
Decision Support System for Sentiment Analysis of Youtube Comments on Government Policies

I Putu Agus Eka Darma Udayana^{1)*}, I Gusti Agung Indrawan²⁾, I Putu Dwi Guna Ambara Putra³⁾

^{1*)2)}Institut Bisnis dan Teknologi Indonesia (INSTIKI), Indonesia

^{1)*}agus.ekadarma@gmail.com, ²⁾agung.indrawan@instiki.ac.id, ³⁾dwiguna81@gmail.com

ABSTRACT

Sentiment analysis is the process of classifying a text dataset as positive, negative, or neutral. YouTube is one of the popular media used to provide responses to a problem. In the Jokowi era, infrastructure development was carried out massively and evenly, one of which was in Bali Province, namely the construction of the Mengwi-Gilimanuk Toll Road. The construction of the Mengwi-Gilimanuk Toll Road consumed a lot of people's agricultural land, which resulted in various pro and con responses from the community. From these problems, sentiment analysis is carried out to get community reviews related to the object being analyzed by utilizing algorithms to be able to classify opinions, in the construction of this system the naïve bayes algorithm is used with testing methods namely accuracy, precision, and recall. Naïve Bayes method used in this research because, in previous studies that compared KNN and Naive Bayes, the Naïve Bayes method has advantages in analyzing sentiment analysis cases. From the sentiment analysis conducted by utilizing 18 video links on YouTube with 701 comments, it produces positive sentiment as much as 50.64%, negative sentiment as much as 7.70% and neutral sentiment as much as 39.23%.

Keywords: Decision Support System; Sentiment Analysis; Government Policy; Text-Mining; Naïve Bayes

1. INTRODUCTION

Indonesia implements a democratic system of governance in which decision-making always includes the people. In the implementation of democracy, every individual is given equal rights in expressing their opinions. In Indonesia, freedom of speech is regulated in Article 28 of the 1945 Constitution and Article 28E paragraph 3 of the 1945 Constitution. Expressing opinions can be done directly during deliberations or indirectly through social media. Social media is an online media that is useful to facilitate users in interacting. As for social media that is popularly used to provide responses, one of them is YouTube. The reason why people switch from television is because YouTube can be accessed anywhere and anytime. We can have an opinion on YouTube by commenting according to the context of the video being watched, in commenting there are several rules that must be followed, namely not to spam, practice fraud and scam (Bhuiyan et al., 2017; Cunha et al., 2019). In 2 terms of office, President Jokowi has several targets to be achieved including human resource development, investment, bureaucratic reform, economic transformation and infrastructure development. In the Jokowi era, infrastructure development was carried out massively and evenly, one of which was in Bali Province. To realize the province of Bali to be more prosperous, the Bali provincial government has a vision of Nangun Sat Kerthi Loka Bali (Udayana et al., 2017; Wiana, 2018). In realizing this vision, the Bali Provincial government will take 22 Bali development missions. One of these missions is the development of land, sea and air infrastructure. One of the land infrastructure developments being planned by the Bali Provincial government is the Mengwi-Gilimanuk Toll Road. Work on the Mengwi-Gilimanuk Toll Road project will begin in December 2021.

* Corresponding author



The Mengwi-Gilimanuk Toll Road has a length of 96.21 KM by passing through 3 districts including Badung, Tabanan and Jembrana (Krisnananda & Kartika, 2021). With the length of the toll road that has been planned, it is estimated that it will sacrifice a lot of agricultural land, from the results of digitization and going to the field, it was found that the area of agricultural land that will be violated is 480.54 Ha with a total of 97 subak being violated. Seeing the construction of the Mengwi-Gilimanuk Toll Road which uses a lot of agricultural land, it raises various pro and contra responses from the public in various media, including on YouTube. There are more than 50 videos that discuss the construction of the toll road, the author took 18 videos from 10 YouTube channels to study. Of the 18 videos used, there were 617 comments and an average of 34 comments on each video. With the problem of pro and con comments related to local government policies, comments on YouTube social media can be used as a reference to assess how people respond to policies issued by the government and from the many pro and con responses to the construction of the Mengwi-Gilimanuk Toll Road, the research goal is to build a decision support system that will be useful for the government in making policies. system in the form of sentiment analysis of YouTube comments on government policies, especially related to the Mengwi-Gilimanuk Toll Road Development by applying the Naive Bayes Algorithm in classifying comment text. This method is due to research conducted by Syarifuddin, which analyzed public opinion sentiment regarding Covid-19. In his research comparing the methods between Naive Bayes and KNN, the Naive Bayes method obtained an accuracy test result of 63.21% while KNN obtained an accuracy test result of 58.10%, so in doing sentiment analysis, it is better to use Naive Bayes (Syarifuddin, 2020). The features of the decision support system made contain crawling YouTube comment data, retrieving comment datasets, classifying comments, weighting words on comments, implementing algorithms and testing for measurement of accuracy, precision and recall (Gao et al., 2019; Laksono et al., 2019). The research results showed that from 18 video links with 701 total comments, the results of positive sentiments collected from YouTube comments on government policies related to the construction of the Mengwi-Gilimanuk Toll Road got the largest percentage of 50.64% compared to negative 7.70% and neutral 39.23%, so the dominant community agrees with the plan to build the Mengwi-Gilimanuk Toll Road.

2. LITERATURE REVIEW

The research that the author developed was actually inspired by previous studies and other sources. The author found several studies that encouraged him to raise the theme of a decision support system for sentiment analysis of youtube comments on government policies (case study of the Mengwi-Gilimanuk toll road construction). The research discusses topics related to the author's research and the algorithms used by the author. Research that has been done related to decision support systems, among others. Research on the Implementation of Sentiment Analysis of Government Service Video Channel Comments on Youtube Using the Naive Bayes Algorithm Method conducted (Saputra et al., 2019) using 3 criteria, namely positive, negative and neutral. This research produces an application that can provide information on how to measure the level of public satisfaction with government services through YouTube. Further research was conducted by (Salam et al., 2018) with the title Sentiment Analysis of Facebook Social Media Comment Data with K-Nearest Neighbor (Case Study on J&T Ekspres Indonesia Goods Expedition Service Account). This research was conducted so that J & T could find out how the public responded to the company's performance. This test results in the classification of sentiment analysis of comment documents on freight forwarding services (J&T) on Facebook using the K-Nearest Neighbor algorithm which has the highest accuracy of 79.21% with 101 training documents while the lowest accuracy is 70.3% with 101 training documents. Other previous research conducted by (Dar et al., 2021), in his research entitled Implementation of the Naive Bayes Classifier (NBC) Algorithm for Sentiment Analysis of Full Day School Policy Comments. In his research, he explained that there are pros and cons of the full day school policy issued by the Indonesian Minister of Education and Culture Muhadjir Effendy. So the author wants to create a system to analyze the problem so that the government can follow up on the policies that have been issued (Wongkar & Angdresey, 2019). Other related research that compares the Naive Bayes method with K-Nearest Neighbor was conducted by (Syarifuddin, 2020), in his research entitled Sentiment Analysis of Public Opinion Regarding Covid-19 on Twitter Using the Naive Bayes and KNN Methods. In his research, he explained that in conveying the aspirations of the community, there are obstacles to the limitations of time and space so that people use social media as a channel for aspirations to be known by the government. The test results using the Naive Bayes and KNN methods show that public opinion tends to be positive (I. P. A. E. D. Udayana et al., 2022). Of the two methods, the Naive Bayes method is the most accurate classification model with an accuracy value of 63.21% while the KNN method has an accuracy value of 58.10%.

* Corresponding author



The research conducted by (Gunawan et al., 2018) with the title Sentiment Analysis System on Product Reviews Using the Naive Bayes Method which discusses the number of reviews and opinions so much that it will be difficult and time consuming to read as a whole. The test results obtained using the Naive Bayes method produced the lowest accuracy value in testing 5 classes using a dataset of 80% training and 20% test data of 52.66%, while in testing 3 classes using a dataset of 90% training data and 10% test data had the highest accuracy of 77.78%.

3. METHOD

3.1. Data Collection Method

This research was made by collecting data through scraping, literature study and observation of the object of research. Observation, the observation carried out is to observe the public sentiment expressed in YouTube social media comments. Scrapping, at this stage data collection is carried out by taking comments on YouTube on videos that show the construction of the Mengwi-Gilimanuk Toll Road. The process of taking comments is done using the YouTube comments scraper feature provided in the python programming language. Literature study, literature is taken from the internet and books that can support research. From the literature study, the author found several studies that encourage to raise the theme of sentiment decision support system analysis of YouTube comments on government policies.

3.2. Naïve Bayes Method Flow

Comment sentiment analysis has several methods that can be used. This research will use the naïve bayes algorithm method in performing sentiment analysis on comments. The following is the flow of the naïve bayes algorithm:

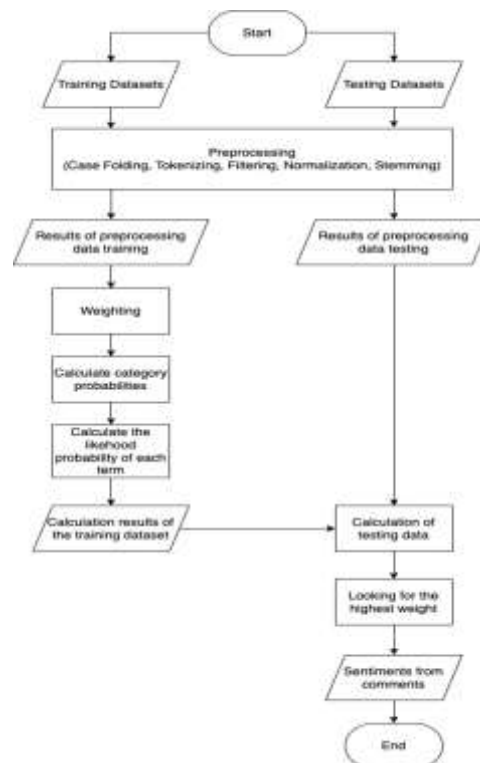


Fig. 1 Naïve Bayes Algorithm Flow

Based on Figure 1, the stages of the Naïve Bayes method in the research can be explained, namely:

a) Data Analysis

From the observations that have been made, data is obtained in the form of various YouTube comments with a

* Corresponding author



total of approximately 617 comments. Then using the youtube comment scraping method, the author uses the python programming language to create a data scraping application. From scraping done on 18 videos, 701 comments are obtained that can be analyzed.

b) Data Preprocessing

After obtaining data from scraping results, it will proceed to the preprocessing process. Preprocessing itself serves to improve the quality of data that will be processed in the naïve bayes method. The stages of preprocessing are as follows:

- 1) Case Folding is a process to convert all comment sentences into lowercase letters.
- 2) Tokenizing is the process of breaking sentences into groups of words to make it easier to weight the data.
- 3) Filtering serves to clean up characters or symbols that are not needed, such as emojis and other characters.
- 4) Normalization is a stage to correct text that is spelled incorrectly.
- 5) Stemming is a step to return words that contain suffixes or prefixes back to the base word.

c) Weighting

Weighting is the stage of giving value to data that has passed the previous stages. At the weighting stage the author uses the term frequency / number of occurrences of words to give value, this step is to give weight to each word according to their respective sentiments.

3.3. Naïve Bayes Algorithm

Naïve Bayes is a popular machine learning technique for text classification, as it is very simple, efficient and performs well in many domains. Here is the formula of the naïve bayes calculation(Alam et al., 2022):

$$P(X|H) = \frac{p(X|H) \times p(H)}{p(X)} \quad (1)$$

Information:

P(H|X) = Probability of hypothesis H based on condition X (a posteriori probability)

P(H) = Probability of hypothesis H (prior probability)

P(X|H) = Probability of X based on the condition in hypothesis H

P(X) = Probability of X

Testing the results of the Naive Bayes method calculation is done by testing accuracy, precision and recall. Accuracy calculation is done by dividing the amount of correct sentiment data by the total data and test data. The accuracy calculation formula is as follows(Tang et al., 2020):

$$\frac{TP+TNG+TNT}{TP+TNG+TNT +FP+FNG+FNT} \times 100\% \quad (2)$$

Information :

TP : True Positive

TNG : True Negative

TNT : True Neutral

FP : False Positive

FNG : False Negative

FNT : False Neutral

The calculation of precision is done by dividing the number of correct data with a positive value divided by the number of correct data with a positive value and incorrect data with a positive value. The precision calculation formula is as follows:

$$\text{Precision} = \frac{\text{true positive}}{\text{true positive}+\text{false positive}} \quad (3)$$

* Corresponding author



The recall calculation is done by dividing the positive correct data by the sum of the positive correct data and the negative wrong data. The recall calculation formula is as follows:

$$\text{Recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (4)$$

4.

RESULT

4.1. System Interface Implementation

Implementation of the system interface is the result of the system that has been built. The following is a display of the Decision Support System for Sentiment Analysis of Youtube Comments on Government Policy (Case Study of Mengwi-Gilimanuk Toll Road Development).

a) Login Page

The first page that will appear in the system when run is the user login page. On this page the user is required to log in by entering a username and password before entering the system to protect the system from being misused by irresponsible people.



Fig. 6 Login Page

b) Register Page

Prospective system users are required to register or register to become system users. So that they can log in and enter the system.



Fig. 7 Register Page

c) Main Page

* Corresponding author



The main page of the system is the first display when the user successfully enters the system. On this page users can manage the links that will be scraped.



Fig. 8 Main Page

d) Dataset Page

The dataset page serves to display comment data obtained from scraping video links inputted on the main menu page, the comment data will be processed in the preprocessing process, the comment data consists of training datasets and test datasets.



Fig. 9 Dataset page

e) Preprocessing page

The preprocessing page serves to display the results of comment processing at the preprocessing stage, from the results of the comment processing, the value weighting is then carried out on the word preprocessing results.



Fig. 10 Preprocessing Page

f) Calculation Results Page

* Corresponding author



The calculation results page is a page that displays the results of the calculation of the naïve bayes algorithm which contains the results of negative, positive, and neutral sentiments from previously processed comment data.



Fig. 11 Calculation Result Page

g) PDF Result

The results of the analysis that have been processed by the system can be presented in PDF form which contains information about the links analyzed, the percentage of positive, negative or neutral and displays a graph of the percentage of sentiment.



Fig.12 PDF Results Page

h) CSV File Results

In addition to creating a PDF file, the system also creates a CSV file from the results of the analysis that has been done, the CSV file presents sentiment and comment information that has been processed in the naïve bayes algorithm.

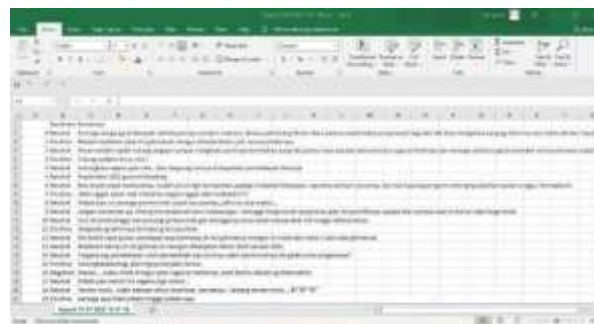


Fig. 13 CSV File Result

4.2. System Testing Results

* Corresponding author



System testing serves to find out whether the system built can function properly and according to plan. In this sentiment analysis decision support system in determining the sample, the slovin formula is used and for system testing the accuracy, precision and recall methods are used. The following are the results of system testing that has been done.

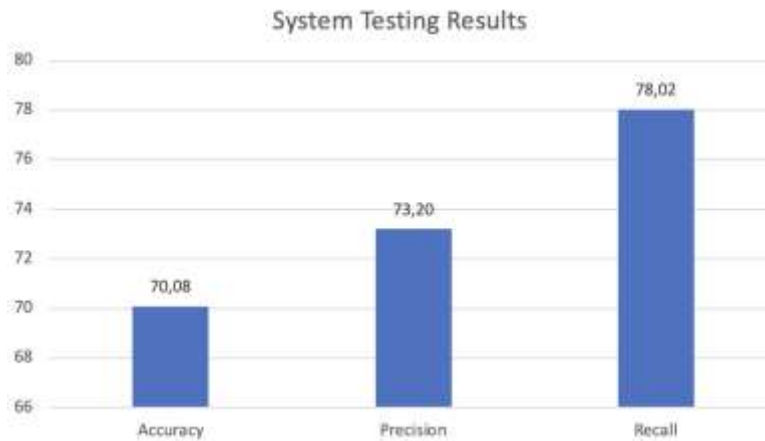


Fig. 14 System Testing Results

h) Expert Team Analysis Results

The author asks for help from a team of experts, namely Indonesian language teachers on behalf of Umar Effendi, S. Pd as expert team 1 and I Nengah Artama, S. Pd as expert team 2, the expert team has a role to check the analysis that has been done by the system. The analysis was carried out on 2 videos with 82 comments.



Fig. 15 Expert Team Analysis Results

There are differences in the analysis results between the expert team and the system, this is due to the dataset used to perform sentiment analysis on YouTube comments is not perfect, because some comments written by the community use local languages. The dataset used to perform sentiment analysis on this system only focuses on Indonesian and in some comments there is sarcasm while the system has not been able to classify comments using majas or not. Because there are several obstacles, it results in differences in analysis results between the expert team and the system.

DISCUSSIONS

This research aims to apply the Naïve Bayes algorithm to analyze public sentiment expressed on the youtube platform related to the construction of the Mengwi-Gilimanuk Toll Road. there are several processes based on the

* Corresponding author



research stages, namely:

1. Data Preprocessing Results

The calculation process using the naïve bayes algorithm begins with the data preprocessing stage which has several sub-processes carried out including Case Folding, Tokenizing, Filtering, Normalization and Stemming. The following is an example of data preprocessing results.

Table 1
 Data Preprocessing Results

| Process Stages | Result |
|----------------|--|
| Sentimen Asli | Moga Incar sesuai rencana, Terimakasih |
| Case Folding | moga Incar sesuai rencana, terimakasih |
| Tokenizing | moga Incar sesuai rencana , terimakasih |
| Filtering | moga Incar sesuai rencana terimakasih |
| Normalization | semoga lancar sesuai rencana terima kasih |
| Stemming | semoga lancar sesuai rencana terima kasih |

2. Context Diagram

Context diagram is a general description of a system to be built. In the context diagram of the Youtube Comment Sentiment Analysis Decision Support System for Government Policy consists of one entity, namely the system user.

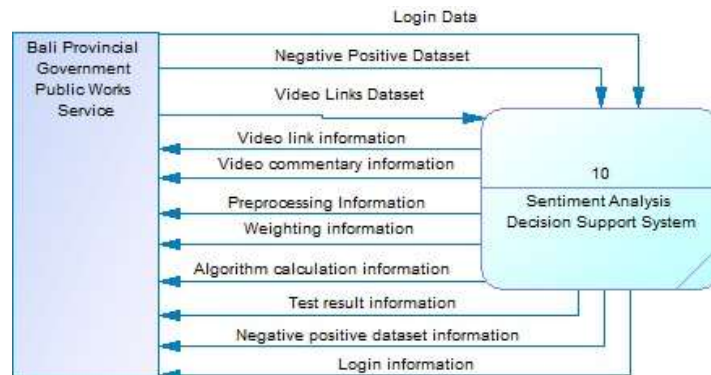


Fig. 2 Context Diagram

3. Physical Data Model

Physical Data Model of the Decision Support System for Sentiment Analysis of Youtube Comments on Government Policies (Case Study of Mengwi-Gilimanuk Toll Road Development) from the conceptual data model where the results of the relationship of all tables used.

* Corresponding author



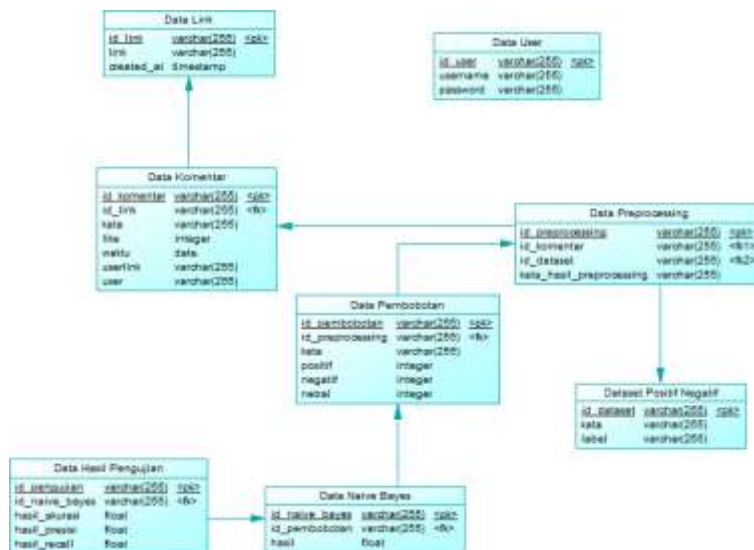


Fig. 5 Physical Data Model

5. CONCLUSION

Based on the descriptions previously described, the conclusions that the author draws from the design and development of the Decision Support System for Sentiment Analysis of Youtube Comments on Government Policies (Case Study of Mengwi-Gilimanuk Toll Road Construction) include the research methodology applied in this study is able to analyze the sentiment of youtube comments on government policies (case study of the construction of the Mengwi-Gilimanuk Toll Road). From 18 video links with 701 total comments, the results of positive sentiment collected from YouTube comments on government policies related to the construction of the Mengwi-Gilimanuk Toll Road got the largest percentage of 50.64% compared to negative 7.70% and neutral 39.23%, so the dominant community agrees with the plan to build the Mengwi-Gilimanuk Toll Road. Then based on system testing carried out using the accuracy, precision and recall methods, the percentage of each test is obtained, namely accuracy 70.08%, precision 73.20% and recall 78.02%. This proves that sentiment analysis using the naïve bayes method is good for sentiment analysis.

6. REFERENCES

Alam, S., Resmi, M. G., & Masriyah, N. (2022). Classification of Covid-19 vaccine data screening with Naive Bayes algorithm using Knowledge Discovery in database method. *Journal of Computer Networks, Architecture and High Performance Computing*, 4(2), 177–185.

Bhuiyan, H., Ara, J., Bardhan, R., & Islam, M. R. (2017). Retrieving YouTube video by sentiment analysis on user comment. *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 474–478.

Cunha, A. A. L., Costa, M. C., & Pacheco, M. A. C. (2019). Sentiment analysis of youtube video comments using deep neural networks. *International Conference on Artificial Intelligence and Soft Computing*, 561–570.

Dar, M. H., Sihombing, V., & Utami, Y. A. D. (2021). *Implementasi Algoritma Naive Bayes Classifier (NBC) untuk Analisis Sentimen Komentar Kebijakan Full Day School*. 6(1), 61–

* Corresponding author



66.

- Gao, H., Zeng, X., & Yao, C. (2019). Application of improved distributed naive Bayesian algorithms in text classification. *The Journal of Supercomputing*, 75(9), 5831–5847.
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2), 113. <https://doi.org/10.26418/jp.v4i2.27526>
- Krisnananda, J., & Kartika, A. A. G. (2021). Analisis Kelayakan Ekonomi dan Finansial Pembangunan Jalan Tol Mengwi-Gilimanuk. *Jurnal Teknik ITS*, 10(2), E253–E258.
- Laksono, R. A., Sungkono, K. R., Sarno, R., & Wahyuni, C. S. (2019). Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes. *Proceedings of 2019 International Conference on Information and Communication Technology and Systems, ICTS 2019, July*, 49–54. <https://doi.org/10.1109/ICTS.2019.8850982>
- Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia). *Prosiding SINTAK*, 480–486.
- Saputra, P. Y., Subhi, D. H., & Winatama, F. Z. A. (2019). Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naive Bayes. *Jurnal Informatika Polinema*, 5(4), 209–213. <https://doi.org/10.33795/jip.v5i4.259>
- Syarifuddin, M. (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn. *Inti Nusa Mandiri*, 15(1), 23–28.
- Syarifuddin, M. (2020). ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19 PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN KNN. *INTI Nusa Mandiri*, 15(1), 23–28. <https://doi.org/10.33480/inti.v15i1.1347>
- Tang, X., Li, J., Liu, M., Liu, W., & Hong, H. (2020). Flood susceptibility assessment based on a novel random Naive Bayes method: A comparison between different factor discretization methods. *Catena*, 190, 104536.
- Udayana, I. P. A. E. D., Kherismawati, N. P. E., & Sudipa, I. G. I. (2022). Detection of Student Drowsiness Using Ensemble Regression Trees in Online Learning During a COVID-19 Pandemic. *Telematika: Jurnal Informatika Dan Teknologi Informasi*, 19(2), 229–244.
- Udayana, I., Sudarma, M., & Kumara, I. N. S. (2017). Balinese Latin Text Becomes Aksara Bali Using Rule Base Method. *Int. J. Res. IT, Manag. Eng*, 7(05), 1–7.
- Wiana, K. (2018). Sad Kertih: Sastra Agama, Filosofi, dan Aktualisasinya. *Bali Membangun Bali, Jurnal Bappeda Litbang*, 1, 159–179.
- Wongkar, M., & Angdresey, A. (2019). Sentiment analysis using Naive Bayes Algorithm of the data crawler: Twitter. *2019 Fourth International Conference on Informatics and Computing (ICIC)*, 1–5.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).