e-Navi

Available online at http://www.e-navi.kr/
e-Navigation Journal

Original article

Influencing factors analysis and prediction of truck appointment no-shows in the container terminal with Truck Appointment System: A data-driven approach

Mengzhi MA [a], Xianglong LI [b], Jian SUN [c*]

[a] Transportation Engineering College, Dalian Maritime University, China, mengzhi1059@dlmu.edu.cn

[b] Transportation Engineering College, Dalian Maritime University, China, lixianglong2022@163.com

[c*] Transportation Engineering College, Dalian Maritime University, China, dreamsj@dlmu.edu.cn, Corresponding Author

Abstract

Truck no-show behavior has posed significant disruptions to the planning and execution of port operations. By delving into the key factors that contribute to truck appointment no-shows and proactively predicting such behavior, it becomes possible to make preemptive adjustments to port operation plans, thereby enhancing overall operational efficiency. Considering the data imbalance and the impact of accuracy for each decision tree on the performance of the random forest model, a model based on the Borderline Synthetic Minority Over-Sampling Technique and Weighted Random Forest (BSMOTE-WRF) is proposed to predict truck appointment no-shows and explore the relationship between truck appointment no-shows and factors such as weather conditions, appointment time slot, the number of truck appointments, and traffic conditions. In order to illustrate the effectiveness of the proposed model, the experiments were conducted with the available dataset from the Tianjin Port Second Container Terminal. It is demonstrated that the prediction accuracy of BSMOTE-WRF model is improved by 4%-5% compared with logistic regression, random forest, and support vector machines. Importance ranking of factors affecting truck no-show indicate that (1) The number of truck appointments during specific time slots have the highest impact on truck no-show behavior, and the congestion coefficient has the second-highest impact on truck no-show behavior and its influence is also significant; (2) Compared to the number of truck appointments and congestion coefficient, the impact of severe weather on truck no-show behavior is relatively low, but it still has some influence; (3) Although the impact of appointment time slots is lower than other influencing factors, the influence of specific time slots on truck no-show behavior should not be overlooked. The BSMOTE-WRF model effectively analyzes the influencing factors and predicts truck no-show behavior in appointment-based systems.

Keywords: Container terminal; Prediction of truck appointment no-shows; Influencing factors; Borderline-SMOTE; Weighted random forest

https://doi.org/10.52820/j.enavi.2023.21.028

1. Introduction

With the continuous expansion of seaborn trade carried by container ships and the development of large-scale ships, the collection and distribution systems of container hub-ports are facing great challenges. Especially for those ports that rely primarily on road transportation as the main mode of collection and distribution, and the concentrated arrival of trucks at peak hours has aggravated the problems of terminal congestion, resource waste and increased greenhouse gas emissions of container trucks. To address the issues, ports such as Los Angeles-Long Beach Port, Tianjin Port, Shanghai Port, and Shenzhen Yantian Port have all implemented the Truck Appointment System (TAS) to manage truck arrivals at the port. The operating practice of these ports has demonstrated that TAS can effectively alleviate port congestion during peak hours and improve port resources utilization. However, occasional disruptions such as traffic congestion, inclement weather, and human factors can sometimes lead to truck appointments delay, i.e., the truck no-show. The truck no-shows significantly disrupt the execution of port operation plan. The thorough analysis of the predominant factors influencing truck no-shows and accurate prediction of such behavior not only facilitates proactive adjustments to port operation plans but also enhances the overall efficiency of port operations. Thus, addressing the challenge of truck appointment no-shows has emerged as a major concern for port operators when implementing TAS.

TAS is a system that utilizes modern information technology and network communication technology to establish an effective link between the port and trucking companies, and reasonably manages truck arrivals (Xu et al. 2020). Truck appointment process can be divided into two stages: static appointment and dynamic appointment. In the static appointment stage, the port operators publicly disclose the appointment quotas of each time slot in the information system. Then trucking companies can make appointments specifying the appointment time slot they plan to deliver or pick up containers. In the dynamic appointment stage, the terminal operator evaluates the queue time based on existing appointments in real time. If a long queue may form, the appointment request will be rejected; otherwise, it is accepted. The rejected trucking companies can resubmit their appointment requests until the system accepts the request and adds it to the existing appointments (Ding et al. 2021). TAS can effectively manage truck arrivals, which alleviates port congestion to a certain extent (Xu et al. 2020). In addition, the container terminal can obtain partial knowledge of operational requirements through TAS before the arrive of trucks. Container terminal can utilize truck appointment information to make efficient storage space allocation and yard cranes scheduling plan in advance (Ma et al. 2018). Truck appointment information can also help container terminal perform real-time optimization of container relocation and retrieving sequence (Chen et al. 2021). Previous studies have shown that even partial knowledge of operational requirements can effectively reduce the number of rehandling operations and improve the operational efficiency of the container terminal yard. However, truck no-shows will disrupt the operation plan and diminish operation efficiency of the container terminal. Proactive scheduling can create a robust baseline yard operational plan that safeguards against potential disruptions during operational plan execution. However, due to the lack of systematic analysis on factors influencing truck no-show behavior and the difficulty in predicting truck no-shows in advance, it becomes challenging to make proactive scheduling. Therefore, there is an urgent need to further explore the key factors influencing truck no-show behavior and utilize the monitoring data of these factors to accurately predict truck no-show behavior.

In daily life, there are many no-show events that may not receive widespread attention. For example, library readers who made appointments but didn't show up, football fans who bought tickets but didn't attend the match, passengers who purchased flight tickets but failed to board the flight on time, or patients who made appointments but didn't show up. These individual low-probability events can have a significant impact on service systems. Similarly, although it is a small probability event for the trucks, it is inevitable from the overall point of view. Once the truck no-show behavior happens, it often disrupts the scheduling plan of the container terminal and leads to a decrease in port resource utilization. However, by identifying the factors that contribute to truck no-shows and accurately predicting these behaviors, the impact of truck appointment no-shows can be minimized. Currently, researchers commonly use methods such as logistic regression (LR), decision tree (DT), and random forest (RF) to study no-show behavior in various industries and domains. Wu et al. (2018) combined a binary response model with DT to identify significant features that have an impact on no-show behavior, and these features were used as classification nodes to construct a decision tree and ultimately obtain the feature combination of no-show patients. Gong et al. (2014) employed χ^2 tests for single-factor analysis to identify meaningful factors, and then used LR model to perform a multivariate analysis on the factors influencing patient appointment no-shows. Su et al. (2019) utilized the Apriori algorithm, an association rule mining algorithm, to analyze the correlation between patient no-shows and various factors, followed by empirical analysis. Based on the in-depth analysis of the factors influencing no-show behavior, providing an accurate prediction of no-show behavior is also of significant value. Briggs et al. (2021) used a binary LR model to assess the likelihood of no-show.

Through a review of existing research both domestically and internationally, it has been found that in-depth analysis of the factors influencing truck no-show behavior and accurate prediction of truck no-show can effectively enhance the robustness of port operation plan. However, there is currently a lack of relevant research on these two issues. Therefore, we refer to the methods already used in the study of no-show behavior from other fields. This paper proposes a truck appointment no-show prediction model based on the Borderline Synthetic Minority Over-Sampling Technique and Weighted Random Forest (BSMOTE-WRF), aiming to explore the underlying relationships between weather conditions, appointment time slots, the number of truck appointments, traffic conditions, and the occurrence of truck no-show behavior. This paper aims to predict future truck appointment no-shows by utilizing the knowledge of these influencing factors and assist port authorities in proactively adjusting their original scheduling plans. And this will help mitigate the disruption caused by truck no-show behavior to container terminal production operations and enhance the resilience of port production operations.

The remainder of this paper is organized as follows. In Sect. 2, we review the related works. In Sect. 3, we briefly introduce sources of data and how they are processed, and then provide a detailed description of the proposed BSMOTE-WRF model. In Sect. 4, the prediction result of BSMOTE-WRF model is compared with three other models, and the model performance is evaluated based on criteria such as accuracy. In Sect. 5, based on the experimental results, we discuss the management insights to the container terminal based on the importance ranking of influencing factors. Finally, the conclusions are presented in Sect. 6.

2. Related works

This section reviews the related works from three aspects. Firstly, it discusses the methods of forecasting in container terminal scheduling. Secondly, the relevant research on classification models is introduced, with a focus on the application of RF model, highlighting the limitations of traditional RF model in the presence of imbalanced data. Finally, it provides an overview of the application of oversampling techniques in addressing imbalanced data.

2.1 The methods of forecasting in Container Terminal Scheduling

The research on predictive methods related to production scheduling in container terminals primarily utilizes quantitative forecasting methods. Quantitative forecasting can be categorized into two main types: those based on statistical approaches and those employing machine learning techniques. Commonly used statistical forecasting methods include time series analysis and causal analysis. Time series forecasting predicts future trends based on past time series data and includes methods like moving averages, exponential smoothing, and trend forecasting. Causal analysis forecasts future trends based on the cause-and-effect relationships of phenomena, employing methods such as regression analysis and input-output analysis. Zhao and Zhou (2018) combined seasonal ARIMA and vector autoregressive models to forecast port throughput. Lian (2019) used single and double exponential smoothing to fit vessel traffic trends and improve vessel traffic forecast accuracy. Compared to traditional methods, machine learning-based forecasting not only enhances efficiency and accuracy but also achieves relatively precise predictions when data trends are unclear. Major methods include decision trees, support vector machines, and random forests. Abreu et al. (2023) developed a ship dwell time prediction model using a decision tree-based algorithm. Additionally, Awah et al. (2021) employed random forest and multilayer perceptron models to forecast optimal container throughput for ports. In the area of truck appointment system, Ma (2019) focused on the process of exporting container assembly at ports and developed a prediction model for the arrival volume of trucks at container terminals based on adaptive boundary kernel estimation. Based on this, she researched the coordinated optimization of the shipping assembly time window and the scheduling of yard cranes. Sun et al. (2022) proposed a data-driven method for designing the booking quota system for external truck appointments at ports. They established a regression model between the number of external truck arrivals and their total turnover time to implement the quota design for the truck appointment system. However, there is a research gap in modeling truck appointment no-show behavior, which needs to be addressed in this paper.

Reviewing existing research reveals that a few studies have begun to apply data-driven approaches to the prediction of terminal operation scheduling issues. In the era of big data, terminal managers can easily obtain data such as the actual arrival times and reservation volumes from the truck appointment system, and engage in selection, processing, and analysis to extract valuable information. This provides a data foundation for analyzing truck no-show behavior. Therefore, it is worth considering the use of data-driven concepts to analyze the historical data of truck arrivals at terminals and predict future trends. By taking into account various factors in actual operations, it's possible to analyze future truck no-show behavior, providing important references for terminal planning or the formulation of future operational plans.

2.2 Classification prediction model

LR, DT, support vector machines (SVM), and RF are commonly used classification models for analyzing influencing factors and predicting binary outcomes of events. LR and DT can achieve desirable modeling effects when the frequency of the two categories of the dependent variable in the sample is equal. Nevertheless, LR and DT models exhibit limited capability in effectively harnessing the limited samples of the minority class. Consequently, these models become vulnerable to the impacts of data imbalance, thereby increasing the risk of overfitting when performing predictions. Within the context of imbalanced datasets, LR and DT models tend to produce prediction outcomes that are biased. To be specific, these models consistently demonstrate a tendency to underestimate the likelihood of event occurrences while overestimating the probabilities of event non-occurrences, consequently manifesting clear biases in the estimation of probabilities.

SVM is a prevalent machine learning algorithm utilized for classification and regression tasks. However, selecting suitable parameters for SVM poses a considerable challenge. Optimal parameter selection plays a pivotal role in determining the model's performance and generalization ability, necessitating a certain level of experience and experimentation. And the decision boundaries generated by SVM often possess intricate characteristics, hindering intuitive comprehension of the model's decision-making process. RF combines the concepts of Bagging ensemble learning theory and random subspace methods (Breiman L 2002), and is composed of a multitude of decision trees, with the final prediction result being determined by aggregating the votes of all decision trees computed using Eq. 1. The utilization of averaging in RF enables it to mitigate the issue of overfitting to a certain extent, consequently compared to LR and DT, RF exhibits superior predictive performance. Compared to SVM, RF demonstrates enhanced interpretability to some extent. It enables the examination of feature importance for each decision tree within the RF as well as the splitting rules for each feature, thereby facilitating an understanding of the model's decision-making process. Furthermore, RF exhibits robust capabilities in handling high-dimensional data and large-scale datasets. It proficiently manages a substantial number of features and samples while preserving accuracy.

$$H(X)=\underset{y=1,2,\ldots,C}{\arg max}\left\{\sum_{t=1}^{T} I\left(h_t(X)=y\right)\right\} \tag{1}$$

The notation is explained as follows. $h_t(X)$ represents the output of the t-th decision tree, X denotes the test set, C is the number of classes, and T represents the total number of decision trees. $I(\cdot)$ is the discriminant function, If the parameter $h_t(X)=y$ in the function is true, the function value is equal to 1; otherwise, it is equal to 0.

Therefore, when studying binary classification problems, many researchers choose to use the RF model. In conventional RF approaches, the training subsets for each decision tree are generated by randomly sampling from the training set with replacement. During the splitting process of each tree's nodes, a random subset of features is selected from the entire feature set, and the optimal splitting feature is determined for tree construction. However, under imbalanced data scenarios, decision trees with insufficient learning of class-specific features may exhibit subpar classification performance. The collective impact of erroneous votes from these trees can degrade the overall classification ability of the RF. To address this issue, Weighted Random Forest (WRF) has been proposed. WRF evaluates the classification effectiveness of decision trees during the training phase and assigns appropriate weights to each tree. During the voting stage of the decision trees, the introduction of weighted voting permits a more nuanced decision-making process, ultimately leading to enhanced overall classification performance. Furthermore, in the presence of imbalanced data, numerous researchers have employed oversampling techniques to address the issue of data imbalance, which subsequently enhances the classification capability of the model and boosts its predictive accuracy.

2.3 The application of oversampling in imbalanced datasets

Due to the existence of imbalanced datasets obtained from the terminal operating system, one approach could involve utilizing oversampling methods to preprocess the data and subsequently enhance the predictive performance of the models. However, currently, there is no literature that employs oversampling methods specifically for handling truck appointment scheduling data. Conversely, there is a significant body of literature that delves deeper into the study of oversampling methods for other issues.

Oversampling techniques are widely employed as a common approach to tackle class imbalance problems. The primary goal is to augment the quantity of samples from the minority class, thereby enhancing the predictive performance of the model towards the minority class. Oversampling techniques can be broadly categorized into two categories: synthetic data techniques and replication of samples techniques. Synthetic data techniques encompass several prominent approaches such as Synthetic Minority Over-sampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), Over-sampling Technique and Edited Nearest Neighbors (SMOTE-ENN), and Borderline Synthetic Minority Over-Sampling Technique (Borderline-SMOTE). And the replication of samples technique mainly includes Random Oversampling. Qing et al. (2022) introduced the ADASYN and local outlier factor algorithm as a solution for tackling the imbalance issue in tornado sample sets using radar data. Wang (2022) employed machine learning and the SMOTE-ENN technique to achieve precise prediction of postoperative facial nerve function recovery, effectively addressing the lack of auxiliary diagnosis methods in the field of acoustic neuroma treatment.

Among the oversampling techniques, borderline-SMOTE stands out as it not only addresses the issue of inadequate recognition of minority class samples resulting from random oversampling, but also effectively avoids the problem of generating an excessive number of samples in the minority class caused by the imbalanced nature of the SMOTE. Borderline-SMOTE successfully tackles the problem of class imbalance, thus improving the overall balance of the dataset. Furthermore, it places particular emphasis on samples that are susceptible to misclassification, thereby enhancing the classifier's performance on these samples. Moreover, it effectively resolves the challenge of producing redundant synthetic samples, consequently mitigating the risk of overfitting. As a result, the borderline-SMOTE is chosen in this paper to handle the imbalanced dataset, with the aim of enhancing the classification performance and predictive accuracy of subsequent models.

More and more scholars have integrated these techniques with other algorithms or models. This integration not only addresses the issue of imbalanced datasets, allowing predictive models to achieve superior learning and prediction capabilities, but also enhances the model's generalization capacity. Furthermore, it improves the predictive accuracy for minority class samples, consequently elevating the overall predictive accuracy of the model. Accordingly, this paper first selects the borderline-SMOTE and combines it with the proposed enhanced predictive model, thereby effectively addressing the research problem at hand in a more comprehensive manner.

3. Data and Model Construction

3.1 Data Collection and Processing

Through the terminal operating system of Tianjin Port Second Container Terminal, the actual arrival time of 14,244 truck appointments for 504 appointment time slots over a period of 42 days were extracted. If the actual arrival time is later than the reserved time slot, mark the truck appointments as a no-show. Then, we check whether there was any truck appointment no-show in each appointment time slot. The occurrence of truck appointment no-shows during a specific time slot is considered as the dependent variable Y . If there are truck appointment no-shows in that time slot, $Y=1$. If there are no truck appointment no-shows in that time slot, $Y=0$. Due to the lack of systematic analysis research on the factors influencing truck appointment no-show behavior, interviews were conducted with container terminal operators, truck drivers, and other relevant individuals. The interviewees generally believe that truck appointment no-show behavior is associated with factors such as weather conditions, appointment time slots, the number of truck appointments, and traffic conditions. Therefore, this paper selects weather conditions, appointment time slots, the number of truck appointments, and traffic conditions as the influencing factors. The weather conditions commonly include sunny, cloudy, rainy, foggy, and snowy, while severe weather like rainy, foggy and snowy can significantly affect driving safety. Therefore, weather conditions are converted into a binary variable by categorizing them as either "severe" or "not severe". To analyze the impact of different time slots on the dependent variable Y , appointment time slots were transformed into 12 corresponding binary variables. The detailed information for each influencing factor is shown in Table 1.

Table 1：Detailed information for each influencing factor

Factors	Variable names	Variable description	Data sources
Weather conditions	Severe weather	No=0；Yes=1	Weather website□Https://tianqi.2345.com/□
Appointment time slots	Time slot 1（0:00-2:00）	No=0；Yes=1	Terminal operating system
	Time slot 2（2:00-4:00）	No=0；Yes=1	
	Time slot 3（4:00-6:00）	No=0；Yes=1	
	Time slot 4（6:00-8:00）	No=0；Yes=1	
	Time slot 5（8:00-10:00）	No=0；Yes=1	
	Time slot 6（10:00-12:00）	No=0；Yes=1	
	Time slot 7（12:00-14:00）	No=0；Yes=1	
	Time slot 8（14:00-16:00）	No=0；Yes=1	
	Time slot 9（16:00-18:00）	No=0；Yes=1	
	Time slot 10（18:00-20:00）	No=0；Yes=1	
	Time slot 11（20:00-22:00）	No=0；Yes=1	
	Time slot 12（22:00-24:00）	No=0；Yes=1	
The number of truck appointments	The number of truck appointments	Integer variable	Terminal operating system
Traffic conditions	Congestion coefficient	Continuous variable	Baidu Maps Traffic and Transportation Big Data Platform □https://jiaotong.baidu.com/□

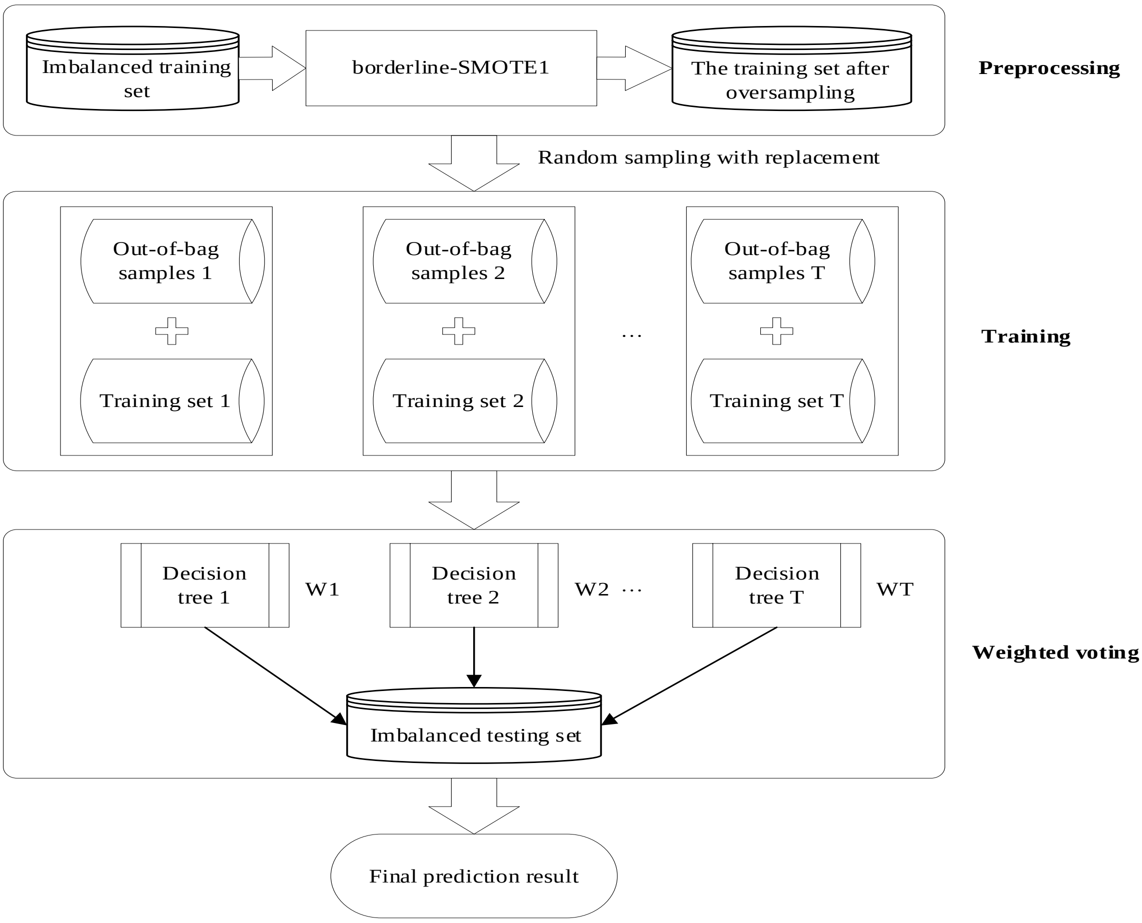

Figure 1：The BSMOTE-WRF model structure diagram

3.2 BSMOTE-WRF model

According to the analysis of historical data obtained from the terminal operating system, it is observed that the sample data suffers from a significant imbalance, with a majority of instances belonging to the category $Y = 0$ and a minority of instances belonging to the category $Y = 1$. This class imbalance issue poses a challenge to achieve high predictive accuracy with a RF model. To address this concern and improve the performance of the predictive model, this paper proposes a truck appointment no-show prediction model based on the BSMOTE-WRF. As mentioned earlier, borderline-SMOTE demonstrates exceptional performance within imbalanced training sets, offering a superior solution to the imbalance present in this dataset. Combining borderline-SMOTE with the proposed WRF in this paper resolves issues within the imbalanced dataset, enabling the predictive model to achieve superior learning and predictive capabilities. This integration strengthens the model's ability to generalize and further enhances predictive accuracy for minority class samples, ultimately elevating the overall predictive accuracy of the model. Therefore, the model proposed in this paper can effectively addresses the issues encountered when predicting truck appointment no-show behavior.

The BSMOTE-WRF model and TAS are two distinct concepts in the fields of data analysis and management. Their combined use can significantly enhance the operational efficiency and predictive accuracy of port management. Firstly, TAS relies on accurate predictions of variables such as truck arrivals to optimize scheduling and resource allocation. It needs to adapt to various factors, such as traffic conditions, unexpected no-shows, and operational restrictions. Secondly, the predictive insights of BSMOTE-WRF can enable dynamic scheduling, adjusting appointments in real-time based on predicted conditions. Through accurate predictions, ports can better allocate resources, such as cranes and personnel, thereby improving efficiency and reducing operational costs. Additionally, TAS can feed real-time data back into the BSMOTE-WRF model, allowing it to continuously learn and improve its predictions based on the latest patterns. In summary, integrating the BSMOTE-WRF model into TAS allows for an appropriate method to handle imbalanced data. WRF assigns different weights to categories or individual samples, making it more sensitive to certain instances and emphasizing minority classes, thus enhancing the model's sensitivity to these classes and improving the accuracy of predictions, thereby increasing the efficiency of port operations. This synergistic effect not only alleviates congestion and optimizes resource allocation but also adapts to constantly changing conditions, ensuring a robust and responsive management system.

The structure of the BSMOTE-WRF model is illustrated in Figure 1, and the construction steps of the model are as follows:

Step 1. Based on the original classification imbalanced training set, borderline-SMOTE1 is used to synthetic new samples to increase the number of samples with truck no-shows to obtain a balanced training set. Firstly, Suppose that the minority class is P, the majority class is N, and $P=\{p_1, p_2, ..., p_{pnum}\}$, $N=\{n_1, n_2, ..., n_{nnum}\}$. The K nearest neighbors method is employed to identify the K samples that are closest in distance to a minority class sample. Next, for all minority class samples, the borderline-SMOTE1 algorithm categorizes them into three categories: Safe, Danger and Noise, based on their distances to the K nearest neighbors. If the sample point x_i has m neighbors and among those m neighbors, m' sample points belong to the majority class, then

$$x_i \ \textit{is classified as} \begin{cases} \text{Noise}, & m' = m \\ \text{Danger}, & \frac{m}{2},, m' < m \\ \text{Safe}, & 0,, m' < \frac{m}{2} \end{cases} \tag{2}$$

Suppose that the Danger is D, the examples in D are the borderline data of the minority, and $D \subseteq P$. Suppose that $D=\{p'_1, p'_2, ..., p'_{dnum}\}$, $0 \le dnum \le pnum$. For each $p'_i (i=1,2,...,dnum)$ in D, we calculate its k nearest neighbors, and randomly select $s (1 \le s \le k)$ nearest neighbors from its k nearest neighbors in P. Then s new synthetic minority examples are generated between p'_i and its nearest neighbors. Finally, repeat the above procedure for each p'_i in D and can attain $s \times dnum$ synthetic examples, as shown in Eq. 3.

$$X_j = p'_i + r_j \times dif_j, \qquad j = 1,2,...,s \tag{3}$$

The notation is explained as follows. $X_j (j=1,2,...,s)$ are new synthetic examples, r_j is a random number between 0 and 1. dif_j is the differences between p'_i and its s nearest neighbors from P.

Step 2. From the training set obtained in Step 1, samples are randomly drawn with replacement to form the training set for each tree. The number of samples drawn is equal to the size of the oversampled training set. The remaining samples that are not drawn are referred to as out-of-bag samples (Baboo et al. 2018). These out-of-bag samples are used as the test set for each tree to evaluate the classification performance.

Step 3. At each node, f features are randomly selected to form a candidate feature set, where $f < F$ (F being the total number of features). The splitting ability of each candidate feature is evaluated based on the Gini coefficient, and the best feature is chosen for node splitting. Node splitting is then performed recursively on each child node until a complete decision tree is generated.

Step 4. The classification performance of each decision tree is evaluated using the out-of-bag samples, by calculating the Kappa coefficient (CK) according to Eq. 3. And Steps 2-4 are repeated until T decision trees are generated.

$$CK = \frac{ACC - CK_c}{1 - CK_c} \tag{4}$$

Where ACC represents the overall classification accuracy, and CK_c represents the chanced concordance rate for classification. ACC and CK_c (Canbek et al. 2017) are calculated using Eqs. 5-6, respectively.

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \quad (5)$$

$$CK_c = \frac{(TP + FN)(TP + FP) + (FP + TN)(FN + TN)}{(TP + FN + FP + TN)^2} \quad (6)$$

In the Equations, TP (True Positive) represents the number of samples that are 1 and predicted as 1. FN (False Negative) represents the number of samples that are 1 but predicted as 0. FP (False Positive) represents the number of samples that are 0 but predicted as 1. TN (True Negative) represents the number of samples that are 0 and predicted as 0.

Step 5. Based on the CK values obtained in step 4, the weight values for each decision tree can be calculated according to Eq. 7, ensuring that decision trees with higher CK values are assigned larger weights, thus exerting a greater influence on the final voting result.

$$w_t = \ln \frac{1 + CK_t}{1 - CK_t}, t = 1, 2, \cdots, T \quad (7)$$

Where w_t represents the weight value for the t-th decision tree, represents the CK value for the t-th decision tree.

Step 6. The final prediction result of the RF classifier is determined by calculating the weighted votes of all decision trees.

$$\hat{H}(X) = \underset{y=1,2,\cdots,C}{\arg max} \left\{ \sum_{t=1}^{T} \left(I\left(h_t(X) = y\right) \times w_t \right) \right\} \quad (8)$$

4. Result analysis

This section first compares the performance of the proposed model with other models, including the LR model, RF model, and SVM model, to validate the reliability of the BSMOTE-WRF model. Then, based on the average impurity reduction, the importance of influencing factors is ranked, laying the groundwork for the discussion in the next section. The results demonstrate that the BSMOTE-WRF model has strong classification capabilities and can significantly improve prediction accuracy. Therefore, utilizing the BSMOTE-WRF model to analyze various factors is the most appropriate approach.

4.1 Performance evaluation for BSMOTE-WRF model

4.1.1 Analysis of prediction results

The imbalanced dataset was randomly divided into a training set (75%) and a testing set (25%). The BSMOTE-WRF model was constructed using the oversampled training set. Additionally, LR model, RF model, and SVM model were constructed using the imbalanced training set. Subsequently, the model predictions were calculated using the imbalanced testing set, as shown in Table 2. In the given context, “D” represents the actual time periods with truck no-shows, i.e., actual value “$Y = 1$”; “~D” represents the actual time periods without truck no-shows, i.e., actual value “$Y = 0$”. “+” indicates the time periods predicted by the model as having truck no-shows, i.e., predicted value “$Y = 1$”; “-” indicates the time periods predicted by the model as not having truck no-shows, i.e., predicted value “$Y = 0$”. The correct and incorrect rates of the model predictions are shown in Table 3. All the experiments are implemented in python 3.10 and run on the same test machine with Intel Core i5 CPU computer at 3.0 GHz, 8.0 GB of RAM and Win 11 operating system.

Table 2: Comparison of prediction results of different models

		LR			RF			SVM			BSMOTE-WRF		
		Actual value		Total	Actual value		Total	Actual value		Total	Actual value		Total
Category		D	~D		D	~D		D	~D		D	~D	
Predicted value	+	**35**	14	49	**37**	16	53	**41**	21	62	**44**	19	63
	-	15	62	77	13	60	73	9	55	64	6	57	53
Total		50	76	126	50	76	126	50	76	126	50	76	126

Table 3: Comparison of the correct and incorrect prediction rates of different models

Category	Expression	LR	RF	SVM	BSMOTE-WRF
Recall	+ | D	70.00%	74.00%	82.00%	88.00%
Accuracy	—	76.98%	76.98%	76.19%	80.16%

Due to the maximum likelihood estimation method used in LR models, they possess certain characteristics such as consistency, asymptotic efficiency, and asymptotic normality. However, the LR model's prediction recall and accuracy for truck no-show are only 70.00% and 76.98%, which is not ideal. The LR model significantly overestimates the accuracy rate of samples without truck no-shows, while underestimating the accuracy rate of samples with truck no-shows. The classification accuracy of decision trees in RF may vary due to different training sets for each tree. Additionally, the imbalance in the data can further deteriorate the classification effectiveness of decision trees, thereby affecting the overall classification ability of RF. The conclusions drawn from the model's results are statistically credible only when the model's prediction accuracy is high.

The results presented in the table exhibit the superior performance of the BSMOTE-WRF model in comparison to the LR, RF, and SVM models. The BSMOTE-WRF model not only mitigates the biases inherent in the LR model but also enhances the classification capabilities of the RF and SVM models. Notably, the BSMOTE-WRF model achieves significantly higher prediction recall of 88.00% and accuracy of 80.16% for truck no-show behavior, surpassing the prediction recalls and accuracies of the LR model, RF model, and SVM model. This improvement in prediction accuracy indicates the substantial enhancement of the BSMOTE-WRF model in accurately predicting truck no-show behavior. The utilization of the borderline-SMOTE1 during the training phase allows subsequent classification algorithms to gain valuable information from minority class samples, thereby bolstering the classification accuracy of such samples. Consequently, the improved accuracy of minority class samples contributes to the overall enhancement in classification performance. Furthermore, the application of the WRF mitigates the influence of decision trees with low training accuracy on the overall model, thus further elevating the overall prediction accuracy.

4.1.2 Model performance evaluation

This paper evaluates the model using Accuracy, CK, and ROC curve. Accuracy represents the ratio of correctly predicted samples to the total number of samples, ranging from 0 to 1. A higher value indicates a more accurate model. The evaluation criteria for CK are as follows (Monserud and Leemans 1992): $CK \le 0.4$, Fail; $0.4 < CK \le 0.55$, Fair; $0.55 < CK \le 0.7$, Good; $0.7 < CK \le 0.85$, Very good; $CK \ge 0.85$, Excellent. ROC curve is commonly used to measure the performance of models in binary classification problems (Swets 1988), AUC (Area Under the Curve) is a quantitative metric of the ROC curve, and the evaluation criteria for AUC are as follows [0]: $AUC \le 0.5$, Fail; $0.5 \le AUC < 0.7$, Good; $0.7 \le AUC < 0.9$, Very good; $AUC \ge 0.9$, Excellent.

The performance of the LR model, RF model, SVM model, and BSMOTE-WRF model constructed based on the training set was evaluated using the test set. The results are shown in Table 4 and Figure 2.

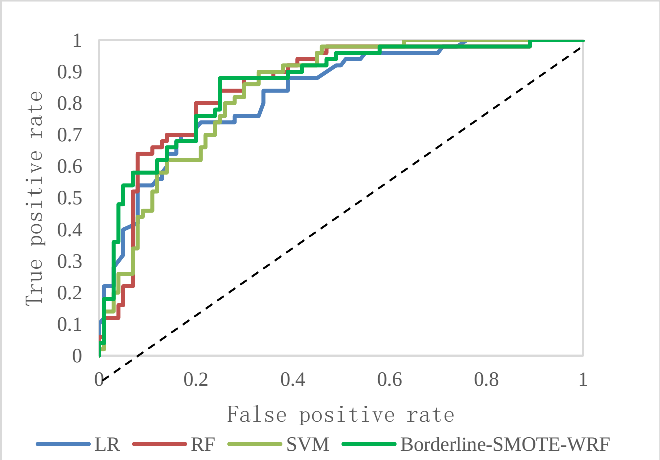

Figure 2: ROC curve plot

As shown in Table 4 and Figure 2, the BSMOTE-WRF model has an Accuracy of 80.16%, a CK of 0.603, and an AUC of 0.859. This indicates that the BSMOTE-WRF model has a higher prediction accuracy, good consistency between predicted and actual results, and good discrimination ability. Overall, the BSMOTE-WRF model outperforms the LR model, RF model, and SVM model in terms of Accuracy, CK, and AUC as comprehensive evaluation metrics.

Based on the above analysis, the degree of improvement of the BSMOTE-WRF model in terms of Accuracy, CK value and AUC compared to the LR, RF and SVM models is shown in Table 4, where Gap = (BSMOTE-WRF model metrics assessment result / other model metrics assessment result - 1)*100%.

Table 4: Calculation results of model evaluation metrics

Targets	LR		RF		SVM		The model of this paper
	Results	*Gap*	Results	*Gap*	Results	*Gap*	Results
Accuracy	76.98%	4%	76.98%	4%	76.19%	5%	80.16%
CK	0.518	16%	0.524	15%	0.522	16%	0.603
AUC	0.832	2%	0.856	1%	0.840	2%	0.859

4.2 Importance ranking of factors affecting truck no-show

The BSMOTE-WRF model demonstrates superior performance and accuracy in predicting the occurrence of truck appointment no-shows compared to other models, namely the LR model, RF model, and SVM model. It effectively captures the influence of various factors, including weather conditions, traffic conditions, appointment time slots, and the number of truck appointments, on truck no-show behavior. Hence, the explanations of these factors in this paper are grounded in the findings of the BSMOTE-WRF model. To identify the key factors that significantly impact truck no-show behavior, the relative importance of these factors is determined based on their average impurity reduction, as depicted in Figure 3.

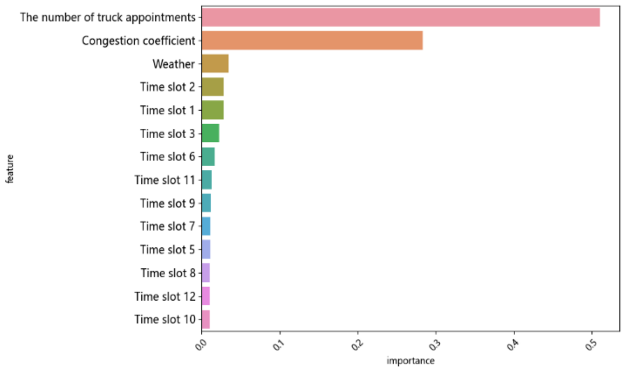

Figure 3: Ranking the importance of factors influencing truck no-show

The analysis reveals that two key factors, namely the number of truck appointments and congestion coefficient, exert a substantial influence on truck no-show occurrences, with importance scores of 0.517 and 0.305, respectively. In comparison, the impact of severe weather conditions such as rainy, foggy, and snowy is relatively minor, as indicated by an importance score of 0.030. Additionally, appointment time slots demonstrate a relatively weaker influence on truck no-show behavior. However, among all the time slots, slots 1, 2, 3, and 6 exhibit a relatively more significant impact compared to other time slots, with respective importance scores of 0.018, 0.023, 0.017, and 0.015.

5. Discussion

The importance ranking of influencing factors confirms that the factors mentioned by the surveyed individuals, such as container terminal operators and truck drivers, during the interviews, are indeed important in influencing truck no-show behavior. Based on the actual situation, the following conclusions can be drawn:

(1) The impact of the number of truck appointments on truck no-show behavior is significant during specific time slots. According to the predictive model, as the number of truck appointments increasing within a time slot, the likelihood of truck no-show occurrences also rises. The port’s truck fleet maintains a relatively consistent collection and distribution capacity over a certain period. However, when the number of appointments surpasses the transport capacity of the truck fleet during a time slot, it becomes challenging to ensure timely delivery of all containers to the port within

the reservation period. Moreover, the congestion caused by a large concentration of trucks arriving at the port amplifies traffic congestion in both the port area and collection and distribution routes, further increasing the probability of truck no-shows during the reservation time slot.

(2) The congestion coefficient is identified as the second most influential factor affecting truck no-show behavior, and its impact is substantial. A higher congestion coefficient signifies a greater degree of traffic congestion along the collection and distribution routes connecting to and from the port. During periods of congestion, the speed of truck travel is reduced, leading to queue formations and unpredictable transportation times. The consequence is the inability to guarantee the timely arrival of trucks at the port. Consequently, the probability of truck no-shows is escalated.

(3) In comparison to the number of truck appointments and congestion coefficient, the influence of severe weather conditions on truck no-show behavior is relatively moderate, yet still noteworthy. Severe weather conditions contribute to deteriorated road traffic conditions. Consequently, truck drivers often opt to reduce their driving speeds for safety concerns. Furthermore, the safety risks associated with traffic accidents significantly elevate during severe weather conditions in comparison to clear weather, leading to an increased likelihood of accidents. In the event of a traffic accident, congestion may occur for a certain duration, resulting in trucks being unable to guarantee timely arrival at the port and an amplified probability of no-show occurrences.

(4) Although the impact of appointment time slots on truck no-show behavior may be relatively lower compared to other influencing factors, the significance of specific time slots should not be disregarded. Time slots 1 and 2 exhibit a higher propensity for truck no-show occurrences primarily due to some truck drivers being in a resting state during these periods, resulting in a reduced supply of transport capacity compared to other time slots. Additionally, the efficiency of nighttime operations at the inland container depot is diminished and the unpredictable operation times, leading to a decline in truck transport organization efficiency. Time slot 3 is susceptible to truck no-shows due to the prevalence of heavy fog and frost during this time slot, which gives rise to severe weather-related no-show incidents. Moreover, containers can only be transported from the inland container depot to the container terminal after the completion of customs clearance procedures. As a result, there is a small surge in port collection activities following the opening of the customs office (9:00). This surge leads to a higher number of truck appointments during time slot 6, rendering it more vulnerable to truck no-show events.

In conclusion, the model proposed in this paper can effectively analyze the influencing factors of truck appointment no-shows, providing a theoretical basis for and justification of container terminal appointment scheduling. In practical applications, based on various information obtained from weather forecasts, traffic big data platforms, and container terminal appointment systems, the model can be utilized for reasonable predictions. By adjusting the original scheduling plans based on the predicted results in advance, the impact of truck appointment no-shows can be reduced, ensuring the normal operation of the port.

6. Conclusion

In order to address the limitations of traditional RF models that exhibit subpar classification performance in the presence of data imbalance, this paper proposes a truck appointment no-show prediction model based on the BSMOTE-WRF. It is specifically developed for the prediction and analysis of truck no-show behavior. The model is evaluated using the data from terminal operating system, traffic congestion coefficients, and weather information. The findings demonstrate that:

(1) In terms of the comprehensive evaluation of Accuracy, CK, and AUC, the BSMOTE-WRF model demonstrates a substantial improvement in performance. When compared to the LR, RF, and SVM models, it is demonstrated that the Accuracy of BSMOTE-WRF model is improved by 4%-5%, the CK of BSMOTE-WRF model is improved by 15%-16%, and the AUC of BSMOTE-WRF model is improved by 1%-2%.

(2) The number of truck appointments, congestion coefficient, severe weather conditions, and appointment time slots all have an impact on truck no-show.

To effectively leverage the proposed method for predicting truck no-show behavior and the analysis results of influencing factors, and to bolster the port’s resilience in tackling uncertainties and enhancing operational efficiency, the following recommendations are proposed:

(1) Real-time monitoring of forecast data for key influencing factors and effective prediction of truck no-show behavior is crucial. One significant factor contributing to truck no-shows is traffic congestion. Therefore, ports should proactively monitor real-time traffic conditions and obtain traffic congestion coefficients from reliable platforms such as Baidu Map Traffic Travel Big Data Platform and Amap Open Platform. In the event of traffic congestion, prompt actions should be taken through scientifically designed online appointment systems to minimize losses and mitigate traffic-related consequences. By leveraging information from weather forecasts and truck appointment system, accurate predictions can be generated using appropriate models. These predictions can then be used to adjust the original scheduling plan in advance, thus reducing the impact of no-shows and ensuring the seamless operation of the port.

(2) The allocation of the number of appointment quotas based on truck no-show predictions is crucial for effective management. Consideration of the number of appointments is essential when establishing the number of appointment quotas to minimize the occurrence of no-show behavior. Additionally, to mitigate wastage of yard operation capacity, it is recommended to allocate additional flexible appointment quotas for subsequent time slots during periods where no-show incidents are likely to occur.

(3) The implementation of penalty or reward measures can serve as effective incentives to ensure timely fulfillment of truck appointments. Management can consider implementing such measures to encourage truck drivers to adhere to their scheduled appointments. For instance, penalties in the form of additional fees can be imposed for instances of no-show behavior, while drivers who consistently fulfill their appointments on time can be rewarded with priority services or other forms of incentives.

(4) The implementation of flexible management practices and the provision of effective truck reminder services can contribute to the improvement of appointment fulfillment rates and the mitigation of no-show incidents in port operations. By leveraging automated messages and other communication channels, timely reminders should be dispatched following truck appointment confirmation and a predetermined time slot before the scheduled arrival at the port. These reminders should encompass comprehensive information pertaining to rescheduling options, appointment cancellation procedures, and protocols to address instances of no-shows. In the event of truck no-show, prompt adjustments to strategies can be made to minimize any adverse effects on port operations. Predictive services can also be harnessed to proactively send messages to scheduled trucks, facilitating early preparation and potential departure ahead of schedule to ensure punctual arrival.

However, it is important to acknowledge the limitations of this paper. Firstly, the existing research literature on the factors influencing truck no-show behavior is relatively scarce, leading to a limited understanding of these factors. As a result, the regression and prediction outcomes may not fully capture the complexities and nuances of this phenomenon. Secondly, while this paper provides a brief analysis of the predictive results for truck no-shows, it falls short of fully utilizing these findings in practical applications. It is necessary to conduct further research considering more influencing factors, such as the behavior of shippers, as well as the differences between weekends and weekdays, among others, to perform a comprehensive analysis of the factors leading to truck no-show behavior. These future studies would contribute to a more robust understanding of truck appointment management and enhance the practical implementation of predictive methods in mitigating no-show incidents.

Acknowledgements

All authors would like to sincerely express gratitude to editor and anonymous reviewers for their valuable comments and insightful suggestions that have helped us to improve this paper. In addition, this research was funded by the Social Science Fund Planning Key Project of Liaoning Province, grant number L20AGL017.

References

Xu, B. G., Liu, X. Y., Yang, Y. S. and Li, J. J.(2020), Multi-constraint scheduling model and algorithm for truck appointment system considering the influence of morning and evening peak periods, Computer Integrated Manufacturing Systems, 26(10):2851-2863

Ding, Y. and Zhang, C. C.(2021), External trucks' appointment optimization of delivering containers from outside at container terminal, Computer Engineering and Applications, 57(19):267-273.

Ma, M. Z., Fan, H., Ji, M. J. and Guo, Z. F.(2018), Integrated optimization of truck appointment for export containers and crane deployment in a container terminal, Journal of Transportation Systems Engineering and Information Technology, 18(03):202-209.

Li, N., Chen, G., Ng, M., et al.(2020), Optimized appointment scheduling for export container deliveries at marine terminals, Maritime Policy & Management, 47(4):456-478.

Fan, H., Peng, W., Ma, M., et al.(2021), Storage space allocation and twin automated stacking cranes scheduling in automated container terminals, IEEE Transactions on Intelligent Transportation Systems, 23(9):14336-14348

Chen, S. M. and Zeng, Q. C.(2021), Optimization of Relocation for import containers in container terminals considering real-time pickup sequence adjustment, Industrial Engineering and Management, 26(06):138-145.

Feng, Y. J., Zeng, Q. C., Li, H. Y. and Yang, Z. Z.(2017), Method for integrated Optimizing of rehandling and retrieving sequence for inbound containers in container terminals, Operations Research and Management Science, 26(05):110-118.

Li, N., Chen, G., Govindan, K., et al.(2018), Disruption management for truck appointment system at a container terminal: A green initiative, Transportation Research Part D: Transport and Environment, 61:261-273.

Wu, Y., Xu, C. J., Cheng, Z. T., et al.(2018), The application of binary response models and decision trees in the study of outpatient appointment no-show behavior, Chinese Hospital Management, 38(10):36-38.

Gong, Y. N., Lu, Y., Lu, C. and Wang, F. S.(2014), Analysis of influencing factors for bidirectional missed appointments in a grade 3 hospital from Xinjiang, Chinese Journal of Health Statistics, 31(06):961-964.

Su, W., Zhu, C. L., Zhang, X. and Xu, F.(2019), Research on factors of missed online appointment for out-patient based on the association rules algorithm, China Journal of Information systems, (02):40-52.

Briggs, M. S., Ulses, C., VanEtten, L., et al.(2021), Predictive Factors for Patients' Failure to Show for Initial Outpatient Physical Therapist Evaluation, Physical Therapy, 101(5):1-8.

Azhar, N. A., Pozi, M. S. M., Din, A. M., et al.(2022), An investigation of SMOTE based methods for imbalanced datasets with data complexity analysis, IEEE, Transactions on Knowledge and Data Engineering.

Qing, Z., Zeng, Q., Wang, H., et al.(2022), ADASYN-LOF Algorithm for Imbalanced Tornado Samples, Atmosphere, 13(4):544.

Wang, J.(2022), Prediction of postoperative recovery in patients with acoustic neuroma using machine learning and SMOTE-ENN techniques, Mathematical Biosciences and Engineering: MBE, 19(10):10407-10423.

Song, L., Xu, Y., Wang, M., et al.(2021), PreCar_Deep: A deep learning framework for prediction of protein carbonylation sites based on Borderline-SMOTE strategy, Chemometrics and Intelligent Laboratory Systems, 218: 10442.

Zhao, S. W., Zhou J. H.(2018), Forecasting China's Port Container Throughput: Based on Combined Time Series, Journal of Systems Science & Mathematical Sciences, 38(02): 210-219.

Lian, J.(2019), Port Ship Traffic Forecast Based on Exponential Smoothing Technique, Ship Science and Technology, 41(24): 202-204.

Abreu, L.R., Maciel, I.S., Alves, J.S., Braga, L.C., and Pontes, H.L. (2023). A decision tree model for the prediction of the stay time of ships in Brazilian ports, Eng. Appl. Artif. Intell., 117, 105634.

Awah, P.C., Nam, H., and Kim, S. (2021). Short Term Forecast of Container Throughput: New Variables Application for the Port of Douala, Journal of Marine Science and Engineering.

Ma, M. Z.(2019), Optimization of Container Terminal Assembly Operations Based on Truck Arrival Volume Prediction, Dalian: Dalian Maritime University.

Sun, S. C., D, Yao., Zheng Y.(2022), A Data-Driven Method for Designing the Booking Quota System for External Truck Appointments at Ports, Journal of Dalian Maritime University, 48(03): 39-45.

Chaabane, I., Guermazi, R., Hammami, M.(2020), Enhancing techniques for learning decision trees from imbalanced data, Advances in Data Analysis and Classification, 14:677-745.

Tsai, C. A. and Chang, Y. J.(2023), Efficient Selection of Gaussian Kernel SVM Parameters for Imbalanced Data. Genes, 14(3):583.

Breiman, L.(2001), Random forests, Machine learning, 45:5-32.

Chawla, N. V., Bowyer, K. W., Hall, L. O., et al.(2002), SMOTE: synthetic minority over-sampling technique, Journal of artificial intelligence research, 16:321-357.

Han, H., Wang, W. Y., Mao, B. H.(2005), Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning, In: International conference on intelligent computing, pp 878-887.

Gajowniczek, K., Grzegorczyk, I., Ząbkowski, T., et al.(2020), Weighted random forests to improve arrhythmia classification, Electronics, 9(1):99.

Baboo, S. S., Iyyapparaj, E.(2018), A classification and analysis of pulmonary nodules in CT images using random forest, In: International Conference on Inventive Systems and Control (ICISC), IEEE, pp 1226-1232 .

Canbek, G., Sagiroglu, S., Temizel, T., et al.(2017), Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights, In: International Conference on Computer Science and Engineering (UBMK), IEEE, pp 821-826.

Monserud, RA., Leemans, R.(1992), Comparing global vegetation maps with the Kappa statistic, Ecological Modelling, 62(4):275-293.

Zhou, C., Fang, X. Q., Wu, X. J., Wang, Y. C.(2019), Risk assessment of mountain torrents based on three machine learning algorithms, Geo-Information Science, 21(11):1679-1688.

Swets, J. A.(1988), Measuring the accuracy of diagnostic systems, Science, 240(4857):1285-1293.

Received	**04 December 2023**
1st Revised	**12 December 2023**
2nd Revised	**22 December 2023**
Accepted	**25 December 2023**