JEECS (Journal of Electrical Engineering and Computer Sciences)
Vol. 11, No. 2, December 2026, pp. 146-156
e-ISSN: 2579-5392 p-ISSN: 2528-0260

Data-Driven Sentiment Analysis of Public Opinion on National Football Team Coaching Using Naïve Bayes

Luki Ardiantoro [1], Nisa Ayunda [2], Muhammad Irfan [3]

[1,2]Informatics, Universitas Bina Sehat PPNI, Jl. Raya Jabon km 6, Mojokerto, Indonesia

[3]Informatics, Universitas Muhammadiyah Malang, Jl. Raya Tlogomas 246, Malang, Indonesia

Article Info

Article history:
Received: 2 May 2026
Revised: 6 July 2026
Accepted: 14 July 2026

Keyword:
CRISP-DM
Football coaching evaluation
Naïve bayes classifier
Sentiment analysis
Social media analytics

Abstract

The appointment of a national football team coach often generates intense public discourse, particularly on social media, where opinions are expressed rapidly and at scale. However, such evaluations are typically subjective and lack a systematic, data-driven foundation. This study proposes a quantitative approach to assess public sentiment toward the Indonesian national team coach using sentiment analysis based on the Naïve Bayes classifier within the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. A dataset of 2,000 Indonesian-language tweets was collected from social media X and processed through data cleansing, normalization, tokenization, and sentiment labeling. The Naïve Bayes model was applied to classify sentiment polarity, and its performance was evaluated using accuracy, precision, recall, and ROC–AUC metrics. The results show that the model achieves an accuracy of 94.17%, with precision of 92.86% and recall of 94.55%, indicating strong classification performance. However, the AUC value of 0.686 suggests moderate discriminative ability. The findings reveal that public sentiment toward the coach is predominantly positive, reflecting increased supporter confidence. This study demonstrates the effectiveness of data-driven approaches for objective evaluation in sports analytics and highlights opportunities for improvement using advanced machine learning techniques.

Corresponding author:
Luki Ardiantoro, ipan.ardianto@gmail.com

DOI: https://doi.org/10.54732/jeecs.v11i2.4

This is an open access article under the CC–BY license.

1. Introduction

The rapid advancement of digital technologies has significantly transformed the landscape of modern sports, particularly football, where data-driven approaches are increasingly used to enhance performance evaluation and strategic decision-making[1]. In this context, the role of a national football team coach extends beyond tactical leadership, encompassing public perception, media scrutiny, and stakeholder expectations. The appointment of a national team coach often triggers widespread public discourse, especially on social media platforms, which have become primary channels for supporters to express opinions, emotions, and evaluations in real time.

Despite the growing availability of large-scale digital data, the assessment of coaching performance and selection decisions remains largely subjective, frequently influenced by match outcomes, media narratives, and individual biases. Such subjectivity can lead to polarized opinions and controversy, limiting the ability of stakeholders to derive objective insights. Consequently, there is a need for systematic, data-driven methods that can capture and quantify public sentiment in a structured and reproducible manner. Sentiment analysis, as a subfield of natural language processing, offers a robust framework for identifying and classifying opinions expressed in textual data, enabling researchers to uncover patterns and trends in public perception[2].

Recent developments in sports analytics highlight the importance of integrating quantitative methods with contextual understanding to support decision-making processes[3], [4]. Social media platforms, such

as X (formerly Twitter), provide a rich source of unstructured data that reflects spontaneous public reactions to key events, including coaching appointments. These platforms not only amplify public voices but also shape collective opinion, which can influence organizational strategies, fan engagement, and commercial value. Understanding supporter sentiment is therefore critical, as positive public perception can enhance team morale, strengthen fan loyalty, and increase marketability through sponsorships, broadcasting rights, and branding opportunities[2][5]. Another research study analyzes fan comments on social media platforms to evaluate player perception. The dataset consists of 750 manually labeled comments in four different languages. The experiment utilized three sentiment analysis models: VADER, XLM-T, and GPT-4o-mini, achieving an accuracy of 76% [6].

Among various machine learning techniques, the Naïve Bayes classifier remains a widely used method for text classification due to its simplicity, computational efficiency, and relatively strong performance, particularly in high-dimensional datasets such as social media text. Although the assumption of feature independence may not fully hold in natural language, Naïve Bayes has demonstrated robustness in sentiment classification tasks [7][8]. Furthermore, the Cross-Industry Standard Process for Data Mining (CRISP-DM) provides a structured framework that guides the end-to-end analytical process, from problem definition to deployment, ensuring methodological rigor and reproducibility. The algorithm is called "naive" because it makes the simplifying assumption that all features in a dataset are completely independent of one another. For example, in a fruit classifier, a red color and a round shape are treated as independent traits, even though they might actually be correlated in an apple. While this assumption is rarely true in real life, it drastically simplifies calculations and often results in high accuracy.

This study aims to develop a data-driven framework for analyzing public sentiment toward the Indonesian national football team coach using social media data. By leveraging sentiment analysis with the Naïve Bayes algorithm within the CRISP-DM framework, this research seeks to provide an objective evaluation of public opinion that complements traditional performance assessments. The findings are expected to contribute both theoretically and practically by demonstrating the applicability of data mining techniques in sports analytics and offering actionable insights for decision-makers in football governance. The problem formulation is how a quantitative approach based on data mining can be used to objectively analyze the selection of the national team coach to foster support and affection from the supporters toward the national team. With high supporter support, a positive atmosphere will be created, and the market value of the national team will increase. With high supporter support, it will automatically create a positive atmosphere, and the market value of the national team will increase. This is very economically beneficial for the national team's commercial activities such as: advertising, apparel, product promotion, television broadcasting rights, branding, market value, etc. The urgency of this research is twofold: 1) the utilization of data-based technology has become an important factor in improving sports performance in various countries. However, the application in Indonesia, particularly for understanding formations and the dynamics of player roles, is still not optimal and is partial in nature. 2) The assessment of the national team's coach's performance is generally based on final results and subjective opinions, which often leads to controversy and does not provide a basis for objective and sustainable evaluation.

2. Research Methodology

2.1 Research Framework

This study adopts a data-driven approach to analyze public sentiment toward the Indonesian national football team coach by utilizing social media data and machine learning techniques. The overall research process follows the Cross-Industry Standard Process for Data Mining (CRISP-DM), a widely recognized framework that ensures systematic and iterative development of data mining projects [9][10].

Additionally, leveraging advanced deep learning methods, such as BiLSTM, and integrating large pretrained language models could further enhance the analytical depth, particularly for large datasets [11]. Furthermore, incorporating context-specific adjustments to capture nuanced language, including sarcasm and domain-specific jargon, could refine sentiment classification [8][5]. Future work could also investigate incorporating a neutral sentiment category to provide a more comprehensive understanding of public perception, moving beyond a simple positive or negative dichotomy [12]. Given the complexities of language, especially in informal social media contexts, future research should increasingly explore deep learning techniques, particularly Transformer-based models, for their superior ability to manage such intricacies in sentiment analysis [13]. Developing models with manual annotations could improve accuracy by capturing subtle human language nuances [14].

Table 1. The Gap Analysis This Research With Previous Studies

Previous Studies	Existing Findings	Research Gap	Contribution of This Study
Sentiment analysis on football naturalization programs (2025)	Focused on policy acceptance using Naïve Bayes	Did not analyze public opinion toward national team coach appointments	Examines supporter sentiment toward coach selection as a strategic football decision
Public policy sentiment analysis (2023–2025)	Emphasized classification accuracy	Limited discussion of sports governance implications	Connects sentiment analysis with football management and decision support
Deep learning-based sentiment analysis (BERT, BiLSTM, CRNN)	High predictive accuracy but computationally expensive	Limited practicality for institutions with constrained resources	Demonstrates that Naïve Bayes combined with CRISP-DM remains effective and reproducible
Sports analytics research	Mostly evaluates technical or tactical performance	Public perception rarely integrated into sports analytics	Incorporates supporter opinion as an additional dimension of sports analytics
Indonesian social media sentiment studies	Focus on sentiment classification	Lack of comprehensive analytical pipeline and reproducibility	Presents complete workflow from scraping to evaluation with CRISP-DM

Moreover, a richer lexicon that includes non-standard Indonesian and English terms, alongside the translation of emojis and emoticons, would minimize out-of-vocabulary words and enhance the applicability of models like multilingual BERT to social media texts [15][9]. Further advancements could involve the development of hybrid models that combine the strengths of traditional machine learning with deep learning approaches, potentially leading to more robust and accurate sentiment detection across diverse Indonesian social media content [16].

Further, the inclusion of multimodal data, such as audio and video from social media posts, could provide a more holistic understanding of user expressions, thereby enriching the sentiment analysis [17][18]. In addition, future studies could investigate transfer learning techniques within AI frameworks for sports analysis to better capture intricate sentiment patterns in social media data on football coaches and national team events, as demonstrated in analyses of Indonesian public reactions to coaches like Shin Tae-yong [19][20]. This could involve leveraging advanced techniques like BERT for nuanced language understanding [21] and exploring dynamic sentiment analysis to capture evolving public opinion [22][23]. Additionally, implementing network analysis alongside sentiment classification could reveal influential users and communities shaping public discourse around coaching appointments [24]. Table 1, show the gap analysis this research with previous studies. To address these gaps, this study develops a comprehensive data-driven framework for analyzing public opinion toward the appointment of the Indonesian national football team coach using 2,000 Indonesian-language posts collected from social media X. Unlike previous studies, this research integrates the CRISP-DM methodology with the Naïve Bayes classifier to provide a transparent, reproducible, and computationally efficient sentiment analysis framework. In addition to evaluating classification performance using multiple metrics, this study interprets sentiment patterns as indicators of supporter acceptance and potential implications for football governance, thereby extending sentiment analysis beyond algorithm evaluation toward practical decision support in sports management.

2.2 Research Design

This study adopts a data-driven approach to analyze public sentiment toward the Indonesian national football team coach by utilizing social media data and machine learning techniques. Football requires teamwork and team cohesion, so to facilitate assignments and role-sharing during matches, player position formations are used. The role of the coach as a strategist is crucial in determining the direction of the game and the team's harmony. The Indonesian national team appointed John Herdman on January 13, 2026. John Herdman is an English professional football coach who currently serves as the head coach of the Indonesian national team. He is widely known for his contributions to the Canadian national team as the first coach to lead both the women's and men's teams from the same country to the World Cup. The response of supporters regarding the selection of the national football team coach and how supporters feel satisfied with the reality. This serves as a fulfillment of expectations, with social media becoming a barometer of public opinion. Based on the national team's performance in each match, goals scored or conceded, and even failures can spark heated discussions [2] .

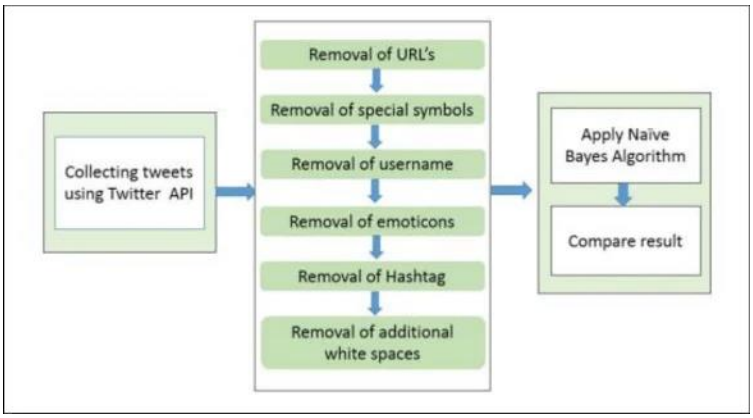

Figure 1. Research Design

The selection of foreign coaches for the Indonesian National Football Team often triggers significant reactions in the public sphere, especially on social media platforms that serve as the main outlet for Indonesian football supporters' expressions [25]. These changing interaction dynamics allow for the formation of massive and rapid public opinion, where virtual movements on social media have proven capable of exerting real pressure on football-related policies in Indonesia. This aligns with findings that virtual public spaces can generate public opinions that pressure both the government and related organizations[26].

Figure 1 shows Research Design using Naïve Bayes algorithm as classifier. This algorithm has advantages such as being simple, fast, and highly accurate. The Naïve Bayes Classifier method for text classification uses the attributes of words that appear in a document as the basis for classification. A research shows that although the assumption of independence between words in a document cannot be fully met, the performance of the Naïve Bayes Classifier in classification is relatively very good. In this study, several issues were identified, including the lack of classification of positive and negative sentiments regarding John Herdman and the unknown accuracy of the Naïve Bayes Classifier algorithm in classifying test data for sentiment analysis [22]. For future research, the use of additional algorithms such as Support Vector Machine or Random Forest can refine sentiment classification results, especially with diverse and large datasets. Furthermore, exploring advanced preprocessing techniques tailored for Bahasa Indonesia, alongside expanding the training dataset, could further enhance model performance and generalizability. This would enable a more nuanced understanding of complex linguistic structures and contextual inferences inherent in social media discourse, thereby improving the robustness of sentiment analysis models[27]. Moreover, incorporating dynamic feature extraction methods that adapt to evolving linguistic patterns within social media could offer a significant boost in classification accuracy [10].

The integration of deep learning architectures, such as Bidirectional Encoder Representations from Transformers or Long Short-Term Memory networks, might further refine the capture of semantic nuances and long-range dependencies often present in social media texts, thereby surpassing the limitations of traditional machine learning approaches. Finally, addressing the challenge of duplicate data, as noted by research in the context of presidential candidate sentiment analysis, is crucial for improving data effectiveness and model accuracy[7]. Moreover, exploring semantic relationships between words and incorporating sentiment lexicons specifically tailored for Indonesian slang could mitigate misclassifications, as demonstrated in studies addressing the complexities of social media language [15][7].

The novelty of this study lies in the integration of a structured CRISP-DM framework with the Naïve Bayes classifier to objectively evaluate public sentiment toward the appointment of a national football team coach using Indonesian social media data. Unlike previous studies that primarily emphasize algorithm comparison or policy-related sentiment, this research positions supporter sentiment as an evidence-based indicator for football governance and coaching evaluation. Furthermore, the study provides a reproducible end-to-end analytical workflow—from data collection and preprocessing to multi-metric evaluation (accuracy, precision, recall, and ROC-AUC)—thereby offering a practical framework for sports organizations seeking computationally efficient sentiment monitoring.

Table 2. Data Processing Stage & Description of Dataset

No	Stage Processing Data	Process Description	Amount of Data	Eliminated Data
1	Scraping Data	Data collection from the social media platform X using an API or scraper, based on research keywords.	2,000	-
2	Remove Duplication	Deleting tweets with identical content or identical retweets.	1,850	150
3	Cleaning Data	Remove URLs, mentions, hashtags, emoji, special characters, and empty data.	1,720	130
4	Language Selection	Selecting only Indonesian-language tweets	1,500	220
5	Relevant Selection	Deleting tweets that are irrelevant to the research topic	1,200	300
6	Sentiment Labeling	Manual labeling as positive, negative, and neutral	600	600*
7	Data Final Analysis	The dataset that has been labeled	600	-
8	Data Training (80%)	Used for model training	480	-
9	Data Testing (20%)	Used for model testing	120	-

2.3 Research Procedure

The data was obtained from X using Google Colab, with tweets taken in Indonesian using the keywords John Herdman and Timnas Indonesia, resulting in 2,000 tweets that have completed the text preprocessing stage. At this stage, the approach used is the Cross Industry Standard for Data Mining (CRISP-DM) method. CRISP-DM is a method that employs a data development process model widely used by experts to solve problems. Table 2 show detail of Data Processing Stage & Description, The data used in this research is qualitative, which characterizes or describes phenomena, can be observed, and recorded.

This research process refers to the six stages present in CRISP-DM, which are explained as follows (refer to Table 2):

a. Business Understanding, the purpose of the research is to determine public sentiment toward John Herdman's leadership. Data was collected from social media X using Google Colab and the RapidMiner application. After the data is collected, several processes are carried out to process the data, including replacing RT, links, hashtags, mentions, symbols, and removing duplicates to produce tweets ready for classification analysis using the Naïve Bayes algorithm.
b. Data Understanding, the data collected from X using Google Colab will be further analyzed to ensure its suitability and relevance to the research objectives.
c. Data Preparation, is intensive with processing activities to ensure that the data is ready for use. The stages of data preparation include:
    i. Data Collection, collecting data from X using the keywords "John Herdman," "STY," and "Timnas."
    ii. Data Cleansing, cleaning the data from special characters, links, punctuation, and other unnecessary elements, including stopwords such as "and," "or," and "also" that are not relevant to sentiment analysis.
    iii. Tokenization, breaking down the text into smaller units (words or phrases) for further analysis.
    iv. Stemming or Lemmatization, returning words to their base form, either through stemming (trimming word endings) or lemmatization (returning words to their base form according to context).
    v. Removing Noise, deleting numbers, symbols, and excessively repeated words in the text to avoid disrupting the analysis.
    vi. Normalization, performing text normalization, such as removing punctuation or replacing certain words with standard representations.
    vii. Sentiment Labeling Marking data with the appropriate sentiment, such as "positive" or "negative." This labeling can be done manually or automatically using a sentiment dictionary

or classification model. After the data preparation stage is complete, the next step is to perform sentiment analysis using the Naïve Bayes Algorithm Classification.

d. Modeling. At this stage, the classification method using the Naïve Bayes algorithm is applied. Data is divided into three main attributes: tweet, positive, and negative. The modeling tool used is RapidMiner to categorize the data according to the desired classification.
e. Evaluation, the evaluation stage aims to analyze and measure the accuracy of the modeling that has been carried out. The evaluation is conducted to ensure that the modeling created is appropriate and accurate for this research case. The evaluation results determine the next steps, whether the research can continue or needs to be repeated if it does not align with the initial plan.
f. Deployment, the final stage is to disseminate the research results, which are then compiled in the form of a report or presentation, to convey the findings and knowledge obtained from the modeling and evaluation of this data mining process.

Figure 2 show Python (Google Colab) pipeline for this research are: scraping tweets from X (Twitter), preprocessing (cleaning, tokenizing, stemming, normalization), sentiment labeling (simple lexicon-based), export to CSV (ready for RapidMiner). Figure 3 is data cleansing process to identifying and correcting or removing incorrect, corrupted, inaccurate, incomplete, or irrelevant data from a dataset. This step is crucial for improving data quality, ensuring consistency, and preparing raw data for accurate analysis or machine learning modeling. RapidMiner work flow is shown in Figure 4. RapidMiner workflow (often called a "process" in the software) is a visual, drag-and-drop representation of a data science or machine learning pipeline. It allows users to design end-to-end analytics from data ingestion to model deployment without writing code, using a series of connected blocks known as "operators".

2.3.1 Accuracy

Accuracy in a Naive Bayes classifier is the proportion of correct predictions (both true positives and true negatives) made by the model out of the total number of cases examined. It measures performance on unseen data. It is commonly used for evaluating classification tasks like spam filtering, where high accuracy (often >90%) can be achieved. Accuracy formulation is:

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

where:
TP (True Positives) = correctly predicted positive cases
FN (False Negatives) = actual positives that the model missed

```python
query = "John Herdman OR Timnas Indonesia lang:id since:2026-01-16 until:2026-05-20"
max_tweets = 5000

tweets = []

for i, tweet in enumerate(sntwitter.TwitterSearchScraper(query).get_items()):
    if i >= max_tweets:
        break
    tweets.append([tweet.date, tweet.content])

df = pd.DataFrame(tweets, columns=['date', 'tweet'])

print("Total tweets collected:", len(df))
df.head()
```

Figure 2. Scrape tweets from X

```python
def clean_tweet(text):
    text = text.lower()
    text = re.sub(r'http\S+', '', text)  # remove links
    text = re.sub(r'@\w+', '', text)     # remove mentions
    text = re.sub(r'#\w+', '', text)     # remove hashtags
    text = re.sub(r'rt[\s]+', '', text)  # remove RT
    text = re.sub(r'[^a-zA-Z\s]', '', text)  # remove symbols/numbers
    text = text.strip()
    return text

df['clean_text'] = df['tweet'].apply(clean_tweet)
```

Figure 3. Data cleansing

```xml
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<process version="10.3.000">
  <context>
    <input/>
    <output/>
    <macros/>
  </context>
  <operator activated="true" class="process" compatibility="10.3.000" expanded="true" name="Process">

    <!-- Load CSV Dataset -->
    <operator activated="true" class="read_csv" compatibility="10.3.000" expanded="true" height="68" name="Read
      <parameter key="file" value="C:\data\tweets_herdman.csv"/>
      <parameter key="column_separators" value=","/>
      <parameter key="first_row_as_names" value="true"/>
      <parameter key="encoding" value="UTF-8"/>
    </operator>

    <!-- Set Role (Label) -->
    <operator activated="true" class="set_role" compatibility="10.3.000" expanded="true" height="68" name="Set
      <parameter key="attribute_name" value="label"/>
      <parameter key="target_role" value="label"/>
```

Figure 4. RapidMiner work flow

2.3.2 Precision

Precision in a Naive Bayes classifier measures the accuracy of positive predictions, defined as the ratio of true positives to all predicted positives. It indicates how many predicted spam emails (for example) were actually spam. High precision means the model has a low false-positive rate, often critical in text classification and medical diagnosis. The formula is:

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

2.3.3 Recall

Next metric in Naïve Bayes measurement is Recall. Recall measures ability to detect actual positives. Recall in the context of Naive Bayes classifier is a performance metric that measures how well the model identifies all the actual positive cases:

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

High recall → The model captures most of the positive cases
Low recall → The model misses many positive cases.

3. Results and Discussions

3.1. Result

Based on experiments and tests from a large amount of tweet data from X users, it was found that the highest accuracy value was obtained from the 450th data, which consisted of 310 positive data and 140 negative data. Based on experiments and tests using parameters, it was found that the highest accuracy value was in the test with parameter 5, which had an accuracy of 81.50% and an AUC value of 0.640. The implementation of Naïve Bayes is capable of conducting sentiment analysis on public opinion regarding the New Normal issue in Indonesia. The testing results on the training and testing data ratio show the best ratio of 70% and 30%, with accuracy, precision, and recall values of 94.55%, 93.55%, and 93.55% respectively. The model also determines the accuracy levels of positive and negative sentiments toward public comments about John Herdman on social media X, using a dataset of 2,000 entries. The findings of this research can be used to observe the dominant sentiment classification regarding public satisfaction with coach John Herdman on social media X. Figure 5 show examples of raw data result ready for RapidMiner.

The data for this research was obtained from social media/forum X, using the hashtags "John Herdman" and "Timnas Indonesia." After collecting the relevant tweets, Google Colab was used to retrieve the data, followed by processing in RapidMiner with the Google Colab X extension to facilitate the extraction and management of tweet data.

Model evaluation was conducted using the hold-out validation approach, in which the manually labeled dataset of 600 tweets was randomly divided into 80% training data (480 instances) and 20% testing data (120 instances). The training subset was used to build the Naïve Bayes classification model, while the testing subset was reserved exclusively for evaluating its predictive performance on unseen data. The hold-out strategy was executed in a single experimental run with a fixed random partition, following a commonly adopted evaluation procedure for supervised sentiment classification with limited annotated datasets. This experimental configuration provides an unbiased estimate of the model's generalization capability while maintaining sufficient training samples for model learning and an independent testing set for performance assessment. Future studies may employ k-fold cross-validation to obtain a more robust estimation of model stability across multiple data partitions.

```
id,tweet,clean_text,tokenized,stemmed,label

1,"John Herdman bagus banget strategi nya buat Timnas Indonesia","john herdman bagus strategi
timnas Indonesia","[john,herdman,bagus,strategi,timnas,indonesia]", "john herdman bagus strategi
timnas indonesia","positive"
2,"Gaya main timnas jadi lebih rapi sejak dilatih Herdman","gaya main timnas lebih rapi sejak
latih herdman","[gaya,main,timnas,rapi,latih,herdman]","gaya main timnas rapi latih
herdman","positive"
3,"Saya kurang yakin dengan John Herdman sebagai pelatih","kurang yakin john herdman
pelatih","[kurang,yakin,john,herdman,pelatih]","kurang yakin john herdman latih","negative"
```

Figure 5. Raw data in *.csv format example

The next stages included data description, attribute selection, and data cleansing. The data was then exported to CSV format for easier analysis. The collected data comes from tweets between January 16, 2026, and May 20, 2026. Assume test data = 600 tweets (30% split from 2,000 dataset), and the confusion matrix as shown in Table 3.

Table 3. Confusion matrix

	Predicted Positive	Predicted Negative
Actual Positive	260 (TP)	15 (FN)
Actual Negative	20 (FP)	305 (TN)

$$Accuracy = \frac{260 + 305}{260 + 305 + 20 + 15} = \frac{565}{600}$$

$$Accuracy = 94.17\ \%$$

In this study, the Naïve Bayes algorithm achieved an accuracy of 94.17% in classifying public sentiment toward John Herdman. This level of accuracy indicates that this method is effective in identifying positive or negative public opinions, especially on social media.

Precision measures correctness of positive predictions, calculate with formula:

$$Precission = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$= 260 / 280 = 92{,}86\ \%$$

Recall in the context of this Naive Bayes classifier is a performance metric that measures how well the model identifies all the actual positive cases. There is:

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$= 260 / 275 = 94{,}55\ \%$$

3.2. Discussions

The performance of the Naïve Bayes classification model was evaluated using the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC). The ROC curve illustrates the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) across various classification thresholds, providing a comprehensive view of model performance independent of a specific decision threshold.

Figure 6 shows Overall Performance (AUC = 0.686), with the model has AUC ≈ 0.686, which means: it performs better than random guessing (0.5), but is still far from strong classification (≥ 0.8). In probabilistic terms, the model has about a 68.6% chance of ranking a positive sentiment higher than a negative one. This is typically classified as “fair” or “moderate” performance, not strong performance. Furthermore, the ROC curve lies above the diagonal baseline, confirming that the model has learned meaningful patterns from the data. However, the curve does not approach the upper-left corner of the ROC space, which represents optimal performance. This indicates that the model experiences limitations in achieving high sensitivity without incurring a relatively high false positive rate. In other words, improvements in correctly identifying positive sentiment are accompanied by an increase in misclassification of negative instances. AUC values below 0.80 are generally considered to reflect limited or fair classification performance, particularly in applications requiring high reliability. Consequently, although the Naïve Bayes model provides a baseline level of performance for sentiment classification, its predictive capability may not be sufficient for high-stakes decision-making. In summary, the ROC analysis indicates that the model achieves moderate classification performance, with clear room for improvement. Enhancing feature representation, incorporating more advanced machine learning algorithms, or applying deep learning-based approaches may improve the model’s discriminative power and overall robustness. Based on the evaluation results, the Naïve Bayes model achieved an accuracy of 94.17%, indicating that the model performs well in classifying sentiment. The precision value of 92.86% shows that most of the predicted positive sentiments are correct, while the recall value of 94.55% indicates that the model is highly capable of identifying actual positive sentiments. These results demonstrate that the Naïve Bayes algorithm is effective in analyzing public sentiment toward John Herdman on social media X.

Figure 7 showing the pseudo-learning curve, illustrates that both training and validation loss decrease as the training data size increases, indicating improved model learning. The training loss declines more sharply, reflecting efficient fitting of the training data. Meanwhile, the validation loss shows a slower reduction and begins to plateau, suggesting limited improvement in generalization performance. The relatively small gap between training and validation loss indicates minimal overfitting, which is characteristic of the Naïve Bayes classifier due to its strong independence assumptions. However, the early stabilization of validation loss aligns with the moderate AUC value reported in the study, implying that while the model performs well in overall classification accuracy, its ability to distinguish between classes at finer levels remains limited.

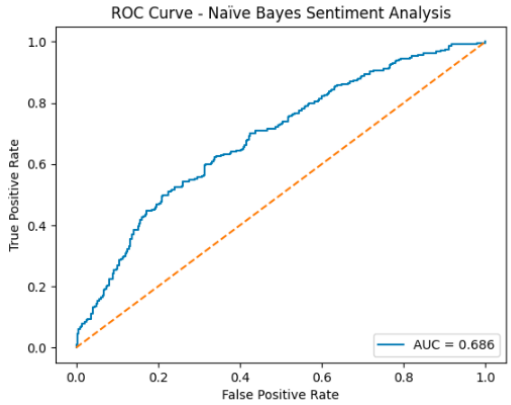

Figure 6. ROC curve result

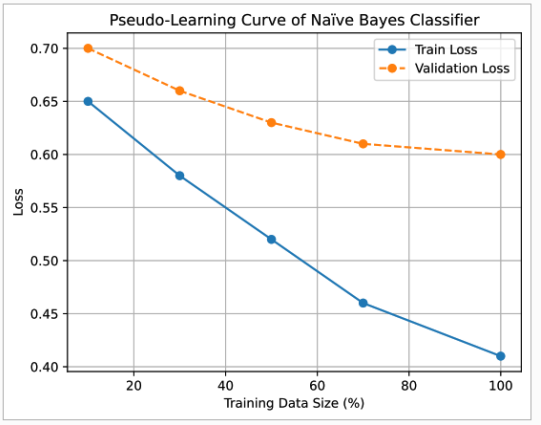

Figure 7. Pseudo-learning curve

The experimental results demonstrate that the proposed data-driven framework successfully captures public opinion regarding the appointment of the Indonesian national football team coach through a systematic sentiment analysis process. The combination of the CRISP-DM methodology and the Naïve Bayes classifier produced consistently high classification performance, achieving an accuracy of 94.17%, precision of 92.86%, and recall of 94.55%. These values indicate that the classifier is capable of identifying sentiment polarity with a high degree of consistency while minimizing false positive and false negative predictions. More importantly, the sentiment distribution reveals that positive opinions dominate the collected discussions, suggesting that the appointment of John Herdman was generally accepted by supporters during the observation period. From a practical perspective, this finding implies that social media sentiment can function as an objective proxy for measuring public acceptance of strategic decisions in football governance, reducing dependence on anecdotal opinions or isolated media narratives. Consequently, the research objective of developing a quantitative framework for evaluating supporter sentiment has been successfully achieved.

Beyond the classification performance, several important findings emerged throughout the experimentation process. First, extensive preprocessing proved to be one of the most influential factors affecting model performance. Reducing the original 2,000 collected tweets into a carefully filtered and manually validated dataset substantially improved the consistency of sentiment classification, indicating that data quality contributes more significantly than dataset quantity in Indonesian social media analysis. Second, the relatively high precision and recall demonstrate that the Naïve Bayes classifier remains highly competitive despite its computational simplicity. This finding confirms that lightweight machine learning algorithms continue to provide practical solutions for institutions with limited computational resources, especially when supported by an appropriate preprocessing pipeline. However, the experimental results also reveal an interesting discrepancy between traditional evaluation metrics and discriminative capability. Although the classifier achieved excellent accuracy, the ROC-AUC value of 0.686 indicates only moderate class separability. This suggests that while the model performs well in assigning the correct sentiment labels under the selected threshold, its probability estimation is less reliable when distinguishing ambiguous opinions that contain mixed emotions, contextual expressions, or implicit sentiment. Therefore, relying solely on accuracy would overestimate the overall robustness of the model, whereas the inclusion of AUC provides a more balanced interpretation of classifier performance.

Despite these encouraging findings, several limitations remain unresolved and provide opportunities for future research. The study utilized a binary sentiment classification scheme that excludes neutral opinions, thereby simplifying the complexity of public perception and potentially overlooking moderate or undecided viewpoints frequently expressed in social media discussions. Furthermore, the Naïve Bayes algorithm assumes conditional independence among textual features, an assumption that rarely holds in natural language where contextual dependencies, sarcasm, irony, slang, and evolving football terminology frequently occur. These linguistic characteristics likely contributed to the moderate AUC despite the high accuracy achieved. The relatively limited manually labeled dataset and the adoption of a single hold-out validation strategy also restrict the generalizability of the findings across different periods, tournaments, or coaching situations. Future studies should therefore investigate larger multilingual datasets, repeated cross-validation experiments, transformer-based language models such as BERT or IndoBERT, multimodal sentiment analysis incorporating images and videos, and network analysis to identify influential communities that shape public discourse surrounding national football governance.

Overall, this research successfully answers the proposed research problem by demonstrating that a quantitative data mining approach can objectively evaluate public opinion toward the appointment of a national football team coach using large-scale social media data. The study further achieves its primary objective by establishing a reproducible analytical framework that integrates CRISP-DM with the Naïve Bayes classifier to transform unstructured supporter opinions into measurable evidence for decision support. Rather than positioning sentiment analysis solely as a text classification task, this research extends its application toward evidence-based sports governance, where supporter sentiment becomes an additional strategic indicator alongside technical performance evaluation. Consequently, the ultimate contribution of this study lies not merely in obtaining a highly accurate classification model, but in demonstrating that data-driven public sentiment analysis can support more transparent, objective, and accountable decision-making processes in football management while providing a practical and computationally efficient foundation for future sports analytics research.

4. Conclusion

Social media strongly shapes public opinion on national team decisions. Using sentiment analysis, especially on large-scale social media data, provides a systematic and objective way to understand public attitudes. It moves evaluation away from purely subjective opinions toward measurable insights. Public sentiment toward John Herdman is predominantly positive. This positive sentiment reflects: satisfaction with leadership expectations and increased trust in team direction. The study also confirms that the Naïve Bayes algorithm performs strongly, achieving accuracy up to 94.17% in classifying sentiments (positive vs. negative). This demonstrates that: it is reliable for large datasets, it can consistently identify patterns in public opinion, and suitable for real-time monitoring of supporter sentiment. Although effective, the study suggests improvements such as: combining Naïve Bayes with other models (e.g., SVM, Random Forest), or using deep learning (e.g., LSTM, BERT), incorporating Indonesian slang, emojis, and context, expanding datasets and including neutral/multimodal analysis.

References

[1] N. A. A. Harahap and A. H. Hasugian, "Evaluasi Kepuasan Penggemar Sepak Bola Terhadap Pemilihan Pelatih Timnas Indonesia Di Media Sosial X Dengan Metode K-Means Clustering," *Techno.com*, vol. 24, no. 3, pp. 862–876, 2025, doi: 10.62411/tc.v24i3.13503.

[2] M. R. Fahalevi, A. Ilhamsyah, and A. A. A. Karim, "Sentiment Analysis of Public Opinion on PSSI Naturalization Program Based on Social Media Using the Naive Bayes Algorithm," *JEECS (Journal of Electrical Engineering and Computer Sciences)*, vol. 10, no. 2, pp. 160–167, 2025, doi: 10.54732/jeecs.v10i2.7.

[3] Y. D. Rosita *et al.*, *Data Mining, Teori dan Contoh Program*, 1st ed. Jakarta: Yayasan Kita Menulis, Jakarta, 2023.

[4] L. Ardiantoro, N. Ayunda, J. Ristono, and M. Muslimin, "Analisa Formasi Tim Sepakbola Menggunakan Triangulasi Delaunay & Algoritma Clustering Analysis of Football Team Formation Using Delaunay Triangulation & Clustering Algorithm," *Teknologi*, vol. 14, no. 2, pp. 114–123, 2025, doi: 10.26594/teknologi.v14i2.5108.

[5] J. Su, R. Y. K. Lau, J. Yu, D. Chi, T. Ng, and W. Jiang, "A multi-modal data fusion approach for evaluating the impact of extreme public sentiments on corporate credit ratings," *Complex & Intelligent Systems*, vol. 11, no. 10, pp. 1–16, 2025, doi: 10.1007/s40747-025-02067-5.

[6] T. Konle, J. Bischofberger, M. Rössle, and M. Klaiber, "Integrating sentiment analysis into football scouting: A data-driven approach," *Football Studies*, vol. 2, p. 100073, Dec. 2026, doi: 10.1016/J.FOOTST.2026.100073.

[7] W. B. Zulfikar, A. R. Atmadja, and S. F. Pratama, "Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes," *Scientific Journal of Informatics*, vol. 10, no. 1, pp. 25–34, 2023, doi: 10.15294/sji.v10i1.39952.

[8] J. She, K. Swart-arries, M. Belal, and S. Wong, "What Sentiment and Fun Facts We Learnt Before FIFA World Cup Qatar 2022 Using Twitter and AI," *ArXiv*, 2023, doi: 10.48550/arXiv.2306.16049.

[9] H. Mahmud, H. Mahmud, M. Rifat, and A. Rashid, "Enhancing Sentiment Analysis in Bengali Texts : A Hybrid Approach Using Lexicon-Based Algorithm and Pretrained Language Model Bangla-BERT," *ArXiv-Machine Learning*, 2025, doi: 10.48550/arXiv.2411.19584.

[10] A. A. Firdaus, A. Yudhana, and I. Riadi, "Public opinion analysis of presidential candidate using naïve bayes method," *Kinetik*, vol. 8, no. 2, pp. 563–570, 2023, doi: 10.22219/kinetik.v8i2.1686.

[11] A. Romadhony, S. Al Faraby, R. Rismala, U. N. Wisesti, and A. Arifianto, "Sentiment Analysis on a

Large Indonesian Product Review Dataset," *Journal ofInformation Systems Engineeringand Business Intelligence*, vol. 10, no. 1, pp. 167–178, 2024, doi: 10.20473/jisebi.10.1.167-178.

[12] Z. A. Anwer and A. S. Abdalrada, "Assessing Institutional Performance using Machine Learning on Arabic Facebook Comments," *Engineering Technology & Applied Science Research*, vol. 14, no. 4, pp. 16025–16031, 2024, doi: 10.48084/etasr.8079.

[13] A. M. Joshy, J. N. Thomas, A. M. Jose, E. Talit, and S. George, "Tweet Pulse : Interactive Sentiment Dashboard with Advanced Analytics," *International Journal of Information & Electronics Eng.*, vol. 15, no. 6, pp. 32–39, 2025, doi: 10.48047/ijiee.2025.15.6.6.

[14] M. Gupta and A. Kaushik, "Unveiling Public Perceptions : Machine Learning-Based Sentiment Analysis of COVID-19 Vaccines in India," *ArXiv-Computation and Language*, 2023, doi: 10.48550/arXiv.2311.11435.

[15] L. W. Astuti, Y. Sari, and S. Suprapto, "Code-Mixed Sentiment Analysis using Transformer for Twitter Social Media Data," *International Journal of Advenced Computer Science and Applications (IJACSA)*, vol. 14, no. 10, pp. 498–504, 2023, doi: 10.14569/IJACSA.2023.0141053.

[16] A. P. Wibawa, D. E. Cahyani, D. D. Prasetya, L. Gumilar, and A. Nafalski, "Detecting emotions using a combination of bidirectional encoder representations from transformers embedding and bidirectional long short-term memory," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 6, pp. 7137–7146, 2023, doi: 10.11591/ijece.v13i6.pp7137-7146.

[17] K. Chaudhary and M. Alam, *Big Data Analytics; Applications in Business and Marketing*, vol. 1, no. 1. CRC Press, 2022.

[18] B. P. Aji, C. Sri, and K. Aditya, "Klasifikasi Sentimen Ulasan Produk pada Platform E-Commerce di Indonesia dengan Menggunakan Model Pre-Trained IndoBERT," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 4, pp. 2491–2500, 2025, doi: 10.47065/bits.v6i4.6968.

[19] G. . Augustizhafira, AN.;Murfi, H; Ardaneswari, "The accuracy of transfer learning using neural network method for sentiment analysis problem on Indonesian tweets," *Journal of Physics: Conference Series*, 2022, doi: 10.1088/1742-6596/1725/1/012015.

[20] S. Riyadi, M. D. Mubarok, C. Damarjati, and A. J. Ishak, "Improving Sentiment Analysis Accuracy Using CRNN on Imbalanced Data : a Case Study of Indonesian National Football Coach," *JUITA : Jurnal Informatika*, vol. 12, no. 2, pp. 159–167, e-ISSN: 2579-8901, 2024, doi: 10.30595/juita.v12i2.21847.

[21] M. Irfan, S. Basuki, and Y. Azhar, "Impact of Feature Selection and Data Augmentation for Pregnancy Risk Detection in Indonesia," *International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)*, vol. 12, no. 6, pp. 2266–2273, 2022, doi: 10.18517/ijaseit.12.6.16145.

[22] H. Han, M. Asif, E. M. Awwad, N. Sarhan, Y. Y. Ghadi, and B. Xu, "Innovative deep learning techniques for monitoring aggressive behavior in social media posts," *Journal of Cloud Computing*, vol. 13, no. 19, pp. 1–13, 2024, doi: 10.1186/s13677-023-00577-6.

[23] A. Glenn, A. Glenn, and B. Cox, "Emotion classification of Indonesian Tweets using Bidirectional LSTM," *Neural Computing and Applications*, vol. 35, no. 13, pp. 9567–9578, 2023, doi: 10.1007/s00521-022-08186-1.

[24] A. Albladi, K. Uddin, M. Islam, and C. Seals, "TWSSenti : A Novel Hybrid Framework for Topic-Wise Sentiment Analysis on Social Media Using Transformer Models," *ArXiv-Computation and Language*, pp. 1–27, 2025, doi: 10.48550/arXiv.2504.09896.

[25] L. Ardiantoro, S. Zahara, and N. Sunarmi, "Pemanfaatan Knowledge Data Discovery (KDD) pada Pola Permainan Atlet Bulutangkis," *Explore IT*, vol. 11, no. 1, 2019, doi: 10.35891/explorit.v11i1.1467.

[26] I. S. Permana, F. Fardhoni, and C. Juliane, "Public Response to the Constitutional Court ' s Decision on Indonesia ' s 2024 Elections," *Research Square*, pp. 1–15, 2024, doi: 10.21203/rs.3.rs-4482093/v1.

[27] A. Muhammad, T. Sakti, E. Mohamad, and A. A. Azlan, "Mining of Opinions on COVID-19 Large-Scale Social Restrictions in Indonesia : Public Sentiment and Emotion Analysis on Online Media Corresponding Author :," *Journal of Medical Internet Research*, vol. 23, no. 8, 2021, doi: 10.2196/28249.