applied sciences

MDPI

Article

Fraud Detection of Bulk Cargo Theft in Port Using Bayesian Network Models

Rongjia Song [1,2,*], Lei Huang [1,*], Weiping Cui [3], María Óskarsdóttir [4] and Jan Vanthienen [2]

[1] Department of Information Management, Beijing Jiaotong University, Beijing 100044, China
[2] Department of Decision Sciences and Information Management, KU Leuven, 3000 Leuven, Belgium; jan.vanthienen@kuleuven.be
[3] State Grid Energy Research Institute, State Grid Corporation of China, Beijing 102200, China; cuiweiping@sgeri.sgcc.com.cn
[4] Department of Computer Science, Reykjavik University, 101 Reykjavik, Iceland; mariaoskars@ru.is
* Correspondence: rjsong@bjtu.edu.cn (R.S.); lhuang@bjtu.edu.cn (L.H.)

Received: 3 January 2020; Accepted: 1 February 2020; Published: 5 February 2020

Featured Application: A novel data-driven methodology with multiple ranking and classification techniques combined and compared is proposed for proactively detecting the fraud of cargo loss risk. Various binary classifiers are compared to derive the suitable predictive model. Bayesian network performs best overall and visually shows the dependencies between fraud features.

Abstract: The fraud detection of cargo theft has been a serious issue in ports for a long time. Traditional research in detecting theft risk is expert- and survey-based, which is not optimal for proactive prediction. As we move into a pervasive and ubiquitous paradigm, the implications of external environment and system behavior are continuously captured as multi-source data. Therefore, we propose a novel data-driven approach for formulating predictive models for detecting bulk cargo theft in ports. More specifically, we apply various feature-ranking methods and classification algorithms for selecting an effective feature set of relevant risk elements. Then, implicit Bayesian networks are derived with the features to graphically present the relationship with the risk elements of fraud. Thus, various binary classifiers are compared to derive a suitable predictive model, and Bayesian network performs best overall. The resulting Bayesian networks are then comparatively analyzed based on the outcomes of model validation and testing, as well as essential domain knowledge. The experimental results show that predictive models are effective, with both accuracy and recall values greater than 0.8. These predictive models are not only useful for understanding the dependency between relevant risk elements, but also for supporting the strategy optimization of risk management.

Keywords: fraud detection; predictive modeling; bulk cargo theft; Bayesian network; port

1. Introduction

The operation in port is closely connected with other organizations, and organizational boundaries are frequently crossed, which leads to a complicated operation environment. Cargo theft is referred to as a silent crime, as it accounts for huge losses that frequently go unreported, and can even occur without the owner being aware until weeks or longer after the theft occurred [1]. In terms of bulk cargo, the situation can be worse since their cargo yards are normally huge and exposed without uniform packaging such as the coal yard. Even though more bulk cargo is illegally delivered, port and cargo owners rarely notice the cargo theft immediately. Most cases are detected in cases where there is not enough cargo to pick up for the final delivery. Some other cases are detected by prevention countermeasures of ports such as sampling inspection in terminals, or informer's accusations, such as witness drivers.

According to the China Police Daily, Chinese Public Security Traffic Management Authorities has cracked more than 4000 criminal cases of port logistics from 2013 to 2017, saving hundreds of millions Chinese Yuan [2]. This paper aims to detect the fraud risk of bulk cargo theft during truck delivery in port. More specifically, this means that drivers deliver more bulk cargo during transportation than the amount in the bill of lading, e.g., by cheating or other illegal means. Up until now, the most common and effective means against fraud was through sampling inspection during the operation of truck delivery, which is pristine, reactive, and at enormous expense. According to interviews with Guangzhou Port Group and Guangzhou Port Security Bureau, delivery trucks mostly execute cargo theft by reforming or retrofitting trucks to carry heavy filler such as liquid tanks, cement, stones or scrap when entering the terminal and weighing empty trucks. Then the reformed trucks furtively unload these fillers in the terminal and 'steal' bulk cargo due to the weight deviation of the empty trucks. Moreover, some common characteristics seem to exist among frauds, such as operational timing, weather, and locale of truck license. However, bulk cargo theft is difficult to detect manually, since (a) port terminals are vast, busy and dynamic; (b) various elements contribute to operational uncertainty, such as cargo type, truck information, and external environment; (c) automatic implements reduce manual intervention, such as the entry allowance using RFID; and (d) different organizations are involved in truck delivery, including the cargo owner, the port company, the delivery company, etc.

As IT technologies including IoT are accommodated in port operations, massive data is captured with implications for the external environment and system behavior [3]. The added value can be obtained by applying data analytics coupled to IoT, which is the cornerstone technology for proactively detecting bulk cargo theft. Research has also demonstrated that the amount of data produced and communicated in the logistics and supply chain is significantly increasing, thereby creating challenges for organizations that would like to reap the benefits of analyzing real operational data [4,5]. Hence, a data-driven approach based on Bayesian network is proposed for predicting fraud risk as well as analyzing risk elements to support better-informed management. The constructed model can be utilized to help cargo interests and security bureau to understand the inter-relations between the risk components of bulk cargo theft in port, further supporting early fraud detection and managerial strategy optimization. A real-world case study for a Chinese bulk port is presented to further evaluate the proposed data-driven method.

Thus, the contribution of this paper is fourfold:

- We propose a novel data-driven methodology for proactively detecting bulk cargo theft in port based on inspection records and operational data. The methodology combines and compares various ranking and classification techniques for fraud feature selection since algorithms may significantly differ on different datasets. Then various binary classifiers are compared to derive the suited predictive model for the theft fraud.
- We tackle the serious business issue by early detecting and deeply understanding the theft fraud of bulk cargo delivery in port. Bayesian networks perform best overall for effectively predicting the fraud, which can also visually show the dependencies between fraud features for optimizing the management strategy.
- The proposed methodology is better than traditional methods at detecting cargo theft for less reliance on expert expertise and survey-based knowledge and with visualization for adaptation.
- A case-based evaluation in real world that shows the successful application of the proposed methodology in a bulk port.

The remainder of this paper is organized as follows. Section 2 gives an overview of the existing research on port risk management. Section 3 proposes the data-driven methodology and Section 4 presents the empirical study on a port in China. Section 5 provides discussions in terms of the methodology application and experimental results. Finally, Section 6 concludes the paper and sets future perspectives.

2. Related Work on Port Risk Management

Strategic collaborations in logistics networks, of which ports are a significant node, reduce risks for all organizations in the network [6]. In ports, high-quality risk management is absolutely necessary for their sustainable development [7]. The present focus of risk management in ports are rather high level, mostly on development, management, organization and commercial issues of ports and terminals. Nevertheless, a few studies have tried to solve risk problems in practical operation of cargo delivery. In terms of the risk analysis process, both qualitative and quantitative techniques can be used. Nowadays a variety of techniques of risk analysis have been developed in industrial settings, normally in response to practical business problems. These techniques, such as physical inspections, flow charts, safety review, checklist analysis, relative ranking, cause–consequence analysis, 'what-if' analysis, failure modes, and so forth, all excessively depend on expert knowledge and manual preprocessing. As a result, in complex operational environments, experts may not be able to be fully aware of all situations, which may have negative influence on the result of risk detection and analysis.

The widespread use of digital technologies has led to the intelligent port operation paradigm, the amount of data collected during operational processes is growing at a rapid pace. Data-driven risk management is emerging to provide companies with better means to obtain early warning from an increasingly massive amount of data and gain a powerful competitive advantage, especially in the operational context [8,9]. More specifically, data from surveys, interviews and expert classification plays a significant role in extant research on port risk analysis. For instance, Ref. [10] qualitatively assessed the potential risks for ports using data from interviews with the administrations of a container terminal in an empirical study. Furthermore, Ref. [11] identified 19 port-centric disruptive events based on the literature and used data from surveys to generate risk metrics for analyzing port-centric supply chain disruption threats. In [7], fault tree analysis and event tree analysis were used to assess the risk factors associated with sea ports and offshore terminals operations and management based on experts' judgement.

Regarding current research on cargo theft, most studies focus on the problem of theft prevention by introducing software systems, hardware facilities or a set of managerial measures. For instance, some algorithms for specific business settings and monitoring methods such as text messaging have been proposed as patents. A cargo theft information processing system is introduced in [12]. Moreover, the development of fast-working, nonintrusive X-ray and detection devices has been suggested to help ensure cargo. In addition to academic papers, there are also some regulations aiming at theft prevention, for instance the Freight Security Requirement and the Trucking Security Requirement issued by Technology Assets Protection Association. However, many fail to realize that the programs designed are not effective enough for preventing theft. To our knowledge, there are no data-driven methods to cope with cargo theft risk analytics in the real-world operational scenario in general and bulk cargo port in particular.

This paper aims at proposing a data-driven method in order to apply it to a real-world bulk cargo port in the south of China to analyze the relations between risk components of cargo theft and early detection of theft risk. These predictive models are not only useful for understanding the dependency between relevant elements of fraud risk, but also for supporting the strategy optimization of preventing bulk cargo theft.

3. Methods

3.1. Bayesian Network

Bayesian networks are acyclic directed graphs that represent the conditional dependencies between a set of variables [13]. Bayesian networks are also considered to be a graphical inference technique used to express the causal relationship among the variables. Bayesian networks are able to perform both predictive analysis and diagnostic analysis [14]. Hence, they are widely used in modeling of complex systems and risk analysis of a wide variety of accidents based on probabilistic and uncertain knowledge thanks to its flexible structure and reasoning engine.

Many authors have investigated different techniques based on Bayesian network to analyze risks in logistics or port scenarios, but there is a lack of analyzing the risk of cargo theft using real operational data. For instance, an expert system for maritime safety management based on Bayesian network was introduced in [15]. The study [16] provides a Bayesian network-based solution to the problem of offshore piracy ranging from the detection of a potential threat to the implementation of a response. A method to generate a Bayesian network from a process model is proposed in [17] to analyze the lateness risk of container handling. [18] presents a Bayesian network-based model to predict accident consequences in the Tianjin port.

Clearly, Bayesian networks have been used extensively for port risk analysis where the model parameters are based on expert elicitation and learning from historical data. It is a powerful technique for analyzing relations between risk components, visualizing specific risk structures and reasoning risk potential as well. As a result, Bayesian networks are used in this paper for the risk analysis of bulk cargo theft based on real theft records from a port security bureau and operational data from a bulk cargo port. The constructed model can be utilized to help cargo interests and security bureau to understand the inter-relations between risk components of bulk cargo theft in ports and further support the risk early detection and risk control strategy optimization.

3.2. Methodology

In this study, we consider cargo theft detection to be a two-class prediction problem. As a supervised learning problem, the performance of a classifier relies on how effectively it can learn from the training set. Another aspect is the proper adjustment of the learning process. Our proposed methodology, shown in Figure 1, consists of three main parts: (a) feature selection; (b) Bayesian network learning; and, (c) model validation, testing and comparative analysis.

3.2.1. Step A. Exploring Event Logs

The historical records of bulk cargo theft are in the form of unstructured data, such as text, which were created during the investigation of previous frauds. Hence, word segmentation techniques are needed to extract meaningful features as structured data, such as license plate numbers, timestamps and others.

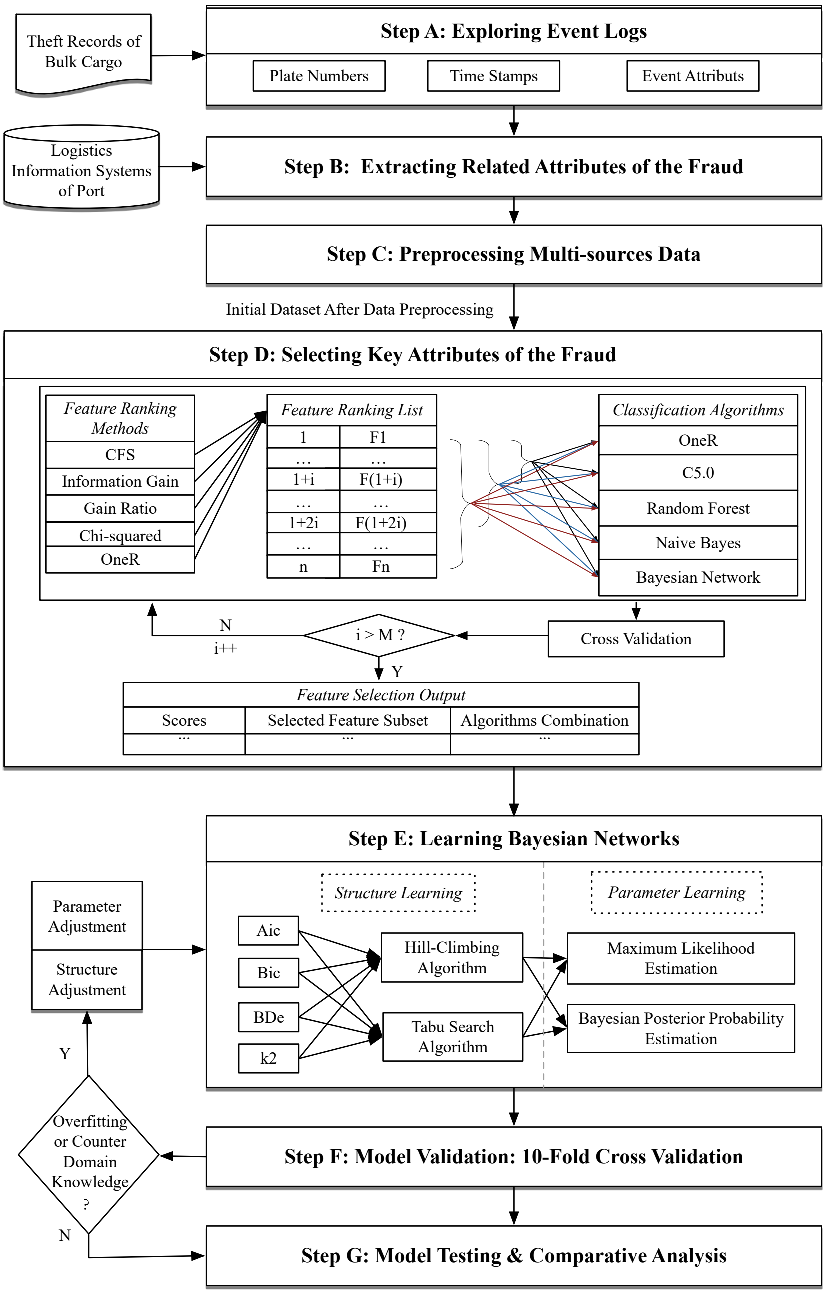
3.2.2. Step B. Extracting Related Attributes of the Fraud

Related operational logs in the logistics information system need to be extracted in terms of license plate numbers and timestamps, which provide complementary features of the fraud. All irrelevant features need to be further filtered out from the extracted operational logs. As a result, the initial dataset involving all relevant features is obtained.

3.2.3. Step C. Preprocessing Multi-Sources Data

After data exploration and data collection, we need to preprocess the initial dataset, which mainly consists of two tasks: data balancing and data discretization.

From a model-refinement perspective, the fraud fact of bulk cargo theft is the focus of prediction but has a much smaller sample size than normal operations. Synthetic Minority Oversampling TEchnique (SMOTE) [19] method is used as the imbalanced classification technique to improve the model sensitivity on the minority class. On the other hand, we keep binary variables and categorical variables in the dataset, but use an entropy-based technique: the Minimum Description Length Principle (MDLP) [20] to discretize continuous variables. More specifically, bulk cargo theft risk is set as the target variable for the supervised data discretization.



Theft Records of Bulk Cargo
Step A: Exploring Event Logs
Plate Numbers
Time Stamps
Event Attributs
Logistics Information Systems of Port
Step B: Extracting Related Attributes of the Fraud
Step C: Preprocessing Multi-sources Data
Initial Dataset After Data Preprocessing
Step D: Selecting Key Attributes of the Fraud
Feature Ranking Methods
CFS
Information Gain
Gain Ratio
Chi-squared
OneR
Feature Ranking List
1 F1
1+i F(1+i)
1+2i F(1+2i)
n Fn
Classification Algorithms
OneR
C5.0
Random Forest
Naive Bayes
Bayesian Network
Cross Validation
i > M ?
N
i++
Y
Feature Selection Output
Scores
Selected Feature Subset
Algorithms Combination
Step E: Learning Bayesian Networks
Structure Learning
Parameter Learning
Aic
Bic
BDe
k2
Hill-Climbing Algorithm
Tabu Search Algorithm
Maximum Likelihood Estimation
Bayesian Posterior Probability Estimation
Parameter Adjustment
Structure Adjustment
Y
Overfitting or Counter Domain Knowledge ?
N
Step F: Model Validation: 10-Fold Cross Validation
Step G: Model Testing & Comparative Analysis

Figure 1. The methodology of the proposed method for constructing the predictive model of fraud detection on bulk cargo theft in port. (For a detailed explanation, see previously in Section 3.2.).

3.2.4. Step D. Selecting Key Attributes of the Fraud

After data preprocessing, feature selection is the next step to reduce the dimensionality of the feature space by removing redundant, irrelevant, or noisy data. A Bayesian network that is constructed directly from data involving all attributes can be rather complicated and thus lack readability because too much unimportant data is involved. Feature selection is therefore crucial to obtain the most effective feature set for the model construction. It supports a faster learning process, improves the data quality and thereby the performance of prediction, as well as increasing the comprehensibility of the analytical results [21–23]. Research [24] also shows that there is no best ranking index for different datasets, since the function of the number of features used may significantly differ. The only way to be sure that the highest accuracy is obtained in practical problems is testing a given classifier on several feature subsets, obtained from different ranking indices. Hence, we use feature ranking methods and feature selection methods with two basic steps of subset generation and subset evaluation for the ranking of each feature in every sub-dataset. Ranking methods with different classification algorithms might give different results for the classification accuracy. Comparative analysis is then performed based on the result of cross-validation. Finally, we choose the most effective composition of ranking method and classification technique to select and ensure a subset of features giving the highest effectiveness.

3.2.5. Step E. Learning Bayesian Networks

Learning Bayesian networks includes two major tasks: learning the graphical structure, and learning the parameters for that structure. We designed comparative experiments for the structure learning by utilizing combinations of four score functions and heuristic structure search algorithms including Hill Climbing and Tabu Search. More specifically, score functions include Akaike information criterion (Aic), Bayesian information criterion (Bic), Bayesian Dirichlet equivalent score (BDe) and k2 algorithm. In addition, the goal of parameter learning is to obtain the optimal parameter values for a given structure with a given corpus of complete data. Thus, we could simply use the empirical conditional frequencies from the data. As parameter learning is straightforward, we will mainly focus on learning the Bayesian network structure.

3.2.6. Step F. Model Validation

Then, k-fold cross-validation is used to evaluate the performance of classifiers. The classification error is chosen as the loss function to estimate how accurately the model will perform in practice, in order to limit problems like overfitting and give insights on how the model will generalize to an independent dataset.

3.2.7. Step G. Model Testing and Comparative Analysis

All optimal models of different combinations with score functions and heuristic structure search algorithms need to be tested by applying them to the separate test dataset. The output is a confusion matrix, which can be used for analyzing the accuracy, the sensitivity and the specificity in order to choose the optimum predictive model for detecting bulk cargo theft.

4. Empirical Study

In this section, we apply the proposed methodology to a real-word dataset which were extracted from the sampling inspection of Port Security Bureau and the logistics information system of Port Company. We present the empirical study, which consists of the feature selection, the structure learning, the parameter learning, stratified cross-validation and model testing. Because of the limited space, we use the abbreviations HC for Hill Climbing; HC.Aic for the combination of Hill Climbing and Aic and TABU.Aic for Tabu Search combined with Aic. For convenience, all combinations adhere to this naming format.

4.1. Data Description and Preprocessing

The data used in this study originates from Guangzhou Port Group and Guangzhou Port Security Bureau in China. It consists of inspection records of delivery trucks from September 2013 until May 2014. In total, 5320 records are used, of which 4898 are reports of normal deliveries and 422 are records of delivery with theft, which we also refer to as fraud delivery. Table 1 presents two examples of fraud delivery reports from sampling inspection.

First, we apply text segmentation to obtain features with structured storage using Rapidminer tool. Instead of using text storage with meaningless redundancy and difficult implementation, we segment these inspection records as structured features including data about the driver, truck, delivery cargo, operational time, fraud cargo and theft mode, and then store it in respective fields of databases. Then we further extract related operational records from the logistics information system in terms of license plate numbers and timestamps. Based on these data, the domain experts of the practitioner and the police help us to identify 38 relevant elements of fraud risk as the initial feature set (Table 2).

Table 1. Examples of fraud delivery records from sampling inspection.

ID	Event Description
1	On the night of 14 March 2014, driver Zhou drove the truck with plate number XiangX. He stole 3.88 tons of brown coal (Indonesia coal, Lower Heating Value: 4400 Kcal/Kg) from terminal A of Guangzhou Port by reforming or retrofitting the water tank of the truck. The truck came into the terminal A with water in the tank and released the water before loading. Then he secretly transported the stolen cargo, identified to have a value of 455 yuan per ton and 1765.4 yuan in total, to the factory B in DongGuan.
2	On the night of 11 April 2014, driver Cao drove the truck with plate number YueY, and stole 3.89 tons of coal (Indonesian Steam Coal, Lower Heating Value: 4651 Kcal/Kg) from terminal C of Guangzhou Port by reforming the car hopper of the truck. The truck came into the terminal C with objects in the car hopper and that were released before loading. Then he secretly transported this cargo from the theft, identified to have a value of 518 yuan per ton and 2011.3 yuan in total, to the factory D in DongGuan.

Table 2. Relevant elements of fraud risk on bulk cargo theft in port.

ID	Category	Features	Explanation
1	Truck Driver	Age	continuous variable
2		Education Background	1: High School and Above, 0: others
3		Credit in Port	0: Low, 1: Medium, 2: High
4		Region	0: Native, 1: Nonnative
5	Bulk Cargo	Cargo Type	0-3: top 4 of largest amount of cargo processed in port, 4: others
6		Trade Type	0: domestic trade, 1: foreign trade
7		Import or Export	0: import, 1: export
8		Customs Supervision	0: unsupervised, 1: supervised
9		Cargo Price	continuous variable
10	Delivery Truck	Truck Manufacturer	0-3: top 4 of largest number of trucks in port, 4: others
11		Year of Manufacture	continuous variable
12		Registration Region	0: province inside, 1: others
13		Truck Category	0-3: top 4 of most common truck registered, 4: others
14		Reformed or Not	0: reformed, 1: unreformed
15		Ownership	0: individual, 1: company
16		Times of Filing	1: three times and above, 0: others
17		Blacklist	1: ever, 0: never
18	Cargo Storage Yard	Yard Category	0: outdoor yard, 1: warehouse, 2: others
19		Specialized Yard	0: non-specialized, 1: specialized
20		Cargo Volume	continuous variable
21	Consignor	Consignor Level	0-3: classified based on the volume of business
22		Contract Category	0: for a single vessel, 1: long-term, 2: others
23		Means of Payments	0: after operation completion, 1: monthly payment, 2: payment in advance
24	Operational Delivery	Arrival time	0: 6:00-13:59, 1: 14:00-21:59, 2: others
25		Leave time	0: 6:00-13:59, 1: 14:00-21:59, 2: others
26		Operation Period	0: 2 hours and less, 1: 4 hours and above, 2: others
27		Operation Period Deviation	continuous variable
28		Operation Period Fluctuation	continuous variable
29		Cargo Weight Deviation	continuous variable
30		Cargo Weight Fluctuation	continuous variable
31		Weighbridge of Empty Cargo Truck	serial number of weighbridges
32		Weighbridge of Heavy Cargo Truck	serial number of weighbridges
33		Arrival Date	0: work day, 1: holiday
34		Shift Category	0: three shifts, 1: two shifts
35		Number of Workers	continuous variable, number of workers in operation
36		Number of Machineries	continuous variable, number of machineries in operation
37	Others	Weather	0: rain, 1: no rain
38		Busy Degree	continuous variable, operational amount per hour

After data preprocessing, which includes data cleaning, data balancing using SMOTE method and data discretization using MDLP method, we were left with 9796 records in structure storage with 38 features in addition to the response feature, i.e., normal of risk delivery.

4.2. Feature Selection

Diverse feature ranking and feature selection techniques have been proposed in the machine learning literature [25]. The purpose of these techniques is to discard irrelevant or redundant features from a given feature vector. Research also shows that there is no best ranking index for different datasets and different classifiers accuracy curves, as the function of the number of features used may significantly differ [24]. In this case, we use different methods to rank the features. Then feature subsets are generated based on the results of the feature ranking method. These feature subsets are further evaluated by using various classification algorithms. Finally, we obtain the selected feature subset which has the optimal performance for classification.

4.2.1. Exploring Event Logs: Feature Subset Generation Based on Ranking Method

We examine the practical usefulness of following commonly used methods including CFS (Correlation-based Feature Selection), IG (Information Gain), GR (Gain Ratio), Chi (Chi-squared) and OneR to rank the features, in other words, features in the initial feature set and output different ranking results. These feature ranking methods are statistical and entropy-based [26], with good performance in various domains [27]. In Table 3, a lower number means a more relevant feature for cargo theft fraud.

Table 3. Feature ranking results.

Feature	**CFS**	**IG**	**GR**	**Chi**	**OneR**
1—Age	8	7	***9***	8	10
2—Education Background	34	34	34	34	34
3—Credit in Port	14	16	***15***	15	12
4—Region	32	33	31	27	29
5—Cargo Type	18	17	***16***	16	19
6—Trade Type	30	29	28	33	3
7—Import or Export	38	38	38	38	35
8—Customs Supervision	7	5	***7***	5	9
9—Cargo Price	33	32	33	32	28
10—Truck Manufacturer	19	19	21	21	20
11—Year of Manufacture	9	10	***10***	11	13
12—Registration Region	27	31	32	30	33
13—Truck Category	12	13	***13***	14	14
14—Reformed or Not	1	3	***3***	3	5
15—Ownership	26	28	29	29	31
16—Time of Filing	15	18	***17***	18	18
17—Blacklist	11	12	***11***	9	8
18—Yard Category	6	6	***6***	7	4
19—Specialized Yard	21	25	24	25	24
20—Cargo Volume	29	26	25	24	27
21—Consignor Level	13	11	***12***	13	17
22—Contract Category	20	22	19	19	26
23—Means of Payment	5	8	***4***	4	6
24—Arrival Time	2	2	***2***	6	1
25—Leave Time	22	20	20	22	21
26—Operation Period	4	4	***5***	2	7
27—Operation Period Deviation	37	36	36	36	36
28—Operation Period Fluctuation	28	30	26	28	30
29—Cargo Weight Deviation	3	1	***1***	1	2
30—Cargo Weight Fluctuation	31	27	30	31	32
31—Weighbridge of Empty Cargo Truck	35	35	35	35	36
32—Weighbridge of Heavy Cargo Truck	36	37	37	37	38
33—Arrival Date	24	24	22	20	22
34—Shift Category	25	23	27	26	23
35—Number of Workers	17	15	18	17	16
36—Number of Machines	23	21	23	23	25
37—Weather	10	9	***8***	10	15
38—Degree of Busyness	16	14	***14***	12	11

Based on the feature rankings in Table 3, we find that:

- Different ranking results are obtained by using various feature ranking methods. Some of these outcomes are similar, especially using IG, GR and Chi.
- The features 'Cargo Weight Deviation', 'Reformed or Not', and 'Arrival Time' always rank the highest independent of the method used. This means that these three features are closely relevant to the theft risk of bulk cargo.
- The features 'Education Background', 'Weighbridge of Empty Cargo Truck', 'Weighbridge of Heavy Cargo Truck' and 'Operation Period Deviation' always rank the lowest. This means that these four features are the most irrelevant features to the theft risk of bulk cargo.

Finally, feature subsets can be generated based on the how relevant the feature is to the theft risk of bulk cargo. We have 38 features in the initial feature set, so 38 feature subsets are generated according to the feature ranking result of each ranking method and the top *i* relevant features constitute multiple subsets, respectively (Figure 2). For instance, the feature ranking results based on the GR method are shown in bold and italic in Table 3 and 38 feature subsets are generated including S1 = {R29}, S2 = {R29, R24}, S3 = {R29, R24, R14}, S4 = {R29, R24, R14, R23}, . . . , S38 = {R1, R2, . . . , R38}. Hence, evaluating and selecting the most effective subset of features for classification is the next task.

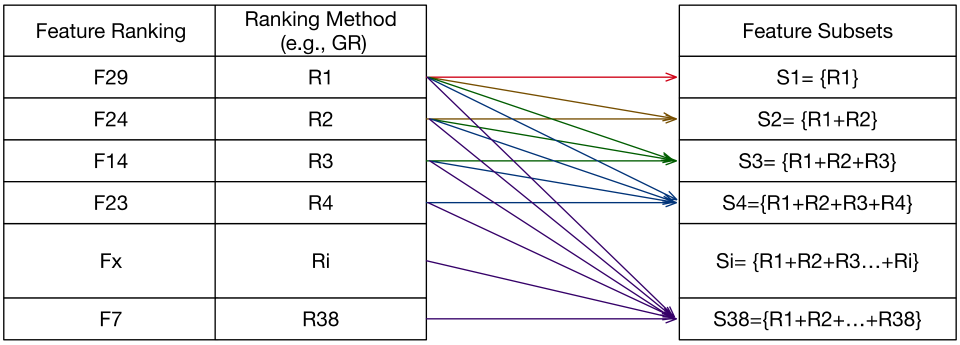


Figure 2. The method of feature subsets generation.

4.2.2. Feature Subset Evaluation Based on Classification Algorithms

A wide range of classification algorithms is available for validating feature subsets, each with its strengths and weaknesses [28]. We use OneR, Decision Tree (DT), Random Forest (RF), Naive Bayesian (NB) and Bayesian Network (BN) as classifiers to evaluate the subsets, which are widely used and perform well in various domains, for instance in [29]. More specifically, OneR is one of the simplest algorithms for use as a classifier, and its classification result can be used as a benchmark for others. Additionally, DT is simple to interpret and RF is an evolved version that performs better with large amounts of data in a short time. An advantage of the NB classifier is that it requires a small training dataset to estimate the statistical parameters necessary for classification, and BN is an evolved version for dealing with uncertainty and reasoning. Classification accuracy is estimated using ten-fold cross-validation. In this practical problem, we test a given classifier on several feature subsets, obtained from different ranking indices, using classic evaluation indices based on the determined operations of classification confusion matrix [30].

Specifically, we use accuracy, precision and recall to measure the effectiveness of classification results. In particular, recall is the key evaluation index because it measures what percentage of all frauds are correctly predicted. Moreover, variance is used for evaluating the reliability of classification results. We process 100 rounds of ten-fold cross-validation in total. In each round, a given classifier is applied to a feature set of one of the ranking methods (5×38×5 times cross-validation in one round).

We decide the final feature set must fulfill the following three criteria: (a) the accuracy must be above 70%; (b) the precision and recall must be above the zeroR baseline (0.5); (c) the variance of the accuracy, precision and recall must be between −20% and 20%. In addition, in every round, one combination of classifier, selected feature set and ranking method must have the most effective performance in terms of the recall. More specifically, as the number of selected features increases in a given composition of classifier and ranking method, the recall of the classification increases until it reaches a peak and then slightly decreases again. The classification results using the GR ranking method are presented in Figure 3 as an example.

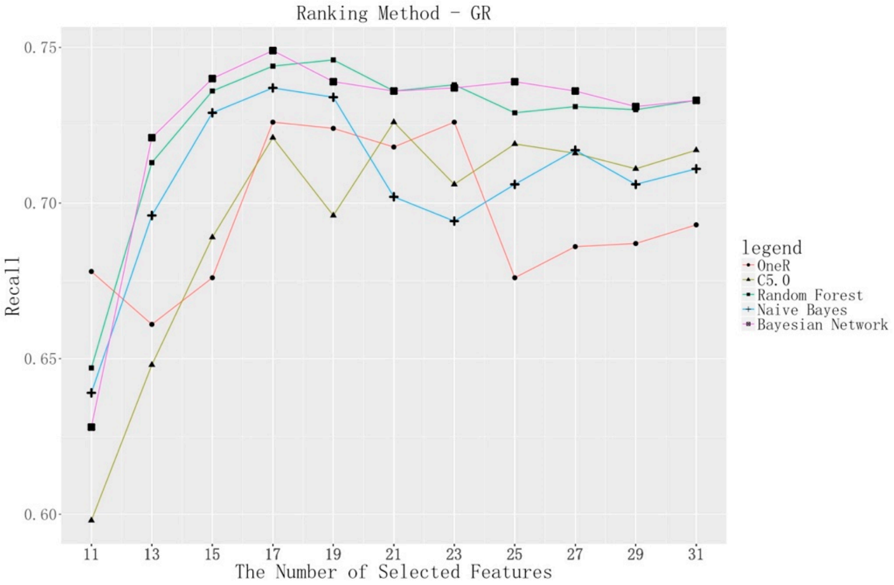


Figure 3. The classification results using the GR ranking method based on the Recall index.

Table 4 presents the compositional classification results of all round 10-fold cross-validation, which shows that the quality of classification for this dataset is obviously influenced by the choice of ranking indices. When 17 features are selected, most rounds obtain the best performance of cross-validation with 61 of 100 rounds. Moreover, 47% rounds gain the best performance by using the combination of GR and Bayesian network. Moreover, classification recall with GR ranking method is the highest for Bayesian network, and also quite high for other classification algorithms, depicted in Figure 3. Hence, we select these 17 features using the GR ranking method as the most effective feature set for classification to construct the predictive model (features in bold and italic of Table 3).

Table 4. Compositional classification results of all rounds.

Rounds	Number of Selected Features	Ranking Method	Classifier
35	17	GR	Bayesian Network
21	19	OneR	Random Forest
21	17	IG	Naïve Bayesian
12	15	GR	Bayesian Network
6	16	GR	Random Forest
5	17	CFS	OneR

4.3. Initial Parameter Setting

4.3.1. Parameter Setting within the Structure Learning

Table 5 presents the specific parameter setting in the methodology applied in this empirical study.

- 'Random restarts' is a configurable number of perturbing operations and a proceeded initial network structure, which can be used to avoid poor local maximization or optimization [31]. We utilize 'random restarts' by setting the parameters of 'restart' and 'perturb', respectively. More specifically, 'restart' makes it possible to probe more area of the search space and increase the chance of finding a structure that fits the data better. Meanwhile, we consider the limit of running time and memory and set them as integers.

- We set the parameter 'mapx' for continuously adjusting the maximum of father nodes in order to analyze the relationship between the maximum of father nodes and the accuracy of predictive models. This setting is used to prevent the learnt structure from becoming too complicated, so 7 is set as the maximum. In Section 4.4.1, 'mapx' is determined as 5, with the lowest training error and cross-validation error during the network structure learning (Figure 4).
- We measure the strength of directed arcs, compare different network structures and further correct the dependencies between nodes, which facilitates a more understandable network structure.

Table 5. Parameter setting in the methodology.

Step	Parameter	Value	Explanation
Structure Learning	restart	100	For avoiding local optimization
	perturb	5	Times of inserting, deleting and inverting directed edges at restart
	mapx	2–7	Gradually increasing the maximum number of father nodes
K-fold Stratified Cross-validation	target	Theft Risk	Setting the theft risk of bulk cargo as the target node of this prediction
	k	10	Every time processing the cross-validation, the sample dataset needs to be divided by k
	runs	10	Run 10 times 10-fold cross-validation

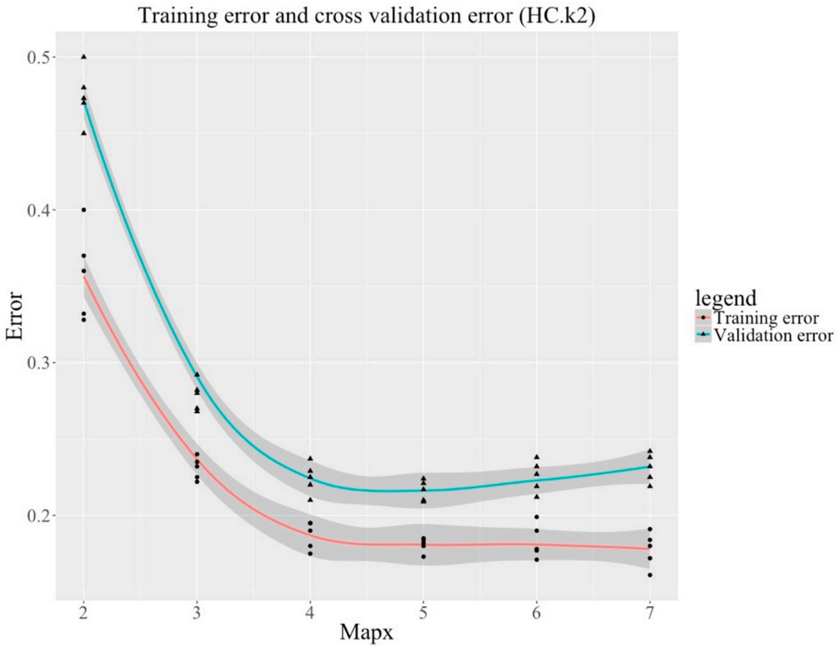


Figure 4. The training error and cross-validation error using the structure search algorithm with HC, the score function with k2, the parameter learning algorithm with Bayesian posterior probability estimation.

4.3.2. 10-Fold Stratified Cross-Validation

On basis of our datasets and extant literatures, we set k as 10 in this study. In every sub-dataset, the proportion of label data is almost the same as it in the original sample by using stratified cross-validation. We process 10 times 10-fold stratified cross-validation and calculate the average accuracy, average sensitivity and average specificity.

4.4. Experimental Results

4.4.1. Structure Learning and Analysis

We first equally divide the whole dataset into three subsets, keeping the same proportion of risk delivery in each subset as in the original dataset. Then, two subsets are used as the training set for learning and validating the model. The last subset is used as the test set for testing the model performance. Bayesian networks are able to graphically present the dependency between features through its structure topology. However, it could be difficult to understand or explain in case of overfitting, which means a very complicated network structure. Hence, in this study, we try to choose an optimal value for the maximum number of father nodes while learning the network structure, in order to insure that the structure learned is both explainable and effective. As the parameter 'mapx' increases, errors of training and cross-validation decline sharply and the level of using combinations of structure search algorithms, score functions and parameter learning algorithms as well. Moreover, in the case of a maximum of five father nodes, errors are at their minimum independent of the combination. Thus, we determine the maximum number of father nodes as five. Using the structure search algorithm with HC, the score function with k2, the parameter learning algorithm with Bayesian posterior probability estimation is presented in Figure 4 as an example. Eight Bayesian network structures were built using the structure search algorithm with HC and TABU, as well as various score functions. More specifically, two network structures learnt with HC.k2 and TABU.k2 are presented as examples in Figures 5 and 6, and others are included as Figures A1–A6 in Appendix A. These score-based learning algorithms are general-purpose heuristic optimization algorithms that rank network structures with respect to a goodness-of-fit score [31]. The network provides a very clear view of the relevant features of cargo theft risk during the delivery process, as well as the relationship between the features. In the figures, a directed arc between nodes indicates the dependency relationship between the two features. More specifically, the thicker the arc is the stronger the dependency. Moreover, we measure arc strength in the network structure according to the score gain/loss which would be caused by the arc's removal, while keeping the rest of the network structure fixed. In other words, it is the difference between the score of the network with and without the arc. Negative values correspond to decreases in the network score and positive values correspond to increases in the network score (the stronger the relationship is, the more negative the difference). In addition, the arcs in red dotted lines highlight differences between these network structures.

Even when applying the same structure search algorithms, the learnt Bayesian network structures are different due to different score functions. Nevertheless, all network structures derived are rather similar, also in terms of arc strength, especially when HC is applied. Moreover, meaningful insights are graphically provided in these network structures, especially coupling these results from quantitative information with qualitative knowledge offered by domain experts. The network structures need to be further refined based on the domain knowledge and expert expertise, as well as adding countermeasures. For instance, these network structures indicate that the fraud risk of bulk cargo theft is directly dependent on the features 'Cargo Weight Deviation', 'Reform or Not', 'Arrival Time', 'Means of Payment' and 'Weather'. The departments of interest should therefore check information on these features before delivery trucks leave the port. In particular, the feature 'Cargo Weight Deviation' is closely related to the theft risk of bulk cargo. At this point, this insight conforms to in-use practical tactics of recognizing abnormal cargo weight in delivery trucks for detecting the cargo theft. Moreover, the dependency relationship between the features 'Means of Payment', 'Cargo type', 'Customs Supervision' and 'Arrival Time' is slightly different in four network structures of applying HC.

In addition, the network structures that apply TABU indicate that there is a dependency relationship between the features 'Consignor Level' and 'Year of Manufacture' (highlighted by the yellow circle in Figure 6). This relationship counters the domain knowledge, which should be removed to refine these network structures.

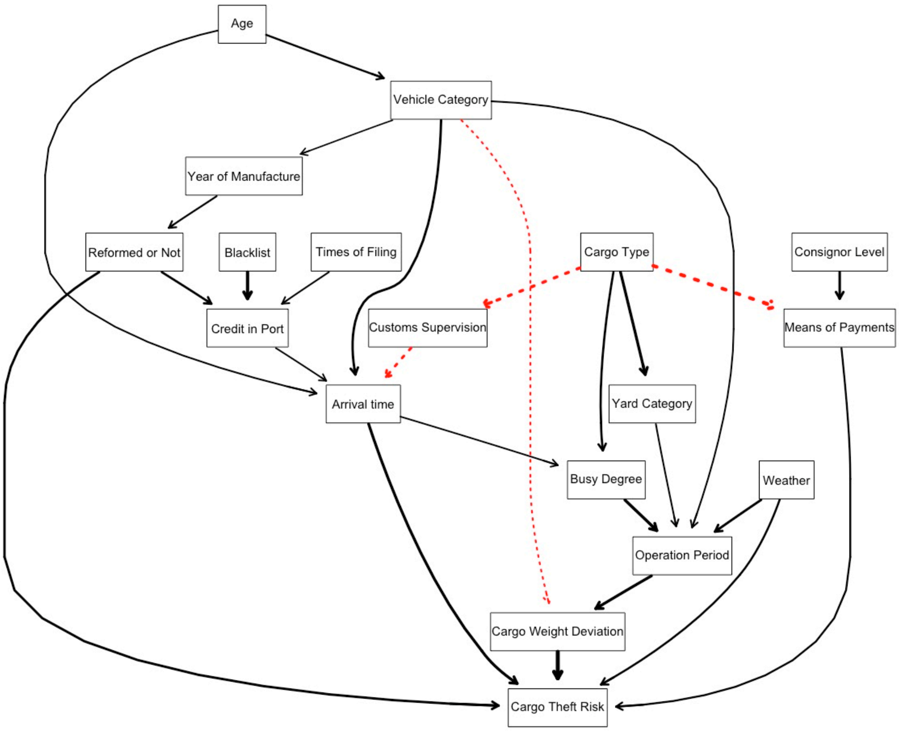


Figure 5. The Bayesian network structure derived using combinations of heuristic structure search algorithm with HC and score function with k2.

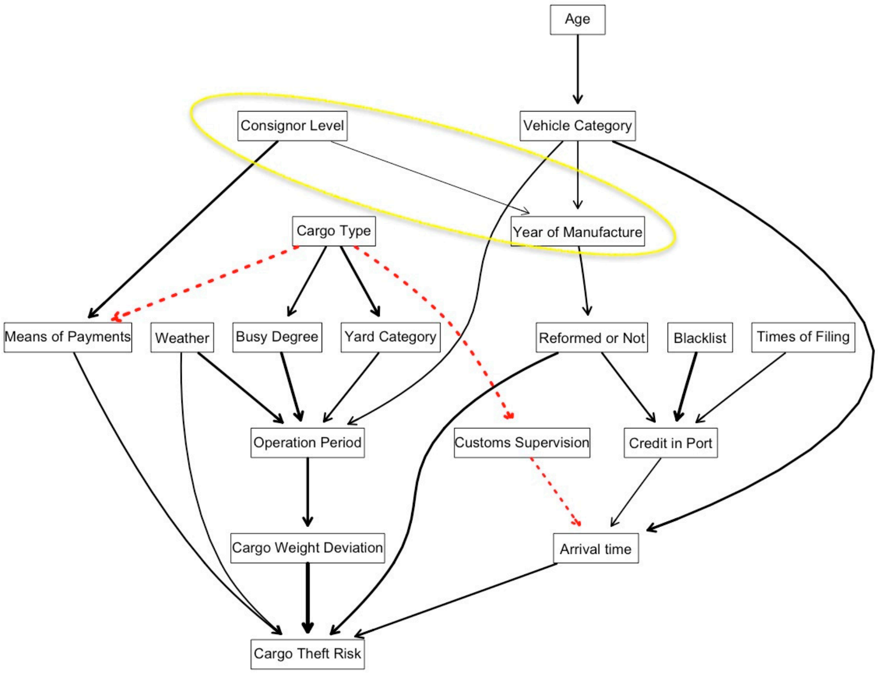


Figure 6. The Bayesian network structure derived using combinations of heuristic structure search algorithm with TABU and score function with k2.

4.4.2. Parameter Learning and Analysis

To learn from discrete data, Bnlearn provides two methods for parameter learning: Maximum Likelihood estimation and Bayesian posterior probability estimation [31]. These two methods are the most popular and effective methods for finding the 'best fitting' model from a specific class of models based on a particular dataset, but how they go about it is somewhat different. Maximum Likelihood estimation views the parameters as quantities whose values are fixed but unknown, and estimates parameter values by maximizing the likelihood (probability) of observing the actual training examples. However, Bayesian posterior probability estimation views parameters as random variables having some known prior distribution and transforms the parameters' prior distribution into posterior distribution by observing the actual training examples. Moreover, in Bayesian posterior probability estimation, when considering new samples, the posterior density function gains sharper performance, guaranteeing the accuracy of the nearby peak parameters.

In cases where the training dataset is very large, the results of the parameter learning are the same using these two methods. However, in reality, and especially in our case, the training set is rather small. Hence, we should consider the choice of these two methods from two perspectives of computational complexity and computational veracity. On the one hand, Maximum Likelihood estimation with differential calculus has lower computational complexity than Bayesian posterior probability estimation with multiple integrals. On the other hand, Bayesian posterior probability estimation normally exhibits better computational veracity than Maximum Likelihood estimation.

Since the training set is not big, Bayesian posterior probability estimation performs better in this case. We present an example to analyze experimental results of the parameter learning. For instance, Figure 7 shows the conditional probability distribution of the 'Cargo Theft Risk' node using the structure learning algorithm with HC, the score function with k2, the parameter learning algorithm with Bayesian posterior probability estimation and five as the maximum number of father nodes.

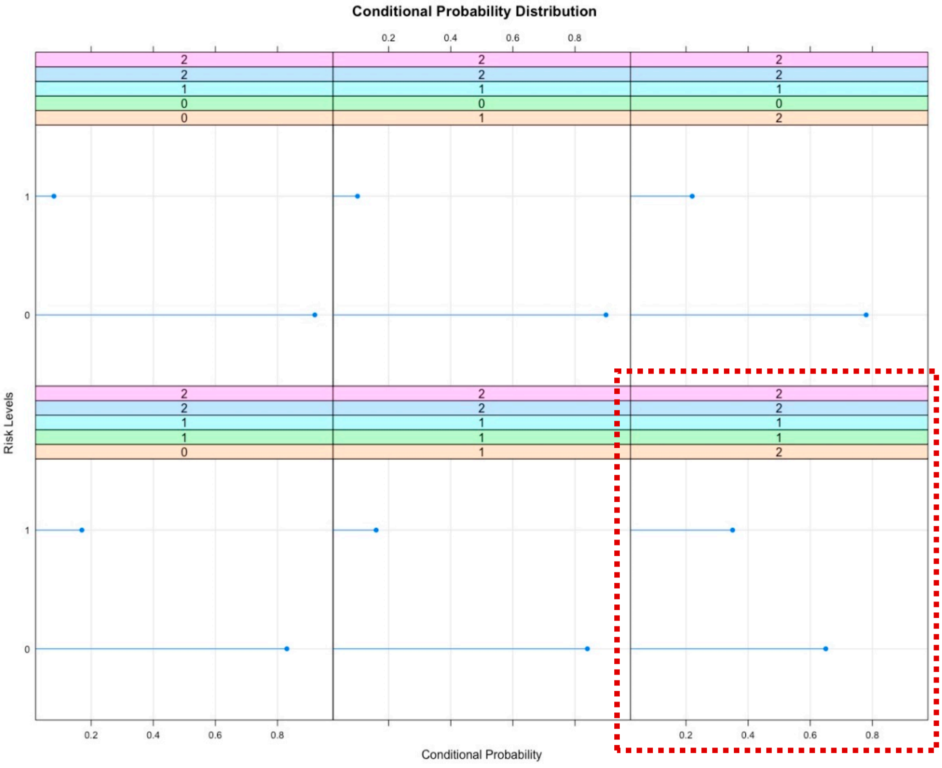


Figure 7. The conditional probability distribution of 'Cargo Theft Risk' node using the combination of structure learning algorithm with HC, the score function with k2, the parameter learning algorithm with Bayesian posterior probability estimation and five as the maximum number of father nodes.

The subfigure in the bottom right corner (highlight in the red and dotted rectangle) of Figure 7 indicates that the cargo theft risk is highest when 'Cargo Weight Deviation' equals 2, 'Means of Payments' equals 2, 'Reformed or Not' equals 1, 'Arrival time' equals 2 and 'Weather' equals 1. In other words, this means that when a reformed truck arrives at the port between 10 p.m. and 6 a.m. to pick up a delivery of prepaid cargo and the cargo weight deviates greatly from the normal value, then there is a significant fraud risk of bulk cargo theft.

4.4.3. Model Validation and Model Testing

Cross-validation is performed for model validation. Figures 8 and 9 respectively present the predictive error of cross-validation using Bayesian posterior probability estimation and Maximum Likelihood estimation.

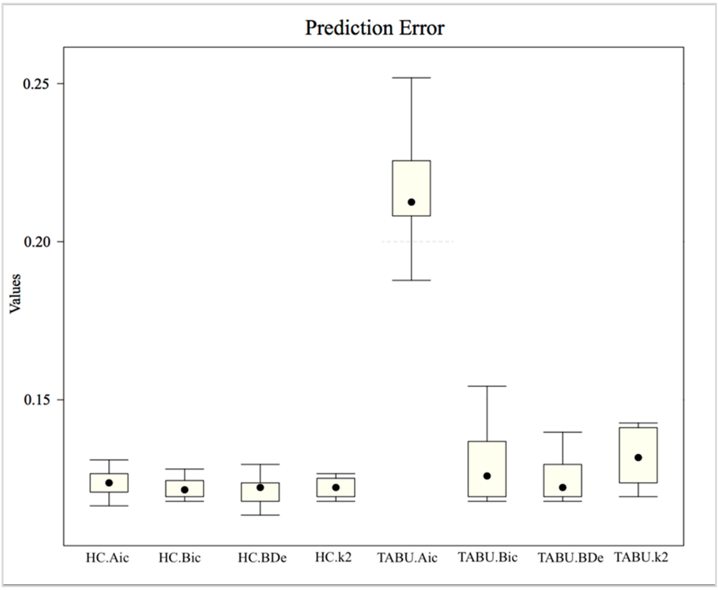


Figure 8. The predictive error of cross-validation using Bayesian posterior probability estimation.

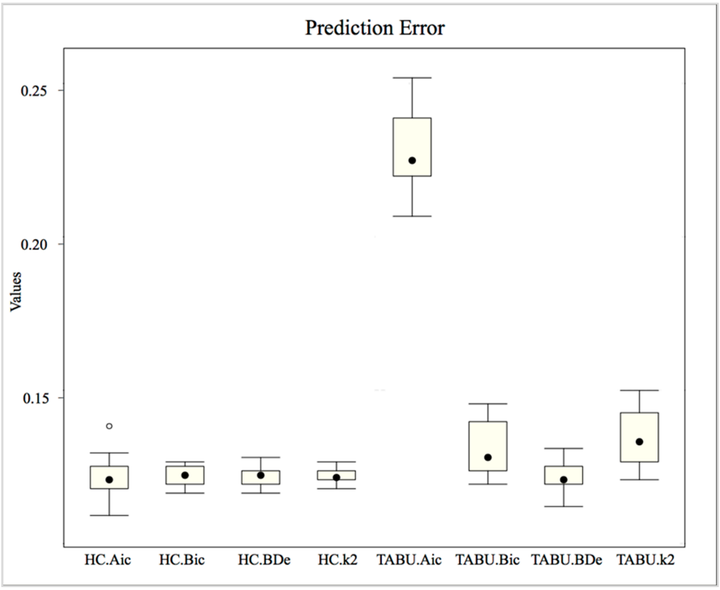


Figure 9. The predictive error of cross-validation using Maximum Likelihood estimation.

In the case of the structure learning algorithm with HC, the medians of the predictive errors with Aic, Bic, BDe and k2 are similar, and the range is relatively small. Nevertheless, there is an outlier in the predictive errors of HC.Aic, which indicates the volatility. Moreover, the predictive error of HC.k2 has the smallest range while also being most stable. In the case of the structure learning algorithm with TABU, the ranges of predictive error are obviously larger than with HC. More specifically, the TABU.BDe performs best when the structure learning algorithm is TABU.

By comparing the medians of the predictive errors, we see that Bayesian posterior probability estimation performs better than Maximum Likelihood estimation as a whole. Moreover, HC performs better than TABU, and the HC.k2 with Bayesian posterior probability estimation performs the best.

Then, Figure 10 presents the results of the model testing with various combinations of structure learning algorithms and score functions. All models except for TABU.Aic have an accuracy above 80%, which indicates that the risk prediction model based on Bayesian network has high reliability and accuracy. The predictive model of HC.k2 has the highest accuracy and recall, and the predictive model of HC.BDe has the highest specificity. Considering three indices of accuracy, recall, specificity, the predictive models using HC performs better than models using TABU.

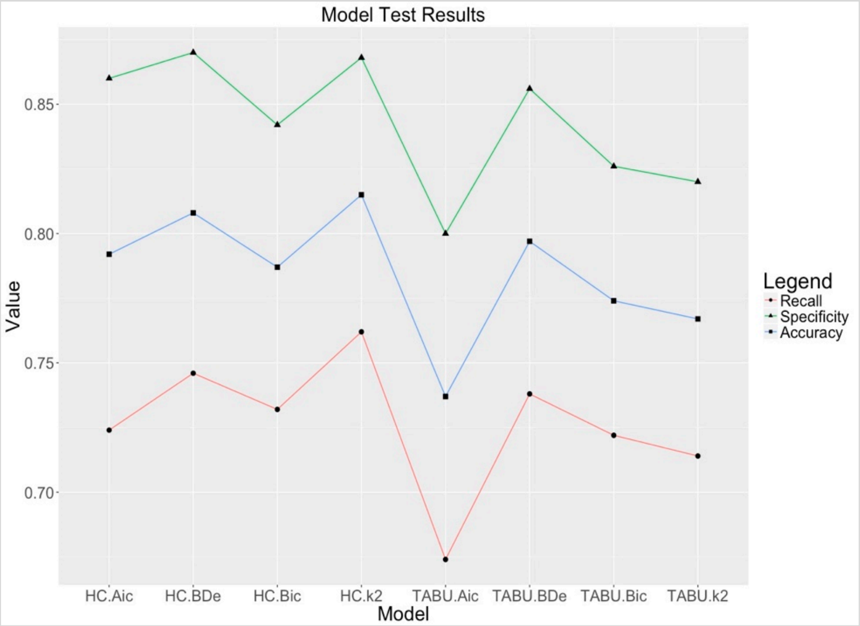


Figure 10. The results of the model testing.

4.4.4. Comparison of Classification Methods

In the phase of feature selection, we compared the classification effectiveness of One R, C5.0, Random Forests, Naive Bayesian and Bayesian network, and Bayesian network performed best in this case. When applying a predictive model for detecting the cargo theft in practice, comprehensibility and interpretability of the applied techniques is important, because it helps the port workers and the port authority to identify risky deliveries faster since they know exactly which features to look for. Therefore, we compare our proposed technique to the commonly known and widely used classification technique, i.e., logistic regression which is known for being both intuitive and easily interpretable, showing clear indication of how the features affect the prediction. As the dataset we used in this case is imbalanced, i.e., that bulk cargo theft has a much smaller sample size than normal operations, the ROC curve is utilized to illustrate the performance of the binary classification. Noticeably, Bayesian network performed better than logistic regression. The results can be seen in Figure 11, which shows values at the tangent point of ROC curve, indicating the higher sensitivity and specificity of Bayesian network.

The AUC of Bayesian network, at 0.834, is higher than the AUC of logistic regression, with 0.771. This result also indicates that Bayesian network performs better for this binary classification.

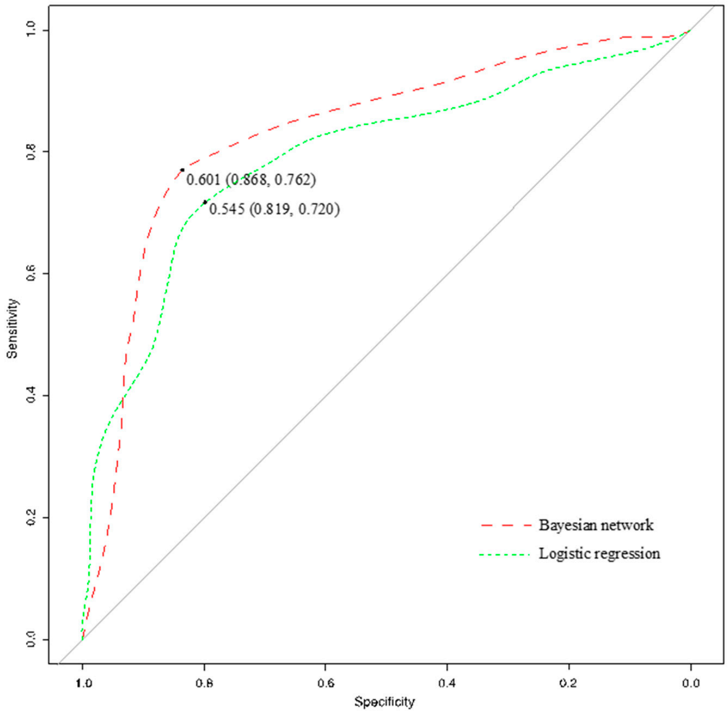


Figure 11. The comparison results of Bayesian network and logistic regression using ROC curve as the performance measure.

5. Discussion

In this empirical study, we considered the risk prediction of bulk cargo theft as a binary classification problem. In a real-world setting it is meaningful to focus on the fraction of risky instances that are correctly predicted as being risky. Therefore, we emphasize that recall is important when evaluating the effectiveness of our models. According to our results, the predictive model of bulk cargo theft risk using HC, k2 and Bayesian posterior probability estimation performs best.

Traditional approaches of risk management are mainly based on expert knowledge or data from interviews and surveys, which may lead to subjectivity and uncertainty. The approach we propose in this paper uses Bayesian network to construct predictive models from historical data of inspection records and operational data, which enables us to reduce the risk prediction deviation while also being less dependent on expert knowledge. However, we do exploit the knowledge of experts in order to refine the model learned from data and to add countermeasures.

The validation and testing results with indices of accuracy, recall, specificity all above 80% indicate that the predictive models are capable of effectively predicting the theft risk of bulk cargo in port. In particular, the optimal predictive model is selected using a comparative analysis considering a real-world scenario. This predictive model can be utilized for organizations of interest to pre-control the theft risk and narrow down the list of suspect trucks.

In addition, the predictive model based on Bayesian network provides a visually understandable representation of the cargo theft risk, with the relations between relevant elements involved. This makes it possible to translate and exploit the interdependencies and interrelationships among the large number of variables. Rather than the 'black box' characteristic of some classification methods, Bayesian network provides better interpretability, which is essential for management optimization. The network structure is intuitive for understanding the theft risk of bulk cargo, and is useful to help decision makers optimize the strategy of risk management. For instance, this predictive model of bulk cargo theft is not only useful for the Port Public Security Bureau to optimize the sampling inspection tactic, but also useful for the port company to optimize the leave-port check in this case. With the support of the predictive model, the cost of time and people to identify and prevent the cargo theft risk would be significantly reduced.

6. Conclusions and Future Perspectives

In this paper, we presented a novel data-driven approach to predict the theft risk of bulk cargo in port. First, key features of theft risk were selected using combinations of feature ranking methods and classification algorithms. Then Bayesian networks were constructed from the event data of bulk cargo theft extracted from sampling inspection records of delivery trucks and the operational records in logistics information systems of ports. Furthermore, we evaluated and comparatively analyzed these predictive models derived using various structure learning algorithms and parameter learning algorithms considering real-world scenarios. Finally, domain knowledge and expert expertise were explored to refine the model and add countermeasures. Potential future work directions are: (a) Test and validate the predictive model on multiple years of data. (b) Refine the predictive model via expert interview and practical application. (c) Include additional features in the model to improve the accuracy and the effectiveness, e.g., currently, we have not considered the communication of driver of delivery trucks, while these theft risk actually spread through social contact among truck drivers.

Author Contributions: Formal analysis, R.S., W.C. and M.Ó.; Methodology, R.S. and W.C.; Project administration, L.H.; Supervision, J.V.; Writing—original draft, R.S. and W.C.; Writing—review & editing, L.H., W.C., M.Ó. and J.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Appendix A

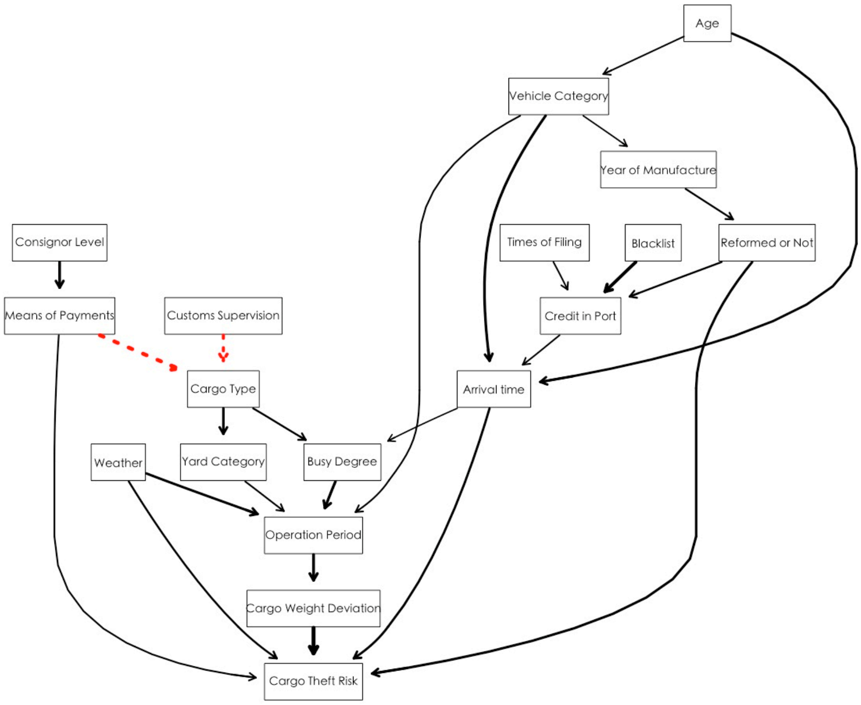


Figure A1. The Bayesian network structure derived using combinations of heuristic structure search algorithm with HC and score function with Aic.

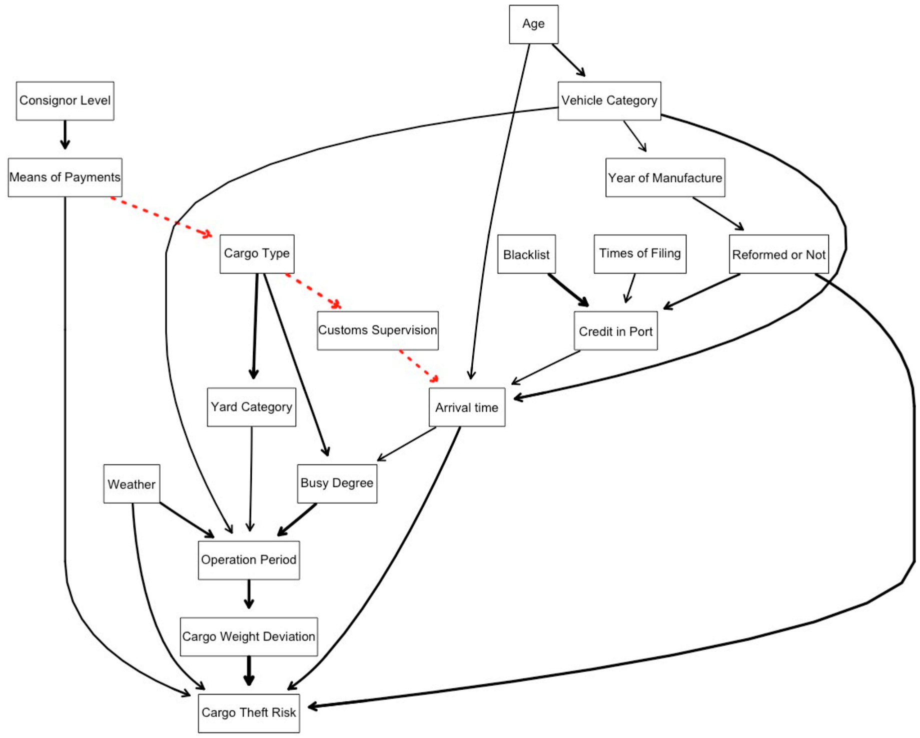


Figure A2. The Bayesian network structure derived using combinations of heuristic structure search algorithm with HC and score function with Bic.

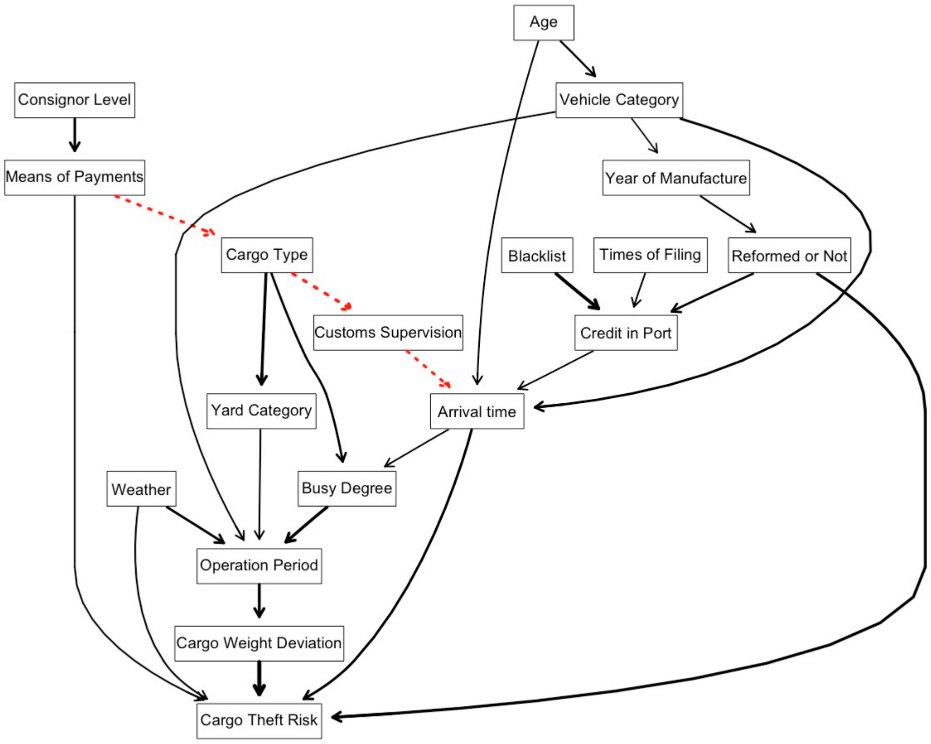


Figure A3. The Bayesian network structure derived using combinations of heuristic structure search algorithm with HC and score function with BDe.

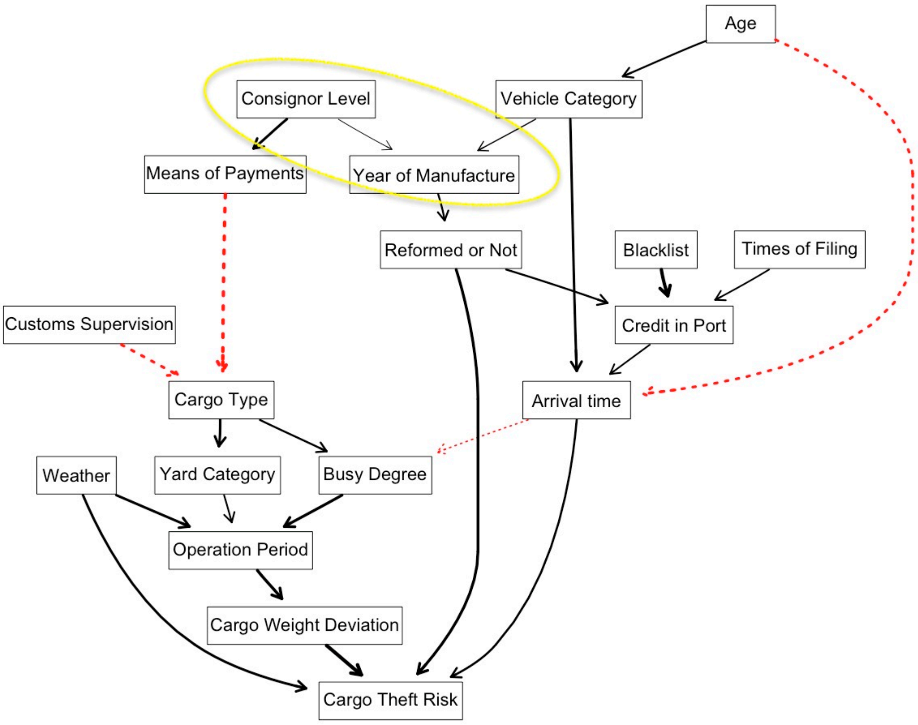


Figure A4. The Bayesian network structure derived using combinations of heuristic structure search algorithm with TABU and score function with Aic.

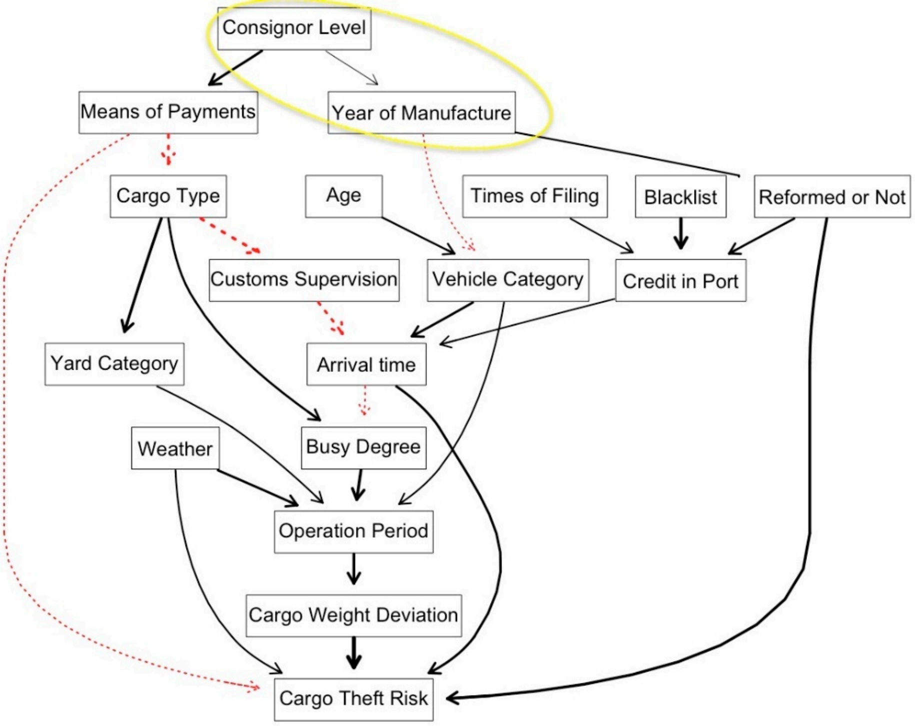


Figure A5. The Bayesian network structure derived using combinations of heuristic structure search algorithm with TABU and score function with Bic.

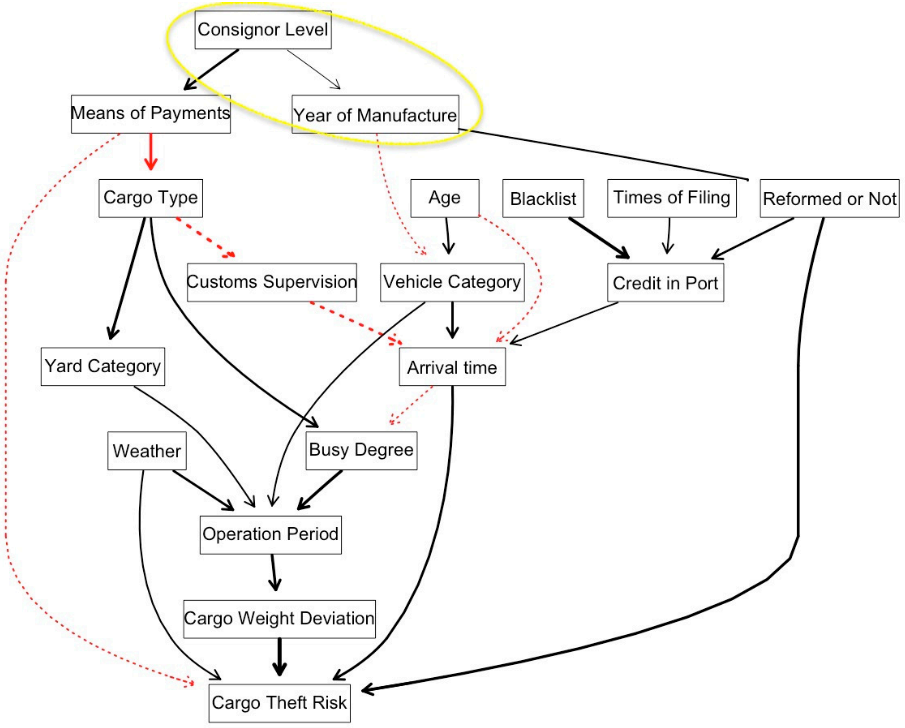


Figure A6. The Bayesian network structure derived using combinations of heuristic structure search algorithm with TABU and score function with BDe.

References

1. Burges, D. Cargo theft, loss prevention, and supply chain security. In *Cargo Theft Loss Prevention & Supply Chain Security*; Butterworth-Heinemann: Oxford, UK, 2012; Volume 79, pp. 267–269.
2. Guangzhou Port Police Authority. Cracking Down on Logistic Crimes, China Police Daily. Available online: http://epaper.cpd.com.cn/szb.html?t=jtzk&d=20180227&p=t (accessed on 10 June 2015).
3. Yang, Y.; Zhong, M.; Yao, H.; Yu, F.; Fu, X.; Postolache, O. Internet of things for smart ports: Technologies and challenges. *IEEE Instrum. Meas. Mag.* **2018**, *21*, 34–43. [CrossRef]
4. Fosso Wamba, S.; Gunasekaran, A.; Papadopoulos, T.; Ngai, E. Big data analytics in logistics and supply chain management. *Int. J. Logist. Manag.* **2018**, *29*, 478–484. [CrossRef]
5. Mincuzzi, N.; Falsafi, M.; Modoni, G.E.; Sacco, M.; Fornasiero, R. Managing Logistics in Collaborative Manufacturing: The Integration Services for an Automotive Application. In *Working Conference on Virtual Enterprises*; Springer: Berlin, Germany, 2019; pp. 355–362.
6. Haugstetter, H.; Cahoon, S. Strategic intent: Guiding port authorities to their new world? *Res. Transp. Econ.* **2010**, *27*, 30–36. [CrossRef]
7. Mokhtari, K.; Ren, J.; Roberts, C.; Wang, J. Application of a generic bow-tie based risk analysis framework on risk management of sea ports and offshore terminals. *J. Hazard. Mater.* **2011**, *192*, 465–475. [CrossRef] [PubMed]
8. Chen, H.; Chiang, R.H.L.; Storey, V.C. Business intelligence and analytics: From big data to big impact. *MIS Q.* **2012**, *36*, 1165–1188. [CrossRef]
9. Wang, G.; Gunasekaran, A.; Ngai, E.W.T.; Papadopoulos, T. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *Int. J. Prod. Econ.* **2016**, *176*, 98–110. [CrossRef]
10. Chlomoudis, C.I.; Pallis, P.L.; Tzannatos, E.S. A Risk Assessment Methodology in Container Terminals: The Case Study of the Port Container Terminal of Thessalonica, Greece. *J. Traffic Transp. Eng.* **2016**, *4*, 251–258.
11. Hui, S.L.; Thai, V.V.; Wong, Y.D.; Yuen, K.F.; Zhou, Q. Portfolio of port-centric supply chain disruption threats. *Int. J. Logist. Manag.* **2017**, *28*, 1368–1386.

12. Toth, G.E. CargoTIPS: An innovative approach to combating cargo theft. *Proc. SPIE Int. Soc. Opt. Eng.* **1998**, *3575*, 315–319.
13. Nielsen, T.D.; Jensen, F.V. *Bayesian Networks and Decision Graphs*, 2nd ed.; Springer Science & Business Media: Berlin, Germany, 2009.
14. Khakzad, N.; Khan, F.; Amyotte, P. Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 925–932. [CrossRef]
15. Hänninen, M.; Valdez Banda, O.A.; Kujala, P. Bayesian network model of maritime safety management. *Expert Syst. Appl.* **2014**, *41*, 7837–7846. [CrossRef]
16. Bouejla, A.; Chaze, X.; Guarnieri, F.; Napoli, A. A Bayesian network to manage risks of maritime piracy against offshore oil fields. *Saf. Sci.* **2014**, *68*, 222–230. [CrossRef]
17. Sutrisnowati, R.A.; Bae, H.; Song, M. Bayesian network construction from event log for lateness analysis in port logistics. *Comput. Ind. Eng.* **2015**, *89*, 53–66. [CrossRef]
18. Zhang, J.; Teixeira, Â.P.; Soares, C.G.; Yan, X.; Liu, K. Maritime Transportation Risk Assessment of Tianjin Port with Bayesian Belief Networks. *Risk Anal.* **2016**, *36*, 1171. [CrossRef]
19. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
20. de Sá, C.R.; Soares, C.; Knobbe, A. Entropy-based discretization methods for ranking data. *Inf. Sci.* **2016**, *329*, 921–936. [CrossRef]
21. Chandrashekar, G.; Sahin, F. A survey on feature selection method. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]
22. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature Selection: A Data Perspective. *ACM Comput. Surv.* **2017**, *50*, 1–45. [CrossRef]
23. Tang, J.; Alelyani, S.; Liu, H. Feature selection for classification: A review. In *Data Classification: Algorithms and Applications*; CRC Press: Boca Raton, FL, USA, 2014; p. 37.
24. Novaković, J. Toward optimal feature selection using ranking methods and classification algorithms. *Yugosl. J. Oper. Res.* **2016**, *21*, 119–135. [CrossRef]
25. Yang, Y.; Pedersen, J.O. A comparative study on feature selection in text categorization. *ICML* **1997**, *97*, 412–420.
26. Abe, N.; Kudo, M. Entropy criterion for classifier-independent feature selection. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 689–695.
27. Bagherzadeh-Khiabani, F.; Ramezankhani, A.; Azizi, F.; Hadaegh, F.; Steyerberg, E.W.; Khalili, D. A tutorial on variable selection for clinical prediction models: Feature selection methods in data mining could improve the results. *J. Clin. Epidemiol.* **2016**, *71*, 76–85. [CrossRef] [PubMed]
28. Ang, J.C.; Mirzal, A.; Haron, H.; Hamed, H.N.A. Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE ACM Trans. Comput. Biol. Bioinform.* **2015**, *13*, 971–989. [CrossRef] [PubMed]
29. Mukherjee, S.; Sharma, N. Intrusion Detection using Naive Bayes Classifier with Feature Reduction. *Procedia Technol.* **2012**, *4*, 119–128. [CrossRef]
30. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, Informedness, Markedness and Correlation. 2011. Available online: https://dspace2.flinders.edu.au/xmlui/handle/2328/27165 (accessed on 2 January 2020).
31. Scutari, M. Learning Bayesian Networks with the bnlearn R Package. *J. Stat. Softw.* **2010**, *35*, 1–22. [CrossRef]