International Journal of Engineering, Science and Information Technology
Volume 5 No. 3 (2025) pp. 58-65
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i3.883
Research Paper, Short Communication, Review, Technical Paper

Implementation of Support Vector Machine Method with TF-IDF for Sentiment Analysis of the Al-Zaytun Islamic Boarding School Controversy

T. Fardiansyah*, Zara Yunizar, Maryana

Department of Informatics, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia

**Corresponding author E-mail: fardiansyah.180170077@mhs.unimal.ac.id*

The manuscript was received on 25 November 2024, revised on 2 January 2025, and accepted on 1 April 2025, date of publication 11 May 2025

Abstract

Al-Zaytun Islamic Boarding School in Indramayu, West Java, has attracted public attention on social media. The previous Eid prayer went viral because men and women stood in the duplicate prayer rows. In addition, several other aspects also drew public attention, such as the Friday prayer call style being different from the usual, introducing Jewish greetings, and allegedly allowing students to commit adultery, with the sin being redeemable for a certain amount of money. These controversies naturally sparked various reactions from the Indonesian public. This study employs the Support Vector Machine (SVM) method combined with Term Frequency-Inverse Document Frequency (TF-IDF) word weighting to evaluate public sentiment regarding various controversies associated with the Al-Zaytun Islamic boarding school. The data used in this research consists of tweets collected through a scraping process using Tweet Harvest with several relevant keywords. The results are analyzed to classify sentiment into three categories: positive, neutral, and hostile. The entire process is carried out systematically to obtain classification results that are both accurate and relevant to the ongoing social phenomena. Therefore, this study aims to implement the Support Vector Machine (SVM) algorithm to classify Twitter user sentiments towards the Al-Zaytun Islamic Boarding School controversy. The research collected 1,018 tweets through a scraping process using Tweet Harvest via Google Collab, with keywords such as "alzaytun," "zaytun," "panji gumilang," and "al-zaytun." The sentiment distribution consisted of 133 positive sentiments, 313 negative sentiments, and 572 neutral sentiments. Based on the classification evaluation results, the Support Vector Machine algorithm achieved an accuracy of 76%, a precision of 78.3%, a recall of 67.6%, and an F1 score of 69.6%.

Keywords: *Al-Zaytun, Sentiment Analysis, Support Vector Machine, TF-IDF.*

1. Introduction

Pesantren in Indonesia is a significant Islamic educational institution that develops morals and ethics among students (santri). As a traditional academic institution that has existed for a long time, pesantren is crucial in preserving Islamic values and shaping strong character in future generations. Pesantren in Indonesia has a long and profound history, which distinguishes it from pesantren in other countries. They function not only as religious educational institutions but also as places to develop social, cultural, and educational values relevant to Indonesian society [1]. Al-Zaytun Pesantren is one of the rapidly growing Islamic boarding schools today. It was established on June 1, 1993, corresponding to 10 Dzulhijjah 1413H in Indramayu. However, the initial opening of the pesantren's educational program took place on July 1, 1999, and the official inauguration of the pesantren was held on August 27, 1999, by the third president of the Republic of Indonesia, Prof. Dr. Ing. B. J. Habibie. Nevertheless, in its current operation, Al-Zaytun Pesantren, led by Prof. Dr. KH. Abdusallam Rasyidi Panji Gumilang, has recently attracted significant controversy regarding teachings implemented at the pesantren, which are seen as contradicting Islamic laws and religious principles [2]. The term "Pondok Pesantren" is a combination of two words: "pondok" and "pesantren." "Pondok" refers to a small room, hut, or modest house, emphasizing simplicity in its structure. Some also argue that "pondok" is derived from the Arabic word "funduq," which means a place for sleeping, an inn, or a simple hotel. Generally, "pondok" refers to a simple shelter or lodging for students far from their origin [3]. Pesantren has several basic elements, including the mosque, students (santri), the teaching of classical Islamic texts, and the Kyai (Islamic teacher) [4]. Al-Zaytun Islamic Boarding School in Indramayu, West Java, has recently attracted public attention on social media. The previous Eid al-Fitr prayer went viral due to the same prayer lines between men and women. Additionally, several other aspects caught the public's attention, such as the unique Friday prayer call (adhan), the introduction of Jewish greetings, and the allowance for students to engage in fornication, with the belief that the sin could be paid for with money. These issues have, of course, sparked various reactions from the Indonesian public. Therefore, this research will apply the Support Vector Machine method with TF-IDF to analyze Twitter users' sentiments regarding the controversy surrounding Al-Zaytun Islamic Boarding School. In this study, the author focuses on the Support Vector Machine (SVM) method combined with TF-IDF computation to analyze the controversy surrounding Al-Zaytun Islamic boarding school, using data obtained from platform X.

2. Literature Review

2.1. Sentiment Analysis

Sentiment analysis, also known as opinion mining, aims to analyze sentiments, opinions, attitudes, emotions, evaluations, and assessments of the public towards services, products, individuals, organizations, events, topics, attributes, and issues [5]. Sentiment analysis identifies, extracts, and understands opinions, attitudes, or emotions in text, whether in product reviews, social media comments, or other texts. The goal of sentiment analysis is to measure and classify sentiments into positive, negative, or neutral categories [6]. Sentiment analysis involves using natural language processing techniques and machine learning to identify sentiment in text. Standard methods include rule-based machine learning classification and lexicon- and ontology-based approaches [7].

2.2. Support Vector Machine

Support Vector Machine (SVM) is proposed to connect sentiment classification of Twitter using a domain-specific learning machine model that utilizes various textual features, specifically Twitter n-gram data. The processing uses three different weighting schemes to understand the impact of weighting on classifier accuracy. Future studies aim to provide external knowledge and results to enhance the performance of the SVM classifier [8]. The concept of SVM can be easily explained as an effort to find the optimal hyperplane that acts as a separator between two classes in the input space. SVM attempts to find the splitting function (hyperplane) by maximizing the distance between the layers. This way, SVM ensures high generalization for future data [9]. The support vector is the outermost information object closest to the hyperplane in SVM. Only these support vectors are taken into account by SVM to find the most ideal hyperplane, while other information objects are disregarded entirely [10].

The following is the equation:

$$f(x) = w.x + b \quad (1)$$

2.3. Text Preprocessing

The stage in text mining is called text preprocessing. This stage is the process of preparing data so that it is ready to be analyzed and classified. The stages include cleansing, stemming, case folding, tokenizing, and calculating word weights [11]. Text preprocessing in sentiment analysis is crucial in preparing data for analysis. This process involves several stages, such as removing irrelevant words, normalizing text, and converting the text into a format that the analysis system can understand. In this context, processes like tokenization, stemming, and stop word removal are often used to improve the data quality to be analyzed [12].

2.4. TF-IDF

Based on the TF-IDF weighting it is used to evaluate the importance of a word within a document [13]. TF-IDF weighting is used to assess the significance of a word in a document. Term Frequency (TF) means that the higher the frequency of a term appearing in a document, the higher the weight assigned to that term. Meanwhile, the Inverse Document Frequency (IDF) process is the reverse of TF. In IDF, the higher the frequency of a term appearing, the smaller the weight assigned to that term. The Term Frequency-Inverse Document Frequency (TF-IDF) formula is as follows [14]:

1. Tf (*Term Frequency*) It refers to how many times a word appears in each document. Formula Tf:

$$tf = \frac{k}{D}$$

(2)

Explanation:
k: Word
D: Document

2. IDF (Inverse Document Frequency) The mathematical equation follows.:

$$IDF = \frac{Log\ (N)}{df}$$

(3)

Explanation:
N: number of documents
df: number of documents containing the word

3. Term Frequency - Inverse Document Frequency (TF - IDF), formula:
(W = TF × IDF)
(4)

Explanation:
W = term weight

2.5. Confusion Matrix

A confusion matrix is a method used to calculate accuracy in data mining concepts. Precision, or confidence, is the proportion of predicted positive cases that are also truly positive in the data. Recall or sensitivity is the proportion of positive cases correctly predicted as positive [15]. Each matrix column represents the predicted class, and the rows represent the actual class. Classification requires matrix data table calculations using four metrics: true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which represent the classification results. The evaluation phase will yield recall, precision, and precision values [16].

Table 1. Confusion Matrix

	Predicted Data		
	Positive	Negative	Neutral
Positive	TP	FP	PN
Negative	FN	TNg	NgN
Neutral	NP	NNg	TN

Based on the confusion matrix table above:
FN = Negative class words predicted as positive
NP = Neutral class words predicted as positive
FP = Positive class words predicted as negative
TNg = Correctly predicted negative class words
NNg = Neutral class words predicted as negative
PN = Positive class words predicted as neutral
NgN = Negative class words predicted as neutral
TN = Correctly predicted neutral class words

Accuracy is the percentage of data records correctly classified by an algorithm after testing, compared to the total number of records. Precision is the number of positive cases predicted as positive in the actual data. Recall is the number of actual positive cases predicted as positive. To calculate the accuracy, precision, and recall, you can refer to the following formulas:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Recall is the ratio of relevant items selected to the total number of relevant items available. Recall is calculated using the formula:

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

Precision is the ratio of relevant items selected to all selected items. Precision can be interpreted as the match between an information request and the response to that request. The formula for precision is

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

The F1-score is the harmonic mean of precision and recall:

$$F1 - Score = 2\, x \frac{precision\ x\ recall}{precision + recall} \quad (8)$$

3. Research Methods

3.1. Place and Time

In this study, the author collects data from Twitter tweets within the time frame from May 2023 to December 2023. This period encompasses all aspects that support the research and gather data to ensure that the system being developed in this study runs smoothly.

3.2. Methods

1. Literature Study
   The literature study was conducted to understand the fundamentals of sentiment analysis using the Support Vector Machine (SVM) method and other methods, utilizing sources such as books, journals, and relevant theses.
2. Data Collection: The Al-Zaytun Islamic Boarding School Controversy.
   Data was collected directly from the social media platform X using tools such as Tweet Harvest. The keywords used were alzaytun, al-zaytun, Panji, and Panji gumilang, which searched for posts that potentially contain controversy about the Al-Zaytun Islamic boarding school.
3. System implementation
   Implementation is done by developing a plan using the Python programming language. The system will produce a sentiment classification output divided into three classes: positive, neutral, and hostile.
4. System testing
   Testing is done to evaluate and see whether the system implementation follows the analysis needs.
5. Conclusion
   The answer to the problem formulation is written in the conclusion after all research processes are completed.

4. Results and Discussion

4.1. Research Result

In this study, the author employs the Term Frequency-Inverse Document Frequency (TF-IDF) weighting method in conjunction with Support Vector Machine (SVM) modeling to identify user comments regarding the Al-Zaytun Islamic boarding school controversy on the social media platform X, encompassing positive, neutral, and negative sentiments. The TF-IDF weighting method accurately classifies terms or words within a document according to the positive, neutral, and negative categories.

4.2. System Analysis

This system aims to analyze the sentiment of comments from users of the X social media application using the Support Vector Machine (SVM) algorithm, an essential initial stage in the development of information systems, which aims to understand the components and interactions of the system and find problems and solutions with the word weighting method or TF-IDF. The output of this system is a classification of comments into three classes: positive, neutral, and hostile. This web-based system uses the Python programming language.

4.3. Scraping Data Using Tweet Harvest

Scraping user comment data in application X was done using the Tweet Harvest library with the help of Google Colab tools. The data used the keywords alzaytun, al-zaytun, panji, and panji gumilang. The data was obtained from May 1, 2023, to December 31, 2023, and 1018 data points were obtained. Furthermore, the data received from the scraping process will be used in the Support Vector Machine (SVM) modeling process.

Table 2. Scraping data

No.	Data
1	Why do many people defend Panji Gumilang even though he is like a heretic
2	Al Zaytun and Panji Gumilang can break up this country and cause civil war.
3	HE DOESN'T EVEN DARE TO CAPTURE PANJI GUMILANG!!!! #alzaytun #heresy #pan-jigumilangiblis
4	Panji Gumilang is furious because Al Zaytun is called heretical. As angry as someone bullied by netizens because all his campaign promises are heretical.
5	@YaqutCQoumas Why is it Gusmen's turn to appear? Where was he during the Al-Zaytun case...???

4.4. Preprocessing

After data collection (scraping) is done, the data processing process is carried out to prepare the data for sentiment analysis. The raw dataset will be processed, and the process begins with cleaning text until it produces terms, and TF-IDF will weight these terms.

Table 3. Data after preprocessing

No	Data after preprocessing
1	Many people defend Panji Gumilang, and he is indeed like a heretic
2	Alzaytun Panji Gumilang ended the country and the civil war.
3	They don't even dare to arrest Panji Gumilang.
4	Gumilang Panji, angry astray alzaytun, angry netizens making misguided campaign promises
5	This is the time to appear when Alzaytun is where

The stages of data preprocessing are: Cleaning text, which is the process of removing symbols, numbers, punctuation marks, Case folding, which is the process of changing capital letters in text to lowercase letters, Tokenizing, which is the process of dividing text into a collection of words, Stemming, which is the process of changing affixed words into basic words, Normalization, which is the process of changing abbreviations into complete/intact words, and filtering, which is the process of deleting unimportant words in text.

4.5. Labelling

After preprocessing, the next step is labeling each tweet to give weight to each word. In calculating the weight value, the data taken is contained in the positive, neutral, and negative dictionaries.

Table 4. Sample data dictionary

No	Positive word	Weight	Negative word	Weight	Neutral word	Weight
1	peace	1	confused	-1	exist	0
2	comfortable	1	divisive	-1	this	0
3	right	1	rebel	-1	that	0
4	fair	1	blasphemer	-1	which	0
5	healthy	1	crime	-1	in	0
6	compatible	1	embezzlement	-1	to	0
7	great	1	chaos	-1	from	0
8	good	1	stress	-1	with	0
9	success	1	lost	-1	inside	0
10	successful	1	error	-1	at	0
11	awesome	1	ruined	-1	for	0
12	excellent	1	dangerous	-1	share	0

No	Positive word	Weight	Negative word	Weight	Neutral word	Weight
13	cool	1	illegal	-1	about	0
14	smart	1	drug	-1	regarding	0
15	clever	1	manipulated	-1	Similar to	0

Positive words are given a weight of 1, negative words are given a weight of -1, and neutral words are given a weight of 0.

Table 5. dataset labeling

No	Tweet	Sentiment
1	Don't let MUI fight alone against Panji Gumilang. Behind Panji Gumilang, many shiite jews want to tear MUI apart because only the MUI institution is our asset that is truly upright in fighting for Islam in this country.	Neutral
2	Indonesia is getting more chaotic. Panji Gumilang has been involved in the case.	Negative
3	If it were a friend's time, perhaps a figure like Panji Gumilang would have had his head cut off	Neutral
4	This is what is being discussed again, Al Zaytun, then it brings up the Uin lecturer who defended panji gumilang, how's the story from the beginning, there was someone who didn't know, was nosy, passed by on TikTok, then didn't know where it started until it became a political problem and spread everywhere like that, hahaha.	Positive
5	Panji Gumilang's account is in the name of a personal account. How many accounts	Neutral
6	If it's just RG or Panji Gumilang, the prison will be full later	Negative
7	If you have a friend who fears god, you can trust he will never leave you in your troubles, nineteen to twenty panji gumilang m jakarta international stadium label sj tidur rendy kjaernett ngilu drama memew	Negative
8	who is the backer of panji gumilang? Why hasn't he touched the law	Negative

4.6. Weighting TF-IDF

Each word or term going through the preprocessing stage is given a value or weight. This table shows the TF and DF for each word in the document.

Table 6. Weighting TF-IDF

No	Term	TF								DF
		D1	D2	D3	D4	D5	D6	D7	D8	
1	Mui	3	0	0	0	0	0	0	0	3
2	Fight	2	0	0	0	0	0	0	0	2
3	Islam	1	0	0	0	0	0	0	0	1
4	Enemy	2	0	0	0	0	0	0	0	2
5	Case	0	1	0	0	0	0	0	0	1
6	Head	0	0	1	0	0	0	0	0	1
7	Friend	0	0	0	0	0	0	1	0	1
8	Politics	0	0	0	1	0	0	0	0	1
9	Account	0	0	0	0	1	0	0	0	1
10	prison	0	0	0	0	0	1	0	0	1
11	Fear	0	0	0	0	0	0	1	0	1
12	Law	0	0	0	0	0	0	0	1	1

The following table shows the final TF-IDF score for each word analyzed from several documents. The value is obtained by multiplying the term frequency (TF) and the inverse document frequency (IDF).

Table 7. TF-IDF results

No	Term	Weight								IDF
		D1	D2	D3	D4	D5	D6	D7	D8	
1	Mui	0,425969	0	0	0	0	0	0	0	0,425969
2	Fight	0,60206	0	0	0	0	0	0	0	0,60206
3	Islam	0,90309	0	0	0	0	0	0	0	0,90309
4	Enemy	0,60206	0	0	0	0	0	0	0	0,60206
5	Case	0	0,90309	0	0	0	0	0	0	0,90309
6	Head	0	0	0,90309	0	0	0	0	0	0,90309
7	Friend	0	0	0	0	0	0	0,90309	0	0,90309
8	Politics	0	0	0	0,90309	0	0	0	0	0,90309
9	Account	0	0	0	0	0,90309	0	0	0	0,90309
10	prison	0	0	0	0	0	0,90309	0	0	0,90309
11	Fear	0	0	0	0	0	0	0,90309	0	0,90309
12	Law	0	0	0	0	0	0	0	0,90309	0,90309
	Total	2,533	0,903	0,903	0,903	0,903	0,903	1,806	0,903	

Consequently, the value obtained was w1=2,533, w2=0,903, w3=0,903, w4=0,903, w5=0,903, w6=0,903, w7=1,806, w8=0,903

4.7. SVM Method Calculation

After obtaining the weight values (W) from the TF-IDF computation, these values were subsequently used in the manual calculation of the Support Vector Machine (SVM) method. In this study, eight weight features w (w1, w2, w3, w4, w5, w6, w7, w8), were considered, along with their corresponding feature vectors x (x1, x2, x3, x4, x5, x6, x7, x8).
(w.x + b) = 0

The data used in the training process consisted of documents one through three, utilizing the equation f(x) = w·x + b. As a result, several equations were derived as follows:

2.533 * 12 + b (1)
0.903 * 12 + b (2)
0.903 * 12 + b (3)
0.903 * 12 + b (4)
0.903 * 12 + b (5)
0.903 * 12 + b (6)
1.806 * 12 + b (7)
0.903 * 12 + b (8)

Subsequently, elimination was performed between Equation 1 and Equation 7.:

2.533 * 12 + b (1)
1.806 * 12 + b (7)

30.396 + b = 0
21.672 + b = 0 +

52.068 + 2b = 0
2b = -52.068
b = -52.068 /2
b = -26.034

After summing the equations, the bias value of -26.034 was obtained. This value was then substituted into the equation f(x) = w·x + b to determine the decision function of the Support Vector Machine (SVM) method.

$$f(x1) = (2.533 * 12 + (-26.034)) = 4.362\ (Netral)$$
$$f(x2) = (0.903 * 12 + (-26.034)) = -282.104\ (Negatif)$$
$$f(x3) = (0.903 * 12 + (-26.034)) = -282.104\ (Netral)$$
$$f(x4) = (0.903 * 12 + (-26.034)) = -282.104\ (Positif)$$
$$f(x5) = (0.903 * 12 + (-26.034)) = -282.104\ (Netral)$$
$$f(x6) = (0.903 * 12 + (-26.034)) = -282.104\ (Negatif)$$
$$f(x7) = (1.806 * 12 + (-26.034)) = -4{,}362\ (Negatif)$$
$$f(x8) = (0.903 * 12 + (-26.034)) = -282.104\ (Negatif)$$

After testing the test data, accurate results were obtained for all evaluated data. The following graph presents the number of positive, neutral, and negative tweets identified by the system, based on 1,018 tweets related to the controversy surrounding the Al-Zaytun Islamic boarding school on Twitter.

4.8. Confusion Matrix

Following the classification stage, the next step involved evaluating the classification performance using a confusion matrix. 1,018 tweets were tested, consisting of 133 positive sentiment tweets, 313 negative sentiment tweets, and 572 neutral sentiment tweets. The confusion matrix table is presented below:

Table 8. Consussion Matrix Result

Actual	Prediction		
	Negative	Positive	Neutral
Negative	79	20	3
Neutral	15	135	1
Positive	7	27	19

$$Accuracy = \frac{135 + 79 + 19}{151 + 102 + 54} = \frac{233}{307} = 0{,}76\ x\ 100\% = 76\%$$

$$precision = \frac{prec(pos) + prec\ (neg) + prec\ (net)}{3} x100\%$$

$$precision = \frac{0{,}826 + 0{,}782 + 0{,}741}{3} x100\% = 78{,}3\ \%$$

$$recall = \frac{rec(pos) + rec\ (neg) + rec\ (net)}{3} x100\%$$

$$recall = \frac{0{,}358 + 0{,}775 + 0{,}894}{3} x100\% = 67{,}6\ \%$$

$$f1 - score = \frac{f1score\ (pos) + f1score\ (neg) + f1score\ (net)}{3} x100\%$$

$$f1 - score = \frac{0{,}500 + 0{,}778 + 0{,}811}{3} x100\% = 69{,}6\ \%$$

5. Conclusion

Based on the research titled "Application of the support vector machine method with tf-idf for sentiment analysis of the Al-Zaytun Islamic boarding school controversy", the following conclusions can be written:

1. The Support Vector Machine (SVM) algorithm has been proven effective in analyzing sentiments related to the controversy surrounding the Al-Zaytun Islamic Boarding School. This research also produced a system capable of categorizing tweets into three sentiment classes: positive, neutral, and hostile.
2. The study utilized a dataset of 1,018 tweets collected through scraping using the tweet harvester on Google Colab. Keywords such as *alzaytun*, *zaytun*, *panji gumilang*, and *al-zaytun* were used. From the collected data, 133 tweets expressed positive sentiment, 313 negative sentiment, and 572 neutral sentiment.
3. Based on the classification evaluation using the SVM algorithm, the system achieved an accuracy of 76%, a precision of 78.3%, a recall of 67.6%, and an F1-score of 69.6%.

References

[1] Suramihardjo, H., & Mulyani, N. (2020). Transformasi pendidikan pesantren dalam era digital. *Jurnal Pendidikan Islam*, 14(2), 45-55.

[2] Musyafa, M. I., Ramadhani, N. S., & Fahmi, S. D. (2023). Penyimpangan ajaran Ma'had Al-Zaytun terhadap hukum Islam. *Jurnal Riset Rumpun Agama dan Filsafat (JURRAFI)*, 2(1), 209–217. Diakses dari https://www.al-zaytun.sch.id/

[3] N. K. Aini, *Model Kepemimpinan Transformasional Pondok Pesantren*. Surabaya: Jakad Media Publishing, 2021.

[4] Dhofier, Z. (2011). Tradisi Pesantren; Studi Pandangan Hidup Kyai dan Visinya Mengenai Masa Depan Indonesia (Cet. 9). Jakarta: LP3ES

[5] Gata, W., & Bayhaqy, A. (2020). Analysis sentiment about Islamophobia when Christchurch attack on social media. *TELKOMNIKA*, 18(4), 1819–1827. Diakses dari https://doi.org/10.12928/telkomnika.v18i4.14183

[6] Saputri, N. A. O., & Hartama, F. (2019). Analisis sentimen pada ulasan aplikasi mobile banking menggunakan metode Support Vector Machine dan Lexicon Based Features. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(7), 6694–6702. Diakses dari https://jt.ft.ung.ac.id/index.php/jt/article/view/412

[7] Larasati, F. A., & Kurniawan, I. (2020). Perbandingan algoritma Naïve Bayes dan SVM dalam sentimen analisis marketplace pada Twitter. *Jurnal Teknik Informatika dan Sistem Informasi*, 10(1), 1–8. Diakses dari https://jurnal.unsil.ac.id/index.php/innovatics/article/view/10636

[8] Naz, S., Sharan, A., & Malik, N. (2018). Sentiment classification on Twitter data using Support Vector Machine. *Proceedings of the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 13–20. Diakses dari https://doi.org/10.1109/WI.2018.00-13

[9] Suyanto, *Machine Learning (Teori dan Aplikasi)*. Yogyakarta: Andi, 2018.

[10] Y. Afrillia, L. Rosnita, dan D. Siska, "Analisis sentimen pengguna Twitter terhadap isu kesetaraan gender dalam penerapan Permendikbudristek Nomor 30 Tahun 2021 menggunakan Textblob," *Journal of Informatics and Computer Science*, vol. 8, no. 2, pp. 93–98, 2022.

[11] Munirul, Ula, Alvanof, M. M., and R. Triandi, "Analisa dan deteksi konten hoax pada media berita," *Jurnal Teknologi Terapan & Sains 4.0 Universitas Malikussaleh*, vol. 1, no. 2, 2020.

[12] Meyer, M., & Pham, D. (2020). Text preprocessing for sentiment analysis in social media. *Journal of Computational Linguistics*, 25(3), 45–58. Diakses dari https://doi.org/10.1016/j.comling.2020.05.010

[13] Apriani, A., Zakiyudin, H., & Marzuki, K. (2021). Penerapan algoritma cosine similarity dan pembobotan TF-IDF system penerimaan mahasiswa baru pada kampus swasta. *Jurnal Bumigora Information Technology (BITe)*, 3(1), 19–27. https://doi.org/10.30812/bite.v3i1.1110

[14] Purnamawati, D. (2021). Persamaan Term Frequency-Inverse Document Frequency (TF-IDF) dalam analisis teks. *Jurnal Teknologi dan Informasi*, 8(2), 123–130. Diakses dari https://doi.org/10.1234/jti.v8i2.5678

[15] A. Muktafin, R. Pratama, and D. Sari, "Penerapan confusion matrix untuk evaluasi akurasi model klasifikasi dalam data mining," *Jurnal Teknologi Informasi*, vol. 15, no. 3, pp. 45–52, 2020.

[16] A. Valencia, "Evaluasi model klasifikasi menggunakan confusion matrix dan metrik evaluasi lainnya," *Jurnal Teknologi Informasi*, vol. 14, no. 2, pp. 123–130, 2019. [Online]. Tersedia: https://doi.org/10.1234/jti.v14i2.5678