Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Analysis of Public Perception of the Digital Branding of the "Content Governor" on X Social Media using Naïve Bayes

Dwi Febrianti[1*], Martanto[2], Fatihanursari Dikananda[3], Mulyawan[4], Irfan Ali[5]

[1,2,3,4,5] *Teknik Informatika, STMIK IKMI CIREBON*

dwifebriiantii@gmail.com[1*], martantomusijo@gmail.com[2], fatiha.dikananda@gmail.com[3], wm7488748@gmail.com[4], irfanaali0.0@gmail.com[5]

Abstract

The development of social media, particularly the X (Twitter) platform, has created a digital discussion space that influences the formation of public opinion on political issues. The "Content Governor" phenomenon, often associated with Dedi Mulyadi, has generated various responses from users, making systematic sentiment analysis necessary to understand these response patterns. This study aims to map public sentiment tendencies while developing a sentiment-classification model using a Naïve Bayes approach combined with TF–IDF weighting. Data were collected through the Twitter API based on several relevant keywords, yielding 2,133 tweets, which, after a selection process and manual labeling, were narrowed down to 1,023 labeled data points consisting of positive, neutral, and negative sentiments. The research stages include preprocessing (case folding, cleaning, normalization, tokenization, stopword removal, and stemming), data balancing using SMOTE, TF–IDF feature construction, and training a Multinomial Naïve Bayes model with an 80:20 data split ratio. Evaluation was carried out using accuracy, precision, recall, F1-score, and a confusion matrix. The results show that the model achieved an accuracy level of 74.15%, with the strongest performance in the negative sentiment category. The distribution of public sentiment also indicates a dominance of negative sentiment (812 items) compared to neutral and positive sentiments, suggesting that netizens' responses to the "Content Governor" issue tend to be more critical and negative in tone. These findings are expected to enrich studies on digital political communication and provide methodological contributions to sentiment analysis research in the Indonesian language.

Keywords: *sentiment analysis; Naïve Bayes; X (Twitter); public opinion; content governor*

1. Introduction

The development of information and communication technology has transformed how society interacts and expresses their views on public issues. Social media, particularly X (Twitter), has become a primary platform for the public to voice political opinions and shape broader public discourse [1]. In the context of Indonesian politics, the ability to analyze sentiment from millions of tweets generated daily is crucial for understanding public perception. Informatics, through sentiment analysis and machine learning, offers computational approaches capable of extracting valuable insights from this data [2].

The "content governor" issue has triggered significant polarization on social media, but the large volume of data and the informal nature of the Indonesian language make manual analysis ineffective and prone to bias [3]. The real-time dynamics of social media data require fast and accurate analytical methods. Machine learning, particularly Naïve Bayes, provides a solution for automatically classifying sentiment, yet its application in the Indonesian context still requires further investigation.

Previous studies have demonstrated the effectiveness of Naïve Bayes in sentiment analysis, with accuracy levels ranging from 78% to 82%. However, most of these studies are focused on English or Arabic texts and have not addressed the linguistic complexities of the Indonesian language [4]–[6]. This highlights a research gap concerning Indonesian-language sentiment analysis for contemporary political issues.

This study aims to analyze public sentiment toward the "content governor" issue using the Naïve Bayes algorithm to offer an objective depiction of public opinion. It also aims to develop an accurate model for processing Indonesian-language tweets, especially those containing informal language. The findings are expected to contribute to the advancement of Indonesian-language sentiment analysis methodologies and provide valuable insights for policymakers, political analysts, and researchers in understanding real-time public opinion dynamics.

Theoretically, this research contributes to the development of sentiment analysis methodologies that are more aligned with the linguistic and sociocultural characteristics of Indonesian. Practically, the findings of this study may be utilized by the government and policy institutions to formulate more responsive public communication strategies, while political analysts, the technology industry, media, and academics can use this research framework to understand public opinion more accurately, efficiently, and data-driven.

Political communication theory explains how political figures shape identities, manage images, and build relationships with the public through digital media. According to [7], political social media marketing integrates political marketing concepts such as brand equity and stakeholder value into digital platforms to create emotional attachment and public trust [8]. In this context, political branding is viewed not only as a campaign tool but also as a strategic governance device that shapes leadership imagery [9].

In this study, sentiment analysis and NLP theory are used to measure public perception of political figures through social media text. (K. L. Tan et al., 2023) explains that the sentiment analysis process consists of several stages, including text preprocessing (tokenization, stopword removal, stemming/lemmatization), feature representation such as TF-IDF, and model evaluation using metrics such as accuracy, recall, and F1-score to produce contextual and reliable analysis [10].

2. Method

2.1. Research Stages

This research employs a quantitative approach by using the Naïve Bayes algorithm as the primary technique for classifying tweet sentiments related to the "content governor" issue. The study focuses on public responses and comments on social media that reflect how the political figure builds a digital image, representing forms of personal branding and political communication strategies used by a regional leader in the virtual space. The research process begins with data collection through the Twitter API using relevant keywords, followed by a series of preprocessing stages such as text cleaning, case folding, normalization, tokenization, stopword removal, and stemming to prepare the text data for machine learning algorithms. After preprocessing, the text features are transformed into numerical representations using the Term Frequency–Inverse Document Frequency (TF-IDF) technique to enable the model to identify patterns within the tweets. The dataset is then split into training and testing sets with an 80:20 ratio to evaluate the Naïve Bayes model's ability to classify sentiment into positive, negative, and neutral categories. Model performance is assessed using metrics such as accuracy, precision, recall, and F1-score, while a confusion matrix is used to examine the types of prediction errors occurring during classification.
The implementation of the analysis is carried out using the Python programming language, supported by libraries such as scikit-learn for machine learning algorithms, NLTK or Sastrawi for Indonesian text processing, and pandas for dataset manipulation. Visualization results are presented through sentiment distribution charts, word clouds to identify the most dominant words in each category, and time-series analysis to monitor changes in public sentiment over time. Overall, this study relies on the combination of Naïve Bayes and TF-IDF as the foundation of the quantitative analysis. The research flowchart can be seen in Figure 1.

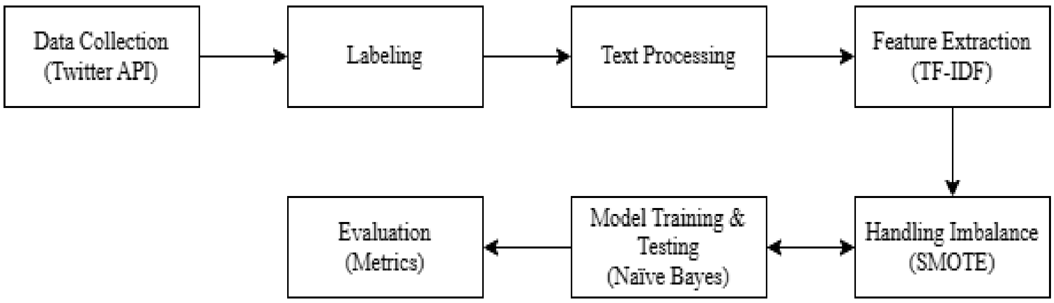


Fig. 1. Research Flowchart

2.2. Dataset

The dataset used in this study consists of 2,133 tweets. The tweets were collected from the X (Twitter) platform with topics related to the "content governor." The labeling process was carried out manually by the researcher by reading each tweet and assigning a sentiment label based on words with positive, negative, or neutral meanings according to the Indonesian Dictionary (KBBI). After the manual labeling process, the number of usable tweets became 1,023. The sentiment labels were classified into three categories: positive, negative, and neutral. There were 84 positive tweets, 812 negative tweets, and 127 neutral tweets. This labeled dataset served as the foundation for training and testing the sentiment analysis model.

2.3. Preprocessing

Data preprocessing is carried out to clean and prepare the data before entering the feature extraction and sentiment analysis stages. The first stage of preprocessing is cleaning and case folding, which involves converting all text to lowercase for consistency and removing certain characters. The second stage is tokenization, which breaks sentences into word units. The third stage is stopword removal to eliminate common words that do not contribute to meaning. Next, the fourth stage is normalization, which includes correcting spelling, expanding slang, and removing punctuation marks. The fifth stage is stemming, which reduces words to their root forms. The output of the preprocessing stages is tweet data that has been transformed from its raw form into a more structured format and is ready to be converted into numerical features using the TF-IDF method.

2.4. Naïve Bayes

In this study, the Naïve Bayes algorithm is used to classify sentiment in comments that have undergone preprocessing. After being cleaned, the comments are transformed into numerical representations using the TF-IDF method, which are then used as input for the classification model. Before training the model, the SMOTE technique is applied to address the imbalance between the positive and negative classes. Through this technique, synthetic data is generated for the class with fewer samples, resulting in a more balanced data distribution. With this strategy, the Naïve Bayes model is expected to more proportionally learn sentiment patterns from both classes and produce more accurate predictions.

2.5. Evaluation

Model performance evaluation is conducted using accuracy and the confusion matrix to assess the model's ability to distinguish sentiment in the test data. Accuracy measures the proportion of correct predictions compared with the total test data, while the confusion matrix provides a more detailed overview of how many predictions are correct or incorrect in each sentiment category. To generate a more comprehensive understanding, the analysis is further strengthened with precision, recall, and F1-score metrics, which are particularly important when evaluating models on datasets with imbalanced class distributions. These three metrics help identify the model's level of precision, sensitivity, and balanced performance in recognizing positive, negative, and neutral sentiments more objectively and measurably, ensuring more accurate evaluation. In addition to numerical evaluation, this research also includes visualizations in the form of word clouds to display the most dominant words appearing in comments labeled as positive, negative, and neutral. Accuracy is defined as the ratio between the number of correctly identified predictions across positive, negative, and neutral categories and the total amount of analyzed data, indicating the model's overall correctness in performing classification. The accuracy value can be obtained using equation (1).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

For multi-class, the formula is:

$$accuracy = \frac{\sum TPi}{Total\ Data}$$

Precision is a measure that assesses the proportion of correct positive predictions from all cases categorized as positive by the model. In other words, this indicator explains the extent to which the system is able to distinguish data that truly belongs in the positive class from the overall predicted data. The precision value can be obtained by applying equation (2).

$$precision = \frac{TP}{TP + FP} \quad (2)$$

For multi-class, calculated per class:

$$precision = \frac{TPi}{TPi + FPi}$$

Recall is the ratio of the number of correct positive predictions to the total data that is factually in the positive class. This metric describes the extent to which the model is able to identify all relevant instances as part of the positive category without missing any. The recall value can be obtained by applying equation (3).

$$recall = \frac{TP}{TP + FN} \quad (3)$$

For multi-class:

$$recall = \frac{TPi}{TPi + FNi}$$

The F1-score is a harmonic combination of precision and recall values that produces a more balanced evaluation measure for assessing the performance of a classification model. The F1-score value can be obtained by applying equation (4).

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

For multi-class, per class:

$$F1 - score = 2 \times \frac{Precision\ i \times Recall\ i}{Precision\ i + Recall\ i}$$

3. Results

3.1. Data Reading

During the data reading stage, the dataset is read and an initial check (data exploration) is performed. The dataset is successfully loaded into the system and displays the initial data structure, which includes the number of samples, class distribution, and the presence of missing or duplicate values. This data reading process ensures that all data is available and ready for processing in the next stage. The total number of samples in the dataset reading is 1023.

3.2. Preprocessing Data

The preprocessing stage includes a series of processes ranging from case folding, text cleaning, normalization, tokenization, stopword removal, to stemming. After all these steps are completed, the raw text is successfully converted into data that is neater, more structured, and ready to be used in the next feature extraction stage. This cleaning process involves removing various irrelevant elements such as punctuation marks, mentions, links, numbers, special characters, and common words that do not contribute significantly to sentiment analysis. Normalization and letter standardization also help to unify spelling variations so that the meaning of words is more consistent. This transformation simplifies sentences and focuses on core keywords such as “governor,” “content,” and “KDM,” thereby improving the accuracy of Naïve Bayes-based sentiment analysis. A visual representation of the preprocessing results is shown in Figure 2 as an illustration of the transformation from raw data to processed data.

HASIL PREPROCESSING (Perbandingan Before-After):

Original Text	After Preprocessing
Tai...gubernur konten KDM aja lu banggain	tai gubernur konten kdm lu banggain
@JendralKepitin9 @BosPurwa @DediMulyadi71 Ah a...	ah aing orang jawa barat sunda teu resep ka gu...
@PresidenKopi @DediMulyadi71 Gini kalau gubern...	gin gubernur mabok konten
Gubernur doyan konten KDM sama dengan yang dul...	gubernur doyan konten kdm masuk gorong
Gubernur konten KDM	gubernur konten kdm
Papa aku hari hari nontonin konten KDM jir sam...	papa nontonin konten kdm jir sampe mewek bjir ...
Sebagai gubernur kan tugasnya membangun daerah...	gubernur tugas bangun daerah youtuber bikin ko...
@sofiesyarief @DediMulyadi71 @farhanpenyiar De...	dedi sibuk bikin konten gubernur
Wkwkwkwkkw emak di rumah kena nih sama konten ...	wkwkwkwkkw emak rumah kena nih konten gubernur...
apakah tidur mu pulas? nyaman dan luas? mimpi ...	tidur mu pulas nyaman luas mimpi mu indah liat...

Fig. 2. Data preprocessing results

3.3. Feature Extraction with TF-IDF

Feature extraction using the TF-IDF (Term Frequency-Inverse Document Frequency) method to convert text data into numerical representations. This method calculates the weight of each word based on its frequency of occurrence in the document and its uniqueness across the entire dataset. The feature extraction results show a TF-IDF matrix that represents each document in the form of a numerical vector, with dimensions that correspond to the number of features extracted. Systematically, the TF-IDF value is obtained by converting the occurrence of a word in a particular document (TF) with a logarithmic value representing the inverse of the number of documents containing that word (IDF). The calculation formula can be obtained from equation (5)

$$TF - IDF(t,d) = TF(t,d) \times IDF(T) \quad (5)$$

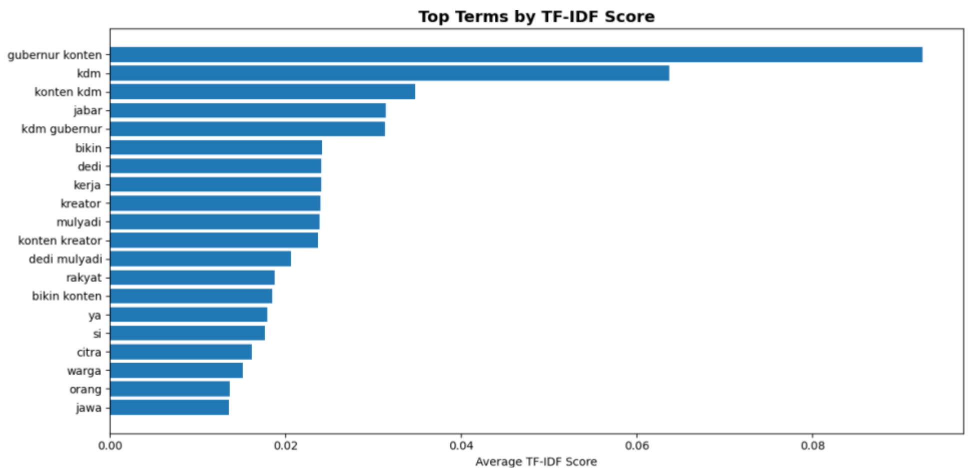


Fig. 3. Feature Extraction Results Using TF-IDF

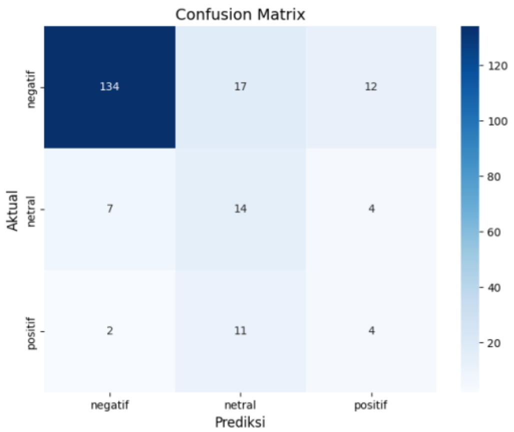
The results of word weight visualization using the TF-IDF (Term Frequency–Inverse Document Frequency) method show the most important and frequently appearing words in the tweet data collection regarding “KDM Content Governor”. Based on the horizontal bar graph, it can be seen that the term “content governor” has the highest TF-IDF score, indicating that this word is the most dominant and most influential in the text corpus. This is followed by the words “kdm”, “kdm content”, and “jabar,” indicating that the focus of X (Twitter) users' conversations largely revolves around the figure of “KDM” and the context of the West Java region. Other words such as “bikin,” “dedi,” “kerja,” and “kreator” also have fairly high scores, indicating the presence of topics related to the creative activities and roles of the figure being discussed. Meanwhile, the words “people,” “image,” and “citizens” reflect public opinion related to the community's view of this figure. Overall, this graph shows that the topic of public conversation on X (Twitter) is highly focused on the issue of digital content and the figure of the “Content Governor” associated with KDM.

3.4. Sentiment Analysis

The sentiment analysis stage involved several key steps. First, manual labeling was performed on 2,133 initial data sets, resulting in 1,023 data sets that were classified as positive, negative, and neutral sentiments. The results showed a predominance of negative comments, followed by neutral comments, and the lowest number of positive comments, reflecting the public's tendency to respond to the issue critically. The results show a predominance of negative comments, followed by neutral comments, and the lowest number of positive comments, eflecting the public's tendency to respond critically to the issue. Second, the labeled data is then divided into training data (80%) and test data (20%) using stratified sampling to maintain a balanced class distribution. This process involves randomizing the data and using random status to ensure consistent results. Third, SMOTE is applied to the training data to address class imbalance by generating synthetic samples for the minority class, ensuring a more balanced dataset for all three sentiment categories without sacrificing the original data. Fourth, the Multinomial Naïve Bayes model is trained using TF-IDF-based text representation. The model learns word occurrence patterns in each sentiment class and generates classification probabilities based on the weights of relevant words. Fifth, the testing phase is conducted using test data to assess the model's ability to predict sentiment in new data, so that the model's performance and generalization can be measured objectively.

3.5. Naïve Bayes Model Evaluation

The Naïve Bayes model was evaluated to assess its performance in classifying public sentiment toward the issue of content governance based on test data. The model evaluation produced a confusion matrix showing that the model was most accurate in recognizing the negative class with 134 correct predictions, while errors still occurred in the neutral and positive classes, which were often interchanged. The evaluation results were also reinforced by the metrics of the Naïve Bayes model performance evaluation. Based on the evaluation metrics, the model obtained an accuracy of 74.15%, a precision value of 80.23%, a recall of 74.15%, and an F1-Score value of 76.53%. Overall, the research results show that the Naïve Bayes model is capable of classifying public sentiment towards the issue of “Content Governor” with fairly good performance, although there are still challenges in detecting non-dominant sentiment due to data imbalance and context similarity between classes. The confusion matrix results and model evaluation metrics can be seen in Figure 4.



* Metrik Kinerja Model:

Metrik	Nilai (%)
Accuracy	74.15
Precision	80.23
Recall	74.15
F1-Score	76.53

Tingkat Akurasi Akhir: 74.15%

Fig. 4. Naïve Bayes Model Evaluation Results

3.6. Visualization

To complement the sentiment analysis, word clouds were also used to identify the words that appeared most frequently in each sentiment category. In positive sentiment, words such as “governor,” “content,” “kdm,” “work,” and “benefits” indicated public appreciation. Negative sentiment is dominated by similar words but accompanied by critical terms such as “image,” “lie,” and “stupid,” which convey dissatisfaction. Meanwhile, neutral sentiment contains descriptive words such as “position,” “program,” and “West Java,” which are informative without emotional connotations. Overall, this visualization provides an overview of the focus of public discussion on the issue of “KDM Content Governor” based on variations in sentiment. The visualization results can be seen in Figure 5.

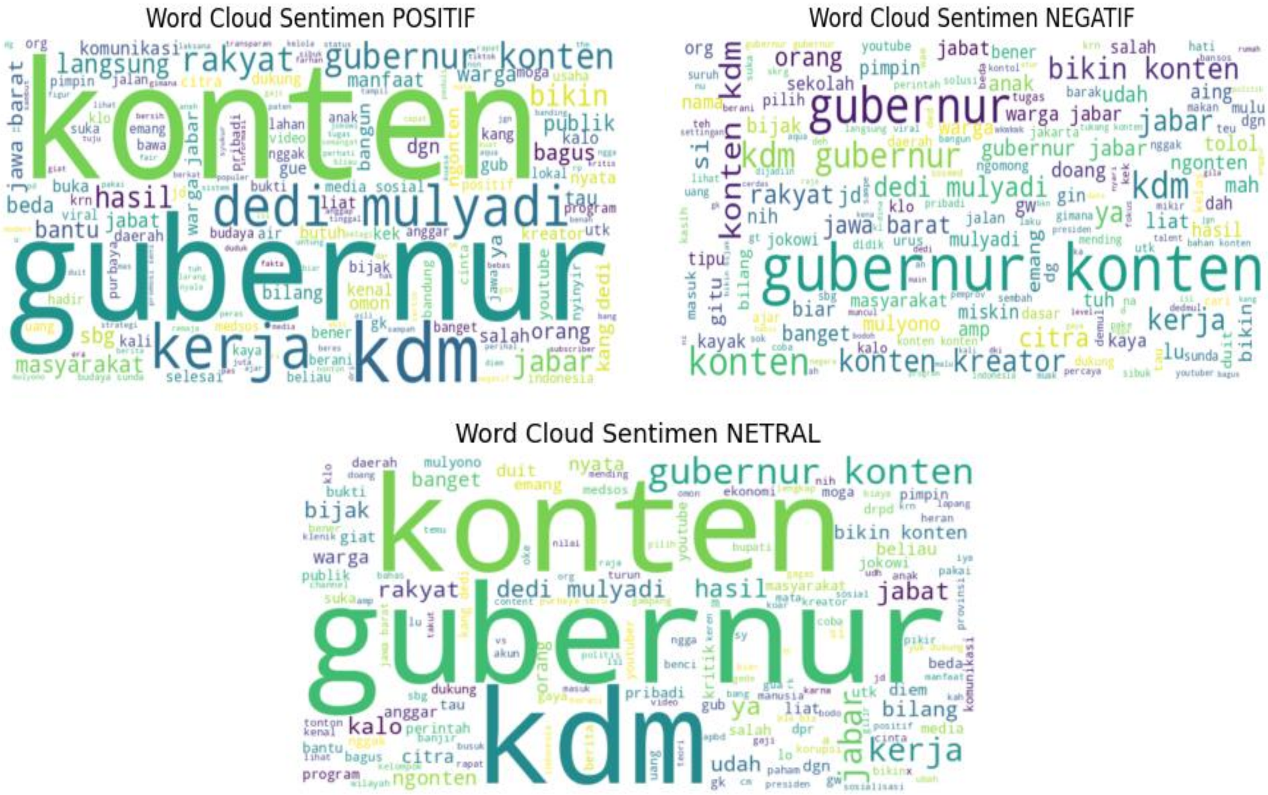


Fig. 5. Word Cloud Sentiment Results

3.7. Discussion

The analysis results show that the Naïve Bayes algorithm is capable of performing sentiment classification tasks with an accuracy rate of 74.15%. When viewed from other evaluation metrics, namely precision, recall, and f1-score, it appears that the model's best performance appears in the negative sentiment category, with precision reaching 0.94, recall of 0.82, and f1-score of 0.88. Conversely, the model's ability to recognize neutral and positive classes is still relatively low, as reflected in the f1-score, which is only around 0.42 for the neutral class and 0.22 for the positive class. This situation indicates that the model is more responsive to the class with the most dominant data, namely the negative category, so that the tendency for bias towards the majority class still occurs and affects the consistency of the classification results. This phenomenon aligns with the findings of (Al-Walid et al., 2025), which state that class imbalance is one of the main factors affecting the performance of classification models, especially when the amount of data in the minority class is much smaller, preventing the model from learning representative patterns. In addition, the TF-IDF text representation method plays an important role in feature vector formation, as it is able to provide proportional weights to words that have distinguishing meanings between documents. This is consistent with the results of a study (Jawad et al., 2024) that compared TF-IDF with hashing methods and found that TF-IDF provides more stable and accurate results in text classification tasks because it considers word frequency at both the document level and the corpus as a whole. However, TF-IDF has limitations in capturing semantic context and word relationships, so the model tends to struggle when encountering sentences that contain sarcasm, ambiguity, or ironic nuances, as often seen in social media texts. The confusion matrix results show that most negative data can be classified correctly, but errors still occur in the neutral and positive classes, which are often confused with the negative class. The macro average and weighted average values in the classification report also show a performance gap between classes, indicating the dominance of results in the majority class. This condition is in line with the view (W. Li et al., 2024) which explains that the Naïve Bayes model has limitations in handling unbalanced data due to its simple probabilistic approach. From a linguistic context, the word cloud results show that the words “governor,” “content,” and “citizens” are the most frequently appearing terms in the dataset, indicating that public discourse tends to focus on certain issues with more negative opinions. This supports the classification results that show the dominance of negative sentiment. In research conducted by (Kumar et al., 2025), it is explained that public perception of social issues in online media often leans toward the negative because users tend to express dissatisfaction or criticism through social media. Overall, the application of the SMOTE method has been proven to help reduce data imbalance at the training stage, but it has not fully improved the performance of the minority class. This is in line with the findings of (Jawad et al., 2024), which emphasize that although balancing techniques can improve data distribution, their effectiveness still depends on the representational ability of the features used.

4. Conclusion and Recommendations

This study successfully implemented a sentiment analysis pipeline for the issue of “Governor of Content,” starting with the data collection process, followed by the preprocessing stage to clean and normalize the text before balancing the data using the SMOTE technique to make the class distribution more proportional. After that, the text features were converted into numerical form through the TF-IDF method, which allows the model to read linguistic patterns more systematically, before finally being trained using the Multinomial Naïve Bayes algorithm. This TF-IDF-based model achieved an accuracy of 74.15%, with the most stable performance in the negative sentiment category, while positive and neutral sentiments showed less optimal results due to the imbalance in the amount of data in each class. Sentiment analysis also confirmed that negative comments dominated, indicating a public tendency to express criticism more frequently than appreciation. This study maintained ethical considerations by only using anonymous public data and evaluating model performance through standard evaluation metrics. In the future, it is highly recommended that research development involve comparison algorithms such as Support Vector Machine (SVM) or Random Forest in order to obtain sentiment models that are more accurate and adaptive to data variations, or deep learning models such as IndoBERT and BiLSTM and use contextual embedding to overcome the limitations of TF-IDF. Data imbalance can be corrected by adding data variation and coverage. Analysis can be expanded by linking types of political content with public responses and using longitudinal or cross-platform approaches to obtain a more complete picture of the dynamics of public opinion.

References

[1] H. Kaur and A. Sharma, "A review on sentiment analysis and emotion detection from text," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, pp. 1–19, 2020, doi: 10.1007/s13278-020-00673-2.
[2] I. Ahmad, M. Alqarni, M. M. Alam, M. A. Alam, and R. A. Khan, "Twitter sentiment analysis: How to hedge your bets in the stock markets," *Expert Syst. Appl.*, vol. 185, p. 115611, 2021, doi: 10.1016/j.eswa.2021.115611.
[3] A. Z. Arifin, I. P. A. Mahendra, and E. M. Yuniarno, "Indonesian sentiment analysis based on deep learning using word embedding," *J. Telecommun. Digit. Econ.*, vol. 7, no. 4, pp. 70–84, 2019, doi: 10.18080/jtde.v7n4.209.
[4] W. Zhang, H. Xu, and W. Wan, "Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis," *Expert Syst. Appl.*, vol. 186, p. 115770, 2021, doi: 10.1016/j.eswa.2021.115770.
[5] F. Rustam, I. Ashraf, A. Mehmood, S. Ullah, and G. S. Choi, "Tweets classification on the base of sentiments for US airline companies," *Entropy*, vol. 21, no. 11, p. 1078, 2019, doi: 10.3390/e21111078.
[6] M. E. Alzahrani, T. H. Aldhyani, and S. N. Alsubari, "Developing an intelligent system with deep learning algorithms for sentiment analysis of E-commerce product reviews," *IEEE Access*, vol. 8, pp. 26938–26952, 2020, doi: 10.1109/ACCESS.2020.2971415.
[7] A. Abid, S. K. Roy, J. Lees-Marshment, B. L. Dey, S. S. Muhammad, and S. Kumar, "Political social media marketing: A systematic literature review and agenda for future research," *Electron. Commer. Res.*, 2023.
[8] S. Effendi Pratama, J. Lucretia, H. Irsyad, A. Rahman, U. Multi Data Palembang, and S. Selatan, "Opini Masyarakat Terhadap Bonus Demografi Pada Kanal Youtube Dengan Metode Tf-Idf, Naïve Bayes Dan Smote," vol. 11, no. 2, pp. 243–253, 2025.
[9] B. R. Bialas, "Political branding: Subterfuge or the new mode of governance?," *J. Public Gov.*, vol. 64, no. 2, 2023.
[10] K. L. Tan, C. P. Lee, and K. M. Lim, "A survey of sentiment analysis: Approaches, datasets, and future research," *Appl. Sci.*, vol. 13, no. 7, p. 4550, 2023.