

Research Article

Classification of Sentiment Analysis and Community Opinion Modeling Topics for Application of ICT in Government Operations

Andi Akram Nur Risal, Fathahillah, Dwi Rezky Anandari Sulaiman

Department of Informatics and Computer Engineering, Universitas Negeri Makassar, Makassar 90224, Indonesia

Received: February 18, 2023; Accepted: April 25, 2023; Published: April 29, 2023

Abstract: Utilizing information systems is very useful in the current era. Digitizing administration in the Village is beneficial in the service process to the public. This is seen as a change in service that can make it easier or more difficult for the people of Sanrobone Village to take care of administration at the village office. This study aims to analyze public opinion regarding the use of e-government, predict public opinion regarding the use of e-government, and analyze modeling topics related to the use of e-government. This research applies a text mining algorithm with a sentiment analysis method to see positive, negative, and neutral public perceptions and also uses topic modeling to get the most frequently appearing topics in the data. Stages in this study include Data Collection, Text Pre-processing, Sentiment Analysis, Topic Modelling, Classification, and Evaluation. The results obtained are the ten words that appear most often in the responses of the Village community: easy 122, help 96, village 80, accessed 80, letter 80, permit 77, resident 73, manage 60, service 52, and the person with 52 words. The sentiment analysis is positive, with 411 opinions, 37 negative opinions, and 152 neutral opinions. Finally, the performance of the Naive Bayes algorithm in predicting classification results is excellent, with an accuracy rate of 98 percent.

Keywords: Naïve Bayes; Opinion; Prediction; Sentiment Analysis; Topic Modeling.

1. Introduction

The development of Information and Communication Technology in Governance (e-government) based on advances in information and communication technology, especially the Internet, has provided new opportunities for the government to provide services and interact with the public efficiently and transparently. With the adoption of this technology, the government can overcome physical and time limitations in delivering public services. The demands of modern society that expect easy access and efficiency in interacting with the government are also reasons the government demands the development of e-government. The application of e-government can

increase transparency and accountability in government services. With an electronic government platform, the government can provide easier access to public information and streamline service time that was previously done manually and required a long time.

The management of e-government-based public services is proliferating, covering provinces, cities, and districts down to the village level. The management of public services at the village level measures the success of government services. The village head leads the government to build and prosper village communities through physical and digital community activities and public services. The ability of a village to manage and improve public services reflects that village communities

have advanced, productive, innovative, creative, and prosperous views so that good governance is formed [1].

Technology in the Village helps the process of making respectable and quality courtiers. The transparency of government activities and public services supports this. Transparency can be said to facilitate access and freedom for the public to access information related to government processes in their regions, where this element of transparency has an informative and open feature [2]. Sanrobone Village, Takalar Regency, located in South Sulawesi Province, implements an e-government-based system. This system is considered very helpful in administering government in Village. In addition to using the E-Government application in the Village, there are positive and negative perceptions from the village community.

Sentiment analysis can be used as a way to get information about the products or services being offered. Sentiment analysis uses perceptions or opinions and then analyzes them to find the meanings and emotions individuals or groups express. Sentiment analysis, also known as Opinion mining, uses NLP – Natural Language Processing to follow public opinion about a particular topic for any product or service [3]–[5].

Sentiments come from people's perceptions that have positive, negative, and neutral meanings related to something. The application of sentiment analysis can also be considered perception mining, such as analyzing opinions, evaluations, attitudes, and emotions toward something that impacts the environment [6]. Sentiment analysis and opinion mining offer tremendous promise for obtaining valuable insights and information from educational data. This study also identifies some of the challenges that researchers and practitioners face when applying sentiment analysis and opinion mining to educational data, such as the diversity of data types, the complexity of academic texts, complex label annotations, and privacy and ethical concerns, to provide valuable guidance for researchers and practitioners in understanding the potential and challenges of using these techniques in educational contexts [7].

According to Chen et al., the direction of user sentiment in social media affects how well sentiment analysis performs. Sentiment analysis methods can categorize sentiments in information published on social media more accurately if they are based on individual sentiment preferences and patterns. Sentiment analysis may grow using sentiment orientation as a development [8]. Feng et al. [9] added that customized text can increase the performance of sentiment analysis. By generating more varied training data, the sentiment analysis model can improve its recognition and classification of sentiments in new or previously unknown texts. Adding

text can improve the accuracy and precision of sentiment analysis.

Xu et al. [10] highlighted numerous developing trends in sentiment analysis on social media in their study. Trends include using machine learning methods and advanced natural language processing techniques, developing user-driven approaches to understanding individual contexts and preferences, using cross-platform data to improve comprehensive sentiment analysis, and a greater emphasis on multilingual sentiment analysis. Xu also discusses the difficulties in social media sentiment analysis, such as noise, data fluctuation, and slang, deemed non-standard language.

The method for doing sentiment analysis and detecting false reviews uses ontologies. Ontologies model knowledge about a specific area, including its ideas, relations, and attitudes. This study uses ontology to model understanding of the product or service under consideration. An ontology-based technique for identifying false reviews can be helpful. Sentiment analysis models may detect abnormalities and inconsistencies in studies and suspicious patterns that signal fake reviews by exploiting information structured in ontologies [11].

For e-commerce product reviews, naive Bayes can yield reliable sentiment classifications. Models learning new information continuously can cope with shifting sentiment patterns and enhance sentiment categorization performance on more recent evaluations. Feng et al. (2020) employ naive Bayes to construct a sentiment analysis system to help e-commerce consumers better comprehend product reviews and make more informed judgments [12]. Regarding classification accuracy and capacity to identify sentiment at the sentence level, S2SAN outperforms other sentiment analysis approaches. The S2SAN method assists in comprehending the many components of online reviews and describes the feelings linked with each statement in greater depth [13].

Mukherjee et al. collected data that included statements with varied negations regarding positive and negative attitudes. The use of dissolution in phrases may substantially impact sentiment analysis and polarity identification and significantly modify the meaning of a sentiment. It, therefore, offers a better knowledge of the intricacies of sentiment analysis in dealing with sentence negation [14].

The Naive Bayes method aids in forecasting user attitudes toward specific films. This sentiment data may subsequently be input into the film recommendation system to deliver more personalized and user-specific suggestions [15]. Sentiment analysis to extract sentiment from movie user reviews. Furthermore, the LDA approach assesses the review's main issues. Film recommendation systems can give consumers more accurate and relevant

suggestions using emotional information and thematic subjects [16].

Ossia et al. suggested a method for analyzing user evaluations and replies to diabetic mobile applications that combines DNN and LDA. DNN is used to predict sentiment in reviews, whereas LDA is used to examine thematic issues in reviews. The findings of Ossia's study may successfully forecast sentiment and perform theme analysis on diabetic mobile app evaluations. This study gives a deeper understanding of the attitudes and demands of diabetic mobile app users by using the strength of DNN in investigating sentiment patterns and the power of LDA in identifying thematic subjects [17].

The data source for sentiment analysis in this study is a review regarding the use of e-government in Sanrobone Village, Takalar Regency. Based on the introduction above, this study uses a sentiment analysis algorithm to see public opinion, a naïve Bayes algorithm to predict public opinion, and an algorithm in topic modeling, namely LDA (Latent Dirichlet Allocation), related to the use of e-government in Sanrobone Village, Takalar Regency.

2. Material & Methods

2.1. Research Approach

This study applies a text mining algorithm with a sentiment analysis method to see positive, negative, and neutral public perceptions and also uses topic modeling to get the most frequently appearing topics in the data. The stages of this research are as follows:

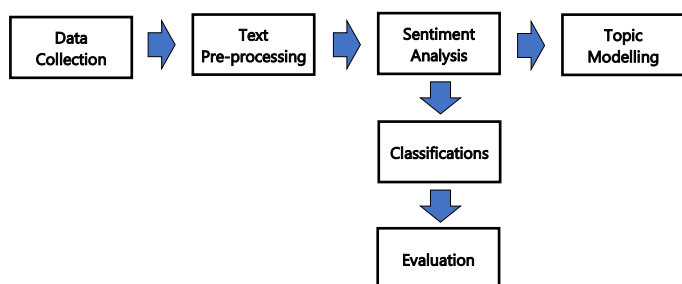


Figure 1. Research Stages.

2.2. Data Collection

The data for this study was collected through interviews with Sanrobone Village residents. The data collection process involved gathering two main data attributes: the name of the participants and their responses related to the usage and perception of e-government services. To ensure the reliability and validity of the data, a total of 600 data points were collected, consisting of 500 training data and 100 test data. This division allowed for comprehensive coverage of the community's perspectives and ensured a robust dataset for analysis and evaluation.

The interviews were conducted using a structured questionnaire that included specific questions related to the ease of using e-government services, the need for assistance or support, the association of e-government with village management, the accessibility of services, the utilization of letters and permits, the role of residents in e-government, and the expected impact on village management and service provision. The data collection process aimed to capture a diverse range of opinions and experiences from the residents of Sanrobone Village, providing valuable insights into their perceptions and expectations regarding e-government services.

Table 1. Data attributes

No.	Attribute	Information
1	Subject "x"	Community as a source of information
2	Opinion	The residents of Sanrobone village's thoughts on the usage of e-government

In this case, the source of knowledge is "the people of Sanrobone Village." This signifies that the data and information utilized in the study or analysis came directly from the citizens of the Village. This data was gathered through surveys and conversations with members of the Sanrobone Village community. In this scenario, the residents of the Village's attitudes toward using e-government are being examined. This viewpoint might be gathered by a survey that contains questions or remarks about the usage of e-government information systems in the Village.

This information may include people's perspectives, opinions, or preferences for using e-government and their experiences and impressions of the system's advantages, problems, or achievements. Information originating from the inhabitants of the Village has the benefit of offering direct insight into the community's viewpoints and real-world experiences with e-government.

2.3. Text Pre-processing

At this stage, data processing is carried out, and the results of data processing are used as a dataset to be processed to proceed to the sentiment analysis process [6] [18]. The stages of text pre-processing start from:

1. Transformation

Case folding and cleaning are part of the data transformation. This step is carried out to convert each character into lowercase text, remove all unnecessary accents, detect HTML tags, parse the text, and finally delete the URL if it is contained in the text.

2. Tokenization

At the tokenization stage, sentences are broken down into words. Tokenization also removes the symbols attached to penalties.

3. Filtering

The filtering process is used to get the essential words in a sentence from the tokenization results. This stage uses stop words from literature. Literary stop words are Indonesian stop words. The results of the process at this stage can be used as a benchmark that has an essential meaning in the sentence text that will be used.

4. N-Gram Range

The N-gram stage uses a range of 2 words, meaning that the results from the previous phase will be selected into two words. The value of the N-Gram range is significant for getting the keywords from the processed sentences.

2.4. Sentiment Analysis

There are three layers of sentiment analysis: document-level sentiment analysis, phrase-level sentiment analysis, and aspect-level sentiment analysis. Opinion detection at the document and sentence levels is in high demand and commonly employed. However, it is highly challenging at the aspect level due to deeper detection [19].

Document sentiment level analysis is a type of sentiment analysis in which all of the sentiments or attitudes included in a document, such as a news story, product review, or social media post, are evaluated. Document sentiment level analysis aims to comprehend the general opinion or overall appraisal of a topic or entity discussed in the document [20].

This method of phrase-level sentiment analysis allows you to comprehend the sentiments in each sentence independently. This is beneficial in situations such as product opinion analysis, identifying sentiment in consumer evaluations, and monitoring social media interactions. Sentence-level sentiment analysis is frequently paired with document-level sentiment analysis or more contextual sentiment analysis to acquire a complete picture of the sentiments in a text [21].

The aspect-based sentiment analysis approach recognizes relevant textual characteristics, such as product attributes, particular subjects, or entities mentioned. The text is then examined to discover the emotions connected with these features. The technique using contrastive learning developed by Xu et al. [22] can increase the performance of aspect-based sentiment analysis. The sentiment analysis model may better distinguish attitudes connected to certain features by considering the difference between distinct text pairings. This method aids

in increasing the precision and accuracy of sentiment analysis about certain parts of the text.

The stages of sentiment analysis are used to process the opinions of the people of the Village. Sentiment analysis is a method used to analyze people's responses. The analysis is divided into three parts: positive, negative, and neutral [3]. Sentence sentiment analysis attributes of reactions to the data will be analyzed to find out the opinion of the people of the Village regarding the use of e-government.

2.5. Topic Modelling

This LDA model, which generates document topics, has three layers: the word, subject, and document layers. LDA is an unsupervised machine learning model that may be utilized for topic extraction across different languages and domains. LDA can better characterize document content than conventional keyword extraction algorithms wholly and correctly. Additionally, LDA is swift and precise while doing brief text topic extraction tasks [16].

Topic modeling in this study uses the LDA (Latent Dirichlet Allocation) model. LDA can generally find the topic of a statement in a document. LDA creates a factorization matrix from documents with many topics, from which word probability is formed from the word distribution [23], [24]. LDA work steps are as follows [25]:

1. The first stage is initialization; this stage aims to determine the frequency of the appearance of words in all text files.
2. The following process is the sampling stage. This process can help define a new topic for every document word.
3. The final stage is the calculation of the last parameters. This process calculates the number of documents on all topics and the number of words on all issues according to the matrix of topic words and topic documents.

2.6. Classification

The Naive Bayes technique can handle issues with many features and minimal computing cost, making it appropriate for text classification and sentiment analysis. This algorithm excels at quickly and successfully deriving patterns from training data [26].

Based on how frequently the characteristics connected to each class appear in the training dataset, Naive Bayes determines the likelihood of each type. Next, the most likely category for a specific new instance is predicted using these probabilities. The classification process uses naïve Bayes. The naïve Bayes method is used to label the class responses of the Village community, namely positive and negative reactions. The Naïve Bayes equation is as follows [27]:

$$P(B|G) = \frac{P(G \setminus B)P(B)}{P(G)} \quad (1)$$

In the equation above, variable B describes the input of data classified with positive and negative labels; variable G is data that does not have a class. P(G) is the probability of G, P(G \setminus B) The probability of the hypothetical condition.

2.7. Evaluation

This process will visualize the results of the previous stages and will also be tested based on the confusion matrix. The confusion matrix serves to see the performance of the classification results. The results obtained can provide information regarding errors in classifying inappropriate data. The confusion matrix compares the actual value and the predicted results from the classification results [28]. This three-way confusion matrix may be used to compute more comprehensive classification assessment metrics such as accuracy, precision, recall, sensitivity, F1-score, and others [29].

3. Result and Discussions

3.1. The Result of Text Pre-processing

The results of the pre-processing of the data are shown in Figure 2.



Figure 2. Word Cloud Text Pre-processing.

Based on the results of the pre-processing of the data, the ten words that most often appear in the responses of the people of Sanrobone Village are: easy (122), help (96), Village (80), accessed (80), letter (80), permit (77), resident (73), manage (60), service (52), and people with (52) words.

The term "easy" in the context of e-government adoption in Village suggests that the community views the acquisition and application of e-government services as uncomplicated and direct. The positive perception of the community regarding the accessibility and usability of e-

government services is reflected in this favorable impression.

Utilizing the term "help" implies that certain community members experience a requirement for aid or reinforcement while availing of e-government services. This observation suggests the possibility of encountering obstacles or complexities in comprehending or employing the service. This statement underscores the significance of furnishing sufficient direction and assistance to guarantee the efficient utilization of e-government services.

The term "village" connotes that the e-government's reception by the populace is intrinsically linked to their respective Village or community. This implies that the e-government services are tailored and aimed toward rural communities. This statement suggests that the utilization of e-government services significantly impacts the management and administration of villages.

The term "accessed" denotes the active utilization or access of e-government services by the community in the Village. This exemplifies a propensity to adopt technological advancements and the advantageous features and availability of the e-governance services provided. The statement highlights the community's proactive utilization of digital platforms for accessing available services.

The terminology employed, namely "letter," implies that the e-government services offered in Sanrobone Village pertain to utilizing or manipulating written correspondence. This may pertain to tasks such as submitting application letters or transmitting notifications to the village administration via the e-government portal. The text emphasizes the application of electronic communication in the context of administrative functions.

The term "permit" connotes that the utilization of e-government is linked to the acquisition of licenses or permits, as perceived by the community. The statement suggests that the e-government services offered in Sanrobone Village encompass features that enable its inhabitants to submit applications for permits or licenses via the e-government portal. The statement suggests that the community anticipates a streamlined and efficient protocol for acquiring requisite authorizations.

The term "resident" implies a correlation between the community's response towards e-government and their position as inhabitants of the locality. This suggests that the e-government services are tailored to meet the needs and involvement of the inhabitants of Sanrobone Village. This statement highlights the significance and pertinence of e-government in catering to the distinct demands and prerequisites of the regional populace.

The community anticipates the utilization of e-government services to facilitate the management of diverse elements or activities in the Village, as implied by

the term "manage." This may entail various responsibilities, including supervising administrative functions, delivering public amenities, or coordinating communal gatherings within a village. The statement denotes the community's anticipation of the e-government's involvement in augmenting village administration and improving service provision.

The community commonly perceives e-government as a means of providing services to its residents, as implied by the term "service." The statement suggests that the community anticipates e-government services to provide valuable and useful services tailored to their requirements. This highlights the significance of providing proficient and impactful services via the e-government platform to meet the anticipations of the populace.

3.2. Sentiment Analysis Results

The sentiment analysis results are based on 500 training and 100 test data. The prediction results are obtained using the Naïve Bayes algorithm as follows:

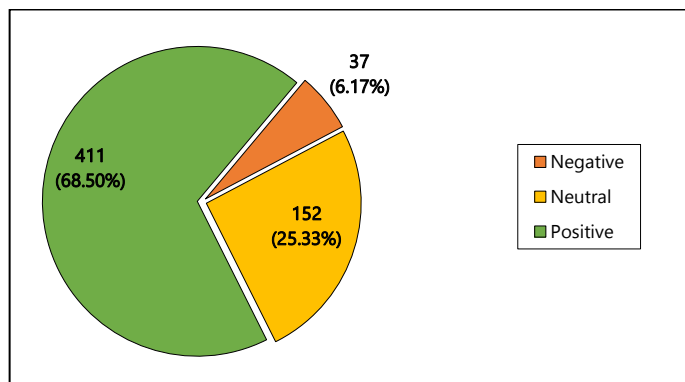


Figure 3. The Results of Sentiment Analysis on the Use of E-Government in the Village.

The sentiment analysis revealed that out of the total dataset, 563 public opinions were classified as positive, while 37 opinions were categorized as negative. This indicates that most public opinions positively opposed using e-government services in Sanrobone Village.

Subset of 10 data points from the positive sentiment category was randomly selected for further analysis. These data points were subjected to sentiment analysis to provide more detailed insights into the positive public opinions. The results of this analysis confirmed that all 10 data points exhibited a positive sentiment toward the use of e-government services.

These findings highlight the overall positive perception and satisfaction of the community in Village regarding the e-government services offered. The sentiment analysis results reinforce the notion that e-government has been well-received by the residents,

indicating its effectiveness in meeting their needs and expectations.

Table 2. Result of 10 sentiment analysis data on the "Positive" training data.

No.	Results	Opinions
1	Positive	Easily accessible and excellent response
2	Positive	It is easy to make a power of attorney
3	Positive	Friendly and easy to get the married permit
4	Positive	Easy licensing
5	Positive	Facilitate licensing
6	Positive	The app is handy for me
7	Positive	Easy licensing
8	Positive	Online services facilitate the community
9	Positive	Can improve the quality of village services
10	Positive	Robust digital services

Opinions in Table 2 raise sentences that have positive meanings from the words accessible, excellent, good, facility, very useful, improving quality, and robust in detection as supporters of positive opinions regarding the use of the e-government system.

Table 3. The results of sentiment analysis data on the "Negative" training data.

No.	Results	Opinions
1	Negative	difficult for parents to access
2	Negative	a challenge for parents

Table 3 explains that the word was difficult and challenged a term with a negative value.

The phrase "difficult for parents to access" indicates impediments or challenges parents have while attempting to use e-government services. This might be due to challenges with technical accessibility, a lack of knowledge or skills in utilizing e-government systems, or a lack of proper assistance or advice.

The phrase "a challenge for parents" indicates how difficult it might be for parents to use e-government. This might indicate that parents find it difficult to use or comprehend e-government services, which would prevent them from doing so efficiently.

The conclusion derived from these negative statements is that not all parents or the community in Village feel comfortable or capable of utilizing e-government services. They must overcome numerous difficulties or problems to use these services. It is critical to pay attention to and respond to negative criticism since it

can give insight into areas that need to be changed or enhanced in Village's e-government implementation. Efforts must be made to improve accessibility, give enough aid or advice, and solve issues parents or communities encounter while using e-government.

Opinions in Table 4 raise sentences with neutral meanings from the words, excellent, clear," and reaction, detected as supporting neutral opinions regarding using the e-government system.

Table 4. The results of sentiment analysis data on the "Neutral" training data.

No.	Results	Opinions
1	Neutral	Difficult access, maybe because of the network
2	Neutral	The information provided is constantly updated
3	Neutral	The resident mail service is excellent and clear
4	Neutral	BLT system information is quite clear.
5	Neutral	reaction time is not obvious

Based on the analysis of public sentiment regarding the use of e-government in the Village, it is excellent. The community's positive response is expressed in the word "easy" and can improve the quality of service. Negative reactions are described by the word "difficult," which underlies the word "difficult" to be negative response is people who do not understand IT. A neutral response is expressed by the word "maybe," which only provides information objectively without giving a specific assessment or evaluation.

3.3. Classification

During the classification stage, two distinct data types are employed: training and test data. The former is utilized as the learning data to generate predictive outcomes, while the latter is employed to predict the results. The Naïve Bayes algorithm is utilized to predict and categorize public sentiments into either positive or negative classifications. Figure 4 displays the outcomes of the forecast derived from the examination data.

Figure 4 employs a sample size of 100 data points for classification. The anticipated consequences of these viewpoints are as follows: The data set consisted of 100 opinions, which were analyzed and classified according to their sentiment. Specifically, 82 opinions were assigned to the positive category, five were allocated to the negative category, and 13 were classified as neutral. The outcomes derived from the classification procedure offer valuable perspectives on the prevailing attitudes of the general populace towards the implementation of e-government in Village. The preponderance of anticipated outcomes

suggests a favorable disposition, with a considerable proportion of viewpoints being categorized as affirmative. Nonetheless, a minor fraction of viewpoints was categorized as unfavorable, implying the presence of dissent or apprehensions among the populace.

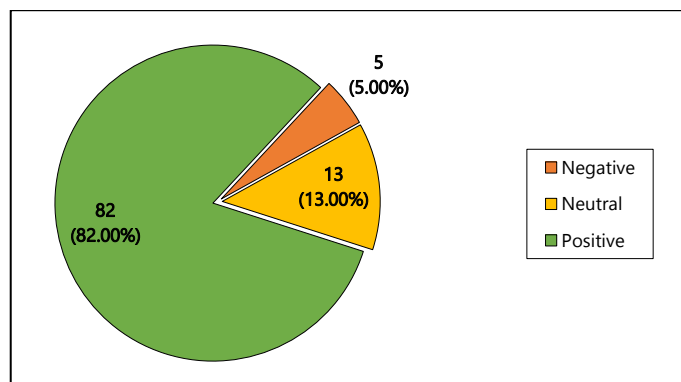


Figure 4. Prediction results of the classification of the naïve Bayes algorithm for using e-Government.

The categorization also identified a particular group of viewpoints classified as impartial. The viewpoints above could suggest a dearth of definitive attitudes or equivocal perspectives regarding providing e-governance facilities in the Village. The classification outcomes provide significant insights into the distribution of sentiments among the general public's views on e-government services. The findings above serve as a foundation for additional examination and comprehension of the populace's viewpoint and acceptance of electronic government services in Village.

3.4. Topic Modeling

The topic modeling results using the LDA algorithm produce three topic keywords—results in Table 5.

Table 5. Topic Modeling Results Using the LDA Algorithm.

No.	Topic	Topic Keywords
1	Topic 1	Manage, Access, Help, Village, Information, Easy, Service, Network, Info, BLT System
2	Topic 2	People, Parents, Old, Residents, Village, Helping, Helping Residents, Difficult, Difficult People, Access
3	Topic 3	Easy, Mail, service, Licensing, Easy Licensing, Accessible, Response, Accessing Response, Easy Accessing, Steady Digital

Evaluation of the topic modeling results using the log perplexity. The results of the confusion log with a value of 20.39 mean that it cannot predict the appearance of words in the community opinion data of Village optimally. This occurs because the sentences in the data are not correctly

structured according to the structure of the Indonesian language.

Fellnhofer's research found that higher degrees of optimism and attentiveness had a favorable link with information-finding ability [30]. According to Fellnhofer's findings, the degree of positivity in the usage of e-government in the Sanrobone village demonstrates that information and village administration services are highly beneficial to the community and increase the quality of village services.

3.5. Evaluation

Evaluation of the results in this study using a confusion matrix; the first is to measure the performance of the naïve Bayes algorithm in classifying community opinion in Village. The accuracy of the comparison of the correct prediction value on the data obtained is 98%, precision is the comparison of the accurate positive prediction with all positive prediction results with a value of 98%, recall is the comparison of the correct positive prediction with all the data that is correctly positive with a value of 98%, and F1-Score is the comparison of the average mean precision and recall obtained with a value of 98%. The results of measuring the performance of the Naive Bayes algorithm serve as a benchmark in testing the results obtained using the confusion matrix values of accuracy, precision, recall, and F1-Score, so the actual evaluation and prediction of the classification results can be seen in Table 6.

Table 6. The Results of The Naïve Bayes Algorithm Confusion Matrix on Community Opinion Data.

		Predicted			Σ
		Negative	Neutral	Positive	
Actual	Negative	100.0%	0.0%	0.0%	37
	Neutral	0.0%	94.4%	0.0%	152
	Positive	0.0%	5.6	100.0%	411
	Σ	37	161	402	600

The results in Table 6 explain that the accuracy of the Naive Bayes algorithm in classifying public opinion is excellent by accurately predicting the negative results of 37 data with a value of 100% according to the actual data and positively predicting 402 data with a predictive value of 100%. Even though the neutral prediction results have a prediction error of 5.6% from 94.4%, the data should be included in the positive classification. The total community opinion data in Village is 600.

The results of this study show that the opinion of the people of Village regarding the use of the e-government information system is that 411 people have a positive opinion, 37 have a negative opinion, and 152 have a neutral opinion. Furthermore, the prediction results

obtained using the naïve Bayes algorithm from 100 test data points showed the predicted results of 82 people with a positive opinion, 5 with a negative opinion, and 13 with a neutral opinion.

The results of applying the LDA algorithm obtained three topic keywords: topic 1 (Managing, Accessing, Helping, Village, Information, Easy, Service, Network, Info, BLT System), topic 2 (Helping, Helping Residents, Difficult, Difficult People, Access), and topic 3 (Permissions, Accessible, Response, Accessible Response, Easily Accessible, Steady Digital). The test results and scores related to the performance evaluation of the naïve Bayes algorithm obtained 98% accuracy, 98% precision, 98% recall, and a 98% F1 score. The confusion matrix results were excellent, predicting both positive and negative opinions with an accuracy of 100%, even though there was an error in prediction on a neutral opinion of 5.6 percent of the results obtained 94.4%. Based on the results obtained, the naïve Bayes algorithm can classify very well and accurately the opinions of the people of the Village.

The majority of the residents of the Village had a favorable opinion on using e-government information technologies, according to this survey. Additionally, opinion and topic modeling relevant to e-government are analyzed and forecasted using naïve Bayes and LDA algorithms. The naïve Bayes method produces very good results with high levels of accuracy, precision, recall, and F1 scores when its performance is evaluated. It should be highlighted. Nevertheless, that neutral opinion can be inaccurately predicted. These results offer insightful information on how the Village community views and assesses the usage of e-government and the efficacy of the algorithm employed in this study.

4. Conclusion

Based on the findings of this study, it is possible to infer that most Village people favor using e-government information systems. The naïve Bayes method utilized in this study provides a high degree of accuracy in categorizing public opinion, predicting negative and positive attitudes. The naïve Bayes algorithm performance evaluation findings demonstrate a high degree of accuracy, precision, recall, and an F1 score of 98%. Despite flaws in identifying neutral thoughts, the confusion matrix findings reveal a fairly accurate prediction in distinguishing positive and negative opinions. The use of the LDA algorithm yields three subject keywords relating to the use of e-government. Overall, this study gives useful insights into community attitudes and evaluations of the usage of e-government.

Acknowledgments

We thank the Makassar State University Research and Community Service Institute for supporting this research. Their resources and guidance have been instrumental in completing this work. We would also like to thank the Sanrobone Village Government in Sanrobone District, Takalar Regency, Indonesia. Their cooperation and collaboration have provided this study's necessary data and insights. The contributions and support from these institutions have greatly contributed to the quality and comprehensiveness of this research. We are grateful for their involvement and assistance throughout the project.

References

- [1] S. Sulaiman, A. Nizam, F. Farid, T. M. Kesuma, I. Madjid, and R. Siregar, "The role of good government governance principles in enhancing the performance of public organization in Aceh, Indonesia," in *1st Aceh Global Conference (AGC 2018)*, 2019, pp. 731–738.
- [2] H. Addink, *Good governance: Concept and context*. Oxford University Press, 2019.
- [3] P. Mehta and S. Pandya, "A review on sentiment analysis methodologies, practices and applications," *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 601–609, 2020.
- [4] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics*, vol. 9, no. 3, p. 483, 2020.
- [5] N. Leelawat *et al.*, "Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning," *Heliyon*, vol. 8, no. 10, p. e10894, 2022.
- [6] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5731–5780, 2022.
- [7] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, "Sentiment analysis and opinion mining on educational data: A survey," *Nat. Lang. Process. J.*, vol. 2, p. 100003, 2023, doi: <https://doi.org/10.1016/j.nlp.2022.100003>.
- [8] J. Chen, N. Song, Y. Su, S. Zhao, and Y. Zhang, "Learning user sentiment orientation in social networks for sentiment analysis," *Inf. Sci. (Ny.)*, vol. 616, pp. 526–538, 2022.
- [9] Z. Feng, H. Zhou, Z. Zhu, and K. Mao, "Tailored text augmentation for sentiment analysis," *Expert Syst. Appl.*, vol. 205, p. 117605, 2022, doi: <https://doi.org/10.1016/j.eswa.2022.117605>.
- [10] Q. A. Xu, V. Chang, and C. Jayne, "A systematic review of social media-based sentiment analysis: Emerging trends and challenges," *Decis. Anal. J.*, p. 100073, 2022.
- [11] D. U. Vidanagama, A. T. P. Silva, and A. S. Karunananda, "Ontology based sentiment analysis for fake review detection," *Expert Syst. Appl.*, vol. 206, p. 117869, 2022, doi: <https://doi.org/10.1016/j.eswa.2022.117869>.
- [12] F. Xu, Z. Pan, and R. Xia, "E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework," *Inf. Process. Manag.*, vol. 57, no. 5, p. 102221, 2020.
- [13] P. Wang, J. Li, and J. Hou, "S2SAN: A sentence-to-sentence attention network for sentiment analysis of online reviews," *Decis. Support Syst.*, vol. 149, p. 113603, 2021, doi: <https://doi.org/10.1016/j.dss.2021.113603>.
- [14] P. Mukherjee, Y. Badr, S. Doppalapudi, S. M. Srinivasan, R. S. Sangwan, and R. Sharma, "Effect of Negation in Sentences on Sentiment Analysis and Polarity Detection," *Procedia Comput. Sci.*, vol. 185, pp. 370–379, 2021, doi: <https://doi.org/10.1016/j.procs.2021.05.038>.
- [15] N. Pavitha *et al.*, "Movie recommendation and sentiment analysis using machine learning," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 279–284, 2022, doi: <https://doi.org/10.1016/j.gltp.2022.03.012>.
- [16] Y. Zhang and L. Zhang, "Movie Recommendation Algorithm Based on Sentiment Analysis and LDA," *Procedia Comput. Sci.*, vol. 199, pp. 871–878, 2022, doi: <https://doi.org/10.1016/j.procs.2022.01.109>.
- [17] C. I. Ossai and N. Wickramasinghe, "Sentiments prediction and thematic analysis for diabetes mobile apps using Embedded Deep Neural Networks and Latent Dirichlet Allocation," *Artif. Intell. Med.*, vol. 138, p. 102509, 2023, doi: <https://doi.org/10.1016/j.artmed.2023.102509>.
- [18] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, 2020.
- [19] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Syst.*, vol. 226, p. 107134, 2021.
- [20] F. Alshuwaier, A. Areshey, and J. Poon, "Applications and Enhancement of Document-Based Sentiment Analysis in Deep learning Methods: Systematic Literature Review," *Intell. Syst. with Appl.*, p. 200090, 2022.
- [21] M. Araújo, A. Pereira, and F. Benevenuto, "A comparative study of machine translation for multilingual sentence-level sentiment analysis," *Inf. Sci. (Ny.)*, vol. 512, pp. 1078–1102, 2020.
- [22] L. Xu and W. Wang, "Improving aspect-based sentiment analysis with contrastive learning," *Nat. Lang. Process. J.*, p. 100009, 2023.
- [23] M. C. Rahmadan, A. N. Hidayanto, D. S. Ekasari, and B. Purwandari, "Sentiment analysis and topic modelling using the lda method related to the flood disaster in jakarta on twitter," in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 2020, pp. 126–130.
- [24] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [25] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimed. Tools Appl.*, vol. 78, pp. 15169–15211, 2019.
- [26] A. Dag, A. Z. Dag, A. Asilkalkan, S. Simsek, and D. Delen, "A Tree Augmented Naïve Bayes-based methodology for classifying cryptocurrency trends," *J. Bus. Res.*, vol. 156, p. 113522, 2023, doi: <https://doi.org/10.1016/j.jbusres.2022.113522>.
- [27] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowledge-Based Syst.*, vol. 192, p.

- 105361, 2020.
- [28] E. Sutoyo and A. Almaarif, "Educational data mining for predicting student graduation using the naïve bayes classifier algorithm," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 4, no. 1, pp. 95–101, 2020.
- [29] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf. Sci. (Ny)*, vol. 507, pp. 772–794, 2020, doi: <https://doi.org/10.1016/j.ins.2019.06.064>.
- [30] K. Fellnhofner, "Positivity and higher alertness levels facilitate discovery: Longitudinal sentiment analysis of emotions on Twitter," *Technovation*, vol. 122, p. 102666, 2023.



© 2023 by the authors. Licensee by Three E Science Institute (International Journal of Environment, Engineering & Education).

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 (CC BY SA) International License. (<http://creativecommons.org/licenses/by-sa/4.0/>).