

ANALISIS SENTIMEN TERHADAP BOIKOT BRAND PRO-ISRAEL PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES (STUDI KASUS: STARBUCKS)

Nur Fadhilah Az-haari, Didi Juardi, Asep Jamaludin

Program Studi Informatika S1, Universitas Singaperbangsa Karawang

Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Telp. (0267) 641177

2010631170024@student.unsika.ac.id

ABSTRAK

Palestina-Israel merupakan dua negara yang masih terlibat dalam konflik peperangan yang berlangsung hingga saat ini. Sejak pecahnya konflik antara Israel dan Hamas pada 7 Oktober 2023, banyak produk yang dianggap mendukung Israel, termasuk produk-produk Amerika seperti Starbucks yang mengalami dampaknya. Penurunan harga ekuitas perusahaan besar kemungkinan terkait menanggapi krisis kemanusiaan yang timbul akibat serangan militer terhadap warga sipil Palestina, dilakukan boikot terhadap produk yang mendukung Israel. Penelitian ini menerapkan sistem klasifikasi untuk mengkategorikan sentimen dari tweet yang menggunakan hashtag #BoikotStarbucks dan #BoikotProIsrael. Hasil pengolahan opini atau sentimen dari penerapan algoritma *Naïve bayes* yang meruakan salah satu algoritma klasifikasi yang sederhana namun sangat efektif dalam pengolahan data teks, seperti klasifikasi sentimen pada tweet. Klasifikasi yang digunakan adalah metodologi KDD yang terbagi kedalam tiga kelas yaitu Positif, Netral dan Negatif melalui media sosial Twitter menggunakan penerapan algoritma *Naïve bayes* diperoleh hasil yaitu 130 label positif, 156 label negatif, dan 611 label netral. Nilai accuracy tertinggi terdapat pada model 90:10 dengan 71%, precision 24%, dan recall 33%. Selain menggunakan confusion matrix model diuji dengan grafik ROC yang menghasilkan nilai AUC tertinggi pada model 90:10 dengan nilai 0.689 yang dapat diartikan model ini kualitasnya good classification

Kata kunci : Analisis Sentimen, Boikot Starbucks, Naïve Bayes, Twitter, Lexicon

1. PENDAHULUAN

Perselisihan antara Palestina dan Israel merupakan salah satu pertikaian yang paling lama dan kompleks di dunia. Konflik ini telah berlangsung selama bertahun-tahun dan melibatkan berbagai aspek, termasuk isu wilayah, agama, sejarah, dan politik. seorang akademisi dan peneliti ekonomi dari FEB Universitas Indonesia, boikot produk *pro*-Israel dianggap beralasan karena Israel telah secara konsisten menunjukkan kebijakan apartheidnya terhadap Palestina selama lebih dari tujuh decade. Perhatian publik terfokus pada perang antara Israel dan Hamas karena insiden kekerasan dan dugaan genosida di wilayah Palestina. Dampaknya terlihat dalam perubahan harga saham dan minat beli masyarakat terhadap produk-produk yang mendukung Israel[1]. Sejak pecahnya konflik antara Israel dan Hamas pada 7 Oktober 2023, banyak produk yang dianggap mendukung Israel, termasuk produk-produk Amerika seperti *Starbucks* yang mengalami dampaknya.

Media social *Twitter* dapat berperan sebagai sarana komunikasi yang menghubungkan antara pemerintah dan masyarakat. Oleh karena itu peran aspirasi masyarakat sangat signifikan dalam kesuksesan kebijakan pemerintah. Keberhasilan suatu kebijakan yang diterapkan oleh pemerintah dapat dinilai, antara lain, dari tanggapan yang diberikan oleh masyarakat [2]. Pada penelitian ini diterapkan sistem klasifikasi untuk mengkategorikan sentimen dengan tujuan menganalisis polaritas suatu *tweet* dengan hastag #BoikotStarbucks dan #BoikotProIsrael,

apakah termasuk dalam kategori positif, negatif, atau netral. *Classifier* atau klasifikasi merupakan suatu metode yang dapat mengelompokkan data menjadi beberapa kelas dengan menggunakan metode *Naïve Bayes*

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian sebelumnya dilakukan oleh Kurniawan & Waluyo, (2022) menggunakan Analisis sentimen terhadap masyarakat Indonesia terkait pemindahan ibukota melalui Twitter yang dilakukan menggunakan metode *Naïve Bayes* dengan total 822 tweet. [3]

Pada penelitian lain terkait analisis sentimen yang dilakukan oleh Mahardhika & Zuliarso (2018) mempelajari Analisis sentimen terhadap pemerintahan Joko Widodo di platform Twitter dilakukan dengan menggunakan algoritma *Naïve Bayes Classifier* yang menunjukkan akurasi 97% saat melakukan analisis sentiment pada objek penelitian terhadap pemerintahan Joko Widodo.[4]

2.2. Analisis Sentimen

Analisis sentimen merupakan solusi yang memfasilitasi ekstraksi pendapat yang disederhanakan atau rincian perasaan terkait suatu topik atau konteks dari sejumlah besar sumber data (Bordoloi & Biswas, 2023). Analisis sentiment dijelaskan sebagai suatu prosedur untuk mengenali, mengidentifikasi, dan/atau mengkategorikan emosi atau pendapat pengguna

terkait dengan layanan tertentu, seperti film, produk, acara, atau atribut lainnya, pengategorian tersebut bisa mencakup dimensisebagai positif, negatif, atau netral [5].

2.3. Aplikasi X

Salah satu brand di dunia digital yang baru-baru ini melakukan rebranding adalah aplikasi X yang dulunya bernama Twitter. X merupakan aplikasi berupa platform media sosial asal Amerika Serikat sebagai penyedia layanan mikroblog agar penggunanya dapat saling berkomunikasi, berekspresi, dan menyebarkan informasi dengan berbagi ide, perspektif, dan pengalaman mereka sendiri, pengguna dapat membangun rasa koneksi dan komunitas dengan orang lain di platform

2.4. Data Mining

Data Mining merupakan salah satu bidang yang mengalami perkembangan pesat karena adanya kebutuhan besar akan nilai tambah dari database dengan skala besar. Definisi data mining adalah proses ekstraksi informasi atau pengetahuan dari data yang sebelumnya tidak diketahui, bersifat tertutup, dan dianggap tidak berguna, menjadi informasi, pengetahuan, atau pola dari data yang berskala besar [6]

2.5. Naïve Bayes Classifier

(Nurtikasari et al., 2022) mendeskripsikan Naive Bayes sebagai algoritma klasifikasi probabilistik yang sangat efisien dalam melakukan analisis sentimen pada teks. Algoritma ini berasaskan pada teorema Bayes, suatu metode untuk memperkirakan kemungkinan terjadinya suatu kejadian berdasarkan data yang ada. [7]

2.6. Confusion Matrix

Confusion matrix merupakan alat yang digunakan dalam mengevaluasi model klasifikasi untuk mengestimasi keberhasilan dan kegagalan prediksi objek. Sebuah matriks yang membandingkan prediksi dengan kelas aktual atau memberi informasi mengenai nilai yang sesungguhnya serta prediksi dalam suatu pengklasifikasian. [8]

Tabel 1. Rumus Confusion Matrix

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP (True Positive)	FP (False Positive)
	FALSE	FN (False Negative)	TN (True Negative)

Keterangan:

- TP : Jumlah kasus positif yang terklasifikasi benar
- FP : Jumlah kasus negatif yang terklasifikasi positif
- TN : Jumlah kasus negatif yang terklasifikasi benar
- FN : Jumlah kasus negatif yang terklasifikasi negative

2.7. Python

Python sebagai bahasa pemrograman yang menjadi salah satu yang paling diminati dan umum diterapkan di berbagai bidang, khususnya dalam analisis data. Python dibuat pada tahun 1990 oleh Guido van Rossum, yang menonjol karena fleksibilitas dan kemudahan penggunaannya, menjadikannya pilihan utama bagi banyak ilmuwan data, peneliti, dan pengembang perangkat lunak. Kelebihan Python melibatkan ketersediaan berbagai pustaka dan kerangka kerja kuat seperti NumPy, Pandas, dan Matplotlib, yang mempermudah proses pengolahan, manipulasi, analisis, dan visualisasi data dengan efisien [9]

2.8. Crawling Data

Pengambilan data merupakan proses pengambilan suatu data dari suatu database. Crawling juga merupakan metode yang diterapkan untuk menghimpun informasi yang berasal dari web. Crawling beroperasi secara otomatis, di mana informasi dikumpulkan berdasarkan kata kunci yang telah diberikan oleh pengguna [10]

2.9. Metode Lexicon

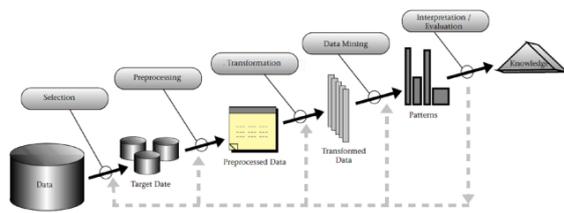
Metode Lexicon adalah metode yang melibatkan frasa, ungkapan, atau isi teks yang umumnya ditemukan dalam percakapan, dialog, pesan, ulasan, dan sejenisnya. Pendekatan Lexicon menggunakan daftar kata sentiment yang terdiri dari kata-kata positif, netral, dan negatif yang dikaitkan dengan kata-kata dalam kalimat untuk menentukan tingkat polaritasnya [11] Dalam kamus tersebut setiap kata telah diberikan skor polaritas yang berkisar dari -1 untuk kategori negatif hingga +1 untuk kategori positif [12]).

2.10. Receiver Operating Characteristic (ROC)

Kurva ROC (Receiver Operating Characteristic) adalah salah satu metode evaluasi yang sering digunakan untuk mengukur performa sistem klasifikasi. Untuk membandingkan dan menentukan keunggulan classifier, digunakan penghitungan AUC (Area Under Curve) yang merupakan luas area di bawah kurva. AUC selalu memiliki nilai di antara 0 hingga 1, dihitung berdasarkan luas gabungan trapesium dari titik-titik sensitivitas dan spesifisitas([13]

3. METODE PENELITIAN

Knowledge Discovery in Databases (KDD) dapat digunakan dalam melakukan proses data mining. KDD merupakan suatu proses data mining untuk menghasilkan informasi berharga, serta rancangan yang terkandung di dalam data. Untuk mengidentifikasi pola-pola yang ada dalam data dapat menggunakan algoritma KDD [14].



Gambar 1. Alur penelitian KDD

Langkah utama dalam proses KDD tergambar dalam gambar 1, yang melibatkan serangkaian tahapan sebagai berikut:

3.1. Selection

Proses selection melibatkan pengumpulan data, di mana dataset terdiri dari 1887 data yang diperoleh dari Twitter melalui teknik crawling di Google Colab. Untuk melakukan crawling Twitter, peneliti perlu memperoleh auth-token terlebih dahulu. Auth-token ini merupakan kode unik yang digunakan untuk mengidentifikasi dan mengautentikasi pengguna Twitter. Google Colab, sebuah platform cloud yang memungkinkan eksekusi kode Python secara online, dapat dimanfaatkan oleh peneliti untuk memperoleh auth-token tersebut. Proses pengambilan data dari Twitter dapat dilakukan langsung di Google Colab.

3.2. Preprocessing

Preprocessing adalah tahap di mana data dibersihkan dan disesuaikan. Saat mengumpulkan data, informasi yang diperoleh biasanya tidak terstruktur dan mengandung banyak karakter yang tidak relevan. Tujuan dari preprocessing adalah menghilangkan noise tersebut. Pada tahap ini, terdapat beberapa langkah yang perlu dilakukan:

- Cleaning* adalah proses menghapus data yang tidak relevan, seperti emotikon, hashtag, nama pengguna, URL, atau karakter lain yang tidak berguna.
- Case Folding* adalah tahap mengubah teks menjadi huruf kecil.
- Tokenization* adalah tahap memecah kalimat menjadi kata-kata.
- Filter by length* adalah salah satu metode untuk menyeleksi data berdasarkan kriteria tertentu.
- Stopword* adalah tahap menghapus kata-kata yang tidak memiliki nilai pada proses data mining, seperti kata sambung.
- Stemming* adalah tahap menghapus imbuhan pada kata.
- Pelabelan adalah tahap memberikan label pada kalimat agar memiliki nilai sentimen.

3.3. Transformation

Transformasi data adalah proses yang dilakukan untuk merubah struktur atau format data sehingga sesuai untuk diolah dalam proses data mining yang memiliki tujuan untuk memastikan bahwa

karakteristik data yang akan digunakan dalam analisis memenuhi persyaratan algoritma atau model yang akan digunakan. Pada tahap ini, akan dilakukan penghitungan bobot kata menggunakan metode TF-IDF.

3.4. Data Mining

Proses penemuan pola atau informasi yang menarik dengan menerapkan teknik, metode, atau algoritma khusus. Variasi metode, algoritma, atau teknik yang ada dalam data mining cukup besar. Dalam penelitian ini, kami memilih algoritma Naïve Bayes untuk mengklasifikasikan data. Sebelum melakukan data mining, data dibagi menjadi dua bagian: data pelatihan dan data pengujian, dengan rasio 90% untuk pelatihan dan 10% untuk pengujian. Pada tahap ini, analisis sentimen dilakukan menggunakan algoritma Naive Bayes Classifier untuk menghitung nilai *accuracy*, *precision*, *recall*, dan *F-measure* dari pengklasifikasi.

3.5. Evaluation

Peneliti menggunakan metode pengujian untuk mengevaluasi kualitas model klasifikasi yang dikembangkan. Model klasifikasi diuji dengan menghitung akurasi, yang merupakan rasio antara jumlah rekaman yang diklasifikasikan dengan benar dan jumlah total rekaman. Hasil pengujian disajikan dalam bentuk tabel matriks konfusi yang mencakup empat nilai: true positif (TP), false negatif (FN), true negatif (TN), dan false positif (FP). Peneliti memanfaatkan matriks konfusi sebagai alat evaluasi untuk mengukur seberapa baik model dapat membedakan antara klasifikasi yang benar atau salah. Selain menggunakan matriks konfusi, performansi model juga dapat diukur dengan nilai AUC (Area Under Curve) dan kurva ROC (Receiver Operating Characteristics). Nilai AUC untuk semua model Naive Bayes Multinomial dapat ditemukan dalam Tabel Berikut:

Tabel 2. Kategori Pengklasifikasian AUC

Nilai AUC	Kategori Pengklasifikasian
0.80 - 1.00	<i>Excellent Classification</i>
0.60 - 0.80	<i>Good Classification</i>
0.40 - 0.60	<i>Fair Classification</i>
0.20 - 0.40	<i>Poor Classification</i>
0.00 - 0.20	<i>Failur Classification</i>

4. HASIL DAN PEMBAHASAN

4.1. Crawling Data

Hasil dari proses seleksi data dilakukan dengan melakukan crawling data dari media sosial Twitter menggunakan kata kunci "Boikot Starbucks" dalam bahasa Indonesia. Proses crawling data ini menggunakan akses ke Twitter API. Contoh dataset awal yang dihasilkan dari crawling ini dapat dilihat pada Gambar 2.

Gambar 2. Hasil Crawling Data

Jumlah data awal yang dihasilkan dari proses *crawling* ini adalah sebanyak 1887, waktu yang digunakan pada saat proses *crawling* data mulai dari tanggal 1 Januari hingga 1 Maret 2024. Dataset awal ini terdiri dari beberapa atribut, termasuk *username*, *created_at*, dan *full_text*.

4.2. Cleaning

Cleaning atau pembersihan adalah tahap dalam preprocessing data di mana data mentah dibersihkan dari informasi yang tidak relevan dan menghapus data yang hilang, menormalkan teks (mengonversi teks menjadi huruf kecil), menghapus tanda baca, mengatasi kata-kata yang salah eja, atau menghapus entitas yang tidak relevan. Hasil dari proses *cleaning* dapat dilihat pada tabel 3 berikut :

Tabel 3. Hasil Cleaning

Sebelum	Sesudah
@bharianmy Teruskan boikot sampai Starbucks keluar dri Malaysia!!!!	Teruskan boikot sampai Starbucks keluar dri Malaysia

4.3. Case Folding

Case Folding yaitu mengubah semua huruf kapital dalam dataset menjadi huruf kecil, untuk menormalkan karakter dalam dataset, sehingga pada proses tokenizing dapat dilakukan dengan lebih mudah. Hasil penenrapan tahap *Case Folding* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Case Folding

Sebelum	Sesudah
Teruskan boikot sampai Starbucks keluar dri Malaysia	terus boikot starbucks malaysia

4.4. Tokenizing

Pada tahap *tokenizing*, data *tweet* yang sebelumnya dalam bentuk kalimat akan dipecah menjadi kata-kata individual untuk memudahkan saat memasuki tahap transformasi, karena tidak lagi memproses kalimat secara keseluruhan tetapi memproses kata-kata secara terpisah dari kalimat tersebut. Hasil dari proses *Tokenizing* dapat dilihat pada tabel 5 berikut :

Tabel 5. Hasil Tokenizing

Sebelum	Sesudah
Teruskan boikot sampai Starbucks keluar dri Malaysia	['terus', 'boikot', 'starbucks', 'malaysia']

4.5. Stopword

Tahap *stopword* untuk kata-kata yang tidak relevan akan disaring untuk proses klasifikasi selanjutnya. Misalnya, kata-kata seperti "yang", "dan", "di", "dari", dan "nya" biasanya tidak memberikan kontribusi signifikan dalam klasifikasi dan akan dihapus dengan menggunakan dokumen yang berisi daftar kata-kata *stopwords* untuk membantu dalam menyaring kata-kata yang tidak diperlukan dalam proses klasifikasi. Hasil dari proses *Stopword* dapat dilihat pada tabel 6 berikut :

Tabel 6. Hasil Stopword

Sebelum	Sesudah
Teruskan boikot sampai Starbucks keluar dri Malaysia	['terus', 'boikot', 'starbucks', 'malaysia']

4.6. Stemming

Stemming melibatkan penghapusan imbuhan dalam sebuah kata yang mungkin terletak di awal, akhir, atau kombinasi keduanya untuk mengurangi frekuensi dari sebuah kata turunan. Dalam tahap *Stemming* ini, biasanya digunakan *library* Sastrawi pada *Python* untuk membantu proses ini. Berikut *source code* yang digunakan pada tahap *stemming*. Hasil dari proses *Stemming* dapat dilihat pada tabel 7 berikut :

Tabel 7. Hasil Stemming

Sebelum	Sesudah
Teruskan boikot sampai Starbucks keluar dri Malaysia	['terus', 'boikot', 'starbucks', 'malaysia']

4.7. Labeling

Setelah proses *stemming* selesai langkah selanjutnya adalah melakukan labeling pada data menggunakan *lexicon Vader*. *Lexicon Vader* menghasilkan nilai *compound* yang merupakan rata-rata tertimbang dari sentimen yang terdapat dalam ulasan[15] Pembobotan sentimen dengan menggunakan kamus *lexicon* dan *negative words*, berhasil mengelompokkan total 897 data ke dalam tiga kelas. Di antaranya, terdapat 130 data yang termasuk ke dalam kelas positif, 156 data masuk ke dalam kelas negatif, dan 611 data diklasifikasikan ke dalam kelas netral.

4.8. Transformation Data

Melakukan pembobotan kata dengan menerapkan metode *Term Frequency – Inverse Document Frequency (TF-IDF)*. Proses ini menggunakan fungsi *TfidfVectorizer* yang tersedia dalam pustaka *sklearn*. Berikut adalah hasil dari proses pembobotan *TF-IDF* yang diterapkan pada dataset, yang dapat ditemukan pada Gambar 5 berikut:

- [2] E. M. Thoriq, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Opini Publik pada Media Sosial Twitter terhadap Vaksin Covid-19 menggunakan Algoritma Support Vector Machine dan Term Frequency-Inverse Document Frequency," 2021. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [3] A. Kurniawan and S. Waluyo, "Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Pemindahan Ibukota Pada Twitter Application Of Naive Bayes Algorithm In Capital Movement Sentiment Analysis On Twitter," 2022. [Online]. Available: <https://senafiti.budiluhur.ac.id/index.php>
- [4] Y. S. Mahardhika and E. Zuliarso, *ANALISIS SENTIMEN TERHADAP PEMERINTAHAN JOKO WIDODO PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVES BAYES CLASSIFIER*. 2018.
- [5] S. Pandya and P. Mehta, "A Review On Sentiment Analysis Methodologies, Practices And Applications," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 9, p. 2, 2020, [Online]. Available: www.ijstr.org
- [6] M. A. Hairudin, H. Zainuddin, and Y. Wabula, "Rekomendasi Strategi Sosialisasi Program Studi Melalui Jalur Undangan Menggunakan Algoritma ID3 dan K-Means," *JITCE (Journal of Information Technology and Computer Engineering)*, vol. 6, no. 01, pp. 14–18, Mar. 2022, doi: 10.25077/jitce.6.01.14-18.2022.
- [7] Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes," *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 1, no. 4, pp. 411–423, Aug. 2022, doi: 10.55123/insologi.v1i4.770.
- [8] S. Juniarsih, E. F. Ripanti, and E. E. Pratama, "Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 3, p. 239, Jul. 2020, doi: 10.26418/justin.v8i3.39118.
- [9] Regina Lo *et al.*, "Penggunaan Bahasa Pemrograman Python dalam Menganalisis Hubungan Kualitas Kopi dengan Lokasi Pertanian Kopi," *Jurnal Publikasi Teknik Informatika*, vol. 2, no. 2, pp. 100–109, May 2023, doi: 10.55606/jupti.v2i2.1752.
- [10] A. Agrani and B. Rikumahu, "PERBANDINGAN ANALISIS SENTIMEN TERHADAP DIGITAL PAYMENT 'GO-PAY' DAN 'OVO' DI MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES DAN WORD CLOUD COMPARISON OF SENTIMENT ANALYSIS AGAINST DIGITAL PAYMENT 'GO-PAY' AND 'OVO' IN SOCIAL MEDIA TWITTER USING NAIVE BAYES ALGORITHM AND WORD CLOUD," *Agustus*, vol. 7, no. 2, p. 2534, 2020.
- [11] R. I. Syah, H. Hoiriyah, and M. Walid, "ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TERHADAP APLIKASI M-HEALTH PEDULI LINDUNGI DENGAN METODE LEXICON BASED DAN NAIVE BAYES," *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 6, no. 1, Jun. 2023, doi: 10.21927/ijubi.v6i1.3275.
- [12] M. Z. Nafan and A. E. Amalia, "Kecenderungan Tanggapan Masyarakat terhadap Ekonomi Indonesia berbasis Lexicon Based Sentiment Analysis," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 3, no. 4, p. 268, Oct. 2019, doi: 10.30865/mib.v3i4.1283.
- [13] B. Tri, R. Doni, S. Susanti, and A. Mubarak, "12~19 Naskah diterima 23 Januari 2021; direvisi 17 Februari," *JURNAL RESPONSIF*, vol. 3, no. 1, 2021, [Online]. Available: <http://ejournal.ars.ac.id/index.php/jti>
- [14] M. Agustina, "IMPLEMENTASI ALGORITMA K-MEANS PADA PERISTIWA GEMPA BUMI DI WILAYAH JAWA BARAT," 2023.
- [15] R. Sanjaya, E. Tohidi, E. Wahyudi, and M. Kec Kesambi Kota Cirebon, "ANALISIS SENTIMEN TERHADAP BERHENTINYA TIKTOKSHOP PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES," 2024.