

4. Exploratory or Discovery Analytics

This process uncovers unforeseen connections among variables within extensive datasets. The collection and analysis of data from diverse sources opens up new avenues for gaining insights and making serendipitous discoveries. One of its major applications involves the identification of patterns in customers' behavior by companies through sources like feedback, tweets, blogs, Facebook data, emails, and sales trends. By deciphering customer behavior, companies can potentially predict actions like renewing a magazine subscription, switching mobile phone service providers, or canceling a hotel reservation. Armed with this information, companies can devise appealing offers aimed at altering the anticipated course of action by the customer [16].

1.3 Indicators

Indicators play a crucial role in assessing and monitoring carbon emissions in the maritime shipping industry. They provide valuable insights into the environmental performance of vessels, facilitate comparisons between different ships or fleets, and help track progress towards emission reduction targets. By measuring various aspects of emissions and energy efficiency, these indicators enable stakeholders to identify opportunities for improvement and implement effective strategies to mitigate the environmental impact of shipping operations.

In this section, we will discuss several key indicators commonly used in the monitoring and evaluation of carbon emissions in maritime shipping. These indicators cover a range of factors, including carbon intensity, energy efficiency, fuel consumption, and cargo transport work. Each indicator offers a unique perspective on emissions, providing researchers, policymakers, and industry stakeholders with valuable information to support decision-making and foster sustainable practices.

It is important to note that the selection and use of indicators may vary depending on the specific research objectives, data availability, and regulatory frameworks in place. The combination of different indicators allows for a comprehensive assessment of emissions and enables a deeper understanding of the efficiency and environmental performance of shipping activities.

Below is the list of indicators discussed in this section:

1. Carbon Intensity Indicator (CII)
2. Energy Efficiency Operational Indicator (EEOI)
3. Energy Efficiency Design Index (EEDI)
4. Energy Efficiency eXisting ship Index (EEXI)

1.3.1 Carbon Intensity Indicator (CII)

The International Maritime Organization (IMO) has introduced a new carbon intensity (CII) measure for ships, which is a more accurate way to evaluate a vessel's environmental impact than total carbon emissions. CII is calculated using the Annual Efficiency Ratio (AER) formula, taking into account a ship's fuel consumption, CO2 emission factor, annual distance sailed, and design deadweight.

To calculate CII in the most basic form:

$$\text{CII} = \frac{\text{Carbon Emission}}{\text{Distance Travelled} \times \text{Cargo Capacity}} \tag{1.1}$$

Vessels are rated A to E based on their CII results, and those with a D or E rating for three consecutive years or an E rating in one year must submit a corrective action plan. The IMO will enforce CII regulations for all ships over 5,000 GT and require an enhanced Ship Energy Efficiency Management Plan (SEEMP) with CII-related content from January 2023. The

SEEMP must include the ship's required annual operational CII target and an implementation plan to achieve it over the next three years. [2].

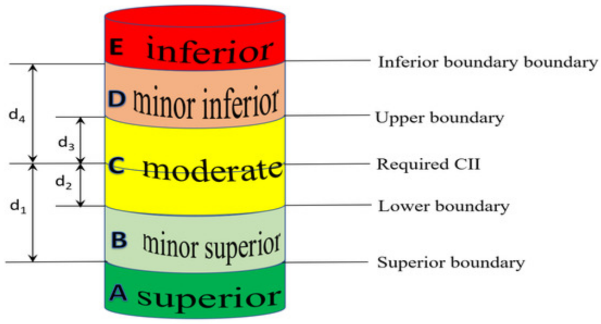

Figure 1.4: Schematic diagram of the CII ratings and boundaries. [26]

The article [18], discusses the increasing popularity of Energy and Car- bon Intensity in- dicators among policy makers, which are calculated as units of energy or mass of emissions per unit of Gross Domestic Product (GDP). These indicators are gaining momen- tum due to support from think tanks, public organizations, consulting groups, and academics focused on energy and climate change policy development. The ways in which intensity in- dicators are framed and perceived in public debates generated by these intermediaries are important as they can influence policy making and the development of better sets of indicators to assess how well countries are addressing climate change and resource efficiency policies. Intensity indicators are appealing for emerging economies as they are not incompatible with high rates of economic growth and do not imply the imposition of absolute emission/energy caps.

1.3.2 Energy Efficiency Operational Indicator (EEOI)

The Energy Efficiency Operational Indicator (EEOI) is a tool used to measure the CO2 gas emissions per unit of transport work, indicating the operational efficiency of a ship. The EEOI is calculated annually and is subject to changes after each voyage due to various external factors, such as navigation conditions, sea area, weather, temperature, and cargo weight.

The EEOI provides an accurate measure for each voyage, and its unit depends on the type of cargo or transport work, such as tons CO2/(tons/nautical miles), tons CO2/(TEU/nautical miles), or tons CO2/(person/nautical miles). The formula for calculating the EEOI is represented by formula (1.2), where a lower value indicates a more energy-efficient ship [15].

$$\text{EEOI} = \frac{\text{Carbon Emission}}{\text{Performed Transport Work}} \tag{1.2}$$

For the calculation of EEOI for a specific voyage, formula (1.3) is used.

$$\text{EEOI} = \frac{\sum_j F_{Cj} \cdot C_{Fj}}{m_{\text{cargo}} \cdot D_j} \tag{1.3}$$

However, when dealing with a large number of ships, formula (1.3) is expressed as equation (1.4), taking into account parameters such as fuel type, voyage number, fuel consumption, fuel-to-CO2 conversion factor, cargo weight, and distance traveled [25].

$$\text{Average}_{\text{EEOI}} = \frac{\sum_i \sum_j (F_{C_i} \cdot C_{F_j})}{\sum_i (m_{\text{cargo},i} \cdot D_i)} \tag{1.4}$$

where:

j : Fuel type used
i : Navigation voyage number
FC_{ij} : Mass of consumed fuel j at voyage i
CF_j : Fuel mass to CO2 mass conversion factor with fuel j
m_{cargo} : Weight of cargo carried (tons) on ship
D_i : Distance of voyage i (nautical miles)

The fuel-to-CO2 conversion factor (CF) is a non-dimensional factor that converts fuel consumption, measured in grams, to CO2 gas emissions, also measured in grams, based on the carbon content. The below table

1.1 is showed the certain value of CF follows the type of fuel.

No.	Type of fuel	Reference	Carbon content	CF (t-CO2/t-Fuel)
1	Diesel/gas oil	ISO 8217 Grades DMX through DMC	0.875	3.206000
2	Light fuel oil (LFO)	ISO 8217 Grades RMA through RMD	0.86	3.151040
3	Heavy fuel oil (HFO)	ISO 8217 Grades RME through RMK	0.85	3.114400
4	Liquefied petroleum gas (LPG)	Propane, butane	0.819, 0.827	3.000000, 3.030000
5	Liquefied natural gas (LNG)		0.75	2.750000

Table 1.1: The value of CF (t-CO2/t-Fuel) [25].

1.3.3 Energy Efficiency Design Index (EEDI)

Energy Efficiency Design Index (EEDI) is a legislation proposed by the International Maritime Organization (IMO) to estimate the energy efficiency of ships and calculate their CO2 emissions per unit of transport work done during the ship design phase. EEDI is based on a complex formula that takes into account the ship's emissions, capacity, and speed, and the lower the ship's EEDI index, the less CO2 emissions it produces.

EEDI is a non-prescriptive mechanism that allows the shipping industry to use the latest technologies for designing commercial vessels as long as they meet the required energy efficiency levels and parameters. It lays down a minimum energy efficiency level, per capacity mile, for different ship types and sizes, including tankers, bulk carriers, gas carriers, general cargo ships, container ships, refrigerated cargo carriers, and combination carriers.

The EEDI formula has two components: attained EEDI and required EEDI. The attained EEDI is calculated using a complex formulation based on the vessel's emissions, capacity, and speed, while the required EEDI is the minimum level of energy efficiency that a ship must meet as per its ship type and size. The attained EEDI is verified based on the ship's design and construction, and the required EEDI is the target that the ship must achieve during its operation [17].

EEDI calculation module as part of Marpol Annex VI, following the directive MEPC.1/Circ.681 at the MEPC meeting conducted by the IMO in 2011. This regulation came into effect on January 1, 2013. The EEDI formula (Equation 1) specified by IMO (2011) is represented by the equation (1.5) [24].

$$EEDI = \frac{P \cdot \mathrm{SFC} \cdot \mathrm{Cf}}{DWT \cdot \mathrm{Vref}} \tag{1.5}$$

where:

P : 70% of the power of the engine (main and auxiliary) in kW
SFC : Amount of fuel burned by the engines in kW (specific fuel consumption)
Cf : Emission rate of fuel used by the ship (presented in Table 1)
DWT : Ship's capacity (in tons)
V_{ref} : Speed of the ship (in knots)

1.3.4 Energy Efficiency eXisting ship Index (EEXI)

The Energy Efficiency Existing Ship Index (EEXI) is a regulation introduced by the International Maritime Organization (IMO) aimed at improving the energy efficiency of existing ships. It is part of the broader effort to reduce greenhouse gas emissions from the shipping industry and combat climate change. The EEXI is designed to complement the Energy Efficiency Design Index (EEDI), which focuses on new ship designs [4].

The EEXI is part of a comprehensive framework that includes short-term, mid-term, and long-term measures. The short-term measures focus on technical and operational improvements, such as retrofitting ships with energy-efficient technologies. The mid-term measures involve market-based mechanisms to incentivize emission reductions, while the long-term measures explore alternative fuels and propulsion systems [3].

The calculation of the Energy Efficiency Existing Ship Index (EEXI) can be optimized by considering the ship's maximum continuous rating (MCR) at 100% capacity. This approach ensures that improvements in technical efficiency closely align with the ship's actual operational fuel use. By accounting for the engine power limits (EPLs) within the engine margin, which have minimal impact on ship operations, a more accurate assessment of energy efficiency can be achieved. Currently, the proposed calculation methods for the EEXI involve using either 75% of the limited MCR (MCRlim), similar to the Energy Efficiency Design Index (EEDI), or a higher value of 87% MCRlim, which only considers the engine margin. However, utilizing ship characteristics data from IHS Markit and applying the appropriate calculation method, the attained EEXI score can still be estimated, providing valuable insights into a ship's energy performance [19].

$$\text{Attained EEXI} = 3.1144 \times \frac{MESFOC \times \sum_{i=1}^{nME} P_{ME,i} + AESFOC \times P_{AE}}{\text{Capacity} \times \text{Vref}} \tag{1.6}$$

The estimation of the attained EEXI score using ship characteristics data from IHS Markit, as outlined in Equation (1.6), contributes to a more comprehensive understanding of a ship's energy efficiency. This information supports decision-making processes related to optimizing operational fuel consumption, implementing retrofit measures, and promoting environmental sustainability in the maritime industry. By calculating the EEXI at 100% MCR, the assessment takes into account the EPLs within the engine margin, which are not expected to significantly affect ship operations. This approach ensures that technical efficiency improvements are properly aligned with the ship's actual operational fuel use, enabling informed decision-making for enhancing operational fuel consumption and reducing environmental impact. Through the utilization of ship characteristics data and the appropriate calculation method, the attained EEXI score serves as a valuable tool for assessing energy efficiency and driving advancements in the maritime sector [19].

1.4 Problem Statement

Carbon emissions from maritime shipping have been identified as a major contributor to global greenhouse gas emissions, with the International Maritime Organization estimating that shipping is responsible for around 3% of global CO2 emissions [10]. To address this issue, the shipping industry has set targets to reduce its carbon footprint, and governments and international organizations have introduced policies and regulations to encourage emissions reduction.

However, measuring and monitoring carbon emissions from maritime shipping can be challenging due to the complexity of the industry and the lack of reliable data. The Energy Efficiency Operational Indicator (EEOI) and the Carbon Intensity Indicator (CII) have been proposed as two metrics to assess the carbon efficiency of ships and enable comparison between different vessels and fleets [32, 3]. However, there is a need to better understand the relationship between EEOI and carbon emissions, as well as to identify the factors that influence this metrics.

Therefore, the aim of this thesis is to conduct a big data analysis of carbon emissions in maritime shipping, using EEXI as the main metric. Specifically, the study will:

- Calculate EEOI for a sample of vessels using real-world data on fuel consumption and other operational parameters.
- Analyze the relationship between EEOI, CII, and carbon emissions, using statistical methods and machine learning algorithms.
- Identify the factors that influence EEOI and CII, such as vessel age, size, speed, and route, and examine their impact on carbon emissions.
- Evaluate the usefulness of EEOI and CII as metrics for monitoring and reducing carbon emissions in maritime shipping, and recommend potential improvements to these metrics.

Overall, the findings of this thesis will contribute to a better understanding of the carbon efficiency of maritime shipping and inform the development of policies and strategies for emissions reduction in this sector.

1.5 Research Question

This theis will focus on answering following research questions:

1. What is the relationship between vessel age and carbon emissions in maritime shipping?
2. How do shipping routes affect carbon emissions in maritime shipping?
3. What role do fuel types and engine technologies play in carbon emissions in maritime shipping?
4. How can EEOI and CII be used to monitor and reduce carbon emissions in maritime shipping?

1.6 Report Outline

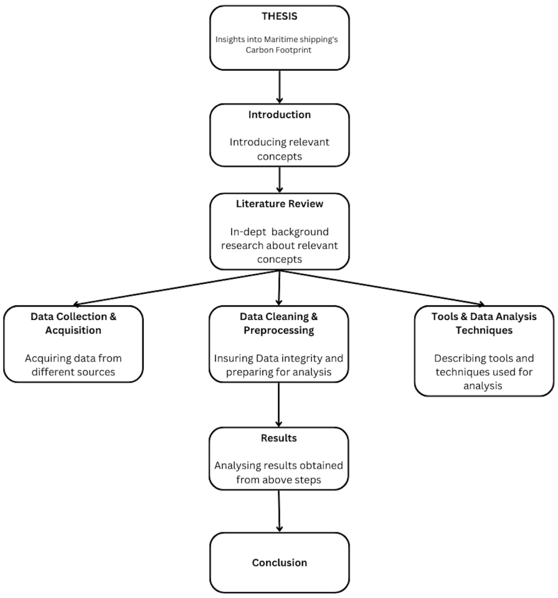

Figure 1.5: Outline of the thesis

Chapter 2: Litrature Review: This chapter covers the background information and literature review of the thesis. this section covers comprehensive review of existing papers and research related to carbon emissions in maritime shipping. The aim is to provide a comprehensive overview of the current state of research in this field and identify any gaps or opportunities for further exploration.

Chapter 3: Data Collection and Understanding: In this chapter, the focus will be on gathering and understanding the data required to perform analysis to understand emissions in martime shipping. Various data sources will be explored, including industry databases, research publications, and government reports. The aim is to acquire a comprehensive dataset that covers different aspects of carbon emissions in the maritime sector. Additionally, this chapter will delve into the intricacies of the collected data, understanding its structure, variables, and potential limitations.

Chapter 4: Data Cleaning and Preprocessing: Before conducting any data analysis, it is crucial to ensure the quality and integrity of the dataset. This chapter will discuss the steps involved in cleaning and preprocessing the data. This process may involve handling missing values, dealing with outliers, standardizing formats, and resolving inconsistencies. By performing these necessary data cleaning procedures, the dataset will be prepared for further analysis, ensuring reliable and accurate results.

Chapter 5: Data Analysis Techniques: In this chapter, various data analysis techniques specific to big data will be explored and applied to the cleaned dataset. These techniques may include statistical analysis, machine learning algorithms, and data visualization methods. The goal is to extract meaningful insights and patterns from the data to gain a comprehensive understanding of carbon emissions in maritime shipping. Additionally, this chapter will discuss the tools and technologies utilized for data analysis and highlight any specific challenges encountered during the process.

Chapter 6: Evaluation of Results: After performing the data analysis, this chapter will focus on evaluating and interpreting the obtained results. The findings will be compared against existing literature, industry benchmarks, and regulatory standards to assess the significance and implications of the analysis. The strengths and limitations of the analysis approach will be discussed, and recommendations for future research or practical applications will be provided. This chapter aims to provide a comprehensive evaluation of the insights gained from the data analysis and their potential impact on the maritime shipping industry.

Conclusion: The conclusion chapter will summarize the key findings and contributions of the thesis. It will highlight the significance of the conducted big data analysis on emissions in maritime shipping and its implications for sustainability and environmental initiatives. The conclusion will also discuss any potential limitations or challenges encountered during the research and suggest avenues for further exploration in this field.

Chapter 2

Literature Review

In response to the urgent need to reduce carbon emissions and combat climate change, researchers and industry stakeholders have focused on developing and implementing strategies to reduce carbon emissions in maritime shipping.

Figure 2.1 shows that the number of publications on energy efficiency and emission reduction in the maritime industry has grown exponentially since 2016. The number of publications from 2006 to 2015 was 76, while from 2016 to 2021, there were 260 publications, indicating a significant increase in interest in decarbonization in the maritime industry. [8]

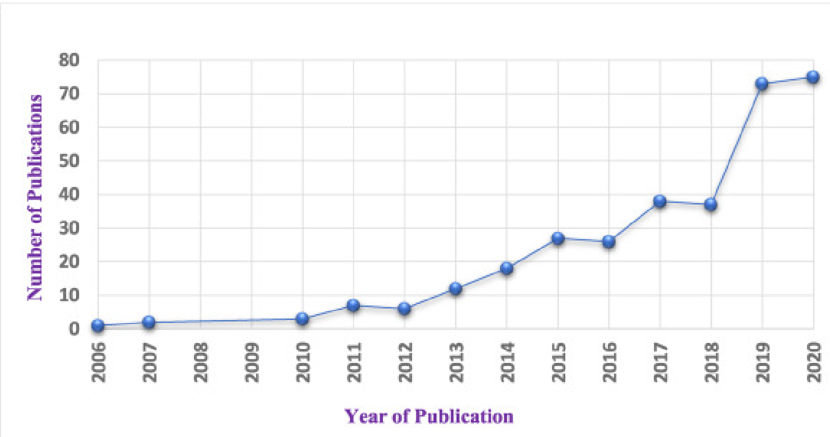

Figure 2.1: Number of publications per year in energy efficiency and emission reduction in the maritime domain

One promising area of research is the use of big data analysis to measure and improve carbon efficiency in maritime shipping. Big data analysis involves the collection and analysis of large and complex data sets to identify patterns, trends, and insights. In the context of maritime shipping, big data analysis can be used to measure carbon emissions and identify opportunities for improvement.

The purpose of this literatsure review is to examine the current state of research on carbon emissions in maritime shipping, with a focus on the Energy The Energy Efficiency eXisting ship Index (EEXI) and Carbon Intensity Indicator (CII) as key metrics for measuring carbon efficiency. The review will provide an overview of the current state of research on these metrics, their strengths and limitations, and their relevance for the maritime shipping industry.

The review will begin by exploring the importance of reducing carbon emissions in maritime shipping and the regulatory and policy frameworks that have been established to address this issue. It will then provide an overview of the EEXI and CII metrics, including their definitions, methodologies for calculating them, and their role in measuring carbon efficiency.

The literature review will also examine the current research on the relationship between EEXI, CII, and carbon emissions in maritime shipping, with a particular focus on the use of big data analysis to measure and improve carbon efficiency. It will explore the potential for big data analysis to provide more accurate and comprehensive data on carbon emissions, and to identify opportunities for operational and technological improvements.

Overall, this literature review will provide a comprehensive overview of the current state of research on carbon emissions in maritime shipping, with a focus on the EEXI and CII metrics and the potential for big data analysis to guide and inform strategies for improving carbon efficiency in the industry.

2.1 Litrature Review

Review by Issa, Ilinca, and Martini [7] shows that Maritime shipping is a crucial aspect of global trade and the global economy, with over 85% of the volume of global trade in goods transported by sea. However, maritime transport also has significant environmental impacts, including carbon emissions. Approximately 3.3% of the world's carbon dioxide (CO2) emissions are attributable to maritime transport, with emissions from marine diesel oil (MDO), marine fuel oil (MFO), and heavy fuel oil (HFO) all contributing to the problem. Reducing carbon emissions in the maritime shipping industry is a significant challenge, but there are a range of strategies that can be used to achieve this goal. Alternative fuels, energy efficiency improvements, and operational measures all have the potential to reduce emissions, but they also have significant economic and resource constraints.

Paper by Grzelakowski, Herdzik, and Skiba [6] mentions that Despite its significant contribution to global economic growth, maritime transport also generates negative externalities, primarily in the form of greenhouse gas (GHG) emissions. They discus how digitalization and the use of artificial intelligence (AI) are being explored as potential ways to reduce emissions in maritime shipping. AI algorithms can optimize shipping routes, reduce fuel consumption, and minimize emissions. Additionally, digitalization can enable better data collection and analysis, which can facilitate more accurate emissions reporting and monitoring.

According to Kao, Chung, and Chen [9], the use of automatic identification system (AIS) to estimate ship emissions, which is an advantage due to the system's ability to provide real-time navigational information. Studies have been conducted utilizing AIS data to estimate ship emissions in different regions, such as Hong Kong and the Pearl River Delta, Las Palmas Port, Qingdao Port, Tianjin port, Naples port, and unidentified vessels with missing ship parameters. The studies have focused on macro-scale spatial and temporal resolution, high-resolution ship emission inventory, high temporal-spatial ship emission inventory, higher spatial-temporal resolution, and real-time ship emission monitoring. It proposes simulation model based on AIS data, specification and what-if scenarios as shown in Figure 2.2

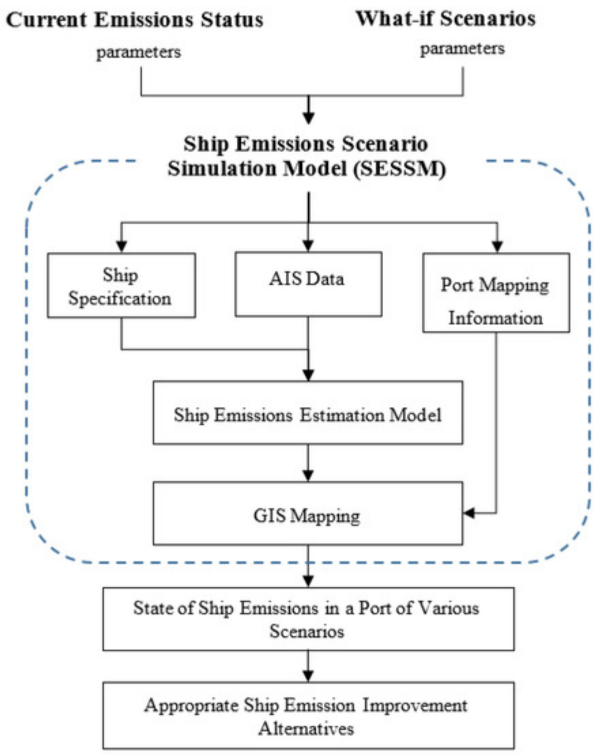

Figure 2.2: Simulation framework

Research by Sou et al. [21], discusses the need for carbon intensity indicators (CIIs) as performance monitoring tools in the shipping industry, particularly for tracking energy efficiency trends and progress towards climate targets. The review highlights the lack of consensus on suitable CIIs, as proposed by various countries to the International Maritime Organization (IMO), and the need for a more comprehensive understanding of global progress towards carbon intensity targets from both demand and supply side indicators. The study aims to address this issue by analyzing CIIs for shipping and the factors that influence the carbon intensity of shipping at the global level. Index decomposition analysis (IDA) is used to quantify the contribution of various factors, including energy efficiency, to changes in carbon intensity from 2012 to 2018.

According to report by Stevenson [23], The International Maritime Organization will introduce Energy Efficiency eXisting ship Index (EEXI) and Carbon Intensity Indicator (CII) regulations in 2023 as part of the wider decarbonisation goals for shipping. More than three-quarters of the existing fleet will not initially meet EEXI baselines and will need to take action to achieve compliance, with overridable engine power limitations (oEPL) expected to be a popular option. However, the effect on vessel operations over a year will be quite small due to the relatively small number of hours where steaming speeds would exceed oEPL limits. The compliance with EEXI can result in modest improvements in AER, CII, and annual CO_2. emissions.

The International Maritime Organization (IMO) has made maritime decarbonization a pri-

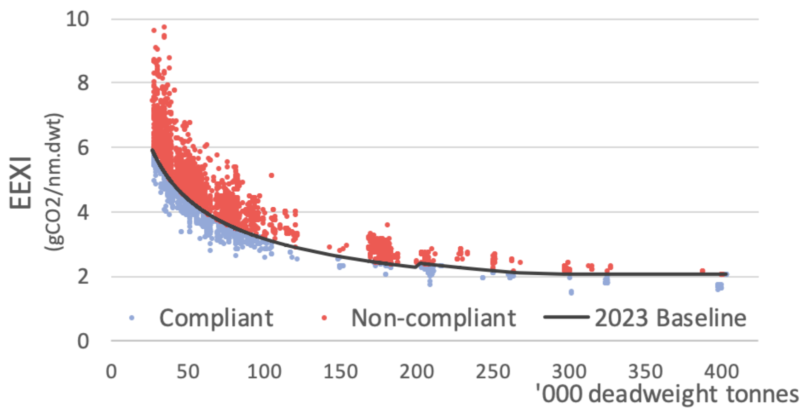

Figure 2.3: EEXI BULKER ESTIMATES VS. 2023 BASELINE

ority, setting targets to reduce greenhouse gas emissions from ships. To achieve these targets, the IMO has adopted mandatory measures, including the carbon intensity indicator (CII), which measures carbon emissions per unit transport work for each ship. But Wang, Psaraftis, and Qi [29] argues there are potential paradoxes with the CII, as it may increase carbon emissions in some situations. There are at least four potential versions of the CII, including supply-based, demand-based, distance-based, and sailing time-based, but the IMO has not yet agreed on which to use. More elaborate models and indicators should be developed to analyze the potential impacts of the CII and achieve utmost carbon emissions reduction.

In "Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions" [12], author Munim et al. explains how Big data and artificial intelligence (AI) have become essential components of data-driven decision-making in most industries. However, the maritime industry still relies on intuition more than on data, mainly because of the vast size of its network and planning problems. The maritime industry generates large amounts of data that, if appropriately utilised in decision-making, can improve maritime safety, reduce environmental impacts, and minimise cost. In this review, we focus on studies that deal with big data and AI applications within the maritime context to map the conceptual structure of the field and identify future research avenues. AIS data to investigate the impact of speed reduction on fuel consumption and carbon emissions in the shipping industry. The study found that a 10% reduction in ship speed could result in a 17% reduction in fuel consumption and a corresponding reduction in carbon emissions. The authors suggested that reducing ship speed is an effective way to reduce fuel consumption and carbon emissions in the maritime industry.

The study by Acomi and Acomi [1], delves into the realm of maritime environmental conservation, emphasizing the Energy Efficiency Operational Index (EEOI) as a pivotal tool. Addressing concerns of marine pollution encompassing water and air aspects, the paper underscores international efforts for emission reduction in shipping. The EEOI, designed to measure carbon emissions per unit of transport work, aids ship-owners and operators in enhancing energy efficiency during operational voyages. The research demonstrates how commercial software and a custom-developed program estimate EEOI values pre-voyage and onboard, revealing the influence of unpredictable factors on energy efficiency. This analysis not only showcases the EEOI's significance in curbing emissions but also highlights its vital role in the

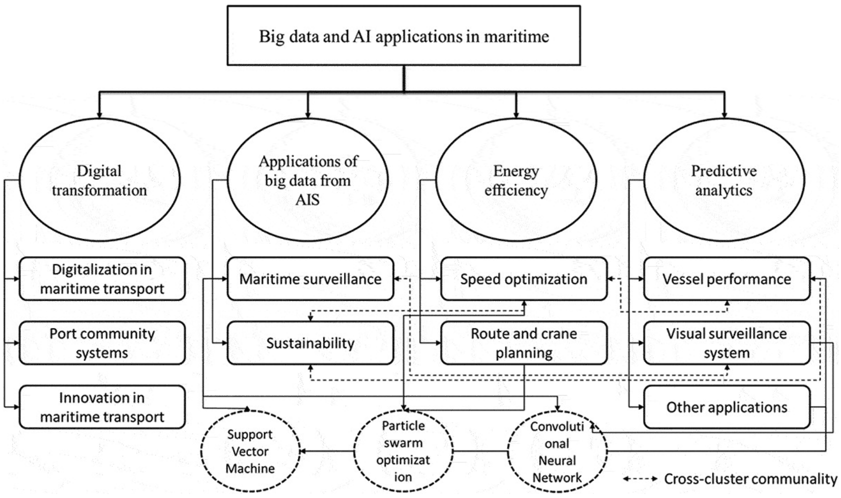

Figure 2.4: Application of Big dada and AI in maritime industry

broader context of maritime sustainability and environmental protection.

2.1.1 Conclusion

In this section, we have reviewed the literature on carbon emissions in maritime shipping, with a focus on the EEOI and CII metrics and the potential for big data analysis to guide and inform strategies for improving carbon efficiency in the industry.

In conclusion, the literature reviewed emphasizes the importance of addressing the significant environmental impact of carbon emissions in the maritime shipping industry. While there are various strategies to reduce emissions, such as alternative fuels, energy efficiency improvements, and operational measures, they have significant economic and resource constraints. Digitalization and the use of artificial intelligence (AI) are being explored as potential ways to reduce emissions by optimizing shipping routes, reducing fuel consumption, and minimizing emissions. Furthermore, the use of automatic identification system (AIS) data can facilitate real-time emissions monitoring, while carbon intensity indicators (CIIs) can be used as performance monitoring tools. The International Maritime Organization (IMO) has made maritime decarbonization a priority by setting targets to reduce greenhouse gas emissions from ships and introducing regulations like EEXI and CII. However, potential paradoxes with the CII and lack of consensus on suitable CIIs highlight the need for more elaborate models and indicators to achieve utmost carbon emissions reduction. Big data and AI applications have the potential to improve maritime safety, reduce environmental impacts, and minimize costs. Overall, more research is needed to address the challenges of monitoring and reducing carbon emissions in the maritime shipping industry while meeting global trade demands.

Chapter 3

Data Collection and Acquisition

data collection and understanding play a pivotal role in extracting meaningful insights and deriving accurate conclusions. The success of any analytical endeavor heavily relies on the quality, comprehensiveness, and relevance of the data used. In the case of carbon emissions in maritime shipping, gathering and understanding the data is crucial for capturing the intricacies of this complex domain. By exploring various data sources from industry databases, a comprehensive dataset can be acquired, encompassing diverse dimensions of carbon emissions in the maritime sector. Furthermore, understanding the structure, variables, and limitations of the collected data is essential for ensuring the validity and reliability of subsequent analyses. This includes examining the completeness of data, identifying any biases or data gaps, and verifying the accuracy of measurements. Ultimately, a thorough and informed understanding of the data sets the foundation for conducting rigorous data analysis and generating actionable insights for addressing carbon emission challenges in maritime shipping.

Collaboration with Astrup Fearnleys Code has been instrumental in enhancing my research on carbon emissions in maritime shipping. As a leading firm in the maritime shipping industry, their expertise and industry connections have provided me with invaluable support and access to crucial datasets. Specifically, through their collaboration, I was able to gain access to two significant datasets: Automatic Identification System (AIS) data and IHS dataset. Collaborating with Astrup Fearnleys Code and leveraging their industry expertise has not only provided me with valuable datasets but also allowed me to gain insights into the complex dynamics of the maritime shipping industry.

3.1 IHS Maeiket dataset

The IHS Markit dataset provides valuable vessel specification data for big data analysis in maritime shipping. This dataset offers comprehensive information on vessel characteristics, including dimensions, tonnage, engine capacity, and ownership. By leveraging this dataset, researchers can gain insights into the diverse specifications of vessels operating in the maritime industry.

Analyzing vessel specifications from the IHS Markit dataset allows for a deeper understanding of the maritime shipping landscape. Researchers can explore correlations between vessel characteristics and various factors such as fuel efficiency, cargo capacity, or operational performance. These insights can aid in optimizing vessel selection, fleet management, and decision-making processes related to vessel operations.

By utilizing the vessel specification data from the IHS Markit dataset, this research contributes to enhancing operational efficiency and optimizing vessel-related decisions in the maritime shipping industry. This dataset equipps this research to identify trends and patterns in the maritime shipping industry related to vessel characteristics. This comprehensive dataset serves as a robust foundation for my research, enabling me to draw meaningful conclusions and make data-driven recommendations for the future of the industry.

(Write more about the dataset source and how it is collected)

3.2 The Automatic Identification System (AIS) dataset

AIS (Automatic Identification System) was developed in the 1990s to enhance navigation safety and prevent ship collisions. It allows ships equipped with AIS to communicate with each other and coastal authorities through VHF transmissions. The International Maritime Organization (IMO) mandates that all international voyage ships above 300 gross tonnage and all passenger ships must have an AIS transmitter. Governments and authorities in different nations also enforce AIS applications to improve safety and security.

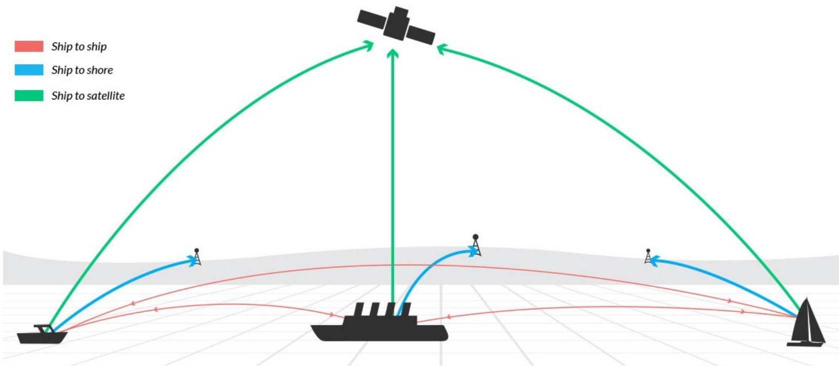

Figure 3.1: Working of AIS

There are two types of AIS transceivers: Class A and Class B. Class A transceivers broadcast more datafields and have higher reporting frequencies. The information broadcasted by a Class A transceiver can be categorized into static information, dynamic information, and voyage-related information. Dynamic information is automatically transmitted every 2-10 seconds when the ship is underway and every 3 minutes when anchored. Static and voyage-related information are broadcasted every 6 minutes regardless of navigational status. Class B transponders transmit a reduced set of data and have sparser reporting intervals compared to Class A transponders.

Data field	**Type**	**Description**
AIS identity and location	Static	Maritime Mobile Service Identity (MMSI) and the location of the system's antenna on board
Ship identity	Static	Ship name, IMO number, type, and call sign of the ship
Ship size	Static	Length and width of the ship
Ship position	Dynamic	Latitude and longitude (up to 0.0001 min accuracy)
Speed	Dynamic	Ranging from 0 knot to 102 knots (0.1 knot resolution)
Rate of turn	Dynamic	Right or left (ranging from 0 to 720° per minute)
Timestamp	Dynamic	Timestamp of the message in UTC

Table 3.1: AIS message data fields [14]

Table 3.1 shows the data fields transmitted by AIS messages. Combining AIS data with other databases can provide additional information. For example, port-to-port average speed can be calculated based on voyage distance and time stamps reported at the two ports. Cargo weight can be estimated using draught and ship sizes. Technical ship specifications, such as DWT (deadweight tonnage), capacity, design speed, and design draught, can be obtained from fleet databases using the IMO number. Port-to-port bunker consumption can be estimated based on speed, distance, and technical ship specifications like DWT and capacity [30].

(Write more about the dataset source and how it is collected)

Chapter 4

Data Cleaning and Preprocessing

Data cleaning and preprocessing play a pivotal role in big data analysis, particularly in the context of maritime shipping. As the volume and complexity of data continue to grow, ensuring data quality and reliability becomes crucial. Data cleaning involves identifying and rectifying inconsistencies, errors, and missing values, while preprocessing involves transforming raw data into a suitable format for analysis. These steps are vital for improving the accuracy and effectiveness of subsequent analysis techniques, such as predictive modeling or pattern recognition. By meticulously cleaning and preprocessing the data, this research ensures the reliability and integrity of the findings, enabling more accurate insights into vessel characteristics, operational efficiency, and decision-making processes in the maritime shipping industry.

The combination of AIS signals dataset and IHS vessel specification data enables the processing of vessel signals from the year 2022, leading to the creation of trade flows for various vessel segments. The removal of duplicate or erroneous vessel position data is of utmost importance in this process. By ensuring the accuracy and reliability of the vessel position data, the subsequent trade flow analysis can provide valuable insights into the movement of different types of vessels, their routes, and the flow of goods across various segments of the maritime shipping industry.

4.1 AIS Tradeflows

AIS Trade Flow systems utilize data transmitted by vessels worldwide to offer real-time and historical insights into maritime trade activities. The objective is to establish a system that defines trade between ports based on AIS signals. Defining port areas within AIS Trade Flow systems is a complex process that takes into account both geographical and operational considerations. Geographically, a port area is typically defined by a specific set of coordinates that delineate the physical boundaries of the port and its surrounding waters. This definition may encompass berths, anchorages, and sometimes even approach channels. Determining whether a ship is within a port area often involves comparing the ship's current AIS-reported coordinates with the defined boundaries of the port. If the ship's position falls within these boundaries, it is considered to be inside the port area. However, the determination is not solely based on geographical location. Operational factors also come into play. For instance, a ship may be within the geographical boundaries of a port, but if it is merely passing through without stopping or engaging in port activities, it may not be considered "in port" from an operational perspective.

To address these complexities, AIS Trade Flow systems frequently employ sophisticated algorithms to accurately establish port boundaries and classify vessel behavior. These algorithms take into account various factors such as the ship's speed, course, and historical behavior patterns. By integrating these diverse data points, these systems can provide a highly accurate depiction of port activities and vessel movements. Astrup Feanley Code has developed a system based on these general principles, resulting in a dataset that encompasses AIS signals as well as information on port stops, loading, and unloading.

Vessels typically travel from one port to another, and these types of journeys are called port-to-port voyages. Port-to-port voyages are categorized into two types: laden and ballast. A laden voyage refers to a maritime journey undertaken by a vessel when it is carrying cargo. This means that the ship is not empty but is loaded with goods or freight that is being transported from one location to another. On the other hand, a ballast voyage refers to a maritime journey in which a vessel travels without any cargo. In this case, the ship is not carrying freight and is being transported from one location to another.

Thus, trade flow is comprised of one or more port-to-port voyages, which can either be laden or ballast.

(Explain figure showing tradeflow)

To achive this, the AIS data is processed to identify port-to-port voyages and then stored them in a database based on segments as shown in the Figure 4.1.

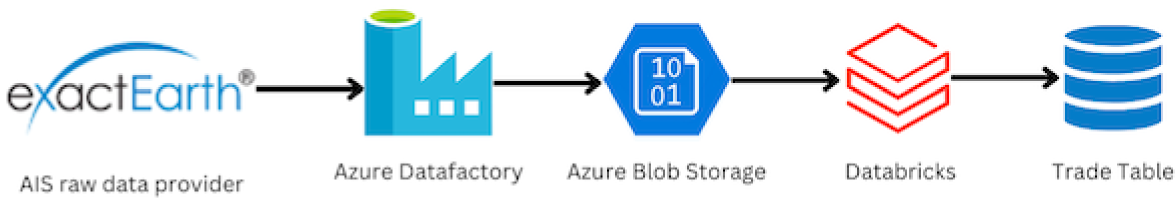

Figure 4.1: AIS Raw Data Processing

4.1.1 Processing Raw AIS data to Trade Flow

Astrup fearnley receives raw AIS data from external vendor called Exact Earth. For this there is a pipeline setup on Azure Data Factory which updates raw data every hour in CSV files.

Using databricks following informations are extracted from the raw data and processed further:

Column	Description	Data Type
IMO	International Maritime Organization 2	Number
TS_POS_UTC	Date and Time of Last Position AIS Message in UTC	YYYYMMDDHHmmSS
POSITION	WGS84 Point, Geographic Location	Geometry
LONGITUDE	WGS 84 Longitude Coordinate 2	Number, Decimal Degrees
LATITUDE	WGS 84 Latitude Coordinate	Number, Decimal Degrees
SOG	Speed over Ground	Number, Knots
HEADING	Heading	Number, Degrees
NAV_STATUS	Navigational Status	Text
DESTINATION	Port of Destination	Text
ETA	Month, Day, Hour, and Minute of Estimated Time of Arrival in UTC	MMDDHHmm
DRAUGHT	Vessel Draught	Number, Metres

Table 4.1: Important data from raw AIS data

Shipstops

To detect a ship stop, the speed is estimated by calculating the distance and time interval between the current and previous AIS signal:

$$\Delta d = \text{geodesic distance between latlong-coordinates in the current and previous signals.} \tag{4.1}$$

$$\Delta t = t_2 - t_1, \text{where } t_2 \text{ and } t_1 \text{ are the current and previous timestamps respectively.} \tag{4.2}$$

$$v = \frac{\Delta t}{\Delta d}, \text{estimated velocity in knots.} \tag{4.3}$$

If the estimated velocity is higher than 30 knots, the row is filtered out from the table because none of the ships can reach this velocity. A shipstop is defined as having a reported speed or estimated speed less than 3 knots, and the tmp_shipstopped0 table is created with these rows.

To distinguish between arrivals and departures, two shipstopped-tables are created:

Shipstopped: tmp_shipstopped0 is joined with itself on latlong1-coordinates and RowID = RowID - 1. This gives the current and previous observation of the ship inside a latlong-square with the same one-decimal coordinates. This table (b) is again left-joined with tmp_shipstopped0 and filtered on b.RowID is null. So we are left with the rows where the ship changes latlong1-coordinates in each observation.

Shipstopped2 (Departures): Same as Shipstopped but tmp_shipstopped0 is joined with itself on latlong1-coordinates and RowID = RowID + 1.

The following figure will be used from here: Ekaterina (IMO: 9196644) trade between Fujairah and Zhanjiang.

lat1	lon1	draught	speed	mdt	destination	destination_raw	eta	destinationdistance	movestatus	heading	lat	lon	deltaDistance	deltaTime	estSpeed	RowID	estSpeedBefore
25.4	56.6	11.5	4.8	2021-08-07 17:35:53	fujairah	KFK UAE	2021-08-10 14:00:00	null	Under Way Using Engine	237	25.409251666666666	56.59437666666667	2506.9738572002257	0.8611111111111111	1.6	606	3.6
25.5	57.1	11.5	3	2021-08-13 20:53:19	fujairah	KFK UAE	2021-08-07 14:00:00	null	Under Way Using Engine	146	25.53799666666667	57.109073333333335	3221.1666685038595	0.6344444444444445	2.7	566	7.8
24.9	57.6	11.5	3.8	2021-08-14 03:50:10	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	158	24.868885	57.58421333333333	4244.683602250046	1.1238888888888889	2	560	3.5
24.8	57.6	11.5	0.1	2021-08-14 04:57:36	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	158	24.833373333333334	57.59999333333333	7.8506277460473095	0.8544444444444445	0	559	2
24.9	57.6	20.5	3.1	2021-08-20 21:56:44	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	336	24.896903333333334	57.56818666666667	7710.434619688637	0.49972222222222223	8.3	438	2.8
25.9	56.9	20.5	2.2	2021-08-21 07:00:54	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	2	25.94421	56.868426666666664	3972.903545294383	1.1283333333333334	1.9	430	4.9
26	56.9	20.5	2.2	2021-08-21 08:08:36	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	2	25.980041666666665	56.87002	970.6593799863429	2.7938888888888886	0.2	429	1.9
9.9	76	20.6	5.1	2021-08-28 01:58:38	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	100	9.910853333333334	76.03030166666667	3122.6436468882644	0.9694444444444444	1.7	369	7.3
9.9	76.1	20.6	0.1	2021-08-28 02:56:48	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	297	9.908451666666666	76.05867166666667	245.36253276348776	0.9886111111111112	0.1	368	1.7
9.9	76	20.6	7.5	2021-08-28 06:44:18	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	215	9.881008333333334	76.035035	83816.3719262108	4.936111111111112	9.2	364	2.5
20.9	111.3	20.6	2.8	2021-09-11 09:58:38	zhanjiang	ZHANGIANG CHINA	2021-09-12 00:01:00	null	Under Way Using Engine	348	20.945585	111.28397666666666	1394.398420105131	0.9502777777777778	0.8	317	9.2
21	111.3	20.6	0.3	2021-09-11 10:55:39	zhanjiang	ZHANGIANG CHINA	2021-09-12 00:01:00	null	At Anchor	56	20.957088333333335	111.27851833333334	329.9013695919444	0.8505555555555555	0.2	316	0.8
21.1	110.6	20.5	10.2	2021-09-19 00:47:45	zhanjiang	ZHANJIANG, CHINA	2021-09-12 00:01:00	null	Under Way Using Engine	289	21.067565	110.57458333333334	16801.913451569675	1.0991666666666666	8.3	298	1.5
21.1	110.4	20.5	5.5	2021-09-19 01:53:42	zhanjiang	ZHANJIANG, CHINA	2021-09-12 00:01:00	null	Under Way Using Engine	315	21.120783333333332	110.42314833333333	3966.916320319417	1.05	2	297	8.3
21.2	110.4	20.5	0	2021-09-19 02:56:42	zhanjiang	ZHANJIANG, CHINA	2021-09-12 00:01:00	null	Under Way Using Engine	0	21.153373333333334	110.40728166666666	2292503.2125236765	221.8525	5.6	296	2

Figure 4.2: Shipstopped(Arrivals) for Ekaterina (IMO: 9196644)

lat1	lon1	draught	speed	mdt	destination	destination_raw	eta	destinationdistance	movestatus	heading	lat	lon	deltaDistance	deltaTime	estSpeed	RowID	estSpeedBefore
25.4	56.6	11.5	8.2	2021-08-13 17:30:35	fujairah	KFK UAE	2021-08-07 14:00:00	null	Under Way Using Engine	72	25.40253166666667	56.603745	32652.130534143795	1.9613888888888888	9	568	1.2
25.5	57.1	11.5	3	2021-08-13 21:31:23	fujairah	KFK UAE	2021-08-07 14:00:00	null	Under Way Using Engine	146	25.513931666666668	57.12706	16308.969735389586	1.4161111111111111	6.2	565	2.7
24.9	57.6	11.5	3.8	2021-08-14 03:50:10	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	158	24.868885	57.58421333333333	4244.683602250046	1.1238888888888889	2	560	3.5
24.8	57.6	20.5	3.1	2021-08-20 20:31:17	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	336	24.83553	57.5989	7473.4143525966165	1.4241666666666668	2.8	439	0.2
24.9	57.6	20.5	3.1	2021-08-20 21:56:44	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	336	24.896903333333334	57.56818666666667	7710.434619688637	0.49972222222222223	8.3	438	2.8
25.9	56.9	20.5	2.2	2021-08-21 07:00:54	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	2	25.94421	56.868426666666664	3972.903545294383	1.1283333333333334	1.9	430	4.9
26	56.9	20.5	0.1	2021-08-21 11:57:14	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	174	25.9857	56.863128333333336	61460.21065166287	10.964722222222223	3	427	0
9.9	76	20.6	5.1	2021-08-28 01:58:38	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	100	9.910853333333334	76.03030166666667	3122.6436468882644	0.9694444444444444	1.7	369	7.3
9.9	76.1	20.6	0.2	2021-08-28 05:55:26	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	316	9.90655	76.05769166666667	3762.403574221099	0.8144444444444444	2.5	365	0
9.9	76	20.6	7.5	2021-08-28 06:44:18	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	215	9.881008333333334	76.035035	83816.3719262108	4.936111111111112	9.2	364	2.5
20.9	111.3	20.6	2.8	2021-09-11 09:58:38	zhanjiang	ZHANGIANG CHINA	2021-09-12 00:01:00	null	Under Way Using Engine	348	20.945585	111.28397666666666	1394.398420105131	0.9502777777777778	0.8	317	9.2
21	111.3	20.5	0.1	2021-09-17 22:38:55	zhanjiang	ZHANJIANG, CHINA	2021-09-12 00:01:00	null	At Anchor	118	20.959655	111.27915333333333	74215.69468999672	26.147222222222222	1.5	299	0
21.1	110.6	20.5	10.2	2021-09-19 00:47:45	zhanjiang	ZHANJIANG, CHINA	2021-09-12 00:01:00	null	Under Way Using Engine	289	21.067565	110.57458333333334	16801.913451569675	1.0991666666666666	8.3	298	1.5
21.1	110.4	20.5	5.5	2021-09-19 01:53:42	zhanjiang	ZHANJIANG, CHINA	2021-09-12 00:01:00	null	Under Way Using Engine	315	21.120783333333332	110.42314833333333	3966.916320319417	1.05	2	297	8.3
21.2	110.4	20.5	0	2021-09-19 02:56:42	zhanjiang	ZHANJIANG, CHINA	2021-09-12 00:01:00	null	Under Way Using Engine	0	21.153373333333334	110.40728166666666	2292503.2125236765	221.8525	5.6	296	2

Figure 4.3: Shipstopped2(Departures) Ship Stops for Ekaterina (IMO: 9196644)

Ship Callport

To determine whether the arrival and/or departure observations are part of a port stop, we follow these steps:

1. Left Join with LatLongPort3 Table: The **LatLongPort3** table is left-joined with the union of the ShipStopped tables. This results in a new table containing all the arrival and departure measurements along with their corresponding destination names based on the latlong-position.

2. Obtain ShipCallPort Table: The **ShipCallPort** table is obtained by lagging the MDT (Mean Draft) and draught columns. This process allows us to retrieve the timestamps and draught values before and after the port stop.

3. Filter Stops Less Than 3 Hours: Finally, we filter out the stops that have a duration of less than 3 hours, as they are likely not significant port stops.

Overall, this approach helps us identify and analyze port stops in the data, enabling further insights into ship movements and activities in specific ports.

lat1	lon1	arriveDraught	departDraught	arrive	depart	arriveOuter	departOuter	destination	destination_raw	TrueDestination	Eta	DestinationDistance	MoveStatus	Heading	country
25.4	56.6	11.5	11.5	2021-08-07 17:35:53	2021-08-13 17:30:35	2021-08-01 07:50:07	2021-08-13 20:53:19	fujairah	KFK UAE	fujairah	2021-08-07 14:00:00	null	Under Way Using Engine	72	AE
25.5	57.1	11.5	11.5	2021-08-13 20:53:19	2021-08-13 21:31:23	2021-08-13 17:30:35	2021-08-14 03:50:10	fujairah	KFK UAE	null	2021-08-07 14:00:00	null	Under Way Using Engine	146	null
24.8	57.6	11.5	20.5	2021-08-14 04:57:36	2021-08-20 20:31:17	2021-08-14 03:50:10	2021-08-20 21:56:44	sohar	SOHAR	null	2021-08-10 23:00:00	null	Under Way Using Engine	336	null
24.9	57.6	20.5	20.5	2021-08-20 21:56:44	2021-08-20 21:56:44	2021-08-20 21:56:44	2021-08-21 07:00:54	sohar	SOHAR	null	2021-08-10 23:00:00	null	Under Way Using Engine	336	null
25.9	56.9	20.5	20.5	2021-08-21 07:00:54	2021-08-21 07:00:54	2021-08-20 21:56:44	2021-08-21 07:00:54	sohar	SOHAR	null	2021-08-10 23:00:00	null	Under Way Using Engine	2	null
26	56.9	20.5	20.5	2021-08-21 08:08:36	2021-08-21 11:57:14	2021-08-21 07:00:54	2021-08-28 01:58:38	sohar	SOHAR	null	2021-08-10 23:00:00	null	Under Way Using Engine	174	null
9.9	76.1	20.6	20.6	2021-08-28 02:56:48	2021-08-28 05:55:26	2021-08-28 01:58:38	2021-08-28 06:44:18	singapore	SINGAPORE	kochi	2021-09-05 14:00:00	null	Under Way Using Engine	316	IN
9.9	76	20.6	20.6	2021-08-28 06:44:18	2021-08-28 06:44:18	2021-08-28 06:44:18	2021-09-11 09:58:38	singapore	SINGAPORE	kochi	2021-09-05 14:00:00	null	Under Way Using Engine	215	IN
21	111.3	20.6	20.5	2021-09-11 10:55:39	2021-09-17 22:38:55	2021-09-11 09:58:38	2021-09-19 00:47:45	zhanjiang	ZHANJIANG, CHINA	null	2021-09-12 00:01:00	null	At Anchor	118	null

Figure 4.4: ship callport

Ship Travel

The **Shiptravel** table is created from the **Shiptravel_** table left-joined with the **ShipCallPort** table on arrival timestamp per IMO-number. This process provides an overview of the total travel route and the ports where the ship called. The destination name, country, and arrival timestamp are concatenated into an array column called **TrueDestinationArray**.

Subsequently, the **ShipTravelTrue** table is created by selecting the last value of **TrueDestinationArray** per IMO-number. This table offers a concise representation of the ship's true travel destinations.

From the **ShipTravelTrue** table, the **export-table** is derived by selecting the last portstop and concatenating the destination name and MDT (Mean Draft) to create the **leg-column**. This results in a table that presents the final destination and the MDT value for the ship's last port stop, providing valuable insights into the ship's final leg of the journey.

lat1	lon1	Draught	speed	MDT	destination	destination_raw	Eta	destinationdistance	MoveStatus	Heading	Lat	Lon	TrueDestinationArray	deltadistance
10.5	75	20.5	9.5	2021-08-27T17:45:48.000+0000	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	108	10.5239	74.97308	kochi□IN□2021-08-28 02:56:48	33145.52462488484
10.4	75.2	20.5	9.5	2021-08-27T19:40:02.000+0000	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	113	10.367151666666667	75.231095	kochi□IN□2021-08-28 02:56:48	68480.45749279462
10.1	75.8	20.5	6.8	2021-08-27T23:33:36.000+0000	sohar	SOHAR	2021-08-10 23:00:00	null	Under Way Using Engine	112	10.067461666666667	75.77800833333333	kochi□IN□2021-08-28 02:56:48	32638.1874332692
9.9	76	20.6	5.1	2021-08-28T01:58:38.000+0000	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	100	9.910853333333334	76.03030166666667	kochi□IN□2021-08-28 02:56:48	3122.6436468882644
9.9	76.1	20.6	0.1	2021-08-28T02:56:48.000+0000	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	297	9.908451666666666	76.05867166666667	kochi□IN□2021-08-28 02:56:48	245.36253276348776
9.9	76.1	20.6	0.2	2021-08-28T03:56:07.000+0000	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	314	9.90654	76.05753666666666	kochi□IN□2021-08-28 06:44:18	25.453247372716717
9.9	76.1	20.6	0.2	2021-08-28T04:07:58.000+0000	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	316	9.906453333333333	76.05775166666666	kochi□IN□2021-08-28 06:44:18	12.554634961819298
9.9	76.1	20.6	0.2	2021-08-28T05:55:26.000+0000	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	316	9.90655	76.05769166666667	kochi□IN□2021-08-28 06:44:18	3762.403574221099
9.9	76	20.6	7.5	2021-08-28T06:44:18.000+0000	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	215	9.881008333333334	76.035035	kochi□IN□2021-08-28 06:44:18	83816.3719262108
9.2	75.6	20.6	10.1	2021-08-28T11:40:28.000+0000	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	145	9.24563	75.61895666666666	fujairah□AE□2021-10-11 05:54:04	38302.992377495015
8.9	75.8	20.6	9.8	2021-08-28T13:45:13.000+0000	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	145	8.947631666666666	75.796475	fujairah□AE□2021-10-11 05:54:04	38224.07301469262
8.7	76	20.6	10	2021-08-28T15:51:14.000+0000	singapore	SINGAPORE	2021-09-05 14:00:00	null	Under Way Using Engine	136	8.651696666666666	75.97593	fujairah□AE□2021-10-11 05:54:04	32705.198151543307

Figure 4.5: shiptraveltrue Table

From Figure 4.5, we can observe that the destination-column is not necessarily the same as the TrueDestinationArray.

Trade Table

Tradetable13Bx is derived from ShipTravel through several steps of data processing and calculations. Initially, **Tradetabletmp** is obtained by estimating the speed and deltaDistance between observations in **ShipTravelTrue**, filtering out rows with estimated speed above 30 knots, and extracting TrueDestinationArray components. **Tradetable0** is defined at the arrival timestamp and unioned with **tmp_etatradetableadditions** to create Tradetable. The **tmp_etatradetableadditions** table is created by examining ETA and nextETA values in **Tradetabletmp** and setting flag values accordingly.

Next, **Tradetable_C** is derived by adjusting the distanceTravelledNm-column using row numbers for each IMO and TrueDestination to identify trips to specific destinations. **Tradetable_collapsed** is then created by left joining the **ShipCallPort** table with **Tradetable_C** and adjusting columns based on individual trip logic.

Further processing is done in **Tmp_tradetable0**, including lagging columns and left joining with **ShiftDraughtChange** table. **Tmp_tradetable1** adds additional rows to **tmp_details** table. Finally, **Tradetable13x** is derived by modifying columns based on certain thresholds and rules, and **Tradetable13Bx** is obtained by aggregating information from **Tradetable13x** and calculating sumLegDistance and HoursTravelled_.

In summary, **Tradetable13Bx** is obtained through a series of data manipulations and calculations starting from ShipTravel.

loadPort	loadRegion	loadCountry	loadHoursMin	loadHoursMax	unloadRegion	unloadCountry	unloadPort	nextUnloadRegion	nextUnloadCountry	nextUnloadPort	maxDraughtChangeS	portStops	EstimatedDischargePct	RouteDistanceAB	c	sumDistance
BASRAH	Middle East Gulf	IRQ	100	165	Asia	CHN	ZHANJIANG	Asia	CHN	ZHANJIANG	9.5	7	-95.8	1475	124	11638.5
FUJAIRAH	Middle East Gulf	ARE	143	301	Asia	CHN	ZHANJIANG	Asia	CHN	ZHANJIANG	9.6	5	0	1467	31	1467
FUJAIRAH	Middle East Gulf	ARE	143	301	Asia	CHN	ZHANJIANG	Asia	CHN	ZHANJIANG	9.6	5	1	1968	9	3435
FUJAIRAH	Middle East Gulf	ARE	143	301	Asia	CHN	ZHANJIANG	null	null	null	9.6	5	99	1329	33	4764

Figure 4.6: Tradetable13Bx

Chapter 5

Tools and Data Analysis Techniques

In this section, we will use tradeflow data for the year 2022, along with specifications from the IHS dataset, to perform data analysis.

5.1 Tools

To perform data analysis, we will utilize the following tools:

- Python
- Databricks
- Apache Spark
- MSSQL
- Pandas
- Astrup Fearnleys Code Emission Calculator API

5.1.1 Databricks

Databricks is a leading platform for big data analysis, specifically designed to handle large-scale datasets efficiently and effectively. Leveraging Apache Spark as its core engine, Databricks provides a distributed computing framework that enables processing massive volumes of data in parallel across multiple nodes. This distributed architecture allows for significant performance improvements, reducing the time it takes to process and analyze big data compared to traditional approaches. With its ability to handle both batch and real-time data processing, Databricks empowers data engineers and analysts to extract valuable insights from vast datasets, enabling them to make data-driven decisions at scale.

One of the key advantages of using Databricks for big data analysis is its ease of use and collaborative features. The platform offers interactive notebooks (Figure 5.1) that allow data professionals to write, execute, and share code seamlessly. This enables collaborative data exploration and simplifies the iterative process of data analysis and model development. Moreover, Databricks provides a rich set of built-in libraries and integrations with popular big data tools and machine learning frameworks, streamlining the development and deployment of complex data pipelines and advanced analytical models. By abstracting the complexities of distributed data processing, Databricks empowers data teams to focus on the analysis and interpretation of results, accelerating the time-to-insight for big data projects and ultimately driving business growth and innovation. One of the key benefits of using Databricks for big data analysis is its ability to handle large volumes of data with ease. Whether you're working with terabytes, petabytes, or even exabytes of data, Databricks can scale to meet your needs without sacrificing performance or reliability. This makes it an ideal solution for organizations looking to perform complex big data analysis tasks, such as machine learning, data warehousing, and real-time streaming analytics [5].

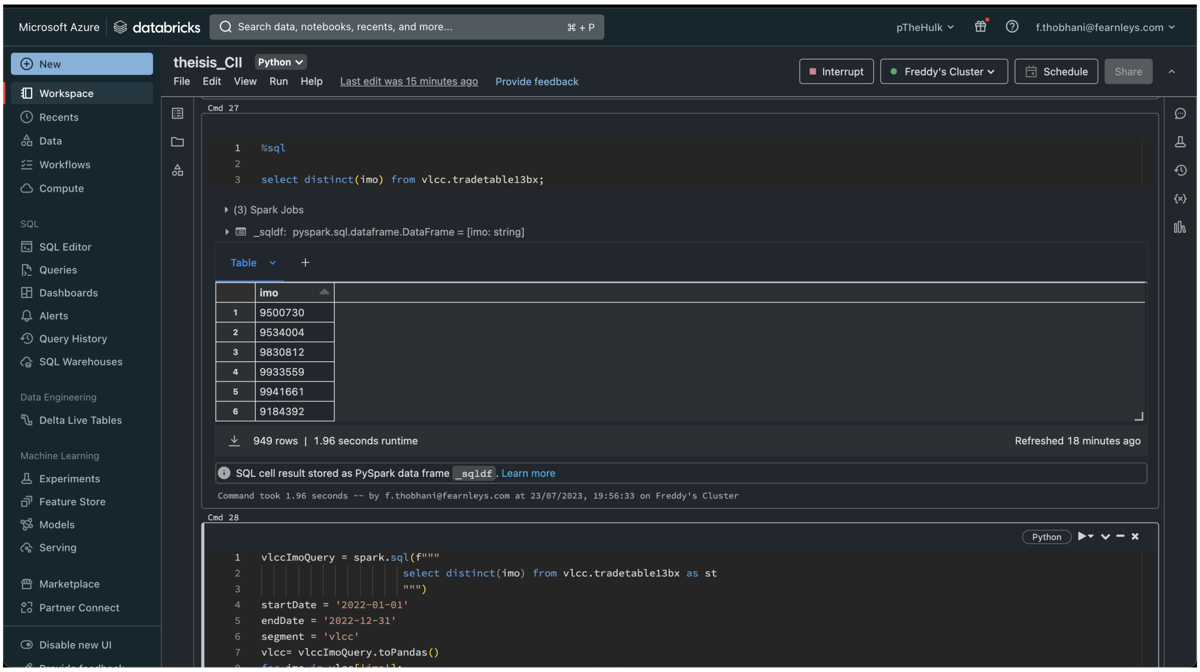

Figure 5.1: Databricks Notebook

Overall, Databricks is a highly capable platform for big data analysis that is well-suited to a wide range of use cases, from simple data processing tasks to complex machine learning and statistical analysis.

The entire analysis will be conducted on Databricks with the following configuration:

- Databricks Runtime Version: 11.3 LTS (includes Apache Spark 3.3.0, Scala 2.12)
- 8GB Memory, 4 Cores with 1 driver and 1 worker node.
- Python 3
- Elastic Disk

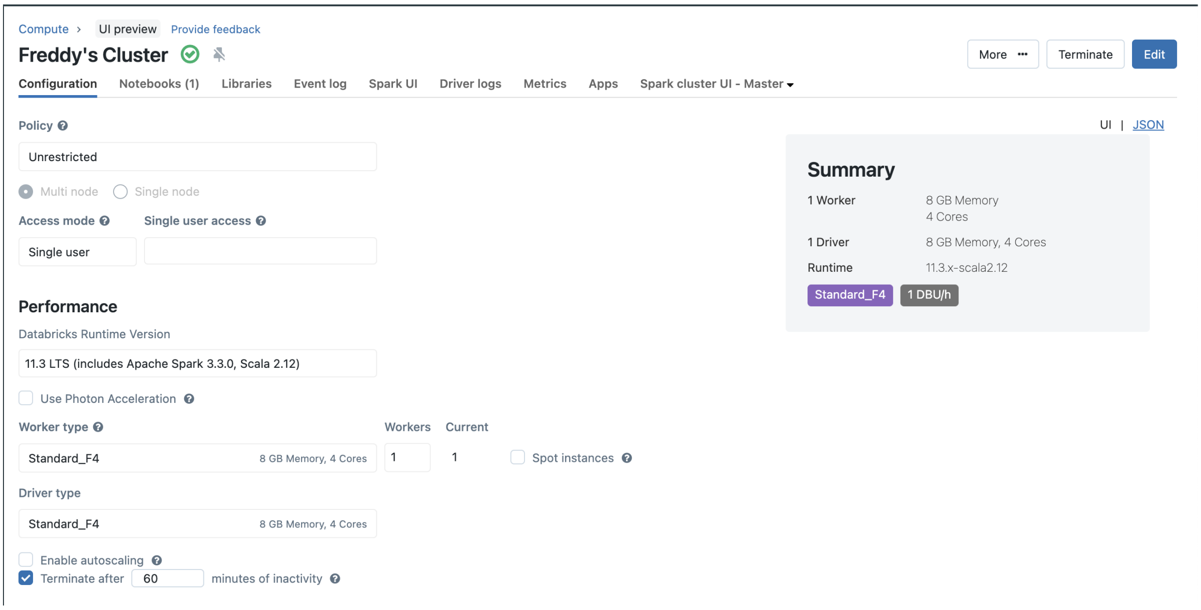

Figure 5.2: Databricks Configuration

5.1.2 Apache Spark

Apache Spark is an open-source distributed computing framework designed for processing and analyzing large-scale datasets in a highly efficient and parallel manner. It provides a unified platform that supports various data processing tasks, including batch processing, real-time streaming, machine learning, and graph processing. Spark's core abstraction is a resilient distributed dataset (RDD), which allows data to be distributed across multiple nodes in a cluster, enabling data processing operations to be executed in parallel [22].

Databricks leverages Apache Spark as its underlying engine to offer a powerful and scalable data analytics platform. By integrating Spark into its infrastructure, Databricks provides users with a seamless and interactive environment for collaborative data engineering and data science tasks. Databricks' interactive notebooks allow data professionals to write and execute Spark-based code, making it easier to perform data manipulations, transformations, and analysis in real-time. The platform also offers support for various programming languages such as Python, Scala, R, and SQL, providing users with flexibility and familiarity in their preferred language.

Furthermore, Databricks enhances Apache Spark by providing additional features and optimizations to improve performance and ease of use. The platform offers auto-scaling capabilities, enabling resources to be automatically allocated and released based on the workload demand, ensuring optimal performance and cost-efficiency. Databricks also provides built-in libraries and tools that simplify complex tasks such as machine learning and data visualization, allowing data scientists to focus on building and deploying advanced analytical models with ease.

In summary, Apache Spark forms the backbone of Databricks, enabling the platform to handle large-scale datasets efficiently and deliver a collaborative and user-friendly environment for data analysis, making it a popular choice for organizations seeking to harness the potential of big data analytics.

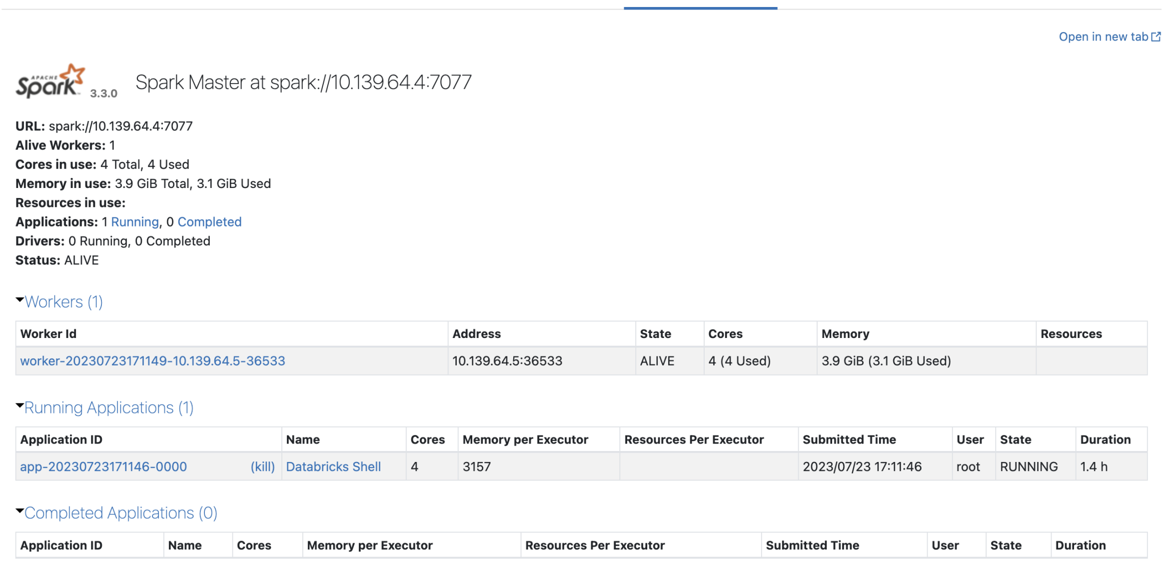

Figure 5.3: Apache Spark Configuration

5.1.3 Astrup Fearnleys Code Emission Calculator API

Astrup Fearnleys Code Emission Calculator API is a service that provides a simple interface for calculating the emissions of a given vessel. The API is built on top of the Astrup Fearnleys Code Emission Calculator, It is internal tool developed to provide a simple interface for calculating the emissions of a given vessel for researchers.

It requireds the following parameters to be passed as a input:

Parameter	Type	Description
imo	integer	IMO of a ship that will require emission calculations.
avg_speed_kn	number	The vessel's intended average sailing speed in knots on the route. Optional if duration_h is given.
distance_nm	number	Expected length of the voyage in Nautical Miles.
cargo_unit	string	Cargo unit used to submit cargo_amt. Currently, the only possible value is tons. Used to calculate EEOI and transport_work.
cargo_amt	number	Amount of cargo the ship is carrying (in the given cargo_unit). Used to calculate EEOI and transport_work.
me_fuel_type	string	Fuel type to be considered for the main engine consumption for the emissions estimate. Overwriting the model assumption.
ae_fuel_type	string	Fuel type to be considered for the auxiliary engine consumption for the emissions estimate. Overwriting the model assumption.
boiler_fuel_type	string	Fuel type to be considered for the boiler consumption for the emissions estimate. Overwriting the model assumption.
duration_h	number	Expected duration of the voyage in hours. Optional if avg_speed_kn is given.
load_cond	string	Loading condition of the vessel to be considered in the power estimation.
me_co2_factor	number	CO2 factor for the main engine fuel. Main engine fuel type must be specified.
ae_co2_factor	number	CO2 factor for the auxiliary engine fuel. Auxiliary engine fuel type must be specified.
boiler_co2_factor	number	CO2 factor for the boiler engine fuel. Boiler engine fuel type must be specified.

Table 5.1: Request Parameters for Emission Calculations

In return, it provides the JSON output in following format:


```
{
  "imo": 1234567,
  "avg_speed_kn": 10.045,
  "load_cond": "ballast",
  "cargo_unit": "tons",
  "cargo_amt": 544385.543,
  "me_fuel_type": "MGO",
  "ae_fuel_type": "MGO",
  "boiler_fuel_type": "MGO",
  "duration_h": 354.5,
  "me_fuel_cons_metric_metric_tons": 0.1,
```

```
    "ae_fuel_cons_metric_metric_tons": 0.1,
    "boiler_fuel_cons_metric_metric_tons": 0.1,
    "total_co2_emission_metric_tons": 0.1,
    "total_so2_emission_metric_tons": 0.1,
    "total_particulate_matter_emission_metric_tons": 0.1,
    "total_nox_emission_metric_tons": 0.1,
    "total_nmvoc_emission_metric_tons": 0.1,
    "total_ch4_emission_metric_tons": 0.1,
    "total_n2o_emission_metric_tons": 0.1,
    "total_co_emission_metric_tons": 0.1,
    "total_black_carbon_emission_metric_tons": 0.1,
    "total_organic_carbon_emission_metric_tons": 0.1,
    "transport_work": 0.1,
    "eeoi": 0.1,
    "me_co2_factor": 2.1,
    "ParamValidation": {
      "status": "OK"
    },
    "LimitMessage": {
      "message_description": "You are within your call limit"
    }
  }
```


Listing 5.1: JSON output format from emission api

5.2 Methodlogy

In this section, using tradeflow data for the year 2022, along with specifications from the IHS dataset, we will perform data analysis.

Tradetable data is available in the `tradetable13bx` dataset. For each segments there is `tradetable13bx` and thereforre we will have to perform the analysis for each segment separately.

The steps involved in the analysis are as follows:

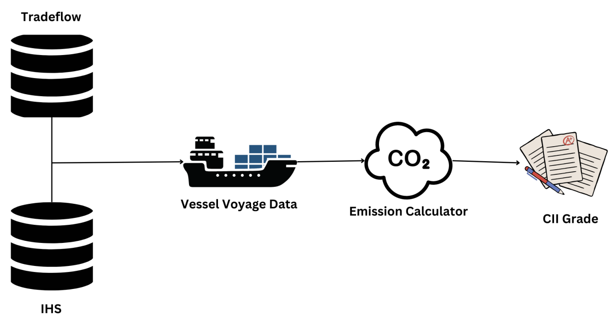

Figure 5.4: Methodlogy Overview

1. Getting distinct IMO numbers from tradetable13bx

Distinct IMO numbers can be obtained from the `tradetable13bx` dataset by querying the `imo` column. This will give us a list of all ships that are present in the dataset.

```
vlccImoQuery = spark.sql(f"""
    SELECT DISTINCT(imo) FROM vlcc.tradetable13bx
""")
```

Listing 5.2: SQL Query to get distinct IMO

2. Getting necessary information for each IMO number from IHS and tradetable13bx

For each IMO number obtained in the previous step, query the IHS dataset or join it with `tradetable13bx` to gather relevant information about each ship, such as its specifications, fuel type, engine types, etc.

```
        shipTrade = spark.sql(f"""
            SELECT
                st.Imo,
                st.AverageSpeedAB,
                st.LoadAtB,
                st.UnloadAtB,
                st.ShipName,
                st.FromPortA,
                st.ToPortB,
                st.DepartA,
                st.ArriveB,
                st.HoursAtAMax,
                st.HoursAtBMax,
                st.ArriveDraughtB,
                st.DraughtChangeB,
                st.DistanceAB,
                st.HoursAB,
                st.EstimatedDischargePct,
                ihs.deadweight,
```

```
                ihs.draught as maxDraught,
                sp.ballast_threshold_pct as ballastThreshold,
                ABS(round((st.ArriveDraughtB * 100) /ihs.draught)) as
                  ↪ draughtPercentage,
                ABS(round(ihs.deadweight*st.EstimateddischargePct
                  ↪ /100,0)) as cargoEst,
                ABS(round((ihs.deadweight*st.ArriveDraughtB)/ihs.
                  ↪ draught)) as cargoEst2,
                  CASE
                    WHEN
                      st.ArriveDraughtB >= (ihs.draught * sp.
                        ↪ ballast_threshold_pct / 100.0)
                    THEN 'laden'
                    ELSE 'ballast'
                  END AS loadCond
              FROM vlcc.tradetable13bx as st
              JOIN ihs_ship_data_ ihs on ihs.lrimoshipno = st.Imo
              JOIN (
                  select * from parameters.segment_parameters where
                    ↪ segment = '{segment}'
                ) sp on true
              WHERE st.Imo = {imo} AND st.DepartA between '{startDate}'
                ↪ AND '{endDate}' AND st.ArriveDraughtB IS NOT null
              ORDER BY st.DepartA

              """)
```

Listing 5.3: SQL Query to get relevent data from `ihs` and `tradetable13bx`

Here, we have estimated cargo based on draught noted on ports and maximum draught from IHS dataset. Also to deptermine whether the tradeflow is laden or ballast, we have used ballast threshold percentage from segment parameters.

Threshold for each sections are as showin in figure 5.5:

`shipTrade` obtained is converted into pandas dataframe and then processed further.

```
shipTradePandas = shipTrade.toPandas()
```

Listing 5.4: shipTrade pandas dataframe

Further trade is divided into dataframes for laden and ballast tradeflows.

```
ladenTrade = shipTradePandas[shipTradePandas["loadCond"] == 'laden']
ballastTrade = shipTradePandas[shipTradePandas["loadCond"] == '
    ↪ ballast']
```

Listing 5.5: laden and ballast tradeflows

segment	ballast_threshold_pct
panamax_bulk	70
suezmax	70
supramax	70
vlcc	70
vlgc	85
vloc	70
aframax	70
capesize	70
lgc	85
lng	85
mgc	85
panamax	70

Figure 5.5: Ballast Threshold Percentage

Then we extract information like distance, time and average speed for both ballast and laden tradeflows.

```
    totalLadenDistance = ladenTrade["DistanceAB"].sum()
    totalBallastDistance = ballastTrade["DistanceAB"].sum()
    totalDistance = shipTradePandas["DistanceAB"].sum()

    totalLadenHours = ladenTrade["HoursAB"].sum()
    totalBallastHours = ballastTrade["HoursAB"].sum()
    totalHours = shipTradePandas["HoursAB"].sum()

    ladenAvgSpeed = ladenTrade["AverageSpeedAB"].mean()
    ladenAvgSpeed2 = totalLadenDistance / totalLadenHours

    ballastAvgSpeed = ballastTrade["AverageSpeedAB"].mean()
    ballastAvgSpeed2 = totalBallastDistance / totalBallastHours

    cargo2 = ladenTrade["cargoEst2"].sum()
```

Listing 5.6: Determing distance, time and average speed

3. Cleaning data

Data cleaning involves handling missing values, removing duplicates, standardizing units, and correcting any inconsistencies in the dataset.

When determining average speed, speed determined from AIS data might be missing for some tradeflows. Also sometimes, it was noticed that averge speed was very high for some tradeflows. In such cases, we can use the distance and time values to calculate average speed.

```
if (ladenAvgSpeed < 0 or ladenAvgSpeed > ladenAvgSpeed2):
  ladenAvgSpeed = ladenAvgSpeed2
if (ladenAvgSpeed > 15):
  ladenAvgSpeed = 15
```

Listing 5.7: Handling missing speed

Sometimes for certain trades average distance was missing, this trades were excluded.

4. Calculating emissions for laden and ballast tradeflows.

Using the information extracted in the previous steps, emissions is calculated using AFC emission calculator API.

For this `getEmission` function is used which takes IMO number, cargo amount, distance, load condition and estimated speed as input and returns emissions. As showin in Listing 5.8.

```
def getEmission(imo, cargo_amt, distance_nm, load_cond, est_speed):
    token = get_token()

    data_string = f"imo={imo}&cargo_amt={cargo_amt}&distance_nm={distance_nm}&duration_h={duration_h}&load_cond={load_cond}"

    url = url_start + data_string

    payload={}
    headers = {
    'Ocp-Apim-Subscription-Key': 'xxxxxxxxxxx',
    'Authorization': 'Bearer ' + token
    }

    response = requests.request("GET", url, headers=headers, data=payload)
    time.sleep(0.04)
    try:
        response_dict = json.loads(response.text)
        return response_dict
    except BaseException as e:
```

```
            print("error", e)
            print("error in url", url)
            return None
```

Listing 5.8: Handling missing speed

5. Calculating CII, CII required, and CII reference.

Carbon Intensity Indicator (CII) is calculated as the ratio of emissions (CO_2) to transport work (distance traveled $\times$ cargo amount). CII required is based on regulatory standards, while CII reference can be derived from historical data or benchmarks. Calculate these indicators for each tradeflow using the emission data and transport work values.

CII reference is calculated using the following formula:

$$CII_{\text{ref}} = a \cdot \text{Capacity}^{-C} \tag{5.1}$$

where a and C are constants according to Figure 5.6, and Capacity is the deadweight of the ship.

Ship type		capacity	a	c
Bulk Carrier	DWT ≥ 279,000	279,000	4745	0.622
	DWT < 279,000	DWT	4745	0.622
Gas Carrier	DWT ≥ 65,000	DWT	14405E+7	2.071
	DWT < 65,000	DWT	8104	0.639
Tanker		DWT	5247	0.610
Container ship		DWT	1984	0.489
General cargo ship	DWT ≥ 20,000	DWT	31948	0.792
	DWT < 20,000	DWT	588	0.3885
Refrigerated cargo carrier		DWT	4600	0.557
Combination carrier		DWT	5119	0.622
LNG Carrier	DWT ≥ 100,000	DWT	9.827	0
	100,000 > DWT ≥ 65,000	DWT	14479E+10	2.673
	DWT < 65,000	65,000	14479E+10	2.673
Ro-ro cargo ship (VC)	GT ≥ 57,700	57,700	3627	0.590
	57,700 > GT ≥ 30,000	GT	3627	0.590
	GT < 30,000	GT	330	0.329
Ro-ro cargo ship		GT	1967	0.485
Ro-ro passenger ship	Ro-ro passenger ship	GT	2023	0.460
	High-speed craft	GT	4196	0.460
Cruise passenger ship		GT	930	0.383

Figure 5.6: a and C constants

CII required can be calculated using the following formula:

$$\text{Requited CII} = \left(\frac{100 - Z}{100}\right) \times \text{CIIRef} \tag{5.2}$$

In Equation 5.2, the reduction factor Z represents the initial value of 5% in the year 2023, with an annual increment of 2%. Additionally, for the years 2027 to 2030, the Z factors are subject to enhancement and refinement, guided by the evaluation of the short-term measure's effectiveness.

Year	Reduction Factor (Z)
2023	5%
2024	7%
2025	9%
2026	11%
2027	**
2028	**
2029	**
2030	**

Table 5.2: Reduction Factors Over the Years

CII reference is calculated using the function shown in Listing 5.9.

```
def calculateCIIRef(deadweight, shipType):
    a = 0
    c = 0
    capacity = deadweight
    if (shipType == "Bulk␣Carriers"):
        a = 4745
        c = 0.622
        if (deadweight >= 279000):
        capacity = 279000
    elif (shipType == "Tankers"):
        a = 5247
        c = 0.61
    elif (shipType == "LNG␣carriers"):
        if capacity > 100000:
        a = 9.827
        c = 0
        elif (capacity >= 65000 and capacity < 100000):
        a = 14479E+10
        c = 2.673
        else:
        capacity = 65000
        a = 14479E+10
        c = 2.673
    elif (shipType == "Container␣ship"):
        a = 1984
        c = 0.489
    return a * (capacity**(-c))
```

Listing 5.9: Function to calculate CII reference

6. Calculating CII grade.

Based on the calculated CII and CII required, determine the CII grade for each tradeflow, indicating its compliance with emissions regulations. This can be done using predefined thresholds or standards to categorize tradeflows into different CII grades.

Taking into account that vessels will experience similar voyages in 2023 as they did in 2022, the CII grade for the end of 2023 can be estimated with a reduction factor Z set at 5.

The ”*dd*” vector is established by calculating the ratio between Attained CII and Required CII. This ratio is subsequently compared against the thresholds for CII grades to ascertain the appropriate grade. The larger the deviation of the ratio, the lower the grade assigned, with A representing the best and E signifying the poorest performance.

Ship type		d1	d2	d3	d4
Bulk Carrier		0.86	0.94	1.06	1.18
Gas Carrier	>=65,000DWT	0.81	0.91	1.12	1.44
	<65,000DWT	0.85	0.95	1.06	1.25
Tanker		0.82	0.93	1.08	1.28
Container ship		0.83	0.94	1.07	1.19
General cargo ship		0.83	0.94	1.06	1.19
Refrigerated cargo carrier		0.78	0.91	1.07	1.20
Combination carrier		0.87	0.96	1.06	1.14
LNG Carrier	>= 100,000DWT	0.89	0.98	1.06	1.13
	<100000DWT	0.78	0.92	1.10	1.37
Ro-ro cargo ship (VC)		0.86	0.94	1.06	1.16
Ro-ro cargo ship		0.76	0.89	1.08	1.27
Ro-ro passenger ship		0.76	0.92	1.14	1.30
Cruise passenger ship		0.87	0.95	1.06	1.16

Figure 5.7: *dd* vectors for determining the rating boundaries of ship types

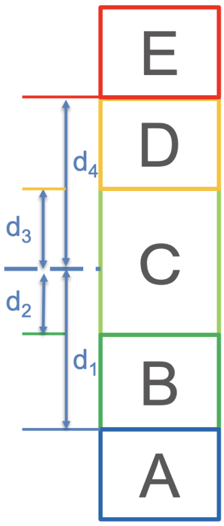

Figure 5.8: CII grading based on *dd* vector

Based on figure 5.7, the CII grade can be determined as represented in figure 5.8.

For Bulk Carriers CII grade is determined as shown in Listing 5.10.

```
def getGrade(cii, ciiRequired):
    ciiRatio = cii / ciiRequired
    if (ciiRatio < 0.82):
        return "A"
    elif (ciiRatio < 0.93):
        return "B"
    elif (ciiRatio < 1.08):
        return "C"
    elif (ciiRatio < 1.28):
        return "D"
    else:
        return "E"
```

Listing 5.10: Function to determine CII grading for Bulk Carriers

Following above 6 steps, we can calculate emissions, CII, CII required, CII reference and CII grade for each vessel. Results can be stored in a SQL table for further analysis. For each segment, we can perform above steps as showing in Listing 5.11

```
capsizeImoQeury = spark.sql(f"""
                 select distinct(imo) from capesize.tradetable13bx as st
                 """)
startDate = '2022-01-01'
endDate = '2022-12-31'
segment = 'capesize'
capeSize = capsizeImoQeury.toPandas()
for imo in capeSize['imo']:
imoCIIData = getCIIGradeByImo(imo, startDate, endDate, segment, 5)
df = pd.DataFrame([imoCIIData])
spark.createDataFrame(df).write.mode("append").saveAsTable("emissions.capesize_cii_2022_v3")
time.sleep(5)
```

Listing 5.11: Analysis for capsize segment

In listing 5.11, we are getting distinct IMO numbers for capsize segment. Then for each IMO number, we are calling `getCIIGradeByImo` function to get emissions, CII, CII required, CII reference and CII grade. Using $spark.createDataFrame(df).write()$ we are storing results in SQL table.

Based on the data obtained so far, we will analysis the results in the next chapter.

Bibliography

[1] Nicoleta Acomi and Ovidiu Cristian Acomi. "Improving the voyage energy efficiency by using EEOI". In: *Procedia-Social and Behavioral Sciences* 138 (2014), pp. 531–536.

[2] Lai Fatt Chuah et al. "Implementation of the energy efficiency existing ship index and carbon intensity indicator on domestic ship for marine environmental protection". In: *Environmental research* 222 (2023), p. 115348.

[3] Lai Fatt Chuah et al. "Implementation of the energy efficiency existing ship index and carbon intensity indicator on domestic ship for marine environmental protection". In: *Environmental Research* 222 (2023), p. 115348. ISSN: 0013-9351. DOI: https://doi.org/10.1016/j.envres.2023.115348. URL: https://www.sciencedirect.com/science/article/pii/S0013935123001408.

[4] Ernest Czermański et al. "Implementation of the Energy Efficiency Existing Ship Index: An important but costly step towards ocean protection". In: *Marine Policy* 145 (2022), p. 105259.

[5] *Data Lakehouse Architecture and AI Company - Databricks*. URL: https://www.databricks.com/.

[6] Andrzej S. Grzelakowski, Jerzy Herdzik, and Sławomir Skiba. "Maritime Shipping Decarbonization: Roadmap to Meet Zero-Emission Target in Shipping as a Link in the Global Supply Chains". In: *Energies* 15.17 (2022). ISSN: 1996-1073. DOI: 10.3390/en15176150. URL: https://www.mdpi.com/1996-1073/15/17/6150.

[7] Mohamad Issa, Adrian Ilinca, and Fahed Martini. "Ship Energy Efficiency and Maritime Sector Initiatives to Reduce Carbon Emissions". In: *Energies* 15.21 (2022). ISSN: 1996-1073. DOI: 10.3390/en15217910. URL: https://www.mdpi.com/1996-1073/15/21/7910.

[8] Veronica Jaramillo Jimenez, Hyungju Kim, and Ziaul Haque Munim. "A review of ship energy efficiency research and directions towards emission reduction in the maritime industry". In: *Journal of Cleaner Production* 366 (2022), p. 132888. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2022.132888. URL: https://www.sciencedirect.com/science/article/pii/S0959652622024817.

[9] Sheng-Long Kao, Wu-Hsun Chung, and Chao-Wei Chen. "AIS-Based Scenario Simulation for the Control and Improvement of Ship Emissions in Ports". In: *Journal of Marine Science and Engineering* 10.2 (2022). ISSN: 2077-1312. DOI: 10.3390/jmse10020129. URL: https://www.mdpi.com/2077-1312/10/2/129.

[10] Anthony King. *Emissions-free sailing is full steam ahead for ocean-going shipping*. https://ec.europa.eu/research-and-innovation/en/horizon-magazine/emissions-free-sailing-full-steam-ahead-ocean-going-shipping. Sept. 2022.

[11] Tatiana Lukoianove and Victoria Rubin. "Veracity Roadmap: Is Big Data Objective, Truthful and Credible?" In: *ASIST 2013: 76th Annual Meeting of the Association for Information Science and Technology: Beyond the Cloud: Rethinking Information Boundaries* 24 (Nov. 2013). DOI: 10.7152/acro.v24i1.14671.

[12] Ziaul Haque Munim et al. "Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions". In: *Maritime Policy & Management* 47.5 (2020), pp. 577–597. DOI: 10.1080/03088839.2020.1788731. eprint: https://doi.org/10.1080/03088839.2020.1788731. URL: https://doi.org/10.1080/03088839.2020.1788731.

[13] Harry E Pence. "What is big data and why is it important?" In: *Journal of Educational Technology Systems* 43.2 (2014), pp. 159–171.

[14] Heather M Perez et al. "Automatic identification systems (AIS) data use in marine vessel emission estimation". In: *18th Annual International Emission Inventory Conference*. Vol. 14. 2009, e17.

[15] Katarzyna Prill et al. "A new method of determining energy efficiency operational indicator for specialized ships". In: *Energies* 13.5 (2020), p. 1082.

[16] V Rajaraman. "Big data analytics". In: *Resonance* 21 (2016), pp. 695–716.

[17] Huilin Ren, Yu Ding, and Congbiao Sui. "Influence of EEDI (Energy Efficiency Design Index) on ship–engine–propeller matching". In: *Journal of Marine Science and Engineering* 7.12 (2019), p. 425.

[18] Miguel Rodriguez, Mario Pansera, and Pablo Cabanelas Lorenzo. "Do indicators have politics? A review of the use of energy and carbon intensity indicators in public debates". In: *Journal of Cleaner Production* 243 (2020), p. 118602.

[19] Dan Rutherford, Xiaoli Mao, and Bryan Comer. "Potential CO2 reductions under the energy efficiency existing ship index". In: *International council on clean transportation. Working paper* 27 (2020).

[20] Marco Siddi et al. "The European Green Deal: Asseasing its current state and future implementation". In: *UPI REPORT* 114 (2020).

[21] Weng Sut Sou et al. "Reducing the carbon intensity of international shipping – The impact of energy efficiency measures". In: *Energy Policy* 170 (2022), p. 113239. ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2022.113239. URL: https://www.sciencedirect.com/science/article/pii/S030142152200458X.

[22] *SparkTM - Unified Engine for large-scale data analytics*. URL: https://spark.apache.org/.

[23] Alastair Stevenson. *Bulker EEXI compliance boosts CII ratings*. SSY online, June 2021. URL: https://www.ssyonline.com/media/2165/eexi-effects-bulkers-jun22rev-1.pdf.

[24] Aydın TOKUŞLU. "Analyzing the Energy Efficiency Design Index (EEDI) performance of a container ship". In: *International Journal of Environment and Geoinformatics* 7.2 (2020), pp. 114–119.

[25] Tien Anh Tran. “A research on the energy efficiency operational indicator EEOI calculation tool on M/V NSU JUSTICE of VINIC transportation company, Vietnam”. In: *Journal of Ocean Engineering and Science* 2.1 (2017), pp. 55–60.

[26] Yuh-Ming Tsai and Cherng-Yuan Lin. “Effects of the Carbon Intensity Index Rating System on the Development of the Northeast Passage”. In: *Journal of Marine Science and Engineering* 11.7 (2023), p. 1341.

[27] *UN body adopts climate change strategy for shipping*. imo.org, 2018. URL: https://www.imo.org/en/MediaCentre/PressBriefings/Pages/06GHGinitialstrategy.aspx.

[28] *UN body adopts climate change strategy for shipping*. imo.org. URL: https://www.imo.org/en/MediaCentre/PressBriefings/Pages/06GHGinitialstrategy.aspx.

[29] Shuaian Wang, Harilaos N. Psaraftis, and Jingwen Qi. “Paradox of international maritime organization’s carbon intensity indicator”. In: *Communications in Transportation Research* 1 (2021), p. 100005. ISSN: 2772-4247. DOI: https://doi.org/10.1016/j.commtr.2021.100005. URL: https://www.sciencedirect.com/science/article/pii/S2772424721000056.

[30] Dong Yang et al. “How big data enriches maritime research–a critical review of Automatic Identification System (AIS) data applications”. In: *Transport Reviews* 39.6 (2019), pp. 755–773.

[31] Ibna Zaman et al. “Challenges and Opportunities of Big Data Analytics for Upcoming Regulations and Future Transformation of the Shipping Industry”. In: *Procedia Engineering* 194 (2017). 10th International Conference on Marine Technology, MARTEC 2016, pp. 537–544. ISSN: 1877-7058. DOI: https://doi.org/10.1016/j.proeng.2017.08.182. URL: https://www.sciencedirect.com/science/article/pii/S1877705817333386.

[32] Shuang Zhang et al. “An alternative benchmarking tool for operational energy efficiency of ships and its policy implications”. In: *Journal of Cleaner Production* 240 (2019), p. 118223. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2019.118223. URL: https://www.sciencedirect.com/science/article/pii/S0959652619330938.