Jurnal Informatika dan Teknologi Interaktif
Volume 2, Nomor 1, April 2025

DOI : 10.63547/jiite.v2i1.45
e-ISSN : 3048-1023

Analisis Sentimen Pengguna X Terhadap Kasus Harvey Moeis Menggunakan Algoritma Naive Bayes

Hamid Abdul Rozak[1], Yosi Sofyan Pangestu[2], Yusdi Fathudin[3] , Hanrifki Pratama[4], Rizki[5]
[1,2,3,4]Fakultas Ilmu Komputer, Jurusan Informatika, Universitas Amikom Purwokerto, Purwokerto, Indonesia
[5]Fakultas Teknik dan Sains, Jurusan Teknik Informatika, Universitas Muhammadiyah Purwokerto, Purwokerto, Indonesia
surel:[1]hamidrozak24@gmail.com,[2]yosisofyan.p@gmail.com,[3]yusdipathudin@gmail.com, [4]ahanrifkipratama@gmail.com, [5]rizkivivoy308@gmail.com

Info Artikel

Sejarah artikel:

Diterima 18-01-2025
Revisi 02-02-2025
Diterima 03-04-2025

Kata kunci:

Analisi Sentimen
Naive Bayes
Kasus Korupsi
Pengguna X

ABSTRAK

Penelitian ini menganalisis sentimen pengguna media sosial X terhadap kasus korupsi Harvey Moeis menggunakan algoritma Naive Bayes. Data dikumpulkan melalui proses scraping dan diproses melalui teknik preprocessing seperti stemming, penghapusan stopwords, dan tokenisasi. Data awal sebanyak 110 entri kemudian diseimbangkan menggunakan teknik SMOTE *(Synthetic Minority Oversampling Technique)* untuk memastikan distribusi yang adil di antara kelas sentimen positif, negatif, dan netral. Model Multinomial Naive Bayes digunakan untuk klasifikasi, dengan evaluasi berdasarkan metrik seperti akurasi, precision, recall, dan f1-score.Hasil penelitian menunjukkan akurasi model mencapai 91%, dengan performa terbaik pada sentimen negatif (-1) dan positif (1). Distribusi sentimen menunjukkan kategori netral (0) paling dominan, mengindikasikan mayoritas opini publik tidak menunjukkan emosi kuat baik positif maupun negatif. Sentimen negatif memiliki jumlah signifikan, sedangkan sentimen positif relatif jarang. Penelitian ini menegaskan efektivitas algoritma Naive Bayes dalam analisis sentimen berbasis teks, serta memberikan wawasan penting tentang persepsi publik terkait kasus korupsi yang dianalisis.

Penulis Korespondensi:

Hamid Abdul Rozak
Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Purwokerto
Email: hamidrozak24@gmail.com

1. PENDAHULUAN

Korupsi merupakan tantangan global yang melintasi batas negara dan berbagai sektor pembangunan. Masalah ini tidak hanya mengancam stabilitas ekonomi, tetapi juga menjadi hambatan signifikan dalam mewujudkan pembangunan berkelanjutan. Secara etimologis, istilah korupsi berasal dari kata Latin *corruption*, yang dalam bahasa Inggris disebut *corruption* dan dalam bahasa Belanda *corruptive*. Secara umum, istilah ini merujuk pada tindakan yang rusak, tidak jujur, dan berkaitan dengan masalah keuangan[1]

Analisis sentimen merupakan proses otomatis untuk memahami dan mengolah data teks guna memperoleh informasi tertentu. Proses ini bertujuan untuk mendeteksi opini mengenai subjek atau objek tertentu, seperti individu, organisasi, atau produk, yang terdapat dalam suatu kumpulan data[2]. Dengan menganalisis sentimen, dapat diketahui apakah opini yang diungkapkan bersifat positif, negatif, atau netral. Informasi ini sangat berguna untuk memahami

persepsi publik, mengevaluasi kepuasan, atau bahkan memprediksi tren berdasarkan pola opini yang muncul dalam data tersebut[3]. Dalam konteks ini, analisis sentimen terhadap opini masyarakat mengenai kasus korupsi Harvey Moeis dapat memberikan gambaran yang lebih jelas mengenai bagaimana persepsi publik terhadap kasus tersebut. Informasi ini dapat menjadi masukan bagi pihak-pihak terkait, seperti pemerintah, media, atau organisasi non-pemerintah, untuk memahami respons masyarakat dan merumuskan langkah yang tepat[4].

X merupakan salah satu platform media sosial yang digemari oleh masyarakat di seluruh dunia. Hal ini terlihat dari peningkatan jumlah pengguna X yang tercatat secara global, termasuk di Indonesia. Selain sebagai sarana untuk bersosialisasi dan berkomunikasi, X juga dimanfaatkan menyampaikan pendapat serta merepresentasikan berbagai isu yang sedang terjadi di tengah masyarakat[5]. Hal ini dapat dijadikan sebagai sebuah acuan untuk dilakukan analisis sentimen masyarakat terhadap kasus korupsi Harvey Moeis. Analisis sentimen adalah metode untuk mengklasifikasikan dokumen teks berdasarkan opini dan pandangan dari kelompok masyarakat tertentu. Proses ini dilakukan dengan memahami, mengolah, hingga mengekstrak data dalam bentuk dokumen teks secara otomatis[6].

Naive Bayes adalah salah satu algoritma yang sering digunakan dalam analisis sentimen karena kesederhanaannya dan efisiensinya dalam memproses data teks. Algoritma ini berbasis pada teori probabilitas yang mengasumsikan bahwa setiap fitur independen terhadap fitur lainnya[7]. Meskipun sederhana, Naive Bayes telah terbukti efektif dalam berbagai studi analisis sentimen. Dalam konteks analisis sentimen, algoritma Naive Bayes digunakan untuk mengklasifikasikan teks ke dalam kategori sentimen, seperti positif, negatif, atau netral. Proses klasifikasi ini dilakukan dengan menghitung probabilitas kemunculan kata-kata tertentu dalam setiap kategori. Sebagai contoh, kata-kata seperti baik atau luar biasa memiliki peluang lebih tinggi untuk diklasifikasikan ke dalam kategori sentimen positif. Sebaliknya, kata-kata seperti buruk atau mengecewakan cenderung dikaitkan dengan sentimen negatif[8].

Berdasarkan uraian di atas penelitian ini bertujuan menganalisis opini publik melalui media sosial X untuk menemukan klasifikasi positif, negatif dan netral terhadap kasus korupsi Harvey Moeis, adapun metode klasifikasi data yang digunakan oleh penulis yaitu Naïve Bayes.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis pembelajaran mesin. Pemilihan metode ini didasarkan pada kemampuannya dalam menganalisis data tekstual dalam jumlah besar secara sistematis dan objektif[9]. Algoritma Naive Bayes dipilih karena efektivitasnya dalam klasifikasi teks dan analisis sentimen, seperti yang ditunjukkan dalam penelitian sebelumnya[10].

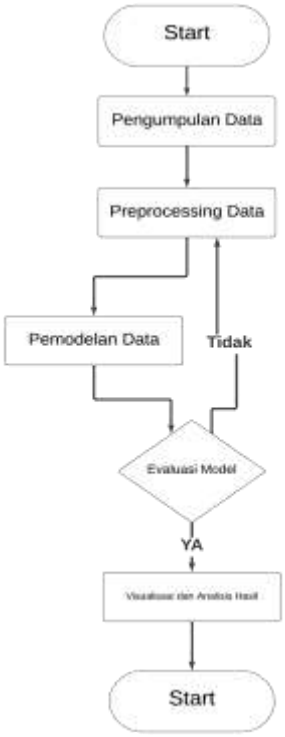
2.1. Desain penelitian

Penelitian dirancang dengan pendekatan cross-sectional, yaitu pengumpulan data dilakukan dalam satu waktu tertentu tanpa adanya intervensi terhadap variabel. Pendekatan ini dipilih karena sesuai untuk menggambarkan kondisi sentimen masyarakat pada periode tertentu secara deskriptif dan analitis[11]. Pengumpulan data dilakukan dengan metode web scraping pada aplikasi X menggunakan kata kunci Harvey Moeis. Metode ini digunakan karena mampu mengotomatisasi proses pengambilan data dalam skala besar dengan waktu yang efisien[12].

Populasi penelitian mencakup seluruh unggahan yang mengandung kata kunci tersebut, sementara sampel sebanyak 100 entri dipilih dengan teknik *purposive sampling*. Teknik ini dipilih karena memungkinkan peneliti menentukan kriteria tertentu yakni relevansi konten terhadap topik korupsi Harvey Moeis sehingga data yang diperoleh lebih fokus dan sesuai dengan tujuan penelitian[13].

2.2. Prosedur Penelitian

Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1 , yang menggambarkan tahapan dari pengumpulan data hingga analisis hasil.



Gambar 1. *Flowchart*

2.3. Spesifikasi Perangkat

Tabel 1. Spesifikasi perangkat

Kompnen	Spesifikasi
Prosesor	AMD Ryzen 2600
RAM	16GB
Sistem Operasi	Windows 10
Software Analisis	Google Collab
Pustaka Python	Pandas, NLTK, Scikit-learn, TextBlob, Matplotlib, Seaborn

2.4. Metode Pengumpulan dan Analisis Data

Data dikumpulkan menggunakan teknik web scraping yang diimplementasikan melalui API aplikasi X. Dataset terdiri dari 100 entri dengan atribut: Teks unggahan, Jumlah balasan, Jumlah suka, Lokasi dan Tanggal unggahan. Analisis data dilakukan dalam beberapa tahap:

2.4.1. Proses *preprocessing*

Dilakukan dengan menerapkan teknik *Natural Language Processing* (NLP) yang mencakup pembersihan data (*cleansing*), konversi huruf menjadi bentuk seragam (*case folding*), tokenisasi, penghapusan kata tidak penting (*stopwords removal*), serta stemming menggunakan pustaka Sastrawi. Langkah ini bertujuan untuk mengurangi kompleksitas teks serta mempertahankan fitur-fitur penting yang relevan bagi proses klasifikasi sentimen[14]

2.4.2. Transformasi fitur teks

Menggunakan pendekatan TF-IDF (*Term Frequency–Inverse Document Frequency*), yang memberikan bobot tinggi pada kata-kata yang sering muncul di dokumen tertentu tetapi jarang di keseluruhan korpus. Metode ini memungkinkan sistem mengenali istilah penting yang mewakili konten teks secara signifikan dan mengurangi pengaruh kata-kata umum[15].

2.4.3. Algoritma Naive Bayes

Karena kemampuannya dalam menangani klasifikasi data teks dengan efisien dan akurat. Data dibagi menggunakan teknik *stratified splitting* untuk memastikan distribusi label kelas tetap seimbang antara data latih dan data uji. Teknik ini penting untuk mencegah model bias terhadap kelas mayoritas selama proses pelatihan[16].

2.4.4. Penerapan teknik SMOTE

Untuk mengatasi ketidakseimbangan data, Teknik SMOTE (*Synthetic Minority Oversampling Technique*) dipilih untuk menangani distribusi yang tidak seimbang antar kelas sentimen. SMOTE bekerja dengan menghasilkan data sintetis berdasarkan fitur dari data minoritas, sehingga model tidak bias terhadap kelas mayoritas dan mampu belajar secara seimbang dari seluruh label kelas[17]

2.4.5. Evaluasi performa model

Dilakukan menggunakan metrik klasifikasi standar, yaitu akurasi, precision, recall, dan F1-score. Selain itu, digunakan juga *classification report* dan *confusion matrix* untuk mengetahui efektivitas prediksi pada tiap kategori sentimen. Untuk memastikan hasil valid, digunakan *k-fold cross-validation* sehingga model diuji secara menyeluruh dengan pembagian data yang merata[18].

Validitas hasil penelitian dijamin melalui penerapan teknik cross-validation dan penggunaan metrik evaluasi yang komprehensif. Keandalan model diuji menggunakan dataset yang belum pernah dilihat sebelumnya (data uji) untuk memastikan generalisasi yang baik.

3. HASIL DAN PEMBAHASAN

Hasil yang diperoleh dari berbagai tahapan penelitian dipaparkan dalam bentuk grafik, tabel, dan penjelasan yang mendukung agar memudahkan pembaca memahami keseluruhan proses dan hasilnya. Penelitian ini berfokus pada analisis sentimen berbasis teks dari media sosial Twitter. Hasil yang disajikan meliputi Deskripsi Data, Training Model, evaluasi kinerja model yang digunakan, serta Analisis Distribusi Sentimen.

3.1 Deskripsi Data

Pada penelitian ini, data yang digunakan berasal dari hasil crawling data di Twitter. Data awal mencakup teks mentah yang diproses melalui beberapa langkah penting, termasuk stemming, penghapusan stopword, dan tokenisasi menggunakan pustaka Sastrawi. Data yang diambil berjumlah 110 data. Setelah dilakukan preprocessing dengan rasio tertentu. Untuk menangani ketidakseimbangan jumlah antara kelas sentimen (positif, negatif, netral), digunakan metode memastikan setiap kelas memiliki distribusi data yang seimbang, yang sangat penting untuk meningkatkan model. Gambar berikut ini menunjukan hasil sebelum di SMOTE dan sesudah di SMOTE.

Tabel 2. Metode SMOTE

Kelas Sentimen	Sebelum SMOTE	Sesudah SMOTE
Positif	9	74
Negatif	27	74
Netral	74	74

Setelah melakukan SMOTE dan pembagian data kita mendapatkan data latih dan data uji sebagai berikut yang ada pada tabel ini.

Tabel 3. Pembagian data

Dataset	Jumlah Data
Data Latih	177
Data Uji	45

3.2. Training Model

Multinomial Naive Bayes (MultinomialNB) adalah algoritma klasifikasi probabilistik yang merupakan varian dari Naive Bayes. Model ini khusus digunakan untuk data diskrit atau kategori, seperti jumlah kemunculan kata dalam dokumen atau fitur yang memiliki nilai diskrit. Salah satu aplikasi paling umum adalah klasifikasi teks, terutama dalam masalah seperti spam filtering atau analisis sentimen. Berikut adalah penjelasan lebih mendalam mengenai setiap konsep yang terkait dengan Multinomial Naive Bayes.

3.2.1. Probabilitas Kelas *P(k)*

Probabilitas kelas *k* secara umum (jumlah dokumen dalam kelas k dibagi jumlah total dokumen).

$$P(k)=\frac{\text{Jumlah dokumen dengan kelas k}}{\text{jumlah total dokumen}}$$

3.2.2. Probabilitas Fitur *P(xi|k)*

Probabilitas kata *xi* muncul di kelas k, yang dihitung berdasarkan jumlah kemunculannya dalam dokumen kelas k.

$$P(xi|k)=\frac{\text{jumlah kemunculan kata xi dalam kelas k + a}}{\text{jumlah total kata dalam kelas k + aV}}$$

3.3. Evaluasi Model

3.3.1. *Classification Report*

Hasil Evaluasi *Classification Report* ini menunjukkan performa untuk tiga kelas berbeda yaitu -1, 0, dan 1. Kelas -1 memiliki precision 0.88, recall 1.00, dan f1-score 0.93 dengan support 21 sampel. Kelas 0 mencapai precision 1.00, recall 0.64, dan f1-score 0.78 dengan support 11 sampel. Sementara kelas 1 menunjukkan precision 0.93, recall 1.00, dan f1-score 0.96 dengan support 13 sampel. Secara keseluruhan, model mencapai accuracy 0.91 atau 91%. Untuk rata-rata macro, model memperoleh precision 0.93, recall 0.88, dan f1-score 0.89. Sedangkan untuk weighted average, model mencapai precision 0.92, recall 0.91, dan f1-score 0.90. Total jumlah sampel yang digunakan adalah 45. Model ini menunjukkan performa yang cukup baik secara keseluruhan, dengan kinerja terbaik pada kelas -1 dan 1, sementara kelas 0 memiliki performa yang sedikit lebih rendah terutama pada aspek recall. Gambar ini menunjukan hasil Classification Report.

Classification Report:

	precision	recall	f1-score	support
-1	0.88	1.00	0.93	21
0	1.00	0.64	0.78	11
1	0.93	1.00	0.96	13
accuracy			0.91	45
macro avg	0.93	0.88	0.89	45
weighted avg	0.92	0.91	0.90	45

Accuracy: 0.91

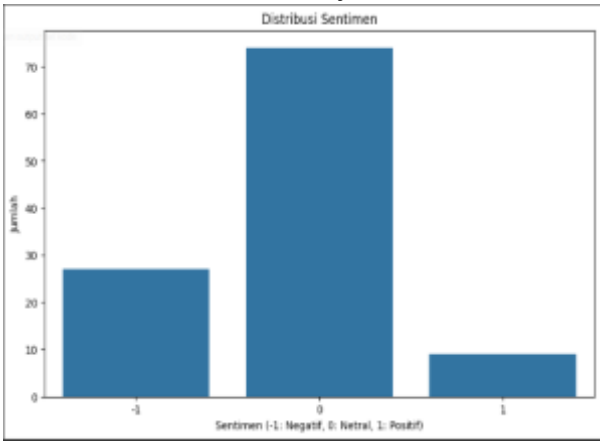
Gambar 2. *Classification Report*

3.3.2. Distribusi Sentimen

Distribusi sentimen dari analisis data yang dibagi menjadi tiga kategori: negatif (-1), netral (0), dan positif (1). Sumbu horizontal menunjukkan kategori sentimen, sedangkan sumbu vertikal menunjukkan jumlah data dalam setiap kategori.

Sentimen netral (0) merupakan kategori yang paling dominan, dengan jumlah lebih dari 70 data. Hal ini menunjukkan bahwa mayoritas data yang dianalisis tidak memiliki kecenderungan emosi yang kuat, baik ke arah positif maupun negatif. Kategori sentimen negatif (-1) menempati posisi kedua dengan jumlah sekitar 30 data, mengindikasikan adanya opini negatif yang cukup signifikan dalam dataset, seperti keluhan atau kritik. Sementara itu,

sentimen positif (1) memiliki jumlah paling sedikit, yaitu kurang dari 10 data, menunjukkan bahwa opini atau emosi positif jarang ditemukan dalam dataset ini. Gambar ini menunjukan hasil evaluasi Distribusi Sentimen.



Gambar 3. Distribusi sentimen

4. KESIMPULAN

Penelitian ini berhasil mengidentifikasi sentimen pengguna media sosial X terhadap kasus korupsi Harvey Moeis menggunakan algoritma Naive Bayes. Data yang diperoleh menunjukkan bahwa mayoritas opini publik bersifat netral, dengan sentimen negatif menempati posisi kedua dan sentimen positif menjadi yang paling sedikit. Model Multinomial Naive Bayes yang digunakan dalam penelitian ini menunjukkan performa yang baik dengan akurasi sebesar 91%, yang didukung oleh teknik balancing data menggunakan SMOTE.Hasil ini menunjukkan bahwa mayoritas masyarakat cenderung tidak mengekspresikan opini yang kuat terkait kasus ini, sementara opini negatif memiliki tingkat signifikan yang mencerminkan kritik atau ketidakpuasan. Penelitian ini memberikan gambaran penting tentang persepsi publik di media sosial dan dapat menjadi rujukan bagi pihak terkait, seperti pemerintah atau organisasi non-pemerintah, dalam memahami dan merespons pandangan masyarakat terhadap isu korupsi.

REFERENSI

[1] T. Pipit Muliyah, Dyah Aminatun, Sukma Septian Nasution, Tommy Hastomo, Setiana Sri Wahyuni Sitepu, "DAMPAK KORUPSI BAGI HAK ASASI MANUSIA DI INDONESIA," *J. GEEJ*, vol. 7, no. 2, pp. 1–12, 2020, doi: 10.8734/CAUSA.v1i2.365.

[2] T. Nurahman, Y. Azhar, and N. Hayatin, "Analisis Sentimen Konten Radikal dalam Kontestasi Politik 2019 di Media Twitter Menggunakan Interjection dan Punctuation," *J. Repos.*, vol. 2, no. 7, p. 905, 2020, doi: 10.22219/repositor.v2i7.868.

[3] F. M. Sarimole and W. Septian, "Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 890–899, 2024, [Online]. Available: http://ejournal.sisfokomtek.org/index.php/saintek/article/view/1359

[4] A. R. Abdillah and F. N. Hasan, "Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier," *Smatika J.*, vol. 13, no. 01, pp. 117–130, 2023, doi: 10.32664/smatika.v13i01.750.

[5] A. Agustian, T. Tukiro, and F. Nurapriani, "Penerapan Analisis Sentimen Dan Naive Bayes Terhadap Opini Penggunaan Kendaraan Listrik Di Twitter," *J. TIKA*, vol. 7, no. 3, pp. 243–249, 2022, doi: 10.51179/tika.v7i3.1550.

[6] S. Suryono and E. Taufiq Luthfi, "Analisis sentimen pada Twitter dengan menggunakan metode Naïve Bayes Classifier," *Jnanaloka*, no. 51, pp. 81–86, 2021, doi: 10.36802/jnanaloka.2020.v1-no2-81-86.

[7] T. Setiawan, S. Liem, and D. M. R. Pribadi, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Komentar Tiktok pada Produk Skincare," *Appl. Inf. Technol. Comput. Sci.*, vol. 3, no. 2, pp. 28–32, 2024, [Online]. Available: https://jurnal.politap.ac.id/index.php/aicoms

[8] D. Alita and R. A. Shodiqin, "Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 1–12, 2023, doi: 10.58602/jaiti.v1i1.20.

[9] V. Giordano, I. Spada, F. Chiarello, and G. Fantoni, "The impact of ChatGPT on human skills: A quantitative study on twitter data," *Technol. Forecast. Soc. Change*, vol. 203, no. April, p. 123389, 2024, doi: 10.1016/j.techfore.2024.123389.

[10] F. N. Hidayat and S. Sugiyono, "Analisis Sentimen Masyarakat Terhadap Perekrutan Pppk Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *J. Sains dan Teknol.*, vol. 5, no. 2, pp. 665–672, 2023, doi: 10.55338/saintek.v5i2.1359.

[11] M. S. Setia, "Methodology series module 3: Cross-sectional studies," *Indian J. Dermatol.*, vol. 61, no. 3, pp. 261–264, 2016, doi: 10.4103/0019-5154.182410.

[12] A. Z. Rizquina and C. I. Ratnasari, "Implementasi Web Scraping untuk Pengambilan Data Pada Website E-Commerce," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 4, pp. 377–383, 2023, doi: 10.47233/jteksis.v5i4.913.

[13] L. A. Palinkas, S. M. Horwitz, C. A. Green, J. P. Wisdom, N. Duan, and K. Hoagwood, "Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research," *Adm. Policy Ment. Heal. Ment. Heal. Serv. Res.*, vol. 42, no. 5, pp. 533–544, 2015, doi: 10.1007/s10488-013-0528-y.

[14] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter.pdf," *INOVTEK Polbeng* -

Seri Inform., vol. 3, no. 1, p. 50, 2018.
[15] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.
[16] Irvandi, B. Irawan, and O. Nurdiawan, "Naive Bayes Dan Wordcloud Untuk Analisis Sentimen Wisata Halal Pulau Lombok," *INFOTECH J.*, vol. 9, no. 1, pp. 236–242, 2023, doi: 10.31949/infotech.v9i1.5322.
[17] R. Peranginangin, E. J. G. Harianja, I. K. Jaya, and B. Rumahorbo, "Penerapan Algoritma Safe-Level-Smote Untuk Peningkatan Nilai G-Mean Dalam Klasifikasi Data Tidak Seimbang," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 4, no. 1, pp. 67–72, 2020, doi: 10.46880/jmika.vol4no1.pp67-72.
[18] Agung Nugroho and Agit Amrullah, "Evaluasi Kinerja Algoritma K-Nn Menggunakan K-Fold Cross Validation Pada Data Debitur Ksp Galih Manunggal," *J. Inform. Teknol. dan Sains*, vol. 5, no. 2, pp. 294–300, 2023, doi: 10.51401/jinteks.v5i2.2506.