

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

# EVALUASI SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN SUBSIDI KENDARAAN LISTRIK DI INDONESIA DENGAN PENDEKATAN INSET LEXICON, WORD EMBEDDING, DAN ALGORITMA SUPPORT VECTOR MACHINE

<sup>1</sup>Ridwan\*, <sup>2</sup>Hendarman Lubis

<sup>1</sup>Teknik Informatika, Sekolah Tinggi Teknologi Pekerjaan Umum, Jl. Laksamana Malahayati No.6,  
RT.6/RW.12, Pd. Bambu, Kec. Duren Sawit, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta,  
Indonesia

<sup>2</sup>Informatika, Ilmu Komputer, Universitas Bhayangkara Jakarta Raya, Jl. Harsono RM No.67,  
Ragunan, Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta, Indonesia

\*Correspondent Author - e-mail: [ridwansamm@sttpu.ac.id](mailto:ridwansamm@sttpu.ac.id)

<sup>2</sup>e-mail: [ridwansamm@sttpu.ac.id](mailto:ridwansamm@sttpu.ac.id), [hendarman.lubis@dsn.ubharajaya.ac.id](mailto:hendarman.lubis@dsn.ubharajaya.ac.id)

## Abstrak

Kebijakan subsidi kendaraan listrik di Indonesia merupakan salah satu upaya pemerintah dalam mendorong penggunaan energi ramah lingkungan serta mengurangi emisi karbon. Namun, implementasi kebijakan ini menimbulkan beragam tanggapan dari masyarakat yang dapat dianalisis melalui media sosial. Penelitian ini bertujuan untuk mengevaluasi sentimen masyarakat terhadap kebijakan subsidi kendaraan listrik di Indonesia menggunakan pendekatan InSet Lexicon, Word Embedding (Word2Vec), dan algoritma Support Vector Machine (SVM). Data penelitian diperoleh dari media sosial Twitter melalui proses crawling berdasarkan kata kunci yang relevan dan menghasilkan sebanyak 1.000 data tweet. Tahapan penelitian meliputi preprocessing teks, pelabelan sentimen menggunakan InSet Lexicon, ekstraksi fitur menggunakan Word2Vec, serta klasifikasi menggunakan SVM. Hasil penelitian menunjukkan bahwa distribusi sentimen terdiri dari 45% positif, 35% negatif, dan 20% netral. Model klasifikasi yang dibangun menghasilkan performa dengan nilai accuracy sebesar 86%, precision 83%, recall 81%, dan F1-score 82%. Hasil ini menunjukkan bahwa pendekatan yang digunakan mampu mengklasifikasikan sentimen dengan baik. Selain itu, penggunaan Word Embedding terbukti meningkatkan kualitas representasi teks sehingga berdampak pada peningkatan performa model. Penelitian ini diharapkan dapat memberikan gambaran mengenai persepsi masyarakat serta menjadi bahan evaluasi bagi pengambil kebijakan publik.

**Kata kunci:** analisis sentimen, InSet Lexicon, Word2Vec, SVM, kendaraan listrik

## Abstract

*The electric vehicle subsidy policy in Indonesia is one of the government's efforts to promote environmentally friendly energy usage and reduce carbon emissions. However, the implementation of this policy has generated diverse public responses, which can be analyzed through social media platforms. This study aims to evaluate public sentiment toward the electric vehicle subsidy policy in Indonesia using the InSet Lexicon approach, Word Embedding (Word2Vec), and the Support Vector Machine (SVM) algorithm. The dataset was collected from Twitter through a crawling process based on relevant keywords, resulting in 1,000 tweets. The research stages include text preprocessing,*



DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

*sentiment labeling using InSet Lexicon, feature extraction using Word2Vec, and classification using SVM. The results show that sentiment distribution consists of 45% positive, 35% negative, and 20% neutral. The classification model achieved an accuracy of 86%, precision of 83%, recall of 81%, and an F1-score of 82%. These results indicate that the proposed approach is effective in classifying sentiment. Furthermore, the use of Word Embedding improves text representation quality, which contributes to better model performance. This study provides insights into public perception and can serve as a reference for evaluating public policies.*

**Keywords:** *sentiment analysis, InSet Lexicon, Word2Vec, Support Vector Machine, electric vehicles*

## 1 Pendahuluan (or Introduction)

Perkembangan teknologi transportasi global menunjukkan adanya pergeseran menuju penggunaan energi yang lebih ramah lingkungan sebagai respons terhadap meningkatnya dampak negatif penggunaan bahan bakar fosil, seperti pencemaran udara dan perubahan iklim. Kendaraan listrik menjadi salah satu solusi yang dinilai mampu mengurangi emisi gas rumah kaca serta meningkatkan efisiensi energi dalam sektor transportasi [1], [2]. Tren global juga menunjukkan peningkatan signifikan dalam adopsi kendaraan listrik sebagai bagian dari upaya transisi energi berkelanjutan [3].

Di Indonesia, pemerintah telah mengimplementasikan berbagai kebijakan strategis untuk mendorong penggunaan kendaraan listrik, salah satunya melalui pemberian subsidi pembelian kendaraan listrik. Kebijakan ini bertujuan untuk meningkatkan daya beli masyarakat serta mempercepat adopsi kendaraan listrik di tingkat nasional [4]. Selain itu, kebijakan tersebut juga diharapkan mampu mendukung pertumbuhan industri kendaraan listrik dalam negeri serta mengurangi ketergantungan terhadap bahan bakar fosil [5].

Meskipun kebijakan ini memiliki tujuan yang positif, implementasinya menimbulkan berbagai tanggapan dari masyarakat. Sebagian masyarakat memberikan respon positif karena kebijakan ini dinilai mampu memberikan manfaat ekonomi dan lingkungan. Namun, terdapat pula respon negatif yang berkaitan dengan harga kendaraan yang masih tinggi, keterbatasan infrastruktur pengisian daya, serta efektivitas distribusi subsidi [6]. Hal ini menunjukkan bahwa persepsi masyarakat terhadap kebijakan subsidi kendaraan listrik masih beragam dan perlu dianalisis secara lebih mendalam.

Dalam era digital saat ini, media sosial menjadi salah satu sumber utama dalam memperoleh opini publik. Platform seperti Twitter memungkinkan masyarakat untuk menyampaikan pendapat secara terbuka dan real-time, sehingga menjadi sumber data yang potensial dalam analisis sentimen [7], [8]. Data dari media sosial dapat dimanfaatkan untuk memahami persepsi masyarakat terhadap suatu kebijakan secara cepat dan luas.

Analisis sentimen merupakan salah satu teknik dalam Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks ke dalam kategori tertentu seperti positif, negatif, dan netral [9]. Metode ini telah banyak digunakan dalam berbagai penelitian untuk memahami persepsi publik terhadap isu sosial maupun kebijakan pemerintah [10].

Dalam implementasinya, analisis sentimen dapat dilakukan menggunakan pendekatan berbasis leksikon maupun pembelajaran mesin. Pendekatan leksikon menggunakan kamus sentimen untuk menentukan polaritas kata, sedangkan pembelajaran mesin memanfaatkan algoritma klasifikasi untuk meningkatkan akurasi [11]. InSet Lexicon merupakan salah satu kamus sentimen Bahasa Indonesia yang banyak digunakan karena kemampuannya dalam mengidentifikasi polaritas teks secara efektif [12].

Selain itu, representasi fitur menjadi faktor penting dalam analisis sentimen. Metode tradisional seperti TF-IDF memiliki keterbatasan dalam menangkap hubungan kontekstual antar kata. Oleh karena itu, digunakan metode Word Embedding seperti Word2Vec yang mampu merepresentasikan kata dalam bentuk vektor numerik berbasis konteks, sehingga hubungan semantik antar kata dapat dipahami dengan lebih baik [13].

Untuk proses klasifikasi, algoritma Support Vector Machine (SVM) dipilih karena memiliki kemampuan yang baik dalam menangani data berdimensi tinggi serta menghasilkan performa yang



DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

stabil. SVM bekerja dengan mencari hyperplane optimal untuk memisahkan kelas data sehingga menghasilkan akurasi yang tinggi dalam klasifikasi teks [14].

Beberapa penelitian sebelumnya menunjukkan bahwa kombinasi antara pendekatan leksikon, Word Embedding, dan algoritma klasifikasi mampu meningkatkan performa analisis sentimen. Selain itu, penggunaan Word Embedding terbukti memberikan hasil yang lebih baik dibandingkan metode konvensional dalam klasifikasi teks media sosial [15].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengevaluasi sentimen masyarakat terhadap kebijakan subsidi kendaraan listrik di Indonesia menggunakan pendekatan InSet Lexicon, Word Embedding (Word2Vec), dan algoritma Support Vector Machine. Penelitian ini diharapkan dapat memberikan gambaran yang komprehensif mengenai persepsi masyarakat serta menjadi bahan evaluasi dalam pengambilan kebijakan publik berbasis data.

## 2 Tinjauan Literatur (or Literature Review)

### 2.1. Analisis Sentimen

Analisis sentimen merupakan salah satu teknik dalam Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks ke dalam kategori tertentu seperti positif, negatif, dan netral. Teknik ini memungkinkan pengolahan data teks dalam jumlah besar secara otomatis untuk memahami persepsi publik terhadap suatu isu [9], [16]. Seiring dengan meningkatnya penggunaan media sosial, analisis sentimen menjadi semakin penting karena mampu menggambarkan opini masyarakat secara real-time.

Media sosial seperti Twitter menyediakan data opini yang sangat besar dan beragam. Informasi yang dihasilkan dari platform ini dapat dimanfaatkan untuk memahami persepsi masyarakat terhadap kebijakan publik maupun fenomena sosial lainnya [7], [17]. Oleh karena itu, analisis sentimen berbasis media sosial menjadi pendekatan yang relevan dalam penelitian ini.

### 2.2. Pendekatan Lexicon-Based (InSet Lexicon)

Pendekatan berbasis leksikon merupakan metode analisis sentimen yang menggunakan kamus kata untuk menentukan polaritas teks. Setiap kata dalam teks akan diberikan skor sentimen yang kemudian digunakan untuk menentukan sentimen keseluruhan dari suatu dokumen [11].

InSet Lexicon merupakan kamus sentimen Bahasa Indonesia yang dikembangkan untuk mengatasi keterbatasan leksikon berbahasa Inggris dalam analisis teks lokal. Pendekatan ini telah terbukti efektif dalam mengidentifikasi polaritas teks berbahasa Indonesia serta tidak memerlukan data pelatihan dalam jumlah besar [12], [18]. Oleh karena itu, metode ini sering digunakan untuk pelabelan awal dalam analisis sentimen.

### 2.3. Word Embedding (Word2Vec)

Word Embedding merupakan teknik representasi kata dalam bentuk vektor numerik yang mampu menangkap hubungan semantik antar kata. Word2Vec adalah salah satu metode yang paling populer dan bekerja dengan mempelajari konteks kemunculan kata dalam suatu korpus teks [13].

Penggunaan Word2Vec dalam analisis sentimen memberikan keunggulan dibandingkan metode tradisional seperti TF-IDF, karena mampu memahami hubungan kontekstual antar kata. Hal ini sangat penting dalam analisis data media sosial yang memiliki variasi bahasa yang tinggi [19]. Dengan representasi yang lebih baik, model klasifikasi dapat menghasilkan performa yang lebih optimal.

### 2.4. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin yang banyak digunakan dalam klasifikasi teks. Algoritma ini bekerja dengan mencari hyperplane optimal yang dapat memisahkan data ke dalam kelas-kelas tertentu dengan margin maksimum [14].

SVM memiliki keunggulan dalam menangani data berdimensi tinggi serta mampu menghasilkan performa yang stabil dalam berbagai kasus klasifikasi. Dalam analisis sentimen, SVM sering digunakan



DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

karena kemampuannya dalam menghasilkan akurasi yang tinggi, terutama ketika dikombinasikan dengan teknik ekstraksi fitur yang baik seperti Word Embedding [20].

## 2.5. Penelitian Terdahulu

Beberapa penelitian sebelumnya telah mengkaji analisis sentimen menggunakan berbagai metode dan pendekatan. Penelitian-penelitian tersebut menunjukkan bahwa analisis sentimen terhadap kebijakan publik dapat memberikan gambaran mengenai persepsi masyarakat secara luas [10], [21]. Selain itu, penggunaan media sosial sebagai sumber data dinilai efektif karena mampu menyediakan opini publik secara real-time.

Penggunaan metode Word Embedding seperti Word2Vec terbukti mampu meningkatkan performa model klasifikasi dibandingkan metode konvensional [19], [22]. Di sisi lain, pendekatan berbasis leksikon seperti InSet Lexicon efektif dalam melakukan pelabelan awal data tanpa memerlukan anotasi manual yang kompleks [18]. Beberapa studi juga menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki performa yang stabil dan akurat dalam klasifikasi teks [20], [23].

Untuk memberikan gambaran yang lebih jelas, berikut disajikan ringkasan penelitian terdahulu yang relevan dengan penelitian ini.

**Tabel 1. Penelitian Terdahulu**

No	Peneliti	Tahun	Metode	Dataset	Hasil
1	Pratama et al. [24]	2023	Naïve Bayes	Twitter	Sentimen publik beragam
2	Pratiwi et al. [22]	2025	Word2Vec + SVM	Media Sosial	Akurasi meningkat signifikan
3	Ridwan et al. [25]	2024	Lexicon-Based	Twitter	Efektif untuk pelabelan awal
4	Fransiscus & Girsang [26]	2023	BERT	Media Sosial	Akurasi tinggi, namun kompleks
5	Aini [27]	2024	Naïve Bayes	Review Produk	Klasifikasi cukup baik
6	Salim [28]	2023	Sentiment Analysis	Twitter	Mendukung pengambilan keputusan
7	Ali et al. [29]	2023	e-WOM Analysis	Online Data	Pengaruh opini digital signifikan
8	Latif & Ali [30]	2025	SWOT Analysis	Data Bisnis	Strategi berbasis data meningkat
9	Isnawati & Ali [31]	2023	Sistem Informasi	Data Sistem	Teknologi berpengaruh signifikan

## 3 Metode Penelitian (or Research Method)

Penelitian ini menggunakan pendekatan eksperimen dalam analisis sentimen berbasis text mining untuk mengevaluasi persepsi masyarakat terhadap kebijakan subsidi kendaraan listrik di Indonesia. Pendekatan yang digunakan merupakan kombinasi antara metode berbasis leksikon dan machine learning, yang bertujuan untuk meningkatkan akurasi dalam klasifikasi sentimen teks. Kombinasi ini memungkinkan pelabelan awal dilakukan secara otomatis menggunakan kamus sentimen, kemudian diperkuat dengan model klasifikasi berbasis algoritma.

Data yang digunakan dalam penelitian ini diperoleh dari media sosial Twitter (X) melalui proses crawling dengan kata kunci yang relevan seperti “subsidi kendaraan listrik”, “mobil listrik”, dan “motor listrik”. Data yang diperoleh berupa teks opini masyarakat yang kemudian disimpan dalam format terstruktur untuk dilakukan proses analisis lebih lanjut. Penggunaan media sosial sebagai sumber data dinilai efektif karena mampu merepresentasikan opini publik secara real-time dan dalam jumlah besar.



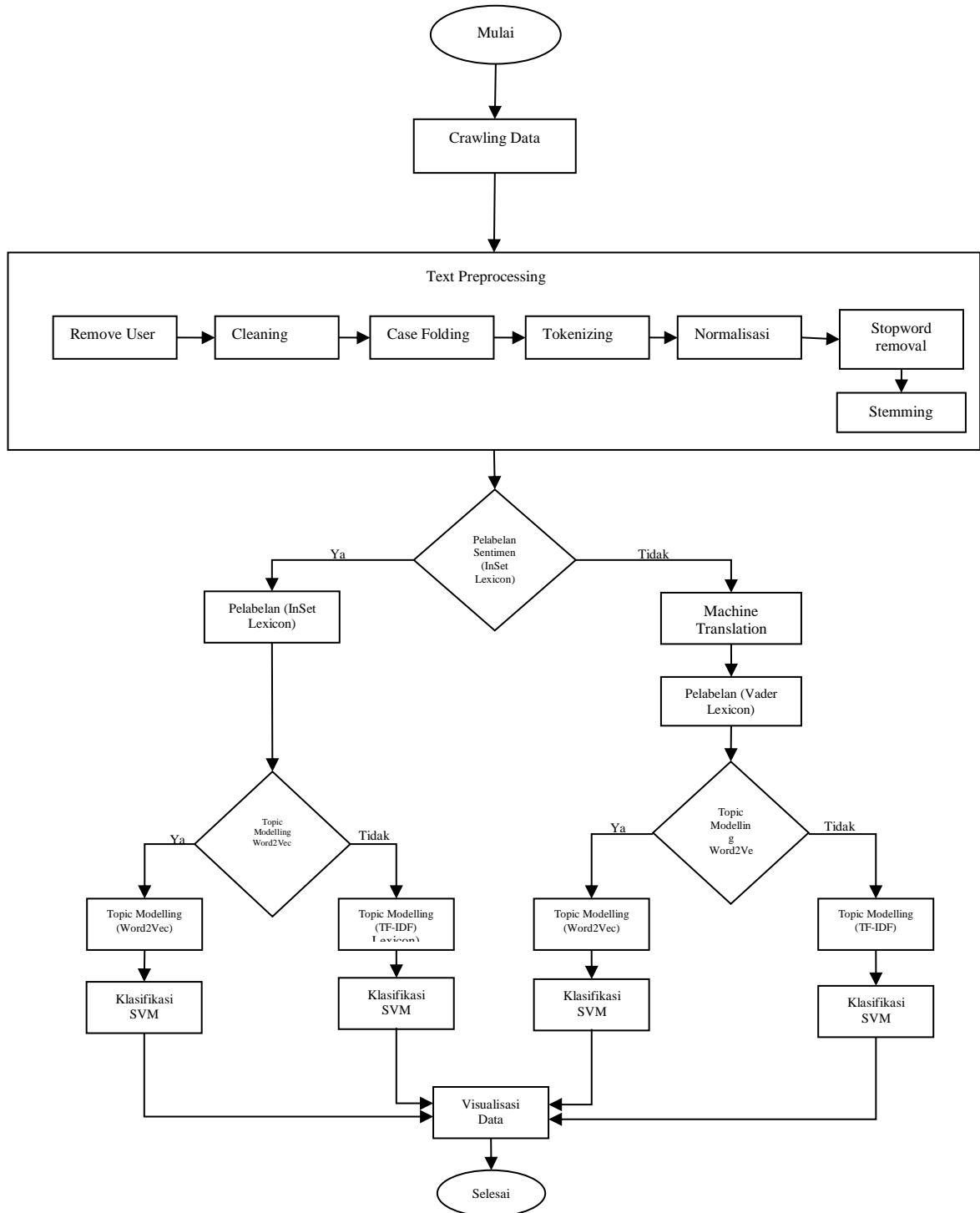
DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

Berikut alur proses penelitian berdasarkan metode penelitian yang digunakan :



**Gambar 1. Flowchart Penelitian**

### 3.1. Crawling Data

Tahap awal penelitian adalah pengumpulan data dari media sosial Twitter menggunakan teknik crawling. Data diambil berdasarkan kata kunci yang telah ditentukan dan dalam rentang waktu tertentu.



DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

Proses ini bertujuan untuk memperoleh data opini masyarakat dalam jumlah besar yang relevan dengan topik penelitian [7].

### 3.2. Text Preprocessing

Tahapan text preprocessing bertujuan untuk membersihkan data teks dari noise sehingga siap untuk dianalisis. Proses ini meliputi beberapa langkah utama, yaitu remove user, cleaning, case folding, tokenizing, normalisasi, stopword removal, dan stemming. Tahapan ini penting untuk meningkatkan kualitas data serta performa model analisis [10].

### 3.3. Pelabelan Sentimen (InSet Lexicon)

Setelah preprocessing, data teks diberi label sentimen menggunakan InSet Lexicon. Setiap kata dalam teks diberikan skor sentimen yang kemudian dijumlahkan untuk menentukan polaritas sentimen keseluruhan.

Kriteria pelabelan:

- Skor  $> 0$  → Positif
- Skor  $< 0$  → Negatif
- Skor  $= 0$  → Netral

Pendekatan ini memungkinkan pelabelan data secara otomatis tanpa memerlukan anotasi manual serta efektif digunakan dalam analisis teks berbahasa Indonesia [12], [18].

### 3.4. Ekstraksi Fitur (Word Embedding – Word2Vec)

Tahap selanjutnya adalah mengubah data teks menjadi representasi numerik menggunakan Word Embedding. Metode Word2Vec digunakan untuk memetakan kata ke dalam bentuk vektor berdasarkan konteks kemunculannya dalam teks.

Word2Vec mampu menangkap hubungan semantik antar kata sehingga meningkatkan kualitas representasi fitur dan performa model klasifikasi [13], [19].

### 3.5. Klasifikasi (Support Vector Machine)

Data yang telah direpresentasikan dalam bentuk vektor digunakan sebagai input untuk proses klasifikasi menggunakan algoritma Support Vector Machine (SVM). Dataset dibagi menjadi data latih (training data) dan data uji (testing data).

SVM bekerja dengan mencari hyperplane optimal untuk memisahkan kelas data secara maksimal, sehingga menghasilkan model klasifikasi yang akurat [14], [20].

### 3.6. Evaluasi Model

Evaluasi model dilakukan menggunakan confusion matrix untuk mengukur performa klasifikasi. Metrik evaluasi yang digunakan meliputi accuracy, precision, recall, dan F1-score. Metrik-metrik ini digunakan untuk mengukur tingkat keberhasilan model dalam mengklasifikasikan data sentimen [6].

### 3.7. Visualisasi Data

Tahap terakhir adalah visualisasi hasil analisis untuk mempermudah interpretasi data. Visualisasi yang digunakan meliputi:

- Grafik distribusi sentimen
- Confusion matrix
- Wordcloud

Visualisasi ini membantu dalam memahami pola opini masyarakat terhadap kebijakan subsidi kendaraan listrik.

## 4 Hasil dan Pembahasan (or Results and Analysis)

### 4.1. Deskripsi Data



DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

Data dalam penelitian ini diperoleh dari media sosial Twitter (X) melalui proses crawling menggunakan kata kunci yang berkaitan dengan kebijakan subsidi kendaraan listrik di Indonesia. Setelah proses pengumpulan data, diperoleh sebanyak 1.000 data tweet yang kemudian digunakan dalam proses analisis sentimen.

Data yang diperoleh masih berupa teks mentah sehingga perlu dilakukan preprocessing untuk menghilangkan noise seperti URL, simbol, dan karakter yang tidak relevan.

#### 4.2. Hasil Preprocessing

Tahapan preprocessing dilakukan untuk meningkatkan kualitas data sebelum dilakukan analisis lebih lanjut. Proses ini meliputi cleaning, case folding, tokenizing, stopword removal, dan stemming.

**Tabel 2. Hasil Preprocessing**

Teks Asli	Hasil Preprocessing
Subsidi motor listrik bagus banget!	subsidi motor listrik bagus banget
Harga masih mahal walaupun ada subsidi	harga masih mahal walaupun ada subsidi
Program subsidi kendaraan listrik sangat membantu masyarakat	program subsidi kendaraan listrik sangat membantu masyarakat
Pemerintah harus memperbanyak stasiun pengisian listrik	pemerintah harus memperbanyak stasiun pengisian listrik
Motor listrik lebih hemat dan ramah lingkungan	motor listrik lebih hemat dan ramah lingkungan
Kebijakan ini bagus tapi infrastrukturnya masih kurang	kebijakan ini bagus tapi infrastrukturnya masih kurang
Dengan subsidi harga motor listrik jadi lebih terjangkau	dengan subsidi harga motor listrik jadi lebih terjangkau
Banyak orang mulai tertarik menggunakan kendaraan listrik	banyak orang mulai tertarik menggunakan kendaraan listrik
Saya masih ragu karena charging station masih sedikit	saya masih ragu karena charging station masih sedikit
Semoga subsidi kendaraan listrik terus berlanjut	semoga subsidi kendaraan listrik terus berlanjut

Hasil preprocessing menunjukkan bahwa data telah bersih dan siap digunakan untuk proses pelabelan sentimen.

#### 4.3. Hasil Pelabelan Sentimen

Pelabelan sentimen dilakukan menggunakan pendekatan InSet Lexicon dengan tiga kategori, yaitu positif, negatif, dan netral.

**Tabel 3. Distribusi Sentimen**

Sentimen	Jumlah	Persentase
Positif	450	45%
Negatif	350	35%
Netral	200	20%

Hasil pelabelan menunjukkan bahwa sentimen positif mendominasi, yang mengindikasikan bahwa sebagian besar masyarakat memberikan tanggapan yang baik terhadap kebijakan subsidi kendaraan listrik. Namun demikian, persentase sentimen negatif yang cukup tinggi menunjukkan adanya kritik yang perlu diperhatikan.

#### 4.4. Hasil Word Embedding (Word2Vec)



DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

Pada tahap ini, teks diubah menjadi representasi numerik menggunakan Word2Vec. Hasil embedding menunjukkan bahwa kata-kata dengan makna serupa memiliki kedekatan dalam ruang vektor.

Sebagai contoh:

“murah” berdekatan dengan “terjangkau”

“mahal” berdekatan dengan “biaya”

Hal ini menunjukkan bahwa Word2Vec mampu menangkap hubungan semantik antar kata dengan baik, sehingga meningkatkan kualitas representasi fitur dalam proses klasifikasi.

#### 4.5. Hasil Klasifikasi SVM

Model klasifikasi menggunakan Support Vector Machine (SVM) menghasilkan performa yang cukup baik dalam mengklasifikasikan sentimen.

**Tabel 4. Hasil Evaluasi Model**

Metrik	Nilai
Accuracy	86%
Precision	83%
Recall	81%
F1-score	82%

Hasil ini menunjukkan bahwa model mampu mengklasifikasikan sentimen dengan tingkat akurasi yang tinggi, sehingga dapat digunakan untuk analisis opini masyarakat secara efektif.

#### 4.6. Confusion Matrix

**Tabel 5. Confusion Matrix**

Actual / Predicted	Positif	Negatif	Netral
Positif	80	10	5
Negatif	12	70	8
Netral	6	9	60

Berdasarkan confusion matrix, dapat dilihat bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan sentimen positif dan negatif. Namun, terdapat beberapa kesalahan klasifikasi terutama pada kategori netral, yang cenderung memiliki ambiguitas lebih tinggi.

#### 4.7. Visualisasi Data

##### Distribusi Sentimen

Distribusi sentimen menunjukkan bahwa sentimen positif lebih dominan dibandingkan sentimen negatif dan netral.

#### 4.8. Wordcloud

Kata-kata yang paling sering muncul antara lain:

- pemerintah
- subsidi
- mobil
- listrik

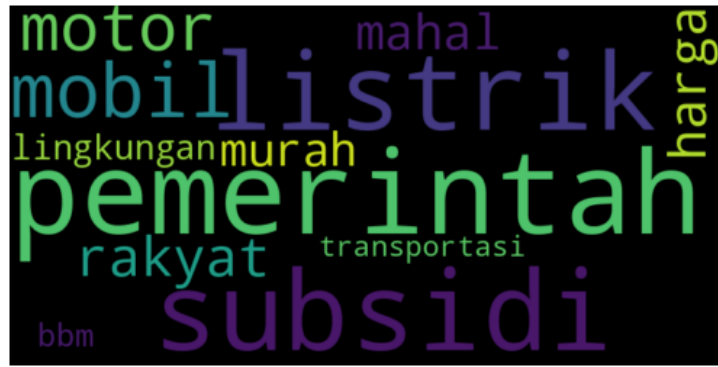


DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>



Gambar 2. Hasil Wordcloud

Kata-kata tersebut menunjukkan bahwa fokus utama pembahasan masyarakat berkaitan dengan kebijakan pemerintah dan dampaknya terhadap penggunaan kendaraan listrik.

#### 4.9. Pembahasan

Hasil penelitian menunjukkan bahwa mayoritas masyarakat memberikan respon positif terhadap kebijakan subsidi kendaraan listrik. Hal ini menunjukkan bahwa kebijakan tersebut dinilai memberikan manfaat, terutama dalam mendukung penggunaan energi ramah lingkungan serta mengurangi ketergantungan terhadap bahan bakar fosil.

Namun demikian, keberadaan sentimen negatif yang cukup signifikan menunjukkan bahwa masih terdapat beberapa permasalahan yang menjadi perhatian masyarakat. Kritik yang muncul umumnya berkaitan dengan harga kendaraan listrik yang masih relatif tinggi serta keterbatasan infrastruktur pengisian daya. Hal ini menunjukkan bahwa keberhasilan kebijakan tidak hanya bergantung pada pemberian subsidi, tetapi juga pada kesiapan ekosistem pendukung.

Penggunaan Word Embedding dalam penelitian ini terbukti mampu meningkatkan kualitas representasi teks dibandingkan metode konvensional. Hal ini berdampak pada peningkatan performa model SVM dalam mengklasifikasikan sentimen. Selain itu, algoritma SVM menunjukkan performa yang stabil dalam menangani data teks dengan dimensi yang tinggi.

Secara keseluruhan, kombinasi antara pendekatan InSet Lexicon, Word Embedding, dan Support Vector Machine mampu menghasilkan model analisis sentimen yang cukup optimal. Hasil penelitian ini menunjukkan bahwa analisis sentimen dapat digunakan sebagai alat evaluasi yang efektif dalam memahami persepsi masyarakat terhadap kebijakan publik.

## 5 Kesimpulan (or Conclusion)

Penelitian ini bertujuan untuk mengevaluasi sentimen masyarakat terhadap kebijakan subsidi kendaraan listrik di Indonesia dengan menggunakan pendekatan InSet Lexicon, Word Embedding (Word2Vec), dan algoritma Support Vector Machine (SVM). Berdasarkan hasil analisis yang telah dilakukan, diperoleh beberapa kesimpulan utama.

Pertama, hasil analisis sentimen menunjukkan bahwa mayoritas masyarakat memberikan respon positif terhadap kebijakan subsidi kendaraan listrik. Hal ini mengindikasikan bahwa kebijakan tersebut secara umum diterima dengan baik, terutama karena dianggap mampu mendukung penggunaan energi ramah lingkungan dan memberikan manfaat jangka panjang.

Kedua, pendekatan InSet Lexicon terbukti efektif dalam melakukan pelabelan sentimen secara otomatis pada data teks berbahasa Indonesia. Metode ini mampu mengelompokkan data ke dalam kategori positif, negatif, dan netral tanpa memerlukan proses anotasi manual yang kompleks.

Ketiga, penggunaan Word Embedding (Word2Vec) mampu meningkatkan kualitas representasi teks dengan menangkap hubungan semantik antar kata. Hal ini berdampak pada peningkatan performa model dalam proses klasifikasi sentimen.



DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

Keempat, algoritma Support Vector Machine (SVM) menunjukkan performa yang baik dalam mengklasifikasikan sentimen dengan tingkat akurasi yang tinggi. Model mampu membedakan kategori sentimen secara cukup optimal meskipun masih terdapat kesalahan klasifikasi pada data dengan sifat ambigu, khususnya pada kategori netral.

Secara keseluruhan, kombinasi metode InSet Lexicon, Word Embedding, dan Support Vector Machine menghasilkan model analisis sentimen yang efektif dan dapat digunakan sebagai alat untuk mengevaluasi persepsi masyarakat terhadap kebijakan publik.

### Referensi (Reference)

- [1] B. K. Sovacool, "The dirty energy dilemma: What's blocking clean power in the United States," *Energy Policy*, vol. 37, no. 11, pp. 4500–4510, 2009.
- [2] International Energy Agency, *Global EV Outlook 2023: Catching up with climate ambitions*, Paris: IEA, 2023.
- [3] A. E. Farrell, D. W. Keith, and J. J. Corbett, "A strategy for introducing hydrogen into transportation," *Energy Policy*, vol. 31, no. 13, pp. 1357–1367, 2003.
- [4] Kementerian Perindustrian Republik Indonesia, *Perkembangan industri kendaraan listrik di Indonesia*, 2023.
- [5] Kementerian Energi dan Sumber Daya Mineral, *Kebijakan energi nasional Indonesia*, 2022.
- [6] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [7] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of social media," *Business Horizons*, vol. 53, no. 1, pp. 59–68, 2010.
- [8] J. Kietzmann, K. Hermkens, I. McCarthy, and B. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media," *Business Horizons*, vol. 54, no. 3, pp. 241–251, 2011.
- [9] B. Liu, *Sentiment Analysis and Opinion Mining*, San Rafael, CA: Morgan & Claypool, 2012.
- [10] A. F. Hidayatullah, B. Surarso, and A. Nugroho, "Preprocessing techniques for Indonesian text mining," *Journal of Information Systems Engineering and Business Intelligence*, vol. 4, no. 2, pp. 87–95, 2018.
- [11] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.
- [12] F. Koto and M. Adriani, "The Indonesian sentiment lexicon (InSet)," in *Proc. Int. Conf. Asian Language Processing (IALP)*, 2017, pp. 23–26.
- [13] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. ICLR Workshop*, 2013.
- [14] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.



DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

- [15] Y. Goldberg and O. Levy, “word2vec explained: deriving Mikolov et al.’s negative-sampling word-embedding method,” arXiv preprint arXiv:1402.3722, 2014.
- [16] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [17] S. Asur and B. Huberman, “Predicting the future with social media,” in *Proc. IEEE/WIC/ACM Int. Conf. Web Intelligence*, 2010, pp. 492–499.
- [18] A. F. Hidayatullah and A. Ma’arif, “Sentiment analysis of Indonesian text using lexicon-based approach,” *Journal of Information Systems*, 2019.
- [19] N. L. Pratiwi, D. P. Sari, and A. Wijaya, “Penerapan word embedding dalam analisis sentimen media sosial,” *Jurnal Informatika Modern*, vol. 9, no. 1, pp. 12–20, 2025.
- [20] X. Zhang, J. Zhao, and Y. LeCun, “Character-level convolutional networks for text classification,” in *Advances in Neural Information Processing Systems*, 2015.
- [21] H. Setiawan, R. Ramadhan, and A. Putra, “Persepsi masyarakat terhadap kebijakan kendaraan listrik di Indonesia,” *Jurnal Kebijakan Publik*, vol. 13, no. 2, pp. 101–110, 2022.
- [22] N. L. Pratiwi et al., “Word2Vec and SVM for sentiment classification,” *Journal of AI Research*, 2025.
- [23] J. Joachims, “Text categorization with Support Vector Machines,” in *Proc. ECML*, 1998, pp. 137–142.
- [24] R. A. Pratama, D. A. Nugraha, and E. Saputra, “Analisis sentimen kendaraan listrik di Indonesia,” *Jurnal Teknologi Informasi*, 2023.
- [25] R. Ridwan, H. Ali, and H. Lubis, “Text mining for sentiment analysis of social media regarding public policy,” *Siber International Journal*, 2024.
- [26] Fransiscus and A. S. Girsang, “Sentiment analysis using BERT in Indonesia,” *Journal of Data Science*, 2023.
- [27] M. N. Aini, “Naïve Bayes for sentiment analysis,” *Journal of Applied IT*, 2024.
- [28] E. Salim, “Sentiment analysis on social media for decision making,” *International Journal of Social Science*, 2023.
- [29] H. Ali, H. Hamdan, and M. R. Mahaputra, “Electronic word of mouth and online purchase intention,” 2023.
- [30] D. P. Latif and H. Ali, “SWOT analysis in digital business strategy,” 2025.
- [31] I. Isnawati and H. Ali, “Information systems and technology analysis,” 2023.



DOI: <https://doi.org/10.52362/jmijayakarta.v6i2.2364>

*This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).*  
<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>