

Analisis Sentimen Opini Publik Tentang Kendaraan Listrik di Indonesia

Dzaky Candy Fahrezy¹, Fandy Setyo Utomo²

¹Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Jawa Tengah, Indonesia
surel:¹20sa1013@mhs.amikompurwokerto.ac.id,²fandy_setyo_utomo@amikompurwokerto.ac.id

Info Artikel

Sejarah artikel:

Diterima 17-02-2024

Revisi 27-02-2024

Diterima 03-08-2024

Kata kunci:

Analisis sentiment,
Opini public,
Kendaraan listrik
Naïve Bayes,
Support Vector Machine.

ABSTRAK

Kendaraan listrik menjadi perhatian utama dalam industri otomotif global dan juga di Indonesia, sebagai upaya untuk mengurangi dampak negatif lingkungan dan ketergantungan terhadap bahan bakar fosil. Namun, penerimaan kendaraan listrik oleh masyarakat sangat dipengaruhi oleh opini publik yang beragam. Oleh karena itu, analisis sentimen opini publik menjadi penting untuk memahami bagaimana masyarakat Indonesia merespons kendaraan listrik. Penelitian ini bertujuan untuk menerapkan dua algoritma klasifikasi yaitu Naïve Bayes dan Support Vector Machine, kemudian menganalisa hasil klasifikasi penerapan kinerja kedua algoritma tersebut. Data yang digunakan dalam penelitian ini adalah data komentar yang diperoleh dari media social Twitter berjumlah 3.043 data. Hasil penelitian menunjukkan bahwa Support Vector Machine memiliki kinerja paling baik dengan nilai accuracy 82.95% sementara Naive Bayes sebesar 73%. Dengan akurasi tersebut, model yang dibuat mampu mengklasifikasi kategori positif dan negatif pada suatu dokumen dengan baik.

Penulis yang sesuai:

Dzaky Candy Fahrezy

Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Purwokerto

Email: 20sa1013@mhs.amikompurwokerto.ac.id

1. PENDAHULUAN

Pada perkembangan teknologi yang semakin pesat, ada teknologi yang mengubah bahan bakar minyak menjadi tenaga listrik, yang dapat menstabilkan tingkat CO₂ di atmosfer. Penggunaan kendaraan bertenaga listrik merupakan langkah pertama menuju pencemaran udara global [1]. Diharapkan penggunaan kendaraan listrik di seluruh dunia akan meningkat, meskipun saat ini masih sedikit. Saat ini, ada 28.188 unit kendaraan listrik di Indonesia, menurut kementerian perhubungan Republik Indonesia [2].

Naive Bayes adalah salah satu metode klasifikasi teks yang populer dan efektif untuk analisis sentimen. Metode ini didasarkan pada teorema Bayes, yang menyatakan bahwa probabilitas suatu kelas (positif, negatif, atau netral) terhadap teks dapat dihitung dengan menggunakan probabilitas prior kelas dan probabilitas kondisional fitur (kata-kata) terhadap kelas. Metode ini dianggap naif karena menganggap bahwa

fitur-fitur dalam teks tidak bergantung satu sama lain. Namun, metode ini memiliki kelebihan karena mudah, cepat, dan akurat [19]. Support Vector Machine (SVM) adalah salah satu algoritma pembelajaran mesin yang paling populer untuk klasifikasi. Dalam sepuluh tahun terakhir, SVM telah berkembang menjadi alat yang kuat untuk pola klasifikasi dengan tingkat keberhasilan yang tinggi saat digunakan di berbagai industri. Karena SVM dapat menangani berbagai masalah pembelajaran, banyak komunitas pembelajaran mesin tertarik untuk mempelajari dan mengembangkannya. Metode pembelajaran mesin yang dikenal sebagai SVM bertujuan untuk menemukan hyperplane terbaik yang dapat memisahkan dua kelas pada ruang input. Dengan menggunakan pelatihan data, algoritma klasifikasi SVM membuat model klasifikasi. Model ini digunakan untuk memprediksi kelas data baru yang belum pernah ada sebelumnya, yang disebut uji data [20].

Penelitian ini dilatar belakangi pada penelitian terdahulu oleh Salman Alfarizi dan Eka Fitriani pada tahun 2022 mengenai penggunaan algoritma Naive Bayes dengan pemilihan fitur terbukti efektif dalam menganalisis sentimen publik terhadap kendaraan listrik menggunakan data Twitter. Selanjutnya penelitian oleh Atha Fitrah Riyadi, dkk pada tahun 2020 mengenai sentimen publik terhadap kendaraan listrik di Indonesia cenderung positif, dengan metode Klasifikasi Naive Bayes memberikan hasil terbaik. Penelitian ini juga merekomendasikan penelitian masa depan dengan cakupan waktu dan volume tweet yang lebih besar, serta penggunaan analisis sentimen leksikon untuk membantu klasifikasi mesin. Selain itu, ekstraksi data dari platform media sosial lain seperti Facebook, Instagram, dan YouTube dapat memajukan pengembangan penelitian. Selain itu pada penelitian yang dilakukan oleh Dedi Darwis, dkk pada tahun 2020 mengenai penggunaan algoritma naive bayes untuk melakukan analisis sentimen terhadap data Twitter terkait badan meteorologi berhasil diterapkan dengan tingkat akurasi sebesar 69,97%. Proses pengujian melibatkan penghapusan stopword dan tokenisasi, serta menggunakan 1179 data Tweet untuk pelatihan dan pengujian. Algoritma Naive Bayes digunakan untuk mengklasifikasikan tweet ke dalam sentimen positif, negatif, dan netral berdasarkan probabilitas tertinggi dari masing-masing kategori.

Pada penelitian ini, alasan penulis menggunakan algoritma Naive Bayes karena mudah digunakan, namun akurasi hasil yang diperoleh tetap tinggi jika kumpulan data dibersihkan secara optimal dan kelas negative dan positif dari kumpulan data tersebut seimbang, dan penulis telah meninjau dalam beberapa penelitian, nilai akurasi Naive Bayes adalah antara 70% hingga 90% [3]. Selain itu, penulis juga menggunakan metode SVM karena memiliki kemampuan untuk mengurangi kesalahan dan memisahkan dua kumpulan data dari dua kelas yang berbeda. Metode SVM juga memiliki kelebihan lain, yaitu kemampuan untuk menentukan jarak, yang membuat proses komputasi lebih cepat [4].

2. METODE

Penelitian ini memiliki alur penelitian yang dimulai dengan melakukan pengumpulan data ulasan pengguna media social Twitter dengan menggunakan Teknik *Web Scraping*. Web scraping adalah proses mengekstraksi dokumen semi terstruktur dari Internet, biasanya berupa halaman web dalam bahasa markup seperti HTML (*Hypertext Markup Language*) atau XHTML (*Extensible Hypertext Markup Language*) dan parsing dokumen untuk mendapatkan beberapa data dari halaman untuk digunakan untuk tujuan lain. Bahasa markup adalah bahasa komputer yang menggunakan tag untuk mengidentifikasi elemen dalam dokumen [5]. Data ulasan ini diambil pada rentan waktu antara tanggal 1 Januari 2020 hingga 15 Oktober 2023 dengan menggunakan kata kunci/hashtag “#kendaraanlistrik, #motorlistrik, #mobillistrik”. Dengan demikian jumlah data ulasan yang berhasil dikumpulkan berjumlah 3.043 data. Berikut metode penelitian yang dilakukan oleh penulis





Gambar 1. Alur Penelitian

Topik kendaraan listrik ini diambil dengan menganalisa melalui situs *Google Trends*. *Google Trends* adalah alat analitik gratis yang diperkenalkan oleh Google pada tahun 2006 yang memungkinkan pengguna mengukur dan memantau data pencarian web dan melacak popularitas istilah atau topik tertentu dari waktu ke waktu [6]. Pemrosesan awal data adalah langkah pertama dalam pembelajaran mesin untuk memastikan kualitas data dan mengekstrak informasi berguna dari kumpulan data. Ini mencakup penanganan nilai nol, pengkodean nilai kategorikal, normalisasi, transformasi, serta ekstraksi dan seleksi fitur [7]. Proses pelipatan mengacu pada proses kebalikan dalam menentukan perkembangan polihedron atau permukaan yang dapat dikembangkan. Ini lebih kompleks dan melibatkan penyelesaian sistem persamaan aljabar. Hasil pelipatan dapat berhubungan dengan beberapa atau bahkan banyak sekali polihedra atau permukaan yang tidak kongruen [8]. Normalisasi teks adalah tugas mengubah bahasa non-standar, seperti transkripsi ucapan atau komunikasi melalui komputer, menjadi tulisan standar [9]. Tokenisasi adalah proses memecah serangkaian data menjadi elemen dasar atau tokennya. Ini digunakan dalam berbagai konteks seperti bahasa pemrograman, jaringan saraf tiruan, pemrosesan bahasa alami, dan keuangan digital. Dalam bahasa pemrograman dan jaringan saraf tiruan, tokenisasi merupakan langkah awal dalam pengolahan data. Dalam pemrosesan bahasa alami [10], Stemming dapat dipahami sebagai proses untuk memotong imbuhan atau mengembalikan suatu kata berimbuhan menjadi kata dasar [11]. Hal ini bertujuan agar setiap kata yang memiliki kata dasar sama dapat dikelompokkan menjadi satu kelompok. Dalam penelitian ini proses untuk melakukan labeling menggunakan Lexicon berbahasa Indonesia. Berbasis leksikon adalah pendekatan yang mencakup frasa, ekspresi, atau konten dalam bentuk teks, yang biasa ditemukan dalam obrolan, percakapan, postingan, ulasan, dll. Lexicon Based merupakan pendekatan yang menggunakan kamus sentimen yang berisi kata positif dan negatif yang dicocokkan dengan kata dalam kalimat untuk menentukan derajat polaritasnya. Klasifikasi sentimen berbasis kosakata merupakan klasifikasi berdasarkan kata-kata positif dan negatif yang muncul dalam tweet bersih [12]. Kemudian untuk melakukan preprocessing digunakan *Library python Sastrawi*. Perpustakaan Sastrawi adalah suatu perangkat lunak NLP (Natural Language Processing) yang khusus dibuat untuk bahasa Indonesia. Meskipun awalnya dikembangkan untuk bahasa pemrograman PHP, popularitasnya mendorong pengembangan perpustakaan ini untuk mendukung Python. Fungsinya utamanya adalah untuk melakukan proses stemming pada teks berbahasa

Indonesia, seperti yang dijelaskan dalam penelitian yang Anda baca. Dengan keberadaan Sastrawi, peneliti dapat dengan mudah melakukan stemming pada teks berbahasa Indonesia [13]. Kemudian peneliti juga menggunakan *WordCloud*. *Word cloud* atau awan kata, merupakan metode visualisasi data yang memungkinkan seseorang dengan mudah memahami konten dari sejumlah besar dokumen teks dan mengidentifikasi domain subjeknya dalam waktu singkat. Dengan teknik ini, sebuah gambaran visual dibuat dari kata-kata yang paling sering muncul dalam dokumen teks, di mana kata-kata yang paling sering muncul ditampilkan dalam font yang lebih besar dan warna yang berbeda. Penggunaan word cloud umumnya bertujuan mempermudah pemahaman tentang topik atau tema yang terkait dengan koleksi dokumen teks yang besar [14]. *Textblob* merupakan sebuah perpustakaan (library) dalam bahasa pemrograman Python yang menyediakan berbagai fitur untuk memproses bahasa alami (Natural Language Processing/NLP). Beberapa dari fitur yang disediakan oleh *Textblob* mencakup ekstraksi frasa kata benda, penandaan bagian ucapan, analisis sentimen, klasifikasi, tokenisasi, frekuensi kata dan frasa, parsing, n-grams, infleksi kata, lemmatisasi, koreksi ejaan, dan integrasi dengan *Wordnet*. Selain itu, library ini juga mendukung penambahan model atau bahasa baru melalui ekstensi, memberikan fleksibilitas dalam mengatasi berbagai jenis tugas pemrosesan bahasa alami [15]. *Pandas* adalah sebuah pustaka pengolahan data yang ditulis dalam bahasa pemrograman Python. Keunggulan pustaka ini terletak pada fleksibilitasnya, ketepatan penggunaannya, dan statusnya sebagai pustaka sumber terbuka (open source), yang membuatnya dapat diakses dan digunakan secara luas dalam manipulasi data. *Pandas* memiliki kemampuan untuk mengelola data dalam format numerik dan teks, dan menyediakan tiga bentuk varian penyajian data, yaitu series, data frame, dan data panel. Pustaka ini sangat berguna dalam manajemen data tabular atau array informasi, memberikan kemudahan dalam analisis dan transformasi data untuk berbagai keperluan seperti ilmu data (data science) dan pengolahan data secara umum [16]. *Google Colab* merupakan layanan gratis dari Google yang memungkinkan pengguna untuk membuat dan menjalankan notebook Jupyter secara daring. Dalam layanan ini, pengguna dapat menulis dan mengeksekusi kode Python, menyimpan serta berbagi notebook, serta memanfaatkan sumber daya komputasi yang luas seperti CPU, GPU, dan TPU. Kelebihan *Google Colab* terletak pada ketersediaan sumber daya komputasi yang besar secara gratis, membuatnya sangat berguna bagi peneliti dan pengembang yang ingin mengakses daya komputasi tinggi tanpa perlu membeli atau mengelola infrastruktur mereka sendiri [17]. Kemudian untuk menampilkan visualisasi hasil peneliti menggunakan *Confusion Matrix*. *Confusion matrix* adalah sebuah tabel yang digunakan untuk mengevaluasi kinerja suatu model klasifikasi. Tabel ini membandingkan hasil prediksi model dengan nilai sebenarnya dari data yang diuji [18].

3. HASIL DAN PEMBAHASAN

3.1. Identifikasi Masalah

Pada tahapan ini dilakukan proses identifikasi masalah yang akan diteliti dengan mempelajari literatur dan yang bersumber dari pustaka lain. Tahap identifikasi masalah adalah tahap awal yang dilakukan dengan menentukan identifikasi masalah topik Kendaraan Listrik melalui situs *Google Trends*

3.2. Web Scraping

Pada tahap pengumpulan data ini dilakukan teknik Web Scraping untuk mendapatkan data komentar dari media *social Twitter*, data yang diambil berupa ulasan pengguna media *social Twitter* pada tanggal 1 Januari 2020 hingga 15 Oktober 2023. metode *scraping* yang bernama *Tweet-Harvest*, kemudian dilakukan pengambilan data dengan metode crawling menggunakan beberapa kata kunci/hastag seperti “#kendaraanlistrik, #motorlistrik, #mobillistrik”

```

* crawl data
filename = 'datatagmotorlistrik.csv'
search_keyword = '#motorlistrik lang:id until:2023-10-15 since:2019-01-01'
limit = 2500

!npx --yes tweet-harvest@latest -o "{filename}" -s "{search_keyword}" -l {limit} --token ""

```

Gambar 2. Program Crawling Data



Tabel 1. Hasil Pengumpulan Data Komentar

No	Komentar
1	@KompasTV Memang sudah waktunya beralih ke Kendaraan Listrik, supaya tidak ada lagi yang beli minyak dari negara arab.
2	Dengan beralih ke kendaraan listrik masyarakat berkontribusi mengurangi emisi karbon sebesar 57,5%. PLN juga telah menyediakan fasilitas pendukung charging station yang memadai dan terjangkau. #PLNDisjatim
3	inilah mengapa gue masi skeptis dengan kendaraan listrik. dah bener pake bensin aja dulu selama masi pake PLTU mah. kalo udah PLTN bolehdeh dipikirkan ðŸ™€
4	Kalau bisa kami sebagai rakyat bercerita pak Presiden @jokowi harga BBM yg saat ini saja kami anggap mahal, gimana kalau naik.... yang pasti kami rakyat akan sangat dibebankan.... Yaa Allah semoga tdk disuruh beli kendaraan listrik, Aamiin....
5	Pasar kendaraan listrik di Indonesia, semakin menarik bagi produsen luar negeri. Salah satunya, adalah VinFast yang tengah berencana untuk menggelontorkan investasi di Indonesia. @Dr_Moeldoko #MoeldokoPenjagaNKRI https://t.co/xelZABkVQo

3.2.3. Pre-Processing Data

3.2.3.1 Case Folding

Pada tahap ini data komentar yang masih memiliki huruf kapital akan diubah menjadi huruf kecil

Tabel 2. Hasil Proses Case Folding

Data Komentar Asli	Hasil Case Folding
Dengan beralih ke kendaraan listrik masyarakat berkontribusi mengurangi emisi karbon sebesar 57,5%. PLN juga telah menyediakan fasilitas pendukung charging station yang memadai dan terjangkau. #PLNDisjatim	dengan beralih ke kendaraan listrik masyarakat berkontribusi mengurangi emisi karbon sebesar 57,5%. pln juga telah menyediakan fasilitas pendukung charging station yang memadai dan terjangkau. #plndisjatim
Gembor gembor kendaraan listrik, tapi sering pemadaman listrik ðŸ™€	gembor kendaraan listrik, tapi sering pemadaman listrik ðŸ™€
Bener sih klo bandingkan kendaraan energi listrik (BEV) Vs kendaraan energi fosil (bensin/solar), hasilnya kendaraan energi listrik minim emisi karbon (polusi udara). TETAPI", energi listrik sebagian besar bersumber dari PLTU. Coba hitung ² an & bandingkan. Smart Choice yaðŸ™€"	bener sih klo bandingkan kendaraan energi listrik (bev) vs kendaraan energi fosil (bensin/solar), hasilnya kendaraan energi listrik minim emisi karbon (polusi udara). tetapi", energi listrik sebagian besar bersumber dari pltu. coba hitung ² an &

	bandingkan. smart choice yaðy~..."
Wowww keren abis nich Indonesia yg mulai tampil beda karena para investor kendaraan Listrik tujuannya bekerjasama KSP @Dr_Moeldoko #MoeldokoPenjagaNKRI https://t.co/o5ZvyF6V9R Bahas kerja sama dengan pemerintah Kota Chongqing China, KSP Moeldoko menyambut baik kerja sama pengembangan kendaraan listrik dan digitalisasi yang menjadi fokus utama pembangunan di Indonesia saat ini. Cc. @Dr_Moeldoko #MoeldokoPenjagaNKRI https://t.co/sHrKz7clyM	woww keren abis nich indonesia yg mulai tampil beda karena para investor kendaraan listrik tujuannya bekerjasama ksp #moeldokopenjagankri https://t.co/o5zvyf6v9r bahas kerja sama dengan pemerintah kota chongqing china, ksp moeldoko menyambut baik kerja sama pengembangan kendaraan listrik dan digitalisasi yang menjadi fokus utama pembangunan di indonesia saat ini. cc. #moeldokopenjagankri https://t.co/shrkz7clym

3.2.3.2 Normalisasi Teks

Proses normalisasi dilakukan untuk mengubah kata-kata yang disingkat menjadi bentuk lengkap dan memiliki makna yang sama berdasarkan Kamus Besar Bahasa Indonesia (KBBI). Hal ini bertujuan agar informasi tersebut dapat diolah dengan lebih mudah, contohnya menggantikan "utk" dengan "untuk", "yg" dengan "yang", dan sejenisnya, pada penelitian ini penulis menggunakan *slangword* bahasa Indonesia.

Tabel 3. Hasil Proses Normalisasi Teks

Hasil Case Folding	Hasil Normalisasi
pt pln (persero) ajak masyarakat merubah kebiasaan dengan beralih menggunakan kendaraan listrik. hal ini disampaikan langsung pada seminar nasional implementasi program kendaraan listrik dan green energy dalam mendukung pelayanan publik" rabu (11/10). #unleashingenergyandbeyond https://t.co/2rntv26x4	perseroan terbatas pln persero ajak masyarakat merubah kebiasaan dengan beralih menggunakan kendaraan listrik hal ini disampaikan langsung pada seminar nasional implementasi program kendaraan listrik dan green energy dalam mendukung pelayanan publik rabu 11/10 unleashingenergyandbeyond https://t.co/2rntv26x4
gw sebagai perantau yg kerja di jakarta dan kost di tangerang pakai kendaraan pribadi (motor),makan paling minim 30rb 3x makan sehari,bensin 30rb/minggu,kost 600rb/bln + listrik 50rb.. itu dah gak jajan2 ya	saya sebagai perantau yang kerja di jakarta dan indekos di tangerang pakai kendaraan pribadi motor makan paling minim 30rb 3x makan sehari bensin 30rb/minggu indekos 600rb/bln listrik 50rb itu sudah tidak jajan2 ya kalau jajan beda



Hasil Case Folding	Hasil Normalisasi
klo jajan bed a lagi. jadi klo dia bilang 150 rbu ngaco bngt.	lagi jadi kalau dia bilang 150 rbu ngaco banget
bener sih klo bandingkan kendaraan energi listrik (bev) vs kendaraan energi fosil (bensin/solar), hasilnya kendaraan energi listrik minim emisi karbon (polusi udara). tetapi", energi listrik sebagian besar bersumber dari pltu. coba hitung ² an & bandingkan. smart choice ya ² ..."	benar sih kalau bandingkan kendaraan energi listrik bev vs kendaraan energi fosil bensin/solar hasilnya kendaraan energi listrik minim emisi karbon polusi udara tetapi energi listrik sebagian besar bersumber dari pltu coba hitung ² an amp bandingkan smart choice ya ² ...
guys. prabowo kalo bikin pltn, indonesia juaraa. listrik murah dan keinginan jokowi untuk memakai kendaraan listrik jadi masuk akal.	guys prabowo kalau bikin pltn indonesia juaraa listrik murah dan keinginan jokowi untuk memakai kendaraan listrik jadi masuk akal
jadi bersyukurlah temen2 ojol masih bisa ngojol karena masih pny modal kendaraan bermotor. sedangkan saya tidak pny motor. sudah berkali2 nge-tag beliau, mungkin pernah ada kerjasama motor listrik beberapa waktu lalu, agar saya bisa daftar sbg ojol. tp no respon	jadi bersyukurlah teman teman ojol masih bisa ngojol karena masih punya modal kendaraan bermotor sedangkan saya tidak punya motor sudah berkali2 nge-tag beliau mungkin pernah ada kerja sama motor listrik beberapa waktu lalu agar saya bisa daftar sebagai ojol tapi nomor respons

3.2.3.3 Tokenizing

Tahapan ini bertujuan untuk memecah teks berupa kalimat, paragraf atau dokumen menjadi token – token atau bagian – bagian tertentu berdasarkan tiap kata.

Tabel 4. Hasil Proses Tokenizing

Hasil Normalisasi	Hasil Tokenizing
memang sudah waktunya beralih kendaraan listrik supaya tidak lagi yang beli minyak dari negara arab	['memang', 'sudah', 'waktunya', 'beralih', 'kendaraan', 'listrik', 'supaya', 'tidak', 'lagi', 'yang', 'beli', 'minyak', 'dari', 'negara', 'arab']
lima spklu siap layani pengguna kendaraan listrik sumbar https://jxusxb5cjf	['lima', 'spklu', 'siap', 'layani', 'pengguna', 'kendaraan', 'listrik', 'sumbar', 'https', './jxusxb5cjf']
dengan beralih kendaraan listrik masyarakat berkontribusi mengurangi emisi karbon sebesar , juga telah menyediakan fasilitas pendukung charging station yang memadai terjangkau plndisjatim	['dengan', 'beralih', 'kendaraan', 'listrik', 'masyarakat', 'berkontribusi', 'mengurangi', 'emisi', 'karbon', 'sebesar', ',', 'juga', 'telah', 'menyediakan', 'fasilitas', 'pendukung', 'charging', 'station', 'yang', 'memadai', 'terjangkau', 'plndisjatim']
perseroan terbatas persero ajak masyarakat merubah kebiasaan	['perseroan', 'terbatas', 'persero', 'ajak', 'masyarakat', 'merubah',



Hasil Normalisasi	Hasil Tokenizing
dengan beralih menggunakan kendaraan listrik disampaikan langsung pada seminar nasional implementasi program kendaraan listrik green energy dalam mendukung pelayanan publik rabu / unleashingenergyandbeyond https ./2rontv26x4	'kebiasaan', 'dengan', 'beralih', 'menggunakan', 'kendaraan', 'listrik', 'disampaikan', 'langsung', 'pada', 'seminar', 'nasional', 'implementasi', 'program', 'kendaraan', 'listrik', 'green', 'energy', 'dalam', 'mendukung', 'pelayanan', 'publik', 'rabu', '/', 'unleashingenergyandbeyond', 'https', './2rontv26x4']
lebih bersih mobil listrik atau mobil konvensional kementerian perindustrian indonesia resmi menyebut emisi karbon dari mobil listrik ternyata lebih tinggi ketimbang kendaraan hybrid ataupun konvensional kenapa sumber listrik mobil listrik masih dari energi fosil https ./ndcj8bnhls	['lebih', 'bersih', 'mobil', 'listrik', 'atau', 'mobil', 'konvensional', 'kementerian', 'perindustrian', 'indonesia', 'resmi', 'menyebut', 'emisi', 'karbon', 'dari', 'mobil', 'listrik', 'ternyata', 'lebih', 'tinggi', 'ketimbang', 'kendaraan', 'hybrid', 'ataupun', 'konvensional', 'kenapa', 'sumber', 'listrik', 'mobil', 'listrik', 'masih', 'dari', 'energi', 'fosil', 'https', './ndcj8bnhls']

3.2.3.4 Stopword

Pada tahapan ini bertujuan untuk menghilangkan kata-kata yang tidak memiliki arti dan pengaruh dalam analisis sentiment seperti kata “di”, “itu”, “yang” dan lain sebagainya.

Tabel 5. Hasil Proses Stopword

Hasil Tokenizing	Hasil Stopword
['memang', 'sudah', 'waktunya', 'beralih', 'kendaraan', 'listrik', 'supaya', 'tidak', 'lagi', 'yang', 'beli', 'minyak', 'dari', 'negara', 'arab']	['beralih', 'kendaraan', 'listrik', 'beli', 'minyak', 'negara', 'arab']
['lima', 'spklu', 'siap', 'layani', 'pengguna', 'kendaraan', 'listrik', 'sumbar', 'https', './jxusxb5cjf']	['spklu', 'layani', 'pengguna', 'kendaraan', 'listrik', 'sumbar', 'https', './jxusxb5cjf']
['dengan', 'beralih', 'kendaraan', 'listrik', 'masyarakat', 'berkontribusi', 'mengurangi', 'emisi', 'karbon', 'sebesar', 'juga', 'telah', 'menyediakan', 'fasilitas', 'pendukung', 'charging', 'station', 'yang', 'memadai', 'terjangkau', 'plndisjatim']	['beralih', 'kendaraan', 'listrik', 'masyarakat', 'berkontribusi', 'mengurangi', 'emisi', 'karbon', 'menyediakan', 'fasilitas', 'pendukung', 'charging', 'station', 'memadai', 'terjangkau', 'plndisjatim']
['perseroan', 'terbatas', 'persero', 'ajak', 'masyarakat', 'merubah', 'kebiasaan', 'dengan', 'beralih', 'menggunakan', 'kendaraan', 'listrik', 'disampaikan', 'langsung', 'pada', 'seminar', 'nasional', 'implementasi', 'program', 'kendaraan', 'listrik', 'green', 'energy', 'mendukung', 'pelayanan', 'publik', 'rabu', '/',	['perseroan', 'terbatas', 'persero', 'ajak', 'masyarakat', 'merubah', 'kebiasaan', 'beralih', 'kendaraan', 'listrik', 'langsung', 'seminar', 'nasional', 'implementasi', 'program', 'kendaraan', 'listrik', 'green', 'energy', 'mendukung', 'pelayanan', 'publik', 'rabu', '/',



Hasil Tokenizing	Hasil Stopword
'energy', 'dalam', 'mendukung', 'pelayanan', 'publik', 'rabu', '/', 'unleashingenergyandbeyond', 'https', './2rontv26x4']	'unleashingenergyandbeyond', 'https', './2rontv26x4']
['inilah', 'mengapa', 'saya', 'masih', 'skeptis', 'dengan', 'kendaraan', 'listrik', 'skeptis', 'dengan', 'kendaraan', 'pakai', 'bensin', 'pakai', 'pltu', 'listrik', 'sudah', 'benar', 'pakai', 'pltn', 'bolehdeh', 'dipikirkan', 'bensin', 'saja', 'dahulu', 'selama', 'masih', 'pakai', 'pltu', 'kalau', 'sudah', 'pltn', 'bolehdeh', 'dipikirkan', '£']	['skeptis', 'kendaraan', 'listrik', 'pakai', 'bensin', 'pakai', 'pltu', 'pltn', 'bolehdeh', 'dipikirkan', '£']

3.2.3 Labeling

Setelah melalui tahapan pre-processing, selanjutnya data akan diberi label dengan 2 kategori kelas yaitu positif dan negative.

Tabel 6. Hasil Proses Labeling

No	Hasil Preprocessing	Skor	Sentiment
1	memang sudah waktunya beralih kendaraan listrik supaya tidak lagi yang beli minyak dari negara arab	-1	Negatif
2	lima spklu siap layani pengguna kendaraan listrik sumbar https ./lk0ihetzes	1	Positif
3	bahas kerja sama dengan pemerintah kota chongqing china moeldoko menyambut baik kerja sama pengembangan kendaraan listrik digitalisasi yang menjadi fokus utama pembangunan indonesia saat moeldokopenjagankri https ./shrkz7clym	12	Positif
4	beli sembilan bahan pokok bayar listrik bayar pajak kendaraan bayar nyapu halaman rumah tiap pagi bayar cicilan	-14	Negatif
5	kadang baterai cacat produksi handphone saja kadang beli baterai baru tapi karena cacat produksi akhirnya drop parahnya reviewer kita tidak jujur cuma nyebutin bagusnya saja sama greenwashing tentunya ioniq5 mobillistrik electricvehicles https ./fax5zjekf4	-18	Negatif

3.2.4 Pembobotan Kata

Langkah selanjutnya dalam penelitian ini adalah penerapan pembobotan kata menggunakan metode TF-IDF (*Term Frequency - Inverse Document Frequency*).

```
[8.32778054 8.32778054 8.32778054 ... 8.32778054 8.32778054 8.32778054]
(0, 6728) 0.5741135856244586
(0, 6610) 0.29159022032210447
(0, 4128) 0.09162148741209777
(0, 4027) 0.3343820116502946
(0, 3670) 0.15474926789673418
(0, 3475) 0.6034967917223729
(0, 2878) 0.07943437325053514
(0, 2617) 0.26588977858419693
```

Gambar 3. Hasil Proses *TF-IDF*

Dalam gambar tersebut, terdapat hasil perhitungan TF-IDF (Term Frequency-Inverse Document Frequency) pada suatu dokumen. Nilai IDF (Inverse Document Frequency) untuk setiap term dalam dokumen dihitung menggunakan library Python dan ditampilkan dalam array. Nilai IDF digunakan untuk mengukur seberapa penting sebuah kata dalam kumpulan dokumen, semakin tinggi nilai IDF, semakin unik dan penting kata tersebut.

3.2.5 Klasifikasi Algoritma

Tahap ini merupakan proses dimana dibagikannya dataset menjadi dua kelompok, yaitu data *Testing* dan data *Training*.

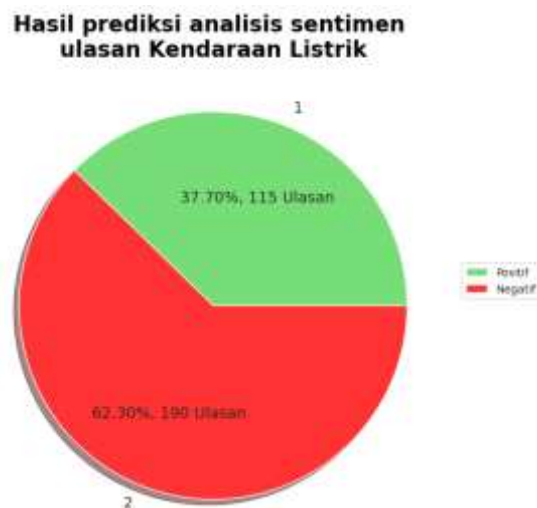
3.2.5.1 Klasifikasi Algoritma Naïve Bayes

Pada proses ini dilakukan setelah membagi data uji dan data latih, selanjutnya melakukan klasifikasi kata menggunakan algoritma naive bayes.

HASIL classification report Naive Bayes Classifier				
	precision	recall	f1-score	support
1	0.79	0.61	0.69	148
2	0.70	0.85	0.77	157
accuracy			0.73	305
macro avg	0.75	0.73	0.73	305
weighted avg	0.74	0.73	0.73	305

Gambar 4. Hasil Klasifikasi Algoritma Naïve Bayes

Hasil klasifikasi dari algoritma Naive Bayes dalam gambar 4.17 tersebut yaitu terdapat dua kelas yang diberi label sebagai 1 dan 2. Kelas 1 memiliki presisi sebesar 0.79, recall sebesar 0.61, f1-score sebesar 0.69, dan support sebesar 148. Kelas 2 memiliki presisi sebesar 0.70, recall sebesar 0.85, f1-score sebesar 0.77, dan support sebesar 157. Akurasi keseluruhan ditunjukkan sebagai rata-rata skor sekitar 0.73. Pada hasil ini penulis fokus pada hasil akurasi yaitu 0.73 yang berarti 73%. Selanjutnya melakukan proses klasifikasi sentiment menggunakan algoritma *Naïve Bayes*.



Gambar 5. Pie Chart Hasil Analisis Sentiment Algoritma Naïve Bayes

Pada gambar merupakan hasil prediksi analisis sentimen ulasan Kendaraan Listrik menunjukkan bahwa 37.70% atau 115 ulasan adalah positif, sedangkan 62.30% atau 190 ulasan adalah negatif. Dalam

konteks ini, analisis sentimen adalah teknik pengolahan bahasa alami yang digunakan untuk menentukan apakah suatu teks (dalam hal ini, ulasan kendaraan listrik) mengandung sentimen positif, negatif, atau netral. Dalam hal ini, sekitar 38% ulasan kendaraan listrik dianggap positif, sementara 62% dianggap negatif.

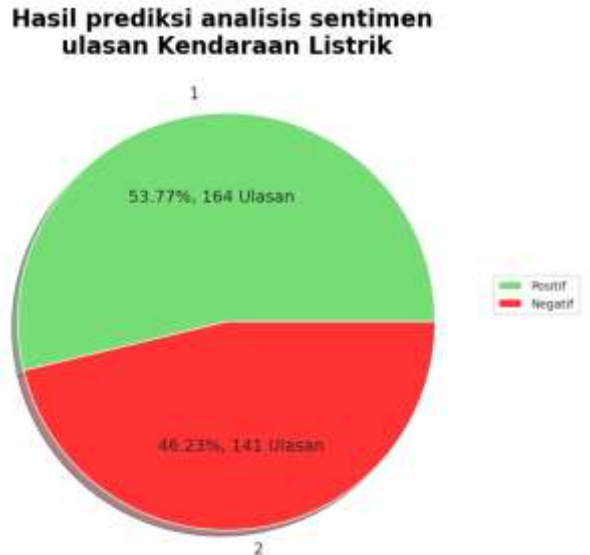
3.2.5.2 Klasifikasi *Support Vector Machine*

Pada tahapan klasifikasi selanjutnya menggunakan algoritma *Support Vector Machine* dengan menggunakan Python Library *Scikit-Learn* serta pemanggilan model SVM menggunakan `svm`, dan `accuracy_score`.

HASIL classification report SVM				
	precision	recall	f1-score	support
1	0.79	0.88	0.83	148
2	0.87	0.78	0.83	157
accuracy			0.83	305
macro avg	0.83	0.83	0.83	305
weighted avg	0.83	0.83	0.83	305

Gambar 6. Hasil Klasifikasi *Support Vector Machine*

Hasil klasifikasi algoritma *Support Vector Machine* (SVM) dalam gambar menunjukkan bahwa model tersebut memiliki akurasi sebesar 83%. Dalam tabel klasifikasi, terdapat dua kelas yang diberi label sebagai "1" dan "2". Kelas "1" memiliki presisi sebesar 0.79, recall sebesar 0.88, dan f1-score sebesar 0.83 dengan dukungan dari 148 sampel data. Sementara itu, kelas "2" memiliki presisi lebih tinggi yaitu 0.87, namun recall lebih rendah yaitu 0.78. Meskipun demikian, f1-score tetap sama dengan kelas "1" yaitu 0.83 dengan dukungan dari 157 sampel data. Pada hasil ini penulis focus pada hasil akurasi yaitu 0.83% yang berarti 83%. Selanjutnya melakukan analisis sentimen menggunakan metode algoritma *Support Vector Machine*.



Gambar 7. Pie Chart Hasil Analisis Sentimen SVM

Pada gambar tersebut merupakan hasil prediksi analisis sentimen ulasan Kendaraan Listrik menunjukkan bahwa 53.77% (164 ulasan) adalah positif dan 46.23% (141 ulasan) adalah negatif. Dalam konteks ini, analisis sentimen adalah teknik pengolahan bahasa alami yang digunakan untuk menentukan apakah suatu teks (dalam hal ini, ulasan kendaraan listrik) mengandung sentimen positif, negatif, atau netral. Dalam hal ini, sekitar 54% ulasan kendaraan listrik dianggap positif, sementara 46% dianggap negative.

4. KESIMPULAN

Berdasarkan hasil dari penelitian yang telah dilakukan pada data komentar pengguna media sosial twitter terkait opini publik mengenai Kendaraan Listrik di Indonesia yang berjumlah 3.043 data maka dapat ditarik kesimpulan bahwa penerapan algoritma Navie Bayes Classifier (NBC) dan algoritma Support Vector Machine (SVM) sudah berhasil menganalisis komentar dengan 2 kategori kelas sentiment yaitu positif dan negatif. Dengan diterapkannya metode algoritma Naive Bayes Classifier (NBC) untuk menganalisis sentimen Kendaraan Listrik di Indonesia menunjukkan hasil performa akurasi berjumlah 0.73 yang berarti berjumlah 73%. Sedangkan hasil dari penerapan metode algoritma Support Vector Machine (SVM) menunjukkan hasil performa akurasi berjumlah 0.82 atau sekitar 82.95%.

Berdasarkan hasil dari klasifikasi algoritma Naïve Bayes Classifier (NBC) menunjukkan hasil presentase sentiment berdasarkan data uji yaitu berjumlah 37.7% sentiment positif dengan total 115 data uji dan 62.3% sentiment negative dengan total 190 data uji. Sementara hasil dari klasifikasi algoritma Support Vector Machine (SVM) menunjukkan hasil presentase sentiment berdasarkan data uji yaitu berjumlah 53.77% sentiment positif dengan total 164 data uji dan 46.23% sentiment negative dengan total 141 data uji. Dari hasil klasifikasi kedua metode algoritma yang sudah diterapkan dapat ditarik kesimpulan bahwa dengan menerapkan algoritma Support Vector Machine (SVM) mendapatkan hasil dengan akurasi yang lebih baik dibandingkan dengan algoritma Naïve Bayes Classifier (NBC).

REFERENSI

REFERENSI

- [1] A. Agustian, T. Tukiro, and F. Nurapriani, “Penerapan Analisis Sentimen Dan Naive Bayes Terhadap Opini Penggunaan Kendaraan Listrik Di Twitter,” *Jurnal TIKA*, vol. 7, no. 3, pp. 243–249, 2022, doi: 10.51179/tika.v7i3.1550.
- [2] Y. Pratama, D. T. Murdiansyah, and K. M. Lhaksana, “Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression dan Principal Component Analysis,” *Jurnal Media Informatika Budidarma*, vol. 7, no. 1, pp. 529–535, 2023, doi: 10.30865/mib.v7i1.5575.
- [3] A. Rahman, E. Utami, and S. Sudarmawan, “Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naïve Bayes dan Algoritma Genetika,” *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, no. 1, pp. 60–71, 2021, doi: 10.31603/komtika.v5i1.5188.
- [4] K. A. Rokhman, B. Berlilana, and P. Arsi, “Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online,” *Journal of Information System Management (JOISM)*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.24076/joism.2021v3i1.341.
- [5] S. Kulkarni, Srikanth., Buradkar, Ayush., Ghadge, Pratiksha., Khainar, “Web Scraping : Extracting Insights from the,” no. May, 2023.
- [6] D. Ø. Madsen and E. S. Silva, “15 years of research on Google Trends : A bibliometric review and future research directions,” 2023.
- [7] D. Varma, A. Nehansh, and P. S. Guide, “Data Preprocessing Toolkit : An Approach to Automate Data Preprocessing,” pp. 1–5, 2023, doi: 10.55041/IJSREM18270.
- [8] H. Stachel, “On the Computation of Foldings,” pp. 268–275, 2017, doi: 10.5937/fmet1702268S.
- [9] M. Lusetti and G. Anne, “Encoder-Decoder Methods for Text Normalization,” no. August, pp. 18–28, 2018.
- [10] R. Friedman, “Tokenization in the Theory of Knowledge,” pp. 380–386, 2023.



- [11] R. Hakim, "Topic Modelling Dokumen Skripsi Menggunakan Metode Latent Semantic," pp. 1–62, 2020.
- [12] M. Al Khadafi *et al.*, "PENERAPAN METODE NAÏVE BAYES CLASSIFIER DAN LEXICON BASED UNTUK ANALISIS SENTIMEN CYBERBULLYING PADA BPJS," *JATI(Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 725–733, 2022.
- [13] M. U. Albab, Y. K. P, and M. N. Fawaiq, "Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic," *TRANSFORMATIKA*, vol. 20, no. 2, pp. 1–10, 2023.
- [14] Y. Kalmukov, "USING WORD CLOUDS FOR FAST IDENTIFICATION," *Proc. Univ. Ruse*, vol. 60, pp. 114–119, 2021.
- [15] N. Kadek, R. Sari, I. M. Agus, D. Suarjaya, and P. Wira, "Perbandingan Translation Library Pada Python (Studi Kasus : Analisis Sentimen Penyakit Menular Di Indonesia)," *JITTER-Jurnal Ilm. Teknol. dan Komput.*, vol. 2, no. 3, 2021.
- [16] I. Albanna, R. Tri, J. S. Informasi, I. Teknologi, and A. Tama, "Implementasi Pandas Data frame sebagai Agregasi dan Tabulasi Penyajian Data Luaran Survei Kepuasan Pengguna Proses Pembelajaran dalam Pendidikan Tinggi," *Semin. Nas. Sains dan Teknol. Terap. X 2022 Inst. Teknol. Adhi Tama Surabaya*, pp. 1–6, 2022.
- [17] F. Regis, V. Alves, R. Passos, and M. Vieira, "The Newton Fractal ' s Leonardo Sequence Study with the Google Colab," *Int. Electron. J. Math. Educ.*, vol. 15, no. 2, 2020.
- [18] M. Ahp and E. D. A. N. Topsis, "SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN MAHASISWA BERPRESTASI DI UNIVERSITAS DHYANA PURA MENGGUNAKAN METODE AHP, ELECTRE DAN TOPSIS," *J. Ilmu Komput. Indones.*, no. 1, pp. 22–33, 2019.
- [19] I. Santoso, "ANALISIS SENTIMEN PADA TWITTER TERHADAP GAGALNYA PELAKSANAAN PIALA DUNIA DI INDONESIA MENGGUANAKAN METODE NAÏVE BAYES," 2023. [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/issue/archive>
- [20] E. Suryati, A. Ari Aldino, N. Penulis Korespondensi, and E. Suryati Submitted, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," vol. 4, no. 1, pp. 96–106, 2023, doi: 10.33365/jtsi.v4i1.2445.