



# ANALISIS SENTIMEN TWEET VAKSIN COVID 19 SINOVAC MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*

**Bitu Parga Zen<sup>1)</sup>, Danang Wicaksana<sup>2)</sup>, Halim Alfidzar<sup>2)</sup>**

<sup>1,2,3</sup> Teknik Informatika, Institut Teknologi Telkom Purwokerto

<sup>1,2,3</sup> JL. D.I Panjaitan 128 Purwokerto, Jawa Tengah, Indonesia

Email: <sup>1</sup>bita@ittelkom-pwt.ac.id, <sup>2</sup>danangwicaksana.st@gmail.com, <sup>3</sup>18102195@ittelkom-pwt.ac.id

## **Abstract**

*In early 2020, the world was shocked by the incidence of severe infections with unknown causes, the Covid-19 virus that is currently sweeping the world which causes diseases in humans and animals. Childhood people still feel sad, afraid, angry or feel happy with the spread of this disease and there has not been a drug or vaccine that has succeeded in paralyzing this virus, so far there have been several institutions or companies that are developing the Covid-19 vaccine, one of which is Sinovac. Twitter is a social networking service website that is in great demand by internet users which is very often used by the public to provide opinions about this Sinovac vaccine. Using algorithms Support Vector Machine (SVM) is a classification method that predicts class based on a model or pattern from the results of the training process, sentiment analysis is carried out to determine the response of most of the community, mostly about this Sinovac vaccine. And got a good accuracy of 74%.*

**Keyword:** *Sentiment Analysis, Twitter, Vaccine, Sinovac, Support Vector Machine (SVM)*

## **Abstrak**

Di awal tahun 2020 dunia dikagetkan dengan kejadian infeksi berat dengan penyebab yang belum diketahui, virus Covid-19 yang sedang melanda dunia saat ini yang menyebabkan penyakit pada manusia dan hewan. Kebanyakan orang masih merasa sedih, takut, marah ataupun merasa gembira dengan masih menyebarnya penyakit ini dan belum ditemukanya obat atau vaksin yang berhasil melumpuhkan virus ini yang sejauh ini sudah ada beberapa institusi atau perusahaan yang sedang mengembangkan vaksin Covid-19 salah satunya Sinovac. Twitter merupakan situs web layanan jejaring sosial yang banyak diminati pengguna internet yang sangat sering digunakan masyarakat dalam memberikan pendapat mengenai vaksin Sinovac ini. Menggunakan algoritma *Support Vector Machine (SVM)* merupakan salah satu metode klasifikasi yang memprediksi kelas berdasarkan model atau pola dari hasil proses training, analisis sentiment dilakukan untuk mengetahui respon sebagian masyarakat sebagian besar mengenai vaksin Sinovac ini. Dan didapatkan akurasi yang baik sebesar 74%.

**Kata Kunci:** Analisis Sentimen, Twitter, Vaksin, Sinovac, *Support Vector Machine (SVM)*

## 1. PENDAHULUAN

Di awal tahun 2020 dunia dikagetkan dengan kejadian infeksi berat dengan penyebab yang belum diketahui, tepatnya dikota Wuhan, Provinsi Hubei, China di akhir tahun 2019 dengan dugaan wal hal ini terkait dengan pasar basah yang menjual kan, hewan laut dan hewan lainnya [1]. Virus Corona COVID 19 sedang melanda dunia saat ini dikarenakan corona virus merupakan keluarga besar virus yang menyebabkan penyakit pada manusia dan hewan, jika pada manusia biasanya menyebabkan penyakit infeksi saluran pernapasan, mulai flu biasa hingga penyakit yang serius seperti *Middle East Respiratory Syndrome* (MERS) dan Sindrom Pernafasan Akut Berat/ *Severe Acute Respiratory Syndrome* (SARS), hingga saat ini ada 213 negara yang terjangkit virus corona covid 19 [2].

Kebanyakan orang masih merasa sedih, takut, marah ataupun merasa gembira dengan masih menyebarnya penyakit ini dan belum ditemukanya obat atau vaksin yang berhasil melumpuhkan virus ini. Pembuatan vaksin baru pada umumnya membutuhkan waktu yang lama yaitu dapat mencapai 10 tahun, Sejauh ini sudah banyak institusi atau perusahaan yang sedang mengembangkan vaksin untuk covid-19 salah satunya yaitu vaksin Sinovac [3].

Masyarakat memberikan pendapat dan opininya melalui berbagai media, salah satu media yang efektif untuk menampung opini mengenai vaksin Sinovac ini adalah Twitter. Twitter merupakan situs web layanan jejaring sosial yang banyak diminati pengguna internet sebagai media komunikasi dan mendapatkan informasi yang berupa pertanyaan, opini atau komentar, baik yang bersifat positif maupun negative [4]. Namun masih belum banyak aplikasi dan metode analisa sentimen yang dikembangkan untuk menganalisis tweet vaksin Sinovac ini. Penelitian analisis sentimen ini dilakukan untuk mengetahui sentimen publik mengenai sesuatu dengan menggunakan pendekatan dalam *machine learning* yang dikenal dengan nama *Support Vector Machine* [5].

*Support Vector Machine* (SVM) merupakan salah satu metode klasifikasi dengan menggunakan *machine learning* (*supervised learning*) yang memprediksi kelas berdasarkan model atau pola dari hasil proses training [6]. Oleh karena itu, dalam penelitian ini digunakan metode SVM yang diharapkan sesuai dengan permasalahan. Dengan dilakukannya analisis sentimen ini, diharapkan dapat mengetahui respon sebagian besar masyarakat terhadap Vaksin Sinovac ini dan dapat menggiring masyarakat untuk tetap berfikir positif sampai pandemik ini selesai.

## 2. METODE PENELITIAN

Metode penelitian yang dilakukan dalam penelitian ini terbagi menjadi beberapa tahapan, berikut ini adalah penjelasannya.

### 2.1 Analisis Sentimen

Analisis Sentimen, yang disebut juga dengan *opinion mining* merupakan salah satu cabang ilmu dari data mining yang bertujuan untuk menganalisis, memahami, mengolah, dan mengekstrak data tekstual yang berupa opini terhadap entitas seperti produk, organisasi, individu, dan topik tertentu [7]. Analisis ini digunakan untuk mendapatkan suatu informasi tertentu dari suatu kumpulan data yang ada.

### 2.2 Pengumpulan Data

Pengumpulan data merupakan salah satu bagian yang penting untuk proses analisis sentiment. Pada kegiatan ini, peneliti mengumpulkan data mengenai komentar tentang vaksin sinovac dari media sosial Twitter. *Dataset* yang digunakan sebagai input dari sistem ini yaitu *dataset* mengenai komentar

tentang vaksin sinovac yang sudah dilabeli dengan jenis respon masyarakat. Jenis respon yang digunakan yaitu *joy, sad, fear, anger*. *Dataset* yang dikumpulkan sebanyak 880 *record* data komentar Twitter yang telah diberi label *joy, sad, fear, anger* yang kemudian di klasifikasikan menjadi 2 kelas yaitu kelas normal (*joy, fear*) dan kelas unnormal (*sad, anger*).

### 2.3 Support Vector Machine (SVM)

Algoritma *Support Vector Machine* (SVM) merupakan metode yang dapat menyelesaikan permasalahan secara linear maupun non-linear sehingga bisa digunakan untuk kasus klasifikasi data [6]. Kelebihan dari algoritma *Support Vector Machine* adalah dari kemampuannya untuk menerapkan pemisah linear pada input data non-linear berdimensi tinggi, dan ini diperoleh dengan menggunakan fungsi kernel yang dipilih dan diterapkan berdasarkan karakteristik data. Banyak peneliti telah melaporkan bahwa *Support Vector Machine* metode yang paling akurat untuk teks klasifikasi [8]. Dalam pemodelan klasifikasi *Support Vector Machine* memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik klasifikasi lainnya. *Support Vector Machine* digunakan untuk mencari *hyperlane* terbaik dengan memaksimalkan jarak antar kelas. *Hyperlane* adalah sebuah fungsi yang dapat digunakan untuk pemisah antar kelas.

### 2.4 Python

Langkah-langkah pemrograman menggunakan Python, diawali dengan persiapan *dataset* yang telah dikumpulkan. Selanjutnya membuat file baru, kemudian memanggil *library-library* yang diperlukan dalam pembuatan program seperti: a) *Pandas*, merupakan *library* yang digunakan untuk membaca berbagai format data seperti file *.txt, .csv*, atau lainnya. b) *Sklearn*, merupakan *library* untuk berbagai metode dan algoritma yang digunakan untuk membuat *machine learning* [9].

### 2.5 TF-IDF

TF-IDF adalah suatu metode algoritma yang berguna untuk menghitung bobot setiap kata yang umum digunakan. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap kata di setiap dokumen. [10]. Dimana sebelumnya dihitung terlebih dahulu pada *Term Frequency* (TF), yaitu frekuensi kata yang ada di kemunculan istilah di setiap dokumen. Kemudian dihitung *Inverse Document Frequency* (IDF) adalah bobot dari suatu *term* nilai yang dihitung dari frekuensi suatu istilah yang muncul dalam beberapa dokumen [11].

### 2.6 Evaluasi

Tahap evaluasi dilakukan untuk mengetahui keakuratan dari pemodelan yang telah diterapkan pada data training.

## 3. HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai proses yang dijalankan dalam penelitian ini. Ada beberapa langkah yang diterapkan pada penelitian ini.

### 3.1 Preprocessing Data

Setelah data terkumpul maka tahapan selanjutnya yaitu memasukan data dan membuat sistem klasifikasi untuk mendeteksi respon masyarakat. Untuk memutuskan respon masyarakat dengan menggunakan metode *Support Vector Machine* (SVM) yaitu dengan cara mengklasifikasi data yang diawali dengan melakukan pelatihan data. Proses training ini dilakukan agar model yang dibuat dapat mengenali objek yang diinginkan. Kemudian dilakukan data *test* untuk mengetahui seberapa besar akurasi yang didapatkan. Pembagian data *training* dan data *test* yang digunakan pada penelitian ini sebanyak 880 data. Data *training* yang digunakan sebanyak 616 sedangkan data *test* yang digunakan sebanyak 264.

**Tabel 1.** Pembagian Data Pelatihan

Klasifikasi Respon	Jumlah Data Training	Jumlah Data Test
Normal	444	179
Unnormal	172	85
Jumlah	616	264

### 3.2 Pembuatan Model

Pada pembuatan model, proses yang dilakukan terdiri dari pembuatan data *training* sebesar 70% dan data *testing* sebesar 30%. Algoritma klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM). Penelitian ini menggunakan *tools Jupyter Notebook* dengan Python versi 3.8.2.

Tahapan pertama yang dilakukan adalah mengimport *pandas* kemudian dilanjutkan dengan mengimport file *training* dan juga *testing* serta memperlihatkan data sample sebanyak 5 seperti yang dilakukan pada gambar 1. Tahapan kedua yaitu melakukan *TF-IDF Vectorizer* dengan memanfaatkan *sklearn feature, min\_df* yang digunakan sebanyak 5 dan *max\_df* yang digunakan sebanyak 0.8 dapat dilihat di gambar 2. Tahapan ketiga yaitu melakukan klasifikasi dan juga evaluasi model dengan memanfaatkan *sklearn* seperti yang dilakukan pada gambar 3.

```
In [43]: import pandas as pd
In [44]: trainData = pd.read_csv("D:\Temu Kembali Informasi\sinovac2.csv")
In [45]: testData = pd.read_csv("D:\Temu Kembali Informasi\sinovac22.csv")
In [46]: trainData.sample(frac=1).head() # shuffle the df and pick first 5
Out[46]:
```

	Text	LABEL
325	RT @denni_sauya: Jateng Gelar Vaksinasi Covid ...	Normal
357	RT @kianinara1: Provinsi Jateng kembali mengge...	Unnormal
289	RT @G3__nakula: Ada beberapa alasan mengapa pe...	Normal
80	RT @humassijunjung: Vaksin covid19 sinovac yan...	Unnormal
447	RT @kianinara1: Walkot Malang, Sutaji memasib...	Normal

**Gambar 1.** Import data dan menampilkan data

```
In [47]: from sklearn.feature_extraction.text import TfidfVectorizer
# Create feature vectors
vectorizer = TfidfVectorizer(min_df = 5,
                             max_df = 0.8,
                             sublinear_tf = True,
                             use_idf = True)
train_vectors = vectorizer.fit_transform(trainData['Text'])
test_vectors = vectorizer.transform(testData['Text'])
```

**Gambar 2.** TF-IDF Vectorizer

```
In [48]: import time
from sklearn import svm
from sklearn.metrics import accuracy_score, f1_score, precision_score, r
# Perform classification with SVM, kernel=Linear
classifier_linear = svm.SVC(kernel='linear')
t0 = time.time()
classifier_linear.fit(train_vectors, trainData['LABEL'])
t1 = time.time()
prediction_linear = classifier_linear.predict(test_vectors)
t2 = time.time()
time_linear_train = t1-t0
time_linear_predict = t2-t1
# results
print("Training time: %fs; Prediction time: %fs" % (time_linear_train, t
report = classification_report(testData['LABEL'], prediction_linear)
print(report)
<
>
```

```
Training time: 0.097883s; Prediction time: 0.035174s
precision    recall  f1-score   support

   Normal    0.77    0.88    0.82     179
  Unnormal    0.63    0.45    0.52      85

 accuracy
macro avg   0.70    0.66    0.67     264
weighted avg 0.73    0.74    0.72     264
```

**Gambar 3. Pembuatan Model**

Hasil akurasi yang didapatkan dari model *SVM* dengan mengklasifikasikan dataset tweet vaksin sinovac sebesar 74%. Tahapan selanjutnya adalah mereview tweet sebanyak 5 kali, seperti yang dilakukan gambar 4.

```
In [13]: review = "eeruntungnya indonesia punya ransapin yang medat..masiua vaksin..asparem
review_vector = vectorizer.transform([review]) # vectorizing
print(classifier_linear.predict(review_vector))
<
>
['Unnormal']
```

```
In [15]: review = ""Sorenooni berlebihan vaksinasi masal Covid-19, yang memperlakukan vaksin
review_vector = vectorizer.transform([review]) # vectorizing
print(classifier_linear.predict(review_vector))
<
>
['Normal']
```

```
In [19]: hasibuan, pulang2 ngecek hp ternyata dapat slot vaksin sinovac besok di posyandu
<
>
['Normal']
```

```
In [16]: review = ""Efikasi Vaksin Covid-19 Sinovac Sebesar 65.3 Persen, Apa Maksudnya?
??
"""
review_vector = vectorizer.transform([review]) # vectorizing
print(classifier_linear.predict(review_vector))
<
>
['Unnormal']
```

```
In [20]: review = ""Zhao Zhenhong adalah yg menciptakan VAKSIN SINOVAc usia 53 tahun men
Saking capeknya..""
review_vector = vectorizer.transform([review]) # vectorizing
print(classifier_linear.predict(review_vector))
<
>
['Unnormal']
```

**Gambar 4. Prediksi Teks**

Hasil dari *review* yang dilakukan dapat memprediksi dengan benar sebanyak 3 dari 5 teks, berikut merupakan tabel *review*.

**Tabel 2. Tabel Review**

Review		
Teks	Actual	Predicted

<p>Beruntungnya Indonesia punya Pemimpin yang Hebat..Walau vaksin diperebutkan banyak negara Indonesia mampu mendapatkan berbagai macam vaksin, wajar jika warga Malaysia cemburu dengan warga Indonesia...Setelah Sinovac, kini giliran orang tua yang dapat vaksin AstraZeneca <a href="https://t.co/sPDvmNzYPL">https://t.co/sPDvmNzYPL</a></p>	Normal	Unnormal
<p>Seremoni berlebihan vaksinasi massal Covid-19, yang memperlakukan vaksin Sinovac bak dewa penyelamat</p>	Unnormal	Normal
<p>lagi hujan ketiduran enak bgt rasanya kayak minum kopi good day dibuatin nadine. Abis minum berasa dilempar ke langit ke 7 ketemu bidadari surga sebelum lanjut ikutan kloter travel haji annisa hasibuan, pulang2 ngecek hp ternyata dapat slot vaksin sinovac besok di posyandu</p>	Normal	normal
<p>Efikasi Vaksin Covid-19 Sinovac Sebesar 65.3 Persen, Apa Maksudnya??</p>	Normal	Normal
<p>Zhao Zhenhong adalah yg menciptakan VAKSIN SINOVAK usia 53 tahun meninggal dunia terkena serangan jantung. Saking capeknya...</p>	Unnormal	Unnormal

*Actual* merupakan hasil dari sentiment yang sebenarnya, sedangkan *predicted* merupakan hasil prediksi dari model yang dipakai.

#### 4. KESIMPULAN

Dengan analisis dan pembahasan diatas, dapat ditarik kesimpulan :

1. Metode *Support Vector Macine* (SVM) baik digunakan sebagai analisis sentiment pada klasifikasi.
2. Mayoritas masyarakat menanggapi vaksin sinovac secara positif.
3. Tweet bersentiment Normal lebih banyak dibandingkan dengan tweet Unnormal.
4. Dari pemodelan metode *Support Vector Macine* (SVM) menghasilkan akurasi sebesar 74%.

## REFERENCES

- [1] P. D. O. Davies, "Multi-drug resistant tuberculosis," *CPD Infect.*, vol. 3, no. 1, pp. 9–12, 2002.
- [2] R. Pakpahan and Y. Fitriani, "JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)," vol. 4, no. 2, pp. 30–36, 2020.
- [3] I. P. Sari and S. Sriwidodo, "Perkembangan Teknologi Terkini dalam Mempercepat Produksi Vaksin COVID-19," *Maj. Farmasetika*, vol. 5, no. 5, p. 204, 2020, doi: 10.24198/mfarmasetika.v5i5.28082.
- [4] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1750–1757, 2017.
- [5] N. D. Putranti and E. Winarko, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 8, no. 1, p. 91, 2014, doi: 10.22146/ijccs.3499.
- [6] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeding Eng.*, vol. 2, no. 1, pp. 1–7, 2015.
- [7] Morgan, *Sentiment Analysis and Opinion Mining*. Claypool.
- [8] I. Santoso, Windu Gata, and Atik Budi Paryanti, "Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 364–370, 2019, doi: 10.29207/resti.v3i3.1084.
- [9] E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python," vol. 7, no. 2, pp. 156–165, 2020.
- [10] W. Setianto, "Analisis Sentimen Twitter Untuk Mengetahui Dampak Dari Pandemi Covid-19 Terhadap Masyarakat Jakarta," Github, pp. 1–17, 2020, [Online]. Available: <https://github.com/Hyuto/Analisis-Sentimen-Corona-DKI-Jakarta> ABSTRAK.