

IMPLEMENTASI ALGORITMA KLASIFIKASI SUPPORT VECTOR MACHINE UNTUK ANALISA SENTIMEN PENGGUNA TWITTER TERHADAP KEBIJAKAN PSBB

¹Heart Parasian PR Zuriel, ²Achmad Fahrurozi
^{1,2}Fakultas Teknologi Industri, Universitas Gunadarma
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat
¹simanungkalitparasian@gmail.com, ²achmad.fahrurozi12@gmail.com

Abstrak

Pada saat ini penggunaan Twitter semakin luas. Pengguna Twitter dapat dengan bebas untuk berpendapat dan membagikan sudut pandang mereka mengenai isu tren dunia. Hal ini membuat konten Twitter menjadi beragam dan menarik untuk dianalisa, termasuk dengan tren ini membuat pemerintah mengeluarkan kebijakan yang bertujuan untuk menekan laju pertambahan orang yang terinfeksi virus. Kebijakan ini diberi nama Pembatasan Sosial Berskala Besar atau yang dikenal PSBB. Kebijakan ini pun hangat diperbincangkan di berbagai sosial media, tak terkecuali Twitter. Analisa sentimen dilakukan dengan menggunakan Support Vector Machine (SVM) sebagai algoritma klasifikasi pada data tweet yang berjumlah 22.335 data. Pelabelan data dalam penelitian ini dilakukan menggunakan metode Lexicon Based. Pada penelitian ini terdapat 4 model SVM yang dibangun menggunakan 4 fungsi kernel yaitu kernel Linear, RBF, Polinomial dan Sigmoid. Hasil klasifikasi dari masing-masing model diukur performanya menggunakan k-fold cross validation. Berdasarkan perhitungan, diperoleh bahwa performa model klasifikasi SVM dengan kernel RBF merupakan yang terbaik dibanding kernel lainnya dalam kasus penelitian analisa sentimen ini. Nilai accuracy, precision, recall, dan f1-score-nya yang diperoleh model klasifikasi dengan kernel RBF secara berturut-turut adalah 95.94%, 94.41 %, 97.8%, dan 96.08%. Model klasifikasi dengan kernel RBF ini memberikan mengklasifikasikan 11.764 (52.7%) data tweet ke dalam kelas positif dan 10.571 (47.3%) data tweet ke dalam kelas negative. Hasil tersebut memberikan kesimpulan bahwa pengguna Twitter cenderung bersentimen positif terhadap kebijakan PSBB.

Kata Kunci: Analisa Sentimen, Fungsi Kernel, PSBB, Support Vector Machine, Twitter

Abstract

Nowadays, the use of Twitter is increasingly widespread. Twitter users can freely express their opinions and share their point of view on world trending issues. This makes Twitter content diverse and interesting to analyze, including the trend of government policies that are widely discussed in Indonesia. The emergence of the Covid-19 pandemic has prompted the government to issue a policy that aims to reduce the rate of increase in people infected with the virus. This policy is called Large-Scale Social Restrictions or what is known as PSBB. This policy is also hotly discussed on various social media, including Twitter. Sentiment analysis is carried out using the Support Vector Machine (SVM) as a classification algorithm for tweet data, totaling 22,335 data. Data labeling in this study done by using Lexicon Based method. In this study, there are 4 SVM models built using 4 kernel functions, namely the Linear kernel, RBF, Polinomial and Sigmoid. In this study, there are 4 SVM models built using 4 kernel functions, namely the Linear kernel, RBF, Polinomial and Sigmoid. The results of the classification of each model are measured for their performance using k-fold cross validation. Based on the calculations, it is found that the performance of the SVM classification model with the RBF kernel is the best compared to other kernels in the case of this sentiment analysis research. The accuracy, precision, recall, and f1-score values obtained by the classification

model with the RBF kernel were 95,94%, 94,41 %, 97,8%, and 96,08%, respectively. This classification model with the RBF kernel provides classifying 11.764 (52,7%) tweet data into a positive class and 10.571 (47,3%) tweet data into a negative class. These results conclude that Twitter users tend to have a positive sentiment towards the PSBB policy.

Keywords: Kernel Function, PSBB, Sentiment analysis, Support Verctor Machine, Twitter

PENDAHULUAN

Internet saat ini sudah menjadi kebutuhan bagi masyarakat terutama di Indonesia dimana penggunaan internet setiap tahunnya terus bertambah. Sebuah *survey* yang dilakukan oleh Asosiasi Penyelenggaraan Jasa Internet Indonesia (APJII) menunjukkan ada peningkatan 10,12% penggunaan internet pada tahun 2019 dibandingkan dengan tahun 2018. Pengguna internet yang sudah masif ini turut bertambah dengan penggunaan sosial media, dimana sosial media ini menjadi wadah masyarakat untuk saling berinteraksi satu sama lain tanpa batas waktu maupun tempat. Sosial media yang populer dikalangan masyarakat salah satunya adalah Twitter. Menurut data yang dimiliki PT. Bakrie Telkom, Twitter yang merupakan situs *microblogging* populer ini memiliki pengguna di Indonesia pada tahun 2013, terdapat 19,5 juta dari 500 juta pengguna secara global.

Twitter menjadi populer karena dirasa lebih simpel dalam hal menyampaikan opini atau pendapat terhadap produk, isu sosial, hingga kebijakan pemerintah yang menjadi topik hangat dalam bentuk teks singkat yang disebut dengan *tweet*. Kebijakan yang sempat ramai diperbincangan di Twitter salah satunya

adalah kebijakan PSBB. Kebijakan PSBB ini merupakan kebijakan pemerintah yang bertujuan untuk menekan angka penyebaran kasus Covid-19 di tanah air. Hal ini disebabkan karena pertumbuhan angka kasus infeksi virus semakin hari semakin meningkat.

Data yang diperoleh dari website gugus tugas percepatan penanggulangan Covid-19, per tanggal 19 Mei 2020 jumlah kasus terinfeksi covid-19 sebesar 43,803 kasus dengan penambahan kasus perharinya 1.041 kasus [1]. Kebijakan PSBB menuai pro kontra di media sosial Twitter dan menjadi topik yang hangat untuk di bahas dan komentar-komentar dari masyarakat ini pun memiliki sentimen positif maupun negatif terhadap kebijakan ini yang menjadi menarik untuk dilakukan analisa sentimen pengguna terhadap kebijakan PSBB ini.

Analisa sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang [2]. Analisa sentimen diimplementasikan pada berbagai situs *review* pengguna layanan jasa hotel [3], *review* pengunjung lokasi wisata [4], maupun pada situs *microblogging* Twitter [5], bertujuan untuk mengetahui tingkat kepuasan konsumen, *tweet* yang mengarah ke *cyber bullying* [6], sentimen masyarakat terhadap

calon Gubernur [7], maupun sentimen terhadap penerapan *social distancing* dalam masa pandemi [8]. Secara umum, penelitian terkait analisa sentimen dapat dibedakan berdasarkan metode pelabelan *dataset*, metode klasifikasi yang digunakan, serta jumlah kelas yang didefinisikan.

Berdasarkan metode klasifikasinya, terdapat tiga pendekatan utama dalam analisa sentimen, yaitu pendekatan *machine learning*, pendekatan *lexicon-based*, dan *hybrid* antar keduanya. Pendekatan *lexicon-based* sendiri dibagi ke dalam 2 kelompok besar, yaitu pendekatan *dictionary-based* (didasarkan pada kata-kata dalam kamus, misalnya WordNet, SenticNet, dan lain-lain) dan pendekatan *corpus-based* (menggunakan data *corpus* data yang dibangun, yang dibagi lagi ke dalam pendekatan *statistical* dan *semantic*) [9]. Algoritma *machine learning* yang umumnya digunakan dalam analisa sentimen dengan pendekatan *machine learning* adalah *Naive-Bayes classifier* [10], SVM [8], LSTM [11], maupun *ensemble classifier* [4].

T. Ahn Le, dkk [11] menggunakan 4 ribu *tweet* yang dilabeli manual sebagai *dataset*, terdiri dari *tweet* yang mengandung banyak kata-kata singkatan atau istilah tidak formal dalam bahasa Indonesia. Pada penelitian tersebut didefinisikan 3 kelas (positif, negatif, dan netral), menggunakan algoritma *Long Short Term Memory* (LSTM) sebagai *classifier* dan memperoleh tingkat akurasi sebesar 73,2%. U. Khaira, dkk [6] juga melakukan pelabelan secara manual dan

mendefinisikan 3 kelas, sementara klasifikasi dilakukan dengan menggunakan *lexicon-based* dan SentiStrength sebagai pendekatannya. Dari *dataset* berupa 454 *tweet*, diperoleh bahwa 161 *tweet* tidak terindikasi *bullying* (35.4%), 87 *tweet* netral (19.1%), dan 206 *tweet* terindikasi *bullying* (45.4%). Model klasifikasi dalam penelitian ini memiliki tingkat akurasi sebesar 60.5%. Tingkat akurasi yang rendah kemungkinan diperoleh karena menggunakan *dictionary* yang tidak sesuai dengan karakteristik *dataset* yang dimiliki.

Sementara itu C. Shofiya dan S. Abidi [8] menggunakan pendekatan lainnya dalam melakukan analisa sentimen, yaitu pendekatan *machine learning*. Penelitian tersebut menggunakan total 629 *tweet* sebagai *dataset*, dengan persentase *tweet* netral, negatif, dan positif berturut-turut adalah 40%, 35%, dan 25%.

Proporsi data latih dan data uji yang diterapkan adalah 80% berbanding 20%, sementara tingkat akurasi dari klasifikasi dengan 3 kelas tersebut adalah 71% menggunakan kernel RBF. C. Shofiya dan S. Abidi mengubah jumlah kelas menjadi 2 dan memperoleh peningkatan akurasi menjadi 81% jika menggunakan kernel Linier, namun tidak meningkat jika menggunakan kernel RBF. Di sisi lain, Murni, dkk [4] menggunakan pendekatan *hybrid* dalam melakukan analisa sentiment terhadap review pengunjung tempat wisata. Secara singkat, pendekatan *hybrid* tersebut menggunakan metode *lexicon-based* untuk melabelkan

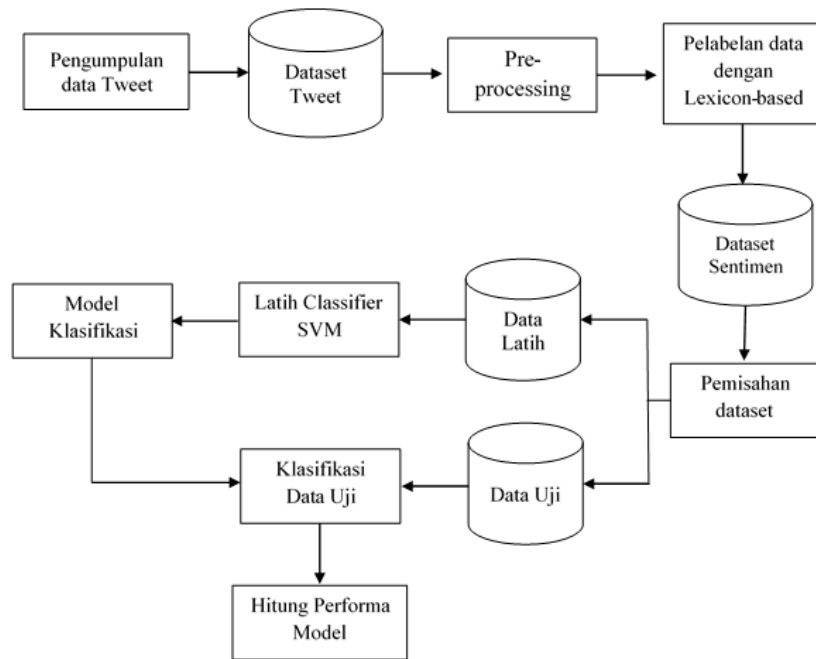
dataset, sehingga tidak perlu dilakukan secara manual, sementara metode *machine learning* digunakan sebagai *classifier*.

Analisa sentimen dinilai dapat alat bantu untuk mengetahui bagaimana sentimen masyarakat dalam opininya terhadap sebuah layanan, kebijakan, dan lain sebagainya. Oleh karena itu diperlukan proses otomatis untuk melakukan proses klasifikasi dan sentimen analisis setiap opini atau *dataset* yang diambil. Berdasarkan penelitian terdahulu, SVM merupakan algoritma *machine learning* yang sering digunakan dan memberikan hasil yang cukup mumpuni dalam riset analisa sentimen. SVM juga menyediakan pilihan yang banyak dalam pembentukan model klasifikasinya, terkait dengan fungsi kernel yang dapat digunakan untuk membangun modelnya.

Pemilihan fungsi kernel berbeda dapat meningkatkan akurasi [8] pada karakteristik data tertentu. Penelitian ini bertujuan untuk menerapkan algoritma SVM untuk menganalisa sentimen pengguna Twitter terhadap kebijakan PSBB dengan mempertimbangkan fungsi kernel yang digunakan dalam membangun model klasifikasinya.

METODE PENELITIAN

Analisa sentimen dilakukan menggunakan *dataset* berupa *tweet* yang didapat dari sosial media sosial Twitter. *tweet* diambil berdasarkan kata kunci 'psbb' yang diambil pada tanggal 28 Mei, 16 September dan 27 September 2020. Data *tweet* mentah diperoleh menggunakan akses Twitter *API* dengan cara *crawling*. Setelah itu data mentah melalui tahap *pre-processing* seperti pembersihan *tweet* dari simbol, angka, username, RT, hashtag, tanda baca, url dan lainnya yang dirasa mengganggu proses ekstraksi fitur dan klasifikasi *tweet*. Setelah proses itu, data akan dilabeli dengan positif dan negatif menggunakan *Lexicon Based*, kemudian data yang telah dilabeli tersebut akan menjadi *dataset* untuk diolah pada tahap *processing*. Pada tahap *processing* data diklasifikasikan menggunakan model algoritma *Support Vector Machine*, diberikan juga hasil dari pelatihan dan uji model algoritma. Pada tahap akhir merupakan penyajian hasil model algoritma klasifikasi yang divisualisasikan dalam bentuk tabel maupun diagram. Tahapan dalam melakukan analisa sentimen ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Proses Analisa Sentimen

Pengumpulan Data

Data diperoleh dengan menggunakan teknik *crawling data* pada situs www.twitter.com. Penulis menggunakan fasilitas dari Twitter berupa Twitter API untuk melakukan request data. Sebelum melakukan pengumpulan data dengan *crawling* pada situs Twitter langkah pertama yang harus dilakukan adalah mendaftarkan akun menjadi pengembang pada situs <https://developer.twitter.com/en/apps> dan melakukan pendaftaran *Apps*. Adapun data diambil pada tanggal 28 Mei, 16 September, dan 27 September 2020.

Setelah melalui semua proses dan diterima oleh pihak Twitter selanjutnya dilakukan pembuatan *Apps* untuk mendapatkan *consumer API key*, *consumer API secret key*, *access token*, dan *access token secret*, dimana 4 hal itu dibutuhkan untuk

menjadi *credential* untuk melakukan *request* pada Twitter *API*. *Crawling* data dilakukan dengan menggunakan kata kunci ‘psbb’ pada tanggal 28 Mei, 16 September dan 27 September 2020. Setelah menjalankan naskah kode diatas diperoleh data sebanyak 24,704 dan kemudian disimpan kedalam file berformat csv.

Preprocessing Data

Preprocessing data merupakan tahap untuk mengubah bentuk data menjadi lebih terstruktur sehingga dapat sesuai dengan kebutuhan untuk proses lebih lanjut. Secara garis besar tahapan *preprocessing* disajikan pada Gambar 2, dengan input berupa data *tweet*. Terdapat 4 proses secara terurut yang dilakukan pada tahap ini, yaitu *Case Folding*, *Cleaning*, *Tokenizing*, dan *Stopword Removal*. Berikut *Case folding* merupakan proses

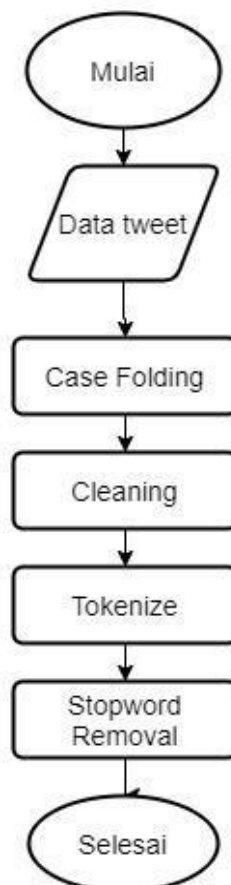
merubah huruf besar alfabet atau *uppercase* menjadi huruf kecil (*lowercase*), sementara *Cleaning* merupakan proses membersihkan data dari komponen- komponen yang tidak diperlukan dalam proses analisa sentimen. Hal yang dibersihkan pada proses ini adalah tautan URL, simbol RT, simbol @, digit, *whitespace*, *tags*, dan *hashtags*. Selanjutnya masuk ke proses *Tokenize*, merupakan penguraian kalimat menjadi token-token atau kata tunggal. Proses terakhir yaitu *Stopword Removal*, merupakan salah satu proses penting dalam melakukan analisa sentimen.

Pembobotan Kata dan Pelabelan Data

Pembobotan kata adalah proses pemberian nilai terhadap setiap kata yang ada pada setiap opini yang sudah melewati proses data *preprocessing*. Pada penelitian ini digunakan metode TF-IDF dengan menggunakan persamaan (1) dan persamaan (2).

$$IDF(w) = \log\left(\frac{N}{DF(w)}\right) \quad (1)$$

$$TFIDF(w,d) = TF(w,d) \times IDF(w) \quad (2)$$



Gambar 2. Alur Tahapan *Preprocessing* Data

Proses pelabelan data pada penelitian ini dilakukan dengan menggunakan metode *Lexicon-based*. Prosedur untuk menentukan kelas sentimen dengan menggunakan *Lexicon-based* adalah dengan menghitung skor jumlah bobot kata positif ditambah skor jumlah bobot kata negatif dalam setiap kalimat.

Kalimat yang memiliki skor lebih besar dari nol diklasifikasikan ke dalam kelas positif, sedangkan kalimat yang memiliki skor sama dengan nol akan diklasifikasikan ke dalam kelas netral, sedangkan kalimat yang memiliki skor lebih kecil dari diklasifikasikan ke dalam kelas negatif. Bobot positif dan negatif diperoleh dari kamus *lexicon* yang diunduh dari situs <https://github.com/fajri91/InSet> [12].

Tahap Pelatihan dan Pengujian Model Klasifikasi

Pada Model klasifikasi sentimen terlebih dahulu dilatih dengan data latih yang telah diberikan bobot dengan TF-IDF atau yang disebut tahap pelatihan, tahap pelatihan bertujuan untuk menemukan nilai a , nilai w dan nilai konstanta b sebagai parameter untuk menemukan *hyperplane* terbaik. Data latih yang telah diberikan bobot diubah menjadi bentuk vector dengan $y \in (-1, 1)$. Untuk mencari margin terbesar digunakan persamaan (3), dengan kendala yang dinyatakan oleh pertidaksamaan (4).

$$\min r(w) = \frac{1}{2} \|w\|^2 \quad (3)$$

$$y_i(w \cdot x_i + b) - 1 \geq 0, \forall i \quad (4)$$

Model matematis dari permasalahan di atas dapat diselesaikan dengan lebih mudah dengan mengubahnya menjadi fungsi Lagrangian dan disederhanakan menjadi persamaan (5), dengan variabel $\bar{a}_i \geq 0$ sebagai Lagrange *multiplier*.

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i y_i (w^T \cdot x_i + b) + \sum_{i=1}^n a_i \quad (5)$$

Nilai optimum dari persamaan (5) dapat diperoleh dengan meminimumkan L terhadap w sekaligus terhadap b dan a , sehingga masalah Lagrange dapat dinyatakan dengan persamaan (6) berikut:

$$\min L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i y_i (w^T \cdot x_i + b) + \sum_{i=1}^n a_i \quad (6)$$

Selanjutnya akan diperoleh nilai \bar{a}_i yang digunakan untuk menemukan nilai bobot (w). Terdapat nilai \bar{a}_i untuk setiap data latih dalam masalah klasifikasi. Nilai \bar{a}_i yang lebih besar dari merupakan *support vector* dan dapat digunakan untuk mencari *hyperplane* (b). Untuk memisahkan data yang sulit dipisahkan secara linear maka digunakan fungsi kernel untuk melakukan transformasi data. Persamaan kernel yang digunakan contohnya adalah kernel linier, seperti pada persamaan (7) [13].

$$K(x_i, x_j) = (x_i^T x_j) \quad (7)$$

Pada penelitian ini digunakan 4 macam kernel, yaitu kernel Linear, RBF, Polinomial dan Sigmoid. Setelah pembentukan model klasifikasi, kemudian model diuji dengan data uji yang sudah diperoleh. Pengujian ini dilakukan untuk menghitung performa model klasifikasi yang dibuat dan melihat prediksi yang dilakukan oleh model klasifikasi terhadap data uji.

HASIL DAN PEMBAHASAN

Hasil Pengumpulan Data

Berdasarkan penjelasan sebelumnya, setelah data berhasil dikumpulkan dengan menggunakan teknik *crawling* data, kemudian data disimpan dalam bentuk csv. Dari proses pengumpulan data berhasil dikumpulkan

24,704 data *tweet*. Tabel 1 menyajikan beberapa contoh dari hasil pengumpulan data.

Hasil *Preprocessing*

Setelah data dikumpulkan, data masuk ke dalam tahap *preprocessing* sebelum data tersebut dapat digunakan untuk melatih dan menguji model klasifikasi.

Pada tahap *preprocessing* ini, data melewati proses *casefolding*, *cleaning*, *tokenizing*, dan *Stopword Removal*. Tabel 2 menyajikan beberapa hasil dari tahap *preprocessing* data.

Hasil Pelabelan Data

Setelah dilakukan *preprocessing* dan pembobotan kata pada setiap data *tweet* yang diperoleh, dilakukan pemberian label terhadap tiap data tersebut dengan menggunakan metode *Lexicon Based*.

Tabel 1. Contoh Hasil Pengumpulan Data

No.	<i>Tweet</i>
1	@Lyonskylight Sabar woee lagi psbb wkww
2	RT @indo35meme: wah psbb diperpanjang https://t.co/1hxzevkOtW
3	Udah mau 3x psbb tidak ada perubahan. https://t.co/KwFAbOYQuH

Tabel 2. Hasil Data *Preprocessing*

Data Sebelum Tahap <i>Preprocessing</i>	Data Sesudah Tahap <i>Preprocessing</i>
@Lyonskylight Sabar woee lagi psbbwkww	sabar woee psbb wkww
RT @indo35meme: wah psbb diperpanjang https://t.co/1hxzevkOtW	psbb diperpanjang
Udah mau 3x psbb tidak ada perubahan. https://t.co/KwFAbOYQuH	udah psbb tidak perubahan

Tabel 3 menyajikan beberapa contoh hasil pelabelan data. Dari 24,704 data yang telah dilabeli dengan *Lexicon Based*, terdapat 11.357 data berlabel positif, 10.978 data berlabel negatif, dan 1.739 data berlabel netral. Pada penelitian ini, analisa sentimen dilakukan hanya pada kelas positif dan negatif, sehingga data berlabel netral dihapus dari *dataset*. Dengan demikian, jumlah data berlabel yang digunakan dalam tahap pelatihan dan pengujian berjumlah total 22.335 data.

Hasil Analisa Sentimen

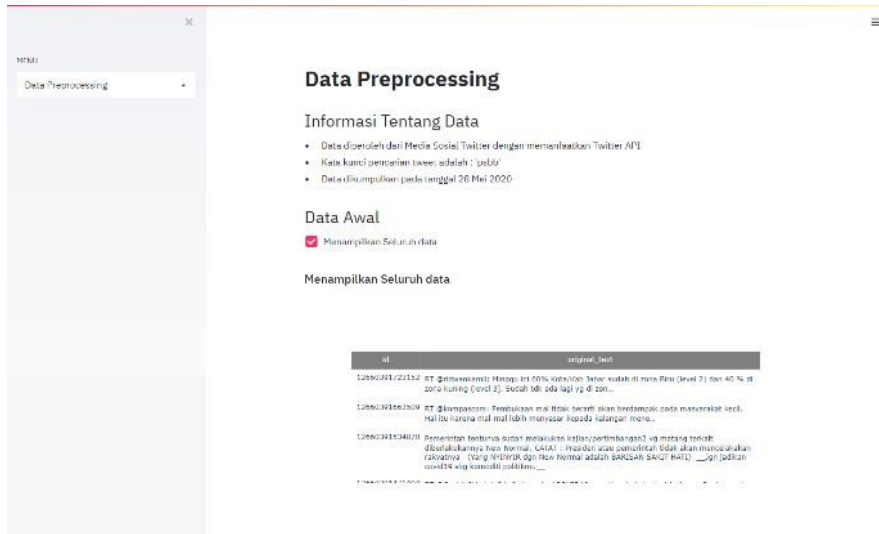
Hasil analisa sentimen pada penelitian ini disajikan ke dalam web interaktif yang memungkinkan *user* untuk memilih fungsi kernel yang digunakan dalam membangun model klasifikasi SVM. Secara garis besar, *website* yang dibuat memiliki 4 buah halaman, yaitu halaman Tentang, halaman Data *Preprocessing*, halaman

Klasifikasi Sentimen, dan halaman Hasil Klasifikasi. Halaman Tentang berfungsi untuk menampilkan judul dari penelitian, abstrak, dan data diri pembuat web ini. Pada halaman tentang ini hampir keseluruhan isinya berupa teks, namun terdapat juga checkbox untuk menampilkan data diri pembuat web.

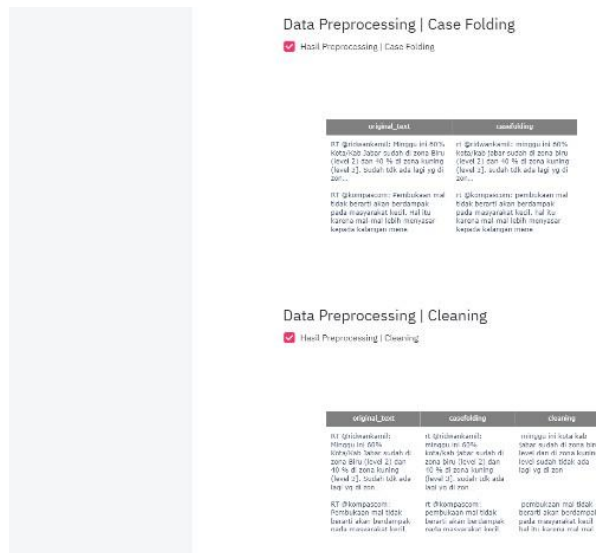
Halaman Data *Preprocessing* dirancang untuk menampilkan tahapan yang terjadi pada tahap *preprocessing* data. Halaman ini menampilkan data awal yang disajikan dalam bentuk tabel, berisi hasil dari pengumpulan data yang telah dilakukan. Pada halaman Data *Preprocessing* juga ditampilkan tabel data hasil dari proses *case-folding*, *cleaning*, *tokenizing*, maupun *Stopword Removal*. Tampilan halaman Data *Preprocessing* dapat dilihat pada Gambar 3, Gambar 4, dan Gambar 5 yang berturut-turut menampilkan data awal, data hasil proses *case-folding* dan *cleaning*, serta data hasil proses *tokenizing* dan *Stopword Removal*.

Tabel 3. Contoh Hasil Pelabelan pada Data Penelitian

Kode	Data	Sentimen
D1	sabar woe psbb wkwkw	1
D2	psbb diperpanjang	1
D3	udah psbb tidak perubahan	-1



Gambar 3. Tampilan Data Awal Pada Halaman Data *Preprocessing*



Gambar 4. Tampilan Data Hasil Proses *Case-Folding* dan *Cleaning* Pada Halaman Data *Preprocessing*

Pada halaman Klasifikasi Sentimen disajikan hasil evaluasi dari model klasifikasi yang dibangun. Selain itu ditampilkan tabel data yang digunakan untuk pelatihan dan pengujian model klasifikasi. Data yang ditampilkan merupakan data yang sudah dilabeli dengan metode *Lexicon Based*. Pada

halaman ini terdapat juga histogram frekuensi sentimen dari basis data penelitian dan selectbox untuk memilih fungsi kernel untuk membangun model klasifikasi SVM yang diinginkan. Tampilan bagian awal dan contoh isi dari halaman Klasifikasi Sentimen dapat dilihat pada Gambar 6 dan Gambar 7.

Data Preprocessing | Tokenizer

Hasil Preprocessing | Tokenizer

Original_text	word_tokenize	tokenizer	word
01 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	01 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	tokenizer	01 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini
02 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	02 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	tokenizer	02 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini
03 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	03 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	tokenizer	03 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini

Data Preprocessing | Stopword Removal


Hasil Preprocessing | Stopword Removal

Original_text	word_tokenize	tokenizer	word
01 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	01 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	tokenizer	01 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini
02 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	02 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	tokenizer	02 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini
03 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	03 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	tokenizer	03 @Pemerintah negara ini akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini

Gambar 5. Tampilan Data Hasil Proses *Tokenizing* dan *Stopword Removal* pada Halaman *Data Preprocessing*

Menu
Klasifikasi Sentimen
Pilih Koneksi
Break

Klasifikasi Sentimen

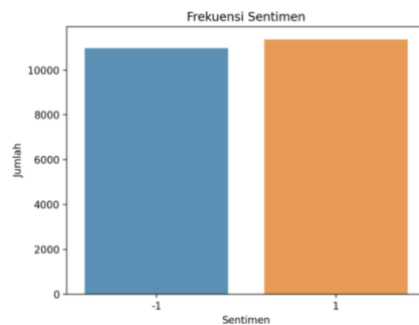


Klasifikasi sentimen dilakukan dengan menggunakan algoritma klasifikasi Support Vector Machine (SVM)

Data

id	original_text	class_sentimen	sentiment
1	Indonesia akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	0	negatif
2	Indonesia akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	1	positif
3	Indonesia akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	0	negatif
4	Indonesia akan menjadi negara yang lebih maju dan lebih sejahtera dengan adanya kegiatan ini	1	positif

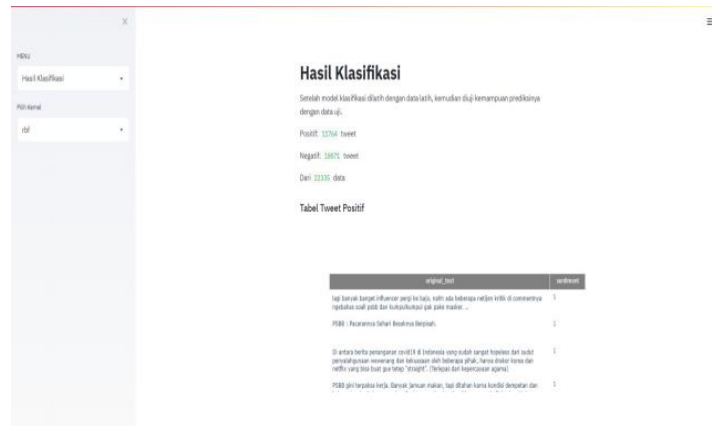
Gambar 6. Tampilan Awal Halaman Klasifikasi Sentimen



Gambar 7. Histogram Frekuensi Sentimen Basis Data pada Halaman Klasifikasi Sentimen

Sementara itu, halaman Hasil Klasifikasi menampilkan hasil dari prediksi yang dilakukan oleh model klasifikasi SVM yang dibangun sesuai dengan kernel yang dipilih oleh *user*. Hasil klasifikasi ini disajikan dalam bentuk tabel dan *pie chart*. Tabel

menyajikan data uji yang termasuk ke dalam kelas positif maupun kelas negatif, sementara *pie chart* menyajikan persentase data uji yang masuk ke kelas positif dan negatif. Tampilan halaman Hasil Klasifikasi dapat dilihat pada Gambar 8 dan Gambar 9.



Gambar 8. Tampilan Bagian Awal dari Halaman Hasil Klasifikasi

Tabel Tweet Negatif

original_text	sentiment
RT @KementerianKesehatan: New Zealand melanjutkan PHIS dengan serius, sistematis, berbasis data & pandemi. Perdana Menteri Jacinda Ardern adalah s...	+1
RT @Sindonews: Kasus meninggalnya Pak Sekda DKI, harusnya bisa jadi bahan riset oleh Kementerian kesehatan. Bagaimana pasien yang tidurnya O...	-1
RT @antabangkukul: Rerang barang di Mall pada berjam-jam. Baru beberapa bulan PSBB dan rakyat awam menahan diri tidak ke Mall. Berarti jk...	-1
RT @bakapau_putih: Hari ini jogsaaa mau expo pibb #Avallogia #realtibayar #jogsapendit #jogjainapak #exp2020 #Avallogia #Avallogia https://t.co/0U6...	-1

Pie Chart Sentimen



Gambar 9. Bagian Akhir dari Halaman Hasil Klasifikasi

Tabel 4. Hasil Pengukuran Performa Model Klasifikasi Berdasarkan Fungsi Kernel

Parameter Performa	Linear	RBF	Polinomial	Sigmoid
Akurasi	94.93%	95.94%	94.55%	93.01%
Precision	95.0%	94.41%	90.59%	93.06%
Recall	95.03%	97.8%	99.62%	93.20%
F1-score	95.01%	96.08%	94.89%	93.13%

Performa Model Klasifikasi

Performa model klasifikasi meliputi empat buah parameter, yaitu akurasi, *precision*, *recall* dan *f1-score*. Pengukuran performa model didasarkan pada confusion

matrix dan dengan menggunakan data berjumlah 22.335. Pengukuran performa dilakukan terhadap 4 buah model SVM yang dibangun dengan 4 fungsi kernel berbeda. Tabel 4 menyajikan hasil pengukuran

performa model SVM berdasarkan fungsi kernelnya. Performa yang disajikan merupakan rata-rata dari hasil perhitungan tiap iterasi dengan metode *k-fold cross validation*, dengan nilai *k* yang dipilih adalah 5.

Setelah dilakukan pengukuran performa sebanyak 5 kali dengan *5-cross validation* diperoleh bahwa model klasifikasi SVM dengan kernel RBF merupakan model klasifikasi yang memberikan performa keseluruhan terbaik dengan *dataset* pada penelitian ini, dengan nilai akurasi, *precision*, *recall* dan *f1-score* secara berturut-turut adalah 95.94%, 94.41 %, 97.8%, dan 96.08%. Model klasifikasi ini memprediksi terdapat sentimen positif sebanyak 11.764 *tweet* dan sentimen negatif sebanyak 10.571 *tweet*, seperti dapat dilihat pada *pie chart* dalam Gambar 11. Hasil tersebut menunjukkan bahwa pengguna Twitter cenderung memiliki sentimen positif terhadap kebijakan PSBB menurut model klasifikasi sentimen yang dibuat pada penelitian ini.

KESIMPULAN DAN SARAN

Berdasarkan hasil pengukuran performa model klasifikasi SVM dengan kernel RBF mendapatkan hasil keseluruhan yang paling baik dibanding model dengan kernel lainnya. Nilai akurasi, *precision*, *recall*, dan *f1-score* yang dimiliki oleh model tersebut secara berturut-turut adalah 95.94%, 94.41 %, 97.8%, dan 96.08%. Model

klasifikasi memprediksi terdapat sentimen positif sebanyak 11.764 *tweet* (52.7%) dan sentimen negatif sebanyak 10.571 *tweet* (47.3%). Dari hasil ini disimpulkan bahwa dari 22.335 *tweet* pengguna Twitter yang diambil pada tanggal 28 Mei, 16 September dan 27 September 2020, cenderung bersentimen positif terhadap kebijakan PSBB di Indonesia. Adapun saran untuk penelitian selanjutnya diantaranya adalah melakukan tahap *preprocessing* data yang lebih banyak dan menggunakan data selain dari sosial media Twitter. Penggunaan algoritma klasifikasi lain juga disarankan untuk pengembangan penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] Gugus Tugas Percepatan Penanganan Covid 19, "Peta Sebaran", 2020, <https://covid19.go.id/peta-sebaran>, diakses 19 Juni 2020.
- [2] B. Pang dan L. Lee, "Opinion Mining and Sentiment Analysis", *Foundation and Trends Information Retrieval*, Vol. 2 (No. 1-2), hal. 1-135, 2008.
- [3] G. Markopoulos, G. Mikros, A. Iliadi, dan M. Lontos, "Sentiment Analysis of Hotel Reviews in Greek: A Comparison of Unigram Features", *Springer Proceedings in Business and Economics*, 2015, hal. 373-382.
- [4] Murni, T. Handhika, A. Fahrurozi, I. Sari, D.P. Lestari, "Hybrid Method for Sentiment Analysis Using

- Homogeneous Ensemble *Classifier*", 2nd International Conference of Computer and Informatics Engineering (IC2IE), 2019, hal. 232-236.
- [5] O.Y. Adwan, M. Al-Tawil, A.M. Huneiti, R.A. Shahin, A. Zayed, R.H. Al-Dibsi, "Twitter Sentiment Analysis Approaches: A Survey", *iJET*, Vol. 15, No. 15, hal. 79-92, 2020.
- [6] U. Khaira, R. Johanda, P. Utomo, dan T. Suratno, "Sentiment Analysis of Cyberbullying on Twitter Using SentiStrength", *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, Vol. 3, No.1, hal. 21-27, 2020.
- [7] G.A. Buntoro, "Sentiment Analysis to Prediction DKI Jakarta Governor 2017 on Indonesian Twitter", *International Journal Of Science, Engineering, And Information Technology*, Vol. 2, No. 1, hal. 36-40, 2017.
- [8] C. Shofiya dan S. Abidi, "Sentiment Analysis on COVID-19-Related Social Distancing in Canada Using Twitter Data", *Int. J. Environ. Res. Public Health*, 18, 5993, 2021.
- [9] A. Sadia, F. Khan, dan F. Bashir, "An Overview of *Lexicon-Based* Approach For Sentiment Analysis", 2018 3rd International Electrical Engineering Conference (IEEC 2018), 2018, hal. 1-6.
- [10] Y. Cahyono dan Saprudin, "Analisis Sentiment Tweets Berbahasa Sunda Menggunakan Naive Bayes *Classifier* dengan Seleksi Feature Chi Squared Statistic", *Jurnal Informatika Universitas Pamulang*, Vol. 4, No. 3, hal. 87-94, 2019.
- [11] T. Anh Le, D. Moeljadi, Y. Miura, dan T. Ohkuma, "Sentiment Analysis for Low Resource Languages: A Study on Informal Indonesian Tweets", *Proceedings of the 12th Workshop on Asian Language Resources*, 2016, hal. 123–131.
- [12] F. Koto dan G. Y. Rahmanytyas, "InSet *Lexicon*: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs", 2017 International Conference on Asian Language Processing (IALP), hal. 391-394, 2017.
- [13] S. S. Chaeikar, "PFW: Polygonal Fuzzy Weighted—An SVM Kernel for the Classification of Overlapping Data Groups," *Electronics*, hal. 1-14, 2020.