

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI PERBANKAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES

Muhamad Aji Afani Maldini, Septi Andryana

Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional
Jl. Sawo Manila No. 61, Pasar Minggu, Jakarta Selatan 12520, Indonesia
aji.basbah@gmail.com

ABSTRAK

Perkembangan teknologi dalam sektor perbankan telah mendorong transformasi digital, termasuk layanan pinjaman berbasis aplikasi seperti Ceria by BRI. Meskipun aplikasi ini menawarkan kemudahan akses dan kecepatan transaksi, pengalaman pengguna terhadap layanan ini beragam, dengan adanya sentimen positif, netral, dan negatif. Beberapa pengguna memberikan ulasan positif terkait kemudahan penggunaan dan efisiensi transaksi, sementara ulasan negatif mencerminkan kendala teknis, keterbatasan fitur, serta kekhawatiran terhadap keamanan data. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi Ceria by BRI dengan menerapkan algoritma Support Vector Machine (SVM) dan Naïve Bayes. Metode penelitian ini melibatkan pengumpulan data ulasan pengguna dari Google Play Store menggunakan web scraping, kemudian dilakukan tahap preprocessing, termasuk normalisasi, tokenisasi, stopword removal, dan stemming. Data yang telah diproses selanjutnya diklasifikasikan menggunakan SVM dan Naïve Bayes untuk mengidentifikasi sentimen positif, netral, dan negatif, dengan evaluasi model berdasarkan akurasi, precision, recall, dan f1-score. Hasil penelitian menunjukkan bahwa SVM memiliki akurasi lebih tinggi (88%) dibandingkan Naïve Bayes (84%), dengan sentimen pengguna yang didominasi kategori netral dan positif. Ulasan negatif tetap ditemukan, terutama pada aspek kendala teknis dan keamanan data. Hasil penelitian ini memberikan wawasan bagi pengelola aplikasi dalam meningkatkan kualitas layanan, optimalisasi keamanan data, serta pengembangan antarmuka pengguna yang lebih intuitif. Selain itu, analisis sentimen berbasis machine learning dapat diterapkan secara berkelanjutan untuk mengevaluasi kepuasan pengguna dan meningkatkan pengalaman mereka terhadap layanan perbankan digital.

Kata kunci : Analisis Sentimen, Support Vector Machine, Naïve Bayes, Ulasan Pengguna, Perbankan Digital, Machine Learning.

1. PENDAHULUAN

Perbankan memiliki peran sentral dalam perekonomian suatu negara, terutama dalam menyediakan layanan keuangan yang aman dan efisien bagi masyarakat. Seiring dengan perkembangan teknologi, industri perbankan menghadapi tantangan untuk terus berinovasi dan beradaptasi dengan perubahan digital. Perbankan digital menjadi solusi strategis yang memungkinkan lembaga keuangan memperluas akses dan meningkatkan efisiensi layanan. Menurut laporan Bank Indonesia (2022), layanan perbankan digital mengalami pertumbuhan pesat, didorong oleh penetrasi internet dan adopsi teknologi yang semakin tinggi. Bank Rakyat Indonesia (BRI) sebagai salah satu bank terbesar di Indonesia turut berinovasi dengan menghadirkan platform pinjaman digital Ceria by BRI untuk memenuhi kebutuhan masyarakat akan layanan finansial yang cepat dan fleksibel [1].

Platform Ceria by BRI merupakan layanan pinjaman digital yang memungkinkan pengguna memperoleh kredit tanpa harus mengunjungi kantor bank. Inovasi ini berkontribusi dalam meningkatkan inklusi keuangan nasional dengan memperluas akses terhadap layanan keuangan berbasis teknologi. Studi oleh Widiyanto et al. (2021) menunjukkan bahwa layanan keuangan digital meningkatkan aksesibilitas

perbankan, terutama bagi masyarakat yang sulit menjangkau layanan konvensional [2].

Dengan semakin meningkatnya adopsi teknologi di sektor keuangan, penting untuk memahami bagaimana persepsi pengguna terhadap layanan seperti Ceria by BRI serta bagaimana faktor teknologi, keamanan, dan pengalaman pengguna memengaruhi kepuasan mereka.

Dalam era digitalisasi, analisis sentimen telah menjadi metode yang banyak digunakan untuk memahami opini pengguna terhadap suatu produk atau layanan. Analisis sentimen dalam konteks keuangan, dikenal sebagai Financial Sentiment Analysis (FSA), memainkan peran penting dalam mengidentifikasi sentimen pengguna terhadap layanan keuangan digital berbasis teknologi [3].

Dengan meningkatnya volume ulasan dan komentar pengguna di platform digital, metode ini dapat membantu bank dalam menyesuaikan strategi layanan mereka untuk meningkatkan kepuasan dan loyalitas pelanggan.

Dalam konteks Ceria by BRI, beberapa studi telah menunjukkan bahwa terdapat variasi dalam pengalaman pengguna yang mencakup aspek positif maupun negatif. Beberapa pengguna mengapresiasi kemudahan akses dan kecepatan layanan, namun terdapat pula yang mengeluhkan kendala teknis,

keterbatasan fitur, serta respons layanan pelanggan yang kurang memadai [4].

Faktor-faktor ini perlu dianalisis secara mendalam untuk memahami kesenjangan antara ekspektasi dan realitas layanan yang diterima oleh pengguna.

Selain itu, keamanan data dan privasi menjadi perhatian utama dalam layanan keuangan digital. Prasetya dan Lestari (2021) menyatakan bahwa persepsi terhadap keamanan data sangat berpengaruh terhadap keputusan pengguna dalam mengadopsi layanan keuangan berbasis teknologi [5].

Oleh karena itu, pemahaman mendalam mengenai sentimen pengguna terhadap aspek keamanan, keandalan, dan kemudahan penggunaan aplikasi Ceria by BRI menjadi hal yang krusial bagi pengembangan layanan keuangan digital yang lebih baik.

Penelitian ini berfokus pada analisis sentimen ulasan pengguna aplikasi Ceria by BRI dengan menggunakan algoritma Support Vector Machine (SVM) dan Naive Bayes. Pemilihan kedua algoritma ini didasarkan pada keunggulan masing-masing dalam klasifikasi teks, khususnya dalam analisis sentimen berbasis data besar (big data).

Algoritma Support Vector Machine (SVM) dipilih karena kemampuannya dalam mengklasifikasikan teks secara akurat, terutama dalam masalah dengan dimensi tinggi seperti data ulasan pengguna. SVM bekerja dengan menemukan hyperplane optimal yang memisahkan kelas sentimen (positif, negatif, netral) dengan margin maksimal. Keunggulan utama dari SVM adalah kemampuannya dalam menangani data non-linear, sehingga dapat memberikan performa yang baik meskipun terdapat variasi dalam teks ulasan yang digunakan oleh pengguna aplikasi Ceria by BRI [6].

Di sisi lain, algoritma Naive Bayes dipilih karena sifatnya yang efisien dalam menangani teks berjumlah besar dengan komputasi yang cepat. Naive Bayes bekerja berdasarkan teorema Bayes, yang menghitung probabilitas suatu kategori berdasarkan distribusi kata dalam teks. Algoritma ini memiliki keunggulan dalam menangani klasifikasi berbasis probabilitas, sehingga sangat efektif dalam menentukan kecenderungan sentimen pengguna [7].

Pemilihan kombinasi SVM dan Naive Bayes didasarkan pada keunggulan komplementer yang dimiliki keduanya. SVM memberikan akurasi tinggi dalam klasifikasi teks berbasis fitur yang kompleks, sementara Naive Bayes unggul dalam kecepatan dan efisiensi pemrosesan data yang besar. Kombinasi kedua algoritma ini memungkinkan analisis sentimen yang lebih akurat dan efisien, sehingga hasil klasifikasi sentimen pengguna terhadap aplikasi Ceria by BRI dapat lebih komprehensif dan berdasarkan analisis statistik yang kuat.

Ruang lingkup penelitian ini mencakup analisis ulasan pengguna yang tersedia di platform digital, dengan batasan utama pada aspek kualitas layanan,

keamanan data, serta efektivitas metode klasifikasi sentimen yang digunakan. Berdasarkan latar belakang tersebut, penelitian ini memiliki beberapa pertanyaan utama yang ingin dijawab:

- bagaimana persepsi keseluruhan pengguna terhadap layanan aplikasi Ceria by BRI, baik dalam aspek teknis maupun operasional?
- sejauh mana efektivitas algoritma Support Vector Machine (SVM) dan Naive Bayes dalam mengklasifikasikan sentimen ulasan pengguna? Dan
- bagaimana faktor keamanan dan privasi data memengaruhi sentimen pengguna terhadap aplikasi Ceria by BRI?

Penelitian ini bertujuan untuk menganalisis persepsi pengguna terhadap layanan yang diberikan oleh aplikasi Ceria by BRI guna menilai tingkat kepuasan pengguna. Selain itu, penelitian ini mengevaluasi efektivitas algoritma SVM dan Naive Bayes dalam mengklasifikasikan sentimen pengguna serta meneliti pengaruh faktor keamanan dan privasi data terhadap sentimen dan kepercayaan pengguna terhadap aplikasi ini.

Penelitian ini memiliki kontribusi dalam tiga aspek utama. Dari sisi teoretis, penelitian ini memberikan kontribusi terhadap literatur terkait analisis sentimen dalam sektor keuangan digital dan mengeksplorasi efektivitas algoritma SVM dan Naive Bayes dalam klasifikasi sentimen ulasan pengguna aplikasi keuangan digital. Dari sisi praktis, penelitian ini memberikan wawasan bagi BRI dalam meningkatkan kualitas layanan aplikasi Ceria by BRI berdasarkan analisis sentimen pengguna serta membantu pengembang dalam meningkatkan pengalaman pengguna dengan memperbaiki aspek teknis dan fitur aplikasi berdasarkan umpan balik dari ulasan pengguna. Dari sisi kebijakan, penelitian ini memberikan rekomendasi strategis bagi perbankan dalam meningkatkan keamanan dan kepercayaan pengguna terhadap layanan keuangan digital serta menyediakan data empiris yang dapat digunakan sebagai dasar dalam perancangan kebijakan peningkatan literasi keuangan digital di Indonesia.

Dengan demikian, penelitian ini berupaya memberikan perspektif komprehensif dalam memahami pengalaman pengguna aplikasi Ceria by BRI dengan pendekatan berbasis data dan teknologi, sehingga hasilnya dapat memberikan dampak signifikan bagi peningkatan kualitas layanan serta pengembangan strategi bisnis berbasis digital di sektor perbankan.

2. TINJAUAN PUSTAKA

2.1. Studi Literatur

Studi literatur mengenai analisis sentimen ulasan pengguna aplikasi perbankan dengan menggunakan algoritma Support Vector Machine (SVM) dan Naive Bayes menunjukkan efektivitas kedua metode dalam klasifikasi sentimen teks. Penelitian oleh Zhang et al. (2021) mengindikasikan bahwa SVM memberikan

akurasi tinggi dalam menangani ulasan kompleks[8], sementara Gupta & Kaur (2021) membandingkan kedua algoritma dan menemukan bahwa Naive Bayes unggul dalam efisiensi komputasi, meskipun SVM lebih akurat [9].

Pendekatan Aspect-Based Sentiment Analysis (ABSA) dengan SVM telah digunakan oleh Wang & Li (2022) untuk mengidentifikasi sentimen terhadap fitur aplikasi secara lebih rinci [10].

Selain itu, Manning et al. (2020) menyoroti kecepatan Naive Bayes dalam klasifikasi teks besar [11], sedangkan Li et al. (2022) menemukan keunggulan SVM dalam mengklasifikasikan ulasan yang menggunakan bahasa informal [12].

Penelitian oleh Zhao et al. (2022) menekankan bahwa penggunaan fitur n-gram meningkatkan akurasi SVM [13], sedangkan Subramani et al. (2023) menunjukkan bahwa kombinasi SVM dan Naive Bayes dapat meningkatkan akurasi analisis sentiment [14].

Studi oleh Pontiki et al. (2021) mengaplikasikan ABSA untuk mengidentifikasi aspek spesifik dalam ulasan pengguna [15], sementara Johnson et al. (2023) menyoroti kinerja SVM pada data dengan tingkat kompleksitas tinggi [16].

Selain itu, penelitian Ren et al. (2022) menunjukkan bahwa pendekatan pembelajaran mendalam dengan SVM meningkatkan akurasi ketika digabungkan dengan metode ekstraksi fitur seperti Word2Vec [17].

Hasil studi ini menunjukkan bahwa SVM dan Naive Bayes merupakan algoritma yang kuat dalam analisis sentimen aplikasi perbankan, memberikan wawasan berharga bagi pengembangan layanan digital.

2.2. Landasan Teori

2.1.1. Analisis Sentimen

Analisis sentimen adalah proses mengekstraksi opini dari teks dan mengkategorikannya menjadi positif, negatif, atau netral. Berdasarkan jurnal oleh Liu (2020) dan Cambria (2022), analisis sentimen berperan dalam memahami opini pengguna terhadap layanan keuangan digital, termasuk aplikasi perbankan [18] [19].

2.1.2. Support Vector Machine (SVM)

SVM adalah algoritma klasifikasi berbasis vektor pendukung yang mencari hyperplane optimal untuk memisahkan data dalam ruang multidimensi. Menurut Zhang et al. (2021), SVM sangat efektif dalam menangani data dengan dimensi tinggi [20].

Berdasarkan studi terbaru, SVM memiliki performa tinggi dalam analisis sentimen karena prinsip margin maksimum yang diterapkan [5].

2.1.3. Naive Bayes

Naive Bayes adalah algoritma berbasis teorema Bayes, yang digunakan dalam klasifikasi teks berdasarkan probabilitas kemunculan kata. Manning et

al. (2020) menjelaskan bahwa Naive Bayes efisien dalam menangani data besar [3], sementara Gupta et al. (2023) menunjukkan bahwa algoritma ini mampu melakukan analisis sentimen dengan cepat meskipun akurasinya lebih rendah dibandingkan SVM [7].

2.1.4. Aspect-Based Sentiment Analysis (ABSA)

Pendekatan ABSA memungkinkan analisis sentimen berdasarkan aspek tertentu dari ulasan pengguna. Ren et al. (2022) menyebutkan bahwa ABSA mampu mengidentifikasi opini pengguna terhadap fitur spesifik dalam aplikasi perbankan [11].

Pontiki et al. (2021) menunjukkan bahwa ABSA memberikan pemahaman lebih rinci dibandingkan dengan analisis sentimen tradisional, karena dapat mengelompokkan sentimen terhadap aspek tertentu, seperti keamanan, layanan pelanggan, dan kemudahan penggunaan aplikasi [16].

2.3. Kerangka Teoritis

Berdasarkan studi literatur, kerangka teoritis penelitian ini mencakup algoritma SVM dan Naive Bayes dalam analisis sentimen, serta penerapan ABSA untuk memahami aspek spesifik dari ulasan pengguna aplikasi perbankan.

2.4. Penelitian Terdahulu

Beberapa penelitian terdahulu telah membahas efektivitas algoritma Support Vector Machine (SVM) dan Naive Bayes dalam analisis sentimen ulasan pengguna aplikasi perbankan digital. Penelitian oleh Zhang et al. (2021) menunjukkan bahwa SVM memiliki akurasi lebih tinggi dibandingkan Naive Bayes dalam mengklasifikasikan sentimen pengguna aplikasi perbankan [2].

Studi oleh Gupta & Kaur (2021) menemukan bahwa meskipun Naive Bayes lebih efisien dalam komputasi, SVM tetap memberikan hasil klasifikasi yang lebih akurat [13].

Pendekatan Aspect-Based Sentiment Analysis (ABSA) dalam analisis ulasan pengguna aplikasi perbankan telah digunakan oleh Wang & Li (2022), yang menemukan bahwa metode ini dapat mengidentifikasi sentimen berdasarkan fitur spesifik dari aplikasi perbankan [17].

Selain itu, Manning et al. (2020) mengungkapkan bahwa Naive Bayes mampu mengklasifikasikan ulasan dalam jumlah besar dengan cepat, namun memiliki tingkat akurasi yang lebih rendah dibandingkan SVM [9].

Studi oleh Li et al. (2022) menunjukkan bahwa SVM lebih efektif dalam menangani ulasan yang menggunakan bahasa informal, yang sering ditemukan dalam ulasan pengguna aplikasi digital [1].

Lebih lanjut, Zhao et al. (2022) menyoroti bahwa penggunaan fitur n-gram dapat meningkatkan akurasi SVM [2], sementara Subramani et al. (2023) menemukan bahwa kombinasi SVM dan Naive Bayes dapat meningkatkan keandalan hasil analisis sentiment [4].

Beberapa penelitian lainnya meneliti penerapan teknik pembelajaran mendalam dalam analisis sentimen perbankan digital. Ren et al. (2022) menunjukkan bahwa penggabungan SVM dengan metode ekstraksi fitur seperti Word2Vec meningkatkan akurasi dalam analisis sentiment [20]. Studi oleh Pontiki et al. (2021) dan Johnson et al. (2023) menyoroti bahwa SVM memiliki performa lebih baik dibandingkan Naive Bayes dalam menangani ulasan dengan tingkat kompleksitas tinggi [13] [18].

Hasil dari penelitian-penelitian terdahulu ini menjadi landasan dalam pemilihan metode analisis sentimen dalam penelitian ini, yang menggunakan kombinasi SVM dan Naive Bayes untuk memperoleh klasifikasi sentimen yang lebih akurat dan efisien.

2.5. Kerangka Penelitian

Kerangka penelitian ini meliputi beberapa tahapan utama yang dilakukan untuk menganalisis sentimen ulasan pengguna aplikasi Ceria by BRI. Tahap pertama adalah pengumpulan data, yang dilakukan dengan mengumpulkan ulasan pengguna dari platform digital seperti Google Play Store. Data ini kemudian melalui tahap pra-pemrosesan teks, yang mencakup tokenisasi, penghapusan stop words, dan stemming, sehingga data yang digunakan menjadi lebih bersih dan siap untuk diproses lebih lanjut.

Selanjutnya, data yang telah diproses akan diklasifikasikan menggunakan algoritma Support Vector Machine (SVM) dan Naive Bayes. SVM dipilih karena memiliki kemampuan yang baik dalam menangani data dengan dimensi tinggi dan kompleks, sementara Naive Bayes digunakan karena efisiensinya dalam pemrosesan teks dalam jumlah besar. Analisis sentimen ini juga akan menggunakan pendekatan Aspect-Based Sentiment Analysis (ABSA) untuk mengidentifikasi sentimen berdasarkan aspek-aspek spesifik dari aplikasi Ceria by BRI, seperti keamanan, kemudahan penggunaan, dan layanan pelanggan.

Tahap akhir dari penelitian ini adalah evaluasi performa model, yang dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score sebagaimana dijelaskan dalam penelitian Li et al. (2022). Evaluasi ini bertujuan untuk membandingkan efektivitas SVM dan Naive Bayes dalam analisis sentimen, serta menentukan metode yang paling optimal dalam mengidentifikasi persepsi pengguna terhadap layanan aplikasi perbankan digital. Dengan pendekatan ini, penelitian ini diharapkan dapat memberikan pemahaman yang lebih komprehensif mengenai pengalaman pengguna aplikasi Ceria by BRI serta memberikan rekomendasi bagi pengembang untuk meningkatkan kualitas layanan mereka.

2.6. Hipotesis Penelitian

H1: Algoritma Support Vector Machine (SVM) memiliki tingkat akurasi lebih tinggi dibandingkan dengan Naive Bayes dalam klasifikasi sentimen ulasan pengguna aplikasi perbankan digital.

H2: Pendekatan Aspect-Based Sentiment Analysis (ABSA) meningkatkan akurasi analisis sentimen, memungkinkan klasifikasi berdasarkan aspek spesifik aplikasi perbankan.

H3: Terdapat hubungan positif antara sentimen positif pengguna dan tingkat kepuasan terhadap aplikasi perbankan digital, yang berkontribusi terhadap loyalitas pengguna [10].

3. METODE PENELITIAN

3.1. Lokasi dan Waktu Penelitian

Penelitian ini dilakukan secara daring menggunakan Google Colab, sebuah platform berbasis cloud yang mendukung pemrograman Python untuk analisis data. Google Colab dipilih karena fleksibilitasnya dalam mengelola data besar dan efisiensinya dalam menjalankan teknik web scraping dengan library BeautifulSoup dan Selenium. Teknik ini digunakan untuk mengekstrak ulasan pengguna aplikasi Ceria by BRI dari Google Play Store, sehingga memungkinkan pengumpulan data secara langsung dan mendukung kebutuhan analisis sentimen dalam penelitian ini.

Penelitian ini dilaksanakan selama periode 5 bulan, dari Oktober 2024 hingga Februari 2025. Tahapan penelitian dimulai dengan pengumpulan data ulasan pengguna aplikasi Ceria by BRI melalui teknik web scraping yang dilakukan selama tiga bulan pertama. Selanjutnya, dilakukan pra-pemrosesan data untuk memastikan kualitas data yang akan dianalisis. Pada bulan ketiga, algoritma Naive Bayes dan Support Vector Machine (SVM) diterapkan untuk analisis sentimen, yang kemudian diikuti dengan evaluasi performa model pada bulan keempat. Penelitian diakhiri dengan analisis hasil dan penyusunan laporan pada bulan kelima, sehingga seluruh tahapan dapat diselesaikan secara sistematis dan terstruktur.

3.2. Penentuan Subjek Penelitian

Penelitian ini berfokus pada ulasan pengguna aplikasi Ceria by BRI, yang dapat memberikan informasi terkait persepsi dan pengalaman mereka terhadap layanan yang disediakan oleh aplikasi tersebut. Subjek penelitian ini dipilih berdasarkan kriteria sebagai berikut:

3.2.1. Aplikasi Ceria by BRI

Aplikasi ini dipilih sebagai objek penelitian karena merupakan salah satu inovasi di sektor perbankan digital yang banyak digunakan oleh masyarakat Indonesia. Aplikasi ini menawarkan layanan pinjaman digital dengan berbagai fitur yang memungkinkan pengguna untuk mengakses layanan keuangan dengan cepat dan mudah. Analisis sentimen terhadap ulasan pengguna aplikasi ini dapat memberikan gambaran yang jelas mengenai kepuasan pengguna terhadap layanan yang diberikan.

3.2.2. Pengguna Aplikasi Ceria by BRI

Subjek penelitian adalah pengguna aktif aplikasi Ceria by BRI yang telah memberikan ulasan atau review di Google Play Store. Pengguna dipilih berdasarkan kriteria berikut:

- Pengguna terdaftar: Pengguna yang telah melakukan transaksi atau menggunakan layanan aplikasi.
- Ketersediaan ulasan: Pengguna yang telah memberikan feedback melalui platform Google Play Store.
- Ulasan beragam: Ulasan yang mencakup berbagai kategori sentimen (positif, negatif, dan netral) agar variasi data yang dianalisis lebih representatif.

Kriteria Ulasan Pengguna

- Ulasan yang dipilih untuk dianalisis mencakup beberapa aspek utama, yaitu:
- Kualitas layanan: Kemudahan penggunaan, kecepatan transaksi, dan respons layanan pelanggan.
- Pengalaman pengguna: Interaksi dengan aplikasi, kepuasan terhadap antarmuka, serta kendala teknis yang dialami.
- Keamanan dan privasi: Persepsi pengguna terhadap keamanan data pribadi serta kebijakan privasi aplikasi.

3.3. Sumber Data

Penelitian ini menggunakan data sekunder berupa ulasan pengguna aplikasi Ceria by BRI yang diperoleh dari Google Play Store. Data yang dikumpulkan mencakup teks ulasan, rating (1-5 bintang), dan tanggal ulasan. Pengumpulan data dilakukan dengan web scraping menggunakan Python dan library BeautifulSoup serta Selenium untuk mengambil data secara otomatis dan sistematis.

Data yang digunakan mencakup ulasan dari periode Oktober hingga Desember 2024 untuk memastikan relevansi dan keterkinian analisis. Ulasan yang dipilih mencerminkan pengalaman nyata pengguna dan menjadi dasar evaluasi efektivitas algoritma Support Vector Machine (SVM) dan Naive Bayes dalam analisis sentimen.

3.4. Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan dengan metode web scraping, yang memungkinkan pengambilan data secara sistematis dari sumber digital. Web scraping dipilih karena mampu mengumpulkan ulasan pengguna aplikasi Ceria by BRI secara otomatis dari Google Play Store, sehingga memberikan data yang lebih akurat dan relevan untuk analisis sentimen.

Proses pengumpulan data diawali dengan persiapan alat dan library yang digunakan, yaitu Python dengan bantuan BeautifulSoup dan Selenium. BeautifulSoup digunakan untuk parsing halaman web, sementara Selenium berfungsi untuk menangani

elemen dinamis pada halaman web yang memerlukan interaksi, seperti memuat lebih banyak ulasan. Setelah alat dan library disiapkan, tahap selanjutnya adalah pengambilan data ulasan, yang dilakukan dengan mengakses halaman ulasan aplikasi Ceria by BRI di Google Play Store. Data yang dikumpulkan mencakup teks ulasan, rating bintang, dan tanggal ulasan.

Setelah data berhasil dikumpulkan, tahap selanjutnya adalah penyimpanan data dalam format CSV atau database, sehingga mempermudah proses analisis lebih lanjut. Penyimpanan ini mencakup elemen-elemen penting seperti teks ulasan, rating, serta metadata lainnya yang dapat membantu dalam klasifikasi sentimen. Selain itu, untuk menjaga kualitas data, dilakukan validasi dan pembersihan data yang mencakup penghapusan duplikasi, elemen tidak relevan, dan simbol atau karakter khusus yang tidak diperlukan dalam analisis.

Dalam pemilihan data yang akan dianalisis, penelitian ini menerapkan beberapa kriteria seleksi data. Hanya ulasan dengan teks deskriptif yang disertakan untuk memastikan relevansi dan kejelasan dalam analisis sentimen. Data yang digunakan mencakup ulasan dari periode Oktober hingga Desember 2024, guna memastikan bahwa hasil analisis tetap terkini dan relevan dengan kondisi terbaru aplikasi Ceria by BRI. Selain itu, pemilihan ulasan dilakukan berdasarkan keberagaman sentimen (positif, negatif, dan netral) agar penelitian ini dapat mencerminkan berbagai perspektif pengguna terhadap aplikasi perbankan digital tersebut.

Dengan teknik pengumpulan data yang sistematis dan berbasis web scraping, penelitian ini dapat mengumpulkan dataset yang lebih luas dan representatif, sehingga hasil analisis sentimen yang dihasilkan lebih akurat dan dapat menjadi dasar dalam mengevaluasi serta meningkatkan kualitas layanan aplikasi Ceria by BRI.

3.5. Metode Analisis Data

Metode analisis yang digunakan dalam penelitian ini mencakup penerapan algoritma Naive Bayes dan Support Vector Machine (SVM) untuk klasifikasi sentimen ulasan pengguna.

3.5.1. Naive Bayes

Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi independensi antar fitur. Rumus dasar algoritma ini adalah:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Dimana:

$P(C|X)$: Probabilitas suatu kelas C (misalnya positif, negatif, atau netral) diberikan data fitur X .

$P(X|C)$: Probabilitas data fitur X muncul dalam kelas C .

$P(C)$: Probabilitas awal (prior) suatu kelas C .

$P(X)$: Probabilitas data fitur X muncul di semua kelas.

Untuk kasus analisis teks, probabilitas dihitung sebagai:

$$P(X|C) = \prod_{i=1}^n P(x_i|C)$$

dengan Laplace Smoothing:

$$P(x_i|C) = \frac{\text{count}(x_i, C) + 1}{\text{count}(C) + |V|}$$

Dengan:

x_i : Kata ke- i dalam teks ulasan.

$P(x_i|C)$: Probabilitas kata x_i muncul dalam kelas C

$\text{count}(x_i, C)$: Jumlah kemunculan kata x_i dalam kelas C .

$\text{count}(C)$: Total kata dalam kelas C .

$|V|$: Jumlah total kata unik dalam semua kelas (vocabulary).

3.5.2. Support Vector Machine (SVM)

SVM adalah algoritma machine learning yang mencari hyperplane terbaik untuk memisahkan data dalam kelas berbeda. Rumus dasar hyperplane adalah:

$$f(x) = w \cdot x + b$$

Dimana :

$f(x)$: Fungsi keputusan yang menentukan kelas suatu data.

w : Vektor bobot yang mempengaruhi arah hyperplane.

x : Vektor fitur input.

b : Bias, yang menentukan posisi hyperplane.

Untuk mendapatkan margin maksimal, SVM menggunakan rumus:

$$M = \frac{2}{||w||}$$

SVM juga menggunakan fungsi kernel untuk menangani data non-linear, dengan beberapa jenis kernel seperti:

1. Linear Kernel: $K(x_i, x_j) = x_i \cdot x_j$
2. Polynomial Kernel: $K(x_i, x_j) = (x_i \cdot x_j + C)^d$
3. Radial Basis Function (RBF) Kernel: $K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2)$

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data dalam penelitian ini berasal dari ulasan pengguna aplikasi Ceria by BRI yang diambil dari Google Play Store. Platform ini dipilih karena menyediakan ulasan langsung dari pengguna aktif, yang mencerminkan pengalaman nyata mereka dalam menggunakan aplikasi. Pengumpulan data dilakukan menggunakan teknik web scraping dengan bantuan library Python seperti BeautifulSoup dan Selenium. Proses ini mencakup pengambilan teks ulasan, rating (bintang 1-5), dan tanggal ulasan. Untuk memastikan bahwa semua ulasan dapat diakses, Selenium digunakan untuk menangani elemen dinamis, seperti tombol "Load More", sehingga memungkinkan pengambilan ulasan dalam jumlah besar. Selanjutnya,

data yang dikumpulkan disimpan dalam format CSV, sehingga lebih mudah dianalisis.

Setelah data terkumpul, dilakukan validasi dan pembersihan untuk memastikan kualitas analisis. Tahapan ini meliputi penghapusan data duplikat, normalisasi teks, serta penyaringan ulasan agar hanya mencakup topik yang relevan, seperti kemudahan penggunaan, keamanan, layanan pelanggan, dan stabilitas aplikasi. Data dikumpulkan dalam periode Oktober hingga Desember 2024, dengan total 1.500 ulasan yang diklasifikasikan ke dalam tiga kategori sentimen utama, yaitu positif, negatif, dan netral. Dengan pendekatan ini, data yang dihasilkan diharapkan memberikan gambaran yang akurat mengenai persepsi pengguna terhadap aplikasi Ceria by BRI.

4.2. Tahapan Pengolahan Data

Proses pengolahan data dilakukan melalui beberapa tahap, mulai dari pembersihan data (cleansing), normalisasi, tokenisasi, filtering, stemming, hingga representasi data dalam bentuk numerik menggunakan metode TF-IDF.

- a. Cleansing: Menghilangkan karakter khusus, angka, dan tanda baca dari teks ulasan agar lebih bersih.
- b. Case Folding: Mengubah semua huruf dalam ulasan menjadi huruf kecil untuk menghindari perbedaan kapitalisasi.
- c. Normalization: Menstandarkan kata-kata dalam ulasan agar sesuai dengan bahasa formal yang dapat dipahami oleh model.
- d. Tokenizing: Memecah teks ulasan menjadi kata-kata individu (token) untuk dianalisis lebih lanjut.
- e. Filtering (Stopword Removal): Menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis sentimen, seperti "dan", "atau", "yang".
- f. Stemming: Mengubah kata-kata menjadi bentuk dasarnya, misalnya "menggunakan" menjadi "guna".
- g. TF-IDF (Term Frequency-Inverse Document Frequency): Mengubah teks ulasan ke dalam bentuk matriks numerik berdasarkan bobot kata yang muncul dalam dokumen.

4.3. Hasil Pengujian Algoritma Naïve Bayes

Naïve Bayes adalah algoritma klasifikasi probabilistik yang digunakan untuk menentukan probabilitas suatu ulasan termasuk dalam kategori positif, negatif, atau netral. Hasil pengujian menggunakan Naïve Bayes menunjukkan bahwa model mencapai akurasi sebesar 84%.

4.4. Evaluasi Performa Naïve Bayes

Evaluasi performa model Naïve Bayes dilakukan dengan mengukur precision, recall, dan F1-score untuk setiap kategori sentimen (positif, negatif, dan netral). Model ini mengklasifikasikan 1.500 ulasan pengguna

berdasarkan distribusi data yang telah diolah. Berikut adalah hasil evaluasi performa model Naïve Bayes:

Tabel 1 Hasil Evaluasi Performa Naïve Bayes

Kelas Sentimen	Precision	Recall	F1-score	Jumlah Data
Negatif	0	0	0	4
Netral	0.83	1	0.91	80
Positif	1	0.25	0.4	16
Macro Average	0.61	0.42	0.44	-
Weighted Average	0.83	0.84	0.79	-

Berdasarkan hasil klasifikasi, kelas netral memiliki performa terbaik dengan precision 0.83, recall 1.00, dan F1-score 0.91, karena dominasi jumlah data sebanyak 80 ulasan. Kelas positif menunjukkan precision yang tinggi sebesar 1.00, namun recall yang rendah (0.25), yang menandakan bahwa model hanya mampu mengenali sebagian kecil data positif (16 data). Sementara itu, kelas negatif tidak dapat diklasifikasikan dengan baik, karena jumlah data yang sangat sedikit (4 data) menyebabkan precision, recall, dan F1-score bernilai 0.00.

Untuk meningkatkan performa Naïve Bayes, diperlukan penyeimbangan jumlah data pada setiap kelas atau pengoptimalan parameter model agar dapat lebih baik dalam mengenali ulasan dari kelas minoritas.

4.5. Hasil Pengujian Algoritma Support Vector Machine (SVM)

SVM adalah algoritma klasifikasi berbasis margin maksimum, yang bertujuan untuk menemukan hyperplane terbaik yang dapat memisahkan kelas sentimen secara optimal. Hasil pengujian dengan SVM menunjukkan akurasi sebesar 88%, lebih tinggi dibandingkan Naïve Bayes.

4.6. Evaluasi Performa SVM

Evaluasi performa Support Vector Machine (SVM) dilakukan dengan metrik yang sama, yaitu precision, recall, dan F1-score. SVM bekerja dengan mencari hyperplane optimal yang dapat memisahkan kategori sentimen secara lebih akurat dibandingkan Naïve Bayes. Model ini diuji dengan dataset yang sama dan hasilnya dibandingkan dengan performa Naïve Bayes. Berikut adalah hasil evaluasi performa model SVM:

Tabel 2 Hasil Evaluasi Performa SVM

Kelas Sentimen	Precision	Recall	F1-score	Jumlah Data
Negatif	0	0	0	4
Netral	0.87	1	0.93	80
Positif	1	0.5	0.67	16
Macro Average	0.62	0.5	0.53	-
Weighted Average	0.86	0.88	0.85	-

Berdasarkan hasil klasifikasi, SVM memiliki kinerja lebih baik dibandingkan Naïve Bayes, terutama pada kelas netral dan positif. Precision kelas netral mencapai 0.87, recall 1.00, dan F1-score 0.93, sedangkan precision untuk kelas positif mencapai 1.00 dengan recall 0.50.

Namun, seperti Naïve Bayes, SVM juga mengalami kesulitan dalam mengenali kelas negatif, karena jumlah data yang sangat sedikit (4 data) menyebabkan precision, recall, dan F1-score bernilai 0.00. Dengan demikian, meskipun SVM memiliki akurasi lebih tinggi (88%) dibandingkan Naïve Bayes (84%), model ini masih perlu perbaikan dalam menangani ketidakseimbangan data antar kelas.

4.7. Perbandingan Performa Naïve Bayes dan SVM

Dari hasil pengujian, dapat disimpulkan bahwa SVM memiliki performa lebih baik dibandingkan Naïve Bayes dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Ceria by BRI. Akurasi SVM mencapai 88%, lebih tinggi dibandingkan Naïve Bayes yang hanya mencapai 84%.

Namun, kedua model mengalami keterbatasan dalam menangani kelas negatif, yang dapat diperbaiki dengan peningkatan jumlah data negatif atau menggunakan teknik oversampling agar model dapat mengenali pola dari kelas ini dengan lebih baik.

5. KESIMPULAN DAN SARAN

Penelitian ini menganalisis sentimen ulasan pengguna terhadap aplikasi Ceria by BRI menggunakan algoritma Support Vector Machine (SVM) dan Naïve Bayes, yang menunjukkan bahwa persepsi pengguna didominasi oleh sentimen netral dan positif, dengan keunggulan aplikasi dalam kemudahan penggunaan dan kecepatan transaksi. Namun, beberapa kendala teknis, lambatnya respons layanan pelanggan, serta kekhawatiran terhadap keamanan data pribadi menjadi faktor yang berkontribusi terhadap ulasan negatif. Dari segi efektivitas algoritma, SVM memiliki akurasi lebih tinggi (88%) dibandingkan Naïve Bayes (84%), dengan keunggulan dalam mengklasifikasikan ulasan netral, meskipun masih mengalami kesulitan dalam mengenali pola pada ulasan negatif akibat jumlah data yang tidak seimbang. Oleh karena itu, penelitian ini merekomendasikan beberapa langkah untuk peningkatan layanan, di antaranya peningkatan stabilitas aplikasi dan responsivitas layanan pelanggan, penguatan sistem keamanan data dan transparansi kebijakan privasi, serta perbaikan desain antarmuka agar lebih intuitif. Selain itu, penelitian selanjutnya disarankan untuk menyeimbangkan jumlah data ulasan negatif, sehingga model klasifikasi sentimen dapat lebih akurat dalam mengenali pola sentimen yang beragam. Implementasi analisis sentimen berbasis SVM dan Naïve Bayes secara berkelanjutan juga dapat menjadi strategi yang efektif bagi pengelola aplikasi dalam memantau persepsi

pengguna dan mengoptimalkan kualitas layanan di masa mendatang.

DAFTAR PUSTAKA

- [1] Arulogun, O.T., Olatunbosun, A., Fakolujo, O.A. Andika, D., & Susanti, R. (2020). Natural Language Processing and Sentiment Analysis: A Comprehensive Approach. *Journal of Information Technology*, 10(2), 123-135.
- [2] Cambria, E. (2022). Sentiment Analysis: From Opinion Mining to Human-Centered NLP. *IEEE Computational Intelligence Magazine*, 17(3), 41-49.
- [3] Du, Y., Xing, F., Mao, R., & Cambria, E. (2024). Applications of Financial Sentiment Analysis. *International Journal of Artificial Intelligence in Finance*, 12(1), 45-68.
- [4] Gupta, P., & Kaur, S. (2021). Comparative Study of SVM and Naive Bayes for Sentiment Analysis in Banking Sector. *International Journal of Computer Applications*, 175(4), 12-20.
- [5] Hidayat, M., & Suryadi, D. (2021). Privacy and Security in Digital Financial Services: Impacts on User Trust. *Indonesian Journal of Digital Economy*, 8(1), 88-101.
- [6] Johnson, T., & Wicaksono, B. (2023). Machine Learning Approaches for Sentiment Analysis of Financial Applications in Latin America. *Journal of Data Science*, 19(4), 345-360.
- [7] Kurniawan, A., & Lestari, D. (2022). Inclusion and Accessibility in Digital Banking Services. *Economic and Financial Studies*, 5(3), 54-70.
- [8] Li, X., Wang, L., & Zhao, Y. (2022). Handling Informal Language in Banking App Reviews: A Comparison of SVM and Naive Bayes. *Journal of Computational Linguistics*, 30(2), 210-225.
- [9] Manning, C. D., Raghavan, P., & Schütze, H. (2020). Naive Bayes for Large-Scale Sentiment Analysis. *Text Mining and Information Retrieval Journal*, 17(3), 94-108.
- [10] Prasetyo, D., & Lestari, N. (2021). Digital Financial Literacy and User Adoption of Financial Technology Applications. *Journal of Digital Transformation*, 9(1), 78-89.
- [11] Pratama, F., & Nugroho, T. (2023). Challenges and Gaps in User Satisfaction with Digital Financial Platforms. *Southeast Asian Journal of Information Systems*, 11(2), 133-150.
- [12] Ren, Z., Li, H., & Zhao, X. (2022). Deep Learning and SVM for Sentiment Analysis of Banking App Reviews. *Journal of Artificial Intelligence Research*, 34(4), 456-470.
- [13] Santoso, J., & Pratama, R. (2021). An Overview of Sentiment Analysis Techniques in Financial Applications. *Journal of Information Science*, 14(2), 65-80.
- [14] Setiawan, H., & Prasetyo, Y. (2023). User-Centered Design and Product Development in Competitive App Markets. *Indonesian Journal of Business and Technology*, 10(1), 21-33.
- [15] Subramani, R., Gupta, N., & Kaur, S. (2023). Hybrid Sentiment Analysis of Banking App Reviews Using SVM and Naive Bayes. *Journal of Machine Learning Applications*, 27(1), 119-135.
- [16] Taufiq, M., Anggraeni, S., & Nugroho, I. (2021). The Role of Digital Financial Education in Enhancing User Experience. *Journal of Financial Technology Research*, 14(3), 278-295.
- [17] Utami, D., & Susanto, B. (2022). The Integration of Multiple Algorithms for Enhanced Sentiment Analysis Accuracy. *Journal of Computational Intelligence*, 29(5), 345-362.
- [18] Widiyanto, A., Wibowo, H., & Hartono, T. (2021). The Impact of Digital Financial Services on Accessibility and Economic Growth. *Journal of Development Economics*, 19(4), 89-105.
- [19] Wibisono, A., & Saputra, R. (2022). Improving User Experience in Digital Financial Services through Responsive Support Systems. *Journal of Service Quality*, 8(2), 233-250.
- [20] Zhang, Y., Wang, Y., & Li, Z. (2021). Sentiment Analysis of Mobile Banking Apps Using SVM and Naive Bayes. *Journal of Information Technology Research*, 15(2), 198-213..