

# Application of data-mining techniques to predict and rank maritime non-conformities in tanker shipping companies using accident inspection reports

Beatriz Navas de Maya<sup>a</sup>, Ozcan Arslan<sup>b</sup>, Emre Akyuz<sup>b</sup>, Rafet Emek Kurt<sup>a</sup> and Osman Turan<sup>a</sup>

*a. Department of Naval Architecture, Ocean & Marine Engineering. Strathclyde University, Glasgow G1 1XW, UK*

*b. Department of Maritime Transport and Management Engineering. Istanbul Technical University*

Corresponding author: Rafet Emek Kurt: [rafet.kurt@strath.ac.uk](mailto:rafet.kurt@strath.ac.uk)

## Abstract

The application of data mining techniques is an extended practice in numerous domains; however, within the context of maritime inspections, the aforementioned methods are rarely applied. Thus, the application of data-mining techniques for the prediction and ranking of non-conformities identified during vessel inspections could be of significant managerial contribution to the safety of shipping companies, as non-conformities could potentially lead to real accidents if not addressed adequately. Hence, specific data mining methods are investigated and applied in this paper to predict and rank non-conformities on oil tankers using a database recorded by tanker shipping companies in Turkey from 2006 to 2019. The results of this study reveal that specific non-conformities, for instance, inadequate ice operations or inadequate general appearance and condition of hull, superstructure and external weather decks, are not company-based problems, rather they are industry wide issues for all tanker shipping companies.

Keywords: Maritime safety, non-conformities, maritime inspections; data mining techniques; WEKA

## 1. Introduction

Safety is of paramount importance for ships in maritime transportation since more than 90% of world trading is performed by maritime transport (Chauvin et al., 2013). Considering the hazardous nature of the work conducted on the ships, maritime safety is of paramount importance, as maritime accidents are responsible for economic losses and social impact (Eliopoulou, Papanikolaou, & Voulgarellis, 2016), and are usually associated with injuries or fatalities, loss of goods and cargo, and defined as a source of damage for the environment (Wang, 2002; Luo & Shin, 2016). Analysing the available literature, it becomes evident that humans have played a major role in past maritime accidents (Smith, Veitch, Khan, & Taylor, 2017). Statistical analyses on industrial casualties indicate that Human Factors (HFs) are the major causes of at least 66% of the accidents, and more than 90% of the incidents in various strategic industries such as aerospace or aviation (O'Hare, Wiggins, Batt, & Morrison, 1994; Ali Azadeh & Zarrin, 2016). However, they are not isolated sectors, as more than 80% of accidents in maritime transportation are due to human factors (Akyuz and Celik, 2018; Turan et al. 2016). In order to prevent

an accident at sea, the maritime regulatory body, the International Maritime Organisation (IMO), has been adopting a set of rules and regulations. The SOLAS Convention (International Convention for the Safety of Life at Sea), for instance, was adopted by IMO in 1974 to specify minimum standards for the construction, equipment and operation of ships. Likewise, MARPOL 73/78 (The International Convention for the Prevention of Pollution from Ships) came into force in response to a spate of tanker accidents in 1976-1977. The aim of the convention is to prevent pollution from ships caused by operational or accidental causes. The COLREG (The International Regulations for Preventing Collisions at Sea), which are a set of rules to be followed by navigation officers to avoid collisions at sea, was adopted in 1972. Although numerous conventions were adopted by the IMO to minimize maritime accident and enhance safety at sea, maritime accidents are still occurring. Particularly, tanker ship accidents pose acute dangers to life and marine environment considering the hazardous nature of the cargoes carried by these types of vessels.

Moreover, various occupational accidents are likely to occur on tanker ships due to the operations or deficiencies in the systems or procedures. Therefore, a set of inspections and audits are required to make sure that potential hazards are identified, and necessary mitigating actions are taken by the company to maximise the safety on board tanker vessels. For instance, SIRE (Ship Inspection Report Programme) is a standard assessment carried out by tanker owners and operators to identify operational deficiencies and non-conformities (Arslan et al., 2016). The SIRE inspection is used to determine whether a tanker follows SIRE standards set by the OICMF (The Oil Companies International Marine Forum). OICMF aims to make sure that vessels are well managed and maintained to enhance the operational safety of tankers, barges and offshore vessels (Grbic et al., 2018). The SIRE inspection can be performed every six months, and the programme involves a review of the tanker's documentation, followed by an inspection of the bridge, communications, main deck, mooring stations, pump room, cargo control room, engine, accommodation and steering compartments (OICMF, 2014). The non-conformities, which is defined as "an observed situation where objective evidence indicates the non-fulfilment of a specified requirement", are determined during SIRE inspection (Christensen, 2013). The non-conformities in the SIRE inspection report do not result in the detention of the ship but allow the charterer to see the condition of the ship. The most common non-conformities detected in SIRE inspections are related to human factors, managerial deficiencies, navigational deficiencies and cargo equipment (Arslan et al., 2016). The reviewing of the non-conformities can be used as a means to determine the focal points, which may affect the operational safety of tanker ships. Therefore, this research study will systematically investigate aforementioned non-conformities with the main aim of determining the important areas which may affect the operational safety of tanker ships

While the context and objective of this study have been introduced above, the rest of this paper is structured as follows: Section 2 provides an initial review of the previous analyses of non-conformities in the maritime sector, a detailed literature review of data-mining techniques and their applications in the maritime sector, and a description of the WEKA software. Section 3 describes the methodology which is adopted and includes the results of this research. Finally, Section 4 and Section 5 provide the discussion and the conclusions, respectively.

## **2. Literature Review**

### **2.1 Non-conformities in the maritime sector**

Performing regular SIRE Inspections can assist in improving tankers' safety and in ensuring consistent quality of service in the tanker shipping industry. The key to preventing accidents on-board tanker ships is to understand the non-conformities. The consequences of the non-conformities can potentially harm humans, the marine environment and cargo since it may cause maritime accidents. Thus, maritime safety researchers focused on analysing non-conformities and deficiencies detected on tanker ships. For instance, Raman and Shankaranarayanan (2018) discussed the importance of operational aspects of the vessel assessment Questionnaire (VIQ) inside the SIRE programme. In their paper, the authors used the Fuzzy Analytic Hierarchy Process (FAHP) method to compose a planned timetable for VIQ execution in the tanker shipping industry. Likewise, detailed SIRE inspection data between years 2006-2015 was reviewed and analysed to investigate non-conformities and the root causes of the non-conformities with tanker ships by Arslan et al. (2016). Heij et al. (2011) performed comprehensive data mining covering 400,000 ship arrivals between 2002–2007, in order to investigate the safety gains of inspection rules as well as options for further developments. The data utilised in this study was composed of the following sources: port state control and industry inspections and casualty investigations. Cebi et al. (2009) prioritized the operational items of the VIQ within the SIRE programme and recommended additional safety precaution for tankers, combination carriers, shuttle tankers, chemical tankers, and gas carriers. Obtaining detailed historical data regarding the maritime incidents such as non-conformities, deficiencies or near misses is very challenging. Therefore, the approaches dealing with dataset would be highly appreciated by the maritime safety researchers.

### **2.2 Data mining techniques in the maritime sector**

Data mining can be defined as a set of mechanisms and techniques, realized in software, utilised to extract implicit, previously hidden and potentially useful information from large databases (Frawley et al., 1991). Data mining techniques, involving classification trees, Bayesian networks, decision rules, support vector machines and logistic regression, have been adopted in different industries for decades. Data mining can initiate a level of algorithm which may allow software applications to create more accurate in forecasting outputs (Akyuz et al., 2019). There are limited data mining techniques applied in the maritime industry since most of the equipment operated on-board ship is capable of capturing dynamic data. Bridge navigational equipment such as GPS, AIS, ECDIS and ARPA radar generate a dynamic dataset that can be utilised for forecasting. The AIS (Automatic Identification System) dataset, for instance, has been widely used in the maritime industry to determine efficient waypoints for voyage planning, voyage optimisation, bunker saving application, etc. (Yang et al., 2019; Mao et al., 2016; Dobrkovic et al., 2015; Tang and Shao, 2009).

There were also some studies undertaken to analyse the VDR (Voyage Data Recorder) dataset to improve maritime security and accident investigation (Jovic et al., 2019 Jian-she, 2010). Based on the basic literature review of the data mining application in the maritime industry, it appears that the topic has not received the amount of attention it deserves.

### 2.3 The WEKA software

The WEKA (Waikato Environment for Knowledge Analysis) was introduced for data analysis and predictive modelling (Witten and Frank, 2005) and including a variety of data mining algorithms. It is generally used for teaching, research, and industrial applications. The WEKA includes a plethora of built-in tools for standard machine learning tasks and additionally provides transparent access to toolboxes. A range of different learning algorithms is provided along with techniques for pre- and post-processing. It is capable of evaluating the results of the learning assesses schemes on any given dataset (Witten et al., 2016). The WEKA utilises numerous machine learning algorithms to capture and analyse datasets, which include data pre-treatment, classification, cluster class, association rule mining, regression and visualization (Zhong, 2011).

Although considerable research on the maritime domain has been undertaken in the past decade, those dedicated to WEKA-based data mining have remained rather limited. There are a few studies, which were performed by using WEKA. For instance, Vanneschi et al. (2015) used WEKA software to estimate the position of vessels, based on information related to the vessel's past positions in a specific time interval. Analysis of marine traffic flow characteristics was determined based on WEKA data mining (Zheng et al., 2009). Another study was performed to predict the ship emergency preparedness level of commercial vessels in the case of fire (Tac et al., 2018). In the paper, authors analysed fire drills carried out on-board ships.

## 3 Methodology

The data which is analysed and reported in this paper includes records of non-conformities from Turkish shipping companies for the period 2006-2019, as part of the Ship Inspection Report (SIRE) programme. The aforementioned period of time was selected as no data on non-conformities was provided by the companies included in this study before 2006. In addition, specific inclusion criteria for this study are displayed in Table 1.

*Table 1. Criteria for initial inclusion in the analysis*

| Criterion                         | Value  |
|-----------------------------------|--|
| Period                            | 2006-2019  |
| Inspection guides                 | SIRE   |
| Companies                         | Companies with more than 250 non-conformities reported |
| No of non-conformities considered | 2207   |

The research design consisted of three stages: 1) Initial data screening, 2) Problem description and available data, and 3) Application of data mining techniques to identify and rank most frequent maritime non-conformities. A more concise description of the steps conducted at each stage is provided below.

### 3.1 Stage 1: Initial data screening

This stage involved an initial screening of the non-conformities database in defining the sampling plan, and this resulted in the exclusion of 4303 cases. The excluded non-conformities from this study are due to the specific constraints applied to the non-conformities database as follows: First, non-conformities which were not reported under the SIRE programme (e.g. inspection conducted by the Chemical Distribution Institute (CDI), terminal inspections, etc.). Second, cases in which the non-conformities were not reported under the selected period of

time, or the aforementioned information was missing. Third, non-conformities which did not refer to a specific chapter on the SIRE programme. Fourth, cases where no company data was available (i.e. missing data). And fifth, cases where the minimum number of entries for a specific company was insufficient. The aforementioned minimum number of entries to perform this analysis was set at 250 entries due to two main reasons. First, this was considered by the authors as a reasonable data sample to obtain reliable results. Second, multiple companies within the database did not include enough data entries. Hence, due to this unbalance distribution and the effect that it might cause in the results, it was decided to set a minimum requirement in terms of entries, and those companies which did not include enough data entries were considered out of the scope of this paper. After the screening process, the non-conformities database under the scope of this study contained N=2207 samples, as shown in Table 2.

Table 2. Number of non-conformities considered in this study after each constraint

| No         | Constraint                                     | Cases | Excluded | Total excluded | Retained |
|------------|--|-------|----------|----------------|----------|
|            | Initial database with non-conformities         | 7012  | 0        | 0              | 100.00%  |
| <b>I</b>   | Non-conformities obtained under SIRE programme | 4390  | 2622     | 2622           | 62.61%   |
| <b>II</b>  | Period 2006-2020                               | 4261  | 129      | 2751           | 60.77%   |
| <b>III</b> | Entries that has missing chapters data         | 3759  | 502      | 3253           | 53.61%   |
| <b>IV</b>  | Companies with insufficient number of entries  | 3668  | 91       | 2842           | 52.31%   |
| <b>V</b>   | Data that falls under the scope of this study  | 2207  | 1461     | 4303           | 31.47%   |

### 3.2 Stage 2: Problem description and available data

Addressing the most frequent non-conformities will be of significant managerial contribution to the safety of the tanker shipping industry, as non-conformities could potentially lead to real accidents if additional safety barriers in place fail to perform their functions. Hence, efforts must be focused on developing a suitable approach which is able to identify those non-conformities with a significant appearance in shipping companies. Thus, specific strategies can be developed and adopted by shipping companies to prevent reoccurrences of the aforementioned non-conformities, in order to mitigate them and to enhance overall company safety.

Therefore, this paper proposes to apply specific data mining techniques to a non-conformities database for the prediction and ranking of main non-conformities in shipping companies. The full list of non-conformities in the database and their frequencies included in this analysis is provided in Table 3.

Table 3. List of non-conformities identified from the database

| Abb  | Non-conformity   | %     | Abb  | Non-conformity   | %     |
|------|--|-------|------|--|-------|
| NC36 | Inadequate ice operations  | 8.84% | NC62 | Inadequate ship security   | 0.68% |
| NC60 | Inadequate safety management related to engine and steering compartments                       | 5.62% | NC63 | Inadequate shipboard oil and marine pollution emergency plans                              | 0.68% |
| NC53 | Inadequate policies, procedures and documentation related to engine and steering compartments  | 4.94% | NC70 | Inadequate collaging , sampling and closed operations related to cargo and ballast systems | 0.68% |
| NC33 | Inadequate general appearance and condition of hull, superstructure and external weather decks | 4.67% | NC71 | Inadequate venting arrangements related to cargo and ballast systems                       | 0.68% |
| NC54 | Inadequate policies, procedures and navigational documentation                                 | 3.90% | NC06 | Inadequate ballast water management  | 0.63% |
| NC38 | Inadequate lifesaving equipment  | 3.76% | NC27 | Inadequate firefighting equipment related to engine and steering compartments              | 0.63% |

| Abb  | Non-conformity   | %     | Abb  | Non-conformity   | %     |
|------|--|-------|------|--|-------|
| NC11 | Inadequate cargo operations and related safety management  | 3.40% | NC65 | Inadequate stability and cargo loading limitations                         | 0.63% |
| NC28 | Inadequate firefighting equipment related to safety management   | 3.35% | NC50 | Inadequate planned maintenance related to engine and steering compartments | 0.59% |
| NC39 | Inadequate machinery status related to engine and steering compartments  | 3.22% | NC55 | Inadequate publications  | 0.59% |
| NC47 | Inadequate navigation equipment  | 3.17% | NC31 | Inadequate general appearance and condition of accommodation areas         | 0.54% |
| NC58 | Inadequate safety management   | 3.17% | NC03 | Inadequate access  | 0.50% |
| NC10 | Inadequate cargo operations and deck area pollution prevention   | 2.76% | NC68 | Inadequate structural condition  | 0.50% |
| NC07 | Inadequate cargo handling and monitoring equipment   | 2.54% | NC42 | Inadequate monitoring of non-cargo spaces                                  | 0.45% |
| NC35 | Inadequate hot work procedures   | 2.40% | NC25 | Inadequate engine and steering compartments                                | 0.41% |
| NC13 | Inadequate charts and publications   | 2.27% | NC51 | Inadequate policies and procedures for maritime security                   | 0.41% |
| NC43 | Inadequate mooring equipment   | 2.27% | NC64 | Inadequate single point moorings   | 0.41% |
| NC73 | Statutory certificates are not valid, or the annual and intermediate surveys have been not carried out within the required range dates | 2.13% | NC67 | Inadequate steering compartments   | 0.41% |
| NC17 | Inadequate crew management   | 1.95% | NC21 | Inadequate drug and alcohol policy   | 0.36% |
| NC69 | Inadequate survey and repair history   | 1.95% | NC41 | Inadequate Material Safety Data Sheets (MSDS)                              | 0.36% |
| NC44 | Inadequate mooring equipment documentation and management  | 1.90% | NC56 | Inadequate pump rooms and oil discharge monitors                           | 0.36% |
| NC32 | Inadequate general appearance and condition of electrical equipment  | 1.77% | NC05 | Inadequate anti-pollution  | 0.32% |
| NC40 | Inadequate manifold arrangements   | 1.72% | NC08 | Inadequate cargo hoses   | 0.32% |
| NC18 | Inadequate crew qualifications   | 1.54% | NC12 | Inadequate cargo pump room   | 0.32% |
| NC30 | Inadequate gas analysing equipment   | 1.54% | NC14 | Inadequate communication equipment   | 0.27% |
| NC48 | Inadequate oil record books for pollution prevention   | 1.18% | NC23 | Inadequate emergency towing arrangements                                   | 0.27% |
| NC45 | Inadequate mooring procedures  | 1.13% | NC09 | Inadequate cargo lifting equipment   | 0.23% |
| NC04 | Inadequate anchoring equipment   | 1.09% | NC19 | Inadequate cyber security  | 0.23% |
| NC24 | Inadequate enclosed space and pump room entry procedures   | 1.09% | NC16 | Inadequate Condition Assessment Scheme                                     | 0.18% |
| NC57 | Inadequate safety equipment  | 1.00% | NC34 | Inadequate general appearance and condition of internal spaces             | 0.18% |
| NC46 | Inadequate navigation  | 0.91% | NC72 | Inadequate vessel structure  | 0.18% |
| NC66 | Inadequate static electricity precautions  | 0.91% | NC26 | Inadequate enhance survey programme  | 0.09% |
| NC02 | General vessel information is missing  | 0.82% | NC49 | Inadequate personnel management related to cargo and ballast systems       | 0.09% |
| NC20 | Inadequate drills, training and familiarisation  | 0.82% | NC61 | Inadequate sample arrangements   | 0.09% |
| NC52 | Inadequate policies, procedures and documentation related to cargo and ballast systems   | 0.77% | NC01 | General appearance and condition of the accommodation areas                | 0.05% |
| NC29 | Inadequate garbage management  | 0.72% | NC22 | Inadequate Dynamic Positioning (DP) Operations                             | 0.05% |
| NC59 | Inadequate safety management and inadequate operator's procedures manuals  | 0.72% | NC37 | Inadequate inert gas system  | 0.05% |
| NC15 | Inadequate communication procedures  | 0.68% |      |  |       |

In the available database, only five companies are considered for the study as they comply with the fifth constraint defined at Stage 1 (i.e. the minimum number of required entries 250). Thus, the number of non-conformities for each company is displayed in Figure 1.

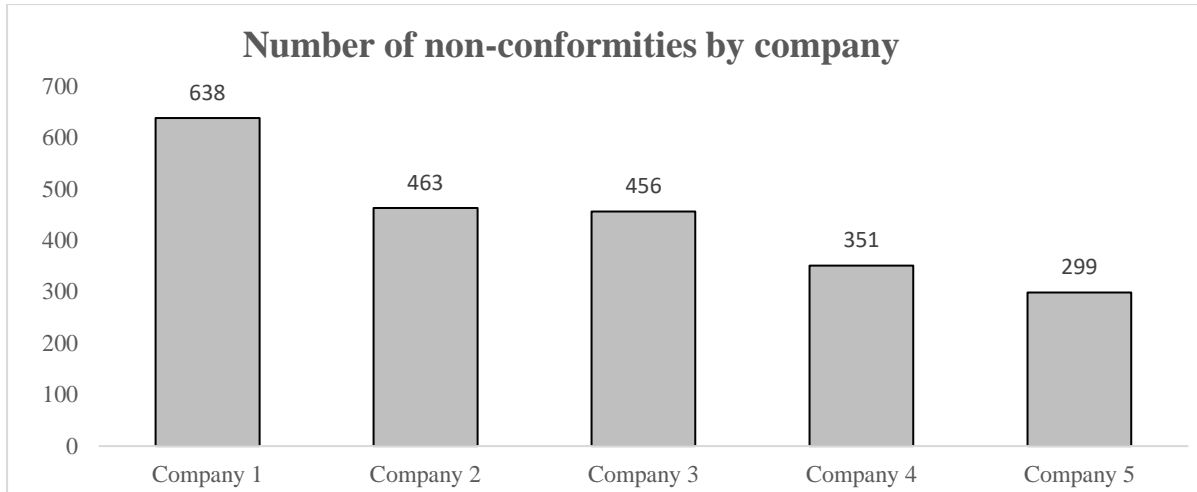


Figure 1. Number of non-conformities by each company

In addition, the prediction and ranking of main non-conformities proposed in this paper are applied to different regroupings of the above companies, aiming to demonstrate our hypothesis, which states that the main non-conformities remain significantly constant and are independent of the company analysed. The final aim is to demonstrate that non-conformities are not a company-based problem, as same non-conformities were identified again and again. Hence, efforts should be directed towards addressing these non-conformities. Therefore, the method is applied to three different groupings (see Table 4) of the above displayed companies as given below:

- **Group 1 (G1):** The five companies have been considered as independent categories. However, it should be noted that the number of data entries in each category differ from one company to another. Thus, an unbalanced distribution is observed amongst categories (i.e. there is a difference of 339 data entries between Company 1 and Company 5).
- **Group 2 (G2):** In order to mitigate the unbalancing effect between companies and to increase the number of data points for each category, Company 4 and Company 5 have been aggregated into one main category (C4+C5). However, although the maximum difference amongst data entries between companies has been reduced, the distribution between categories still remains unbalanced.
- **Group 3 (G3):** Finally, in order to mitigate the unbalancing effect among categories and to increase the number of data points for each category, the companies have been aggregated into two groups, with a similar data distribution, the first group incorporate Company 1 and Company 2, while the second group incorporates the remaining companies (C3+C4+C5).

Table 4. Proposed aggregation and sizes

| Abb | Companies | Size | Abb | Companies | Size | Abb | Companies | Size |
|-----|-----------|------|-----|-----------|------|-----|-----------|------|
| G1  | C1        | 638  | G2  | C1        | 638  | G3  | C1+C2     | 1101 |
|     | C2        | 463  |     | C2        | 463  |     | C3+C4+C5  | 1106 |
|     | C3        | 456  |     | C3        | 456  |     |           |      |
|     | C4        | 351  |     | C4+C5     | 650  |     |           |      |
|     | C5        | 299  |     |           |      |     |           |      |

### 3.3 Stage 3: Application of data mining techniques

In order to address the specific prediction and ranking problem described in the previous section, the authors will first investigate data mining techniques to identify the most suitable techniques for the formulated problem. Given the available data, the problem under consideration is to predict the most influential non-conformities which have been identified on board the vessels which are operated by the above companies for the period under analysis. Then, those data mining techniques with the higher success rate for prediction will be applied to rank the non-conformities for each of the proposed group sets, aiming to demonstrate that non-conformities are not a company-based problem, but an overall problem within the tanker shipping industry.

#### Evaluation of various forecasting data mining techniques

Some of the data mining techniques most commonly used in the literature are classification trees, Bayesian networks, decision rules, and logistic regression (Rivas et al., 2011). In order to evaluate the most suitable technique for prediction, all of the database will be used for both training and testing. Table 5 displays the success rates for each data mining technique for forecasting non-conformities in the provided database. It can be seen that the Logistic, the Tree-J48, and the Rule (OneR) are the techniques with a higher success rate.

Table 5. Success rates for forecasting the non-conformity by WEKA for the first group (G1)

| WEKA model  | Success rate | WEKA model                     | Success rate |
|-------------|--------------|--------------------------------|--------------|
| Logistic    | 35.30%       | Tree (LMT)                     | 34.98%       |
| Tree (J48)  | 35.30%       | Function Multilayer Perceptron | 31.85%       |
| Rule (OneR) | 35.30%       | Rule (PART)                    | 30.99%       |
| Naïve Bayes | 30.99%       |                                |              |

Therefore, the above techniques are further investigated. Table 6, Table 7, and Table 8 report the confusion matrices for the Logistic, the Tree-J48, and the Rule (OneR) techniques on the different group problems described above. Thus, the aforementioned confusion matrices provide a further insight into the abilities of each data mining technique to predict a non-conformity in a specific company. For instance, the Logistic technique in the first aggregation problem (Table 6) will be able to predict successfully 416 non-conformities related to Company 1. In addition, in 196 cases, this technique will produce a false negative result, which consists of a wrong prediction of a non-conformity (a non-conformity from Company 2 is predicted as a non-conformity from Company 1). Moreover, in 103 cases, the logistic technique will produce a false positive result, which consists of misclassifying a non-conformity (in this case, a non-conformity from Company 1 is wrongly predicted as a non-conformity from Company 2).

Table 6. Confusion matrices of G1 on the different data mining techniques (Logistic in the left, Tree-48 in the middle, and Rule (OneR) in the right)

|    | Logistic |     |     |    |    |    | Tree-J48 |     |     |    |    |    | Rule (OneR) |     |     |    |    |
|----|----------|-----|-----|----|----|----|----------|-----|-----|----|----|----|-------------|-----|-----|----|----|
|    | C1       | C2  | C3  | C4 | C5 |    | C1       | C2  | C3  | C4 | C5 |    | C1          | C2  | C3  | C4 | C5 |
| C1 | 416      | 103 | 77  | 32 | 10 | C1 | 426      | 109 | 76  | 17 | 10 | C1 | 426         | 109 | 76  | 17 | 10 |
| C2 | 196      | 183 | 60  | 17 | 7  | C2 | 199      | 192 | 58  | 7  | 7  | C2 | 199         | 192 | 58  | 7  | 7  |
| C3 | 237      | 85  | 110 | 21 | 3  | C3 | 237      | 89  | 109 | 18 | 3  | C3 | 237         | 89  | 109 | 18 | 3  |



| Logistic |     |    |    |    | Tree-J48 |    |     |    |    | Rule (OneR) |    |    |     |    |    |    |    |
|----------|-----|----|----|----|----------|----|-----|----|----|-------------|----|----|-----|----|----|----|----|
| C1       | C2  | C3 | C4 | C5 | C1       | C2 | C3  | C4 | C5 | C1          | C2 | C3 | C4  | C5 |    |    |    |
| C4       | 163 | 74 | 56 | 55 | 3        | C4 | 176 | 79 | 56 | 37          | 3  | C4 | 176 | 79 | 56 | 37 | 3  |
| C5       | 170 | 53 | 43 | 18 | 15       | C5 | 174 | 58 | 41 | 11          | 15 | C5 | 174 | 58 | 41 | 11 | 15 |

Table 7. Confusion matrices of G2 on the different data mining techniques (Logistic in the left, Tree-48 in the middle, and Rule (OneR) in the right)

| Logistic |     |     |       |     | Tree-J48 |     |     |       |     | Rule (OneR) |     |     |       |     |
|----------|-----|-----|-------|-----|----------|-----|-----|-------|-----|-------------|-----|-----|-------|-----|
| C1       | C2  | C3  | C4+C5 |     | C1       | C2  | C3  | C4+C5 |     | C1          | C2  | C3  | C4+C5 |     |
| C1       | 299 | 83  | 43    | 213 | C1       | 305 | 84  | 51    | 198 | C1          | 305 | 84  | 51    | 198 |
| C2       | 149 | 159 | 37    | 118 | C2       | 153 | 162 | 47    | 101 | C2          | 153 | 162 | 47    | 101 |
| C3       | 162 | 67  | 64    | 163 | C3       | 164 | 67  | 80    | 145 | C3          | 164 | 67  | 80    | 145 |
| C4+C5    | 189 | 89  | 39    | 333 | C4+C5    | 194 | 92  | 56    | 308 | C4+C5       | 194 | 92  | 56    | 308 |

Table 8. Confusion matrices of G3 on the different data mining techniques (Logistic in the left, Tree-48 in the middle, and Rule (OneR) in the right)

| Logistic |          |     | Tree-J48 |          |     | Rule (OneR) |          |     |
|----------|----------|-----|----------|----------|-----|-------------|----------|-----|
| C1+C2    | C3+C4+C5 |     | C1+C2    | C3+C4+C5 |     | C1+C2       | C3+C4+C5 |     |
| C1+C2    | 664      | 437 | C1+C2    | 664      | 437 | C1+C2       | 664      | 437 |
| C3+C4+C5 | 456      | 650 | C3+C4+C5 | 456      | 650 | C3+C4+C5    | 456      | 650 |

### Application of data mining techniques for ranking of non-conformities

The specific problem described here is to rank the non-conformities based on their ability to be influential in predicting the company or group which is affected in the set of data available (i.e. those non-conformities with a higher successful prediction rate in a company or group will be ranked higher). Thus, each of the above selected data mining techniques will be applied to the proposed aggregation sets to demonstrate the hypothesis that, based on the available data, the top non-conformities in the ranking are the most influential in being identified by both, particular shipping companies and all tanker shipping companies. This will demonstrate that non-conformities are not a company-based problem but an overall problem in the maritime industry.

Table 9. Top non-conformities ranked by both, Logistic and Rule (OneR) on the different data groups

| G1       |      | G2       |      | G3       |      |
|----------|------|----------|------|----------|------|
| Position | NC   | Position | NC   | Position | NC   |
| 1        | NC36 | 1        | NC36 | 1        | NC36 |
| 2        | NC33 | 2        | NC39 | 2        | NC33 |
| 3        | NC42 | 3        | NC66 | 3        | NC60 |
| 4        | NC48 | 4        | NC60 | 4        | NC53 |
| 5        | NC60 | 5        | NC33 | 5        | NC38 |
| 6        | NC53 | 6        | NC53 | 6        | NC54 |
| 7        | NC11 | 7        | NC38 | 7        | NC28 |
| 8        | NC38 | 8        | NC54 | 8        | NC39 |

| G1       |      | G2       |      | G3       |      |
|----------|------|----------|------|----------|------|
| Position | NC   | Position | NC   | Position | NC   |
| 9        | NC54 | 9        | NC28 | 9        | NC11 |
| 10       | NC28 | 10       | NC11 | 10       | NC47 |

Table 10. Top non-conformities ranked by Tree-J48 on the different data groups

| G1       |      | G2       |      | G3       |      |
|----------|------|----------|------|----------|------|
| Position | NC   | Position | NC   | Position | NC   |
| 1        | NC36 | 1        | NC36 | 1        | NC36 |
| 2        | NC33 | 2        | NC39 | 2        | NC33 |
| 3        | NC42 | 3        | NC66 | 3        | NC60 |
| 4        | NC48 | 4        | NC60 | 4        | NC53 |
| 5        | NC39 | 5        | NC33 | 5        | NC38 |
| 6        | NC60 | 6        | NC53 | 6        | NC54 |
| 7        | NC53 | 7        | NC38 | 7        | NC28 |
| 8        | NC38 | 8        | NC54 | 8        | NC47 |
| 9        | NC54 | 9        | NC28 | 9        | NC11 |
| 10       | NC28 | 10       | NC11 | 10       | NC58 |

#### 4 Discussion

In this study, the authors applied various data mining techniques to non-conformities recorded by tanker shipping companies in Turkey from 2006 to 2019. Looking at the findings, the data mining used under Logistic and Rule (OneR) gives the same results for all aggregations. Table 9 and Table 10 shows that the top non-conformities ranked by Logistic, Rule (OneR) and Tree-J48 on the different data aggregations are almost the same. It shows that the techniques used for data mining and obtained results are consistent and the approach adopted for analysis is reasonable. Also, it proves that the Logistic, the Rule (OneR) and the Tree-J48 techniques present a higher success rate in predicting the ranking of the non-conformities when compared to additional techniques such as Naïve Bayes or Tree (LMT). Based on the findings, inadequate ice operations (NC36) is the most significant non-conformities observed during SIRE inspections in tanker ships. The reason for this is that ice notation was made during the construction of ships but most of them were not equipped with the necessary equipment for ice navigation since they are not operating in frozen water. It may be considered as a managerial level defect.

The second most common non-conformity observed during SIRE inspections is defects related to the condition of hull, superstructure and external weather decks (NC33). This deficiency is due to an inadequate budget allocation by ship owners for overhaul maintenance. Most of the tanker owners minimised their maintenance and repair costs after the 2008 economic crisis, and this is reflected in the condition of the ship structures. Inspection of the hull and superstructure condition involves checking for any evidence of structural problems including collision/jetty contact damage, distortion from heavy weather, permitted deck loads or composite materials. Inadequate safety management related to engine and steering compartments is another significant NC, which is observed during SIRE inspections. The reason for this is that HSEQ (Health, Safety, Environment and Quality) managers and safety superintendents fail to develop necessary policies and procedures to improve operational

safety for both navigational and engine room operations. Likewise, inadequate cargo operations and related safety management appears to be another critical non-conformity for tanker ownerships since it is ranked in the top ten non-conformities. The ISM code requirements as per SOLAS (Safety of Life at Sea) convention are of the utmost importance to enhance safety in tanker ships. This deficiency is due to inadequate safety implementations on-board ship and within the shipping company, and may be attributed to human factors. Another influential non-conformity observed during SIRE inspections is inadequate life-saving equipment which can lead to severe consequences for crew health. Life-saving appliances are mandatory as per chapter 3 of the SOLAS convention and are very important to protect human life on-board tanker ships. More deficiencies related to life-saving appliances were found in the lifeboats and life buoys (Hagström and Nilsson, 2017). Deficiencies or inadequacies of life-saving equipment are due to negligent maintenance performed by the ship crew. This may again be attributed to human factors.

The findings show that the majority of non-conformities detected in the SIRE inspection on-board tanker ships are due to three fundamental defects. The first one is addressing construction and managerial deficiencies which include structural problems, shell plating, decks, compartments and ballast tanks. The second one is failing to develop policies and procedures related to the safe operation of ships and cargo. The third one is related to human factors since the human-based errors (decision-based errors, faulty standards and procedures being followed, poor maintenance, etc.) are major contributors to maritime accidents in the shipping industry (Ye et al., 2020). Human error can lead to fatal situations on tanker ships, but the underlying reasons for human errors may possibly be linked to organisational/management deficiencies.

Finally, it is noted that the non-conformities reported in Table 3 are different from the ones obtained employing the proposed approach (Table 9 and Table 10). For example, while inadequate safety management related to engine and steering compartments (NC60) appears as the second factor classified from a statistical point of view, the same factor has been ranked in a different position when applying data mining techniques (i.e. positions 3 to 6 depending on the group under consideration and the applied data mining technique). Moreover, the results reported in Table 3, based on a simplistic evaluation of the frequency of non-conformities, do not support or show any predictive ability, which is the scope of this study. Therefore, the authors believe that the method presented in this paper is capable of providing a further insight with respect to a simple statistical ranking.

## **5 Conclusions**

This paper has investigated the feasibility of various data mining techniques for handling, predicting and ranking non-conformities on board ships. This information can be very useful for shipping companies, as they can allocate efforts on developing and applying specific actions in order to avoid the most frequent predicted non-conformities.

With this aim in mind, a non-conformities oil tanker database from 2006 to 2019 was first processed in order to transform the inspectors' comments into the correspondent section from the SIRE guidelines. Moreover, an initial data screening of the database above was conducted to select the scope of this study. Then, various methods available from the WEKA software (i.e. classification trees, Bayesian networks, decision rules, and logistic regression) were applied and compared to identify the most successful methods for prediction of the top-ranked non-conformities. Thus, those methods with the best success rates for forecasting non-conformities were utilised in the database. Moreover, those data mining techniques were also utilised to rank the above non-conformities for

various data aggregations. Thus, the analysis, based on the available data, confirms that the top non-conformities in the ranking are the same as the most influential non-conformities identified for a particular shipping company and the most influential when considering the overall tanker shipping industry. Analysing non-conformities identified in the SIRE inspection can be beneficial for tanker ship owners, HSEQ managers and safety superintendents to minimize accidents and enhance safety in tanker ships.

Although the proposed approach is novel, it is worth highlighting some limitations. The accuracy of the data mining techniques was as good as expected in this particular study. This might be a consequence of two different factors. First, in the recorded data sample, it is possible that not enough examples were used for training and testing. Second, the variables in terms of non-conformities that were considered within this research were numerous. Nevertheless, the results can still be utilised in terms of decision-making by tanker shipping companies, and they will be of significant contribution to enhance the overall safety of the tanker shipping industry.

Finally, the developed methodology can be applied to other non-conformity databases from different shipping companies (e.g. passenger vessels, fishing vessels, etc.). Thus, in order to increase the size of the available data, other inspection programmes (e.g. inspection conducted by the Chemical Distribution Institute (CDI), terminal inspections, etc.) could be considered in a future study.

## References

- Akyuz, E., Cicek, K., & Celik, M. (2019). A Comparative research of machine learning impact to future of maritime transportation. *Procedia Computer Science*, 158: 275-280.
- Akyuz, E., & Celik, E. (2018). The role of human factor in maritime environment risk assessment: A practical application on Ballast Water Treatment (BWT) system in ship. *Human and Ecological Risk Assessment: An International Journal*, 24(3), 653-666.
- Arslan, O., Aydemir, B., Kececi, T. (2016). Review of Tanker Ship SIRE Inspection Findings in Turkey. *International Conference on Maritime Safety and Human Factors*, At Glasgow, Scotland, U.K.
- Azadeh, A., & Zarrin, M. (2016). An intelligent framework for productivity assessment and analysis of human resource from resilience engineering, motivational factors, HSE and ergonomics perspectives. *Safety Science*, 89, 55-71. doi:<http://dx.doi.org/10.1016/j.ssci.2016.06.001>
- Cebi, S., Celik, M., & Cicek, K. (2009). Prioritization of the VIQ items within SIRE program for an oil tanker ship. In *2009 International Conference on Computers & Industrial Engineering* (pp. 449-452). IEEE.
- Chauvin, C., Lardjane, S., Morel, G., Clostermann, J. P., & Langard, B. (2013). Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS. *Accident Analysis & Prevention*, 59, 26-37.
- Christensen, M. (2013). A qualitative study of the review and verification process of the Safety Management System within companies servicing the Norwegian Continental Shelf (Master's thesis, Høgskolen i Vestfold).
- Dobrkovic, A., Jacob, M. E., & van Hillegersberg, J. (2015). Using machine learning for unsupervised maritime waypoint discovery from streaming AIS data. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business* (p. 16). ACM.
- Eliopoulou, E., Papanikolaou, A., & Voulgarellis, M. (2016). Statistical analysis of ship accidents and review of safety level. *Safety Science*, 85, 282-292. doi:<http://dx.doi.org/10.1016/j.ssci.2016.02.001>
- Frawley, W., Piatetsky-Shapiro, G., & Matheus, C. (1991). Knowledge discovery in database: an overview, ser. Fayyad, U.M. and Piatetsky-Shapiro, G., Smyth, P. and Ulthurusamy, R. (Eds.), *Knowledge Discovery in Database*. MIT Press, Cambridge, MA, 1991.
- Grbić, L., Čulin, J., & Perković, T. (2018). SIRE Inspections on Oil Tankers. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 12.
- Hagström, K., & Nilsson, M. (2017). Testing and evaluation of lifeboat stability-Experimental stability testing and a study of the responsibility delegation between the parties involved in the classification process of life-saving appliances.
- Heij, C., Bijwaard, G. E., & Knapp, S. (2011). Ship inspection strategies: Effects on maritime safety and environmental protection. *Transportation research part D: transport and environment*, 16(1), 42-48.

- Jian-she, H. U. A. N. G. (2010). On Shipborne Voyage Data Record Reconstruction System Based on VDR and AIS Data Fusion. *Journal of Jimei University (Natural Science)*, (2), 11.
- Jović, M., Tijan, E., Marx, R., & Gebhard, B. (2019). Big Data Management in Maritime Transport. *Pomorski zbornik*, 57(1), 123-141.
- Luo, M., & Shin, S.-H. (2016). Half-century research developments in maritime accidents: Future directions. *Accident Analysis & Prevention*. doi:<http://dx.doi.org/10.1016/j.aap.2016.04.010>
- Mao, S., Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G. B. (2018). An automatic identification system (AIS) database for maritime trajectory prediction and data mining. In *Proceedings of ELM-2016* (pp. 241-257). Springer, Cham.
- O'Hare, D., Wiggins, M., Batt, R., & Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics*, 37(11), 1855-1869. doi:10.1080/00140139408964954
- Oil Companies International Marine Forum, Ship Inspection Report (SIRE) Programme. (2014) Edition Rev.1.2, 2016.
- Raman, R. S., & Shankaranarayanan, G. (2018). Prioritization of the VIQ things inside SIRE program for an oil tanker transport. *International Journal of Mechanical and Production Engineering Research and Development (IJMPERD)*, Vol. 8, Issue 1: 857-862.
- Rivas, T., et al. "Explaining and predicting workplace accidents using data-mining techniques." *Reliability Engineering & System Safety*, vol.96 (2011): 739-747.
- Smith, D., Veitch, B., Khan, F., & Taylor, R. (2017). Understanding industrial safety: Comparing Fault tree, Bayesian network, and FRAM approaches. *Journal of Loss Prevention in the Process Industries*, 45, 88-101. doi:<http://dx.doi.org/10.1016/j.jlp.2016.11.016>
- Tac, O.B., Celik, M., Akyuz, E. (2018). Using WEKA data-mining analysis to determine ship emergency preparedness level in case of fire. 7th International Conference on "Innovations in Learning for the Future": Digital Transformation in Education. Page: 56 – 61. September 11-14, Istanbul.
- Tang, C and Shao, Z. (2009). Data mining platform based on AIS data. *Proceedings of the International Conference on Transportation Engineering 2009 (ICTE 2009)*.
- Turan, O., Kurt, R. E., Arslan, V., Silvagni, S., Ducci, M., Liston, P., . . . Papadakis, G. (2016). Can We Learn from Aviation: Safety Enhancements in Transport by Achieving Human Orientated Resilient Shipping Environment. *Transportation Research Procedia*, 14, 1669-1678. doi:<http://dx.doi.org/10.1016/j.trpro.2016.05.132>
- Vanneschi, L., Castelli, M., Costa, E., Re, A., Vaz, H., Lobo, V., & Urbano, P. (2015). Improving maritime awareness with semantic genetic programming and linear scaling: prediction of vessels position based on AIS data. In *European Conference on the Applications of Evolutionary Computation* (pp. 732-744). Springer, Cham.

- Wang, J. (2002). Offshore safety case approach and formal safety assessment of ships. *Journal of Safety Research*, 33(1), 81-115. doi:[http://dx.doi.org/10.1016/S0022-4375\(02\)00005-1](http://dx.doi.org/10.1016/S0022-4375(02)00005-1)
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Witten, I. H., Frank, E., & Hall, M. A. (2005). *Data mining: Practical machine learning tools and techniques*, (morgan kaufmann series in data management systems). Morgan Kaufmann, June, 104, 113.
- Yang, D., Wu, L., Wang, S., Jia, H., & Li, K. X. (2019). How big data enriches maritime research—a critical review of Automatic Identification System (AIS) data applications. *Transport Reviews*, 39(6), 755-773.
- Ye, X., Chen, B., Lee, K., Storesund, R., & Zhang, B. (2020). An integrated offshore oil spill response decision making approach by human factor analysis and fuzzy preference evaluation. *Environmental Pollution*, 114294.
- Zheng, B., Chen, J. B., Xia, S. S., & Jin, Y. X. (2009). Analysis of marine traffic flow characteristics based on data mining. *Navigation of China*, 32(1): 60-63.
- Zhong, X. Y. (2011). The research and application of web log mining based on the platform WEKA. *Procedia Engineering*, 15: 4073 - 4078.