

Public Sentiment Analysis on Police Service Satisfaction Using Twitter Dataset Based on NLP and SVM

Giri Van Transco^{1*}, Asti Herliana²

^{1,2} Informatics Engineering, Adhirajasa Reswara Sanjaya University
girivant047@gmail.com^{1*}

Abstract

The Indonesian National Police plays an important role in maintaining security and providing services to the public. However, there is still public doubt about the quality of its services. This study aims to analyze public sentiment towards police services using Twitter data with a Natural Language Processing (NLP) approach. A total of 14,718 tweets were collected, and after preprocessing, 13,941 tweets were produced that were worthy of analysis. The data was automatically labeled using the Indonesian lexicon method, resulting in 3,737 positive tweets and 6,869 negative tweets. Text representation was carried out using the Term Frequency–Inverse Document Frequency (TF-IDF) method, then classified with the Support Vector Machine (SVM) algorithm using linear, RBF, and polynomial kernels. The Grid Search results showed that the RBF kernel with parameters $C=1000$ and $\gamma=0.1$ gave the best performance with an accuracy, precision, and recall of 91.36%. Model evaluation on training and test data ratios (70:30, 80:20, and 90:10) showed the highest accuracy of 91.83% at the 90:10 ratio. 10-fold cross-validation produced an average accuracy of 92.31%, precision of 92.29%, and recall of 92.31%. These results indicate that SVM with RBF kernel is effective in classifying text-based sentiment in Indonesian.

Keywords: Sentiment Analysis, Twitter, NLP, SVM, RBF Kernel

1. Introduction

Since its establishment in 1945, the Indonesian National Police (Polri) has experienced various changes and challenges in its efforts to provide optimal public services, including problems such as slow case handling and lack of response to public complaints. The police institution plays an important role in maintaining order, enforcing the law, and providing protection and services to the community. In Indonesia, the role and responsibilities of the police are regulated in Law Number 2 of 2002, which states that the Polri is a state apparatus tasked with maintaining public security and order [1]. Public trust in the police is a key factor in the successful implementation of its duties.

In reality, there are still many public reports regarding the poor performance of services at various levels of the police organization, including the Regional Police, Resort Police, and Sector Police, which are often ignored [2]. For example, on October 6, 2021, the hashtag #PercumaLaporPolisi became a trend on Twitter after being uploaded by Project Multatuli (@projectm_org). The hashtag accompanied the news titled "My Three Children Were Raped, I Reported It to the Police, and They Stopped the Investigation," which sparked a wave of netizen testimonies about their negative experiences with the police. As of October 15, 2021, the tweet had been retweeted 10,600 times, received 1,703 quote tweets, and received 18,300 likes [3]. This phenomenon reflects the public's ongoing dissatisfaction with police services, as shown through negative sentiment on social media.

A similar trend reemerged on Twitter through a song titled "Bayar Bayar Bayar" by the band Sukatani, criticizing alleged illegal fees in police services, with lyrics such as: "Want to make a driving license? Pay the police. Get caught on the road? Pay the police." These types of content have sparked widespread public reactions, particularly on Twitter. Therefore, social media especially Twitter can be a valuable data source for analyzing public opinion regarding police services [4].

Twitter, as one of the largest social media platforms, has become a major outlet for people to voice their opinions [5]. It has changed how society expresses its views on public institutions, including the police. With more than 15.7 million active users in Indonesia, Twitter is a potential data source for understanding public perceptions of the police [6]. Positive and negative sentiments are commonly studied through sentiment analysis, a method used to extract and classify opinions from text, aiming to understand public sentiment patterns [7]. In this study, sentiment analysis will be applied to Twitter data to evaluate public satisfaction with police services. To conduct this analysis, the Support Vector Machine (SVM) method is proposed. SVM has demonstrated reliable performance in text classification tasks, helping to determine the weight of each word and identify its contextual meaning [8].

Previous studies have shown that the SVM algorithm achieves high accuracy in sentiment analysis tasks. For example, in a study on sentiments regarding the relocation of Indonesia's capital city, SVM achieved an accuracy of 96.68%, a precision of 95.82, and an AUC of

0.979 [9]. Similarly, in research analyzing sentiment toward the Shopee application, the algorithm reached 98% accuracy with an F1-score of 0.98 [10].

Based on the explanation above, this study is titled "Sentiment Analysis of Public Satisfaction with Police Services Using Twitter Dataset Based on NLP and SVM." This research is expected to contribute significantly to understanding public perceptions of the police and serve as a reference for developing strategies to improve the quality of public services, thereby providing deeper insight into public opinion that can be used to enhance the performance of police institutions in Indonesia.

2. Research Method

2.1. Research Object

The object of this study is a dataset of comments or tweets from the public on Twitter social media discussing police services in Indonesia. The dataset was collected using certain keywords that are relevant to the topic of public services by the police, such as report services, public complaints, SPKT services, and interactions between police officers and the public. The selection of Twitter as a data source is based on its open and real-time nature, and is often used by the public to express opinions directly, including regarding the quality of public services.

The data obtained is unstructured so that it requires natural language processing (Natural Language Processing/NLP) before sentiment classification is carried out. This study focuses on processing and analyzing sentiment from the data using the Support Vector Machine (SVM) method, with the aim of determining the level of public satisfaction with police services. With this approach, the study is expected to provide an overview of public perceptions that are digitally recorded through social media.

2.2. Research Flow

The research flow is structured to describe the systematic stages taken to achieve the research objectives in a structured manner. This flow aims to provide a comprehensive overview of the research process, from problem identification to the preparation of final results. In this study, the flow of activities is presented in the form of a flow diagram that describes each stage in a structured manner and is equipped with detailed explanations for each step. The study entitled "Analysis of Public Sentiment towards Police Service Satisfaction Using Twitter Dataset Based on NLP and SVM" compiles the research flow logically and systematically so that readers can easily understand the process carried out. The stages in this study are illustrated in the framework of thought that can be seen in the image below.

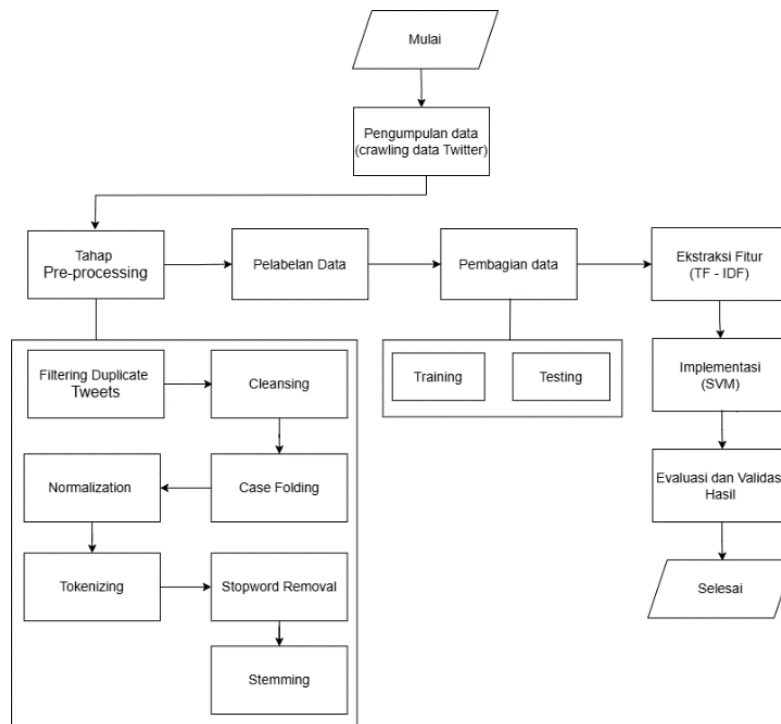


Fig. 1: Framework of thinking

3. Results and Discussion

3.1. Data Collection

Data collection was the initial stage of this research, involving the automated retrieval of tweets containing public opinions about police services using the Twitter API with Python libraries such as Tweepy or Snsrape. Relevant keywords and hashtags, such as "pelayanan polisi," "lapor polisi," and "terimakasih polisi," were used to ensure the collected tweets reflected public perception. A total of 14,718

tweets were gathered between January and April 2025 and stored in CSV format for further processing. All data collected were public and complied with Twitter's data usage and privacy policies. The source code used for crawling was originally developed by Helmi Satria and was modified to suit the specific needs of this study, as illustrated in Figure 2.

```
# Crawl Data

filename = 'pelayanan polisi.csv'
search_keyword = 'pelayanan_polisi since:2025-03-01 until:2025-03-05 lang:id'
limit = 1000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Fig. 2: Data collection process

3.2. Text Preprocessing

After relevant data was collected and key attributes were selected, the next step was preprocessing. This stage aims to clean and prepare raw text data to ensure optimal performance during sentiment analysis. Since data quality directly affects model accuracy, several preprocessing steps were performed, including duplicate removal, text cleaning, stopwords removal, tokenization, and stemming. All processes were implemented in Python using libraries such as re, nltk, and Sastrawi. The following are the preprocessing stages in this study as shown in Figure 3.

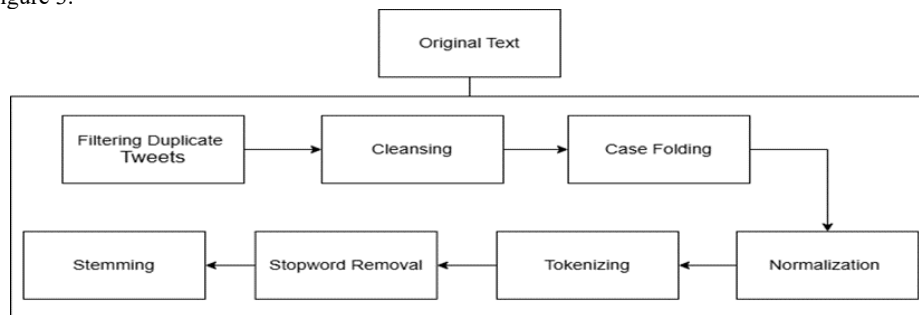


Fig. 3: Preprocessing Stage

1. Filtering Duplicate Tweets

This stage involves removing duplicate entries from the dataset, which commonly occur due to repeated crawling or users posting similar content in a short time. Duplicate tweets can distort sentiment analysis by overweighting specific opinions, leading to biased results. Therefore, duplicates were identified and eliminated to ensure data quality. So the result after the text is duplicated is as in Figures 4.

```
<class 'pandas.core.frame.DataFrame'>
Index: 14311 entries, 0 to 14715
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   id_str       13942 non-null    object
1   tanggal     14309 non-null    object
2   full_text   14310 non-null    object
dtypes: object(3)
memory usage: 447.2+ KB
```

Fig. 4: Duplication Results

2. Cleansing

In this stage, the dataset is cleaned by removing irrelevant elements using Python-based functions. The process includes eliminating URLs, usernames or mentions (starting with "@"), HTML tags, emojis, special characters, and numeric values. These components are considered semantically insignificant and may interfere with sentiment classification.

3. Case Folding

After cleansing, case folding is applied to standardize the text format. All cleaned text is converted to lowercase using the lower() function and stored in a new column called casefolding. This ensures consistency across all words in the dataset before proceeding to the next stage. This step prepares the data for tokenization.

4. Normalization

After case folding, the text proceeds to the normalization stage, where informal or slang words are replaced with their standard forms. This process uses a slang-to-standard dictionary obtained from Kaggle, loaded from a file named kamuskatabaku.xlsx stored in Google Drive. Each word in the case-folded text is checked against the dictionary, and if a match is found, it is replaced with the standard form. This step enhances the consistency of the text before further processing.

5. Tokenizing

After normalization, the next step is tokenization, where each sentence is split into individual words using a simple space-based separator. The resulting list of tokens is stored in a new column called tokenize. This step allows each word to be analyzed separately in subsequent processes such as stopwords removal or stemming.

6. Stopword Removal
After tokenization, irrelevant words that do not contribute significantly to sentiment analysis are removed. Each token is compared against the Indonesian stopwords list from the NLTK library. Common words such as "yang," "dan," and "di" are eliminated, retaining only meaningful words. This step helps improve the focus and accuracy of the analysis.
7. Stemming
After stopword removal, the remaining words are processed through stemming to reduce them to their root forms using the Sastrawi library. For example, words like "pelayanan," "melayani," and "dilayani" are all converted to the root word "layan." This step ensures consistency in word forms for more accurate analysis.

Table 1: Pr Process Results samples

Full text	cleansing	case_folding	Normalisasi	tokenization	stopword_removal	Stemming
@neVerAl0nely Bagaimana kabar polisi yang katanya sudah lapor namun belum ada tindak lanjut dan akhirnya pemilik rental mobil yang bertindak sendiri akhirnya tewas tertembak	Bagaimana kabar polisi yang katanya sudah lapor namun belum ada tindak lanjut dan akhirnya pemilik rental mobil yang bertindak sendiri akhirnya tewas tertembak	bagaimana kabar polisi yang katanya sudah lapor namun belum ada tindak lanjut dan akhirnya pemilik rental mobil yang bertindak sendiri akhirnya tewas tertembak	bagaimana kabar polisi yang katanya sudah lapor namun belum ada tindak lanjut dan akhirnya pemilik rental mobil yang bertindak sendiri akhirnya tewas tertembak	bagaimana,kabar polisi yang katanya sudah lapor namun ada tindak lanjut dan akhirnya pemilik rental mobil yang bertindak sendiri akhirnya tewas tertembak	[kabar,polisi,lapor,tindak,pemilik,rental,mobil,bertindak,tewas,tertembak]	polisi layan gratis sehat masyarakat butuh polisi rutin gelar bagi sembako pelayananpolr iprima abdi jalan wajib terimakasih polisi layan masyarakat

3.3. Labeling

After preprocessing, tweet data were labeled as positive or negative sentiment using an automated Lexicon-Based method. This approach calculates sentiment scores by counting positive words and subtracting negative words in each tweet, enabling efficient and accurate sentiment classification for large datasets. Based on the review in Table 2, there are 4 positive words and 10 negative words identified using the lexicon dictionary. Therefore, the sentiment score is calculated as follows: $Score = 4 - 10 = -6$. The results are as follows:

Table 2: Sentiment Score Calculation Sample

Ulasan	kata negative	kata positif
kabar polisi lapor tindak milik rental mobil bertindak tewas tembak	tindak,lapor,milik,mobil	tindak,tewas,tembak,
Jumlah	4	-10

To support the automatic labeling process, researchers use the Lexicon-Based approach, where the system calculates the number of positive and negative words in each tweet based on an Indonesian opinion dictionary. This dictionary is obtained from the InSet open repository available at <https://github.com/fajri91/InSet>, with a total of 10,606 words, consisting of 6,869 negative words and 3,737 positive words. Each positive word will increase the sentiment score, while negative words will reduce it. The determination of the class label is based on the final score: if the score ≥ 0 then the review is classified as positive, and if < 0 then it is classified as negative. The entire labeling process is implemented using Google Colab, with the help of Python libraries such as pandas and regex for data processing, as well as reading the dictionary directly from the URL. The results of this automatic labeling process are displayed in table form, which shows tweet data that has gone through the stemming process, the bobot_nilai_total, and the final sentiment label.

	steming data	bobot_nilai_total	label
0	polisi layan gratis sehat masyarakat butuh pol...	-5	negatif
1	lapor polisi keluarga gue tipu perkara motorua...	2	positif
2	polisi buruk apaapa bayar alam polisi dompet t...	-16	negatif
3	kabar polisi lapor tindak milik rental mobil t...	-9	negatif
4	lapor polisi	2	positif
...
13936	terimakasih informasi alami lapor wa kapolres	1	positif
13937	tabrak mobil pas renti lampu merah cuman motor...	-5	negatif
13938	assalamu alaikum wr wb selamat siang terimakasih...	9	positif
13939	hasil raih polisi cilik sat lantas polresta ba...	7	positif
13940	innalillahi wsinnalilahi rojuna moga sih adek ...	-1	negatif

13941 rows x 3 columns

Fig. 5: Automatic Labeling Results

After going through the process of weighting and labeling the review data using the Python programming language, the results of the comparison of the number of data from the sentiment class labeling were obtained as shown in Figure IV.6.

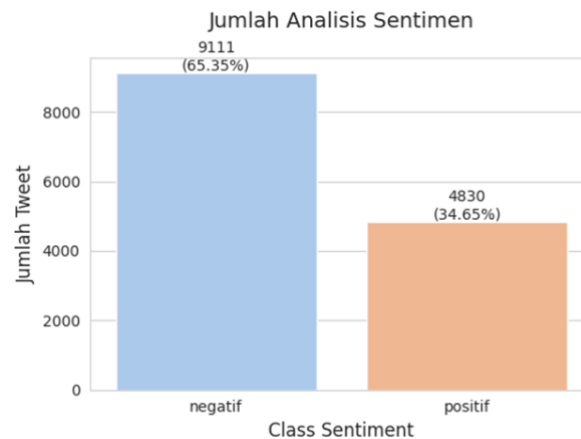


Fig. 6: Comparison Results

3.4. Split Data

The dataset division process is the initial stage in implementing classification. The dataset is divided into two groups, namely training data which functions to train the model and testing data which is used to measure model performance against previously unknown data. This division aims to ensure that model evaluation is carried out objectively and accurately. At this stage, researchers divide the data using the `train_test_split` function from the scikit-learn library. In this study, three comparison scenarios were used between training data and testing data. Researchers implemented three data split ratio scenarios based on a total of 13,941 tweets:

Table 3: Training and Testing Data Sharing

Rasio Data	Data type	Number of Data
70:30	Training	9.759
	Testing	4.182
80:20	Training	11.153
	Testing	2.788
90:10	Training	12.547
	Testing	1.394

This ratio refers to the Pareto principle [11] to ensure that model evaluation is carried out fairly on previously unknown data.

3.5. Text Vectorization with TF-IDF

After the data is divided, the next process is to convert the text data into a numeric representation using the TF-IDF (Term Frequency-Inverse Document Frequency) method. This method is used to measure how important a word is in a document relative to the entire corpus. TF-IDF helps in reducing the influence of words that appear frequently but have less information weight in the context of classification. This transformation is done using the `TfidfVectorizer` library from Scikit-learn. The results of the TF-IDF values can be seen in the table below.

Table 4: TF-IDF Values of Text Vectorization Results

Index	Fitur	TF-IDF
10144	Polisi	0.06939669437858623
6962	Lapor	0.07821304436543648
4726	Hilang	0.2206431408092746
13600	Untung	0.39800185374160413

Based on Table 4, it can be seen that each word that has gone through the tokenization and stemming process is represented by a certain TF-IDF weight value. This value indicates the level of importance of the word in the document. For example, the word "polisi" has a lower TF-IDF value because it tends to appear more frequently in various documents, while words such as "untung" have a higher weight because their existence is less frequent but significant in the context of a particular document. These values are then used as numerical input in the classification process to help the algorithm recognize relevant patterns between documents.

3.6. Classification with Support Vector Machine

After the data is represented with TF-IDF, the classification process is carried out using the Support Vector Machine (SVM) algorithm through the SVC library from Scikit-learn. SVM is effective for high-dimensional data such as text because it is able to form an optimal hyperplane between classes. The model is drilled with training data and tested on testing data to produce predictions. The results of the model performance evaluation are presented in Table 5.

Table 5: SVM Model Performance Classification Results

Metric	Value
Accuracy	89.96 %
Precision	90.06%
Recall	89.96 %

The SVM model achieved an accuracy of 89.96%, indicating a high classification success rate. A precision of 90.06% indicates the accuracy of sentiment prediction, and a recall of 89.96% indicates the model's ability to correctly recognize most of the data. This evaluation is supported by the confusion matrix in Figure 7.

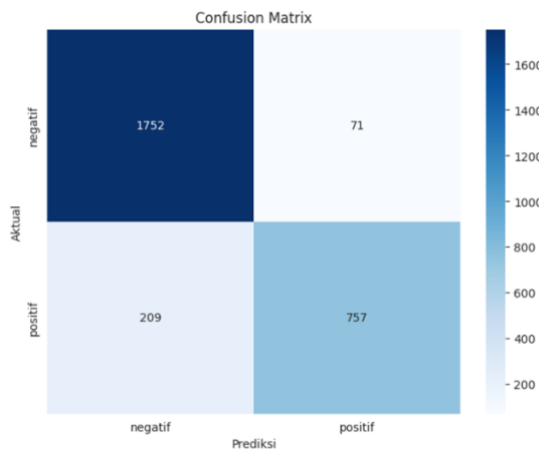


Fig. 7: Confusion Matrix Model SVM

From the confusion matrix, it can be seen:

1. True Negative (TN): 1752 → negative data that is predicted correctly
2. False Positive (FP): 71 → negative data that is wrongly predicted as positive
3. False Negative (FN): 209 → positive data that is wrongly predicted as negative
4. True Positive (TP): 757 → positive data that is predicted correctly

This model is proven to have good performance, especially in detecting negative sentiment. However, there is a slight decrease in recall in the positive class, which means that there are some positive data that are still wrongly classified as negative.

3.6.1. SVM Kernel Implementation and Comparison

To improve performance, three types of SVM kernels were tested, namely linear, RBF, and polynomial, each with a parameter value of C adjusted to its characteristics:

1. Linear: C = 1
2. RBF: C = 100
3. Polynomial: C = 10

The model was trained on training data (80%) and tested on test data (20%) which were separated stratified to maintain label balance. The results of the number of correct predictions from each kernel are shown in Table 6.

Table 6: Number of Correct Predictions Per Kernel

Kernel	Parameter C	Number of Correct Predictions
Linear	1	2535
RBF	100	2539
Polynomial	10	2269

The results show that the RBF kernel provides the best prediction results with 2,539 correct predictions. To improve model performance, Grid Search is performed to find the best parameters of each kernel, namely a combination of C, gamma, and degree. This process aims to optimize the accuracy, precision, and recall of each kernel.

Table 7: Grid Search Results and Evaluation of Each SVM Kernel

Kernel	Best parameters	Best Score	Accuracy %	Precision %	Recall %
Linear	C = 10	0.9071	90.96	91.03	90.96

RBF	C = 1000, gamma = 0.1	0.9101	91.36	91.36	91.36
Polynomial	C = 10, degree = 1	0.9071	91.03	91.03	90.96

The RBF kernel showed the best performance in sentiment classification with accuracy, precision, and recall of 91.36% each, so it is recommended as the main model. After the best parameters were obtained through Grid Search, model performance validation was carried out with three variations of training and testing data ratios (70:30, 80:20, and 90:10), while still using the best parameters of the RBF kernel (C = 1000, gamma = 0.1), to test the stability and consistency of the model against changes in data proportions.

Table 8: Evaluation Results of SVM RBF Model with Various Data Ratios

Training Data Ratio: Test	Accuracy %	Precision %	Recall %
70 : 30	90.65	90.63	90.65
80 : 20	91.36	91.36	91.36
90 : 10	91.83	91.89	91.83

At a training and test data ratio of 90:10, the SVM model with RBF kernel produces a final accuracy of 91.83%, proving its performance consistency and suitability as an optimal sentiment classification model.

3.7. Evaluation and Validation of Results

To test the stability of the model performance as a whole, a 10-fold cross-validation was performed using the SVM model with RBF kernel and optimal parameters (C=1000, gamma=0.1). All data have been represented in the form of TF-IDF vectors. The results of the evaluation of accuracy, precision, and recall for each fold are shown in Table 9.

Table 9: Evaluation Results of 10-Fold Cross-Validation of SVM Model with RBF Kernel

Fold	Akurasi (%)	Presisi (%)	Recall (%)
0	91,83	91,85	91,83
1	91,25	91,23	91,25
2	92,61	92,58	92,61
3	91,18	91,15	91,18
4	92,83	92,79	92,83
5	93,11	93,1	93,11
6	92,18	92,18	92,18
7	91,46	91,43	91,46
8	94,05	94,03	94,05
9	92,61	92,59	92,61
Average	92,31	92,29	92,31

The Support Vector Machine (SVM) model with RBF kernel shows stable and consistent performance based on 10-fold cross-validation. The average evaluation metrics are Accuracy: 92.31%, Precision: 92.29%, Recall: 92.31%. These results indicate the model's ability to recognize patterns and classify sentiments accurately and with low errors. The difference between folds is small, indicating that the model is robust to data variations and has good generalization capabilities. With optimal parameters (C = 1000, gamma = 0.1). Thus, the RBF kernel SVM is proven to be effective in sentiment classification in this study.

4. Conclusions and Suggestions

4.1. Conclusions

This study shows that sentiment analysis of police service satisfaction based on Twitter data can be carried out effectively with the Natural Language Processing (NLP) approach and the Support Vector Machine (SVM) algorithm. The data collection process successfully obtained 14,718 tweets, which were then processed through a text preprocessing stage to produce clean data that is suitable for classification. Automatic sentiment labeling using an Indonesian lexicon-based approach was able to classify tweets into two main categories, namely positive and negative. Data representation using the TF-IDF method has been proven to be able to increase the effectiveness of classification by highlighting words that are relevant to sentiment. The SVM classification model with RBF kernel and optimal parameters (C = 1000, gamma = 0.1) gave the best performance, with an accuracy of 91.83% on a 90:10 data division and an average accuracy of 92.31% based on 10-fold cross-validation. These results indicate that the model used has high generalization capabilities and can be relied on to analyze public sentiment towards police services.

4.2. Suggestions

Based on the results of the research that has been conducted, it is recommended that the development of a sentiment analysis system in the future include expanding data sources to other platforms such as Instagram, Facebook, or online news portals in order to obtain a more comprehensive coverage of opinions. In addition, exploration of other classification algorithms such as Random Forest, Naive Bayes, or Deep Learning-based approaches such as LSTM and BERT needs to be done to compare performance with the SVM method used in this study. Implementation of the system in the form of a web-based or mobile application is also recommended to increase the accessibility and usefulness of the analysis results for related agencies and the wider community.

References

- [1] N. K. S. Adnyani, "Kewenangan diskresi kepolisian Republik Indonesia dalam penegakan hukum pidana," *Jurnal Ilmiah Ilmu Sosial*, vol. 7, no. 2, pp. 135–144, 2021.
- [2] H. Lahaling, K. Makkulawuzar, R. Rahmawati, R. Nur, D. Darmawati, and N. Insani, "Perspektif Masyarakat Terhadap Perilaku Oknum Polisi Mempengaruhi Kualitas Pelayanan Publik Di Wilayah Kepolisian Daerah Gorontalo," *Al Daulah: Jurnal Hukum Pidana dan Ketatanegaraan*, vol. 12, no. 1, pp. 78–89, 2023.
- [3] U. R. Handaningtias, P. A. Praceka, and I. A. Indriyany, "Kepercayaan Publik (Public Trust) Terhadap Polisi: Studi Mengenai Wacana Public Dalam# Percumalaporpolisi Dengan Pendekatan Big Data Analysis," *International Journal of Demos*, vol. 4, no. 3, pp. 940–953, 2022.
- [4] R. M. Nusantara, "Analisis Sentimen Masyarakat terhadap Pelayanan Bank Central Asia: Text Mining Cuitan Satpam BCA pada Twitter," *Co-Value Jurnal Ekonomi Koperasi dan kewirausahaan*, vol. 15, no. 9, 2025.
- [5] A. R. Padri, A. Asro, and C. Chairuddin, "Klasifikasi Kemacetan Lalu Lintas di Indonesia Menggunakan Metode Naive Bayes Classification," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 14, no. 2, pp. 297–310, 2023.
- [6] R. Darman, "Analisis Sentimen Respons Twitter terhadap Persyaratan Badan Penyelenggara Jaminan Sosial (BPJS) di Kantor Pertanahan," *Widya Bhumi*, vol. 3, no. 2, pp. 113–136, 2023.
- [7] E. R. Lidinillah, T. Rohana, and A. R. Juwita, "Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine," *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 154–164, 2023.
- [8] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021.
- [9] P. Arsi and R. Waluyo, "Analisis sentimen wacana pemindahan ibu kota Indonesia menggunakan algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput*, vol. 8, no. 1, p. 147, 2021.
- [10] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 1, pp. 32–35, 2023.
- [11] A. S. Biyantoro and B. Prasetyo, "Penerapan Decision Tree untuk Klasifikasi Status Kesehatan dengan perbandingan KNN dan Naive Bayes: Application of Decision Tree for Health Status Classification, Compared to KNN and Naive Bayes," *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, vol. 4, no. 1, pp. 47–55, 2024.