

Sentiment Analysis of SpeedCash Application Reviews on Google Play Store Using Naïve Bayes and Support Vector Machine Methods

[Analisis Sentimen Ulasan Aplikasi SpeedCash di Google Play Store Menggunakan Metode Naive Bayes dan Support Vector Machine]

Naufal Ainurrochman ¹⁾, Yulian Findawati ^{*,2)}

¹⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: yulianfindawati@umsida.ac.id

Abstract. *The rapid development of digital financial services in Indonesia has driven the widespread use of electronic payment applications such as SpeedCash. User reviews on the Google Play Store serve as a crucial data source for evaluating service quality and user satisfaction. However, manual analysis of these reviews is time-consuming and inefficient. This study aims to analyze user sentiment regarding the SpeedCash application and compare the classification performance of the Naïve Bayes (Multinomial Multinomial) and Support Vector Machine (SVM with linear kernel) algorithms. Research data were collected via web scraping from the Google Play Store, yielding 30.0967 raw reviews and 24.922 valid data points following data cleaning and text preprocessing (case folding, cleaning, tokenization, stopword removal, and Sastrawi stemming). Feature extraction was performed using Term Frequency-Inverse Document Frequency (TF-IDF), with dataset split into (60%:40%, 70%:30%, dan 80%:20%). In classification testing, the Naïve Bayes Bernoulli method with a 70:30 ratio achieved an overall accuracy of 73,47%, with a peak precision of 73,77%, recall of 73,47%, and a F1-score 73,57%. Meanwhile, the SVM RBF method with a 70:30 ratio achieved an accuracy of 89,07%, a peak precision of 89,03%, recall of 89,07%, and a F1-score. Overall performance analysis demonstrates that SVM outperforms Naïve Bayes in effectively identifying and classifying user sentiment trends for the SpeedCash application.*

Keywords - Speedcash, Naïve Bayes, SVM, Sentiment Analysis, E Wallet

Abstrak. *Pesatnya perkembangan layanan keuangan digital di Indonesia mendorong tingginya penggunaan aplikasi pembayaran elektronik seperti SpeedCash. Ulasan pengguna pada platform Google Play Store menjadi sumber data krusial untuk mengevaluasi kualitas layanan dan kepuasan pengguna. Namun, analisis ulasan secara manual membutuhkan waktu yang lama dan kurang efisien. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi SpeedCash serta membandingkan performa metode Naïve Bayes (Multinomial Multinomial) dan Support Vector Machine (SVM dengan kernel linear). Data penelitian dikumpulkan melalui teknik web scraping dari Google Play Store, menghasilkan 30.0967 ulasan mentah dan 24.922 data valid setelah melalui tahap data cleaning dan preprocessing (case folding, cleaning, tokenizing, stopword removal, dan stemming Sastrawi). Ekstraksi fitur dilakukan menggunakan Term Frequency-Inverse Document Frequency (TF-IDF), kemudian data dibagi menjadi (60%:40%, 70%:30%, dan 80%:20%). Dalam pengujian klasifikasi, metode Naïve Bayes Bernoulli dengan rasio 70:30 menghasilkan tingkat akurasi sebesar 73,47%, presisi 73,77%, recall 73,47%, dan F1-score 73,57%. Sementara itu, metode SVM RBF dengan rasio 70:30 menghasilkan akurasi sebesar 89,07%, presisi 89,03%, recall 89,07%, dan F1-score 88,99%. Berdasarkan analisis keseluruhan, metode SVM terbukti memberikan performa yang lebih unggul dalam mengidentifikasi dan mengklasifikasikan kecenderungan sentimen pengguna aplikasi SpeedCash secara akurat.*

Kata Kunci - Speedcash, Naïve Bayes, SVM, Sentiment Analysis, E Wallet

I. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi yang semakin pesat telah mendorong pertumbuhan layanan keuangan digital di Indonesia. Kemajuan teknologi tersebut memberikan kemudahan bagi masyarakat dalam melakukan berbagai aktivitas transaksi secara cepat, praktis, dan efisien melalui aplikasi berbasis mobile. Salah satu bentuk perkembangan tersebut adalah hadirnya aplikasi dompet digital dan layanan pembayaran elektronik yang kini semakin banyak digunakan oleh masyarakat. Aplikasi keuangan digital tidak hanya berfungsi sebagai alat transaksi, tetapi juga menjadi bagian penting dalam mendukung aktivitas ekonomi masyarakat di era digital saat ini[1]. Salah satu aplikasi layanan keuangan digital yang cukup dikenal adalah . Aplikasi ini menyediakan berbagai layanan transaksi seperti pembayaran tagihan, pembelian pulsa, transfer uang, dan layanan pembayaran lainnya. Sebagai aplikasi yang tersedia di Google Play Store, SpeedCash menerima berbagai ulasan dan penilaian dari pengguna.

Ulasan tersebut berisi pengalaman pengguna terhadap kualitas layanan aplikasi, baik berupa komentar positif maupun negatif[1].

Ulasan pengguna pada Google Play Store menjadi sumber informasi yang penting bagi pengembang aplikasi maupun calon pengguna. Melalui ulasan tersebut dapat diketahui tingkat kepuasan pengguna terhadap performa aplikasi, kemudahan penggunaan, keamanan, serta kualitas layanan yang diberikan. Namun, jumlah ulasan yang sangat banyak menyebabkan proses analisis secara manual menjadi kurang efektif dan membutuhkan waktu yang lama. Oleh karena itu, diperlukan suatu metode otomatis yang mampu menganalisis opini pengguna secara sistematis dan efisien.

Analisis sentimen merupakan salah satu teknik dalam bidang Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini atau penilaian pengguna terhadap suatu produk maupun layanan ke dalam kategori tertentu, seperti sentimen positif dan negatif. Dengan menggunakan analisis sentimen, data ulasan pengguna dapat diolah menjadi informasi yang bermanfaat untuk mengetahui kecenderungan opini masyarakat terhadap aplikasi SpeedCash[2].

Berbagai penelitian sebelumnya menunjukkan bahwa metode klasifikasi berbasis machine learning mampu memberikan hasil yang baik dalam analisis sentimen. Penelitian tersebut menunjukkan bahwa metode Naive Bayes memiliki performa yang baik dalam klasifikasi sentimen ulasan aplikasi dengan tingkat akurasi yang cukup tinggi. menunjukkan bahwa metode Support Vector Machine (SVM) mampu menghasilkan tingkat akurasi yang lebih stabil dalam pengolahan data teks berdimensi besar[3].

Metode Naive Bayes merupakan algoritma klasifikasi yang sederhana namun efektif dalam pengolahan data teks karena memiliki proses perhitungan yang cepat dan mudah diterapkan. Sementara itu, Support Vector Machine (SVM) merupakan metode klasifikasi yang mampu bekerja dengan baik pada data berdimensi tinggi dan memiliki tingkat akurasi yang tinggi dalam proses klasifikasi sentimen. Oleh karena itu, kedua metode tersebut dipilih untuk dibandingkan performanya dalam penelitian ini[4].

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk menganalisis sentimen ulasan pengguna aplikasi SpeedCash di Google Play Store menggunakan metode Naive Bayes dan Support Vector Machine (SVM), serta membandingkan tingkat akurasi kedua metode tersebut guna menentukan metode yang paling efektif dalam klasifikasi sentimen pengguna aplikasi SpeedCash.

Tinjauan Pustaka

Analisis Sentimen

Analisis sentimen (sentiment analysis) atau opinion mining adalah proses komputasi untuk mengidentifikasi dan mengekstrak informasi subjektif dari teks, seperti opini, emosi, dan sikap penulis terhadap suatu topik tertentu[5], analisis sentimen dapat didefinisikan sebagai bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, topik, dan atribut mereka.

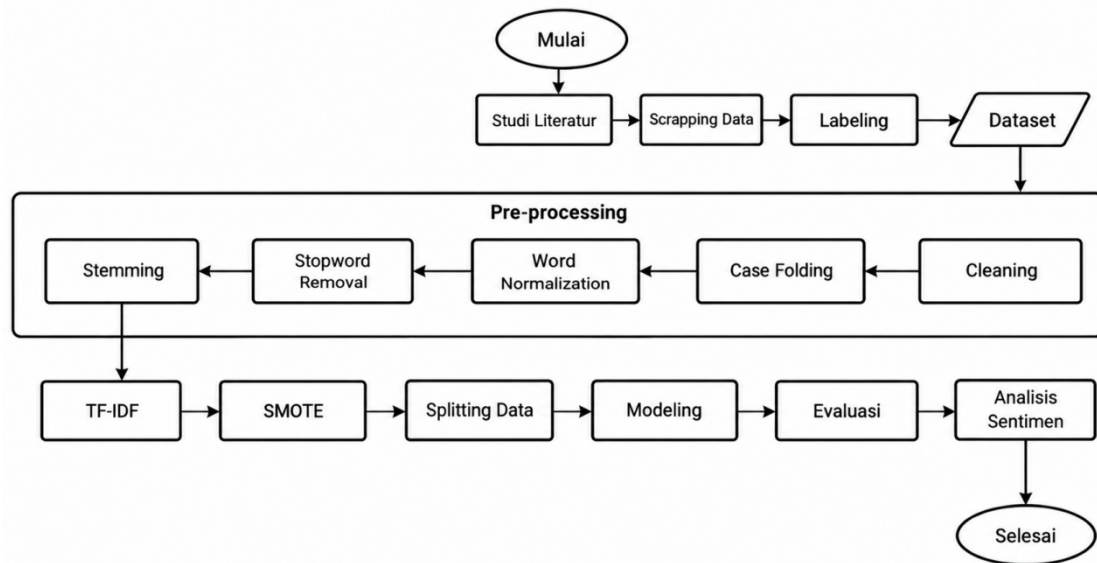
Terdapat tiga level analisis sentimen, yaitu: (1) Document-level sentiment analysis, yang bertujuan mengklasifikasikan sentimen keseluruhan dokumen; (2) Sentence-level sentiment analysis, yang mengklasifikasikan sentimen setiap kalimat; dan (3) Aspect-level sentiment analysis, yang mengidentifikasi sentimen terhadap aspek atau fitur spesifik dari suatu entitas[4]. Dalam penelitian ini, analisis sentimen dilakukan pada level dokumen, di mana setiap ulasan atau rating dianggap sebagai satu dokumen yang akan diklasifikasikan ke dalam kategori sentimen positif, negatif, atau netral.

Text Mining

Text mining atau text data mining adalah proses mengekstrak informasi dan pengetahuan yang bermanfaat dari data tekstual yang tidak terstruktur atau semi-terstruktur[6]. mendefinisikan text mining sebagai proses penemuan pengetahuan yang sebelumnya tidak diketahui dari data tekstual.

Proses text mining umumnya melibatkan beberapa tahapan, yaitu: (1) Text preprocessing, yang meliputi cleaning, tokenization, filtering, dan stemming; (2) Feature extraction, untuk mengubah teks menjadi representasi numerik; (3) Model building, menggunakan algoritma machine learning; dan (4) Evaluation, untuk mengukur performa model[7].

II. METODE



Gambar 1. Alur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan sistematis guna mencapai hasil yang diharapkan, sebagaimana ditampilkan pada Gambar 1 dijelaskan pada tahap 2.1 sampai 2.10.

Dataset

Dataset dalam penelitian ini merupakan kumpulan ulasan pengguna aplikasi SpeedCash yang diperoleh dari Google Play Store diperoleh sebanyak 30.0967 **data ulasan** yang digunakan sebagai dataset penelitian. Setiap data ulasan memiliki label sentimen yang terdiri atas **positif, netral, dan negatif**. Dataset tersebut selanjutnya digunakan dalam proses preprocessing, pembobotan kata menggunakan TF-IDF, penyeimbangan data menggunakan SMOTE pada data latih, serta proses klasifikasi menggunakan metode Naïve Bayes dan Support Vector Machine (SVM).

Study Literatur

Metode komparatif digunakan dalam penelitian ini dengan pendekatan kuantitatif karena berfokus pada pengolahan data numerik hasil klasifikasi sentimen. Berbagai penelitian terdahulu telah menerapkan algoritma Naïve Bayes dan Support Vector Machine (SVM) untuk analisis sentimen pada berbagai topik, seperti ulasan aplikasi, permainan digital, hingga opini pengguna media sosial[3]. Kedua metode tersebut umumnya dipadukan dengan tahapan preprocessing teks dan pembobotan TF-IDF sebelum proses klasifikasi dilakukan. Hasil penelitian sebelumnya menunjukkan bahwa SVM cenderung menghasilkan performa klasifikasi yang lebih baik dibandingkan Naïve Bayes pada beberapa kasus analisis sentimen, terutama dalam menangani data teks dengan variasi kata yang tinggi.

Namun, kajian yang secara khusus membandingkan kedua metode pada sentimen masyarakat terhadap aplikasi Speedcash di platform google play store masih terbatas. Oleh karena itu, penelitian ini dilakukan untuk membandingkan kinerja Naïve Bayes dan SVM dalam mengklasifikasikan sentimen pengguna terhadap aplikasi speedcash sebagai bahan evaluasi peningkatan kualitas layanan.

Scraping data

Scraping data Data yang digunakan dalam penelitian ini ialah data sekunder yang berbentuk 30.0967 ulasan publik yang diambil dari platform google play store. Data tersebut berisi beragam opini, keluhan, kritik, dan apresiasi masyarakat terkait layanan Aplikasi Speedcash. Pemilihan data dilakukan dengan mencari komentar yang memiliki hubungan dengan topik penelitian, serta hanya diambil ulasan yang menggunakan bahasa Indonesia agar sesuai dengan konteks analisis sentimen yang dilakukan[8].

Pelabelan Data

Pada tahap pelabelan data, ulasan yang telah melalui proses scraping akan diberi label secara otomatis ke dalam tiga kategori sentimen yaitu positif, negatif, dan netral. Proses pelabelan ini dilakukan oleh peneliti untuk memastikan kualitas dan akurasi data latih yang digunakan dalam proses klasifikasi[5].

Sentimen positif diberikan pada ulasan yang berisi apresiasi atau kepuasan terhadap pelayanan Aplikasi Speedcash, sentimen negatif diberikan pada ulasan yang mengandung keluhan atau kritik, sedangkan sentimen netral

diberikan pada ulasan yang bersifat informatif atau tidak menunjukkan kecenderungan emosi tertentu. Dengan dilakukan pelabelan secara manual, diharapkan data yang digunakan dalam proses pelatihan model memiliki tingkat keakuratan yang lebih tinggi sehingga dapat meningkatkan performa model klasifikasi yang dihasilkan.

Preprocessing Data Teks

Pada tahapan preprocessing dilakukan pembersihan data dan juga dilakukan perapian agar teks ledih rapih[2]. Pada tahapan ini dilakukan beberapa tahapan seperti, case folding bertujuan untuk mengubah semua huruf menjadi huruf kecil, cleansing bertujuan untuk menghapus elemen-elemen yang tidak diperlukan, tokenizing memiliki tujuan memecah kalimat menjadi token/kata, stopword removal bertujuan untuk menghapus kata-kata yang dianggap tidak memiliki makna penting, serta stemming mengubah kata ke bentuk dasar menggunakan Sastrawi. Tahap tersebut dilakukan bertujuan guna menghasilkan data teks yang bersih dan sudah siap untuk digunakan pada proses ekstraksi fitur

Pembagian Data

Pada tahapan ini dilakukan pembagian dataset menjadi dua bagian, yaitu data latih dan data uji. Pada penelitian ini digunakan tiga skema pembagian data, yaitu rasio 80:20, 70:30, dan 60:40. Penggunaan tiga skema ini bertujuan untuk melihat secara komprehensif pengaruh proporsi ketersediaan data latih terhadap kinerja model klasifikasi. Kegunaan dari data latih ialah untuk membangun dan melatih model klasifikasi, sedangkan data uji dipakai semata mata untuk mengukur kemampuan model dalam mengklasifikasikan sentimen ulasan yang belum pernah dilihat sebelumnya.

Salah satu tantangan dalam analisis sentimen adalah distribusi jumlah kelas yang tidak seimbang imbalanced data, yang dapat menyebabkan model menjadi bias ke arah kelas mayoritas[7]. Untuk mengatasi hal ini, diterapkan teknik Oversampling menggunakan metode SMOTE (Synthetic Minority Over-sampling Technique). Proses SMOTE dilakukan dengan membangkitkan data sintesis baru pada kelas minoritas berdasarkan kedekatan antar data yang memiliki karakteristik serupa pada ruang fitur, misalnya hasil representasi TF-IDF. Berbeda dengan duplikasi data secara langsung, SMOTE menghasilkan sampel baru yang dibentuk dari kombinasi nilai fitur antara data minoritas yang berdekatan sehingga distribusi data menjadi lebih seimbang. Penting untuk dicatat bahwa proses oversampling ini hanya diterapkan pada data latih (training data) setelah tahap pembagian data (data splitting) selesai. Data uji (test data) tetap dibiarkan dalam kondisi asli tanpa penambahan data sintesis. Tujuan dari langkah ini adalah agar model dapat mempelajari pola dari distribusi kelas yang lebih seimbang tanpa menimbulkan kebocoran data (data leakage) pada proses pengujian, sehingga hasil evaluasi model menjadi lebih objektif dan dapat dipercaya.

Pembobotan TF-IDF

Ekstraksi fitur teks dilakukan dengan menggunakan sebuah metode bernama Term Frequency–Inverse Document Frequency (TF-IDF) digunakan untuk merepresentasikan data teks ke dalam bentuk numerik[2]. Pembobotan TF-IDF digunakan untuk memberikan penekanan pada kata-kata yang memiliki tingkat kepentingan lebih tinggi dalam suatu dokumen dibandingkan keseluruhan korpus. Secara konseptual, pembobotan TF-IDF dapat dirumuskan sebagai berikut:

$$TF - IDF(t, d) = TF(t, d) \times \log \left(\frac{N}{DF(t)} \right)$$

Pada persamaan tersebut, $TF(t, d)$ merujuk pada frekuensi kemunculan kata t di dalam dokumen d , $DF(t)$ menunjukkan banyaknya dokumen yang memuat kata t , sedangkan N adalah jumlah keseluruhan dokumen yang terdapat dalam korpus. Dalam penelitian ini, tahap ekstraksi fitur dijalankan menggunakan tools dengan parameter TF-IDF dari library scikit-learn. Proses ini diterapkan secara konsisten pada data latih maupun data uji dengan mengacu pada vocabulary yang dibangun eksklusif dari data latih, kemudian diberlakukan pada data uji tanpa modifikasi apapun[9]. Keluaran dari tahap ini berupa matriks numerik yang selanjutnya diumpankan sebagai input ke dalam proses pelatihan dan evaluasi kedua model klasifikasi sentimen. dengan penekanan bahwa input TF-IDF berasal dari data latih yang sudah di-oversampling.

Pelatihan Model Naive Bayes

Naïve Bayes Multinomial merupakan metode klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen[10]. terbukti efektif dalam berbagai aplikasi, khususnya klasifikasi teks [11]. Metode ini digunakan untuk menghitung jumlah probabilitas suatu dokumen yang termasuk ke dalam sebuah kelas tertentu berdasarkan kata-kata yang terkandung di dalamnya. Berikut adalah rumusnya:

$$P(H|X) = (P(X|C) \times P(C)) / P(X)$$

Persamaan tersebut menegaskan bahwa penetapan kelas sentimen pada sebuah ulasan ditentukan oleh dua faktor yang saling melengkapi. Nilai $P(X|C)$ mencerminkan seberapa besar kemungkinan suatu fitur atau kata muncul dalam kelas sentimen tertentu, sementara $P(C)$ merepresentasikan bobot awal tiap kelas berdasarkan proporsinya di dalam keseluruhan data. Adapun $P(X)$ berperan sebagai konstanta penyeimbang yang memastikan keluaran probabilitas tetap

berada dalam rentang yang valid[12]. Dalam pelaksanaan penelitian ini, kalkulasi tersebut tidak dijalankan secara manual mengikuti persamaan di atas, melainkan diserahkan sepenuhnya kepada implementasi komputasional dengan memanfaatkan library Naïve Bayes dari paket scikit-learn. Untuk memastikan eksperimen yang mendalam, penelitian ini menguji 4 parameter Naïve Bayes, yaitu: $\alpha = 0.1$, $\alpha = 0.5$, $\alpha = 1.0$, dan $\alpha = 2.0$. Berkebalikan dengan SVM yang mengandalkan margin spasial, Naïve Bayes bekerja penuh berdasarkan teori hitung peluang (probabilitas statistik). Setelan tambahan berupa `fit_prior = True` juga dideklarasikan agar model tetap patuh pada persentase awal (prior probability) kelas yang masuk saat fase pelatihan.

Pelatihan Model SVM

Support Vector Machine (SVM) merupakan metode klasifikasi yang bekerja dengan mencari hyperplane optimal yang dapat memisahkan data ke dalam kelas yang berbeda[13]. SVM dikenal efektif dalam menangani data berdimensi tinggi, seperti data teks pada analisis sentimen. Metode ini berusaha memaksimalkan margin atau jarak antara data terdekat dari masing-masing kelas terhadap hyperplane. SVM dimodifikasi untuk menyertakan fungsi kernel untuk menemukan hasil dengan cepat[11]. Berikut adalah rumusnya:

$$f(x) = \text{sign}(w \cdot x + b)$$

Persamaan di atas merepresentasikan fungsi keputusan yang bekerja pada Support Vector Machine, di mana setiap vektor fitur x dipetakan ke dalam ruang berdimensi tinggi melalui kombinasi linier antara bobot w dan nilai bias b guna menetapkan posisi data relatif terhadap hyperplane pemisah. Hasil dari fungsi $f(x)$ kemudian dijadikan penentu kelas sentimen, di mana polaritas keluarannya positif atau negatif secara langsung mencerminkan ke kelas mana suatu data diklasifikasikan. Dalam konteks penelitian ini, penerapan algoritma SVM tidak dikerjakan melalui kalkulasi matematis secara eksplisit, melainkan diimplementasikan menggunakan library SVM dari paket scikit learn. Pada rumpun algoritma Support Vector Machine, fondasi pemisah kelas (hyperplane) sangat dipengaruhi oleh jenis kernel yang dipilih. Dalam studi komparatif ini, empat fungsi kernel diuji secara simultan guna melihat fleksibilitas SVM dalam menghadapi teks yang berdimensi tinggi[14]. Nilai toleransi penalti atau parameter C ditetapkan secara seragam di angka 1.0 agar margin yang terbentuk bersifat moderat (tidak terlalu lunak dan tidak terlalu kaku). Adapun siklus literasi pencarian (`max_iter`) dikunci pada batas maksimal 2000 putaran guna menghindari beban komputasi laten.

Evaluasi Model

Evaluasi model dilakukan menggunakan confusion matrix dengan metrik Accuracy, recision, Recall, dan F1 Score. Tujuan dari tahapan ini ialah untuk mengetahui hasil evaluasi dari kedua algoritma, yang kemudian akan dilakukan perbandingan guna mengetahui metode mana yang memberikan hasil dan performa terbaik dalam analisis sentimen pelayanan Aplikasi Speedcash. Untuk mengetahui kinerja model klasifikasi yang digunakan, dilakukan proses evaluasi menggunakan confusion matrix. Berdasarkan confusion matrix, diperoleh beberapa metrik evaluasi seperti accuracy, precision, recall, dan f1-score. Accuracy menunjukkan tingkat ketepatan model dalam mengklasifikasikan data secara keseluruhan[15].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

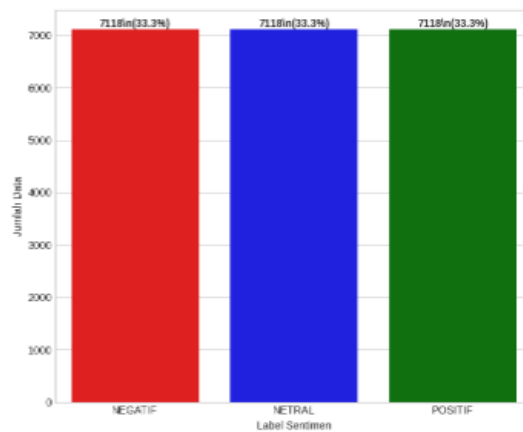
$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Dari sisi teoritis, accuracy mencerminkan sejauh mana model mampu menghasilkan prediksi yang tepat secara agregat terhadap seluruh data uji. Precision mengukur tingkat presisi model ketika menetapkan suatu data ke dalam kelas tertentu, sementara recall menilai kapasitas model dalam menjangkau seluruh data yang sejatinya termasuk ke dalam kelas yang dimaksud. Adapun F1-score hadir sebagai representasi harmonik antara keduanya, memberikan gambaran yang lebih berimbang ketika terdapat ketimpangan antara nilai precision dan recall[16]. Pada implementasinya dalam penelitian ini, keempat metrik tersebut tidak dihitung secara manual, melainkan diperoleh secara otomatis melalui fungsi `classification report` dan confusion matrix yang disediakan oleh scikit-learn, sehingga konsistensi dan akurasi proses evaluasi dapat terjaga di setiap skenario pengujian.

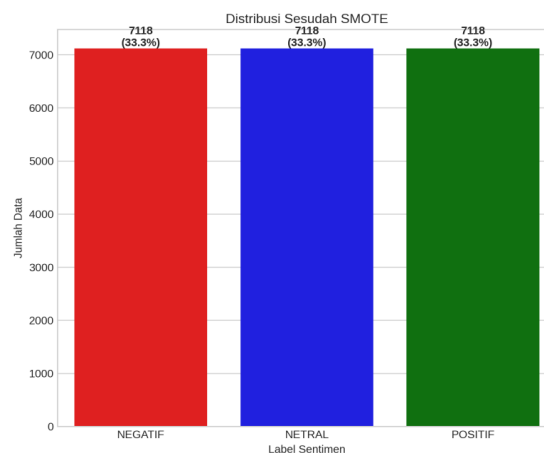
III. HASIL DAN PEMBAHASAN

Distribusi Dataset

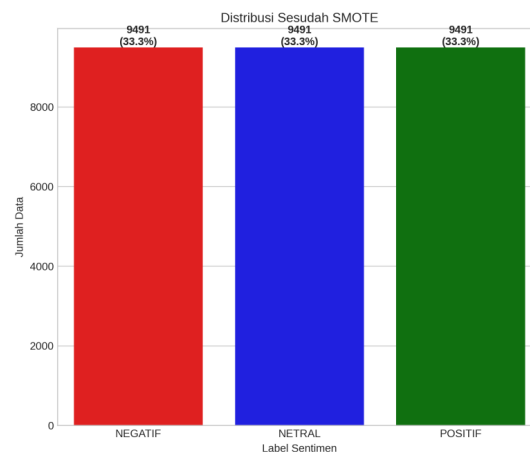
Dataset akhir sebanyak 24.876 ulasan dibagi menjadi 14.925 data latih dan 9.951 data uji untuk rasio 60:40, 17.413 data latih dan 7.463 data uji untuk rasio 70:30, dan untuk rasio 80:20 19.900 data latih dan 4.976 data uji. Kedua algoritma, yaitu Naive Bayes dan SVM, dilatih menggunakan data latih yang sama dan diuji pada data uji yang sama pula agar perbandingan performa kedua model dapat dilakukan secara adil. Distribusi label sentimen pada dataset akhir relatif seimbang antar ketiga kelas, sebagaimana ditunjukkan pada Gambar 2,3, dan 4 sehingga model tidak bias terhadap kelas mayoritas tertentu. Hasil evaluasi kedua model pada gambar 2,3, dan 4



Gambar 2. Distribusi Label Sentimen Dataset Rasio 60:40 (Total 21.354 ulasan)

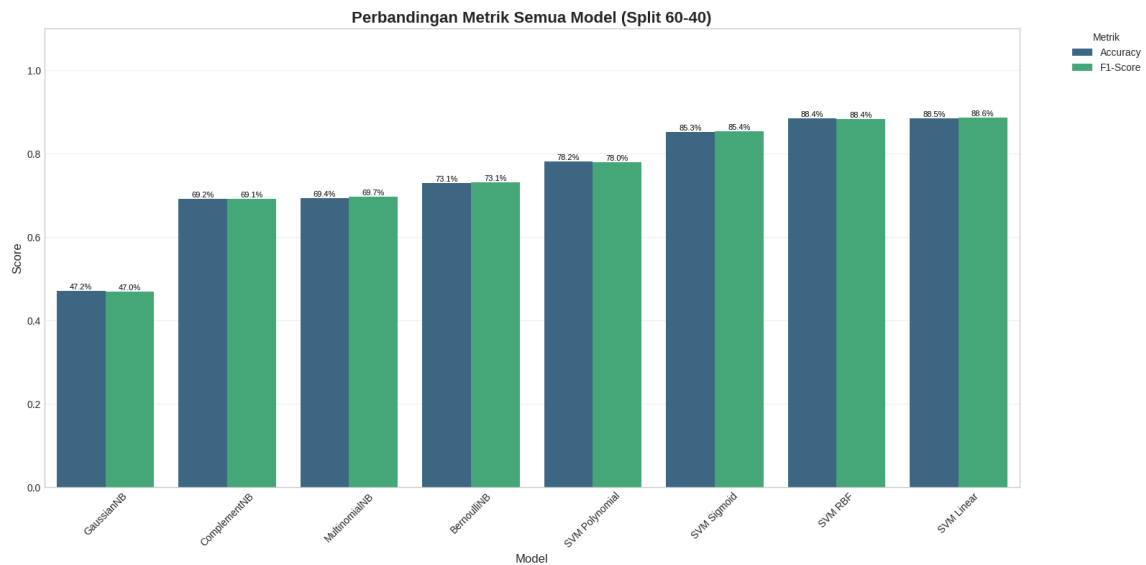


Gambar 3. Distribusi Label Sentimen Dataset Rasio 70:30 (Total 24.915 ulasan)



Gambar 4. Distribusi Label Sentimen Dataset Rasio 80:20 (Total 28.473 ulasan)

Hasil Evaluasi Model

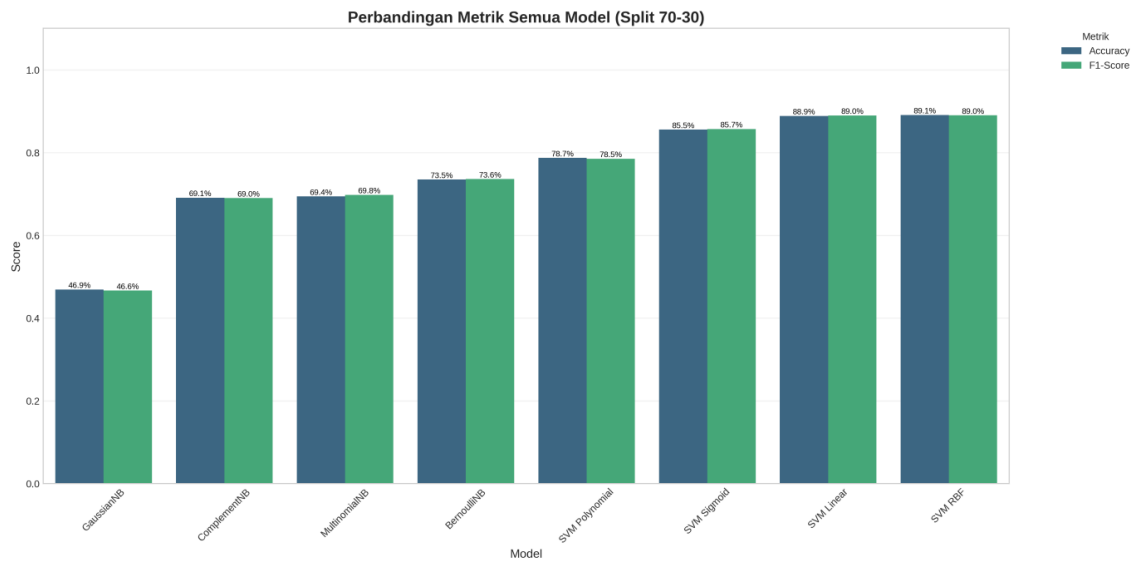


Gambar 5. Perbandingan Persentase Accuracy, Precision, Recall, dan F1-Score Naive Bayes dan SVM dengan rasio 60:40

Tabel 1. Hasil evaluasi model pada rasio 60% Latih dan 40% Uji

Model	Kernel / Varian	F1-Score (Weighted)	Accuracy	Keterangan
SVM	Linear	88,64%	88,52%	Terbaik F1 dan acc
SVM	RBF	88,35%	88,44%	-
SVM	Sigmoid	85,42%	85,27%	-
SVM	Polynomial	77,95%	78,21%	-
Naïve Bayes	Bernoulli	73,14%	73,08%	Terbaik F1 dan acc (nb)
Naïve Bayes	Multinomial	69,70%	69,42%	-
Naïve Bayes	Complement	69,14%	69,23%	-
Naïve Bayes	Gaussian	46,70%	47,21%	Terendah

Pada skenario pembagian data 60% data latih dan 40% data uji, seluruh model tetap mampu memberikan hasil klasifikasi yang baik setelah diterapkan Synthetic Minority Over-sampling Technique (SMOTE) pada bagian data latih saja. SVM (Linear) memperoleh F1-Score tertinggi sebesar 88,64% sekaligus Accuracy tertinggi sebesar 88,52%. SVM (RBF) menghasilkan F1-Score sebesar 88,35% dan Accuracy sebesar 88,44%, sedangkan SVM (Sigmoid) memperoleh F1-Score sebesar 85,42% dan Accuracy sebesar 85,27%. SVM (Polynomial) menunjukkan performa terendah pada kelompok SVM dengan F1-Score sebesar 77,95% dan Accuracy sebesar 78,21%. Pada kelompok Naïve Bayes, Bernoulli Naïve Bayes memberikan performa terbaik dengan F1-Score sebesar 73,14% dan Accuracy sebesar 73,08%, diikuti oleh Multinomial Naïve Bayes dengan F1-Score sebesar 69,70% dan Accuracy sebesar 69,42%. Selanjutnya, Complement Naïve Bayes memperoleh F1-Score sebesar 69,14% dan Accuracy sebesar 69,23%, sedangkan Gaussian Naïve Bayes menjadi varian dengan performa terendah dengan F1-Score sebesar 46,97% dan Accuracy sebesar 47,21%. Berdasarkan confusion matrix pada Gambar 5 dan tabel 1.

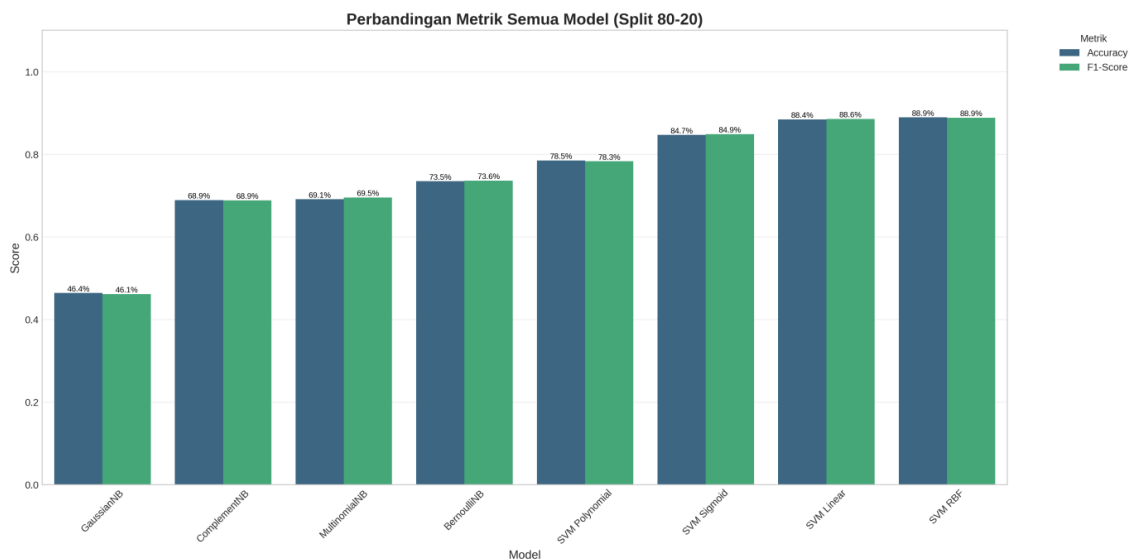


Gambar 6. Perbandingan Persentase Accuracy, Precision, Recall, dan F1-Score Naive Bayes dan SVM dengan rasio 70:30

Tabel 2. Hasil evaluasi model pada rasio 70% Latih dan 30% Uji

Model	Kernel / Varian	F1-Score (Weighted)	Accuracy	Keterangan
SVM	Linear	88,96%	88,85%	-
SVM	RBF	88,99%	89,07%	Terbaik F1 dan acc
SVM	Sigmoid	85,69%	85,54%	-
SVM	Polynomial	78,49%	78,71%	-
Naïve Bayes	Bernoulli	73,57%	73,47%	Terbaik F1 dan acc (nb)
Naïve Bayes	Multinomial	69,76%	69,40%	-
Naïve Bayes	Complement	68,99%	69,07%	-
Naïve Bayes	Gaussian	46,64%	46,90%	Terendah

Pada skenario pembagian data 70% data latih dan 30% data uji, seluruh model tetap mampu memberikan hasil klasifikasi yang baik setelah diterapkan Synthetic Minority Over-sampling Technique (SMOTE) pada bagian data latih saja. SVM (RBF) memperoleh F1-Score tertinggi sebesar 88,99% sekaligus Accuracy tertinggi sebesar 89,07%. SVM (Linear) menghasilkan F1-Score sebesar 88,96% dan Accuracy sebesar 88,85%, sedangkan SVM (Sigmoid) memperoleh F1-Score sebesar 85,69% dan Accuracy sebesar 85,54%. SVM (Polynomial) menunjukkan performa terendah pada kelompok SVM dengan F1-Score sebesar 78,49% dan Accuracy sebesar 78,71%. Pada kelompok Naïve Bayes, Bernoulli Naïve Bayes memberikan performa terbaik dengan F1-Score sebesar 73,57% dan Accuracy sebesar 73,47%, diikuti oleh Multinomial Naïve Bayes dengan F1-Score sebesar 69,76% dan Accuracy sebesar 69,40%. Selanjutnya, Complement Naïve Bayes memperoleh F1-Score sebesar 68,99% dan Accuracy sebesar 69,07%, sedangkan Gaussian Naïve Bayes menjadi varian dengan performa terendah dengan F1-Score sebesar 46,64% dan Accuracy sebesar 46,90%. Berdasarkan confusion matrix pada Gambar 6 dan tabel 2.



Gambar 7. Perbandingan Persentase Accuracy, Precision, Recall, dan F1-Score Naive Bayes dan SVM dengan rasio 80:20

Tabel 3. Hasil evaluasi model pada rasio 80% Latih dan 20% Uji

Model	Kernel / Varian	F1-Score (Weighted)	Accuracy	Keterangan
SVM	Linear	88,56%	88,44%	-
SVM	RBF	88,85%	88,93%	Terbaik F1 dan acc
SVM	Sigmoid	84,87%	84,69%	-
SVM	Polynomial	78,30%	78,50%	-
Naïve Bayes	Bernoulli	73,61%	73,47%	Terbaik F1 dan acc (nb)
Naïve Bayes	Multinomial	69,53%	69,09%	-
Naïve Bayes	Complement	68,85%	68,91%	-
Naïve Bayes	Gaussian	46,15%	46,40%	Terendah

Pada skenario pembagian data 80% data latih dan 20% data uji, seluruh model tetap mampu memberikan hasil klasifikasi yang baik setelah diterapkan Synthetic Minority Over-sampling Technique (SMOTE) pada bagian data latih saja. SVM (RBF) memperoleh F1-Score tertinggi sebesar 88,85% sekaligus Accuracy tertinggi sebesar 88,93%. SVM (Linear) menghasilkan F1-Score sebesar 88,56% dan Accuracy sebesar 88,44%, sedangkan SVM (Sigmoid) memperoleh F1-Score sebesar 84,87% dan Accuracy sebesar 84,69%. SVM (Polynomial) menunjukkan performa terendah pada kelompok SVM dengan F1-Score sebesar 78,30% dan Accuracy sebesar 78,50%. Pada kelompok Naïve Bayes, Bernoulli Naïve Bayes memberikan performa terbaik dengan F1-Score sebesar 73,61% dan Accuracy sebesar 73,47%, diikuti oleh Multinomial Naïve Bayes dengan F1-Score sebesar 69,53% dan Accuracy sebesar 69,09%. Selanjutnya, Complement Naïve Bayes memperoleh F1-Score sebesar 68,85% dan Accuracy sebesar 68,91%, sedangkan Gaussian Naïve Bayes menjadi varian dengan performa terendah dengan F1-Score sebesar 46,15% dan Accuracy sebesar 46,40%. Berdasarkan confusion matrix pada Gambar 7 dan tabel 3.

Perbandingan Rasio Data Latih dan Data Uji

Tabel 4 berikut merangkum perbandingan lengkap seluruh hasil evaluasi keenam belas konfigurasi pengujian berdasarkan F1-Score dan Accuracy tertinggi per skenario:

Tabel 4. Rekap Perbandingan Kinerja Seluruh Model di Tiga Skenario SVM

Rasio	Model	Accuracy	Precision	Recall	F1-Score
60:40	SVM Linear	88,52%	88,89%	88,52%	88,64%
70:30	SVM Linear	88,85%	89,16%	89,16%	88,96%
80:20	SVM Linear	88,44%	88,78%	88,44%	88,56%
60:40	SVM RBF	88,44%	88,41%	88,44%	88,35%
70:30	SVM RBF	89,07%	89,03%	89,07%	88,99%
80:20	SVM RBF	88,93%	88,89%	88,93%	88,85%
60:40	SVM Sigmoid	85,27%	85,90%	85,27%	85,42%
70:30	SVM Sigmoid	85,54%	86,12%	85,54%	85,69%
80:20	SVM Sigmoid	84,69%	85,37%	84,69%	84,87%
60:40	SVM Polynomial	78,21%	78,72%	78,21%	77,95%
70:30	SVM Polynomial	78,71%	79,20%	78,71%	78,49%
80:20	SVM Polynomial	78,50%	79,17%	78,50%	78,30%

Tabel 4 menunjukkan hasil evaluasi SVM pada tiga skenario rasio pembagian data, yaitu 80:20, 70:30, dan 60:40. SVM memperoleh accuracy tertinggi pada ketiga skenario dengan performa terbaik dicapai pada rasio 80:20 (accuracy 68,45%), diikuti rasio 70:30 (70,55%), dan rasio 60:40 (74,81%). Pola ini menunjukkan bahwa semakin besar proporsi data latih yang tersedia, semakin baik pula kemampuan SVM dalam mengenali pola sentimen pada data ulasan.

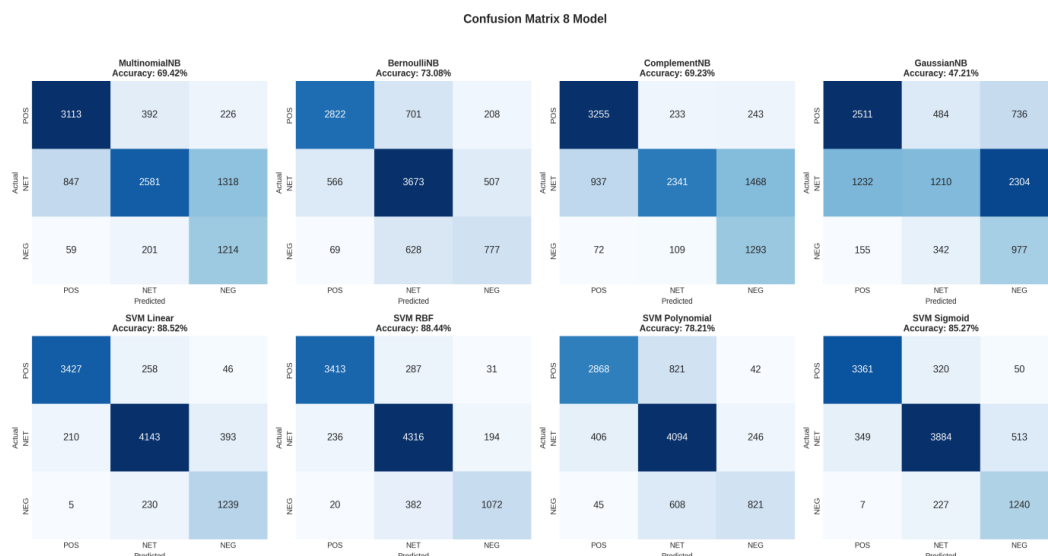
Tabel 5. Rekap Perbandingan Kinerja Seluruh Model di Tiga Skenario Naive Bayes

Rasio	Model	Accuracy	Precision	Recall	F1-Score
60:40	Bernoulli NB	73,08%	73,34%	73,08%	73,14%
70:30	Bernoulli NB	73,47%	73,77%	73,47%	73,57%
80:20	Bernoulli NB	73,47%	73,83%	73,47%	73,61%
60:40	Multinomial NB	69,42%	74,34%	69,42%	69,70%
70:30	Multinomial NB	69,40%	74,57%	69,40%	69,76%
80:20	Multinomial NB	69,09%	76,66%	69,09%	69,53%
60:40	Complement NB	69,23%	76,61%	69,23%	69,14%
70:30	Complement NB	69,07%	76,95%	69,07%	68,99%
80:20	Complement NB	68,91%	76,66%	68,91%	68,85%
60:40	Gaussian NB	47,21%	56,10%	44,21%	46,70%
70:30	Gaussian NB	46,90%	56,73%	46,90%	46,64%
80:20	Gaussian NB	46,40%	56,19%	46,40%	46,15%

Tabel 5 menunjukkan hasil evaluasi Naive Bayes Naive Bayes justru memperoleh accuracy tertinggi pada rasio 80:20 (69,49%), lebih baik dibandingkan accuracy tertinggi pada rasio 70:30 (69,01%) dan accuracy tertinggi pada rasio 60:40 (68,90%). Hal ini mengindikasikan bahwa performa Naive Bayes pada penelitian ini tidak selalu

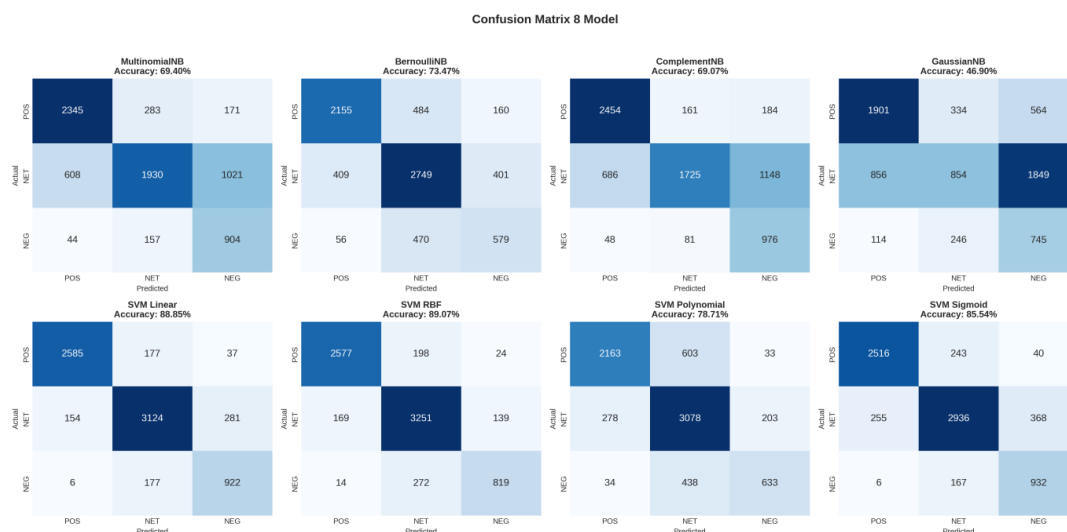
meningkat seiring bertambahnya data latih, kemungkinan karena asumsi independensi antar fitur pada Naive Bayes membuat model kurang sensitif terhadap penambahan volume data dibandingkan SVM.

Confusion Matrix Rasio



Gambar 8. Confusion Matrix Model Naive Bayes SVM pada Rasio 60% Latih dan 40% Uji

Confusion matrix menunjukkan bahwa SVM Linear paling baik mengenali kelas Positif (3.427 dari 3.731 data uji diklasifikasikan dengan benar) dan kelas Netral (4.143 dari 4.746 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 1.239 dari 1.474 sedangkan untuk Naive Bayes Bernoulli paling baik mengenali kelas Positif (2.822 dari 3.731 data uji diklasifikasi dengan benar) dan kelas Netral (3.673 dari 4.746 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 777 dari 1.474 data yang diklasifikasikan dengan benar dan sisanya lebih sering keliru diklasifikasikan sebagai Negatif atau Netral. Kesalahan ini kemungkinan disebabkan oleh kemiripan kosakata antara ulasan Positif dan Netral yang bersifat informatif, seperti penyampaian apresiasi yang disampaikan secara tidak eksplisit. Visualisasi confusion matrix tersebut disajikan pada Gambar 8.



Gambar 9. Confusion Matrix Model Naive Bayes SVM pada Rasio 70% Latih dan 30% Uji

Confusion matrix menunjukkan bahwa SVM RBF paling baik mengenali kelas Positif (2.577 dari 2.799 data uji diklasifikasikan dengan benar) dan kelas Netral (3.251 dari 3.559 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 819 dari 1.105 sedangkan untuk Naive Bayes Bernoulli paling baik mengenali kelas Positif (2.155 dari 2.799 data uji diklasifikasi dengan benar) dan kelas Netral (2.749 dari 3.559 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 579 dari 1.105 data yang diklasifikasikan dengan benar

dan sisanya lebih sering keliru diklasifikasikan sebagai Negatif atau Netral. Kesalahan ini kemungkinan disebabkan oleh kemiripan kosakata antara ulasan Positif dan Netral yang bersifat informatif, seperti penyampaian apresiasi yang disampaikan secara tidak eksplisit. Visualisasi confusion matrix tersebut disajikan pada Gambar 9.



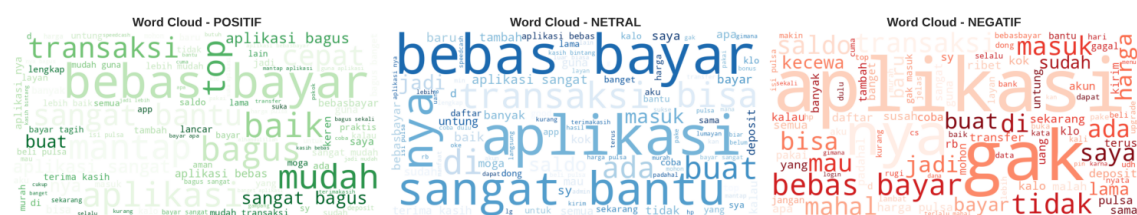
Gambar 10. Confusion Matrix Model Naive Bayes SVM pada Rasio 80% Latih dan 20% Uji

Confusion matrix menunjukkan bahwa SVM RBF paling baik mengenali kelas Positif (1.716 dari 1.866 data uji diklasifikasikan dengan benar) dan kelas Netral (2.162 dari 2.373 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 547 dari 737 sedangkan untuk Naive Bayes Bernoulli paling baik mengenali kelas Positif (1.430 dari 1.866 data uji diklasifikasi dengan benar) dan kelas Netral (1.817 dari 2.373 data), namun masih cukup banyak kesalahan pada kelas Negatif, di mana hanya 409 dari 737 data yang diklasifikasikan dengan benar dan sisanya lebih sering keliru diklasifikasikan sebagai Negatif atau Netral. Kesalahan ini kemungkinan disebabkan oleh kemiripan kosakata antara ulasan Positif dan Netral yang bersifat informatif, seperti penyampaian apresiasi yang disampaikan secara tidak eksplisit. Visualisasi confusion matrix tersebut disajikan pada Gambar 10.

WordCloud

Berdasarkan wordcloud Gambar 5 (a) pada kelas Positif, terlihat bahwa kata-kata yang paling sering muncul adalah Aplikasi, bayar, bagus, mudah, sangat, bebas, mantap, dan transaksi. Kata “Aplikasi” menjadi kata yang paling dominan, yang menunjukkan kepuasan pengguna Aplikasi SpeedCash. Selain itu, munculnya kata seperti “nya”, dan “baik” menunjukkan adanya apresiasi user terhadap aplikasi serta bermanfaatnya fitur yang sudah disediakan. Berdasarkan wordcloud Gambar 5 (b) pada kelas netral, terlihat bahwa kata-kata yang paling sering muncul adalah aplikasi, bayar, sangat, bantu, bebas, nya, pulsa, dan baru. Kemunculan kata-kata tersebut menunjukkan bahwa sebagian besar ulasan netral membahas informasi atau kondisi layanan aplikasi SpeedCash tanpa memberikan opini yang terlalu positif maupun negatif. Selain itu, terdapat kata seperti mau, dan masuk yang menunjukkan pembahasan terkait penggunaan aplikasi dan fitur di dalamnya. Wordcloud ini menggambarkan bahwa sentimen netral didominasi oleh percakapan informatif mengenai layanan dan aplikasi SpeedCash.

Berdasarkan wordcloud Gambar 5 (c) pada kelas negatif, terlihat bahwa kata-kata yang paling sering muncul adalah aplikasi, gak, bayar, pulsa, nya, mahal, masuk, dan harga. Kata “saldo” dan “tidak” menunjukkan adanya ketidakpuasan user terhadap fitur dan pelayanan aplikasi SpeedCash. Selain itu, munculnya kata seperti saldo, dan tidak menggambarkan adanya error atau masalah pada fitur yang tersedia. Wordcloud ini menunjukkan bahwa sentimen Negatif didominasi oleh tanggapan user yang merasa tidak puas karena adanya kesalahan pada fitur pada saat digunakan.



Gambar 11. WordCloud Negatif Netral Positif

IV. SIMPULAN

Hasil klasifikasi analisis sentiment pengguna aplikasi speedcash cukup beragam. Sesuai data yang ada ulasan Netral mendominasi ulasan yang ada pada google play store. Dengan 11.864 ulasan lebih pengguna aplikasi speedcash berkomentar Netral, dan terdapat 9.328 ulasan lebih pengguna yang berkomentar Positif, serta lebih dari 3.684 ulasan berkomentar negatif. Pengujian dilakukan terhadap delapan model, yaitu empat kernel Support Vector Machine (SVM) dan empat varian Naïve Bayes, pada tiga rasio pembagian data (60%:40%, 70%:30%, dan 80%:20%) dengan menerapkan Synthetic Minority Over-sampling Technique (SMOTE) pada data latih. Hasil penelitian menunjukkan bahwa metode SVM secara konsisten memberikan kinerja yang lebih baik dibandingkan Naïve Bayes pada rasio pengujian (60%:40%), Performa terbaik diperoleh SVM Linear dengan F1-Score sebesar 88,64% dan Accuracy 88,52% dan untuk naive bayes Bernoulli dengan F1-Score sebesar 73,14% dan Accuracy 73,08%, pada rasio pengujian (70%:30%), Performa terbaik diperoleh SVM RBF dengan F1-Score sebesar 88,99% dan Accuracy 89,07% dan untuk naive bayes Bernoulli dengan F1-Score sebesar 73,57% dan Accuracy 73,47%, dan untuk rasio pengujian (80%:20%), Performa terbaik diperoleh SVM RBF dengan F1-Score sebesar 88,85% dan Accuracy 88,93% dan untuk naive bayes Bernoulli dengan F1-Score sebesar 73,61% dan Accuracy 73,47%.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada semua pihak yang telah memberikan dukungan, bimbingan, dan bantuan selama proses penyelesaian penelitian ini, baik dalam aspek akademik maupun nonakademik. Kontribusi berupa arahan ilmiah, saran yang membangun, serta penyediaan fasilitas pendukung sangat membantu penulis dalam menyelesaikan penelitian hingga tahap akhir. Selain itu, dukungan moral dan motivasi yang diberikan selama pelaksanaan penelitian juga menjadi faktor penting dalam menjaga kelancaran serta kualitas penyusunan karya ini. Penulis berharap hasil penelitian ini dapat memberikan manfaat serta menjadi referensi bagi pengembangan penelitian pada masa yang akan datang.

REFERENSI

- [1] U. Nandiroh, M. Bastomi, R. A. Nutkhofifah, and M. Z. Abdillah, "Optimalisasi penggunaan dompet digital sebagai solusi efisiensi transaksi," *J. Inov. Has. Pengabd. Masy.*, vol. 7, no. 1, pp. 11–19, 2023, doi: 10.33474/jipemas.v7i1.20405.
- [2] Endar Nirmala and Andri Fahmi, "Comparison of LSTM and Naïve Bayes in Google Play Store App Review Sentiment Analysis," *J. Inotera*, vol. 11, no. 1, pp. 171–181, 2026, doi: 10.31572/inotera.vol11.iss1.2026.id653.
- [3] N. E. Salassa, A. Patanduk, A. Yusupa, and Y. D. Y. Rindengan, "Comparative Analysis of Naive Bayes and Support Vector Machine for Sentiment Classification of Indonesian-Language Mobile Application Reviews on Google Play Store," *J. Ilm. Inform. dan Komput.*, vol. 3, no. 1, pp. 20–25, 2026, doi: 10.69533/informatech.volume3number1.515.
- [4] R. Rahmat, A. Rahim, and A. Arbansyah, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Alibaba Di Google Play Store," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 3050–3057, 2025, doi: 10.36040/jati.v9i2.13269.
- [5] B. Liu and C. Cardie, "Book Reviews Sentiment Analysis and Opinion Mining," 2014, doi: 10.1162/COLI.
- [6] A. A. Romadhoni, A. Rachmadany, and B. H. Prasjo, "Sentiment Analysis of Indrive App Usage Reviews on Google Playstore Using Support Vector Machine (Svm) and Naïve Bayes Algorithm," *Int. J. Artif. Intell. Digit. Mark.*, vol. 2, no. 10, pp. 102–113, 2025, doi: 10.61796/ijaidf.v2i10.421.
- [7] Tom Michael Mitchell, "*Machine Learning*". 1997.
- [8] O. M. Nurfauzi, S. S. Hilabi, F. Nurapriani, and B. Huda, "Analisis Sentimen, Grab Indonesia Analisis Sentimen Grab Indonesia Pada Ulasan Google Play Store Menggunakan Algoritma Naïve Bayes Dan SVM," *SMARTICS J.*, vol. 11, no. 1, pp. 8–13, 2025, doi: 10.21067/smartics.v11i1.11789.
- [9] A. Suharman and M. Kamayani Sulaeman, "Analisis Sentimen Pengguna Aplikasi Livin' by Mandiri Menggunakan Metode Support Vector Machine (SVM) dengan Ekstraksi Fitur TF-IDF dan Word2Vec," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 8, pp. 2201–2212, 2025, doi: 10.52436/1.jpti.941.

- [10] A. N. Puspitasari, Y. Findawati, and Y. Rahmawati, "ANALISIS SENTIMEN TWEET PENGGUNA E-COMMERCE DENGAN MENGGUNAKAN METODE KLASIFIKASI NAIVE BAYES," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 3, pp. 1123–1132, Aug. 2024, doi: 10.29100/jipi.v9i3.4939.
- [11] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA J.*, vol. 10, no. 02, pp. 71–76, Dec. 2020, doi: 10.32664/smatika.v10i02.455.
- [12] Ratih Puspitasari, Y. Findawati, and M. A. Rosid, "SENTIMENT ANALYSIS OF POST-COVID-19 INFLATION BASED ON TWITTER USING THE K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE CLASSIFICATION METHODS," *J. Tek. Inform.*, vol. 4, no. 4, pp. 669–679, Aug. 2023, doi: 10.52436/1.jutif.2023.4.4.801.
- [13] C. Cardie, "Book Reviews: Sentiment Analysis and Opinion Mining by Bing Liu," *Vol. 40, Issue 2 - June 2014*, pp. 511–513, 2014, doi: 10.1162/COLI.
- [14] F. T. Admojo, S. Risnanto, A. W. Windiawati, M. Innuddin, and D. Mualfah, "Comparison of Naïve Bayes and Random Forest Algorithm in Webtoon Application Sentiment Analysis," *Innov. Res. Informatics*, vol. 6, no. 1, May 2024, doi: 10.37058/innovatics.v6i1.10636.
- [15] A. A. Aysha, M. P. Aji, E. S. Wijaya, and E. A. Pambudi, "Perbandingan Kinerja Algoritma Naïve Bayes, Decision Tree, dan Support Vector Machine dalam Deteksi Serangan Siber Berdasarkan Log Sistem di Universitas Muhammadiyah Purwokerto," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 12, pp. 3620–3629, Jan. 2026, doi: 10.52436/1.jpti.1196.
- [16] L. B. Ilmawan and M. A. Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 154–161, Aug. 2020, doi: 10.33096/ilkom.v12i2.597.154-161.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.