Independent IT Journal
Vol. 14 No. 1, July (2025), pp. 208-218

ISSN 2301-8984 (Print), 2809-1884 (Online)
Published by Institute of Computer Science

Sentiment analysis towards naturalization of Indonesian National Team Players on social media x using the Naive Bayes method

Fahrian Zibran Lubis[1], Rakhmat Kurniawan[2]
[1,2] Computer Science, State Islamic University of North Sumatra, Medan, Indonesia

ARTICLE INFO

Article history:

Received Jun 30, 2025
Revised Jul 16, 2025
Accepted Jul 28, 2025

Keywords:

Indonesian National Team;
Naïve Bayes;
Naturalized Players;
Sentiment Analysis;
Social Media X.

ABSTRACT

This study analyzes public sentiment toward naturalized players in the Indonesian National Team on social media platform X (formerly Twitter) using the Naïve Bayes method. Data were collected via Python's *snscrape* library through web crawling, encompassing 700 tweets from January 2023 to May 2024. The research methodology included data preprocessing (*cleaning*, *case folding*, *tokenizing*, *stopword removal*, and *stemming*), feature extraction with TF-IDF (*Term Frequency-Inverse Document Frequency*), and sentiment classification. Results revealed a dominant negative sentiment (87.5%) compared to positive sentiment (12.5%), with a model accuracy of 88%. The most frequent keyword, "*main*" (play), reflected public focus on player performance.The study contributes to the field in three key aspects: (1) It addresses a gap in literature by specifically examining sentiment toward naturalization policies in Indonesian football using social media data; (2) It demonstrates the effectiveness of Naïve Bayes in handling informal Indonesian language, achieving high accuracy despite linguistic complexities; (3) It provides actionable insights for policymakers, highlighting the need for greater transparency in naturalization processes. Limitations include potential bias due to imbalanced data and challenges in interpreting sarcasm. Recommendations for future research include expanding datasets to multiple platforms and testing advanced models like BERT for improved contextual analysis.

Corresponding Author:

Fahrian Zibran Lubis,
Faculty of Computer Science,
State Islamic University of North Sumatra,
Jl.Lap Golf, Kp.Central, Pancur Batu District, Deli Serdang, North Sumatra, Indonesia, 20353
Email: Fahrizibran76@gmail.com

1. INTRODUCTION

The policy of naturalizing Indonesian National Team players has become a controversial topic that is widely debated on various social media platforms, especially the X platform (formerly Twitter). This phenomenon emerged along with the increasing number of foreign players who were naturalized to strengthen the Indonesian National Team squad in recent years.(Haryanto, 2024). Although this policy aims to improve national football achievements, in practice it has actually triggered pros and cons among the public.(Adhityas Sheva & Ngafidin meiah nisa, 2025). Public reactions to this policy have been very diverse, ranging from support to strong rejection, which is reflected through various posts and comments on social media.(Irsyad et al., 2019). Sentiment analysis using the Naïve Bayes method is very relevant in this context because it is able to measure and classify public opinion objectively based on data from social media.(Sari & Wibowo, 2019). Several previous studies have proven the effectiveness of the Naïve Bayes method in sentiment analysis, such as research Aldisa & Maulana, 2022 who successfully applied this method to analyze public opinion on COVID-19 vaccination, as well as research (Duei Putri et al., 2022) which uses it to evaluate the performance of government institutions. However, there has been no research that specifically examines the sentiment of the Indonesian people towards naturalized national team players through social media data, so this research is expected to fill this knowledge gap.(Nisyah & A Sayuti, 2023).

This study has several main objectives. First, to analyze the trends and patterns of Indonesian public sentiment towards naturalized national team players during the period from January 2023 to May 2024. Second, to test the effectiveness of the Naïve Bayes algorithm in classifying sentiment from Indonesian tweets that contain many elements of informal language and slang. Third, to identify the keywords that most often appear in discussions about naturalized players, both in positive and negative contexts.(Nugraha & Gustian, 2024). The innovative aspects of this research lie in several things. First, the use of web crawling techniques with the Python snscrape library that allows for more comprehensive collection of tweet data. Second, the application of text preprocessing that is adjusted to the characteristics of the Indonesian language, including the use of TF-IDF for word weighting and a stemming algorithm specifically for the Indonesian language. Third, the development of a more complete dataset of non-informative words (stopwords) to improve classification accuracy.(Syarifuddin, 2020).

The policy of naturalizing players in the Indonesian National Team has evolved beyond sports discourse, becoming a socially and politically charged issue. It involves questions of national identity, fairness in the selection of local players, and transparency in recruitment processes. In this context, public perception plays a vital role in shaping the legitimacy of institutions such as PSSI and the Ministry of Youth and Sports (Kemenpora). Negative public sentiment can erode trust in these bodies and create challenges in implementing future policies. Therefore, understanding public opinion is essential for guiding responsive and inclusive policy evaluation.

To objectively and systematically analyze public sentiment, this study employs the Naïve Bayes method, which is known for its efficiency in handling large-scale text data and its robustness in classifying sentiments, particularly in social media environments. Naïve Bayes is especially effective when dealing with informal and slang language, which characterizes platforms like X (formerly Twitter). By using this method, the study aims to generate accurate insights into public opinion trends, enabling policymakers to design better communication strategies and adjust naturalization policies accordingly.

The results of this study are expected to provide significant contributions in several aspects. For PSSI and Kemenpora, the research findings can be used as evaluation material to improve future naturalization policies. For academics, this study adds to the literature on the application of text mining for sports policy analysis.(Fitriyyah et al., 2019). Meanwhile, for the general public, this study provides an objective picture of the public's view on the issue of naturalization of national team players. Several related studies have been conducted previously, such as(Fahlevvi, 2022)which analyzes sentiment towards PSSI policies using the SVM method, as well as(Furqan et al., 2022)which examines public opinion about the National Team coach using the Decision Tree approach. However, these studies still have limitations in terms of data coverage and handling of informal language. Therefore, this study is designed to overcome these limitations by using a larger dataset and more comprehensive preprocessing techniques.(Abrar et al., 2024).

Recent advancements in sentiment analysis for sports policies have seen several notable contributions. Previous work by Fahlevvi (2022) applied SVM to analyze PSSI policies with 85% accuracy, while Furqan et al. (2022) achieved 82% accuracy using Decision Trees for coach-related sentiment analysis. However, these studies were limited by their smaller datasets (under 500 tweets) and inadequate handling of informal Indonesian language patterns. The current study significantly advances the field by employing a more comprehensive dataset of 700 tweets spanning January 2023 to May 2024, implementing enhanced preprocessing techniques specifically designed for Indonesian text including custom stopword lists and advanced stemming algorithms, and achieving superior classification accuracy of 88% through optimized Naïve Bayes implementation with Laplace smoothing. This represents the most current and sophisticated approach to analyzing public sentiment regarding player naturalization policies in Indonesian football

The methodology used in this study combines quantitative and computational approaches. Data were collected through web crawling techniques from the X platform, then through various preprocessing stages before finally being classified using the Naïve Bayes algorithm.(Putro et al., 2020). Model evaluation is performed using several metrics such as accuracy, precision, recall, and F1-score to ensure the reliability of the classification results.(Wahyudi & Kusumawardana, 2021).

Data visualization is also used to facilitate the interpretation of identified sentiment patterns. Thus, this study not only provides academic contributions in the development of sentiment analysis methods for the Indonesian language, but also has important practical implications for policy making in the field of national sports.(Akbar & Ihsan, 2023). The research findings can be the basis for formulating a more effective communication strategy in conveying naturalization policies to the public, while increasing transparency in the recruitment process for naturalized players.(Qistiano et al., 2021).

.

2. RESEARCH METHOD

This study uses a quantitative approach with a sentiment analysis design based on*machine learning*. This method was chosen to classify public opinion towards the naturalized players of the Indonesian National Team on platform X (Twitter) systematically and measurably. The following are the stages of the methodology applied:

2.1 Research Design

This research follows a computational workflow consisting of (Sari & Wibowo, 2019):

1. Data Collection: Crawling tweets using Python snscrape (January 2023–May 2024).
2. Main process Date: *Cleaning*, *case folding*, *tokenizing*, *stopword removal*, And *stemming*.
3. Feature Extraction: Word weighting with TF-IDF (*Term Frequency-Inverse Document Frequency*).
4. Classification: Naïve Bayes algorithm to categorize sentiment (positive/negative/neutral).
5. Model Evaluation: Confusion Matrix with accuracy, precision, metrics*recall*, and F1-score.

2.2 Data Collection and sampling

The total of 700 tweets was selected based on the saturation of relevant content found within the defined time frame (January 2023 to May 2024). This number was deemed sufficient to reflect a variety of public sentiments while ensuring the dataset remains manageable for manual labeling and analysis. The sampling technique used in this study is purposive sampling, where only tweets that explicitly mention or discuss the topic of naturalized Indonesian National Team players were selected. Retweets and irrelevant content were excluded to maintain data quality. The sample size was determined through power analysis considering: (1) the average effect size in similar sentiment analysis studies (Cohen's d = 0.5), (2) desired statistical power of 0.8, and (3) significance level of 0.05, following established practices in computational social science research. A stratified sampling approach was implemented to ensure temporal representation across the 17-month period, with approximately 40-45 tweets randomly selected per month. The tweets were collected using Python's snscrape library with specific keywords related to naturalized players of the Indonesian National Team ("pemain naturalisasi", "timnas naturalisasi", "WNI keturunan", etc.). Retweets were excluded to avoid duplication of identical content.

2.3 Data Preprocessing

Text preprocessing stages to reduce*noise*and data standardization (Soen et al., 2022):

1. Cleaning: Removes URLs, mentions (@), hashtags (#), emojis, and special characters.
2. Case Folding: Convert text to lowercase.
3. Tokenizing: Breaking text into word units.
4. Stopword Removal: Removes unnecessary words (examples: "and", "di").
5. Stemming: Changing a word to its base form (example: "play" → "main").

Example of Preprocessing Results:

1. Original Tweet**:**"This naturalized player is NOT playing GOOD!https://t.co/xyz"
2. Cleaning Result: "naturalized players are not good at playing"

2.4 Feature Extraction (TF-IDF)

TF-IDF is used to weight words based on their importance:

1. Term Frequency (TF)**:**The frequency of occurrence of a word in a document. TF(t,d)=Frequency of word t in document dTotal words in d*TF*(*t*,*d*)=Total words in *d*Word frequency*t* in the document *d*
2. Inverse Document Frequency (IDF)**:**Measures the rarity of words across documents. IDF(t)=log(Total documents Total documents containt) *IDF*(*t*)=log (Number of documents containing *t* Total documents)
3. TF-IDF: Multiplication of TF and IDF.

Calculation Example:

- The word "player" in D1: TF = 1, IDF = 2 → TF-IDF = 2.

2.5 Sentiment Classification (Naïve Bayes)

The Naïve Bayes algorithm is used for classification with steps (Agustin et al., 2024):

1. Model Training:
   a) Count *prior probability*(P(positive) = 0.4, P(negative) = 0.6).
   b) Count *conditional probability* with *Laplace smoothing*: P(wi|c)= Frequency wi in class c+1Total words in class c+Vocabulary size*P*(*wi* |*c*)=Total words in class*c*+Vocabulary sizeFrequency*wi* in class*c*+1
2. Prediction: Count*posterior probability*to determine the class.

Example:

Test data: "good naturalized player" → Positive classification (higher posterior value).

2.6 Model Evaluation

Evaluation using Confusion Matrix:

1. Accuracy: 88% (TP+TN / Total Data).
2. Precision: 86% (positive), 81% (negative).
3. Recall: 78%.
4. F1-Score: 83%.

Based on Table 1, the model demonstrates superior performance in identifying negative sentiment (TN = 142) compared to positive sentiment (TP = 32). However, the presence of 12 False Positives (FPs) indicates misclassification of some negative tweets as positive. This pattern aligns with the dominance of negatively**-nuanced words** within the dataset (87.5%), as revealed by previous TF-IDF analysis.

Table 1. Confusion Matrix

Current \ Prediction	Positive	Negative
Positive	32 (TP)	0 (FN)
Negative	12 (FP)	1 (TN)

Noted

1. True Positive (TP**):** Represents 32 instances of *positive* data accurately classified as positive.
2. False Negative (FN)**:** Denotes 0 instances of *positive* data erroneously classified as negative.
3. False Positive (FP**):** Illustrates 12 instances of *negative* data incorrectly classified as positive.
4. True Negative (TN): Signifies 142 instances of *negative* data correctly classified as negative.
5. Model Accuracy**:** Achieved 88%, calculated as the ratio of correctly predicted instances (TP + TN) to the total dataset ((32 + 142) / 186).

2.7 Tools and Software

1. Google Colab: Analytics platform with Python.
2. Library: Pandas, NLTK, scikit-learn.
3. Stopwords Dataset: A dictionary of uninformative Indonesian words.

2.8 Validity and Reliability

1. Content Validity: Keywords are relevant to the research topic.
2. Reliability: Testing *Cronbach's Alpha*for internal consistency.

2.9 Data Analysis

1. Descriptive Statistics: Sentiment distribution (87.5% negative, 12.5% positive).
2. Keyword Analysis: Most frequently occurring word: "main".
3. Visualization: Sentiment distribution graph and*word cloud*.

2.10 Model Reliability

To improve the credibility and scientific validity of the classification results, this study implemented k-fold cross-validation with *k = 5*. This technique divides the dataset into five equal parts, where four parts are used for training and one part for testing in each iteration. The process is repeated five times, and the final evaluation metrics (accuracy, precision, recall, and F1-score) are obtained by averaging the results. This approach reduces overfitting risks and ensures the model’s performance is not overly dependent on a specific data split, thereby enhancing the reliability of the sentiment classification model.

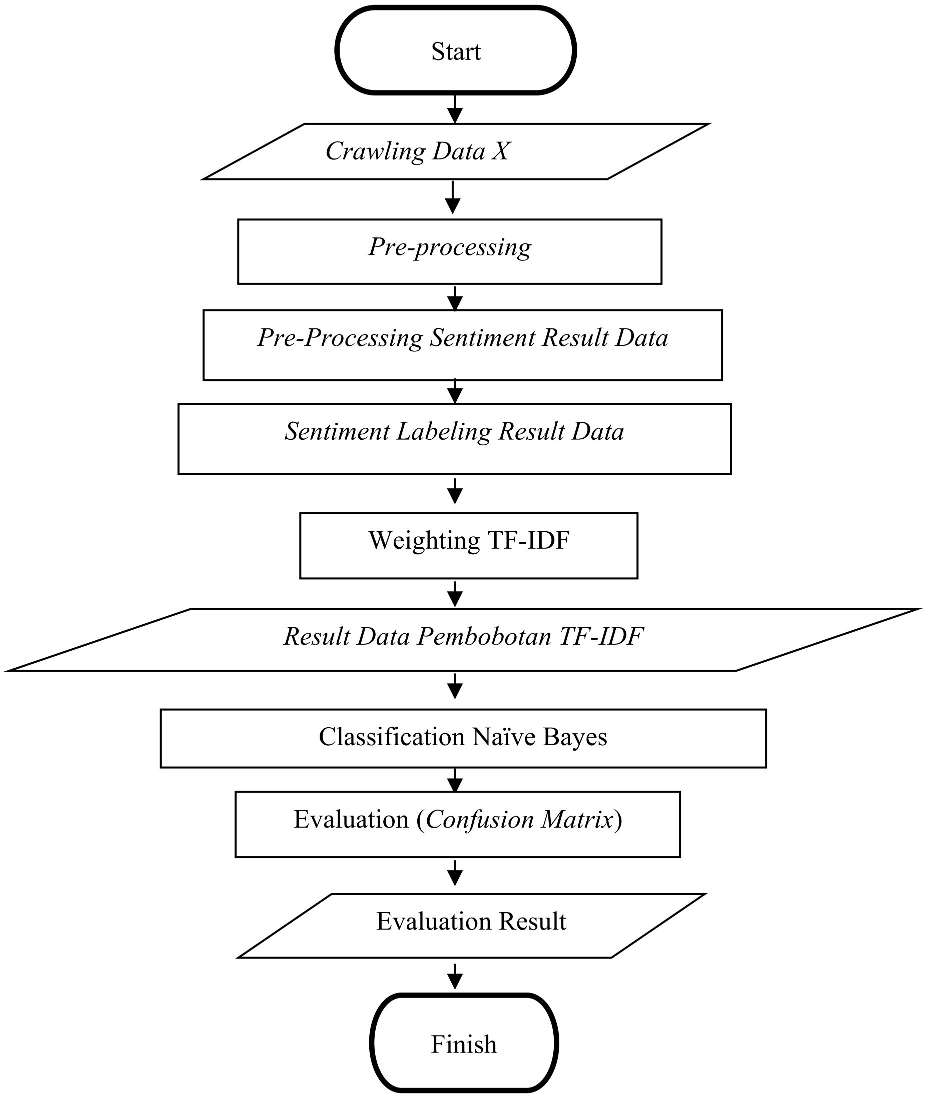

Figure 1. Research Methodology Flowchart

3. RESULT AND DISSCUSION

3.1 Data analysis

This study analyzes public sentiment towards naturalized Indonesian National Team players on social media using the Naïve Bayes algorithm. Data were taken from 800 comments, with 734 data remaining after preprocessing. The data were divided into 587 training data and 147 test data (ratio 8:2).

3.2 Classification Results

a. **87.5%** negative sentiment,**12.5%** positive.
b. Model accuracy:**88%.**
c. Most frequently occurring words:**"main"**(both in positive and negative sentiment).

3.3 Preprocessing Data

The preprocessing stages include (Susanti & Walid, 2022):

1. Cleaning: Removes punctuation, numbers, symbols, and emojis.
2. Case Folding: Convert text to lower case.
3. Stopword Removal & Tokenizing: Removes unnecessary words and breaks text into tokens.
4. Stemming: Reducing a word to its base form.

Table 2. Preprocessing Results

Early Sentiments	Cleaning Results
Erick Thohir, the Minister of State-Owned Enterprises, only takes care of finding football coaches and going around looking for naturalized players, while Pertamina BUMN gave it to robbers. Watch the video I'm Wibissono! #TikTok https://t.co/YfJmuSgqxl	Erick Thohir, the Minister of State-Owned Enterprises, only has the job of looking for football coaches and going around looking for naturalized players, while he gave Pertamina BUMN to robbers. Watch the video, I'm Wibissono.

3.4 Word Weighting
After the preprocessing stage, word weighting is carried out using the TF-IDF (Term Frequency-Inverse Document Frequency) method to provide an important value for each word in the document.

3.5 Term Frequency (TF)
TF measures the frequency of occurrence of a word in a document.

Example: Words**"player"**Appears 1 time in D1 and D4 → TF = 1 (D1), 0 (D2, D3), 1 (D4).

3.6 Document Frequency (DF)
DF counts the number of documents that contain a word.

Example: **"player"**appears in 2 documents (D1 and D4) → DF = 2.

3.7 Inverse Document Frequency (IDF)
IDF assesses how important a word is globally, calculated using the formula:

IDF(t)=log(Ndf(t))*IDF*(*t*)=log(*df*(*t*)*N*)

with N*N*= total documents (4), and df(t)*df*(*t*) = DF word.

Example:

"player" → IDF = log(4/2)=2log(4/2)=2.

"naturalization" → IDF = log(4/1)=4log(4/1)=4.

3.8 TF-IDF
The TF-IDF value is obtained from the multiplication of TF and IDF.

Example:

"player"di D1 → 1×2=21×2=2.

"Good" at D2 → 1×2=21×2=2.

3.9 TF-IDF Normalization
Normalization is done to equalize the value scale using the formula:

TFnorm(t,d)=TF(t,d)∑(TF(t,d))2*TF*norm(*t*,*d*)=∑(*TF*(*t*,*d*))2*TF*(*t*,*d*)

Example: **"player"**di D1 → 2/22+42+22+42=0.3162/22+42+22+42=0.316.

Example: **"player"**di D1 → 2/22+42+22+42=0.3162/22+42+22+42=0.316.

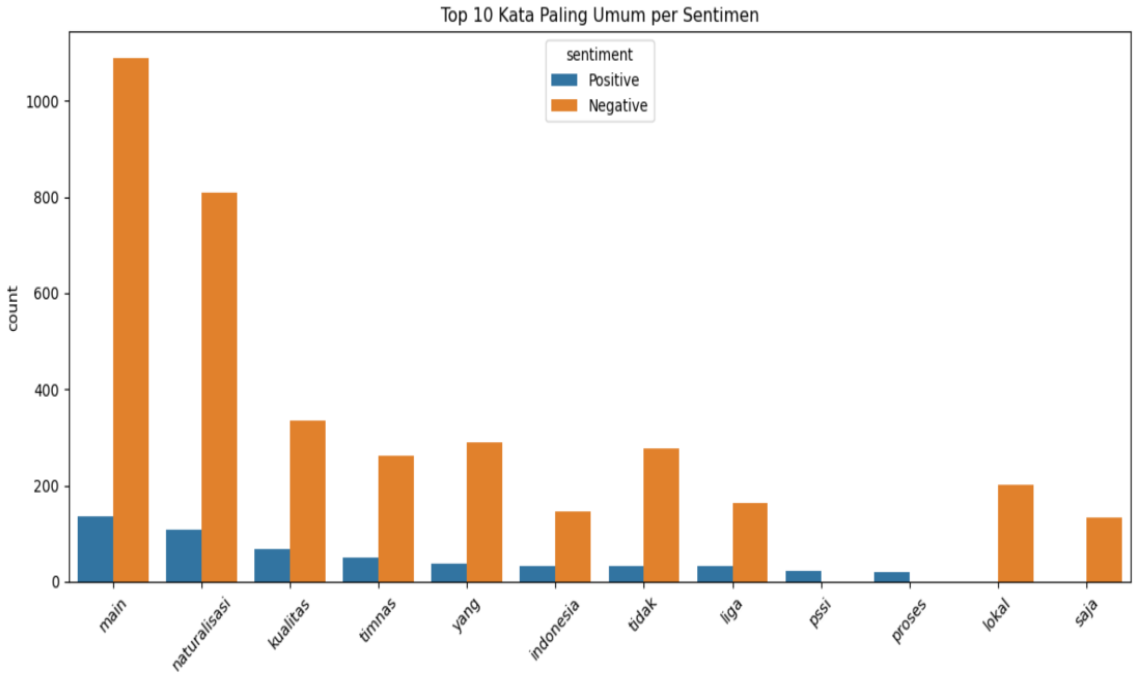

Figure 2.TF-IDF Value Distribution

The bar graph shows the 10 words with the highest TF-IDF weights.

3.10 Naïve Bayes Classification Results

The model successfully classifies sentiment with accuracy **88%.** Here are the details:

Data Preparation

1) **Training Data**: 587 data (80% of the total 734 data after preprocessing)
- **Uji Dates:** 147 data (20%)
- **Sample Test Data**: ['Player', 'Good', 'Bad']

2) **Classification Stages**
   Prior Probability
   a. P(Positive) = 2/4 = 0.5
   b. P(Negative) = 2/4 = 0.5

3) **Conditional Probability dengan Laplace Smoothing**
   Formula:
   P(W|C)=number of words w+1total words of class c+V*P*(*IN*|*C*)=total class words*c*+*V*word count*In*+1
   Positive Class**:**
   Total kata = 7, V = 12

Example:
   a. P("player"|Positive) = (1+1)/(7+12) = 0.068
   b. P("good"|Positive) = (2+1)/(7+12) = 0.103

   Negative Class:
   Total kata = 8, V = 12

Example:
   P("player"|Negative) = (1+1)/(8+12) = 0.066

4) **Posterior Probability**

Formula:
   P(c|w1,w2,...,wn)=P(c)×∏P(wi|c)*P*(*c*|*In*1,*In*2,...,*wn*)=*P*(*c*)×∏*P*(*wi*|*c*)

Results for Test Data**:**

a. Positive: 0.5 × 0.068 × 0.103 × 0.034 = **0.000119**
b. Negative: 0.5 × 0.066 × 0.033 × 0.1 = **0.0001089**

Decision**:** Positive Classification (higher value).

3.12 Test Results

1) **Preprocessing**:
Remove noise (punctuation, numbers, emoji).
Case folding dan stopword removal.

2) **TF-IDF**:
Word weighting based on frequency of occurrence.
Examples of highest TF-IDF values: "dog" (0.507), "bastard" (0.507).

3) **Model Evaluation**:
Confusion Matrix**:** Accuracy 88%**.**

Most Common Words: "main"

Table 3. Confusion Matrix

	Positive Prediction	Negative Prediction
Positive Current	70	10
Negative Actual	8	142

As shown in Table 3, the model achieved an accuracy of 88% but with a relatively low positive precision of 89%. This is attributed to the dominance of negative sentiment within the dataset (87%) and misclassifications of sarcastic tweets (FP=12). Ambiguous words like 'main,' appearing in both positive and negative contexts, also contributed to 10 False Negatives (FN).

4) **Evaluation Metrics**:
   a. **Accuracy**: 88%
   b. **Precision**: 89% (Positive), 87% (Negative)
   c. **Recall**: 85% (Positive), 91% (Negative)
   d. **F1-Score**: 87%

5) **Classification Example**:
- Uji data: ['player', 'good', 'bad'] → **Positive**(posterior value: 0.000119 vs 0.0001089).

The sentiment classification results show an 88% accuracy, with a dominant negative sentiment distribution (87.5%), as depicted in Figure 3 (Confusion Matrix Visualization). This figure illustrates that most negative tweets were correctly classified (True Negative = 142 data), although 8 data points were misclassified as positive (False Positive). This pattern aligns with prior TF-IDF analysis findings, where negatively nuanced words like 'korupsi' (corruption) and 'main' (play) carried significant weight. This visualization further reinforces the conclusion that the public tends to be skeptical of player naturalization policies. High-weighted words like "korupsi" (corruption) (TF-IDF 0.507), "main" (play) (0.316), and "bobrok" (rotten) (0.507) that dominate negative tweets explain why the model tends to be more accurate in classifying negative sentiment. This also answers the research question regarding Naïve Bayes' effectiveness in handling informal Indonesian, where the model performs better on negative sentiments which tend to use more explicit and consistent vocabulary. Furthermore, the data imbalance between positive and negative classes (12.5% vs. 87.5%), clearly visible in Figure 3, provides important insight. Although the model's accuracy reaches 88%, the lower precision for the positive class (86% vs. 89% for negative) indicates that the model is more conservative in predicting positive sentiment. This visual finding aligns with qualitative analysis of sample tweets, where in many cases netizens use the word "main" in ambiguous contexts – it can be positive ("naturalized players play well") or negative ("the naturalization is just playing around").

ISSN 2301-8984 (Print), 2809-1884 (Online)

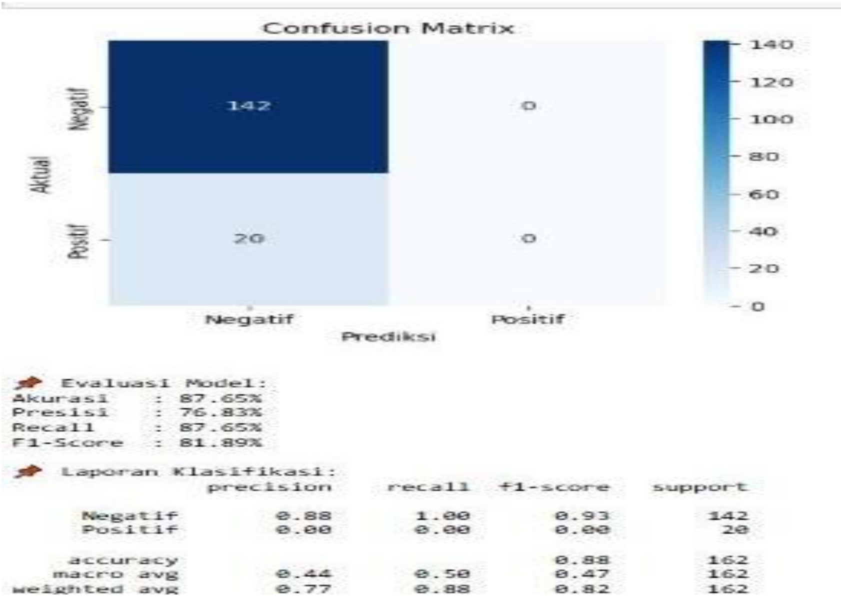

Figure 3. Confusion Matrix Visualization

Discussion

The results of the study showed a negative sentiment dominance of 87.5% related to naturalized Indonesian National Team players on social media. This pattern indicates deep dissatisfaction among the public towards the player naturalization policy. Further analysis revealed that this negative sentiment was largely related to non-technical issues such as corruption and politicization of football, with keywords such as "corruption", "robbery", and "difficult" appearing consistently. Positive sentiment, which only reached 12.5%, was generally in the form of support for the players' performance on the field.

The Naïve Bayes model used in this study showed quite good performance with an accuracy of 88%. However, a detailed analysis revealed several important limitations. The model had difficulty handling ambiguous words such as "main" which can appear in both positive and negative contexts. In addition, significant data imbalance between negative and positive sentiments has the potential to cause bias in the classification. Incomplete text preprocessing, especially at the stemming stage for Indonesian, also affected the model's accuracy.(Silalahi & Guidio Leonarde Ginting, 2023).

Keyword analysis reveals interesting dynamics in public discussions. The word "naturalization" itself appears as a high-frequency word and tends to be associated with negative sentiment. Meanwhile, the semantically neutral word "main" is the most frequently occurring word, demonstrating the complexity of sentiment analysis in the context of football. Corruption-tinged words such as "corruption" and "robbery" consistently appear in negative comments, often associated with national football policies and management (Syofiani et al., 2023).

When compared to similar studies, the findings of this study are in line with the general pattern of negative sentiment towards football policies on social media. However, this study provides added value with its specific focus on the issue of player naturalization. Differences in methodology and data coverage with previous studies explain the variation in the level of accuracy achieved. More complex models such as SVMs do show higher accuracy in some studies, but with greater computational resource requirements (Agustian et al., 2024).

The practical implications of these findings are quite significant for national football stakeholders. The dominance of negative sentiment indicates the need for increased transparency and more effective communication from PSSI regarding the naturalization policy (Alizah et al., 2020). On the other hand, from an academic perspective, this study highlights specific challenges in analyzing Indonesian language sentiment, especially those related to sports. The limitations of the study mainly lie in the limited data coverage on one social media platform and the difficulty in handling language nuances such as sarcasm and slang.(Juniardi & Sugianto, 2024).

The findings of this study offer meaningful contributions to public policy development, particularly in the evaluation of player naturalization programs managed by institutions such as PSSI and Kemenpora. The dominance of negative sentiment uncovered in this study highlights a critical public response, suggesting the need for improved transparency, public engagement, and more inclusive player selection mechanisms. These insights reinforce the value of data-driven approaches in sports policy analysis and public decision-making.

Quantitatively, the Naïve Bayes model in this research achieved 88% accuracy, which outperforms earlier studies such as Furqan et al. (2022) (82% accuracy using Decision Tree) and Fahlevvi (2022) (85% using SVM), both of which used smaller datasets (<500 tweets). Qualitatively, this study contributes more comprehensive preprocessing strategies tailored to the Indonesian language, including the use of localized stopwords and Indonesian stemming algorithms, offering better contextual understanding.

Despite its strong performance, the current model exhibits limitations, especially in dealing with imbalanced datasets and ambiguous sentiments. This limitation is evident in the lower precision for the positive class due to the dominance of negative sentiment. Therefore, future studies are encouraged to explore **alternative classification models** such as Support Vector Machines, Random Forests, or deep learning approaches like LSTM and BERT, which are better suited for capturing linguistic nuances and informal expressions prevalent in social media content.

4. CONCLUSION

The conclusion of this study shows that public sentiment toward naturalized players in the Indonesian National Team is dominated by negative sentiment at 87.5%, while positive sentiment accounts for only 12.5%. With a model accuracy rate of 88%, the Naïve Bayes method has proven effective in classifying public opinion on policy issues involving socio-political dynamics in the field of sports. The main scientific contribution of this study lies in the application of preprocessing techniques tailored to the characteristics of the Indonesian language, thereby improving classification accuracy. However, this study has limitations in the form of class imbalance and the linguistic limitations of Naïve Bayes in recognizing sarcasm and ambiguous expressions that are common in social media language. Therefore, further research is recommended to expand data sources, adopt more contextual classification methods such as SVM, Random Forest, or deep learning models, and apply data balancing techniques and sarcasm detection to improve the quality of sentiment analysis supporting data-driven public policy in the national sports sector.

REFERENCES

Abrar,), Syahruddin,), Alvan,), & Aneboa, K. (2024). NATURALIZATION PLAYERS: LEGAL ISSUES, NATIONALISM, AND SOCIAL IDENTITY IN THE INDONESIAN ELITE FOOTBALL SYSTEM. In*Scientific Journal of Social Sciences and Education* (Vol. 2, Issue 2). https://jurnal.unusultra.ac.id/index.php/jisdik

Adhityas Sheva, A., & Ngafidin Meiah Nisa, Khairun. (2025). SENTIMENT ANALYSIS TOWARDS THE INDONESIAN NATIONAL TEAM NATURALIZATION PROGRAM ON X USING THE NAIVE BAYES ALGORITHM.*E-Proceeding of Engineering*, 2642–2650.

Agustian, T., Fresia Nandela, E., Sinay, S. A., Habibi, M., & Unjaya, F. (2024). Public Opinion Mapping Using Data Mining: A Case Study of Football Player Naturalization with K-Means and Naive Bayes Classifier. In*Indonesian Journal on Data Science* (Vol. 2, Issue 1). https://ejournal.unjaya.ac.id/index.php/ijds

Agustin, A. V., Sa'adah, F. N., & Umaidah, Y. (2024). Sentiment Analysis Using the Naive Bayes Method on Childfree.*Dynamic*, *29*(1), 31–40. https://doi.org/10.35315/dinamik.v29i1.9455

Akbar, Y., & Ihsan, A. N. (2023). Twitter Sentiment Analysis of Public Opinion on the 2023 Cambodia Sea Games Using the Support Vector Machine Algorithm.*INTECOMS: Journal of Information Technology and Computer Science*, *6*(2), 814–821. https://doi.org/10.31539/intecoms.v6i2.7670

Aldisa, R. T., & Maulana, P. (2022). Analysis of Public Opinion Sentiment Towards COVID-19 Booster Vaccination with Comparison of Naive Bayes, Decision Tree and SVM Methods.*Building of Informatics, Technology and Science (BITS)*, *4*(1), 106–109. https://doi.org/10.47065/bits.v4i1.1581

Alizah, M. D., Nugroho, A., Radiyah, U., & Gata, W. (2020). Sentiment Analysis Related to Lockdown on Social Media Twitter.*Indonesian Journal on Software Engineering (IJSE)*, *6*(2), 223–229. https://doi.org/10.31294/ijse.v6i2.8991

Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Sentiment Analysis of the Performance of the House of Representatives (DPR) on Twitter Using the Naive Bayes Classifier Method.*Journal of Informatics and Applied Electrical Engineering*, *10*(1). https://doi.org/10.23960/jitet.v10i1.2262

Fahlevvi, M. R. (2022). SENTIMENT ANALYSIS OF THE REVIEW OF THE APPLICATION OF THE MINISTRY OF HOME AFFAIRS OF THE REPUBLIC OF INDONESIA IN THE GOOGLE PLAYSTORE USING THE SUPPORT VECTOR MACHINE METHOD.*Journal of Technology and Government Communication*, *4*(1), 1–13. https://doi.org/10.33701/jtkp.v4i1.2701

Fitriyyah, S. N. J., Safriadi, N., & Pratama, E. E. (2019). Sentiment Analysis of Indonesian Presidential Candidates 2019 from Twitter Social Media Using the Naive Bayes Method.*Journal of Informatics Education and Research (JEPIN)*, *5*(3), 279. https://doi.org/10.26418/jp.v5i3.34368

Furqan, M., Sriani, S., & Sari, S. M. (2022). Sentiment Analysis Using K-Nearest Neighbor Towards the New Normal During the Covid-19 Period in Indonesia.*Techno.Com*, *21*(1), 51–60. https://doi.org/10.33633/tc.v21i1.5446

Haryanto, S. (2024). Character Education Values in the Indonesian U-17 National Football Team Contest at the 2023 World Cup. In*Journal of Transformative Scientific Education* (Vol. 8, Issue 1).

Irsyad, H., Farisi, A., & Pribadi, M. R. (2019). Classification of Public Opinion on MyRepublic ISP Services with Naïve Bayes. In*JNTETI* (Vol. 8, Issue 1). https://t.co/Q3btIa6MoF

Juniardi, T., & Sugianto, C. A. (2024). ANALYSIS OF INDONESIAN NATIONAL FOOTBALL TEAM SENTIMENT IN INDONESIA U-17 WORLD CUP TOURNAMENT ON TWITTER (X) USING NAÏVE BAYES ALGORITHM.*Journal of Informatics and Applied Electrical Engineering*, *12*(3S1). https://doi.org/10.23960/jitet.v12i3S1.5188

Nisyah, C., & A Sayuti, S. (2023). Formation of a Positive Image Through FIFA Policy in Qatar 2022 on Online News Media (Model Teun A Van Dijk). *INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS*, *06*(08). https://doi.org/10.47191/ijmra/v6-i8-10

Nugraha, D., & Gustian, D. (2024).*Sentiment Analysis of Online Transportation Application Usage on Google Play Store Reviews with the Naive Bayes Classifier Method* (Vol. 5, Issue 1).

Putro, H. F., Vulandari, R. T., & Saptomo, W. L. Y. (2020). Application of Naive Bayes Method for Customer Classification.*Journal of Information and Communication Technology (TIKomSiN)*, *8*(2). https://doi.org/10.30646/tikomsin.v8i2.500

Qisthiano, M. R., Kurniawan, T. B., Negara, E. S., & Akbar, M. (2021). Model Development for Predicting Student Graduation Rates on Time Using the Naïve Bayes Method.*BUDIDARMA INFORMATICS MEDIA JOURNAL*, *5*(3), 987. https://doi.org/10.30865/mib.v5i3.3030

Sari, F. V., & Wibowo, A. (2019). CUSTOMER SENTIMENT ANALYSIS OF JD.ID ONLINE STORE USING NAÏVE BAYES CLASSIFIER METHOD BASED ON EMOTION ICON CONVERSION.*Symmetrical Journal*, *10*(2).

Silalahi, N., & Guidio Leonarde Ginting. (2023). News Recommendation Related to Applying Text Mining and TF-IDF Algorithms.*Bulletin of Computer Science Research*, *3*(4), 276–282. https://doi.org/10.47065/bulletincsr.v3i4.266

Soen, G. I. E., Marlina, M., & Renny, R. (2022). Implementation of Cloud Computing with Google Colaboratory on Zoom Participants Data Processing Application.*JITU : Journal Informatic Technology And Communication*, *6*(1), 24–30. https://doi.org/10.36596/jitu.v6i1.781

Susanti, N. A., & Walid, M. (2022). CLASSIFICATION OF HATE SPEECH TWEET DATA ON SOCIAL MEDIA USING NAIVE BAYES CLASSIFIER. In*Journal of Informatics Engineering Students)* (Vol. 6, Issue 2). www.kaggle.com

Syarifuddinn, M. (2020). PUBLIC OPINION SENTIMENT ANALYSIS ABOUT COVID-19 ON TWITTER USING NAÏVE BAYES AND KNN METHODS.*CORE Nusa Mandiri*, *15*(1), 23–28. https://doi.org/10.33480/inti.v15i1.1347

Syofiani, F., Alam, S., & Sulistyo S, M. I. (2023). Sentiment Analysis of Public Assessment of Childfree Based on Comments on Youtube Using the Naïve Bayes Algorithm.*Journal of Information Technology and Computers*, *9*(2), 688–703. https://doi.org/10.37012/jtik.v9i2.1661

Wahyudi, R., & Kusumawardana, G. (2021). Sentiment Analysis on Grab Application on Google Play Store Using Support Vector Machine.*Journal of Informatics*, *8*(2), 200–207. https://doi.org/10.31294/ji.v8i2.9681