Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16261 p-ISSN : 2541-044X

Comparison of Naive Bayes and Support Vector Machine for Sentiment Analysis of BPJS Health Service Deactivation

Vitriayanti Payung Allo[1)*], Marlinda Sanglise[2)], Julius Panda Putra Naibaho[3)]
[1,2,3)]University of Papua, Indonesia
[1)]vitriyantipayungallo@gmail.com, [2)] msanglise@unipa.ac.id, [3)]j.naibaho@unipa.ac.id

Submitted :May 21, 2026 | **Accepted** : June 13, 2026 | **Published** : July 5, 2026

Abstract: The deactivation of BPJS Health services has become a topic of public discussion on social media, particularly on platform X, where users frequently express opinions regarding healthcare access and membership status. This study aims to analyze public sentiment toward the deactivation of BPJS Health services and compare the performance of Naive Bayes and Support Vector Machine (SVM) for sentiment classification. The dataset consisted of 8,357 tweets collected from social media X, of which 7,546 tweets were retained after preprocessing, including data cleaning, case folding, tokenizing, stopword removal, and stemming. TF-IDF and FastText were employed as text representation techniques, while model evaluation was conducted using 5-Fold Cross Validation, Grid Search Cross Validation for hyperparameter optimization, and a paired t-test for statistical significance analysis. Classification performance was measured using accuracy, precision, recall, and F1-score metrics. The results showed that negative sentiment dominated public opinion, accounting for 70.63% of the dataset, followed by neutral sentiment (26.48%) and positive sentiment (2.89%). The SVM model with TF-IDF achieved the highest performance, with an accuracy of 80.97%, precision of 80.16%, recall of 80.97%, and F1-score of 79.70%, outperforming Naive Bayes with TF-IDF (79.01%), Naive Bayes with FastText (64.26%), and SVM with FastText (80.31%). Furthermore, a paired t-test confirmed that the performance difference between Naive Bayes and SVM was statistically significant ($p = 0.011$). These findings indicate that SVM combined with TF-IDF is more effective for sentiment classification of high-dimensional social media text data and provide empirical evidence regarding the effectiveness of different text representation approaches for healthcare policy-related sentiment analysis.

Keywords: 5-Fold Cross Validation, BPJS Health, FastText, Naive Bayes, Sentiment Analysis, Social Media X, Support Vector Machine, TF-IDF.

INTRODUCTION

The development of information and communication technology has encouraged people to actively express opinions through social media. Social media has become one of the important data sources for understanding public opinion regarding various issues and government policies. Sentiment analysis is a text mining technique used to identify opinions or emotions in text into positive, negative, or neutral categories. Recent survey studies have shown that sentiment analysis remains one of the most widely used approaches in text mining and social media analytics, with applications ranging from public opinion monitoring to policy evaluation (Li et al., 2022; Mao et al., 2024; Ghatora et al., 2024).

One issue that has attracted considerable public attention is the deactivation of BPJS Health services. This policy has generated various responses from the public, particularly regarding healthcare access, changes in membership status, and the lack of clear information related to service deactivation. Many users on social media X expressed complaints, criticism, and support regarding the policy through posts and comments. The high level of public opinion activity on social media makes platform X a relevant data source for analyzing public perceptions of BPJS Health services.

Various previous studies have been conducted on sentiment analysis using machine learning methods. Commonly used methods include Naive Bayes and Support Vector Machine (SVM) because they provide good

*name of corresponding author

performance in text classification(Iskandar & Nataliani, 2021). Previous research showed that the SVM method performs better than Naive Bayes in Twitter sentiment classification regarding BPJS services (Amandasari & Damayanti, 2022). Recent studies have also demonstrated that machine learning algorithms combined with effective feature extraction techniques can significantly improve sentiment classification performance on social media data (Elgeldawi et al., 2021; Ghatora et al., 2024).

Although previous studies have compared Naive Bayes and Support Vector Machine for sentiment analysis, most studies focused on different domains such as marketplace reviews, public services, educational assistance programs, and corruption-related issues (D. Iskandar & Y. Nataliani, 2021; Kurniawan et al., 2023; Sholekhah & Muntahanah, 2025). Furthermore, many previous studies relied primarily on TF-IDF feature representation and conventional machine learning algorithms without investigating alternative text representation approaches such as word embeddings. In addition, model evaluation was often limited to classification accuracy or simple train-test split strategies, which may not provide a robust assessment of model generalization performance. Therefore, there remains a need to evaluate sentiment classification models using more reliable validation techniques and to examine the effectiveness of different text representation approaches for healthcare policy-related sentiment analysis.

This study investigates public sentiment regarding the deactivation of BPJS Health services on social media X by comparing Naive Bayes and Support Vector Machine classifiers as well as TF-IDF and FastText text representation approaches using 5-Fold Cross Validation. The study aims to evaluate not only the effectiveness of different classification algorithms but also the impact of different text representation techniques on sentiment classification performance. FastText was included because previous studies have demonstrated that semantic-based word embedding approaches can improve text representation by capturing contextual and semantic relationships among words (Bojanowski et al., 2017; Elgeldawi et al., 2021; Ghatora et al., 2024). The preprocessing stages included data cleaning, case folding, tokenizing, stopword removal, and stemming (Han et al., 2012). Model performance was evaluated using accuracy, precision, recall, and F1-score metrics (A. Fauzi & D. Nugroho, 2021; M. Ridwan & F. Rizki, 2022) .

This study contributes by comparing the performance of Naive Bayes and Support Vector Machine classifiers as well as TF-IDF and FastText text representation approaches for sentiment analysis of BPJS Health service deactivation discussions. Unlike previous studies that mainly focused on comparing classification algorithms using TF-IDF features and simple evaluation strategies, this study incorporates FastText word embeddings, applies hyperparameter optimization through Grid Search Cross Validation, and validates the results using statistical significance testing (Elgeldawi et al., 2021; Hastie et al., 2009). Therefore, this study contributes not only to classifier selection but also to the evaluation of TF-IDF and FastText representations, hyperparameter optimization, and statistically validated model performance for healthcare policy-related sentiment analysis.

LITERATURE REVIEW

Many studies on sentiment analysis have been conducted using machine learning methods, particularly Naive Bayes and Support Vector Machine (SVM). (Iskandar & Nataliani, 2021),compared the Naive Bayes, SVM, and k-NN methods in aspect-based sentiment analysis and showed that SVM achieved the best classification performance. (Kurniawan et al., 2023) also demonstrated that the SVM method achieved higher accuracy than Naive Bayes on social media X data. Meanwhile, the Naive Bayes method is still widely used because it has a simple and efficient computational process (Ridwan & Rizki, 2022).

Recent studies have also demonstrated that machine learning algorithms combined with effective feature extraction techniques can significantly improve sentiment classification performance on social media data (Elgeldawi et al., 2021;Ghatora et al., 2024).

Sentiment analysis research on social media X has also been conducted on various public policy issues such as mass layoffs (Saddam, D, & Indra, 2023), KIP-K educational assistance programs, confiscation of corruptors' assets (Sholekhah & Muntahanah, 2025), and BPJS Health services(Amandasari & Damayanti, 2022). Previous studies indicated that the Support Vector Machine method tends to provide better classification performance than Naive Bayes on social media text data(Iskandar & Nataliani, 2021).

In addition, recent studies showed that the use of TF-IDF can improve the quality of text feature representation in sentiment classification, thereby enhancing the performance of machine learning algorithms(Agustina et al., 2024; S. Akuma et al., 2022). Other studies also indicated that Support Vector Machine is highly effective for social media text data with sparse and high-dimensional features (Sholekhah & Muntahanah, 2025; Sebastiani, 2002).

Recent advances in sentiment analysis have introduced alternative text representation approaches, such as word embedding techniques including Word2Vec and FastText, which are capable of capturing semantic relationships between words more effectively than traditional term-frequency-based methods(Mikolov et al., 2013; Bojanowski et al., 2017). Several studies reported that word embeddings can improve sentiment classification performance, particularly for datasets containing contextual and semantically related expressions(E. Elgeldawi et al., 2021; P.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16261 p-ISSN : 2541-044X

S. Ghatora et al., 2024). Nevertheless, TF-IDF remains widely used due to its simplicity, interpretability, and competitive performance in many text classification tasks(S. Akuma et al., 2022). As a result, the comparative evaluation between TF-IDF and semantic-based representations such as FastText remains an important research issue, particularly for short social media texts and healthcare-related discussions.

However, research on sentiment analysis regarding the deactivation of BPJS Health services on social media X remains relatively limited. In addition, previous studies have primarily focused on classification performance using TF-IDF and conventional machine learning algorithms, while comparisons involving alternative text representation approaches and more robust evaluation strategies are still limited(E. Elgeldawi et al., 2021; P. S. Ghatora et al., 2024). Therefore, this study compares the performance of Naive Bayes and Support Vector Machine classifiers under different text representation settings, namely TF-IDF and FastText. In addition, 5-Fold Cross Validation, hyperparameter optimization through Grid Search Cross Validation, and statistical significance testing were employed to provide a more robust and reliable assessment of model performance. This approach enables a comprehensive evaluation of both classifier effectiveness and text representation effectiveness, while also providing a more robust experimental framework through hyperparameter optimization and statistical significance testing.

METHOD

The research stages included data collection, preprocessing, TF-IDF weighting, FastText word embedding generation, sentiment classification using Naive Bayes and Support Vector Machine (SVM), hyperparameter optimization through Grid Search Cross Validation, model evaluation using 5-Fold Cross Validation, and statistical significance testing. The study was designed to evaluate both classifier performance and text representation effectiveness for healthcare policy-related sentiment analysis (Elgeldawi et al., 2021; Ghatora et al., 2024; Hastie et al., 2009).

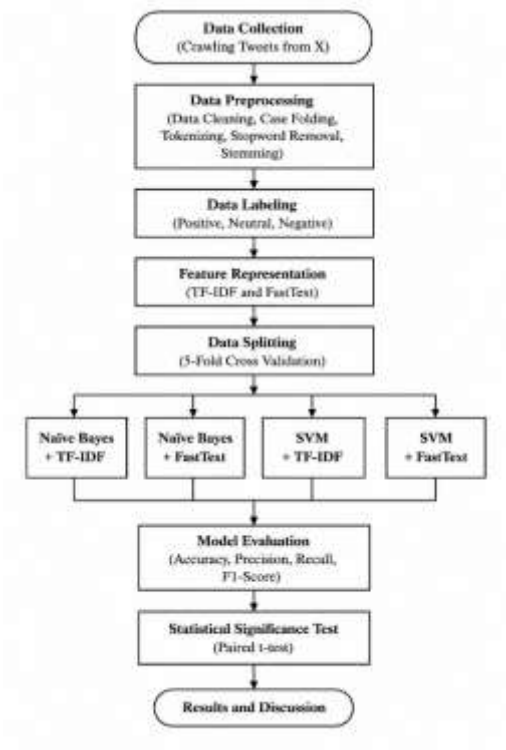

Fig. 1. Research Flowchart

Data Collection

The research data were obtained from social media X (Twitter) using a data crawling technique. Data collection was conducted using several keywords such as “BPJS”, “KIS”, “BPJS nonactive”, and “BPJS deactivation”. The collected data consisted of Indonesian-language tweets related to the research topic. The dataset was then filtered to remove duplicate and irrelevant data. The data processing and classification stages in this study were carried out using the Python programming language through the Google Colab platform with the assistance of Scikit-learn, Pandas, and Sastrawi libraries.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16261 p-ISSN : 2541-044X

Data Preprocessing

Preprocessing is the initial stage in sentiment analysis aimed at cleaning and preparing text data for further processing. This stage is necessary because social media data still contain considerable noise such as symbols, punctuation marks, emojis, and irrelevant words. The preprocessing stages in this study included:

a. Data Cleaning

Data cleaning is the process of removing unnecessary characters such as punctuation marks, numbers, symbols, and URLs. This stage aims to eliminate noise in the data so that the text becomes cleaner and easier to process.

b. Case Folding

Case folding is the process of converting all letters in the text into lowercase letters. This process is performed to standardize the data and avoid differences between uppercase and lowercase letters during text analysis (Han, Kamber, & Pei, 2012).

c. Tokenizing

Tokenizing is the process of splitting sentences into word units or tokens. This stage aims to facilitate the analysis of words in the text(Han, Kamber, & Pei, 2012).

d. Stopword Removal

Stopword removal is the process of eliminating common words that do not contribute significant meaning, such as “and”, “in”, and “the”. This stage aims to improve the relevance of words in the classification process.

e. Stemming

Stemming is the process of converting words into their root forms. For example, the word “helping” becomes “help”. This stage aims to simplify word variations and improve classification accuracy.

Data Labeling

Data labeling was performed by grouping tweets into positive, neutral, and negative sentiment categories based on the opinions expressed by social media X users regarding the deactivation of BPJS Health services. The labeling process was carried out semi-manually using a lexicon-based approach, followed by manual verification to ensure label consistency and sentiment accuracy.

TF-IDF Weighting

In this study, word weighting was conducted using the Term Frequency-Inverse Document Frequency (TF-IDF) method to convert text data into numerical representations based on the importance level of words within documents (Wibowo & Pratama, 2023; Salton & Buckley, 1988). The TF-IDF weighting results were then used as features in the sentiment classification process.

Table 1. TF-IDF Weighting Results

Term	TF-IDF Weight
kis	0.204
sehat	0.135
bpjs	0.121
layanan	0.098
nonaktif	0.067

Table 1 presents examples of terms with the highest TF-IDF weights obtained from the dataset. The results indicate that terms related to BPJS Health services, such as *kis*, *bpjs*, and *nonaktif*, contributed significantly to the sentiment classification process.

FastText Word Embedding

This study employed two text representation approaches, namely TF-IDF and FastText. TF-IDF was used as a traditional frequency-based representation, while FastText was used as a semantic-based representation capable of capturing subword and contextual information (Bojanowski et al., 2017). Previous studies reported that semantic-based word embeddings can improve text classification performance by representing contextual relationships among words more effectively than conventional frequency-based methods (Elgeldawi et al., 2021; Ghatora et al., 2024). The comparison between these two approaches was conducted to evaluate their effectiveness in

*name of corresponding author

representing sentiment information within BPJS Health service deactivation discussions. The FastText model was trained using a vector size of 100, window size of 5, and 20 training epochs. Document vectors were generated by averaging word vectors within each tweet before being used as input for classification.

Data Splitting and Cross Validation

The dataset that had undergone preprocessing, data labeling, TF-IDF weighting, and FastText word embedding generation was evaluated using the 5-Fold Cross Validation method (Kohavi, 1995). In this approach, the dataset was divided into five folds, where four folds were used for training and one fold was used for testing in each iteration. This process was repeated until all folds had been used as testing data. The final evaluation score was obtained by averaging the results of all folds.

Classification Methods

a. Naive Bayes

Naive Bayes is a probabilistic classification algorithm that works based on Bayes' theorem with an independence assumption among features(Han, Kamber, & Pei, 2012). This algorithm is widely used in sentiment analysis because it has a simple and efficient computational process for text classification.

b. Support Vector Machine

Support Vector Machine (SVM) is a classification algorithm that works by finding the optimal hyperplane to separate data classes(Joachims, 1998; Sebastiani, 2002). In this study, SVM used a linear kernel because it is suitable for handling high-dimensional text data generated from TF-IDF weighting. The linear kernel was selected because it can provide more stable classification performance for sparse and high-dimensional text data (Sholekhah & Muntahanah, 2025).

Hyperparameter optimization was conducted using Grid Search Cross Validation to determine the optimal value of parameter C for the Support Vector Machine model. Four values of C (0.1, 1, 10, and 100) were evaluated.(Elgeldawi et al., 2021).

Model Evaluation

Model evaluation was performed to measure the performance of the classification methods using a confusion matrix with accuracy, precision, recall, and F1-score parameters(Hastie, Tibshirani, & Friedman, 2009).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Accuracy was used to determine the overall correctness of the model in classification. Precision was used to measure the correctness of positive predictions generated by the model. Recall was used to measure the model's ability to identify all actual positive instances. F1-score is the harmonic mean of precision and recall used to balance both metrics. In addition to accuracy, the use of precision, recall, and F1-score is important because the dataset has an imbalanced sentiment distribution. Using multiple evaluation metrics helps provide a more objective assessment of model performance under imbalanced data conditions.

RESULT

This section presents the results of sentiment analysis on public opinions regarding the deactivation of BPJS Health services on social media X. The results include dataset distribution, preprocessing outcomes, sentiment distribution, classification performance, model evaluation, and wordcloud visualization.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16261 p-ISSN : 2541-044X

Preprocessing Results

The dataset used in this study consisted of 8,357 tweets collected from social media X. After preprocessing, duplicate removal, and data cleaning, the dataset was reduced to 7,546 tweets. Approximately 811 tweets (9.71%) were excluded because they contained duplicate, noisy, or irrelevant information.The preprocessing stages included data cleaning, case folding, tokenizing, stopword removal, and stemming.

Table 2. Preprocessing Results

Stage	Before	After
Cleaning	BPJS KIS saya dinonaktifkan 😭 https://t.co/abc123	BPJS KIS saya dinonaktifkan
Case Folding	BPJS KIS saya dinonaktifkan	bpjs kis saya dinonaktifkan
Tokenizing	bpjs kis saya dinonaktifkan	[bpjs, kis, saya, dinonaktifkan]
Stopword Removal	[bpjs, kis, saya, dinonaktifkan]	[bpjs, kis, dinonaktifkan]
Stemming	[bpjs, kis,dinonaktifkan]	[bpjs, kis, nonaktif]

Sentiment Distribution

The sentiment distribution results showed that negative sentiment dominated public opinion regarding the deactivation of BPJS Health services with a percentage of 70.63%, while neutral sentiment accounted for 26.48% and positive sentiment only 2.89%.

Table 3. Sentiment Distribution

Sentiment	Total Tweets	Percentage
Negative	5,330	70.63%
Neutral	1,998	26.48%
Positive	218	2.89%

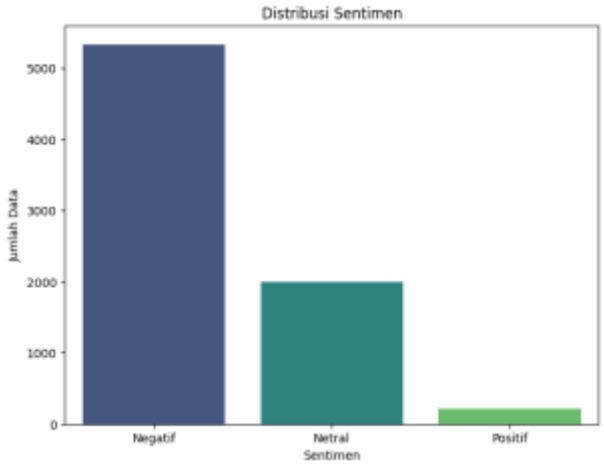

Fig. 2. Sentiment Distribution

This result indicates that most social media users expressed negative opinions toward the BPJS Health deactivation policy.

Classification Results

The classification process was carried out using four experimental settings: Naive Bayes with FastText, Support Vector Machine with TF-IDF, and Support Vector Machine with FastText..

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16261 p-ISSN : 2541-044X

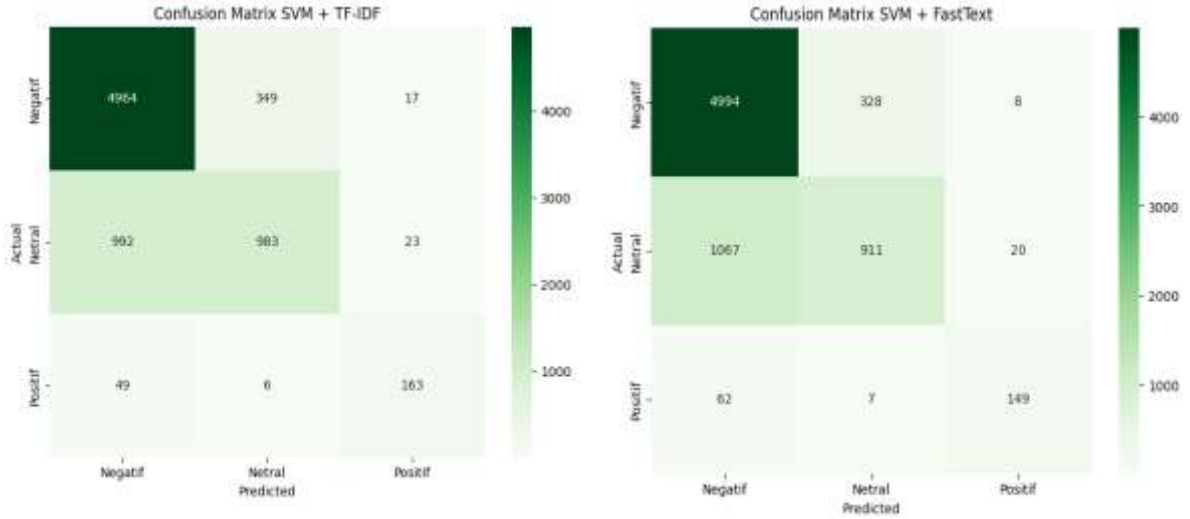

Fig. 3. Confusion Matrix Comparison of SVM Using TF-IDF and FastText

Figure 3 shows the confusion matrices of SVM using TF-IDF and FastText representations. The TF-IDF representation achieved slightly better performance, resulting in fewer classification errors and the highest accuracy of 80.97%.

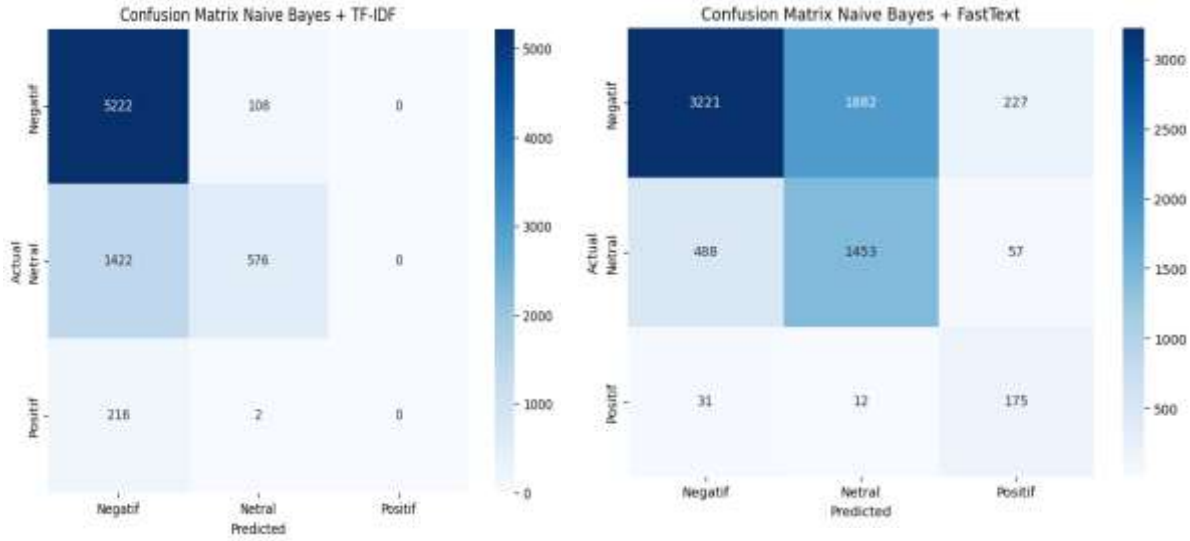

Fig. 4. Confusion Matrix Comparison of Naive Bayes Using TF-IDF and FastText

Figure 4 presents the confusion matrices of Naive Bayes using TF-IDF and FastText representations. A substantial decrease in performance was observed when FastText was used, with accuracy dropping from 79.01% to 64.26%. This result indicates that FastText representations were less compatible with the probabilistic assumptions of Naive Bayes for the BPJS Health sentiment dataset.

Grid Search Cross Validation Results

Grid Search Cross Validation was performed using a Linear SVM with parameter C values of 0.1, 1, 10, and 100 to identify the optimal parameter configuration for sentiment classification.

Table 4. Grid Search Cross Validation Results

C Value	Accuracy (%)
0.1	80.89
1	80.97
10	80.68
100	80.49

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16261 p-ISSN : 2541-044X

Table 4 presents the results of hyperparameter optimization using Grid Search Cross Validation. Among the evaluated parameter values, C = 1 achieved the highest average accuracy of 80.97%, outperforming the other parameter configurations. Therefore, C = 1 was selected as the optimal parameter for the final SVM classification model. The results indicate that proper hyperparameter tuning contributed to improving classification performance and identifying the most suitable model configuration for the BPJS Health sentiment dataset.

Model Evaluation Results

The classification performance varied across the four experimental settings. Both classifier selection and text representation approaches influenced the overall sentiment classification results. Therefore, the performance of Naive Bayes and Support Vector Machine was evaluated using both TF-IDF and FastText representations.

Table 5. Accuracy Results of 5-Fold Cross Validation

Fold	Naïve Bayes + TF-IDF	Naïve Bayes + FastText	SVM+ TF-IDF	SVM + FastText
Fold 1	77.42%	64.65%	79.60%	80.60%
Fold 2	80.25%	64.15%	80.58%	80.25%
Fold 3	78.53%	63.62%	81.44%	80.38%
Fold 4	78.79%	64.28%	81.25%	80.45%
Fold 5	80.05%	64.28%	81.97%	80.52%

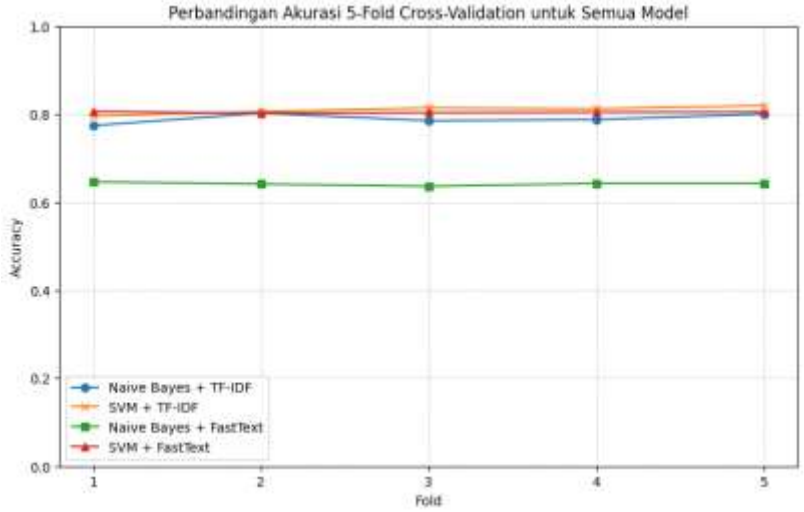

Fig. 5. Accuracy Comparison of 5-Fold Cross Validation

Based on the 5-Fold Cross Validation results, the SVM + TF-IDF model consistently achieved the highest accuracy across all folds, with an average accuracy of 80.97%. The SVM + FastText model achieved a comparable performance with an average accuracy of 80.31%. Meanwhile, Naive Bayes with TF-IDF achieved an average accuracy of 79.01%, whereas Naive Bayes with FastText produced the lowest performance with an average accuracy of 64.26%. The results indicate that classifier selection and text representation approaches both influenced sentiment classification performance. In general, TF-IDF representations produced better results than FastText for both Naive Bayes and Support Vector Machine models. Furthermore, Support Vector Machine consistently outperformed Naive Bayes under both text representation settings.

The comparison also demonstrates that FastText did not necessarily improve classification performance for this dataset. Although FastText is capable of capturing semantic relationships among words, the sentiment information contained in short social media texts appeared to be sufficiently represented by TF-IDF features. Consequently, TF-IDF remained the most effective representation approach for the BPJS Health sentiment dataset.

Table 6. Model Evaluation Results

Method	Accuracy	Precision	Recall	F1-Score
Naïve Bayes + TF-IDF	79.01%	78.81%	79.01%	76.76%
Naïve Bayes + FastText	64.26%	73.45%	64.26%	66.06%

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16261 p-ISSN : 2541-044X

SVM + TF-IDF	80.97%	80.16%	80.97%	79.70%
SVM + FastText	80.31%	79.51%	80.31%	78.66%

Table 6 presents the overall evaluation results of the four experimental settings. The SVM + TF-IDF model achieved the highest performance, with an accuracy of 80.97%, precision of 80.16%, recall of 80.97%, and F1-score of 79.70%. The SVM + FastText model achieved a comparable accuracy of 80.31%, indicating that FastText remained a competitive semantic-based representation approach. Meanwhile, Naïve Bayes + TF-IDF achieved an accuracy of 79.01%, whereas Naïve Bayes + FastText produced the lowest performance with an accuracy of 64.26%.

These results further confirm that Support Vector Machine consistently outperformed Naïve Bayes under both TF-IDF and FastText representations. In addition, TF-IDF provided better classification performance than FastText for both classifiers, suggesting that term-frequency-based representations remained highly effective for sentiment classification of short social media texts related to BPJS Health service deactivation.

Statistical Significance Test

To determine whether the performance difference between the best-performing Naive Bayes model (TF-IDF) and the best-performing Support Vector Machine model (TF-IDF) was statistically significant a paired t-test was conducted using the accuracy scores obtained from the 5-Fold Cross Validation process(Hastie et al., 2009). The test produced a t-statistic value of 4.462 and a p-value of 0.011. Since the p-value was lower than the significance level of 0.05, the null hypothesis (H0), which states that there is no significant difference between the performance of the two classifiers, was rejected. Therefore, the performance difference between Naive Bayes and Support Vector Machine was statistically significant. These results indicate that the superior performance achieved by Support Vector Machine was not caused by random variation in the dataset but reflected a meaningful improvement over Naive Bayes.

Wordcloud Visualization

The wordcloud visualization results indicated that several dominant words frequently appeared in public discussions regarding BPJS Health deactivation, such as "nonaktif", "bpjs", "kis", and "health".

Fig. 6. Wordcloud of Public Sentiment Toward BPJS Health Service Deactivation

The dominant terms in the wordcloud indicate that public discussions mainly focused on BPJS membership status, KIS, healthcare access, and service deactivation issues.

DISCUSSIONS

The research results showed that the Support Vector Machine method provided better classification performance than Naive Bayes in sentiment analysis on social media X data. The SVM model achieved an accuracy of 80.97%, while Naive Bayes achieved an accuracy of 79.01%. This finding is consistent with previous studies which stated that SVM is more effective for high-dimensional and sparse text data(Joachims, 1998; Sholekhah & Muntahanah, 2025).

The higher performance of SVM was influenced by its ability to find an optimal hyperplane in separating sentiment classes. In addition, the use of TF-IDF weighting improved text feature representation, allowing important words in the dataset to contribute more significantly to the classification process(Wibowo & Pratama, 2023; Salton & Buckley, 1988).

Although previous studies have generally reported that Support Vector Machine tends to outperform Naive Bayes in sentiment classification tasks(Iskandar & Nataliani, 2021; Kurniawan et al., 2023), this study provides additional evidence within the context of public healthcare policy, particularly BPJS Health service deactivation.

*name of corresponding author

The findings demonstrate that SVM remains effective in handling sentiment classification on healthcare-related social media discussions characterized by high-dimensional and sparse textual features. Furthermore, the use of 5-Fold Cross Validation and statistical significance testing provides a more reliable evaluation by reducing the potential bias associated with a single train-test split approach and confirming that the observed performance differences were statistically significant (Kohavi, 1995; Hastie et al., 2009). In addition, the comparison between TF-IDF and FastText representations contributes further insight into the effectiveness of different text representation approaches for healthcare-related sentiment analysis.

The comparison between TF-IDF and FastText representations showed that the SVM model with TF-IDF achieved the highest accuracy (80.97%), while the SVM model with FastText achieved an accuracy of 80.31%. Although FastText is capable of capturing semantic relationships between words through distributed vector representations, TF-IDF remained more effective for this dataset. This finding suggests that sentiment information in BPJS Health discussions was sufficiently represented by term-frequency patterns, making TF-IDF a competitive feature representation approach for short social media texts. The comparison results showed that TF-IDF outperformed FastText on the BPJS Health sentiment dataset, indicating that traditional term-frequency-based representations can still provide competitive performance for healthcare-related social media sentiment analysis.

A different trend was observed for the Naïve Bayes classifier. While Naïve Bayes with TF-IDF achieved an accuracy of 79.01%, the use of FastText reduced the accuracy to 64.26%. This result suggests that semantic-based vector representations were less compatible with the probabilistic assumptions of Naïve Bayes in this dataset. Consequently, TF-IDF remained a more suitable representation approach for Naïve Bayes in BPJS Health sentiment classification.

Meanwhile, although Naive Bayes has a simpler and faster computational process, its assumption of feature independence often reduces classification performance on social media text data where contextual relationships among words are strongly interconnected.

The sentiment distribution results indicated that negative sentiment dominated public opinion regarding the deactivation of BPJS Health services, accounting for 70.63% of the total dataset. This result reflects that many social media users expressed complaints, concerns, and criticism regarding healthcare access, membership status changes, and administrative issues related to BPJS Health services.

The highly imbalanced sentiment distribution may also explain the superior performance of Support Vector Machine compared to Naive Bayes. Previous studies have reported that SVM tends to be more robust when handling high-dimensional text data with uneven class distributions, whereas Naive Bayes may be more sensitive to class imbalance due to its probabilistic assumptions(Joachims, 1998; Sholekhah & Muntahanah, 2025). This finding suggests that SVM is a suitable alternative for sentiment classification tasks involving imbalanced social media datasets.

The dominance of negative sentiment suggests that public concern regarding the deactivation policy remains relatively high. Therefore, the findings of this study can provide useful insights for policymakers and related institutions in evaluating and improving healthcare service policies.

This study contributes to the sentiment analysis literature in two ways. First, it provides an empirical comparison between TF-IDF and FastText representations under both Naïve Bayes and Support Vector Machine classifiers within the context of healthcare policy discussions. Second, the study demonstrates that semantic-based representations such as FastText do not always outperform traditional TF-IDF features, particularly for short social media texts. These findings provide additional evidence regarding the relationship between text representation approaches and classifier performance in sentiment analysis tasks.

One limitation of this study is the imbalanced sentiment distribution, where negative sentiment dominated the dataset. This condition may affect the classification performance of minority sentiment classes. Although FastText representation was evaluated, this study did not explore more advanced deep learning approaches such as BERT, IndoBERT, CNN, or LSTM. Future studies are encouraged to investigate these approaches, apply data balancing techniques, and evaluate their effectiveness on healthcare policy-related sentiment datasets.

CONCLUSION

This study successfully compared the performance of the Naive Bayes and Support Vector Machine methods for sentiment analysis of public opinions regarding the deactivation of BPJS Health services on social media X. The research process included data collection, preprocessing, TF-IDF weighting, FastText word embedding generation, sentiment classification, and model evaluation using 5-Fold Cross Validation.

The results showed that negative sentiment dominated public opinion regarding the deactivation of BPJS Health services. Based on the evaluation results, the Support Vector Machine method outperformed Naive Bayes, achieving an accuracy of 80.97%, precision of 80.16%, recall of 80.97%, and F1-score of 79.70%. In comparison, the Naive Bayes method achieved an accuracy of 79.01%.

These findings indicate that the Support Vector Machine method is more suitable for sentiment classification of high-dimensional social media text data. Furthermore, this study contributes empirical evidence regarding the

*name of corresponding author

comparison of Naive Bayes and Support Vector Machine under different text representation approaches, namely TF-IDF and FastText, for healthcare policy-related sentiment analysis. The statistical significance test confirmed that the superior performance of Support Vector Machine was not caused by random variation but represented a meaningful improvement over Naive Bayes. In addition, the comparison results demonstrated that TF-IDF outperformed FastText on the BPJS Health sentiment dataset, indicating that traditional term-frequency-based representations can still provide competitive performance for short social media texts. These findings contribute to the understanding of how classifier selection and text representation approaches influence sentiment classification performance in healthcare-related social media discussions. The findings also demonstrate that social media can serve as a valuable source of information for understanding public responses to healthcare policies and may assist policymakers and related institutions in evaluating and improving BPJS Health service policies. Future studies are encouraged to explore more advanced deep learning approaches such as BERT, IndoBERT, CNN, LSTM, or hybrid models to further improve sentiment classification performance.

REFERENCES

A. Fauzi, & D. Nugroho. (2021). Analisis Sentimen Media Sosial Menggunakan Machine Learning. *Jurnal Sistem Informasi*, *11*(2), 98–107.

A. Wibowo, & R. Pratama. (2023). Implementasi TF-IDF dan Support Vector Machine pada Analisis Sentimen Media Sosial Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *7*(1), 120–128.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics, 5, 135–146.
https://doi.org/10.1162/tacl_a_00051

C. A. Nurhaliza Agustina, R. Novita, Mustakim, & N. E. Rozanda. (2024). The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm. *Procedia Computer Science*, *234*, 156–163.

D. Iskandar, & Y. Nataliani. (2021). Perbandingan Naive Bayes, SVM, dan K-NN untuk Analisis Sentimen Berbasis Aspek. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(6), 1120–1126.

E. Elgeldawi, A. Sayed, A. R. Galal, & A. M. Zaki. (2021). Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. *Informatics*, *8*(4).

F. Z. Tala. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. University of Amsterdam.

Salton, G., & Buckley, C. (1988). Term-Weighting Approaches in Automatic Text Retrieval. Information Processing & Management, 24(5), 513–523.
https://doi.org/10.1016/0306-4573(88)90021-0

Indra, K., Apri Lia, H., Shofa Shofia, H., Agustia, H., Bayu, P., & Aviv Yuniar, R. (2023). Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter. *Jurnal Teknik Informatika Dan Sistem Informasi*, *10*(1). Retrieved from http://jurnal.mdp.ac.id

Jiawei Han, Micheline Kamber, & Jian Pei. (2012). *Data Mining: Concepts and Techniques* (3rd Edition). USA: Morgan Kaufmann.

Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Proceedings of ECML-98, 137–142.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI), 1137–1143.

M. Ridwan, & F. Rizki. (2022). Klasifikasi Sentimen Twitter Bahasa Indonesia Menggunakan TF-IDF dan Support Vector Machine. *Jurnal Informatika*, *8*(3), 210–219.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.P. S. Ghatora, S. E. Hosseini, S. Pervez, M

J. Iqbal, & N. Shaukat. (2024). Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained LLM. *Big Data and Cognitive Computing*, *8*(12), 199.

Q. Li, & et al. (2022). A Survey on Text Classification: From Traditional to Deep Learning. *ACM Transactions on Intelligent Systems and Technology*, *13*(2), 31.

R. Amandasari, & N. Damayanti. (2022). Analisis Sentimen Pelayanan BPJS Kesehatan Menggunakan Metode Support Vector Machine dan Naive Bayes. *Jurnal Media Informatika Budidarma*, *6*(3), 1455–1463.

S. Akuma, T. Lubem, & I. T. Adom. (2022). Comparing Bag of Words and TF-IDF with Different Models for Hate Speech Detection from Live Tweets. *International Journal of Information Technology*, *14*(7), 3629–3635.

saddam, Mohd. A., D, E. K., & Indra. (2023). Analisis Sentimen Fenomena PHK Massal Menggunakan Naive Bayes dan Support Vector Machine, *8*(3).

Sebastiani, F. (2002). *Machine Learning in Automated Text Categorization*. Retrieved from www.ira.uka.de/bibliography/Ai/automated.text.

*name of corresponding author

Sholekhah, A., & Muntahanah, M. (2025). Perbandingan Naïve Bayes dan Support Vector Machine Dalam Analisa Sentimen Tentang Penyitaan Aset Koruptor di Twitter. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *5*(3), 981–989. doi:10.57152/malcom.v5i3.2068

Trevor Hastie, Robert Tibshirani, & Jerome Friedman. (2009). *The Elements of Statistical Learning* (2nd Edition). New York: Springer.

Y. Mao, Q. Liu, & Y. Zhang. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, *36*(4), 102048.

*name of corresponding author