# Implementation of Text-Mining For Sentiment Analysis Twitter With Support Vector Machine Algorithm

Indrico Jowensen Jasumin [1*], Aditiya Hermawan[2]

[1,2] Department of Informatics Engineering, Buddhi Dharma University
Indrico.jowensen@gmail.com

**Abstract**: Text Mining can make it possible to analyze sentiments quickly. In this implementation, the process of capturing text is done with the help of the API. Twitter is used as an object of research because Twitter provides a medium for the process of retrieving data quickly by using keywords in the form of a word or hashtag. The form of implementation made in this study is the website. This website is made with the Django framework with the Python programming language. To produce sentiments, a data mining process is needed. This data mining process uses a support vector machine algorithm and this process takes place in the backend of a website. The level of accuracy generated by this data mining process is 73%. The purpose of this sentiment analysis is used so that the data that has been collected can be used as useful information.

**Keywords**: Support Vector Machine, Text-Mining, Twitter, Django.

## 1. Introduction

Indonesia is one of the countries that have the largest growth of social media users in the world in one year, which is 20 million users in 2019 (Kemp, 2019, p. 71). The growth of Social Media can occur because it is increasingly easy to get internet access both from gadgets and using computers in all corners of the country.

With the increase of active users of social media, the amount of data contained on the internet in the form of text (text), images, and videos has also increased, and can even be said to increase exponentially. This is evidenced in 2013 the total number of data scattered on the internet was 4.4 ZB. Then in 2020 the amount of data is predicted to be 44 ZB (ZB = 10007 bytes) ) (Visual Capitalist, 2019) increased by 10 times in 7 years. With a large amount of data and increasing the faster it causes a "flood of information" that occurs where the data has value if it can be analyzed. This vast amount of information is not possible to analyze manually one by one without the aid of computer-based data analysis technology.

To analyze large amounts of data you can use text mining techniques. Text Mining is the process of obtaining high-quality information from texts (Purbo, 2019). Text mining can retrieve large amounts of data and then it is processed to get useful information from a collection of texts. The application can be done by text mining in the form of sentiment analysis on social media, information extraction to classify text, and text summarization to summarize long text into 1 or 2 paragraph form.

The object of the application of text mining that will be carried out in this study is social media. Texts on social media can be retrieved quickly with Text mining. After the data is taken, the data is cleaned first. The process of cleaning the data is called pre-processing, then after cleaning, the data is then processed by the method to be used.

A text on social media does not only convey information on information. But having information is sentiment. Sentiments are opinions or views based on excessive feelings towards something (Kamus Bahasa Indonesia, 2008 : 1319). The sentiment analysis process is carried out to get information in the form of sentiments about an object. Sentiments that are generally raised can be positive as people like, praise about the object and negative which can be a scolding or blasphemy against an object.

Twitter was chosen as the object of research because Twitter has an active number of users within one month of 330 million users in 2019 (Kemp, 2019 : 46). Then the text contained on Twitter can be easily mined because Twitter itself already provides a public API that can be used by anyone. Besides, there have been several studies that have made Twitter the subject of research.

Support Vector Machine is a technique for finding hyperplane that can separate two data sets from two different classes (Vapnik, 1995). This technique has been used to search mining texts with high accuracy as in the research conducted by Ira Zulfa and Edi Winarko. They mentioned that using the Support Vector Machine method can produce an accuracy of 92.18% compared to the Naive Bayes Classifier Method with 79.10% accuracy for Indonesian Tweet Sentiment Analysis (Zulfa & Winarko, 2017).

With the high accuracy generated by the Support Vector Machine Algorithm in the research that has been done, the authors motivated to implement sentiment analysis with the Support Vector Machine algorithm. The implementation is carried out in the form of a website application that can be used practically to analyze sentiments, especially in Indonesian.

This website is made so that users can get public sentiment about a particular object that is conveyed on Twitter and can be a measuring tool whether the objects contained in social media have a positive, neutral, or negative sentiment.

## 2. Material and Methods

The scope of this implementation are:

1. Data to be retrieved from Twitter can only be retrieved from the last 30 days due to the limitation of the Free Twitter API.
2. Sentiments were analyzed in Indonesian.
3. The results of text mining are positive, neutral, and negative sentiments from the tweet.
4. In analyzing text mining, emoticons are not included in the analysis and will be removed during the preprocessing process

### 2.1 Data Collection

In this study there are 2 types of data were collected. Training Data in the form of words containing positive or negative sentiment and Testing Data from tweets collected by the Twitter API. The purpose of this training data collection is so that the machine can learn words that have sentiments so that the machine can classify the data on Twitter as having positive or negative sentiments.

The training data used originally came from Bing Liu Opinion Words List (Bing, et al., 2005) which had been modified by Wahid, D.H into Indonesian (Wahid & Azhari, 2016). These positive and negative words can be downloaded at https://github.com/masdevid/US-OpinionWords. The number of negative words to be trained is 2436 words and the number of positive words is 1197 words so that the total training data used is 3633 words. Both positive and negative words that are used do not have capital letters, symbols so when you want to do classification, the data to be tested must first be converted into lowercase letters. The following table are samples of positive words used as training data

**Table 1.** Positive and Negative Sentiment

| No | Positive Sentiment | Negative Sentiment |
|---|---|---|
| 1 | acungan jempol | acungan jempol |
| 2 | adaptif | adaptif |
| 3 | adil | adil |
| 4 | afinitas | afinitas |
| 5 | afirmasi | afirmasi |

Then the testing data collected in this study are primary. This testing data is used as a test of the classification performed. This data is collected using the python tweepy library. The tweepy library in python can be used to access the Public API provided by Twitter.

During the data retrieval process using the Twitter API, a retweet filter process is performed. This retweet filter aims to avoid data redundancy because when collecting data with the Twitter API there is data that has the same contents. The Twitter API considers retweeting to be like a normal tweet. Tweets that are Retweeted begin with RT writing in front of them. The retweet filter process is done by adding "–filter: retweets" when using api.search from tweepy. In addition to the retweet filter, the lang = id parameter is used to retrieve tweets in Indonesian.

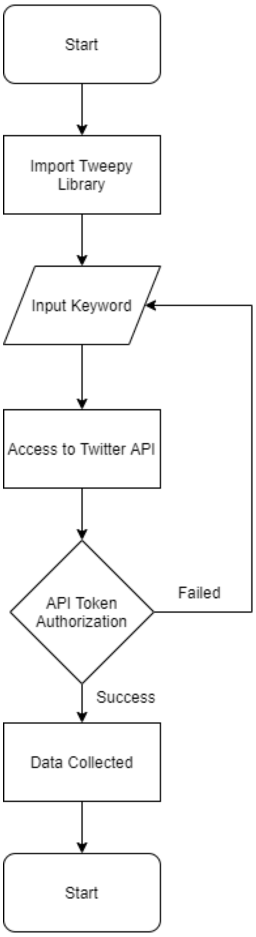


Figure 1 Process Data collection using Tweepy

## 2.2 Preprocessing

After the data has been collected, the available tweet data must be cleaned so the data can be classified to get the desired output. This data cleaning includes removing the word noise removal, lower case, stopword removal.

Lower case is done so that all data to be used is lowercase without any uppercase. This is because the computer distinguishes if the text is uppercase and lowercase. The training data to be analyzed has all lowercase letters. So if there is data with a capital letter, it will be missed by the existing training data.

**Table 2.** Lowercasing

| Raw | Lowercased |
|---|---|
| AMAN<br>AmAn<br>Aman | aman |
| ANJING<br>Anjing<br>anjinG | anjing |

Noise removal is done to eliminate the symbols before analysis. When retrieving data with the Twitter API there are many symbols, for example, \n which means break line, then question mark symbols, emoticons are removed so that tweets can be analyzed by training data. In addition to the symbol on the noise removal link removal, which is taken by the Twitter API because the tag will burden the analysis process.

**Table 3.** Noise Removal

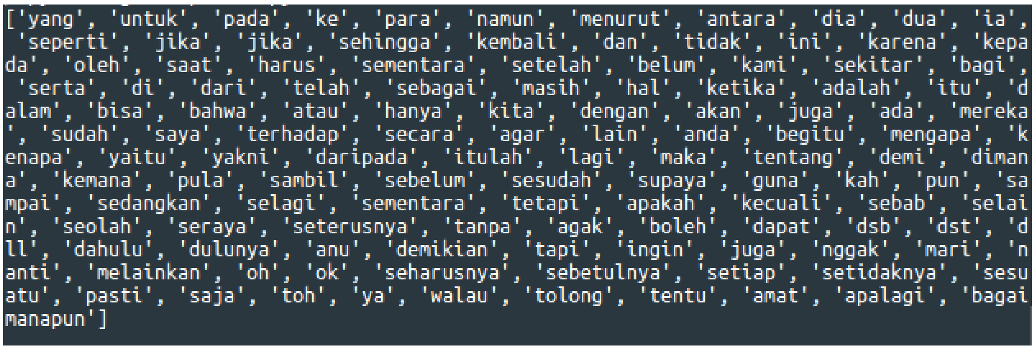
| Tweet | No Noise Tweet |
|---|---|
| @MVSOLAR Ada psbb segala sih �😭 | mvsolar ada psbb segala sih |
| @steffislsbl Kalo di tempat gue udh ga psbb jd dah banyak yang nongki | steffislsbl alo di tempat gue udh ga psbb jd dah banyak yang nongki |

Tokenization is the process of cutting a string in a text into a simpler form called a token. Tokens can be in the form of words, sentences, numbers and punctuation in a text. Tokenization can separate existing writing by looking for a barrier in a word. Barring a word that is commonly used is a space mark (Whitespace). Tokenization is done so that every piece of text in a tweet can be cleaned.

After the noise removal is done, the stopword removal process is performed. The purpose of removing the stopword is to reduce the amount of text that will be processed by

NLP so that the process is done faster. Stopword contains words that are less meaningful to be analyzed and can reduce the quality of the data being analyzed.

To eliminate stopwords in Indonesian, the Python library Sastrawi is used. Sastrawi is an NLP module in Indonesian created by Hanif Amal Robbani. This library can be downloaded via https://pypi.org/project/Sastrawi/. The library also has a stopword removal feature that can be used with the StopWordRemoverFactory class.



```
['yang', 'untuk', 'pada', 'ke', 'para', 'namun', 'menurut', 'antara', 'dia', 'dua', 'ia',
 'seperti', 'jika', 'jika', 'sehingga', 'kembali', 'dan', 'tidak', 'ini', 'karena', 'kepa
da', 'oleh', 'saat', 'harus', 'sementara', 'setelah', 'belum', 'kami', 'sekitar', 'bagi',
 'serta', 'di', 'dari', 'telah', 'sebagai', 'masih', 'hal', 'ketika', 'adalah', 'itu', 'd
alam', 'bisa', 'bahwa', 'atau', 'hanya', 'kita', 'dengan', 'akan', 'juga', 'ada', 'mereka
', 'sudah', 'saya', 'terhadap', 'secara', 'agar', 'lain', 'anda', 'begitu', 'mengapa', 'k
enapa', 'yaitu', 'yakni', 'daripada', 'itulah', 'lagi', 'maka', 'tentang', 'demi', 'diman
a', 'kemana', 'pula', 'sambil', 'sebelum', 'sesudah', 'supaya', 'guna', 'kah', 'pun', 'sa
mpai', 'sedangkan', 'selagi', 'sementara', 'tetapi', 'apakah', 'kecuali', 'sebab', 'selai
n', 'seolah', 'seraya', 'seterusnya', 'tanpa', 'agak', 'boleh', 'dapat', 'dsb', 'dst', 'd
ll', 'dahulu', 'dulunya', 'anu', 'demikian', 'tapi', 'ingin', 'juga', 'nggak', 'mari', 'n
anti', 'melainkan', 'oh', 'ok', 'seharusnya', 'sebetulnya', 'setiap', 'setidaknya', 'sesu
atu', 'pasti', 'saja', 'toh', 'ya', 'walau', 'tolong', 'tentu', 'amat', 'apalagi', 'bagai
manapun']
```

**Figure 2 List of Stopword**

The following table is a comparison before and after preprocessing

**Table 4. Preprocessing Result**

| Tweet | No Noise Tweet |
|---|---|
| @Polreslamongan1 Kapolri ajak masyarakat bersama tingkatkan kedisiplinan Indonesia Menuju New Normal\n\n #PolriDukungNewNormal | polreslamongan kapolri ajak masyarakat bersama tingkatkan kedisiplinan indonesia menuju new normal polridukungnewnormal |
| @_kangcilung @rheeechan @hrdbacot Masih mending yaa terima email doang. Jam 9 malem masih standby meeting zoom. New… https://t.co/kT4yaPHOJt | _kangcilung rheeechan hrdbacot mending yaa terima email doang jam malem standby meeting zoom new |

## 2.3 Data Mining

After the text data to be analyzed has been done the pre-processing process, the text can be analyzed. But before data classification can be done, the data of positive and negative sentences must first be trained. Training positive and negative data using the Support Vector Machine algorithm in the Scikit-learn. library.

This classification process begins by creating a feature vector for training data. This Vector Feature is made to convert writing into vector form so that the classification process can be carried out by Support Vector Machine. This process uses the library owned by sklearn with the feature_extraction.text module in the TfidfVectorizer.fit_transform function.

After the training data is converted into vectors, the vectors are located somewhere to study the pattern with the Support Vector Machine. The classification used in this study uses Nu Support Vector Regression. NuSVR has the same properties as ordinary Support Vector

Regression, but in NuSVR in the process of regression, the nu parameter is used to control the number of support vectors.

In the training data used positive sentences have a value of 1, while negative sentences have a value of -1. Because using NuSVR, the value of the classification results have a maximum value of 1 and the smallest value of -1. Giving values on positive and negative words is done so that in classifying it can issue neutral sentiment output when there are no positive or negative words in the sentence. The calculation results in SVM for each tweet have an output in the form of decimal numbers. Based on the value issued, a conversion table is made to determine the tweet has positive, neutral or negative sentiments.

**Table 5** Value Conversion

| Classification | Value Threshold |
|---|---|
| Positive | x $\geq$ 0.5 |
| Neutral | -0.5 $\leq$ x $\leq$ 0.5 |
| Negative | x $\leq$ -0.5 |

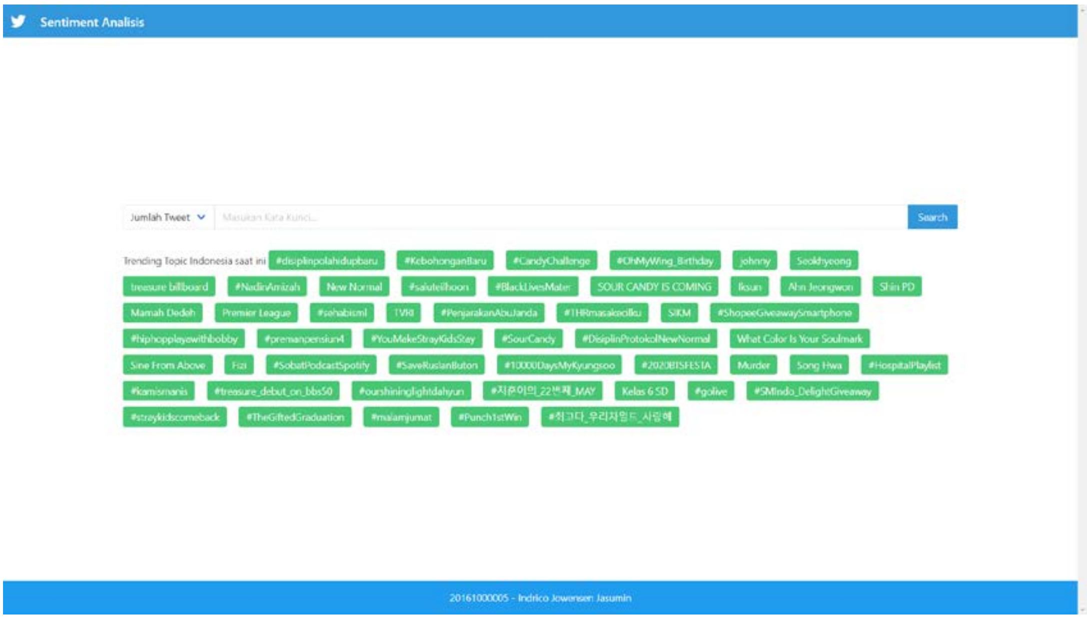
After the data is trained, the process of classifying data from tweets is done. Testing data in the form of tweets that have been cleaned are converted into vector form as well. Once converted into a vector the data will be classified. If the tweet has negative sentiment words then the placement of vector support is in an area that has a negative value such as -0.6 whereas if it has a positive sentiment the value of positive value is 0.56.

The support vector values from the tweet are strongly influenced by the training data used and the existing support vector is very much related to one another. For example, there is 1 positive sentiment that has a value of 1. If there is a word that is not sentimental or negative sentiment, the value of the existing vector support is shifted. If there is only one sentiment in a tweet sentiment, but the rest does not send at all in large numbers, overall the tweet is considered to have no sentiment. Therefore, the removal of the preposition (stopword) is done so that the classification results are better.

Data that has been trained will be stored using pickle. Pickle is used to reusing models and vectors from the training data that has been done so that when you want to re-classify the data, you don't need to re-train so the data classification process can be done faster. This pickle process is done by dumping a model that has been trained and an existing vector. This training data dump is called classifier.sav and vectorizer.sav dump vector shapes.

## 3. Result and Discussion

This website is created using a Python-based framework Django as Back-End. The version of Python used in making this website is Python 3.7. HTML, CSS and Javascript as the display interface (Front-End). The back end will process user input from the interface. Then the input will be used to process the text mining. The process of text mining occurs in the Back-End part of a website. After the data has been processed, from the back-end of a website it will output sentiment to the interface so that users can see the results. The figure below is a website interface


Sentiment Analisis

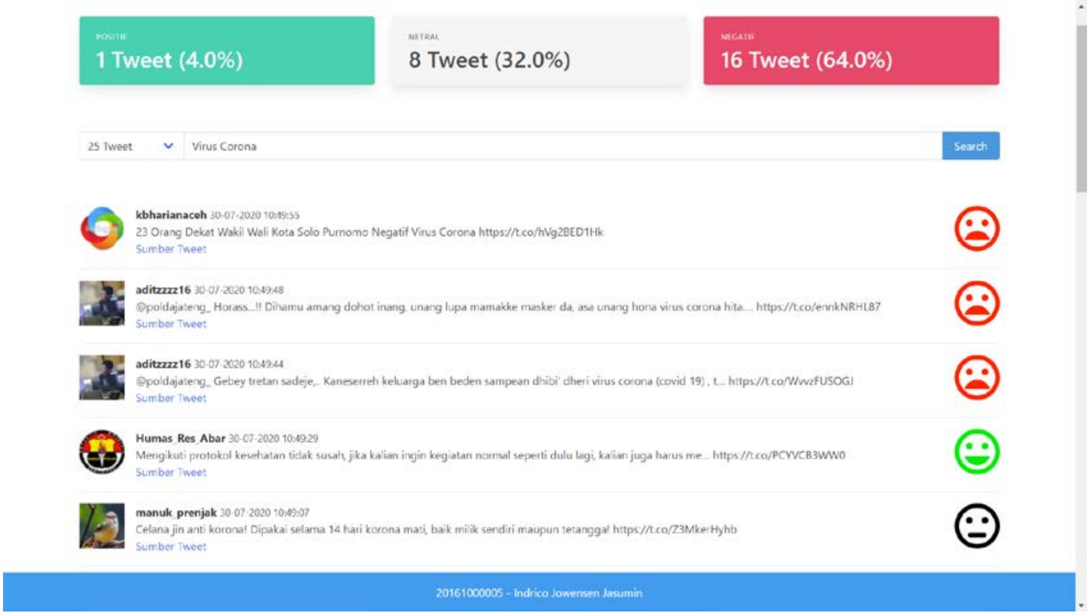
Jumlah Tweet | Masukan Kata Kunci... | Search

Trending Topic Indonesia saat ini #disiplinpolahidupbaru #KebohonganBaru #CandyChallenge #OhMyWing_Birthday johnny Seokhyeong treasure billboard #NadinAmizah New Normal #saluteilhoon #BlackLivesMater SOUR CANDY IS COMING Iksun Ahn Jeongwon Shin PD Mamah Dedeh Premier League #sehabisml TVRI #PenjarakanAbuJanda #THRmasakecilku SIKM #ShopeeGiveawaySmartphone #hiphopplaywithbobby #premanpensiun4 #YouMakeStrayKidsStay #SourCandy #DisiplinProtokolNewNormal What Color Is Your Soulmark Sine From Above Fizi #SobatPodcastSpotify #SaveRusianButon #10000DaysMyKyungsoo #2020BTSFESTA Murder Song Hwa #HospitalPlaylist #kamismanis #treasure_debut_on_bbs50 #ourshininglightdahyun #지훈이의_22번째_MAY Kelas 6 SD #golive #SMIndo_DelightGiveaway #straykidscomeback #TheGiftedGraduation #malamjumat #Punch1stWin [illegible]

20161000005 - Indrico Jowensen Jasumin

Figure 3 Website Interface

In the main view the user must input the number of tweets to be analyzed, the number of tweets that can be inputted by the user in the form of a dropdown which has parameters of 10 Tweets, 25 Tweets, 50 Tweets and 100 Tweets to be collected. 100 Tweets is the maximum usable limit due to the limitations of the Free API provided by Twitter. In addition to the number of tweets, the user must input keywords to be searched for. At the bottom, there are trending topics on Twitter in Indonesia. This trending topic was taken using tweepy library. It's accessed twitter API then collected current trending topic. By clicking on the trending topic below, the keywords section will automatically with the help of Javascript. After the search button is clicked, the display will appear as follows.


POSITIF
1 Tweet (4.0%)

NETRAL
8 Tweet (32.0%)

NEGATIF
16 Tweet (64.0%)

25 Tweet | Virus Corona | Search

**kbharianaceh** 30-07-2020 10:49:55
23 Orang Dekat Wakil Wali Kota Solo Purnomo Negatif Virus Corona https://t.co/hVg2BED1Hk
Sumber Tweet

**aditzzzz16** 30-07-2020 10:49:48
@poldajateng_ Horass...!! Dihamu amang dohot inang. unang lupa mamakke masker da, asa unang hona virus corona hita.... https://t.co/ennkNRHL87
Sumber Tweet

**aditzzzz16** 30-07-2020 10:49:44
@poldajateng_ Gebey tretan sadeje,. Kaneserreh keluarga ben beden sampean dhibi' dheri virus corona (covid 19) , t... https://t.co/WvvzFUSOGJ
Sumber Tweet

**Humas_Res_Abar** 30-07-2020 10:49:29
Mengikuti protokol kesehatan tidak susah, jika kalian ingin kegiatan normal seperti dulu lagi, kalian juga harus me... https://t.co/PCYVCB3WW0
Sumber Tweet

**manuk_prenjak** 30-07-2020 10:49:07
Celana jin anti korona! Dipakai selama 14 hari korona mati, baik milik sendiri maupun tetangga! https://t.co/Z3MkerHyhb
Sumber Tweet

20161000005 - Indrico Jowensen Jasumin

Figure 4 Result Page

In the results display, users can see the total number of tweets with positive, neutral and negative sentiments, user profile photos, user names, date and time the tweet was made, the contents of the tweet and analysis of the sentiment results on the tweet. The sentiment results in the following figure are emoticons. The smile emoticon and green color in the image mean that the tweet has a positive sentiment. The flat and black emoticon means the tweet has no sentiment or neutral and the sad red emoticon means the tweet is negative. The tweet source button when clicked will open a new tab to redirect the original tweet source on Twitter.

The process of evaluating data mining is done by conducting experiments 10 times with different keywords and data retrieval time then evaluated with a confusion matrix to measure the level of accuracy generated between the experiments, the precision obtained, the value of recall and prevelance. Each experiment will evaluate 50 tweets so that the total data to be evaluated will be 500 tweets. Keywords used to carry out this evaluation include "Jokowi", "Virus Corona", "Ahok", and "New Normal".

The accuracy in this research is calculated by conducting an independent evaluation of each tweet tested. This amount of accuracy is calculated by calculating the total predicted correct tweets divided by the total tweets tested

The following table is list the experiment and time the analysis has been done

**Table 6** List the evaluation conducted

| No | Keyword | Time |
|---|---|---|
| 1 | Jokowi | June 29, 2020 12:00 UTC+7 |
| 2 | Virus Corona | June 29, 2020 13:00 UTC+7 |
| 3 | Ahok | June 29, 2020 13:00 UTC+7 |
| 4 | Tangerang | June 29, 2020 14:00 UTC+7 |
| 5 | Jokowi | June 30, 2020 16:00 UTC+7 |
| 6 | New Normal | June 30, 2020 16:00 UTC+7 |
| 7 | Virus Corona | June 30, 2020 16:00 UTC+7 |
| 8 | Jokowi | July 1, 2020 01:00 UTC+7 |
| 9 | New Normal | July 1, 2020 01:00 UTC+7 |
| 10 | PSBB | July 1, 2020 23:00 UTC+7 |

The following table is a sample tweet that was successfully classified

**Table 7** Correct Classification

| Tweet | Prediction | Actual |
|---|---|---|
| jokowi marah ancam reshuffle mardani jangan mengeluh depan rakyat darisuara | Negative | Negative |
| pihak istana melalui deputi bidang protokol pers media sekretariat presiden ri bey triadi machmudin menjelask | Neutral | Neutral |
| narasinewsroom jokowi mentri jokowi aja degil gimana rakyat ga degil pemerintahan sendiri ngajarin degil | Negative | Negative |
| robert__moses pak jokowi fokus mengatasi pandemi yg diharapkan kpd menteri nya dukung terus tolakadudombafokuspandemi | Positive | Positive |
| null perlukah jokowi reshuffle menteri ekonomi presiden jokowi memerintahkan jajaran | Neutral | Neutral |

The first tweet in the table above gives negative predictions according to the actual sentiment that occurred. Giving negative sentiment to the tweet is influenced by the word "ancam" because the word "ancam" itself is already a negative sentiment. Then the second tweet has no sentiment at all because the sentence contains only news descriptions and there are no words that have sentiment.

The third tweet contained the word " degil " which had negative sentiments and the word was repeated several times so that the sentiment on the tweet was negative. Then for the fourth tweet, there are the words "mengatasi" and "dukung" which make the sentiment in the tweet to be positive. Then for the fifth tweet contains only questions that have no sentiment.

However, in this experiment, there was a misclassification. The following table is an example of misclassified tweets

**Table 8** Misclasification

| Tweet | Prediksi | Aktual |
|---|---|---|
| berikut pidato lengkap jokowi sentil kinerja menteri hingga singgung reshuffle | Neutral | Negative |
| jansen_jsp jokowi janji tunjangan buat tenaga medis belum terealisasi | Neutral | Negative |
| jokowi marah ancam reshuffle menteri layak diganti | Neutral | Negative |
| video tersebut presiden jokowi tampak kesal lantaran merasa menternya | Neutral | Negative |

| | | |
|---|---|---|
| mempunyai perasaan tuju | | |
| ratas tersebut jokowi menekankan pentingnya komunikasi sosialisasi masyarakat soal berbagai langk | Neutral | Positive |

The first tweet on table above there are the words "sentil" and "singgung" which should be negative sentiment, but the training data does not contain these words so that the tweet is considered to have no sentiment. Then in the second tweet there is the word "belum terealisasi" which is also not found in the training data. However, for the third tweet, the result of the predictive sentiment became neutral because there was the word "layak" which had a positive sentiment. When there is a negative and positive sentiment in 1 sentence it causes the resulting sentiment to be neutral.

For the fourth tweet the word "kesal" has a negative sentiment, but because there is the word "perasaan" the sentiment contained in the sentence is predicted to be neutral and the fifth tweet has "menekankan" which according to the training data used is negative sentiment and "pentingnya komunikasi" has a positive sentiment. so that the tweet is predicted to be neutral sentiment.

After 10 experiments that have been carried out calculated the average of accuracy, precision recall and prevalence. The following table is the overall results of the experiments that have been carried out.

**Table 9** Overall Evaluation

| No | Accuracy | Precision | Recall | Prevalence |
|---|---|---|---|---|
| 1 | 74% | 83% | 63% | 16% |
| 2 | 66% | 57% | 50% | 16% |
| 3 | 74% | 95% | 72% | 58% |
| 4 | 70% | 17% | 50% | 8% |
| 5 | 74% | 75% | 50% | 24% |
| 6 | 66% | 20% | 14% | 14% |
| 7 | 76% | 100% | 75% | 16% |
| 8 | 74% | 86% | 75% | 16% |
| 9 | 76% | 100% | 67% | 30% |
| 10 | 78% | 33% | 25% | 8% |
| Average | 73% | 67% | 54% | 21% |

The resulting accuracy values range from 66% to 78% with an average accuracy of 73%. For precision values vary greatly from 17% to 100%. This precision value is greatly influenced by the amount of prevalence that exists. Can be seen in the fourth and tenth trials which have the lowest prevalence value. The resulting precision is also low. The average precision produced is 66%.

The resulting recall value has a range ranging from 14% to 75% with an average produced is 54%. Then the prevalence value produced has a range ranging from 8% to 30% with an average yield of 21%.

## 4. Conclusion

By using text mining, sentiments on social media can be analyzed quickly when compared to seeing one by one tweet on social media by determining the sentiments contained in the tweet, but this will not happen if Twitter does not have an API that can publicly accessed and a collection of words that contain positive and negative sentiments in the Indonesian language that are used as training data.

Support Vector Machine can be implemented in the form of a website using the Django web framework. The accuracy value generated by the SVM algorithm applied in this study for sentiment analysis in the form of Indonesian tweets is 73%.

However, the level of accuracy, precision and recall produced in this study is greatly influenced by the quality of the training data used. If the training data used is not good using any algorithm, the results are isn’t desirable.

## 5. References

Bing, L., Minqing, H. & Cheng, J., 2005. Opinion Observer: Analyzing and Comparing Opinions on the Web." Proceedings of the 14th International World Wide Web Conference (WWW-2005). *WWW '05 Proceedings of the 14th international conference on World Wide Web*, pp. 342-351.

Kamus Bahasa Indonesia, (2008). *Kamus Bahasa Indonesia*. 5th red. Jakarta: Pusat Bahasa.

Kemp, S., (2019). Digital 2019: Global Digital Overview, Vancouver: Hootsuite.

Purbo, O. W., 2019. *Text Mining Analisis MedSos, Kekuatan Brand & Intelijen di Internet.* Yogyakarta: Penerbit ANDI.

Vapnik, V., (1995). *The Nature of Statistical Learning Theory*. Berlin: Springer.

Visual Capitalist, 2019. *How Much Data is Generated Each Day?.* [Online] Available at: https://www.visualcapitalist.com/wp-content/uploads/2019/04/data-generated-each-day-full.html
[Använd 22 10 2019].

Wahid, D. H. & Azhari, S. N., 2016. Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity. *Indonesian Journal of Computing and Cybernetics Systems*, 10(2), pp. 207-218.

Zulfa, I. & Winarko, E., (2017). Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), Volym 11, pp. 187-198.