

SENTIMEN ULASAN KLASIFIKASI GAME BUS SIMULATOR INDONESIA DENGAN METODE TF-IDF NAÏVE BAYES

Daeng Indar Kurniawan, B. Satria Bakti, Ulya Anisatur Rosyidah, Miftahur Rahman, Qurrota A'Yun

Teknik Informatika, Universitas Muhammadiyah Jember

Jl. Karimata No. 49, Sumbersari, Kabupaten Jember, Jawa Timur, Indonesia.

Daengindar123@gmail.com

ABSTRAK

Industri hiburan digital lokal, khususnya game mobile seperti Bus Simulator Indonesia (BUSSID), berkembang pesat dan menarik banyak pengguna. Namun, tingginya volume ulasan pengguna di Google Play Store menyulitkan pengembang dalam mengevaluasi umpan balik secara manual untuk meningkatkan kualitas game. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pengguna BUSSID ke dalam kategori positif, negatif, dan netral. Metode yang digunakan adalah data preprocessing (cleaning, case folding, normalisasi, tokenizing, stemming), pelabelan awal dengan Indonesia Sentiment Lexicon, pembobotan Term Frequency-Inverse Document Frequency (TF-IDF), dan klasifikasi menggunakan algoritma Naïve Bayes. Hasil penelitian menunjukkan kombinasi TF-IDF dan Naïve Bayes sangat efektif. Pada skenario data terbaik (70:30), model mencapai akurasi sebesar 89,92%. Performa kelas positif memperoleh precision 82%, recall 98%, F1-score 89%; kelas negatif mencapai precision 98%, recall 90%, F1-score 94%; serta kelas netral mencatat precision 95%, recall 73%, F1-score 83%.

Kata kunci : *sentiment analysis, naïve bayes, tf-idf, bus simulator indonesia, web scraping.*

1. PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi yang berlangsung sangat cepat selama dua dekade terakhir telah membawa perubahan signifikan pada berbagai bidang kehidupan, termasuk sektor hiburan digital [1].

Di antara berbagai bentuk hiburan digital, industri game menjadi salah satu sektor yang menunjukkan pertumbuhan paling menonjol, terutama pada kategori game yang dimainkan melalui perangkat 6 [2].

Salah satu contoh keberhasilan dalam industri tersebut adalah Bus Simulator Indonesia (BUSSID), sebuah game karya Maleo yang berhasil menembus pasar internasional dan memperoleh lebih dari seratus juta unduhan melalui Google Play Store [3].

Sebagai produk lokal unggulan, eksistensi BUSSID menghasilkan volume umpan balik yang sangat besar dari penggunaannya setiap hari.

Namun, besarnya volume ulasan pengguna yang masuk dalam bentuk teks tidak terstruktur menjadi tantangan serius bagi pihak pengembang. Proses evaluasi kualitas layanan, identifikasi kendala teknis, dan analisis kepuasan pemain secara manual menjadi sangat tidak efisien dan memerlukan waktu yang lama. Akibatnya, banyak informasi berharga atau keluhan krusial dari pengguna yang terlewatkan. Kebutuhan mendesak akan sebuah sistem otomatis yang dapat mengekstrak informasi dari jutaan ulasan ini menjadi permasalahan utama yang harus diselesaikan agar pengembang dapat merespons preferensi pengguna secara cepat dan tepat.

Untuk mengatasi permasalahan tersebut, rencana penyelesaian masalah yang diajukan dalam penelitian ini adalah membangun sebuah sistem klasifikasi sentimen otomatis berbasis komputasi. Berdasarkan

penelitian terdahulu, analisis sentimen terbukti efektif dalam mengevaluasi ulasan aplikasi secara otomatis. Penggunaan algoritma Naïve Bayes dipilih karena kesederhanaannya serta efisiensi komputasinya yang tinggi dalam menangani dataset teks berukuran besar [4].

Guna mengoptimalkan performa klasifikasi, algoritma tersebut dikombinasikan dengan teknik ekstraksi fitur Term Frequency-Inverse Document Frequency (TF-IDF) untuk memberikan bobot pada kata-kata penting dalam korpus [5].

Meskipun penelitian sentimen pada aplikasi umum sudah banyak dilakukan, studi mendalam yang membagi ulasan game lokal seperti BUSSID ke dalam tiga kategori (positif, negatif, dan netral) masih memerlukan eksplorasi lebih lanjut.

Berdasarkan situasi tersebut, penelitian ini bertujuan untuk menerapkan metode TF-IDF dan algoritma Naïve Bayes dalam mengidentifikasi serta mengelompokkan sentimen yang terkandung dalam ulasan pengguna BUSSID di Google Play Store secara akurat. Melalui pencapaian tujuan ini, diharapkan hasil klasifikasi dapat memberikan rekomendasi dan dasar pengambilan keputusan yang objektif bagi pihak pengembang dalam melakukan perbaikan aplikasi serta pengembangan fitur di masa mendatang.

2. TINJAUAN PUSTAKA

2.1. Studi Literatur

Sebagai landasan dalam penelitian ini, berbagai sumber literatur ilmiah ditelaah untuk memperoleh pemahaman yang komprehensif mengenai analisis sentimen, proses klasifikasi pada teks ulasan pengguna, serta implementasi algoritma Naïve Bayes dalam pengolahan data tekstual. Literatur tersebut mencakup penelitian mengenai penggunaan ekstraksi

fitur TF-IDF untuk meningkatkan akurasi klasifikasi ulasan pada aplikasi marketplace. Selain itu, kajian juga diarahkan pada analisis sentimen pengguna aplikasi dengan memanfaatkan rating dan komentar di Google Play Store untuk memahami pola kepuasan konsumen. Pelaksanaan studi literatur dalam penelitian ini diarahkan untuk membangun dan memperkuat dasar teoritis yang mendukung proses analisis sentimen otomatis terhadap ulasan pengguna game simulator Indonesia, sekaligus memberikan pemahaman mendalam mengenai kinerja algoritma klasifikasi dalam menangani data teks tidak terstruktur.

2.2. Pengumpulan Data

Data penelitian dikumpulkan melalui proses ekstraksi data terhadap ulasan aplikasi Bus Simulator Indonesia di Google Play Store. *Dataset* yang didapat berupa teks ulasan tidak terstruktur yang mencerminkan opini pengguna dalam berbagai bahasa dan format penulisan.

2.3. Preprocessing Data

Tahap ini memiliki tujuan untuk membersihkan data agar lebih optimal saat diproses oleh algoritma. Proses ini meliputi *cleaning* untuk menghapus karakter non-alfabet, *case folding* untuk menyeragamkan seluruh karakter menjadi huruf kecil, normalisasi guna mengonversi kata-kata tidak baku ke bentuk yang sesuai, *tokenizing* untuk memisahkan teks menjadi unit kata, *stemming* untuk mengembalikan kata ke bentuk dasarnya, serta *labeling* yang bertujuan mengklasifikasikan data ke dalam kategori sentimen yang telah ditetapkan. Setelah seluruh tahapan tersebut selesai dilakukan, data kemudian diproses lebih lanjut pada tahap ekstraksi fitur menggunakan metode TF-IDF [6].

$$TF(t, d) = \frac{\text{jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{jumlah total kata dalam dokumen } d} \quad (1)$$

Selanjutnya, dilakukan perhitungan *inverse document frequency* (IDF) untuk menentukan tingkat kepentingan kata berdasarkan jumlah dokumen yang mengandung kata tersebut dalam seluruh koleksi dokumen (N). Untuk menghindari pembagian dengan nol pada kata yang tidak muncul dalam kelas tertentu, diterapkan teknik *smoothing* sesuai dengan rumus (2).

$$IDF_{sm}(t) = \log \left[1 + \frac{D}{1 + DF(t)} \right] \quad (2)$$

Hasil akhir pembobotan diperoleh dengan mengalikan nilai TF dengan IDF menggunakan rumus (3).

$$W(t, d) = TF(t, d) \times \log \left(\frac{D}{DF(t)} \right) \quad (3)$$

Pada rumus tersebut, W merepresentasikan bobot akhir atau skor TF-IDF dari kata t pada dokumen d di dalam korpus D. Sementara itu, df(t) menunjukkan jumlah dokumen dalam koleksi yang mengandung kata t. Kata dengan nilai W tinggi dianggap sebagai

fitur yang paling kuat dalam merepresentasikan karakteristik ulasan game BUSSID.

2.4. Naïve Bayes

Proses klasifikasi pada penelitian ini menggunakan algoritma Multinomial Naïve Bayes, yaitu metode yang memanfaatkan prinsip Teorema Bayes untuk menentukan kategori data. Dalam penerapannya, algoritma ini mengasumsikan bahwa setiap fitur yang digunakan dalam proses klasifikasi tidak saling bergantung atau bersifat independen satu sama lain. Probabilitas kelas dihitung berdasarkan kemunculan kata pada data latih yang telah diberi bobot oleh TF-IDF [8].

Secara umum, rumus dasar teorema Bayes dapat dilihat pada rumus (4).

$$P(C | X) = \frac{P(X | C) \cdot P(C)}{P(X)} \quad (4)$$

Pada persamaan tersebut, P(C|X) menunjukkan probabilitas posterior yang menggambarkan kemungkinan data X berada pada kelas C setelah mempertimbangkan informasi yang tersedia. Sementara itu, P(X|C) merepresentasikan *likelihood*, yaitu peluang kemunculan data X apabila diketahui berasal dari kelas C. Nilai P(C) merupakan probabilitas awal (*prior*) suatu kelas sebelum dilakukan pengamatan terhadap data, sedangkan P(X) menyatakan probabilitas keseluruhan dari data yang umumnya tidak diperhitungkan dalam proses penentuan kelas karena nilainya sama untuk seluruh kelas yang dibandingkan. Penentuan nilai *prior* dilakukan dengan membagi banyak data pada kelas C dengan total seluruh data (N) sesuai rumus (5).

$$P(C) = \frac{n(C)}{N} \quad (5)$$

Untuk menghitung *likelihood*, digunakan teknik *Laplace Smoothing* untuk menghindari nilai probabilitas nol yang dapat merusak hasil perhitungan. Penilaian kata tidak hanya didasarkan pada kepentingan dokumen secara umum, tetapi juga relevansinya dalam konteks kelas tertentu melalui integrasi nilai TF-IDF sebagai *likelihood* yang dinyatakan dalam rumus (6).

$$P(t|C) = \frac{TF-IDF(t | C) + 1}{|V| + \sum TF-IDF(C)} \quad (6)$$

Pada rumus tersebut, P(t|C) adalah probabilitas kata t muncul di kelas C, sedangkan V merupakan jumlah kosa kata unik dalam seluruh dokumen. Dalam pendekatan ini, probabilitas kelas diperkirakan dengan menggunakan informasi dari variabel input yang relevan guna menentukan kategori sentimen yang paling mungkin. Apabila nilai *posterior* yang dihasilkan terlalu kecil, dilakukan transformasi menggunakan operasi logaritma untuk mempermudah perbandingan skor antar kelas sesuai rumus (7).

$$\log \{P(C) | X_1, X_2, X_3, \dots, X_n\} \propto \log P(C) + \sum_{i=1}^n \log P(X_i | C) \quad (7)$$

2.5. Confusion Matrix

Evaluasi performa model klasifikasi dilakukan menggunakan *confusion matrix*, yaitu metode yang membandingkan hasil prediksi model dengan data aktual untuk melihat tingkat ketepatan klasifikasi. Confusion matrix terdiri atas empat elemen utama, yaitu True Positive (TP) yang menunjukkan data positif yang berhasil diprediksi dengan benar, True Negative (TN) yang merepresentasikan data negatif yang terklasifikasi secara tepat, False Positive (FP) yang menggambarkan data negatif namun diprediksi sebagai positif, serta False Negative (FN) yang menunjukkan data positif tetapi diklasifikasikan sebagai negatif. Berdasarkan keempat komponen tersebut, berbagai ukuran evaluasi dapat dihitung untuk memberikan gambaran menyeluruh mengenai kinerja model yang dihasilkan.

Metrik pertama adalah *accuracy* yang dimanfaatkan untuk mengukur seberapa sesuai model memprediksi data secara keseluruhan, baik positif, negatif maupun netral, terhadap total data yang diprediksi. Perhitungan nilai *accuracy* dilakukan sesuai dengan rumus (8).

$$Accuracy = \frac{\text{Banyak data prediksi benar}}{\text{Total seluruh data}} \quad (8)$$

Selanjutnya, *Precision* digunakan untuk mengukur tingkat ketepatan model dalam mengidentifikasi kelas positif, di mana nilai yang tinggi menunjukkan bahwasannya prediksi tersebut memiliki probabilitas kebenaran yang besar sesuai rumus (9).

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

Recall digunakan untuk menilai sejauh mana model mampu mengenali seluruh data yang benar-benar termasuk ke dalam kelas positif. Nilai ini menggambarkan persentase data positif aktual yang berhasil diidentifikasi secara tepat oleh model, sebagaimana dinyatakan pada rumus berikut (10).

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

Sebagai metrik keseimbangan, *f1-score* menggabungkan *Precision* serta *recall* ke dalam satu nilai untuk mengukur stabilitas performa model secara rata-rata. Nilai *f1-score* dihitung berdasarkan rumus (11).

$$F1\text{-Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (11)$$

3. METODE PENELITIAN

3.1. Game Bus Simulator Indonesia (BUSSID)

Bus Simulator Indonesia (BUSSID) merupakan *game* simulasi berkendara bus berbasis Android yang dikembangkan oleh Maleo. *Game* ini mengintegrasikan suasana lokal Indonesia, mulai dari arsitektur jalanan, trayek antarkota, hingga klakson khas "Om Telolet Om". Keunikan elemen lokal ini menjadi daya tarik utama yang menghasilkan volume

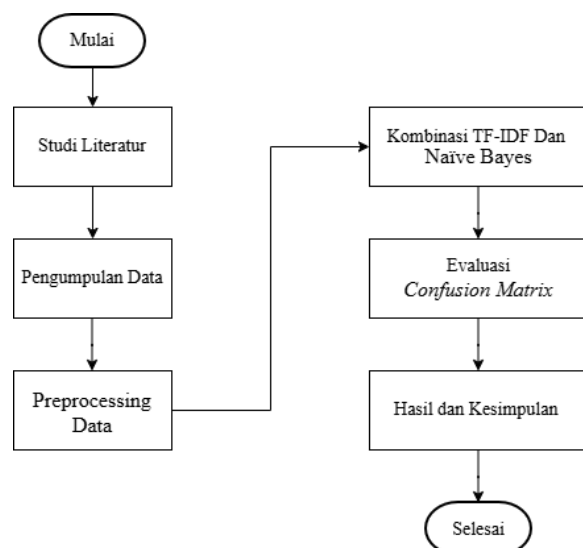
ulasan masif di Google Play Store. Antarmuka dan visualisasi dari *game* BUSSID disajikan pada Gambar 1.



Gambar 1. Game Bus Simulator Indonesia

3.2. Alur Penelitian

Metode penelitian ini mencakup beberapa tahapan sistematis untuk menyelesaikan masalah klasifikasi sentimen ulasan *game* BUSSID. Tahapan dimulai dari pengumpulan data, preprocessing data, ekstraksi fitur memanfaatkan TF-IDF, hingga proses klasifikasi dengan algoritma Naïve Bayes. Secara visual, bagan alir penelitian ini disajikan pada Gambar 1.



Gambar 2. Bagan Alir Penelitian

3.3. Pengumpulan Data

Dataset yang digunakan totalnya adalah 1.852 ulasan yang telah melalui tahap *preprocessing data*. Distribusi label sentimen terdiri dari kategori positif, negatif, dan netral. Data ini kemudian dibagi menjadi data latih dan data uji untuk mengukur keandalan model klasifikasi. Guna mendukung transparansi penelitian, keseluruhan *Dataset* ulasan yang digunakan dalam penelitian ini dapat diakses secara terbuka melalui tautan: https://drive.google.com/file/d/1pZQD_FB_byl0UTJSQUfOFqsf3gCVDyR/view.

3.4. Hasil Preprocessing Data

Tahap *preprocessing data* dilaksanakan melalui serangkaian tahapan sistematis untuk memastikan data ulasan siap digunakan dalam proses analisis sentimen.

Proses ini mencakup data *cleaning* untuk menghilangkan elemen non-teks dan karakter yang tidak relevan, *case folding* untuk menyeragamkan seluruh huruf menjadi huruf kecil, normalisasi untuk mengubah kata tidak baku menjadi kata baku, *tokenizing* untuk memecah teks menjadi unit kata, serta *stemming* untuk mengembalikan kata ke bentuk dasarnya. Setelah tahap *stemming*, proses *labeling* sentimen dilakukan menggunakan kamus *lexicon* (*lexicon-based approach*) dengan mengklasifikasikan ulasan ke dalam kategori sentimen positif, negatif, dan netral. Hasil dari seluruh tahapan *preprocessing data* ulasan dapat diakses dan diperiksa melalui tautan berikut:

<https://drive.google.com/drive/u/0/folders/17gOzyNaE531HKe8Zcai0rkRFRQfwtWmh>.

3.5. Rekapitulasi Perhitungan TF-IDF

Pembobotan TF-IDF dilakukan untuk menentukan tingkat kepentingan fitur dalam *Dataset* sebelum memasuki tahap klasifikasi. Proses ini diawali dengan penghitungan *term frequency* (TF) yang berfungsi mengukur frekuensi kemunculan suatu kata dalam sebuah dokumen. Perhitungan nilai TF pada penelitian ini dilakukan sesuai dengan persamaan rumus (1), yang sampel hasilnya disajikan pada Tabel 1.

Tabel 1. Sampel Perhitungan *Term Frequency*

IDD	<i>Term Frequency</i> (TF)		
D1	Kata pada <i>stemming</i> ulasan: (“sangat”, “bagus”)		
	TF (sangat) = 0,5	TF (bagus) = 0,5	
D2	Kata pada <i>stemming</i> ulasan: (“sangat”, “senang”)		
	TF (sangat) = 0,5	TF (senang) = 0,5	
D3	Kata pada <i>stemming</i> ulasan: (“tidak”, “bagus”)		
	TF (tidak) = 0,5	TF (bagus) = 0,5	
D4	Kata pada <i>stemming</i> ulasan: (“seru”)		
	TF (seru) = 1		
D5	Kata pada <i>stemming</i> ulasan: (“tidak”, “urus”, “lagi”, “apa”, “bagaimana”)		
	TF (sangat) = 0,5	TF (bagus) = 0,5	TF (bagus) = 0,5
	TF (sangat) = 0,5	TF (bagus) = 0,5	

Setelah nilai TF diperoleh, dilakukan perhitungan *inverse document frequency* (IDF) menggunakan teknik *smoothing* sesuai dengan persamaan rumus (2). Langkah ini bertujuan untuk menyeimbangkan bobot kata yang jarang muncul sekaligus menghindari kegagalan komputasi akibat nilai nol pada pembagi. Sampel hasil perhitungan IDF *smoothing* ditunjukkan pada Tabel 2.

Tabel 2. Sampel Perhitungan *Inverse Document Frequency Smoothing*

No	Kata	Kelas Positif	Kelas Negatif	Kelas Netral
1	aa	7,5250	6,3230	7,6373
2	aallaikumtuk	7,5250	6,3230	7,6373
3	abang	4,5806	5,2244	4,8040
4	abrek	7,5250	6,3230	7,6373
5	accesoris	7,5250	7,0162	6,9441

Hasil akhir pembobotan diperoleh dengan mengalikan nilai *term frequency* (TF) dan *inverse document frequency* (IDF) sesuai dengan persamaan (3) untuk menghasilkan skor bobot. Sampel nilai akhir TF-IDF disajikan pada Tabel 3.

Tabel 3. Sampel Perhitungan TF-IDF

No	Kata	Kelas Positif	Kelas Negatif	Kelas Netral
1	aa	.0	0.7019	0.0
2	aallaikumtuk	0.0	0.3920	0.0
3	abang	8.3779	6.5880	11.1068
4	abrek	0.0	0.2150	0.0
5	accesoris	0.0	0.0	0.4097

3.6. Hasil Evaluasi

Tahap klasifikasi dilakukan dengan menentukan peluang kemunculan setiap kata pada masing-masing kategori sentimen, sehingga dapat diketahui kecenderungan suatu data untuk masuk ke kelas sentimen tertentu. Implementasi TF-IDF membantu menonjolkan kata-kata yang memiliki bobot informasi tinggi bagi model. Berdasarkan hasil pengujian yang dilakukan, model mampu membedakan sentimen ulasan dengan cukup baik. Ringkasan performa model yang diukur melalui *confusion matrix* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Pengujian Model Klasifikasi

<i>Metrix Kelas</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Positif	83%	94%	88%
Negatif	81%	85%	83%
Netral	95%	73%	83%
Akurasi total			90%

4. HASIL DAN PEMBAHASAN

Berdasarkan hasil pengujian, nilai *Precision* tertinggi dicapai oleh kelas netral yaitu sebesar 95%.

Hal ini menunjukkan bahwasannya model memiliki akurasi prediktif yang sangat tinggi ketika mengidentifikasi ulasan bersentimen netral. Kendati demikian, nilai *recall* pada kelas netral (73%) cenderung lebih rendah dibandingkan kelas lainnya, yang mengindikasikan bahwasannya sebagian ulasan netral masih salah diklasifikasikan ke dalam kategori positif atau negatif.

Sebaliknya, kelas positif mencatat nilai *recall* yang tinggi sebesar 94%, membuktikan sensitivitas model yang sangat baik dalam mengenali ulasan bernada apresiatif. Secara keseluruhan, kombinasi metode TF-IDF dan algoritma *Naïve Bayes* terbukti efektif untuk *Dataset* ini, dengan capaian rata-rata *f1-score* yang stabil pada rentang 83% hingga 88%. Hasil pengujian ini sekaligus menjawab tantangan efisiensi terkait evaluasi ulasan pengguna secara otomatis bagi pihak pengembang Maleo.

5. KESIMPULAN DAN SARAN

Implementasi teknik pembobotan kata berdasarkan frekuensi dan keunikan kata dalam dokumen. dan algoritma *Naïve Bayes* terbukti efektif dalam mengklasifikasikan sentimen ulasan *game* Bus Simulator Indonesia (BUSSID) dengan performa yang stabil pada kategori positif, negatif, dan netral. Hasil pengujian menunjukkan tingkat *Precision* tertinggi dicapai oleh kelas netral sebesar 95%, sementara ulasan positif mendominasi dengan nilai *recall* tertinggi mencapai 94%. Penggunaan TF-IDF sebagai ekstraksi fitur mampu memberikan pembobotan kata yang akurat sehingga mengoptimalkan kinerja *Naïve Bayes* dalam mengenali karakteristik ulasan meskipun *Dataset* memiliki keragaman bahasa yang tinggi. Berdasarkan hasil tersebut, penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan data (*data balancing*) yang lebih bervariasi atau mengeksplorasi algoritma *ensemble learning* guna meningkatkan nilai *recall* pada kelas netral yang masih relatif rendah.

DAFTAR PUSTAKA

- [1] M. E. Andriansyah, U. A. Rosyidah, and L. Handayani, "Perancangan Game 'Nusan Guesser' Bertema Sejarah Berbasis Android Menggunakan Metode Game Development Life Cycle," *Explorit: Jurnal Keilmuan dan Aplikasi Teknik Informatika*, vol. 17, no. 1, 2025. doi: 10.35891/explorit.v17i1.
- [2] D. R. Azzami, B. Huda, A. Hananto, and Tukino, "Analysis of User Comment Sentiment on the Siwaslu Application Using the Naive Bayes Method," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, no. 3, pp. 2235-2240, 2025. doi: 10.59934/jaiea.v4i3.1132.
- [3] S. J. Angelina, A. B. Putra Negara, and D. P. Hastyadi, "Sentimen Analisis Aplikasi M-Paspor Menggunakan Metode Naïve Bayes," *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 1, pp. 56-62, 2022. doi: 10.30871/jaic.v6i1.3533.
- [4] P. A. Kharisma and R. Reski, "Analisis Sentimen Pengguna Aplikasi TikTok Menggunakan Metode Naïve Bayes dan K-Nearest Neighbor," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 2, pp. 345-352, 2023.
- [5] F. Nurhuda and S. W. Iriananda, "Analisis Sentimen Ulasan Pengguna Aplikasi Marketplace Menggunakan Ekstraksi Fitur TF-IDF dan Naïve Bayes Classifier," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 5, no. 1, pp. 12-20, 2024.
- [6] M. R. Syafapri, E. Haerani, I. Iskandar, and L. Afriyanti, "Klasifikasi sentimen terhadap larangan pernikahan beda agama menggunakan metode Naive Bayes Classifier," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 5, no. 1, pp. 10-18, 2024. doi: 10.37859/coscitech.v5i1.6889.
- [7] A. Q. Surbakti, R. Hayami, and J. Al-Amien, "Analisa tanggapan terhadap PSBB di Indonesia dengan algoritma decision tree pada twitter," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 2, no. 2, pp. 91-97, 2021. doi: 10.37859/coscitech.v2i2.2851.
- [8] S. Ulya, Ridwan, A. Cholid, and F. M. Hana, "Text Mining Sentimen Analisis Pengguna Aplikasi Marketplace Tokopedia Berdasar Rating dan Komentar pada Google Play Store," *Jurnal Bisnis Digital dan Sistem Informasi*, vol. 3, no. 2, 2025.
- [9] P. H. Susilo, M. G. Rohman, A. B. Laksono, and A. Bachri, "Sistem pakar penentuan kualitas jagung menggunakan metode Naive Bayes," *Insearch: Information System Research Journal*, vol. 4, no. 2, pp. 152-160, 2024.
- [10] M. K. Suryadewiansyah and T. E. Eng Tju, "Klasifikasi Naïve Bayes Untuk Analisis Alert Pada Intrusion Detection System Berdasarkan SKKNI Bidang Data Science," *Jurnal Nasional Teknologi dan Sistem Informasi (TEKNOSI)*, vol. 8, no. 2, 2022. doi: 10.25077/TEKNOSI.v8i2.2022.81-88.
- [11] D. A. Agustina, S. Subanti, and E. Zukhrinah, "Implementasi Text Mining pada Analisis Sentimen Pengguna Twitter terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine," *Indonesian Journal of Applied Statistics*, vol. 3, no. 2, 2020. doi: 10.13057/ijas.v3i2.44337.
- [12] D. Zhafira, "Analisis Sentimen Ulasan Aplikasi di Google Play Store Menggunakan Algoritma Klasifikasi," *Jurnal Teknologi Informasi*, vol. 5, no. 1, pp. 45-52, 2021.