VOL. 03. NO. 01 JANUARI 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI QUORA MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES DAN TF-IDF

Nur Farisah Patimah[1], Rudi Kurniawan[2], Bani Nurhakim[3], Aris Pratama Putra[4], Saeful Anwar[5].

Program Studi Teknik Informatika [15]
Program Studi Rekayasa Perangkat Lunak[2]
Program Studi Manajemen Informatika[34]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
nurfarisah149@gmail.com

(*) Corresponding Author : nurfarisah149@gmail.com;
Published : 30 Januari 2026

Abstract— *This study aims to analyze the sentiment of user reviews of the Quora application by developing a classification model using the Term Frequency–Inverse Document Frequency (TF-IDF) method and the Multinomial Naïve Bayes algorithm. A total of 2,000 reviews were collected through web scraping from the Google Play Store, followed by manual filtering and labeling, resulting in 1,631 reviews suitable for analysis. The preprocessing stage included text cleaning, case folding, normalization, tokenization, stopword removal, and stemming to ensure data consistency. The processed data were then divided into training and testing sets before undergoing feature extraction using TF-IDF. The Multinomial Naïve Bayes model was trained on the labeled training data and evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The results show that the model achieved an accuracy of 80%, with the highest performance in the positive class (F1-score 0.89), moderate performance in the negative class (F1-score 0.62), and the lowest performance in the neutral class (F1-score 0.47). These findings indicate that the combination of TF-IDF and Multinomial Naïve Bayes is effective for classifying sentiment in Indonesian-language application reviews, although neutral sentiment remains challenging due to its ambiguous nature. This study is expected to contribute to the development of automated sentiment analysis systems for continuous monitoring of application service quality.*

Keywords: *Sentiment Analysis, Multinomial Naïve Bayes, TF-IDF, Quora, User Review.*

Abstrak— Penelitian ini bertujuan menganalisis sentimen pada ulasan pengguna aplikasi Quora dengan membangun model klasifikasi menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) dan algoritma *Multinomial Naïve Bayes*. Sebanyak 2.000 ulasan dikumpulkan melalui *web scraping* dari Google Play Store, kemudian diseleksi dan diberi label secara manual hingga menghasilkan 1.631 ulasan yang layak dianalisis. Tahap *preprocessing* meliputi pembersihan teks, *case folding*, normalisasi, tokenisasi, penghapusan *stopword*, dan *stemming* untuk menjaga konsistensi data. Data yang telah diproses kemudian dibagi menjadi data latih dan data uji sebelum melalui ekstraksi fitur menggunakan TF-IDF. Model *Multinomial Naïve Bayes* dilatih pada data latih dan dievaluasi menggunakan akurasi, *precision*, *recall*, F1-score, dan *confusion matrix*. Hasil evaluasi menunjukkan bahwa model mencapai akurasi 80%, dengan performa tertinggi pada kelas positif (F1-score 0.89), performa sedang pada kelas negatif (F1-score 0.62), dan performa terendah pada kelas netral (F1-score 0.47). Temuan ini menunjukkan bahwa kombinasi TF-IDF dan *Multinomial Naïve Bayes* efektif untuk klasifikasi sentimen ulasan berbahasa Indonesia, meskipun sentimen netral masih menjadi tantangan karena sifatnya yang ambigu. Penelitian ini diharapkan dapat mendukung pengembangan sistem analisis sentimen otomatis untuk pemantauan kualitas layanan aplikasi secara berkelanjutan.

Kata Kunci: Analisis Sentimen, Multinomial Naïve Bayes, TF-IDF, Quora, Ulasan Pengguna.

INTRODUCTION

Perkembangan teknologi informasi dalam sepuluh tahun terakhir telah mengubah cara manusia berkomunikasi, bekerja, dan mencari informasi. Kemajuan di bidang *Artificial Intelligence*, *Machine Learning*, dan *Big Data Analytics* memberi

efisiensi pada berbagai proses serta membuka peluang untuk memahami perilaku pengguna melalui data yang tersedia dalam jumlah besar [1]. Salah satu sumber data yang penting dalam menilai kualitas layanan digital adalah ulasan pengguna. Ulasan tersebut menggambarkan persepsi, pengalaman, dan emosi pengguna terhadap suatu aplikasi, sehingga dapat digunakan untuk meningkatkan kualitas layanan dan mendukung pengambilan keputusan berbasis data [2]. Namun ulasan pengguna sering berjumlah besar, tidak terstruktur, dan menggunakan bahasa informal sehingga sulit dianalisis secara manual.

Pada aplikasi komunitas seperti Quora, pengguna menyampaikan pertanyaan, informasi, dan opini yang beragam. Di sisi lain, ulasan Quora pada Google Play Store memberikan gambaran nyata tentang kualitas aplikasi, mulai dari kemudahan penggunaan, stabilitas, keamanan data, hingga fitur komunitas. Jumlah ulasan yang banyak dan tidak terstruktur membuat analisis manual kurang efektif dan berisiko menimbulkan bias. Kondisi ini menjadikan analisis sentimen berbasis pembelajaran mesin sebagai pendekatan yang tepat untuk membaca pola opini pengguna secara otomatis [3], [4].

Algoritma *Naïve Bayes*, terutama varian *Multinomial Naïve Bayes*, menjadi salah satu metode yang paling sering digunakan untuk analisis sentimen teks berbahasa Indonesia. Banyak penelitian yang mengombinasikan TF-IDF sebagai metode ekstraksi fitur menunjukkan bahwa model ini mampu mengklasifikasikan teks pendek secara akurat dan efisien [5]. Penelitian pada ulasan aplikasi Starlink menunjukkan bahwa model ini mampu memetakan sentimen publik secara efektif, sedangkan penelitian pada aplikasi BRImo membuktikan kemampuannya dalam mengenali pola opini nasabah layanan perbankan digital [6]. Studi pada aplikasi Digitalent Mobile dan Flip juga menunjukkan bahwa kombinasi TF-IDF dan *Naïve Bayes* menghasilkan klasifikasi yang stabil meskipun data tidak homogen [7].

Metode ini juga terbukti efektif pada domain lain. Penelitian pada ulasan iPusnas mampu mengidentifikasi opini pengguna layanan perpustakaan digital dengan baik [8]. Penelitian pada aplikasi Female Daily menunjukkan bahwa TF-IDF dan *Naïve Bayes* dapat digunakan untuk *aspect-based sentiment analysis* sehingga membantu menilai opini pengguna pada fitur tertentu. Tinjauan sistematis mengenai penelitian analisis sentimen di Indonesia juga menyatakan bahwa *Multinomial Naïve Bayes* banyak dipilih karena performanya stabil dan mudah diterapkan [9].

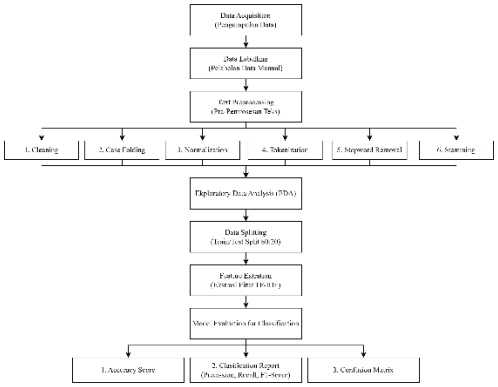
Penelitian pada bidang hiburan dan gaya hidup semakin menguatkan bukti efektivitas metode ini. Analisis ulasan game eFootball 2024 menunjukkan kecenderungan sentimen positif, sedangkan analisis ulasan makanan manis di platform X menunjukkan pola sentimen berdasarkan preferensi pengguna [8], [9]. Penelitian lain yang membandingkan ulasan berbahasa Inggris dan Indonesia memperlihatkan bahwa *Multinomial Naïve Bayes* tetap unggul ketika digabungkan dengan TF-IDF dan *n-grams* [10]. Pada aplikasi e-commerce dan layanan keuangan, metode ini mampu menangani teks panjang maupun pendek, serta bahasa informal, sehingga mendukung proses klasifikasi skala besar [11].

Beberapa penelitian juga menyoroti tantangan pada model *Naïve Bayes*, terutama masalah ketidakseimbangan kelas dan kesulitan dalam mengklasifikasikan sentimen netral. Penelitian pada Tokopedia, BRImo, dan MyPertamina menemukan bahwa sentimen netral sering salah diklasifikasikan karena jumlahnya lebih sedikit dan memiliki ekspresi yang minim [12]. Struktur bahasa Indonesia yang mengandung negasi atau sarkasme juga dapat menurunkan akurasi model jika tidak diolah dengan teknik *preprocessing* yang tepat[13]. Meskipun terdapat tantangan tersebut, berbagai studi tetap menyimpulkan bahwa *Naïve Bayes* merupakan metode yang kompetitif untuk analisis sentimen, terutama pada data ulasan yang ringkas dan padat informasi [14], [15].

Walaupun analisis sentimen pada berbagai aplikasi populer telah banyak dilakukan, belum ada penelitian yang secara khusus membahas ulasan pengguna aplikasi Quora di Google Play Store menggunakan TF-IDF dan *Multinomial Naïve Bayes*. Padahal ulasan pada aplikasi Quora bersifat ringkas, beragam, dan informatif sehingga sesuai untuk dianalisis dengan pendekatan ini. Pemahaman mengenai persepsi pengguna aplikasi Quora juga berpotensi memberi wawasan penting bagi pengembang dalam meningkatkan kualitas layanan. Berdasarkan kesenjangan tersebut, penelitian ini bertujuan membangun dan mengevaluasi model klasifikasi sentimen pada ulasan pengguna aplikasi Quora di Google Play Store dengan menggunakan TF-IDF dan *Multinomial Naïve Bayes*. Penelitian ini juga bertujuan mengidentifikasi kecenderungan opini pengguna, memetakan distribusi sentimen, serta menyediakan masukan berbasis data untuk pengembangan layanan aplikasi Quora di masa mendatang.

MATERIALS AND METHODS

Penelitian ini menerapkan metode kuantitatif dengan desain eksperimen komputasional menggunakan pendekatan *Knowledge Discovery in Database* (KDD) untuk menganalisis sentimen ulasan pengguna secara sistematis. Pendekatan ini dipilih karena mampu mengolah data teks tidak terstruktur menjadi informasi yang terukur melalui tahapan yang terintegrasi, mulai dari pengumpulan data via *web scraping*, pelabelan sentimen yang divalidasi ahli, pra-pemrosesan teks, ekstraksi fitur TF-IDF, hingga klasifikasi menggunakan algoritma *Multinomial Naïve Bayes* dan evaluasi model. Adapun tahapan penelitian secara rinci dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

A. Data Acquisition

Data dikumpulkan melalui web scraping ulasan aplikasi Quora di Google Play Store menggunakan *library* Python. Data mentah berupa teks ulasan, nama pengguna, rating, dan waktu unggah.

B. Data Labelling

Data yang terkumpul diberi label sentimen secara manual ke dalam tiga kategori, yaitu positif, negatif, dan netral. Pelabelan dilakukan dengan membaca setiap ulasan untuk memastikan kesesuaian makna dan konteks.

C. Text Preprocessing

Text preprocessing dilakukan untuk membersihkan dan menstandarkan teks dalam dataset. Prosesnya dimulai dengan *cleaning* untuk menghapus emoji, angka, tautan, dan simbol yang tidak diperlukan, dilanjutkan dengan *case folding* agar seluruh teks menjadi huruf kecil. Setelah itu dilakukan *normalization* untuk menyamakan kata tidak baku atau slang, kemudian *tokenization* untuk memecah kalimat menjadi kata. Tahap berikutnya adalah *stopword removal* guna menghilangkan kata umum yang tidak relevan, serta *stemming* menggunakan Sastrawi untuk mengembalikan kata ke bentuk dasarnya.

D. Exploratory Data Analyst (EDA)

Memahami karakteristik awal data ulasan sebelum masuk ke proses pemodelan. Analisis ini mencakup pemeriksaan distribusi rating, pola sentimen, serta frekuensi kemunculan kata dalam ulasan. EDA membantu mengidentifikasi kondisi data, seperti ketidakseimbangan jumlah ulasan pada setiap kategori sentimen, keberadaan kata yang dominan, serta potensi noise yang dapat memengaruhi kualitas model.

E. Data Splitting

Membagi dataset menjadi dua bagian utama, yaitu data latih dan data uji. Pembagian ini bertujuan untuk memastikan bahwa model dapat dilatih menggunakan sebagian data, sementara performanya dievaluasi menggunakan data yang tidak pernah dilihat sebelumnya. Pada penelitian ini, dataset dibagi dengan proporsi umum 80% sebagai data latih dan 20% sebagai data uji. Data latih digunakan untuk membangun model Multinomial Naïve Bayes, sedangkan data uji digunakan untuk mengukur kemampuan generalisasi model melalui metrik evaluasi seperti akurasi, presisi, recall, dan F1-score.

F. Feature Extration (TF-IDF)

Mengonversi teks ulasan ke dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (*TF-IDF*). Berfungsi memberikan bobot pada setiap kata berdasarkan seberapa sering kata tersebut muncul dalam suatu ulasan. Menonjolkan kata-kata yang relevan serta mengurangi pengaruh kata yang terlalu umum.

G. Model Evaluation for Classification

Evaluasi performa model klasifikasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai kemampuan model dalam mengidentifikasi sentimen secara tepat. *Accuracy* mengukur rasio prediksi benar terhadap seluruh sampel uji, sementara *precision* menilai ketepatan prediksi pada tiap kelas. *Recall* digunakan untuk melihat sejauh mana model mampu menangkap seluruh sampel yang benar dalam suatu kelas. *F1-score*, sebagai rata-rata harmonis antara *precision* dan

recall, memberikan penilaian yang lebih stabil terutama pada data yang tidak seimbang. Analisis tambahan melalui *confusion matrix* digunakan untuk memeriksa pola kesalahan prediksi pada masing-masing label sentimen, sehingga performa model dapat dievaluasi secara lebih komprehensif.

RESULTS AND DISCUSSION

A. Hasil Data Acquisition

Pada tahap pengambilan data, terkumpul 2.000 entri ulasan pengguna aplikasi Quora dari Google Play Store melalui proses *web scraping*. Setiap entri berisi lima atribut utama, yaitu *Review ID*, *Username*, *Rating*, *Review Text*, dan *Date*. Seluruh data mentah yang diperoleh kemudian disimpan dalam berkas *hasil_scraper_ulasan_app_Quora.csv* agar dapat diproses lebih lanjut pada tahap *preprocessing* dan analisis berikutnya.

Tabel 1. Hasil Data Acquisition

Review ID	Username	Rating	Review Text	Date
a95a03ed-327b-4b50-940c-616bca947988	Pengguna Google	5	bagus banget	2025-11-03
42da2b3f-6050-4bde-b894-8df36813fc69	Pengguna Google	5	aplikasi bagus	2025-11-01
7889b9fd-4879-4e74-a3a5-96335e1e185e	Pengguna Google	2	aplikasi ga bisa dibuka	2025-10-31
0a2eef4f-e9c9-44a9-9820-3b5213fe0870	Pengguna Google	5	sangat mengedukasi	2025-10-30

B. Hasil Data Labelling (Validasi Ahli)

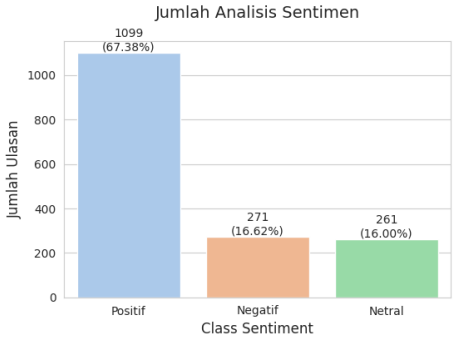
Dari 2.000 ulasan mentah yang diperoleh pada tahap *data acquisition*, dilakukan proses pelabelan manual untuk mengelompokkan setiap ulasan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Proses pelabelan dilakukan secara teliti dan divalidasi oleh ahli bahasa agar anotasi sentimen pada setiap teks tetap akurat dan konsisten. Setelah melalui tahap seleksi dan pembersihan, sebanyak 1.631 ulasan dinyatakan valid dan digunakan sebagai *dataset* berlabel.

Tabel Hasil Data Labelling

	Date	Username	Rating	Label	Ulasan
0	03/11/2025 19.05	yulianus doo	5	Positif	bagus banget
1	01/11/2025 12.07	Irawan Nugroho	5	Positif	aplikasi bagus
2	31/10/2025 11.23	M Reza Agusetiawan	2	Negatif	aplikasi tidak bisa dibuka, tanpa internet ter...
3	30/10/2025 14.09	Raden Nuna Daya Al Najabahari Jayawijaya	5	Positif	sangat mengedukasi. .

Hasil pelabelan menunjukkan distribusi sentimen sebagai berikut: 1.099 ulasan berkategori positif, 261 ulasan termasuk kategori netral, dan 271 ulasan berada pada kategori negatif. Komposisi ini memperlihatkan bahwa ulasan positif memiliki jumlah yang lebih dominan dibandingkan dua kategori lainnya.

Gambar



Gambar 2. Jumlah Analisis Sentimen

C. Hasil Text Preprocessing

1. Hasil Cleaning

Hasil cleaning menunjukkan bahwa berbagai karakter yang tidak relevan, seperti tanda baca berlebihan, emotikon, angka, URL, serta duplikasi huruf berhasil dihapus dari teks. Proses ini menghasilkan teks yang lebih rapi dan terkontrol, sehingga mengurangi noise yang dapat memengaruhi tahap pemodelan.

Tabel 2. Hasil Cleaning

Cleaning
GILA ISINYA MIND BLOWING BANGET baru hari download apk ini rasanya langsung kebuka pikiran ku banyak banget hal yang bisa dipelajarin disinii terus adu argumennya juga berbobot jugaa
ini aplikasi kenapa sih setiap mau jawab pertanyaan blank dlu ke kredensial profil bikin lemot aja ya Ga jelas bgt deh

2. Hasil Case Folding

Tahap *case folding* mengubah seluruh huruf menjadi huruf kecil. Proses ini mencegah duplikasi fitur akibat kapitalisasi berbeda dan membuat distribusi kata lebih konsisten untuk tahap pemrosesan selanjutnya.

Tabel 3. Hasil Case Folding

Case Folding

gila isinya mind blowing banget baru hari download apk ini rasanya langsung kebuka pikiran ku banyak banget hal yang bisa dipelajarin disinii terus adu argumennya juga berbobot jugaa
ini aplikasi kenapa sih setiap mau jawab pertanyaan blank dlu ke kredensial profil bikin lemot aja ya ga jelas bgt deh

3. Hasil Normalization

Normalization berfungsi untuk membuat data lebih konsisten. Proses ini mengubah kata-kata yang tidak baku, singkatan, atau bahasa gaul menjadi kata baku sesuai kaidah bahasa Indonesia.

Tabel 4. Hasil Normalization

Normalization
gila isinya mind blowing banget baru hari download aplikasi ini rasanya langsung kebuka pikiran ku banyak banget hal yang bisa dipelajarin disinii terus adu argumennya juga berbobot juga
ini aplikasi kenapa sih setiap mau jawab pertanyaan blank dulu ke kredensial profil bikin lemot saja ya tidak jelas banget deh

4. Hasil Tokenization

Hasil *tokenization* menunjukkan bahwa setiap ulasan berhasil dipecah menjadi kata-kata individual yang siap untuk diproses lebih lanjut. Proses ini memudahkan perhitungan frekuensi kata, analisis distribusi, serta penerapan *stopword removal* dan *stemming*. Tahap ini memastikan data telah berubah dari bentuk kalimat menjadi bentuk kata yang lebih terstruktur.

Tabel 5. Hasil Tokenization

Tokenization
['gila', 'isinya', 'mind', 'blowing', 'banget', 'baru', 'hari', 'download', 'aplikasi', 'ini', 'rasanya', 'langsung', 'kebuka', 'pikiran', 'ku', 'banyak', 'banget', 'hal', 'yang', 'bisa', 'dipelajarin', 'disinii', 'terus', 'adu', 'argumennya', 'juga', 'berbobot', 'juga']
['ini', 'aplikasi', 'kenapa', 'sih', 'setiap', 'mau', 'jawab', 'pertanyaan', 'blank', 'dulu', 'ke', 'kredensial', 'profil', 'bikin', 'lemot', 'saja', 'ya', 'tidak', 'jelas', 'banget', 'deh']

5. Hasil Stopword Removal

Tahap *stopword removal* bertujuan menghapus kata-kata umum yang tidak berpengaruh besar terhadap makna sentimen, seperti “yang”, “dan”, “di”, dan “ke”. Setelah proses ini, teks menjadi lebih ringkas dan fokus pada kata-kata yang bermakna. Penghapusan *stopword* mengurangi jumlah fitur yang kurang penting sehingga meningkatkan efektivitas model pada tahap klasifikasi

Tabel 6. Stopword Removal

Stopword Removal
['gila', 'isinya', 'mind', 'blowing', 'banget', 'download', 'aplikasi', 'langsung', 'kebuka', 'pikiran', 'ku', 'banget', 'dipelajarin', 'disinii', 'adu', 'argumennya', 'berbobot']
['aplikasi', 'sih', 'blank', 'kredensial', 'profil', 'bikin', 'lemot', 'ya', 'banget', 'deh']

6. Hasil Stemming

Stemming mengubah kata ke bentuk dasarnya sehingga variasi kata yang berbeda tetapi memiliki makna sama dapat digabungkan. Proses ini membantu mengurangi sparsitas data dan membuat representasi fitur lebih efisien. Tahap ini memastikan setiap kata berada dalam bentuk dasar yang konsisten sebelum dilakukan ekstraksi fitur menggunakan TF-IDF.

Tabel 7. Hasil Stemming

Stemming
gila isi mind blowing banget download aplikasi langsung buka pikir ku banget dipelajarin sini adu argumen bobot
aplikasi sih blank kredensial profil bikin lot ya banget deh

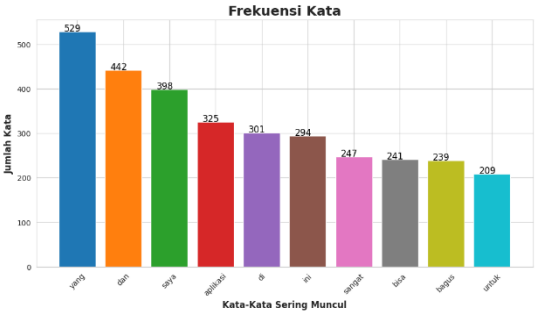
7. Hasil Word Cloud dan Frekuensi Kata

Wordcloud dan frekuensi kata dilakukan untuk mengidentifikasi pola kemunculan kata yang dominan pada dataset, baik sebelum maupun sesudah proses *text preprocessing*. Tahap ini bertujuan memberikan representasi visual dan statistik mengenai karakteristik linguistik yang muncul dalam ulasan pengguna aplikasi Quora, sehingga dapat memperkuat pemahaman terhadap kecenderungan topik serta pola penggunaan bahasa yang paling sering ditemukan.

Gambar 3. Word Cloud Sebelum Text Preprocessing

Gambar Word Cloud Sebelum Text Preprocessing

(Keterangan: Visualisasi menunjukkan kata-kata yang paling sering muncul pada data mentah sebelum dilakukan pembersihan teks.)



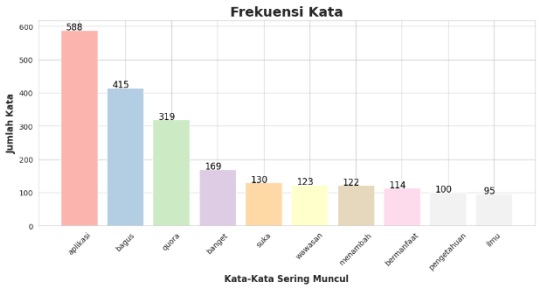
Gambar 4. Frekuensi Kata Sebelum Text Preprocessing

Gambar Frekuensi Kata Sebelum Text Preprocessing

(Keterangan: Menampilkan distribusi kata dominan pada data mentah yang masih mengandung kata tidak baku dan simbol non-teks.)

Gambar 5. Word Cloud Setelah Text Preprocessing

(Keterangan: Visualisasi menunjukkan kata-kata dominan setelah dilakukan pembersihan teks dan normalisasi kata, yang lebih merepresentasikan konteks sentimen pengguna.)



Gambar 6. Frekuensi Kata Setelah Text Preprocessing

(Keterangan: Menunjukkan frekuensi kata dominan setelah pembersihan data, dengan perbedaan yang lebih jelas antara sentimen positif, netral, dan negatif.)

D. Hasil Exploratory Analyst Data (EDA)

Analisis *Exploratory Data Analysis (EDA)* pada penelitian ini berfokus pada penggunaan visualisasi *wordcloud* untuk setiap kategori sentimen, yaitu positif, negatif, dan netral. Visualisasi *wordcloud* membantu menampilkan kata yang paling sering muncul pada masing-masing kategori sehingga memberi gambaran pola bahasa yang digunakan pengguna dalam menuliskan ulasan pada aplikasi Quora. Pendekatan ini relevan karena dapat mendukung proses identifikasi kecenderungan ekspresi dan karakteristik isi ulasan, yang selanjutnya berkaitan dengan rumusan masalah mengenai pola sentimen serta tujuan penelitian untuk memahami struktur kata yang dominan pada tiap kelompok sentimen.

1. Word Cloud Sentimen Positif

Wordcloud pada kategori positif memperlihatkan bahwa kata "bagus", "manfaat", "suka", dan "keren" muncul dengan frekuensi paling tinggi. Kemunculan kata yang bersifat apresiatif dan bernada emosional positif ini menunjukkan bahwa banyak pengguna menilai aplikasi Quora sebagai layanan yang bermanfaat, memberikan informasi yang mereka butuhkan, dan mudah digunakan.

Gambar 7. Word Cloud Sentimen Positif

2. Word Cloud Sentimen Negatif

Pada kategori negatif, *wordcloud* menunjukkan dominasi kata "bug", "hilanh", "spam", "susah", dan "loading". Kata tersebut menggambarkan keluhan yang berkaitan dengan hambatan teknis, seperti aplikasi yang gagal dibuka, gangguan selama penggunaan, atau kinerja yang lambat.

Gambar 8. Word Cloud Sentimen Negatif

3. Word Cloud Sentimen Netral

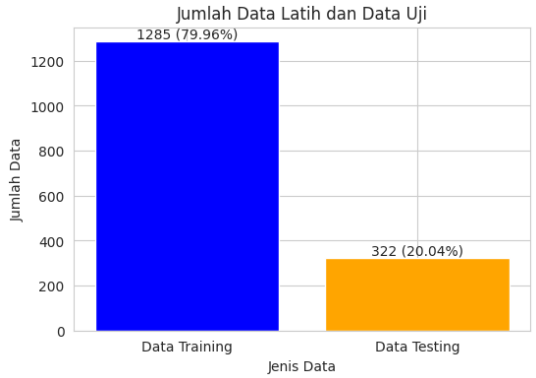
Wordcloud pada sentimen netral menampilkan kata yang muncul tanpa menunjukkan kecenderungan emosional tertentu. Kata tersebut menggambarkan isi ulasan yang bersifat deskriptif, berupa penjelasan, pertanyaan, atau komentar yang tidak memuat penilaian positif maupun negatif.

Gambar 9. Word Cloud Sentimen Netral

E. Data Splitting (80:20)

Setelah tahap *preprocessing* selesai, terdapat 1.607 ulasan yang dinyatakan valid dan siap digunakan untuk proses pemodelan. Data kemudian dipisahkan menjadi 1.285 ulasan (80%) sebagai data latih untuk membangun model *Multinomial Naïve Bayes*, serta 322 ulasan (20%) sebagai data uji untuk menilai kinerja model melalui metrik akurasi, *precision*, *recall*, dan *F1-score*. Proses pemisahan dilakukan dengan *stratified splitting* agar proporsi label sentimen

tetap terjaga pada kedua kelompok data. Pendekatan ini memastikan model memperoleh data latih yang seimbang dan menghasilkan evaluasi yang lebih tepat pada data uji.



Gambar 10. Hasil Data Splitting

F. Hasil Feature Extration

Menghasilkan representasi numerik berdimensi tinggi untuk setiap ulasan, dengan proses *fitting* dilakukan hanya pada data latih guna mencegah *data leakage* setelah pemisahan data. Hasil ekstraksi menunjukkan bahwa TF-IDF membentuk 652 fitur unik sebagai kosakata akhir setelah melalui proses pembersihan teks, *case folding*, normalisasi, tokenisasi, penghapusan *stopword*, dan *stemming*. Matriks TF-IDF yang dihasilkan memiliki dimensi (1285, 652) untuk data latih dan (322, 652) untuk data uji, yang menandakan bahwa kedua kelompok data direpresentasikan menggunakan jumlah fitur yang konsisten berdasarkan parameter dari data latih.

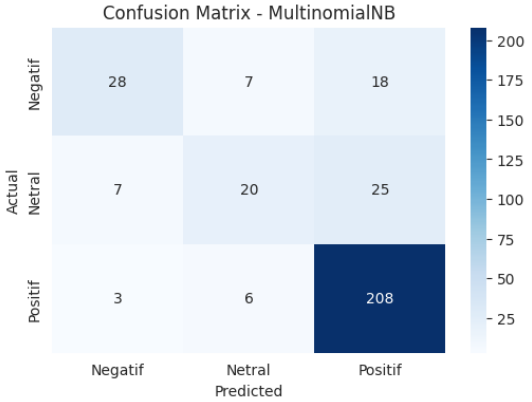
G. Hasil Model Evaluation for Classification

Berdasarkan hasil evaluasi model yang ditampilkan pada tabel accuracy, algoritma *Multinomial Naïve Bayes* berhasil mencapai tingkat akurasi sebesar **80%** dalam mengklasifikasikan 322 ulasan pada data uji. Nilai ini menunjukkan bahwa sebagian besar prediksi yang dihasilkan model sesuai dengan label sebenarnya, sehingga dapat dikatakan bahwa model memiliki performa yang baik untuk tugas analisis sentimen. Akurasi yang cukup tinggi ini turut didukung oleh nilai *weighted average* precision dan F1-score yang stabil, masing-masing sebesar 0.78, yang mengindikasikan bahwa model mampu menangani ketidakseimbangan kelas secara efektif. Dengan demikian, hasil akurasi ini memperkuat bahwa kombinasi TF-IDF dan *Multinomial Naïve Bayes* mampu menghasilkan klasifikasi sentimen yang akurat pada ulasan pengguna aplikasi Quora.

Tabel 8. Hasil Evaluasi Model Naïve Bayes

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.74	0.53	0.62	53
Netral	0.61	0.38	0.47	52
Positif	0.83	0.96	0.89	217
Accuracy			0.80	322
Macro Avg	0.72	0.62	0.66	322
Weighted Avg	0.78	0.80	0.78	322

Gambar *confusion matrix* memperlihatkan sebaran prediksi model untuk tiap kelas sentimen. Model mengenali kelas positif dengan baik, ditunjukkan oleh jumlah prediksi benar yang dominan pada diagonal. Kesalahan terbesar muncul pada kelas netral karena banyak sampel tercampur ke kelas positif atau negatif. Kelas negatif menunjukkan tingkat kesalahan sedang, terutama akibat kemiripan kata dan konteks yang kurang jelas. Secara keseluruhan, *confusion matrix* menunjukkan bahwa model kuat pada kelas positif namun masih lemah dalam membedakan sentimen netral.



Gambar 11. Confusion Matrix

CONCLUSION

Berdasarkan penelitian berjudul *"Analisis Sentimen Ulasan Pengguna Quora Menggunakan Multinomial Naïve Bayes dan TF-IDF"*, dapat disimpulkan bahwa model *Multinomial Naïve Bayes* mampu bekerja cukup baik dengan akurasi 80% dan *weighted average F1-score* 0.78. Hasil ini menunjukkan bahwa sebagian besar ulasan pada data uji dapat diprediksi dengan benar meskipun distribusi kelas tidak seimbang. Model paling akurat dalam mengenali sentimen positif dengan F1-score 0.89 dan *recall* 0.96, sehingga hampir seluruh ulasan positif berhasil diidentifikasi. Untuk sentimen negatif, F1-score 0.62 menunjukkan bahwa sebagian besar sampel dapat diklasifikasikan dengan tepat meskipun masih ada kesalahan ke kelas netral. Kelas netral menjadi yang paling sulit dibedakan dengan F1-score 0.47 karena banyak ulasan bersifat ambigu dan minim ekspresi emosional.

Secara keseluruhan, kombinasi TF-IDF dan *Multinomial Naïve Bayes* terbukti efektif untuk analisis sentimen berbahasa Indonesia pada ulasan aplikasi Quora, namun model masih menunjukkan kecenderungan bias terhadap kelas mayoritas. Perbaikan dapat dilakukan melalui penyeimbangan data, penambahan variasi ulasan, atau penggunaan metode representasi semantik yang lebih kuat agar performa pada kelas netral dan negatif meningkat.

REFERENCE

[1] S. Khoerunnisa, "Application of the Naive Bayes Algorithm with TF-IDF and Cross Validation for Public Sentiment on Starlink," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, 2025.

[2] A. A. P. Simarmata, M. S. Hutagalung, A. A. S. Simarmata, and A. H. Koto, "Sentiment Analysis on BRImo Application Reviews Using Naïve Bayes," *J. Apl. Inform.*, 2025.

[3] I. E. Team, "Sentimen Analisis Aplikasi Digitalent Mobile Menggunakan Naïve Bayes," *INTECOM J. Inf. Technol. Comput. Sci.*, 2024.

[4] S. A. Helmayanti, "Penerapan Algoritma TF-IDF dan Naïve Bayes untuk Analisis Sentimen Aspek Ulasan Aplikasi Flip," *J. IT/MIK*, 2023.

[5] H. N. Zuhdi and B. Prasetiyo, "Sentiment Analysis on Ipusnas Application Reviews in Google Play Store Using Naive Bayes Classifier," *Int. J. IRSE*, 2025.

[6] C. H. Yutika, Adiwijaya, and S. Al Faraby, "Aspect-Based Sentiment Analysis on Female Daily Reviews Using TF-IDF and Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, 2021.

[7] A. Nurul, W. Pradnyan, and A. Wibawa, "A Review of Sentiment Analysis Applications in Indonesia," *Unesa J. Inf. Syst.*, 2022.

[8] R. N. Rahman, "Analisis Sentimen Ulasan Game eFootball 2024 Menggunakan Naïve Bayes dan TF-IDF," *J. Inform.*, 2025.

[9] N. A. Murnastiti, "Analisis Sentimen Terhadap Makanan Manis di Platform X Menggunakan TF-IDF dan Naïve Bayes," *J-PTIIK*, vol. 9, 2025.

[10] A. Serlina, "Comparative Analysis of Naïve Bayes Algorithm in English and Indonesian Reviews," *Tek. J.*, 2025.

[11] M. F. Pratama and D. Sari, "Sentiment Analysis on E-commerce and Financial App Reviews Using Naïve Bayes and TF-IDF," in *Proceedings of National Conference on Informatics*, 2023.

[12] S. Lailiyah, M. Fahrurozi, and R. A. Nugroho, "Sentimen Analisis Ulasan Playstore Menggunakan Naïve Bayes: Studi pada Tokopedia, BRImo, dan MyPertamina," *J. Ilm. Teknol. Inf.*, 2024.

[13] S. Purwanto and R. Winarto, "Sentiment Analysis of Public Service Applications Using Multinomial Naïve Bayes," *J. Sist. Inf. dan Teknol.*, 2023.

[14] M. Rachman, T. A. Fadillah, and I. R. Kartika, "Comparison of Naïve Bayes Variants for Indonesian Text Reviews," *J. Inform. Glob.*, 2022.

[15] L. H. Permana and A. Widodo, "TF-IDF Based Sentiment Classification on Playstore Application Reviews," *Int. J. Comput. Inf. Res.*, 2024.