

# Sentiment Analysis Od Face To Face School Policy On Twitter Social Media With Support Vector Machine(SVM)

Amir Mahmud Husein<sup>1\*</sup>, Berninto Sipahutar<sup>2</sup>, Ramonda Jumpatua Dashuah<sup>3</sup>, Eben Ezer Hutaeruk<sup>4</sup>  
<sup>1,2,3,4</sup> Universitas Prima Indonesia, Sumatera Utara, Indonesia  
[amirmahmud@unprimdn.ac.id](mailto:amirmahmud@unprimdn.ac.id)

**Submitted :** Nov 28, 2022 | **Accepted :** Dec 16, 2022 | **Published :** Jan 1, 2023

**Abstract:** Twitter social media is one way to get fast information, especially related to face-to-face learning system where during covid-19 pandemic learning is held online. In this case government has informed related to the face-to-face learning system as well as the community or students gave an enthusiastic response to the policies provided by the government including giving a good response to these policies and some of them disagreeing with these policies. In this case, the researcher analyzes public opinion on government policies related to face-to-face learning on Twitter social media using the Support Vector Machine algorithm. By doing an analysis related to government policies regarding learning during the COVID-19 pandemic, the government can find out how the public responds and can make decisions. Based on a series of processes that have been carried out previously using the Support Vector Machine method by applying the TF-IDF weighting function, the results can reach 93%. To see the level of accuracy of the proposed method, the researchers made a comparison by applying several other methods. The accuracy results obtained from the support vector machine method are 93%, based on the accuracy obtained, it can be determined that the level of accuracy using the Support Vector Machine method is quite high in classifying sentiment data, but when compared to other methods, namely nave Bayes, which obtains an accuracy of 94%, Logistic Regression which obtained 93% accuracy, and K-NN which obtained 90% accuracy. Thus, the accuracy results of four methods are not too different.

**Keywords:** Sentiment Analysis, Machine learning, Twitter Social Media, Support Vector Machine, Face-to-face School.

## 1. INTRODUCTION

Information Technology is a technology that has various uses in processing raw data, including processing, obtaining, compiling, storing results, and manipulating data in different way(Cholik, 2021). Social media is one of the very rapid developments of information technology because with social media information and communication can be conveyed quickly, there are several kinds of social media including Facebook, Instagram, YouTube, Twitter, etc. Twitter is one of the social media that is quite commonly used to convey aspirations and information and can be a means of communicating and expressing oneself through uploaded work(Nurhadi, 2017). Twitter social media is one way to get fast information, especially related to the face-to-face learning system where during the COVID-19 pandemic learning is held online. In this case the government has informed related to the face-to-face learning system as well as the community or students gave an enthusiastic response to the policies provided by the government including giving a good response to these policies and some of them disagreeing with these policies. Using data collection from Twitter and then analyzing public opinion on government policies based on data that is processed by analyzing sentiment so that all opinions can be grouped into 2 classes, namely negative and positive(Fitriana, Utami & Fatta, 2021). Supports Vector Machine can perform text classification quickly and effectively, SVM is also able to analyze data and perform pattern recognition(Irfani, 2020). This classification was chosen during training as a unique hyper plane that separates known positive instances from negative instances(A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, 2021), in SVM classification it has an important advantage in its theoretical approach which justifies the over fitting problem which allows it to work well(Pratama, Wihandika & Ratnawati, 2018).

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sentiment analysis is method for classifying a person's emotional level as positive or negative. One of the sentiment analysis applications applied to Twitter by utilizing a program system that will automatically retrieve data through posts or tweets from users and the program system will also perform a classification to assess the overall post whether it contains positive or negative opinion (Fitriana, Utami & Fatta, 2021). After clarification on the data obtained results, we can know and assess certain topics and can make decisions. The Support Vector Machine (SVM) algorithm is a commonly used method for solving classifications. The SVM concept stems from a two-class classification problem that requires positive and negative training dataset (Abbas, 2016). The Support Vector Machine (SVM) method is an algorithm that is often used in classifying data (Lusiana, 2018). Therefore, the researcher conducted a study on "Analysis of Sentiment Against Face-to-Face School Policies on Twitter Social Media with Support Vector Machine (SVM)".

## 2. LITERATURE REVIEW

According to Lestari Lusiana in research (Lusiana, 2018) entitled Twitter Sentiment Analysis by using the Support Vector Machine (SVM) Algorithm. To do SVM algorithm testing to find various public opinions on issues which is widely discussed on Twitter. This research aims to analyzing sentiment on tweets data. According to Kirelli et al. in research (Kirelli & Arslankaya, 2020) entitled Sentiment Analysis of Shared Tweets on Global Warming on Twitter with Data Mining Methods: A Case Study on Turkish Language. In this study analyze the Turkish text yet much researched in the literature. Sentiment analysis on global warming and climate change determined by machine learning methods. According to H. Setiawan et al. in research (Setiawan, Utami, & Sudarmawan, 2021) entitled Sentiment Analysis Twitter Post-Covid-19 Online Lecture Using Support Vector Algorithm Machine and Naive Bayes. Using the SVM Algorithm where in the algorithm weighting process is carried out by establishing a line pattern for carried out weighting and classification. While using the Naïve Algorithm Bayes where in this algorithm the system will classify through the probabilities of the possible from the data obtained. This research will try conduct sentiment analysis on social media twitter to measure opinion and public opinion regarding post-covid-19 online lectures.

## 3. METHOD

The research using quantitative research. In this study, research analyzed data contained in social media related to government policies regarding face-to-face learning using a support vector machine (SVM). SVM has a high degree of accuracy in the field of deep learning artificial neural networks and will be applied in this study where the dataset used for testing is sourced from the netlityc.org website. The dataset is obtained from the results of data crawling through the website <https://netlityc.org>, the next step is to analyze the dataset using the support vector machine (SVM) method and then present the accuracy of the classification results of data analysis.

### a. Work Procedures

Netlytic is cloud-based text analyzer and social network that can automatically summarize textual data and find communication networks from publicly accessible social media posts (Kirelli & Arslankaya, 2020).

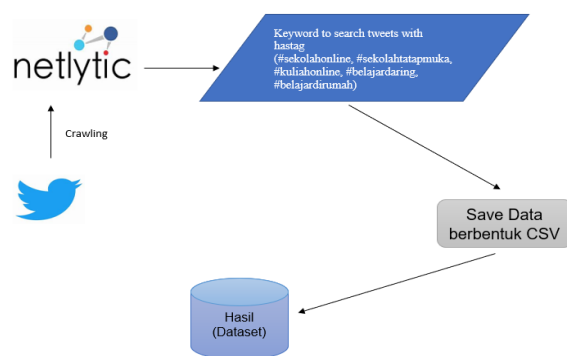


Figure 1. The process of retrieving datasets from <https://netlityc.org>

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

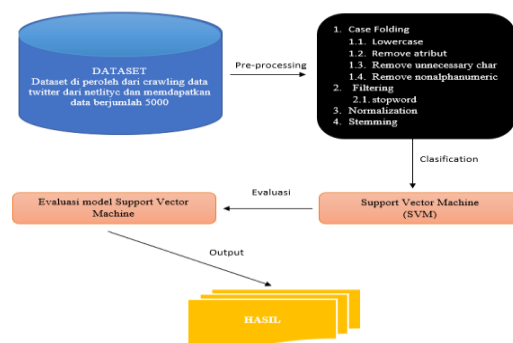


Figure 2. Process of data classification and evaluation

## b. Dataset

The dataset is a collection of data taken from Twitter data crawling through the <https://netlityc.org> website using the hashtag online school, face-to-face schools, online lectures, online learning, learning at home with a total of 5035 text tweets.

## c. Work Procedures

The work procedure in this study, are: Dataset is a collection of data taken from Twitter data crawling through the <https://netlityc.org> website using the hashtag online school, face-to-face schools, online lectures, online learning, learning at home with a total of 5035 text tweets; Tweets usually have no formal writing standards and therefore each text is refined by applying steps to create a purer model. The goal of data processing is to produce more acceptable data by reducing the feature size; SVM is a supervised learning method and is generally used for the classification process (Setiawan, Utami, & Sudarmawan, 2021). Usually used to help solve classification fixation problems. SVM is useful for separating data sets into two groups in the most appropriate way and these data groups are then referred to as hyperplanes; At this stage the relevant data is identified to measure whether the results obtained are in accordance with the objectives to be achieved.

## 4. RESULT

The final research result is obtained by carrying out several processes including data collection, cleaning, pre-processing, and classifying the model using the Support Vector Machine to find the accuracy of the model used by the researcher to be compared by applying several models, are Naïve Bayes Classifier, logistic regression, KNN.

## a. Data Preparing

In collecting dataset, researchs crawled Twitter data from the website <https://netlytic.orgwhere> researchers look for data with predetermined hashtags.

Table 1. Dataset

Author	Pubdate	Description	Tweet_type
Asterachest	4/6/2022 4:56	@Mahasiswa UMS daring ap luring nih? aku ada nih kalo mau	reply
xtJaemin	4/6/2022 2:57	@jupfter daring yaa? gue luring ?Ã´Ã± Ã±	reply
Manisangulajawa	4/6/2022 2:22	@neverseeu Bucin e ae luring, kuliah e tetap daring	reply
Sweetaswinee	4/6/2022 2:18	ky kesamber gledek siang2 dosen tbth ngasi pilihan luring atau daring	original
.....	.....	.....	.....

## b. Testing

In the data testing phase, the researcher applies the TF-IDF weighting function to calculate how important words are in a document. Then, perform the data classification process with the support vector machine method to get the results of data analysis based on accuracy.

1. Term Frequency Inverse Document Frequency weighting.

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Steps to perform TF-IDF weighting are: Determine the keywords to use; Determining the value of TF can use binary TF, that is, if a word is contained in a document it will be assigned a value of 1 (one), otherwise it will be assigned a value of 0 (zero) Fitriana, Utami & Fatta, 2021). Find the value of df by counting the number of documents that contain a certain word. Calculates the Inverse Document Frequency (IDF) value which is useful for knowing how important words are in a document(Netlix, 2022). IDF is formulated as follows:

$$IDF = \log D / df$$

Note : Number of document containing term

D = Total number of documents

After that, calculate the IDF TF by combining the TF with IDF calculations as follows.

$$W_{ij} = tf \times IDF$$

$$W_{ij} = tf \times \log D / df$$

Description:

$W_{ij}$  = the weight of words in each document

Tf = the number of occurrences of a word in a document.

After that add up each keyword weight in each document.

## 2. Support Vector Machine (SVM)

At this stage, a SVM model is created and classification is carried out to determine the accuracy results obtained (R. Sulaeman, 2022). Then, the accuracy results obtained are compared with several models including Naïve Bayes Classifier, logistic regression, and K-NN. The classification model that has been formed can already be used for predict the label of the test data.

After the TF-IDF weighting process is complete, then tested the data into the classification model. The data used in the classification model is only the response column part without the label column because the label part will be predicted by the classification model. Based on the classification process carried out, the accuracy of the support vector machine method can be generated and make comparisons with several methods as a measure of the level of accuracy and success of the proposed method.

Below is accuracy the classification model of the SVM method and its comparison with the Naïve Bayes Classifier, Logistic Regression, K-NN method as follows:

### 1. Naive Bayes

**Table 2.** Classification of the Naïve Bayes Model

Model	Sentiment	precision	Recall	f1-score	Support
Naive Bayes	Positive	0,9615	0,9202	0,9404	163
	Negative	0,9316	0,9672	0,9491	183
	Accuracy			0,9672	346
	Macro avg	0,9466	0,9437	0,9448	346
	Weighted avg	0,9457	0,9451	0,9450	346

### 2. Logistics Regression

**Table 3.** Classification of Logistic Regression Models

Model	Sentiment	precision	Recall	f1-score	Support
Logistics Regression	Positive	0,9167	0,9448	0,9305	163
	Negative	0,9494	0,9235	0,9363	183
	Accuracy	-	-	0,9335	346
	Macro avg	0,9331	0,9341	0,9334	346
	Weighted avg	0,9340	0,9335	0,9336	346

### 3. K-Nearest Neighbors (K-NN)

**Table 4.** K-Nearest Neighbors (K-NN)

Model	Sentiment	precision	Recall	f1-score	Support
K-Nearest Neighbors (K-NN)	Positive	0,9062	0,8896	0,8978	163
	Negative	0,9032	0,9180	0,9106	183
	Accuracy	-	-	0,9046	346
	Macro avg	0,9047	0,9038	0,9042	346
	Weighted avg	0,9047	0,9046	0,9046	346

### 4. Support Vector Machine(SVM)

**Table 4.** Support Vector Machine(SVM)

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Model	Sentiment	precision	Recall	f1-score	Support
Support Vector Machine(SVM)	Positive	0,9029	0,9693	0,9349	163
	Negative	0,9708	0,9071	0,9379	183
	Accuracy	-	-	0,9364	346
	Macro avg	0,9368	0,9382	0,9364	346
	Weighted avg	0,9388	0,9346	0,9365	346

### 5. Confusion Matrix

Confusion matrix is a table that research is often used to measure how models and classifications perform in machine learning. This table describes in more detail the amount of data classified correctly or incorrectly (Zalyhaty, 2021). In this study, a confusion matrix was applied to calculate the performance results and the level of accuracy that must be successful from the proposed classification process. With the application of the confusion matrix, researchers can display and compare methods and get actual values and model prediction values that can be used to generate evaluation metrics.

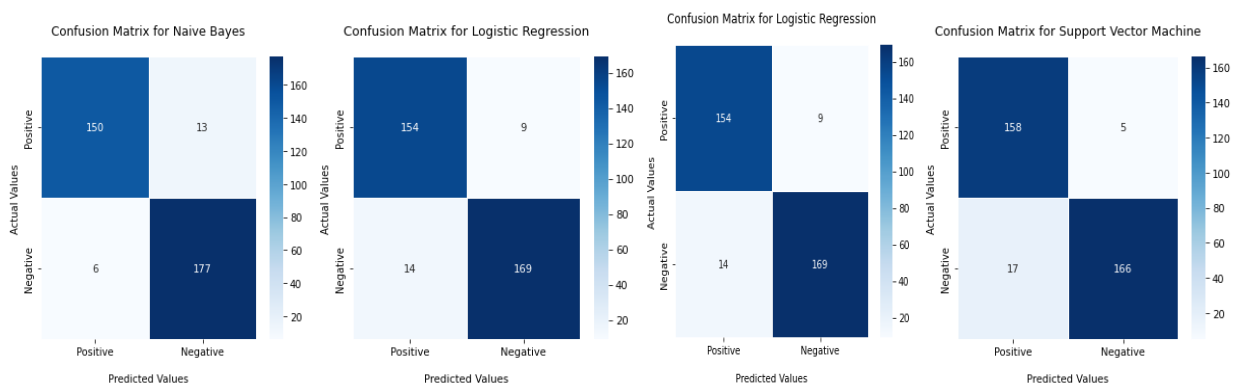


Figure 8. Confusion Matrix

To get the accuracy value by calculating the number of positive data obtained and predicted to be positive and the negative data obtained and predicted to be negative divided by the total number of overall data in the dataset. As for the explanation of each formula regarding Accuracy (Fitriana & Sibaroni, 2020) above as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

TP = The amount of data that has a positive prediction and has a positive value.

TN = The amount of data that is predicted to be negative and has a positive value.

FP = The amount of data that is predicted to be positive but has a negative value.

FN = The amount of data that is predicted to be negative but has a positive value (Naf'an, Burhanuddin, & Riyani, 2019).

## 5. DISCUSSIONS

In research (Kelvin et al, 2022) compared the performance of Logistic Regression and Support methods Vector Machine (SVM) using the TF-IDF research weighting technique this is the superior accuracy of the Support Vector Machine (SVM) method compared to the Logistic Regression method in classifying data Twitter sentiment about the Tweet of the CoronaVirus Disease-2019 (Covid-19) case with accuracy value of 91.15%. Based on the results of (Arsi & Waluyo, 2021) the testing process that has been carried out on tweets the sentiment of moving the capital city from social media twitter as many as 1,236 tweets (404 positive and 832 negative) using SVM, the results of the initial hypothesis match the results finally, SVM is better than the previous method (BM25 + KNN and Naïve Bayes). With accuracy = 96.68%, precision = 95.82%, recall = 94.04% and AUC = 0.979. In research (Styawati et al., 2021) also obtained results from the classification using the method Support Vector Machine which is divided into three neutral classes as much as 98.34%, class negative as much as 0.99%, and positive class as much as 0.66%. In contrast to research this uses the division of two classes, namely positive and negative so that the data negative and positive obtained more with a high degree of accuracy is 93%.

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



## 6. CONCLUSION

Support vector machine (SVM) algorithm method proposed the study is to classify data from public sentiment towards policies towards face-to-face schools. The testing phase is carried out using a dataset obtained from the results of data crawling through the <https://netlytic.org> website. Then the process dataset by selecting the attributes used then performs pre-processing, TF-IDF weighting and classification with the proposed method.

Accuracy results obtained from Support Vector Machine method are 93%, based on the accuracy obtained, It is concluded that the support vector machine method is quite high in classifying sentiment data, but when compared to other methods, namely nave Bayes, which obtains an accuracy of 94%, Logistic Regression which obtained 93% accuracy, and K-NN which obtained 90% accuracy. Thus, the accuracy results of the four methods are not too different, so they can be considered enough to be used in conducting sentiment analysis on face-to-face school policies on Twitter social media with vector machine support.

## 7. REFERENCES

- Abbas, I (2016). Penerapan Metode Moving Average (MA) Berbasis Algoritma Support Vector Machine (SVM) untuk Membandingkan Pola Kurva dengan Trend Kurva pada Trading Forex Online. *Ilk. J. Ilm.* 8(1):37–43. doi: 10.33096/ilkom.v8i1.20.37-43.
- Arsi. P & Waluyo. R (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *J. Teknol. Inf. dan Ilmu Komput.* 8(1):147, 2021. doi: 10.25126/jtiik.0813944.
- Cholik, C. A (2021). Perkembangan Teknologi Informasi Komunikasi / ICT Dalam Berbagai Bidang. *Jurnal Fakultas Teknik UNISA Kuningan.* 2(2)
- Fitriana, D. N & Sibaroni, Y (2020). Klasifikasi Data Tweet dengan Menggunakan Metode Klasifikasi Multi-Class Support Vector Machine (SVM) (Studi Kasus : PT.KAI). *e-Proceeding Eng.* 7(2): 8493–8505 [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/12746>
- Fitriana, F., Utami, E & Fatta, H. A (2021). Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes,” *J. Komtika (Komputasi dan Inform.* 5(1): 19–25. doi: 10.31603/komtika.v5i1.5185.
- Irfani, F. F., Triyanto M, Hartanto, A. D & Kusnawi (2020). Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine,” *Jurnal Bisnis, Manajemen dan Informatika.* 16(3): 258–266
- Isnain, A. R., Sakti, A. I., Alita D. & Marga, N. S (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm,” *Jdmsi.* 2(1)1: 31–37 [Online]. Available: <https://t.co/NfhmfMjtXw>
- Kelvin, Banjarnahor, J., Indra, E., Nababan, M. N. K & Sinurat S.H (2022). Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM),” *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA).* 5(2): 47–52. doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2365
- Kirelli, Y & Arslankaya, S (2020). Sentiment Analysis of Shared Tweets on Global Warming on Twitter with Data Mining Methods: A Case Study on Turkish Language,” *Comput. Intell. Neurosci.,* vol. 2020. doi: 10.1155/2020/1904172.
- Lusiana, L (2018). Analisis Sentimen Twitter Dengan Menggunakan Algoritma Support Vector Machine (SVM) (Studi Kasus: 3556 Data Tweets dengan Kata Kunci Cadar dan Hijab).
- Naf'an, M. Z., Burhanuddin, A. & Riyani, A (2019). Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen,” *J. Linguist. Komputasional.* 2(1): 23–27
- Netlytic - social media text and social networks analyzer. <https://netlytic.org/> (accessed Jul. 28, 2022).
- Nurhadi, Z. F (2017). Model Komunikasi Sosial Remaja Melalui Media Twitter. *Jurnal ASPIKOM.* 3(3): 539–549
- Pratama, A., Wihandika, R. C & Ratnawati, D. E (2018). Implementasi algoritme support vector machine (SVM) untuk prediksi ketepatan waktu kelulusan mahasiswa. *J. Pengemb. Teknol. Inf. dan Ilmu Komput.* 2(4):1704–1708
- Setiawan, H., Utami, E & Sudarmawan, S (2021). Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan 19 Naive Bayes. *J. Komtika (Komputasi dan Inform.* 5(1): 43–51. doi: 10.31603/komtika.v5i1.5189.
- Styawati., Hendrastuty, N., Isnain, A. R & Rahmadhani, A. Y (2021). Analisis 20 Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT.* 6(3):150–155. [Online]. Available: <http://situs.com>

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- 
- Sulaeman, R(2022). Analisis Algoritma Support Vector Machine Dalam Klasifikasi Penyakit Stroke Support Vector Machine Algorithm Analysis In Stroke Disease Classification. e-Proceeding of Engineering. 9(3): 922–928
- Zalyhaty, L. Q (2021). Analisis Sentimen Tanggapan Masyarakat Terhadap Vaksin Covid-19 Menggunakan Algoritma Support Vector Machine (SVM). Repositori Universitas Dinamika

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.