Comparative Analysis of SVM and NB Algorithms in Evaluating Public Sentiment on Supreme Court Rulings

Putri Dwi Rahayu Maulidiana[1], Anik Vega Vitianingsih[2]*, Slamet Kacung[3], Anastasia Lidya Maukar[4], David Hermansyah[5]
Informatics Department, Universitas Dr. Soetomo, Surabaya, Indonesia[1], [2], [3], [5]
Industrial Engineering Department, President University, Bekasi, Indonesia[4]
putridwir7@gmail.com[1], vega@unitomo.ac.id[2], slamet@unitomo.ac.id[3], almaukar@president.ac.id[4], david.hermansyah@unitomo.ac.id[5]

***Abstract*— The legal events that happened to Ferdy Sambo and the Supreme Court's decision in the cassation triggered emotional reactions and various opinions among the public, especially on social media sites such as Xapps. Some comments reflect people's concerns about fairness in the legal system. They doubted the integrity of legal institutions or believed that decisions were unfair or in line with vested interests. This research aims to analyze public perceptions of Supreme Court decisions. The research process includes data collection, preprocessing, labeling, weighting, classification using Support Vector Machine and Naïve Bayes, and performance evaluation using a confusion matrix. A dataset of 624 was taken from X apps using the Twitter scraping technique. The lexicon method is used for data labeling, dividing the data into positive, negative, and neutral classes. The analysis results show 46 tweets categorized as positive sentiment, 133 tweets categorized as negative sentiment, and 422 tweets categorized as neutral sentiment. Based on testing with a data ratio of 80:20, both SVM and NB methods show good performance. The SVM criteria showed an accuracy of 0.84, precision of 0.61, recall of 0.78, and f1-score of 0.66, while the NB criteria showed an accuracy of 0.73, precision of 0.37, recall of 0.57, and f1-score of 0.35.**

Keywords*— Sentiment Analysis, Naïve Bayes, SVM, Ferdy Sambo, Xapps.

I. INTRODUCTION

The Supreme Court is a judicial institution that serves as the highest court in a country or jurisdiction[1]. The judicial system is crucial in maintaining justice and public trust in the law. The issue of corruption has become a troubling problem in many countries, including Indonesia. Corruption committed by government officials impacts state financial losses and erodes public trust in government institutions. When public figures are involved in corruption cases, this can give the impression that corruption has penetrated various layers of society, including those who are supposed to be role models.

One of the cases that caused tension and raised serious questions about the honesty and integrity of public figures was the murder of Brigadier Yosua on July 8, 2022, involving high-ranking police officers. This case shook the community and made headlines in the media. The murder caused an emotional response and public concern because it involved the life of a police officer who was carrying out his duties. The legal process that followed the murder of Brigadier Yosua became an ongoing public spotlight. This legal struggle culminated when Ferdy Sambo filed an appeal to the Supreme Court as a last resort to change his fate. The Supreme Court ruling issued on 18-08-2023[2] is back in the news. The Supreme Court's decision in Ferdy Sambo's appeal has clear implications for the death penalty and the justice system in Indonesia.

The involvement of a public figure in a controversial legal case tarnishes his excellent name and creates public distrust of the legal system and law enforcement officials. Twitter, now known as X apps, has become a popular platform worldwide that allows people to get the latest news, share views or thoughts, and participate in various online discussions[3]. The rapid development of social media makes it easy for people to engage and participate in discussions on these topics. Sentiment analysis is an appropriate approach to understanding how the public responds to Supreme Court decisions in these cases. This research can provide insights for policymakers or law enforcers regarding public opinion on these decisions and help improve the quality of legal decisions and public confidence in the justice system.

This study aims to understand the responses and opinions of the public regarding the Supreme Court's decision in Ferdy Sambo's cassation and identify the pattern of public opinion, including positive, negative, or neutral aspects. SVM and NB methods are used to compare the performance of both and aim to determine which method provides better results when analyzing sentiment. The data used in this study was obtained from the Twitter social media platform, currently known as X apps, through a crawling process. The next step involves data preprocessing, which includes cleaning, case folding, tokenization, normalization, stopword removal, and stemming processes. Once the preprocessing stage is complete, the next step is to label the sentiment classes of attributes such as positive, negative, and neutral, using a Lexicon-based dictionary. The next process includes TF-IDF weighting, after which the system will process the data using SVM and NB methods. The final step is model evaluation, which uses a confusion matrix to generate accuracy, precision, and recall metrics.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2116, Copyright ©2024
Submitted : March 7, 2024, Revised : April 29, 2024, Accepted : May 2, 2024, Published : June 15, 2024

Several previous studies have conducted sentiment analysis research that discusses public opinion regarding the image of the Police institution. Research by researchers [4] obtained 1100 datasets using the SVM method with an accuracy of 85.5%, precision of 86%, recall of 64%, and the NB method with an accuracy value of 81.25%, precision of 91%, and recall of 51%. In addition, another study [5] with 269 datasets using the NB Classifier method resulted in an accuracy rate of 98.51%, precision of 98.97%, and recall of 97.40%. Similar research related to other subjects [6] with 2000 datasets using the NB method resulted in an accuracy of 95%, precision of 90.82%, and recall of 98.89%. Not only that, other researchers [7] with 1000 datasets using the SVM method got 88% accuracy, Extra Trees Classifier with 86% accuracy, Logistic Regression with 85% accuracy, Random Forest Classifier with 85% accuracy, K-Nearest Neighbor with 83% accuracy, and Multinomial Naïve Bayes with 78% accuracy. In another study [8], 1200 datasets using the SVM method performed well in the 90:10 data scenario with an accuracy value 0.82.

Based on the previously presented literature review, the authors take a sentiment analysis approach by applying natural language processing techniques. The methods applied include SVM and NB, where SVM proves effective in managing complex or high-dimensional data, while NB offers a simpler and more efficient approach in processing text. With the application of SVM which again provides optimal results, this research successfully identifies public opinion with a higher level of accuracy, provides readers with an in-depth understanding of public responses to Ferdy Sambo's cassation verdict, and provides relevant policy recommendations based on the results of sentiment analysis.

Sentiment analysis is crucial in understanding individual opinions and perceptions on a particular topic[9]. Overall, sentiment analysis provides valuable insights for various organizations and entities to make better decisions, respond more effectively to customer feedback, and understand the changing dynamics of public opinion[10]. Within the framework of this study, sentiment analysis provides comprehensive insights into the diverse perspectives and sentiments of the public concerning the Supreme Court's decision in Ferdy Sambo's cassation case. Therefore, the sentiment analysis results provide a more detailed understanding of the public's positive, negative, or neutral feelings.

II. RESEARCH METHODS

This research uses a comparative approach that pits the Support Vector Machine (SVM) with the Naïve Bayes (NB) algorithm to measure the accuracy of public sentiment toward Ferdy Sambo's cassation case. The implementation of this comparative model involves several stages to achieve optimal accuracy. Figure 1 shows the various steps in the research process.

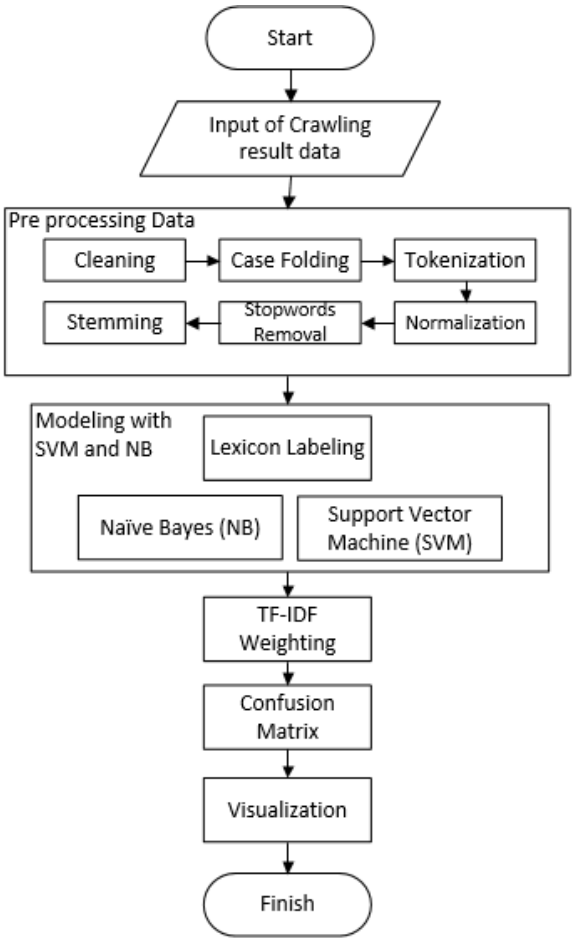

Fig 1. Research Process Design

The research begins by collecting data by crawling Twitter. Then, the crawl results will go through preprocessing, including cleaning, case folding, tokenization, normalization, stopword removal, and stemming. Next, data labeling is done with a lexicon to distinguish positive, negative, and neutral class categories. After labeling, feature extraction is performed using the Term Frequency-Inverse Document Frequency method (TF-IDF). The classification is performed with SVM and NB algorithms. The confusion matrix will evaluate the outcomes to guarantee F1-Score, accuracy, precision, and recall.

A. Data Collection

The dataset utilized in this research came from comments or tweets seen in the *detikcom* and *kumparan* account posts on 08-08-2023[11][12]. The dataset was collected using the Python library and Google Collaboratory crawling techniques. The total data collected reached 624 tweets, which were then saved in **.csv* file format.

B. Preprocessing Data

In this process, data preprocessing is performed before proceeding to sentiment analysis. The data obtained through the crawling process is a raw dataset, which is still unstructured text and contains a lot of noise[13]. Preprocessing is a critical step in text analysis that aims to improve data quality and remove

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2116, Copyright ©2024
Submitted : March 7, 2024, Revised : April 29, 2024, Accepted : May 2, 2024, Published : June 15, 2024

irrelevant text efficiently[14]. Therefore, processing is needed to make the data more structured and easily understood. The following preprocessing stages refer to Figure 2 [14]–[16].

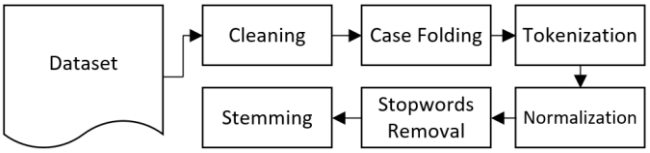

Fig 2. Preprocessing Flow

1. *Cleaning is a process where text data is cleaned from irrelevant information or noise.*
2. *Case folding refers to changing every capital letter to a lowercase letter.*
3. *Tokenization breaks down text into the smallest units, such as sentences or words.*
4. *Normalization is changing less efficient words into words that follow the standards in the KBBI (Kamus Besar Bahasa Indonesia).*
5. *Stopword removal is taking essential words in a sentence or removing words without meaning.*
6. *Stemming, which is converting words into primary forms (root words) to reduce word variations.*

C. Data Labeling

After the preprocessing stage, the data consists only of cleaned opinions. After that, each sentence is scored for positive, negative, and neutral labeling. The automatic data labeling process will involve using a lexicon dictionary to calculate the score value. The score calculation results show that the sentence is categorized as a positive class if the score is > 0, a negative class if the score is < 0, and a neutral class if the score = 0 [17]. Calculating the sentence sentiment score can use Equation 1[17].

$$Skor = (\sum kata\ positif - \sum kata\ negatif) \quad (1)$$

D. Word Weighting

The correlation of a word (term) in a document can be given significance using the TF-IDF (Term Frequency-Inverse Document Frequency) method[18]. This method combines two fundamental weight calculation concepts: the frequency of a word appearing in a certain document and the inverse frequency of documents containing that word[19].

The occurrence rate of a word in a given document serves as a measure of the word's significance in a given document. In contrast, the frequency of documents that include the word provides insight into the word's overall prevalence across documents. Accordingly, when a word appears frequently in a document and is not frequently found in other papers, the association between the term and the document is strongest[19]. TF-IDF weight calculation can be done using Equation 2.

$$TF.IDF_{std}(t) = tf_d^t\ x\ log\frac{N}{df^t} \quad (2)$$

Where the tf_d^t $v$ariable is the number of times the phrase "t" appears in the document $d.N$ denotes the total number of quantity in the corpus, while df^t Specifically, it refers to the number that contains the *t* variable.

E. Model Analysis

The algorithm model analysis involves a thorough comparison between the Support Vector Machine and Naïve Bayes, which is performed separately and aims to distinguish which model produces the most optimal accuracy value.

1) Support Vector Machine (SVM)

SVM is a computational algorithm applied to categorize or partition a data set into different groups [20]. It is one of the machine learning algorithms commonly used for classification and regression. In the classification stage, SVM can provide optimal results even with limited datasets. SVM's capabilities include handling overfitting problems by applying regularization techniques [20]. The basic formula of Support Vector Machine can be seen in Equations 3 and 4.

$$f(\emptyset(x)) = w.\emptyset + b \quad (3)$$
$$f(x) = w.x + b \quad (4)$$

Where w is the weight vector, x is the available data, and b is the switch value.

2) Naïve Bayes (NB)

NB consists of a set of classification algorithms derived from the principles of Bayes' Theorem[21]. This method assumes that each feature or attribute used in classification is independent. The naive Bayes formula can be seen in Equation 5

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)} \quad (5)$$

The formula of Bayes' Theorem in Equation 5. Where *c,d* denote events. $P(c|d)$ signifies the likelihood of C given D, conditionally. The conditional probability of D given C is represented by $P(d|c).P(c), P(d)$ indicates the likelihood of either C or D occurring.

F. Evaluation

In machine learning, evaluation measures the system's accuracy and tests the classification findings [22]. After obtaining a classification method using SVM and Naïve Bayes, an evaluation stage is required before applying the method to test data. Evaluation is done by applying the Confusion Matrix method. The confusion matrix is one of the techniques used to evaluate how effective or ineffective a classification process performs [23]. The confusion matrix considers the level of accuracy, precision, and recall. The ratio of accurate predictions—positive, neutral, and negative—to the total data determines accuracy. The ratio of True Positive (TP) predictions to total positive predictions is used to calculate the precision. The ratio of True Positive (TP) predictions to all positive data is known as recall. The confusion matrix, outlined in the (6),(7), and (8), contains all this information[23].

$$precision = \frac{TP}{TP+FP} \quad (6)$$

$$recall = \frac{TP}{TP+FN} \quad (7)$$

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2116, Copyright ©2024
Submitted : March 7, 2024, Revised : April 29, 2024, Accepted : May 2, 2024, Published : June 15, 2024

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

TP, True Negative events by TN, False Negative events by FN, and False Positive events by FP denote True Positive events. As opposed to TP, TN indicates that both the expert and the system produce unfavorable outcomes. When the expert provides a positive result but the system offers a negative outcome, this is known as FN. Conversely, FP happens when the expert offers a positive result, but the system provides a negative result[23].

III. Results and Discussion

A. Data Collection

By applying crawling methods, this research successfully collected 624 tweets from public comments related to the case on the X apps detikcom and *kumparan* accounts. Table I. results from the data collection process carried out by crawling Twitter.

TABLE I. Sample Comment Data

No.	Comment
1.	*Tidak ada yg tidak mungkin buktinya hukuman mati kena diskon flash sale 8.8*
2.	*Tidak ada yg tidak mungkin untuk hukum dinegara Konoha ini pak*
3.	*Another Kejadian paling lucu di 2023* ☝😂

B. Data Labeling

The data that has been collected is then subjected to a labeling stage to classify whether the comment belongs to the negative, positive, or neutral sentiment category.

TABLE II. Labeling Results

No.	Comment	Score	Sentiment
1.	*Tidak ada yg tidak mungkin buktinya hukuman mati kena diskon flash sale 8.8*	-1	Negative
2.	*Tidak ada yg tidak mungkin untuk hukum dinegara Konoha ini pak*	0	Neutral
3.	*Another Kejadian paling lucu di 2023* ☝😂	1	Positive

Table II is the result of labeling using the lexicon formula. The number -1 indicates a negative label, 0 indicates neutral, and 1 indicates positive. The entire dataset of 624 went through this process. As a result, there are 422 comments with neutral sentiments, 133 negative comments, and 46 positive comments. Some data does not fit into this category, probably because the comment contains # that are considered unimportant and are not detected by the machine.

C. Data Preprocessing

As described in the research method, the preprocessing stage of this research involves several steps, including cleansing, case folding, tokenization, normalization, and stemming. Table III. is the result of the data-cleaning process. The table shows that the data has undergone cleansing, case folding, tokenization, normalization, stopword, and stemming.

TABLE III. Preprocessing Results

Dataset	
Another *Kejadian paling lucu di 2023* ☝😂	
Preprocessing	
Cleansing	*Another Kejadian paling lucu di*
Case folding	*another kejadian paling lucu di*
Tokenize	*"another", "kejadian", "paling", "lucu", "di"*
Normalized	*"another", "kejadian", "paling", "lucu", "di"*
Stopword	*"another", "kejadian", "paling"*
Stemming	*"another", "jadi", "paling"*

D. SVM Classification

The next stage involves data classification, achieved by applying the Support Vector Machine (SVM) algorithm method. Figure 3 shows the results of the algorithm method with a ratio of 80:20, which obtained 3 diagram results, with 4 neutral comments, 17 negative comments, and 100 positive comments.

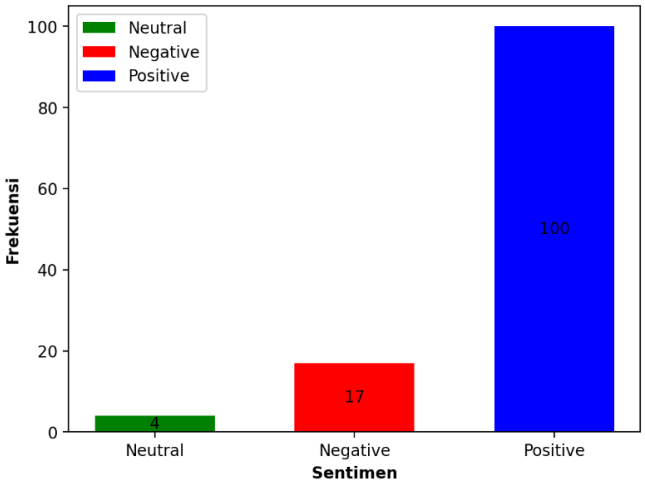

Fig 3. SVM classification result

E. NB Classification

In the next stage, in addition to data classification through NB method. Figure 4 shows the results of algorithm method with a ratio of 80:20 obtained 2 diagram results, with 3 negative comments and 118 positive comments.

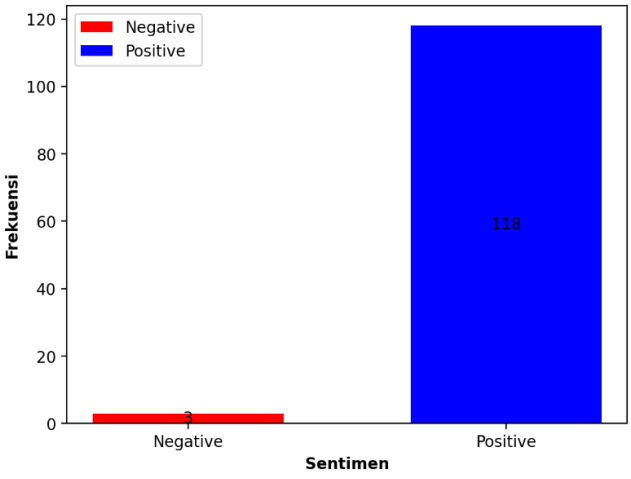

Fig 4. NB classification result

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2116, Copyright ©2024
Submitted : March 7, 2024, Revised : April 29, 2024, Accepted : May 2, 2024, Published : June 15, 2024

F. Model Evaluation

After completing a thorough set of tests against all the data, the next step involves evaluating the performance of the implemented model. This evaluation uses a confusion matrix to calculate accuracy, precision, and recall values. The model performance using the confusion matrix for the SVM method can be found in Figures 5 and 6

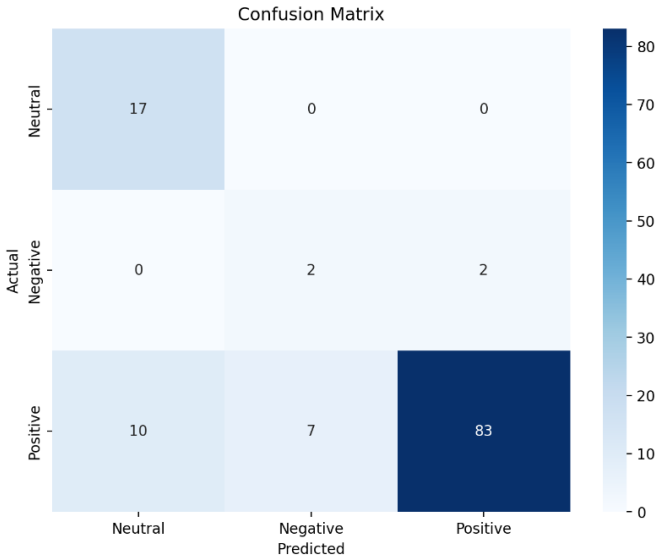

Fig 5. SVM Confusion Matrix Result

classification report :	precision	recall	f1-score	support
-1	0.63	1.00	0.77	17
0	0.22	0.50	0.31	4
1	0.98	0.83	0.90	100
accuracy			0.84	121
macro avg	0.61	0.78	0.66	121
weighted avg	0.90	0.84	0.86	121

Fig 6. SVM Confusion Matrix Result

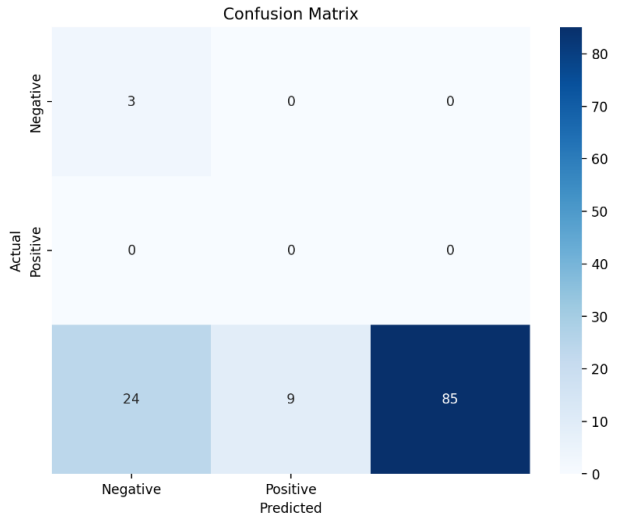

Fig 7. NB Confusion Matrix Result

Naive Bayes classification report :	precision	recall	f1-score	support
-1	0.11	1.00	0.20	3
0	0.00	0.00	0.00	0
1	1.00	0.72	0.84	118
accuracy			0.73	121
macro avg	0.37	0.57	0.35	121
weighted avg	0.98	0.73	0.82	121

Fig 8. NB Confusion Matrix Result

Based on the confusion matrix results illustrated in Figures 5 and 6, the SVM method obtained an accuracy value of 0.84, a precision level of 0.61, a recall level of 0.78, and an f1-score of 0.66. Furthermore, Figures 7 and 8 display the confusion matrix with the NB method. Based on the confusion matrix results in Figures 7 and 8, the NB method obtained an accuracy value of 0.73, precision of 0.37, recall of 0.57, and f1-score of 0.35.

G. Visualization Word Clouds

As described in the Research Methods, this study successfully generated a visualization of a three-word cloud: word clouds for positive, negative, and neutral comments. An image of the word cloud showing the words that appear frequently in all the data with positive sentiments can be seen in Figure 9.

Fig 9. WordCloud Positive

Figure 9 represents the words often appearing in people's tweets with positive sentiments about this case. A wide variety of word expressions contain good sentences. Table IV is some examples of positive sentences in the dataset. Based on the words in the word cloud:

TABLE IV. EXAMPLES OF POSITIVE SENTENCES IN THE DATASET

Sentences in the dataset
respect hakim adil pihak sisi adil
Pengadilan dunia hanyalah sebuah tempat dimana hakim selaku aparat penegak hukum memiliki andil untuk dapat mewakili tuhan dalam menyelesaikan suatu perkara. Akan ada masanya sebenar-benarnya pengadilan di hari akhir menampakan wujudnya-

An image of a word cloud showing frequently occurring words within all data with negative sentiments can be seen in

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2116, Copyright ©2024
Submitted : March 7, 2024, Revised : April 29, 2024, Accepted : May 2, 2024, Published : June 15, 2024

Figure 10.

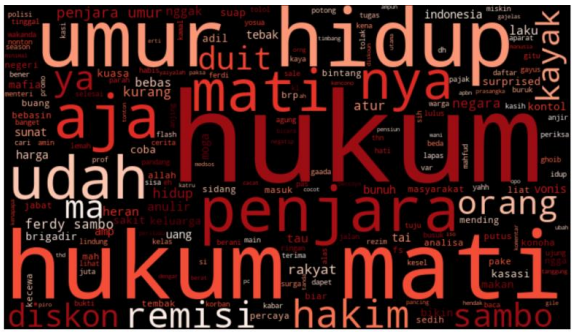

Fig 10. WordCloud Negative

Figure 10 can represent the words that often appear in the tweets of people with negative sentiments about this case. Various kinds of word expressions contain unfavorable sentences. Table V shows some examples of negative sentences in the dataset. Based on the words in the word cloud.

TABLE V. EXAMPLES OF NEGATIVE SENTENCES IN THE DATASET

Sentences in the dataset
Hukuman mati aja bisa jd d buat seumur hidup Apalagi remisi Rakyat udah dong" kemana arahnya Udah kenyang dikibulin Cuma g' bs ngebantah aja Ntr makin di persulit hidupnya🤔"
Penjara itu menyakitkan bagi orang kecil yang masuk kalau bagi orang kaya dan pejabat menyakitkan atau biasa aja pak?

The word cloud image showing the words that frequently appear in all the data with neutral sentiments can be seen in Figure 11.

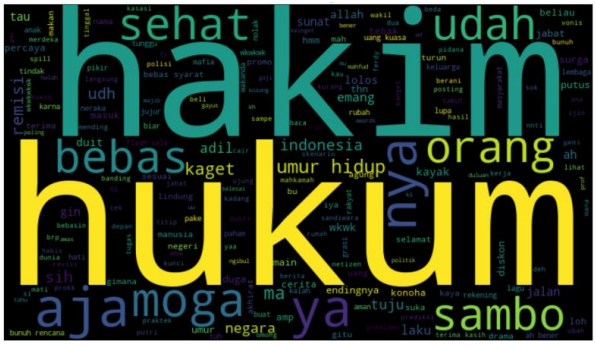

Fig 11. WordCloud Neutral

Figure 11 represents the words often appearing in people's tweets with positive sentiments about this case. A wide variety of word expressions contain neither positive nor negative sentences. Table VI shows some examples of neutral sentences in the dataset. Based on the words in the word cloud.

TABLE VI. EXAMPLES OF NEUTRAL SENTENCES IN THE DATASET

Sentences in the dataset
Dasar hakim nerima kasasi sambo apa?
Yg 3 hakim lainnya bedebah

IV. CONCLUSION

The sentiment analysis results of this research refer to public opinions related to the Supreme Court's decision in the Ferdy Sambo cassation case using SVM and NB methods. In this study, both methods were tested to see their performance. Data was collected through scraping from platform X with 624 entries. Of the 601 data cleaned through the preprocessing stage, there are 422 comments with neutral sentiments, 133 negative comments, and 46 positive comments. In addition, some data does not fit into the category, probably because it contains the character "#" which is considered irrelevant by the machine. In the 80:20 ratio test, the SVM method generated 100 positive comments, 17 negative comments, and 4 neutral comments, while NB generated 118 positive comments and 3 negative comments.

The evaluation results show that the SVM method has an accuracy of 0.84, precision of 0.61, recall of 0.78, and f1-score of 0.66. In contrast, the NB method has an accuracy of 0.73, precision of 0.37, recall of 0.57, and f1-score of 0.35. From these test results, it can be concluded that the SVM method is superior to the NB method. Based on the results of this study, several suggestions emerged as potential focus areas that future researchers could explore. Although SVM outperformed NB in terms of accuracy, it is suggested that this approach be extended to include more machine learning techniques or even try a hybrid approach to improve model dependability. Future researchers are also encouraged to examine the possibility of employing oversampling or undersampling methodologies to enhance the system's accuracy in the classification of sentiments.

REFERENCES

[1] D. I. Prof. Dr. Sulistiyono S, *Sistem Peradilan Di Indonesia Dalam Teori Dan Praktik*. Tapos - Cimanggis, Depok: PRENADAMEDIA GROUP, 2018.

[2] K. M. A. R. Indonesia, "Direktori Putusan Mahkamah Agung Republik Indonesia," *Publikasi Dokumen Elektronik Putusan seluruh Pengadilan di Indonesia*. .

[3] and A. F. O. P. D. Darwis, E. S. Pratiwi, "Data Twitterkomisi Pemberantasan Korupsi Republik Indonesia," *J. Ilm. Edutic*, vol. 7, no, 2020.

[4] Lisa R, "Digital Repository Universitas Jember Digital Repository Universitas Jember," *Anal. Sentimen Opini Masy. Terhadap Institusi Polri Pada Media Sos. Twitter Menggunakan Metod. Support Vector Mach. dan Naïve Bayes*, vol. 3, no. 3, pp. 69–70, 2022.

[5] R. N. Fahmi, N. Nursyifa, and A. Primajaya, "Analisis Sentimen Pengguna Twitter Terhadap Kasus Penembakan Laskar Fpi Oleh Polri Dengan Metode Naive Bayes Classifier," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 61–66, 2021.

[6] S. Nurhaliza, D. Febriawan, and F. N. Hasan, "Sentiment Analysis of Society Towards the Child-free Phenomenon (Life Without Children) on Twitter Using Naïve Bayes Algorithm," vol. 13, pp. 40–48, 2024.

[7] K. Hasanah, "Comparison of Sentiment Analysis Model for Shopee Comments on Google Play Store," vol. 13, pp. 21–30, 2024.

[8] E. R. Kaburuan and N. R. Setiawan, "Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 105–116, 2023.

[9] H. R. Nugraha F, Haranii N, *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Bandung: Kreatif Industri Nusantara, 2020.

[10] D. Asri Y, Kuswardanu D, "MACHINE LEARNING & DEEP LEARNING: Analisis Sentimen Menggunakan Ulasan Pengguna

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2116, Copyright ©2024
Submitted : March 7, 2024, Revised : April 29, 2024, Accepted : May 2, 2024, Published : June 15, 2024

Aplikasi." Uwais Inspirasi Indonesia, Ponorogo, 2019.

[11] "MA Anulir Vonis Mati Ferdy Sambo, Jadi Penjara Seumur Hidup!," *detikcom*, 2023. .

[12] "MA kabulkan kasasi Ferdy Sambo jadi divonis seumur hidup," *kumparan*, 2023. .

[13] I. S. Swastika R, Mukodimah S, Susanto F, Muhamad M, *IMPLEMENTASI DATA MINING (Clastering, Association, Prediction, Estimation, Classification)*. Indramayu: Penerbit Adab, 2023.

[14] P. D. Oleh Fitri Marisa, S.Kom., M.Pd. , Anastasia Lidya Maukar, S.T., M.Sc., M.M.T. , Dr. Tubagus Mohammad Akhriza, S.Si., M.M.S.I., *Data Mining Konsep Dan Penerapannya*. Yogyakarta: Deepublish, 2021.

[15] J. Hirschberg and C. D. Manning, "Advances in natural language processing," *Science (80-.).*, vol. 349, no. 6245, pp. 261–266, 2015.

[16] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. 2015.

[17] D. Solimun, Fernandes A, *Metodologi Penelitian*. Tim UB Press, 2023.

[18] M. S, *FAST TEXT DAN WORD2VEC PADA QUERY KESAMAAN SEMANTIK SISTEM TEMU KEMBALI INFORMASI*. Sleman: Deepublish, 2022.

[19] S. T. H. Faisal R, Kartini D , Arrahimi R A, *Belajar Data Science: Text Mining Untuk Pemula I*. Kalimantan Selatan: Scripta Cendekia, 2022.

[20] Yahya, *DATA MINING*. Bojong genteng, Sukabumi: Tim CV Jejak, 2022.

[21] D. Permana A, Wahyuddin S, *MACHINE LEARNING*. PT Global Eksekutif Teknologi, 2023.

[22] D. Junaidi S, Beno I, *MACHINE LEARNING*. Jambi: PT. Sonpedia Publishing Indonesia, 2024.

[23] D. Mustika, Ardilla Y, *DATA MINING DAN APLIKASINYA*. Bandung: Widina Bhakti Persada Bandung (Grup CV. Widina Media Utama), 2021.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2116, Copyright ©2024
Submitted : March 7, 2024, Revised : April 29, 2024, Accepted : May 2, 2024, Published : June 15, 2024