

Article

Shipping Accidents Dataset: Data-Driven Directions for Assessing Accident's Impact and Improving Safety Onboard

Panagiotis Panagiotidis * , Kyriakos Giannakis, Nikolaos Angelopoulos and Angelos Liapis

Konnekt-able Technologies Ltd., X91 W0XW Waterford, Ireland; kgiannakis@konnektable.com (K.G.); nangelopoulos@konnektable.com (N.A.); aliapis@konnektable.com (A.L.)

* Correspondence: ppanagiotidis@konnektable.com

Abstract: Recent tragic marine incidents indicate that more efficient safety procedures and emergency management systems are needed. During the 2014–2019 period, 320 accidents cost 496 lives, and 5424 accidents caused 6210 injuries. Ideally, we need historical data from real accident cases of ships to develop data-driven solutions. According to the literature, the most critical factor to the post-incident management phase is human error. However, no structured datasets record the crew's actions during an incident and the human factors that contributed to its occurrence. To overcome the limitations mentioned above, we decided to utilise the unstructured information from accident reports conducted by governmental organisations to create a new, well-structured dataset of maritime accidents and provide intuitions for its usage. Our dataset contains all the information that the majority of the marine datasets include, such as the place, the date, and the conditions during the post-incident phase, e.g., weather data. Additionally, the proposed dataset contains attributes related to each incident's environmental/financial impact, as well as a concise description of the post-incident events, highlighting the crew's actions and the human factors that contributed to the incident. We utilise this dataset to predict the incident's impact and provide data-driven directions regarding the improvement of the post-incident safety procedures for specific types of ships.

Keywords: marine accidents; dataset; accident reports; human error; incident's impact; contingency plans; machine learning



Citation: Panagiotidis, P.; Giannakis, K.; Angelopoulos, N.; Liapis, A. Shipping Accidents Dataset: Data-Driven Directions for Assessing Accident's Impact and Improving Safety Onboard. *Data* **2021**, *6*, 129. <https://doi.org/10.3390/data6120129>

Academic Editor: Giuseppe Ciaburro

Received: 22 October 2021

Accepted: 29 November 2021

Published: 3 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to the European Maritime Safety Agency, thousands of people were injured and hundreds died in marine accidents during the last decades, indicating the importance of safety onboard [1]. Crew members and passengers have been wounded in two ways, either immediately by the accident's impact or the post-incident crisis. Despite the established plans/protocols [2], many omissions have occurred when applied by the crew members [3] due to the post-incident turmoil. Specifically, post-incident management is a stressful process, especially for the crew, leading to bad decision-making [4].

Safety for people onboard and accident management constitute timeless issues, resulting in the establishment of several organisations, such as the International Maritime Organisation (IMO) (<https://www.imo.org/>, accessed on 20 November 2021) that introduced the Safety of Life at Sea Convention (SOLAS). In recent decades, technological devices, e.g., sensors, are used for an accurate ship evacuation analysis to provide the crew with a more precise assessment of the ship's situation [5]. However, recent accident cases, such as the Costa Concordia, caused the loss of many human lives, indicating the need for developing new, more effective safety procedures and emergency management systems [6].

There are many cases where the crew members underestimated or did not properly assess the accident conditions during the implementation of emergency procedures in the post-incident phase. A representative case is a fire that burst out aboard on 17 August 2016 at the Caribbean Fantasy passenger vessel. During the evacuation, the ineffective

crew's actions caused the injuries of 49 passengers. Specifically, the crew members did not consider the wind speed, wind direction, or height of the waves, resulting in a steep slope of the Marine Evacuation System and the lifeboats' hitting to the side of the Caribbean Fantasy [7]. However, these data are in text format (i.e., reports), and their further usage requires the prior process of extracting this information. Creating such structured datasets for marine incidents is challenging, and there are many restrictions to the use of the already existing ones [2].

Our work is developed in the context of the Palaemon (EUROPEAN UNION's Horizon 2020) project, whose vision is to build a sophisticated mass centralised evacuation system (<https://palaemonproject.eu/about-palaemon/>, accessed on 20 November 2021). The main contributions of this work are the following. First, we provide a high-quality dataset that combines attributes such as the ship's technical characteristics, the weather conditions, a description of the accident including the crew's actions applied in the post-incident phase, the human factors that contributed to the incident, and the attributes related to the environmental/financial impact of each incident. To the of our knowledge, this is the first time a dataset includes characteristics related to the accident's conditions (e.g., weather, cause, etc.), the post-incident management process (e.g., successful/failed evacuation of the ship, the crew actions, etc.), the human factors that contributed to each incident's occurrence, and the corresponding environmental/financial impact. We also describe the framework followed by domain experts to convert all the unstructured information in accident reports into a structured format. This way, we provide domain experts/professionals with guidelines for enhancing the proposed dataset with new cases from additional reports. The dataset also provides the opportunity to improve the safety procedures during marine accidents. In particular, we identify specific types of ships and accidents with weaknesses in post-incident management procedures that need to be updated. Consequently, we pave the way for researchers and domain experts to introduce data-driven emergency management systems, (e.g., discovering effective/ineffective action patterns based on weather conditions, causal inference regarding accident conditions, human errors, etc.) [8]. Finally, we present an experimental study for the prediction of the incidents' environmental/financial impact using machine learning algorithms achieving remarkable performance.

The paper's structure is as follows. Section 2 presents existing datasets (Section 2.1) and their role in developing data-driven solutions (Section 2.3), highlighting the differences between the previous datasets and the proposed one (Section 2.2). Section 3.1 describes the dataset's construction process step by step, and Section 3.2 presents a statistical analysis of it. Section 4 presents the results for the prediction of the incidents' environmental/financial impact (Section 4.1) and the findings of the data analysis for the weaknesses of the post-incident management procedures (Section 4.2). Finally, Section 5 discusses conclusions and future data-driven directions to strengthen the safety onboard.

2. Related Work

2.1. Previous Datasets

Many governmental organisations, such as the National Transportation Safety Board (NTSB) (<https://www.nts.gov/Pages/default.aspx>, accessed on 20 November 2021), Japan Transportation Safety Board (JTSB) (<https://www.mlit.go.jp/jtsb/marrep.html>, accessed on 20 November 2021), and Marine Accident Investigation Branch (MAIB) (<https://www.gov.uk/government/organisations/marine-accident-investigation-branch>, accessed on 20 November 2021) record detailed reports regarding marine accidents. These organisations provide reports to their databases describing each accident; these reports are an important source to retrieve complete and reliable accident data [9]. However, such information is unstructured. There are also limited open-source structured datasets for marine accidents; in any case, the contact with the respective authorities is a prerequisite for the dataset's availability providing details regarding the usage purposes. In [10], the authors mentioned that they manage to receive structured datasets after several months of contact-

ing the responsible authorities. In the scientific literature, these datasets are used mostly for statistical analysis [11].

One of the most complete and detailed databases is the SOS, one of the SMD's (Swedish Maritime Department) databases. This database operates on the Microsoft SQL Server and contains approximately 6000 marine accident reports and provides the corresponding information for each accident in a structured format [12]. The description of each accident includes the accident's date, location, the total number of the crew members and passengers onboard, the event type (e.g., fire), the weather conditions at the accident time, information about the ship's cargo (e.g., chemicals), the number of deaths or injuries, the extent of the environmental damage, other technical ship's characteristics (e.g., length, construction material). Furthermore, the IMO (International Marine Organisation) provides an open-source database with well-structured information about each accident recording variables such as the accident time and coordinates, the initial event, the causality type, the ship type, the weather conditions, a summary of each accident events, etc. [13]. Our work is similar to the one of [14], which manually extracts information from 500 marine accident reports using multiple sources to identify the accident consequences and related contributing factors.

2.2. Comparison of the Existing Datasets with the Proposed One

To the best of our knowledge, this is the first time that a dataset includes the *human factors* that contributed to each incident's occurrence as well a short description of the accident that *focuses on the crew's actions* in addition to the weather conditions, the ship's technical characteristics, the incident's location, etc. The literature review illustrates that human factors are the primary reasons for maritime accidents [15]. The proposed dataset is an excellent opportunity for researchers to focus on human-related activities in ships' emergency management operations. This way, it could contribute to establishing new, more effective procedures targeting the increase of the safety level [16]. Even though IMOs, SOS, and other databases are valuable sources of information, they do not record the crew's actions during the incident or the human factors that contribute to the incident's occurrence, preventing researchers from analysing the contribution of the human factor (which is the most important ones [17]) to the accident occurrence and management. Furthermore, the proposed dataset is open without any limitations in usage. Our ambition is to create an open-source database to which anyone following the instructions (described in Section 3.1) could add new accident cases in a structured format. This process may encourage researchers to establish new emergency management systems by applying several techniques (e.g., information retrieval and machine learning) to overcome safety issues (e.g., real-time decision making). Thus, the proposed dataset bridges the gap between safety onboard and the maritime transportation industry providing operational data for safety analysis [18].

2.3. The Role of Data in Onboard Safety Enhancement

Recent evacuation systems use real-time data to facilitate the evacuation process. For instance, Reference [19] proposes an approach that uses data from smart bracelets and sensors (e.g., collected via smart cameras) to identify the passengers' position at the ship. However, domain experts have highlighted the usefulness of historical data by introducing various emergency management systems for ships that combine real-time data and data from past accident cases. For instance, the Smart Escape Support System combines real-time sensor data with historical data of the ship's escape routes to generate a faster and safer route for each passenger to reach the master station [5]. Furthermore, the Decision Support System (DSS) of [20], for navigation under rough weather, compares motion-related real-time data with data (e.g., rolling, pitch) from a database to support the captain to identify possible consequences of each navigation action, e.g., the estimation of the damage caused to the hull due to high waves considering the ship's route changes. However, the lack of operational or experimental data is still an open issue. In [19], the authors mention

that although we could estimate the time to evacuate up until the master station, it is impossible to do the same for the lifeboat boarding stage without adequate operational or historical data.

3. Data Collection

3.1. Dataset Description

The National Transportation Safety Board (NTSB), the Japan Transportation Safety Board (JTSB), and the Marine Accident Investigation Branch (MAIB) databases were utilised to retrieve the accidents' reports. The reports' structure is more or less the same. They begin with a summary, which briefly explains the event and the investigators' findings. The reports continue with a general description of the vessel and the accident, including the causes. The conclusion presents the event's chain, the accident's underlying factors, and some recommendations to improve maritime safety (see Figures A1 and A2 and [21]).

Five experts from the naval industry and safety science field collaborated to the Naval dataset creation. Specifically, each report was inspected by at least two experts (i.e., findings cross-check). The experts created the dataset in two stages. The process followed during the first stage encoded the basic characteristics of the 348 reports, i.e., the unique id of each report (Unique Id), the accident type (Accident Type), the vessel's name (Vessel Name), the date of the accident (Date), the vessel's length (Vessel Length), the vessel's type (Vessel Type), and the total number of persons onboard (Persons Onboard), ignoring further details about the weather conditions, crew's actions, etc. Table 1 shows the final set of attributes and their data types in a structured way. The first version of the dataset is named Naval dataset.

Table 1. The first version of the structured dataset. The first row shows the attributes names and the second row is a description of the corresponding data type.

Unique Id	Vessel Name	Vessel Type	Vessel Length	Accident Type	Persons Onboard	Date
A Serial Number	String	String	Numeric in meters	String	Integer	MM/DD/YYYY

Our motivation to create a dataset that includes characteristics related to *both the accident's conditions and the post-incident management process* guided us to the second stage. At this stage, the experts created a refined (final) version of the initial dataset, the Naval_v2 dataset, based on the information extracted in the previous step (creation of Naval dataset). This version of the dataset focuses on extracting accident characteristics related to safety, e.g., crew's actions during an evacuation, the human factors contributing to the accident's occurrence, and other details, e.g., the weather conditions. Specifically, the experts kept only the Naval dataset's accident cases whose value of the attribute Persons Onboard is greater than zero, i.e., 249 out of 348 reports were again inspected for a more detailed recording of their characteristics. This filtering criterion allows us to identify the effectiveness of the crew's actions based on the number of human casualties at the post-incident management phase.

In this second stage, we create a second version of the dataset (more refined compared to the first one) that includes all the available attributes of the previous version, as well as the weather conditions, a short description of the accident including the crew's actions, the number of crew members and passengers onboard, the number of deaths and injuries, the place that the accident happened, the accident's economic/environmental impact, and the human factors that contributed to each incident's occurrence.

Table 2 shows the separation of the information into fourteen primary categories: Unique Id, Date, Ship Attributes, Weather Attributes, Accident Type, Impact Attributes, Accident Description, Effectiveness, Place, Secondary effects of the initial incident, General human and organisational factors, Human and organisational factors based on incident type, Environmental Pollution, and Economical Impact. Specifically, the Unique Id attribute is the identifier of each accident. Furthermore, Ship Attributes consist of six sub-categories: Length (indicating the ship's length in meters), the Vessel Type (e.g., cargo, fishing, cruise),

the No. of Crew Members (that is, the number of crew members onboard), the No. of Passengers (that is, the number of passengers onboard), the No. of Persons Onboard (that is, the total number of persons onboard), and the Vessel Name.

Weather Attributes has seven sub-categories. The Rain attribute takes the boolean value 1 if it rains; otherwise it is equal to 0. The Wind Speed attribute is a numeric value that indicates the wind speed gusts in m/s at the accident time. Moreover, the Visibility attribute consists of single numbers that indicate the maximum value in meters that the crew could see. Additionally, in a few cases, there is a string description of the visibility situation. The Water Temperature and the Air Temperature attributes consist of single numbers that indicate the temperature in Kelvin at the accident time. Furthermore, the Wind Direction attribute is a string value that indicates the wind's direction. The Sea State attribute indicates an interval with minimum and maximum values regarding the height of the waves in meters and also, in a few cases, a string description of the sea's situation.

The Accident Type category describes the event type (e.g., fire, grounding). The Impact Attribute, the Accident Description, the General human and organisational factors, and the Human and organisational factors based on incident type categories are directly connected with the safety onboard. The Impact Attribute shows the number of deaths and injuries for each accident. The Accident Description attribute indicates a short description of the accident; the crew's actions are separated in brackets (e.g., ... [The securities check if any water tide off.] The ship remains in red condition until it gets to the dock. [The engineers check if everything's okay.]). The General human and organisational factors are factors that contribute to the accident's occurrence. Each of the twenty-four factors has a unique encoding that consists of three parts. The first and the second parts are the same for all the factors. These two parts indicate that the categorisation did not take into consideration the accident type and begin with HFACS-MA (i.e., Human Factors Analysis and Classification System for Maritime Accidents) according to [15]. The last part of the encoding consists of a unique number for each factor from one to twenty-four (see Table A1). Furthermore, the Human and organisational factors refer to groundings, collisions, machinery space fires, and explosions accident types. Furthermore, each of these factors has a unique encoding that consists of three parts. The first and the second part are the same for the factors referring to a specific incident type, i.e., HFACS-Ground for the groundings, HFACS-Coll for the collisions, and HFACS-MSS for the machinery space fires and explosions according to [22–24]. The last part of the encoding consists of a unique number for each of these three accident types, a number from one to twenty-four for the groundings, from one to twenty for the collisions, and from one to twenty-six for the machinery space fires and explosions (see Table A2).

The Place category consists of three attributes: the first is a text description of the place where the accident occurred; the second is the Location Type and takes five categorical values (i.e., 0: The accident happened at the open sea, 1: The accident happened at the port, 2: The accident happened at a gulf or a canal, 3: The accident happened at the river and 4: The accident happened at the lake); and the last attribute is the Place Geo-location that indicates the coordinates of the place where the accident happened. The Secondary effects of the initial incident are the effects that the incident had on the ship, e.g., if the ship sunk after a grounding, the secondary effect is the ship's sinking. The Environmental Pollution takes the value 1 if environmental damage occurred after the accident, else it is 0. The Economic Impact consists of two attributes. The first is the Damage to a vessel, which indicates the damage to the vessel in dollars. The second is the Damage to facilities, showing the damage caused to infrastructures in dollars (e.g., damage to the port). The Date attribute corresponds to the date that the accident happened. The Effective attribute takes the boolean value 1 if no one was injured, else it is 0. Based on the previous definition, the Naval_v2 dataset consists of 144 Effective and 105 Ineffective cases.

Table 2. The attribute categories, sub-categories, type, and measurement unit of Naval_v2 dataset.

Basic Attributes	Basic Attributes Categories	Attributes Type	Measurement Unit
Unique Id	A Serial Number	Numeric	-
Date	-	MM/DD/YYYY	-
Ship Attributes	Length	Numeric	Meters
	Vessel Type	String	-
	No of Crew Members	Numeric	-
	No of Passengers	Numeric	-
	No of Person Onboard	Numeric	-
	Vessel Name	String	-
Weather Attributes	Rain	Boolean	1 or 0
	Wind Speed	Numeric	m/s
	Wind Direction	String	-
	Water Temperature	Numeric	K
	Air Temperature	Numeric	K
	Visibility	Numeric	Meters
	Sea State	Numeric	Meters
Accident Type	-	String	-
Impact Attributes	No of Deaths	Numeric	-
	No of Injuries	Numeric	-
Accident Description	-	String	-
Effective	-	Boolean	1 or 0
Place	A brief description of the place	String	-
	Location Type	Categorical	0–4
	Place Geo-location	Longitude, Latitude	-
Secondary effects of the initial incident	-	String	-
General Human and organisational factors	-	String	-
Human and organisational factors based on incident type	-	String	-
Environmental Pollution	-	Boolean	1 or 0
Economic Impact	Damage on vessel	Numeric	Dollars
	Damage on facilities	Numeric	Dollars

To the best of our knowledge, this is the first time that a dataset (i.e., Naval_v2) enables researchers to discover the relation between the accident's conditions and the post-incident management or between the accident and the economic/environmental consequences, etc. This process may result in the establishment of new, more effective emergency management systems. The two versions of the dataset, i.e., the preliminary Naval dataset and the final Naval_v2 dataset are available here (<https://zenodo.org/record/5592999>, accessed on 20 November 2021).

3.2. Statistical Analysis of the Dataset

This section presents a statistical analysis of the Naval_v2 dataset. This analysis uses charts, a timeline analysis, a map with the accidents' places, and statistics about the human factors contributing to the accident. Figure 1 shows the percentages of the ship types in the dataset: 29.7% of the sample are Fishing Vessels, 16.2% Towing Vessels, 14.9% Passenger Vessels, 9.2% General Cargoes, 7.6% Bulk Carriers, 6% Tankers, and 3.2% Cruise Ships, and smaller percentages correspond to other ship types (13.2%). Figure 2a shows the distribution per accident type in the dataset: 22.1% of cases are collisions, 20.5% are machinery fires and explosions, 16.5% are groundings, and 8.4% heavy weather damages, and the other cases belong to other accident types (32.5%). Figure 2b shows the distribution of the location types that the accident occurred in the dataset: 37.3% of cases occurred at open seas, 26.5% of the cases occurred at gulfs or canals, 16.5% at rivers, 16.1% at ports, and 3.2% at lakes.

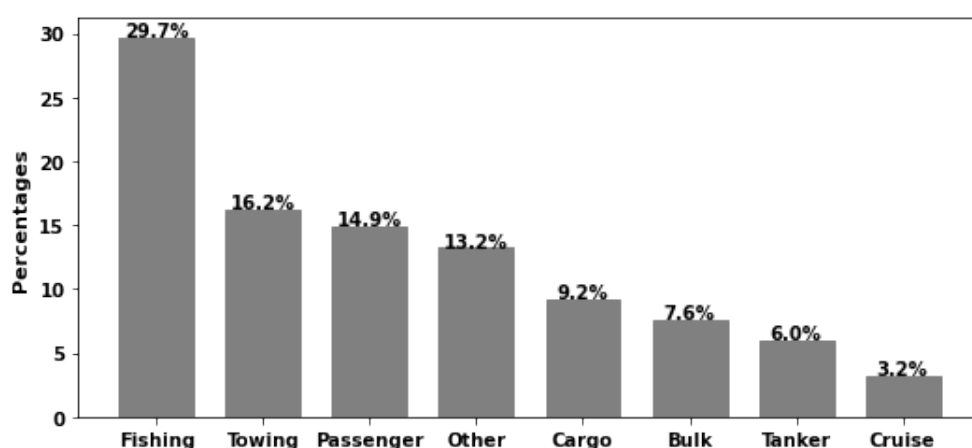


Figure 1. Distribution of ship types in the dataset.

Figure 3 shows the distribution of deaths and injuries with respect to wind speed. In this case, we split the dataset into two categories. The first category includes the accidents with light or moderate wind speed conditions, i.e., (0, 34] Knots, whereas the second one includes the cases with strong wind speed conditions (<https://www.canada.ca/en/environment-climate-change/services/general-marine-weather-information/understanding-forecasts/beaufort-wind-scale-table.html>, accessed on 20 November 2021), i.e., (34, 130] Knots. In the first category, the majority (85%) of the persons on board are injured, while the deaths' percentage is equal to 15%. On the contrary, the percentage of deaths in the second category is much higher (79%) due to the worse wind conditions.

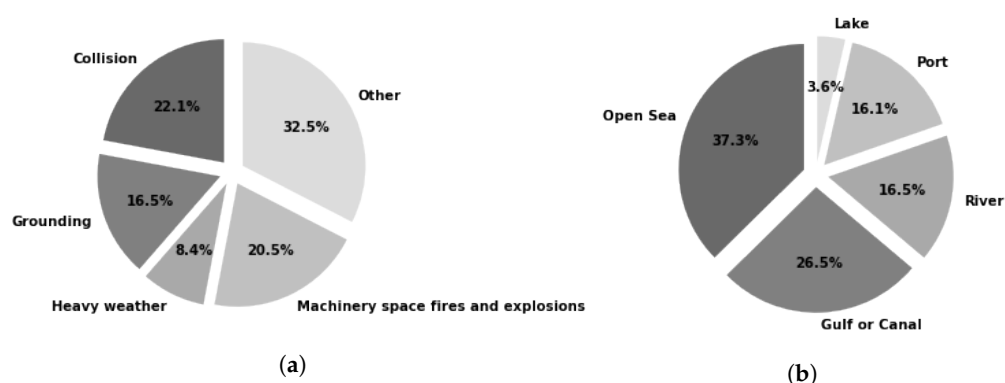


Figure 2. (a) Distribution per accident type in the dataset. (b) Distribution of the location types where the accidents occurred.

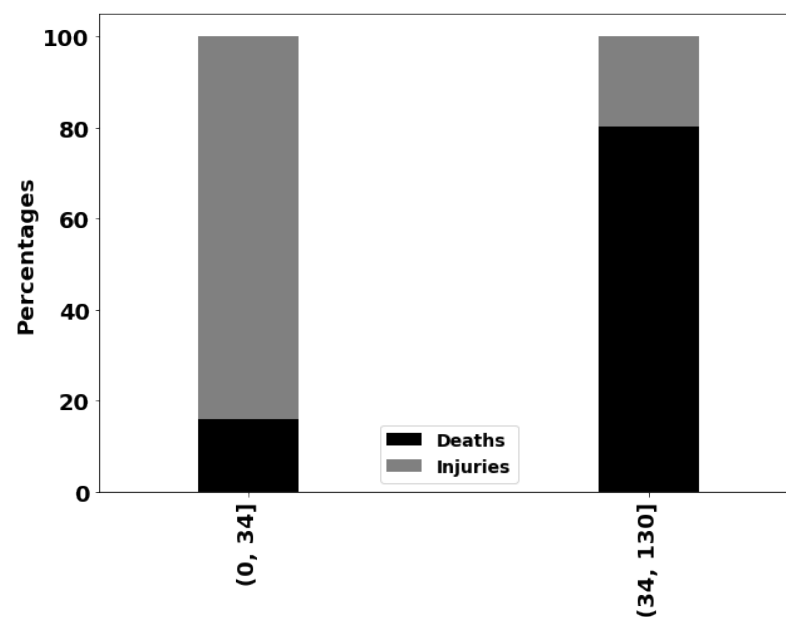


Figure 3. Distribution of deaths and injuries regarding the wind speed conditions.

Figure 4 shows the distribution of deaths and injuries based on visibility conditions during the accident. In this case, the dataset is split into four categories: good (more than 5 nautical miles), moderate (between 2 and 5 nautical miles), poor (between 1 and 2 nautical miles), and very poor or foggy visibility conditions (less than 1 nautical mile (<https://www.metoffice.gov.uk/weather/guides/coast-and-sea/glossary>, accessed on 20 November 2021)). For the third and especially for the fourth category, deaths overcome the injuries percentage, with 88% of persons onboard suffering deadly injuries with very poor or foggy visibility conditions. In conclusion, Figures 3 and 4 confirm that the visibility and wind speed conditions are strongly related to fatality [14], as accidents that happened under adverse weather conditions had more human losses.

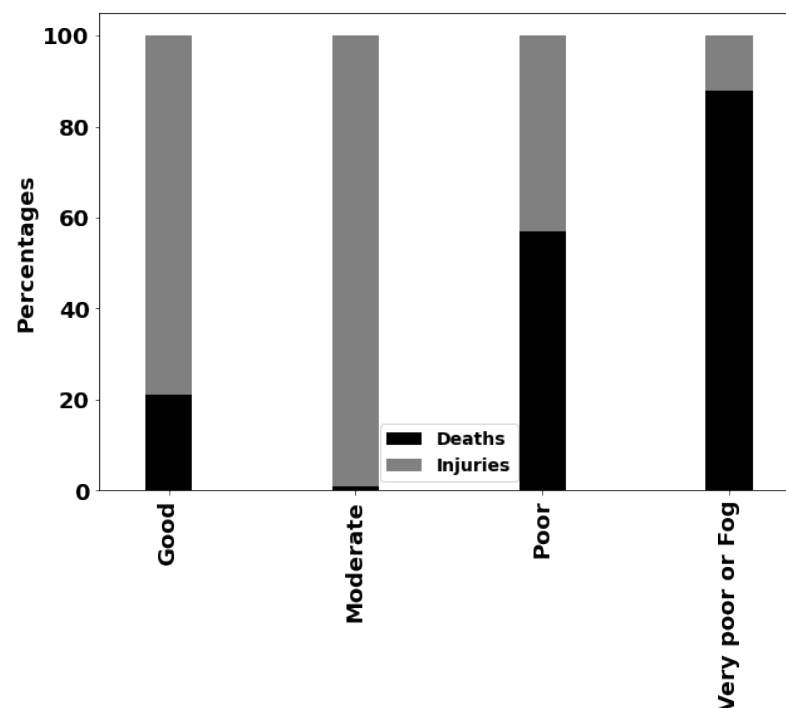


Figure 4. Distribution of deaths and injuries regarding the visibility conditions.

Figure 5 shows the number of accidents for the period 1983–2020. There is a clear increasing trend to the number of accidents from the year 2010. The increasing trend may be due to the improvement of the reporting procedures, resulting in an increased number of reports [11]. Figure 6 indicates all the places where accidents happened. This figure shows each accident place, including rivers, lakes, and sea areas.

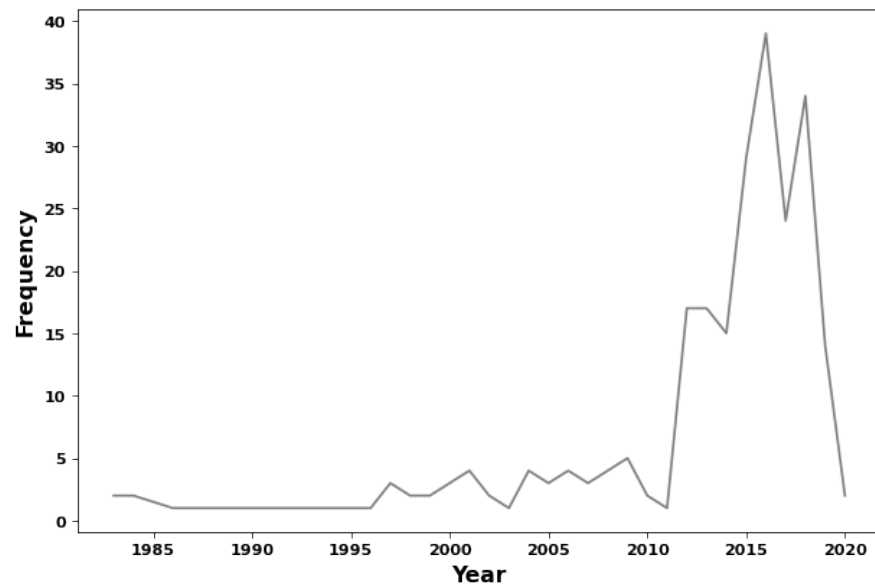


Figure 5. Timeline for the total number of accidents for the period 1983–2020.

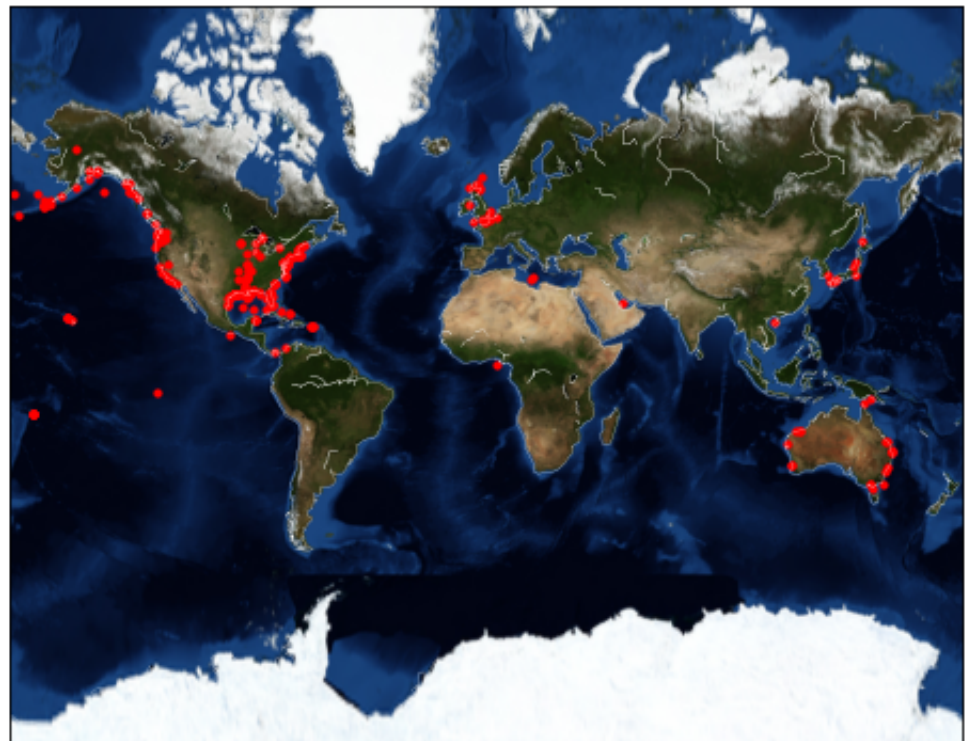


Figure 6. Map of the accident locations for the period 1983–2020.

Finally, we provide some statistics about the human factors that contribute to the incident's occurrence. Table 3 shows the percentages of the general human and organisational factors that contribute to incidents' occurrence. The most frequent factor is asset management (HFACS-MA-5) with 17.32%, followed by the decision errors (HFACS-MA-22)

with 15.58%. In the other ranking positions, we see the planned inappropriate operation factor (HFACS-MA-9) with 12.55%, the organisational process factor (HFACS-MA-7) with 8.23%, the resource management factor (HFACS-MA-17) with 6.49%, the skill-based errors factor (HFACS-MA-20) with 6.06%, the physical environment factor (HFACS-MA-12) with 4.76%, and smaller percentages corresponding to the other factors (29.01%).

Table 4 shows the percentages of the factors that contribute to grounding incidents. The most frequent factors are judgement/decision (HFACS-Ground-2) with 12.2% and resource management (HFACS-Ground-20) with 12.2%, followed by the skill-based factor (HFACS-Ground-1) with 9.76%, the inappropriate planned operations factor (HFACS-Ground-17) with 9.76%, the physical/mental limitations factor (HFACS-Ground-12) with 7.31%, the perceptual (HFACS-Ground-13) factor with 7.31%, the physical environment factor with 4.76%, and smaller percentages corresponding to the other factors (41.46%). Table 5 shows the percentages of the factors that contribute to the machinery fire engine and explosion incidents. The most frequent factor is equipment/facility (HFACS-MSS-5) resources with 34.78% followed by the technological environment factor (HFACS-MSS-17) with 23.91%, the skill-based errors factor (HFACS-MSS-22) with 10.87%, and smaller percentages corresponding to the other factors (30.44%). Table 6 shows the percentages of factors that contribute to the collision incidents. The most frequent factor is decision errors (HFACS-Coll-2) with 23.21% followed by planned inappropriate operation (HFACS-Coll-14) with 17.86%, ship resource mismanagement (HFACS-Coll-11) with 10.71%, perceptual errors violations (HFACS-Coll-3) with 8.93%, organisational process (HFACS-Coll-19) with 8.93%, and smaller percentages corresponding to the other factors (30.36%).

Table 3. Percentages of general human and organisational factors that contribute to the incidents' occurrence.

Factor	Percentage (%)
HFACS-MA-5	17.32
HFACS-MA-22	15.58
HFACS-MA-9	12.55
HFACS-MA-7	8.23
HFACS-MA-17	6.49
HFACS-MA-20	6.06
HFACS-MA-12	4.76

Table 4. Percentages of the human factors that contribute to grounding incidents.

Factor	Percentage (%)
HFACS-Ground-2	12.2
HFACS-Ground-20	12.2
HFACS-Ground-1	9.76
HFACS-Ground-17	9.76
HFACS-Ground-12	7.31
HFACS-Ground-13	7.31

Table 5. Percentages of the human factors that contribute to machinery fire engine and explosion incidents.

Factor	Percentage (%)
HFACS-MSS-5	34.78
HFACS-MSS-17	23.91
HFACS-MSS-22	10.87

Table 6. Percentages of the human factors that contribute to collision incidents.

Factor	Percentage (%)
HFACS-Coll-2	23.21
HFACS-Coll-14	17.86
HFACS-Coll-11	10.71
HFACS-Coll-3	8.93
HFACS-Coll-19	8.93

4. Experimental Study

This section provides deeper insights into the proposed dataset, highlighting its usefulness in the naval domain. Specifically, Section 4.1 presents the experimental results of two different classification tasks. First, we are interested in predicting whether an incident that occurs under specific conditions (e.g., weather) results in environmental pollution (e.g., oil spill) or not. In the second task, we try to estimate the size of the financial damage to a vessel (i.e., low, moderate, and high-cost) due to the accident. Our study in both tasks aims to show the utility of the Naval_v2 dataset in the prediction of an incident's environmental/financial impact based on informative attributes of the dataset without applying extensive data pre-processing and model tuning. We experiment with the following machine learning algorithms: Random Forest [25], Support Vector Machines [26], K Nearest Neighbours, Logistic Regression [27], Bagging [28] with Decision Tree as base estimator, and Decision Trees [29]. Tables 7 and 8 show the experimental results. We use overall Accuracy as an evaluation measure with 10-fold cross-validation. We also give the standard deviation in parentheses for each model. Finally, Section 4.2 presents clustering results, according to the K-means algorithm [30], to highlight the usefulness of the raw dataset's attribute (which briefly describes the incident and the crew's actions) towards improving safety onboard. We use the scikit-learn (<https://scikit-learn.org/stable/>, accessed on 20 November 2021) python library in our study.

Table 7. Mean and standard deviation of the overall accuracy of the classification algorithms using 10-fold cross-validation for the prediction of the environmental impact.

Algorithm	Overall Accuracy (std.)
Random Forest	0.78 (0.08)
Support Vector Machines	0.64 (0.01)
K Nearest Neighbours	0.67 (0.08)
Logistic Regression	0.66 (0.08)
Bagging	0.76 (0.08)
Decision Trees	0.69 (0.10)

Table 8. Mean and standard deviation of the overall accuracy of the classification algorithms using 10-fold cross-validation for the prediction of the financial impact.

Algorithm	Overall Accuracy (std.)
Random Forest	0.71 (0.07)
Support Vector Machines	0.69 (0.02)
K Nearest Neighbours	0.71 (0.08)
Logistic Regression	0.69 (0.02)
Bagging	0.72 (0.08)
Decision Trees	0.63 (0.10)

4.1. Prediction of the Incidents' Environmental/Financial Impact

First, we experimented with various classifiers to predict whether the accidents caused environmental pollution. Intuitively, the Ship Type attribute is strongly related to the possibility of environmental pollution (e.g., tankers that store liquids or gases are more threatening than fishing vessels), as well as the Ship Length attribute, which gives a sense of the ship's size. Moreover, weather conditions, described by Visibility and Wind Speed attributes, usually play an important role in both the accident's occurrence and the post-incident management to restrict its consequences. Finally, information related to the location type (i.e., Location Type attribute) and the number of crew members (i.e., No of Crew Members attribute) indicates the availability of human and technical resources to restrain the accident's impact (e.g., the timely intervention of the authorities and a sufficient number of crew members enable the immediate intervention in different parts of the ship where there are damages). We converted the Accident Type variable into a categorical variable (i.e., 1: Collision, 2: Grounding, 3: Heavy weather, 4: Machinery space fires and explosions, and 5: Other), as well as the Ship Type variable (i.e., 1: Fishing, 2: Towing, 3: Passenger, 4: Other, 5: Cargo, 6: Bulk, 7: Tanker, and 8: Cruise). Missing values for the categorical and numerical variables are replaced with the highest frequency value and the mean value of each variable, respectively. Random Forest achieves the best performance, i.e., 0.78, outperforming all the other classifiers due to its ability to deal with small sample sizes (as in our case) [31] (see Table 7). Bagging follows in the final ranking, also achieving high overall accuracy (i.e., 0.76).

Table 8 shows the results for the prediction of the size of the financial damage to the vessel (i.e., low, moderate, and high-cost) due to the accident. To predict the financial cost, we used the Economic impact damage on vessel dataset's attribute to create three categories, i.e., the first category contains all the damages that cost from \$0 to \$500,000, the second from \$500,000 to \$5,000,000, and the last one \$5,000,000 and higher. So, in this case, we predict which of these three categories the financial cost will belong to. These categories represent low-, moderate-, and high-cost damages. Intuitively, the number of deaths (i.e., No. of Deaths attribute) is strongly related to the financial damage because it implicitly reveals the accident's severity. In this vein, the number of passengers (i.e., No. of Passengers attribute) is a complementary but equally important element that the model considers along with the number of deaths to infer the magnitude of human loss. Moreover, the No. of Passengers is a quite informative attribute in many ways; e.g., it also gives a sense for the ship's type (for example, large ships such as cruise ships carry a large number of passengers, and even minor damage to these ships can be costly). Furthermore, weather conditions, specifically the wind (i.e., Wind Speed attribute), usually play an important role in both the accident's occurrence and the post-incident management to restrict the consequences (e.g., for fire accidents). Finally, information related to the location type (i.e., Location Type attribute) indicates the availability of human and technical resources to restrain the accident's impact (timely intervention). In this task, Bagging achieves the best performance, i.e., 0.72, outperforming all the other classifiers due to its proven ability to deal with small sample sizes [32]. K Nearest Neighbours Classifier and Random Forest follow in the final ranking, also achieving high accuracy (i.e., 0.71).

4.2. Identifying Specific Types of Ships and Accidents with Weaknesses in Post-Incident Procedures

First, we represent each accident's text description, appearing at the Raw attribute of the dataset, as a TF-IDF vector [33]. The TF-IDF is measures the importance of a word by comparing the number of times the word appears in a document with the number of documents where the word appears and is defined as:

$$w_{i,j} = TF_{i,j} \times \log\left(\frac{N}{DF_i}\right) \quad (1)$$

where $w_{i,j}$ is the TF-IDF score for term i in the document j , N is the number of documents in the collection, $TF_{i,j}$ is the term frequency of the term i in document j , and DF_i is the document frequency, which is equal to the number of documents in the collection that contain the term i [34]. Then, we use the clustering algorithm K-means to group the TF-IDF vectors (First, we convert all uppercase characters into lowercase and remove stopwords and punctuation. Then, we use the scikit-learn Python library for the TF-IDF vector representation) into three clusters (<https://github.com/ContextLab/hypertools>, accessed on 20 November 2021) (see Figure 7). The three clusters that have been created contain the following number of instances: cluster 1 has 93 instances, cluster 2 has 96, and cluster 3 has 60.

Specifically, cluster 1 includes 93 instances, and the mean length of the vessels in this cluster is 136.31 m, which is the larger mean compared to the other two clusters (see Figure 8). More than half of the vessels in this cluster are Cargoes, Bulk, Tankers, and Cruises (i.e., 47 out of 93, see Figure 9), so this cluster represents the larger vessels. Furthermore, 65 out of 93 (i.e., approximately 70%) in this cluster are Collisions or Groundings (see Figure 10). Finally, 38 out of 93 cases cost human injuries or deaths (i.e., approximately 40%).

Cluster 2 includes 96 instances whose mean vessels' length is 42.18 m, which is the smaller mean compared to the other two clusters (see Figure 8). More than half of the vessels in this cluster are Fishing (i.e., 51 out of 96, see Figure 9), representing the smaller vessels. Furthermore, in this cluster, 47 out of 96 (i.e., approximately 50%) belong to the Other type of accident (see Figure 10). Finally, 39 out of 96 cases cause human injuries or deaths (i.e., 37.5%).

Cluster 3 includes 60 instances whose mean vessel length is 94.09 m (see Figure 8). Approximately half of the vessels in this cluster are Towing or Passenger (i.e., 27 out of 60—see Figure 9), representing the middle-sized vessels. Furthermore, in this cluster, 57 out of 60 (i.e., 95%) belong to machinery space fire and explosion accidents (see Figure 10). Finally, 28 out of 60 cases caused human injuries or deaths (i.e., approximately 50%).

This data-driven study is consistent with the literature (e.g., [2]), highlighting the fact that existing post-incident management plans include some common steps. Specifically, collision and grounding accidents response contain common steps, e.g., the captain sent crew members to assess the damage and identify any water inflow. On the other hand, as the nature of the accident is different during machinery space fires and explosion accidents, another action plan is followed. In this case, it is reasonable why the TF-IDF vectors of the collision and grounding accidents are in the same cluster and the machinery space fires and explosion accidents in another cluster (see Figure 10).

To sum up, there is a need for more effective and well-defined post-incident management plans for the collisions, groundings, and machinery space fires and explosions accidents. As we identified above, a considerable number of grounding or collision accidents, as well as machinery space fire and explosion accidents, caused human injuries or deaths, i.e., 39 out of 96 and 28 out of 60 cases, respectively. The data analysis indicated that the instructions and actions for such accident types were ineffective in protecting human life. Hence, there is a need for updated contingency plans, especially for the collisions and groundings accidents for large vessels and fire accidents for middle-sized vessels. Experience from such historical data could effectively contribute to accidents response by improving safety protocols.

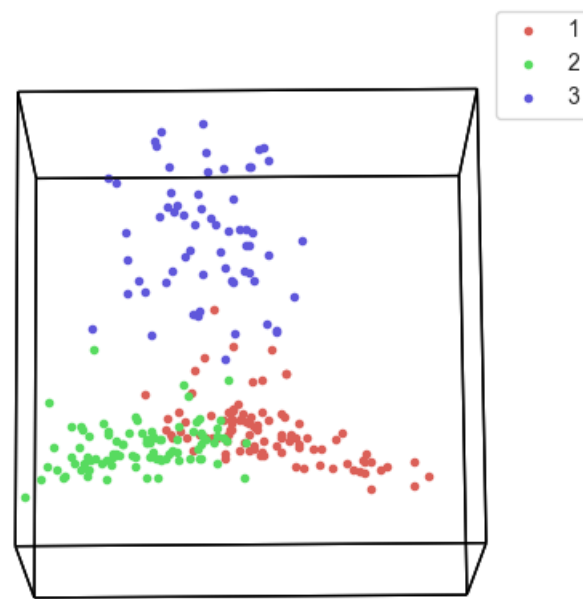


Figure 7. A 3D projection of the dataset's TF-IDF vector representations using PCA. Each point represents an accident's description. There are three clusters (according to K-means algorithm) represented by different colours.

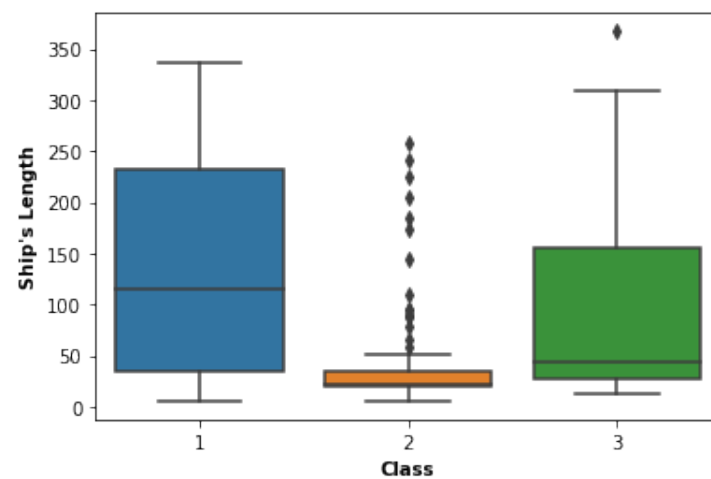


Figure 8. The length of the ships in each cluster.

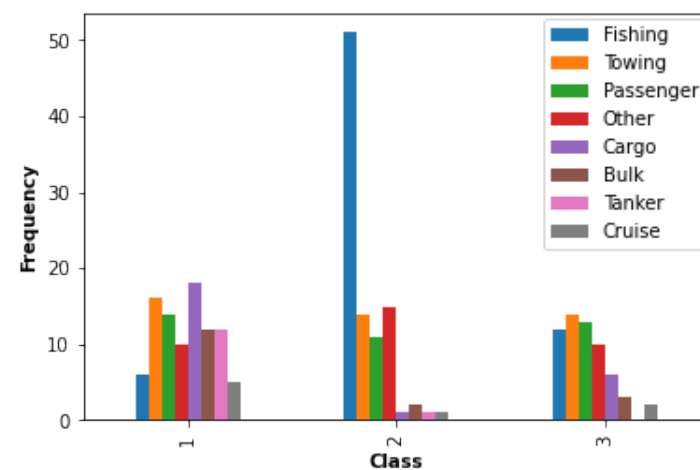


Figure 9. The ship types in each cluster.

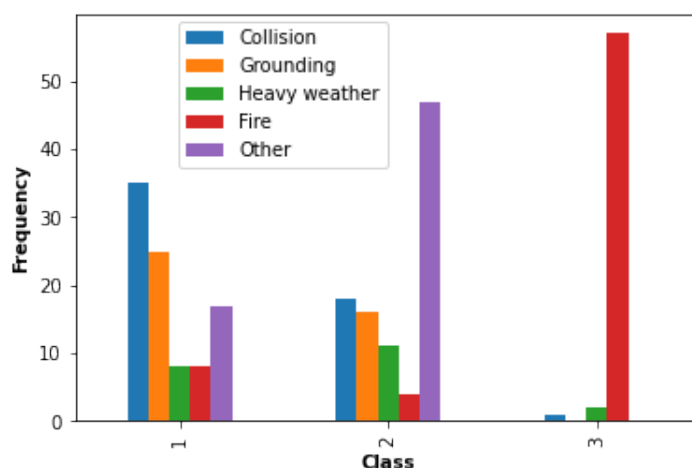


Figure 10. The accident types in each cluster.

5. Conclusions and Future Directions

In this work, we provide a high-quality dataset, called Naval_v2, that combines characteristics related to accident conditions, the post-incident management process, the human factors contributing to each incident's occurrence, and the corresponding environmental/financial impact. Our experimental study indicates a need for updated contingency plans regarding collisions and groundings accidents for large vessels and fire accidents for middle-sized vessels. Furthermore, the dataset enables us to predict with remarkable accuracy (i.e., 0.78) whether an incident causes environmental pollution or not and the economic impact of the accident to the vessel with satisfactory accuracy (i.e., 0.72) without applying extensive data pre-processing and models' tuning, indicating that the datasets' attributes are very informative.

Furthermore, we plan to enrich the Naval_v2 dataset using more accident cases from the Japan Transportation Safety Board (<https://www.mlit.go.jp/jtsb/marrep.html>, accessed on 20 November 2021) and UK Government (<https://www.gov.uk/maib-reports>, accessed on 20 November 2021) databases. Experience from such historical data could effectively contribute to accident responses by improving safety protocols. Specifically, data-driven Artificial Intelligence approaches could be built to support the captain in better decision-making. We also aim to develop a real-time decision support system using machine learning techniques (e.g., Classifier Chains [35]) and the experience of past accidents to find the most appropriate set of actions based on the dedicated situation (e.g., weather conditions, ship technical characteristics, etc.) for an effective (i.e., without human losses and injuries) post-incident management.

Author Contributions: Conceptualization, P.P. and A.L.; methodology, P.P. and A.L.; investigation P.P., K.G. and N.A.; data curation, P.P. and A.L.; writing, P.P., K.G., N.A. and A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 814962 (PALAEMON).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used for this paper can be found at <https://zenodo.org/record/5592999>, accessed on 22 October 2021.

Acknowledgments: I would like to thank all the partners that participated in Palaemon's project for their useful feedback.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Example of accident reports and the information that they provide.

Accident type	Fire/Explosion	No. DCA18FM038
Vessel name	Jacob Kyle Rusthoven	
Location	Lower Mississippi River, mile 673.8; near West Helena, Arkansas ¹ 34°38.77' N, 090°35.12' W	
Date	September 12, 2018	
Time	1005 central daylight time (coordinated universal time – 5 hours)	
Injuries	None	
Property damage	\$1.5 million est.	
Environmental damage	None reported	
Weather	Clear, visibility 10 miles, winds light and variable, air temperature 70°F	
Waterway information	According to the Helena Gage at mile 663.1, located 10 miles south of the accident, the river height was 19.1 feet and rising.	

Formerly the *Capt. George Brumley*, the 85-foot-long towing vessel *Jacob Kyle Rusthoven*

spread, three of the barges broke away from the tow, and one rolled over and lost its cargo. All six crewmembers abandoned the vessel onto the barges, from where they were rescued by a Good

Figure A1. Extracting features related to the unique id of each report, the accident type, the vessel's name, the date of the accident, the vessel's length, the vessel's type, and the persons onboard [36].

the area generating strong winds and waves reportedly 3–5 feet high with the highest wind gust recorded at 73 miles per hour (mph). The *Stretch Duck 54*, which entered the lake about 2 minutes

lasting windstorm is associated with a continuous band of rapidly moving showers or intense thunderstorms and is characterized by damaging strong straight-line wind.⁶⁰ The swath of the storm

Type of Injury	Crew	Passengers	Total
Fatal	1	16	17
Serious	0	2	2
Minor	1	3	4
None	0	8	8

Outdoor temperature greater than 95°F

for fetch and the wave height on the lake. It concluded that Table Rock Lake has a very complex shoreline, so fetch was highly dependent on the direction of the wind, which was generally flowing from the north. Fetch distances ranged from as short as 1.55 miles to as long as 3.7 miles.

know where I was. I was hitting, trying to get out." She escaped from a depth where she described the water as "icy cold." The captain and a young passenger recalled escaping through the front

crew reported visibility of 2 miles

Figure A2. Extracting data related to the sea state, the wind speed, the existence of rain at the accident time, the number of crew members, the number of deaths and injuries, the air temperature, the wind direction, the water temperature, and the visibility [37].

Appendix B

Table A1. The HFACS-MA factors for all the type of accidents.

All Incidents	
Factors	Dataset Encoding
Legislation gaps	HFACS-MA-1
The deficiencies in the administration	HFACS-MA-2
Flaws in design	HFACS-MA-3
Others	HFACS-MA-4
Asset management	HFACS-MA-5
Organizational climate	HFACS-MA-6
Organizational process	HFACS-MA-7
Inadequate supervision	HFACS-MA-8
Planned inappropriate operation	HFACS-MA-9
Failure to correct known problems	HFACS-MA-10
Violations in supervision	HFACS-MA-11
Physical environment	HFACS-MA-12
Technological environment	HFACS-MA-13
Adverse mental states	HFACS-MA-14
Adverse physical conditions	HFACS-MA-15
Physical or mental limitations	HFACS-MA-16
Resource management	HFACS-MA-17
Readiness for the task	HFACS-MA-18
Communication (ships and VTS)	HFACS-MA-19
Skill-based errors	HFACS-MA-20
Perception errors	HFACS-MA-21
Decision errors	HFACS-MA-22
Routine violations	HFACS-MA-23
Exceptional violations	HFACS-MA-24

Table A2. The HFACS-MSS factors for the machinery space fire and explosion accidents, the HFACS-Ground factors for the grounding accidents, and the HFACS-Coll factors for the collision accidents.

Machinery Space Fires and Explosions		Groundings		Collisions	
Factors	Dataset Encoding	Factors	Dataset Encoding	Factors	Dataset Encoding
International standards	HFACS-MSS-1	Skill-based	HFACS-Ground-1	Skill-based errors	HFACS-Coll-1
Flag State implementation	HFACS-MSS-2	Judgment Decision	HFACS-Ground-2	Decision errors	HFACS-Coll-2
Human resources	HFACS-MSS-3	Perceptual error	HFACS-Ground-3	Perceptual errors violations	HFACS-Coll-3
Technological resources	HFACS-MSS-4	Routine	HFACS-Ground-4	Routine violations	HFACS-Coll-4
Equipment/facility resources	HFACS-MSS-5	Exceptional	HFACS-Ground-5	Exceptional violations	HFACS-Coll-5
Structure	HFACS-MSS-6	Physical environment	HFACS-Ground-6	Physical environment	HFACS-Coll-6
Policies	HFACS-MSS-7	Technological environment	HFACS-Ground-7	Technological environment	HFACS-Coll-7
Culture	HFACS-MSS-8	Infrastructures	HFACS-Ground-8	Adverse mental state	HFACS-Coll-8
Operations	HFACS-MSS-9	Cognitive factors	HFACS-Ground-9	Adverse physiological state	HFACS-Coll-9
Procedures	HFACS-MSS-10	Psycho-behavioral factors	HFACS-Ground-10	Physical/mental limitations	HFACS-Coll-10
Oversight	HFACS-MSS-11	Adverse physiological state	HFACS-Ground-11	Ship Resource Mismanagement	HFACS-Coll-11
Shipborne and shore supervision	HFACS-MSS-12	Physical/Mental limitations	HFACS-Ground-12	Personal readiness	HFACS-Coll-12
Shipborne operations	HFACS-MSS-13	Perceptual factors	HFACS-Ground-13	Inadequate leadership	HFACS-Coll-13
Shipborne related shortcomings	HFACS-MSS-14	Coordination	HFACS-Ground-14	Planned inappropriate operation	HFACS-Coll-14
Shipborne violations	HFACS-MSS-15	Communication Planning	HFACS-Ground-14	Failed to correct problem	HFACS-Coll-15
		Personal readiness	HFACS-Ground-15		

Table A2. Cont.

Machinery Space Fires and Explosions		Groundings		Collisions	
Factors	Dataset Encoding	Factors	Dataset Encoding	Factors	Dataset Encoding
Physical environment	HFACS-MSS-16	Inadequate supervision	HFACS-Ground-16	Leadership violations (non compliance with Safety Management System SMS)	HFACS-Coll-16
Technological environment	HFACS-MSS-17	Planned inappropriate operations	HFACS-Ground-17	Resource management	HFACS-Coll-17
Cognitive factors	HFACS-MSS-18	Failed to correct known problems	HFACS-Ground-18	Organisational climate	HFACS-Coll-18
Physiological state	HFACS-MSS-19	Supervisory violations	HFACS-Ground-19	Organisational process	HFACS-Coll-19
Crew interaction	HFACS-MSS-20	Resource management	HFACS-Ground-20	Outside factors	HFACS-Coll-20
Personal readiness	HFACS-MSS-21	Organizational climate	HFACS-Ground-21		
Skill-based errors	HFACS-MSS-22	Organizational process	HFACS-Ground-22		
Decision and judgment errors	HFACS-MSS-23	Regulation gaps	HFACS-Ground-23		
Perceptual errors	HFACS-MSS-24	Other factors	HFACS-Ground-24		
Routine	HFACS-MSS-25				
Exceptional	HFACS-MSS-26				

References

1. Szubrycht, T. Marine accidents as potential crisis situations on the Baltic Sea. *Arch. Transp.* **2020**, *54*, 125–135. [\[CrossRef\]](#)
2. Karahalios, H. The contribution of risk management in ship management: The case of ship collision. *Saf. Sci.* **2014**, *63*, 104–114. [\[CrossRef\]](#)
3. Chauvin, C. Human factors and maritime safety. *J. Navig.* **2011**, *64*, 625. [\[CrossRef\]](#)
4. Wu, B.; Zong, L.; Yip, T.L.; Wang, Y. A probabilistic model for fatality estimation of ship fire accidents. *Ocean Eng.* **2018**, *170*, 266–275. [\[CrossRef\]](#)
5. Choi, J.; Yang, C.S. Smart Escape Support System for Passenger Ship: Active Dynamic Signage & Real-time Escape Routing. In *Proceedings of the Korean Institute of Navigation and Port Research Conference*; Korean Institute of Navigation and Port Research: Seoul, Korea, 2017; pp. 79–85.
6. Sun, S. *Research on Improving Maritime Emergency Management Based on AI and VR in Tianjin Port*; World Maritime University: Malmö, Sweden, 2020.
7. NTSB. Fire aboard Roll-on/Roll-off Passenger Vessel Caribbean Fantasy Atlantic Ocean, 2 Miles Northwest of San Juan, Puerto Rico 17 August 2016. In *Marine Accident Report NTSB/MAR-18/01 PB2018-101068*; NTSB: Washington, DC, USA, 2016.
8. Pine, J.C. Research needs to support the emergency manager of the future. *J. Homel. Secur. Emerg. Manag.* **2003**, *1*. [\[CrossRef\]](#)
9. Wang, H.; Liu, Z.; Wang, X.; Graham, T.; Wang, J. An analysis of factors affecting the severity of marine accidents. *Reliab. Eng. Syst. Saf.* **2021**, *210*, 107513. [\[CrossRef\]](#)
10. Mullai, A.; Larsson, E.; Norrman, A. A study of marine incidents databases in the Baltic sea region. In *Marine Navigation and Safety of Sea Transportation*; Taylor & Francis Group: London, UK, 2009; pp. 247–253.
11. Ventikos, N.; Koimtzoglou, A.; Louzis, K.; Eliopoulou, E. Statistics for marine accidents in adverse weather conditions. *Marit. Technol. Eng.* **2014**, *1*, 243.
12. Mullai, A.; Paulsson, U. A grounded theory model for analysis of marine accidents. *Accid. Anal. Prev.* **2011**, *43*, 1590–1603. [\[CrossRef\]](#)
13. Zhang, Z.; Li, X.M. Global ship accidents and ocean swell-related sea states. *Nat. Hazards Earth Syst. Sci.* **2017**, *17*, 2041–2051. [\[CrossRef\]](#)
14. Zhang, L.; Wang, H.; Meng, Q.; Xie, H. Ship accident consequences and contributing factors analyses using ship accident investigation reports. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2019**, *233*, 35–47. [\[CrossRef\]](#)
15. Wang, X.; Zhang, B.; Zhao, X.; Wang, L.; Tong, R. Exploring the underlying causes of Chinese Eastern Star, Korean Sewol, and Thai Phoenix ferry accidents by employing the HFACS-MA. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4114. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Bowo, L.P.; Furusho, M. Human error assessment and reduction technique for reducing the number of marine accidents in Indonesia. In *Applied Mechanics and Materials*; Trans Tech Publications Ltd.: Bâche, Switzerland, 2018; Volume 874, pp. 199–206.
17. Uğurlu, Ö.; Yıldırım, U.; Başar, E. Analysis of grounding accidents caused by human error. *J. Mar. Sci. Technol.* **2015**, *23*, 748–760.
18. Akyuz, E.; Celik, M. A hybrid decision-making approach to measure effectiveness of safety management system implementations on-board ships. *Saf. Sci.* **2014**, *68*, 169–179. [\[CrossRef\]](#)
19. Stefanidis, F.; Boulougouris, E.; Vassalos, D. Ship evacuation and emergency response trends. In *Design and Operation of Passenger Ships*; The Royal Institution of Naval Architects: London, UK, 2019.
20. Perera, L.; Rodrigues, J.; Pascoal, R.; Soares, C.G. Development of an onboard decision support system for ship navigation under rough weather conditions. In *Sustainable Maritime Transportation and Exploitation of Sea Resources*; Rizzuto, E., Guedes Soares, C., Eds.; Taylor & Francis Group: London, UK, 2012; pp. 837–844.

21. BMA. Report of the investigation into a fire at sea May 2013. In *Bahamas Maritime Authority Official Number 8000400*; Bahamas Maritime Authority: London, UK, 2014.
22. Chauvin, C.; Lardjane, S.; Morel, G.; Clostermann, J.P.; Langard, B. Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS. *Accid. Anal. Prev.* **2013**, *59*, 26–37. [[CrossRef](#)]
23. Schröder-Hinrichs, J.U.; Baldauf, M.; Ghirxi, K.T. Accident investigation reporting deficiencies related to organizational factors in machinery space fires and explosions. *Accid. Anal. Prev.* **2011**, *43*, 1187–1196. [[CrossRef](#)]
24. Mazaheri, A.; Montewka, J.; Nisula, J.; Kujala, P. Usability of accident and incident reports for evidence-based risk modeling—A case study on ship grounding reports. *Saf. Sci.* **2015**, *76*, 202–214. [[CrossRef](#)]
25. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
26. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
27. Cramer, J.S. *The Origins of Logistic Regression*; Tinbergen Institute: Amsterdam, The Netherlands, 2002.
28. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
29. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
30. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [[CrossRef](#)]
31. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [[CrossRef](#)]
32. Chawla, N.V.; Moore, T.E.; Hall, L.O.; Bowyer, K.W.; Kegelmeyer, W.P.; Springer, C. Distributed learning with bagging-like performance. *Pattern Recognit. Lett.* **2003**, *24*, 455–471. [[CrossRef](#)]
33. Daniel, J.; James, H.M. *Speech and Language Processing*; Prentice Hall: Hoboken, NJ, USA, 2000.
34. Zhang, W.; Yoshida, T.; Tang, X. A comparative study of TF* IDF, LSI and multi-words for text classification. *Expert Syst. Appl.* **2011**, *38*, 2758–2765. [[CrossRef](#)]
35. Read, J.; Pfahringer, B.; Holmes, G.; Frank, E. Classifier chains for multi-label classification. *Mach. Learn.* **2011**, *85*, 333–359. [[CrossRef](#)]
36. NTSB. Engine Room Fire aboard Towing Vessel Jacob Kyle Rusthoven, Lower Mississippi River, near West Helena, Arkansas 12 September 2018. In *Marine Accident Report NTSB/MAR-18/01 PB2018-101068*; NTSB: Washington, DC, USA, 2018.
37. NTSB. Sinking of Amphibious Passenger Vessel Stretch Duck 7 Table Rock Lake, near Branson, Missouri July 19, 2018. In *Marine Accident Report NTSB/MAR-20/01 PB2020-101002*; NTSB: Washington, DC, USA, 2018.