

Analisis Sentimen Vaksin COVID-19 pada Twitter Menggunakan Algoritma Support Vector Machine

Fikri Adams¹, Iin Ernawati², Nurul Chamidah³

Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. R.S Fatmawati No. 1 Pondok Labu Jakarta Selatan 12450

fikriadams@upnvj.ac.id¹, iin_ernawati@yahoo.com², nurul.chamidah@upnvj.ac.id³

Abstrak. Indonesia saat ini mengalami pandemi COVID-19, Angka penyebaran kasus COVID-19 sangat besar sehingga berdampak pada ekonomi, sosial, budaya dan kesehatan. Pemerintah Indonesia saat ini mengambil langkah kebijakan untuk mengatasi masalah penyebaran kasus COVID-19, salah satunya yaitu dengan melakukan vaksin pada masyarakat. Dengan adanya kebijakan tersebut, twitter menjadi salah satu sumber informasi yang dapat mempengaruhi pada kebijakan vaksin COVID-19 karena masih banyak masyarakat pro dan kontra terhadap vaksin COVID-19. Pada penelitian ini menganalisis sentimen menggunakan metode algoritma support vector machine dengan kernel radial basis function. Data yang digunakan berasal dari sosial media twitter dengan topik opini masyarakat terhadap vaksin covid-19. Percobaan pada tanggal 13 Januari sampai dengan 20 Januari 2021 mendapatkan nilai akurasi sebesar 82.6%. Hasil pemodelan klasifikasi diperoleh cukup baik dalam menganalisis sentimen terhadap vaksin covid-19 pada tweet positif dan negatif. Dari pelabelan data sentimen positif sebesar 251 dan sentimen negatif sebesar 237, respon masyarakat terhadap vaksin COVID-19 di sosial media twitter masih mendominasi sentimen positif karena banyak masyarakat mendukung dan mengajak adanya vaksin COVID-19.

Kata Kunci: Vaksin, COVID-19, analisis, sentimen, Support Vector Machine, twitter

1 Pendahuluan

Indonesia saat ini mengalami fase wabah *Corona Virus Disease 2019* (COVID-19). Virus ini menyerang sistem kekebalan tubuh dengan menginfeksi saluran pernapasan sehingga menyebabkan sindrome pernapasan akut berat/*Severe Acute Respiratory Syndrome* (SARS). Virus corona baru ditemukan di Wuhan Cina pada bulan Desember 2019, kemudian virus ini diberi nama *Severe Acute Respiratory Syndrome Virus Corona 2* (SARS-CoV-2). Gejala umum jika tertular pada virus ini biasanya ditandai seperti demam, batuk pilek, letih lesu, sakit tenggorokan dan gangguan pernapasan namun gejala umum tersebut bisa mengakibatkan gejala yang fatal sehingga akan berdampak kematian [1]. Penularan virus ini sangat cepat, karena virus ini menyebar melalui tetesan cairan yang dihasilkan seseorang melalui batuk dan bersin. Namun disisi-lain, virus ini menyebar melalui kontak seperti menyentuh benda yang terkontaminasi oleh virus dan berjabat tangan [1].

Angka penyebaran kasus virus ini sangat besar sehingga berdampak pada ekonomi, sosial, budaya dan kesehatan. Pemerintah Indonesia saat ini mengambil langkah kebijakan untuk mengatasi masalah penyebaran kasus COVID-19 ini, salah satunya yaitu dengan melakukan vaksinasi pada setiap masyarakat. Bergulirnya dari waktu ke waktu sebagian masyarakat banyak sekali pro dan kontra pada vaksin.

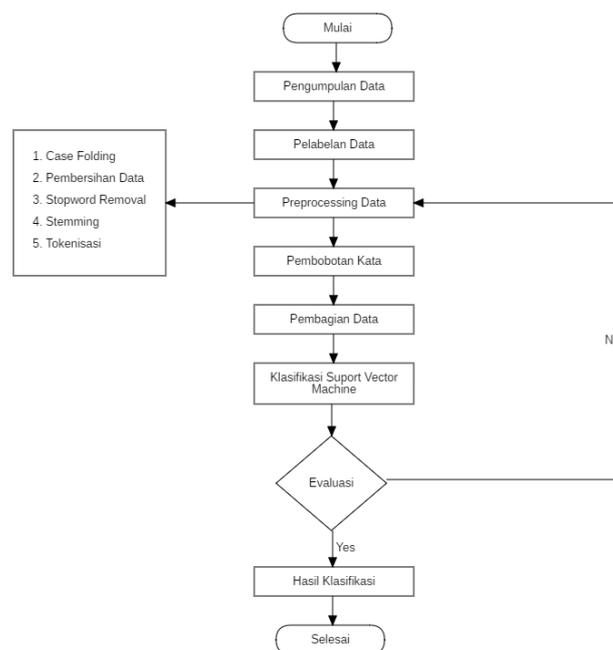
Berdasarkan dari kasus tersebut, opini masyarakat terhadap vaksin COVID-19 merupakan topik yang sangat hangat pada media sosial *twitter* karena memberikan pengaruh pada suatu informasi, sehingga masyarakat bebas dalam menuangkan pikiran dan pendapatnya. Dari informasi tersebut pemerintah harus mempertimbangkan untuk melakukan vaksinasi atau imunisasi pada masyarakat, agar dapat memberikan efek samping yang baik dan jaminan keamanan vaksin kepada masyarakat sehingga masyarakat antusias untuk mau menggunakan vaksin COVID-19. Maka dari itu, perlu mendeteksi *tweet* yang mengandung opini masyarakat terhadap vaksin COVID-19 di sosial media *twitter*.

Metode *support vector machine* merupakan algoritma klasifikasi yang salah satunya digunakan untuk mendeteksi sentimen pada *tweet*. Metode *support vector machine* juga digunakan pada penelitian yang membahas tentang review film menggunakan 1000 data yang memiliki kelas positif dan negatif. Metode klasifikasi yang digunakan adalah *support vector machine* dengan kernel *radial basis function*. Pada penelitiannya menggunakan percobaan parameter cost dan gamma dari kernel radial basis function, di mana menggunakan nilai parameter cost (C) sebesar 0.01, 0.1, 1, 10, 100, 1000, sedangkan untuk nilai gamma (γ) sebesar 0.01, 0.1, 1, 10, 100, 1000. Berdasarkan dari hasil percobaan parameter kernel *radial basis function* didapatkan parameter baik cost (C) sebesar 1000 dan gamma (γ) sebesar 1, sehingga hasil akurasi yang diperoleh adalah 84.3% [2].

Berdasarkan hal tersebut, penelitian ini menggunakan metode klasifikasi dengan algoritma *support vector machine* (SVM) untuk mendeteksi *tweet* yang mengandung opini masyarakat terhadap vaksin COVID-19 di sosial media *twitter*.

2 Metodologi Penelitian

Ada beberapa tahapan penelitian untuk analisis sentimen menggunakan algoritma *support vector machine* sebagai berikut:



Gambar 1. Merupakan *flowchart* pada tahap analisis sentimen menggunakan algoritma *support vector machine* sehingga bertujuan untuk mencapai penelitian.

2.1 Pengumpulan Data

Proses pengumpulan data dilakukan dengan cara meng-*crawling* data *tweet* menggunakan API pada sosial media *twitter*. Data yang diambil pada *tweet* berupa hashtag #vaksincovid19 dan #vaksincorona yang dimulai pada tanggal 13 januari sampai dengan 20 Januari 2021 berdasarkan pada tanggal 13 januari 2021 merupakan jadwal pertama kali vaksin COVID-19 yang dilakukan oleh bapak presiden Joko Widodo sehingga banyak masyarakat yang beropini vasin COVID-19 dimedia sosial *twitter*. Kemudian data tersebut dibagi menjadi data latih dan data

uji yang digunakan untuk sebuah pengujian dalam memprediksi model dengan menggunakan metode klasifikasi pada algoritma *support vector machine*.

2.2 Pelabelan Data

Proses pelabelan pada data *tweet* menggunakan 2 orang annotator secara manual, masing-masing kata *tweet* yang mengandung opini masyarakat terhadap vaksin COVID-19 diberikan label sentimen positif dan negatif, yang mana label positif merupakan opini pro dan label negatif merupakan opini kontra. Berikut Tabel 1 adalah contoh pelabelan data *tweet* secara manual menggunakan 2 orang *annotator*:

Tabel 1. Contoh pelabelan data manual

Data <i>Tweet</i>	<i>Annotator</i> 1	<i>Annotator</i> 2	Hasil Akhir
Vaksin covid kalo bukan @Anyaselalubenaar yang nyuntik, gua kaga mau #VaksinCovid19 #VaksinCorona	Negatif	Negatif	Negatif
Liat drama kebohongan bro @jokowi jadi takut di vaksin sinovac! #VaksinCovid19 #VaksinCorona	Negatif	Negatif	Negatif
Pemerintah Harus Keluarkan Aturan Jika Vaksin Berdampak Buruk Bagi Rakyat #VaksinCovid19 #VaksinCorona	Positif	Negatif	Negatif

Dari Tabel 1 menggunakan 2 orang *annotator* masih terdapat kesalahan pada pelabelan data *tweet*, maka dari itu perlu digunakan metode penilaian *kappa* value untuk mengukur tingkat kesepakatan antar variabel kelas. Berikut adalah rumus 1 dari persamaan *kappa* value:

Rumus Persamaan (1) dengan *Kappa* Value

$$Kappa = \frac{P_o - P_e}{1 - P_e} \quad Kappa = \frac{P_o - P_e}{1 - P_e} \quad (1)$$

Keterangan:

Kappa : Koefisien dari nilai kesepakatan di mana 0 untuk persetujuan secara kebetulan, 1 untuk persetujuan total.

$P_o P_o$: Proporsi frekuensi pengamatan.

$P_e P_e$: Peluang kesepakatan antar pengamat.

Di mana persamaan (2) adalah $P_e P_e$:

$$P_e = P(\text{positif})^2 + P(\text{negatif})^2 \quad (2)$$

Bahwa hasil menggunakan *kappa* value dapat dikriteria objektif untuk menilai nilai sebuah kesepakatan [3]. Berikut ini adalah kriteria dalam penilaian *kappa* value untuk kesepakatan, dapat dilihat pada Tabel 2:

Tabel 2. Nilai kesepakatan pada *Kappa* value

Kesepakatan	Nilai k
Rendah (<i>poor</i>)	$k < 0.00$
Kurang (<i>slight</i>)	0.00 - 0.20
Lumayan (<i>fair</i>)	0.21 – 0.40
Cukup (<i>moderate</i>)	0.41 – 0.60
Bagus (<i>substantial</i>)	0.61 – 0.80
Sangat kuat (<i>almost perfect</i>)	$k > 0.81$

2.3 Preprocessing Data

Pada tahapan digunakan pembersihan data untuk mengklasifikasi pada analisis sentimen. Dengan tujuan untuk mengurangi dan mengatasi *noise* data. Data yang dipakai pada tahap ini adalah data mentah yang masih terdapat *noise* sehingga data tersebut dilakukan *preprocessing* agar memberikan hasil yang baik pada proses pengklasifikasian. Berikut ini proses untuk *preprocessing* data :

2.3.1 Case Folding

Case folding merupakan sebuah proses untuk mengubah kata yang masih terdapat huruf kapital (*uppercase*) menjadi sebuah huruf kecil (*lowercase*). Sehingga pada tahapan *case folding* data *tweet* yang masih terdapat huruf kapital diubah menjadi huruf kecil semua [4].

2.3.2 Pembersihan Data

Pembersihan data digunakan untuk menghilangkan *value* yang tidak memiliki nilai, seperti menghapus data *noise* pada karakter tulisan, tanda baca, *link URL*, *hashtag* dan mention id *username*.

2.3.3 Stopword Removal

Stopword dapat didefinisikan sebagai proses menghilangkan kata pada dokumen yang memiliki makna tidak ada [5]. Pada proses *stopword* menghilangkan kata *tweet* seperti “yang”, “ke”, “di”, dan lain-lain.

2.3.4 Stemming

Stemming merupakan sebuah proses menghilangkan kata pada dokumen yang terdapat imbuhan *affix* seperti *prefix* dan *suffix*, sehingga menjadi kata dasar [5]. Dalam proses *stemming* menghilangkan kata imbuhan seperti di-, me-, -kan, -lah dan lain-lain.

2.3.5 Tokenisasi

Tokenisasi digunakan untuk proses pemecahan kalimat teks menjadi frasa, kata, simbol, atau elemen bermakna. Tujuan dari tokenisasi adalah eksplorasi kata-kata dalam sebuah kalimat [6]. Sehingga data *tweet* yang masih terdapat kalimat satu kemudian dipecah menjadi potongan kata.

2.4 Pembobotan Kata

Pembobotan kata digunakan untuk proses pengindeksan teks dalam menilai setiap dokumen. Tujuan pembobotan kata digunakan untuk memberikan frekuensi bobot pada kemunculan kata disuatu dokumen [6]. Metode perhitungan pembobotan kata digunakan *Term Frequency* (TF) dan *Inverse Document Frequency*. Berikut ini rumus persamaan (3) dan (4) untuk perhitungan pembobotan kata :

Rumus persamaan (3) pembobotan kata (Term Weighting)

$$W_{t,d} = TF_{t,d} * IDF_t \quad (3)$$

Keterangan :

$W_{t,d}$: bobot $t(term)$ dalam suatu dokumen

$TF_{t,d}$: frekuensi kemunculan $t(term)$ dalam dokumen d

IDF_d : *inverse document frequency*

Rumus persamaan (4) *inverse document frequency*

$$IDF_t = \log \left(\frac{N}{n_t} \right) \quad (4)$$

Keterangan :

N : Jumlah semua dokumen

n_t : jumlah dokumen yang mengandung $term t$

2.5 Pembagian Data

Pembagian data menggunakan persentase data latih 80% dan data uji 20%. Dari skenario yang telah dicoba pada penelitian ini, pembagian data 80:20 merupakan pembagian data yang baik karena setelah di evaluasi hasil model performanya memberikan nilai maksimal. Cara kerja pembagian data ini dilakukan secara *stratified random sampling* agar pembagian data harus memperhatikan keseimbangan data antara label positif dan label negatif untuk melakukan proses data latih.

2.6 Support Vector Machine

Support Vector Machine merupakan sub kategori *supervised learning* yang digunakan untuk sebuah proses klasifikasi dan regresi. Tujuan SVM (*support vector machine*) digunakan untuk memaksimalkan margin. *Margin* didefinisikan sebagai jarak antara *hyperplane* pemisah (batas keputusan) dan sampel titik pelatihan yang paling dekat dengan *hyperplane*, yang disebut *support vector* [7]. Pada tahapan ini digunakan persamaan rumus 5, 6 dan 7 sebagai berikut [8] :

Rumus persamaan (5), (6) dan (7) pada *support vector machine*

$$f(x) = w \cdot x + b \quad (5)$$

Atau

$$f(x) = \sum_{j=1}^n c_j y_i K(x, x_i) + b \quad (6)$$

Keterangan :

w : Parameter *hyperplane* dengan mencari garis lurus antara garis *hyperplane* pada titik data.

- x : Titik data.
 $c_i c_i$: Bobot pada nilai disetiap titik data.
 $K(x, x_i)K(x, x_i)$: Fungsi kernel.
 bb : Nilai bias.

Yang mana dalam tahap ini menggunakan kernel *radial basis function*. Dengan rumus sebagai berikut:

$$K(x_t, x_{tn}) = \exp\left(-\frac{\|x_t - x_{tn}\|^2}{2\sigma^2}\right) \quad (7)$$

Keterangan :

- $K(x, x_{tn})$: Nilai kernel dari data.
 \exp : Nilai eksponensial.
 x : Fitur data untuk *training*.
 c : Nilai konstanta.
 σ : Standar deviasi kernel *radial basis function*.

2.7 Evaluasi

Evaluasi digunakan untuk mengetahui performa dari model yang sudah dibuat berdasarkan hasil dari klasifikasi. [9]. Berikut adalah persamaan rumus 8, 9 dan 10 untuk evaluasi dengan menganalisa akurasi sebagai berikut [10] :

Persamaan rumus (8) akurasi

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (8)$$

Persamaan rumus (9) *recall*

$$recall = \frac{TP}{TP + FN} \quad (9)$$

Persamaan rumus (10) *specificity*

$$specificity = \frac{TN}{TN + FP} \quad (10)$$

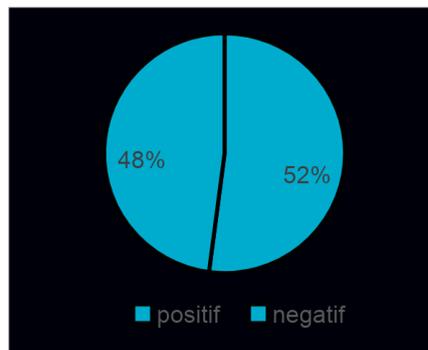
2.8 Visualisasi

Visualisasi merepresentasikan hasil penelitian sentimen positif maupun negatif pada sentimen masyarakat terhadap vaksin COVID-19 menggunakan metode *wordcloud*. *Wordcloud* merupakan representasi visual dari frekuensi kata yang sering muncul pada sumber teks. Dengan adanya visualisasi data maka dapat mempermudah proses menganalisis.

3 Hasil Pembahasan

Data yang diperoleh menggunakan metode hasil *crawling tweet* sebanyak 488 data. Kata kunci yang dicari adalah '#vaksinovid19' dan '#vaksinocorona' yang dimulai pada tanggal 13 Januari sampai dengan 20 Januari 2021, berdasarkan pada tanggal 13 Januari 2021 merupakan jadwal pertama kali vaksin COVID-19 yang dilakukan oleh Presiden Joko Widodo sehingga banyak masyarakat yang beropini vaksin COVID-19 di media sosial *twitter*.

Setelah data terkumpul dilakukan pelabelan secara manual menggunakan 2 orang *annotator*. Teknik pelabelan data menggunakan nilai kesetaraan *kappa value* untuk mengukur tingkat kesepakatan antar kelas label. Pada proses pelabelan data masing-masing kata *tweet* yang mengandung opini masyarakat terhadap vaksin COVID-19 diberikan label sentimen positif dan negatif, yang mana label positif merupakan opini pro dan label negatif merupakan opini kontra. Hasil pelabelan data yang didapat sebanyak 237 negatif dan 251 positif, sehingga dapat dipersentase menggunakan Gambar 2 sebagai berikut:



Gambar 2. Persentase Sentimen Terhadap Vaksin COVID-19

Berdasarkan hasil visualisasi pelabelan data di atas bahwa persentase opini masyarakat yang pro terhadap vaksin COVID-19 yaitu 52% sedangkan opini masyarakat yang kontra terhadap vaksin COVID-19 yaitu 48%. Persentase tersebut, sentimen masyarakat terhadap vaksin COVID-19 di sosial media *twitter* masih mendominasi opini pro karena banyak masyarakat yang beropini untuk mengajak dan mengikuti program vaksin COVID-19.

Setelah pelabelan, data *tweet* dilakukan *preprocessing* dengan tujuan untuk membersihkan, mengurangi dan mengatasi *noise* pada data sehingga ketika masuk tahap model klasifikasi harapannya dapat memberikan performa yang baik. Hasil *preprocessing* data semua dokumen *tweet* diubah menjadi kalimat yang dipecah (tokenisasi), sehingga dapat dilihat pada Tabel 3 :

Tabel 3. *Preprocessing* data

Data <i>tweet</i>
['vaksin', 'covid', 'bukan', 'nyuntik', 'gua', 'kaga', 'mau']
['liat', 'drama', 'bohong', 'bro', 'jadi', 'takut', 'vaksin', 'sinovac']
['pasca', 'suntik', 'vaksin', 'sinovac', 'ariel', 'noah', 'aku', 'cepat', 'kantuk']
['vaksin', 'covid', 'bukan', 'nyuntik', 'gua', 'kaga', 'mau']
['liat', 'drama', 'bohong', 'bro', 'jadi', 'takut', 'vaksin', 'sinovac']

Pada tahapan ini pembobotan kata menggunakan hasil *preprocessing* data pada tokenisasi sebanyak 1764 kata (*term*) dengan jumlah dokumen 488 dari data *tweet*. Perhitungan menggunakan TF-IDF (*term frequency- inverse document frequency*). Berikut adalah perhitungan TF-IDF yang dirangkum pada Tabel 4 :

Tabel 4. Pembobotan kata TF-IDF

Term	Dokumen			DF	IDF	TF-IDF		
	D1	D2	D3			D1	D2	D3
vaksin	1	1	1	3	0	0	0	0
covid	1	0	0	1	0,477	0,477	0	0
kalo	1	0	0	1	0,477	0,477	0	0

kaga	1	0	0	1	0,477	0,477	0	0
mau	1	0	0	1	0,477	0,477	0	0
liat	0	1	0	1	0,477	0	0,477	0
drama	0	1	0	1	0,477	0	0,477	0
bohong	0	1	0	1	0,477	0	0,477	0
bro	0	1	0	1	0,477	0	0,477	0
jadi	0	1	0	1	0,477	0	0,477	0
sinovac	0	1	1	2	0,176	0	0,176	0,176
pasca	0	0	1	1	0,477	0	0	0,477
suntik	0	0	1	1	0,477	0	0	0,477
ariel	0	0	1	1	0,477	0	0	0,477
noah	0	0	1	1	0,477	0	0	0,477
aku	0	0	1	1	0,477	0	0	0,477
cepat	0	0	1	1	0,477	0	0	0,477
kantuk	0	0	1	1	0,477	0	0	0,477

Dokumen yang telah diberi bobot menggunakan TF-IDF masuk pada tahap pembagian data. Pada proses pembagian data dilakukan secara *stratified random sampling*, di mana data latih sebanyak 80 % data data uji sebanyak 20 %. Berikut adalah pembagian data dapat dilihat pada Tabel 5 :

Tabel 5. Pembagian data

	Label Positif	Label Negatif	Total
Data Latih	201	189	390
Data Uji	50	48	98
Total	251	237	488

Setelah dilakukan pembagian data masuk tahap klasifikasi. Pada proses klasifikasi menggunakan fungsi kernel *radial basis function* (RBF) dari *support vector machine* (SVM). Parameter yang digunakan untuk kernel *radial basis function* (RBF) adalah *cost* (C) dan *gamma* (γ), di mana optimasi dilihat dari parameter untuk mengetahui performa model klasifikasi pada *support vector machine*. Percobaan untuk parameter *cost* (C) adalah C = [1, 10, 50, 100] dan *gamma* (γ) = [1, 2, 3, 4, 5]. Diperoleh hasil akurasi dari percobaan parameter kernel *radial basis function* dapat dilihat pada Tabel 6 :

Tabel 6. Akurasi percobaan parameter *cost* dan *gamma*

Parameter	$\gamma=1$	$\gamma=2$	$\gamma=3$	$\gamma=4$	$\gamma=5$
C=1	80.6 %	81.6 %	81.6 %	55.1 %	53.1 %
C=10	81.6 %	82.7 %	81.6 %	63.3 %	54.1 %
C=50	81.6 %	82.7 %	81.6 %	63.3 %	54.1 %

C=100	81.6 %	82.7 %	81.6 %	63.3 %	54.1 %
-------	-----------	-----------	-----------	-----------	-----------

Dari hasil proses klasifikasi yang telah diuji menggunakan parameter *cost* (C) dan *gamma* (γ) didapat parameter terbaik yaitu C = 10 dan $\gamma = 2$ dalam membuat model klasifikasi. Sehingga hasil evaluasi berupa *confusion matrix* untuk mengukur performa model klasifikasi. Berikut ini merupakan hasil evaluasi dengan menggunakan *confusion matrix* yang dirangkum pada Tabel 7 :

Tabel 7. Confusion matrix

Prediksi	Aktual	
	Positif	Negatif
Positif	40 (TP)	8 (FP)
Negatif	9 (FN)	41 (TN)

Dari Tabel 7 maka dapat dihitung menggunakan persamaan rumus (8), (9) dan (10) pada *confusion matrix* sebagai berikut :

$$accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad accuracy = \frac{TP+TN}{TP+FN+FP+TN} = \frac{40+41}{40+9+8+41} = \frac{40+41}{40+9+8+41} = 82.7\%$$

$$recall = \frac{TP}{TP+FN} \quad recall = \frac{TP}{TP+FN} = \frac{40}{40+9} = 81.6\%$$

$$specificity = \frac{TN}{TN+FP} \quad specificity = \frac{TN}{TN+FP} = \frac{41}{41+8} = 83.8\%$$

Pemodelan menggunakan kernel RBF dari *Support Vector Machine* (SVM), diperoleh bahwa 98 data uji (*testing*) mendapatkan akurasi sebesar 82.6% dari data positif dan negatif. Hasil prediksi data positif yang terdeteksi benar sebanyak 41 data dan prediksi data negatif yang terdeteksi benar sebanyak 40 data. Sedangkan data negatif diprediksi sebagai data positif. Sebanyak 8 data, dan data positif diprediksi sebagai data negatif sebanyak 9 data.

Kemudian tahap terakhir pada penelitian ini melakukan visualisasi di mana berfungsi untuk memvisualisasikan hasil sentimen positif dan negatif pada opini masyarakat terhadap vaksin COVID-19 agar mudah dipahami oleh pembaca. Visualisasi digunakan berupa *wordcloud* sentimen positif dan negatif terhadap opini masyarakat terhadap vaksin COVID-19. Hasil visualisasi *wordcloud* diambil berdasarkan kata yang sering muncul pada dokumen *tweet*.

Dari hasil penelitian yang sudah dilakukan didapatkan kesimpulan yaitu data yang diperoleh pada tanggal 13 sampai dengan 20 Januari 2021 dari media sosial *twitter* terkumpul sebanyak 488 dokumen yang dilabeli 2 orang *annotator* dengan jumlah label positif 251 dan label negatif 237. Sehingga hasil pelabelan mengenai opini masyarakat terhadap vaksin COVID-19 memiliki persentase 52% pro dan 48% kontra. Berdasarkan dari hasil persentase tersebut respon masyarakat terhadap vaksin COVID-19 masih banyak bersentimen pro dibandingkan dengan sentimen kontra. Opini yang sering dibahas masyarakat terhadap vaksin COVID-19 biasanya seperti kehalalan suatu vaksin, efektivitas vaksin, program vaksin dan dampak vaksin.

Hasil pemodelan menggunakan algoritma *support vector machine* (SVM) dari *kernel radial basis function* (RBF) dengan pembagian data latih dan data uji sebanyak 80% dan 20% menggunakan secara *stratified random sampling*, memiliki akurasi sebesar 82.7%, *recall* sebesar 81.6% dan *specificity* sebesar 83.8%.

4.2 Saran

Dari hasil penelitian yang telah dibuat ada beberapa saran untuk pengembangan penelitian, yaitu:

- Pada penelitian berikutnya diharapkan menabahkan kata *stopword* pada fungsi *python*, karena *library sastrawi* Indonesia masih belum lengkap penggunaan kata *stopword*.
- Untuk penelitian kedepannya diharapkan pelabelan data menggunakan ahli tata Bahasa tanpa harus menggunakan 2 orang *annotator*
- Perlu menambahkan normalisasi kata untuk menghindari kata singkatan dan *slang word*, sehingga dengan adanya normalisasi hasil preprocessing dapat maksimal.
- Untuk kedepannya dalam penelitian ini dapat menggunakan algoritma klasifikasi yang lain seperti *K-Nearest Neighbor*, *Neural Network*, *Naive Bayes*, *decision tree* dan lain-lain untuk perbandingan model.

Referensi

- [1] S. K. Saxena, *Coronavirus Disease 2019 (COVID-19): Epidemiology, Pathogenesis, Diagnosis, and Therapeutics*, no. 6. 2020.
- [2] Irene Mathilda Yulietha, S. Al Faraby, and Adiwijaya, "Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine," *e-Proceeding Eng.*, vol. 4, no. 3, pp. 4740–4750, 2017.
- [3] C. S. Aviva Petrie, *Medical Statistics at a Glance, 3rd Edition*. Wiley-Blackwell, 2013.
- [4] F. F. Irfani, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *JBMI (Jurnal Bisnis, Manajemen, dan Inform.)*, vol. 16, no. 3, p. 258, 2020, doi: 10.26487/jbmi.v16i3.8607.
- [5] D. Sarkar, *Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data*. 2016.
- [6] T. Verma, R. Renu, and D. Gaur, "Tokenization and Filtering Process in RapidMiner," *Int. J. Appl. Inf. Syst.*, vol. 7, no. 2, pp. 16–18, 2014, doi: 10.5120/ijais14-451139.
- [7] S. Raschka and V. Mirjalili, *Python Machine Learning Third edition*, vol. 53. Packt Publishing Ltd, 2019.
- [8] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [9] 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.
- [10] J. LING, I. P. E. N. KENCANA, and T. B. OKA, "Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *E-Jurnal Mat.*, vol. 3, no. 3, p. 92, 2014, doi: 10.24843/mtk.2014.v03.i03.p070.