

Perbandingan SVM dan Decision Tree untuk Analisis Sentimen Ulasan Wuling Air EV

Jefri Handanu Wijaya Putra^{*1}

1. S1 Informatika, Teknik dan Ilmu Komputer, Universitas Muhammadiyah Pekajangan Pekalongan, Indonesia

Article Info

Kata Kunci: Analisis Sentimen; Decision Tree; SVM; Wuling Air EV.

Keywords: *Decision Tree; Sentiment Analysis; SVM; Wuling Air EV.*

Article history:

Received 17 August 2018

Revised 15 February 2019

Accepted 4 April 2019

Available online 4 April 2019

DOI :

<https://doi.org/10.29100/jipi.v4i1.781>

* Corresponding author.

Jefri Handanu Wijaya Putra

E-mail address:

jefriputra2004@gmail.com

ABSTRAK

Perkembangan kendaraan listrik di Indonesia semakin meningkat seiring dengan upaya pemerintah dalam mendorong penggunaan transportasi ramah lingkungan. Meskipun demikian, masih terdapat berbagai persepsi masyarakat terkait aspek teknis kendaraan listrik, khususnya mengenai daya tahan baterai. Persepsi tersebut sering kali disampaikan melalui media sosial, salah satunya melalui kolom komentar pada *platform* YouTube, yang dapat dimanfaatkan sebagai sumber data untuk analisis sentimen. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap daya tahan baterai mobil listrik berdasarkan komentar pengguna YouTube serta membandingkan kinerja metode klasifikasi *Support Vector Machine* (SVM) dan *Decision Tree*. Data penelitian diperoleh dari komentar pada video pengujian daya tahan baterai kendaraan listrik Wuling Air EV. Proses pengolahan data dimulai dari tahap pelabelan sentimen secara manual ke dalam tiga kategori, yaitu positif, negatif, dan netral. Selanjutnya dilakukan *text preprocessing* yang meliputi *case folding*, *tokenization*, *stopword removal*, dan *stemming* menggunakan pustaka Sastrawi. Representasi fitur dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Dataset kemudian dibagi menjadi data latih dan data uji untuk membangun model klasifikasi menggunakan algoritma SVM dan *Decision Tree*. Hasil pengujian menunjukkan bahwa metode SVM menghasilkan tingkat akurasi yang lebih tinggi dibandingkan *Decision Tree*. Namun, *Decision Tree* memiliki keunggulan dalam interpretasi model karena mampu menampilkan aturan keputusan yang lebih mudah dipahami. Selain itu, analisis distribusi sentimen menunjukkan bahwa mayoritas komentar pengguna memiliki kecenderungan sentimen positif terhadap daya tahan baterai mobil listrik. Temuan ini menunjukkan bahwa analisis sentimen pada komentar media sosial dapat memberikan gambaran awal mengenai persepsi publik terhadap teknologi kendaraan listrik.

ABSTRACT

The development of electric vehicles in Indonesia has continued to increase in line with the government's efforts to promote the use of environmentally friendly transportation. However, various public perceptions still exist regarding the technical aspects of electric vehicles, particularly concerning battery durability. These perceptions are often expressed through social media, one of which is the comment section on the YouTube platform, which can be utilized as a data source for sentiment analysis. This study aims to analyze public sentiment toward the battery durability of electric cars based on

YouTube user comments and to compare the performance of the Support Vector Machine (SVM) and Decision Tree classification methods. The research data were obtained from comments on videos testing the battery durability of the Wuling Air EV electric vehicle. The data processing began with the manual labeling of sentiments into three categories: positive, negative, and neutral. Subsequently, text preprocessing was performed, including case folding, tokenization, stop-word removal, and stemming using the Sastrawi library. Feature representation was carried out using the Term Frequency Inverse Document Frequency (TF-IDF) method. The dataset was then divided into training data and testing data to build classification models using the SVM and Decision Tree algorithms. The testing results show that the SVM method produces a higher level of accuracy compared to the Decision Tree. However, the Decision Tree has an advantage in model interpretability because it can present decision rules that are easier to understand. In addition, the sentiment distribution analysis indicates that the majority of user comments tend to have positive sentiment toward the battery durability of electric vehicles. These findings suggest that sentiment analysis of social media comments can provide an initial overview of public perception toward electric vehicle technology.

1. PENDAHULUAN

Industri otomotif di Indonesia berkembang dengan cepat, yang dapat dilihat dari banyaknya jenis kendaraan yang tersedia serta meningkatnya permintaan konsumen terhadap alat mobilitas modern [1]. Berdasarkan data Kementerian ESDM, jumlah pengguna kendaraan listrik di Indonesia meningkat pesat dari 2.012 pada tahun 2020 menjadi 7.679 pada tahun 2022, menunjukkan bahwa mobil listrik semakin banyak digunakan di Indonesia [2]. Keberhasilan penjualan Wuling Air EV didukung oleh strategi pemasaran *omnichannel* yang diterapkannya. Melalui kombinasi berbagai kanal *online* dan *offline*, Wuling berhasil memperluas titik interaksi mereka dengan publik dalam hal promosi kendaraan listriknya [3]. Teknologi EV (*electric vehicle*) dianggap lebih ramah lingkungan serta berpotensi dapat mengurangi polusi udara di area perkotaan padat penduduk [4]. Mobil listrik memiliki biaya perawatan rutin dan operasional yang jauh lebih murah dibanding mobil konvensional berkat sistem kerja mesinnya yang sederhana [5].

YouTube adalah sebuah *platform* internet yang menyediakan berbagai macam informasi berupa video sehingga pengguna dapat mengakses konten yang memuat berita terkini [6]. *Platform* YouTube menyediakan berbagai cara mekanisme sosial untuk mengukur opini dan penayangan pengguna terhadap sebuah konten melalui voting, *rating*, favorit, *share*, dan komentar negatif [7]. [8] menerapkan algoritma *Support Vector Machine* (SVM) pada analisis sentimen terkait mobil listrik dan memperoleh performa klasifikasi yang cukup baik dengan akurasi mencapai 86,07%. Penelitian oleh [9]

membuktikan bahwa komentar YouTube dapat dimanfaatkan sebagai sumber data analisis sentimen terkait mobil listrik. Berdasarkan penelitian terdahulu, sebagian besar penelitian analisis sentimen mengenai mobil listrik masih berfokus pada opini publik secara umum dan menggunakan *platform* seperti Twitter. Penelitian yang secara khusus membahas sentimen publik terhadap daya tahan baterai mobil listrik berbasis komentar YouTube menggunakan perbandingan algoritma SVM dan *Decision Tree* masih terbatas. Oleh karena itu, penelitian ini dilakukan untuk menganalisis sentimen publik terhadap daya tahan baterai Wuling Air EV menggunakan metode SVM dan *Decision Tree*. Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap daya tahan baterai Wuling Air EV menggunakan algoritma SVM dan *Decision Tree* serta membandingkan performa kedua metode tersebut.

2. METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini berasal dari komentar video pada kanal Ridwan Hanif dengan judul “Bawa Wuling Air Ev Ke Bandung” di *platform* YouTube dengan jumlah komentar sebanyak 1210 komentar yang dikumpulkan pada tanggal 19 November 2025. Pengumpulan data dilakukan menggunakan metode *Scraping* dengan python serta YouTube API, yang berfungsi untuk mendapatkan data komentar dari video kanal YouTube Ridwan Hanif dengan judul “Bawa Wuling Air Ev Ke Bandung” secara resmi dan terstruktur. Target data yang dikumpulkan dalam

penelitian ini adalah komentar tingkat awal yang memiliki makna sentimen positif, negatif, maupun netral yang relevan dengan penelitian ini. [10] menjelaskan bahwa pengumpulan data dari media sosial, termasuk YouTube, perlu dilakukan dengan cara yang terstruktur melalui API atau audit data agar informasi yang diperoleh tetap relevan, representatif, dan layak digunakan dalam penelitian. Selanjutnya, data di sortir berdasarkan relevansinya terhadap konteks penelitian yang kemudian diberi label sentimen oleh secara manual lalu disimpan dalam bentuk format CSV. Pelabelan manual dipilih karena menurut [11], pelabelan secara manual mampu menghasilkan kualitas label yang lebih tepat, sehingga model *Machine Learning* dapat mencapai kinerja yang lebih unggul dibandingkan menggunakan label otomatis berbasis lexicon.

#	A	B	C
1	komentar		sentimen
2	Disaat nt semua memakai kendaraan EV tiba2 listrik dishutdown bs trjadi bencana alam atau srsuatu yg dibuat tere		negatif
3	Kecil harga ny mahal juga		negatif
4	100thn ke depan manusia akan bingung tentang mengatasi limbah baterai 0V...0V...0V...		negatif
5	Mobil listrik gk bisa kena hujan n mungkin bengkelnya khusus n jauh x		negatif
6	INTINYA INI MOBIL NYUSAHIN , BODY DOANK BAGUS RIBET NYA MINTA AMPUN.		negatif
7	Yang masih aku pikirkan kl pake mobil listrik tu, mengingat Indonesia banyak jalanan kampung yang masih jarang b		negatif
8	Mobil listrik ada minusnya juga sih kalau melalui jin yg banjur bakal komlet kelistrakan nra, malahan ada yg gak b		netral
9	sampah begini dipakai, berkendara itu citra kenyamanan keamanan keselamatan paling utama		negatif
10	Electric car, mutlak hanya utk mobil ke 2. Harga airev dgn stargazer/xpander beda tipis. Kalau mau tenang ke luar k		netral
11	Sbenarnya mobil listrik inovasi yg futuristik. Tpi sayangnya masih cukup mahal buat masyarakat saat ini		netral
12	Mungkin perlu dimodif ada teknologi kalo jalan bisa sambil ngecas batremnya sendiri		netral
13	Yg jd masalah kendaraan elektrik tr ngecasnya yg lama.		negatif
14	Siapa bilang mobil listrik lebih murah biaya BBM nya? Hahahah pada ketipu kalian... kalian akan rasakan setelah pe		negatif
15	Ngeri kalo lewat jalan banjur		netral
16	pengen beli tapi nunggu subsidi pemerintah 0V...0V"		netral
17	Awalnya ngerasa mobil aneh, tp skrg malah inlove, berharap bisa beli. Aamin		positif
18	Pengen bgt irit bgt mobil listrik 0V"		positif

Gambar 1. Struktur Data

2.2 Preprocessing

Tahap *preprocessing* dilakukan untuk membersihkan data teks sebelum proses klasifikasi. Menurut [12], penerapan teknik pra-pemrosesan teks seperti segmentasi kata, penghilangan *stopword* khusus, dan pembersihan simbol secara sistematis terbukti meningkatkan performa model klasifikasi teks, karena input menjadi lebih bersih dan representatif. Tahapan yang diterapkan meliputi *case folding*, *cleaning*, *tokenization*, *stopword removal*, dan *stemming*.

2.3 TF-IDF dan Split Data

Setelah *preprocessing*, data teks diubah menjadi representasi numerik menggunakan metode TF-IDF. Metode ini digunakan untuk memberikan bobot pada kata berdasarkan tingkat kepentingannya dalam dokumen dibandingkan dengan seluruh korpus [13]. Pembagian data dilakukan menggunakan fungsi `train_test_split` dari *library* *scikit-learn*, yang masing - masingnya adalah 80% data latih dan 20% data uji. Karena, menurut [11], rasio pembagian data 80:20 memiliki pengaruh paling besar terhadap kinerja model ketika jumlah data latih adalah 200.

2.4 Klasifikasi

Penelitian ini menggunakan *library* *LinearSVC* dari *scikit-learn* dengan parameter *default*. Karena, menurut [14], *Support Vector Machine* (SVM) adalah algoritma klasifikasi yang bekerja dengan tujuan untuk menemukan *hyperplane* optimal dengan *margin* maksimum sebagai pemisah dua kelas yang berbeda. Ditembakkan metode *Decision Tree* sebagai pembanding

karena menurut [15], *Decision Tree* adalah salah satu metode yang efektif dan banyak diterapkan di berbagai bidang, seperti *machine learning*, pengolahan visual, serta pengenalan pola.

2.5 Evaluasi

Evaluasi model dilakukan untuk mengukur performa algoritma SVM dan *Decision Tree* dalam mengklasifikasi sentimen komentar YouTube. Pengukuran performa dilakukan menggunakan *confusion matrix* serta beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* digunakan untuk mengukur tingkat ketepatan prediksi model secara keseluruhan, sedangkan *precision*, *recall*, dan *F1-score* digunakan untuk mengevaluasi performa klasifikasi pada masing-masing kelas sentimen.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan & Labeling Data

Data yang dikumpulkan hanya berupa komentar tanpa mengambil keterangan identitas pengguna, waktu *posting*, ataupun jumlah *like*, yang menjadikan penelitian ini tetap menjaga privasi bagi para pengguna. Berikut adalah gambar kode sintaks untuk memulai tahap *scrapping* data:

```

from googleapiclient.discovery import build
import pandas as pd
API_KEY = "AIzaSyAP4XV1kg@9kkmR6MNSAAIoo4MPrYw"
VIDEO_ID = "Pc69a10e5p1"
youtube = build('youtube', 'v3', developerKey=API_KEY)
def get_comments_only(video_id):
    comments = []
    next_page_token = None
    while True:
        response = youtube.commentThreads().list(
            part="snippet",
            videoId=video_id,
            maxResults=100,
            textFormat="plainText",
            pageToken=next_page_token
        ).execute()
        for item in response["items"]:
            top_comment = item["snippet"]["topLevelComment"]["snippet"]["textDisplay"]
            comments.append(top_comment)
        next_page_token = response.get("nextPageToken")
        if not next_page_token:
            break
    return comments
comments = get_comments_only(VIDEO_ID)
df = pd.DataFrame(comments, columns=["comment"])
df.head()
df.to_csv("komentar_youtube.csv", index=False)
print("Total komentar diambil:", len(df))

```

Gambar 2. Sintaks Scraping Data

Setelah itu, setiap komentar diberi label sentimen secara manual yang dibagi menjadi tiga kelas yaitu positif, negatif, dan netral seperti gambar berikut;

#	A	B	C
1	komentar		sentimen
2	Disaat nt semua memakai kendaraan EV tiba2 listrik dishutdown bs trjadi bencana alam atau srsuatu yg dibuat tere		negatif
3	Kecil harga ny mahal juga		negatif
4	100thn ke depan manusia akan bingung tentang mengatasi limbah baterai 0V...0V...0V...		negatif
5	Mobil listrik gk bisa kena hujan n mungkin bengkelnya khusus n jauh x		negatif
6	INTINYA INI MOBIL NYUSAHIN , BODY DOANK BAGUS RIBET NYA MINTA AMPUN.		negatif
7	Yang masih aku pikirkan kl pake mobil listrik tu, mengingat Indonesia banyak jalanan kampung yang masih jarang b		negatif
8	Mobil listrik ada minusnya juga sih kalau melalui jin yg banjur bakal komlet kelistrakan nra, malahan ada yg gak b		netral
9	sampah begini dipakai, berkendara itu citra kenyamanan keamanan keselamatan paling utama		negatif
10	Electric car, mutlak hanya utk mobil ke 2. Harga airev dgn stargazer/xpander beda tipis. Kalau mau tenang ke luar k		netral
11	Sbenarnya mobil listrik inovasi yg futuristik. Tpi sayangnya masih cukup mahal buat masyarakat saat ini		netral
12	Mungkin perlu dimodif ada teknologi kalo jalan bisa sambil ngecas batremnya sendiri		netral
13	Yg jd masalah kendaraan elektrik tr ngecasnya yg lama.		negatif
14	Siapa bilang mobil listrik lebih murah biaya BBM nya? Hahahah pada ketipu kalian... kalian akan rasakan setelah pe		negatif
15	Ngeri kalo lewat jalan banjur		netral
16	pengen beli tapi nunggu subsidi pemerintah 0V...0V"		netral
17	Awalnya ngerasa mobil aneh, tp skrg malah inlove, berharap bisa beli. Aamin		positif
18	Pengen bgt irit bgt mobil listrik 0V"		positif

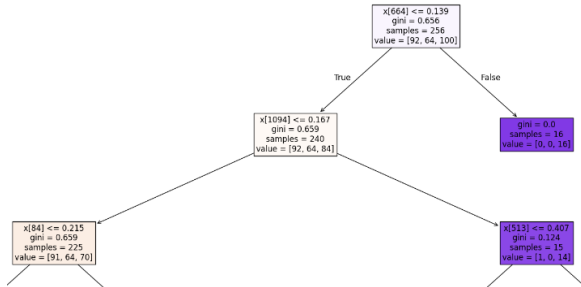
Gambar 3. Pelabelan Data


```

*** Accuracy: 0.5846153846153846
      precision  recall  f1-score  support
negatif    0.58    0.46    0.51     24
netral     0.38    0.55    0.44     11
positif    0.70    0.70    0.70     30

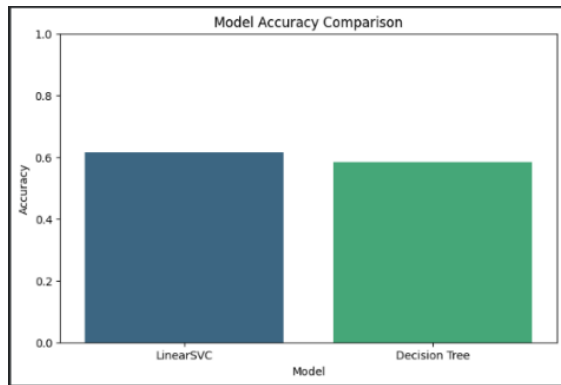
accuracy          0.58     65
macro avg         0.55     65
weighted avg      0.60     65
    
```

Gambar 11. Hasil Klasifikasi Decision Tree



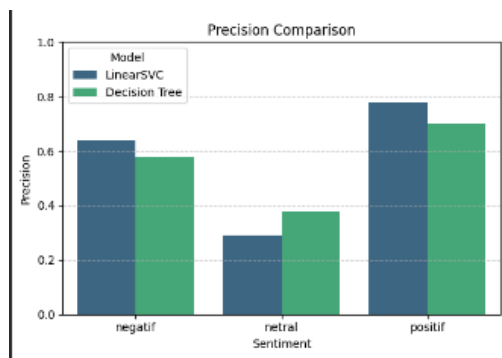
Gambar 12. Tampilan Pohon Keputusan

3.6. Perbandingan Performa SVM & Decision Tree



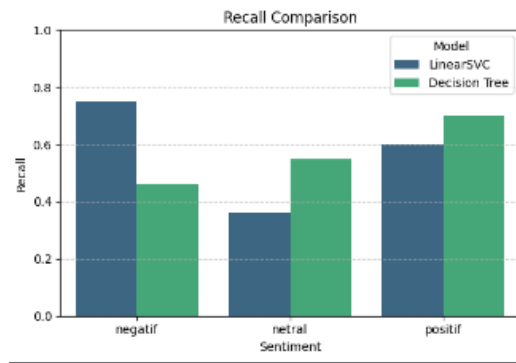
Gambar 13. Perbandingan Akurasi

Gambar 13 menunjukkan bahwa performa metode SVM sedikit lebih baik dibanding Decision Tree.



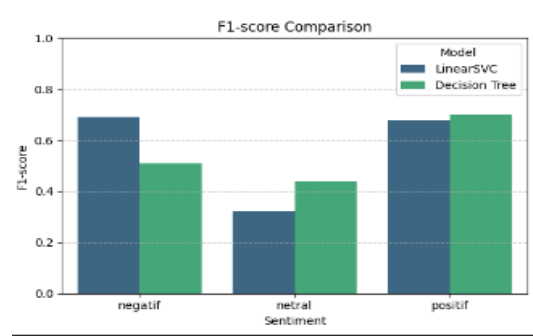
Gambar 14. Perbandingan Precision

Gambar 14 menunjukkan bahwa performa metode SVM pada kelas positif dan negatif lebih baik dibanding Decision Tree. Namun, metode Decision Tree unggul pada kelas netral.



Gambar 15. Perbandingan Recall

Gambar 15 menunjukkan bahwa performa matriks recall Decision Tree pada kelas netral dan positif lebih baik dibanding SVM. Namun, metode SVM jauh lebih unggul pada kelas negatif.



Gambar 16. Perbandingan F1-Score

Gambar 16 menunjukkan bahwa performa matriks F1-Score dari SVM hanya unggul pada kelas negatif. Sebaliknya, metode Decision Tree yang lebih baik pada kelas netral dan positif.

3.7. Pembahasan

Hasil penelitian menunjukkan bahwa pemilihan algoritma berpengaruh terhadap performa klasifikasi sentimen. Metode SVM, sebagai model linear dengan margin maksimum, efektif pada ruang fitur TF-IDF yang berdimensi tinggi. Decision Tree memberikan transparansi keputusan, membantu memahami pola yang diambil model. Namun, kedua model sama-sama menghasilkan performa rendah pada kelas netral, mengindikasikan perlunya pertimbangan untuk menambah sampel dataset atau melakukan balancing data.

Tabel 1. Perbandingan Kelas Netral

Metode	Akurasi	Precision	Recall	F1-Score
SVM	61.53%	0.29	0.36	0.32
DT	58.46%	0.38	0.55	0.44

4. KESIMPULAN

Berdasarkan hasil penelitian, metode Support Vector Machine (SVM) menunjukkan performa yang lebih baik dibandingkan Decision Tree dalam klasifikasi

sentimen komentar YouTube terkait daya tahan baterai mobil listrik Wuling, dengan akurasi masing-masing sebesar 61,53% dan 58,46%. SVM lebih efektif dalam mengklasifikasikan sentimen positif dan negatif, sedangkan *Decision Tree* memiliki performa yang lebih baik pada kelas netral. Hasil penelitian juga menunjukkan bahwa komentar YouTube dapat dimanfaatkan sebagai sumber data untuk menganalisis persepsi publik terhadap mobil listrik. Secara umum, sentimen publik terhadap daya tahan baterai mobil listrik Wuling cenderung positif, meskipun masih terdapat sentimen negatif dan netral yang mencerminkan adanya kekhawatiran serta sikap menunggu dari sebagian masyarakat. negatif dan netral yang mencerminkan adanya kekhawatiran serta sikap menunggu dari sebagian masyarakat.

REFERENSI

- [1] D. Ardiyanti, F. Kurniawan, U. Raokter, and R. Wikansari, "Analisis Penjualan Mobil Listrik Di Indonesia Dalam Rentang Waktu 2020-2023," *ECOMA: Journal of Economics and Management*, vol. 1, no. 3, pp. 114–122, Dec. 2023, doi: 10.55681/ecoma.v1i3.26.
- [2] S. Rofi'ah and E. Pranajaya, "Consumer Behavior and the Adoption of Electric Vehicle in Indonesia," 2025.
- [3] Erwin Permana, Siti Khalisa Naurah Rahayu, Shifa Sabilla Hanum, and Syamsurizal, "Penerapan Strategi Pemasaran Produk Wuling Air Ev dalam Menarik Minat Pembelian Konsumen Terhadap Mobil Listrik," *Jurnal Ekonomi, Akuntansi, dan Perpajakan*, vol. 1, no. 3, pp. 243–257, Jul. 2024, doi: 10.61132/jeap.v1i3.307.
- [4] I. Siti Aprillia *et al.*, "KEBIJAKAN MOBIL LISTRIK DI INDONESIA: TANTANGAN DAN PELUANG DALAM MEWUJUDKAN MOBILITAS RAMAH LINGKUNGAN," *Jurnal Pendidikan Sejarah dan Riset Sosial Humaniora*, vol. 4, no. 3, 2024.
- [5] I. Veza, M. Z. Asy'ari, M. Idris, V. Epin, I. M. Rizwanul Fattah, and M. Spraggon, "Electric vehicle (EV) and driving towards sustainability: Comparison between EV, HEV, PHEV, and ICE vehicles to achieve net zero emissions by 2050 from EV," *Alexandria Engineering Journal*, vol. 82, pp. 459–467, Nov. 2023, doi: 10.1016/j.aej.2023.10.020.
- [6] H. Hidayat, F. Santoso, and L. F. Lidimillah, "Analisis Sentimen Pengguna YouTube Tentang Rohingya Menggunakan Algoritma SVM (Support Vector Machine)," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 1729–1738, Jul. 2024, doi: 10.33379/gtech.v8i3.4497.
- [7] R. Singh and A. Tiwari, "YOUTUBE COMMENTS SENTIMENT ANALYSIS," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, p. 5, 2021, [Online]. Available: www.ijsrem.com
- [8] A. Santoso, A. Nugroho, and A. S. Sunge, "Analisis Sentimen Tentang Mobil Listrik Dengan Metode Support Vector Machine Dan Feature Selection Particle Swarm Optimization," 2022.
- [9] A. Karimah and G. Dwilestari, "ANALISIS SENTIMEN KOMENTAR VIDEO MOBIL LISTRIK DI PLATFORM YOUTUBE DENGAN METODE NAIVE BAYES," 2024. [Online]. Available: <https://www.kaggle.com/datasets/billycemerson/analisi>
- [10] J. Ohme *et al.*, "Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking," *Commun. Methods Meas.*, vol. 18, no. 2, pp. 124–141, 2024, doi: 10.1080/19312458.2023.2181319.
- [11] M. Hayaty and A. H. Pratama, "Performance of Lexical Resource and Manual Labeling on Long Short-Term Memory Model for Text Classification," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 1, pp. 74–84, Feb. 2023, doi: 10.26555/jiteki.v9i1.25375.
- [12] D. Zhang, J. Li, Y. Xie, and A. Wulamu, "Research on performance variations of classifiers with the influence of pre-processing methods for Chinese short text classification," *PLoS One*, vol. 18, no. 10 October, Oct. 2023, doi: 10.1371/journal.pone.0292582.
- [13] K. Taha, P. D. Yoo, C. Yeun, D. Homouz, and A. Taha, "A comprehensive survey of text classification techniques and their research applications: Observational and experimental insights," Nov. 01, 2024, *Elsevier Ireland Ltd.* doi: 10.1016/j.cosrev.2024.100664.
- [14] F. Adi Artanto, "Support Vector Machine Berbasis Particle Swarm Optimization Pada Analisis Sentimen Anggota KPPS," *JURNAL FASILKOM*, vol. 14, no. 1, pp. 75–79, 2024.
- [15] B. T. Jijo and A. M. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 1, pp. 20–28, Jan. 2021, doi: 10.38094/jastt20165.