



## Classification Models for Assessing the Severity of Marine Accidents Based on Machine Learning

Rossi Passarella<sup>1\*</sup>, Arinda I. Safitri<sup>1</sup>, Nyayu Latifah Husni<sup>2</sup>, Rifka Widyastuti<sup>3</sup>, Harumi Veny<sup>4</sup>

<sup>1</sup> Department of Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya, Palembang 30128, Indonesia

<sup>2</sup> Department of Electrical Engineering, Politeknik Negeri Sriwijaya, Palembang 30139, Indonesia

<sup>3</sup> Department of Information Technology, Universitas Siber Indonesia, Jakarta 12530, Indonesia

<sup>4</sup> College of Engineering, Universiti Teknologi MARA (UiTM), Shah Alam 40450, Malaysia

Corresponding Author Email: [passarella.rossi@unsri.ac.id](mailto:passarella.rossi@unsri.ac.id)

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijssse.140418>

### ABSTRACT

**Received:** 14 April 2024

**Revised:** 5 June 2024

**Accepted:** 11 June 2024

**Available online:** 30 August 2024

#### Keywords:

*classification, marine transport, marine accident, machine learning, modelling*

Marine transport is still famous and claimed to be part of human civilization, but in practice, marine vessels still experience accidents quite frequently, which can result in large losses. Therefore, this research aims to integrate multiple data sources on marine accidents, classify them to identify patterns, and create a model to forecast and prevent future accidents. The first step in the methodology is to connect several variables from multiple data sources and generate target variables. We then feed this ready data set into 10 machine learning algorithms to determine which one best suit the data type and quality. The training results provided four algorithms with the best performance, namely label spreading, label propagation, random forest, and XGB classifier algorithms. After comparing the training and testing results, we found that XGB performed slightly better than the other three models, where the developed model and dataset only had a performance of 70%-74% in predicting marine accidents in the corresponding class.

## 1. INTRODUCTION

Ocean transportation plays a crucial role in international trade because it can increase economic growth, human mobility, and national security. As such, all global activities rely heavily on maritime transportation, with approximately 90% of commodity goods being transported by sea [1]. More than 50,000 merchant ships are involved in international trade every day [2], which causes the level of shipping traffic to increase, especially in narrow waters, and creates a great opportunity for marine accidents [3]. Based on the records of the European Maritime Safety Agency Report [4], in 2014–2021, the total number of marine accidents that occurred in the world was 23,623 cases, and according to the Allianz Global Corporate & Specialty Report [1], during the last decade, the total number of marine accidents in the world was 44,264 cases. The types of accidents that occurred included sinking or reversing ships, collisions, fires or explosions, and aground, with varying degrees of severity. From the records of these accidents, it is necessary to conduct further studies on the severity of marine accidents to improve the security and safety of shipping around the world.

Previously, the assessment of the severity of marine accidents has been carried out by several researchers, in which Weng and Yang [5] analyzed the development of the binary logistic regression model and the zero-truncated binomial model using data sourced from the database managed by Lloyd's List Intelligence Company from 2001-2011. Meanwhile, Wang and Yang [6] developed the Naïve Bayes

model on the Chinese Maritime Safety Administration (MSA) investigation report data from 1979-2015. Both of these studies focus on the severity of marine accidents based on contributing factors such as the type of accident, crew's educational background, accident location, and vessel age. Furthermore, in Lu et al. [7], a model using the Random Oversampling (RO) technique was developed to predict the severity of non-traditional security (NTS) incidents or piracy on the sea route, with the most contributing factors being accident time and ship type using Global Integrated Shipping Information System (GISIS) data from 2015–2020.

Then, Cakir et al. [8] with the Decision Tree (DT) model was used to predict the severity of oil spills due to marine accidents, where this level is influenced by the type of accident and type of ship, using USCG (United States Coast Guard) data from 2002–2015. Then, in the research conducted by Wang et al. [9], a Zero Inflated Ordered Probit (ZIOP) model was developed to predict the severity factor of marine accident injuries. The data used was sourced from the results of the TSB (Transportation Safety Board of Canada), MAIB (Marine Accident Investigation Branch), ATSB (Australian Transport Safety Board), NTSB (National Transportation Safety Board), JTSA (Japan Transport Safety Board), MSA (Maritime Safety Administration), and BSU (The Federal Bureau of Maritime Casualty) investigation reports from 2000–2019. Finally, a research study conducted by Zhou et al. [10] applied a spatial fuzzy multi-criteria evaluation model to assess and map the risk level of marine transportation hazard zones in China's seas with five levels: very high, high, medium, low, and very

low risk zones, using various data sources including governments, international organizations, and commercial companies from 1980-2019.

Mullai and Paulsson [11] conducted research on marine accidents in 2011 for the benefit of Swedish shipping, developing a conceptual model to analyze marine accidents using empirical data from the Swedish Maritime Administration database. The model is based on eleven categories, including non-metric and metric variables such as fatalities, ship property, the number of people on board, and accidents. The model's combined predictive power accounts for 65% of the fatality variance. Other datasets can apply the model, which holds theoretical and practical value. The results of this 2011 study will serve as a comparison to the final findings of the study we developed.

Marine accident research using the data-driven Bayes method is used to analyze vessels on the Istanbul Strait, which is a narrow and busy waterway. using only 418 accident data, with the type of accident that is targeted.

As for the results obtained, the data shows that all vessels, especially those under 300 GRT, are more likely to experience adrift accidents. Thus, research using a small amount of data that is based on the facts of the incident can be a research dataset [12].

This marine accident research has also been reviewed and collected from publications on the WOS (Web of Science) database from 2000 to 2022, with the conclusion that the researchers found that emerging accident analysis methods such as machine learning and big data mining have also shown powerful insights in the analysis of marine accidents [13].

Therefore, to look deeper into the analysis of marine accidents that occur, it is necessary to machine learning and to classify the severity of accidents by focusing on contributing features, including ship type, accident type, accident factors, and severity. Thus, it is expected to improve the effectiveness of prevention strategies for future shipping safety. The data used is sourced from the official websites of the investigation reports of the NTSB, ATSB, NTSB, TSB, MAIB, and JTSB of marine accidents that occurred from 2003 to 2022.

In addition, there is also research using data-driven Bayesian network models (BN) to analyze the relationship between marine accident severity and relevant accident-influencing factors (AIF). This study uses data based on marine accident investigation reports involving 1,294 vessels from 2000 to 2019. The severity of marine accidents is classified, and a database of factors affecting the severity of marine accidents is created. It only uses the Tree Augmented Naive Bayesian (TAN) algorithm to create a data-driven BN model [14]. This distinguishes the research conducted by the author by comparing several machine learning algorithm models.

The structure of this research paper consists of an introductory section that explains the importance of sea

transportation for the economic growth of countries around the world and its impact. In the data and methods section, data processing is performed and research methods are implemented. The next section presents the results of data processing and analyses the factors that contribute to the severity of marine accidents. The last section contains conclusions that summarize the research findings.

## 2. DATA AND METHOD

In this research, there are several stages that must be performed on the data and methods used to suit the research objectives. These stages are shown in Figure 1.

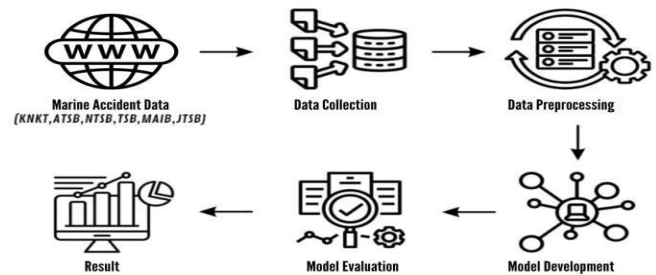


Figure 1. Research stages

### 2.1 Marine accident data and data collection

The data used in this research were obtained from the official website of the results of marine accident investigation reports in several countries, as shown in Table 1. The accident data that occurred consisted of the results of the NTSB report of 220 cases, the ATSB report of 165 cases, the NTSB report of 333 cases, the TSB report of 228 cases, and the MAIB report of 715 cases recorded from 2003–2022. Meanwhile, 107 cases of JTSB were recorded from 2008 to 2020.

Then, all of these data were combined so that the total accident data became 1,768 cases (observations) with 9 variables; one of the targeted variables was severity, as shown in Figure 2.

Date	Month	Year	Ship Type	Type of Accident	Ship Age	Factor	Country	Severity Level
0	28	1 2022	Bulk carrier	Collision	Old	Human	AUS	Serious
1	25	1 2022	Tanker	Other	Beginner	Human	AUS	Less Serious
2	4	6 2021	Cargo	Other	Beginner	Human	AUS	Very Serious
3	25	3 2021	Cargo	Fire or explosion	Beginner	Human	AUS	Serious
4	19	1 2021	Passenger Ship	Aground	Beginner	Human	AUS	Serious
5	11	12 2020	Cargo	Other	Beginner	Human	AUS	Very Serious
6	24	5 2020	Container	Other	Beginner	Nature	AUS	Very Serious
7	29	2 2020	Cargo	Collision	Old	Human	AUS	Serious
8	22	6 2019	Tugboat	Other	Old	Human	AUS	Serious
9	23	3 2019	Passenger Ship	Other	Beginner	System	AUS	Serious

Figure 2. Dataset overview

Table 1. Dataset sources

Organization	Source of Data	Period	Num of Data
KNKT	<a href="https://knkt.go.id">https://knkt.go.id</a>	2003-2022	220
ATSB	<a href="https://www.atsb.gov.au/">https://www.atsb.gov.au/</a>	2003-2022	165
NTSB	<a href="https://www.nts.gov">https://www.nts.gov</a>	2003-2022	333
TSB	<a href="https://www.tsb.gc.ca">https://www.tsb.gc.ca</a>	2003-2022	228
MAIB	<a href="https://www.gov.uk">https://www.gov.uk</a>	2003-2022	715
JTSB	<a href="https://www.mlit.go.jp">https://www.mlit.go.jp</a>	2008-2020	107
<b>Total</b>			<b>1,768</b>

Marine accident severity categories are based on the International Maritime Organization (IMO) guidelines [15], which are divided into the following four categories:

(1) In the "very serious" accident category, when the victim ship is lost, washed away, or suffers loss of life or death, pollution occurs due to oil spills of more than 500 tons.

(2) "Serious" accident category, when the crew suffers injuries and damage to the ship, such as system damage, environmental damage, hull damage, or severe structural damage to the ship, that resulting in the ship being unable to continue the voyage.

(3) "Less serious" accident category, when the accident victim does not meet the criteria in the very serious and serious accident categories.

(4) "Marine incident" category, when it is directly related to the operation of a vessel that has not been repaired, thus endangering shipping activities and the environment.

However, this research uses only three categories of severity: less serious, serious, and very serious. After the data are collected, they are converted into.csv (comma-separated values) format. This is done to facilitate the process of processing data into programming software, and the dataset can be used in the next stage.

## 2.2 Data preprocessing

This stage is the most important for data preparation before further analysis [16, 17]. The first step of the dataset must go through the data cleaning process. From this process, the dataset dimensions were obtained with as many as 1,768 observations with nine variables, including date, month, year, ship type, type of accident, ship age, factor, country, and severity level, as shown in Table 2.

According to Table 2, a missing label with the number 0 indicates the absence of missing values in each data variable. The next step is to identify the distribution of each variable. The date variable contains numeric data ranging from 1 to 31. The data distribution reveals that accidents occur frequently from the beginning to the middle of the month, with the average occurring on the 15th. Secondly, the variable 'month' contains numerical data ranging from 1 to 12, with the highest frequency of marine accidents occurring between June and July. Thirdly, the variable 'year' contains numerical data ranging from 2003 to 2022, with the highest frequency of marine accidents occurring between 2012 and 2013.

Fourth, the ship type variable has categorical data consisting of nine categories: fishing vessel, passenger ship, cargo, tanker, bulk carrier, container, tugboat, barge, and other types of ships. These variables show that the type of ship that is prone to marine accidents is fishing vessels. Fifth, the type of accident variable has categorical data consisting of six categories: ship sinking or turning, collision, aground, fire or explosion, contact, and others. This variable shows that the most common type of

accident is other types of accidents (various types of accidents not covered by the previous categories). Sixth, the ship age variable has categorical type data consisting of two categories: beginner ship age (< 30 years of operation) and old ship age (> 30 years of operation). This variable shows that the age of the ship that has more accidents is the age of the beginner ship.

Seventh, the factor variable has categorical data consisting of four categories: human factors, systems, nature, and overload. This variable shows that humans are the most common factor that causes accidents. Eighth, the country variable has categorical data consisting of six categories namely the UK (United Kingdom), USA (United States of America), Canada, Indonesia, AUS (Australia), and Japan. This variable shows that most ship accidents occur in UK waters. Finally, the severity variable, which has categorical data consisting of 3 categories, namely less serious, very serious, and serious severity, shows the impact of accidents that occur frequently, resulting in serious severity.

The next step is the feature selection stage. This stage is carried out to select the nine dataset variables used so that the most relevant and informative variables are obtained in accordance with the research objectives. The selection of these variables was based on the results of the correlation between the target severity variable and other variables.

The variables with the strongest association with the first severity are the accident type variable, which produces a correlation value of 0.24; the factor variable, which produces a correlation value of 0.1; the year variable, which produces a correlation value of 0.078; the date variable, which produces a correlation value of 0.014; and the ship type variable, which produces a correlation value of 0. The country variable, with a correlation value of -0.096, was the least significant of the variables. The month variable, with a correlation value of -0.016, was also not a significant contributor. Finally, the ship age variable, with a correlation value of -0.0016, was the least significant of the variables. Based on the correlation results, the research will focus on four variables that significantly contributed to the occurrence of marine accidents. These variables are the accident type, factor, ship type, and severity.

Based on data on marine accidents during the period between 2003 and 2022, it was found that the highest type of vessel involved in accidents was fishing vessel with 344 cases (19.5%), and the lowest type of vessel was barge with 38 cases (2.1%). Then, other types of accidents are the highest type, recorded at 428 cases (24.2%). This, type of accident tends to be more diverse and cannot be classified, while the lowest type is contact, recorded at only 195 cases (11.0%). Furthermore, the highest factor that can cause accidents is human, recorded in 890 cases (50.3%), and the lowest accident factor is overload, recorded in only 44 cases (2.5%). Finally, in terms of severity, the highest number of crashes were serious, with 808 cases (45.7%), while the lowest severity was less serious, with only 459 cases (26.0%), as shown in Table 3.

**Table 2.** Dataset descriptive statistic of each variable

Variable	Data Type	Mean	Median	Dispersion	Min	Max	Missing Data
Date	Numerical	15.27	15	0.56	1	31	0
Month	Numerical	6.54	7	0.52	1	12	0
Year	Numerical	2012.56	2013	0	2003	2022	0
Ship Type	Categorical	Fishing vessel	Fishing vessel	2.06	-	-	0
Type of Accident	Categorical	Other	Other	1.76	-	-	0
Ship Age	Categorical	Beginner	Beginner	0.533	-	-	0
Factor	Categorical	Human	Human	1.08	-	-	0
Country	Categorical	UK	UK	1.6	-	-	0
Severity Level	Categorical	Serious	Serious	1.07	-	-	0

**Table 3.** Description of the variable selected for analysis

Variables	Description	Frequency	Percentage
Ship Type	Bulk carrier	137	7.7
	Cargo	326	18.4
	Container	130	7.4
	Fishing	344	19.5
	Passenger	335	18.9
	Tanker	151	8.5
	Barge	38	2.1
	Tugboat	128	7.2
	Other	179	10.1
Type of Accident	Aground	282	16
	Fire	256	14.5
	Contact	195	11
	Other	428	24.2
	Collision	284	16.1
Factor	Sinking	323	18.3
	Nature	257	14.5
	Human	890	50.3
	Overload	44	2.5
Severity Level	Systems	577	32.6
	Less serious	459	26
	Very serious	501	28.3
	Serious	808	45.7

Next, proceed with the data transformation stage. This stage is carried out because the four variables selected above are categorical. At this stage, these variables are converted into numerical data, which is carried out so that the dataset can later be processed into the analysis method used by changing the ship type variables (features), which include the bulk carrier category coded "0", cargo coded "1", container coded "2", other coded "3", fishing vessel coded "4", passenger ship coded "5", tanker coded "6", barge coded "7", and tugboat coded "8". Accident type variables (features), which include the run-aground category, are coded "0", fire or explosion is coded "1", contact is coded "2", other is coded "3", collision is coded "4", and sinking or turning is coded "5".

**Table 4.** Dataset transformation

Variables	Status	Category	Output Code
Ship Type	Features	Bulk carrier	0
		Cargo	1
		Container	2
		Other	3
		Fishing vessel	4
		Passenger ship	5
		Tanker	6
		Barge	7
		Tugboat	8
Type of Accident	Features	Aground	0
		Fire or explosion	1
		Contact	2
		Others	3
		Collision	4
Factor	Features	Sinking or turning	5
		Nature	0
		Human	1
		Overload	2
Severity Level	Target	System	3
		Less serious	0
		Very serious	1
		Serious	2

Factor variables (features) that include the categories of nature are coded "0", human is coded "1", overload is coded "2", and system are coded "3". Finally, labelling was

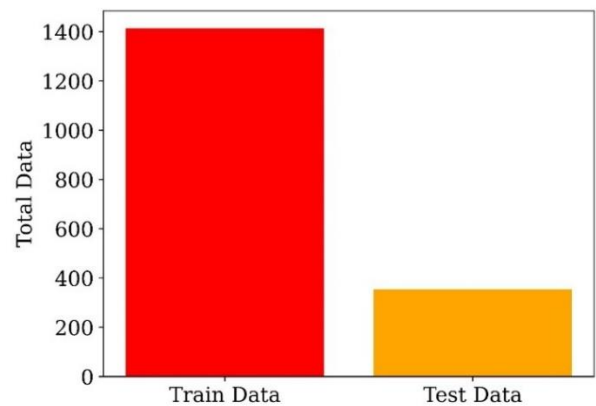
performed on the marine accident severity variable (target), including the less serious class labelled "0", the very serious class labelled "1", and the serious class labelled "2", as shown in Table 4.

Based on Table 5, the distribution of the severity target variable has a total of 1,768 data points, showing imbalanced data classes where class "0" amounts to 459 data points, class "1" amounts to 501 data points, and class "2" amounts to 808 data points. Therefore, the focus of model selection is the f1-score value. This unbalanced dataset is then separated for the machine learning process. Before entering the process, the dataset must be separated into two datasets, namely training data and testing data, using the data splitting method of 80:20. The visualization of splitting dataset are shown in Figure 3, where the training data amounted to 1,414 data points, whereas the testing data was 354 data points.

From the above process, the pre-processing stage has ended, and the dataset can be processed for the machine learning model development stage.

**Table 5.** Target variable dataset distribution

Total Data	Severity Level		
	0	1	2
1,768	459	501	808



**Figure 3.** Splitting data

### 2.3 Method

To classify the severity of marine accidents, this research applies machine learning methods using the *lazypredict* classifier library [18, 19]. From the prediction process, 10 best models were obtained to be tested on the dataset. These models include AdaBoostClassifier (AD), LabelPropagation (LP), DecisionTreeClassifier (DT), ExtraTreeClassifier (ET), ExtraTreesClassifier (ETS), XGBClassifier (XGB), LGBMClassifier (LGBM), LabelSpreading (LS), BaggingClassifier (BC), and RandomForestClassifier (RF).

### 3. RESULT AND DISCUSSION

The results of data processing and analysis of factors contributing to the severity of marine accidents are presented based on several stages that have been carried out.

#### 3.1 Dataset analysis

In this section, the dataset description is explained in depth

to gain insight into the dataset used so that, when modelling, the meaning of the classification model development can be known.

Figure 4 displays a graph of marine accident data from 2003-2022, illustrating the types of accidents experienced by ships. The most common type of accident was 'other' (24.2%), followed by sinking or turning (18.3%), collision (16.1%), aground (16%), fire or explosion (14.5%), and contact with floating objects on the sea surface (11%). During this timeframe, many marine accidents also occurred in 2018, totaling 137 cases, with the highest type of accident being sinking or turning. In 2019, there were 122 cases, with the highest type of accident being collisions, and the lowest accident occurred in 2022, with 29 cases, with the most frequent type of accident being fire or explosion.

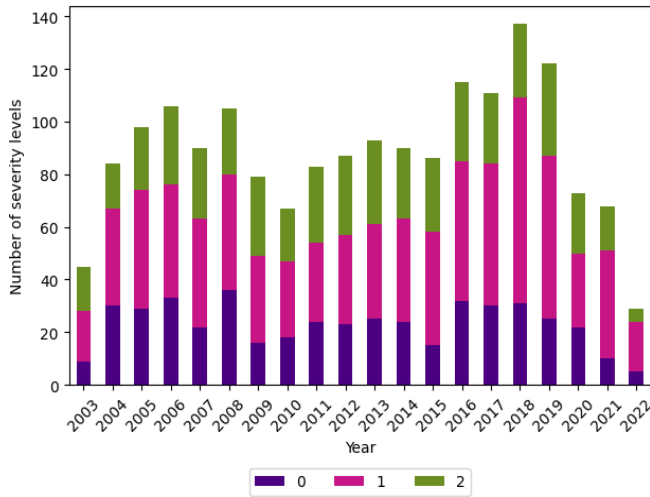


Figure 4. Accident data by year

Figure 5 displays the number of accidents based on the severity level and year. The highest number of accidents in the less serious class occurred in 2008, while the lowest occurred in 2022. In the very serious class, the highest number of accidents occurred in 2019, and the lowest occurred in 2022. The most severe accidents happened in 2018, and the least severe occurred in 2003. Therefore, the data indicates that there are numerous marine accidents resulting in serious severity every year.

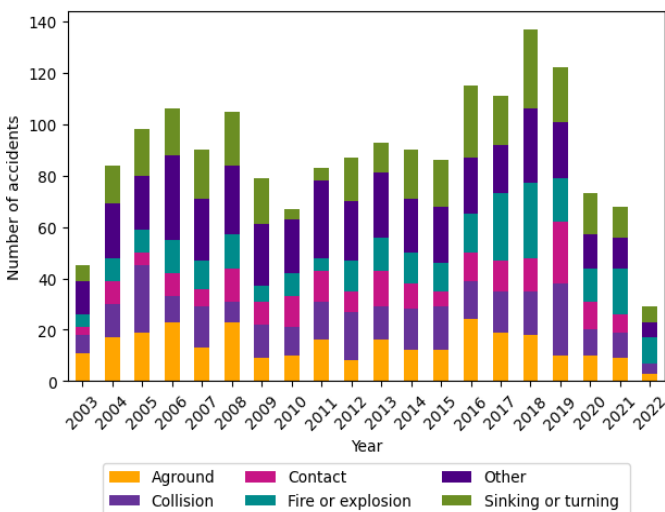


Figure 5. Accident severity by year

Figure 6 shows data on the type of accident and its severity. Based on the data, grounding tends to have a less serious severity impact, whereas other types of accidents tend to have a very serious severity impact, and types of accidents such as collision and fire or explosion tend to have a serious severity impact.

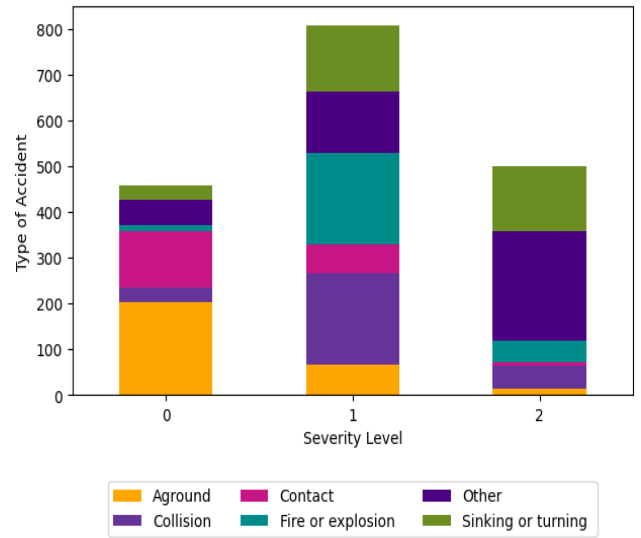


Figure 6. Types of accidents and their severity

Figure 7 shows the distribution of factors causing marine accidents by severity. During the period from 2003 to 2022, the highest contributing factor to accidents was human factors (50.3%), followed by ship system failure or damage (32.6%) and nature (14.5%). The three factors provide varying degrees of severity, ranging from less serious to serious and to very serious, while the overloading factor (2.5%), which exceeds the safety standardization limit, contributes little to the occurrence of marine accidents but tends to have an impact with a very serious severity.

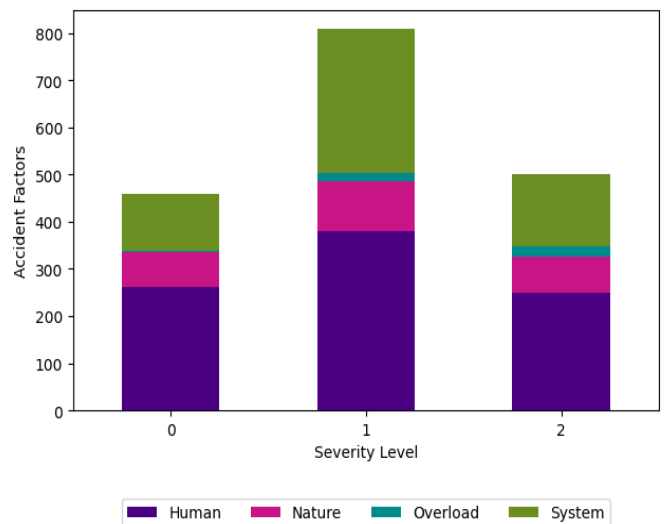


Figure 7. Accident factors and their severity

Figure 8 shows a type of ship that is prone to marine accidents based on their severity. Based on this data, the types of vessels that are prone to the highest number of marine accidents are fishing vessels (19.5%), which tend to have very serious impacts, while the types of vessels that tend to have serious impacts include passenger ships (18.9%), cargo

(18.4%), tankers (8.5%), containers (7.4%), and tugboats (7.2%). Ship types that tend to have less serious impacts include bulk carriers (7.7%), barges (2.1%), and other ship types (10.1%).

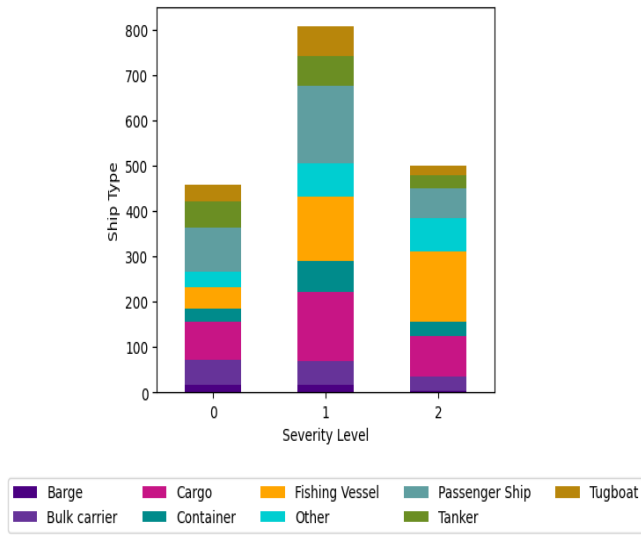


Figure 8. Vessel types and their severity

Apart from that, when observing the age of ships that have experienced accidents (referring to Table 2), based on the dataset, 78% (1,373 rows of observation data) are beginner ships (ship age less than 30 years), while the remaining 22% (395 rows of data) are old ships that are more than 30 years old. The results of the analysis show that ships with an age of less than 30 years have many accidents due to the most dominant human factor, followed by disruption of the shipping system. The details of the data visualization on this matter are shown in Figure 9. With visualization results such as Figure 9, it raises the question of whether new ships under 30 years old with better technology than old ships have complex systems, so that humans such as captains and crew find it difficult to take avoidance actions before an accident occurs.

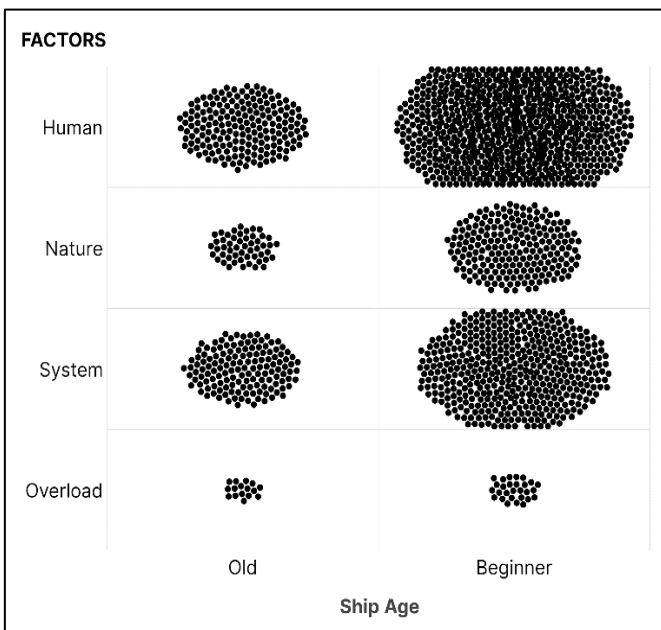


Figure 9. Visual relation between Ship age and factors contributes to accident

This question is answered when looking at Figure 10. It can be seen that system disturbances that cause accidents on ships are mostly caused by fires or explosions, while human-caused factors mostly cause collisions between ships, shipwrecks (aground), and collisions with objects other than ships.



Figure 10. Visualization data between type of accident versus ship age

### 3.2 Development of the model

After analyzing the dataset used, the next step is to develop a model for classifying the severity of marine accidents. At this stage, the research target is the severity level consisting of 3 classes, namely the less serious, serious, and very serious classes. However, the model development process only focuses on the serious class f1-score value. This is because the serious class is more likely to occur in marine accidents than the less serious and very serious classes. Furthermore, the model is analyzed using training and testing data so that the results of the calculation of the performance.

#### 3.2.1 Model analysis using training data

The performance results of each model are shown in table 6. In this machine learning model development stage, using the 10 best models obtained from the *lazypredict* classifier library. The model with the highest F1 score for class 2 was selected as the target model based on the results (Table 6). The best model is chosen from a pool of four models, which are then summarized and presented in Table 7. The best 4 models out of 10 trained in classifying the severity of marine accidents focusing on the serious class were obtained, and the best models included LP, XGB, LS, and RF, with the highest f1-score value in that class of 0.71. Meanwhile, a serious class classification with an f1-score value of 0.70 was generated from the DT, ET, ETS, LGBM, and BC models. Finally, at serious class classification with the lowest f1-score value of 0.65 was generated from the AD model.

#### 3.2.2 Model analysis using testing data

After the training model is obtained, model testing is then carried out using data that has been separated previously for

testing, namely 354 observation data points (Figure 3), while the results of the test are shown in Table 8.

Table 8 shows the result of comparing the model evaluation matrix using testing data. After obtaining the four best models generated from training data, the models were tested using testing data (20%). This is done to determine the performance of each model when tested using test data. Furthermore, we analyze the performance of the f1-score value of each model, as shown in Table 9. Based on this table, the results of testing the serious class severity classification model using testing data show that the LP, RF, and LS models produced the highest f1-score value of 0.76. Then, the XGB models produced a f1-score value of 0.75.

**Table 6.** Training data evaluation matrix comparison

Model	Accuracy	Severity	Precision	Recall	F1-Score
AD	0.65	0	0.69	0.71	0.70
		1	0.55	0.69	0.61
		2	0.72	0.59	0.65
		<b>Mean/total</b>	<b>0.65</b>	<b>0.66</b>	<b>0.65</b>
LP	0.69	0	0.75	0.69	0.72
		1	0.64	0.64	0.64
		2	0.69	0.72	0.71
		<b>Mean/total</b>	<b>0.69</b>	<b>0.68</b>	<b>0.69</b>
DT	0.69	0	0.71	0.75	0.73
		1	0.63	0.66	0.65
		2	0.72	0.68	0.70
		<b>Mean/total</b>	<b>0.69</b>	<b>0.70</b>	<b>0.70</b>
ET	0.69	0	0.71	0.75	0.73
		1	0.63	0.66	0.65
		2	0.72	0.68	0.70
		<b>Mean/total</b>	<b>0.69</b>	<b>0.70</b>	<b>0.70</b>
ETS	0.69	0	0.71	0.75	0.73
		1	0.63	0.66	0.65
		2	0.72	0.68	0.70
		<b>Mean/total</b>	<b>0.69</b>	<b>0.70</b>	<b>0.70</b>
XGB	0.69	0	0.72	0.73	0.73
		1	0.64	0.62	0.64
		2	0.70	0.71	0.71
		<b>Mean/total</b>	<b>0.69</b>	<b>0.69</b>	<b>0.70</b>
LGBM	0.69	0	0.72	0.71	0.72
		1	0.64	0.62	0.63
		2	0.69	0.71	0.70
		<b>Mean/total</b>	<b>0.69</b>	<b>0.68</b>	<b>0.69</b>
LS	0.69	0	0.75	0.69	0.72
		1	0.64	0.64	0.64
		2	0.69	0.73	0.71
		<b>Mean/total</b>	<b>0.70</b>	<b>0.69</b>	<b>0.69</b>
BC	0.69	0	0.74	0.70	0.72
		1	0.62	0.66	0.64
		2	0.70	0.70	0.70
		<b>Mean/total</b>	<b>0.69</b>	<b>0.69</b>	<b>0.69</b>
RF	0.69	0	0.75	0.69	0.72
		1	0.65	0.63	0.64
		2	0.69	0.73	0.71
		<b>Mean/total</b>	<b>0.70</b>	<b>0.68</b>	<b>0.69</b>

**Table 7.** Performance of the f1-score models on training data

Model	Less Serious (0)	Very Serious (1)	Serious (2)
LP	0.72	0.64	0.71
XGB	0.73	0.64	0.71
LS	0.72	0.64	0.71
RF	0.72	0.64	0.71

### 3.2.3 Analysis of training data and testing data result

To determine the extent of the performance generated from

the 4 best machine learning models in classifying the severity of sea accidents using training and testing data, further analysis will be carried out. To determine the best model, the average value produced by each model is compared. A comparison of the results of training data and testing data is shown in Table 10. Of the four models, the same average results were obtained. The models include LP, LS, and RF, with training data results of 0.69, whereas the XGB model is 0.70. Thus, the XGB model is superior to the other 3 models, where this model produces an average training data model performance of 0.70 and test data of 0.74. The Table 10 also shows that the performance of all models is over fitted. Figure 11 shows the classification results of training data and test data from the best model obtained from 10 model tests, namely the XGB model.

**Table 8.** Comparison of the testing data evaluation matrix

Model	Accuracy	Severity	Precision	Recall	F1-score
LP	0.74	0	0.85	0.60	0.71
		1	0.70	0.78	0.74
		2	0.73	0.81	0.76
		<b>Mean/total</b>	<b>0.76</b>	<b>0.73</b>	<b>0.74</b>
XGB	0.74	0	0.77	0.70	0.73
		1	0.70	0.78	0.74
		2	0.76	0.75	0.75
		<b>Mean/total</b>	<b>0.74</b>	<b>0.74</b>	<b>0.74</b>
LS	0.74	0	0.84	0.61	0.71
		1	0.70	0.78	0.74
		2	0.73	0.80	0.76
		<b>Mean/total</b>	<b>0.76</b>	<b>0.73</b>	<b>0.74</b>
RF	0.69	0	0.82	0.64	0.72
		1	0.70	0.78	0.74
		2	0.74	0.79	0.76
		<b>Mean/total</b>	<b>0.75</b>	<b>0.73</b>	<b>0.74</b>

**Table 9.** F1-score performance of the models of testing

Model	Less Serious (0)	Very Serious (1)	Serious (2)
LP	0.71	0.74	0.76
XGB	0.73	0.74	0.75
LS	0.71	0.74	0.76
RF	0.72	0.74	0.76

**Table 10.** Model performance results of training and testing data

Model	Training Data				Testing Data			
	0	1	2	Mean	0	1	2	Mean
LP	0.72	0.64	0.71	0.69	0.71	0.74	0.76	0.74
XGB	0.73	0.64	0.71	0.70	0.73	0.74	0.75	0.74
LS	0.72	0.64	0.71	0.69	0.71	0.74	0.76	0.74
RF	0.72	0.64	0.71	0.69	0.72	0.74	0.76	0.74

Apart from considering Table 10, these four algorithms also have advantages and disadvantages. The advantages of XGB are high performance, scalability, and flexibility, while the disadvantages of XGB are complexity, blackbox tendency, and computational cost. When looking at applicability, XGB is very suitable for structured data and data that has complex relationships.

For RF, it is suitable for various data types and robust to noise. The advantages of RF are that it is interpretable and can handle high-dimensionality data. Next is the LP algorithm. The LP algorithm is a machine learning algorithm in the semi-supervised learning category. The advantages of the LP

algorithm are simplicity, ease of implementation, and interpretability, while the disadvantages are that it is very sensitive to initial labels and graph construction, has limited model complexity, and has no guarantee of convergence.

The next algorithm is LS. This algorithm is almost the same as LP, which is part of machine learning for semi-supervised learning. The use of LS is very suitable for situations where very little amount of data is labeled, and the data used can be described naturally in the form of a graph. The advantages of the LS algorithm are that it is very effective for unlabeled data, flexibility in handling different types of data, simplicity, and interpretability. Then the disadvantages of LS are sensitivity to the similarity graph, limited model complexity, and the potential for label bias.

True Label	Less Serious	67	10	19
	Very Serious	3	77	19
	Serious	17	23	119
		Less Serious	Very Serious	Serious
		Predicted Label		

(a) Training data

True Label	Less Serious	265	36	62
	Very Serious	15	251	136
	Serious	87	100	462
		Less Serious	Very Serious	Serious
		Predicted Label		

(b) Testing data

**Figure 11.** Classification result of the training and testing data from the XGB models

#### 4. LIMITATIONS

This research has limitations on the amount of data used to develop a model. We did not augment the data because the data taken from several accident databases is factual and natural. If we add data in this way, the pattern of events based on the date and year can become more biased. To avoid this, Then the researcher keeps the data intact according to the facts from the database that has been collected.

In addition, the parameter tuning process is not carried out because all algorithms used are collected in the “lazypredicts” library. So that the selection of the best model is only based on the results of training and testing and considering imbalanced class data so that to represent a balanced class distribution, the F1-Score value is chosen [20, 21].

#### 5. THE PRACTICAL APPLICATION

Using machine learning to classify the severity of marine accidents has practical applications. This research is highly valuable for enhancing maritime safety and improving response efforts. Here is an analysis of its practical uses: First, Improved Emergency Response: By rapidly evaluating the seriousness of accidents using machine learning algorithms, authorities can distribute resources with greater efficiency. This can result in the expedited deployment of search and rescue teams, the implementation of pollution control measures, and the provision of medical help for emergency cases. The second improvement is better accident investigation. Using machine learning, huge amounts of data from accident reports can be analyzed, showing complex patterns and links between different factors that cause serious accidents. This can assist investigators in identifying the fundamental reasons and formulating precise preventive actions. Third, regulatory authorities can develop targeted safety laws by identifying the key elements that have the greatest impact on serious accidents. This allows for the creation of regulations that specifically address the most significant hazards. Implementing a data-driven strategy can lead to more efficient safety standards. Fourth, Insurance Risk Assessment: Insurance firms can use these models to improve the precision of evaluating boat risk profiles and then modify premiums accordingly. This can motivate shipowners to give priority to safety measures. Fifth, Enhanced Route Planning: By combining machine learning models with weather forecasting and other marine data, it is possible to pinpoint locations with a high-risk factor and provide recommendations for safer routes for vessels. This can greatly diminish the probability of accidents.

#### 6. CONCLUSIONS

The research analyzed historical and investigative marine accident data from 2003 to 2022, recording a total of 1,768 cases. The findings indicate that ship-to-ship collisions tend to be more severe, while ship grounding tend to be less serious. Other types of accidents result in very serious severity. Fishing vessels are more prone to serious marine accidents compared to other types of vessels. The accident investigation report data reveals that human factors account for approximately 50.3% of accidents and have significant consequences. Additionally, a marine accident severity classification model was developed, and one of the ten machine learning models used yielded the best performance. The XGB classifier model is superior in classifying the severity of marine accidents due to its higher average F1 score compared to other models. While it can be used to predict future marine accidents, its reliability percentage is only 70-74%.

The results of this study show a better performance of about 5%-10% compared to the study developed by Mullai and Paulsson, which showed a performance of 65%. This improvement is very significant in the development of marine transportation safety science so that accidents and misfortunes can be prevented. Thus, marine transportation safety is not just about reducing the risk of accidents but also about ensuring better access, economic development, and overall societal well-being. All of this contributes to the achievement of sustainable development goals.



## REFERENCES

- [1] Allianz. (2015). Safety and shipping review 2015. <http://www.agcs.allianz.com/assets/PDFs/Reports/Shipping-Review-2015.pdf>, accessed on Dec. 11, 2023.
- [2] Chen, J., Bian, W., Wan, Z., Yang, Z., Zheng, H., Wang, P. (2019). Identifying factors influencing total-loss marine accidents in the world: Analysis and evaluation based on ship types and sea regions. *Ocean Engineering*, 191: 106495. <https://doi.org/10.1016/j.oceaneng.2019.106495>
- [3] Ozturk, U., Cicek, K. (2019). Individual collision risk assessment in ship navigation: A systematic literature review. *Ocean Engineering*, 180: 130-143. <https://doi.org/10.1016/j.oceaneng.2019.03.042>
- [4] UNCTAD. (2021). Review of Maritime Report 2021. [https://unvtad.org/en/PublicationsLibrary/rmt2015\\_en.pdf](https://unvtad.org/en/PublicationsLibrary/rmt2015_en.pdf), accessed on Jan. 21, 2024.
- [5] Weng, J., Yang, D. (2015). Investigation of shipping accident injury severity and mortality. *Accident Analysis & Prevention*, 76: 92-101. <https://doi.org/10.1016/j.aap.2015.01.002>
- [6] Wang, L., Yang, Z. (2018). Bayesian network modelling and analysis of accident severity in waterborne transportation: A case study in China. *Reliability Engineering & System Safety*, 180: 277-289. <https://doi.org/10.1016/j.ress.2018.07.021>
- [7] Lu, J., Su, W., Jiang, M., Ji, Y. (2022). Severity prediction and risk assessment for non-traditional safety events in sea lanes based on a random forest approach. *Ocean & Coastal Management*, 225: 106202. <https://doi.org/10.1016/j.ocecoaman.2022.106202>
- [8] Cakir, E., Sevgili, C., Fiskin, R. (2021). An analysis of the severity of oil spills caused by vessel accidents. *Transportation Research Part D: Transport and Environment*, 90: 102662. <https://doi.org/10.1016/j.trd.2020.102662>
- [9] Wang, H., Liu, Z., Wang, X., Huang, D., Cao, L., Wang, J. (2022). Analysis of the injury severity outcomes of maritime accidents using a zero-inflated ordered probit model. *Ocean Engineering*, 258: 111796. <https://doi.org/10.1016/j.oceaneng.2022.111796>
- [10] Zhou, X., Cheng, L., Li, M. (2020). Assessing and mapping maritime transportation risk based on spatial fuzzy multi-criteria decision making: A case study in the South China Sea. *Ocean Engineering*, 208: 107403. <https://doi.org/10.1016/j.oceaneng.2020.107403>
- [11] Mullai, A., Paulsson, U. (2011) A grounded theory model for analysis of marine accidents. *Accident Analysis and Prevention*, 43(4): 1590-1603. <https://doi.org/10.1016/j.aap.2011.03.022>
- [12] Kamal, B., Çakır, E. (2022). Data-driven Bayes approach on marine accidents occurring in Istanbul strait. *Applied Ocean Research*, 123: 103180. <https://doi.org/10.1016/j.apor.2022.103180>
- [13] Cao, Y., Wang, X., Yang, Z., Wang, J., Wang, H., Liu, Z. (2023). Research in marine accidents: A bibliometric analysis, systematic review and future directions. *Ocean Engineering*, 284: 115048. <https://doi.org/10.1016/j.oceaneng.2023.115048>
- [14] Cao, Y., Wang, X., Wang, Y., Fan, S., Wang, H., Yang, Z., Liu, Z., Wang, J., Shi, R. (2023). Analysis of factors affecting the severity of marine accidents using a data-driven Bayesian network. *Ocean Engineering*, 269: 113563. <https://doi.org/10.1016/j.oceaneng.2022.113563>
- [15] IMO. (2014). Casualties. <https://www.imo.org/en/OurWork/MSAS/Pages/Casualties.aspx>, accessed on Feb. 11, 2024.
- [16] Zhang, F., Zhang, Z., Keung, JW., Tang, X., Yang, Z., Yu, X., Hu, W. (2024). Data preparation for deep learning based code smell detection: A systematic literature review. *Journal of Systems and Software*, 216: 112131. <https://doi.org/10.1016/j.jss.2024.112131>
- [17] Passarella, R., Iqbal, M.D., Buchari, M.A., Veny, H. (2023). Analysis of commercial airplane accidents worldwide using K-means clustering. *International Journal of Safety and Security Engineering*, 13(5): 813-819. <https://doi.org/10.18280/ijss.130505>
- [18] Chan, KR. (2021). How to use the Lazy Predict library to select the best machine learning model: Using the best machine learning tools to answer the right questions. *Omics Diary: Medium*. <https://medium.com/omics-diary/how-to-use-the-lazy-predict-library-to-select-the-best-machine-learning-model-65378bf4568e>, accessed on Jan. 25, 2024.
- [19] Passarella, R., Nurmaini, S., Rachmatullah, M.N., Veny, H., Hafidzoh, F.N.N. (2024). Development of a machine learning model for predicting abnormalities of commercial airplanes. *Data Science and Management*. <https://doi.org/10.1016/j.dsm.2024.03.002>
- [20] Feki, R. (2022) Imbalanced data: Best practices: A guide to deliver great results on ML models with imbalanced datasets. <https://rihab-feki.medium.com/imbalanced-data-best-practices-f3b6d0999f38>, accessed on Apr. 11, 2024.
- [21] Brownlee, J. (2021). Tour of evaluation metrics for imbalanced classification. <https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/>, accessed on Apr. 12, 2024.