

PERBANDINGAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE DAN NAÏVE BAYES PADA ANALISIS SENTIMEN KENDARAAN LISTRIK

Ni Wayan Ernawati¹, I Nyoman Satya Kumara², Widyadi Setiawan²

¹Mahasiswa Program Studi Teknik Elektro, Fakultas Teknik, Universitas Udayana

²Dosen Program Studi Teknik Elektro, Fakultas Teknik, Universitas Udayana

Jl. Raya Kampus Unud Jimbaran, Kec. Kuta Sel., Kabupaten Badung, Bali

niwayanernawati1206@gmail.com

ABSTRAK

Kendaraan listrik merupakan salah satu solusi yang dapat digunakan untuk menghadapi permasalahan emisi gas rumah kaca. Peralihan ke kendaraan listrik bisa menjadi solusi yang efektif karena kendaraan listrik memiliki banyak keunggulan. Namun, penerimaan kendaraan listrik di Indonesia tergantung pada opini atau sentimen yang diberikan oleh masyarakat. Analisis sentimen dapat memberikan gambaran secara spesifik tentang bagaimana sentimen atau opini yang diberikan oleh masyarakat Indonesia terhadap kendaraan listrik. Proses analisis sentimen dilakukan menggunakan bahasa pemrograman Python. Penelitian ini membandingkan metode SVM dan *Naïve Bayes* dalam analisis sentimen dari segi akurasi dan efisiensi waktu. Sebanyak 717 data digunakan sebagai data uji dan SVM mengklasifikasikan dengan benar sebanyak 150 data negatif, 152 data netral, dan 277 data positif. Sedangkan *Naïve Bayes* mengklasifikasikan dengan benar sebanyak 166 data negatif, 143 data netral, dan 282 data positif. Melalui proses evaluasi diperoleh akurasi SVM sebesar 81% dan *Naïve Bayes* sebesar 82%. Waktu pelatihan yang dibutuhkan untuk metode SVM yakni 37,42 detik sedangkan *Naïve Bayes* 0,10 detik. *Naïve Bayes* menjadi metode terbaik dalam penelitian ini karena akurasi yang tinggi dan waktu pelatihan yang cepat.

Kata kunci : Analisis Sentimen, *Support Vector Machine*, *Naïve Bayes* , Kendaraan Listrik

ABSTRACT

Electric vehicles are one of the solutions that can be used to deal with the problem of greenhouse gas emissions. Switching to electric vehicles can be an effective solution because electric vehicles have many advantages. However, the acceptance of electric vehicles in Indonesia depends on the opinions or sentiments given by the community. Sentiment analysis can provide a specific picture of how the sentiment or opinion given by the Indonesian people towards electric vehicles. The sentiment analysis process is carried out using the Python programming language. This research compares SVM and Naïve Bayes methods in sentiment analysis in terms of accuracy and time efficiency. A total of 717 data were used as test data and SVM correctly classified 150 negative data, 152 neutral data, and 277 positive data. Meanwhile, Naïve Bayes correctly classified 166 negative data, 143 neutral data, and 282 positive data. training time required for the SVM method is 37.42 seconds while Naïve Bayes is 0.10 seconds. Naïve Bayes is the best method in this study because of its high accuracy and fast training time.

Key Words : *Sentiment Analysis, Support Vector Machine, Naïve Bayes , Electric Vehicle*

1. PENDAHULUAN

Kendaraan listrik merupakan salah satu solusi yang dapat digunakan untuk menghadapi permasalahan emisi gas rumah kaca. Penyumbang emisi gas rumah kaca terbesar di Indonesia berasal dari sektor

transportasi [1]. Sehingga peralihan ke kendaraan listrik bisa menjadi solusi yang efektif karena kendaraan listrik memiliki banyak keunggulan, seperti lebih ramah lingkungan, efisiensi energi, dan biaya operasional yang lebih rendah [2]. Namun,

keberadaan dari kendaraan listrik sendiri masih menjadi perdebatan di kalangan masyarakat [3].

Sentimen dari masyarakat bisa menjadi tolok ukur dari tingkat adopsi kendaraan listrik [4]. Salah satu cara untuk mengetahui sentimen yang diberikan oleh masyarakat yakni melalui media sosial. Twitter menjadi salah satu media sosial yang sering digunakan oleh masyarakat Indonesia untuk membagikan pendapat atau opini terkait topik tertentu. Analisis sentimen dapat dilakukan dengan menggunakan metode klasifikasi seperti *Support Vector Machine* (SVM), *Naïve Bayes Classifier*, *K-Nearest Neighbor*, dan lain-lain [5],[6],[7]. Penelitian oleh Salman Alfarizi dan Eka Fitriani dengan judul “Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma *Naive Bayes* dengan Seleksi Fitur *Information Gain* dan *Particle Swarm Optimization*”, memperoleh akurasi 79,43% dan AUC 0,639. Selain itu, setelah menggunakan algoritma *Naive Bayes* dengan seleksi fitur *Information Gain* dan *Particle Swarm Optimization* akurasi yang diperoleh yakni 84,54% dan AUC 0,729 [2].

Penelitian yang dilakukan oleh Adittia Agustian, Tukino, dan Fitria Nurapriani dengan judul “Penerapan Analisis Sentimen dan *Naive Bayes* terhadap Opini Penggunaan Kendaraan Listrik di Twitter”, memperoleh akurasi data sebesar 80%. Penelitian ini menggunakan kata kunci “Kendaraan Listrik” dan data yang dikumpulkan mulai dari 08 Juli 2019 sampai 10 Oktober 2022. Berdasarkan hasil klasifikasi, diperoleh respon masyarakat terhadap kendaraan listrik lebih mengarah positif [8].

Penelitian yang dilakukan oleh Imam Adi Nata, Damar Wicaksono, dan D. Jayus Nor Salim dengan judul “Analisis Sentimen Publik Indonesia Terhadap Motor Listrik pada Media Sosial Twitter”, mendapatkan hasil akurasi tertinggi oleh model SVM yakni 91%. Sedangkan *Naive Bayes*, *K-Nearest Neighbors* (KNN), dan *random forest* memperoleh akurasi terbesar kedua yakni 88%. Terakhir model *decision tree* memperoleh akurasi 83%. Penelitian ini menggunakan kata kunci “Motor Listrik” dan data yang dikumpulkan mulai dari Januari hingga April 2023. Berdasarkan hasil klasifikasi, diperoleh respon positif mendominasi sentimen masyarakat terhadap motor listrik [9].

Metode *Naive Bayes* menjadi metode yang sering digunakan untuk menyelesaikan permasalahan analisis sentimen karena memiliki algoritma yang sederhana. Selain itu, SVM juga menjadi salah satu algoritma yang cocok digunakan untuk klasifikasi teks karena kemampuannya untuk menemukan *hyperplane* yang optimal [10].

Tujuan penelitian ini adalah untuk mengetahui secara spesifik sentimen atau opini yang diberikan oleh masyarakat Indonesia terhadap kendaraan listrik dengan membandingkan dua metode klasifikasi. Penelitian ini membandingkan metode SVM dan *Naive Bayes* untuk mengetahui metode yang paling bagus dari segi akurasi dan efisiensi waktu. Analisis sentimen dalam penelitian ini dilakukan dengan membagi sentimen menjadi tiga kelas yakni sentimen positif, negatif, dan netral. Adapun data yang digunakan yakni data yang berasal dari Twitter dengan kata kunci kendaraan listrik, motor listrik, dan mobil listrik.

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen atau yang sering disebut sebagai *opinion mining* merupakan prosedur yang digunakan untuk memahami dan mengklasifikasikan informasi yang terkandung dalam tulisan dengan menggunakan alat analisis teks [11]. Analisis sentimen digunakan sebagai metode untuk mengevaluasi dan memahami perasaan atau opini yang terdapat dalam ulasan, komentar, dan umpan balik dari berbagai sumber.

Sentimen atau pendapat yang diekspresikan dalam teks dapat dikelompokkan menjadi tiga kategori: baik (positif), netral (tidak memiliki sentimen kuat), dan negatif. Analisis sentimen bertujuan untuk mengidentifikasi pikiran, emosi, dan sikap yang diungkapkan dalam sebuah teks atau sekelompok teks [12].

2.2 Kendaraan Listrik

Kendaraan listrik adalah kendaraan yang ramah lingkungan karena tidak menggunakan bahan bakar fosil, tetapi menggunakan baterai sebagai sumber tenaga untuk menggerakkan motornya. Hal ini berarti kendaraan listrik tidak menghasilkan polusi udara seperti kendaraan bermotor konvensional. Komponen utama dari kendaraan listrik meliputi motor listrik, baterai, alat pengisian

ulang, kontrol kecepatan, dan sistem manajemen energi [13]. Beberapa jenis kendaraan listrik yang dikembangkan sampai saat ini yakni sepeda motor listrik, mobil listrik, sepeda listrik, bus listrik, dan skuter listrik.

2.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi. Secara sederhana, SVM mencari garis keputusan atau batas terbaik untuk memisahkan dua kelas yang berbeda dalam data [14].

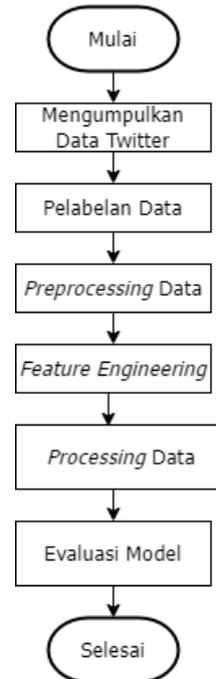
Tujuan utama dari SVM adalah untuk menemukan *hyperplane* yang memiliki jarak terbesar antara dua kelas yang disebut sebagai margin. SVM berusaha untuk memaksimalkan margin ini sehingga dapat memberikan keputusan klasifikasi yang paling akurat untuk data yang belum pernah dilihat sebelumnya. Selain itu, SVM juga memperhitungkan titik-titik yang paling dekat dengan margin ini. Titik-titik ini disebut sebagai *support vector* yang membantu SVM dalam menentukan posisi dan orientasi *hyperplane* [15].

2.4 Naïve Bayes

Naïve Bayes adalah metode yang sederhana dalam klasifikasi probabilitas dengan mengandalkan teorema Bayes. Pada metode ini, asumsi utama adalah bahwa setiap fitur atau variabel dalam data adalah independen, tidak memiliki hubungan satu sama lain. Meskipun asumsi ini terkadang tidak realistis dalam beberapa kasus, metode *Naïve Bayes* tetap efektif dan sering digunakan dalam berbagai aplikasi. Metode ini mengukur seberapa mungkin suatu data masuk ke dalam kelas tertentu berdasarkan nilai fitur yang diamati. Kelebihan utama dari *Naïve Bayes* adalah tidak memerlukan jumlah data yang besar dalam proses pelatihan model, sehingga cocok untuk situasi di mana ketersediaan data terbatas [16].

3. METODE PENELITIAN

Penelitian ini dilaksanakan di Program Studi Teknik Elektro Fakultas Teknik Universitas Udayana, mulai dari bulan Januari sampai Mei 2023. Tahapan dari penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.1 Mengumpulkan Data Twitter

Pustaka Snsrape digunakan untuk mengumpulkan informasi dari Twitter. Snsrape merupakan salah satu pustaka Python untuk mengumpulkan informasi dari situs jejaring sosial, seperti profil pengguna, tagar, atau unggahan yang relevan sesuai dengan pencarian yang dilakukan. Snsrape membutuhkan *query* untuk mendapatkan *tweet* sesuai dengan topik yang diinginkan. *Query* dapat ditentukan dengan menggunakan fitur *advance search* pada Twitter [17]. Data Twitter yang dikumpulkan merupakan data *tweet* yang mengandung kata kunci kendaraan listrik, motor listrik, dan mobil listrik. Data dikumpulkan mulai dari 01 September 2022 sampai 28 Februari 2023.

3.2 Pelabelan Data

Proses pelabelan data dilakukan untuk menentukan kelas sentimen positif, negatif, dan netral pada dataset. Proses pelabelan ini dilakukan secara manual oleh dua orang annotator. Untuk memastikan bahwa prosedur pelabelan telah dikerjakan secara objektif, annotator kedua akan memeriksa ulang label yang diberikan oleh annotator pertama.

3.3 Preprocessing Data

Preprocessing data perlu dilakukan untuk mengubah data *tweet* yang tidak

terstruktur menjadi data yang lebih terstruktur [18] sehingga lebih mudah untuk diproses untuk keperluan analisis sentimen kendaraan listrik. Tahap *preprocessing* meliputi *case folding*, *cleansing text*, *word normalization*, *tokenization*, *stemming*, dan *stopword removal*. *Case folding* digunakan untuk mengubah semua huruf dalam dataset menjadi huruf kecil [18]. Proses ini dilakukan dengan tujuan untuk menghindari perbedaan penulisan yang tidak konsisten dalam satu dokumen. *Cleansing text* dilakukan untuk membersihkan teks dari tanda baca, angka, *emoticon*, serta *white space* [19]. *Word normalization* digunakan untuk mengubah kata tidak baku menjadi kata baku [20]. *Tokenization* digunakan untuk membagi kalimat menjadi bagian-bagian kecil berupa kata. *Stemming* digunakan untuk merubah sebuah kata menjadi bentuk dasarnya dengan cara menghapus awalan dan akhiran kata tersebut [21]. *Stopword removal* digunakan untuk mengeliminasi kata-kata yang tidak relevan [20] seperti kata tanya, kata depan, dan kata penghubung.

3.4 Feature Engineering

Penelitian ini mengimplementasikan *feature engineering* dengan kombinasi TF-IDF dan N-Gram dalam proses pembobotan kata. Selain itu dilakukan juga proses seleksi fitur menggunakan metode Chi-Square, sehingga fitur atau kata yang digunakan pada saat pemodelan atau analisis data adalah fitur yang benar-benar relevan. Proses *feature engineering* dilakukan menggunakan *library* Scikit-Learn pada Python.

3.5 Processing Data

Penelitian ini menggunakan algoritma klasifikasi *Support Vector Machine* (SVM) dan *Naïve Bayes* untuk melakukan analisis sentimen kendaraan listrik. Pada tahap *processing data*, dilakukan proses *tuning hyperparameter* dengan menggunakan *GridSearch* untuk menemukan kombinasi parameter yang optimal. Metode *GridSearch* secara sistematis mencoba semua kombinasi yang mungkin dari nilai *hyperparameter* yang telah ditentukan sebelumnya [22]. Model SVM dan *Naïve Bayes*, selanjutnya digunakan untuk melakukan pelatihan menggunakan data latih (*training*) dan pengujian pada data uji (*testing*), sehingga diperoleh hasil prediksi

berupa kelas sentimen, yaitu sentimen positif, negatif, atau netral.

3.6 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui performa dari model *machine learning* [17]. Evaluasi model SVM dan *Naïve Bayes* untuk analisis sentimen kendaraan listrik dilakukan dengan menggunakan *confusion matrix* untuk mengetahui nilai *accuracy*, *precision*, *recall*, dan *f1-score* dari masing-masing metode. Dengan membandingkan data yang diklasifikasikan dengan seluruh jumlah data yang disediakan, maka dapat diketahui akurasi dari model. *Precision* dapat diketahui dengan membandingkan data yang berhasil diprediksi dengan data yang diklasifikasikan dengan benar. Perbandingan data yang diklasifikasikan dengan benar dan total data yang sebenarnya disebut *recall*. Sedangkan *F1-Score* adalah metrik yang digunakan untuk menunjukkan bagaimana presisi dan *recall* seimbang [23].

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Data yang dikumpulkan merupakan data teks dengan kata kunci kendaraan listrik, motor listrik, dan mobil listrik yang diperoleh dengan menggunakan *library* SnsCrape. Data yang berhasil dikumpulkan yakni sebanyak 4.000 data *tweet* berbahasa Indonesia yang selanjutnya disimpan dalam format CSV. Data yang berhasil dikumpulkan ini merupakan data mentah, dan masih terdapat data yang tidak sesuai dengan topik serta masih banyak data yang terduplikat. Sehingga perlu dilakukan pembersihan data dengan menghapus *tweet* yang duplikat dan menghapus data yang tidak sesuai dengan topik. Melalui proses ini, data yang tersisa yakni 3.582 data yang siap digunakan untuk proses *preprocessing*. Contoh data *tweet* yang berhasil dikumpulkan ditunjukkan pada Tabel 1.

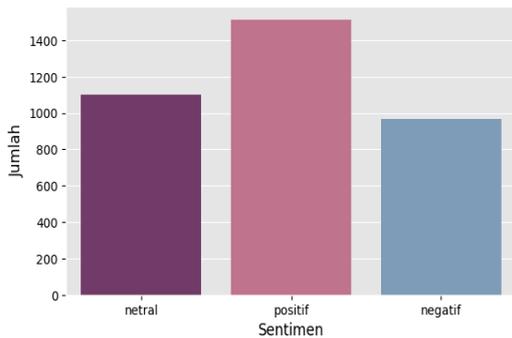
Tabel 1. Contoh Data Tweet

No	Teks Tweet
1	Agak gimana gitu ya... infrastruktur jalanan aje masih berantakan. Rusak dikit, gak diberesin. Rusak banyak, lama benerinya. Ini disuruh percepatan kendaraan listrik ? Jadi berasa baru masuk 4G hari Kamis, disuruh nerapin 5G hari Jumat. Cuma

	keuntungan doangan sih yang dipikirin!
2	Kalau pemerintah sudah menggencarkan program yang menarik agar warga beralih ke kendaraan listrik tak lupa sarana prasarannya juga disediakan #pakemotorlistrikyuk #pakemobillistrikyuk #pakebislistrikyuk #lebihhemat
3	@penjaga_nkri45 Agenda Presidensi G20 bisa dijadikan tempat untuk memamerkan industri kendaraan listrik Tanah Air. #NKRIMembangun

4.2 Pelabelan Data

Proses pelabelan data dilakukan secara manual untuk menentukan kelas sentimen positif, negatif, dan netral pada dataset. Proses pelabelan ini dilakukan oleh dua orang annotator. *Tweet* yang mengandung kalimat dukungan atau ajakan menggunakan kendaraan listrik dikategorikan sebagai sentimen positif, adapun *tweet* yang mengandung kata-kata tidak setuju ataupun kata-kata kasar dikategorikan sebagai sentimen negatif, dan *tweet* berupa informasi ataupun berita tentang kendaraan listrik dikategorikan sebagai sentimen netral. Perbandingan jumlah sentimen dari proses pelabelan ditunjukkan pada Gambar 2.



Gambar 2. Perbandingan Jumlah Sentimen

Pada Gambar 2 dapat diketahui bahwa dari total 3.582 data *tweet* yang dikumpulkan sebanyak 1.511 data memiliki label positif, 1.102 data berlabel netral, dan 969 data memiliki label negatif.

4.3 Preprocessing Data

Preprocessing data penting untuk dilakukan untuk menyelaraskan format pada data agar memudahkan saat proses analisis data.

a) Case Folding

Pada tahap ini teks *tweet* yang mengandung huruf kapital diseragamkan menjadi huruf kecil [24]. Contoh data hasil *case folding* ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil Case Folding

<i>Tweet Asli:</i> @penjaga_nkri45 Agenda Presidensi G20 bisa dijadikan tempat untuk memamerkan industri kendaraan listrik Tanah Air. #NKRIMembangun
<i>Case Folding:</i> @penjaga_nkri45 agenda presidensi g20 bisa dijadikan tempat untuk memamerkan industri kendaraan listrik tanah air. #nkrimembangun

b) Cleansing Text

Pada tahap ini teks *tweet* dibersihkan dari tanda baca, angka, tautan, emoji, serta *white space* sehingga dihasilkan teks *tweet* yang bersih dari karakter-karakter tidak penting. Contoh data hasil *cleansing text* ditunjukkan pada Tabel 3.

Tabel 3. Contoh Hasil Cleansing Text

<i>Tweet Asli:</i> @penjaga_nkri45 Agenda Presidensi G20 bisa dijadikan tempat untuk memamerkan industri kendaraan listrik Tanah Air. #NKRIMembangun
<i>Cleansing Text:</i> agenda presidensi bisa dijadikan tempat untuk memamerkan industri kendaraan listrik tanah air

c) Word Normalization

Tahap *word normalization* dilakukan untuk mengubah kata yang tidak baku menjadi kata baku [20]. Contoh data hasil *word normalization* ditunjukkan pada Tabel 4.

Tabel 4. Contoh Hasil Word Normalization

<i>Tweet Asli:</i> @penjaga_nkri45 Agenda Presidensi G20 bisa dijadikan tempat untuk memamerkan industri kendaraan listrik Tanah Air. #NKRIMembangun
<i>Word Normalization:</i> agenda presidensi bisa dijadikan tempat untuk memamerkan industri kendaraan listrik tanah air

d) Tokenization

Pada tahap ini teks *tweet* yang masih berupa kalimat utuh, dipecah menjadi potongan kata-kata. Contoh data hasil *tokenization* ditunjukkan pada Tabel 5.

Tabel 5. Contoh Hasil *Tokenization*

<p><i>Tweet Asli:</i> @penjaga_nkri45 Agenda Presidensi G20 bisa dijadikan tempat untuk memamerkan industri kendaraan listrik Tanah Air. #NKRIMembangun</p>
<p><i>Tokenization:</i> ['agenda', 'presidensi', 'bisa', 'dijadikan', 'tempat', 'untuk', 'memamerkan', 'industri', 'kendaraan', 'listrik', 'tanah', 'air']</p>

e) *Stemming*

Stemming dilakukan untuk mengubah suatu kata menjadi kata dasarnya dengan menghilangkan imbuhan yang terdapat pada kata tersebut [21]. Contoh data hasil *stemming* ditunjukkan pada Tabel 6.

Tabel 6. Contoh Hasil *Stemming*

<p><i>Tweet Asli:</i> @penjaga_nkri45 Agenda Presidensi G20 bisa dijadikan tempat untuk memamerkan industri kendaraan listrik Tanah Air. #NKRIMembangun</p>
<p><i>Stemming:</i> ['agenda', 'presidensi', 'bisa', 'jadi', 'tempat', 'untuk', 'pamer', 'industri', 'kendara', 'listrik', 'tanah', 'air']</p>

f) *Stopword Removal*

Tahap ini dilakukan untuk membersihkan data *tweet* dari kata-kata yang dianggap tidak penting atau tidak relevan [20]. Contoh data hasil *stopword removal* ditunjukkan pada Tabel 7.

Tabel 7. Contoh Hasil *Stopword Removal*

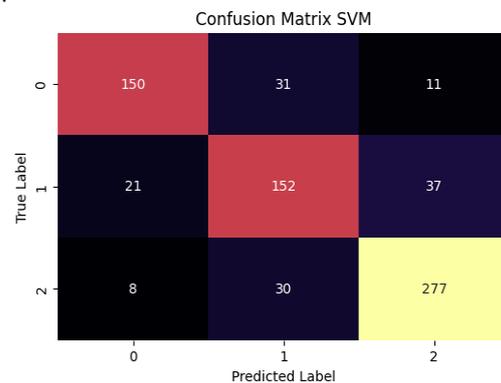
<p><i>Tweet Asli:</i> @penjaga_nkri45 Agenda Presidensi G20 bisa dijadikan tempat untuk memamerkan industri kendaraan listrik Tanah Air. #NKRIMembangun</p>
<p><i>Stopword Removal:</i> ['agenda', 'presidensi', 'pamer', 'industri', 'kendara', 'listrik', 'tanah', 'air']</p>

4.4 Feature Engineering

Data *tweet* yang sudah melalui tahap pelabelan dan *preprocessing* akan digunakan pada tahap selanjutnya yakni *feature engineering*. Dalam tahap ini dilakukan pembobotan kata menggunakan kombinasi TF-IDF dengan N-Gram (1,2), serta dilakukan juga seleksi fitur menggunakan Chi-Square. Dari proses *feature engineering* ini diperoleh sebanyak 5.000 kata unik yang siap digunakan untuk pemodelan.

4.5 Support Vector Machine (SVM)

Model SVM dibangun menggunakan *library* Scikit-learn dengan modul SVC (*Support Vector Classifier*). Sebanyak 3582 data pada dataset dibagi menjadi set pelatihan dan set pengujian dengan perbandingan 80:20. Sehingga jumlah data yang digunakan untuk pelatihan yakni 2.865 data dan untuk pengujian sebanyak 717 data. Model SVM menggunakan GridSearch untuk menemukan kombinasi parameter terbaik yang akan digunakan untuk pelatihan dan pengujian. Dengan menggunakan GridSearch diperoleh parameter terbaik yakni parameter kernel dengan nilai "rbf", parameter C=10, gamma=0,5, dan decision_function_shape dengan nilai "ovr". Hasil evaluasi untuk model *Support Vector Machine* ditunjukkan pada Gambar 3.

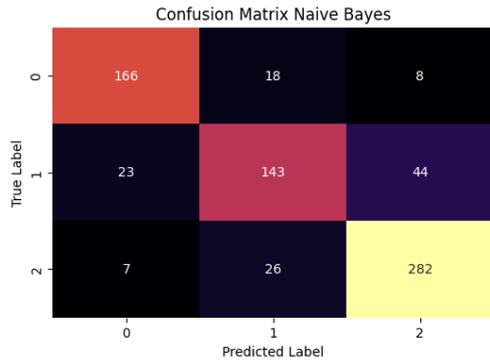


Gambar 3. Confusion Matrix SVM

Gambar 3 menunjukkan *confusion matrix* berukuran 3 x 3 untuk model SVM yang mewakili setiap kelas sentimen positif, negatif, dan netral. Berdasarkan *confusion matrix* tersebut dapat diketahui bahwa dari 717 data yang diuji, model SVM mengklasifikasikan dengan benar sebanyak 150 data negatif, 152 data netral, dan 277 data positif.

4.6 Naïve Bayes

Model *Naïve Bayes* dibuat dengan menggunakan MultinomialNB dalam *library* Scikit-learn. Data yang digunakan untuk membangun model *Naïve Bayes* sama dengan data yang digunakan untuk membangun model SVM. Model *Naïve Bayes* diatur dengan parameter alpha = 0,01 dan fit_prior bernilai True. Hasil evaluasi untuk model *Naïve Bayes* ditunjukkan pada Gambar 4.

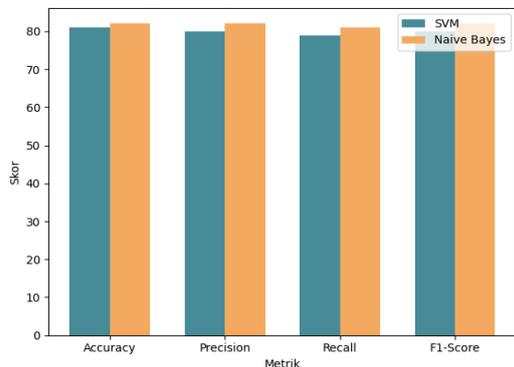


Gambar 4. Confusion Matrix Naive Bayes

Gambar 4 menunjukkan *confusion matrix* berukuran 3 x 3 untuk model *Naive Bayes* yang mewakili setiap kelas sentimen positif, negatif, dan netral. Berdasarkan *confusion matrix* tersebut dapat diketahui bahwa dari 1717 data yang diuji, model *Naive Bayes* mengklasifikasikan dengan benar sebanyak 166 data negatif, 143 data netral, dan 282 data positif.

4.7 Perbandingan Metode SVM dan Naive Bayes

Berdasarkan hasil evaluasi untuk analisis sentimen kendaraan listrik menggunakan SVM dan *Naive Bayes* diperoleh akurasi 81% untuk metode SVM dan 82% untuk metode *Naive Bayes*. Selain itu, dengan menggunakan *confusion matrix*, dapat diketahui juga nilai *precision*, *recall* dan *f1-score* untuk masing-masing metode. Perbandingan Performa SVM dan *Naive Bayes* ditampilkan pada Gambar 5.

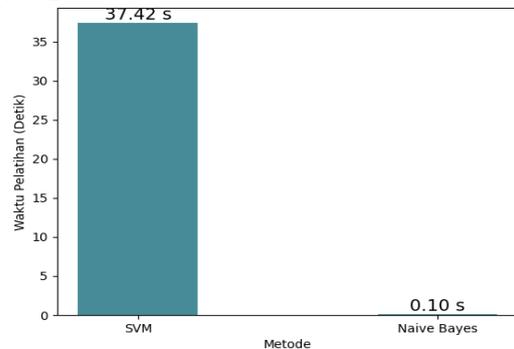


Gambar 5. Perbandingan Performa SVM dan Naive Bayes

Metode SVM memperoleh *precision* sebesar 80%, *recall* 79%, dan *f1-score* 80%. Sedangkan metode *Naive Bayes* memperoleh *precision* sebesar 82%, *recall* 81%, dan *f1-score* 82%. Hasil evaluasi ini menunjukkan bahwa kedua metode telah

berhasil mengklasifikasikan sentimen dengan baik dengan memperoleh akurasi lebih dari 80%, dengan akurasi tertinggi dicapai oleh metode *Naive Bayes* yakni sebesar 82%.

Adapun perbandingan dari segi waktu pelatihan untuk kedua metode dapat dilihat pada Gambar 6. Durasi waktu pelatihan tercepat pada penelitian ini adalah metode *Naive Bayes*, yaitu selama 0,10 detik. Metode *Naive Bayes* memiliki waktu pelatihan yang cepat karena model matematika dari metode *Naive Bayes* sangatlah sederhana dibandingkan dengan metode lain sehingga memungkinkan proses pelatihan yang cepat [25]. Sedangkan metode SVM memiliki durasi pelatihan selama 37,42 detik. Metode SVM memiliki durasi waktu yang lebih lama karena SVM lebih kompleks daripada metode *Naive Bayes*. Hal ini bisa terjadi karena metode SVM melibatkan perhitungan optimalisasi matematika yang lebih rumit daripada metode *Naive Bayes* yang lebih sederhana.



Gambar 6. Perbandingan Durasi Waktu Pelatihan

5. KESIMPULAN

Berdasarkan hasil dan pembahasan yang sudah dilakukan dalam penelitian ini maka dapat disimpulkan bahwa kedua metode telah memperoleh hasil akurasi yang baik untuk mengklasifikasikan sentimen kendaraan listrik. Metode *Naive Bayes* memperoleh akurasi sebesar 82% dan SVM memperoleh akurasi sebesar 81%, terdapat perbedaan hasil akurasi hanya sebesar 1%. Metode *Naive Bayes* memiliki kelebihan dari segi durasi waktu pelatihan yang lebih cepat yakni 0,10 detik dibandingkan dengan metode SVM yang memiliki durasi waktu pelatihan 37,42 detik. Pada penelitian ini dapat diketahui bahwa metode *Naive Bayes* merupakan metode terbaik untuk analisis sentimen kendaraan

listrik dari segi efisiensi waktu dan tingkat akurasi.

6. DAFTAR PUSTAKA

- [1] Subekti, R. 2022. "Urgensi Regulasi Kendaraan Listrik untuk Pengendalian Iklim dan Penggunaan Energi Terbarukan (Analisis Komparatif Antara Indonesia, China, dan Amerika Serikat)". *Jurnal Rechts Vinding Media Pembn. Huk. Nas. Vol. 11, No. 3*.
- [2] Alfarizi, S., & Fitriani, E. 2023. "Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur Information Gain dan Particle Swarm Optimization". *Indonesian Journal of Software Engineering Vol. 9, No. 1*.
- [3] Santoso, A., Nugroho, A., & Sunge, A. S. 2022. "Analisis Sentimen Tentang Mobil Listrik Dengan Metode Support Vector Machine Dan Feature Selection Particle Swarm Optimization". *Jurnal Practical Computer Science Vol. 2, No. 1*.
- [4] Christidis, P., & Focas, C. 2019. "Factors Affecting the Uptake of Hybrid and Electric Vehicles in the European Union". *Energies Vol. 12, No. 18*.
- [5] Thakur, R. K. 2019. "Kernel Optimized-Support Vector Machine and MapReduce Framework for Sentiment Classification of Train Reviews". *International Journal of Uncertainty, Fuzziness Knowledge-Based Systems Vol. 27, No. 6*
- [6] Wang, Q., Liu, K., & Ma, K. 2019. "Emotional Analysis of Public Opinions in Colleges and Universities: Based on Naive Bayesian Classification Method". *Journal of Physics Vol. 1187, No. 5*.
- [7] Furqan, M., Sriani, & Sari, S. M. 2022. "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia". *Techno.Com Vol. 21, No. 1*.
- [8] Agustian, A., Tukino, & Nurapriani, F. 2022. "Penerapan Analisis Sentimen dan Naive Bayes terhadap Opini Penggunaan Kendaraan Listrik di Twitter". *Jurnal Teknologi Informasi dan Komunikasi Aceh Vol. 7, No. 3*.
- [9] Nata, I. A., Wicaksono, D., & Salim, D. J. N. 2023. "Analisis Sentimen Publik Indonesia Terhadap Motor Listrik pada Media Sosial Twitter". *Jurnal Electrical Engineering, Computer, and Information Technology Vol. 4, No. 1*.
- [10] Indriyani, E. R., Paradise, & Wibowo, M. 2022. "Perbandingan Metode Naïve Bayes dan Support Vector Machine Untuk Analisis Sentimen Terhadap Vaksin Astrazeneca di Twitter". *Jurnal Media Informasi Budidarma Vol. 6, No. 3*.
- [11] Irawan, F. R., Jazuli, A., & Khotimah, T. 2022. "Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearest Neighbors". *Jurnal Informasi dan Komputer (JIKO), Vol. 5, No. 1*.
- [12] Amrani, Y. A. L., Lazaar, M., & EL Kadiri, K. E. 2018. "Random Forest and Support based Hybrid Approach to Sentiment Analysis". *Procedia Computer Science Vol. 127*.
- [13] Suranto, D. D., Anwar, S., Nuruddin, M., Rofi'i, A., & Zain, A. T. 2023. "Analisa Perancangan dan Pengujian Kendaraan Listrik Roda Dua dengan Variasi Pembebanan". *J-Proteksion Jurnal Kajian Ilmiah dan Teknologi Teknik Mesin Vol. 7, No. 2*.
- [14] Wafa, H. S., Hadiana, A. I., & Umbara, F. R. 2022. "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)". *Informatics and Digital Expert Vol. 4, No. 1*.
- [15] Uddin, S., Khan, A., Hossain, E., & Moni, M. A. 2019. "Comparing different supervised machine learning algorithms for disease prediction". *BMC Medical Informatics and Decision Making, Vol. 19, No. 281*
- [16] Suryani, N. P. S. M., Linawati, & Saputra, K. O. 2019. "Penggunaan Metode Naive Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia". *Majalah Ilmiah Teknologi Elektro Vol. 18, No. 1*.
- [17] Safira, A., & Hasan, F. N. 2023. "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier". *Jurnal Sistem Informasi Vol. 5, No. 1*.
- [18] Arsi, P., Wahyudi, R., & Waluyo, R. 2021. "Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia". *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) Vol. 5, No. 2*.

- [19] Aulia, T. M. P., Arifin, N., & Mayasari, R. 2021. "Perbandingan Kernel Support Vector Machine (SVM) Dalam Penerapan Analisis Sentimen Vaksinasi Covid-19". *SINTECH (Science Information Technology Journal) Vol. 4, No. 2*.
- [20] Aulia, G. N., & Patriya, E. 2019. "Implementasi Lexicon Based dan Naive Bayes Pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019". *Jurnal Ilmiah Informasi Komputer Vol. 24, No. 2*.
- [21] Alhaq, Z., Mustopa, A., Mulyatun, S., & Santoso, J. D. 2021. "Penerapan Metode Support Vector Machine untuk Analisis Sentimen Pengguna Twitter". *Jurnal Information Systems and Management Vol. 3, No. 1*.
- [22] Apriani, H., Jaman, J. H., & Adam, R. I. 2022. "Optimasi SVM menggunakan algoritme grid search untuk identifikasi citra biji kopi robusta berdasarkan circularity dan eccentricity". *Jurnal Teknologi dan Sistem Komputer Vol. 10, No. 1*.
- [23] Rivaldi, A. A., Azra, B., Ziaulhaq, Y. I., & Rakhmawati, N. A. 2022. "Analisis Karakteristik Akun Twitter Berdasarkan Sentimen Pendapat Terkait Undang-Undang PSE". *Jurnal Sains dan Teknologi Informasi Vol. 8, No. 2*.
- [24] Samsir, Ambiyar, Verawardina, U., Edi, F., & Watrianthos, R. 2021. "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes". *Jurnal Media Informasi Budidarma Vol. 5, No. 1*.
- [25] Setiawan, Y. A., Divayana, Y., & Widiadha, W. 2022. "Perbandingan Algoritma Supervised Machine Learning Untuk Sistem Penghindaran Halangan Pada Robot Assistant Udayana 02 (Ratna02)". *Jurnal SPEKTRUM Vol. 9, No. 3*.