AICOMS
Applied Information Technology and Computer Science
e-ISSN: 2964-7703
Vol. 4 No. 2, Desember-2025, pp. 55-61

Opini Publik terhadap Isu Pengoplosan Pertamax di Youtube Menggunakan Metode Naive Bayes

Adikara Alif Nurrahman [1*], Earlando Moza [2], Ramanda md [3], Muhamad Rizvi Roshan [4], Ahmad Rizky [5] and Hafiz Irsyad [6]

[1] Universitas Multi Data Palembang; adikaraalf@mhs.mdp.ac.id
[2] Universitas Multi Data Palembang; earlandomoza_2226250096@mhs.mdp.ac.id
[2] Universitas Multi Data Palembang; ramandamd@mhs.mdp.ac.id
[2] Universitas Multi Data Palembang; rizviroshan10@mhs.mdp.ac.id
[2] Universitas Multi Data Palembang; ahmadrizky@mhs.mdp.ac.id
[2] Universitas Multi Data Palembang; hafizirsyad@mdp.ac.id
* Korespondensi: adikaraalf@mhs.mdp.ac.id

Info Artikel:
Dikirim: 07 November 2025
Direvisi: 13 November 2025
Diterima: 20 November 2025

Abstract: This study aims to explore public perceptions regarding the issue of Pertamax fuel adulteration, a topic that has sparked widespread discussion on YouTube, by employing sentiment analysis techniques based on the Naive Bayes algorithm. This issue has attracted significant public attention and become a trending topic on social media, particularly on the YouTube platform. The data analyzed in this research consist of user comments responding to the issue. The Naive Bayes algorithm is used to classify sentiments in the comments into three categories: positive, negative, and neutral. To address the imbalanced distribution of data, the Synthetic Minority Over-sampling Technique (SMOTE) is applied. The results show that before applying SMOTE, the model achieved an accuracy of only 48%, with a precision of 0.48, recall of 0.36, and an F1-score of 0.41 for the negative category, as well as a precision of 0.48, recall of 0.56, and an F1-score of 0.52 for the positive category. After implementing SMOTE, the model's accuracy increased significantly to 88%, with a precision of 0.91, recall of 0.93, and an F1-score of 0.92 for the negative category. For the positive category, precision improved to 0.80, although recall decreased to 0.75, yielding an F1-score of 0.77. The average precision, recall, and F1-score (macro average) after applying SMOTE reached 0.85, 0.84, and 0.85, respectively, representing a substantial improvement compared to the results before SMOTE. This study highlights the importance of using SMOTE to enhance sentiment analysis accuracy, particularly in addressing class imbalance issues within the dataset.

Keywords: Sentiment Analysis; Naive Bayes; Public Opinion; Pertamax Adulteration; SMOTE; YouTube.

Intisari: Penelitian ini bertujuan untuk mengeksplorasi pandangan publik terkait isu pengoplosan Pertamax yang ramai diperbincangkan di YouTube, dengan menggunakan teknik analisis sentimen berbasis algoritma Naive Bayes. Isu ini menarik perhatian masyarakat dan menjadi topik hangat yang banyak dibicarakan di media sosial, terutama di platform YouTube. Dalam penelitian ini, data yang dianalisis berasal dari komentar-komentar pengguna YouTube yang menanggapi isu tersebut. Algoritma Naive Bayes digunakan untuk mengkategorikan sentimen dalam komentar ke dalam tiga kategori: positif, negatif, dan netral. Untuk menangani ketidakseimbangan distribusi data, teknik Synthetic Minority Over-sampling Technique (SMOTE) diterapkan. Hasil penelitian menunjukkan bahwa sebelum penerapan SMOTE, akurasi model hanya mencapai 48%, dengan presisi 0,48, recall 0,36, dan skor F1 0,41 untuk kategori negatif, serta presisi 0,48, recall 0,56, dan skor F1 0,52 untuk kategori positif. Setelah penerapan SMOTE, akurasi meningkat signifikan menjadi 88%, dengan presisi 0,91, recall 0,93, dan skor F1 0,92 untuk kategori negatif. Untuk kategori positif, presisi meningkat menjadi 0,80, meskipun recall menurun menjadi 0,75, yang menghasilkan skor F1 sebesar 0,77. Rata-rata presisi, recall, dan F1 (rata-rata makro) setelah

SMOTE mencapai 0,85, 0,84, dan 0,85, menunjukkan peningkatan yang signifikan dibandingkan sebelum penerapan SMOTE. Penelitian ini menegaskan pentingnya penggunaan SMOTE dalam meningkatkan akurasi analisis sentimen, khususnya dalam menangani masalah ketidakseimbangan kelas dalam data

Kata Kunci: Analisis Sentimen; Naive Bayes; Opini Publik; Pengoplosan Pertamax; SMOTE; YouTube

1. Pendahuluan

Di era teknologi informasi dan komunikasi yang sudah berkembang ini, telah membawa perubahan besar terhadap cara masyarakat mengakses serta menyebarkan informasi. Salah satu media sosial yang kini banyak dimanfaatkan adalah YouTube. Platform ini tidak hanya berfungsi sebagai sarana hiburan, tetapi juga sebagai media penyebaran isu-isu aktual dan sosial yang sedang ramai dibicarakan. Melalui kolom komentar, para pengguna bisa menyampaikan tanggapan dan pendapat mereka secara langsung, menjadikan YouTube ruang terbuka bagi publik untuk menyuarakan opini mereka secara bebas.

Opini publik merupakan pendapat publik atau respon publik atas suatu masalah yang bisa dikatakan kontroversial. Opini publik juga dapat dikatakan sebagai beberapa pendapat atau opini perorangan yang terkumpul menjadi satu, dimana pendapat atau opini ini berawal dari suatu fenomena atau masalah [1]. Opini publik berkembang seiring dengan penyebaran informasi yang cepat melalui berbagai media sosial, terutama platform Youtube. Oleh karena itu, opini publik juga dapat terpengaruh dari lingkungan sekitar terhadap isu yang sedang terjadi.

Salah satu topik yang sempat mencuri perhatian publik mengenai isu tuduhan adanya pengoplosan bahan bakar jenis Pertamax. Pertamax merupakan jenis bahan bakar dengan angka oktan 92. Bensin pertamax dianjurkan digunakan untuk kendaraan bahan bakar bensin yang mempunyai perbandingan kompresi tinggi (9,1:1 sampai 10,0:1) [2]. Isu ini memicu berbagai tanggapan dari masyarakat, mulai dari kekhawatiran terhadap mutu bahan bakar yang beredar, hingga menurunnya kepercayaan terhadap pihak penyedia BBM. Selain itu, informasi mengenai kasus ini turut menyebar luas melalui media sosial, sering kali disertai berbagai spekulasi. Pandangan dan sikap masyarakat terhadap isu tersebut dapat terlihat dari kolom komentar pada video-video YouTube yang membahas topik ini, mencerminkan opini publik secara langsung dan spontan.

Untuk memahami lebih dalam mengenai pandangan masyarakat terhadap isu tersebut, dibutuhkan pendekatan analisis opini publik yang berbasis data. Salah satu teknik yang cukup populer dalam pengelompokan opini atau analisis sentimen adalah algoritma Naive Bayes. Algoritma naive bayes adalah algoritma pembelajaran mesin untuk masalah klasifikasi yang terutama digunakan untuk klasifikasi teks yang melibatkan kumpulan data pelatihan berdimensi tinggi [3].

Beberapa Penelitian mengenai analisis sentimen sudah banyak dilakukan sebelumnya, salah satunya yang berjudul *"Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER"* [4]. penelitian ini membahas tentang analisis sentimen akun Twitter terkait game Apex Legends menggunakan metode VADER dan aplikasi Orange Data Mining. Penelitian dilakukan terhadap 500 tweet dan menunjukkan bahwa persentase sentimen positif, negatif, dan netral yang diperoleh dari VADER cukup mendekati hasil penilaian pakar. Hasil analisis menunjukkan distribusi sentimen: positif 18%, negatif 4,8%, dan netral 77,2%. Analisis oleh pakar menunjukkan: positif 27%, negatif 10,8%, dan netral 62,2%. Akurasi VADER dibandingkan dengan penilaian pakar adalah 65,2%. Dari 10 sampel data, hanya 1 yang hasil analisis VADER dan pakar sama. Banyak kalimat yang menurut VADER positif/negatif, tetapi pakar menafsirkannya berbeda tergantung konteks game. Kesimpulannya, metode VADER efektif untuk analisis awal, tetapi perlu penyesuaian konteks agar lebih akurat dalam domain game seperti Apex Legends.

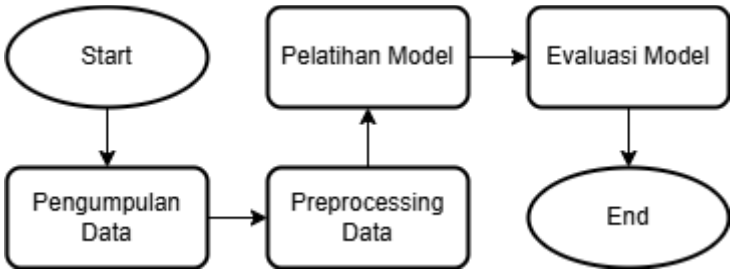
Adapun penelitian lainnya berjudul "*Analisis sentimen aplikasi gojek menggunakan support vector machine dan k nearest neighbor"* [5]. penelitian ini membahas tentang analisis sentimen terhadap ulasan pengguna aplikasi Gojek di Google Playstore menggunakan metode *Support Vector Machine* (SVM) dan K Nearest Neighbor (KNN). Tujuannya adalah membandingkan tingkat akurasi kedua metode dalam mengklasifikasi ulasan menjadi positif atau negatif. Hasilnya menunjukkan bahwa metode SVM dengan kernel linear dan parameter C=1 mencapai akurasi tertinggi sebesar 87,98%, sedangkan KNN dengan K=22 mencapai 82,14%. Kesimpulannya, SVM lebih efektif dalam klasifikasi sentimen ulasan pengguna Gojek. Kemudian untuk penelitian terakhir yang terkait dengan opini publik dengan judul "*Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine*" [6]. Hasil dari penelitian ini mendapatkan akurasi sebesar 84% untuk algoritma *Support Vector Machine* (SVM) dan mencapai 79% untuk algoritma Naive Bayes.

Penelitian ini mengimplementasikan algoritma Naive Bayes sebagai metode klasifikasi, dengan mempertimbangkan adanya ketidakseimbangan dalam dataset. Untuk menangani hal ini, digunakan teknik *Synthetic Minority*

Oversampling Technique (SMOTE), yang bertujuan untuk menyeimbangkan distribusi antar kelas dengan cara menambah jumlah data pada kelas minoritas melalui pembuatan sampel sintetik. Ketidakseimbangan dalam data dapat mempengaruhi kinerja model klasifikasi karena model lebih cenderung mempelajari data dari kelas mayoritas. Oleh karena itu, penerapan SMOTE diharapkan dapat memperbaiki distribusi kelas dan meningkatkan presisi serta akurasi model, karena hasil yang optimal tidak hanya bergantung pada algoritma yang digunakan, tetapi juga pada kondisi dan karakteristik dataset itu sendiri.

2. METODE PENELITIAN

Metode penelitian ini dilakukan dengan menggunakan Google Collab sebagai platform berbasis cloud untuk proses pengujian, dengan bahasa pemrograman Python. Algoritma Naïve Bayes digunakan sebagai pendekatan dalam metode machine learning guna melakukan klasifikasi sentimen. Adapun flowchart yang ditampilkan pada Gambar 1. menggambarkan rangkaian tahapan yang diterapkan dalam penelitian ini. Tahapan dalam penelitian ini diawali dengan proses data mining, kemudian dilanjutkan dengan tahap preprocessing data, visualisasi informasi, pemisahan data, pemberian bobot pada data, pembangunan model menggunakan Naïve Bayes, hingga tahap akhir berupa evaluasi untuk menilai performa model yang dihasilkan secara menyeluruh.



Gambar 1. Tahapan Proses

2.1. Pengumpulan Dataset

Tahapan penelitian ini dimulai dengan Pengumpulan data. Data ini dikumpulkan pada tanggal 07 NOvember 2025, Sumber data berasal dari komentar pengguna di platform YouTube, khususnya pada video yang membahas isu pengoplosan Pertamax. Komentar-komentar tersebut dikumpulkan dan disimpan dalam format CSV. Setiap entri dalam dataset mencakup sejumlah atribut seperti ID komentar, isi komentar, dan label sentimen (positif, negatif, atau netral). Hasil pengumpulan dapat dilihat pada tabel 1.

Tabel 1. Dataset

Nama Akun	Komentar
@fajarabdullatief1950	Jangan gunakan produk pertamina Gunakan lah produk shell
@hakimhakim6144	Sudah saatnya kita sebagai nitizen dan masyarakat indonesia menuntut hukuman mati kepada pejabat yang curang ayo kita ramai suarakan😢
@zelld8381	Kepercayaan masyarakat itu bukan terus menurun, tp sudah gk percaya
@yekdayat4622	Selama para koruptor tidak di hukum mati dan di siarkan langsung di TV maka selamanya para pejabat di negeri ini korupsi berjamaah
@chifond	Dulu katanya bensin eceran di Oplos karena botolnya di kocok dulu. eh taunya bapaknya yang ngoplos duluan 🚤🚤🚤

2.2. Preprocessing Data

Pre-processing data merupakan tahap awal yang penting untuk mempersiapkan data yang telah dikumpulkan agar siap diproses lebih lanjut [7]. Setelah menetapkan label pada setiap dataset yang diperoleh, langkah selanjutnya adalah implementasi pra-pemrosesan teks. Tujuan utama dari fase ini adalah untuk mengorganisasi data agar informasi yang tidak terstruktur dapat diubah menjadi format yang lebih terstruktur. Proses ini bertujuan untuk mengolah data teks menjadi informasi yang lebih mudah dianalisis, khususnya untuk analisis sentimen.

2.3. Pelatihan Model

Data dibagi ke dalam data pelatihan dan data pengujian, dimana 85% data digunakan untuk pelatihan dan 15% sisanya untuk pengujian. Teks dalam data pelatihan dan pengujian diubah menjadi format numerik melalui CountVectorizer. CountVectorizer merupakan sebuah metode atau proses pengolahan dokumen atau teks yang mengubah teks menjadi representasi vektor [8]. Setelah teks dikonversi ke format numerik, oversampling diterapkan pada data pelatihan dengan memanfaatkan teknik SMOTE (*Synthetic Minority Over-Sampling Technique*). SMOTE berperan dalam mengatasi isu ketidakseimbangan kelas pada data pelatihan dengan cara membuat contoh-contoh baru dari kelas yang lebih sedikit, sehingga menciptakan data pelatihan yang lebih seimbang.

2.3.1. Metode Naive Bayes

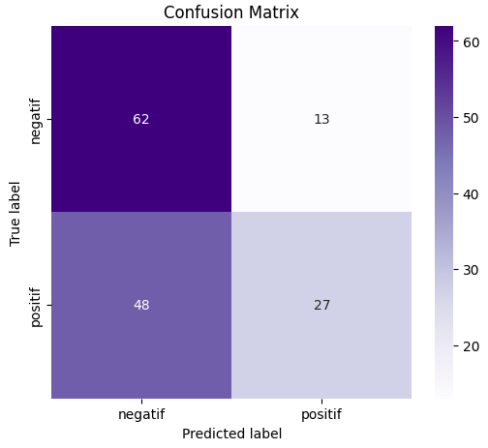
Metode Naive Bayes Classifier merupakan salah satu pengklasifikasi statistik, dimana klasifikasi ini dapat memprediksi probabilitas keanggotaan kelas suatu data yang akan masuk ke dalam kelas tertentu, sesuai dengan perhitungan probabilitas [9]. Penelitian ini memanfaatkan algoritma Naive Bayes, tepatnya tipe Multinomial Naive Bayes, yang dikenal efektif untuk tugas klasifikasi teks. Algoritma ini bekerja dengan menggunakan data berupa fitur-fitur diskrit seperti jumlah kemunculan kata dalam sebuah dokumen. Salah satu ciri khas dari metode ini adalah anggapannya bahwa setiap kata dalam teks tidak saling bergantung satu sama lain, yang dikenal sebagai naive assumption. Berdasarkan frekuensi kata dalam data pelatihan, model ini kemudian menghitung kemungkinan suatu teks termasuk dalam kategori sentimen tertentu.

2.3.2. SMOTE

SMOTE adalah salah satu penerapan dari metode oversampling, sehingga salah satu kelebihan metode ini adalah tidak akan menyebabkan adanya informasi yang hilang, dikarenakan tidak ada pengurangan data seperti yang dilakukan pada metode undersampling [10].

2.4. **Evaluasi Model**

Evaluasi model menggunakan Confusion Matrix. Confusion Matrix yang digunakan berukuran 2x2 untuk mewakili hasil klasifikasi biner pada dataset. Berbagai persamaan umum untuk komputasi efektivitas klasifikasi yang mencakup ringkasan hasil prediksi yang mencakup nilai akurasi, presisi, dan recall dapat ditampilkan dalam presentase [11].



Gambar 2. *Confusion Matrix*

Gambar 2, Confusion matrix adalah tabel yang berfungsi untuk menilai kinerja model klasifikasi. Matriks ini terdiri dari dua baris dan dua kolom, di mana setiap sel mengandung angka. Baris melambangkan label yang sebenarnya, sedangkan kolom menunjukkan label yang diprediksi. Ada empat sel:

1. Sel kiri atas (*true negative*) berisi angka 62, menunjukkan bahwa 62 sentimen dengan benar diprediksi sebagai "negatif".
2. Sel kanan atas (*false positive*) berisi angka 13, menunjukkan bahwa 13 sentimen salah diprediksi sebagai "positif" padahal sebenarnya "negatif".

3. Sel kiri bawah (*false negative*) berisi angka 48, menunjukkan bahwa 48 sentimen salah diprediksi sebagai "negatif" padahal sebenarnya "positif".
4. Sel kanan bawah (*true positive*) berisi angka 27, menunjukkan bahwa 27 sentimen dengan benar diprediksi sebagai "positif".

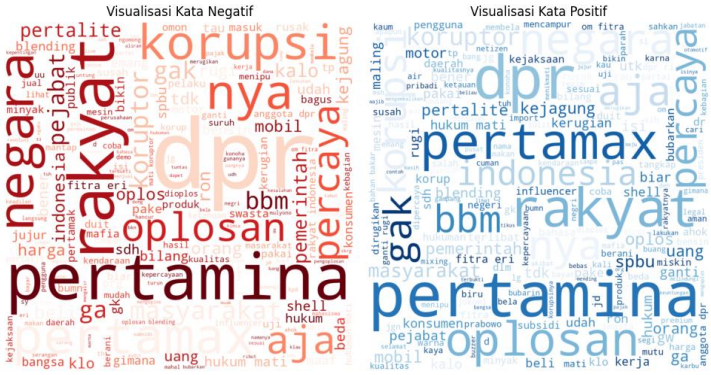
Confusion matrix memberikan pemahaman mengenai seberapa efektif model dalam membedakan antara kelas 'positif' dan 'negatif'. Selain menggambarkan jumlah prediksi yang akurat, matriks ini juga menjelaskan tipe kesalahan yang muncul.

3. Hasil dan Pembahasan

3.1. **Hasil Preprocessing**

Adapun teknik filtering yang digunakan pada tahap ini di antaranya [12] [13] :

1. Pembersihan Teks : Tahap ini untuk menghilangkan karakter yang tidak relevan dan mengonversi seluruh huruf menjadi format huruf kecil.
2. Normalisasi : Tahap ini dengan mengganti kata-kata tidak baku atau istilah gaul menjadi bentuk baku menggunakan kamus normalisasi khusus.
3. Stopword Removal : kata-kata umum yang tidak memiliki kontribusi signifikan dalam analisis sentimen akan dihapus dengan menggunakan pustaka Sastrawi.
4. Tokenisasi dan Stemming : Memecah kalimat menjadi kata-kata dasar (token) dan mengembalikan kata ke bentuk dasarnya menggunakan Sastrawi Stemmer.
5. Penyimpanan Data Bersih : Data hasil pembersihan disimpan ke dalam file CSV untuk digunakan pada tahap berikutnya.



Gambar 3. Visualisasi hasil data preprocessing

Seperti yang sudah dilihat pada Gambar 3, dataset yang sudah di-preprocessing akan ditampilkan dalam bentuk word cloud yang ditampilkan dalam bentuk token di setiap pengelompokannya yang dimana kata "pertamina" muncul di kedua kelompok data kata yang merupakan kata itu sendiri adalah kata pokok atau utama dalam penelitian ini.

3.2. **Naïve Bayes**

Algoritma Naive Bayes digunakan untuk menganalisis sentimen terhadap komentar yang diperoleh pada YouTube mengenai perilaku oposisi. data yang diperoleh dikelompokkan menjadi 2 kategori sentimen, yaitu positif dan negatif. Untuk mengatasi masalah ketidakseimbangan data, penelitian ini menerapkan model SMOTE yang bertujuan untuk meningkatkan akurasi dari 48% menjadi 88%. Dengan demikian, algoritma Naive Bayes memberikan hasil yang konsisten dan menguntungkan setelah dilakukan perubahan data.

Tabel 2. Naive Bayes tanpa *SMOTE*

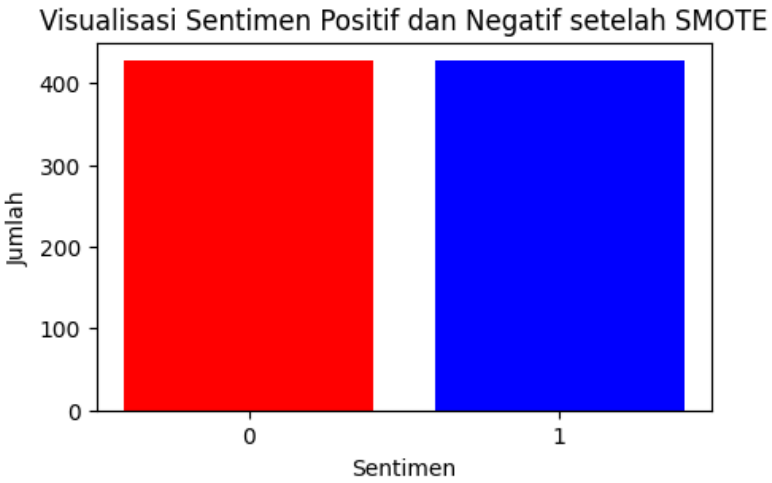
Akurasi Model Laporan Klasifikasi	**Naive Bayes : 0.48**			
	precision	recall	fi-score	support
negatif	0.48	0.36	0.41	75
positif	0.48	0.56	0.52	75
accuracy			0.48	150
macro avg	0.48	0.48	0.48	150
weighted avg	0.48	0.48	0.48	150

Tabel 3. Naive Bayes dengan *SMOTE*

Akurasi Model Laporan Klasifikasi	**Naive Bayes : 0.88**			
	precision	recall	fi-score	support
negatif	0.91	0.93	0.92	145
positif	0.80	0.75	0.77	55
accuracy				0.88
macro avg	0.85	0.84	0.85	200
weighted avg	0.88	0.88	0.88	200

3.3. **SMOTE**

SMOTE adalah salah satu penerapan dari metode oversampling, sehingga salah satu kelebihan metode ini adalah tidak akan menyebabkan adanya informasi yang hilang, dikarenakan tidak ada pengurangan data seperti yang dilakukan pada metode undersampling [11].



Gambar 4. Visualisasi *SMOTE*

Gambar 4 Menjelaskan distribusi sampel dalam setiap kelas (0 mewakili negatif dan 1 mewakili positif) setelah implementasi SMOTE. Terbukti dari grafik bahwa aplikasi pasca-SMOTE, ada jumlah sampel yang sama di kelas positif dan negatif, dengan setiap kelas berisi 400 sampel.

4. Kesimpulan

Penelitian ini dengan efektif mengidentifikasi sudut pandang publik mengenai pembahasan masalah isu pengoplosan Pertamax melalui analisis sentimen terhadap komentar YouTube dengan memanfaatkan algoritma Naive Bayes dan teknik SMOTE. Hasil evaluasi mengindikasikan bahwa penerapan SMOTE meningkatkan akurasi model dari 48% menjadi 88%. Walaupun terdapat penurunan recall pada kategori positif, presisi justru meningkat untuk kategori itu. SMOTE terutama memperbaiki daya ingat pada kategori negatif, menunjukkan bahwa kemampuan model dalam mendeteksi kategori minoritas bertambah setelah penerapan metode ini. Penelitian ini menekankan signifikansi metode pengambilan sampel ulang seperti SMOTE dalam menangani ketidakseimbangan kelas pada dataset,

khususnya dalam analisis sentimen. Hasil penelitian ini dapat menjadi landasan untuk menambah wawasan yang lebih mendalam terhadap opini publik dari berbagai isu yang sedang terjadi.

Daftar Pustaka

[1] D. A. Ramadhanti, "*Opini Publik terhadap Drama SBS Racket Boys di Media Sosial*," vol. 1, no. 11. 2021.
[2] S. Mulyono, G. Gunawan, and B. Maryanti, "Pengaruh Penggunaan dan Perhitungan Efisiensi Bahan Bakar Premium dan Pertamax Terhadap Unjuk Kerja Motor Bakar Bensin," *JTT (Jurnal Teknol. Terpadu)*, vol. 2, no. 1, pp. 28–35, 2014, doi: 10.32487/jtt.v2i1.38.
[3] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
[4] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Y. Yusra, "Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 3, pp. 423–431, 2022, doi: 10.32672/jnkti.v5i3.4382.
[5] M. N. Muttaqin and I. Kharisudin, "Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor," *UNNES J. Math.*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: http://journal.unnes.ac.id/sju/index.php/ujm
[6] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.
[7] A. Y. Simanjuntak, I. S. S. Simatupang, and Anita, "Implementasi Data Mining Menggunakan Metode Naïve Bayes Classifier Untuk Data Kenaikan Pangkat Dinas," *J. Sci. Soc. Res.*, vol. 4307, no. 1, pp. 85–91, 2022.
[8] R. Vincent, I. Maulana, and O. Komarudin, "Perbandingan Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Analisis Sentimen Dengan Multiclass Di Twitter," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2496–2505, 2024, doi: 10.36040/jati.v7i4.7152.
[9] O. V. Putra, T. Harmini, and A. Saroji, "Outlier Detection On Graduation Data Of Darussalam Gontor University Using One-Class Support Vector Machine," *Procedia Eng. Life Sci.*, vol. 2, no. 2, pp. 89–92, 2021, doi: 10.21070/pels.v2i0.1139.
[10] N. P. Y. T. WIJAYANTI, E. N. KENCANA, and I. W. SUMARJAYA, "Smote: Potensi Dan Kekurangannya Pada Survei," E-Jurnal Mat., vol. 10, no. 4, p. 235, 2021, doi: 10.24843/mtk.2021.v10.i04.p348.
[11] I. D. Id, *MACHINE LEARNING : Teori, Studi Kasus dan Implementasi Menggunakan Python*. Unri Press. [Online]. Available: https://books.google.co.id/books?id=JvBPEAAAQBAJ
[12] H. I. Putra Ganda Dewata1, Azzar Rizky2, "Jurnal Rein," *Anal. Sentimen Terhadap Boikot Prod. Isr. Menggunakan Algoritm. Naive Bayes Dan SMOTE*, vol. 1, no. 1, pp. 16–21, 2024.
[13] D. Darmanto, N. I. Pradasari, and E. Wahyudi, "Sistem Deteksi Plagiarisme Tugas Akhir Mahasiswa Berbasis Natural Language Processing Menggunakan Algoritma Jaro-Winkler dan TF-IDF," Smart Comp: Jurnalnya Orang Pintar Komputer, vol. 13, no. 1, pp. 201–211, 2024.