Norwegian School of Economics
Bergen, Autumn, 2018

Predictive Power of AIS on Marine Insurance:
A demonstration of how activity level and operational patterns of the merchant fleet can be used to predict P&I insurance claims using machine learning

Eirik Møller & Jonas Sletten
Supervisor: Walter Pohl

Master's in Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Preface

This thesis is written to fulfil our graduation at the Norwegian School of Economics. The process has been incredibly motivating and provided invaluable insight into the potential and limitations of machine learning. The thesis is written in cooperation with Safetec and Skuld, and we are grateful of all the support both companies have provided throughout the process.

The work aims at investigating the potential for machine learning in P&I insurance, by combining AIS-information with data on insurance claims. The objective has been to provide Skuld with background information on machine learning and insight into the quality and limitations of AIS-information and to provide Safetec with new perspectives on how to implement AIS-analysis in risk management. The extensive section on AIS and pre-processing as well as the comprehensive discussion on further work is a result of this.

We want to thank Andreas Sylthe and Otto Rendedal at Skuld for daily support throughout the process, and we are grateful for the extensive resources Skuld has put aside for this research. We would also like to thank Asbjørn Lein Aalberg and Peter Ellevseth at Safetec for valuable inputs and pointing us in the right direction along the way. In addition, a significant number of employees at both Skuld and Safetec has contributed with information and support, and the quality of the thesis would not have been the same without these contributions. Last, we would like to thank our supervisor Walt Pohl for valuable input and support, and for ensuring that we always had our focus on the most critical issues.

Bergen, 19.20.2018

Eirik Møller

Eirik Møller

Jonas Sletten

Jonas Sletten

Abstract

Digitalisation is making a growing appearance across all sectors, and traditional P&I insurance is no exception. Marine insurance is said to be several years behind traditional land-based insurance when it comes to digitalisation. This thesis is attempting to narrow the gap, by investigating the potential of applying machine learning on AIS-information against the extensive database on P&I insurance claims from the P&I Club Skuld.

The thesis aims at investigating the potential to predict P&I insurance claims based on variables retrieved from AIS. AIS-information from 2013-2017 and Skuld's claims data for the same period was combined, and a total of five machine learning methods were tested to assess the predictive power of AIS-information. An extensive pre-processing was executed to make the data available for machine learning, and this section provides detailed information to anyone that aims at utilising AIS in their research.

The research finds that AIS-information has predictive power for claims, as it links claims to activity level and operational patterns of the merchant fleet. The findings have implication for two fields in marine insurance; risk assessment/ pricing and loss prevention. In relation to loss prevention; average distance sailed, number of unique ports visited, and total distance sailed were found to have the most predictive power. Regarding risk assessment, the strongest model was able to predict 79 % of all cargo claims for Bulk & Cargo small.

The research has revealed that machine learning has potential to create significant value in P&I insurance and that an extensive amount of data is ready to be applied in the pursuit of more accurate risk assessments and more precise loss prevention measures. Estimates vary between a potential yearly reduction in claims of 7-14%, in addition to increased revenue as a result of correct pricing.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIS	Automatic Identification System
AB	AdaBoost
B&CS	Bulk Cargo Small
GNSS	Global Navigational Satellite System
GPS	Global positioning system
H&M	Hull & Machinery
IG	International Group
IMO	International Maritime Organization
KTS	Knots – nautical miles per hour
NB	Naive Boyes
NM	Nautical miles (1nm = 1852 meters)
NMIP	Nordic Marine Insruance Plan
NMIP	Nordic Marine Insruance Plan
NNET	Neural Network
OOW	Officer of the watch
P&I	Protection and indemnity
RDC	Running Down Clause
RF	Random Forest
SIS	Signal in Space
TEU	Twenty-foot equivalent unit
USD	United States Dollar

Contributors

This thesis is written in cooperation with Skuld P&I Insurance Club and Safetec Nordic AS. Skuld, and Safetec has defined the minimum limits of the scope and provided all available data. In addition, both companies have contributed with expertise within their domain.

Skuld

“Skuld is a world leading marine insurance provider with a strong financial position and an 'A' rating with Standard & Poor's. We cater to the needs of shipowners, the offshore and energy sector, charterers and traders, ports and terminals, cargo and the superyacht community.

Skuld was established in Oslo in 1897 as a P&I club for Scandinavian shipowners. Since then, we have grown into a diversified marine insurer and now offer bespoke covers through Skuld P&I, Skuld SMA and Skuld 1897, a syndicate at Lloyd's.”

(Skuld, 2018)

Safetec

“Safetec is a leading provider of risk management services. With an integrated approach to technical, human and organisational aspects, Safetec is a full range supplier of services within risk, reliability and emergency preparedness. Safetec has provided services to worldwide offshore, maritime and land-based industries since 1984.”

(Safetec, 2018)

1 Introduction

Marine insurance is paramount to secure efficient shipping markets and allow shipowners and charters to take on the necessary risk to ensure global trade (Ajigboye, 2016, UNCTAD, 1982). A single accident could lead to bankruptcy for shipowner's, if not for well-functioning marine insurance industry. The cost of replacing a vessel perished at sea only constitutes for a fraction of the possible liabilities shipowners face. Claims related to pollution, loss of lives, wreck removal and cargo claims can add to many times the value of the vessel.

The current marine insurance market is dominated by three insurance products; Hull & Machinery (H&M), Cargo and Protection & Indemnity (P&I). H&M and Cargo belong to the traditional marine insurance products with origins dated back to the 14th century. P&I insurance developed later, and primarily through the mid-19th century in London, as to counter third-party liabilities. Today thirteen P&I clubs hold liability for 90% of the world's ocean-going tonnage (IG, 2018). The clubs are in close cooperation and mutually responsible for each other's claims through the International Group of P&I Clubs (IG). At the same time, they compete among each other, in a market with a limited number of vessels and operators. Mutuality makes it challenging to compete on price, and other elements such as service and support are essential factors in the competition to attract shipowners.

The primary driver of cost in marine insurance is the frequency and severity of marine accidents, and damage of equipment or cargo, while under maritime transportation. Accidents have the potential to create environmental disasters, loss of life and tremendous financial loss. Accidents are defined as unfortunate incidents that happened unexpectedly (Oxford Dictionaries, 2018) and are unpredictable by nature. However, research reveals that accidents have common factors, and with growing available data, further patterns might arise.

Accurate risk assessment is an essential part of any marine insurance company to attain competitive advantage. Increased understanding of the contributing factors of marine insurance claims, can be achieved through in-depth data analyses. In the last decade, computer power has increased rapidly along with the amount of available data. Approximately 50,000 merchant vessels trade internationally (AGCS 2018), and transmits on AIS continuously. The market for analysing AIS-information is expected to reach $ 225 million in 2020 (Markets&Markets, 2014), and the potential values for companies involved in shipping are likely higher. In general, shipping lags in new tech adoptions compared to other sectors, but changes are inevitable (Latarche, 2018), and this research aims at narrowing the gap.

Most research on marine accidents is related to collisions, groundings, machinery breakdown, fire on board and personal injuries. These categories constitute for a large part of P&I insurance claims. However, the largest category of insurance claims is related to cargo, and the vessels does not need to have taken any damage. This research contributes to expanding the scope of current research, to not only describe accidents at sea and also explain any claim related to vessels activity and operation.

Standard & Poor's rating report for Skuld (2018) explicitly identifies that the P&I structure of the marine insurance industry has high barriers to entry. They assess the risk of new competitors to enter the industry as low, and the overall assessment of Skuld gives the company an A rating. Further, Standard & Poor's evaluate a downside scenario where Skuld is unable to maintain its strong competitive position. One key factor to sustain the current industry position, is to develop successful growth by taking full advantage of the existing information obtained in the company and create new solutions to increase the effectiveness of operations.

The purpose of this thesis is to analyse the potential of exploiting a combination of AIS-information from Safetec and claims data from Skuld. More specifically it aims at utilising machine learning to create a model for predicting insurance claims in three categories; Cargo, Contact & Collisions and Human. Machine learning has the potential to analyse large amounts of data to retrieve patterns that are not possible to detect for any person, and findings can be used to develop more accurate risk models.

Increased understanding of the relationship between vessels operational pattern and claims, has the potential to create value in three fields of marine insurance; loss prevention, pricing, and member support. Loss prevention is the field of insurance aimed at avoiding accidents before they happened. If the underlying factors behind an insurance claim are identified, monitoring these factors can potentially lead to the right measures being taken, and accidents can be avoided. The insight of underlying factors might also be used to adjust risk models, and implement new pricing strategies. Current price models in the P&I markets are based on a variety of factors, but to the knowledge of the researchers, activity level is not one of them. A possible future scenario is activity-based pricing as opposed to current static models.

1.1 Problem Definition

The objective of this thesis is two folded; examine the potential for machine learning and AIS-information in relationship to P&I insurance, while specifically designing a model to predict claims and identify predictors of importance. The specific research question defined are:

1. (a) To what extent can AIS-information combined with machine learning algorithms predict claims; (b) What are the most important predictors retrieved from AIS-information?
2. What do these findings suggest about the potential for implementing machine learning and AIS-information in P&I insurance?

The first research question defines the boundaries of the machine learning methods used in this thesis. The path from pre-processing data, to the final output of the models, is all purposely aimed at answering the first research question. The analysis only specifically address research question one. To answer the specific question machine learning will be applied to variables retrieved from AIS-information, relating to vessels activity level and operational pattern, such as nautical miles sailed, and the number of ports visited. The second research question is primarily answered in the discussion, as the nature of the subject makes it more suitable for a qualified assessment than a quantitative analysis.

While machine learning and AIS-information is frequently applied in research of marine accidents, there are to the knowledge of the researchers of this thesis, no studies applying machine learning to AIS-information, with the purpose of predicting insurance claims. For any insurance company, the production cost of the insurance is not known at the time of sale. Determining the correct premium is challenging, and the price is calculated based on estimated risk from customer knowledge, historical data and predictive models. As this thesis aims at investigating the predictive power of AIS-information, it may lead to more accurate risk assessments. Identifying underlying factors of insurance claims may ensure more precise and timely loss prevention measures, and possibly reduce the total number of claims.

1.2 Limitations

The research is limited to P&I insurance, claims within three categories; Cargo, Contact & Collision and Human, and nine specific types of vessels. P&I insurance refers to Shipowners P&I insurance as opposed to other P&I product, such as Charters P&I. The particular claims categories represent the most frequent or costly categories of claims. The limitations are set based on available data and available computational power. The methods used are transmissible to any vessel or type of cargo. A broader list of arguments and limitation is discussed throughout Chapter 3, 4 and 5.

1.3 Defining Marine Accidents

Marine accidents and claims may refer to a variety of situations and conditions. To ensure an unambiguous interpretation, the following definitions of a P&I insurance claim, marine incident and marine accident are used.

Marine P&I Insurance claims

Protection & Indemnity insurance claim is a formal request to a P&I insurance company for coverage or compensation for a covered loss (Brækhus & Rein, 1979).

As this thesis is limited to owners´ P&I, the request will come from a Shipowner. The term claim refers to a P&I insurance claim unless otherwise specified.

Marine Incident

"Marine Incident means abnormal events occurring in the course of operation of sea-going ships and likely to cause danger to man, ships, architectural work or the environment." (Kuehmayer, 2008).

Marine Accident

"Marine Accident means one or more than one marine undesired incident which results in personal injury, damage or loss. Accidents include loss of life or major injury to any person on board, the actual or presumed loss of a ship, her abandonment or material damage to her, collision or grounding, disablement, and also material damage caused by a ship. It is the duty of every master or skipper to examine any accident occurring to, or on board, his ship." (Kuehmayer, 2008)

1.4 Litterateur Review

To the knowledge of the researchers, there is no research studying the relationship between AIS-information and P&I insurance claims. On the contrary, an extensive amount of research on risk and causes related to maritime accidents exists. As claims related to maritime accidents constitutes to a large part of insurance claims, relevant literature related to such accidents are included in this section. The research can broadly be separated in four branches; studies of frequency and type of accidents, studies of underlying causes of accidents, studies of models to analyse marine accidents, and development of risk models of ship accidents. In the next paragraphs selected contributions to the research of marine accidents is presented.

A study of Kujala (2009) on accidents in the Gulf of Finland concluded that the most frequent type of accidents was related to groundings, collisions, fire and machinery damage. Additional research and statistics from EMSA (2018) and AGCS (2018) support the findings of Kujala and adds foundering to the list. While foundering tops the statistics in many databases, it is normally a result of other underlying factors.

The contributing factors to maritime accidents can be separated into human factors, vessel specifications, route characteristics, climatic factors, weather conditions and situational factors (Mazheri, 2017). Each of the elements could be studied separately; however, there has been a shift in focus of research related to maritime accidents from technical aspects to the human and organisational factors (Grech et al., 2008).

Multiple models for human and organisational factors have been developed; SHEL (software, hardware, environment, liveware), Swiss Cheese model, Human Factors Analysis and Classification System (HFACS), (Guizhen, Z, 2016). Harrald, et al. (1998) states that human error is the primary cause of most transport-related accidents according to all research studies and investigation reports. Through the HFACS framework, Chauvin et al. (2013) found that most collisions are caused by human errors due to decision error, because of poor visibility and misuse of instruments. By combining HFACS and cognitive maps Akyuz and Celik (2014) identified unsafe preconditions as the most essential factor for marine accidents. Akyuz and Celik point at lack of organisation on board, the absence of teamwork, and physical and mental tiredness as the main contributing factors in unsafe preconditions.

Mullai and Paulsson (2011) developed a ground theory model for marine accidents based on empirical data from 6000 accidents, and 87 variables. They found that the predictive power for fatality by the exposure element was strong, and the most significant factor was the number of people on board. In contrast, the same research found that the predictive power of the variable marine accidents was very limited in predicting fatality.

Mazaheri (2017) takes ground theory model one step further, and his framework for evidence-based risk modelling of groundings is extensive. Mazaheri argues that most risk models for groundings do not fully utilise the evidence available and they do not reflect reality in the required extent. His model identifies being of course, and loss of control as the two most significant factors resulting in groundings.

In addition to independent research, a large body of investigation boards exists to investigate maritime accidents with the objective to increase maritime safety (Kuehmayer, 2008). Severe incidents are thoroughly studied in detail by these agencies. European Marine Safety Agency (EMSA) is responsible for maintaining the European Marine Casualty Information Platform and investigate incidents involving vessels flying the flag of any member state, or accident that occurs inside the member states territorial waters. Most nations with a large fleet have independent bodies to investigate marine accidents, such as the Norwegian Havarikommisjonen. Their purpose is to increase maritime safety, and they do not take part in the distribution of guilt. The variety of reporting regimes for accidents makes data available for independent researchers. However, data from insurance companies related to the cost of accidents is not as easily accessible. Also, minor accidents and other insurance-related issues are not reported in the public databases. The vast amount of studies are related to collisions, groundings, machinery breakdown, fire and human incidents and the cause of these incidents. Most studies have maritime safety as the overarching topic and not financial implications.

1.5 Thesis Structure

Chapter 2 provides background information on maritime insurance and P&I Clubs. It intends to provide the necessary information to understand the basic principles of P&I insurance, the P&I Club structure, and current risk and loss prevention regimes. Chapter 3 explains the datasets used in the research as well as pre-processing of data. AIS is given special attention as it requires extensive pre-processing and choices of processing method is relevant for the outcome of the final models. Chapter 4 describes the dependent and independent variables used in the analysis. The purpose of the chapter is to explain the reasoning behind the choice of variables, and the expected predictive power. Chapter 5 details the methodology used to analyse data. It contains background information on machine learning and general information about the machine learning techniques used. Chapter 6 presents the final model and results. A total of five models is utilised and compared to identify the most accurate model, and the variables of most importance. Chapter 7 covers the implications of the findings, limitations of the research, and further potential for machine learning in marine insurance. Chapter 8 contains the final conclusions.

2 Marine Insurance

This chapter provides background information on the particularities of P&I insurance, the development and history of the marine insurance industry, as well as different types of products. The aim is not to detail the extensive legal framework of P&I Insurance, and only selected rules and regulations are described.

2.1 History and Development of P&I Insurance

The first marine insurance policies were written in the 13th century in Italy, after centuries of alternative methods for mitigating risk relating to piracy, stormy weather, or onboard fires (Haueter, 2013). In antiquity, risk was often seen through the lens of fate and met with acceptance rather than defiance; however, as ships grew larger and carried more cargo, the value of shipment increased. To protect their investments, seafaring nations mitigated risk by spreading their cargo on several vessels (Haueter, 2013). Diversifying shipments could only reduce a portion of the exposure, and the first marine insurance policies developed as a more sustainable method for limiting risks.

In the next centuries marine insurance developed slowly, until the Italian Lombard family moved to London in the 16th century and popularised the concept. The further development is undeniably related to the history of Lloyds and his coffeehouse in Tower Street, where the insurance market was centred from 1688 onwards (Mutenga & Parson, 2012).

Until the 1850s, the leading marine insurance products were related to hull and cargo, but in the 1850s a new insurance product where introduced, protection insurance. There are several theories as to why the demand for protection insurance arose. A common explanation is that it developed to cover the last one-fourth of collision liability, which hull underwriters refused to take on. An alternative theory relates to the introduction of new and stricter rules of liability related to death and personal injuries (Brækhus & Rein, 1979). The first P&I Club, Shipowners Mutual Protection Society, was established in 1855 in England, and it exists today, under the name Britannia Association.

Several P&I clubs were established in London in the following years, and the insurance product continued to develop. In 1870, after the loss of the vessel Westonhope, where the shipowner was made liable for the loss of cargo, indemnity protection was introduced. The new insurance products were adopted by most clubs, and it is from these two types of insurance products the clubs have their names today.

In the next decades, two P&I clubs where established in Norway, Swedish Hull Club adopted P&I insurance as a separate class, and one club was formed in the US. By 1917, twelve of the thirteen members of the IG of P&I clubs were founded. The latest member, a Japanese club, was established in 1950. Today these thirteen clubs hold 90% of the P&I insurance market (Mutenga & Parson, 2012).

2.2 Market and Products

In the modern market of marine insurance, a wide range of insurance products are available. However, the traditional products are still the most common; H&M, Cargo, and P&I. The included coverage is regulated by the insurance plan where the protection is placed. The Nordic Marine Insurance Plan (NMIP) is one of these plans: NMIP is signed by leading actors in Scandinavia and ensures a common framework for all non-P&I insurances (CEFOR, 2016). An alternative to NMIP is the Institute Time Clauses Hulls (ITC). The major difference between the plans is that NMIP covers everything not explicitly excluded, while the ITC only covers what is explicitly included. P&I is indirectly affected by these plans as NMIP provides full collision liability, while ITC only covers three-fourths. A shipowner under ITC will likely aim at 100% collision cover, and the last one-fourth collision liability is often covered through a Running Down Clause (RDC) under the P&I insurance. An owner under NMIP would not require RDC. In the following sections, H&M and Cargo insurance are mentioned briefly, while P&I is detailed to a more considerable extent.

2.2.1 Hull & Machinery

Skuld (2017) defines Hull & Machinery insurance as an insurance to protect the shipowner's investment in the vessel itself. The insurance typically covers total loss of the vessel, damage to the vessel or equipment on board, explosions and fires, and groundings (Luddeke, 1996). A clause for collision with other vessels and other objects are often added. In addition, a trading warranty usually restricts the area of operations for the vessel, and a breach of the agreement will result in limitations of cover (Skuld, 2017).

In practice the insurance cover is spread on several different insurance companies, where one company typically only covers a limited proportion of the vessel, for example, 5-10%, depending on the vessels value (Luddeke, 1996). The company with the most substantial stake in the vessel, will in case of an accident act as the claims lead, and all underwriters must follow the decisions made by the claims lead. In some circumstances, war insurance and loss of hire are included in the coverage.

2.2.2 Cargo

Cargo insurance is intended to cover the financial exposure of the cargo owner. The responsibility of the carrier depends on the specification in the Bill of lading[1], in addition to established legislation and rules (Luddeke, 1996). The owner of cargo does not need to be the shipowner and are in most cases not (Skuld, 2017). Premiums vary with the type of cargo, and multiple insurance regimes exit. The most common practice is that each cargo owner takes out insurance for his/her proportion of the cargo with one insurance company. For a 10,000 TEU container vessel, hundreds of insurance companies could have liabilities in the cargo. In the case of cargo claims, where the owner of cargo is not the Shipowner, cargo insurers often aim at reclaiming their liabilities from the Shipowner. Shipowners are insured against these claims through their P&I insurance. Most cargo claims are related to an alleged breach of contract (Luddeke, 1996).

2.3 Protection & Indemnity

P&I is a combination of two products; Protection & Indemnity. In contrary to H&M and Cargo, which aims at asset protection, P&I is a third-party liability insurance. P&I insurance can be divided into two categories, contractual liabilities, and legal liabilities. Contractual liabilities are all claims originating from parties with interest in the standard operating practices of the vessel, such as cargo owners. Legal liabilities relate to all claims arising from the jurisdiction where the vessel is operating, such as government, private persons or companies that have no contractual agreement with the ships, but whose property is affected (Gohlish, 2009). Claims in the latter category can be related to the cost of clean up after oil-spills or cost of damage to private property by any means. In the following sections, selected topics associated with P&I insurance is discussed. The two last sections of this chapter are Skuld specific, as it is based on

[1] Bill of laden - Legal document between a Shipowner and the carrier, which specifies terms of the shipment.

information that is regarded confidential by most companies. The information in this section is likely representable for all P&I Clubs.

2.3.1 Mutuality

The P&I market is divided into two main categories; insurers that are part of International Group and insurers that are not (Gohlish, 2009). Furthermore, a common distinction is between fixed and mutual insurance, where the clubs in the IG primarily are mutual. The principles behind mutuality are that the assured also become the insurer (Gohlish, 2009). The assured becomes a member of the insurance organisation, and all members are mutually liable for claims. P&I clubs are in theory non-profits, where premiums for a given policy year is estimated and paid by the members. The final premium is agreed when the insurance policy is closed, which is generally after three years, due to the complexity of settling legal disputes. In the case of remaining funds, the profits are repaid to the members, and in the opposite case, more premiums need to be paid (Skuld, 2017).

Mutual insurers do not suffer from capacity constraints, as each new member also becomes an asset to the organisation. This has enabled P&I clubs to offer higher coverage compared to what a traditional insurance company can provide (Gohlish, 2009). P&I club rules are based on a pay to be paid principal, where the insured first needs to pay the claim before claiming under the insurance policy. The pay to be paid principle limits shipowners ability to take on extensive risk, and reduce the risk of insolvency (Luddeke, 1996).

2.3.2 P&I Clubs

North (2012) defines P&I club as an association of shipowners that have grouped to insure each other on a mutual nonprofit-making basis, for their third-party liabilities. Although P&I clubs, in theory, are non-profits there are differences in how they are managed. At least two common management systems exist; P&I clubs led by a company operating for profit, and clubs managed by directly employed staff. For P&I clubs managed by a company, the P&I club and the management are two separate entities connected by a management contract. The benefit is that shipowners are more likely to only pay for their own claims handling. For P&I clubs managed by directly employed staff or a management company owned by the club, the main advantage is that such management will offer a higher level of enhanced services (North, 2012).

As previously stated, thirteen of the leading P&I companies are part of the IG, which by itself operates on principles of mutuality. All clubs are mutually responsible for the total portfolio.

The system is neatly defined and regulated. In short, the club which insures a vessel covers all individual claims up to $ 10 million, and for claims above this amount, all clubs are mutually liable (International Group, 2013).

2.3.3 P&I Policy Period

The P&I policy period starts noon GMT[2] on the 20^{th} of February. The historical reason is that many of the insured vessels were laid up during the winter months when the Baltic Sea was frozen. 20^{th} of February was assessed as the date that the Baltic would certainly be "ice-free" (North, 2012).

2.3.4 P&I Coverage

P&I coverage varies with the clubs and the insurance plan the club operates. This section comments on claims statistics and standard coverage in the owners P&I covered by Skuld. However, for all clubs listed with the IG, the club rules are very similar (Gohlish, 2009).

P&I insurance is intended to cover liabilities to thirds parties who may have a contractual or legal claim against the vessel (Skuld, 2017). The specific list of coverage is detailed and extensive, but the most frequent areas of claims origins from coverage of death and personal injuries, loss of crew members personal effects, loss of or damage to cargo, and optionally contact & collision through an RDC. Claims covered by wreck removal or pollution are less frequent, despite these types can potentially result in large payouts. In the following sections, coverage of the claims categories used in the analysis is discussed.

Death and Personal Injuries

P&I insurance covers the owner's liability for all deaths, personal injuries, and illnesses which occur on board, including death or injury to crew, passengers, stevedores, pilots and visitors to the vessel (Skuld, 2017). In addition, the insurance covers costs of repatriating crew members who become sick or are injured on board, and crew's hospital bills and expenses of sending replacement personnel to the ship (Skuld, 2017).

[2] Greenwich mean time

Claims covered in this category are among the most frequent and constitutes on average for approximately one-third of all P&I claims. However, claims in this category are usually less expensive than the average claim, and payout represents about one-sixth of total payouts.

Loss of or Damage to Cargo

Coverage in case of breach of contract of carriage is one of the primary functions for Protection and Indemnity insurance (Skuld, 2017). The cargo indemnity insurance is generally triggered as a result of damage or shortage of cargo while under the control of the shipowner. A cargo insurer will pay the owner of the cargo before the underwriter seeks to cover losses from the shipowner. P&I clubs usually handles the claim on behalf of the shipowner (Skuld, 2017). Cargo claims account for approximately 40% of all claims, and one-third of all payouts. Thus, the average cargo claims are more expensive than death and personal injuries, but still less than the average payout for a claim.

Contact & Collision

Coverage for contact & collision is only covered when an RDC and Fixed or Floating object (FFO) clause has been included (Skuld, 2017). In the Nordic Plan, collision is normally included under H&M insurance. Standard P&I coverage offered by Skuld are not including RDC and FFO, as several of the members are insured under the Nordic Plan. RDC and FFO are offered as a separate agreement.

Contact & collision are among the least frequent claims, but when they occur the payouts have the potential to be extremely high. Less than 5% of all claims are related to contact & collision, but payouts constitute for close to 40% of total payouts made by Skuld in the period studied. More detailed investigation reveals that two vessels alone constitute for 60% of total payouts in the category. Most of these payouts are not directly related to the contact, collision or grounding, but the aftermath of the incident, such a wreck removal or pollution.

The low frequency, but extreme payouts for contact & collisions emphasise the importance of the mutuality principles, both within the specific club as well as among the members of the international group.

2.3.5 Premium and Risk Determination

The two main parts related to pricing in P&I insurance are; premiums and deductibles. The current pricing structure at Skuld allows for some flexibility in the determination of deductible, while premiums are the central area of negotiation. The premium is negotiated based on market prices and a risk-adjusted price model.

Skuld's current price model works by assessing risk, based on owners claims history, and static vessel information such as gross tonnage, age, type, and flag. The price model acts as a management tool in the negotiation between the representative of the shipowner, and the underwriter at Skuld. In practice, 80% of all negotiations are done by brokers, while the remaining 20% is negotiating directly between the underwriter and the Shipowner. In the latter case, there is usually a long relationship between the actors.

Brokers have up to date information on market prices, which in most cases is the determining factor in agreeing on a premium. In a case with a significant deviation between the price model and the acceptable price of the broker, the insurance will not be approved by Skuld.

The underwriter's primary role is to act as a salesperson for the P&I Club, while at the same time ensure reliable risk assessments of entered vessels. Brokers act as an agent to provide the best possible price and conditions for the shipowner.

2.3.6 Loss Prevention

Loss prevention is performed on multiple levels, where the two main areas are information and vessel safety surveys[3]. The primary purpose of loss prevention is to reduce the number of claims, by avoiding them before they occur. An example of current measures is information brochures on how to prevent liquefication of iron ore during transport, as this is a very costly affair when it occurs. Surveys are used to assess a vessel before it enters the Club, and to identify shortcomings in any procedures the vessels undertake. Skuld also holds lectures and courses for shipowners and crew, to ensure they are up to date on the last safety requirements and best practice procedures.

2.3.7 P&I Summary

Gohlish (2009) argues that no organisation survives for over 100 years unless it serves a specific need, that is either unavailable or more expensive in alternative markets. He continues that the combined value of seaborne trade is about $ 8 trillion, while the combined P&I market generates a collective insurance premium of over $ 2.7 billion, which of about half is allocated for cargo

[3] Survey – technical control to assess safty of a vessel

liabilities. This means that the close to all seaborne cargo is insured for about 0.015% of its value. According to Gohlish, this proves that the current P&I market is remarkedly effective.

3 Data

The analysis is based on two primary sources of information; claims data provided by Skuld and AIS-information provided by Safetec. Aditional supporting dataets are utilised including world list of ports and list of flag of convenience. To answer research questions one, and build models for predicting and identifying variables related to P&I insurance claims, an extensive amount of data was required. Section 3.1 covers basics of the dataset, while section 3.2 – 3.5 covers the data in detail, as well as the extensive pre-processing that were required to make the data usable for machine learning algorithms. Section 3.6 explains the process connecting AIS-information to the claims data.

3.1 Data Collection

The entire AIS dataset provided by Safetec consist of 166 million lines of AIS-information for about 5,800 ships in the period January 2013 to December 2017. The data consists of satellite-based AIS-information that ensures global coverage. The resolution of AIS-information is one hour and was chosen by Skuld to limit the cost and amount of data.

Unstructured AIS-information contains limited to no information about aggregated activity level and operational patterns for a vessel, and to retrieve this information extensive pre-processing was necessary. The final output from pre-processing and filtering was 52 million lines of AIS-information, and 2.4 million unique transits and port visits were identified. This information was further processed to determine the aggregated activity level and operational pattern for 3,500 vessels. Claims data from Skuld were extracted directly from Skuld database, and 18,541 claims were identified in the period. Depending on the cataloguing system in use, more than 60 unique types of claims exists. In order to make the dataset usable for machine learning, claims data were categorised into four groups; Cargo, Contact & Collision, Human and Other.

3.2 Vessel Selection

The original dataset for Skuld, contained close to 5,800 insured vessels, of 19 different types. Vessels are classified by their design, purpose or type of cargo. Each class has unique characteristics, and not all classes are suitable to be analysed through the same framework. To solve this problem, all vessels were reviewed to determine suitable vessels. Nine classes were removed, and the ten remaining classes contained approximately 3,500 vessels. Also, vessels under 1,000 Gross Tonnage was removed, as smaller vessels tend to perform shorter port visits, and the one-hour resolution of AIS-information is unsuitable for detecting port visits of less than 2 hours. The largest categories of vessels not included were offshore related vessels and ferries. Offshore vessels, such as supply, operates on or in the vicinity of offshore installations, and as the list of ports does not contain information about the location of these, transit-related information would not be comparable with vessels in transit between identified ports. Ferries were removed as their port visits are too short for our algorithms to detect the port. Vessels included in the dataset are presented in Figure 1. Numbers on the sector diagram represent vessels of each type, while the percentage in the legend indicates the proportion by type in the final dataset.

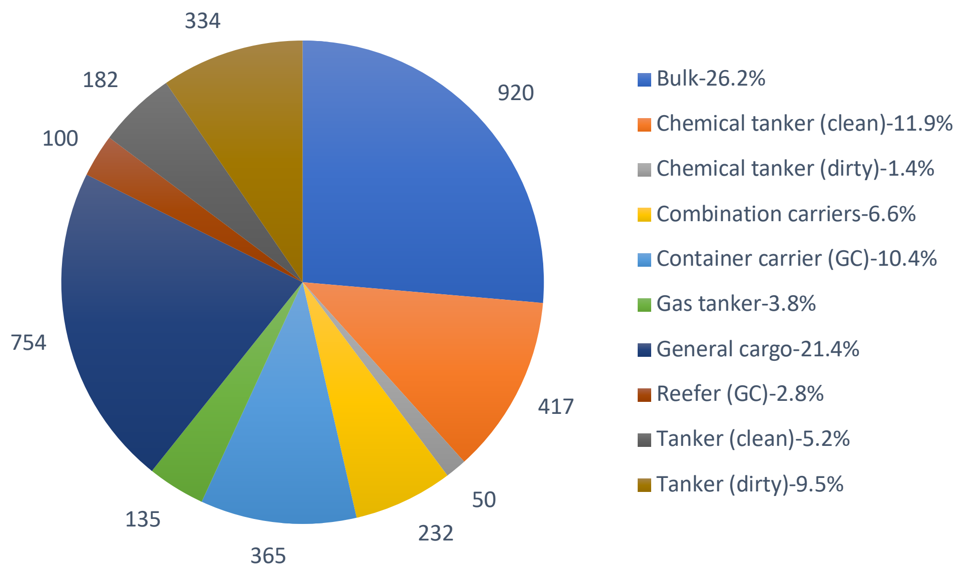

Figure 1 – Vessels Statistics by Type

Through initial testing of the algorithms, it was assessed that certain categories included too few observations for the machine learning models to work. In collaboration with insurance specialists in Skuld, a broader classification of vessels was agreed. Vessels with similar risk, assessed by the specialists, were placed in the same group, and a total of five groups were established. The categories were designed to consider the type and size of vessels. Figure 2 shows the five groups, while Table 1 summaries the criteria for each group.

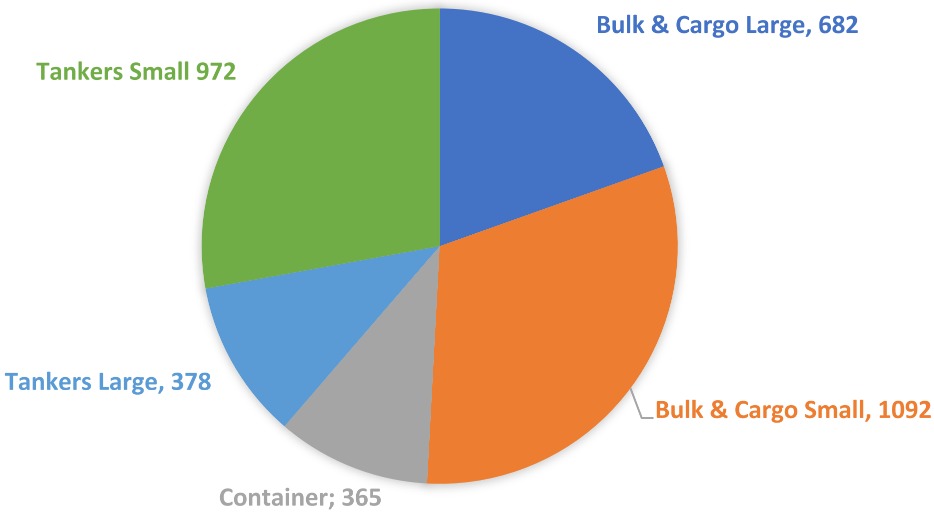

Figure 2 - Vessels by Group

Group	Size	Type
Bulk & Cargo *Large*	> 30 000 GT	Bulk, General Cargo, *Reefer*
Bulk & Cargo *Small*	< 30 000 GT	
Tankers Large	> 40 000 GT	Gas Tanker, Chemical Tanker (clean dirty), Tankers (clean & dirty), *Combination Tankers*
Tankers *Small*	< 40 000 GT	
Container	All	Container Vessels

Table 1 - Criteria for Groups

Both datasets contained information on all of Skuld's portfolio. In addition, the AIS-information contained five years of history for every vessel, not taking into consideration if the vessel were insured for only a shorter period during these years. Skuld's claims database only contains claims history for a vessel in the period of which it is insured with Skuld; thus, vessels were filtered by dates of actual insurance policy. Also, vessels with a combined policy period of fewer than six months during the five years were removed, as it was assessed as the operational statistics retrieved for these vessels were insufficient to provide value in the models.

3.2.1 Static Variables

A long list of static variables for each unique vessel is possible to retrieve. Skuld's database contains detailed information about ownership such as age, type, physical dimensions. Furthermore, as the aim of research question one is to analyse the predictive power of AIS-variables, only a limited number of static variables was selected to be utilised in the models; *Flagg of convenience, Age and Gross Tonnage*. Age and gross tonnage were directly extracted from the dataset, while *Flag of convenience* was identified by examining each vessel against the International Transport Workers' Federation (ITF) list of flags of convenience, which include 34 different countries (ITF, 2018). To travel under flag of convenience are defined as a vessel which is operated or taxed under the laws of a country different from its home country in order to save money (Cambridge Dictionary, 2018). Table 2 is a summary of the static variables used. The reasons for as to why the variables were included is covered in Chapter 4.

Variable	Description
Age	Age is calculated as the difference between the insurance year and the year the vessels where built. At the aggregate level, age was represented as age of the vessels in 2013. For vessels insured after 2013, the age at entry is used.
Gross Tonnage	Gross-tonnage is measured as the overall internal volume of a vessel. It includes space for cargo as well as crew recreational areas.
Flagg of convenience	Flagg of convenience is indicated as a factor of Yes or No. Where Yes indicates that the vessel is registered in a country of convenience. At the aggregate level, flag status in the last recording is used to calculate the variable. Information about vessels that change flag during the period is not detected.

Table 2 - Static Vessel Variables

3.3 Claims data

Claims data are collected from Skuld's insurance system. Each claim is entered into the system by a professional claims handler, typically a highly educated solicitor with extensive maritime experience. Inputs are manually entered into the system based on information the claims handler collects from representatives of the shipowner and the opposing actor. This is done in a combination of drop-down menus and free text. Only inputs from drop-down menus and amounts in USD[4] are utilised in this thesis.

[4] United States Dollar

A large part of registered claims results in zero payments, either because the deductible is higher than the claim, the claim is fought and won, or Skuld's claims handlers are able to reimburse the costs from other actors. For this thesis, a claim is defined as any claim registered by Skuld, regardless of the size of payout. Zero-payout claims were included, as claims handlers at Skuld argued that in most cases there is a fragile line between these claims, and claims resulting in a payout.

Data is some of the most sensitive and valuable information an insurance company possess, system managers at Skuld granted permission to an external database, where the information could be extracted without access to the complete system. The raw data retrieved was unfinished and insufficiently structured to be directly utilised in machine learning algorithms. Pre-processing and categorising the data was needed to get the final dataset.

3.3.1 Processing and cleaning of claims data

The original data was structured in multiple sub-datasets, and an extensive process was required to merge all datasets into one single dataset.

Each vessel was identified through the unique IMO-number[5] assigned to the vessel, and each claim registered is given a unique event number. Each vessel is possibly linked to multiple event numbers; however, each event number is only connected to one single vessel. Also, a single event can consist of numerous cases. A simple example illustrates the relations between IMO-number, event number and case; Two vessels collide, and two unique events are registered by their insurers. One vessel catch fire after the collision, and as a result, one person is injured, and nearby waters are polluted by oil. For the given vessel the insurance company will register one event type; collision, and three cases; fire on board, human injury, and pollution. The event type is used as the classification for claims as it is deemed to be the cause of the claim.

To begin the processing, all unwanted vessels included in the sample was removed, as explained in section 3.2. Furthermore, the data included multiple transactions for accounting purposes, and payouts for each event had to be processed to retrieve the final payout for a given claim. A large proportion of claims from 2016 and 2017 were not yet closed, and the reserved amount (an estimate of total payouts) were used for open claims.

[5] International Maritime Organization identification number

A total of 64 event types were identified. Certain categories contained few claim observations. To ensure a sufficient number of incidents in each category, the 64 types were categorized into four groups; Cargo, Contact & Collision, Human and Other. Table 3 describes each category.

Claim type	**Description**
Cargo	Any cargo related claim (Shortage, damage, loss of, and more.)
Contact & Collision	Collision or contact between vessels, vessel and pier or other objects, and groundings
Human	Human accidents or injuries (crew and passenger)
Other	The remainder of claims, not related to the three defined categories

Table 3 - Claim Description

The category Other contained a vast range of different types of claims, with no apparent correlation between them. As any results from this category would be highly biased, the category was not included in further analysis.

Only events that occurred between 2013-2017 where kept. The final claims dataset contained the variables: *IMO-number, Event number, Event type, Incident date, Total amount claim, Incident country* and *Incident place*. The total number of unique claims was 18,541.

3.3.2 Claims output

The following section presents the statistical findings from the claims dataset. Table 4 shows the USD amount for payouts made by Skuld in the registered period, and an average for claims by type. Contact & Collision are the categories with fewest claims, but the average amount per claim is significantly higher than the rest of the categories.

Type	**Total Amount in USD**	**Number of claims**	**Average USD per claim**
Cargo	$ 951,166,645	7,465	$ 127,417
Contact & Collision	$ 705,653,230	908	$ 777,151
Human	$ 510,667,798	5,395	$ 94,656
Other	$ 464,438,504	4,773	$ 97,306
Total	**$ 2,631,926,177**	**18,541**	**$ 141,952**

Table 4 - Claims Statistics 2013-2017

Figure 3 describes the variation in the number of claims by year, and total payout for each year. As visualised by the columns, the year to year change in the number of claims are relatively stable. As shown by the lines in the plot, payout varies considerably. An example is the limited increase of four claims for Contact & Collision from 2015 to 2016, while total payout increased from $ 72 million to $ 214 million.

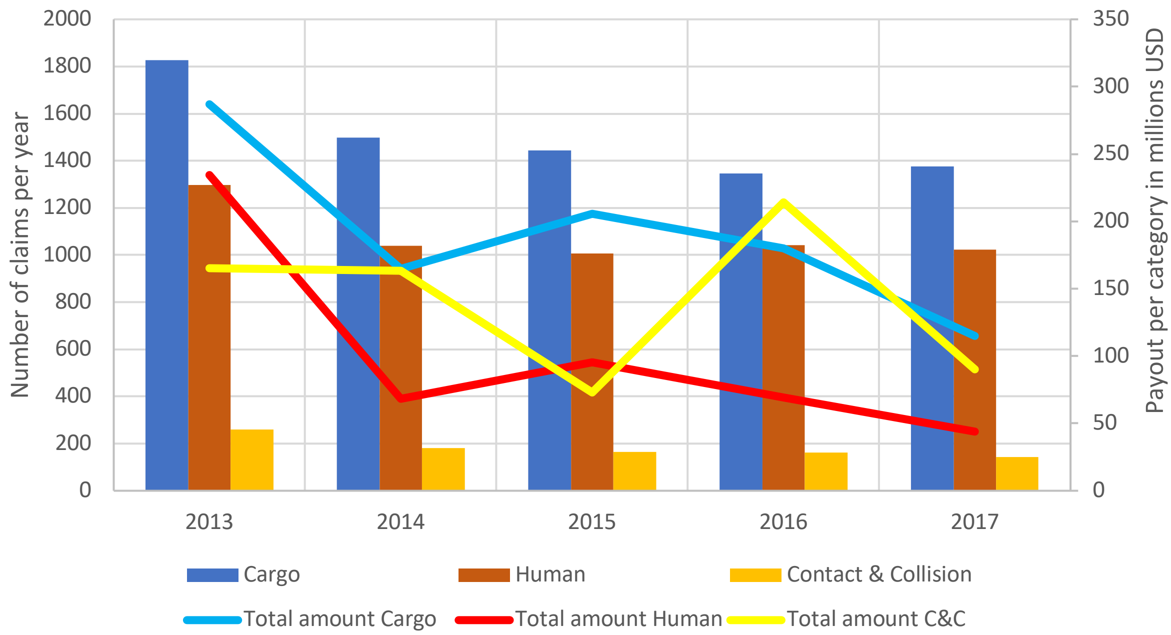

Figure 3 - Frequency and Size of Claims

Figure 4 shows the distribution of claims studies in the period. Cargo is the largest group followed by Human. Contact & Collison constitutes for a small part of a total number of accidents, but a significant portion of payout, and because of that it was decided to keep the category in the dataset despite the limited frequency.

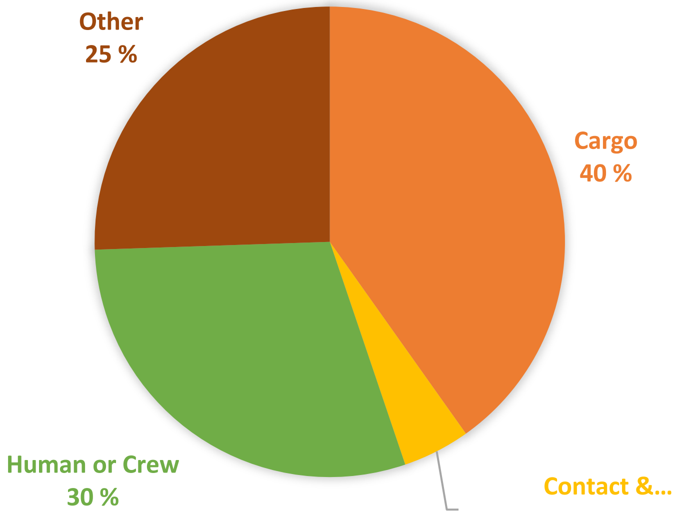

Figure 4 - Distribution of Claims by Type

The final variables retrieved from pre-processing of claims is presented in Table 5. Alternative variables are discussed in Chapter 7.

Variable	**Description**
Cargo Yes/No	Each variable represents the factor Yes / No and refers to whether or not the vessel had any claims in the specific category during the period.
Contact & Collison Yes/No	
Human Yes/No	

Table 5 - Variables Extracted from Claims

3.4 AIS-data

Automatic Identification System (AIS) is an anti-collision system designed to prevent accidents at sea (Kystverket, 2015). Currently, two systems coexist, Class-A and Class-B. Class-A is intended for vessels engaged in trade, while Class-B is designed for pleasure crafts and land stations. The main differences are concerning legal requirements, data contained in the specific AIS-message, and the frequency of update. All vessels studied in this thesis require Class-A AIS, and all information below primarily relates to this class.

In year 2000, the International Maritime Organization (IMO), adopted a new requirement that made it mandatory for all vessels above 300 gross tonnage engaged in international trade, vessels above 500 gross tonnage not involved in international trade, and all passenger vessels regardless of size to fit AIS on board (IMO, 2018). According to regulations, AIS equipment shall broadcast the following information; *ship identity, type, position, course, speed, navigational status and other safety-related information.* Detailed AIS-requirements can be found in Annex17, Chapter 5 in the SOLAS regulation.

AIS-information can be classified as vessel and dynamic information. Vessel information is entered upon installation of the system or automatically transmitted from sensors. Static information entered upon installation are; Maritime Mobile Service Identity (MMSI) number, IMO vessel number, Radio call sign, Name of ship, Type of ship, Dimensions, and reference for the position of the electronic position fixing device (EPFD) antenna. In theory, this data should be protected by a password, and only modified in case of changes in ownership or technical changes to the vessel. Dynamic information regarding the movement of the ship is automatically updated through other systems, or directly by the AIS-system itself. This information includes position, course, speed, heading and rate of turn. Transit-related data is manually entered by the officer of the watch (OOW) and are prone to error. Common manual inputs are; Navigational status, destination, cargo and ETA. The frequency of update is regulated by IMO. Table 6 presents IMO requirements for update frequency of dynamic AIS information. Static data is updated every 6 minutes.

Type of ship	**General Reporting interval**
Ship at anchor	3 min
Ship 0-14 knots	12 sec
Ship 0-14 knots and changing course	4 sec
Ship 14-23 knots	6 sec
Ship 14-23 knots and changing course	2 sec
Ship >23 knots	3 sec
Ship >23 knots and changing course	2 sec

Table 6 – Update Frequency of AIS

Table 7 lists all variables included in a standard AIS-message.

Type	Information	Source
Dynamic **Update rate determined by speed and course alteration**	Rate of turn	Automatically updated from the ship's ROT sensor or derived from the gyro.
	AIS-Navigational Status	Manually entered by OOW. (Examples: -underway by engines, at anchor, moored, constrained by draught, aground
	Speed over Ground	Automatically updated from the position sensor connected to AIS.
	Position Coordinates	Automatically updated from the position sensor connected to AIS.
	Course over ground	Automatically updated from ship's main position sensor connected to AIS
	Heading	Automatically updated from ship's main position sensor connected to AIS
	Position time stamp, UTC	Automatically updated from the position sensor connected to AIS.
	MMSI Number	Set on installation
Information item **Updated rate every 6 minute**	IMO Number	Set on installation
	Call Sign	Set on installation
	Name	Set on installation
	Length and beam	Set on installation
	Type of ship	Set on installation (Examples: Fishing, Cargo, Passenger)
	Location of GNSS antenna	Set on installation
	Draught	Manually entered by OOW
Voyage-related **Update rate every 6 minute**	Destination	Manually entered by OOW
	Hazardous cargo	Manually entered by OOW (Examples: Dangerous goods, Harmful substances, Marine pollutants)
	ETA	Manually entered by OOW
	Route plan	Manually entered by OOW

Table 7 – AIS-Information

The research in this thesis utilises a larger variety of information contained in AIS, while the essential information is the recorded positions of the vessels. The information is used to determine if the vessel is alongside, or in transit. The position of the vessel is retrieved from a global navigational satellite system (GNSS) as required by IMO. AIS suppliers are free to choose the preferred GNSS; however, GPS is most commonly used. All GNSS encounter the same performance issues. The following section discusses the limitations of GPS specifically, but the implication is in large valid for any GNSS.

3.4.1 GPS

The following information is based on official sources the National Coordination Office for Space-Based Positioning, who maintains online resources regarding GPS for the US Government. In specific, the sections rely on the GPS Standard positioning service (SPS) Performance Standard 4th edition published by Department of defence (2008).

Global positioning system (GPS) was developed and is maintained, by the United States (US) government. Initially, the system was developed for military applications. However, as the potential for civilian use is tremendous, the system was soon opened for civilian users. Today the US government is committed to maintaining the system, and a set of specific commitments regarding availability and precision for SPS is contained in the GPS Performance Standard. SPS is a part of GPS available for civilian users. The list of commitments is detailed and included accuracy standards for Signal in Space (SIS), which excluded errors originating from signal disturbances in the earth's atmosphere. Among the current SIS requirements are a 95% global user range error of fewer than 7.8 meters and a 99.94% use range error of fewer than 30 meters. User range error is the distance measured from a satellite to the receiver, and not user accuracy which is defined as a radius in meters from the GPS receiver's correct position. As all US Government commitments are related to SIS, independent performance analysis is necessary to identify user accuracy. A performance analysis performed in late 2016 found that within a 95% confidence interval, the average GPS accuracy for the twenty-eight test sites was less than 2 meters (Federal Aviation Administration, 2017). When recording positions for a stationary antenna over time, the position will fluctuate around the correct position. The precision has implantations for the algorithm used in this thesis to determine if a vessel is alongside or not, as each recording deviates by a few cm or meters. The aggregated recorded movement of a vessel in port can potentially reach several nautical miles for an extended port visit.

3.4.2 Waypoints

Waypoint is a commonly used term in navigational planning. In this thesis, the term waypoint is used to refer to the recorded position of a vessel at a given time. A twenty-four hours transit with one recording per hour will consequently have twenty-four recorded waypoints.

3.5 Processing AIS-information

Before utilising AIS-information in machine learning algorithms, it was necessary to process the data. The final output from pre-processing was a dataset containing information on aggregated activity level and operational patterns. Total time for each vessel is defined by the number of months insured in the period 2013-2017, limiting to a maximum of five years.

Pre-processing included; cleaning the data for unwanted information, identifying individual transits, and calculating and extracting necessary variables. In the following sections, each particular part of the pre-processing is detailed. This information is included, as pre-processing of AIS is a challenging matter, and there are several methods to achieve similar, but not equal results. Also, the section aims at providing the necessary background information for Skuld to determine future business decisions regarding AIS-information.

3.5.1 Data-cleaning

Errors in the original dataset were identified by manually reading AIS-information and plotting sections of the original dataset on maps to visualise possible sources of error. Three primary sources of errors were identified and removed through pre-processing; position outliers, spoofing and missing or wrong inputs.

Position outliers

Position outliers can emerge from a variety of sources, where the most common are atmospheric disturbances, multipath, or signalling errors (Department of Defence, 2008). Position outliers can deviate from the correct position with hundreds or thousands of miles.

Figure 5- Example of a Single Outlier

Figure 5 shows movements for the cargo vessel Vistigue, in the period 24-28 February 2016. 26th of February, a single position was recorded in southern Russia. Outliers of this type are identified and removed by calculating the average speed between every waypoint in the dataset. If the average speed between one waypoint and the next is above 50 knots, the position is dropped. As data is recorded every hour, removing one waypoint results in a two-hour gap between positions. For overseas voyage this does not cause any challenges; however, for short inshore transits, a two-hour resolution might be too low to identify the particular port of calls.

Spoofing and manipulation

Windward (2014) published in 2014 an analysis of data manipulations at sea, the research was based on observed AIS-data from 2012 to 2014. Windward concluded that four manipulation practices are commonly used at sea; Identify Fraud, "Going Dark" (Obscuring Activities), GPS manipulation, and AIS spoofing. Table 8 lists significant findings in the report.

AIS - Manipulation
1% of all ships use fake identification information.
On average final port of call was reported in only 41% of the time
25% of the vessels turn off AIS 10% of the time

Table 8 - Findings in Windwards 2014 report on AIS-manipulation

Implications for this thesis is that fake identification makes it challenging to differentiate between the positions that belong to the actual vessel and the one that belongs to the fraudulent vessel. If both vessels are kept in the dataset, variables such as total distance sailed and average speed will be wrong, and in most cases, extreme and unrealistic high speeds will be calculated. The underlying reason is that the method for calculating total distance sailed, where all data are sorted by date to calculate the distance between one waypoint and the next, and finally summarises all distances for each transit. Table 9 exemplifies the challenge. Rows marked with red indicates the fraudulent vessel, and the last column shows whether the algorithm for outliers would remove the waypoints or not. As the table shows, a position from the fraudulent vessel and a position form the actual vessel is removed. For single outliers, this is solved by calculating average speed between both the previous and the next waypoint. However, as spoofing have no logic pattern, this method is not possible. Running the algorithm over and over, would remove all spoofing, but also remove correct waypoints and introduce considerable gaps in the data.

IMO	Position	Time	Distance between WP	Average Speed	Removed by algorithm
9039121	5.41N 1.101E	12:00	NA	NA	NO
9039121	5.42N 1.102E	13:00	5 nm	5 kts	NO
9039121	9.13N 78.89W	14:00	2000 nm	2000 kts	YES
9039121	9.15N 78.92W	15:00	2 nm	2 kts	NO
9039121	5.42N 1.10E	16:00	2000 nm	2000 kts	YES
Totals			**4007 nm**	**1002 kts**	

Table 9 - Example of Spoofing

Table 9 shows that total distance sailed could be extremely high when spoofing is present. By removing rows with extreme values, the problem will be shifted one row down, and the process needs to be performed until all extreme values are discarded. If ten consecutive rows belong to a fraudulent vessel, the process needs to be completed ten times, and the final data will contain one correct position at time T and the next correct position at time T+10. In some cases, both the fraudulent vessel and the actual vessels AIS-information is recorded for every hour, and gaps can be avoided. The method described above also requires that the first recorded AIS-information must be the actual vessel. If not, the actual vessel will be removed, and the fraudulent vessel will remain in the dataset.

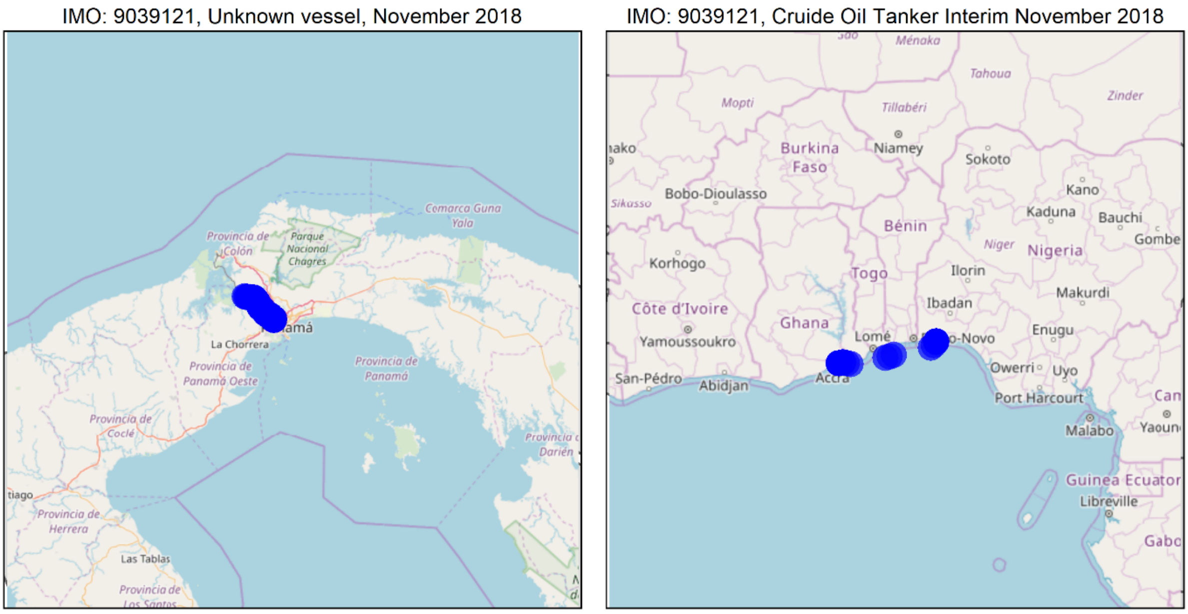

Figure 6 - Example of Spoofing

Figure 6 shows the recorded positions for the Crude Oil Tanker Interim, in November 2018. AIS shows the vessels are operating in the Gulf of Guinea, while at the same time transiting the Panama Canal. According to Skuld data, the correct positions of Interim are shown in the figure to the right, while the true identity of the vessel to the left remains unknown. The gap between the correct positions of Interim is due to the data has been replaced by positioning data from the unknown vessel. Data is recorded for every hour for the IMO-number, though it seems to be random which of the two vessels position that is registered to the database.

Identifying spoofing in a visual plot is a simple task for any human. However, as the AIS-information for the unknown vessels is identical to the one of the actual vessel, it is challenging to write an algorithm to detect and remove spoofing.

The method decided to use for removing spoofing were based on predicting the position of the next waypoint and compare it to the recorded position. If the recorded position deviates by a set value from the predicted position, the recorded position is dropped. This method assumes that the first vessels observed in the dataset are the actual vessel. The method was found to be efficient to remove spoofing. The downside of the method are that it demanded intensive computer power, and would break down when the two vessels operated in the same area. The algorithm could switch to the unknown vessel and discharge all correct positions. The discargement of a vessel was observed and fixed once in the dataset, but the problem might be more frequent. Ultimately, the algorithm cannot verify that all vessels with fake identification have been removed, and the dataset is likely to include some fraudulent vessels.

Going dark

Windward (2014) identified that 25% of vessels turn of AIS, 10% of the time. The findings were controlled for scheduled and unscheduled loss of GPS availability. Windward found that larger vessels were more likely to turn off AIS, in order to conceal their activities. When a vessel turn of AIS, it is not possible to observe the movement of the vessel. In this thesis, the waypoints are connected regardless of the time between the positions. The connection is made by combining the positions to a straight line, without taking into consideration the feasibility of the transit along the line. Possible port visits and deviations from routes are not observed, and the outcome variables could contain errors.

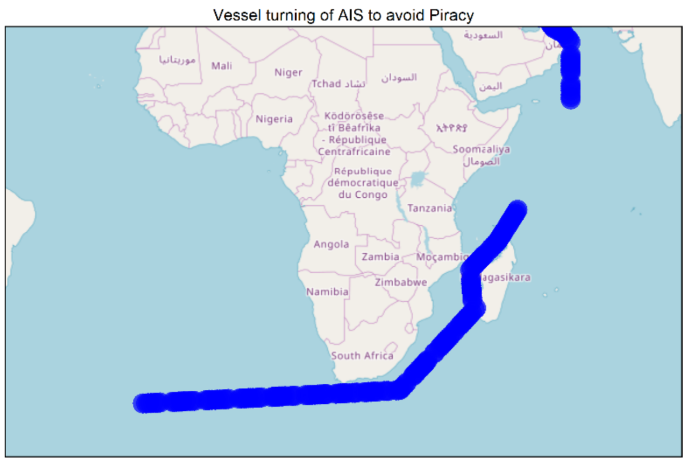

Figure 7 - Vessel "Going Dark"

Figure 7 presents the recorded positions of the Bulk Carrier vessel Magsenger 11, transiting from Argentina to the Gulf. The plot shows that Magsenger 11 turned off AIS after passing Madagascar. The recordings are from 2013, during a period were piracy in Aden still was making headlines. Several shipowners advised their vessels to turn off AIS when transiting these waters, despite the recommendation from IMO's Best Management Practices for Protection against Somalia Based Piracy to keep AIS on (BMP4, 2011). In theory, the vessel could have performed a port visit in Somalia during the four days without AIS. In this specific case that is highly unlikely, but it creates uncertainties in the dataset.

Missing or wrongful data

Windward (2014) identified that only 41% of the recorded waypoints contained information about the final port of call. All manual AIS-inputs are prone to error, either by deliberate misleading, poor procedures or by accident. The variables for navigational status and final port of call could be of great use; however, the variables are not given any special attention as the data is prone to error, and multiple inconsistency was observed in the manual inputs. No manual data was utilised in the final models.

3.5.2 Defining individual routes

Unstructured AIS contain large extent of information, and it is possible to extract operational data, such as total distanced sailed, total time at sea, average speed. However, to exploit the dataset further and identify variables such as the number of port calls, number of transits, the data needs to be structured into individual transits.

For the identification of individual transits, a series of methods were used. An individual transit is defined as the movement between one port and the next. This definition is not detailed enough to ensure unambiguous routes through machine learning, and the technical description of a transit is described as follow; A transit starts when a vessels navigational status changes from static to moving and ends once the vessel is stationary. Additionally, the total distance of the transit needs to be longer than 1 nm. The additional requirement was added to avoid separate transits when a vessel performs harbour movement, like changing pier to get fuel, give place to another vessel, or turn around to load from the opposite side.

The process of extracting transits is simple in theory, while in practice several methods had to be performed. The next section describes the general process of how transits were derived.

Determine navigational status

As previously mentioned, the AIS-navigational status depends on manual updates from the OOW and will in several cases be wrong or delayed. To identify the proper navigational status of each waypoint, three parameters where used; distance, speed, and change in heading between one waypoint and the next. For this thesis, two navigational status were defined and used; stationary and moving.

The logic behind the method is that a vessel at port does not move more than a few meters due to tides, current and wind. The heading of the ship will not change more than a few degrees based on the same logic, and the log[6] will not record high speeds while alongside. Two combinations of these criteria were used to determine if the vessel is stationary or moving.

Combination one:

- Navigational status *stationary* was decided, based on the following criteria:
  - Change in the heading of fewer than 3 degrees
  - Speed below 0.6 knots
  - Distance between two waypoints of less than 0.015nm (27.78 meters)

Combination two:

Navigational status *stationary* was decided, based on the following criteria:

- Speed below 0.2 knots
- Distance between two waypoints of less than 0.005nm (9.2 meters)

For combination one, a length of less than 0.015 nm (27.78 meters) was chosen based on the GPS theory presented in section 3.4.2, and detailed studies of the dataset. Change in the heading of fewer than 3 degrees and speed of fewer than 0.6 knots, were based on values found in the recordings of vessels in port.

Combination two were included to counter for sudden change in heading of certain vessels when moored to a pier. In some cases, the vessels turned around to bring the other side of the vessel alongside, while in other cases it is assessed that the gyro-compass[7] were turned off, causing temporary deviation in heading.

Heading were explicitly chosen to differentiate between vessels at anchor and in port. Anchorage could qualify as an independent navigational status, and there are arguments to support both that a vessel at anchor should be regarded as a vessel in port, or as a vessel at sea. For the purpose of this thesis, anchorage is classified as vessels at sea. In most cases, this is achieved. However, as the behaviour of vessels at anchorage and port are very similar,

[6] A vessels log indicates the speed of the vessel measured in nautical miles per hours, knots.
[7] Electronic compass used by vessels in order to detect true north, steer, and find positions and record courses.

anchorage could in some cases be classified as port visits. These classification errors can especially be found for vessels anchoring in still waters, or inside breakwaters.

Assigning transit number to individual transits

Once navigational status was determined, a transit was identified by assigning a common transit number to a sequence of waypoints with similar navigational status. All waypoint with the same number belongs to the same transit or port visit. The last waypoint of a port visit was assigned both the port stay and the next transit, to ensure that the transit could be named by the name of the port of departure and the port of arrival. The data were corrected for small movement of less than 1 nm or transits with the same port of departure and arrival of length sailed less than 8 nm.

Assigning port of departure and port of arrival

Safetec provided a list of approximately 5000 ports, which were used to assign port of arrival and port of departure for each individual transit.

Each port was represented by the name of the port and a single position. To identify the port of arrival and port of departure for each transit, the first and last waypoint of each individual transit were extracted. The distance between each individual waypoint and all ports were calculated, and the port with the shortest distance to the waypoint were assigned. The calculations were extremely computer intensive, and data had to be split in multiple sections in order to achieve the calculations.

The calculated distance between a vessel and the port will varies by the size of the port, and a maximum distance between a vessel and a port were assigned to ensure that the vessels were within the port area. A short radius ensures that few vessels are assigned the wrong port, but a large number of vessels would not get any port assigned. Rotterdam harbour area extends for 20 nm, while the port of Oslo extends less than 2 nm. In this thesis, a vessel is assigned a port if it is within 8 nm of the port. If the vessel is outside the 8 nm radius, the country name will be assigned instead of the port name.

3.5.3 Variables Retrieved from AIS

After each transit was identified, information was aggregated to describe the activity level and operational patterns of each vessel. Table 10 includes all variables extracted from AIS and provides information about the technical methods used for the calculations. All variables are calculated for the aggregated insurance period for the vessels. A detailed argumentation for the choice of variables is included in Chapter 4.

Variable	Description
Number of transits per month	Aggregated number of all transits for each vessel, scaled by the number of months with available data for the vessel.
Total distance sailed per month	Aggregated total distance sailed for each vessel, scaled by the number of months with available data for the vessel.
Total time at sea per month	Aggregated time spent at sea for each vessel, scaled by the number of months with available data for the vessel.
Average distance at sea	Aggregated distance sailed for each vessel, divided by the number of transits for each vessel.
Average time at sea	Aggregated time sailed for each vessel, divided by the number of transits for each vessel.
Number of Unique ports per month	Aggregated number of unique port visits, scaled by the number of months with available data for the vessel. For unknown ports, country is used, and the actual number of unique ports is likely higher for certain vessels.
Proportion at sea vs. port	The total percentage of time spent at sea versus time spent in port. 100% indicates that a vessel has not been in port during the period. As explain above, vessels at anchor are in most cases classified as vessels at sea.

Table 10 - Variables Extracted from AIS-Information

3.6 Combining AIS-information and claims data

In order to analyse the explanatory power of AIS against claims data, the datasets had to be merged. At an aggregated level AIS-information and claims data can be combined by IMO-number; however, to maintain valuable information contained in the table of transits, AIS and claims were combined by using both IMO-number and incident date.

The major challenge in combining the datasets was related to the uncertainty in the reported incident date. While AIS-information contains extremely accurate timestamps, the incident date of a claim is dependent on when the damage was identified, as well as the accuracy of each claims handler. Specific claims such as collisions, groundings and personal injuries are recorded in the vessels logbook, and one can assume that the incident date is somewhat accurate. Damaged cargo might not be identified before the cargo is handed over to the owner; thus, the incident date could deviate by months from the correct date.

The limitations related to incident date is only of concern when studying variables for individual ports or transits. At an aggregated level, such as total transits performed or total distance sailed, the impact of a wrong incident date only matters the claim are assigned to the wrong year.

3.7 Summary

Pre-processing is time-consuming and computationally intensive. The balance between available time, data power and accuracy of the results were a constant factor through the work. The extensive pre-processing finally resulted in 3 categorical variables, 4 static variables and 7 AIS-variables.

For static vessel variables, the accuracy is assessed as high, and only changes of flag of convenience are deemed as a source of error. For claims related data, no particular concerns associated with the final data is identified, except for accuracy of incident date, which cannot be assessed without detailed analysis of other sources. Pre-processing of AIS was the most challenging task, and the AIS variables are likely to contain sources of error. However, as the variables are used at an aggregated level, the impact of errors are assumed to have limited effect.

4 Variables

While pre-processing was elaborated in Chapter 3, this chapter aims at explaining choice of variables. The overall aim of this thesis is to identify the potential for machine learning and AIS in P&I insurance and develop a model to predict claims, and the choice of variables is of great importance to develope a feasible model. The choice of variables is based on previous research, recommendations from Skuld and Safetec, and feasibility to extract the variables from AIS.

4.1 Dependent Variables

Information from claims data provides the opportunity to research a large variety of dependent variables. During the exploration of models, a range of variables was tested, and three variables were found to have the potential for machine learning models; *Cost of Claims*, *Claim/No Claim*, and *Number of Claims*. *Claim/No Claim* was chosen as the preferred variable, and a discussion on the alternatives is found in section 7.6.

4.1.1 Assessment of Claim/No Claim

General statistics from the dataset reveals that the majority of vessels experience some sort of claim during the period. However, when studying the four groups of claims; Cargo, Contact & Collison, Human and Other, the frequency of claims is more widely spread. The assessment of whether a vessel will have a claim or not is a useful variable and can be used to assess risk factors and identify variables of importance. If a vessel has any cargo claims during the period, the variable will return "Yes" for the category Cargo claims. The variable does not take number of claims or time occurrence into consideration.

4.2 Independent variables

This section provides reasoning for choosing the variables included in the models. In addition to the seven variables related to vessel operations, four variables were included; *Age*, *Age²*, *Flag of Convenience* and *Gross Tonnage*. The four variables were chosen on recommendations by Skuld, based on previous findings and the fact that they are all part of Skuld current pricing model. All variables (except *Flag of Convenience*) is numeric, and most variables are continuous, except for *Proportion of time at sea*, which is a percentage. The following sections assess how the independent variables presumably will react and a short explanation as to why.

Several of the machine learning algorithms are complex and studies the relationship between independent variables as well as the relationship between the independent and dependent variables. There might be indications for certain variables that can contributes to more claims and fewer claims at the same time, depending on the relationship between the variable and other variables. The assessment in the following sections is based on an all else equal methodology, and not taking into consideration relations between the independent variables.

Age

Age is included to identify possible relationship with age and claims frequency, as age can be regarded as a proxy for vessel quality. The most intuitive impact is that increased age results in more accidents as the equipment on board are more likely to fail (Butt et al., 2012). However, an older ship might be more familiar to the crew, and result in less accidents. The relationship could alternatively be assumed to be quadratic with more frequent claims, primarily related to Human, in the first period when the ship is new, and a declining number of claims as the familiarity with the vessel increases. Thus, the variable *Age²* were included.

Gross Tonnage

A variety of methods to describe a vessels size, weight, or carrying capacity exists; displacement, deadweight tonnes, gross registered tonnage and gross tonnage is some the most commonly used. The difference measurement relates to carrying capacity, volume and absolute weight. As several of the vessels, types insured by Skuld does not move goods in bulks, gross tonnage was chosen as the preferred measurement to compare vessel size. *Gross Tonnage* is defined as the moulded volume of all enclosed spaces of a vessel and selected as the preferred measurement (Oxford Dictionaries, 2018).

Larger vessels have a different operational pattern than smaller vessels. Large vessels tend to conduct long and few transits and are likely to move more cargo than smaller vessels. Consequently, it is more likely that larger vessels are more exposed to cargo claims, while possibly less exposed to human accidents. For smaller vessels, one can assume operations are performed more manually, which might lead to an increased number of human accidents.

Flag of convenience

A vessel sailing under *Flags of convenience* could be exposed to higher levels of risk than vessel sailing under the flag of the home country. *Flags of convenience* are possibly linked to countries with fewer requirements than other countries; thus, maintenance and condition of vessels might be worse than for other vessels (Butt et al., 2012). Thereby, vessels under flag of convenience could be more exposed to accidents. Certain professional at Skuld, argued strongly that flag of convenience is not of importance, as it is merely a result of financial decisions (primarily related to taxation) and does not reveal information about the quality of the shipowner.

Number of unique ports per month

Number of unique ports per month represents the operational pattern of vessels. One could expect that a high number of unique ports would results in more frequent incidents as the areas and operators the vessel interact with are unfamiliar. Vessels on long contracts normally visit fewer ports, compared to vessels that trade in the spot marked. As 15% of the port visits are defined as unknown, the actual number of unique ports are likely higher than the variable express and could be a source of bias. To reduce the impact from unknown ports, country of departure or arrival is used to compensate for unknown ports, and thus ensures higher precision in the variable.

Number of transits per month

Number of transits per month relates to activity level and/or length of each individual transit. A high activity level could indicate increased risk, and vessels with few and short transits are likely more exposed to risk. A high number of transits also indicates more regional travel, and possible more transits in high traffic areas. High traffic density is known to increase risk (Stornes, 2015)

Average distance at sea

Average distance at sea indicates the length of each transit at sea. The variables relates to operational patterns of the vessel, more than activity level. However, a high average could be a result of both few and long transits or short and faster transits. The prior example would likely result in fewer claims, while a high average due to high speeds and low turnover could indicate increased risk, as it also relates to increased activity level.

Total distance per month

Total distance per month is related to activity level and/or length of each individual transit. High activity level as a result of multiple short transits is likely increase risk, while fewer and longer transits will likely reduce risk.

Proportion at sea

Proportion at sea is measured as the time at sea compared to time in port, and a level of 100 indicates 100% time spent at sea and no time in port. Increased proportion at sea indicates a higher activity level and possibly increased risk. The variable also reveals information of the operational pattern of the vessels, as a percentage close to 100% will indicate that the vessel is operating between offshore installations or in layup. The latter can be controlled for by measuring total distance sailed.

Average time at sea

Average time at sea indicates the average time of each individual transit. The variable is related to activity level and operational pattern of the vessel. High average indicates high activity. A low average indicates lower activity, and the vessel are more likely operating in regional waters.

Total time at sea per month

Total time at sea per month indicates the activity level of vessels. If a vessel has a high time at sea, the proportion at port will be lower. Increased time at sea is likely to imply increased risk as it relates to an high activity level. Exceptions are vessels at anchor or in lay-up that are currently not in trade, and their risk exposure is low.

5 Methodology

Machine learning is a broad terminology, and a wide range of algorithms and methods are included in the term. This chapter briefly explains the terminology, before it elaborates on methods and algorithms specifically used, and how model performance is determined. The last section describes the specific process for how the analysis was performed.

5.1 Machine learning models

Machine learning is one of the most up and coming fields of study, but the term has been around for quite some time. Arthur Samuel (1959) defined the area as a "Field of study that gives computers the ability to learn without being explicitly programmed". Since then, there have been numerous ways to explain the area. For simplicity, the term in this thesis describe learning algorithms used for modelling and prediction. The two main types of statistical machine learning models are classification and regression. The difference is that the output of a regression model is quantitative, while for a classification model the output is qualitative (James et al., 2013). All models presented and evaluated in this thesis are classification models; however, regression models were utilised in the process of identifying possible dependent variables.

5.2 Supervised and unsupervised machine learning

Statistical and machine learning problems are often categorised into two groups; supervised learning and unsupervised learning. Unsupervised learning is defined as when the measurement or predicting variables are known, while the outcome variable is unknown. The process can be seen as "learning without a teacher" since the prediction cannot be associated with a response variable (Hastie et al., 2009). Unsupervised learning is primarily used for analysis, estimation and clustering.

In supervised learning, the response variable is known, and the primary goal is to explain or predict the outcome, based on the predictors (James et al., 2013). By acting as a guide, the intent is to teach the algorithm the result it should present. The requirement for the algorithm to work is that the output is known, and training data are labelled. Supervised machine learning algorithms can be used to solve both classification and regression problems (Hastie et al., 2009).

Since the response variable for the dataset used in this thesis is known, the process is classified as supervised learning.

5.3 Splitting data and K-fold Cross-Validation

To assess the performance of machine learning models, a method for validating the model is essential. A conventional approach is to split the dataset into two sections; training set to make the algorithm, and a test set to assess the performance of the model on new data (Burger, 2018). A random split is used, where the algorithm ensures that a minimum number of outcomes are included in each set. This is achieved by specifying the dependent variable for the split to minimise the possibility that the observations in either the training or test set are skewed. Data in this thesis are split by 80% for training and 20% for testing. By assigning most of the data to the training set, the purpose is to build the best possible model to predict unseen data from the test set.

Cross-validation is a resampling procedure to evaluate the model while performing model training (Hastie et al., 2009). Training data is split into k-folds, where k-1 folds are used for training, and one-fold are used for validation. The purpose of cross-validation is to minimise error while training the model and to avoid overfitting the unseen test data (Hastie et al., 2009). For this thesis, 5-fold cross validation has been performed on the data.

Figure 8 visualise the process of data splitting and cross-validation used in the analysis.

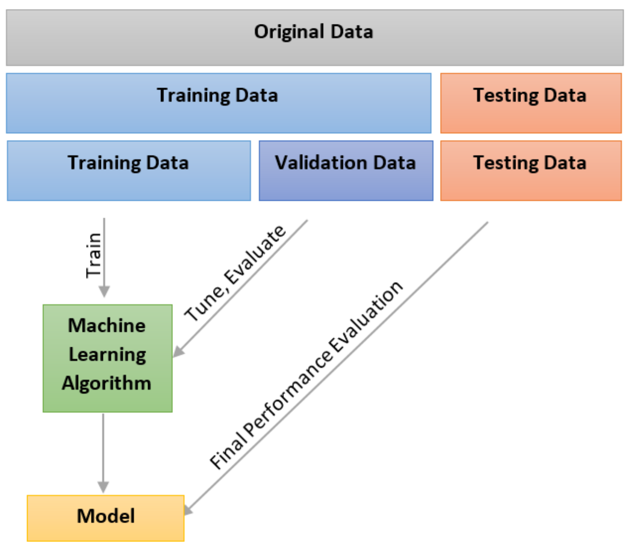

Figure 8 – Model-Validation (Eissa, 2018)

The following sections give a brief introduction to machine learning models used in the analysis in Chapter 6. The complexity of models varies, and a broad understanding of the models are only achieved through a detailed mathematical description. This section aims at providing the minimum knowledge necessary to interpret the output of the models.

5.3.1 Logistic Classification model

Logistic regression models are similar to linear models in predicting, but instead of fitting a linear output, they classify the dependent variable as a categorical value. While the output in linear regression is a numeric value, logistic regression models print the probability that the dependent variable belongs to a particular category (James et al., 2013). By using the logistic function, the response variables are given a value between 0 and 1, which is referred to as a binomial distribution (James ett. All 2013). The regression is then performed, using a method called maximum likelihood, which solves a mathematical formula to maximise the chance of predicting the correct outcome for all response variables. The outcome of the regression is a value between 0 and 1 for each response variable, and a threshold is determined to classify the response variable in one of two categories. A threshold of >0.5 will ensure that all response variables with a value of less than 0.5 will be assigned one category, and the variables with a value of 0.5 or more will be assigned the other category. The specific logistic regression model used in this thesis is a generalized linear model with a binomial probability distribution, and in the remaining of the work referred as logistic classification model.

5.3.2 Random Forest

Random forest is one of the most popular and influential algorithms used in machine learning. The model is built on a method called bagging or bootstrap aggregation, a technique for reducing the variance of an estimated prediction function (James et al., 2013). The bagging method creates multiple subsamples of the training data, which are trained to predict an outcome. The algorithm creates decision trees based upon each subsample and aggregates the results. Instead of using the same predictors each time, the model randomly samples predictors for each new tree. Random sampling adds more diversity and reduces variance, at the cost of equal or higher bias. The process create a more robust and powerful model.

For classification, Random Forest uses the most frequent result from all trees, known as the majority vote. One of the most valuable features of the random forest is that it reduces overfitting by creating multiple decision trees (Burger, 2018). Single trees are often prone to overfitting, primarily if a tree consists of many branches. In random forest two main features are decided by the researcher, number of trees and the number of variables randomly sampled as candidates at each split.

5.3.3 Naive Bayes

Naive Bayes is a widely used algorithm to solve classification problems. The conditions of the model are based on Bayes theorem of probability. The term naive springs from the simplifying assumptions for how the algorithm threats interaction between predictors. The first assumption of the algorithm is that there is strong independence between predictors; in reality, this is generally not true (Hastie et al., 2009). The second assumption is that the formal value of the predictor is irrelevant, and the algorithm produces a probability of each outcome. Based on this assumption, the algorithm produces a probability model and chooses the outcome with the highest probability. Despite these rather optimistic assumptions, Naive Bayes classifiers often outperform far more sophisticated alternatives (Hastie et al., 2009).

5.3.4 AdaBoost

Adaptive boosting, or more commonly known as AdaBoost, is a boosting algorithm. Boosting is a method which aims at converting "weak" learners into "strong" learners by combining multiple weak learners (Hastie et al., 2009). A weak learner is a predictor that is only partially correlated with the correct output, but provides better results than pure guessing. Strong learners are well-correlated with the output, and are naturally more suited to predict the correct value (Hastie et al., 2009).

AdaBoost performs this process by sequentially using the weak classification algorithm on a repeatedly modified version of the training set. In each iteration, both classifiers and data points are weighted to force the algorithm to focus on observations that are difficult to predict. After a certain number of iterations, the final weight of the predictors is calculated by combining the result of all previous attempts, using the majority vote (Hastie et al., 2009).

Boosting is in theory applicable to numerous models, and multiple variations to the original AdaBoost is available. In the analysis a version of AdaBoost applied to classification trees is used. The specific model utilises the original AdaBoost algorithm in addition to two alternative versions; Gentle and Real AdaBoost. The difference is the specific mathematical formulas applied when boosting, while the general principle of making "weak" learners "strong" is the same. The model can be tuned by specifying the number of threes and the depth of each tree. Learning parameter can also be set to avoid overfitting and limit the use of computer power.

5.3.5 Neural Network

Neural Networks are nonlinear statistical models, with the ability to study the relationship among variables in addition to the relationship between independent and the dependent variable. The term neural network comes from the biological term neuron, which is a nerve cell that receives, process and transmits information. Neural network algorithms are built on the same basis, with artificial neurons that solve for the input they receive. An artificial neuron is a number between 0 and 1, depending on what it represents (Burger, 2018).

A neural network consists of multiple layers, where the output from one layer is used as input in the next, until the last layer, where the neuron with the highest number represents the final outcome or answer to the classification problem. Like the logistic model, the final output is estimated by maximum likelihood (Hastie et al., 2009). All layers between the first and the last are referred to as hidden layers. The complexity of a network is dependent on the number of hidden layers and the number of neurons in each layer. In theory, all neurons are connected, and if it were not for the simple mathematical formulas of linear combinations, developing neural network would be an extremely computer-intensive task. The specific algorithm used for the neural network is the NNET, which consist of one single hidden layer.

5.4 Evaluation of model performance

In order to assess the performance of a classification model or to compare two models, a set of statistics must be applied. In the analysis Accuracy and Kappa statistics were utilised, as both statistics are commonly used in classification problems (Akosa, 2017, Viera et al., 2005).

For each model a confusion matrix can be extracted; the matrix provides information about the model’s performance, by presenting the distribution of how the model classified all observations compared to the true result. Information from the matrix can be used to calculate both Accuracy and Kappa statistics. Table 11 presents an example matrix, and how the information should be interpreted.

Prediction		Reference: NO	Reference: YES
	NO	TN	FP
	YES	FN	TP

Table 11 - Example of a Confusion Matrix

In application to the specific analysis performed, True Positive (TP) is the correctly predicted values for true claims, True Negative (TN) is the correctly predicted values for *No Claim*. False Positive (FP) and False Negative (FN) is predictions the model got wrong. To make a comparison between models more intuitive, the Confusion Matrix were scaled into percentages. The primary function of this adjustment is that each category of claims and vessels contain a different number of observations, and a percentage would make it easier to compare the results.

$$TN = \frac{TN}{TN + FN} \qquad FN = \frac{FN}{TN + FN} \qquad TP = \frac{TP}{TP + FP} \qquad FP = \frac{FP}{TP + FP}$$

The value for TP should be read as the percentage of true claims the model got right. The sum of each column adds to 100% and represent the total number of true claims and the total number of no claims. The sum of each row has no logic interpretation. Accuracy is calculated by the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy of 1 indicates a perfect model, while a model with accuracy closer to 0.5 has very limited predictive power. Accuracy has severe limitations in imbalanced datasets (Akosa, 2017). Kappa aims at overcoming this limitation by taking into consideration Expected Accuracy. Expected Accuracy is calculated by the formula below, where OBS equals the number of total observations.

$$Expected\ Accuracy: \frac{\frac{(TN + FN) * (TN + FP)}{\mathrm{OBS}} + \frac{(FP + TP) * (FN + TP)}{\mathrm{OBS}}}{\mathrm{OBS}}$$

Kappa is further derived by the formula for Accuracy and Expected Accuracy, and calculated by the following formula:

$$Kappa = \frac{Accuarcy - Expected\ Accuracy}{1 - Expected\ Accuracy}$$

Kappa values are interpreted on a scale from < 0 to 1, where a value of 1 indicates perfect agreement, and 0 or less indicates less agreement than chance (Viera et al., 2005). Table 12 summaries the interpretations of Kappa.

Kappa	**Agreement**
< 0	Less than chance agreement
0.01 - 0-20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.60 - 0.80	Substantial agreement
0.81 - 0.99	Almost perfect agreement

Table 12 - Interpretation of Kappa

Accuracy and Kappa can be calculated for all classification models and compared among all models used in this thesis.

5.5 Identifying the most suitable model

As previously stated, accidents are unexpected events, and multiple factors are of importance when analysing accidents or insurance claims. Thus, it is not possible to aggregate all the data and expect good results. From the discussion in Chapter 4, one dependent variable was chosen; however, three types of events were assessed as the minimum to ensure useable results. This makes for three different dependent variables of the same classifier. In Chapter 3 the logic for selection of vessels by type was outlined, and the uniqueness of each type of vessel makes it challenging to make a single model containing vessels from different classes. To limit the number of necessary models, all vessels were grouped into five categories. Last, five algorithms, including the benchmark model, were chosen to build the strongest model. Thus, to overcome the issues of vessel and claims specifics a total of 75 models[8] was assessed.

The extensive number of models made for a real challenge in how to compare, evaluate and not least, present the findings. To solve for these challenges the following sequence and method were used in the analysis, to stepwise assess the performance of the models. The process is outlined in Chapter 6.

6 Analysis and Results

In the following chapter, results from the analysis are presented. The analysis aims at answering research question 1a and 1b. A combination of tables, figures, and written explanations are used to ensure readability. Findings in section 6.1 and 6.2 emphasise model performance, where section 6.1 address model specific and 6.2 group specific variations. Bulk & Cargo Small vessels are presented as illustrative examples. Section 6.3 conveys the variable importance retrieved from the Random Forest model. Section 6.4 present conclusion for the analysis.

6.1 Models Performance

Assessing the model performance, a logarithmic classification model was used as a benchmark. The reason behind using the logarithmic model as a benchmark is due to the simplicity and high interpretability of the model. All models are performed with the use of 5-fold cross-validation to for tuning and optimising the results for testing and avoid overfitting. Only results from Bulk & Cargo Small (B&CS) have been explicitly presented for each model, and Appendix A-E contains results for the remaining vessel groups. For each model; Accuracy, Kappa and the Confusion Matrix is displayed and discussed.

For Contact & Collision, Kappa values are near zero for all models and are found to have no predictive power. Results from this category are thereby only discussed under the section for the benchmark model. Also, Kappa values for Container are close to zero for most models and findings are just mentioned briefly.

---

[8] 3 Claims categories * 5 group of vessels * 5 machine learning algorithms = 75 models

6.1.1 Model Performance of Benchmark Model

Results from the benchmark model for B&CS are highlighted in Table 13. The model achieves a prominent score for Accuracy on Cargo claims, with a Kappa value deemed as fair. Contact & Collision produces the best Accuracy for all models, but the Kappa value is near zero, which indicates that the model has low or none predictive value. For Human, the results are weak, and Kappa indicates only a slight agreement.

Model Logistic Classification	Validation Performance		Test Performance	
	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.7661	0.3420	0.7622	0.2718
Contact & Collision	0.7729	0.0039	0.8432	0.0452
Human	0.6212	0.1824	0.5730	0.1174

Table 13 - Model Performance - Logistic Classification

The results from B&CS are representative for the remaining groups, with one notable exception. Kappa values for Container in predicting Cargo claims is negative and indicates no agreement. Except for Container, the general trend is that the model predicts Cargo claims better for Bulk & Cargo and Human better for Tankers. Additionally, the model is unable to predict claims for Contact & Collision for any of the groups.

	Cargo			Contact&Collision			Human	
	No	Yes		No	Yes		No	Yes
No	23 %	4 %	No	99 %	100 %	No	64 %	43 %
Yes	77 %	96 %	Yes	1 %	0 %	Yes	36 %	57 %

Table 14 - Confusion Matrix - Logistic Classification

The output presented in Table 14 is the Confusion Matrix of B&CS. The result for Contact & Collision explains the lows Kappa values. The model gets 99% correct of *No Claim*, while all predictions for Claim are incorrect. As there are significantly more *No Claim* than *Claim*, the total Accuracy is high, while Kappa correcting for, by providing a low score. Furthermore, the model for B&CS predicts 96% correct *Claim* for Cargo, while only 23% the *No Claim*. For Human, correct and incorrect predictions are near evenly distributed around 60%, which are the explanation for the low Accuracy. The trend is similar for the remaining groups, except Tankers large, where the model is better at predicting *Claim* for Cargo than *No Claim*. For all groups, the models perform better in predicting *No Claim* than *Claim* for Human.

6.2 Model performance for advanced machine learning models

In this section, advanced algorithms are used to identify models with enhanced predictive power. The machine learning models have considerable more tuning abilities to be optimised for retrieving better results. A variety of machine learning algorithms were tested and assessed, and four models are presented. The performance measure for each machine learning model is assesed against the benchmark model, to ensure comparability in the results.

6.2.1 Random Forest

This section presents the results for the Random Forest (RF) model. Each model was tuned to grow 1000 decision trees and randomly sample 3 variables for each split. The general performance of RF exceeds the benchmark model, both on Accuracy and Kappa. As seen in Table 15 for B&CS, Accuracy is slightly higher for Cargo and significantly higher for Human. Kappa values for both models are significantly higher compared to the benchmark model.

Model	Validation Performance		Test Performance	
Random Forest	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.7718	0.3605	0.7798	0.3928
Contact & Collision	0.8144	0.0441	0.7982	-0.0571
Human	0.6749	0.3442	0.6422	0.2795

Table 15 - Model Performance - Random Forest

The trend for the remaining groups is similar to B&CS. Except for Tankers Small and Tankers Large for predicting Human, where Kappa values of all models achieve a significantly higher score than the benchmark. Overall, RF achieves significantly higher Kappa agreement and a slightly higher Accuracy.

	Cargo		Contact & Collision		Human	
	No	Yes	No	Yes	No	Yes
No	44 %	9 %	96 %	100 %	67 %	39 %
Yes	56 %	91 %	4 %	0 %	33 %	61 %

Table 16 - Confusion Matrix - Random Forest

As seen in Table 16, RF for B&CS most accurately predicts *Claims* for Cargo, while for *No Claim* on Cargo, less than half of the predictions is correct. For all groups, RF performs more accurate forecasts for *No Claim* than the benchmark, which results in higher Accuracy. For Human, the distribution is generally better for all the models, and especially Tankers Large are able to predict claims for Human at a significantly higher rate than the benchmark.

6.2.2 Naive Bayes

Naive Bayes (NB) is the model with the most similarities to the benchmark in the way the algorithm works. There are performed any tuning for the final model, and it is performed with standard settings. Table 17 shows a significant increase in Kappa and a slight increase in Accuracy for all models. It is the only model that provides a slight agreement for Contact & Collision.

Model	Validation Performance		Test Performance	
Naive Bayes	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.7429	0.3327	0.7706	0.4009
Contact & Collision	0.7439	0.0492	0.7523	0.1018
Human	0.6362	0.2737	0.6193	0.2477

Table 17 - Model Performance - Naive Bayes

For the remaining groups, the model gives the same indication as for B&CS. The exceptions are Human in Tanker Large and Tanker Small, where the benchmark model outperforms NB on both Accuracy and Kappa.

	Cargo		Contact & Collision		Human	
	No	**Yes**	**No**	**Yes**	**No**	**Yes**
No	51 %	13 %	86 %	76 %	54 %	29 %
Yes	49 %	87 %	14 %	24 %	46 %	71 %

Table 18 - Confusion Matrix - Naive Bayes

Assessing the Confusion Matrix of NB, the model gives slightly lower correct predictions for *Claim*, compared to the benchmark model. However, the NB models manage to predict *No claim* significantly better than the benchmark. The increased Accuracy in the models is thereby a result of the more accurate prediction of *No Claim*.

6.2.3 AdaBoost

To work with AdaBoost (AB) the dataset had to be normalized and converted to a matrix. The parameters for the model were set to a depth of 3, and a learning rate of 0.1. No restrictions for iterations were set, and the best fit for the majority of the models used 150 iterations. AdaBoost performs significantly better results for B&CS on both Accuracy and Kappa for Cargo and significantly better Kappa for Human.

Model	Validation Performance		Test Performance	
AdaBoost	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.7659	0.3016	0.7936	0.4007
Contact & Collision	0.8261	0.0000	0.8303	0.0000
Human	0.6800	0.3568	0.6284	0.2489

Table 19 - Model Performance - AdaBoost

For most groups AB perform significantly better results for Cargo. For Bulk & Cargo a significant increase in Kappa for Human are observed. For results for Tankers against Human are poor, with a decrease in performance for Tankers Small, and negative Kappa values for Tankers Large. Container has significant Kappa values for Human, while the Accuracy is on the lower end of acceptable values.

	Cargo		Contact & Collision		Human	
	No	**Yes**	**No**	**Yes**	**No**	**Yes**
No	39 %	5 %	100 %	100 %	68 %	43 %
Yes	61 %	95 %	0 %	0 %	32 %	57 %

Table 20 - Confusion Matrix - AdaBoost

AB's Confusion Matrix from Table 20 shows more accurate predictions for *No Claim* compared to the benchmark model. The result of this is increased Kappa for the majority of the models. Exceptions are Human for Tankers Large, where AB predicts *Claim* significantly better than the benchmark, while the benchmark outperforms AB on *No Claim*. For Tankers Small the benchmark also outperforms AB for Human on both predictions.

6.2.4 Neural Network

The following section presents the results for Neural Network (NNET). As with Adaboost the data were converted into a matrix and data normalised. Tuning parameters for learning parame ter were set to 0.1. NNET outperforms the benchmark model on Cargo, while underperforms f or Human.

Model	Validation Performance		Test Performance	
Neural Network	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.7670	0.3635	0.7838	0.3662
Contact & Collision	0.7820	0.0000	0.8432	0.0000
Human	0.6528	0.2687	0.5568	0.0913

Table 21 - Model Performance - Neural Network

The remaining groups show variable performance when comparing the benchmark and NNET. For Bulk & Cargo Large/Small the prediction for Cargo are almost similar, while for Human model performance is low. For Tankers Small, the benchmark outperforms NNET for Human, while for Cargo the models have relatively similar results. The benchmark model outperforms both models for Tanker Large.

	Cargo			Contact & Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	36 %	7 %	**No**	100 %	100 %	**No**	59 %	40 %
Yes	64 %	93 %	**Yes**	0 %	0 %	**Yes**	41 %	60 %

Table 22 - Confusion Matrix - Neural Network

The NNET model predicts *No Claim*s for Cargo at a better rate than the benchmark, while the rest of the parameters differ only a few per cent as shown in Table 22. For the remaining groups, most of the models are outperformed by the benchmark model.

6.2.5 Overall Assessment of Model specific variation

Overall model performance is as expected low to moderate. However, several models predict *Claim* for Cargo and Human significantly better than the benchmark model. Random Forest, AdaBoost and Naive Bayes provide the best overall results. Variation in performance across these three models is low. However, for particular groups and categories, extensive variations are observed, and a combination of the three models could provide the best overall result. Neural Network produces the weakest results among all models.

6.2.6 Model Performance by Group

Previously model specific variations were addressed, while in the following section, group-specific variations are highlighted. In the first part group variations for Random Forest are presented, while the following parts address the overall variation across all groups for Cargo.

Random Forest - Cargo				
Group	Validation Performance		Test Performance	
	Accuracy	Kappa	Accuracy	Kappa
Bulk & Cargo Small	0. 7718	0. 3605	0. 7798	0. 3928
Bulk & Cargo Large	0.7269	0.3733	0.7132	0.3203
Tanker Small	0.6938	0.3603	0.6546	0.2794
Tanker Large	0.6243	0.1231	0.7067	0.3529
Container	0.5618	0.0809	0.5000	-0.0791

Table 23 - Model Performance - Cargo with Random Forest all Groups

Table 23 shows that predictions for both groups of Bulk & Cargo, as well as Tankers Small, provided the best overall results in predicting claims for Cargo using Random Forest. Accuracy and the average Kappa for Tankers are generally weaker than the score of Bulk & Cargo. Negative Kappa and the low accuracy for Container indicates no agreement.

Random Forest – Human				
Group	Validation Performance		Test Performance	
	Accuracy	Kappa	Accuracy	Kappa
Bulk & Cargo Small	0.6749	0.3442	0.6422	0.2795
Bulk & Cargo Large	0. 6284	0. 2119	0. 6103	0. 1897
Tanker Small	0.6798	0.3256	0.5928	0.1494
Tanker Large	0.6276	0.1189	0.6933	0.2694
Container	0.6381	0.2665	0.5972	0.1675

Table 24 - Model Performance - Human with Random Forest all Groups

Table 24 shows the results for RF against Human for all groups. Performance statistics are on average lower when predicting Human. However, all groups show agreements of slight or fair.

Figure 9 visualises the model performance for predicting Cargo for all models and all groups (negative values for Kappa are not included). The plot indicates the highest Kappa agreement for Bulk & Cargo, and it illustrates that no models can predict Cargo for Container vessels. The plot also shows that AB, NB and RF are the models with the most predictive power. A similar plot for Human is found in the Appendix F.

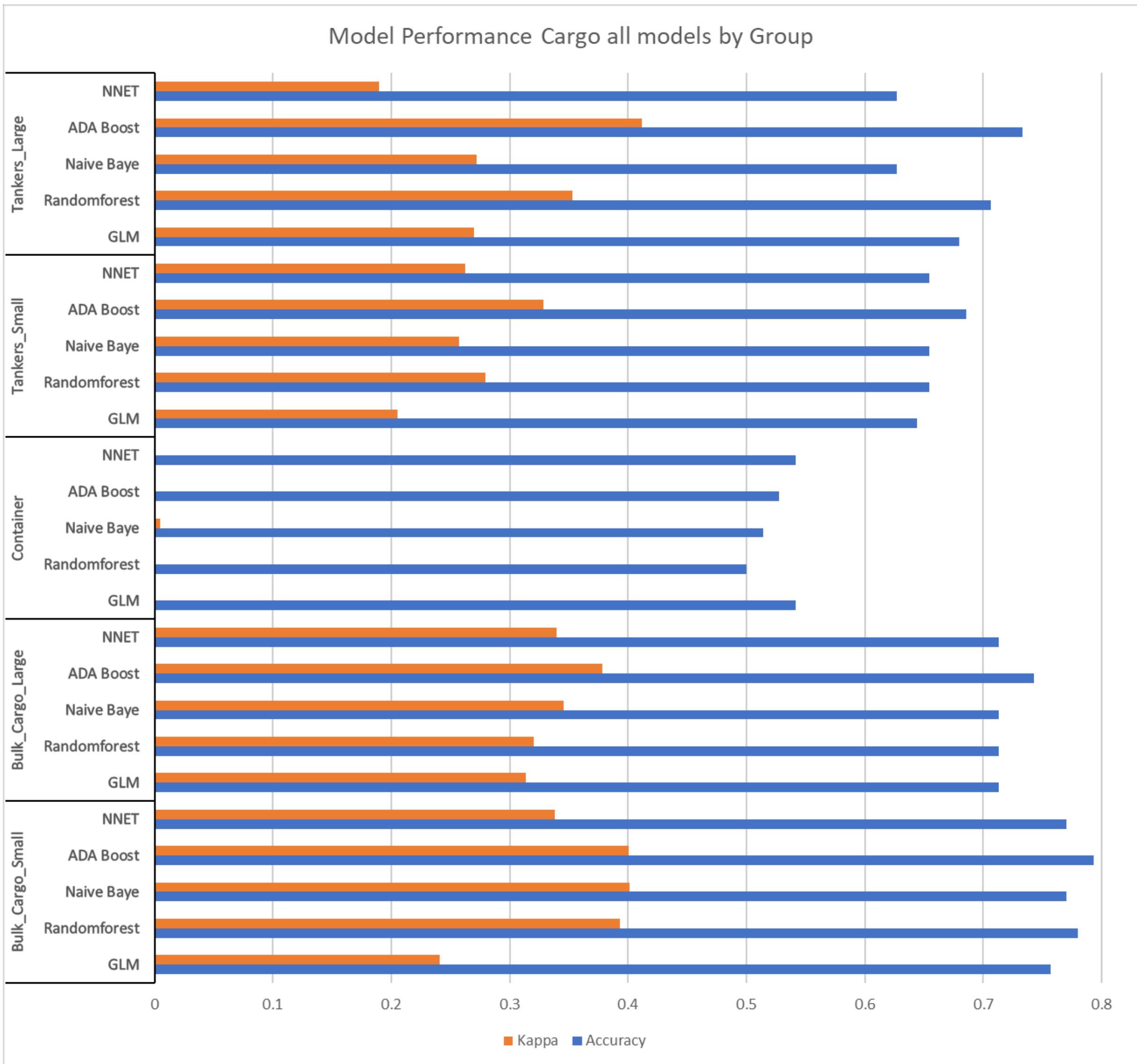

Figure 9 - Model Performance - Cargo all groups

6.2.7 Summary Model Performance

On average, all models outperformed the benchmark model. However, beating the benchmark is not the most prominent insight. The important conclusion of this section is that the models are able to predict claims based on AIS-information. Three models were found to provide the highest agreement overall; Random Forest, Naive Bayes and AdaBoost. The models provided no value in two particular situations; predicting claims in Contact & Collision, and Cargo claims for Container. Table 25 summarises the findings in section 6, by visualising the average predictive power in each model.

Predictive Power

Green = Strong, Yellow = Medium, Red = Weak

	Cargo Claims	Contact & Collision	Human
GLM			
Random Forest			
Naive Baye			
ADA Boost			
NNET			

Table 25 - Predictive Power of Models

6.3 Variable Importance

Prediction power of models is essential but has limited contribution when it comes to loss prevention. To impliment the correct loss preventing measures, identifying the most important predictors is necessary. Several machine learning models have limitations when it comes to extracting variables of importance. The “black box” of machine learning refers to a model where inputs go into a box and decisions comes out on the other side (Hastie et al., 2009). One of the critiques towards “black box” models are that the process between input and output is not observable for the researcher (Burger, 2018). In this thesis, there was deliberately selected a mixture between “black box” and simpler models, as variable importance is of significant interest for the research. In the following sections findings for variable importance of Random Forest is presented. Random Forest was selected, due to high overall performance, and variable importance can easily be extracted from the model.

6.3.1 Random forest - Variable Importance

In the following section, variable importance is presented for Cargo and Human for all vessel groups. Variable importance for Contact & Collision is omitted, as the model performance in section 6.1 and 6.2 were assessed as too low to provide any value. The variables are weight by their importance, and the most important variable is listed at top. The concept of benchmarking variable importance is of limited value, so results from logarithmic classification are omitted in this part.

Figure 10 shows variable importance from Random Forest. The plots show that variable importance varies among all vessel groups, as well as between claims categories for each group of vessels. An essential feature of that Random Forest is that it is not an adaptive model, and the plots presented below does not indicate the direction of how the variable affects claims statistics, but it is a measure of the strength of each predictor. The decrease in accuracy is used as a measure of the importance of a variable in Random Forest (Hastie et al., 2009).

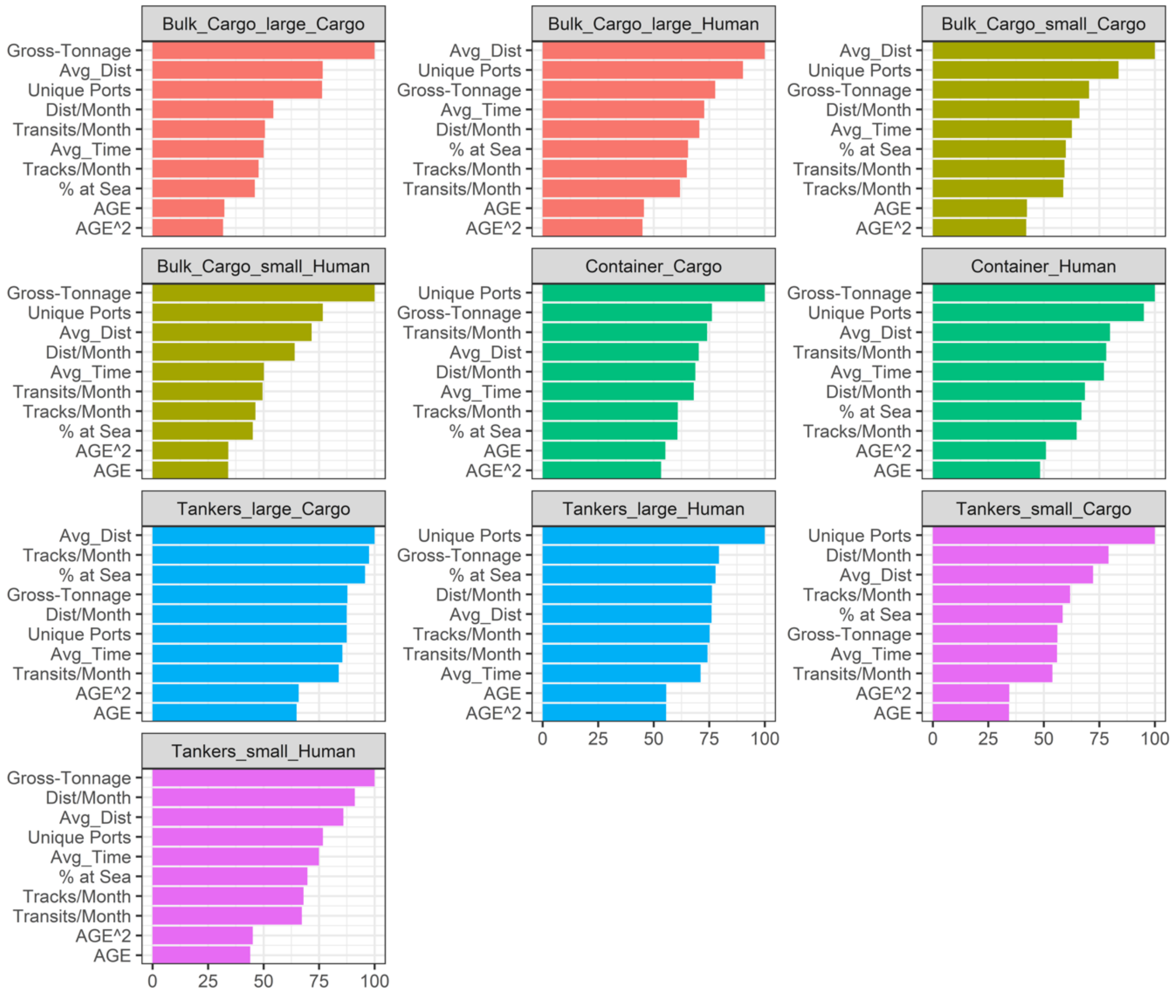

Figure 10 - Variable Importance for Random Forest

Four variables score higher than the others on average, regardless of vessel group; *Gross Tonnage*, *Average distance sailed per Transit*, *Distance Sailed per Month*, and *Number of Unique Ports per Month*.

There are several possible reasons for *Number of Unique Ports per Month* to be among the most important factors. *Number of Unique Ports per Month* is likely related to transiting in new and unknown waters, and lack local knowledge by the navigator might increase the rate of accidents. On the contrary, good navigators would pay special attention when transiting in new waters, and utilise pilots when taking the vessel into port. Thus, *Number of Unique Ports per Month* could also indicate fewer accidents. Another reason might be related to the communication and relationship between the crew and operators in the port. Each port has unique procedures and methods, and without an established relationship between the actors, misunderstanding might occur, and cargo and persons can take damage or be injured during port operations.

Average distance sailed per transit indicates the operational pattern of the vessel. Ocean crossing vessels will have a high average distance, and regional or domestic vessels will normally have a shorter average distance. Ocean crossing vessels are often larger than regional and domestic vessels. Increased *Average distance sailed per transit* can possibly lead to fewer accidents; one explanation is that machinery is better off with running more extended periods, then short intervals, and results in less maintenance and damage to machinery. Shorter *Average distance sailed per transit* could result in increased frequency of claims as the vessels are likely in areas with higher traffic density and with more port activities.

Distance Sailed per Month acts as a proxy for activity level. Two operational patterns lead to high distance sailed; vessels with short to medium transits with high speeds and short turn-over in port, and vessels with long transits and short turnover time in port. For both types, the high activity level could indicate increased risk and more claims.

The variable *Age* and *Age²* scores very low in all models and groups. A possible reason could be that Skuld does not insure vessels above 30 years of age, unless the vessel is of particular interest. Age of maximum 30 years is likely chosen by Skuld, based on historical data, and could represent a threshold for risk. The short time span of age in the sample is likely explaining the limited importance of age in the models.

In variable importance plots variables is sometimes omitted due to low importance. In Figure 10, F*lag of convenience* was excluded by the model. The importance of *Flag of convenience* was revealed in a secondary plot, and the overall importance was less than *Age and Age²*.

Table 26 shows a summary of findings for variable importance and the prediction strength of each predictor retrieved from AIS-information.

Predictive Power

Green = Strong, Yellow = Medium, Red = Weak

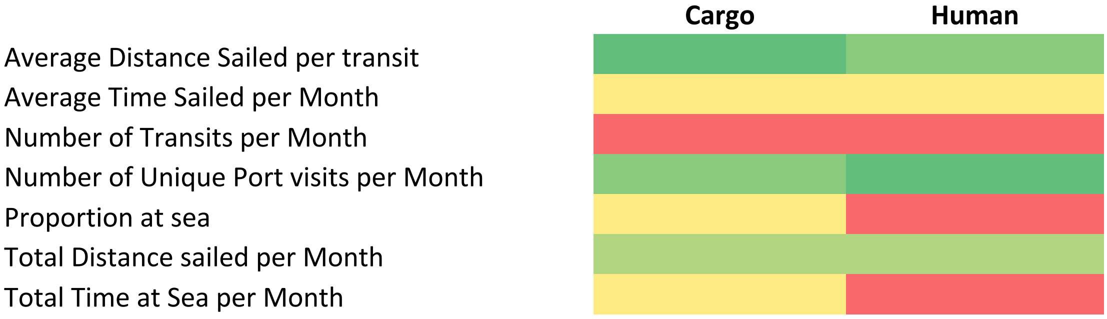

Table 26- Predictive Power of AIS-Variables

6.4 Conclusion Analysis

Chapter 6 aims at answering research question 1; To what extent can AIS-information combined with machine learning algorithms predict claims, and what is the most important predictors retrieved from AIS-information?

Results from the analysis indicate that models based on AIS-variables are able make predictions for most claim categories. The overall results for Cargo are fair to moderate, for Human it is slight to fair, while predictive power for Contact & Collision is close to zero. AdaBoost, Naive Bayes and Random Forest provide the best overall results, and on average their predictive power is relatively similar. The results from predicting Cargo claims achieve on average on the highest Accuracy and Kappa score. The models predict approximately 75 % of Cargo claims for Bulk & Cargo groups and 68 % for Tankers.

Chauvin et al. (2013) found that most collisions are caused by human errors due to decision error, because of poor visibility and misuse of instruments. The models used to predict claims do not contain any predictors for human characteristics, visibility conditions or systems specifications for vessels. This could be one of the main reasons why all models for Contact & Collision provide poor results.

Results for variable importance makes it hard to give a single conclusion. The overall tendency is that *Number of* Unique *Ports per Month*, *Total Distance Sailed per Month* and *Average Distance Sailed per Month* is assigned most importance.

In relation to research question 2, the analysis provides evidence that machine learning models based on AIS-variables have predictive power for Cargo and Human, and could provide value in risk assessment and determination of insurance premium. Regarding loss prevention, machine learning can identify the most important AIS-variables and provide additional support in determining measures to prevent claims.

7 Discussion

The purpose of this thesis, as previously stated, is two-folded; investigate the opportunities for machine learning in marine insurance, while specifically making a model to predict claims based on AIS-information. The following sections are structured to discuss the two research objectives, based on statistical findings from the analysis. The questions are strongly connected. However, for the purpose of clarity, they have been separated into two sections; statistical findings, and potential for machine learning in P&I Insurance. In addition, the chapter includes sections on financial impact, limitations, recommendations, and possible improvements of models.

7.1 Statistical Findings

As outlined in Chapter 6, the overall model performance varies by vessel group and type of claim. Predicting claims is a challenging task, and any statistically significant results are of potential value. Multiple models achieve fair to moderate results in predicting Cargo claims, and fair results in predicting Human claims. Certain models have no predictive power; in particular, the models fail to predict claims related to Contact & Collision for all vessel groups, and Cargo claims for Container vessels.

The model accuracy of close to zero for Contact & Collision is not surprising, as predicting collisions are exceptionally challenging. Multiple researchers such as Harrald et al. (1998) and Chauvin et al. (2013) conclude that human factors are the main triggering factor related to collision and groundings. Activity level might indirectly relate to physical and mental tiredness, which is one of the factors that are likely to increase the risk of accidents (Akyuz, 2014). However, most AIS-variables primarily relates to activity level and operational pattern, and not directly to human factors, and the low-performance is reasonable. In addition, claims for Contact & Collision are underrepresented in the dataset, and the algorithm has fewer observations to build a strong machine learning model.

The results for Container vessels are surprising. One possible factor could be that Container is the smallest group of vessels, resulting in fewer observations to build models. On the contrary, the group Tankers Large only consists of a few more vessels than Container and provides significantly better results. Comments from claims handlers at Skuld has not provided any further insight. Inspecting the data for Container did not indicate any explanations for the differing results. Further analysis of the group Container should be assessed to reveal possible explanations for the low performance.

Increases activity are likely to induce more Cargo claims, as higher activity level translates to more shipment of cargo. The models predict approximately 75 % of Cargo claims for Bulk & Cargo groups, and 68 % for Tankers. The findings represent valuable insight for Skuld, as activity level is not a part of current risk-assessments. A further discussion for implementing the results are included in the section 7.2.

Models for Human scores on average 62 % for Bulk & Cargo, and 66 % for Tankers. Activity and the operational pattern seem to have less explanatory power for Human claims. A reason could be that the relationship between cargo transported, and cargo damage is linear. While for Human, higher activity could lead to more claims, but also more experienced workers and familiar procedures, resulting in fewer claims.

Findings for variable importance are primarily identified by their weight in the Random Forest model. The results reveal minor disagreement as compared to the expected results discussed in Chapter 4. The limited significance of *Age* and *Age²* defy the expectations, and indicates that the variables are not suited for risk assessment. Research by Clarkson (1991) finds that maintenance seems to be more important than age when it comes to quality of a vessel, which support the low relevance for *Age* and *Age²* in the models. The low importance of *flag of convenience* is in agreement with statements from several professionals at Skuld, who claims that flag is only a measure to optimise taxes, and that does not reflect the quality of the shipowner. However, the finding contradicts the results of Butt et al. (2012) analysis on marine accidents.

The high importance of *Gross Tonnage* may indicate that the groups defined in Chapter 3 are not optimal, and more precise limits of size should be adjusted before further modelling. The low importance of *Number of transits per month* and moderate importance of *Total time sea per month* are more surprising. The moderate importance for *Total time at sea per month* could be explained by the inclusion of vessels at anchor in the aggregated time, and the accuracy in relation to activity level is less than variable related to distance. No logical explanation for the low importance for *Number of transits per month* is identified.

The high importance of *Number of* Unique *Ports per Month*, *Total Distance Sailed per Month* and *Average distance sailed per Transit* agreed with the expectations, and their importance was discussed in section 6.3.1.

7.2 Potential for machine learning in P&I Insurance

Insights from analysing claims have two main areas for utilisation; identify potential loss prevention measures, and increased accuracy in determining risk. The following sections discuss possible implementations for the specific findings, in addition, the sections take a broader view on how machine learning can contribute to P&I Insurance.

7.2.1 The potential for loss prevention

In Chapter 6 machine learnings ability to identify variables of importance were demonstrated. Knowledge of these variables can be exploited in different ways in loss prevention. The research from this thesis has identified some areas where machine learning can be implemented.

Monitor fleet activity / Targeted loss prevention

The machine learning models can bring Skuld valuable information into which factors should be monitored. By implementing a service where Skuld and the members can monitor their activities, Skuld can inform members when potential threshold values are reached and suggest measures to minimise risk. An example could be that a member vessel, reached a certain threshold for average sailing distance. Skuld can notify the member and suggest that a survey should be performed in the next port, to check for possible errors which can lead to a claim. Implementing thresholds for the predictors identified in this thesis could provide additional support to members and possibly differentiate Skuld from other P&I companies. An alternative to informing shipowners are to impose restrictions on their activity levels. P&I insurance companies have a limited history with restricting vessels operations, and only certain limitations in the area of trade-related to war zones are common. Imposing limitations based on a maximum total distance sailed or the number of ports visited is absent, and the market is likely, not prepared for such restrictions.

Education and training

A current area of loss prevention is information campaigns, training and education. The curriculum is determined through findings in claims statistics and reports. Results from the analysis can be implemented in a more targeted training for shipowners, and possibly implement new subjects in the curriculum. A specific example is to inform shipowners of the increased risk vessels that trade in multiple ports is exposed to compared to liner traffic between fewer ports. The loss prevention measures could include a recommendation for the use of pilots, or specific naming preferred operators in the most risk exposed ports.

CSR & Sustainability

Corporate Social Responsibility (CSR) and sustainability are not directly related to loss prevention, but negative publicity can result in financial loss. Machine learning can identify unfavourable trading patterns, or trading that violates international treaties, and shipowners can be sanctioned or expelled from the club based on their behaviour.

7.2.2 The potential for risk assessment and pricing

The process to determine premiums were addressed in section 2.3.5, where two means are used, a risk assessment model and market prices. Machine learning has the potential to increase accuracy in risk/price models by implementing more variables and exploit complex relationships between activity levels and claims.

Accurate pricing of segments

Identifying under or over-priced segments are of substantial value and can be used to identify unwanted risk or to increase market shares. If a vessel segment is under-priced, Skuld can reduce prices to gain market-shares. If a vessel segment is over-priced, Skuld should increase premiums in the following negotiation period. If a segment has to high risk after new assessments, Skuld could evaluate to drop the segment. By identifying miss-pricing in the insurance portfolio, Skuld can lower the overall risk, which can result in lower claim amounts and more stable premium incomes.

Activity-based insurance

Current risk models are as discussed based on historical performance of the member and static information about the vessel. Current models are likely developed based on what records were available when P&I insurance started to develop, and until electronic navigation systems and AIS became available, the vessels activity levels were close to impossible to monitor. With easily available AIS-information and possibly other external sensors, the activity can now be continuously monitored. Findings in Chapter 6 makes a strong argument for activity-based pricing. Activity can be linked to pricing in multiple ways. An interval pricing structure could work as a simple framework, where the premium is partly based on total distance sailed each year. If the agreed interval is breached, a new premium is determined, or the deductible is increased.

Automatisation of Marine Insurance

P&I insurance premiums are primarily determined by negotiations between brokers and underwriters. It is a manual process, often conducted person to person, and including extensive travelling. A manual process is usually costlier than an automatic process, and by developing a sophisticated automatic price model, the use of brokers can be decreased. In a possible future scenario, shipowners buy insurance online through a Skuld portal, and not through a broker. Automated pricing reduces costs for all actors.

7.3 Financial impact

In order to emphasise the relevance of the work, this section aims at setting a value in USD of the findings. Calculating potential savings is challenging, as predicting claims is one part of the equation, but avoiding the claim is something different. To overcome these issues, several simplifications are made, and the calculation is based on payouts for Cargo and Human claims. As stated in Table 4, costs related to Cargo claims in the period were $ 951,166,645, and for Human $ 510,667,798. The average payout is $ 127,417 for Cargo claims and $ 94,656 for Human claims.

On average, the models are able to predict, 73% of all Cargo claims and 64% of all Human claims. Even though models predict a large proportion of claims, it is challenging to determine a number on how many of these claims can be avoided. Two of the major actors in AIS risk-analysis Skuld have met with, estimates a yearly reduction in claims between 7-14%. In USD this is equivalent of saving between $ 20 - $ 40 million, based on anual payouts.

In addition to savings from loss prevention, Skuld can use the information from the predictors in their pricing models to better calculate premiums and increase earnings. More accurate pricing also reduces the risk of having to collect additional premiums from members during the policy period. Predictable prices are of importance to all members, as it allows for a better allocation of funds throughout the year.

An explicit value of the findings is impossible to calculate without further research. By comparing the agreed premiums with an alternative premium based on activity-level for the five year period, a potential profit of using activity-level could be uncovered. These calculation should Skuld consider doing, before deciding on implementing activity-based pricing system. At this stage, the best estimation for savings in loss prevention, which the established AIS risk-analysis actors presents.

7.4 Limitations

Limitations in scope are discussed in section 1, this section address limitations in the sense of weaknesses with the work. Most of the arguments are related to limitations as a result of choices during pre-processing of data.

Several sources of error were not entirely eliminated through the pre-processing. Two of the issues relates to spoofing and identification of ports, and both are thoroughly discussed in Chapter 3. An additional limitation as a result of the pre-processing is the treatment of vessels at anchor. The impact of grouping vessels at anchor with vessels at sea was not revealed before the analysis was complete, and as discussed variables relate to time are heavily biased by vessels at anchor. In hindsight vessels at anchor should have been treated as a particular case, and the accuracy of the activity level as a function of time would likely increase significantly. An adjustment should be implemented in future work. In general, all bias and errors in dataset will affect model performance (Hastie et al., 2009).

This thesis is based on the assumption that most claims are caused by accidents, and not deliberately insurance fraud. However, it is likely that claims in our dataset are a result of fraud, and Skuld professionals confirm that they identify fraud from time to time. In particular claims such as shortage of cargo, might be related to theft and not damage or accidental mistakes. Other variables would likely be more suited to predict fraud, such as the inclusion of operators, specific ports or crew composition.

Regarding claims data, three event categories were used. As explained in Chapter 3, each claim may have several cases, and by exploiting case type in addition to Event type, a more profound insight could be achieved.

Choice of models is also likely to limit the performance. Algorithms that are easy to interpret and allow for extraction of variable importance were deliberately chosen. More advanced models could lead to higher predictive power, at the cost of the "black box" effect discussed previously. A clear distinction between models for use in loss prevention and risk identification could be utilised to solve for this, and by separating the models based on what they intend to find, overall performance could be increased.

7.5 Recommendations for model enhancement

In this section recommendation for further work and measures to increase model performance is addressed. The recommendations are based on observations and limitations revealed throughout the process.

Increased number of observations

Model performance would presumably increase with more observations (Burger, 2018). Two methods to increase the number of observations are possible, expand scope or increase resolution of available data. Expanding scope can be done by including more years of AIS-information or more vessels. This is in practice not a feasible solution as AIS-information before the year 2012 is limited, due to the lack of satellite-based AIS-systems, and Skuld is not able to "buy" claims history from other clubs[9]. Increased resolution can be achieved by splitting the data into shorter time periods. The current analysis is performed on an aggregated level for all years. There are several reasons as to why this method was chosen. The potential lag of activity towards claims, e.g. increased such as increased sailing in one year, may not result in more claims before next year, as maintenance has been put on hold. The other reason is related to the number of claims. 18,541 claims on 3,500 vessels for four event groups, leaves on average just above one claim per vessel for each category. Increasing the resolution would result in a very skewed dataset. Other reasons are related to the uncertainty of the incident date, the problem with vessels in lay-up during the winter seasons, and vessels starting long transits just before a new year.

Several of the issues mentioned above could be solved, and by increasing the resolution to one year, the number of observations would increase by about 3 times. The most favourable resolution would be to study claims on individual transits. Activity levels would need to be altered, by possibly including variables on activity level the last six months before the given voyage, or number of ports visited last month. A resolution per voyage could also allow for

[9] A project to collaborate on data analysis between Clubs was attempted, but failed.

including other variables such as operators, type of cargo transported and whether pilots were used or not.

Possible Gains from Increase Accuracy in AIS-Information

Windward (2014) identifies a limitation in manual updates of AIS-information. Model performance could increase significantly by more accurate AIS-information, such as the type of Cargo, next port of call and navigational status. The potential gain from increased accuracy in navigational status is outlined in section 3.4. An incentive from P&I Clubs to motivate Shipowner and OOW to increase attention on accurate AIS updates could lead to more accurate model predictions. An incentive could be implemented as a bonus system with a reduction in premium based on the accuracy of the AIS-information. Calculations on the value of accurate data should be performed to decide if the incentives-system is profitable.

Research Port-risk

Through this research, the object of study is vessels activity and operational patterns. It is possible to tweak the methods used, to study risk related to ports to a more considerable extent. By placing individual ports as the object of study, one could assess the risk of individual ports by analysing variables extracted from AIS-information related to traffic density, navigational bottlenecks, type of vessels and time in port. Thus, machine learning and AIS could be used to create an index including risk for all ports of interest.

7.6 Alternative variables

Through the research, multiple predictors have been reviewed, and two alternative dependent variables were considered. In this section, both alternative dependent and independent variable are discussed.

7.6.1 Alternative dependent variables

In the process of building the models, the variable *Claim/No Claim* where used as the dependent variable. Following are an elaboration of the two alternative variables retrieved from the claims data, with potential in machine learning modelling.

Cost of Claims

Cost of claims was tested in the initial phase and did not yield any useful results. In discussion with Skuld professionals, two major causes were assessed as the likely reasons. Firstly, the five most costly accidents account for 60 % of the total payouts during 2013-2017, and no correlation based on AIS were identified between these five accidents. Even if these five vessels were removed, the element of randomness was deemed too large. A simple example is that a grounded vessel could ground on a bank of sand, rocks or coral reef. Grounding on rocks compared to sand would naturally cause more damage to the vessel. Grounding on a coral reef would likely result in similar damage as on rocks, while the environmental effects would likely be higher, and claims from government related to clean-up could result in significantly higher costs. Thus, small but significant changes in where the accidents occur might result in considerable differences in liabilities.

Further reasoning why this variable was unsuitable is due to the time bar for certain claims, that can be 2-3 years. A large proportion of the claims from 2016 and 2017 are still open, meaning the final payout has not been decided. However, the potential of a model able to predict the size of claims would be of high value, and building a model predicting size of claims by including more variables should be exploited.

Number of Claims

An alternative to *Cost of Claims* and the categorical variable was *Number of Claims*. The benefit of the variable as compared to the categorical variable is the reflection of number of claims and allows for regression algorithms. In practice, the variable would only differ slightly from the categorical variable, as claims frequency is overall low, and for vessels with one claim the factor variable YES/NO would yield the same result. It was decided that the increased value from adding *Number of Claims* as a dependent variable would not be of great importance at this stage. However, exploiting number of claims further could be of great value in claims assessment.

7.6.2 Alternative predictors

Model performance could increase by including new variables, and this section discussed several alternative predictors. Some of the alternative variables require the analysis to be performed on transit level as opposed to the five year aggregated approach.

Information of Ownership

Previous claims statistics, financial strength, size of fleet and type of trade could all be relevant factors in predicting claims. Some of these factors are already utilised in determining premiums or when accepting new members into the club. By including predictors related to owners, the model would assess member specific issues, and new findings could be revealed.

Information on Trade
Information about the specific trade of each transit could be of potential value as certain types of goods are believed to be more exposed to risk. One example is the transportation of grain compared to oil, where the risk of pollution arises when carrying oil. By adding the type of goods to the model, the vessel-specific risk could be further assessed when predicting claims. Also, trade specific patterns could reveal additional factors connected to trading routes.

Information on Weather
Weather is known to impact trade at sea, and every year vessels are lost due to heavy weather (AGCS, 2018). Several variables are suitable to represent the possible impact of weather while sea state and wind are the most common. Sea state represents the wave height. Both variables are possible to extract from multiple sources; the challenge is the practical implementation in the model. At an aggregated level weather is not useful, and even on a transit-based level, the uncertainty of the incident date and time, makes it difficult to implement in practice.

Information on Crew:
Crew composition and level of training and experience could be a relevant factor to exploit. Multi-ethnicity crews are common, and language barriers could induce risk (Badawi & Halawa, 2014). In many types of trade, crew turnover is high, and the data is most suitable to be used on a transit-based level. However, this will introduce costs in data collection, and might not be feasible in practice.

Common Challenges
A common challenge to most of these predictors is the time lag between the build-up to an accident and the trigger of the accident. An underqualified crew in the previous year could lead to accidents in the following year, as maintenance was not performed according to standards. For activity level, this can be solved by making a variable for activity in a given period before a voyage, but for other variables, this is not possible.

8 Conclusion

This thesis aimed at determining the potential of applying machine learning in P&I insurance, by analysing the relationship between activity level and operational patterns against P&I insurance claims. A comprehensive pre-processing of AIS-information were performed, and variables were extracted from 52 million data points, combined with data on 5,000 ports. The process allowed for analysing five years of insurance data, including information on 3,500 vessels. The work contributes to current research by expanding the scope of literature on maritime accidents, to include any incident related to vessel activity.

To answer the research questions, five categorical machine learning algorithms were applied to predict the outcome of whether a vessel would encounter a claim or not in three different categories. A total of 75 models were built and compared to analyse algorithm and vessel group-specific variations. This allowed to control for multiple sources of errors and make the presidency for a robust conclusion. Findings from the analysis imply that activity level and operational patterns of a vessel have predictive power on P&I insurance claims. Three variables were found to contain the strongest predictive power; *Number of unique ports per month*, *Total distance sailed per month*, and *Average distance sailed per month*. The former variable is related to the operational pattern of vessels, while the two latter is a combination of operational pattern and activity-level of the vessel. Two major shortcomings of the analysis are the lack of predictive power for claims in the category Contact & Collison and the lack of predictive power for Container vessels.

A major challenging are to compute an economic value on the research, as predicting claims is one part of the equation, preventing claims from occurring are more complex. However, through business presentations, the two major providers of AIS analysis, Windwards and Concirrus, gives a point in the direction of potential savings. They assess that machine learning can reduce claims by 7-14%. This translate to a yearly saving of between \$ 20 - \$ 40 million. In addition to savings from loss prevention, the increased knowledge from AIS-information could help Skuld to create more accurate price models; thus, increase revenues from premiums.

While Skuld will continue to explore the potential of AIS and increased integration of fleet monitoring and activity-based decisions, the primary recommendation on further work is to rethink the processing of vessels at anchor, and possibly apply three states of operation; transit, port and anchor. In addition, future research should aim at increasing the resolution of the current data, to investigate activity level per transit. This could enhance the model performance by expanding number of observations and by facilitating for additional variables related to ownership, crew, weather, trade, and more.

Findings from the analysis are directly applicable in P&I insurance and can be implemented in automated real-time loss prevention, or directly as a parameter in risk assessment and calculation of insurance premiums. The research makes a strong argument for implementing activity-based pricing in P&I Insurance. The P&I Club Skuld have as a result of this research investigating how to practically implement risk parameters from the AIS-variables found in this thesis in their pricing strategy. Knowledge and experience attained through the process were directly applied when Skuld decided to enter into an agreement with the AIS analysis company Concirrus in December 2018.

9 References

AGCS (2018) Safety and shipping review 2018: An annual review of trends and developments in shipping losses and safety. Germany: Allianz global corporate & speciality

Ajigboye, O. (2016) A review of the doctrine of insurable interest under the marine insurance act in Nigeria. The gravitas review of business & property law, Vol.7, No.3.

Akosa, J. S. Predictive (2017) Accuracy: A Misleading Performance Measure for Highly Imbalanced Data, Oklahoma State University.

Akyuz, E., Celik, M. (2014) Utilization of cognitive map in modelling human error in marine accidents analysis and prevention. Safety Science, Volume 70, Pages 19-28.

Badawi, E., Halawa, A.M. (2014) Maritime Communication: The Problem of Cross Cultural Multilingual Crews. Alexandria: College of Maritime Transport & Technology

BMP4 (2011) Best Management Practices for Protection against Somalia Based Piracy. Version 4. Edinburgh: Whiterby Publishing Group Ltd.

Brækhus, S. & Rein, A. (1979) Handbook of P&I Insurance. 2nd ed. Arendal: Assuranceforeningen Gard-Gjensidig.

Burger, S.V. (2018) Introduction to Machine Learning with R. Sebastopol: O´Reilly Media.

Butt, N., Johnson, D., Pike K., Pruce-Roberts, N., Vigar, N. (2012) 15 Years of Shipping Accidents: A review for WWF. Southampton: Southampton Solent University

Cambridge Dictionary (2018). FLAG OF CONVENIENCE | meaning in the Cambridge English Dictionary. [online] Dictionary.cambridge.org. Available at: https://dictionary.cambridge.org/dictionary/english/flag-of-convenience [Accessed 4 Dec. 2018].

Cambridge Dictionary (2018) Insurance claim | Definition of "insurance claim" from the Cambridge Business English Dictionary. [online] Available at: https://dictionary.cambridge.org/dictionary/english/insurance-claim [Accessed 28 Nov. 2018].

CEFOR (2016) Global Risk Global Cover: An introduction to The Nordic Marine Insurance Plan of 2013, Version 2016. Oslo: The Nordic Association of Marine Insurers.

Chauvin, C., Lardjane, S., Morel, G., Clostermann, J. P., & Langard, B. (2013) Human and organizational factors in maritime accidents: Analysis of collisions at sea using the HFACS. Accident Analysis & Prevention, 59, 26-37.

Clarkson (1991). The Clarkson VLCC Quality Survey, Clarkson Research Studies Ltd., London.

Department of Defense (2008) Global Positioning System Standard Positioning Service Performance Standard. 4th ed. Washington D.C.: Department of Defense.

Eissa, M. (2018). Introduction to Machine Learning - CodeProject. [online] Codeproject.com. Available at: https://www.codeproject.com/Articles/1146582/Introduction-to-Machine-Learning [Accessed 19 Nov. 2018].

Federal Aviation Administration (2017) Global Positioning System (GPS) Standard Positioning Service (SPS) Performance Analysis Report. WAAS T&E Team. Report number: 96.

Gohlish, H.E .(2009) Charters' Liability Insurance: Essential Best Practice. London: Seamanship International Ltd.

Grech, M., Horberry, T., Koester, T. (2008) Human Factors in the Maritime Domain. CRC Press

Guizhen, Z., Thai, V. V. (2016) Expert elicitation and Bayesian Network modelling for shipping accidents: A literature review, Safety Science, Volume 87, Pages 53-62,

Harrald, J.R., Mazzuchi, T.A., Spahn, J., Van Dorp, R., Merrick, J., Shrestha, S., Grabowski, M. (1998) Using system simulation to model the impact of human error in a maritime system, Safety Science, Volume 30, Issues 1–2, Pages 235-247.

Hastie T., Tibshirani R., Friedman J. (2009) The Elements of Statistical Learning - Data Mining, Inference, and Prediction
(2nd ed.). New York: Springer.

James G., Witten D., Hastie T., Tibshirani R. (2013) An introduction to statistical learning with application in R. New York: Springer.

Haueter N. V. (2013) A History of Insurance. Switzerland: Swiss Reinsurance Company Ltd.

IMO (2018). Automatic Identification Systems (AIS). [online] Imo.org. Available at: http://www.imo.org/en/OurWork/Safety/Navigation/Pages/AIS.aspx [Accessed 4 Dec. 2018].

IG (2013) International Group Agreements 2013. International Group of P&I Clubs. Available at: https://static.mycoracle.com/igpi_website/media/article_attachments/IGA2013.pdf [Accessed 4 Dec. 2018].

ITF (2018). Flags of convenience. [online] International Transport Workers' Federation. Available at: http://www.itfglobal.org/en/transport-sectors/seafarers/in-focus/flags-of-convenience-campaign/ [Accessed 4 Dec. 2018].

Kuehmayer, J. R. (2008) Marine Accident and Casualty Investigation Boards. Australian Marine Equipment Manufacturers.

Kujala, P., Hanninen, M., Arola, T., Ylitalo, J. (2009) Analysis of the marine traffic safety in the Gulf of Finland, Volume 94, Issue 8, Pages 1349-1357.

Kystverket (2013) Hva er AIS? Havbase: Kystverket. Available at: https://havbase.kystverket.no/havbase_report/doc/AIS.pdf [Accessed 2 Nov. 2018].

Latarche, M. (2018) Shipping lags behind in new tech adoption, but changes is inevitable. Ship Insight, Retrieved from: https://shipinsight.com/articles/shipping-lags-behind-in-new-tech-adoption-but-change-is-inevitable-posidonia-2018-experts-agree [Accessed, 21.Nov.2018]

Lloyd´s List Intelligence (2017) Understanding AIS. London: Maritime Intelligence. Available at: https://maritimeintelligence.informa.com/~/media/informa-shop-window/mnl/files/lloyds_list_intelligence/understanding_ais.pdf [Accessed, 13.Nov.2018]

Luddeke, C. F. (1996) Marine Claims: A guide for handling and prevention of marine claims. 2nd edition, London: Lloyd's of London Press LTD.

MAIB (2004) Bridge Watchkeeping Safety Study. Southampton: Marine Accident Investigation Branch

Markets&Markets (2014). Automatic Identification System Market. [online] Marketsandmarkets.com. Available at: https://www.marketsandmarkets.com/Market-Reports/automatic-identification-system-market-138357200.html [Accessed 21 Nov. 2018].

Mazaheri, A. (2017) A framework for evidence-based risk modelling for ship groundings. Helsinki: Aalto University. Available at: https://aaltodoc.aalto.fi/browse?type=author&value=Mazaheri,%20Arsham [Accessed, 13.Nov.2018]

Mullai, A., Paulsson, U. (2011) A ground theory model for analyzing of marine accidents. Accidents Analysis & Prevention, Volum 43, Issue 4, Pages 1590-1603.

Mutenga, S. & Parsons, C. (2012) Marine Insurance. Tally, W. K. ed. Maritime Economics. West Sussex: Blackwell Publishing Ltd, p. 452-470.

North (2012) An introduction to P&I insurance and loss prevention. 2nd ed. North of England P&I Association

Oxford Dictionaries (2018) Accident | Definition of accident in English by Oxford Dictionaries. [online] Available at: https://en.oxforddictionaries.com/definition/accident [Accessed 22 Nov. 2018].

Oxford Dictionaries (2018) Accident | Definition of gross tonnage in English by Oxford Dictionaries. [online] Available at: https://en.oxforddictionaries.com/definition/accident [Accessed 01 Nov. 2018].

Skuld Samuel, A.L (1959) *Some studies in machine learning using the game of checkers*. IBM Journal of Research and Development, vol 3.

Skuld (2017) *An Introduction to P&I Insurance for Mariners*. Oslo: Assuranceforeningen Skuld

SOLAS, Safety of Navigation, Annex 17 – Automatic Identification Systems (AIS). Available at: http://solasv.mcga.gov.uk/annexes/Annex17.htm [Accessed 20 Nov. 2018].

Standard & Poor's report September 2018 - Assuranceforeningen SKULD (Gjensidig). Available at: https://www.skuld.com/contentassets/72ebc555ca5a491b81860558ebd5716b/standard_and_poors_report_sept2018.pdf [Accessed 25 Oct. 2018]

Stornes, P (2015) Risk influencing factors in maritime accidents, An explotory statistical analysis of the Norwegian Maritime Authority incident database. Trondheim: NTNU Social Research

UNCTAD (1982), Legal and documentary aspects of marine insurance contract. Geneva: United Nations Conference on Trade and Development. Available at: http://www.fortunes-de-mer.com/mer/images/documents%20pdf/divers/unctad%20legal%20aspects%20marine%20insurance.pdf [Accessed 15 Nov. 2018].

Viera, A. J., Joanne, M.D., Garrett, M. (2005) Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine, Vol. 37, No. 5 page 360-363.

Windward (ed.) (2014) AIS Data on the High Seas: An Analysis of the Magnitude and Implications of Growing Data Manipulation at Sea. Windward.

10 Appendix

A. Bulk & Cargo Small

Model performance - Logistic Classification Model

Model	Validation Performance		Test Performance	
Logistic Classification	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.7283	0.1459	0.7569	0.2408
Contact & Collision	0.8172	0.0090	0.8257	-0.0090
Human	0.6627	0.3180	0.6055	0.2056

Confusion Matrix - Logistic Classification Model

	Cargo			Contact&Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	23 %	4 %	**No**	99 %	100 %	**No**	64 %	43 %
Yes	77 %	96 %	**Yes**	1 %	0 %	**Yes**	36 %	57 %

Model Performance - Random Forest

Model	Validation Performance		Test Performance	
Random Forest	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.7718	0.3605	0.7798	0.3928
Contact & Collision	0.8144	0.0441	0.7982	-0.0571
Human	0.6749	0.3442	0.6422	0.2795

Confusion Matrix - Random Forest

	Cargo			Contact&Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	44 %	9 %	**No**	96 %	100 %	**No**	67 %	39 %
Yes	56 %	91 %	**Yes**	4 %	0 %	**Yes**	33 %	61 %

Model Performance - Naïve Bayes

Model	Validation Performance		Test Performance	
Naive Bayes	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.7429	0.3327	0.7706	0.4009
Contact & Collision	0.7439	0.0492	0.7523	0.1018
Human	0.6362	0.2737	0.6193	0.2477

Confusion Matrix - NB

	Cargo			Contact&Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	51 %	13 %	**No**	86 %	76 %	**No**	54 %	29 %
Yes	49 %	87 %	**Yes**	14 %	24 %	**Yes**	46 %	71 %

Model Performance - AdaBoost

Model	Validation Performance		Test Performance	
AdaBoost	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.7659	0.3016	0.7936	0.4007
Contact & Collision	0.8261	0.0000	0.8303	0.0000
Human	0.6800	0.3568	0.6284	0.2489

Confusion Matrix - AdaBoost

	Cargo			Contact&Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	39 %	5 %	**No**	100 %	100 %	**No**	68 %	43 %
Yes	61 %	95 %	**Yes**	0 %	0 %	**Yes**	32 %	57 %

Model Performance – Neural Network

Model	Validation Performance		Test Performance	
Neural Network	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.7362	0.2487	0.7706	0.3380
Contact & Collision	0.8261	0.0008	0.8303	0.0000
Human	0.6570	0.3122	0.5963	0.1920

Confusion Matrix – Neural Network

	Cargo			Contact & Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	36 %	7 %	**No**	100 %	100 %	**No**	59 %	40 %
Yes	64 %	93 %	**Yes**	0 %	0 %	**Yes**	41 %	60 %

B. Bulk & Cargo Large

Model performance - Logistic Classification Model

Model	Validation Performance		Test Performance	
Logistic Classification	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.6915	0.2631	0.7132	0.3137
Contact & Collision	0.8050	-0.0057	0.8088	0.0000
Human	0.5911	0.0999	0.6029	0.1292

Confusion Matrix – Logistic Classification Model

	Cargo			Contact & Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	39 %	10 %	**No**	100 %	100 %	**No**	30 %	18 %
Yes	61 %	90 %	**Yes**	0 %	0 %	**Yes**	70 %	82 %

Model Performance - Random Forest

Model	Validation Performance		Test Performance	
Random Forest	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.7269	0.3733	0.7132	0.3203
Contact & Collision	0.7952	0.0050	0.8088	0.0868
Human	0.6284	0.2119	0.6103	0.1897

Confusion Matrix – Random Forest

	Cargo			Contact & Collision			Human	
	No	Yes		No	Yes		No	Yes
No	41 %	11 %	No	98 %	92 %	No	49 %	30 %
Yes	59 %	89 %	Yes	2 %	8 %	Yes	51 %	70 %

Model Performance Naïve Bayes

Model	Validation Performance		Test Performance	
Naive Bayes	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.6873	0.2859	0.7132	0.3458
Contact & Collision	0.7615	0.0070	0.7868	0.0811
Human	0.6048	0.1274	0.6250	0.2046

Confusion Matrix – Naïve Bayes

	Cargo			Contact & Collision			Human	
	No	Yes		No	Yes		No	Yes
No	49 %	16 %	No	95 %	88 %	No	44 %	24 %
Yes	51 %	84 %	Yes	5 %	12 %	Yes	56 %	76 %

Model Performance - AdaBoost

Model	Validation Performance		Test Performance	
AdaBoost	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.7215	0.3507	0.7426	0.3779
Contact & Collision	0.8095	0.0000	0.8088	0.0000
Human	0.6434	0.2255	0.6397	0.2358

Confusion Matrix - AdaBoost

	Cargo No	Cargo Yes	Contact & Collision No	Contact & Collision Yes	Human No	Human Yes
No	41 %	7 %	100 %	100 %	46 %	23 %
Yes	59 %	93 %	0 %	0 %	54 %	77 %

Model Performance - Neural Network

Model	Validation Performance		Test Performance	
Neural Network	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.7161	0.3482	0.7132	0.3396
Contact & Collision	0.8095	0.0000	0.8088	0.0000
Human	0.6291	0.2119	0.5956	0.1633

Confusion Matrix – Neural Network

	Cargo No	Cargo Yes	Contact & Collision No	Contact & Collision Yes	Human No	Human Yes
No	47 %	15 %	100 %	100 %	49 %	33 %
Yes	53 %	85 %	0 %	0 %	51 %	67 %

C. Tankers Small

Model performance - Logistic Classification Model

Model	Validation Performance		Test Performance	
Logistic Classification	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.6388	0.2028	0.6443	0.2050
Contact & Collision	0.7957	-0.0040	0.7990	0.0000
Human	0.6355	0.2234	0.6701	0.2972

Confusion Matrix – Logistic Classification Model

	Cargo		Contact & Collision		Human	
	No	Yes	No	Yes	No	Yes
No	27 %	8 %	100 %	100 %	48 %	19 %
Yes	73 %	92 %	0 %	0 %	52 %	81 %

Model Performance - Random Forest

Model	Validation Performance		Test Performance	
Random Forest	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.6938	0.3603	0.6546	0.2794
Contact & Collision	0.7819	0.0251	0.8041	0.0685
Human	0.6798	0.3256	0.5928	0.1494

Confusion Matrix – Random Forest

	Cargo		Contact & Collision		Human	
	No	Yes	No	Yes	No	Yes
No	52 %	25 %	99 %	95 %	46 %	31 %
Yes	48 %	75 %	1 %	5 %	54 %	69 %

Model Performance – Naïve Bayes

Model	Training Performance		Test Performance	
Naive Bayes	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.6739	0.3014	0.6546	0.2572
Contact & Collision	0.7275	0.0626	0.7165	0.0340
Human	0.6513	0.2589	0.6186	0.1990

Confusion Matrix – Naïve Bayes

	Cargo		Contact & Collision		Human	
	No	Yes	No	Yes	No	Yes
No	41 %	17 %	85 %	82 %	47 %	27 %
Yes	59 %	83 %	15 %	18 %	53 %	73 %

Model Performance – AdaBoost

Model	Validation Performance		Test Performance	
AdaBoost	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.6949	0.3503	0.6856	0.3283
Contact & Collision	0.7943	.	0.7990	0.0000
Human	0.6700	0.3004	0.6134	0.1838

Confusion Matrix – AdaBoost

	Cargo		Contact & Collision		Human	
	No	Yes	No	Yes	No	Yes
No	41 %	17 %	85 %	82 %	47 %	27 %
Yes	59 %	83 %	15 %	18 %	53 %	73 %

Model Performance – Neural Network

Model	Validation Performance		Test Performance	
Neural Network	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.6856	0.3298	0.6546	0.2622
Contact & Collision	0.7943	0.0000	0.7990	0.0000
Human	0.6526	0.2533	0.5979	0.1308

Confusion Matrix – Neural Network

	Cargo			Contact & Collision			Human	
	No	Yes		No	Yes		No	Yes
No	44 %	19 %	No	98 %	91 %	No	35 %	22 %
Yes	56 %	81 %	Yes	2 %	9 %	Yes	65 %	78 %

D. Tankers Large

Model performance - Logistic Classification Model

Model	Validation Performance		Test Performance	
Logistic Classification	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.6051	0.1020	0.6800	0.2701
Contact & Collision	0.8554	-0.0110	0.8667	0.0000
Human	0.6764	0.2125	0.7333	0.3590

Confusion matrix - Logistic Classification Model

	Cargo			Contact & Collision			Human	
	No	Yes		No	Yes		No	Yes
No	81 %	56 %	No	25 %	33 %	No	25 %	33 %
Yes	19 %	44 %	Yes	75 %	67 %	Yes	75 %	67 %

Model performance – Random Forest

Model	Validation Performance		Test Performance	
Random Forest	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.6243	0.1231	0.7067	0.3529
Contact & Collision	0.8439	0.0558	0.8533	-0.0248
Human	0.6276	0.1189	0.6933	0.2694

Confusion Matrix – Random Forest

	Cargo			Contact & Collision			Human	
	No	Yes		No	Yes		No	Yes
No	79 %	44 %	No	98 %	100 %	No	37 %	13 %
Yes	21 %	56 %	Yes	2 %	0 %	Yes	63 %	88 %

Model Performance – Naïve Bayes

Model	Validation Performance		Test Performance	
Naive Bayes	Accuracy	Kappa	Accuracy	Kappa
Cargo	0.6111	0.2349	0.6267	0.2723
Contact & Collision	0.7707	0.1047	0.7733	-0.0241
Human	0.6734	0.2047	0.7067	0.2949

Confusion Matrix – Naïve Bayes

	Cargo			Contact & Collision			Human	
	No	Yes		No	Yes		No	Yes
No	56 %	26 %	No	88 %	90 %	No	37 %	10 %
Yes	44 %	74 %	Yes	12 %	10 %	Yes	63 %	90 %

Model Performance – AdaBoost

Model	Validation Performance		Test Performance	
AdaBoost	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.6565	0.2097	0.7333	0.4118
Contact & Collision	0.8548	0.0210	0.8667	0.0000
Human	0.6498	0.0672	0.6400	0.0203

Confusion Matrix – AdaBoost

	Cargo No	Cargo Yes		Contact & Collision No	Contact & Collision Yes		Human No	Human Yes
No	85 %	52 %	**No**	100 %	100 %	**No**	4 %	6 %
Yes	15 %	48 %	**Yes**	0 %	0 %	**Yes**	96 %	94 %

Model Performance – Neural Network

Model	Validation Performance		Test Performance	
Neural Network	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.6291	0.1771	0.6267	0.1898
Contact & Collision	0.8548	0.0000	0.8667	0.0000
Human	0.6862	0.2450	0.7067	0.3309

Confusion Matrix – Neural Network

	Cargo No	Cargo Yes		Contact & Collision No	Contact & Collision Yes		Human No	Human Yes
No	85 %	52 %	**No**	100 %	100 %	**No**	48 %	17 %
Yes	15 %	48 %	**Yes**	0 %	0 %	**Yes**	52 %	83 %

E. Container

Model Performance – Logistic Classification Model

Model	Validation Performance		Test Performance	
Logistic Classification	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.5849	0.1293	0.5417	-0.0017
Contact & Collision	0.8371	-0.0117	0.8333	-0.0261
Human	0.5792	0.1530	0.5972	0.1714

Confusion Matrix - Logistic Classification Model

	Cargo			Contact & Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	19 %	20 %	**No**	98 %	100 %	**No**	42 %	26 %
Yes	81 %	80 %	**Yes**	2 %	0 %	**Yes**	58 %	74 %

Model Performance - Random Forest

Model	Validation Performance		Test Performance	
Random Forest	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.5618	0.0809	0.5000	-0.0791
Contact & Collision	0.8352	-0.0005	0.8472	0.1121
Human	0.6381	0.2665	0.5972	0.1675

Confusion Matrix – Random Forest

	Cargo			Contact & Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	19 %	27 %	**No**	98 %	91 %	**No**	39 %	23 %
Yes	81 %	73 %	**Yes**	2 %	9 %	**Yes**	61 %	77 %

Model Performance - Naive Bayes

Model	Validation Performance		Test Performance	
Naive Bayes	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.5710	0.1142	0.5139	0.0047
Contact & Collision	0.7867	0.0040	0.8056	0.1173
Human	0.6219	0.2333	0.5833	0.1449

Confusion Matrix - Naive Bayes

	Cargo			Contact & Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	42 %	41 %	**No**	92 %	82 %	**No**	42 %	28 %
Yes	58 %	59 %	**Yes**	8 %	18 %	**Yes**	58 %	72 %

Model Performance - AdaBoost

Model	Validation Performance		Test Performance	
AdaBoost	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.5983	0.1300	0.5278	-0.0364
Contact & Collision	0.8464	0.0000	0.8472	0.0000
Human	0.6116	0.2091	0.6389	0.2589

Confusion Matric Ada-Boost

	Cargo			Contact & Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	16 %	20 %	**No**	100 %	100 %	**No**	48 %	23 %
Yes	84 %	80 %	**Yes**	0 %	0 %	**Yes**	52 %	77 %

Model Performance - Neural Network

Model	Validation Performance		Test Performance	
Neural Network	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Cargo	0.5901	0.1313	0.5417	-0.0277
Contact & Collision	0.8464	0.0000	0.8472	0.0000
Human	0.6065	0.2000	0.4861	-0.0423

Confusion Matrix - Neural Network

	Cargo			Contact & Collision			Human	
	No	**Yes**		**No**	**Yes**		**No**	**Yes**
No	10 %	12 %	**No**	100 %	100 %	**No**	39 %	44 %
Yes	90 %	88 %	**Yes**	0 %	0 %	**Yes**	61 %	56 %

F. Model Performance Cargo for all Models and Groups