

Implementation of Support Vector Machine Method for Twitter Sentiment Analysis Related to Cancellation of u-20 World Cup in Indonesia

Muhammad Armanda¹, Fenina Adline Twince Tobing^{2*}

^{1,2}Informatics Department, Universitas Multimedia Nusantara, Indonesia

¹muhammad.armanda@student.umn.ac.id

²fenina.tobing@umn.ac.id

Accepted 25 June 2024

Approved 03 July 2024

Abstract-The cancellation of the U-20 world cup in Indonesia in 2023 has become a hot debate among the Indonesian people because the reasons for the cancellation are still unclear. The number of pro and con opinions uploaded by the Indonesian people on twitter social media makes these opinions can be used as data to assess opinions which are divided into three categories, namely positive, negative and neutral. After being divided into three categories, sentiment analysis will then be carried out using the SVM method and comparing linear, polynomial and rbf kernels to get the best performance of existing kernels in the support vector machine method. By using confusion matrix to measure the performance of the classification, accuracy, precision, recall and f1-score can be assessed. It was found that the 80:20 data ratio had the highest accuracy of the linear, polynomial, rbf kernel and the rbf kernel had better results than the linear and polynomial kernels, namely Accuracy 78.15%, F1-Score, 76.30%, Precision 77.37% and Recall 75.58%. In addition, the data obtained also succeeded in analyzing Indonesian texts that were input externally and categorized into positive, neutral and negative. From the results that have been obtained, the support vector machine method has been successfully implemented in sentiment analysis of the U-20 world cup cancellation in Indonesia in 2023 on twitter social media.

Keywords-Sentiment Analysis, confusion matrix, U-20 World Cup, Support Vector Machine, Twitter

I. INTRODUCTION

The U-20 world cup is a world soccer event established by FIFA (Federation Internationale de Football Association) and followed by countries from all parts of the world by conducting qualifying rounds and players under the age of 20 years. The U-20 world cup is held every 2 years. The U-20 world cup was first held in 1977 in Tunisia [1]. In 2019 at the FIFA board meeting in Shanghai, China it was determined that Indonesia would host the U-20 world cup beating other candidates such as Peru and Brazil. The U-20 world cup

in Indonesia was supposed to be held in 2021, but the event was canceled due to the unfinished Covid-19 pandemic. Furthermore, the U-20 world cup in Indonesia was determined to have a new match schedule on May 20, 2023 - June 11, 2023 [2].

When the time of the U-20 World Cup in Indonesia was approaching, the Governor of Bali and the Governor of Central Java gave a statement rejecting the presence of the Israeli national team [2]. Reporting from CNBC Indonesia, the rejection was because Israel did not have diplomatic relations and colonized the Palestinian people [3]. Then on March 26, 2023, PSSI (Indonesian Football Association) confirmed that the U-20 world cup phase match drawing schedule was canceled and the drawing was planned to be held on March 31, 2023 in Bali. Reporting from the official FIFA website on March 29, 2023 it was announced that the U-20 world cup in Indonesia was canceled [4]. Of course, this decision raises a variety of pro and con opinions regarding the cancellation of the U-20 world cup to be held in Indonesia which have sprung up to social media. One of the social media used for opinion is Twitter.

Twitter is a microblog social media that allows users to interact with each other and various topics of discussion [5]. Reporting from Data report, the total active twitter users in Indonesia reached 24 million by 2023 [6]. Reporting from We Are Social Twitter is the sixth most social media of total social media users in Indonesia. This number shows that twitter is a social media that has become one of the media in Indonesian society to get information and have opinions in the public sphere [7]. With the many opinions conveyed on twitter social media, these opinions can be used as an assessment material to determine public sentiment regarding the cancellation of the U-20 world cup in Indonesia.

Sentiment analysis is the process of understanding, extracting and processing textual data automatically to find out sentiment information contained in a word [8]. Therefore, from the many opinions conveyed by the public, this sentiment analysis can be useful for assessing and knowing the sentiments held by the public on twitter social media related to the cancellation of the U-20 world cup in Indonesia. Then, the data obtained will be classified into three opinion categories, namely positive, neutral and negative opinions.

Similar research has been conducted by [9] from the study found that the SVM algorithm gets an accuracy of 88.76% and the K-NN algorithm gets an accuracy of 88.1% from these results the SVM algorithm gets better results than K-NN. Similar research has also been conducted by [10] for sentiment analysis using the Naive Bayes, SVM and K-NN algorithms from the results of this study the SVM algorithm has a higher accuracy than the Naive Bayes and K-NN algorithms.

Based on the background of the problems that have been found, sentiment analysis research on twitter social media related to the cancellation of the U-20 world cup to be held in Indonesia aims to find out what sentiments are contained in opinions expressed by the public through twitter social media and the selection of the Support Vector Machine algorithm because from previous research this algorithm can provide better accuracy than other algorithms such as K-NN and Naive Bayes. SVM (Support Vector Machine) algorithm used to classify data in machine learning which aims to find a dividing line or hyperplane with the largest margin, which serves to classify data sets optimally [11].

II. THEORETICAL FOUNDATION

A. U-20 World Cup

The U-20 World Cup is an international soccer event formed by FIFA (Federation Internationale de Football Association) this event is intended for players under the age of 20 years and is held every 2 years. The U-20 World Cup procurement was first held in 1977 and the tournament was held in Tunisia [12]. The first time this sporting event was held was known as the FIFA Junior Cup and was attended by sixteen U-20 national teams from around the world. In 1985 FIFA changed the name of this sporting event to the FIFA World Championship for young players and added a player age limit to 20 years and under [12].

B. Sentiment Analysis

Sentiment analysis is the process of understanding, extracting, and processing textual data to reveal sentiment information contained in a sentence [8]. Sentiment analysis is part of Text mining which means sentiment analysis performs computational research based on sentiment, emoticons, comments and every expression through text. Sentiment analysis is divided into two classifications, namely the classification of opinion or fact documents and the classification of

documents into negative, positive and neutral groups. [13].

C. TF-IDF

TF-IDF method is a method to give weight value to words in a document or sentence. The TF-IDF method combines two concepts in its calculation, namely term frequency and inverse document frequency. Term Frequency is the frequency of occurrence of a word in a particular document or opinion while inversed document frequency is the frequency of documents that have that word. The frequency of occurrence of a word is given to show how important the word is in a document. The frequency of documents containing a word can show how common the word is. This causes the relationship between a word and a document to be high if the frequency of the word is high in a document or sentence and the frequency of all documents containing the word is low in the document collection [14]. To calculate TF-IDF, you can use the following:

- 1) TF calculation formula.

$$tf_{(t,d)} = tf / \max(tf)$$

- 2) IDF calculation formula

$$idf_{(t)} = \log(D / df_t)$$

- 3) TF-IDF calculation formula

$$W_{t,d} = tf_{(t,d)} \times idf_t$$

Description:

$tf_{(t,d)}$ = Term frequency (TF).

$\max(tf)$ = Total of all words in a document.

tf = Term occurrence in the document.

D = Total of all documents.

$idf(t)$ = the occurrence weight of term t across documents.

$W_{t,d}$ = the weight of a word in a document

df_t = number of documents that have term t

D. Support Vector Machine

Support vector machine is an algorithm that uses hypothesis space as a learning system. The Support Vector Machine algorithm is part of data classification that can classify multi-class data. support vector machine algorithm classification uses a hyperplane line as a dividing line if the dividing line is straight called linear and if the dividing line is not straight it is called non-linear [15].

In the classification of models using support vector machines, there are several parameters that are useful for improving the performance of modeling, namely gamma, cost (C) and kernel. The gamma parameter is a parameter that serves to determine how far the influence of the sample dataset is trained and in the

gamma parameter low values mean far while high values mean close, the cost (C) parameter is a parameter used as an optimization of the SVM method to avoid errors in classification on the trained data [16] and the use of kernels in the support vector machine algorithm is useful for transforming data into high-dimensional space [16]. There are several choices of kernel functions that can be used in the classification of the Support Vector Machine method, namely:

1) Linear Kernel

Linear kernel is a simple function in SVM modeling. Linear kernels are useful for analyzing linearly separated data [17]. Equation 2.4 is a linear equation and Figure 1 is an example of using a linear kernel.

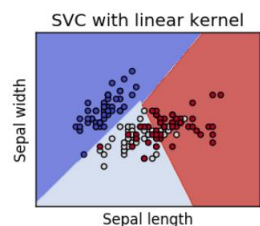


Fig 1. Linear Kernel
Source: [18]

$$K(x, xi) = \text{sum}(x * xi) \quad (2.4)$$

In the kernel equation formula, xi is the training data and x is the Support Vector Machine test data.

2) Polynomial Kernel

Polynomial kernels are used when the dividing line is not linear. Polynomial kernels are used to solve the classification problem of training datasets that have been normalized [17]. Formula 2.5 is a polynomial equation and Figure 2 is an example of using a polynomial kernel.

SVC with polynomial (degree 3) kernel

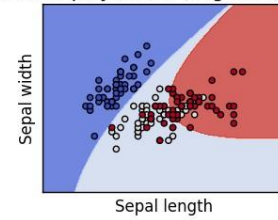


Fig 2. Polynomial kernel
Source: [18]

$$K(x, xi) = 1 + \text{sum}(x * xi)^d \quad (2.5)$$

The polynomial kernel has a degree parameter (d) the function of this parameter is to find the optimal value of the dataset and the greater the degree used, the less stable the resulting performance.

3) Kernel Rbf

Rbf kernel is used when the data used is non-linear. The rbf kernel produces a smaller error value than other kernels [17] Equation 2.6 is the rbf equation and Figure 3 is an example of using the rbf kernel.

SVC with polynomial (degree 3) kernel

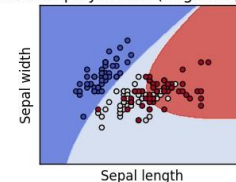


Fig 3. Rbf kernel
Source: [18]

$$K(x, xi) = \exp(-\gamma * \text{sum}((x - xi)^2)) \quad (2.6)$$

RBF kernel has a gamma parameter that serves to determine how far the influence of a sample dataset is trained.

E. Confusion Matrix

Confusion Matrix is used to measure performance in machine learning modeling. Confusion Matrix can consist of two or more classes [19]. Confusion Matrix has a table of four different combinations as seen in Figure 4.

		True Class	
		Positif	Negatif
Predict Class	Positif	TP	FP
	Negatif	FN	TN

Fig 4. Confusion Matrix

Description:

- TP: The prediction result is positive and correct.
- TN : The prediction result is negative and correct.
- FP: The prediction result is positive and false.
- FN : The prediction result is negative and false.

Confusion Matrix has several formulas in its calculations which are used to calculate accuracy, precision, recall and F1-score.

1) Accuracy

Accuracy in Confusion Matrix is a calculation to describe how accurate the classification model that has been made correctly. The following is the formula for calculating Accuracy:

$$Accuracy = (TP+TN) / (TP+TN +FP+FN)$$

2) Precision

Precision is a formula to describe the accuracy between data and classification results with the model. The following is the formula for calculating precision:

$$Precision = TP / (TP+FP)$$

3) Recall

Recall is a depiction of the success of the model applied in retrieving information. The following is the formula for calculating recall:

$$Recall = TP / TP+FN$$

4) F1-Score

F1-score is the average comparison of precision and recal results that have been weighted. The following is the formula for calculating F1-score:

$$F1-Score = 2 \times ((precision \times recall) / (precision+recall))$$

III. METHODOLOGY

In this section, we will describe the flowchart of the system that has been created. The flowchart includes a flowchart of the design overview, Pre-Processing flowchart, Apply TF-IDF flowchart and Apply Support Vector Machine flowchart.

A. Overview of the Design

Overview of the Design is the steps of the design which are organized and described using Flowchart diagrams from the beginning to the end of the design. Seen in Figure 5 the flow of this design starts from crawling data, labeling, Text pre-processing, Train-Test split data, Apply TF-IDF, Apply SVM, Evaluation.

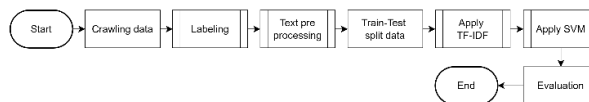


Fig 5. Flowchart of the Design Overview

B. Crawling Data

Data Crawling is a method for retrieving information on a website [20]. In this study, data crawling was carried out using the snsrape library to get the tweets of a number of users with the research topic. In this research, crawling data uses the keyword "U20 World Cup" and the data retrieved is only in Indonesian starting from March 30, 2023 to April 05, 2023. The date was chosen since FIFA announced the cancellation of the u-20 world cup which will be in Indonesia in 2023 and from the results of data collection obtained as much as 14807 tweet data.

C. Data Labeling

Data labeling is used to give sentiment to tweets that have been obtained from the data crawling process. At the labeling stage will use the Valence Aware Dictionary and Sentiment Reasoner (VADER). VADER is a tool in sentiment analysis based on rules specifically tailored to the sentiment being expressed. VADER consists of a sentiment dictionary of words that are generally labeled with positive, negative and neutral words [21].

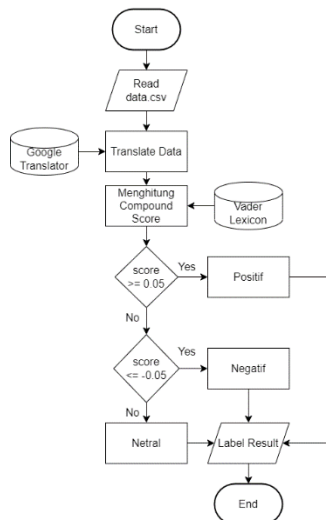


Fig 6. Data Labeling Flowchart

Figure 6 is a flowchart in performing the labeling process. Starting from reading the data that has been obtained from the crawling results then translated data. Translated data is done because VADER cannot read data that uses languages other than English. In the Translated Data process using the Google Translator library. After the data has been successfully translated, the next process is to calculate the compound score using VADER. If the score ≥ 0.05 then the sentiment is positive, if the score ≤ -0.05 the sentiment is negative and if the score is > -0.05 and < 0.05 then the sentiment given is neutral.

D. Text Preprocessing

Text preprocessing stage is a process to select text data so that it becomes structured. In the research conducted by [22] found that the best stage of doing text preprocessing is to do the normalization process first before doing the stopword and stemming stages. Figure 7 is the stage of text preprocessing.

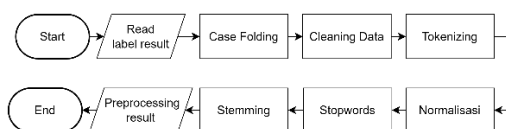


Fig 7. Text Preprocessing Flowchart

1) Case Folding

Case Folding is a process that generalizes the text in the data to lower case.

2) Cleaning Data

This stage performs selection by cleaning the text from usernames, hastags, symbols, emoticons, links, single char, numbers, whites pace and punctuation.

3) Tokenizing

Tokenizing stage will separate the sentence into a list of words. This is done for the next stage, namely the Normalization, Stopwords and Stemming stages.

4) Normlasasi

Normalization aims to correct words that are wrong in writing. Examples of the use of normalization such as the word "sdg" becomes the word "sedang".

5) Stopwords

The stopwords stage serves to eliminate words that have no meaning in sentences such as conjunctions.

6) Stemming

Stemming is the stage of extracting the basic word form by removing affix words.

E. Train-Test Split Data

Train-Test split data is a technique in machine learning to split data into training and testing data. This process is done to assess modeling performance in machine learning. Train-Test Split Data is done before weighting because it is to see the data that is divided still into text so that it can read sentences and compare the labels given before and after modeling with the support vector machine algorithm.

F. Apply TF-IDF

Before classifying the model with the support vector machine algorithm, TF-IDF weighting will be carried out which functions to provide weight values to the data. In Figure 8 after the data has gone through the Train-Test Split Data stage, the next word weighting will be carried out using TF-IDF and produce Train (TF-IDF) and Test (TF-IDF).

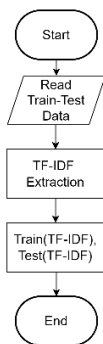


Fig 8. Flowchart of Apply TF-IDF

	Datetime	Username	Text
0	2023-04-05 23:57:19+00:00	halintarbungkam	@ch_cholimah2 @Dennysiregar7 @erickthohir @jok...
1	2023-04-05 23:54:25+00:00	kirarachelsea	Gagalnya perhelatan piala dunia U-20 di Indone...
2	2023-04-05 23:29:31+00:00	teodorikgultom	@tatakujiyati Tapi @aniesbaswedan kan sering m...
3	2023-04-05 23:22:27+00:00	03_nakula	Dicoret sebagai tuan rumah piala dunia U-20, b...
4	2023-04-05 23:18:25+00:00	Aprilia_wiji	Ketua umum Erick Thohir mengungkapkan alasan F...
...
14802	2023-03-30 00:01:26+00:00	PRFMnews	Piala Dunia U-20 Batal Digelar di Indonesia ht...
14803	2023-03-30 00:00:41+00:00	CNNIndonesia	4 Tahun Perjuangan Indonesia Jadi Tuan Rumah P...
14804	2023-03-30 00:00:08+00:00	Eka12Febby	@Metro_TV 3043 guru PI PPPK se-Indonesia dan T...
14805	2023-03-30 00:00:01+00:00	Bolanet	Pemain Timnas Indonesia U-20 ramal-ramal tumpa...
14806	2023-03-30 00:00:00+00:00	voidotdi	Politikus PDIP Budiman Sudjatmiko mengatakan, ...

14807 rows x 3 columns

Fig 10. Data Crawling Results

G. Apply SVM

After the text data has been weighted using TF-IDF, the next step is to implement the support vector machine method to classify the data.

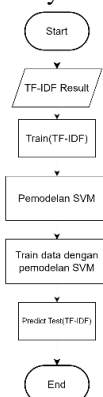


Fig 9. SVM Apply Flowchart

Figure 10 is the result of crawling data that has been done and obtained as many as 14807 tweets. It can be seen that the data frame contains three columns that store dates, usernames and tweets from the results of crawling twitter. After successfully collecting data, the next step is to label the data that has been collected.

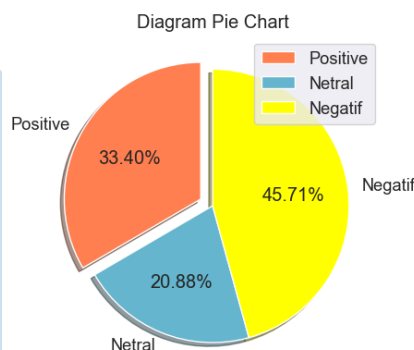


Fig 11. Labeling Results with VADER

Figure 9 the Train(TF-IDF) data goes through a modeling process so that the results of the modeling are used to predict the Test(TF-IDF) data.

H. Evaluation

The evaluation stage will use Confusion Matrix to measure the performance in classifying the data that has been done by the Support Vector Machine algorithm using a table consisting of positive, neutral and negative. The table has a True value which means the true label and predict label predict.

IV. RESULT

A. Implementation

Data collection related to this research will use the snsrape library and pandas library which is useful for storing the data into csv file dataframes.

Figure 11 is the result of labeling using VADER from 14807 data that has been collected. The data shows that 33.40% or 4946 data has positive sentiment, 45.71% or 6769 data has negative sentiment and 20.88% or 3092 has neutral sentiment. After successfully labeling the next step is to do Text Preprocessing. This stage is a process for selecting text data so that it becomes structured. Because the text to be analyzed for sentiment is in Indonesian, the columns that will be stored in the data are Indonesian text and labels. Stages in performing Text Pre-processing are case folding, data cleaning, tokenizing, normalization, stopwords and stemming. Table I is an example of the results of data that has been done a series of text preprocessing.

TABLE I. Example of text preprocessing results

Teks	Text Preprocessing
Intinya dalam salah satu pasal persyaratan ikut FIFA itu bahwa sepakbola ini hanya murni sebatas olahraga jangan ada Intervensi dari pihak pemerintah. ajang piala dunia U-20 sebagai pemersatu anak dari pelbagai bangsa yang hanya mau bermain bola tanpa pedulikan agama atau politik	inti salah pasal syarat fifa sepakbola murni batas olahraga intervensi perintah ajang piala dunia satu anak bangsa main bola peduli agama politik

After the text preprocessing stage has been carried out, the next step is to divide the training data and testing data to measure the performance of the modeling, then weight the data that has been collected using TF-IDF. TF-IDF serves to convert text into numeric data so that the support vector machine algorithm can understand the input text and after the weighting stage is carried out, then classify the data using the SVM method.

B. Testing

The test phase will be carried out with test scenarios and each scenario will be evaluated so that it can be seen which one can provide the best performance. The sentiment analysis test scenario of this research is to compare linear, polynomial and rbf kernels in the Support Vector Machine method and also compare training and testing data as much as 60 to 40, 70 to 30 and 80 to 20. This test is conducted to find which kernel can produce better performance behavior and how influential the comparison of training and testing data is in sentiment analysis. After testing kernels that produce better performance, modeling will be made for training all data that has been collected to test the SVM method in analyzing Indonesian text that is entered or inputted externally.

TABLE II is the result of kernel comparison and division of 14807 data that has been collected as 60% training data and 40% testing data.

TABLE II. 60:40 Dataset Comparison

Matrix %	60:40		
	Linear	Polynomial	Rbf
Accuracy	74.52%	69.42%	76.12%
F1-Score	72.87%	66.11%	74.33%
Precision	73.13%	74.40%	75.51%
Recall	72.63%	63.82%	73.54%

TABLE III is the result of kernel comparison and division of 14807 data that has been collected as 70% training data and 30% testing data.

TABLE III. 70:30 Dataset Comparison

Matrix %	70:30		
	Linear	Polynomial	Rbf
Accuracy	75.39%	70.08%	76.68%
F1-Score	73.87%	66.93%	74.91%
Precision	74.11%	74.72%	76.06%
Recall	73.64%	64.58%	74.12%

TABLE IV is the result of kernel comparison and division of 14807 data that has been collected as 80% training data and 20% testing data.

TABLE IV. 80:20 Dataset Comparison

Matrix %	80:20		
	Linear	Polynomial	Rbf
Accuracy	76.02%	71.47%	78.15%
F1-Score	74.40%	68.65%	76.30%
Precision	74.59%	76.06%	77.37%
Recall	74.23%	66.36%	75.58%

From TABLE II, TABLE III, TABLE IV after the testing process by comparing and dividing the training data and testing data, the results show that the linear, polynomial and rbf kernels get the best assessment of the 80:20 dataset division and the rbf kernel gets better results than the linear and polynomial kernels by getting the results of Accuracy 78.15%, F1-Score 76.30%, Precision 77.37% and Recall 75.58%. From the results of these trials, it can also be seen that the more training data used, the better the performance of the modeling results.

After testing, the rbf kernel obtained better results than the polynomial and linear kernels. The next step is to model sentiment analysis using the rbf kernel and train the model with a total of 14807 data.

	Text	Label-SVM
0	gagalnya piala dunia membuat kita kecewa	negatif
1	tetap semangat buat kedepannya	positif
2	piala dunia membuat masyarakat kecewa	negatif
3	dengan batalnya piala dunia dapat membuat kita...	positif
4	ada apa dengan piala dunia di indonesia ?	netral

Fig 12. Indonesian text testing results

Figure 12, the model can analyze the sentiment contained in Indonesian text and classify it into positive, negative or neutral classes.

V. CONCLUSION

Based on the results of the trials that have been carried out in Twitter Sentiment Analysis Related to the Cancellation of the U-20 World Cup in Indonesia, several conclusions and suggestions can be drawn as follows:

- 1) Twitter sentiment analysis research related to the cancellation of the U-20 World Cup in Indonesia using the Support Vector Machine method has been successfully carried out and can be used for sentiment analysis related to Indonesian language discussion topics. Of the 14807 data that has been collected, labeling is done using VADER, the results of the labeling are 33.40% or 4946 positive sentiment data, 45.71% or 6769 negative sentiment data and 20.88% or 3092 neutral sentiment.
- 2) The results of kernel comparison and data comparison measurements that have been carried out using confusion matrix show that the rbf kernel gets better performance than the linear kernel and polynomial kernel by showing the results of Accuracy 78.15%, F1-Score 76.30%, Precision 77.37% and Recall 75.58%. The division of training data and testing data in the test scenario shows that the more data that is trained in modeling, the better the accuracy results will be.

REFERENCES

- [1] R. Mustajab, "Deretan tuan rumah piala dunia u-20 sejak 1977-2019," 2023. [Online]. Available: <https://dataindonesia.id/ragam/detail/deretan-tuanrumah-piala-dunia-u20-sejak-19772019>
- [2] Y. M. Hendarto, "Membedah kegagalan menjadi tuan rumah piala dunia u20, dari sepakbola hingga politik," 2023. [Online]. Available: <https://www.kompas.id/baca/riset/2023/04/01/membedah-kegagalanmenjadi-tuan-rumah-piala-dunia-u20-dari-sepakbola-hingga-politik>
- [3] C. Indonesia, "Apa penyebab ri batal jadi tuan rumah u-20? ini kata fifa," March 2023. [Online]. Available: <https://www.cnbcindonesia.com/news/20230330092601-4-425726/apa-penyebab-ri-batal-jadi-tuan-rumah-u-20-ini-kata-fifa>
- [4] FIFA, "Fifa removes indonesia as host of fifa u20 world cup 2023™," March 2023. [Online]. Available: <https://www.fifa.com/about-fifa/organisation/media-releases/fifaremoves-indonesia-as-host-of-fifa-u-20-world-cup-2023-tm>
- [5] S. Fransiska and Yolanda, "Analisis sentimen twitter untuk review film menggunakan algoritma naive bayes classifier (nbc) pada sentimen r programming," Siliwangi, vol. 5, no. 2, January 2019.
- [6] S. Kem, "Digital 2023: Indonesia," February 2023. [Online]. Available: <https://datareportal.com/reports/digital-2023-indonesia>
- [7] C. Saskia, "15 medsos favorit orang indonesia, nomor 1 bukan instagram," February 2023. [Online]. Available: <https://tekno.kompas.com/read/2023/02/14/10300097/15-medsosfavorit-orang-indonesia-nomor-1-bukan-instagram?page=all>
- [8] B. Liu, "Sentiment Analysis and Subjectivity," in Handbook of Natural Language Processing, Second Edition, 2010.
- [9] M. R. A. Nasution and M. Natalian, "Perbandingan akurasi dan waktu proses algoritma k-nn dan svm dalam analisis sentimen twitter," INFORMATIKA, vol. 6, no. 2, December 2019.
- [10] J. W. Iskandar and Y. Natalian, "Perbandingan naive bayes, svm, dan knn untuk analisis sentimen gadget berbasis aspek," Siliwangi, vol. 5, no. 6, December 2021.
- [11] N. Yunita, "Analisis sentimen berita artis dengan menggunakan algoritma support vector machine dan particle swarm optimization," SISTEM INFORMASI STMIK ANTAR BANGSA, vol. V, no. 2, December 2016.
- [12] P. Putra, "Sejarah piala dunia u-20: Turnamen paling bergengsi untuk pemain muda," April 2023. [Online]. Available: <https://sepakpojok.republika.co.id/posts/208721/sejarah-pialadunia-u20-turnamen-paling-bergengsi-untuk-pemain-muda>
- [13] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Yusra, "Analisis sentimen akun twitter apex legends menggunakan vader," Nasional Komputasi dan Teknologi Informasi, vol. 5, no. 3, June 2022.
- [14] M. Nurjannah and I. F. Hamdani and Astuti, "Penerapan algoritma term frequency-inverse document frequency (tf-idf) untuk text mining," Informatika Mulawarman, vol. 8, no. 3, December 2013.
- [15] Styawati, N. Hendrastuty, A. R. Isnain, and A. Y. Rahmadhani, "Analisis sentimen masyarakat terhadap program kartu prakerja pada twitter dengan metode support vector machine," Informatika, vol. 6, no. 3, December 2021.
- [16] S. Fransiska and Yolanda, "Komparasi fungsi kernel metode support vector machine untuk analisis sentimen instagram dan twitter (studi kasus : Komisi pemberantasan korupsi)," Sarjana Teknik Informatika, vol. 9, no. 2, Juni 2021.
- [17] Trivusi, "Apa itu kernel trick? pengertian dan jenis-jenis fungsi kernel svm," July 2022. [Online]. Available: <https://www.trivusi.web.id/2022/04/fungsikernel-svm.html>
- [18] scikit learn, "Plot different svm classifiers in the iris dataset." [Online]. Available: https://scikit-learn.org/0.18/auto-examples/svm/plot_iris.html
- [19] M. S. Anggreany, "Confusion matrix," November 2020. [Online]. Available: <https://socs.binus.ac.id/2020/11/01/confusion-matrix/>
- [20] D. Brahma, "Perbedaan antara crawling dan scraping," Feb 2020. [Online]. Available: <https://medium.com/@dede.brahma2/perbedaan-antara-crawlingdan-scraping-98e64e0c6439>
- [21] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," Proceedings of the Eighth International AAAI Conference, 2015.
- [22] S. Shevira, I. M. A. D. Suarjaya, and P. W. Buana, "Pengaruh kombinasi dan urutan pre-processing pada tweets bahasa indonesia," JITTER- Jurnal Ilmiah Teknologi dan Komputer, vol. 3, no. 2, Agustus 2022.