

## **PENERAPAN ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK KLASIFIKASI SENTIMEN VAKSIN BOOSTER PADA TWITTER**

**Lutfi Fauzi<sup>1</sup>, Moh. Dasuki<sup>2</sup>, Lutfi Ali Muharom<sup>3</sup>**

<sup>1</sup> Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember, Indonesia

<sup>2</sup> Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember, Indonesia

<sup>3</sup> Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember, Indonesia

### **Info Artikel**

#### **Riwayat Artikel:**

Diterima : **02-09-2023**

Direvisi : **16-10-2023**

Disetujui : **10-12-2023**

#### **Kata Kunci:**

*Support Vector Machine,*

*Vaksin Booster,*

*Sentimen analisis,*

*Twitter*

### **ABSTRAK**

Rendahnya cakupan vaksin *booster* di Indonesia dipengaruhi oleh banyak faktor dan menuai pro kontra dari masyarakat Indonesia. Cukup banyak masyarakat yang tidak mau melakukan vaksin *booster*. Hal ini dikarenakan kalangan masyarakat Indonesia masih meragukan keamanan, efektivitas dan juga kemampuan vaksin *booster*. Pro dan kontra tentang vaksin *booster* ini salah satunya tertuang pada media sosial twitter. Masyarakat mengungkapkan opininya pada media twitter dengan menggunakan *hashtag*. Studi komputasional dari opini-opini orang yang menjadi satu disebut dengan sentiment analisis atau *opinion minning*. Sentiment analisis merupakan sebuah teknik atau cara yang digunakan untuk mengidentifikasi bagaimana sebuah sentiment diekspresikan sebagai sentiment positif maupun sentiment negatif. Teknik yang digunakan untuk mengklasifikasikan sentiment analisis tersebut yaitu *Support Vector Machine (SVM)*. Tujuan dari penelitian ini adalah menerapkan metode algoritma *Support Vector Machine (SVM)* untuk mengklasifikasikan sentiment vaksin booster pada social media twitter. Hasil dari penelitian ini diperoleh pada proses K-fold cross validation nilai akurasi tertinggi yang dihasilkan menggunakan metode *Support Vector Machine (SVM)* berdasarkan langkah uji ketiga di KF=10 yaitu 84% dan nilai akurasi terendah yaitu 52% dengan menggunakan KF=10 pada langkah uji ini mendapat nilai akurasi sebesar 73% dan lebih tinggi daripada langkah uji yang menggunakan K-fold 2 dan K-fold 5.

### **Keywords:**

*Support Vector Machine,*

*Booster Vaccine,*

*Sentiment Analysis,*

*Twitter*

### **ABSTRACT**

*The low coverage of booster vaccines in Indonesia is influenced by many factors and has drawn agree and disagree from Indonesian society. Quite a lot of people don't want to take the booster vaccine. This is because Indonesian people still underestimate the safety, effectiveness and efficacy of booster vaccine. One of the agree and disagree about this booster vaccine is stated on twitter. People express their opinions on twitter using hashtags. The computational study of people's opinions coming together is called sentiment analysis or opinion minning. Sentiment analysis is a technique or method used to identify how a sentiment gives rise to positive sentiment or negative sentiment. The technique used to classify sentiment analysis is support vector machine (SVM). The purpose of this research is apply the Support Vector Machine (SVM) algorithm method to classify booster vaccine sentiment on twitter. The result of this research were obtained from K-fold cross validation process, the highest accuracy value was produced using the Support Vector Machine (SVM) method based on the third test step at KF=10, namely 84% and the lowest accuracy value was 52% using KF=10 in the step. This test received an accuracy value of 73% and was higher than the test steps that used k-fold 2 and k-fold 5.*

### **Penulis Korespondensi:**

Lutfi Fauzi

Program Studi Teknik Informatika

Universitas Muhammadiyah Jember

Email: fauzilutfi147@gmail.com

## 1. PENDAHULUAN

Vaksin *booster* merupakan vaksin yang diberikan kepada seseorang yang telah lengkap vaksinasi primernya, namun seiring berjalannya waktu kadar antibodi akan perlahan menurun sehingga harus mendapatkan *booster*. Program vaksin *booster* ini bermaksud meningkatkan kembali kadar antibodi yang semakin hari semakin menipis sehingga kekebalan tubuh saat terserang infeksi covid 19 meningkat. Cakupan vaksin dosis 1 yaitu 202.144.603 dosis (97,06%), dosis 2 yaitu 169.759.138 dosis (81,51%), dan vaksin *booster* yaitu 54.148.579 (26,00%) [4].

Rendahnya cakupan vaksin *booster* ini dipengaruhi oleh banyak faktor dan menuai pro kontra dari masyarakat Indonesia. Cukup banyak masyarakat yang tidak mau melakukan vaksin *booster*. Hal ini dikarenakan kalangan masyarakat Indonesia masih meragukan keamanan, efektivitas, dan juga kemampuan vaksin *booster*. Ditambah lagi dengan menyebarnya berita hoax mengenai vaksin *booster*, membuat masyarakat Indonesia kurang percaya terhadap program vaksinasi *booster* yang dilakukan oleh pemerintah. Penelitian ini berfokus pada opini masyarakat tentang vaksin *booster* yang tertuang dalam mediasosial salah satunya media twitter.

Studi komputasional dari opini-opini orang yang menjadi satu disebut dengan *sentiment* analisis atau *Opinion Mining*, *sentiment* dan emosi melalui entitas atau atribut yang dimiliki yang diekspresikan dalam bentuk teks [1]. Teknik yang digunakan pada *sentiment* analisis dalam mengklasifikasikan tersebut yaitu *Support Vector Machine* (SVM). Pengenalan tulisan tangan dan kategorisasi teks dengan akurasi yang tinggi merupakan hasil dari SVM yang sangat menjanjikan. *Support Vector Machine* (SVM) juga dapat digunakan pada data berdimensi banyak dan menghindari kesulitan dari permasalahan dimensionalitas [2].

. Pada penelitian terdahulu berjudul "Penerapan Alogaritma SVM Untuk Analisis Sentimen Pada Twitter Komisi Pemberantasan Korupsi Republik Indonesia" menggunakan metode klasifikasi SVM dan metode TF-IDF yakni sebanyak 2000 data hasil crawling, penelitian ini menghasilkan 1890 data dan 3846 term/kata dari hasil *preprocessing* kemudian diakumulasi nilai dari kata yang muncul untuk *labelling* yang menghasilkan *sentiment* positif, negatif dan netral. Berdasarkan hasil uji, penerapan metode SVM mendapatkan hasil sebesar 82% dan hasil *sentiment* negatif dengan jumlah 77%, label positif 8% dan label netral 25%. [3]

Berdasarkan latar belakang diatas, akan dilakukan penelitian dengan judul 'Penerapan Alogaritma *Support Vector Machine* (SVM) Untuk Klasifikasi Sentimen Vaksin *Booster* Pada Sosial Media Twitter'.

## 2. METODE PENELITIAN

Tahapan yang dilakukan pada penelitian ini adalah menghilangkan teks, *Cleansing*, *case folding*, *filtering*, *stemming*, normalisasi bahasa dan juga *tokenizing*. Pengumpulan data dihimpun dari cuitan pengguna di twitter tentang Vaksin Booster. Data dibagi mejnadi dua yaitu sentimen positif dan sentimen negatif. Sejumlah 250 data akan digunakandengan metode *Support Vector Machine*. Data yang diambil dari penelitian ini adalah data berbahasa Indonesia.

### 2.1 Eksperimen dan pengujian metode

Setelah melewati tahap *preprocessing text*, tahap selanjutnya yaitu melakukan ekstraksi fitur atau pembobotan pada setiap term yang terdapat pada komentar. Perangkingan dokumen dilakukan jika bobot setiap kata dilakukan dengantepat. Proses pembobotan kata yang dipilih pada penelitian ini menggunakan Metode TF-IDF karena metode tersebutterkenal efisien, mudah dan memiliki hasil akurasi yang tinggi.

### 2.2 Evaluasi dan Validasi

Evaluasi bertujuan menyediakan informasi tentang seberapa jauh kegiatan tercapai, dan terdapatperbedaan pencapaian dengan suatu standar tertentu untuk mengetahui selisih manfaat yang dikerjakan dan dibandingkan dengan ekspektasi yang diperoleh. *K-fold Cross Validation* digunakan dengan tujuan akhiruntuk menggeneralisasikan model atau independen.

## 3. HASIL DAN ANALISIS

Pada bagian ini, membahas implementasi tahapan-tahapan yang digunakan oleh peneliti. Pada bagian ini juga menjelaskan hasil pengumpulan data dan analisis yang dilakukan sesuai dengan metodologi penelitian. Hasil analisis akan dibahas dalam bab ini berdasarkan pengolahan data tweet dari Twitter yang telah di filter berdasarkan keyword "#vaksinbooster".

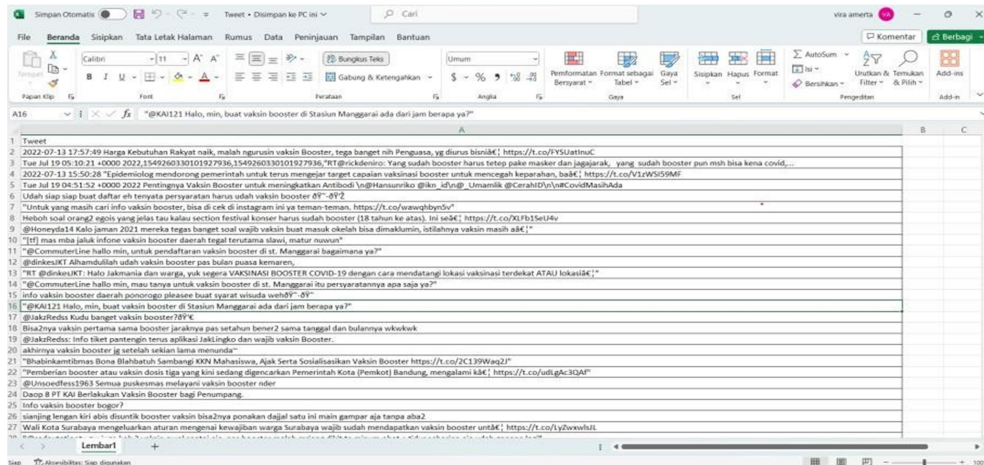
### 3.1 Analisis

Analisis dimulai dengan melakukan tahapan penambangan data/*crawling*, filter data, pelabelan, *text pre-processing*, pembobotan kata TF-IDF, klasifikasi *Support Vector Machine*, evaluasi dan validasi. Penelitian ini dilakukan dengan Google Colab. Kemudian spesifikasi hardware yang

digunakan dalam penelitian ini yaitu dengan processor Intel(R) Core i5-1035G1 dan RAM 4GB, Operating System Windows 1164-bit.

### 3.2 Penambahan Data

Dalam penelitian ini dilakukan dengan cara crawling. Crawling data twitter dengan menggunakan bahasa pemrograman python pada tools google colab. Kemudian data tweet dengan keyword “#vaksinbooster”. Proses pengumpulan data tweet hanya mengambil data yang berbentuk teks berbahasa indonesia, yang kemudian data hasil crawling akan disimpan dalam bentuk ekstensi excel (.xls). Hasil dari crawling data adalah sebagai berikut.



Gambar 1. Hasil crawling data

### 3.3 Filter Data

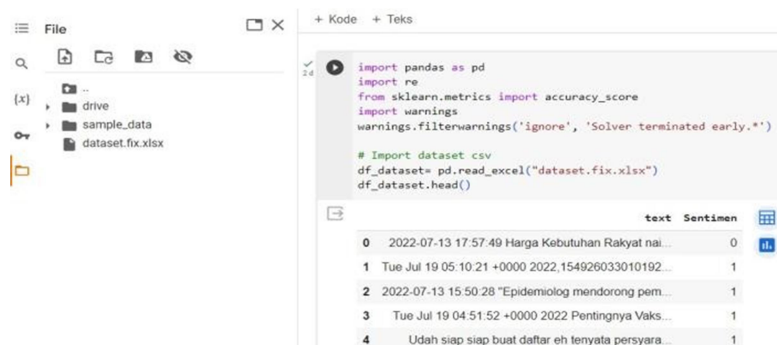
Setelah melakukan crawling data maka selanjutnya dilakukan filter secara manual untuk memilah data mentah dengan memilah tweet yang menggunakan Bahasa selain Bahasa Indonesia dan selanjutnya siap untuk melakukan pelabelan

### 3.4 Pelabelan

Pelabelan dilakukan pada 250 data dan memberikan kelas pada masing-masing data yaitu kelas positif “1” dan negative “0”. Selanjutnya divalidasi oleh salah satu guru Bahasa Indonesia SMK Muhammadiyah 8 Siliragung.

### 3.5 Preprocessing Text

Selanjutnya untuk melakukan proses text pre-processing pada python data set akan di upload terlebih dahulu kedalam cloud drive Google colab dengan cara berikut:



Gambar 2. Upload data set ke drive

Selanjutnya untuk import data dalam program python memanfaatkan library pandas untuk code import dataset. Setelah data di import, label pada set yaitu “1” dan “0” yang dimana melambangkan nilai 1 adalah positif dan 0 melambangkan negatif.

	text	Sentimen
0	2022-07-13 17:57:49 Harga Kebutuhan Rakyat nai...	0
1	Tue Jul 19 05:10:21 +0000 2022,154926033010192...	1
2	2022-07-13 15:50:28 "Epidemiolog mendorong pem...	1
3	Tue Jul 19 04:51:52 +0000 2022 Pentingnya Vaks...	1
4	Udah siap siap buat daftar eh ternyata persyara...	1

Gambar 3. Data set hasil import

Kemudian proses text pre-processing. Pada program text pre-processing diawali dengan penghapusan atribut yang tidak diperlukan yang berasal dari Twitter seperti emoticon, link, dan sebagainya. Berikut hasil dari cleansing.

Tabel 1. Hasil Cleansing

No	Text	Cleansing
1	Harga Kebutuhan Rakyat naik, malah ngurusin vaksin Booster, tega banget nih Penguasa	Harga Kebutuhan Rakyat naik malah ngurusin vaksin Booster tega banget nih Penguasa
2	Yang sudah booster harus tetep pakemasker dan jaga jarak, yang sudah booster pun msh bisa kena covid	Yang sudah booster harus tetep pake masker dan jaga jarak yang sudah booster pun msh bisa kena covid
3	Epidemiolog mendorong pemerintah untuk terus mengejar target capaian vaksinasi booster untuk mencegah keparahan	Epidemiolog mendorong pemerintah untuk terus mengejar target capaian vaksinasi booster untuk mencegah keparahan
4	Pentingnya vaksin booster untuk meningkatkan antibodi	Pentingnya vaksin booster untuk meningkatkan antibodi
250	Lama tak dengar corona. Apa kabar? Apa kabar juga para pengusaha dan sales vaksin? Bisnis booster dan PCR nya.	Lama tak dengar corona Apa kabar Apa kabar juga para pengusaha dan sales vaksin bisnis booster dan PCR nya

Setelah tahap cleansing selanjutnya dilakukan tahap casefolding. Berikut hasil dari proses casefolding:

Tabel 2. Hasil casefolding

No	Cleansing	Casefolding
1	Harga Kebutuhan Rakyat naik malah ngurusin vaksin Booster tega banget nihsePenguasa	harga kebutuhan rakyat naik malah ngurusin vaksin booster tega banget nih penguasa
2	Yang sudah booster harus tetep pake masker dan jaga jarak yang sudah booster pun msh bisa kena covid	yang sudah booster harus tetep pake masker dan jaga jarak yang sudah booster pun msh bisa kena covid
3	Epidemiolog mendorong pemerintah untuk terus mengejar target capaian vaksinasi booster untuk mencegah keparahan	epidemiolog mendorong pemerintah untuk terus mengejar target capaian vaksinasi booster untuk mencegah keparahan
4	Pentingnya Vaksin Booster untuk meningkatkan antibodi	Pentingnya vaksin booster untuk meningkatkan antibodi
250	Lama tak dengar corona Apa kabar Apa kabar juga para pengusaha dan salesvaksin bisnis booster dan PCR nya	Lama tak dengar corona apa kabar apa kabar juga para pegusaha dan sales vaksin bisnisbooster dan pcr nya

Selanjutnya dilanjutkan stemming, berikut hasil dari stemming:

Tabel 3. Hasil Stemming

No	Casefolding	Stemming
1	harga kebutuhan rakyat naik malah ngurusin vaksin booster tega banget nihpenguasa yang sudah booster harus tetap pake	harga butuh rakyat naik malah ngurusin vaksinbooster tega banget nih kuasa  yang sudah booster harus tetap pake masker
2	masker dan jaga jarak yang sudahbooster pun msh bisa kena covid	dan jaga jarak yang sudah booster pun mshbisa kena covid
3	epidemiolog mendorong pemerintah untuk terus mengejar target capaian vaksinasi booster untuk mencegah keparahan	epidemiolog dorong perintah untuk terus kejar target capai vaksinasi booster untuk cegah parah
4	Pentingnya vaksin booster untuk meningkatkan antibodi	penting vaksin booster untuk tingkat antibodi
250	Lama tak dengar corona apa kabar apakah juga parapengusaha dan sales vaksin bisnis booster dan pcr nya	lama tak dengar corona apa kabar apa kabar juga para usaha dan sales vaksin bisnis booster dan pcr nya

Kemudian dilanjutkan tahap stopword removal dimana menghapus kata-kata yang tidak memilikimaknya. Berikut hasil dari filtering/stopword removal:

Tabel 4. Hasil Filtering/Stopword Removal

No	Stemming	Filtering/Stopword Removal
1	harga butuh rakyat naik malah ngurusinvaksin booster tega banget nih kuasa	harga butuh rakyat ngurusin vaksin boostertega banget nih kuasa
2	yang sudah booster harus tetap pake masker dan jaga jarak yang sudah booster pun msh bisa kena covid	booster tetap pake masker jaga jarak boostermsh covid
3	epidemiolog dorong perintah untukterus kejar target capai vaksinasi booster untuk cegah parah	epidemiolog dorong perintah kejar target capaivaksinasi booster cegah parah
4	penting vaksin booster untuk tingkatantibodi	vaksin booster tingkat antibodi
250	lama tak dengar corona apa kabar apa kabar juga para usaha dan sales vaksin bisnis booster dan pcr nya	dengar corona kabar kabar usaha sales vaksinbisnis booster pcr

Tabel 5. Hasil Normalisasi Bahasa

No	Normalisasi Bahasa
1	harga butuh rakyat ngurusin vaksin booster tega banget nih kuasa
2	booster tetap pake masker jaga jarak booster msh covid
3	epidemiolog dorong perintah kejar target capai vaksinasi booster cegah parah
4	vaksin booster tingkat antibodi
250	dengar corona kabar kabar usaha sales vaksin bisnis booster pcr

Tabel 3.6 Hasil Tokenizing

No	Tokenizing
1	['harga', 'butuh', 'rakyat', 'ngurusin', 'vaksin', 'booster', 'tega', 'banget', 'nih', 'kuasa']
2	['booster', 'tetep', 'pake', 'masker', 'jagajarak', 'booster', 'msh', 'covid']
3	['epidemiolog', 'dorong', 'perintah', 'kejar', 'target', 'capai', 'vaksinasi', 'booster', 'cegah', 'parah']
4	['vaksin', 'booster', 'tingkat', 'antibodi']
250	['dengar', 'corona', 'kabar', 'kabar', 'usaha', 'sales', 'vaksin', 'bisnis', 'booster', 'pcr']

### 3.6 Pebobotan kata dengan TF-IDF

Setelah komentar melalui tahap pre-processing, selanjutnya adalah proses pembobotan kata dengan menggunakan TF-IDF. Berikut hasil perhitungan TF-IDF:

Tabel 3.7 Hasil Perhitungan TF-IDF

No	Text	TF-IDF
1	Harga butuh, rakyat, ngurusin, vaksin, booster, tega, banget, nih, kuasa, urus, bisnis	0.004409320166969666, 0.006434060025389578, 0.05595531271586845, 0.05720145491667862, 0.0, 0.22946795841216072, 0.0, 0.24864676324227944, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
2	booster, tetep, pake, masker, jaga, jarak, booster, masih, covid	[0.011023300417424164, 0.0, 0.0, 0.0, 0.1148605159558928, 0.0,

[illegible]

		0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
--	--	------------------------------------

Kemudian setelah mendapatkan nilai TF-IDF, untuk tahap selanjutnya mencoba mengambil sentiment dimana sentiment tersebut akan menguji coba dengan perhitungan manual setelah nilai sentiment tersebut dipilih, selanjutnya melakukan proses vektorisasi, proses ini adalah untuk memberikan sebuah nilai pada kata pada suatu dokumen sehingga ketika dokumen tersebut dapat diolah dan pola dari kata tersebut mudah dicari ketika masuk dalam klasifikasi. Proses dimana klasifikasi dilakukan pada data tersebut untuk membuat model yang nantinya model ini akan digunakan untuk pengujian data. Berikut hasil nilai perhitungan keputusan:

Tabel 3.8 Hasil Perhitungan Keputusan

Perhitungan Fungsi Keputusan		
	Sentimen Awal	Sentimen Akhir
D1	-1	-1
D2	1	1
D3	1	1
D4	1	1
D5	-1	-1

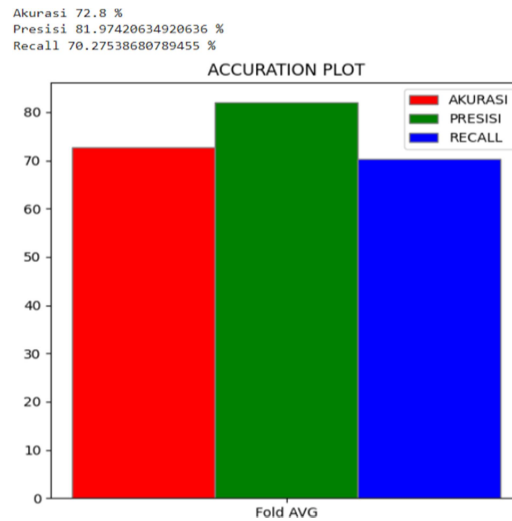
### 3.7 Evaluasi dan Validasi

Selanjutnya pada tahap evaluasi dilakukan untuk mengetahui hasil akurasi, waktu pemrosesan dan kinerja dari algoritma Support Vector Machine. Kemudian validasi dilakukan untuk mengetahui hasil akurasi dari algoritma Support Vector Machine. Berikut nilai akurasi Support Vector Machine:

Tabel 3.9 Hasil Nilai Akurasi Support Vector Machine (SVM)

K-fold cross	Langkah Uji	Support Vector Machine (SVM) (Data Training)
2-fold	Langkah Uji 1	62%
	Langkah Uji 2	59%
	Langkah Uji 1	58%
5-fold	Langkah Uji 2	82%
	Langkah Uji 3	80%
	Langkah Uji 4	74%
	Langkah Uji 5	62%
	Langkah Uji 1	52%
10-fold	Langkah Uji 2	84%
	Langkah uji 3	84%
	Langkah uji 4	80%
	Langkah uji 5	72%
	Langkah uji 6	72%
	Langkha uji 7	80%
	Langkah uji 8	68%
	Langkah uji 9	68%
	Langkah uji 10	68%

Grafik hasil dari confusion matrix dapat dilihat pada gambar berikut:



Gambar 3. Hasil Confusion Matrix

#### 4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan oleh penulis diperoleh kesimpulan. Bahwa pada penelitian ini diperoleh pada proses K-fold cross validation nilai akurasi tertinggi yang dihasilkan oleh metode Support Vector Machine (SVM) berdasarkan langkah uji ketiga di KF = 10 yaitu 84% dan nilai akurasi terendah yaitu 52% dengan menggunakan KF = 10 pada langkah uji ini mendapat nilai akurasi sebesar 73% dan lebih tinggi dari pada langkah uji yang menggunakan k-fold 2 dan k-fold 5.

#### REFERENSI

- [1] Aditya, Syakuro 2017. *Sentimen Analisis dalam Open Mining*. Jurnal Akutansi Pemerintah: Vol.2.No.1.PP.79-91
- [2] Darmawan. 2018. *Optimasi Algoritma SVM*. Coorporate Governance: An International Review. Vol.14.No.2. PP 107-125
- [3] Darwis, Dedi dkk. 2020. *Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia*. Jurnal Ilmiah Edutic. Vol.7,No.1
- [4] Komite Covid. *Efektivitas Pelaksanaan Kegiatan Satgas Covid-19*. Jakarta: Rineka Cipta. 2022