

Sentiment Analysis of Mobile Banking Reviews Using Naïve Bayes and Support Vector Machine on Play Store

Dewi Masitoh^{1*}, R. Rhoedy Setiawan², Soni Adiyono³

^{1,2,3}Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus, Kudus, Indonesia

Email: ¹*dewimsthhh@gmail.com, ²rhoedy.setiawan@umk.ac.id, ³soni.adiyono@umk.ac.id

(*: corresponding author)

Abstract - The increasing adoption of mobile banking has generated a large volume of user reviews on the Google Play Store, providing valuable insights into customer satisfaction with digital banking services. This study analyzes user sentiment toward four mobile banking applications BRimo, Livin' by Mandiri, BCA Mobile, and SeaBank—and compares the performance of the Naïve Bayes and Support Vector Machine algorithms for sentiment classification. A dataset of 40,000 Play Store reviews was collected through web scraping. Reviews were labeled based on user ratings, followed by preprocessing, TF-IDF feature extraction, and sentiment classification. Model performance was evaluated using the train-test split method and 5-fold cross-validation. The results indicate that Naïve Bayes outperformed and demonstrated greater stability than SVM across all evaluation metrics, achieving an accuracy of 89.23%, precision of 89.76%, recall of 89.23%, and F1-score of 89.37%, while SVM achieved an accuracy of 88.39%. Sentiment analysis revealed that SeaBank received the highest number of positive reviews, whereas BCA Mobile recorded the highest proportion of negative sentiment. These findings provide a comparative evaluation of Naïve Bayes and SVM for mobile banking sentiment classification and offer practical guidance for selecting appropriate classification algorithms while supporting the continuous improvement of digital banking services through user feedback.

Keywords: Google Play Store, Mobile banking, Naïve Bayes, Sentiment Analysis, Support Vector Machine, TF-IDF.

Abstrak - Meningkatnya penggunaan *mobile banking* telah menghasilkan sejumlah besar review aplikasi di Play Store yang diberikan oleh pengguna serta berfungsi sebagai alat untuk menilai tingkat kepuasan pengguna pada layanan perbankan digital. Fokus penelitian diarahkan analisis sentimen yang diberikan pengguna terhadap aplikasi perbankan digital, yaitu BRimo, Livin' by Mandiri, BCA Mobile, dan SeaBank serta mengevaluasi kinerja algoritma Naïve Bayes dan Support Vector Machine dalam klasifikasi sentimen. Penelitian menggunakan 40.000 ulasan hasil *scraping* dari Play Store. Data diberi label berdasarkan rating pengguna, kemudian melalui tahapan *preprocessing* dan ekstraksi fitur memakai TF-IDF sebelum dilakukan klasifikasi. Evaluasi model dilakukan memakai *train-test split* dan *5-fold cross validation*. Temuan penelitian ini mengindikasikan bahwa algoritma Naïve Bayes dan Support Vector Machine memiliki kinerja lebih unggul dan stabil daripada Support Vector Machine pada metrik evaluasi secara keseluruhan. Naïve Bayes memperoleh akurasi sebesar 0.8923, presisi 0.8976, recall 0.8923, dan F1-score 0.8937, sedangkan Support Vector Machine mendapatkan akurasi sebesar 0.8839. Analisis sentimen menunjukkan SeaBank memiliki tingkat kepuasan tertinggi dengan 8.745 ulasan positif, sedangkan BCA Mobile memiliki sentimen negatif tertinggi. Penelitian ini berkontribusi dengan menyajikan evaluasi komparatif algoritma Naïve Bayes dan Support Vector Machine pada klasifikasi ulasan *mobile banking* sehingga memberikan referensi dalam pemilihan algoritma klasifikasi yang sesuai. Temuan ini diharapkan dapat membantu pengembangan layanan perbankan digital dalam memanfaatkan umpan balik pengguna untuk meningkatkan kualitas layanan secara berkelanjutan.

Kata Kunci: Analisis Sentimen, Google Play Store, Perbankan Seluler, Naïve Bayes, Support Vector Machine, TF-IDF.

1. PENDAHULUAN

Berbagai inovasi teknologi yang terus berkembang telah membawa perubahan atas banyak sektor, termasuk pada sektor perbankan. Dalam upaya meningkatkan kualitas layanan sesuai kebutuhan nasabah yang makin beragam, perbankan mulai mengadopsi teknologi digital dalam berbagai aktivitas layanan yang sebelumnya dilakukan secara konvensional [1]. Salah satunya adalah layanan aplikasi *mobile banking* yang memudahkan pengguna untuk menjalankan berbagai aktivitas transaksi secara langsung menggunakan *smartphone* [2]. Peningkatan penggunaan layanan *mobile banking* menunjukkan bahwa masyarakat semakin bergantung pada teknologi digital dalam melakukan aktivitas keuangan karena dianggap lebih praktis, cepat, dan efisien dibandingkan layanan perbankan konvensional [3]. Sejalan dengan hal tersebut, perkembangan layanan perbankan digital juga mendukung pencapaian *Sustainable Development Goals*, salah satunya pada aspek SDG 8 (*Decent Work and Economic Growth*) melewati peningkatan inklusi serta efisiensi layanan keuangan, serta SDG 9 (*Industry, Innovation and Infrastructure*) melalui dorongan terhadap transformasi digital serta inovasi dalam sektor perbankan [4]. Keberhasilan implementasi layanan teknologi digital dipengaruhi oleh kesiapan teknologi dan penerimaan pengguna terhadap sistem yang digunakan [5, 6].

Tingginya penggunaan aplikasi *mobile banking* menyebabkan persaingan antar bank dalam menghadirkan layanan digital yang berkualitas semakin meningkat [7]. Setiap bank terus berupaya mengembangkan fitur, meningkatkan keamanan sistem, serta memperbaiki kualitas layanan aplikasi agar mampu memberikan pengalaman penggunaan yang optimal bagi nasabah [8]. Beberapa aplikasi *mobile banking* yang memiliki tingkat

popularitas tinggi dan jumlah pengguna besar di Indonesia antara lain BRI Mo, Livin' by Mandiri, BCA Mobile, dan SeaBank. Tingginya jumlah pengguna pada aplikasi tersebut menghasilkan banyak data ulasan pengguna Play Store. Ulasan tersebut umumnya memuat opini, persepsi, kritik, dan saran terkait standar layanan pada aplikasi, misalnya antarmuka aplikasi yang mudah digunakan, kecepatan dalam memproses transaksi, performa sistem yang stabil, proteksi aplikasi, serta kendala teknis yang dialami selama penggunaan aplikasi[9]. Informasi dari pengguna memiliki nilai penting karena mencerminkan tingkat kepuasan secara langsung berdasarkan pengalaman nyata pengguna [10]. Ulasan pengguna dapat digunakan sebagai bahan evaluasi layanan *mobile banking* [11].

Banyaknya data ulasan pengguna menjadikan proses analisis manual kurang efektif untuk dilakukan[12]. Data ulasan juga bersifat tidak terstruktur sehingga sulit dianalisis tanpa metode pengolahan teks yang tepat. Analisis manual juga memiliki kelemahan karena bersifat subjektif dan bergantung pada interpretasi peneliti[13] [14], analisis sentimen dipakai untuk mengatasi hal tersebut, yaitu penerapan metode *text mining* berfungsi agar menganalisis persepsi pengguna berdasarkan data teks[15]. Melalui analisis sentimen, opini pengguna dapat diketahui melalui klasifikasi sentimen positif dan negatif agar dapat diketahui kecenderungan persepsi pengguna terhadap kualitas layanan aplikasi yang dipakai[16].

Riset mengenai analisis sentimen umumnya memanfaatkan metode *machine learning* dalam klasifikasi teks secara otomatis. Meskipun sudah banyak penelitian yang membahas analisis persepsi pengguna melalui ulasan aplikasi *mobile banking* maupun aplikasi digital lainnya memakai algoritma Naïve Bayes serta Support Vector Machine, umumnya masih berorientasi satu jenis pada aplikasi atau belum melakukan perbandingan antar aplikasi *mobile banking* di Indonesia secara komprehensif, serta menunjukkan perbedaan hasil akurasi akibat variasi *preprocessing*, fitur ekstraksi, dan karakteristik dataset [17] sehingga masih terdapat *research gap* terkait konsistensi performa algoritma dan kebutuhan evaluasi yang lebih menyeluruh. Mengacu pada kondisi tersebut, penelitian kali ini membandingkan Naïve Bayes serta Support Vector Machine menggunakan dataset berjumlah 40.000 ulasan empat aplikasi *mobile banking* yang dipilih karena mewakili layanan *mobile banking* yang banyak digunakan di Indonesia. Perbandingan antar aplikasi diharapkan memberikan gambaran sentimen pengguna yang lebih komprehensif sekaligus mengevaluasi performa algoritma pada karakteristik yang berbeda [18].

Algoritma *machine learning* dipilih karena memiliki kapabilitas mengatasi dataset dalam kuantitas tinggi serta menghasilkan klasifikasi sentimen prediksi akurasi yang optimal [19]. Studi ini menerapkan algoritma Naïve Bayes serta Support Vector Machine. Pendekatan Naïve Bayes dibangun berdasarkan Teorema Bayes dengan independensi pada setiap fitur [20] yang unggul dalam kemudahan penggunaan, efisiensi, serta klasifikasi teks [21]. Support Vector Machine mengklasifikasikan data dengan menentukan *hyperplane* optimal sebagai pemisah antar kelas [22]. Penelitian ini berkontribusi dengan menyajikan evaluasi komparatif kedua algoritma menggunakan dataset yang relatif besar dari empat aplikasi *mobile banking* di Indonesia sehingga dapat menjadi referensi dalam pemilihan algoritma klasifikasi sentimen. Penggunaan kedua algoritma tersebut diharapkan mampu membandingkan performa klasifikasi untuk menentukan algoritma dengan akurasi terbaik dalam analisis ulasan *mobile banking* [23].

Melalui penelitian ini, dilakukan kajian terhadap data sentimen guna mengetahui kecenderungan opini yang disampaikan para pengguna *mobile banking* BRI Mo, Livin' by Mandiri, BCA Mobile serta SeaBank pada Play Store melalui Naïve Bayes serta Support Vector Machine diterapkan untuk melakukan klasifikasi, diharapkan dapat memberikan informasi performa kedua algoritma serta menjadi masukan bagi pihak perbankan dalam melakukan peningkatan kualitas layanan aplikasi *mobile banking* guna menciptakan pengalaman pengguna yang lebih optimal

2. METODE PENELITIAN

Penelitian ini menerapkan penerapan analisis sentimen dalam mengidentifikasi persepsi pengguna mengenai layanan *mobile banking* yaitu, BRI Mo, Livin' by Mandiri, BCA Mobile, dan SeaBank melalui data ulasan pada Play Store. Proses penelitian dilakukan dengan memanfaatkan konsep *text mining* serta *machine learning* agar membedakan sentimen yang diberikan pengguna mengenai kelas positif maupun negatif. Naïve Bayes dan Support Vector Machine diterapkan saat tahap klasifikasi, melalui proses representasi fitur mengadopsi pembobotan TF-IDF. Tahapan penelitian dilakukan rangkaian tahapan berupa perolehan data, pelabelan, *preprocessing*, *data split*, pembobotan TF-IDF, klasifikasi sentimen dan evaluasi model memakai *confusion matrix*. Tujuan penelitian untuk mengukur performa akurasi algoritma Naive Bayes dan Support Vector Machine saat klasifikasi sentimen serta memberikan informasi terkait tingkat pengguna merasa puas mengenai layanan *mobile banking*.

2.1. Pengumpulan Data

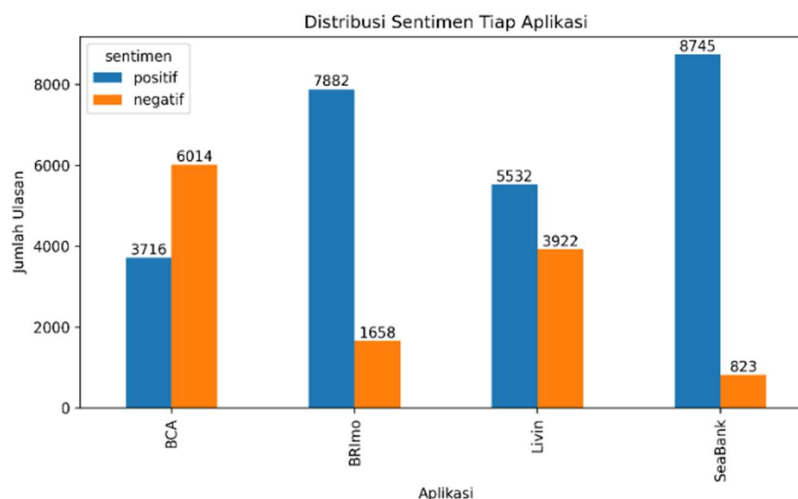
Penelitian dimulai melalui proses pengambilan *review* pengguna BRI Mo, Livin by Mandiri, BCA mobile, dan SeaBank melalui *platform* Play Store. Data ulasan pengguna diperoleh dengan memanfaatkan metode *scraping* memakai *library google-play-scraper* pada visual code guna mengekstraksi ulasan pengguna. Data yang dipakai berupa *username*, rating, tanggal ulasan, serta isi komentar pengguna. Penelitian ini menggunakan data sebanyak

10.000 ulasan yang diperoleh dari keempat aplikasi *mobile banking* tersebut dengan total keseluruhan 40.000 data. Penggunaan data ulasan dari Play Store dipilih karena platform tersebut menyediakan opini langsung dari pengguna secara aktual terhadap layanan *mobile banking* yang digunakan.

Proses *scraping* dilakukan secara otomatis menggunakan bahasa pemrograman Python dengan mengambil ulasan berbahasa Indonesia (*lang* = "id") dan wilayah Indonesia (*country* = "id"), menggunakan urutan ulasan terbaru (*Sort.NEWEST*). Seluruh rating pengguna (1-5) disertakan dalam proses pengambilan data, sedangkan ulasan yang kosong tidak dimasukkan ke dalam. Setelah data berhasil dikumpulkan, penyimpanan dataset dilakukan dalam format CSV agar mengoptimalkan proses pengolahan data ditahap selanjutnya. Dataset mentah yang diperoleh masih mengandung berbagai elemen yang tidak memiliki relevansi dalam data, seperti simbol, ikon, tautan, serta kata-kata yang tidak sesuai sehingga memerlukan tahap *preprocessing* sebelum dilakukan analisis sentimen.

2.2. Pelabelan Data

Tahap selanjutnya dilakukan proses pemberian label sentimen terhadap data ulasan pengguna aplikasi *mobile banking*. Pelabelan sentimen didasarkan pada rating pengguna, rating 4 serta 5 rating tinggi dan dikategorikan positif karena menunjukkan kepuasan, rating 1,2 dan 3 dikategorikan negatif karena mencerminkan ketidakpuasan pengguna. Setelah proses pelabelan selesai dilakukan, diperoleh distribusi sentimen positif serta negatif di tiap aplikasi *mobile banking*. Hasil distribusi sentimen berfungsi menjadi variabel target di tahap pelatihan dan pengujian model klasifikasi. Hasil distribusi sentimen berdasarkan pelabelan rating ditunjukkan pada Gambar 1.



Gambar 1. Distribusi Perbandingan Sentimen Tiap Aplikasi

Berdasarkan Gambar 1, distribusi sentimen hasil pelabelan menunjukkan bahwa Seabank memiliki jumlah data berlabel positif tertinggi yaitu sebanyak 8.745 ulasan, sedangkan jumlah berlabel negatif sebanyak 823 ulasan. BRImo memiliki 7.882 ulasan berlabel positif dan 1.658 ulasan berlabel negatif. Livin by Mandiri memiliki 5.532 ulasan positif dan 3.922 ulasan negatif. Sementara itu, BCA Mobile memiliki 3.716 ulasan positif dan 6.014 ulasan negatif, sehingga menjadi aplikasi dengan jumlah data berlabel negatif terbanyak di antara aplikasi yang dianalisis. Hasil pelabelan tersebut kemudian dijadikan sebagai kelas acuan pada tahap pelatihan serta evaluasi model klasifikasi. Distribusi hasil pelabelan menunjukkan bahwa jumlah data sentimen positif serta negatif di tiap aplikasi tidak seimbang (*class imbalance*). Karakteristik dataset ini perlu diperhatikan karena distribusi kelas yang tidak merata berpotensi memengaruhi proses pembelajaran model pada tahap klasifikasi.

2.3. Preprocessing

Preprocessing dilakukan untuk meningkatkan kualitas teks dengan mengurangi *noise* sebelum tahap klasifikasi menggunakan *machine learning* [24]. Beberapa tahapan dipakai pada *preprocessing*, *case folding*, *cleaning*, normalisasi kata tidak baku, *tokenizing*, *stopword removal*, dan *stemming*. *Case folding* berfungsi untuk menyamakan format kalimat dengan mengganti huruf menjadi *lower case* [23, 24]. Setelah dilakukan tahapan *cleaning* menghapus angka numerik, tanda baca, simbol, emoji, dan link URL dan karakter yang tidak memiliki keterkaitan dalam proses pengolahan teks. Selanjutnya dilakukan normalisasi kata untuk menghasilkan kosakata baku dengan mengonversi kata tidak baku serta bahasa slang. Selanjutnya dilakukan normalisasi kata, *stopword*

removal, dan stemming menggunakan library sastrawi untuk menghasilkan kata dasar sebelum proses klasifikasi [25]

2.4. Split Data

Dataset hasil tahap *preprocessing* kemudian diterapkan pada tahap pemisahan data latih serta data uji melalui teknik *train-test split*. 80% data pelatihan (*training data*), 20% Data uji (*testing data*) selebihnya dipakai. Penggunaan parameter *random_state* diterapkan pada tahap pembagian dataset untuk menjaga konsistensi hasil pengujian model meskipun proses pemisahan data dilakukan secara acak.

2.5. Pembobotan TF-IDF

Teks *preprocessing* yang dihasilkan selanjutnya direpresentasikan dikonversi menjadi vektor angka melalui pembobotan TF-IDF agar dapat menjadi masukan pada proses klasifikasi menggunakan *machine learning* [27]. TF-IDF membobotkan dari frekuensi dan tingkat kepentingan kata dalam dataset [28]. Penelitian ini, proses ekstraksi fitur menggunakan TF-IDF menerapkan parameter *max_features* = 5000 dan *ngram_range* = (1,2) untuk merepresentasikan kata tunggal (*unigram*) dan pasangan kata (*bigram*) ke dalam bentuk vektor numerik sebagai masukan bagi model klasifikasi

2.6. Klasifikasi Sentimen

Naïve Bayes serta Support Vector Machine dipakai sebagai klasifikasi sentimen. Naïve Bayes digunakan sebab mampu melakukan klasifikasi teks secara efektif dengan proses yang sederhana dan efisien [29], sedangkan Support Vector Machine diterapkan sebagai pembanding karena efektif dalam mengelola teks berdimensi tinggi [30]. Dalam penelitian, Naïve Bayes memakai metode Multinomial Naïve Bayes ($\alpha=1,0$), sedangkan model Support Vector Machine menggunakan linier *support vector classification* (LinearSVC) dengan parameter $C = 1,0$ dan *loss* = *squared hinge*. Parameter lainnya menggunakan nilai bawaan (*default*) dari *library scikit-learn*. Kedua model dilatih dengan fitur TF-IDF dan *preprocessing*. Kemudian model dipakai agar memproyeksikan kategori sentimen pada data uji sehingga diperoleh hasil akhir klasifikasi berupa sentimen positif dan negatif

2.7. Evaluasi Model

Model dievaluasi pada tahap akhir penelitian memakai *confusion matrix* dan metrik *accuracy*, *precision*, *recall*, dan *f1-score* dipakai menkinerja klasifikasi dengan komprehensif. Akurasi berperan dalam menilai akurasi klasifikasi model. *Precision* mengevaluasi ketepatan prediksi positif [31]. *Recall* memperlihatkan kapasitas model untuk mengenali semua data yang termasuk ke dalam kelas positif, sedangkan *f1-score* menilai keseimbangan nilai *precision* dan *recall* [32]. Rumus evaluasi model ditunjukkan pada persamaan (1) hingga persamaan (4).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

$$F1-Score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall} \right) \times 100\% \quad (4)$$

Keterangan: Komponen pada *confusion matrix* berisi mengenai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). TP menggambarkan banyaknya data positif telah berhasil diklasifikasikan secara tepat, prediksi yang tepat terhadap data berkategori negatif dipresentasikan oleh nilai *True Negatif* (TN). FP menggambarkan data negatif salah dikenali menjadi positif, dan FN prediksi salah data positif menjadi negatif

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data diambil menggunakan metode *scraping* memakai bantuan visual studio code (VS Code). Data diambil seperti *username* pengguna, rating, tanggal ulasan, dan isi komentar pengguna aplikasi. Masing-masing aplikasi *mobile banking* dikumpulkan sebanyak 10.000 ulasan sehingga diperoleh total 40.000 data ulasan pengguna.

Tabel 1. Jumlah Data pada Setiap Tahap Pengolahan

Tahap Pengolahan Data	Jumlah Data
Dataset hasil scraping	40.000
Dataset setelah <i>preprocessing</i>	38.292

Dataset tersebut lalu diolah melewati rangkaian proses *preprocessing* berupa *case folding*, *cleaning*, normalisasi kata, *tokenizing*, *stopword removal*, serta *stemming*.

Tabel 2. Sampel Data Mentah Ulasan Pengguna

Score	Ulasan	Tanggal
5	puasss pake bangett	2026-03-30 14:24:36
5	cukup memuaskan dan mudah	2026-03-30 14:18:01
2	kenapa setiap login selalu di verifikasi terus bikin darah tinggi	2026-03-30 13:36:04
5	keren sekali	2025-10-27 17:47:57
5	sangat puas menggunakan seabank	2026-03-30 14:25:08

3.2. Pelabelan Data

Setelah *preprocessing* selesai, data ulasan memasuki tahap pelabelan sentimen menggunakan data penilaian pengguna di Play Store. Rating 4 serta 5 dikelompokkan sentimen positif karena mencerminkan kepuasan terhadap aplikasi *mobile banking*, sedangkan rating 1,2, dan 3 diklasifikasikan sebagai sentimen negatif karena menunjukkan adanya keluhan, kritik atau ketidakpuasan terhadap layanan yang diberikan. Pelabelan sentimen dilakukan untuk menghasilkan kelas target yang selanjutnya menjadi dasar klasifikasi memakai algoritma Naïve Bayes dan Support Vector Machine. Hasil pelabelan menunjukkan dataset terbagi menjadi dua kelas sentimen, yaitu positif dan negatif.

Tabel 3. Hasil Pelabelan Data

Score	Sentimen	Ulasan
5	Positif	puasss pake bangett
2	Negatif	tiap login selalu verifikasi terus bikin darah tinggi
5	Positif	keren sekali
5	Positif	baik
5	Positif	sangat puas guna seabank

Berdasarkan Tabel 4, terlihat bahwa jumlah sentimen positif dan negatif berbeda pada setiap aplikasi *mobile*. Perbedaan tersebut menunjukkan adanya variasi tingkat kepuasan pada masing-masing aplikasi dan selanjutnya dimanfaatkan sebagai acuan saat pelatihan diproses serta model klasifikasi diuji.

3.3. Preprocessing

Preprocessing dipakai demi memperoleh data ulasan hasil *scraping* yang baik sebelum proses klasifikasi sentimen. Data yang berhasil diambil melalui proses *scraping* masih mengandung berbagai elemen yang dapat mengganggu analisis, seperti huruf kapital, tanda baca, emoji, URL, dan kata tidak baku. Oleh karena itu, data diproses melalui *case folding*, *cleaning*, normalisasi kata, *tokenizing*, *stopword removal*, dan *stemming*. Tabel 4 memperlihatkan *preprocessing* menghasilkan ulasan pengguna aplikasi BRImo, Livin by Mandiri, BCA Mobile, dan SeaBank. Tahapan *preprocessing* diterapkan pada seluruh data ulasan yang dipakai pada penelitian ini agar memberikan dataset yang cenderung bersih, terstruktur, serta siap dipakai pada proses pembobotan TF-IDF serta klasifikasi sentimen.

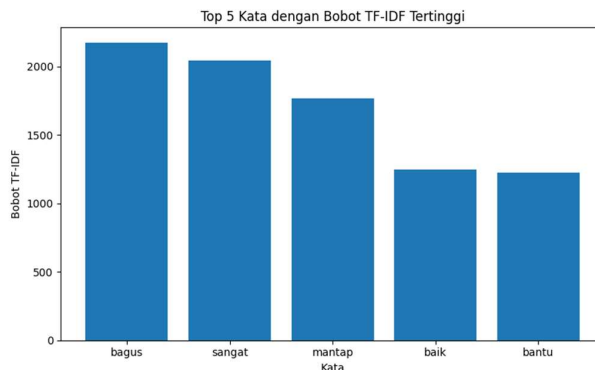
Tabel 4. Hasil *Preprocessing*

Teks Asli	Case Folding dan Cleaning	Normalisasi	Tokenizing	Stopword removal	Stemming	Hasil <i>Preprocessing</i>
Bank BRI Terbaik 🍌 100 paling top 🍌🍌🍌🍌	bank bri terbaik paling top	bank bri terbaik paling top	['bank', 'bri', 'terbaik'] ['paling', 'top']	['bank', 'bri', 'terbaik'] ['paling', 'top']	['bank', 'bri', 'baik'] ['paling', 'top']	bank bri baik paling top

Berdasarkan Tabel 4, proses *preprocessing* dapat mengeliminasi karakter yang tidak diperlukan, menyeragamkan penulisan kata, mengubah kata tidak baku menjadi bentuk dasar. Hasil *preprocessing* menghasilkan data ulasan yang lebih terstruktur sehingga lebih sesuai untuk proses ekstraksi fitur memakai TF-IDF. Pada tahap ini meminimalkan *noise*, peningkatan kinerja Naïve Bayes serta Support Vector Machine dalam mengklasifikasikan data berdasarkan kategori sentimen dipengaruhi oleh kondisi tersebut.

3.4. Pembobotan TF-IDF

Selanjutnya pada tahap ekstraksi fitur memakai metode TF-IDF (*Term Frequency-Inverse Document*) untuk memberikan hasil representasi teks data. Melalui metode tersebut, setiap istilah data dikonversi menjadi nilai numerik berdasarkan perhitungan frekuensi kemunculan dan tingkat kepentingan kata dalam kumpulan data yang digunakan. Proses pembobotan akan menghasilkan nilai bobot cenderung lebih tinggi terhadap kata yang mempunyai tingkat relevansi berskala besar dalam merepresentasikan dokumen karena lebih efektif dalam membedakan sentimen, sedangkan kata yang paling banyak muncul pada hampir seluruh ulasan akan diberi bobot lebih rendah. Pada penelitian ini, pembobotan dilakukan menggunakan *TfidfVectorizer* dengan jumlah fitur maksimum sebanyak 5.000 kata.



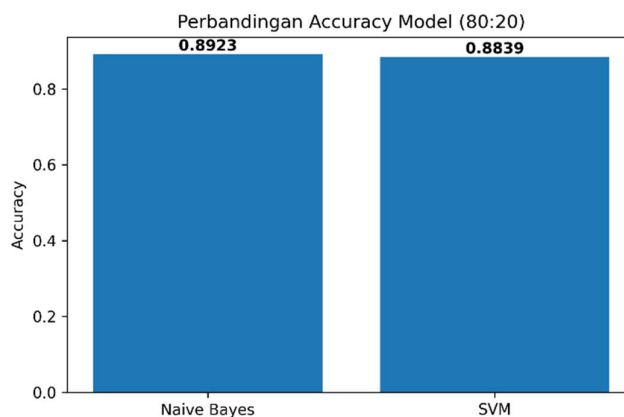
Gambar 2. Grafik 5 kata bobot TF-IDF tertinggi

3.5. Klasifikasi Sentimen

Naïve Bayes dan Support Vector Machine diterapkan sebagai klasifikasi teks, mengingat kedua algoritma tersebut mampu bekerja secara efektif terhadap karakteristik data teks serta banyak diterapkan pada penelitian analisis sentimen. Naïve Bayes melakukan klasifikasi mengadopsi pendekatan metode probabilistik serta asumsi independensi antar fitur, sedangkan Support Vector Machine digunakan sebagai pembandingan karena mampu menentukan *hyperplane* terbaik untuk memisahkan kelas sentimen.

3.6. Evaluasi Model

Performa kualitas hasil kerja Naïve Bayes serta Support Vector Machine dievaluasi memakai data uji didapatkan saat pembagian dataset dengan rasio data latih 80% dan data uji 20%. Selain itu, penerapan *5-fold cross validation* digunakan agar mendapatkan model evaluasi yang lebih konsisten serta objektif.

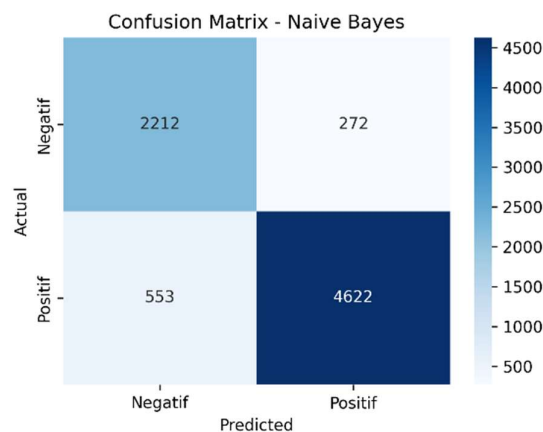


Gambar 3. Grafik Perbandingan Kinerja Model

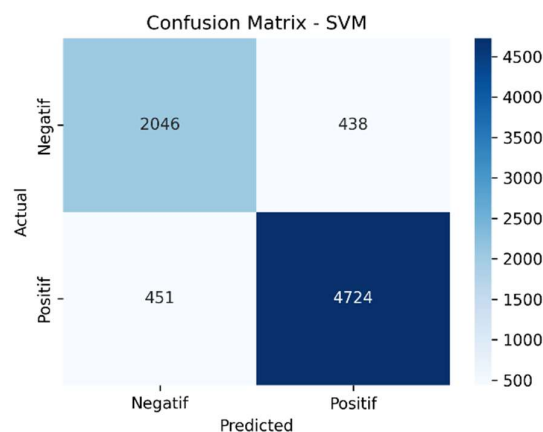
Tabel 5. Perbandingan Kinerja Model

Model	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	0.8923	0.8976	0.8923	0.8937
Support Vector Machine	0.8839	0.8841	0.8839	0.8840

Hasil pengujian pada Tabel 5 dengan penerapan metode *train-test* Split 80:20, memperlihatkan bahwa algoritma Naïve Bayes mendapatkan *accuracy* nilai 0.8923, *precision* 0.8976, *recall* 0.8923, dan *F1-score* 0.8937. Adapun algoritma Support Vector Machine menghasilkan *accuracy* 0.8839, *precision* 0.8841, *recall* 0.8839, dan *F1-score* 0.8840. Perbandingan tersebut memperlihatkan bahwa Naïve Bayes memiliki kemampuan performa cenderung lebih baik melebihi Support Vector Machine pada metrik evaluasi secara keseluruhan. Kondisi tersebut diduga dipengaruhi oleh kemampuan Naïve Bayes dalam memodelkan distribusi probabilitas kata pada representasi TF-IDF sehingga mampu mengenali pola sentimen secara lebih efektif. Temuan ini memperkuat hasil penelitian [13] yang melaporkan Naïve Bayes lebih baik dibandingkan Support Vector Machine yang membuat karakteristik dataset menjadi salah satu faktor yang mempengaruhi perbedaan performa kedua algoritma dalam analisis sentimen, proses *preprocessing*, dan juga representasi fitur yang dipakai dalam proses klasifikasi.



Gambar 4. Confusion matrix Algoritma Naïve Bayes



Gambar 5. Confusion matrix Support Vector Machine

Hasil pengujian memperlihatkan sebagian besar data telah berhasil diidentifikasi secara benar oleh kedua model klasifikasi. Pada model Naïve Bayes berhasil mengklasifikasikan 2.212 data negatif serta 4.622 data positif dengan error sebanyak 272 False Positive (FP) serta 553 False Negative (FN). Sementara itu, model Support Vector Machine menghasilkan 2.046 data negatif dan 4.724 data positif benar, namun menunjukkan angka kesalahan yang lebih tinggi pada kelas negatif sebesar 438 FP dan 451 FN pada kelas positif. Meskipun dataset memiliki ketidakseimbangan kelas (*class imbalance*), kedua model tetap menunjukkan performa klasifikasi yang baik. Hal ini mengindikasikan bahwa Naïve Bayes maupun Support Vector Machine masih mampu mempelajari pola pada kedua kelas, meskipun jumlah data positif dan negatif tidak seimbang.

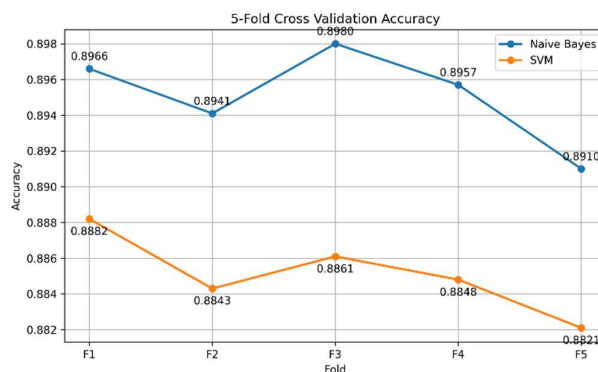
Performa Naïve Bayes yang lebih baik menunjukkan bahwa algoritma ini lebih efektif dalam memanfaatkan distribusi fitur TF-IDF pada data ulasan mobile banking. Karakteristik data yang berupa teks pendek dengan kata-kata yang memiliki kecenderungan kuat terhadap sentimen positif maupun negatif membuat pendekatan probabilistik Naïve Bayes lebih optimal, sedangkan LinearSVC menghasilkan akurasi yang sedikit lebih rendah pada dataset penelitian ini. Penemuan ini memperkuat hasil penelitian [33] yang menunjukkan bahwa performa

klasifikasi yang dihasilkan oleh Naïve Bayes lebih baik pada saat analisis sentimen. Perbedaan performa kedua algoritma juga dapat terpengaruh dari ciri khas dataset, tahapan *preprocessing*, serta representasi fitur yang dipakai selama proses klasifikasi

Confusion matrix menunjukkan bahwa masih terdapat sejumlah data yang mengalami salah klasifikasi, yang ditunjukkan oleh nilai *false positive* dan *false negative* pada kedua model. Hal ini mengindikasikan bahwa variasi karakteristik teks ulasan masih menjadi tantangan dalam proses klasifikasi sentimen. Meskipun demikian, Naïve Bayes menghasilkan jumlah kesalahan yang lebih rendah dibandingkan Support Vector Machine sehingga dapat menunjukkan kinerja lebih baik pada penelitian ini.

3.7. Cross Validation

Pengujian memakai metode *5-fold cross validation* dilakukan agar dapat menilai kestabilan dan kemampuan generalisasi model. Hasil validasi silang tersebut disajikan pada Gambar 6.



Gambar 6. Grafik *5-fold cross validation*

Berdasarkan hasil pengujian *5-fold cross validation*, algoritma Naïve Bayes secara konsisten mengungguli Support Vector Machine pada seluruh iterasi (fold). Akurasi tertinggi Naïve Bayes dicapai pada F3 sebesar 0,8980 dan terendah pada F5 sebesar 0,8910. Sementara itu, nilai F1 tertinggi pada Support Vector Machine mencapai 0,882 dan titik terendah pada F5 sebanyak 0,8821. Hasil akhir ini menegaskan algoritma Naïve Bayes menghasilkan stabilitas dan performa akurasi cenderung lebih optimal dibandingkan Support Vector Machine. Hasil pengujian mengindikasikan keunggulan bahwa Naïve Bayes lebih dan lebih stabil daripada Support Vector Machine pada dataset penelitian.

Berdasarkan pengujian *train-test split* tunggal, Naïve Bayes mendominasi seluruh metrik evaluasi dengan keunggulan utama pada kemampuan meminimalkan salah klasifikasi ulasan negatif (False Positive). Konsistensi keunggulan tersebut diperkuat oleh hasil pengujian *5-fold cross validation*, Naïve Bayes secara stabil mempertahankan nilai akurasi tertinggi di setiap iterasi dibandingkan model Support Vector Machine. Dengan demikian, dapat disimpulkan bahwa karakteristik algoritma Naïve Bayes lebih adaptif dan optimal dalam mengenali pola distribusi fitur teks bersentimen pada penelitian ini.

Meskipun demikian, penelitian ini menggunakan rating pengguna yang dijadikan acuan pelabelan sentimen, yaitu rating 1-3 kategori negatif serta rating 4-5 kategori positif. Pendekatan ini memungkinkan proses pelabelan dilakukan secara efisien pada dataset berskala besar, namun terdapat kemungkinan ketidaksesuaian antara rating dan isi ulasan yang dapat menimbulkan label noise. Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan anotasi manual atau validasi label untuk meningkatkan kualitas data pelatihan

4. KESIMPULAN

Mengacu pada hasil penelitian mengenai 40.000 ulasan pengguna aplikasi BRimo, Livin' by Mandiri, BCA Mobile, dan SeaBank di Play Store, dapat disimpulkan bahwa sentimen pengguna terhadap layanan *mobile banking* berhasil diklasifikasikan melalui penerapan ekstraksi fitur TF-IDF yang dipadukan dengan algoritma Naïve Bayes serta Support Vector Machine. Proses evaluasi menghasilkan temuan Naïve Bayes mendapatkan akurasi sebanyak 89,23%, lebih optimal daripada Support Vector Machine dan mendapatkan hasil akurasi sebanyak 88,39%, sehingga menunjukkan performa yang lebih baik dan lebih konsisten dalam proses klasifikasi sentimen. Selain itu, hasil analisis menunjukkan bahwa SeaBank memperoleh dominasi sentimen positif tertinggi, sedangkan BCA Mobile memiliki jumlah sentimen negatif terbanyak. Penelitian ini berkontribusi dengan menyajikan evaluasi komparatif kedua algoritma menggunakan dataset yang relatif besar dari empat aplikasi

mobile banking di Indonesia sehingga dapat menjadi referensi dalam pemilihan algoritma klasifikasi sentimen. Temuan ini juga memberikan implikasi praktis bagi penyedia layanan perbankan digital dalam memanfaatkan ulasan pengguna sebagai dasar evaluasi demi mendukung peningkatan layanan secara berkelanjutan.

Beberapa keterbatasan masih ada pada penelitian ini, yaitu pelabelan sentimen masih didasarkan pada rating pengguna sehingga berpotensi menimbulkan ketidaksesuaian antara rating dan isi ulasan, serta hanya membandingkan dua algoritma klasifikasi tanpa melakukan uji signifikansi statistik terhadap perbedaan performanya. Selain itu, sebaran data antar kelas pada dataset yang digunakan belum seimbang (*class imbalance*), sehingga kondisi tersebut berpotensi memengaruhi proses pembelajaran model dan hasil klasifikasi. Karakteristik dataset yang diperoleh dari ulasan pengguna di Play Store juga memungkinkan adanya variasi gaya bahasa, penggunaan kata tidak baku, serta perbedaan antara isi ulasan dan rating yang diberikan, sehingga masih menjadi tantangan dalam proses analisis sentimen. Dengan demikian, penelitian berikutnya dianjurkan agar menerapkan teknik pelabelan sentimen yang lebih akurat, seperti membandingkan pelabelan manual atau validasi oleh anotator, melakukan uji signifikansi statistik untuk memperkuat hasil perbandingan model, sekaligus membuat perbandingan performa algoritma mengadopsi model *deep learning* yang dibangun menggunakan arsitektur *transformer*, seperti IndoBERT guna memperoleh hasil pengklasifikasian yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] A. Huda, M. Arifin, and S. Adiyono, "Perancangan Antarmuka Sistem Informasi Employee Self Service Berbasis Mobile Menggunakan Pendekatan Design Thinking Pada Haen Teknologi Nusantara," *Jurnal SITECH : Jurnal Sistem Informasi dan Teknologi*, vol. 7, no. 2, pp. 51–60, 2024, doi: 10.24176/sitech.v7i2.13436.
- [2] S. P. Halim, M. N. Azhad, and I. Puspitadewi, "Pengaruh Kemudahan, Kemanfaatan, Dan Risiko Terhadap Minat Nasabah Dalam Menggunakan Mobile Banking (Studi Kasus PadaBank Syariah Indonesia Cabang Jember)," *Jurnal Mahasiswa Entrepreneur (JME)*, vol. 1, no. 10, pp. 2035–2042, 2022, doi: 10.36841/jme.v1i10.2252.
- [3] M. M. Suri, P. R. Leony, R. A. Keluanan, M. E. Im, M. Klau, and Y. P. Lian, "Peran Aplikasi Mobile Banking dalam Kemudahan Transaksi Harian," *Jurnal Pendidikan Tambusai*, vol. 8, no. 2, pp. 32127–32130, 2024.
- [4] A. D. Satya, Cupian, and Y. A. Faisal, "Digital Financing Business Models for MSMEs in Indonesia: Effect on Financial Access, Productivity, and Job Creation toward SDG 8," *Jurnal Ilmiah Akuntansi Kesatuan*, vol. 13, no. 6, pp. 1651–1662, 2025, doi: 10.37641/jiakes.v13i6.4336.
- [5] S. Adiyono, N. Latifah, and D. L. Fithri, "Analysis of Digital Readiness in the Social Assistance Distribution System with the Unified Theory of Acceptance and Use of Technology (UTAUT)," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 2, pp. 483–489, 2025, doi: 10.30871/jaic.v9i2.9070.
- [6] K. Y. Dewi, I. D. A. I. Saraswati, Y. M. Kiuk, G. R. Dantes, G. Indrawan, and I. M. A. O. Gunawan, "Benchmarking User Experience on Mobile Banking Applications in Indonesia: A Comparative Study Using the User Experience Questionnaire," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 19, no. 1, pp. 1336–1352, 2026, doi: 10.24036/jtip.v19i1.1087.
- [7] O. S. Jelni, M. L. Radhitya, G. W. Wardhana, N. W. J. K. Dewi, and N. M. M. Desmayani, "Sentiment Analysis of BRImo Reviews on Google Play Store Using SVM and KNN," *Indonesian Journal of Data and Science*, vol. 6, no. 3, pp. 548–562, 2025, doi: 10.56705/ijodas.v6i3.365.
- [8] J. Yi, J. Kim, and Y. K. Oh, "Uncovering the quality factors driving the success of mobile payment apps," *Journal of Retailing and Consumer Services*, vol. 77, p. 103641, 2024, doi: 10.1016/J.JRETCONSER.2023.103641.
- [9] V. Anggraini, I. Budi, A. B. Santoso, and P. K. Putra, "Measuring mobile banking service quality using Topic Modeling and Term Ranking: A case study of an Indonesian digital bank," *The Indonesian Journal of Computer Science*, vol. 13, no. 6, pp. 9136–9152, 2024, doi: 10.33022/ijcs.v13i6.4517.
- [10] Z. A. Zulkifly, N. Brasit, M. S. Alhaqqi, and S. Adelia, "Analisis Peningkatan Kualitas Layanan Mobile Banking dengan Pendekatan Metode E-Servqual," *JBMI (Jurnal Bisnis, Manajemen, dan Informatika)*, vol. 19, no. 1, pp. 61–79, 2022, doi: 10.26487/jbmi.v19i1.21337.
- [11] N. D. Septiyanti, M. I. Luthfi, and N. T. Romadloni, "Komparasi Metode Klasifikasi Dalam Analisis Sentimen Ulasan Pengguna Aplikasi KRL Access Di Google Play Store," *Journal Computer Science and Information Systems : J-Cosys*, vol. 4, no. 1, pp. 64–75, 2024, doi: 10.53514/jco.v4i1.495.
- [12] A. Aryanusa, R. M. Akbar, and Y. Nur, "Evaluasi Kinerja Logistic Regression Dan Multinomial Naïve Bayes Dalam Klasifikasi Sentimen Mobile Banking," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3S1, pp. 463–474, Oct. 2025, doi: 10.23960/jitet.v13i3S1.7687.
- [13] Md. S. Islam, M. N. Kabir, N. A. Ghani, and K. Z. Zamli, "Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach," *Artif. Intell. Rev.*, vol. 57, no. 3, pp. 62–79, 2024, doi: 10.1007/s10462-023-10651-9.
- [14] P. W. Harjita, F. M. Putra, and D. A. Tyas, "Sentimen Analisis Pengguna Media Sosial Berdasarkan Metode Ekstraksi Fitur dan Klasifikasi," *Jurnal Ilmu Komputer*, vol. 16, no. 2, pp. 88–96, 2023, doi: 10.24843/JIK.2023.v16.i02.p02.
- [15] A. U. Safitra, A. M. B. Aji, B. O. Lubis, and B. Santoso, "Evaluasi Mobile Banking Dengan Pendekatan Use Questionnaire Dan Importance Performance Analysis," *Jurnal Teknologi Informatika dan Komputer*, vol. 8, no. 2, pp. 148–163, 2022, doi: 10.37012/jtik.v8i2.833.

- [16] A. Parlindungan Zai and S. Saleh, "Perbandingan Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pengguna Jobstreet," *Idealis: Indonesia Journal Information System*, vol. 9, no. 1, pp. 30–38, 2026, doi: 10.36080/idealis.v9i1.3663.
- [17] A. Palanivinaiyagam, C. Z. El-Bayeh, and R. Damaševičius, "Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review," *Algorithms*, vol. 16, no. 5, p. 236, 2023, doi: 10.3390/a16050236.
- [18] K. Raoufuddin Ahmed, S. A. Zura Jalil, and S. Usman, "Improved Tuna Swarm-based U-EfficientNet: Skin Lesion Image Segmentation by Improved Tuna Swarm Optimization," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, pp. 903–913, 2023, doi: 10.14569/IJACSA.2023.0140595.
- [19] A. Lukito, F. Nugraha, and N. Latifah, "Analisis Sentimen Ulasan Di Playstore Terhadap Aplikasi Marketplace Menggunakan Naive Bayes dan Logistic Regression," *Jurnal Unitek*, vol. 18, no. 2, pp. 322–334, 2025, doi: 10.52072/unitek.v18i2.1625.
- [20] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 375–384, 2024, doi: 10.57152/malcom.v4i2.1206.
- [21] F. A. Irawan, A. R. Atmadja, and A. Wahana, "Analisis Sentimen Ulasan Aplikasi Bank Digital Menggunakan Algoritma Naive Bayes," *EXPLORER: Journal of Computer Science and Information Technology*, vol. 4, no. 2, pp. 60–68, 2024, doi: 10.47065/explorer.v4i2.1181.
- [22] S. Manda and V. Vardhan, "A Systematic Literature Review on Sentimental Analysis using Machine Learning for Preprocessing, Feature Extraction and Classification," *SSRN Electronic Journal*, 2024, doi: 10.2139/ssrn.5061012.
- [23] D. Rifaldi, A. Fadlil, and Herman, "Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet 'Mental Health,'" *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 161–171, 2023, doi: 10.51454/decode.v3i2.131.
- [24] A. Sasi Kirana, R. Roeswidiyah, and A. Pudoli, "Analisis Sentimen Pada Media Sosial Terhadap Layanan Samsat Digital Nasional Dengan Support Vector Machine," *Idealis: Indonesia Journal Information System*, vol. 8, no. 1, pp. 53–63, 2025, doi: 10.36080/idealis.v8i1.3276.
- [25] D. Septiani and I. Isabela, "Analisis Term Frequency Inverse Document Frequency (Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks," *SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia*, vol. 1, no. 2, pp. 81–88, 2022.
- [26] A. Arasy, S. Agustian, L. Handayani, and I. Iskandar, "Klasifikasi Sentimen Menggunakan Metode Multilayer Perceptron dengan Fitur TF-IDF," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, pp. 908–919, 2025, doi: 10.57152/malcom.v5i3.2052.
- [27] M. I. Burhan, A. N. Ali, A. I. Auliyah, and M. Hading, "Sentiment Analysis of Bapenda South Sulawesi Mobile Application on Google Play Store Using Support Vector Machine," *Journal of Mathematics and Applied Statistics*, vol. 2, no. 2, pp. 89–96, 2024, doi: 10.35914/mathstat.v2i2.244.
- [28] A. Mudya Yolanda and R. Tri Mulya, "Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store," *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, vol. 6, no. 2, pp. 76–83, 2024, doi: 10.35580/variensiumm258.
- [29] A. J. Tobing and A. Febriandirza, "Analisis Sentimen Aplikasi Mobile Banking Bca Pada Ulasan Pengguna di Google Play Store Menggunakan Metode Naive Bayes," *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 998–1005, 2024, doi: 10.47065/josh.v5i4.5485.
- [30] O. M. Wulandari, I. Maulana, F. Syamsudin, and R. Waluyo, "Perbandingan Algoritma Naive Bayes dan SVM dalam Analisis Sentimen Twitter terhadap Isu Ijazah Jokowi Palsu," *Jurnal Manajemen Informatika, Sistem Informasi dan Teknologi Komputer (JUMISTIK)*, vol. 4, no. 1, pp. 392–400, 2025, doi: 10.70247/jumistik.v4i1.145.
- [31] M. Aulia and A. Hermawan, "Analisis Perbandingan Algoritma SVM, Naïve Bayes, dan Perceptron untuk Analisis Sentimen Ulasan Produk Tokopedia," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, p. 1850, 2023, doi: 10.30865/mib.v7i4.6839.