

Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python

Risa Wati^{*1}, Siti Ernawati²

¹Program Studi Sistem Informasi Universitas Bina Sarana Informatika; Jl. Kramat Raya No.98, Senen, Jakarta Pusat 10450, telp. (021)23231170

²Program Studi Sistem Informasi Universitas Nusa Mandiri; Jl. Jatiwaringin Raya No.02, RT.08/RW.013, Kel. Cipinang Melayu, Kec. Makassar, Jakarta Timur, telp (021) 28534236
e-mail: ^{*}risawati.rwx@bsi.ac.id, ²siti.ste@nusamandiri.ac.id

Abstrak

Covid-19 merupakan penyakit menular melalui mulut dan hidung seseorang yang terinfeksi saat sedang berbicara, batuk maupun bersin dan menyebar secara luas didunia sehingga ditetapkan sebagai pandemi. Banyak upaya pemerintah yang dilakukan untuk menekan penyebaran Covid-19, salah satunya adalah penerapan PPKM untuk wilayah Jawa dan Bali. Pemberlakuan PPKM ini menimbulkan pro-kontra antar masyarakat, ada yang setuju dan ada yang tidak setuju diberlakukannya PPKM. Oleh sebab itu peneliti melakukan penelitian sentimen masyarakat terhadap pemberlakuan PPKM wilayah Jawa dan Bali. Komentar masyarakat diambil dari media sosial yaitu twitter berupa komentar positif dan negatif, kemudian data diolah menggunakan text editor Jupyter dan bahasa pemrograman Python serta menggunakan algoritma SVM. Penelitian ini memiliki tujuan apakah algoritma SVM dapat menjadi pengklasifikasi teks yang baik untuk analisis sentimen pemberlakuan PPKM, membandingkan Kernel pada SVM antara Kernel Linier dengan Kernel RBF, serta menilai apakah penerapan PPKM untuk wilayah Jawa dan Bali terbukti berhasil menekan angka penyebaran virus Covid-19. Algoritma SVM dengan kernel linier terbukti menjadi algoritma pengklasifikasi text yang baik pada analisis sentimen pemberlakuan PPKM wilayah Jawa dan Bali dengan nilai akurasi sebesar 86%. Serta Dilihat dari hasil analisis sentimen penerapan PPKM untuk wilayah Jawa dan Bali terbukti berhasil menekan angka penyebaran Covid-19.

Kata Kunci —PPKM, Algoritma SVM, Python

Abstract

Covid-19 is an infectious disease through the mouth and nose of someone who is infected when talking, coughing, or sneezing and spread widely in the world so set as a pandemic. Many governments attempt made to suppress the spread of Covid-19, one of them is the application of PPKM to Java and Bali. The implementation of PPKM raises the pros and cons between the community, there are agreed and there is no agreed enactment of PPKM. Therefore, researchers conducted the research community sentiment towards the implementation of PPKM Java and Bali. People's comments are taken from social media, namely Twitter in the form of positive and negative comments, then the data is processed using a text editor Jupyter and the Python programming language and use the algorithm of SVM. This research has a purpose whether the algorithm of SVM can be a classifier is a good text for sentiment analysis implementation of PPKM, compare the Kernel in SVM between the Linear Kernel with the RBF Kernel, as well as assess whether the application of PPKM to Java and Bali proved successful in suppressing the number of the spread of the virus Covid-19. The algorithm of SVM with the linear kernel proved to be the algorithm for classifying text on sentiment analysis implementation of PPKM Java and Bali with a value accuracy of 86%. As Seen from the results of the sentiment analysis application PPKM to Java and Bali proved to successfully reduce the number of Covid-19.

Keywords — PPKM, SVM Algorithm, Python

1. PENDAHULUAN

Penyakit Corona Virus Disease 2019 (Covid-19) adalah penyakit menular yang disebabkan oleh virus SARS-CoV-2. Virus dapat menyebar dari mulut atau hidung orang yang terinfeksi saat

sedang berbicara, batuk maupun bersin. Bukti menunjukkan bahwa virus dapat menyebar terutama diantara orang-orang yang melakukan kontak erat, berada dalam ruangan dengan ventilasi udara yang buruk atau dalam keramaian [1]. Covid-19 menyebar secara luas didunia sehingga Covid-19 ditetapkan sebagai pandemi. Banyak upaya pemerintah dilakukan untuk menahan tingkat penyebaran Covid-19 salah satunya yaitu Instruksi Menteri Dalam Negeri Nomor 42 Tahun 2021 Tentang Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) Level 4, Level 3, dan Level 2 di Wilayah Jawa dan Bali [2]. Pemberlakuan PPKM menimbulkan pro dan kontra pada masyarakat [3]. Banyak dari masyarakat yang menolak pemberlakuan PPKM karena dianggap memberatkan bagi masyarakat kecil, masyarakat pelaku UMKM maupun dibidang pariwisata, dan ada pula masyarakat yang setuju diberlakukannya PPKM karena dinilai efektif untuk menekan penyebaran Covid-19.

Saat ini media sosial menjadi bagian dari kehidupan sebagian besar masyarakat dan sebagai sumber utama untuk mengetahui tentang Covid-19 [4]. Salah satu tempat yang digunakan untuk menyampaikan tanggapan dan keluhan adalah dengan menggunakan media sosial, salah satunya yaitu twitter [5]. Twitter merupakan media sosial yang digunakan untuk mencari informasi tentang bisnis, hiburan, ekonomi, politik, dan lainnya [6]. Selain itu Twitter juga merupakan salah satu media sosial yang paling populer yang digunakan untuk sumber data pada analisis teks [7]. Twitter semakin menyerukan pandemi Covid-19 di dunia maya dengan munculnya hashtag mengenai pandemi Covid-19 sehingga menjadi trending topik [4].

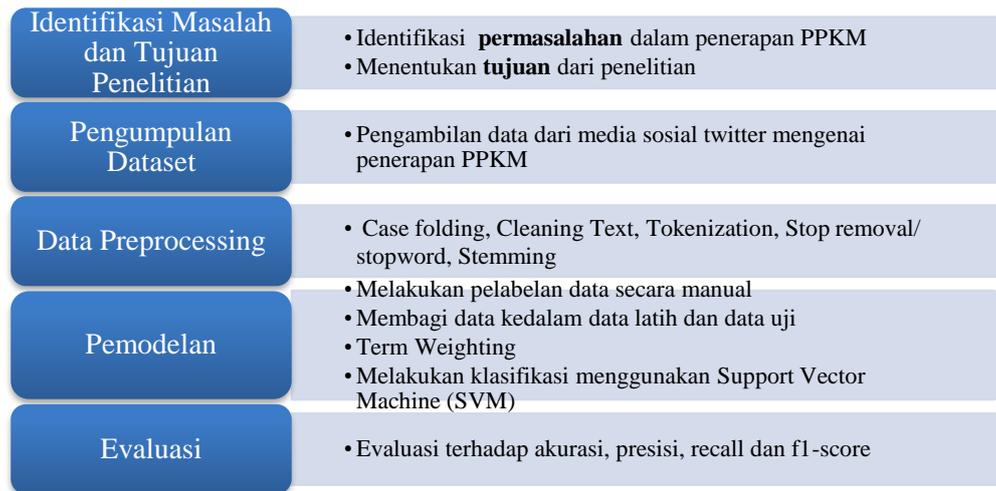
Analisis sentimen merupakan salah satu cara untuk mengumpulkan pendapat orang banyak terhadap sesuatu seperti layanan publik, isu, kinerja pemerintahan atau hal lainnya [8]. Analisis sentimen menjelaskan sentimen seseorang berupa sikap, pendapat, dan emosi terhadap produk, individu, topik, organisasi, maupun jasa [9]. Proses pelatihan pada analisis sentimen lebih sulit dari pada bidang machine learning lain karena data pada analisis sentimen bersifat subjektif yaitu berupa opini yang nilainya tidak kongkrit [7].

Pada penelitian ini, analisis sentimen pemberlakuan PPKM wilayah Jawa dan Bali dikumpulkan dari media sosial Twitter. diolah menggunakan text editor Jupyter dengan bahasa pemrograman Python serta menggunakan algoritma SVM. Jupyter merupakan text editor yang bersifat open source yang mendukung untuk semua bahasa pemrograman [10]. Algoritma SVM, Naïve Bayes, dan Neural Networks memiliki akurasi yang tinggi untuk analisis sentimen [11]. Penelitian memiliki tujuan apakah algoritma SVM dapat menjadi pengklasifikasi yang baik untuk sentimen pemberlakuan PPKM wilayah Jawa dan Bali, membandingkan Kernel pada SVM antara Kernel Linier dengan Kernel RBF (*Gaussian Radial basis Function*), serta penelitian ini dapat menilai apakah penerapan PPKM untuk wilayah Jawa dan Bali terbukti berhasil untuk menekan angka penyebaran Covid-19.

Berikut adalah beberapa penelitian terkait mengenai analisis sentimen yaitu Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berskala Besar Menggunakan Algoritma SVM, terbukti algoritma SVM berhasil dalam pengklasifikasian sentimen masyarakat [12], Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers [3], Analisis Sentimen Berdasarkan Komentar Youtube Menggunakan Support Vector Machine, pada penelitian ini menggunakan Lexicon Based dan Confusion Matrix untuk mengetahui hasil pembobotan dengan nilai True Positive sebesar 91.1% [13], Analisis Klasifikasi Sentimen Sekolah Daring pada Twitter Menggunakan SVM, model terbaik yang digunakan yaitu Logistic Regression dengan nilai akurasi sebesar 87%, nilai presisi sebesar 0,88, nilai recall sebesar 0,87 dan nilai f1-score sebesar 0,88 [5], dan Analisis Sentimen Layanan Indihome Berdasarkan Twitter Menggunakan Metode SVM, memperoleh akurasi sebesar 87% [6].

2. METODE PENELITIAN

Dalam melakukan penelitian terdapat beberapa tahapan diantaranya adalah sebagai berikut



Gambar 1. Tahapan Penelitian

2.1 Identifikasi Masalah dan Tujuan Penelitian

Permasalahan pada penelitian yaitu terjadinya pro-kontra mengenai sentimen masyarakat terhadap pemberlakuan PPKM wilayah Jawa dan Bali. Sedangkan Penelitian memiliki tujuan apakah algoritma SVM dapat menjadi pengklasifikasi yang baik untuk sentimen pemberlakuan PPKM wilayah Jawa dan Bali, membandingkan Kernel pada SVM antara Kernel Linier dengan Kernel RBF (*Gaussian Radial basis Function*), serta penelitian ini dapat menilai apakah penerapan PPKM untuk wilayah Jawa dan Bali terbukti berhasil untuk menekan angka penyebaran virus Covid-19.

2.2 Pengumpulan Dataset

Pada penelitian ini dataset diambil dari <https://netlytic.org> berupa komentar masyarakat pada media sosial yaitu twitter mengenai penerapan PPKM di wilayah Jawa dan Bali. Total data yang digunakan yaitu berjumlah 1514 data dengan kategori 757 kelas Positif dan 757 kelas Negatif. Data disimpan dalam bentuk CVS (*Comma Separated Values*). Dalam proses pengolahan data teks menggunakan aplikasi Python 3.9. Python adalah bahasa pemrograman interpretative yang mudah dipelajari, berfokus pada keterbacaan kode serta python bersifat multi platform [14]. Pada penelitian ini Text Editor yang digunakan adalah Jupyter. Jupyter dipilih karena bersifat open source dan mendukung disemua bahasa pemrograman [10]. Langkah selanjutnya pada proses penelitian yaitu memberikan label positif dan negatif terhadap sentimen secara manual, dalam proses ini diambil beberapa kata yang menjadi acuan dalam proses pelabelan. Tabel 1 merupakan kumpulan kata yang menjadi acuan dalam proses pelabelan.

Tabel 1. Kumpulan Kata dalam Proses Pelabelan

Positif	Negatif
Semangat	Lengah
Sembuh	Longgar
Sehat	Kendor
Disiplin	Pandemi
Vaksinansi	Kena

2.3 Data Preprocessing

Setelah dataset dikumpulkan, proses selanjutnya adalah preprocessing. Data yang diperoleh dari twitter tidak terstruktur serta banyak memiliki noise maka perlu dilakukan proses preprocessing. Preprocessing merupakan tahap awal untuk penyeleksian atau membuang kata yang tidak diperlukan pada teks komentar tweet [15]. Berikut adalah beberapa tahapan preprocessing.



Gambar 2. Proses Data Preprocessing

2.3.1 Cleaning Text

Pada proses cleaning text digunakan untuk membersihkan kata dari yang tidak diperlukan seperti @ (at) untuk tautan, url atau situs web seperti http, url, www, # (hashtag), RT (retweet) [16].

2.3.2 Case Folding

Mengkonversi keseluruhan teks pada dokumen menjadi suatu bentuk standar. Proses ini mengkonversi uppercase menjadi lowercase [17].

2.3.3 Tokenization

Pada proses tokenization digunakan untuk memisahkan kata atau huruf dari tanda baca dan simbol [17].

2.3.4 Stop Removal/ Stopword

Pada proses Stop Removal/ Stopword digunakan untuk menghilangkan kata yang dianggap tidak perlu dalam pengolahan data review [17].

2.3.5 Stemming

Proses stemming biasanya berisi pengenalan dan penghapusan prefix, sufiks dan pluralisasi yang tidak sesuai [16]

2.4 Pembagian Data Latih dan Data Uji

Data yang digunakan dalam penelitian ini berjumlah 1514. Dibagi kedalam data uji dan data latih. Dengan perbandingan jumlah data uji sebesar 152 dan data latih sebesar 1362.

2.5 Pemodelan

Melakukan term weighting atau pembobotan setiap dokumen. Term weight adalah pembobotan kata untuk menilai kata yang terkandung dalam sebuah dokumen [18]. Metode yang digunakan adalah TF-IDF (*Term Frequency-Inverse Document Frequency*) yaitu metode untuk memberikan bobot setiap kata untuk menentukan seberapa jauh keterhubungan kata terhadap dokumen [18]. Melakukan klasifikasi menggunakan Support Vector Machine (SVM). SVM adalah metode pembelajaran yang menganalisis data dan mengenali pola yang digunakan untuk klasifikasi [16]. SVM juga merupakan model yang berasal dari teori pembelajaran statistika yang memberikan hasil lebih baik dibandingkan metode lain [15]. Langkah selanjutnya yaitu melakukan eksperimen pada kernel diantaranya adalah Linear Kernel, Polymonical Function, Sigmoid, dan RBF (*Gaussian Radial basis Function*).

2.6 Evaluasi

Tahapan terakhir pada penelitian ini yaitu melakukan evaluasi terhadap nilai akurasi, presisi, recall dan f1-score. Semakin tinggi nilai akurasi menggunakan algoritma Support Vector Machine (SVM) berarti algoritma SVM terbukti mampu menjadi pengklasifikasi text yang baik untuk pengujian sentimen terhadap penerapan PPKM wilayah Jawa dan Bali.

3. HASIL DAN PEMBAHASAN

3.1. Data Preprocessing

Tahapan dasar yang dilakukan dalam sentimen analisis adalah preprocessing yang berguna untuk meningkatkan kinerja dan merupakan proses penting dalam penambangan data. Preprocessing yang dilakukan dalam penelitian ini diantaranya cleaning teks, case folding, tokenization, stop removal/ stopword dan stemming. Tabel 2 menunjukkan hasil preprocessing dari data yang digunakan.

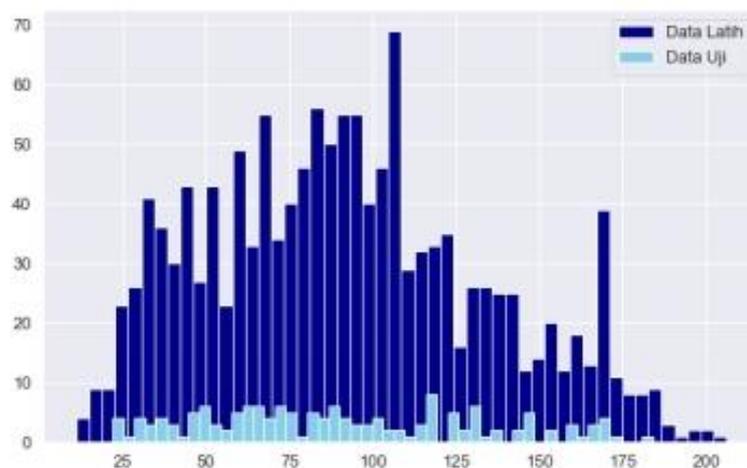
Tabel 2. Hasil Preprocessing

Data Tweet	Preprocessing	Hasil
level PPKM mungkin turun kawan, tapi semangat kita untuk melawan pandemi ini tidak boleh turun ya tetap harus taat prokes , kesehatan yang utama #PPKMlevelturun Prokes Jangan Kendor https://t.co/0LSjzITs2e	Case Folding	level ppkm mungkin turun kawan, tapi semangat kita untuk melawan pandemi ini tidak boleh turun ya tetap harus taat prokes , kesehatan yang utama #ppkmlevelturun prokes jangan kendor https://t.co/0lsjzits2e
	Cleaning Text	level ppkm mungkin turun kawan tapi semangat kita untuk melawan pandemi tidak boleh turun tetap harus taat prokes kesehatan yang utama prokes jangan kendor
	Tokenization	['level', 'ppkm', 'mungkin', 'turun', 'kawan', 'tapi', 'semangat', 'kita', 'untuk', 'melawan', 'pandemi', 'tidak', 'boleh', 'turun', 'tetap', 'harus', 'taat', 'prokes', 'kesehatan', 'yang', 'utama', 'prokes', 'jangan', 'kendor']
	Stop Removal/ Stopword	['level', 'ppkm', 'turun', 'kawan', 'semangat', 'melawan', 'pandemi', 'turun', 'taat', 'prokes', 'kesehatan', 'utama', 'prokes', 'kendor']
	Stemming	level ppkm turun kawan semangat lawan pandemi turun taat prokes sehat utama prokes kendor

3.2. Pemodelan

Proses yang dilakukan sebelum pemodelan adalah pemberian label terhadap data yang digunakan. Pelabelan dilakukan secara manual serta label yang digunakan yaitu positif dan negatif. Kemudian melakukan term weighting atau pembobotan setiap dokumen. Metode yang digunakan dalam pembobotan ini menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*). Pembobotan ini merupakan ukuran statistik untuk mengevaluasi kepentingan sebuah kata bagi dokumen [5].

Proses selanjutnya yaitu dilakukan pembagian data antara data latih dan data uji yaitu 90% data latih dan 10% data uji. Gambar 3 menunjukkan visualisasi pembagian data latih dan data uji yang digunakan dalam penelitian.



Gambar 3. Pembagian Data Latih dan Data Uji

Gambar 4 merupakan visualisasi wordcloud dari data penelitian yang digunakan. Visualisasi wordcloud ini dapat dimanfaatkan untuk menghitung frekuensi kata dalam teks [19]. Semakin besar kata yang muncul dalam teks maka semakin besar frekuensi kata tersebut muncul dalam sentiment

Gambar 5 menunjukkan confusion matriks dari hasil pengujian. Dari gambar tersebut dapat terlihat persentase dari setiap nilai *true positive*, *true negative*, *false positive* dan *false negative*. Dari keempat nilai tersebut dapat dihasilkan nilai akurasi dengan menggunakan rumus akurasi. Nilai akurasi yang diperoleh adalah 0.86 atau 86%.

Berikut ini merupakan hasil pengukuran.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$= \frac{44 + 42}{44 + 42 + 8 + 6}$$

$$= 0,86 = 86\%$$

$$Precision = \frac{TP}{TP + FP}$$

$$= \frac{44}{44 + 8}$$

$$= 0,85 = 85\%$$

$$Recall = \frac{TP}{TP + FN}$$

$$= \frac{44}{44 + 6}$$

$$= 0,88 = 88\%$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

$$= 2 * \frac{85 * 88}{85 + 88}$$

$$= 0,86 = 86\%$$

3.3. Evaluasi

Pada tahap pemodelan menggunakan SVM dengan kernel linear menghasilkan nilai akurasi sebesar 86%, Precision sebesar 85%, Recall sebesar 88% dan F1-Score sebesar 86%. Tahapan terakhir setelah pemodelan adalah evaluasi yang merupakan proses untuk menarik kesimpulan dari hasil penelitian yang telah dilakukan.

4. KESIMPULAN

1. Dataset yang digunakan dalam penelitian berupa tweet dengan jumlah tweet 757 kelas positif dan 757 kelas negatif.
2. Pengolahan data teks menggunakan aplikasi Python 3.9 dan Jupyter sebagai Text Editor.
3. Metode klasifikasi yang digunakan adalah SVM dengan membandingkan dua kernel yaitu kernel linear dan kernel RBF.
4. Pengukuran kinerja dari model menggunakan akurasi, precision, recall dan F1-score.
5. Hasil analisis, pengujian dan eksperimen telah dilakukan dalam penelitian ini, sehingga dapat ditarik kesimpulan bahwa metode SVM dapat mengklasifikasikan teks dengan baik.
6. Dilihat dari hasil analisis sentimen penerapan PPKM untuk wilayah Jawa dan Bali terbukti berhasil untuk menekan angka penyebaran virus Covid-19.

DAFTAR PUSTAKA

- [1] 2021, Coronavirus disease (COVID-19). [Online]. Available: https://www.who.int/health-topics/coronavirus#tab=tab_1. [Accessed: 04-Oct-2021].
- [2] 2021, Instruksi Menteri Dalam Negeri Nomor 42 Tahun 2021. [Online]. Available: <https://covid19.go.id/p/regulasi/instruksi-menteri-dalam-negeri-nomor-42-tahun-2021>. [Accessed: 04-Oct-2021].
- [3] Krisdiyanto T, 2021 Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.* **7**, 1 p. 32–37.
- [4] Fauziyyah A K, 2020 Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python *J. Ilm. SINUS* **18**, 2 p. 31.
- [5] Savitri N L P C Rahman R A Venyutzky R and Rakhmawati N A, 2021 Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning *J. Tek. Inform. dan Sist. Inf.* **7**, 1 p. 47–58.
- [6] Tineges R Triayudi A and Sholihati I D, 2020 Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM) *J. Media Inform. Budidarma* **4**, 3 p. 650.
- [7] Ferdiana R Jatmiko F Purwanti D D Ayu A S T and Dicka W F, 2019 Dataset Indonesia untuk Analisis Sentimen *J. Nas. Tek. Elektro dan Teknol. Inf.* **8**, 4 p. 334.
- [8] Suryono S and Taufiq Luthfi E, 2021 Analisis sentimen pada Twitter dengan menggunakan metode Naïve Bayes Classifier *Jnanaloka* p. 81–86.
- [9] Oktavianto, Andri and S F P, 2021 Persepsi Publik Tentang Pembelajaran Daring di Indonesia: Studi Menggunakan ELK Stack dan Python untuk Analisis Sentimen di Twitter *J. Tek. ITS* **9**, 2 p. 170–175.
- [10] 2021, About Jupyter. [Online]. Available: <https://jupyter.org/about>. [Accessed: 05-Oct-2021].
- [11] Mehta P and Pandya S, 2020 A review on sentiment analysis methodologies, practices and applications *Int. J. Sci. Technol. Res.* **9**, 2 p. 601–609.
- [12] Tuhuteru H, 2020 Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine *Inf. Syst. Dev.* **5**, 2 p. 7–13.
- [13] Tanesab F I Sembiring I and Purnomo H D, 2017 Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine *Int. J. Comput. Sci. Softw. Eng.* **6**, 8 p. 180–185.
- [14] JUD, 2016 *Pemrograman Python untuk Pemula* Jubilee Enterprise.
- [15] Darwis D Pratiwi E S and Pasaribu a F O, 2020 Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia *Eduitic - Sci. J. Informatics Educ.* **7**, 1 p. 1–11.
- [16] Basari A S H Hussin B Ananta I G P and Zeniarja J, 2013 Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization *Procedia Eng.* **53** p. 453–462.
- [17] Ernawati S Wati R Nuris N Marita L S and Yulia E R, 2020 Comparison of Naïve Bayes Algorithm with Genetic Algorithm and Particle Swarm Optimization as Feature Selection for Sentiment Analysis Review of Digital Learning Application *J. Phys. Conf. Ser.* **1641**, 1.
- [18] Deolika A Kusri K and Luthfi E T, 2019 Analisis Pembobotan Kata Pada Klasifikasi Text Mining *J. Teknol. Inf.* **3**, 2 p. 179.
- [19] Fahmi A Ramadhan I Studi P Informasi S and Komputer F I, 2020 Analisis Sentiment Masyarakat Selama Bulan Ramadhan Dalam Menghadapi Pandemi Covid-19 *J. Inform. dan Sist. Inf.* **1**, 1 p. 608–617.
- [20] 2021, Support vector machines (SVMs). [Online]. Available: <https://scikit-learn.org/stable/modules/svm.html>. [Accessed: 12-Oct-2021].
- [21] Permana D S and Silvanie A, 2021 Prediksi Penyakit Jantung Menggunakan Support Vector Machine dan Python Pada Basis Data Pasien **2**, 1 p. 29–34.
- [22] Sun M, 2014, Support Vector Machine for Classification, in *Encyclopedia of Business Analytics and Optimization*, p. 2395–2409.