

Analisis Sentimen Subsidi Kendaraan Listrik di Aplikasi X menggunakan Support Vector Machine

Bayu Anggoro^{*1}, Aditia Yudhistira²

^{1,2}Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia
Email: ¹bayu_anggoro@teknokrat.ac.id, ²aditiayudhistira@teknokrat.ac.id

Abstrak

Kebijakan subsidi kendaraan listrik memicu berbagai tanggapan di masyarakat yang menjadi topik diskusi hangat di media sosial, termasuk aplikasi X. Masalah utama yang dihadapi adalah beragamnya opini masyarakat, dari yang mendukung kendaraan listrik sebagai solusi defisit migas hingga yang menyoroti perlunya persiapan matang, khususnya infrastruktur. Penelitian ini bertujuan untuk mengidentifikasi kecenderungan opini masyarakat (positif, negatif, atau netral) terkait kebijakan tersebut. Penelitian menggunakan metode algoritma Support Vector Machine (SVM) yang dikenal unggul dalam kinerja klasifikasi, penanganan ketidakseimbangan data, dan data berdimensi tinggi. Dataset terdiri dari 1.812 tweet yang, setelah melalui tahapan preprocessing, dibagi menjadi 1.449 data latih dan 363 data uji. Hasil analisis menunjukkan bahwa 87,9% (319 tweet) bersentimen netral, 5,2% (19 tweet) negatif, dan 6,9% (25 tweet) positif, menandakan masyarakat belum memiliki pandangan tegas terhadap kebijakan ini. Metode SVM menghasilkan performa yang baik dengan akurasi 86,43%, precision positif 83,33%, recall 87,30%, dan f1-score 85,27%. Dampak penelitian ini diharapkan memberikan wawasan mendalam tentang persepsi masyarakat terhadap subsidi kendaraan listrik sehingga dapat mendukung pengambilan kebijakan yang lebih efektif dan tepat sasaran.

Kata kunci: *analisis sentiment, subsidi kendaraan listrik, support vector machine.*

Sentiment Analysis of Electric Vehicle Subsidy in Application X Using Support Vector Machine

Abstract

Electric vehicle subsidy policies have sparked various responses from the public, becoming a hot topic of discussion on social media, including Application X. The main issue lies in the diverse public opinions, ranging from those who support electric vehicles as a solution to the oil and gas deficit to those who emphasize the need for thorough preparation, particularly regarding infrastructure. This study aims to identify the tendency of public opinions (positive, negative, or neutral) toward the policy. The study employs the Support Vector Machine (SVM) algorithm, known for its excellence in classification performance, handling data imbalance, and processing high-dimensional data. The dataset consists of 1,812 tweets, which, after preprocessing, are divided into 1,449 training data and 363 testing data. The analysis results show that 87.9% (319 tweets) are neutral, 5.2% (19 tweets) are negative, and 6.9% (25 tweets) are positive, indicating that the public does not yet have a definitive stance on this policy. The SVM method demonstrates good performance with an accuracy of 86.43%, positive precision of 83.33%, recall of 87.30%, and an F1-score of 85.27%. The impact of this research is expected to provide deeper insights into public perceptions of electric vehicle subsidy policies, thereby supporting more effective and well-targeted policymaking.

Keywords: *Electric Vehicles Subsidies, Sentiment Analysis, Support Vector Machine*

1. PENDAHULUAN

Pada tahun 2023 Jumlah kendaraan bermotor di Indonesia terus mengalami pertumbuhan yang pesat. Berdasarkan data dari Badan Pusat Statistik (BPS), terdapat lebih dari 132 juta unit sepeda motor, naik dari 106 juta unit pada tahun 2018. emisi gas rumah kaca, terutama karbon monoksida (CO), yang berasal dari kendaraan berbahan bakar fosil. Selain itu, jumlah mobil juga bertambah dari 14 juta menjadi 18 juta unit selama periode yang sama. Kenaikan ini mencerminkan tingginya ketergantungan masyarakat Indonesia pada kendaraan pribadi untuk mobilitas sehari-hari. Namun, lonjakan jumlah kendaraan bermotor juga menimbulkan tantangan

serius, termasuk kemacetan lalu lintas, peningkatan emisi gas rumah kaca, dan polusi udara yang memengaruhi kualitas lingkungan dan Kesehatan[1].

Peningkatan jumlah kendaraan bermotor di Indonesia memberikan dampak signifikan pada berbagai aspek lingkungan dan kesehatan masyarakat. Salah satu efek utamanya adalah peningkatan Gas rumah kaca, khususnya karbon monoksida (CO), yang dihasilkan oleh kendaraan dengan bahan bakar fosil. Emisi ini berkontribusi pada polusi udara yang memperburuk kualitas udara di perkotaan. Di kota besar seperti Jakarta, misalnya, polusi udara dari kendaraan bermotor menjadi salah satu faktor utama dalam menurunkan kualitas hidup masyarakat, dengan risiko meningkatnya penyakit pernapasan, jantung, hingga stroke. Selain itu, lonjakan jumlah kendaraan bermotor juga memperparah kemacetan lalu lintas, yang berdampak pada tingginya konsumsi bahan bakar dan pelepasan emisi tambahan akibat kendaraan yang diam terlalu lama dalam antrean lalu lintas. Kondisi ini tidak hanya merusak lingkungan tetapi juga menyebabkan kerugian ekonomi akibat waktu yang hilang dalam perjalanan dan konsumsi bahan bakar yang tidak efisien[2]. Upaya untuk mengatasi dampak ini termasuk mengembangkan kendaraan listrik sebagai alternatif ramah lingkungan dan mendorong pembangunan infrastruktur pendukung seperti stasiun pengisian daya listrik. Dengan langkah ini, diharapkan emisi dari sektor transportasi dapat ditekan, sekaligus memperbaiki kualitas udara dan kesehatan masyarakat[3].

Seiring dengan perkembangan teknologi dan perhatian terhadap isu lingkungan, kebijakan subsidi kendaraan listrik yang diterapkan oleh pemerintah Indonesia merupakan salah satu langkah penting dalam upaya mengurangi emisi karbon dan ketergantungan pada bahan bakar fosil. Subsidi ini bertujuan untuk mendorong adopsi kendaraan listrik yang ramah lingkungan, namun masih menghadapi tantangan dalam penerapannya, terutama terkait persepsi masyarakat yang beragam mengenai efektivitas dan manfaat dari kebijakan ini[4], [5].

Dengan tujuan mendorong percepatan adopsi kendaraan listrik dan mempercepat menuju transisi yang berkelanjutan, kemudian pemerintah Indonesia melalui Kementerian Keuangan mengeluarkan regulasi subsidi mobil listrik dan bus listrik. Untuk Kendaraan Bermotor Tenaga Baterai (KLBBB) yang memenuhi persyaratan, konsumen hanya membayar 1% Pajak Pertambahan Nilai (PPN), dari 11% yang seharusnya dibayar. Relaksasi ini tertuang dalam Peraturan Menteri Keuangan Nomor 38 Tahun 2023 tentang Pajak Pertambahan Nilai atas penyediaan KBLBB roda empat tertentu dan bus KBLBB tertentu yang dibiayai pemerintah. Syarat untuk mendapatkan subsidi ini adalah kendaraan harus memiliki Tingkat Komponen Nasional (TKDN) minimal 40%[6].

Aplikasi x merupakan salah satu platform media sosial yang paling banyak digunakan di Indonesia. Sebagai sarana komunikasi digital, Aplikasi x memberikan kebebasan kepada penggunanya untuk berbagi informasi, berpendapat, dan mengikuti topik-topik terkini dalam berbagai bentuk, termasuk *tweet*, *retweet*, dan komentar[7]. Setiap pengguna dapat menyampaikan opini mereka tentang berbagai isu, termasuk kebijakan pemerintah, seperti subsidi kendaraan listrik, dalam bentuk *tweet* singkat yang bisa diakses oleh publik secara langsung. Selain menjadi ruang untuk berbagi pendapat, Aplikasi x juga menjadi sumber data yang sangat berharga untuk analisis sentimen. Banyak penelitian yang telah memanfaatkan data dari Aplikasi x untuk mengevaluasi opini publik tentang isu-isu tertentu, karena platform ini memungkinkan pengumpulan data secara real-time dari berbagai kalangan masyarakat. Melalui proses *Crawling* yang dilakukan mulai dari tanggal 18 Oktober 2024 sampai 25 November 2024 yang menghasilkan *output* berupa *tweets* sejumlah 2139, *tweet-tweet* terkait dikumpulkan untuk kemudian dianalisis lebih lanjut, sehingga menghasilkan wawasan yang dapat digunakan untuk memahami persepsi masyarakat terhadap suatu kebijakan atau topik tertentu[8].

Analisis sentimen adalah proses untuk mengidentifikasi opini atau perasaan yang terkandung dalam teks, dengan tujuan untuk mengklasifikasikan sentimen menjadi Positif, negatif, ataupun netral. Pada platform media sosial seperti Aplikasi X, analisis sentimen berguna untuk menilai opini publik terkait isu-isu atau kebijakan tertentu. Beragam metode pengolahan bahasa alami (Natural Language Processing/NLP) dan algoritma pembelajaran mesin, seperti Support Vector Machine (SVM) dan Naive Bayes, sering digunakan dalam analisis ini. Penelitian oleh Purnama et al.[9] menunjukkan bahwa analisis sentimen dapat memberikan pemahaman yang lebih mendalam mengenai reaksi masyarakat terhadap kebijakan pemerintah, termasuk kebijakan transportasi.

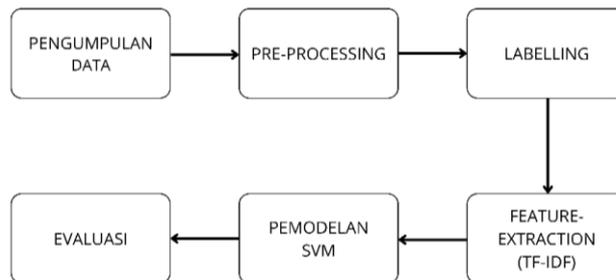
Penelitian sebelumnya, seperti "Analisis Sentimen Kebijakan Pemberian Subsidi Motor Listrik Menggunakan Metode Support Vector Machine," hanya berfokus pada sentimen masyarakat terhadap subsidi motor listrik sebagai salah satu jenis kendaraan yang mendapat subsidi. Meskipun penelitian tersebut memberikan wawasan penting tentang penerimaan publik terhadap subsidi motor listrik, ruang lingkupnya terbatas pada satu jenis kendaraan[10]. Penelitian ini, sebagai pengembangan lebih lanjut, memperluas cakupan dengan menganalisis sentimen masyarakat terhadap berbagai jenis kendaraan listrik yang mendapat subsidi, termasuk mobil listrik dan kendaraan lainnya. Dengan lingkup yang lebih luas, penelitian ini bertujuan untuk memberikan gambaran yang lebih komprehensif tentang persepsi masyarakat terhadap kebijakan subsidi kendaraan listrik secara keseluruhan. Pendekatan yang digunakan tetap berbasis algoritma Support Vector Machine (SVM), yang dikenal memiliki performa tinggi dalam analisis sentimen berbasis teks.

Berdasarkan penjelasan di atas, Penelitian ini bertujuan untuk menganalisis sentimen atau opini masyarakat Indonesia mengenai kebijakan subsidi kendaraan listrik. Data sentimen diperoleh dari media sosial Aplikasi X dan diklasifikasikan ke dalam tiga kategori yakni, sentimen positif, negatif, dan netral dengan menggunakan algoritma *Support Vector Machine*.

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Tahapan penelitian adalah proses atau langkah-langkah yang dilakukan secara sistematis untuk menyelesaikan penelitian dan mencapai tujuan yang telah ditetapkan. Gambar 1 di bawah ini menunjukkan diagram yang diajukan dalam penelitian ini.



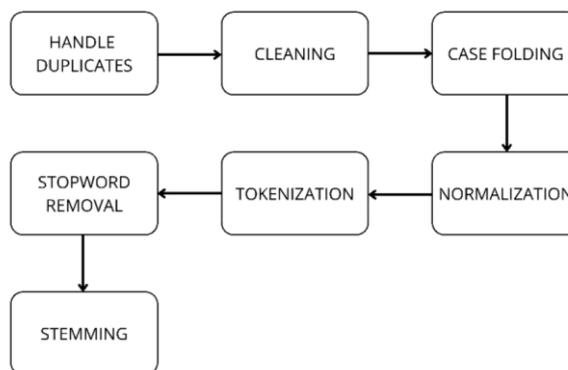
Gambar 1. Diagram alur penelitian
Sumber: Br Sinulingga & Sitorus, 2024

2.2. Crawling Data

Dalam penelitian ini, data dikumpulkan menggunakan metode *Crawling* data, yaitu teknik otomatis untuk mengambil data dari platform digital, khususnya aplikasi X [11]. Proses *Crawling* dilakukan dengan memanfaatkan *Application Programming Interface* (API) dari platform tersebut untuk mengakses data unggahan pengguna yang relevan. Kata kunci yang digunakan dalam pencarian meliputi “subsidi kendaraan listrik”, “motor listrik” dan “mobil listrik” guna memastikan data yang diperoleh sesuai dengan topik penelitian. Data yang diambil mencakup teks unggahan, waktu unggahan, *username*,serta atribut lain yang relevan untuk analisis sentimen[6]. Proses pengumpulan data dilakukan secara sistematis agar data yang terkumpul dapat mendukung tujuan penelitian dengan optimal.

2.3. Preprocessing Text

Tahap *preprocessing* dilakukan guna pengolahan data mentah menjadi koleksi data yang siap digunakan. Secara garis besar terdapat dua tahapan *preprocessing* dalam penelitian ini, yakni *preprocessing* secara umum dan *text processing*. *Preprocessing* secara umum dilakukan dengan tujuan pembersihan data dari jumlah data yang ganda (*handling duplicate*) sementara *text processing* dilakukan untuk seleksi data yang akan diproses pada dokumen [12].Beberapa tahapan *preprocessing* yang dilakuakan pada penelitian ini dapat dilihat pada Gambar 2 berikut.



Gambar 2. Tahap *preprocessing*
Sumber : Br Sinulingga & Sitorus, 2024

Berikut ini merupakan tahap *preprocessing* sesuai dengan Gambar 2 :

- a) *Handle duplicates*: dilakukan untuk menghapus data yang berulang dalam dataset, sehingga analisis lebih akurat dan efisien.
- b) *Cleaning*: menghapus elemen yang tidak relevan dalam data, seperti *Uniform Resource Locator* (URL), *mention*, *hashtag*, angka, dan simbol, untuk meningkatkan kualitas data yang akan dianalisis.
- c) *Case folding*: mengubah semua teks menjadi huruf kecil untuk memastikan konsistensi data dan mengurangi variasi yang disebabkan oleh perbedaan penggunaan huruf kapital.
- d) *Normalization*: Mengembalikan bentuk penulisan kata sesuai dengan Kamus Besar Bahasa Indonesia (KBBI).
- e) *Tokenization*: memecah teks menjadi unit-unit kecil seperti kata, frasa, atau kalimat, yang disebut token, untuk memudahkan analisis lebih lanjut.
- f) *Stopword removal*: menghilangkan kata-kata yang tidak relevan. Kata-kata tersebut tidak memberikan kontribusi signifikan terhadap informasi dalam kalimat. Contohnya adalah kata hubung “yang”, “akan”, “di”, “pada”, dan “dan”. Dengan menghapus kata-kata ini, informasi dalam kalimat tetap dapat dipertahankan tanpa pengaruh yang signifikan [11].
- g) *Stemming*: mengubah kata-kata menjadi bentuk dasarnya (*root word*) dengan menghapus akhiran atau awalan, untuk mengurangi variasi kata yang tidak perlu, seperti mengubah “berjalan” menjadi “jalan”..

2.4. Labelling

Pelabelan data adalah tahap di mana dataset dikelompokkan ke dalam kategori-kategori tertentu. Ini merupakan langkah awal dalam pekerjaan terkait *machine learning*. Dalam analisis sentimen, fokus utamanya adalah mengelompokkan teks dalam kalimat atau dokumen ke dalam kategori positif atau negatif, serta menilai opini yang terkandung dalam teks tersebut [13].

2.5. Weighting

TF-IDF (*Term Frekuensi-Invers Dokumen Frekuensi*) adalah metode yang digunakan untuk menghitung nilai bobot setiap kata yang diekstraksi. Metode ini digunakan untuk menghitung kata-kata umum dalam pengambilan informasi [10]. Model pembobotan TF-IDF merupakan metode yang mengintegrasikan model frekuensi kata (TF) dan model aliran teks (IDF). Frekuensi kemunculan (TF) adalah metode penghitungan jumlah kemunculan kata-kata yang dianggap umum dan tidak relevan dalam sebuah teks [10].

2.6. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. Algoritma ini bekerja dengan mencari hyperplane optimal yang memisahkan data ke dalam dua kelas dengan margin maksimum [14]. Margin ini didefinisikan sebagai jarak antara hyperplane dan data titik terdekat dari masing-masing kelas, yang disebut support vector [15]. Dalam kasus data non-linier, SVM menggunakan fungsi kernel seperti linear kernel, polynomial kernel, dan radial basis function (RBF) untuk memetakan data ke dalam dimensi yang lebih tinggi agar dapat dipisahkan dengan lebih baik. SVM memiliki keunggulan dalam menangani data berdimensi tinggi dan sering digunakan untuk analisis teks seperti analisis sentimen karena menghasilkan akurasi yang tinggi dalam memprediksi polaritas sentimen, baik positif, negatif, maupun netral [16].

3. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil penelitian terkait analisis sentimen pengguna aplikasi x terhadap kebijakan subsidi kendaraan listrik dengan memanfaatkan algoritma *Support Vector Machine* (SVM), yang disusun berdasarkan tahapan yang telah diuraikan pada bagian sebelumnya.

3.1. Pengumpulan Data

Pada penelitian ini, data dikumpulkan menggunakan Teknik *Crawling* data yang diimplementasikan menggunakan Bahasa pemrograman Python. Untuk melakukan *Crawling*, digunakan *library* Python seperti *tweet* Harvesting, *Node.js*, dan *Pandas* [6]. *Keyword* yang digunakan adalah “subsidi kendaraan listrik”, “motor listrik” dan “mobil listrik”. Proses pengumpulan data dilakukan dari tanggal 18 Oktober 2024 sampai 25 November 2024 dengan total perolehan *tweets* sejumlah 2,139 *tweets* yang kemudian disimpan dalam format CSV, jumlah data hasil *crawling* untuk setiap kata kunci dapat dilihat pada Tabel 1. dibawah ini.

Tabel 1. Data hasil *crawling* berdasarkan kata kunci

kata kunci	jumlah tweet
Subsidi kendaraan listrik	959
Mobil listrik	610
Motor listrik	586
Total	2,139

3.2. Preprocessing

Pada tahap ini dilakukan pengolahan terhadap hasil *crawling* yang formatnya tidak terstruktur sehingga menjadi kata dasar. *Preprocessing* dilakukan memakai bantuan library di bahasa pemrograman Python menggunakan Google Colaboratory. Hal pertama yang dilakukan adalah melakukan *handle duplicates*, *cleaning*, *case folding*, *normalization*, *tokenization*, *stopword removal* serta *stemming* sehingga dapat menghasilkan data bersih dan siap untuk lanjut pada proses berikutnya[11].

3.2.1. Handle Duplicates

Bertujuan untuk menjamin kualitas data yang digunakan dalam analisis tetap terjaga dengan menghilangkan baris data yang identik atau berulang. Data duplikat dapat terjadi karena kesalahan dalam proses pengumpulan data, seperti pengunduhan berulang atau redundansi dalam sistem pencatatan[17]. Tabel 2 berikut menunjukkan jumlah tweet sebelum dan sesudah dilakukan proses *handle duplicates*.

Tabel 2. Jumlah data sebelum dan sesudah proses *handle duplicates*

sebelum proses <i>handle duplicates</i>	sesudah proses <i>handle duplicates</i>
2,139	1,812

3.2.2. Cleaning

Cleaning data bertujuan untuk membersihkan dataset dari elemen yang tidak relevan, seperti *retweet*, URL, tanda baca, atau karakter khusus, sehingga data lebih terstruktur dan siap digunakan dalam analisis. Proses ini memastikan data berkualitas, konsisten, dan bebas dari *noise* yang dapat memengaruhi hasil analisis atau pelatihan model. Tabel 3 dibawah merupakan hasil dari tahapan *cleaning*.

Tabel 3. Tahap *cleaning*

sebelum <i>cleaning</i>	sesudah <i>cleaning</i>
@Svarnadiputera @menghanyurkan Astra internasional nanti merajuk karena penjual lesu lalu meminta pemerintah menerbitkan pengurangan pajak kendaraan/meminta subsidi seperti mobil/motor listrik	Astra internasional nanti merajuk karena penjual lesu lalu meminta pemerintah menerbitkan pengurangan pajak kendaraanmeminta subsidi seperti mobilmotor listrik

3.2.3. Case Folding

Case folding menghasilkan teks yang konsisten dengan mengubah seluruh huruf menjadi format kecil (*lowercase*). Proses ini memastikan data lebih seragam dan menghindari perbedaan makna yang disebabkan oleh penggunaan huruf besar dan kecil, sehingga mempermudah analisis lanjutan seperti tokenisasi dan pengolahan teks. Tabel 4 berikut merupakan simulasi hasil dari tahap *case folding*.

Tabel 4. *Case folding*

sebelum <i>case folding</i>	sesudah <i>case folding</i>
Astra internasional nanti merajuk karena penjual lesu lalu meminta pemerintah menerbitkan pengurangan pajak kendaraanmeminta subsidi seperti mobilmotor listrik	astra internasional nanti merajuk karena penjual lesu lalu meminta pemerintah menerbitkan pengurangan pajak kendaraanmeminta subsidi seperti mobilmotor listrik

3.2.4. Normalization

Normalization menghasilkan teks yang lebih standar dengan mengganti kata-kata tidak baku, singkatan, atau istilah slang menjadi bentuk baku atau formal. Proses ini membantu meningkatkan konsistensi data dan

memastikan setiap kata dapat dikenali dan diproses dengan baik oleh model analisis, sehingga hasil analisis menjadi lebih akurat dan bermakna[18],dapat dilihat pada Tabel 5.

Tabel 5. Tahap *normalization*

sebelum <i>normalization</i>	sesudah <i>normalization</i>
astra internasional nanti merajuk karena penjual lesu lalu meminta pemerintah menerbitkan pengurangan pajak kendaraanmeminta subsidi seperti mobilmotor listrik	astra internasional nanti merajuk karena penjual lesu lalu meminta pemerintah menerbitkan pengurangan pajak kendaraanmeminta subsidi seperti mobilmotor listrik

3.2.5. *Tokenization*

Tokenization menghasilkan teks yang terpecah menjadi unit-unit kecil, seperti kata, frasa, atau kalimat, yang disebut token. Proses ini memudahkan analisis data teks dengan memisahkan setiap elemen sehingga dapat diproses secara individu dalam tahap lanjutan, seperti penghitungan frekuensi kata atau analisis sentimen. Hasil dari tahapan *tokenization* ditunjukkan pada Tabel 6.

Tabel 6. Hasil *tokenization*

sebelum <i>tokenization</i>	sesudah <i>tokenization</i>
['astra','internasional','nanti','merajuk','karena','penjual','lesu','lalu','meminta','pemerintah','menerbitkan','pengurangan','pajak','kendaraanmeminta','subsidi','seperti','mobilmotor','listrik']	['astra','internasional','nanti','merajuk','karena','penjual','lesu','lalu','meminta','pemerintah','menerbitkan','pengurangan','pajak','kendaraanmeminta','subsidi','seperti','mobilmotor','listrik']

3.2.6. *Stopword Removal*

Stopword removal menghasilkan teks yang lebih relevan dengan menghapus kata-kata umum yang tidak memiliki makna signifikan, seperti "dan," "yang," atau "dari." Proses ini membantu mengurangi *noise* pada data dan memastikan analisis hanya berfokus pada kata-kata yang benar-benar penting untuk konteks tertentu.Tabel 7 merupakan output dari proses *Stopword removal*.

Tabel 7. Hasil *stopword removal*

sebelum <i>stopword removal</i>	sesudah <i>stopword removal</i>
['astra','internasional','nanti','merajuk','karena','penjual','lesu','lalu','meminta','pemerintah','menerbitkan','pengurangan','pajak','kendaraanmeminta','subsidi','seperti','mobilmotor','listrik']	['astra','internasional','nanti','merajuk','karena','penjual','lesu','lalu','meminta','pemerintah','menerbitkan','pengurangan','pajak','kendaraanmeminta','subsidi','seperti','mobilmotor','listrik']

3.2.7. *Stemming*

Stemming menghasilkan kata-kata dalam bentuk dasar (*root word*) dengan menghapus imbuhan seperti awalan, akhiran, atau sisipan. Proses ini membantu menyederhanakan teks sehingga kata-kata dengan makna serupa, seperti "berlari," "lari," atau "pelari," direpresentasikan sebagai satu bentuk dasar, yaitu "lari." Hal ini meningkatkan konsistensi data untuk analisis lebih lanjut. Hasil dari tahap *stemming* dapat dilihat pada Tabel 8 berikut.

Tabel 8. Hasil *stemming*

sebelum <i>stemming</i>	sesudah <i>stemming</i>
['astra','internasional','nanti','merajuk','karena','penjual','lesu','lalu','meminta','pemerintah','menerbitkan','pengurangan','pajak','kendaraanmeminta','subsidi','seperti','mobilmotor','listrik']	['astra','internasional','rajuk','jual','lesu','perintah','terbit','kurang','pajak','kendaraanmeminta','subsidi','mobilmotor','listrik']

3.3. *Labelling*

Pelabelan data merupakan suatu tahapan untuk menentukan respon *tweet* di kumpulan data. Setiap *tweet* dalam file CSV dianalisis menggunakan TextBlob, sebuah pustaka Python yang memudahkan akses ke berbagai fungsi Pemrosesan Bahasa Alami (NLP). TextBlob menyediakan antarmuka yang sederhana untuk berbagai aktivitas, seperti ekstraksi Frasa Nomina, penandaan Part-of-Speech (PoS), serta analisis sentimen, dan masih

banyak lagi(10). Hasil akhir menunjukkan jumlah *tweet* yang terklasifikasi sebagai positif, negatif, atau netral, seperti yang dapat dilihat pada Tabel 9 berikut.

Tabel 9. Hasil *labelling data*

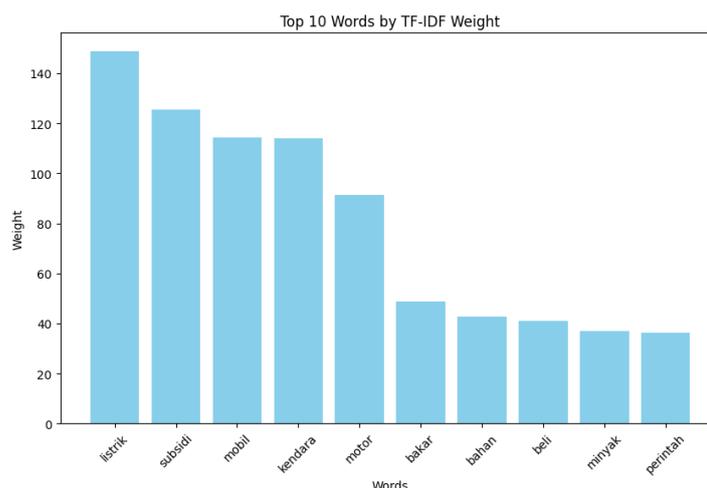
Sentimen	Jumlah
Netral	1,641
Positif	97
Negatif	74

3.4. Pembobotan TF-IDF

Pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah teknik yang digunakan dalam pemrosesan teks untuk menilai seberapa penting suatu kata dalam sebuah dokumen relatif terhadap seluruh kumpulan dokumen. Proses ini disebut ekstraksi fitur atau vektorisasi, yang menjadi langkah krusial dalam analisis teks. Pada tahap ini, kata-kata diubah menjadi representasi numerik, baik dalam bentuk bilangan bulat maupun desimal, sesuai dengan algoritma pembelajaran mesin yang digunakan. Dalam penelitian ini, dua metode populer untuk melakukan vektorisasi teks adalah Count Vectorizer dan TF-IDF Vectorizer[19].

TF-IDF terdiri dari dua langkah utama. Langkah pertama adalah menghitung seberapa sering suatu kata muncul dalam setiap dokumen. Langkah kedua adalah menentukan bobot kata tersebut di seluruh dokumen yang ada. Perhitungan TF-IDF membantu mengukur pentingnya suatu kata dalam dokumen tertentu dengan membandingkan kehadirannya di keseluruhan kumpulan dokumen[20]. Di bawah ini ditampilkan sepuluh kata yang paling sering muncul. Hasil ini divisualisasikan menggunakan grafik batang, yang menampilkan kata-kata tersebut pada sumbu x (kata) dan frekuensinya pada sumbu y (frekuensi), yang terlihat pada gambar 4.

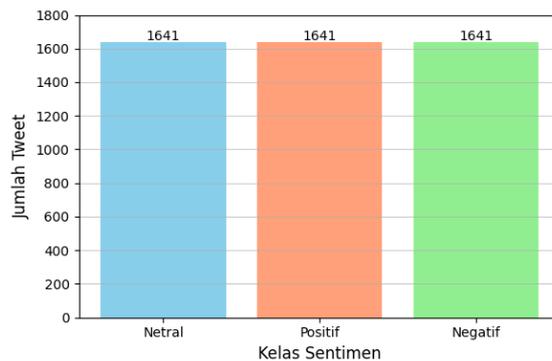
Proses vektorisasi teks dilakukan dengan menggunakan *TfidfVectorizer* dari pustaka *scikit-learn*, yang mengonversi teks yang terdapat pada kolom 'NORMALIZED' menjadi representasi numerik. Representasi ini dihitung berdasarkan bobot kata-kata yang didasarkan pada rumus TF-IDF (*Term Frequency-Inverse Document Frequency*). Bobot yang dihasilkan akan mencerminkan seberapa penting suatu kata dalam konteks suatu dokumen dibandingkan dengan seluruh kumpulan dokumen. Setelah memperoleh representasi numerik tersebut, teknik SMOTE (*Synthetic Minority Over-sampling Technique*) digunakan untuk menangani ketidakseimbangan kelas yang ada pada data. SMOTE bertujuan untuk meningkatkan jumlah sampel pada kelas minoritas dengan cara membuat sampel sintetis agar distribusi kelas menjadi lebih seimbang. Proses ini diharapkan dapat meningkatkan akurasi model klasifikasi dalam memprediksi kelas minoritas yang sebelumnya kurang ter-representasi, seperti yang dapat dilihat pada Tabel 10 dan Gambar 5 yang menunjukkan distribusi jumlah data positif dan negatif sebelum dan setelah penerapan SMOTE.



Gambar 4. Kata dengan tingkat kemunculan tertinggi

Tabel 10. Hasil implementasi SMOTE

	Sebelum melakukan SMOTE	Sesudah melakukan SMOTE
Netral	1,641	1,641
Positif	97	1,641
Negatif	74	1,641



Gambar 5. Hasil data setelah implemmentasi SMOTE

3.5. Model Support Vector Machine (SVM)

Tujuan penelitian adalah mengimplementasikan algoritma SVM pada sentimen di media sosial aplikasi x terkait topik pemindahan ibu kota untuk mengetahui seberapa besar tingkat akurasi, recall dan precision dari analisis sentimen yang dilakukan. Komputasi model SVM terhadap bobot tiap kata dalam sentimen sangat menentukan kelas negatif dan positif, serta akurasi yang dihasilkan dari proses pemodelan[12].

Support Vector Machine (SVM) adalah algoritma yang digunakan dalam penelitian ini untuk mengklasifikasikan sentimen pengguna aplikasi terhadap kebijakan subsidi kendaraan listrik menjadi tiga kategori yakni positif, negatif, dan netral. SVM bekerja dengan mencari hyperplane optimal yang memisahkan data berdasarkan fitur yang dihasilkan, seperti representasi TF-IDF dari teks[6]. Model ini dipilih karena efektivitasnya dalam menangani data berukuran besar dan kompleks, terutama pada tugas klasifikasi teks, seperti yang telah dibuktikan oleh Rahmawati et al. dalam penelitian mereka pada analisis sentimen media sosial menggunakan SVM[21].

Hasil dari penerapan model SVM menunjukkan akurasi keseluruhan sebesar 90% yang dapat dilihat pada Gambar 6 dibawah, meskipun kinerja pada kelas minoritas (positif dan negatif) memerlukan perbaikan. Hal ini sejalan dengan temuan Pratama et al., yang menyebutkan bahwa SVM memiliki kekuatan pada data dengan dimensi tinggi, tetapi performanya dapat bergantung pada distribusi data antar kelas[22].

```
[ ] # Evaluasi performa model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.2f}')
```

↔ Accuracy: 0.90

Gambar 6. Akurasi model

3.6. Evaluasi (Confusion Matrix)

Untuk mengevaluasi performa model yang telah dibangun, digunakan beberapa metrik evaluasi seperti accuracy, precision, recall, dan f1-score. Hasil evaluasi model SVM berdasarkan data testing[10] yang telah dilakukan dapat dilihat pada Tabel 11 berikut.

Tabel 11. Confusion matrix dan hasil evaluasi model

Kelas actual	Negative(0)	Netral(1)	Positif(2)	Total
Negatif	19	0	0	19
Netral	0	319	0	319
Positif	0	0	25	25
total	19	319	25	363

Dari Tabel 11 dapat dilihat bahwa distribusi kelas yang diprediksi oleh model tidak seimbang, dengan jumlah prediksi kelas netral yang jauh lebih banyak dibandingkan kelas negatif dan positif. Hal ini mengarah pada nilai accuracy yang cukup tinggi, namun dengan nilai precision, recall, dan f1-score yang tidak merata di setiap kelas, sesuai Tabel 12 dibawah ini.

Table 12. Hasil Evaluasi Model SVM

Metrik	Nilai
Accuracy	0.90
Precision	Negatif: 1.00, Netral: 0.89, Positif: 1.00
Recall	Negatif: 0.11, Netral: 1.00, Positif: 0.16
F1-score	Negatif: 0.19, Netral: 0.94, Positif: 0.28

Hasil diatas menunjukkan bahwa model memberikan performa terbaik dalam mengklasifikasikan kelas netral, dengan nilai *accuracy* mencapai 90%. Namun, untuk kelas negatif dan positif, performa model cukup rendah, terutama pada nilai *recall* yang sangat rendah pada kelas negatif dan positif. Hal ini menunjukkan bahwa meskipun model memiliki tingkat ketepatan yang tinggi pada beberapa kelas, kemampuannya dalam mengidentifikasi kelas yang kurang dominan masih perlu diperbaiki.

4. KESIMPULAN

Berdasarkan penelitian ini, analisis sentimen terhadap kebijakan subsidi kendaraan listrik di Indonesia menunjukkan bahwa mayoritas opini masyarakat bersifat netral. Dari total 363 tweet yang dianalisis, sekitar 319 tweet (87.9%) termasuk dalam kategori sentimen netral, sementara 19 tweet (5.2%) memiliki sentimen negatif dan 25 tweet (6.9%) memiliki sentimen positif. Hal ini mengindikasikan bahwa masyarakat belum memiliki pandangan yang tegas mengenai kebijakan subsidi kendaraan listrik, dengan kecenderungan lebih banyak bersikap netral atau tidak memberikan penilaian kuat.

Metode Support Vector Machine (SVM) terbukti efektif dalam mengklasifikasikan sentimen pada data tweet ini dengan tingkat akurasi yang tinggi, yaitu sebesar 90%. Meskipun demikian, hasil evaluasi menunjukkan bahwa nilai *precision* dan *recall* untuk sentimen negatif dan positif relatif rendah, yang menandakan bahwa model kurang optimal dalam mendeteksi sentimen negatif dan positif secara tepat. Perbaikan lebih lanjut dalam model, seperti penyesuaian parameter dan penggunaan teknik lain seperti pengurangan dimensi atau fine-tuning model, dapat membantu meningkatkan performa dalam mengidentifikasi berbagai sentimen secara lebih akurat.

Penelitian ini menggunakan 2,139 dataset yang telah di bersihkan menjadi 1,812 tweet setelah dilakukan pembersihan terhadap data yang tidak relevan. Dengan menggunakan dataset yang terfilter ini, model SVM mampu menghasilkan evaluasi kinerja yang cukup baik, namun masih perlu pengembangan lebih lanjut untuk mencapai performa yang lebih optimal dalam klasifikasi sentimen pada topik-topik serupa.

Sebagai rekomendasi untuk pengembangan lebih lanjut, penelitian mendatang dapat mempertimbangkan penggunaan dataset yang lebih besar dan lebih beragam untuk meningkatkan generalisasi model terhadap berbagai sentimen. Selain itu, eksplorasi algoritma lain, seperti Transformer-based models (contohnya BERT), dapat memberikan hasil yang lebih baik dalam memahami konteks sentimen secara mendalam. Peningkatan performa juga dapat dicapai dengan mengintegrasikan teknik feature engineering yang lebih kompleks, seperti penambahan fitur berbasis emosi atau konteks historis pengguna. Pengembangan aplikasi berbasis web atau mobile untuk memantau sentimen secara real-time juga disarankan, sehingga hasil penelitian dapat dimanfaatkan langsung oleh pemerintah atau pemangku kebijakan dalam mengevaluasi kebijakan publik. Terakhir, kolaborasi dengan ahli kebijakan publik dan psikologi sosial dapat memberikan sudut pandang yang lebih holistik dalam menganalisis hasil sentimen masyarakat terhadap kebijakan tertentu.

DAFTAR PUSTAKA

- [1] Auto2000, "Mengenal Program Subsidi Mobil Listrik dengan Lengkap," Auto2000. [Online]. Available: <https://auto2000.co.id/berita-dan-tips/subsidi-mobil-listrik>. Accessed: Dec. 11, 2024.
- [2] A. Herdito, "Dampak Kendaraan Bermotor dalam Polusi Udara," MERTANI. [Online]. Available: <https://www.mertani.co.id/post/dampak-kendaraan-bermotor-dalam-polusi-udara>. Accessed: Dec. 18, 2024.
- [3] M. P. Lingkungan, F. Pascasarjana, and U. G. Mada, "Kajian Pencemaran Lingkungan terhadap Kesehatan Masyarakat akibat Gas Buangan CO Kendaraan Bermotor di Kawasan Universitas Gadjah Mada," vol. 5, no. 1, pp. 16–38, 2024.
- [4] M. I. Nurhidayat, "Persepsi Masyarakat Terhadap Karakteristik Kepemilikan Kendaraan Listrik Berbasis Baterai (Studi Kasus di Jakarta, Indonesia)." [Online]. Available: <https://repository.its.ac.id/99103/>. Accessed: Jan. 03, 2025.
- [5] K. A. Simanjuntak, M. Koyimatu, and Y. P. Ervanisari, "Analisis Perubahan Opini Publik Terhadap

- Kendaraan Listrik di Indonesia Melalui Komentar YouTube: Pendekatan Topic Modeling BERTopic,” 2024, doi: 10.37817/jurnalinovasikewirausahaan.v1i3.3789.
- [6] A. S. Putra, D. Anubhakti, and L. L. Hin, “Penerapan Algoritma K-Nearest Neighbor Untuk Analisis Sentimen Pengguna Twitter Terhadap Kebijakan Subsidi Kendaraan Listrik,” *Semin. Nas. Mhs. Fak. Teknol. Inf.*, vol. 2, no. 2, pp. 736–744, 2023. [Online]. Available: <http://senafti.budiluhur.ac.id/index.php/senafti/article/view/871>.
- [7] G. Robbini, “X Lebih dari Sekedar Cuitan: Jelajahi Dunia Baru dalam Komunitas.” [Online]. Available: <https://www.kompasiana.com/ginarobbini/670b963b34777c486d720613/x-lebih-dari-sekedar-cuitan-jelajahi-dunia-baru-dalam-komunitas>. Accessed: Jan. 03, 2025.