

# Analysis of Public Sentiment on Twitter Social Media the Design of the Latest Jersey of the Indonesian Football Team using the Support Vector Machine (SVM) Method

Gusti Alfianda Akbar<sup>1\*</sup>, Relita Buaton<sup>2</sup>, Magdalena Simanjuntak<sup>3</sup>

<sup>1,2,3</sup> Informatics Engineering, STMIK KAPUTAMA

Jl. Veterans No. 4A-9A, Binjai, North Sumatra, Indonesia

Corresponding: [gustialfianda31@gmail.com](mailto:gustialfianda31@gmail.com)<sup>1\*</sup>, [bbcbuaton@gmail.com](mailto:bbcbuaton@gmail.com)<sup>2</sup>, [magdalena.simanjuntak84@gmail.com](mailto:magdalena.simanjuntak84@gmail.com)<sup>3</sup>

## Abstract

Twitter has become a major platform for real-time public expression, including reactions to the Indonesian national football team's new jersey released by Erspo on January 23, 2025. The previous edition had received strong criticism, creating the need to examine how the public responded to the new design. This study aims to analyze the distribution of sentiments on Twitter and evaluate the performance of the chosen classification method. The research employs Support Vector Machine (SVM) with a linear kernel to classify Indonesian-language tweets into positive and negative categories. Data were collected through crawling and processed using text preprocessing techniques such as case folding, tokenizing, filtering, and stemming, with features extracted using Term Frequency–Inverse Document Frequency (TF-IDF). The model's performance was assessed based on accuracy, precision, and recall. Results show that public sentiment comprised 308 positive and 437 negative tweets. The SVM model achieved an accuracy of 82.35%, with 76% precision for positive and 86% precision for negative classifications. These results indicate that public responses tended to be negative, though positive appreciation was still evident. Overall, SVM proved effective for sentiment analysis and can provide valuable insights for decision-makers and jersey developers.

**Keywords:** Jersey, National Team, Sentiment Analysis, Support Vector Machine, Twitter

## 1. Introduction

Social media has become the main means for the public to express opinions quickly and openly. Twitter, with its short-form messaging characteristics and real-time dissemination of information, is a relevant platform for monitoring public opinion on a wide range of issues, including sports [1]. One of the issues that is widely discussed is the release of the latest design of the Indonesian national team jersey on January 23, 2025. Previous editions have been criticized for being considered less attractive and not representing national identity well [2]. This condition raises questions about how the public response to the latest edition design has again sparked debate on social media.

Sentiment analysis is the right approach to understand large amounts of public opinion. By utilizing Natural Language Processing (NLP), opinions can be grouped into positive or negative categories [3]. Previous studies have shown that the Support Vector Machine (SVM) algorithm is capable of providing high accuracy in text classification and is often used in sports-related sentiment analysis [4]. However, studies that specifically highlight public opinion on the design of the Indonesian national team's jerseys are still limited. This opens up research opportunities to contribute to understanding people's preferences.

This study aims to analyze the distribution of public sentiment on Twitter regarding the latest design of the Indonesian national team jersey and evaluate the effectiveness of the SVM method with a linear kernel. By applying the text preprocessing stages (case folding, tokenizing, filtering, and stemming) as well as TF-IDF weighting, the resulting classification model achieved an accuracy of 82.35%. These findings suggest that the majority of sentiment tends to be negative, although positive appreciation nonetheless emerges. The results of the research are expected to be input for designers and decision-makers in designing communication strategies and developing national team jersey designs in the future.

## 2. Theoretical Foundation

### 2.1. Sentimen Analysis

Sentiment analysis or *opinion mining* is a method of analyzing text data to identify and classify the opinions contained in a document. The main goal of sentiment analysis is to determine the polarity of the text, whether positive, negative, or neutral, so as to capture the public's

perception of an issue (Amin & Salma, 2024)[1]. On social media such as Twitter, sentiment analysis is often used to measure people's responses to certain topics, events, or products.

## 2.2. Natural Language Processing (NLP)

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on the processing of human language by computers. In sentiment analysis, NLP is used to prepare text before it is further analyzed. The preprocessing stage includes case folding, tokenizing, filtering, and stemming, which aims to clean and simplify the text to preserve key information[5].

## 2.3. Text Representation with TF-IDF

One popular method for converting text into numerical forms is Term Frequency–Inverse Document Frequency (TF-IDF). This method assigns weight to each word based on the frequency with which it appears in a document compared to the entire document collection. Words that appear frequently in one document but rarely in another will acquire higher weight[6].

The TF-IDF formula is:

$$TFIDF(t,d) = TF(t,d) \times IDF(t) \quad (1)$$

With  $IDF(t) = \log(\frac{N}{df(t)})$ , where  $N$  is the total number of documents and  $df(t)$  is the number of documents containing the term.

## 2.4. Support vector machine(SVM)

Support Vector Machine (SVM) is a supervised learning algorithm used for classification, including sentiment analysis. SVM works by building an optimal *hyperplane* that separates data into two classes with maximum margins [7]. The advantage of SVM is its ability to handle the high-dimensional, *sparse data* that is common in text [4].

The SVM decision function is expressed as:

$$f(x) = w \cdot x + b \quad (2)$$

where  $w$  is the weight vector,  $x$  is the feature, and  $b$  is the bias [8].

# 3. Analysis and Design

## 3.1. Research Method

This study uses a quantitative approach with *the text mining method* to analyze public sentiment on social media Twitter. The goal is to classify Indonesian-language tweets related to the release of the Indonesian national team's jersey into positive and negative categories. The research stages include data collection, preprocessing, feature extraction, classification, and evaluation.

## 3.2. System Planning

The system design follows a linear data processing workflow, as illustrated in Figure. 1. This flow starts from the raw text data input, which then passes through a series of pre-processing modules. The cleaned text is then converted into a numeric vector through TF-IDF weighting. This vector is then the input for the SVM model that has been trained to generate sentiment class predictions.

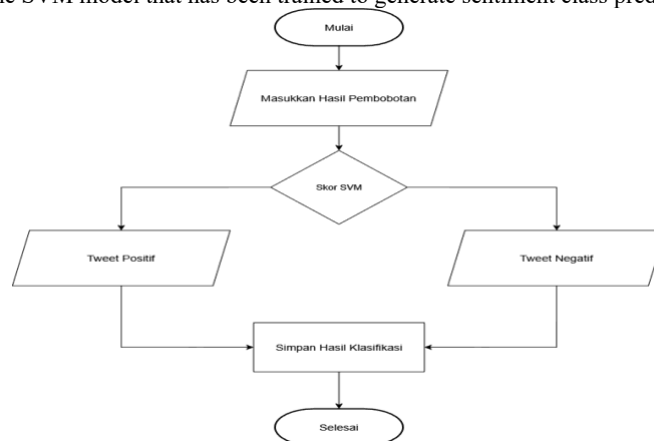


Fig. 1: SVM Classification System Flowchart

## 3.3. Preprocessing Data

The preprocessing stage is carried out to make the text cleaner and uniform before entering the analysis process. These stages include:

1. Case Folding: converts the entire text to lowercase.
2. Tokenizing: breaking down text into words (tokens).
3. Filtering: removes *unnecessary stopwords* and characters.
4. Stemming: returns the word to its basic form.

### 3.4. Support Vector Machine Processing

To provide a more concrete understanding of the workflow of the applied method, manual calculations use a small sample of data. This illustration covers the entire stage from pre-processing to normalization of feature vectors.

#### 3.4.1. Training Data and Test Data

In actual implementation, the dataset will be divided into training data (e.g., 80% of the total data) and test data (20%). However, for illustrative purposes, the following 4 data samples will be used.

1. D1 (Positive): This time the design of the national team is coolnnnn
2. d2 (Negative): It's too much to say that this is the time of the teamwear jersey last year to make the national team jersey what is out of stock?
3. d3 (Positive): the new jersey of the national team is much better than the previous tp why are there small things disturbing
4. d4 (Positive): hopefully with his new jersey other positive things can be achieved by our national team

#### 3.4.2. Preprocessing Text

The following shows how each pre-processing stage converts raw text into a clean list of terms. These tables serve to transparently demonstrate how the system handles linguistic variations such as slang, abbreviations, and suffixes, which is a crucial step for standardization of features.

Table 1: Preprocessing Text

No	Teks Asli	Hasil Case Folding	Hasil Tokenizing	Hasil Filtering	Hasil Stemming (Kata Dasar)
D1	Kali ni desain jersey timnas kerennnn	kali ni desain jersey timnas kerennnn	[kali, ni, desain, jersey, timnas, kerennnn]	desain, jersey, timnas, kerennnn	[desain, jersey, timnas, keren]
D2	Keterlaluhan sih ini masa jersey teamwear thn yg lalu buat jersey timnas apa habis stok	keterlaluhan sih ini masa jersey teamwear thn yg lalu buat jersey timnas apa habis stok	[keterlaluhan, sih, ini, masa, jersey, teamwear, thn, yg, lalu, buat, jersey, timnas, apa, habis, stok]	[keterlaluhan, jersey, teamwear, thn, lalu, jersey, timnas, habis, stok]	[lalu, jersey, teamwear, thn, lalu, jersey, timnas, habis, stok]
D3	Jersey baru timnas jauh lebih bagus dr sebelumnya tp kenapa ada hal kecil ganggu	jersey baru timnas jauh lebih bagus dr sebelumnya tp kenapa ada hal kecil ganggu	[jersey, baru, timnas, jauh, lebih, bagus, dr, sebelumnya, tp, kenapa, ada, hal, kecil, ganggu]	[jersey, baru, timnas, jauh, bagus, kecil, ganggu]	[jersey, baru, timnas, jauh, bagus, kecil, ganggu]
D4	Semoga dengan jersey baru nya halhal positif lain nya dapat diraih oleh timnas kita	semoga dengan jersey baru nya halhal positif lain nya dapat diraih oleh timnas kita	[semoga, dengan, jersey, baru, nya, halhal, positif, lain, nya, dapat, diraih, oleh, timnas, kita]	[jersey, baru, halhal, positif, raih, timnas]	[jersey, baru, hal, positif, raih, timnas]

#### 3.4.3. Weighting (TF-IDF)

From the results of *the stemming*, vocabulary for 4 sample documents was obtained. Next, the TF-IDF weights for each term are calculated. Table 2 shows this calculation, where N=4. The IDF value is calculated using the formula  $idf = \log(N/df)$

Table 2: TF-IDF Calculation

Term	TF (d1)	TF (d2)	TF (d3)	TF (d4)	df	idf ( $\log(4/df)$ )	TF-IDF (d1)	TF-IDF (d2)	TF-IDF (d3)	TF-IDF (d4)
desain	1	0	0	0	1	0.602	0.602	0	0	0
jersey	1	2	1	1	4	0	0	0	0	0
timnas	1	1	1	1	4	0	0	0	0	0
keren	1	0	0	0	1	0.602	0.602	0	0	0
Term	TF (d1)	TF (d2)	TF (d3)	TF (d4)	df	idf ( $\log(4/df)$ )	TF-IDF (d1)	TF-IDF (d2)	TF-IDF (d3)	TF-IDF (d4)
lalu	0	2	0	0	1	0.602	0	1.204	0	0
teamwear	0	1	0	0	1	0.602	0	0.602	0	0
habis	0	1	0	0	1	0.602	0	0.602	0	0

stok	0	1	0	0	1	0.602	0	0.602	0	0
baru	0	0	1	1	2	0.301	0	0	0.301	0.301
jauh	0	0	1	0	1	0.602	0	0	0.602	0
bagus	0	0	1	0	1	0.602	0	0	0.602	0
kecil	0	0	1	0	1	0.602	0	0	0.602	0
ganggu	0	0	1	0	1	0.602	0	0	0.602	0
hal	0	0	0	1	1	0.602	0	0	0	0.602
positif	0	0	0	1	1	0.602	0	0	0	0.602
raih	0	0	0	1	1	0.602	0	0	0	0.602

### 3.4.4. Normalization

After obtaining the TF-IDF vector, the normalization process is carried out. An example of the normalization weight of  $t_i$  in the illustration of the weighting of the term "design" in the first document (d1) is as follows.

$$\frac{0.602}{\sqrt{0.602^2 + 0.602^2}} = \frac{0.602}{\sqrt{0.7248}}$$

$$\frac{0.602}{0.8513} = 0.707$$

All weighting results were normalized. Table. 3 explains the normalization obtained from the weighting calculation.

**Table 3: TF-IDF Vector Normalization Results**

Term	Norm (d1)	Norm (d2)	Norm (d3)	Norm (d4)
desain	0.707	0	0	0
keren	0.707	0	0	0
lalu	0	0.756	0	0
teamwear	0	0.378	0	0
habis	0	0.378	0	0
Term	Norm (d1)	Norm (d2)	Norm (d3)	Norm (d4)
stok	0	0.378	0	0
baru	0	0	0.243	0.277
jauh	0	0	0.485	0
bagus	0	0	0.485	0
kecil	0	0	0.485	0
ganggu	0	0	0.485	0
hal	0	0	0	0.555
positif	0	0	0	0.555
raih	0	0	0	0.555

These normalized vectors are the final input to train and test the *Support Vector Machine classification model*.

### 3.4.5 Calculation of Support Vector Machine (SVM)

**Table 4: Normalized TF-IDF Input Vectors and Assigned Labels**

Dokumen	Label (y)	desain	keren	lalu	teamwear	habis	stok	baru	jauh	bagus	kecil
d1 (x1)	1	0.707	0.707	0	0	0	0	0	0	0	0
d2 (x2)	-1	0	0	0.756	0.378	0.378	0.378	0	0	0	0
d4 (x4)	1	0	0	0	0	0	0	0.277	0	0	0

Document d3 is not included in this table because it will be used as test data, the determination of the sentiment label is as follows:

1. d1: "This time the national team jersey design is cool" → Positive (+1)
2. d2: "It's too much to say that this is the time of the teamwear jersey last year to make the national team jersey what is out of stock" → Negative (-1)
3. d4: "Hopefully with the new jersey other positive things can be achieved by our national team" → Positive (+1)

Document d3 ("The new jersey of the national team is much better than the previous one tp why is there a small thing disturbed") contains mixed sentiment signals ("good" vs "disturbing"). This document will be used as test data for model validation. Thus, the problem of *SVM Hard-Margin* optimization can be formally stated as follows

Minimize:

$$\frac{1}{2} \|w\|^2$$

With Limitations:

$$y_i(w \cdot x_i + b) \geq 1, \quad \text{untuk } i = 1, 2, 4$$

Where  $x_i$  is the normalized TF-IDF feature vector for document  $i$ , and  $y_i$  is its class label (+1 or -1), as presented in Table . The Lagrangian function (L) combines the function of purpose and constraint into a single equation:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_i \alpha_i [y_i(w \cdot x_i + b) - 1]$$

To find a solution, we take the partial derivative of L against the primal variables (w and b) and equalize it by zero:

$$\begin{aligned} \frac{\partial L}{\partial w} &= w - \sum_i \alpha_i y_i x_i = 0 \Rightarrow w = \sum_i \alpha_i y_i x_i \\ \frac{\partial L}{\partial b} &= - \sum_i \alpha_i y_i = 0 \Rightarrow \sum_i \alpha_i y_i = 0 \end{aligned}$$

Maximize:

$$L_D(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

With limitations:

$$\sum_i \alpha_i y_i = 0 \quad \text{dan} \quad \alpha_i \geq 0$$

Then, Identify Support Vectors through KKT Conditions. The Karush-Kuhn-Tucker (KKT) condition is a set of conditions necessary for an optimal solution to a non-linear optimization problem with constraints. One of the most important KKT conditions for SVM is *complementary slackness*, which is expressed as:

$$\alpha_i [y_i(w \cdot x_i + b) - 1] = 0$$

This condition states that for each data point  $x_i$ :

1. If  $y_i(w \cdot x_i + b) > 1$ , then the point is outside the margin and is correctly classified. For the above equation to be met, its Lagrange multiplier must be  $\alpha_i = 0$ . This point is not a *support vector*.
1. If  $y_i(w \cdot x_i + b) = 1$ , then that point is located just above the margin. The Lagrange multiplier can be  $\alpha_i > 0$ . This point is the *support vector*.

Solving Equations for Lagrange Multipliers ( $\alpha$ )

Calculate all the required dot products ( $x_i \cdot x_j$ ) from the data

$$\begin{aligned} x_1 \cdot x_1 &= (0.707)^2 + (0.707)^2 \\ &= 0.4998 + 0.4998 = 0.9996 \approx 1.0 \\ x_2 \cdot x_2 &= (0.756)^2 + (0.378)^2 + (0.378)^2 + (0.378)^2 \\ &= 0.5715 + 0.1428 + 0.1428 + 0.1428 \approx 1.0 \\ x_4 \cdot x_4 &= (0.277)^2 + (0.555)^2 + (0.555)^2 + (0.555)^2 \\ &= 0.0767 + 0.308 + 0.308 + 0.308 \approx 1.0 \\ x_1 \cdot x_2 &= 0 \quad (\text{there is no such term}) \\ x_1 \cdot x_4 &= 0 \quad (\text{there is no such term}) \\ x_2 \cdot x_4 &= 0 \quad (\text{there is no such term}) \end{aligned}$$

Substitution into the KKT equation for each  $w = \sum \alpha_j y_j x_j$  support vector:

1. For  $x_1$  (with  $y_1 = +1$ ):

$$\begin{aligned} \alpha_1(x_1 \cdot x_1) - \alpha_2(x_1 \cdot x_2) + \alpha_4(x_1 \cdot x_4) + b &= 1 \\ \alpha_1(1.0) + b &= 1 \Rightarrow \alpha_1 + b = 1 \end{aligned}$$

2. For  $x_2$  (with  $y_2=-1$ ):

$$\begin{aligned} -[\alpha_1(x_2 \cdot x_1) - \alpha_2(x_2 \cdot x_2) + \alpha_4(x_2 \cdot 4) + b] &= 1 \\ -[\alpha_2(1.0) + b] &= 1 \Rightarrow \alpha_2 - b = 1 \end{aligned}$$

3. For  $x_4$  (with  $y_4=+1$ ):

$$\begin{aligned} \alpha_1(x_4 \cdot x_1) - \alpha_2(x_4 \cdot x_2) + \alpha_4(x_4 \cdot x_4) + b &= 1 \\ \alpha_4(1.0) + b &= 1 \Rightarrow \alpha_4 + b = 1 \end{aligned}$$

4. Tops of dual problems:

$$\alpha_1 y_1 + \alpha_2 y_2 + \alpha_4 y_4 = 0$$

$$\alpha_1(1) + \alpha_2(-1) + \alpha_4(1) = 0 \Rightarrow \alpha_1 - \alpha_2 + \alpha_4 = 0$$

4 Linear equations with 4 variables ( $\alpha_1, \alpha_2, \alpha_4, b$ ).

from (1) and (3),  $\alpha_1 = \alpha_4$

$$\text{Substitution to (4): } \alpha_1 - \alpha_2 + \alpha_1 = 0 \Rightarrow 2\alpha_1 = \alpha_2$$

b :

$$\text{From (1): } b = 1 - \alpha_1$$

$$\text{From (2): } b = \alpha_2 - 1$$

Then,  $1 - \alpha_1 = \alpha_2 - 1$ . Replace  $\alpha_2$  with  $2\alpha_1$ :

$$1 - \alpha_1 = (2\alpha_1) - 1 \Rightarrow 2 = 3\alpha_1 \Rightarrow \alpha_1 = \frac{2}{3} \approx 0.667$$

Other values:

$$1. \quad \alpha_4 = \alpha_1 = 0.667$$

$$2. \quad \alpha_2 = 2\alpha_1 = 2 \times \left(\frac{2}{3}\right) = \frac{4}{3} \approx 1.333$$

$$3. \quad b = 1 - \alpha_1 = 1 - \frac{2}{3} = \frac{1}{3} \approx 0.333$$

Semua nilai  $\alpha$  positif, semua titik adalah *support vectors*.

Setelah nilai-nilai  $\alpha$  ditemukan, dapat menghitung parameter model akhir.

Menghitung Vektor Bobot  $w$ : Vektor bobot  $w$  dihitung menggunakan formula

$$w = \sum \alpha_i y_i x_i = \alpha_1(1)x_1 + \alpha_2(-1)x_2 + \alpha_4(1)x_4.$$

$$w = 0.667x_1 - 1.333x_2 + 0.667x_4$$

**Table 5:** Final SVM Parameter Calculation (w and b)

Term (Fitur)	Komponen Vektor w	Perhitungan
desain	w1	$0.667 \times 0.707 = 0.471$
keren	w2	$0.667 \times 0.707 = 0.471$
lalu	w3	$(-1.333) \times 0.756 = -1.008$
Term (Fitur)	Komponen Vektor w	Perhitungan
teamwear	w4	$(-1.333) \times 0.378 = -0.504$
habis	w5	$(-1.333) \times 0.378 = -0.504$
stok	w6]	$(-1.333) \times 0.378 = -0.504$
baru	w7	$0.667 \times 0.277 = 0.185$
jauh	w8	0
bagus	w9	0
kecil	w10	0
ganggu	w11	0
hal	w12	$0.667 \times 0.555 = 0.370$
positif	w13	$0.667 \times 0.555 = 0.370$

raih	w14	0.667×0.555=0.370
Bias	b	0.333

The SVM decision function that has been trained can now be written explicitly. For each new input vector  $x_{test}$ , its classification is determined by a sign (positive or negative) of the output of the following function:

$$f(x_{test}) = \text{sign}(w \cdot x_{test} + b)$$

If  $w \cdot x_{test} + b > 0$ , then  $x_{test}$  is classified as a Positive (+1 class). If  $w \cdot x_{test} + b < 0$ , then it is classified as a class Negative (-1). The value of  $w \cdot x_{test} + b$  itself is often referred to as a *decision score* or *confidence score*. Testing the model on the d3 document: "The new jersey of the national team is much better than the previous but why are there small things disturbing". Normalized TF-IDF vector for d3 ( $x_3$ )

Vector  $x_3$  has the following components:

1. New: 0.243
2. Away: 0.485
3. Good: 0.485
4. Small: 0.485
5. Disturb: 0.485
6. All other terms: 0

Next is to calculate *the dot product*  $w \cdot x_3$ :

$$w \cdot x_3 = \sum_{k=1}^{17} w_k \cdot (x_3)_k$$

Adding multiplication for non-zero components on  $x_3$ :

$$w \cdot x_3 = (w_7 \times 0.243) + (w_8 \times 0.485) + (w_9 \times 0.485) + (w_{10} \times 0.485) + (w_{11} \times 0.485)$$

Using the  $w$  value from Table 2:

$$w \cdot x_3 = (0.185 \times 0.243) + (0 \times 0.485) + (0 \times 0.485) + (0 \times 0.485) + (0 \times 0.485)$$

$$w \cdot x_3 = 0.045 + 0 + 0 + 0 + 0 = 0.045$$

$$\text{Decision score} = w \cdot x_3 + b = 0.045 + 0.333 = 0.378$$

$$\text{Classification} = \text{sign}(0.378) = +1$$

Table 7: d3 Document Classification Test

Langkah	Deskripsi	Nilai
1	Vektor Input $x_3$	[0, 0, 0, 0, 0, 0, 0.243, 0.485, 0.485, 0.485, 0, 0, 0, 0]
2	Hitung Dot Product $w \cdot x_3$	0.045
3	Tambahkan Bias $b$	0.045+0.333=0.378
4	Terapkan Fungsi $\text{sign}()$	$\text{sign}(0.378)$
5	Hasil Klasifikasi Akhir	Positif (+1)

## 4. Implementation

### 4.1. Testing

In the implementation, it will be explained about the results of the data test using *the Support Vector Machine (SVM)* method using python and google colab to get the results of the accuracy of public sentiment data on social media twitter against the latest jersey design of the Indonesian national football team.

4.1.1. Program Analysis

Dalam analisa program ini hal yang dilakukan untuk mengetahui berapa jumlah data yang akan di analisis dari twitter serta menentukan sentimen pada setiap data.

1. Determine the sentiment on each text data

```
polarity
negative    437
positive    308
Name: count, dtype: int64
```

tweet_tokens	tweet_tokens_fdist	tweet_tokens_MSH	tweet_normalized	tweet_baku	tweet_tokens_stemmed	text	polarity_score	polarity
['info', 'bell', 'jersey', 'timnas']	<FreqDist with 4 samples and 4 outcomes>	['info', 'bell', 'jersey', 'timnas']	['info', 'bell', 'jersey', 'timnas']	['info', 'bell', 'jersey', 'timnas']	['info', 'bell', 'jersey', 'timnas']	info bell jersey timnas	-4	negative
['saya', 'ingin', 'membeli', 'jersey', 'timnas...']	<FreqDist with 7 samples and 7 outcomes>	['membeli', 'jersey', 'timnas', 'baru']	['membeli', 'jersey', 'timnas', 'baru']	['membeli', 'jersey', 'timnas', 'baru']	['bell', 'jersey', 'timnas', 'baru']	bell jersey timnas baru	-1	negative
['ada', 'netjen', 'negeri', 'sebelah', 'yg', ...]	<FreqDist with 25 samples and 30 outcomes>	['netjen', 'negeri', 'sebelah', 'membandingkan...']	['netjen', 'negeri', 'sebelah', 'membandingkan...']	['warganet', 'negeri', 'sebelah', 'membandingkan, jer...']	['warganet', 'negeri', 'belah', 'banding', 'jersey', 'timnas je...']	warganet negeri belah banding jersey timnas je...	-15	negative
['seberapa', 'bagus', 'rating', 'jersey', 'ter...']	<FreqDist with 8 samples and 8 outcomes>	['bagus', 'rating', 'jersey', 'terbaru', 'timn...']	['bagus', 'rating', 'jersey', 'terbaru', 'timn...']	['bagus', 'rating', 'jersey', 'terbaru', 'timnas', 'erspo']	['bagus', 'rating', 'jersey', 'baru', 'timnas']	bagus rating jersey baru timnas erspo	2	positive
['keterlalaian', 'sih', 'ini', 'masa', 'jersey...']	<FreqDist with 14 samples and 15 outcomes>	['jersey', 'teamwear', 'thin', 'jersey', 'timna...']	['jersey', 'teamwear', 'thin', 'jersey', 'timna...']	['jersey', 'teamwear', 'tahun, jersey, timnas, habi...']	['jersey', 'teamwear', 'tahun', 'jersey', 'tim...']	jersey teamwear tahun jersey timnas habis stok	-3	negative
...	...	...	...	...	...	...	...	...
['itu', 'yg', 'terakhir', 'ngapain', 'ani', 'an...']	<FreqDist with 9 samples and 9 outcomes>	['ngapain', 'anjing', 'jersey', 'timnas', 'malay']	['ngapain', 'anjing', 'jersey', 'timnas', 'malay']	['apa', 'anjing', 'jersey', 'timnas', 'malay']	['apa', 'anjing', 'jersey', 'timnas', 'malay']	apa anjing jersey timnas malay	-7	negative
['akhirnya', 'bisa', 'lihat', 'jersey', 'timnas...']	<FreqDist with 11 samples and 15 outcomes>	['lihat', 'jersey', 'timnas', 'keren', 'lihat', ...]	['lihat', 'jersey', 'timnas', 'keren', 'lihat', ...]	['lihat', 'jersey', 'timnas', 'keren, lihat, jersey, ...']	['lihat', 'jersey', 'timnas', 'keren', 'lihat', ...]	lihat jersey timnas keren lihat jersey timnas ...	9	positive
['jersey', 'bagusnya', 'moda...']	<FreqDist with 10 samples and 10 outcomes>	['jersey', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya']	['jersey', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya']	['jersey', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya', 'bagusnya']	['jersey', 'bagus', 'setaras', 'sama', 'setaras', 'sama', 'setaras', 'sama', 'setaras', 'sama']	jersey bagus setaras sama sama	6	positive

Fig. 2: Determine the sentiment on each text data

2. Visualization of sentiment with pie chart  
Sentiment visualization with pie chart is an image that shows sentiment with a picture or diagram so that it is easy to understand

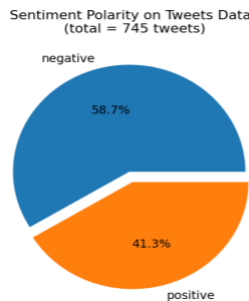


Fig. 3: Visualization of Sentiment with Pie Chart

3. Visualization of sentiment with bar charts

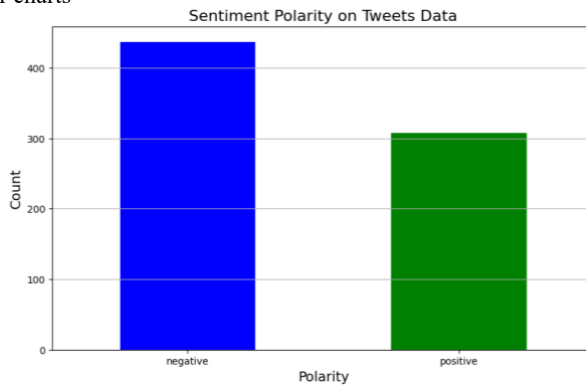


Fig. 4: Sentiment Visualization with Bar Chart

4. Displaying Data with the Highest Positive Sentiment

	tweet_baku	polarity_score	polarity
1	[paham, bocah, cilik, sumber, daya, manusia, fan, timnas, terendah, sedunia, jersey, tim, menyingkan, signature,ri, khas, jati, bangsa, ditonjokkan, puas, lihat, tampilan, jersey, punya, nilai, makna]	23	positive
2	[erspo, resmi, merilis, jersey, terbaru, kandang, timnas, indonesia, mengusung, tema, indonesia, pusaka, semarak, kebanggaan, indonesia, jersey, menceritakan, wujud, semangat, kekuatan, bangsa, indonesia, diwariskan, generasi, generasi, elemen, tcojndonyq]	21	positive
3	[erspo, resmi, meluncurkan, jersey, terbaru, timnas, indonesia, laga, kualifikasi, piala, dunia, maret, tema, indonesia, pusaka, desain, jersey, menggabungkan, ornamen, bunga, edelweis, simbol, kekayaan, alam, lambang, Garuda]	20	positive
4	[erspo, merilis, jersey, timnas, indonesia, pusaka, desainnya, mencerminkan, kekayaan, budaya, alam, respon, positif, wagernet, tcoelynrvz]	20	positive
5	[produksi, apparel, indonesia, erspo, resmi, merilis, jersey, terbaru, timnas, indonesia, langkah, menandai, pencapaian, erspo, penyedia, apparel, resmi, timnas, sambutan, mengantisipasi, mills, jersey, ariyar, tcojnqyrtbo]	19	positive
6	[jersey, terbaru, timnas, indonesia, erspo, penyedia, jersey, apparel, pasi, meluncurkan, jersey, kandang, timnas, indonesia, jersey, dijual, februari, nilai, jersey, timnas, indonesia, foto, dok, erspo, tcoawstfno]	19	positive
7	[erspo, resmi, merilis, jersey, timnas, indonesia, bertema, indonesia, pusaka, desainnya, memukau, mencerminkan, kekayaan, budaya, semangat, nasionalisme, jersey, debut, laga, kualifikasi, piala, dunia, maret, selengkapnya, tcozefhyot, tcozspnyqzw]	19	positive
8	[desain, jersey, timnas, cakap, secara, bisnis, oke, ditrik, negaralub, alhasil, merek, lokal, lembus, pasar, internasional, setaras, prestasi, timnas, ngebangun, dukungan, fan]	17	positive
9	[resmi, dirilis, sekian, menunggu, jersey, terbaru, timnas, indonesia, dirilis, erspo, resmi, merilis, sore, Kamis, Januari, jersey, terbaru, timnas, mengusung, tema, indonesia, pusaka, agendakan, lanjutan, laga, tcooetvayst]	16	positive
10	[erspo, rilis, jersey, timnas, keren, gilaak, menambah, keren, pas, lihat, foto, shayne]	15	positive

Fig. 5: Data with the Highest Positive Sentiment

5. Showing the 10 data with the Highest Negative Sentiment

	tweet_baku	polarity_score	polarity
1	[jersey, timnas, indonesia, bagus, long, sleeve, cuma, celananya, tidak, usah, merahnya, sampingnya, bagus, putih, polos, menyatu, warna, putih, bajunya, tcovedux]	-31	negative
2	[malaysia, jelek, parah, desain, jersey, timnas, erspo, maaf, malaysia, kritik, keras, boikot, penjualannya, federasi, sadar, fan, malaysia, sekadar, angka, diperas, uang, tiket, san, merch, suaranya, dianggap]	-16	negative
3	[meminimalisir, drama, patut, apresiasi, ksruh, jersey, cape, kemarn, drama, pelatih, timnas]	-15	negative
4	[warganet, negeri, sebelah, membandingkan, jersey, timnas, jersey, timnas, komen, lucu, marah, lucu, terus, marah, sampai, jersey, sesua]	-15	negative
5	[bagus, jelek, design, masih, relatif, jersey, edisi, kritik, tetap, bagus, terpenting, buka, bagusjelek, desainya, pencapaian, timnas, pakai, jersey, jersey, erspo, edisi, prestasi, melek]	-15	negative
6	[bahasnya, out, topic, malu, bodoh, dibanggakan, haha, olimpic, pemah, emas, pelatih, bullang, ambil, dari, negeri, jiran, jersey, keset, timnas, main, kat, marapun, main, cuma, kat, piala, merdeka, j]	-14	negative
7	[cakap, garis, putih, kelihatan, agak, retro, long, sleeve, kayaknya, beli, jersey, timnas]	-13	negative
8	[memang, benar, jelek, jersey, kuning, bayangkan, jersey, timnas, mengamuk]	-13	negative
9	[jersey, erspo, bagus, banget, jersey, kenangan, menangis, menonton, bola, gara, gara, timnas, menang, lawan, korse, piala, asia, pas, penaltinya, arhan, tcoinoomjnt]	-12	negative
10	[terjawab, erspo, nlls, jersey, timnas, cepat, mengalihkan, podkes, botak, dagu, miring, mengalihkan, isu]	-12	negative

Fig. 6: 10 data with the Highest Negative Sentiment

6. Building SVM Models

SVM Accuracy: 0.8235294117647058

SVM Accuracy: 82.35%

SVM Classification Report:

	precision	recall	f1-score	support
negative	0.86	0.84	0.85	114
positive	0.76	0.79	0.78	73
accuracy			0.82	187
macro avg	0.81	0.82	0.82	187
weighted avg	0.83	0.82	0.82	187

Fig. 7: Building SVM Models

7. Confusion Matrix

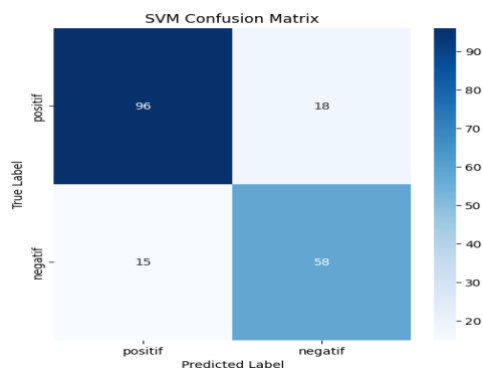


Fig. 8: Confusion Matrix

5. Conclusion

Based on the results of the analysis of public sentiment on social media Twitter on the latest jersey design of the Indonesian national football team using the Support Vector Machine (SVM) method, a total of 308 positive sentiments and 437 negative sentiments were obtained. The SVM classification model showed quite good performance with an accuracy of 82.35%, a margin of error of 17.65%, a precision of 76% for positive sentiment, and 86% for negative sentiment. The accuracy, precision, and recall values achieved indicate that the SVM method can be used effectively in text-based sentiment analysis. These findings also indicate that public opinion tends to be more negative than positive towards the release of the latest jersey design of the Indonesian national team, although positive appreciation is still present among some Twitter users.

References

[1] M. R. Amin and A. N. Salma, "Analisis Sentimen Opini Publik Mengenai Tragedi Kanjuruhan Pada Media Sosial Twitter Menggunakan Brand24," *eProceedings Manag.*, vol. 11, no. 6, pp. 147–153, 2024.

- [2] S. Fauzi, A. P. Siswono, and A.-N. F. Prayitno, "Analisis Sentimen Jersey Erspo Timnas Indonesia Menggunakan Model Klasifikasi dan Algoritma Naïve Bayes," *J. Tekno Kompak*, vol. 19, no. 1, p. 25, 2024, doi: 10.33365/jtk.v19i1.4397.
- [3] D. Prastyo, D. Irawan, and I. H. Mursyidin, "Klasifikasi Sentimen Komentar YouTube dengan NLP pada Debat Pilkada Banten 2024," *bit-Tech*, vol. 7, no. 2, pp. 413–421, 2024, doi: 10.32877/bt.v7i2.1833.
- [4] M. G. R. Lubis, D. S. Sitompul, T. M. Giovanni, F. Ramadhani, and S. Dewi, "Evaluasi Kinerja Algoritma Support Vector Machine (SVM) Dalam Analisis Sentimen Publik Terhadap Naturalisasi Timnas Indonesia di Twitter," *J. Account. Law Commun. Technol.*, vol. 2, no. 1, pp. 81–89, 2024, doi: 10.57235/jalakotek.v2i1.4180.
- [5] S. Styawati, N. Hendrastuty, and A. R. Isnain, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, doi: 10.30591/jpit.v6i3.2870.
- [6] S. D. Wahyuni and R. H. Kusumodestoni, "Optimalisasi Algoritma Support Vector Machine ( SVM ) Dalam Klasifikasi Kejadian Data Stunting," vol. 5, no. 2, pp. 56–64, 2024, doi: 10.47065/bit.v5i2.1247.
- [7] N. Dalifah, N. Suarna, and W. Prihartono, "ANALISIS DATA SENTIMEN NEGATIF PADA OPINI PENGGUNA TWITTER TERHADAP BERITA SEPAK BOLA LIGA 1 TAHUN 2022 DENGAN PENERAPAN SUPPORT VECTOR MECHINE," vol. 8, no. 1, pp. 209–214, 2024.
- [8] A. D. Cahyadi, M. R. Roshan, and M. R. Pribadi, "Applied Information Technology and Computer Science Analisis Sentimen Komentar Twitter Tentang Perfoma Manchester United Dengan Menggunakan Algoritma Support Vector Machine ( SVM )," vol. 3, no. 1, pp. 1–8, 2024.