The Indonesian Journal of Computer Science
www.ijcs.net
Volume 14, Issue 3, June 2025
https://doi.org/10.33022/ijcs.v14i3.4533

The Effects of Data Sampling and Feature Selection on Public Service Satisfaction Using an Ensemble Classifier Algorithm

Agus Hidayat[1], Joniwarta[2], Tyastuti Sri Lestari[3], and Wowon Priatna[4]
agus.hidayat@dsn.ubharajaya.ac.id[1], joniwarta@dsn.ubharajaya.ac.id[2], tyas@ubharajaya.ac.id[3], wowon.priatna@dsn.ubharajaya.ac.id
[1,2,3,4] Departement of Informatics, Faculty of Computer Science, Bhayangkara Jakarta Raya University

Article Information

Received : 26 Dec 2024
Revised : 26 Apr 2025
Accepted : 9 Jun 2025

Keywords

level of satisfaction, ensemble classifier, Particle Swarm Optimization, Random Under sampling, Classification

Abstract

Customer satisfaction is an important factor that determines quality. User satisfaction analysis can identify the service quality and measure quality through an evaluation process to improve services. This research aims to measure the performance of services provided by the village government. Villages and sub-districts offer services based on the community's specific needs. Nevertheless, by delivering impeccable service, it is possible to satisfy the community without causing physical or material harm. An essential requirement is the development of a service user classification methodology to enhance service quality, efficiently address service user grievances, detect recurring trends, and promptly offer feedback to enhance the offerings of products and services. Machine learning approaches can be used to quantify public service satisfaction in the analytical process. Machine learning is an algorithmic approach used to assess and prioritize satisfaction with public services offered by service providers. The main approach for machine learning is an ensemble classifier. The data was analyzed using Excel; then, the data was processed first to create a classification model. At the preprocessing stage, the data is grouped to obtain labels/targets to be processed based on algorithmic classification. The classification uses the Classifier aggregation algorithm. Type improvements using optimization features using the Particle Swarm Optimization (PSO) sampling algorithm and random subsampling techniques. This research produced an accuracy value before adding sampling techniques and a PSO accuracy value of 92.68. After adding sampling techniques and PSO optimization, an accuracy value of 100% was obtained.

A. Introduction

Public services are services provided to people who are citizens or legal residents of the country concerned. Its implementation supports public service objectives Law No.32 of 2004 concerning regional government. The government's performance within the framework of regional autonomy is expected to prioritize the community's interests, especially in the provision of public services and public administration. The village government is the village head, whom village officials assist in running the village government. Along with society's current development, people need increasingly complex services and better, faster, and more accurate services. Subprime is a government management system with the authority to manage and provide services that meet community needs.

Customer satisfaction is a critical factor that determines quality. User satisfaction analysis can identify a given service's quality and measure quality through an evaluation process to improve service[1]. The evaluation procedure can be security, comfort, hygiene, and ease of access to a public service[2]. User satisfaction analysis can also evaluate the occurrence of faults during service provision[3]. Currently, villages and sub-districts provide services according to community needs. However, if they fully provide optimal service, it can create satisfaction for the community without harming the community either physically or materially. To enhance service quality and efficiently address service user grievances, it is necessary to implement a classification approach for service users. This method will help discover patterns and provide timely feedback to improve the products and services offered[4]. The analysis process that has been carried out to measure the level of public satisfaction with services can use machine learning methods[5]. Machine learning is a method of ranking satisfaction with public services provided by service providers[6][7].

There are some studies in the classification of public service satisfaction dieters from previous research: Research [8] Classification of village service satisfaction using the decision tree algorithm yielded an accuracy of 90.66%. Classification of village service satisfaction using the decision tree algorithm produces an accuracy of 90.66%. Type of complaint reports using LDA-SVM makes an accuracy of 79.85%[9]. Classification of public service complaints using Naïve Bayes[6]. Using the Naïve Bayes, KNN, SVM, and boosting algorithms produces the best SVM accuracy compared to other methods[10]. Classification models will produce high accuracy values if there is no uneven distribution of classes[11], which can lead to bias in applications and hinder machine learning from creating classification models. One technique to reduce imbalanced classes uses data sampling techniques that can divide majority classes into balanced classes[12]. The data sampling technique succeeded in increasing the accuracy produced by the Ensemble Classifier[13], and the data sampling technique can anticipate unbalanced class data[14].

Selected from the total dataset[15]; feature selection is a process where part of the feature space is selected according to its relevance by considering the output of the classification[16]. PSO is a feature optimization algorithm that can increase the resulting accuracy values. PSO has been used to optimize accuracy increases for Random Forest, Decision Tree, Naïve Bayes and KKN for diabetes dataset classification[17]. PSO can be combined with data sampling techniques to increase

the accuracy of the ensemble classifier algorithm in classification [18]. This research focuses on and provides the latest contributions, including creating a public service satisfaction classification model using an ensemble classifier, testing the effect of feature selection on ensemble classifier classification, testing the effect of feature selection on the ensemble classifier algorithm and finally, a combination of data sampling techniques and feature selection on Ensemble classifier algorithm performance.

B. Research Method

The data set used in this research is a survey by sending Google forms distributed to the community in the village located in the conservation of the southern Tambun of Bekasi. Ten thousand twenty-nine respondents completed the survey with 15 questions to measure the level of satisfaction with the public service provided by the service provider, the village. The following variables and dimensions, descriptions, scales, and intervals of the research data used are presented in Table 1. The values in Table I correspond to the following categories: 1 represents satisfied (S) and 2 represents non satisfied (NS).

Table 1. List Variable, Measurement and Value

Variable	Description	Possible Value
X1	Facilities and infrastructure	1= S, 2= NS
X2	Cleanliness	1= S, 2= NS
X3	Employee appearance	1= S, 2= NS
X4	On Time	1= S, 2= NS
X5	Employee appearance	1= S, 2= NS
X6	Conformity	1= S, 2= NS
X7	Responsive	1= S, 2= NS
X8	Accuracy of information	1= S, 2= NS
X9	Attitude	1= S, 2= NS
X10	Service Suitability	1= S, 2= NS
X11	Cost Suitability	1= S, 2= NS
X12	Easy to contact	1= S, 2= NS
X13	Attention	1= S, 2= NS
X14	Fair	1= S, 2= NS
X15	Friendly	1= S, 2= NS

One can assess the suitability of a study measuring instrument by examining its critical criteria, which include (a) validity and (b) reliability or rehabilitation[19]. The validity of an instrument can be assessed by calculating the correlation coefficient between the test instrument and the criteria/test of each element using Equation 1.

$$r_{xy} = \frac{\boldsymbol{n}(\sum \boldsymbol{xiyi}) - (\sum \boldsymbol{xi})(\sum \boldsymbol{yi})}{\sqrt{(\boldsymbol{n}(\sum \boldsymbol{xi^2}) - (\boldsymbol{xi})^2(\boldsymbol{n}\sum \boldsymbol{yi^2}) - (\boldsymbol{yi})^2))}} \quad (1)$$

The calculated correlation coefficient is then checked with the values in the table. An item is considered valid if its count is greater than or equal to the value of r in the table. If the count of r is less than the count of r in the table, then the item is considered invalid. For this study, 30 data points from fully answered questionnaires were utilized to determine the validity and reliability values of the students. The Cronbach's Alpha coefficient was computed using Equation 2 to assess

the instrument's dependability. The calculated alpha coefficient findings are then compared with the table values. If the alpha value is more than or equal to the critical value of r from the table, then the research instrument is considered dependable. If the alpha value is less than the critical value in the r table, then the research instrument is considered unreliable.

$$\mathrm{ri} = \frac{\mathrm{k}}{(\mathrm{k}-1)}\{1 - \frac{\sum \mathrm{si}^2}{\mathrm{st}^2}\} \quad (2)$$

This research uses 4 (four) ensemble classifier algorithms, including Random Forest (RF), XGBOOST, Catboost and Lightgbm, to classify public service satisfaction. Side data techniques using the RUS algorithm are used to anticipate class balance. PSO is used as an algorithm for feature selection and classification algorithm optimization. The modelling will be carried out in this research; Creating a public service satisfaction classification model using an ensemble classifier, testing the influence of feature selection on ensemble classifier classification, testing the influence of feature selection on the ensemble classifier algorithm and finally, a combination of data sampling techniques and feature selection on the performance of the ensemble classifier algorithm.

The Random Forest (RF) algorithm employs a recursive binary approach to arrive at the ultimate node in a tree structure, relying on tree classification and regression [20]. Three things must be known about the random forest algorithm. First, bootstrap sampling in making a prediction tree; on each decision tree, the prediction is carried out randomly, and the prediction is produced through a combination of each decision tree with a majority vote in classifying it [21]. The RF algorithm is constructed using a three-step process: (1) Sampling k training subsets, (2) Creating a decision tree model for each subset, and (3) Combining the k decision trees into an RF model. Applying the RF algorithm to big, unbalanced datasets yields high-performance outcomes and rapid execution times[21][22].

Extreme Gradient Boosting (XGBoost) is a type of machine learning used for prediction and classification with a decision tree structure[23]. XGBoost is a boosting method consisting of several decision trees where the previous and the following trees depend on each other [19]. When classifying, XGBoost will update the weights for each tree built to obtain a strong classification tree [24]. Equation (3) is calculating XGBoost.

$$\bar{\mathrm{Y}}\mathrm{i} = \sum_k^K f_k(x_i)\,, f_k \epsilon F \quad (3)$$

CatBoost is a machine learning algorithm that is still part of the Gradient Boosted Decision Trees (GBDT) family within the scope of ensemble learning[25]. CatBoost is an algorithm that is generally open to continue to be developed in the scope of two combinations: Ordered Target Statistics and Ordered Boosting, the combination of the two is called catboost[26].

LightGBM classification aims to find an additive model that minimizes the loss function[27]. The algorithm equation for improving regression trees can be generalized in equation 4, where the final model is a simple gradual addition model of the b value[28].

$$F(x) = \sum_{b=1}^{B} f^b(x) \quad (4)$$

The purpose of the feature selection stage is to eliminate superfluous and inconsequential features in the data set [29] It is crucial for the examination of sample data with a large number of dimensions [30]. The feature selection used in this research is PSO in the learning stage, which aims to optimize each data generated [30]. The stages of the PSO method are stated in Algorithm 1.

Algorithm 1. Particle Swarm Optimization [31]

1: Let N represent the size of the group or swarm in terms of the number of particles.
2: Create an initial population X by randomly selecting values within the range of X(B) and X(A), resulting in X1. Subsequently, denote particle j and its speed in iteration I as X(i) j and V (i) j, respectively. Thus, the initial particles are represented as X1(0), X2(0), ..., XN (0).
3: Determine the velocity of every individual particle. Assign the value of 1 to the iteration variable i.
4: On iteration i, find two important parameters for each particle j.
5: The velocity of particle j on iteration 1 is calculated as follows: vi,m=w. vi,m+c 1∗R*(pbesti,m−xi,m) +c 2∗R∗(gbestm−xii,m)
6: The velocity of particle j on iteration I is calculated as vi,m = w * vi,m + c1 * R * (pbesti,m - xi,m) + c2 * R * (gbestm - xi,m).

The accuracy rating is determined by the frequency at which the model produces correct results, as measured by a confusion matrix [32]. Classification evaluation assesses its performance by utilizing recall and precision metrics. The accuracy can be computed using equation (5), precision can be determined using equation (6), and recall can be evaluated using equation (7) for the Area Under the Curve (AUC)[33].

Accuracy = (TP+TN)/(TP+TN+FP+FN) (5)
Precision = (TP)/(TP+FP) (6)
Recall = (TP) / (TP+FN) (7)

C. Result and Discussion

The Python programming tool utilizes a machine learning package to construct an ensemble classifier algorithm classification model while also addressing feature optimization.

The first stage is the stage where processing must be carried out before classifying the model so that the dataset can be modelled using a classification algorithm. Survey data processed using Microsoft Excel amounted to 10029 responses. Of the 10,029 responses, 7,073 were satisfied, and 2,956 were dissatisfied. The dataset is processed using Google Collaboratory to handle data preprocessing and transformation processes so that the dataset is ready to create a classification model.

The Python programme utilises an ensemble classifier technique to classify user service level satisfaction. The training data is divided into 70% for training and

30% for testing. The classification model underwent testing using a confusion matrix to derive accuracy, precision, and recall metrics. Table 2 displays the test results. The classification findings indicate that all ensemble learning algorithms, as presented in Table 2, exhibit identical accuracy, recall, precision, and AUC values.

Table 2. Algorithm Ensemble Classification Results

Algorithm	Accuracy	Recall	Precision	AUC
Random Forest	92.68%	1.00	0.89	0.941
XGBOOST	92.68%	1.00	0.89	0.941
Chatboost	92.68%	1.00	089	0.941
LightGBM	92.68%	1.00	089	0.941

Table 2 shows an unbalanced class where the target or label for satisfied is 7043 and dissatisfied is 2956. So, in this study, the Random Under Sampling technique was used so that the class scores were balanced, satisfied at 7073 and dissatisfied at 7073. Next are the results from the RUS. A classification model was carried out using the Classifier ensemble algorithm. The modelling results, which then carried out the classification test, are shown in Table 3. After carrying out the sampling technique, the accuracy value for all algorithms increased to 100% and the AUC value to 1.00.

Table 3. Combination Classification Results with Sampling

Algorithm	Accuracy	Recall	Precision	AUC
Random Forest + RUS	100%	1.00	0.89	1.00
XGBOOST+RUS	100%	1.00	0.89	1.00
Chatboost+RUS	100%	1.00	089	1.00
LightGBM+RUS	100%	1.00	089	1.00

To increase accuracy, feature selection was also done using the PSO algorithm with 50 particles of 15 dimensions and a position value >=0.5, resulting in 8 features after selection from the initial 15 features. The following are the results of PSO optimization followed by creating a classification model. The modelling results, which were then carried out by the classification test, are shown in Table 4. The classification results shown in Table 4 show the accuracy and AUC values after feature selection, which increased to 100 and 1.

Table 4. Combination Classification Results with Feature Selection and Sampling

Algorithm	Accuracy	Recall	Precision	AUC
Random Forest + RUS+PSO	100%	1.00	0.89	1.00
XGBOOST+RUS+PSO	100%	1.00	0.89	1.00
Chatboost+RUS+PSO	100%	1.00	0.89	1.00
LightGBM+RUS+PSO	100%	1.00	0.89	1.00

Based on the accuracy results from testing the classification model used in this research, using the Random Under sampling technique and feature selection with PSO can maximize the accuracy and AUC values. The values produced by all ensemble classifier algorithms show that accuracy, recall, precision, and AUC are the same, even after combining them with sampling techniques and feature selection. This shows the classification model built by all the algorithms recommended for measuring satisfaction with public services provided by village governments.

D. Conclusion

With a sampling technique using the RUS algorithm to anticipate unbalanced classes, the accuracy value increases to 100% for all algorithms. The optimization results with PSO, and a combination of classification algorithms and sampling techniques get the maximum value. Utilizing Particle Swarm Optimization (PSO) for feature selection has effectively increased the accuracy of each model, thereby improving the classification process for each model. Random Forest, XGBOOST, Chartboost, LightGBM optimized using PSO to identify features or variables that impact public service satisfaction most. The models built have been proven to increase the accuracy of each algorithm, allowing classification models to assess the level of satisfaction with public services.

E. References

[1] J. Zhang, W. Chen, N. Petrovsky, And R. M. Walker, "The Expectancy-Disconfirmation Model And Citizen Satisfaction With Public." 2022.

[2] F. Osei, G. Ampomah, C. Kankam-Kwarteng, D. Opoku Bediako, And R. Mensah, "Customer Satisfaction Analysis Of Banks: The Role Of Market Segmentation," *Sci. J. Bus. Manag.*, Vol. 9, No. 2, P. 126, 2021, Doi: 10.11648/J.Sjbm.20210902.19.

[3] H. Bui, "Factors Affecting Customer Satisfaction And Word Of Mouth About Fresh Agricultural Products In Markets And Supermarkets In Long Xuyen City, An Giang Province," *Sci. J. Tra Vinh Univ. Issn 2815-6072; E-Issn 2815-6099*, Vol. 1, Pp. 1–10, 2021, Doi: 10.35382/18594816.1.45.2021.779.

[4] D. Banga And K. Peddireddy, "Artificial Intelligence For Customer Complaint Management," *Int. J. Comput. Trends Technol.*, Vol. 71, No. 3, Pp. 1–6, 2023, Doi: 10.14445/22312803/Ijctt-V71i3p101.

[5] Z. Xu, G. Zhu, N. Metawa, And Q. Zhou, "Machine Learning Based Customer Meta-Combination Brand Equity Analysis For Marketing Behavior Evaluation," *Inf. Process. Manag.*, Vol. 59, P. 102800, 2022, Doi: 10.1016/J.Ipm.2021.102800.

[6] K. Wabang, Oky Dwi Nurhayati, And Farikhin, "Application Of The Naïve Bayes Classifier Algorithm To Classify Community Complaints," *J. Resti (Rekayasa Sist. Dan Teknol. Informasi)*, Vol. 6, No. 5, Pp. 872–876, 2022, Doi: 10.29207/Resti.V6i5.4498.

[7] S. Khedkar And S. Shinde, "Deep Learning And Ensemble Approach For Praise Or Complaint Classification," *Procedia Comput. Sci.*, Vol. 167, Pp. 449–458, 2020, Doi: 10.1016/J.Procs.2020.03.254.

[8] M. Abdurohman, R. Husna, I. Ali, G. Dwilestari, And N. Rahaningsih, "Penerapan Model Klasifikasi Dalam Tingkat Kepuasan Layanan Publik Kelurahan Karyamulya Dengan Menggunakan Algoritma Decision Tree," *Inf. Manag. Educ. Prof. J. Inf. Manag.*, Vol. 6, No. 1, P. 81, 2022, Doi: 10.51211/Imbi.V6i1.1678.

[9] M. Alkaff, A. R. Baskara, And I. Maulani, "Klasifikasi Laporan Keluhan Pelayanan Publik Berdasarkan Instansi Menggunakan Metode Lda-Svm," *J. Teknol. Inf. Dan Ilmu Komput.*, Vol. 8, No. 6, Pp. 1265–1276, 2021, Doi: 10.25126/Jtiik.2021863768.

[10] F. Caldeira, L. Nunes, And R. Ribeiro, "Classification Of Public Administration Complaints," *Openaccess Series In Informatics*, Vol. 104, No. 9. Schloss Dagstuhl

– Leibniz-Zentrum Für Informatik, Dagstuhl Publishing, Germany, Pp. 9:1-9:0, 2022. Doi: 10.4230/Oasics.Slate.2022.9.
[11] Z. Salekshahrezaee, J. L. Leevy, And T. M. Khoshgoftaar, "The Effect Of Feature Extraction And Data Sampling On Credit Card Fraud Detection," *J. Big Data*, Vol. 10, No. 1, 2023, Doi: 10.1186/S40537-023-00684-W.
[12] B. Lliu And G. Tsoumakas, "Dealing With Class Imbalance In Classifier Chains Via Random Undersampling," *Knowledge-Based Syst.*, Vol. 192, P. 105292, 2020, Doi: 10.1016/J.Knosys.2019.105292.
[13] Y. E. Kurniawati And Y. D. Prabowo, "Model Optimisation Of Class Imbalanced Learning Using Ensemble Classifier On Over-Sampling Data," *Iaes Int. J. Artif. Intell.*, Vol. 11, No. 1, Pp. 276–283, 2022, Doi: 10.11591/Ijai.V11.I1.Pp276-283.
[14] L. Liu, X. Wu, S. Li, Y. Li, S. Tan, And Y. Bai, "Solving The Class Imbalance Problem Using Ensemble Algorithm: Application Of Screening For Aortic Dissection," *Bmc Med. Inform. Decis. Mak.*, Vol. 22, No. 1, Pp. 1–16, 2022, Doi: 10.1186/S12911-022-01821-W.
[15] E. Purnamasari, D. Palupi Rini, And Sukemi, "Seleksi Fitur Menggunakan Algoritma Particle Swarm Optimization Pada Klasifikasi Kelulusan Mahasiswa Dengan Metode Naive Bayes," *J. Resti (Rekayasa Sist. Dan Teknol. Informasi)*, Vol. 1, No. 3, Pp. 469–475, 2020.
[16] S. A. Alsenan, I. M. Al-Turaiki, And A. M. Hafez, "Feature Extraction Methods In Quantitative Structure-Activity Relationship Modeling: A Comparative Study," *Ieee Access*, Vol. 8, Pp. 78737–78752, 2020, Doi: 10.1109/Access.2020.2990375.
[17] A. Fauzi And A. H. Yunial, "Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K – Nearest Neighbor, Dan Random Forest Menggunakan Algoritma Particle Swarm Optimization Pada Diabetes Dataset," *J. Edukasi Dan Penelit. Inform.*, Vol. 8, No. 3, P. 470, 2022, Doi: 10.26418/Jp.V8i3.56656.
[18] D. Zheng, C. Qin, And P. Liu, "Adaptive Particle Swarm Optimization Algorithm Ensemble Model Applied To Classification Of Unbalanced Data," *Sci. Program.*, Vol. 2021, No. 1, 2021, Doi: 10.1155/2021/7589756.
[19] H. H. Sinaga And S. Agustian, "Pebandingan Metode Decision Tree Dan Xgboost Untuk Klasifikasi Sentimen Vaksin Covid-19 Di Twitter," *J. Nas. Teknol. Dan Sist. Inf.*, Vol. 8, No. 3, Pp. 107–114, 2022, Doi: 10.25077/Teknosi.V8i3.2022.107-114.
[20] L. E. O. Breiman, "Random Forests," Pp. 5–32, 2001.
[21] W. Yustanti And N. Rochmawati, "Analisis Algoritma Klasifikasi Untuk Memprediksi Karakteristik Mahasiswa Pada Pembelajaran Daring," *J. Edukasi Dan Penelit. Inform.*, Vol. 8, No. 1, Pp. 57–61, 2022.
[22] Yoga Religia, Agung Nugroho, And Wahyu Hadikristanto, "Klasifikasi Analisis Perbandingan Algoritma Optimasi Pada Random Forest Untuk Klasifikasi Data Bank Marketing," *J. Resti (Rekayasa Sist. Dan Teknol. Informasi)*, Vol. 5, No. 1, Pp. 187–192, 2021, Doi: 10.29207/Resti.V5i1.2813.
[23] I. Pratama, A. Y. Chandra, And P. T. Presetyaningrum, "Seleksi Fitur Dan Penanganan Imbalanced Data Menggunakan Rfecv Dan Adasyn," *J. Eksplora Inform.*, Vol. 11, No. 1, Pp. 38–49, 2022, Doi: 10.30864/Eksplora.V11i1.578.
[24] Z. Salam Patrous, "Evaluating Xgboost For User Classification By Using

Behavioral Features Extracted From Smartphone Sensors," P. 67, 2018, [Online]. Available: Https://Www.Diva-Portal.Org/Smash/Record.Jsf?Pid=Diva2%3a1240595&Dswid=-6444

[25] A. N. A. Aldania, A. M. Soleh, And K. A. Notodiputro, "A Comparative Study Of Catboost And Double Random Forest For Multi-Class Classification," *J. Resti (Rekayasa Sist. Dan Teknol. Informasi)*, Vol. 7, No. 1, Pp. 129–137, 2023, Doi: 10.29207/Resti.V7i1.4766.

[26] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, And A. Gulin, "Catboost: Unbiased Boosting With Categorical Features," *Adv. Neural Inf. Process. Syst.*, Vol. 2018-December, No. Section 4, Pp. 6638–6648, 2018.

[27] S. Touzani, J. Granderson, And S. Fernandes, "Gradient Boosting Machine For Modeling The Energy Consumption Of Commercial Buildings," *Energy Build.*, Vol. 158, Pp. 1533–1543, 2018, Doi: 10.1016/J.Enbuild.2017.11.039.

[28] J. Christian, I. Ernawati, And ..., "Implementasi Penggunaan Algoritma Categorical Boosting (Catboost) Dengan Optimisasi Hiperparameter Dalam Memprediksi Pembatalan Pesanan Kamar Hotel," *… Mhs. Bid. Ilmu …*, No. 2021, Pp. 641–651, 2022, [Online]. Available: Https://Conference.Upnvj.Ac.Id/Index.Php/Senamika/Article/View/2230

[29] Y. Wanli Sitorus, P. Sukarno, S. Mandala, F. Informatika, And U. Telkom, "Analisis Deteksi Malware Android Menggunakan Metode Support Vector Machine & Random Forest," *E-Proceeding Eng.*, Vol. 8, No. 6, Pp. 12500–12518, 2021.

[30] L. Zhang, "A Feature Selection Algorithm Integrating Maximum Classification Information And Minimum Interaction Feature Dependency Information," *Hindawi Comput. Intell. Neurosci.*, Vol. 2021, P. 10, 2021.

[31] U. Indahyanti, N. L. Azizah, And H. Setiawan, "Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa," *J. Sains Dan Inform.*, Vol. 8, No. 2, Pp. 160–169, 2022, Doi: 10.34128/Jsi.V8i2.459.

[32] R. C. Chen, C. Dewi, S. W. Huang, And R. E. Caraka, "Selecting Critical Features For Data Classification Based On Machine Learning Methods," *J. Big Data*, Vol. 7, No. 1, 2020, Doi: 10.1186/S40537-020-00327-4.

[33] P. Sedgwick, "How To Read A Receiver Operating Characteristic Curve," *Bmj*, Vol. 350, 2015, Doi: 10.1136/Bmj.H2464.