Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 3, Maret 2025, Hal. 655-666
DOI: https://doi.org/10.52436/1.jpti.691
p-ISSN: 2775-4202
e-ISSN: 2775-4180

Perbandingan Metode SVM dan Naïve Bayes untuk Analisis Sentimen pada Twitter tentang Obesitas di Kalangan Gen Z

Nanda Arif Maulana[1], Dedi Darwis[*2]

[1,2]Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia
Email: [1]nanda_arif_maulana@teknokrat.ac.id, [2]darwisdedi@teknokrat.ac.id

Abstrak

Pada era digital yang berkembang pesat, media sosial menjadi platform utama dalam menyampaikan opini dan informasi, termasuk mengenai isu kesehatan seperti obesitas, terutama di kalangan Generasi Z. Penelitian ini bertujuan untuk membandingkan efektivitas metode Support Vector Machine (SVM) dan Naïve Bayes dalam mengklasifikasikan sentimen publik terhadap obesitas berdasarkan data dari Twitter. Dataset penelitian terdiri dari 4.056 tweet yang dikumpulkan melalui proses scraping, kemudian diproses melalui tahap pembersihan, tokenisasi, dan stemming. Hasil penelitian menunjukkan bahwa algoritma SVM memiliki akurasi 89,23%, lebih tinggi dibandingkan dengan Naïve Bayes yang hanya mencapai akurasi 72,14%. Keunggulan SVM terletak pada kemampuannya memisahkan kelas sentimen dengan hyperplane optimal, menjadikannya metode yang lebih andal untuk analisis sentimen berbasis media sosial. Hasil penelitian ini dapat membantu pembuat kebijakan kesehatan dalam memahami opini publik terhadap obesitas serta merancang strategi komunikasi yang lebih efektif berbasis data.

Kata kunci: *analisis sentimen, obesitas, generasi Z, SVM, naïve bayes*

Comparison of SVM and Naive Bayes Methods for Sentiment Analysis on Twitter About Obesity Among Gen Z

Abstract

In the rapidly developing digital era, social media has become the main platform for conveying opinions and information, including on health issues such as obesity, especially among Generation Z. This study aims to compare the effectiveness of the Support Vector Machine (SVM) and Naïve Bayes methods in classifying public sentiment towards obesity based on data from Twitter. The research dataset consists of 4,056 tweets collected through the scraping process, then processed through the cleaning, tokenization, and stemming stages. The results of the study showed that the SVM algorithm had an accuracy of 89.23%, higher than Naïve Bayes which only achieved an accuracy of 72.14%. The advantage of SVM lies in its ability to separate sentiment classes with optimal hyperplanes, making it a more reliable method for social media-based sentiment analysis. The results of this study can help health policymakers understand public opinion towards obesity and design more effective data-based communication strategies.

Keywords: *generation Z, naïve Bayes, obesity, sentiment analysis, SVM.*

1. PENDAHULUAN

Pada masa digital yang senantiasa berkembang, media sosial sudah dijadikan salah satu platform utama guna menyampaikan opini, berbagi pengalaman, dan menyuarakan berbagai isu yang relevan di masyarakat, termasuk isu kesehatan. Salah satu masalah kesehatan yang banyak diperbincangkan adalah obesitas, khususnya di kalangan Generasi Z. Generasi ini kerap kali menghadapi tantangan unik terkait pola makan, gaya hidup, dan persepsi masyarakat terhadap tubuh ideal. Diskusi tentang obesitas di media sosial, seperti Twitter, dapat mencerminkan beragam pandangan publik, mulai dari kesadaran terhadap pentingnya kesehatan hingga stigma sosial yang mungkin menyertainya [1].

Analisis sentimen merupakan pendekatan yang efektif untuk memahami persepsi publik berdasarkan data teks, seperti tweet. Dengan memanfaatkan metode ini, kita dapat mengevaluasi opini masyarakat, baik yang positif, negatif, maupun netral terhadap obesitas di kalangan Gen Z. Naive Bayes dikenal sebagai metode probabilistik yang sederhana namun efisien untuk mengklasifikasikan teks, sementara SVM menawarkan akurasi tinggi dalam menangani data dengan dimensi yang besar [2].

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Penelitian sebelumnya menunjukkan bahwa algoritma SVM sering kali memberikan hasil yang lebih unggul daripada Naive Bayes dalam analisis sentimen. Sebagai contoh, sebuah penelitian yang menganalisis opini masyarakat tentang layanan e-commerce menggunakan data dari Twitter menggambarkan bahwa SVM mampu menembus akurasi sejumlah 85%, sementara Naive Bayes hanya menembus 78% [3]. Selain itu, studi lain yang mengkaji sentimen terhadap kebijakan pemerintah di Indonesia menemukan bahwa SVM memberikan akurasi hingga 88%, sementara Naive Bayes mencatatkan akurasi sejumlah 82% [4]. Hal tersebut menunjukkan bahwa SVM memiliki potensi yang lebih optimal untuk mengatasi data teks melalui pola yang kompleks dibandingkan Naive Bayes.

Metode klasifikasi teks Naive Bayes dan SVM, mempunyai keunggulan masing-masing. Maka dari itu, penelitian memiliki tujuan guna membandingkan kinerja kedua algoritma untuk menganalisis sentimen terkait obesitas di kalangan Gen Z berdasarkan data Twitter. Selain itu juga untuk meningkatkan Nilai Akurasi. Melalui penelitian ini, diharapkan dapat ditemukan metode yang paling efektif untuk memahami opini publik, yang dapat membantu dalam merancang kebijakan kesehatan yang lebih baik dan relevan. Aplikasi X telah menjadi wadah bagi masyarakat untuk berbagi informasi, opini, dan kekhawatiran mereka terkait obesitas. Tanggapan yang muncul beragam, mulai dari kekhawatiran tentang meningkatnya angka obesitas, pandangan mengenai penyebab dan dampaknya terhadap kesehatan, hingga berbagi tips tentang diet dan olahraga. Namun, di samping diskusi yang informatif, banyak juga disinformasi yang beredar, seperti hoaks tentang diet instan, metode penurunan berat badan yang tidak sehat, serta klaim tentang suplemen yang dapat "membakar lemak" tanpa bukti ilmiah. Hal ini memperumit pemahaman terhadap masalah obesitas dan mengaburkan upaya untuk mengedukasi masyarakat tentang cara-cara yang tepat untuk mencegah dan mengelola obesitas. Berdasarkan World Health Organization (WHO), obesitas ialah bentuk dari masalah kesehatan global yang terus meningkat dan dapat menyebabkan beragam penyakit kronis, misalnya diabetes tipe 2, penyakit jantung, dan hipertensi (WHO, 2021). Maka dari itu, penting untuk selalu mencari informasi yang akurat dan berbasis bukti dari sumber yang terpercaya.

Tanggapan masyarakat di Aplikasi X sangat beragam. Beberapa pengguna memberikan komentar informatif atau netral, sementara yang lain menyampaikan kekhawatiran atau kepanikan terkait obesitas di kalangan Gen Z. Selain itu, ada juga yang menunjukkan dukungan dan optimisme terhadap upaya penanganan masalah obesitas. Namun, banyak pula yang memberikan respons negatif, seperti kritik terhadap gaya hidup modern yang dianggap memicu obesitas atau kurangnya edukasi kesehatan yang memadai [5]. Dalam situasi seperti ini, analisis sentimen dijadikan alat yang sangat penting dalam mempelajari bagaimana masyarakat merespons suatu isu [6]. Analisis sentimen memungkinkan peneliti untuk mendeteksi emosi atau pandangan yang tersirat dalam teks, baik dalam bentuk sentimen positif, negatif, maupun netral. Dengan menganalisis data besar dari platform media sosial, peneliti dan pembuat kebijakan dapat memperoleh wawasan yang lebih mendalam mengenai persepsi publik terhadap isu obesitas [7]. Ini dapat membantu dalam merumuskan kebijakan yang lebih baik serta mendeteksi potensi disinformasi yang beredar di masyarakat. Analisis sentimen juga memungkinkan pemetaan pola persepsi publik secara lebih cepat dan efektif, sehingga pihak terkait dapat mengambil tindakan yang lebih responsif dalam menangani kekhawatiran masyarakat [8].

Analisis sentimen yang membahas Tentang Obesitas di Media Twitter ini menggunakan metode Support Vector Machine (SVM) serta menerapkan kernel Radial Basis Function. Hasil terbaik diperoleh dengan akurasi 65%. Metode SVM menggambarkan akurasi yang cukup tepat untuk mendeteksi sentimen terkait obesitas [9]. Penelitian ini membahas tentang sentimen terkait COVID-19 di Twitter, yang banyak dibahas pada awal tahun 2020 di Indonesia. Sentimen yang terkandung dalam 1652 tweet dianalisis dengan menerapkan metode Naïve Bayes dan Support Vector Machine. Hasil dari analisis ini yang menunjukkan metode SVM meraih akurasi 81,6%, sementara Naïve Bayes mencatat akurasi sebesar 78,3% [10]. Dua metode yang populer dan sering diterapkan pada klasifikasi sentimen ialah Support Vector Machine (SVM) dan Naïve Bayes. SVM adalah metode yang beroperasi dengan memisahkan kelas-kelas sentimen melalui hyperplane yang paling optimal, sementara Naïve Bayes ialah algoritma berbasis probabilitas yang menerapkan teorema Bayes guna memprediksi kelas sentimen berdasarkan distribusi fitur yang terdapat dalam data [11]. Kedua metode ini memiliki karakteristik dan keunggulan masing-masing dalam hal akurasi, kecepatan komputasi, dan performa pada berbagai jenis dataset.

Perbandingan performa algoritma Naïve Bayes dan SVM untuk klasifikasi sentimen tanggapan pengguna game Genshin Impact. Dengan data ulasan dari Google Playstore, SVM mencapai akurasi 83%, lebih tinggi daripada Naïve Bayes dengan 75,5%. SVM direkomendasikan karena hasilnya lebih unggul dalam klasifikasi sentimen ulasan game online [12]. Analisis sentimen ulasan pengguna aplikasi pinjaman online di Twitter dilakukan menerapkan algoritma Naïve Bayes dan K-NN. Naïve Bayes menunjukkan presisi lebih tinggi 83% dibandingkan K-NN 74%, meskipun akurasi K-NN sedikit lebih baik. Hal ini menjelaskan bahwa Naïve Bayes lebih unggul pada klasifikasi sentimen [13]. Penelitian menggunakan algoritma Naïve Bayes Classifier model Gaussian Bedrtujuan guna mengklasifikasikan status gizi balita dilihat dari data Jenis Kelamin, Berat Badan, dan Tinggi Badan. Status gizi dikategorikan dalam enam kelas, termasuk gizi buruk dan obesitas. Hasil menggambarkan bahwa algoritma ini berhasil memperoleh akurasi sejumlah 81,85% [14]. Dalam usaha untuk

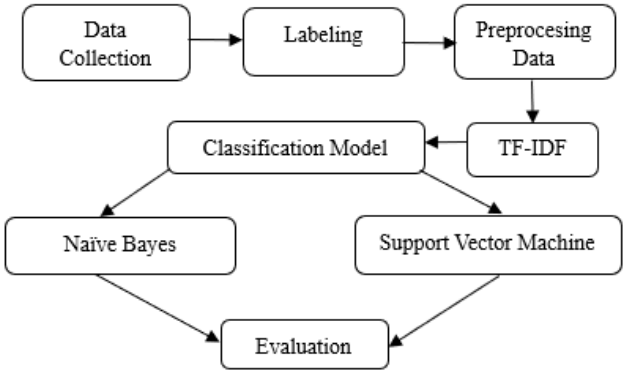
p-ISSN: 2775-4227
e-ISSN: 2775-4219

memahami sentimen publik secara digital, analisis sentimen algoritma SVM dan Naïve Bayes sering digunakan guna mengklasifikasikan teks berdasarkan polaritas sentimen [15]. SVM memanfaatkan hyperplane optimal untuk melakukan klasifikasi, sedangkan Naïve Bayes mengandalkan teorema Bayes guna perhitungan probabilitas kelas sebuah data [16].

Penelitian bertujuan untuk membandingkan kinerja kedua algoritma dalam mengklasifikasikan sentimen masyarakat mengenai obesitas dikalangan Gen Z di Aplikasi X serta untuk menentukan model yang lebih efektif dalam analisis sentimen ini. Melalui perbandingan tersebut, temuan diharapkan akan memperkaya ilmu yang lebih detail mengenai efektivitas masing-masing model. Hasil penelitian juga diharapkan dapat memberikan rekomendasi mengenai model klasifikasi sentimen yang paling optimal untuk diterapkan pada platform digital, sehingga dapat digunakan untuk meningkatkan pemahaman dan respons terhadap isu kesehatan masyarakat.

2. METODE PENELITIAN

Tujuan penelitian untuk menganalisis data terkait obesitas di kalangan Gen Z. Tahapan meliputi pengumpulan data dari berbagai sumber, preprocessing data untuk menghilangkan noise dan inkonsistensi, serta labeling data untuk mengidentifikasi entitas yang relevan. Proses ini bertujuan untuk mempersiapkan data yang berkualitas tinggi sebagai input untuk model pembelajaran mesin [17]. Setelah itu, dilakukan pembobotan kata untuk menentukan pentingnya setiap kata dalam analisis sentimen. Terakhir, mengevaluasi kinerja model Naïve Bayes dan SVM dalam mengklasifikasikan sentimen terkait Obesitas di kalangan Gen Z.



Gambar 1. Prosedur Penelitian

2.1. Pengumpulan Data

Dalam penelitian, data dikumpulkan melalui data scraping melalui ekstraksi ulasan pengguna aplikasi X. Dataset yang berhasil dikumpulkan sebanyak 5070 tweet, berisikan empat kolom yang terdiri: ’no’, ’created_at’, ’username’, dan ’full_text’ menggunakan pengembang Google Colab Kolom 'full_text' digunakan untuk menyimpan teks lengkap dari setiap entri, yang kemudian dianalisis menggunakan pendekatan berbasis sentimen. Analisis sentimen dilakukan dengan memanfaatkan fungsi getPolarity, yang didasarkan pada TextBlob untuk mengekstraksi skor polaritas dari setiap teks. Skor polaritas ini untuk mengukur sentimen positif atau negatif yang terkandung dalam teks. Setiap skor polaritas dihitung dan disimpan dalam kolom baru bernama ’Polarity’, sehingga dapat dianalisis lebih lanjut dalam penelitian ini [18].

2.2. Labeling

Proses labeling melibatkan penandaan data dengan kategori sentimen yang sesuai, seperti positif, negatif, atau netral. Penelitian bertujuan untuk memfasilitasi analisis sentimen selanjutnya dan menyediakan data pelatihan yang terstruktur untuk model klasifikasi. Untuk proses pelabelan, fungsi getLabel digunakan untuk mengonversi kategori sentimen menjadi format numerik. Dalam fungsi ini, label 'Positif' dikonversi menjadi 1, label 'Netral' menjadi 2, dan label lainnya, seperti 'Negatif', menjadi 0. Setelah itu, kolom baru bernama 'Label' ditambahkan ke dalam dataset dengan menerapkan fungsi ini pada kolom 'Analysis'. Hasilnya dapat dilihat pada beberapa baris pertama dataset dengan menggunakan data.head().

p-ISSN: 2775-4227
e-ISSN: 2775-4219

2.3. Preprocessing Data

Pengumpulan data dari aplikasi X dilakukan dengan mengekstrak informasi mengenai pengguna dan tweet yang relevan melalui kata kunci tertentu. Ekstraksi ini memanfaatkan Application Programming Interface (API) yang disediakan oleh platform X. Dalam penelitian ini, dataset dikumpulkan menggunakan teknik web scraping, khususnya dengan kata kunci "Obesitas". Proses pengumpulan ini dilakukan dengan menggunakan library Python yang dirancang untuk scraping tweet, memungkinkan pengambilan data secara otomatis dan efisien.

a. Cleaning Data

Proses menghapus data yang tidak relevan, duplikat, atau noise, guna memastikan kualitas dan integritas dataset yang akan dianalisis [19]. Dengan menghilangkan elemen yang dapat mengganggu analisis, model dapat dilatih menggunakan data yang lebih bersih dan representatif, yang berkontribusi pada hasil yang lebih akurat.

b. Case Folding

Langkah mengganti seluruh teks menjadi huruf kecil guna menjaga konsistensi dalam analisis teks dan menghindari duplikasi akibat perbedaan huruf kapital. Teknik ini penting karena variasi dalam penggunaan huruf kapital dapat menghasilkan kesalahan dalam pengelompokan kata, sehingga mengurangi keakuratan model.

c. Tokenizing

Tahap memecahkan teks menjadi unit-unit lebih kecil, yaitu kata dan token, yang mempermudah analisis lanjutan. Dengan mengubah teks menjadi token, model dapat lebih mudah mengidentifikasi pola dan hubungan antar kata, meningkatkan kemampuan klasifikasi sentimen.

d. Stopword

Menghapus kata umum yang tidak memberikan makna signifikan pada konteks analisis, sehingga fokus dapat diberikan pada kata-kata yang lebih informatif. Dengan menghilangkan stopword, analisis dapat difokuskan pada elemen-elemen yang lebih relevan dengan konteks, sehingga meningkatkan kekuatan sinyal dalam data yang dianalisis.

e. Stemming

Tahap mengubah kata-kata ke bentuk dasarnya dengan mengurangi variasi kata yang mempunyai makna sama, sehingga meningkatkan efisiensi analisis. Dengan mereduksi kata ke bentuk dasar, teknik ini meningkatkan efisiensi analisis dan mengurangi kompleksitas, memungkinkan model untuk menangkap lebih banyak informasi relevan dari data.

2.4. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk memberikan bobot pada kata-kata dalam dokumen. Metode ini membantu mengidentifikasi kata-kata yang lebih relevan dan signifikan dalam konteks dataset, sehingga meningkatkan kualitas analisis sentiment [20].

2.5. Naïve Bayes

Algoritma klasifikasi yang berbasis pada Teorema Bayes dengan asumsi sederhana bahwa fitur-fitur pada data adalah independen satu dengan lainnya. Walaupun asumsi tersebut terkadang tidak realistis, Naïve Bayes sering memberikan kinerja yang baik, terutama pada dataset dimensi tinggi. Keunggulan Naïve Bayes termasuk kecepatan pelatihan dan prediksi yang tinggi serta kinerja yang baik pada dataset yang cukup besar [21]. Dengan menggunakan Random Forest dan Naïve Bayes sebagai model dalam proses modelling, kita dapat memanfaatkan keunggulan masing-masing model untuk mendapatkan prediksi atau klasifikasi yang akurat dan berguna dalam analisis data. Selain itu, perbandingan kinerja kedua model ini juga dapat memberikan wawasan tambahan untuk menentukan model yang paling selaras untuk dataset dan tujuan analisis yang spesifik.

$$P(X/Y) = \frac{P(Y|X) \times P(X)}{P(Y)} \qquad (1)$$

Keterangan:

- $P(X \mid Y)$: Ialah probabilitas bahwa X terjadi apabila Y sudah terjadi.
- $P(Y \mid X)$: Ialah probabilitas bahwa Y terjadi jika X telah terjadi, dengan P(X) dan P(Y) sebagai probabilitas sebelumnya dari X dan Y secara berturut-turut.

2.6. Support Vector Machine (SVM)

Support Vector Machine ialah algoritma pembelajaran mesin yang kuat dalam melakukan klasifikasi dan regresi. Prinsip kerjanya berfokus pada pencarian hyperplane optimal yang memisahkan data menjadi kelas-kelas yang beragam [22]. Hyperplane ditentukan sedemikian rupa agar jarak antara hyperplane dengan data terdekat dari

p-ISSN: 2775-4227
e-ISSN: 2775-4219

masing-masing kelas (disebut margin) maksimal. Dengan memaksimalkan margin, SVM menghasilkan model yang lebih generalisasi dan robust terhadap noise dalam data. Berikut rumusnya:

$$f\ (x)\ =\ \{sign(w\}^{Tx} + b) \qquad (2)$$

Keterangan:

- w : Vektor bobot yang menentukan arah dan kemiringan hyperplane. Bobot ini diperoleh selama proses pelatihan model.
- x : Vektor fitur dari data baru yang ingin diklasifikasikan.
- b : Bias yang menyesuaikan posisi hyperplane dalam ruang fitur. Bias ini juga dihasilkan selama pelatihan.
- $w\ Tx + b$: Ini adalah fungsi linear yang menghitung jarak data xxx dari hyperplane. Hasilnya bisa positif, negatif, atau nol.
- sign () : Fungsi ini mengembalikan tanda dari nilai yang dihitung. Jika hasilnya lebih besar dari nol, fungsi ini mengembalikan +1 (menunjukkan bahwa data tersebut termasuk dalam kelas positif). Jika hasilnya kurang dari nol, fungsi ini mengembalikan -1 (menunjukkan bahwa data tersebut termasuk dalam kelas negatif).

2.7. Evaluation Model

Evaluasi model dilakukan untuk mengukur akurasi dan kinerja dari kedua metode (Naïve Bayes dan SVM) dalam mengklasifikasikan sentiment [23]. Proses ini mencakup pengujian menggunakan metrik seperti akurasi, presisi, recall, dan F1-Score untuk menilai efektivitas masing-masing model. Akurasi mengukur proporsi prediksi yang benar dari total prediksi, memberikan gambaran umum tentang efektivitas model. Presisi mengukur proporsi prediksi positif yang benar dibandingkan dengan total prediksi positif, penting untuk meminimalkan kesalahan dalam mengklasifikasikan data positif. Recall, atau sensitivitas, menilai proporsi prediksi positif yang benar terhadap seluruh data positif yang sebenarnya, sehingga sangat penting untuk menangkap sebanyak mungkin kasus positif. Dengan menggunakan metrik-metrik ini, evaluasi model memberikan pemahaman yang komprehensif tentang kinerja masing-masing algoritma dalam mengklasifikasikan sentimen, serta membantu dalam menentukan model yang paling sesuai untuk analisis lebih lanjut.

3. HASIL DAN PEMBAHASAN

Berdasarkan hasil analisis sentimen ini, diperoleh wawasan penting mengenai persepsi masyarakat terhadap Obesitas Dikalangan Gen Z. Pembahasan berikut akan menguraikan implikasi dari temuan tersebut dan hal yang melatarbelakanginya, dimulai dengan penjelasan mengenai proses pelabelan dan preprocessing data, yang kemudian diikuti oleh penerapan metode pemodelan Naïve Bayes dan SVM.

3.1. Pembagian Data

Penelitian menggunakan dataset yang berasal dari Twitter, berisi tweet yang mengandung kata “Obesitas” dari November 2023 - Oktober 2024. Proses pengumpulan data dilaksanakan melalui teknik scraping menerapkan Tweet Harvest, dan berhasil mengumpulkan total awal sebanyak 5070 record. Setelah menjalani proses pembersihan data, jumlah record yang tersisa adalah 4056, memastikan kualitas data yang digunakan dalam analisis. Selanjutnya, proses pelabelan dilakukan menggunakan library Python, TextBlob. TextBlob adalah library untuk pemrosesan teks yang menyediakan alat untuk analisis sentimen, pengenalan entitas, klasifikasi teks, dan banyak lag. Hasil pelabelan menunjukkan bahwa terdapat 4465 tweet dalam kelas netral, 403 tweet dalam kelas positif, dan 202 tweet dalam kelas negatif. Dataset yang telah dilabeli dilakukan pembagian 80:20 dianggap sebagai praktik terbaik yang banyak diterapkan dalam penelitian di bidang pembelajaran mesin (Zhang, 2021), di mana 80% dari total data (4056 record) dialokasikan untuk pelatihan model (training) dan 20% (1014 record) digunakan untuk menguji performa model (testing). Pembagian ini dirancang untuk memberikan evaluasi yang objektif terhadap kemampuan generalisasi model untuk memahami pola dan sentimen dari data baru yang tidak pernah dinilai sebelumnya, sehingga output dapat diandalkan untuk analisis lebih lanjut terkait Obesitas Dikalangan Gen Z.

3.2. Klasifikasi Menggunakan Algoritma Naïve Bayes dan SVM

Hasil pengujian model Naïve Bayes menggambarkan bahwa model mempunyai performa yang sangat baik untuk mengklasifikasikan sentimen netral. Terlihat pada tabel 2 dengan precision sebesar 0.01, recall 0,03 dan f1-score 0,01. Tetapi, pada kelas sentimen negatif dan positif, performa model jauh lebih baik. Sedangkan sentimen negatif, precision hanya mencapai

p-ISSN: 2775-4227
e-ISSN: 2775-4219

0.59 dan recall 1.00, sedangkan pada sentimen positif, precision sebesar 0,77 dan recall 0.99. Secara keseluruhan, model mencapai akurasi 71%, dengan macro average f1-score sebesar 0.48, yang mengindikasikan bahwa model kesulitan dalam mengklasifikasikan kelas minoritas. Sementara itu, weighted average f1-score sebesar 0.75 menunjukkan performa yang lebih baik karena kelas netral yang dominan dalam dataset.

Tabel 1. Hasil Naïve Bayes

Sentimen	***Precision***	***Recall***	***F1-Score***
Negatif	59%	100%	74%
Positif	77%	99%	87%
Netral	100%	94%	97%

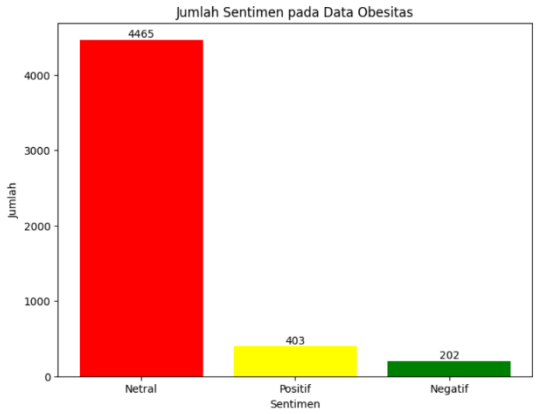
Tabel 3 hasil uji model SVM menggambarkan akurasi keseluruhan sejumlah 90%. Model memiliki performa sangat baik untuk mengklasifikasikan sentimen netral dengan precision sebesar 0,90 recall 1,00 dan f1-score 0.95. Pada kelas sentimen negatif, precision mencapai 1.00 dengan recall 0.00 dan f1- score 0.00, yang menunjukkan model cukup baik dalam mengidentifikasi sentimen negatif meskipun recall masih rendah. Namun, untuk sentimen positif, walaupun precision sempurna (1,00), recall hanya mencapai 0.17, menghasilkan f1-score 0.29, yang menunjukkan model kesulitan dalam mengenali sentimen positif secara konsisten.

Tabel 2. Hasil SVM

Sentimen	***Precision***	***Recall***	***F1-Score***
Negatif	100%	00%	00%
Positif	100%	17%	29%
Netral	90%	100%	95%

3.3. Pelabelan data

Proses selanjutnya adalah tahap memberi label untuk data, yaitu menentukan sentimen positif atau negatif berdasarkan kriteria tertentu. Untuk proses pelabelan ini, digunakan pustaka *Python* seperti *TextBlob*, yang mampu melakukan analisis sentimen berdasarkan nilai polaritas (*polarity*)



Gambar 2. Hasil Pelabelan

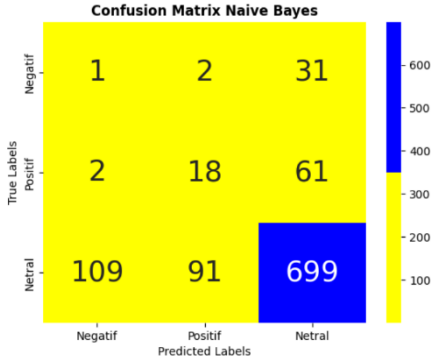
Berdasarkan gambar 2 yang menunjukan Analisis distribusi sentimen berdasarkan diagram menunjukkan adanya ketidakseimbangan yang signifikan dalam dataset. Sentimen netral berjumlah 4465 *tweet*, sentimen Positif dengan jumlah 403 *Tweet* dan Negatif dengan jumlah 202 *Tweet.*

3.4. Tahap Pengujian

Dalam tahap pengujian mengevaluasi kinerja dua algoritma klasifikasi, yaitu Naïve Bayes dan SVM, dengan menggunakan metode cross-validation. Menunjukkan hasil bahwa Naïve Bayes menghasilkan skor cross-validation dengan nilai 0.7080, 0.7179, 0.7199, 0.7228, dan 0.73865, yang memberikan akurasi rata-rata sebesar 72,14%. Sementara itu, Support Vector Machine berhasil menunjukkan kinerja terbaik dengan skor 0.9003, 0.8915, 0.8777, 0.8934, dan 0.8984, menghasilkan akurasi rata-rata tertinggi sebesar 89,23%.

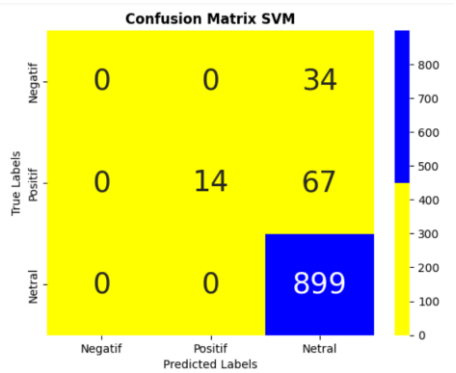
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Hasil menandakan bahwa SVM ialah model yang sangat efektif dalam klasifikasi data dalam penelitian ini, diikuti oleh Naïve Bayes. Walaupun Naïve Bayes mempunyai akurasi yang lebih rendah, model ini tetap akan menyajikan kinerja yang optimal dalam konteks analisis sentimen. Temuan tersebut memberikan wawasan penting dalam pemilihan model yang tepat untuk aplikasi analisis sentimen, serta menunjukkan bahwa pemilihan algoritma yang sesuai berperan signifikan terhadap kinerja model dalam pengolahan data. Dalam menentukan model algoritma yang paling optimal, penulis melaksanakan eksperimen dengan membandingkan nilai-nilai dalam confusion matrix dari dua algoritma yang diterapkan. Hasil perbandingan confusion matrix pada setiap model algoritma akan mengacu pada gambar 2 sebagai berikut:



Gambar 3. Confusion Matrix Naïve Bayes

Confusion matrix pada model Naïve Bayes menggambarkan bahwa dari total data yang diuji, model berhasil mengklasifikasikan 1 teks sebagai negatif dengan benar, tanpa ada teks positif yang salah diklasifikasikan sebagai negatif (false positives). Namun, terdapat 31 teks negatif yang salah dikategorikan sebagai netral atau positif (false negatives). Untuk kategori positif, model berhasil mengidentifikasi 18 teks dengan benar, tetapi terdapat 61 teks positif yang salah diklasifikasikan sebagai netral atau negatif. Sementara itu, pada kategori netral, model menunjukkan kinerja yang sangat baik dengan 699 teks yang benar-benar diidentifikasi sebagai netral, meskipun terdapat 109 teks negatif dan 91 teks positif yang salah dikategorikan sebagai netral. Hasil dari confusion matrix ini mengindikasikan bahwa model Naïve Bayes lebih efektif dalam mengenali teks positif dan netral daripada dengan teks negatif.



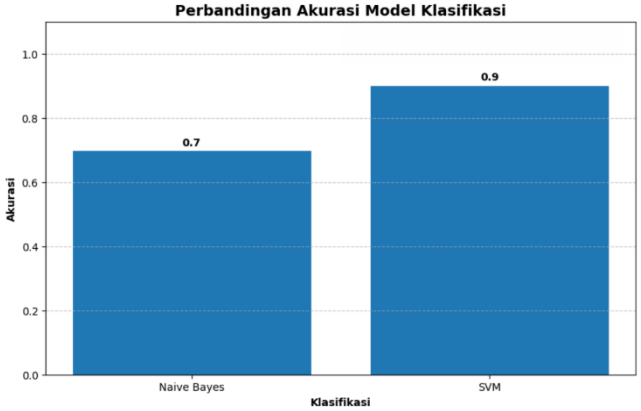
Gambar 4. Confusion Matrix SVM

Confusion matrix untuk model Support Vector Machine menggambarkan kinerja klasifikasi pada tiga kategori sentimen: negatif, positif, dan netral. Dari hasil pengujian, model berhasil mengklasifikasikan 0 data negatif dengan benar, tanpa kesalahan dalam mengidentifikasi data negatif sebagai positif. Namun, terdapat 34 data negatif yang salah diklasifikasikan sebagai netral. Model juga menunjukkan akurasi yang baik dalam mengidentifikasi data positif, dengan 14 data positif terklasifikasi dengan tepat, meskipun 67 data positif salah dikategorikan sebagai netral. Kategori netral memiliki performa yang sangat baik, dengan 600 prediksi benar, dan tidak ada kesalahan klasifikasi untuk kategori ini. Hasil ini menunjukkan bahwa meskipun Support Vector Machine efektif dalam mengenali sentimen positif dan netral, ada ruang untuk perbaikan dalam mengidentifikasi sentimen negatif.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.5. Perbandingan Akurasi

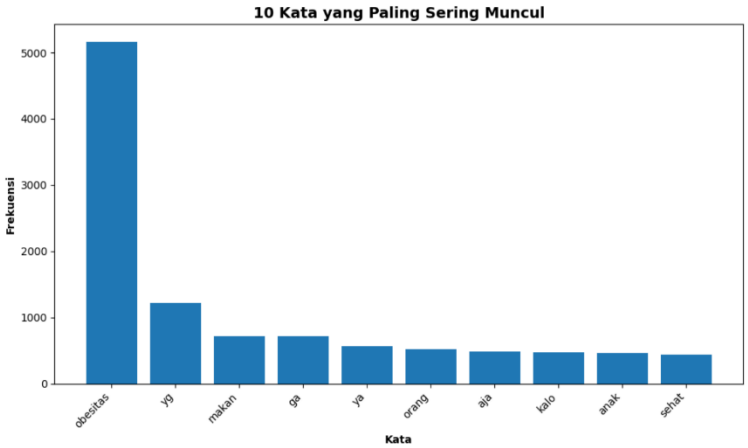
Visualisasi perbandingan adalah alat yang berguna untuk menggambarkan perbedaan atau perubahan antara beberapa kelompok data atau kondisi tertentu, sehingga mempermudah pemahaman dan pengambilan keputusan. Dalam konteks analisis data, visualisasi sering digunakan untuk menyoroti dampak dari suatu perubahan atau intervensi, seperti penerapan teknik tertentu pada data. Hasil perbandingan akurasi akan mengacu pada gambar 5.



Gambar 5. Perbandingan Akurasi Model Klasifikasi

3.6. Frekuensi Kata Teringgi

Frekuensi Kata Tertinggi adalah visualisasi yang menunjukkan kata-kata yang paling sering dijumpai pada dataset yang telah dianalisis. Dengan menggunakan grafik batang atau diagram lingkaran, visualisasi ini mengurutkan kata berdasarkan jumlah kemunculannya, sehingga kata-kata yang paling sering dijumpai akan terlihat jelas.



Gambar 6. Frekuensi Kata

3.7. WordCloud Sentimen

Penelitian ini menggunakan visualisasi wordcloud selaku alat dalam meneliti sekumpulan data teks. Visualisasi wordcloud pada analisis sentimen merupakan representasi grafis yang menunjukkan kata-kata yang paling sering dijumpai pada kumpulan teks. Masing-masing kata ditunjukkan dengan ukuran font yang bervariasi, sehingga sebuah kata yang lebih besar menggambarkan frekuensi kemunculan yang lebih tinggi daripada dengan kata yang lebih kecil.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Gambar 7. Semua Sentimen

Gambar 8. Sentimen Negatif

Gambar 9. Sentimen Netral

Gambar 10. Sentimen Positif

4. DISKUSI

Berdasarkan hasil penelitian, media sosial, khususnya Aplikasi X, telah menjadi platform penting dalam memahami persepsi publik terhadap obesitas di kalangan Generasi Z. Penelitian ini menunjukkan bahwa Aplikasi X tidak hanya menjadi tempat berbagi informasi tetapi juga menjadi medium yang rentan terhadap penyebaran disinformasi, seperti klaim palsu tentang diet instan atau suplemen penurunan berat badan tanpa dasar ilmiah. Hal ini mencerminkan kompleksitas pemahaman isu kesehatan masyarakat melalui media sosial.

Dalam penelitian, algoritma Support Vector Machine (SVM) dan Naïve Bayes (NB) diterapkan guna mengklasifikasikan sentimen dari 5.070 tweet, yang kemudian disaring menjadi 4.056 data bersih. Hasil evaluasi menggambarkan bahwa SVM unggul dengan akurasi tertinggi sejumlah 89,23%, sementara Naïve Bayes

p-ISSN: 2775-4227
e-ISSN: 2775-4219

mencatatkan rata-rata akurasi sebesar 72,14%. Perbedaan kinerja ini mengonfirmasi keunggulan SVM dalam menangani data teks yang kompleks dan beragam, terutama dalam mengklasifikasikan sentimen netral yang mendominasi dataset.

Hasil penelitian ini mendukung studi sebelumnya yang membahas implementasi metode Support Vector Machine (SVM) dan Naïve Bayes dalam klasifikasi data teks. Sebagai contoh, penelitian oleh Apriyani dan Kurniati (2020) yang menganalisis klasifikasi diabetes melitus dengan SVM dan Naïve Bayes menggambarkan bahwa algoritma SVM dengan kernel polynomial memperoleh akurasi tertinggi sejumlah 96,27%, sementara Naïve Bayes menghasilkan akurasi yang lebih rendah Hal ini menguatkan temuan bahwa SVM unggul dalam menangani data teks yang kompleks..

Penelitian ini menegaskan bahwa SVM adalah pilihan algoritma yang lebih andal untuk tugas analisis sentimen, terutama pada isu sensitif seperti obesitas, yang melibatkan opini publik yang sangat beragam. Untuk penelitian masa depan, disarankan untuk mengeksplorasi kombinasi metode atau pendekatan ensemble guna meningkatkan akurasi dan kemampuan generalisasi model pada data yang lebih besar dan kompleks.

5. KESIMPULAN

Dari hasil penelitian, maka ditarik kesimpulan media sosial, khususnya Aplikasi X, memainkan peran penting dalam mencerminkan pandangan publik terhadap isu kesehatan seperti obesitas di kalangan Generasi Z. Namun, tantangan utama yang dihadapi adalah maraknya disinformasi, seperti klaim palsu tentang metode penurunan berat badan, yang menyulitkan dalam memahami sentimen masyarakat secara akurat. Untuk mengatasi hal ini, algoritma Support Vector Machine (SVM) dan Naïve Bayes (NB) diterapkan guna analisis sentimen pada 4.056 data tweet. Hasil menunjukkan bahwa SVM lebih unggul dibandingkan NB dengan akurasi tertinggi sebesar 89,23%, sementara NB hanya mencapai rata-rata akurasi 72,14%. Temuan menggambarkan bahwa SVM lebih optimal untuk menangani data teks yang kompleks dan beragam, terutama dalam kategori sentimen netral yang mendominasi dataset. Penelitian ini memberikan dampak signifikan dalam membantu pembuat kebijakan kesehatan memahami pola sentimen masyarakat dan merancang strategi komunikasi yang lebih efektif. Pendekatan ini juga relevan untuk diterapkan dalam analisis sentimen isu kesehatan lainnya di masa depan, guna meningkatkan kesadaran masyarakat terhadap masalah kesehatan global. Penelitian mendatang dapat mengeksplorasi platform lain seperti Facebook atau Instagram untuk mendapatkan wawasan yang lebih luas. Selain itu, penggunaan model lain seperti deep learning dapat dieksplorasi untuk meningkatkan akurasi klasifikasi sentimen.

DAFTAR PUSTAKA

[1] F. F. Mailoa, “Analisis sentimen data Twitter menggunakan metode text mining tentang masalah obesitas di Indonesia,” J. Inf. Syst. Public Heal., vol. 6, no. 1, p. 44, 2021, doi: 10.22146/jisph.44455.

[2] Y. Yusra and M. Fikry, “Klasifikasi Tweet E-Commerce dengan Menggunakan Metode Support Vector Machine,” J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf., vol. 4, no. 2, p. 50, 2018, doi: 10.24014/coreit.v4i2.5205.

[3] W. Ningsih, B. Alfianda, R. Rahmaddeni, and D. Wulandari, “Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia,” MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.

[4] Ardiyansah and Parjito, “Perbandingan Metode Naïve Bayes dan Support Vector Machine Dalam Analisis Sentimen Terhadap Tokoh Publik,” Media Online, vol. 4, no. 6, pp. 2813–2821, 2024, doi: 10.30865/klik.v4i6.1928.

[5] D. Ananda and R. R. Suryono, “Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes,” J. Media Inform. Budidarma, vol. 8, no. 2, pp. 748–757, 2024, doi: 10.30865/mib.v8i2.7517.

[6] A. S. Rahayu, A. Fauzi, and R. Rahmat, “Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify,” J. Sist. Komput. dan Inform., vol. 4, no. 2, p. 349, 2022, doi: 10.30865/json.v4i2.5398.

[7] H. Herlawati, R. T. Handayanto, P. D. Atika, F. N. Khasanah, A. Y. P. Yusuf, and D. Y. Septia, “Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine,” J. Komtika (Komputasi dan Inform., vol. 5, no. 2, pp. 153–163, 2021, doi: 10.31603/komtika.v5i2.6280

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[8] A. E. Octaviyani, A. Y. Rangan, S. Informasi, K. Samarinda, R. Film, and F. B. Leksikon, "Analisis Sentimen X Untuk Review Film Agak Laen X Sentiment Analysis For Somewhat Laen Movie Reviews Using Naïve Bayes Method," vol. 28, no. 2, pp. 1–8, 2024, doi: 10.46984/sebatik.v28i2.0000.

[9] E. D. Nurcahya, "Klasifikasi Penyakit Ayam Menggunakan Metode Support Vector Machine," VOLT J. Ilm. Pendidik. Tek. Elektro, vol. 2, no. 1, p. 45, 2017, doi: 10.30870/volt.v2i1.1316.

[10] H. Sujadi, "Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine," INFOTECH J., vol. 8, no. 1, pp. 22–27, 2022, doi: 10.31949/infotech.v8i1.1883.

[11] M. Azlan, S. Putra, I. Permana, and M. Afdal, "Sentiment Analysis of Childfree Campaign on X Social Media Using NBC and SVM Algorithms Analisis Sentimen Masyarakat Mengenai Gerakan Childfree di Media Sosial X Menggunakan Algoritma NBC dan SVM," vol. 4, no. 1, pp. 1189–1198, 2024.

[12] M. Safrudin, M. Martanto, and U. Hayati, "Perbandingan Kinerja Naïve Bayes Dan Support Vector Machine Untuk Klasifikasi Sentimen Ulasan Game Genshin Impact," JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 3, pp. 3182–3188, 2024, doi: 10.36040/jati.v8i3.8415.

[13] R. Afandi, M. Afdal, and R. Novita, "Analisis Sentimen Masyarakat Terhadap Pinjaman Online di Twitter Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest N eighbor," vol. 6, no. 2, pp. 596–605, 2024, doi: 10.47065/bits.v6i2.5300.

[14] H. Kurniawan, A. Rahim, T. Azhima, and Y. Siswa, "Implementasi Algoritma Gaussian Naïve Bayes Dalam Klasifikasi Status Gizi Pada Balita," vol. 6, no. 2, pp. 627–635, 2024.

[15] Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes," Edumatic J. Pendidik. Inform., vol. 8, no. 1, pp. 183–192, 2024, doi: 10.29408/edumatic.v8i1.25667.

[16] Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter, Indones. J. Comput. Sci., vol. 12, Aug. 2023, doi: 10.33022/ijcs.v12i4.3341.

[17] Hidayat, "Metode Machine Learning untuk Klasifikasi Data Gizi Balita dengan Algoritma Naïve Bayes, KNN dan Decision Tree," J. SIMETRIS, vol. 15, no. 1, pp. 57– 68, 2024, [Online]. Available: https://jurnal.umk.ac.id/index.php/simet/article/view/10679.

[18] S. Rita and A. Pambudi, "Penggunaan Support Vector Machine Untuk Analisis Sentimen Ulasan Aplikasi Truecaller Dan Getcontact Using Support Vector Machine for Sentiment Analysis of Truecaller and Getcontact App Reviews," vol. 20, no. 2, pp. 131–140, 2023.

[19] A. A. Munandar, F. Farikhin, and C. E. Widodo, "Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM," JOINTECS (Journal Inf. Technol. Comput. Sci., vol. 8, no. 2, p. 77, 2023, doi: 10.31328/jointecs.v8i2.4747.

[20] Dania Siregar, Faroh Ladayya, Naufal Zhafran Albaqi, and Bintang Mahesa Wardana, "Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter," J. Stat. dan Apl., vol. 7, no. 1, pp. 93–104, 2023, doi: 10.21009/jsa.07109.

[21] Sri Diantika, Windu Gata, and Hiya Nalatissifa, "Komparasi Algoritma SVM Dan Naive Bayes Untuk Klasifikasi Kestabilan Jaringan Listrik," Elkom J. Elektron. dan Komput., vol. 14, no. 1, pp. 10–15, 2021, doi:10.51903/elkom.v14i1.319.

[22] T. A. Dewi and E. Mailoa, "Perbandingan Implementasi Metode Smote Pada Algoritma Support Vector Machine (Svm) Dalam Analisis Sentimen Opini Masyarakat Tentang Mixue," J. Indones. Manaj. Inform. dan Komun., vol. 4, no. 3, pp. 849–855, 2023, doi:10.35870/jimik.v4i3.289.

[23] H. Apriyani and K. Kurniati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," J. Inf. Technol. Ampera, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.isssue3.year20 20.page133-143.

[24] Tundo, Eldina, R., Setiawan, K., & Fajri, R. (2024). Sentiment Analysis of Cigarette Use Based on Opinions from X Using Naive Bayes and SVM. Jurnal Indonesia : Manajemen Informatika Dan Komunikasi, 5(3), 2561-2569. https://doi.org/10.35870/jimik.v5i3.947

[25] Supian, Acuan, Bagus Tri Revaldo, Nanda Marhadi, Lusiana Efrizoni and Rahmaddeni Rahmaddeni. "Perbandingan Kinerja Naïve Bayes Dan Svm Pada Analisis Sentimen Twitter Ibukota Nusantara." Jurnal Ilmiah Informatika (2024): n. pag.

[26] Syahputra, R., Yanris, G.J., & Irmayani, D. (2022). SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter. Sinkron.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[27] Agustiana, N., Pratiwi, O. N., & Fakhrurroja, H. (2023). Comparison Of Sentiment Analysis Of Traveloka And Tiket.Com Applications On Twitter Using The Naive Bayes Method. ITEJ (Information Technology EngineeringJournals), 8(2),73-83. https://doi.org/10.24235/itej.v8i2.119

[28] Fauzi, F., Setiayani, W., Utami, T.W., Yulianto, E., & Harmoko, I.W. (2023). Comparison Of Random Forest And Naïve Bayes Classifier Methods In Sentiment Analysis On Climate Change Issue. Barekeng: Jurnal Ilmu Matematika dan Terapan.

[29] Zulfikar, W.B., Atmadja, A.R., & Pratama, S.F. (2023). Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes. Scientific Journal of Informatics.

[30] Onwuegbuche, F.C., Wafula, J.M., & Mung'atu, J. (2019). Support Vector Machine for Sentiment Analysis of Nigerian Banks Financial Tweets. Journal of Data Analysis and Information Processing.

[31] Samsir, Irmayani, D., Edi, F., Harahap, J.M., Jupriaman, Rangkuti, R.K., Ulya, B., & Watrianthos, R. (2021). Naives Bayes Algorithm for Twitter Sentiment Analysis. Journal of Physics: Conference Series, 1933.

[32] Hadianti, S., & Tember, F.Y. (2022). Analisis Sentiment Covid-19 Di Twitter Menggunakan Metode Naive Bayes Dan Svm.