

Analisis Sentimen Komentar Pengguna Aplikasi Klik Indomaret Di Google Playstore Menggunakan Metode Support Vector Machine (SVM)

Dea Andini Andriati¹, Rani Laple Satria Putra², Galih Dika Saputra³
Prodi Teknik Informatika, Universitas Dian Nusantara, Jakarta Barat , Indonesia

ABSTRACT

Keywords:

Analisis Sentimen,
Aplikasi Klik Indomaret,
Komentar Pengguna,
Google Playstore.
SVM.

Perkembangan teknologi digital telah mendorong pergeseran perilaku konsumen, khususnya dalam penggunaan aplikasi belanja daring. Salah satu aplikasi yang banyak digunakan adalah Klik Indomaret, yang memungkinkan pengguna berbelanja produk ritel secara praktis melalui perangkat mobile. Seiring meningkatnya jumlah pengguna, banyak ulasan yang terkumpul di Google Play Store yang dapat dimanfaatkan untuk mengevaluasi kualitas layanan. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap komentar pengguna aplikasi Klik Indomaret menggunakan algoritma Support Vector Machine (SVM). Sebanyak 4.841 komentar pengguna dikumpulkan secara otomatis melalui teknik web crawling berbasis Python. Data kemudian melalui tahapan prapemrosesan yang meliputi case folding, pembersihan teks, tokenisasi, penghapusan stopword, normalisasi, dan stemming untuk menghasilkan data yang terstruktur. Metode Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk pembobotan kata sehingga dapat mengidentifikasi kata yang paling berpengaruh dalam dokumen. Dataset yang sudah siap kemudian diuji menggunakan algoritma SVM untuk mengklasifikasikan sentimen ke dalam kategori positif dan negatif. Hasil evaluasi menunjukkan bahwa model SVM memiliki akurasi keseluruhan sebesar 87% dari 3.874 data uji. Kelas negatif memberikan performa tinggi dengan precision 0,82, recall 0,94, dan f1-score 0,87, sedangkan kelas positif mencatat precision 0,92, recall 0,85, dan f1-score 0,89. Namun, kelas netral belum terdeteksi dengan baik karena jumlah data yang sedikit. Uji 10-fold cross-validation memperlihatkan akurasi yang konsisten pada rentang 85,4% hingga 87,6%, menandakan kemampuan generalisasi model yang baik tanpa indikasi overfitting.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license



Corresponding Author:

Dea Andini Andriati
Universitas Dian Nusantara
Jakarta Barat
dea.andini.andriati@dosen.undira.ac.id

INTRODUCTION

Dalam satu dekade terakhir, perkembangan transformasi digital telah memberikan dampak signifikan terhadap perubahan perilaku konsumen, terutama dalam cara mereka berinteraksi dengan layanan ritel (Giovani et al., 2020). Salah satu bentuk nyata dari perubahan ini adalah pemanfaatan aplikasi mobile oleh perusahaan ritel sebagai sarana distribusi dan komunikasi yang lebih efektif dan efisien (Sipayung et al., 2016). Di Indonesia, fenomena ini tercermin melalui meningkatnya jumlah aplikasi belanja daring, seperti aplikasi (Dewi, 2022) Klik Indomaret yang dikembangkan oleh jaringan ritel Indomaret sebagai alternatif belanja yang lebih praktis (Darwis et al., 2021).

Aplikasi Klik Indomaret menyediakan berbagai fitur yang mencakup pemesanan produk, informasi promosi, serta layanan pengantaran, yang semuanya dirancang untuk meningkatkan kenyamanan pengguna dalam berbelanja tanpa harus mengunjungi toko fisik (Petiwi et al., 2022).

Dengan bertambahnya jumlah pengguna aplikasi, turut meningkat pula interaksi pengguna yang diwujudkan dalam bentuk ulasan dan komentar pada platform Google Playstore. Respons-respons ini merefleksikan pengalaman serta persepsi pengguna selama menggunakan aplikasi tersebut (Tanggraeni & Sitokdana, 2022).

Ulasan yang diberikan oleh pengguna merupakan salah satu sumber informasi yang bernilai bagi pengembang aplikasi dalam melakukan evaluasi terhadap performa dan kualitas layanan (Ghani Ardiansyah et al., 2025). Komentar yang bernada positif dapat diinterpretasikan sebagai indikator keberhasilan fitur aplikasi, sedangkan komentar negatif mengindikasikan aspek yang masih memerlukan perbaikan (Buntoro, 2016). Namun demikian, banyaknya komentar yang tersedia menimbulkan tantangan tersendiri dalam proses pengolahan dan interpretasi data secara manual, sehingga diperlukan pendekatan analisis yang sistematis dan efisien (Herlinawati et al., 2020).

Analisis sentimen merupakan salah satu pendekatan yang sering digunakan dalam mengolah data teks yang mengandung opini (Melisa Nur Aini et al., 2024). Tujuan utama dari teknik ini adalah untuk menentukan polaritas dari suatu pernyataan, apakah mengandung sentimen positif, negatif, atau netral. Metode ini banyak diterapkan dalam bidang pengolahan bahasa alami karena mampu memberikan gambaran umum mengenai kecenderungan opini publik terhadap suatu produk, layanan, maupun isu tertentu.

Dalam implementasi analisis sentimen, algoritma Support Vector Machine (SVM) dikenal sebagai salah satu metode pembelajaran mesin yang efektif dalam tugas klasifikasi data teks (Zaidah, 2021). SVM beroperasi dengan mencari hyperplane terbaik yang dapat memisahkan data ke dalam beberapa kategori secara optimal (Rahman et al., 2024). Kemampuannya dalam menangani data berdimensi tinggi menjadikannya sangat cocok digunakan dalam konteks klasifikasi teks, khususnya untuk menganalisis komentar pengguna aplikasi (Imawan & Mude, 2020).

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap komentar pengguna aplikasi Klik Indomaret yang diperoleh dari platform Google Playstore dengan menggunakan algoritma SVM. Sebelum tahap klasifikasi dilakukan, teks perlu melalui proses pra-pemrosesan agar dapat diproses oleh algoritma secara optimal. Tahapan ini mencakup proses case folding, pembersihan data (cleaning), tokenisasi, penghapusan stopword, hingga normalisasi. Proses ini bertujuan untuk mengurangi gangguan data (noise) dan mempertahankan informasi relevan yang berkontribusi terhadap akurasi klasifikasi sentimen.

METHODS

Analisis dan Perancangan Sistem

Penelitian ini bertujuan untuk mengembangkan sistem analisis sentimen yang mengimplementasikan algoritma Support Vector Machine (SVM), dengan fokus pada ulasan berbahasa Indonesia dari aplikasi yang tersedia di Google Play Store (Sentimen et al., 2025). Sistem ini dirancang untuk mengevaluasi efektivitas SVM dalam mengklasifikasikan sentimen pengguna terhadap aplikasi. Proses dimulai dengan pemanfaatan data pelatihan yang diproses guna membangun model klasifikasi. Model ini kemudian digunakan untuk mengelompokkan komentar pengguna berdasarkan sentimen. Sebelum proses klasifikasi dilakukan, data ulasan melalui tahap prapemrosesan yang mencakup normalisasi huruf (case folding), penghapusan kata-kata tidak bermakna (stopwords), penghilangan tanda baca, serta pemberian tag khusus pada kalimat yang menunjukkan sentimen negatif (Syamia & Furqan, 2025).

Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini dikumpulkan dari ulasan pengguna terhadap aplikasi Duolingo melalui platform Google Play Store. Teknik web scraping digunakan untuk mengotomatisasi proses ekstraksi data. Selanjutnya, data yang diperoleh diproses menggunakan Google Colaboratory dengan bantuan pustaka TensorFlow. Dataset yang telah dikumpulkan disimpan dalam format

spreadsheet pada media penyimpanan lokal atau dalam basis data, sebelum kemudian diproses lebih lanjut dalam tahap prapemrosesan.

Tahap Prapemrosesan Data

Prapemrosesan merupakan tahapan krusial dalam proses analisis data, bertujuan untuk memastikan bahwa data yang akan dianalisis berada dalam kondisi yang bersih, konsisten, dan layak untuk digunakan dalam pemodelan machine learning. Langkah-langkah prapemrosesan yang diterapkan dalam penelitian ini meliputi:

- a) **Penggunaan Ekspresi Reguler**
Digunakan untuk membersihkan teks dari pola yang tidak relevan seperti karakter khusus dan tanda baca dengan menggunakan teknik ekspresi reguler (regular expressions).
- b) **Penghapusan Tautan (URL)**
Seluruh URL yang terdapat dalam teks dihapus karena tidak memberikan kontribusi berarti dalam analisis sentimen.
- c) **Penghilangan Simbol Media Sosial**
Simbol seperti "@" yang umumnya digunakan untuk menandai akun media sosial dihapus karena tidak mengandung informasi sentimen yang signifikan.
- d) **Eliminasi Karakter Angka**
Seluruh angka yang tidak berkaitan langsung dengan konteks sentimen dihapus dari teks.
- e) **Tokenisasi**
Teks diuraikan menjadi unit linguistik terkecil berupa token (kata atau frasa) untuk memudahkan proses analisis selanjutnya.
- f) **Stemming**
Proses ini mengubah kata-kata menjadi bentuk dasarnya dengan menghapus imbuhan, baik di awal maupun di akhir kata. Hal ini bertujuan untuk menyederhanakan variasi morfologis kata.
- g) **Normalisasi Huruf**
Seluruh huruf dalam teks diubah menjadi huruf kecil guna memastikan konsistensi dalam pemrosesan dan mencegah perbedaan makna yang disebabkan oleh penggunaan huruf kapital.
- h) **Penghapusan Stopwords**
Kata-kata umum yang tidak memiliki nilai informatif dalam analisis, seperti "yang", "dan", atau "di", dihapus berdasarkan daftar stopwords.

Pelabelan Sentimen

Hanya data yang mengandung sentimen positif atau negatif yang digunakan dalam analisis. Data dengan sentimen netral dikecualikan, berdasarkan nilai polaritas yang telah dihitung sebelumnya.

TF-IDF (Term Frequency–Inverse Document Frequency)

Langkah berikutnya adalah melakukan pembobotan kata untuk mengubah teks ke dalam bentuk numerik sehingga lebih mudah diproses oleh algoritma machine learning. Pada penelitian ini digunakan TF-IDF Vectorizer dari pustaka scikit-learn, yang bekerja dengan memberi bobot pada setiap kata berdasarkan seberapa sering kata tersebut muncul dalam suatu dokumen serta relevansinya dibandingkan dengan keseluruhan kumpulan data. Dengan metode ini, kata yang jarang muncul tetapi memiliki arti penting akan mendapatkan bobot yang lebih besar.

Pemodelan Data

Setelah tahap prapemrosesan selesai, langkah selanjutnya adalah membangun model klasifikasi menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM). Pemodelan dilakukan

melalui teknik k-fold cross-validation dengan jumlah lipatan sebanyak 10 (10-fold). Dalam setiap iterasi, satu subset data digunakan sebagai data uji, sementara sembilan subset lainnya berperan sebagai data pelatihan. Tujuan dari pemodelan ini adalah untuk mengidentifikasi model klasifikasi yang paling optimal dalam membedakan ulasan pengguna ke dalam dua kategori sentimen utama: positif dan negatif. Proses ini dilaksanakan dalam lingkungan cloud-based Google Colaboratory dengan menggunakan pustaka TensorFlow dalam bahasa pemrograman Python.

Evaluasi Kinerja Model

Tahap akhir dari proses analisis ini melibatkan evaluasi performa model yang telah dikembangkan. Evaluasi dilakukan untuk mengukur akurasi dan kemampuan generalisasi model dalam mengklasifikasikan data secara tepat. Dua metrik evaluasi utama digunakan, yaitu accuracy. Penilaian dilakukan menggunakan pendekatan 10-fold cross-validation untuk memperoleh hasil yang representatif dan obyektif. Evaluasi ini berfokus pada kemampuan model dalam membedakan secara efektif antara dua kelas sentimen dan dalam memberikan prediksi yang akurat terhadap data baru.

RESULTS AND DISCUSSION

Pengumpulan dan Pengolahan Data Awal

Penelitian ini memanfaatkan kumpulan data berupa 4.841 komentar pengguna sebagai data mentah, yang diperoleh dari ulasan aplikasi Klik Indomaret di Google Play Store. Data dikumpulkan secara otomatis menggunakan teknik web crawling yang diimplementasikan melalui bahasa pemrograman Python. Setelah proses akuisisi data selesai, dilakukan tahapan prapemrosesan untuk mengubah data mentah menjadi format yang lebih terstruktur dan layak digunakan dalam analisis lanjutan. Data yang telah diproses kemudian dianotasi berdasarkan polaritas sentimen, yaitu kategori positif dan negatif. Selain itu, metode Term Frequency-Inverse Document Frequency (TF-IDF) diterapkan untuk melakukan pembobotan terhadap kata-kata dalam dokumen. Dataset yang telah melalui proses ini selanjutnya digunakan dalam pengujian algoritma Support Vector Machine (SVM), dengan tujuan mengevaluasi tingkat akurasi serta menentukan kernel yang paling optimal dalam klasifikasi sentimen.

Deskripsi Data Awal

Data yang digunakan dalam penelitian ini berasal dari ulasan teks pengguna aplikasi Klik Indomaret di platform Google Play Store. Data diperoleh secara otomatis melalui skrip berbasis Python. Untuk menjamin kualitas dan relevansi data, dilakukan proses penyaringan awal yang mencakup penghapusan entri duplikat dan komentar yang tidak memiliki keterkaitan langsung dengan topik atau konteks aplikasi Klik Indomaret. Sampel Komentar Pengguna Aplikasi Klik Indomaret dapat dilihat pada Table 1.

Table 1. Komentar Pengguna Aplikasi Klik Indomaret di Google Playstore

Username	Komentar
1	sangat membantu kalau lagi pergi kemana mana malas keluar rumah
2	dari bulan juni ini aplikasi eror mulu pake web bulan juni bisa sekarang malah ikutan eror web pun.. coba aplikasi nya lebih di urus lagi..
3	dengan adanya aplikasi ini bisa menghemat waktu untuk berbelanja
4	Alhamdulillah pelayanan baik...
5	cepat dan memuaskan

Tahapan Prapemrosesan Teks

Prapemrosesan merupakan komponen esensial dalam analisis data teks, yang bertujuan untuk mentransformasi data tidak terstruktur menjadi bentuk yang dapat diolah oleh algoritma pembelajaran mesin. Karena data yang diperoleh bersifat tidak berlabel (unsupervised), dibutuhkan serangkaian proses transformasi agar data tersebut dapat digunakan dalam proses klasifikasi sentimen.

Case Folding

Pada tahap ini, seluruh karakter dalam dokumen teks dikonversi menjadi huruf kecil (lowercase) untuk menjaga konsistensi penulisan serta meminimalkan perbedaan representasi kata akibat variasi kapitalisasi. Hasil sampel case folding dapat dilihat pada Table 2.

Table 2. Hasil Case Folding

Sebelum	Sesudah
sangat membantu kalau lagi pergi kemana mana malas keluar rumah	sangat membantu kalau lagi pergi kemana mana malas keluar rumah
dari bulan juni ini aplikasi eror mulu pake web bulan juni bisa sekarang malah ikutan eror web pun.. coba aplikasi nya lebih di urus lagi..	dari bulan juni ini aplikasi eror mulu pake web bulan juni bisa sekarang malah ikutan eror web pun.. coba aplikasi nya lebih di urus lagi..
dengan adanya aplikasi ini bisa menghemat waktu untuk berbelanja	dengan adanya aplikasi ini bisa menghemat waktu untuk berbelanja
Alhamdulillah pelayanan baik...	alhamdulillah pelayanan baik...
cepat dan memuaskan	cepat dan memuaskan

Pembersihan Teks (Cleaning)

Dilakukan penghapusan elemen-elemen yang tidak mendukung analisis seperti tanda baca, simbol, karakter non-alfabet, serta kata-kata yang tidak relevan dengan konteks analisis. Hasil Sampel Cleaning Text dapat dilihat pada Table 3.

Table 3. Hasil Pembersihan Text (Cleaning)

Sebelum	Sesudah
sangat membantu kalau lagi pergi kemana mana malas keluar rumah	sangat membantu kalau lagi pergi kemana mana malas keluar rumah
dari bulan juni ini aplikasi eror mulu pake web bulan juni bisa sekarang malah ikutan eror web pun.. coba aplikasi nya lebih di urus lagi..	dari bulan juni ini aplikasi eror mulu pake web bulan juni bisa sekarang malah ikutan eror web pun coba aplikasi nya lebih di urus lagi
dengan adanya aplikasi ini bisa menghemat waktu untuk berbelanja	dengan adanya aplikasi ini bisa menghemat waktu untuk berbelanja

alhamdulillah pelayanan baik...	alhamdulillah pelayanan baik
cepat dan memuaskan	cepat dan memuaskan

Tokenisasi

Teks dibagi menjadi unit-unit kata yang lebih kecil (token), dengan tujuan mempermudah proses ekstraksi fitur dan klasifikasi pada tahap selanjutnya. Hasil sampel tokenizing dapat dilihat pada Table 4.

Table 4. Hasil Tokenisasi

Sebelum	Sesudah
sangat membantu kalau lagi pergi kemana mana malas keluar rumah	'sangat', 'membantu', 'kalau', 'lagi', 'pergi', 'kemana', 'mana', 'malas', 'keluar', 'rumah'
dari bulan juni ini aplikasi eror mulu pake web bulan juni bisa sekarang malah ikutan eror web pun coba aplikasi nya lebih di urus lagi	'dari', 'bulan', 'juni', 'ini', 'aplikasi', 'eror', 'mulu', 'pake', 'web', 'bulan', 'juni', 'bisa', 'sekarang', 'malah', 'ikutan', 'eror', 'web', 'pun', 'coba', 'aplikasi', 'nya', 'lebih', 'di', 'urus', 'lagi'
dengan adanya aplikasi ini bisa menghemat waktu untuk berbelanja	'dengan', 'adanya', 'aplikasi', 'ini', 'bisa', 'menghemat', 'waktu', 'untuk', 'berbelanja'
alhamdulillah pelayanan baik	'alhamdulillah', 'pelayanan', 'baik'
cepat dan memuaskan	'cepat', 'dan', 'memuaskan'

Penghapusan Stopword

Kata-kata umum yang memiliki frekuensi tinggi namun rendah nilai semantisnya, seperti kata sambung dan kata bantu, dihapus dari korpus karena tidak memberikan kontribusi signifikan terhadap interpretasi sentimen. Hasil sampel stopwords removal dapat dilihat pada Table 5.

Table 5. Hasil Stopword

Sebelum	Sesudah
sangat membantu kalau lagi pergi kemana mana malas keluar rumah	membantu pergi kemana malas keluar rumah
dari bulan juni ini aplikasi eror mulu pake web bulan juni bisa sekarang malah ikutan eror web pun coba aplikasi nya lebih di urus lagi	bulan juni aplikasi eror mulu pake web bulan juni bisa sekarang ikutan eror web coba aplikasi urus
dengan adanya aplikasi ini bisa menghemat waktu untuk berbelanja	aplikasi bisa menghemat waktu berbelanja
alhamdulillah pelayanan baik	alhamdulillah pelayanan baik
cepat dan memuaskan	cepat memuaskan

Normalisasi

Proses normalisasi dilakukan untuk menyamakan kata-kata yang tidak baku atau yang salah eja dengan bentuk yang sesuai dengan standar Bahasa Indonesia. Hasil sampel normalisasi dapat dilihat pada Table 6.

Table 6. Hasil Normalisasi

Sebelum	Sesudah
membantu pergi kemana malas keluar rumah	membantu pergi kemana malas keluar rumah
bulan juni aplikasi eror mulu pake web bulan juni bisa sekarang ikutan eror web coba aplikasi urus	bulan juni aplikasi error terus pakai web bulan juni bisa sekarang ikutan error web coba aplikasi urus
aplikasi bisa menghemat waktu berbelanja	aplikasi bisa menghemat waktu berbelanja
alhamdulillah pelayanan baik	alhamdulillah pelayanan baik
cepat memuaskan	cepat memuaskan

Stemming

Pada tahap ini, setiap kata dikembalikan ke bentuk dasarnya dengan menghilangkan imbuhan awalan, sisipan, maupun akhiran. Proses ini penting untuk menyederhanakan variasi morfologis kata, sehingga meningkatkan efektivitas representasi teks dalam proses klasifikasi. Hasil sampel stemming dapat dilihat pada Table 7.

Table 7. Hasil Stemming

Sebelum	Sesudah
membantu pergi kemana malas keluar rumah	bantu pergi kemana malas keluar rumah
bulan juni aplikasi error terus pakai web bulan juni bisa sekarang ikutan error web coba aplikasi urus	bulan juni aplikasi error terus pakai web bulan juni bisa sekarang ikut error web coba aplikasi urus
aplikasi bisa menghemat waktu berbelanja	aplikasi bisa hemat waktu belanja
alhamdulillah pelayanan baik	alhamdulillah layan baik
cepat memuaskan	cepat puas

Hasil Pelabelan Sentimen

Setelah tahapan pra pemrosesan selesai, langkah selanjutnya adalah memberi label pada pola data. Pada studi ini, pelabelan dilakukan dengan analisis sentimen terhadap seluruh komentar dalam dataset, yang diklasifikasikan ke tiga kategori: positif, netral, dan negatif. Sampel hasil pelabelan ditampilkan pada Figure 1.

	content	sentimen
0	bantu pergi kemana malas keluar rumah	Netral
1	bulan juni aplikasi error terus pakai web bula...	Negatif
2	aplikasi bisa hemat waktu belanja	Netral
3	alhamdulillah layan baik	Positif
4	cepat puas	Positif

Figure 1. Hasil Sampel Pelabelan Data

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

Hasil perhitungan TF-IDF (Term Frequency–Inverse Document Frequency) pada tabel memberikan gambaran mengenai bobot kepentingan setiap kata dalam kumpulan dokumen. Kata “klik” muncul paling sering dengan frekuensi 45.497 dan nilai tf-idf sebesar 0.058134 dan nilai idf 1.135749. Kata “indomaret” memiliki frekuensi 43.461 dengan nilai tf-idf 0.057771. Kata “terima” memiliki frekuensi 28.447 dengan nilai idf 1.605353 dan tf-idf 0.051377. Kata “kasih” memiliki pola yang mirip dengan “terima”, dengan frekuensi 28.354, idf 1.608627, dan tf-idf 0.051314. Terakhir, kata “hi” muncul 25.751 kali dengan nilai idf tertinggi (1.704922) dan tf-idf 0.049393. Hasil perhitungan TF-IDF dapat dilihat pada Figure 2.

	term	frequency	idf	tf	tfidf
0	klik	45497	1.135749	0.051186	0.058134
1	indomaret	43461	1.181531	0.048895	0.057771
2	terima	28447	1.605353	0.032004	0.051377
3	kasih	28354	1.608627	0.031899	0.051314
4	hi	25751	1.704922	0.028971	0.049393

Figure 2. Proses Perhitungan TF-IDF

Hasil Permodelan Data

Hasil evaluasi model Support Vector Machine (SVM) menunjukkan bahwa kinerja model secara keseluruhan cukup baik dengan akurasi sebesar 87% dari total 3.874 data uji. Pada kategori negative, model menghasilkan nilai precision 0,82, recall 0,94, dan f1-score 0,87. Hal ini mengindikasikan bahwa model mampu mengenali ulasan dengan sentimen negatif dengan tingkat keberhasilan yang tinggi, meskipun terdapat beberapa kesalahan klasifikasi. Untuk kategori positive, performa model juga menunjukkan hasil yang baik dengan precision 0,92, recall 0,85, dan f1-score 0,89. Nilai precision yang tinggi menunjukkan sebagian besar prediksi positif benar, namun recall yang sedikit lebih rendah mengindikasikan adanya sebagian data positif yang tidak terdeteksi dengan baik.

Hasil permodelan data dapat dilihat pada Figure 3.

SVM - Classification Report:				
	precision	recall	f1-score	support
negative	0.82	0.94	0.87	1770
neutral	0.00	0.00	0.00	124
positive	0.92	0.85	0.89	1980
accuracy			0.87	3874
macro avg	0.58	0.60	0.59	3874
weighted avg	0.84	0.87	0.85	3874

Figure 3. Hasil Pengujian SVM

Di sisi lain, kategori neutral memiliki nilai precision, recall, dan f1-score 0,00. Hal ini disebabkan oleh jumlah data netral yang relatif kecil (124 sampel) sehingga model kesulitan mengenali kelas ini secara akurat. Nilai macro average untuk precision, recall, dan f1-score masing-masing adalah 0,58, 0,60, dan 0,59. Angka ini menggambarkan bahwa kinerja antar kelas tidak seimbang karena buruknya performa pada kelas netral. Namun, nilai weighted average yang mencapai 0,84 (precision), 0,87 (recall), dan 0,85 (f1-score) tetap menunjukkan performa yang baik karena mempertimbangkan dominasi data negatif dan positif yang jumlahnya lebih besar. Confusion Matrix SVM dapat dilihat pada Figure 4.

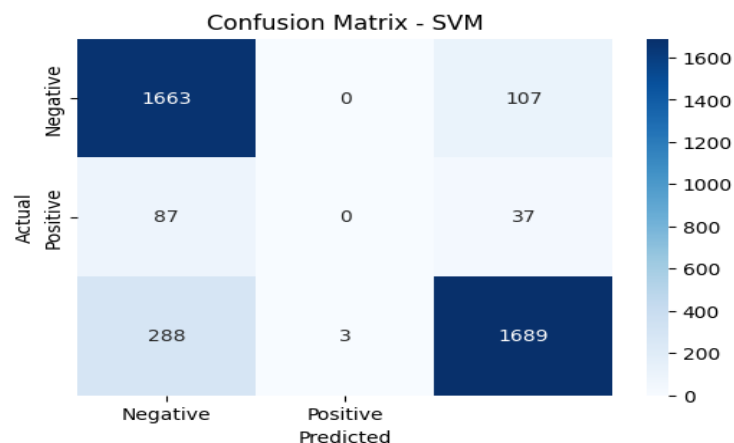


Figure 4. Confusion Matrix SVM

Hasil Evaluasi Kinerja Model

Berdasarkan hasil grafik pada Figure 5 Grafik menunjukkan hasil evaluasi akurasi model Support Vector Machine (SVM) menggunakan metode 10-fold cross-validation. Setiap titik pada grafik mewakili nilai akurasi pada masing-masing fold. Secara umum, akurasi model berada pada rentang 85,4% hingga 87,6%, yang menunjukkan performa yang cukup stabil di berbagai subset data.

Nilai akurasi tertinggi terlihat pada fold ke-5 dengan capaian sekitar 87,6%, sedangkan akurasi terendah terjadi pada fold ke-6 dengan nilai sekitar 85,4%. Perbedaan akurasi antar fold relatif kecil, sehingga dapat disimpulkan bahwa model memiliki kemampuan generalisasi yang baik dan tidak mengalami overfitting yang signifikan. Sebagian besar fold juga memiliki akurasi di atas 86%, yang menandakan konsistensi kinerja model.

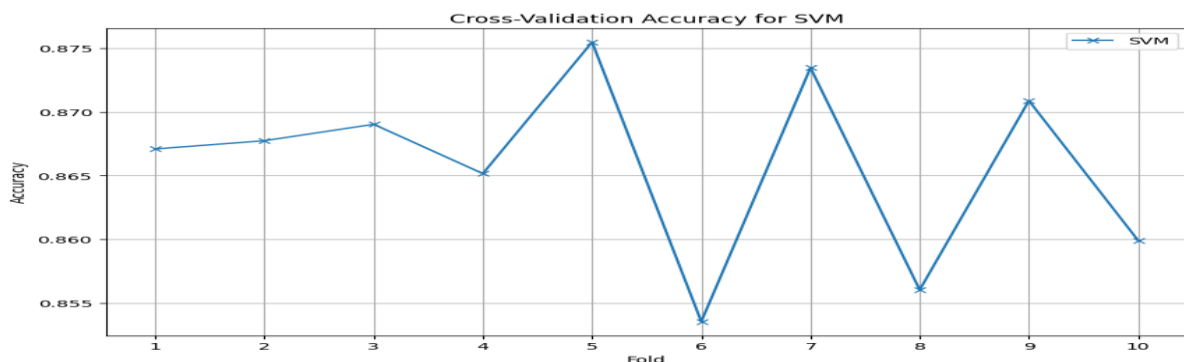


Figure 5. Hasil Akurasi Cross-Validation Model SVM Pada 10 fold.

CONCLUSION

Hasil evaluasi menunjukkan bahwa model SVM memiliki performa yang cukup baik dengan akurasi 87%. Kelas negatif menunjukkan precision 0,82, recall 0,94, dan f1-score 0,87, sementara kelas positif mencatat precision 0,92, recall 0,85, dan f1-score 0,89. Namun, model belum mampu mengklasifikasikan kelas netral dengan baik (precision, recall, dan f1-score = 0,00) karena jumlah data netral yang relatif sedikit. Hasil 10-fold cross-validation memperlihatkan bahwa akurasi model cukup konsisten, berkisar antara 85,4% hingga 87,6%, yang menunjukkan kemampuan generalisasi yang baik tanpa indikasi overfitting yang signifikan.

Penelitian di masa mendatang juga dapat mempertimbangkan penggunaan algoritma lain seperti ensemble learning (Random Forest, XGBoost) atau model berbasis deep learning (misalnya LSTM atau BERT) untuk membandingkan kinerja dengan SVM. Penambahan jumlah data netral yang lebih banyak dan beragam akan membantu model memahami karakteristik kelas ini sehingga hasil klasifikasi dapat lebih seimbang di seluruh kategori.

REFERENCE

- Buntoro, G. A. (2016). Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine. *Jurnal Dinamika Informatika*, 5(2), 1–13.
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>
- Dewi, A. K. (2022). Analisis Sentimen Ekspedisi Sicepat Dari Ulasan Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 796–805. <https://doi.org/10.35957/jatisi.v9i2.1802>
- Ghani Ardiansyah, A., Hidayati, N., Sari Sinaga, N., Criston Purba, D., Yeng Sinaga, E., Tunas Bangsa Pematangsiantar, S., Ghani Ardiansyah STIKOM Tunas Bangsa Pematang Siantar, A., & Sumatra, N. (2025). Design of Building Shop Information System Abghan22 Web-Based Using The Naïve Bayes. *Jatilima: Journal of Multimedia and Information Technology*, 07. <https://doi.org/10.54209/jatilima.v7i01.978>
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., & Samudi, S. (2020). Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 293. <https://doi.org/10.24114/cess.v5i2.18186>
- Ilmawan, L. B., & Mude, M. A. (2020). Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. *ILKOM Jurnal Ilmiah*, 12(2), 154–161. <https://doi.org/10.33096/ilkom.v12i2.597.154-161>
- Melisa Nur Aini, Rita Yulfani, & Nurul Jariah. (2024). Penerapan Metode Naïve Bayes Untuk Analisis Sentimen Ulasan Produk Sunscreen Berdasarkan Female Daily Review. *Jurnal Multimedia Dan Teknologi Informasi (Jatilima)*, 6(01), 24–34. <https://doi.org/10.54209/jatilima.v6i01.421>
- Petiwi, M. I., Triayudi, A., & Sholihati, I. D. (2022). Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine. *Jurnal Media Informatika Budidarma*, 6(1), 542. <https://doi.org/10.30865/mib.v6i1.3530>
- Rahman, Z., Sakinah, P., Hendra, Y., Satria, B., Maulana, F., & Ayun, A. Q. (2024). *Sentiment Analysis of Gojek App Reviews on Google Play Store with Natural Language Processing Using Naive Bayes' Algorithm-Zumardi Rahman et.al Sentiment Analysis of Gojek App Reviews on Google Play Store with Natural Language Processing Using Naive Bayes' Algorithm*. 06. <https://doi.org/10.54209/jatilima.v6i03.1189>

- Sentimen, A., Pelanggan, U., Hokben, A., Google, D., Menggunakan, P., Support, M., Vikri, R., Al, A. E., Alif, R. V., & Hasugian, A. H. (2025). *Analisis Sentimen Ulasan Pelanggan Aplikasi Hokben Di Google Playstore Menggunakan Metode Support Vector Machine*. 07(02). <https://doi.org/10.54209/jatilima.v7i02.1501>
- Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Informasi (JSI) UNSRI*, 8(1), 958–965.
- Syamia, N., & Furqan, M. (2025). *Sentiment Analysis On The Change Of Coach Of The Indonesian National Team Using Support Vector Machine Algorithm : A Case Study Of Twitter Data-Nanda Syamia et.al Sentiment Analysis On The Change Of Coach Of The Indonesian National Team Using Support Vector Machine Algorithm : A Case Study Of Twitter Data*. 07(02). <https://doi.org/10.54209/jatilima.v7i02.1510>
- Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 785–795. <https://doi.org/10.35957/jatisi.v9i2.1835>
- Zaidah, A. R. (2021). *Analisis Sentimen Berbasis Aspek Pada Aplikasi Peduli Lindungi Menggunakan Latent Dirichlet Allocation dan Support Vector Machine*. 1–63. <http://digilib.uinsby.ac.id/id/eprint/52007>