

Analisis Opini Vaksin COVID-19 menggunakan SVM Berbasis PSO pada Data *Twitter*

Amrina Rosyada¹, Yuni Yamasari²

^{1,2}Jurusan Teknik Informatika, Universitas Negeri Surabaya

¹amrina.17051204037@mhs.unesa.ac.id

²yuniyamasari@unesa.ac.id

Abstrak— Infeksi yang disebabkan Sars-Cov-2 mengakibatkan pandemi penyakit pernafasan di seluruh dunia dan belum ada pengobatan yang pasti hingga saat ini. Pemberian vaksin adalah salah satu intervensi untuk mencegah penyakit ini. Di Indonesia, pemerintah mewajibkan untuk melakukan vaksinasi. Pemberian vaksin ini mendapatkan banyak tanggapan dari berbagai kalangan masyarakat. Efektivitas dan efek samping pasca pemberian vaksin serta status penggunaan darurat pada vaksin menjadi salah satu yang menjadi pro-kontra di masyarakat. Hal-hal tersebut yang melandasi penelitian ini sangat perlu untuk dilakukan. Penelitian akan memfokuskan pada analisis sentimen tentang vaksin COVID-19 dengan algoritma SVM. Namun, algoritma ini mempunyai kelemahan dalam pemilihan fitur parameter yang menyebabkan penurunan kinerja dari model yang dibangun. Penerapan fitur *Particle Swarm Optimization* (PSO) dapat ditambahkan untuk melakukan optimasi dengan pemilihan dan penyetingan fitur parameter. Berkaitan dengan dataset, penelitian ini membagi kedalam 2 sentimen yaitu positif dan negatif. Data yang digunakan berasal dari media sosial twitter dengan pencarian *tweet* yang mengandung kata vaksin COVID-19 sebanyak 646 *tweet* yang telah dipilih dengan opini positif sebanyak 57,93% dan negative sebanyak 42,07%. Pengujian dilakukan dengan 2 cara yaitu dengan *split* dan *cross-validation*. Pada pengujian *cross-validation*, penelitian ini menghasilkan akurasi yang tinggi pada $K=15$. Berturut-turut level akurasi dari SVM dan SVM-PSO adalah sebesar 76,52% dan 80,63%. Hasil akurasi tertinggi pada teknik *split* dicapai dengan rasio 1:9. Level akurasi pada SVM dan SVM-PSO, berturut-turut, adalah sebesar 83,30% dan 84,80%. Hal ini mengindikasikan bahwa penerapan PSO pada SVM mampu menaikkan level akurasi sebesar 4,11% dan 1,5% dengan pengujian *cross-validation* dan *split*, secara berturut-turut.

Kata Kunci— *Text Mining, Vaksin COVID-19, Twitter, Support Vector Machine, Particle Swarm Optimization.*

I. PENDAHULUAN

Infeksi yang disebabkan *Severe Acute Respiratory Syndrome Coronavirus* (Sars- Cov-2) sering disebut COVID-19 telah menyebabkan pandemi penyakit pernafasan diseluruh dunia [1]. Dilansir dari laman WHO, penyebaran COVID-19 telah mencapai lebih dari 97 juta kasus di seluruh dunia dan menyebabkan lebih dari 2 juta meninggal dunia. Data yang diterima pada 25 Januari 2021, di Indonesia tercatat sebanyak 989.262 kasus [2]. Hingga saat ini pengobatan yang pasti untuk penyakit COVID-19 ini belum ada. Namun, penyakit ini bisa

dicegah dengan kehadiran melalui vaksin. Pemberian vaksin adalah salah satu intervensi dalam kesehatan yang paling sukses di dunia. Bahkan dapat menyelamatkan nyawa mencapai angka 3 juta setiap tahunnya. Setiap negara berlomba-lomba dalam menemukan formula vaksin COVID-19. Pengembangan vaksin COVID-19 mencetak rekor dengan tempo waktu yang tidak lazim karena belum pernah terjadi sebelumnya. Lazimnya pengembangan secara total sebuah vaksin membutuhkan waktu lebih dari 10 tahun hingga bisa berhasil melewati uji klinis dan digunakan oleh manusia. Dengan fase normal, penelitian dan penemuan vaksin minimal mencapai rentang waktu lebih dari 2-5 tahun [3].

Di Indonesia, pemerintah mengadakan Program vaksinasi nasional untuk memutus mata rantai virus COVID-19. Lebih dari 400 juta dosis disiapkan untuk 181,5 juta jiwa dengan beberapa klaster masyarakat. Ada 4 jenis vaksin yang akan diberikan dengan perbandingan 1:1 dari 4 jenis vaksin yang akan digunakan. Dapat dinyatakan dengan jumlah diantaranya 100 juta vaksin sinovac, 100 juta vaksin *Novavax*, 100 juta vaksin *AstraZeneca* dan 100 juta vaksin *Pfizer*. Pemberian vaksin dilakukan 4 tahap dengan kelompok prioritas penerima yang telah ditentukan. Vaksin COVID-19 yang akan digunakan di Indonesia telah dinyatakan aman dan halal dengan diterbitkan EUA (*Emergency Use Authorization*) atau persetujuan penggunaan dalam kondisi darurat yang dikeluarkan oleh Badan Pengawas Obat dan Makanan (BPOM) dan fatwa dari MUI. BPOM mengizinkan penggunaan dengan beberapa kriteria efikasi dan keamanan penggunaan vaksin COVID-19 setelah beberapa fase uji klinis [4].

Pemberian vaksin di Indonesia dilakukan dengan mempertimbangkan berbagai aspek diantaranya kelayakan, efek samping dan prosedur pelaksanaan pemberian vaksin COVID-19. Penerapan kebijakan kesehatan tersebut yang dilakukan secara efektif diharapkan menghindari terjadinya kemungkinan buruk di Indonesia. Hal tersebut salah satu aspek yang menjadi *concern* dari pemerintah [5]. Sosialisasi pelaksanaan program vaksinasi nasional sangat gencar diberitakan diberbagai media massa. Pemerintah mewajibkan

masyarakat untuk melakukan vaksinasi COVID-19 terutama kelompok prioritas penerima vaksin. Kelompok yang diprioritaskan masyarakat Indonesia berusia 18 tahun keatas. Bahkan, bagi masyarakat yang menolak pemberian vaksin akan diberi sanksi. Aturan tersebut dituangkan dalam Perpres 14/2021 pasal 13A tentang sanksi bagi penerima sasaran yang memenuhi kriteria untuk menerima vaksin [6]. Hal ini mendapatkan banyak tanggapan dari berbagai kalangan masyarakat. Tanggapan tersebut terbagi menjadi 2 yaitu berupa tanggapan positif dan negatif di masyarakat. Faktor-faktor yang mendapatkan respon dari masyarakat diantaranya keraguan masyarakat terhadap status keamanan dalam penggunaan vaksin karena status vaksin hanya dapat digunakan dalam kondisi darurat dan bahan baku pembuatan vaksin yang dipertanyakan kehalalannya oleh masyarakat. Selain itu, keefektifan vaksin dan efek samping pasca pemberian vaksin mendapatkan sorotan di kalangan masyarakat. Mereka berbondong-bondong memberikan opininya di media sosial. Salah satu media sosial yang dijadikan ajang mencurahkan berbagai pendapat masyarakat yaitu media sosial *twitter*.

Pemanfaatan pengambilan data untuk survei menggunakan media sosial *twitter* adalah salah satu terobosan alternatif di era global saat ini. Mengingat bahwa, Indonesia adalah salah satu negara yang mengalami peningkatan penggunaan internet ditandai dengan peningkatan penggunaan gadget di kehidupan sehari-hari yang hampir sebagian waktunya dihabiskan untuk bermedia sosial [7]. *Twitter* merupakan salah satu media sosial yang sangat diminati di Indonesia. Sehingga Indonesia tercatat menduduki 5 peringkat teratas pengguna media sosial *twitter* yang mencapai angka 19.5 juta pengguna. Indonesia memiliki pertumbuhan pengguna *tweet* aktif harian yang sangat besar [8]. *Twitter* menghimpun Isu-isu dan pendapat yang sedang populer. Melalui fitur *hashtag* para pengguna *twitter* dapat mengetahui topik yang sedang dibahas secara *realtime*. Tidak jarang jika *twitter* memicu suatu berita muncul karena *headline* tidak jarang di media online dibahas lebih dulu di *twitter*. Faktor ini yang dapat dijadikan acuan untuk melakukan analisis opini dengan pengolahan data dari cuitan-cuitan di *twitter*.

Analisis opini merupakan metode untuk memperoleh data dari sentimen masyarakat yang berasal berbagai *platform* yang tersedia di internet. Analisis opini lebih sering dikenal dengan nama lain analisis sentimen. Opini merupakan pendapat atau pikiran dari seseorang. Semakin banyaknya opini yang disampaikan maka akan membentuk gambaran persepsi masyarakat. Analisis opini dapat bernilai positif dan negatif [9]. Metode *machine learning* yang lazim digunakan seperti SVM, Naïve Bayes dan *Random Forest* pada penerapan analisis suatu sentimen. Metode-metode tersebut diujikan dengan dataset yang sama menghasilkan akurasi berturut-turut dengan

persentase akurasi 97%,90.20% dan 88% [10]. Salah satu penelitian yang memanfaatkan media *Twitter* dalam analisis sentimen dengan metode *machine learning* adalah Buntoro, dalam penelitiannya digunakan data studi kasus *hatespeech* dengan menggunakan 2 algoritma yaitu SVM dan NBC. Hasil akurasi yang diperoleh dari SVM memiliki akurasi tertinggi yaitu 85,7% [11]. Sementara itu, pada tahun 2018, Tuheru dan Iriana melakukan analisis sentimen terkait PLN cabang ambon menggunakan metode SVM dan NBC. Penelitian tersebut menghasilkan akurasi 81,67 % dan 67,20% [12]. Pada tahun 2021, Sayam Majumder melakukan analisis sentimen studi kasus reaksi selama lockdown dengan menggunakan metode SVM dan *logistic regression*. Penelitian tersebut menghasilkan 92,5% dan 87.75% [13].

Pada beberapa penelitian sebelumnya, SVM merupakan salah satu algoritma dalam *machine learning* yang memperoleh akurasi tertinggi dari beberapa studi kasus yang berbeda. Untuk itu, penulis akan menggunakan metode SVM untuk diterapkan pada *tweet* mengenai vaksin COVID-19 ini. Klasifikasi kelas sentimen yang dieksplorasi dari media sosial *Twitter*. *Support Vector Machine* (SVM) sendiri merupakan metode terbaik diantara beberapa metode lainnya karena mampu mengkomputasi data dengan dimensi tinggi sehingga tingkat akurasi yang dihasilkan lebih baik [14]. SVM memiliki keunggulan yaitu mampu mengidentifikasi yang terpisah *hyperplane* dan yang memaksimalkan *margin* antara dua kelas berbeda serta dapat memproses secara efektif koleksi data. SVM memiliki kekurangan pada pemilihan parameter atau fitur yang mempengaruhi hasil akurasi secara signifikan.

Penerapan fitur *Particle Swarm Optimization* (PSO) dapat ditambahkan untuk melengkapi kekurangan pada metode SVM dalam pemilihan fitur. Penambahan fitur PSO dapat meningkatkan akurasi metode SVM dalam analisis sentiment (Idrus et al., 2019). Penelitian yang dilakukan oleh Arifin melakukan komparasi 2 fitur seleksi antara GA dan PSO pada metode SVM diperoleh fitur PSO yang dapat meningkatkan akurasi lebih besar dibandingkan GA [15]. *Particle Swarm Optimization* (PSO) merupakan suatu teknik optimasi yang biasa digunakan untuk menerapkan dan memodifikasi beberapa parameter. Sehingga, fitur berbasis *Particle Swarm Optimization* (PSO) merupakan sinergi yang baik saat dikolaborasi dengan algoritma *Support Vector Machine* (SVM) untuk diterapkan pada analisis opini mengenai vaksin COVID-19.

Berdasarkan beberapa penelitian yang sudah dilakukan sebelumnya, penulis akan melakukan penelitian sejenis dengan menerapkan SVM yang dioptimasi dengan PSO dalam melakukan analisis opini vaksin COVID-19. Kombinasi ini diharapkan menghasilkan akurasi yang baik. Setelah itu, proses

perbandingan dilakukan antara dua metode yaitu antara metode SVM biasa dan SVM-PSO untuk mengetahui metode mana yang menghasilkan akurasi yang lebih tinggi. Selain itu, kecenderungan sentimen publik akan diketahui dalam menanggapi peristiwa yang sedang ramai di masyarakat mengenai vaksin COVID-19. Opini tersebut dikategorikan menjadi 2 yaitu menjadi kategori opini negatif dan positif berdasarkan data *twitter* yang diperoleh.

II. METODOLOGI

Penelitian ini menggunakan metode eksperimen atau percobaan dalam menyelesaikan studi kasus vaksin COVID-19. Pemecahan masalah terhadap studi kasus tersebut menggunakan metode SVM dengan penambahan PSO untuk mendapatkan akurasi terbaik dalam melakukan analisis opini cuitan di *Twitter*.

Tujuan dari penelitian ini yaitu menerapkan salah satu metode dari *data mining* klasifikasi yaitu SVM yang dikolaborasikan dengan fitur PSO untuk melakukan analisis opini mengenai vaksin COVID-19. Agar dapat mencapai tujuan yang diinginkan, penelitian ini melakukan tahapan-tahapan yang terdiri dari beberapa langkah secara terstruktur sebagaimana ditunjukkan Gbr.1. Pada penelitian ini akan melakukan eksperimen untuk mengetahui pengaruh fitur PSO pada metode SVM.

A. Pengambilan Data

Proses pengambilan data pada penelitian ini berasal dari *tweet* media sosial *twitter*, pengumpulan data dilakukan dengan cara *crawling twitter* dari unggahan pengguna media sosial *twitter* terkait isu vaksin COVID-19.

Pengambilan data dilakukan dengan menggunakan pencarian *by keyword* yang mengandung kata “Vaksin Covid-19”. Kemudian, data *tweet* tersebut diseleksi serta disimpan kedalam *file excel* untuk diolah dalam pengujian selanjutnya. Rentang yang diambil pada 1 Januari 2021 Januari hingga bulan 31 Maret 2021 dengan tipe popular. Data disimpan dalam file excel. Alur pengambilan data ditunjukkan pada Gbr.2. Data yang terkumpul dibersihkan dari data duplikasi dari penghapusan RT (‘retweet’) sehingga data yang tersisa sebanyak 656 *tweet* diproses ketahap selanjutnya.

B. Pre-Processing

Pre-Processing yaitu tahap pengolahan data awal dilakukan sebelum data diproses oleh metode atau algoritma. Tahapan pra-pengolahan data harus terlebih dulu dilakukan. Tahapan ini bertujuan untuk membersihkan *tweet* yang *noise* agar akurasi yang dihasilkan dari analisis opini akan lebih akurat. Tahap ini terdiri dari langkah-langkah sebagai berikut:

1. Labeling

Bertujuan untuk melakukan labeling koleksi data agar mudah di proses ke tahap selanjutnya. *Cleaning* Bertujuan menghapus *mention*, *emoji*, *hashtag*, URL dan simbol yang tidak relevan.

2. Case Folding

Bertujuan mengubah menjadi huruf kecil.

3. Tokenization

Proses yang bertujuan melakukan pemisahan string input setiap kata.

4. Remove stopwords

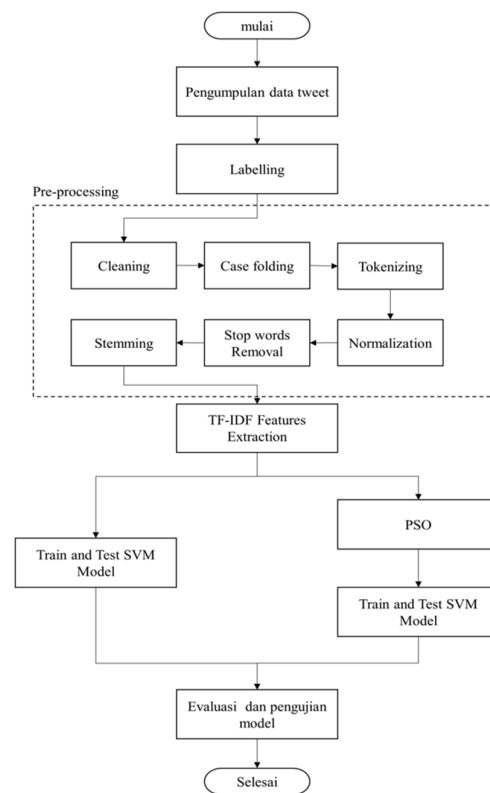
Proses menghapus *stopwords* (kata umum yang tidak memberikan informasi penting).

5. Normalization

Proses melengkapi kata-kata singkatan menjadi kata yang sempurna.

6. Stemming

Proses yang bertujuan untuk mereduksi kata ke bentuk kata dasar.

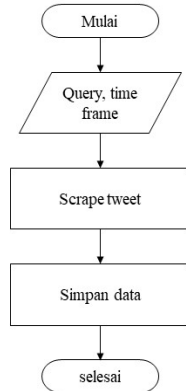


Gbr 1. Tahapan penelitian analisis opini

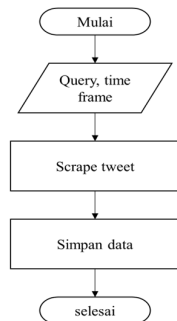
C. Pembobotan TF-IDF

Tahap pembobotan setiap kata atau *term weighting* menggunakan TF-IDF bertujuan untuk mendapatkan nilai dari kata dasar yang berhasil diekstrak, kemudian kata-kata dasar tersebut dikonversi menjadi sebuah vektor yang mewakili kata yang bersangkutan. Pembobotan dilakukan setelah tahap *pre-processing*.

Pembobotan menggunakan TF-IDF mengharuskan koleksi data diubah ke bentuk numerik untuk memudahkan dalam pembobotan. Tujuan pembobotan agar dapat meningkatkan kemampuan analisis sentiment [16]. Kerja proses pembobotan dengan TF-IDF dapat dilihat pada Gbr 3.



Gbr 2. Gambaran pengambilan data twitter



Gbr 3. Proses pembobotan

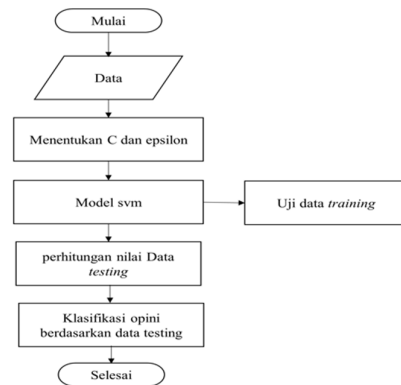
D. Support Vector Machine

Support vector machine adalah sebuah metode seleksi yang membandingkan standar seperangkat nilai diskrit yang disebut kandidat ser, dan mengambil salah satu yang memiliki akurasi klasifikasi terbaik [17]. Metode SVM berfungsi untuk membandingkan hasil sentimen positif, dan sentimen negatif. Pada SVM ada data latih yaitu data yang melatih algoritma dan data uji yaitu data untuk menguji data yang dilatih sebelumnya. SVM akan mengubah teks menjadi vektor model. *flowchart* SVM dapat dilihat pada Gbr 4.

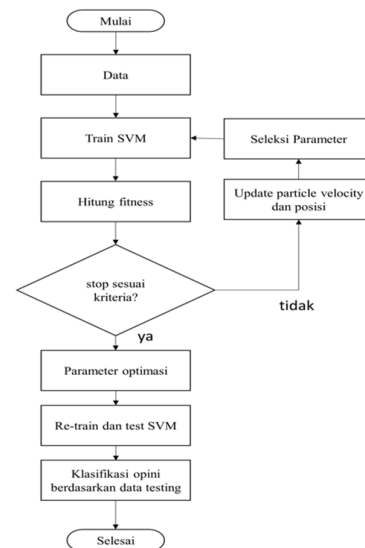
E. Particle Swarm Optimization

Particle swarm optimization adalah teknik optimasi stokastik berbasis populasi yang digunakan untuk memecahkan berbagai masalah optimasi kontinu dan diskrit. PSO berisi segerombolan partikel dan setiap partikel yang sesuai dengan solusi potensial dapat terbang dengan kecepatannya di ruang pencarian.

Dalam Penelitian ini, penggunaan PSO bertujuan untuk membantu SVM dalam melakukan optimasi sehingga dapat meningkatkan akurasi dari metode klasifikasi yang diusulkan. Proses kinerja optimasi SVM dengan PSO dapat dilihat pada Gbr 5.



Gbr 4. Flowchart proses SVM



Gbr 5 Flowchart SVM -PSO

Sistem kerja *Particle Swarm Optimization* (PSO) dapat ditunjukkan dengan *pseudocode* berikut ini:

```
def fitness_function(position):
    y_train_pred = svrRegressor.predict(X_train)
    y_test_pred = svrRegressor.predict(X_test)
    mse_test = mean_squared_error(y_test, y_test_pred)
    mse_train = mean_squared_error(y_train, y_train_pred)
    retVal= [rmse_train, rmse_test]
    optimizer = ps.single.GlobalBestPSO(n_particles=2,
    dimensions=2,options=options, bounds = bounds)
    gamma_opt, C_opt = optimizer.optimize(fitness_function,
    iters=10)
    classifier = SVC(gamma=gamma_opt, C=C_opt[0],
    probability=True)
    model_pso = classifier.fit(X_train, y_train)
    self.prob = model_pso.predict_proba(X_test)[:, 1]
    y_pred = model_pso.predict(X_test)
    prediction = model_pso.predict(validation)
    self.y_true = y_test
    if prediction[0] == 1:
        self.classLabel.setText('Tweet Classified as
        Positive Sentiment')
    else:
        self.classLabel.setText('Tweet Classified as
        Negative Sentiment')
    self.lb = np.unique(y_test)
```

- F. Eksperimen untuk menguji model dengan pengujian *cross validation* dan *split* data.
- G. Mengevaluasi hasil penelitian yang telah di uji pada metode SVM dan SVM-PSO yang disajikan dalam bentuk *confusion matrix*.

III. HASIL DAN PEMBAHASAN

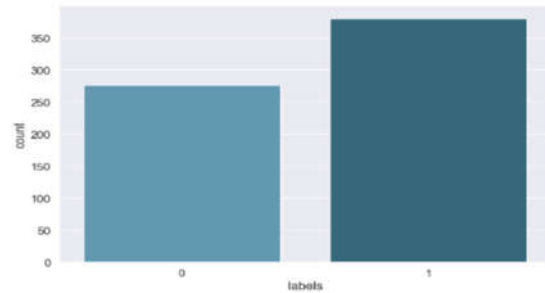
Penelitian ini melakukan analisis sentimen mengenai vaksin COVID-19 dengan dua algoritma yaitu *Support Vector Machine* (SVM) biasa dan *Support Vector Machine* (SVM) dengan fitur *Particle Swarm Optimization* (PSO). Tujuannya untuk membandingkan hasil akurasi yang disajikan dengan Tabel skenario pengujian.

Langkah awal sebelum melakukan klasifikasi analisis opini yaitu mengumpulkan data yang akan dijadikan sebagai dataset dari media sosial. Proses pengambilan data ini dinamakan *crawling*. Pada penelitian ini, *crawling* dilakukan pada media social *Twitter* dengan kata kunci pencarian “Vaksin Covid-19”. Setelah data terkumpul dilakukan tahap *pre-processing* dan seterusnya sebagaimana alur yang sudah dijelaskan pada bagian metodologi penelitian ini.

A. Hasil Pengumpulan data

Proses pengumpulan data teks *tweet* pada media sosial *Twitter* mengenai vaksin COVID-19 yang diambil dengan kata kunci tertentu ini diperoleh 656 data *tweet*. Dataset yang telah terkumpul kemudian dilakukan proses *labelling* sentimen, dimana semua *tweet* akan dikelompokkan menjadi sentimen berlabel positif dan negatif dengan menggunakan bantuan *library textblob* dan *vader*. Kedua *library* tersebut akan diambil irisan diantara keduanya kemudian diambil skor sentimen dari masing-masing *library* bernilai positif dan negatif. Skor negatif akan diberikan label sentimen “negatif” dan skor positif akan

diberikan pada label sentimen “positif”. Sehingga label sentimen negatif sebanyak 276 data dan positif sebanyak 380 data. Jika dinyatakan dengan persentase opini negatif sebesar 42,07% dan opini positif lebih banyak dibandingkan negatif sebesar 57,93%. Sentimen netral dihilangkan karena dapat mengakibatkan ambiguitas opini yang artinya belum bisa menjelaskan kearah negatif atau positif. Representasi secara visual dari perbandingan tersebut dapat dilihat pada Gbr 6.



Gbr 6. Diagram jumlah dataset vaksin COVID-19

B. Pre-processing

Berikut ini merupakan tahapan-tahapan yang dilakukan pada *pre-processing*:

1. Tahap *Cleaning*
Tahap proses *Cleaning* ini merupakan proses dimana dilakukan menghapus *mention*, *emoji*, *hashtag*, *URL* dan simbol yang tidak relevan. Hasilnya diperlihatkan kan pada Tabel I. Tahapan ini pada contoh *tweet* melakukan penghapusan titik dan koma.
2. Tahap *Case Folding*
Tahapan selanjutnya yaitu tahapan *case folding*. Pada tahapan ini dilakukan mengubah huruf menjadi kecil. Contohnya kata Jika yang diawali dengan huruf kapital pada tahap *Cleaning* diubah menjadi huruf kecil.
3. Tahap *Tokenizing*
Tahapan *tokenizing* untuk melakukan pemisahan kata. Dapat dilihat pada Tabel I.
4. Tahap *Normalization*
Tahapan *Normalization* untuk melakukan proses normalisasi kata singkatan sehingga kata yang akan diproses menjadi jelas. Contohnya dapat dilihat pada Tabel I. Disini, tahap ini melakukan perbaikan kata adlh yang dinormalisasi menjadi adalah.
5. Tahap *Stopword Removing*
Stopword Removing untuk melakukan proses penghapusan kata kata umum yang tidak memberikan informasi seperti kata sambung dan kata depan. Misalnya

kata “Jika” pada proses *Normalization* akan dihilangkan pada proses *stopword removing*.

6. Tahap *Stemming*

Tahapan terakhir dalam *pre-processing* yaitu tahapan *stemming* yang bertujuan untuk mereduksi kata ke bentuk kata dasar. Contohnya adalah kata melacak pada tahapan *stopword removing* dan kata melacak diubah menjadi kata lacak pada tahapan *stemming*.

tersebut berfungsi menampilkan kata-kata yang sering muncul dalam dataset yang diolah.

Tampilan *wordcloud* tersebut dibagi menjadi dua tampilan *wordcloud* dari hasil labelling data crawling. Berikut dibawah ini merupakan tampilan *wordcloud* dari hasil *crawling* data *tweet* tentang vaksin COVID-19 Pada Gbr. 7 merupakan tampilan *wordcloud* sentimen positif. Pada Gbr. 8 merupakan tampilan *wordcloud* sentimen negatif.

Tabel I
 HASIL PRE-PROCESSING

Tweet	
Jika ada seseorang yang mengatakan bahwa covid adlh konspirasi dari Illuminati Yahudi dan elit global vaksin, chip yang ditanam untuk melacak kita ke bumi datar, otomatis saya blacklist dari daftar calon bini.	
Tahap Pre-processing	Hasil
Cleaning	Jika ada seseorang yang mengatakan bahwa covid adlh konspirasi dari Illuminati Yahudi dan elit global vaksin chip yang ditanam untuk melacak kita ke bumi datar otomatis saya blacklist dari daftar calon bini
Case Folding	jika ada seseorang yang mengatakan bahwa covid adlh konspirasi dari illuminati yahudi dan elit global vaksin chip yang ditanam untuk melacak kita ke bumi datar otomatis saya blacklist dari daftar calon bini
Tokenizing	[jika, ada, seseorang, yang, mengatakan, bahwa, covid, adlh, konspirasi, dari, illuminati, yahudi, dan, elit, global, vaksin, chip, yang, ditanam, untuk, melacak, kita, ke, bumi, datar, otomatis, saya, blacklist, dari, daftar, calon, bini]
Normalization	[jika, ada, seseorang, yang, mengatakan, bahwa, covid, adalah, konspirasi, dari, illuminati, yahudi, dan, elit, global, vaksin, chip, yang, ditanam, untuk, melacak, kita, ke, bumi, datar, otomatis, saya, blacklist, dari, daftar, calon, bini]
Stopword Removing	[covid, konspirasi, illuminati, yahudi, elit, global, vaksin, chip, ditanam, melacak, bumi, datar, otomatis, blacklist, daftar, calon, bini]
Stemming	covid konspirasi illuminati yahudi elit global vaksin chip tanam lacak bumi datar otomatis blacklist daftar calon bini

C. Wordcloud

Hasil klasifikasi analisis sentimen data teks juga dapat divisualisasikan melalui *wordcloud*, dimana *wordcloud*



Gbr 7. Wordcloud sentimen positif



Gbr 8. Wordcloud sentimen negatif

D. Hasil Pengujian

Dengan mengkolaborasikan algoritma SVM dengan fitur PSO dimana PSO adalah fitur yang baik dalam melakukan optimasi dalam pemilihan parameter pada SVM sehingga menghasilkan akurasi SVM-PSO. Hasil akurasi dari kedua model yaitu SVM tanpa PSO dan SVM dengan fitur optimasi PSO akan dibandingkan. Pada penelitian ini sebelumnya dilakukan dua kali pengujian yaitu dengan *cross validation* dan split data dengan perbandingan data uji dan data latih yang ditentukan. Dua pengujian tersebut akan diambil pengujian mana yang lebih baik untuk penelitian ini.

Pengujian pertama yaitu, *cross validation* merupakan metode yang mirip split data namun melakukan latih ulang data

latih yang diubah menjadi data uji dan sebaliknya. Metode pembagian data dengan cross validation sering dikenal dengan nama *k-fold cross validation* yang artinya K disini sebagai jumlah sub pelatihannya. Seperti pada penelitian ini mengambil contoh dari penggunaan *10-fold* yang artinya masing-masing grup split terdiri dari data latih dan data uji, jika rata-rata hasil akurasi sama dengan 10 kali cek akurasi dengan 10 grup split yang berbeda.

Pengujian yang kedua, pembagian data uji dan latih dengan split data dengan rasio yang ditentukan. Pada penelitian ini menggunakan 5 rasio perbandingan data 1:9, 2:8, 3:7, 4:6 dan 5:5. seperti pada penelitian ini akan mengambil contoh rasio split data 1:9 yang artinya perbandingan pembagian 1 bagian atau 10 % data uji dan 9 bagian atau 90% data latih dari dataset yang digunakan secara acak.

Berikut dibawah ini merupakan tabel dan grafik perbandingan hasil akurasi dari pengujian pada metode *Support Vector Machine (SVM)* biasa tanpa PSO dan *Support Vector Machine (SVM)* yang menerapkan PSO untuk analisis opini pada penelitian ini.

1. *Pengujian dengan Cross Validation*

Pengujian pertama menggunakan pengujian cross validation dengan nilai variasi K yang digunakan hingga mendapatkan akurasi yang baik. Pada pengujian K ditentukan nilai variasi K yang ditentukan dari *2-fold* hingga *15-fold* untuk menghasilkan tingkat akurasi dari masing-masing variasi nilai k yang akan digunakan dapat dilihat dari tabel dan grafik hasil perbandingan nilai akurasi sebagai berikut:

a. *Support Vector Machine*

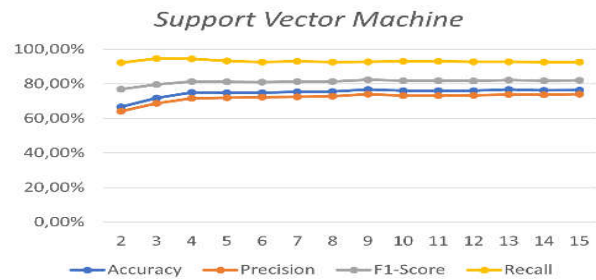
Berikut ini merupakan Tabel hasil perbandingan nilai akurasi dari masing-masing nilai *K-fold cross validation* yang digunakan pada pengujian pada metode *Support Vector Machine (SVM)* biasa tanpa PSO.

Tabel II
HASIL PENGUJIAN K-FOLD PADA SVM

K	Accuracy	Precision	F1-Score	Recall
2	66,77%	64,24%	77,05%	92,32%
3	71,95%	68,81%	79,65%	94,74%
4	75,15%	71,72%	81,48%	94,47%
5	75%	71,99%	81,23%	93,42%
6	75,00%	72,35%	81,14%	92,65%
7	75,46%	72,60%	81,48%	93,15%
8	75,61%	72,91%	81,49%	92,61%
9	76,83%	74,05%	82,36%	92,89%
10	76,07%	73,21%	81,86%	93,17%

K	Accuracy	Precision	F1-Score	Recall
11	76,07%	73,21%	81,86%	93,17%
12	76,06%	73,36%	81,84%	92,90%
13	76,83%	73,97%	82,21%	92,86%
14	76,37%	73,71%	81,94%	92,63%
15	76,52%	74,02%	82,06%	92,61%

Berdasarkan hasil pengujian pada Tabel II diatas untuk memudahkan melihat perbandingan hasil pengujian pada SVM maka akan divisualisasikan melalui grafik pada Gbr 10. Berdasarkan Gbr. 9, diketahui dari hasil pengujian model SVM dengan tanpa fitur PSO diatas bahwa nilai akurasi yang paling tinggi dengan nilai K=9 menghasilkan nilai akurasi 76,83%.



Gbr. 9. Grafik hasil pengujian K-fold pada model SVM

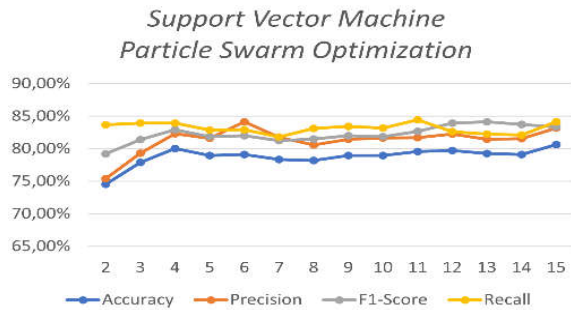
b. *Support Vector Machine dengan Particle Swarm Optimization*

Berikut ini merupakan Tabel hasil perbandingan nilai akurasi dari masing-masing nilai *K-fold cross validation* yang digunakan pada pengujian dengan menerapkan *Particle Swarm Optimization (PSO)* pada model *Support Vector Machine (SVM)*.

Tabel III
HASIL PENGUJIAN K-FOLD PADA SVM-PSO

K	Accuracy	Precision	F1-Score	Recall
2	74,54%	75,42%	79,20%	83,68%
3	77,89%	79,32%	81,42%	83,95%
4	80,03%	82,30%	82,93%	83,95%
5	78,97%	81,62%	81,86%	82,86%
6	79,12%	84,12%	82,02%	82,92%
7	78,34%	81,76%	81,22%	81,81%
8	78,20%	80,58%	81,48%	83,13%
9	78,96%	81,47%	82,01%	83,42%
10	78,95%	81,67%	81,87%	83,16%
11	79,56%	81,70%	82,67%	84,44%
12	79,73%	82,28%	83,96%	82,63%
13	79,27%	81,45%	84,16%	82,28%
14	79,11%	81,54%	83,71%	82,13%
15	80,63%	83,16%	83,32%	84,16%

Berdasarkan hasil pengujian pada Tabel III diatas untuk memudahkan melihat perbandingan hasil pengujian pada SVM maka akan divisualisasikan melalui grafik pada Gbr 10.



Gbr. 10 Grafik hasil pengujian pengujian K-fold pada model SVM-PSO

Berdasarkan Gbr. 10, diketahui dari hasil pengujian model SVM dengan fitur PSO diatas bahwa nilai akurasi yang paling tinggi dengan nilai K=15 menghasilkan nilai akurasi 80,63%.

2. Pengujian Split Latih dan Uji

Pengujian kedua pada penelitian ini menggunakan split latih dan uji dengan rasio yang telah ditentukan di atas. Pengujian ini melakukan pembagian dataset menjadi 2 bagian yaitu data latih dan data uji dengan perbandingan jumlah data uji lebih kecil dibandingkan data latih.

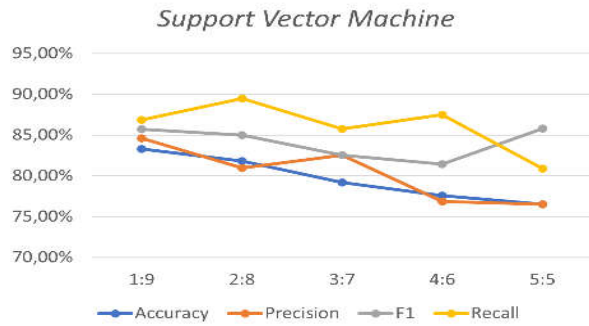
a. Support Vector Machine

Berikut ini merupakan Tabel hasil perbandingan nilai akurasi dari masing-masing rasio perbandingan data latih dan data uji yang digunakan pada pengujian dengan tanpa menerapkan PSO pada model SVM.

Tabel IV
HASIL PENGUJIAN DATA SPLIT DENGAN RASIO PADA SVM

Rasio	Accuracy	Precision	F1	Recall
1:9	83,30%	84,62%	85,71%	86,84%
2:8	81,81%	80,95%	85,00%	89,47%
3:7	79,18%	82,55%	82,55%	85,77%
4:6	77,56%	76,87%	81,45%	87,50%
5:5	76,52%	76,53%	85,79%	80,89%

Berdasarkan hasil pengujian pada Tabel IV diatas untuk memudahkan melihat perbandingan hasil pengujian pada SVM maka akan divisualisasikan melalui grafik pada Gbr. 11.



Gbr. 11. Grafik hasil pengujian data split pada model SVM

Berdasarkan Gbr.11, diketahui dari hasil pengujian model SVM tanpa fitur PSO. Dari pembagian dataset dengan metode tersebut menghasilkan akurasi yang beragam. Pada penelitian ini didapatkan nilai akurasi yang paling tinggi dengan pembagian data dengan perbandingan rasio 1:9 yang artinya dataset dibagi menjadi 1 bagian atau 10 % sebagai data uji dan 9 bagian atau 90 % sebagai data latih. Pengujian data dengan split data dengan perbandingan rasio tertentu terhadap metode SVM tanpa PSO menghasilkan akurasi yang paling tinggi sebesar 83,30%.

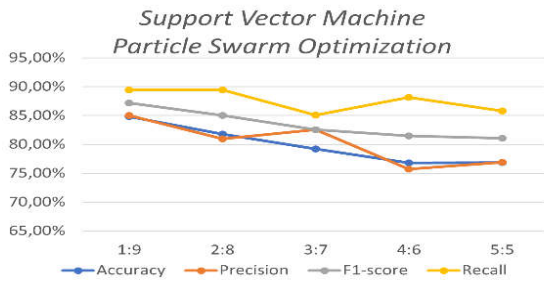
b. Support Vector Machine dengan Particle Swarm Optimization

Berikut ini merupakan tabel hasil perbandingan nilai akurasi dari masing-masing rasio perbandingan data latih dan data uji yang digunakan pada pengujian dengan menerapkan PSO pada SVM.

Tabel V
HASIL PENGUJIAN DATA SPLIT DENGAN RASIO PADA SVM-PSO

Rasio	Accuracy	Precision	F1-score	Recall
1:9	84,80%	85,00%	87,18%	89,47%
2:8	81,81%	80,95%	85,00%	89,47%
3:7	79,19%	82,55%	82,55%	85,09%
4:6	76,81%	75,71%	81,46%	88,16%
5:5	76,83%	76,89%	81,09%	85,80%

Berdasarkan hasil pengujian pada Tabel V diatas untuk memudahkan melihat perbandingan hasil pengujian pada SVM-PSO maka akan divisualisasikan melalui grafik pada Gbr. 12.



Gbr. 12 Grafik hasil pengujian data split dengan rasio pada model SVM-PSO

Pada grafik tersebut dapat dilihat bahwa pengujian dengan menggunakan split data dengan perbandingan rasio data yang diterapkan pada Metode *Support Vector Machine* (SVM) dengan fitur PSO menghasilkan akurasi beragam. Pada penelitian ini menghasilkan akurasi paling tinggi pada pembagian data 1:9 dengan data uji 1 bagian atau 10 % dan data latih 9 bagian atau 90 % dari dataset yang digunakan. Akurasi yang paling tinggi dengan menggunakan pengujian tersebut dapat menghasilkan akurasi sebesar 84,80%.

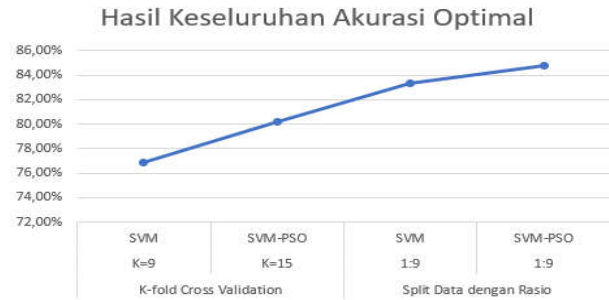
Tentunya akurasi tersebut merupakan akurasi yang paling tinggi dibandingkan dengan pengujian yang dilakukan sebelumnya yaitu dengan pengujian *k-fold cross validation* dapat dilihat pada Tabel VI. Pada Tabel tersebut menyajikan hasil akurasi optimal keseluruhan dari teknik pembagian data baik *k-fold cross validation* maupun menggunakan teknik split data dengan rasio yang diterapkan pada metode SVM tanpa PSO dan SVM dengan PSO. Berdasarkan tabel tersebut hasil akurasi yang paling tinggi pada split data dengan rasio baik yang diterapkan pada SVM dengan PSO maupun tanpa PSO.

Berdasarkan Tabel VI menyatakan pada pengujian *k-fold cross validation* menghasilkan akurasi paling tinggi dibandingkan lainnya pada pengujian ini yang diterapkan pada SVM pada nilai K=9 dengan akurasi 76,83% dan nilai K=15 pada SVM dengan PSO.

Tabel VI
HASIL KESELURUHAN AKURASI OPTIMAL

Pengujian	Classifier	Accuracy
<i>K-fold Cross Validation</i>	K=9 SVM	76,83%
	K=15 SVM-PSO	80,63%
<i>Split Data dengan Rasio</i>	1:9 SVM	83,30%
	1:9 SVM-PSO	84,80%

Berdasarkan hasil pengujian pada Tabel VI diatas untuk memudahkan melihat perbandingan hasil keseluruhan akurasi optimal maka akan divisualisasikan melalui grafik pada Gbr. 13.



Gbr. 13 Grafik hasil pengujian keseluruhan akurasi optimal

Confusion matrix hasil akurasi tertinggi pada metode SVM disajikan pada Tabel VII dan *Confusion matrix* hasil akurasi tertinggi pada metode SVM-PSO disajikan pada Tabel VIII.

Tabel VII
CONFUSION MATRIX SVM

accuracy: 83,30%			
	true negative	true positive	class precision
pred. negative	22	6	81,50%
pred. positive	5	34	84,60%
class recall	78,60%	86,80%	

Tabel tersebut menunjukkan nilai akurasi sebesar 83,30% dengan *true positive* 33 records dan *true negative* 22 records. Hal ini data yang diklasifikasikan menjadi 33 data sentimen positif dan 22 data sentimen negatif. Sedangkan data yang seharusnya bersentimen positif tetapi di prediksi negatif (*false negative*) sebesar 6 records dan data yang seharusnya bersentimen negatif namun diklasifikasikan kedalam sentimen positif (*false positive*) sebesar 5 records.

Tabel VIII
CONFUSION MATRIX SVM-PSO

accuracy: 84,80%			
	true negatif	true positif	class precision
pred. negatif	22	6	84,60%
pred. positif	4	34	85%
class recall	78,60%	89,50%	

Pada metode *Support Vector Machine* (SVM) dengan PSO menghasilkan akurasi tertinggi 84,80% yang disajikan dengan *confusion matrix* yang ditunjukkan pada Tabel VIII. Performa pada SVM-PSO dengan *true positive* 34 records dan *true negative* 22 records yang artinya 22 records yang diklasifikasikan kedalam sentimen negatif dan 34 records kedalam sentimen positif. Sedangkan data yang seharusnya

bersentimen positif tetapi di prediksi negatif (*false negative*) sebesar 6 *records* dan data yang seharusnya bersentimen negatif namun diklasifikasikan kedalam sentimen positif (*false positive*) sebesar 4 *records*.

Berdasarkan perbandingan hasil akurasi grafik yang ditampilkan pada Gbr.14 diatas menunjukkan bahwa evaluasi menggunakan SVM-PSO memiliki akurasi yang lebih tinggi dibandingkan yang tanpa PSO dengan level akurasi sekitar 84.80%.



Gbr. 14 Grafik hasil perbandingan akurasi

IV. KESIMPULAN

Penerapan teknik optimasi PSO dalam analisis opini vaksin COVID-19 dapat mengatasi kelemahan SVM. PSO mampu melakukan pemilihan parameter mana yang paling baik dengan beberapa pengkondisian untuk melakukan analisis opini. PSO yang diterapkan pada SVM membuat kinerja analisis sedikit lebih baik dibandingkan dengan SVM tanpa PSO. Hal ditunjukkan dengan hasil pengujian dengan *split data dan cross-validation* mencapai akurasi yang lebih tinggi pada SVM-PSO.

Level akurasi yang dihasilkan dalam analisis opini vaksin COVID-19 ini pada metode *Support Vector Machine (SVM)* mendapatkan akurasi sebesar 83,30%, dengan nilai *recall* sebesar 86,80%, *F1 score* sebesar 85,71% dan nilai *precision* nya sebesar 84,62%. Sedangkan SVM dengan penerapan optimasi fitur PSO akurasi dapat meningkat sebesar 1,50% dari semula yaitu sebesar 84,84%, dengan nilai *recall* sebesar 89,47%, *F1 score* sebesar 87,18% dan nilai *precision* nya sebesar 85,00%.

UCAPAN TERIMA KASIH

Puji syukur kepada Allah SWT atas rahmat Nya dan shalawat tercurahkan selalu atas junjungan Nabi Muhammad SAW. Saya ucapkan terima kasih kepada orang tua yang telah membimbing baik moril dan materil. Ucapan terima kasih juga kepada dosen pembimbing yang telah membimbing dalam penelitian ini dan semua pihak yang mendukung jalanya penelitian ini.

REFERENSI

- [1] (2021) The Johns Hopkins website [Online], <https://www.hopkinsmedicine.org/Health/Conditions/And-Diseases/Coronavirus>, tanggal akses: 25 Januari 2021.
- [2] (2020) The WHO website.[Online], <https://covid19.who.int/table>, tanggal akses:25 Januari 2021.
- [3] (2021) The World Economic Forum Website. [Online], www.weforum.org/agenda/2020/06/vaccine-development-barriers-coronavirus, tanggal akses:25 Januari 2021.
- [4] (2021) The BPOM Website. [Online], <https://www.pom.go.id/new/view/more/pers/584/Penerbitan-Persetujuan-Penggunaan-Dalam-Kondisi-Darurat-Atau-Emergency-Use-Authorization--EUA--Pertama-Untuk-Vaksin-COVID-19.html>. tanggal akses:25 Januari 2021.
- [5] D. R. Prehanto, Suryono, and Sularto, "Sentiment Analysis with LDA Algorithm for Government Policy Analysis using Twitter," *Pjate*, vol. 10, no. 8, pp. 699–711, 2020.
- [6] (2021) The Kompas Website. [Online], <https://www.kompas.com/tren/read/2021/02/19/085700865/rincian-sanksi-untuk-penolak-vaksin-covid-19-dalam-perpres-nomor-14-tahun>. Tanggal akses: 23 Maret 2021
- [7] (2019) The BBC News Indonesia Website. [Online], <https://www.bbc.com/indonesia/majalah-49630216>.,, 2019. <https://www.bbc.com/indonesia/majalah-49630216>. tanggal akses:28 januari 2021
- [8] (2012) The Kemenkominfo website. [Online], https://kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan_media. tanggal 1 maret 2021
- [9] Y. W. Syaifudin and R. A. Irawan, "Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means," *J. Inform. Polinema*, vol. 4, no. 3, p. 189, 2018, doi: 10.33795/jip.v4i3.205.
- [10] A. Sheshaayee and G. Thailambal, "Comparison of Classification Algorithms in Text Mining," 2017.
- [11] G. A. Buntoro, "Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine," *J. Din. Inform.*, vol. 5, no. September, pp. 1–21, 2016.
- [12] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.
- [13] S. Majumder, A. Aich, and S. Das, "Sentiment Analysis of People During Lockdown Period of COVID-19 Using SVM and Logistic Regression Analysis," *SSRN Electron. J.*, 2021, doi: 10.2139/ssrn.3801039.
- [14] R. Siringoringo and J. Jamaludin, "Text Mining dan Klasterisasi Sentimen Pada Ulasan Produk Toko Online," *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 1, pp. 41–48, 2019, doi: 10.34012/jutikomp.v2i1.456.
- [15] Y. T. Arifin, "Komparasi Fitur Seleksi Pada Algoritma Support Vector Machine Untuk Analisis Sentimen Review," *J. Inform. UBISI*, vol. 3, no. September, pp. 191–199, 2016.
- [16] Z. H. Deng, K. H. Luo, and H. L. Yu, "A study of supervised term weighting scheme for sentiment analysis," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3506–3513, 2014, doi: 10.1016/j.eswa.2013.10.056.
- [17] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*,

