FRAUNHOFER-CENTER FÜR MARITIME LOGISTIK UND DIENSTLEISTUNGEN CML

WHITE PAPER

MACHINE LEARNING IN MARITIME LOGISTICS

Lutz Kretschmann | Miriam Zacharias | Steffen Klöver | Tina Hensel

WHITE PAPER

MACHINE LEARNING IN MARITIME LOGISTICS

Lutz Kretschmann | Miriam Zacharias | Steffen Klöver | Tina Hensel

Fraunhofer
CML

IMPRESS

Fraunhofer Center for Maritime Logistics and Services CML

Am Schwarzenberg-Campus 4, Gebäude D
21073 Hamburg
Germany

Phone +49 (0)40 42878-4450
Mail info@cml.fraunhofer.de

www.cml.fraunhofer.de

TABLE OF CONTENTS


ML

PREFACE

Prof. Dr.-Ing. Carlos Jahn
Head of Fraunhofer CML

Dear Reader,

Data has become an exceedingly valuable resource that is used to power much of the transformative technology we see today. But – although inherently valuable – data requires careful processing and analysis to extract meaningful information. Thus, making the most of available data is a crucial task for future-oriented, successful management.

Due to modern IT systems and IT technologies used today in fleet management, ship operation and port logistics, the amount of data available to maritime companies is constantly expanding. Thoroughly analyzing this data reveals important cause-and-effect relationships, provides reliable forecasts and thus enables optimal decision making. The list of possible data-driven improvements in business practice is long and the potential for change is enormous.

For a comprehensive realization of data-based innovations, the classical methods of statistics are increasingly reaching their limits. At the same time, new concepts from the field of artificial intelligence are on the rise. In particular, machine learning techniques are very promising for various applications. But how can a company take advantage of the benefits that machine learning offers? Is sufficient data available and does it contain the right information? And what is a good starting point for machine learning? This white paper provides orientation by addressing these questions and more. It shows best practice solutions and approaches by describing examples from our projects with partners from the maritime industry and highlights critical success factors.

At Fraunhofer CML, we have gained knowledge and experience from numerous projects that use machine learning to optimize processes and systems in the maritime industry. Starting with the analysis of existing data and the identification of most promising machine learning use cases, through the development and testing of pilot solutions to the implementation of innovative applications, we offer customer-specific solutions. You are very welcome to contact us about opportunities for cooperation.

I hope you enjoy reading the white paper »Machine Learning in Maritime Logistics« and gain valuable insights!

The Fraunhofer Center for Maritime Logistics and Services CML supports the maritime industry in exploiting the potential of data to optimizes processes and systems along the maritime supply chain. To that end we develop, pilot and implement digital solutions for our customers that leverage the value of data with machine learning and enable data-driven decision making.

1 MANAGEMENT SUMMARY

Future-oriented maritime services will increasingly rely on the intelligent use of data from maritime supply chains, transport markets as well as fleet management and port operations. The comprehensive availability of digital data in real time, seamless data exchange and systematic analysis through machine learning technologies as a basis of data-driven decision making can lead to significant changes in the operating principles and business logic of maritime logistics.

Machine learning is considered a key technology for the optimization of business processes in maritime logistics and shipping. No doubt, it offers opportunities to improve cost-efficiency, safety and sustainability of transporting goods in maritime supply chains. While the excitement about machine learning continues to grow, far-reaching promises about the disruptive potential of data-based innovations are widespread, and it can sometimes be difficult to distinguish fact from fiction.

The white paper »Machine Learning in Maritime Logistics« attempts to clear the fog around artificial intelligence (AI) in maritime logistics and shed some light on the buzzword machine learning. At first, key factors of maritime digitalization – including opportunities and challenges – are presented. This provides an overview of the interaction between data, technology and machine learning while considering the specific nature of the maritime industry. Subsequently, the concepts of AI and machine learning are discussed in order to provide a good understanding of the fundamental principles involved and thus pave the way for a successful application in different parts of maritime logistics.

Although companies across the board are aware of the potential value of data, they are still cautious when it comes to implementing machine learning techniques. Providing guidance – in terms of an appropriate approach to data-based projects – can help a wider adoption of machine learning in practice. Starting from a description of data-based decision making, the white paper therefore presents important elements for a successful implementation of machine learning in maritime applications. First of all, this concerns finding those use cases for which machine learning methods are actually promising. Second, a process model is introduced according to which data-based projects can be implemented. Finally, different roles within a data science project are described that help maximize the overall chance for success.

Within the last part, the white paper aims to close the gap between theoretical advances in machine learning and a real-world practical use of data-based methods. To this end, actual examples of machine learning in different applications from the maritime industry are described. Drawing inspiration from such good practices can facilitate the identification of opportunities on a company level. First, practical examples for identifying suitable machine learning use cases and implementing a proof of concept are discussed. Then, best practices are introduced on how machine learning can be used for predictive analytics and optimization. Last, important insights for applying deep learning, reinforcement learning, and hybrid models in a maritime context are presented. The white paper concludes by describing possibilities for cooperation with Fraunhofer CML in the realization of data-based projects.

2 DIGITALIZATION IN THE MARITIME INDUSTRY

For several years the maritime industry has been in the middle of a transition process. At the end of this digital transformation stand promises of improved logistic processes, optimized cargo handling in ports as well as enhanced efficiency, safety and environmental sustainability of maritime transport. A key factor in realizing these opportunities and enabling the next efficiency leap in maritime logistics is the use of data on a much larger scale than in the past. Innovations in maritime connectivity are central to making this data available and exchanging it seamlessly within maritime transport networks. Creating value from this data with advanced analytics and machine learning can take place along maritime supply chains, in ports and shipping companies (see Figure 1).

A new level of real-time visibility along the maritime transport can lead to more flexible and dynamic digital supply chains that reduce inefficiencies and synchronize operations. Digital ports, in turn, create value by enabling a better handling of goods at the interface between ships and hinterlands which lowers waiting times and improves capacity utilization. Lastly, digital ships boost productivity of shipping companies by making fleet-wise voyage optimization possible or reducing operating cost via smart predictive maintenance strategies. However, even though the rewards are significant, there are some obstacles to overcome on the way. Thus, both opportunities and challenges of digitalization in the maritime industry will be outlined to conclude this section.

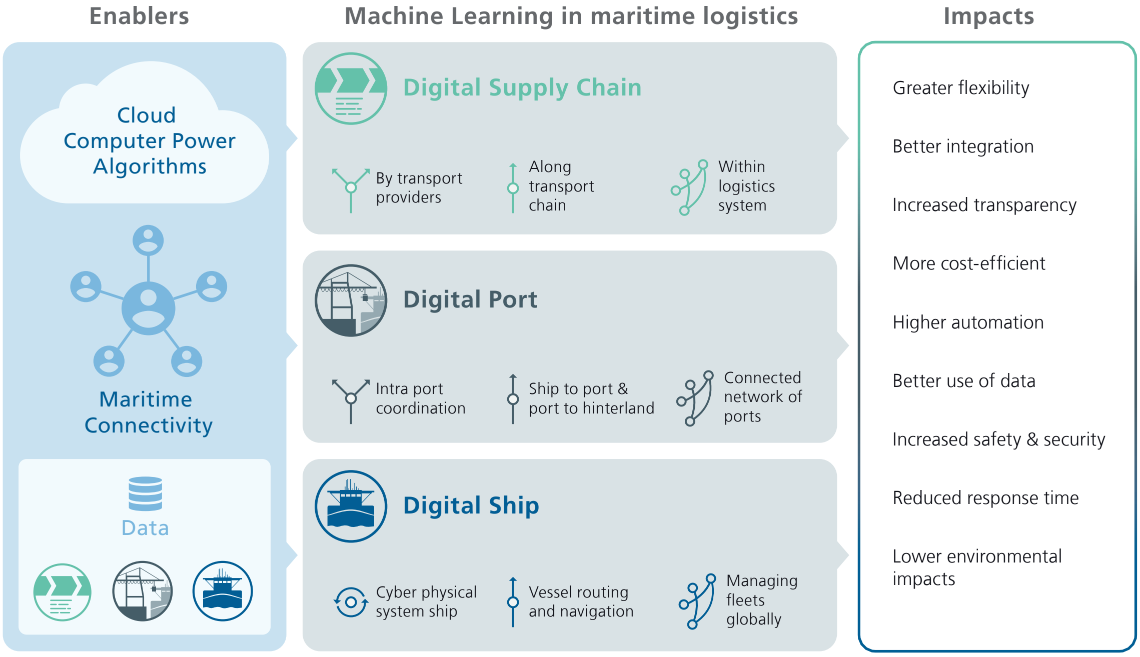

Figure 1: Machine Learning in Maritime Logistics – Enablers, Applications and Impacts

2.1 Maritime Connectivity

Data-based innovations that utilize machine learning methods are founded on the premise that sufficiently large amounts of data are available. Traditionally, limited connectivity or a lack of means to transfer larger data volumes altogether has been a major issue in maritime transport. Accordingly, innovative solutions were constrained by narrow bandwidth and limited data accessibility from shore and in some cases by ignorance of the importance of the details and information visible in the data.

However, this is changing rapidly as maritime connectivity has already seen significant improvements over the recent past and a continuation of this trend is predicted for the future. Nowadays, 3G or 4G is commonly available near shore and in ports where in some cases broadband connections can be used as well if necessary. [1], [2] In future, 5G will provide another flexible solution in ports. On the open sea, obviously, data exchange is only possible via satellite communications, which makes data exchange still rather expensive. Higher data transmission speeds and declining communication costs are expected and provide opportunities for innovative applications, which require frequent exchange of large amounts of data between ship and shore. [3]

Another milestone in the digitalization of maritime logistics is the obligation to equip merchant ships with transmitters of the Automatic Identification System (AIS). The automated exchange of position, speed, course and other data via AIS from ship to ship or ship to shore station has significantly increased both the efficiency and safety of maritime traffic since the introduction of AIS. At the same time, the availability of AIS data over many years has created a unique database that reflects the worldwide movement patterns of the entire merchant fleet as well as the navigational behavior of individual ships. [4] Both offer opportunities for completely new data-based solutions. This is further amplified by new ship-to-ship and ship-to-shore communication frameworks such as VHF data exchange system (VDES). [5]

Overall, current developments in the maritime area are leading towards a seamless interconnectivity between sea and shore and real-time exchange of large and complex maritime data sets. Together, this enables an intensified use of machine learning applications in maritime logistics.

2.2 Digital Supply Chains

Generating, sharing and using information within transport networks from the source of raw materials to the final consumer is a key function of today’s supply chain management. In this context, data and its systematic analysis can have a large impact. By leveraging technologies such as the Internet of Things and utilizing advanced analytics, digital supply chains can contribute towards reaching a new level of operational efficiency in maritime logistics.

Important functionalities in this regard include real-time visibility and reliable forecasts of the movement of goods as well as automated and data-driven decision making. Shipping companies, freight forwarders and transport providers can all benefit by offering more flexible, cost-efficient and better coordinated transport services. Additional potential lies in simplifying the transfer and processing of documents, information or financial transactions within digital supply chains, especially in maritime transport networks where a large number of geographically dispersed players work together. [6]

However, the highly competitive environment and business models that benefit from limited transparency can also hinder collaboration in digital supply chains. [7]

Nevertheless, shippers are increasingly focusing on a simultaneous optimization of entire transport networks to create added value. Here, technical innovations and worldwide connectivity enable seamless information exchange throughout a shipper's global supply chain. Machine learning can be especially useful to achieve increased visibility of individual shipments by extracting valuable information from large data volumes provided by cargo tracking devices or AIS. Also, data-driven decision-making – enabled by digital supply chains – can address the inherent complexity and dynamics of logistics processes much better than current solutions. At the same time new solutions help ensure security in digital supply chains by making data exchange tamper-proof and trackable. [8]

2.3 Digital Ports

Ports have always been locations where technological innovations play a major role. As intermodal hubs, ports must react quickly to changing conditions, ensure an efficient handling and storage of goods, enable value-added logistics services and connect with various hinterland transport modes. All this is happening under increasing transport volumes due to the global division of labour, higher safety and security requirements and the obligation to protect the environment. [9]

Port centred innovations can especially be found in the physical handling of goods, loading units and modes of transport as well as in information technology to optimize port calls. Of particular importance is the extension of the planning horizon of port stakeholders through intra and inter-port collaboration, ship-to-port collaboration and port-to-hinterland collaboration. Advantages result from the exchange of event data related to port calls among all relevant stakeholders. This allows more efficient use of resources and shorter vessel turnaround times. [10] Ship arrivals and departures command the heartbeat of every seaport. Tuning this rhythm of international transport across multiple smart ports promises further benefits. Through information exchange and data-based predictions, handling and transport capacities to be better coordinated, opening up opportunities to reduce costs and increase efficiency along digital supply chains.

In addition to connecting ports as a basis for optimizing maritime transport networks, autonomous vehicles represent a second innovation driver with the potential to disrupt port operations as we see them today. Promising use cases can be found in land-based terminal and intra-port transport, where automatically guided vehicles are already state of the art. Beyond that, additional innovations focus on the water side, where remote-controlled or partially autonomous tugs support mooring and unmooring manoeuvres. Furthermore, small unmanned surface vessels can carry out parts of intra-port transport. If equipped with environmental sensors such as radar, cameras or sonar, they also serve as a data collection platform to help inspection and predictive maintenance of port infrastructure. [9]

2.4 Digital Ships

Building upon the ideas of Industry 4.0 and the Internet of Things, a multitude of opportunities open up around the digitization of ships. Today, new vessels increasingly resemble cyber-physical systems, where on-board software components connect mechanical and electronic parts via a dedicated data infrastructure. As a result, current data from all systems, components and information processes on board is permanently available as a basis for improving ship operation. [1]

The use of data collected during the operational phase of a vessel is increasingly seen as the main driving force for the next leap in maritime transport efficiency and, along with that, a reduction in ship operating costs. With the comprehensive availability of digital data in real time, its seamless exchange and systematic analysis, far reaching changes in the practice of ship operation are expected. Fields of innovation include the individual systems on board, the entire operational spectrum of a ship and the way entire fleets are managed.

Optimization is not limited to the most economical operation in terms of fuel consumption but also considers ensuring a seamless transport from door to door or providing higher safety by collision avoidance. Moreover, maintenance can be improved by progressing from rigid "Planned Maintenance" programs to "Condition-based Maintenance" systems. Another important innovation field is navigation and ship control, with developments such as smart nautical assistance systems, augmented reality for remotely-controlled vessels and even completely autonomous maritime surface ships.

Developments in other industries have shown that digitalization is not only fundamentally changing business processes, but also creates entirely new business models as more data becomes available. The maritime industry is no exception. Emerging data-based services and solutions are already beginning to disrupt the current industry logic. Advanced analytics and machine learning often play a key role in these new digital functionalities and services by ship managers, the shipbuilding and marine equipment industry, classification societies and other maritime service providers. [1]

2.5 Opportunities and Challenges

Digitalization initiatives by ports and ships in recent years – be it the obligation to equip AIS transmitters, improved network architectures or better connectivity within maritime supply chains – have created a unique maritime and continuously growing database that offers wide-ranging opportunities for improving maritime logistics. Leveraging the value of this data with tools provided by AI and machine learning is seen as an increasingly important aspect of gaining and maintaining competitive advantages in all segments of the maritime sector from shipping, ports, shipbuilding and the maritime supply industry to offshore wind and marine engineering. It allows companies to:

- Reduce costs through optimized operations and data-driven decision making,
- Improve quality control through digital monitoring solutions,
- Increase safety through incident predictions,
- Capture knowledge hidden in past business records,

- Identify decision-relevant information in large data sets,Reduce effort for handling and processing of documents,
- Automatize processes using intelligent assistants,
- Improve safety with autonomous or remotely-controlled operations, and
- Establish new business models and products.

By using machine learning and AI methods to automatically process large amounts of data and convert them into valuable information, a wide range of practical problems in maritime logistics can be addressed with data-driven decision support, as this white paper will show. Nevertheless, besides opportunities, there are also a number of challenges that companies face in this context as described in Figure 2.

OPPORTUNITIES

- Identify decision-relevant information in large data sets
- Establish new business models and offers
- Automatize processes using intelligent assistants
- Reduce costs through optimized operations & data-driven decision making
- Capture knowledge hidden in past business records
- Enhance safety with autonomous or remotely-controlled operations
- Improve quality control
- Reduce effort for handling and processing of documents
- Increase safety through incident predictions

CHALLENGES

- Data not available in sufficiently large quantity
- Insufficient data quality can result in a high workload of cleansing and pre-processing
- Expectations for AI and machine learning can rise to unsustainable heights
- Incomprehensibility and traceability of machine learning algorithms
- Missing explainability of machine learning algorithms in practical solutions
- Computational power needed to solve some machine learning tasks
- Missing competences and talents in companies

Figure 2: Opportunities and Challenges of Machine Learning in Maritime Logistics

Per definition, data science and machine learning projects revolve around data. Accordingly, data is the most obvious limitation that companies are confronted with. Machine learning can only produce convincing results if the problem at hand is actually solvable based on the given data. That almost sounds trivial, but the current hype around machine learning and AI has raised some expectations to unsustainable heights. Data analytics provides a powerful set of tools but the underlying data must be representative of the problem and data sets must be sufficiently large. For many business use cases, creating or obtaining such data sets can be difficult. Moreover, insufficient data quality is a related challenge that companies have to address. In consequence, a large part of the time and effort in data driven projects is actually spent on collecting, structuring and storing data as well as cleaning and preprocessing it.

Besides data related challenges, another issue for a practical application is known as the interpretability problem of AI algorithms. Both the input data and the output or decision of a trained machine learning model can be observed, but the exact mechanisms how and why the output was generated remain hidden. Not knowing how such a black box model behaves can be a problem in practical applications and achieving sufficient "explainability" is of particular concern for high-stake decisions. [11]

Moreover, some applications of machine learning require considerable computational power, which fortunately can be offset today – especially in smaller companies – by using external data centers. [12] The acquisition of required competences and talents for using machine learning can cause far greater headaches. Qualified data scientist and data analysts are in high demand across industries. Where companies have difficulties in recruiting skilled personnel, or to further explore the capabilities of their own datasets, one alternative is an increased cooperation with external partners and research organizations. With that in mind, the following sections will give an insight into conducting machine learning projects successfully.

3 ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Artificial intelligence is a wide field in the area of computer science concerned with building machines that are capable of performing tasks on par with human intelligence. AI is considered as one of the transformational technologies of the information age with impacts in all industrial sectors.

AI covers a broad spectrum of possible abilities. Depending on how close these abilities compare to cognitive functions of humans, it is either classified as weak AI or strong AI. [13] The ultimate aim of AI is a system that can perform cognitive tasks like perceiving, learning or problem solving in any environment or domain. Such a strong or general AI would be able to generalize knowledge between domains, transfer insights from one domain to another and adapt flexibly to changes in the environment.

As of today, this strong version of AI equivalent to human capabilities remains in the realm of science fiction. In comparison, existing real-world applications of AI used in industry and consumer-facing products have a narrower scope than general human-level intelligence. They are commonly labeled as weak AI. Nevertheless, the term weak does not imply that the system is not powerful. These applications are mostly highly specialized and can achieve human-level performance in perceptual tasks, like image and speech recognition. [14], [15]

The development of AI faced several falls and rises throughout its history. After being established as a research field in the 1950s there were some setbacks due to unfulfilled expectations in addressing real-world problems. Early AI programs that relied on combinatorial exploration failed in scaling to larger problems due to exploding complexity. [13] Many approaches used explicit rules handcrafted by a programmer to solve problems, which were suitable for solving well defined problems, like playing chess, but were unable to deal with fuzzy, unbounded problems like image- and speech recognition. [16] Also, some early forms of learning-based methods were lacking the capacity to sufficiently represent knowledge.

In the 21th century and especially after 2010 the research field AI gained new popularity. Three major factors contributed to significant progress in AI since then: the availability of data, hardware and algorithmic advances (see Figure 3).

The distribution of data over the internet provided the basis for AI to learn from ever-increasing datasets. Data can be present in structured form like tabular records of transactions or unstructured data such as collections of text-documents, images and videos. Processing ever larger amounts of data was made possible by the wide availability of graphical processing units (GPUs) that replaced large clusters of CPUs for executing parallelizable algorithms often used in AI. Learning transformations of data to arrive at meaningful representations was the key to solve many real-world problems. A representation is simply a different view on the input data that is more useful to solve a given task. In the domain of machine learning, complex structures used to simulate intelligent behavior – like deep artificial neural networks – were developed as well as advanced algorithms for efficient learning and inference.

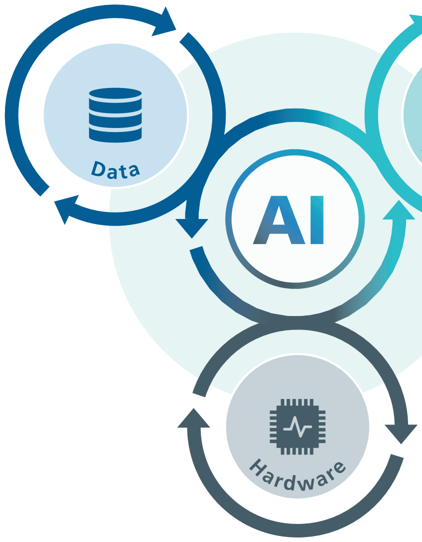

Figure 3: Enablers of Artificial Intelligence

3.1 Machine Learning as a Subset of Artificial Intelligence

Machine learning is a form of artificial intelligence that enables a system to learn from data rather than through explicit programming. In this paradigm, rules are not handcrafted by a programmer to obtain an answer from consuming data but they are the outcome of the implemented machine learning algorithm itself (see Figure 4). These automatically generated rules can then be applied to new data to produce answers. Machine learning thereby provides an AI with the capability to adapt to new circumstances and detect patterns. [13] Patterns in data can be anything, from sequences of words in a text-document to groups of pixels depicting an object in images. The goal of machine learning is to find a model that fits both learned and unseen data. The capacity of a model largely determines how much different patterns can be learned. Learning misleading or irrelevant patterns should be prevented in order for the model to generalize well to new data. [16]

In other words, machine learning is a form of applied statistics with an emphasis on the use of computers to approximate complicated functions. [17] A machine learning algorithm learns from recognizing patterns in data. The outcome is a model that is able to transform data into a meaningful output, i.e. an answer to a specific problem. This model could be comprised of a simply linear function or might be a complex graph structure like an artificial neural network. In the training phase, the parameters of the model are adjusted or learned by optimizing a performance measure. Afterwards, the model is ready to be deployed into production for making predictions. With more (high-quality) data available for the training of a machine learning model, insights can become more precise and more complex problems can be addressed.

One of the most recent, very impressive advances in machine learning can be found in deep learning, often implemented in form of artificial neural networks. Artificial neural networks represent complex nonlinear functions with a network of linear threshold units. [13] "Deep" in this context refers to putting an emphasis on learning successive layers of increasingly meaningful representations. [16] The

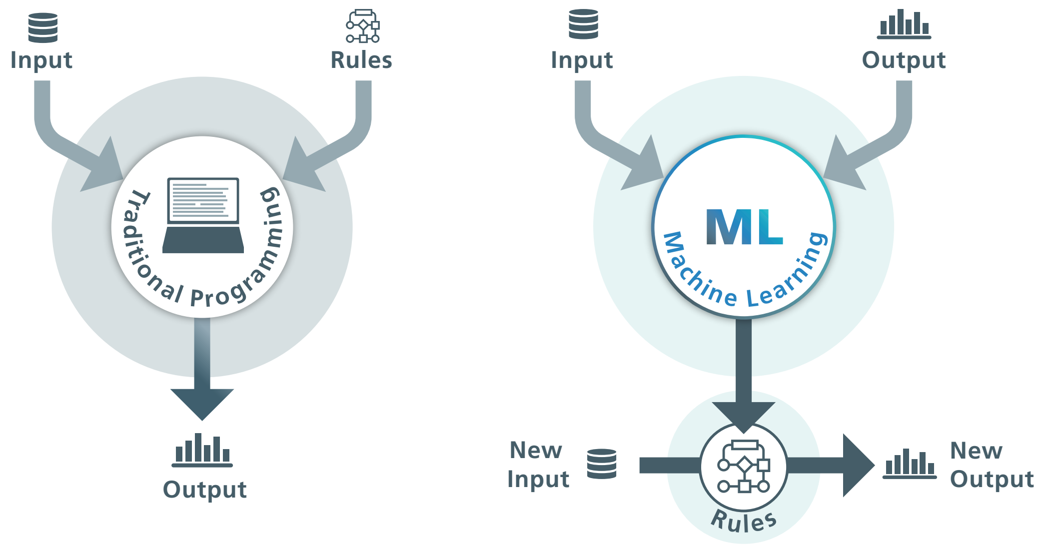

Figure 4: Paradigm Shift from Traditional Programming to Machine Learning [16]

depth of an artificial neural network is characterized by the number of stacked layers. Perceptional tasks, like image, text and sound recognition, are unbounded problems which require a high level of model capacity with millions of adjustable parameters, making deep learning models relatively large compared to traditional machine learning models.

Recurrent- and convolutional neural networks are widely used architectures in this field. Recurrent neural networks (RNNs) have the ability to store information from previous outputs to be used as inputs, providing them the capability of learning from complex sequences, such as text or time series data. Convolutional neural networks (CNNs) are specifically designed to extract important features from data, making them perfect for tasks like image classification, where the relevant objects are naturally surrounded by a background, irrelevant to the task.

3.2 Types of Machine Learning

Machine learning can be subdivided into three main types of learning problems: supervised, unsupervised and reinforcement learning (see Figure 5). They vary largely with regards to areas of application, used methods and data requirements. [18]

In supervised learning, a model is trained by learning the relationship between input data samples (explanatory variables or features) and their corresponding target variables (output or label attribute). This means, that for every input pattern the output value is known. Both are combined in a labeled example. In a classification task, the model learns to map the input variables to a finite set of categories or labels. For an image recognition application, this might refer to identifying objects of a specific type, i.e. different kind of ships at sea shown on a picture. If the desired output is continuous, the task is called regression. Predicting the yield of a plant based on sensory data or the expected arrival time of ships based on AIS-data would be examples of regression problems. Supervised machine learning is the most commonly used and successful type of machine learning today. [19] Popular algorithms include decision trees and artificial neural networks.

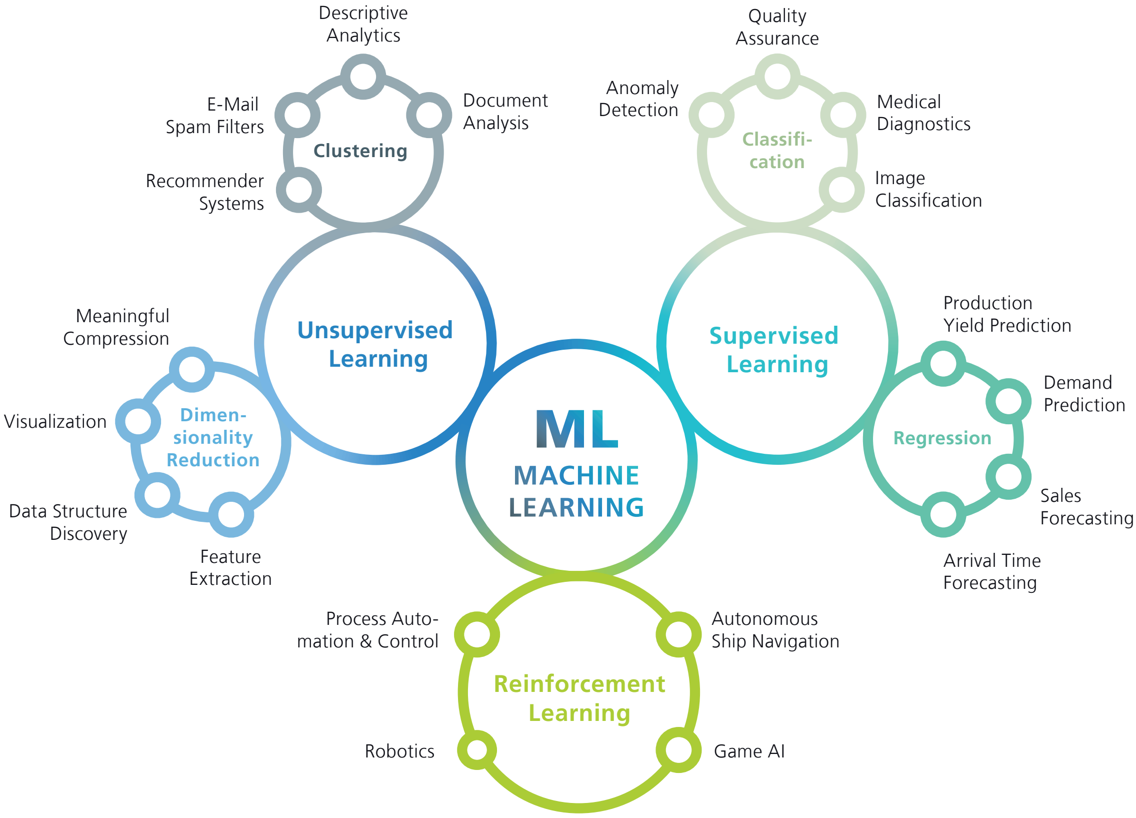

Figure 5: Types of Machine Learning and Areas of Application

If there are no corresponding target variables that describe the input data and create labelled examples, a machine learning algorithm can only learn patterns within the input data that are characteristic and describe its structure. These types of learning problems are called unsupervised, since there is no expected output provided. A prominent example are clustering algorithms, which are able to detect groups in data based on some similarity measure. These might be route patterns of vessels, used for trajectory prediction.

Reinforcement learning is the last type of learning problems. Here the algorithm learns to find suitable actions for a given input in order to maximize a reward. In contrast to supervised learning, the optimal outcome is not provided but instead discovered by a sequence of trial and error. [20] In robotics, learning the correct sequence motions to complete a task is an example of applied reinforcement learning. Furthermore, the approach is popular in creating AI systems, capable of outperforming humans in games like chess but increasingly in technical or business-related settings as well. [21]

3.3 Applications of Machine Learning

Defining the desired solution under consideration of the available data is usually the first step when thinking about applying machine learning in logistical or industrial processes. The desired outcome is often characterized as some kind of optimization in terms of cost, time or safety improvements. Depending on the type of problem, there might be massive amounts of data available, but the information needs to be extracted in an intelligent manner to be useful. In other cases, it might be necessary to install cameras or additional sensory equipment to obtain input data in sufficient quantity and quality or to find a way to integrate knowledge of human experts. Subsequently, a number of successful applications of machine learning are described to give an idea of the many possibilities in maritime logistics.

Machine learning enables the automation of work processes which currently contain manual steps carried out by humans in order to improve efficiency, speed or accuracy. Monitoring processes through machine learning applications is able to provide increased quality, reduced downtime and workloads. Descriptive statistics and unsupervised learning methods such as clustering can be used to gain new insights for optimization. Even control of processes can be advanced through machine learning. Reinforcement learning algorithms can adjust parameters of industrial processes and learn through immediate feedback.

Anomaly detection is a subcategory of classification, where it is determined whether a specific input is out of the ordinary. This might include a predictive maintenance task, where a model is trained based on sensory data, such as vibrations, and then able to detect readings that might suggest a defect. Other possible applications are fraud or intrusion detection in IT-Systems.

Computer vision applications deal with extracting useful information from digital images. From face recognition in mobile applications to autonomous driving, there are a variety of successful applications in industry and consumer facing products. Powered by deep learning and convolutional neural networks, objects in industrial or logistical settings can be recognized with high precision and accuracy from image- or video data.

Recommender systems are about discovering relevant items. These items might be documents related to a search query. In contrast to a simple search based on captions or file naming, the recommender systems are able to compare documents by content. This might be based on relative word frequencies or even more complex on semantic comparison. Here, clustering algorithms and natural language processing approaches are used to provide an intelligent search.

Forecasting events or potential outcomes with machine-learning-based predictive analytics provides a new level of accuracy and enables a better management of uncertainties in operations. Examples are demand predictions, arrival time forecasts or remaining life time estimations. Recurrent neural networks can learn complex relationships from time series data und make predictions for decision support or real-time applications. In addition, combining the predictive power of machine learning with mathematical optimization brings substantial scope for efficiency improvements in operational logistics decisions, such as finding optimal routes for vehicles or exploiting storage capacities in an optimized manner.

Not every machine learning technique is suitable for every business problem. Identifying the most suitable algorithm and model framing for a given real-world application can be a challenging task. Additionally, machine learning models have to be integrated with existing systems in most cases. Therefore, applying machine learning in real-world maritime applications involves a high degree of customization. Fraunhofer CML supports customers from the maritime industry in this context by realizing machine learning applications specified to their individual needs and goals. The developed solution can be integrated with existing infrastructures, business applications or mobile devices.

4 A SYSTEMATIC APPROACH TO DATA-BASED PROJECTS

Before getting started with machine learning projects in practice, it is helpful to take a step back and define a structured and systematic approach for achieving a specific project goal. Advancing digitalization and the increasing data availability have also given rise to new project management models that take into account the specific requirements of data-based projects. Considering latest scientific findings, this section introduces important elements that facilitate a successful implementation of machine learning in maritime logistics.

4.1 Data Science and Data-driven Decision Making

Although companies have proven to be successful when making decisions based on experience and "gut instinct", [22] the value of data-based decision making is undoubtedly agreed upon and an important factor to improve performance. [23] It would hardly be possible, for example, to reach todays levels of process automation without utilizing data. A fully automated production line requires a variety of sensor-data from different work stations and other sources to keep track of the system's state and to feed the manufacturing execution systems with production data collected from the shop floor. [24]

Moreover, providing required information at the right time is of immense value for decision makers. For example, shipping companies have to consider different influencing factors for freight rate quotes, including bunker fluctuations, currency rates, or available ship capacities. Advanced analytics and machine learning can provide reliable forecasts for such influencing factors, thereby enabling carriers to optimize freight rates across channels. Furthermore, patterns in historic shipment data can be analysed to identify underused capacity or enable customers to choose their preferred mode of transport and routing, depending on price points and transit times. [25]

By combining data from different sources in supply chains or logistics networks, transparency can be increased significantly and a single source of clean and pre-processed data allows for predictive insights. In this context, data-driven decision-making assumes the function of combining, analysing and visualizing data from different sources and providing it as an input for decision-making processes (see Figure 6), utilizing methods from different areas including data science, machine learning and advanced analytics.

provide reliable forecasts for such influencing factors, thereby enabling carriers to optimize freight rates across channels. Furthermore, patterns in historic shipment data can be analysed to identify underused capacity or enable customers to choose their preferred mode of transport and routing, depending on price points and transit times. [25]

By combining data from different sources in supply chains or logistics networks, transparency can be increased significantly and a single source of clean and pre-processed data allows for predictive insights. In this context, data-driven decision-making assumes the function of combining, analysing and visualizing data from different sources and providing it as an input for decision-making processes

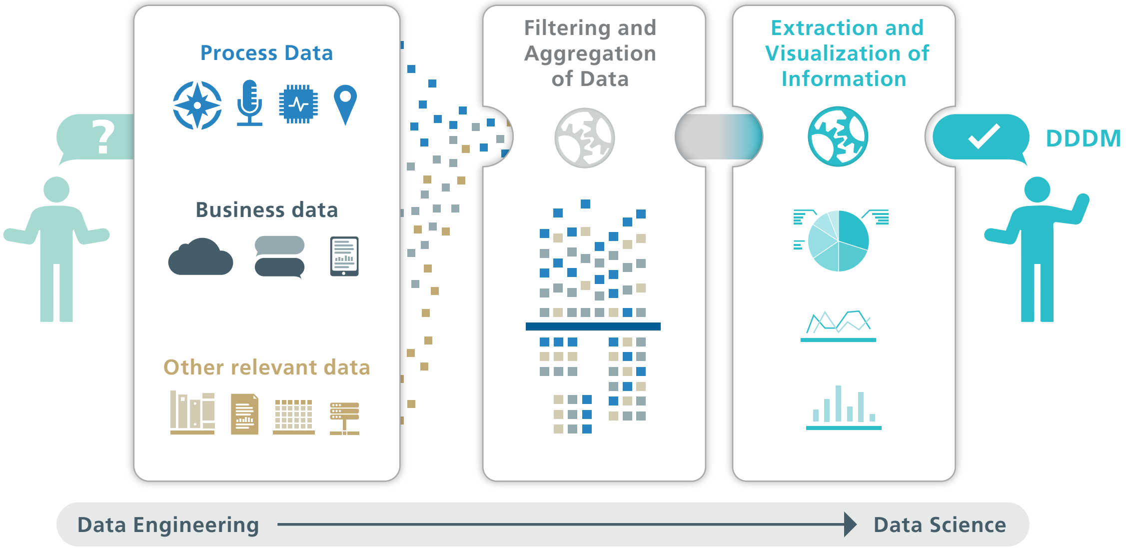

Figure 6: Data Processing Steps for Data-Driven Decision Making

(see Figure 6), utilizing methods from different areas including data science, machine learning and advanced analytics.

It can be beneficial for practitioners to understand how methods from these scientific fields interact in order to define a suitable project framework and to bring together the right qualifications for a data-driven project. Data science aims at understanding a problem by making use of principles, processes and techniques for (automated) analysis of data, with the underlying goal to improve decision making. It is closely linked to other data-related processes in the organization. In particular, data engineering plays an important role by processing, filtering and aggregating data from different sources, which is a prerequisite for extraction and visualization of information by data science tasks. [26]

Machine learning algorithms are well suited and often used for data-driven decision making. Although machine learning receives a lot of attention lately, it is often left out that it requires high-quality data and often a lot of data engineering. Thus, it benefits significantly from a suitable information architecture which defines clear processes for collecting data from different sources, filtering, standardizing, pre-processing and merging it in a common platform and checking its quality and fitness for purpose. This does not only ensure a well-managed database but also makes it easier to gather valuable data throughout the company and its supply chain network. Depending on the level of data maturity achieved in a company, applications can be limited to rather simple descriptive techniques, permit more sophisticated predictive analytics or – once a high maturity is achieved – make extensive use of advanced concepts and machine learning methods (see Figure 7).

Data-driven decision making provides powerful solutions to advance from a system based primarily on intuition, experience and ad-hoc processes. This does not mean that it is the best solution for every case or an all-or-nothing approach replacing experience and domain knowledge. On the contrary, the goal is to design and implement data-driven decision-making systems which complement expert knowledge and enlarge the horizon of the decision maker. Fraunhofer CML supports companies from the shipping and logistics industry in defining a data-driven decision-making strategy as well as establishing the necessary preconditions.

DATA MATURITY LEVEL – LOW	DATA MATURITY LEVEL – MEDIUM	DATA MATURITY LEVEL – HIGH
Use descriptive methods and statistics for:	Use predictive analytics methods for:	Use advanced analytics methods for:
• Initial analysis of status-quo • evaluation of potential data-based improvements	• Implementation of first data-driven decision making use cases	• Optimization of processes • strategic use of data-driven decision making

Figure 7: Data Maturity Level as Prerequisite for Data-Driven Decision Making

4.2 Applicability of Data-driven Methods

Although data-driven decision making is recognized as a key to success in a variety of applications, a vast majority of companies has not yet started exploiting AI in its different forms. According to a recent study, only 20 % of German logistics companies were using big data analytics while the figure was an even lower 6 % when asked about AI technologies. [27] Meanwhile, 90 % regard AI as very important to simplify business processes and increase productivity in the transport and logistics sector. [28]

This supports the conclusion that there are a number of potential applications in shipping and logistics, where data-driven optimization and decision making can be introduced to great benefit. Naturally, the first challenge is to identify use cases that are most promising for data-based methods such as machine learning. A rather obvious requirement for applying data-driven methods is the availability of data or the possibility to collect data throughout the project duration respectively. Data can be directly process-related, like sensor measurements or historical business-related data, but might also consist of external information from sources outside the company.

However, even before available data is systematically evaluated in terms of quantity and quality, the underlying problem deserves special attention. Just because a data-based approach to solving a problem seems promising does not mean that the use case is actually suitable. A thorough screening and review of interesting candidates helps to concentrate valuable resources where a solution is ultimately feasible. A simple consideration to test whether machine learning can solve a problem is to look for cases that a data scientist should be able to manage, assuming that time is not an issue.

There are also a number of criteria that can help to identify use cases where designing reliable forecasting or prediction models is promising. These include: [29]

- Causal factors are measurable and known,
- Large amount of historical data is available,
- Forecasts do not influence the real-world outcome,
- Future developments follow the same pattern as the past.

The more these criteria are met for a particular forecasting problem, the greater the chance of successful implementation in practice. To give an example, some natural events – like tide times – can be predicted with high accuracy (see Figure 8). In this case, factors causing the event are well known and data of past tide phases is available. Furthermore, there is no interrelation between the prediction of high and low tide and the actual occurrence, and it is safe to assume that the physical laws responsible for changing water levels will not change in the future.

Problem area	Causal factors	Big data	No interrelation	Future = past
Stock price	✗	✓	✗	✓ short-term
Customer demand	✓ short-term	✓	✓	✓ short-term
Tide times	✓	✓	✓	✓

Figure 8: Criteria to Identify Promising Problems for Prediction Models

4.3 Approach in Data-based Projects

Once the initial considerations on the applicability of data-based methods are completed, project planning is the next step. Like other industrial projects, data-based projects require a structured approach throughout all five project management phases from conception and initiation to planning, execution, performance monitoring, and lastly project closure.

The approach, usually implemented in data-based projects of Fraunhofer CML, is presented subsequently. Setting up the project plan takes reference from the Cross-Industry Standard Process for Data Mining (CRISP-DM). [32] Following this approach when analyzing large data sets with machine learning and statistical techniques – in other words data mining – ensures that all relevant aspects are considered, provides a uniform vocabulary and allows for an agile data mining process.

CRISP-DM breaks down a data mining project into six individual phases which are described further in the following (see Figure 9). These phases are not linear and strictly subsequent but have multiple interdependencies. That way individual phases can be repeated several times until an overall satisfactory result is achieved. For example, if it turns out that the amount of data is insufficient, the process goes back to the Data Understanding phase in order to broaden the data base for a next iteration.

The first part of a data mining project – which usually represents between 10–20 % of overall project duration – actually has little to do with data. Instead, the aim of this Business Understanding phase is to gain a consistent understanding throughout the project team regarding background, business objective and success criteria which define the area of investigation. Depending on the case at hand, this may require an assessment of requirements and constraints, risks and contingencies, costs and benefits and aspired business value that data mining is supposed to provide. It further includes discovering what data is available to the project and which resources can be used for its completion. On this basis, the next step is defining project objectives and success criteria clearly, e. g. predict the expected number of shipments per day to a certain level of accuracy. Subsequently, the last step translates the project objective into a project plan and establishes a project timeline.

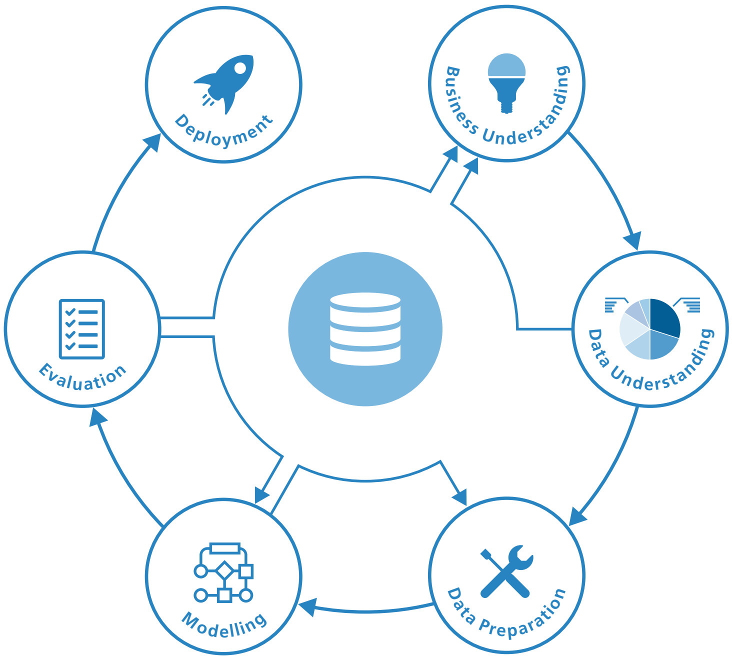

Figure 9: Phases of the Cross-Industry Standard Process for Data Mining [33]

In contrast to the first, the second phase is all about data. It starts with collecting data from different sources inside and outside of the company, describing its properties and merging it in a common database. The phase further includes visualizing and comparing different data attributes in order to verify that relevant project requirements are met. Subsequently, an explorative data analysis is carried out with basic data science and statistical methods – including data visualization in form of histograms, correlation matrices, or box plot diagrams and variable extraction techniques – to provide a basic understanding of the data and generate first findings. Closely associated herewith is the final step of the Data Understanding phase – combined about 20–30 % of project duration – in which the data quality is checked, including completeness, missing or implausible values and outliers.

With 50–70 % the next Data Preparation phase usually takes up most of the time and effort in a data mining project. It includes all necessary steps to construct the final data set used for actual knowledge discovery with advanced statistical or machine learning methods. Not all available data can or should always be used and thus the first step is selecting data based on considerations such as type, relevance, quality or volume. Where data quality issues were identified previously, adequate approaches to clean the data or estimate missing values need to be designed and implemented now. Before data from different sources is combined and the final data set is constructed, feature engineering (constructing new attributes from existing values e. g. through a combination), pre-processing (e. g. normalization or standardization) and transformations (e. g. representing categorical data as numeric values) may be necessary. Overall, the individual steps of data preparation depend on both, the methods and modelling tools used in the project and on the question that data mining is supposed to answer.

During the Modelling phase – which covers about 10–20 % of project duration – the project team choses different data mining techniques and determines their performance with the given data set. Usually different algorithms – e. g. artificial neural networks or random decision forests – can be used for one problem. Once these are identified and selected, their performance is determined empirically within a specified test design by using quality measures that reflect the defined project objectives and success criteria. It can also be helpful to identify suitable benchmarks against which to compare the performance of complex data-based models (e. g. a naive classifier that predicts the majority class or classical statistical methods such as moving average). To ensure the quality and validity of a model, the available data is usually divided into two parts; one part is used to build the model (training data set) while the other is used to check validity and accuracy of predictions made with that model (test data set). Finally, the last step is a discussion of model performance between both, data scientists and domain experts.

Subsequently, the Evaluation phase – typically 10–20 % of time – reviews results of the modelling exercise with particular emphasis on the initial problem statement and business objective defined in the first phase. In most cases, this will require technical, domain and data science expertise, but above all management involvement is important, because it leads to the decision how to use the data mining results in the business context. That means concluding the project and using the gained insights or moving on to the next and final Deployment phase – 5–10 % of time and effort – where a first demonstrator solution is developed further, integrated with legacy systems and used "live" within the organization's decision-making processes. Additionally, deployment includes drafting a maintenance plan for the developed solution, monitoring its correct functioning and collecting feedback for future improvements. These ideas are also a starting point to begin with the next data mining project.

4.4 Roles in Data Science Projects

A special feature of data-based projects is the interdisciplinary expertise they require. Roles can be divided into four main areas (see Figure 10). These need to be staffed with appropriately qualified team members to maximize the overall chances of project success. To address the problem from all sides, it is necessary to involve both a domain expert and a data scientist. These are supported by an IT specialist and a technical expert who is familiar with the hardware and processes under consideration.

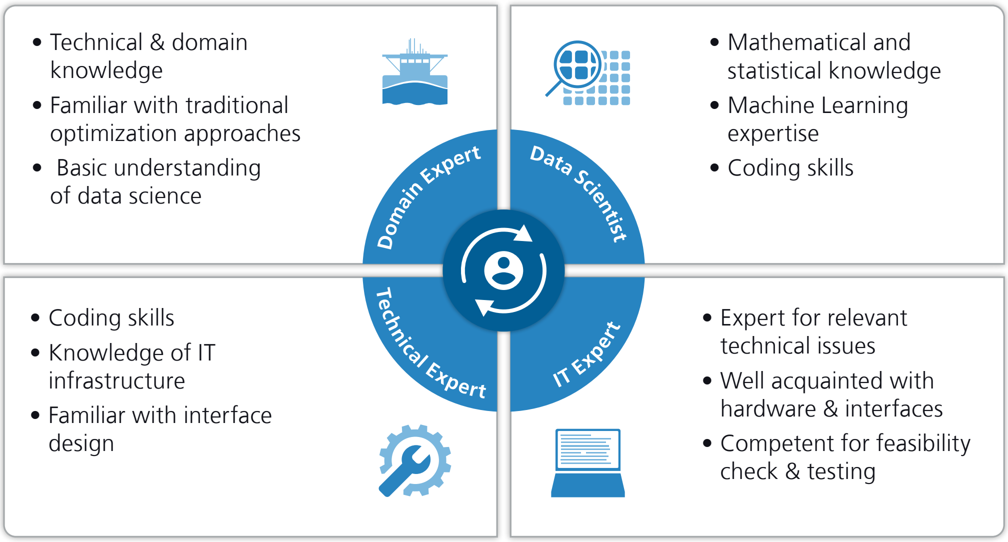

Figure 10: Roles in Industrial Data Science Projects

In most cases, a large amount of valuable knowledge is present inside companies, from process manager to IT support or machine operators. Ideally, an experienced employee who has access to this knowledge is chosen as domain expert and project lead. He or she cooperates closely with all other roles and provides the data scientists with required process knowledge that enables defining a suitable machine learning approach. The domain expert validates the proposed approach, provides feedback throughout the project phases and leads the evaluation of achieved results.

The implementation of statistical methods and machine learning models including prior data preparation and analysis lies primarily in the hands of the data scientist. Obviously, this requires in depth knowledge of advanced analytics methods. As specific mathematical or machine learning expertise is not always available inside the company, external partners or research institutes can support or take over the data science role in a project.

Lastly, IT and technical experts are essential support functions not only for building data-driven solutions but also for integrating machine learning models into the existing IT infrastructure. They should already be involved at an early project stage, during the requirements specification phase, so that the developed software can be integrated easily into existing systems.

5 MACHINE LEARNING USE CASES IN MARITIME INDUSTRY PRACTICE

After introducing essential principles of machine learning and describing important considerations for the implementation of data-based projects in the previous chapters, the following sections focus on actual applications of machine learning in maritime logistics. By presenting various examples that have been implemented or are currently being developed, the multitude of opportunities for machine learning in maritime logistics is shown. Drawing inspiration from these best practice projects can help practitioners in defining a company-specific approach to implementing machine learning and data-driven decision making.

The first section revolves around machine learning use case definition and describes how this task can be completed successfully. Next, a proof of concept – which checks whether a selected concept has practical potential – is addressed as the logical step before actual implementation. Last, practical applications of machine learning in five important areas are discussed (see Figure 11).

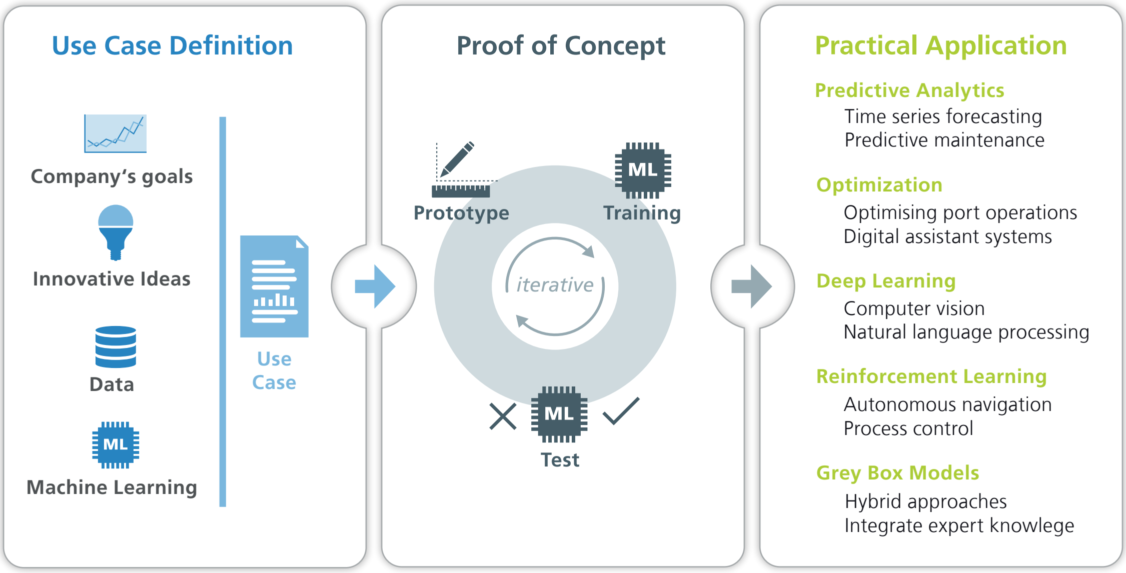

Figure 11: From Use Case Definition to Practical Applications of Machine Learning in Maritime Logistics

5.1 Identification of Use Cases

Although companies see great potential in data-driven optimization, they sometimes face difficulties in getting started with machine learning. In this context, a sensible first step is a dedicated analysis of given potentials for machine-learning-driven improvements on an individual company level. The approach has a low entrance barrier and aims at quick but evidence-based results.

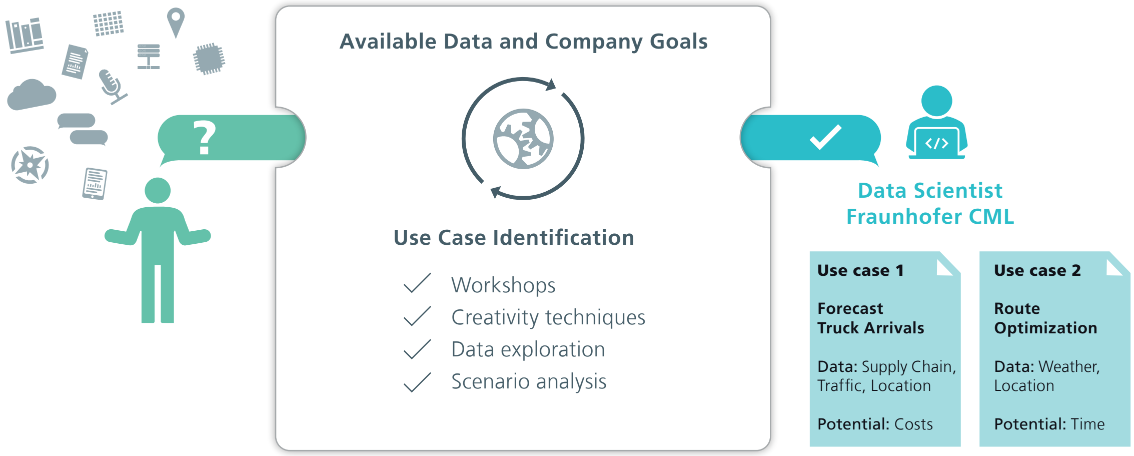

Figure 12: Analysis of Machine Learning Potentials and Use Case Identification

Besides an exploratory analysis of the current data base, moderated workshops with management and staff can greatly facilitate the process of finding a suitable match between the individual company's goals and business model, the available data and the capabilities of different machine learning techniques (see Figure 12). Results are not only a basis for strategic decisions on how to leverage the value of data in the respective business context, but frequently a nucleus for innovations in a broader sense as well.

Together with industry partners, Fraunhofer CML has conducted several studies and research projects on the possible applications of machine learning in maritime logistics. A good example is related to AIS data. The introduction of the Automatic Identification System was an important milestone of maritime digitalization. First and foremost, it had direct positive effects on maritime safety. However, by recording and storing worldwide vessel movement data on a daily basis, a unique maritime data set was established over the past years, which is associated with a multitude of potential use cases.

The key question is, how data-based techniques and machine learning can lead to innovative applications based on AIS data. The investigation carried out by Fraunhofer CML[1] in one example revealed a large number of different scenarios including automated vessel movement and traffic pattern analyses, automated assessment of traffic densities and frequencies and risk assessments of encounter situations with two or more ships. Moreover, data-based safety assessment for restricted waters such as shipping channels or fairways, anomaly detections regarding course deviations, leaving of traffic separation schemes and early detection of abnormal behavior to prevent collisions or groundings were identified as well. In consultation with experts from the maritime industry, vessel movement and traffic pattern analyses as well as assessment of traffic frequencies were selected as particularly promising and have since been addressed in follow-up projects.

1 The goal of Fraunhofer CML in the project TINA was to investigate potential data driven applications in shipping and to validate their feasibility. In addition to AIS data, weather and environmental data was used. The project received funding from the German Federal Ministry of Transport and Digital Infrastructure under project number 19F1043A.

Overall, workshops with relevant stakeholders are an essential part of identifying suitable machine learning use cases. They are used to specify the practical challenges for each considered scenario and evaluate potential benefits from data-based solutions. In preparation of the workshop, the general scope is defined and main technological trends as well as market developments are determined. During the workshop itself, different machine learning use cases in shipping and maritime logistics are identified with the help of creativity techniques. In order to prepare decision-making, these are documented in a consistent way and can be ranked by the participants regarding their practical potential.

The second starting point for use case identification is data. After identifying available data sources and databases, these are checked in terms of nature, scope and quality. Moreover, an exploratory and visual data analysis provides an initial understanding of the status quo and reveals important interdependencies. Both helps to determine to what degree the available data is suitable for machine learning and which use cases are promising based on given data sets. Together with different industrial partners, Fraunhofer CML has achieved good results with this collaborative approach that considers both the potential value-added and available data sources.

5.2 Proof of Concept

Once a promising use case for machine learning has been identified and described in terms of the desired outcome and business value, a proof of concept is a sensible next step. It examines whether it is feasible in practice to exploit a data-driven optimization potential with machine learning. A main question will usually be: is the available data actually sufficient in terms of quantity and quality. Further, it is important to test different data-based methods at this point in order to identify which machine learning technique archives the best results and if the performance meets minimum requirements for a practical application. In sum, a successful proof of concept creates the necessary basis for decision-making regarding a full-scale implementation and also allows an identification of main commercial and technical risks.

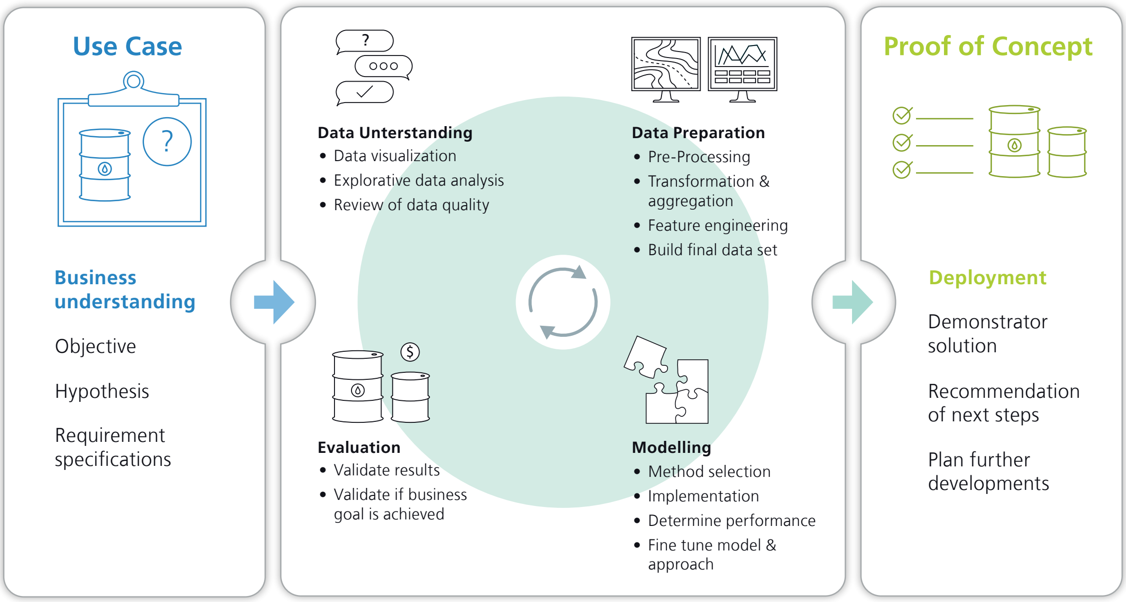

Figure 13: Proof of Concept for Machine Learning Solutions in Maritime Logistics

A well-established approach to a proof of concept begins with narrowing down the precise project scope and specifying the associated requirements (see Figure 13). Next, the data set is compiled and evaluated in detail regarding quality and usability for the particular research question. Often, data preparation followed by an initial analysis of the data represents a significant part of the overall workload associated with the proof of concept.

Once the data set is cleaned, pre-processed and aggregated, it can be used to train different machine learning models and compare their performance on a separate test data set. Subsequently, the achieved model accuracy is checked against minimum requirements for implementation in the concerned business process. Based on the results, recommendations for further action can be derived.

A first proof of concept example in which Fraunhofer CML used machine learning methods to improve maritime logistics concerns ship arrivals in port. The exact arrival and departure times of ocean-going vessels are an important factor for several port actors as this knowledge allows them to optimize their logistics and business processes. However, since compliance with official sailing schedules is low – a recent study finds only half of services arrive on time [33] – and available predictions inaccurate at best, precise capacity planning is difficult in many cases.

Against this background, it was expected that a data-based algorithm should be able to provide a better and non-discriminatory prediction of arrival and departure times for all interested port actors. The subsequent proof of concept study[2] carried out by Fraunhofer CML confirmed this. Technically, the chosen approach is based on a correlation of position and motion data from AIS with meteorological and hydrographic parameters such as wind, wave, current and water level using machine learning. Between 3 and 4 million AIS data records were processed daily to provide a forecast of up to 72 hours. Accuracy of predicted arrival times in all major German ports is in the range of +/- 1h on a time horizon of 24 h and +/- 3h when looking 72h ahead, while departure times can be determined correctly in 90 percent or cases. Both accuracy levels met requirements defined by stakeholders. Since then, the algorithms were developed further and implemented successfully in a number of similar problems and follow-up projects.

A second proof of concept example is associated with the assessment and minimization of ship emissions.[3] In this context, Fraunhofer CML is developing and verifying an approach for a quantitative estimation of CO2 emissions from vessels in European waters. Two sets of data are used to build a realistic emission assessment model. The first one contains publicly available data broadcasted via AIS (including movement of vessels, ship sizes and speeds) plus environmental data (ocean conditions and weather information). The second data subset covers ship-specific information related to main engine consumption and resulting emissions. Since this information is unknown or not readily available for most oceangoing vessels, a clustering approach in combination with a detailed resistance model is used to estimate the specific fuel consumption and emissions values for each ship. Machine learning and clustering algorithms are well suited in this context to identify classes with similar features and estimate missing values on ships main engines. Subsequently, opportunities to reduce emissions are identified with the help of an optimization model which provides benchmarks and suggests the most

2 In the VESTVIND project Fraunhofer CML developed a AIS-based real-time maritime traffic prediction tool with a forecast period of up to 72 hours.

3 In the project EmissionSEA Fraunhofer CML is working in a consortium to verify a concept of calculating CO2-emissions based on AIS data and other information with machine learning. The project receives funding from the German Federal Ministry of Transport and Digital Infrastructure under project number 19F2062C.

efficient routing alternatives (with the lowest emissions). The proof of concept therefore investigates a combination of machine learning and mathematical optimization.

5.3 Predictive Analytics

Decision making in maritime logistics often involves external, fluctuating factors that are stochastic in nature and therefore complicate planning and identifying well-considered actions. This underlines the importance of reliable forecasts for making business decisions under uncertainty. With machine-learning-based predictive analytics methods, forecasts can reach a new level of accuracy and reliable predictions for business-relevant influencing factors are achievable. The approach identifies correlations and patterns in historical data and uses these dependencies to predict future developments or events. In this way, both scope and accuracy of information presented to decision makers on a daily basis can be increased. In addition, time series forecasts, such as demand and freight volume predictions, arrival time forecasts, or remaining life time estimations, enable a better handling of uncertainties in operations.

Machine learning methods used for time series forecasting are perhaps one of the most interesting areas for data-based predictions in the coming years. First of all, this is due to the fact that predicting decision-relevant variables such as arrival times, demand values or freight volumes is quite interesting in many logistical contexts. Further, machine learning promises to be quite effective in cases where classical time series forecasting – which is already used extensively in supply chain management and logistics – reaches its limits. In particular, this refers to problems where multiple input variables and thus large amounts of data are available and complex nonlinear relationships between different parameters can be expected.

One example of how predictive analyses and machine learning can improve maritime logistics is associated with the arrival of trucks at container depots or ships at port terminals (see Figure 14). At logistical nodes such as port terminals, distribution centers or container depots, unforeseen waiting times are quite common if trucks deliver or pick up containers and goods. In addition, the duration of loading and unloading operations at the ramp also varies in many cases. These uncertainties prevent an optimal disposition of trucks by forwarding agents and trucking companies.

However, even though waiting times differ and seem unexpected, they are usually far from unpredictable. As past projects of Fraunhofer CML have shown, advanced forecasting techniques prove very capable in predicting truck arrivals and waiting times.[4] Both enable trucking companies to make better dispatching and routing decisions and also helps terminals to plan sufficient equipment and personnel to meet demand and cope better with peaks.

A second example how machine-learning-based predictions can lead to significant improvements is related to ship operations. Due to cost advantages and reduced risks of unexpected breakdowns, conventional time-based preventive maintenance is increasingly replaced by condition based or even predictive maintenance strategies. In condition-based maintenance systems, sensors continuingly monitor the equipment (e.g. temperature, vibration, acoustic emission) and an algorithm automatically

4 Fraunhofer CML and a Hamburg based container depot and service provider cooperated to predict the utilization of an empty container depot in the project LILIE.

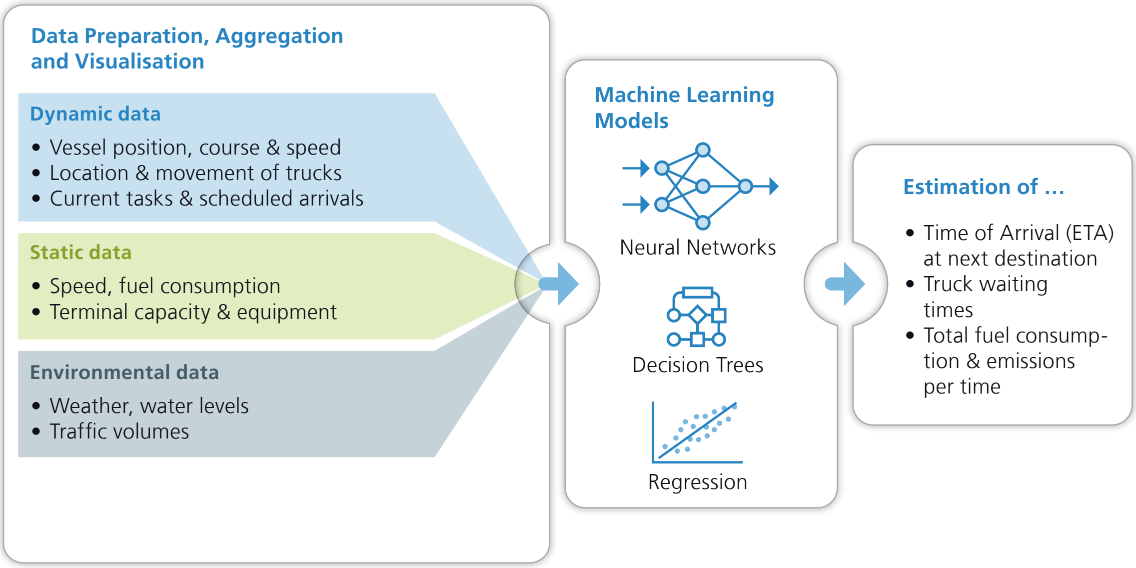

Figure 14: Reliable Forecasts in Maritime Logistics

produces a condition assessment from this data. Predictive maintenance goes one step further by also predicting the remaining lifetime of the equipment. In both cases, machine learning algorithms are used very successfully if a sufficiently large amount of data is available. Based on the experience with using machine learning for predictive maintenance, Fraunhofer CML has developed a framework that guides the development of algorithms for innovative maintenance strategies for maritime systems. [34]

5.4 Optimization and Machine Learning

As industrial processes often evolved through experience over time, they are sometimes characterized by "operational blindness" and inefficiencies. Against this background, operations research has proven to be effective in optimizing operational logistic decisions, e.g. at container terminals, for crane scheduling, or in ship routing and fleet management. Typical methods applied to solve such decision problems are (meta-)heuristics, priority rules, or simulation models. With the help of machine learning there are different ways to optimize industrial processes even further building upon and extending classical operations research methods.

First and foremost, this concerns integrating machine learning into traditional optimization algorithms. For example, traditional optimization models are largely influenced by the quality of input data. Utilizing the predictive power of machine learning to provide more accurate input data for multivariate optimization of complex processes helps finding better solutions and thus improving efficiency. As mentioned before, truck and vessel arrival times are important input variables for planning and optimizing port and terminal operations. A machine-learning-based prediction model – which provides fast and accurate arrival time or freight volume estimations – can be integrated into control algorithms and decision support systems of the entire terminal to improve overall planning and control process. Same holds true for the prediction of other relevant influencing factors, including the availability of and demand for empty containers[5] or expected container dwell times at the terminal.

5 In the research project C-TIMING Fraunhofer CML is using machine learning methods to forecast empty container demand and supply worldwide. The project receives funding from the German Federal Ministry of Education and Research under project number 01IS20004B.

The more precise such parameters are estimated using machine learning, the more efficient and seamless can handling and transshipment take place.

In the aforementioned examples, machine learning models were used to provide traditional planning algorithms with better input data. In a different setup, machine learning can also support optimizing maritime operations directly. For instance, it can be utilized to go through large amounts of past routing, weather and hydrodynamic data in order to learn the underlying patterns between environmental conditions and the best route for such conditions. Thus, the machine learning model learns to identify the optimal course for given hydrodynamic states and can use this relationship to recommend routing alternatives to the ship's crew. These provide different options characterized by performance measures such as travel time, fuel consumption, emissions and safety criteria.

Likewise, similar approaches of machine-learning-driven optimization can be used in port and terminal operations or support planning along maritime supply chains. In this context, machine learning algorithms are responsible for automatically evaluating the increasing amount of data from maritime transport and make it available as valuable contextual information that extend current planning horizons. By analysing vessel-specific information in combination with shipping markets data, self-learning digital assistance systems can thus provide enhanced visibility and forecasts to support decisions of forwarders, shipping companies, or terminal operators alike.

5.5 Deep Learning Models

Digital images, text, or audio files contain immense informational value, which is easy to access by humans but difficult for computers to comprehend. Currently, this changes as deep learning enables machines to understand previously inaccessible information of digital content. As a consequence, work processes which currently contain manual steps carried out by humans – such as tracking and inspection of cargo units – can be automated in future with associated advantages of improved efficiency, speed, or accuracy.

A first example for utilizing this kind of artificial perception in the context of maritime logistics is computer vision (see Figure 15). Modern computer vision applications based on deep learning are able to automatically identify and segment defects such as coating failures, corrosion, or structural damage on a ship hull or cargo by analyzing images captured in dive through portals or by remote inspection

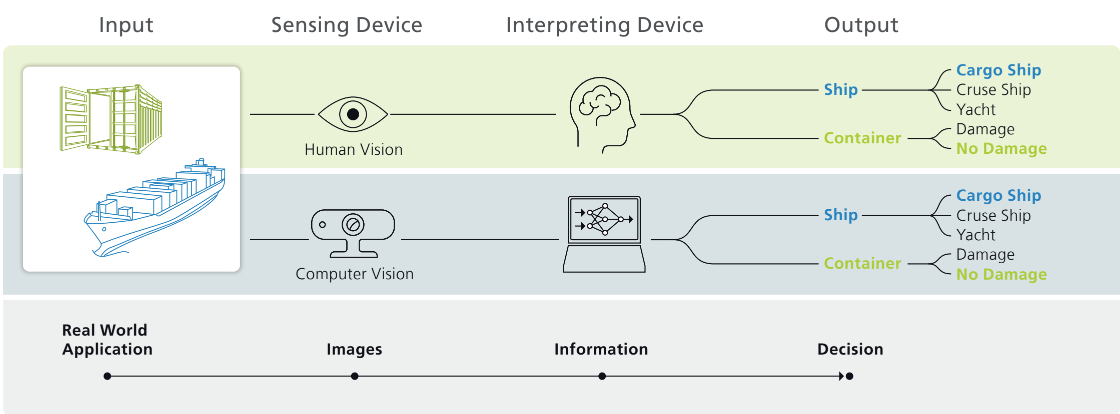

Figure 15: Application of Computer Vision in Maritime Logistics

technologies such as drones. Fraunhofer CML is currently working on a practical application of computer vision related to damage assessment in context of empty container handling.[6] Visual inspection of freight containers en-route helps to ensure that they are in a suitable condition for loading and safe transport. As this process is currently done manually, it can be both error prone and time-consuming. An automation of the inspection process that makes use of the capabilities of computer vision could increase the efficiency of inspection processes at the gate and also improve the ability to plan subsequent terminal processes.

Other possible application of computer vision in maritime logistics include identification and localization of objects or optical character recognition. Together, these can support an optimization of operations in warehouses or on cargo terminals by automatically identifying and localizing transport or cargo units through image-based detection. Moreover, computer vision models can be used to automatically process large numbers of cargo and transport documents with routine inputs in order to reduce both costs and workload. Besides a comparatively simple digitalization of text or recognition of objects, deep learning is also able to extract semantic information from text, such as identifying the context of a document or the intent of an e-mail written by a customer.

Lastly, innovations such as maritime autonomous surface ships would not be feasible without environmental perception provided by computer vision. By processing data from different sensors computer vision models are capable of identifying obstacles and more generally perceive the environment of an autonomous vessel and changes within it. Tracking objects like ships in proximity through cameras or radar will also support human lookout in navigation and collision avoidance as an intermediate step towards full autonomy.

Speech recognition and natural language processing are already widely used in today's electronic communication products such as phones or home devices and also another area where deep learning models can have an impact in shipping and maritime logistics. A concrete example concerns radio communication on the bridge, which can be difficult in a marine context as background noise, harsh environmental conditions or dialects interfere. This is not only a nuisance but can lead to mistakes that affect maritime safety. To help prevent this in the future, Fraunhofer CML is working in a project[7] which aims at automatically transcribing maritime VHF communication using deep learning models. This speech recognition system is supposed to make voice messages easier to understand, dissolve linguistic ambiguities and provide a chronological documentation of what was said, which can be of particular benefit in emergency situations.

Beyond the mentioned use case, there are many more promising applications of natural language processing in maritime logistics. For instance, voice- or chatbots could handle certain inquiries automatically. Moreover, machine learning offers an analytics toolbox for the interpretation of unstructured text data in emails, contracts, bills of lading, or invoices. With the help of these technologies freight documents could be handled automatically, e.g. to identify anomalies that require special attention or extract and process certain information. Going one step further, machine learning also offers ways to fully automate certain repetitive administrative and clerical tasks associated with document

6 In research project COOKIE Fraunhofer CML is working on image-based damage detection on containers with computer vision. The project is supported by the German Federal Ministry of Transport and Digital Infrastructure under project number 19H19006B.

7 The ARTUS project Fraunhofer CML works on an automatic transcription solution for maritime radio communication. It receives funding from the German Federal Ministry of Education and Research under project number 13N15019.

handling and information transfer. In this way, natural language processing provides the necessary input for more complex tasks like interpretation and decision making.

Overall, deep learning is considered as one of the key technologies in AI with much potential for future business gains. With its partners, Fraunhofer CML committed to develop innovative deep learning solutions for a growing number of applications in maritime logistics to enable process automation and optimization.

5.6 Reinforcement Learning

Where even a combination of optimization or control engineering methods with machine learning algorithms reaches its limits, reinforcement learning can be a promising approach to deal with particularly complex real-world problems. In cooperation with industry partners, Fraunhofer CML is currently working on a number of examples that use reinforcement learning in maritime logistics.

In a reinforcement learning approach, an "agent" (i.e. an autonomous software module) carries out a specific task that requires sequential decisions on what action to take in a given state (see Figure 16). The term "reinforcement" refers to a trial-and-error learning method, where a reward mechanism for advantageous combinations of actions taken and the feedback provided by the observation of the environment – in other words a "reward" – is required. The aim is to learn a strategy or "policy" that maximizes the expected long-term reward by suggesting the best action to take for every state the agent can experience.

Reinforcement learning is able to learn complex processes autonomously by experimenting without requiring big data sets with the "right" answers. It has been successfully applied to complex problems like robots learning to walk, computers beating the world champion in the strategy game Go, or flying a helicopter autonomously. [35] As many decisions in maritime logistics involve complex problems with sequential, or recurrent, decisions that enable experimenting, e. g. the assignment of containers

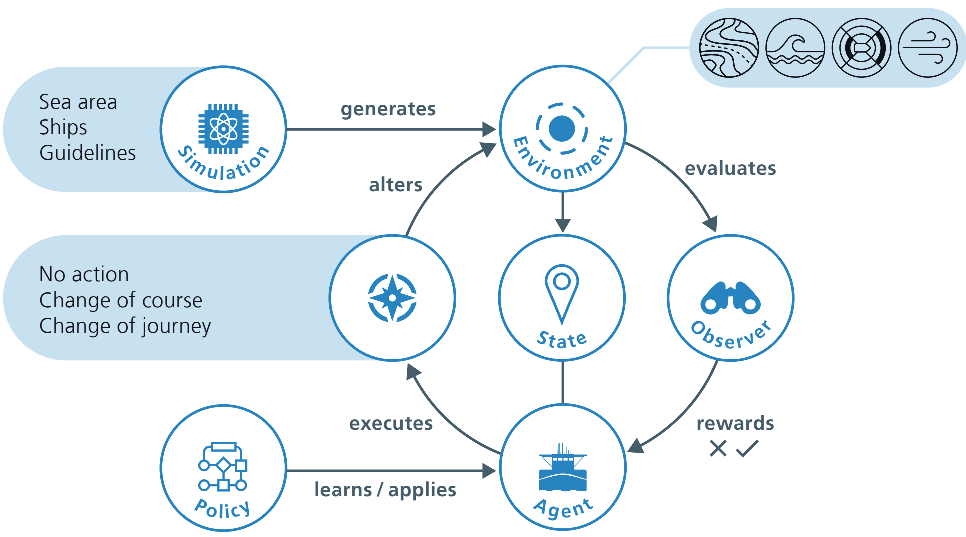

Figure 16: Reinforcement Learning Approach to Maritime Routing Problems

to equipment in a terminal or adjusting a vessel's course throughout its journey, it is worth to consider reinforcement learning as a promising solution method.

Not all problems are equally suitable for reinforcement learning approaches. To determine where reinforcement learning is likely to produce good results, the following factors are indicators:

- An optimal action or response within a complex control problem needs to be identified.
- Traditional (engineering) methods are of limited use to identify this optimal action or response.
- It is possible to apply the principle of trial and error (feedback loop) in a sequential manner.
- Short processes with a high number of repetitions are particularly promising.
- It is advantageous if the process can be executed in a simulated environment so that an action can be evaluated in terms of its impact on the system, without the training of the optimization algorithm interfering with actual real-world processes.
- A simulation model and offline learning can be indispensable where real-world trials cause high cost or can compromise process safety.

A first example from maritime logistics using reinforcement learning is related to the (partially) autonomous navigation of ships. The demand for smart and autonomous navigation assistance systems in maritime transport is growing. Due to its inherent ability to solve complex and "hard to describe" problems in changing environments, reinforcement learning is well suited for providing nautical decision support.[8] Similar ideas apply to autonomous guided vehicles in other maritime logistics use cases such as terminal and intra-port transportation.

One possible approach is to train a reinforcement learning algorithm to anticipate collisions and groundings for nautical situations with multiple participants or to avoid dangerous harsh weather within a defined set of constraints. Moreover, a system based on reinforcement learning methods can be trained to show "normal" navigational behaviour. These simulated normal operations can be compared with actual behaviour of individual ships (e. g. route, course and speed) in order to identify critical and anomalous situations and thus increase safety at sea. To this end, sensor data – e. g. from AIS, radar and positioning systems – are used to parametrise a simulation environment. The simulation approach allows the modelling of any number of scenarios, in which even small changes in environmental parameters can be represented. Subsequently, the environment is used to train a reinforcement learning agent within a multitude of realistic nautical scenarios. Over a large number of trial and error iterations that reward good behaviour, the agent learns to take adequate actions (combination of course change, speed change, or no action) and thus adjusts the predefined route to ensure a safe passage. [36]

Besides assisted navigation and autonomous vehicles, a second promising area of using reinforcement learning is process control on container terminals and in ports. Possible use cases – which can be assessed based on the above-mentioned criteria – are of a very different nature. Initial studies show, for example, that reinforcement learning can be used to optimize a ship's stowage plan by automatically allocating containers to slots and thereby minimizing reshuffling and yard crane shifts.

8 The research project B ZERO promotes the development of a partly autonomous navigation system which serves as a link between manned and autonomous navigation. It receives funding from the German Federal Ministry of Economic Affairs and Energy under project number 03SX500A.

[37] Another promising example is yard crane scheduling, where the sequence of drayage trucks served by a crane is specified by a reinforcement learning algorithm with the objective of minimizing waiting time. Decisions taken by the reinforcement learning agent in this case define which truck to serve next, based on e. g. the distances to the crane and the current waiting times. [38] Beyond that, reinforcement-learning-based methods can also support routing decisions. An interesting example covers routing of automated guided vehicles in a guide-path network. The method aims at finding the shortest-time route on the terminal instead of the shortest-distance route for each delivery by considering the congestion at intersections and bidirectional path segments. Experiments for this use case show that travel times could be reduced successfully compared to traditional solutions that rely on shortest-distance routes. [39]

5.7 Hybrid and Grey Box Models

Due to the complexity of optimization problems in logistical and maritime applications it can be useful to combine different approaches, methods and algorithms. In addition, the integration of experience and domain knowledge of experts into data-based algorithms often helps to improve the results by a large extend. Such holistic approaches are known as "hybrid and grey box models". They try to combine different methods in a way that weaknesses of individual algorithms are compensated while the respective strengths – e. g. quality, computation time, robustness, or transparency – are bundled to support the overall use case.

Machine learning algorithms generally rely on a completely non-parametric approach, known as the "black box" character of machine learning. In contrast, traditional optimization techniques – including mathematical modelling and simulation – are based on parametric models and therefore summarized under the term "white box". "Grey box" methods are hybrid models combining the strengths of both approaches (see Figure 17). To that end, they make use of both qualitative or quantitative knowledge of domain experts and large data sets in order to exploit all relevant information for a given decision problem. Moreover, they also represent a solution to overcome the "black box" character of machine learning by making model predictions more explainable and thus increase their acceptance by practitioners.

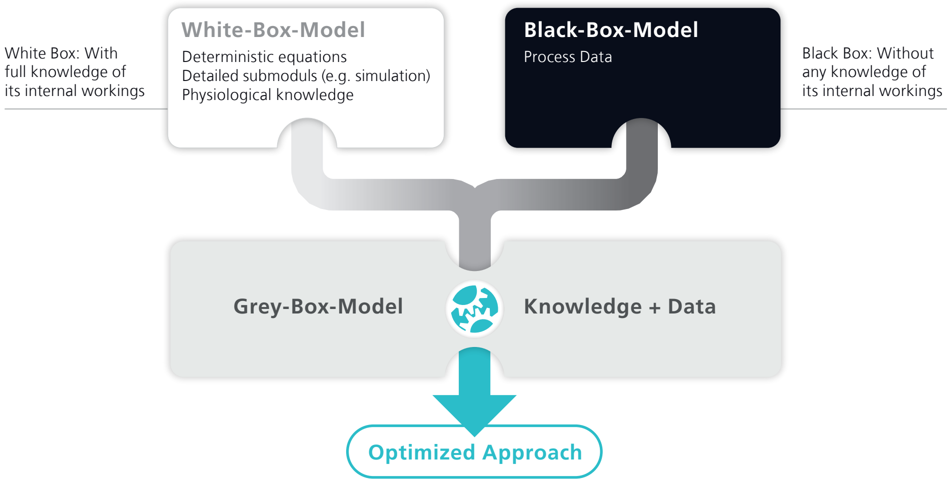

Figure 17: Combination of Knowledge and Data-based Insights in Grey-Box-Models

Visualization has proven to be a very useful technique in gaining user acceptance of data-based decision support systems. For example, intelligent software solutions that visualize the energy intake of all consumers over a period of time, e. g. on a port terminal, make it easier to develop suitable measures for detecting operations with unnecessarily high consumption. To this end, power consumption can be compared with benchmark values to identify extreme and unnecessary high energy use. However, average values will not reflect the dynamic pattern of energy consumption typical for port operations adequately. This requires forecasting of the expected load profiles for different operational situations e. g. ship arrivals and associated cargo handling activities.

In this context, a hybrid approach is ideal, since it allows an evaluation of energy consumption data with machine learning algorithms and also incorporates domain expertise about the loading and unloading processes. In this way, machine learning can complement the individual knowledge of experts. As a result, an expected power consumption calculated with a hybrid model is a more dynamic and thus accurate reference point for energy management and a decision support system designed on that basis can contribute to a long-term reduction of energy costs and CO2-emissions in ports.

Another hybrid approach is associated with classical optimization methods – like metaheuristics and simulations – which can be time consuming when applied to real-world data and problems. With the help of machine learning, it is possible to replace some particularly time-consuming tasks of those classical methods by machine learning model predictions to speed up the computation. For example, artificial neural networks can be used to select scheduling rules in real time based on the current system's state. Looking at a yard crane or vehicle scheduling problem, this could mean that a machine learning model chooses the respective priority rule (e. g. based on waiting time or minimal distance) that determines the selection of the container to be loaded/unloaded or transported next. Moreover, machine learning can also be used to optimize hyperparameters of (meta-)heuristics or to estimate system performance measures for different operational strategies instead of executing a simulation run.

Lastly, there are a number of interesting maritime operation applications which combine simulation and machine learning in a hybrid model setting. Such an approach, for instance, could help to increase safety in marine navigation. Port calls are as demanding as they are risky, since limited fairways and turning points are less fault-tolerant than operations on the open sea. Combining machine learning with simulation in a hybrid model and training this model on AIS-based movement patterns can serve as the basis for a digital assistant which proposes obstacle-free and safe routes for each vessel. In addition, information fusion and automatic evaluations achieved in this way can contribute to a more comprehensive understanding of the current traffic situation.

6 OPPORTUNITIES FOR COOPERATION WITH FRAUNHOFER CML

Fraunhofer CML supports companies in closing the gap between theoretical advances in maritime digitalization and practical application of data-based methods. CML's independent consulting considers the respective company status and develops an individually suitable path towards becoming a data-driven organization. Fraunhofer CML has substantial experience in building machine learning solutions, especially for data-based forecasting and decision support systems in maritime logistics.

The CML offers comprehensive support, initial feasibility studies, concrete proof of concept and first applications to lasting implementations in the form of modular use case specific applications and holistic decision support systems. The process of integrating machine learning into practical applications involves a high degree of customization. Solutions are specifically tailored to the requirements of the maritime industry and individually adapted to the needs of the customer. The range of professional services offered by Fraunhofer CML in this context includes the following:

Level 1: Getting startet

WHO

Companies with

- Low level of data utilization
- Limited experience with machine learning so far
- Limited data science expertise in-house

WHY

- Benefit from low entrance barrier and quick but evidence-based results
- Gain knowledge from available data
- Determine how machine learning can be used most effectively

WHAT

- Facilitate the identification of company specific opportunities for data-driven decision making and the definition of associated strategic goals
- Identify existing data sources and databases; transfer, harmonize and integrate them to build a common basis for data-driven solutions
- Review quality of available data sets to verify their applicability and validity for machine learning models and intelligent assistance systems
- Conduct descriptive and explorative data analyses for knowledge enrichment
- Identify the most promising use cases for machine learning in a collaborative approach taking into consideration both potential value-added and available data sets

Level 2: First Applications

WHO

Companies with

- First experience with data based solutions
- Sufficient data for testing first use cases
- Interest in benefiting quickly from proof of concepts

WHY

- Explore possibilities of machine learning with minimal financial expenditure
- Use demonstrators to interact with customers
- Utilize machine learning as a future competitive advantage

WHAT

- Develop and implement interactive dashboards for real-time data visualization covering everything from ship routes to selected key performance indicators
- Carry out proof of concept studies for selected machine learning use cases in maritime logistics
- Develop demonstrator solutions and give recommendations on next steps
- Identify innovative products and services that leverage the value of data and design related business models
- Estimate business value of machine learning solutions both in terms of operative improvements and return on investment
- Recommend a suitable information architecture which supports machine learning and data-driven decision making

Level 3: Lasting Realization

WHO

Companies with

- High data maturity level and information architecture in place
- Ideas, which machine learning solutions should be developed
- The intention of gaining a partner to implement their AI strategy

WHY

- Fully exploit the possibilities of machine learning in your organization
- Introduce digital products and business models based on machine learning
- Benefit from the scientific excellence and maritime industry know-how of Fraunhofer CML

WHAT

- Develop and implement machine learning solutions in a company specific context and integrate them with legacy systems
- Build customized data-based prediction models to support decision making in real-time, including arrival times, demand values or freight volumes
- Integrate machine learning modules in traditional optimisation solutions, e. g. for adaptive weather routing or improved terminal planning
- Apply reinforcement learning techniques to solve complex real-world problems, e. g. for nautical assistance systems or automated decision making
- Access the information in images, text, or audio files with deep learning models, e. g. to automate tasks or simplify document processing
- Develop hybrid machine learning approaches to exploit both the extensive knowledge of employees and experts as well as insights from large data sets
- Implement grey box models to make machine learning predictions explainable and thus increase their acceptance by practitioners


ML

7 AUTHORS

Dipl.-Ing.oec. Lutz Kretschmann
Head of Team »Marine Operation Management«

Expert for maritime logistics, digital business models and predictive analytics

Dr.-Ing. Miriam Zacharias
Senior Scientist

Expert for optimization, reinforcement learning and hybrid models

Steffen Klöver, M.Sc.
Machine Learning Specialist

Expert for deep learning, computer vision and data analytics

Tina Hensel, M.Sc.
Head of Team »Nautical Solutions and Data Science«

Expert for AIS-data analytics and data science, nautical assistance systems and data visualization

8 REFERENCES

[1] L. Kretschmann, F. Deichmann and C. Jahn, "White Paper »Ship Operation 4.0«," Fraunhofer Verlag, 2019.

[2] Fraunhofer CML; VDMA, "Digitalisierung des Service in der maritimen Zulieferindustrie," 2017.

[3] M. Wingrove, "Satellite investment boosts cruise ship connectivity," 24. 05. 2020. [Online]. Available: https://www.rivieramm.com/news-content-hub/news-content-hub/satellite-investment-boosts-cruise-ship-connectivitynbsp-58659. [Accessed 07. 09. 2020].

[4] International Telecommunication Union, "Technical characteristics for an automatic identificatio nsystem using time division multiple access in the VHF maritime mobile frequency band; Recommendation ITU-R M.1371-5," M Series; Mobile, radiodetermination, amateur and related satellite services, Genua, 2014.

[5] F. Lázaro, R. Raulefs, W. Wang, F. Clazzer and S. Plass, "VHF Data Exchange System (VDES): an enabling technology for maritime communications," CEAS Space Journal, no. 11, pp. 55-63, 2019.

[6] V. Carlan, C. Sys, A. Calatayud and T. Vanelslander, "Digital innovation in maritime supply chains: experiences from Northwestern Europe," IDB, 2018.

[7] F. Parola, G. Satta, N. Buratti und F. Vitellaro, „Digital technologies and business opportunities for logistics centres in maritime supply chains," Maritime Policy & Management, pp. 1-17, 2020.

[8] UNCTAD Policy Brief No. 75, Digitalization in Maritime Transport: Ensuring Opportunities for Development, UNCTAD/PRESS/PB, 2019.

[9] K. Brümmerstedt, R. Fiedler, V. Flitsch, C. Jahn, H. Roreger, B. Sarpong, S. Saxe and B. Scharfenberg, Digitalization of Seaports – Visions of the Future, Hamburg: Fraunhofer Verlag, 2017.

[10] Sea Traffic Managmenet Validation Project, Final Report, 2019.

[11] IDG Research Service, "Studie Machine Learning / Deep Learning 2019," 2019.

[12] YOUSE, EHI, DFKI, "Perspektiven der Künstlichen Intelligenz für den Einzelhandel in Deutschland," 2019.

[13] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 3rd Edition, Upper Saddle River: Pearson, 2010.

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Frei-Frei, "ImageNet Large Scale Visual Recognition Challenge," Int. J. Comput. Vis. 115, pp. 211-252, 2015.

[15] W. Xiong, L. Wu, F. Alleva, J. Droppo, X. Huang and A. Stolcke, "The Microsoft 2017 Conversational Speech Recognition System," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5934-5938, Calgary, AB, 2018.

[16] F. Chollet, Deep Learning with Python, Shelter Island: Manning Publications Co., 2017.

[17] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, Cambridge: MIT Press, 2016.

[18] C. M. Bishop, Pattern Recognition and Machine Learning, New York: Springer Science+Business Media, 2006.

[19] A. C. Müller and S. Guido, Introduction to Machine Learning with Python, Sebastopol: O'Reilly Media, Inc, 2017.

[20] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, Cambridge: The MIT Press, 1998.

[21] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik and e. al., "Grandmaster level in StarCraft II using multi-agent reinforcement learning," Nature 575, p. 350–354, 2019.

[22] S. Bonnell, "Inc.com," 22. 01. 2018. [Online]. Available: https://www.inc.com/sunny-bonnell/how-to-follow-your-instincts-in-business-even-when-people-say-youre-crazy.html. [Accessed 07. 08. 2020].

[23] Capgemini, Economist Intelligence Unit, "The Deciding Factor: Big data and decision-making," [Online]. Available: https://www.capgemini.com/wp-content/uploads/2017/07/The_Deciding_Factor__Big_Data___Decision_Making.pdf. [Accessed 07. 08. 2020].

[24] J. Finnerty, "The fabricator," 30. 09. 2008. [Online]. Available: https://www.thefabricator.com/thefabricator/article/shopmanagement/automated-collection-of-real-time-production-data. [Accessed 07. 08. 2020].

[25] A. Dichter, M. Rothkopf, F. Bauer, M. Bäuml, J.-F. Rösch, C. Bärwind, M. Schuster, M. Lommer and H. Friedrich, "Travel and logistics: data drives the race for customers," McKinsey & Company: Travel, Transport & Logistics, May 2018.

[26] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," Big data, vol. 1, no. 1, pp. 51-59, 2013.

[27] bitkom, "Digitalisierung der Logistik," Dr. Bernhard Rohleder, Berlin, 2019.

[28] BMWi, „Einsatz von Künstlicher Intelligenz in der Deutschen Wirtschaft,“ Berlin, 2020.

[29] R. J. Hyndman, "Exploring the boundaries of predictability: What can we forecast and when should we give up?," Sunnyvale, California, 2015.

[30] M. Gilliland, L. Tashman and U. Sglavo, Business forecasting: Practical problems and solutions, Hoboken, New Jersey: John Wiley & Sons, 2016.

[31] P. M. Catt, "Entropy as an A Priori Indicator of Forecastability," Department of Computing and Information Technology, UNITEC, Auckland 1025, New Zealand.

[32] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," Journal of Data Warehousing, vol. 5, no. 4, 2000.

[33] S. Sundboell, "container-news.com," 17. 01. 2020. [Online]. Available: https://container-news.com/zim-tops-liner-shipping-punctuality-list/. [Accessed 11. 11. 2020].

[34] J. Poschmann, "Zustandsbasierte Instandhaltung in der Schifffahrt," Schiff & Hafen Nr. 8, August 2017.

[35] JAXenter Editorial Team, "jaxenter.com," Machine Learning Conference session Dr. Christian Hidber: Reinforcement learning: a gentle introduction and industrial application, 30. 09. 2019. [Online]. Available: https://jaxenter.com/reinforcement-learning-2-161473.html. [Accessed 11. 11. 2020].

[36] C. Chen, X. Q. Chen, F. Ma, X. J. Zeng and J. Wang, "A knowledge-free path planning approach for smart ships based on reinforcement learning," Ocean Engineering, vol. 189, p. 106299, 2019.

[37] Y. Shen, N. Zhao, M. Xia and X. Du, "A deep q-learning network for ship stowage planning problem," Polish Maritime Research, vol. 24(s3), pp. 102-109, 2017.

[38] F. Fotuhi, N. Huynh, J. M. Vidal and Y. Xie, "Modeling yard crane operators as reinforcement learning agents," Research in transportation economics, vol. 42(1), pp. 3-12, 2013.

[39] S. M. Jeon, K. H. Kim and H. Kopfer, "Routing automated guided vehicles in container terminals through the Q-learning technique," Logistics Research, vol. 3(1), pp. 19-27, 2011.

www.cml.fraunhofer.de