

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Pada bagian ini dijelaskan studi sebelumnya yang berasal dari hasil studi sebelumnya dan digunakan sebagai referensi dalam menyusun penelitian ini.

Tabel 2.1 Penelitian terdahulu

Penelitian terdahulu 1	
Judul	Analisis Sentimen Tentang Mobil Listrik Dengan Metode Support Vector Machine Dan Feature Selection Particle Swarm Optimization
Nama Penulis	Ahmad Nugroho, Agung Sunge, Aswan S
Jurnal	Journal of partical computer science [22]
Tahun	2022
Permasalahan	Banyaknya tanggapan dari masyarakat Indonesia dalam mengevaluasi mobil listrik menjadi indikator penting bagi peneliti dalam melakukan analisis sentimen masyarakat terhadap mobil listrik di platform media sosial Twitter.
Algoritma	Support Vector Machine Dan Feature Selection Particle Swarm Optimization
Temuan	Temuan yang diperoleh dengan menerapkan metode Support Vector Machine tanpa seleksi fitur Particle Swarm Optimization, dan pengujian dengan metode Support Vector Machine yang melibatkan seleksi fitur Particle menunjukkan peningkatan nilai AUC sebesar 2,13% dari 0,844 menjadi 0,862.
Pembahasan	Hasil penelitian menunjukkan bahwa analisis dalam sentimen masyarakat terhadap mobil listrik di media sosial Twitter, dari total 452 data tweet yang diuji, pada pengujian pertama dengan menggunakan metode Support Vector Machine, terdapat 436 data tweet yang menunjukkan sentimen positif dan 16 data tweet yang menunjukkan sentimen negatif . Sedangkan pada pengujian kedua dengan menggunakan metode Support Vector Machine dan seleksi fitur Particle Swarm Optimization, terdapat 426 data tweet yang menunjukkan sentimen positif dan 26 data tweet yang menunjukkan sentimen negatif.
Relevansi	Meskipun hasil analisis tersebut tidak menggunakan algoritma naive Bayes, namun dapat memberikan beberapa relevansi dengan topik analisis sentimen masyarakat Indonesia terhadap mobil listrik yang menggunakan algoritma naive bayes. penggunaan fitur <i>selection particle swarm optimization</i> dalam pengujian kedua menunjukkan upaya untuk meningkatkan akurasi klasifikasi sentimen. Meskipun menggunakan algoritma yang berbeda, konsep peningkatan metode analisis sentimen ini dapat diadopsi atau diterapkan dalam penggunaan algoritma naive bayes untuk mendapatkan hasil yang lebih baik
Penelitian Terdahulu 2	
Judul	<i>Exploring the Sentiment Analysis of Electric Vehicles Social Media Data by Using Feature Selection Methods</i>
Nama Penulis	Costello, Francis Joseph Lee, Kun Chang
Jurnal	Journal of Digital Convergence [23]

Tahun	2020
Permasalahan	Saat melakukan klasifikasi kami menemukan masalah dimensi dan juga mengeksplorasi penggunaan model pemilihan fitur (FS) untuk mereduksi data dimensi set. Kami menemukan bahwa penggunaan tiga model FS menampilkan hasil yang paling menjanjikan bila digunakan bersama mesin logistik dan dukungan vector algoritma klasifikasi
Algoritma	<i>Feature Selection, Text Mining, dan K-Fold Cross Validation</i>
Temuan	Metode dan klasifikasi <i>Feature selection</i> yang dilakukan menghasilkan pengurangan dimensi dataset CFS dan PSO FS menunjukkan pengurangan dataset paling besar dengan CFS memilih kurang dari 10% fitur aslinya. IG dan CHI memilih jumlah fitur yang sama dengan 840 fitur terpilih dan RFF memilih total 825 fitur. Kumpulan data asli terdiri dari 1508 atribut dan seperti yang terlihat, telah dikurangi secara signifikan oleh semua teknik feature selection.
Pembahasan	Hasil studi telah menunjukkan bahwa sentimen terhadap kendaraan listrik saat ini lebih positif dibandingkan negatif. Kami telah menemukan bahwa kendaraan listrik sedang menjadi tren teknologi yang mendapatkan tanggapan positif dari masyarakat berdasarkan hasil yang diperoleh dari analisis sentimen.
Relevansi	Penelitian ini relevan karena menggunakan text mining memiliki kemampuannya dalam mengidentifikasi reaksi dalam kalimat terutama telah bekerja dengan baik dengan teks media sosial.
Penelitian Terdahulu 3	
Judul	Analisis Sentimen Pengguna Twitter di Indonesia terhadap Kenaikan Harga BBM Menggunakan Algoritma Naïve Bayes
Nama Penulis	Alfa Nugraha, Joel
Jurnal	UMN [14]
Tahun	2023
Permasalahan	Peningkatan harga bahan bakar terjadi pada bulan Februari 2023. Kenaikan harga bahan bakar ini menyulut beragam tanggapan dan pandangan dari masyarakat di Indonesia, mulai dari sikap negatif dan penolakan, hingga respon positif dan dukungan. Tanggapan-tanggapan ini banyak diungkapkan melalui berbagai platform media sosial, termasuk Twitter.
Algoritma	Naïve bayes
Temuan	Model Multinomial Naïve Bayes mendapatkan precision data negatif yang 24% lebih tinggi dari Complement Naïve Bayes, namun dengan recall dan f1-score yang 45% dan 32% lebih rendah.
Pembahasan	Berdasarkan hasil studi ini Model Naive Bayes terbaik dan stabil antara precision, recall, dan f-1 score adalah model Complement Naive Bayes yang mendapatkan akurasi sebesar 82%. Model memiliki precision tertinggi sebesar 59% untuk data negatif dan 89% untuk data positif.
Relevansi	Penelitian ini relevan karena menerapkan analisis sentiment walaupun dalam konteks yang berbeda tetapi dapat menjadi referensi bagi penelitian, dalam hal membandingkan keefektivitas Model multinominal naïve bayes dan Model complement naïve bayes.
Penelitian Terdahulu 4	
Judul	Analisis Sentimen terhadap Kereta Cepat Jakarta-Bandung dan Kereta Panoramic pada Media Sosial Twitter menggunakan Algoritma Naive Bayes
Nama Penulis	Haryanto Ruswandi, Jason
Jurnal	UMN [24]
Tahun	2023
Permasalahan	Munculnya kereta cepat Jakarta-Bandung dan juga kereta panoramic ini mengakibatkan banyaknya pro dan kontra pada masyarakat. Waktu yang singkat dan kapasitas tambahan (opsi lain kereta api Argo Parahyangan) pada

	kereta cepat Jakarta-Bandung serta pemandangan yang indah pada kereta panoramic merupakan alasan masyarakat mendukung adanya dua kereta tersebut.
Algoritma	Naïve bayes
Temuan	Penelitian analisis sentimen mengenai kereta cepat Jakarta-Bandung dan kereta panoramic dengan menggunakan algoritma Naive Bayes menghasilkan nilai akurasi terbaik sebesar 77.65%.
Pembahasan	Hasil analisis ini menunjukkan respon positif Masyarakat mengenai kereta cepat Jakarta – Bandung, dengan Teknik algoritma Naïve Bayes menghasilkan nilai akurasi terbaik .
Relevansi	Penggunaan teknik Naïve bayes untuk analisis sentimen mengenai kereta cepat Jakarta-Bandung dan kereta panoramic, relevan untuk penelitian dalam analisis sentimen mobil listrik terhadap masyarakat menunjukkan teknik yang dapat meningkatkan akurasi analisis.
Penelitian Terdahulu 5	
Judul	<i>Multinomial Naive Bayes Algorithm for Classifying Indonesian Sentiments Regarding Jakarta International Stadium (JIS)</i>
Nama Penulis	Rizki Surya Pratama, Daffa Munandar, Fadhilla Ramdhanisa, Khairunnisa.
Jurnal	International Journal of Information Technology and Computer Science Applications [25]
Tahun	2024
Permasalahan	Analisis sentimen publik kehadiran stadion jakarta international terhadap standar internasional menggunakan data ulasan Google Maps dengan tantangan mengenali sentimen secara akurat. jika pandangan ini dianalisis dengan tepat, maka dapat memberikan informasi evaluatif untuk pengambilan keputusan oleh pengelola JIS.
Algoritma	Naïve bayes, Randomforest
Temuan	Temuan dari pemodelan klasifikasi algoritma multinomial Naive Bayes mengungguli secara akurat memprediksi 733 titik data uji, terdiri dari 292 nilai positif dan 317 nilai negatif. Nilai presisi kelas positif dan negatif sebesar 83%.
Pembahasan	Algoritma multinomial Naive Bayes lebih unggul dari pendekatan Random Forest terutama pada bagian Tingkat akurasi yang lebih tinggi.
Relevansi	Penelitian relevan ini menunjukkan efisiensi penggunaan Algoritma Naïve bayes dan Randomforest dalam analisis sentiment, yang bisa diterapkan dalam menganalisis sentiment Masyarakat terhadap Mobil Listrik.
Penelitian Terdahulu 6	
Judul	ANALISIS SENTIMEN TERHADAP LAYANAN APLIKASI GRAB INDONESIA MENGGUNAKAN METODE NAÏVE BAYES
Nama Penulis	Rifa'I, Ahmad Ardhani, Risma Pratama, Denni Fatihanursari, Fatihanursari
Jurnal	JATI (Jurnal Mahasiswa Teknik Informatika) [26]
Tahun	2024
Permasalahan	Transportasi umum memiliki sejumlah permasalahan, antara lain fasilitas yang tidak memadai, pelayanan yang buruk, dan perencanaan kegiatan yang tidak efektif.
Algoritma	Naïve bayes
Temuan	Hasil confusion matrix menunjukkan terdapat 1.677 data yang terklasifikasi benar, 6 data terklasifikasi benar tetapi terklasifikasi (prediksi negatif), 69 data terklasifikasi salah (data positif yang seharusnya masuk kategori negatif), dan 248 data salah. klasifikasi yang salah bersifat negatif tetapi masuk dalam kategori positif. Akurasi, presisi, dan recall perhitungan matriks konfusi manual masing-masing sebesar 84,36%, 99,64%, dan 87,12%.

Pembahasan	Hasil studi tersebut menemukan bahwa pengguna aplikasi Grab memberikan ulasan positif. Dari jumlah tersebut, total hasil prediksi klasifikasi sentimen positif sebanyak 1677 atau 83,85%, dan sentimen negatif sebanyak 69 atau 3,45% dari data teks. Hal ini menunjukkan bahwa meskipun banyak pengguna yang memberikan komentar positif pada aplikasi Grab, tidak sedikit pula pengguna yang memberikan komentar negatif.
Relevansi	Meskipun analisis sentimen terfokus pada layanan aplikasi Grab, temuan ini relevan dengan topik analisis sentimen masyarakat Indonesia terhadap mobil listrik dengan algoritma Naïve Bayes. Studi ini mengungkapkan bahwa sebagian besar ulasan dari pengguna aplikasi Grab adalah positif, meskipun terdapat juga sejumlah kecil ulasan negatif. Relevansinya terletak pada penggunaan metode Naïve Bayes untuk mengklasifikasikan sentiment.
Penelitian Terdahulu 7	
Judul	Analisis Sentimen Pengguna Transportasi <i>Online</i> Maxim Pada Instagram Menggunakan Naïve Bayes Classifier dan K-Nearest Neighbor
Nama Penulis	Warrailhan, Dzul Asfi Permana, Inggih Mustakim, Mustakim Novita, Rice
Jurnal	Jurnal Media Informatika [27]
Tahun	2023
Permasalahan	Maxim membutuhkan opini pengguna terkait aplikasi dan pelayanannya. Saat ini, Maxim mendapatkan opini tersebut melalui email, website, atau bertemu langsung dengan pengguna.
Algoritma	Naïve bayes dan K-Nearest Neighbor
Temuan	Pada algoritma KNN menggunakan 5 buah parameter K yaitu K = 3, K = 5, K = 7, K = 9, dan K = 11. Perbandingan performa algoritma dilihat dari nilai accuracy tertinggi. Hasil accuracy pada data sentimen terkait aplikasi menggunakan algoritma NBC yaitu 81,03% dan pada algoritma KNN dengan nilai k=3 yaitu sebesar 80,72%. Sedangkan pada data sentimen terkait layanan menghasilkan nilai accuracy pada algoritma NBC yaitu 94% dan Algoritma KNN dengan k=3 yaitu 84%.
Pembahasan	memberikan kontribusi yang berarti dalam pemahaman tentang pandangan pengguna terhadap layanan transportasi online dan memberikan dasar yang kuat untuk penggunaan NBC dalam analisis sentimen di platform media sosial seperti Instagram, dengan potensi implikasi yang besar untuk perencanaan strategi dan pengambilan keputusan di industri transportasi online.
Relevansi	Penelitian relevan ini menunjukkan meskipun dilakukan pada platform Instagram dengan fokus pada layanan transportasi online Maxim, hasil yang menunjukkan keunggulan Naïve Bayes Classifier (NBC) dalam mengatasi masalah ketidakseimbangan kelas pada data sentimen memberikan wawasan yang berharga bagi penelitian mengenai mobil listrik. Dalam konteks analisis sentimen terhadap mobil listrik di Indonesia, penggunaan algoritma NBC dapat menjadi alat yang efektif untuk memahami pandangan dan persepsi masyarakat terhadap mobil listrik.
Penelitian Terdahulu 8	
Judul	ANALISIS SENTIMEN ULASAN PENUMPANG MASKAPAI PENERBANGAN DI INDONESIA DENGAN ALGORITMA RANDOM FOREST DAN KNN
Nama Penulis	Ramadhansyah, Donny Asrofiq, Ahmad Yunefri, Yogi Kuning
Jurnal	Jurnal Sistem Informasi [28]
Tahun	2024
Permasalahan	Analisis sentimen ini dilakukan berdasarkan ulasan maskapai dan ulasan penumpang penerbangan yang nantinya dapat membantu penumpang memilih maskapai yang sesuai dengan kebutuhan mereka.
Algoritma	Random Forest dan K-Nearest Neighbor

Temuan	Hasil evaluasi menunjukkan bahwa baik Random Forest (akurasi 83%) maupun KNN (akurasi 82%) memberikan performa yang baik dengan nilai presisi, recall, dan F1-Score yang memuaskan. Komparasi antara keduanya menunjukkan hasil yang kompetitif. Secara keseluruhan, model ini efektif dalam mengklasifikasikan sentimen ulasan maskapai penerbangan, dengan catatan penting untuk memperhatikan kualitas data dan kemungkinan fine-tuning pada model.
Pembahasan	Dalam penelitian ini, dua metode yaitu Random Forest dan K-Nearest Neighbors (KNN) digunakan untuk menganalisis sentimen data dari Kaggle yang diambil dari www.airlinequality.com . Hasil menunjukkan bahwa Random Forest sedikit lebih akurat daripada KNN, karena KNN mencari nilai K data terdekat dalam ruang fitur untuk menentukan label data.
Relevansi	Meskipun berfokus pada analisis sentimen ulasan maskapai penerbangan, penelitian dengan algoritma Random Forest dan KNN memberikan pengetahuan tentang penggunaan algoritma dalam memahami pandangan masyarakat, relevan untuk analisis sentimen terhadap mobil listrik.

Penelitian Terdahulu 9

Judul	Analisis Sentimen Pengguna Twitter Terhadap Bus Listrik Menggunakan Naïve Bayes
Nama Penulis	Ike Verawati, Syarif Nurwahid Jaelani
Jurnal	Jurnal Media Informatika [21]
Tahun	2024
Permasalahan	Masyarakat berpendapat bahwa infrastruktur pendukung operasional bus listrik masih kurang dan adanya kekhawatiran akan kehabisan daya baterai saat kondisi macet.
Algoritma	Naïve Bayes
Temuan	Klasifikasi dilakukan sebanyak 4 kali dengan rasio pembagian data latih dan data uji yang berbeda. Rasio pembagian data yang dipilih adalah 9:1, 8:2, 7:3, dan 6:4, rasio 9:1 berarti 90% dataset dijadikan data latih dan 10% sisanya menjadi data uji, selanjutnya data latih dan data uji diimplementasikan kedalam algoritma naïve bayes.
Pembahasan	Berdasarkan hasil penelitian yang telah dilakukan pada model dari 4 kali percobaan pada dataset tweet sebanyak 2412 diperoleh kelas sentimen terbanyak adalah sentimen positif yaitu sebanyak 77,32% sentimen positif dan 22,69% sentimen negatif, yang diperoleh pada rasio pembagian data 6:4. Sentimen positif juga mendominasi pada rasio pembagian data lainnya, hal ini menunjukkan mayoritas pengguna twitter mendukung baik terhadap bus listrik
Relevansi	Meskipun difokuskan pada analisis sentimen pengguna Twitter terhadap bus listrik, hasil penelitian tersebut memiliki relevansi dengan topik analisis sentimen masyarakat Indonesia terhadap mobil listrik dengan algoritma Naïve Bayes. Dari hasil empat kali percobaan dengan dataset tweet sebanyak 2412, mayoritas sentimen yang ditemukan adalah positif, mencapai 77,32%, sementara sentimen negatif hanya sebesar 22,69%. Temuan ini menunjukkan mayoritas pengguna Twitter mendukung bus listrik, memberikan pemahaman tentang pandangan masyarakat terhadap teknologi ramah lingkungan seperti mobil listrik. Meskipun berfokus pada bus listrik, hasil ini memberikan wawasan yang dapat diterapkan dalam memahami sentimen masyarakat terhadap kendaraan listrik secara lebih luas.

Penelitian Terdahulu 10

Judul	Penerapan Naive Bayes Terhadap Sentimen Analisis Media Sosial Twitter Pengguna Kereta Cepat Jakarta-Bandung (Whoosh)
Nama Penulis	Agustiranti, Tifani, Izzati Kurdiana, Aulia Khalfani, Al Ghiffari, Bilal Dwi Juniar, Elza Gita Purnama, Diki
Jurnal	Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI) [29]

Tahun	2024
Permasalahan	Analisis ini dilakukan untuk memahami persepsi masyarakat terhadap layanan transportasi yang krusial ini serta memberikan dasar bagi perbaikan dan pengembangan layanan yang lebih baik di masa
Algoritma	Naïve Bayes
Temuan	Hasil sentimen yang digunakan dalam perhitungan algoritma Naive Bayes Classifier menggunakan Python. Akurasi yang dicapai sebesar 88%, artinya hasil yang dicapai sangat baik.
Pembahasan	Hasil dari tanggapan masyarakat mengenai penggunaan Kereta Cepat Jakarta-Bandung positif dan bekerja dengan baik dengan menghasilkan nilai akurasi cukup tinggi yaitu 88%, presisi 82%, dan recall 88%. Dengan melalui proses normalisasi kata yang sudah diubah sesuai dengan pengujian kamus kata baku. Meskipun begitu, masih banyak kata data yang tidak terklasifikasi dengan baik atau dapat dikatakan sebagai data yang salah klasifikasi. Terdapat total 396 data yang belum terklasifikasi secara lengkap pada data berlabel netral yang dianggap sulit dideteksi.
Relevansi	Analisis sentimen pengguna kereta cepat Jakarta-Bandung, hasil penelitian ini memiliki relevansi dengan topik analisis sentimen masyarakat Indonesia terhadap mobil listrik dengan algoritma Naïve Bayes. Tanggapan positif masyarakat terhadap penggunaan Kereta Cepat karena data berlabel netral yang sulit dideteksi. Hal ini menyoroti tantangan dalam proses klasifikasi sentimen, yang dapat memberikan wawasan dalam analisis sentimen masyarakat terhadap mobil listrik.

Dari penelitian sebelumnya, penulis menyusun kerangka penelitian untuk menganalisis sentimen masyarakat Indonesia terhadap mobil listrik. Melalui berbagai penelitian yang menggunakan berbagai metode analisis sentimen, kita dapat memahami pandangan dan tanggapan masyarakat terhadap berbagai topik, mulai dari layanan keuangan hingga proyek infrastruktur dan kebijakan publik. Analisis sentimen dapat memberikan wawasan yang berharga bagi peneliti untuk membuat keputusan yang lebih baik, merancang strategi komunikasi yang efektif, dan meningkatkan pengalaman pengguna. Dengan menggunakan teknik-teknik seperti Naive Bayes, serta penggunaan seleksi fitur dan optimisasi, kita dapat menghasilkan hasil analisis sentimen yang lebih akurat dan informatif, dalam penelitian sentimen masyarakat Indonesia terhadap mobil listrik.

Pembeda dari penelitian sebelumnya adalah pendekatan yang digunakan dalam penelitian ini, yaitu fokus pada analisis sentimen khusus terhadap mobil listrik di Indonesia dengan metode naïve bayes. Studi ini memberikan kontribusi berharga dengan mengkhususkan analisis pada topik yang relevan dengan tren dan isu terkini di masyarakat, yaitu adopsi dan persepsi terhadap mobil listrik.

Penggunaan teknik-teknik khusus seperti naïve bayes dan optimisasi seleksi fitur juga menunjukkan pendekatan metodologis yang lebih dan terperinci dalam menghasilkan hasil analisis sentimen yang lebih diperinci dan terstruktur. Dengan demikian, penelitian ini memberikan pemahaman yang lebih mendalam tentang sikap dan respons masyarakat Indonesia terhadap inovasi mobil listrik, membedakannya dari penelitian sentimen umum sebelumnya yang mungkin tidak sefokus atau sespesifik dalam konteks topik yang sama.

2.2 Teori yang digunakan

2.2.1 Mobil Listrik

Pada awalnya, penggunaan mobil listrik telah diperkenalkan sejak tahun 1828, dengan produksi pertama dilakukan pada tahun 1884. Pada periode antara 1897 hingga 1900, mobil listrik mencapai pangsa pasar sebesar 28% dari total kendaraan yang tersedia di pasaran [30]. Mobil listrik merupakan kendaraan transportasi yang mengandalkan sumber energi dari listrik bukan dari bahan bakar fosil atau minyak bumi. Mobil ini beroperasi dengan menggunakan tenaga listrik yang disimpan dalam baterai atau media penyimpanan energi lainnya. Industri mobil telah mengalami perkembangan yang signifikan dalam pengembangan mobil listrik, dengan berbagai produsen berlomba-lomba untuk menarik minat konsumen dan bersaing di pasaran.

Kemajuan pada komponen dan perangkat elektronik dalam mobil listrik sangat mempengaruhi harga dan kualitasnya. Semakin tinggi kualitas perangkat elektronik, semakin mahal pula mobil tersebut. Mobil listrik dilengkapi dengan perangkat elektronik pendukung yang memiliki manfaat besar, seperti lampu utama yang membantu meningkatkan keselamatan saat berkendara di malam hari. Penempatan perangkat elektronik dalam mobil listrik harus dilakukan dengan cermat dan efisien untuk memastikan kinerja optimal [31].

Mobil listrik dipercaya dapat berkontribusi untuk mengurangi emisi CO₂ dari perjalanan, terutama ketika produksi dan penggunaannya

ditenagai dengan energi terbarukan. Mobil ramah lingkungan sedang berkembang saat ini, dengan mobil listrik sebagai salah satu solusi yang dipilih, terutama di Indonesia, yang menduduki peringkat keempat dalam hal jumlah penduduk di dunia. Dengan berbagai isu tentang mobil listrik mencakup masalah ketersediaan infrastruktur pengisian daya, keterbatasan jarak tempuh, dan biaya awal yang tinggi. Selain itu, keberlanjutan energi dan dampak lingkungan dari pembuatan baterai juga menjadi perhatian. Indonesia sedang berusaha untuk mengadopsi teknologi mobil listrik sebagai bagian dari transportasi umum. Keputusan untuk membawa mobil listrik ke Indonesia diharapkan dapat membawa dampak positif dalam mengurangi ketergantungan pada bahan bakar minyak dan memperlambat perubahan iklim [32].

2.2.2 Analisis Sentimen

Analisis sentimen merupakan sebuah metode *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi sentimen atau opini yang tersirat dalam sebuah teks, terutama dalam konteks penelitian ini yang difokuskan pada komentar di platform media sosial twitter. Analisis sentimen memiliki kontribusi besar dalam memahami ekspresi subjektif manusia di media sosial, karena dengan memahami hal tersebut, kita dapat mengetahui emosi, perasaan, pendapat, dan sikap pribadi dari pengguna media sosial tersebut [33].

Hasil analisis sentimen memiliki tujuan untuk mengevaluasi emosi, sikap, pendapat, dan penilaian yang diekspresikan oleh pembicara atau penulis terkait dengan produk. Karena hal ini, dalam beberapa penelitian, terutama pada ulasan produk, langkah awal sering kali melibatkan identifikasi unsur-unsur produk yang akan dianalisis sebelum memulai proses ekstraksi opini [34]. Dengan analisis sentimen, kita dapat menentukan apakah suatu kalimat memiliki polaritas positif atau negatif berdasarkan isinya [35]. Contoh kalimat yang mengandung kalimat positif adalah “Mobil listrik memiliki keunggulan karena ramah lingkungan”, lalu

untuk kalimat negatifnya “Mobil listrik membutuhkan waktu pengisian daya yang sangat lama”.

2.2.3 Twitter

Twitter merupakan *platform* jejaring sosial yang populer yang memungkinkan pengguna untuk me-retweet apapun dalam batasan 140 karakter. Sekitar 300 juta orang telah mendaftar untuk menggunakan Twitter, dengan lebih dari 500 juta pengikut baru yang mendaftar setiap harinya. Twitter telah menjadi salah satu sumber data yang paling diminati oleh pengguna karena kemudahannya. Fokus utamanya adalah pada pengalaman atau pandangan pengguna terhadap peristiwa terkini, termasuk foto, video, tautan, dan konten lain yang dapat dengan mudah dibagikan ke akun pengguna [36]. Twitter menjadi alat menarik bagi berbagai kalangan untuk mengikuti minat pengguna terhadap berbagai situasi secara langsung. Hal ini menjadikan Twitter sebagai sumber data potensial yang dapat dimanfaatkan oleh jutaan orang, dengan menggunakan teknik *Crawling*, *streaming mining*. Dengan prinsip ini, informasi dapat diperoleh mengenai opini publik secara umum.

Beberapa dari fitur yang dapat ditemukan di Twitter meliputi [36]:

1. Halaman Utama

Pada tampilan awal, kita dapat melihat Tweets yang diposting oleh orang-orang yang menjadi teman atau diikuti (*following*).

2. Profil

Halaman yang memuat informasi pribadi dari pengguna Twitter.

3. *Followers*

Pada fitur ini pengguna lain yang ingin menjadi teman. Jika pengguna lain memilih untuk mengikuti akun seseorang, maka Tweets dari akun yang diikuti tersebut akan muncul di halaman utama.

4. *Following*

Pada fitur ini pengguna mengikuti akun seseorang agar Tweets yang diposting oleh akun tersebut akan muncul di halaman utama.

5. Mentions

Konten ini merujuk pada balasan dalam percakapan, memungkinkan pengguna untuk langsung menandai orang yang diajak bicara.

Twitter menyediakan Antarmuka Pemrograman Aplikasi (API) untuk memfasilitasi pengambilan data secara *Crawling*. API mempermudah pengguna untuk mengakses data Tweet secara *real-time*.

2.2.4 Web Scraping

Untuk melakukan *web scraping* pada mesin pencari Google, berbagai teknologi dapat digunakan, Dengan menggunakan bahasa pemrograman Python bersama dengan modul BeautifulSoup dan Selenium, sebuah program akan menjalankan pencarian pada halaman Google dan mengekstrak teks dari hasil pencarian. Setelah itu, teks tersebut akan diolah menjadi struktur data yang lebih mudah untuk dikelola dan disusun [37].

Pengambilan data dilakukan menggunakan *tools* google colab dengan memanfaatkan library google-play-scraper. google colab adalah *platform cloud-based* yang disediakan oleh google untuk melakukan tugas Python, di lingkungan yang mirip dengan Jupyter Notebook. Google colab dapat digunakan untuk melakukan *web scraping* dengan menggunakan Selenium, sebuah alat yang digunakan untuk pengujian, tugas otomatisasi *web*, dan tugas *web scraping* [38].

2.2.6 Precision, Recall, dan Accuracy

Precision dan *recall* adalah kinerja yang digunakan untuk mengevaluasi performa analisis. *Precision* mencerminkan sejauh mana data yang diambil sesuai dengan informasi yang diperlukan, sementara *recall* menunjukkan seberapa baik analisis dalam menemukan kembali informasi yang diminta. Sedangkan untuk *accuracy* menggambarkan

sejauh mana nilai yang diperoleh sesuai dengan nilai yang sebenarnya [42].

2.2.7 F1-Score

F1-Score adalah perbandingan rata-rata antara *precision* dan *recall* dalam suatu metrik evaluasi. Meskipun diperoleh dari performa algoritma pada dataset yang beragam, evaluasi yang lebih keseluruhan memerlukan perbandingan terhadap beragam parameter. Representasi visual, seperti grafik, dapat memperjelas perbandingan ini untuk analisis yang lebih mendalam dan komprehensif [42].

2.2.7 Pelabelan

Vader Lexicon dalam Python digunakan untuk otomatis memberi label pada sentimen. Setiap kata dianalisis untuk melihat seberapa banyak aspek positif dan negatif yang terkandung di dalamnya. Metode ini menggunakan sistem yang rumit untuk menentukan apakah kata-kata tersebut memiliki implikasi positif atau negatif dalam penilaian Vader Lexicon. Jika dua kata memiliki konotasi positif dan negatif, penggabungannya cenderung menghasilkan kalimat yang negatif. Ini menunjukkan bahwa setiap kalimat dan kata memiliki nilai tersendiri yang ditentukan berdasarkan struktur kalimatnya. Jika struktur kalimat lebih cenderung ke arah negatif, maka akan dikategorikan sebagai kalimat dengan sentimen negatif, begitu pula sebaliknya [43].

2.3 Teori Framework dan Algoritma yang digunakan

2.3.1 Naïve Bayes

Naïve bayes *classifier* (NBC) merupakan suatu metode dengan klasifikasi yang sesuai dengan teorema Bayes. Dalam metode naïve Bayes terkenal dengan kesederhanaan dan efisiensinya dalam klasifikasi teks dan data, yang biasanya digunakan untuk analisis sentimen [44]. Pendekatan ini sering diterapkan pada masalah prediksi berbasis klasifikasi. Diketahui bahwa algoritma ini memiliki akurasi yang sangat tinggi [45].

Fungsinya untuk menghitung dan mencari nilai probabilitas tertinggi untuk pengklasifikasian data uji ke dalam kategori yang sesuai. Secara umum, rumus Naive Bayes dapat dinyatakan sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.1)$$

Rumus 2.1 Rumus Naïve Bayes

Keterangan:

X = Data dengan class yang belum diketahui

H = Hipotesis data suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X

P(H) = Probabilitas hipotesis H

P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) = Probabilitas X

Pada rumus 2.1 merupakan rumus naïve bayes, rumus ini sederhana, namun algoritma naïve bayes memanfaatkan asumsi "naïve" bahwa semua atribut adalah independen satu sama lainnya, meskipun dalam kenyataannya seringkali tidak demikian. Meskipun demikian, Naïve Bayes tetap menjadi algoritma yang sangat populer karena sifatnya yang sederhana dan kinerja yang baik, terutama dalam klasifikasi teks dan kategori yang berbasis teks. Algoritma naïve bayes dapat diimplementasikan dalam berbagai konteks, termasuk dalam analisis sentimen.

Cara kerja naïve bayes untuk proses analisis sentiment dibagi kedalam 2 proses, yaitu proses training dan testing. Proses training digunakan untuk menghasilkan model analisis sentimen yang nantinya akan digunakan sebagai acuan untuk mengklasifikasikan sentimen dengan data testing atau data mentah yang baru [46].

2.3.2 Text Mining

Text mining merupakan suatu teknik analisis yang digunakan untuk mengeksplorasi dan menarik informasi berharga dari berbagai dokumen atau teks dalam jumlah besar. Dalam metode ini, algoritma serta pendekatan statistik diterapkan untuk mengolah, mengatur, dan menyajikan data teks agar dapat dipahami dan dianalisis baik oleh manusia maupun sistem komputer. Tujuan utama dari *text mining* adalah untuk memperoleh informasi yang berharga dari kumpulan dokumen tersebut. Karenanya, sumber data yang dimanfaatkan dalam *text mining* adalah teks yang bersifat tidak terstruktur atau setidaknya semi-terstruktur. dalam *text mining* meliputi pengkategorisasian dan pengelompokan teks [47].

2.3.3 Term Frequency Inverse Document Frequency (TF-IDF)

TF-IDF merupakan metode dalam pemrosesan bahasa alami yang bertujuan untuk menilai kepentingan relatif suatu kata dalam satu dokumen atau sekelompok dokumen yang disebut korpus. Konsep ini mempertimbangkan dua faktor utama: Frekuensi Kata (TF) dan Inverse Document Frequency (IDF). TF mengukur seberapa sering suatu kata muncul dalam satu dokumen dengan membandingkannya dengan jumlah total kata dalam dokumen tersebut. IDF, di sisi lain, menilai seberapa unik atau langka suatu kata di seluruh korpus dokumen dengan membandingkan logaritma dari total dokumen dengan jumlah dokumen yang mengandung kata tersebut. Formula TF-IDF menggabungkan kedua faktor ini untuk menghasilkan skor yang mencerminkan relevansi suatu kata terhadap dokumen dalam konteks korpus secara keseluruhan. Dengan pendekatan ini, kata-kata yang sering muncul dalam satu dokumen tetapi juga umum di seluruh korpus akan memiliki bobot yang lebih rendah, sementara kata-kata yang muncul jarang dalam satu dokumen tetapi unik di seluruh korpus akan memiliki bobot yang lebih tinggi. Penggunaan TF-IDF umumnya meliputi ekstraksi fitur dalam

analisis teks, pengelompokan dokumen, dan pengembangan sistem rekomendasi [48].

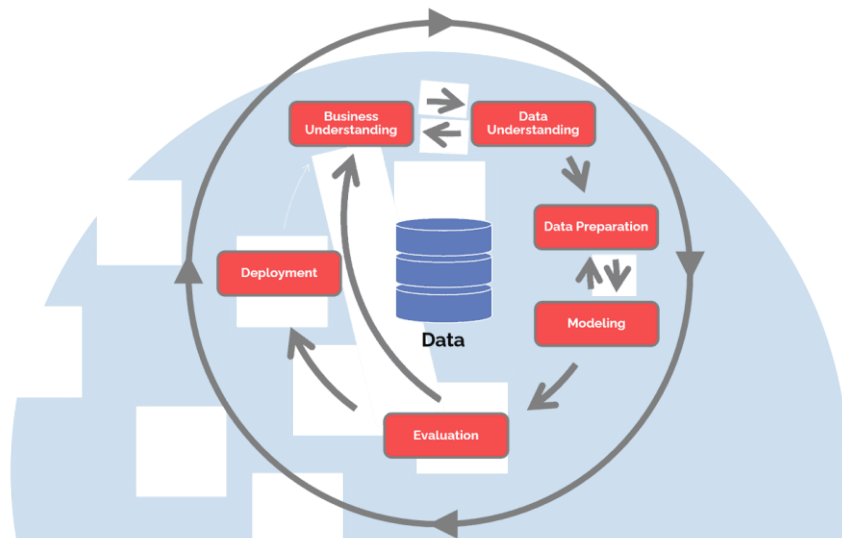
2.3.4 Text pre-processing

Text pre-processing adalah langkah awal dalam mempersiapkan teks agar siap untuk diolah lebih lanjut. Proses ini melibatkan beberapa tahap, yang dijelaskan sebagai berikut [49]:

1. *Tokenizing*: Proses ini melibatkan pemisahan string input menjadi token berdasarkan setiap kata yang membentuknya.
2. *Case Folding*: Pada tahap ini, semua huruf dalam dokumen dikonversi menjadi huruf kecil. Hanya karakter dari "a" sampai "z" yang tetap dipertahankan, sementara karakter lain dianggap sebagai pembatas dan dihilangkan.
3. *Cleaning*: Tahap ini mencakup proses eliminasi kata-kata yang tidak relevan untuk mengurangi noise pada proses klasifikasi. Ini termasuk simbol, angka, URL, hashtag (#), dan mention (@username).
4. *Normalization*: Normalisasi melibatkan pengubahan kata-kata non-baku menjadi baku dengan menggunakan kamus yang berisi pasangan kata non-baku dan baku.

2.3.5 CRISP-DM

Dalam Metode penelitian yang digunakan adalah CRISP-DM (*Cross Industry Standard Process for Data Mining*). Metode ini menerapkan 6 tahapan dalam mengolah suatu masalah mengenai data mining, yaitu *Business Understanding, Data Understanding, Data preparation, Modeling, Evaluation, Deployment* [50].



Gambar 2.1 Metode CRISP-DM

Sumber:[51]

Berikut ini merupakan tahapan dari metode CRISP-DM, yaitu[51]:

2.3.5.1 Business Understanding

Dalam kerangka CRISP-DM Tahap ini dimulai dengan memahami tujuan bisnis dari proyek *data mining*. Fokus pada identifikasi tujuan yang spesifik dan pertanyaan bisnis yang ingin dijawab dengan menggunakan data. Dalam tahap ini, tim proyek berinteraksi dengan pemangku kepentingan bisnis untuk memahami kebutuhan dan ekspektasi mereka.

2.3.5.2 Data Understanding

Setelah tujuan bisnis diklarifikasi, langkah berikutnya adalah memahami data yang tersedia untuk proyek. Ini meliputi pengumpulan data, pemeriksaan kualitas data, eksplorasi data awal, dan pemahaman karakteristik dan hubungan antar variabel dalam dataset.

2.3.5.3 Data Preparation

Tahap ini melibatkan pra-pemrosesan atau pembersihan data untuk mempersiapkan data mentah agar siap digunakan dalam proses

analisis. Aktivitas yang dilakukan termasuk pembersihan data (*cleaning*), transformasi data, penggabungan data dari berbagai sumber, dan pemilihan fitur.

2.3.5.4 Modeling

Setelah data dipersiapkan, langkah selanjutnya adalah membangun model untuk menganalisis data. Berbagai teknik dan algoritma data mining dapat diterapkan pada tahap ini, seperti regresi, klasifikasi, *clustering*, atau aturan asosiasi, tergantung pada tujuan analisis dan jenis data yang digunakan.

2.3.5.5 Evaluation

Tahap evaluasi dilakukan untuk mengevaluasi kinerja model yang telah dibangun. Metrik evaluasi digunakan untuk mengukur seberapa baik model dapat memecahkan masalah yang ada atau memenuhi tujuan bisnis yang ditetapkan. Hasil evaluasi dapat digunakan untuk memilih model terbaik atau menentukan langkah selanjutnya dalam proyek.

2.3.5.6 Deployment

Tahap terakhir dalam CRISP-DM adalah implementasi model dan hasil analisis ke lingkungan produksi. Ini melibatkan integrasi model ke dalam sistem bisnis atau pengambilan keputusan, serta pelaporan hasil kepada pemangku kepentingan.

2.4 Tools yang digunakan

2.4.1 Jupyter Notebook

Penggunaan aplikasi Jupyter notebook telah menjadi kunci bagi para ilmuwan data sejak diluncurkan pada tahun 2015. Keberadaan data memiliki peranan penting dalam berbagai konteks saat ini, dimana banyak perusahaan memberikan peluang karir dalam bidang analisis dan presentasi data untuk memperkuat citra perusahaan dan membuat keputusan strategis untuk masa depan. Kebijakan perusahaan seringkali

bergantung pada informasi yang diperoleh dari data awal, sehingga publikasi artikel dalam dunia akademis menjadi krusial untuk memberikan wawasan khususnya dalam hal pengindeksan. Aplikasi Jupyter notebook sendiri mengambil singkatan dari bahasa pemrograman Julia (Ju), Python (Py), dan R. Jupyter notebook merupakan platform online yang gratis untuk membuat, berbagi, serta menyajikan kode, hasil perhitungan, visualisasi analisis, dan narasi komputasi [52].

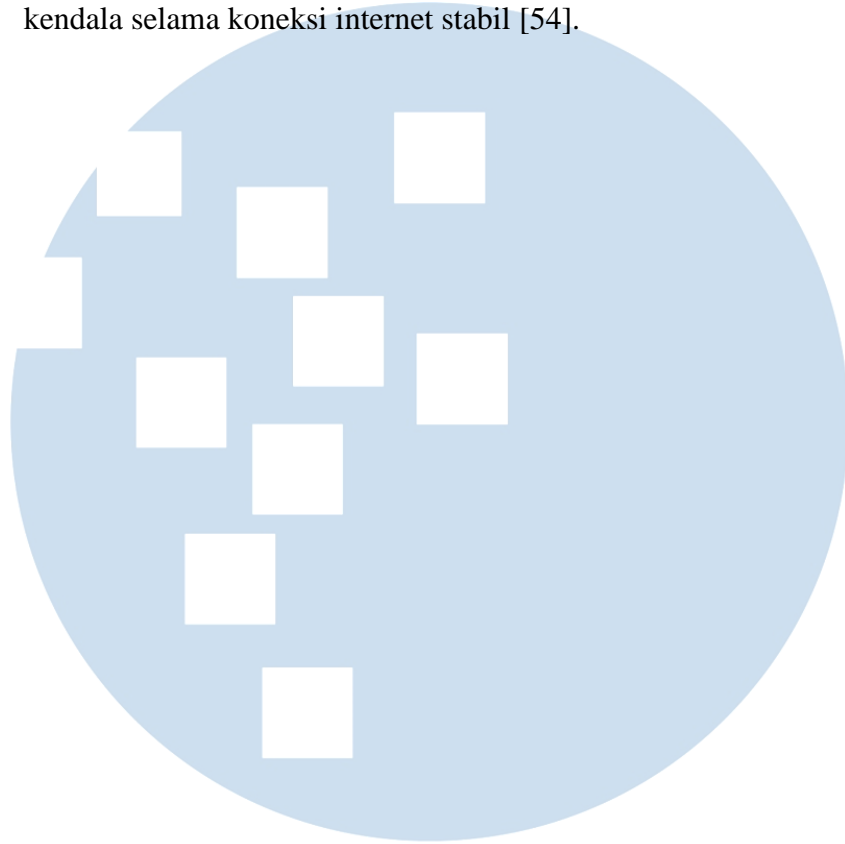
2.4.2 Python

Python adalah bahasa pemrograman serbaguna yang bersifat interpretatif, dengan filosofi desain yang mengutamakan kejelasan kode. Python diakui sebagai bahasa yang menggabungkan kemampuan dan kegunaan, dengan sintaksis kode yang sangat mudah dipahami, serta didukung oleh berbagai pustaka standar yang luas dan komprehensif. Python dapat dianggap sebagai bahasa pemrograman yang diciptakan untuk tujuan umum, namun dengan fokus khusus pada kejelasan dalam pembacaan kode. Selain itu, Python juga menyediakan beragam pustaka lengkap yang memungkinkan para pengembang untuk membuat aplikasi yang canggih dengan menggunakan kode yang terlihat sederhana [53].

2.4.3 Google Colab

Google colab adalah suatu lingkungan pengembangan bahasa pemrograman Python di mana eksekusi kode dilakukan oleh infrastruktur server tinggi performa milik Google. Dari segi perangkat lunak, Google colab menyediakan sebagian besar pustaka yang dibutuhkan, termasuk, *tensorFlow*, NumPy, Pandas, dan pustaka pendukung lainnya seperti Matplotlib untuk visualisasi data. Selain itu, Google colab juga menyediakan beragam versi dari *TensorFlow* (baik 1.x maupun 2.x) dan Python (mulai dari versi 2.x hingga 3.x). Dari sisi perangkat keras, Google colab menyediakan layanan penyimpanan yang terintegrasi dengan Google drive, serta pilihan prosesor CPU, GPU, dan TPU, beserta kapasitas RAM yang disesuaikan. Dengan infrastruktur server yang

stabil, hampir semua proses dapat dijalankan dengan lancar tanpa kendala selama koneksi internet stabil [54].



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA