

Applying Naive Bayes Classifier Method for Sentiment Classification of Electric Cars

Nurul Afifah, Dony Permana*, Dodi Vionanda, dan Dina Fitria

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: donypermana@fmipa.unp.ac.id

Submitted : 30 Mei 2023

Revised : 25 Juli 2023

Accepted : 08 Agustus 2023

ABSTRACT

In recent years, electric cars have become increasingly popular as an alternative to environmentally friendly vehicles in the automotive industry. These vehicles use electric power as an energy source that can mitigate the reliance on fossil fuels contribute to efforts to minimize greenhouse gas emissions and air pollution. However, the presence of electric cars raises pro and con opinions from the public. the conversation about electric cars has become one of the hot on social media. Twitter is a social media microblogging that permits its users to create short messages and share them easily and quickly. These opinions require sentiment analysis. The purpose of conducting sentiment analysis is to find out how people's perceptions and opinions on electric cars are leading in a favorable or unfavorable direction. Thus, sentiment analysis can help companies marketing strategies, and better business decisions. Then the opinions will be classified based on positive and negative categories. This investigation employs the naive classifier method to generate positive and negative sentiment towards electric cars on Twitter. The accuracy results of naive bayes obtained by using a confusion matrix in this research are 77.8%, with a dataset split composition of 70%:30%.

Keywords: *electric cars, machine learning, naïve bayes, sentiment analysis.*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Mobil listrik merupakan jenis kendaraan bermotor yang memanfaatkan energi listrik sebagai sumber tenaga untuk beroperasi. Mobil listrik hadir sebagai solusi dari melonjaknya harga minyak bumi yang menyebabkan naiknya harga BBM yang merupakan sumber penggerak kendaraan konvensional (Hrp & Aslami, 2022). Penggunaan minyak bumi tidak hanya untuk kendaraan, tetapi juga diolah sebagai produk lain seperti oli dan aspal. Sehingga, jika bertambahnya pemakaian minyak bumi, otomatis persediaan BBM juga terus berkurang karena minyak bumi termasuk energi yang tak dapat diperbaharui.

Berdasarkan data dari www.dataindonesia.com, “terjadi peningkatan penjualan mobil listrik di Indonesia dari tahun 2021 ke tahun 2022 sebesar 383,46%”. Melihat perkembangan tersebut, mobil listrik memungkinkan untuk menarik perhatian positif dari masyarakat. Namun, kehadiran mobil listrik tetap menuai pro dan kontra. Masyarakat masih memperdebatkan keuntungan dan kerugian antara memakai mobil konvensional atau mobil listrik. Beberapa hal yang dipertimbangkan adalah harga mobil listrik yang lebih mahal dari mobil konvensional, tempat pengisian bahan bakar yang masih minim di Indonesia, serta jarak tempuh dan kecepatan mobil listrik yang masih terbatas (Parinduri *et al.*, 2018). Analisis sentimen mengenai persepsi masyarakat terhadap mobil listrik sangat penting bagi produsen mobil listrik dan pemerintah dalam merancang strategi pemasaran dan kebijakan yang efektif. Melalui sentimen positif dan negatif tersebut, produsen dapat melakukan evaluasi untuk meningkatkan kualitas dan daya tarik dari mobil listrik. Oleh karenanya, diperlukanlah analisis sentimen mengenai mobil listrik.

Analisis sentimen bertujuan untuk memperoleh opini, mengidentifikasi perasaan seseorang yang disalurkan melalui teks, yang kemudian diklasifikasikan berdasar nilai-nilai yang terkandung di dalam teks (Asnawi *et al.*, 2021). Opini tersebut diperoleh dengan memanfaatkan sosial media twitter. Twitter adalah sosial media berbasis *microblogging*, yang sering digambarkan sebagai “jurnalisme amatir”, dimana para pengguna saling berbagi konten mengenai peristiwa dan situasi terkini yang sedang hangat diperbincangkan (Pozzi *et al.*, 2017:16).

Untuk mengklasifikasikan opini-opini tersebut, diperlukan metode *Naive Bayes Classifier* yang merupakan salah satu teknik klasifikasi *machine learning*. *Naive Bayes Classifier* dikenal dengan kesederhanaan dan

efisiensinya ketika melakukan komputasi, sehingga sangat cocok untuk diterapkan pada analisis sentimen dengan data yang berjumlah besar dan bersifat *real-time*. Dimana metode ini didasarkan pada *probability* dan hipotesis Teorema Bayes dengan anggapan setiap variabel X sifatnya *independent* (bebas). Pada riset yang dilaksanakan oleh Yulita *et al.*, (2021), diperoleh bahwa pengujian dengan *Naïve Bayes* menghasilkan akurasi sebesar 93%. Selain itu, pada penelitian Fajar Ratnawati (2018), *Naïve Bayes* menunjukkan performa yang tinggi dengan tingkat akurasi 90%.

Dari penjelasan tersebut, maka dilakukanlah penelitian yang bertujuan untuk mengklasifikasikan sentimen mobil listrik kedalam dua kategori yaitu positif dan negatif menggunakan data twitter dengan teknik klasifikasi *naïve bayes*.

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data yang digunakan merupakan data sekunder yang diperoleh dengan melakukan *crawling* atau penarikan data twitter menggunakan bantuan *google colab*. Data ini mengenai *tweet-tweet* masyarakat Indonesia mengenai mobil listrik pada bulan Desember 2022. Variabel penelitiannya yaitu variabel prediktor yang berupa *tweet* dari para pengguna twitter dan variabel respon yang berupa klasifikasi sentimen *tweet* positif dan negatif.

B. Tahapan Analisis Data

Tahapan analisis data sentimen dengan metode klasifikasi *naïve bayes* adalah sebagai berikut.

1. Melakukan Analisis Sentimen

Analisis sentimen merupakan bidang studi yang mengkaji pendapat, penilaian, sikap, dan emosi orang terhadap layanan, produk, organisasi, individu, atau lainnya, dalam ruang lingkup masalah besar yang diungkapkan dalam teks tertulis. Sentimen ialah tentang perasaan dan keyakinan subjektif, dimana hal tersebut adalah pusat psikologi manusia yang memberikan pengaruh utama dari perilaku yang dilakukan (Liu, 2015:1). Sebelum data sentimen diolah, perlu dilakukan *text preprocessing* terlebih dahulu.

Text preprocessing bertujuan menyiapkan data teks mentah agar menjadi data terstruktur. Tahapan yang dilakukan antara lain (Asnawi *et al.*, 2021), pertama *cleaning* data yakni membersihkan teks dari elemen-elemen yang tidak berpengaruh terhadap hasil klasifikasi, seperti tanda baca, simbol-simbol khusus, url, angka, dan sebagainya. Kedua yaitu *case folding*, proses perubahan semua karakter dalam teks menjadi *lower case* (huruf kecil). Ketiga, *tokenization* merupakan suatu metode pemisahan teks menjadi unit-unit yang lebih kecil dalam bentuk kata, sehingga hasilnya disajikan dalam bentuk data matriks. Keempat, *stopword removal* dimana *stopword* merupakan kosakata yang bukan kata informatif pada suatu dokumen, sehingga *stopword removal* adalah proses pembuangan kata yang termasuk dalam kategori tidak penting atau tidak relevan. Terakhir, *stemming* berfungsi untuk mengganti kata-kata yang berimbuhan menjadi bentuk dasarnya.

2. Pelabelan Data

Pelabelan data dilakukan dengan memanfaatkan metode *lexicon based*. Analisis sentimen berbasis leksikon adalah suatu analisis data yang dilakukan dengan menggunakan kata-kata dan frasa opini tanpa ada pengetahuan sebelumnya (Verma & Thakur, 2018). Kata-kata yang mengandung opini tersebut disusun dan dikumpulkan. Kata-kata positif dan negatif di dalam suatu pendapat disebut leksikon opini yang akan dievaluasi berdasarkan leksikon opini untuk menentukan orientasi dan sentimen teks.

3. Melakukan TF-IDF

Suatu teknik pembobotan yang bertujuan untuk mengukur tingkat relevansi antara kata dan dokumen dengan pemberian bobot pada setiap kata. Tujuan pembobotan yaitu untuk mengubah data teks menjadi data numerik. TF-IDF adalah kombinasi antara *term-frequency* (TF) yaitu seberapa kerap sebuah kata muncul dalam sebuah dokumen dan *inverse document frequency* (IDF) yang mengindikasikan seberapa umum kata tersebut di seluruh dokumen (Widyasanti *et al.*, 2018). Berikut persamaan TF, IDF, dan TF-IDF.

$$tf_{ij} = \frac{freq_i(d_j)}{\sum_{i=1}^k freq_i(d_j)} \quad (1)$$

Persamaan (1) digunakan untuk menggambarkan seberapa penting peran kata dalam sebuah dokumen yang memberikan nilai bobot yang tinggi pada kata-kata yang sering muncul dalam sebuah dokumen.

$$idf_i = \log \frac{N}{df_i} \quad (2)$$

Rumus IDF untuk mengukur seberapa penting sebuah kata dalam korpus dokumen secara keseluruhan yang memberikan nilai bobot yang lebih tinggi pada kata-kata yang jarang muncul. IDF membantu dalam mengidentifikasi kata-kata unik dalam dokumen tertentu.

$$(tfidf)_{ij} = tf_{ij} \times idf_i \quad (3)$$

TF-IDF memberikan skor bobot yang lebih tinggi pada kata-kata yang sering muncul dalam sebuah dokumen (TF tinggi) dan jarang muncul dalam dokumen lain dalam korpus tinggi (IDF tinggi). TF-IDF berguna untuk menghitung bobot kata yang memberikan representasi yang lebih baik tentang seberapa pentingnya sebuah kata dalam konteks dokumen-dokumen yang ada.

Keterangan:

$freq_i(d_j)$: jumlah kejadian *term* ke-*i* dalam dokumen ke-*j*

idf_i : logaritma dari perbandingan antara total dokumen dengan jumlah dokumen yang memiliki *term* yang dimaksud

N : total koleksi dokumen dalam database

df_i : total dokumen yang mencakup *term*-*i*

$(tfidf)_{ij}$: nilai bobot *term* ke-*i* yang terkait dokumen ke-*j*

i : 1, 2, 3, ..., k dengan k adalah banyaknya *term* yang terbentuk

j : 1, 2, 3, ..., n dimana n adalah banyak dokumen

4. Pembagian Data *Training* dan *Testing*

Data terbagi ke dalam data *training* dan data *testing* dengan rasio pembagian 70%:30%, 80%:20%, dan 90%:10%. Tiga komposisi pembagian ini dilakukan karena tidak ada panduan yang jelas mengenai rasio mana yang terbaik atau yang optimal dalam pembagian data menjadi data *training* dan data *testing* (Joseph, 2022). Sehingga, untuk memperoleh rasio yang optimal maka pembagian dataset akan diujicobakan dengan tiga rasio tersebut.

5. Klasifikasi Sentimen dengan *Naïve Bayes*

Naïve Bayes Classifier dikenal dengan kemudahan dan kesederhanaan dalam melakukan pengolahan data (Asnawi *et al.*, 2021). *Naïve bayes* bekerja dengan memanfaatkan Teorema *Bayes* dengan asumsi antar variabel prediktor bersifat saling bebas. Asumsi saling bebas mengimplikasikan bahwa atribut tertentu pada suatu kelas tidak bergantung pada keberadaan atribut lainnya.

Naïve Bayes memerlukan data pelatihan untuk memperoleh probabilitas. Setelah hasil probabilitas diperoleh, langkah selanjutnya adalah proses klasifikasi, dimana probabilitas setiap atribut data pada setiap label kelas dihitung menggunakan data *testing*. Kemudian, kelas dengan hasil probabilitas tertinggi akan menjadi label untuk data *testing* tersebut. Berikut persamaan yang digunakan (Santoso & Wibowo, 2022).

$$P(V_j|a_i) = \frac{P(a_i|V_j)P(V_j)}{P(a_i)} \quad (4)$$

dimana,

$P(V_j|a_i)$: probabilitas kategori- j ketika *term*- i muncul

$P(a_i|V_j)$: probabilitas *term*- i masuk ke dalam kategori- j

$P(V_j)$: probabilitas jumlah kejadian kategori- j

$P(a_i)$: probabilitas jumlah munculnya kata ke- i

i : penomoran kata dimulai dari 1 hingga ke- k

j : label kategori dimulai dari 1 hingga ke- n

Teorema bayes pada persamaan (4) digunakan untuk menghitung probabilitas kelas sentimen berdasarkan kemunculan kata-kata dalam teks. Setiap *tweet* pada *naïve bayes* diwakili oleh rangkaian atribut (a_1, a_2, \dots, a_k) dengan a_1 sebagai kata pertama, a_2 sebagai kata kedua dan seterusnya. Pada saat pengklasifikasian, *naïve bayes* akan menentukan kategori berdasar probabilitas yang tertinggi dengan menginputkan atribut (a_1, a_2, \dots, a_k) dengan

V sebagai himpunan kelas. Dimana V_{MAP} sebagai kategori yang memiliki probabilitas tertinggi. Berikut rumus yang digunakan.

$$V_{MAP} = \underset{V_j \in V}{argmax} P(V_j) \times \prod_{i=1} P(a_i|V_j) \tag{5}$$

Persamaan (5) digunakan untuk memprediksi kategori suatu data berdasarkan kemunculan atribut. Dimana, $P(V_j)$ dan probabilitas a_i untuk masing-masing kelas $P(a_i|V_j)$ dihitung berdasarkan data *training*, yang diungkapkan sebagai berikut.

$$P(a_i|V_j) = \frac{n_i + 1}{n + kosakata} \tag{6}$$

Persamaan (6) digunakan untuk menghitung probabilitas kemunculan kata dalam kelas sentimen yang berbeda. Probabilitas yang diperoleh digunakan secara bersama dengan probabilitas kelas sentimen yang lain sehingga diperoleh probabilitas secara keseluruhan yang akan digunakan untuk mengklasifikasikan dokumen baru ke dalam kelas sentimen yang ditentukan. Dimana n_i adalah frekuensi kata a_i dalam kategori V_j , n adalah total kosakata yang ada dalam kategori V_j dan kosakata adalah total data secara keseluruhan.

6. Evaluasi Model

Proses klasifikasi memerlukan evaluasi kinerja dari sistem klasifikasi. Evaluasi kinerja klasifikasi dapat menggunakan *confusion matrix* berupa tabel dalam menggambarkan kinerja algoritma secara jelas. Setiap baris dari matriks menggambarkan kelas sebenarnya dari data sedangkan setiap kolom menggambarkan kelas prediksi. Berikut tabel *confusion matrix* yang digunakan (Saputro & Sari, 2019).

Tabel 1. Confusion Matrix

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	<i>True Negative (TN)</i>	<i>False Positive (FP)</i>
Aktual Positif	<i>False Negative (FN)</i>	<i>True Positive (TP)</i>

Rumus untuk menghitung akurasi adalah sebagai berikut.

$$Accuracy = \frac{TP + TN}{Total} \tag{7}$$

Perhitungan akurasi berguna untuk mengukur tingkat keberhasilan dari suatu model klasifikasi, yakni dengan menghitung persentase dari jumlah prediksi yang benar dibandingkan dengan total prediksi yang dilakukan.

III. HASIL DAN PEMBAHASAN

A. Preprocessing Data

Teks preprocessing terdiri dari tahap *cleaning* data yang mencakup menghilangkan *username*, *url*, *hashtags* dan tanda baca. Selanjutnya yaitu tahapan *case folding*, *stopword removal*, *tokenization*, dan *stemming*. Tampilan data yang telah melalui *text preprocessing* dapat dilihat pada tabel berikut.

Tabel 2. Hasil Cleaning Data

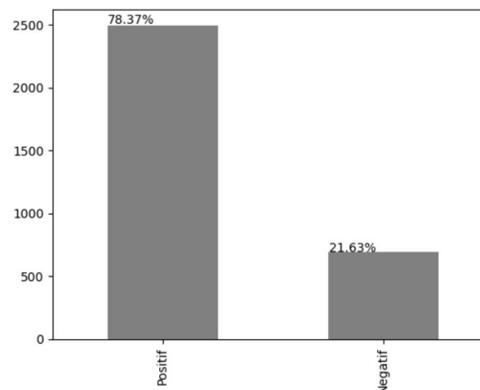
Sebelum Cleaning	Setelah Cleaning	Sentimen
Ekspektasi ku kirain mobil listrik tuh bakal slow gitu kaan.. TERNYATA real nya kendaraan listrik ga kalah gesit dan lincah loh, jujur baru pertama kali coba tapi langsung jatuh cinta! #pakemotorlistrikyuk @CNNIndonesia Lucu. Subsidi rp 80 juta per orang buat pembeli mobil?	ekspektasi kirain mobil listrik slow gitu kaan real kendaran listrik kalah gesit lincah loh jujur kali coba langsung jatuh cinta pakemotorlistrikyuk	Positif
	lucu subsidi rupiah juta orang beli mobil korupsi subsidi rakyat miskin mending	Negatif

Sebelum <i>Cleaning</i>	Setelah <i>Cleaning</i>	Sentimen
Ini korupsi. Subsidi itu untuk rakyat miskin. Mending buat subsidi listrik rumah yg saat ini mencekik leher	subsidi listrik rumah cekik leher	

Berdasar Tabel 2, tampak bahwa tweet yang belum dilakukan *preprocessing* tidak terstruktur secara baik. Sehingga diperlukanlah *cleaning* data untuk menghilangkan tanda baca, simbol-simbol, dan angka. Kemudian melakukan *case folding* yakni pengubahan karakter huruf menjadi *lower case*. Menghapus kata-kata yang tidak bermakna yang termasuk dalam tahap *stopword removal*. Selanjutnya, proses mengubah kata ke bentuk dasarnya yang disebut dengan tahap *stemming*.

B. Pelabelan Data

Pelabelan data dilakukan secara otomatis dengan menggunakan metode *lexicon based*. Hasil pelabelan tersebut dapat dilihat melalui visualisasi sebaran dari analisis sentimen berdasar kategori positif dan negatif pada Gambar 1.



Gambar 1. Distribusi Kelas Sentimen

Berdasar Gambar 1, tampak bahwa sentimen masyarakat terhadap mobil listrik cenderung ke arah positif dibanding ke arah negatif, yang mengindikasikan bahwa sebagian besar masyarakat memiliki pandangan yang positif terhadap mobil listrik. Sentimen positif yang dominan menunjukkan adanya ketidakseimbangan kelas. Dimana persentase kategori positif sebesar 78.37% dan kategori negatif sebesar 21.63%.

C. Naive Bayes Classifier

Pada bahasa pemrograman *python*, penerapan *naive bayes* memanfaatkan *library sklearn.naive_bayes* dengan memuat modul *MultinomialNB* terlebih dahulu. Sedangkan untuk cara kerja manual dapat dijabarkan sebagai berikut.

Tabel 3. Data Training

NO	Tweet	Kelas
1	ekspektasi kirain mobil listrik slow gitu real kendaraan listrik tidak kalah gesit	P
2	lucu subsidi rupiah juta orang beli mobil korupsi subsidi untuk rakyat miskin	N

Untuk menghitung probabilitas setiap kata pada data *training* menggunakan persamaan (6). Perhitungan probabilitas kata 'ekspektasi' adalah sebagai berikut.

$$P(\text{ekspektasi} | V_{\text{positif}}) = \frac{1 + 1}{12 + 21} = \frac{2}{33}$$

$$P(\text{ekspektasi} | V_{\text{negatif}}) = \frac{0 + 1}{12 + 21} = \frac{1}{33}$$

Probabilitas setiap kata dapat dilihat pada Tabel 4 dan 5.

Tabel 4. Probabilitas Tweet 1

Kelas	P(V _j)	P(a _i V _j)										
		ekspektasi	kirain	mobil	listrik	slow	gitu	real	kendaraan	tidak	kalah	gesit
P	0,5	0,06	0,06	0,06	0,09	0,06	0,06	0,06	0,06	0,06	0,06	0,06
N	0,5	0,03	0,03	0,06	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03

Tabel 5. Probabilitas Tweet 2

Kelas	P(V _j)	P(a _i V _j)										
		lucu	subsidi	rupiah	juta	orang	beli	mobil	korupsi	untuk	rakyat	miskin
P	0,5	0,03	0,03	0,03	0,03	0,03	0,03	0,06	0,03	0,03	0,03	0,03
N	0,5	0,06	0,09	0,06	0,06	0,06	0,06	0,06	0,06	0,06	0,06	0,06

Setelah probabilitas dihitung, selanjutnya menggunakan data testing untuk memprediksi kelas dari tweet.

Tabel 6. Data Testing

Tweet	Kelas
orang kaya disuruh beli mobil listrik subsidi	Belum Diketahui

Untuk kelas positif.

$$\begin{aligned}
 P(\text{tweet} | V_{\text{positif}}) &= P(V_{\text{positif}}) \times P(\text{orang} | V_{\text{positif}}) \times P(\text{kaya} | V_{\text{positif}}) \times P(\text{disuruh} | V_{\text{positif}}) \times P(\text{beli} | V_{\text{positif}}) \times P(\text{mobil} | V_{\text{positif}}) \\
 &\quad \times P(\text{listrik} | V_{\text{positif}}) \times P(\text{subsidi} | V_{\text{positif}}) \\
 &= 0,5 \times 0,03 \times 1 \times 1 \times 0,03 \times 0,06 \times 0,09 \times 0,03 \\
 &= 0,00000007
 \end{aligned}$$

Untuk kelas negatif.

$$\begin{aligned}
 P(\text{tweet} | V_{\text{negatif}}) &= P(V_{\text{negatif}}) \times P(\text{orang} | V_{\text{negatif}}) \times P(\text{kaya} | V_{\text{negatif}}) \times P(\text{disuruh} | V_{\text{negatif}}) \times P(\text{beli} | V_{\text{negatif}}) \times \\
 &\quad P(\text{mobil} | V_{\text{negatif}}) \times P(\text{listrik} | V_{\text{negatif}}) \times P(\text{subsidi} | V_{\text{negatif}}) \\
 &= 0,5 \times 0,06 \times 1 \times 1 \times 0,06 \times 0,06 \times 0,03 \times 0,09 \\
 &= 0,00000029
 \end{aligned}$$

Melalui perhitungan probabilitas setiap kelas menggunakan persamaan (5), maka data testing yang kelasnya belum diketahui termasuk ke dalam kelas negatif, sebab nilai probabilitas kelas negatif lebih besar dibanding kelas positif.

D. Performansi Algoritma

Confusion matrix digunakan untuk melihat performansi dari sebuah algoritma. Tabel *confusion matrix* untuk pembagian data training dan testing 70%:30% dapat dilihat sebagai berikut.

Tabel 7. *Confusion Matrix*

Kelas Prediksi	Kelas Aktual	
	Negatif	Positif
Negatif	4	0
Positif	212	740

Berdasar Tabel 7, diketahui model mengklasifikasikan secara benar sebesar 740 data sebagai positif dan 4 data sebagai negatif. Selain itu, 212 data negatif diprediksi positif dan 0 data positif diprediksi negatif. Dimana perhitungan akurasi menggunakan persamaan (7) adalah sebagai berikut.

$$Akurasi = \frac{740 + 4}{740 + 4 + 0 + 212} = \frac{744}{956} = 0,778$$

Hasil akurasi untuk pengujian semua komposisi pembagian data training dan testing dapat dilihat pada Tabel 8.

Tabel 8. Pengujian Algoritma

Pengujian	Accuracy
70%:30%	0,778
80%:20%	0,757
90%:10%	0,733

Berdasar Tabel 8, pembagian data training dan data testing dengan komposisi 70%:30% memperoleh nilai akurasi tertinggi. Dimana perolehan nilai akurasi sebesar 77,8%.

IV. KESIMPULAN

Dari hasil dan pembahasan yang telah disajikan, disimpulkan bahwa sentimen masyarakat terhadap kehadiran mobil listrik cenderung ke arah positif dengan persentase sentimen positif sebesar 78,37% dan sentimen negatif sebesar 21,63%. Dimana tingkat akurasi yang dihasilkan dengan penerapan komposisi pemisahan data pelatihan dan uji sebesar 70%:30% adalah 77,8%. Saran untuk penelitian selanjutnya yaitu diperlukan referensi khusus berbahasa Indonesia untuk mengidentifikasi kata-kata yang tidak relevan pada tahap *text preprocessing*, terutama dalam pemakaian kata yang tidak baku, agar hasil analisis sentimen yang diperoleh lebih akurat.

DAFTAR PUSTAKA

- Asnawi, M. H., Firmansyah, I., Novian, R., & Pontoh, R. S. (2021). Perbandingan Algoritma Naive Bayes, K-NN, dan SVM dalam Pengklasifikasian Sentimen Media Sosial. *Seminar Nasional Statistika X*.
- Joseph, V. R. (2022). *Optimal Ratio for Data Splitting*. Georgia Institute of Technology: USA.
- Hrp, G. R., & Aslami, N. (2022). Analisis Dampak Kebijakan Perubahan Publik Harga BBM terhadap Perekonomian Rakyat Indonesia. *Ilmu Komputer, Ekonomi, dan Manajemen*, 2(1), 1464-1474.
- Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. New York: Cambridge University Press.
- Mustajab, R. (2023, January 18). *Penjualan Mobil Listrik di Indonesia Capai 15.437 Unit pada 2022*. Retrieved from dataindonesia.id: <https://dataindonesia.id/sektor-riil/detail/penjualan-mobil-listrik-di-indonesia-capai-15437-unit-pada-2022>
- Parinduri, L., Yusmartato, & Parinduri, T. (2018). Kontribusi Konversi Mobil Konvensional ke Mobil Listrik dalam Penanggulangan Pemanasan Global. *Journal of Electrical Technology*, 3(2), 116-120.
- Pozzi, F. A., Fersini, E., Messina, E., & Liu, B. (2017). *Sentiment Analysis in Social Networks*. United States: Elsevier
- Ratnawati, Fajar. (2018). Implementasi Algoritma Naïve Bayes terhadap Analisis Sentimen Opini Film pada Twitter. *Jurnal Inovtek Polbeng Seri Informatika*, 3(1).
- Santoso, D. P., & Wibowo, W. (2022). Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store. *Jurnal Sains dan Seni ITS*, 11(2).
- Saputro, I. W., & Sari, B. W. (2019). Uji Performa Algoritma Naive Bayes untuk Prediksi Masa Studi Mahasiswa. *Citec Journal*, 6(1).

- Verma, B., & Thakur, R. S. (2018). Sentiment Analysis Using Lexicon and Machine Learning-Based Approaches A Survey. *Proceedings of International Conference on Recent Advancement on Computer and Communication*.
- Widyasanti, N. K., Putra, I. K., & Rusjayanthi, N. K. (2018). Seleksi Fitur Bobot Kata dengan Metode TF-IDF untuk Ringkasan Bahasa Indonesia. *Jurnal Evolusi*, 4(1), 25-31.
- Yulita, W., Nugroho, E. D., & Algifari, M., H. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Data Mining dan Sistem Informasi*, 2(2), 1-9.