Indonesian Journal of Computer Science
ISSN 2302-4364 (*print*) dan 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam, No. 1, Padang, Indonesia, Telp. (0751) 7056199, 7058325
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Sentiment Analysis on the FIFA U-20 World Cup in Argentina Using Support Vector Machine

Achmad Warsito Sujatmiko[1], Anik Vega Vitianingsih[2*], Slamet Kacung[3], Dwi Cahyono[4], Anastasia Lidya Maukar[5]
Email titojatmiko91@gmail.com[1], vega@unitomo.ac.id[2*], slamet@unitomo.ac.id[3], dwikk@unitomo.ac.id[4], almaukar@president.ac.id[5]
[1,2,3,4]Informatics Department, Universitas Dr. Soetomo, Surabaya, Indonesia
[5]Industrial Engineering Department, President University, Bekasi, Indonesia

Article Information

Submitted : 6 May 2024
Reviewed : 11 May 2024
Accepted : 15 Jun 2024

Keyword

FIFA World Cup U-20 Argentina, Soundtrack World Cup U-20 2023, Sentiment Analysis, Support Vector Machine

Abstract

The decision made by FIFA regarding the selection of the soundtrack and the host country for the FIFA U-20 World Cup has sparked emotional reactions among the public and raised concerns about the event, especially on social media platform X. This is due to FIFA's decision to choose a soundtrack not from the host country, Argentina, but from the previous host, Indonesia. FIFA should advocate for the creation of a soundtrack by the host country to reflect its distinctive characteristics or atmosphere. Concerns about the U-20 World Cup in Argentina have also been fueled by the country's economic crisis, which is feared to affect the facilities and infrastructure for the young players representing their nations. This research focuses on filtering public responses to FIFA's decisions regarding the soundtrack selection and the host country for the U-20 World Cup into positive, neutral, and negative categories using the Support Vector Machine (SVM) method. The research aims to provide policy recommendations regarding the host selection process and cultural representation in international sports events. Additionally, this study is expected to provide a deeper understanding of the preferences and values held by the public regarding international sports. The research steps include data collection, pre-processing, labeling, weighting, and classification using a Support Vector Machine. The data for this research were obtained through crawling on social media platform X, totaling 2400 data points. The performance evaluation of the SVM algorithm using a 50:50 ratio of training and testing data yielded an average accuracy of 85.71%, Precision of 85.98%, Recall of 85.71%, and F1-score of 85.58%.

A. Introduction

The advancement of Technology and Information has laid a strong foundation for social change, particularly in communication and the global dissemination of information, especially in Indonesia, where Internet usage, as demonstrated by the Indonesian Internet Service Providers Association (APJII) survey from 2019 to 2020, is most common among the age group of 15-19 years old with 91%, followed by the age group of 20-24 years old with 88.5%, with a focus on communication (32.9%) and social media (51.5%)[1]. The We Are Social report revealed that 213 million people in Indonesia were online in January 2023, comprising 77% of the population, marking a 5.44% increase from the previous year and a total of 142.5 million new users since 2013. The advancement of Technology and Information has led to the emergence of social media platforms such as Facebook, Twitter, and Instagram, where users engage, share content, and make social connections[2]. The impact of social media on communication, information sharing, and social relationships is very significant. The global social media users currently reach 4.76 billion, forming 59.4% of the world's population. In Indonesia, the percentage of social media users reaches 60.4% of the country's total population for the same period[3]. Twitter has emerged as a popular mode of communication in Indonesian society due to its growing user base and rapid dissemination of trending information and news[4]. Twitter users can receive real-time feedback like status updates, known as "tweets." Tweets contain catalogs of individual preferences, reluctance, and feedback on various subjects[5]. According to We Are Social, there were 564.1 million Twitter users globally in July 2023. Global Twitter users increased by 16.1% year-on-year and 51.3% quarter-on-quarter. Indonesia ranked fourth globally, up from sixth in May 2023, with 25.25 million users. The United States remained the top country with 98.5 million Twitter users in July 2023. Japan ranked second with 67.9 million users, followed by India with 26.45 million users. Brazil and the United Kingdom had 22.95 million and 22.7 million users respectively.

The FIFA World Cup is an international-level football competition that gathers national teams from around the world. All teams that successfully qualify will be able to participate in this tournament to compete for the title of champion. The FIFA U-17 and U-20 World Cups are international football tournaments for youths classified by age. Football players in the national team who are under 20 years old are referred to as "U-20"[6]. The FIFA World Cup is participated by the senior men's national teams from every eligible country. FIFA is the highest football organization in the world that organizes this championship[7]. As one of the biggest and most anticipated sports events in the world, the FIFA World Cup has become a prestigious platform for the host countries[8]. The host country selected by FIFA will hold an event, featuring a special soundtrack or song for each FIFA World Cup. The FIFA World Cup showcases a traditional highlight known as the soundtrack or theme song for the event, typically created by the host country. FIFA has chosen to use a soundtrack produced by an Indonesian music group called Weird Genius titled "Glorious," sparking comments from the general public, especially football fans. This decision has not only triggered various public responses regarding the soundtrack but also raised concerns about various aspects such as Argentina, as the host

country, experiencing an economic crisis that will impact facilities and infrastructure.

To ensure FIFA and the host of the FIFA U-20 World Cup are effectively steering their actions towards their objectives, Twitter discourse provides a plethora of comments ripe for analysis. These remarks can shed light on how FIFA's choices resonate with the tournament's image, influence the music scene, and impact the host nation of the FIFA U-20 World Cup. Therefore, sentiment analysis is needed to understand the public's response more deeply to FIFA's decisions and the host's role in organizing the U-20 World Cup. By considering various comments on Twitter, patterns of positive, negative, or neutral sentiments that emerge in online discussions can be identified. This can provide valuable insights for FIFA and the host to adjust their strategies to better support their shared goals in running this global football event.

This study analyzes sentiment in Twitter comments using the Support Vector Machine (SVM) algorithm. To produce representations of different classes in multidimensional space, the SVM approach was chosen [9]. The generated hyper lane functions as a boundary and assists in classifying and separating between classes [9]. The Support Vector Machine (SVM) is capable of effectively delineating classes by creating a linear boundary with a substantial margin and is also able to handle feature vectors of unlimited dimensions [10]. Furthermore, labeling using the lexicon method is used to ensure the magnitude of expressed sentiment, where words are weighed beforehand [11]. This study aims to ascertain the mood that was prevalent during the FIFA U-20 World Cup in Argentina and to impart knowledge to audiences that are unfamiliar with the use of the SVM algorithm as a tool to help FIFA and the nation hosting the tournament accomplish their objectives. The expected outcome of this research is to provide further information on Twitter users' responses and reactions to various topics, such as Argentina's decision to select a song for the FIFA U-20 World Cup soundtrack and how it impacts the country's image, local music industry, and tournament organization.

B. Research Methodology

This research methodology involves a series of stages for sentiment analysis, as illustrated in Figure 1. The initial stage of this sentiment analysis research method is when the user inputs links from Twitter or platform X social media for dataset extraction using crawling to obtain raw datasets. These raw datasets are then entered into the Pre-Processing stage to be processed into clean datasets. The stages in Pre-Processing include Data Cleaning, Case Folding, Tokenizing, Language Normalization, Stopword Removal, Stemming, and Labeling. After Pre-Processing is done, the dataset will be entered into the Weighting stage using the TF-IDF method. This stage aims to extract the necessary TF-IDF features to identify and extract the most important keywords in a document or dataset. Next, after obtaining the weighting results, the dataset will be automatically processed for classification using the Support Vector Machine method. After the classification process is completed, an Evaluation of the SVM Method and sentiment classification results on the dataset will be provided.

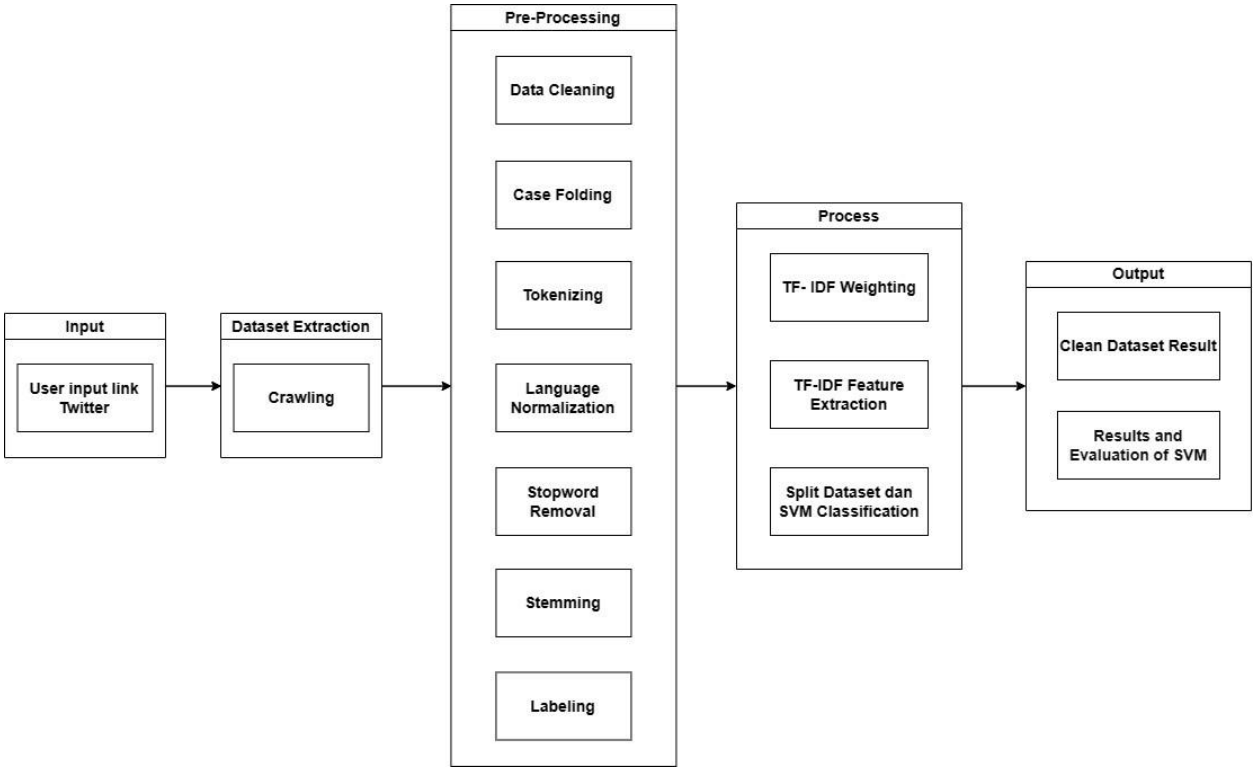

Figure 1. Research Methodology

1. Dataset

The data used originates from Twitter URLs or X https://twitter.com/, which serve as the primary data source. The main data for this research was obtained from collecting Twitter data through the Twitter API, focusing on approximately 2,400 tweets related to the FIFA U-20 World Cup held in Argentina. This dataset uses tweets Indonesian language text.

2. Pre-Processing

The pre-processing phase in text analysis involves various steps, including Data Cleaning, Case Folding, Tokenizing, Language Normalization, Stopword Removal, Stemming, and Labeling. This stage is crucial in preparing text data for further analysis because it allows for noise removal, format standardization, identification of information units, reduction of lexical variations, elimination of irrelevant words, normalization of words, reduction of feature dimensions, and providing appropriate labels for the desired analysis purposes.

Step-1: The Data Cleaning stage aims to prepare data for analysis by removing invalid, incomplete, or irrelevant data and addressing issues such as duplicates and outliers that can affect the integrity and interpretation of the data. Proper data cleaning ensures a clean and consistent dataset ready for processing, enhancing the quality of analysis, and reducing errors as seen in the results in Table 1.

Table 1. Cleaning Data

Dataset	Cleaning Data
@FaktaSepakbola logonya pildun ini justru lebih bagusan yang Argentina ketimbang di Indo (:	*logonya pildun ini justru lebih bagusan yang Argentina ketimbang di Indo*

Step 2: Case Folding is a text standardization process that converts all characters to lowercase, excluding non-alphabetic characters. This process aims to remove URLs, numbers, and punctuation marks for text document consistency. The Case Folding stage aims to change letters to lowercase or uppercase as seen in the results in Table 2.

Table 2. Case Folding

Data Cleanning	Case Folding
logonya pildun ini justru lebih bagusan yang Argentina ketimbang di Indo	*logonya pildun ini justru lebih bagusan yang argentina ketimbang di indo*

Step-3: Tokenization is the process of dividing text into smaller pieces known as tokens. Depending on the precise requirements of the text processing activity being performed and the complexity of the analysis, tokens might be words, phrases, or characters. As demonstrated by the findings in Table 3, tokenization also makes it possible to normalize words and eliminate superfluous characters, which facilitates text processing and increases analysis accuracy.

Table 3. Tokenizing

Case Folding	Tokenizing
logonya pildun ini justru lebih bagusan yang argentina ketimbang di indo	*['logonya', 'pildun', 'ini', 'justru', 'lebih', 'bagusan', 'yang', 'argentina', 'ketimbang', 'di', 'indo']*

Step 4: Language Normalization is the process of representing words, including converting words in the text to standard forms to ensure consistency for computational systems. Language Normalization involves changing the format and removing unnecessary elements such as punctuation and stop words as seen in the results in Table 4.

Table 4. Language Normalization

Tokenizing	Normalisasi Bahasa
['logonya', 'pildun', 'ini', 'justru', 'lebih', 'bagusan','yang','argentina', 'ketimbang', 'di', 'indo']	*['logonya', 'piala dunia', 'ini', 'justru', 'lebih', 'bagusan', 'yang', 'argentina', 'daripada', 'di', 'indonesia']*

Step-5: Stopword Removal is a process aimed at improving computational efficiency by eliminating unimportant words. Its focus is on removing keywords such as "and" and "or" for clearer text analysis. Removing keywords enhances analysis relevance by allowing focus on important words. This process aids in understanding content and extracting accurate information. Additionally, it enhances text processing performance by reducing feature space. Eliminating stopwords is crucial for efficient text analysis and interpretation, as evidenced by the results in Table 5.

Table 5. Stopword Removal

Normalisasi Bahasa	Stopword Removal
['logonya', 'piala dunia', 'ini', 'justru', 'lebih', 'bagusan', 'yang', 'argentina', 'daripada', 'di', 'indonesia']	*['logonya', 'piala dunia', 'bagusan', 'argentina', 'indonesia']*

Step 6: Stemming is a process aimed at reducing words to their base form by removing inflections and affixations. Stemming helps identify words with the same

root, enhancing efficiency in text analysis. Its goal is to simplify text comparison and effectively extract meaning. Stemming simplifies words such as "run" to enhance text analysis without losing meaning. It also improves consistency and reduces complexity in language models, as evident in the results in Table 6.

Table 6. Stemming

Stopword Removal	Stemming
['logonya', 'piala dunia', 'bagusan', 'argentina', 'indonesia']	*['logo', 'piala dunia', 'bagus', 'argentina', 'indonesia']*

Step 7: Labeling, the process of assigning labels to sentences or words based on the expressed emotions or attitudes, known as sentiment analysis. The main goal is to identify whether the text conveys positive, negative, or neutral sentiments, sometimes using Lexicon-based methods as seen in the results in Table 7.

Table 7. Labeling

Dataset Clean	Score	Label
logo piala dunia bagus argentina indonesia	-4	Negatives

3. TF-IDF Weighting

The method known as Term Frequency-Inverse Document Frequency (TF-IDF) plays a crucial role in transforming textual data into numerical vectors, thereby enhancing the efficiency and accuracy of various text processing activities [11]. Renowned for its effectiveness, user-friendliness, and precision, the TF-IDF method involves careful calculations of Term Frequency (TF) and Inverse Document Frequency (IDF) for individual words across documents, enabling the determination of the frequency of each word in the document as seen in the results in Table 8.

Table 8. TF-IDF Weighting Result

	aja	akal	akibat	argentina	arti	audit
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0.227648

4. Support Vector Machine

Support Vector Machine (SVM) is a machine learning technique designed to identify the optimal hyperplane that divides two classes in the input domain, and its classification algorithm constructs a model based on training data to predict the classification of new unseen test data samples [12]. The distance between each class's nearest data point and the hyperplane, also referred to as the support vector, defines the term margin [13]. The prominent lines indicate the optimal hyperplane equidistant from two classes. [13]. Support Vector Machine (SVM) works especially well in high-dimensional spaces, showing good performance when the number of dimensions is greater than the number of samples. Additionally, SVM's ability to utilize subsets of training points enhances memory efficiency [14].

The classification outcomes of the Support Vector Machine (SVM) model function as the central point for an extensive analysis to fully comprehend the underlying patterns in the dataset. This evaluation process often necessitates careful examination of both positive and negative sentiments expressed in the text, as well as the identification of key features that significantly influence sentiment determination. Additionally, it involves delving deeper into the reasons behind inaccuracies observed in the classification results, to enhance the overall understanding of the classification process.

6. Evaluation Model

After the Support Vector Machine (SVM) modeling process, the results are evaluated using the Confusion Matrix. The accuracy value for the data is determined by calculating the ratio of correctly predicted data to the total data, including both successful and unsuccessful predictions. The classification outcomes are represented by four terms: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Equation (1) will be used to assess the accuracy, which is the frequency with which the classification model correctly predicts the data class. Based on Equation (2), we will assess the classification model's predictive accuracy for positive outcomes. Equation (3) may be used to calculate the recall, or the percentage of real positive data that the model correctly predicts to be positive. Equation (4) is used to calculate the F1-score, a statistic that uses the harmonic mean to balance recall and accuracy.

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

$$Presisi = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\text{-}Score = 2\ x\ \frac{Presisi\ x\ recall}{Presisi+Recall} \quad (4)$$

C. Result and Discussion

The method used for sentiment analysis in this study involves employing a Support Vector Machine with lexicon-based labeling. In the data collection stage, the author reviews the information available on social media platforms such as Twitter or Xapps. In this research, using the data collection stage, the author conducted a review of information available on social media platforms such as Twitter or X. This study successfully gathered a dataset consisting of 2400 user reviews related to the FIFA U-20 World Cup in Argentina found on various social media accounts on the platform using the Crawling method, as seen in Table 9. It's important to note that the dataset used in this study consists of Indonesian language texts.

This dataset comprises a variety of user comments and perspectives on the event. A total of 2,400 review data were utilized in the study. These were

subsequently divided evenly, with 50% allocated for training data and the remaining 50% for test data, ensuring balanced representation throughout the analysis phase.

Table 9. Dataset

created_at	full_text
Sat May 20 10:49:00 +0000 2023	*@FaktaSepakbola Sad*
Sat May 20 09:17:23 +0000 2023	*@FaktaSepakbola Indonesia dapet emas sea games palestina ikut seneng? Ga juga.*
Sat May 20 06:50:21 +0000 2023	*@FaktaSepakbola Opening Ceremony di Gbk jam berapa eyy?*
Sat May 20 13:10:40 +0000 2023	*@FaktaSepakbola seharusnya sekarang Palembang udah rame bule bule tapi ah sudahlah sakit tapi tak berdarah*
Sat May 20 08:48:40 +0000 2023	*@FaktaSepakbola Mungkin Guatemala dan New Zealand ngga akan pernah ikut piala dunia senior*
Sat May 20 07:34:43 +0000 2023	*@FaktaSepakbola Salahkan ganjar dan partainy*
Sat May 20 12:34:11 +0000 2023	*@FaktaSepakbola Ganjar dan koster nanti yang nyerahin piala dunia bagi yang juara*
Sat May 20 09:50:56 +0000 2023	*@FaktaSepakbola Yok gagal jd presiden bisa yoook @ganjarpranowo*
Sat May 20 06:34:46 +0000 2023	*@FaktaSepakbola Padahal udh mau move on malah di ingetin lagi*
Sat May 20 06:35:41 +0000 2023	*@FaktaSepakbola gara gara capres ubanan sama roller coaster*

To maximize the results of sentiment analysis, researchers utilize various processes. The first process is pre-processing aimed at cleaning and organizing text data for easier analysis. In this pre-processing, there are several stages aimed at improving the quality of sentiment analysis by preparing the data appropriately. The stages include data cleaning, lowercase conversion, tokenization, language normalization, removal of common words, stemming, and labeling.

The second process is weighting using TF-IDF on the cleaned dataset to obtain numeric representations of words in the text. By using TF-IDF, researchers can identify the most important or unique words in each document, which can assist in sentiment analysis by giving appropriate weights to those words. After TF-IDF weighting, feature extraction is performed followed by classification using the Support Vector Machine (SVM) method with a linear kernel and equipped with oversampling techniques commonly used to address imbalance issues using oversampling RandomOverSampler (random_state=50). The accuracy obtained before using oversampling methods is 50%, but after using oversampling methods, the accuracy increases to 85%.

Based on the sentiment analysis results, it can be stated that the performance of the SVM method using a linear kernel with Twitter user review data or X for research related to the U-20 World Cup in Argentina has been successfully implemented. In this investigation, preprocessing is conducted using 2,400 data points for cleaning and classification, employing the Support Vector Machine (SVM) approach. A 50:50 ratio of training data to test data is utilized. The training data handling procedure involves generating synthetic samples from minority classes using the SMOTE approach to address overfitting issues in imbalanced data settings. The categorization outcomes are presented in Figure 2's Confusion Matrix Table 10,

which also displays the model performance results. Additionally, Figure 3 illustrates the categorization findings through a pie chart.

Actual Negatif:

- Predicted Negatif: 25 (25 correct predictions)
- Predicted Netral: -20 (20 incorrect predictions)
- Predicted Positif: -15 (15 incorrect predictions)

Actual Neutral:

- Predicted Negatif: 3 (3 incorrect predictions)
- Predicted Neutral: 25 (25 correct predictions)
- Predicted Positif: 25 (25 incorrect predictions)

Actual Positif:

- Predicted Negatif: 1 (1 incorrect prediction)
- Predicted Neutral: 10 (10 incorrect predictions)
- Predicted Positif: 20 (20 correct predictions)

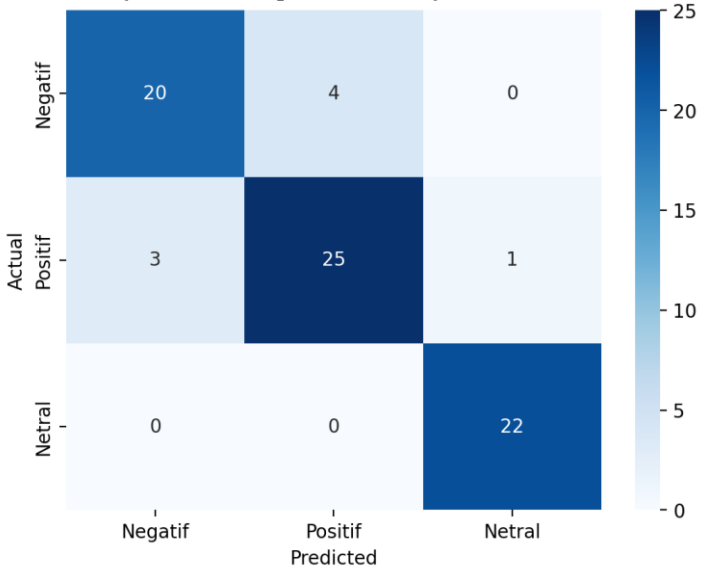

Figure 2. Confusion Matrix

Table 10. Performance results of Support Vector Machine

Metric	Score	Percentage
Accuracy	0.857142857	85.71%
Precision	0.859790507	85.98%
Recall	0.857142857	85.71%
F1-Score	0.855819111	85.58%

The pie chart in the above image, presented in Figure 3, depicts the results obtained from sentiment analysis conducted using the Support Vector Machine (SVM) algorithm with a linear kernel and Oversampling technique. The analysis

shows that negative sentiment accounts for 63.2%, while neutral sentiment is 20.8%, and positive sentiment has the smallest value at 15.9%.

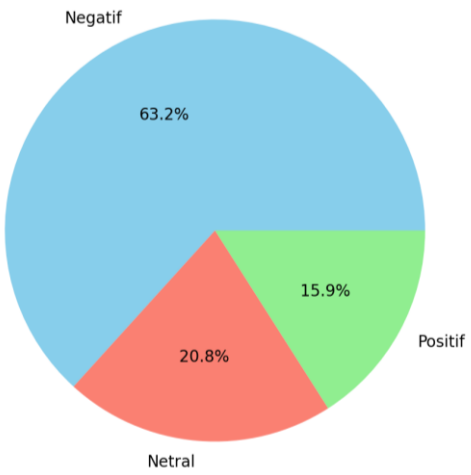

Figure 3. Classification Visualization for Actual Class

D. Conclusion

This research conducted sentiment analysis using the Twitter social media platform, resulting in a dataset consisting of 2400 comments. Following the sequence of steps: Pre-Processing, TF-IDF weighting, and implementation of the Support Vector Machine (SVM) classification model with a 50:50 split between training and test data. The SVM algorithm demonstrated the ability to categorize sentiment with an accuracy rate of 85.71%, precision of 85.98%, recall of 85.71%, and an F1-Score of 85.58%. From the sentiment analysis using the SVM algorithm with a linear kernel, it was found that the obtained negative sentiment was 63.2%, neutral sentiment was 20.8%, and positive sentiment had the smallest proportion at 15.9%. The results of this research are expected to provide insights into the world of football, especially for football organizations and countries hosting football competitions, aiming to enhance satisfaction among fans and sports enthusiasts. Therefore, the SVM algorithm emerges as a highly efficient instrument in examining sentiment in Twitter comments, showing potential for further utilization in designing solutions related to improving football fan satisfaction.

E. Acknowledgment

This research was supported by the Informatics Department, Faculty of Engineering Universitas Dr. Soetom. We would like to thank GitHub for helping us out and for allowing us to use their dataset as the foundation for our study.

F. References

[1] Y. Sari and D. H. Prasetya, "Literasi Media Digital Pada Remaja, Ditengah Pesatnya Perkembangan Media Sosial," *J. Din. Ilmu Komun.*, vol. 8, no. 1, pp. 12–25, 2022.

[2] A. N. Z. Muhammad and A. S. Sitanggang, "ANALISIS DAMPAK SOSIAL MEDIA DALAM PENGEMBANGAN SISTEM INFORMASI," *J. Ilm. Indones.*, vol. 3, no. April, pp. 386–393, 2023.

[3] Fauzan Baehaqi and Nuri 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, [Online]. Available: http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135

[4] Tiara Danirmala and Y. S. Nugroho, "Analisis Sentimen Terhadap Topik Kenaikan Harga Bahan Bakar Minyak (BBM) pada Media Sosial Twitter," *Indones. J. Comput. Sci.*, vol. 12, no. 3, pp. 1258–1268, 2023, doi: 10.33022/ijcs.v12i3.3199.

[5] N. Cahyono and Dewi Setiyawati, "Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Perokok Di Indonesia," *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 262–272, 2023, doi: 10.33022/ijcs.v12i1.3154.

[6] M. A. Maulana and M. L. Hakim, "Politik, Olahraga, dan Islam Studi Kasus Pembatalan RI Menjadi Tuan Rumah Piala Dunia U-20 2023," vol. 1, pp. 16–24, 2023.

[7] P. R. Saputra Andri , Subing Mulia, "Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Mengenai Piala Dunia Fifa 2022," *TEKNOMATIKA*, vol. 13, no. 01, pp. 22–31, 2023, [Online]. Available: http://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/616

[8] G. A. Chairunnisa, M. M. Muchlis, U. A. Indonesia, and U. I. Jakarta, "Transformasi Ekonomi Melalui Piala Dunia FIFA 2022: Analisis Meningkatnya Pendapatan Per Kapita di Qatar sebagai Tuan Rumah," vol. 2, no. 4, pp. 771–778, 2023.

[9] E. R. Lidinillah, T. Rohana, and A. R. Juwita, "Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 154–164, 2023, doi: 10.37373/tekno.v10i2.440.

[10] U. I. Arsyah, M. Pratiwi, and A. Muhammad, "Twitter Sentiment Analysis of Public Space Opinions using SVM and TF-IDF Methods," *Indones. J. Comput. Sci.*, vol. 13, no. 1, pp. 387–394, 2024, doi: 10.33022/ijcs.v13i1.3594.

[11] N. M. Farhan and B. Setiaji, "Komparasi Metode Naive Bayes dan SVM pada Sentimen Twitter Mengenai Persoalan Perpu Cipta Kerja," *Indones. J. Comput. Sci.*, vol. 12, no. 5, pp. 2718–2727, 2023, doi: 10.33022/ijcs.v12i5.3375.

[12] E. Suryati, Styawati, and A. Ari Aldino, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," *J. Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 96–106, 2023.

[13] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.

[14] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *J. Inform. Upgris*, vol. 7, no. 1, pp. 36–40, 2021, doi: 10.26877/jiu.v7i1.7099.