DOI: 10.15849/IJASCA.230720.07

Int. J. Advance Soft Compu. Appl, Vol. 15, No. 2, July 2023
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Prediction of Ship Departure Delay Using Supervised Learning

Sri Mumpuni Retnaningsih, Iis Dewi Ratih and Kharin Octavian Ranto

Business Statistics, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia
kharinoctavian77@gmail.com
Business Statistics, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia
mumpuni@statistika.its.ac.id
Business Statistics, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia
iisdewiratih@gmail.com

Abstract

The loading and unloading activities of a ship start from the activities of the ship berthing until the ship departs from the wharf. The arrival of ships to the wharf of a terminal has been scheduled or known as a window, but in reality there are still many delays in ship departure which result in an increase in the length of the ship berth so that it affects the low value of Box/Ship/Hour (BSH), and can result in the schedule of the next ship's schedule experiencing a delay berthing. This affects the income of a terminal and becomes the focus of attention for the terminal to be handled. This study was conducted to predict ship departure delay in the hope that it can help provide consideration to the terminal in compiling the next ship berthing schedule and the operational department can carry out anticipatory activities to prevent ship departure delay from ships that are predicted to be delayed through the allocation of loading and unloading facilities. The research was conducted by comparing several supervised learning methods, including the K-Nearest Neighbor (KNN), Naïve Bayes Classifier (NBC), NBC with Bagging Ensemble Classifier, Decision Tree, and Ordinal Logistic Regression. It was found that the best classification result is the KNN method (k = 5) with an accuracy value of 91%, but for the delay ≤ 4 hours and > 4 hours it has a sensitivity value ≤ 50%, it means that ships with a delay ≤ 4 hours and > 4 hours have have less accurate predictions, and the use of the bagging method on the NBC method is proven to improve the classification performance of the NBC method.

Keywords: *Bagging Ensemble Classifier, Classification, Decision Tree, Delay, K-Nearest Neighbor, Naïve Bayes Classifier, Ordinal Logistic Regression.*

Received 1 September 2022; Accepted 30 March 2023

1 Introduction

Loading and unloading activities start from the ship berthing until the ship departs from the wharf, which is called berthing time. The arrival of ships to the wharf of a terminal has been scheduled or known as a window, namely the cooperation of wharf services between the terminal and shipping agents through the provision of ship berthing guarantees only for certain ships whose arrival schedule has been determined [1].

The window system makes the ship's arrival schedule neatly and compactly arranged, but in reality there are still many delays in ship departure from the predetermined schedule. The ship is said to experience a departure delay when the ship leaves the wharf more than the Estimated Time of Departure (ETD) which is the time when the ship must leave the wharf. This causes the next ship to not be able to berth on time or experience a berthing delay and result in a longer berthing time, thus affecting the low value of BSH (Box/Ship/Hour).

The higher the BSH value, the shorter the service time, which affects the tariffs to be paid. A high BSH value will attract shipping agents to berth their ships at the terminal. The BSH value that does not reach the target causes the satisfaction of shipping agents to decrease and causes the potential to move to another terminal with a higher BSH value. This causes losses for the terminal due to the loss of ships that should be able to be served. This condition can be handled by predicting the ship departure delay. This prediction can help the operational department to find out the length of time for ship departure delays so that anticipation can be made in the form of allocating loading and unloading facilities or as consideration for the preparation of the next ship berthing schedule.

Prediction of ship departure delay will be carried out using supervised learning methods, including KNN, NBC, and NBC with bagging ensemble, Decision Tree, and Ordinal Logistics Regression. These various methods will be compared based on the best classification performance in predicting ship departure delays. Supervised learning is a learning method that aims to predict the target variable of a new data (data testing) based on a model that has been trained using a labeled data set (training data) [2].

KNN is a classification method that is carried out quickly, is suitable for small dimensions and data, and is very well applied to data with multi-class targets [3]. NBC was chosen because this method does not require large training data, is able to handle quantitative data and discrete data. The bagging ensemble method needs to be done to increase the accuracy value of the NBC method. Decision trees have several advantages, namely having good accuracy and being able to find unexpected combinations of data [4]. Ordinal logistic regression is a regression analysis between response variables with ordinal data scale and predictor variables with categorical and continuous data types [5]. The bagging method is able to increase accuracy significantly greater than using only individual classification models, and is stronger against the effects of noise and overfitting of training data[6].

2 Related Work

A study by Ramos and Ismail entitled Analysis of the Causes of Delay in Loading and Unloading Activities at PT. Pelabuhan Indonesia I Dumai, stated that the cause of the delay in loading and unloading was the condition of the equipment and weather factors [7]. A study by Safrianda and Fatnanta entitled Analysis of the Causes of Delays in

Unloading Goods Due to Equipment Factors (Case Study: Wharf A Dumai Port), stated that the main factor causing loading and unloading delays was equipment damage [8]. Another study by Azis et al which aims to compare the performance of classification methods on multiclass datasets, get the results that the KNN method has the highest level of accuracy compared to the Support Vector Machine (SVM), Neural Network (NN), Random Forest Classifier (RFC), Ada Boost Classifier (ABC) and Quadratic Discriminant Analysis (QDC) [9]. Fitriyani stated that the NBC algorithm with the bagging method (NBC+BG) provides higher performance results than using the NBC algorithm alone, with it being proven that the bagging method can not only solve class imbalance problems, but also improves the predictive model of thoracic surgery [10]. Hana with the research title Classification of Diabetes Patients Using the Decision Tree C4.5 Algorithm stated that the Decision Tree has an accuracy of 97.12%, Precision is 93.02%, and Recall is 100.00% [11]. Nurmalasari et al with the research title Analysis of Factors Affecting Human Development Index (HDI) Using Ordinal Logistic Regression and Ordinal Probit Regression Methods (Case Study of Districts/Cities in Central Java in 2014) stated that ordinal logistic regression has an accuracy of 80% [12].

3 Problem Formulations or Methodology

3.1 Dataset

The data used was obtained from the Commercial and Business Development Department of PT XYZ with the research unit is international ships that did not experience berthing delay as many as 171 ships during the period January 1, 2021-December 31, 2021. The data will be divided into 2 parts randomly, namely training data and testing data with a proportion of 80% training data and 20% testing data. The research variables used in this study are described in Table 1.

Table 1: Research Variables

Variable	Description	Unit	Category
Y	Ship departure delay	-	0: No delay 1: Delay ≤ 4 hours 2: Delay > 4 hours
X_1	Ship type	-	0: Feeder 1: Direct
X_2	Wharf	-	1: Wharf 1 2: Wharf 2 3: Wharf 3 4: Wharf 4
X_3	Number of Container Crane (CC)	Unit	-
X_4	Number of discharge containers	Boxes	-
X_5	Number of loading containers	Boxes	-
X_6	Total breakdown time	Minute	-
X_7	Number of hatches	Unit	-
X_8	Number of load shifting in Container Yard (CY)	Boxes	-
X_9	Duration of Waiting Approval Agent (WAG)	Minute	-
X_{10}	Duration of bad weather	Minute	-

3.2 K-Nearest Neighbor

KNN is a lazy learning algorithm or known as instance-based learning, that is, there is no processing of the training data so there is no model creation process [13]. KNN is a method that determines categories based on the majority of categories in k-nearest neighbors [3]. The distance calculation is carried out between the testing data and the training data, where the training data is X_i ($i = 1, 2, \ldots, h$) where h is the amount of data in the training data and the testing data is X_j ($j = 1, 2, \ldots, m$) where m is the amount of data in the data testing, and the variable is symbolized by l ($l = 1, 2, \ldots, p$) where p is the number of variables. The distance calculation is carried out using the Euclidean distance described in Theorem 3.1.

Theorem 3.1 *Euclidean distance.*

$$d(X_i, X_j) = \sqrt{\sum_{l=1}^{p} (\text{diff}(X_{il}, X_{jl}))^2} \tag{1}$$

$\text{Diff}(X_{il}, X_{jl})$ is the distance value between training data and testing data, where the calculation of the distance value depends on the data type of the independent variable described in Table 2 [6].

Table 2: Distance Calculation Based on Data Type

Data Type	Distance Formula
Categorical	$\text{diff}(X_{il}, X_{jl}) = \begin{cases} 0 \text{ jika } X_{il} = X_{jl} \\ 1 \text{ jika } X_{il} \neq X_{jl} \end{cases}$
Numerical	$\text{diff}(X_{il}, X_{jl}) = \lvert X_{il} - X_{jl} \rvert$

3.3 Naive Bayes Classifier

NBC is a method by predicting the probability of future events based on past events. The theorem is combined with "naïve" which assumes conditions between attributes are independent, with the intention that a variable is not related to the presence or absence of other variables in the data [14]. The NBC formula is explained in Theorem 3.2.

Theorem 3.2 *NBC formula.*

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \tag{2}$$

Description:
X: Data with unknown class
Y: Hypothesis data X is a specific class
P(X): Probability of the observed sample
P(Y): Hypothesis probability Y
P(Y|X): Probability of hypothesis Y based on condition X
P(X|Y): Probability of X based on condition on hypothesis Y
The relationship between hypothesis (Y) and condition (X) with classification is, the hypothesis in Bayes' theorem is the class label which is the target attribute, while the condition is the set of attributes that are input in the classification model [15]. The probability of P(X) is always constant, so in the calculation it can be ignored and only calculates the part of P(X|Y)P(Y).

3.3 Bagging Ensemble Classifier

The bagging method has two main processes, namely boostrap and aggregating. The first process is bootstrapping, where a sampling process is carried out by resampling the training data so that new data variations are formed and use them as new learning sets using a predetermined base learner [16]. Then the testing data will be tested into each model from each sample, so that each model has different intelligence from one another. Next is the aggregating process, which is the merging of many predicted values from k models into one estimated value [17]. This is done using majority vote, which is determining the class estimate from the most classification results. The illustration of the stages of the bagging process is described in Fig 1.

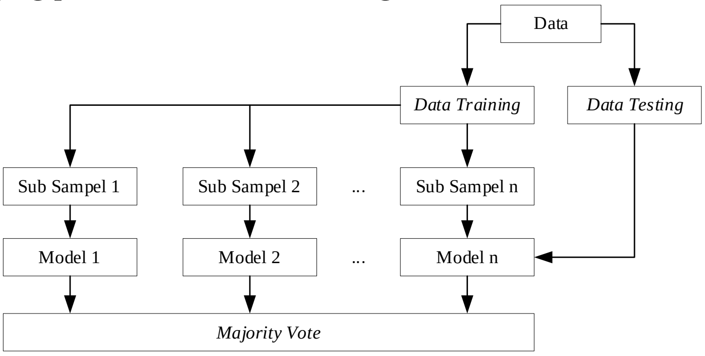

Fig 1: Bagging Process Stages

3.4 Decision Tree

Decision tree is a classification method using a tree structure, where each node shows the variable and the branch shows the value of the variable, and the leaves indicate the target class. The top node is called the root [4]. Decision tree performs top down classification. The classification process is carried out by tracing the path from root to leaf, then predicting the class from the data [18]. The illustration of the decision tree algorithm is described in Fig 2.

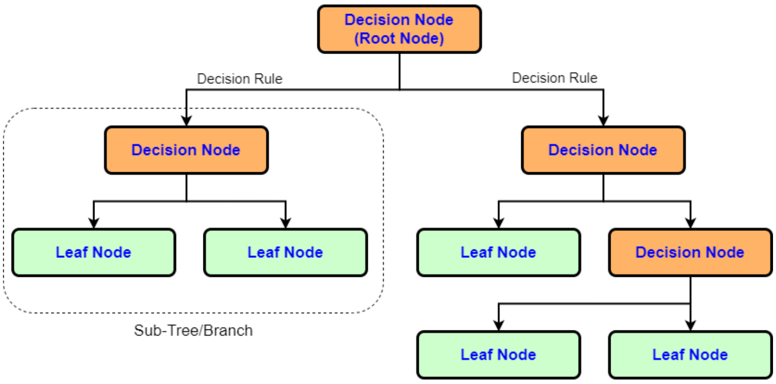

Fig 2: Decision Tree Algorithm (Source: satishgunjal.com/decision_tree)

3.5 Ordinal Logistic Regression

Ordinal logistic regression is a regression analysis in which the response variable is data with an ordinal scale that has more than two categories, with the predictor variables being either categorical data or continuous data [5]. The model used is the logit model, obtained by comparing the cumulative probability, namely the probability of less than the same as the q^{th} response category on p predictor variables.

Theorem 3.3 *Cumulative probability.*

$$P(Y \le q \mid x) = \frac{\exp[\alpha_q + \sum_{l=1}^{p} \beta_l x_l]}{1 + \exp[\alpha_q + \sum_{l=1}^{p} \beta_l x_l]} \tag{3}$$

Where q = 1, 2, …, Q is the response category.

Theorem 3.4 *Logit function.*

$$\text{logit } P(Y \le q \mid x) = \log\left[\frac{P(Y \le q \mid x)}{P(Y > q \mid x)}\right] \tag{4}$$

By substituting Equations 3 and 4 we get.

Theorem 3.5 *Ordinal logistic regression model.*

$$\text{logit } P(Y \le q \mid x) = \log\left[\frac{P(Y \le q \mid x)}{1 - P(Y \le q \mid x)}\right] \tag{5}$$

$$= \alpha_q + \sum_{l=1}^{p} \beta_l x_l$$

3.6 Confusion Matrix

The confusion matrix is a two-dimensional matrix between the actual class and the predicted class which serves to record the classification performance [19]. The confusion matrix with multi-class targets is presented in Table 3.

Table 3: Confusion Matrix

		Prediction Class (t)			Total
		Class A	Class B	Class C	
Actual Class (s)	Class A	TP_A	E_{AB}	E_{AC}	$FN_A=E_{AB}+E_{AC}$
	Class B	E_{BA}	TP_B	E_{BC}	$FN_B=E_{BA}+E_{BC}$
	Class C	E_{CA}	E_{CB}	TP_C	$FN_C=E_{CA}+E_{CB}$
Total		$FP_A=E_{BA}+E_{CA}$	$FP_B=E_{AB}+E_{CB}$	$FP_C=E_{AC}+E_{BC}$	n

Table 3 shows the correct prediction results are called True Positive (TP) and the rest shows the wrong predictions. Classification evaluation with multi-class targets is carried out based on each existing target class, where the evaluation values used are accuracy, precision, and sensitivity which are described as follows [20].

Accuracy

The accuracy value can be obtained through Theorem 3.6.

Theorem 3.6 *Accuracy.*

$$\text{Accuracy} = \frac{TP_A + TP_B + TP_C}{n} \tag{6}$$

Precision

The precision value can be obtained through Theorem 3.7.

Theorem 3.7 *Precison.*

$$\text{Precision}_{(s)} = \frac{TP_S}{TP_S + FP_S} \tag{7}$$

Sensitifity

The sensitivity value can be obtained through Theorem 3.8.

Theorem 3.8 *Sensitivity.*

$$\text{Sensitifity}_{(s)} = \frac{TP_S}{TP_S + FN_S} \tag{8}$$

3.7 Berthing Time

The Decree of the General Director of Sea Transportation Number UM.002/38/18/ DJPL-11 concerning Port Operational Service Performance Standards explains that docking time is the time used by ships while at the dock to carry out loading and unloading activities, which is calculated from the time the first rope is tied to the dock (mooring), until the last mooring rope is released from the dock (unmooring) [21]. The calculation of the berthing time is described in Theorem 3.7.1 [22].

Theorem 3.7.1 *Berthing time.*

$$BT = \frac{\text{Number of Loading and Unloading Continers}}{\text{BSH}} \tag{9}$$

3.8 Estimation Time of Departure (ETD)

Estimation Time of Departure (ETD) is the estimated time of ship's departure from the dock after loading and unloading [23].

4 The Proposed Method

The analysis steps for predicting ship departure delay are explained as follows.

1. Pre-processing of ship data by categorizing the length of the ship's departure delay.
2. Describe the characteristics of the ship.
3. Divide the data into training data (80%) and testing data (20%) randomly.
4. Classify ship departure delays using the KNN, NBC, NBC with Bagging, Decision Tree, and Ordinal Logistics Regression.
5. Choose the classification method with the best performance
6. Interpret the results of the analysis.
7. Draw conclusions and suggestions.

5 Results, Analysis and Discussions

This chapter will discuss the results of the descriptive analysis of ship departure delay on 171 ships and the results of the classification analysis of ship departure delay.

5.1 Characteristics of Ship Departure Delay

The proportion of ships based on departure delays at PT XYZ in 2021 is 127 ships (74.27%) not experiencing departure delays, 36 ships (21.05%) experiencing departure delays > 4 hours, and 8 ships (4.68%) experiencing departure delays ≤ 4 hours. The proportion of ships based on departure delay is presented in Fig 3.

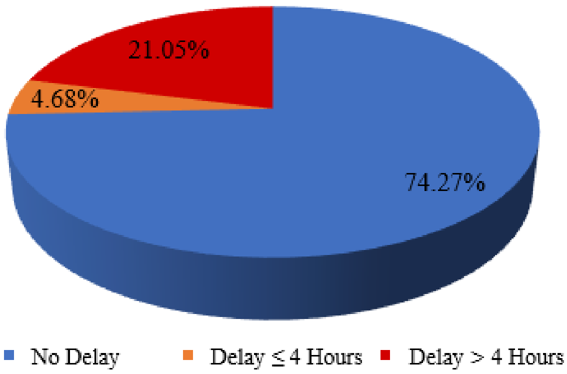

Fig 3: Characteristics of Departure Delay

5.2 Classification of Ship Departure Delay

The results of the classification of ship departure delays in each classification method will be presented in the form of precision, sensitivity, and accuracy values.

Precision describes the percentage of ships that are correctly predicted for each category of delay based on the category in each prediction, while sensitivity describes how many percent of ships that are correctly predicted for each category of delay to the overall category of actual ships. Accuracy describes the percentage of ships that were predicted correctly compared to the overall ship data. Precision, sensitivity, and accuracy are used to determine the performance of the classification results.

The three indicators of the goodness of the model are obtained using Equations 6, 7, and 8. The classification performance of each classification method is presented in Table 5.

Table 5: Classification Performance

Methods	Delay Category	Precision	Sensitivity	Accuracy
KNN (k = 5)	No Delay	0,91	1	0,91
	Delay ≤ 4 Hours	1	0,5	
	Delay > 4 Hours	1	0,33	
NBC	No Delay	0,89	0,83	0,74
	Delay ≤ 4 Hours	0	0	
	Delay > 4 Hours	0,17	0,33	
NBC with Bagging	No Delay	0,86	1	0,86
	Delay ≤ 4 Hours	0	0	
	Delay > 4 Hours	0	0	
Decision Tree	No Delay	0,87	0,67	0,63
	Delay ≤ 4 Hours	0,5	0,5	
	Delay > 4 Hours	0,1	0,33	
Ordinal Logistic Regression	No Delay	0,79	0,93	0,77
	Delay ≤ 4 Hours	0	0	
	Delay > 4 Hours	0,15	0,36	

Table 5 shows that the KNN method with k = 5 has the best classification performance which is selected based on the sensitivity value. The accuracy value is 0.91, which means that 91% of the data can be classified correctly. The classification results have high accuracy, but when the ship experiences a delay in departure, the probability of being classified correctly is low, this is due to an imbalance in data classes.
The non-delay category has a precision value of 0.91 which means that there are 91% of ships that do not experience delays of all ships that are predicted to experience no delay, while a sensitivity value of 1 means that all ships that do not experience delays are predicted to not experience delays.
The delay category ≤ 4 hours has a precision value of 1, which means that all ships experiencing a delay ≤ 4 hours out of all ships predicted to experience a delay ≤ 4 hours, while the sensitivity value is 0.5, which means that 50% of ships experiencing a delay ≤ 4 are predicted correctly experienced a delay ≤ 4 hours.
The delay category > 4 hours has a precision value of 1, which means that all ships experiencing delays > 4 hours out of all ships predicted to experience delays > 4 hours, while the sensitivity value is 0.33 which means that 33% of ships experiencing delays > 4 are predicted correctly experienced a delay of > 4 hours.
In the case of ship departure delay classification, the sensitivity value is preferred. The terminal will prefer to predict the ship will experience a delay in departure even though the ship is not delayed, compared to predicting the ship will not experience a delay when the ship should be delayed. This is because if the actual ship is experiencing delays but is not predicted to experience delays, then the terminal will not need to anticipate operational activities to speed up loading and unloading services and do not schedule the next ship, which will result in delays for the next ship to dock.
The classification performance results are obtained from calculations based on the confusion matrix. The confusion matrix classification using KNN with k = 5 is presented in Table 6.

Table 6: Confusion Matrix KNN k = 5

		Prediction Class			Total
		No Delay	Delay ≤ 4 Hours	Delay > Hours	
Actual Class	No Delay	30	0	0	30
	Delay ≤ 4 Hours	1	1	0	2
	Delay > 4 Hours	2	0	1	3
Total		33	1	1	35

Table 6 shows that the results of the classification of the KNN method with k = 5 there are 30 ships that did not experience a departure delay, all of which were correctly classified as ships that did not experience a departure delay. Meanwhile, of the 2 ships that experienced a departure delay ≤ 4 hours, 1 ship was correctly classified as having a departure delay ≤ 4 hours. Furthermore, 1 out of 3 ships that experience a departure delay of > 4 hours is correctly classified as having a departure delay of > 4 hours.

6 Conclusion

The conclusions obtained based on the results of the analysis and discussion are.

1. The KNN method has the best performance in predicting ship departure delay, with an accuracy value of 91% and precision > 90%, but for the delay ≤ 4 hours and > 4 hours it has a sensitivity value ≤ 50%, it means ships with a delay ≤ 4 hours and > 4 hours have less accurate predictions.
2. The use of the bagging method on the NBC method is proven to improve the classification performance of the NBC method.

ACKNOWLEDGEMENTS
Thank you to all those who have supported the completion of this journal.

References

[1] B. A. Setiono. (2012). Dampak Perubahan Rencana Kedatangan Kapal Terhadap Arus Barang dalam Sistem Windows di Terminal Nilam Multipurpose di PT. Pelabuhan Indonesia III Cabang Tanjung Perak Surabaya. *J. Apl. Pelayaran dan Kepelabuhanan,* 2, 121.
[2] Y. Heryadi and E. Irwansyah. (2020). Deep Learning: *Aplikasinya di Bidang Geospasial*. (pp. 4) AWI Technology Press.
[3] X. Wu *et al.*. (2008). Top 10 Algorithms in Data Mining. *Knowl. Inf. Syst,* 14(1), 1–37.
[4] D. A. Pratiwi, R. M. Awangga, and M. Y. H. Setyawan. (2020). *Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Menggunakan Metode Naïve Bayes*. (pp. 37). Kreatif Industri Nusantara.
[5] Kurniawan. (2019). *Analisis Data Menggunakan Stata Se 14 (Panduan Analisis, Langkah Lebih Cepat, Lebih Mudah Dan Paling Praktis)*. (111). Deepublish.
[6] J. Han, J. Pei, and M. Kamber. (2011). *Data Mining: Concepts and Techniques*. (pp. 461). United States: Elsevier.
[7] H. Ramos and H. S. Ismail. (2020). Analisa Peyebab Keterlambatan Kegiatan Bongkar Muat pada PT. Pelabuhan Indonesia I Dumai. *J. Marit. Educ.*, 2(1), 64–69.
[8] M. Safrianda and F. Fatnanta. (2016). Analisis Penyebab Keterlambatan Bongkar Muat Barang Akibat Faktor Peralatan (Studi Kasus: Dermaga A Pelabuhan Dumai). *Jom FTEKNIK*, 3(2).
[9] H. Azis, F. T. Admojo, and E. Susanti. (2020). Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah. *Techno. Com*, 19(3), 286–294.
[10] F. Fitriyani. (2018). Metode Bagging Untuk Imbalance Class Pada Bedah Toraks Menggunakan Naive Bayes. *J. Kaji. Ilm.*, 18(3), 278–282.
[11] F. M. Hana. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4. 5. *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, 4(1), 32–39.
[12] R. Nurmalasari, D. Ispriyanti, and S. Sudarno.(2017). Analisis Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia (IPM) Menggunakan Metode Regresi Logistik Ordinal Dan Regresi Probit Ordinal (Studi Kasus Kabupaten/Kota Di Jawa Tengah Tahun 2014). *J. Gaussian*, 6(1), 111–120.
[13] U. Sa’adah, M. Y. Rochayani, D. W. Lestari, and D. A. Lusia. (2021). *Kupas Tuntas Algoritma Data Mining dan Implementasinya Menggunakan R*. (pp. 121). Malang: UB Press.

[14] N. L. W. S. R. Ginantra *et al*.. (2021). *Data Mining dan Penerapan Algoritma*. (pp. 119). Medan: Yayasan Kita Menulis.
[15] R. A. Putra. (1996). Penerapan Naïve Bayes Classifier dengan Gaussian Function Untuk Menentukan Kelompok UKT. *J. Inform. Glob.*, 9(2).
[16] L. Breiman. (1996). Bagging predictors. *Mach. Learn.*, 24(2), 123–140.
[17] W. Agwil, H. Fransiska, and N. Hidayati. (2020). Analisis Ketepatan Waktu Lulus Mahasiswa dengan Menggunakan Bagging CART. *FIBONACCI J. Pendidik. Mat. dan Mat.*, 6(2), 155–166.
[18] Y. Ardilla *et al*.. (2021). *Data Mining dan Aplikasinya*. (pp. 123). Penerbit Widina.
[19] E. Prasetyowati. (2017). *Data Mining: Pengelompokan Data untuk Informasi dan Evaluasi*. (pp. 97). Pamekasan: Duta Media Publishing.
[20] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis. (2021). Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. *Technologies*, 9(4), 81.
[21] Decree of the General Director of Sea Transportation UM.002/38/18/ DJPL-11. 2011. *Port Operational Service Performance Standards*. Ministry of Transportation.
[22] M. Y. Jinca. (2019). *Transportasi Laut Indonesia*. (pp. 117). Sidoarjo: Brilian Internasional.
[23] A. Susilo, B. Mukhsen, A. Fuadhi, and F. K. S. W. (2018). *Buku Pintar Dunia Ekspor (Seri 1)*. (pp. 161). Semarang: CV. Oxy Consultant.

Notes on contributors

Sri Mumpuni Retnaningsih. *She is an Associate Lecturer at the Department of Business Statistics, Faculty of Vocational, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Her main research is on productivity measurament, capability process, and operation management.*

Iis Dewi Ratih. *She is an Associate Lecturer at the Department of Business Statistics, Faculty of Vocational, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Her main research is on statistical modeling, multivariate, and extreme value theory.*

Kharin Octavian Ranto. *She is a student at the Department of Business Statistics, Faculty of Vocational, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. She has experience as a lecture assistant for 2 years, an internship focused on data analyst for 5 months, and has an interest in data analytics. Her main research is on statistical modeling for industry.*