

SENTIMENT ANALYSIS COVID VACCINE ON TWITTER USING SVM ALGORITHM

¹Ratna Akila Reta, ²Rosita Herawati

^{1,2}Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Katolik Soegijapranata
²rosita@unika.ac.id

ABSTRACT

Currently there are still many people who have not been vaccinated. many people comment on social media, especially the twitter application, to provide opinions and comments about the covid vaccine in Indonesia. In order to find out which tweets appeared including positive comments or negative comments, an analysis was carried out to find out how many negative and positive comments were by sentiment analysis. This study analyzes comments taken from the twitter application using the crawling method. Furthermore, pre-processing will be carried out such as case folding, tokenization, and stopwords filtering. Then before the classification is carried out, data labeling and data split will be carried out to facilitate classification. After that, classification will be carried out using the SVM algorithm method. The final result of this research project is that there are 9981 data obtained from the crawling method. This data proves that 92.4% of Indonesians gave a neutral response to the topic of the Covid vaccine on the Twitter application. The SVM algorithm is easy to use in the implementation of sentiment analysis. And the accuracy results obtained using the SVM method are very high at 90.5%.

Keywords: SVM, covid, vaccine, twitter

BACKGROUND

The covid vaccine has been implemented in early 2021. But until now there are still many people who have not been vaccinated. Some people refuse to be vaccinated because they think that Covid doesn't exist, are worried about side effects after vaccines, and don't believe in certain vaccines. Nowadays people are free to express their opinions and opinions on social media. Twitter is one of the social media that is often used to express opinions. Currently, many Twitter users provide opinions or comments about vaccines in Indonesia. But the tweets that appear are still arranged randomly which makes it difficult for readers to find out more negative or positive comments that appear on the topic of this covid vaccine.

Based on the explanation above, a sentiment analysis research was conducted on Twitter to classify Twitter user comments about the Covid vaccine in Indonesia. To conduct this research, firstly, a data crawling process will be carried out whose function is to retrieve tweet data, after that the data will be pre-processed, then proceed with classifying tweets into 3 classes, namely positive, neutral and negative. This classification uses the Support vector Machine algorithm. This classification can make it easier for users to find out positive, neutral, and negative comments.

By using the Support Vector Machine algorithm for sentiment analysis classification on Twitter, users can find out comments about this covid vaccine that have more positive or negative tweets easily.

RESEARCH METHODOLOGY

Literature Study

There are 10 journals used in this study. From all these journals, it was concluded that most of the journals used the naive Bayes algorithm and the rest used the SVM and K-NN algorithms in classifying sentiment analysis. Therefore, in this study, the SVM algorithm is used because it is easy to use for sentiment analysis on Twitter which will be examined.

Dataset Preparation

The dataset used in this study is a collection of data about vaccines and covid taken from Twitter which was taken using the data crawling method. The data taken is approximately 9000 tweet data. The data taken is datetime, name, and tweet.

To be able to retrieve Twitter data, users must have an API Key and a customer key to make it easier for users to retrieve tweet data in real time. The API key can be retrieved on the twitter developer site. To be able to get the API key the user must have a twitter account. Twitter will share the API Key number once certain terms and conditions have been met. After that, the user can start crawling after the API Key is obtained. The dataset is taken from opinions and comments that come from the twitter application. To get this data, you must have a twitter account and have registered as a developer account to get an API key. The data is taken using the crawling method. The data obtained are datetime, name, and tweet.

Examples of datasets obtained after crawling data :

Table 1. dataset after data crawling process

no	datetime	name	Tweet
1.	2021-11-23 13:06:35	polresjembrana	Bhabinkamtibmas Kelurahan Tegalcangkring, Polsek Mendoyo Pengamanan Dan Pemantauan Vaksin Covid 19 Door to Door. https://t.co/GtkalyHXIH
2.	2021-11-23 13:05:54	kemitris	@AldoBabeh Anies gagal dalam urusan vaksin, masih kurang mumpuni jadi pemimpin. Walau debatable, covid 19 bukan wab... https://t.co/78ToKcfr6r
3.	2021-11-23 13:05:37	tvOneNews	Mobil Vaksin Covid-19 Keliling, Ridwan Kamil: Kami Mengapresiasi Armada Vaksin Ini https://t.co/S1vxaAAtZa
4.	2021-11-23 13:05:29	VvVwWwLq	RT @TedHilbert: Mau tahu kenapa orang2 seperti @dr_koko28 @blogdokter @tonangardyanto @__Sridiana_3va

			@drpriono1 dan masih banyak lagi meny...
5.	2021-11-23 13:06:35	polresjembrana	Bhabinkamtibmas Kelurahan Tegalcangkring, Polsek Mendoyo Pengamanan Dan Pemantauan Vaksin Covid 19 Door to Door. https://t.co/GtkalyHXIH

Pre-processing

1. Case folding: Converts all letters to lowercase. Characters other than letters a to z will be omitted. Then delete numbers and punctuation that have nothing to do with what will be analyzed.
2. Tokenization: The process of separating text into chunks of words so that they can then be analyzed.
3. Filtering stopwords: the process of taking important words from the tokenization results. Common words that often appear but have no meaning in the data will be deleted and classified as general words.

TF-IDF

The data that was previously pre-processed will be calculated using the tf-idf method after which the data will be tested. First, calculate the frequency in the training document. After that the results will show how often the word appears in the document. Then, do this method to test some other data.

SVM Algorithm

The classification process will be carried out using the SVM algorithm, which stands for Support Vector Machine. This classification process is carried out using negative and positive data originating from a review of data that has been taken from twitter through the crawling method and has been pre-processed, then the data has negative and positive comment data that have previously been through the labeling process, then will be used. SVM algorithm to study data patterns based on the characteristics of the data in each class. After that, the results of this SVM algorithm will be tested using test data, so that after that get the results of the level of accuracy and prediction results.

This study uses the SVM algorithm to determine negative, neutral and positive sentiment in the classification of sentiment analysis, comments and opinions on the Twitter application. In this study the programming language used is Python. By using the google colab text editor to run the project.

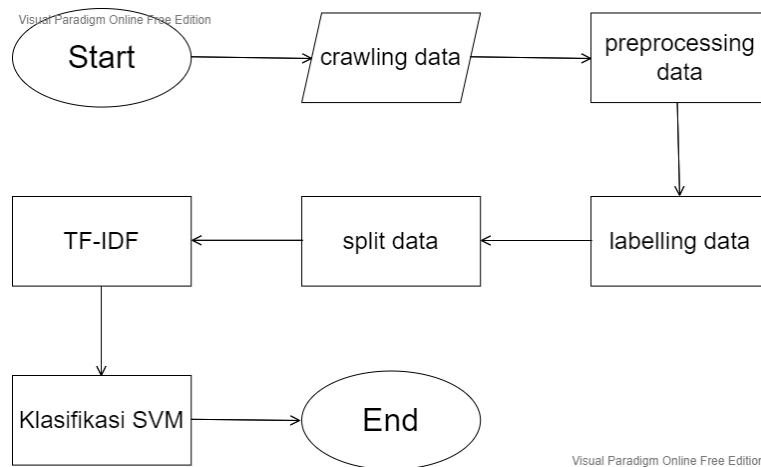


Figure 1. Process flowchart

The first step in this research is to take the dataset using data crawling, then the data that has been obtained must go through the pre-processing stage, namely the case folding stage, tokenization, filtering and stopwords. After that, it goes into the labelling process to separate negative, neutral and positive sentiment data. Then split the data to get test data and train data. Then, enter the TF-IDF process, and after that enter the SVM Algorithm process.

IMPLEMENTATION AND RESULT

In this study, it was shown that 92.4% of Twitter application users gave a neutral response to the topic of the Covid vaccine. But there are 0.8% of the people who gave a negative response to the covid vaccine. Meanwhile, 6.7% of people in Indonesia gave a positive response to the topic of the Covid vaccine in Indonesia. In this study, before classifying the data, the thing that must be done is to retrieve data using the crawling method by having a twitter account to get the twitter API which will be used to retrieve data from twitter using the crawling method. Here are some data results that have been successfully retrieved from Twitter. The picture below is some of the data results that were successfully retrieved through the crawling method by having an API key that contains `consumer_key`, `consumer_secret`, `access_token`, and `access_secret`. The data taken are datetime, name, and tweet. The data that was successfully retrieved were 9981 data.

After the data has been successfully retrieved by the crawling method, then the data will go through a pre-processing process. In this process, case folding will be carried out which serves to convert all letters into lowercase letters, only letters a to z are accepted. Other than that character will be omitted and considered delimiter. At this stage only use the modules available in python and do not use any external libraries. In addition to changing the letters to lowercase. The result of this process is in the TWEET column. The following is the result of case folding which converts letters into lower case.

	A	B	C
1	datetime	name	tweet
2	2021-11-23 13:06:35	poiresjembrana	Bhabinkamtibmas Kelurahan Tegalcangkriing, Polsek Mendoyo Pengamanan Dan Pemantauan Vaksin Covid 19 Door to Door. https://t.co/GtkalgjHXIH
3	2021-11-23 13:05:54	kemitris	@AldoBabeh Anies gagal dalam urusan vaksin, masih kurang mumpuni jadi pemimpin. 'Walaupun debatable, covid 19 bukan wab... https://t.co/78ToKof6r
4	2021-11-23 13:05:37	tvOneNews	Mobil Vaksin Covid-19 Keliling, Ridwan Kamil: Kami Mengapresiasi Armada Vaksin Ini https://t.co/S1vvaAAZa
5	2021-11-23 13:05:29	VvVvWwLq	RT @TedHilbert: Mau tahu kenapa orang2 seperti @dr_koko28 @blogdokter @tonangardjanto @__Sridiana_3va @drpriono1 dan masih banyak lagi meny...
6	2021-11-23 13:04:18	jeff_ikhlas	RT @TedHilbert: Mau tahu kenapa orang2 seperti @dr_koko28 @blogdokter @tonangardjanto @__Sridiana_3va @drpriono1 dan masih banyak lagi meny...
7	2021-11-23 13:04:06	StopPlandemit	RT @mas_prasetyo: Tidak usah sok superior seperti itu. Kalian yg percaya pd vaksin covid silak an, tidak ada melindungi ini itu, masing2 in...
8	2021-11-23 12:59:21	kab_upp	UPP Kab.Bantaeng melaksanakan giat pengamanan & amp; pendampingan vaksin covid-19 di kantor desa pattaneteang Kec.Tompob... https://t.co/QoJleL0L
9	2021-11-23 12:59:08	Primaditaaaa	RT @NarasiNewsroom: Sebelas jenis vaksin COVID-19 yang mendapatkan izin penggunaan darurat dari Badan Pengawas Obat dan Makanan (BPOM) bisa...
10	2021-11-23 12:58:21	kab_upp	UPP Kab.Bantaeng melaksanakan giat pengamanan & amp; pendampingan vaksin covid-19 di kantor lurah bonto sunggu Kec.Bissa... https://t.co/O72IHuP796
11	2021-11-23 12:56:54	apien_k19	RT @TedHilbert: Must watch!nI Dokter sejati yang mengerti tentang Covid.Dokter sales vaksin di Indonesia seharusnya merasa malu pada dir...
12	2021-11-23 12:56:45	BKabsumedang	Sobat BUMN, Sebagai bentuk kepedulian terhadap percepatan penanganan COVID-19, Pusri melaksanakan penyerahan ban... https://t.co/sogqvAGzA
13	2021-11-23 12:55:38	DinkesOfficial	Data vaksin dan pasien covid-19 hari ini https://t.co/X7ov5J2eH
14	2021-11-23 12:55:30	ubsibekasi	SENTRA VAKSIN COVID-19 UNIVERSITAS BSI KAMPUS BOGOR A (CILEBUT)Hai, Sobat BSI. Catat Informasi pentingnjai Lok... https://t.co/7GT10t6pwz
15	2021-11-23 12:55:27	Abughazy_2007	RT @TedHilbert: Peningkatan 40% Kematian akibat Vaksin Covid-19; Gangguan Jantung di antara orang berusia 15-44 tahun telah melonjak hing...
16	2021-11-23 12:54:16	apien_k19	RT @TedHilbert: Akhirnya mereka melakukan survei serologi. Tidak perlu vaksin berbahaya. Covid sudah endemik seperti flu. Pandemi selesai...

Figure 2. Some of the results of the data taken from twitter

	datetime	name	tweet	TWEET
0	2021-11-23 13:06:35	poiresjembrana	Bhabinkamtibmas Kelurahan Tegalcangkriing, Pols...	bhabinkamtibmas kelurahan tegalcangkriing, pols...
1	2021-11-23 13:05:54	kemitris	@AldoBabeh Anies gagal dalam urusan vaksin, ma...	@aldobabeh anies gagal dalam urusan vaksin, ma...
2	2021-11-23 13:05:37	tvOneNews	Mobil Vaksin Covid-19 Keliling, Ridwan Kamil: ...	mobil vaksin covid-19 keliling, ridwan kamil: ...
3	2021-11-23 13:05:29	VvVvWwLq	RT @TedHilbert: Mau tahu kenapa orang2 seperti...	rt @tedhilbert: mau tahu kenapa orang2 seperti...
4	2021-11-23 13:04:18	jeff_ikhlas	RT @TedHilbert: Mau tahu kenapa orang2 seperti...	rt @tedhilbert: mau tahu kenapa orang2 seperti...
5	2021-11-23 13:04:06	StopPlandemit	RT @mas_prasetyo: Tidak usah sok superior sep...	rt @mas_prasetyo: tidak usah sok superior sep...
6	2021-11-23 12:59:21	kab_upp	UPP Kab.Bantaeng melaksanakan giat pengamanan ...	upp kab bantaeng melaksanakan giat pengamanan ...
7	2021-11-23 12:59:08	Primaditaaaa	RT @NarasiNewsroom: Sebelas jenis vaksin COVID...	rt @narasinewsroom: sebelas jenis vaksin covid...
8	2021-11-23 12:58:21	kab_upp	UPP Kab.Bantaeng melaksanakan giat pengamanan ...	upp kab bantaeng melaksanakan giat pengamanan ...

Figure 3. The results of case folding change all letters to lowercase

Furthermore, in the case folding process, it is also done to remove numeric characters and punctuation marks that have nothing to do with what will be analyzed. In this process python uses the re module to remove this character. The result of this process is in the tweet_clean column. The following is the result of case folding that removes numbers and punctuation marks.

	datetime	name	tweet	TWEET	tweet_clean
0	2021-11-23 13:06:35	poiresjembrana	Bhabinkamtibmas Kelurahan Tegalcangkriing, Pols...	bhabinkamtibmas kelurahan tegalcangkriing, pols...	bhabinkamtibmas kelurahan tegalcangkriing polse...
1	2021-11-23 13:05:54	kemitris	@AldoBabeh Anies gagal dalam urusan vaksin, ma...	@aldobabeh anies gagal dalam urusan vaksin, ma...	aldobabeh anies gagal dalam urusan vaksin mas...
2	2021-11-23 13:05:37	tvOneNews	Mobil Vaksin Covid-19 Keliling, Ridwan Kamil: ...	mobil vaksin covid-19 keliling, ridwan kamil: ...	mobil vaksin covid keliling ridwan kamil kami ...
3	2021-11-23 13:05:29	VvVvWwLq	RT @TedHilbert: Mau tahu kenapa orang2 seperti...	rt @tedhilbert: mau tahu kenapa orang2 seperti...	tedhilbert mau tahu kenapa orang2 seperti koko...
4	2021-11-23 13:04:18	jeff_ikhlas	RT @TedHilbert: Mau tahu kenapa orang2 seperti...	rt @tedhilbert: mau tahu kenapa orang2 seperti...	tedhilbert mau tahu kenapa orang2 seperti koko...
5	2021-11-23 13:04:06	StopPlandemit	RT @mas_prasetyo: Tidak usah sok superior sep...	rt @mas_prasetyo: tidak usah sok superior sep...	mas prasetyo tidak usah sok superior seperti...
6	2021-11-23 12:59:21	kab_upp	UPP Kab Bantaeng melaksanakan giat pengamanan ...	upp kab bantaeng melaksanakan giat pengamanan ...	upp kab bantaeng melaksanakan giat pengamanan ...
7	2021-11-23 12:59:08	Primaditaaaa	RT @NarasiNewsroom: Sebelas jenis vaksin COVID...	rt @narasinewsroom: sebelas jenis vaksin covid...	narasinewsroom sebelas jenis vaksin covid yan...
8	2021-11-23 12:58:21	kab_upp	UPP Kab Bantaeng melaksanakan giat pengamanan ...	upp kab bantaeng melaksanakan giat pengamanan ...	upp kab bantaeng melaksanakan giat pengamanan ...
9	2021-11-23 12:56:54	apien_k19	RT @TedHilbert: Must watch!nI Dokter sejati...	rt @tedhilbert: must watch!nini dokter sejati...	tedhilbert must watch in dokter sejati yang ...
10	2021-11-23 12:56:45	BKabsumedang	Sobat BUMN,\n\nSebagai bentuk kepedulian terh...	sobat bum,\n\n\nsebagai bentuk kepedulian terh...	sobat bum sebagai bentuk kepedulian terhadap ...
11	2021-11-23 12:55:38	DinkesOfficial	Data vaksin dan pasien covid-19 hari ini https://t.co/X7ov5J2eH	data vaksin dan pasien covid-19 hari ini https://t.co/X7ov5J2eH	data vaksin dan pasien covid hari ini https://t.co/X7ov5J2eH
12	2021-11-23 12:55:30	ubsibekasi	SENTRA VAKSIN COVID-19 UNIVERSITAS BSI KAMPUS ...	sentra vaksin covid-19 universitas bsi kampus ...	sentra vaksin covid universitas bsi kampus bog...
13	2021-11-23 12:55:27	Abughazy_2007	RT @TedHilbert: Peningkatan 40% Kematian akib...	rt @tedhilbert: peningkatan 40% kematian akib...	tedhilbert peningkatan kematian akibat vaksin...
14	2021-11-23 12:54:16	apien_k19	RT @TedHilbert: Akhirnya mereka melakukan surv...	rt @tedhilbert: akhirnya mereka melakukan surv...	tedhilbert akhirnya mereka melakukan survei s...

Figure 4. case folding results remove numbers and punctuation marks

After the case folding process is complete, then in the tweet_clean column the tokenization process will be carried out. This process serves to separate the text into pieces called tokens so that they can then be analyzed. The result of this process is in the TOKENIZATION column. The following is the result of tokenization

tweet_clean	TOKENIZATION
bhabinkamtibmas kelurahan tegalcangkring polse...	[bhabinkamtibmas, kelurahan, tegalcangkring, p...
aldobabeh anies gagal dalam urusan vaksin mas...	[, aldobabeh, anies, gagal, dalam, urusan, vak...
mobil vaksin covid keliling ridwan kamil kami ...	[mobil, vaksin, covid, keliling, ridwan, kamil...
tedhilbert mau tahu kenapa orang seperti koko...	[, tedhilbert, mau, tahu, kenapa, orang, seper...
tedhilbert mau tahu kenapa orang seperti koko...	[, tedhilbert, mau, tahu, kenapa, orang, seper...
mas prasetyo tidak usah sok superior seperti...	[, mas, prasetyo, tidak, usah, sok, superior,...
upp kab bantaeng melaksanakan giat pengamanan ...	[upp, kab, bantaeng, melaksanakan, giat, penga...
narasinewsroom sebelas jenis vaksin covid yan...	[, narasinewsroom, sebelas, jenis, vaksin, cov...
upp kab bantaeng melaksanakan giat pengamanan ...	[upp, kab, bantaeng, melaksanakan, giat, penga...
tedhilbert must watch ini dokter sejati yang ...	[, tedhilbert, must, watch, ini, dokter, sejat...

Figure 5. Result of tokenization process

After the tokenization process is complete, then in the TOKENIZATION column, a stopwords filtering process will be carried out. In this process, it functions to retrieve important words from the tokenization results by using a stoplist algorithm that functions to remove less important words and a wordlist to store important words. Words that often appear but have no meaning will be deleted and classified as general words. The result of this process is in the STOP_REMOVAL column. The following is the result of filtering stopword.

TOKENIZATION	STOP_REMOVAL
[bhabinkamtibmas, kelurahan, tegalcangkring, p...	bhabinkamtibmas kelurahan tegalcangkring polse...
[, aldobabeh, anies, gagal, dalam, urusan, vak...	aldobabeh anies gagal urusan vaksin mumpuni p...
[mobil, vaksin, covid, keliling, ridwan, kamil...	mobil vaksin covid keliling ridwan kamil menga...
[, tedhilbert, mau, tahu, kenapa, orang, seper...	tedhilbert orang koko blogdokter tonangardyan...
[, tedhilbert, mau, tahu, kenapa, orang, seper...	tedhilbert orang koko blogdokter tonangardyan...
[, mas, prasetyo, tidak, usah, sok, superior,...	mas prasetyo sok superior percaya vaksin cov...

Figure 6. Stopwords filtering process

Furthermore, after the case folding process has been completed, a labeling process is carried out in the tweet_final column which functions to label sentiments which include negative.

comments, positive comments, and neutral comments. In this process there are 9981 data labeled. The results of the labeling process are in the label column. Here are the results of the labeling process.

	datetime	name	tweet_akhir	label
0	2021-11-23 13:06:35	polresjembrana	bhabinkamtibmas kelurahan tegalcangkring polse...	neutral
1	2021-11-23 13:05:54	kemitris	aldobabeh anies gagal urusan vaksin mumpuni pe...	neutral
2	2021-11-23 13:05:37	tvOneNews	mobil vaksin covid keliling ridwan kamil menga...	neutral
3	2021-11-23 13:05:29	VvVwWwLq	tedhilbert orang koko blogdokter tonangardyant...	neutral
4	2021-11-23 13:04:18	jeff_ikhlas	tedhilbert orang koko blogdokter tonangardyant...	neutral
...
9976	2021-11-17 6:53:51	CapikWite	fahmibinhassan dos penggalak vaksin solusi ide...	positif
9977	2021-11-17 6:53:09	_mfdks	fahmibinhassan dos penggalak vaksin solusi ide...	positif
9978	2021-11-17 6:52:52	FirmanSupriadi7	mediaindonesia pakar penyakit menular terkemuk...	neutral
9979	2021-11-17 6:52:52	EzriIF	fahmibinhassan dos penggalak vaksin solusi ide...	positif
9980	2021-11-17 6:52:25	khaalisz	fahmibinhassan dos penggalak vaksin solusi ide...	positif

9981 rows x 4 columns

Figure 7. The result of the labeling process

Furthermore, after the labeling process, the data will be split. This data will be divided into test data and train data. The test data will be used to determine the performance of the previously trained algorithm and when finding new data that has never been seen before.

	tweet_akhir	label
6907	siswa amerika salah menerima dosis vaksin covi...	neutral
9566	pelanjutan jangka hayat produk vaksin covid je...	neutral
8082	vaksin keluarga kebal covid https pkjoyjxtr	neutral
1861	perkembangan menjadikan kumulatif pemberian va...	neutral
1509	tedhilbert koko data https zzyrvnkr	neutral
...
7890	babinsa koramil reban serda dwi bhabinkamtibma...	neutral
4772	sinaronline khairy data kemasukan hospital pen...	neutral
8171	pemerintah mengajak masyarakat takut divaksin ...	neutral
2405	ayo vaksin cegah penyebaran covid https gwxnth...	neutral
7031	merdekadotcom dosis vaksin covid wilayah das b...	neutral

999 rows x 2 columns

Figure 8. Test data results

While the training data is used to train the algorithm used. After splitting the data, the test data has 999 data while the train data has 8982 data. The following are the results of the test data and train data that have been separated.

```
[10] df_train
```

	tweet_akhir	label
8786	jpenerangan suntikan dos penggalak vaksin covi...	neutral
6024	cek fakta istri ceo pfizer meninggal akibat ko...	neutral
5800	indonesia terima juta dosis vaksin astrazeneca...	neutral
6435	sobat bumh bentuk kepedulian percepatan penang...	neutral
2168	kuala lumpur november malaysia menerima juta t...	neutral
...
9225	kkmputrajaya kenyataan akhbar kementerian kesi...	neutral
4859	vaksin covid dikawal ahli bpom mui https uuwkgzvc	neutral
3264	alexcg hanux tedhilbert micmedsos aldobabeh sh...	neutral
9845	kkmputrajaya pelanjutan jangka hayat produk va...	neutral
2732	terbuka warga kalimantan utara ayo warga bulun...	neutral

8982 rows x 2 columns

Figure 9. training data results

After splitting the data, then a test is carried out using tf-idf which serves to reduce the weight of a term if its occurrence is spread throughout the document. Here are some of the results.

'kelurahan': 1746, 'polsek': 3305, 'pengamanan': 3003, 'pemantauan': 2919, 'vaksin': 4321, 'covid': 695, 'door': 982, 'https': 1360, 'gagal': 1166, 'urusan': 4299, 'wab': 4501, 'mobil': 2515, 'keliling': 1736, 'ridwan': 3555, 'orang': 2730, 'koko': 1889, 'blogdokter': 521, 'tonangardyanto': 4197, 'sridiana': 3895, 'va': 4314, 'drpriono': 1005, 'meny': 2433, 'mas': 2150, 'prasetiyo': 3364, 'sok': 3868, 'superior': 3956, 'percaya': 3070, 'silakan': 3823, 'melindungi': 2214, 'upp': 4296, 'kab': 1602, 'bantaeng': 286, 'melaksanakan': 2197, 'giat': 1214, 'amp': 111, 'pendampingan': 2981, 'kantor': 1653, 'desa': 788, 'pattaneteang': 2830, 'kec': 1705, 'tompob': 4194, 'narasinewsroom': 2597, 'sebelas': 3680, 'jenis': 1552, 'izin': 1492, 'penggunaan': 3020, 'darurat': 743, 'badan': 239, 'pengawas': 3010, 'obat': 2706, 'makanan': 2105, 'bpom': 548, 'lurah': 2079, 'bonto': 537, 'sunggu': 3947, 'bissa': 508, 'must': 2572, 'watch': 4549, 'dokter': 972, 'sejati': 3697, 'mengerti': 2352, 'sales': 3618, 'indonesia': 1420, 'malu': 2120, 'dir': 913, 'sobat': 3864, 'bumn': 580, 'bentuk': 350, 'kepedulian': 1793, 'percepatan': 3073, 'penanganan': 2970, 'pusri': 3456, 'penyerahan': 3053, 'ban': 267, 'data': 748, 'pasien': 2806, 'sentra': 3745, 'universitas': 4286, 'bsi': 560, 'kampus': 1639, 'bogor': 528, 'cilebut': 651, 'hai': 1265, 'catat': 613, 'informasi': 1436, 'peningkatan': 3034, 'kematian': 1760, 'akibat': 63, 'gangguan': 1176, 'jantung': 1519, 'berusia': 463, 'melonjak': 2217, 'hingg': 1335, 'survei': 3962, 'serologi': 3774, 'berbahaya': 363, 'endemik': 1066, 'flu': 1133, 'pandemi': 2777, 'selesai': 3727, 'varian': 4340, 'covovax': 702, 'bohong': 529, 'rakyat': 3487, 'menghentikan': 2366, 'pdip': 2862, 'bonekanya': 536, 'negara': 2616, 'sukangetweet': 3927, 'puanmaharani': 3433

Figure 10. Results from tf-idf

Furthermore, in the process of classifying this data, the process of testing the SVM algorithm is carried out. and the final result of the test is the SVM algorithm has an accuracy of 98.6.

```

Classification report:
              precision    recall  f1-score   support

   negatif      0.80      0.50      0.62         8
   neutral      0.99      1.00      0.99        909
   positif      1.00      0.89      0.94         82

 accuracy              0.99              999
 macro avg      0.93      0.80      0.85              999
 weighted avg   0.99      0.99      0.99              999

Confusion Matrix:
[[ 4  4  0]
 [ 1 908  0]
 [ 0  9 73]]

```

Figure 11. classification report and confusion matrix

below is the confusion matrix table.

Table 2. Confusion matrix

Actual	Prediction		
	negative	neutral	Positive
Negative	4	4	0
Neutral	1	908	0
Positive	0	9	73

From the confusion matrix table above, the data is tested to get the recall, precision, and accuracy values. The performance metrics are calculated as in Table 3 with an accuracy value of 0.99 while the results from all precision are 0.93, while the results from all recall are 79.67.

Table 3. Model Performance

Class	Accuracy	Precision	Recall	F1-Score
Negative	99.5%	80%	50%	62%
Neutral	98.6%	99%	100%	99%
Positive	99.1%	100%	89%	94%
Average	99.07%	93.00%	79.67%	85.00%

CONCLUSION

Sentiment analysis research conducted on tweets with the topic of the covid vaccine has been completed. The SVM algorithm method used can run well. This SVM algorithm is easy to

implement for classification of sentiment analysis on comments and opinions on Twitter. By getting a high level of accuracy of 98.6%. The data used for this study was taken from twitter using the crawling method. By getting 9981 data. This data proves that 92.4% of Indonesians gave a neutral response to the topic of the Covid vaccine on the Twitter application. But 0.8% of the public gave a negative response to the covid vaccine. Meanwhile, 6.7% of the public gave a positive response to the topic of conversation about the covid vaccine in Indonesia.

It is hoped that further research can use other algorithms so that the classification can run even better. And hopefully in the future it can be used for more data.

REFERENCES

- [1] W. Athira, I. Gholissodin, and rizal setya Perdana, “Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 4704–4713, 2018, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3051>.
- [2] A. Syakuro, “Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG) Halaman Judul Skripsi Oleh : Abdan Syakuro,” *Anal. sentimen Masy. terhadap e-commerce pada media Sos. menggunakan Metod. naive bayes Classif. dengan Sel. fitur Inf. gain*, pp. 1–89, 2017, [Online]. Available: <http://etheses.uin-malang.ac.id/11706/>.
- [3] R. Sari, “Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371. <https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/7371>
- [4] S. S. Salim and J. Mayary, “Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighbor,” *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 1–17, 2020, doi: 10.35760/ik.2020.v25i1.2411. <https://ejournal.gunadarma.ac.id/index.php/infokom/article/view/2411>
- [5] N. Y. A. Faradhillah, “Analisis Sentimen Terhadap Kinerja Pelayanan Publik Di Kota Surabaya Berdasarkan Klasifikasi Komentar Di Media Sosial Dengan Menggunakan Algoritma Naïve Bayes,” 2016, [Online]. Available: <https://repository.its.ac.id/id/eprint/63202>.
- [6] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, “Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *Edu tic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779. <https://journal.trunojoyo.ac.id/edutic/article/view/8779>
- [7] H. Rachmi, “Penerapan Principal Component Analysis Dan Genetic Algorithm Pada

- Analisis Sentimen Review Pengiriman Barang Menggunakan Algoritma Support Vector Machine,” *Evolusi J. Sains dan Manaj.*, vol. 5, no. 2, pp. 97–104, 2017, doi: 10.31294/evolusi.v5i2.3130.
<https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/3130>
- [8] A. Deviyanto and M. D. R. Wahyudi, “Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01. <http://ejournal.uin-suka.ac.id/sainstek/JISKA/article/view/31-01>
- [9] A. W. Attabi, L. Muflikhah, and M. A. Fauzi, “Penerapan Analisis Sentimen untuk Menilai Suatu Produk pada Twitter Berbahasa Indonesia dengan Metode Naïve Bayes Classifier dan Information Gain,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4548–4554, 2018, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2984>. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2984>
- [10] A. Herdhianto, *Sentiment Analysis Menggunakan Naïve Bayes Classifier (NBC) Pada Tweet Tentang Zakat*. 2020. <http://repository.uinjkt.ac.id/dspace/handle/123456789/53661>