Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 7, No. 2, September 2024, pp. 570 – 578
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Classification of The Level of Public Satisfaction With the Use of Water Tourism Jetski in Balai Ujung Tanjung Using the Naïve Bayes Algorithm

[1]Nursalimah Isnaina Fatwa, [2]Rakhmat Kurniawan
[1,2]Departement of Science and Technology, State Islamic University of North Sumatera, Indonesia
Email: [1]nursalimah28@gmail.com, [2]rakhmat.kr@uinsu.ac.id

Article Info

Article history:
Received Aug 22th, 2024
Revised Sept 25th, 2024
Accepted Oct 4th, 2024

Keyword:
Classification
Data Mining
Machine Learning
Naïve Bayes Algorithm
Python

ABSTRACT

Jetski water tourism is one of the attractions that is often visited by the public compared to other attractions. One of the factors causing this is because there is no fee charged to visitors. The source of funds used in this tourist attraction is from the local government budget. Be it in terms of assessment to improve facilities, or even comments on whether the Jetski Water Tourism facility is good or bad. Certainly, with the public comments, it will help the government in improving its services to the community, especially in the management of this water tourism Jetski.The sentiment data collected from visitors to this Water Tourism Jetski can be used as a benchmark for the government in improving this Water Tourism Jetski facility. Both in terms of scope and the Jetski media used. By knowing the responses and comments of the community regarding Jetski Wisata Air, the government can evaluate in order to support visitor satisfaction and so that Jetski Wisata Air can last long and compete with other tourist attractions. The Naïve Bayes Algorithm has often been used in a study in the form of sentiment analysis. The Naïve Bayes model shows that the level of public satisfaction with Jetski Water Tourism in Ujung Tanjung Hall, Tanjungbalai City can be predicted with an accuracy of 75%. This indicates that the model is quite effective in identifying the level of user satisfaction, although there is a 25% possibility of inaccuracy in prediction. With this accuracy, the model can provide useful insights for the evaluation and improvement of jetski tourism services, but it should be considered to conduct further analysis to improve accuracy and get a more comprehensive picture of community satisfaction.

Corresponding Author:
Nursalimah Isnaini Fatwa
Departement of Science and Technology, Study Program of Computer Science
State Islamic University of North Sumatera,
Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Indonesia.
Email: nursalimah28@gmail.com

DOI: http://dx.doi.org/10.24014/ijaidm.v7i2.32761

1. INTRODUCTION

Balai Di Ujung Tanjung, that's what the people of Tanjungbalai call a replica building of approximately 20 square meters. This hall is located on the cape, the confluence of 2 rivers namely Asahan and Silau. It is said that this hall is the forerunner of the naming of Tanjungbalai City. It is said that in 1620, the son of Sultan Iskandar Muda (Sultan of the Kingdom of Aceh) named Sultan Abdul Jalil Rahmadsyah was crowned the first sultan of the Asahan Kingdom after the death of his father on December 27, 1620.

Currently, Balai Di Ujung Tanjung is only used when the Tanjungbalai anniversary is held, namely as a place for the traditional procession of the Malay kingdom which is one part of the series of events. This building was inaugurated as a cultural heritage building on November 21, 2021. As one of the tourist attractants, around Ujung Tanjung Hall there is also a Jetski water tour that is crowded with visitors, especially on holidays and weekends. Jetski is a small and fast-moving water vehicle, usually used for recreation on the water. Jetskis

are usually powered by a powerful gasoline engine and equipped with a rudder that can be controlled by the rider. They usually have an aerodynamic and lightweight design, with a shape similar to a motorcycle, but with a jet propulsion system that allows them to accelerate quickly and turn agilely on the water surface. Jetskis are often used for various water activities such as frolicking on the beach, water sports, or even for water rescue missions. They can accommodate one or two people, depending on the model, and are usually equipped with safety systems such as safety harnesses and fire extinguishing systems.

Jetski water tourism is one of the attractions that is often visited by the public compared to other attractions. One of the factors causing this is because there is no fee charged to visitors. The source of funds used in this tourist attraction is from the local government budget. Although classified as a free tourist attraction, the facilities and comfort at the location are well maintained.With the crowds of visitors to Jetski water tourism, there are certainly several opinions and assessments of the free facilities provided by this government. Be it in terms of assessment to improve facilities, or even comments on whether the Jetski Water Tourism facility is good or bad. Certainly with the public comments, it will help the government in improving its services to the community, especially in the management of this water tourism Jetski.

The sentiment data collected from visitors to this Water Tourism Jetski can be used as a benchmark for the government in improving this Water Tourism Jetski facility. Both in terms of scope and the Jetski media used. By knowing the responses and comments of the community regarding Jetski Wisata Air, the government can evaluate in order to support visitor satisfaction and so that Jetski Wisata Air can last long and compete with other tourist attractions. The Naïve Bayes Algorithm has often been used in a study in the form of sentiment analysis [1].

As research conducted by Roy Hendra Tinambunan, Julia Titaleym and Charles E. Mongi (2022), about Naive Bayes Classification in Predicting Student Satisfaction Levels Towards Lecturer Teaching in the Mathematics Study Program FMIPA Sam Ratulangi University Manado. Of all the data, 200 data are used as training data and 100 data are used as testing data. Of the 50 training data with GOOD classification as much as 46 data and NOT GOOD as much as 4 data. Testing using R-Studio software with the Naive Bayes method successfully predicts classification with accuracy 0.96 or 96% and Sensitivity 100% [2].

And also research conducted by Triase (2022), on the Use of the Naïve Bayes Algorithm in Measuring User Satisfaction with the Online System of Adventist University of Indonesia, where the Naïve Bayes Algorithm successfully classifies 122 data correctly from 150 training data, and shows that the classification of determining user satisfaction using the Naïve Bayes Algorithm gets the greatest accuracy value with training data testing, namely with a percentage of 81.3% [3].

Data mining is a process of data collection and data processing which will then be extracted into important information from the data. This process can be done using software that is certainly assisted by calculations such as systematics, mathematics or Artificial Intelligence (AI) technology which is a field of various scientific fields. In data mining, there are several methods that can be used and also applied to life to analyze broad knowledge easily for many people, one of the methods is classification [4]. Data mining, in simple terms, is an extraction step to obtain important information that is implicit and unknown. In addition, data mining has a relationship with various fields including statistics, machine learning, pattern recognition, computing algorithms, database technology, and high performance computing [5].

A collection of techniques that allow users to implement algorithms that are adaptable to automatically organize input data based on similar features and make predictions is known as machine learning [6]. In the past few decades, machine learning has become one of the most common tools used for any job that requires information extraction from large data sets [7]. The term Machine Learning itself refers to an automated process that detects important patterns in data. Scikit-Learn is one of the scientific libraries of choice for algorithm development [8] Scikit-Learn provides various implementations of well-known machine learning algorithms, and is designed to be easy to use [9].

Classification is a process to classify and predict new data from an attribute or variable based on the standards of existing data. Classification has two types of data, namely test data and training data. Training data is data that is based on facts, for example positive data and negative data, which also serves to form a probability table. Meanwhile, test data is data that has been classed or labeled and serves to test or calculate the accuracy value of the data or probability table that has been formed [10].

The results of research conducted by Roy Hendra Tinambunan, Julia Titaleym and Charles E. Mongi (2022), on Naive Bayes Classification in Predicting the Level of Student Satisfaction with Lecturer Teaching in the Mathematics Study Program FMIPA Sam Ratulangi University Manado, where the Naive Bayes Algorithm can be used as one of the methods for classification in determining student satisfaction with the way lecturers teach in the Mathematics Study Program FMIPA Unsrat with the parameters used in this study namely Communication, Building a Learning Atmosphere, Assessment of Students and Delivery of Material. Of all the data, 200 data are used as training data and 100 data are used as testing data. Of the 50 training data with GOOD classification as much as 46 data and NOT GOOD as much as 4 data. Testing using R-Studio software

with the Naive Bayes method successfully predicts classification with accuracy 0.96 or 96% and Sensitivity 100% [2].

The results of research conducted by Cahya Dewi A, Yusril Rahman, and Qurratul Aini (2022), on Shopee Food Service User Satisfaction Using the Naive Bayes Algorithm, which based on the results of the research hypothesis test, it was found that there was a simultaneous influence between application appearance, number of drivers, discounts and payment methods of 0.497 on user satisfaction. This shows that the ability of the independent factors in this study to influence the dependent variable is 49.7%, while the remaining 50.3% is explained by other variables. Thus, the prediction model proposed is included in the moderate category. In addition, the results of the NBC performance test with Rapid Miner obtained an accuracy rate of 100% when the training and testing data were 90% and 10% [11].

The results of research conducted by Gusticentiedina, Muhammad Siddik, and Yenny Desnelita (2020), on the Application of Naïve Bayes to Predict Student Satisfaction Levels with Academic Services, where from the results of the final test conducted from questionnaire sample data with attributes or indicators of tangible, reliability, responsiveness, assurance, and empathy using the Naïve Bayes Algorithm obtained an accuracy rate of 96.71% with a precision value of 96.15% and a recall value of 98.43%. Based on the test results, the Naïve Bayes Algorithm can be recommended to predict the level of student satisfaction with academic services at universities, because the precision and recall values are high [12].

With the various descriptions that have been conveyed above, the author will conduct research as well as fulfill the final project with the title "Classification of Community Satisfaction Levels Towards the Use of Water Tourism Jetski at Balai Ujung Tanjung Using the Naïve Bayes Algorithm".

2. MATERIAL AND METHOD

2.1 Naive Bayes

Naïve Bayes is a classification algorithm that is quite simple and easy to implement so this algorithm is very effective when tested with the correct dataset, especially if naïve bayes is combined with function selection, so naïve bayes can reduce redundant data, besides that naive bayes shows good results when combined with clustering methods [11].

Naïve Bayes is one of the most effective and efficient machine learning and data mining algorithms. Despite the assumption that attributes in data are independent, the performance of naïve bayes classification remains quite high even though the assumption of attribute independence rarely occurs in real data, if violated, the algorithm can still produce good classification results [13]. Naïve Bayes is an implementation of a probabilistic machine learning algorithm using the Bayes theorem calculation technique [14]. The following is the general form of Bayes' theorem on equation 1.

$$P(H|X)^{\wedge =} = \frac{p(H|X).P(H)}{P(X)} \quad (1)$$

2.2 Confusion Matrix

Confusion Matrix is an instrument used to evaluate the performance of the classification model that has been generated. In the confusion matrix, the predicted class results will be compared with the acoldl data class results, these results will then be used to calculate the accuracy, precion, recall, and f-score values [15].

In terms of data classification, sometimes there are errors in classifying it. So, it is necessary to know how much accuracy to compare with other studies [16]. One way to find out the accuracy of text classification is to use accuracy calculations. Accuracy is the degree of closeness between predicted and acoldl values.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \; x \; 100\% \quad (2)$$

TP : True Positive, meaning that the predicted value and the acoldl value are the sum of the values in the positive category.
TN : True Negative, meaning the predicted value and the acoldl value are the number of values in the negative category.
FP : False Positive, meaning the value that has a positive category prediction but the predicted value has a number of negative category values.
FN : False Negative, meaning the predicted value in the number of negative categories but has a positive category in the acoldl value.

Accuracy is defined if all conditions when predicted are correct when predicted. So if the accuracy value gets 100%, it shows that the condition predicts correctly like the original condition [17]. Precision is defined as the accuracy of a system in data classification by calculating the ratio of the number of positive true

conditions to all positive predicted conditions. Recall is the relevance of a system in data classification by calculating the number of true positive conditions against an original positive condition. F-score defines the average of similar values between precision and recall values [18].

2.3 Python

Python is a popular programming language today. It was created by Guido van Rossum, and released in 1991. Python can be used for Web development (server side), Software development or creating applications (software), Solving Math equations, System scripting and Microcontroller Programming (MicroPython) [19]. Some of the functions of Python are that it can be used on a server to create web applications, it can be used alongside software to create workflows, it can connect to database systems, it can also read and modify files, it can be used to handle big data and perform complex math, and it can be used for rapid prototyping, or for production-ready software development [20].

2.4 Data Sample

At this stage, the process of collecting user review data regarding the level of satisfaction with the use of Jet Ski Water Tourism in Ujung Balai, Tanjug Balai City is carried out in 3 stages, namely by conducting observations, interviews, and questionnaires using the Google Form feature. A lot of data is used 200 data, which is then divided back into 80% training data and 20% testing. This means that 160 of these data are training data groups, and the rest are testing data. Covering 12 questions that are considered relevant to measure the level of public satisfaction with jetski tourism in Tanjungbalai. The Table 1 is a list of questions presented to respondents in the questionnaire.

Table 1. List Of Questions on The Questionnaire

No	Questions
1	Age?
2	Gender?
3	How is the safety quality of the Water Tourism Jetski area?
4	What about the natural conditions and scenery at this Jetski Water Tourism location?
5	How are the maintenance of existing facilities such as seating, walkways, and rubbish bins, are they well maintained?
6	What about the operational opening schedule provided, does it match the needs of visitors?
7	How about information services about tours available clearly, either through signage, websites, or officers?
8	What is the level of visitor satisfaction with the staff in directing the use of the Jetski?
9	How do you feel about the quality and comfort of the Jetski?
10	How about the personal protective equipment facilities in using the jetski?
11	With the number of operational jetskis provided, do you think it is in accordance with the number of visitors who come? Give your reason
12	Are you satisfied with the free Jetski facility? Give us your reasons

These questions will be used as a research questionnaire and become the basis for conducting analysis related to the classification of the level of public satisfaction with jetski tourism in Tanjung Balai. Questions number 3 to number 11 will be initialized with labels Q1 to Q9, while question number 10 will be given the initials y. Each question has multiple answer options, namely: A. Dissatisfied, B. Neutral, C. Satisfied. The following is the initial data that has been collected from distributing questionnaires, using the questions described in Table 1 above.

After the data is collected, the process of designing a classification model using the Naïve Bayes algorithm can be started by separating the dataset into training data and testing data. Training data is used to train the model in finding patterns from the dataset, which then becomes the basis for classifying the level of public satisfaction with jetski tourism. After the model is trained, testing is done with testing data to evaluate the performance of the model, ensuring the model is able to predict the level of satisfaction with good accuracy.

In this research, the author will build a classification model using the Naïve Bayes algorithm using 5 main attributes, namely: Gender, Age, Physical and Environmental Aspects, Service Aspects, and Jetski Usage Aspects. These attributes were chosen because they are considered relevant in influencing the level of public satisfaction with jetski tourism. Through this model, the author aims to identify the relationship patterns between these attributes and the classification of satisfaction levels, so that the results of the analysis can provide more in-depth guidance for tourism managers.

The data processing process on Google colab includes several steps, starting from the dataset, Preprocessing, dividing the data into two parts, namely testing and training data. The data processing process carried out will produce a result that will be discussed and produce conclusions in the research process carried out. Data that has been successfully collected will then go through the Preprocessing stage, which is the selection stage of the examination in text mining to perform cleaning, correcting errors in the text and simplifying the text so that the text can be processed further.

After the results are obtained, the TF-IDF weighting stage is carried out again to calculate the weight of the word. TF-IDF is a calculation to measure how important words are in documents and corpus. After the TF-IDF weighting stage is complete, the next stage is the classification of the Naïve Bayes Algorithm. The data classified by the Naïve Bayes Algorithm is the Preprocessing result data which has been labeled according to the sentiment of the opinion data after the Naïve Bayes Algorithm classification results are obtained, then it is finished [21].

2.5 Flowchart Methodology

Flowchart methodology is a step taken in solving the problem discussed, used to assist researchers in compiling a clear framework of stages. There are several stages that must be passed to conduct this research, including the following figure 1.

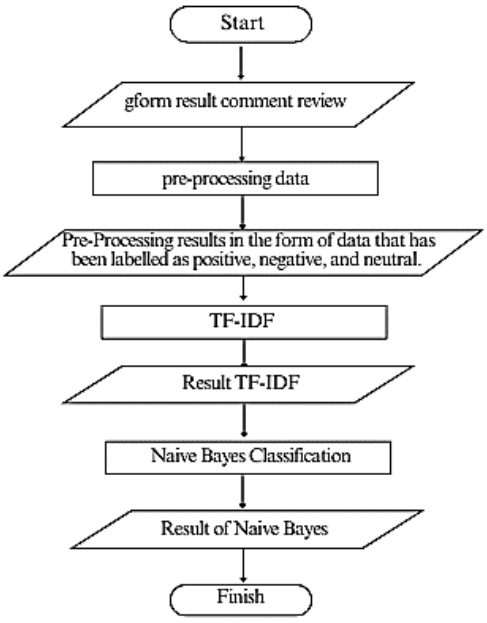

Figure 1. Flowchart Methodology

The first stage is inputting data where the data that has been obtained from the survey results using the Google Form feature. After inputting the data, the next step is Preprocessing data is a step to get relevant training data so that it is ready to use. The software used in preprocessing data is Google Colab by using the phyton library as a storage or library, after preprocessing the data, the results are obtained. After the results are obtained, the TF-IDF weighting stage is carried out again to calculate the weight of the word. TF-IDF is a calculation to measure how important words are in documents and corpus. After the TF-IDF weighting stage is complete, the next stage is the classification of the Naïve Bayes Algorithm. The data classified by the Naïve Bayes Algorithm is the Preprocessing result data which has been labelled according to the sentiment of the opinion data after the Naïve Bayes Algorithm classification results are obtained, then it is finished [22].

3. RESULTS AND ANALYSIS

This section will explain in detail about the application of the Naïve Bayes algorithm in classifying the level of public satisfaction with the use of jetskiing as a water tourism vehicle in Ujung Tanjung Hall, Tanjung Balai City. The main purpose of applying this algorithm is to understand and predict the factors that affect the level of satisfaction based on the various variables available. Through this analysis, a clearer picture of the public's perception of this water tourism facility will be obtained, which is expected to help the management in improving service quality and visitor experience.

In this research, the author will build a classification model using the Naïve Bayes algorithm using 5 main attributes, namely: Gender, Age, Physical and Environmental Aspects, Service Aspects, and Jetski Usage Aspects. These attributes were chosen because they are considered relevant in influencing the level of public satisfaction with jetski tourism. Through this model, the author aims to identify the relationship patterns between these attributes and the classification of satisfaction levels, so that the results of the analysis can provide more in-depth guidance for tourism managers.

After the target class for each data is determined, the initial dataset has been transformed using the attributes and target class that have been determined previously. This transformation process makes the dataset shorter and easier to read, and simplifies calculations when using the Naïve Bayes algorithm model [23]. The following table shows how the data has been grouped into categories according to the attributes and target classes, can view table 2.

Table 2. Data Transformation

No	Attribute A	Attribute B	Attribute C	Attribute D	Attribute E	y
1	1	2	2	2	2	1
2	1	2	2	2	2	1
3	1	2	2	2	2	1
4	1	2	2	2	2	0
5	1	2	2	2	2	1
6	1	2	2	1	2	0
…	…	…	…	…	…	..
200	1	1	2	2	2	1

3.1. Application of Naive Bayes Classification

3.1.1. Calculating The Prior Probability

From the training data consisting of 160 data, the prior probability or probability of occurrence of the target variable for each class can be calculated. In this study, no data falls into the "Unsatisfied" class category, while 25 people fall into the "Neutral" class, and 135 people fall into the "Satisfied" class. Therefore, the prior probability calculation will be done based on these distributions. The Prior Probability can view table 3.

Table 3. The Prior Probability

Variable	Class	Amount	Probability
y	-1	0	0,0000
	0	25	0,1666
	1	135	0,8333

Based on the table above, out of a total of 160 data in the training data, the highest probability is found in the "Satisfied" class with a value of 0.8333333. Below it, the "Neutral" class has a probability of 0.1666667. Meanwhile, the "Dissatisfied" class has the lowest probability, which is 0. After calculating the prior probability value, the next step is to calculate the probability of each predetermined attribute. In this research, there are 5 attributes, namely attribute A to attribute E.

3.1.2. Calculating classification using the naive bayes model

After the probability for each attribute is calculated, the next step is to apply the Naive Bayes model to the testing data. This testing data is 20% of the total dataset, which amounts to 40 data. This model will be used to predict the target class based on the previously calculated probabilities.

Table 4. Data Testing

No	Gender	Age Category	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
1	F	Adult	Normal	Normal	Good	Good	Good	Good	Good	Good	Good
2	M	Adult	Good	Good	Normal	Good	Good	Good	Good	Normal	Good
3	M	Adult	Good	Good	Normal	Good	Normal	Good	Good	Good	Good
4	M	Adult	Good	Good	Good	Good	Good	Good	Good	Good	Good
5	M	Adult	Good	Good	Good	Good	Good	Good	Normal	Normal	Normal
6	F	Adult	Good	Normal	Normal	Good	Good	Good	Good	Good	Normal
…	…	…	…	…	…	…	…	…	…	…	…
40	M	Teenager	Good	Good	Good	Good	Good	Normal	Normal	Good	Normal

The testing data will be transformed into numerical form to facilitate processing using the Naive Bayes model. This transformation is done so that categorical attributes can be processed by the algorithm efficiently. More details about this data transformation can be seen in the following table 5.

Table 5. Transformation of Data Testing

No	Gender	Age Category	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
1	1	1	2	2	3	3	3	3	3	3	3
2	0	1	3	3	2	3	3	3	3	2	3
3	0	1	3	3	2	3	2	3	3	3	3
4	0	1	3	3	3	3	3	3	3	3	3

No	Gender	Age Category	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
5	0	1	3	3	3	3	3	3	2	2	2
6	1	1	3	2	2	3	3	3	3	3	2
…	…	…	…	…	…	…	…	…	…	…	…
40	0	0	3	3	3	3	3	2	2	3	2

The gender variable will be converted into attribute A, while the Age Category variable will be converted into attribute B. Q1 to Q3 variables will be merged and converted into attribute C, Q4 to Q6 variables into attribute D, and Q7 to Q9 variables will be converted into attribute E. After the transformation of the testing data is complete, the classification process can be carried out using the probability value of each attribute that has been calculated previously.

3.2. Resullts

This section presents the results of the implementation of the Naive Bayes model program that has been designed. The main focus of this discussion is the analysis of the results and testing of the model, as well as displaying the level of accuracy of the application of the model that has been developed. At this stage, 200 questionnaire data, with 5 attributes namely Initials, Gender, Age, Physical and Environmental Aspects, Service Aspects, and Jetski Usage Aspects will be processed. For classification purposes, the values in the 'y' column, which may represent certain categories of respondents, will be converted into numeric numbers using LabelEncoder.

After all attributes have been set to build a classification model using the Naïve Bayes algorithm, the next step is to use y = the level of public satisfaction on the collected dataset, which consists of 200 data. In this study, the author has defined three categories for y, namely: Dissatisfied, Neutral, and Satisfied. More details about this target can be seen in the table 6.

Table 6. Output Encoding Label

	Atribute A	Atribute B	Atribute C	Atribute D	Atribute E	y
0	0	1	2	2	2	1
1	1	1	2	2	2	1
2	1	0	2	2	2	1
3	1	1	2	2	2	0
4	1	1	2	2	2	1

The division of training data and testing data is an important step in building machine learning models. Training data is used to train the model to recognize patterns, while testing data is used to evaluate the model's performance on new data to measure its ability to make accurate predictions. The data is separated into two parts: features (independent variables) and target variables (dependent variables). Features are the columns used as input for the model, while the target variable is the result to be predicted by the model.

The trained model is used to predict the class of the testing data. The model classifies the new data into predefined categories based on patterns learned from the training data. The result is a label that indicates the predicted class for each sample of testing data. The outcomes will be acquired following the conclusion of the Naïve Bayes algorithm testing procedure. where the results under consideration are labels from test data that the model received during training. The accuracy, precision, recall, and f1-score values of the model employed on thedataset will be revealed once the test data classification results sentiment classes are separated from the real class data, on the table 7.

Table 7. Confussion Matrix

	Precission	Recall	F1-score	Support
0	50.00%	20.00%	28.75%	10.00 %
1	77.78%	93.33%	84.85%	30.00 %
Accuracy	75.00%	75.00%	75.00%	75.00%
Macro Avg	63.89%	56.67 %	56.71%	40.00 %
Weighted Avg	70.83%	75.00%	70.78%	40.00%

4. CONCLUSION

The Naïve Bayes model shows that the level of public satisfaction with Jetski Water Tourism in Ujung Tanjung Hall, Tanjungbalai City can be predicted with an accuracy of 75%. This indicates that the model is quite effective in identifying the level of user satisfaction, although there is a 25% possibility of inaccuracy in prediction. With this accuracy, the model can provide useful insights for the evaluation and improvement of jetski tourism services, but it should be considered to conduct further analysis to improve accuracy and get a more comprehensive picture of community satisfaction. The model shows varying performance between the

two categories. For the “Satisfied” category, the model achieved a precision of 77.78% and a recall of 93.33%, indicating a strong ability to identify and predict “Satisfied” instances with high accuracy. This means that the model is quite Reliable in capturing all instances that are truly “Satisfied” and also has a relatively low error rate in this prediction category. However, for the “Neutral” category, the model shows lower precision, namely 50%, and very low recall, namely 20%. Low precision indicates that only a portion of the “Neutral” predictions are actually correct, while low recall indicates that the model fails to recognize most of the examples that are actually “Neutral”. This indicates that the model is less effective in identifying the "Neutral" category, which may be due to data integration or features that are less representative for this category.

To increase the effectiveness of the model in identifying people's satisfaction levels, it is recommended to expand the amount of data used. By adding more data, the model will have a wider variety of answers and can improve its accuracy and predictive ability.

REFERENCES

[1] F. Fitroh and F. Hudaya, “Systematic Literature Review: Analisis Sentimen Berbasis Deep Learning,” *TEKNOSI*, vol. 9, no. 2, pp. 132–140, Aug. 2023, doi: 10.25077/TEKNOSI.v9i2.2023.132-140.

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

[3] G. C. Triasis, D. Arisandi, and T. Sutrisno, “ANALISIS KEPUASAN PENGGUNAAN APLIKASI SHOPEE MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *jiksi*, vol. 10, no. 1, Mar. 2022, doi: 10.24912/jiksi.v10i1.17857.

[4] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, T. Tukino, and B. Priyatna, “Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter,” *JOINTECS*, vol. 8, no. 1, p. 11, Mar. 2023, doi: 10.31328/jointecs.v8i1.4265.

[5] A. I. Putri and M. Furqan, “Application of Data Mining to Predict Birth Rates in Medan City Using the K-Nearest Neighbor Method,” *Journal of Computer Science*, vol. 5, no. 1.

[6] M. Furqan, S. Sriani, and S. M. Sari, “Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia,” *tc*, vol. 21, no. 1, pp. 51–60, Feb. 2022, doi: 10.33633/tc.v21i1.5446.

[7] S. Harlina, S. Suryani, and M. Oton Kadang, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi kelayakan Calon Nasabah Kredit Berbasis Web,” *SINTaKS*, vol. 1, no. 1, Aug. 2022, doi: 10.35842/sintaks.v1i1.18.

[8] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, “Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile: Analisis Sentimen,” *petir*, vol. 15, no. 2, pp. 264–275, Nov. 2022, doi: 10.33322/petir.v15i2.1733.

[9] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, “Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung,” *Informatik : Jurnal Ilmu Komputer*, vol. 18, no. 3, p. 239, Dec. 2022, doi: 10.52958/iftk.v18i3.4694.

[10] S. Hilda Kusumahadi, H. Junaedi, and J. Santoso, “Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine,” *jpit*, vol. 4, no. 1, pp. 54–60, Jan. 2019, doi: 10.30591/jpit.v4i1.1125.

[11] C. D. A. A. P. Chrishariyani, Y. Rahman, and Q. Aini, “Kepuasan Pengguna Layanan Shopee Food Menggunakan Algoritma Naive Bayes,” *J. Sistem Info. Bisnis*, vol. 12, no. 2, pp. 98–105, Dec. 2022, doi: 10.21456/vol12iss2pp98-105.

[12] G. Gustientiedina, M. Siddik, and Y. Deselinta, “Penerapan Naïve Bayes untuk Memprediksi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademis,” *J.InfoMedia*, vol. 4, no. 2, p. 89, Jan. 2020, doi: 10.30811/jim.v4i2.1892.

[13] Sonia Wanda Mafriza and Armansyah, “Klasifikasi karir mahasiswa bidang web developer menggunakan algoritma naïve bayes,” *infotech*, vol. 4, no. 2, pp. 270–280, Dec. 2023, doi: 10.37373/infotech.v4i2.907.

[14] A. Safira and F. N. Hasan, “ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER,” *Jurnal Sistem Informasi*, vol. 5, no. 1, 2023.

[15] S. Wirma, “Data Mining Dengan Metode Naïves Bayes Classifer dalam Memprediksi Tingkat Kepuasan Pelayanan Dokumen Kependudukan,” *INFEB*, pp. 156–160, Sep. 2022, doi: 10.37034/infeb.v4i3.155.

[16] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Y. Yusra, “Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER,” *JNKTI*, vol. 5, no. 3, pp. 423–431, Jun. 2022, doi: 10.32672/jnkti.v5i3.4382.

[17] Suparyanto, “Klasifikasi Kepuasan Layanan Akademik Di STMIK El Rahma Menggunakan Metode Algoritma Naive Bayes,” *FAHMA*, vol. 20, no. 2, pp. 100–111, May 2022, doi: 10.61805/fahma.v20i2.37.

[18] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. Santoso Dwi, “Penerapan Metode Support Vector Machine untuk Analisis Sentimen Pengguna Twitter,” *Journal of Information System Management*, vol. 3, 2021.

[19] 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.

[20] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, “Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing,” *eksplora*, vol. 10, no. 1, pp. 32–42, Sep. 2020, doi: 10.30864/eksplora.v10i1.390.

[21] E. R. Lidinillah, T. Rohana, and A. R. Juwita, "Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine," *tekno*, vol. 10, no. 2, pp. 154–164, Jul. 2023, doi: 10.37373/tekno.v10i2.440.
[22] V. P. Virza, G. Tri Pranot, and F. Eko Putra, "Klasifikasi Kebutuhan Sparepart Dengan Algoritma K-Nearest Neighbor Untuk Meningkatkan Penjualan Sparepart," *bit*, vol. 4, no. 3, pp. 287–293, Sep. 2023, doi: 10.47065/bit.v4i3.729.
[23] M. Siddik, H. Hendri, R. N. Putri, Y. Desnelita, and G. Gustientiedina, "Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naïve Bayes," *INTECOMS*, vol. 3, no. 2, pp. 162–166, Nov. 2020, doi: 10.31539/intecoms.v3i2.1654.

BIBLIOGRAPHY OF AUTHORS

Nursalimah Isnaini Fatwa 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.

Rakhmat Kurniawan is a lecturer at the Faculty of Science and Technology at UINSU. Who is also the supervisor of Siti Nurul Aini in completing this final project research. Rakhmat Kurniawan has more than 20 articles related to computer science.