JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 5, No. 1, Januari 2026, pp. 563 - 573
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/169

Development of a Sentiment Analysis Application on XYZ Social Media for the Semilir Japan Matsuri 3 Event Announcement using SVM Method

Rheyza Avta Naufaldy*[1], Migunani[2], Maya Utami Dewi[3], Cherlina Helena Purnamasari Panjaitan[4]
[1,2,3,4] Fakultas Studi Akademik, Universitas Sains dan Teknologi Komputer
Jl. Majapahit No. 605, 50192, Kota Semarang, Jawa Tengah, Indonesia
e-mail: rheyzaavtanaufaldy@gmail.com*[1], migunani@stekom.ac.id[2], maya@stekom.ac.id[3], cherlina@stekom.ac.id[4]

ARTICLE INFO

Article History:

Received: 11-11-2025
Revised: 21-12-2025
Accepted: 14-01-2026
Published: 21-01-2026

ABSTRACT

The exponential growth of social media platforms has led to an overwhelming volume of user-generated textual data, making manual sentiment identification increasingly impractical and time-consuming. This condition highlights the urgency of developing automated sentiment analysis systems that are accurate, efficient, and easy to use. The objective of this study is to design and implement an automatic sentiment analysis application capable of classifying textual opinions in real time. The research adopts a Research and Development (R&D) approach, and the system is modeled using Object-Oriented Programming (OOP) principles and the Unified Modeling Language (UML) to ensure a structured, scalable application design. Sentiment classification is performed using the Support Vector Machine (SVM) algorithm, which is well known for its effectiveness in text classification. The application is developed in Python and deployed using the Streamlit framework to provide an interactive and user-friendly interface. Text data are processed through standard preprocessing stages before being classified into sentiment categories. The results demonstrate that the proposed system can automatically analyze input text and deliver sentiment predictions in real time with reliable performance. In conclusion, integrating SVM with Streamlit offers a practical solution for automated sentiment analysis, particularly for applications that require rapid interpretation of social media text data.

Keywords: *Sentiment Analysis, Machine Learning, Support Vector Machine, Streamlit.*

Abstrak

Perkembangan media sosial yang pesat telah menghasilkan volume data teks yang sangat besar dalam bentuk opini dan ulasan pengguna. Analisis manual terhadap data ini untuk memahami sentimen publik tidak lagi efisien. Hal inilah yang mendasari dibuatnya sebuah aplikasi analisis sentimen otomatis menggunakan metode machine learning *Support Vector Machine* (SVM). Untuk mendukung penelitian yang dilakukan, peneliti menggunakan metode penelitian *Research and Development* (R&D) serta perancangannya menggunakan metode *Object Oriented Programming* (OOP) menggunakan *Unified Modeling Language* (UML). Aplikasi ini dibangun menggunakan bahasa pemrograman Python dengan framework Streamlit untuk menyediakan antarmuka pengguna yang interaktif dan mudah diakses. Aplikasi yang dihasilkan

DOI: https://doi.org/10.51903/48shas61

p-ISSN : 2803-1507 e-ISSN : 2803-1531

berhasil mengotomatiskan seluruh proses, memungkinkan pengguna untuk memasukkan teks baru dan mendapatkan hasil analisis sentimen secara real-time. Penelitian ini menunjukkan bahwa kombinasi SVM dan Streamlit dapat menjadi solusi yang efektif dan praktis untuk analisis sentimen otomatis.

Kata Kunci: *Analisis Sentimen, Pembelajaran Mesin, Support Vector Machine, Streamlit*.

1. PENDAHULUAN

Era Industri 4.0 mengubah interaksi sosial secara fundamental dengan menempatkan media sosial sebagai medium utama [1]. Di Indonesia, fenomena ini sangat signifikan; data awal tahun 2024 menunjukkan bahwa 94,8% dari 185 juta pengguna *internet* aktif berinteraksi di media sosial. Salah satu *platform* yang mendominasi adalah media sosial XYZ, yang menempati posisi pertama sebagai aplikasi dengan rata-rata waktu penggunaan tertinggi, mencapai 38 jam per bulan per pengguna di Indonesia. Tingginya durasi penggunaan ini memicu produksi data opini publik yang masif dan cepat (*high velocity*) [2]. Hal ini menjadikannya sumber data potensial untuk analisis.

Fenomena produksi data yang cepat ini teramati pada konten pengumuman *event* Semilir Japan Matsuri 3 yang diunggah oleh akun media sosial XYZ resmi Dusun Semilir. Konten tersebut berhasil menarik perhatian publik secara luas, dibuktikan dengan pencapaian 3,1 juta kali tontonan dan lebih dari 1000 komentar. Ketersediaan data komentar yang melimpah ini merefleksikan opini dan sentimen publik secara langsung. Saat ini, proses analisis umpan balik (*feedback*) masih dilakukan secara manual, yang dinilai tidak efisien dan menghambat pengambilan keputusan strategis. Metode *manual* ini tentu tidak efisien, menghabiskan banyak waktu dan tenaga [3], serta menghambat pengambilan keputusan strategis yang cepat.

Untuk mengatasi masalah ketidakefisienan tersebut, diperlukan otomatisasi analisis sentimen menggunakan algoritma *Machine Learning* (ML). Analisis sentimen adalah proses mengolah data tekstual secara otomatis untuk mengekstrak informasi subjektif atau sentimen [4]. Salah satu algoritma ML yang terbukti sangat efektif untuk tugas klasifikasi teks adalah *Support Vector Machine* (SVM), yang secara konsisten menunjukkan performa superior [5]. Beberapa penelitian terdahulu telah menerapkan SVM untuk analisis sentimen di berbagai *platform*. Penelitian oleh Yolanda dan Mulya [6] menerapkan SVM pada ulasan aplikasi Sayurbox di Google Play Store dengan akurasi 89,29%. Penelitian lain oleh Firdaus *et al.* [7] dan Pamungkas *et al.* [8] juga berfokus pada *platform* seperti *Google Review* dan *Twitter*. Meskipun demikian ditemukan celah penelitian di mana mayoritas penelitian terdahulu berfokus pada platform dengan struktur bahasa yang relatif baku.

Berdasarkan latar belakang, fenomena, dan celah penelitian yang telah diidentifikasi, tujuan dari penelitian ini adalah untuk bertujuan membangun sistem analisis sentimen otomatis untuk mengklasifikasikan opini publik pada unggahan video pengumuman *event* di media sosial XYZ. Lebih lanjut media sosial XYZ yang dikenal memiliki karakteristik unik seperti penggunaan bahasa informal dan *slang* [9], membuat penelitian ini juga bertujuan untuk mengukur performa metode SVM dalam menangani *dataset* komentar berbahasa Indonesia yang berasal dari media sosial XYZ

Kontribusi atau *novelty* utama dari penelitian ini terletak pada penerapan metode SVM pada *dataset* yang belum banyak dieksplorasi dalam penelitian sejenis, yaitu pada data komentar media sosial XYZ berbahasa Indonesia mengenai sebuah *event* budaya dengan karakteristik yang lebih menantang seperti penggunaan bahasa informal dan *slang*. Secara praktis, penelitian ini berkontribusi memberikan sebuah model dan prototipe sistem otomatis bagi pihak penyelenggara (Dusun Semilir) untuk mengevaluasi sentimen publik secara efisien, sebagai bahan pertimbangan strategis untuk penyelenggaraan *event* di masa depan. Secara akademis, hasil penelitian ini akan memperkaya literatur analisis sentimen di Indonesia dengan menyajikan data performa SVM terhadap teks komentar pada media sosial XYZ yang tidak terstruktur, sekaligus menjadi referensi bagi penelitian selanjutnya.

2. TINJAUAN PUSTAKA

2.1. *Pengembangan Sistem*

Pendekatan pengembangan perangkat lunak dalam penelitian akademis seringkali mengadopsi model terstruktur untuk memastikan luaran yang fungsional. Peneliti [10] menekankan bahwa pengembangan sistem yang efektif harus menerjemahkan kebutuhan pengguna yang abstrak menjadi spesifikasi teknis yang teruji. Dalam konteks analisis data sosial, pengembangan sistem tidak hanya berfokus pada konstruksi kode, tetapi juga pada integrasi komponen pengolahan data (*data processing*) ke dalam antarmuka yang dapat

p-ISSN : 2803-1507 e-ISSN : 2803-1531

digunakan oleh pengguna awam untuk pengambilan keputusan [11]. Penelitian ini mengadopsi prinsip tersebut untuk menjembatani kompleksitas algoritma sentimen dengan kebutuhan praktis penyelenggara event.

2.2. *Analisis Sentimen*

Analisis sentimen atau *opinion mining* telah berkembang dari sekadar klasifikasi dokumen menjadi analisis mendalam terhadap subjektivitas teks [12]. Studi literatur menunjukkan pergeseran tren dari pendekatan berbasis leksikon (kamus) menuju pendekatan berbasis *Machine Learning* karena kemampuan adaptasinya yang lebih baik terhadap konteks bahasa yang dinamis [13]. Dalam penerapannya, analisis sentimen dapat dilakukan pada berbagai tingkatan analisis, mulai dari tingkat dokumen hingga tingkat aspek (*aspect-based*), yang memungkinkan identifikasi opini spesifik terhadap fitur tertentu dari suatu entitas [14].

2.3. *Machine Learning*

Dalam ranah pengolahan teks, *Machine Learning* (ML) menjadi pendekatan dominan karena kemampuannya mendeteksi pola non-linear dalam data berdimensi tinggi [15]. Berbeda dengan pemrograman konvensional, pendekatan *supervised learning* dalam ML memungkinkan sistem untuk "belajar" dari data berlabel (*ground truth*) untuk melakukan prediksi pada data baru. Literatur menunjukkan bahwa keberhasilan model ML sangat bergantung pada kualitas pra-pemrosesan data dan pemilihan fitur yang relevan, terutama dalam menangani data teks yang tidak terstruktur [16].

2.4. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) secara konsisten diakui dalam literatur sebagai salah satu algoritma klasifikasi teks paling *robust* [5]. Keunggulan teoritis SVM terletak pada prinsip *Structural Risk Minimization* (SRM) yang bertujuan meminimalkan batas atas kesalahan generalisasi, bukan hanya kesalahan pada data latih. Hal ini dicapai dengan mencari *hyperplane* optimal yang memisahkan kelas data dengan *margin* maksimal. Studi komparatif oleh [5] serta penelitian lanjutan menunjukkan bahwa SVM seringkali mengungguli algoritma lain seperti Naive Bayes dalam konteks klasifikasi sentimen karena ketahanannya terhadap *curse of dimensionality* yang umum terjadi pada representasi data teks.

2.5. *Media Sosial*

Karakteristik data pada platform media sosial modern menghadirkan tantangan unik dalam analisis komputasi. [17] dan [18] menyoroti bahwa konten media sosial (*User-Generated Content*) memiliki sifat *high-velocity* dan seringkali informal. Khusus pada platform berbasis video pendek seperti XYZ. Literatur terbaru mencatat adanya tantangan linguistik spesifik berupa penggunaan *slang* yang masif, ambiguitas, sarkasme, dan ketergantungan pada konteks visual yang tidak ditemukan pada platform teks konvensional [19]. Hal ini menuntut adaptasi metode pra-pemrosesan yang lebih agresif dibandingkan analisis pada teks formal [20].

2.6. *Evaluasi Model*

Untuk mengukur reliabilitas model klasifikasi, literatur standar merekomendasikan penggunaan *Confusion Matrix* sebagai basis evaluasi yang disajikan pada Tabel 1. Metrik-metrik seperti *Accuracy, Precision*, *Recall*, dan *F1-Score* menjadi krusial untuk memberikan gambaran performa yang lebih holistik [21]. Terutama pada dataset dengan distribusi kelas yang tidak seimbang (*imbalanced dataset*) yang sering ditemui dalam data media sosial.

Tabel 1. Confusion Matrix

		Actual Values	
		Positive	**Negative**
Predicted Values	**Positive**	TP	FP
	Negative	FN	TN

2.7. *Alat Pendukung Penelitian*

Penelitian ini mengintegrasikan serangkaian perangkat lunak *open-source* untuk efisiensi pengembangan. *Google Colaboratory* dipilih sebagai lingkungan eksekusi *Python* berbasis *cloud* yang memfasilitasi

p-ISSN : 2803-1507 e-ISSN : 2803-1531

komputasi data berat tanpa beban pada perangkat lokal. Untuk manajemen dataset awal, Microsoft Excel digunakan karena kemudahan fitur pengeditan. Sementara itu, *Streamlit* dimanfaatkan sebagai kerangka kerja (*framework*) utama untuk *deployment* model. Literatur teknis menyoroti keunggulan *Streamlit* dalam memfasilitasi *rapid prototyping* aplikasi data interaktif tanpa memerlukan pengembangan *backend* yang kompleks, sehingga sangat sesuai untuk validasi model secara *real-time*.

2.8. *Penelitian Terdahulu*

Sejumlah penelitian terdahulu telah membuktikan efektivitas *Support Vector Machine* (SVM) untuk analisis sentimen pada beragam platform media sosial. Studi yang dilakukan oleh [6] menerapkan SVM untuk menganalisis ulasan aplikasi *Sayurbox* yang diambil dari *Google Play Store* dan berhasil mencatatkan tingkat akurasi 89.29%. Penelitian lain oleh [7] juga menggunakan SVM, namun berfokus pada analisis sentimen berbasis aspek pada ulasan restoran di *Google Review*, dengan hasil akurasi yang tinggi yaitu 92%. Selain itu, [8] telah menerapkan SVM pada data *Twitter* terkait isu benih lobster, dan penelitian tersebut mampu menghasilkan akurasi sebesar 84,21%.

2.9. *Kebaruan Penelitian*

Meskipun studi-studi terdahulu berhasil menunjukkan keunggulan *Support Vector Machine* (SVM), terdapat kesenjangan penelitian yang jelas teridentifikasi. Perbandingan posisi kebaruan penelitian ini dibandingkan dengan studi terdahulu yang disajikan pada Tabel 2.

Tabel 2. Kebaruan Penelitian

Peneliti	Platform	Objek Data	Struktur Bahasa	Konteks
Yolanda, A. M. dan Mulya, R. T.	Google Play	Ulasan Aplikasi	Terstruktur dan kalimat lengkap	Eksplisit berupa ulasan produk
Firdaus M. F. dkk	Google Review	Ulasan Restoran	Terstruktur dan kalimat lengkap	Eksplisit berupa ulasan layanan
Pamungkas, B. dkk.	Twitter	Opini	Terstruktur dan kalimat lengkap	Eksplisit berupa opini
Penelitian Ini	Media Sosial XYZ	Komentar Video Event	Tidak terstruktur	Implisit berupa Reaksi terhadap video

Mayoritas penelitian sebelumnya, seperti [6], [7], dan [8], cenderung berfokus pada analisis sentimen di platform media sosial yang sudah umum diteliti, seperti *Twitter*, *Google Play Store*, atau *Google Review* yang tidak memiliki kompleksitas dataset seperti dataset yang diperoleh dari *platfrom* XYZ seperti :

1. Bahasa *Slang* dan Singkatan Ekstrem, Berbeda dengan ulasan Google Play yang menggunakan bahasa Indonesia baku atau semi-baku, komentar XYZ didominasi oleh singkatan ekstrem (misal: "bgt" untuk banget, "gmn" untuk bagaimana, "ygy" untuk ya guys ya) yang menyulitkan proses tokenisasi standar.
2. Ambiguitas Emoji dan Sarkasme, Pengguna XYZ sering menggunakan emoji sebagai pengganti kata sifat utama. Selain itu, budaya komentar di platform ini sarat dengan sarkasme (contoh: "Antriannya mantap sekali" padahal bermaksud mengeluh), yang seringkali gagal dideteksi oleh metode klasifikasi konvensional tanpa pra-pemrosesan khusus.

Dengan kompleksitas dataset pada platform XYZ tersebut, penelitian ini diharapkan dapat mengisi kesenjangan penelitian dengan tantangan kompleksitas dataset yang belum ditemukan oleh penelitian sejenis. Oleh karena itu, penelitian ini mengisi kesenjangan tersebut dengan menerapkan metode SVM secara khusus untuk menganalisis sentimen publik dari data komentar XYZ, guna menyediakan sistem analisis otomatis bagi pihak penyelenggara *event*.

3. METODOLOGI PENELITIAN

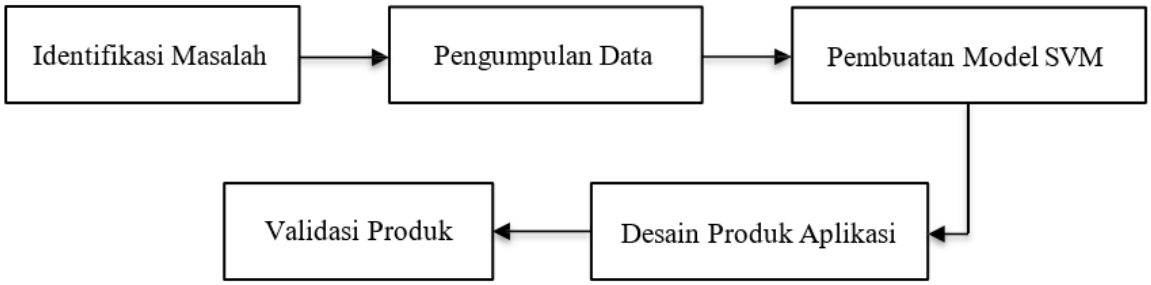
3.1. *Jenis Penelitian*

Jenis penelitian yang digunakan adalah *Research and Development* (R&D). Penelitian R&D, atau penelitian dan pengembangan, adalah strategi atau metode penelitian yang berfokus pada produksi produk baru atau penyempurnaan produk yang sudah ada, yang kemudian divalidasi efektivitasnya [22]. Tujuan dari pendekatan ini adalah untuk menghasilkan sebuah produk spesifik, yang dalam konteks penelitian ini adalah sebuah aplikasi *website* untuk analisis sentimen otomatis. Produk ini dirancang untuk dapat mengklasifikasikan sentimen dalam data komentar secara fungsional dan teruji.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

3.2. *Alur Penelitian*

Tahapan penelitian dilaksanakan secara sistematis meliputi identifikasi masalah, pengumpulan data, pengembangan model, hingga validasi produk. Tahap inti dari penelitian ini adalah pengembangan model *Machine Learning* (ML), yang mencakup ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), pembagian data, dan pelatihan model *Support Vector Machine* (SVM). Setelah model dievaluasi dan dianggap optimal, model tersebut diimplementasikan ke dalam sistem aplikasi *website* menggunakan *framework streamlit*. Sistem yang sudah jadi kemudian diuji secara fungsional dengan validasi oleh para ahli untuk mengukur kelayakannya. Alur penelitian secara visual disajikan pada Gambar 1.



Gambar 1. Alur Penelitian

3.3. *Teknik Pengumpulan Data*

Pengumpulan data dalam penelitian ini dilakukan dengan teknik *scraping* data digital dari *platform* media sosial XYZ. Objek data yang diambil adalah komentar-komentar publik yang ditinggalkan oleh pengguna pada unggahan video spesifik, yaitu pengumuman *event* "Semilir Japan Matsuri 3". Proses *scraping* ini memanfaatkan sebuah *tool* dari *platform Apify* untuk mengunduh data komentar secara otomatis.

3.4. *Metode Analisis Sentimen*

Metode analisis data dalam penelitian ini berfokus pada pengembangan model klasifikasi sentimen. Proses ini terdiri dari beberapa tahapan teknis yang krusial.

1. Pelabelan Data (*Data Labeling*), Dilakukan secara manual untuk memberikan label 'Positif' atau 'Negatif' pada setiap komentar. Selanjutnya data berlabel ini digunakan sebagai ground truth dalam pelatihan model.
2. Pra-pemrosesan Data (*Data Preprocessing*), Bertujuan membersihkan dan menstandarisasi teks agar dapat diproses oleh mesi n[23]. Tahapan ini meliputi: *Case Folding*, *Normalization* (menggunakan kamus kustom untuk menangani *slang*), *Tokenizing*, *Stopword Removal*, dan *Stemming*.
3. Ekstraksi Fitur dan Pemodelan, Teks direpresentasikan secara numerik menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) untuk menghitung bobot kata [24]. Vektor yang dihasilkan menjadi input bagi algoritma Support Vector Machine (SVM).
4. Pelatihan Model, *Dataset* yang telah di-vektorisasi dibagi menjadi dua bagian: 80% data digunakan sebagai data latih (*training data*) dan 20% data digunakan sebagai data uji (*testing data*). Proses pelatihan dan pengujian model ini dilakukan menggunakan bahasa pemrograman *Python* di lingkungan *Google Colaboratory*. Beberapa *library* utama yang digunakan adalah *Scikit-learn* untuk implementasi TF-IDF dan SVM, serta *Pandas* dan *Numpy* untuk manipulasi data.
5. Pengujian Model, Pengujian model dilakukan melalui pengujian model (*model evaluation*) yang dilakukan secara kuantitatif menggunakan *Confusion Matrix* terhadap data uji. Dari *Confusion Matrix* ini, dihitung metrik performa utama yaitu *Accuracy* (akurasi), *Precision* (presisi), *Recall* (sensitivitas), dan *F1-Score* untuk mengukur seberapa baik model dapat menggeneralisasi data baru [21].

3.5. *Implementasi Sistem*

Implementasi model SVM yang telah dilatih diwujudkan dalam bentuk aplikasi *website*. Arsitektur sistem ini dikembangkan menggunakan *Streamlit*, sebuah *framework open-source Python* yang dirancang untuk membangun dan menyebarkan aplikasi *web* berbasis data dan *Machine Learning* secara interaktif. *Streamlit* menangani integrasi antara model *backend* dan antarmuka pengguna (*frontend*), memungkinkan model SVM yang telah disimpan untuk diakses dan dieksekusi secara langsung. Aplikasi ini bertugas menerima masukan

p-ISSN : 2803-1507 e-ISSN : 2803-1531

file berformat CSV dari pengguna melalui komponen antarmuka yang disediakan, memproses masukan tersebut melalui alur pra-pemrosesan dan TF-IDF yang sama persis dengan saat pelatihan, dan kemudian memprediksi sentimennya. Hasil prediksi ('Positif' atau 'Negatif') kemudian ditampilkan secara dinamis kepada pengguna langsung di antarmuka *web*.

3.6. *Teknik Validasi*

Untuk memastikan kelayakan dan efektivitas produk aplikasi yang dikembangkan, dilakukan tahap validasi produk. Validasi ini menggunakan pendekatan penilaian ahli (*expert judgment*) yang melibatkan dua kelompok validator [10]. Kelompok pertama adalah Ahli Internal, yang merupakan ahli pakar atau dosen yang memiliki keahlian di bidang pengembangan *software* dan *machine learning*. Kelompok kedua adalah Ahli Eksternal, yang bertindak sebagai penguji lapangan atau pengguna potensial dari aplikasi tersebut.

Instrumen yang digunakan untuk validasi adalah kuesioner yang dirancang menggunakan skala *Likert*. Kuesioner ini disesuaikan untuk mengukur aspek fungsionalitas, kegunaan, dan keandalan sistem analisis sentimen yang dibangun [25]. Hasil dari kuesioner kedua kelompok validator kemudian dianalisis secara deskriptif untuk menentukan tingkat kelayakan produk, yang dikategorikan menjadi "Sangat Baik", "Baik", "Cukup", atau "Kurang".

4. HASIL DAN PEMBAHASAN

4.1. *Hasil Pengumpulan dan Prapemrosesan Data*

Pengumpulan data komentar dikumpulkan melalui teknik *web scraping* dari media sosial XYZ. Hasil dari pengumpulan data ini adalah berupa dataset berjumlah 2133 data dari semua komentar pada konten Semilir Japan Matsuri 3 di akun resmi media sosial XYZ Dusun Semilir. Setiap komentar kemudian diperlukan proses pelabelan manual. Pelabelan dilakukan dengan fungsi sebagai *ground truth* bagi model *Machine Learning* (ML) menggunakan *Microsoft excel*. Tahap prapemrosesan data dilakukan pada platfrom Google Colaboratory. Pembersihan data menghasilkan 607 data komentar bersih yang siap digunakan untuk proses selanjutnya seperti *case folding*, normalisasi, tokenisasi penghapusan stopword, *stemming* pada Gambar 2.

	text	sentimen	text_casefolding	Normalized	tokenized	stopword_removed	stemmed
0	rumahku deket ama dusun semilir, malahan dekee...	positif	rumahku deket ama dusun semilir, malahan dekee...	rumahku dekat sama dusun semilir, malahan deka...	['rumahku', 'dekat', 'sama', 'dusun', 'semilir...	['rumahku', 'dekat', 'sama', 'dusun', 'semilir...	['rumah', 'dekat', 'sama', 'dusun', 'milir', '...
1	abisssss padahal dekettt	negatif	abisssss padahal dekettt	habis padahal dekat	['habis', 'padahal', 'dekat', ' ']	['habis', 'padahal', 'dekat', ' ']	['habis', 'padahal', 'dekat', '']
2	ka 26 mei ada ya tp aku ga boleh kesana sedih ...	positif	ka 26 mei ada ya tp aku ga boleh kesana sedih ...	ka 26 mei ada ya tapi aku tidak boleh kesana s...	['ka', '26', 'mei', 'ada', 'ya', 'tapi', 'aku'...	['ka', '26', 'mei', 'aku', 'kesana', 'sedih', ...	['ka', '26', 'mei', 'aku', 'kesana', 'sedih', ...
3	hehe ak dibolehin sm ortu	positif	hehe ak dibolehin sm ortu	hehe saya dibolehin sama orang tua	['hehe', 'saya', 'dibolehin', 'sama', 'orang'...	['hehe', 'dibolehin', 'sama', 'orang', 'tua', ...	['hehe', 'dibolehin', 'sama', 'orang', 'tua', '']
4	deket sih,tapi gada uang	negatif	deket sih,tapi gada uang	dekat sih,tapi tidak ada uang	['dekat', 'sih', ',', 'tapi', 'tidak', 'ada', ...	['dekat', 'sih', ',', 'uang']	['dekat', 'sih', '', 'uang']
...	...	...	...	...	...	...	...
602	udah sold out kah baru bangun tidur lupa kalau...	negatif	udah sold out kah baru bangun tidur lupa kalau...	sudah sold out kah baru bangun tidur lupa kala...	['sudah', 'sold', 'out', 'kah', 'baru', 'bangu...	['sold', 'out', 'baru', 'bangun', 'tidur', 'lu...	['sold', 'out', 'baru', 'bangun', 'tidur', 'lu...
603	aslinya mau kesana tapi ga ada dana??	negatif	aslinya mau kesana tapi ga ada dana??	aslinya mau kesana tapi tidak ada dana??	['aslinya', 'mau', 'kesana', 'tapi', 'tidak', ...	['aslinya', 'mau', 'kesana', 'dana', '?', '?']	['asli', 'mau', 'kesana', 'dana', '', '']
604	mau tapi gabisa	negatif	mau tapi gabisa	mau tapi gabisa	['mau', 'tapi', 'gabisa']	['mau', 'gabisa']	['mau', 'gabisa']
605	AAAAAA AKU DAPET PLATINUM ??????????????????	positif	aaaaaa aku dapet platinum ??????????????????	aaaaaa aku dapet platinum ??????????????????	['aaaaaa', 'aku', 'dapet', 'platinum', '?', '?...	['aaaaaa', 'aku', 'dapet', 'platinum', '?', '?...	['aaaaaa', 'aku', 'dapet', 'platinum', '', ''...
606	Mauuu:(, Udah ada duitnya tapi belum ada yang ...	negatif	mauuu:(, udah ada duitnya tapi belum ada yang ...	mauuu:(, sudah ada duitnya tapi belum ada yang...	['mauuu', ':', '(', ',', 'sudah', 'ada', 'duit...	['mauuu', ':', '(', ',', 'duitnya', 'nganterin...	['mauuu', '', '', '', 'duit', 'nganterin', ''...

607 rows × 7 columns

Gambar 2. Hasil Prapemrosesan Data

4.2. *Hasil Pengujian Model Klasifikasi SVM*

Model *Support Vector Machine* (SVM) yang telah dilatih menggunakan data latih kemudian diuji performanya menggunakan data uji. Evaluasi model dilakukan dengan membandingkan hasil prediksi model terhadap label *ground truth* dari data uji. Hasil evaluasi disajikan secara rinci dalam bentuk *Confusion Matrix* pada Tabel 3, yang menunjukkan bagaimana model mengklasifikasikan data uji dengan benar maupun salah.

Tabel 3. Hasil Confusion Matrix

	Prediksi Positif	Prediksi Negatif
Aktual Positif	62	5
Aktual Negatif	20	35

Berdasarkan *Confusion Matrix* tersebut, kinerja model diukur menggunakan empat metrik evaluasi utama: *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Model SVM yang dikembangkan berhasil mencapai tingkat *Accuracy* (akurasi) keseluruhan sebesar 79.51%, yang menunjukkan kemampuannya dalam mengklasifikasikan sentimen secara umum. Nilai *Precision* (presisi) tercatat sebesar 76%, sementara nilai

p-ISSN : 2803-1507 e-ISSN : 2803-1531

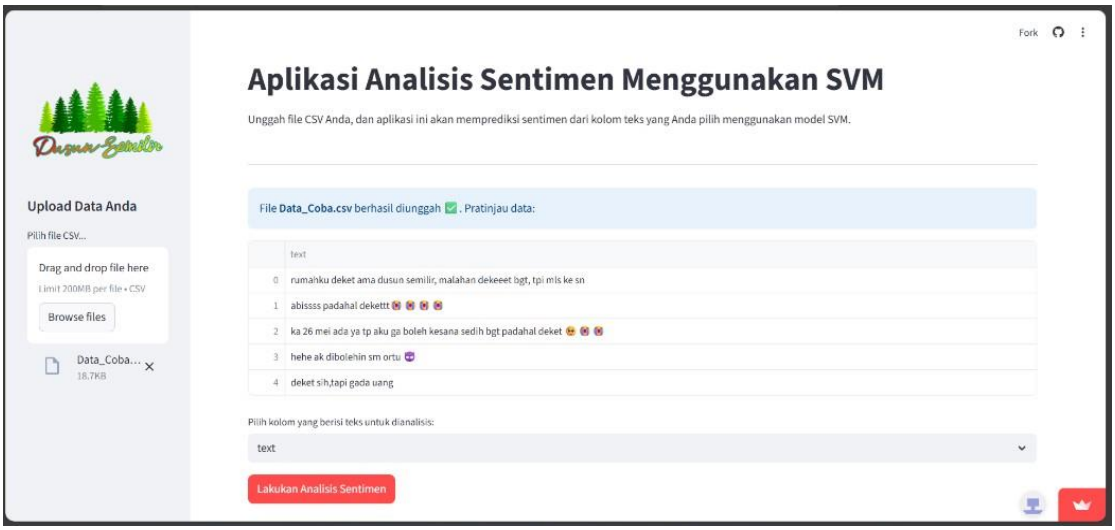
Recall (sensitivitas) mencapai 93%. Nilai *F1-Score*, yang merupakan rata-rata harmonik dari *Precision* dan *Recall*, adalah sebesar 83%.

4.3. *Hasil Pengembangan Sistem*

Hasil bagian dari penelitian ini adalah model ML yang telah teruji diimplementasikan ke dalam sebuah produk aplikasi *website*. Sistem ini dikembangkan menggunakan *Streamlit*, sebuah *framework open-source Python* yang dirancang untuk membangun dan menyebarkan aplikasi *web* secara interaktif [26]. Aplikasi ini memungkinkan pengguna untuk memasukkan data dari kumpulan teks komentar berformat CSV melalui sebuah komponen upload. Data tersebut kemudian diproses langsung oleh aplikasi, melalui alur pra-pemrosesan dan ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) yang sama persis dengan yang digunakan saat pelatihan model. Model SVM kemudian memprediksi sentimen dari vektor TF-IDF tersebut dan mengembalikan hasilnya ('Positif' atau 'Negatif') untuk ditampilkan secara dinamis kepada pengguna di antarmuka *web*. Tampilan antarmuka utama dari aplikasi yang dikembangkan dapat dilihat pada Gambar 3, Gambar 4, dan Gambar 5.

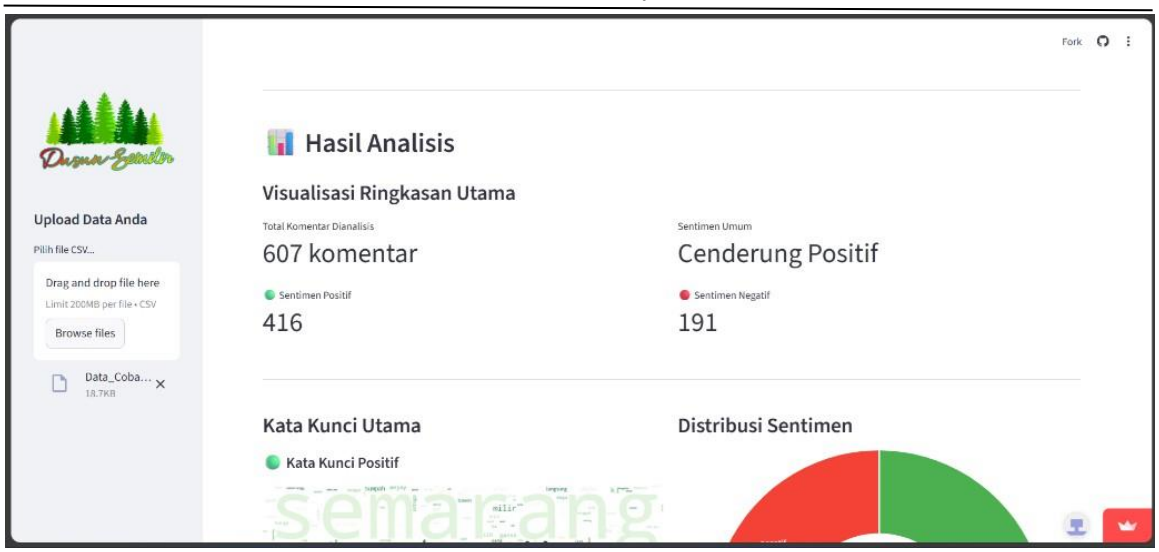


Gambar 3. Tampilan Awal Halaman



Gambar 4. Tampilan Setelah Upload Data

p-ISSN : 2803-1507 e-ISSN : 2803-1531



Gambar 5. Tampilan Setelah Analisis Data

4.4. *Hasil Validasi Sistem*

Validasi sistem dilakukan dengan menyebarkan kuesioner skala *Likert* kepada ahli internal dan eksternal. Hasil validasi sistem dianalisis menggunakan standar usabilitas ISO 9241-11. Standar ini mendefinisikan usabilitas dalam tiga komponen utama: Efektivitas (kemampuan pengguna mencapai tujuan secara akurat), Efisiensi (sumber daya yang dikeluarkan untuk mencapai tujuan), dan Kepuasan (kenyamanan pengguna). Hasil validasi sistem ini disajikan pada Tabel 4.

Tabel 4. Hasil Validasi Sistem

Kriteria	Indikator	Skor Ahli Internal	Skor Eksternal
Efektifitas	Program berjalan dengan baik	4	3
	Program sesuai sistem yang dirancang	4	3
	Rancangan UML tepat	4	3
	Program menyediakan database	3	2
	Input & output sesuai rancangan	4	2
	Gambar & informasi sesuai lapangan	4	3
Efisiensi	Kemudahan akses program	4	2
	Kemudahan edit data	3	3
	Desain program mudah dipahami	4	2
Kepuasan	Kenyamanan tampilan (interface)	4	3
Total Skor		**38**	**26**

Hasil validasi mengungkapkan perbedaan signifikan khususnya pada aspek efisiensi dan efektivitas. Ahli internal memberikan penilaian hampir sempurna pada seluruh indikator, yang mengindikasikan bahwa secara teknis, fungsionalitas sistem, logika algoritma SVM, dan struktur basis data telah memenuhi standar. Namun, perspektif penguji eksternal menunjukkan penurunan skor. Secara spesifik, indikator kemudahan akses dan pemahaman desain dinilai sangat tinggi oleh ahli internal dengan skor 4, namun dinilai cukup atau dengan skor 2 oleh pengguna lapangan. Hal ini mengisyaratkan adanya beban kognitif yang lebih tinggi pada pengguna awam dibandingkan ahli, di mana antarmuka sistem dirasa kurang intuitif meskipun fungsional. Selain itu, penurunan skor pada indikator input/output dari skor 3 oleh ahli internal menjadi 2 oleh ahli eksternal menandakan bahwa meskipun sistem mampu memproses data secara akurat, mekanisme umpan balik kepada pengguna mengenai status data masih perlu dipertegas.

Hasil validasi yang disajikan pada Tabel 4. menunjukkan total skor Ahli Internal (Pakar) memberikan skor total 38 dari 40, yang berdasarkan kriteria penilaian masuk dalam kategori "Sangat Baik". Sementara itu, Ahli Eksternal (Penguji Lapangan/Pengguna) memberikan skor total 26 dari 30, yang diklasifikasikan sebagai "Baik". Temuan ini menegaskan bahwa meskipun sistem memiliki validitas konten yang tinggi sehingga layak digunakan, aspek interaksi manusia-komputer memerlukan penyempurnaan untuk menjembatani kesenjangan antara pengembang dan pengguna akhir.

5. DISKUSI

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Hasil penelitian menunjukkan bahwa model SVM berhasil diterapkan untuk menganalisis sentimen pada data komentar media sosial XYZ dengan akurasi 79.51%. Jika dibandingkan dengan penelitian terdahulu yang menggunakan SVM di *platform* lain, seperti penelitian Yolanda, A. M. dan Mulya, R. T. (89.29% di *Google Play Store*) [6] dan Firdaus M. F. dkk. (92% di *Google Review*) [7], capaian akurasi dalam penelitian ini terlihat lebih rendah. Penurunan akurasi ini merupakan konsekuensi dari karakteristik data XYZ yang menghadirkan tantangan analitis yang tidak ditemukan pada penelitian lain dengan menggunakan platform dengan ulasan formal. Karakteristik data komentar pada media sosial XYZ didominasi oleh variasi ejaan tak baku yang ekstrem seperti: "bgt", "bgtt", "banget", "bngttt". Variasi ejaan tak baku ini belum diatasi dengan kamus kata yang digunakan pada tahap normalisasi sehingga mengakibatkan penurunan akurasi SVM.

Terlepas dari tantangan analitis tersebut, validasi sistem membuktikan keberhasilan implementasi model menjadi produk aplikasi yang fungsional. Skor validasi pakar (38/40) menegaskan kehandalan arsitektur sistem yang bebas *error*, sementara skor pengguna (26/40) menandai perlunya penyederhanaan antarmuka untuk pengguna awam. Hal ini membuktikan bahwa meskipun akurasi model terpengaruh oleh *noise* data, integrasi SVM dengan *framework* Streamlit tetap menghasilkan solusi praktis yang valid ("Sangat Baik" secara teknis) untuk pemantauan sentimen bagi penyelenggara *event*.

Secara teoretis, penelitian ini memberikan kontribusi penting dengan mengisi kesenjangan penelitian terkait penerapan SVM pada *platform* media sosial XYZ, membuktikan bahwa metode tersebut tetap relevan meskipun dihadapkan pada data yang lebih *noisy*. Secara praktis, penelitian ini menghasilkan dua keluaran. Pertama, hasil analisis sentimen (68.5% positif) memberikan validasi kuantitatif bagi penyelenggara *event* bahwa pengumuman mereka diterima dengan baik oleh audiens media sosial XYZ. Kedua, produk aplikasi R&D yang telah divalidasi "Baik" dan "Sangat Baik" oleh para ahli menyediakan sebuah *tool* praktis yang dapat digunakan oleh praktisi atau peneliti lain untuk melakukan analisis sentimen cepat pada data serupa.

6. KESIMPULAN DAN SARAN

6.1 *Kesimpulan*

Berdasarkan hasil penelitian dan pembahasan yang telah dipaparkan, dapat ditarik beberapa kesimpulan yang menjawab rumusan masalah. Pertama, sistem analisis sentimen otomatis ini telah berhasil diimplementasikan menggunakan metode *Machine Learning Support Vector Machine* (SVM). Penerapan metode ini pada data komentar *event* Semilir Japan Matsuri 3 menunjukkan bahwa sentimen audiens secara umum cenderung positif, yang dibuktikan dengan 68.5% dari keseluruhan komentar terklasifikasi sebagai positif. Kedua, sistem otomatis yang dibangun dapat berfungsi dengan baik dalam menganalisis sentimen, yang ditunjukkan oleh kinerja model *machine learning* pada data uji. Model tersebut berhasil mencapai tingkat akurasi sebesar 79.51%, *recall* sebesar 93%, *precision* sebesar 76%, dan *F1-Score* sebesar 83%.

Selanjutnya, aplikasi sistem ini telah melalui tahap uji validasi kelayakan oleh para ahli, yang terdiri dari ahli internal (ahli pakar) dan ahli eksternal (penguji lapangan). Hasil dari uji validasi ahli internal (Ahli Pakar) memberikan skor 38, yang tergolong dalam kategori "sangat baik". Sementara itu, hasil uji validasi dari ahli eksternal (Penguji Lapangan) memberikan skor 26, yang masuk dalam kategori "baik". Dengan demikian, berdasarkan kedua hasil pengujian tersebut, Sistem Analisis Sentimen Otomatis menggunakan metode *Machine Learning Support Vector Machine* (SVM) ini dapat dinyatakan valid, efektif, dan layak untuk digunakan.

6.2 *Saran*

Berdasarkan kesimpulan dari hasil penelitian, terdapat beberapa saran yang dapat dijadikan bahan pertimbangan untuk pengembangan sistem serupa di masa depan. Saran pertama terkait dengan pemeliharaan model. Sistem analisis sentimen otomatis ini disarankan untuk selalu diperbaharui model *machine learning*-nya secara berkala. Proses pembaharuan ini diperlukan agar hasil analisis yang dikeluarkan oleh sistem dapat tetap relevan dan memiliki akurasi yang semakin baik seiring waktu. Saran kedua berkaitan dengan pengalaman pengguna, di mana desain sistem aplikasi perlu ditingkatkan agar tampilannya bisa lebih dinamis. Peningkatan desain ini bertujuan agar aplikasi menjadi lebih menarik secara visual dan lebih mudah digunakan oleh pengguna.

DAFTAR PUSTAKA

[1] T. Fajriah and E. R. Ningsih, “Pengaruh Teknologi Komunikasi terhadap Interaksi Sosial di Era Digital,” *Merdeka Indonesia Jurnal International*, vol. 4, no. 1, pp. 149–158, Jun. 2024, doi:

p-ISSN : 2803-1507 e-ISSN : 2803-1531

10.5555/miji.v4i1.99.

[2] Agustin, N., & Rahayu, S., "The Role of Social Media as a Micro-Ecosystem in Supporting Community-Based E-Learning Platforms: A Systematic Literature Review," *Journal of Technology Informatics and Engineering*, vol. 4, no. 3, pp. 369–390, Dec. 2025, doi: 10.51903/jtie.v4i3.

[3] I.-D. Anastasiei and M. R. Georgescu, "Automated vs Manual Content Analysis – A Retrospective Look," *Scientific Annals of Economics and Business*, vol. 67, no. Special Issue, pp. 57–67, Dec. 2020, doi: 10.47743/saeb-2020-0025.

[4] Z. A. Nurlaila, "Algoritma Support Vector Machine (SVM) dan Naïve Bayes pada Analisis Sentimen Sosial Media Twitter: Studi Kasus Kuliah Luring Setelah Pandemi Covid-19," Universitas Pendidikan Indonesia, 2023. [Online]. Available: http://repository.upi.edu/101281

[5] Rahardjo, D. A. S., Andriana, M., & Prasastono, S. H., "Consumer Behavior and Purchase Intentions: A Machine Learning-Based Sentiment Analysis of Social Media Data," *Journal of Management and Informatics*, vol. 3, no. 1, pp. 157-173, 2025, doi: 10.51903/jmi.v3i1.45.

[6] A. M. Yolanda, R. T. Mulya, and R. T. Mulya, "Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store," *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 6, no. 02, pp. 76–83, Sep. 2024, doi: 10.35580/variansiunm258.

[7] M. F. Firdaus, D. E. Ratnawati, and N. Y. Setiawan, "Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Menggunakan Algoritma Support Vector Machine (Studi Kasus: Depot Bamara)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1265–1272, 2024, doi: 10.25126/jtiik.2024117564.

[8] B. Pamungkas, M. E. Purbaya, and D. J. A.K, "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 3, no. 2, pp. 10–20, 2021. doi: 10.20895/inista.v3i2.243.

[9] A. Febrianti, Y. R. Agustin, D. P. Pascha F, A. Rusdianti, and E. Sholihatin, "Semantic Analysis of Slang Varieties by Gen Z in TikTok Application," *Journal of Humanities and Social Studies*, vol. 1, no. 03, pp. 1257–1264, 2023. [Online]. Available: https://humasjournal.my.id/index.php/hj/article/view/231

[10] R. S. Pressman and B. R. Maxim, *Software Engineering: A Practitioner's Approach*, 8th ed. New York, NY, USA: McGraw-Hill, 2014. [Online]. Available: https://www.biblio.com/book/software-engineering-practitioners-approach-eighth-edition/d/1669690803

[11] K. E. Kendall and J. E. Kendall, *Systems Analysis and Design, Global Edition*. Pearson Education, 2019. [Online]. Available: https://books.google.co.id/books?id=F8AoEAAAQBAJ

[12] B. Liu, *Sentiment Analysis and Opinion Mining*, Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2012. doi: 10.2200/s00416ed1v01y201204hlt016.

[13] Junaedi, A. Hendra Gunawan, V. Kuswanto, and Jonathan, "Eksplorasi Algoritma Support Vector Machine untuk Analisis Sentimen Destinasi Wisata di Indonesia," *bit-Tech*, vol. 7, no. 2, pp. 323–330, Dec. 2024, doi: 10.32877/bt.v7i2.1810.

[14] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '04. New York, NY, USA: Association for Computing Machinery, 2004, pp. 168–177. doi: 10.1145/1014052.1014073.

[15] I. Id, *Machine Learning : Teori, Studi Kasus dan Implementasi Menggunakan Python*. 2021. doi: 10.5281/zenodo.5113507.

[16] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. in Adaptive Computation and Machine Learning series. MIT Press, 2012. [Online]. Available: https://books.google.co.id/books?id=RC43AgAAQBAJ

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[17] Pratama, N., Izzudin, M. A., & Ilham, I., "The Effect of User Satisfaction (EUCS) on User Loyalty of Spotify among Surabaya Students," *Jurnal Ilmiah Sistem Informasi (JUISI)*, vol. 5, no. 1, pp. 186–199, Jan. 2026, doi: 10.51903/s15fn419.

[18] Danah M. Boyd and N. B. Ellison, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, Oct. 2007, doi: 10.1111/j.1083-6101.2007.00393.x.

[19] D. Purnamasari *et al.*, *Pengantar Metode Analisis Sentimen*. Depok: Penerbit Gunadarma, 2023. [Online]. Available: https://penerbit.gunadarma.ac.id/2023/09/13/pengantar-metode-analisis-sentimen

[20] B. Batrinca and P. C. Treleaven, "Social Media Analytics: A Survey of Techniques, Tools and Platforms," *AI Soc*, vol. 30, no. 1, pp. 89–116, 2015, doi: 10.1007/s00146-014-0549-4.

[21] V. KP, R. AB, G. HL, V. Ravi, and M. Krichen, "A Tweet Sentiment Classification Approach using an Ensemble Classifier," *International Journal of Cognitive Computing in Engineering*, vol. 5, pp. 170– 177, 2024, doi: 10.1016/j.ijcce.2024.04.001.

[22] M. Waruwu, "Metode Penelitian dan Pengembangan (R&D): Konsep, Jenis, Tahapan dan Kelebihan," *Jurnal Ilmiah Profesi Pendidikan*, vol. 9, no. 2, pp. 1220–1230, May 2024, doi: 10.29303/jipp.v9i2.2141.

[23] Y. Peng, G. Wang, J. Guo, and Z. Wang, "Application of Short Text Classification Model Based on GPT-3.5 in E-Commerce," *Journal of Organizational and End User Computing*, vol. 36, no. 1, 2024, doi: 10.4018/joeuc.356500.

[24] F. E. Cahyanti, Adiwijaya, and S. Al Faraby, "On the Feature Extraction for Sentiment Analysis of Movie Reviews Based on SVM," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, 2020, pp. 1–5. doi: 10.1109/icoict49345.2020.9166397.

[25] J. Nielsen, *Usability Engineering*. in Interactive Technologies. Elsevier Science, 1994. [Online]. Available: https://books.google.co.id/books?id=95As2OF67f0C

[26] R. Yadav and H. Raheman, "Machine Learning-Based Estimation of Agricultural Tyre Sinkage: A Streamlit Web Application," *J Terramech*, vol. 119, p. 101055, 2025, doi: 10.1016/j.jterra.2025.101055.