Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 7, No. 2, September 2024, pp. 561 – 569
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Analysis of The Level of Satisfaction of Electric Bus Passengers in Medan Using C5.0 Algorithm

[1]Azizah Oktarina Ritonga, [2]Sriani
[1,2]Study Program of Computer Science, Departement of Science and Technology,
State Islamic University of North Sumatera, Indonesia
Email: [1]azizahoktarinaritonga@gmail.com , [2]sriani@uinsu.ac.id

Article Info

Article history:
Received Aug 23th, 2024
Revised Sep 26th, 2024
Accepted Sep 30th, 2024

Keyword:
C5.0 Algortima
Data Mining
Decision Tree
Electric Bus
Satisfaction

ABSTRACT

The C5.0 method was successfully implemented in the analysis of passenger satisfaction level of Medan City Electric Bus using 500 passenger data divided into 70% for training data (350 data) and 30% for test data (150 data). The steps include creating a decision tree based on training data, where the model learns to identify patterns that affect passenger satisfaction. After the decision tree is formed, the model is tested using testing data to measure the prediction accuracy. The evaluation results show that the C5.0 model is able to classify the testing data effectively, providing an accurate picture of passenger satisfaction levels in the tested context. Based on manual calculations with 10 testing data against training data, the following prediction results are obtained: out of 10 testing data, the model predicts 7 data as Satisfied (satisfied) and 3 data as Dissatisfied (not satisfied). This result shows that the model managed to classify most of the testing data correctly, giving a positive indication of the accuracy and reliability of the model in identifying passenger satisfaction levels. Based on the evaluation results of the C5.0 model using R, the training data showed accuracy, precision, and recall each reached 100%. This indicates that the C5.0 model was perfectly successful in classifying all training data, with no errors in prediction or classification. This result confirms that the model is very effective and reliable in analyzing the satisfaction level of Medan City Electric Bus passengers, demonstrating its ability to provide accurate and consistent predictions.

Corresponding Author:
Azizah Oktarina Ritonga
Study Program of Computer Science, Departement of Science and Technology,
State Islamic University of North Sumatera, Indonesia
Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Indonesia.
Email: azizahoktarinaritonga@gmail.com

DOI: http://dx.doi.org/10.24014/ijaidm.v7i2.32785

1. INTRODUCTION

Medan is the capital of North Sumatra Province, Medan city is the third largest city in Indonesia after DKI Jakarta and Surabaya, often called a metropolitan city and its population is large, as one of the cities with a large area of course people need a lot of transportation use, Transportation is a very influential element in the wheels of the economy. All aspects of national life depend on this one sector, which functions as a driver, supporter and driver of economic growth. One aspect of transportation that concerns the livelihood of many people is public transportation. Public transportation is the transfer of people or goods from one place to another by using vehicles in urban areas related to fixed and regular routes. Public transportation that is widely used and needed by the community is land transportation services such as buses.

Satisfaction can be said to be a condition where someone gets something in accordance with their expectations. Where this satisfaction can be obtained, one of which is through the services provided by a service provider company. Services can be defined as activities that can be carried out or provided by individuals, groups or organizations directly or indirectly to meet service needs, services are activities that can be carried

out or provided by government and private parties which have the same goal of improving the welfare or rate of living of the community. Service can also be defined as a thing or action, treatment carried out with the aim of helping, helping or satisfying and providing what is needed by others.

According to [1] Transportation is one of the components that is very meaningful in improving the wheels of this economy is an inseparable and indispensable part of society in every day life. Public transportation has many advantages for those who use it. Not only does it provide safety at an affordable price, it also improves mobility by reducing congestion on the roads. Transportation has a huge influence on the economic sector.

Public transport is a necessity for people to support their daily activities and is an important part of overall urban development. Due to its large carrying capacity, public transport is essential for improving community mobility and reducing traffic congestion. In addition, using efficient public transportation can reduce the use of private cars. To improve the accessibility of public transport, additional improvements are needed. These include clarifying departure times, prohibiting public transport from picking up and dropping off passengers at certain locations, and improving services at certain locations, and improving terminal services, security, and convenience [2]. According to the Regulation of the Minister of Transportation of the Republic of Indonesia Number 26 of 2017 concerning the Implementation of Transportation of People by Public Motor Vehicles not on a Route, the definition of a bus car is a motorized vehicle for transportation of people who have seats for more than 8 people, including the driver, weighing more than 3,500 kg.

This Electric Bus is one part of Bus Rapid Transit (BRT). This Electric Bus is a Buy the Sevice program of the Medan mayor's government to develop public road transport in urban areas, using reliable cashless telecommunication technology to improve safety and comfort in traveling. The ultimate goal is to provide economical, practical, reliable, and convenient transportation for the community. Medan Mayor, North Sumatra, Bobby Nasution launched the operation of four units of Medan-Binjai-Deli Serdang (Mebidang) Bus Rapid Transit (BRT) service electric buses in this area. In addition, he continued, the Mebidang BRT service electric bus does not use the Medan City APBD with the number of electric corridor 1 fleets will increase periodically.

The electric bus corridor 1 has a route from J-City Housing Complex to Street of Karya Wisata, Street of AH Nasution, Street of Jamin Ginting, Street of Sudirman, Street of Diponegoro, and Street of Kapten Maulana Lubis. Then to Street of Balai Kota, Street of Putri Hijau, Street of Guru Patimpus, Street of Gatot Subroto, Street of Iskandar Muda, Jamin Ginting, Street of AH Nasution and back to Street of Karya Wisata. BRT Mebidang has a 21-kilometer track, connected to 31 bus stops with 17 routes reaching Medan, Binjai and Deli Serdang using a fleet of 515 buses supported by depots, bus stops and special bus lanes.

The main problem as one of the transportation services is whether the services provided satisfy the users. Therefore, this Electric Bus is required to maintain consumer confidence by providing facilities to improve service quality and increase customer satisfaction [3]. To analyze passenger satisfaction through the servqual dimension specifically seen from the attributes that have or have not met the expectations of the passengers. Then use the C5.0 Algortima method to find out what percentage of passenger satisfaction levels have been met.

This research applies the C5.0 algorithm method, a development of the C4.5 and ID3 algorithms. The C5.0 algorithm shows a higher level of accuracy in calculations compared to other algorithms. And it is able to create decision trees with more optimal accuracy because it continuously updates the weighting on the training data used as a reference, so that the decision tree is always updated using the latest training data. This ensures that the resulting decision tree model is always the latest [4].

Previous research conducted by Riyono, (2023) [5], this research focuses on analyzing the sentiments expressed by passengers through the Trans Metro Deli Medan social media account. and the current research focuses on analyzing the level of satisfaction of electric bus passengers in Medan City. The data comes from comments, reviews, or posts on social media related to Trans Metro Deli Medan while the data collection carried out by the current researcher is by questionnaire and interview. previous research using the Naïve Bayes Method Used in sentiment analysis to classify text based on sentiment polarity (positive, negative, or neutral). while the method used by the current researcher C5.0 algorithm Used in satisfaction analysis to create decision trees and predictive models from passenger satisfaction data. Thus, the two studies differ in terms of the focus of the analysis, both studies seek to measure the level of passenger satisfaction, but use different means to achieve this goal.

In the research of Nurfitrayani et al, (2023), where the calculation results of the two algorithms have an accuracy value of 91.89%, a precision value of 93.75%, a recall value of 96.77%, an F1 - Score value of 95.24%, an AUC value of 0.8172 and is included in the good classification category [6]. Research conducted by Asih et al, (2022) in which in this study, the C5.0 algorithm proved to be a fairly accurate algorithm because it produced an accuracy value of 0.74 or 74%. Experienced an increase from previous research which was 66%. For the average value of precision 56% and recall and f-measure equal 45%. In the accuracy value of precision,

recall, and f-Measure in the positive class has a high accuracy, it means that tweet data or twitter user opinions regarding the assistance of the Ministry of Education and Culture respond positively or are useful for them [7]. Research by Indriyani et al, (2022) which in this study aims to determine the level of consumer satisfaction with Pinkan Bakery & Cake bread products. By knowing the level of customer satisfaction, the shop can improve and increase customer satisfaction factors if there are weaknesses and shortcomings. The method used in this research with data mining is the C4.5 algorithm. The accuracy rate produced by the method is 93% [8].

2. MATERIAL AND METHOD

2.1 Data Mining

Data mining is a tool that allows users to quickly access large amounts of data as well as a process of extracting or extracting large, previously unknown, but understandable and useful data and information from large databases and is used to make very important business decisions [9]. Data Mining is defined as data mining or an attempt to extract valuable and useful information from very large databases. The most important thing in data mining techniques is the rule for finding high-frequency patterns among a collection of items called the Association Rules function [10]. In data mining, there are many algorithms/ methods/ techniques for finding knowledge or information. Each algorithm/ method/ technique has a different function and purpose [11].

Data classification is a process of finding properties that are common to a set of objects in a database and classifying them into different classes according to a defined classification model. The purpose of this classification is to find a model from the training set that distinguishes attributes into appropriate categories or classes, the model is then used to classify attributes whose classes are not previously known. Classification techniques are divided into several techniques [12]

Classification is a very important task in data mining. A classification is formed from a set of training data and has predefined classes. Classification techniques are systematic approaches to building classification models from a set of input data [13]. For example, decision tree techniques, artificial neural networks (backpropagation), techniques based on the concept of mining association rules and other techniques (K-Nearest neighbor, genetic algorithms, techniques with rough and fuzzy set approaches) [14]. Each technique also has its own advantages and disadvantages. Data with certain profiles may be most optimal if classified with certain techniques or in other words, certain profiles can support the utilization of the advantages of this technique [15].

2.2 Decision Tree

Decision tree is an application of clustering techniques which is the process of determining an objective function that maps each set of behaviors to a group of previously defined classes. Decision trees can find hidden relationships between a number of input values and an output value [16]. And decision trees combine data mining and data modeling. With its ability to control complex decision-making processes to be simpler. And decision making is a thought process in the context of solving a problem to get the final result to be implemented [17].

The decision tree process is to convert the data form (table) into a tree model, convert the tree model into rules and simplify the rules. In forming a decision tree with the C5.0 algorithm, entropy and information gain are used to determine the root node. The gain with the highest value will be the root node of the smallest entropy of each attribute [18].

2.3 C5.0 Algorithm

The C5.0 algorithm is a refinement algorithm of the C4.5 algorithm. The process is almost the same as the C4.5 algorithm, but the C5.0 algorithm has advantages over the C4.5 algorithm [19]. The C5.0 algorithm is a refinement of previous algorithms formed by Ross Quinlan in 1987, namely ID3 and C4.5. In selecting attributes to break objects into several classes, the attribute that produces the greatest information gain will be selected as the root for the next node [20]. This algorithm starts with all the data being the root of the decision tree while the selected attribute will be the divisor for the sample. C5.0 produces a tree with a variable number of branches per node. C5.0 requires continuous variables the same as CART does, but for categorical variables C5.0 treats categorical variable values as splitters. In the process of forming the C5.0 algorithm classification tree, the first step is to determine the root node, then determine the branches for each node[21].

The C5.0 algorithm builds a decision tree with steps similar to the C4.5 algorithm. In both algorithms, the initial steps taken are calculating fairness occurrence, entropy, and information gain [22]. However, the C5.0 algorithm does not stop at the calculation of information gain, but instead continues by calculating the gain ratio using information gain and entropy. Gain ratio is used to select attributes at each node and tree. The attribute with the largest gain ratio will become the parent for the next node [23]. To find the entropy value, the formula is used:

$$Entropy\ (S) = \sum_{i=1}^{n} -pi * log2\ pi \quad (1)$$

To find the value of information gain:

$$InformationGain\ (S, A) = Entropy\ (S) - \sum_{i=1} n\ \frac{|Si|}{|Si|} \quad (2)$$

2.4 Data Sample

Data for the analysis of passenger satisfaction level of Medan City Electric Bus was collected through interviews and questionnaires involving 500 passengers, each identified by a unique ID. The data collected covered four main variables: Facility, Price, Location, and Service. Facility variables were categorized as complete or incomplete; Price as cheap or expensive; Location as strategic or not strategic; and Service as good or not good. The output variable, Passenger Satisfaction, was categorized as satisfied or dissatisfied. This data was analyzed using the C5.0 Algorithm to identify the relationship between the input variables and the level of passenger satisfaction. This analysis aims to understand the key factors affecting passenger satisfaction and provide insights to improve trolleybus services in Medan City. The complete data of this survey can be seen in Table 1.

Table 1. Data Sample

No	Passenger ID	Facility	Price	Location	Service	Satisfaction
1	PNP-0001	Incompleted	Costly	Strategic	Good	Satisfied
2	PNP-0002	Incompleted	Costly	Unstrategic	Not Good	Dissatisfied
3	PNP-0003	Incompleted	Costly	Strategic	Not Good	Dissatisfied
…	PNP-0004	Incompleted	Costly	Strategic	Not Good	Dissatisfied
500	PNP-0500	Completed	Costly	Strategic	Not Good	Dissatisfied

2.5 Flowchart Methodology

In conducting research, a systematically arranged investigation process is needed which is aimed at conveying information in solving a problem. Therefore, a research framework is needed so that the researcher runs according to what is expected. This research framework is a preliminary survey to identify passenger satisfaction with Medan City Electric Bus. The research methodology can view figure 1.

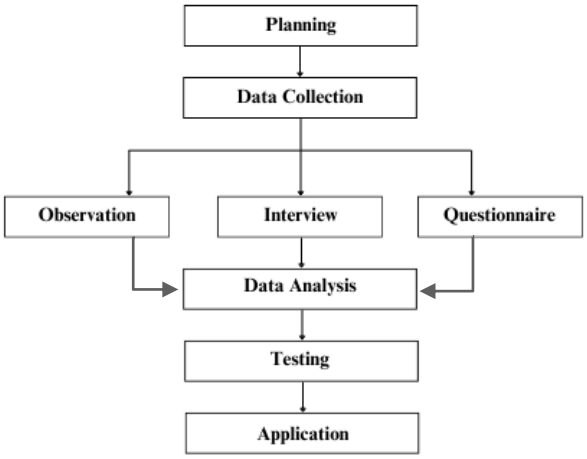

Figure 1. Research Methodology

1. Planning
   In the initial process of research carried out by determining the topic to be discussed. In this research, the topic to be discussed is related to the level of passenger satisfaction regarding the Medan city electric bus. This bus has been operating since January 2024.

2. Data Collection
   To obtain data collection and information in data processing, data collection is necessary, data collection is carried out in 3 ways, namely, interviews, observations and distributing questionnaires or questionnaires. This research was conducted by distributing questionnaires to passengers of electric buses in the city of Medan.

a. Interview
It is a data collection technique carried out by researchers with sources directly, this technique can be done if researchers want to know things from respondents that are more in-depth.

b. Questionnaire
Data collection techniques that are carried out by giving a set of questions using a form link or can be in writing to respondents to answer. The questionnaire is a very efficient data collection technique if the researcher is certain about the variables to be measured and knows what can be expected from the respondent. This study uses a sample draw, the number of samples must be representative (representative) so that the calculation does not require a sample size table, and the research results can be generalised, so that it can be done with simple formulas and calculations. Where the opinion of Hair at all says the minimum sample size is many research indicators multiplied by five or ten.

c. Observation
Observation is a data collection technique by making direct observations of the object under study so that researchers obtain accurate results.

3. Data Analysis
Data analysis is the process of analysing, interpreting and presenting data that has been collected during research. In this study, data analysis is the preparation of respondent data on electric bus users obtained from the results of the questionnaire using R. After the data is collected, the next thing to do is to analyse the data.

4. Testing
The testing stage aims to find out whether the system process that has been designed is in accordance with its function. Testing is carried out on the pattern of passenger satisfaction levels where the input data used is based on the results of the calculation value of the data, namely observations and interviews or questionnaires. And the calculation process uses training data and testing data. After the calculation process, the output is obtained which is a decision tree rule. Testing is done to find out whether the system runs properly in accordance with the objectives that have been made.

5. Application
At this stage of application, the service prediction process will be carried out with the results of predicting very satisfied, satisfied, moderately satisfied and dissatisfied. For data samples taken as many as 500 respondents. At this stage the data will also be processed using the c5.0 algorithm calculation. Followed by applying it to the R language.

3. RESULTS AND ANALYSIS

The dataset of 500 entries is divided using the train-test split technique, with a proportion of 70% for training data (350 data) and 30% for test data (150 data). This split is important to ensure that the developed model has enough data to learn as well as to be tested properly. Training data is used to train the model so that it can understand the relationship patterns between features (such as Facility, Price, Location, and Service) and the target (Satisfaction). Meanwhile, test data serves to measure the model's ability to predict new data that has never been seen before, ensuring the model does not experience overfitting and can generalize patterns well. By doing this division, the model is expected to be able to work effectively when faced with new data, and evaluation of model performance can be done fairly through various evaluation metrics such as accuracy. Furthermore, this dataset will be implemented into the C5.0 algorithm, which will build a decision tree based on the training data and generate classification rules that can be used to predict passenger satisfaction on the test data. The C5.0 algorithm is very efficient in handling categorical data, such as that contained in this dataset, and is able to produce interpretative and accurate models to predict the target variable.

In the data analysis of the satisfaction level of Medan City Electric Bus passengers, the passenger data identified with a unique ID is processed using the C5.0 Algorithm. This data includes four main input variables: Facility (complete or incomplete), Price (cheap or expensive), Location (strategic or not strategic), and Service (good or not good), with the output variable being Passenger Satisfaction (satisfied or dissatisfied). Prior to analysis, the data will be cleaned and processed to address any missing values or outliers' that may exist. After that, the data is divided into training set and testing set to build and test the classification model. The C5.0 algorithm will be used to generate a decision tree that describes the relationship between input and output variables, allowing researchers to identify the most significant factors affecting passenger satisfaction. Model

validation is performed using cross-validation techniques to ensure the reliability of the results. The performance of the model will be evaluated based on metrics such as accuracy, precision, and recall. The results of this analysis will provide in-depth insights into service areas that need to be improved and assist trolleybus service managers in making more informed decisions to improve overall passenger satisfaction.

The implementation stage of the C5.0 algorithm for analyzing data on the level of satisfaction of Medan City Electric Bus passengers starts with data collection. Then the data that has been collected will be analyzed with the steps of selecting the best attributes for data separation based on the size of the specified criteria, such as gain ratio or gain information, to divide the dataset into tree branches. This process continues recursively by building additional branches until the data within each branch becomes homogeneous or meets the stopping criteria. Once the tree is built, the next step is to prune the tree to reduce complexity and overfitting by simplifying the tree structure without compromising accuracy. Finally, the pruned decision tree is used to classify new data, by referring to the rules obtained from the tree structure.

If the highest gain ratio is found in the attributes "Non-Strategic Service, Low Price, and Non-Strategic Location" with a gain ratio value of 1, then a perfect separation has been achieved. All data in the subset has been homogeneously divided into their respective classes, so no further splitting is required. Thus, the decision tree formation process on this branch has been completed.

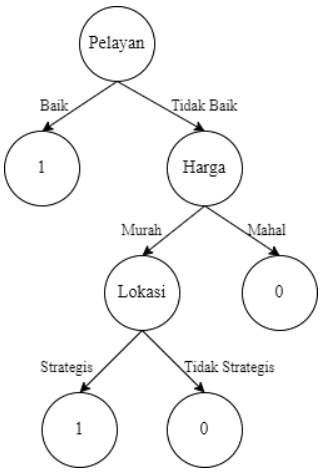

Figure 2. Final Result Decision Tree

The decision tree above has been obtained, the next step is pruning. In the application of the C5.0 algorithm, the decision tree formation process produces a decision tree that is already quite simple and easy to understand. Although the algorithm provides a pruning option to reduce complexity and prevent overfitting, in this case, pruning is not necessary because the decision tree produced is optimal. As such, the resulting model was able to provide accurate predictions without the need for additional steps to simplify the tree structure.

At this stage, the C5.0 model will be implemented using R software, where the data used is the same as that used in the previous manual calculation. The purpose of this implementation is to validate the results of manual calculations with the results generated by the automated model, as well as to identify patterns or decision rules that are more efficient in analyzing the satisfaction level of electric bus passengers in Medan City. This implementation also allows for comparisons between manual and automated methods in terms of speed, accuracy, and interpretation of results, thus providing greater insight into the effectiveness of the C5.0 model in data analysis.

This test is carried out by applying the prediction results of the decision tree to the testing data that has been prepared. The purpose of testing this model is to ensure that the model is able to classify new data appropriately, identify the variables that have the most influence on the level of passenger satisfaction, and provide insights that can be used in making more effective decisions related to improving electric bus services in Medan City.

Table 2. Prediction Result Display

No	Passenger ID	Facility	Price	Location	Service	Satisfaction
1	PNP-0001	Incompleted	Costly	Strategic	Good	Satisfied
2	PNP-0002	Incompleted	Costly	Unstrategic	Not Good	Dissatisfied
3	PNP-0003	Incompleted	Costly	Strategic	Not Good	Dissatisfied
…	PNP-0004	Incompleted	Costly	Strategic	Not Good	Dissatisfied
500	PNP-0500	Completed	Costly	Strategic	Not Good	Dissatisfied

To assess how well the model predicts outcomes when applied to test data. In this study, the evaluation was performed using the confusion matrix with 350 training data in R, providing an accurate picture of the accuracy and prediction error of the model. The following is the result of the evaluation display on training data with the C5.0 method using R:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 126 & 0 \\ 1 & 0 & 224 \end{bmatrix}$$

$$Accuracy = \frac{number\ of\ correctly\ predicted\ data}{total\ number\ of\ predictions}\ x\ 100\%$$

$$Accuracy = \frac{350}{350} x100\%$$

$$Accuracy = 100\%$$

$$Precission = \frac{many\ correct\ predictions}{many\ correct\ predictions + many\ predictions\ were\ wrong}\ x\ 100\%$$

$$Precission = \frac{350}{350 + 0} x100\%$$

$$Precission = 100\%$$

$$Recall = \frac{many\ correct\ predictions}{many\ correct\ predictions + number\ of\ prediction\ data\ error}\ x\ 100\%$$

$$Recall = \frac{350}{350 + 0} x100\%$$

$$Recall = 100\%$$

Blackbox testing next focuses on assessing whether the program units meet the specified functional requirements, based on system specifications such as functions, menu options, and compatibility of the models used in this study. This approach involves the execution of the developed program and observation to ensure its conformance to the desired requirements. The following table 3 displays the system performance metrics.

Table 3. Blackbox Testing

No	Scenario	Expected Results	Test Results	Description
1	Inputting Training data processed using R.	Users can input Training data processed using R.	The user has successfully input Training Data processed using R according to the input.	Validation
2	Inputting Testing data and processed using R	User can input Testing Data and processed by R.	The user has successfully input Training Data and processed by RapidMiner according to the input.	Validation
3	Implementing the C5.0 model Using R	Users can implement the C5.0 model using R.	Users have successfully implemented the C5.0 model using R.	Validation
4	Training data can generate or create a Decision Tree.	The system can process Training Data to generate or create a Decision Tree.	The system successfully processes the training data to generate or create a Decision Tree.	Validation
5	Testing data is predictable.	The system can process the Testing data to generate predictions.	The system successfully processes the Testing data to generate predictions.	Validation

4. CONCLUSION

The C5.0 method was successfully implemented in the analysis of passenger satisfaction level of Medan City Electric Bus using 500 passenger data divided into 70% for training data (350 data) and 30% for test data (150 data). The steps include creating a decision tree based on training data, where the model learns to identify patterns that affect passenger satisfaction. After the decision tree is formed, the model is tested using testing data to measure the prediction accuracy. The evaluation results show that the C5.0 model is able to classify the testing data effectively, providing an accurate picture of passenger satisfaction levels in the tested context. Based on manual calculations with 10 testing data against training data, the following prediction results are obtained: out of 10 testing data, the model predicts 7 data as Satisfied (satisfied) and 3 data as Dissatisfied

(not satisfied). This result shows that the model managed to classify most of the testing data correctly, giving a positive indication of the accuracy and reliability of the model in identifying passenger satisfaction levels. Based on the evaluation results of the C5.0 model using R, the training data showed accuracy, precision, and recall each reached 100%. This indicates that the C5.0 model was perfectly successful in classifying all training data, with no errors in prediction or classification. This result confirms that the model is very effective and reliable in analyzing the satisfaction level of Medan City Electric Bus passengers, demonstrating its ability to provide accurate and consistent predictions. Although the C5.0 model shows excellent performance in this study, exploration and comparison with other classification models. Parameter adjustment and alternative model analysis can further improve the accuracy and efficiency of the model in the context of real applications.

REFERENCES

[1] S. A. Sianipar and H. Herman, "PENGARUH KUALITAS PELAYANAN DAN FASILITAS TERHADAP KEPUASAN PENGGUNA JASA TRANSPORTASI UMUM TRANS BATAM," vol. 4, no. 3, 2020.

[2] A. Soimun, A. Prima Gilang Rupaka, N. Wayan Putu Sueni, and Hendrialdi, "Identifikasi Aksesibilitas Angkutan Umum Dan Terminal Kawasan Metropolitan Sarbagita," *j.keselam.transportasiStreet of*, vol. 8, no. 1, pp. 62–76, Jun. 2021, doi: 10.46447/ktj.v8i1.309.

[3] M. H. Tinambunan, A. Hasibuan, S. Wahyuni, and A. S. Wibowo, "KLASIFIKASI TINGKAT KEPUASAN MAHASISWA TERHADAP FASILITAS PADA FTIK UNIVERSITAS DHARMAWANGSA MEDAN DENGAN ALGORITHM NAIVE BAYES," *BN*, vol. 6, no. 1, pp. 208–215, Jun. 2023, doi: 10.46576/bn.v6i1.3356.

[4] E. Hartati, "Penggunaan Klasifikasi Sayur Segar dan Sayur Busuk Mengunakan Algorithm Support Vector Machine," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. Vol 07, No 03, pp. 678–687, Dec. 2020.

[5] J. Riyono, A. L. R. Putri, and C. E. Pujiastuti, "Early Detection of COVID-19 Disease Based on Behavioral Parameters and Symptoms Using Algorithm-C5.0," *IJAIDM*, vol. 6, no. 1, p. 47, Apr. 2023, doi: 10.24014/ijaidm.v6i1.22074.

[6] N. Nurfitrayani, I. Islamiyah, and A. P. Azam Masa, "Penerapan Klasifikasi Algorithm C4.5 Dan Algorithm C5.0 Untuk Mengetahui Tingkat Kepuasan Mahasiswa Terhadap Website Sistem Informasi Terpadu Layanan Program Studi (SIPLO)," *mib*, vol. 7, no. 4, p. 1877, Oct. 2023, doi: 10.30865/mib.v7i4.6433.

[7] N. M. Asih, J. H. Jaman, and Y. Umaidah, "Analisis Sentimen Terhadap Bantuan Kuota Internet Dari Kemendikbud Dimasa Covid-19 Menggunakan Algorithm C5.0," *INTECOMS*, vol. 5, no. 2, pp. 1–9, Sep. 2022, doi: 10.31539/intecoms.v5i2.2793.

[8] Novita Indriyani, Heru Satria Tambunan, and Zulia Almaida Siregar, "Analisis Faktor Kepuasan Konsumen Terhadap Produk Roti Pinkan Bakery & Cake dengan Algorithm C4.5," *JURRITEK*, vol. 1, no. 2, pp. 76–90, Oct. 2022, doi: 10.55606/jurritek.v1i2.413.

[9] L. S. Hasibuan, S. Nabila, N. Hudachair, and M. A. Istiadi, "Evaluation of F-Measure and Feature Analysis of C5.0 Implementation on Single Nucleotide Polymorphism Calling," *IJAIDM*, vol. 1, no. 1, p. 1, Mar. 2018, doi: 10.24014/ijaidm.v1i1.4616.

[10] D. Selywita and Hamdani, "SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN SUPPLIER OBAT MENGGUNAKAN METODE FUZZY TSUKAMOTO," *Jurnal Ilmiah SISFOTENIKA*, vol. 3, no. 1, 2023.

[11] M. A. Hidayatuloh, K. P. Kartika, and D. F. H. Permadi, "Implementasi Algorithm C4.5 Untuk Memprediksi Capaian Pembelajaran Daring (Studi Kasus Siswa MAN 3 Blitar)," *algoritme*, vol. 3, no. 1, pp. 33–47, Oct. 2022, doi: 10.35957/algoritme.v3i1.3292.

[12] R. A. Manullang and F. A. Sianturi, "Penerapan Algorithm K-Nearest Neighbour Untuk Memprediksi Kelulusan Mahasiswa," 2021.

[13] J. Jasmir, X. Sika, M. Mulyadi, and R. Amelia, "Klasifikasi Kelayakan Pemberian Kredit Pada Calon Debitur Menggunakan Naïve Bayes," *Jur. Ris. Kom.*, vol. 9, no. 6, p. 1833, Dec. 2022, doi: 10.30865/jurikom.v9i6.5131.

[14] A. H. Lubis, L. P. A. Lubis, and Sriani, "Sentiment analysis on twitter about the death penalty using the support vector machine method," *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, vol. 11, no. 2, pp. 312–321, 2024, doi: 10.37373.

[15] B. Sugianto and G. P. Utama, "IMPLEMENTASI ALGORITHM PATHFINDING DAN DECISION TREE DALAM PEMBUATAN VIDEO GAME BERGENRE THIRD PERSON SHOOTER," *SKANIKA Budi Luhur*, vol. 4, no. 2, pp. 7–14, Jul. 2021, doi: 10.36080/skanika.v4i2.1825.

[16] R. A. Putra, M. A. M. Putri, S. M. Sinaga, S. F. Octavia, and R. C. Rachman, "Implementation of Association Rules Algorithm to Identify Popular Topping Combinations in Orders," *PREDATECS*, vol. 1, no. 2, pp. 95–101, Feb. 2024, doi: 10.57152/predatecs.v1i2.863.

[17] M. R. Anugrah, N. A. Al-Qadr, N. Nazira, and N. Ihza, "Implementation of C4.5 and Support Vector Machine (SVM) Algorithm for Classification of Coronary Heart Disease," *PREDATECS*, vol. 1, no. 1, pp. 20–25, Jul. 2023, doi: 10.57152/predatecs.v1i1.805.

[18] A. I. Putri *et al.*, "Implementation of K-Nearest Neighbors, Naïve Bayes Classifier, Support Vector Machine and Decision Tree Algorithms for Obesity Risk Prediction," *PREDATECS*, vol. 2, no. 1, pp. 26–33, Apr. 2024, doi: 10.57152/predatecs.v2i1.1110.

[19] M. S. Sungkar and M. T. Qurohman, "Penerapan Algorithm C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer," *mib*, vol. 5, no. 3, p. 1166, Jul. 2021, doi: 10.30865/mib.v5i3.3116.

[20] S. Sinaga and A. M. Husein, "Penerapan Algorithm Apriori dalam Data Mining untuk Memprediksi Pola Pengunjung pada Objek Wisata Kabupaten Karo," *JUTIKOMP*, vol. 2, no. 1, pp. 49–54, Apr. 2019, doi: 10.34012/jutikomp.v2i1.461.
[21] R. W. Pratiwi, S. F. H, D. Dairoh, D. I. Af'idah, Q. R. A, and A. G. F, "Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)," *INISTA*, vol. 4, no. 1, pp. 40–46, Dec. 2021, doi: 10.20895/inista.v4i1.387.
[22] R. Pratiwi, M. N. Hayati, and S. Prangga, "PERBANDINGAN KLASIFIKASI ALGORITHM C5.0 DENGAN CLASSIFICATION AND REGRESSION TREE (STUDI KASUS : DATA SOSIAL KEPALA KELUARGA MASYARAKAT DESA TELUK BARU KECAMATAN MUARA ANCALONG TAHUN 2019)," *BAREKENG*, vol. 14, no. 2, pp. 273–284, Sep. 2020, doi: 10.30598/barekengvol14iss2pp273-284.
[23] A. Damayanti, "SISTEM INFORMASI ADMINISTRASI KEPENDUDUKAN DENGAN ALGORITHM C5.0 UNTUK KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN (PKH) (STUDI KASUS: DESA PERNING, KECAMATAN JATIKALEN, KABUPATEN NGANJUK)," vol. 8, no. 1, 2024.

BIBLIOGRAPHY OF AUTHORS

Azizah Oktarina Ritonga is computer science student at UINSU who started her undergraduate education in 2020. This research is a form of final assignment to complete his undergraduate study period.

Sriani is a lecturer at the Faculty of Science and Technology at UINSU. Who is also the supervisor of Azizah Oktarina Ritonga in completing this final project research. Sriani has more than 20 articles related to computer science.