

Analisis Sentimen Tindakan Pemerintah Indonesia Dalam Penanganan Covid-19 Menggunakan Metode Support Vector Machine

Ariesta Damayanti^{1,*}, Helda Ludya Safitri¹, Rudy Cahyadi²

¹Teknologi Informasi, Informatika, Universitas Teknologi Digital Indonesia, Bantul, Indonesia

²Politeknik Negeri Media Kreatif, Jakarta, Indonesia

Email: ^{1,*}ariesta@utdi.ac.id, ²heldaludya5@gmail.com, ³masrudyc@polimedia.ac.id

Email Penulis Korespondensi: ariesta@utdi.ac.id

Submitted: 12/12/2022; Accepted: 27/12/2022; Published: 31/12/2022

Abstrak—Corona Virus Disease 2019 (Covid-19) yang melanda dunia termasuk Indonesia sejak awal tahun 2020 merupakan wabah yang menjadi ancaman yang serius bagi kesehatan dunia. Pemerintah Indonesia melakukan berbagai tindakan untuk menangani permasalahan ini, sedangkan masyarakat dengan adanya media sosial banyak memberikan tanggapan terhadap kebijakan pemerintah tersebut. Twitter adalah salah satu media sosial yang banyak digunakan masyarakat untuk menyampaikan komentar baik berupa tanggapan, saran, hingga kritik kepada pemerintah yang berkaitan dengan penanganan Covid-19. Komentar yang muncul tersebut selanjutnya digunakan pemerintah sebagai bagian acuan dalam mengevaluasi suatu kebijakan atau tindakan yang dilakukan dalam penanganan Covid-19. Sehingga salah satu cara yang bisa digunakan dalam menangani hal tersebut adalah salah satu metode yang ada dalam domain teks mining yaitu analisis sentimen. Penelitian ini dilakukan dengan menganalisis sentimen menggunakan metode Support Vector Machine (SVM) dengan Kernel Radial Basis Function (RBF). Tweet akan diklasifikasi menjadi sentimen positif, negatif, dan netral, sehingga dapat diketahui seberapa banyak persentase dari masing-masing kategori opini. Penelitian ini menggunakan data sebanyak 600 tweet yang diperoleh dari hasil scraping menggunakan twitterscraper. Hasil dari penelitian ini adalah tingkat akurasi pelatihan sebesar 77% dalam melakukan klasifikasi sentimen positif, negatif, dan netral. Dari hasil klasifikasi data, diperoleh sebagian besar tweet terdiri dari sentimen negatif.

Kata Kunci: Covid-19; Radial Basis Function; Sentimen; Support Vector Machine; Twitter

Abstract—Corona Virus Disease 2019 (Covid-19) which has hit the world including Indonesia since the beginning of 2020 is an outbreak that has become a serious threat to world health. The Indonesian government is taking various actions to deal with this problem, while the public, with the existence of social media, has provided many responses to these government policies. Twitter is one of the social media that is widely used by the public to convey comments in the form of responses, suggestions, to criticism of the government regarding the handling of Covid-19. The comments that appear should be used by the government as part of the reference in evaluating a policy or action taken in handling Covid-19. So that one way that can be used to deal with this is one of the methods that exist in the domain of text mining, namely sentiment analysis. This research was conducted by analyzing sentiment using the Support Vector Machine (SVM) method with the Kernel Radial Basis Function (RBF). Tweets will be classified into positive, negative and neutral sentiments, so that the percentage of each opinion category can be known. This study uses data of 600 tweets obtained from the results of scraping using a Twitter scraper. The result of this study is that the training accuracy rate is 77% in classifying positive, negative, and neutral sentiments. From the results of the data classification, it was found that most of the tweets consisted of negative sentiments.

Keywords: Covid-19; Radial Basis Function; Sentiment; Support Vector Machine; Twitter

1. PENDAHULUAN

Corona Virus Disease 2019 (Covid-19) yang melanda dunia termasuk Indonesia sejak awal tahun 2020 merupakan wabah yang menjadi ancaman yang serius bagi kesehatan dunia. *World Health Organization* (WHO) menyampaikan bahwa *collaps*-nya penanganan dari fasilitas kesehatan yang ada di setiap negara dikarenakan terus naiknya angka kasus secara eksponensial dari yang kasus ringan, sedang, berat hingga kematian [1]. Untuk menangani semakin tingginya kasus Covid-19 ini beberapa negara kemudian memberlakukan kebijakan pembatasan kegiatan dan akses bagi warga negaranya.

China sebagai negara tempat awal terjangkitnya Covid-19 ini kemudian melakukan tindakan berupa penyediaan fasilitas kesehatan khusus pasien Covid-19, mengubah gedung olahraga, aula, sekolah, dan juga hotel menjadi rumah sakit sementara, melakukan *rapid-test* ataupun *polymerase chain reaction* (PCR) pada banyak warga, hingga mengimplementasikan metode mengisolasi kota (*lockdown*). Negara Korea Selatan tepatnya di wilayah Daegu melakukan proses deteksi dini melalui rapid test yang dilakukan secara massal dengan tujuan melokalisasi individu yang terpapar Covid-19. Tindakan lain yang dilakukan sebagai sebagai upaya preventif untuk meminimalkan penyebaran virus Covid-19 adalah dengan meliburkan sekolah dan kampus, dan juga melaksanakan kebijakan *lockdown* [2].

Indonesia sebagai negara di wilayah Asia Tenggara, menurut WHO menjadi salah satu dari 5 (lima) negara di wilayah Asia Tenggara yang memiliki angka kasus dan kematian tertinggi sejak pertama kali kasus Covid-19 ditemukan [3]. Kenaikan angka tersebut tentu saja menjadi perhatian pemerintah khususnya Kementerian Kesehatan terkait penanganan kasus di Indonesia. Beberapa tindakan pemerintah Indonesia yang diterapkan terkait pembatasan akses bagi seluruh kegiatan masyarakat yang dikenal dengan Pembatasan Sosial Berskala Besar

(PSBB) [4]. Penerapan kebijakan PSBB ini merupakan salah satu kebijakan yang merubah pola kehidupan masyarakat Indonesia [5]. Perubahan pola kegiatan masyarakat inilah yang kemudian mengundang berbagai pendapat dalam masyarakat. Pendapat-pendapat itu tidak saja yang dikemukakan secara langsung tapi juga yang disampaikan melalui media sosial baik berupa tanggapan, saran, hingga kritik kepada pemerintah [6]. Berdasarkan data dari *Global Digital Statistic* di *We Are Social* tahun 2020, pengguna media sosial di Indonesia khususnya Twitter mencapai lebih dari 160 juta orang [7]. Jumlah ini merupakan representasi masyarakat yang lebih memilih untuk menyampaikan pendapat atau komentarnya dengan lebih terbuka dan bisa saja anonim [8].

Komentar yang diberikan oleh masyarakat ini bisa saja memiliki kecenderungan komentar yang positif, negatif ataupun netral. Komentar yang muncul tersebut selayaknya digunakan pemerintah sebagai bagian acuan dalam evaluasi suatu kebijakan atau tindakan yang dilakukan dalam penanganan Covid-19. Sehingga salah satu cara yang bisa digunakan dalam menangani hal tersebut adalah salah satu metode yang ada dalam domain teks mining yaitu analisis sentimen. Analisis ini bertujuan untuk menentukan pendapat dan subyektifitas dari pemberi pendapat [9]. Analisis sentimen merupakan suatu bidang data mining yang berfungsi untuk menganalisis, mengolah, dan mengekstrak data tekstual pada entitas, seperti layanan, produk, individu, organisasi, peristiwa, atau masalah dan topik tertentu [10].

Analisis ini dapat mengklasifikasikan data berdasarkan sentimen yang ada di dalamnya. Secara umum sentimen yang ada tersebut dapat diklasifikasikan ke dalam dua grup yaitu sentimen positif dan negatif atau tiga grup yaitu kelas sentimen positif, negatif serta netral [11]. Secara teknis, sentimen analisis dapat dibagi menjadi empat metode pendekatan [9]: pendekatan pertama adalah pendekatan *machine learning*, dimana dalam pendekatan ini menggunakan algoritma pembelajaran untuk menyusun suatu model training data. Pendekatan kedua adalah pendekatan *lexicon-based*. Pendekatan *lexicon-based* ini melibatkan perhitungan dari polaritas sentimen tersebut berdasarkan orientasi semantik dari kata atau kalimat yang ada pada dokumen. Pendekatan *rule-based* adalah mencari opini dari kata-kata kemudian mengklasifikasikannya berdasarkan kata positif dan negatif. Sedangkan untuk pendekatan model statistik adalah pendekatan dengan menjadikan pendapat apapun sebagai aspek analisis dan peringkat.

Pada pendekatan *machine learning*, salah satu algoritma klasifikasi yang biasa digunakan pada sentimen analisis adalah Support Vector Machine (SVM). SVM digunakan pada penelitian-penelitian sebelumnya untuk analisis sentimen mengenai penggunaan transportasi *online* [12]. Penelitian tersebut menggunakan metode SVM dengan total data tweet sebesar 1852 menghasilkan dua klasifikasi sentimen yaitu sentimen positif sebesar 62% dan sentimen negatif sebesar 38%. Penelitian lain dengan menggunakan data tweet mengenai pendapat masyarakat mengenai kinerja Komisi Pemberantasan Korupsi (KPK).

Berdasarkan *crawling* data yang dilakukan menghasilkan 1890 data dan 3846 term/kata dari hasil *preprocessing*. Klasifikasi untuk sentimen dengan label negatif memberikan hasil 77%, untuk sentimen dengan label positif sebesar 8% dan sentimen dengan label netral 25% [13]. Kajian tentang analisis sentimen dengan data yang berkaitan dengan permasalahan Covid-19 juga telah dilakukan. Penelitian mengenai analisis sentimen menggunakan SVM juga dilakukan mengenai kebijakan pemerintah untuk Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) dengan data sejumlah 2000 tweet menghasilkan sentimen positif sebesar 57-60% dan sisanya berupa sentimen negatif [14].

Kebijakan pemerintah yang lain mengenai Covid-19 yaitu tentang vaksin Covid-19 juga menjadi topik dalam analisis sentimen dengan SVM. Penelitian yang dilakukan dengan data sebanyak 488 dokumen menghasilkan analisis sentimen dengan jumlah label positif 251 dan label negatif 237. Sehingga hasil pelabelan mengenai opini masyarakat terhadap vaksin COVID-19 memiliki persentase sebesar 52% pro dan 48% kontra [15].

Berdasarkan penelitian terkait, model SVM yang digunakan dapat menggunakan salah satu kernel yang ada pada SVM yaitu, fungsi linear, *polynomial*, *gaussian Radial Base Function* (RBF), dan sigmoid. Penelitian analisis sentimen terhadap opini maskapai penerbangan yang ada pada Twitter menggunakan algoritma SVM menunjukkan bahwa klasifikasi dengan fitur *Lexicon Based* menghasilkan kelas positif dan negatif dari tweet berbahasa Inggris. Penelitian dengan parameter C bernilai 10 dan learning rate 0,03 serta *Lexicon Based Features* dengan iterasi sebanyak 50 kali memberikan hasil akurasi sebesar 40%, *precision* 40%, *recall* 100%, dan *f-measure* sebesar 57,14% [16]. Hasil analisis sentimen menggunakan algoritma SVM dengan kernel RBF diyakini memiliki akurasi yang baik dalam analisis sentimen dengan mencapai akurasi di atas 80% [17].

Pada penelitian lain mengenai klasifikasi tweet yang tercampur iklan atau dengan menggunakan SVM dan Kernel Radial Basis Function (RBF) bahkan menghasilkan akurasi sebesar 99,12% [18]. Kernel RBF juga diyakini sebagai pemodelan yang adaptif dan fleksibel sehingga dapat mengurangi beban komputasi pada sistem [19]. Berdasarkan hal tersebut kemudian akan diuji mengenai kemampuan SVM dengan kernel RBF ini untuk melakukan analisis sentimen pada data komentar masyarakat Indonesia di Twitter mengenai kebijakan pemerintah dalam penanganan Covid-19.

Sehingga untuk menguji hal tersebut perlu dilakukan pemodelan yang tepat untuk memproses opini pengguna Twitter tentang kebijakan pemerintah dalam penanganan kasus Covid-19 di Indonesia. Kemudian juga diuji mengenai SVM dengan kernel RBF untuk menghasilkan klasifikasi sentimen sesuai dengan kelas yang dibentuk. Demikian pula dengan tingkat akurasi yang dihasilkan dari SVM kernel RBF ini akan memberikan hasil yang optimal.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

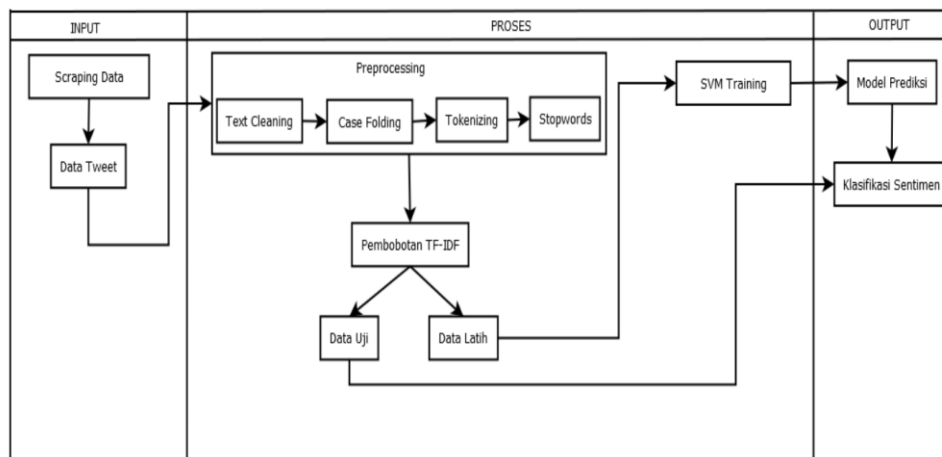
Dalam setiap proses yang ada pada text mining dalam hal ini adalah analisis sentimen, perlu digambarkan suatu mekanisme dalam menyelesaikan masalah tersebut. Blok diagram sistem merupakan suatu gambaran bagaimana sistem pada analisis sentimen ini bekerja. Gambar blok diagram sistem yang digunakan pada penelitian ini seperti ditunjukkan pada gambar 1.

Data pada penelitian ini adalah data tweet yang diambil dari Twitter. Data diambil menggunakan *scrapper twitterscraper*. Data yang diambil berupa teks mentah sehingga harus dilakukan preprocessing sehingga data dapat diolah lebih lanjut untuk menghasilkan kelas yang sesuai. Data yang digunakan merupakan tweet komentar yang diambil mulai kurun waktu Maret 2020 tentang opini publik terhadap tindakan pemerintah terkait penanganan Covid-19. Pengambilan data dilakukan melalui hashtag #viruscoronaindonesia, #menkes, #indonesiaterserah, #PSBBJakarta dan keyword ‘psbtotal’, ‘lockdownindonesia’.

2.2 Prosedur Pengambilan Data

Langkah-langkah yang dilakukan pada proses pengambilan data dengan scraping menggunakan Twitterscraper yang disediakan oleh Python adalah sebagai berikut,

1. Memasang modul Twitterscraper pada Python dengan package pip.
2. Memasukkan kata kunci berupa hashtag, batasan tanggal, dan batasan jumlah data yang ingin diambil.
3. Data yang didapat dalam bentuk file csv yang kemudian dilakukan pelabelan secara manual.



Gambar 1. Blok Diagram Sistem

2.3 Analisis dan Pemodelan

Secara lebih rinci dijelaskan bagaimana sistem ini berjalan sesuai dengan blok diagram yang ada. Untuk langkah pertama pada proses input, data tweet hasil dari scraping menggunakan twitterscraper diberikan pelabelan secara manual untuk membentuk pemodelan pelatihan. Setiap data tweet akan diberikan tiga label kelas yaitu positif, negatif, netral. Kemudian selanjutnya dilakukan tahap *preprocessing* data yang bertujuan untuk menghilangkan noise, menyeragamkan bentuk kata dan mengurangi volume kosa kata. Tahapan pada *preprocessing* data ini meliputi *text cleaning*, *case folding*, *tokenizing*, *stopwords removal*. Dari hasil preprocessing didapatkan data fitur untuk diproses pada tahap selanjutnya. Hasil *preprocessing* pada data ini ditunjukkan pada tabel 1. Setelah mendapatkan fitur hasil dari *preprocessing*, maka dilakukan pembobotan TF-IDF pada masing-masing fitur berdasarkan indeks. TF menyatakan jumlah keberadaan suatu term (t) dalam dokumen (d), DF merupakan perhitungan banyak dokumen dalam suatu term (t) yang muncul dan IDF untuk mengurangi bobot suatu term jika kemunculannya banyak tersebar diseluruh dokumen. Data hasil pembobotan TF-IDF akan diubah kedalam bentuk vektor untuk dilakukan proses SVM *training*. Hasil dari nilai bobot tf-idf menghasilkan fitur dalam bentuk vektor yang akan digunakan untuk proses pelatihan klasifikasi SVM dengan data latih dan data uji yang sudah dibagi. Proses pelatihan model atau SVM *training* dilakukan untuk memberikan pembelajaran kepada *machine learning* agar dapat melakukan prediksi terhadap data yang diujikan.

Tabel 1. Hasil *Preprocessing*

Proses	Tweet Asli	Hasil Praproses
<i>Text Cleaning</i>	Ini bukti pemerintah pusat gak koordinasi dr awal terkait wabah corona. Terlalu santai dan selalu menganggap kita zero kasus. Hrsnya dr awal semua kepala daerah dikumpulin dan diberi	Ini bukti pemerintah pusat gak koordinasi dr awal terkait wabah corona Terlalu santai dan selalu menganggap kita zero kasus Hrsnya dr awal semua kepala

Proses	Tweet Asli	Hasil Proposes
Case Folding	arahan. Keliatan gk kompak. Payah! @mohmahfudmd @jokowi #VirusCoronaindonesia	daerah dikumpulin dan diberi arahan keliatan gk kompak payah
	Ini bukti pemerintah pusat gak koordinasi dr awal terkait wabah corona. Terlalu santai dan selalu menganggap kita zero kasus. Hrsnya dr awal semua kepala daerah dikumpulin dan diberi arahan. Keliatan gk kompak. Payah! @mohmahfudmd @jokowi #VirusCoronaindonesia	ini bukti pemerintah pusat gak koordinasi dr awal terkait wabah corona terlalu santai dan selalu menganggap kita zero kasus hrsnya dr awal semua kepala daerah dikumpulin dan diberi arahan keliatan gk kompak payah
Tokenizing	Ini bukti pemerintah pusat gak koordinasi dr awal terkait wabah corona. Terlalu santai dan selalu menganggap kita zero kasus. Hrsnya dr awal semua kepala daerah dikumpulin dan diberi arahan. Keliatan gk kompak. Payah! @mohmahfudmd @jokowi #VirusCoronaindonesia	['ini', 'bukti', 'pemerintah', 'pusat', 'gak', 'koordinasi', 'dr', 'awal', 'terkait', 'wabah', 'corona', 'terlalu', 'santai', 'dan', 'selalu', 'menganggap', 'kita', 'zero', 'kasus', 'hrsnya', 'dr', 'awal', 'semua', 'kepala', 'daerah', 'dikumpulin', 'dan', 'diberi', 'arahan', 'keliatan', 'gk', 'kompak', 'payah']
Stopword Removal	Ini bukti pemerintah pusat gak koordinasi dr awal terkait wabah corona. Terlalu santai dan selalu menganggap kita zero kasus. Hrsnya dr awal semua kepala daerah dikumpulin dan diberi arahan. Keliatan gk kompak. Payah! @mohmahfudmd @jokowi #VirusCoronaindonesia	bukti pemerintah pusat gak koordinasi dr terkait wabah corona santai menganggap zero hrsnya dr kepala daerah dikumpulin arahan keliatan gk kompak payah

Proses *training* menggunakan SVM non linear yaitu dengan fungsi kernel Radial Basis Function (RBF) dengan optimasi parameter C dan gamma (γ). Untuk mendapatkan fitur baru berdimensi tinggi, dilakukan kernelisasi menggunakan fungsi persamaan 1, dimana $x_i - x$ adalah data input.

$$K = (x_i, x) \exp(-\gamma ||x_i - x||^2) \tag{1}$$

$$Ld = \sum_i^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \tag{2}$$

dimana hasil nilai α harus memenuhi : - Syarat 1 : $\alpha_1 - \alpha_2 - \alpha_3 = 0$
- Syarat 2 : $\alpha_1, \alpha_2, \alpha_3 \geq 0$

Persamaan 1 menghasilkan matriks K yang akan berkorelasi dengan α_{ij} dalam dualitas *Lagrange Multiplier* (Ld max). *Lagrange Multiplier* digunakan untuk menentukan sejumlah *support vector* dengan menggunakan Quadratic Programming pada persamaan 2. yang selanjutnya dihasilkan model prediksi yang digunakan untuk klasifikasi pada data uji. Cara pengubahan data teks menjadi data vektor dari hasil pembobotan TF-IDF yaitu dengan mengambil term dari tiga dokumen tweet kelas positif, negatif, netral secara acak. Data vektor pada setiap dokumen tweet diurutkan berdasarkan indeks kata terkecil hingga terbesar. Tahap ini akan terus berulang hingga semua term pada data tweet sudah berubah ke format data vektor. Analisis *training one vs all* digunakan untuk klasifikasi k-kelas dengan ditemukan k fungsi pemisah dimana k adalah banyaknya kelas.

Tahap selanjutnya adalah menghitung nilai x. Nilai x yang didapat dari nilai term pada dokumen akan digunakan untuk perhitungan kernel. Selanjutnya dilakukan kernelisasi pada set data dari fitur dimensi lama sehingga mendapatkan set data fitur baru dimensi tinggi. Proses klasifikasi data uji dilakukan berdasarkan model pelatihan yang didapatkan melalui proses SVM *training* setelah dilakukan preprocessing dan pembobotan TF-IDF. Klasifikasi data yang diuji menghasilkan 3 kelas sentimen yaitu positif, negatif, netral yang diperoleh dari hasil nilai w dan *hyperplane data training*. SVM testing digunakan untuk mengukur kinerja atau evaluasi model dari 34 proses training dengan menghasilkan akurasi.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi

Untuk pelaksanaan implementasi ini digunakan bahasa pemrograman Python beserta *library*nya untuk membangun dan memproses analisis sentimen dengan SVM. Proses pertama dimulai dari proses *scrapping* data yang dilakukan untuk mengumpulkan data dari Twitter dengan batasan waktu dan jumlah data. Proses ini dilakukan dengan menggunakan Twitterscraper yang tersedia pada Python. Hasil dari proses *scrapping* yaitu data dalam bentuk file csv. Data tweet dari hasil *scrapping* selanjutnya dilakukan pelabelan data. Pada tahap ini, setiap tweet akan diklasifikasikan ke dalam tiga kelas yaitu positif, negatif, dan netral. Proses pelabelan dilakukan secara

manual dengan melihat satu per satu tweet berupa opini untuk digunakan pada proses pelatihan. Beberapa hasil pelabelan yang dilakukan seperti ditunjukkan pada tabel 2.

Setelah pelabelan data, selanjutnya adalah melakukan praproses terhadap data. Tahap ini terdiri dari 4 tahapan yaitu *text cleaning*, *case folding*, *tokenizing*, *stopwords removal* yang bertujuan untuk mempersiapkan dan membersihkan data sebelum diproses. *Stopword* menggunakan library nltk dan Sastrawi Bahasa Indonesia serta menghasilkan daftar kata yang akan dihapus pada dokumen tweet. Beberapa contoh *stopword* dapat dilihat pada tabel 3. Setelah dilakukan tahap *preprocessing*, selanjutnya fitur diekstraksi ke dalam bentuk vektor melalui pembobotan TF-IDF terhadap setiap kata (term). Sentimen pada data tweet terlebih dahulu diubah ke dalam bentuk numerik sehingga menjadi 1 untuk sentimen positif, 0 untuk sentimen netral, dan -1 untuk sentimen negatif. Penelitian ini menggunakan *TfidfVectorizer* dari modul *sklearn Python* untuk menghitung nilai bobot term. TF-IDF melakukan pembobotan lebih rendah terhadap kata-kata yang sering muncul dan bobot lebih tinggi untuk kata yang jarang muncul pada dokumen. *Tfidfvectorizer* bekerja dengan menghitung jumlah kata yang muncul, nilai IDF dan TF-IDF secara bersamaan menjadi matriks fitur. Penelitian ini menghasilkan 2903 fitur berupa term yang akan diketahui nilai bobotnya.

Tabel 2. Contoh hasil pelabelan manual

Tweet	Sentimen
Kalau pemerintah cuma menghimbau ini itu, jangan keluar dll, susah masyarakat kalau cuma dihibmau, kalau memang lockdown, ya udah lockdown, memang bakal ada resiko dampaknya, tapi ya harus memilih.	Negatif
Yukk... Gotong royong mencegah corona untuk indonesia yang aman dan tentram #CoronaIndonesia #Menkes #indonesiabisa #COVID19 #basmicorona #VirusCinta #President #GotongRoyongRakyat	Netral
Pemerintah sudah melakukan segala hal untuk menanggulangi #COVID19 .Jika belum maksimal dan terkesan gagal,masyarakat berperan penting dalam kegagalan itu,sebab masyarakat tak pedulikan protokol kesehatan	Positif

Pembagian data latih dan data uji dilakukan dengan menggunakan *K-Fold Cross Validation* dengan cara melipat data sebanyak K pengulangan (iterasi). Data yang digunakan adalah 600 tweet, dibagi menjadi data latih dan data uji menjadi 10 bagian (K=10), sehingga data yang diperoleh adalah 540 data latih dan 60 data uji. Pengujian menggunakan data yang sudah dipartisi akan diulang sebanyak 10 kali (K=10) dengan posisi data uji berbeda disetiap iterasi. Tahap ini dilakukan menggunakan library python yaitu *scikit learn model_selection*. Proses pembagian data latih dan data uji dilakukan dengan mengacak data tweet secara random. Proses pembagian iterasi pertama dimulai pada indeks ke-0 hingga indeks ke-59 sebagai data uji dan sisa indeks sebagai data latih. Iterasi akan berulang sampai 10 kali hingga semua record data mendapatkan bagian menjadi data uji. Pada tabel 4 ditunjukkan hasil representasi 10-Fold Cross Validation.

Tabel 3. Stopword

Stopword
ada
adalah
bahwa
biasa
cukup
demi
juga
jika
kalau
mau

3.2 Optimasi Hyperparameter SVM

Penelitian ini menggunakan optimasi *hyperparameter grid search* untuk mencari parameter terbaik pada proses SVM training. Parameter yang digunakan yaitu C dan gamma pada kernel RBF. *Scikit learn python* menyediakan kelas *GridSearchCV* yang dapat menghasilkan parameter terbaik dari proses training untuk prediksi hasil klasifikasi. Berdasarkan kernel yang digunakan yaitu RBF, maka pada penelitian ini nilai C dan gamma akan diuji coba adalah 0.1, 1, 10, 100. Hasil uji coba tersebut seperti ditunjukkan pada tabel 4. Split atau pembagian data sebanyak 10 kali iterasi dengan 10-Fold Cross Validation menghasilkan skor rata-rata yang terbaik yaitu 0,524 pada parameter C = 10 dan gamma = 0,1.

Tahap pelatihan model atau training merupakan tahap utama dalam penelitian ini. Tahap ini bertujuan untuk melatih sistem agar dapat mengklasifikasi sentimen dengan tingkat akurasi yang baik. Proses pelatihan model untuk keseluruhan data training pada penelitian ini menggunakan *LibSVM Support Vector Classifier (SVC)*. Parameter yang digunakan yaitu parameter terbaik dari tahap optimasi hyperparameter dengan *GridSearchCV*.

Data yang digunakan untuk pelatihan model yaitu data latih meliputi x_train sebagai tweet dan y_train sebagai label sentimen.

Tabel 4. Hasil Uji Coba Hyparameter SVM

Parameter	Spli t 1	Spli t 2	Spli t 3	Spli t 4	Spli t 5	Spli t 6	Spli t 7	Spli t 8	Spli t 9	Spli t 10	mean_test_score	Rank_test_score
{'C': 10, 'gamma': 0.1, 'kernel': 'rbf'}	4	0.5	0.4	0.5	0.5	0.5	0.4	0.5	0.5	0.5	0,524	1
{'C': 10, 'gamma': 1, 'kernel': 'rbf'}	4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0,520	2
{'C': 100, 'gamma': 1, 'kernel': 'rbf'}	4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0,520	2
{'C': 100, 'gamma': 0.1, 'kernel': 'rbf'}	2	0.4	0.4	0.5	0.5	0.5	0.4	0.4	0.5	0.5	0,507	4
{'C': 1, 'gamma': 1, 'kernel': 'rbf'}	2	0.5	0.4	0.5	0.5	0.5	0.5	0.5	0.4	0.4	0,50	5
{'C': 0.1, 'gamma': 0.1, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,477	6
{'C': 0.1, 'gamma': 1, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,477	6
{'C': 0.1, 'gamma': 10, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,477	6
{'C': 0.1, 'gamma': 100, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,477	6
{'C': 1, 'gamma': 0.1, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,477	6
{'C': 1, 'gamma': 10, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,475	11
{'C': 1, 'gamma': 100, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,475	11
{'C': 10, 'gamma': 10, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,475	11
{'C': 10, 'gamma': 100, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,475	11
{'C': 100, 'gamma': 10, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,475	11
{'C': 100, 'gamma': 100, 'kernel': 'rbf'}	8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0,475	11

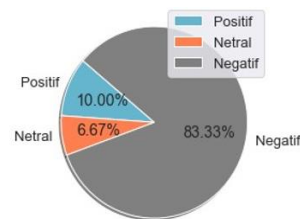
3.3 Klasifikasi Sentimen (SVM Testing)

Model hasil pelatihan akan digunakan untuk mengklasifikasi sentimen dari data uji tweet yang berjumlah 60 data. Proses klasifikasi ini bertujuan untuk menentukan apakah sebuah tweet merupakan opini positif, negatif, atau netral berdasarkan hasil prediksi model SVM. Pada tabel 5 dapat dilihat hasil dari perbandingan antara klasifikasi sentiment yang dihasilkan oleh SVM dengan RBF dan analisis sentimen yang diberikan label secara manual. Word cloud sentimen positif menggambarkan frekuensi kemunculan kata terbanyak yaitu pada kata pemerintah, psbb, indonesia, covid, virus corona pada data berlabel positif. Untuk Word Cloud pada sentimen negatif menggambarkan frekuensi kemunculan kata terbanyak yaitu pada kata pemerintah, psbb, indonesia, covid, penyebaran cepat pada data berlabel negatif. Sedangkan Word Cloud pada sentimen netral menggambarkan frekuensi kemunculan kata terbanyak yaitu pada kata pemerintah, psbb, indonesia, covid, penyebaran cepat pada data berlabel netral.

Tabel 5. Perbandingan Klasifikasi Sentimen

Tweet	Hasil Sentimen	Kelas Manual
katanya psbb total tetapi di ralat.. akhirnya jd gagal total alias Gatot. ekonomi hancur	Negatif	Negatif
Awal2 psbb total ganjil genap di tiadakan tetep sepi, senayang-pondok kelapa 30 menit, ngantuk dijalan	Negatif	Netral
Wow, dana penanggulangan pandemi covid-19 Jakarta triliunan besar sekali, tapi PSBB Jakarta gagal total	Negatif	Negatif
wew masih kudet juga , udah di bilangin berkali kali klo pusat setuju dgn PSBB total DKI"	Positif	Negatif
Kebijakan tanpa data, hanya tata kata yg berakibat PSBB gagal total"	Negatif	Negatif

Pada penelitian ini visualisasi data tampak pada gambar 2 yang terdiri dari persentase sentimen positif, negatif, dan netral. Diagram menunjukkan persentase 10% sentimen positif, 83.33% sentimen negatif, dan 6.67% sentimen netral dari 600 data yang kemudian diambil 540 untuk data latih dan 60 data uji. Kelas negatif memiliki nilai persentase maksimum artinya merupakan sentimen tweet terbanyak dari keseluruhan data atau sebagian besar tweet terdiri dari sentimen negatif terhadap pemerintah dalam kasus Covid-19. Pengukuran kinerja klasifikasi bertujuan untuk mengetahui apakah suatu model memiliki akurasi yang baik atau tidak dalam melakukan klasifikasi sentimen. Penelitian ini menggunakan *confusion matrix* untuk mengukur kinerja klasifikasi dengan menghitung nilai True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP) pada setiap kelas sentimen.



Gambar 2. Persentase Hasil Analisis Sentimen

Nilai True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP) yang sudah diperoleh dari masing-masing kelas kemudian akan dihitung nilai Presisi, Recall, dan F1-Score. Hasil dari pengukuran nilai-nilai tersebut seperti ditunjukkan pada tabel 6.

Tabel 6. Nilai Presisi, Recall, F1-Score

Kelas	Presisi	Recall	F1-Score
-1	0,84	0,93	0,88
0	0,25	0,20	0,22
1	0,50	0,30	0,37

Presisi menyatakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Pada kelas -1 diperoleh nilai presisi yang menyatakan bahwa persentase sentimen yang benar negatif dari keseluruhan sentimen yang diprediksi negatif sebesar 84%. Hasil presisi kelas 0 menyatakan bahwa persentase sentimen yang benar netral dari keseluruhan sentimen yang diprediksi netral sebesar 25%. Sedangkan presisi kelas 1 menyatakan bahwa persentase sentimen yang benar positif dari keseluruhan sentimen yang diprediksi positif sebesar 50%. Recall menyatakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Pada kelas -1 diperoleh nilai recall yang menyatakan bahwa persentase sentimen yang diprediksi negatif dari keseluruhan sentimen yang sebenarnya negatif sebesar 93%. Kelas 0 menyatakan bahwa persentase sentimen yang diprediksi netral dari keseluruhan sentimen yang sebenarnya netral sebesar 20%. Kelas 1 menyatakan bahwa persentase sentimen yang diprediksi positif dari keseluruhan sentimen yang sebenarnya positif sebesar 30%. F1-Score menyatakan perbandingan recall dan presisi dari masing-masing kelas. Pada kelas -1 diperoleh persentase nilai F1-Score sebesar 88%, kelas 0 diperoleh persentase F1-Score sebesar 22%, dan kelas 1 diperoleh persentase F1-Score sebesar 37%. Penelitian ini menggunakan akurasi sebagai acuan performansi kinerja klasifikasi. Akurasi diperoleh dari keseluruhan kelas yang mampu diprediksi dengan benar (TP) dibandingkan total data dengan nilai akurasi sebesar 77%.

4. KESIMPULAN

Berdasarkan hasil dari implementasi sistem untuk menganalisis sentimen opini publik tentang tindakan pemerintah Indonesia terhadap kasus Covid-19 menggunakan SVM dapat diambil kesimpulan bahwa untuk analisis sentimen tindakan pemerintah terhadap penanganan Covid-19 diklasifikasikan dalam 3 kategori yaitu, yaitu sentimen positif, sentimen negatif, dan netral. Berdasarkan hasil pengujian klasifikasi yang telah dilakukan, dari 60 data uji yang ada, didapatkan hasil sebesar 10% untuk sentimen positif, kemudian 83.33% sentimen negatif, dan 6.67% sentimen netral. Hasil pengujian tersebut menunjukkan sebagian besar tweet yang ada merupakan tweet dengan sentimen negatif terhadap pemerintah terkait penanganan kasus Covid-19. Akurasi sebagai acuan performansi kinerja klasifikasi pada penelitian ini menghasilkan nilai akurasi sebesar 77%.

REFERENCES

- [1] A. Sarkar, G. Liu, Y. Jin, Z. Xie, and Z. J. Zheng, "Public health preparedness and responses to the coronavirus disease 2019 (COVID-19) pandemic in South Asia: a situation and policy analysis," *Global Health Journal*, vol. 4, no. 4, pp. 121-132, Dec. 2020, doi: 10.1016/j.glohj.2020.11.003.

- [2] “South Korea’s Cults and Conservatives Sabotaged Virus Response.” <https://foreignpolicy.com/2020/02/27/coronavirus-south-korea-cults-conservatives-china/> (accessed Dec. 08, 2022).
- [3] Who, “Coronavirus disease (COVID-19) Global epidemiological situation.”
- [4] K. Kesehatan and L. Negara, “Tambahan Lembaran Negara Republik Indonesia Nomor 4723).”
- [5] S. Mujani and D. Irvani, “Sikap dan Perilaku Warga terhadap Kebijakan Penanganan Wabah Covid-19,” *Politika: Jurnal Ilmu Politik*, vol. 11, no. 2, pp. 219–238, Oct. 2020, doi: 10.14710/politika.11.2.2020.219-238.
- [6] E. Esyudha Pratama and H. Sastypratiwi, “JIP (Jurnal Informatika Polinema) Analisis Kecenderungan Informasi Terkait Covid-10 Berdasarkan Big Data Sosial Media dengan Menggunakan Metode Data Mining”.
- [7] “DIGITAL 2020 INDONESIA.”
- [8] P. Seminar and N. Geotik, “Analisis Sentiment Pada Twitter Dengan Menggunakan Metode Naïve Bayes Classifier,” 2018.
- [9] A. A. Collomb, D. Joyeux, O. Hasan, and L. Brunie, “A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation.”
- [10] B. Liu, “Sentiment analysis and opinion mining,” *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–184, 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.
- [11] W. C. F. Mariel, S. Mariyah, and S. Pramana, “Sentiment analysis: A comparison of deep learning neural network algorithm with SVM and naïve Bayes for Indonesian text,” in *Journal of Physics: Conference Series*, Apr. 2018, vol. 971, no. 1. doi: 10.1088/1742-6596/971/1/012049.
- [12] V. Kevin, S. Que, : Analisis, S. Transportasi, A. Iriani, and H. D. Purnomo, “Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization),” 2020. [Online]. Available: www.tripadvisor.com,
- [13] D. Darwis, E. Shintya Pratiwi, A. Ferico, and O. Pasaribu, “Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” 2020.
- [14] A. Putra, D. Haeirudin, H. Khairunnisa, and R. Latifah, “Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma Svm,” 2021.
- [15] F. Adams, L. Ernawati, and N. Chamidah, *Analisis Sentimen Vaksin COVID-19 pada Twitter Menggunakan Algoritma Support Vector Machine*. 2021.
- [16] A. M. Pravina, I. Cholissodin, and P. P. Adikara, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM),” 2019. [Online]. Available: <http://j-ptiik.uib.ac.id>
- [17] A. P. Gopi, R. N. S. Jyothi, V. L. Narayana, and K. S. Sandeep, “Classification of tweets data based on polarity using improved RBF kernel of SVM,” *International Journal of Information Technology 2020*, pp. 1–16, Jan. 2020, doi: 10.1007/S41870-019-00409-4.
- [18] I. A. Muis and M. Affandes, “Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet,” *Jurnal Sains, Teknologi dan Industri*, vol. 12, no. 2, pp. 189–197, 2015, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/sitekin>
- [19] X. hui Tan, W. hua Bi, X. liang Hou, and W. Wang, “Reliability analysis using radial basis function networks and support vector machines,” *Comput Geotech*, vol. 38, no. 2, pp. 178–186, Mar. 2011, doi: 10.1016/J.COMPGEO.2010.11.002.