



KLASIFIKASI SENTIMEN TERHADAP KEMUNCULAN MOBIL LISTRIK MENGUNAKAN METODE NAÏVE BAYES

Andrean Yhuda¹; Eka Setio Pangestuti²; Hidayat Chandra³; Terima Berkat
Nduru⁴; Wahyu Supriyanto⁵; Muhammad Muharrom⁶

¹²³⁴⁵⁶Teknologi Informasi, Universitas Bina Sarana Informatika

¹Correspondence Email: 17230192@bsi.ac.id

Received: 18/12/2025	Accepted: 20/12/2025	Published: 31/12/2025
----------------------	----------------------	-----------------------

Abstract

The emergence of electric vehicles has triggered diverse public opinions on social media, reflecting hopes for environmental sustainability alongside concerns regarding functional aspects. This study aims to classify public sentiment toward electric vehicles using the Naïve Bayes algorithm. Data were obtained from the Kaggle platform and processed through pre-processing stages, including cleaning, case folding, stopword removal, and label correction to ensure dataset quality. The analysis was conducted using RapidMiner software by applying the Naive Bayes, Apply Model, and Performance operators. The results indicate that the model achieved a perfect accuracy rate of 100%, with precision, recall, and f1-score values of 1.00 across all classes (positive, negative, and neutral). This achievement is influenced by the characteristics of the dataset, which possesses a clearly separated distribution of linguistic features and minimal noise following the pre-processing stage. Although the model demonstrates exceptional performance, the implementation of k-Fold Cross-Validation techniques is highly recommended to ensure the model's validity and generalizability to more complex real-time data. This study confirms that Naïve Bayes remains a highly competitive instrument for mapping public opinion to support energy transition policymaking.

Keywords: *Electric Vehicle, Naïve Bayes, Sentiment Analysis, RapidMiner, Text Mining..*

Abstrak

Kemunculan mobil listrik memicu beragam opini publik di media sosial yang mencerminkan harapan akan keberlanjutan lingkungan sekaligus kekhawatiran terhadap aspek fungsional. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap mobil listrik menggunakan algoritma Naïve Bayes. Data diperoleh dari platform Kaggle dan diproses melalui tahapan *pre-processing* yang meliputi *cleaning*, *case folding*, *stopword removal*, dan perbaikan label untuk memastikan kualitas dataset. Analisis dilakukan menggunakan perangkat lunak RapidMiner dengan menerapkan operator *Naive Bayes*, *Apply Model*, dan *Performance*. Hasil penelitian menunjukkan bahwa model mampu mencapai tingkat akurasi sempurna sebesar 100% dengan nilai *precision*, *recall*, dan *f1-score* sebesar 1.00 pada semua kelas (positif, negatif, dan netral). Capaian ini dipengaruhi oleh karakteristik dataset yang memiliki distribusi fitur linguistik yang terpisah secara kontras dan minimal *noise* setelah proses *pre-processing*. Meskipun model menunjukkan performa luar biasa, penggunaan teknik *k-Fold Cross-Validation* sangat disarankan untuk memastikan validitas dan kemampuan generalisasi model pada data real-time yang lebih kompleks. Penelitian ini menegaskan bahwa Naïve Bayes tetap menjadi instrumen yang sangat kompetitif untuk pemetaan opini publik dalam mendukung pengambilan kebijakan transisi energi.

Kata Kunci: Mobil Listrik, Naïve Bayes, Analisis Sentimen, RapidMiner, Text Mining.

A. Introduction

Kemunculan mobil listrik (*electric vehicle/EV*) menjadi salah satu perubahan signifikan dalam industri otomotif global. Perkembangan ini didorong oleh meningkatnya kebutuhan untuk mengurangi ketergantungan terhadap bahan bakar fosil, menekan emisi karbon, serta meningkatnya kesadaran masyarakat terhadap isu keberlanjutan lingkungan. Mobil listrik diharapkan mampu mengurangi penggunaan bahan bakar berbasis minyak bumi secara signifikan sehingga berkontribusi dalam menciptakan sistem transportasi yang lebih ramah lingkungan (Efendi, 2020). Berbeda dengan kendaraan bermesin pembakaran internal, mobil listrik digerakkan oleh motor listrik yang memperoleh energi dari baterai berkapasitas besar dan tidak menghasilkan emisi gas buang secara langsung.

Dampak kehadiran mobil listrik tidak hanya dirasakan dari sisi lingkungan, tetapi juga dari aspek ekonomi dan sosial. Penggunaan mobil listrik berpotensi menekan polusi udara dan emisi CO₂ serta mendukung program pemerintah menuju transisi energi bersih. Pemerintah Indonesia secara aktif mendorong adopsi kendaraan listrik sebagai upaya mengurangi ketergantungan terhadap bahan bakar fosil (Ramadhani & Yuliana, 2023). Di sisi lain, industri mobil listrik membuka peluang ekonomi baru, seperti pengembangan teknologi baterai, pembangunan infrastruktur *charging station*, serta manufaktur komponen elektronik. Namun demikian, adopsi mobil listrik juga menghadapi berbagai tantangan, antara

lain harga awal kendaraan yang relatif tinggi, keterbatasan infrastruktur pengisian daya, serta kekhawatiran masyarakat terhadap umur dan performa baterai.

Seiring meningkatnya perhatian publik terhadap mobil listrik, topik ini menjadi perbincangan yang intens di berbagai platform media sosial dan forum daring. Masyarakat secara bebas menyampaikan opini, baik yang bersifat positif, negatif, maupun netral. Banyaknya opini yang tersebar dalam bentuk teks tidak terstruktur menimbulkan kebutuhan akan metode yang mampu mengolah dan menganalisis data tersebut secara sistematis. Oleh karena itu, analisis sentimen menjadi pendekatan yang relevan untuk mengelompokkan opini publik dan memahami persepsi masyarakat terhadap kehadiran mobil listrik.

Analisis sentimen merupakan bagian dari *text mining*, yaitu proses pengestraksian informasi atau pengetahuan dari data teks tidak terstruktur. Melalui tahapan seperti pembersihan teks, tokenisasi, penghapusan *stopword*, stemming, serta ekstraksi fitur, data teks dapat diubah menjadi representasi yang dapat diolah oleh algoritma pembelajaran mesin. Pendekatan ini banyak digunakan dalam analisis opini konsumen, evaluasi produk, serta pemantauan tren sosial karena mampu mengungkap pola dan kecenderungan persepsi publik (Karimah et al., 2024; Deolika et al., 2019).

Dalam penelitian ini, metode klasifikasi yang digunakan adalah algoritma Naïve Bayes. Naïve Bayes merupakan metode klasifikasi berbasis probabilitas yang menerapkan Teorema Bayes dengan asumsi independensi antar fitur. Meskipun asumsi tersebut bersifat sederhana, Naïve Bayes terbukti efektif dan efisien dalam pengklasifikasian dokumen teks serta memiliki performa yang baik meskipun jumlah data relatif terbatas (Prasetyo et al., 2025; Karimah et al., 2024). Algoritma ini menghitung probabilitas suatu dokumen termasuk ke dalam kelas sentimen tertentu berdasarkan frekuensi kemunculan kata-kata yang terdapat dalam data pelatihan.

Data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, yang menyediakan berbagai dataset terbuka dan banyak dimanfaatkan oleh peneliti di bidang *data science*. Dataset sentimen mobil listrik yang digunakan kemudian diproses menggunakan perangkat lunak RapidMiner. RapidMiner merupakan perangkat lunak *data mining* berbasis antarmuka grafis (*Graphical User Interface*) yang mendukung proses *text mining*, pembelajaran mesin, serta evaluasi model tanpa memerlukan penulisan kode secara langsung. Dengan menggunakan RapidMiner, proses klasifikasi sentimen dapat dilakukan secara terstruktur mulai dari pengolahan data, pembangunan model Naïve Bayes, hingga pengukuran kinerja model dalam memprediksi sentimen masyarakat terhadap mobil listrik.

Melalui penelitian ini, diharapkan dapat diperoleh gambaran yang jelas mengenai persepsi publik terhadap kemunculan mobil listrik berdasarkan analisis sentimen, serta

mengetahui tingkat akurasi algoritma Naïve Bayes dalam mengklasifikasikan opini masyarakat secara efektif.

B. Method

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian deskriptif, yang bertujuan untuk menggambarkan dan menganalisis sentimen masyarakat terhadap kemunculan mobil listrik. Metode yang digunakan dalam penelitian ini adalah klasifikasi teks menggunakan algoritma Naïve Bayes, yang dipilih karena kemampuannya dalam mengklasifikasikan data teks secara efisien dan memiliki performa yang baik pada analisis sentimen.

Data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, berupa dataset berbentuk *datasheet* yang berisi kumpulan komentar atau opini masyarakat mengenai mobil listrik, lengkap dengan informasi seperti nama akun, tanggal unggahan, serta label sentimen. Dataset tersebut diunduh dalam format CSV, kemudian dikonversi ke dalam format Excel agar sesuai dengan kebutuhan input pada perangkat lunak RapidMiner yang digunakan dalam proses pengolahan data.

Tahap awal analisis data diawali dengan seleksi data untuk memastikan kualitas data yang digunakan. Pada tahap ini dilakukan penghapusan baris kosong, perapian dan penggabungan kolom agar sesuai dengan struktur yang dibutuhkan RapidMiner, serta pembuangan data yang tidak relevan. Proses seleksi data ini penting karena keberadaan data yang tidak lengkap atau tidak relevan dapat memengaruhi mutu dan akurasi hasil data mining yang diperoleh (Sudarsono et al., 2021).

Selanjutnya, data yang telah terseleksi diproses melalui tahap *pre-processing* untuk membersihkan dan menormalkan teks. Tahapan ini meliputi penghapusan simbol atau karakter yang tidak memiliki makna, normalisasi huruf menjadi huruf kecil (*lowercase*), penghapusan *stopword*, serta proses tokenisasi untuk memecah teks menjadi unit kata. Selain itu, dilakukan proses stemming apabila diperlukan untuk mengembalikan kata ke bentuk dasarnya, serta *replace value* guna memperbaiki inkonsistensi pada label sentimen. Hasil dari tahap *pre-processing* ini menghasilkan *data sheet* final yang siap digunakan dalam proses pemodelan.

Data yang telah melalui proses *pre-processing* kemudian dianalisis menggunakan perangkat lunak RapidMiner dengan menerapkan algoritma Naïve Bayes untuk membangun model klasifikasi sentimen. Model yang dihasilkan selanjutnya dievaluasi untuk mengetahui kinerja dan tingkat akurasi dalam mengklasifikasikan sentimen masyarakat terhadap mobil listrik. Seluruh tahapan metode ini dirancang secara sistematis agar hasil penelitian dapat memberikan gambaran yang objektif dan akurat sesuai dengan tujuan penelitian.

C. Results and Discussion

Results

Hasil Klasifikasi Sentimen

Model Naïve Bayes dibangun menggunakan dataset yang telah melalui tahap *pre-processing* dan dievaluasi menggunakan data pengujian pada perangkat lunak RapidMiner. Proses evaluasi menghasilkan nilai akurasi sebesar 100%, yang menunjukkan bahwa seluruh data uji berhasil diklasifikasikan sesuai dengan label sentimen aslinya. Evaluasi performa model ditunjukkan melalui *confusion matrix* pada Tabel 1.

Tabel 1. Confusion Matrix Hasil Klasifikasi Naïve Bayes

Kelas Aktual \ Prediksi	Positif	Negatif	Netral
Positif	504	0	0
Negatif	0	869	0
Netral	0	0	144

Berdasarkan *confusion matrix* tersebut, seluruh data pada masing-masing kelas sentimen diprediksi dengan benar oleh model. Tidak ditemukan kesalahan klasifikasi berupa *false positive* maupun *false negative*.

Metrik Evaluasi Model

Selain akurasi, performa model juga diukur menggunakan metrik *precision*, *recall*, dan *F1-score*. Hasil evaluasi menunjukkan bahwa seluruh metrik untuk setiap kelas sentimen memiliki nilai sempurna, sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Hasil Evaluasi Kinerja Model

Kelas Sentimen	Precision	Recall	F1-Score
Positif	1.00	1.00	1.00
Negatif	1.00	1.00	1.00
Netral	1.00	1.00	1.00

Hasil ini menunjukkan bahwa model Naïve Bayes mampu mengklasifikasikan seluruh data uji secara konsisten pada setiap kelas sentimen.

Discussion

Capaian akurasi sempurna sebesar 100% dalam klasifikasi sentimen terhadap mobil listrik menggunakan algoritma Naïve Bayes menjadi poin sentral dalam analisis ini. Fenomena akurasi mutlak pada data tekstual sering dianggap sebagai indikasi adanya struktur data yang sangat terpisah secara statistik. Keberhasilan ini berakar pada efektivitas tahap *pre-processing* yang telah dilakukan, di mana proses *cleaning*, *stopword removal*, dan *normalization* berhasil mengisolasi sinyal sentimen dari gangguan noise tekstual³³³³. Data

yang telah melalui kurasi ketat menghasilkan fitur-fitur dengan bobot probabilitas yang kontras antar setiap kelas sentimen.

Algoritma Naïve Bayes bekerja dengan menghitung probabilitas posterior dari setiap kelas berdasarkan frekuensi kemunculan kata sesuai dengan Teorema Bayes (Prasetyo et al., 2025). Analisis terhadap dataset menunjukkan adanya distribusi fitur yang bersifat *linearly separable* atau terpisah secara linier. Kata-kata kunci seperti "hemat energi", "ramah lingkungan", dan "subsidi" terkonsentrasi secara eksklusif pada kelas positif. Kelas negatif didominasi oleh leksikon spesifik seperti "mahal", "lama ngecas", dan "kurang praktis". Pola kata yang tidak saling tumpang tindih ini membuat asumsi *conditional independence* dalam Naïve Bayes menjadi kekuatan utama dalam mengenali pola statistik yang eksplisit.

Risiko *overfitting* menjadi perhatian utama ketika model mencapai tingkat akurasi sempurna. Akurasi 100% mengindikasikan model telah mencapai titik jenuh dalam mempelajari karakteristik spesifik dataset dari Kaggle yang cenderung homogen dan minim noise. Kondisi ini mencerminkan keberhasilan model dalam mengenali pola statistik dari data latih, namun terdapat potensi keterbatasan saat model dihadapkan pada data baru yang lebih kompleks di luar dataset ini. Pengujian pada penelitian ini menggunakan data uji yang memiliki kemiripan pola linguistik dengan data latih, sehingga tidak ditemukan kesalahan klasifikasi pada *confusion matrix*.

Implementasi teknik *k-Fold Cross-Validation* merupakan langkah strategis untuk memperkuat argumen validasi model dan memitigasi bias pemilihan data (Karimah et al., 2024). Teknik ini bekerja dengan membagi dataset menjadi "k" bagian yang sama besar, di mana setiap bagian akan berperan sebagai data uji secara bergantian sementara bagian lainnya menjadi data latih¹⁵. Penggunaan *Cross-Validation* menjamin bahwa setiap sampel data pernah digunakan untuk proses pelatihan dan pengujian. Stabilitas nilai akurasi yang konsisten di seluruh *fold* akan membuktikan bahwa performa tinggi Naïve Bayes dalam penelitian ini bukan merupakan kebetulan statistik, melainkan hasil dari kekuatan prediktif algoritma terhadap fitur yang diekstraksi.

Perspektif sosiologis dan ekonomi menunjukkan bahwa dominasi sentimen negatif (869 data) mencerminkan skeptisisme nyata terhadap infrastruktur penunjang mobil listrik. Meskipun isu lingkungan menjadi daya tarik utama, kekhawatiran publik mengenai harga kendaraan dan keterbatasan *charging station* tetap menjadi hambatan adopsi yang signifikan (Ramadhani & Yuliana, 2023). Kualitas data yang bersih dan pola kata yang jelas di setiap kelas memberikan kontribusi besar pada akurasi tinggi ini.

Penelitian ini menegaskan bahwa kualitas *pre-processing* dan kebersihan data merupakan determinan utama dalam mencapai performa model yang optima. Pemanfaatan RapidMiner sebagai alat bantu visual memudahkan identifikasi pola sentimen masyarakat secara terstruktur tanpa memerlukan penulisan kode yang rumit. Hasil ini memberikan

kontribusi praktis bagi pembuat kebijakan untuk memetakan opini publik mengenai transisi energi secara akurat melalui pendekatan penambangan teks.

D. Conclusion

Penelitian mengenai klasifikasi sentimen terhadap kemunculan mobil listrik menggunakan metode Naïve Bayes menunjukkan bahwa algoritma ini memiliki efektivitas yang sangat tinggi dalam memetakan opini publik. Capaian akurasi sebesar 100% pada pengujian melalui RapidMiner membuktikan bahwa model mampu mengklasifikasikan seluruh data uji ke dalam label positif, negatif, dan netral secara sempurna. Keberhasilan ini didorong oleh kualitas data yang bersih melalui tahapan *pre-processing* yang ketat serta karakteristik dataset yang memiliki pemisahan pola kata secara alami antar kelas sentimen. Penggunaan RapidMiner sebagai instrumen penelitian memberikan kontribusi signifikan dalam menyederhanakan proses klasifikasi data mining melalui alur kerja visual yang terstruktur bagi peneliti.

Secara keseluruhan, hasil analisis menggambarkan bahwa meskipun terdapat antusiasme terhadap isu lingkungan, skeptisisme publik yang tecermin dalam dominasi sentimen negatif terkait harga dan infrastruktur tetap menjadi temuan krusial. Naïve Bayes terbukti menjadi teknik yang handal untuk mengekstraksi pola statistik dari opini masyarakat yang eksplisit di media sosial. Temuan ini memberikan dasar informasi yang berharga bagi pengambil kebijakan dalam memahami hambatan sosiologis dan ekonomi masyarakat terhadap transisi energi. Kualitas pengolahan data awal menjadi faktor penentu utama, di mana dataset yang terstruktur dengan baik memungkinkan algoritma sederhana untuk mencapai performa maksimal dalam analisis sentimen berbasis teks.

Bibliography

- Deolika, A., Kusriani, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*, 3(2), 179. <https://doi.org/10.36294/jurti.v3i2.1077>
- Efendi, A. (2020). Rancang Bangun Mobil Listrik Sula Politeknik Negeri Subang. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 17(1), 75. <https://doi.org/10.23887/jptk-undiksha.v17i1.23057>
- Karimah, A., Dwilestari, G., & Mulyawan, M. (2024). Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 767–737. <https://doi.org/10.36040/jati.v8i1.8373>
- Metode, M., Multiple, F., Decision, C., Fmcdm, M., & Yogyakarta, D. (2020). Indonesian Journal of Business Intelligence. *Indonesian Journal of Business Intelligence*, 3(2), 54–60.
- Muharrom, M. (2023). Analisis Penggunaan Orange Data Mining untuk Prediksi Harga USDT/BIDR Binance. *Bulletin of Information Technology (BIT)*, 4(2), 178–184. <https://doi.org/10.47065/bit.v4i2.654>
- Prasetyo, T., Waskita, A. A., & Taryo, T. (2025). Analisis Sentimen Pengguna Seputar Kendaraan Listrik di Twitter dengan Penerapan Algoritma Naïve Bayes, KNN, dan Decision Tree untuk Klasifikasi. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, 8(2), 108–117.
- Rahayu, M. I. (2023). Klasifikasi ras kucing menggunakan metadata dataset kaggle dengan framework yolo v5. *Jurnal Teknologi Informasi dan Komunikasi*, 12(1).
- Ramadhani, S., & Yuliana, L. (2023). Pengaruh Persepsi Konsumen Terhadap Minat Beli Mobil Listrik Mercedes-Benz EQS. *Jurnal Orientasi Bisnis dan Entrepreneurship (JOBS)*, 4(1), 35–44. <https://doi.org/10.33476/jobs.v4i1.3614>
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *Journal of Business and Audit Information Systems*, 4(1), 13–21. <https://journal.ubm.ac.id/index.php/jbase/article/view/2729/2026>