

PENERAPAN METODE NAÏVE BAYES UNTUK MENGANALISIS SENTIMEN PUBLIK TERHADAP LAYANAN TRANSPORTASI ONLINE DI INDONESIA BERBASIS DATA MEDIA SOSIAL X

Fahri Nasution¹, Zuli Agustina Gultom²

^{1,2}Universitas Muhammadiyah Sumatera Utara, Indonesia

Email: fahrinasution105@gmail.com

ABSTRAK

Layanan transportasi online seperti Gojek dan Grab telah menjadi bagian penting dari mobilitas masyarakat di Indonesia. Tingginya penggunaan layanan ini memicu banyaknya opini publik yang tersebar di media sosial X, baik berupa apresiasi maupun keluhan. Penelitian ini bertujuan untuk menerapkan text mining dan metode Naïve Bayes dalam menganalisis dan memetakan sentimen publik terhadap layanan transportasi online agar dapat dimanfaatkan sebagai bahan evaluasi kualitas layanan berbasis data. Dataset yang digunakan berjumlah 1.930 tweet mengenai Gojek dan Grab yang dikumpulkan dari Januari 2023 hingga Desember 2025, lalu dilabeli secara manual ke dalam kelas positif dan negatif. Data diproses melalui tahapan preprocessing (termasuk *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*), lalu dibobotkan menggunakan metode TF-IDF sebelum diklasifikasikan menggunakan algoritma Multinomial Naïve Bayes dengan rasio pembagian data latih dan uji sebesar 80:20. Hasil pengujian model menunjukkan performa yang sangat baik dengan tingkat accuracy sebesar 87,31%, precision 89,53%, recall 83,24%, F1-score 86,27%, dan nilai ROC-AUC mencapai 0,9574. Hasil pemetaan klasifikasi sentimen menunjukkan bahwa terdapat 923 tweet (47,8%) bersentimen positif dan 1.007 tweet (52,2%) bersentimen negatif. Hal ini membuktikan bahwa metode Naïve Bayes dengan pembobotan TF-IDF sangat efektif dalam menganalisis opini teks, serta menunjukkan bahwa opini masyarakat saat ini terhadap layanan transportasi online di Indonesia masih didominasi oleh keluhan yang sedikit lebih besar dibandingkan apresiasi.

Kata Kunci: Analisis Sentimen, Naïve Bayes, TF-IDF, Transportasi Online, Media Sosial X

ABSTRACT

Online transportation services such as Gojek and Grab have become an essential part of community mobility in Indonesia. The high usage of these services triggers massive public opinion spread on X social media, in the form of both appreciation and complaints. This study aims to apply text mining and the Naïve Bayes method to analyze and map public sentiment towards online transportation services so that it can be utilized as data-driven evaluation material for service quality. The dataset used consists of 1,930 tweets regarding Gojek and Grab collected from January 2023 to December 2025, which were manually labeled into positive and negative classes. The data were processed through preprocessing stages (including *cleaning*, *case folding*, *tokenizing*, *stopword removal*, and *stemming*), then weighted using the TF-IDF method before being classified using the Multinomial Naïve Bayes algorithm with a 80:20 ratio of training and testing data. The model evaluation results demonstrated excellent performance with an accuracy of 87.31%, precision of 89.53%, recall of 83.24%, F1-score of 86.27%, and an ROC-AUC value reaching 0.9574. The sentiment mapping results revealed that there were 923 tweets (47.8%) with positive sentiment and 1,007 tweets (52.2%) with negative sentiment. This proves that the Naïve Bayes method combined with TF-IDF weighting is highly effective in analyzing text opinions, and also indicates

that current public opinion regarding online transportation services in Indonesia is still dominated by complaints, which are slightly higher than appreciation.

Keywords: Sentiment Analysis, Naïve Bayes, TF-IDF, Online Transportation, X Social Media.

PENDAHULUAN

Perkembangan teknologi informasi, internet, dan perangkat bergerak telah membawa perubahan besar dalam berbagai aspek kehidupan masyarakat, termasuk pada sektor transportasi. Perubahan tersebut mendorong lahirnya layanan transportasi online yang menawarkan kemudahan, kecepatan, dan efisiensi melalui pemanfaatan aplikasi digital. Di Indonesia, layanan seperti Gojek dan Grab telah menjadi bagian penting dari mobilitas masyarakat karena mampu memberikan alternatif yang lebih praktis dibandingkan transportasi konvensional. Kondisi ini menunjukkan bahwa transportasi online tidak lagi sekadar pilihan tambahan, tetapi telah berkembang menjadi bagian dari kebutuhan harian masyarakat modern (Khoiruddin et al., 2023).

Selain Gojek, Grab juga merupakan salah satu penyedia layanan transportasi online terbesar di Indonesia yang memiliki basis pengguna sangat luas. Grab menyediakan berbagai layanan, seperti transportasi, pengantaran makanan, pengiriman barang, dan pembayaran digital yang digunakan masyarakat dalam aktivitas sehari-hari. Seiring meningkatnya jumlah pengguna, volume ulasan dan opini terhadap layanan Grab juga terus bertambah. Penelitian (Nurfauzi et al., 2025) menunjukkan bahwa ribuan ulasan pengguna Grab dapat dimanfaatkan sebagai sumber informasi untuk mengevaluasi kualitas layanan, mengidentifikasi permasalahan operasional, serta memberikan masukan bagi pengembangan layanan. Oleh karena itu, analisis terhadap opini pengguna Grab menjadi penting untuk memahami persepsi masyarakat terhadap salah satu layanan transportasi online terbesar di Indonesia.

Seiring dengan meningkatnya penggunaan layanan transportasi online, volume opini publik terhadap layanan tersebut juga semakin besar. Opini tersebut muncul dalam bentuk ulasan aplikasi maupun unggahan di media sosial yang memuat pengalaman pengguna, kritik, keluhan, serta apresiasi terhadap layanan yang mereka gunakan. (Yuniar & Kismiantini, 2023) menunjukkan bahwa dari 4.198 ulasan Gojek yang dianalisis, data cenderung memiliki sentimen positif dan metode Naïve Bayes menghasilkan akurasi tertinggi sebesar 89,9%. Penelitian (L et al., 2023) juga menemukan bahwa dari 6.485 ulasan pelanggan Gojek, terdapat 4.387 sentimen positif dan 2.098 sentimen negatif, dengan akurasi model Naïve Bayes sebesar 88,5%. Sementara itu, penelitian (Rifa'I et al., 2024) pada aplikasi Grab menghasilkan 1.677 sentimen positif dan 69 sentimen negatif, dengan akurasi sebesar 84,36%. Hasil tersebut diperkuat oleh penelitian (Nurfauzi et al., 2025) yang menganalisis 2.140 ulasan pengguna Grab Indonesia menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM). Penelitian tersebut menunjukkan bahwa algoritma Naïve

Bayes memperoleh akurasi sebesar 92%, sedangkan SVM memperoleh akurasi sebesar 93%. Meskipun SVM memiliki akurasi sedikit lebih tinggi, Naïve Bayes tetap dinilai memiliki keunggulan dari sisi efisiensi komputasi sehingga masih layak digunakan dalam analisis sentimen terhadap layanan transportasi online. Temuan-temuan ini menunjukkan bahwa opini publik terhadap layanan transportasi online merupakan sumber informasi yang penting untuk menilai kualitas layanan secara lebih terukur (D. Budiman et al., 2024; Rifa'l et al., 2024; Yuniar & Kismiantini, 2023).

Di sisi lain, banyaknya data opini yang tersedia di media sosial menimbulkan persoalan dalam proses pengolahannya. Data yang terkumpul umumnya berupa teks tidak terstruktur, tersebar dalam jumlah besar, dan sulit dianalisis secara manual. (Yuniar & Kismiantini, 2023) menjelaskan bahwa jumlah ulasan pengguna yang terus meningkat menyebabkan perusahaan kesulitan mendapatkan informasi dari seluruh ulasan yang ada, sehingga diperlukan teknik pengolahan data dalam jumlah besar. Pada penelitian yang sama juga ditemukan bahwa ulasan Gojek memuat topik yang berulang pada sentimen negatif, sehingga informasi dari opini pengguna sebenarnya dapat dimanfaatkan sebagai dasar evaluasi layanan. Artinya, opini publik terhadap layanan transportasi online tidak cukup hanya dibaca secara manual, tetapi perlu diproses dengan pendekatan yang sistematis agar pola sentimennya dapat dipetakan secara lebih akurat (Yuniar & Kismiantini, 2023).

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan teknologi yang mampu mengolah data teks secara otomatis. Dalam hal ini, text mining dan analisis sentimen menjadi pendekatan yang relevan karena dapat digunakan untuk mengekstraksi opini, mengelompokkan polaritas sentimen, dan menghasilkan informasi yang lebih mudah dipahami. (Yuniar & Kismiantini, 2023) menyatakan bahwa analisis sentimen merupakan metode yang secara otomatis mempelajari dan memproses teks untuk memperoleh informasi sentimen yang terkandung dalam kalimat opini. Selain itu, (D. Budiman et al., 2024) menunjukkan bahwa analisis sentimen pada data Twitter dapat dilakukan melalui tahapan pengumpulan data, pelabelan, praproses, transformasi TF-IDF, klasifikasi, evaluasi model, dan visualisasi hasil. Dengan demikian, penerapan text mining dan analisis sentimen dapat membantu memetakan persepsi masyarakat terhadap layanan transportasi online secara lebih cepat, objektif, dan terstruktur (D. Budiman et al., 2024; M. A. Irawan & Aryanny, 2025; Yuniar & Kismiantini, 2023).

Selain Naïve Bayes, beberapa algoritma lain juga banyak digunakan dalam penelitian analisis sentimen, di antaranya Support Vector Machine (SVM) dan Random Forest. Support Vector Machine merupakan algoritma klasifikasi yang dikenal efektif dalam mengolah data teks karena mampu membentuk batas pemisah (hyperplane) yang optimal antar kelas sehingga sering menghasilkan tingkat akurasi yang tinggi. Sementara itu, Random Forest merupakan algoritma berbasis ensemble learning yang bekerja dengan menggabungkan banyak pohon keputusan (decision tree) untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting. (Fahmi et al., 2023) menjelaskan bahwa Random Forest memiliki performa yang baik pada

data yang memiliki banyak fitur serta mampu menghasilkan tingkat akurasi yang tinggi dalam tugas klasifikasi sentimen. Selain itu (B. Irawan & Nurdiawan, 2023) juga menerapkan Random Forest dalam analisis sentimen pengguna Gojek dan Grab pada media sosial Twitter, yang menunjukkan bahwa algoritma tersebut dapat digunakan secara efektif untuk mengidentifikasi kecenderungan opini pengguna terhadap layanan transportasi online. Meskipun SVM dan Random Forest mampu menghasilkan performa yang baik, kedua metode tersebut umumnya memiliki proses pelatihan yang lebih kompleks serta memerlukan optimasi parameter yang lebih banyak dibandingkan Naïve Bayes. Sebaliknya, Naïve Bayes memiliki proses komputasi yang lebih sederhana, efisien, dan tetap mampu menghasilkan tingkat akurasi yang kompetitif pada berbagai penelitian analisis sentimen layanan transportasi online. Oleh karena itu, penelitian ini memilih menggunakan metode Naïve Bayes karena menawarkan keseimbangan antara tingkat akurasi, efisiensi komputasi, kemudahan implementasi, serta kesesuaiannya dalam mengolah data teks yang berasal dari media sosial X.

Model SVM memperoleh akurasi sebesar 96% dan mengungguli Random Forest yang memperoleh akurasi sebesar 93%. Namun, Naïve Bayes terbukti efektif dalam pengolahan teks karena mampu menghasilkan akurasi tinggi, seperti pada penelitian (E et al., 2024) yang memperoleh akurasi 97,65%, precision 95,23%, dan recall 98,74%, sehingga model naïve bayes dapat digunakan dalam penelitian tentang layanan transportasi online di Indonesia. Layanan tersebut memiliki ketergantungan dengan opini publik yang tersebar di media sosial X. Data opini tersebut sangat bernilai untuk memahami persepsi masyarakat, tetapi sulit diolah secara manual karena bersifat tidak terstruktur dan berjumlah besar. Oleh sebab itu, penelitian mengenai analisis sentimen publik terhadap layanan transportasi online di Indonesia berbasis data media sosial X dengan menggunakan metode Naïve Bayes perlu dilakukan. Penelitian ini diharapkan mampu memberikan gambaran mengenai kecenderungan sentimen masyarakat sekaligus memperkuat penerapan text mining dalam pengolahan data opini publik.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis *text mining* dan *machine learning* untuk mengidentifikasi persepsi publik terhadap layanan transportasi daring Gojek dan Grab. Data penelitian berupa opini atau *tweet* pengguna yang diperoleh dari media sosial X (Twitter). Pengumpulan data dilakukan melalui proses *crawling* secara terpisah berdasarkan kata kunci yang berkaitan dengan masing-masing layanan, seperti "Gojek", "GoRide", "GoCar", "GoFood", "GoPay", "driver Gojek", "Gojek error", "Grab", "GrabBike", "GrabCar", "GrabFood", "GrabExpress", "driver Grab", dan kata kunci lain yang merepresentasikan pengalaman pengguna. Data dari kedua layanan kemudian digabungkan menjadi satu dataset untuk dianalisis. Pemilihan media sosial X sebagai sumber data didasarkan pada karakteristiknya yang memungkinkan pengguna

menyampaikan pengalaman, keluhan, kritik, maupun apresiasi secara langsung terhadap layanan transportasi daring.

Data yang telah dikumpulkan selanjutnya melalui tahap *preprocessing* untuk menghilangkan unsur-unsur yang tidak relevan dan menyiapkan teks agar dapat diproses oleh algoritma klasifikasi. Tahapan *preprocessing* meliputi *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*. Proses *cleaning* digunakan untuk menghapus URL, *mention*, *hashtag*, simbol, tanda baca, emoji, serta unsur lain yang tidak diperlukan, sedangkan *case folding* dilakukan dengan mengubah seluruh huruf menjadi huruf kecil. Selanjutnya, *tokenizing* digunakan untuk memecah teks menjadi satuan kata, *stopword removal* untuk menghilangkan kata umum yang kurang memberikan informasi terhadap proses klasifikasi, dan *stemming* untuk mengubah kata berimbuhan menjadi bentuk dasar. Setelah proses tersebut, data diberi label ke dalam dua kategori sentimen, yaitu positif dan negatif. Dataset kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20. Data teks yang telah dibersihkan selanjutnya ditransformasikan ke dalam bentuk numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), yang memberikan bobot terhadap kata berdasarkan frekuensi kemunculannya dalam dokumen dan tingkat kemunculannya dalam keseluruhan dataset.

Tahap selanjutnya adalah melakukan klasifikasi sentimen menggunakan algoritma Naïve Bayes. Model dibangun berdasarkan data latih untuk mempelajari pola kata yang berkaitan dengan kategori sentimen positif dan negatif, kemudian digunakan untuk memprediksi kategori sentimen pada data uji. Kinerja model dievaluasi menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Evaluasi tersebut digunakan untuk mengetahui kemampuan model dalam melakukan klasifikasi terhadap opini pengguna. Selain evaluasi numerik, hasil analisis disajikan melalui visualisasi berupa distribusi sentimen, tabel hasil klasifikasi, dan informasi performa model sehingga kecenderungan persepsi publik terhadap layanan Gojek dan Grab dapat diinterpretasikan secara lebih mudah. Implementasi penelitian dilakukan menggunakan bahasa pemrograman Python dengan dukungan pustaka Pandas untuk pengolahan data, Scikit-learn untuk proses TF-IDF dan Naïve Bayes, Sastrawi untuk *stemming* bahasa Indonesia, serta Matplotlib dan Seaborn untuk visualisasi hasil analisis.

HASIL DAN PEMBAHASAN

1. Implementasi Sistem

Pada tahap ini, rancangan sistem yang telah dijelaskan pada BAB III diimplementasikan menjadi sebuah aplikasi web menggunakan bahasa pemrograman Python dengan framework Streamlit. Implementasi sistem mengikuti lingkungan pengembangan yang telah ditentukan sebelumnya, yaitu menggunakan pustaka Pandas untuk manipulasi data, Scikit-learn untuk pembobotan TF-IDF dan klasifikasi Naïve Bayes, Sastrawi untuk proses *stemming* bahasa Indonesia, serta SQLite sebagai basis data penyimpanan.

a. Implementasi Halaman Login

Halaman Login merupakan halaman pertama yang tampil ketika sistem dijalankan. Admin diwajibkan memasukkan username dan password yang valid sebelum dapat mengakses halaman-halaman lain di dalam sistem. Apabila kombinasi username dan password yang dimasukkan benar, sistem akan mengarahkan Admin menuju halaman Dashboard. Sebaliknya, jika kombinasi yang dimasukkan salah, sistem akan menampilkan pesan kesalahan dan tetap berada pada halaman Login.



Gambar 1 Tampilan Halaman Login

b. Implementasi Halaman Dashboard

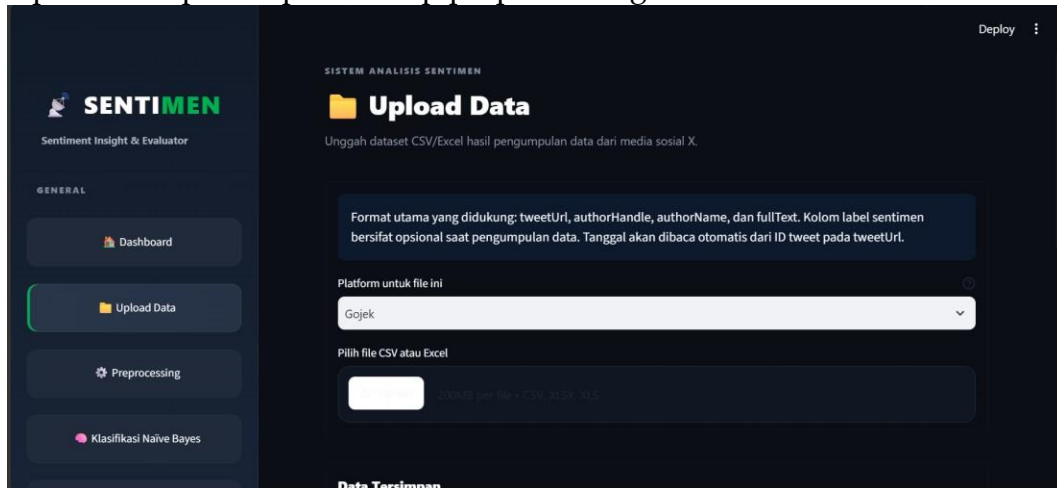
Halaman Dashboard merupakan tampilan utama yang muncul setelah Admin berhasil login. Halaman ini berfungsi sebagai pusat navigasi sistem yang menyediakan akses ke fitur-fitur utama, yaitu Upload Data, Preprocessing, Klasifikasi Naïve Bayes, dan Hasil Analisis. Dashboard tidak melakukan proses pengolahan data secara langsung, melainkan mengarahkan Admin ke tahapan analisis yang ingin dijalankan. Dengan adanya halaman ini, alur penggunaan sistem menjadi lebih terstruktur dan mudah dipahami. Pada kondisi data terkini, Dashboard menampilkan ringkasan sebanyak 1.930 total data, terdiri atas 1.015 data Gojek dan 915 data Grab, dengan 923 tweet bersentimen positif dan 1.007 tweet bersentimen negatif.



Gambar 2. Tampilan Halaman Dashboard

c. Implementasi Halaman Upload Data

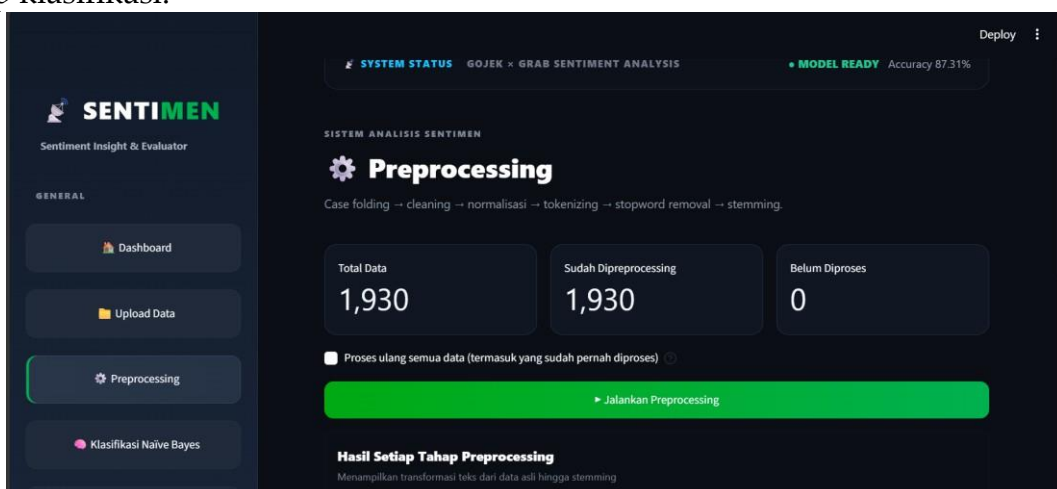
Halaman Upload Data digunakan oleh Admin untuk mengunggah dataset hasil crawling dalam format .csv. Sistem menyediakan pilihan platform (Gojek atau Grab) sebelum berkas diunggah, kemudian melakukan validasi terhadap kolom yang wajib tersedia pada berkas, seperti kolom teks tweet, tautan tweet, dan nama pengguna. Apabila validasi berhasil, data akan disimpan ke dalam tabel tweets pada basis data dan siap untuk diproses pada tahap preprocessing.



Gambar 3 Tampilan Halaman Upload Data

d. Implementasi Halaman Preprocessing

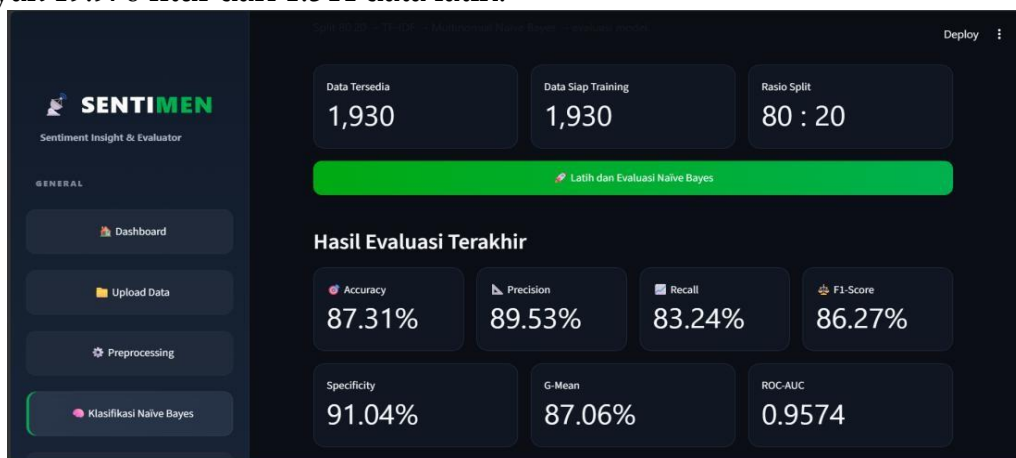
Halaman Preprocessing menjalankan tahapan pembersihan teks yang telah dirancang pada BAB III, yaitu case folding, cleaning, normalisasi kata tidak baku, tokenizing, stopword removal, dan stemming menggunakan Sastrawi. Kata negasi seperti "tidak", "bukan", dan "belum" tetap dipertahankan agar makna sentimen pada kalimat tidak berubah. Hasil dari setiap tahapan preprocessing ditampilkan dalam bentuk tabel sehingga Admin dapat memverifikasi kualitas data sebelum masuk ke tahap klasifikasi.



Gambar 4 Tampilan Halaman Preprocessing

e. Implementasi Halaman Klasifikasi Naïve Bayes

Halaman Klasifikasi Naïve Bayes merupakan halaman inti dari sistem, tempat Admin menjalankan proses pelatihan dan pengujian model. Ketika tombol "Latih dan Evaluasi Naïve Bayes" ditekan, sistem akan membagi data bersih menjadi data latih dan data uji dengan rasio 80:20 secara stratifikasi, melakukan pembobotan kata menggunakan TF-IDF hanya pada data latih untuk mencegah kebocoran data (data leakage), kemudian melatih model MultinomialNB. Model dan objek TF-IDF yang telah dilatih disimpan dalam berkas model_nb.pkl dan tfidf.pkl agar dapat digunakan kembali tanpa perlu melatih ulang setiap kali sistem dijalankan. Pada pelatihan terakhir, sistem menghasilkan kosakata (fitur kata dan pasangan kata/bigram) sebanyak 19.976 fitur dari 1.544 data latih.



Gambar 5 Tampilan Halaman Klasifikasi

f. Implementasi Halaman Hasil

Halaman Hasil menampilkan luaran akhir dari proses klasifikasi, yaitu grafik distribusi sentimen publik terhadap layanan Gojek dan Grab ke dalam kategori positif dan negatif, serta tabel data uji yang memuat teks tweet, label sentimen aktual, hasil prediksi model, probabilitas kelas positif, dan status prediksi (Benar/Salah). Tabel ini memungkinkan Admin melakukan penelusuran terhadap data yang berhasil maupun gagal diklasifikasikan dengan benar oleh model.



Gambar 6 Tampilan Halaman Hasil Analisis

2. Implementasi Basis Data

Basis data sistem diimplementasikan menggunakan SQLite dengan nama berkas database_sentimen.db, yang terdiri atas dua tabel utama. Tabel tweets menyimpan data tweet beserta seluruh tahapan preprocessing (case_folding, cleaning, normalisasi, tokenizing, stopword, text_bersih), label sentimen, pembagian data (data_split), hasil prediksi, probabilitas kelas positif, dan status prediksi. Tabel evaluasi_model menyimpan riwayat hasil evaluasi setiap kali proses pelatihan dijalankan, meliputi jumlah data, jumlah data latih dan uji, serta nilai accuracy, precision, recall, F1-score, specificity, G-Mean, AUC, dan komponen confusion matrix (TN, FP, FN, TP). Struktur ini memungkinkan sistem menampilkan hasil evaluasi terakhir pada halaman Dashboard tanpa perlu menjalankan ulang proses klasifikasi. Hingga tahap penelitian ini, tabel tweets menyimpan 1.930 baris data yang seluruhnya telah memiliki label sentimen manual.

a. Pengujian Sistem (*Black Box Testing*)

Pengujian sistem dilakukan menggunakan metode black box testing, yaitu pengujian yang berfokus pada fungsionalitas sistem tanpa memeriksa struktur kode program secara langsung. Pengujian ini bertujuan untuk memastikan setiap fitur pada sistem berjalan sesuai dengan rancangan yang telah dibuat. Hasil pengujian dapat dilihat pada Tabel dibawah ini.

Tabel 1. Hasil Pengujian Black Box

No	Skenario Pengujian	Hasil yang Diharapkan	Hasil Pengujian
1	Login dengan username dan password benar	Sistem masuk ke halaman Dashboard	Sesuai
2	Login dengan username atau password salah	Sistem menampilkan pesan kesalahan	Sesuai
3	Upload dataset .csv dengan format kolom sesuai	Data tersimpan ke tabel tweets	Sesuai
4	Upload dataset dengan kolom teks tidak ditemukan	Sistem menampilkan pesan error	Sesuai
5	Menjalankan proses preprocessing	Seluruh tahapan preprocessing tampil di tabel	Sesuai
6	Menjalankan pelatihan dan evaluasi Naïve Bayes	Model terlatih dan metrik evaluasi tampil	Sesuai
7	Menampilkan halaman Hasil setelah klasifikasi	Grafik dan tabel hasil klasifikasi tampil	Sesuai

Berdasarkan hasil pengujian black box pada Tabel 1, seluruh skenario pengujian menunjukkan hasil yang sesuai dengan hasil yang diharapkan, sehingga dapat disimpulkan bahwa sistem telah berjalan sesuai dengan fungsional yang dirancang.

b. Hasil Pengujian Model Naïve Bayes

Bagian ini menyajikan hasil pengujian performa model Naïve Bayes yang telah dilatih menggunakan dataset terbaru, mulai dari pembagian data latih dan data uji,

confusion matrix, hasil evaluasi berdasarkan metrik accuracy, precision, recall, F1-score, specificity, G-Mean, dan AUC, hingga analisis terhadap contoh hasil klasifikasi yang benar maupun keliru, serta perhitungan manual sebagai verifikasi cara kerja model.

1) Pembagian Data Latih dan Data Uji

Dataset yang digunakan dalam pengujian model berjumlah 1.930 tweet yang telah melalui tahap pelabelan manual, terdiri atas 1.015 tweet layanan Gojek dan 915 tweet layanan Grab, dengan komposisi 923 sentimen positif (47,8%) dan 1.007 sentimen negatif (52,2%). Dataset kemudian dibagi menggunakan rasio 80:20 secara stratifikasi, sehingga proporsi kelas positif dan negatif tetap seimbang baik pada data latih maupun data uji.

Tabel 2. Pembagian Data Latih dan Data Uji

Keterangan	Jumlah
Total data	1.930 tweet
Data latih (80%)	1.544 tweet
Data uji (20%)	386 tweet
Data uji - positif	185 tweet
Data uji - negatif	201 tweet

2) Confusion Matrix

Hasil prediksi model Naïve Bayes terhadap 386 data uji dibandingkan dengan label aktual untuk membentuk confusion matrix, sebagaimana ditampilkan pada Tabel 3.

Tabel 3. Confusion Matrix Hasil Klasifikasi

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	TN = 183	FP = 18
Aktual Positif	FN = 31	TP = 154

Berdasarkan Tabel 3, model berhasil mengklasifikasikan 183 data negatif dan 154 data positif secara tepat (*True Negative* dan *True Positive*), sedangkan 18 data negatif diprediksi salah sebagai positif (*False Positive*) dan 31 data positif diprediksi salah sebagai negatif (*False Negative*).

3) Hasil Evaluasi Model

Berdasarkan confusion matrix pada Tabel 3, dihitung nilai accuracy, precision, recall, F1-score, specificity, G-Mean, dan AUC menggunakan rumus yang telah dijelaskan. Hasil evaluasi model secara lengkap ditampilkan pada Tabel 4.

Tabel 4. Hasil Evaluasi Model Naïve Bayes

Metrik	Nilai
Accuracy	87,31%
Precision	89,53%

Recall (Sensitivity)	83,24%
F1-Score	86,27%
Specificity	91,04%
G-Mean	87,06%
ROC-AUC	0,9574

Nilai accuracy sebesar 87,31% menunjukkan bahwa model Naïve Bayes mampu mengklasifikasikan sentimen dengan tepat pada sebagian besar data uji, sekaligus telah melampaui target akurasi minimal 80% yang ditetapkan dalam penelitian ini. Nilai precision 89,53% menunjukkan bahwa dari seluruh tweet yang diprediksi positif, sebagian besar memang benar-benar bersentimen positif. Nilai recall 83,24% menunjukkan kemampuan model mengenali tweet bersentimen positif secara aktual, sedangkan nilai specificity 91,04% menunjukkan kemampuan model mengenali tweet bersentimen negatif secara aktual. Selisih yang relatif kecil antara recall dan specificity, yang tercermin pula pada nilai G-Mean sebesar 87,06%, mengindikasikan bahwa performa model cukup seimbang dalam mengenali kedua kelas sentimen dan tidak condong berlebihan ke salah satu kelas, meskipun model sedikit lebih baik dalam mengenali tweet bersentimen negatif dibanding positif. Nilai ROC-AUC sebesar 0,9574 berada pada rentang 0,9-1,0 yang menurut kategorisasi pada BAB III tergolong "sempurna", yang berarti model memiliki kemampuan pemisahan (separability) antara kelas positif dan negatif yang sangat baik.

4) Analisis Hasil Klasifikasi

Untuk memberikan gambaran lebih rinci mengenai kinerja model, berikut disajikan beberapa contoh hasil klasifikasi yang benar dan yang keliru pada Tabel 5.

Tabel 5. Contoh Hasil Klasifikasi Data Uji

Cuplikan Tweet	Label Aktual	Prediksi	Status
"...gilaa cepet bgt drivernya dtg, blm ada 3 menit udh nyampe depan rumah..."	Positif	Positif	Benar
"...kok bs y saldo kepake buat trip org lain, aneh bgt sistem billing nya..."	Negatif	Negatif	Benar
"Keajaiban dunia dpt gojek cepet jam segini didaerah plosok ini."	Positif	Negatif	Salah
"Pertanyaan pertama ketika bapak gojek jemput aku, ga ada babibu langsung 'tinggi nya berapa to mbak?'"	Negatif	Positif	Salah

Berdasarkan Tabel 5, kesalahan klasifikasi umumnya terjadi pada tweet yang bermakna implisit atau menggunakan gaya bahasa tidak langsung tanpa kata kunci sentimen yang eksplisit, seperti "Keajaiban dunia dpt gojek cepet jam segini didaerah plosok ini" yang sebenarnya bersentimen positif namun tidak memuat kata-kata pujian yang umum dipelajari model (misalnya "bagus", "keren", atau "puas"). Kesalahan lain terjadi pada tweet yang bermuatan sindiran atau kekesalan implisit tanpa kata negatif eksplisit, seperti tweet mengenai pertanyaan pribadi dari pengemudi Gojek, sehingga model cenderung salah menafsirkan bobot kata secara keseluruhan. Pola ini menunjukkan bahwa Naïve Bayes, yang mengasumsikan independensi antar kata, memiliki keterbatasan dalam menangkap konteks kalimat secara menyeluruh, khususnya pada opini yang bersifat implisit maupun mengandung sarkasme.

5) Perhitungan Manual

Untuk memverifikasi bahwa proses klasifikasi pada sistem telah berjalan sesuai dengan konsep TF-IDF dan Naïve Bayes yang dijelaskan pada BAB III, berikut disajikan contoh perhitungan manual menggunakan sebagian kecil data asli dari dataset penelitian terbaru. Perhitungan tidak dilakukan terhadap keseluruhan 1.544 data latih karena akan menghasilkan ribuan kombinasi kata yang tidak mungkin ditelusuri secara manual, sehingga digunakan 6 dokumen latih (3 data positif dan 3 data negatif) beserta 1 dokumen uji sebagai ilustrasi tahapan perhitungan. Keenam dokumen latih tersebut adalah D1 "nasehat bagus bapak gojek hari" (Positif), D2 "driver gojek ngasih gw bintang" (Positif), D3 "driver gojek wangi anugrah" (Positif), D4 "gojek lg error" (Negatif), D5 "kelaperan eh malah gojek error" (Negatif), dan D6 "gojek susah bget dapetnyaa" (Negatif), sedangkan dokumen uji Dt adalah "gojek error anj" dengan label aktual Negatif.

a) Tahap Pembobotan TF-IDF (Persamaan 6)

Dari 6 dokumen latih terbentuk kosakata (vocabulary) sebanyak 19 kata unik. Dokumen uji Dt terdiri atas 3 kata, yaitu "gojek", "error", dan "anj". Kata "anj" tidak pernah muncul pada data latih sehingga tidak memiliki kolom bobot pada model (out-of-vocabulary) dan diabaikan pada saat prediksi. Dengan demikian, hanya kata "gojek" dan "error" yang berperan dalam perhitungan.

$$idf("gojek") = \ln\left(\frac{D}{df}\right) = \ln\left(\frac{6}{6}\right) = \ln(1) = 0,0000$$

$$idf("error") = \ln\left(\frac{D}{df}\right) = \ln\left(\frac{6}{2}\right) = \ln(3) = 1,0986$$

Kata "gojek" memperoleh bobot idf sebesar 0 karena muncul di seluruh dokumen, sehingga dianggap tidak informatif. Sebaliknya, kata "error" hanya muncul pada 2 dari 6 dokumen (keduanya berkelas negatif), sehingga memperoleh bobot idf yang lebih tinggi dan lebih diskriminatif terhadap kelas negatif.

b) Tahap Klasifikasi Naïve Bayes (Persamaan 7)

Probabilitas prior masing-masing kelas dihitung dari proporsi dokumen latih:

$$P(\text{Positif}) = 3/6 = 0.5 ; P(\text{Negatif}) = 3/6 = 0.5$$

Likelihood tiap kata dihitung menggunakan penghalusan Laplace terhadap total bobot TF-IDF kata pada tiap kelas. Total bobot semua kata pada kelas Positif adalah 18,3231 dan pada kelas Negatif adalah 14,7395. Kata "gojek" memiliki total bobot 0,0000 pada kedua kelas, sehingga diperoleh $P(\text{gojek}|\text{Positif}) = 0.0268$ dan $P(\text{gojek}|\text{Negatif}) = 0.0296$. Kata "error" memiliki total bobot 0,0000 pada kelas Positif sehingga $P(\text{error}|\text{Positif}) = 0.0268$, dan sebesar 2,1972 pada kelas Negatif sehingga $P(\text{error}|\text{Negatif}) = 0.0948$. Nilai $P(\text{kata} | \text{Kelas})$ diperoleh dari rumus $P(w|C) =$

$$\frac{\sum \text{bobot } w \text{ pada kelas } C + 1}{\text{total bobot semua kata pada kelas } C + V}$$

Skor akhir tiap kelas kemudian dihitung dengan mengalikan prior dengan likelihood seluruh kata yang dikenali pada dokumen uji (yaitu "gojek" dan "error"):

$$\text{Skor}(\text{Positif}) = P(\text{Positif}) \times P(\text{gojek}|\text{Positif}) \times P(\text{error}|\text{Positif}) = 0,5 \times 0,0268 \times 0,0268 = 0,000359$$

$$\text{Skor}(\text{Negatif}) = P(\text{Negatif}) \times P(\text{gojek}|\text{Negatif}) \times P(\text{error}|\text{Negatif}) = 0,5 \times 0,0296 \times 0,0948 = 0,001404$$

Kedua skor kemudian dinormalisasi agar berjumlah 1 sehingga diperoleh probabilitas akhir tiap kelas:

$$P(\text{Positif}|d) = \frac{0.000359}{0.000359 + 0.001404} = 0,2036 \text{ (20,36\%)}$$

$$P(\text{Negatif}|d) = \frac{0.001404}{0.000359 + 0.001404} = 0,7964 \text{ (79,64\%)}$$

Karena $P(\text{Negatif}|d)$ lebih besar daripada $P(\text{Positif}|d)$, dokumen uji "gojek error anj" diklasifikasikan sebagai NEGATIF melalui perhitungan manual.

Hasil ini konsisten dengan keluaran sistem untuk dokumen data uji yang sama: sistem memprediksi kelas negatif dengan probabilitas positif sebesar 0,0856, sesuai dengan label aktual (negatif), sehingga berstatus "Benar". Perbedaan nilai probabilitas antara perhitungan manual (20,36%) dan sistem (8,56%) merupakan hal yang wajar, karena sistem menghitung TF-IDF menggunakan seluruh 1.544 dokumen data latih dengan 19.976 fitur kata dan pasangan kata (bigram), serta menerapkan smoothing idf dan normalisasi L2 bawaan pustaka Scikit-learn, sedangkan perhitungan manual pada bagian ini hanya menggunakan 6 dokumen latih sebagai ilustrasi sederhana agar setiap langkah perhitungan dapat ditelusuri. Meskipun demikian, kedua pendekatan menghasilkan kesimpulan klasifikasi akhir yang sama, sehingga dapat disimpulkan bahwa implementasi sistem telah berjalan sesuai dengan prinsip metode TF-IDF dan Naïve Bayes yang digunakan dalam penelitian ini.

Selain memverifikasi proses klasifikasi, dilakukan pula perhitungan manual untuk membuktikan validitas nilai accuracy, precision, recall, F1-score, specificity, dan G-Mean yang dihasilkan oleh sistem sebagaimana ditampilkan pada Tabel 4. Perhitungan dilakukan menggunakan rumus

standar evaluasi klasifikasi terhadap nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) yang diperoleh dari confusion matrix pada Tabel 4.3, yaitu TP = 154, TN = 183, FP = 18, dan FN = 31, dengan total keseluruhan 386 data uji.

c) Akurasi (Accuracy)

Akurasi dihitung dengan membagi jumlah prediksi yang benar (TP + TN) dengan keseluruhan data uji.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{154 + 183}{386} = \frac{337}{386} = 0,8731$$

Hasil perhitungan manual menunjukkan nilai akurasi sebesar 0,8731 (87,31%), yang sama persis dengan nilai accuracy pada Tabel 4.4.

d) Precision

Precision dihitung dari proporsi prediksi positif yang benar-benar merupakan data bersentimen positif.

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{154}{154 + 18} = \frac{154}{172} = 0,8953$$

Nilai precision sebesar 0,8953 (89,53%) ini sesuai dengan nilai pada Tabel 4.4, yang menunjukkan bahwa dari seluruh tweet yang diprediksi positif oleh model, sebagian besar memang benar-benar bersentimen positif.

e) Recall (Sensitivity)

Recall dihitung dari proporsi data bersentimen positif secara aktual yang berhasil dikenali dengan benar oleh model.

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{154}{154 + 31} = \frac{154}{185} = 0,8324$$

Nilai recall sebesar 0,8324 (83,24%) ini sama dengan nilai pada Tabel 4.4, sehingga membuktikan kemampuan model dalam mengenali data bersentimen positif secara actual.

f) F1-Score

F1-score dihitung sebagai rata-rata harmonik antara precision dan recall.

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times 0,8953 \times 0,8324}{0,8953 + 0,8324} = \frac{1,4906}{1,7277} = 0,8627$$

Hasil perhitungan manual menunjukkan nilai F1-score sebesar 0,8627 (86,27%), yang sama persis dengan nilai pada Tabel 4.4.

g) Specificity

Specificity dihitung dari proporsi data bersentimen negatif secara aktual yang berhasil dikenali dengan benar oleh model.

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{183}{183 + 18} = \frac{183}{201} = 0,9104$$

Nilai specificity sebesar 0,9104 (91,04%) ini sesuai dengan nilai pada Tabel 4.4, yang menunjukkan kemampuan model mengenali data bersentimen negatif secara actual.

h) G-Mean

G-Mean dihitung sebagai akar kuadrat dari hasil kali recall dan specificity, untuk menggambarkan keseimbangan performa model dalam mengenali kedua kelas sentiment.

$$G\text{-Mean} = \sqrt{(\text{Recall} \times \text{Specificity})} = \sqrt{\left(\frac{154}{185} \times \frac{183}{201}\right)} = \sqrt{(0.8324 \times 0.9104)} = \sqrt{(0.7579)} = 0.8706$$

Hasil perhitungan manual menunjukkan nilai G-Mean sebesar 0,8706 (87,06%), yang sama dengan nilai pada Tabel 4.

Nilai ROC-AUC sebesar 0,9574 pada Tabel 4.4 tidak diverifikasi melalui perhitungan manual pada bagian ini, karena perhitungan AUC memerlukan nilai skor probabilitas prediksi pada seluruh data uji di berbagai ambang batas (threshold), bukan hanya nilai TP, TN, FP, dan FN pada satu ambang batas sebagaimana digunakan pada confusion matrix Tabel 4.3. Oleh karena itu, nilai ROC-AUC sepenuhnya dihitung oleh sistem menggunakan fungsi `roc_auc_score` pada pustaka Scikit-learn.

Secara keseluruhan, hasil perhitungan manual pada bagian ini — baik terhadap proses klasifikasi Naïve Bayes maupun terhadap accuracy, precision, recall, F1-score, specificity, dan G-Mean — menghasilkan nilai yang sama atau konsisten dengan nilai yang dihasilkan oleh sistem. Hal ini membuktikan bahwa implementasi klasifikasi maupun perhitungan metrik evaluasi pada sistem telah dilakukan dengan benar dan dapat dipertanggungjawabkan.

Pembahasan

Hasil accuracy sebesar 87,31% yang diperoleh dalam penelitian ini menunjukkan performa yang kompetitif apabila dibandingkan dengan beberapa penelitian terdahulu yang telah diuraikan pada BAB I. Nilai ini berada di atas hasil penelitian (Rifa'I et al., 2024) yang memperoleh akurasi 84,36% pada data yang sangat tidak seimbang (1.677 positif berbanding 69 negatif), namun sedikit lebih rendah dibandingkan penelitian (Yuniar & Kismiantini, 2023) sebesar 89,9% dan (Nurfauzi et al., 2025) sebesar 92%. Perbedaan ini wajar terjadi mengingat sumber data, jumlah data, dan karakteristik bahasa pada masing-masing penelitian berbeda. Data pada penelitian ini berasal dari media sosial X yang cenderung memuat bahasa informal, singkatan, dan gaya penulisan yang lebih beragam dibandingkan ulasan pada Google Play Store, sehingga tingkat kesulitan klasifikasinya relatif lebih tinggi. Meskipun demikian, akurasi yang diperoleh tetap berada pada kategori baik dan telah memenuhi target penelitian, serta mengalami peningkatan dibandingkan hasil pengujian pada tahap sebelumnya seiring bertambahnya jumlah data latih dari 870 menjadi 1.544 tweet.

Berdasarkan hasil pengujian, Pertama, text mining dan metode Naïve Bayes berhasil diterapkan untuk menganalisis sentimen publik terhadap layanan Gojek dan Grab melalui tahapan pengumpulan data, preprocessing, pelabelan, pembobotan TF-IDF, klasifikasi, dan evaluasi model yang terintegrasi dalam satu sistem berbasis web. Kedua, performa metode Naïve Bayes dalam mengklasifikasikan sentimen menunjukkan hasil yang baik, dengan accuracy 87,31%, precision 89,53%, recall

83,24%, dan F1-score 86,27%. Ketiga, hasil pemetaan sentimen publik terhadap layanan transportasi online menunjukkan komposisi 923 tweet (47,8%) bersentimen positif dan 1.007 tweet (52,2%) bersentimen negatif dari total 1.930 tweet yang dianalisis, yang mengindikasikan bahwa opini masyarakat terhadap layanan Gojek dan Grab masih cenderung diwarnai keluhan dalam proporsi yang sedikit lebih besar dibandingkan apresiasi.

KESIMPULAN

Berdasarkan hasil penelitian bahwa *text mining* dan metode Naïve Bayes berhasil diterapkan untuk menganalisis sentimen publik terhadap layanan transportasi online di Indonesia, khususnya Gojek dan Grab, melalui sistem berbasis web yang mengintegrasikan tahapan pengumpulan data, *preprocessing*, pelabelan, pembobotan TF-IDF, klasifikasi, dan evaluasi model secara berurutan. Model Naïve Bayes menunjukkan performa yang baik dalam mengklasifikasikan sentimen terhadap 1.930 *tweet*, dengan nilai *accuracy* sebesar 87,31%, *precision* 89,53%, *recall* 83,24%, *F1-score* 86,27%, *specificity* 91,04%, *G-Mean* 87,06%, dan *ROC-AUC* 0,9574. Hasil tersebut menunjukkan bahwa model mampu mengidentifikasi sentimen positif dan negatif dengan kinerja yang baik serta telah melampaui target akurasi minimal 80% yang ditetapkan dalam penelitian. Pemetaan sentimen menunjukkan bahwa dari 1.930 *tweet*, sebanyak 923 *tweet* (47,8%) memiliki sentimen positif dan 1.007 *tweet* (52,2%) memiliki sentimen negatif. Komposisi tersebut menunjukkan bahwa opini publik terhadap layanan Gojek dan Grab masih cenderung didominasi oleh sentimen negatif, meskipun selisihnya relatif kecil dibandingkan dengan sentimen positif.

DAFTAR PUSTAKA

- Ahmad Rifa'i, A., Ardhani, R., Pratama, D., & Fatihanursari. (2024). Analisis sentimen terhadap layanan aplikasi Grab Indonesia menggunakan metode Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 303–313.
- Budiman, D., et al. (2024). Analisis sentimen aplikasi Gojek pada Twitter menggunakan algoritma Naïve Bayes. *JAMASTIKA*, 3(2), 107–116.
- Fahmi, M., Yuningsih, Y., & Puspita, A. (2023). Analisis sentimen ulasan aplikasi Gojek menggunakan Naïve Bayes. *Jurnal Teknik Informatika & Sistem Informasi*, 4(2), 50–60.
- Heristian, S., Napih, M., & Erawati, W. (2025). Analisis sentimen ulasan pelanggan menggunakan algoritma Naïve Bayes pada aplikasi Gojek. *Computer Science (CO-SCIENCE)*, 5(1), 35–44.
- Irawan, B., & Nurdiawan, D. (2023). Analisis sentimen pengguna Gojek dan Grab pada media sosial Twitter menggunakan Random Forest. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(5), 3614–3619.
- Islami, A. I., & Choerunnisa, A. (2025). Digital disruption in transportation: Gojek's entry and growth in Makassar (2015–2020). *Yupa: Historical Studies Journal*, 9(1), 60–71.

- Khoiruddin, Yulianto, & Santoso. (2023). Analisis sentimen ulasan aplikasi Gojek menggunakan SVM dan Random Forest. *Jurnal Teknologi dan Sistem Informasi Bisnis*, 4(1), 1–20.
- L., M., Febriani, F. W., Tan, R., & Ayuni, M. (2023). Manajemen strategik bisnis PT Gojek Indonesia. *Journal of Business Finance and Economic (JBFE)*, 4(1), 119–132.
- Petiwi, M. I., Triayudi, A., & Sholihati, I. D. (2022). Analisis sentimen GoFood berdasarkan Twitter menggunakan metode Naïve Bayes dan Support Vector Machine. *Jurnal Media Informatika Budidarma*, 6(1), 542–550.
- Nurfauzi, et al. (2025). Analisis sentimen Grab Indonesia pada ulasan Google Play Store menggunakan algoritma Naïve Bayes dan Support Vector Machine. *SMARTICS Journal*.
- Patiger Satria, Harfianti, & Dayanti, K. D. (2024). Analisis pengaruh kualitas layanan dan harga terhadap kepuasan pelanggan pada platform transportasi online di Indonesia. *Jurnal Ilmiah Mahasiswa Merdeka EMBA*, 3(2), 451–459.
- Yuniar, P., & Kismiantini. (2023). Analisis sentimen ulasan Gojek menggunakan Naïve Bayes Classifier. *Statistika*, 23(2), 164–175.
- Saputra, A. D., Budiman, D. P., Reynanda, R. M., & Sari, A. P. (2024). Analisis sentimen aplikasi Gojek pada Twitter menggunakan algoritma Naïve Bayes. *JAMASTIKA*, 3(2), 107–116.
- Yutika, C. H., Adiwijaya, & Faraby, S. A. (2021). Analisis sentimen berbasis aspek pada review Female Daily menggunakan TF-IDF dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 422–430.