

IJID_Fathurrohman, Afandi, Hasan - cek.01

by Cek Turnitin UHAMKA

Submission date: 27-May-2024 02:44PM (UTC+0700)

Submission ID: 2389089074

File name: Fathurrohman,_Afandi,_Hasan_-_cek.01.docx (616.82K)

Word count: 5948

Character count: 34036

Sentiment Analysis of Public Opinion on Twitter Towards the Indonesian National Team in the Asian Cup U-23 2024 Using Naive Bayes and SVM Algorithms

Sewin Fathurrohman ^{1*}
Department of Information Systems
Universitas Gunadarma
line 3: Depok, Indonesia
sewin.fathur@gmail.com

⁶ Irfan Ricky Afandi ²
Department of Informatics Engineering
Universitas Muhammadiyah Prof. Dr. Hamka
Jakarta, Indonesia
richky99@gmail.com

⁶ Firman Noor Hasan ³
Department of Informatics Engineering
Universitas Muhammadiyah Prof. Dr. Hamka
Jakarta, Indonesia ⁸
firman.noorhasan@uhamka.ac.id

Article History

Received month dd, yyyy
Revised month dd, yyyy
Accepted month dd, yyyy
Published month, yyyy

Abstract— The high interest and overwhelming enthusiasm of the Indonesian people towards football make the participation of the Indonesian National Team in international competitions a highly anticipated moment. One of the prestigious events that is eagerly awaited is the Asian Cup U-23 2024. The Indonesian National Team participates in this competition, carrying the high hopes of many that the players will deliver their best performance. Social media serves as an effective communication platform where the public can express their opinions and feelings. Many Indonesians show their high enthusiasm through support, opinions, and reactions to the performance of the Indonesian National Team in the 2024 U-23 Asian Cup, especially on Twitter. The sentiments expressed by the public can be a valuable source of data to understand their responses to the team's performance in the competition. Therefore, proper, fast, and accurate data processing is required to analyze these sentiments. This research focuses on the sentiment analysis of the public on Twitter regarding the performance of the Indonesian National Team players in the 2024 U-23 Asian Cup. Public sentiments are classified into negative, neutral, and positive categories then will implementation Naive Bayes and Support Vector Machine (SVM) algorithms, aiming to determine which algorithm provides the best accuracy in sentiment classification. The study successfully collected a dataset of 3367 entries. It classified them into three different categories: positive sentiment with 1597 entries, negative sentiment with 386 entries, and neutral sentiment with 1384 entries, with the help of the TextBlob library. The testing results using the Naive Bayes algorithm obtained an accuracy rate of 82.81%, while the SVM algorithm achieved an accuracy rate of 80.42%. This indicates that most of the public are satisfied with the performance of the Indonesian National Team players in the 2024 U-23 Asian Cup, while some are less satisfied. The implementation of Naive Bayes and SVM algorithms was able to provide excellent and accurate sentiment classification. This analysis is expected to provide a deeper understanding of public perception towards the team's performance in the competition, as well as serve as an evaluation and development material to improve the quality of the Indonesian National Team's performance in the future.

Keywords—football sports; indonesian national team; naive bayes; sentiment analysis; support vector machine

1 INTRODUCTION

The dynamic evolution in the information technology sector has resulted in beneficial effects on how people acquire, organize, and distribute information efficiently and in a timely manner [1]. Social media has become an effective form of digital communication media that is highly favored by the Indonesian public for obtaining and sharing information [2]. Currently, many people use social media in their daily lives as a platform to share views on various topics, making it a precious source of data [3]. Therefore, effective data processing and documentation are essential to produce information that can be used in the decision-making process [4]. Twitter has become a favorite social media platform among Indonesians for sharing views and information on various topics [5] where users can connect with family, friends, and loved ones through the use of computers or mobile phones [6]. Twitter also facilitates users with features like hashtags and trending topics [7] allowing people to participate in various conversations and events that are currently being widely discussed [8] making it considered a social network that provides broader and more accurate data compared to other platforms [9].

Football is a popular topic for discussion and receives many opinions from the public [10]. Indonesia is also known as a country in Asia with very high enthusiasm for football [11]. This indicates that football has become an essential element for the Indonesian people, creating great expectations for its development in the future [12]. International football competitions are highly anticipated events by the Indonesian public [13]. The participation of the Indonesian National Team in international sports competitions consistently garners significant attention from the public. It has become a hot topic of discussion, especially on social media platforms like Twitter [14]. One of the prestigious football competitions eagerly awaited by Indonesians is the 2024 U-23 Asian Cup, where the Indonesian National Team is also participating. Various opinions and sentiments from the public emerge on social media as the Indonesian National Team progresses in this competition.

Manual data management on social media requires significant effort and time, thus necessitating a classification model that can automatically categorize these opinions into sentiment labels [15]. Machine Learning can extract knowledge from abundant data sources and make decisions or predictions without direct user intervention, making it an effective tool for automating data processing [16]. Sentiment analysis is a field of development within Machine Learning commonly used to understand public views on specific issues or events through social media [17] by extracting valuable information from unstructured data [18]. The results of data processing with sentiment analysis techniques can produce information in the form of public opinions classified into positive, neutral, or negative sentiment categories [19].

Several previous studies have focused on the topic of sentiment analysis in football. For example, the study [20] analyzed sentiments towards football news on Twitter using the Support Vector Machine (SVM) algorithm. This study

obtained 509 data points classified into three different categories. It achieved an accuracy rate of up to 91%, demonstrating the success of applying the SVM algorithm in assisting the sentiment classification process of public opinion. Another study [21] on sentiment analysis regarding changes in football policy on Twitter used the Naive Bayes algorithm, where 1800 data points were grouped into two different categories and achieved an accuracy rate of up to 96%, also showing the effectiveness of this algorithm in sentiment classification of public opinion. This research aims to understand public perception on Twitter regarding the involvement of the Indonesian National Team in the 2024 U-23 Asian Cup by automatically categorizing these opinions into negative, neutral, and positive classes using the TextBlob library. Furthermore, this study will compare the effectiveness of two methods, the Support Vector Machine (SVM) and Naive Bayes algorithms, by determining which algorithm has the best accuracy in sentiment classification.

2 METHOD

The researchers conducted a series of steps in the data processing procedure, as illustrated in the diagram provided in Figure 1. These steps outline the process that must be followed to achieve the desired results in the data analysis.

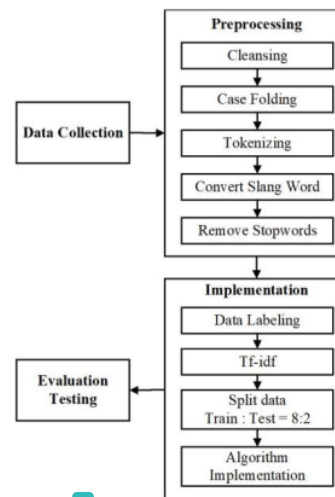


Figure 1. Stages of Research.

2.1 Data Collection

In this step, the researchers accessed a dataset from the social media platform Twitter regarding public opinions on the participation of the Indonesian National Team in the Asian Cup U-23 2024 competition using the Python programming language with the `twitter_auth_token`. The keyword phrase used was "timnas Indonesia u23". Data was collected over the period from April 15 to May 3, 2024, with texts in Indonesian, resulting in 3746 tweets, as shown in Figure 2. The dataset was saved in CSV file format.



conversation_id_str	created_at	favorite_count	full_text	id_str
0	1786183300731818120 Thu May 02 23:57:39 +0000 2024	0	hasil skor 2-1 yang memandakan Irak memenangkan ...	1786183300731818120
1	1786179651734647067 Thu May 02 23:43:56 +0000 2024	0	Timnas Indonesia U23 Ditebak Irak 1-2 Tiket OK...	1786179651734647067
2	1786176654093746201 Thu May 02 23:31:14 +0000 2024	1	Timnas U23 Indonesia kalah dari Irak pada laga...	1786176654093746201
3	1786176342608007563 Thu May 02 23:30:06 +0000 2024	0	Link siaran ulang Indonesia vs Irak dalam laga...	1786176342608007563
4	1786170161114995774 Thu May 02 23:05:26 +0000 2024	0	Sejumlah pemain Guinea U23 calon timnas TI...	1786170161114995774
...
3744	177966025216353200 Mon Apr 15 03:33:06 +0000 2024	3	Timnas U23 Indonesia vs Qatar AFC 2024 Live TV...	177966025216353200
3745	1779660114211180706 Mon Apr 15 03:20:40 +0000 2024	0	Nathan Tjoe-Aon Gabung Timnas Indonesia...	1779660114211180706
3746	177966036587200818 Mon Apr 15 03:20:22 +0000 2024	0	PRE-ANNOUNCEMENT: Nathan Tjoe-Aon Resmi Gabung...	177966036587200818

3747 rows x 5 columns

Figure 2. Results of Data Retrieval.

The next step is to remove unnecessary columns from the table and retain only the columns containing public opinions. Table 1 presents examples of tweets to provide a clearer picture of the variation in expressions and sentiment analysis used in this research.

Table 1. Several Opinions from Twitter

Tweet
Semangat Timnas Indonesia U23 #TimNasDay jelek sekali permainan Timnas Indonesia U23 malam ini. Sangat kecewa 🙄.
Terima kasih untuk Timnas Indonesia U23 perjuangan kalian sampai sejauh ini patut diapresiasi 🙌 Tetap semangat untuk berjuang di laga selanjutnya ya 🙏 @KompasBola #PlayoffOlimpiade #OlimpiadeParis2024 #TimnasIndonesia #KitaGaruda #TimnasDay #Guinea #ErickThohir
RESMI : Nathan Tjoe telah mendapatkan izin dari Heerenveen untuk kembali memperkuat timnas Indonesia di ajang Piala Asia U23. #timnasindonesiau23 #shintayong #pelatih #timnasindonesia https://www.viva.co.id/bola/liga-indonesia/1708088-alasan-heerenveen-izinkan-nathan-tjoe-a-on-kembali-ke-timnas-indonesia-u-23
RT@BolaBolaAja Pertama kali gw lihat Timnas Indonesia U23 semalam main dan jelek banget mainnya, gak becus, menang euforia doang di medsos sama akun2 alay.



This article is distributed under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/). See for details: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Tweet

Indonesia tidak akan lolos ke Olimpiade 🙄.

2.2 Preprocessing Data

The next step is preprocessing the CSV dataset. Before entering this stage, the dataset with similar or duplicate opinion values must be removed first, resulting in a total dataset of 3367 entries for the subsequent processes. Data preprocessing is the process of transforming unstructured text into a more structured form, which will aid in data processing[22]. Several steps in data preprocessing include cleansing, case folding, tokenizing, converting slang words, and removing stopwords.

2.2.1 Cleansing: The cleansing process is an essential step in data preprocessing. Cleansing aims to remove all URL links, symbols, mentions, retweets, hashtags, and emoticons from the dataset [23]. This step is necessary to ensure that the data used for further analysis or processing is clean and free from irrelevant or distracting characters [24]. By performing cleansing, the dataset becomes cleaner and ready for the next preprocessing steps, where the text results have already passed through the cleansing stage, as shown in Table 2.

Table 2. The Outcome of Data Cleansing

Before	After
Semangat Timnas Indonesia U23 #TimNasDay	Semangat Timnas Indonesia
jelek sekali permainan Timnas Indonesia U23 malam ini. Sangat kecewa 🙄.	jelek sekali permainan Timnas Indonesia malam ini Sangat kecewa
Terima kasih untuk Timnas Indonesia U23 perjuangan kalian sampai sejauh ini patut diapresiasi 🙌 Tetap semangat untuk berjuang di laga selanjutnya ya 🙏 @KompasBola #PlayoffOlimpiade #OlimpiadeParis2024 #TimnasIndonesia #KitaGaruda #TimnasDay #Guinea #ErickThohir	Terima kasih untuk Timnas Indonesia perjuangan kalian sampai sejauh ini patut diapresiasi Tetap semangat untuk berjuang di laga selanjutnya ya
RESMI : Nathan Tjoe telah mendapatkan izin dari Heerenveen untuk kembali memperkuat timnas Indonesia di ajang Piala Asia U23. #timnasindonesiau23 #shintayong #pelatih #timnasindonesia https://www.viva.co.id/bola/liga-indonesia/1708088-alasan-heerenveen-izinkan-nathan-tjoe-a-on-kembali-ke-timnas-	RESMI Nathan Tjoe telah mendapatkan izin dari Heerenveen untuk kembali memperkuat timnas Indonesia di ajang Piala Asia

Before	After
indonesia-u-23	
RT@BolaBolaAja Pertama kali gw lihat Timnas Indonesia U23 semalam main dan jelek banget mainnya, gak becus, menang euforia doang di medsos sama akun2 alay. Indonesia tidak akan lolos ke Olimpiade 🙄.	Pertama kali gw lihat Timnas Indonesia semalam main dan jelek banget mainnya gak becus menang euforia doang di medsos sama akun alay Indonesia tidak akan lolos ke Olimpiade

2.2.2 *Case Folding*: The next step is case folding, which is the process of converting all uppercase letters in the dataset to lowercase letters with the aim of standardizing the text because uppercase letters can affect subsequent analysis processes [25]. The text results that have undergone the case folding stage are shown in Table 3.

Table 3. The Outcome of Case Folding

Before	After
Semangat Timnas Indonesia	semangat timnas indonesia
jelek sekali permainan Timnas Indonesia malam ini Sangat kecewa	jelek sekali permainan timnas indonesia malam ini sangat kecewa
Terima kasih untuk Timnas Indonesia perjuangan kalian sampai sejauh ini patut diapresiasi Tetap semangat untuk berjuang di laga selanjutnya ya	terima kasih untuk timnas indonesia perjuangan kalian sampai sejauh ini patut diapresiasi tetap semangat untuk berjuang di laga selanjutnya ya
RESMI Nathan Tjoe telah mendapatkan izin dari Heerenveen untuk kembali memperkuat timnas Indonesia di ajang Piala Asia	resmi nathan tjoe telah mendapatkan izin dari heerenveen untuk kembali memperkuat timnas indonesia di ajang piala asia
Pertama kali gw lihat Timnas Indonesia semalam main dan jelek banget mainnya gak becus menang euforia doang di medsos sama akun alay Indonesia tidak akan lolos ke Olimpiade	pertama kali gw lihat timnas indonesia semalam main dan jelek banget mainnya gak becus menang euforia doang di medsos sama akun alay indonesia tidak akan lolos ke olimpiade

2.2.3 *Tokenizing*: The next stage in data preprocessing is the tokenization process, which involves splitting a sentence into several words or tokens [26] allowing for more effective processing of individual words and facilitating the subsequent preprocessing stages such as Convert Slang Word and Remove Stopwords. The text results that have undergone the tokenization stage are as shown in Table 4.

Table 4. The Outcome of Tokenizing

Before	After
semangat timnas indonesia	semangat, timnas, indonesia
jelek sekali permainan timnas indonesia malam ini sangat kecewa	jelek, sekali, permainan, timnas, indonesia, malam, ini, sangat, kecewa
terima kasih untuk timnas indonesia perjuangan kalian sampai sejauh ini patut diapresiasi tetap semangat untuk berjuang di laga selanjutnya ya	terima, kasih, untuk, timnas, indonesia, perjuangan, kalian, sampai, sejauh, ini, patut, diapresiasi, tetap, semangat, untuk, berjuang, di, laga, selanjutnya, ya
resmi nathan tjoe telah mendapatkan izin dari heerenveen untuk kembali memperkuat timnas indonesia di ajang piala asia	resmi, nathan, tjoe, telah, mendapatkan, izin, dari, heerenveen, untuk, kembali, memperkuat, timnas, indonesia, di, ajang, piala, asia
pertama kali gw lihat timnas indonesia semalam main dan jelek banget mainnya gak becus menang euforia doang di medsos sama akun alay indonesia tidak akan lolos ke olimpiade	pertama, kali, gw, lihat, timnas, indonesia, semalam, main, dan, jelek, banget, mainnya, gak, becus, menang, euforia, doang, di, medsos, sama, akun, alay, indonesia, tidak, akan, lolos, ke, olimpiade

2.2.4 *Converting slang words*: The following process is converting slang words, which involves removing all affixes from words to obtain their base form in the dataset [27]. The Sastrawi library facilitates further analysis processes. The text results that have undergone the convert slang word stage are shown in Table 5.

Table 5. The Outcome of Converting Slang Words

Before	After
semangat, timnas, indonesia	semangat, timnas, indonesia
jelek, sekali, permainan, timnas, indonesia, malam, ini, sangat, kecewa	jelek, sekali, main, timnas, indonesia, malam, ini, sangat, kecewa
terima, kasih, untuk, timnas, indonesia, perjuangan, kalian, sampai, sejauh, ini, patut, diapresiasi, tetap, semangat, untuk, berjuang, di, laga, selanjutnya, ya	terima, kasih, untuk, timnas, indonesia, juang, kalian, sampai, jauh, ini, patut, apresiasi, tetap, semangat, untuk, juang, di, laga, lanjut, ya
resmi, nathan, tjoe, telah, mendapatkan, izin, dari, heerenveen, untuk, kembali, memperkuat, timnas, indonesia, di, ajang, piala, asia	resmi, nathan, tjoe, telah, dapat, izin, dari, heerenveen, untuk, kembali, kuat, timnas, indonesia, di, ajang, piala, asia
pertama, kali, gw, lihat,	pertama, kali, saya, lihat,



Before	After
terima kasih timnas indonesia juang patut apresiasi semangat juang laga lanjut	thank you indonesian national team deserves appreciation for their fighting spirit in the next match
resmi nathan tjoie dapat izin heerenveen kembali kuat timnas indonesia ajang piala asia	nathan tjoie officially gets permission for heerenveen to return strong for the indonesian national team in the asian cup
lihat timnas indonesia malam main jelek banget main tidak becus menang euforia medsos akun alay indonesia tidak lolos olimpiade	saw the indonesian national team playing really bad tonight didn't they play because they won the euphoria of social media the indonesian alay account didn't qualify for the olympics

10 The next step is to label the data using the TextBlob library. TextBlob is a Python programming library that facilitates the automatic annotation of text data. This is achieved by computing the polarity values of the words within the text [30]. These polarity values can yield insights into the sentiment or emotional content present within the text. Table 8 below shows the dataset before and after the labeling process.

Table 8. The Outcome of Data Labeling

Text	Label
spirit of the indonesian national team	Positive
bad to play the indonesian national team at night disappointed	Negative
thank you indonesian national team deserves appreciation for their fighting spirit in the next match	Positive
nathan tjoie officially gets permission for heerenveen to return strong for the indonesian national team in the asian cup	Neutral
saw the indonesian national team playing really bad tonight didn't they play because they won the euphoria of social media the indonesian alay account didn't qualify for the olympics	Negative

Using the TextBlob library, each text in the dataset has been assigned a sentiment label based on its analysis of the words contained within it. This allows for further analysis of the sentiment of these texts, such as evaluating customer feedback or understanding public opinion on a product or service. To facilitate researchers, the results of sentiment classification that have been conducted can be visualized in the form of a pie chart, as shown in Figure 5 below.

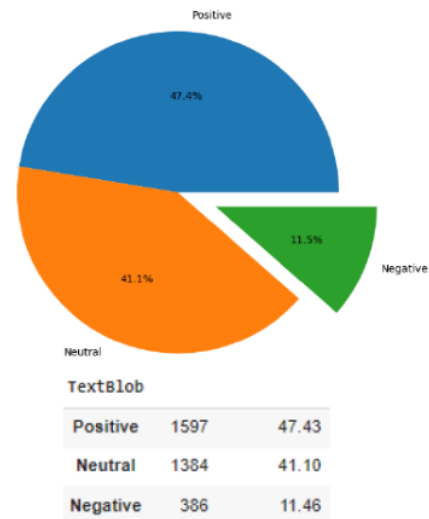


Figure 5. Visualization of Word Frequency in Bar Chart.

This visualization helps researchers quickly see the distribution of sentiment within the dataset graphically, making it easier to analyze and understand the overall sentiment classification results. By looking at the diagram, it's easier to identify the majority sentiment as positive, negative, or neutral and gain additional insights into any patterns that may exist in the data. Based on the classification results using the TextBlob library, there are 1597 public opinions categorized as positive, representing 47.43% of the total. Furthermore, there are 1384 opinions considered neutral, covering 41.10% of the overall data. Meanwhile, negative opinions amount to 386, with a proportion of 11.46%. Therefore, the majority of public opinions in the dataset tend to be positive, followed by neutral opinions, with a smaller proportion being negative.

3.2 TF-IDF Weighting

TF-IDF weighting aims to assign values to data based on the frequency of word occurrence in a document. The frequency of each word in the document is calculated and then given a weight based on the frequency of occurrence of that word relative to the total number of documents containing that word [31]. This process utilizes the 'CountVectorizer' library in the Python programming language to convert text into metric form or numerical values.

3.3 Data Splitting into Training and Testing

21 After the data weighting process using the TF-IDF method, the next step is to split the data into two groups: the training data group and the testing data group. This process is an essential step in preparation for further analysis. In this study, the data splitting process is done with a particular proportion, where 80% of the entire dataset is allocated as training data, and the remaining 20% is used as testing data.



3.4 Algorithm Implementation

The evaluation stage is a crucial component in every system development process, aiming to assess the success level based on specific metrics. In this study, the researchers applied the Naive Bayes and Support Vector Machine algorithms to the training and testing data in the dataset. The next step is for the researchers to process the data using a confusion matrix on the 3367 data points that have been split into training and testing data from the application of the Naive Bayes algorithm. The visualization results of the confusion matrix process can be seen in Figure 6 below.

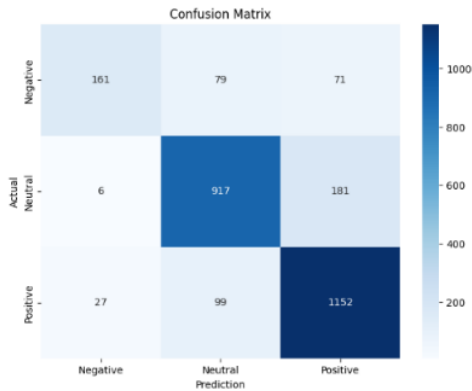


Figure 6. Confusion Matrix of Naive Bayes Implementation Result.

The following table 8 explains the description provided in the above Figure 6 regarding the results of the confusion matrix obtained by applying the Naive Bayes algorithm.

Table 9. Results from the Naive Bayes Method's Confusion Matrix

Description	Value Data
True Positive (TPos)	1152
False Neutral Positive (FNetPos)	181
False Negative Positive (FNegPos)	71
False Positive Neutral (FPosNet)	99
True Neutral (TNET)	917
False Negative Neutral (FNegNet)	79
False Positive Negative (FPosNeg)	27
False Neutral Negative (FNetNeg)	6
True Negative (TNeg)	161

The next step is to conduct the accuracy testing process for the implementation of the Naive Bayes algorithm that has been performed, where the result of this process yields an accuracy value of 82.81% as shown in the following Figure 7.

```
#Accuracy Score Naive Bayes Algorithm
accuracy = accuracy_score(y_test, predictions)
print("Accuracy score NaiveBayes: {:.2f}%".format(accuracy*100))

Accuracy score NaiveBayes: 82.81%
```

Figure 7. Confusion Matrix of Naive Bayes Implementation Result.

Next step the data will be processed using a confusion matrix on the 3367 data points that have been divided into training and testing data from the application of the SVM algorithm. The visualization results of the confusion matrix process can be seen in Figure 8 below.

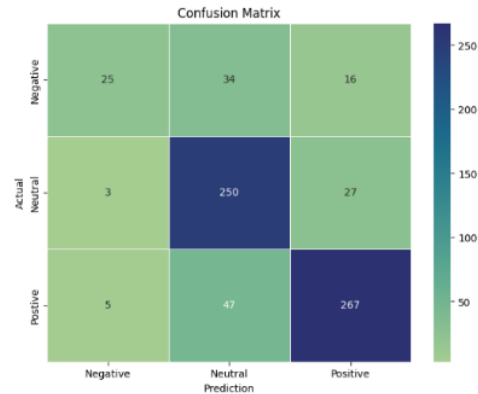


Figure 8. Confusion Matrix of Naive Bayes Implementation Result.

The description provided in the Figure 8 above regarding the results of the confusion matrix obtained by applying the SVM algorithm is explained in the following table 9.

Table 10. Results from the SVM Method's Confusion Matrix

Description	Value Data
True Positive (TPos)	1152
False Neutral Positive (FNetPos)	181
False Negative Positive (FNegPos)	71
False Positive Neutral (FPosNet)	99
True Neutral (TNET)	917
False Negative Neutral (FNegNet)	79
False Positive Negative (FPosNeg)	27
False Neutral Negative (FNetNeg)	6
True Negative (TNeg)	161

The next step is to conduct the accuracy testing process for the implementation of the SVM algorithm that has been performed. The result of this process yields an accuracy value of 80.04%, as shown in the following Figure 9.




```
#Accuracy Score SVM Algorithm
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy score svm: {:.2f}%".format(accuracy*100))
```

Accuracy score svm: 80.42%

Figure 9. Confusion Matrix of SVM Implementation Result.

3.5 Evaluation Testing

In this study [11], researcher have implemented two distinct algorithms: Naive Bayes and Support Vector Machine (SVM). Each algorithm offers unique and effective methods for processing data. The Naive Bayes algorithm is rooted in Bayes' theorem and provides a probability-based approach to classification [32]. On the other hand, SVM is a robust technique that operates by seeking the optimal hyperplane to separate data classes [33]. Evaluation testing is a critical step in the process of developing and assessing models in machine learning. At this stage various evaluation metrics such as F1-score, recall, and precision are employed to assess the performance of the implemented model. The following table 11 shows the results of the F1-score, recall, and precision values of the Naive Bayes algorithm implementation and the following table 12 shows the results of the F1-score, recall, and precision values of the SVM algorithm implementation in this study.

Table 11. Evaluation Testing of Naive Bayes Implementation Result

	Precision	Recall	F1-score	support
Negative	0.83	0.52	0.64	311
Neutral	0.84	0.83	0.83	1104
Positive	0.82	0.90	0.86	1278
Accuracy			0.83	2693
Macro avg	0.83	0.75	0.78	2693
Weighted avg	0.83	0.83	0.82	2693

In the table 11 above, the evaluation results show accuracy values for negative sentiment with an F1-score of 64%, a recall of 52%, and a precision of 83%. For neutral sentiment, the F1-score reaches 83%, recall is 83%, and precision is 84%. Meanwhile, for positive sentiment, there is an F1-score of 86%, recall of 90%, and precision of 82%.

Table 12. Evaluation Testing of SVM Implementation Result

	Precision	Recall	F1-score	support
Negative	0.76	0.33	0.46	75
Neutral	0.76	0.89	0.82	280
Positive	0.86	0.84	0.85	319
Accuracy			0.80	674
Macro avg	0.79	0.69	0.71	674
Weighted avg	0.81	0.80	0.79	674

In the table 12 above, the evaluation results show accuracy values for negative sentiment with an F1-score of 46%, a recall of 33%, and a precision of 76%. For neutral sentiment, the F1-score reaches 82%, recall is 89%, and

precision is 76%. Meanwhile, for positive sentiment, there is an F1-score of 85%, recall of 89%, and precision of 86%.

4 CONCLUSION

Out of the total data collected from the Twitter social media platform regarding public sentiment towards the participation of the Indonesian national team in the Asian Cup U23 football competition, only 3367 unique or non-duplicate data were obtained and used in this study. The sentiment classification process on the 3367 datasets using the TextBlob library resulted in 1597 data classified as positive sentiment, 1384 as neutral sentiment, and 386 as negative sentiment. These data indicate that the majority of public opinions on Twitter regarding the participation of the Indonesian National Team in the Asian Cup U-23 are either positive or neutral. This reflects significant support from the Indonesian community, although there are still some negative opinions. The use of the Naive Bayes algorithm on 3367 tweet data yielded an accuracy score of 82.81%, while the SVM algorithm resulted in an accuracy score of 80.42%. In the evaluation stage, the Naive Bayes algorithm obtained an accuracy values for negative sentiment with an F1-score of 64%, a recall of 52%, and a precision of 83%. For neutral sentiment, the F1-score reaches 83%, recall is 83%, and precision is 84%. Meanwhile, for positive sentiment, there is an F1-score of 86%, recall of 90%, and precision of 82%. In the SVM algorithm obtained an accuracy values for negative sentiment with an F1-score of 46%, a recall of 33%, and a precision of 76%. For neutral sentiment, the F1-score reaches 82%, recall is 89%, and precision is 76%. Meanwhile, for positive sentiment, there is an F1-score of 85%, recall of 89%, and precision of 86%. This indicates that the implementation of both algorithms can classify sentiment with a relatively high level of accuracy. The outcomes of the dataset processing that has been conducted can be applied as a reference in the decision-making process, or as material for evaluation and development to enhance the performance quality of the Indonesian National Team in the forthcoming period.

AUTHOR'S CONTRIBUTION

Sewin Fathurrohman conceptualization of the research, data collection and processing, implementation of the Naive Bayes and SVM algorithms, and writing and revising the draft article. Irfan Ricky Afandi assisted in data collection and preprocessing, wrote the introduction, conducted a literature review, and composed the research conclusion. Firman Noor Hasan validated the model and conducted a statistical analysis of the classification results, structured the methodology and results, and assisted in manuscript editing.

COMPETING INTERESTS

In accordance with the publication ethics of this journal Sewin Fathurrohman, Irfan Ricky Afandi, Firman Noor Hasan as the authors of this article declare that this article is free from conflict of interest (COI) and conflict of interest (CI).



ACKNOWLEDGMENT (HEADING 5)

The authors would like to express sincere gratitude to all parties who have provided support and contributions to this research. We deeply appreciate the assistance provided by the institutions where we are affiliated, as well as our fellow researchers who have offered valuable advice and input throughout the research process. We also extend our thanks to the Twitter users who participated by providing their opinions, and to those who granted access to the necessary data. Without the contributions and support from all these parties, this research would not have been possible.

REFERENCES

- [1] I. R. Afandi, N. Pratiwi, A. A. Rizki, M. Irva, and M. F. Aulia, "Perancangan Sistem Informasi Pelayanan Pembuatan Surat Online Di Desa Ciangsana Bebas Website," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 571–577, 2022, doi: 10.36040/jati.v6i2.5318.
- [2] O. B. Islamuddin and I. Yuadi, "ANALISIS TEXT CLUSTERING PADA DATA MINING MENGGUNAKAN METODE K- NEAREST NEIGHBOR (KNN) DAN DECISION TREE," *J. KHATULISTIWA Inform.*, vol. 11, no. 2, pp. 128–134, 2023, doi: https://doi.org/10.31294/jki.v11i2.16072.g6323.
- [3] I. M. Noor and M. Turan, "Sentiment Analysis on New Currency in Kenya using Twitter Dataset," *IJID (International J. Informatics Dev.)*, vol. 8, no. 2, p. 81, 2020, doi: 10.14421/ijid.2019.08206.
- [4] P. Afikah, I. R. Afandi, and F. N. Hasan, "Implementasi Business Intelligence Untuk Menganalisis Data Kasus Virus Corona di Indonesia Menggunakan Platform Tableau," *Pseudocode*, vol. 9, no. 1, pp. 25–32, 2022, doi: 10.33369/pseudocode.9.1.25-32.
- [5] J. Annissa and A. Ariesta, "Analisis Semiotika Makna Nasionalisme melalui Text Mining pada Media Sosial Twitter di Kejuaraan AFF Tahun 2020," *J. IPTEKKOM J. Ilmu Pengetah. Teknol. Inf.*, vol. 25, no. 1, pp. 69–84, 2023, doi: 10.17933/iptekkom.25.1.2023.69-84.
- [6] R. Sulastiyono, A. Setiawan, and S. Nugroho, "Sentimen Analisis Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 Menggunakan Metode Naive Bayes," *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1387–1394, 2023, doi: 10.47065/josh.v4i4.3737.
- [7] A. D. Dayani, Yuhandri, and W. G. Nurcahyo, "Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *J. KomtekInfo*, vol. 11, no. 1, pp. 1–10, 2024, doi: 10.35134/komtekinfo.v11i1.439.
- [8] L. Lestari, M. D. R. Wahyudi, and U. Kiftiyani, "Veil and Hijab: Twitter Sentiment Analysis Perspective," *IJID (International J. Informatics Dev.)*, vol. 9, no. 1, p. 52, 2020, doi: 10.14421/ijid.2020.09108.
- [9] T. Y. Widiyanti, T. B. Adji, and I. Hidayah, "The Social Engagement to Agricultural Issues using Social Network Analysis," *IJID (International J. Informatics Dev.)*, vol. 10, no. 1, pp. 1–7, 2021, doi: 10.14421/ijid.2021.2185.
- [10] E. R. Setyaningsih, "Sentiment Classification untuk Opini Berita Sepak Bola," *J. Intell. Syst. Comput.*, vol. 3, no. 2, pp. 93–98, 2021, doi: 10.52985/insyst.v3i2.193.
- [11] M. Habibi, D. Handoko, D. Kurniawan, Rasman, and R. Anggrani, "Analisis Framing Robert Entman Pemberitaan Tragedi Kanjuruhan di Media Asing," *Perspekt. Komun. J. Ilmu Komun. Polit. dan Komun. Bisnis*, vol. 7, no. 1, pp. 43–64, 2023, doi: https://doi.org/10.24853/pk.7.1.43-64.
- [12] A. Damayanti and A. K. Maulidiyah, "Analisa Sentimen Pada Pembatalan Piala Dunia U20 di Indonesia Menggunakan Metode Support Vector Machine," *JITSI J. Ilm. Ter. Sains dan Teknol.*, vol. 1, no. 2, pp. 97–103, 2023, doi: 10.25139/jitsi.v1i2.6607.
- [13] M. Kholilullah, M. Martanto, and U. Hayati, "Analisis Sentimen Pengguna Twitter(X) Tentang Piala Dunia Usia 17 Menggunakan Metode Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 392–398, 2024, doi: 10.36040/jati.v8i1.8378.
- [14] R. P. Ratnaningtyas and Y. A. Muhammad, "Analisis Pemberitaan Timnas Indonesia pada Media Daring," *MUKASI J. Ilmu Komun.*, vol. 2, no. 1, pp. 45–52, 2023, doi: 10.54259/mukasi.v2i1.1492.
- [15] S. Rahmawati and M. Habibi, "Public Sentiments Analysis about Indonesian Social Insurance Administration Organization on Twitter," *IJID (International J. Informatics Dev.)*, vol. 9, no. 2, pp. 87–93, 2020, doi: 10.14421/ijid.2020.09205.
- [16] R. Pebrianto, S. N. Nugraha, A. Latif, and M. R. Firdaus, "Analisis Sentimen Twitter Terhadap Menteri Indonesia Dengan Algoritma Support Vector Machine Dan Naive Bayes," *E-Link J. Tek. Elektro dan Inform.*, vol. 17, no. 1, p. 1, 2022, doi: 10.30587/e-link.v17i1.3965.
- [17] F. Maulana, M. A. Abdullah, J. Sari, D. Z. Siddik, M. Agustinus, and D. D. Saputra, "Sentiment Analysis on the Twitter Pssi Performance Using Text Mining With the Naive Bayes Algorithm," *J. Comput. Inf. Syst.*, vol. 18, no. 2, pp. 211–216, 2022, doi: 10.33480/pilar.v18i2.3938.
- [18] M. A. Hanafi and A. Solichin, "Analisis Sentimen Terhadap Pssi Atas Tragedi Kanjuruhan Menggunakan Multinomial Naive Bayes," *Telemat. MKOM*, vol. 15, no. 1, p. 21, 2023, doi: 10.36080/telematikamkom.2292.
- [19] A. E. Nanada, A. N. Sihananto, and A. M. Rizki, "Analisis Sentimen Pada Pembatalan Tuan Rumah Indonesia Di Piala Dunia U-20 Menggunakan Fasttext Embeddings Dan Algoritma Recurrent Neural Network," *J. Tek. Inform. Sains dan Ilmu Komun.*, vol. 2, no. 2, 2024, doi: https://doi.org/10.59841/saber.v2i2.1000.
- [20] N. Dalifah, N. Suarna, and W. Prihartono, "Analisis Data Sentimen Negatif Pada Opini Pengguna Twitter Terhadap Berita Sepak Bola Liga 1 Tahun 2022 Dengan Penerapan Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 209–214, 2024, doi: 10.36040/jati.v8i1.8303.
- [21] E. L. Rara and E. Mailoa, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Perubahan Piala Dunia U-20," *Progresif J. Ilm. Komput.*, vol. 20, no. 1, pp. 259–268, 2024, doi: https://doi.org/10.35889/progresif.v20i1.1550.
- [22] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 375–384, 2024, doi: 10.57152/malcom.v4i2.1206.
- [23] A. T. Susilawati, N. A. Lestari, and P. A. Nina, "Analisis Sentimen Publik Pada Twitter Terhadap Boikot Produk Israel Menggunakan Metode Naive Bayes," *J. Ilm. Mhs.*, vol. 2, no. 1, pp. 26–35, 2024, [Online]. Available: https://doi.org/10.59603/niantanasikka.v2i1.240
- [24] I. R. Afandi, F. Noor Hasan, A. A. Rizki, N. Pratiwi, and Z. Halim, "Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes," *J. Linguist. Komputasional*, vol. 5, no. 2, pp. 63–70, 2022, doi: https://doi.org/10.26418/jlk.v5i2.107.
- [25] D. A. Nugroho and F. N. Hasan, "Analisis Sentimen Kegiatan Pembersihan Sampah Pada Media Sosial X Menggunakan SVM dan Naive Bayes," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. April, pp. 843–853, 2024, doi: 10.30865/mib.v8i2.7562.
- [26] M. Samantri and Afiyati, "Perbandingan Algoritma Support Vector Machine dan Random Forest untuk Analisis Sentimen Terhadap Kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 8, no. 1, pp. 1–9, 2024, doi:



DOI:

IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448

- 10.35870/jtik.v8i1.1202.
- [27] G. Radiena and A. Nugroho, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine," *J. Pendidik. Teknol. Inf.*, vol. 6, no. 1, pp. 1–10, 2023, doi: 10.37792/jukanti.v6i1.836.
- [28] Fauzan Baehaqi and N. Cahyono, "Analisis Sentimen Terhadap Cyberbullying Pada Komentar Di Instagram Menggunakan Algoritma Naïve Bayes," *Indones. J. Comput. Sci.*, vol. 13, no. 1, pp. 1051–1063, 2024, doi: 10.33022/ijcs.v13i1.3301.
- [29] I. A. Ricky, I. F. Hanif, F. N. Hasan, E. S. Sinduningrum, Z. Halim, and N. Nunik, "Analisis Sentimen Opini Masyarakat Terkait Penyelenggaraan Sistem Elektronik Menggunakan Metode Logistic Regression," *J. Linguist. Komputasional*, vol. 5, no. 2, p. 77, 2022, [Online]. Available: <https://t.co/23c4krbjp>
- [30] N. L. Putri, B. Warsito, and B. Surarso, "Pengaruh Klasifikasi Sentimen Pada Ulasan Produk Amazon Berbasis Rekayasa Fitur dan K-Nearest Neighbor," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 65–74, 2024, doi: 10.25126/jtiik.20241117376.
- [31] P. W. Hardjita, Nurochman, and R. Hidayat, "Sentiment Analysis of Tweets on Prakerja Card using Convolutional Neural Network and Naive Bayes," *IJID (International J. Informatics Dev.*, vol. 10, no. 2, pp. 82–91, 2022, doi: 10.14421/ijid.2021.3007.
- [32] A. Karimah, G. Dwilestari, and Mulyawan, "Analisis Sentimen Komentar Video Mobil Listrik Di Platform YOUTUBE DENGAN METODE NAIVE BAYES," *J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 767–773, 2024, doi: <https://doi.org/10.36040/jati.v8i1.8373>.
- [33] L. A. Susanto, "Komparasi Model Support Vector Machine Dan K-Nearest Neighbor Pada Analisis Sentimen Aplikasi Polri Super App," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4152.



This article is distributed under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/). See for details: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

ORIGINALITY REPORT

11%

SIMILARITY INDEX

9%

INTERNET SOURCES

5%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1	www.ruetir.com Internet Source	2%
2	thesis.eur.nl Internet Source	1%
3	ejurnal.seminar-id.com Internet Source	1%
4	repository.uhamka.ac.id Internet Source	1%
5	Tiara Sari Ningsih, Teguh Iman Hermanto, Imam Ma'ruf Nugroho. "Sentiment Analysis of Mobile Provider Application Reviews Using Naive Bayes Algorithm and Support Vector Machine", Sinkron, 2024 Publication	1%
6	Asrul Sani, Samuel, Djaka Suryadi, Firman Noor Hasan, Ade Davy Wiranata, Siti Aisyah. "Predicting the Success of Garment Sales on Transaction Data using the Classification Method with the Naïve Bayes Algorithm", 2023 International Conference on Computer	1%

Science, Information Technology and Engineering (ICCoSITE), 2023

Publication

7	Submitted to SDM Universitas Gadjah Mada Student Paper	<1 %
8	discovery.researcher.life Internet Source	<1 %
9	projectswiki.eleceng.adelaide.edu.au Internet Source	<1 %
10	inacl.id Internet Source	<1 %
11	www.science.gov Internet Source	<1 %
12	garuda.kemdikbud.go.id Internet Source	<1 %
13	Y-C Chang. "Shape optimization of mufflers hybridized with multiple connected tubes using the boundary element method, neural networks, and the genetic algorithm", Proceedings of the Institution of Mechanical Engineers Part C Journal of Mechanical Engineering Science, 01/01/2010 Publication	<1 %
14	Raúl Felipe Parra Aramburo, Miguel Ângelo Lellis Moreira, Luiz Paulo Lopes Fávero, Igor Pinheiro de Araújo Costa et al. "Data Science	<1 %

in Social Politics with Particular Emphasis on Sentiment Analysis", Procedia Computer Science, 2022

Publication

15

Tetrian Widyanto, Ina Ristiana, Arief Wibowo. "Komparasi Naïve Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter", SINTECH (Science and Information Technology) Journal, 2023

Publication

16

www.breakinglatest.news

Internet Source

17

Maulana Rizki Hidayat, Mulia Sulistiyono. "Comparison of Accuracy and Time Of Naïve Bayes Algorithm with Support Vector Machine Algorithm in Twitter Sentiment Analysis of Peduli Lindungi Application", 2022 5th International Conference on Information and Communications Technology (ICOIACT), 2022

Publication

18

Submitted to ITESM: Instituto Tecnológico y de Estudios Superiores de Monterrey

Student Paper

19

Rizky Afrinanda, Lusiana Efrizoni, Wirta Agustin, Rahmiati Rahmiati. "Hybrid Model for Sentiment Analysis of Bitcoin Prices using Deep Learning Algorithm", MATRIK : Jurnal

<1 %

<1 %

<1 %

<1 %

<1 %

Manajemen, Teknik Informatika dan Rekayasa Komputer, 2023

Publication

20	ejournal.uin-suka.ac.id Internet Source	<1 %
21	journal.universitaspahlawan.ac.id Internet Source	<1 %
22	journal.yp3a.org Internet Source	<1 %
23	www.ejournal.radenintan.ac.id Internet Source	<1 %
24	www.stmik-budidarma.ac.id Internet Source	<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On