e-ISSN: 3108-9828 | p-ISSN: XXXX-XXXX
Volume 1 No 2 November 2024 | Page: 30-34
DOI: https://doi.org/10.62712/jocsaic.v1i2.18

Journal of Computer Science Artificial Intelligence and Communications

Journal Homepage: https://journals.raskhamedia.or.id/index.php/jocsaic

Sentiment Analysis of Social Media towards Public Services Using Naive Bayes and Text Mining

Rusmin Saragih[1,*], Mardiah[2], Deni Apriadi[3]

[1]prodi Sistem Informasi, STMIK Kaputama, Indonesia
[2]prodi sistem informasi, Universitas Nahdlatul Ulama Sumatera Utara, Indonesia
[3]Prodi Sistem Informasi, STMIK Bina Nusantara Jaya Lubuklinggau, Indonesia
Author(s) Email: [1]evitha12014@gmail.com, [2]mardiahindin23@gmail.com, [3]denidrv@gmail.com

ARTICLE INFO

Article history:

Received July 10, 2024
Revised July 12, 2024
Accepted July 12, 2024
Publish November 30, 2024

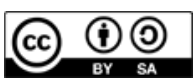
ABSTRACT

The rapid development of information and communication technology has driven the increased use of social media as a means of interaction between the public and service providers. Social media has become a platform for the public to express their opinions on the quality of services they receive, whether in the form of praise, suggestions, or complaints. Therefore, sentiment analysis of social media data can be a strategic tool in evaluating the performance of public services. This research aims to analyze public sentiment towards public services by utilizing text mining techniques and the Naive Bayes Classifier algorithm. The data used was collected from social media platforms such as Twitter and Facebook, followed by a text preprocessing stage that included tokenizing, stopword removal, and stemming. Subsequently, the data was analyzed to classify sentiment into positive, negative, and neutral categories. The test results show that the Naive Bayes algorithm is capable of classifying data with a satisfactory level of accuracy, making it an efficient method for monitoring public perception in real-time. This research contributes to supporting decision-making by government agencies regarding the improvement of public service quality based on publicly available feedback from social media.

Keywords:

Sentiment Analysis, Social Media, Public Services, Naive Bayes, Minería de textos

Corresponding Author:
Rusmin Saragih,
prodi sistem informasi, Universitas Nahdlatul Ulama Sumatera Utara, Indonesia
Email: evitha12014@gmail.com



1. INTRODUCTION

The development of digital technology and the internet has brought significant changes in the way society interacts, expresses opinions, and responds to various matters, including public services[1]. One of the results of this change is the increased use of social media such as Twitter, Facebook, Instagram, and similar platforms as a means of communication for the public with government agencies, public service providers, and related parties[2]. Social media is not only used for ordinary social activities but has also become a platform for the public to voice opinions, criticisms, complaints, or appreciation regarding the performance of public services[3]. Therefore, social media has now become a very rich and highly valuable source of data in measuring public perception of the quality of services they receive[4]. Public services

are all forms of services provided by government or private institutions to meet the basic needs and rights of the community[5]. The quality of public services has become one of the main indicators in assessing the effectiveness and efficiency of government administration[6]. In this context, it is important for the government to actively evaluate and improve the services they provide[7]. However, conventional mechanisms for obtaining feedback from the public, such as surveys or suggestion boxes, often have limitations in terms of reach, time, and data representation. This causes the results of public service evaluations to often not reflect the overall and actual conditions or perceptions of the community[8].

As an alternative, the utilization of social media data for sentiment analysis offers a more responsive and dynamic approach[9]. Sentiment analysis is the process of identifying and categorizing opinions expressed in text, specifically to determine whether the author's attitude towards a particular topic is positive, negative, or neutral[10]. By using this technique, government institutions can automatically filter and analyze thousands to millions of comments from the public spread across various social media platforms, and conclude the general public's perception of the services provided[11]. In this study, a text mining approach and the Naive Bayes Classifier algorithm were used to analyze public sentiment towards public services[12]. Text mining is the process of extracting valuable information from unstructured text data. This process includes several stages, starting from data collection, preprocessing (text cleaning), feature extraction, to classification. One of the main challenges in text mining is managing the large volume of data and the diversity of languages used on social media, such as abbreviations, emoticons, slang, and non-standard spelling. Therefore, the preprocessing stage becomes very important to ensure that the data being analyzed is relevant and clean[13].

The Naive Bayes algorithm is one of the probabilistic-based classification methods often used in text analysis due to its simplicity, efficiency, and ability to handle large-sized data. Although it uses the assumption of independence between features, Naive Bayes has proven to be quite reliable in various sentiment analysis studies. In the context of this research, the Naive Bayes algorithm is used to classify public opinions contained in social media texts into three categories: positive, negative, and neutral. The results of this classification can then be used as evaluation material and decision-making by government agencies[14]. The use of social media as a data source also provides advantages in terms of speed and data continuity. Compared to conventional survey methods that require high time and cost, social media data analysis can be conducted in real-time and continuously[15]. This allows the government to quickly detect changes in public opinion and respond to issues that arise within society. Thus, social media-based sentiment analysis can become a strategic tool in achieving public services that are more responsive, transparent, and oriented towards public satisfaction.

This research has a high level of urgency considering the importance of improving the quality of public services in realizing good governance. In addition, the increasing dependence of society on social media as a means of communication also demands the optimal utilization of that data for policy and service improvement. By utilizing the Naive Bayes algorithm and text mining techniques, it is hoped that the results of this research can provide a more objective and comprehensive picture of public sentiment towards public services, as well as serve as valuable input for policymakers.

Overall, this research aims to: (1) collect public opinion data from social media related to public services; (2) preprocess the text data to prepare it in an analyzable format; (3) apply the Naive Bayes algorithm in the sentiment classification process; and (4) evaluate the classification results as a basis for analyzing public perception of government services. It is hoped that this research can serve as a reference and foundation for the development of sentiment analysis systems that can be used by government agencies to improve the quality and effectiveness of public services in the current digital era.

2. RESEARCH METHODOLOGY

This research uses a quantitative approach with an exploratory method to analyze text data from social media related to public sentiment towards public services. This approach was chosen because it can systematically identify patterns, opinions, and societal trends through the processing of unstructured text data.
The steps of this research method generally consist of:

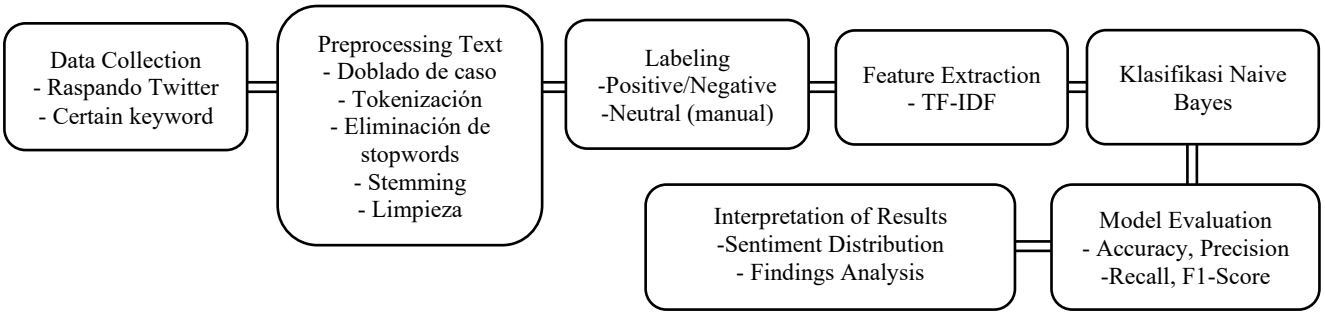


Figure 1. Research Flow Diagram

2.1 Data Collection

Data is collected from social media (e.g., Twitter) using web scraping techniques with relevant keywords such as "public service," "government services," "service complaints," and others. The data collected includes comment text, upload time, and usernames anonymously.

2.2 Preprocessing Text Data

Before analysis is conducted, the text data needs to be cleaned and processed. The preprocessing stages include:

a. Case Folding: Converting the entire text to lowercase.
b. Tokenizing: Breaking a sentence into words.
c. Stopword Removal: Removing common words that do not have significant meaning (such as "and", "that", "in", etc.).
d. Stemming: Changing a word to its base form (for example: "pelayanan", "melayani" becomes "layan").
e. Cleaning: Removing irrelevant numbers, symbols, URLs, and punctuation.

2.3 Sentiment Labeling

The data is manually labeled (training set) as positive, negative, or neutral sentiment. This is used as training data for the classification process.

2.4 Feature Extraction

Using the TF-IDF (Term Frequency-Inverse Document Frequency) method to convert text data into a numerical representation so that it can be processed by machine learning algorithms.

2.5 Classification with Naive Bayes

Using the Naive Bayes Classifier algorithm, the data is analyzed to classify the sentiment of each tweet or comment into three classes:

a. Positive
b. Negative
c. Neutral.

2.6 Model Evaluation

Evaluation is conducted using testing data by calculating the value:

a. Accuracy
b. Precision
c. Recuerda
d. F1-score

2.7 Model Evaluation

The classification results are analyzed to understand public perception of public services. Sentiment distribution diagrams or graphs will be used to visually display the results.

3. RESULT AND DISCUSSION

This research successfully classified public opinions on public services taken from social media, particularly Twitter. A total of 1,500 comment data were collected using keywords related to public services, such as "government services," "public complaints," "department responses," and others. After preprocessing, 1,200 data were used as training data and 300 data as testing data.

After going through the text mining and classification process using the Naive Bayes algorithm, the sentiment classification results are obtained as shown in Table 1 below:

Here is an example of the association rules obtained:

Table 1. Sentiment Classification Results on Test Data (n = 300)

Sentiment	Number of Data	Percentage (%)
Positif	110	36,67%
Negatif	130	43,33%
Netral	60	20,00%
Total	300	100,00%

Berdasarkan tabel di atas, dapat dilihat bahwa sentimen negatif mendominasi opini masyarakat terhadap layanan publik, diikuti oleh sentimen positif, dan netral. Hal ini menunjukkan bahwa masyarakat masih banyak mengeluhkan layanan yang diberikan oleh instansi pemerintah, seperti keterlambatan pelayanan, buruknya respons pegawai, atau ketidakjelasan prosedur.

Selanjutnya dilakukan pengujian akurasi model Naive Bayes terhadap data uji. Evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score. Berikut adalah hasil evaluasinya:

Table 2. Evaluation of Naive Bayes Algorithm Performance

Metrik	Positif	Negatif	Netral	Macro Average
Precision	0,85	0,80	0,76	0,80
Recall	0,82	0,84	0,72	0,79
F1-Score	0,83	0,82	0,74	0,80
Total Accuracy				0,81 (81%)

The evaluation results show that Naive Bayes can classify data with an accuracy of 81%, which is quite good for text classification. The highest F1-score is found in the positive and negative classes, while the neutral class has a slightly lower score. This is likely due to the ambiguity or subjectivity in neutral comments, as well as the algorithm's limitations in deeply understanding emotional context. Further analysis shows that positive comments generally appear on fast and responsive online administrative services, such as online licensing services or digital queue systems. On the other hand, negative sentiments are often found in comments regarding complaints about slow face-to-face services, lack of transparency, and poor attitudes of field staff.

These findings provide a real picture of how the public assesses the performance of public services through social media. By utilizing these analysis results, government agencies can identify service areas that need improvement and use social media as a tool for real-time monitoring of service quality.

4. CONCLUSIONS

This research shows that social media-based sentiment analysis can be an effective tool in evaluating public perception of public services. Through a text mining approach and the application of the Naive Bayes algorithm, public opinions contained in text data can be automatically classified into three sentiment categories: positive, negative, and neutral. The classification results on 300 test data show that the majority of public opinions are negative (43.33%), followed by positive opinions (36.67%) and neutral (20%). This reflects that there are still many complaints and dissatisfaction with public services circulating on social media. From a performance perspective, the Naive Bayes algorithm shows good performance with an accuracy of 81%, as well as relatively high precision and recall values across all sentiment categories. This indicates that this method is quite reliable and can be implemented in an automated public opinion monitoring system. In addition, the classification results can serve as an initial indicator to detect issues in public services and can be used as evaluation material by government agencies. By leveraging the vast potential of real-time social media data, the government can respond more quickly to emerging issues in society. This research also opens up opportunities for the development of an artificial intelligence-based public sentiment monitoring system that is directly integrated with public service policies.

REFERENCES

[1] A. M. A. Ausat, "The role of social media in shaping public opinion and its influence on economic decisions," *Technol. Soc. Perspect.*, vol. 1, no. 1, pp. 35–44, 2023.
[2] V. Mabillard, R. Zumofen, and M. Pasquier, "Local governments' communication on social media platforms: Refining and assessing patterns of adoption in Belgium," *Int. Rev. Adm. Sci.*, vol. 90, no. 1, pp. 65–81, 2024.
[3] D. Hertati, "Improving the Quality of Public Services through Innovation of Online Complaints Based on the Wargaku Application," *J. La Soc.*, vol. 4, no. 3, pp. 97–102, 2023.
[4] M. Hasanuddin, "Journal of Computer Science Artificial Intelligence Optimization of Computer Network Performance Using Heuristic Algorithms," vol. 1, no. 1, pp. 12–17, 2024.
[5] C. K. Sanders and E. Scanlon, "The digital divide is a human rights issue: Advancing social inclusion through social work advocacy," *J. Hum. rights Soc. Work*, vol. 6, no. 2, pp. 130–143, 2021.
[6] T. Felício, A. Samagaio, and R. Rodrigues, "Adoption of management control systems and performance in public sector organizations," *J. Bus. Res.*, vol. 124, pp. 593–602, 2021.
[7] P. Vassilakopoulou, A. Haug, L. M. Salvesen, and I. O. Pappas, "Developing human/AI interactions for chat-based customer services: lessons learned from the Norwegian government," *Eur. J. Inf. Syst.*, vol. 32, no. 1, pp. 10–22, 2023.
[8] N. K. Marutlulle, "A critical analysis of housing inadequacy in South Africa and its ramifications," *Africa's Public Serv. Deliv. Perform. Rev.*, vol. 9, no. 1, p. 372, 2021.
[9] A. S. George and T. Baskar, "Leveraging big data and sentiment analysis for actionable insights: A review of data mining approaches for social media," *Partners Univers. Int. Innov. J.*, vol. 2, no. 4, pp. 39–59, 2024.
[10] F. Mehraliyev, I. C. C. Chan, and A. P. Kirilenko, "Sentiment analysis in hospitality and tourism: a thematic and methodological review," *Int. J. Contemp. Hosp. Manag.*, vol. 34, no. 1, pp. 46–77, 2022.

[11] J. C. Lyu and G. K. Luli, "Understanding the public discussion about the Centers for Disease Control and Prevention during the COVID-19 pandemic using Twitter data: Text mining analysis study," *J. Med. Internet Res.*, vol. 23, no. 2, p. e25108, 2021.
[12] W. B. Zulfikar, A. R. Atmadja, and S. F. Pratama, "Sentiment analysis on social media against public policy using multinomial naive bayes," *Sci. J. Informatics*, vol. 10, no. 1, pp. 25–34, 2023.
[13] S. Khodijah, C. A. Rizki, and M. Hasanuddin, "Journal of Computer Science Artificial Intelligence," vol. 1, no. 1, pp. 1–6, 2024.
[14] M. H. Shahidzadeh and S. Shokouhyar, "Shedding light on the reverse logistics' decision-making: a social-media analytics study of the electronics industry in developing vs developed countries," *Int. J. Sustain. Eng.*, vol. 15, no. 1, pp. 161–176, 2022.
[15] M. Hasanuddin, B. E. Susanto, S. Ginting, and F. Rizaldi, "Analisis Minat Siswa Kelas 1 SMK Pada Ekstrakulikuler Sepak Bola Dengan Metode Technology Acceptance Model," vol. 4, no. 1, pp. 52–58, 2025.