

Analisis Komparatif Algoritma *Supervised Learning* Untuk Klasifikasi Sentimen *Multiclass* Terhadap Tren Kendaraan Listrik

Verra Budhi Lestari¹⁾; Sestri Novia Rizki²⁾; Vani Maharani Nasution³⁾

^{1,2)}Program Studi Sistem Informasi, Universitas Media Nusantara Citra Jakarta

³⁾Program Studi Rekayasa Perangkat Lunak, Universitas Jakarta Internasional Jakarta

¹⁾vera.budhi@mncu.ac.id; noviasestri@gmail.com ²⁾; vanimaharainasution@gmail.com ³⁾

ABSTRACT

This research is motivated by the high complexity of public opinion regarding electric vehicle (EV) trends, which can no longer be adequately represented through binary classification; however, a gap remains in the literature regarding the most efficient multiclass classification models within this domain. The study aims to conduct a comparative analysis of Support Vector Machine (SVM), Logistic Regression (LR), Multinomial Naive Bayes (MNB), and K-Nearest Neighbor (KNN) to determine the best model based on accuracy, precision, recall, and computational efficiency. Data consisting of 1,517 textual public opinions from social media were processed through stages including data cleaning, tokenization, stopword removal, and TF-IDF feature extraction. The results indicate that SVM achieved the best performance with an accuracy of 0.781 and an F1-score of 0.595, reflecting model stability and a good balance between precision and recall. Logistic Regression demonstrated superior precision (0.843) but lower recall, while MNB showed good computational efficiency despite moderate performance. Conversely, KNN yielded the lowest performance due to limitations in handling high-dimensional and sparse data. Further analysis reveals that all models struggled with the neutral class, indicating data imbalance and class similarity. This study contributes to the limited literature on multiclass sentiment evaluation in the EV domain and provides strategic insights into the trade-offs between model complexity, efficiency, and performance. These findings serve as a foundation for developing effective sentiment analysis systems to support decision-making related to electric vehicle trends.

Keywords: *Sentiment Analysis, Naive Bayes, Support Vector Machine, Logistic Regression, Electric Vehicle*

I. PENDAHULUAN

Perkembangan kendaraan listrik (*electric vehicle/EV*) dalam beberapa tahun terakhir menunjukkan peningkatan yang signifikan sebagai bagian dari upaya global dalam mengurangi emisi karbon dan ketergantungan terhadap bahan bakar fosil [1][2]. Seiring dengan meningkatnya adopsi EV, opini publik terhadap teknologi ini juga berkembang pesat, terutama melalui media sosial yang menjadi sumber utama ekspresi persepsi masyarakat. Analisis terhadap opini tersebut menjadi penting karena dapat memberikan wawasan strategis bagi pengambil kebijakan dan pelaku industri dalam memahami tingkat penerimaan masyarakat terhadap kendaraan listrik [3].

Analisis sentimen merupakan salah satu teknik dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam bentuk teks. Dalam beberapa tahun terakhir, penelitian di bidang ini berkembang pesat dengan memanfaatkan berbagai pendekatan *machine*

learning dan *deep learning* untuk meningkatkan akurasi dan kualitas prediksi. Selain itu, kemajuan teknologi NLP, termasuk penggunaan model berbasis transformer, semakin memperkuat peran analisis sentimen dalam memahami opini publik secara lebih mendalam [4].

Sejumlah penelitian sebelumnya telah mengkaji analisis sentimen pada topik kendaraan listrik menggunakan berbagai algoritma, seperti *Support Vector Machine* (SVM), *Naive Bayes*, dan *Random Forest*. Hasil penelitian menunjukkan bahwa SVM sering memberikan performa yang unggul dalam klasifikasi teks dibandingkan metode lainnya [5][6].

Namun demikian, terdapat kesenjangan penelitian (*research gap*) yang nyata dalam literatur saat ini, yaitu sebagian besar penelitian tersebut masih berfokus pada klasifikasi sentimen biner, yaitu hanya membedakan antara sentimen positif dan negatif [5][6]. Pendekatan ini dinilai belum mampu

seungguhnya merepresentasikan kompleksitas opini publik, karena dalam praktiknya terdapat pula sentimen netral yang memiliki peran penting dalam menggambarkan persepsi masyarakat secara lebih objektif. Selain itu, penelitian yang secara khusus membandingkan beberapa algoritma *supervised learning* dalam konteks klasifikasi sentimen *multiclass* pada topik kendaraan listrik masih relatif terbatas.

Dalam melakukan analisis sentimen, pemilihan algoritma klasifikasi menjadi faktor krusial yang memengaruhi performa model. Algoritma seperti *Support Vector Machine* (SVM), *Multinomial Naive Bayes* (MNB), *Logistic Regression* (LR), dan *K-Nearest Neighbor* (KNN) memiliki karakteristik yang berbeda, terutama dalam menangani data teks yang umumnya berdimensi tinggi dan bersifat *sparse*. Oleh karena itu, diperlukan analisis komparatif yang mendalam untuk mengidentifikasi model yang paling optimal dalam menjaga keseimbangan performa antar kelas pada kasus *multiclass* yang memiliki tingkat kompleksitas lebih tinggi.

Selain itu, tahapan pra-pemrosesan data seperti *case folding*, pembersihan karakter non-alfabetik (*cleansing*), tokenisasi, *filtering*, dan *stemming* berperan penting dalam meningkatkan kualitas data [7]. Representasi fitur menggunakan TF-IDF juga menjadi pendekatan yang umum digunakan karena mampu merepresentasikan bobot kata secara efektif dalam dokumen teks [8].

Evaluasi kinerja model dilakukan menggunakan *confusion matrix* dengan metrik akurasi, presisi, *recall*, dan *F1-score*, yang memberikan gambaran menyeluruh terhadap performa model. Pendekatan evaluasi ini penting untuk mengidentifikasi keseimbangan performa antar kelas, terutama dalam kasus klasifikasi *multiclass* yang memiliki tingkat kompleksitas lebih tinggi [9].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk melakukan analisis komparatif terhadap beberapa algoritma *supervised learning* dalam klasifikasi sentimen *multiclass* terhadap tren kendaraan Listrik, kelas dalam penelitian yang akan dilakukan terdiri atas sentimen positif, netral, dan negatif. Dengan memasukkan kelas netral sebagai bagian dari klasifikasi, penelitian ini diharapkan dapat memberikan representasi sentimen yang lebih komprehensif. Hasil penelitian ini diharapkan dapat menjadi dasar dalam pengembangan sistem analisis sentimen yang efektif dan efisien guna mendukung pengambilan

keputusan terkait perkembangan kendaraan listrik di masa depan.

II. TINJAUAN PUSTAKA

Machine learning (ML) merupakan kemampuan sistem untuk belajar dari data dalam menemukan pola tersembunyi serta mengotomasi pembuatan model analitis untuk menyelesaikan berbagai tugas [10]. Teknologi ini memungkinkan mesin belajar dari data dan pengalaman tanpa instruksi langsung manusia [11][12], sehingga data menjadi komponen utama dalam proses pembelajarannya [13]. Selain itu, ML mampu mendukung pengambilan keputusan secara otomatis melalui insight prediktif. Secara umum, ML merupakan bagian dari kecerdasan buatan yang memanfaatkan algoritma pembelajaran, baik terawasi maupun tidak terawasi, untuk menghasilkan prediksi berbasis data [14][15]. Kinerja modelnya sangat dipengaruhi oleh proses *hyperparameter tuning* yang tepat [16].

Dalam ML terdapat beberapa tipe pembelajaran, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. Berbagai penelitian telah memanfaatkan pendekatan ini, salah satunya penggunaan algoritma Random Forest untuk membangun model klasifikasi yang mampu mendeteksi lima jenis penyakit pada kucing (cacingan, Feline Herpesvirus, FLUTD, diabetes melitus, dan ginjal kronis) berdasarkan data gejala klinis [12].

Analisis sentimen merupakan salah satu metode untuk mengklasifikasi objek ke dalam dua kategori sentiment positif dan negatif [17]. Berdasarkan berbagai sumber dan penelitian, *machine learning* memiliki karakteristik sebagai berikut, *Machine learning* memiliki kemampuan untuk beradaptasi secara otomatis terhadap data baru dan perubahan lingkungan, sehingga model dapat terus belajar tanpa intervensi manual yang signifikan. Teknologi ini juga mampu mengelola data dalam jumlah besar secara efisien serta mengidentifikasi pola yang kompleks. Selain itu, *machine learning* dapat melakukan prediksi pada data baru melalui kemampuan generalisasi dari data pelatihan. Kinerja model juga dapat terus ditingkatkan melalui proses optimasi dan pembaruan parameter. Fleksibilitasnya memungkinkan penggunaan berbagai algoritma yang dapat disesuaikan dengan kebutuhan dan karakteristik data. [15].

Hasil penelitian menunjukkan bahwa algoritma stemming Sastrawi memiliki kinerja

terbaik, diikuti oleh Nazief-Adriani, sementara algoritma Arifin-Setiono masih memerlukan perbaikan karena belum optimal dalam mengembalikan kata berimbuhan ke bentuk dasarnya [18]. Eksperimen lain yang menggunakan dataset Aplikasi serta Film dan TV dari Amazon menunjukkan bahwa metode LSIBA-ENN menghasilkan performa terbaik dibandingkan SVM, Naive Bayes, dan ENN berdasarkan metrik presisi, recall, F1-score, dan akurasi. Penggunaan metode pembobotan fitur seperti W2V, TF, TF-IDF, TF-DFS, dan LTF-MICF juga berpengaruh signifikan, di mana LTF-MICF memberikan hasil paling optimal [19]. Selain itu, penelitian pada berbagai dataset populer seperti IMDB, Yelp, Amazon Product, dan ulasan makanan menunjukkan bahwa pemilihan metode representasi fitur sangat berpengaruh terhadap kualitas klasifikasi sentimen, sehingga menjadi faktor penting dalam proses analisis [20][21].

Penelitian menunjukkan bahwa model BiLSTM yang dikombinasikan dengan CNN menghasilkan performa terbaik, meskipun CNN memiliki waktu pelatihan yang lebih cepat tetapi kurang optimal pada kelas dengan jumlah data sedikit [7]. Pada studi lain, algoritma SVM dan KNN terbukti dapat digunakan untuk analisis sentimen, dengan SVM menunjukkan performa lebih baik (akurasi 0,743) dibandingkan KNN (0,64) serta lebih stabil dalam mengklasifikasikan kelas mayoritas, meskipun masih terdapat kesalahan pada kelas minoritas [22]. Permasalahan ketidakseimbangan data juga memengaruhi kinerja model, di mana penerapan teknik seperti SMOTE dan ADASYN mampu meningkatkan keseimbangan hasil klasifikasi, meskipun pada beberapa kasus dapat menurunkan presisi dan F1-score [23]. Selain itu, media sosial seperti TikTok memiliki peran penting dalam membentuk opini dan persepsi pengguna, khususnya Generasi Z, melalui interaksi sosial dan penyajian konten yang menarik, termasuk konten edukatif yang disampaikan secara singkat dan efektif [24][25].

Berdasarkan tinjauan terhadap berbagai literatur di atas, terdapat konsistensi temuan bahwa algoritma *Support Vector Machine* (SVM) cenderung memberikan performa yang lebih stabil dan unggul dibandingkan algoritma berbasis jarak seperti KNN dalam tugas klasifikasi teks. Namun, secara kritis dapat diidentifikasi bahwa performa model klasifikasi sangat dipengaruhi oleh kualitas data teks yang umumnya bersifat *sparse* dan berdimensi tinggi. Hal ini menuntut pemilihan metode ekstraksi fitur yang tepat, seperti TF-IDF, serta

algoritma *stemming* yang akurat seperti Sastrawi untuk mereduksi kompleksitas kata.

Kesenjangan penelitian (*research gap*) yang ditemukan adalah mayoritas studi terdahulu masih berfokus pada klasifikasi biner (positif dan negatif). Penggunaan kelas netral sering kali diabaikan, padahal kelas ini krusial untuk merepresentasikan opini publik secara objektif, khususnya pada isu teknologi seperti kendaraan listrik. Selain itu, masalah ketidakseimbangan data (*imbalanced dataset*) tetap menjadi tantangan besar yang menyebabkan penurunan presisi dan F1-score pada label minoritas. Penelitian ini mensintesis temuan-temuan tersebut dengan melakukan analisis komparatif *multiclass* (positif, negatif, netral) menggunakan empat algoritma berbeda guna menemukan solusi yang paling seimbang antara performa klasifikasi dan efisiensi komputasi.

III. METODE PENELITIAN

Penelitian ini merupakan studi komparatif yang membandingkan kinerja beberapa algoritma *supervised learning* dalam klasifikasi sentimen terhadap opini masyarakat mengenai tren kendaraan listrik. Proses penelitian dilakukan secara sistematis melalui tahapan pengumpulan data, pelabelan sentimen, pra-pemrosesan teks, ekstraksi fitur, pembagian dataset, implementasi model, dan evaluasi performa. Data berupa teks opini publik diklasifikasikan ke dalam tiga kategori, yaitu positif, negatif, dan netral. Secara keseluruhan, penelitian ini bertujuan membandingkan performa algoritma dalam klasifikasi sentimen *multiclass*. Diagram alur penelitian ditampilkan pada Gambar 1 sebagai gambaran umum.

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan dataset publik yang diperoleh dari situs Kaggle. Dataset ini memuat 1.517 opini masyarakat dalam Bahasa Indonesia yang dikumpulkan dari kolom komentar platform media sosial terkait topik kebijakan insentif kendaraan Listrik.

dilihat sebelumnya. Pembagian ini menghasilkan proporsi *Data Training* sebanyak 1.061 sampel, dan *Data Testing* sebanyak 456 sampel. Pemilihan rasio ini bertujuan untuk memberikan porsi data latih yang cukup besar agar model dapat mempelajari variasi bahasa yang beragam dalam 1.517 komentar tersebut.

3.6 Classification

Tahap klasifikasi merupakan inti penelitian, di mana fitur hasil ekstraksi TF-IDF digunakan sebagai input untuk membangun model prediksi. Penelitian ini menerapkan pendekatan *supervised learning* dengan membandingkan empat algoritma klasifikasi yang memiliki karakteristik berbeda dalam menangani data teks.

1. Support Vector Machine

Support Vector Machine (SVM) adalah algoritma klasifikasi yang mampu menangani data linier maupun nonlinier dengan mencari hyperplane optimal yang memaksimalkan jarak antar kelas. Ketika data tidak terpisah secara linier, SVM menggunakan fungsi kernel untuk memetakan data ke ruang berdimensi lebih tinggi. Dengan pendekatan ini, SVM memiliki kemampuan generalisasi yang baik dan efektif digunakan dalam analisis sentimen [9].

Secara prinsip, SVM bekerja dengan cara mencari *hyperplane* atau bidang pemisah optimal yang mampu memaksimalkan jarak (*margin*) antar kelas yang berbeda. Untuk data yang dapat dipisahkan secara linier, posisi *hyperplane* tersebut dapat ditentukan melalui Persamaan 4.

$$f(x) = \omega \times x + b \quad (4)$$

Dalam menentukan hyperplane optimal, SVM menggunakan beberapa komponen utama, yaitu x sebagai vektor fitur hasil TF-IDF, w sebagai bobot yang menentukan arah hyperplane, dan b sebagai bias yang mengatur posisi hyperplane terhadap titik asal. Jika data tidak dapat dipisahkan secara linier, SVM memanfaatkan fungsi kernel untuk memetakan data ke ruang berdimensi lebih tinggi sehingga pemisahan kelas dapat dilakukan dengan lebih efektif [9][22].

2. Naive Bayes

Naive Bayes merupakan teknik klasifikasi berbasis probabilistik yang efektif digunakan pada data dengan dimensi fitur yang tinggi. Pada penelitian ini digunakan varian Multinomial Naive Bayes, yang umum diterapkan pada data teks. Model ini bekerja dengan menghitung probabilitas suatu data termasuk ke dalam setiap kelas, kemudian mengklasifikasikannya ke kelas dengan

probabilitas posterior tertinggi. Naive Bayes mengasumsikan bahwa setiap fitur bersifat independen terhadap fitur lainnya dalam suatu kelas, sehingga proses perhitungan menjadi lebih sederhana dan efisien untuk dataset berskala besar [8][25]. Meskipun memiliki asumsi yang sederhana, metode ini tetap banyak digunakan dalam analisis sentimen dan mampu memberikan performa yang kompetitif, khususnya pada data tekstual. Naive Bayes didasarkan pada Teorema Bayes dengan asumsi independensi yang kuat antar fitur dengan menghitung probabilitas (Persamaan 5).

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (5)$$

Perhitungan pada algoritma ini menggunakan komponen probabilistik utama untuk menentukan kelas sentimen. Probabilitas posterior $P(C|X)$ menunjukkan peluang suatu data termasuk ke kelas C berdasarkan fitur X . Nilai ini dipengaruhi oleh probabilitas prior $P(C)$ dan likelihood $P(X|C)$, sedangkan $P(X)$ berfungsi sebagai faktor normalisasi dalam perhitungan probabilitas secara keseluruhan.

3. Logistic Regression

Logistic Regression merupakan metode *supervised machine learning* yang umum digunakan untuk klasifikasi biner. Metode ini merupakan pengembangan regresi linier yang menggunakan fungsi logistik untuk membatasi output pada rentang 0 hingga 1, sehingga dapat merepresentasikan probabilitas suatu kelas secara efektif [8]. Selain itu, model ini tidak memerlukan hubungan linier langsung antara variabel input dan target karena menerapkan transformasi nonlinier pada *odds ratio*. Probabilitas suatu kelas dalam Logistic Regression dapat dinyatakan melalui Persamaan 6.

$$P(y = 1|X) = \frac{1}{1 + e^{-(\omega \cdot x + b)}} \quad (6)$$

Dalam perhitungan Logistic Regression, terdapat beberapa komponen utama yang memengaruhi hasil prediksi. Probabilitas $P(y=1|x)$ menunjukkan peluang suatu data termasuk ke dalam kelas tertentu berdasarkan fitur yang diberikan. Variabel x merupakan vektor fitur hasil TF-IDF, w adalah bobot yang menentukan pengaruh tiap fitur, dan b adalah bias yang mengatur posisi model. Sementara itu, e merupakan konstanta *Euler* yang digunakan dalam fungsi logistik. Dengan mekanisme ini, model dapat menentukan batas

keputusan untuk membedakan kelas sentimen secara efektif.

4. K-Nearest Neighbors

K-Nearest Neighbors (KNN) merupakan algoritma klasifikasi berbasis *instance* yang menentukan kelas suatu data berdasarkan mayoritas kelas dari k tetangga terdekatnya. Metode ini sederhana dan mudah diterapkan, serta sering digunakan sebagai pembandingan dalam analisis sentimen, meskipun memiliki keterbatasan pada waktu komputasi untuk dataset besar [22]. Untuk mengukur kedekatan antar data, penelitian ini menggunakan jarak Euclidean sebagaimana ditunjukkan pada Persamaan 7.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (7)$$

Dalam perhitungan KNN, jarak $d(x,y)$ menunjukkan kedekatan antara data uji x dan data latih y . Nilai x_i dan y_i merepresentasikan bobot fitur ke- i (TF-IDF), sedangkan n adalah jumlah dimensi fitur. Dengan menghitung selisih kuadrat pada setiap dimensi, KNN menentukan posisi data dalam ruang vektor dan mengklasifikasikannya berdasarkan mayoritas tetangga terdekat.

3.7 Model Evaluation

Evaluasi performa model dalam penelitian ini menggunakan *confusion matrix*, yang membandingkan label aktual dengan hasil prediksi model. Evaluasi ini didasarkan pada empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [9].

Akurasi mengukur sejauh mana model mampu mengklasifikasikan seluruh data secara benar (Persamaan 8). Presisi menggambarkan tingkat keakuratan antara data yang diprediksi positif dengan data aktual yang benar-benar positif (Persamaan 9). Metrik ini krusial untuk meminimalkan kesalahan prediksi positif semu (*False Positive*). *Recall* mengukur kemampuan model dalam menemukan kembali seluruh informasi kelas tertentu dari total data aktual kelas tersebut (Persamaan 10). Metrik ini digunakan untuk melihat seberapa banyak data yang luput dari klasifikasi (*False Negative*). *F1-Score* merupakan rata-rata harmonik dari presisi dan *recall* (Persamaan 11). Metrik ini memberikan gambaran yang lebih objektif dan seimbang mengenai performa model, terutama pada kondisi dataset yang memiliki ketidakseimbangan jumlah data antar kelas (*imbalanced data*).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

$$Presisi = \frac{TP}{TP+FP} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \quad (11)$$

IV. HASIL DAN PEMBAHASAN

4.1. Analisis Karakteristik Data

Hasil observasi terhadap 1.517 data menunjukkan adanya ketidakseimbangan kelas (*class imbalance*), di mana sentimen Negatif (869 data) memiliki jumlah hampir dua kali lipat dari sentimen Positif, dan jauh lebih banyak dibandingkan sentimen Netral yang hanya berjumlah 144 data. Hal ini mengindikasikan bahwa diskursus publik di platform tersebut lebih banyak berisi kritik atau kekhawatiran terkait kebijakan insentif mobil listrik.

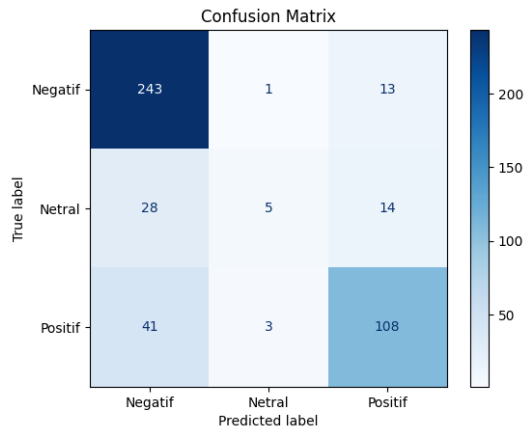
4.2. Hasil Klasifikasi Model

Berdasarkan hasil pengujian terhadap empat algoritma klasifikasi, yaitu Support Vector Machine (SVM), Multinomial Naive Bayes (MNB), Logistic Regression (LR), dan K-Nearest Neighbor (KNN), terlihat adanya perbedaan kinerja yang cukup signifikan sebagaimana ditunjukkan pada Tabel 1. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score untuk menilai performa model secara komprehensif.

Tabel 1. Perbandingan Performa Model

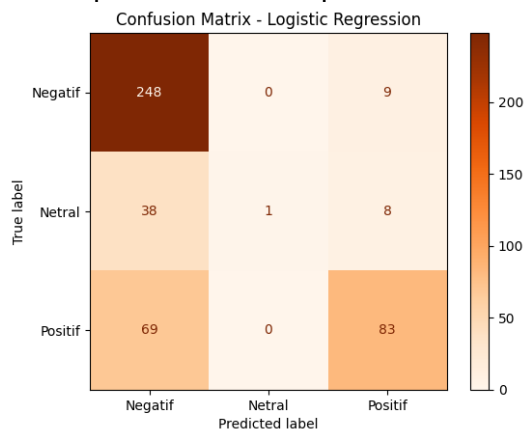
Model	Akurasi	Presisi	Recall	F1-Score
SVM	0.781	0.711	0.587	0.595
MNB	0.649	0.480	0.422	0.395
LR	0.728	0.843	0.511	0.507
KNN	0.404	0.500	0.410	0.356

Hasil pengujian menunjukkan bahwa model SVM memiliki performa terbaik dibandingkan model lainnya, dengan nilai akurasi sebesar 0,781, recall sebesar 0,587, dan F1-score sebesar 0,595. Nilai tersebut menunjukkan bahwa SVM mampu memberikan keseimbangan yang baik antara kemampuan dalam mengidentifikasi data secara benar (recall) dan ketepatan dalam melakukan prediksi (presisi) seperti yang disajikan pada Gambar 4. Dengan demikian, SVM dinilai paling optimal dalam mengklasifikasikan data pada penelitian ini. Keunggulan ini dapat disebabkan oleh kemampuan SVM dalam memaksimalkan margin antar kelas, sehingga model mampu memisahkan data dengan lebih baik, terutama pada data yang memiliki pola kompleks seperti teks.



Gambar 4. Confusion Matrix SVM

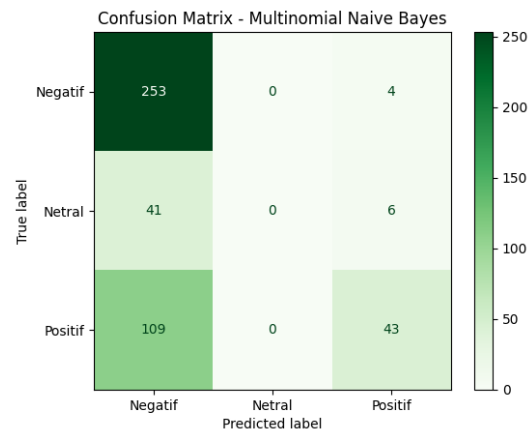
Model LR menunjukkan keunggulan pada metrik presisi dengan nilai sebesar 0,843, yang merupakan yang tertinggi dibandingkan model lainnya. Tingginya nilai presisi menunjukkan bahwa model LR memiliki tingkat kesalahan positif (false positive) yang relatif rendah, sehingga prediksi yang dihasilkan cenderung lebih akurat ketika model memutuskan suatu kelas, seperti pada Gambar 5. Namun demikian, nilai recall LR yang hanya sebesar 0,511 menunjukkan bahwa model ini masih kurang optimal dalam menangkap seluruh data yang seharusnya terklasifikasi dalam kelas tertentu. Hal ini mengindikasikan adanya trade-off antara presisi dan recall pada model LR.



Gambar 5. Confusion Matrix LR

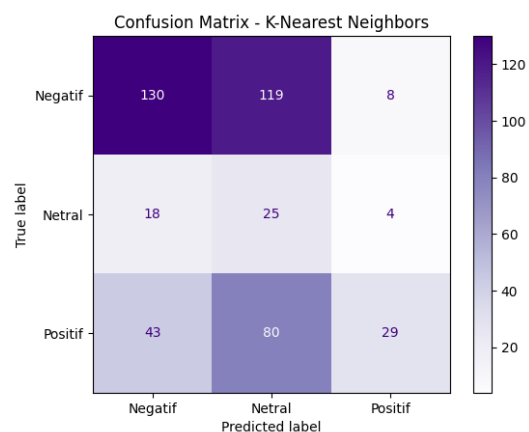
Sementara itu, model MNB menunjukkan kinerja yang lebih rendah dibandingkan SVM dan LR, dengan akurasi sebesar 0,649, presisi 0,480, recall 0,422, dan F1-score 0,395. Rendahnya performa ini dapat disebabkan oleh asumsi independensi antar fitur yang digunakan oleh MNB, yang dalam praktiknya sering kali tidak terpenuhi, terutama pada data teks yang memiliki keterkaitan antar kata. Hal ini menunjukkan kecenderungan bias terhadap kelas mayoritas. Walaupun MNB mencatatkan jumlah prediksi benar tertinggi pada kelas

negatif, model ini banyak melakukan kesalahan klasifikasi pada kelas positif, di mana sejumlah data signifikan justru terdeteksi sebagai sentimen negatif, seperti yang dapat dilihat pada Gambar 6.



Gambar 6. Confusion Matrix Naïve Bayes

Sementara itu, berdasarkan Gambar 7, model KNN menunjukkan performa paling rendah dibandingkan seluruh model yang diuji. Hal ini terlihat dari sebaran prediksi yang cenderung sporadis dan tidak terkonsentrasi pada diagonal utama, khususnya pada kelas positif dan netral. Kondisi tersebut mengindikasikan bahwa KNN mengalami kesulitan dalam membedakan kelas sentimen secara tepat. Keterbatasan ini menunjukkan bahwa algoritma berbasis jarak kurang efektif ketika diterapkan pada fitur TF-IDF yang bersifat sparse dan berdimensi tinggi. Selain itu, tingginya dimensi data juga memengaruhi perhitungan jarak antar data (*curse of dimensionality*), yang pada akhirnya berkontribusi terhadap rendahnya akurasi model.



Gambar 7. Confusion Matrix KNN

Secara statistik, dominasi performa SVM dibandingkan algoritma lainnya mengindikasikan kemampuan model ini dalam menjaga keseimbangan yang optimal antara

precision dan recall pada klasifikasi *multiclass*. Hal ini membuktikan bahwa SVM memiliki reliabilitas yang lebih tinggi dalam memproses karakteristik dataset opini publik terkait tren kendaraan listrik yang kompleks. Sebaliknya, meskipun LR menunjukkan keunggulan pada aspek presisi yang berarti sangat akurat dalam meminimalkan kesalahan prediksi *false positive*, nilai *recall*-nya yang lebih rendah menunjukkan adanya bias statistik di mana model cenderung gagal menangkap banyak data aktual yang relevan. Oleh karena itu, LR hanya dapat menjadi alternatif apabila fokus utama sistem adalah pada tingkat presisi yang sangat tinggi. Sementara itu, rendahnya performa KNN dan hasil MNB yang kurang maksimal membuktikan secara empiris bahwa kedua algoritma tersebut kurang optimal dalam menangani dimensi data teks yang bersifat *sparse* dan berdimensi tinggi.

Berdasarkan analisis lebih mendalam pada seluruh model, terungkap adanya keterbatasan utama dalam mengklasifikasikan kelas netral. Kelas netral merupakan kategori yang paling sulit diprediksi oleh semua model, di mana sering kali terjadi misklasifikasi dengan memprediksi data netral ke dalam kategori negatif atau positif. Keterbatasan ini secara statistik dipengaruhi oleh rendahnya frekuensi data netral dibandingkan kelas lainnya, yang dikombinasikan dengan adanya ambiguitas fitur bahasa yang tumpang tindih. Kemiripan karakteristik kata antar kelas menyebabkan model kesulitan menentukan batas keputusan (*decision boundary*) yang tegas pada opini yang bersifat moderat. Fenomena ini menegaskan bahwa meskipun SVM terpilih sebagai model paling andal, diperlukan upaya pengembangan lanjutan untuk mengatasi keterbatasan tersebut, seperti penerapan teknik penyeimbangan data melalui *oversampling* atau *undersampling* serta optimasi hyperparameter untuk meningkatkan keandalan model pada kelas-kelas minoritas di masa depan.

V. PENUTUP

Pemilihan algoritma klasifikasi berpengaruh signifikan terhadap performa analisis sentimen pada data teks. Berdasarkan hasil pengujian empat model *supervised learning*, yaitu Support Vector Machine (SVM), Logistic Regression (LR), Multinomial Naive Bayes (MNB), dan K-Nearest Neighbor (KNN), SVM menunjukkan kinerja terbaik dengan nilai akurasi dan F1-score tertinggi serta keseimbangan yang baik antara presisi dan

recall. Hal ini menunjukkan bahwa SVM memiliki kemampuan generalisasi yang lebih baik dalam menangani data berdimensi tinggi seperti fitur TF-IDF.

Logistic Regression menunjukkan performa yang cukup kompetitif, terutama pada nilai presisi yang tinggi, namun masih memiliki keterbatasan pada recall. Sementara itu, Multinomial Naive Bayes memberikan hasil yang moderat, sedangkan KNN menunjukkan performa paling rendah akibat keterbatasannya dalam menangani data berdimensi tinggi dan sifat *sparse*.

Penelitian ini memberikan kontribusi utama berupa evaluasi komparatif klasifikasi sentimen *multiclass* yang menegaskan efektivitas SVM dalam menangani ambiguitas opini publik pada domain kendaraan listrik di Indonesia. Temuan ini menjadi dasar penting bagi pengembangan sistem pemantauan persepsi masyarakat yang lebih komprehensif dibandingkan model biner konvensional.

Meskipun demikian, hasil analisis juga menunjukkan bahwa model masih mengalami kesulitan dalam mengklasifikasikan kelas tertentu, khususnya kelas netral, yang mengindikasikan adanya ketidakseimbangan distribusi data dan kemiripan karakteristik antar kelas. Oleh karena itu, diperlukan upaya lebih lanjut seperti optimasi model, melalui tuning hyperparameter, penerapan teknik penyeimbangan data seperti *oversampling* atau *undersampling*, serta eksplorasi metode lain seperti ensemble learning atau pendekatan berbasis deep learning untuk meningkatkan performa klasifikasi, khususnya dalam menangani kelas yang sulit diprediksi.

DAFTAR PUSTAKA

- [1] A. J. Paramitha and S. Kristiyanto, "EV Adoption in Surabaya: Social, Economic, Environmental Impact," *Jurnal Samudra Ekonomika*, vol. 9, no. 2, pp. 1-15, Sep. 2025, doi: 10.33059/jse.v9i2.11514.
- [2] M. L. Souma, I. P. Hakim, R. N. Rizki, N. L. Fitriani, and Mushawir, "Dampak Electric Vehicle Terhadap Lingkungan dan Ekonomi Berkelanjutan," *Journal of Economics, Business and Entrepreneurship (JEBER)*, vol. 2, no. 4, pp. 12-18, Agt. 2025, doi: 10.69714/03cm9274.
- [3] R. Hidayat, Imran, and R. Ramadhan, "Peran Media Sosial dalam Mengkonstruksi Opini Publik Terkait Kebijakan Pemerintah: Studi Kasus Wacana Publik Tahun 2025," *CORE:*

- Journal of Communication Research*, vol. 3, no. 2, pp. 62-75, Jul. 2025, doi: 10.47650/core.v3i2.2223.
- [4] H. M. U. Ali, Q. Farooq, A. Imran, et al., "A systematic literature review on sentiment analysis techniques, challenges, and future trends," *Knowl Inf Syst*, vol. 67, pp. 3967–4034, 2025, doi: 10.1007/s10115-025-02365-x.
- [5] Y. A. Firdaus, M. N. Annabil, E. P. P. A. Wibawa, and M. Y. T. Sulistyono, "Perbandingan Algoritma Naive Bayes dan Decision Tree pada Analisis Sentimen Media Sosial Terhadap Kendaraan Listrik Indonesia Menggunakan Metode TF-IDF," *Jurnal MNEMONIC*, vol. 8, no. 2, pp. 88-94, Sep. 2025, doi: 10.36040/mnemonic.v8i2.14873.
- [6] Sehabudin, A. I. Purnamasari, A. Bahtiar, and E. Wahyudin, "Peningkatan Model Klasifikasi Sentimen Publik di Youtube CNBC Indonesia Terhadap Mobil Listrik Menggunakan Algoritma Support Vektor Machine," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 2680-2687, Agt. 2025, doi: 10.36040/jati.v9i4.14184.
- [7] V. B. Lestari, E. Utami, and Hanafi, "Combining Bi-LSTM And Word2vec Embedding for Sentiment Analysis Models of Application User Reviews," *Indonesian Journal of Computer Science (IJCS)*, vol. 13, no. 1, pp. 311-325, Feb. 2024, doi: 10.33022/ijcs.v13i1.3647.
- [8] V. B. Lestari and C. A. Hutagalung, "Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes," *J-KOMA (Journal of Computer Science and Applications)*, vol. 8, no. 1, pp. 36-44, Jun. 2025, doi: 10.21009/j-koma.v8i1.05.
- [9] F. Farasalsabila, V. B. Lestari, D. D. N. Cahyo, Hanafi, T. Lestari, F. R. A. Islami, and M. A. Maulana, "Sentiment Analysis for IMDb Movie Review Using Support Vector Machine (SVM) Method," *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 8, no. 2, pp. 60-66, Jul. 2023, doi: 10.25139/inform.v8i2.5700.
- [10] E. S. Shindy, "Evaluasi Kinerja Machine Learning dalam Memprediksi Kemampuan Adaptasi Mahasiswa pada Lingkungan Pembelajaran Daring," *Eval. Kinerja Mach. Learn. dalam Memprediksi Kemamp. Adapt. Mhs. pada Lingkung. Pembelajaran Daring*, vol. 5, no. 1, pp. 50–57, 2025, doi: 10.55382/jurnalpustakaai.v5i1.901.
- [11] I. Akbar, F. Supriadi, and D. Indra Junaedi, "Pemanfaatan Machine Learning Di Bidang Kesehatan," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 1, pp. 1744–1749, 2025, doi: 10.36040/jati.v9i1.12663.
- [12] M. Ichsan, "Machine Learning Deteksi Penyakit Pada Kucing Menggunakan RANDOM FOREST," vol. 13, no. 3, 2000, doi: 10.23960/jitet.v13i3S1.8164.
- [13] I. K. N. Ananda, N. P. N. P. Dewi, N. W. Marti, and L. J. E. Dewi, "Klasifikasi Multilabel Pada Gaya Belajar Siswa Sekolah Dasar Menggunakan Algoritma Machine Learning," *Journal of Applied Computer Science and Technology*, vol. 5, no. 2, pp. 144–154, 2024, doi: 10.52158/jacost.v5i2.940.
- [14] R. G. Wardhana, G. Wang, and F. Sibuea, "Penerapan Machine Learning dalam Prediksi Tingkat Kasus Penyakit di Indonesia," *Journal of Information System Management (JOISM)*, vol. 5, no. 1, pp. 40-45, 2023, doi: 10.24076/joism.2023v5i1.1136.
- [15] M. Mirafuddin, "Penerapan Machine Learning pada Aplikasi Kesehatan HEALTIME," vol. 7, no. 6, pp. 1601–1606, 2024, doi.org/10.32672/jnkti.v7i6.8233
- [16] Y. D. Amritha, N. Luh, P. Ika, and W. P. Dananjaya, "Model Machine Learning yang Dioptimalkan untuk Prediksi Penyakit Jantung Menggunakan R Shiny," vol. 8, no. 01, pp. 1–10, 2026, doi: 10.1109/ICCPCT.2014.7054837
- [17] H. Utami, "Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network," *Indones. J. Appl. Stat.*, vol. 5, no. 1, p. 31, 2022, doi: 10.13057/ijjas.v5i1.56825.
- [18] D. Mustikasari, I. Widaningrum, R. Arifin, and W. H. E. Putri, "Comparison of Effectiveness of Stemming Algorithms in Indonesian Documents," *Proc. 2nd Borobudur Int. Symp. Sci. Technol. (BIS-STE 2020)*, vol. 203, pp. 154–158, 2021, doi: 10.2991/aer.k.210810.025.
- [19] H. Zhao, Z. Liu, X. Yao, and Q. Yang, "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach," *Inf. Process. Manag.*, vol. 58, no. 5, p. 102656, 2021, doi: 10.1016/j.ipm.2021.102656.
- [20] A. Iqbal, R. Amin, J. Iqbal, R. Alroobaea, A. Binmahfoudh, and M. Hussain, "Sentiment

- Analysis of Consumer Reviews Using Deep Learning," *Sustain.*, vol. 14, no. 17, 2022, doi: 10.3390/su141710844.
- [21] C. Colón-Ruiz and I. Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drug reviews," *J. Biomed. Inform.*, vol. 110, no. February, p. 103539, 2020, doi: 10.1016/j.jbi.2020.103539.
- [22] R. A. Nugroho, D. Remawati, T. Susyanto, W. Laksito, and Y. Saptomo, "Perbandingan Metode Support Vector Machine dan K-Nearest Neighbors dalam Analisis Sentimen Ulasan Aplikasi MyXL," vol. 14, no. 1, 2026, doi: 10.30646/tikomsin.v14i1.1062.
- [23] W. Hidayat, M. Ardiansyah, and A. Setyanto, "Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 11–20, 2021, doi: 10.29408/edumatic.v5i1.3125.
- [24] F. I. R. Firamadhina and H. Krisnani, "PERILAKU GENERASI Z TERHADAP PENGGUNAAN MEDIA SOSIAL TIKTOK: TikTok Sebagai Media Edukasi dan Aktivisme," *Share Soc. Work J.*, vol. 10, no. 2, p. 199, 2021, doi: 10.24198/share.v10i2.31443.
- [25] W. T. Atmojo, E. Keisya, and A. T. Ayunda, "Implementasi Algoritma Naïve Bayes dalam Analisis Sentimen terhadap Trend TikTok," *Jurnal TIKomSiN*, vol. 13, no. 2, pp. 25-34, Okt. 2025, doi: 10.30646/tikomsin.v13i2.1015.
- [26] D. Remawati, H. Wijayanto, Y. R. W. Utami, and B. D. Raharja, "Pengelompokan Film Trending di Youtube Menggunakan TF-IDF dan K-Means Clustering," *Jurnal Sistem Informasi TGD*, vol. 4, no. 1, pp. 65-74, Jan. 2025, doi: 10.53513/jursi.v4i1.10614.
- [27] D. Remawati, E. Noersasongko, Pujiono, and A. Marjuni, "Mental Health Detection with TF-IDF Feature Extraction," dalam *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, Surakarta, Indonesia, 2024, hlm. 1-6, doi: 10.1109/AIMS61812.2024.10512480.