

# ANALISIS SENTIMENT DATA TWITTER MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES TERHADAP PEMBERITAAN PERKEMBANGAN PANDEMIK CORONA

*TWITTER DATA SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE AND NAIVE  
BAYES ALGORITHM ON DEVELOPMENT NEWS CORONA PANDEMIC*

**Dahri Yani Hakim Tanjung<sup>1</sup>, Silvia Lestari<sup>2</sup>**

*<sup>1,2</sup>Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama*

*<sup>1,2</sup>Universitas Potensi Utama, K.L. Yos Sudarso KM 6,5 No. 3A Tj. Mulia - Medan*

*Email : [notashapire@gmail.com](mailto:notashapire@gmail.com)<sup>1</sup> [silvialestai.via96@gmail.com](mailto:silvialestai.via96@gmail.com)<sup>2</sup>*

## **Abstrak**

*Twitter merupakan salah satu media sosial yang sangat banyak digunakan untuk memberikan kritik akan sebuah pemberitaan atau masalah pribadi mereka. Pada saat ini sedang banyak pembicaraan mengenai perkembangan virus corona di Indonesia maupun di dunia, banyak media dunia yang memberikan pemberitaan melalui Twitter dan ada pula para pengguna yang memberikan tanggapan terhadap pemberitaan tersebut. Maka dari itu pada penelitian ini dilakukannya analisis terhadap tweet atau tanggapan para pengguna Twitter terhadap perkembangan atau isu pandemik Korea. Penelitian ini dilakukan dengan cara memisahkan sentiment atau tanggapan yang negatif dan positif menggunakan algoritma Naive Bayes, dan Support Vector Machine (SVM). Hasil pengujian yang dilakukan dapat menunjukkan bahwa tingkat akurasi dengan term frequency memberikan hasil akurasi yang lebih baik daripada akurasi dengan fitur TF-IDF. Metode Support Vector Machine menghasilkan akurasi yang lebih baik daripada metode Naive Bayes baik dalam klasifikasi sentimen maupun dalam klasifikasi kategori. Namun demikian, secara keseluruhan penggunaan metode Support Vector Machine dan Naive Bayes sama-sama memiliki performansi yang cukup baik untuk melakukan klasifikasi tweet.*

**Kata kunci :** Analisis Sentiment, Naive Bayes, Support Vector Machine, Twitter, Pandemi Corona

## **Abstract**

*Twitter is one of the social media that is very widely used to provide criticism of a news or their personal problems. At this time there is a lot of talk about the development of the corona virus in Indonesia and in the world, many world media that provide news via Twitter and there are also the use of which gave a response to the news. Therefore in this study an analysis of tweets or Twitter users' responses to the development or issue of the Corona pandemic. This research was conducted by separating negative and positive sentiments or responses using the Naive Bayes algorithm, and Support Vector Machine (SVM). The results of tests carried out can show that the accuracy with the term frequency gives better results than the accuracy with the TF-IDF feature. The Support Vector Machine method produces better accuracy than the Naive Bayes method both in sentiment classification and in category classification. However, overall the use of the Support Vector Machine and Naive Bayes methods both have fairly good performance to classify tweets.*

**Keywords:** Sentiment Analysis, Naive Bayes, Support Vector Machine, Twitter, Corona Pandemic

## 1. PENDAHULUAN

Twitter merupakan salah satu media sosial yang sangat banyak digunakan untuk memberikan kritik akan sebuah pemberitaan atau masalah pribadi mereka yang memungkinkan para penggunaanya untuk mengirim pesan hanya dengan penggunaan 140 karakter akan tetapi setelah twitter di perbaharui para pengguna dapat mengetik teks sampai pada dengan 280 karakter. Awal mula twitter berdiri pada Maret 2006 oleh Jack Dorsey, dan diluncurkan pada bulan Juli. sejak kemunculannya, twitter menjadi salah satu situs yang paling sering dikunjungi oleh pengguna sehingga twitter dijuluki dengan “pesan singkat dari internet”. Meningkatnya penggunaan Twitter menyebabkan layanan ini telah dimanfaatkan untuk berbagai aktifitas dalam berbagai aspek, seperti sebagai sarana protes, kampanye politik, sarana pembelajaran, dan sebagai media komunikasi darurat. Selain itu, twitter juga memiliki bnyak permasalahan dan kontroversi seperti masalah keamnan dan privasi pengguna, serta penyensoran.

Munculnya penelitian ini mencoba memanfaatkan Twitter sebagai media yang digunakan untuk menganalisis cuitan berbahasa Indonesia yang membahas mengenai perkembangan pandemik covid-19 yang berkembang baik di negara indonesia maupun di seluruh dunia. Banyak media dunia yang memberikan pemberitaan melalui twitter dan ada pula para penggunaan yang memeberikan tanggapan terhadap pemberitaan tersebut. Pembahasan yang di analisis merupakan beberapa tanggapan terhadap media – media online, maupun tanggapan mereka secara langsung terhadap perkembangan pandemik korona yang berada di daerah tempat tinggal mereka. Seperti yang kita ketahui pada saat ini karena pesatnya perkembangan virus corona yang menyerang warga indonesia membuat pemerintah membuat beberapa peraturan guna untuk memutuskan mata rantai virus corona dimulai dilakukannya *Sosial Distancing* yaitu dengan melakukan semua kegiatan dari rumah mulai dari proses mengajar dan belajar, mengerjakan pekerjaan dari rumah serta akhir-akhir ini pemerintah memberlakukan adanya PSBB yaitu pembatasan sosial berskala besar yang dimana pemerintah menganjurkan semua warga nya untuk tetap dirumah yang membuat siklus ekonomi warga indonesia menjadi terhambat serta banyaknya warga yang kelaparan akibat diberlakukannya PSBB yang membuat mereka tidak bisa mencari nafkah, oleh karena itu berbagai tanggapan atau *tweet* dari warga baik positif maupun negatif.

Problem solving yang digunakan pada penelitian ini adalah menggunakan algoritma naïve bayes classifier untuk mengklasifikasikan tweet yang kemudian dikombinasikan dengan beberapa fitur seperti term frequency serta TF-IDF, kemudian menggunakan Metode Support Vector Machine untuk menghasilkan akurasi yang lebih baik daripada metode Naive Bayes baik dalam klasifikasi sentimen maupun dalam klasifikasi kategori.

## 2. METODE PENELITIAN

### 2.1 Support Vector Machine

Support Vector Machine (SVM) pertama kali disusulkan oleh cortes dan vapnik yang merupakan sebuah mesin *learning* baru untuk masalah klasifikasi dua kelompok. Mesin ini secara konseptual mengimplementasikan ide dari sebuah input vektor yang kemudian dipetakan secara non-linear ke ruang fitur berdimensi sangat tinggi [1]. SVM merupakan salah satu algoritma supervised machine learning dengan performa yang luar biasa digunakan dalam banyak pengenalan pola dan masalah estimasi regresi, peramalan dan membangun mesin cerdas [2][3].

SVM telah menunjukkan kinerja yang dapat diterima jika dibandingkan dengan metode lain, terutama bila digunakan untuk klasifikasi[4]. Pada prinsipnya SVM dapat secara efisien melakukan klasifikasi linier dan non-linier, dan metode didasarkan pada minimalisasi risiko struktural yang paling akurat untuk mengklasifikasikan teks karena berfokus pada kelas terpisah[5]. SVM berusaha mencari hyperplane terbaik, dimana Hyperplane merupakan

garis tengah yang memisahkan antara 2 kelas antara titik-titik yang mewakili kategori dari hal positif dan dari titik-titik yang mewakili hal negative dalam sebuah klasifikasi[6][7]

Fungsi deskriptif dari SVM dapat dilihat pada persamaan berikut[7]:

$$h(X) = z x \phi(X) + c \quad (1)$$

dimana:

X : fitur vector,

z : vector dari bobot yang berbeda,

$\phi$ : fungsi pemetaan non-linier,

c : vector bias

z dan c diperoleh secara otomatis dari dataset training

## 2.2 Naive Bayes Classifier

Naive Bayes Classifier merupakan pengklasifikasi probabilistik sederhana berdasarkan teorema Bayesian. Ini mengasumsikan fungsi atau fitur independen dari kumpulan data. Fitur di kelas tidak tergantung pada ada dan tidak adanya fitur lainnya. Keuntungan utama dari classifier ini adalah membutuhkan sejumlah kecil data pelatihan untuk menghitung rata-rata dan varians variabel untuk klasifikasi. Varians label ditentukan dari seluruh matriks kovarians karena fitur dari kelas tidak tergantung satu sama lain[8]

Misalkan A adalah tupel data dan S adalah hipotesis sehingga tupel data A termasuk dalam kelas C. Untuk mengklasifikasikan P (S|A) dapat dilihat pada persamaan berikut.

$$P(S|A) = \frac{P(A|S) P(S)}{P(A)} \quad (2)$$

Dimana:

P(S|A) : Probabilitas S dengan bukti A

P(S) : Probabilitas S

P(A|S) : Probabilitas A berpengaruh terhadap S

P(A) : Probabilitas A

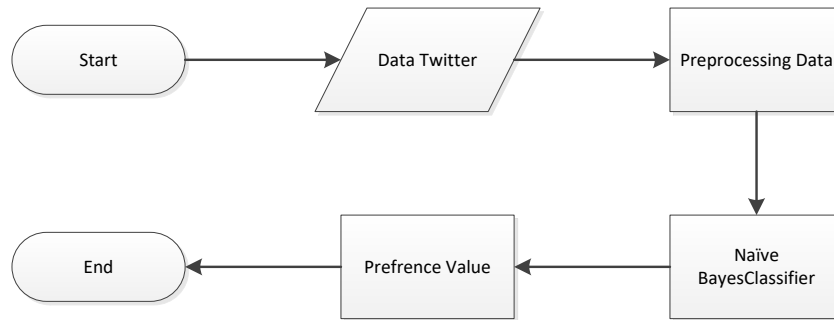
## 2.3 Analisis Sentiment

Analisis sentimen adalah proses otomatis untuk memahami pendapat tentang subjek tertentu dari bahasa lisan atau tulisan. Analisis sentimen menjadi topik tren dari berbagai penelitian dan *text mining* telah dilakukan dalam beberapa tahun terakhir[9]. Analisis Sentimen dan *Opinion mining* adalah subbidang dari *text mining* yang bertujuan untuk mengidentifikasi subjek. Informasi subjektif ini diekstraksi dari teks dengan menggunakan kombinasi mesin pembelajaran dan pemrosesan bahasa alami teknik[10]. Analisis sentimen kini telah menjadi pendekatan dominan yang digunakan untuk mengekstrak sentimen dan penilaian tentang topik tertentu dari sumber online, yang dibagi menjadi tiga kategori yaitu positif, negatif, dan netral[11].

## 2.4 Sistem Yang Dibangun

Pada penelitian ini bertujuan untuk mengklasifikasikan tanggapan – tanggapan dan sentimen – sentimen yang terdapat pada setiap postingan di twitter mengenai perkembangan virus corona yang sedang marak pada saat ini di negara indonesia. Adapun cuitan – cuitan dapat berupa cuitan – cuitan yang positif dan negatif.pada analisis sentimen ini menggunakan algoritma Naive Bayes dan Support Vector machine yang dimana dari kedua algoritma ini dibutuhkannya tingkat akurasi yang tepat serta algoritma mana yang lebih baik digunakan dalam mengklasifikasikan sentimen – sentimen baik positif maupun negatif.sebelum dilakukannya proses pengklasifikasian diperlukan preprocessing data untuk menyederhanakan data.kemudian kalsifikasi data menjadi data training dan data testing setelah itu data akan diuji nilai akurasinya menggunakan cross validation,dilanjutkan dengan mencari preference value

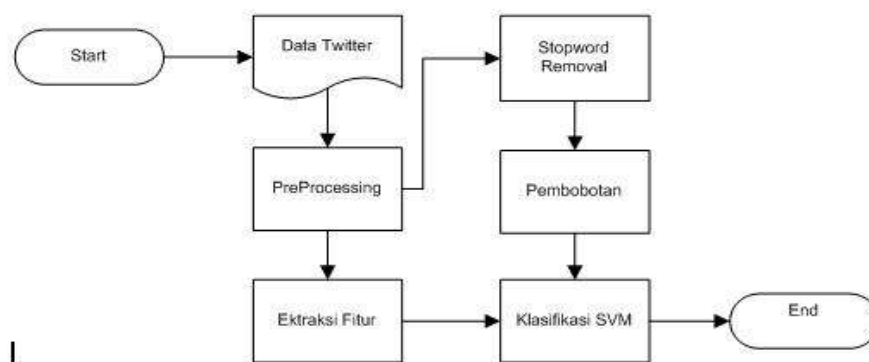
yaitu dengan menghitung nilai preference value-nya. sedangkan pada Support Vector Machine ( SVM ) adapun gambaran dari sistem yang di bangun dapat dilihat pada gambar 1 dan 2 yaitu :



Gambar 1. Flowchart Algoritma Naive Bayes Classifier

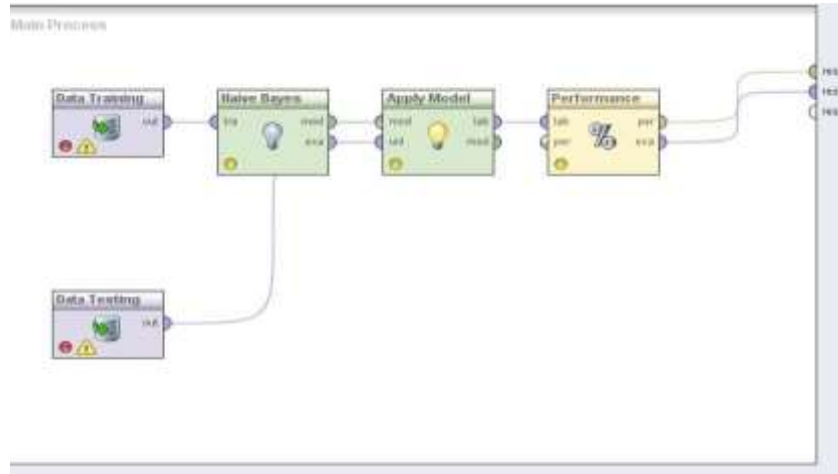
Adapun penjelasan pada gambar 1 yaitu :

1. Data Twitter  
Data twitter diambil langsung melalui media twitter yang dimana berupa postingan – postingan serta tweet para pengguna twitter
2. Preprocessing Data  
Preprocessing data merupakan salah satu proses yang bertujuan untuk membuat bentuk data agar sesuai dengan kebutuhan pada saat pengklasifikasian dan memudahkan pada saat pemrosesan data dalam sistem.
3. Klasifikasi Naive Bayes  
Pada proses ini data di klasifikasikan menjadi beberapa sentimen yang biasanya terdiri data sentimen positif dan sentimen negatif.
4. Preference Value  
Preference Value merupakan suatu proses yang dimana pada proses ini setiap kandidat akan dihitung nilai preference valuenya melalui tweet yang dikumpulkan berbeda dengan data training. yang dimana data dikumpulkan akan dimasukkan pada proses klasifikasi .



Gambar 2. Flowchart Algoritma SVM





Gambar 4. Proses Klasifikasi Sentiment Twitter

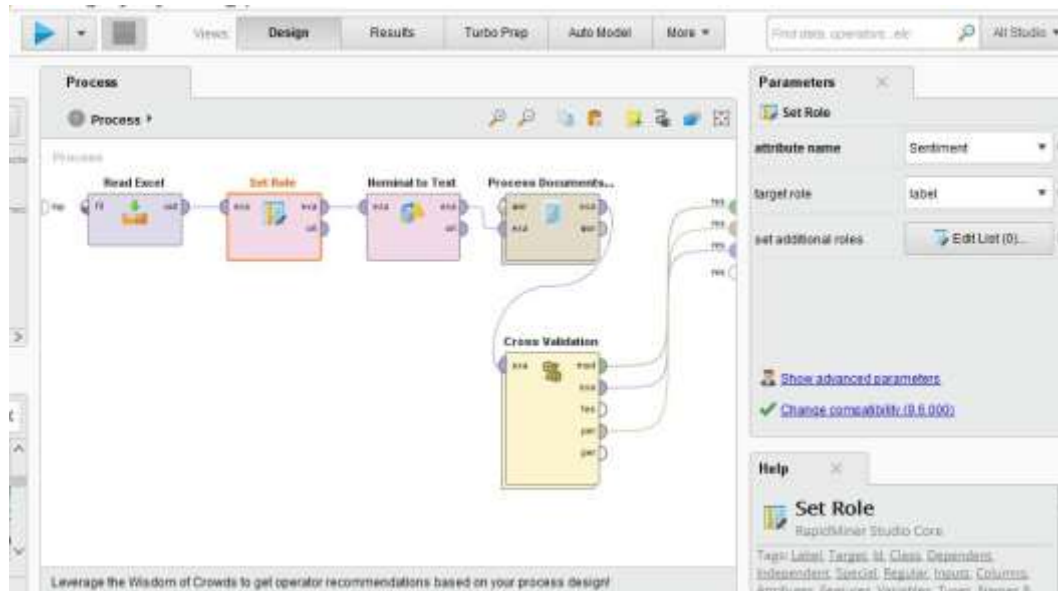
Pada gambar 4 menunjukkan bagaimana cara mengklasifikasi sentiment menggunakan RapidMiner dengan menggunakan algoritma naive bayes yaitu dengan menginput data training dan data testing kemudian koneksikan dengan metode naive bayes, setelah itu gunakan operator apply model yang berfungsi untuk mendapatkan prediksi pada data yang tidak terlihat atau mengubah data dengan menerapkan model preprocessing. Kemudian koneksikan dengan operator performance yang bertujuan untuk evaluasi kinerja. Ini memberikan daftar nilai kriteria kinerja. Kriteria kinerja ini secara otomatis ditentukan agar sesuai dengan jenis tugas pembelajaran. Adapun yang menunjukkan data testing yang telah di proses terdapat pada gambar 5 yaitu :

Row No.	Sentiment	confidence	conf_kasat	conf_bayes	Text
1	Negatif	0.975	0.014	0.011	Negatif esoknya
2	Negatif	0.975	0.014	0.011	Negatif anuhnya2E
3	Positif	0.014	0.975	0.011	Positif kompasan
4	Negatif	0.975	0.014	0.011	Negatif Diambil_
5	Negatif	0.975	0.014	0.011	Negatif terpedobu
6	Positif	0.014	0.975	0.011	Positif kompasan
7	Positif	0.014	0.975	0.011	Positif ini, it, and
8	Positif	0.014	0.975	0.011	Positif pntika, Tuk
9	Positif	0.014	0.975	0.011	Positif Sudmaya97
10	Positif	0.014	0.975	0.011	Positif wisanggan,
11	Positif	0.014	0.975	0.011	Positif _S_0_S_ M
12	Positif	0.014	0.975	0.011	Positif Kik, Sastra
13	Positif	0.014	0.975	0.011	Positif DuluW, Jta
14	Positif	0.014	0.975	0.011	Positif HutanPati
15	Positif	0.014	0.975	0.011	Positif SistemAccu
16	Negatif	0.975	0.014	0.011	Negatif Sblm, PE
17	Positif	0.014	0.975	0.011	Positif Dulu2337-
18	Positif	0.014	0.975	0.011	Positif saye, wE, b
19	negatif	0.018	0.975	0.954	negatif Harlan822

Gambar 5. Example set Data Testing

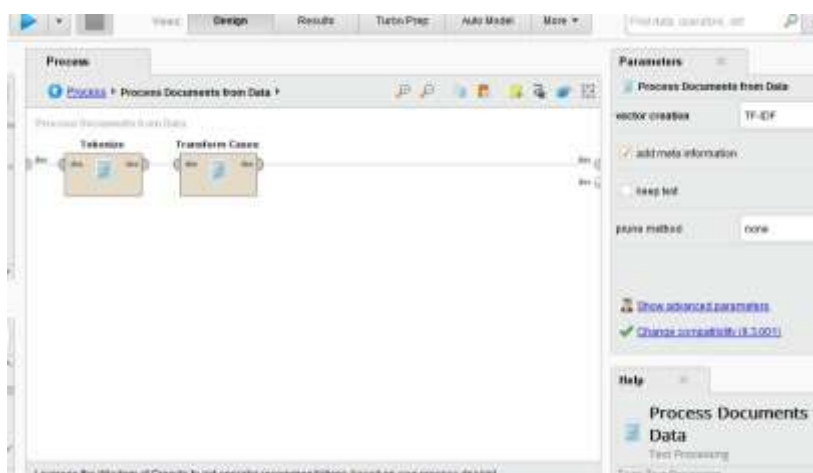
Pada Gambar 6 menunjukkan nilai akurasi pada setiap sentiment positif dan sentiment negatif yang menghasilkan nilai akurasi 100%.





Gambar 8. Proses Support Vector Machine

Pada gambar 8 dijelaskan pada tahap awal dilakukan dengan cara menginput data testing kemudian dikoneksikan dengan data set role yang berfungsi untuk mengubah peran satu Atribut atau lebih. kemudian dikoneksikan dengan operator nominal to text yang bertujuan untuk berfungsi untuk mengubah jenis atribut nominal yang dipilih menjadi teks. Operator ini juga memetakan semua nilai dari atribut ini ke nilai string yang sesuai. Data teks diproses untuk mengekstrak kata-kata dan mengirimkan kata -vektor (numerik untuk mewakili teks). Kemudian memilih operator Process Documents from Data berfungsi menghasilkan kata vektor dari atribut string. kemudian pilihlah operator Cross Validation berfungsi melakukan validasi silang untuk memperkirakan kinerja statistik operator pembelajaran (biasanya pada kumpulan data yang tidak terlihat). Adapun beberapa tahapan SVM yang dilakukan selanjutnya yaitu terdapat pada gambar 9, 10, 11 yaitu :



Gambar 9. Proses Support Vector Machine





Gambar 10. Proses SVM dengan Cross Validation

	true Negatif	true Positif	class precision
pred. Negatif	0	0	0.00%
pred. Positif	12	37	75.51%
class recall	0.00%	100.00%	

Gambar 11. Proses Support Vector Machine

#### 4. KESIMPULAN

Berdasarkan pembahasan dari bab-bab sebelumnya yang telah dilakukan maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Metode Naive Bayes menghasilkan akurasi sebesar 100% antara sentimen positif dan negatif sedangkan pada Metode Support Vector Machine menghasilkan akurasi 75,50% +/- 8,32% ( Micro Average 75,51 %)
2. Perbandingan antara kedua metode Naive Bayes dan Support Vector Machine didapatkan hasil SVM lebih baik dibandingkan dengan Naive bayes karena menampilkan nilai vektor yang lebih jelas.

#### 5. SARAN

Adapun saran untuk penelitian yang akan datang adalah :

1. Peneliti Selanjutnya seharusnya menggunakan data yang lebih banyak, agar data text lebih jelas pengklasifikasiannya
2. Peneliti selanjutnya seharusnya menggunakan attribute yang lebih banyak agar dapat menghasilkan data sentiment yang lebih akurat lagi.

## DAFTAR PUSTAKA

- [1] C. Cortes and Vladimir Vapnik, "Photonit neural networks and learning machines the role of electron-trapping materials," *Support. Networks*, vol. 7, no. 5, pp. 63–72, 1992, doi: 10.1109/64.163674.
- [2] S. Styawati and K. Mustofa, "A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 3, p. 219, 2019, doi: 10.22146/ijccs.41302.
- [3] J. Nayak, B. Naik, and H. S. Behera, "A Comprehensive Survey on Support Vector Machine in Data Mining Tasks: Applications & Challenges," *Int. J. Database Theory Appl.*, vol. 8, no. 1, pp. 169–186, 2015, doi: 10.14257/ijtda.2015.8.1.18.
- [4] A. Abubakar et al., "A support vector machine classification of computational capabilities of 3D map on mobile device for navigation aid," *Int. J. Interact. Mob. Technol.*, vol. 10, no. 3, pp. 4–10, 2016, doi: 10.3991/ijim.v10i3.5056.
- [5] N. N. A. Sjarif, Y. Yahya, S. Chuprat, and N. H. F. M. Azmi, "Support vector machine algorithm for SMS spam classification in the telecommunication industry," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, no. 2, pp. 635–639, 2020, doi: 10.18517/ijaseit.10.2.10175.
- [6] I. Ernawati, "Naïve Bayes Classifier Dan Support Vector Machine Sebagai Alternatif Solusi Untuk Text Mining," *J. Int. Ti2*, vol. 5, no. 1, pp. i–viii, 2016, [Online]. Available: <http://tip.pj.unp.ac.id/index.php/tip/article/view/219/122>.
- [7] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.
- [8] T. k and M. Wadhawa, "Analysis and Comparison Study of Data Mining Algorithms Using Rapid Miner," *Int. J. Comput. Sci. Eng. Appl.*, vol. 6, no. 1, pp. 9–21, 2016, doi: 10.5121/ijcsea.2016.6102.
- [9] I. M. Noor and M. Turan, "Sentiment Analysis using Twitter Dataset," *IJID (International J. Informatics Dev.)*, vol. 8, no. 2, p. 81, 2020, doi: 10.14421/ijid.2019.08206.
- [10] L. Batista and S. Ratté, "Multi-Classifer System for Sentiment Analysis and Opinion Mining," in *Encyclopedia of Social Network Analysis and Mining*, New York, NY: Springer New York, 2017, pp. 1–9.
- [11] A. Pertiwi, A. Triayudi, and E. T. E. Handayani, "Sentiment Analysis of the Impact of Covid-19 on Indonesia's Economy through Social Media Using the ANN Method," *J. Mantik*, vol. 4, no. May, pp. 605–612, 2020.