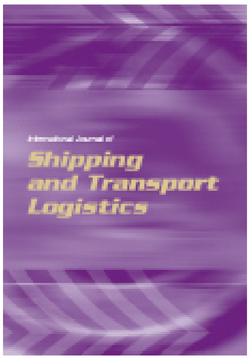


International Journal of Shipping and Transport Logistics

ISSN online: 1756-6525 - ISSN print: 1756-6517
https://www.inderscience.com/ijstl

Service anomaly detection in dry bulk terminals: a machine learning approach

Iñigo L. Ansorena

DOI: 10.1504/IJSTL.2023.10051280

Article History:

Received:	12 June 2021
Accepted:	12 September 2022
Published online:	09 November 2023

Service anomaly detection in dry bulk terminals: a machine learning approach

Iñigo L. Ansorena

Universidad Internacional de La Rioja,
c/ Avenida de la Paz, 137, Logroño, La Rioja, Spain
Email: ilopezans@gmail.com

Abstract: Bulk terminals are complex environments due to a number of variables that affect terminal performance. Although the analysis of big datasets is destined to become an important component of terminal management, previous research has not addressed this issue yet. This paper aims to shed new light on the operation of dry bulk terminals through a two-stage method based on unsupervised machine learning techniques. The first step gives an overview of the terminal's performance, revealing the strongest associations between the variables, while the second calculates an anomaly score for each vessel through an optimised implementation of the isolation forest. As a result, we detect anomalous services which could be directly attributable to the terminal operator. This method can be used to increase transparency in service and assist the terminal operator and ship agents in future contracts.

Keywords: bulk cargo terminals; terminal performance; machine learning; association discovery; anomaly detection; anomalous service; inefficient service; association rules.

Reference to this paper should be made as follows: Ansorena, I.L. (2023) 'Service anomaly detection in dry bulk terminals: a machine learning approach', *Int. J. Shipping and Transport Logistics*, Vol. 17, No. 3, pp.281–302.

Biographical notes: Iñigo L. Ansorena is a Civil Engineer and holds a PhD from the Polytechnic University of Madrid. He is the main author of more than 20 studies, which have been published in scientific journals. His research interests focus on port issues, including machine learning, logistics and operations research. He is currently at the Universidad Internacional de La Rioja (Unir) where he is supervising several master theses.

1 Introduction

UNCTAD (2019) estimated that seaborne trade of dry bulks (major and minor bulks) was 5.2 billion tons in 2018. This volume exceeded the total volume of containerised shipping. However, dry bulk terminals receive considerable less attention than container terminals in the port literature. The planning of operations in a dry bulk terminal is a complex task due to multiple interdependencies between the terminals and the ships, such as storage of materials, type and volume of materials, ship size, number of cargo holds, stresses in the ship, etc. In broad terms, the operation of a bulk carrier can be described by the following parameters: deadweight tonnage, shipload, number of cargo holds, number and type of materials (e.g., iron ore, coking coal, steam coal, etc.), and last but not least the structural stresses during the loading/discharging process. The first parameters vary from ship to ship and the terminal cannot avoid such variations. Only the structural considerations are somehow in the hands of the terminal operator, see IACS (2018).

In this context, the contract (charter party) between the charterer (operator) and the ship-owner sets a laytime period which is the agreed period of time for loading or unloading the ship. The maximum laytime is a limit to be avoided, since an unexpected delay above this maximum will increase the total cost of freight and cause a loss to the seller. Alternatively, if the charterer operates the vessel in a shorter period than the laytime allowed, then the charterer may require the ship-owner to pay a sum of money (despatch) for the time saved. The question that arises is in which cases the delay of the ships is attributable to the service of the terminal, i.e. when the delay is the fault of the terminal operator. Surprisingly this crucial question has not been addressed before in the scientific literature. Moreover, anomalies in dry bulk terminals have not been addressed with the appropriate use of existing data yet. Therefore, to shed new light on this issue we propose a two-step approach based on the discovery of associations and anomalies in service.

The rest of the paper is structured as follows. Section 2 presents the literature that inspired this research. Section 3 introduces the context of the study with a brief reference to the database, which describes the operations of nearly 800 vessels at a multi-user dry bulk terminal in Europe. Section 4 brings the methodology, which is divided into three subsections. The first focuses on the discovery of associations or main rules of service, the second focuses on anomaly detection and the third explains the steps to evaluate the robustness of the method. In Section 5, the methodology is applied to the terminal in order to detect the main service rules and anomalous vessels. It also includes the results of the evaluation and a brief discussion of the implications for industry. Finally, conclusions and future work are presented in Section 6.

2 Literature review

2.1 Service in dry bulk terminals

There are several studies that have successfully analysed the special characteristics of dry bulk terminals (van Vianen et al., 2011), the arrival and service process (Hess et al., 2007) and (van Vianen et al., 2012), the required stockyard size for dry bulk terminals (van Vianen et al., 2014), and the design of the conveyor network (van Vianen et al., 2016). By applying the models and design parameters proposed in the aforementioned studies, it is possible to make decisions on how to optimise the service process and increase the efficiency of bulk terminals.

In broad terms, the service process includes the following steps: entering to a port > pilotage service > tugging > waiting > berthing > unloading. And each of these steps can significantly affect the overall dwell time of bulk carriers in the port. In this regard, there are many factors that can delay service in bulk ports. For example, Othman et al. (2019) have recently concluded after an analytical hierarchy process (AHP) method that 'foul weather and tide prediction' is the most significant sub-cause of delay in Malaysia.

Despite its importance, few researchers have taken into account the limitations of tides. This could be partly explained by the fact that tidal conditions are only important in those bulk ports where draught conditions depend on high tide conditions. To the best of our knowledge, only Wang (2018) and Zhang and Zheng (2021) have scheduled available berths for arriving vessels taking into account a multi-tidal planning horizon.

The service process in bulk terminals depends on many aspects, but perhaps the most important is the size of the vessel. Bulk carriers are classified according to the deadweight tonnage as very large bulk carriers (VLBC), Capesize class, Panamax class, Handysize class, or Handymax class, see ROM 2.0-11 (2011). The size of the ship is crucial not only from a technical point of view (it affects tugging, berthing, etc.), but also from an economic point of view. In this regard, Acik and Baser (2020) have recently analysed the interactions between the prices of three commodities (coal, wheat and steel) and freight rates as a function of vessel size.

Despite the previous studies there is still considerable vagueness concerning operational bottlenecks. According to VALE (2011), an operational bottleneck is a "failure or inability to attain full system capacity, due to internal limiting processes within the macro process". It can therefore be understood as an anomaly in the service process. But, what leads to the anomaly or port failure? And how the terminal operator can detect it?

Cutrim et al. (2013) focused on the first question and found that 77% of the stoppages in the ship loading process at the Port of Tubarão were related to operational causes, i.e., waits/stops and operational blockages. In addition, they found the following causes behind the operational blockages:

1 ship-loaders blocking (when it was not possible to operate two ship-loaders at the same time due its prior positioning or relative priority)

2 ship blocking (when there was no product route for loading due stacking priority, or when the route was being used to supply another berth or to perform product transferring among yard cells)

3 stockyard blocking (when it was not possible to recover the iron ore from the yard due to structural issues, equipment and routes provision).

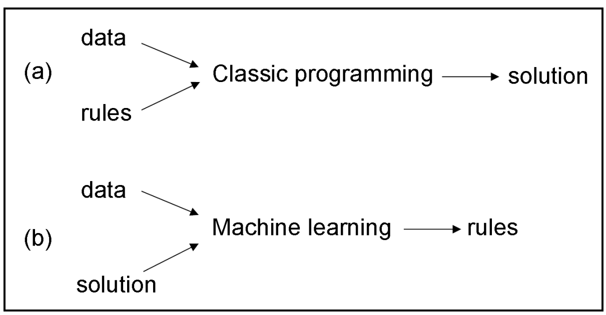
2.2 Berth allocation problem

The berth allocation problem (BAP), i.e. the allocation of berthing space to ships, is a popular problem in port literature. Although the problem was originally proposed to cope with berthing at container terminals, some authors have moved the BAP to dry bulk terminals. The constraints (such as inter-vessel and end-berth clearance distance) that have been adopted in many studies dealing with container terminals are complemented by constrains related to bulk piles. As examples, Barros et al. (2011) used a linear programming model based on the transportation problem to represent the BAP in bulk ports with stock level constraints, and Robenek et al. (2014) proposed a branch-and-price algorithm to address not only the BAP but also the yard assignment problem. The introduction of technical constraints brings the problem formulation closer to real-world conditions, although it increases the complexity of the problem.

In de León et al. (2017), a novel direction was proposed to address the BAP in bulk terminals. The problem was tackled using a machine learning-based system which provided a ranking of BAP-solver algorithms sorted by suitability. The computational study showed an increase in the quality of the provided solutions when the algorithm was selected according to the features of the instance. According to Heilig et al. (2020), this novel approach demonstrates that data mining and machine learning can not only be used to analyse operational data, but also to select appropriate planning methods and tools.

Another interesting aspect of the above-mentioned study is the alternative way that machine learning provides to address berthing problems. In the simplest berthing problem, there are two basic inputs: an objective function to be optimised (a data function representing the vessel's total service times for example) and a set of conditions (rules) that constraint the problem. The typical direction to solve the problem is through heuristic algorithms, linear programming, etc. In contrast, machine learning uses the known solutions of the problem as input to find rules that are consistent with the solution of the problem; see Figure 1. The deduced rules can then be used to better understand the behaviour of the berthing line. Machine learning is therefore a useful and under-explored direction for building models and extracting hidden knowledge from port databases.

Figure 1 (a) Classic programming vs. (b) machine learning



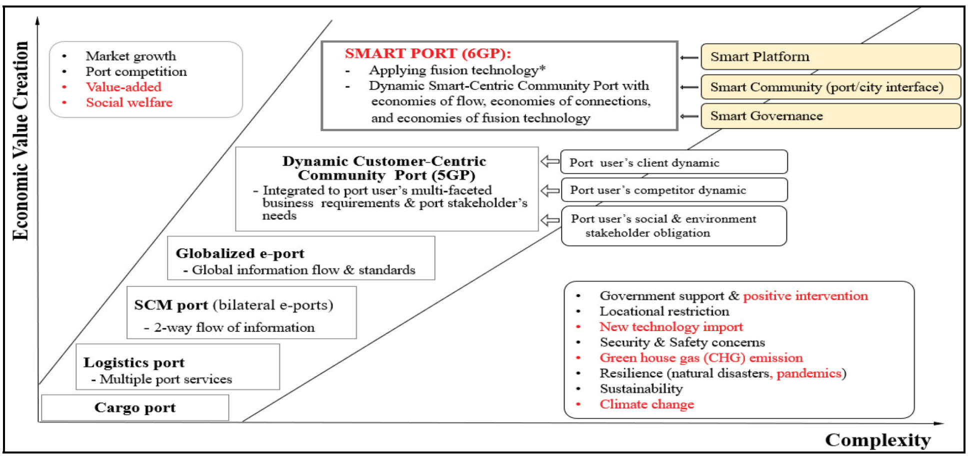
2.3 Machine learning

Machine learning is a scientific discipline in the field of Artificial Intelligence that creates systems that learn automatically. It could therefore be included in the select group of fusion technologies that will characterise the port of the future. According to Lee (2020a, 2020b) the application of fusion technologies is the key factor that will make the difference for smart ports (sixth generation ports). Although the use of these technologies is very complex, it also brings the highest value creation, see Figure 2.

In the context of dry bulk terminals, finding hidden associations in the data (rules) and detecting anomalous vessels are two crucial tasks. A familiar problem encountered in association mining exercises is that they often produce a large numbers of association rules, and this makes it difficult for users to identify those that are of interest. To address this problem and facilitate the extraction of useful knowledge from learned rules, Zhao et al. (2009) provided a systemic collection of the up-to-date techniques for reducing the numbers of association rules, after mining.

The problem of association rule mining in large databases can be defined as a set of binary attributes I (i_1, i_2, i_3, …, i_n) called items, which are connected via a set of transactions T (t_1, t_2, t_3, …, t_n). Each transaction has a unique identity and connects (contains) a specific subset of the items in I. The idea is to analyse specific rules that contain two items from I (X → Y). Agrawal et al. (1993) popularised the concept of association rules. In particular they were interested in those the rules that have minimum transactional support and minimum confidence at the same time. Support is an indication of how frequently the item set appears in the dataset, while confidence is an indication of how often the rule has been found to be true. These are the most common thresholds, but there are more thresholds in previous research, see, e.g., leverage level in Piatetsky-Shapiro and Frawley (1991) or the lift level in Brin et al. (1997).

Figure 2 Future port model: Sixth generation ports as Smart port (2020) (see online version for colours)



Notes: *BT, IT, NT, AI, blockchain, big data and soft computing. BT – bio-technology, IT – information technology, NT – nanotechnology, AI – artificial intelligence.

Source: Lee (2020b)

Given a fixed threshold, there are several learning algorithms that can be used to find associations. The most famous is the k2 algorithm originally introduced in Cooper and Herskovits (1992, 1993). This is a score-based algorithm in Bayesian network, which has been successfully used in many scientific applications in the last two decades, see, e.g., Webb (2000), Webb and Zhang (2005) and Sammut and Webb (2011). Specifically, it recovers the underlying distribution in the form of a decision analysis graph efficiently. In the field of seaports, k2 algorithm has been used together with other machine learning techniques to manage uncertainty in container terminals (Parolas, 2016) as well as in ferry terminals (Ansorena, 2019, 2020).

The detection of anomalies within a big dataset is based on similar principles. Thus, it uses machine learning techniques to detect rare items, events or observations, which raise suspicions by differing significantly from the majority of the data (Zimek and Schubert, 2018). A well known technique is based on the isolation forest; see Liu et al. (2008, 2016). The basic idea is that anomalous instances in a large dataset are more susceptible to be isolated than normal instances when a decision tree approach is used. Each observation in the dataset has its own anomaly score. The detection algorithm then uses the scores to reveal anomalous data points.

In sum, we can conclude that previous research has tended to focus on container terminals rather than bulk terminals and many studies did not make an efficient use of big datasets. There are few databases available in ports literature and to the best of author's knowledge, databases describing operations at bulk terminals are difficult to find, or at least, they are not free or publicly available. However, the search for new measures and techniques to model datasets and extract information from them is today an active research trend in many fields. Despite this active interest, no one to the best of our knowledge has brought these techniques to bulk cargo terminals. Consequently, we bring two of these techniques (association rule learning and anomaly detection) to bridge this gap.

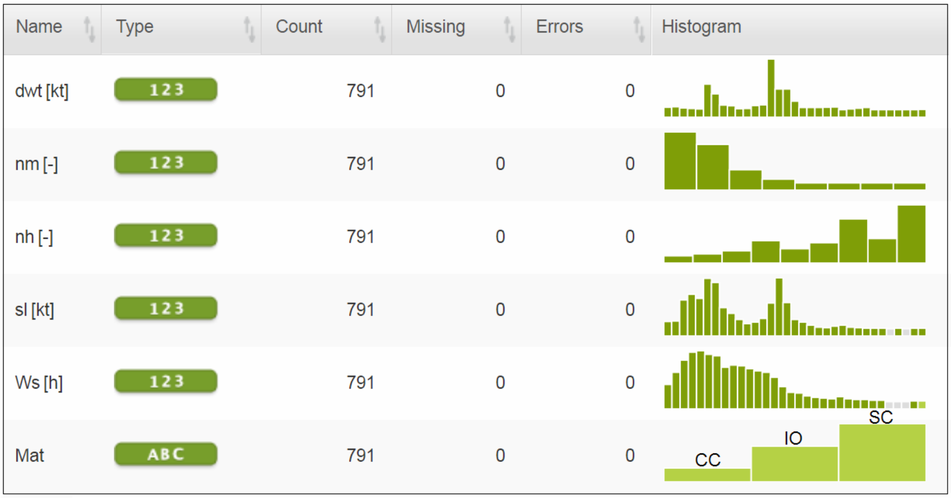
3 Study context: database

Van Vianen (2015) database is used with the aim of gaining a better comprehension of operations in dry bulk cargo terminals. The database describes the service at a multi-user import terminal. The terminal operator did not want to be named explicitly due to commercial interests and the author replaced its name by T2. This case study was selected because it includes a practical and complete database describing the operation of 791 bulk carriers, see Figure 3. The database has no missing values or errors and the descriptive fields are:

- Ship's deadweight [kt]: the typical arrival has a DWT between 161.82 and 173.73 kt (there is a peak with 209 instances in the histogram).
- Number of materials carried in the ship [-]: the typical vessel carries only one material.
- Number of unloaded holds [-]: there is a majority of vessels with nine cargo holds.
- Shipload [kt]: there are two peaks in the histogram. The first includes 98 instances with a shipload between 61.24 and 72.09 kt, while the second includes 99 instances with a shipload between 158.88 and 169.73 kt.
- Service time [h]: the histogram of the registered service time has a long tail and a peak with 71 instances between 31.82 and 37.27 h.
- Materials: unloaded materials are iron ore (IO), coking coal (CC) and steam coal (SC), which is the most common material.

As explained in the previous section, there are other variables that can affect the dwell time, e.g., the performance of unloading equipment, the existence of cranes on board, personnel, weather and tidal conditions, etc. For the sake of simplicity we have not considered all these variables, although they are also crucial to understand the operation of the terminal.

Figure 3 Dataset overview (see online version for colours)



Name	Type	Count	Missing	Errors	Histogram
dwt [kt]	123	791	0	0	
nm [-]	123	791	0	0	
nh [-]	123	791	0	0	
sl [kt]	123	791	0	0	
Ws [h]	123	791	0	0	
Mat	ABC	791	0	0	CC IO SC

Source: Data from van Vianen (2015, pp.177–183, Appendix B – Bulk ships)

4 Method and concepts

4.1 Association discovery

In the first step, a well-known unsupervised machine learning technique is used to find relevant associations in datasets. The association discovery technique is based on the filtered-top-k2 algorithm implemented in BigML, which is a machine learning platform that provides the software for manipulating and analysing data. The k2 algorithm pseudo-code for learning Bayes net structures as introduced in Cooper and Herskovits (1992) is presented in Figure 4.

The basic inputs are a set of n nodes, an upper bound u on the number of parents a node may have, and a database D containing m cases. The algorithm heuristically searches for the most probable belief-network structure given the database of cases. It uses equation (1).

$$f(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} \alpha_{ijk}! \tag{1}$$

where

π_i set of parents of node x_i

$$q_i = |\phi_i|$$

Φ_1 list of all possible instantiations of the parents of x_i in database D. That is, if $p_1, \ldots, p_s$ are the parents of x_i then Φ_i is the Cartesian product $\{v_1^{p_1}, \ldots, v_{r_{p1}}^{p_1}\} \times \ldots \times \{v_1^{p_s}, \ldots, v_{r_{ps}}^{p_s}\}$ of all the possible values of attributes p_1 through p_s.

$$r_i = |V_i|$$

V_i list of all possible values of the attribute x_i

α_{ijk} number of cases (i.e., instances) in D in which the attribute x_i is instantiated with its k^{th} value, and the parents of x_i in π_i are instantiated with the j^{th} instantiation in Φ_i.

$$N_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$$

That is, the number of instances in the database in which the parents of x_i in π_i are instantiated with the j^{th} instantiation in Φ_i.

The informal intuition here is that $f(i, \pi_i)$ is the probability of the database D given that the parents of x_i are π_i.

Figure 4 How the k2 algorithm works

```
procedure K2;
{Input: A set of n nodes, an ordering on the nodes, an upper bound u on the
        number of parents a node may have, and a database D containing m cases.}
{Output: For each node, a printout of the parents of the node.}
for i:= 1 to n do
    π_i := ∅;
    P_old := f(i, π_i); {This function is computed using Equation 1.}
    OKToProceed := true;
    While OKToProceed and |π_i| < u do
        let z be the node in Pred(x_i) - π_i that maximizes f(i, π_i ∪ {z});
        P_new := f(i, π_i ∪ {z});
        if P_new > P_old then
            P_old := P_new;
            π_i := π_i ∪ {z};
        else OKToProceed := false;
    end {while};
    write('Node: ', x_i, ' Parent of x_i: ',π_i);
end {for};
end {K2};
```

As a result, we obtain the parents of each node in the database. In other words, the application of k2 algorithm shows the learned topology of the network. Thus the cases lead to a set of rules that connect *n* nodes, but what are the most interesting rules? The goal now is to find them through statistics. To that end, we use the four statistical parameters: support, confidence, leverage and lift. In the following paragraphs we give the definition and precise formulae to compute them in a specific rule (A → C), which means that if the antecedent is the node 'A' then 'C' appears as the consequence.

- Support is the proportion of instances (node) in the dataset that contain a specific itemset (A and C). In other words, the portion of instances that contain the rule's antecedent and rule's consequent together, divided by the total number of instances (nodes: n) in the dataset. It gives a measure of the prevalence of the rule in the dataset.

$$\text{Support}(A \Rightarrow C) = \text{Support}(A \cup C) = P(A \cap C) \tag{2}$$

- Confidence (or strength) is the percentage of instances that contain the consequent and antecedent together over the number of instances that only contain the antecedent. Confidence is computed using the support of the association rule over the coverage of the antecedent (support of A). It gives an estimate of the probability that the consequent will occur in case the antecedent occurs.

$$\text{Confidence}(A \Rightarrow C) = \frac{\text{Support}(A \Rightarrow C)}{\text{Support}(A)} = \frac{P(A \cap C)}{P(A)} \tag{3}$$

- Leverage is the difference between the probability of the rule and the expected probability if the items were statistically independent. Leverage ranges between [–1, 1]. A leverage of 0 suggests there is no association between the items. Higher positive leverage values suggest a stronger positive association between the antecedent and consequent. Negative values for leverage suggest a negative relationship.

$$\text{Leverage}(A \Rightarrow C) = \text{Support}(A \Rightarrow C) - [\text{Support}(A) \times \text{Support}(C)] \tag{4}$$

$$\text{Leverage}(A \Rightarrow C) = P(A \cap C) - [P(A) \times P(C)] \tag{5}$$

- Lift is a measure that indicates how many times more often antecedent and consequent occur together than expected if they were statistically independent, e.g., a lift of 5 for the following rule (A → C) means that A makes it 5 times more likely C. Lift is always a real positive number, 1 suggests there is no association between the items, while a lift between 0 and 1 indicates a negative correlation. Higher values suggest stronger relationships between the items.

$$\text{Lift}(A \Rightarrow C) = \frac{\text{Support}(A \Rightarrow C)}{\text{Support}(A) \times \text{Support}(C)} = \frac{P(A \cap C)}{P(A) \times P(C)} \tag{6}$$

As a brief example, let consider a theoretical dataset with n = 100 nodes connected by rules. Now suppose that the node A appears 19 times, the node C appears 10 times and both nodes appear 8 times altogether following the rule (A) → (C). Therefore, P(A) = 19/100 = 0.19; P(C) = 10/100 = 0.1. Following the previous formulae we obtain:

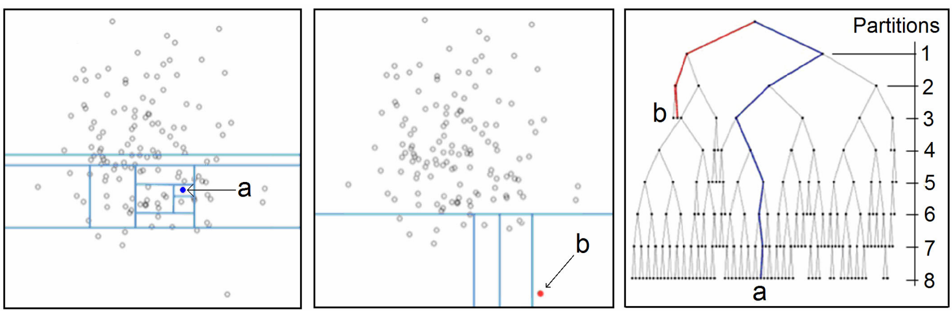
- Support (A → C) = P(A ∩ C) = 8/100 = 0.08.
- Confidence (A → C) = P(A ∩ C)/P(A) = 0.08/0.19 = 0.42
- Leverage (A → C) = P(A ∩ C) – P(A) ∗ P(C) = 0.08 – (0.19 ∗ 0.1) = 0.061
- Lift (A → C) = P(A ∩ C)/P(A) ∗ P(C) = 0.08/(0.19 ∗ 0.1) = 4.21.

4.2 Anomaly detection

In the second step, we use an optimised implementation of the isolation forest algorithm proposed in Liu et al. (2016) to detect anomalous instances in the dataset. This unsupervised machine learning technique can efficiently deal with high-dimensional datasets. As explained in previous sections, the basic idea is that rare instances are more susceptible to be isolated than normal instances when a decision tree approach is used. Therefore, we elaborate an ensemble of binary (isolation) trees that deliberately over-fits each single tree to isolate each instance (ship call) from the rest. Each decision tree is developed by selecting a feature (monitored variable) and a random split. Thus the space is recursively partitioned randomly until the instances are isolated.

Figure 5 shows an example to better understand the idea. The blue data point (a) in Figure 5(a) is a nominal point that took eight partitions to be isolated, while the red one (b) in Figure 5(b) is an anomalous point that took only three partitions, see Figure 5(c). As a result, b has a higher anomaly score than a. The graphic representation illustrates how anomalous instances can be isolated easier than normal instances. In other words, the closer to the tree root the more anomalous the instance.

Figure 5 (a) A normal data point (b) An anomalous data point (c) Partitions (see online version for colours)



(a) (b) (c)

Once all instances (ship calls) have been isolated through the Isolation Forest algorithm, we calculate an anomaly score for each ship by averaging the number of splits needed to isolate it. As explained above, fewer partitions will result in a higher score. These averages are then normalised to obtain a final score that measures how anomalous a ship call is. The isolation forest anomaly score is calculated through:

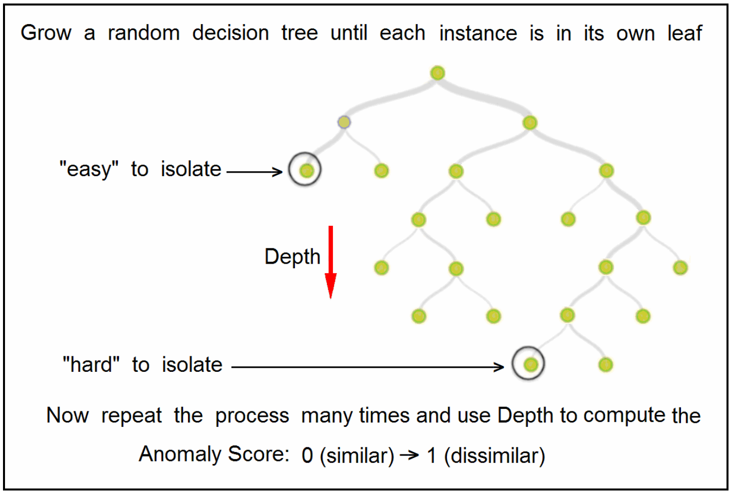
$$s(x, n) = 2^{-\cdot\frac{E(h(x))}{c(n)}} \tag{7}$$

where $h(x)$ is the path length of observation x, $c(n)$ is the average path length of unsuccessful search in a binary search tree and n is the number of external nodes.

As a result, we rank data by dissimilarity. Each ship call has its own an anomaly score between 0 and 1 (Figure 6). A score close to 1 indicates a rare observation or outlier, while a score much lower than 0.5 indicates a normal observation. If all scores are below or close to 0.5 then the sample has no clear anomalies. We are now able to discern between suspicious and common patterns in the database.

In summary, we propose a machine learning-based method to find associations and detect rare observations in a dataset. The originality lies in the fact that we apply these techniques to the analysis of a dry bulk terminal. It should be noted that so far there is no application of this method on this type of terminal in the literature.

Figure 6 Isolation forest: concept (see online version for colours)



4.3 Robustness of the method

The workflow in Figure 7 is used to assess the robustness of the method. First, an ‘anomaly score’ feature is added to the initial dataset (with the score obtained in the second stage). Then, through feature engineering, the ‘anomaly score’ feature is transformed into a category feature, namely ‘anomaly category’, which indicates whether the vessel has an anomaly score higher than 0.5, i.e. whether the service is suspicious of being anomalous. We lower the bar (0.5 instead of 0.6) to ensure that the anomalous vessels are correctly identified in their class.

The new dataset is then split into two disjoint datasets with 80% and 20% of data. A basic tree model is built with the first subset of data (80% of the data). First, the tree model is trained to predict whether the vessel is anomalous, i.e., if it has an anomaly score higher than 0.5. The model is then tested with the second subset of data, i.e. data that the model has not seen before. To asses the performance of the tree model, the ratios of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) are measured. Finally, several evaluation parameters are calculated including accuracy, precision and recall.

- Accuracy (the fraction of vessels that were rightly categorised) is given by equation (8).

$$\text{Accuracy} = (\text{TN} + \text{TP})/(\text{TP} + \text{FP} + \text{FN} + \text{TN}) \tag{8}$$

- Precision (the count of vessels either normal or rare that were rightly categorised) is computed using equation (9).

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{9}$$

- Recall (also referred as the true positive rate or sensitivity) is a measure of how well a test can identify true positives. It is computed using equation (10).

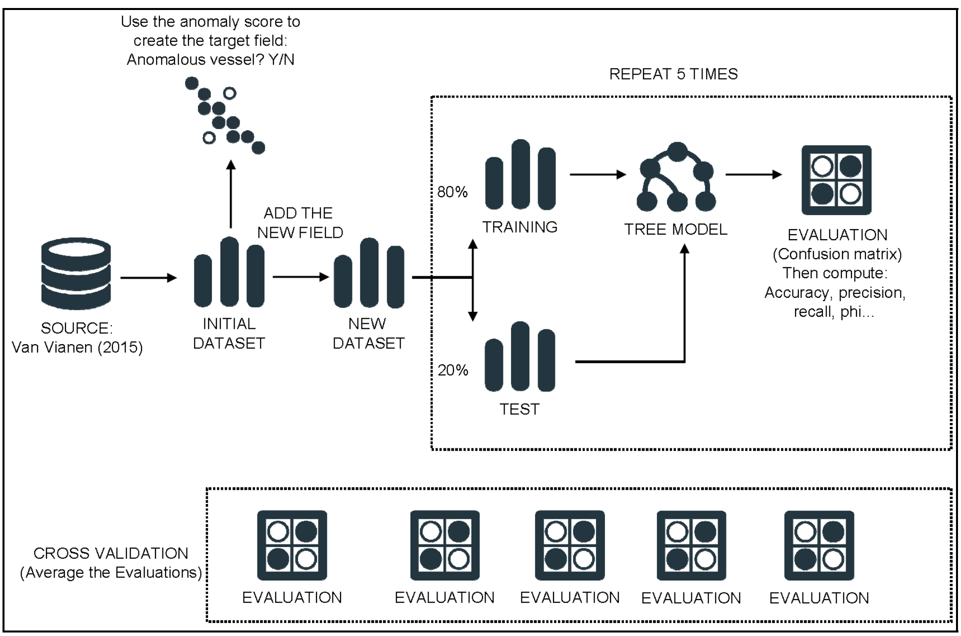
$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{10}$$

A script in Whizzml code is used to asses the robustness of the model. The script performs a 5-fold cross validation of the model built from the dataset. This technique is widely used in artificial intelligence projects to validate the generated models. Thus, the only input required for the script to run is the dataset used to train and test the models in the cross validation. The algorithm:

- divides the dataset in five parts
- holds out the data in one of the parts and builds a model with the rest of data
- evaluates the model with the hold out data.

The second and third steps are repeated with each of the five splits, so that five evaluations are generated. Finally, the evaluation metrics are averaged to obtain the cross-validation metrics.

Figure 7 Workflow to asses the robustness of the method (see online version for colours)



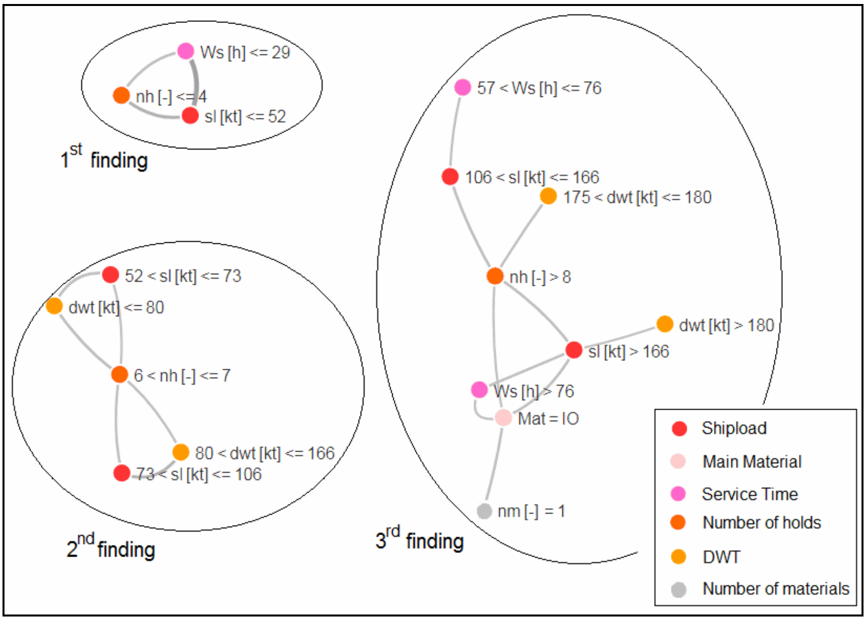
5 Results and discussion

5.1 Association discovery

The first stage of the study aims to classify the terminal service according to the measured variables, i.e., size of the ship, number of cargo holds, number and type of materials, shipload and service time. As a first result, a synthesised graph showing the main associations between the variables is obtained by means of the k2 algorithm, see

Figure 8. The parameters of coverage, support, confidence, leverage and lift of the association rules are presented in the Appendix (Table A1).

Figure 8 Main associations deduced from the database (leverage > 5%) (see online version for colours)



The links between each pair of nodes are representing an association rule with a leverage level over 5%. This level means that antecedent and consequent occur together 5% more often than if they were statistically independent. As explained in the previous section a higher leverage value means a stronger positive association between the antecedent and consequent. Therefore, the following findings can be deduced from Figure 8.

- Firstly, a ship with less than four cargo holds is linked to a shipload of less than 52,000 t (Handymax class) and an average operating time of less than 29 hours.
- Secondly, a ship with 6–7 cargo holds is linked to two different classes. On one hand, we see a connection with the Panamax class, i.e., vessels with a deadweight tonnage of less than 80,000 t and shiploads between 52,000 and 73,000 t. On the other hand, we see another connection with the Capesize class, i.e., vessels with a deadweight tonnage between 80,000 t and 166,000 t and shiploads between 73,000 t and 106,000 t.
- Thirdly, a ship with more than eight cargo holds is linked to a deadweight tonnage between 175,000 and than 180,000 t (the limit of Capesize class), a shipload between 106,000 and 166,000 t, and a service time between 57 and 76 hours. Additionally, ships with more than 8 cargo holds are also linked to a deadweight tonnage over 180,000 t (very large bulk carriers – VLBC) and a shipload over 166,000 t. In this case, the average service time is now over 76 hours and the material is iron ore.

If we focus on the parameters of service and the type of ship, we obtain the links in Table 1. Here we can see that there are some relations that are not detected yet. In this sense, we can deduce more association rules through a reduction of the leverage parameter. However, it is important to note that these new associations will have less statistical significance. It is also important to note that the association rules are obtained from a limited number of bulk carriers in a single terminal. Consequently these results should be treated with caution.

Table 1 Strongest links deduced from the analysis (leverage > 5%)

Bulk carriers	*DWT (kt)**	*DWT (kt)*	*Shipload (kt)*	*Throughput (t/h)*	*Ws (h)*
VLBC	200	>180	>166	~2,200	>76
Capesize (max)	180	175-180	106-166	1,400-2,900	57-76
Capesize	100-150	80-166	73-106	Undiscovered	Undiscovered
Panamax	70	80	52 - 73	Undiscovered	Undiscovered
Handymax	50	52	Undiscovered	~1,800	<29
Handysize	10-30	Undiscovered	Undiscovered	Undiscovered	Undiscovered

Note: *According to ROM 2.0-11 (2011).

Source: Own elaboration

A more detailed discussion about these association rules goes beyond the aim of the paper, although it might be interesting from the perspective of the terminal operator.

5.2 Anomaly detection

The associations found in the previous section are important because they show how the terminal works. However, as explained in the methodology section, this first step does not detect rare vessels, as it cannot be deduced whether a specific vessel experienced an anomalous service. For this reason, we apply the anomaly detection technique to find rare instances in the dataset. The resulting *anomaly* scores together with the contribution of each variable to the anomaly score can be found in the Table A2 – Appendix. With this new information, we have obtained a more complete picture of terminal's performance by describing not only the association rules, but also pointing out 'noise' that did not follow the general pattern of service.

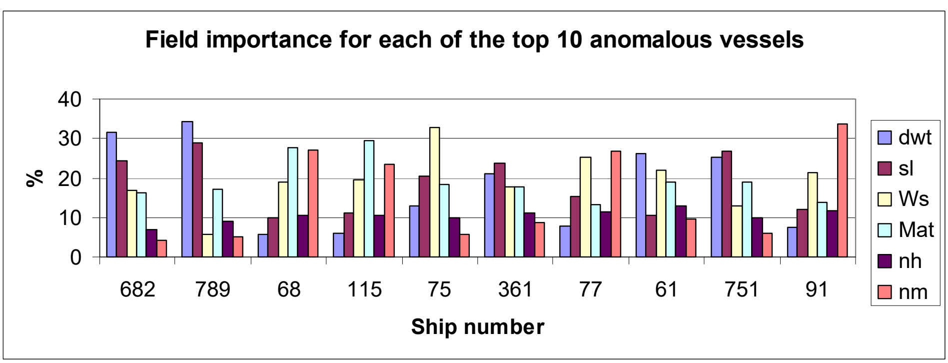
A short list containing the top anomalous elements found in the database ranked by their anomaly score is presented in Table 2. As explained in the methodology section, the anomaly score ranges from 0 (0% minimum anomaly score) to 1 (100% maximum anomaly score). A higher anomaly score indicates a more anomalous operation. In general, a score over 0.5 is a signal that the observation could be an uncommon observation and a score of 0.6 or higher is a solid basis for a given observation to be considered anomalous. Following this criterion we have detected 10 bulk carriers that are close to the limit and only 8 with an anomaly score over 60%, see Table 2. Taking into consideration that the database includes 791 bulk carriers, we can affirm that the terminal provides a predictable and reliable service.

Figure 9 shows the contribution of each variable (field) to the anomaly score. The importance of each field in the histogram is calculated through the evaluation of the Isolation Forest. We normalise the sums of the instances (per field) partitioned by each split and this yields a percentage (ranging from 0% to 100%). The percentage is the relative contribution of each field to the anomaly score. For instance, we find that the Ws field has a larger contribution to the anomaly score in the 75, 77, 61 and 91 ship calls. These are suspect ships in terms of service. Due to this, the terminal manager should identify the special causes that led to such rare services. In contrast, ship number 789 is not as suspicious as the previous ones, since the Ws field has a very low contribution to the anomaly score. In other words, the service time in this case more or less follows the general pattern of service in the terminal.

Table 2 Short list of anomalous services

Ship nr.	*Dwt [kt]*	*sl [kt]*	*Ws [h]*	*Mat*	*nh [-]*	*nm [-]*	*Anomaly score*
682	388	363	138	IO	8	1	67.44%
789	400	365	74	IO	6	1	66.13%
68	172	160	129	CC	9	8	65.09%
115	179	171	123	CC	9	7	62.72%
75	249	241	190	IO	9	1	62.49%
361	297	286	116	IO	5	4	61.91%
77	181	191	149	SC	9	7	60.37%
61	311	166	136	IO	5	4	60.17%
751	317	310	103	IO	6	3	59.83%
91	171	163	123	SC	8	8	59.77%

Figure 9 Field-importance for each rare case (see online version for colours)



In short, we have been able to uncover key insights with the help of association discovery, and the anomaly detection stage has helped us to isolate anomalous vessels. It is now clear that we have turned a complex problem into something that can be visualised and understood, but is this methodology robust enough, and what are the implications for industry? These questions are discussed in the next subsection.

5.3 *Robustness of the method and implications*

As explained in the methodology, we apply the cross-validation technique to evaluate the results and ensure that they are independent of the partition separating training and test data. The idea is to repeat the process five times and obtain an average of the evaluations.

As we can see in Table 3, the learning model shows significant robustness when dealing with new data (i.e., data unknown to the model). Detailed results for each evaluation are presented in the Appendix (Tables A3 and A4). The results show that the method is sufficiently robust, so we have demonstrated that machine learning can be used to deduce not only the main association rules, but also to detect, or anticipate, anomalies in dry bulk terminals.

Table 3 Cross-validation

Target class	*All classes*	*0 class*	*1 class (anomalous vessels)*
Average accuracy	90.52%	90.52%	90.52%
Average precision	80.45%	94.17%	66.73%
Average recall	79.23%	94.86%	63.59%
Average F-measure	0.8	0.95	0.65
Average phi	0.6	0.6	0.6

In terms of implications, this methodology, based on a machine learning system, can be used to increase transparency and assist the terminal operator and its customers in their agreements. It can also be extended to other terminals (and processes) following the same customised workflow that was embodied in the BigML platform for traceability, easy retraining and automation. The port of the future (sixth generation port) will have to move from paperless processes to smart processes and machine learning is a good tool in this move. We strongly believe that machine learning (artificial intelligence) and other fusion technologies that are increasingly being applied in container terminals should also reach dry bulk terminals, although there is still a long way to go.

This study is only a first step and should therefore be treated with caution. We are aware that it has limitations. Perhaps the most important one is the number of variables included in the study, which reflects the difficulty of collecting data. Indeed, we have used hundreds of records, but these records were collected on a limited number of variables. As mentioned in the previous sections, more variables are needed to determine the special causes behind some long dwell times, for example, were adverse weather conditions the cause of a delay? Although this future work could easily be carried out, access to new records describing weather/tidal conditions, cranes on board, work force, yard throughput, conveyor breakdowns, etc. is not an easy task. Many terminal operators are reluctant to share this information with their competitors, (or perhaps there databases are not ready yet).

Today, transparency is essential to bring a better and more predictable service. For example, in the event of a sudden breakdown of the conveyor belt, the system must be repaired, as soon as possible, and the remaining material must take an alternative route. This has an impact on the unloading process and could explain some of the rare cases observed in Figure 9. In general, terminal operators have little interest in providing this information.

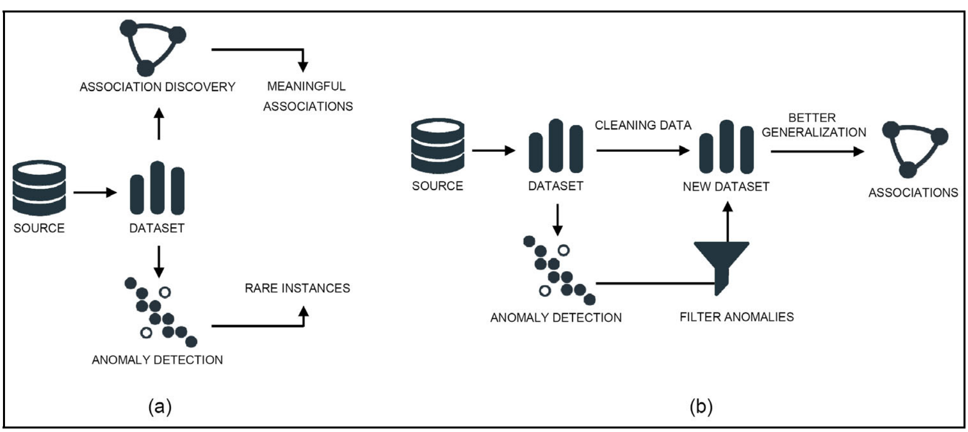
Understandably, this methodology cannot work alone. It is also necessary to implement a data sharing system. Both systems working together can help the dry cargo industry move forward. Better visibility and predictability of ‘terminal performance’ can facilitate stakeholder strategies. This may be the case for slow steaming practises in the maritime leg; or more efficient use of rail corridors in inland transport. Machine learning and data sharing systems should also be a lever to attract more customers. Terminal operators would therefore also benefit from this.

6 Conclusions and future work

This study proves that bulk terminals can be analysed through machine learning techniques. Following a basic principle, "learn from the past to find the answer", we have determined the vessels that received a low level of service. The strength of the methodology lies in the twofold approach. On the one hand, we have obtained a comprehensive picture that describes the performance of the terminal (association rules). On the other hand, we have determined the anomalous vessels (those with anomaly scores over 60%). Some of them did not receive a typical service and this could be (in same cases) directly attributable to the operator. We believe that the proposed methodology can be used as a tool to increase transparency and assist the terminal operator and ship agents in future agreements.

The study has gone some way towards our understanding of a dry bulk terminal in north-Europe. Unfortunately, it was not possible to investigate the causes that led to rare observations because the database had a limited number of variables. However, we believe that the methodology can easily be applied to a larger database of this terminal. If it were not possible to add new variables to van Vianen (2015) database, we would invite other authors to try other leaning algorithms (for association search and anomaly detection) and compare their results with ours.

Figure 10 (a) Basic workflow vs. (b) alternative workflow



In regard to the methodology, a basic workflow was followed in this study [Figure 10(a)] in which the two steps are independent, but an alternative workflow should be used in the future [Figure 10(b)]. This second workflow removes the anomalies before the association discovery stage and should therefore improve the generalisation of the model. We hope that further testing will confirm this hypothesis.

Finally, a third via could be the application of the same method to other types of terminals with other layouts and operational procedures (e.g., export terminals focused on ship loading).

Acknowledgements

The open access fee of this work is supported by China Merchants Energy Shipping.

References

Acik, A. and Baser, S.O. (2020) 'Asymmetric causality from commodity prices to shipping markets: an empirical research on ISTFIX region', *World Review of Intermodal Transportation Research*, Vol. 9, No. 1, pp.47–62.

Agrawal, R., Imielinski, T. and Swami, A. (1993) 'Mining associations between sets of items in large databases', *ACM SIGMOD 1993 Conference: Proceedings of the ACM SIGMOD. Int. Conference on Management of Data*, Washington DC, USA, pp.207–216.

Ansorena, I.L. (2019) 'Forecasting of traffic flows at ferry terminals. A hybrid model', *Int. J. Agile Systems and Management*, Vol. 12, No. 1, pp.27–43.

Ansorena, I.L. (2020) 'Managing uncertainty in ferry terminals: a machine learning approach', *Int. J. Business Information Systems*, Vol. 33, No. 2, pp.285–297.

Barros, V.H., Costa, T.S., Oliveira, A.C.M. and Lorena, L.A.N. (2011) 'Model and heuristic for berth allocation in tidal bulk ports with stock level constraints', *Computers & Industrial Engineering*, Vol. 60, No. 4, pp.606–613.

Brin, S., Motwani, R., Ullman, J. D. and Tsur, S. (1997) 'Dynamic itemset counting and implication rules for market basket data', *ACM SIGMOD 1997 Conference: Proceedings of the ACM SIGMOD. Int. Conference on Management of Data*, Tucson Arizona, USA, pp.265–276.

Cooper, G. and Herskovits, E. (1992) 'A Bayesian method for the induction of probabilistic networks from data', *Machine Learning*, Vol. 1992, No. 9, pp.330–347.

Cooper, G. and Herskovits, E. (1993) *A Bayesian Method for the Induction of Probabilistic Networks from Data*, Technical Report KSL-91-02, Knowledge Systems Laboratory, Medical Computer Science, Stanford University School of Medicine, Stanford, CA 94305-5479, Updated November 1993 [online] http://smiweb.stanford.edu/pubs/SMI Abstracts/SMI-91-0355.html.

Cutrim, S.S., Robles, L.T., Merigueti, B.A. and Assis, K.M. (2013) 'Port efficiency analysis: port of Tubarão's Pier I case study', Paper presented at the *IAME 2013 Conference*, 3–5 July, Marseille, France.

de León, A.D., Lalla-Ruiz, E., Melián-Batista, B. and Moreno-Vega, J.M. (2017) 'A machine learning-based system for berth scheduling at bulk terminals', *Expert Systems with Applications*, Vol. 87, No. 11, pp.170–182.

Heilig, L., Stahlbock, R. and Voß, S. (2020) 'From digitalization to data-driven decision making in container terminals', *Handbook of Terminal Planning*, pp.125–154, Springer, Cham.

Hess, M., Kos, S. and Hess, S. (2007) 'Queuing system in optimization function of the port's bulk unloading terminal', *Promet – Traffic & Transportation*, Vol. 19, No. 2, pp.61–70.

IACS (2018) *Bulk Carriers. Guidance and Information on Bulk Cargo Loading and Discharging to Reduce the Likelihood of Over-stressing the Hull Structure*, IACS Report, IACS Ltd, 36 Broadway, London, UK

Lee, P.T-W. (2020a) *The 6th Generation Ports and Smart Ports*, Working Paper No. 2020-01, Maritime Logistics and Free Trade Islands Research Center, Ocean College, Zhejiang University, Zhoushan, China, 20pp.

Lee, P.T-W. (2020b) 'Policy guidelines and action plans for sustainable and resilient port development in Asia and the Pacific', *UNESCAP Capacity Building Workshop on 'Sustainable and Resilient Port Development to Support Sustainable Maritime Connectivity in Asia and the Pacific' (Virtual)*, 8 December, Bangkok, Thailand.

Liu, F.T., Ting, K.M. and Zhou, Z.H. (2008) 'Isolation forest', *Proceedings of the Eighth IEEE. International Conference on Data Mining (ICDM)*, Pisa, Italy, pp.413–422.

Liu, F.T., Ting, K.M. and Zhou, Z.H. (2016) *Isolation Forest Algorithm* [online] https://feitonyliu.files.wordpress.com/2009/07/liu-iforest.pdf (accessed 1 February 2019).

Othman, M.K., Sanusi, I.A., Arof, A.M., and Ismail, A. (2019) 'Evaluation of delay factors on dry bulk cargo operation in Malaysia: a case study of Kemaman Port', *The Asian Journal of Shipping and Logistics*, Vol. 35, No. 3, pp.127–137.

Parolas, I. (2016) *ETA Prediction for Containerships at the Port of Rotterdam Using Machine Learning Techniques*, Master thesis, Delft University of Technology and TNO.

Piatetsky-Shapiro, G. and Frawley, W. (1991) 'Discovery, analysis, and presentation of strong rules', in Piatetsky-Shapiro, G. and Frawley, W. (Eds.): *Knowledge Discovery in Databases*, The MIT Press, Cambridge, MA, pp.229–248.

Robenek, T., Umang, N., Bierlaire, M., and Ropke, S. (2014) 'A branch-and-price algorithm to solve the integrated berth allocation and yard assignment problem in bulk ports', *European Journal of Operational Research*, Vol. 235, No. 2, pp.399-411.

ROM 2.0-11 (2011) *Recomendaciones para el proyecto y ejecución en Obras de Atraque y Amarre*, Ente Público Puertos del Estado, Madrid.

Sammut, C. and Webb, G.I. (2011) *Encyclopedia of Machine Learning*, Springer Science & Business Media, Berlin.

UNCTAD (2019) *Review of Maritime Transport 2019*, Technical Report, United Nations Publication, New York, USA.

VALE (2011) *PDO Plano Diretor Operacional do Terminal de Minério de Tubarão 2012 a 2016*, Departamento de Operação de Portos e Rebocadores, VALE. Vitória, Brazil.

van Vianen, T.A. (2015) *Simulation-integrated Design of Dry Bulk Terminals*, TRAIL Thesis Series T2015/1, . Appendix B – Bulk ships, pp.177–183, the Netherlands TRAIL Research School, Delft University of Technology.

van Vianen, T.A., Ottjes, J.A. and Lodewijks, G. (2011) 'Dry bulk terminal characteristics', Paper presented at the *Bulk Solids Handling Conference*, 6–8 April 2011. Mumbai, India.

van Vianen, T.A., Ottjes, J.A. and Lodewijks, G. (2012) 'Modeling the arrival process at dry bulk terminals', Paper presented at the *Bulk Ports, Terminals and Logistics Conference*, 20–22 May, Amsterdam, The Netherlands.

van Vianen, T.A., Ottjes, J.A. and Lodewijks, G. (2014) 'Simulation based determination of the required stockyard size for dry bulk terminals', *Simulation Modelling Practice and Theory*, Vol. 42, No. 3, pp.119–128.

van Vianen, T.A., Ottjes, J.A. and Lodewijks, G. (2016) 'Belt conveyor network design using simulation', *Journal of Simulation*, Vol. 10, No. 3, pp.157–165.

Wang, R. (2018) *Berth Scheduling at Seaports: Meta-Heuristics and Simulation*, Faculty of Engineering and Technology, Liverpool John Moores University and the European EC-China Research Network on Integrated Container Supply Chains (ENRICH) project.

Webb, G.I. (2000) 'Efficient search for association rules', *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston, Massachusetts, USA, pp.99–107.

Webb, G.I. and Zhang, S. (2005) 'K-optimal rule discovery', *Data Mining and Knowledge Discovery*, Vol. 10, No. 1, pp.39–79.

Zhang, B. and Zheng, Z. (2021) 'Model and algorithm for vessel scheduling optimisation through the compound channel with the consideration of tide height', *Int. J. Shipping and Transport Logistics*, Vol. 13, Nos. 3–4, pp.445–461.

Zhao, Y., Zhang, C. and Cao, L. (2009) *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*, Information Science Reference (an imprint of IGI Global), Hershey, New York.

Zimek, A. and Schubert, E. (2018) 'Outlier detection', in Liu, L. and Özsu, M.T. (Eds.): *Encyclopedia of Database Systems*, 2nd ed., pp.1–5, Springer-Verlag New York.

Appendix

The experiment data file is in the Mendeley repository (doi: 10.17632/bm5jdh3jdk.1).

Table A1 Main association rules deduced from the initial dataset (ordered by leverage)

Antecedent	*Consequent*	*Coverage*	*Support*	*Confidence*	*Leverage*	*Lift*
sl [kt] ≤ 52	Ws [h] ≤ 29	20.48%	14.67%	71.61%	10.60%	3.61
nh [-] ≤ 4	sl [kt] ≤ 52	21.75%	12.01%	55.23%	7.56%	2.70
73 < sl [kt] ≤ 106	80 < dwt [kt] ≤ 166	20.23%	10.49%	51.88%	6.86%	2.89
6 < nh [-] ≤ 7	dwt [kt] ≤ 80	21.37%	10.75%	50.30%	6.51%	2.53
dwt [kt] ≤ 80	52 < sl [kt] ≤ 73	19.85%	10.11%	50.96%	6.35%	2.69
sl [kt] > 166	Mat = IO	17.70%	12.14%	68.57%	6.27%	2.07
106 < sl [kt] ≤ 166	nh [-] > 8	22.63%	12.64%	55.87%	6.26%	1.98
nh [-] ≤ 4	Ws [h] ≤ 29	21.75%	10.49%	48.26%	6.18%	2.43
sl [kt] > 166	nh [-] > 8	17.70%	11.13%	62.86%	6.14%	2.23
Ws [h] > 76	sl [kt] > 166	19.60%	9.48%	48.39%	6.01%	2.73
Mat = IO	nm [-] ≤ 1	33.12%	19.47%	58.78%	5.99%	1.44
6 < nh [-] ≤ 7	73 < sl [kt] ≤ 106	21.37%	9.99%	46.75%	5.67%	2.31
6 < nh [-] ≤ 7 and 80 < dwt [kt] ≤ 166	73 < sl [kt] ≤ 106	8.47%	7.33%	86.57%	5.62%	4.28
nh [-] > 8	175 < dwt [kt] ≤ 180	28.19%	13.65%	48.43%	5.60%	1.70
6 < nh [-] ≤ 7 and 73 < sl [kt] ≤ 106	80 < dwt [kt] ≤ 166	9.99%	7.33%	73.42%	5.54%	4.09
106 < sl [kt] ≤ 166	57 < Ws [h] ≤ 76	22.63%	10.24%	45.25%	5.41%	2.12
6 < nh [-] ≤ 7 and 52 < sl [kt] ≤ 73	dwt [kt] ≤ 80	7.21%	6.70%	92.98%	5.27%	4.68
Mat = IO and nh [-] > 8	sl [kt] > 166	11.13%	7.21%	64.77%	5.24%	3.66
Mat = IO and Ws [h] > 76	sl [kt] > 166	8.47%	6.70%	79.10%	5.20%	4.47
73 < sl [kt] ≤ 106 and 80 < dwt [kt] ≤ 166	6 < nh [-] ≤ 7	10.49%	7.33%	69.88%	5.09%	3.27
dwt [kt] > 180	sl [kt] > 166	14.16%	7.59%	53.57%	5.08%	3.03

Note: A lower leverage ratio means a less significant rule.

Table A2 Anomaly detection

DATA COLLECTED							*TARGET**	*RESULTS OF THE EXPERIMENT (ANOMALY DETECTION PHASE)*						
Ship nr.	*dwt [kt]*	*nm [-]*	*nh [-]*	*sl [kt]*	*Ws [h]*	*Mat.*	*Anomalous vessel?*	*Anomaly score*	*dwt [kt] importance*	*nm [-] importance*	*nh [-] importance*	*sl [kt] importance*	*Ws [h] importance*	*Mat. importance*
1	23	1	7	22	63	SC	1	0.503	0.26	0.10	0.13	0.20	0.13	0.17
2	176	2	8	41	92	IO	0	0.493	0.09	0.10	0.16	0.19	0.18	0.27
3	171	2	1	16	34	CC	1	0.581	0.07	0.06	0.34	0.14	0.09	0.32
4	170	2	4	39	77	SC	0	0.485	0.11	0.11	0.24	0.18	0.18	0.18
5	23	1	7	21	41	SC	0	0.476	0.26	0.10	0.14	0.20	0.12	0.18
6	22	1	4	19	37	SC	1	0.513	0.25	0.09	0.19	0.19	0.11	0.17
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
789	400	1	6	365	74	IO	1	0.661	0.34	0.05	0.09	0.29	0.06	0.17
790	176	1	9	173	34	IO	0	0.413	0.09	0.11	0.18	0.19	0.16	0.27
791	181	1	9	176	35	IO	0	0.418	0.10	0.10	0.18	0.20	0.15	0.27

Note: *Target field (category) for assessing the robustness of the method.

Table A3 Detailed results for each evaluation

Evaluation	*1*	*2*	*3*	*4*	*5*	*Average*
Accuracy	90.6%	89.2%	89.2%	91.8%	91.8%	90.52%
Precision	93.6%	94.1%	93.5%	95.6%	94.2%	94.17%
Recall	95.6%	93.4%	94.2%	94.9%	96.3%	94.86%
F-measure	0.9459	0.9373	0.9382	0.9524	0.952	0.95
phi	0.5824	0.5597	0.5238	0.6503	0.6654	0.6

Note: Positive class: the vessel is 0 class (normal vessel).

Table A4 Detailed results for each evaluation

Evaluation	*1*	*2*	*3*	*4*	*5*	*Average*
Accuracy	90.6%	89.2%	89.2%	91.8%	91.8%	90.52%
Precision	68.4%	60.9%	60.0%	68.2%	76.2%	66.73%
Recall	59.1%	63.6%	57.1%	71.4%	66.7%	63.59%
F-measure	0.6342	0.6222	0.5854	0.6977	0.711	0.65
phi	0.5824	0.5597	0.5238	0.6503	0.6654	0.6

Note: Positive class: the vessel is 1 class (anomalous vessel).

Table A5 Cross-validation: all classes

Target class	*All classes*	*0 class*	*1 class (anomalous vessels)*
Average accuracy	90.52%	90.52%	90.52%
Average precision	80.45%	94.17%	66.73%
Average recall	79.23%	94.86%	63.59%
Average F-measure	0.8	0.95	0.65
Average phi	0.6	0.6	0.6