



# Analisis Sentimen Publik Terhadap Implementasi Kurikulum Merdeka Melalui Data Komentar Youtube Menggunakan Algoritma Klasifikasi *K-Nearest Neighbors* dan *Naïve Bayes*

Rinaldi Satria Pratama<sup>1\*</sup>, Sutedi<sup>2</sup>

rinaldi.2121210002@mail.darmajaya.ac.id<sup>1</sup>, sutedi@darmajaya.ac.id<sup>2</sup>

<sup>1,2</sup> Institut Informatika dan Bisnis Darmajaya

## Abstrak

*This research aims to conduct a public sentiment analysis on the implementation of the Merdeka Curriculum through YouTube comment data using the K-Nearest Neighbors (K-NN) and Naïve Bayes classification algorithms. The Merdeka Curriculum is an educational program launched by the Indonesian government in 2020 to improve the quality of education in Indonesia. The research method used is quantitative, with stages including data collection, data preprocessing, and sentiment classification modeling using the K-NN and Naïve Bayes algorithms. The research results show that the K-NN algorithm performs better than the Naïve Bayes algorithm in classifying public sentiment towards the implementation of the Merdeka Curriculum. The K-NN algorithm with the best k value of 11 achieved an accuracy of 79.70%, precision of 77.20%, recall of 79.70%, and an f1-score of 77.65%, with an AUC value of 0.91, indicating excellent classification performance. In contrast, the Naïve Bayes algorithm achieved an accuracy of 53.82%, precision of 72%, recall of 53.82%, an f1-score of 59%, and an AUC value of 0.75, which falls into the fair category.*

## Status Artikel:

Diterima: 17-7-2024

Direvisi: 19-7-2024

Diterima: 29-7-2024

## Kata Kunci:

*Sentiment Analysis, Merdeka Curriculum, K-Nearest Neighbors, Naïve Bayes, YouTube*



© 2024 Ulfa arma putri, Yulina

This work is licensed under a

[Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

## 1. PENDAHULUAN

Kurikulum Merdeka adalah program pendidikan yang diluncurkan oleh pemerintah Indonesia pada tahun 2020 untuk meningkatkan kualitas pendidikan. Program ini berfokus pada hal-hal esensial, dengan lebih banyak waktu untuk pengembangan kompetensi dan karakter, serta bertujuan mengatasi krisis pembelajaran di Indonesia (Rahayu et al., 2022). Meskipun akan diterapkan secara nasional pada tahun 2024 namun, Kurikulum Merdeka masih menjadi pemilihan bagi Lembaga pendidikan (Kemendikbudristek, 2022).

Dalam konteks media sosial seperti YouTube, yang merupakan platform populer dengan 170 juta pengguna di Indonesia, pendapat masyarakat mengenai Kurikulum Merdeka dapat dianalisis melalui komentar di video (Harpizon et al., 2022). Komentar pada Youtube dapat menjadi informasi yang bermakna dan dapat dilakukan klasifikasi mengikuti alur arus data

*mining*. *Data mining* ialah cara yang dapat digunakan dalam menemukan suatu pengetahuan tersembunyi dalam sebuah database (Khusnul Khotimah dan Husniyati, 2020). Teknik *data mining* bertujuan agar memberikan kemudahan-kemudahan saat pola dalam teks dapat menerapkan konsep *text mining*. *Text mining* dipakai saat melakukan klasifikasi polaritas, yang dalam hal ini adalah klasifikasi dokumen (Anggara et al., 2022). Tujuannya untuk menemukan sumber data yang dapat digunakan pada saat tertentu. *Text mining* diproses untuk berbagai tujuan, seperti peringkasan, pencarian dokumen teks, dan analisis sentimen (Anggara et al., 2022).

Untuk mengkategorikan percakapan pengguna YouTube menjadi informasi bermakna terkait Kurikulum Merdeka, analisis sentimen dapat digunakan. Analisis sentimen adalah proses data teks untuk memahami dan mengklasifikasikan opini publik (Rakhman, 2017). Teknik data mining, termasuk asosiasi, klasifikasi, pengelompokan, dan prediksi, digunakan untuk mengidentifikasi pola dan tren dalam data (Fuad Nur Hasan, 2018). Beberapa penelitian sebelumnya juga menggunakan metode klasifikasi seperti Naive Bayes (NB), Support Vector Machine (SVM), dan K-Nearest Neighbor (KNN) untuk analisis sentimen di media sosial. Hasilnya menunjukkan bahwa algoritma SVM berbasis memberikan akurasi tinggi (Sujadi, 2022). Hasil penelitian Wandani (2021) pada tahun 2021, menggunakan algoritma K-NN, Random Forest, dan Naive Bayes untuk melakukan survei analisis sentimen pengguna Twitter pada acara flash sale. Dari hasil yang diperoleh kinerja algoritma Naive Bayes lebih tinggi sehingga hal ini menjadikan proses baru untuk peneliti dapat mengetahui data yang berbeda pada kinerja optimal K-NN dan Naive Bayes.

Berlandaskan latar belakang tersebut, digunakanlah sebuah metode KNN dan Naive Bayes dalam analisis sentimen komentar YouTube terkait implementasi Kurikulum Merdeka. Penelitian ini akan memaparkan proses dari preprocessing hingga analisis sentimen, serta menguji kualitas hasil analisis dari masing-masing algoritma. Judul penelitian yang diusulkan adalah "Analisis Sentimen Pada Media Sosial YouTube Terhadap Implementasi Kurikulum Merdeka Pada Satuan Pendidikan di Indonesia Menggunakan Metode K-Nearest Neighbors dan Naive Bayes".

## II. METODE

Metode dalam penelitian ini adalah pendekatan kuantitatif yang akan dilalui beberapa tahapan. Tahap pertama mengambil komentar melalui situs Youtube.com. Tahap kedua mengekstrak dan menyimpan komentar di direktori google colabs dan tahap ketiga mengekstrak data sentiment positif, sentiment negatif dan sentiment netral dari data komentar menggunakan algoritma K-Nearest Neighbors dan Naive Bayes Classifier.

### A. Tahap pengumpulan data

Data yang digunakan ialah komentar pengguna media sosial Youtube yang di dapatkan dari beberapa video youtube menggunakan kata kunci kurikulum merdeka. Data dikumpulkan (*crawling*) secara langsung terhubung dari media sosial Youtube dengan menggunakan google colabs, yang kemudian disimpan dalam bentuk file csv.

### B. Tahap Pengolahan Data

Tujuan dari tahap ini ialah agar mengekstrak dan membersihkan data yang diperoleh selama eksplorasi sebelumnya. Langkah-langkah dalam tahap ini: 1) penghapusan data berulang

(*remove duplicate*), 2) pemilihan *atribute* 3) penghapusan tanda baca (*replace*) 4) *filterisasi stopword*, 5) *tokenisasi*, dan 6) mencari frekuensi kata-kata yang sering muncul. (Nur Adinda Salsabila, 2022)

### C. Tahap Analisis Sentimen

Pada tahap ini, data diproses sehingga pelabelan dan penambahan atribut dapat dilakukan untuk penentuan nilai sentiment. Dari pengelompokan data tersebut (negatif, positif, atau netral). Adapun tahapan-tahapan yang dilakukan pada saat melakukan analisis sentimen adalah sebagai berikut (Wandani, 2021).

1. Tahap 1: mencari sentimen negatif, positif, dan netral di setiap baris.
2. Tahap 2: menganalisis sentimen seluruh dokumen sebagai negatif, positif, serta netral.
3. Tahap 3: menerapkan pengelompokan untuk mengumpulkan semua atribut yang ada dengan skor sentimen yang sama.
4. Tahap 4: menggunakan visualisasi data dari analisis sentimen untuk interaksi antar pengguna.

### D. Tahap Implementasi Algoritma

Setelah melakukan proses klasifikasi maka hasilnya akan di implementasikan menggunakan sebuah algoritma. Pada algoritma KNN dan *Naïve Bayes* sebagai pembanding algoritma mana yang memiliki akurasi lebih tinggi. KKN adalah salah satu algoritma klasifikasi yang dapat memeriksa jarak antara data terdekat dengan data baru. Algoritma ini menemukan jarak terdekat antara set pelatihan dan k tetangga terdekat dari set pengujian. Algoritma KNN mengklasifikasikan dokumen uji X sebagai berikut (Wandani, 2021).

- a. Tentukan parameter K, yang merupakan jumlah tetangga terdekat.
- b. Pada data training dilakukan, hitung kuadrat dari jarak Euclidean.
- c. Urutkan data tersebut dalam urutan menaik.
- d. Gabungkan kategori Y sesuai dengan parameter K..

Keunggulan metode *Naïve Bayes* pada proses klasifikasi ada sejumlah kecil data pelatihan yang diperlukan untuk penentuan estimasi parameter. Perbedaan yang lebih mendasar antara *teorema Bayes* untuk metode lain, parameter Bayesian dianggap sebagai variabel acak, sedangkan dalam statistik klasik parameternya tidak dapat diketahui. Hubungan antara probabilitas bersyarat kejadian H dan X disebut teorema Bayes. Nilai akurasi yang dihasilkan lebih tinggi daripada metode lainnya (Nur Adinda Salsabila, 2022).

Data yang dikumpulkan akan diproses terlebih dahulu agar tidak menjadi tidak akurat atau tidak relevan, sehingga hasil prediksi dapat diperhitungkan. Pada langkah pengklasifikasian teks atau sentimen akan dilalui *preprocessing* agar teks yang mempunyai isi tidak sempurna seperti salah ketik, tidak valid, data tidak relevan dan data hilang sehingga dari data yang memiliki kecacatan akan dihapus hal ini disebabkan karena data tersebut dapat mengurangi nilai akurasi. Data tersebut dibuang karena keberadaannya bisa mengurangi mutu atau akurasi. Adapun terdapat teks yang belum diproses biasanya terdapat *noise*, struktur teks yang berantakan dan dimensinya akan tinggi.

### E. Evaluasi Metode

Skenario yang diterapkan untuk melakukan evaluasi dan validasi agar memastikan validitas metode dan hasil penelitian, diusulkan model dengan menerapkan algoritma KNN dan *Naïve*

*Bayes*. Algoritma KNN dan *Naïve Bayes* diyakini dapat menghasilkan akurasi yang baik. Pada penelitian ini, *performance* akan diukur dengan *Accuracy* dan AUC dan akan ditampilkan berbentuk kurva ROC.  
(Wandani, 2021).

### III. Hasil dan Pembahasan

#### A. Pengumpulan Data

Data dikumpulkan melalui scraping web yang dilakukan pada laman video youtube dengan kata kunci “Kurikulum Merdeka” pada bulan Oktober 2023. Peneliti mengumpulkan data menggunakan bahasa pemrograman Python dan mengeksekusinya menggunakan Google Colab. *Library* yang digunakan untuk melakukan proses *crawling* adalah *googleapiclient*, dimana nantinya diminta untuk memasukkan *API key* dan *id* video *Youtube* yang akan dilakukan *crawling*. Setelah dilakukan proses penarikan data maka didapatlah data sebanyak 26.539 komentar.

#### B. Pengolahan Data

Pada tahapan ini, akan dilakukan proses pembersihan data komentar dari kata-kata dan symbol yang tidak memiliki makna adapun proses dalam tahapan ini adalah: 1) penghapusan data berulang (*remove duplicate*), 2) pemilihan *attribute* 3) penghapusan tanda baca (*replace*) 4) *filterisasi stopword*, 5) *tokenisasi*, dan 6) mencari frekuensi kata-kata yang sering muncul.

**Tabel 3.1** Hasil *Preprosesing Data*

| No | Sebelum Proses pembersihan  | Setelah Proses pembersihan  |
|----|---|---|
| 1  | Kurikulum merdeka<br>1. Pendidikan berpusat pada peserta didik<br>2. Pengalaman belajar yang menyenangkan<br>3. Pembelajaran yang mengembangkan kreatif dan inovatif <br>4. Pembelajaran yang kontekstual<br>5. Pembelajaran yang inklusi | kurikulum merdeka baru pendidikan berpusat pada peserta didik baru pengalaman belajar yang menyenangkan baru pembelajaran yang mengembangkan kreatif dan inovatif baru pembelajaran yang kontekstual baru pembelajaran yang inklusi |

|   |  |  |
|---|--|--|
| 2 | Sebagai guru SD sy berterimakasih untuk penerapan kurikulum merdeka. Karena dapat menghasilkan peserta didik yang mandiri, kreatif dan inovatif. | sebagai guru sudah saya berterimakasih untuk penerapan kurikulum merdeka karena dapat menghasilkan peserta didik yang mandiri kreatif dan inovatif |
| 3 | Isinya buzzer semua kah ? Komenan kagak ada yang kreatif semua tinggal copas   | isinya buzzer semua kah komenan tidak ada yang kreatif semua tinggal copas   |
| 4 | Kurikulumnya merdeka muridnya sengsara   | kurikulumnya merdeka muridnya sengsara   |
| 5 | Pembelajaran pada kurikulum merdeka terpusat pada siswa, sehingga diharapkan menjadi generasi muda yang lebih mandiri, kreatif dan inovatif      | pembelajaran pada kurikulum merdeka terpusat pada siswa sehingga diharapkan menjadi generasi muda yang lebih mandiri kreatif dan inovatif          |
| 6 | Kurikulum merdeka bertujuan memperbaiki sistem pendidikan Indonesia yang mengedepankan konsep pembelajaran yang berpusat pada peserta didik.     | kurikulum merdeka bertujuan memperbaiki sistem pendidikan indonesia yang mengedepankan konsep pembelajaran yang berpusat pada peserta didik        |

### C. Analisis Sentimen

Setelah dilakukan perbersihan data komentar maka proses selanjutnya adalah pemberian lebel positif, negatif dan netral pada setiap komentar, pelebelan menggunakan metode pelebelan otomatis dengan kamus *lexicon* bahasa Indonesia pada *library phton*.

**Tabel 3.2**  
 Pemberian Kategori Sentimen

| <b>Sebelum Dibersihkan</b>   | <b>Setelah Dibersihkan</b>  | <b>Kategori</b> |
|--|---|-----------------|
| Kurikulum merdeka<br>1. Pendidikan berpusat pada peserta didik<br>2. Pengalaman belajar yang menyenangkan<br>3. Pembelajaran yang mengembangkan kreatif dan inovatif<br>4. Pembelajaran yang kontekstual<br>5. Pembelajaran yang inklusi | kurikulum merdeka baru pendidikan berpusat pada peserta didik baru pengalaman belajar yang menyenangkan baru pembelajaran yang mengembangkan kreatif dan inovatif baru pembelajaran yang kontekstual baru pembelajaran yang inklusi | positif         |
| Sebagai guru SD sy berterimakasih untuk penerapan kurikulum merdeka. Karena dapat menghasilkan peserta didik yang mandiri, kreatif dan inovatif.   | sebagai guru sudah saya berterimakasih untuk penerapan kurikulum merdeka karena dapat menghasilkan peserta didik yang mandiri kreatif dan inovatif  | positif         |
| Isinya buzzer semua kah ?<br>Komenan kagak ada yang kreatif semua tinggal copas  | isinya buzzer semua kah komenan tidak ada yang kreatif  | negatif         |

|  |   |         |
|--|---|---------|
|  | semua tinggal copas   |         |
| Kurikulumnya merdeka muridnya sengsara   | kurikulumnya merdeka muridnya sengsara  | negatif |
| Pembelajaran pada kurikulum merdeka terpusat pada siswa, sehingga diharapkan menjadi generasi muda yang lebih mandiri, kreatif dan inovatif  | pembelajaran pada kurikulum merdeka terpusat pada siswa sehingga diharapkan menjadi generasi muda yang lebih mandiri kreatif dan inovatif   | netral  |
| Kurikulum merdeka bertujuan memperbaiki sistem pendidikan Indonesia yang mengedepankan konsep pembelajaran yang berpusat pada peserta didik. | kurikulum merdeka bertujuan memperbaiki sistem pendidikan indonesia yang mengedepankan konsep pembelajaran yang berpusat pada peserta didik | netral  |

#### D. Implementasi Algoritma

Pada tahap ini peneliti membagi terlebih dahulu data uji dan data latih dari dataset yang sebelumnya sudah diproses. penulis membagi rasio pembagian data latih dan sebanyak 80% dan data uji sebanyak 20%. Pembagian rasio ini dilakukan pada kedua model klasifikasi yaitu Naïve bayes dan K-Nearest Neighbor. Model klasifikasi K-Nearest Neighbor (KNN) setelah pembagian rasio tersebut kemudian akan dilakukan uji variasi nilai k. Nilai k terbaik dicari dari bantuan algoritma KNN, pembobotan word2vec dan *10-fold cross validation*(Yanti, 2021). Nilai k terbaik akan terlihat dari nilai *error* terkecil dengan variasi k yang akan diuji yaitu 10 nilai ganjil pertama dimulai dari k=1 sampai k=19.

Tahap terakhir yaitu tahap *asses* pada penelitian ini. Tahap ini dilakukan evaluasi dari 2 model klasifikasi yaitu Naïve Bayes dan K-Nearest Neighbor. Evaluasi tersebut berisikan nilai *confusion matrix*, nilai akurasi, nilai presisi, nilai *recall*, dan *f1-score* dari data uji pada masing masing model klasifikasi yaitu Naïve Bayes dan KNN. Kemudian metode KNN selanjutnya akan dilakukan *10-fold cross validation* dalam pencarian nilai variasi k terbaik. Hasil nilai akurasi, presisi, *recall*, dan *f1-score* dari masing-masing metode klasifikasi. Berikut hasil rangkuman hasil nilai akurasi, presisi, *recall*, dan *f1-score*.

**Tabel 3.3**  
 Pengujian Algoritma *Naïve Bayes*

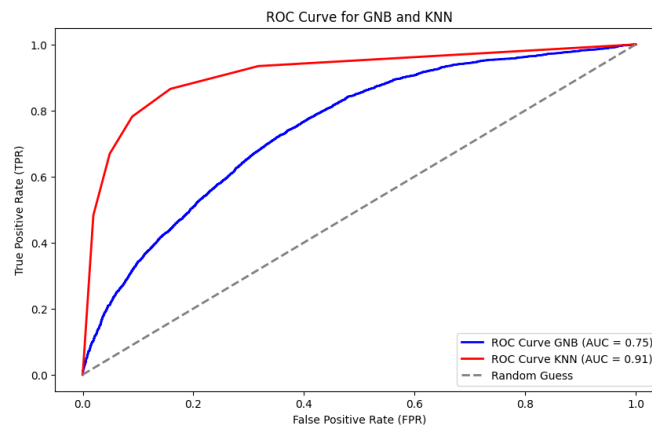
| Rasio   | Naïve Bayes |         |               |           |
|---------|-------------|---------|---------------|-----------|
|         | Akurasi     | Presisi | <i>Recall</i> | <i>f1</i> |
| 80 : 20 | 53.82 %     | 72%     | 53.82%        | 59%       |

**Tabel 3.4** Pengujian Algoritma KNN

| Rasio   | K-Nearest Neighbor |         |               |           |
|---------|--------------------|---------|---------------|-----------|
|         | Akurasi            | Presisi | <i>Recall</i> | <i>f1</i> |
| 80 : 20 | 79.70 %            | 77.20 % | 79.70 %       | 77.65%    |

**E. Evaluasi Metode**






Untuk mengevaluasi kinerja dari algoritma *Naïve Bayes* dan KNN akan dihitung menggunakan Kurva ROC dan menghitung luas area di bawah Kurva ROC atau disebut juga *Area Under Curve* (AUC). Kurva ROC menggambarkan hubungan antara *observed class* dan *predicted class*(Lara, 2022).



Gambar 3.1 Kurva ROC dengan Nilai AUC

Berdasarkan Gambar 4.18, didapatkan nilai AUC sebesar 0,75 pada algoritma *gaussian naïve bayes* sehingga klasifikasi nya masuk kedalam kategori cukup sedang untuk algoritma KNN mendapatkan nilai AUC sebesar 0,91 maka hasil klasifikasi algoritma KNN masuk kedalam kategori sangat baik. Pernyataan ini dapat kita lihat pada Tabel 3.5.

**Tabel 3.5** Nilai AUC

| <i>AUC</i>   | <i>Meaning</i>                  | <i>Symbol</i>   |
|--------------|---------------------------------|---|
| 0.90 – 1.00  | <i>Excellent classification</i> |  |
| 08.80 – 0.90 | <i>Good classification</i>      |  |
| 0.70 – 0.80  | <i>Fair classification</i>      |  |
| 0.60 – 0.70  | <i>Poor classification</i>      |  |
| <0.60        | <i>Failure</i>                  |  |

Sumber: Gorunescu dalam Wahono (2014) dalam (Fuad Nur Hasan, 2018)

Sehingga dapat kita simpulkan dalam penelitian ini algoritma KNN mampu bekerja lebih baik dibanding dengan algoritma *Gaussian Naïve Bayes*

## SIMPULAN

Kesimpulan dari hasil analisa yang dilakukan dalam penelitian ini peneliti tentang analisis sentimen mengenai implementasi kurikulum merdeka pada satuan pendidikan di platform video YouTube melalui algoritma *Naïve bayes* dan *K-Nearest Neighbor* peneliti menemukan algoritma terbaik dalam penelitian ini adalah *K-Nearest Neighbor* dengan akurasi 79,70%, presisi 77,20%, recall 79,70%, dan *f1-score* 77,65%. Sementara *Naïve Bayes* kinerjanya lebih rendah dengan nilai akurasi, presisi, *recall*, dan *f1-score* adalah sebesar 53.82%, 72%, 53.82% dan 59% hal ini disebabkan oleh asumsi independensi fitur tidak terpenuhi, karena dalam dataset penelitian ini banyak kata-kata yang sifatnya saling terikat dan ketidak seimbangan kelas. Dari 26.539 komentar, 19.979 positif, 5.108 negatif, dan 1.398 netral, menunjukkan sentimen positif lebih dominan.

## REFERENSI

- Anggara, E. D., Widjaja, A., & Suteja, B. R. (2022). Prediksi Kinerja Pegawai sebagai Rekomendasi Kenaikan Golongan dengan Metode Decision Tree dan Regresi Logistik. *Jurnal Teknik Informatika Dan Sistem Informasi*, 8(1), 218–234. <https://doi.org/10.28932/jutisi.v8i1.4479>
- Fuad Nur Hasan. (2018). *Analisis Sentimen Artikel Berita Tokoh Sepak Bola Dunia Menggunakan Algoritma Support Vector Machine Dan Naive Bayes Berbasis Particle Swarm Optimization*. Sekolah Tinggi Manajemen Informatika Dan Komputer Nusa Mandiri.

- Harpizon, H. A. R., Kurniawan, R., Iwan Iskandar, Salambue, R., Budianita, E., & Syafria, F. (2022). Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes. *JNKTI (Jurnal Nasional Komputasi Dan Teknologi Informasi)*, 5(1), 131–140.
- Kemendikbudristek. (2022). Buku Saku: Tanya Jawab Kurikulum Merdeka. *Kementerian Pendidikan, Kebudayaan, Riset Dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset Dan Teknologi*, 9–46.
- Khusnul Khotimah dan Husniyati. (2020). PENGELOLAAN BIG DATA PERPUSTAKAAN DENGAN TEKNIK DATA MINING. *Jurnal SIENNA*, 1(2), 1–5.
- Lara. (2022). ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI TANGERANG LIVE MENGGUNAKAN LATENT DIRICHLET ALLOCATION DAN NAÏVE BAYES. In *repository.uinjkt.ac.id* (Issue 8.5.2017).
- Nur Adinda Salsabila. (2022). Analisis Sentimen Pada Media Sosial Twitter Terhadap Tokoh Gus Dur Menggunakan Metode Naïve Bayes dan Support Vector Machine(SVM). In *repository.uinjkt.ac.id* (Issue 8.5.2017).
- Rahayu, R., Rosita, R., Rahayuningsih, Y. S., Hernawan, A. H., & Prihantini, P. (2022). Implementasi Kurikulum Merdeka Belajar di Sekolah Penggerak. *Jurnal Basicedu*, 6(4), 6313–6319. <https://doi.org/10.31004/basicedu.v6i4.3237>
- Rakhman, A. (2017). Prediksi Ketepatan Kelulusan Mahasiswa Menggunakan Metode Decision Tree Berbasis Particle Swarm Optimization (PSO). *Smart Comp :Jurnalnya Orang Pintar Komputer*, 6(1), 193–197.
- Sujadi, H. (2022). Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine. *INFOTECH Journal*, 8(1), 22–27. <https://doi.org/10.31949/infotech.v8i1.1883>
- Wandani, A. (2021). Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 651–665.
- Yanti, H. A. (2021). PENGOLAHAN DATA SEDERHANA MENGGUNAKAN R STUDIO. *Jurnal SIENNA*, 2, 1–9.