Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2023. Vol. 3. No. 1; e-ISSN : 2808-4519

Sentiment Analysis Using Text Mining Techniques On Social Media Using the Support Vector Machine Method Case Study Seagames 2023 Football Final

Muhammad Rifa'i[1], Relita Buaton[2], I Gusti Prahmana[3]

[1,2,3]*Information Systems Study Program, STMIK Kaputama, Binjai*
*Mhd.rifai1406@gmail.com**

Abstract

This thesis aims to analyze sentiment on text data from social media related to the 2023 SEA Games, especially in the final match of the soccer sport. The method used is the Text Mining Technique with the SVM (Support Vector Machine) algorithm to classify user sentiment as positive or negative regarding the match. Text data is retrieved from various social media platforms during and after the game. The results of the sentiment analysis are expected to provide insight into the public's view of the sporting event. This research can contribute to the understanding of public sentiment towards the 2023 SEA Games final football match through the analysis of text data from social media.

Keywords: *Sentiment Analysis, Text Mining techniques, SVM, SEA Games, Football.*

1. Introduction

Sentiment analysis on social media is becoming increasingly important for understanding public opinion and consumer behavior in today's digital era. Text mining techniques can be used to process and classify large text data from social media into positive and negative sentiment categories. The Support Vector Machine (SVM) method is one of the popular machine learning algorithms for sentiment analysis. The Support Vector Machine (SVM) method can take into account the subjectivity factor in text data and can assist in identifying sentiment patterns and trends. By using text mining techniques and SVM methods, companies and organisms can process and analyze text data from social media efficiently and accurately to assist in decision making [1].

Current technological developments make it easier for many people to express opinions on online social media as a form of response and response to something. Social Media has also become a very important platform for spreading information and views.

Text mining techniques are used in sentiment analysis to process text data from social media, such as Tiktok, tweets, Facebook statuses, or Instagram comments. This technique involves data collection, data cleaning, data processing, and data classification [2].

Based on the background of the problems above, the formulation of the problem in this study is as follows:

1. How to do sentiment analysis using text mining techniques on text data from social media?
2. How to get large enough text data to train and test the Support Vector Machine model in recognizing and classifying sentiments from text data on social media?
3. How to make a sentiment analysis system from text data on social media using the Support Vector Machine method?

In solving a problem it is necessary to make a problem boundary so as not to deviate from the existing problem, therefore the problem boundaries in this writing are:

1. The SVM(Support Vector Machine) method is a machine learning algorithm that is widely used for data classification in sentiment analysis.
2. The program used is Python software.
3. The variables used are: positive and negative language.
4. The media used is Twitter.
5. This study uses data taken from the Twitter platform related to seagames 2023, especially focusing on the finals of the Football competition.
6. Text sentiment is analyzed into several categories such as positive and negative. Aims to understand the reactions and opinions of social media users.
7. The data taken is in the form of comments related to the case study. This data will be in the form of text in Indonesian.
8. Using a database with excel format.

The objectives of this study are as follows:

1. Finding and collecting large enough text data sources to train and test Support Vector Machine models.
2. Using text mining techniques to process and analyze text data in different languages or non-standard languages.
3. To determine the test results of the Support Vector Machine method with positive and negative resulting data.
4. to analyze the sentiments of social media users regarding the SeaGames 2023 final in the football sport using text mining techniques and the SVM method.
5. This study aims to understand the views and responses of people on social media and evaluate the effectiveness of the SVM model in classifying texts into positive and negative sentiment categories.
6. The results of the research are expected to provide insight into the public's response to the sporting event and the potential for improvement in interpreting sentiment on social media.

The benefits of this research are as follows:

1. Helping decision making in the results of sentiment analysis of various things, such as understanding a person's response, and knowing the views of the public on a particular topic or event.
2. Able to group sentiment data using the Support Vector Machine method.
3. Using the Support Vector Machine method can make it easier to group sentiment data more effectively and efficiently.
4. The benefit of this research is to understand public opinion and sentiment towards the 2023 SEA Games final in sports football through text analysis on social media.

2. Research Methodology

2.1. Data Mining

Data mining is a process of finding meaningful relationships, patterns, and trends by examining large sets of data stored in storage using pattern recognition techniques such as statistical and mathematical techniques. Extracting data information is carried out based on the classification of the data mining method to be used. There are several methods and data mining algorithms that are used to extract data information, including: association methods, clustering methods, classification methods, prediction methods, and estimation methods [3].

2.2. Sentiment Analysis

According to[2]Sentiment analysis is also known as opinion extraction, opinion mining, sentiment mining, and subjectivity analysis. Sentiment analysis is a computational study of people's opinions, appraisals and emotions through entities, events and attributes.

2.3. Text Mining

Text mining is a data mining technique in the form of text. Text taken from social media which of course does not use standard language really needs to be processed using text mining. Typical text mining processes include text categorization, text clustering, concept/entity extraction, granular taxonomy production, sentiment analysis, document inference, and entity relationship modeling [4].

2.4. Twitter

Twitter is a website, owned and operated by Twitter Inc., which offers a social network in the form of a microblog that enables its users to send and read tweets.

2.5. Pythons

Python is an interactive object-oriented programming language that provides high-level data structures. Python is a multipurpose interpretive programming language with a design that focuses on code readability so that the syntax used is easier to understand. Python is said to be a language that combines capabilities, capabilities, with a very clear code syntax [5].

2.6. Steps to the Support Vector machine(SVM) method

Create Word Weighting

$$W_{ij\,ij} = tf \,.\, idf$$

$$w_{permit} = tf \,.\, \log\left(\frac{n}{df}\right)$$

Where:

Wij : term weight (tj) to document (d1),

tfij : The number of occurrences of the term (tj) in (d1),

n : The number of all documents in the database,

df : The number of documents that contain the term

3. Flow Chart

The flowchart system to be designed is as follows:

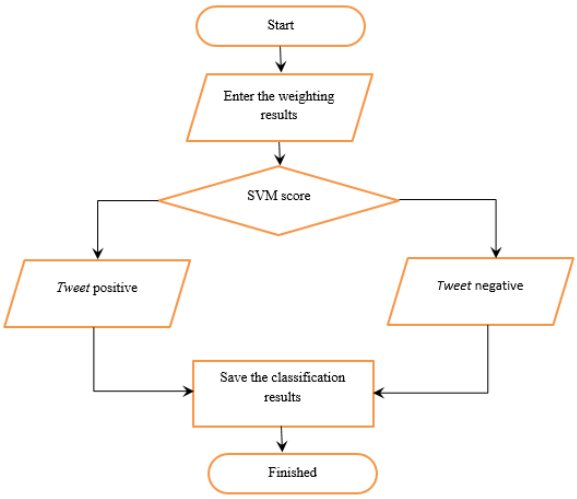

Figure 1: Flowchart flow

Case in point: @FaktaSepakbola @Changsuek_TH @ThaiFootball @TL_Central Asean football level role model
Preprocessing steps [6], [7]:

1. Case folding: The process of this step is to reduce letters or words (lowercase)
2. Tokenizing: To remove certain characters such as punctuation marks. While the space character is used as a delimiter to break sentences into groups of words.
3. Filtering: Used to select stopword words. If there is a stopword, then the word will be automatically deleted.
4. Stemming: the process of filtering words that are changed into basic word forms by using certain rules.
5. If the sentence contains joy, appreciation, pleasure, it is a positive sentence
6. If the sentence contains disappointment, frustration, dissatisfaction including negative sentences

Table 1: Example of Training Data

No	Keyword	Tweets	date	Source	Sentiment
1	@MhmmdRauf13	@FaktaSepakbola @Changsuek_TH @ThaiFootball @TL_Central A role model for ASEAN football	18 june 2023	ticktock	positive

Table 2: Example of Testing Data

No	*Tweets*
1	@FaktaSepakbola @Changsuek_TH @ThaiFootball @TL_Central A role model for ASEAN football

Table 3: Case Folding Process

No	Tweets	Casefolding
1	@FaktaSepakbola @Changsuek_TH @ThaiFootball @TL_Central A role model for ASEAN football	@FactsFootball @changsuek_th @thaifootball @tl_central a role model for ASEAN football

Table 4: Tokenizing Process

No	Tweets	tokenizing
1	@FactsFootball @changsuek_th @thaifootball @tl_central a role model for ASEAN football	@factsfootball @changsuek_th @thaifootball @tlcentral role model levels football asean

Table 5: Filtering Process

No	Tweets	filtering
1	@FactsFootball @changsuek_th @thaifootball @tl_central a role model for ASEAN football	role model levels football asean

Table 6: Stemming Process

No	Tweets	filtering
1	@FactsFootball @changsuek_th @thaifootball @tl_central a role model for ASEAN football	role model levels football asean

3.1. Weighting

In this process, count the number of terms or words that appear in the tweet (tf), count the number of tweets that contain the term (df), calculate the inverse document frequency (idf), and multiply tf by idf as the weight of the term in each tweet [8], [9].

Formula*Tf-idf*

$W_{ij\ ij} = tf\ .\ idf$

$w_{permit} = tf\ .\log\left(\frac{n}{df}\right)$

Wij : term weight (tj) to document (d1),

tfij : The number of occurrences of the term (tj) in (d1),

n : The number of all documents in the database,

df : The number of documents that contain the term

Example of sentences :Asean football level role model

Table 7: Word Weighting Results

Term	tf				df	n/df	idf	tf-idf			
	t1	t2	t3	t4				t1	t2	t3	t4
role model	1	0	0	0	1	4	0.60	0.60	0	0	0
levels	1	0	0	0	1	4	0.60	0.60	0	0	0
football	1	0	1	1	3	1,3	0.11	0.60	0	0.60	0.60
asean	1	0	0	0	1	4	0.60	0.60	0	0	0

4. Program Discussion

4.1. Display Sentiment Analysis

As for the view to analyze using Anaconda Python Jupyter Notebook

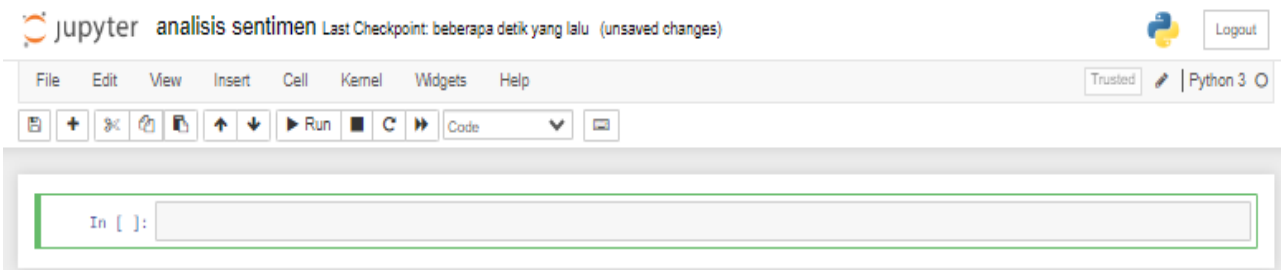

Figure 2: Initial View of the Jupyter Project

4.2. Data Used in Sentiment Analysis

Before starting to enter the syntax, first upload the data that will be used in sentiment analysis.

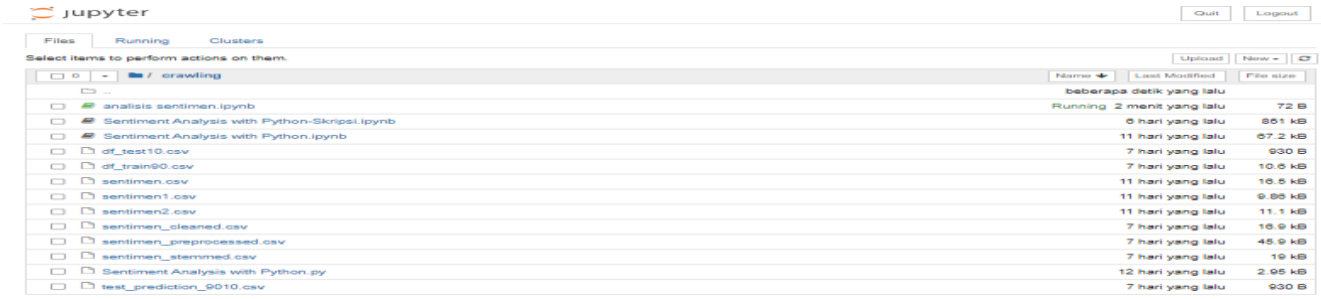

Figure 3: Display of Data Filled in the Sentiment Analysis folder

4.3. Calling Libraries in Python Jupyter Notebook

To load the library in Jupyter Notebook's Python environment, you can use the import command. Here are some examples of calling the library used.

Figure 4: Library view on jupyter

4.4. Preprocessing Stage

Following are the 4 stages of preprocessing syntax and the results of calling data that has gone through 4 stages of processing.

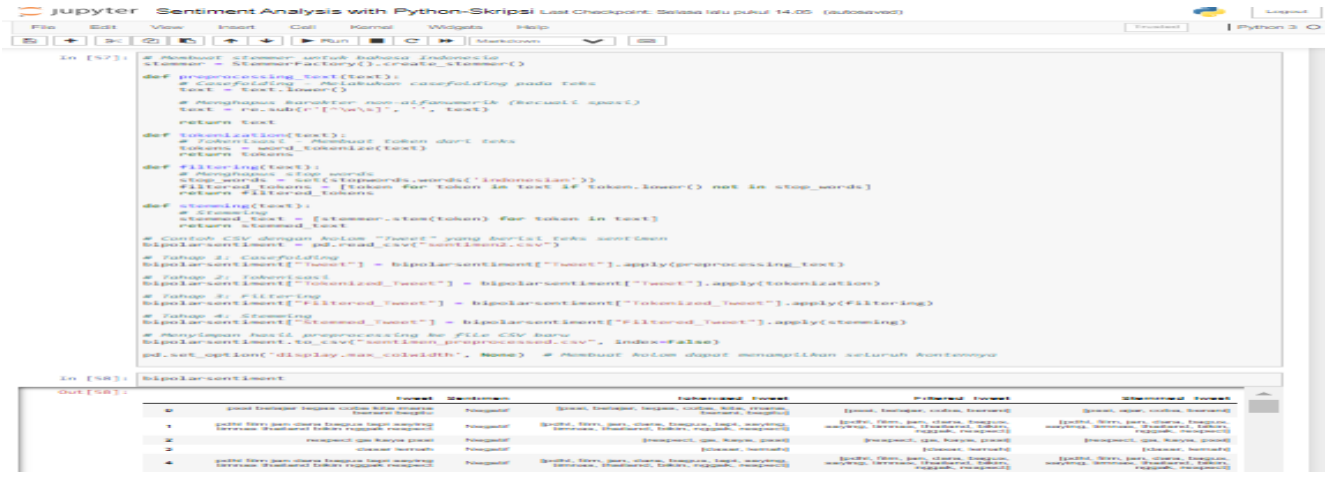

Figure 5: Display of the Preprocessing Stage

4.5. Converting Text Sets to Number Variables

Converting a set of text to a numerical representation is usually necessary in text analysis, especially when you want to apply machine learning algorithms that require numeric input.

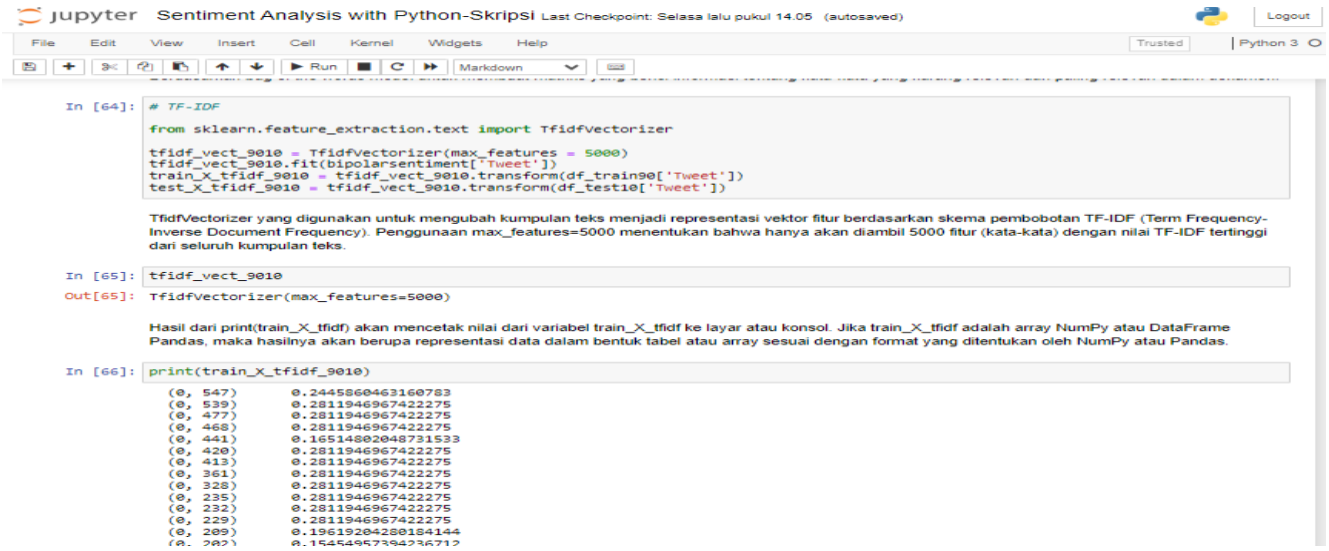

Figure 6: Converting Text to Numbers

4.6. Splitting Document Sentences into Features/Words

Splitting document sentences into features or words is an important step in text preprocessing. You can use a variety of techniques to do this, including tokenization, vocabulary creation, and vector representation.

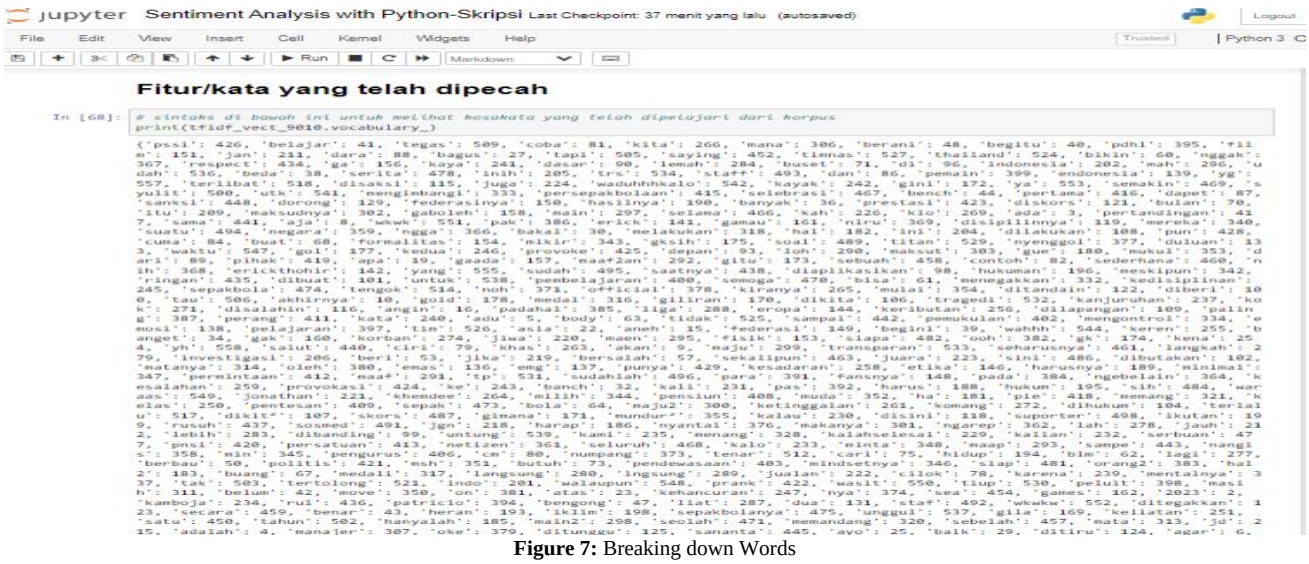

Figure 7: Breaking down Words

5. Conclusion

While the sentiment analysis results provide a snapshot of how social media users respond, we recognize that data from the Twitter social media platform may have limitations. Therefore, these results need to be interpreted with caution and should not be taken as representing absolute truth. Therefore, the core conclusions of the research are as follows:

1. This study succeeded in conducting a sentiment analysis of the responses of social media users regarding the SEA Games 2023 final in the sport of football.
2. The application of text mining techniques allows the identification and extraction of sentiments from social media texts.
3. The SVM method is used as a classification algorithm to predict sentiment from text data.
4. SVM models are evaluated using accuracy, precision, recall, and F1-score metrics to measure model performance.
5. In addition to sentiment analysis, this research also identifies related entities that are often discussed by social media users.
6. The results of the analysis provide valuable insights for stakeholders in understanding social media responses to major sporting events.

Reference

[1] Abdullah, w. F. (2021). Sentiment analysis for handling Covid-19 with a support vector machine: evaluation. 3(2), 160–170.
[2] Alhaq, Z., Mustopa, A., & Santoso, J. D. (n.d.). Application of the support vector machine method for sentiment analysis of Twitter users.
[3] Maulana, m. A., Setyanto, A., & Kurniawan, M. P. (2018). Amikom university social media sentiment analysis. 7–12.
[4] Romadoni, f., Umaidah, Y., & Sari, b. N. (2020). Text mining for analysis of customer sentiment towards electronic money services using a support vector machine algorithm. 09, 247–253.
[5] Saputra, t. I., & Arianty, R. (2019). Implementation of the k-means clustering algorithm on sentiment analysis of Indosat user complaints. Scientific journal of computer informatics, 24(3), 191–198. Https://doi.org/10.35760/ik.2019.v24i3.2361
[6] Sentiment, a., airline, p., husada, h. C., & paramita, a. S. (2021). Sentiment analysis of airlines on the Twitter platform uses the support vector machine (svm) algorithm. 10(1), 18–26. Https://doi.org/10.34148/teknika.v10i1.311
[7] System, r., main, h. S., rosiyadi, d., prakoso, b. S., & ariadarma, d. (2021). Resti journal sentiment analysis of the odd-even system on the Bekasi toll road using. 1(10), 2–8.
[8] Syaifuddin, y. W., & Irawan, R. A. (2018). Implementation of clustering analysis and Twitter data sentiment on beach tourism opinion using the k-means method. Polinema informatics journal, 4(3), 189. Https://doi.org/10.33795/jip.v4i3.205
[9] Wati, R., Ernawati, S., Study, P., Information, S., Bina, U., Informatics, S., Study, P., Information, S., Nusa, U., & Timur, J. (2021). Public perception sentiment analysis regarding ppkm on svm-based twitter using python. 06, 240–247.