

ANALISIS SENTIMEN KENDARAAN LISTRIK PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER

Riska Indrayana^{1*}, Mukhammad Solikhin, M.Si²

Matematika, Fakultas MIPA, Universitas Negeri Malang, riska.indrayana.2003126@students.um.ac.id¹

Matematika, Fakultas MIPA, Universitas Negeri Malang, mukhammad.solikhin.fmipa@um.ac.id²

*Email : riska.indrayana.2003126@students.um.ac.id

Abstrak

Analisis sentimen twitter adalah teknik yang digunakan untuk mengidentifikasi sentimen atau pendapat seseorang dalam sebuah tweet dan kemudian mengategorikannya ke dalam tweet positif, negatif, ataupun netral. Penelitian kali ini akan mengangkat salah satu topik yang dibahas pada sosial media twitter mengenai kendaraan listrik, motor listrik, dan mobil listrik. Kendaraan listrik ini menuai banyak komentar dari masyarakat yang menimbulkan pro dan kontra di sosial media twitter. Tujuan dari penelitian ini adalah untuk mengetahui sentimen dari masyarakat terhadap kendaraan listrik, apakah pendapat tersebut lebih mengarah positif, negatif, atau netral. Metode yang digunakan yaitu Naive Bayes, analisis sentimen, dan teks mining. Berdasarkan analisis, dapat diamati bahwa sebagian besar tweet memiliki sentimen positif (67%), sementara jumlah tweet yang netral (8%) dengan nilai akurasi 92,20% dan jumlah tweet yang menentang (25%). Dari hasil analisis sentimen dari medis sosial Twtter dengan menggunakan aplikasi Rapidminer terlihat hasil tingginya sentimen positif menandakan bahwa masyarakat tertarik dan setuju dengan perkembangan dan peralihan kendaraan bahan bakar fosil ke kendaraan listrik.

Kata kunci : kendaraan listrik, analisis sentimen, sentimen, Naive Bayes

Abstract

Sentiment analysis on Twitter is a technique used to identify the sentiment or opinion of an individual in a tweet and categorize it as positive, negative, or neutral. This research focuses on one of the topics discussed on the social media platform Twitter regarding electric vehicles, electric motorcycles, and electric cars. Electric vehicles have sparked numerous comments from the public, leading to both positive and negative opinions on Twitter. The objective of this research is to determine the sentiment of the public towards electric vehicles, whether the opinions lean towards positive, negative, or neutral. The methods employed include Naive Bayes, sentiment analysis, and text mining. Based on the analysis, it can be observed that the majority of tweets have a positive sentiment (67%), while a small percentage is neutral (8%) with accuracy value 92,20% and a significant portion expresses opposing views (25%). The sentiment analysis results from Twitter, conducted using the Rapidminer application, show a high level of positive sentiment, indicating that the public is interested in and supportive of the development and transition from fossil fuel-powered vehicles to electric vehicles.

Keyword : Electric vehicles, sentiment analysis, sentiment, Naive Bayes

PENDAHULUAN

Kendaraan listrik telah menjadi tren utama dalam dunia otomotif modern. Dengan meningkatnya kepedulian terhadap lingkungan dan kebutuhan untuk mengurangi emisi gas rumah kaca, kendaraan listrik dipandang sebagai solusi potensial untuk dampak negatif transportasi terhadap lingkungan. Namun, penerapan kendaraan listrik bergantung tidak hanya pada teknis dan ekonomi, tetapi juga pada persepsi dan pengguna terhadap kendaraan tersebut.

Teknologi komunikasi dan informasi saat ini semakin berkembang membuat informasi yang beredar saat ini dapat langsung diketahui dengan cepat melalui layanan internet. Menurut survei mengenai penetrasi pengguna internet yang disajikan Asosiasi Jasa Internet Indonesia (AJII) menyatakan bahwa pengguna internet Indonesia mencapai sebanyak 210 juta pengguna, dan lebih dari pengguna internet menggunakan media sosial [1]. Media sosial sendiri menjadi layanan internet yang paling sering diakses Indonesia, salah satunya adalah Twitter. Dengan semakin banyaknya pengguna internet dalam menggunakan media sosial, dibangunlah sebuah sistem yang dapat menganalisis opini-opini yang ada suatu topik atau bahasan yang disebut *opini mining*. *Sentiment analysis* atau *opini mining* merupakan pembelajaran komputasi yang berkaitan dengan opini orang-orang, *sentiment*, dan emosi melalui teks [3].

Menurut penelitian yang dilakukan oleh Aditta Agustian, Tukino, dan Fitria Nurapriani dengan judul “Penerapan Analisis Sentimen dan Naïve Bayes Terhadap Opini Penggunaan Kendaraan Listrik di Twitter”. Pada penelitian ini peneliti mendapatkan akurasi 80% pada penelitiannya[2]. Pada penelitian yang dilakukan Francis Joseph Costello, dan Kun Chang Lee yang berjudul “Exploring the Sentiment Analysis of Electric Vehicle Social Media Data by Using Feature Selection Methods”, dimana peneliti menemukan bahwa penggunaan 3 feature selection (Chi Square, Information Gain, dan ReliefF) menunjukkan hasil yang baik ketika digunakan dengan Logistic Regression atau support vector machine sebagai algoritma klasifikasi nya dengan akurasi 70,05%/70,04%/70,03% untuk support vector machine dan 70,40%/70,39%/70,51% untuk Logistic Regression[4].

Dalam konteks ini, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kendaraan listrik menggunakan pendekatan analisis teks dan algoritma klasifikasi. Lebih khusus lagi, algoritma *Naive Bayes* akan digunakan untuk mengklasifikasikan sentimen positif, negatif, dan netral berdasarkan beberapa opini masyarakat dalam media sosial twitter terhadap kendaraan listrik. Dengan memahami sentimen pengguna, kita dapat memperoleh wawasan berharga mengenai faktor yang memengaruhi adopsi kendaraan dan merumuskan strategi pemasaran yang tepat. Tujuan utama dari penelitian ini adalah untuk mengembangkan model klasifikasi sentimen menggunakan algoritma *Naive Bayes* untuk menganalisis ulasan dan pendapat masyarakat tentang kendaraan listrik di Indonesia. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi aspek-aspek tertentu yang berkontribusi dalam sentimen positif, negatif, atau netral terhadap kendaraan listrik. Dengan demikian, penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik mengenai persepsi masyarakat terhadap kendaraan listrik.

Untuk mencapai tujuan penelitian ini, kami akan mengumpulkan data ulasan dan opini masyarakat mengenai kendaraan listrik dari media sosial Twitter. Data yang nanti di kumpulkan akan diproses terlebih dahulu serta teks yang dianalisis nanti akan mengekstraksi karakteristik penting. Kemudian, algoritma *Naive Bayes* akan diterapkan untuk melatih model klasifikasi sentimen. Model yang dikembangkan akan dievaluasi menggunakan parameter yang sesuai.

Penelitian ini diharapkan dapat memberikan salah satu kontribusi bagi pengembangan kendaraan listrik dengan memahami sentimen pengguna. Hasil penelitian ini dapat digunakan oleh produsen industri otomotif untuk meningkatkan desain, pemasaran, dan strategi layanan mereka. Selain itu, penelitian ini juga dapat menjadi dasar untuk penelitian lanjutan dalam analisis sentimen dan penggunaan algoritma klasifikasi lainnya dalam konteks kendaraan listrik.

METODE

Ada beberapa tahapan yang perlu dilakukan untuk mendapatkan hasil dari analisis yang terbaik dalam menganalisis *sentiment*. Langkah – langkahnya terdiri dari Analisis sentimen, *Teks mining*, dan Metode *Naive Bayes*.

Analisis Sentimen

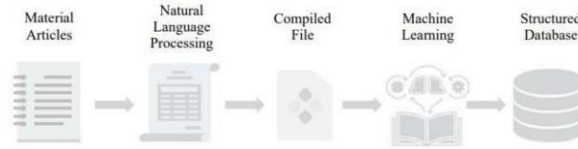
Analisis sentimen merupakan studi komputasi mengenai pendapat, perilaku dan emosi seseorang pada sebuah ungkapan berbentuk teks. Analisis sentimen pada dasarnya mencoba menyimpulkan sentimen orang berdasarkan ekspresi bahasa mereka ke dalam beberapa kategori seperti emosi positif, negatif, atau netral. Saat ini menjadi peneliti sering menggunakan analisis sentimen sebagai bagian dari penelitian ilmu komputer. Jaringan analisis sentimen sering memanfaatkan media sosial seperti Twitter untuk menentukan persepsi audiens.

Analisis sentimen juga bisa disamakan dengan *opinion mining* karena berfokus pada opini yang ada diekspresikan secara positif atau negatif [5]. Analisis sentimen melibatkan penambangan data untuk dianalisis, memproses, dan mengekstrak data tekstual dalam satu entitas, seperti layanan, produk, individu, fenomena, atau topik tertentu [7]. Proses analisis dapat mencakup ulasan teks, forum, tweet, atau blog dengan data *preprocessing* meliputi proses *tokenization*, *stopword*, penghapusan, *stemming*, identifikasi sentimen, dan klasifikasi sentimen [6].

Teks Mining

Teks mining adalah proses dalam menganalisis teks atau informasi untuk yang diambil dari suatu sumber atau dokumen baik secara terstruktur maupun tidak terstruktur. *Teks mining* bertujuan untuk mengekstrak

informasi dari dokumen, *teks mining* dapat menghasilkan informasi melalui pemrosesan, pengelompokan, dan analisis data tidak terstruktur dalam jumlah besar. Proses pengambilan informasi dalam *teks mining* dapat menghasilkan analisis perasaan yang secara emosional mengidentifikasi pernyataan ketika pernyataan itu positif atau negatif. Gambar 1 menunjukkan teks terstruktur menjadi data material terstruktur dalam *teks mining* selanjutnya disimpan dalam baris data terstruktur [8].



Gambar 1. Proses Teks Mining [8]

Metode Naive Bayes

Naive Bayes adalah algoritma pembelajaran mesin berbasis model probabilistik yang digunakan dalam kategorisasi teks dalam kasus penggunaan sebelumnya. Metode ini termasuk kedalam kategori umum Algoritma *Bayesian* dalam *machine learning*. Metode ini berpotensi baik pada kategorisasi dalam hal presisi dan komputasi data. *Naive Bayes* banyak digunakan dalam teknik kategorisasi terutama Twitter menggunakan beberapa metode seperti *Unigram Naive Bayes*, *Multinomial Naive Bayes*, dan *Maximum Entropy Classification*. Perhitungan kategori probabilitas pada *Naive Bayes* menggunakan pendekatan algoritma Bayes menggunakan persamaan:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \tag{I}$$

Persamaan (I) menunjukkan Y suatu kelas spesifik, X merupakan data pada kelas yang belum diketahui, $P(Y|X)$ adalah probabilitas hipotesis berdasarkan kondisi, $P(Y)$ dan $P(X|Y)$ adalah probabilitas sebelumnya dari kelas berdasarkan kondisi hipotesis, sedangkan $P(X)$ merupakan probabilitas Y. Dalam klasifikasi *Naive Bayes*, persamaan I dikembangkan menjadi persamaan (II) :

$$P(Y|X_1, X_2, \dots, X_n) = \frac{P(X_1, X_2, \dots, X_n|Y)P(Y)}{P(X_1, X_2, \dots, X_n)} = \frac{P(X_1|Y)P(X_2|Y)\dots P(X_n|Y)P(Y)}{P(X_1, X_2, \dots, X_n)} \tag{II}$$

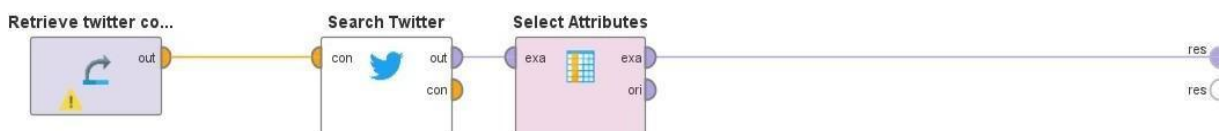
$P(Y|X_1, X_2, \dots, X_n)$ merupakan hasil hitung dari semua probabilitas posterior pada nilai X untuk semua nilai di Y, sehingga kategorisasi *Naive Bayes* akan membuat prediksi berdasarkan probabilitas maksimum dari probabilitas *posterior laplace* pada persamaan (III) dimana c adalah jumlah nilai pada Y [9]:

$$P(X_i|Y) = \frac{N_{ic}+1}{N_c+c} \tag{III}$$

Pada klasifikasi *Naive Bayes* ada 4 tahapan penelitian yang akan di lakukan yaitu *Data Crawling* yaitu tahap pengumpulan data dan pemberian label pada data, lalu *Preprocessing* yaitu tahap pembersihan, penghapusan, pemberian *stopwords*, tokenisasi, dan *stemming* data, selanjutnya Ekstrasi Fitur yaitu tahap perhitungan kata, *vector*, dan *TF-idf*, serta yang terakhir ada *Naive Bayes Classifier* yaitu tahap pengujian dan evaluasi model. Data awal dalam bentuk *tweet* di-*crawling* dari Twitter dan disimpan dalam bentuk *CSV file*. Data dibagi menjadi dua dataset yaitu data latih dan data uji. Pelabelan akan diberikan untuk membedakan *tweet* positif dan negatif. Metode *Naive Bayes* digunakan pada tahap klasifikasi sentimen dan interpretasi hasil analisis sentimen.

HASIL DAN PEMBAHASAN

Dalam proses *Crawling* data Twitter dengan menggunakan *Rapid Miner* kami menggunakan kata kunci ‘kendaraan listrik’, ‘motor listrik’, dan ‘mobil listrik’ yang difilter dengan kata kunci ‘listrik’ dan ‘kendaraan’ pada *tweet-tweet* dalam bahasa Indonesia. Proses ini dilakukan secara manual menggunakan akses token yang didapatkan dari Twitter API dengan menggunakan *tools Rapidminer* untuk mendapatkan data dan informasi yang dibutuhkan. Karena keterbatasan *crawling* secara manual sehingga hanya menghasilkan data sebanyak 141 *tweet* yang berkaitan dengan kata kunci yang dipilih pada proses *crawling* data.



Gambar 2. Proses Crawling Data Menggunakan Rapidminer

Gambar 2 menunjukkan proses *crawling* data menggunakan *Rapid Miner*. Operator-operator yang digunakan seperti *Retrieve twitter* dan *search twitter* untuk menghubungkan ke Twitter dengan memasukkan kode akses token didapat dari Twitter API, lalu yang kedua ada *select attributes* untuk mengambil atribut yang dibutuhkan seperti *username* dan teks, dan operator terakhir yaitu *write CSV* yang digunakan untuk menyimpan data ke dalam format *CSV*.

Setelah proses klasifikasi analisis sentimen telah selesai dilakukan, langkah terakhir setelahnya adalah menyempurnakan proses yang telah dilakukan dengan mengevaluasi hasil. Untuk hasil pengambilan data menggunakan *Rapidminer* ditunjukkan pada tabel 1 dimana terdapat atribut *username* dan teks.

Tabel 1. Data Crawling

Username	Text Tweet
@arkaاتمaja7	Sy setuju soal subsidi listrik tidak tepat apalagi ada conflic of interest dari beberapa pejabat yg punya bisnis...
@vidi_A050	Kendaraan listrik baik mobil listrik atau pun motor listrik dinilai sebagai kendaraan masa depan karena ramah lingkungan Namun kehadiran electric vehicle EV masih dikritisi oleh sejumlah pihak
@OposisiCerdas	Anies Nilai Subsidi Mobil Listrik Tidak Tepat: Kendaraan Umum Lebih Butuh Subsidi https://t.co/rLQpeMFqIM
@Baiqhibatt	Kendaraan listrik makin banyak dijumpai di jalan raya belakangan ini Bahkan di ecommerce Tokopedia penjualan kendaraan listrik dan transaksi motor listrik naik signifikan
@willsarana	STOP MOBIL DAN MOTOR LISTRIK MUDAH TERBAKAR Kita digiring untuk beli kendaraan listrik ternyata malah mudah terbakar

Proses *labelling* data dilakukan untuk menentukan tweet termasuk ke dalam kelas positif yang berisi dukungan, pujian, saran yang positif. Sedangkan kelas negatif berisikan keluhan, tolakan, kritik yang mengandung makna kekecewaan, amarah, ataupun sindiran. Hasil pemberian label di contohkan pada tabel 2.

Tabel 2. Labelling Data

Label	Text Tweet
Positif	Sy setuju soal subsidi listrik tidak tepat apalagi ada conflic of interest dari beberapa pejabat yg punya bisnis...
Positif	Kendaraan listrik baik mobil listrik atau pun motor listrik dinilai sebagai kendaraan masa depan karena ramah lingkungan Namun kehadiran electric vehicle EV masih dikritisi oleh sejumlah pihak
Negatif	Anies Nilai Subsidi Mobil Listrik Tidak Tepat: Kendaraan Umum Lebih Butuh Subsidi #majukantransportasi #Jakartabisa https://t.co/rLQpeMFqIM
Positif	Kendaraan listrik makin banyak dijumpai di jalan raya belakangan ini Bahkan di ecommerce Tokopedia penjualan kendaraan listrik dan transaksi motor listrik naik signifikan #GoGreen #EV
Negatif	STOP MOBIL DAN MOTOR LISTRIK MUDAH TERBAKAR Kita digiring untuk beli kendaraan listrik ternyata malah mudah terbakar

Setelah melakukan proses di atas, didapatkan hasil sebagai berikut :



Gambar 4. Hasil Penelitian

Hasil yang di dapatkan dari penelitian ini adalah masyarakat Indonesia khususnya pada media sosial Twitter rata – rata memberikan respond yang positif terkait adanya kendaraan listrik di Indonesia dengan persentase data sebesar 67% dan jumlah data sebanyak 141 data yaitu 94 data. Respond negatif dengan adanya kendaraan listrik di Indonesia lebih kecil dibanding respond positif sehingga membuktikan hanya sedikit masyarakat yang menentang dengan adanya kendaraan listrik di Indonesia ini. Hal ini ditunjukkan dengan nilai respond negatif sebanyak 25% yaitu 35 data dan persentase netral sebanyak 8% yaitu 11 data. Nilai akurasi yang dihasilkan:

Tabel 5. Tingkat Akurasi

Kategori	Prediksi Benar	Data
Positif	90	94
Negatif	30	35
Netral	10	11
Total	130	141

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi benar} \times 100\%}{\text{Jumlah Total Data}}$$

$$= \frac{130}{141} \times 100\% = 92,20\%$$

PENUTUP

Penelitian ini menggunakan penerapan algoritma *Naive Bayes* untuk klasifikasi sentimen. Analisis sentimen dalam study ini menggunakan data Twitter dengan kata kunci ‘kendaraan listrik’, ‘motor listrik’, dan ‘mobil listrik’. Dalam penelitian ini, data ulasan dan opini masyarakat mengenai kendaraan listrik yang diambil dari media sosial Twitter menggunakan aplikasi *Rapidminer* menghasilkan analisis sentimen yang menunjukkan bahwa dari 141 data yang didapat, sentimen positif ada sebanyak 67%, sentimen negatif sebanyak 25%, dan sentimen netral sebanyak 8%.

Kesimpulan yang dapat diambil dari penelitian ini yaitu mayoritas masyarakat di Indonesia memiliki sentimen positif terhadap adanya kendaraan listrik di Indonesia. Hal ini menunjukkan adanya minat dan dukungan terhadap pengembangan dan peralihan dari kendaraan bahan bakar fosil ke kendaraan listrik. Namun, penelitian ini hanya dari hasil analisis terhadap data dari media sosial Twitter saja. Oleh karena itu, untuk penelitian selanjutnya bisa mengambil sampel data yang lebih luas dan dapat memberikan gambaran

yang lebih komprehensif mengenai sentimen masyarakat terhadap kendaraan listrik di Indonesia.

DAFTAR RUJUKAN

- [1] APJII, (2022), “Survei Penetrasi dan Perilaku Penggunaan Internet “, [Online]. Tersedia: <https://apjii.or.id/survei/surveiprofilinternetindonesia2022-21072047> [Diakses 21 Desember 2022]
- [2] A. Agustian, T. Tukiro, dan F. Nurapriani, “Analisis Sentimen, Text Mining Penerapan Analisis Sentimen Dan Naive Bayes Terhadap Opini Penggunaan Kendaraan Listrik Di Twitter”, *TIKA*, vol. 7, no. 3, hlm. 243-249, 2022, doi: <https://doi.org/10.51179/tika.v7i3.155>
- [3] F. Handayani, dan M. Mustikasari, “SENTIMENT ANALYSIS OF ELECTRIC CARS USING RECURRENT NEURAL NETWORK METHOD IN INDONESIAN TWEETS”, *Jurnal Ilmiah KURSOR*, vol. 10, no. 04, hlm. 153-158, 2020, doi: <https://doi.org/10.21107/kursor.v10i4.233>
- [4] F. J. Costello and K. C. Lee, “Exploring the Sentiment Analysis of Electric Vehicles Social Media Data by Using Feature Selection Methods,” *Journal of Digital Convergence*, vol. 18, no. 2, pp. 249–259, Feb. 2020, doi: <https://doi.org/10.14400/JDC.2020.18.2.249>
- [5] C. B. Saputra, A. Muzakir, and D. Udariansyah, “ANALISIS SENTIMEN MASYARAKAT TERHADAP #2019GANTIPRESIDEN BERDASARKAN OPINI DARI TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER,” in *Bina Darma Conference on Computer Science*, 2019, pp. 403—413
- [6] R. Watrianthos, S. Suryadi, D. Irmayani, M. Nasution, and E. F. S. Simanjorang, “Sentiment Analysis Of Traveloka App Using Naïve Bayes Classifier Method,” *Int. J. Sci. Technol. Res.*, vol. 8, no. 07, pp. 786–788, 2019.
- [7] B. Liu, *Sentiment Analysis and Opinon Mining*. California, USA: Morgan & Claypool Publishers, 2012.
- [8] X. Gao, R. Tan, and G. Li, “Research on Text Mining of Material Science Based on Natural Language Processing,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 768, p. 072094, Mar. 2020, doi: 10.1088/1757- 899X/768/7/072094
- [9] Z. Dong, X. Guo, S. Rajana, and B. Chen, “Understanding 21st century bordeaux wines from wine reviews using naïve bayes classifier,” *Beverages*, 2020, doi: 10.3390/beverages6010005
- [10] D. Setian and I. Seprina, “ANALISIS SENTIMEN MASYARAKAT TERHADAP DATA TWEET LAZADA MENGGUNAKAN TEXT MINING DAN ALGORITMA NAIVE BAYES CLASSIFIER,” in *Bina Darma Conference on Computer Science*, 2020, pp. 998–1004.
- [11] P. Tiwari, P. Yadav, S. Kumar, B. K. Mishra, G. N. Nguyen, S. P. Gochhayat, et al., "Sentiment analysis for airlines services based on Twitter dataset" in *Social Network Analytics*, New York, NY, USA:Academic, pp. 149- 162, 2019.