ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
SCHOOL OF BUSINESS

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
DEPARTMENT OF MANAGEMENT SCIENCE & TECHNOLOGY

THESIS

XXXXXXX XXXXXXX

ACADEMIC SUPERVISOR:

DR. XXXXX XXXXX

INTERNSHIP SUPERVISOR:

XXXXXX XXXXXXX

12/07/2021

THESIS

DATA-DRIVEN ADVANCEMENTS IN THE SHIPPING INDUSTRY, A MACHINE LEARNING MODEL DEVELOPMENT ON FUEL CONSUMPTION FORECASTING

Abstract

The shipping sector is the backbone of the worldwide trade, offering a smooth and economical transportation of goods, and is of perennial importance to the Greek economy. However, recently, it has been facing numerous crises and mounting pressure from regulators, and is struggling to recoup its once robust financial prospects.

This thesis examines the opportunities that the shipping industry can capitalize on, through the use of data and data-related technologies. The opportunities of the latter mainly revolve around blockchain (effectually implemented in the TradeLens platform), and automation (in the form of robotics and autonomous shipping). Conversely, its challenges are forcing the sector to adopt stricter cybersecurity policies, and more capable, higher throughput data transfer methods, such as 5G. Regarding data, the data lifecycle framework used splits data into three main parts: sources, engineering, and applications. The main data sources include the Automatic Identification System, Internet of Things applications (both vessels' and ports'), and weather forecasts. Data engineering is the practice of secure, efficient, instantly accessible, and reliable storing of properly cleaned, and reliable data. Lastly, the thesis analyses that through data applications, sector members improve efficiency, and make more factually informed decisions, mostly through the use of predictive maintenance, freight value forecasting, and energy efficiency analysis. Overall, the results show that it would be of utmost importance for the once traditional, laggard, and bureaucratic sector to become highly efficient and technologically advanced.

In order to highlight the value of the aforementioned, a case study is conducted on vessels' fuel consumption forecasting, using noon report data. The study showcases the data cleaning methods commonly utilised in data engineering processes, the data exploration methods that allow easy extraction of business value from the data, and, ultimately, the development of machine learning, predictive analytics, models for the consumption estimation.

Table of Contents

Research Question and Methodology

The research question, upon which this thesis will be based, is the following:

"How can the Shipping Industry utilise the advancements in data and data-related technologies?"

The above query can be further subdivided into multiple research questions, that will form the structure of this thesis. Firstly, the question that the reader will be able to answer is "What are the main data-related challenges and opportunities that the Shipping Industry is facing?". These chapters pave the way for the following data-lifecycle chapter, which is based on the essence of the research question: "What are the main artificial intelligence applications, on what data sources are they based on, and how can those be correctly managed?".

Ultimately, based on the above, the reader will also be able to answer the following questions: "How well has the Shipping Industry utilised data analysis advancements?" and, subsequently, "Which are the data analysis advancements that the Shipping Industry has utilized and which can be utilised further, going forward?".

The methodology followed was the initial search for the main topics concerning the research question, to form the initial bibliographic structure. Upon finding the main research areas and before their inclusion in this thesis, the volume and quality of their bibliography was assessed by searching prestigious online databases and content aggregators, such as Google Scholar, Elsevier's Science Direct, and IEEE.

To structure the bibliography of the data and data-related applications used in the industry, a search to find the most appropriate data lifecycle model was conducted. The chosen one was the comprehensive scenario-agnostic data lifecycle (COSA-DLC) framework, which allowed the division of the data-related literature into data sources, data engineering, and data applications.

The prioritization of the most academically sound and credible sources was of utmost importance. Firstly, most of the references were- in order- books, and published journals (white literature), followed by prestigious conferences, government and institutional reports, lectures, theses, and well-known websites (grey literature), with the use of blog posts and questionable sources (black literature) being restricted.

Furthermore, it is important to disambiguate between the terms "Maritime Industry" and "Shipping Industry", that are found in the bibliography. The former is defined as the collection of any activity and object that is related to the sea, seafaring, vessels, navigation, ocean; thus, it is a general term, describing an abundance of things. Conversely, the latter is defined as the specific act of transporting cargo (or "freight") from a point (for example, port) to another, using a vessel, and thus it is considered a subset of the Maritime Industry. Sometimes they are used interchangeably, and shipping can also refer to the non-maritime part of shipping (e.g., train freight). In this thesis, the term shipping will always refer to maritime shipping, unless otherwise stated.

1. Introduction

Shipping intelligence will be a force shaping the industry moving forward and, thus, is becoming the centre of attention in the industry. The sheer amount of diverse data allows analyses of correlations, patters, and trends. It improves performance monitoring and prediction, transparency, decision-making support, and human error reduction, but comes with multiple challenges, such as data engineering and processing, reliability, and data security (Zaman et al., 2017).

McKinsey & Company's Jie et al. (2020) recently offered an extensive report on how data will be the main factor to decide success, moving forward to the next normal in the future of shipping, under the prism of the coronavirus crisis and inauspicious commodity trends. They describe that the medium-term crisis that the industry is experiencing will allow the companies leveraging data (including market insights data) to come out of it unscathed, as market leaders. Even in this challenging environment, data accessibility (concerning the market insights, analytics, and customers) will allow an industry with a historically instinctive and opinion-based decision-making approach to transition into a fact-based, analytical one. As is evident by this thesis, the investment (of time and money) for the aforementioned will be sizeable, but necessary, in order to come out of the current crises as winners and market leaders.

Furthermore, the authors also described that being a market leader in the future of the sector is becoming extremely important, since the competition is expected to become fiercer. Due to the business nature of McKinsey, the authors offer insights into the business perspective of analytical applications, focusing on areas like attractive subsector and niche markets analysis, portfolio risk-benefit optimization, commercial choice improvement, and vessels operation optimization, which are mostly based on the applications mentioned in the relevant chapter of this thesis.

McKinsey also forecasted a chasm between the growing supply (increased capacity) and the staggering growth of the freight demand, leading to decreasing freight rates and sustained oversupply. This might prove to be a good forecast in the long-term, but, short-term, it seems that the charter rates are skyrocketing (reaching a 150% increase, in some cases), forming a bull run of a ten-year high. This is attributed to a domino effect, that started in the pandemic. It started by the cancellation of man charters, due to diminishing rates (attributed to the decreased demand, during the initial stages of the lockdown). The bottleneck, that resulted in the steeply-risen rates, was created by the increased consumer demand of the latter lockdowns, leading to the demand for containerized goods greatly surpassing the available supply. Lastly, the short-term chasm is also amplified by the limited vessel deliveries during the pandemic, as a result of decreased shipbuilding operations (Lademan, 2021).

Another report showcasing the importance of the use of data and data-related technologies was published by the United Kingdom's Department for Transport (*MARITIME 2050: Navigating the Future*, 2019). In the "Technology" chapter, it stresses how big data analytics, digitalization, and advanced communications are essential to improving connectedness, efficiency, and minimizing cost; but do not come without risks, such as cyberattacks. The report outlines UK's mobilization in the quest of becoming the world's leader in the use of smart shipping technologies. The proposed steps are legislative advances to accommodate modern technological applications (such as autonomous vessels), improved funding of state-of-the-art projects, increased efforts and funding in the relevant Research & Development hubs, and incorporation of its- increasing- smart ports in this endeavor.

Tian et al. (2017) divide big data into three main categories: business data (concerning financial information, such as markets, rates, and indices), management data (concerning operations of vessels and ports), and supervision data (concerning the government and customs, operational safety, and credit risk). This thesis offers a business perspective on the technical applications and, as such, will mostly mention the management data, as data sources, but will mention the others in their relevant applications. Future work could include a more extensive look into the other categories.

2. Shipping Industry

2.1 Introduction to the Shipping Industry

The process of transporting goods has been a main focus of humans' economic activities for many millennia. Starting with the Egyptians and sail ships (3.200BC), when it started being the central support of global goods' trade, to the Egyptians' expansion of it to reach Sumatra (1.200BC) and then the Chinese establishment of shipping routes on the South China Sea and the Indian Ocean on the 10th century. Continuing with the took over of shipping activities by European countries (such as Portugal, Spain, and England) that considered shipping of utmost strategic importance for their trade and colonial aspirations, around the 16th century (Rodrigue, 2020).

Nowadays, even though global trade networks are facing incredible challenges (such as the COVID-19 pandemic and trade wars) global shipping has withstood and reached a volume of 11.08 billion tons for 2020, albeit expanding by only 0,5% in 2019 (in contrast to 2,8% in 2018) (Asariotis et al., 2020).

According to the International Chamber of Shipping (2021), the shipping industry is responsible for 80% of the total trade (exports and imports) by volume and 50% by total value, namely 14 trillion USD. Similarly, CoS states that the total number of seafarers reaches 1,65 million. Therefore, the vast scale of shipping becomes immediately apparent and with such high numbers, even a small improvement in the business model, will yield increased profits that may total billions.

2.2 Greece as a key player

It is worth noting that the Shipping Industry has been of utmost importance for the Greek economy and society in general. Conversely, Greece has an important place in the global shipping sphere. While Greece has faced numerous financial problems, its shipping industry never ceased to be an important pillar of the Greek economy, being one of the largest in the world, both in terms of fleet and of gross tonnage; mainly based on Greece's main port, Piraeus (Pallis, 2007). Even though the Greek shipping industry withstands some competitive disadvantages, such as a weak local demand (in contrast to the Japanese shipping operations, for example), the main elements that contribute to its success are the country's location, strong local institutions, and a favourable local tax regime (Icaza et al., 2009).

The numbers above might be misleading when one thinks about the impact that shipping has on the Greek economy. Icaza et al. also describe that, even though the above information sound impressive, the impact the shipping industry has on the Greek economy is only moderate. The goods transported mainly do not originate or move towards Greece and the tax schemes do not allow a big portion of the revenue to help the struggling local economy. However, Theodoropoulos & Tassopoulos (2014) describe that even though the seaborne transportation services do offer substantial earnings to the economy, more balanced shipping activities could enhance the impact that shipping has on it.

It is worth noting that the Greek shipping industry does not only affect the industry worldwide financially, but also by the volume of research Greek scientists and institutions produce. An intriguing meta-research conducted by Munim et al. (2020), specifically in the field of Big Data and Artificial Intelligence, indicated that a big portion of the research comes from Greek institutions, such as University of the Aegean and University of Piraeus, and from Greek scientists; we need only look at the bibliography used in this thesis.

2.3 Challenges and Opportunities

The industry has seen great expansion throughout the last few decades, going from being responsible for 23% of the world total volume in 1980, to ~85% in the last few years. The total TEUs (Twenty-Foot equivalent units) and total vessel capacity have been growing exponentially, due to the growing demand for global trade, the shifts in the global geopolitics and, most importantly for this thesis, the continuous technological advancements upon which it has capitalized (Caschili and Medda, 2012). After all, each industrial revolution has greatly impacted the shipping industry and the 4th one does not differ in that aspect. Estimates are foreseeing seismic changes to consumer demands and to the way goods, services, and energy are transported (Rex et al., 2017).

The 4th industrial revolution and the digitization of the modern age (that new shipping business models will rely on) offer great opportunities for increased profits and efficiency. However, it is worth noting the challenges that will be faced are important and should be swiftly and thoroughly addressed, to avoid destabilizing the current normal functioning system, that keeps the world's economies afloat. From cyber-security, to the legalities of an autonomous ship, there are numerous problems that need to be addressed before it is too late (Alop, 2019).

However, for today's standards of constant change and adaption, the industry has lagged behind in adapting to new challenges and opportunities that constantly arise in the world trade environment. With that being said, change is still being carried out and the industry as a whole is picking up steam on this front. That is mainly due to the monumental challenges and opportunities that the world is facing currently, which will be presented below and will pave the way- later in this thesis- for the analysis of specific new technologies that are being adapted.

2.3.1 Global Warming

The world is facing a change in its weather patterns and temperature like never seen before. This change has especially affected the Shipping Industry in two specific areas, predicting and preparing for ever-worsening natural phenomena (like thunderstorms and ice movements), in order to be able to keep their operations safe and efficient, and adjusting to the constant push for less carbon emissions (Wan et al., 2017). Additionally, the European Council for maritime Applied R&D, in its Maritime Technology Challenges 2030 report, states that- besides digitization- decarbonatization is its vision for the future of the European Waterborne Sector.

2.3.1.1 Weather Fluctuations

Vessels operate in the open seas, being victims of one of the harshest environments on Earth. Thus, the shipping industry is always trying to minimize the chance of something going wrong due to the weather, which could hinder the crew's safety, the shipping company's bottom line and- in some extreme cases- the world's trade. Weather fluctuations do not only affect long haul international trade routes, but also short sea shipping (de Osés and Castells, 2008).

Consequently, shipping companies and institutions need to increase their forecasting capabilities and provide better instructions to the vessel's crew. This is especially important in an era when the world's climate has been altered greatly (leading to sudden and unexpected conditions) and in hazardous territories, such as the Artic, where sea ice forecasts play a crucial role (Gascard et al., 2017). The ship routing based on weather forecasts has also had a great impact on the total operating cost of the vessel (*RESOLUTION A.893(21)*, 1999). As such, Perera & Soares (2017) have proposed a routing system, that suggests transportation routes, before and during a vessel's voyage, based on weather forecasts, vessel location, orientation, and speed conditions. It can be used in future bridge systems, to facilitate the need to reach optimal and safe navigation, while minimizing environmental impact. As will be discussed below, data and data analysis can play an integral part in implementing applications such as the aforementioned one.

2.3.1.2 Decarbonization

Decarbonization has been the epicentre of the race to reach a carbon neutral footprint and reverse the effects of human intervention on the environment, before it is too late. While the Shipping Industry had enjoyed a relative freedom in this aspect, due to its nature of operating in alienated oceans, nowadays there have been numerous attempts to regulate its environmental footprint. Many worldwide organisations (such as the European Council for maritime Applied R&D (2030), the European Parliamentary Research Service (Pape, 2020) and a joint Deloitte and Shell report (Shell; Deloitte, 2020)) believe decarbonization to be one of the biggest challenges that the industry ought to address. However, the light at the end of the tunnel is "not yet" here (Psaraftis and Kontovas, 2020).

Wan et al. (2017) explain that, with their current pace, vessel-related Greenhouse Gasses (GHG) emissions have been estimated to surge by 250% from 2012 to 2050, due to the ever-increasing demand for cheap freight. The authors also provide a critical view into the steps already carried out by international organisations and shipping companies, and advise as to how to implement better planning on decarbonization. Specifically, better planning can be achieved by making sure that future solutions are interdisciplinary robust and meaningful (from a market, technical and operational standpoint), in order to be feasible. At the same time, world synergies should be formed to facilitate this move going forward. Another pathway is also paved by Rex et al. (2020), who suggest that, while uncertainty about possible feasible decarbonization solutions remains, the key to shipping decarbonization will be a regulatory spark, coupled with profitable and well-thought-out business plans (that aim to make zero-carbon-fuelled vessels economically and strategically attractive).

While the need for decarbonization may be clear, as is evident by the above, the path to achieve it is convoluted and filled with bureaucratic, financial, political and technical hurdles. Hence, there needs to be an evaluation of the proposals. A common solution circulating, is that of Market-Based Measures (MBMs), which are mechanisms that urge emissions' decrease through economic variables and penalties, therefore adding a financial cost to polluting. Though the current ones have been found to not be drastic and swift enough (Lagouvardou et al., 2020; Psaraftis et al., 2021). In addition, measures like the Emissions Trading System (ETS), which is expected to include the shipping industry, have faced opposition from it, with various alternatives being proposed, such as the direct increase of the fuel cost, based on its environmental impact.

A more technical evaluation has been conducted by Mallouppas & Yfantis (2021), in which they proposed technical and operational strategies to achieve IMO's 2050 targets. They concluded that poly-fuels, LNG and Biofuels are promising fuel types that have a reduced environmental impact, albeit needing iterations in their development. Similarly, renewable types, like wind, solar, and biomass seem promising. Thus, a blended fuel mixture, coupled with CO_2 abatement might be the short-term way forward.

Besides mechanical and chemical solutions, another field of engineering will come into play in achieving a carbon neutral footprint; data science. Data is already utilized in this problem's possible solutions, but, moving forward, leaps are expected in its implementation. From estimating air emissions from vessels (Miola and Ciuffo, 2011) to using Bayesian simulations & time series forecasting (Jiahui Liu et al., 2021) and machine learning & optimization methods (Gkerekos et al., 2019a; Lu et al., 2013; Safaei et al., 2019a), there are countless proposals for immediate applications and not only will they be analysed below, but a case study on this very subject will be presented.

Covid-19 has proved to be another hurdle in the decarbonization process. The shipping industry has hit a roadblock, in reduced demand for freight, resulting in initial reductions in overall GHG emissions. Still, that means that the reduction in fuel cost makes fuel consumption, and- in turn- emissions, an issue of reduced significance for shipping companies, thus forcing them to not focus on it (Lagouvardou et al., 2020).

2.3.2 Covid-19

The Covid-19 pandemic came unexpectedly in 2019 and had a vast impact both socially and financially. Every "challenges" section in current literature reviews has a "Covid-19" subsection. As such, this thesis would be incomplete without presenting the impact the latest coronavirus has had in the Shipping Industry.

In the effort of governments to limit the spread of the virus, strict restrictions have been imposed, which have been found to have severely hampered global trade, with some lockdown tactics being more fruitful and less damaging than others (Guan et al., 2020). However, Verschuur et al. (2021) suggested that the estimated damage is- on occasions- exaggerated, and some countries have even briefly benefited (like Vietnam increasing manufacturing in the wake of China's strict lockdown).

Those large-scale, but necessary, containment actions greatly affected ports, especially in Western Europe, China and the Middle-East, together with certain supply chains, such as oil and vehicle manufacturing. The sector suffered a great impact, with an estimated ~8% reduction in total trade (or ~246 million tonnes or ~320 billion USD in losses), during the first eight months of 2020 alone (Verschuur et al., 2021b). Not only was the size of worldwide trade reduced, but the structure of also it changed, with the interconnectedness and density of the trade networks having experienced alterations by the pandemic (Vidya and Prabheesh, 2020).

However, the damage seems to vary, depending on a lot of factors (Michail and Melas, 2020). One of those factors is the type of commodity being transported. For example, it has affected the dry bulk and dirty tanker segments, but not the clean tanker one. Due to the diminished price of oil, an important step was storing oil in the vessels, to defend the oil market and wait out the price drops. Additionally, the authors identified a strong relationship between freight rates and coronavirus cases, and noted that the added volatility will especially hurt shipping hubs, such as Greece, with reductions in output and job losses.

The crisis might sound familiar, since the industry had already faced a financial crisis in 2008, but the effects have been found to be quite different to the current crisis, with the industry showing increased resilience this time (Notteboom et al., 2021), even though the trade contraction is greater than the 2008 crisis (Majid, 2021). As a result, in order to counteract the effects of the crisis, the shipping industry had to quickly adapt fast and adopt leaner and more efficient operations, which it did with better results than the aftermath of 2008. However, most importantly, the pandemic helped the shipping industry adapt to more technologically advanced operations, jumpstarting the technological transformation of certain operations (like the paperwork between the vessels and the ports) (Notteboom et al., 2021; Notteboom and Haralambides, 2020). This paved the way and facilitated the adoption of the applications that will be presented next in this paper.

2.4 New Technologies

Even though this paper is mostly focused on how data and data applications are used to transform the industry, it is worth noting some data-related technologies that are helping shape the future of it. The 4th industrial revolution has forced the industry to adopt those technologies, not just proactively but, also, reactively in some cases.

2.4.1 Cybersecurity

The 4th industrial revolution has accelerated the integration and interconnection of systems, such as ERPs, between their components (for example, vessel and Headquarters), each other and the Internet. However, this opportunity presents a major risk: if every detail of our operations is easily accessible by our organization, and the companies with which we do business, then it may, also, be accessible by unwanted third-parties. Thus, we need a modern defense mechanism, i.e., Cybersecurity.

Even though this sector is critical in a company's security mechanisms, surveys show that, even though many organizations have experienced costly cyber-attacks, they still perceive the chance of it happening as quite low. This is forcing the shipping sector to trail other industries in cybersecurity measures, such as more robust systems and networks, and better guidelines, standards and training (Alcaide and Llave, 2020; Heering, 2020). Certainly, ports are also facing similar threats and need to follow a well-planned-out cybersecurity hygiene (Senarak, 2021).

Tam & Jones (2018) excellently detail the defense mechanisms that can be used against their corresponding attacks, focusing on the vessels. They divide the possible attacks into four categories. First, malware vulnerabilities, such as malware on the bridge via USB and software update attack on an autonomous ship. Second, jamming vulnerabilities, like close proximity jamming aided with social engineering and shore-based jamming, to prevent or delay operations. Third, denial of service vulnerabilities, like denial of sensor readings for critical operations and chokehold traffic jams. Fourth, spoofing vulnerabilities, such as GPS spoofing and AIS misdirection. The authors suggest detailed policies for combating those attacks, from cyber-physical to purely cybersecurity.

Specifically for the case of autonomous vessels, which are becoming increasingly popular, the danger of fatal outcomes from a cyber-attack is formidable. Therefore, their vulnerability for external intrusion must be minimized, partly by using optimal and robust firewalls (Shapo, 2018).

2.4.2 5G

Vessels are independent entities that constantly travel the Earth's oceans. This means that one of the most important aspects of their operations is communication. Additionally, with the 4th industrial revolution (including the data and data analysis applications that are being adopted), the amount of data that needs to be transported is exponentially increasing (Yuen et al., 2020). The need for ultra-fast communication speeds is expected to be met by a state-of-the-art technology, 5th generation cellular networks (5G), which are revolutionising global communication networks (Agiwal et al., 2016; Navarro-Ortiz et al., 2020).

Once the shipping industry realises its potential, 5G will allow great advancements in its operations. These include- but are not limited to- communications, maintenance, operations, navigation and the advancement of- semi or fully- autonomous vessels; even if its adoption poses new risks for the industry (Agarwala and Guduru, 2021). Additionally, it will facilitate the revision of data acquisition, data processing and data applications, the development of system applications, infrastructure, management and control systems, and the formation of integrated information services, between the vessels, ports, company's headquarters and the customs (Jing and Zheng, 2020). The 5G network could, also, revolutionize the sea safety of people and goods onboard vessels (Gajewska, 2019).

Lastly, it should be noted that 5G is not a magical protocol that will allow vessels to instantaneously communicate with other entities, in the middle of the ocean, due to the high cost and low proximity of 5G antennas. The enablement of the 5G technology in remote areas has been an interesting research topic. To combat this problem, and decrease the speed differences between the 5G network and maritime communication networks, Li et al. (2020) propose forming a hybrid network. It will combine pre-existing (or advanced iterations of) shore-based terrestrial stations and marine satellites with unmanned aerial vehicles, taking advantage of the fact that vessel movement is based on standard sea lanes. Similarly, some of the previous paper's authors expanded on the hybrid communication idea (focusing on terrestrial in conjunction with high-throughput satellite hubs, like Starlink) for the requirements of the increasing IoT devices, in order to prevent the performance loss current systems experience. Additionally, they divided the technologies into three categories, boosting transmission efficiency, growing network coverage performance, and technologies responsible for providing maritime-specific services effectively (Wei et al., 2021). It is worth noting that the latest research is even starting to discuss the applications of 6G (R. W. Liu et al., 2021).

2.4.3 Blockchain

Another new technology that is gaining traction in the shipping industry is blockchain. Blockchain is a decentralized, distributed database of transactions, or events, that has been shared and agreed upon between the corresponding parties, thus becoming a robust source for secure and verifiable records, that cannot be changed or "hacked". While it may be a popular buzzword due to the popularity of cryptocurrencies (such as Bitcoin), its opportunities are limitless (Crosby et al., 2016; Underwood, 2016).

The blockchain technology has not completely matured, but it is poised to become exceptionally useful in certain applications. These include digitalizing paperwork, accelerating customs' clearance and management, and reinforcing cross-border trade security and safety, offering leaner, faster and more secure paperwork and operations (Yang, 2019).

As can easily be deduced by the above, fighting bureaucracy is one of the hardest tasks one can undertake. The proposed applications have been estimated to offer vast cost advantages (~25% cost reduction, ~40% reduced delivery time). However, the market readiness for blockchain is lacking, with the lack of a single standard and the reluctancy of senior shipowners, local actors and regulators posing as a roadblock to its swift and universal adoption (Bavassano et al., 2020). Even countries that have suffered from bureaucracy, like Greece, which have the most to gain from the adoption of blockchain, seem to be very sceptical of it (Papathanasiou et al., 2020). The problems for this mechanism do not end there, with Jabbar & Bjørn (2018) describing that an infrastructural grind exists between the existing shipping information infrastructure and the proposed blockchain one. However, this reluctancy and infrastructural problems seem to be abating, mainly with the introduction and promotion of the TradeLens platform. TradeLens is a blockchain platform, made from IBM and Maersk, aiming to incorporate the blockchain technology in the sector, in a move usually described as sector revolutionization.

As explained by Todd Scott (2018), Blockchain Vice President at IBM, TradeLens, which was introduced in 2018, is a global trade platform, that aims to decrease the global shipping cost, increase trust and transparency with trading partners, and expand visibility between supply chains (mainly, but not exclusively, in the shipping sector). Additionally, it aims to reduce human errors, eliminate bureaucratic barriers, and optimize processes; i.e., modernize supply chains by digitizing them. It is built from the start as an open blockchain platform, based on multiple stakeholder’s feedback, which was crucial in their pursuit of onboarding as many organisations as possible. These include carriers, ports, and companies who need huge amounts of their products transported daily, resulting in the creation of not just a technology solution, but a vast interconnected ecosystem. Its members upload huge amounts of data to the platform, through the platform’s APIs, with millions of shipping events and containers had already been uploaded in less than a year after the platform’s introduction. TradeLens provides members with an encrypted, integrated, shared access to the transaction data, if- of course- they are allowed by the respective owners of the data.

TradeLens faces three unsolved barriers to global trade: costs, uncertain lead times, and transactional security. Extensive research has been conducted to combat them, without having found any effective solutions. Nevertheless, it provided the foundations for the TradeLens platform, which- if adopted adequately, as is expected- has the potential to transform the global trade (Jensen et al., 2019). The adoption has been wide but, while the advanced technologies used (such as blockchain, hybrid cloud storage, and modern data exchange standards) will transform the industry, the organisations are challenged with internal obstacles of adoption. They must be overcome by moving to a proactive scheme of real-time data supply in a modern logistical data network (Louw-Reimer et al., 2021). Thus, to combat those technological problems, the big players in the industry need to actively participate in the effort to overcome the technological challenges of the technology’s developments (as Maersk is doing). Subsequently, the smaller players, who do not have the resources to help in the former, should focus on developing the capabilities to be prepared for its adoption, since it is vital for their future viability (Jacob et al., 2019).

In addition, the stakeholder analysis and platform development discussion should actively involve policy creators, to provide information and be informed enough to adequately form relevant and appropriate regulation and useful legislation (Jugović et al., 2019). Policy creators should also be involved because the platform is expected to create great value for the government, since it also revolutionises the business-government information sharing (mostly through customs) (Rukanova et al., 2021).

Lastly, it is apparent that the implementation of blockchain needs to be thoroughly studied, allowing all of the stakeholders to be a part of it. After conducting a stakeholder analysis, Loklindt et al. (2018) investigate four possible approaches of implementation and propose a set of eight design principles, namely: immutability, decentralization, security, privacy, compatibility, scalability, inclusiveness and territoriality.

2.4.4 Automation

The last new technology that will be presented is automation. The major leaps that Internet of Things, artificial intelligence and robotics have taken, have allowed automation to grow and mature, which has simultaneously matured the arguments against it (Acemoglu and Restrepo, 2019). This has allowed the automation of certain shipping processes to be at the forefront of the technological advancements that the industry is undergoing (European Council for maritime Applied R&D, 2017). The use of automation in the industry is twofold; it can be used to automate certain procedures with the use of robots (for example, the transfer of goods in ports), and it can, also, be used to fully automate a vessel’s navigation, making it fully autonomous.

Firstly, the use of robotics has some critics, due to its dependency on artificial intelligence, the loss of jobs that it may cause, the over-reliance on sensors, reduced situational awareness, and increased complacency (Hannaford and Hassel, 2021). However, it is expected to offer increased speed and safety for the crews in various applications, such as the transfer of supplies to the vessels which are near the port (Muhammad et al., 2018) or the order of object removal from a container (Mojtahedzadeh et al., 2015). The use of robotics will be further analysed in the ports’ Internet of Things chapter.

Secondly, a great amount of research is being conducted on autonomous vessels and their future impact in the industry (Munim et al., 2020). The introduction of autonomous vessels will open up a new era and improve cost efficiency, accident prevention and human resources, if the stakeholders manage to adjust safety, security and environmental protection regulations and conventions (Kim et al., 2020). The benefits can also be divided into social (for the staff), economical (with cargo increase and better space allocation), environmental (with optimized routing and speed) and safety-related; achieving overall better operational performance (Batalden et al., 2017). The numerous regulatory, operational, and quality assurance challenges can be overcome with careful planning (Komianos, 2018).

The transition between the current environment and an autonomous one should be carefully planned out. The planning has to include a formulation of clear roles, responsibilities and definition of autonomous vessels’ operator abilities, to certify that, even with a high level of autonomy, the future operations will have a human-centred element, to ensure higher safety (Praetorius et al., 2020).

The increased safety is an important aspect of this endeavour. Porathe et al. (2018) state that human error is the cause of 70-90% of maritime accidents and that this number will be diminished with automation. However, other dangers, which humans currently effectively combat, may arise with the introduction of autonomous vessels. Thus, a hybrid system is the best way forward, with dynamic risk assessment and other various real time monitoring tools. De Vos et al. (2021), also state that an increase in maritime safety will mainly be attributed to the reduction of people that will be onboard the autonomous vessel (especially on smaller ones).

3. Shipping Data Lifecycle

The previous chapter included the introduction of the Shipping Industry and- focusing on the data aspects- the challenges and opportunities faced. It also included the new, promising, data-related technologies that already are- or will shortly be- utilised. This chapter will present the actual data side of the advancements that the Shipping Industry is undergoing, following the data lifecycle as a structural pathway.

Multiple data lifecycle models have been proposed- depending on the volume and the sector of the data being used. They provide an effective framework by focusing on the stages the data go through, in order to improve its robustness, security, interconnectability, quality and reliability (Plale and Kouper, 2017). However, in this chapter, it will be used solely for structural purposes, following a similar approach to the comprehensive scenario-agnostic data lifecycle (COSA-DLC) framework (Sinaeepourfard et al., 2017). The components of this framework are data acquisition, data preservation, and data processing. They will be translated into three parts. First, an analysis of the various data sources that offer datasets of great value for the industry. Second, an analysis of the data engineering methods that are being deployed to handle the ever-increasing volume of data. Third, an overview of the various data analysis/ machine learning applications that are expected to revolutionise the way the industry operates and makes decisions, respectively.

That overview is of great significance, because it emphasizes the value of the data created, and the endless analysis opportunities. Shipowners should be aware of them, since the data gathered on vessels might be increasing, but the involvement of it in the decision-making process is still not as extensive as it should be (Aiello et al., 2020).

3.1 Data Sources

The first step in the data lifecycle is the process of collecting the data that will be stored and, subsequently, analysed to extract value; i.e., data acquisition. The data that the industry produces, and/or uses, come from an amalgamation of both pre-existing data sources, that the industry has had before the modern data-related advancements. They include the Automatic Identification System (AIS) and new state-of-the-art technologies, such as the Internet of Things (IoT) devices and sensors, that are increasingly found onboard the vessels. In addition, there are various business data (e.g., budgets) that every company can integrate into their database, but this will not be the focus of this chapter.

3.1.1 Automatic Identification System (AIS)

The Automatic Identification System is used by vessels which transmit data (either to AIS base stations positioned on shore, or on satellites) related to the vessel's characteristics (name, size, and type), the position (coordinates), speed, direction (heading), and voyage-related data, such as draught, origin port and destination ports. Initially, it was used to prevent collisions (an issue in the centre of attention due to the dangers such an event poses to human lives and revenues). Lately, the data has proved to be of immeasurable value to researchers who seek to analyse vessel movements, and, of course, to the shipowners, who seek to analyse their vessel's positions and routes (Z. Yan et al., 2020). It is mandatory for vessels (of more than 300 gross tonnage) to have the capability and transmit AIS data. Still, in certain situations they might stop transmitting, due to some international agreements, rules and standards, or in cases where navigational information protection is important (IALA, 2004).

Besides preventing collisions and analysing vessel routes, data are important in detecting instances of illegal activities. Since every large vessel's route is monitored, carrying loads that contain illegal substances, or ignoring international sanctions (e.g., transferring fuel to North Korea), needs a spoofing mechanism, in order to circumvent tracking. Thus, a mechanism has been developed to monitor cases where AIS is predicted to have been spoofed (Prasad et al., 2021).

The analysis applications that the AIS enable are endless. Due to this reason, D. Yang et al. (2019) divided the academic applications into seven categories: AIS data mining, navigation safety, vessel behaviour analysis, environmental impact assessment, trade analysis, vessel and port performance, and anything related to shipping in the Arctic. Besides the applications, the AIS data sources and analysis methodologies are abundant, so a study of data source and methodology assessment has been developed (Tu et al., 2018). To understand the essence of the AIS applications, some of the most promising ones will be presented below.

Firstly, a destination prediction machine learning model is proposed, basing the prediction in the comparison between the vessel's trajectory and historical trajectory data, using a random forest-based model. The model helps improve the destination data, and the ETA of the vessels to the ports, enabling better scheduling by them (Zhang et al., 2020), since the destination port information is often not reliable (with erroneous data reaching an estimated 40%)(Yang et al., 2021). Similarly, using the vessel's trajectory from AIS data, a model has been developed for vessel route grouping and anomaly detection, to improve the collision avoidance efforts (Rong et al., 2020). Also, another outlier detecting model has been developed, using IoT data, transferred by 6G, with the use of bidirectional long short-term memory (BLSTM)-based models (R. W. Liu et al., 2021). The global shipping density metrics are quite useful, both for safety and route value assessment, so there has been research on correctly mapping shipping density, from AIS data (Wu et al., 2017). Additionally, the data can help in monitoring global oil trade routes (Z. Yan et al., 2020), which is extremely valuable for oil trade analysts.

It should be noted that AIS's implementation and working protocols are prone to cyber-security threats (such as spoofing, altered weather forecasts, replay attacks, and availability disruptions), so shipowners should be alert about the increasing possibility of attack (Balduzzi et al., 2014; Help Net Security, 2013).

3.1.2 Internet of Things

The Internet of Things (IoT) is a concept that has risen to the forefront of the novel applications related to the 4th industrial revolution. In essence, IoT is the integrated online presence of a plethora of objects, using RFID tags, sensors, actuators, and- generally- devices enabling data collection and instantaneous transmission to the Internet, while achieving greater interconnectedness between them, to achieve common goals (Atzori et al., 2010). This forms a dynamic worldwide network of connected devices, allowing for the development of many novel applications, based on real-time gathered data. Its development spans the fields of networks, computer architecture, data processing, management, security, and analysis (Borgia, 2014).

One of the most promising and valuable fields of applications of IoT is logistics, with the shipping sector having the opportunity for a significant increase in its efficiency. For example, it is able to improve the transfer efficiency between the loading and unloading ports (by the use of RFID to transfer freight information instead of paper documents), container monitoring during transportation, and automating the process of loading and unloading the containers. But achieving maximum added value from IoT will happen once we overcome existing problems, such as lack of standards, costs, and security risks (Song et al., 2012).

The applications of IoT in the shipping industry are two-fold. Firstly, it enables the interconnectedness of the sensors onboard a vessel, allowing data from the equipment and systems to be collected and transmitted- live- to the shipowners. Next, it enables the ports to collect the data that are necessary for streamlining their operations, reducing paperwork, and increasing autonomous procedures, while exchanging data with the shipowners to achieve faster and leaner operations.

3.1.2.1 Vessels' IoT

Since the vessels are invaluable machinery roaming the oceans, it is very important for shipowners to monitor their status (consequently, the onboard equipment's status) and apply real time analytics applications, while storing the data for future applications that will be based on historical data, increasing efficiency, transparency, and safety. The IoT data onboard vessels is experiencing exponential growth, with researchers analysing whether 6G is needed to harness the ocean of IoT data, and whether modern GPU-enabled computational methods are enough (R. W. Liu et al., 2021). Therefore, it is essential for shipbuilders to incorporate these new technologies and create efficient, safe, and sustainable vessels, even in the building phase (Muñoz and Pérez, 2017) and use its data in the design phase of the vessels (Oneto et al., 2016).

The onboard sensors and data acquisition systems (DAQ) can be utilised to collect data that can be analysed to evaluate vessel performance, under various conditions, and can be a part of the ship energy efficiency management plan (SEEMP) (Lokukaluge P. Perera and Mo, 2016; Perera and Mo, 2017a). Ship performance monitoring and analysis systems (PM&A) do not, historically, have robustly accurate data, but many improvements have been proposed (Hasselaar, 2011). Since the DAQ systems are not reliable, noon reports are continuously used for vessel operational performance monitoring, but they have extensive limitations (such as limited frequency and features). Thus, the DAQ systems need to be improved and coupled with sensor data (Soner et al., 2018a).

Besides the occasional erroneous data, which is mostly related to the manual input from crew that is not experienced with technology, two other issues need to be resolved, to reach optimum data collection from the DAQ and sensors. Firstly, the retention of data is only short-term onboard the vessels, which does not allow for historical data analysis, or machine learning model training. Secondly, the sampling rate is sparse, with systems that produce rapidly changing data (like engine revolutions) not being monitored properly (Tinga et al., 2017).

Integrated bridge systems (IBS) contain the navigation systems (including the Electronic Chart Display and Information System (ECDIS), autopilot system, and radar) and automation systems (including power management, engine and propulsion controls), and are fitted with the DAQ and sensors to create the "onboard Internet of Things" (Perera and Mo, 2017b). The communication of those sensors and systems is also important, with many standards existing, mainly from the National Marine Electronics Association, which enable consolidation, arrangement and display of multiple sources (Lee et al., 2009). Besides intra-vessel data communication, inter-vessel and shore-based communication has been studied, with a unified machine-type communication (MTC) system having been proposed, to accommodate IoT data exchange (M. M. Wang et al., 2020).

The data collected from the vessels can be plentiful and can, also, be separated into two main categories. Firstly, those collected from onboard equipment, such as Voyage data recorder, Engine Data Logger, and Ballast Control System. Secondly, those collected from onboard and shore applications (such as optimum trim, engine monitoring, and energy managements system), with their integration being a main focus. It is worth noting that the integration of a single data lake, to which every data source feeds data, is important for the vessel's data management. That is due to the fact that implementations of data architecture still offer sparse data integration, without having a single master database (Yoshida, 2016); as will be analyzed in the data engineering chapter.

It was argued above that IoT will enable the shipowners and the crew to be more aware of the current status of their vessels, and the shipowners to perform state-of-the-art analytics, since the stream of data will be sizeable. Concurrently, besides the information transfer between the vessel and the shipowners (or crew), IoT will allow greater sharing of status data between the vessels themselves. That will make them more implicitly interconnected, and allow better coordination and explicit quantification- into data- of current "good seamanship" practises; necessary for the implementation of autonomous vessels (Heikkilä et al., 2020).

The authors above have also proposed six technology levels for marine traffic coordination (Wahlström et al., 2019), which can act as a basis for the assessment of marine coordination, stemming from the vessel's IoT interconnection:

- Level 0: "I can see you", meaning that the information sharing is based on visual and vocal cues, thus necessarily using piloting services in unknown situations; i.e., the baseline
- Level 1: "I can see your flags", meaning basic technology usage, limited to shared signalling tools and lights
- Level 2: "My radar sees you", meaning the use of technologies that allow signal transmission, such as positioning via satellites, radar, and communication
- Level 3: "I can see data about you", meaning sharing stored data between vessels

- Level 4: "My robot sees you", meaning the use of Artificial Intelligence and Machine Learning models, for various use cases, such as object detection and autonomous navigation
- Level 5: "My robot sees what your robot sees", the most advanced level, allowing Level 4 systems from various vessels to automatically exchange data, forming a self-organising system, without the need for human intervention; the ultimate cornerstone of technology usage in the sector

3.1.2.2 Port's IoT

The Internet of Things has not only disrupted the traditional way that the vessels functioned, but also the ports' operations, which has led the research community to refer to the modern ports as "smart ports". Starting the transition in 2000, ports have started to become "smarter", using new technologies to streamline the data exchange and communication between the governments, freight managers, shipping companies, and shipowners, increasing the ports' service level, and decreasing operational times (Wang and Liu, 2012), with this progression seeming unstoppable (Karas, 2020).

This is achieved by the effective, real-time, and rapid collection, processing, exchange, and use of data, due to the use of state-of-the-art technologies, such as IoT (most importantly), cloud computing and AI. This allows ports to overcome problems of the past, like low efficiency- due to bureaucratic obstacles-, cumbersome data integration and inter-connectedness, and slow response times, to reach optimum levels of efficiency, accuracy, and safety (Li et al., 2018). Similarly to the vessels' technology levels above, the previous paper's authors have proposed five levels of port intelligence, starting from fully manual, and ending with smart, and fully, all-weather, automated port operations.

Presently, some smart ports have already incorporated automatic container terminals, equipped with smart sensing systems, such as RFID, for localization, and cameras that use image classifying ML applications. The goal is to integrate a greater level of IoT systems where needed, and reliable wireless networking and Internet connectivity, which has been an obstacle (Yang et al., 2018). Another obstacle has been the information management, due to the huge increase of the gathered data from the aforementioned sources, which was investigated by Heilig & Voß (2017), who proposed a framework for information management and categorization. These multiple diverse data sources, that ports need to manage, also need an IoT integration platform, with Bracke et al. (2021) having proposed Obelisk. It focuses on the three following main capabilities:

- interoperability between wireless technologies, data sources and formats
- multi-tenancy enabling, securely, multiple stakeholder usage
- scalability, maximizing performance and stability, irrespective of data volume

The latest iteration of smart ports is characterised by five main pillars: smart services (e.g., vessel and container management), integrated modern technologies, shipping companies cluster management, development of smart port hubs (to achieve greater collaboration between port clusters), and sustainable technology (environmentally friendlier operations) (Yau et al., 2020).

The focus on environmentally friendly, energy efficient solutions, highlights the contemporary approach to this new opportunity. New energy efficient algorithms are being developed for smart ports (Ozturk et al., 2018), and the application of micro-grids is being studied. These offer higher energy savings, lower energy dependency, lower emissions, and better downtime management (Molavi et al., 2020).

The main applications of IoT, in order of use, are the quay crane (container loading/unloading), yard trailer (autonomous container transportation throughout the port), rubber tire gantry crane (container stacking), and the terminal operating system (the "brain" of the operation, controlling everything) (Yau et al., 2020). Another interesting application is the automatic mooring ("parking") of the vessels, by the automatic detection of vacant spots, from the placed sensors, and the subsequent data transfer to the vessels waiting to moor (Kamolov and Park, 2019). Those applications can be coupled with the container IoT systems, that have already been developed (Salah et al., 2020).

Due to its value and implementation complexity, many academics are focusing on how IoT can be integrated in their local ports, such as Croatia (Jovic et al., 2019), Le Havre (Belfkih et al., 2017), Hamburg (where space limitations demand the IoT enablement of efficient operations) (Hanschke et al., 2016), Spain (González et al., 2020; González-Cancelas et al., 2020), Tunisia (Bessid et al., 2020), China (Yao and Yuan, 2018), and the regions covered by the Indian Ocean (Sakhuja, 2019), and the North and Baltic Seas (Karas, 2020). The value is also evident by the fact that the Covid-19 pandemic, which took the world by storm (and the industry itself, as explained above), has not managed to genuinely decelerate the transition to smarter ports (Majid, 2021).

Moreover, the geographical distribution of the smart ports worldwide was investigated by Molavi et al. (2019), who proposed a Smart Port Index (SPI), which was based on four main pillars: smart operations, environmental performance, energy, and safety and security; with their respective KPIs. Their analysis concluded that the three smartest ports (from those studied) are Hamburg's, Rotterdam's, and Antwerp's (which explains the other conclusion of Europe being the continent with the most advanced smart ports), followed by Busan's, Singapore's, and Los Angeles' port.

As discussed in the cybersecurity section of this report, the integration of digital, online tools and technologies (IoT-related in this case) has arisen the demanding issue of securing the data and systems of the smart ports. Since the sector has lagged behind in the digitalization, it is only now realising and acting on this issue (both the industry and academia). That is due to the attacks that have been reported lately and the realisation of deficiencies in critical infrastructures. At the forefront of the securing endeavour is the development of methodologies for attack risk analysis, effects mitigation, protection and recovery plan designing, while policymakers need to intensify their efforts in the development of cybersecurity policies (de la Peña Zarzuelo, 2021). At the beginning of this focus on port cybersecurity, a framework was proposed, to assess data security in smart ports, from a privacy management perspective. The main focus was the security and privacy requirements on different levels and domains of the proposed framework (Heilig and Voß, 2016).

3.1.3 Weather Data

As mentioned in the "Challenges" chapter, weather data are extremely important for shipping companies, because they operate in the world's oceans, which are extremely affected by the weather. From typical dimensions, such as wind and temperature, to extreme weather conditions, such as hurricanes. Weather data collection is not a novel data source, like IoT, but it still offers a great way for the shipping sector to monitor the vessels, in order to increase their safety, and, also, build prediction algorithms, which need weather data to adequately predict the vessels' behaviour.

There are many ways to collect weather data, but the most common way to collect multiple data points is by using Application Programming Interfaces (APIs). An example of weather data collection can be found in the weather routing system that was developed by Cui et al. (2016). In it they collected the necessary weather forecast files from the European Centre for Medium-Range Weather Forecasts (ECMWF), for the waves and wind data, and from the National Oceanic and Atmospheric Administration (NOAA), for the currents. The data contained time, latitude, and longitude, which are the three basic dimensions for weather data collection (basically they provide a point in the four basic dimensions, since the height is always at the surface level).

The data can either be historical (past weather), live (present weather), or forecasted (future weather). They can be acquired for free, with more detailed weather data being usually costly, but are necessary when areas that are influenced by smaller scale geographic features (such as islands and narrow ocean currents) are being analysed (Rødseth et al., 2016).

3.2 Data Engineering

The sheer volume of data (commonly known as big data) require proper processes to ensure that the data are clean, reliable, and stored properly, efficiently, securely, safely, and can be instantly accessible. This process is necessary in various stages of the data lifecycle. For example, from the moment a sensor sends data to a receiver, the moment the receiver aggregates or stores the data, the moment the vessel's or port's database receives the data from the receiver, to the moment the shipowner or shipping line receives the data from the port, the vessel, the API, etc. This chapter will focus on the handling of the data once they reach a more central part of the data lifecycle, for example, once the data reaches the main database of the shipowner.

The need for proper shipping data engineering can also be explained through the use of the 5 Vs of Big Maritime Data:

- volume (ever increasing, reaching petabytes of total data)
- velocity (through real-time streaming data; e.g., from sensors, weather stations, transmitters)
- variety (multiple heterogenous sources; e.g., historical, streaming, forecasts)
- veracity (varying degrees of accuracy, inconsistency, uncertainty, and ambiguity)
- value (can be utilized and exploited in multiple, diverse shipping data-driven applications) (Lytra et al., 2017)

To combat the five challenges, the authors proposed a Big Maritime Data Architecture, which is based on a data lake (containing all the diverse data sources). The data are passed through an integration pipeline, containing data curation, data integration, and semantic enrichment, in order to be in the correct format, to be used by the query processing engine. The system is also equipped with access controls, security mechanisms, privacy policies, and analytics and visualization platforms.

The master database of the stored data can use state-of-the-art technologies, such as distributed file systems (e.g., Hadoop), NoSQL key-value stores (e.g., Redis), NoSQL document stores (e.g., MongoDB), NoSQL wide-columnar stores (e.g., Cassandra), and Spatiotemporal systems (e.g., PostGIS) (Herodotou et al., 2021). Its creation can be based on the multiple research papers that aim to create shipping databases, with clearly defined steps, that can be used by shipowners as a structural pathway in their data engineering steps. One example is developed by Etienne et al. (2021), who provide a step-by-step guide for the development of a shipping database, for analytical traffic and vessel behavior purposes, with PostgreSQL and PostGIS. Similarly, the datAcron is a project that aims to assist in the shipping data architecture, management, and integration (Vouros et al., 2018).

Big data management and engineering is always a difficult undertaking for any company- or sector in general- that is not used to handling a vast assortment of data, as is evident by the above. There are multiple challenges that the industry is facing in this regard. Concerning data quality, the frequent lack of it is due to the multiple errors found in the data, attributed to either sensor faults or human errors- willful or not. It can be addressed either during the entry (with automated data entry, sensor fault detection, and validity or integrity checks), or after it, during the context-aware data cleaning phase (using anomaly and conflicting values detection, or incomplete data filling). On the other hand, data quantity results in two main challenges, data transfer (as analysed in the 5G chapter) and data storage, with the capacity requirements constantly increasing, and standardisation of data becoming a necessity (Mirović et al., 2018).

To reduce storage requirements, the database engineers can utilise statistical methods, which remove some of the data noise, but keep the main components and variation, such as Principal Component Analysis (PCA) (Perera and Mo, 2017b). Thus, the data pre-processing steps can be divided into data curation & cleaning (probing for missing, noisy, or inconsistent data), data integration (consolidation, federation, and propagation), data transformation (normalization, aggregation, generalization, and enhancement), and data reduction, similarly to the example above (Herodotou et al., 2021). The previous steps are suitable -and probably necessary- not just from the shipowner's perspective, but also from the vessel's (Perera & Mo, 2016).

In addition to the above, the challenges include data governance and security, and the lack of data specialists in the field is hindering the effective tackling of the challenges that are being faced (Mirović et al., 2018). Also, the complexity of the measured phenomena (such as the environmental impact, and the varying measurements of water, wind, and wave speed around the ship, depending on the placement of the sensor), the cost of interfaces, lack of standards, and limited ownership of the underlying data (e.g., limited access to data related to vessel machinery monitoring) (Rødseth et al., 2016).

Lastly, besides data infrastructure and data processes themselves, an additional factor that shipping companies should address is the integration of their data infrastructure with the systems from their business partners. Every member of the sector has invaluable data, whose value can only be increased through sharing with other members (e.g., ports' data sharing with shipowners and vice versa), creating close, transparent, and highly-collaborative business relationships, and increasing the sector's situational awareness (Lind et al., 2021). This requires the development of certain standards and protocols, with the TradeLens platform being an example.

3.3 Artificial Intelligence Applications

The collection of shipping data, as was described above, might have many applications, but the pinnacle of those, the quintessence of data analysis, is the use of artificial intelligence (usually machine learning) applications. The sector can utilise the multiple available algorithms to build models that have been trained on the rich variety and volume of shipping data being collected. These will have extreme added value for businesses, not just by automating operations, but, also, for assisting in its business and operation processes and overall efficiency. Artificial intelligence can be the solution to the mounting regulatory requirements (such as environmental laws), and market pressure. That is why it is gaining ground rapidly, expanding the efficiency boundaries of shipping companies, changing decision-making procedures, and challenging their traditional management strategies (Yuen et al., 2020). This chapter will also include quantitative statistical models, since their value is still quite high and relevant, and should not be overlooked.

3.3.1 Predictive Maintenance

Vessels are costly equipment, not just when it comes to their initial purchase, but also their operating costs, mostly through their pricey maintenance procedures, both for the onboard equipment and the machinery. As a result, multiple companies delay the maintenance, which can have significant consequences for the safety, environmental impact, and operational stability of the vessel, resulting in business losses, fines (due to the increasing scrutiny by the regulatory bodies) and, possibly, deaths. Inversely, proper maintenance leads to increased efficient and effective operations, efficient fuel consumption, and regulatory compliance. Thus, better maintenance strategies should be followed, with AI presenting a great tool in this regard (Michala et al., 2015).

While the level of the predictive maintenance applications is limited to predicting failures that are not unanticipated or inexplicable (which- of course- does not mean that they will not improve in this area), the added safety and intelligence that such an application presents to the traditional human-centered monitoring and risk analysis approach is substantial (Jimenez et al., 2020). Due to the aforementioned, the vessel's maintenance planning is an area that has extensive bibliography, combining the fields of mechanical engineering and data science.

The first step of building this application is the data collection and pre-processing, focusing on structural and machinery measurements. That is done by the use of accelerometers, hull monitoring systems, machinery sensor measurements, such as pressure, vibration, deflections clearances, and temperature (Raptodimos et al., 2016). It has been proved that the temperature measurement can also be carried out effectively with the use of infrared thermography (Molenda and Charchalis, 2019). Additionally, the use of vibration data can prove to be quite useful but simultaneously quite complex. It might contain both time-related and frequency-related features (measurements), with high multicollinearity between them and, thus, dimensionality reduction is often used with PCA (Gkerekos et al., 2016).

It is apparent that this application mostly uses the vessel's IoT systems as a data source, but the data sources that can be used in this application are more diverse than initially though. An effective and more advanced out-of-the-box example is the use of text analytics on dock indent documents (possibly taken from a platform such as TradeLens), to analyse associations and patterns of parts and equipment, whose maintenance is commonly carried out simultaneously or consecutively (Hwang et al., 2021). After the data are collected, there are multiple possible implementations to reach a predictive maintenance system.

The following papers are based on the Inspection Capabilities for Enhanced Ship Safety (INCASS) project, which aims to enhance vessel's safety. That is done by implementing a Decision Support Systems (DSS) based on sensors' data collection, to inspect and evaluate the maintenance needs of a vessel's machinery and equipment, based on a risk analysis (Lazakis et al., 2016). Firstly, together with modern AI algorithms, a condition-based maintenance approach can be used, which can couple a Dynamic Fault Tree Analysis (DFTA) with data clustering, to assess machinery conditions, which can be passed through Fuzzy Logic (FL) risk indices, to conclude the proper action to take (Cheliotis & Lazakis, 2018). Another novel tool, for the INCASS project, incorporates the Condition Based Maintenance and Decision Support Systems, to develop a maintenance tool that offers a detailed overview of the suggested maintenance actions, with a cost-benefit analysis, allowing greater efficiency and control of the vessel (Michala et al., 2015). The vibration data- and similar performance data- can be used as input in anomaly detection and classifier algorithms, such as the Support Vector Machine (SVM), through the use of PCA, as mentioned above (Gkerekos et al., 2017, 2016).

The PCA-SVM combination has also been used in the context of ship yard ballast pumps, in a binary classification model (prediction about whether maintenance is needed). It can be adequately accurate without the use of complex sensor data, which allows the shipowners to assess the application before a sizeable investment in IoT is made (Kimera and Nangolo, 2020). Another approach is to use regression models, such as OLS single linear regression, multiple linear ridge regression, OLS single polynomial regression and multiple polynomial ridge regression with the exponentially weighted moving average for fault detection. By scoring the implementations of the models mentioned, with R^2 and k-fold cross-validation, it is concluded that an impressive 0,96 R^2 is feasible, with the use of the multiple polynomial ridge regression, and the main engine's power, pressure, and speed as input (Cheliotis et al., 2020).

Additional models have been implemented, offering great results. For example, the use of multilayer-perceptron artificial neural networks, using fuzzy logic-based failure mode and effect criticality analysis as criticality assessment (offering an error of less than 5%) (Zaman et al., 2021). Additionally, examples include tree-based models implementing bagging, random forest, and boosting methods (which are not only accurate, but also easily interpretable) (Soner et al., 2018b).

The process of predictive maintenance can be divided into four key areas (Jimenez et al., 2020). These are:

- failure modes identification (failure events and causes association)
- likely consequence of failure (predicting failure risk, which is challenging, since the machinery's operation is usually stopped before a possible failure, and thus the predictive models cannot fully understand the scenarios)
- asset criticality (which is related to "teaching" the algorithm to not only predict failure, but also the business effects of it, thus providing intelligent risk/gain analyses)
- corrective action proposal

3.3.2 Freight Value Forecasting

This chapter transitions from the aforementioned mechanical engineering application, into a financial oriented one. The business model of the shipping industry is based on freight movement, which is its ultimate essence and core income stream. Since artificial intelligence can provide significant financial forecasts, moving into an AI-centred financial sector will prove to be invaluable for shipowners, and the sector in general.

Freight rates are in the centre of the interplay and business relations of the respective members of the shipping industry. Their size and direction constitute the "shipping risk", and are highly elastic, being affected by global freight movement demand and supply. Their forecast, risk analysis, and efficient management is the key to optimal income streams (for the shipowners, who wish to find the timing of the maximum possible freight rate), and freight timing and assignment (for the other parties, who wish to do the opposite). This can either be short-term, with the correct timing of a contract, or long-term, with the correct planning of vessel purchase and shipbuilding (capacity increase), or vessel maintenance and scrapping (capacity decrease) (Schramm and Munim, 2021). An effective planning of it can be based on the cycle of the container market, that every shipowner should be acquainted with (Jeon et al., 2020).

The industry has developed plenty of indices that measure the shipping trends and behaviours, which have even become important in the assessment of the economic stability in general. They can be divided into maritime, economic, environmental, and miscellaneous indices, with the four main ones being the Baltic Dry Index, the Energy Efficiency Design Index, the Baltic Capesize Index, and the New ConTex Index (Karamperidis et al., 2013). The most prominent one, the Baltic Dry Index (a quantitative maritime index) is an index of the average freight rate of dry bulk materials, across certain routes. It is considered a principal indicator of economic health, because changes in it reflect supply and demand fluctuations for important materials used in manufacturing (Kopp, 2021). The influence in the relationship is bidirectional, with the commodities markets having a spillover effect in the shipping markets (Sun et al., 2020), and, inversely, the index having a spillover effect on the financial markets (mostly during periods of crisis), and acting as a short-term indicator of them (Lin et al., 2019).

Firstly, a two-tier cross-validation and back-testing procedure has been used to predict the freight rates of China's containerized freight index, with both the system dynamics modelling and conventional time-series procedures. The former was used to adequately model the intricate nonlinear nature of freight rates, and the latter as comparative benchmarking (Jeon et al., 2021). The shipowners, from a financial perspective, need not only focus on the future potential freight rates, but also on the forecasted volatility of them. A reliable model has been developed for this application, combining time-series and machine learning models (as is common in these applications) and offers additional valuable insight into the leverage effect (which is a risk/gain financial term) that is faced (Jiaguo Liu et al., 2021).

As mentioned above, the Baltic Dry Index (BDI) is the most frequently used freight rate index and, thus, it is extensively researched. Firstly, an accurate forecasting model for it is the one developed by Katris & Kavussanos (2021), which uses time-series and machine learning methods (support vector regression, feed-forward fully connected artificial neural networks, and multivariate adaptive regression splines). By using the autoregressive integrated moving average (including some variations of it) and Diebold–Mariano tests, for model assessment, they conclude that the two model approaches have analogous performance, but can be combined for optimum accuracy.

The use of ensembled (combined) models seems to be a great practice for the BDI's forecasting, according to J. Li et al. (2021). The authors propose an ensemble forecasting model with multi-objective programming (integrating the Pareto optimization theory and multi-objective particle swarm optimization), which focuses not only on the minimization of the forecasting error but, also, in the maximization of the model's diversity. Additionally, a multivariate vector autoregressive model with exogenous variables has been developed for the BDI, with its performance being evaluated against a benchmark model, which uses a univariate ARIMA framework, and proven to be superior (Tsioumas et al., 2017). Furthermore, Makridakis et al. (2020) proposed another BDI forecasting model, which uses Bayesian compressed regression and adds lags to the index data, with its key differentiator being the transition functions, that capture changes in the data's time path, by embedding them in the random search procedure of the algorithm.

The revenue is not only affected by the freight rate, but also from the probability of the booking cancellation of the container slot. In addition, this forecast can allow shipping companies to overbook their capacity, by a factor of $\frac{total\ slots}{1 - average\ probability\ of\ cancellation} - safety\ margin$, to increase revenue, while ensuring the needed availability. While predicting the freight rate might be more important than the above probability, it is worth investigating it, to achieve optimum and complete forecasting capabilities. Zhao et al. (2020) propose an implementation for this novel research topic, of a data-driven model, based on a time-to-event modelling technique. Additionally, it includes the quantitative assessment of the internal influential factors of slot booking, in conjunction with the external factors of booking cancellation, while also incorporating a frailty term that takes the regionality into account.

Moreover, in machine learning, field experience has to be taken into account at all times and, as such, it has been proved that the sentiment, judgement, confidence, and apprehension measurements of industry experts greatly improve the forecasting models (Schramm and Munim, 2021). In addition, this application is not only useful during the industry's normal operations, but can also be used in times of crises (like the COVID-19 pandemic), to predict their impact in the industry (Koyuncu et al., 2021).

Shipowners are not only directly assisted in their forecasts and planning (through the aforementioned models), but also indirectly, through the employment of such models in their affiliated ports. The communication of the result of those forecasts to the shipowners is of immeasurable value in determining port loads and bottlenecks, therefore minimizing the waiting time until docking (e.g., if long waiting times are forecasted, the vessel can decrease speed to reduce consumption and emissions, and its remaining itinerary could be similarly automatically adjusted). In a more long-term approach, a forecasting model can inform about the cargo throughput trends that a port will experience, which can assist the ports in capacity planning and development, and the shipowners in determining the ports that will shape the future of the industry.

The short-term application is related to vessel flows' (inflow/outflow) forecasting. A relevant proposed deep learning model integrates a bi-directional long short-term memory (BDLSTM; highly used in similar flow analysis applications) and a convolutional neural network (CNN; highly used in spatiotemporal analyses). It achieves a forecasting accuracy of 77.5% to 80%, which is satisfactory given its constraints (Zhou et al., 2020). A similar application was developed by EY, predicting vessel arrival 14 days in advance, the needed resources for the receival of the incoming vessels, and potential bottlenecks ("choke points"). The application improved workforce planning (making ground operations more predictable), and increased efficiency and planning accuracy by 3%, resulting in US$10.2 million in savings (How Predicting When Ships Arrive Boosts Efficiency for an Entire Port, EY Global).

Dragan et al. (2021) proposed a model for the aforementioned long-term application, that forecasts ports' cargo throughput, with a Dynamic Factor Analysis-ARIMAX (DFA-ARIMAX) modelling approach. External economic factors are feeding information to the DFA (through the use of a PCA, multiple linear regression, and Monte Carlo pipeline), which in turn provides the dynamic factors to the ARIMAX model.

3.3.3 Energy efficiency analysis

The previous chapters analysed the models that facilitate the forecast of the revenue, and the optimisation of the- sizeable- cost of maintenance. This chapter will present the analysis methods used in the fuel consumption of a vessel, which can be divided into cost optimisation and emissions minimization. More specifically, until recently, fuel consumption was only viewed as an expensive and necessary operating cost. Lately, with the decarbonisation efforts due to global warming (as explained in the respective chapter) the industry has shifted more weight in its effort of reducing fuel consumption and, as a result, emissions (as it should). It is worth noting that fuel consumption forecasting will also be the subject of the case study. This application has expanded to include the estimated power consumption of electric vessels, which is of crucial importance in the quest for more environmentally-friendly and efficient vessels (called electric propulsion ships- EPS) (Lim et al., 2019).

Firstly, the majority of the relevant applications (and the case study) use input data derived from the noon reports (singularly or in conjunction with other data). Noon reports are data sheets that are commonly sent from the vessel's captain to the shipowner, usually at noon (hence the name). They contain information relative to the vessel's location, distance travelled, speed, propeller revolutions, cargo, and weather conditions (related to swell direction and height, currents, wind, and temperature). Noon reports are by no means robust data (since they are handled by a person not necessarily familiar with technology, they usually contain errors and missing values), presenting inherent uncertainty that needs to be taken into account when using it as an analytical data source (Aldous et al., 2013). This manual data collection should- and will- be obsolete in the future, since it is becoming increasingly automatic.

The proposed framework that will be presented first is a two-stage fuel consumption prediction and optimisation model, with an ultimate 2-7% in fuel savings. The fuel consumption prediction model is based on a random forest regression, which allowed the assessment of the features' importance, concluding that the sailing speed is the most important factor in fuel consumption. This information is used in the development of the latter part of the framework, the optimisation model, which aims to optimise vessel sailing speed, by using a mixed-integer linear programming model (R. Yan et al., 2020). Therefore, the previous model allows the integration of the cost of speed in route planning, balancing this cost with the added benefit of accelerated operations.

The weather conditions can have a great impact in overall fuel consumption. As discussed in the Data Sources chapter, weather information is invaluable for a shipowner, and this is a proof of this assumption. A proposed model estimates the applied resistance, based on weather data, thus calculating fuel consumption, which is used in conjunction with a route planning algorithm (using an isochrone method) to plan the most economical route, instead of the shorter one (Roh, 2013). A more computationally advanced version of the previous framework contains the use of a deep learning model, using similar inputs as the previous one, achieving an impressive 94% accuracy in its prediction, when provided with a high-quality dataset. The model is combined with an asymmetric travelling salesman problem algorithm, to conclude the optimum route (Bui-Duy and Vu-Thi-Minh, 2021). L. Yang et al. (2020) reiterate the importance of weather (e.g., currents) data on the calculation of fuel consumption. They propose an optimisation solution that uses the currents' data to differentiate between speed through water and speed over ground, reducing the overall error (stemming from the lack of this distinction) by approximately 70%.

The operating cost of fuel is based on two variables, volume (with applications concerning it presented above) and price. The price is related to the fuel bunkering policies that the shipping owners employ, depending on the vessel's itinerary and the fluctuation of fuel prices between ports. Due to this reason, a mixed integer non-linear programming model was developed, for cost minimisation, based on fuel consumption, local fuel prices, and fuel bunkering policies. The model proposed a fuel bunkering strategy, based on the estimated vessel's fuel inventory and the local cost at the port, to estimate the needed volume of bunkered fuel and subsequent cost (De et al., 2021).

As mentioned above, the shift from noon reports into a more automatic and robust data acquisition method has been researched, such as the automated data logging and monitoring systems. It has been found, by the use of advanced machine learning models, that the latter method can increase accuracy by 7% and reduce data collection time by a staggering 90%. Of the used models, the random forest regressor and extra trees regressor models were found to have the best performance, after the ensemble models, which seem to be the ultimate benchmark (Gkerekos et al., 2019b).

Concerning emissions minimisation, a study was developed, offering a comparative analysis of popular methods, combining the energy consumption, with the maintenance conditions of the vessel (offering a glimpse into how the discussed applications can be combined), to calculate the total emissions. It was concluded that the ship traffic emission assessment model (STEAM) was the best method for energy consumption estimation, while Goldsworthy has proposed the best method for emission estimation (Moreno-Gutiérrez et al., 2015). Therefore, it is apparent that all of the discussed applications and data sources are interconnected. In this case besides noon reports, weather data, maintenance data (and relevant estimations), and onboard systems, AIS data can be used, to form a complete dataset, upon which a machine learning method can be employed (Safaei et al., 2019b).

Furthermore, a research paper presents the shift that global shipping patterns might experience, to accommodate environmentally friendlier routes. The trans-Arctic routes seem as a feasible option to change and reduce the pollution patterns, due to the decreasing ice (as a result of the global warming) making the routes more accessible (which is an overall ironic situation). With an average of 26% fuel reduction and 24% emission's reduction, in specific routes, the shipping companies should take note of similar alternative routing scenarios (Z. Wang et al., 2020).

4. Case Study

The previous chapters explained in detail how the data can be utilized by a shipping line, in order to maximize the value that is extracted from them. In order to showcase the above, the value they produce, and the ease with which those applications can be developed, this thesis includes a case study chapter. This chapter presents a report of the code that was developed for the purpose of this thesis, which revolved around the phases presented above, in detail.

More specifically, the case study is based on a noon report dataset, that was kindly provided by the author's internship supervisor (XXXXX XXXXX) and XXXX. The dataset originates from real noon reports, but is masked for confidentiality purposes. Its scope and size are adequate for the analytical purposes of this thesis, but usually shipping lines have access to more detailed data, with more features (data columns) and records.

Namely, the features that the initial dataset contains are related to the noon report itself (starting and ending dates), the voyage (location and speed), and the vessel (name, average speed, deadweight tonnage, propellers' total revolutions, main engine's working hours, and total consumption), for every vessel and studying period (record).

The structure of the case study revolves around the structure of the data lifecycle chapter. Initially, the dataset is explored, to comprehend its contents, and cleaned, which- as analyzed above- is a necessary step to ensure that the analysis will be based on correct, robust data. Consecutively, the data are enhanced, with various calculated features, and with the use of a weather API, that offers historical marine data. However, the use of this API is not extensively used for the whole dataset, since multiple API calls are only available for the premium version. Similarly, the use of an API that calculates the sea distance between two points is not implemented, since they do not offer free versions.

It should be noted that this chapter is a report of the key features of the current version of the case study, providing a brief summary. The extensive analysis, together with the respective code and thorough documentation, is hosted on the author's Github page[1], and can be accessed through NBViewer[2], which is able to render the Jupyter Notebooks.

4.1 Dataset Exploration and Processing

The first stage of analytics is exploring the dataset in order to understand the basic characteristics and principles of it. At the same time, the dataset is transformed into one that will be more suited for the analysis that will follow, and enhanced by the use of both calculated (stemming from the existing data) and external (provided by other sources) data. Throughout this procedure, it is ensured that the correct data-cleaning processes are utilized, certifying that the analysis is based upon robust, clean, and reliable data.

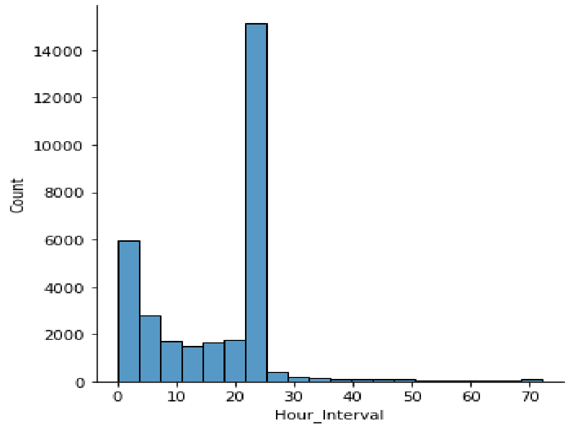


Figure 1: Hour Interval Distribution

Firstly, theoretically, the noon reports concern a 24-hour period, but that is not necessarily the case for this dataset. There are multiple records with an hour interval of less than 24 hours, and some with more. Thus, we need to calculate a common basis; this is the conversion of all the numerical data into an hourly basis (e.g., consumption per hour).

With further exploration, the notebook is able to provide basic dataset statistics. The dataset contains 31,951 total records, out of which 10,007 (31.3%) are Nan (missing/null) consumption data. It contains 68 vessels, covering a total distance of 2,877,391 miles, with a total of 543,280 voyage hours studied. The modest number of vessels (and the high number of Nan values) might reduce the predicting (and generalizing) capabilities of the models, possibly due to overfitting the small dataset.

When exploring the dataset, it is frequently useful to plot the highest and lowest values, for the numerical columns, and monitor their validity. Upon doing that, it was noticed that the consumption data contain a value that is most probably erroneous, and needs to be removed. Since the removal requires the checking of every numerical value of the consumption's column, the Nan consumption (and vessel speed, due to their correlation with consumption) data are simultaneously eliminated, which will further help the next phases of the case study. The rest

[1] https://github.com/nantoniou/Fuel-Consumption-Forecasting-Noon-Reports
[2] https://nbviewer.jupyter.org/github/nantoniou/Fuel-Consumption-Forecasting-Noon-Reports/blob/main/Noon_Report_Analysis.ipynb?flush_cache=true

of the numerical values are similarly checked for outliers, and the aforementioned outlier is examined, in order to understand its source.

Since the dataset contains multiple numerical values that seem correlated, it is worth analyzing the numerical connection between these features (with distribution plots) with the studied feature, namely the fuel consumption.

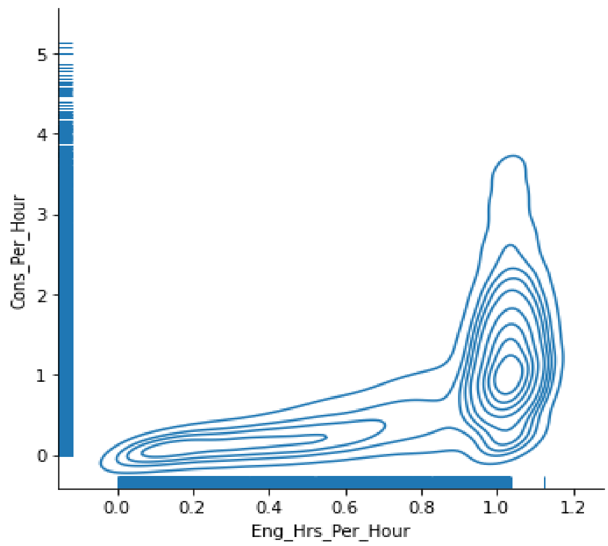


Figure 2: Correlation between Engine Working Hours and Consumption

The first instance of this analysis includes the distributional correlation of the main engine's working hours and the engine's consumption. From the plot above, it can be deduced that the idling vessels have low consumption, which increases slightly, and linearly, when the engine's working hours increase. Additionally, it seems that the majority of the engine's working hours' records lie close to one, with the correlation of those two variables becoming obscure at this point, proving that there are many other factors that influence the consumption. Lastly, it can be noticed that the engine's working hours, per hour, might surpass one, which is probably due to erroneous data (since the engine cannot operate more than one hour per hour).

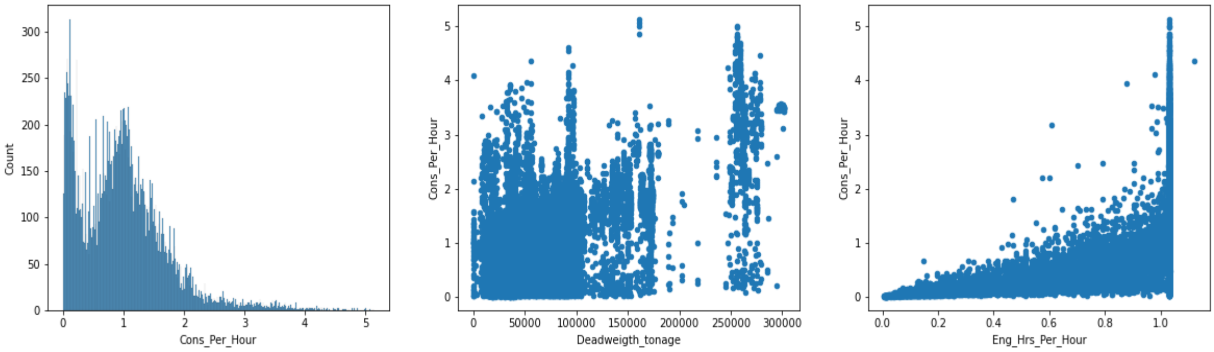


Figure 3: Other Features' Correlation with the Consumption

Delving deeper into the consumption correlation plots, the understanding of the consumption distribution, and the correlation of the consumption with the engine's working hours (in a different format) and vessel's deadweight, is improved. The first deduction of the above plots is that the distribution of the consumption data is mostly centred around one, with a sizeable spike around zero. This indicates that the consumption data can be split into the noon reports that concern mostly idle vessels (zero spike), and the ones that concern moving vessels (distribution around one). The data related to idle vessels will be removed later, since it does not assist the pursuit of an effective consumption forecasting model. Secondly, the deadweight varies, but some vertical data columns (batches) can be seen, proving further that the dataset contains vessels of various types, being most distinct in the 250,000 to 300,000 marks. Lastly, the top values of the consumption increase linearly, in relation to the engine's working hours, per hour, until they reach the point with the highest number of occurrences, when the latter is around one.

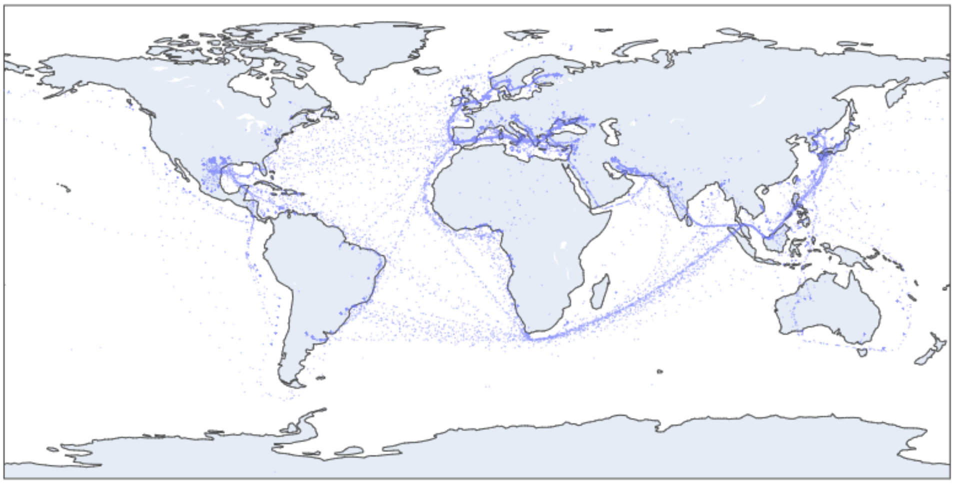
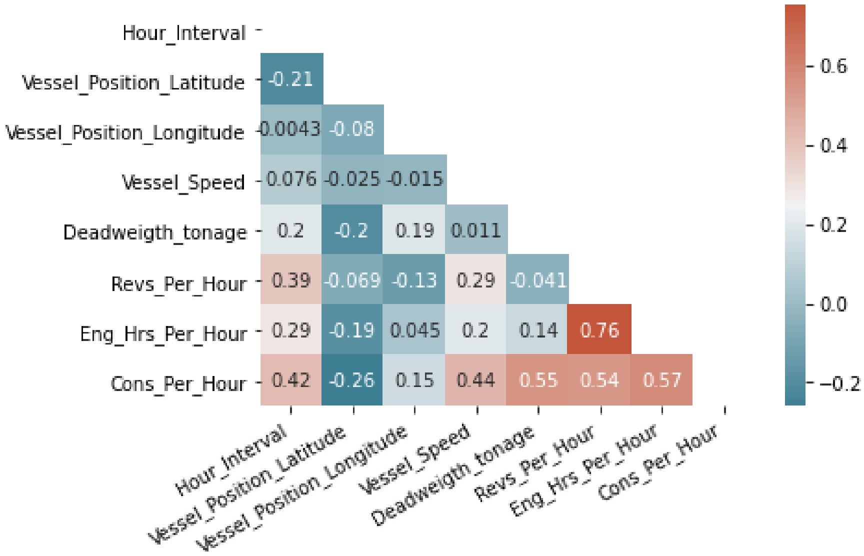


Figure 4: Vessel Positions

Furthermore, since the noon reports describe vessel itineraries, it is worth actually visualizing the positions of the vessels on a map. Every dot is an individual noon report, and the formation of various trading routes is apparent, like the one connecting eastern Asia with the southern point of Africa. This constitutes the first part of the routes towards Europe (not passing through the Suez Canal) and America. Most of the data seem to stem from the Mediterranean Sea. It should be noted that this visualization is interactive, with the end user having the ability to delve deeper in it; sadly, the application used for the rendering does not support it.

Figure 5: Correlations' Plot



Before the training of the models, the correlation of the features needs to be checked, to ensure that they are within the preferred limits, and visualize the patterns. The engine's working hours and the propeller revolutions have a high correlation, which is borderline acceptable, since the features are limited. This might become a hindrance in the later use of linear regression, and generally the algorithms that are seriously affected by highly correlated features.

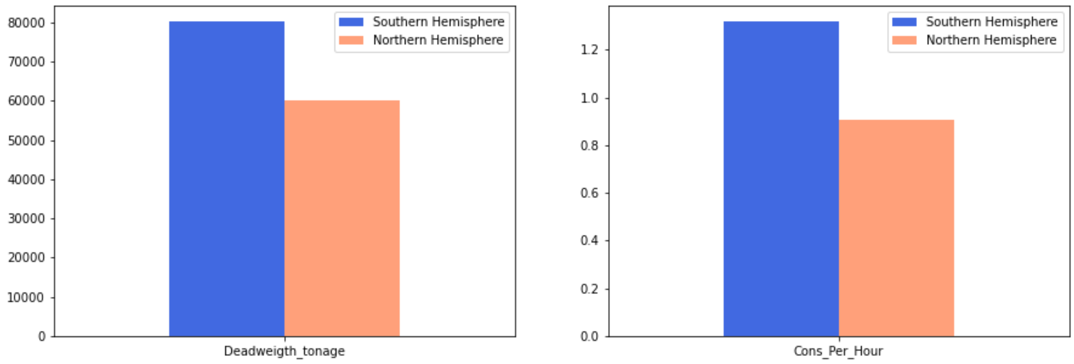


Figure 6: Hemispheres' Comparison

Another interesting analysis, originating from the results of the two previous plots, is the comparison between the northern and southern hemisphere voyages. It seems that there are some discrepancies between them, for certain columns. More specifically, it is quite apparent the deadweights and hourly consumption are higher in the southern hemisphere, probably due to the nature of the respective routes there (mostly long-haul, with heavier cargo).

Additionally, as mentioned above, the use of a weather API is implemented. The API used provides information about historic marine weather data, queried by the location (coordinates) and datetime. The result is returned in JSON format, containing information such as temperature, wind speed and direction, swell height and direction, and water temperature. It is provided by World Weather Online, which offers a limited free version, which is the one used in this case study. Since the version is free, the number of API calls (queries to the server) and the span of the historic weather dataset are limited. Therefore, it is used for completeness purposes, but is not used for the extensive dataset.

4.2 Consumption Forecasting

The value of the consumption forecasting is vast for every shipowner that aims to utilize the data and structure the decision-making process based on data-driven analyses. In the respective chapter, the value that this forecast produces was analyzed, along with the algorithms that are commonly used in this scenario. This might seem remarkably convoluted, requiring great expertise and countless hours of development and implementation, in order to achieve great results; or even just build a seemingly simple model.

Howbeit, in practice, a simple version of a forecasting model can be easily developed, provided that the developer is familiar with basic software engineering and data science processes. Surely, developing robust and exceptionally accurate models, on which decisions of strategical importance can be based, is (or, at least, should be) a product of cooperation by senior data scientists that fully understand the multiple aspects of machine learning algorithms. Nevertheless, in this chapter, multiple machine learning models will be developed to showcase how the aforementioned forecasting looks in practice.

4.2.1 Linear Regression

Currently, in the world of data science, the most advanced machine learning algorithms (such as neural networks) are gaining all the attention, becoming the field's epicentre. This is a trend that is somewhat explained by the sheer computing complexity and often accurate handling of really convoluted problems, such as natural language processing and autonomous driving (or navigating, in the case of vessels). However, the basis of data science, statistics, should not be overlooked. Based on this, this chapter will include linear regression algorithms, which offer a simple solution to regression (i.e., numerical value prediction, in this case consumption). Consecutively, an introduction to feature selection will be presented, showcasing how different features (data columns) can yield different predictors.

The models' development starts with the simple linear regression implementation, that is based on the four main predictors (for now subjectively) of the consumption value; namely vessel's speed, deadweight tonnage, engine's working hours, and revolutions per hour. The R^2 value is ~63%, with the p-values being zero, besides the engine's working hours, which is ~0,12. This means that the inclusion of it in the algorithm is not statistically significant (and can subsequently be omitted). Additionally, the Nan values were removed from the dataset.

During the development of predictive analytics applications, there are three main things that should be optimized:

- The data used for training, making sure that it is of high quality (clean, robust, complete, and reliable). If it is not, multiple data cleaning and data expansion processes should be implemented, as was done above.
- Hyperparameters, which are used to control the learning process and the overall behavior of the algorithm. The hyperparameter tuning will be showcased in the next subchapter.
- Lastly, feature engineering, which is the selection and handling of the multiple features that may or may not be inputted in the model. In the current application, the feature engineering is somewhat limited, due to the limited scope of the dataset.

Therefore, to increase the forecasting accuracy of the linear regression algorithm, the arbitrary selection of the features is changed into a dynamic, objective one. In order to select the features that yield the highest score, a method is implemented, which checks every feature combination, for every number of features. The total combinations that are fitted reach 128, with the best one having the location (both latitude and longitude), speed, deadweight, and the revolutions features. The score is slightly higher, at 64,4% R^2, with all p-values being nil.

4.2.2 Advanced Machine Learning

The score of the linear regression is adequate for analyses of lesser impact and importance. However, if there is a need for more advanced predictions, it seems that the use of a more complex algorithm is necessary. Indeed, using the XGBoost algorithm, which usually yields impressive results for similar applications, we can reach an exceptionally better score. Before analyzing the results, it is worth mentioning that- in contrast to linear regression- the dataset has to be split into training and testing data. This is done to ensure accurate scoring, which is not a result of overfitting (i.e., the incorporation of the specific dataset's noise into the predictions, improving the score in the specific dataset, but giving a wrong impression of its accuracy). Thus, the dataset was split into 80% training and 20% testing data. For computational complexity reasons, the features that will be used are the ones initially used in the linear regression. The R^2 score of the XGBoost algorithm is around 89%, offering an increase of 25%, with a mean squared error ~0,057.

As mentioned above, hyperparameter tuning is a great way to improve the performance of the models. In this instance, Sci-kit Learn's GridSearch library was utilized, which offers a great way to test the model's prediction capabilities with different hyperparameters, and automatically select the best combination. For example, in the case study, the hyperparameters tested were the colsample bytree, the alpha, the maximum depth, and the learning rate. Consecutively, to ensure that the selection is not based on overfitting, it offers cross-validation. Ultimately, a total of 32 fits were made, increasing the R^2 by approximately 1,3%, which is important when reaching scores greater than 90%. Similarly, the mean squared error was slightly decreased.

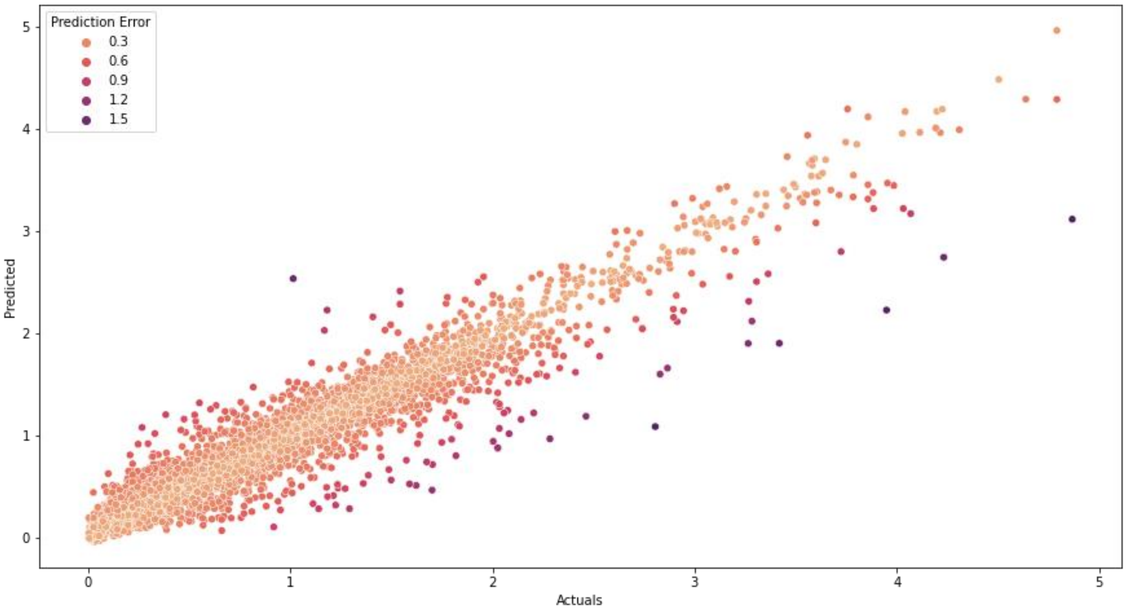


Figure 7: Actual and Predicted Consumption Comparison

In some cases, the numbers themselves do not fully capture the results of the algorithm. In order to better visualise the prediction accuracy, the testing dataset was used, and for every X value in it, a prediction was made, which was subsequently compared with the initial actual value. This visualization is shown above, with the predicted value on the Y axis, the actual value on the Y axis, and their respective difference appearing in a colour scale. The prediction accuracy is now evident, showcasing how the two dimensions are close aligned; when the actual is increased, the predicted is increased too (albeit having some outliers).

Further exploring the machine learning algorithms, multiple models were implemented, to ensure that the final chosen one is the best available. The approximate scores (R^2 and Mean Squared Error) of the *current version* of the developed models are as follows:

Table 1: Models' Scores

Algorithm	R^2	*MSE*
XGBoost Grid Search	91,6%	0.046
Ensemble Voting Regressor	91,4%	0.038
Gradient Boosting	91,2%	0.039
Random Forest	90,8%	0.040
XGBoost	90,5%	0.053
Decision Tree	83,6%	0.072
Linear Regression Best Model	64,4%	0.429
SGD	63,5%	0.160
SKLearn Linear Regression	63,3%	0.157
Linear Regression	63,3%	0.429
SVM	47,4%	0.231
Ada Boost	43,2%	0.249

The best model is the XGBoost model, that the grid search algorithm produced, followed by the ensemble algorithm, which was based on the top performing models. This showcases that, firstly, the hyperparameter tuning is very important to reach optimum performance and, secondly, that the use of ensemble models results in great predictors, as analysed in the respective chapter above. Ultimately, the models can be divided into three categories: the top performing ones (which reach great scores), the mediocre ones (Linear Regression and SGD, which reach adequate, but mostly lacking, predictions), and the inferior ones (SVM and Ada Boost, which trail behind, not being suited for this problem).

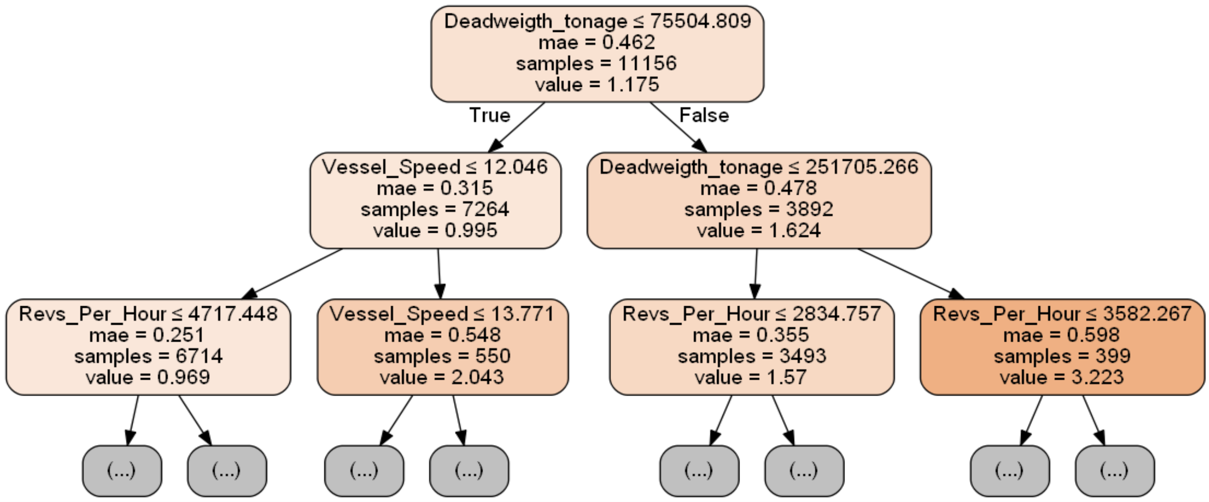


Figure 8: Decision Tree Visualisation

To provide some further insight into what is happening behind the scenes of the algorithms, this case study provides the figure above, which visualises the decision tree model. Generally, one of the reasons that decision trees are commonly used is their intelligible, interpretational "white-box" nature, stemming from their straightforward predictions' structure. This is apparent in the figure, where the main predictors, and their respective effect on the prediction, are clearly visible. The tree's root node can approximately be read as: Does a vessel have a deadweight of less than 75.505? If yes predict 1, if no predict 1,6. Consecutively, such calculations are constantly being made, reaching the last node in the tree (leaf), which produces the most accurate, final, prediction.

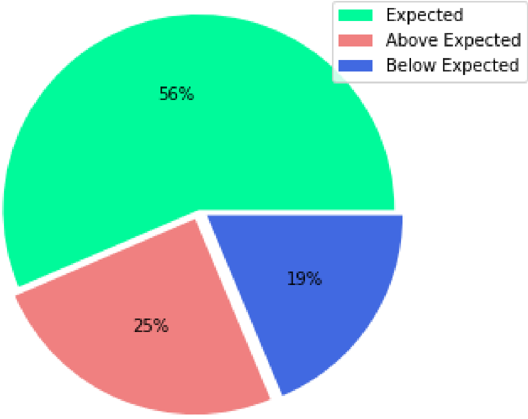


Figure 9: Distribution of Prediction Classes

Lastly, to provide insight into the predictions of the ensemble model, the predicted values were classified into three main categories: below expected (less than 95% of the actual value), above expected (more than 105% of the actual value), and expected (95-105% of the actual value). The distribution of the predicted values' classes appears in the figure above, where it can be observed that most of the predictions are very close to the actual values, followed by above expected, and then below expected.

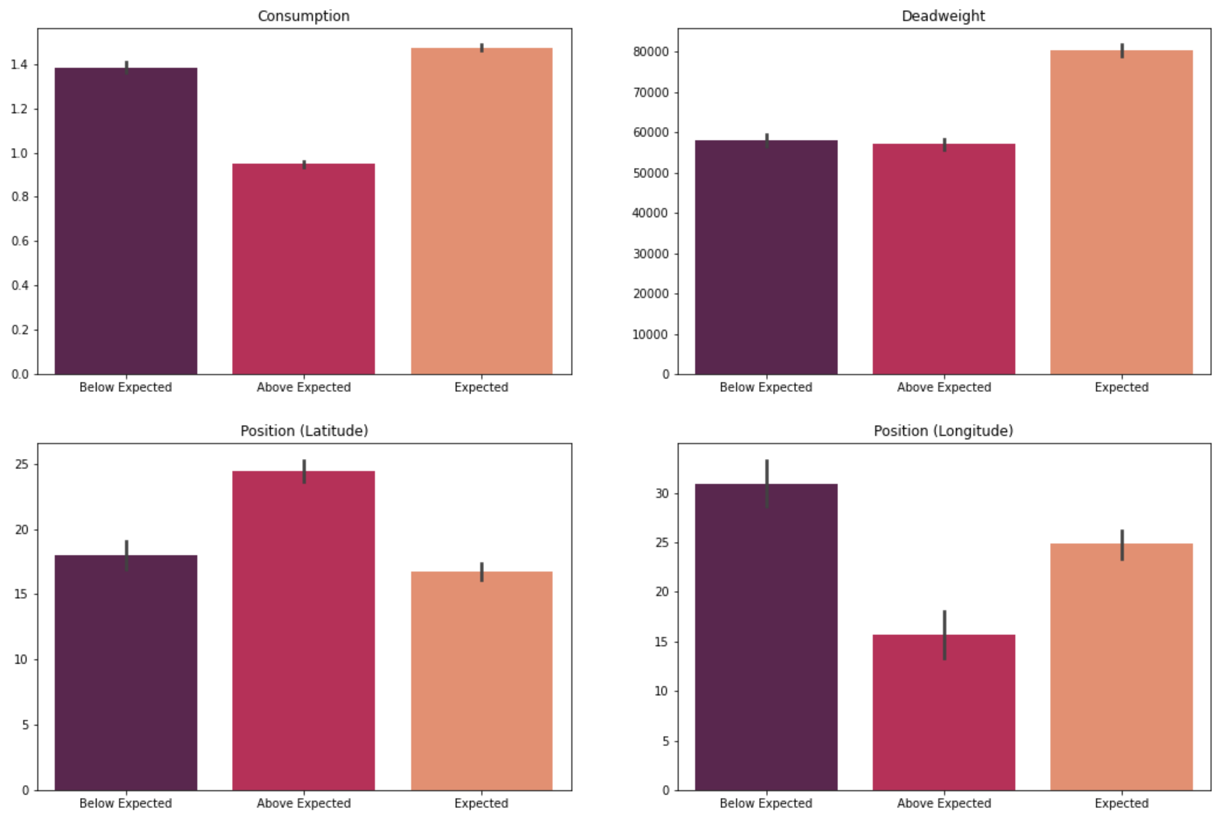


Figure 10: Classes Feature's Comparison

Delving deeper into the classes' differences, the figure above provides comparisons between them, concerning the consumption, deadweight, and vessel position. Firstly, it seems that the records with lower consumption have an even lower consumption than predicted. Secondly, vessels with a lower deadweight have both higher and lower values than predicted (thus possibly providing insight into the feature's deviational properties). Lastly, concerning location, the vessels with a higher latitude (thus located more North) are consuming more than expected and the ones with a higher longitude (more East) consume less than expected, in contrast with the vessels with a lower longitude, that consume more than expected.

Concluding, this chapter provided insight into the process of creating an analytical application based on noon reports. It showcased how the data can be explored (to understand the fleet's characteristics and charters better), cleaned (to increase the data quality), and then transformed into data suitable for further machine learning analyses. Additionally, it provided examples of how different algorithms perform, delving deeper into some basic machine learning processes, such as feature analysis, ensemble models, and hyperparameter tuning. Consecutively, various figures were presented, to give insights and meaning to the case study, not only providing visual analysis for the experienced data scientist, but also visual stimuli, for the business stakeholder to better understand the analysis and the ultimate predictions.

5. Discussion

Throughout this thesis, the goal was to highlight the importance of the use of data and data-related technologies in the shipping sector. The study of the use of those technologies in this specific sector is very valuable for a plethora of reasons. Firstly, it is a sector upon which the global trade is based, and any disruptions or improvements concerning it might turn out to have great effects in the movement of goods and the quality-of-life people enjoy. That is due to the fact that global trade has increased the quality and quantity of products that people around the globe can have access to.

Secondly, since the sector is operating with vast revenues and razor thin profit margins, every rise in its efficiency results in multiple billions in increased profits. Investing in new technologies is risky and sometimes very costly, especially before they reach maturity. However, in this instance, what the shipping sector ought to realise is that the slightest differences in operating efficiency have the potential to make a shipping line lead the market. Conversely, failure to keep up with technology will force it to ineffective and inefficient handling of the crises (and added regulating pressure), resulting in submission to the pressure.

Thirdly, it is useful to also look at this thesis from the perspective of the technologies that are being proposed and expected to be adopted from the shipping industry. Data and data-related technologies have been intensely studied in the last decades, and their value has constantly been proven, thus being hailed as an integral part of the 4th industrial revolution. The aforementioned research, while imperative in the quest to reach technological maturity, has not been adequately addressing the main hinderances regarding the studied technologies, which are adoption feasibility and espousal propensity.

The shipping sector, ever being traditional, laggard and bureaucratic, is the perfect case study for the assessment of the technological maturity of data applications. The sector's implementational and adoption reluctancy is the perfect illustration of the reactance that has been faced. Currently, it seems that the increasing data gathering onboard vessels is not utilised in the decision-making process, thus the value and knowledge contributions are limited (Aiello et al., 2020). However, the sector's first major steps in the adoption of state-of-the-art applications showcases that the core problems facing data technologies are- finally- being outshined by its incredible prospects.

And incredible they are. As this thesis presented above, the opportunities for the introduction of monitoring systems and artificial intelligence applications are endless. The former ones allow the shipowner to have a complete and accurate overview of the fleet, which is crucial when the business is based on the operation of such complex machinery, which roam the oceans. That is achieved through the integration of IoT sensors- and corresponding systems. The research showed that the most valuable AI applications can be divided into three categories. Firstly, predictive maintenance, which is crucial for the safety, high availability, and efficient vessel performance. Secondly, freight value forecasting, allowing more profitable charter structures, and more accurate financial forecasts. Thirdly, energy efficiency analysis, reducing the fuel cost and the emissions, which are crucial in the current highly- and increasingly-regulated market.

Nonetheless, the current unfavourable circumstances in the sector have actually had considerable effects on the adoption momentum. The financial adversity that the industry has been experiencing has increased the chasm between the technological adoption of the bigger players, compared to the smaller ones. The increased funds and technological capabilities (and "know-how") of the former allow them to partake in significant advancements. Simultaneously, the latter are restricted to waiting for the investment cost to drop, and for the technologies to mature, thus lagging behind and increasing the market share and efficiency variances (Lambrou and Ota, 2017).

Concluding the reasons why the adoption is lagging, Joshi (2021) divides them into two main categories: costs (hardware, upgradation, training, and operational), and hesitancy. The latter involves the shipowners' caution against seemingly black-box (lacking explanation) decisions, which most AI solutions produce and cannot be compared to the current transparent decision-making processes. Furthermore, it involves apprehension about the extent to which AI should be allowed to take over operations (human accidents are easier to explain and relate to, than AI's), and trust issues concerning data sources.

6. Conclusions

Overall, this thesis analysed how the sector is moving ahead into what it perceives as unchartered territory, aiming to combat multiple crises with technological advancements, increasing efficiency and streamlining the decision-making process. In that way, its struggling, vying members will be able to adjust in the modern era of razor thin profit margins and achieve a higher market share. In this endeavour, it has faced multiple problems, some being a result of its own operating culture. However, they are starting to subside, making room for true innovation. Relevant innovation comes in the form of data-related technologies and data.

The most important aspects of the data-related technologies were identified and divided into four categories. First, the use of state-of-the-art data communication methods, which will allow the increasing collected data to move quickly and without interruptions; a difficult undertaking for a sector operating in the middle of the ocean. Second, the transferred data is invaluable for the business it belongs to and, thus, its security is of utmost importance, which is achieved through the use of modern cybersecurity protocols and practices. Third, the automation of the sector, which can be further subdivided into autonomous vessels (a milestone into safer operations and increased efficiency and predictability), and robotics (extremely useful in a sector operating with incredibly heavy loads). Last, the use of blockchain is expected to revolutionise the way certain data (e.g., contracts) are transferred, and streamline the current dallying bureaucracy.

The data were divided by the use of the COSA data lifecycle framework, in order to add structure to the convoluted data processes. Firstly, the applications need valuable data, which is found in the form of Internet of Things (coming from the vessels' sensors and systems, and the ports' sensors and status monitoring), weather (past, present, and future), and Automatic Identification System's data. These are later passed through data engineering pipelines, assuring reliability, coherence, consistency, and quality. Through similar processes they are then entered into a database to be readily accessible. Lastly, the aforementioned data are used to build artificial intelligence applications, mainly in the areas of predictive maintenance, freight rate forecasting, and energy efficiency analysis.

Lastly, this thesis provided a specific example of how such an application is developed, and the value it produces. It is apparent, that the data sources are critical in AI applications, and the data engineering part should not be overlooked, since it provides the foundation upon which the subsequent model will be developed. Additionally, the value of simple data analysis and visualisation tools was showcased, to allow the reader to realise that the AI applications might be the endgame of data analysis, but basic analyses should not be overlooked. Furthermore, it provided insight into the development of machine learning models, from basic linear regressions to complex ensemble models, and presented basic machine learning techniques, such as feature selection and hyperparameter tuning.

7. Future Research

Through the preparation of this thesis, it was found that there are specific research areas, concerning the new technologies being proposed by academia, that seem to be lacking the bibliographic volume and quality that is expected (and needed). Specifically, it was discovered that the relevant bibliography has a palpable lack of specific monetary estimations ("hard numbers") about the effects of such an adoption.

That has to be studied further, because the shipping sector operates completely differently from the tech industry, in which every new technology is immediately implemented and thoroughly studied, with its value still being questionable (due to the incredibly nimble investment strategies). The decisions in shipping are made by executives, who are- as a rule- wary of risky and unproven investment strategies.

Their conservative hesitancy about new data solutions is also described by the World Maritime News (2019), by analysing the reasons why shipowners are hesitant about autonomous vessels (which are thoroughly researched and supported by academia). This is mainly due to the fact that they trust their crews more than AI, which has led the companies, who aim to offer relative AI solutions, to approach it in a more hybrid way, incorporating the crew's input in the end product. Concluding, it can become apparent that most of the relevant literature describes the technological feasibility (and respective advancements) of data. Therefore, additional research is needed, with related work focused on shipping digital transformation management, such as the ones conducted by Lambrou et al. (2019), and Maydanova et al. (2020).

References

Acemoglu, D., Restrepo, P., 2019. Artificial Intelligence, Automation, and Work, in: Agrawal, A., Gans, J., Goldfarb, A. (Eds.), The Economics of Artificial Intelligence. University of Chicago Press, pp. 197–236. https://doi.org/10.7208/9780226613475-010

Agarwala, N., Guduru, S.S., 2021. The potential of 5G in commercial shipping. Maritime Technology and Research 3. https://doi.org/10.33175/mtr.2021.248995

Agiwal, M., Roy, A., Saxena, N., 2016. Next generation 5G wireless networks: A comprehensive survey. IEEE Communications Surveys and Tutorials 18, 1617–1655. https://doi.org/10.1109/COMST.2016.2532458

Aiello, G., Giallanza, A., Mascarella, G., 2020. Towards shipping 4.0. A preliminary gap analysis, in: Procedia Manufacturing. Elsevier B.V., pp. 24–29. https://doi.org/10.1016/j.promfg.2020.02.019

Alcaide, J.I., Llave, R.G., 2020. Critical infrastructures cybersecurity and the maritime sector, in: Transportation Research Procedia. Elsevier B.V., pp. 547–554. https://doi.org/10.1016/j.trpro.2020.03.058

Aldous, L., Smith, T., Bucknall, R., 2013. Noon report Data Uncertainty, in: Proceedings of Low Carbon Shipping Conference. Low Carbon Shipping, London, UK.

Alop, A., 2019. The main challenges and barriers to the successful "smart shipping." TransNav 13, 521–528. https://doi.org/10.12716/1001.13.03.05

Asariotis, W.J.R., Assaf, M., Ayala, G., Ayoub, A., Benamara, H., Chantrel, D., Hoffmann, J., Larouche-Maltais, A., Premti, A., Rodríguez, L., Sun, S., Youssef, F., Abbas, H., Bradford, G., Cariou, P., Crowe, T., Davidson, N., Manuel, J., Orejas, D., Dominioni, G., Ducrest, J., Faghfouri, M., Haag, F., Ingebrigtsen, M., Kontou, E., Lane, A., Lind, M., Alcover, A.M., Milne, J., Mooney, T., Notteboom, T., Pallis, A.A., Sánchez, R., Stevenson, A., Stratidakis, S., Teodoro, A., Verhoeven, P., 2020. Review of Maritime Transport 2020, in: United Nations Conference on Trade and Development. Geneva.

Atzori, L., Iera, A., Morabito, G., 2010. The Internet of Things: A survey. Computer Networks 54, 2787–2805. https://doi.org/10.1016/j.comnet.2010.05.010

Balduzzi, M., Pasta, A., Wilhoit, K., 2014. A security evaluation of AIS automated identification system, in: Proceedings of the 30th Annual Computer Security Applications Conference on - ACSAC '14. ACM Press, New York, New York, USA. https://doi.org/10.1145/2664243.2664257

Batalden, B.-M., Leikanger, P., Wide, P., 2017. Towards autonomous maritime operations, in: 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA). IEEE. https://doi.org/10.1109/CIVEMSA.2017.7995339

Bavassano, G., Ferrari, C., Tei, A., 2020. Blockchain: How shipping industry is dealing with the ultimate technological leap. Research in Transportation Business and Management 34, 100428. https://doi.org/10.1016/j.rtbm.2020.100428

Belfkih, A., Duvallet, C., Sadeg, B., 2017. The Internet of Things for smart ports: Application to the port of Le Havre, in: International Conference on Intelligent Platform for Smart Port (IPaSPort 2017).

Bessid, S., Zouari, A., Frikha, A., Benabdelhafid, A., 2020. Smart Ports Design Features Analysis: A Systematic Literature Review, in: 13ème CONFERENCE INTERNATIONALE DE MODELISATION, OPTIMISATION ET SIMULATION (MOSIM2020). Maroc.

Borgia, E., 2014. The internet of things vision: Key features, applications and open issues. Computer Communications. https://doi.org/10.1016/j.comcom.2014.09.008

Bracke, V., Sebrechts, M., Moons, B., Hoebeke, J., de Turck, F., Volckaert, B., 2021. Design and evaluation of a scalable Internet of Things backend for smart ports. https://doi.org/10.1002/spe.2973

Bui-Duy, L., Vu-Thi-Minh, N., 2021. Utilization of a deep learning-based fuel consumption model in choosing a liner shipping route for container ships in Asia. Asian Journal of Shipping and Logistics 37, 1–11. https://doi.org/10.1016/j.ajsl.2020.04.003

Caschili, S., Medda, F.R., 2012. A review of the maritime container shipping industry as a complex adaptive system. Interdisciplinary Description of Complex Systems: INDECS 10, 1–15.

Cheliotis, M., Lazakis, I., Theotokatos, G., 2020. Machine learning and data-driven fault detection for ship systems operations. Ocean Engineering 216. https://doi.org/10.1016/j.oceaneng.2020.107968

Cheliotis, M.F., Lazakis, I., 2018. Ship machinery fuzzy condition based maintenance, in: Proceedings of the 2018 Smart Ship Technology Conference. Royal Institution of Naval Architects.

Crosby, M., Pattanayak, P., Verma, S., Kalyanaraman, V., 2016. Blockchain technology: Beyond bitcoin. Applied Innovation 2, 71.

Cui, T., Turan, O., Boulougouris, E., 2016. Development of a ship weather routing system for energy efficient shipping, in: Annual Conference of the International Association of Maritime Economists 2016. pp. 1–14.

De, A., Choudhary, A., Turkay, M., Tiwari, M.K., 2021. Bunkering policies for a fuel bunker management problem for liner shipping networks. European Journal of Operational Research 289, 927–939. https://doi.org/10.1016/j.ejor.2019.07.044

de la Peña Zarzuelo, I., 2021. Cybersecurity in ports and maritime industry: Reasons for raising awareness on this issue. Transport Policy 100, 1–4. https://doi.org/10.1016/j.tranpol.2020.10.001

de Osés, F.X.M., Castells, Marcel. la, 2008. Heavy Weather in European Short Sea Shipping: Its Influence on Selected Routes. Journal of Navigation 61. https://doi.org/10.1017/S0373463307004468

de Vos, J., Hekkenberg, R.G., Valdez Banda, O.A., 2021. The Impact of Autonomous Ships on Safety at Sea – A Statistical Analysis. Reliability Engineering and System Safety 210, 107558. https://doi.org/10.1016/j.ress.2021.107558

Decarbonising Shipping: ALL HANDS ON DECK, 2020.

Dragan, D., Keshavarzsaleh, A., Intihar, M., Popović, V., Kramberger, T., 2021. Throughput forecasting of different types of cargo in the adriatic seaport Koper. Maritime Policy & Management 48. https://doi.org/10.1080/03088839.2020.1748242

Etienne, L., Ray, C., Camossi, E., Iphar, C., 2021. Maritime Data Processing in Relational Databases, in: Artikis, A., Zissis, D. (Eds.), Guide to Maritime Informatics. Springer, Cham. https://doi.org/10.1007/978-3-030-61852-0_3

Gajewska, M., 2019. Maritime Communication and Sea Safety of the Future — Machine - Type 5G Communication Concept. Scientific Journal of Polish Naval Academy 218. https://doi.org/10.2478/sjpna-2019-0018

Gascard, J.C., Riemann-Campe, K., Gerdes, R., Schyberg, H., Randriamampianina, R., Karcher, M., Zhang, J., Rafizadeh, M., 2017. Future sea ice conditions and weather forecasts in the Arctic: Implications for Arctic shipping. Ambio 46, 355–367. https://doi.org/10.1007/s13280-017-0951-5

Gkerekos, C., Lazakis, I., Theotokatos, G., 2019a. Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study. Ocean Engineering 188, 106282. https://doi.org/10.1016/j.oceaneng.2019.106282

Gkerekos, C., Lazakis, I., Theotokatos, G., 2019b. Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study. Ocean Engineering 188. https://doi.org/10.1016/j.oceaneng.2019.106282

Gkerekos, C., Lazakis, I., Theotokatos, G., 2017. Ship machinery condition monitoring using performance data through supervised learning, in: Smart Ship Technology Conference. London, UK.

Gkerekos, C., Lazakis, I., Theotokatos, G., 2016. Ship machinery condition monitoring using vibration data through supervised learning, in: Proceedings of The International Conference of Maritime Safety and Operations (MSO). Glasgow, UK.

González, A.R., González-Cancelas, N., Serrano, B.M., Orive, A.C., 2020. Smart ports: ranking of Spanish port system. World Scientific News 144, 1–12.

González-Cancelas, N., Molina Serrano, B., Soler-Flores, F., Camarero-Orive, A., 2020. Using the SWOT Methodology to Know the Scope of the Digitalization of the Spanish Ports. Logistics 4, 20. https://doi.org/10.3390/logistics4030020

Guan, D., Wang, D., Hallegatte, S., Davis, S.J., Huo, J., Li, S., Bai, Y., Lei, T., Xue, Q., Coffman, D.M., Cheng, D., Chen, P., Liang, X., Xu, B., Lu, X., Wang, S., Hubacek, K., Gong, P., 2020. Global supply-chain effects of COVID-19 control measures. Nature Human Behaviour 4, 577–587. https://doi.org/10.1038/s41562-020-0896-8

Hannaford, E., Hassel, E. van, 2021. Risks and Benefits of Crew Reduction and/or Removal with Increased Automation on the Ship Operator: A Licensed Deck Officer's Perspective. Applied Sciences 11. https://doi.org/10.3390/app11083569

Hanschke, L., Heitmann, J., Renner, C., 2016. Challenges of WiFi-Enabled and Solar-Powered Sensors for Smart Ports, in: Proceedings of the 4th International Workshop on Energy Harvesting and Energy-Neutral Sensing Systems. ACM, New York, NY, USA. https://doi.org/10.1145/2996884.2996887

Hasselaar, T.W.F., 2011. An investigation into the development of an advanced ship performance monitoring and analysis system.

Heering, D., 2020. Ensuring Cybersecurity in Shipping: Reference to Estonian Shipowners. TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation 14. https://doi.org/10.12716/1001.14.02.01

Heikkilä, E., Wahlström, M., Granholm, G., 2020. Towards maritime traffic coordination in the era of intelligent ships: a systems theoretic study, in: Proceedings of the International Seminar on Safety and Security of Autonomous Vessels (ISSAV) and European STAMP Workshop and Conference (ESWC) 2019. Sciendo. https://doi.org/10.2478/9788395669606-020

Heilig, L., Voß, S., 2017. Port-Centric Information Management in Smart Ports: A framework and categorisation, in: Ports and Networks. Routledge, pp. 236–250.

Heilig, L., Voß, S., 2016. A holistic framework for security and privacy management in cloud-based smart ports, in: 15th International Conference on Computer and IT Applications in the Maritime Industries-COMPIT '16, Lecce Italy.

Help Net Security, 2013. Digital ship pirates: Researchers crack vessel tracking system - Help Net Security [WWW Document]. URL https://www.helpnetsecurity.com/2013/10/16/digital-ship-pirates-researchers-crack-vessel-tracking-system/ (accessed 5.21.21).

Herodotou, H., Aslam, S., Holm, H., Theodossiou, S., 2021. Big Maritime Data Management. https://doi.org/10.1007/978-3-030-50892-0_19

How predicting when ships arrive boosts efficiency for an entire port [WWW Document], n.d. . EY Global. URL https://www.ey.com/en_gl/better-begins-with-you/how-predicting-when-ships-arrive-boosts-efficiency-for-an-entire-port (accessed 6.14.21).

Hwang, T., Youn, I.-H., Oh, J., 2021. Study on Text Analysis of the Liquefied Natural Gas Carriers Dock Specification for Development of the Ship Predictive Maintenance Model. Journal of the Korean Society of Marine Environment and Safety 27. https://doi.org/10.7837/kosomes.2021.27.1.060

IALA Guideline No. 1028 On The Automatic Identification System (AIS), 2004.

Icaza, L., Marzo, S., Popa, T., Sahbaz, U., Saravelos, G., 2009. The Greek shipping cluster. Harvard Business School 7, 1–31.

Jabbar, K., Bjørn, P., 2018. Infrastructural grind: Introducing blockchain technology in the shipping domain, in: Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work. Association for Computing Machinery, pp. 297–308. https://doi.org/10.1145/3148330.3148345

Jacob, B., Porter Jr, J., Khemraj, R., Lasecki, D., Miller Jr, E., 2019. What can we learn from early adopters?, in: Emerging Information Technologies II. Pace University.

Jensen, T., Hedman, J., Henningsson, S., 2019. How TradeLens Delivers Business Value With Blockchain Technology. MIS Quarterly Executive 18. https://doi.org/10.17705/2msqe.00018

Jeon, J.-W., Duru, O., Munim, Z.H., Saeed, N., 2021. System Dynamics in the Predictive Analytics of Container Freight Rates. Transportation Science. https://doi.org/10.1287/trsc.2021.1046

Jeon, J.-W., Duru, O., Yeo, G.-T., 2020. Modelling cyclic container freight index using system dynamics. Maritime Policy & Management 47. https://doi.org/10.1080/03088839.2019.1708984

Jie, Y., Kersing, A., Saxon, S., Xie, Q., 2020. Data will decide success in the next normal of bulk and tanker shipping.

Jimenez, V.J., Bouhmala, N., Gausdal, A.H., 2020. Developing a predictive maintenance model for vessel machinery. Journal of Ocean Engineering and Science 5, 358–386. https://doi.org/10.1016/j.joes.2020.03.003

Jing, M., Zheng, W., 2020. Research and Application of Real-time Ship Traffic Intelligent Analysis and Calculation Platform under 5G Background. IOP Conf. Ser.: Earth Environ. Sci 446, 52093. https://doi.org/10.1088/1755-1315/446/5/052093

Joshi, N., 2021. Why AI adoption is lagging in international shipping [WWW Document]. Allerin. URL https://www.allerin.com/blog/why-ai-adoption-is-lagging-in-international-shipping (accessed 6.17.21).

Jovic, M., Kavran, N., Aksentijevic, S., Tijan, E., 2019. The transition of Croatian seaports into smart ports, in: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019 - Proceedings. Institute of Electrical and Electronics Engineers Inc., Opatija, Croatia, pp. 1386–1390. https://doi.org/10.23919/MIPRO.2019.8757111

Jugović, A., Bukša, J., Dragoslavić, A., Sopta, D., 2019. The Possibilities of Applying Blockchain Technology in Shipping. Pomorstvo 33. https://doi.org/10.31217/p.33.2.19

Kamolov, A., Park, S.H., 2019. An IoT Based Smart Berthing (Parking) System for Vessels and Ports, in: Kim, K.J., Kim, H. (Eds.), Mobile and Wireless Technology 2018. Springer, Singapore, pp. 129–139. https://doi.org/10.1007/978-981-13-1059-1_13

Karamperidis, S., Jackson, E., Mangan, J., 2013. The use of indices in the maritime transport sector. Maritime Policy & Management 40. https://doi.org/10.1080/03088839.2013.776182

Karas, A., 2020. Smart Port as a Key to the Future Development of Modern Ports. TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation 14. https://doi.org/10.12716/1001.14.01.01

Katris, C., Kavussanos, M.G., 2021. Time series forecasting methods for the Baltic dry index. Journal of Forecasting. https://doi.org/10.1002/for.2780

Kim, M., Joung, T.-H., Jeong, B., Park, H.-S., 2020. Autonomous shipping and its impact on regulations, technologies, and industries. Journal of International Maritime Safety, Environmental Affairs, and Shipping 4. https://doi.org/10.1080/25725084.2020.1779427

Kimera, D., Nangolo, F.N., 2020. Improving ship yard ballast pumps' operations: A PCA approach to predictive maintenance. Maritime Transport Research 1, 100003. https://doi.org/10.1016/j.martra.2020.100003

Komianos, A., 2018. The Autonomous Shipping Era. Operational, Regulatory, and Quality Challenges. TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation 12. https://doi.org/10.12716/1001.12.02.15

Kopp, C., 2021. Baltic Dry Index (BDI) [WWW Document]. Investopedia. URL https://www.investopedia.com/terms/b/baltic_dry_index.asp (accessed 6.3.21).

Koyuncu, K., Tavacioğlu, L., Gökmen, N., Arican, U.Ç., 2021. Forecasting COVID-19 impact on RWI/ISL container throughput index by using SARIMA models. Maritime Policy & Management. https://doi.org/10.1080/03088839.2021.1876937

Lademan, D., 2021. Container ship charter rates continue upward momentum as tonnage markets run dry [WWW Document]. S&P Global Platts. URL https://www.spglobal.com/platts/en/market-insights/latest-news/shipping/042321-container-ship-charter-rates-continue-upward-momentum-as-tonnage-markets-run-dry (accessed 6.25.21).

Lagouvardou, S., Psaraftis, H.N., Zis, T., 2020. A literature survey on market-based measures for the decarbonization of shipping. Sustainability (Switzerland) 12. https://doi.org/10.3390/SU12103953

Lambrou, M., Ota, M., 2017. Shipping 4.0: Technology Stack and Digital Innovation Challenges, in: Proceedings of IAME 2017 Conference. Kyoto, Japan, pp. 1–20.

Lambrou, M., Watanabe, D., Iida, J., 2019. Shipping digitalization management: conceptualization, typology and antecedents. Journal of Shipping and Trade 4. https://doi.org/10.1186/s41072-019-0052-7

Lazakis, I., Dikis, K., Michala, A.L., Theotokatos, G., 2016. Advanced Ship Systems Condition Monitoring for Enhanced Inspection, Maintenance and Decision Making in Ship Operations, in: Transportation Research Procedia. Elsevier B.V., pp. 1679–1688. https://doi.org/10.1016/j.trpro.2016.05.133

Lee, C.-U., Kim, D.-Y., Yu, Y.-H., Shin, O.-K., 2009. Development of Embedded Vessel Monitoring System Using NMEA2000. Journal of the Korean Society of Marine Engineering 33. https://doi.org/10.5916/jkosme.2009.33.5.746

Li, J., Hao, J., Feng, Q.Q., Sun, X., Liu, M., 2021. Optimal selection of heterogeneous ensemble strategies of time series forecasting with multi-objective programming. Expert Systems with Applications 166. https://doi.org/10.1016/j.eswa.2020.114091

Li, S., Ma, Z., Han, P., Zhao, S., Guo, P., Dai, H., 2018. Bring Intelligence to Ports Based on Internet of Things. https://doi.org/10.1007/978-3-030-00018-9_12

Li, X., Feng, W., Wang, J., Chen, Y., Ge, N., Wang, C.-X., 2020. Enabling 5G on the Ocean: A Hybrid Satellite-UAV-Terrestrial Network Solution. IEEE Wireless Communications 27. https://doi.org/10.1109/MWC.001.2000076

Lim, C. og, Park, B. cheol, Lee, J. chul, Kim, E.S., Shin, S. chul, 2019. Electric power consumption predictive modeling of an electric propulsion ship considering the marine environment. International Journal of Naval Architecture and Ocean Engineering 11, 765–781. https://doi.org/10.1016/j.ijnaoe.2019.02.011

Lin, A.J., Chang, H.Y., Hsiao, J.L., 2019. Does the Baltic Dry Index drive volatility spillovers in the commodities, currency, or stock markets? Transportation Research Part E: Logistics and Transportation Review 127, 265–283. https://doi.org/10.1016/j.tre.2019.05.013

Lind, M., Ward, R., Jensen, H.H., Chua, C.P., Simha, A., Karlsson, J., Göthberg, L., Penttinen, T., Theodosiou, D.P., 2021. The Future of Shipping: Collaboration Through Digital Data Sharing, in: Lind, M., Michaelides, M., Ward, R., Watson, R.T. (Eds.), Maritime Informatics. Springer, Cham. https://doi.org/10.1007/978-3-030-50892-0_9

Liu, Jiahui, Duru, O., Law, A.W.K., 2021. Assessment of atmospheric pollutant emissions with maritime energy strategies using bayesian simulations and time series forecasting. Environmental Pollution 270, 116068. https://doi.org/10.1016/j.envpol.2020.116068

Liu, Jiaguo, Li, Z., Sun, H., Yu, L., Gao, W., 2021. Volatility forecasting for the shipping market indexes: an AR-SVR-GARCH approach. Maritime Policy & Management. https://doi.org/10.1080/03088839.2021.1898689

Liu, R.W., Nie, J., Garg, S., Xiong, Z., Zhang, Y., Hossain, M.S., 2021. Data-Driven Trajectory Quality Improvement for Promoting Intelligent Vessel Traffic Services in 6G-Enabled Maritime IoT Systems. IEEE Internet of Things Journal 8. https://doi.org/10.1109/JIOT.2020.3028743

Loklindt, C., Moeller, M.-P., Kinra, A., 2018. How Blockchain Could Be Implemented for Exchanging Documentation in the Shipping Industry. https://doi.org/10.1007/978-3-319-74225-0_27

Louw-Reimer, J., Nielsen, J.L.M., Bjørn-Andersen, N., Kouwenhoven, N., 2021. Boosting the Effectiveness of Containerised Supply Chains: A Case Study of TradeLens, in: Lind, M., Michaelides, M., Ward, R., Watson, R.T. (Eds.), Maritime Informatics. Springer, Cham. https://doi.org/10.1007/978-3-030-72785-7_6

Lu, R., Turan, O., Boulougouris, E., 2013. VOYAGE OPTIMISATION: PREDICTION OF SHIP SPECIFIC FUEL CONSUMPTION FOR ENERGY EFFICIENT SHIPPING.

Lytra, I., Vidal, M.E., Orlandi, F., Attard, J., 2017. A big data architecture for managing oceans of data and maritime applications, in: 2017 International Conference on Engineering, Technology and Innovation: Engineering, Technology and Innovation Management Beyond 2020: New Challenges, New Approaches, ICE/ITMC 2017 - Proceedings. Institute of Electrical and Electronics Engineers Inc., Madeira, Portugal, pp. 1216–1226. https://doi.org/10.1109/ICE.2017.8280019

Majid, M.A., 2021. DEVELOPMENT OF SMART PORTS AMID THE COVID-19 PANDEMIC. The Impact of COVID-19 Pandemic on Malaysia’s Maritime Sectors and Way Forward: MIMA Issue Paper 12.

Makridakis, S., Merikas, A., Merika, A., Tsionas, M.G., Izzeldin, M., 2020. A novel forecasting model for the Baltic dry index utilizing optimal squeezing. Journal of Forecasting 39. https://doi.org/10.1002/for.2613

Mallouppas, G., Yfantis, E.Ar., 2021. Decarbonization in Shipping Industry: A Review of Research, Technology Development, and Innovation Proposals. Journal of Marine Science and Engineering 9, 415. https://doi.org/10.3390/jmse9040415

MARITIME 2050: Navigating the Future, 2019. . London, UK.

MARITIME TECHNOLOGY CHALLENGES 2030, 2017.

Maydanova, S., Ilin, I., Jahn, C., Weigell, J., 2020. Global Container Shipping Line Digital Transformation and Enterprise Architecture Modelling. IOP Conference Series: Materials Science and Engineering 940. https://doi.org/10.1088/1757-899X/940/1/012004

Michail, N.A., Melas, K.D., 2020. Shipping markets in turmoil: An analysis of the Covid-19 outbreak and its implications. Transportation Research Interdisciplinary Perspectives 7, 100178. https://doi.org/10.1016/j.trip.2020.100178

Michala, A.L., Lazakis, I., Theotokatos, G., Varelas, T., 2015. Predictive maintenance decision support system for enhanced energy efficiency of ship machinery, in: Proceedings of the International Conference on Shipping in Changing Climates. Glasgow, United Kingdom, pp. 195–205.

Miola, A., Ciuffo, B., 2011. Estimating air emissions from ships: Meta-analysis of modelling approaches and available data sources. Atmospheric Environment 45, 2242–2251. https://doi.org/10.1016/j.atmosenv.2011.01.046

Mirović, M., Miličević, M., Obradović, I., 2018. Big Data in the Maritime Industry. NAŠE MORE: znanstveni časopis za more i pomorstvo 65. https://doi.org/10.17818/NM/2018/1.8

Mojtahedzadeh, R., Bouguerra, A., Schaffernicht, E., Lilienthal, A.J., 2015. Support relation analysis and decision making for safe robotic manipulation tasks. Robotics and Autonomous Systems 71, 99–117. https://doi.org/10.1016/j.robot.2014.12.014

Molavi, A., Lim, G.J., Race, B., Shi, J., 2019. Smart Ports: the Future of Maritime Transportation, in: Texas Hurricane Center for Innovative Technology. Houston.

Molavi, A., Shi, J., Wu, Y., Lim, G.J., 2020. Enabling smart ports through the integration of microgrids: A two-stage stochastic programming approach. Applied Energy 258, 114022. https://doi.org/10.1016/j.apenergy.2019.114022

Molenda, J., Charchalis, A., 2019. Preliminary Research of Possibility of Using Thermovision for Diagnosis and Predictive Maintenance of Marine Engines. Journal of KONBiN 49. https://doi.org/10.2478/jok-2019-0050

Moreno-Gutiérrez, J., Calderay, F., Saborido, N., Boile, M., Rodríguez Valero, R., Durán-Grados, V., 2015. Methodologies for estimating shipping emissions and energy consumption: A comparative analysis of current methods. Energy 86, 603–616. https://doi.org/10.1016/j.energy.2015.04.083

Muhammad, B., Kumar, A., Cianca, E., Lindgren, P., 2018. Improving Port Operations through the Application of Robotics and Automation within the Framework of Shipping 4.0, in: International Symposium on Wireless Personal Multimedia Communications, WPMC. IEEE Computer Society, pp. 387–392. https://doi.org/10.1109/WPMC.2018.8712998

Munim, Z.H., Dushenko, M., Jimenez, V.J., Shakil, M.H., Imset, M., 2020. Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. Maritime Policy & Management 47. https://doi.org/10.1080/03088839.2020.1788731

Muñoz, J.A., Pérez, R., 2017. Design of smart ships for the IoT, in: Proceedings of the Second International Conference on Internet of Things, Data and Cloud Computing. ACM, New York, NY, USA. https://doi.org/10.1145/3018896.3018930

Navarro-Ortiz, J., Romero-Diaz, P., Sendra, S., Ameigeiras, P., Ramos-Munoz, J.J., Lopez-Soler, J.M., 2020. A Survey on 5G Usage Scenarios and Traffic Models. IEEE Communications Surveys and Tutorials 22, 905–929. https://doi.org/10.1109/COMST.2020.2971781

Notteboom, T., Haralambides, H., 2020. Port management and governance in a post-COVID-19 era: quo vadis? Maritime Economics and Logistics. https://doi.org/10.1057/s41278-020-00162-7

Notteboom, T., Pallis, T., Rodrigue, J.P., 2021. Disruptions and resilience in global container shipping and ports: the COVID-19 pandemic versus the 2008–2009 financial crisis. Maritime Economics and Logistics 1–32. https://doi.org/10.1057/s41278-020-00180-5

Oneto, L., Anguita, D., Coraddu, A., Cleophas, T., Xepapa, K., 2016. Vessel monitoring and design in industry 4.0: A data driven perspective, in: 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow (RTSI). IEEE. https://doi.org/10.1109/RTSI.2016.7740594

Ozturk, M., Jaber, M., Imran, M.A., 2018. Energy-Aware Smart Connectivity for IoT Networks: Enabling Smart Ports. Wireless Communications and Mobile Computing. https://doi.org/10.1155/2018/5379326

Pallis, A.A., 2007. Maritime Transport: The Greek Paradigm. Elsevier.

Papathanasiou, A., Cole, R., Murray, P., 2020. The (non-)application of blockchain technology in the Greek shipping industry. European Management Journal 38, 927–938. https://doi.org/10.1016/j.emj.2020.04.007

Pape, M., 2020. Decarbonising maritime transport: The EU perspective.

Perera, L.P., Mo, B., 2017a. Marine Engine-Centered Data Analytics for Ship Performance Monitoring. Journal of Offshore Mechanics and Arctic Engineering 139. https://doi.org/10.1115/1.4034923

Perera, L.P., Mo, B., 2017b. Development of Data Analytics in Shipping. https://doi.org/10.4018/978-1-5225-2486-1.ch011

Perera, Lokukaluge P., Mo, B., 2016. Data Analytics for Capturing Marine Engine Operating Regions for Ship Performance Monitoring, in: Volume 7: Ocean Engineering. American Society of Mechanical Engineers. https://doi.org/10.1115/OMAE2016-54168

Perera, Lokukaluge Prasad, Mo, B., 2016. Machine Intelligence for Energy Efficient Ships: A Big Data Solution, in: Proceedings of the 3rd International Conference on Maritime Technology and Engineering (MARTECH 2016). Taylor and Francis, Lisbon, Portugal, pp. 143–150.

Perera, L.P., Soares, C.G., 2017. Weather routing and safe ship handling in the future of shipping. Ocean Engineering 130, 684–695. https://doi.org/10.1016/j.oceaneng.2016.09.007

Plale, B., Kouper, I., 2017. The Centrality of Data: Data Lifecycle and Data Pipelines, in: Data Analytics for Intelligent Transportation Systems. Elsevier Inc., pp. 91–111. https://doi.org/10.1016/B978-0-12-809715-1.00004-3

Porathe, T., Hoem, Å.S., Rødseth, Ø.J., Fjørtoft, K.E., Johnsen, S.O., 2018. Safety and Reliability – Safe Societies in a Changing World, in: Haugen, S., Barros, A., van Gulijk, C., Kongsvik, T., Vinnem, J.E. (Eds.), . CRC Press. https://doi.org/10.1201/9781351174664

Praetorius, G., Hult, C., Sandberg, C., 2020. Towards Autonomous Shipping – Exploring Potential Threats and Opportunities in Future Maritime Operations. https://doi.org/10.1007/978-3-030-20503-4_57

Prasad, P., Vatsal, V., Roy Chowdhury, R., 2021. Maritime Vessel Route Extraction and Automatic Information System (AIS) Spoofing Detection, in: 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). IEEE. https://doi.org/10.1109/ICAECT49130.2021.9392536

Psaraftis, H.N., Kontovas, C.A., 2020. Decarbonization of Maritime Transport: Is There Light at the End of the Tunnel? Sustainability 13, 237. https://doi.org/10.3390/su13010237

Psaraftis, H.N., Zis, T., Lagouvardou, S., 2021. A comparative evaluation of market based measures for shipping decarbonization. Maritime Transport Research 2, 100019. https://doi.org/10.1016/j.martra.2021.100019

Raptodimos, Y., Lazakis, I., Theotokatos, G., Varelas, T., Drikos, L., 2016. Ship sensors data collection and analysis for condition monitoring of ship structures and machinery systems, in: Smart Ship Technology.

RESOLUTION A.893(21), 1999.

Rex, C., Kristensen, N., Munch, J., Jensen, S., Wergeland, C., 2017. SHIPPING MARKET REVIEW May 2017.

Rex, C., Stenbjerg, J., Hjelmager, T., Hoffmann, J., Müllertz, S., 2020. SHIPPING MARKET REVIEW November 2020.

Rodrigue, J.-P., 2020. The Geography of Transport Systems, 5th ed. Routledge, New York.

Rødseth, Ø.J., Perera, L.P., Mo, B., 2016. Big data in shipping - Challenges and opportunities, in: 15th International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT 2016). Lecce, Italy.

Roh, M.-I., 2013. Determination of an economical shipping route considering the effects of sea state for lower fuel consumption. International Journal of Naval Architecture and Ocean Engineering 5, 246–262. https://doi.org/10.2478/ijnaoe-2013-0130

Rong, H., Teixeira, A.P., Guedes Soares, C., 2020. Data mining approach to shipping route characterization and anomaly detection based on AIS data. Ocean Engineering 198, 106936. https://doi.org/10.1016/j.oceaneng.2020.106936

Rukanova, B., Ubacht, J., van Engelenburg, S., Tan, Y.-H., Geurts, M., Sies, M., Molenhuis, M., Slegt, M., 2021. Realizing value from voluntary business-government information sharing through blockchain-enabled infrastructures: The case of importing tires to the Netherlands using TradeLens, in: Proceedings of 22nd Annual International Conference on Digital Government Research. Association for Computing Machinery (ACM), Omaha, United States.

Safaei, A.A., Ghassemi, H., Ghiasi, M., 2019a. VLCC's fuel consumption prediction modeling based on noon report and automatic identification system. Cogent Engineering 6. https://doi.org/10.1080/23311916.2019.1595292

Safaei, A.A., Ghassemi, H., Ghiasi, M., 2019b. VLCC's fuel consumption prediction modeling based on noon report and automatic identification system. Cogent Engineering 6. https://doi.org/10.1080/23311916.2019.1595292

Sakhuja, V., 2019. The Indian Ocean and Smart Ports 1. Indian Foreign Affairs Journal 14, 207–220.

Salah, K., Alfalasi, A., Alfalasi, M., Alharmoudi, M., Alzaabi, M., Alzyeodi, A., Ahmad, R.W., 2020. IoT-Enabled Shipping Container with Environmental Monitoring and Location Tracking, in: 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC). IEEE. https://doi.org/10.1109/CCNC46108.2020.9045495

Schramm, H.-J., Munim, Z.H., 2021. Container freight rate forecasting with improved accuracy by integrating soft facts from practitioners. Research in Transportation Business & Management. https://doi.org/10.1016/j.rtbm.2021.100662

Scott, T., 2018. TradeLens: how IBM and Maersk are sharing blockchain to build a global trade platform [WWW Document]. IBM.com. URL https://www.ibm.com/blogs/think/2018/11/tradelens-how-ibm-and-maersk-are-sharing-blockchain-to-build-a-global-trade-platform/ (accessed 5.30.21).

Senarak, C., 2021. Port cybersecurity and threat: A structural model for prevention and policy development. Asian Journal of Shipping and Logistics 37, 20–36. https://doi.org/10.1016/j.ajsl.2020.05.001

Shapo, V., 2018. CYBERSECURITY IMPLEMENTATION ASPECTS AT SHIPPING 4.0 AND INDUSTRY 4.0 CONCEPTS REALIZATION. Journal of Scientific Perspectives 2, 1–12. https://doi.org/10.26900/jsp.2018445371

Shipowners Still Not Ready to Give Up Control to Intelligent Vessels - Offshore Energy [WWW Document], 2019. . World Maritime News. URL https://www.offshore-energy.biz/orca-ai-shipowners-still-not-ready-to-give-up-control-to-intelligent-vessels/ (accessed 6.17.21).

Shipping and world trade: driving prosperity [WWW Document], n.d. . International Chamber of Shipping. URL https://www.ics-shipping.org/shipping-fact/shipping-and-world-trade-driving-prosperity/ (accessed 5.1.21).

Sinaeepourfard, A., Garcia, J., Masip-Bruin, X., Marín-Tordera, E., 2017. A comprehensive scenario agnostic Data LifeCycle model for an efficient data complexity management, in: Proceedings of the 2016 IEEE 12th International Conference on E-Science, e-Science 2016. Institute of Electrical and Electronics Engineers Inc., pp. 276–281. https://doi.org/10.1109/eScience.2016.7870909

Soner, O., Akyuz, E., Celik, M., 2018a. Use of tree based methods in ship performance monitoring under operating conditions. Ocean Engineering 166, 302–310. https://doi.org/10.1016/j.oceaneng.2018.07.061

Soner, O., Akyuz, E., Celik, M., 2018b. Use of tree based methods in ship performance monitoring under operating conditions. Ocean Engineering 166, 302–310. https://doi.org/10.1016/j.oceaneng.2018.07.061

Song, X., Huang, L., Fenz, S., 2012. Internet of Things Applications in Bulk Shipping Logistics: Problems and Potential Solutions. https://doi.org/10.1007/978-3-642-32427-7_81

Sun, X., Liu, C., Wang, J., Li, J., 2020. Assessing the extreme risk spillovers of international commodities on maritime markets: A GARCH-Copula-CoVaR approach. International Review of Financial Analysis 68. https://doi.org/10.1016/j.irfa.2020.101453

Tam, K., Jones, K.D., 2018. Maritime cybersecurity policy: the scope and impact of evolving technology on international shipping. Journal of Cyber Policy 3, 147–164. https://doi.org/10.1080/23738871.2018.1513053

Theodoropoulos, S., Tassopoulos, E., 2014. The Greek maritime cluster as a driving force for the economy's restarting.

Tian, Y., Li, T., Chen, C.L.P., 2017. Shipping big data and its application prospects in shipping company, in: 2017 4th International Conference on Information, Cybernetics and Computational Social Systems (ICCSS). IEEE, Dalian, China. https://doi.org/10.1109/ICCSS.2017.8091502

Tinga, T., Tiddens, W., Amoiralis, F., Politis, M., 2017. Predictive maintenance of maritime systems: Models and challenges, in: Safety and Reliability – Theory and Applications. CRC Press, CRC Press Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742. https://doi.org/10.1201/9781315210469-56

Tsioumas, V., Papadimitriou, S., Smirlis, Y., Zahran, S.Z., 2017. A Novel Approach to Forecasting the Bulk Freight Market. Asian Journal of Shipping and Logistics 33, 33–41. https://doi.org/10.1016/j.ajsl.2017.03.005

Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., Huang, G.-B., 2018. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey From Data to Methodology. IEEE Transactions on Intelligent Transportation Systems 19. https://doi.org/10.1109/TITS.2017.2724551

Underwood, S., 2016. Blockchain beyond bitcoin. Communications of the ACM 59. https://doi.org/10.1145/2994581

Verschuur, J., Koks, E.E., Hall, J.W., 2021a. Observed impacts of the COVID-19 pandemic on global trade. Nature Human Behaviour. https://doi.org/10.1038/s41562-021-01060-5

Verschuur, J., Koks, E.E., Hall, J.W., 2021b. Global economic impacts of COVID-19 lockdown measures stand out in high-frequency shipping data. PLOS ONE 16. https://doi.org/10.1371/journal.pone.0248818

Vidya, C.T., Prabheesh, K.P., 2020. Implications of COVID-19 Pandemic on the Global Trade Networks. Emerging Markets Finance and Trade 56, 2408–2421. https://doi.org/10.1080/1540496X.2020.1785426

Vouros, G.A., Doulkeridis, C., Santipantakis, G., Vlachou, A., 2018. Taming Big Maritime Data to Support Analytics, in: Popovich, V., Schrenk, M., Thill, J.-C., Claramunt, C., Wang, T. (Eds.), Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17). Springer, Cham. https://doi.org/10.1007/978-3-319-59539-9_2

Wahlström, M., Heikkilä, E., Granholm, G., 2019. Technology levels for maritime traffic coordination: towards the internet of intelligent ships, in: MTEC/ICMASS 2019.

Wan, Z., Makhloufi, A. el, Chen, Y., el Makhloufi, A., Tang, J., 2017. Decarbonizing the international shipping industry: Solutions and policy recommendations. Marine Pollution Bulletin 126, 428–435. https://doi.org/10.1016/j.marpolbul.2017.11.064

Wang, B., Liu, S.F., 2012. Port process reengineering based on information technology, in: Proceedings of International Conference on Engineering and Business Management.

Wang, M.M., Zhang, J., You, X., 2020. Machine-Type Communication for Maritime Internet of Things: A Design. IEEE Communications Surveys and Tutorials 22, 2550–2585. https://doi.org/10.1109/COMST.2020.3015694

Wang, Z., Silberman, J.A., Corbett, J.J., 2020. Container vessels diversion pattern to trans-Arctic shipping routes and GHG emission abatement potential. Maritime Policy and Management 1–20. https://doi.org/10.1080/03088839.2020.1795288

Wei, T., Feng, W., Chen, Y., Wang, C.-X., Ge, N., Lu, J., 2021. Hybrid Satellite-Terrestrial Communication Networks for the Maritime Internet of Things: Key Technologies, Opportunities, and Challenges. IEEE Internet of Things Journal. https://doi.org/10.1109/JIOT.2021.3056091

Wu, L., Xu, Y., Wang, Q., Wang, F., Xu, Z., 2017. Mapping Global Shipping Density from AIS Data. Journal of Navigation 70. https://doi.org/10.1017/S0373463316000345

Yan, R., Wang, S., Du, Y., 2020. Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship. Transportation Research Part E: Logistics and Transportation Review 138. https://doi.org/10.1016/j.tre.2020.101930

Yan, Z., Xiao, Y., Cheng, L., Chen, S., Zhou, X., Ruan, X., Li, M., He, R., Ran, B., 2020. Analysis of global marine oil trade based on automatic identification system (AIS) data. Journal of Transport Geography 83, 102637. https://doi.org/10.1016/j.jtrangeo.2020.102637

Yang, C.S., 2019. Maritime shipping digitalization: Blockchain-based technology applications, future improvements, and intention to use. Transportation Research Part E: Logistics and Transportation Review 131, 108–117. https://doi.org/10.1016/j.tre.2019.09.020

Yang, D., Wu, L., Wang, S., 2021. Can we trust the AIS destination port information for bulk ships?– Implications for shipping policy and practice. Transportation Research Part E: Logistics and Transportation Review 149, 102308. https://doi.org/10.1016/j.tre.2021.102308

Yang, D., Wu, L., Wang, S., Jia, H., Li, K.X., 2019. How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications. Transport Reviews 39. https://doi.org/10.1080/01441647.2019.1649315

Yang, L., Chen, G., Zhao, J., Rytter, N.G.M., 2020. Ship Speed Optimization Considering Ocean Currents to Enhance Environmental Sustainability in Maritime Shipping. Sustainability 12. https://doi.org/10.3390/su12093649

Yang, Y., Zhong, M., Yao, H., Yu, F., Fu, X., Postolache, O., 2018. Internet of things for smart ports: Technologies and challenges. IEEE Instrumentation & Measurement Magazine 21. https://doi.org/10.1109/MIM.2018.8278808

Yao, X., Yuan, X., 2018. Thoughts on the Construction of Smart Ports in China. Advances in Social Sciences 07. https://doi.org/10.12677/ASS.2018.78205

Yau, K.-L.A., Peng, S., Qadir, J., Low, Y.-C., Ling, M.H., 2020. Towards Smart Port Infrastructures: Enhancing Port Activities Using Information and Communications Technology. IEEE Access 8. https://doi.org/10.1109/ACCESS.2020.2990961

Yoshida, T., 2016. Utilizing big data and the internet of things in shipping. Sea Japan 1–32.

Yuen, K.F., Xu, G., Lam, J.S.L., 2020. Special issue on 'Artificial Intelligence & big data in shipping.' Maritime Policy & Management 47. https://doi.org/10.1080/03088839.2020.1790052

Zaman, I., Pazouki, K., Norman, R., Younessi, S., Coleman, S., 2017. Challenges and opportunities of big data analytics for upcoming regulations and future transformation of the shipping industry, in: Procedia Engineering. Elsevier Ltd, pp. 537–544. https://doi.org/10.1016/j.proeng.2017.08.182

Zaman, M.B., Siswantoro, N., Priyanta, D., Pitana, T., Prastowo, H., Semin, Busse, W., 2021. The Combination of Reliability and Predictive Tools to Determine Ship Engine Performance based on Condition Monitoring. IOP Conference Series: Earth and Environmental Science 698, 012015. https://doi.org/10.1088/1755-1315/698/1/012015

Zhang, C., Bin, J., Wang, W., Peng, X., Wang, R., Halldearn, R., Liu, Z., 2020. AIS data driven general vessel destination prediction: A random forest based approach. Transportation Research Part C: Emerging Technologies 118, 102729. https://doi.org/10.1016/j.trc.2020.102729

Zhao, H., Meng, Q., Wang, Y., 2020. Probability estimation model for the cancellation of container slot booking in long-haul transports of intercontinental liner shipping services. Transportation Research Part C: Emerging Technologies 119. https://doi.org/10.1016/j.trc.2020.102731

Zhou, X., Liu, Z., Wang, F., Xie, Y., Zhang, X., 2020. Using Deep Learning to Forecast Maritime Vessel Flows. Sensors 20. https://doi.org/10.3390/s20061761