



Sentiment Analysis of Indonesian National Team Failure in the 2026 World Cup Qualifications Using Support Vector Machine Algorithm

Muhammad Nouval^{1*}, Fanza Maulana Habibi², Anisya Rahmi³, Muhammad Dawam Amru Bittaqwa⁴, Rizki Agustianto⁵

^{1,2,3,4,5} Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
15230077@bsi.ac.id^{1*}, 15230208@bsi.ac.id², 15230303@bsi.ac.id³, 15230029@bsi.ac.id⁴, 15230292@bsi.ac.id⁵

Abstract

The Indonesian National Team's failure in the 2026 World Cup qualifiers has generated diverse responses on social media, particularly on Ferry Irwandi's YouTube channel. This study aims to analyze public sentiment towards the national team's performance based on YouTube user comments. The method used is a Support Vector Machine (SVM) with stages of data scraping, pre-processing (cleaning, case folding, normalization, tokenization, stopword removal, stemming), lexicon-based automatic labeling, and model evaluation using a confusion matrix. The data consists of 8,353 comments divided with a ratio of 80:20 for training and testing. The results show that the SVM algorithm is able to classify comments into two classes, positive and negative, with an accuracy of 81%, a precision of 82%, a recall of 83%, and an F1-score of 82%. These results demonstrate the effectiveness of SVM in accurately and stably identifying public opinion towards the Indonesian National Team's failure.

Keywords: Sentiment Analysis, Support Vector Machine, Text Mining, Indonesian National Team

1. Introduction

Football is one of the sports that is very popular among the Indonesian people, and has now developed into a very promising industry [1]. Football's popularity is inseparable from its long evolution, from its humble beginnings to a modern sport utilizing sophisticated coaching methods. As a competitive sport, football offers individuals the opportunity to develop physical fitness, technical skills, playing strategies, and mental fortitude, enabling them to compete at both the national and international levels [2].

In the digital age, social media has become a primary platform for people to express their opinions on various issues, including football. One widely used platform is YouTube, where many people upload match videos, analysis, and commentary from football experts. This research took comment data from Ferry Irwandi's YouTube channel, which discussed Indonesia's failure to qualify for the 2026 World Cup. In the comments on Ferry Irwandi's video, there are various responses from viewers regarding the performance of the Indonesian national team, which directly reflect the feelings of the public.

Analyzing YouTube comments can provide an honest picture of public sentiment. Sentiment analysis allows researchers to categorize people's opinions into two categories: positive and negative, based on the text they write. One algorithm proven effective for this classification is the Support Vector Machine (SVM), due to its ability to generalize and improve performance, especially on large data sets. With this method, research can be conducted in an organized and structured manner to determine the extent of public sentiment toward the Indonesian national team's failure in the 2026 World Cup qualifiers.

This study aims to analyze YouTube user sentiment regarding comments on a video titled Indonesia failed to qualify for the 2026 World Cup on Ferry Irwandi's channel. The method used is the Support Vector Machine (SVM) algorithm. The results are expected to provide insight into how the public views the Indonesian national team's performance and serve as a reference for relevant parties in managing Indonesian football.

2. Research Methodology

2.1. Text Mining

Text mining is a scientific process used to extract hidden information, meaning, and knowledge from unstructured text data sets. This process transforms public comments on social media into measurable data that can be analyzed quantitatively. Because information in text is implicit and not directly stored in data structures, text mining is a crucial step in converting public emotional expressions into data that can be further processed [3].

2.2. Sentiment Analysis

Sentiment analysis, or opinion mining, is a subset of text mining, defined as the computational study of people's opinions, sentiments, judgments, attitudes, and emotions toward an entity or attribute. Its purpose is to extract, identify, and analyze the characteristics of a person's sentiments or tendencies in expressing opinions about an issue or object, conducted within a text unit using NLP (Natural Language Processing), statistics, or machine learning methods [3].

2.3. Support Vector Machine

According to Lee & Lee (2015), a Support Vector Machine is a machine learning method that works based on the principle of Structural Risk Minimization (SRM). Its goal is to find the best hyperplane to separate two classes in the input space. The best hyperplane is the one located midway between two groups of objects from two classes. The best separating hyperplane between the two classes can be found by measuring the margin of the hyperplane and finding its maximum value. The margin is the distance between the hyperplane and the closest pattern of each class. This closest pattern is called the support vector [4].

2.4. Python

Python is a high-level programming language that can be run directly, interactively, uses the concept of objects, and can be used on various operating systems such as Mac, Linux, and Windows [5]. According to K. R. V. Kothapalli, M. A. Mohammed, R. Mohammed, and P. Pasam (2024) In addition, Python is easier to use than compiled programming languages because many complex things such as memory management are taken care of automatically. Python's processing speed is a big advantage in product development, especially those that require a fast iteration process [6].

3. Results and Discussion

This research generally consists of several stages, starting with data scraping, data preprocessing, data labeling, data splitting, and Support Vector Machine model evaluation. Figure 1 shows the research stages, which include:

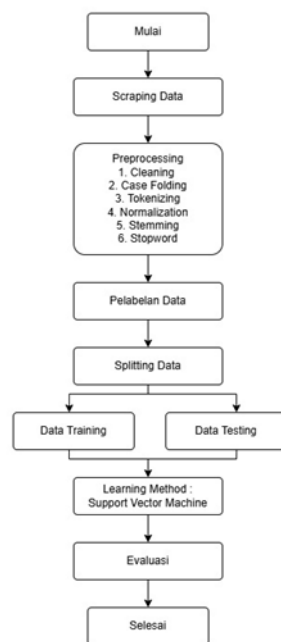


Fig. 1: Research Stages

3.1. Scraping Data

Scraping is a method of data collection that is used to collect information easily and orderly, thus facilitating the data analysis process [7]. The data used for this study comes from comments on Ferry Irwandi's YouTube channel. Python was used to collect 8,353 comments. The data was then preprocessed before further analysis.

Table 1: Scraping Data

Author	Comment Text	Published at
@btoahdo4458	Mundurrrr Cookkk Munduurrrr Kemajooonnn Cooookk	2025-10-16T10:32:02Z
@auliasinarjaya8694	Timnas usia muda yg dibangun Indra Syafri & Fa..	2025-10-16T07:45:29Z
@nikovarv6774	STY dan PK tetaplah jadi kenangan... Jangan Ke...	2025-10-16T07:02:03Z
@JK4R-69	Beneran dipecat Anjir si Patrick... Ferry lagi...	2025-10-16T06:31:29Z
@ryeadhiakhbar8925	Assalamu'alaikum selamat siang Pak ferry, maaf...	2025-10-16T04:42:19Z

3.2. Pre-Processing Data

The next stage is data pre-processing. At this stage, the data is processed into information that can be used in the next stage according to a predetermined format [8]. Pada tahap pre-processing dibagi menjadi 6 langkah yaitu cleaning, case folding, normalisasi tokenisasi, stopword, dan stemming.

a. Cleaning Data

This step aims to clean the text of noise to make the data cleaner. The data cleaning results are shown in Table 2.

Table 2: Cleaning Data

Before	After
Beneran dipecat Anjir si Patrick... Ferry lagi-lagi 100% Winrate 😊	Beneran dipecat Anjir si Patrick Ferry lagi lagi NUM Winrate
"itu kan proyek ET sebagai ketua Federasi#🤔" blog 😊	itu kan proyek ET sebagai ketua Federasi blog jangan samain ET sama Florentino Perez blog
Mungkin alasan terbesar untuk mempertahankan petrick, supaya pemain titipan bisa masuk timnas. MUNGKIN. 😊😊	Mungkin alasan terbesar untuk mempertahankan petrick supaya pemain titipan bisa masuk timnas MUNGKIN
lanjut,,agar makin kompak	lanjut agar makin kompak

b. Case Folding

The next step is case folding, which converts all letters to lowercase. This process makes searching easier, as not all text documents are suitable for uppercase letters [9]. The results of the folding case are in table 3.

Table 3: Case Folding

Before	After
ET ganti Ratu Tisha aja sekalian 😊	et ganti ratu tisha aja sekalian
STY dan PK tetaplah jadi kenangan... Jangan Kembali lagi... masih ada banyak pilihan.. "contoh Alegri yang di pecat, dan kembali ke bianconeri, endingnya juga putus"...	sty dan pk tetaplah jadi kenangan jangan kembali lagi masih ada banyak pilihan contoh alegri yang di pecat dan kembali ke bianconeri endingnya juga putus
STY itu pelatih kelas pildun, ngelatih korsel yang top 25 dunia, pernah ngalahin jerman di pildun waktu jadi head coach korsel, punya pengetahuan, pemahaman dan jam terbang tinggi lawan tim asia yang dimana ini sangat penting karena sebelum ke pildun kita harus lewat kualifikasi yang lawannya tim asia semua, malah dituker sama sekelas PK, sungguh benar-benar mindblowing	sty itu pelatih kelas pildun ngelatih korsel yang top num dunia pernah ngalahin jerman di pildun waktu jadi head coach korsel punya pengetahuan pemahaman dan jam terbang tinggi lawan tim asia yang dimana ini sangat penting karena sebelum ke pildun kita harus lewat kualifikasi yang lawannya tim asia semua malah dituker sama sekelas pk sungguh benar benar mindblowing
Si arya sinalungga pecat	si arya sinalungga pecat

c. Word Normalization

Normalizing words using a non-standard dictionary aims to change non-standard words, including slang, into standard word forms according to the Big Indonesian Dictionary (KBBI) [10]. This process uses a dictionary in Excel format, which contains two columns for non-standard words and standard words. The results of the word normalization are shown in Table 4.

Table 4: Word Normalization

Before	After
Sakit banget hati gw liat potensi besar timnas garuda saat ini yang malah dihancur leburkan oleh orang2 yg lebih mengedepankan EGO !	sakit banget hati gue lihat potensi besar timnas garuda saat ini yang malah dihancur leburkan oleh orang-orang yang lebih mengedepankan ego !
STY itu pelatih kelas pildun, ngelatih korsel yang top 25 dunia, pernah ngalahin jerman di pildun waktu jadi head coach korsel, punya pengetahuan, pemahaman dan jam terbang tinggi lawan tim asia yang dimana ini sangat penting karena sebelum ke pildun kita harus lewat kualifikasi yang lawannya tim asia semua, malah dituker sama sekelas PK, sungguh benar-benar mindblowing	sty itu pelatih kelas pildun ngelatih korsel yang top num dunia pernah mengalahkan jerman di pildun waktu jadi head coach korsel punya pengetahuan pemahaman dan jam terbang tinggi lawan tim asia yang dimana ini sangat penting karena sebelum ke pildun kita harus lewat kualifikasi yang lawannya tim asia semua malah dituker sama sekelas pakai sungguh benar benar mindblowing

<p>kenapa harus seseorang harus mundur? Yang Harus Mundur ya SEMUA Untung dah gak liat semenjak ganti pelatih. Karena wajah pelatih baru tidak meyakinkan dan tidak membangun tim yang baik malah hanya main taktik tai</p>	<p>kenapa harus seseorang harus mundur? yang harus mundur ya semua untung sudah tidak lihat semenjak ganti pelatih karena wajah pelatih baru tidak meyakinkan dan tidak membangun tim yang baik malah hanya main taktik tahi</p>
--	---

d. Tokenizing

This tokenization is done to simplify the subsequent pre-processing stages. The tokenization results in the Python programming language used are placed in a list data structure [11]. The results of tokenizing are in table 5.

Table 5: Tokenizing

Before	After
<p>Mungkin alasan terbesar untuk mempertahankan petrick, supaya pemain titipan bisa masuk timnas. MUNGKIN. 😊😊 STY dan PK tetaplah jadi kenangan... Jangan Kembali lagi... masih ada banyak pilihan.. "contoh Alegri yang di pecat, dan kembali ke bianconeri, endingnya juga putus"...</p> <p>Sakit banget hati gw liat potensi besar timnas garuda saat ini yang malah dihancurkan leburkan oleh orang2 yg lebih mengedepankan EGO !</p> <p>mimpi nii ye....., masuk piala dunia. mundur pak Erik jdi ketua PSSI banyak yg lain mau menjadi ketua PSSI yg mau bangun timnas Indonesia.</p>	<p>Mungkin, alasan, terbesar, untuk, mempertahankan, petrick, supaya, pemain, titipan, bisa, masuk, timnas, mungkin sty, dan, pakai, tetaplah, jadi, kenangan, jangan, kembali, lagi, masih, ada, banyak, pilihan, contoh, alegri, yang, di, pecat, dan, kembali, ke, bianconeri, endingnya, juga, putus sakit, banget, hati, gue, lihat, potensi, besar, timnas, garuda, saat, ini, yang, malah, dihancurkan, leburkan, oleh, orang, orang, yang, lebih, mengedepankan, ego, ! mimpi, ini, ye, masuk, piala, dunia, mundur, pak, erik, jadi, ketua, pssi, banyak, yang, lain, mau, menjadi, ketua, pssi, yang, mau, bangun, timnas, indonesia</p>

e. Stopword Removal

This step involves removing common words that don't have strong sentimental meaning. The goal is to only retain important words that influence the sentiment labeling results. The results for stopwords are shown in Table 6.

Table 6: Stopword Removal

Before	After
<p>Si arya sinalungga pecat STY dan PK tetaplah jadi kenangan... Jangan Kembali lagi... masih ada banyak pilihan.. "contoh Alegri yang di pecat, dan kembali ke bianconeri, endingnya juga putus"...</p> <p>"itu kan proyek ET sebagai ketua Federasi" blog 😊 jangan samain ET sama Florentino Perez blog 😊 kasih kesempatan pastor jadi head coach untuk piala asia dan match day...menghemat uang pssi..</p>	<p>arya, sinalungga, pecat pakai, tetaplah, kenangan, jangan ,kembali, lagi ,masih, banyak, pilihan, contoh, alegri, pecat, kembali, bianconeri, endingnya, juga, putus proyek, ketua, federasi, blog, jangan, samain, sama, florentino, perez, blog kasih, kesempatan, pastor, head, coach, asia, match, day, menghemat, uang, pssi</p>

f. Stemming

The next stage is stemming, which involves converting a word to its base form by removing affixes such as prefixes, infixes, or suffixes. The stemming results are shown in Table 7.

Table 7: Stemming

Before	After
<p>mari kita hancurkan juga mimpi ET di pilpres STY dan PK tetaplah jadi kenangan... Jangan Kembali lagi... masih ada banyak pilihan.. "contoh Alegri yang di pecat, dan kembali ke bianconeri, endingnya juga putus"...</p> <p>Ini pelatih jejak karirnya apa? Kemampuan dia pas jadi pemain bola apa? Kenapa pssi membuang pelatih emasnya padahal pelatih sty jarang di temui didunia dan tipe tipe pelatih yang 10.000 tahun sekali baru ada pelatih macem ini? Apa urgensi dan agenda pak erik tohir memilih pelatih ini? Tanyakan ke pak erik tohir.. Dia pemimpin nya.... Knp hilang bgtu saja...</p>	<p>mari, kita, hancur, juga, mimpi, pilpres pakai, tetap, kenang, jangan, kembali, lagi, masih, banyak, pilih, contoh, alegri, pecat, kembali, bianconeri, endingnya, juga, putus latih, jejak, karirnya, apa, mampu, dia, pas, main, bola, apa, kenapa, pssi, buang, latih, emas, padahal, latih, jarang, temu, dunia, tipe, tipe, latih, tahun, sekali, baru, latih, macem, apa, urgensi, agenda, tohir, pilih, latih tanya, tohir, dia, pimpin, kenapa, hilang, begitu</p>

3.3. Data Labeling

In this process, the pre-processed dataset will be labeled as positive or negative. This labeling is done automatically using a lexicon-based dictionary that assigns a score to each word related to the comment opinion. The results of the data labeling are shown in Table 8.

Table 8: Data Labeling

Comment	Sentiment
pk mundur dan ketum harus mundur	Negative
Mundurrrr Cookkk Munduurrrr Kemajooonnn Cooockk	Negative
meskipun nunggu 5 tahun lagi apa berapa lama pun kalo cara milih pelatihnya masih model sama kayak waktu pemilihan patrick kluivert ya sama aja kita semua ampe	Negative

sekarang kan masih heran "apa sebenarnya alasan patrick kluivert dipilih sebagai pelatih ??"	Positive
Apakah alex pastoor cocok mengganti kan patrick Kluivert sampai piala asia?	Negative
seorang harus mundur? seseorang harus mati!	Negative

3.4. Splitting Data

After the data labeling stage is complete, the next step is to split the data into training and testing data. This data splitting system uses an 80:20 ratio. The results are shown in Table 9.

Table 9: Splitting Data

Rasio Perbandingan	Data Training	Data Testing
80:20	6.480	1.620

3.5. Classification Results

The classification results on the Support Vector Machine algorithm with a ratio of training data and testing data of 80:20. Testing this model uses a confusion matrix by showing the classification results of the comment data category, obtaining the highest precision, recall, F1-score, and support values, namely negative totaling 4,586 with a percentage of 55.8%, while for positive comments totaling 3,630 with a percentage of 44.2%, and model accuracy of 81%. Showing the balance of data, evidenced by the graph of comment classification results as follows.

Classification Report:

	precision	recall	f1-score	support
negatif	0.820	0.843	0.831	897
positif	0.798	0.770	0.784	723
accuracy			0.810	1620
macro avg	0.809	0.807	0.808	1620
weighted avg	0.810	0.810	0.810	1620

Fig. 2: Classification Report

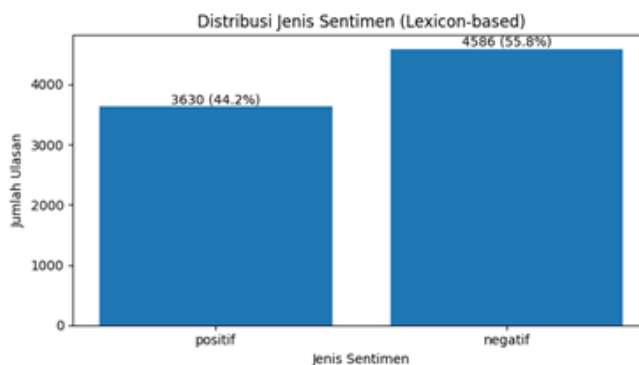


Fig. 3: Classification Chart

3.6. Evaluation Results

The final stage is the evaluation of the model using the Support Vector Machine. This evaluation is conducted to determine process performance using a confusion matrix. According to Putra and Wibowo (2020), a confusion matrix is a table that shows the number of test data correctly grouped and the number of test data incorrectly grouped [12]. Proven by the following results.

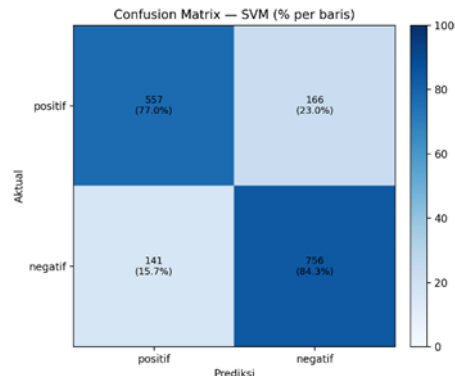


Fig. 6: Confusion Matrix

4. Conclusion

Sentiment analysis of user opinions on Ferry Irwandi's YouTube channel shows diverse responses regarding the failure of the Indonesian national team in the 2026 World Cup qualifiers. This study used the Support Vector Machine (SVM) algorithm. The results and evaluation show that SVM is able to group comments into two types of sentiment, namely positive and negative, with a very good level of accuracy. The data preparation process, which includes data cleaning, case folding, word normalization, tokenizing, stopword removal, and stemming, successfully improved data quality before the qualifications were carried out. From the evaluation results, SVM has proven effective in identifying public opinion with an accuracy rate of 81%, precision 82%, recall 83%, and f1-score 82%. Thus, it can be concluded that the use of the Support Vector Machine algorithm provides accurate and stable results in determining the tendency of public sentiment towards the failure of the Indonesian national team in the 2026 World Cup.

5. Suggestion

This research still has room for further development. As a recommendation, future research could combine the Support Vector Machine (SVM) algorithm with other methods such as Naïve Bayes or Random Forest to improve sentiment classification accuracy. This combination of the two algorithms is expected to leverage their respective strengths, with SVM offering good generalization capabilities while the companion algorithm can assist in determining more optimal class boundaries. Furthermore, the pre-processing stage also needs to be refined to ensure more accurate and consistent labeling and identification of positive and negative sentiment words.

References

- [1] M. Malik, D. A. Pangestu, and M. R. Pribadi, "AICOMS Applied Information Technology and Computer Science Analisis Sentimen Hasil Pertandingan Sepakbola Timnas Indo-nesia di Piala Asia U-23 pada Platform Youtube menggunakan Algoritma Suport Vector Machine (SVM)," vol. 3, no. 1, pp. 38–45, 2024.
- [2] H. Ngatifah and F. A. Irawan, "Kecepatan dalam serangan balik menjadi andalan: Analisis pertandingan Tim Nasional Indonesia U-23 pada Piala Asia 2024," *Sepakbola*, vol. 5, no. 1, pp. 77–86, 2025, doi: 10.33292/sepakbola.v5i1.388.
- [3] H. Eni Tri and S. Ari, "Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 3, pp. 32–37, 2022, [Online]. Available: <http://repository.teknokrat.ac.id/3224/>
- [4] T. Hidayat, N. Nurchim, and N. Nurmalitasari, "Analisis Sentimen Ulasan Pengguna pada Aplikasi SIREKAP 2024 Menggunakan Machine Learning," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 14, no. 1, pp. 109–116, 2025, doi: 10.30591/smartcomp.v14i1.7021.
- [5] R. Kurnia, "Journal of Vocational Education and Information Technology," *J. Vocat. Educ. Inf. Technol.*, vol. 1, no. 1, pp. 1–6, 2020.
- [6] L. Simorangkir, E. Sany, P. Studi, T. Informatika, and U. N. Hamzah, "Penerapan Metode K-Means Untuk Pengelompokan Pariwisata Provinsi Jambi," pp. 35–40, 1907.
- [7] A. T. Susilawati, N. A. Lestari, and P. A. Nina, "Analisis Sentimen Publik Pada Twitter Terhadap Boikot Produk Israel Menggunakan Metode Naïve Bayes," *Nian Tana Sikk. J. Ilm. Mhs.*, vol. 2, no. 1, pp. 26–35, 2024, [Online]. Available: <https://doi.org/10.59603/niantanasikka.v2i1.240>
- [8] R. A. A. Renal, Syariful Alam, and Moch Hafid T, "Komparasi Payment Digital Untuk Analisis Sentimen Berdasarkan Ulasan Di Google Playstore Menggunakan Metode Support Vector Machine," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 118–128, 2023, doi: 10.55123/storage.v2i3.2337.
- [9] K. Ahmad, "Tahapan Pra-pemrosesan Teks (Text Preprocessing)," *J. Mandalika Lit.*, vol. 4, no. 4, pp. 323–332, 2023.
- [10] H. H. Mubaroroh, H. Yasin, and A. Rusgiyono, "Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naïve Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance," *J. Gaussian*, vol. 11, no. 2, pp. 248–257, 2022, doi: 10.14710/j.gauss.v11i2.35472.
- [11] M. A. Nur and N. Wardhani, "Optimasi Normalisasi Kata Pada Data Twitter Untuk Meningkatkan Akurasi Analisis Sentimen (Studi Kasus Respon Masyarakat Terhadap Layanan Teman Bus)," *J. Fokus Elektroda Energi List. Telekomun. Komputer, Elektron. dan Kendali*, vol. 7, no. 4, pp. 237–243, 2022.
- [12] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.