



ANALISIS SENTIMEN TERHADAP PENGGUNA MOBIL LISTRIK BERDASARKAN KOMENTAR YOUTUBE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN TEKNIK ENSEMBLE

Bunga Juliana Azzahra*, Asep Id Hadiana, Agus Komarudin

Program Studi Informatika, Fakultas Sains dan Informatika, Universitas Jendral Achmad Yani,
Jl.Terusan Jendral Sudirman, Kota Cimahi, Jawa Barat, Indonesia

*email: azzahrabungajulianaa@gmail.com

Received: 2025-04-18 Accepted: 2025-06-06 Published: 2025-06-28

Abstrak

Kemajuan teknologi kendaraan listrik telah memicu beragam perdebatan di masyarakat, sehingga penting untuk menganalisis sentimen publik terhadap inovasi ini. Penelitian ini bertujuan untuk mengevaluasi opini masyarakat mengenai mobil listrik melalui analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM) dan teknik *ensemble stacking*. Data yang digunakan berjumlah 1.516 komentar yang diambil dari YouTube. Proses penelitian meliputi pengumpulan data, *pra-pemrosesan* data untuk menyiapkan teks bagi model pembelajaran mesin, penerapan model menggunakan data uji dan data latih, serta evaluasi model. Model yang diterapkan meliputi SVM dengan kernel RBF, polinomial, dan sigmoid sebagai *base learner*, serta regresi logistik sebagai *meta learner* dalam *ensemble stacking*. Hasil penelitian menunjukkan kinerja terbaik pada rasio pembagian data 70:30, dengan *ensemble stacking* mencapai akurasi tertinggi sebesar 76,21%. Model SVM polinomial (75,99%) dan regresi logistik (75,55%) juga memberikan hasil yang unggul. Penelitian ini menunjukkan efektivitas *ensemble stacking* dalam meningkatkan akurasi analisis sentimen serta memberikan wawasan mendalam mengenai pandangan masyarakat terhadap mobil listrik.

Kata kunci: analisis sentimen, Support vector Machine, ensemble stacking, mobil listrik, TF-IDF.

Abstract

The advancement of electric vehicle technology has sparked diverse debates within society, highlighting the importance of analyzing public sentiment toward this innovation. This study aims to evaluate public opinion regarding electric cars through sentiment analysis, employing the *Support Vector Machine* (SVM) algorithm and an *ensemble stacking* technique. A total of 1,516 comments from YouTube were used as data. The research process encompassed several stages: data collection, preprocessing to prepare textual data for machine learning models, model implementation using both training and testing datasets, and model evaluation. The models applied include SVM with RBF, polynomial, and sigmoid kernels serving as *base learners*, with logistic regression acting as the *meta-learner* within the *ensemble stacking* framework. The results indicate optimal performance at a 70:30 train-test data split. At this ratio, the *ensemble stacking* model achieved the highest accuracy of 76.21%. The polynomial SVM (75.99%) and logistic regression (75.55%) models also produced strong results. This study demonstrates the effectiveness of *ensemble stacking* in enhancing sentiment analysis accuracy and offers valuable insights into public perceptions of electric vehicles.

Keywords: sentiment analysis, Support Vector Machine, ensemble stacking, electric vehicles, TF-IDF.

How to cite (in APA style): Azzahra, B. J., Asep Id, H., & Agus, K. (2025). Analisis sentiment terhadap pengguna mobil listrik berdasarkan komentar youtube menggunakan metode support vector machine (svm) dan teknik ensemble. *Jurnal Pendidikan Informatika Dan Sains*, 14(1), 10–23. <https://doi.org/10.31571/saintek.v14i1.8871>



PENDAHULUAN

Munculnya mobil listrik telah menimbulkan beragam pendapat di kalangan masyarakat dan menjadi perbincangan hangat di media sosial. Sebagian pihak melihatnya sebagai solusi potensial untuk mengurangi ketergantungan pada migas, sementara lainnya menyoroti pentingnya kesiapan infrastruktur sebelum beralih ke kendaraan listrik, meskipun teknologi ini terbukti mampu menurunkan emisi gas rumah kaca hingga 38% dibandingkan mobil berbahan bakar bensin (Puricelli et al., 2025). Seiring kemajuan ini, inovasi di sektor otomotif pun berkembang, salah satunya melalui kehadiran mobil listrik yang menggunakan energi listrik sebagai penggerak utama (Momon & Astuti, 2020). Di sisi lain, pesatnya perkembangan teknologi digital telah mengubah pola kerja, komunikasi, dan sistem bisnis (Harahap et al., 2024). Tanggapan masyarakat Indonesia yang beragam terhadap mobil listrik menjadi latar belakang penting bagi para peneliti untuk menggali dan menganalisis opini publik melalui informasi yang tersebar di media sosial.

Dalam hal ini, dilakukannya penggalan opini atau analisis sentimen merupakan proses untuk memahami opini terhadap suatu topik berdasarkan komentar di media sosial. Analisis ini dapat mengklasifikasikan teks ke dalam kategori sentimen seperti positif, negatif, atau netral. (Satria et al., 2022). Pada proses analisis sentimen dilakukan penggunaan pembelajaran mesin untuk mengklasifikasikan data menggunakan Support Vector Machine (SVM) (Himawan et al., 2018). SVM bekerja dengan mencari hyperplane terbaik yang secara optimal memisahkan kelas sentimen. Dalam konteks ini, beberapa fungsi kernel meliputi kernel RBF, sigmoid dan polynomial. Untuk meningkatkan akurasi klasifikasi sentimen secara signifikan, prediksi dari model-model SVM dengan berbagai kernel ini dapat digunakan menggunakan teknik ensemble Stacking.

Pada penelitian sebelumnya menjelaskan tentang analisis sentimen terhadap aplikasi pencari kerja dengan menggunakan pendekatan metode Lexicon Based Features dan Support Vector Machine (SVM) yang menghasilkan akurasi sebesar 81% untuk aplikasi Glints, kemudian 76% untuk kupu, dan untuk aplikasi Pintarnya 82% (Himawan et al., 2018). Dengan menganalisis sentimen juga bisa dilakukan terhadap penyedia layanan internet IndiHome dan First Media menggunakan metode Ensemble Stacking dan model pembelajaran mesin seperti SVM, yang dimana hasil penelitian ini menunjukkan bahwa metode SVM menghasilkan sebesar 96,68% (Arsi & Waluyo, 2021).

Penelitian sebelumnya menjelaskan tentang analisis sentimen terhadap penerbangan pesawat dengan pendekatan ensemble yang menggunakan metode Naïve Bayes, Logistic Regression, Decision Tree, SVM Linear, SVM RBF, SVM Polynomial dan menggunakan word embedding. Sehingga hasil akurasi pelatihan data yang paling besar oleh metode Logistic Regression, SVM Linear, dan SVM RBF (Daniati & Utama, 2023). Selanjutnya, penelitian lain dilakukan dengan pendekatan menggunakan metode Ensemble Learning yang bertujuan untuk mendeteksi adanya serangan Distributed Denial of Service (DDoS), sebagai upaya meningkatkan akurasi dan efektivitas dalam sistem deteksi tersebut. Peneliti menerapkan Soft Voting Classifier untuk menghasilkan kombinasi tertinggi yang diperoleh metode Logistic Regression, Multi Layer Perceptron dan Naïve Bayes, yaitu sebesar 98,32% (Jones et al., 2024). Dalam penelitian terdahulu, metode ensemble stacking yang menggabungkan sclogram images dan ECG grayscale images untuk diagnosis penyakit kardiovaskular, yang menghasilkan akurasi 93.97%, menunjukkan efektivitasnya dalam meningkatkan diagnosis CVD (Yoon & Kang, 2023). Dengan hal ini, teknik ensemble stacking dapat meningkatkan akurasi dengan berbagai algoritma yang tepat.

Analisis sentimen menggunakan teknik ensemble stacking dari berbagai algoritma salah satunya SVM dan Logistic Regression. Dari berbagai algoritma yang di uji, model Logistic Regression sebagai base learner menghasilkan akurasi tertinggi sebesar 74,6%. Setelah penerapan metode ensemble stacking akurasi meningkat menjadi 75,3% (Leidiyana, 2023). Metode ensemble stacking juga dapat menggabungkan tiga algoritma pembelajaran mesin: Random Forest, XGBoost

dan AdaBoost dengan Extra Trees Regressor sebagai meta-learner, dengan membandingkan model dasarnya menunjukkan peningkatan akurasi dalam memprediksi output daya pada fotovoltaik (Abdellatif et al., 2022). Penelitian selanjutnya menyoroti penggunaan metode ensemble stacking yang secara khusus dikembangkan untuk lingkungan Internet of Things (IoT). Metode ini menggabungkan empat model *deep learning* yakni Multi-layer Perceptron (MLP), Deep Neural Network (DNN), Convulutional Neural Network (CNN), dan Long Short- Term Memory (LSTM). Berdasarkan hasil pengujian, model DIS-IoT menunjukkan performa paling unggul dengan akurasi mencapai 99,6%. Karena memang dirancang untuk kebutuhan spesifik di ranah IoT (Lazzarini et al., 2023). Selain itu, pendekatan ensemble stacking juga bisa dikombinasikan dengan model klasik seperti Logistic Regression, Decision Tree, K-Nearest Neighbor, dan Support Vector Machine. dalam penelitiannya, gabungan metode ini berhasil mencapai akurasi hingga 99.9% (Sunarko et al., 2023). Menariknya, jika melihat dari beberapa jurnal sebelumnya, penggunaan metode ensemble untuk analisis sentimen masih tergolong tidak terlalu banyak ditemui.

Penelitian ini dilakukan dari beragamnya opini masyarakat mengenai mobil listrik dimedia sosial menimbulkan kebutuhan akan analisis sentimen untuk menggali data dan memahami opini yang terkandung dalam komentar pengguna media sosial YouTube.

Penelitian ini bertujuan untuk mengaplikasikan beberapa metode klasifikasi, yaitu Support Vector Machine (SVM), Logistic Regression, serta pendekatan berbasis teknik Ensemble. Dalam penerapannya, SVM digunakan dengan berbagai jenis kernel seperti polynomial, RBF, dan sigmoid, guna mengeksplorasi perbedaan performa masing-masing varian. Dalam proses analisis, data diklasifikasikan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Untuk menilai kinerja model, digunakan beberapa metrik evaluasi seperti akurasi, presisi, recall, dan f1-score. Menariknya, penerapan teknik ensemble stacking terbukti mampu meningkatkan performa secara keseluruhan di bandingkan jika hanya menggunakan satu model secara tunggal. Hal ini sekaligus memperlihatkan bahwa pendekatan ensemble cukup efektif untuk menangani permasalahan klasifikasi sentimen yang kompleks.

METODE

Penelitian ini menggunakan dataset sebanyak 1516 data yang diambil dari komentar media sosial dan kemudian berisikan menjadi 3 kelas yaitu sentimen positif, sentimen negatif, dan sentimen netral. Sumber data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, yang dapat diakses melalui tautan berikut: <https://www.kaggle.com/code/billycemerson/analisis-sentimen-intensif-mobil-listrik-svm/notebook>. Untuk memastikan alur penelitian yang sistematis dan terstruktur, proses pelaksanaan penelitian ini dibagi ke dalam beberapa tahapan, mulai dari pengumpulan data hingga evaluasi model. Setiap tahapan dirancang untuk mendukung akurasi dan keandalan hasil klasifikasi sentimen. Rangkaian proses penelitian tersebut dapat dilihat secara visual pada Gambar 1.



Gambar 1. Tahapan Penelitian

Dalam tahap pengumpulan data, digunakan sebuah dataset berbahasa Indonesia yang terdiri dari 1516 data komentar yang diambil dari platform media sosial. Dataset ini memiliki lima kolom utama, yaitu nama_akun, Text_cleaning yang berisikan komentar, serta ada kolom label sentimen yang mencerminkan kategori positif, negatif, dan netral. Kumpulan data ini dibuat pada tahun 2023 dan diperoleh dari situs website Kaggle, dengan informasi lebih lanjut yang dapat diakses melalui tautan URL berikut: <https://www.kaggle.com/code/billycemerson/analisis-sentimen-intensif-mobil-listrik-svm/notebook>

Preprocessing data ini dilakukan dengan memastikan bahwa data yang digunakan dalam analisis sentimen menjadi relevan. Di samping itu data juga tidak dapat mengolah data mentah, sehingga untuk mempermudah dan dimengerti menggunakan Bahasa pemrograman python. Melakukan data cleaning dan penampilan data dalam bentuk diagram lingkaran berisikan persentase sentimen positif, sentimen negatif, dan sentimen netral.

Proses penganan data yang tidak lengkap untuk memperbaiki kualitas data dengan menghilangkan duplikasi dan memperbaiki kesalahan, sehingga data yang dihasilkan menjadi lebih akurat dan dapat dihadapi analisis sentimen pada data yang disebut dengan data cleaning (Putri, 2024). Wordcloud dimanfaatkan sebagai media visual yang memudahkan dalam mengenali kata-kata yang paling sering muncul dalam kumpulan komentar dari media sosial. Visualisasi ini menyajikan kata-kata dengan ukuran yang bervariasi, di mana semakin sering suatu kata muncul dalam data, maka semakin besar pula tampilannya di Wordcloud. Dengan cara ini, pengguna dapat dengan cepat menangkap kesan umum atau nuansa sentimen yang dominan dalam data teks yang sedang dianalisis (Pradana, 2020). Term Frequency-Inverse Document Frequency atau yang biasa dikenal dengan TF-IDF adalah salah satu teknik populer dalam analisis teks. Metode ini digunakan untuk mengukur seberapa penting sebuah kata dalam suatu dokumen dibandingkan dengan kumpulan dokumen lainnya. Teknik ini bekerja dengan mempertimbangkan dua hal, yaitu seberapa sering sebuah kata muncul dalam satu dokumen tertentu (yang disebut Term frequency), serta membandingkan frekuensi kemunculan kata tersebut dengan seluruh kumpulan dokumen yang ada (melalui Inverse Document Frequency). Dengan menggabungkan kedua aspek ini, TF-IDF mampu memberikan penilaian yang lebih akurat terhadap kata-kata yang signifikan, sehingga sangat berguna dalam proses ekstraksi fitur yang relevan saat melakukan analisis atau pemodelan data berbasis teks (Mahendra et al., 2023).

Penelitian ini menggunakan algoritma klasifikasi Support Vector Machine (SVM) untuk menganalisis serta membandingkan kinerja metode dalam mengklasifikasikan sentimen. SVM sendiri adalah salah satu algoritma pembelajaran mesin yang bekerja dengan mencari *hyperplane* terbaik untuk memisahkan data ke dalam kategori sentimen positif, negatif, dan netral. Secara umum, ada dua *hyperplane* yang digunakan linear dan non-linear. Jika data dapat dipisahkan secara sempurna menggunakan *hyperplane* linear, maka metode tersebut di sebut sebagai SVM linear (Rahman et al., 2021). Namun, jika tidak dapat dipisahkan secara sempurna maka diperlukan pendekatan dengan teknik kernel untuk mentransformasikan data ke dalam dimensi fitur yang lebih tinggi agar pemisahan linear menjadi memungkinkan pembentukan *hyperplane* linier (Rabbani et al., 2023). Dalam Penelitian ini digunakan kernel polynomial, sigmoid dan Radial Basis Function (RBF) yang penjelasan dan rumusnya akan dijabarkan pada bagian berikutnya: Kernel Radial Basis Function (RBF) bekerja dengan memanfaatkan fungsi Gaussian yang berperan untuk menilai tingkat kemiripan antara dua vector input dalam ruang fitur, dengan cara mempertimbangkan jarak non-linear, sehingga mampu menangkap pola-pola kompleks dalam data yang tidak dapat dipisahkan secara linear. Fungsi kernel RBF secara matematis dinyatakan dalam Persamaan 1.

$$k(x_i, x_j) = e^{-\gamma |x_i - x_j|^2} \quad (1)$$

Kernel polynomial memodelkan hubungan non-linear dalam ruang asli dengan menggunakan fungsi polynomial, memungkinkan pemisahan data yang tidak dapat dipisahkan secara linear. Rumus kernel polynomial ditunjukkan dalam Persamaan 2.

$$k(x_i, x_j) = (x_i \cdot x_j + c)^d \quad (2)$$

Kernel sigmoid menggunakan fungsi aktivasi sigmoid (tanh) untuk mengukur hubungan antara dua vector ke dalam ruang fitur dan menyerupai cara kerja Neural Network tiruan dalam klasifikasi data. Kernel ini dinyatakan dalam Persamaan 3.

$$k(x_i, x_j) = \tanh(\alpha x_i^T x_j) + c \quad (3)$$

Yang dimana x_i, x_j adalah vector fitur dari data, α adalah parameter scalar untuk mengontrol skala inner product, dan c adalah bias atau konstanta. Fungsi dari aktivasi tanh digunakan dalam jaringan Neural Network yang mengukur hubungan antara dua vector berdasarkan aktivasi non-linear.

Salah satu teknik yang cukup populer dalam analisis data dan machine learning, terutama karena kemampuannya menangani data dalam jumlah besar secara efisien yaitu Logistic Regression. Metode ini digunakan untuk memahami hubungan anatara sejumlah variabel bebas (independent) dengan hasil yang bersifat biner, yaitu hanya memiliki dua kemungkinan. Cara kerjanya adalah dengan memperkirakan seberapa besar kemungkinan suatu peristiwa atau kategori tertentu akan terjadi berdasarkan nilai dari variabel-variabel tersebut. Logistic Regression juga dikenal sebagai pendekatan yang handal dan efektif dalam menilai sejauh mana masing-masing variabel independent memberikan pengaruh terhadap variabel target atau variabel terikat dengan cara mengukur kontribusi unik dari setiap faktor yang dianalisis (Ainunnisa & Sulastri, 2023).

Metode pembelajaran mesin yang menggabungkan beberapa model klasifikasi untuk meningkatkan akurasi prediksi disebut juga dengan Teknik ensemble stacking (Rafrastara et al., 2023). Dalam teknik ini, beberapa algoritma dasar (base models) dilatih secara independent pada dataset yang sama (Majid et al., 2025). Prediksi yang dihasilkan oleh masing-masing model dalam sistem ini selanjutnya dimanfaatkan sebagai input bagi sebuah meta model. Meta model ini berperan penting dalam tahap akhir, karena bertugas untuk mengolah dan menggabungkan seluruh hasil prediksi dari model-model sebelumnya, sehingga dapat menghasilkan satu keputusan akhir yang lebih akurat (Rafrastara et al., 2023). Dalam implementasi teknik ensemble stacking ini, Support Vector Machine (SVM) dengan berbagai konfigurasi kernel akan berperan sebagai base learner. Secara spesifik, tiga model SVM akan dilatih menggunakan kernel sigmoid, kernel polynomial, dan kernel RBF. Masing - masing base learner ini akan menghasilkan prediksi probabilitas kelas untuk setiap sampel data. Selanjutnya, kumpulan prediksi probabilitas dari ketiga base learner SVM tersebut akan digabungkan dan menjadi input fitur bagi meta learner. Dalam skenario ini, Logistic Regression dipilih sebagai meta learner. Meta learner ini akan dilatih untuk mempelajari pola kompleks dari probabilitas yang dihasilkan oleh setiap base learner, mengidentifikasi dari prediksi tersebut untuk menghasilkan prediksi sentimen akhir yang lebih akurat. Dengan demikian, Logistic Regression sebagai meta learner menggunakan masing-masing SVM base learner untuk mengambil keputusan akhir yang lebih akurat.

HASIL DAN PEMBAHASAN

Penelitian ini menganalisis 1.516 komentar dari media sosial YouTube yang membahas isu mobil listrik. Data yang dikumpulkan mencakup ID komentar, nama akun, tanggal unggah, isi komentar, serta label sentimen yang telah diklasifikasikan ke dalam kategori positif, negatif, dan netral. Tampilan data awal tersebut ditunjukkan pada Gambar 2. Untuk memastikan akurasi klasifikasi sentimen, data terlebih dahulu diproses melalui tahap pembersihan (data cleaning) sebagaimana terlihat pada Gambar 3. Proses ini menghasilkan kolom baru seperti `text_cleaning` dan `sentimen encoding`, dengan kode 1 untuk positif, -1 untuk negatif, dan 0 untuk netral.

Distribusi sentimen dari keseluruhan data divisualisasikan pada Gambar 4. Hasilnya menunjukkan bahwa sentimen negatif mendominasi dengan persentase 57,2%, disusul oleh sentimen positif sebesar 33,2%, dan netral sebesar 9,5%. Temuan ini mengindikasikan bahwa persepsi masyarakat terhadap mobil listrik masih cenderung kritis. Untuk memahami lebih dalam karakteristik tiap kategori sentimen, digunakan visualisasi Word Cloud (netral, negatif dan positif) seperti terlihat pada Gambar 5. Word Cloud untuk sentimen negatif didominasi oleh kata-kata seperti banyak, kendaraan, dan beli, yang mencerminkan kekhawatiran masyarakat terhadap harga, infrastruktur, dan kesiapan teknologi. Sebaliknya, pada sentimen positif, kata-kata yang menonjol adalah harga, mahal,

dan mau, yang mengisyaratkan adanya minat yang besar meskipun disertai kekhawatiran terhadap aspek ekonomi. Sementara itu, kata dominan dalam komentar netral antara lain subsidi dan laku, yang menunjukkan bahwa komentar bersifat informatif tanpa muatan emosional kuat. Langkah berikutnya adalah perhitungan bobot Term Frequency–Inverse Document Frequency (TF-IDF), yang divisualisasikan dalam Gambar 6 untuk menunjukkan signifikansi kata-kata dalam dokumen.

index	id_komentar	nama_akun	tanggal	text_cleaning	sentimen;
0	Ugzbl5eyrly3-gdUUJ4AaABAg	Sqn Ldr	2023-08-06 12:54:49+00:00	saran sih bikin harga ionic sama kayak brio insya alloh laris manis	positif;
1	UgzEDUv3OTv943p8p4AaABAg	lushman ace	2023-08-04 12:16:23+00:00	problem subsidi kualitas diturunin harga dinaikin usaha gitu cari cuan subsidi sebab inflasi paling gede	negatif;
2	UgwaJqu6JMF4E2CsVV4AaABAg	Fath Al-Ayyubi	2023-08-04 10:17:57+00:00	baik kualitas kembang dulu baik kualitas motor motor pabrikan jepang	positif;
3	UgyYcCMR1fKwuOjZY14AaABAg	yp office	2023-08-04 08:29:54+00:00	model jelek kualitas bunuk harga mahal croot	negatif;
4	UgxKAclUAWZOOK8es-x4AaABAg	Lembur Kuring	2023-08-04 07:55:37+00:00	syarat ngaco woy anak muda blom punya rumah blom jd umkm bukan serta kur dapet ngaco sia deui sia deui nu narima subsidi teh ari aing iraha	negatif;
5	Ugx-zVY4ktD7JNUB8xV4AaABAg	Syarif Airlangga	2023-08-04 06:58:17+00:00	harga motor mahal masa harga mirip motor beat kualitas bagai langit bumi	positif;
6	Ugz5Wqya394dPhoq14AaABAg	BajuMax	2023-08-04 06:31:56+00:00	mol keren yah berita plus padahal mol lokal merk batrenya juta garansi tahun anggop batre awet tahun rakyat indonesia per tahun sekali keluar juta buat beli batre baru o saran uang subsidi keluar buat batre perintah punya wewenang pertamina kelola contoh gas kg gas kg produk baru misal batre v batre v batre v cabut bayar pasang gas pnp pom sistem swap batre banyak titik swap point indonesia luas	negatif;
7	Ugyy3luBOOHQspWjBIR4AaABAg	Putut Panweto	2023-08-04 01:04:18+00:00	proses kenal produk baru butuh waktu ganti kendara btm jadi kendara butuh waktu tehnologi baru sabar	negatif;
8	Ugww-4D-u_0UFxhRTp4AaABAg	Heru Prasetyo	2023-08-03 23:58:55+00:00	subsidi tepat sasaran	netral;
9	Ugx2K37YO1fch9eEM4d4AaABAg	jonan kikk ass hole	2023-08-03 11:25:57+00:00	adil rata terima subsidi jangan jangan pajak pph msh subsidi perintah khianat pancasila sila lima adil seluruh rakyat indonesia mng mau rata bagus usa subsidi sama sekali untung subsidi rata umum alih kendara btm lebih cepet efek apa kurang konsumsi subsidi btm kurang polusi tingkat minat pihak swasta di ikut partisipasi investasi ekosistem motor	negatif;
10	UgzBFjd9vzYUd7E2bGN4AaABAg	Rendy Ramadhan	2023-08-03 07:20:28+00:00	tepat sasaran donk syarat belum tepat subsidi kurang besar jd warga kurang mampu pikir beli anda subsidi jt mungkin warga kurang mampu lebih berani beli	negatif;

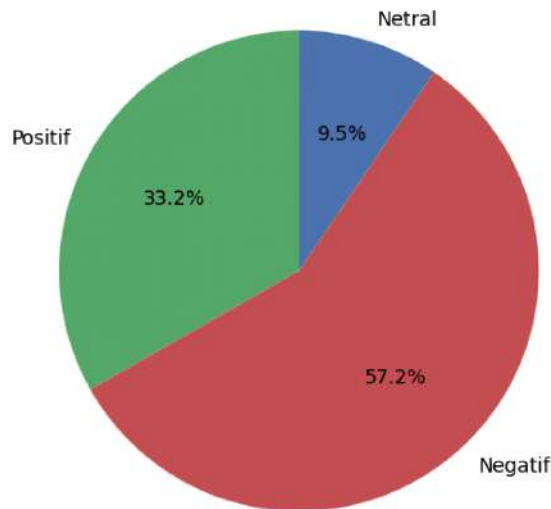
Gambar 2. Tampilan Data Awal

index	text_cleaning	sentimen;	sentimen_encoded
0	saran sih bikin harga ionic sama kayak brio insya alloh laris manis	positif;	1.0
1	problem subsidi kualitas diturunin harga dinaikin usaha gitu cari cuan subsidi sebab inflasi paling gede	negatif;	-1.0
2	baik kualitas kembang dulu baik kualitas motor motor pabrikan jepang	positif;	1.0
3	model jelek kualitas bunuk harga mahal croot	negatif;	-1.0
4	syarat ngaco woy anak muda blom punya rumah blom jd umkm bukan serta kur dapet ngaco sia deui sia deui nu narima subsidi teh ari aing iraha	negatif;	-1.0
5	harga motor mahal masa harga mirip motor beat kualitas bagai langit bumi	positif;	1.0
6	mol keren yah berita plus padahal mol lokal merk batrenya juta garansi tahun anggop batre awet tahun rakyat indonesia per tahun sekali keluar juta buat beli batre baru o saran uang subsidi keluar buat batre perintah punya wewenang pertamina kelola contoh gas kg gas kg produk baru misal batre v batre v batre v cabut bayar pasang gas pnp pom sistem swap batre banyak titik swap point indonesia luas	negatif;	-1.0
7	proses kenal produk baru butuh waktu ganti kendara btm jadi kendara butuh waktu tehnologi baru sabar	negatif;	-1.0
8	subsidi tepat sasaran	netral;	0.0
9	adil rata terima subsidi jangan jangan pajak pph msh subsidi perintah khianat pancasila sila lima adil seluruh rakyat indonesia mng mau rata bagus usa subsidi sama sekali untung subsidi rata umum alih kendara btm lebih cepet efek apa kurang konsumsi subsidi btm kurang polusi tingkat minat pihak swasta di ikut partisipasi investasi ekosistem motor	negatif;	-1.0
10	tepat sasaran donk syarat belum tepat subsidi kurang besar jd warga kurang mampu pikir beli anda subsidi jt mungkin warga kurang mampu lebih berani beli	negatif;	-1.0
11	mungkin subsidi baik arah seragam sekolah bila lama guna seragam sekolah hapus negara ri mungkin subsidi seragam sekolah minimal rasa marah jengkel kpd saudara bangsa n tanah air terus rusak bangsa n negara n dg laku korupsi macam terus ma af jk salah	negatif;	-1.0
12	kampung sekarang banyak banget booi sama oewek pake motor	positif;	1.0
13	banyak kendaraan buat tiangkok rekam bakar letup amerika rana masalah baterai kendaraan bahaya	negatif;	-1.0
14	harga terlalu mahal	positif;	1.0
15	tahan th	negatif;	-1.0
16	iklan kudu kembang	netral;	0.0
17	kendaraan plg rawan bakar mudah padam	negatif;	-1.0
18	buat hari sih sangat guna jalan jauh sih mikir pake charging station keluar kota saudara kota buat charge banyak stasiun charger nya jg kaya bensin tinggal isi beberapa menit langsung jalan butuh waktu sebentar kritis banget mungkin nyolokin minimarket or rumah orang	negatif;	-1.0
19	bapak luhut panjaitan yth baik pakai dulu dinas masyarakat lihat hasil jangan dijsjeli subsidi laku pakai logika maaf masyarakat beli terimakasih	positif;	1.0

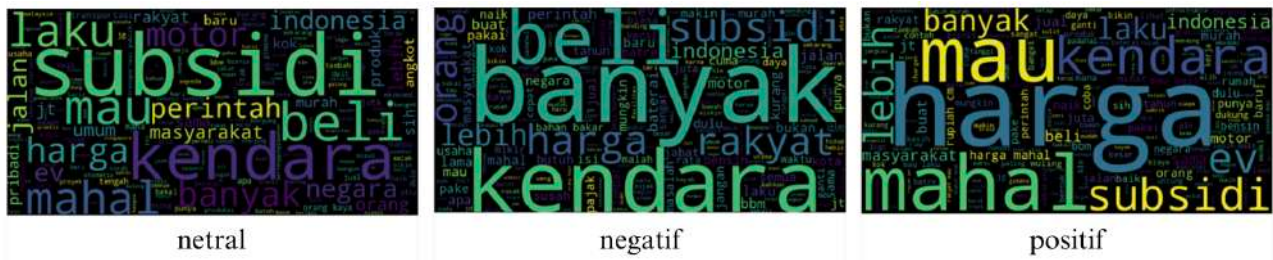
Gambar 3. Hasil dari Proses Data Cleaning

Eksperimen dilakukan dengan membagi data ke dalam tiga skenario proporsi pelatihan dan pengujian, yaitu 60:40, 70:30, dan 80:20 sebagaimana disajikan dalam Tabel 1. Proporsi 70% data latih dan 30% data uji dipilih sebagai skenario utama karena memberikan hasil performa terbaik. Model pertama yang diuji adalah Support Vector Machine (SVM) dengan kernel Radial Basis Function (RBF). Berdasarkan Gambar 7, model ini menghasilkan precision tertinggi pada kelas negatif sebesar 84,6% dan recall sebesar 81,7%, namun rendah pada kelas netral dengan f1-score hanya 24%. Akurasi keseluruhan mencapai 74,9%. Model SVM dengan kernel polynomial (Gambar 8) menunjukkan peningkatan akurasi menjadi 76%, dengan f1-score tertinggi tetap berada pada kelas

negatif. Akan tetapi, performa pada kelas netral masih belum memuaskan dengan f1-score hanya 8,3%.



Gambar 4. Hasil dari persentase sentimen



Gambar 5. Tampilan Word Cloud

index	abis	ac	ada	adil	agya	ahli	air	airev	ajah	ajar	akal	akhir	aki	alibat	akses	alam	alami	alas	alat	alhamdulillah
0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.33022001792643596		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0		0.0	0.0	0.12763569128903128		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
21	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
22	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar 6. Hasil dari TF-IDF

Model SVM dengan kernel sigmoid (Gambar 9) mencatatkan akurasi 68,3%, dengan hasil recall tertinggi pada kelas negatif sebesar 96,9%. Namun, kelas netral gagal diklasifikasikan, yang tercermin dari nilai f1-score sebesar 0%. Sementara itu, model Logistic Regression (Gambar 10) mencatat akurasi sebesar 75,6%, dengan precision tinggi pada kelas netral (100%) tetapi recall sangat rendah (5,7%), menunjukkan bahwa model ini hanya mampu mengenali sebagian kecil data netral secara sangat yakin. Kombinasi dari ketiga model SVM dalam bentuk ensemble stacking dengan Logistic Regression sebagai meta learner (Gambar 11) menunjukkan performa terbaik dengan akurasi

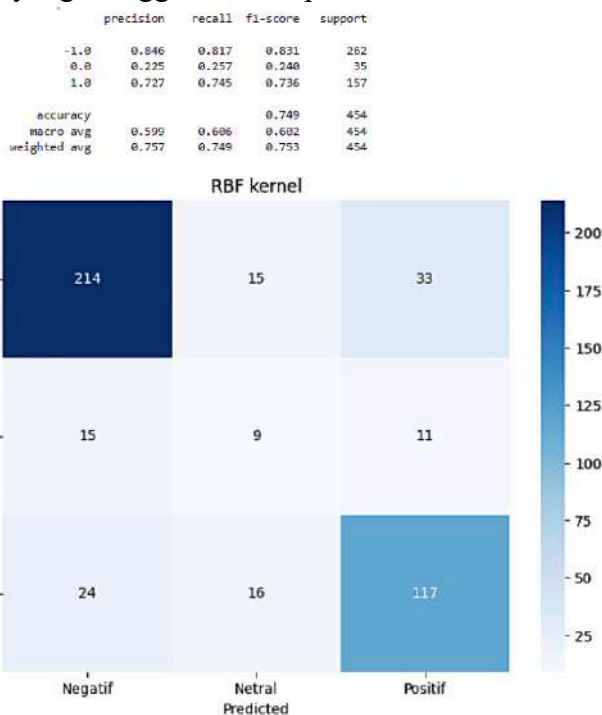
mencapai 76,21%. Nilai precision dan recall tertinggi kembali ditemukan pada kelas negatif (79,4% dan 87%), sedangkan performa untuk kelas netral masih tergolong lemah (f1-score sebesar 5,4%).

Pada setelah hasil TF-IDF selanjutnya, akan dilakukan eksperimen melalui tiga skenario yang berbeda. Skenario pertama akan membagi data menjadi 60% untuk latih dan 40% untuk uji. Selanjutnya, skenario kedua akan menggunakan 70% data latih dan 30% data uji. Ketiga pembagian ini merupakan untuk mengevaluasi stabilitas kinerja model data. Melalui serangkaian skenario pengujian ini, diharapkan dapat diidentifikasi model yang paling optimal dalam memberikan prediksi yang akurat. Dalam pengujian model ini, dilakukan skenario 70% data latih dan 30% data uji.

Table 1. Splitting Data

Rasio	Data Latih	Data Uji	Total Data
60%:40%	910	606	
70%:30%	1061	455	1516
80%:20%	1213	303	

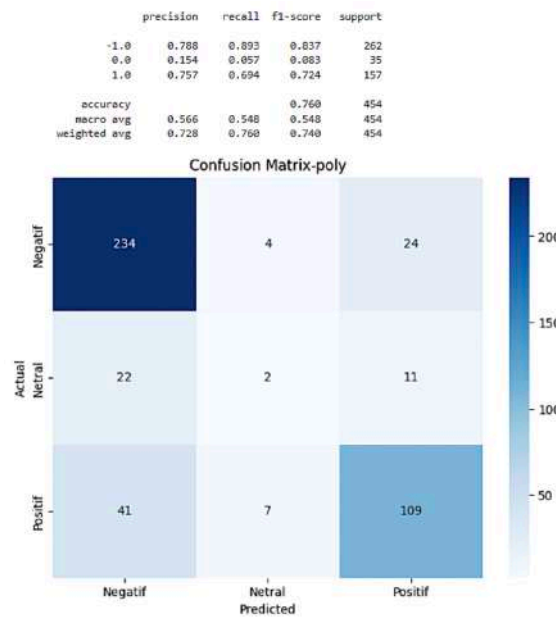
Hasil penerapan algoritma SVM dengan kernel RBF dapat dilihat melalui tampilan Confusion Matrix pada gambar berikut yang menggambarkan performa model dalam mengklasifikasikan data.



Gambar 7. Hasil dari SVM RBF

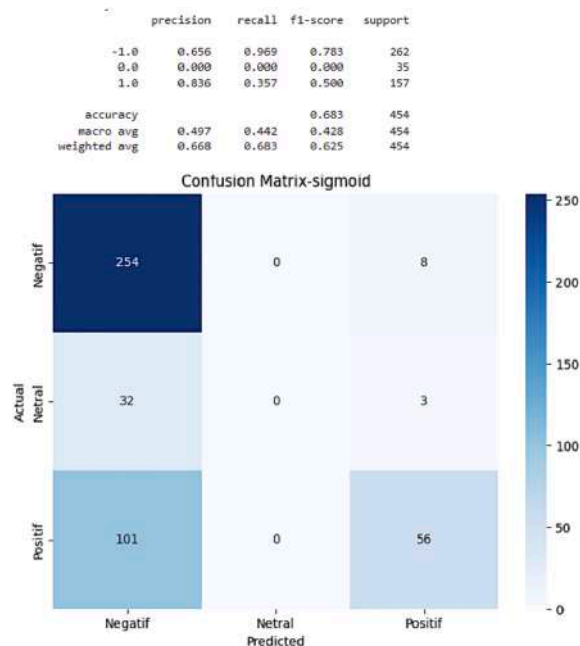
Hasil evaluasi model menunjukkan bahwa precision pada kelas negatif mencapai 84,6%, pada kelas netral sebesar 22,5% dan pada kelas positif sebesar 72,7%. Berdasarkan hasil tersebut, dapat diambil kesimpulannya bahwa sistem mampu mengidentifikasi label dengan cukup baik, dengan akurasi tertinggi terdapat pada kelas negatif sebesar 84,6%. Selain itu, pada hasil recall yang diperoleh untuk kelas negatif sebesar 81,7%, kelas netral 25,7% dan kelas positif 74,5%. Sementara itu, pada hasil f1-score menunjukkan nilai sebesar 83,1% untuk kelas negatif, 24%, untuk kelas netral, dan 73,6% untuk kelas positif. Dari hasil recall dan f1-core, dapat diketahui bahwa kelas negatif memiliki nilai tertinggi, sehingga secara keseluruhan sistem mencapai tingkat akurasi sebesar 74,9%.

Pada hasil selanjutnya, menggunakan penerapan algoritma SVM dengan kernel Polynomial dapat dilihat melalui tampilan Confussion Matrix pada Gambar 8 yang menggambarkan performa model dalam mengklasifikasi data.



Gambar 8. Hasil dari SVM polynomial

Nilai hasil tersebut menunjukkan bahwa sistem mampu mengklasifikasikan skala label dengan benar dan kelas dengan akurasi tertinggi yang didapatkan yaitu kelas negatif 78,8%. Berdasarkan perhitungan, hasil precision pada kelas negatif mencapai 78,8%, pada kelas netral 15,4%, serta kelas positif 75,7%. Sementara untuk recall pada kelas negatif 89,3%, kelas netral 57%, dan kelas positif 69,4%. Namun, pada perhitungan f1-score hasilnya untuk kelas negatif 83,7%, kelas netral 8,3%, dan juga kelas positif 72,4%. Dari hasil recall dan f1-score dapat menemukan hasil tertinggi di peroleh kelas negatif yang dapat demikian hasil akurasi mencapai 76%.

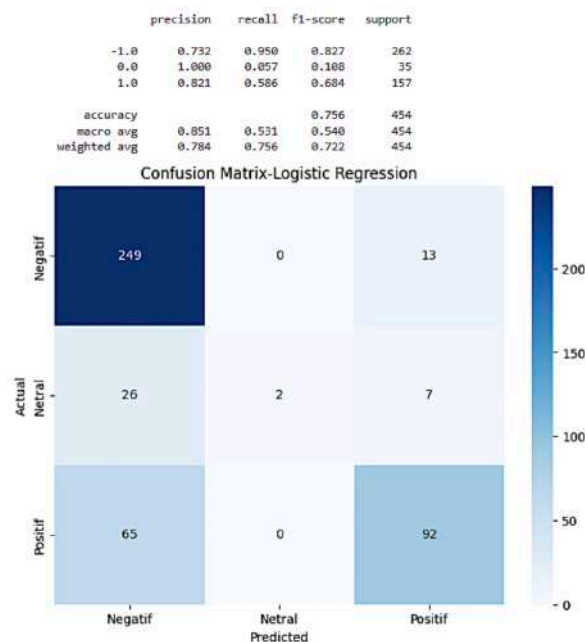


Gambar 9. Hasil dari SVM sigmoid

Hasil penerapan algoritma SVM selanjutnya dengan menggunakan kernel sigmoid dapat diamati melalui Confusion Matrix pada gambar berikut. Tampilan tersebut menunjukkan kinerja model dalam mengklasifikasikan data serta memberikan gambaran mengenai seberapa baik mode membedakan setiap kelas dalam analisis sentimen.

Hasil yang akurat pada precision menunjukkan kelas negatif pada 65,6%, kelas positif dihasilkan 83,6%, sedangkan kelas netral menunjukkan tidak mendapatkan hasil. Dari hasil ini, terbukti bahwa sistem dapat dengan benar mengkategorikan skala label dan tingkat klasifikasi yang paling tinggi dengan hasil 83,6%. Hasil dari recall di kelas negatif 96,9%, kemudian pada kelas positif mencapai 35,7%, sedangkan pada kelas netral menunjukkan tidak mendapatkan hasil. Hasil yang didapatkan dari f1-score pada kelas negatif mencapai 78,3%, pada kelas netral menghasilkan 0%, sedangkan pada kelas positif mendapatkan 50%. Dari hasil recall dan f1-score dapat menemukan hasil tertinggi di peroleh kelas negatif yang dapat demikian hasil akurasi mencapai 68,3%.

Hasil dari penerapan algoritma Logistic Regression dapat di analisis melaui tampilan Confusion Matrix yang terdapat pada Gambar 10.

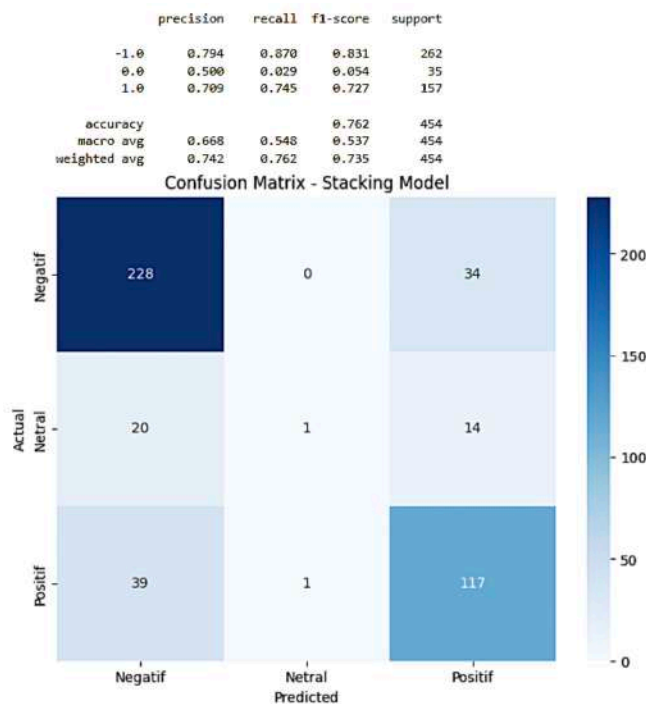


Gambar 10. Hasil dari Logistic Regression

Tingkat precision yang diperoleh menunjukkan bahwa untuk kelas negatif berada di angka persentase 73,2%, kelas netral menunjukkan hasil sempurna sebesar 100%, dan kelas positif mencapai 82,1%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa sistem yang digunakan cukup handal dalam mengelompokkan label sentimen secara tepat, dengan performa terbaik ditunjukkan oleh kelas netral yang berhasil diklasifikasikan secara akurat tanpa kesalahan. Hasil nilai recall pada kelas negatif 95%, pada kelas netral mencapai 5,7%, kemudian pada kelas positif mencapai 58,6%. Hasil yang diperoleh dari f1-score menunjukkan bahwa performa model dalam mengklasifikasikan data untuk kelas negatif adalah 82,7%, untuk kelas netral berada pada angka 10,8%, sedangkan untuk kelas positif menunjukkan nilai sebesar 68,4%. Dari hasil recall dan f1-score dapat menemukan hasil tertinggi di peroleh kelas negatif yang dapat demikian hasil akurasi mencapai 75,6%.

Hasil dari penerapan metode Ensemble Stacking dapat di analisis melaui tampilan Confusion Matrix yang terdapat pada gambar. Berdasarkan hasil yang diperoleh, nilai precision untuk kelas negatif berada pada angka 79,4%, untuk kelas netral menunjukkan nilai sebesar 50%, dan untuk kelas positif menghasilkan nilai precision 70,9%. Dari pencapaian hasil tersebut, dapat disimpulkan sistem telah mampu melakukan klasifikasi label sentimen dengan cukup baik, yang di mana kelas negatif

menunjukkan nilai precision tertinggi dengan persentase sebesar 79,4%. Hasil dari recall pada kelas negatif 87%, pada kelas netral mencapai 2,9%, kemudian pada kelas positif mencapai 74,5%. Hasil yang didapatkan dari f1-score untuk kelas negatif di dapatkan persentase sebesar 83,1%, pada kelas netral mencapai 5,4%, sedangkan untuk kelas positif dengan hasil persentase 72,7%. Dari hasil recall dan f1-score dapat menemukan hasil tertinggi di peroleh kelas negatif yang dapat demikian hasil akurasi mencapai 76,5%.



Gambar 11. Hasil dari Ensemble Stacking

Adapun hasil akurasi dari setiap algoritma SVM RBF, SVM Polinomial, SVM Sigmoid, Logistic Regression, dan teknik ensemble stacking yang dapat di lihat Tabel 2.

Tabel 2. Hasil akurasi dari beberapa model

Model	Akurasi
Logistic Regression	75,55%
SVM RBF	72,69%
SVM Polynomial	75,99%
SVM Sigmoid	68,28%
Ensemble Stacking	76,21%

Hasil evaluasi yang dilakukan menunjukkan bahwa model Ensemble Stacking dalam analisis sentimen, dengan akurasi sebesar 76,21%. Angka ini menunjukkan peningkatan signifikan dibandingkan dengan temuan pada jurnal sebelumnya (Leidiyana, 2023), yang mendapatkan akurasi dibawah 76% untuk ensemble Stacking. Secara lebih lanjut, model SVM dengan kernelnya dan Logistic Regression juga memberikan hasil yang kompetitif, masing-masing dengan akurasi 75,99% dan 75,55%. Di sisi lain, SVM dengan kernel RBF sedikit tertinggal dengan akurasi 72,69%, sementara SVM dengan kernel sigmoid menunjukkan akurasi terendah, yaitu hanya 68,28%.

Model ensemble stacking terbukti memberikan hasil klasifikasi terbaik dalam penelitian ini dengan akurasi sebesar 76,21%, lebih tinggi dibandingkan model SVM Polynomial (75,99%) dan Logistic Regression (75,55%). Hal ini menunjukkan bahwa kombinasi model prediktif dalam

kerangka ensemble memungkinkan peningkatan performa klasifikasi sentimen yang signifikan, terutama ketika menangani data dengan karakteristik kompleks dan tidak seimbang. Teknik ini memanfaatkan kekuatan masing-masing base learner (SVM dengan kernel berbeda) dan menyatukannya melalui meta learner Logistic Regression, sehingga mampu menangkap pola sentimen yang lebih bervariasi. Temuan ini sejalan dengan hasil studi Leidiyana (2023) yang menunjukkan bahwa ensemble stacking meningkatkan akurasi klasifikasi dibandingkan model tunggal. Rafrastara et al. (2023) dan Majid et al. (2025) juga menyatakan bahwa pendekatan stacking memiliki keunggulan dalam meminimalkan bias model tunggal dan memperbaiki ketepatan prediksi, terutama untuk tugas klasifikasi berbasis teks seperti ini.

Kendati begitu, hasil evaluasi menunjukkan bahwa hampir seluruh model masih mengalami keterbatasan dalam mengklasifikasikan komentar bersentimen netral. F1-score untuk kelas netral jauh lebih rendah dibandingkan kelas positif dan negatif. Hal ini mungkin disebabkan oleh rendahnya proporsi data netral dalam dataset serta struktur linguistik komentar netral yang tidak mencerminkan emosi secara eksplisit. Pradana (2020) menegaskan bahwa komentar netral sering kali bersifat faktual, informatif, atau mengandung ambiguitas, sehingga tidak mudah dikenali oleh model berbasis TF-IDF. Selain itu, Rabbani et al. (2023) mencatat bahwa efektivitas SVM sangat tergantung pada pemilihan kernel yang sesuai untuk pola distribusi data. Oleh karena itu, untuk meningkatkan klasifikasi sentimen netral di masa mendatang, pendekatan lanjutan seperti oversampling kelas minoritas, penggunaan word embedding kontekstual seperti BERT, atau pemanfaatan teknik deep learning dapat dipertimbangkan.

Penelitian ini memberikan kontribusi praktis dan teoritis dalam bidang analisis sentimen, khususnya yang berbasis media sosial berbahasa Indonesia. Secara praktis, hasil ini dapat dimanfaatkan oleh industri otomotif dan pengambil kebijakan sebagai alat monitoring opini publik terhadap mobil listrik, sehingga strategi komunikasi, pemasaran, maupun kebijakan insentif dapat disusun secara lebih responsif dan berbasis data. Dominasi sentimen negatif yang ditemukan mengindikasikan pentingnya intervensi kebijakan publik untuk mengatasi resistensi konsumen terhadap adopsi teknologi kendaraan listrik, sebagaimana disorot oleh Puricelli et al. (2025). Secara teoritis, penelitian ini juga memperkaya kajian tentang efektivitas model ensemble stacking dalam konteks analisis teks berbahasa lokal. Mengingat sebagian besar penelitian sebelumnya lebih berfokus pada teks bahasa Inggris, studi ini membuka ruang eksplorasi lanjutan tentang adaptasi teknik pembelajaran mesin terhadap struktur linguistik dan konteks sosial budaya lokal, sebagaimana juga diangkat oleh Daniati & Utama (2023) dan Sunarko et al. (2023).

SIMPULAN

Berdasarkan hasil penelitian, dapat konstatir bahwa penerapan analisis sentimen terhadap wacana mobil listrik di media sosial terbukti efektif apabila menggunakan algoritma Support Vector Machine (SVM) yang dipadu dengan teknik ensemble stacking. Secara empiris, model ensemble stacking menunjukkan kinerja terbaik dengan nilai akurasi sebesar 76,21%, menegaskan keunggulan metode ini dalam menangani kompleksitas dan ketidakseimbangan data teks. Temuan tersebut memiliki implikasi strategis bagi pengembang teknologi dan pemangku kebijakan industri otomotif, terutama dalam merancang kebijakan komunikasi dan strategi pemasaran yang lebih responsif terhadap persepsi publik. Selain memperkaya literatur tentang analisis sentimen pada domain otomotif, studi ini juga membuka peluang aplikatif pada berbagai konteks lain misalnya, evaluasi respons masyarakat terhadap produk baru atau survei opini publik secara umum. Untuk penelitian selanjutnya, disarankan penggunaan korpus data yang lebih besar dan heterogen serta eksplorasi metode berbasis deep learning atau representasi semantik yang lebih maju, guna meningkatkan presisi dan kedalaman analisis sentimen di masa depan.

REFERENSI

- Abdellatif, A., Mubarak, H., Ahmad, S., Ahmed, T., Shafiullah, G. M., Hammoudeh, A., Abdellatef, H., Rahman, M. M., & Gheni, H. M. (2022). Forecasting Photovoltaic power generation with a stacking ensemble model. *Sustainability (Switzerland)*, 14(17), 1–21. <https://doi.org/10.3390/su141711083>
- Ainunnisa, I. R., & Sulastri, S. (2023). Analisis sentimen aplikasi TikTok dengan metode Support Vector Machine (SVM), Logistic Regression dan Naïve Bayes. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 6(3), 423–430. <https://doi.org/10.32493/jtsi.v6i3.31076>
- Arsi, P., & Waluyo, R. (2021). Analisis sentimen wacana pemindahan ibu kota Indonesia menggunakan algoritma support vector machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>
- Daniati, E., & Utama, H. (2023). Analisis sentimen dengan pendekatan ensemble learning dan word embedding pada twitter. *Journal of Information System Management (JOISM)*, 4(2), 125–131. <https://doi.org/10.24076/joism.2023v4i2.973>
- Harahap, F. A. A., Yulandari, M., Asshiddiqi, M. H., & Putri, H. (2024). Skrining fitokimia dan identifikasi senyawa metabolit sekunder tanin secara kromatografi lapis tipis ekstrak etanol daun Afrika (*Vernonia amygdalina* Del.). *Jurnal Kesehatan Unggul Gemilang*, 8(1), 7–15.
- Himawan, H., Kaswidjanti, W., Sentimen, A., Sosial, M., & Based, L. (2018). Metode Lexicon Based dan Support Vector Machine untuk menganalisis sentimen pada media sosial sebagai rekomendasi oleh-oleh favorit. *Seminar Nasional Informatika, 2018*(November), 235–244.
- Joses, S., Quinevera, S., Mardianto, R., Yulvida, D., & Shiddiqi, A. M. (2024). Pendekatan metode ensemble learning untuk deteksi serangan DDoS menggunakan soft voting classifier. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 10(1), 79. <https://doi.org/10.26418/jp.v10i1.73241>
- Lazzarini, R., Tianfield, H., & Charissis, V. (2023). A stacking ensemble of deep learning models for IoT intrusion detection. *Knowledge-Based Systems*, 279, 110941. <https://doi.org/10.1016/j.knosys.2023.110941>
- Leidiyana, H. (2023). Ensemble stacking dalam analisa sentimen reaksi veteran militer AS terhadap pengambilalihan Afghanistan oleh Taliban. *INTI Nusa Mandiri*, 18(1), 23–28. <https://doi.org/10.33480/inti.v18i1.4175>
- Mahendra, M. H., Murdiansyah, D. T., & Lhaksmana, K. M. (2023). Analisis sentimen tweet COVID-19 menggunakan K-Nearest Neighbors dengan TF-IDF dan ekstraksi fitur CountVectorizer. *DIKE : Jurnal Ilmu Multidisiplin*, 1(2), 37–43. <https://doi.org/10.69688/dike.v1i2.35>
- Majid, N. K., Supriyanto, C., & Marjuni, A. (2025). Peningkatan keberagaman data untuk klasifikasi penyakit diabetes berbasis stacking ensemble learning. *Jurnal Pengembangan Informatika dan Teknologi (JPIT)*, 10(1), 1–10. <https://doi.org/10.30591/jpit.v10i1.7375>
- Momon, M., & Astuti, D. (2020). Strategi penurunan emisi gas buang kendaraan di Kota Padang. *Jurnal Kebijakan Pembangunan*, 15(1), 1–10. <https://doi.org/10.47441/jkp.v15i1.34>
- Pradana, M. G. (2020). Penggunaan fitur wordcloud dan document term matrix dalam text mining. *Jurnal Ilmiah Informatika*, 8(01), 38-43.
- Puricelli, S., Eid, M., Casadei, S., Dolci, G., Lixi, S., Oever, V. D., Rigamonti, L., & Grosso, M. (2025). Life cycle assessment of electric vehicle technology: Environmental and policy implications. *Journal of Cleaner Production*, 505, 145347. <https://doi.org/10.1016/j.jclepro.2025.145347>
- Putri, A. (2024). Pentingnya data cleaning sebelum visualisasi: Teknik dan tips. *Teknologipintar.org*, 4(5), 2024–2025.
- Rabbani, S., Safitri, D., Rahmadhani, N., Sani, A. A. F., & Anam, M. K. (2023). P Perbandingan evaluasi kernel SVM untuk klasifikasi sentimen dalam analisis kenaikan harga BBM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 153–160.

<https://doi.org/10.57152/malcom.v3i2.897>

- Rafrastara, F. A., Supriyanto, C., Paramita, C., & Astuti, Y. P. (2023). Deteksi malware menggunakan metode stacking berbasis ensemble. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(1), 11–16. <https://doi.org/10.30591/jpit.v8i1.4606>
- Rahman, O. H., Abdillah, G., & Komarudin, A. (2021). Klasifikasi ujaran kebencian pada media sosial Twitter menggunakan Support Vector Machine. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1), 17–23. <https://doi.org/10.29207/resti.v5i1.2700>
- Satria, R. D., Suarjaya, I. M. A. D., & Pratama, I. P. A. E. (2022). Sentimen analisis antusias masyarakat terhadap sampah plastik dengan menggunakan metode Support Vector Machine (SVM). *JITTER: Jurnal Ilmiah Teknologi dan Komputer*, 3(1), 880–893.
- Sunarko, B., Hasanah, U., Hidayat, S., Muhammad, N., Ardiansyah, M. I., Ananda, B. P., Hakiki, M. K., & Baroroh, L. T. (2023). Penerapan stacking ensemble learning untuk klasifikasi efek kesehatan akibat pencemaran udara. *Edu Komputika Journal*, 10(1), 55–63. <https://doi.org/10.15294/edukomputika.v10i1.72080>
- Yoon, T., & Kang, D. (2023). Multi-modal stacking ensemble for the diagnosis of cardiovascular diseases. *Journal of Personalized Medicine*, 13(2). <https://doi.org/10.3390/jpm13020373>