Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.58, p.696-709. March/April, 2026.
DOI: https://doi.org/10.5935/jetia.v12i58.2912

RESEARCH ARTICLE OPEN ACCESS

THE USE OF MACHINE LEARNING IN ONLINE COURSE SENTIMENT ANALYSIS: A CASE STUDY POST COVID-19 PANDEMIC

Darwan*[1], Sirojudin Wahid[2], Fajar Hardoyono[3], Indra[4], and Muhammad Hikmal[5]

[1]Department of Informatics, Universitas Islam Negeri Siber Syekh Nurjati Cirebon, West Java, Indonesia;
[2,5]Department of Mathematics Education, Universitas Islam Negeri Siber Syekh Nurjati Cirebon, West Java, Indonesia;
[3]Department of Environmental Science, Universitas Islam Negeri Professor Kiai Haji Saifuddin Zuhri, Purwokerto, Central Java, Indonesia;
[4]Information Technology, Universitas Budi Luhur, Tangerang, Indonesia.

[1]https://orcid.org/0000-0003-3744-7351, [2]https://orcid.org/0000-0002-0468-1574, [3]https://orcid.org/0000-0003-2813-9568, [4]https://orcid.org/0000-0002-8180-4849, [5]https://orcid.org/0009-0009-5410-3033

Email: *darwan@uinssc.ac.id, sirojudin.wahid@syekhnurjati.ac.id, hardoyono@uinsaizu.ac.id, indra@budiluhur.ac.id, muhammad.hikmal116@mail.syekhnurjati.ac.id

ARTICLE INFO

Article History
Received: November 8, 2025
Reviewed: December 19, 2025
Accepted: March 10, 2026
Published: April 30, 2026

Keywords:
Online Lectures,
Twitter,
Naïve Bayes,
Decission Tree,
Support Vector Machine.

ABSTRACT

The number of tweets on Twitter containing information related to online lectures in the post-pandemic Covid-19 period or after the government lifted the pandemic period has drawn many pros and cons. This research discusses the best level of accuracy of the three machine learning methods, namely Naïve Bayes (NB), Decision Tree (DT) and Support Vector Machine (SVM) in analyzing Twitter results related to online lectures. The data used is crawled data from Twitter using the keyword "Online Lecture" and 5,978 tweets were obtained after verification. Three category labels were used in this research, namely Positive, Negative and Neutral. Pre-processing was carried out using the stages of Data Cleansing, Tokenizing, Stopword, Normalization and Stemming. While the feature extraction stage uses the TF-IDF stage. The average accuracy results for the NB method are 78.02%, precision is 78.18%, recall is 78.02% and f1-score is 77.45%. DT obtained an average accuracy of 86%, precision is 89.14%, recall is 81.14% and f1-score is 82.31%. The average accuracy obtained by SVM is 89%, precision 90.97%, recall 87.31%, and the F1 score is 85.99%. Among the three algorithms, it is known that SVM achieves very good performance in online lecture sentiment analysis based on Twitter data.

Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

During the Covid-19 pandemic like a few years ago, it had a huge impact on human activities. All activities that were previously carried out as usual had to be done digitally. Among those that were greatly felt and had a big impact was the world of education, for almost two years learning was carried out from home through online learning. Indonesia is no exception, for more than two years the government and related agencies have tried to ensure that learning continues as it should even though it is carried out with online learning. There are many responses from the public regarding the implementation of online lectures, especially during the pandemic. Pros and cons will certainly spark debate in expressing their opinions. As stated by [1] stated that the online learning system carried out during the COVID-19 pandemic is effective and inefficient.

Effective because the conditions require online learning and inefficient because the costs incurred are higher when compared to offline lectures. The costs incurred are mainly to buy internet quota to attend online lectures and the obstacles faced during online lectures are network constraints. Meanwhile, [2] related to obstacles to online learning seen from internal and external factors. Internal factor obstacles are first students and lecturers experiencing miscommunication (25.71%), second decreased motivation (24.28%), and third poor time management (17.15). Meanwhile, the external obstacles are, firstly, related to the internet network signal which is unstable or difficult in the residential environment (55.13%), secondly the availability of internet quota (12.97%), and thirdly the limited facilities/gadgets (9.18%).

The same thing was conveyed by [3] whose conclusion was that students most often experience signal interference, which will later have an impact on unclear images or sound, as well as intermittent signal interruptions. This is almost the same as [4] research, he even added that among them are students without cellphones, lack of student attention, and parents' working being disrupted. Once conveyed almost the same thing, namely regarding inadequate infrastructure, students' difficulties in understanding the material, and the limited ability of lecturers to utilize technology [5]. Although in this case lecturers are not constrained in terms of delivering material to students, it should be underlined that successful learning requires an effort to strengthen students [6]. In addition to the above, there is also research by [7] which states that their findings summarize students' perceptions of learning and their involvement when given a choice about how to engage in content.

Most participants chose asynchronous methods to learn content, citing time constraints as the main reason for not choosing synchronous learning methods. Furthermore, [8] emphasized the importance of considering emotions in distance learning environments and their significant impact on academic performance and participant satisfaction. Even [9] research explains the common challenges and opportunities faced by participants and educators in thirty countries, based on sentiment analysis from these stakeholders, the results show that although digital education is developing into a flexible and student-centered didactic approach, various obstacles in meeting online learning effectively still exist, where the findings reveal the shortcomings, and therefore the need for, appropriate teaching and learning materials and strategies for digital education.

This is also almost the same as that conducted by [10], which found that most educators experience problems during online learning practices. There are even several studies that conduct literature review related to open learning and online learning, such as those done by [11] and [12]. From these limitations, sometimes they complain through social media such as Facebook, Instagram or Twitter. Of the many social media that are often used, Twitter is one of the social media that is often used by its users to interact. Based on the 2024 internet penetration and behaviour survey data issued by APJII (Association of Indonesian Internet Service Providers) [13] which states that the reason for using the internet is to be able to access social media (including Twitter), which is 3.31%. While the social media that is often used for Twitter is 1.06%, and the chat media that is often used in this case is Twitter, which is 0.34%.

Publishing and disseminating news information is very suitable for using Twitter. Very short tweets are very easy for users to post using handphone or other devices. While the 140 characters posted by some users are considered short and easy to spread information [14]. Even [15] conducted research to identify content containing propaganda by collecting data from Twitter. Twitter as we know after crawling will get a very large amount of data information so that data mining is needed. The application of data mining is widely used in various researches, one of which is text mining. The use of text mining is very relevant to analyse text-based case studies, one example of which has been mentioned above is the use of Twitter. In the field of education, it is widely used to analyse sentiment towards the world of education. The sentiment can be positive or negative. With sentiment results like we can process the results of public opinion in the world of education, one of which is online lectures.

In this research, we will examine the sentiment analysis of online learning based on Twitter data by comparing three methods (Naive Bayes, Decision Tree & SVM). The main reason for choosing Naive Bayes is because it is a fairly simple algorithm but in several researches it has a fairly high level of accuracy [16]. While Decision Tree is one of them in decision making which was previously very global and complex can be changed to be simpler and more specific [17], and is easy to present and implement in the form of diagrams [18]. The third is SVM that can perform a more accurate classification process [19]. From the advantages of the three algorithms, researchers try to compare by looking at performance, namely in the form of accuracy percentage. It is expected that the results of this sentiment analysis research based on Twitter data can be a reference and help the world of education in efforts to improve the quality of learning for students in the future, and these results are expected to be a reference for comparing with other methods.

II. LITERATURE REVIEW

Researchers conducted a literature review of several researches related to sentiment analysis based on Twitter data. According to the results of the literature review, it was concluded that research on sentiment analysis based on Twitter is still quite extensive to discuss. Research on Twitter sentiment analysis is still limited to online lectures and there are still few researches that discuss sentiment analysis. This research focuses on sentiment analysis from Twitter data, using the Long Short-Term Memory (LSTM) algorithm and Artificial Neural Network (ANN) with an accuracy rate of 96% for LSTM and 76% for ANN [20]. Research on sentiment analysis for finance using the BERT method achieved an accuracy of 95.29%, precision of 95.37%, and recall of 95.24% [21]. Research uses NLP (Natural Language Programming), his research provides an overview of sentimental investigation or opinion mining related to product surveys [22]. [23] his research is related to sentiment analysis which aims to provide a systematic review of sentiment analysis and related fields for future research. Proposed the use of the Support Vector Machine (SVM) model to develop an automatic Twitter crawling system and achieved an accuracy level of 85% [24].

Used the VADER model to identify people's opinions about online learning [25]. Proposed a machine learning algorithm for Arabic sentiment analysis which was used to find out the opinions of students and teaching staff about online learning during the COVID-19 outbreak in Saudi Arabia [26]. Conducted their research to study online learning during the pandemic situation and the adaptation of online learning worldwide after the pandemic situation using machine learning such as SVM and artificial neural networks (ANN) [27]. According to the results, ANN was better than SVM and achieved 81.97% accuracy with higher precision, recall, f-measure value, and lowest error value.

Research [28] present the results of the application of BERT, these results will be compared with the contribution of the author [29], where the results using BERT are quite good (88.93% accuracy). Research uses the SentiStrenth method, with 54% neutral results, while 85% accuracy rates, and 15% error rate [30]. Research explores the experience of students engaging with recorded lectures and the extent to which such engagement contributes to improved learning [31]. This research involved distributing questionnaires to undergraduate and postgraduate students (n = 660) who had access to lecture recordings at a public university in New Zealand, the use of sentiment analysis shows how multiple data can be triangulated to confirm or contradict research findings.

New ensemble model was proposed for sentiment analysis, the experimental results showed that the sentiment recognition accuracy of the proposed model was higher than that of the other seven comparison models, with an F1score of more than 91%, and the results of recognizing different levels of emotion showed the stability of the proposed ensemble model [32]. Research used the sentiment analysis method to identify and extract student responses in subjective information focusing on positive and negative polarity, and their research findings showed that the results of sentiment analysis using the SVM method obtained an accuracy of 84% [33].

Research used Chi-square and SVM algorithms, the results of their research assist university administrators and lecturers to identify the strengths and weaknesses of lecturers based on textual evaluations conducted by students [34]. [35]Research related to the use of LMS developed at Anadolu University, the results of their research concluded that machine learning techniques can be used to better understand students and how they feel [35]. From several previous researches that have been explained above, the differences can be seen in Table 1. From several researches, various methods are used to obtain better results.

In this research, the researcher tries to compare three machine learning methods, namely Naïve Bayes, Decision Tree, and Support Vector Machine. The three methods will be seen from their performance level, namely which level of accuracy is better. It is because from several previous researches, most researches only use one or two methods, and even if there are, they are not exactly the same using the same method with the same case study. As far as the researcher knows, there has been no research that explicitly discusses sentiment analysis based on Twitter by comparing three methods (Naïve Bayes, Decision Tree & SVM).

Table 1: Literature Review.

Researcher	Years	Methods
Valarmathi et al	2024	- LSTM (Long Short-Term Memory) - ANN (*Artificial Neural Network*)
Adelakun and Baale	2024	BERT
Namitha et al.	2023	NLP (*Natural Language Programming*)
Pulikonde et al.	2023	Sentiment Analysis
Sajinika et al.	2023	VADER
Alqaan and Qamar	2022	- SVM (Support Vector Machine) - KNN (k-Nearest Neighbor) - Random Forest
Senadhira et al.	2022	- SVM (Support Vector Machine) - ANN (*Artificial Neural Network*)
Ngoc and Thi	2021	BERT
Hardian et al.	2021	SentiStrenth
Nkomo and Daniel	2021	Questionnaire
Pu et al.	2021	Ensamble
Kurniawan and Wahyuni	2021	SVM (Support Vector Machine)
Ayeni et al.	2020	SVM (Support Vector Machine)
Osmanoglu et al.	2019	Logistic regression

Source: Authors, (2026).

III. MATERIALS AND METHODS

This research follows the proposed methodology sequence such as data collection, pre-processing, feature extraction, classification and evaluation (Figure 1. Proposed method).

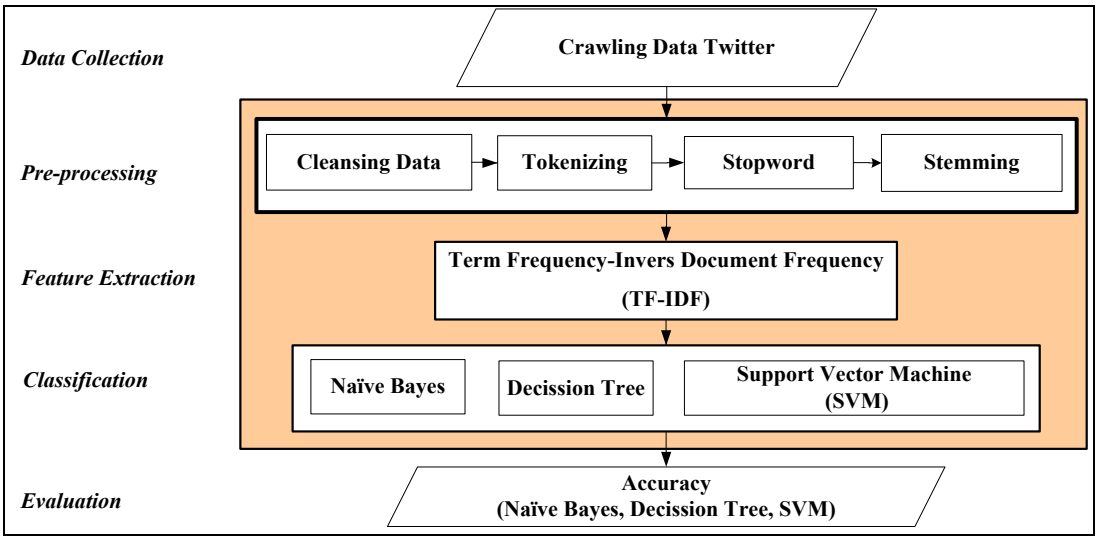

Figure 1: Proposed Method.
Source: Authors, (2026).

III.1 DATA COLLECTION

The first research flow in this research is data collection. The data collection in question is used to obtain a dataset for the analysis process which in this case uses Twitter Harvest or API from Twitter. The keyword used in the data retrieval process is "Kuliah Online" (Indonesian word), which in English means "Online Lectures". The data that will be used is crawling data from Twitter.

III.2 PRE-PROCESSING

Pre-processing in this research is the stage of cleaning data from a sentence or word, link, or things that are not needed to perform sentiment analysis. The dataset that has been collected is still in the form of unstructured or irregular words or sentences, so it still requires a process called data pre-processing to produce clean data, so that the sentiment analysis results are expected to be accurate. The data pre-processing stages consist of cleansing data, tokenizing, stopwords, normalization and stemming. Cleansing data is carried out to reduce unimportant words (noise) in tweet data, such as URLs, hash tags (#), usernames (@username), emoticons (: @,: *,: D), emails, and punctuation [36]. Data changes at this cleaning stage, such as lowercase, can also occur [30].

Tokenizing is a process that functions to break sentences into words. The tokens that have been produced will later be in the form of single words which will make the next process easier [36]. Stopword is a process of deleting or eliminating words resulting from the previous process, namely tokenizing, which have no meaning, such as "di", "yang", "itu", "ke", and so on. Normalization is used to change data into a more standard form [37]. Normalization can help improve accuracy in text analysis and make data easier to process [38]. Stemming is the process of changing words that have affixes to become basic words according to standard Indonesian language rules [30].

III.3 FEATURE EXTRACTION

The third process is feature extraction. At this stage, the TF-IDF (Term Frequency-Inverse Document Frequency) method is used. TF-IDF is used to determine how far the word (term) is related to a document by giving each word a weight. This method will calculate each term's weight in the document. The equation used in this process is [39]:

$$TFIDF(\ t\ ) = TF * log\ \frac{N}{df} \tag{1}$$

Where t is the term, N is the total documents, TF is the number of terms t that appear in the document, and df is the number of documents containing the term t.

III.4 CLASSIFICATION

This stage classification is used to group data or sentences based on their similarities [40]. The classification in the sentiment analysis process is often modelled as two-way (positive/negative) or three-way (positive/negative/neutral) [41]. In this classification stage, three outputs will be classified, namely positive, negative and neutral and will compare three methods by looking at the level of good accuracy, namely Naïve Bayes, Decision Tree and Support Vector Machine.

III.4.1 Naïve Bayes

Naïve Bayes is a simple probabilistic classifier that calculates a set of probabilities by summing the frequencies and combinations of values from a given dataset [42], a simple probability classification process that refers to Bayes' theory. In other words, given an output value, the probability of observing it together is the product of the individual probabilities. The general equation used in the Naïve Bayes algorithm is:

$$\boldsymbol{P(H|X)} = \frac{\boldsymbol{P(X|H).\ P(H)}}{\boldsymbol{P(X)}} \tag{2}$$

Where:
H: Data hypothesis is a specific class
X : Data with unknown classes
P(H) : Probability of hypothesis H
P(X) : Probability X
P(H|X) : Probability of hypothesis H based on condition X
P(X|H) : Probability X based on the conditions in hypothesis X

III.4.2 Decision Tree

Decision tree is a prediction model that can be used to represent classification and regression models in operations for decision making with strategies that are most likely to achieve goals [43]. In developing a Decision tree, the initial stage is to evaluate all existing attributes using a statistical measure (which is commonly used is information gain) to measure the effectiveness of an attribute in classifying a collection of data samples. Attributes with continuous values must be discretized [44]. To find the gain value (information Gain) can be carried out with the concepts of Entrophy, Gini Index and Classification Error [45].

- Information Gain with Entrophy value

$$Entrophy(s) = \sum_{i=1}^{n} p(i|s) log_2 p(i|s) \tag{3}$$

Where:
S : Case sets in a dataset
n : Number of partitions attribute S
Pi : The proposition of Si to S

This entropy value will be used to calculate the gain value. The formula for finding gains is as follows:

$$Gain\ (S,A) = Entrophy(S) = Entrophy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Entrophy\ S_i \tag{4}$$

S : Case set
A : Attribute
n : Number of partitions in attribute A
| *Si* | : Number of cases in partition 1
| *S* | : Number of cases in S

- *Gain information with Gini Index value*

$$Gini\ (t) = 1\ \sum_{i=1}^{n} [p(i|s)]^2 \tag{5}$$

$$Gain(S,A) = Gini(S)\ \sum_{i=1}^{n} \frac{|Si|}{|S|}\ Gini\ (Si) \tag{6}$$

- Gain Information with Classification Error value C. Error is obtained from the value of the attribute that is the smallest of the class label.

$$C.Error(S) = 1\ max_i\ [p(i|s)] \tag{7}$$

$$Gain(S,A) = Error(s) \sum_{i=1}^{n} \frac{|S_i|}{|S|}\ Error\ (S_i) \tag{8}$$

III.4.3 Support Vector Machine

Support Vector Machine can make predictions for both classification and regression [46]. SVM can be used on different types of data sets, namely data sets with linear and non-linear separation. The Support Vector Machine (SVM) method allows calculations for linear problems by applying mathematical transformations to the learning space using kernel functions [47]. Support Vector Machine (SVM) has a central concept in classifying data, namely determining the best hyperplane to provide a distance or separator between two predetermined classes [48]. Hyperplanes that have large margins can classify data that has not been seen well, by providing a large margin; new data can enter each class [49] where the margin itself can be described as in Figure 2 below.

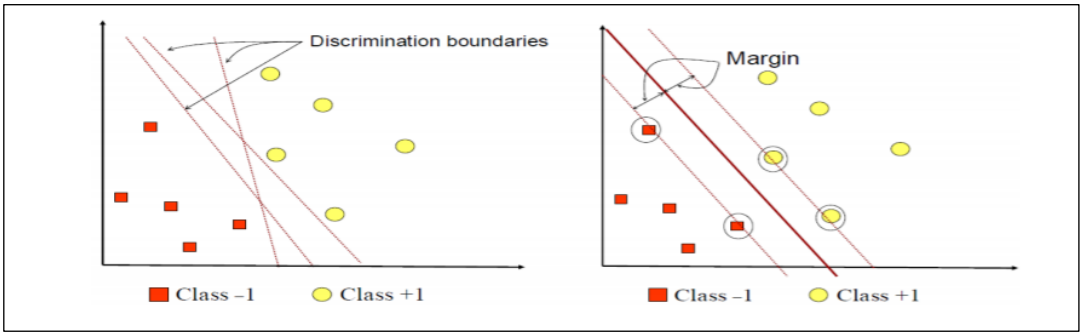

Figure 2: Best Hyperplane.
Source: [50].

In overcoming the Support Vector Machine in non-linear cases that occur, the kernel method is used to overcome it. A function of the kernel is to transform data into a space or other dimension which can be illustrated in Figure 3 below [49].

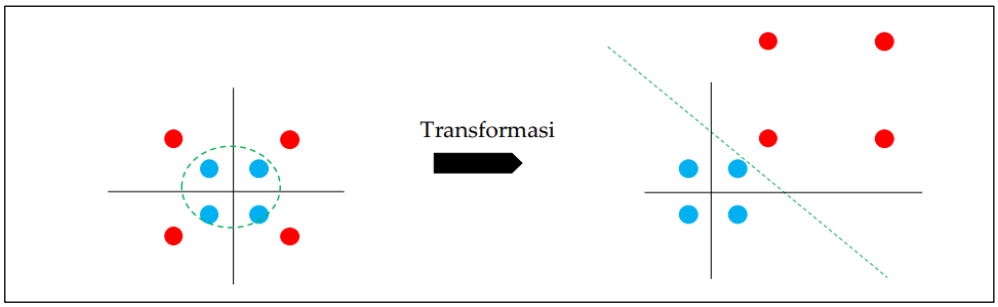

Figure 3: Data Transformation.
Source: [49].

By using the kernel method, a data x is input into a higher-dimensional feature space F. Feature mapping from two-dimensional space to two-dimensional feature space [51].

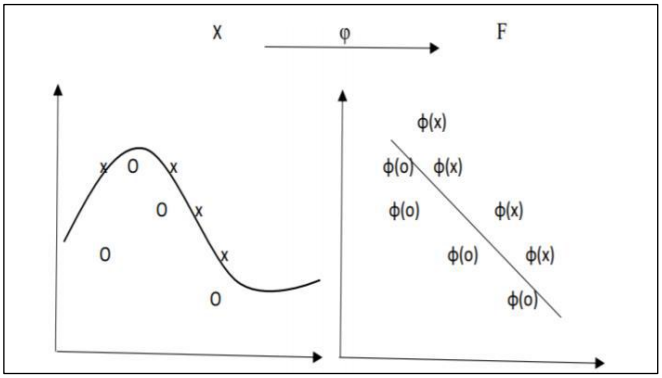

Figure 4: SVM Non Linear Kernel Map.
Source: [51].

There are several kernel functions that are usually used in SVM including linear, polynomial, radial basis function, and sigmoid. The stages in the Support Vector Machine method are as follows [52]:

1. Determine the most frequently occurring words from each document used
2. Determine the initial initialization for the values of α, C, λ, gamma and epsilon.
3. Calculate the matrix with the formula as below:

$$D_{ij} = yi\ yj\ ((xi^{\rightarrow\rightarrow}.xj^{\rightarrow\rightarrow}) + \lambda\ 2) \tag{9}$$

Explanation:
D_{ij} = data matrix element to -ij
yi = class or label of data to -i
yj = class or label of data to -j
λ = theoretical limit derivative
K($xi^{\rightarrow\rightarrow}.xj^{\rightarrow\rightarrow}$) = kernel function

4. For the nth power = 1,2,3,.. n use the equation below:

$$E_i = \sum \alpha i\ n\ j = 1\ Dij \tag{10}$$

$$\delta\alpha i = \min\{\max[\ \gamma(1 - Ei), - \alpha i\ ], C - \alpha i\ \} \tag{11}$$

$$\alpha i = \alpha i + \delta\alpha i \tag{12}$$

Explanation:
Ei = data error value to -i
γ = learning levels
$Max_{(i)}$ D_{ij} = maximum value of the diagonal of the equation matrix

5. Find the bias value (b) using the equation below:

$$b = -\ 1\ 2\ [w\ .\ x\ +\ +\ w\ .\ x\ -] \tag{13}$$

6. Testing on the document being tested
7. Decision calculation

If the decision calculation is more than or equal to 0 then the *h(x)* sign value is +1, then it is included in the positive class and if the decision calculation result is less than 0 then the *h(x)* sign value is -1, then it is included in the negative class. Decision calculation using the equation as below:

$$h(x) = w\ .\ x + b \text{ atau } h(x) = \sum \alpha i\ yi\ K(x, xi\ m\ i = 1) + b \tag{14}$$

III.5 EVALUTION

In this final stage after a series of processes are carried out, the next step is to evaluate. This is carried out to obtain the level of accuracy of a method used in analyzing sentiment from public opinion. In this evaluation, the k-fold cross validation method is used to test the accuracy of the three methods used (Naive Bayes, Decision Tree and SVM). The k-fold cross validation method all data used will be treated the same, namely as testing and training data. In general, k-fold cross validation, where k ≥ 2 is an integer, and the dataset used will be partitioned into k parts [53].

Table 2: Illustration of 3-Fold Cross Validation.

	Experiment 1	Experiment 2	Experiment 3
A (data 1 – 10)	A & B (Data Training)	A (Data Training)	A (Data Testing)
B (data 11 – 20)		B (Data Testing)	B & C (Data Training)
C (data 21 – 30)	C (Data Testing)	C (Data Training)	

Source: Authors, (2026).

From the k-fold cross validation method, the level of accuracy, precision, recall and F1-Score can be obtained using the confusion matrix method. With the confusion matrix we can analyze how well the categories we use can recognize data from different categories [44].

Table 3: Model confusion matrix.

Correct classification	**Classified As**	
	+	-
+	TP	FN
-	FP	TN

Source: Authors, (2026).

Accuracy, precision, recall, f1-score are obtained by equations (14) to (17) [44]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} x\ 100\% \tag{15}$$

$$Precision = \frac{TP}{TP+FP} x\ 100\% \tag{16}$$

$$Recall = \frac{TP}{TP+FN} x\ 100\% \tag{17}$$

$$F1\text{-}Score = \frac{2TP}{2TP+FP+FN} x\ 100\% \tag{18}$$

Where, TP = *True Positive*, TN = *True Negative*, FP = *False Positive,* dan FN = *False Negative*.

IV. RESULTS AND DISCUSSIONS

This section discusses the results and discussions related to the proposed method. This section starts from data collection, data processing, per-processing, feature extraction, and classification.

IV.1 DATA COLLECTION

The data used is data collected from Twitter in the form of tweets. The data collection process is carried out through the data crawling process, in this case using Tweet Harvest (Version 2.6.1). One of the advantages of Tweet Harvest is that it allows users to collect more data and can be run in the form of a Command Line Interface (CLI) and only requires an auth_token [54].

```
#Twitter Auth Token

twitter_auth_token = '[redacted]'
```

Figure 5: Token Authoring.
Source: Authors, (2026).

The keywords used in the data retrieval process refer to “Online lectures”. Data collection was carried out from June 13-25, 2024. The data crawled was tweet data from June 23, 2023 to June 22, 2024. The crawled data retrieval started from June 23, 2024, this was carried out after the issuance of Presidential Decree No. 17 of 2023 concerning the Determination of the End of the Corona Virus Disease 2019 (COVID-19) Pandemic Status in Indonesia, which was set on June 22, 2023. This is very relevant because the tweet data retrieval was carried out after Indonesia was declared free from the pandemic, so that we can obtain tweets from the public regarding the experience of online lectures during the pandemic until the pandemic is over.

	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name
0	1250187762684190720	Tue Apr 14 23:56:49 +0000 2020	0	@masgib___ Anak kuliahan sibuk kuliah online l...	1250211440943718400	NaN	chllghbrn_
1	1250211296345120768	Tue Apr 14 23:56:15 +0000 2020	1	Udah kuliah online ngerjain tugas udah jungker...	1250211296345120768	NaN	NaN
2	1250209036177301504	Tue Apr 14 23:50:50 +0000 2020	0	@bintangemon Mandi pagi mau kuliah online	1250209934739824640	NaN	bintangemon
3	1250209679583506432	Tue Apr 14 23:49:50 +0000 2020	0	Kuliah online tim anak laju bisa leha leha dat...	1250209679583506432	NaN	NaN
4	1250206199376678914	Tue Apr 14 23:36:00 +0000 2020	0	Bingung dengan strategi efektif kuliah online ...	1250206199376678914	https://pbs.twimg.com/media/EUeVShiUYAEwjn9.jpg	NaN

Figure 6: Crawling result data.
Source: Authors, (2026).

Based on the data that has been obtained previously, the data processing process was carried out using 8,411 Twitter crawling data, and after verification, the data used was 5,978 tweets. After the data was obtained, the next step was the labelling process. The labelling process was carried out from August 8, 2024 to September 20, 2024. Data labelling was carried out in three categories, namely positive, negative and neutral (Table 4).

Table 4: Label data Twitter.

Description/Category	Label
Positive	1
Negative	2
Neutral	3

Source: Authors, (2026).

From the 5,978 tweets used, each label was obtained, namely 1,953 positive labels, 2,045 negative labels and 1,980 neutral labels (Figure 7).

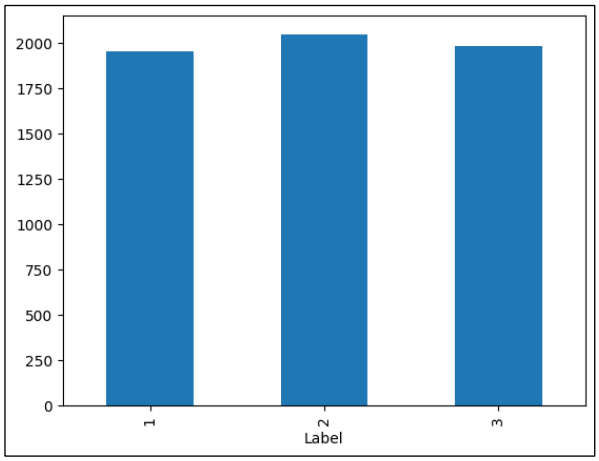

Figure 7: Label data Twitter.
Source: Authors, (2026).

IV.2 PRE-PROCESSING

After the data is collected, the data will then be processed through the pre-processing stage. This stage aims to make the data clean and there are no indicators that interfere with the score value during the implementation process of the Bag of Words weighting method and the Naive Bayes classification method. This stage is very important, because if there is unclean data, the score value will be affected. Pre-processing is carried out to prepare documents in raw text until they are ready to be analysed according to the method used. The purpose of this process is to increase classification accuracy [55].

	conversation_id_str	created_at	full_text	id_str	username	Label
0	1804659268114909952	"Sat Jun 22 23:48:51 +0000 2024"	"[8/8] Daaan...Info ini jangan berhenti di kam...	"1804662866483753112"	"UnivTerbuka"	1
1	1804659268114909952	"Sat Jun 22 23:48:13 +0000 2024"	"[7/8] Universitas Terbuka fleksibel Kuliahnya...	"1804662710090834395"	"UnivTerbuka"	1
2	1804659268114909952	"Sat Jun 22 23:46:10 +0000 2024"	"[6/8] Yuk segera bergabung di UT! Semua Bisa ...	"1804662193692250369"	"UnivTerbuka"	1
3	1804659268114909952	"Sat Jun 22 23:44:26 +0000 2024"	"[5/8] Hadiah: Rp. 600.000.000 (include Beasis...	"1804661757233012998"	"UnivTerbuka"	1
4	1804659268114909952	"Sat Jun 22 23:42:26 +0000 2024"	"[4/8] ï¸□ Jadwal Pertandingan: 10 - 11 Agust'...	"1804661255015436310"	"UnivTerbuka"	3

Figure 8: Crawling data Results Preprocessing.
Source: Authors, (2026).

To assist the results of data crawling, researchers try to visualize the data in the form of a word cloud (Figure 9). Word cloud is a form of visualization of text data that describes a collection of words that are often found in a text analysis. The size of the font in the word cloud indicates that the word appears frequently.

Figure 9: Data Visualization in WordCloud.
Source: Authors, (2026).

In addition to using word clouds, researchers also use graphs to visualize the results of words that appear frequently (Figure 10).

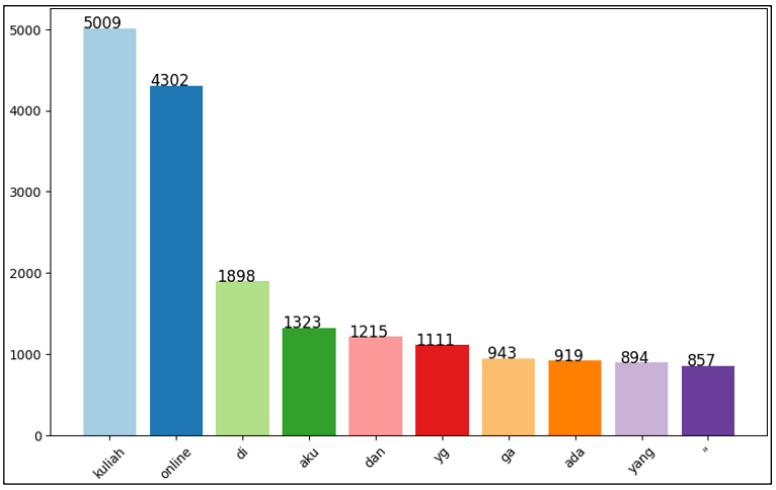

Figure 10: Frequently Appearing Word Graphs.
Source: Authors, (2026).

Before the data is processed and enters the first stage in pre-processing, the Case Folding process is carried out first. Case Folding is used to change all letters especially uppercase letters in a document to lowercase letters, only letters ‘a’ to ‘z’ are accepted. Characters other than letters are removed and considered delimeter. Uppercase or lowercase letters in a word are assumed to be no difference, all uppercase characters are usually changed to lowercase. This aims to make the text more consistent, thus allowing more consistent text processing.

```
Case Folding Result :

0    "[8/8] daaan...info ini jangan berhenti di kam...
1    "[7/8] universitas terbuka fleksibel kuliahnya...
2    "[6/8] yuk segera bergabung di ut! semua bisa ...
3    "[5/8] hadiah: rp. 600.000.000 (include beasis...
4    "[4/8] ï¸ jadwal pertandingan: 10 - 11 agust'...
Name: full_text, dtype: object
```

Figure 11: Case Folding Process.
Source: Authors, (2026).

IV.3 FEATURE EXTRACTION

Feature extraction is the process of converting terms from text into numbers that can be read by a computer. In this research, TF-IDF (Term Frequency-Inverse Document Frequency) is used as a process by giving a value to each word in the training data. The calculated value is given to determine how important a word is for a sentence. The TF-IDF value depends on how often the word appears in the document [56].

```
(0, 4529)     0.3651962248249808
(0, 6755)     0.0974972577683869
(0, 8994)     0.4993300228237758
(0, 9465)     0.10054239548236474
(0, 12533)    0.7731020353864696
(1, 206)      0.39333361187157334
(1, 2751)     0.36162190824250384
(1, 5124)     0.13073626338893005
(1, 6127)     0.2320014980259387
(1, 6755)     0.049603993768355545
(1, 9465)     0.10230655657640196
(1, 10700)    0.529238356284812
(1, 12198)    0.3699631773645535
(1, 12390)    0.36162190824250384
(1, 12837)    0.2879896126494258
(2, 198)      0.11089800034777621
(2, 560)      0.17927193875755346
(2, 1947)     0.3074011965668767
(2, 2035)     0.19327622228550825
(2, 2407)     0.24538027533298465
(2, 2557)     0.17291892158166827
(2, 4035)     0.10683369011216404
```

Figure 12: Results of the Feature Extraction Stage Process with TF-IDF.
Source: Authors, (2026).

Simply put, TF-IDF is used to determine how often/important a word appears in a document. The tweet data in the pre-processing stage is still in the form of text (words), while in the classification stage the data used must be numeric or numbers. So at this feature extraction stage, the data will first be converted into numbers using the TF-IDF method. The results of the TF-IDF process are as shown in Figure 12.

IV.4 CLASSIFICATION

The results obtained at the feature extraction stage will be input data at the classification stage. At this classification stage, three methods are used, namely Naïve Bayes, Decision Tree, and Support Vector Machine. Meanwhile, to measure the performance of the method used by researchers in terms of validating and evaluating using k-fold cross validation and confusion matrix, where k = 10 with a division of 80% training data and 20% testing data. This classification seeks the best level of accuracy from each method. So we can compare which level of accuracy is better from the three methods.

IV.4.1 Naïve Bayes

Classification using *Naïve Bayes* is carried out in this analysis using a ratio of 80:20 between training data and testing data. With validation and evaluation using k-fold cross validation, where the value of k = 10, the level of accuracy is obtained as in Figure 18. From these results, the average level of accuracy obtained is 78.02%, with the highest accuracy at k = 9, which is 80.91%. While the average precision value is 78.18%, recall is 78.02% and F1-Score is 77.45% (Figure 13).

```
Naive Bayes (K-Fold Cross Validation (K=10))
============================================
Accuracy          : [77.08768267 75.99164927 76.88284519 75.05230126 77.24895397 80.17782427
 79.44560669 80.17782427 80.91004184 77.24895397]
Mean Accuracy     : 78.02

Recall            : [77.08768267 75.99164927 76.88284519 75.05230126 77.24895397 80.17782427
 79.44560669 80.17782427 80.91004184 77.24895397]
Mean Recall       : 78.02

Precision          : [77.81964458 75.75784028 77.52005167 75.94284245 77.93378724 80.15934596
 79.63592612 79.83949059 80.74378957 76.40046899]
Mean Precision     : 78.18

F1-Score          : [76.32962836 75.73978488 76.44089662 74.14623553 76.76672629 79.72983041
 78.66874383 79.61048039 80.46512477 76.60495255]
Mean F1-Score     : 77.45
```

Figure 13: Classification Results using Naïve Bayes.
Source: Authors, (2026).

IV.4.2 Decision Tree

By using the same data division 80:20, and validation and evaluation using k-fold cross validation, where the value of k = 10, the accuracy level is obtained as in Figure 19. The average result at the accuracy level is 86%, with the highest accuracy obtained at k = 4 and k = 6, which is 100%. While the average precision value is 89.14%, recall is 81.24% and F1-Score is 82.31% (Figure 14).

```
Decision Tree (K-Fold Cross Validation (K=10))
==============================================
Accuracy          : [ 90.  80.  90. 100.  80. 100.  90.  70.  70.  90.]
Mean Accuracy     : 86.0

Recall            : [ 40.         100.          66.66666667 100.          60.
 100.          80.          80.         100.          85.71428571]
Mean Recall       : 81.24

Precision          : [ 91.66666667  92.          92.         100.          85.71428571
 100.          80.          70.83333333  86.66666667  92.5       ]
Mean Precision     : 89.14

F1-Score          : [ 67.03296703  90.1010101   90.1010101  100.          79.16666667
 100.          80.          69.6969697   56.66666667  90.32967033]
Mean F1-Score     : 82.31
```

Figure 14: Classification Results using Decision Tree.
Source: Authors, (2026).

IV.4.3 Support Vector Machine

By using Support Vector Machine, the accuracy level is obtained as shown in Figure 20. The average accuracy level obtained is 89%, with the highest accuracy at k = 6 and k = 10, which is 100%. While the average precision value is 90.97%, recall is 87.31% and F1-Score is 88.99% (Figure 15).

```
Support Vector Machine (K-Fold Cross Validation (K=10))
=======================================================
Accuracy          : [ 90.  70.  90.  80.  90. 100.  90.  90.  90. 100.]
Mean Accuracy     : 89.0

Recall            : [ 80.          75.          80.          71.42857143 100.
 100.         100.         100.          66.66666667 100.        ]
Mean Recall       : 87.31

Precision          : [ 91.66666667  72.          91.66666667  88.          91.42857143
 100.          91.66666667  92.          91.25       100.        ]
Mean Precision     : 90.97

F1-Score          : [ 89.8989899   70.3030303   89.8989899   80.83333333  89.67032967
 100.          89.8989899   90.1010101   89.33333333 100.        ]
Mean F1-Score     : 88.99
```

Figure 15: Classification Results using Support Vector Machine.
Source: Authors, (2026).

The overall results data of the three methods (Naïve Bayes, Decision Tree, and Support Vector Machine) can be seen in Table 5.

Table 5: Comparison of Naïve Bayes, Decision Tree, and SVM Results.

Methods / K-Fold	Naïve Bayes (NB)				Decision Tree (DT)				Support Vector Machine (SVM)			
	Accuracy	**Recall**	**Precision**	**F1-Score**	**Accuracy**	**Recall**	**Precision**	**F1-Score**	**Accuracy**	**Recall**	**Precision**	**F1-Score**
K=1	77,09 %	77,09 %	77,82 %	76,33 %	90 %	40 %	91,67 %	67,03 %	90 %	80 %	91,67 %	89,89 %
K=2	75,99 %	75,99 %	75,76 %	75,74 %	80 %	100 %	92 %	90,10 %	70 %	75 %	72 %	70,30 %
K=3	76,88 %	76,88 %	77,52 %	76,44 %	90 %	66,67 %	92 %	90,10 %	90 %	80 %	91,67 %	89,89 %
K=4	75,05 %	75,05 %	75,94 %	74,15 %	100 %	100 %	100 %	100 %	80 %	71,43 %	88 %	80,83 %
K=5	77,25 %	77,25 %	77,93 %	76,77 %	80 %	60 %	85,74 %	79,17 %	90 %	100 %	91,43 %	89,67 %
K=6	80,18 %	80,18 %	80,16 %	79,73 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
K=7	79,45 %	79,45 %	79,64 %	78,67 %	90 %	80 %	80 %	80 %	90 %	100 %	91,67 %	89,89 %
K=8	80,18 %	80,18 %	79,84 %	79,61 %	70 %	80 %	70,83 %	69,69 %	90 %	100 %	92 %	90,10 %
K=9	80,90 %	80,91 %	80,74 %	80,47 %	70 %	100 %	86,67 %	56,67 %	90 %	66,67 %	91,25 %	89,33 %
K=10	77,25 %	77,25 %	77,40 %	76,61 %	90 %	85, 71 %	92,50 %	90,33 %	100 %	100 %	100 %	100 %
Mean	**78,02 %**	**78,02 %**	**78,18 %**	**77,45 %**	**86 %**	**81,14 %**	**89,14 %**	**82,31 %**	**89 %**	**87,31 %**	**90,97 %**	**88,99 %**

Source: Authors, (2026).

Table 5 shows the average accuracy of using the Naïve Bayes method of 78.02%, Decision Tree of 86%, and Support Vector Machine of 89%. So from the data, the best accuracy level of the three methods is Support Vector Machine (89%). Thus, sentiment analysis using the algorithm shows significant results as evidenced by its accuracy level of 89% in the Support Vector Machine method.

V. CONCLUSIONS

Based on the final results of the research, it can be concluded that from the 8,411 data that have been obtained, 5,978 data are left to be used in the research. At the classification stage, the performance measurement of the method used by the researcher in terms of validating and evaluating using k-fold cross validation and confusion matrix, where k = 10 with a division of 80% training data and 20% testing data. The final results of using the method obtained accuracy for the Naïve Bayes method of 78.02%, Decision Tree of 86%, and Support Vector Machine of 89%. So in this case, it is seen from the level of accuracy that is superior, namely SVM. This means that the performance of the system's success in sentiment analysis of online lectures based on Twitter data is very good. Future research is expected to add machine learning methods in its classification such as the K-Nearest Neighbor method, Logistic Regression, Random Forest or Artificial Neural Network. This is expected to be able to find out more comprehensively regarding the use of other methods.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Darwan, Sirojudin Wahid, Fajar Hardoyono, Indra, and Muhammad Hikmal.
Methodology: Darwan, Sirojudin Wahid, Fajar Hardoyono, Indra, and Muhammad Hikmal.
Investigation: Darwan, Sirojudin Wahid, Fajar Hardoyono, Indra, and Muhammad Hikmal.
Discussion of results: Darwan, Sirojudin Wahid, Fajar Hardoyono, Indra, and Muhammad Hikmal.
Writing – Original Draft: Darwan, Sirojudin Wahid, Fajar Hardoyono, Indra, and Muhammad Hikmal.
Writing – Review and Editing: Darwan, Sirojudin Wahid, Fajar Hardoyono, Indra, and Muhammad Hikmal.
Resources: Darwan, Sirojudin Wahid, Fajar Hardoyono, Indra, and Muhammad Hikmal.
Supervision: Darwan, Sirojudin Wahid, Fajar Hardoyono, Indra, and Muhammad Hikmal.
Approval of the final text: Darwan, Sirojudin Wahid, Fajar Hardoyono, Indra, and Muhammad Hikmal.

VII. ACKNOWLEDGMENTS

Thanks to the Ministry of Religious Affairs of the Republic of Indonesia, the Rector of IAIN Syekh Nurjati Cirebon and LPPM IAIN Syekh Nurjati Cirebon for funding this research with contract number: 141/In.8/L.I/TL.01/05/2024. Gratitude is also extended to all those who have helped this research.

VIII. REFERENCES

[1] A. N. Bahasoan, Wulan Ayuandiani, Muhammad Mukhram, and Aswar Rahmat, "Effectiveness of Online Learning In Pandemic Covid-19," Int. J. Sci. Technol. Manag., vol. 1, no. 2, pp. 100–106, 2020, doi: 10.46729/ijstm.v1i2.30.

[2] K. Marifa, K. Komariah, S. Hamidah, and S. Sugiyono, "Hambatan Pembelajaran Daring Dalam Perkuliahan Pada Masa Pandemi Covid-19," Pendidik. Tek. Boga, 2021, [Online]. Available: https://journal.uny.ac.id/index.php/ptbb/article/view/44687

[3] N. R. Hanik and T. Wiharti, "TANGGAPAN DAN KENDALA MAHASISWA DALAM PEMBELAJARAN DARING SERTA HARAPANNYA PADA PEMBELAJARAN DAN PRAKTIKUM SELAMA PANDEMI COVID-19," J. Ilm. Profesi Pendidik., vol. 6, no. 4, pp. 609–616, 2021, doi: https://doi.org/10.29303/jipp.v6i4.238.

[4] Suci, W. L. Candra, Murtono, Suryani, and F. Budi, "Constraints in Implementing Online Learning during the Covid-19 Pandemic," J. Phys. Conf. Ser., vol. 1823, no. 1, 2021, doi: 10.1088/1742-6596/1823/1/012088.

[5] Kolivia, Nazaryo Radja, Kelik Wachyudi, and Acep Bahrum Kamil, "Hambatan-Hambatan Yang Dialami Mahasiswa Dalam Pembelajaran Daring Selama Pandemi Covid-19," J. Santiaji Pendidik., vol. 12, no. 2, pp. 80–84, 2022, doi: 10.36733/jsp.v12i2.3634.

[6] N. Nasrah, R. Y. Purwati, and A. M. Nur, "Analisis Perkuliahan Daring Konsep Dasar IPA: Study Kasus," Sainsmat J. Ilm. Ilmu Pengetah. Alam, vol. 12, no. 2, p. 170, 2023, doi: 10.35580/sainsmat122440882023.

[7] R. Henrikson and N. Baliram, "Examining voice and choice in online learning," Int. J. Educ. Technol. High. Educ., vol. 20, no. 1, 2023, doi: 10.1186/s41239-023-00401-w.

[8] B. Abdellaoui et al., "Analyzing emotions in online classes: Unveiling insights through topic modeling, statistical analysis, and random walk techniques," Int. J. Cogn. Comput. Eng., vol. 5, no. May, pp. 221–236, 2024, doi: 10.1016/j.ijcce.2024.05.003.

[9] M. Al-Hail, M. F. Zguir, and M. Koç, "University students' and educators' perceptions on the use of digital and social media platforms: A sentiment analysis and a multi-country review," iScience, vol. 26, no. 8, 2023, doi: 10.1016/j.isci.2023.107322.

[10] G. Anggadwita, N. Indarti, and V. Ratten, "Changes in Indonesian private universities educational practices in the post COVID-19 environment," Int. J. Manag. Educ., vol. 22, no. 1, p. 100905, 2024, doi: 10.1016/j.ijme.2023.100905.

[11] O. Estrada-Molina, J. Mena, and A. López-Padrón, "The Use of Deep Learning in Open Learning: A Systematic Review (2019 to 2023)," Int. Rev. Res. Open Distrib. Learn., vol. 25, no. 3, pp. 370–393, 2024, doi: 10.19173/irrodl.v25i3.7756.

[12] S. R. A. Abdullah and A. Al-Azawei, "Predicting Online Learners' Performance Through Ontologies: A Systematic Literature Review," Int. Rev. Res. Open Distrib. Learn., vol. 26, no. 1, pp. 16–37, 2025, doi: https://doi.org/10.19173/irrodl.v26i1.7921.

[13] APJII, "APJII.pdf." 2024. [Online]. Available: https://survei.apjii.or.id/

[14] R. Rafif, E. B. Setiawan, and I. Kurniawan, "Analisis dan implemenasi algoritma C4.5 dan pembobotan TF-IDF untuk menentukan trending topik pada media sosial twitter," in e-Proceeding of Engineering, Bandung, Indonesia: Telkom University, 2020, pp. 7661–7672. [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/12778/12501

[15] B. Mostefai, T. Boutefara, A. Chahinez, and A. Marwa, "Machine Learning Techniques for Identifying Textual Propaganda on Social Media: Development of a Detecting Digital Manipulation System," J. Eng. Technol. Ind. Appl., vol. 11, no. 53, pp. 43–49, 2025, doi: 10.5935/jetia.v11i53.1443.

[16] R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," J. Inform., vol. 8, no. 2, pp. 111–122, 2021, doi: 10.31294/ji.v8i2.10494.

[17] C. A. S. Sugianto, "Penerapan Teknik Data Mining Untuk Menentukan Hasil Seleksi Masuk Sman 1 Gibeber Untuk Siswa Baru Menggunakan Decision Tree," J. TEDC, pp. 49–54, 2015, [Online]. Available: http://ejournal.poltektedc.ac.id/index.php/tedc/article/download/240/185

[18] E. Priyanti, "Penerapan Decision Tree Untuk Klasifikasi Tingkat Pendapatan," IJCIT (Indonesian J. Comput. Inf. Technol., vol. 7, no. 1, pp. 7–12, 2022, doi: 10.31294/ijcit.v7i1.12506.

[19] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," Jambura J. Electr. Electron. Eng., vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjeee.v5i1.16830.

[20] B. Valarmathi, N. S. Gupta, V. Karthick, T. Chellatamilan, K. Santhi, and D. Chalicheemala, "Sentiment Analysis of Covid-19 Twitter Data using Deep Learning Algorithm," Procedia Comput. Sci., vol. 235, no. 2023, pp. 3397–3407, 2024, doi: 10.1016/j.procs.2024.04.320.

[21] N. O. Adelakun and A. A. Baale, "Sentiment Analysis of Financial News Using The BERT Model," J. Eng. Technol. Ind. Appl., vol. 10, pp. 21–27, 2024, [Online]. Available: https://media.neliti.com/media/publications/427128-infrastructural-development-in-yewa-sout-0d9e6e62.pdf

[22] S. Namitha, P. Sanjan, N. C. Reddy, Y. Srikar, H. Shanmugasundaram, and B. P. Andraju, "Sentiment Analysis: Current State and Future Research Perspectives," Proc. 7th Int. Conf. Intell. Comput. Control Syst. ICICCS 2023, pp. 1115–1119, 2023, doi: 10.1109/ICICCS56967.2023.10142318.

[23] V. A. Pulikonda, D. Vanukuru, M. S. Navali, B. Motepalli, and K. Swathi, "Exploring the Applications, Challenges, and Issues of Sentiment Analysis," Proc. - 7th Int. Conf. Comput. Methodol. Commun. ICCMC 2023, no. Iccmc, pp. 807–812, 2023, doi: 10.1109/ICCMC56507.2023.10083940.

[24] K. Purwandari, R. B. Perdana, J. W. C. Sigalingging, R. Rahutomo, and B. Pardamean, "Automatic Smart Crawling on Twitter for Weather Information in Indonesia," Procedia Comput. Sci., vol. 227, pp. 795–804, 2023, doi: 10.1016/j.procs.2023.10.585.

[25] H. W. K. Sajinika, S. Vasanthapriyan, and P. Wijeratne, "Twitter Sentiment Analysis and Topic Modeling for Online Learning," ICARC 2023 - 3rd Int. Conf. Adv. Res. Comput. Digit. Transform. Sustain. Dev., pp. 262–267, 2023, doi: 10.1109/ICARC57651.2023.10145609.

[26] S. E. Alqaan and A. M. Qamar, "Utilizing Sentiment Analysis to Enhance the Quality of Online Learning," Proc. 2022 5th Natl. Conf. Saudi Comput. Coll. NCCC 2022, pp. 41–46, 2022, doi: 10.1109/NCCC57165.2022.10067560.

[27] K. I. Senadhira, R. A. H. M. Rupasingha, and B. T. G. S. Kumara, "Sentiment Analysis on Twitter Data Related to Online Learning During the Covid-19 Pandemic," Proc. - Int. Res. Conf. Smart Comput. Syst. Eng. SCSE 2022, pp. 131–136, 2022, doi: 10.1109/SCSE56529.2022.9905190.

[28] T. V. Ngoc, M. N. Thi, and H. N. Thi, "Sentiment Analysis of Students' Reviews on Online Courses: A Transfer Learning Method," Proc. Int. Conf. Ind. Eng. Oper. Manag., pp. 306–314, 2021, doi: 10.46254/ap01.20210122.

[29] Z. Kastrati, A. S. Imran, and A. Kurti, "Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs," IEEE Access, vol. 8, no. June, pp. 106799–106810, 2020, doi: 10.1109/ACCESS.2020.3000739.

[30] R. W. Hardian, P. E. Prasetyo, U. Khaira, and T. Suratno, "Analisis Sentiment Kuliah Daring Di Media Sosial Twitter Selama Pandemi Covid-19 Menggunakan Algoritma Sentistrength," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 1, no. 2, pp. 138–143, 2021, doi: 10.57152/malcom.v1i2.15.

[31] L. M. Nkomo and B. K. Daniel, "Sentiment Analysis of Student Engagement with Lecture Recording," TechTrends, vol. 65, no. 2, pp. 213–224, 2021, doi: 10.1007/s11528-020-00563-8.

[32] X. Pu, G. Yan, C. Yu, X. Mi, and C. Yu, "Sentiment analysis of online course evaluation based on a new ensemble deep learning mode: Evidence from chinese," Appl. Sci., vol. 11, no. 23, 2021, doi: 10.3390/app112311313.

[33] C. Kurniawan and F. Wahyuni, "Sentiment analysis of online learning students feedback for facing new semester: A support vector machine approach," Proc. - 2021 7th Int. Conf. Educ. Technol. ICET 2021, no. September, pp. 1–6, 2021, doi: 10.1109/ICET53279.2021.9575116.

[34] O. Abiodun Ayeni, A. Mercy, T. A.F, and M. A.S, "Web-Based Student Opinion Mining System Using Sentiment Analysis," Int. J. Inf. Eng. Electron. Bus., vol. 12, no. 5, pp. 33–46, 2020, doi: 10.5815/ijieeb.2020.05.04.

[35] U. Ö. Osmanoglu, O. N. ATAK, K. ÇAĞLAR, H. KAYHAN, and T. CAN, "Sentiment Analysis for Distance Education Course Materials: A Machine Learning Approach," J. Educ. Technol. Online Learn., vol. 3, no. 1, pp. 31–48, 2020, doi: 10.31681/jetol.663733.

[36] M. Muljono, D. P. Artanti, A. Syukur, A. Prihandono, and D. R. I. M. Setiadi, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," in Konferensi Nasional Sistem Informasi (KNSI), Pangkalpinang, Indonesia, 2018, pp. 165–170. [Online]. Available: https://jurnal.atmaluhur.ac.id/index.php/knsi2018/article/view/353

[37] K. H. Abdulkareem et al., "Realizing an Effective COVID-19 Diagnosis System Based on Machine Learning and IoT in Smart Hospital Environment," IEEE Internet Things J., vol. 8, no. 21, pp. 15919–15928, 2021, doi: 10.1109/JIOT.2021.3050775.

[38] R. R. Salam, M. F. Jamil, Y. Ibrahim, R. Rahmaddeni, S. Soni, and H. Herianto, "Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 3, no. 1, pp. 27–35, 2023, doi: 10.57152/malcom.v3i1.590.

[39] J. Beel, S. Langer, and B. Gipp, "TF-IDuF: A Novel Term-Weighting Sheme for User Modeling based on Users' Personal Document Collections," in Proceedings of the iConference 2017, iSchools, 2017, pp. 1–7. [Online]. Available: https://www.ideals.illinois.edu/items/101808

[40] D. Setyawan and E. Winarko, "Analisis Opini Terhadap Fitur Smartphone Pada Ulasan Website Berbahasa Indonesia," IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 10, no. 2, p. 183, 2016, doi: 10.22146/ijccs.17485.

[41] B. Liu, E. Blasch, Y. Chen, D. Shen, and G. Chen, "Scalable sentiment classification for Big Data analysis using Naïve Bayes Classifier," Proc. - 2013 IEEE Int. Conf. Big Data, Big Data 2013, no. December, pp. 99–104, 2013, doi: 10.1109/BigData.2013.6691740.

[42] M. Ibrahim, E. Bu'ulolo, and I. Lubis, "Penerapan Algoritma Naive Bayes Classifier Untuk Mendeteksi Tingkat Krediblitas Hoax News/ Fake News Pada Sosial Media Di Indonesia Berbasis Android (Studi Kasus : Kantor Tribun Medan)," RESOLUSI Rekayasa Tek. Inform. dan Inf., vol. 1, no. 1, pp. 9–17, 2020, doi: https://doi.org/10.30865/resolusi.v1i1.8.

[43] L. Rokach and O. Maimon, DATA MINING WITH DECISION TREES Theory and Applications 2nd Edition. World Scientific, 2014. doi: https://doi.org/10.1142/9097.

[44] J. Han, M. Kamber, and J. Pei, Data Mining Concepts and Techniques, Third. Morgan Kaufmann Publishers, 2012.

[45] E. A. Guna et al., "Implementasi Algoritma Decision Tree untuk Klasifikasi Data Evaluation Car Menggunakan Python," J. Sist. Inf. dan Ilmu Komput., vol. 1, no. 4, pp. 167–177, 2023, [Online]. Available: https://doi.org/10.59581/jusiik-widyakarya.v1i4.

[46] S. K. Hasna, "Analisis Sentimen Data Ulasan Menggunakan Algoritma Support Vector Machine (Studi Kasus: Aplikasi Iflix)," Universitas Islam Indonesia, Yogyakarta, 2021. [Online]. Available: https://dspace.uii.ac.id/bitstream/handle/123456789/34393/17522136 Salwa Kamilia Hasna.pdf?sequence=1&isAllowed=y

[47] Y. Al-Amrani, M. Lazaar, and K. E. El Kadiri, "Sentiment analysis using hybrid method of support vector machine and decision tree," J. Theor. Appl. Inf. Technol., vol. 96, no. 7, pp. 1886–1895, 2018, [Online]. Available: https://www.jatit.org/volumes/Vol96No7/13Vol96No7.pdf

[48] A. Kowalczyk, Support Vector Machines (Succintctly). Syncfusion, 2017. [Online]. Available: https://www.syncfusion.com/succinctly-free-ebooks/support-vector-machines-succinctly

[49] J. W. G. Putra, Pengenalan Pembelajaran Mesin dan Deep Learning, 1.4. Tokyo, 2020. [Online]. Available: https://wiragotama.github.io/

[50] I. C. R. Drajana, "METODE SUPPORT VECTOR MACHINE DAN FORWARD SELECTION PREDIKSI PEMBAYARAN PEMBELIAN BAHAN BAKU KOPRA," Ilk. J. Ilm., vol. 9, no. 2, pp. 116–123, 2017, doi: https://doi.org/10.33096/ilkom.v9i2.134.116-123.

[51] A. S. Ritonga and E. S. Purwaningsih, "Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan SMAW (Shield Metal Arc Welding)," J. Ilm. Edutic Pendidik. dan Inform., vol. 5, no. 1, pp. 17–25, 2018, doi: https://doi.org/10.21107/edutic.v5i1.4382.

[52] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," J. Media Inform. Budidarma, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.

[53] G. Dong and J. Bailey, CONTRAST DATA MINING : Concepts, Algorithms, and Applications. CRC Press - Taylor & Francis Group, 2013.

[54] D. P. Deandra, "ANALISIS SENTIMEN KEPERCAYAAN MASYARAKAT TERHADAP INSTANSI POLRI DENGAN ALGORITMA SUPPORT VECTOR MACHINE PADA MEDIA SOSIAL TWITTER," Universitas Multimedia Nusantara, Tangerang, 2023. [Online]. Available: https://kc.umn.ac.id/id/eprint/25732

[55] N. M. A. J. Astari, Dewa Gede Hendra Divayana, and Gede Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," J. Sist. dan Inform., vol. 15, no. 1, pp. 27–29, 2020, doi: 10.30864/jsi.v15i1.332.

[56] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 3, no. 3, pp. 2789–2797, 2019, doi: 10.1061/9780784415313.ch06.