

PENGUNAAN METODE SUPPORT VECTOR MACHINE UNTUK PROSES ANALISIS SENTIMEN MENGENAI TANGGAPAN MASYARAKAT TERHADAP PEMBERIAN VAKSIN COVID-19

Riharsya Salsabila dan Karmilasari

Universitas Gunadarma

Jl. Margonda Raya No. 100, Depok, Jawa Barat 16424

riharsyasalsabila@student.gunadarma.ac.id, karmila@staff.gunadarma.ac.id

ABSTRAK

Penyebaran Covid-19 yang disebabkan oleh Virus Corona 2019 berlangsung dengan cepat hingga banyak negara diseluruh dunia terjangkit wabah tersebut. Akibatnya banyak negara diseluruh dunia berlomba menciptakan vaksin guna menghentikan penyebaran Covid-19 sehingga menimbulkan beragam tanggapan masyarakat mengenai varian vaksin yang beredar. Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis data terkait dengan tanggapan masyarakat terhadap varian vaksin Covid-19 dengan metode Support Vector Machine (SVM). Tahapan penelitian ini meliputi : perencanaan, pengumpulan data, pengolahan data (terdiri dari case folding, tokenizing, filtering, dan stemming), pembobotan komentar (TF-IDF), pengelompokan/klasifikasi (menggunakan SVM) dan uji coba. Hasil akhir dari penelitian ini adalah pengguna bisa menginput kalimat secara manual untuk menentukan termasuk tanggapan positif atau negatif dengan tingkat akurasi sebesar 82% dengan hasil confusion matriks 96 True Positive, 34 False Positive, 19 True Negative, dan 151 False Negative. Penelitian ini pun menghasilkan visualisasi dalam bentuk histogram, wordcloud, dan bar diagram.

Kata Kunci : *Vaksin, Tanggapan, Positif, Negatif, dan SVM*

PENDAHULUAN

Virus Covid-19 yang pertama kali ditemukan di Wuhan, Cina pada akhir 2019 telah menjadi anacaman dan menyebar ke seluruh dunia. Pada 9 Maret 2020, *World Health Organization* (WHO) menetapkan virus covid-19 sebagai pandemi dan sejak itu semua negara di seluruh dunia berkomitmen bersama dengan melibatkan pemerintah, perusahaan bioteknologi, ilmuwan, dan akademisi untuk dapat menciptakan vaksin Covid-19 guna menangkal penyebaran virus Corona yang semakin meluas [1]. Hingga saat ini beragam vaksin untuk menangkal virus covid-19 telah berhasil diciptakan, diantaranya : Sinovac dan Sinopharm (Cina), Pfizer dan AstraZeneca (Inggris), Moderna (Amerika Serikat). Adapun vaksin yang disebarkan ke negara yang memerlukan dikirimkan dalam bentuk bahan baku (*bulk*) dan vaksin dalam bentuk jadi (*finished product*). Untuk vaksin dalam bentuk bahan baku dilakukan pemrosesan lanjutan untuk menjadi vaksin siap pakai. Perusahaan Indonesia yang ditunjuk untuk memproses vaksin bahan baku menjadi vaksin jadi adalah Biofarma.

Pada bulan Desember 2020 Indonesia mulai mengimpor vaksin Covid-19 sinovac dari perusahaan sinovac biotech Ltd. Untuk tahap pertama mendatangkan sebanyak 1,2 juta dan hingga Juni 2020, Indonesia telah mendatangkan kurang lebih 70 juta dosis vaksin Sinovac [2]. Mulai bulan Juli 2021, Indonesia mulai mendatangkan pula vaksin AstraZeneca.

Sasaran penerima vaksin diutamakan adalah pelayan kesehatan , depan yang langsung berhadapan dengan pasien covid-19. Selanjutnya adalah pelayan publik seperti TNI, Polri, guru, dosen dan pelayan publik lainnya. Sejak Juli 2021 vaksinasi menasar lansia, masyarakat umum dan anak-anak serta remaja mendapatkan gilirannya. Semua itu merupakan upaya untuk mencapai kekebalan komunal (*herd immunity*) minimal 70% atau sekitar 181 juta dari masyarakat Indonesia.

Pada saat diluncurkannya vaksinasi covid-19, muncul pro-kontra dari masyarakat perihal kehalalan vaksin, efek samping yang muncul, kecocokan tipe vaksin yang dipilih Indonesia dan sebagainya [3]. Keresahan juga muncul

terkait cara masyarakat mendapatkan vaksin dan rentang pemberian vaksin. Tanggapan tersebut wajar dikarenakan pemberian vaksin telah menjadi syarat dalam beragam aktivitas pada *era new normal* ini, seperti syarat untuk berpergian, syarat untuk masuk ke lingkungan seperti kantor, sekolah, mal, syarat mendapatkan fasilitas tertentu dan sebagainya. Berbagai tanggapan tersebut dituangkan masyarakat salah satunya melalui sosial media seperti Twitter. Berdasarkan tanggapan tersebut maka dapat ditelaah melalui pemrosesan bahasa berbasis teks atau bisa disebut analisis sentimen.

Analisis sentimen adalah proses yang bertujuan untuk menentukan isi dari dataset yang berbentuk teks (dokumen, kalimat, paragraf) bersifat positif, negatif atau netral [4]. Setiap hari, internet dibanjiri oleh miliaran data dari berbagai sumber. Opini pengguna ada yang bernilai positif, netral dan negatif, untuk itu diperlukan suatu klasifier. Terdapat banyak klasifier salah satunya yaitu *Support Vector Machine* (SVM) yang dapat mengatasi masalah klasifikasi dan regresi dengan linear maupun *non linear*.

Tulisan ini bertujuan menganalisis data terkait tanggapan masyarakat terhadap varian vaksin Covid-19 dengan metode *Support Vector Machine* (SVM).

METODE PENELITIAN

Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan suatu teknik prediksi, baik prediksi dalam kasus regresi maupun klasifikasi [5]. SVM dapat bekerja pada data non- linier dengan menggunakan pendekatan kernel pada fitur data awal himpunan data. Fungsi kernel digunakan untuk memetakan dimensi awal (dimensi yang lebih rendah) himpunan data ke dimensi baru (dimensi yang relatif lebih tinggi). Beberapa macam fungsi kernel [6] diantaranya:

1. Kernel Gaussian Radial Basic Function(RBF)

$$K(X_i, X_j) = \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right) \quad (1)$$

2. Kernel Polynomial

$$K(X_i, X_j) = ((X_i \cdot X_j) + C)^d \quad (2)$$

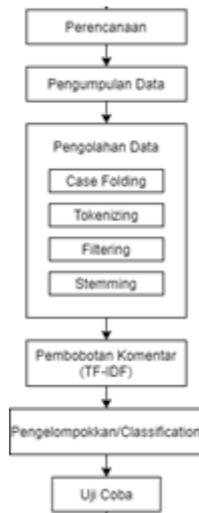
X_i dan X_j adalah pasangan dua data training. Parameter σ , c , $d > 0$ merupakan konstanta. Fungsi kernel yang digunakan untuk substitusi *dot product* di *feature space* sangat tergantung pada data karena fungsi kernel ini akan menentukan fitur baru di mana *hyperplane* akan dicari.

Karakteristik umum SVM [8] diantaranya :

1. Secara prinsip SVM adalah *linear classifier*.
2. *Pattern recognition* dilakukan dengan mentransformasikan data pada ruang input (*input space*) ke ruang yang berdimensi lebih tinggi (*feature space*) dan optimisasi dilakukan pada ruang vector yang baru tersebut.
3. Menerapkan strategi *Structural Risk Minimization* (SRM).
4. Prinsip kerja SVM pada dasarnya hanyamampu menangani klasifikasi dua kelas, namun telah dikembangkan untuk klasifikasi lebih dari dua kelas dengan adanya *pattern recognition*

Tahapan Analisis Sentimen

Dalam melakukan analisis sentimen tanggapan masyarakat terhadap pemberian vaksin covid-19 dengan menggunakan SVM dilakukan tahapan seperti tampak pada Gambar 1.



Gambar 1. Tahapan Analisis Sentimen

1. Perencanaan
Pada tahap ini dilakukan literatur studi mengenai konsep analisis sentimen serta metode yang digunakan.
2. Pengumpulan Data
Data dikumpulkan melalui sosial media yang bersumber dari twitter menggunakan sncrap yaitu library yang memperbolehkan siapa saja melakukan scraptweet tanpa memiliki API keys pribadi
3. Pengolahan Data
Pada tahap ini dilakukan beberapa proses *Text Mining* yaitu *Case Folding*, *Tokenizing*, *Filtering* dan *Stemming*. *Case Folding* mengubah semua karakter huruf pada sebuah kalimat menjadi huruf kecil dan menghilangkan karakter yang dianggap tidak valid seperti angka, tanda baca, dan Uniform Resources Locator (URL) pembersihan kata dari spesial karakter, username, serta *emoticon*. *Tokenizing* bertujuan memisahkan kata per kata. *Filtering* atau *Stopword* bertujuan menghilangkan kata yang tidak baku pada sebuah kalimat. *Stemming* bertujuan untuk menghilangkan kata berimbuhan dengan menginstall library berbahasa Indonesia, pada penelitian ini library yang digunakan adalah Sastrawi.

4. Pembobotan Komentar (TF-IDF)
Pada tahap ini dilakukan pembobotan komentar per-kata. Teknik pembobotan yang digunakan adalah TF-IDF (*Term Frequency-Inverse Document Frequency*). *Term Frequency* (TF) menyatakan jumlah berapa banyak keberadaan suatu term dalam satu dokumen dan kemudian dilogaritmikan agar mengurangi besarnya bilangan. Sedangkan *Inverse Document Frequency* (IDF) berfungsi mengurangi bobot suatu term jika kemunculannya banyak tersebar di seluruh koleksi dokumen. Bobot akhir term adalah dengan mengalikan keduanya yaitu

$$tf-idf(kata_i, komentar_j) = tf(kata_i, komentar_j) \times idf(kata_i) \quad (3)$$

5. Pengelompokan/Klasifikasi
Pada tahap ini, dilakukan pengelompokan berdasarkan data yang telah diolah melalui proses *Text Mining*. Proses analisis akan dilakukan dengan algoritma *Support Vector Machine* (SVM). Hasil analisis akan membagi data yang didapatkan menjadi positif dan negatif.
6. Uji Coba
Melakukan pengujian terhadap program yang telah dibuat dalam implementasi pada dataset di dalam Google Colab. Uji coba perhitungan akurasi klasifikasi dilakukan menggunakan *confusion matrix*, yang mengelompokkan hasil dalam nilai prediksi dan nilai aktual. Representasi hasil meliputi : *True Positive/TP* (prediksi positif yang memang bernilai positif), *True Negative/TN* (prediksi negatif yang bernilai negatif), *False Positive/FP* (prediksi positif yang ternyata bernilai negatif) dan *False Negative/FN* (prediksi negatif yang ternyata bernilai positif). Selain itu dicari juga nilai *Accuracy* (seberapa akurat model mengklasifikasikan dengan benar), nilai *Precision* (akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model) dan nilai *Recall* (keberhasilan model dalam

menemukan kembali sebuah informasi). Rumus dari *accuracy*, *precision* dan *recall* adalah sebagai berikut :

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (6)$$

Hasil akhir divisualisasikan melalui histogram, bar diagram, dan *wordcloud*.

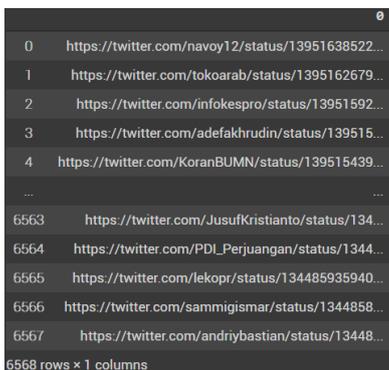
HASIL DAN PEMBAHASAN

Hasil Pengumpulan Data

Hasil pengumpulan data Twitter diperoleh 1500 tweets yang dikumpulkan dengan kurun waktu Maret 2021 – Mei 2021. Data tersebut diambil menggunakan *snscrip* yang di install melalui Command Prompt dengan *pip install snscrip*. Setelah *snscrip* berhasil ter-*install* maka masukkan *link* dari twitter yang data tersebut akan diambil dengan kondisi terdapat kata vaksin, covid, corona atau Indonesia dengan sintaks berikut.:

```
C:\Python WPy64-3880\scripts>snscrip twitter-
search "vaksin (vaksin OR covid OR corona OR
indonesia) -malaysia lang:id until: 2021-05-20 since:
2021-01-01 near:jakarta">tanggapanvaksin.csv
```

Dari data yang diambil menggunakan *snscrip* tersebut masih dalam bentuk link seperti pada Gambar 2.



Gambar 2. Link data hasil *snscrip*

Berikutnya adalah pengambilan data dari tweeep dengan sintaks :

```
import re
```

```
def fetch_tw(ids):
    list_of_tw_status=api.statuses_lookup(ids,
        tweet_mode= "extended")
    empty_data = pd.DataFrame()
    for status in list_of_tw_status:
        tweet_elem={"screen_name":
            status.user.screen_name,
            "tweet":status.full_text,
            "date":status.created_at}
        empty_data= empty_data.append(tweet_elem,
            ignore_index=True)
    empty_data.to_csv("/content/drive/MyDrive/tangg
    apanvaksinII.csv", mode="a")
```

Contoh hasil pengambilan data yang didapat dari Twitter menggunakan *tweeep* sebanyak 2 tweets dari 1500 tweets yang disimpan ke dalam file berformat *.csv* dapat dilihat pada Tabel 1.

Tabel 1. Contoh hasil pengumpulan data twitter

	date	screen_name	tweet
0	5/19/ 2021 22:48	marudiyafu	@missdapur Baru tahu ada aplikasi ini. Jadidi aplikasi ini kita juga bisa tahu ya, lokasi mana saja untuk vaksin.
1	5/19/ 2021 22:30	hariankompas	Keterbukaan penetapan harga vaksin gotong royong dinantikan. Kejelasan perhitungan besaran harga tertinggi dan tarif layanan maksimal vaksinasi gotong royong perlu disampaikan kepada publik. #Humaniora #AdadiKompas @deonisiaarlinta https://t.co/GxXksOV8yY

Hasil Pengolahan Data

Setelah data di kumpulkan tahap selanjutnya adalah pengolahan data meliputi *case folding*, *tokenizing*, *filtering* dan *stemming*. Berikut adalah sintaks dan hasil dari setiap tahap pengolahan data.

```
#CaseFolding
import re
def casefolding(tweet):
    tweet = tweet.lower()
```


Data testing diambil sebanyak 0,2 atau 20% dari data set yang ada secara acak menggunakan *random_state* bernilai 7 agar data tidak monoton dan diacak tiap 7 tweet. Di bawah ini adalah *source code* untuk *splitting data*. Sintaksnya adalah sebagai berikut :

```
#SVM Pemisahan data training dan test
x = link ['tweet']
y = link ['Label']
x_train, x_test, y_train, y_test = train_test_split
(x, y, test_size=0.2, random_state = 7)
```

Langkah berikutnya membangun model yang sesuai agar mendapatkan akurasi skor yang lebih baik. Parameter *c* digunakan untuk mencoba level signifikansi. Adapun level signifikansi yang akan di uji coba terdiri dari 6 macam yaitu 0.01, 0.05, 0.25, 0.5, 0.75 dan 1. Sintaknya sebagai berikut :

```
for c in [0.01, 0.05, 0.25, 0.5, 0.75, 1]:
    svm = LinearSVC(C=c)
    svm.fit(x_train,
            y_train)
    print('Akurasi untuk c = %s: %s' % (c,
        accuracy_score(y_test, svm.predict(x_test))))
```

Hasilnya dapat dilihat pada gambar 7.

```

> Akurasi untuk c = 0.01: 0.7933333333333333
Akurasi untuk c = 0.05: 0.8133333333333334
Akurasi untuk c = 0.25: 0.8233333333333334
Akurasi untuk c = 0.5: 0.8233333333333334
Akurasi untuk c = 0.75: 0.82
Akurasi untuk c = 1: 0.8166666666666667
```

Gambar 7. Hasil skor akurasi

Akurasi tertinggi terdapat pada 0.25 dan 0.5 maka diambil 0.5 sebagai akurasi skor utama bernilai 0.82 dari 1.00 sebagai *model final*.

Proses terakhir adalah evaluasi model menggunakan *confusion matrix* yang di kelompokkan dalam *True Positive*, *False Negative*, *False Positive* dan *True Negative*. Hasil *confusion matrix* berikut nilai *accuracy*, *precision*, *recall* dan *F1 score* dapat dilihat pada gambar 8.

Accuracy of SVM Classifier on test set : 0.82				
	precision	recall	f1-score	support
negatif	0.83	0.74	0.78	130
positif	0.82	0.89	0.85	170
accuracy			0.82	300
macro avg	0.83	0.81	0.82	300
weighted avg	0.82	0.82	0.82	300

Gambar 8. Hasil Confussion matrix

Hasil *confusion matrix* diperoleh nilai 96 *True Positive*, 34 *False Positive*, 19 *True Negative*, dan 151 *False Negative*. Sedangkan nilai akurasinya bernilai 0,82.

Uji Coba

Pada tahapan ini dilakukan pengujian data analisis sentimen mengenai tanggapan masyarakat terhadap pemberian vaksin Covid-19 dengan metode *Support Vector Machine*. Pengujian data ini dilakukan dengan menginput komentar dan mesin akan mengklasifikasikan komentar tersebut termasuk ke dalam tanggapan positif atau negatif.

Bangun model menggunakan *svm.predict* dengan contoh kalimat “efek abis vaksin ngantuk mulu” dan periksa output dari komentar tersebut sudah sesuai dengan hasil antara positif atau negatif. Jika hasilnya negatif maka prediksi sudah benar (gambar 8).

```
[ ] result = svm.predict(vectorizer.transform(["efek abis vaksin ngantuk mulu"]))
result
array(['negatif'], dtype=object)
```

Gambar 8. Hasil prediksi model negatif

Berikutnya bangun model untuk tanggapan positif seperti pada gambar 9 menggunakan sintaks yang sama dengan kalimat yang berbeda.

```
[20] result = svm.predict(vectorizer.transform(["proses vaksin hari ini berjalan lancar"]))
[21] result
array(['positif'], dtype=object)
```

Gambar 9. Hasil prediksi model positif

Visualisasi Data

Pada tahap implementasi ini ditampilkan hasil analisis sentimen secara visual di mana pada hasil visualisasi tersebut berbentuk histogram, word cloud, dan bar diagram. Sintaks untuk membuat histogram adalah dengan mengimport library *seaborn* sebagai penyedia antar-

Jumlah detail dari tweet positif maupun tweet negatif dengan sintaks berikut.

```
link['Label'].value_counts()
```

Hasilnya nilai dari tanggapan positif dan negatif adalah sebagai berikut :

```
positif    882
negatif    618
Name: Label, dtype: int64
```

Gambar 13. Jumlah tanggapan positif dan negatif

PENUTUP

Penelitian ini telah berhasil dilakukan analisis sentimen mengenai tanggapan masyarakat terhadap vaksin Covid-19 menggunakan metode Support Vector Machine di Google Colab menggunakan bahasa pemrograman python. Hasil akurasi yang dicapai dalam penelitian ini adalah 82%. Dari *confussion matrix* diperoleh informasi bahwa nilai *True Positif* 96, nilai *False Positive* 34, nilai *True Negative* 19, dan nilai *False Negative* 151. Hasil divisualisasikan dengan histogram, *word cloud* dan diagra bar. Secara umum sentimen masyarakat menanggapi pemberian vaksin covid-19, 882 orang memberikan tanggapan positif dan 618 memberikan tanggapan negatif. Hasil tersebut dapat menjadi dasar pemerintah dan pengambil kebijakan dalam menentukan langkah selanjutnya dalam penanganan covid-19 terutama dalam hal pemberian vaksin.

Analisis sentimen menggunakan metode SVM sangat bergantung pada jumlah dan keberagaman kata didalamnya sehingga sangat penting untuk memiliki sentimen SVM yang memiliki jumlah kata yang lebih banyak dan beragam. Hubungan antar kata juga perlu di analisis lebih lanjut karena hubungan antar kata dapat merubah polaritas sentimen pada teks. Pada penelitian lebih lanjut, analisis sentimen dapat dikembangkan dengan meningkatkan variasi data dan mencoba metode lain untuk melihat perbandingan agar hasil yang didapatkan lebih akurat.

DAFTAR PUSTAKA

- [1] T. Jaserivic, "Public statement for collaboration on COVID-19 Vaccine development" [online]. Available : <https://www.who.int/news/item/13-4-2020-public-statement-for-collaboration-on-covid-19-vaccine-development>, 2020.
- [2] Allina M.N, "Impor Vaksin Covid-19 Di Indonesia Perspektif Hukum Ekonomi Syariah", Institut Agama Islam Negeri Purwokerto, Purwokerto, 2021.
- [3] Tasnim, "Persepsi Masyarakat Tentang Vaksin Covid-19 Di Wilayah Provinsi Sulawesi Tenggara", Kita Menulis, 2021.
- [4] Kontopoulos E., Berberidis C., Dergiades T., and Bassiliades N, Ontology-based sentiment analysis of twitter posts, *Expert Systems with Applications*, 40(10), 4065–4074, 2013.
- [5] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu, Yogyakarta, 2007.
- [6] Prasetyo. E, *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Andi : Yogyakarta, 2012.
- [7] W. Sanders, "Prediction of Cell Penetrating Peptides by Support Vector Machines", *journal.pcbi.1002101*, 2011.
- [8] A. S. Nugroho, A. B. Witarto, dan D. Handoko, *Support Vector Machine : Teori dan Aplikasinya dalam Bioinformatika*, Kuliah Umum Ilmu Komputer, 2003