ISSN: 2477-0078

Komparasi Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Media Sosial terhadap Wacana Mata Uang BRICS

Nailil Falahah[1] , Muhsi[2], Miftahul Walid[3]
[1,2]Sistem Informasi, Teknik, Universitas islam Madura
[3]Teknik Informatika, Teknik, Universitas islam Madura
[1]naililfalahah6@gmail.com, [2]

ABSTRAK

Kemunculan mata uang *BRICS* memunculkan beragam persepsi publik yang terekam di media sosial, namun belum banyak dikaji secara sistematis. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap wacana mata uang *BRICS* dengan memanfaatkan data dari tiga platform, yaitu X, YouTube, dan TikTok. Proses klasifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* setelah melalui tahapan prapemrosesan teks serta pembobotan *TF-IDF*. Hasil penelitian menunjukkan bahwa SVM menghasilkan akurasi tertinggi, yakni 65,5% pada X, 73,6% pada YouTube, dan 63,8% pada TikTok. Sementara itu, *Naïve Bayes* memberikan akurasi yang lebih rendah tetapi relatif stabil di ketiga platform. Dengan demikian, SVM dinilai lebih efektif dalam mengklasifikasikan sentimen media sosial terkait mata uang *BRICS*. Temuan ini memberikan kontribusi pada pemetaan opini publik sekaligus menjadi rujukan dalam pemilihan metode analisis sentimen berbasis media sosial.

Kata kunci: *BRICS*, analisis sentimen, *support vector machine*, *naïve bayes*, media sosial

ABSTRACT

The emergence of the BRICS currency has generated diverse public perceptions recorded on social media, yet it has not been systematically studied. This research aims to analyze public sentiment towards the BRICS currency by utilizing data from three platforms: X, YouTube, and TikTok. The classification process was carried out using the Support Vector Machine (SVM) and Naïve Bayes algorithms after text preprocessing and TF-IDF weighting. The results show that SVM achieved the highest accuracy, namely 65.5% on X, 73.6% on YouTube, and 63.8% on TikTok. Meanwhile, Naïve Bayes produced lower accuracy but remained relatively stable across the three platforms. Therefore, SVM is considered more effective in classifying social media sentiment regarding the BRICS currency. These findings contribute to the mapping of public opinion and serve as a reference in selecting sentiment analysis methods for social media data.

Keywords: *BRICS, sentiment analysis, Support Vector Machine, Naïve Bayes, social media*

ISSN: 2477-0078

1. PENDAHULUAN

Aliansi ekonomi BRICS yang terdiri dari Brasil, Rusia, India, Tiongkok, dan Afrika Selatan semakin memperkuat pengaruhnya dalam sistem keuangan global. Salah satu inisiatif penting adalah rencana pembentukan mata uang bersama sebagai alternatif dominasi dolar Amerika Serikat [1],[2]. Isu ini tidak hanya berdampak pada perekonomian makro, tetapi juga menciptakan dinamika sosial yang terekam melalui interaksi publik di media sosial [3]. Bagi Indonesia, keterlibatan dalam BRICS menimbulkan implikasi strategis yang berkaitan dengan stabilitas ekonomi nasional dan posisi dalam sistem moneter internasional [4].

Perkembangan media sosial memungkinkan masyarakat tidak hanya menjadi penerima informasi, tetapi juga aktor aktif dalam membentuk narasi global secara real-time [5]. Kondisi ini menjadikan analisis sentimen publik berbasis media sosial relevan untuk memahami persepsi digital terhadap isu strategis seperti dedolarisasi.

Sejumlah penelitian terdahulu memperlihatkan bahwa algoritma Support Vector Machine (SVM) memiliki akurasi lebih tinggi dibandingkan Naïve Bayes dalam mengklasifikasikan opini publik, baik pada isu politik maupun ulasan aplikasi digital [6],[7]. Temuan tersebut menunjukkan keunggulan SVM dalam menangani data teks berdimensi tinggi. Di sisi lain, penelitian lain menegaskan bahwa Naïve Bayes tetap kompetitif dalam analisis teks berbahasa Indonesia karena kesederhanaan dan stabilitas performanya [8][9]. Dengan demikian, kedua algoritma memiliki kelebihan masing-masing, sehingga perbandingan kinerjanya penting dilakukan.

Kendati demikian, sebagian besar kajian terdahulu masih terbatas pada satu platform atau topik spesifik, seperti ulasan aplikasi atau isu politik [10][11]. Belum banyak penelitian yang secara komprehensif memetakan sentimen publik terhadap wacana mata uang BRICS melalui pendekatan multiplatform. Padahal, perbedaan karakteristik bahasa dan gaya komunikasi di X, YouTube, dan TikTok berpotensi menghasilkan distribusi sentimen yang berbeda.

Berdasarkan celah penelitian tersebut, studi ini difokuskan untuk menganalisis sentimen publik terhadap wacana mata uang BRICS dengan memanfaatkan data dari tiga platform media sosial, yaitu X, YouTube, dan TikTok. Perbandingan kinerja algoritma SVM dan Naïve Bayes dilakukan guna memperoleh gambaran komprehensif mengenai persepsi publik digital, sekaligus memberikan kontribusi empiris dalam pengembangan metode analisis sentimen media sosial.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi berbasis *machine learning* untuk menganalisis sentimen masyarakat terhadap wacana mata uang BRICS. Tahapan penelitian disusun secara sistematis mulai dari pengumpulan data hingga evaluasi model, sebagaimana dijelaskan berikut.

2.1 Sumber dan Pengumpulan Data

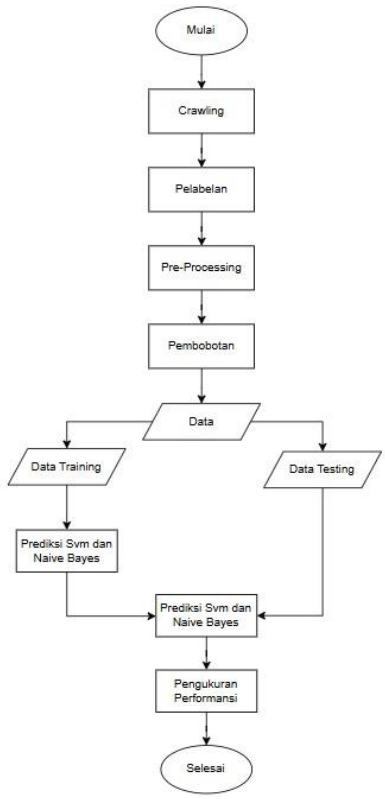
Data yang digunakan berasal dari tiga platform media sosial yang memiliki karakteristik pengguna berbeda, yaitu X (Twitter), YouTube, dan TikTok. Pengumpulan data dilakukan secara otomatis dengan memanfaatkan *tools* dan API yang sesuai, misalnya Tweet Harvest dan *auth token* untuk X, YouTube Data API untuk komentar YouTube, serta Apify untuk komentar TikTok. Komentar yang dikumpulkan difokuskan pada kata kunci yang relevan dengan topik “mata uang BRICS” dan isu dedolarisasi. Periode pengambilan data berlangsung pada Januari–April 2025 dengan total 7.196 komentar. Seluruh data kemudian disimpan dalam format CSV agar mudah diproses lebih lanjut.

2.2 Pra-pemrosesan Data

Pra-pemrosesan data merupakan tahap esensial dalam analisis sentimen berbasis teks, khususnya ketika berhadapan dengan data tidak

ISSN: 2477-0078

terstruktur seperti komentar media sosial. Tujuan utama dari tahap ini adalah menyaring informasi penting dan menghilangkan unsur-unsur yang tidak relevan sehingga data siap dianalisis secara otomatis oleh algoritma machine learning. Tahapan-tahapan pra-pemrosesan penelitian ditunjukkan pada Gambar 1 berikut ini:



Gambar 1. Flowchart Pra-Pemrosesan Data

Komentar yang diperoleh dari media sosial pada umumnya bersifat tidak terstruktur dan mengandung elemen yang tidak relevan. Oleh karena itu, dilakukan tahap pra-pemrosesan data dengan beberapa langkah penting, yaitu pembersihan teks untuk menghapus URL, angka, tanda baca, emotikon, dan karakter khusus. *Case folding* dengan mengubah seluruh teks menjadi huruf kecil. Tokenisasi untuk memecah kalimat menjadi unit kata. Penghapusan *stopword* menggunakan daftar kata umum bahasa Indonesia. Serta *stemming* dengan pustaka Sastrawi agar kata dikembalikan ke bentuk dasarnya. Setelah melalui tahapan ini, data diberi label sentimen secara manual ke dalam tiga kategori, yaitu positif, negatif, dan netral. Pelabelan manual dipilih untuk menjaga ketepatan interpretasi konteks, mengingat komentar media sosial kerap mengandung bahasa informal, sarkasme, maupun ironi.

Menurut [12] serta [13], tahapan pembersihan data teks dan representasi fitur seperti TF-IDF sangat krusial untuk meningkatkan kualitas masukan model machine learning. Oleh karena itu, langkah pra-pemrosesan dalam penelitian ini dirancang sejalan dengan standar yang telah banyak diterapkan pada penelitian serupa.

2.3 Pembobotan

Teks yang telah diproses selanjutnya direpresentasikan dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF).* Metode ini menilai pentingnya suatu kata dalam dokumen relatif terhadap seluruh korpus dengan rumus:

$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t) \quad (1)$$

Dengan:

$$TF(t,d) = \frac{f_{t,d}}{\sum_k f_{k,d}}, \quad IDF(t) = \log \frac{N}{df_t} \quad (2)$$

Yang merepresentasikan frekuensi kata *t* dalam dokumen *d*. Dan mengukur seberapa jarang kata *t* muncul dalam seluruh korpus yang terdiri dari *N* dokumen.

Proses konversi teks dilakukan menggunakan fungsi TfidfVectorizer dari Pustaka scikit learn, sehingga menghasilkan vektor fitur numerik yang siap untuk digunakan dalam proses klasifikasi.

2.4 Pemodelan dan Klasifikasi

Dalam analisis sentimen tersedia berbagai algoritma dengan keunggulan dan keterbatasan masing-masing. Decision Tree mudah

ISSN: 2477-0078

dipahami namun rentan overfitting, sedangkan Random Forest mengatasinya dengan menggabungkan banyak pohon untuk hasil yang lebih stabil. K-Nearest Neighbor (KNN) mengklasifikasikan data berdasarkan kedekatan jarak, sederhana tetapi kurang efisien pada dataset besar [12].

Dalam penelitian ini dipilih SVM dan Naïve Bayes. SVM unggul dalam klasifikasi teks berdimensi tinggi dan terbukti efektif pada data opini public [6]. NB dipilih karena sederhana, cepat, dan tetap stabil untuk teks bahasa Indonesia [10].

Proses klasifikasi dilakukan dengan membandingkan dua algoritma, yaitu *Support Vector Machine* (SVM) dan *Multinomial Naïve Bayes*. SVM membangun *hyperplane* optimal untuk memisahkan kelas dengan fungsi keputusan:

$$f(x) = \text{sign}(w \cdot x + b) \tag{3}$$

Di mana:

- x: vektor fitur hasil TF-IDF
- w: bobot
- b: bias
- fungsi sign menentukan label kelas dari hasil klasifikasi.

Tujuan dari SVM adalah mencari *hyperlane* dengan margin maksimum, yang secara matematis diformulasikan pada Persamaan (4)

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{dengan syarat } y_i(w \cdot x_i + b) \geq 1 \tag{4}$$

Pada penelitian ini, digunakan variasi kernel linear dan rbf serta parameter C dan gamma yang disesuaikan menggunakan *GridSearchCV* untuk memperoleh performa optimal. Kelebihan utama SVM terletak pada kemampuannya menangani data berdimensi tinggi, termasuk data teks hasil transformasi TF-IDF.

Naïve Bayes adalah metode probabilistik yang mengasumsikan independensi antar fitur[14]. Algoritma ini sederhana namun efisien, sehingga mampu bekerja baik pada data teks berukuran besar. Rumus utamanya ditunjukkan pada Persamaan (5).

$$P(C \mid x) = \frac{P(x \mid C) \cdot P(C)}{P(x)} \tag{5}$$

Dalam implementasi, nilai P(x) diabaikan karena konstan pada semua kelas. Prediksi kemudian dilakukan dengan mencari kelas C yang memaksimalkan Persamaan (6).

$$\hat{C} = \arg\max_{C} P(C) \prod_{j=1}^{n} P(x_j \mid C) \tag{6}$$

Dalam penelitian ini, digunakan implementasi *Multinomial Naïve Bayes* dari pustaka scikit-learn.

Dataset dibagi menjadi data latih (80%) dan data uji (20%) untuk menguji kinerja model. Pada SVM, dilakukan eksplorasi parameter (kernel, C, dan gamma) dengan prosedur pencarian terkontrol, sedangkan Naïve Bayes digunakan dengan konfigurasi standar yang sesuai untuk teks.

2.5 Evaluasi Model

Kinerja kedua algoritma diukur dengan empat metrik utama, yaitu akurasi, presisi, recall, dan F1-score. Evaluasi dilakukan dengan bantuan pustaka scikit-learn, dan hasilnya disajikan dalam bentuk tabel maupun grafik untuk memudahkan perbandingan antar algoritma dan antar platform. Secara matematis, metrik evaluasi tersebut didefinisikan sebagai berikut:

Akurasi: Persentase prediksi yang benar dibandingkan dengan total data.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

ISSN: 2477-0078

Presisi (*Precision*): Rasio antara jumlah prediksi positif yang benar terhadap seluruh prediksi positif.

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (6)$$

Recall: Rasio antara jumlah prediksi positif yang benar terhadap seluruh data aktual positif.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

F1-Score: Rata-rata harmonis antara presisi dan recall.

$$F1 = 2 \cdot \frac{\text{Presisi} \cdot \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (8)$$

2.6 Perangkat dan Lingkungan Kerja

Seluruh proses penelitian diimplementasikan menggunakan bahasa pemrograman Python di Google Colaboratory. Beberapa pustaka yang digunakan antara lain scikit-learn untuk klasifikasi, NLTK untuk pemrosesan bahasa alami, serta Sastrawi untuk *stemming*. Untuk pengumpulan data digunakan Apify, YouTube Data API, dan Tweet Harvest. Sementara itu, Microsoft Excel berperan sebagai alat bantu untuk pengecekan data dan anotasi manual.

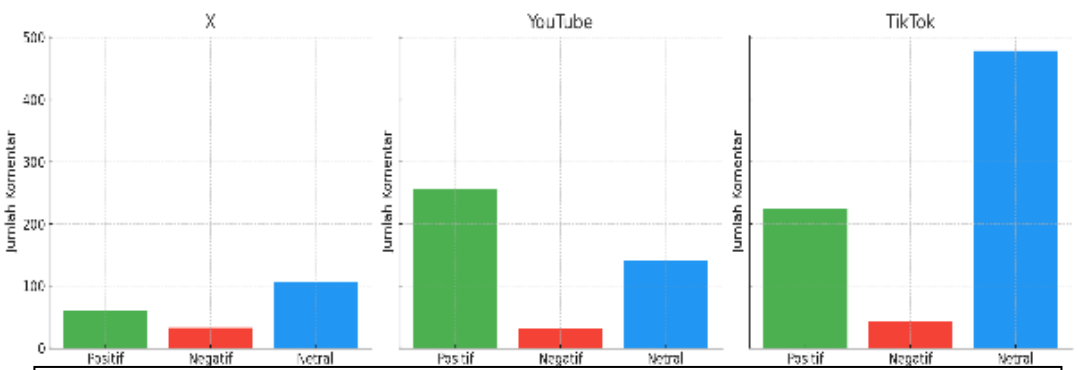
Dengan alur metodologi ini, penelitian diharapkan mampu menghasilkan analisis sentimen yang akurat dan dapat dipertanggungjawabkan secara ilmiah, sekaligus memberikan gambaran komprehensif mengenai persepsi publik terhadap wacana mata uang BRICS di media sosial.

3. HASIL DAN PEMBAHASAN

Penelitian ini berhasil mengumpulkan total 7.196 komentar dari tiga platform media sosial, yaitu X (1.000 data), TikTok (4.055 data), dan YouTube (2.141 data). Jumlah ini menunjukkan bahwa isu BRICS banyak menarik perhatian publik di dunia digital, meskipun dengan karakteristik interaksi yang berbeda di tiap platform.

Setelah melalui tahap pra-pemrosesan dan representasi fitur dengan TF-IDF, data dianalisis menggunakan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes*. Distribusi sentimen yang diperoleh memperlihatkan adanya perbedaan pola di setiap platform. Pada X, sentimen negatif cenderung lebih dominan, sedangkan YouTube menunjukkan dominasi sentimen positif. Sementara itu, TikTok didominasi sentimen netral, yang sebagian besar diekspresikan dalam bentuk konten santai atau humor. Hasil distribusi sentimen tersebut dapat dilihat pada Gambar 2.

Gambar 1. Distribusi Sentimen pada X, YouTube, dan TikTok



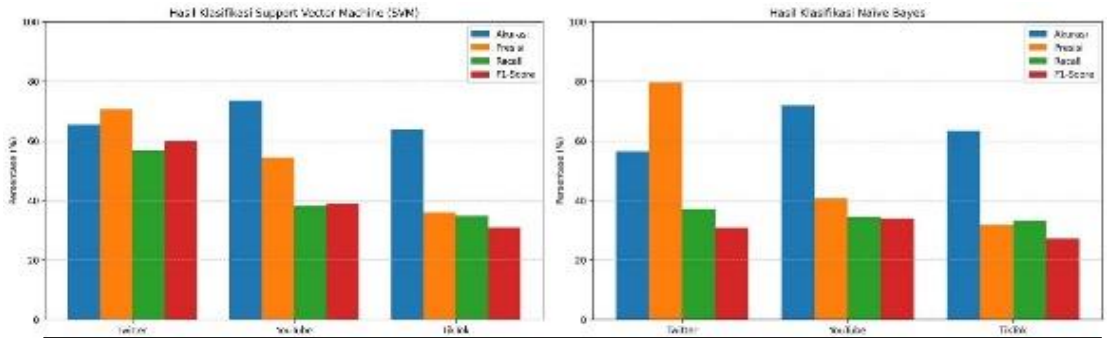
Gambar 2. Distribusi Sentimen pada X, YouTube, dan TikTok

Temuan ini menekankan potensi dampak mata uang BRICS terhadap perekonomian Indonesia, sehingga sentimen publik yang beragam mencerminkan kekhawatiran sekaligus harapan terhadap inisiatif tersebut. Selain itu, pola

ISSN: 2477-0078

dominasi sentimen positif di YouTube memperkuat hasil [5] yang mengaitkan BRICS dengan peluang kerja sama strategis di tingkat global.

Dari sisi performa algoritma, SVM menunjukkan hasil yang lebih baik dibandingkan Naïve Bayes. Pada platform YouTube, akurasi SVM mencapai 73,6%, sementara pada X sebesar 65,5%, dan TikTok 63,8%. Sebaliknya, Naïve Bayes memberikan hasil yang relatif stabil tetapi kurang mampu mendeteksi sentimen positif dan negatif secara akurat, khususnya pada TikTok. Perbandingan kinerja algoritma berdasarkan metrik akurasi, presisi, recall, dan F1-score ditunjukkan pada Gambar 3 dan Tabel 1.



Gambar 3. Perbandingan Kinerja Algoritma SVM dan Naïve Bayes pada Setiap Platform

Platform	*SVM (%)*	*Naïve Bayes (%)*	*Presisi/ Recall (%)*	*F1-Score (%)*
X	*65.5*	*56.5*	70.6 / 56.7	*60.0*
Youtube	*73.6*	*71.9*	*54.4 / 38.1*	38.9
TikTok	*63.8*	*63.4*	*36.0 / 34.8*	30.9

Tabel 1. Perbandingan Kinerja Algoritma SVM dan Naïve Bayes pada Setiap Platform

Hasil ini menunjukkan bahwa SVM lebih efektif dalam menangani data berdimensi tinggi serta variasi bahasa yang kompleks. Namun demikian, perbedaan performa antar platform mengindikasikan bahwa karakteristik bahasa dan gaya komunikasi pengguna sangat berpengaruh terhadap keberhasilan klasifikasi. Analisis multiplatform ini sekaligus memberikan gambaran yang lebih komprehensif dibandingkan penelitian yang hanya berfokus pada satu media sosial.

4. KESIMPULAN

Penelitian ini menganalisis sentimen masyarakat terhadap wacana mata uang BRICS menggunakan data dari X, YouTube, dan TikTok. Hasil penelitian menunjukkan distribusi sentimen yang berbeda, yaitu dominasi negatif pada X, positif pada YouTube, dan netral pada TikTok. Dari sisi performa, algoritma SVM memberikan hasil terbaik dengan akurasi tertinggi 73,6% pada YouTube, 65,5% pada X, dan 63,8% pada TikTok, sedangkan Naïve Bayes cenderung stabil namun kurang efektif pada ekspresi sentimen yang kompleks. Temuan ini berkontribusi pada pengembangan analisis sentimen multiplatform serta menjadi acuan dalam pemilihan algoritma klasifikasi. Untuk penelitian selanjutnya, disarankan memperluas data ke platform lain dan menerapkan model deep learning guna meningkatkan akurasi klasifikasi secara kontekstual.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan, bimbingan, dan bantuan dalam penyusunan artikel ini. Ucapan khusus ditujukan kepada dosen pembimbing atas arahan yang berharga, serta keluarga dan teman-teman atas dukungan moral yang tak ternilai.

ISSN: 2477-0078

DAFTAR PUSTAKA

[1] F. N. Idris, A. M. Dzaky, R. Haq, dan S. Hafsari, "Hegemoni dolar dan potensi kemunculan mata uang brics," *Emerald J. Econ. Soc. Sci. J. Econ. Soc. Sci.*, vol. 1, no. 1, hal. 19–30, 2022.

[2] U. Airlangga dan I. Monetary, "Analisis Potensi Mata Uang BRICS dalam Menantang Dominasi Dolar AS dalam Perspektif Ekonomi Moneter Islam : Keunggulan dan Tantangan," 2022.

[3] A. P. Bate, "Hashtag Activism and Football Tragedy Commemoration: Social Network Analysis In the #100harikanjuruhan Hashtag on Twitter," *J. Komun. Indones.*, vol. 13, no. 1, 2024, doi: 10.7454/jkmi.v13i1.1226.

[4] V. A. S. Braviano dan G. Heriantomo, "Prospek Mata Uang BRICS dan Korelasinya Terhadap Potensi Keamanan Ekonomi di Indonesia," *J. Stud. Islam dan Hum.*, vol. 4, no. 2, hal. 304–321, 2024.

[5] J. Mahasiswa *et al.*, "DAMPAK KERJASAMA BRICS TERHADAP DOMINASI NILAI DOLAR AMERIKA SERIKAT (AS) DI TAHUN 2022," vol. 1, no. 1, hal. 297–314, 2024, doi: 10.36859/dgsj.v1i1.2857.

[6] F. H. Hilmi dan A. D. Indriyanti, "Sentiment Analysis of 2024 Election Fraud Using SVM and Naïve Bayes Algorithms," vol. 5, no. 4, hal. 353–364, 2025.

[7] rahayu deny danar dan alvi furwanti Alwie, A. B. Prasetio, R. Andespa, P. N. Lhokseumawe, dan K. Pengantar, "Tugas Akhir Tugas Akhir," *J. Ekon. Vol. 18, Nomor 1 Maret201*, vol. 2, no. 1, hal. 41–49, 2020.

[8] T. T. Widowati dan M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, hal. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.

[9] S. Kusuma Wardani dan Y. Arum Sari, "Analisis Sentimen menggunakan Metode Naïve Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan Document Frequency Thresholding," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, hal. 5582–5590, 2021, [Daring]. Tersedia pada: http://j-ptiik.ub.ac.id

[10] Y. P. Astuti, A. R. Wibowo, E. Kartikadarma, E. R. Subhiyakto, N. A. Sri Winarsih, dan M. S. Rohman, "Penerapan Metode Naïve Bayes Classifier Untuk Klasifikasi Sentimen Pada Judul Berita," *LogicLink*, vol. 1, no. 1, hal. 1–12, 2024, doi: 10.28918/logiclink.v1i1.7684.

[11] M. M. Danyal, S. S. Khan, M. Khan, M. B. Ghaffar, B. Khan, dan M. Arshad, "Sentiment Analysis Based on Performance of Linear Support Vector Machine and Multinomial Naïve Bayes Using Movie Reviews with Baseline Techniques," *J. Big Data*, vol. 5, no. September, hal. 1–18, 2023, doi: 10.32604/jbd.2023.041319.

[12] D. Papakyriakou dan I. S. Barbounakis, "Data Mining Methods: A Review," *Int. J. Comput. Appl.*, vol. 183, no. 48, hal. 5–19, 2022, doi: 10.5120/ijca2022921884.

[13] N. S. -, R. B. -, dan R. P. -, "A Review on Data Mining Issues, Solution & Techniques," *Int. J. Multidiscip. Res.*, vol. 6, no. 4, hal. 1–9, 2024, doi: 10.36948/ijfmr.2024.v06i04.26654.

[14] N. Z. B. Jannah dan K. Kusnawi, "Comparison of Naïve Bayes and SVM in Sentiment Analysis of Product Reviews on Marketplaces," *Sinkron*, vol. 8, no. 2, hal. 727–733, 2024, doi: 10.33395/sinkron.v8i2.13559.

ISSN: 2477-0078