

## ANALISIS SENTIMEN RESPON MASYARAKAT TERHADAP KABAR HARIAN COVID-19 PADA TWITTER KEMENTERIAN KESEHATAN DENGAN METODE KLASIFIKASI NAIVE BAYES

Eni Tri Handayani<sup>1</sup>, Ari Sulistiyawati<sup>2</sup>

Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia<sup>1</sup>  
Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia<sup>2</sup>

[enitrihandayani2@gmail.com](mailto:enitrihandayani2@gmail.com)<sup>1</sup>, [ari\\_sulistiyawati@teknokrat.ac.id](mailto:ari_sulistiyawati@teknokrat.ac.id)<sup>2</sup>

**Received:** (9 September 2021) **Accepted:** (16 September 2021) **Published:** (30 September 2021)

### Abstract

*The Ministry of Health of the Republic of Indonesia acts as a proprietary group in handling COVID-19 in Indonesia, always providing daily information to provide information on cases of the Covid-19 pandemic every day through the Twitter account of the Ministry of Health of the Republic of Indonesia. However, not all daily tweets provided by the twitter account of the Ministry of Health of the Republic of Indonesia can be consumed and responded to harmoniously by the public. This problem can be solved by conducting research in the field of Sentiment Analysis, which is a field of research that focuses on the computational study of opinions, behavior, and emotions towards an entity that is expressed in text form. This study was conducted to find out how the results of sentiment analysis related to the public response to the daily news of Covid-19 from the twitter of the Ministry of Health of the Republic of Indonesia and classify them into three classes, namely positive, negative, and neutral using the Naïve Bayes Classifier classification method, so that this research is expected to be able to generate sentiment with classes that tend to be positive, negative, or neutral. The results of the conclusions of this study can be seen in the form of graphs. This study also carried out accuracy testing, precision testing, and recall, a score of f-1 to ensure the accuracy of the study.*

**Keywords:** Covid-19, Kemenkes RI, Sentiment Analysis, Text Mining, Naive Bayes Classifier.

### Abstrak

Kementerian Kesehatan Republik Indonesia berperan sebagai gugus terdepan dalam penanganan covid-19 di Indonesia selalu menyajikan kabar harian untuk menyediakan informasi mengenai kasus pandemi covid-19 perharinya melalui akun twitter milik Kementerian Kesehatan Republik Indonesia. Namun, tidak semua tweet harian yang diberikan oleh akun twitter Kementerian Kesehatan Republik Indonesia dapat dikonsumsi dan direspon masyarakat dengan selaras. Adapun masalah ini dapat diatasi dengan melakukan penelitian di bidang Analisis Sentiment, yang mana merupakan bidang penelitian yang berfokus kepada studi komputasi atas opini, tingkah laku, dan emosi terhadap suatu entitas yang dituangkan dalam bentuk teks. Penelitian ini dilakukan untuk mengetahui bagaimana hasil analisis sentimen terkait respon masyarakat dari kabar harian Covid-19 dari twitter Kementerian Kesehatan Republik Indonesia dan mengklasifikasikannya menjadi tiga kelas yaitu positif, negatif, dan netral menggunakan metode klasifikasi Naïve Bayes Classifier, sehingga penelitian ini diharapkan dapat menghasilkan sentimen dengan kelas yang cenderung positif, negatif, atau netral. Hasil dari kesimpulan penelitian ini nantinya dapat dilihat dalam bentuk grafik. Penelitian ini juga melakukan pengujian akurasi, pengujian precision, dan recall, f-1 score untuk memastikan keakuratan dari penelitian.

**Kata Kunci:** Covid-19, Kemenkes RI, Analisis Sentiment, Text Mining, Klasifikasi Naive Bayes.

**To cite this article:**

Handayani, Sulistiyawati (2021). Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan Dengan Metode Klasifikasi Naive Bayes. Vol (2), No. 3, 32 - 37

## 1. Pendahuluan

Informasi yang sedang menjadi perbincangan utama akhir-akhir ini berasal dari masyarakat dunia yang sedang di gemparkan dengan adanya pandemi covid-19 (Corona Virus Disease 2019) yang tengah berlangsung sejak awal maret lalu hingga saat ini, pandemi ini tidak hanya mewabah di Indonesia tetapi juga menyebar hampir di seluruh penjuru dunia [1] [2]. Belum tuntasnya pelaksanaan pemutusan rantai penyebaran virus covid-19 dari pemerintah hingga saat ini, tentu membuat kalangan masyarakat menjadi resah, was-was hingga ketakutan dengan adanya pandemi ini [19] [3]. Seiring dengan terus berlanjutnya pandemi covid-19 hingga saat ini, beragam cara telah dilakukan oleh pemerintah hingga kalangan masyarakat demi memutus rantai penyebaran virus covid-19.

Kementerian Kesehatan Republik Indonesia berperan sebagai gugus terdepan dalam penanganan covid-19 di Indonesia selalu menyajikan kabar harian untuk menyediakan informasi mengenai kasus pandemi covid-19 perharinya melalui akun twitter milik Kementerian Kesehatan Republik Indonesia [9]. Namun, tidak semua tweet harian yang diberikan oleh twitter Kementerian Kesehatan Republik Indonesia dapat dikonsumsi dan direspon masyarakat dengan selaras. Akun twitter milik Kementerian Kesehatan Republik Indonesia dibanjiri dengan beragam komentar yang muncul dari kalangan masyarakat pengguna twitter, komentar yang diberikan oleh pengguna twitter ini merupakan data teks yang dapat digali kembali, diolah, dan dimanfaatkan dan dijadikan sebagai bahan penelitian untuk berbagai keperluan di bidang pengembangan informasi.

Penelitian ini menggunakan data teks berbahasa indonesia dari akun twitter Kementerian Kesehatan Republik Indonesia dengan metode klasifikasi Naive Bayes Classifier. Klasifikasi sendiri di definisikan sebagai proses untuk menyatakan objek data sebagai suatu kategori atau kelas yang sebelumnya telah di definisikan [18] [4] [5]. Berdasarkan pemaparan masalah tersebut, penelitian ini dilaksanakan untuk mengetahui bagaimana hasil analisis sentimen terkait respon masyarakat dari kabar harian Covid-19 dari twitter Kementerian Kesehatan Republik Indonesia dan mengklasifikasikannya menjadi tiga kelas yaitu positif, negatif, dan netral menggunakan metode klasifikasi Naive Bayes Classifier, sehingga penelitian ini diharapkan dapat menghasilkan sentimen dengan kelas yang cenderung positif, negatif, atau netral. Hasil dari kesimpulan penelitian ini nantinya dapat dilihat dalam bentuk grafik.

Penelitian ini juga melakukan pengujian akurasi, pengujian precision, dan recall, f-1 score untuk memastikan keakuratan dari penelitian.

## 2. Tinjauan Pustaka

### 2.1. Twitter

Media sosial twitter merupakan satu dari beberapa bagian perkembangan media komunikasi yang diciptakan dengan tujuan agar pengguna bisa memberikan ekspresi, pendapat, aspirasi, kritik, serta bertukar informasi mengenai suatu informasi yang sedang menjadi perbincangan utama, tanpa keterbatasan waktu dan ruang. Sehingga, pendapat tersebut dapat tersampaikan secara langsung, kegiatan mem-posting pendapat dalam twitter ini disebut sebagai istilah tweet [6] [6] [7]. Hal tersebut menjadikan twitter sebagai salah satu sumber data text yang dapat digali dan dimanfaatkan untuk berbagai keperluan penelitian di bidang teknologi informasi. Twitter telah menyiapkan Structured text (XML dan daftar kategori) di dalamnya selalu terdapat informasi username, timestamp, text, retweet, favorite dan informasi lainnya, hal ini dapat dilihat dari proses implementasi metadata di setiap tweet. Sedangkan unstructured yaitu konten bagian yang tidak memiliki struktur khusus, berbentuk text dilihat dengan bentuk text atau disebut tweet.

### 2.2. Text Mining

Text mining didefinisikan sebagai proses mengekstraksi pengetahuan implisit dari data textual. Karena pengetahuan implisit yang merupakan output dari Text Mining tidak ada dalam penyimpanan yang diberikan, maka harus dibedakan dari informasi yang diambil dari penyimpanan. Text mining adalah salah satu jenis data mining khusus, layaknya seperti relational mining, web mining, dan big data mining [8] [8] [9].

### 2.3. Analisis Sentiment

Analisis sentimen atau opinion mining merupakan salah satu bagian dari bidang studi text mining yang disebut sebagai studi komputasi mengenai pendapat rakyat, sentimen, penilaian, sikap, dan emosi terhadap entitas dan atribut yang fokusnya adalah mengekstraksi, mengidentifikasi dan meneliti dan menemukan karakteristik sentimen atau kecenderungan seseorang terhadap sebuah opini dalam sebuah masalah atau objek oleh seseorang dalam unit teks menggunakan metode NLP (Natural Language Processing), statistik atau machine learning [13] [10] [11] [12].

**2.4. Naive Bayes Classifier**

digunakan untuk mencari nilai probabilitas tertinggi dalam teknik klasifikasi data uji pada kategori yang paling tepat [2]. Saat proses klasifikasi algoritma Naive Bayes akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (Vmap), Perhitungannya dapat dilihat sebagai berikut [18]

$$V_{map} = \underset{V_j \in \{\text{Positif}, \text{"Negatif"}\}}{\text{argmax}} P(V_j) \prod P(x_i | V_j) \quad (2.1)$$

Keterangan :

V<sub>j</sub> : Kategori tweet j=1,2,..dst. Misal j<sub>1</sub> = Kelas sentimen positif, j<sub>2</sub> kelas sentimen negatif.

P(x<sub>i</sub> | V<sub>j</sub>) : Probabilitas x<sub>i</sub> pada kategori V<sub>j</sub>

P(V<sub>j</sub>) : Probabilitas dari V<sub>j</sub>

Untuk P(V<sub>j</sub>) dan P(x<sub>i</sub> | V<sub>j</sub>) dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut:

$$P(V_j) = \frac{|docs_j|}{|contoh|} \quad (2.2)$$

$$P(x_i | v_j) = \frac{n_i + |kosaka|}{nk + 1} \quad (2.3)$$

Keterangan:

|doscs j| : jumlah dokumen setiap kategori j

|contoh| : jumlah dokumen dari semua kategori

nk : jumlah frekuensi kemunculan setiap kata

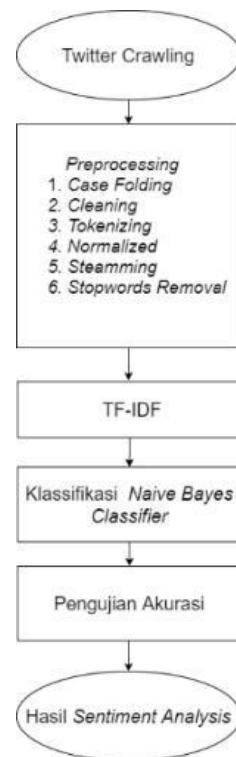
n : jumlah frekuensi kemunculan kata dari setiap kategori

|kosakata| : jumlah semua kata dari semua kategori

**3. Metode Penelitian**

**3.1 Tahapan Penelitian**

Tahapan penelitian ini merancang alur dari penelitian yang akan dilaksanakan, alur ini tersusun secara terstruktur dan disampaikan melalui gambar dengan tahapan-tahapan yang runtun tentang apa saja yang akan dilaksanakan. Tahapan penelitian dapat dilihat pada Gambar 1:



Gambar 1. Tahapan penelitian

**4. Hasil dan Pembahasan**

**4.1 Twitter Crawling**

Dataset dikumpulkan melalui pengambilan data dari twitter Kementerian Kesehatan Republik Indonesia menggunakan python dan memanfaatkan fasilitas twitterscrapper, berupa komentar tweet berbahasa Indonesia sebanyak 2397 data yang diambil mulai dari 1 Maret 2020 hingga 30 Oktober 2020 dengan kata kunci covid-19 dan kemenkes dalam bentuk file .csv.

**4.2 Preprocessing**

Preprocessing merupakan tahap awal pada proses pengolahan data teks hasil ekstraksi crawling sebelum diolah lebih lanjut ke tahap proses klasifikasi hingga nanti masuk ketahap uji dan evaluasi. Di tahap ini data yang semula berjumlah 2000 data. Dengan tujuan menghasilkan term/kata dari data yang sesuai dengan kebutuhan.

**4.3 Pengujian dan Evaluasi**

Pengujian ini dilakukan untuk bertujuan memastikan bahwa penelitian yang dilaksanakan dengan metode Naive Bayes Classifier menghasilkan nilai akurasi dengan

keakuratan yang baik. *Syntax* yang digunakan untuk melakukan perhitungan confusion matrix menggunakan library *skit.learn* yang merupakan library yang sama dengan proses klasifikasi menggunakan *Naive Bayes Classifier*. Hasil dari model training diklasifikasikan dengan testing menghasilkan matrix dengan ukuran 3x3 sebagai representatif kelas aktual dan kelas prediksi.

Tabel 1. *Confusion Matrix*

Data Aktual	Data Prediksi		
	Negatif	Netral	Positif
Negatif	TNg	NgN	FN
Netral	NNg	TN	NP
Positif	FP	PN	TP

Pada Tabel 1 dijelaskan bahwa *confusion matrix* merupakan hasil dari prediksi menggunakan mesin klasifikasi *Naive Bayes Classifier* yang akan diukur performa dari tiap-tiap kelas dengan menghitung *precision*, *recall*, dan F1-Score. Dengan keterangan sebagai berikut:

- TP = Kelas kata terprediksi benar bernilai positif
- PN = Kelas kata positif terprediksi netral
- FP = Kelas kata positif terprediksi negatif
- NP = Kelas kata netral terprediksi sebagai kelas kata positif
- TN = kelas kata terprediksi netral
- NNg = Kelas kata netral terprediksi negatif
- FN = Kelas kata negatif terprediksi positif
- NgN = Kelas kata negatif terprediksi netral
- TNg = Kelas kata negatif terprediksi negatif

Tabel 2. Hasil *Confusion Matrix*

Data Aktual	Data Prediksi		
	Negatif	Netral	Positif
Negatif	347	30	33
Netral	16	4	7
Positif	16	3	24

Berdasarkan tabel didapatkan hasil *confusion matrix* dari mesin melalui library *skit.learn*, dapat terlihat jumlah kelas kata yang diprediksi oleh mesin yaitu: Sebanyak 24 kelas kata terprediksi benar bernilai positif (TP), 3 kelas kata positif terprediksi netral (PN), 16 kelas kata positif terprediksi negatif (FP), 7 kelas kata netral terprediksi sebagai kelas kata positif (NP), 4 kelas kata terprediksi netral (TN), 16 kelas kata netral terprediksi negatif (NNg), 33 kelas kata negatif terprediksi positif (FN), 30 kelas kata negatif terprediksi netral (NgN), dan 347 kelas kata negatif terprediksi negatif (TNg).

**Kelas Negatif:**

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{24}{24 + 16} = 0.92$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{24}{24 + 33} = 0.85$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times 0.92 \times 0.85}{0.92 + 0.85} = 0.88$$

**Kelas Netral:**

$$\text{Precision} = \frac{\text{NP}}{\text{NP} + \text{FP}} = \frac{7}{7 + 16} = 0.11$$

$$\text{Recall} = \frac{\text{NP}}{\text{NP} + \text{FN}} = \frac{7}{7 + 33} = 0.15$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times 0.11 \times 0.15}{0.11 + 0.15} = 0.12$$

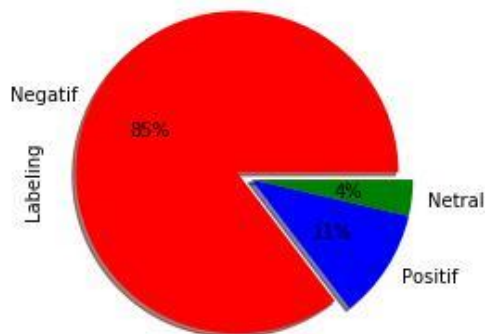
**Kelas Positif:**

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{24}{24 + 16} = 0.38$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{24}{24 + 33} = 0.56$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times 0.38 \times 0.56}{0.38 + 0.56} = 0.45$$

Grafik hasil persentase sentimen dibuat dengan menggunakan library *matplotlib*. Persentase ini dibuat untuk memudahkan pembaca serta mengukur berapa besar sentiment yang diberikan oleh masyarakat terhadap respon kabar harian covid-19 pada twitter Kementerian Kesehatan RI berdasarkan proses labeling.



Gambar 2. Hasil Grafik Sentiment Analisis

Pada gambar ini menunjukkan bahwa hasil dari sentimen analisis yang telah dilakukan menghasilkan persentase 85% negatif, 4% netral dan 11% positif.

**5. Kesimpulan Dan Saran**

**5.1 Simpulan**

Dari penelitian yang telah dilaksanakan dengan penggunaan metode *machine learning* yaitu algoritma *Naive Bayes Clasifier* pada dataset kabar harian covid-19 pada twitter Kementerian Kesehatan republik Indonesia dengan kata kunci Kemenkes dan Covid-19 telah didapatkan dataset hasil crawling sebanyak 2397 dataset,

yang kemudian dilakukan proses Preprocessing kemudian dataset tersebut diolah untuk proses selanjutnya untuk mendapatkan kelas sentimen.

Selanjutnya setelah melakukan klasifikasi sentimen, didapatkan hasil klasifikasi sentimen dengan tiga kelas, yaitu kelas positif sebanyak 11%, kelas negatif sebanyak 85%, dan kelas netral sebanyak 4%. Data hasil klasifikasi ini diperoleh dengan membagi data menjadi data training dan testing, dengan ketentuan jumlah data training sebanyak 80% dan data testing sebanyak 20%.

Dari hasil klasifikasi Naïve Bayes Classifier dan pengujian akurasi, precision, recall, dan F1-Score, dengan demikian dapat disimpulkan bahwa penelitian ini menghasilkan sentimen masyarakat pengguna twitter mengenai respon masyarakat mengenai kabar harian covid-19 yang diberikan oleh twitter Kementerian Kesehatan Republik Indonesia dengan presentase kelas sentimen negatif sebesar 77%. Dengan keakuratan hasil pengujian akurasi sebesar 78%. dan pengujian precision sebesar 92%, Recall sebesar 85%, F1-Score sebesar 88%. Menggunakan pengujian metode Naive Bayes Classifier.

## 5.2 Saran

Dalam penelitian ini masih terdapat beberapa kekurangan yang perlu diperbaiki, adapun beberapa saran yang dapat penulis sampaikan agar penelitian selanjutnya memperoleh hasil yang lebih maksimal, penulis menyarankan untuk peneliti selanjutnya sebagai disarankan beberapa hal berikut, yaitu:

1. Peneliti selanjutnya disarankan untuk tidak mengandalkan kamus sastraawi saja tapi juga membuat kamus sendiri yang lebih lengkap, agar hasil lebih akurat.
2. Disarankan untuk memasukan metode klasifikasi lain dalam penelitian selanjutnya.

## DAFTAR PUSTAKA

- [1] R. Fadilah and H. Kuswoyo, "Transitivity Analysis of News Reports on Covid-19 of Jakarta Post Press," 2021.
- [2] Y. A. Rahman, "Vaksinasi Massal Covid-19 sebagai Sebuah Upaya Masyarakat dalam Melaksanakan Kepatuhan Hukum (Obedience Law)," *Khazanah Huk.*, vol. 3, no. 2, pp. 80–86, 2021, doi: 10.15575/kh.v3i2.11520.
- [3] I. Mahfud and A. Gumantan, "Survey Of Student Anxiety Levels During The Covid-19 Pandemic," *Jp. jok (Jurnal Pendidik. Jasmani, Olahraga dan Kesehatan)*, vol. 4, no. 1, pp. 86–97, 2020.
- [4] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [5] N. Neneng, A. S. Puspaningrum, and A. A. Aldino, "Perbandingan Hasil Klasifikasi Jenis Daging Menggunakan Ekstraksi Ciri Tekstur Gray Level Co-occurrence Matrices (GLCM) Dan Local Binary Pattern (LBP)," *SMATIKA J.*, vol. 11, no. 01, pp. 48–52, 2021.
- [6] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [7] B. M. Pintoko and K. M. L., "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [8] M. Bakri, "Penerapan Data Mining untuk Clustering Kualitas Batu Bara dalam Proses Pembakaran di PLTU Sebalang Menggunakan Metode K-Means," vol. 11, pp. 1–4, 2017.
- [9] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 2, p. 126, 2016, doi: 10.35314/isi.v1i2.131.
- [10] R. R. SURYONO and B. Indra, "P2P Lending sentiment analysis in Indonesian online news," in *Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*, 2020, pp. 39–44.
- [11] M. A. Assuja and S. Saniati, "Analisis Sentimen Tweet Menggunakan Backpropagation Neural Network," *J. Teknoinfo*, vol. 10, no. 2, pp. 48–53, 2016.
- [12] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis
- [13] Alita, D. (2017) *Usulan Penelitian S2 Analisis Sentimen Pada Teks Berbahasa Indonesia Yang Memuat Emoticon Analisis Sentimen Pada Teks Berbahasa Indonesia Yang Memuat Emoticon.*
- [14] Alita, D., Priyanta, S. And Rokhman, N. (2019) "Analysis Of Emoticon And Sarcasm Effect On Sentiment Analysis Of Indonesian Language On Twitter," *Journal Of Information Systems Engineering And Business Intelligence*, 5(2), P. Doi: 10.20473/Jisebi.5.2.100-109.
- [15] Chatrina, Siregar, N., Ruli, A, Siregar, R. And Yoga, Distra, Sudirman, M. (2020) "Implementasi Metode Naive Bayes Classifier (Nbc) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (Pjj)," *Jurnal Teknologi*, 3(1), Pp. 102–110. Available At: <https://Aperti.EJournal.Id/Teknologi/Article/View/67>.

- [16] Feldman, R. And Sanger, J. (2007) *The Text Mining Handbook*.
- [17] Fitri, V. A., Andreswari, R. And Hasibuan, M. A. (2019) "Sentiment Analysis Of Social Media Twitter With Case Of Anti-Lgbt Campaign In Indonesia Using Naïve Bayes, Decision Tree, And Random Forest Algorithm," *Procedia Computer Science*. Elsevier B.V., 161, Pp. 765–772 Doi: 10.1016/J.Procs.2019.11.181.
- [18] Haranto, F. Fat And Sari, B. W. (2019) "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap," 15(2),Pp.171–176.Do:10. 33480/Pilar.V15i2.699.
- [19] Isnain, A. R. Et Al. (2020) "Bidirectional Long Short Term Memory Method And Word2vec Extraction Approach For Hate Speech Detection," *Ijccs (Indonesian Journal Of Computing And Cybernetics Systems)*, 14(2), P. 169 Doi: 10.22146/Ijccs.51743.