JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

Analisis Sentimen Ulasan Pengguna Aplikasi Mobile Banking Bank Kalbar Menggunakan Algoritma Svm Random Forest dan Naive Bayes

Theofilus BA[1*], Panny Agustia Rahayuningsih[2], Muhammad Rezki[3]

Program Studi Informatika, Universitas Bina Sarana Informatika, Pontianak, Kalimantan Barat, Indonesia
Email: [1]15220520@bsi.ac.id, [2]Panny.par@bsi.ac.id, [3]muhammad.mdk@bsi.ac.id
(*Email Corresponding Author: 15220520@bsi.ac.id)
Received: 25 Juni 2026 | Revision: 30 Juni 2026 | Accepted: 1 Juli 2026

Abstrak

Perkembangan dalam layanan perbankan seluler telah meningkatkan jumlah ulasan pengguna di platform seperti Google Play Store, yang menyimpan data berharga terkait kepuasan dan keluhan dari nasabah. Penelitian ini memiliki tujuan untuk menganalisis sentimen dari ulasan pengguna aplikasi mobile banking Bank Kalbar dengan memanfaatkan tiga algoritma machine learning untuk klasifikasi, yaitu Support Vector Machine (SVM), Random Forest, dan Naive Bayes, yang diterapkan dengan representasi fitur berbasis TF-IDF. Data yang digunakan dalam penelitian ini mencakup 2.923 ulasan yang diperoleh melalui web scraping dan secara otomatis diberi label melalui model LLaMA 3.1 8B Instant menggunakan Groq API ke dalam tiga kategori sentimen: positif (49,3%), negatif (41,2%), dan netral (9,5%). Ketidakseimbangan dalam distribusi kelas ditangani dengan metode SMOTE, sedangkan optimisasi hyperparameter dilakukan dengan teknik GridSearchCV menggunakan Stratified 5-Fold Cross Validation. Hasil evaluasi menunjukkan bahwa SVM berperforma sebagai model paling unggul dengan akurasi 87,52%, F1-Weighted 86,80%, dan F1-Macro 79,59%, diikuti oleh Naive Bayes (85,98%) dan Random Forest (85,47%). Namun, analisis statistik McNemar mengindikasikan bahwa perbedaan kinerja di antara ketiga model tidak signifikan secara statistik ($p > 0,05$), yang menunjukkan bahwa ketiga algoritma ini memiliki kemampuan generalisasi yang setara pada dataset ini. Tantangan utama yang teridentifikasi adalah rendahnya tingkat recall pada kelas netral (41–46%) disebabkan oleh ketidakseimbangan data. Hasil dari penelitian ini memberikan wawasan yang jelas mengenai pola sentimen dari pengguna aplikasi mobile banking Bank Kalbar dan dapat menjadi referensi bagi manajemen untuk meningkatkan kualitas layanan digital, terutama dalam hal stabilitas aplikasi dan kemudahan akses.

Kata Kunci: Analisis Sentimen, Mobile Banking, Support Vector Machine, Random Forest, Naive Bayes, TF-IDF, Bank Kalbar.

Abstract

The rise of mobile banking options has resulted in a growing number of user evaluations on sites like the Google Play Store, which hold important insights about user satisfaction and grievances. This research intends to examine the sentiment expressed in user feedback for the Bank Kalbar mobile banking app by employing three machine learning classification methods: Support Vector Machine (SVM), Random Forest, and Naive Bayes, utilizing a TF-IDF feature representation. The dataset comprises 2,923 reviews obtained through web scraping and automatically categorized into three sentiment types—positive (49.3%), negative (41.2%), and neutral (9.5%)—using the LLaMA 3.1 8B Instant model via the Groq API. To tackle the issue of class imbalance, the SMOTE technique was applied, and hyperparameter tuning was performed through GridSearchCV alongside Stratified 5-Fold Cross Validation. The evaluation outcomes indicate that SVM achieved the highest accuracy at 87.52%, with an F1-Weighted score of 86.80% and an F1-Macro measure of 79.59%, succeeding Naive Bayes (85.98%) and Random Forest (85.47%). However, the McNemar statistical analysis reveals that the differences in performance between the three models are not statistically significant ($p > 0.05$), pointing to similar generalization capabilities for all three algorithms with this dataset. The primary challenge noted was the low recall rate for the neutral class (41–46%), attributed to data imbalance. The results of this research offer a clear understanding of user sentiment trends regarding the Bank Kalbar mobile banking application and can assist management in enhancing digital service quality, particularly in terms of application stability and user accessibility.

Keywords: *Sentiment Analysis, Mobile Bankin, Support Vector Machine, Random Forest, Naive Bayes, TF-IDF, Bank Kalbar.*

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

1. PENDAHULUAN

Dunia keuangan di Indonesia telah terintegrasi sepenuhnya dengan kemajuan teknologi digital. Salah satu manifestasi paling jelas dari transformasi ini adalah kehadiran layanan perbankan seluler yang sekarang telah menjadi bagian tak terpisahkan dari aktivitas sehari-hari jutaan pengguna. Dengan smartphone di tangan, pengguna dapat melakukan pengiriman uang, membayar tagihan, memeriksa saldo, serta berbagai transaksi lainnya kapan saja dan di mana saja tanpa harus mengantri di bank atau mesin ATM. Perubahan ini bukan hanya soal kenyamanan, tetapi juga memberikan dampak yang signifikan terhadap efektivitas sektor perbankan itu sendiri. Pramitasari dan Nanggala menunjukkan dengan bukti empiris bahwa bank di Indonesia yang menerapkan sistem perbankan mobile menunjukkan kinerja yang lebih baik dan memiliki risiko keuangan yang lebih minim, bahkan perbankan seluler terbukti memainkan peran krusial dalam menjaga stabilitas finansial di masa-masa sulit [1]. Situasi ini mendorong hampir semua lembaga keuangan di Indonesia, termasuk bank-bank daerah, untuk terus meningkatkan layanan digital mereka guna memenuhi ekspektasi nasabah yang semakin tinggi.

Salah satu lembaga keuangan daerah yang berperan dalam dunia perbankan digital adalah Bank Pembangunan Daerah Kalimantan Barat, dikenal sebagai Bank Kalbar. Sebagai bank yang dimiliki oleh Pemerintah Provinsi Kalimantan Barat, Bank Kalbar telah meluncurkan aplikasi mobile banking yang dapat diakses melalui Google Play Store atau App Store. Dengan aplikasi ini, pelanggan bisa menikmati berbagai layanan mulai dari cek saldo, transfer antar rekening, pembayaran beragam tagihan, pembelian pulsa, sampai pembayaran pajak seperti PBB di berbagai daerah di Kalimantan Barat. Seiring dengan meningkatnya jumlah pengguna, jumlah ulasan di Google Play Store juga mengalami pertumbuhan. Ulasan-ulasan ini mencerminkan berbagai pendapat dari para pengguna; beberapa memberikan pujian atas kemudahan fitur, sementara yang lain mengeluhkan masalah teknis, dan ada juga yang hanya menawarkan pertanyaan atau saran. Seperti yang dinyatakan oleh Arifiyanti dan rekan-rekannya, ulasan dari pengguna adalah umpan balik yang perlu dianalisis oleh pihak pengembang, agar dapat dijadikan sebagai dasar untuk meningkatkan dan memperbaiki kualitas aplikasi[2]. Namun, tantangannya adalah bahwa membaca dan merangkum ratusan hingga ribuan ulasan teks secara manual tentu sangat tidak efisien dan memakan banyak waktu.

Di sinilah analisis sentimen muncul sebagai solusi yang tepat. Analisis sentimen adalah metode dalam Natural Language Processing (NLP) yang memungkinkan mesin untuk secara otomatis membaca, memahami, dan mengelompokkan teks ke dalam kategori opini tertentu, seperti positif, negatif, atau netral. Dalam dunia aplikasi perbankan digital, metode ini telah banyak digunakan dan terbukti ampuh. Misalnya, Sujjada dan rekan-rekannya melakukan penelitian terhadap tiga aplikasi bank digital terkenal, yaitu Bank Jago, Neobank, dan Seabank dengan metode SVM yang menggunakan pembobotan TF-IDF, berhasil meraih akurasi rata-rata 91%, di mana Seabank tercatat mencapai persentase tanggapan positif tertinggi di kalangan penggunanya [3]. Penelitian lain oleh Pratama dan timnya yang membandingkan SVM dengan pendekatan berbasis leksikon pada ulasan BRImo dan BCA Mobile juga menunjukkan hasil yang menggembirakan, dengan akurasi mencapai 94% untuk BRImo dan 95% untuk BCA Mobile [4].

Tidak terbatas pada satu aplikasi, analisis sentimen juga dilakukan secara perbandingan pada beberapa platform mobile banking sekaligus. Munandar dan rekannya mengevaluasi ulasan dari tiga aplikasi mobile banking paling terkenal di Indonesia, yaitu BRImo, BSI Mobile, dan Livin' by Mandiri menggunakan algoritma K-Nearest Neighbor, dan menemukan bahwa BRImo memperoleh sentimen positif tertinggi sebesar 58,25% [5]. Di sisi algoritma tunggal, Bhatara dan Suryono meneliti ulasan dari aplikasi BCA Mobile dengan menggunakan Naïve Bayes dan SVM, yang masing-masing didukung oleh teknik SMOTE untuk mengatasi masalah ketidakseimbangan data. Hasil penelitian menunjukkan bahwa SVM mencapai tingkat akurasi 85%, sedangkan Naïve Bayes 83%, dan meskipun perbedaannya kecil, keduanya menunjukkan variasi dalam kemampuannya untuk mendeteksi sentimen positif dan negatif dari ulasan pengguna [6]. Selain itu, Kurniawan dan Wijaya yang melakukan studi terhadap ulasan aplikasi Blu BCA dengan Naïve Bayes meraih akurasi 85,31%, serta mengungkapkan bahwa meskipun sebagian besar ulasan bersifat positif, keluhan tentang kinerja aplikasi dan keamanan data tetap cukup terlihat [7].

Saat ketiga metode utama, yaitu SVM, Random Forest, dan Naïve Bayes, diuji bersamaan dalam satu desain eksperimen, hasilnya menunjukkan pola yang menarik. Mola dan rekan-rekan membandingkan ketiga metode ini pada ulasan aplikasi Halo BCA yang terdaftar di Google Play Store yang terdiri dari total 6.313 ulasan yang dikelompokkan ke dalam tiga kategori sentimen. Random Forest menunjukkan performa terbaik dengan akurasi 90,01% dan nilai F1 0,90, sedangkan

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

SVM berada di urutan kedua dengan akurasi 86,86%, dan Naïve Bayes berada di urutan ketiga dengan 81,42% [8]. Temuan ini mencerminkan bahwa tidak ada satu metode pun yang selalu lebih baik di seluruh aspek, sehingga perbandingan yang sistematis dan menyeluruh dalam kerangka eksperimen yang konsisten sangatlah penting untuk dilakukan.

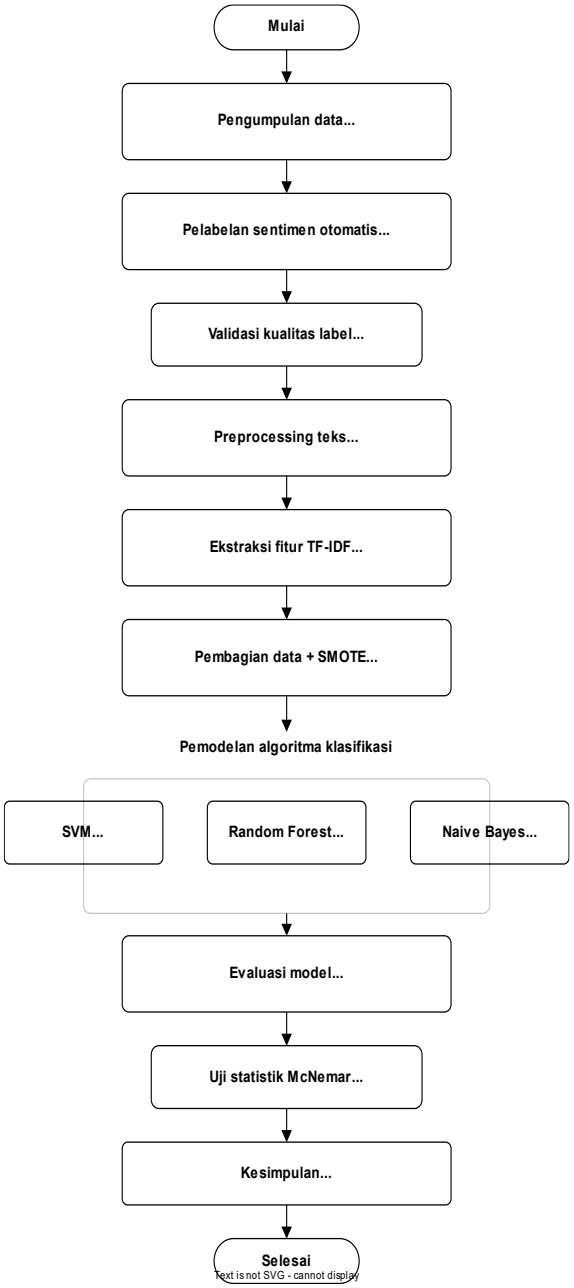
Berdasarkan analisis mengenai berbagai studi sebelumnya, terdapat beberapa pertanyaan yang masih terbuka. Pertama, belum ada penelitian yang secara spesifik meneliti ulasan pengguna aplikasi perbankan seluler Bank Kalbar sebagai institusi keuangan daerah di Kalimantan Barat. Studi-studi yang ada umumnya lebih berkaitan dengan bank-bank nasional besar, sehingga karakteristik pengguna dan pola sentimen yang unik dari bank daerah belum pernah diteliti dengan mendalam. Kedua, banyak penelitian sebelumnya hanya melakukan perbandingan antara satu atau dua algoritma, sementara perbandingan tiga algoritma, yaitu SVM, Random Forest, dan Naïve Bayes, dalam satu penelitian yang konsisten dan adil masih jarang dilaksanakan, terutama dalam konteks mobile banking bank daerah. Ketiga, penelitian ini mengadopsi tiga kategori sentimen, yaitu positif, negatif, dan netral, yang dianggap lebih mencerminkan variasi ekspresi pengguna dalam bahasa Indonesia sehari-hari dibandingkan dengan klasifikasi sederhana positif-negatif yang umum digunakan dalam penelitian sebelumnya.

Berdasarkan latar belakang serta celah yang ada dalam penelitian ini, penelitian ini memiliki tujuan sebagai berikut: (1) melakukan analisis sentimen terhadap tanggapan pengguna aplikasi mobile banking Bank Kalbar yang diambil dari Google Play Store; (2) melakukan perbandingan kinerja antara tiga algoritma yaitu SVM, Random Forest, dan Naïve Bayes dalam mengklasifikasikan sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral; (3) menentukan algoritma mana yang menghasilkan akurasi dan F1-score tertinggi dalam dataset yang digunakan; dan (4) menciptakan rekomendasi berbasis data yang dapat digunakan oleh manajemen Bank Kalbar untuk memahami aspek layanan digital mana yang perlu mendapatkan perhatian lebih berdasarkan pola sentimen nyata dari pengguna.

2.METODOLOGI PENELITIAN

Penelitian ini menerapkan metode kuantitatif dengan desain eksperimen perbandingan untuk mengevaluasi efektivitas tiga algoritma klasifikasi dalam machine learning, yaitu Support Vector Machine (SVM), Random Forest, dan Naïve Bayes, dalam menganalisis sentimen dari ulasan para pengguna aplikasi mobile banking Bank Kalbar. Proses penelitian dilakukan melalui tujuh langkah yang dilakukan secara berurutan, dimulai dari pengumpulan data ulasan menggunakan teknik web scraping, pelabelan sentimen otomatis dengan memanfaatkan Large Language Model (LLM), pengolahan data teks, ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), penanganan ketidakseimbangan kelas dengan teknik Synthetic Minority Over-sampling Technique (SMOTE), pembentukan model dengan ketiga algoritma tersebut, hingga penilaian kinerja model dan penerapan uji statistik McNemar untuk memastikan bahwa perbedaan yang terobservasi dalam kinerja memiliki makna statistik yang valid. Rangkaian penelitian ini dapat dilihat secara keseluruhan pada Gambar 1.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331



Gambar 1. Alur Penelitian

2.1 Pengumpulan dan Preprocessing Data

Data yang digunakan dalam penelitian ini berasal dari komentar pengguna terkait aplikasi mobile banking Bank Kalbar yang terdapat di Google Play Store. Pengumpulan data dilakukan dengan metode web scraping yang berbasis Python melalui library google-play-scraper menggunakan App ID com.xlink. xmobilebankingadkalbar. Elemen-elemen yang diambil mencakup konten (teks dari ulasan), skor (rating antara 1 hingga 5), tanggal ulasan, dan nama pengguna. Penting untuk dicatat bahwa penilaian bintang tidak digunakan secara langsung sebagai indikator sentimen. Hal ini disebabkan oleh fakta bahwa pengguna sering kali memberikan nilai rendah karena faktor eksternal seperti masalah sinyal yang tidak mencerminkan kualitas aplikasi yang sebenarnya[9]. Oleh karena itu, proses pelabelan sentimen dilakukan secara otomatis dengan memanfaatkan Large Language Model (LLM).

Sebelum dimasukkan ke dalam model, teks melalui lima tahap pemrosesan awal secara berurutan. Pertama, dilakukan case folding untuk menyamakan teks menjadi huruf kecil. Kedua, dilakukan pembersihan untuk menghapus

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

karakter yang tidak relevan seperti URL, tanda baca, simbol, dan emoji. Ketiga, dilakukan tokenisasi untuk membagi kalimat menjadi satuan kata. Keempat, dilakukan penghilangan stopword menggunakan pustaka PySastrawi untuk menyingkirkan kata-kata umum yang tidak berarti. Kelima, dilakukan stemming memakai algoritma Nazief-Adriani lewat PySastrawi untuk mengubah kata yang terikat menjadi bentuk dasarnya, misalnya "pembayaran" menjadi "bayar" dan "mentransfer" jadi "transfer".

Tabel 1. Dataset

Nama_pengguna	Ulasan	Rating
Delfi Ramadhani	sudah bagus, tapi better di upgrade dibagian e-wallet disediakan tambahan langsung bisa kedana. Terimakasih	5
Syamsul Arifin	jelek	1
Dinda Auni	pas mau login susah ini kenapa min... pas bagian informasi tekan 'ok' malah ga gerak².. sms belum terima lahh padahal di sms disuruh balik buat regestrasi.. apk tidak menaggapi	3
Fransiska Anita	Sering gangguan.. tolong perbaiki..	5
Bonny Markus	host tidak aktif? mksdnya gmn y??? mau transaksi banking mobile di saat lagi urgent jdi terhalang!	3

2.2 Pelabelan Data

Pelabelan sentimen dalam studi ini dilakukan secara otomatis menggunakan Large Language Model (LLM), khususnya LLaMA 3.1 8B Instant yang dibuat oleh Meta AI dan tersedia melalui Groq API. LLM adalah sistem kecerdasan buatan dengan arsitektur Transformer yang menerapkan mekanisme self-attention, memungkinkan pemahaman mendalam terhadap konteks kalimat [10]. Groq sendiri adalah layanan inference berkinerja tinggi yang memanfaatkan chip Language Processing Unit (LPU), memungkinkan pemrosesan ribuan data dengan efisien. Dalam penelitian ini, LLaMA 3.1 dijalankan dengan teknik zero-shot prompting, di mana model menerima prompt sistem yang mencakup definisi tiga kategori sentimen (POSITIF, NEGATIF, NETRAL) tanpa memerlukan contoh pelatihan tambahan. Metode ini terbukti berhasil untuk tugas klasifikasi sentimen sederhana, karena LLM memiliki keunggulan dibandingkan model tradisional dalam konteks zero-shot [11].

Pemanfaatan LLM sebagai pengganti anotator manual membawa berbagai keuntungan jika dibandingkan dengan sistem pelabelan yang menggunakan bintang atau pelabelan tradisional. Beberapa dari keunggulan tersebut meliputi pemahaman konteks yang lebih baik, hasil yang lebih konsisten karena menerapkan perintah yang sama untuk setiap data, serta kemampuan untuk menangani volume data yang besar dengan cara yang dapat ditingkatkan. Namun, metode ini juga memiliki beberapa kekurangan, termasuk ketergantungan pada keandalan API dan risiko bias yang mungkin terdapat dalam model tersebut.

Tabel 2. Data Pelabelan

No	Nama Pengguna	Teks Bersih	Label Sentimen
1	Delfi Ramadhani	sudah bagus tapi better di upgrade dibagian ewallet disediakan tambahan langsung bisa kedana terimakasih	Positif
2	Syamsul Arifin	jelek	Negatif
3	Dinda Auni	pas mau login susah ini kenapa min pas bagian informasi tekan ok malah tidak gerak sms belum terima lahh padahal di sms disuruh balik buat regestrasi apk tidak menaggapi	Negatif
4	Fransiska Anita	sering gangguan tolong perbaiki	Negatif

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

No	Nama Pengguna	Teks Bersih	Label Sentimen
5	Bonny Markus	host tidak aktif mksdnya gmn y mau transaksi banking mobile di saat lagi urgent jdi terhalang	Negatif

2.3 Preprocessing Teks

Pra-pemrosesan teks adalah langkah yang diambil untuk mempersiapkan data teks mentah agar dapat diproses dalam kegiatan penambangan teks, meliputi pengelolaan stopwords, stemming, dan normalisasi yang berdampak pada ketepatan model[12]. Langkah-langkah yang digunakan dalam studi ini disajikan dalam tabel berikut.

Tabel 3. Tahap Pemprosesan

Tahap	Deskripsi	Contoh
Case Folding	Mengonversi seluruh teks menjadi huruf kecil untuk menyeragamkan variasi penulisan (Rifaldi dkk., 2023)	"BANK KALBAR" → "bank kalbar"
Cleaning	Menghapus karakter tidak relevan seperti URL, tanda baca, simbol, dan *emoji* (Rifaldi dkk., 2023)	"hebat!! 👍 " → "hebat"
Tokenisasi	Memecah kalimat menjadi unit kata individual (Rifaldi dkk., 2023)	"bank kalbar bagus" → ["bank", "kalbar", "bagus"]
Stopword Removal	Menghapus kata umum yang tidak bermakna menggunakan *Sastrawi stoplist* (Rifaldi dkk., 2023)	["aplikasi", "ini", "bagus"] → ["aplikasi", "bagus"]
Stemming	Mereduksi kata berimbuhan ke bentuk dasar menggunakan algoritma Nazief-Adriani via PySastrawi (Rifaldi dkk., 2023)	"pembayaran" → "bayar"
Normalisasi	Mengganti kata tidak baku dan singkatan menggunakan kamus slang Indonesia	"ga" → "tidak", "gws" → "get well soon"

2.4 TF-IDF (Term Frequency–Inverse Document Frequency)

TF-IDF (Term Frequency–Inverse Document Frequency merupakan cara untuk menggambarkan teks dengan pendekatan statistik yang membantu mengubah teks jadi vektor angka dengan menggabungkan dua indikator: Frekuensi Term (TF) yang menilai seberapa sering suatu kata muncul dalam satu dokumen, dan Frekuensi Dokumen Invers (IDF) yang menilai seberapa jarang sebuah kata di seluruh kumpulan dokumen. Angka tinggi pada TF-IDF menunjukkan kata tersebut unik dan kaya informasi, sedangkan angka nol menunjukkan kata terlalu umum sehingga tidak mampu membedakan di antara dokumen-dokumen tersebut[13].

Sebagai contoh, pada dokumen D1 "aplikasi yang menarik dan gampang dipakai", D2 "aplikasi sering kali mengalami kesalahan tidak dapat masuk", dan D3 "aplikasi yang baik transfernya cepat", kata "aplikasi" menghasilkan TF-IDF = 0 karena hadir di ketiga dokumen itu, sementara kata "menarik" memiliki TF-IDF ≈ 0,183 karena hanya muncul di D1 ([13]). Dalam penelitian ini, TF-IDF digunakan setelah teks mengalami pra-pemrosesan dan hasilnya dijadikan fitur untuk algoritma klasifikasi ([14]).

2.5 Pembagian Data, SMOTE, dan Pemodelan Klasifikasi

Dataset dibagi dengan proporsi 80:20 antara data pelatihan dan data pengujian. Rasio ini sering digunakan dalam studi klasifikasi teks karena dapat menyediakan data pelatihan yang cukup, sembari mempertahankan data pengujian yang

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

representatif [15]. Mengingat bahwa distribusi kelas pada dataset ulasan aplikasi sering kali tidak seimbang, teknik SMOTE (Synthetic Minority Oversampling Technique) diterapkan. SMOTE menghasilkan sampel sintetis pada kelas yang lebih sedikit melalui interpolasi antara sampel yang berdekatan, sehingga distribusi kelas lebih seimbang tanpa hanya menduplikasi data [15]. Optimasi hyperparameter dilakukan dengan GridSearchCV dan 5-fold Stratified Cross Validation untuk memastikan rasio kelas tetap terjaga di setiap fold. Untuk SVM, parameter yang disesuaikan adalah C dan kernel; untuk Random Forest, parameter yang diatur adalah n_estimators dan max_depth; sedangkan untuk Naive Bayes, nilai alfa disesuaikan [16]. Sebelum evaluasi akhir, stabilitas model diperiksa menggunakan Stratified K-Fold Cross Validation dengan 5 fold. Pendekatan ini memastikan bahwa setiap fold memiliki proporsi kelas yang representatif, sehingga hasil evaluasi tidak akan bergantung pada satu pembagian data tertentu [17].

Proses pemodelan dilakukan dengan tiga algoritma. SVM dengan kernel Linear/RBF dan strategi One-vs-Rest berusaha menemukan hyperplane yang optimal untuk memisahkan kelas[18]. Random Forest memanfaatkan teknik bagging dengan voting mayoritas dari berbagai pohon keputusan untuk mengurangi risiko overfitting [19]. Naive Bayes tipe Multinomial dengan Laplace smoothing dipilih karena cocok untuk data yang memiliki frekuensi kata seperti TF-IDF [20].

2.6 Evaluasi Model

Tahap penilaian model bertujuan untuk menentukan sejauh mana kemampuan model klasifikasi dalam memprediksi sentimen dengan akurasi. Penilaian dilakukan dengan menggunakan beberapa metrik yang berasal dari confusion matrix, yaitu tabel yang merangkum hasil prediksi model dengan membandingkan label yang diprediksi dengan label yang sebenarnya melalui empat elemen: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) [21]. Dari confusion matrix ini, dihitung empat metrik penilaian utama sebagai berikut: Akurasi mengukur proporsi semua prediksi yang benar dibandingkan total seluruh data uji. Metrik ini memberikan gambaran umum tentang kinerja model, namun kurang mencerminkan ketepatan ketika distribusi kelas tidak merata [15]. Presisi (Precision) menghitung seberapa banyak proporsi prediksi positif yang benar-benar positif. Metrik ini sangat krusial saat perlu mengurangi kesalahan prediksi positif palsu [21]. Recall mengukur seberapa besar proporsi data positif aktual yang berhasil dikenali oleh model. Recall menjadi krusial ketika kesalahan prediksi negatif palsu (false negative) memiliki dampak yang signifikan [15]. F1-Score adalah rata-rata harmonik antara presisi dan recall, sehingga memberikan suatu keseimbangan di antara keduanya. Metrik ini sangat dianjurkan untuk dataset yang tidak seimbang karena dapat mewakili kinerja model lebih adil dibandingkan hanya menggunakan akurasi [21]. Keempat metrik ini digunakan secara serentak untuk mendapatkan gambaran evaluasi yang menyeluruh terhadap ketiga model klasifikasi (SVM, Random Forest, dan Naive Bayes) yang dicoba dalam penelitian ini.

2.7 Uji Statistik McNemar

Uji statistik McNemar diterapkan untuk dengan objektif memeriksa apakah ada perbedaan yang signifikan dalam kinerja antara algoritma yang sedang dibandingkan. Metode ini merupakan teknik statistik non-parametrik yang menyoroti ketidakcocokan prediksi berpasangan antara dua model pada dataset yang sama, menjadikannya ideal untuk menilai perbedaan dalam tingkat kesalahan dari dua klasifikasi [22]. Uji McNemar telah secara luas diterapkan pada banyak aplikasi praktis seperti pengklasifikasian gambar, pengenalan suara, dan analisis sentimen [23]. Proses pengujian ini berlandaskan pada tabel kontingensi 2x2 yang mengilustrasikan kesepakatan dan ketidakcocokan prediksi kedua model. Hipotesis yang digunakan meliputi H_0: tidak ada perbedaan kinerja yang signifikan, dan H_1: terdapat perbedaan kinerja yang signifikan. Statistik pengujian mengikuti distribusi chi-square dengan tingkat signifikansi $p = 0,05$; apabila $p < 0,05$ maka H_0 akan ditolak [24].

3.HASIL DAN PEMBAHASAN

3.1 Hasil Pelabelan Data

Prosedur penandaan otomatis yang memanfaatkan model LLaMA 3.1 8B Instant dengan Groq API telah berhasil mengelompokkan semua ulasan ke dalam tiga kelompok emosi. Dari 2.923 ulasan yang berhasil diberi label, pembagian emosi yang dihasilkan ditampilkan dalam Tabel 4.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

Tabel 4. Distribusi Sentimen Hasil Pelabelan LLM

Sentimen	Jumlah	Persentase (%)
Positif	1.441	49,3
Negatif	1.205	41,2
Netral	277	9,5
Total	**2.923**	**100**

Berdasarkan Tabel 3, ulasan dengan sentimen positif mencapai 1.441 (49,3%), diikuti dengan sentimen negatif yang mencatat 1.205 ulasan (41,2%), sementara sentimen netral menjadi kategori yang paling sedikit dengan 277 ulasan (9,5%). Mayoritas sentimen positif menunjukkan bahwa secara keseluruhan, pengguna cenderung memberikan reaksi yang cukup memuaskan terhadap aplikasi mobile banking Bank Kalbar. Namun, tingginya proporsi sentimen negatif yang mencapai 41,2% menandakan adanya isu yang perlu ditangani dengan serius oleh pengembang [2]. Ketidakseimbangan dalam distribusi kelas, terutama untuk kelas netral yang hanya sebesar 9,5%, menjadi suatu tantangan dalam proses klasifikasi, sehingga penerapan teknik SMOTE diperlukan untuk menyeimbangkan data latihan sebelum pemodelan dilakukan [17].

3.2 Hasil Evaluasi Model

Uji coba dilaksanakan pada tiga algoritma klasifikasi menggunakan 585 sampel data yang terdiri dari 20% dari keseluruhan dataset. Hasil analisis kinerja ketiga model dapat dilihat di Tabel 4.

Tabel 5. Hasil Evaluasi Model

Model	Akurasi (%)	Precision-W (%)	Recall-W (%)	F1-W (%)	F1-Macro (%)
SVM	87,52	87,61	87,52	86,80	79,59
Naive Bayes	85,98	85,93	85,98	85,16	77,27
Random Forest	85,47	85,52	85,47	84,86	78,40

Berdasarkan Tabel 4, SVM muncul sebagai model paling unggul dengan tingkat akurasi 87,52%, Precision-Weighted 87,61%, Recall-Weighted 87,52%, F1-Weighted 86,80%, dan F1-Macro 79,59%. Keunggulan SVM dalam klasifikasi teks dengan fitur tinggi seperti TF-IDF telah banyak diulas dalam berbagai penelitian, karena SVM dapat menentukan hyperplane terbaik yang memaksimalkan jarak antara kelas di ruang berdimensi tinggi [18]. Naive Bayes menempati urutan kedua dengan akurasi 85,98%, sedangkan Random Forest mencatatkan akurasi terendah sebesar 85,47%, meskipun perbedaan antara ketiga model sangat kecil.

Untuk mendapatkan pemahaman yang lebih menyeluruh, Tabel 5 menyuguhkan laporan klasifikasi secara rinci untuk setiap kategori dari tiap model.

Tabel 6. Classification Report per Kelas

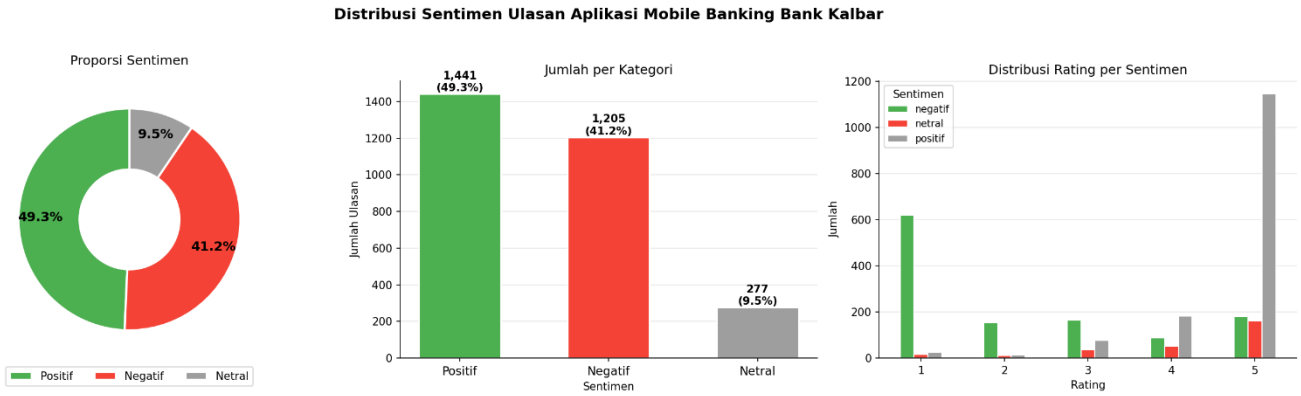
Model	Kelas	Precision	Recall	F1-Score	Support
SVM	Negatif	0,87	0,91	0,89	241
	Netral	0,89	0,45	0,60	56
	Positif	0,88	0,93	0,90	288

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

Model	Kelas	Precision	Recall	F1-Score	Support
Random Forest	Negatif	0,85	0,88	0,87	241
	Netral	0,87	0,46	0,60	56
	Positif	0,86	0,91	0,88	288
Naive Bayes	Negatif	0,85	0,89	0,87	241
	Netral	0,85	0,41	0,55	56
	Positif	0,87	0,92	0,90	288

Berdasarkan Tabel 5, ketiga model menunjukkan hasil yang baik secara konsisten pada kelas positif dan negatif, namun menghadapi tantangan pada kelas netral. Untuk SVM, recall untuk kelas netral hanya mencapai 0,45, yang menunjukkan bahwa hanya 25 dari 56 sampel netral yang sebenarnya dapat diprediksi dengan akurat. Pola yang serupa terlihat pada Random Forest dengan recall netral 0,46, serta Naive Bayes yang mencatat recall netral terendah sebesar 0,41. Rendahnya kinerja pada kelas netral disebabkan oleh jumlah sampel netral yang jauh lebih sedikit dibandingkan dengan kelas lainnya, sehingga model kesulitan dalam menemukan pola yang cukup representatif untuk secara konsisten membedakan kelas netral [21].

Hasil prediksi dari tiga model tersebut ditampilkan secara lebih rinci dalam bentuk confusion matrix pada Gambar 2.



Gambar 2. confusion matrix

Berdasarkan Gambar 2, analisis pada SVM menunjukkan bahwa dari 241 sampel negatif yang ada, 220 berhasil terprediksi dengan akurat, sementara 21 lainnya salah identifikasi sebagai positif. Dalam kategori positif, dari total 288 sampel aktual, 267 terprediksi dengan tepat dan hanya 18 salah diklasifikasikan sebagai negatif. Kesalahan yang paling signifikan terjadi pada kategori netral, di mana dari 56 sampel yang ada, 16 salah diprediksi sebagai negatif dan 15 sebagai positif. Metode Random Forest dan Naive Bayes memperlihatkan pola kesalahan yang sama, dengan sebagian besar kesalahan berada pada kategori netral yang memiliki kesamaan fitur secara semantik dengan kedua kategori lainnya [8].

3.3 Uji Statistik McNemar

Untuk menjamin bahwa perbedaan kinerja antar model bukan hanya disebabkan oleh kebetulan, diuji statistik McNemar dengan tingkat signifikansi $\alpha = 0,05$. Pengujian ini membandingkan ketidakcocokan prediksi yang berpasangan antara dua model pada dataset yang serupa [22]. Hasil dari pengujian ditampilkan pada Tabel 7.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

Tabel 7. Hasil Uji Statistik McNemar

Model A	Model B	Statistik	P-Value	Signifikan
SVM	Random Forest	3,7812	0,0518	Tidak
SVM	Naive Bayes	1,6410	0,2002	Tidak
Random Forest	Naive Bayes	0,0678	0,7946	Tidak

Berdasarkan Tabel 6, hasil dari pengujian McNemar menunjukkan bahwa hipotesis nol diterima untuk semua pasangan model, karena semua nilai p-value lebih tinggi daripada ambang signifikansi 0,05. Perbandingan antara SVM dan Random Forest memberikan nilai p = 0,0518 yang sangat dekat dengan batas signifikansi, namun tetap tidak cukup untuk menolak hipotesis nol. Perbandingan SVM dengan Naive Bayes menghasilkan p = 0,2002, sementara perbandingan Random Forest dengan Naive Bayes menunjukkan p = 0,7946, yang mengindikasikan kesetaraan tertinggi di antara keduanya [23].

Temuan ini menunjukkan bahwa meskipun SVM menunjukkan akurasi tertinggi secara numerik, perbedaan kinerjanya dibandingkan dengan Random Forest dan Naive Bayes tidak signifikan secara statistik. Ketiga algoritma ini menunjukkan kemampuan generalisasi yang sebanding pada dataset ulasan aplikasi mobile banking Bank Kalbar. Oleh karena itu, pemilihan algoritma dalam konteks ini sebaiknya mempertimbangkan faktor lain seperti efisiensi komputasi dan kemudahan dalam penafsiran [8].

3.4 Perbandingan dengan Penelitian Sebelumnya

Untuk menilai letak penelitian ini dalam konteks literatur yang ada, hasil optimal yang diperoleh dibandingkan dengan studi-studi sebelumnya yang relevan. Perbandingan ditampilkan pada Tabel 8.

Tabel 8. Perbandingan dengan Penelitian Terdahulu

Peneliti	Metode	Domain	Kelas	Akurasi Terbaik
Sujjada dkk. [3]	SVM + TF-IDF	Bank Jago, Neobank, Seabank	3	91%
Pratama dkk. [4]	SVM + Lexicon Based	BRImo & BCA Mobile	3	94–95%
Munandar dkk. [5]	K-Nearest Neighbor	BRImo, BSI, Mandiri	3	82,9%
Bhatara & Suryono [6]	SVM + NB + SMOTE	BCA Mobile	3	85% (SVM)
Kurniawan & Wijaya [7]	Naive Bayes	Blu BCA	3	85,31%
Mola dkk. [8]	SVM + RF + NB	Halo BCA	3	90,01% (RF)
Penelitian ini	**SVM + RF + NB + TF-IDF + SMOTE**	**Bank Kalbar**	**3**	**87,52% (SVM)**

Berdasarkan data di Tabel 7, akurasi SVM dalam studi ini mencapai 87,52%, yang lebih tinggi dibandingkan hasil yang diperoleh oleh Bhatara dan Suryono [6] yang mencatatkan akurasi 85% dengan SVM pada ulasan BCA Mobile, serta Kurniawan dan Wijaya [7] yang mendapatkan akurasi 85,31% melalui Naive Bayes pada ulasan Blu BCA. Hal tersebut menunjukkan bahwa proses yang digunakan dalam penelitian ini, yang mengintegrasikan TF-IDF, SMOTE, GridSearchCV, dan Stratified K-Fold Cross Validation, berhasil menciptakan kinerja yang kompetitif dibandingkan dengan penelitian sejenis di bidang mobile banking.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

Namun, akurasi dalam studi ini masih berada di bawah pencapaian Pratama dkk. [4] yang meraih antara 94–95% dan Sujjada dkk. [3] yang mencapai 91%. Ketidaksamaan ini bisa dijelaskan dengan melihat karakteristik dari data ulasan aplikasi mobile banking di bank daerah, seperti Bank Kalbar, yang cenderung menggunakan bahasa yang lebih informal, singkatan tidak resmi, dan variasi ekspresi yang lebih kaya dibandingkan dengan bank besar nasional, sehingga membuat proses klasifikasi lebih sulit [2]. Di sisi lain, jika dibandingkan dengan penelitian Mola dkk. [8] yang juga melakukan perbandingan antara SVM, Random Forest, dan Naive Bayes secara bersamaan, hasil penelitian ini menunjukkan bahwa SVM berperan sebagai model yang paling efektif, sementara dalam studi Mola dkk. , Random Forest menonjol. Perbedaan ini menegaskan bahwa tidak ada satu algoritma yang selalu superior di semua dataset, dan kinerja yang optimal sangat dipengaruhi oleh karakteristik data yang digunakan [8].

4. KESIMPULAN

Penelitian ini menerapkan analisis sentimen pada 2.923 ulasan pengguna aplikasi mobile banking Bank Kalbar dari Google Play Store, mengklasifikasikan sentimen menjadi tiga kategori: positif (49,3%), negatif (41,2%), dan netral (9,5%). Pelabelan otomatis menggunakan LLM LLaMA 3.1 8B Instant via Groq API terbukti efektif untuk memproses data besar secara konsisten dan skalabel. Perbandingan tiga algoritma klasifikasi menunjukkan SVM sebagai model terbaik dengan akurasi 87,52%, F1-Weighted 86,80%, dan F1-Macro 79,59%. Naive Bayes berada di urutan kedua dengan akurasi 85,98%, diikuti Random Forest dengan akurasi 85,47%. Keunggulan SVM pada teks berfitur tinggi berbasis TF-IDF selaras dengan temuan di penelitian sejenis pada domain mobile banking [8]. Namun, uji McNemar menunjukkan perbedaan performa antar model tidak signifikan secara statistik ($p > 0,05$), yang menandakan kemampuan generalisasi ketiganya setara pada dataset ini. Tantangan utama adalah rendahnya recall pada kelas netral (sekitar 41–46%) akibat distribusi kelas yang tidak seimbang. Penerapan SMOTE meningkatkan kemampuan model mengenali semua kelas sentimen, tetapi kelas netral tetap menjadi kendala yang perlu ditangani pada penelitian berikutnya. Berdasarkan peta distribusi sentimen, manajemen Bank Kalbar disarankan memprioritaskan perbaikan pada isu yang memicu sentimen negatif, seperti gangguan teknis, masalah login, dan ketidakstabilan aplikasi yang mendominasi ulasan negatif. Penelitian lanjutan dapat mengeksplorasi pendekatan berbasis transformer seperti IndoBERT untuk menangani nuansa bahasa informal yang lebih kompleks, serta memperluas analisis ke platform lain seperti App Store untuk memperoleh gambaran sentimen yang lebih komprehensif.

REFERENCES

[1] T. D. Pramitasari and A. Y. A. Nanggala, “Dampak Mobile Banking Terhadap Kinerja Dan Stabilitas Keuangan Perbankan di Indonesia,” *Jurnal Manajemen dan Bisnis Indonesia*, vol. 9, no. 2, pp. 241–252, 2023, doi: 10.32528/jmbi.v9i2.855.

[2] A. A. Arifiyanti, N. R. Shantika, and A. O. Syafira, “Analisis Sentimen Ulasan Pengguna Bsi Mobile Pada Google Play Dengan Pendekatan Supervised Learning,” *Jurnal Informatika Polinema*, vol. 9, no. 3, pp. 283–288, 2023, doi: 10.33795/jip.v9i3.1003.

[3] Sujjada et al., “Sentiment Analysis of Digital Bank Reviews on Google Play Store Using the Support Vector Machine (SVM) Method,” *Jurnal Rekayasa Teknologi Nusa Putra*, vol. 9, no. 2, pp. 122–135, 2023.

[4] M. R. Pratama, Y. R. Ramadha, and M. A. Komara, “Analisis Sentimen BRImo dan BCA Mobile Menggunakan Support Vector Machine dan Lexicon Based,” *Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 12, no. 3, p. 1439, 2023, doi: 10.35889/jutisi.v12i3.1431.

[5] D. Munandar, M. Afdal, Z. Zarnelly, and R. Novita, “Analisis Sentimen Ulasan Pengguna Aplikasi Mobile Banking Menggunakan Algoritma K-Nearest Neighbor,” *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 7, no. 3, pp. 1309–1318, 2024, doi: 10.32493/jtsi.v7i3.41409.

[6] R. R. S. Dimas Wahyu Bhatara, “Analisis Sentimen Aplikasi BCA Mobile Menggunakan Algoritma Naive Bayes dan Suport Vector Machine,” *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.

[7] Wahyudi, R. Kurniawan, and Y. A. Wijaya, “Playstore Mengunakan Algoritma Naïve Bayes (Studi Kasus Sentimen Pengguna Terhadap Pengalaman Aplikasi Blu Bca),” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 2511–2517, 2024.

[8] S. A. S. Mola, D. L. B. Baun, I. O. Nunes, and M. M. A. R. Sani, “Analisis Sentimen Aplikasi Halo Bca Di Google Play Store Menggunakan Metode Naive Bayes, Support Vector Machine Dan Random Forest,” *HOAQ (High*

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1252-1263
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/ 10.62712/juktisi.v5i2.1331

Education of Organization Archive Quality) : Jurnal Teknologi Informasi, vol. 15, no. 2, pp. 69–79, 2024, doi: 10.52972/hoaq.vol15no2.p69-79.

[9] M. G. Al Hakim and F. Irwiensyah, "Analisis Sentimen Terhadap Ulasan Pengguna Pada Aplikasi BCA Mobile Menggunakan Metode Naïve Bayes," *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 911–921, 2024, doi: 10.47065/josh.v5i4.5343.

[10] H. Touvron *et al.*, "LLaMA: Open and Efficient Foundation Language Models," 2023, [Online]. Available: http://arxiv.org/abs/2302.13971

[11] W. Zhang, Y. Deng, B. Liu, S. J. Pan, and L. Bing, "Sentiment Analysis in the Era of Large Language Models: A Reality Check," *Findings of the Association for Computational Linguistics: NAACL 2024 - Findings*, pp. 3881–3906, 2024, doi: 10.18653/v1/2024.findings-naacl.246.

[12] D. Rifaldi and A. Fadlil, "DECODE: Jurnal Pendidikan Teknologi Informasi TEKNIK PREPROCESSING PADA TEXT MINING MENGGUNAKAN DATA TWEET 'MENTAL HEALTH,'" *DECODE: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 161–171, 2023.

[13] A. Addiga and S. Bagui, "Sentiment Analysis on Twitter Data Using Term Frequency-Inverse Document Frequency," *Journal of Computer and Communications*, vol. 10, no. 08, pp. 117–128, 2022, doi: 10.4236/jcc.2022.108008.

[14] J. Hunt, "Machine Learning in Python," vol. 12, pp. 633–642, 2023, doi: 10.1007/978-3-031-40336-1_56.

[15] S. N. Almuayqil, M. Humayun, N. Z. Jhanjhi, M. F. Almufareh, and D. Javed, "Framework for Improved Sentiment Analysis via Random Minority Oversampling for User Tweet Review Classification," *Electronics (Switzerland)*, vol. 11, no. 19, pp. 1–17, 2022, doi: 10.3390/electronics11193058.

[16] Achmad Baroqah Pohan, Irmawati, and A. Kurniasih, "Optimization of Classification Algorithm with GridSearchCV and Hyperparameter Tuning for Sentiment Analysis of the Nusantara Capital City," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 3, no. 3, pp. 808–814, 2024, doi: 10.59934/jaiea.v3i3.514.

[17] S. F. Kadir and A. Fairuzabadi, "Analisis Sentimen Ulasan Shopee di Google Play dengan TF-IDF dan Logistic Regression," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 2, pp. 7940–57945, 2025, doi: 10.31004/riggs.v4i2.2850.

[18] R. Obiedat *et al.*, "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution," *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.

[19] I. Setiawan, A. M. Widodo, M. Rahaman, M. B. Ulum, E. Y. Mulyani, and N. Erzed, *Proceedings of the First Mandalika International Multi-Conference on Science and Engineering 2022, MIMSE 2022 (Informatics and Computer Science)*, vol. 2. Atlantis Press International BV, 2022. doi: 10.2991/978-94-6463-084-8.

[20] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information (Switzerland)*, vol. 10, no. 4, pp. 1–68, 2019, doi: 10.3390/info10040150.

[21] M. Heydarian and T. E. Doyle, "MLCM : Multi-Label Confusion Matrix," pp. 19083–19095, 2022.

[22] R. Wang and J. Li, "Block-regularized 5$\times$2 Cross-validated McNemar's Test for Comparing Two Classification Algorithms," vol. 14, no. 8, pp. 1–10, 2023, [Online]. Available: https://arxiv.org/abs/2304.03990v1

[23] J. A. Roldán-Nofuentes, T. S. Sheth, and J. F. Vera-Vera, "Hypothesis Test to Compare Two Paired Binomial Proportions: Assessment of 24 Methods," *Mathematics*, vol. 12, no. 2, 2024, doi: 10.3390/math12020190.

[24] D. Anggraini, I. Gamayanto, and S. Wibowo, "Comparing Decision Tree and Support Vector Machines in Hospital Satisfaction," *Journal of Applied Informatics and Computing*, vol. 9, no. 2, pp. 364–372, 2025, doi: 10.30871/jaic.v9i2.9203.