JGIT
Jurnal Greenation Ilmu Teknik
+62 81210467572 https://greenationpublisher.org/JGIT greenation.info@gmail.com
GREENATION PUBLISHER

DOI: https://doi.org/10.38035/jgit.v4i2.693

Analisis Sentimen Ulasan Aplikasi Livin' by Mandiri pada Google Play Store Menggunakan Metode Naive Bayes

M. Rizky Alfaredho[1], Alan Novi Tompunu[2], Ariansyah Saputra[3]
[1]Politeknik Negeri Sriwijaya, Palembang, Indonesia, rizkyalfaredho22@gmail.com
[2]Politeknik Negeri Sriwijaya, Palembang, Indonesia, alan_nt@polsri.ac.id
[3]Politeknik Negeri Sriwijaya, Palembang, Indonesia, ariansyah@polsri.ac.id

Corresponding Author: rizkyalfaredho22@gmail.com[1]

Abstract: *User reviews on Google Play Store contain valuable information about users' experiences with Livin' by Mandiri, but their large volume and unstructured form make manual analysis inefficient. This study aims to apply text preprocessing and Naive Bayes to classify reviews into positive and negative sentiments and evaluate the model using accuracy, precision, recall, and F1-score. The data consisted of 10,000 Indonesian reviews collected through web scraping. After removing noise and duplicate content, 7,824 reviews remained. The stages included cleaning, case folding, tokenization, normalization, stopword removal, stemming, lexicon-based labeling, an 80:20 train-test split, TF-IDF feature extraction, SMOTE, and Naive Bayes modeling. Neutral reviews were removed, resulting in 7,583 records, including 3,695 positive and 3,888 negative reviews. The selected Multinomial Naive Bayes model achieved 85.63% accuracy, 87.14% precision, 85.63% recall, and 85.44% F1-score. The model was implemented in the SentimenLivin AI website to display sentiment distributions, evaluation results, dominant words, and new-review predictions. The results show that Naive Bayes can classify Livin' by Mandiri reviews effectively.*

Keyword: *Google Play Store, Livin' by Mandiri, Naive Bayes, Sentiment Analysis, TF-IDF.*

Abstrak: Ulasan pengguna pada Google Play Store memuat informasi mengenai pengalaman menggunakan Livin' by Mandiri, tetapi jumlahnya yang besar dan bentuknya yang tidak terstruktur menyebabkan analisis manual menjadi tidak efisien. Penelitian ini bertujuan menerapkan prapemrosesan teks dan Naive Bayes untuk mengklasifikasikan ulasan ke dalam sentimen positif dan negatif serta mengevaluasi model berdasarkan *accuracy*, *precision*, *recall*, dan *F1-score*. Data penelitian terdiri atas 10.000 ulasan berbahasa Indonesia yang dikumpulkan melalui *web scraping*. Setelah penghapusan noise dan duplikasi konten, diperoleh 7.824 ulasan. Tahapan penelitian meliputi *cleaning*, *case folding*, *tokenizing*, normalisasi, *stopword removal*, *stemming*, pelabelan berbasis leksikon, pembagian data 80:20, ekstraksi fitur TF-IDF, SMOTE, dan pemodelan Naive Bayes. Penghapusan data netral menghasilkan 7.583 ulasan, yang terdiri atas 3.695 ulasan positif dan 3.888 ulasan negatif. Model Multinomial Naive Bayes terpilih menghasilkan *accuracy* 85,63%, *precision*

87,14%, *recall* 85,63%, dan *F1-score* 85,44%. Model kemudian diimplementasikan pada situs SentimenLivin AI untuk menampilkan distribusi sentimen, evaluasi model, kata dominan, dan prediksi ulasan baru. Hasil penelitian memperlihatkan bahwa Naive Bayes mampu mengklasifikasikan ulasan Livin' by Mandiri dengan baik.

Kata Kunci: Analisis Sentimen, Google Play Store, Livin' by Mandiri, Naive Bayes, TF-IDF.

PENDAHULUAN

Transformasi digital telah menggeser aktivitas masyarakat dari layanan konvensional menuju layanan berbasis aplikasi. Pada sektor keuangan, perubahan ini ditandai oleh meningkatnya pemanfaatan *mobile banking* yang memungkinkan nasabah melakukan transfer dana, pembayaran tagihan, pengisian dompet digital, pembelian produk finansial, dan pengelolaan rekening melalui perangkat bergerak. Keberhasilan layanan tersebut tidak hanya ditentukan oleh kelengkapan fitur, tetapi juga oleh kemudahan penggunaan, manfaat, keamanan, risiko, kualitas layanan, dan kepercayaan pengguna terhadap institusi perbankan (Alalwan et al., 2017; Souiden et al., 2021; Suharman & Sulaeman, 2025).

Persaingan industri perbankan digital di Indonesia mendorong penyedia layanan untuk terus meningkatkan inovasi, keamanan, kemudahan navigasi, serta stabilitas sistem. Salah satu layanan yang banyak digunakan ialah Livin' by Mandiri, yaitu aplikasi perbankan digital yang mengintegrasikan transaksi finansial, pembayaran, investasi, dan kebutuhan gaya hidup dalam satu platform. Tingginya jumlah unduhan dan ulasan di Google Play Store memperlihatkan luasnya penggunaan aplikasi tersebut. Ulasan pengguna memuat apresiasi, kritik, keluhan teknis, dan saran mengenai kualitas fitur maupun antarmuka aplikasi (Komalasari et al., 2026).

Ulasan aplikasi merupakan sumber umpan balik yang dapat digunakan untuk memahami pengalaman pengguna, mengidentifikasi gangguan, menemukan kebutuhan fitur, mengevaluasi pembaruan, dan menentukan prioritas pengembangan perangkat lunak (Maalej et al., 2016; Martin et al., 2017). Kajian sistematis juga menempatkan ulasan aplikasi sebagai sumber data penting dalam evaluasi kualitas, pemeliharaan sistem, penggalian kebutuhan, dan pengambilan keputusan pengembangan aplikasi (Dąbrowski et al., 2022). Namun, ulasan Google Play Store umumnya berbentuk teks bebas yang tidak terstruktur, menggunakan bahasa sehari-hari, singkatan, kata tidak baku, kesalahan pengetikan, emotikon, dan simbol. Jumlah data yang besar menyebabkan identifikasi opini secara manual menjadi tidak efisien, memerlukan waktu lama, dan rentan terhadap subjektivitas (Hariansyah & Siswanto, 2022).

Permasalahan tersebut dapat diatasi melalui analisis sentimen dengan memanfaatkan *machine learning* dan *Natural Language Processing*. Analisis sentimen memungkinkan sistem mengenali kecenderungan opini dalam teks dan mengelompokkannya ke dalam kategori tertentu secara otomatis. Karakteristik linguistik, kosakata tidak baku, singkatan, dan pola kata dapat diolah menjadi fitur yang dikenali oleh model klasifikasi (Nurzaman et al., 2024). Pendekatan analisis sentimen dapat berbasis leksikon, *machine learning*, metode hibrida, maupun *deep learning*, dengan pemilihan metode yang disesuaikan dengan karakteristik korpus, jumlah data, tujuan klasifikasi, dan sumber daya komputasi (Birjali et al., 2021).

Model *deep learning* mampu mempelajari representasi teks yang kompleks, tetapi umumnya memerlukan data dan sumber daya komputasi yang besar (Minaee et al., 2021). Sebaliknya, algoritma klasik masih banyak digunakan karena lebih sederhana dan efisien. Salah satunya adalah Naive Bayes, yaitu algoritma probabilistik berbasis Teorema Bayes yang memiliki proses pelatihan cepat, kebutuhan komputasi ringan, dan kemampuan

mengolah data teks berdimensi tinggi (Ariany et al., 2026). Meskipun demikian, klasifikasi sentimen tetap menghadapi tantangan berupa ambiguitas bahasa, negasi, sarkasme, kesalahan penulisan, singkatan, dan ketergantungan makna pada konteks kalimat (Wankhade et al., 2022).

Naive Bayes telah diterapkan pada berbagai ulasan aplikasi perbankan. Insan et al. (2023) dan Umair dan Susanto (2024) menggunakannya untuk menganalisis ulasan BRImo, sedangkan Tobing dan Febriandirza (2024) menerapkannya pada ulasan aplikasi BCA Mobile. Pada objek Livin' by Mandiri, Hariansyah dan Siswanto (2022) menggunakan Multinomial Naive Bayes pada data Twitter, sementara Alviyanti et al. (2024) memanfaatkan ulasan Google Play Store. Nirmala dan Sanjaya ER (2024) menggunakan Logistic Regression untuk mendeteksi kecenderungan sentimen berdasarkan kata yang terdapat dalam ulasan. Komalasari et al. (2026) selanjutnya mengombinasikan Naive Bayes dengan pembobotan TF-IDF dan memperoleh akurasi sebesar 88,37%. Temuan berbagai studi tersebut memperlihatkan bahwa ulasan Livin' by Mandiri dapat dianalisis menggunakan beragam algoritma, tetapi performa model tetap dipengaruhi oleh kualitas pelabelan, tahapan *preprocessing*, representasi fitur, dan distribusi kelas.

Pengembangan analisis sentimen pada aplikasi perbankan juga dilakukan melalui algoritma dan teknik tambahan. Simarmata dan Sasongko (2025) memanfaatkan IndoBERT pada ulasan aplikasi BRImo, sedangkan Ginasputri et al. (2025) mengombinasikan Naive Bayes dengan seleksi fitur Chi-Square untuk menganalisis respons pengguna terhadap pembaruan fitur. Akbar dan Pratama (2025) membandingkan Naive Bayes, Random Forest, dan Decision Tree pada ulasan aplikasi MyBCA. Perbedaan pendekatan tersebut memperlihatkan bahwa pemilihan algoritma dan ekstraksi fitur menjadi bagian penting dalam memperoleh performa klasifikasi yang sesuai dengan karakteristik data.

Meskipun telah banyak diterapkan, analisis sentimen terhadap data ulasan aktual masih menghadapi beberapa permasalahan. Salah satunya adalah ketidaksesuaian antara rating bintang dan isi teks. Pengguna dapat memberikan rating tinggi, tetapi menuliskan keluhan, atau memberikan rating rendah dengan komentar yang tidak sepenuhnya bersifat negatif. Anomali tersebut ditemukan dalam analisis ulasan MyPertamina dan JMO, sehingga pelabelan yang hanya didasarkan pada rating berpotensi menghasilkan kategori sentimen yang tidak merepresentasikan isi teks secara tepat (Darmawan et al., 2023; Rizaldi et al., 2023). Beberapa studi juga memilih menghapus data netral untuk menyederhanakan klasifikasi tanpa memberikan evaluasi komparatif yang memadai terhadap dampak penghapusan tersebut (Sayogo et al., 2023).

Permasalahan lain berkaitan dengan ambiguitas bahasa dan ketidakseimbangan distribusi kelas. Model berbasis pembobotan kata dapat mengalami kesulitan ketika ulasan mengandung negasi, sarkasme, istilah tidak baku, atau kombinasi kata yang memiliki makna kontekstual. Selain itu, jumlah data pada setiap kategori sentimen tidak selalu sama sehingga model berpotensi lebih banyak mempelajari kelas dominan. Komalasari et al. (2026) memperlihatkan bahwa penerapan TF-IDF dan Naive Bayes telah menghasilkan performa yang baik, tetapi persoalan ambiguitas makna dan *class imbalance* masih membuka peluang pengembangan lebih lanjut. Berdasarkan kondisi tersebut, celah riset terletak pada mekanisme pelabelan yang lebih berorientasi pada isi teks, penanganan distribusi kelas, dan penyajian interpretasi hasil yang tidak terbatas pada metrik evaluasi.

Riset ini menerapkan pelabelan berbasis aturan atau *rule-based text labeling* dengan memanfaatkan daftar kata positif dan negatif. Pendekatan tersebut digunakan agar kategori sentimen ditentukan berdasarkan kandungan teks ulasan, bukan semata-mata rating bintang. Data selanjutnya diproses melalui tahapan *cleansing*, *case folding*, *tokenization*, normalisasi kata tidak baku, *stopword removal*, dan *stemming*. Hasil pemrosesan direpresentasikan dalam bentuk numerik menggunakan TF-IDF, sedangkan SMOTE diterapkan pada data latih untuk

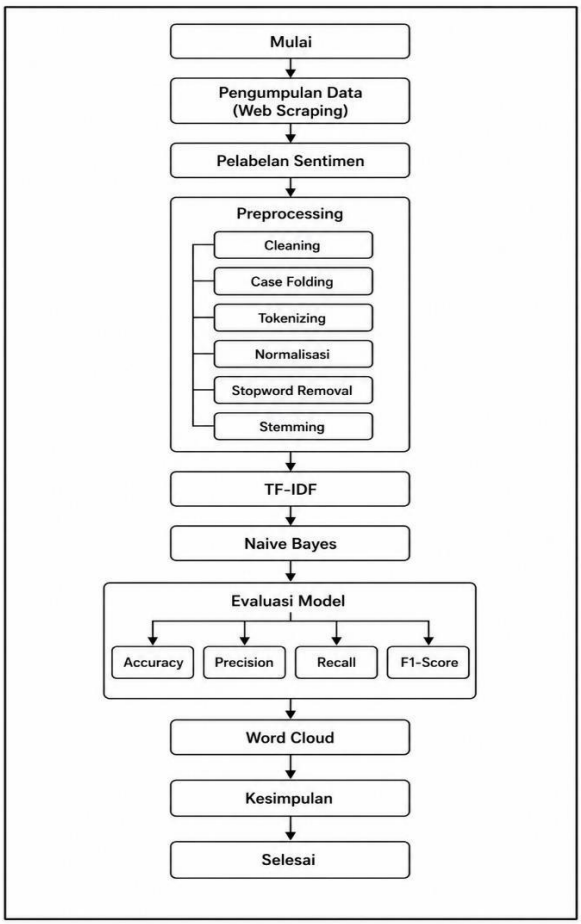
mengurangi potensi bias akibat perbedaan jumlah data antarkelas. Proses klasifikasi dilakukan menggunakan Naive Bayes melalui bahasa pemrograman Python dalam lingkungan Google Colab.

Selain mengevaluasi performa klasifikasi, riset ini menyajikan visualisasi frekuensi kata dan *word cloud* untuk memperlihatkan kosakata yang dominan pada kelompok sentimen positif dan negatif. Visualisasi tersebut digunakan untuk melengkapi metrik kuantitatif agar apresiasi maupun keluhan pengguna dapat diinterpretasikan secara lebih komunikatif. Dengan demikian, keluaran analisis tidak hanya berupa nilai *accuracy*, *precision*, *recall*, dan *F1-score*, tetapi juga informasi mengenai kecenderungan pengalaman pengguna yang dapat dipertimbangkan dalam evaluasi layanan digital.

Riset dibatasi pada ulasan berbahasa Indonesia yang diperoleh secara mandiri dari Google Play Store melalui teknik *web scraping*. Kategori sentimen dibedakan menjadi positif dan negatif, sedangkan analisis difokuskan pada isi teks tanpa menghubungkannya dengan identitas, karakteristik, atau perilaku pengguna. Algoritma yang digunakan adalah Naive Bayes tanpa perbandingan langsung dengan algoritma klasifikasi lainnya. Berdasarkan ruang lingkup tersebut, riset ini bertujuan menerapkan tahapan *preprocessing* dan Naive Bayes untuk mengklasifikasikan sentimen ulasan Livin' by Mandiri serta mengevaluasi performanya berdasarkan *accuracy*, *precision*, *recall*, dan *F1-score*. Temuan yang diperoleh diharapkan dapat memberikan gambaran mengenai persepsi pengguna, menjadi bahan evaluasi bagi pengembangan layanan aplikasi, dan memperkaya referensi mengenai penerapan analisis sentimen pada platform perbankan digital.

METODE

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen pada bidang *machine learning* dan *text mining*. Objek penelitian berupa ulasan pengguna aplikasi Livin' by Mandiri yang diperoleh dari Google Play Store. Data dibatasi pada ulasan berbahasa Indonesia dan dikumpulkan secara mandiri melalui teknik *web scraping* menggunakan Python. Proses penelitian mencakup pengumpulan data, pembersihan awal, *preprocessing* teks, pelabelan sentimen, pembagian data, ekstraksi fitur TF-IDF, penerapan SMOTE, pemodelan Naive Bayes, evaluasi model, visualisasi *word cloud*, dan implementasi hasil pada situs web.



Sumber: Hasil riset, 2026

Gambar 1. Tahapan penelitian

Tahapan penelitian pada Gambar 1 diawali dengan pengumpulan 10.000 ulasan Livin' by Mandiri dari Google Play Store. Data hasil *scraping* memiliki atribut *username*, *content*, *score*, dan *at*. Atribut *content* menjadi teks utama yang dianalisis, *score* memuat penilaian bintang 1 sampai 5, *username* digunakan untuk membantu pemeriksaan data, sedangkan *at* menunjukkan waktu ulasan. Data kemudian dibersihkan dengan menghapus ulasan kosong, ulasan yang tidak memiliki karakter huruf, dan konten yang terduplikasi. Pembersihan dilakukan sebelum proses pengolahan teks agar data yang masuk ke dalam model tidak mengandung baris yang tidak informatif maupun pengulangan ulasan yang dapat memengaruhi pembobotan fitur.

Tahap *preprocessing* terdiri atas *cleaning*, *case folding*, *tokenizing*, normalisasi, *stopword removal*, dan *stemming*. Tahap *cleaning* menghapus URL, emoji, angka, tanda baca, serta karakter khusus. *Case folding* mengubah seluruh huruf menjadi huruf kecil, sedangkan *tokenizing* memecah kalimat menjadi token kata. Normalisasi mengubah singkatan, kata tidak baku, dan kesalahan pengetikan menjadi bentuk yang lebih seragam. *Stopword removal* menghapus kata umum yang tidak memberikan informasi penting terhadap sentimen, sedangkan *stemming* mengubah kata berimbuhan menjadi kata dasar menggunakan pustaka Sastrawi. Hasil setiap tahap digunakan secara berurutan sehingga teks akhir menjadi lebih konsisten sebelum proses pelabelan dan pembobotan.

Pelabelan dilakukan setelah tahap *preprocessing* menggunakan pendekatan *lexicon based* dengan kamus sentimen bahasa Indonesia. Setiap ulasan dihitung berdasarkan jumlah kata bermuatan positif dan negatif. Ulasan diberi label positif apabila skor kata positif lebih besar daripada skor kata negatif, sedangkan label negatif diberikan apabila skor negatif lebih

besar. Data yang tidak memperlihatkan dominasi skor dimasukkan ke kategori netral. Karena penelitian berfokus pada klasifikasi dua kelas, data netral dikeluarkan sebelum proses pemodelan sehingga data yang digunakan hanya terdiri atas sentimen positif dan negatif.

Data dua kelas selanjutnya dibagi menjadi data latih dan data uji menggunakan *random splitting* dengan perbandingan 80:20. Sebanyak 6.066 ulasan digunakan sebagai data latih dan 1.517 ulasan digunakan sebagai data uji. Teks ditransformasikan menjadi bentuk numerik menggunakan TF-IDF dengan *max_features* 5.000 dan *ngram_range* (1,2), sehingga fitur yang terbentuk terdiri atas *unigram* dan *bigram*. TF-IDF dipasang pada data latih, kemudian digunakan untuk mentransformasikan data latih dan data uji. SMOTE diterapkan pada data latih untuk membentuk data sintetis pada kelas dengan jumlah lebih sedikit agar distribusi kelas menjadi seimbang, sedangkan data uji tetap mempertahankan distribusi awal.

Pemodelan dilakukan menggunakan Multinomial Naive Bayes dan Complement Naive Bayes. Konfigurasi yang diuji meliputi Multinomial Naive Bayes dengan nilai *alpha* 1,0; 0,1; dan 0,01 serta Complement Naive Bayes dengan *alpha* 1,0. Pemilihan model didasarkan pada nilai *accuracy* dan *weighted F1-score*. Evaluasi akhir terhadap data uji dilakukan menggunakan *accuracy*, *precision*, *recall*, *F1-score*, serta *confusion matrix*. Model terpilih kemudian diimplementasikan menggunakan Python pada bagian *backend* serta HTML, CSS, dan JavaScript pada bagian *frontend* situs SentimenLivin AI.

HASIL DAN PEMBAHASAN

Hasil Pengumpulan dan Pembersihan Data

Proses *web scraping* menghasilkan 10.000 ulasan pengguna Livin' by Mandiri dari Google Play Store. Distribusi *rating* terdiri atas 2.740 ulasan bintang 1, 694 ulasan bintang 2, 712 ulasan bintang 3, 584 ulasan bintang 4, dan 5.270 ulasan bintang 5. Pada tahap pembersihan awal tidak ditemukan ulasan kosong. Sebanyak 92 data dihapus karena hanya berisi angka, simbol, atau emoji tanpa karakter huruf. Penghapusan 2.084 konten duplikat menghasilkan 7.824 ulasan yang digunakan pada tahap *preprocessing*.

Tabel 1. Ringkasan Hasil Pembersihan Data

Tahap	Jumlah Data	Perubahan
Data mentah hasil scraping	10.000	Data awal
Setelah penghapusan ulasan kosong	10.000	Tidak berkurang
Setelah penghapusan noise tanpa huruf	9.908	Berkurang 92
Setelah penghapusan duplikasi	7.824	Berkurang 2.084
Setelah penghapusan label netral	7.583	Berkurang 241

Sumber: Hasil riset, 2026

Penilaian bintang lima memiliki jumlah terbanyak, yaitu 5.270 ulasan atau 52,70%, sedangkan penilaian bintang satu berjumlah 2.740 ulasan atau 27,40%. Komposisi tersebut memberikan gambaran awal bahwa banyak pengguna memberikan penilaian tinggi, tetapi nilai bintang tidak langsung digunakan sebagai label sentimen. Proses klasifikasi dalam penelitian ini tetap didasarkan pada isi teks ulasan melalui pelabelan berbasis leksikon karena teks memuat bentuk apresiasi, keluhan, dan pengalaman pengguna secara lebih rinci.

Pengurangan data dari 10.000 menjadi 7.824 ulasan memperlihatkan bahwa data hasil *scraping* belum dapat langsung digunakan dalam pemodelan. Ulasan tanpa karakter huruf tidak memberikan informasi yang dapat diolah sebagai fitur, sedangkan konten duplikat dapat menyebabkan kata tertentu memperoleh frekuensi yang berlebihan. Oleh karena itu, pembersihan awal diperlukan untuk mempertahankan ulasan yang benar-benar memiliki isi teks dan mengurangi pengulangan pola yang sama pada dataset.

Hasil *Preprocessing* dan Pelabelan Sentimen

Tahap *preprocessing* menghasilkan teks yang lebih bersih dan seragam. Contoh pada Tabel 2 memperlihatkan perubahan ulasan "knp Livin saya GK bisa di buka. jadinya sulit untuk melakukan transaksi" menjadi "livin tidak buka jadi sulit lakukan transaksi". *Cleaning* menghapus tanda baca, *case folding* menyeragamkan huruf, *tokenizing* memisahkan kata, normalisasi mengubah "knp" menjadi "kenapa" dan "gk" menjadi "tidak", *stopword removal* menghapus kata yang kurang informatif, sedangkan *stemming* mengubah kata berimbuhan ke bentuk dasarnya.

Tabel 2. Contoh Tahapan *Preprocessing* Teks

Tahapan	Hasil
Teks asli	knp Livin saya GK bisa di buka.. jadinya sulit untuk melakukan transaksi
Cleaning	knp livin saya gk bisa di buka jadinya sulit untuk melakukan transaksi
Tokenisasi	knp | livin | saya | gk | bisa | di | buka | jadinya | sulit | untuk | melakukan | transaksi
Normalisasi	kenapa livin saya tidak bisa di buka jadinya sulit untuk melakukan transaksi
Stopword removal	livin tidak buka jadinya sulit melakukan transaksi
Stemming	livin tidak buka jadi sulit lakukan transaksi

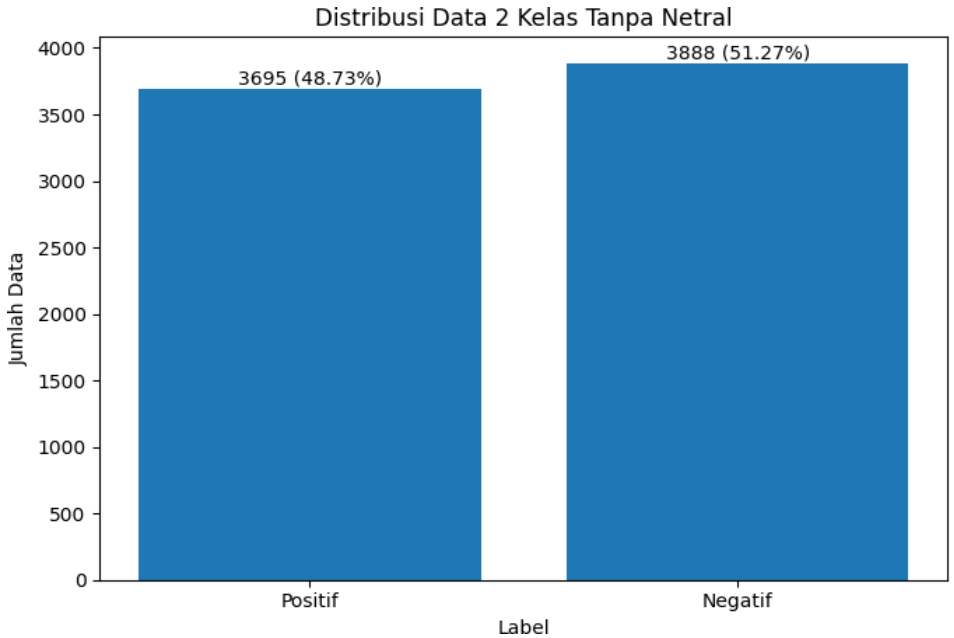
Sumber: Hasil riset, 2026

Pelabelan terhadap 7.824 ulasan menghasilkan kelas positif, negatif, dan netral. Setelah 241 data netral dikeluarkan, diperoleh 7.583 ulasan dua kelas yang terdiri atas 3.695 sentimen positif atau 48,73% dan 3.888 sentimen negatif atau 51,27%. Data tersebut selanjutnya digunakan dalam proses pembagian data, ekstraksi fitur, pelatihan, dan pengujian model.

Tabel 3. Distribusi Sentimen dan Pembagian *Dataset*

Komponen	Positif	Negatif	Total
Dataset dua kelas	3.695 (48,73%)	3.888 (51,27%)	7.583
Data latih sebelum SMOTE	2.956	3.110	6.066
Data uji	739	778	1.517

Sumber: Hasil riset, 2026



Sumber: Hasil riset, 2026

Gambar 2. Distribusi Hasil Pelabelan Sentimen

Sentimen positif menggambarkan pengalaman pengguna yang berkaitan dengan kemudahan transaksi, kecepatan, manfaat, dan kelancaran penggunaan aplikasi. Sentimen negatif memuat keluhan mengenai aplikasi yang tidak dapat dibuka, kegagalan *login*, *error*, transaksi gagal, respons yang lambat, dan kesulitan menggunakan fitur tertentu. Distribusi hasil pelabelan pada Gambar 2 memperlihatkan bahwa kedua kelas relatif berimbang, meskipun jumlah sentimen negatif sedikit lebih tinggi daripada sentimen positif.

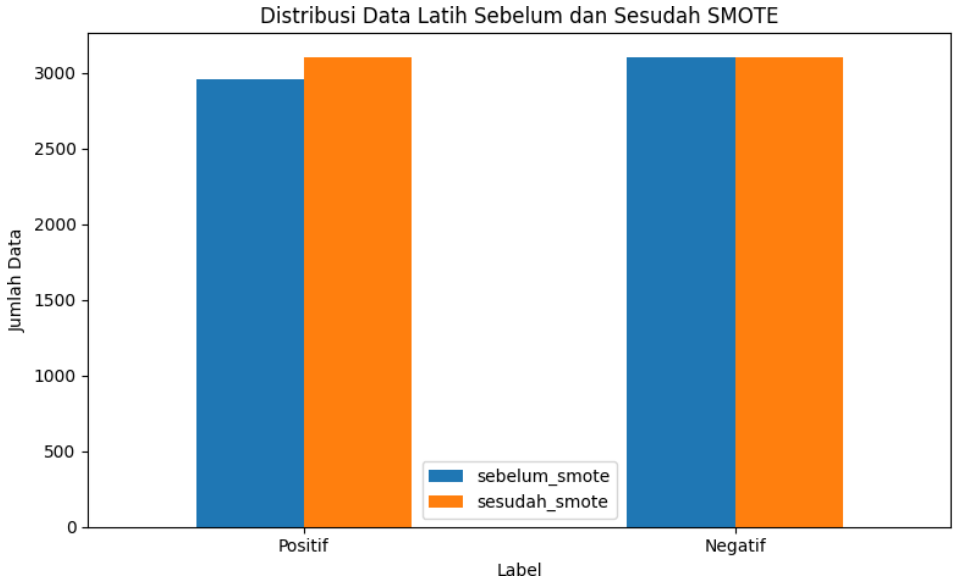
Hasil tersebut menunjukkan bahwa pengalaman pengguna terhadap Livin' by Mandiri tidak hanya berisi apresiasi terhadap kemudahan layanan, tetapi juga memuat keluhan teknis dalam jumlah yang hampir sama. Proporsi negatif sebesar 51,27% menandakan bahwa persoalan akses aplikasi, autentikasi, kestabilan sistem, dan kegagalan transaksi menjadi bagian penting dari opini pengguna. Pada sisi lain, proporsi positif sebesar 48,73% tetap memperlihatkan bahwa kemudahan dan manfaat aplikasi dirasakan oleh banyak pengguna.

Tahap *preprocessing* berperan dalam membantu pelabelan tersebut karena kata tidak baku dan bentuk kata berimbuhan diseragamkan sebelum skor leksikon dihitung. Normalisasi, misalnya, memungkinkan istilah seperti "gk" dikenali sebagai "tidak", sedangkan *stemming* menyatukan bentuk "memudahkan" menjadi "mudah". Penyeragaman tersebut mengurangi variasi penulisan yang dapat menyebabkan kata bermuatan sentimen tidak terbaca secara konsisten.

Hasil Pembagian Data dan Ekstraksi Fitur TF-IDF

Dataset berjumlah 7.583 ulasan dibagi dengan rasio 80:20, yaitu 6.066 data latih dan 1.517 data uji. Hasil ekstraksi fitur TF-IDF menghasilkan 5.000 fitur dengan ukuran matriks data latih 6.066 × 5.000 dan matriks data uji 1.517 × 5.000. Penggunaan *ngram_range* (1,2) memungkinkan sistem membentuk fitur *unigram* dan *bigram* sehingga model dapat mengenali kata tunggal maupun pasangan kata yang muncul pada ulasan.

Sebelum SMOTE, data latih terdiri atas 2.956 ulasan positif dan 3.110 ulasan negatif. Setelah proses SMOTE, jumlah data positif dan negatif menjadi 3.107 pada masing-masing kelas. Penyeimbangan dilakukan hanya pada data latih, sedangkan data uji tetap menggunakan distribusi asli agar evaluasi model dilakukan pada data yang belum pernah digunakan dalam pelatihan.



Sumber: Hasil riset, 2026

Gambar 3. Distribusi Data Latih Sebelum dan Sesudah SMOTE

Penerapan SMOTE bertujuan menambah sampel sintetis pada kelas yang memiliki jumlah lebih sedikit. Distribusi data latih sebelum dan sesudah proses tersebut ditampilkan pada Gambar 3. Penyeimbangan ini ditempatkan setelah pembagian data sehingga data sintetis hanya dibentuk dari data latih. Dengan demikian, data uji tetap berfungsi sebagai data yang terpisah untuk menilai kemampuan model terhadap ulasan yang tidak digunakan dalam pelatihan.

Representasi TF-IDF mengubah kata menjadi bobot numerik berdasarkan frekuensi kemunculannya dalam satu ulasan dan kelangkaannya pada seluruh dokumen. Kata yang sering muncul di hampir seluruh ulasan memperoleh bobot lebih rendah, sedangkan kata yang lebih khas pada ulasan tertentu memperoleh bobot lebih tinggi. Kombinasi *unigram* dan *bigram* juga memungkinkan fitur seperti "gagal login", "tidak buka", atau "mudah transaksi" direpresentasikan sebagai pasangan kata, bukan hanya sebagai kata yang berdiri sendiri.

Jumlah fitur dibatasi menjadi 5.000 agar matriks yang digunakan dalam pelatihan tetap terkelola. Matriks data latih dan data uji memiliki jumlah kolom yang sama karena keduanya menggunakan ruang fitur TF-IDF yang dibentuk dari data latih. Hasil transformasi tersebut menjadi masukan bagi Naive Bayes untuk menghitung probabilitas setiap fitur terhadap kelas sentimen positif dan negatif.

Hasil Pemodelan Naive Bayes

Pengujian dilakukan terhadap empat konfigurasi model. Multinomial Naive Bayes dengan alpha 1,0 memperoleh *accuracy* 85,62% dan *weighted F1-score* 85,43%. Complement Naive Bayes dengan alpha 1,0 menghasilkan nilai yang sama. Multinomial Naive Bayes dengan alpha 0,1 memperoleh *accuracy* 85,56% dan *weighted F1-score* 85,38%, sedangkan alpha 0,01 memperoleh *accuracy* 85,36% dan *weighted F1-score* 85,17%.

Tabel 4. Hasil Pengujian Konfigurasi Naive Bayes

Model	*Accuracy*	*Weighted F1-Score*
Multinomial Naive Bayes (α = 1,0)	85,62%	85,43%
Complement Naive Bayes (α = 1,0)	85,62%	85,43%
Multinomial Naive Bayes (α = 0,1)	85,56%	85,38%
Multinomial Naive Bayes (α = 0,01)	85,36%	85,17%

Sumber: Hasil riset, 2026

Berdasarkan hasil pengujian, Multinomial Naive Bayes dengan *alpha* 1,0 dipilih sebagai model yang digunakan pada evaluasi akhir dan implementasi sistem. Konfigurasi tersebut menghasilkan performa lebih tinggi daripada variasi Multinomial Naive Bayes dengan *alpha* 0,1 dan 0,01. Complement Naive Bayes dengan *alpha* 1,0 memperoleh nilai yang sama, tetapi penelitian menetapkan Multinomial Naive Bayes sebagai model utama sesuai rancangan pemodelan yang digunakan pada data TF-IDF.

Perubahan nilai *alpha* menghasilkan perbedaan performa yang relatif kecil, tetapi kecenderungannya menunjukkan bahwa *alpha* 1,0 memberikan hasil paling baik pada dataset ini. Parameter tersebut berfungsi sebagai *smoothing* untuk mengurangi kemungkinan probabilitas nol ketika suatu fitur tidak muncul pada kelas tertentu. Ketika nilai *alpha* diturunkan menjadi 0,1 dan 0,01, nilai *accuracy* dan *weighted F1-score* ikut menurun, walaupun selisihnya tidak besar.

Hasil Evaluasi Model

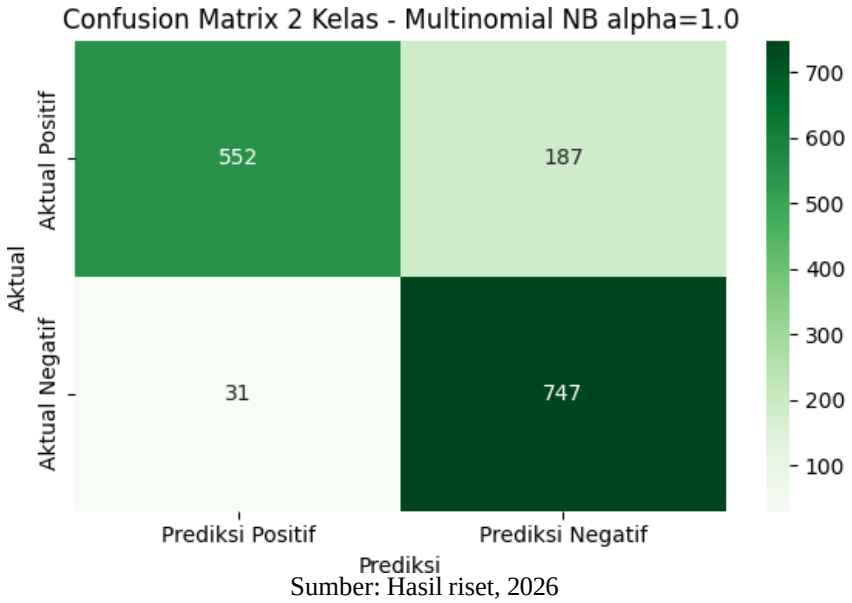
Pengujian terhadap 1.517 data uji menghasilkan *accuracy* sebesar 85,63%, *precision* sebesar 87,14%, *recall* sebesar 85,63%, dan *F1-score* sebesar 85,44%. Nilai tersebut menunjukkan bahwa model mampu memprediksi sebagian besar sentimen pada data uji

dengan benar. Kedekatan nilai *accuracy*, *recall*, dan *F1-score* juga memperlihatkan bahwa performa keseluruhan model tidak hanya bertumpu pada satu metrik evaluasi.

Tabel 5. Hasil Evaluasi Model

Metrik	Nilai
Accuracy	85,63%
Weighted Precision	87,14%
Weighted Recall	85,63%
Weighted F1-Score	85,44%

Sumber: Hasil riset, 2026



Sumber: Hasil riset, 2026

Gambar 4. *Confusion Matrix* Multinomial Naive Bayes

Confusion matrix pada Gambar 4 memperlihatkan bahwa 552 ulasan positif diprediksi dengan benar sebagai positif, 187 ulasan positif diprediksi sebagai negatif, 31 ulasan negatif diprediksi sebagai positif, dan 747 ulasan negatif diprediksi dengan benar sebagai negatif. Secara keseluruhan, model menghasilkan 1.299 prediksi benar dan 218 prediksi salah dari 1.517 data uji.

Jumlah prediksi benar pada kelas negatif lebih besar daripada kelas positif. Dari 778 data negatif, model mampu mengenali 747 data secara benar dan hanya 31 data yang diprediksi sebagai positif. Pada kelas positif, 552 dari 739 data dapat dikenali secara benar, sedangkan 187 data diprediksi sebagai negatif. Pola tersebut memperlihatkan bahwa model lebih kuat dalam mengenali ulasan negatif, terutama ulasan yang memuat kata keluhan yang tegas, dibandingkan ulasan positif yang susunan katanya dapat bercampur dengan istilah bernuansa negatif.

Kesalahan klasifikasi pada kelas positif dapat terjadi ketika ulasan memuat struktur kalimat yang tidak sederhana, negasi, atau kombinasi kata yang tidak cukup kuat memperoleh bobot positif. Sebaliknya, kata seperti “gagal”, “error”, “tidak”, dan “susah” memiliki muatan negatif yang lebih mudah dikenali ketika sering muncul pada data latih. Naive Bayes menentukan kelas berdasarkan probabilitas fitur, sehingga model belum memahami konteks kalimat dengan cara yang sama seperti pembaca manusia.

Tabel 6. Contoh Hasil Prediksi Model

Ulasan (Hasil *Preprocessing*)	Sentimen Asli	Sentimen Prediksi	Keterangan
sangat mudah transaksi cepat	Positif	Positif	Benar
aplikasi error login gagal	Negatif	Negatif	Benar
sering update malah error	Negatif	Positif	Salah
tampilan bagus fitur lengkap	Positif	Positif	Benar

Sumber: Hasil riset, 2026

Contoh hasil prediksi pada Tabel 6 memperlihatkan bahwa ulasan "sangat mudah transaksi cepat" dan "tampilan bagus fitur lengkap" berhasil diprediksi sebagai positif, sedangkan ulasan "aplikasi error login gagal" berhasil diprediksi sebagai negatif. Kesalahan masih ditemukan pada ulasan "sering update malah error" yang diprediksi sebagai positif. Kesalahan tersebut menunjukkan bahwa keberadaan satu kata negatif belum selalu menghasilkan prediksi negatif apabila bobot gabungan fitur lain lebih kuat mengarah pada kelas positif.

Nilai *accuracy* 85,63% memperlihatkan bahwa Naive Bayes dapat digunakan untuk klasifikasi sentimen ulasan Livin' by Mandiri. Hasil ini sedikit lebih tinggi dibandingkan penelitian Komalasari et al. (2026) yang memperoleh akurasi 85,02% menggunakan Naive Bayes dan TF-IDF pada objek yang sama. Selisih tersebut tidak hanya dipengaruhi oleh algoritma, tetapi juga dapat berkaitan dengan jumlah data, periode pengumpulan, tahapan *preprocessing*, mekanisme pelabelan, pembagian dataset, dan konfigurasi model yang digunakan.

Secara keseluruhan, hasil pengujian menjawab tujuan penelitian mengenai performa Naive Bayes berdasarkan *accuracy*, *precision*, *recall*, dan *F1-score*. Model memiliki tingkat ketepatan yang baik pada data uji, tetapi hasil *confusion matrix* memperlihatkan bahwa kemampuan klasifikasi belum sepenuhnya sama pada kedua kelas. Oleh karena itu, nilai agregat perlu dibaca bersama dengan jumlah prediksi benar dan salah pada setiap sentimen.

Analisis Visualisasi *Word Cloud*

Sumber: Hasil riset, 2026

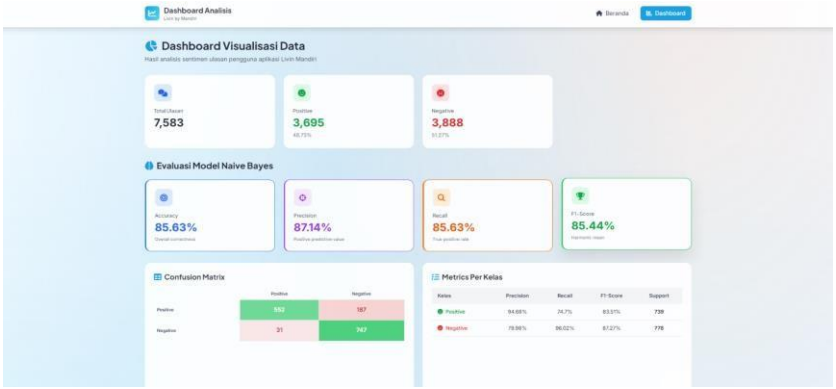
Gambar 5. *Word Cloud* Sentimen Positif dan Negatif

Word cloud pada sentimen positif menampilkan kata yang sering muncul, antara lain "mudah", "livin", "bantu", "cepat", "aman", "lancar", dan "bagus". Kata-kata tersebut menggambarkan kemudahan dan manfaat yang dirasakan pengguna. Pada sentimen negatif, kata yang dominan meliputi "tidak", "buka", "aplikasi", "login", "error", "susah", "gagal", "lelet", dan "ribet", yang menggambarkan kendala akses, autentikasi, stabilitas aplikasi, dan kegagalan transaksi.

Visualisasi *word cloud* membantu menyederhanakan ribuan ulasan menjadi gambaran kata-kata utama yang sering digunakan. Ukuran kata yang lebih besar menunjukkan frekuensi kemunculan yang lebih tinggi. Kata "mudah", "cepat", dan "lancar" memperlihatkan bahwa kemudahan transaksi menjadi bentuk apresiasi utama. Sebaliknya, kata "tidak", "buka", "login", "error", dan "gagal" menunjukkan bahwa akses aplikasi dan keberhasilan transaksi menjadi sumber keluhan yang sering muncul.

Hasil visualisasi tersebut melengkapi evaluasi numerik model karena klasifikasi tidak hanya disajikan sebagai nilai metrik, tetapi juga diterjemahkan menjadi kecenderungan kosakata pengguna. Meskipun demikian, *word cloud* hanya menggambarkan frekuensi kemunculan kata dan tidak menjelaskan hubungan sebab-akibat maupun konteks lengkap setiap ulasan. Oleh sebab itu, interpretasinya digunakan sebagai pendukung untuk mengenali tema umum apresiasi dan keluhan.

Hasil Implementasi *Website*



Sumber: Hasil riset, 2026

Gambar 6. Tampilan *Dashboard* SentimenLivin AI

Model diimplementasikan pada situs SentimenLivin AI menggunakan Python pada bagian *backend* dan HTML, CSS, serta JavaScript pada bagian *frontend*. Situs memiliki bagian *Home*, *About*, *Features*, Coba Analyzer, dan Dasbor Visualisasi Data. Fitur Coba Analyzer memungkinkan pengguna memasukkan ulasan baru secara manual atau mengunggah data dalam bentuk berkas CSV dan Excel untuk memperoleh prediksi sentimen.

Sistem dikembangkan dengan konsep *Single Page Application* menggunakan *top navigation bar*. Dasbor menampilkan total 7.583 ulasan, distribusi sentimen, nilai *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, tabel evaluasi, serta kata dominan pada sentimen positif dan negatif. Implementasi tersebut mengintegrasikan hasil klasifikasi, evaluasi model, dan visualisasi data dalam satu antarmuka.

Pembahasan Hasil Klasifikasi Sentimen

Hasil pembersihan dan pelabelan menghasilkan 7.583 ulasan dua kelas yang terdiri atas 3.695 ulasan positif atau 48,73% dan 3.888 ulasan negatif atau 51,27%. Komposisi ini memperlihatkan bahwa sentimen pengguna relatif seimbang, meskipun proporsi negatif sedikit lebih besar daripada positif. Oleh karena itu, temuan penelitian tidak menunjukkan dominasi sentimen positif, tetapi menggambarkan adanya dua pengalaman pengguna yang hampir berimbang, yaitu apresiasi terhadap manfaat aplikasi dan keluhan terhadap kendala teknis yang masih terjadi.

Sentimen positif terutama berkaitan dengan kemudahan transaksi, kecepatan, keamanan, kelancaran, dan manfaat aplikasi dalam membantu aktivitas perbankan sehari-hari. Kata-kata seperti "mudah", "cepat", "aman", "lancar", "bagus", dan "bantu" yang

muncul pada *word cloud* positif menunjukkan bahwa sebagian pengguna menilai Livin' by Mandiri mampu menyederhanakan proses transfer, pembayaran, pengisian saldo, dan layanan finansial lainnya. Temuan tersebut mencerminkan bahwa integrasi berbagai layanan dalam satu aplikasi telah memberikan efisiensi waktu dan kemudahan akses bagi pengguna.

Di sisi lain, sentimen negatif memuat kata-kata seperti "tidak", "buka", "login", "error", "susah", "gagal", "lelet", dan "ribet". Pola tersebut menunjukkan bahwa keluhan pengguna lebih banyak berhubungan dengan kesulitan mengakses aplikasi, kegagalan autentikasi, lambatnya respons sistem, serta transaksi yang tidak berhasil diproses. Dalam konteks layanan perbankan digital, kendala tersebut menjadi penting karena gangguan pada akses dan transaksi dapat langsung memengaruhi kepercayaan dan kenyamanan pengguna. Hasil ini dapat digunakan sebagai masukan untuk memprioritaskan peningkatan stabilitas aplikasi, keandalan proses login, kapasitas layanan pada waktu sibuk, dan penanganan transaksi gagal.

Perbedaan persepsi pengguna dapat dipengaruhi oleh versi aplikasi, perangkat yang digunakan, kualitas jaringan, waktu akses, dan jenis transaksi. Namun, faktor-faktor tersebut tidak dianalisis karena penelitian dibatasi pada isi teks ulasan. Dengan demikian, hasil klasifikasi menggambarkan kecenderungan opini yang tercermin melalui bahasa pengguna, bukan hubungan langsung antara sentimen dengan karakteristik pengguna atau kondisi teknis tertentu.

Tahap *preprocessing* berperan dalam membentuk data yang lebih konsisten sebelum pelabelan dan pemodelan. *Cleaning* menghapus simbol dan karakter yang tidak mendukung klasifikasi, *case folding* menyeragamkan penggunaan huruf, *tokenizing* memisahkan kalimat menjadi kata, normalisasi mengubah kata tidak baku ke bentuk standar, *stopword removal* mengurangi kata yang kurang informatif, dan *stemming* mengembalikan kata berimbuhan ke bentuk dasar. Normalisasi kata seperti "gk" menjadi "tidak" serta perubahan "memudahkan" menjadi "mudah" membantu sistem mengenali kata bermuatan sentimen secara lebih seragam.

Pelabelan berbasis leksikon memungkinkan kategori sentimen ditentukan berdasarkan kandungan teks, bukan semata-mata berdasarkan rating bintang. Pendekatan ini relevan karena skor bintang tidak selalu sejalan dengan isi ulasan. Meskipun demikian, label yang dihasilkan tetap bergantung pada cakupan kamus dan aturan perhitungan polaritas. Ulasan yang mengandung negasi, sarkasme, atau gabungan apresiasi dan keluhan dapat menghasilkan label yang kurang tepat karena makna kalimat tidak selalu dapat ditentukan hanya dari akumulasi kata positif dan negatif.

Dataset kemudian dibagi menjadi 6.066 data latih dan 1.517 data uji. SMOTE diterapkan pada data latih untuk membentuk distribusi kelas yang seimbang sehingga model dapat mempelajari pola positif dan negatif secara proporsional. Penerapan tersebut dilakukan setelah pembagian data agar sampel sintetis tidak masuk ke dalam data uji. Namun, karena penelitian belum menampilkan perbandingan model sebelum dan sesudah SMOTE, teknik tersebut lebih tepat dipahami sebagai bagian dari skenario pemodelan, bukan sebagai faktor yang telah terbukti secara terpisah meningkatkan akurasi.

Representasi TF-IDF dengan maksimal 5.000 fitur dan rentang *unigram* serta *bigram* membantu mengubah teks menjadi bobot numerik. *Unigram* memungkinkan model mempelajari kata tunggal, sedangkan *bigram* menangkap pasangan kata seperti "gagal login", "tidak buka", dan "mudah transaksi". Penggunaan pasangan kata penting karena makna sentimen sering kali dibentuk oleh hubungan dua kata, sehingga informasi yang diperoleh lebih spesifik dibandingkan penggunaan kata tunggal saja.

Pembahasan Hasil Evaluasi Model

Pengujian pada 1.517 data uji menghasilkan *accuracy* sebesar 85,63%, *precision* tertimbang sebesar 87,14%, *recall* tertimbang sebesar 85,63%, dan *F1-score* tertimbang sebesar 85,44%. Nilai tersebut menunjukkan bahwa Multinomial Naive Bayes mampu mengklasifikasikan sebagian besar ulasan sesuai dengan label acuannya. Kedekatan nilai *accuracy*, *recall*, dan *F1-score* juga memperlihatkan bahwa performa agregat model relatif konsisten dan tidak hanya terlihat baik pada satu metrik.

Confusion matrix mencatat 552 ulasan positif diprediksi benar sebagai positif, 187 ulasan positif diprediksi sebagai negatif, 31 ulasan negatif diprediksi sebagai positif, dan 747 ulasan negatif diprediksi benar sebagai negatif. Dengan demikian, terdapat 1.299 prediksi benar dan 218 prediksi salah. Pola tersebut menunjukkan bahwa model lebih kuat dalam mengenali sentimen negatif daripada sentimen positif. *Recall* kelas negatif mencapai sekitar 96,02%, sedangkan *recall* kelas positif sekitar 74,70%. Sebaliknya, *precision* kelas positif mencapai sekitar 94,68%, yang berarti prediksi positif yang dikeluarkan model umumnya tepat, tetapi sebagian ulasan positif masih terlewat dan dipetakan sebagai negatif.

Kemampuan yang lebih tinggi dalam mendeteksi sentimen negatif dapat dipengaruhi oleh kosakata keluhan yang cenderung tegas dan berulang, seperti “gagal”, “error”, “tidak”, dan “susah”. Kata-kata tersebut memperoleh pola probabilitas yang kuat pada kelas negatif. Ulasan positif dapat ditulis dengan variasi bahasa yang lebih beragam atau memuat kata negatif dalam konteks yang tidak sepenuhnya bernada keluhan. Kondisi ini dapat menyebabkan model lebih berhati-hati dalam memberikan label positif.

Kesalahan prediksi pada contoh “sering update malah error” menunjukkan keterbatasan Multinomial Naive Bayes dalam memahami konteks kalimat. Algoritma menghitung peluang kelas berdasarkan distribusi fitur dan mengasumsikan fitur bersifat independen. Akibatnya, hubungan antarkata, negasi kompleks, sarkasme, dan makna kontekstual belum dapat dipahami seperti pembacaan manusia. Bahasa *slang*, kesalahan pengetikan, serta istilah baru yang belum tercakup dalam kamus normalisasi atau data latih juga dapat memengaruhi ketepatan prediksi.

Multinomial Naive Bayes dengan nilai alpha 1,0 dipilih sebagai model utama karena menghasilkan *accuracy* dan *weighted F1-score* tertinggi dibandingkan variasi alpha 0,1 dan 0,01. Nilai alpha berfungsi sebagai *smoothing* untuk mencegah probabilitas nol ketika suatu fitur tidak ditemukan pada kelas tertentu. Complement Naive Bayes dengan alpha 1,0 memperoleh nilai yang sama, tetapi Multinomial Naive Bayes tetap digunakan sesuai rancangan utama penelitian dan karakteristik data TF-IDF yang diolah.

Akurasi 85,63% sedikit lebih tinggi daripada temuan Komalasari et al. (2026) sebesar 85,02% pada objek yang sama. Selisih tersebut tidak dapat langsung ditafsirkan sebagai keunggulan mutlak karena masing-masing penelitian dapat menggunakan jumlah data, periode pengambilan, mekanisme pelabelan, *preprocessing*, pembagian data, dan konfigurasi model yang berbeda. Meskipun demikian, hasil ini memperkuat bahwa Naive Bayes dengan representasi TF-IDF dapat memberikan performa yang baik untuk klasifikasi sentimen ulasan aplikasi perbankan dengan kebutuhan komputasi yang relatif ringan.

Visualisasi *word cloud* melengkapi metrik evaluasi dengan menunjukkan kata yang paling sering digunakan pada masing-masing kelompok sentimen. Visualisasi tersebut membantu menerjemahkan keluaran klasifikasi menjadi informasi yang lebih mudah dipahami mengenai aspek yang diapresiasi dan dikeluhkan pengguna. Namun, frekuensi kata tidak menjelaskan hubungan sebab-akibat dan tidak selalu menggambarkan konteks lengkap suatu ulasan, sehingga interpretasinya harus digunakan sebagai pendukung, bukan sebagai satu-satunya dasar pengambilan kesimpulan.

Pembahasan Hasil Implementasi Sistem

Model diimplementasikan pada website SentimenLivin AI agar hasil analisis sentimen dapat disajikan secara terintegrasi. Python digunakan pada bagian *backend* untuk membaca data klasifikasi, memuat model Multinomial Naive Bayes dan modul pemrosesan teks, serta menghubungkan proses komputasi dengan antarmuka. Bagian *frontend* dikembangkan menggunakan HTML, CSS, dan JavaScript untuk menampilkan komponen visual dan menyediakan interaksi dengan pengguna.

Sistem menggunakan konsep *Single Page Application* dengan *top navigation bar* sebagai navigasi utama. Konsep ini memungkinkan pengguna berpindah ke setiap bagian melalui *scrolling* atau transisi pada halaman yang sama tanpa *reload*. Bagian *Home* menampilkan identitas proyek dan ringkasan hasil, *About* menjelaskan latar belakang sistem, sedangkan *Features* menyajikan teknologi serta fungsi yang digunakan. Susunan tersebut membuat informasi sistem dapat diakses dalam satu alur tampilan yang ringkas.

Fitur Coba Analyzer menyediakan analisis satuan dan analisis *batch*. Analisis satuan digunakan untuk memproses ulasan yang diketik secara manual, sedangkan mode *batch* digunakan untuk mengunggah kumpulan data dalam format CSV atau Excel. Teks masukan diproses menggunakan tahapan yang sama dengan data penelitian, kemudian ditransformasikan melalui TF-IDF dan diprediksi oleh Multinomial Naive Bayes. Hasil prediksi ditampilkan secara *real-time* sebagai sentimen positif atau negatif.

Dashboard berfungsi sebagai pusat penyajian hasil penelitian. Bagian ini menampilkan jumlah 7.583 ulasan, distribusi positif dan negatif melalui *pie chart* dan *bar chart*, nilai *accuracy*, *precision*, *recall*, dan *F1-score*, tabel evaluasi per kelas, *confusion matrix*, serta visualisasi kata dominan. Integrasi tersebut memudahkan pengguna memahami hasil klasifikasi tanpa harus membaca data mentah atau menjalankan kode secara langsung.

Implementasi SentimenLivin AI menunjukkan bahwa keluaran penelitian tidak berhenti pada eksperimen di Google Colab, tetapi telah diterapkan dalam sistem yang dapat menerima ulasan baru dan menampilkan hasil secara interaktif. Sistem berfungsi sebagai media dokumentasi, pengujian, dan visualisasi, sedangkan keputusan klasifikasi tetap bergantung pada pola yang dipelajari model. Oleh karena itu, kemungkinan kesalahan pada ulasan ambigu, sarkastis, atau menggunakan istilah yang tidak dikenali masih dapat terjadi ketika sistem digunakan.

Secara keseluruhan, penerapan *preprocessing*, pelabelan berbasis leksikon, TF-IDF, SMOTE, dan Multinomial Naive Bayes telah membentuk alur analisis dari data mentah hingga implementasi. Keluaran utama penelitian berada pada klasifikasi dan evaluasi model, sedangkan website berperan sebagai sarana penyajian agar distribusi sentimen, performa model, kata dominan, dan prediksi ulasan baru dapat dipahami secara lebih cepat dan terstruktur. Keterbatasan penelitian terletak pada penggunaan label otomatis, klasifikasi yang hanya terdiri atas dua kelas, belum adanya perbandingan langsung sebelum dan sesudah SMOTE, serta keterbatasan model dalam memahami konteks semantik yang kompleks.

KESIMPULAN

Penelitian ini berhasil menerapkan *preprocessing*, pelabelan berbasis leksikon, TF-IDF, SMOTE, dan Multinomial Naive Bayes untuk mengklasifikasikan sentimen ulasan Livin' by Mandiri pada Google Play Store. Dari 10.000 ulasan awal diperoleh 7.583 ulasan dua kelas yang terdiri atas 3.695 sentimen positif dan 3.888 sentimen negatif, sehingga sentimen negatif memiliki proporsi sedikit lebih besar. Dataset dibagi menjadi 6.066 data latih dan 1.517 data uji serta direpresentasikan menggunakan 5.000 fitur TF-IDF. Multinomial Naive Bayes dengan alpha 1,0 menghasilkan *accuracy* 85,63%, *precision* 87,14%, *recall* 85,63%, dan *F1-score* 85,44%. *Confusion matrix* menunjukkan 1.299 prediksi benar dan 218 prediksi salah, dengan kemampuan pengenalan sentimen negatif yang lebih kuat daripada sentimen

positif. *Word cloud* memperlihatkan bahwa apresiasi pengguna berkaitan dengan kemudahan, kecepatan, keamanan, dan kelancaran, sedangkan keluhan berkaitan dengan kendala membuka aplikasi, login, error, respons lambat, dan kegagalan transaksi. Model kemudian diimplementasikan pada SentimenLivin AI untuk menampilkan prediksi, distribusi sentimen, evaluasi model, dan kata dominan secara terintegrasi.

REFERENSI

Akbar, M. R. M., & Pratama, I. (2025). Sentiment Analysis of MyBCA Application User Reviews Using Naive Bayes, Random Forest, and Decision Tree. *Sistemasi: Jurnal Sistem Informasi*, *14*(5), 2464–2478. https://doi.org/10.32520/stmsi.v14i5.5472

Alalwan, A. A., Dwivedi, Y. K., & Rana, N. P. (2017). Factors Influencing Adoption of Mobile Banking by Jordanian Bank Customers: Extending UTAUT2 with Trust. *International Journal of Information Management*, *37*(3), 99–110. https://doi.org/10.1016/j.ijinfomgt.2017.01.002

Alviyanti, S. H., Purwandira, A., Febiyanti, I., Daniati, E., & Ristyawan, A. (2024). Klasifikasi Sentimen Pengguna Aplikasi Livin by Mandiri pada Playstore Menggunakan Algoritma Naive Bayes. In *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)* (Vol. 8, pp. 1–9). https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/5052

Ariany, F., Tri Prastowo, A., Ahmad, I., & Jafar Adrian, Q. (2026). Algoritma Naïve Bayes dalam Analisis Komperatif Sentimen Dompet Digital. *Jurnal Tekno Kompak*, *20*(1), 364–370. https://doi.org/10.33365/jtk.v20i1.1454

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A Comprehensive Survey on Sentiment Analysis: Approaches, Challenges, and Trends. *Knowledge-Based Systems*, *226*. https://doi.org/10.1016/j.knosys.2021.107134

Dąbrowski, J., Letier, E., Perini, A., & Susi, A. (2022). Analysing App Reviews for Software Engineering: A Systematic Literature Review. *Empirical Software Engineering*, *27*(2). https://doi.org/10.1007/s10664-021-10065-7

Darmawan, G., Alam, S., & Sulistyo, M. I. (2023). Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi MyPertamina pada Google Playstore Menggunakan Metode Naive Bayes. *STORAGE: Jurnal Ilmiah Teknik Dan Ilmu Komputer*, *2*(3), 100–108. https://doi.org/10.55123/storage.v2i3.2333

Ginasputri, H. N. A., Saputri, A. D., Wahid, S. N., Ningrum, A. F., & Haris, M. Al. (2025). Analisis Sentimen Ulasan Pengguna BRImo terhadap Pembaruan Fitur Aplikasi Menggunakan Naive Bayes dengan Seleksi Fitur Chi-Square. *Jurnal Statistika Dan Komputasi*, *4*(2), 56–67. https://doi.org/10.32665/statkom.v4i2.5194

Hariansyah, M. Z., & Siswanto. (2022). Implementasi Metode Multinomial Naive Bayes pada Analisis Sentimen terhadap Layanan Aplikasi Livin by Mandiri. In *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi* (Vol. 1, Issue 1, pp. 517–524). https://senafti.budiluhur.ac.id/senafti/article/view/271

Insan, M. K. K., Hayati, U., & Nurdiawan, O. (2023). Analisis Sentimen Aplikasi BRImo pada Ulasan Pengguna di Google Play Menggunakan Algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 478–483. https://doi.org/10.36040/jati.v7i1.6373

Komalasari, D., Sari, B. N., & Mayasari, R. (2026). Analisis Sentimen Ulasan Aplikasi Livin by Mandiri pada Play Store dengan Algoritma Naive Bayes. *Jurnal Ilmiah Wahana Pendidikan*, *12*(3.A), 14–22. https://jurnal.peneliti.net/index.php/JIWP/article/view/12573

Maalej, W., Kurtanović, Z., Nabil, H., & Stanik, C. (2016). On the Automatic Classification of App Reviews. *Requirements Engineering*, *21*(3), 311–331.

https://doi.org/10.1007/s00766-016-0251-9

Martin, W., Sarro, F., Jia, Y., Zhang, Y., & Harman, M. (2017). A Survey of App Store Analysis for Software Engineering. *IEEE Transactions on Software Engineering*, *43*(9), 817–847. https://doi.org/10.1109/TSE.2016.2630689

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning-Based Text Classification: A Comprehensive Review. *ACM Computing Surveys*, *54*(3), 1–40. https://doi.org/10.1145/3439726

Nirmala, N. M. G. S., & Sanjaya ER, N. A. (2024). Analisis Sentimen dengan Logistic Regression untuk Deteksi Kata pada Livin' by Mandiri. *JNATIA (Jurnal Nasional Teknologi Informasi Dan Aplikasinya)*, *2*(4), 869–878. https://doi.org/10.24843/JNATIA.2024.v02.i04.p25

Nurzaman, N., Suarna, N., & Prihartono, W. (2024). Analisis Sentimen Ulasan Aplikasi Threads di Google Playstore Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 967–974. https://doi.org/10.36040/jati.v8i1.8708

Rizaldi, S. A., Alam, S., & Kurniawan, I. (2023). Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) pada Google Play Store Menggunakan Metode Naive Bayes. *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, *2*(3), 109–117. https://doi.org/10.55123/storage.v2i3.2334

Sayogo, D. S., Irawan, B., & Bahtiar, A. (2023). Analisis Sentimen Ulasan Instagram di Google Play Store Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3314–3319. https://doi.org/10.36040/jati.v7i6.8178

Simarmata, A. A. P., & Sasongko, T. B. (2025). Sentiment Analysis on BRImo Application Reviews Using IndoBERT. *Journal of Applied Informatics and Computing*, *9*(3), 851–862. https://doi.org/10.30871/jaic.v9i3.8162

Souiden, N., Ladhari, R., & Chaouali, W. (2021). Mobile Banking Adoption: A Systematic Review. *International Journal of Bank Marketing*, *39*(2), 214–241. https://doi.org/10.1108/IJBM-04-2020-0182

Suharman, A., & Sulaeman, M. K. (2025). Analisis Sentimen Pengguna Aplikasi Livin' by Mandiri Menggunakan Metode Support Vector Machine (SVM) dengan Ekstraksi Fitur TF-IDF dan Word2Vec. *Jurnal Pendidikan dan Teknologi Indonesia*, *5*(8), 2201–2212. https://doi.org/10.52436/1.jpti.941

Tobing, A. J., & Febriandirza, A. (2024). Analisis Sentimen Aplikasi Mobile Bangking BCA pada Ulasan Pengguna di Google Play Store Menggunakan Metode Naive Bayes. *Journal of Information System Research*, *5*(4), 998–1005. https://doi.org/10.47065/josh.v5i4.5485

Umair, M., & Susanto, E. R. (2024). Analisis Sentimen Ulasan Pengguna pada Aplikasi BRImo BRI Menggunakan Metode Klasifikasi Algoritma Naive Bayes. *Jurnal Media Informatika Budidarma*, *8*(2), 1149–1159. https://doi.org/10.30865/mib.v8i2.7381

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A Survey on Sentiment Analysis Methods, Applications, and Challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1