

Optimizing Library Visitor Satisfaction Analysis With Machine Learning

Yeni Fitria Nurahman^{1,*}, Imam Yuadi²

^{1,2} Universitas Airlangga, Indonesia

Jl. Airlangga No 4-6 Gubeng, Surabaya

^{1,*}yeni.fitria.nurahman-2024@fisip.unair.ac.id, ²imam.yuadi@fisip.unair.ac.id

Article history:

Received 20 August 2025
Revised 10 October 2025
Accepted 28 November 2025
Available online 30 November 2025

Keywords:

Library Services,
Visitor Satisfaction,
Machine Learning,
Predictive Modeling,
Orange Data Mining.

Abstract

In today's increasingly digital era, libraries continue to play a vital role as centers of information, knowledge, and culture. Despite the widespread availability of online information, libraries remain essential for providing diverse resources, services, and convenient facilities. The role of libraries has evolved to meet the needs and expectations of visitors, requiring ongoing innovation in services and amenities to ensure user satisfaction. This study aims to assess the level of visitor satisfaction at UNUSA Library regarding the services provided. The research utilized questionnaire data, initially collected from 802 respondents, of which 224 valid responses were analyzed. Furthermore, this study compares the predictive performance of three machine learning methods—K-Nearest Neighbor, Decision Tree, and Support Vector Machine—to determine which method achieves the highest accuracy in predicting visitor satisfaction. The analysis was conducted using the Orange Data Mining application as the prediction model. The results indicate that library visitors generally report a high level of satisfaction, with certain services rated more positively than others, and that machine learning models can effectively predict satisfaction levels based on visitor feedback.



This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2025 by author.

I. INTRODUCTION

In the digital era that continues developing, libraries still maintain a vital role as centers of information, knowledge, and culture. Progress in technology has made it easier to access information through online networks, but libraries still have their value that is not replaced as a place that offers various sources of convenient and supportive power, services, facilities, and activity literacy. In addition to providing collection prints, libraries are also now expanding digital services to follow the development of technology

^{1*} Corresponding author

and patterns of consumption in the information society. Transformation This makes the library a room for dynamic and inclusive learning, where visitors can access various forms of knowledge and facilities for free [1]. Function library Keep going and develop along with the changing needs and wants of visitors. Realizing the matter, the library sued to innovate in service as well as provide modern facilities, such as discussion rooms, Wi-Fi access, digital library services, and comfortable and calm reading spaces. Steps innovative This is not only for interesting visitors but also to make the library relevant in the middle development of the digital era. According to IFLA (International Federation of Library Associations and Institutions), libraries need to adapt to the needs of modern society through a more interactive and based approach so that they can still become an empowered institution used to support learning throughout life [2].

Research on connection between frequency visits and levels satisfaction visitors library expected can give valuable insights for manager library. This understanding can become runway in designing improvement strategies facilities and services library. With know aspects what is needed improved , manager can more effective in fulfil hope visitors, good those who are regular come or the rare ones come. This can also increase image library as relevant and comfortable place for various circles public . Research on satisfaction visitors library has Lots done For measure to what extent are the facilities and services library fulfil hopes and needs its users. Some study show that factors like completeness collection, comfort space, and quality service staff greatly influences satisfaction visitors [3]. This research indicates that library that provides responsive, convenient and informative service tend own more visitors satisfied. Another study found that visit routine to library relate with more satisfaction tall to aspects certain, such as availability room comfortable learning and facilities that support the learning process. Frequent visitors come to library show level more satisfaction Good compared to with those who are rare visit. This is can caused by the presence of more interaction intense between users and staff, as well as ability user For more understand facilities provided [4]. However, research this No fully delve deeper How variation frequency visit influence dimensions satisfaction other like technology or access digital collection.

The study provided perspective important related quality service as one of the determinant satisfaction visitors library [5]. This research find that friendly and efficient service from staff library own influence significant to satisfaction visitors, regardless from frequency visit. Although study This focus on relationships between service staff and satisfaction, study This Not yet explore whether more visitors often come will show level different satisfaction to same service compared to rare visitor come. Library make it easier access visitors with provide facility online catalogs, online services such as borrowing independent, return independent, extension independent, library suggestions, ordering book, upload independent and facilities for refreshing or healing space with diverse Games. Library try always update and do innovation in accordance with the need for them to feel satisfied with facilities and services library. One of the method library know how much satisfied the visitors is with provide assessment media. The purpose of study This is For know how much satisfied visitors library so that they can maintain trust they to service library. research This use method Data Mining, Extraction and analysis For determine level satisfaction visitors library [6].

In this research, using a model that can reliably predict the level of satisfaction of visitors to the service library. Calculate and evaluate accuracy using three algorithms through statistical techniques and patterns found in data of big size [7]. Supervised learning is algorithms that make it possible to function in certain

ways to produce output with a method to map input to the desired result. Parameters in the algorithm are adjusted by the model to minimize error between results prediction and actual output [8]. The main benefit of supervised learning is that with the data that has been labeled model can produce more accurate and relevant predictions. One of the machine learning classification algorithms that can be used to make predictions is the Decision Tree C4.5 Algorithm and the Naive Bayes Algorithm. The goal of this study is to compare the accuracy and performance of two algorithms to predict cerebrovascular disease or stroke. Based on results study obtained that the C4.5 algorithm obtains a greater level of accuracy, which is 95%, while the Naive Bayes algorithm obtains a level of accuracy of 91% [9]. Development technology advances rapidly, followed by the growing popularity of expanding social media. Social networking networks can create big-data profiles of users, tracking all activities such as participation, delivery messages, and website visits.

Many people now post criticism about something on social media platforms such as Facebook, Twitter, Instagram, and others. Therefore, it is important to comprehend how social media users respond to OpenAI's ChatGPT being made available to the public. Ekman's algorithm and the Orange Data Mining program are used to process positive, negative, and neutral answers in order to understand the emotion in the received Twitter data. The findings indicate that ChatGPT obtained a score of 54.72%, 31.64%, and 13.64% for neutral, positive, and negative responses to speaking English. The analysis sentiment speak Indonesian no far differs from mark reaction neutral by 63.96%, response positive 23.56%, and response negative 12.48%. So, we may conclude that the majority of the public responds neutrally or rejects the release of ChatGPT [10]. The Orange Data Mining Application compares a number of algorithms learning machines. To classify similar types of vehicles in the system, use digital ticketing. This study compares and analyses the algorithms logistic regression, Support Vector Machine (SVM), and Neural Network (NN) for vehicle classification in digital ticketing. The data set was obtained from the Kaggle website, and the approach was validated using a confusion matrix. The research results suggest that in the training process and depending on the dataset used, the algorithms with the highest levels of accuracy are logistic regression, neural networks, and support vector machines. During the testing procedure, all algorithms on the model may do classification with complete accuracy, precisely 100% [11].

The purpose of this study is to use Orange Data Mining software to assess sentiment among Twitter users regarding the debut of the iPhone 16. The data came from 344 English-language tweets in Indonesia that were gathered using a crawling technique and tweet-harvesting tools. The research process includes preprocessing text to eliminate irrelevant elements and using the Ekman algorithm to analyze emotion. The findings of the analysis indicate that the launch of the iPhone 16 elicits a range of emotional responses, with "surprise" (182 tweets) being the most prevalent emotion and "happiness" (110 tweets) coming in second. Results This illustrates how a new product launch can have a big emotional influence on social media users, both positively and negatively [12]. This research predicts graduation students; however, in other universities, there is no single strategy for estimating whether a student can graduate on time. To overcome the matter, a model capable of predicting graduation students was necessary. This study conducted analyses against three methods: Naive Bayes, KNN, and neural network. This study aims to determine whether strategy is more appropriate for predicting graduation. The research also includes a comparison of the third approach mentioned and the best method produced, which is the K-NN method with a mark accuracy of 89% [13].

This article's writing structure consists of numerous components, including The introduction presents the study's context, phenomena, and aims. The research method section describes the analysis procedures employed, including data collection, processing, and evaluation. Furthermore, the results and discussion section present the findings and their interpretation. The debate delves into the ramifications of these findings, concluding with recommendations for educational institutions. The goal is to determine the level of satisfaction that UNUSA library visitors have with the services provided. Based on the preceding information, this study employs the support vector machine, decision tree, and K-nearest neighbour algorithms. This study aims to provide an overview of the role of libraries in the development of UNUSA library services and infrastructure.

II. METHODS

This research will evaluate satisfaction visitors to service library with compare level highest accuracy from the method used namely support vector machine, K-NN and decision tree. This research take data from UNUSA Library survey results via UNUSA Digilib.

Support Vector Machine (SVM) is a regression and classification prediction technology that is used to choose the optimum function separator and separate observations with various target variable values [14].

K-NN, or K-Nearest Neighbor, is an instance-based learning method in which training data is maintained in order to find classification for new records that have not yet been classified by comparing the most similarities in data mining [6].

A visual model called a decision tree works For simplify the manufacturing process decision rational [15]. Decision Tree can consider various choice breakdown problems, including those that affect mark risk and value information in choice decision. Every study need methodology research that can help researcher finish stage methodology study with explanation as following:



Figure 1. Overview of Research Methodology

1. Business Understanding
Do discussion with library For know common problems happened, clarify goals and limitations problem in data mining. With study This can add insight into analysis satisfaction visitors to services available at the library using machine learning Orange Data Mining [16]
2. Data Understanding
Initial data collection process and data description will used. Done analysis on the data namely service as a shared target to be of sufficient value , less satisfied, satisfied, and very satisfied. Review Text as a meta that contains comment from visitors library in form of suggestion or criticism.
3. Data Preparation

Data preparation in this process is the data that will be continued to the next process prepared, done processing of the dataset to be used For undergo a series preprocessing . Raw data changed become ready data used. Data cleaning is the process of removing data that is not relevant, invalid. Done data cleaning especially in the review text section as meta so that the results obtained are later can maximum [13].

4. Data Processing

At the stage This namely the process carried out start data understanding, cleaning so that the data is ready processed in accordance need with easy output understood. At this point, 224 data sets have been processed using Orange machine learning technologies, including support vector machine, decision tree approaches and K-nearest neighbor. Processing results can seen on the Test and Score widget that will appear display mark highest from third the method used [17].

5. Evaluation

Considerations and observations repeat done before reach results end. Evaluation use minimize existing error with see results level the resulting accuracy based on satisfaction data visitors library . Researcher do observation repeat Then do trial and error until obtained results the same ending and become a valid output [18].

III. RESULTS AND DISCUSSIONS

Analysis results For comparison support vector machine, decision tree, and K-nearest neighbor methods in classification evaluation service library as following : Business Understanding, at the stage this is initial data collection was carried out from satisfaction survey results through Digilib UNUSA, spreadsheet, as many as 224 data. Expected can add outlook library related satisfaction visitors to service library using machine learning. Data Understanding, at the stage data understanding is done dataset processing used at this stage Next . Analysis related variables in service data as a shared target to be of sufficient value, less satisfied, satisfied and very satisfied. Review text as a meta that contains comment from visitors library containing evaluation in form of suggestion words or criticism. Understanding the data use understand the data with good and according to the analysis. Data Preparation, data preparation in this process means prepare data for the next process . Done with change the new data collected become ready data used.

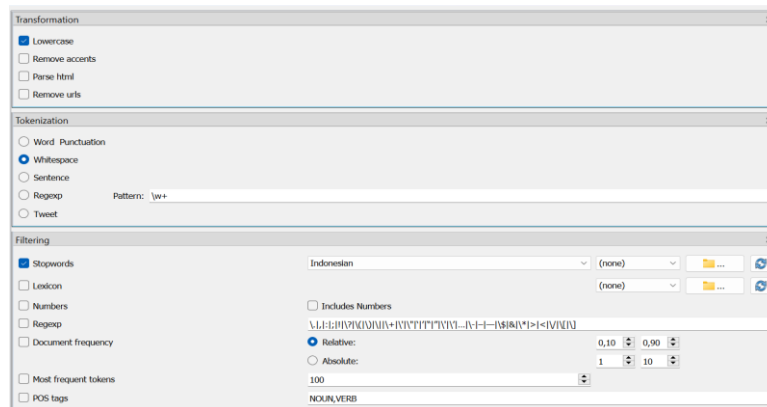


Figure 2. Preprocess in Orange Data Mining

Transformation, initial step in preprocessing text aiming For adjust the text format to make it more easy processed. Example transformation, one of which is Lowercase, namely change all over text become letter small For consistency. This is useful so that words like ‘Economy’ and ‘economics’ are considered the same by the system . The purpose of transformation is to reduce variation dam text and make sure text more uniform before analysis is done. The process of tokenization involves dividing text into discrete pieces known as tokens, which are typically words or other symbols. Whitespaces are break text based on space only. For example, ‘Sharia Economics’ will broken down become ‘Economy’, ‘Sharia’. The goal is for make it easier analysis with convert long text becomes piece more small that can analyzed. Filtering, process or delete certain words or symbols from text For increase quality of the data analyzed . Stopwords that is remove common words that are not own meaning specific in analysis such as ‘and’, ‘which’, ‘is’ in Indonesian. The purpose is remove elements that are not relevant or bother in the data so that the results more accurate.

Data Processing

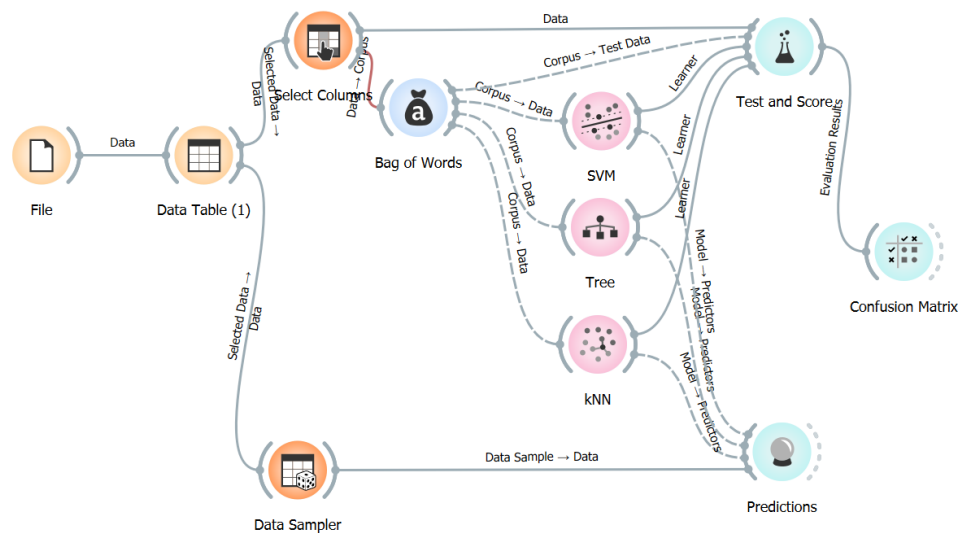


Figure 3. Orange tools plan calculates accuracy

Figure 3 shows a model that compares the algorithm being tested. With using the Test and Score widget, the success process through orange was done. Total of 224 data were processed as input for tested at stage testing and scoring, accordingly with algorithms discussed that is support vector machine, decision tree, and K- methods, nearest neighbor.

	Name	Type	Role	Values
1	Pelayanan yan...	C categorical	target	Cukup, Puas, Sangat Puas
2	Pelayanan yan...	S text	meta	Cukup, Puas, Sangat Puas
3	Lama ...	C categorical	feature	Cukup, Puas, Sangat Puas
4	Pelayanan yan...	S text	skip	Cukup, Puas, Sangat Puas
5	Kenyamanan ...	S text	skip	Cukup, Puas, Sangat Puas
6	Kebersihan ...	S text	skip	Cukup, Puas, Sangat Puas
7	Pencahayaan ...	S text	meta	Cukup, Puas, Sangat Puas

Figure 4. Data labels

The image above is stage the beginning of the data in the file widget will be display all data attributes , with emphasis on the attributes used For process the data. The data will shared become three values, namely enough, satisfied and very satisfied and changed become categorical . There are 224 data that have been cleaned and ready used.

Model	AUC	CA	Prec	Recall
SVM	0.967	0.915	0.917	0.915
Tree	0.888	0.902	0.902	0.902
kNN	0.954	0.906	0.906	0.906

Figure 5. Test and result scores

Figure 5 shows The "Test and Score" widget view located in the Orange tool . As shown in the figure , the K-fold cross validation method (K=10) is used. For operate testing in study. This widget show results from three compared algorithms in predict level satisfaction visitors library. This research only see AUC, CA, Precision, and Recall values. Evaluation results show that each algorithm own level different accuracy. K- Nearest Neighbor get value 0.954 and Support Vector Machine (SVM) gets mark highest 0.967. On the other hand , Decision Tree gets the lowest AUC value is 0.888, which is mark lowest. As a result, it can concluded that the most efficient SVM method For measure level satisfaction visitors library.

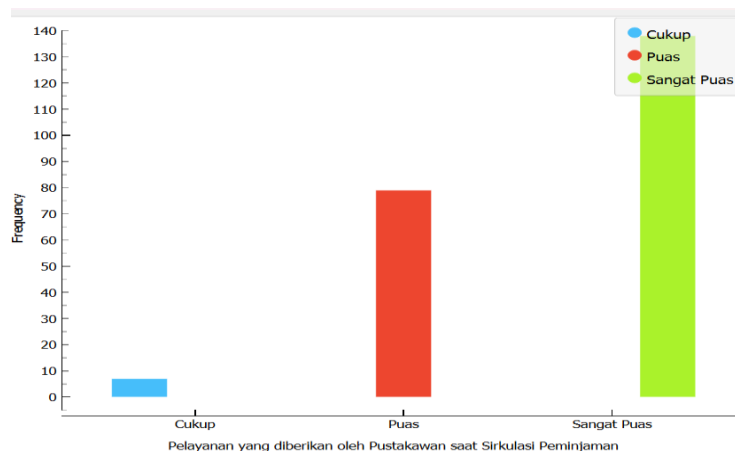


Figure 6. Distribution

Services that have been processed displayed in results distribution. Data is arranged based on category satisfaction visitors, namely enough, satisfied, and very satisfied. Visualization This displayed via the 'distribution' widget to make it easier reader understand results. Figure 6 shows that response visitors library dominated by the very satisfied category, which is quite tall compared to with category others. Visualization results This clear show that visitors library tend give very satisfied review about the services they provide accept, which shows that majority visitors own level very high satisfaction with the services they provide accept.

Evaluation

The Confusion Matrix method is used For count accuracy, recall, precision, and error rate [6]. Confusion matrix in study This covers choice like enough, less satisfied, satisfied, and very satisfied, which describes percentage from every category service.

		Predicted			
		Cukup	Puas	Sangat Puas	Σ
Actual	Cukup	100.0 %	2.3 %	0.8 %	8
	Puas	0.0 %	87.5 %	5.3 %	84
	Sangat Puas	0.0 %	10.2 %	93.9 %	132
Σ		5	88	131	224

Figure 7. SVM confusion matrix results

For describe distribution model prediction for every class, Figure 7 shows the confusion matrix widget which displays proportion prediction. Matrix prediction show that proportion prediction For very satisfied category on the Support Vector Machine algorithm, reaching 93.9%.

		Predicted			
		Cukup	Puas	Sangat Puas	Σ
Actual	Cukup	88.9 %	0.0 %	0.0 %	8
	Puas	11.1 %	90.8 %	10.1 %	84
	Sangat Puas	0.0 %	9.2 %	89.9 %	132
Σ		9	76	139	224

Figure 8. Decision Tree confusion matrix results

Figure 8 shows Confusion matrix widget display showing proportion prediction, namely percentage data prediction on the Decision Tree algorithm reached 89.9%.

		Predicted			
		Cukup	Puas	Sangat Puas	Σ
Actual	Cukup	75.0 %	1.2 %	0.7 %	8
	Puas	25.0 %	89.0 %	6.7 %	84
	Sangat Puas	0.0 %	9.8 %	92.5 %	132
Σ		8	82	134	224

Figure 9. K-NN confusion matrix results

Figure 9 shows Confusion matrix widget display showing proportion prediction, namely percentage data prediction on the K-NN algorithm reached 92.5%.

Model	CA	Prec	Recall
SVM	0.915	0.917	0.915
Tree	0.902	0.902	0.902
kNN	0.906	0.906	0.906

Figure 10. Comparison result method

Figure 10 shows that Support Vector Machine method is more accurate than the other two models, show that Support Vector Machine can become good choice for the classification process. However, the performance every method depends on the existing characteristics in each dataset used..

IV. CONCLUSIONS

Test results show that support vector machine is the most effective method For evaluate satisfaction visitors library to services, compared with decision tree and K-nearest neighbor. K-nearest neighbor get value 0.954 and support vector machine (SVM) gets mark highest 0.967. On the other hand, decision tree gets AUC value of 0.888, which is mark lowest. For further study, recommended To compare other methods and use tools besides Orange. Research results this also shows that part big visitors library very satisfied with services and facilities they provide accept. For get information new that can increase model performance, recommended for add relevant features and use method more data processing sophisticated. This will become recommendation for study upcoming.

V. REFERENCES

- [1] F. Tjiptasari, "Perkembangan Perpustakaan Tradisional menuju Digital," *Media Inf.*, vol. 31, no. no.1, pp. 33–44, 2022, doi: <https://doi.org/10.22146/mi.v31i1.4575>.
- [2] IFLA, "Libraries in the Digital Age," *IFLA*, 2021. [Online]. Available: <https://www.ifla.org/publications/libraries-as-media-redefining-a-library-in-the-digital-age/>
- [3] R. Masriyatun; Fatmawati, "Factors Affecting the User Satisfaction of Online Catalog Information Systems (UNSLA)," *EduLib*, vol. 14, no. 1, pp. 26–34, 2024, doi: 10.17509/edulib.v14i1.71178.
- [4] P. Adeniran, "User Satisfaction with Academic Library Services : Academic Staff and Students Perspective," *Int. J. Libr. Inf. Sci.*, vol. 3, no. 10, pp. 209–216, 2011, doi: 10.5897/IJLIS11.045.
- [5] N. Arsela, Feny; Nurhayani; Yasmin, "Mengukur Kualitas Layanan Perpustakaan di Universitas Potensi utama menggunakan metode Libqual+," *VISA J. Visions Ideas*, vol. 4, no. 3, pp. 948–958, 2024, doi: <https://doi.org/10.47467/visa.v4i3.2409>.
- [6] D. E. Muflikhah, Lailil; Ratnawati, *Buku Ajar Data Mining*. Malang: UB Press, 2018.
- [7] I. Pamungkas, Fajar Sodik; Prasetya Bayu Dwi; Kharisudin, "Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customer menggunakan Python," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 692–697, 2020, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/37875>
- [8] E. Alpaydin, *Introduction to Machine Learning*. MIT Press, 2020.
- [9] Z. Kohsasih, Kelvin Leonardo; Situmorang, "Analisis Perbandingan Algoritma C4.5 dan Naive Bayes Dalam Memprediksi Penyakit Cerebrovascular," *J. Inform.*, vol. 9, no. 1, pp. 13–17, 2022, [Online]. Available: https://www.researchgate.net/profile/Kelvin-Leonardi-Kohsasih/publication/359649144_Analisis_Perbandingan_Algoritma_C45_Dan_Naive_Bayes_Dalam_Memprediksi_Penyakit_Cerebrovascular/links/6246a5dd21077329f2e6c3b4/Analisis-Perbandingan-Algoritma-C45-Dan-Naive
- [10] H. Pahtoni, Tri Yuli; Jati, "Analisis Sentimen Data Twitter terkait ChatGPT menggunakan Orange Data Mining," *J. Teknologi Inf. dan Ilmu Komput.*, vol. 11, no. 2, 2024, doi: <https://doi.org/10.25126/jtiik.20241127276>.
- [11] M. Pranadjaya, Egipta; Pangestu, Evan Sudira; Sereati, Catherine Olivia; Octaviani, Sandra; Darmawan, "Perbandingan Algoritma Macine Learning menggunakan Orange Data Mining untuk Klasifikasi Jenis Kendaraan pada Sistem Tilang Digital," *J. Elektro*, vol. 17, no. 1, 2024, [Online]. Available: https://scholar.google.com/scholar?hl=id&as_sdt=0%2C5&q=artikel+dengan+orange+data+mining&oq=a
- [12] L. S. Hakim, Ichwanul; Harahap, "Analisis Sentimen terhadap iPhone 16 pada Data Twitter menggunakan Orange Data Mining," *J. Sains dan Teknol.*, vol. 5, no. 3, 2024, doi: <https://doi.org/10.3785/kohesi.v5i3.7040>.
- [13] F. Safitri, Dinda; Hilabi, Shofa Shofiah; Nurapriani, "Analisis Penggunaan Algoritma Klasifikasi dalam Prediksi Kelulusan Menggunakan Orange Data Mining," *J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 1, pp. 75–81, 2023, [Online]. Available: <https://jurnal.univrab.ac.id/index.php/rabit/article/view/3009/1366>
- [14] I. A. Idris, Irma Surya Kumala; Mustofa, Yasin Aril; Salihi, "Analisis sentimen terhadap penggunaan aplikasi shopee menggunakan algoritma support vector machine (SVM)," *JAMBURA J. Electr. Electron. Eng.*, vol. 5, no. 1, 2023, [Online]. Available: <https://ejurnal.ung.ac.id/index.php/jjee/article/view/16830>
- [15] In. R. Cahyaningtyas, Cristian; Nataliani, Yessica; Widiasari, "Analisis Sentimen pada Rating Aplikasi Shopee Menggunakan Decision Tree Berbasis SMOTE," *J. Teknol. Inf.*, vol. 18, no. 2, 2021, doi:

- <https://doi.org/10.24246/aiti.v18i2.173-184>.
- [16] N. H. Wahyuningsih, “Perbandingan Metode Klasifikasi dalam Analisis Sentimen Masyarakat terhadap Identitas Kependudukan Digital (IKD),” *J. Ilm. Penelit. dan Pembelajaran Inform.*, vol. 8, no. 4, 2023, doi: DOI: <https://doi.org/10.29100/jipi.v8i4.4155>.
 - [17] A. E. Ardiansyah, Fitrah Rizki; Firmansyah, “Implementasi Algoritma K-Means Clustering menggunakan orange untuk mengelompokkan penjualan Smartphone (studi kasus counter Mutiara cell),” *J. Soc. Sci. Res.*, vol. 4, no. 6, pp. 1371–1379, 2024, [Online]. Available: <https://j-innovative.org/index.php/Innovative/article/view/16415/11097>
 - [18] H. Hartono, Seno; Perwitasari, Anggi; Sujaini, “Komparasi Algoritma Nonparametrik untuk Klasifikasi Citra Wajah Berdasarkan Suku di Indonesia,” *J. Edukasi Penelit. Inform.*, vol. 6, no. 3, 2020, [Online]. Available: <https://jurnal.untan.ac.id/index.php/jepin/article/view/43268>