Proceeding of International Joint Conference on UNESA
Homepage : https://proceeding.unesa.ac.id/index.php/pijcu
e-ISSN : 3032-3762

PIJCU, Vol. 3, No. 1, Desember 2025
Page 1-14
© 2025 PIJCU :
Proceeding of International Joint Conference on UNESA

Sustainable Vocational Innovation in E-Government: Development of a Sentiment Analysis For Evaluating SIREKAP Mobile Performance Using Naive Bayes

Dwi Ramadhaniasari[1*], Andi Iwan Nurhidayat[2], Dimas Novian Aditia Syahputra[3], Ari Kurninawan[4], Moch Deny Pratama[5], Faris Abdi El Hakim[6]
[1*] Universitas Negeri Surabaya, Surabaya, Indonesia

Keywords:
Sentiment Analysis
SIREKAP Mobile
Regional Election
Naïve Bayes
Performance Evaluation

ABSTRACT

The Regional Head Elections (Pilkada) represent a crucial moment in the democratic process that demands transparency and accuracy in vote counting. The SIREKAP Mobile application was designed to support efficient and transparent vote recapitulation in Indonesia. However, during the 2024 elections, the application attracted public criticism due to technical issues and its perceived impact on public trust. This study aims to analyze user sentiment toward the SIREKAP Mobile application during the 2024 Tulungagung Regency elections using the Naïve Bayes algorithm. User reviews were collected from the Google Play Store, TikTok comments, and online surveys, and classified into three sentiment categories: positive, neutral, and negative. Preprocessing steps included tokenization, stopword removal, and stemming, followed by sentiment classification using a semi-supervised approach combining Naïve Bayes with Expectation Maximization (EM). Evaluation results showed strong performance with 83.5% accuracy, 83.6% precision, 83.5% recall, and an F1-score of 83.5%. Sentiment classification revealed that 69.8% of feedback was negative, 15.5% positive, and 14.7% neutral. This study provides insights into public perception of electoral technology and contributes to improving the transparency and accountability of future elections through enhanced application design and implementation.

INTRODUCTION

Elections are a fundamental aspect of democratic governance, providing citizens the opportunity to choose their leaders. In Indonesia, technological innovation has been leveraged to improve transparency, efficiency, and accuracy in elections. One such innovation is the SIREKAP (Recapitulation Information System) Mobile application, designed by the General Elections Commission (KPU) to support digital vote recapitulation. However, during the 2024 Presidential and Regional Elections, SIREKAP faced widespread criticism due to technical issues, connectivity problems, and a lack of trust from the public (Ismail, Asih Widiarti, Dani Muhadiansyah, 2024). Previous studies have reported similar concerns. (Herjanto & Carudin, 2024; Tinanto et al., 2024; Zahra et al., 2024) observed that the majority of user reviews for SIREKAP were negative, indicating general dissatisfaction. (Bahtiar, 2024) also noted that despite its intended role in improving election transparency, technical challenges and limited user readiness remain significant obstacles. These concerns were similarly evident during the 2020 elections, when connection failures prevented data submission from many polling stations. Sentiment analysis has emerged as a valuable approach to systematically assess public perceptions of government technology (Harrison & Johnson, 2019; Serrano-Guerrero et al., 2015). Among various techniques, the Naïve Bayes algorithm is widely used due to its simplicity and high classification performance(Hanafi & R, 2024; Ma, 2025; Sano et al., 2023). Other comparative research has also demonstrated the superior performance of naïve bayes over Support Vector Machine (SVM) and Decision Tree algorithms in certain sentiment analysis tasks, particularly in terms of accuracy and precision(Gifari et al., 2022; Rahayu et al., 2022; Wahid & Saputri, 2022). Despite the growing body of research on sentiment analysis and electoral information systems, previous studies have primarily focused on evaluating classification accuracy without investigating the evolution of user sentiment in specific regional contexts. This study addresses this gap by analyzing sentiment trends toward SIREKAP over time, specifically within Tulungagung Regency, one of the regions that utilized the system during the 2024 Regional Head Elections. Emphasizing this local context is expected to provide deeper insights into user perceptions and system performance, leading to practical and targeted recommendations for improvement. The objective of this research is twofold: (1) to examine how user sentiment toward SIREKAP has evolved over time in Tulungagung Regency, and (2) to develop a web-based sentiment analysis system to support this evaluation. By identifying sentiment patterns, this study aims to provide actionable input for the KPU and system developers, thereby enhancing the credibility, effectiveness, and sustainability of SIREKAP in supporting more transparent and accountable electoral processes. This research also aligns with the principles of sustainable vocational innovation by fostering the development of digital competencies among election officers and supporting the long-term sustainability of e-government services. By providing an adaptable sentiment analysis framework, the study contributes to continuous improvement in public digital services, ensuring that technological innovations in the electoral process remain efficient, transparent, and accessible for future use.

RESEARCH METHOD

1. Research Design

This research adopts a quantitative approach using text classification techniques in sentiment analysis. The study focuses on evaluating user sentiment towards the SIREKAP Mobile application during the 2024 Regional Elections in Tulungagung Regency. The sentiment analysis system was developed using the Naïve Bayes algorithm combined with the Expectation Maximization (EM) method to handle both labeled and unlabeled data.

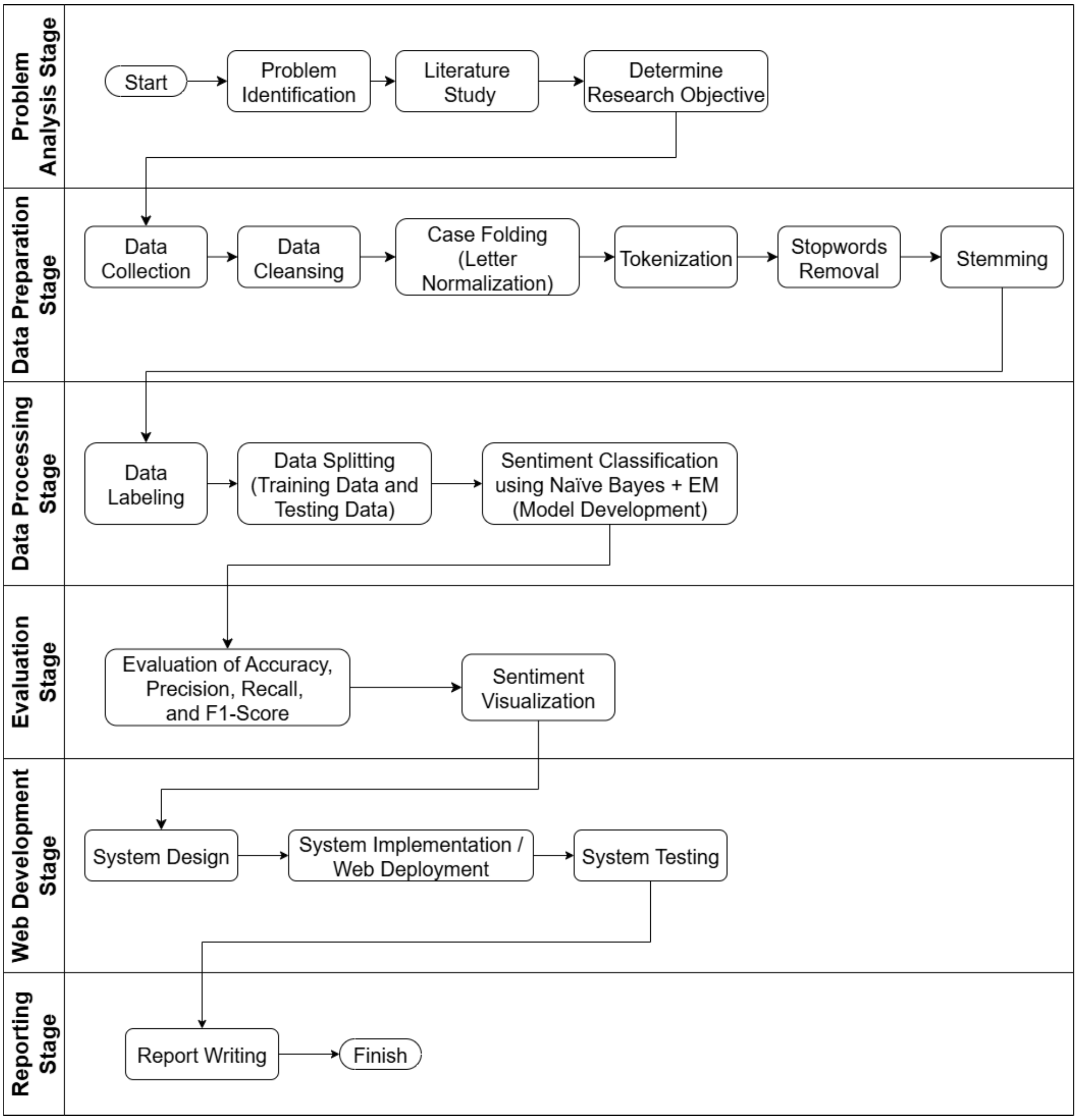


Figure 1. *Research flow of sentiment analysis system for SIREKAP mobile*

2. Data Collection

The dataset for this research was collected from three sources:
Google Play Store: User reviews were collected through web scraping of the official SIREKAP Mobile page. The extracted data included comments, ratings, and dates. User Surveys (Google Forms): Additional data was obtained via an online survey distributed to election officers (PPK, PPS, KPPS) in Tulungagung Regency, focusing on their experiences using the application. TikTok Comments: Comments related to the use of SIREKAP during the 2024 election process were manually collected from relevant TikTok videos. The data collection period spanned from November 2024 to April 2025.

3. Data Preprocessing

After the data collection stage, the next step was data preprocessing, which was conducted to clean and organize the raw text before classification. The process began with case folding to convert all characters into lowercase, ensuring uniformity. Data cleansing removed irrelevant elements such as numbers, special characters, duplicate entries, usernames, timestamps, emojis, and URLs. Normalization standardized informal words and slang using an Indonesian normalization dictionary (e.g., "thx" → "thanks"). The text was then segmented through tokenization and refined via stopword removal to eliminate low-value terms such as "and" and "which". Finally, stemming reduced words to their root form using the Sastrawi library, grouping morphological variants into a single representation. These steps ensured the dataset was consistent, noise-free, and ready for sentiment analysis.

4. Sentiment Classification

The sentiment classification process utilized a Naïve Bayes classifier combined with the Expectation Maximization (EM) algorithm to leverage both labeled and unlabeled data (Bai et al., 2024; Ummah, 2019; van Engelen & Hoos, 2020). Initially, Google Play Store and survey reviews were labeled based on rating scores and then manually validated to ensure accuracy. This validated dataset was used to train the initial Naïve Bayes model. TikTok comments, which lacked rating scores, were classified using EM, which learned from the validated labeled dataset. The final model integrated both the validated labeled data and EM-labeled TikTok data for comprehensive sentiment classification into positive, neutral, and negative categories.

5. Data Analysis

Following the labeling process, the data underwent a feature extraction phase using Term Frequency-Inverse Document Frequency (TF-IDF) (Bounabi et al., 2021). This method is widely used in text mining and natural language processing to convert text data into numerical representations that reflect the importance of words within the documents. The dataset was split 80:20 for training and testing. The Naïve Bayes model was trained, then iteratively improved with EM by incorporating pseudo-labeled data until performance stabilized. Final sentiment distribution provided insights and recommendations for improving the SIREKAP Mobile application.

6. Evaluation Metrics

The performance of the classification model was evaluated using four standard metrics in machine learning: Accuracy, Precision, Recall, and F1-Score.

Accuracy measures the proportion of correctly classified data points over the total data points and is calculated as:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{1}$$

Precision measures the proportion of true positive predictions among all predicted positive cases:

$$Precision = \frac{TN}{TP+FP} \tag{2}$$

Recall assesses the model's ability to correctly identify all relevant positive instances:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

F1-Score is the harmonic mean of precision and recall, providing a balanced evaluation of both metrics:

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

7. System Design Overview

7.1 System Architecture

The system, illustrated in Figure 2, is a web-based sentiment analysis application built using the Flask framework. It provides data management, preprocessing, classification, and visualization functionalities for analyzing user reviews of the SIREKAP Mobile application.

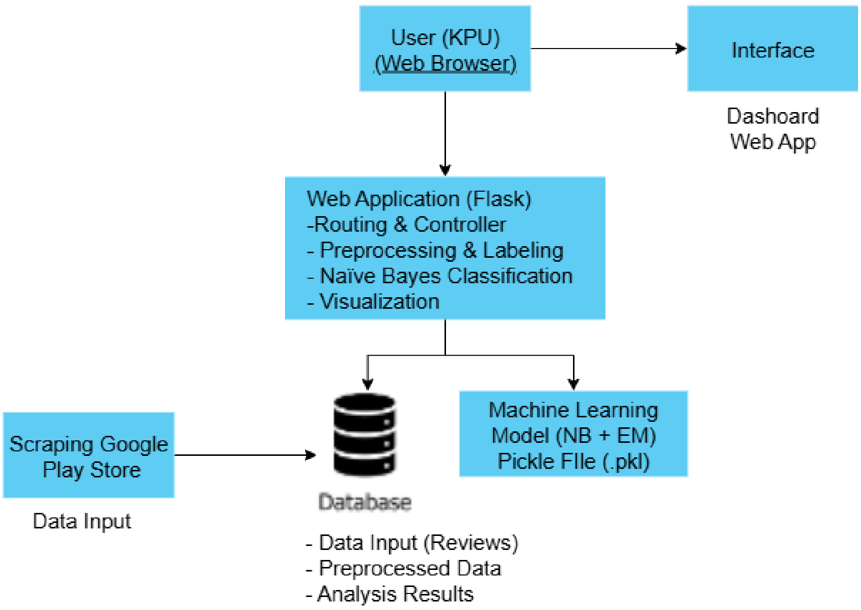


Figure 2. *System architecture*

The architecture comprises:

1 User Interface (KPU / Web Browser) – Authorized users access the system via a dashboard to view sentiment analysis results and manage datasets.
2 Web Application (Flask) – Handles routing, preprocessing, labeling, classification with Naïve Bayes + Expectation Maximization (EM), and visualization.
3 Database – Stores raw reviews, preprocessed data, and sentiment analysis results, ensuring consistency and traceability.
4 Machine Learning Model (NB + EM) – Implemented as a pickle (.pkl) file, used for classifying sentiments.
5 Data Source – User reviews scraped from the Google Play Store as the primary input.
6 Dashboard Interface – Displays analysis results interactively for decision-making

7.2 System Flowchart

The system flowchart illustrated in Figure 3 describes the complete workflow of the sentiment analysis system for processing user reviews of the SIREKAP Mobile application. This flowchart demonstrates the sequential processes starting from user access to the final result generation within the system. The process begins when the user accesses the system through the web application. Upon accessing, the user is required to log in using valid credentials. If the login attempt fails due to invalid credentials, the user will be redirected back to the login interface. Once authenticated, the user is directed to the main dashboard or home menu of the application.

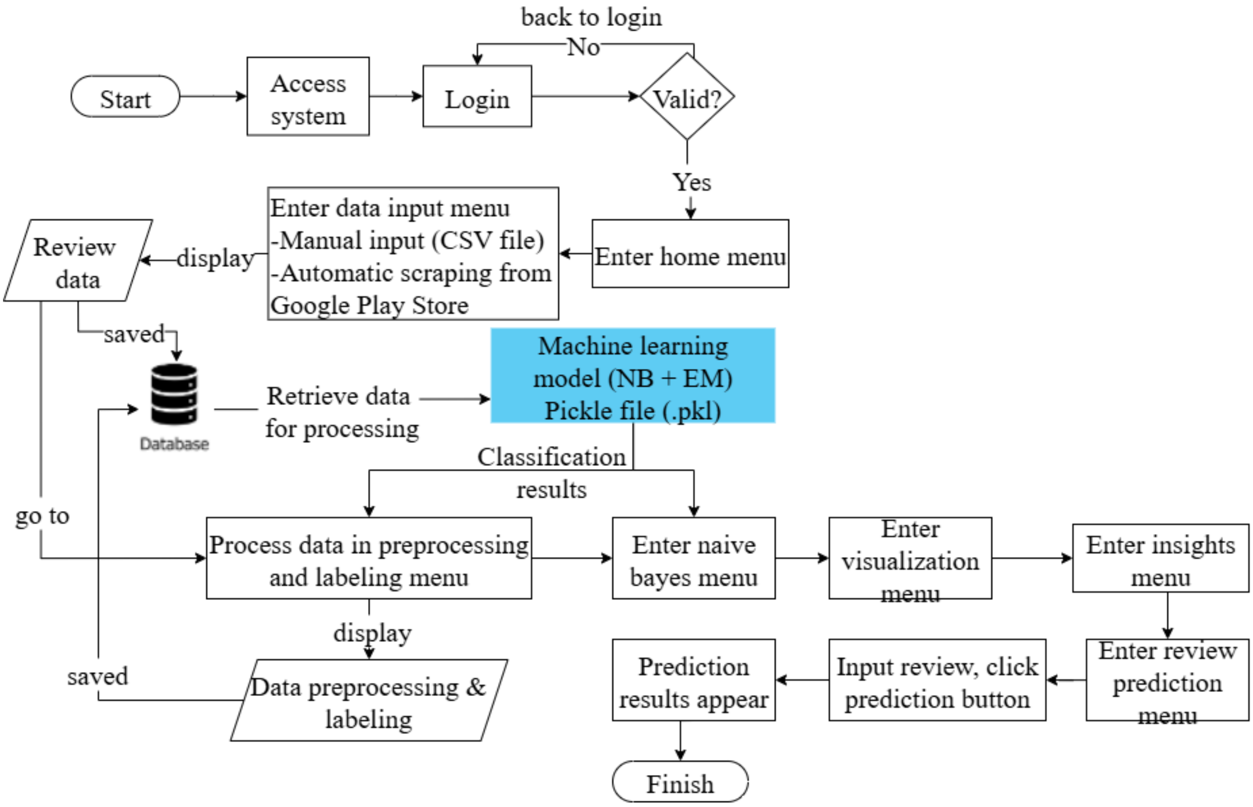


Figure 3. *System flowchart*

The system workflow in Figure 3 begins when the user logs into the application. From the Home menu, data can be entered manually by uploading a CSV file containing user reviews or automatically by retrieving data from the Google Play Store. All data is stored in a database and can be checked before processing. Next, the stored data undergoes pre-processing and labeling, including cleaning, normalization, tokenization, removal of irrelevant words, and stemming, after which it is labeled as positive, neutral, or negative. The processed and labeled data is stored back in the database and classified using a pre-trained Naïve Bayes + Expectation Maximization model saved in a .pkl file. Classification results can be accessed via the Naïve Bayes Menu, the Visualization Menu for graphical sentiment distribution, or the Insights Menu for summary analysis. The Prediction Menu allows users to enter new reviews and immediately obtain sentiment predictions. This workflow ensures efficient data handling, accurate classification, and clear presentation of results.

7.3 Class Diagram

Figure 4 illustrates the modular design of the sentiment analysis web application, showing relationships between classes for user management, data processing, storage, and sentiment classification.

- User: Represents authorized KPU administrators, handling authentication and dashboard access.
- WebAppController: Central controller managing preprocessing, classification, visualization, and model loading.
- Review: Manages user review data from CSV or Google Play scraping, including retrieval, saving, and cleaning.
- PreprocessData: Stores cleaned reviews, sentiment labels, and supports preprocessing and labeling tasks.
- Database: Handles data storage and retrieval for raw and processed reviews.
- MLModel: Loads the trained Naïve Bayes + EM model from a .pkl file and predicts sentiment.

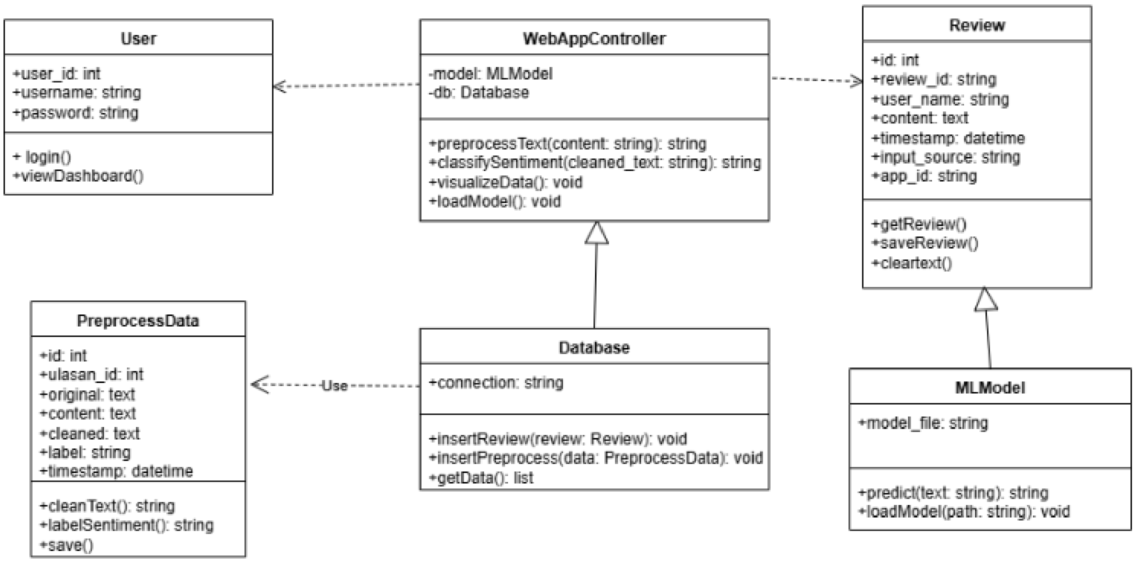


Figure 4. *Class diagram*

RESULTS AND DISCUSSION

1.System Implementation Results

The sentiment analysis system for evaluating user reviews on the SIREKAP Mobile application was successfully developed as a web-based platform using the Flask framework. The system supports multiple functionalities, including data input, text preprocessing, sentiment classification, visualization of analysis results, and real-time sentiment prediction. Users interact with the system through a dashboard interface that facilitates review management, preprocessing operations, sentiment categorization, and visualization of sentiment trends. The system integrates a Naïve Bayes classification model enhanced with the Expectation-Maximization (EM) method to effectively process both labeled and unlabeled datasets. Users can upload datasets manually or obtain them automatically through scraping from the Google Play Store. Afterward, the review data undergoes several preprocessing stages before being classified into three sentiment categories: positive, neutral, and negative. Visualization tools such as pie charts, bar charts, and word clouds are provided to offer comprehensive insights into sentiment trends.

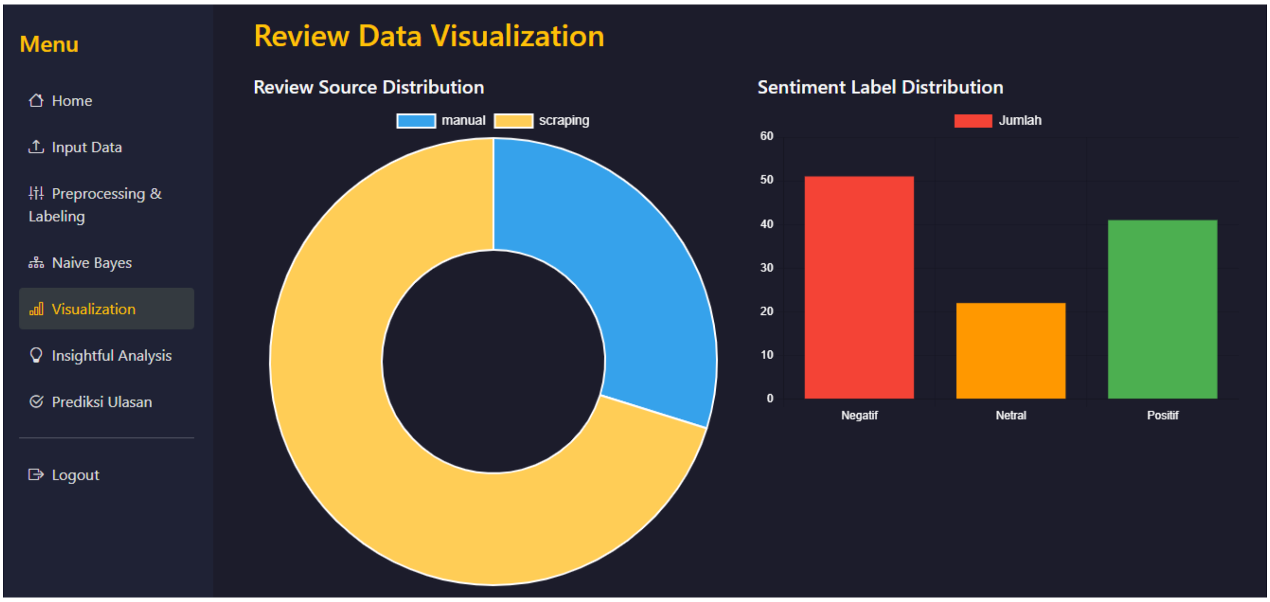


Figure 5. Dashboard system

2. Evaluation of Classification Model

The performance of the sentiment classification model was evaluated using four commonly used machine learning metrics: accuracy, precision, recall, and F1-score. These metrics are essential for measuring the reliability and effectiveness of the classification model in handling real-world sentiment data. The comparison results are presented in Table 1.

Table 1. Comparison of four multinomial naïve bayes models

Model	Accuracy	Precision	Recall	F1-Score
Model 1	84.6%	84.4%	84.6%	83.1%
Model 2	80.5%	80.8%	80.5%	80.3%
Model 3	81.1%	81.9%	81.1%	81.1%
Model 4	83.5%	83.7%	83.5%	83.5%

Results indicate that Model 4, which integrates a semi-supervised Expectation-Maximization (EM) approach, offers the best balance across all metrics (83.5%) and performs consistently across sentiment classes. Although Model 1 achieved the highest accuracy (84.6%), it struggled with neutral sentiment classification, limiting its practical application.

3. Sentiment Distribution Results

Sentiment distribution analysis was conducted to determine the proportions of positive, neutral, and negative sentiments in the classified user reviews. The results of this analysis are illustrated in **Figure 6**, which displays a pie chart representing the overall sentiment composition.

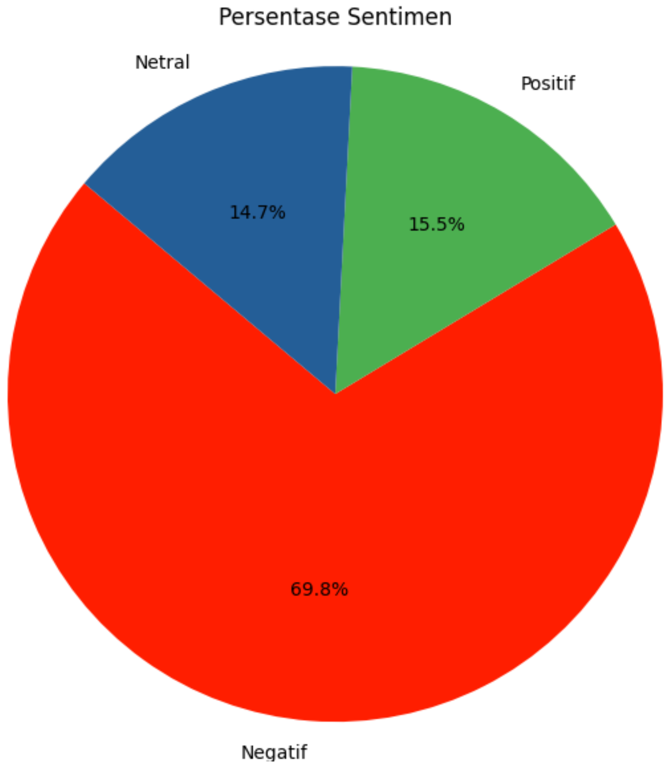
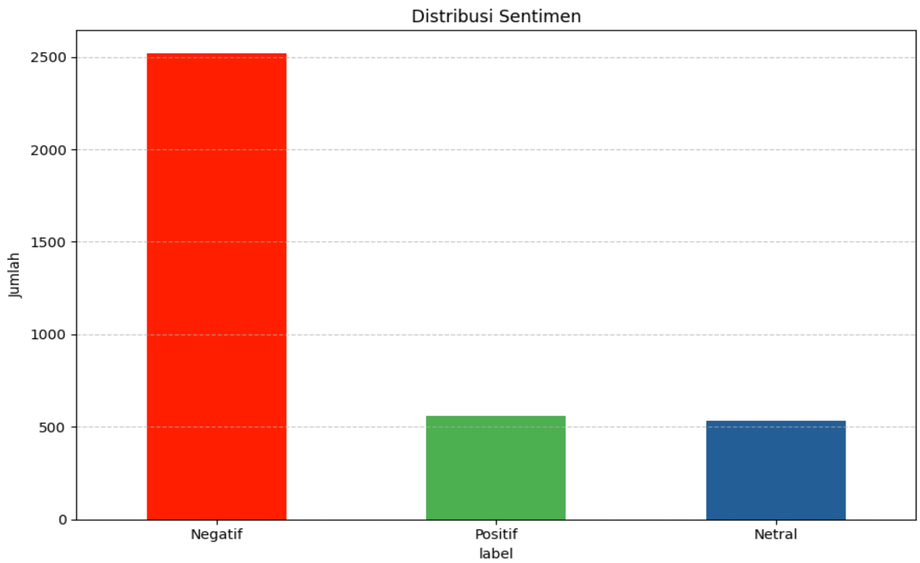


Figure 6. *Sentiment percentage*

The analysis revealed that negative sentiment dominates the reviews, accounting for 69.8% of the total. Meanwhile, positive sentiment represents 15.5%, and neutral sentiment makes up 14.7%. These findings indicate that a substantial portion of users expressed dissatisfaction with the SIREKAP Mobile application, particularly related to issues such as system errors, slow performance, and login difficulties during the election period.




```
label
Negatif    2517
Positif     560
Netral      531
Name: count, dtype: int64
```

Figure 7. *Distribution of sentiment labels*

In addition, figure 7 shows the distribution of sentiment labels across all user reviews in the form of a bar chart. This visualization further emphasizes the imbalance among sentiment categories, with the negative class clearly dominating, represented by more than 2,000 entries. In contrast, both positive and neutral sentiments appear significantly less frequent. This stark contrast highlights the prevalence of user dissatisfaction and underscores the need for deeper investigation into the application's performance issues. The results show that negative sentiment dominated the reviews, accounting for 69.8% of the total. Positive sentiment represented only 15.5%, while neutral sentiment comprised 14.7%. These findings suggest that a significant number of users expressed dissatisfaction with the SIREKAP Mobile application, particularly in relation to issues such as system errors, slow performance, and login difficulties during the election period.

4. Survey-Based Sentiment Ratings (Tulungagung Case Study)

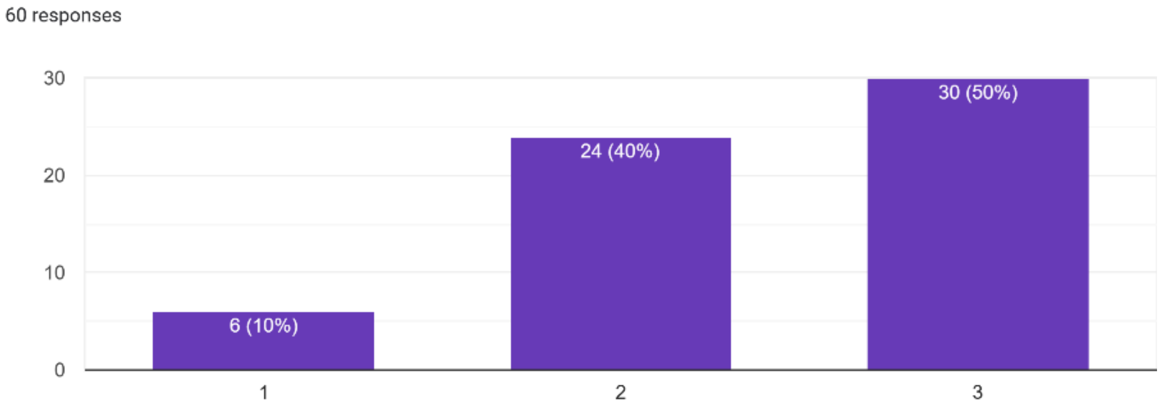


Figure 8. *Survey-based sentiment ratings (Tulungagung case study)*

Figure 8 presents the survey results of user ratings for the SIREKAP Mobile application during the 2024 Regional Election (Pilkada) in Tulungagung Regency. The bar chart is based on feedback from 60 respondents, who provided ratings on a scale from 1 (dissatisfied) to 3 (satisfied).

The results show that:

- 30 respondents (50%) gave a rating of 3 – indicating a positive experience,
- 24 respondents (40%) rated the app as 2 – suggesting a neutral or fair impression,
- while only 6 respondents (10%) rated it 1 – reflecting negative feedback.

These findings reveal that direct field feedback tends to be more favorable compared to the overall sentiment derived from automated text analysis. This discrepancy may be due to differences in expectations, usage context, or the nature of user expression in online reviews versus survey settings.

5. Visualization of Review Content

To further explore the patterns in user feedback, the sentiment classification results were visualized using word cloud diagrams for each sentiment category. Figures 9 to 11 illustrate the most frequently occurring words in reviews classified as negative, neutral, and positive, respectively.

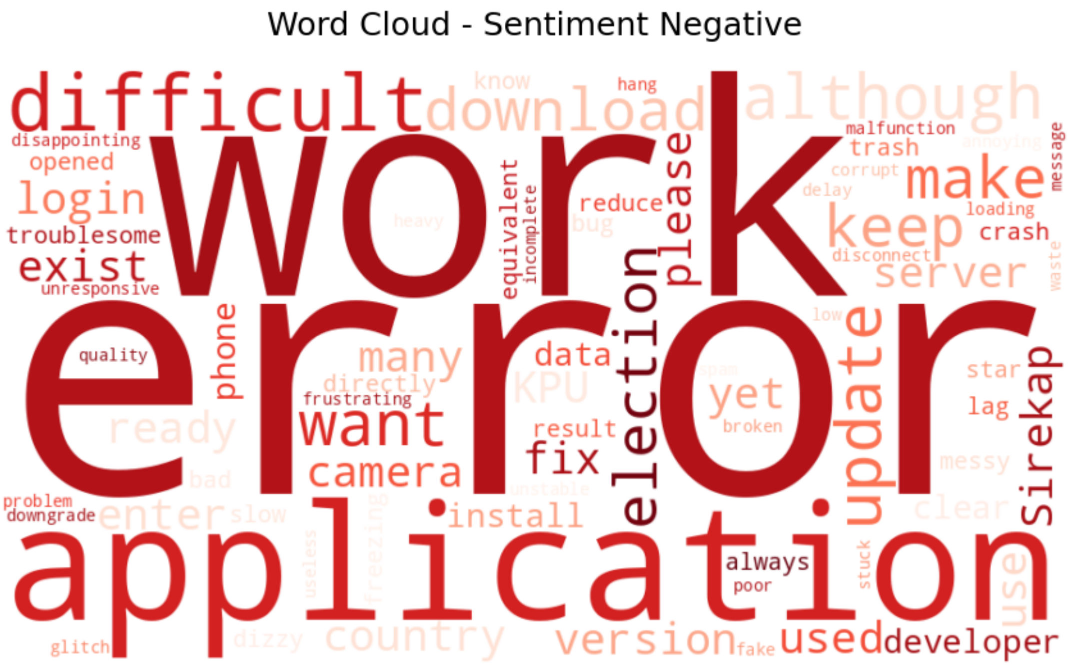


Figure 9. *Word cloud negative*

Figure 9 presents the word cloud generated from reviews labeled as negative. Dominant terms such as "difficult", "error", "login", and "unresponsive" are prominently featured. These keywords reflect common user frustrations with the application's performance, particularly issues related to system errors, access failures, and application crashes during critical moments of the voting process.

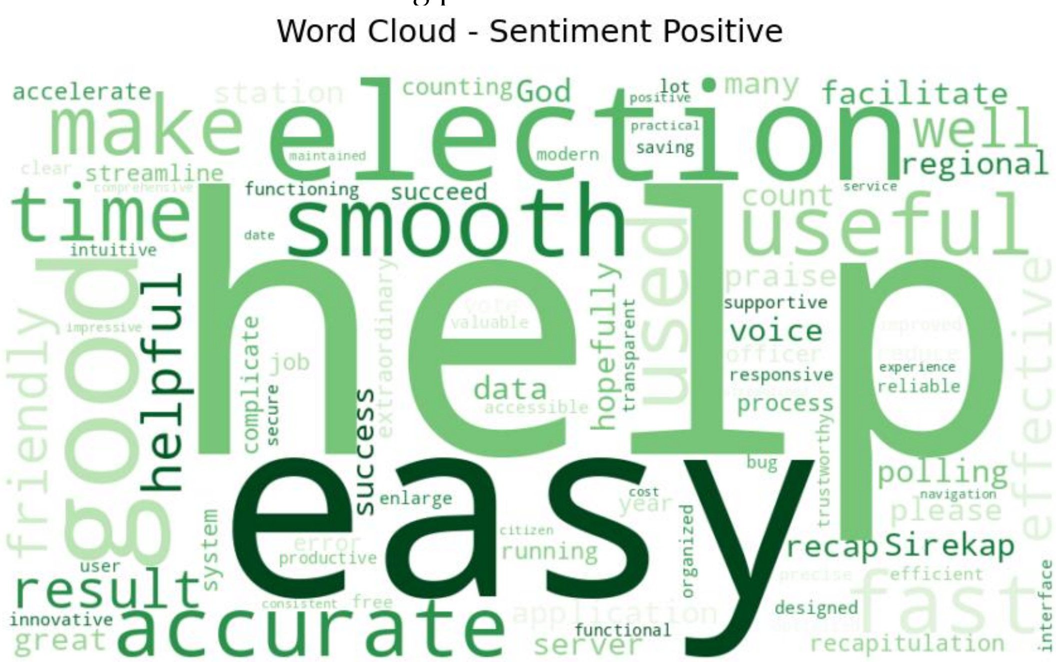


Figure 10. *Word cloud positive*

Figure 10 illustrates the word cloud for positive sentiment. Frequently used words include "easy', "accurate", "good" and "helpful". These terms suggest that users who had successful experiences with the application valued its speed and ease of use, particularly when the system performed as intended.

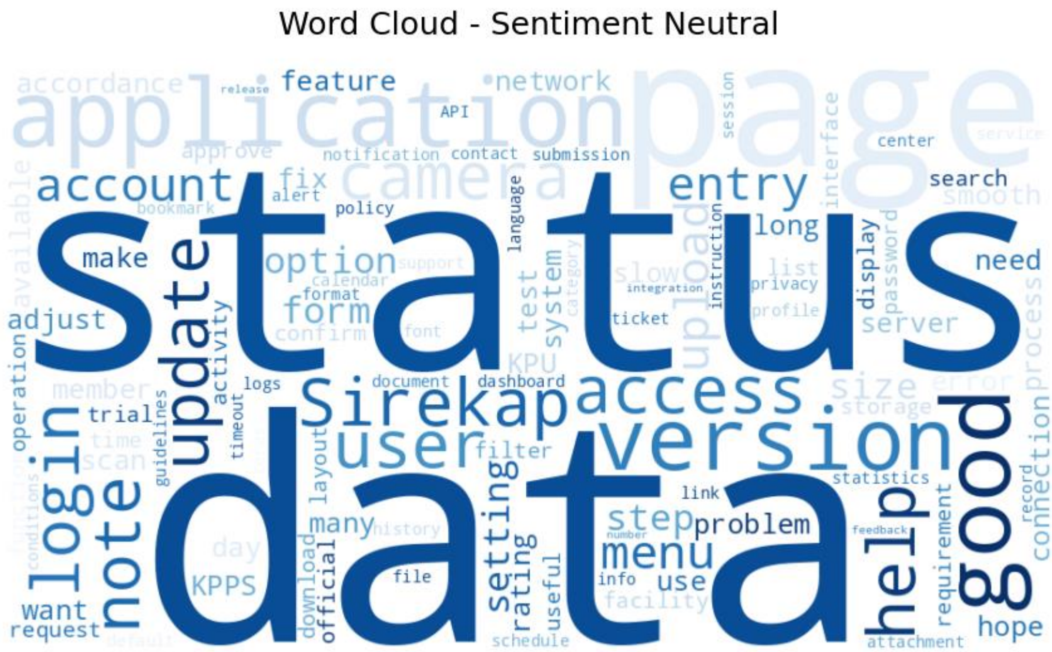


Figure 11. *Word cloud neutral*

Figure 11 shows the word cloud for neutral sentiment. Prominent terms include "status," "data," "page," "application," "access," "update," "version," and "login." These words suggest that neutral comments typically focus on reporting system conditions, usage activities, or providing general feedback without strong positive or negative emotions. The feedback in this category often consists of factual statements or constructive suggestions related to application features, data handling, and interface usability.

Discussion

The study confirms that integrating the Naïve Bayes algorithm with the Expectation-Maximization (EM) method is effective for sentiment analysis, particularly with partially labeled datasets. The optimal configuration, Model 4, achieved balanced precision, recall, and F1-score of 83.5%, demonstrating superior stability compared to other tested models. Although Model 1 attained slightly higher accuracy (84.6%), its poor recall on the neutral class revealed class imbalance. The semi-supervised nature of the chosen model enables strong generalization, making it suitable for real-world applications with limited or costly labeled data. The sentiment distribution analysis revealed that nearly 70% of user reviews for the SIREKAP Mobile application during the 2024 Regional Elections expressed negative sentiment, primarily concerning login failures, slow system performance, and application instability. These findings were reinforced by word cloud visualizations highlighting recurring keywords such as "error," "login," and "difficult," providing empirical evidence for recommendations on technical and user experience improvements. However, field survey results in Tulungagung Regency presented a contrasting view, with 50% of respondents giving the highest satisfaction rating (3/3) and only 10% expressing dissatisfaction. This discrepancy underscores the importance of integrating automated sentiment mining with field survey data to obtain a more comprehensive understanding of public sentiment. The developed system enables real-time sentiment monitoring to support election officials' decision-making. Using TF-IDF feature extraction and structured preprocessing, classification is based on clean and relevant data. The web-based system, built with the Naïve Bayes algorithm combined with the Expectation-Maximization (EM) method, supports inputting user reviews, preprocessing, sentiment classification, visualization, and predicting sentiments for new data. The dataset—sourced from Google Play Store reviews, Google Forms surveys, and TikTok comments was categorized into positive, neutral, and negative sentiment classes.

CONCLUSION

This study successfully developed a sentiment analysis system to evaluate public feedback on the SIREKAP Mobile application during the 2024 Regional Elections by integrating the Naïve Bayes algorithm with the Expectation-Maximization (EM) method. The system achieved an accuracy of 83.5% and effectively handled both labeled and unlabeled data, demonstrating the practical value of semi-supervised learning in public service evaluation. The analysis revealed a predominance of negative sentiment in online reviews, contrasting with more favorable survey responses from Tulungagung, underscoring the importance of integrating multiple data sources for a comprehensive understanding of public sentiment. Overall, this research offers both theoretical and practical contributions to e-government evaluation and provides a foundation for future studies to explore alternative classification techniques, diversify data sources, and extend the framework to other public sector domains.

REFERENCES

Bahtiar, R. A. (2024). *RENCANA PENGGUNAAN SIREKAP PADA PILKADA 2024*. *XVI*(14).

Bai, W., Wang, Y., Chen, W., & Sun, H. (2024). An Expectation-Maximization Algorithm for Training Clean Diffusion Models from Corrupted Observations. *Advances in Neural Information Processing Systems*, *37*(17), 1–25.

Bounabi, M., Elmoutaouakil, K., & Satori, K. (2021). A new neutrosophic TF-IDF term weighting for text mining tasks: text classification use case. *International Journal of Web Information Systems*, *17*(3), 229–249. https://doi.org/10.1108/IJWIS-11-2020-0067

Gifari, O. I., Adha, M., Freddy, F., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, *2*(1), 36–40. https://doi.org/10.46229/jifotech.v2i1.330

Hanafi, M. R., & R, R. K. (2024). *Sentiment Analysis on Sirekap App Reviews on Google Play Using Naive Bayes Algorithm Analisis Sentimen pada Ulasan Aplikasi Sirekap di Google Play Menggunakan Algoritma Naive Bayes*. *4*(October), 1578–1586.

Harrison, S., & Johnson, P. (2019). Challenges in the adoption of crisis crowdsourcing and social media in Canadian emergency management. *Government Information Quarterly*, *36*(3), 501–509. https://doi.org/10.1016/J.GIQ.2019.04.002

Herjanto, M. F. Y., & Carudin, C. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifer. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2), 1204–1210. https://doi.org/10.23960/jitet.v12i2.4192

Ismail, Asih Widiarti, Dani Muhadiansyah, E. K. (2024). *Melihat Teknologi yang*

Digunakan KPU Pada Pemilu 2024. TEMPO Publishing. https://ipusnas2.perpusnas.go.id/book/5c9f6b3b-e55c-46be-91c6-c8c6c08c4667

Ma, Y. (2025). Construction and Data Analysis of a New Media Content Popularity Prediction Model Based on Naive Bayes Algorithm. *Procedia Computer Science*, *261*, 294–302. https://doi.org/10.1016/J.PROCS.2025.04.207

Rahayu, A. S., Fauzi, A., & Rahmat, R. (2022). Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(2), 349. https://doi.org/10.30865/json.v4i2.5398

Sano, A. V. D., Stefanus, A. A., Madyatmadja, E. D., Nindito, H., Purnomo, A., & Sianipar, C. P. M. (2023). Proposing a visualized comparative review analysis model on tourism domain using Naïve Bayes classifier. *Procedia Computer Science*, *227*, 482–489. https://doi.org/10.1016/J.PROCS.2023.10.549

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, *311*, 18–38. https://doi.org/10.1016/J.INS.2015.03.040

Tinanto, B., Zetra, A., & Putera, R. E. (2024). Diffusion of Recapitulation Information System (SIREKAP)Innovation in Realizing the Integrity of the 2024 Elections in MerantiIslands Regency, Riau Province, Indonesia. *JPSI (Journal of Public Sector Innovations)*, *9*(1), 24–33.

Ummah, M. S. (2019). No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title. *Sustainability (Switzerland)*, *11*(1), 1–14. http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI

van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, *109*(2), 373–440. https://doi.org/10.1007/s10994-019-05855-6

Wahid, A., & Saputri, G. (2022). Analisis Sentimen Komentar Youtube Tentang Relawan Patwal Ambulance Menggunakan Algoritma Naïve Bayes dan Decision Tree. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(2), 319. https://doi.org/10.30865/json.v4i2.4941

Zahra, M. S., Mulyadi, D., & Syahla, A. K. (2024). The Impact of Technological Advancement in the Use of the SIREKAP Program in the 2024 Elections is studied according to the Principles of Public Accountability for Democratic Election Calculations. *Al-Musthalah: Jurnal Riset Dan Penelitian Multidisiplin*, *1*(2), 208–220.