

# Perbandingan Support Vector Machine dan Naïve Bayes Terkait Kepuasan Pengguna Bus Listrik Kota Medan

Zery Mesanda\*, Muhammad Ikhsan

Fakultas Sains dan Teknologi, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: <sup>1,\*</sup>zerymesanda23@gmail.com, <sup>2</sup>mhd.ikhsan@uinsu.ac.id

Email Penulis Korespondensi: zerymesanda23@gmail.com

Submitted: 13/07/2024; Accepted: 11/08/2024; Published: 12/08/2024

**Abstrak**—Pemerintah kota telah memperkenalkan inisiatif penggunaan bus listrik sebagai alternatif yang lebih bersih dan berkelanjutan. Keberhasilan suatu sistem transportasi publik tidak hanya ditentukan oleh ketersediaan armada dan infrastruktur, tetapi juga oleh tingkat kepuasan pengguna. Kepuasan pengguna merupakan indikator penting yang mencerminkan sejauh mana layanan tersebut memenuhi harapan dan kebutuhan pengguna. Oleh karena itu, pemahaman mendalam tentang faktor-faktor yang memengaruhi kepuasan pengguna, serta kemampuan untuk memprediksi dan mengelolanya, menjadi kunci dalam meningkatkan kualitas layanan transportasi publik, termasuk penggunaan. Dalam upaya untuk memahami dan meningkatkan kepuasan pengguna bus listrik di Kota Medan, diperlukan pendekatan analisis prediktif yang memadai. Dengan menggunakan metode-metode seperti *Support Vector Machine* (SVM) dan *Naïve Bayes*, kita dapat mengembangkan model prediktif yang dapat mengidentifikasi pola dan tren dalam data pengguna, sehingga memungkinkan pihak terkait untuk mengambil tindakan yang sesuai untuk meningkatkan layanan. Dalam konteks ini, perbandingan antara metode SVM dan *Naïve Bayes* akan memberikan wawasan yang berharga tentang keefektifan masing-masing metode dalam memprediksi kepuasan pengguna bus listrik di Kota Medan. Berdasarkan hasil perbandingan, algoritma *Naïve Bayes* menunjukkan kinerja yang sedikit lebih baik dibandingkan dengan *Support Vector Machine* dalam analisis sentimen ini. Dengan nilai akurasi yang dihasilkan oleh penerapan metode *Naïve Bayes* adalah senilai 58%, sedangkan dengan menerapkan metode *Support Vector Machine* senilai 57%. Meskipun demikian, kedua algoritma memberikan *insight* yang berharga tentang sentimen masyarakat Medan terhadap Bus Listrik.

**Kata Kunci:** Bus Listrik; Penambangan Teks; Support Vector Machine; Naïve Bayes

**Abstract**—The city government has introduced an initiative to use electric buses as a cleaner and more sustainable alternative. The success of a public transportation system is not only determined by the availability of fleet and infrastructure, but also by the level of user satisfaction. User satisfaction is an important indicator that reflects the extent to which the service meets users' expectations and needs. Therefore, an in-depth understanding of the factors that influence user satisfaction, as well as the ability to predict and manage them, is key in improving the quality of public transportation services, including usage. In an effort to understand and improve trolleybus user satisfaction in Medan City, an adequate predictive analysis approach is required. By using methods such as Support Vector Machine (SVM) and Naïve Bayes, we can develop predictive models that can identify patterns and trends in user data, thus enabling relevant parties to take appropriate actions to improve services. In this context, the comparison between SVM and Naïve Bayes methods will provide valuable insights into the effectiveness of each method in predicting the satisfaction of electric bus users in Medan City. Based on the comparison results, the Naive Bayes algorithm shows slightly better performance compared to the Support Vector Machine in this sentiment analysis. The accuracy value generated by applying the Naive Bayes method is 58%, while applying the Support Vector Machine method is 57%. Nonetheless, both algorithms provide valuable insights into the sentiment of Medan people towards Electric Buses.

**Keywords:** Electric Bus; Text Mining; Support Vector Machine; Naïve Bayes

## 1. PENDAHULUAN

Di tengah meningkatnya kesadaran akan keberlanjutan lingkungan dan kebutuhan akan mobilitas yang efisien, transportasi publik menjadi fokus penting dalam pengembangan perkotaan. Kota Medan, sebagai salah satu pusat urbanisasi utama di Indonesia, mengalami tantangan dalam menyediakan layanan transportasi yang efisien dan ramah lingkungan bagi penduduknya. Dalam konteks ini, pemerintah kota telah memperkenalkan inisiatif penggunaan bus listrik sebagai alternatif yang lebih bersih dan berkelanjutan [1].

Keberhasilan suatu sistem transportasi publik tidak hanya ditentukan oleh ketersediaan armada dan infrastruktur, tetapi juga oleh tingkat kepuasan pengguna. Kepuasan pengguna merupakan indikator penting yang mencerminkan sejauh mana layanan tersebut memenuhi harapan dan kebutuhan pengguna. Oleh karena itu, pemahaman mendalam tentang faktor-faktor yang memengaruhi kepuasan pengguna, serta kemampuan untuk memprediksi dan mengelolanya, menjadi kunci dalam meningkatkan kualitas layanan transportasi publik, termasuk penggunaan [2].

Selain Kota Medan, beberapa kota besar lainnya di Indonesia juga telah mengadopsi bus listrik sebagai bagian dari solusi transportasi berkelanjutan. Misalnya, Jakarta, Surabaya, Bandung, dan beberapa kota lainnya telah memperkenalkan armada bus listrik sebagai upaya untuk mengurangi polusi udara dan mempromosikan mobilitas ramah lingkungan [3]. Meskipun tingkat adopsi dan ketersediaan bus listrik bervariasi di setiap kota, tren umumnya menunjukkan peningkatan minat dalam mengadopsi teknologi ini sebagai bagian dari strategi pengurangan emisi [2].

Meskipun ada dorongan untuk mengadopsi bus listrik, ketersediaan dan efektivitas layanan ini di berbagai kota di Indonesia masih dihadapkan pada tantangan tertentu. Salah satu tantangan utama adalah

infrastruktur pengisian daya yang terbatas dan kurangnya pemahaman publik tentang manfaat dan kenyamanan menggunakan transportasi berbasis listrik [4]. Namun demikian, upaya pemerintah dan dukungan dari sektor swasta membuka peluang besar untuk memperluas jaringan bus listrik dan meningkatkan kesadaran masyarakat tentang manfaatnya.

Pemerintah Indonesia telah meluncurkan berbagai program dan insentif untuk mendukung penggunaan bus listrik sebagai bagian dari strategi transportasi berkelanjutan. Ini termasuk penyediaan subsidi untuk pembelian bus listrik, peningkatan infrastruktur pengisian daya, dan pengembangan regulasi yang mendukung pengoperasian bus listrik di jalan raya[5].

Transportasi merupakan salah satu bagian penting yang menunjang aktivitas manusia. Namun, laporan dari *International Energy Agency* (IEA) tahun 2020, menyatakan bahwa kurang lebih 24% emisi CO<sub>2</sub> (emisi gas karbon) global disebabkan oleh sektor transportasi, yang berasal dari pembakaran bahan bakar fosil. Ada satu solusi yang mungkin dilakukan untuk *dekarbonisasi* pada bagian transportasi adalah pergerakan elektro melalui penggunaan kendaraan listrik dengan emisi CO<sub>2</sub> yang rendah [6].

Mengenalkan bus listrik di Kota Medan membawa cara tidak hanya dalam penyediaan armada kendaraan yang ramah lingkungan, tetapi juga dalam peningkatan prasarana yang sesuai, seperti stasiun pengisian daya dan jalur yang dioptimalkan. Walaupun demikian, tantangan seperti biaya pemodal awal, pemeliharaan infrastruktur, dan kesadaran masyarakat tentang manfaat bus listrik tetap menjadi perhatian yang perlu diatasi. Dengan kewajiban untuk menghadapi tantangan lingkungan dan transportasi sebagai inovatif dan berkepanjangan, pemerintah Kota Medan berniat mengangkat bus listrik dapat memberikan pengganti transportasi yang lebih bersih dan efisien bagi warganya.

Data *mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [7]. Data Mining dapat diterapkan pada berbagai bidang yang mempunyai sejumlah data, tetapi karena wilayah penelitian dengan sejarah yang belum lama, dan belum melewati masa ‘remaja’, maka data mining masih diperdebatkan posisi bidang pengetahuan yang memilikinya. Maka Daryl Pregibon menyatakan bahwa “Data Mining adalah campuran dari statistik, kecerdasan buatan, dan riset basis data” yang masih berkembang Data Mining mempunyai empat akar bidang ilmu yaitu Statistik, Kecerdasan Buatan (*Artificial Intelligence*), Pengenalan Pola, dan Sistem Basis Data [8].

Pada Text mining selalu melibatkan pra proses dokumen yaitu melakukan kategorisasi teks, melakukan ekstraksi informasi, dan mengekstraksi kata. Metode ini untuk mengekstraksi informasi yang diambil dari sumber data dengan cara mengidentifikasi serta melakukan eksplorasi pola yang menarik. *Text Mining* merupakan teknik yang digunakan untuk menangani permasalahan klasifikasi, *clustering*, *information extraction* dan *information retrieval* [9]. Tugas Teks *preprocessing* adalah termasuk di dalamnya pemilihan data, klasifikasi dan ekstraksi fitur untuk mengubah dokumen menjadi bentuk perantara, yang harus cocok dengan tujuan pencarian yang berbeda. Bagian utama dari pekerjaan Pekerjaan selanjutnya sebelum pemrosesan memanipulasi data atau informasi terbaru yang berasal dari prose teks *mining*, seperti evaluasi dan pemilihan informasi yang ditemukan, *interpretasi* dan *visualisasi* informasi yang dihasilkan [10].

Analisis sentimen mengolah data secara otomatis untuk memperoleh sentimen sebuah kalimat. Hal ini dilakukan untuk melihat pendapat atau kecenderungan opini masyarakat apakah cenderung berpandangan positif atau negatif [11]. Analisis sentimen ialah proses mengekstraksi, mengolah dan memahami data berupa teks yang tidak terstruktur secara otomatis guna mengambil informasi sentimen yang terdapat pada sebuah kalimat pendapat atau opini. Analisis sentimen dapat diterapkan pada opini semua bidang seperti ekonomi, politik, sosial dan hukum. Media sosial Twitter ini membuka jendela bagi para peneliti untuk mempelajari emosi, suasana hati, dan pendapat publik melalui analisis sentimen [12].

Klasifikasi dapat didefinisikan sebagai pekerjaan yang melakukan pelatihan/pembelajaran terhadap fungsi target yang memetakan setiap set atribut (fitur) kesatu jumlah label kelas yang tersedia, bisa juga diartikan pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari jumlah kelas yang tersedia [13]. Klasifikasi memiliki dua jenis data yaitu data uji (testing data) dan data latih. Data latih adalah data yang memang berdasarkan dari fakta yang contohnya data positif dan data negatif juga berfungsi untuk membentuk tabel probabilitas. Sementara itu data uji adalah data yang telah berkelas atau berlabel dan berfungsi untuk menguji atau menghitung nilai akurasi dari data atau tabel probabilitas yang telah dibentuk [14].

*Support vector machine* (SVM) adalah jenis model vektor berdasarkan pengklasifikasi yang memerlukan konversi teks ke vektor sebelum dapat digunakan untuk klasifikasi [15]. *Key idea* dari *Support vector machine* adalah untuk menemukan permukaan keputusan *Hyperlane* yang maksimal dari setiap titik data. Untuk melakukan *training* mesin yang didukung oleh vektor atau biasa disebut *Support Vector Machine* memerlukan solusi *Quadratic Programming* (QP) yang sangat besar. *Quadratic Programming* adalah masalah matematika untuk menemukan vektor “x” yang meminimalkan fungsi kuadrat, dengan melakukan pembagian kelas menggunakan *hyperlink* maka masing-masing kelas positif dan negatif dapat dibagi berdasarkan area masing-masing sehingga ketika terdapat data baru dapat ditentukan kelasnya berdasarkan area positif dan negatif [16].

*Naive Bayes* adalah algoritma analisis statistik yang menggunakan probabilitas *Bayesian* untuk melakukan pemrosesan data pada data numerik. Klasifikasi *Bayesian* adalah klasifikasi statistik yang dapat memprediksi kelas anggota probabilitas [17]. Dalam klasifikasi *Bayesian* sederhana, yang lebih dikenal sebagai pengklasifikasi *Nave-Bayes*, pengaruh nilai atribut kelas dapat diasumsikan tidak terpengaruh atau tidak

terpengaruh oleh nilai atribut lainnya. Sederhananya, kita berasumsi bahwa kemunculan istilah kata dalam sebuah kalimat tidak dipengaruhi oleh kata lain, jadi dalam analisis sentimen ada bobot untuk setiap kemunculan kata, yang dihitung sebagai jumlah bobot untuk keseluruhan kalimat. Apakah kalimat itu positif, netral, atau negatif [18].

Hasil pengujian klasifikasi dilakukan evaluasi untuk mengetahui nilai akurasi dari algoritma yang akan dianalisa apakah mode klasifikasi yang dibuat layak digunakan [19]. Metode evaluasi yang digunakan untuk mengukur kinerja dari algoritma klasifikasi adalah *confusion matrix*. *Confusion matrix* menampilkan perbandingan hasil klasifikasi yang dilakukan dengan data sebenarnya dalam bentuk tabel *matrix* [20].

Penelitian yang dilakukan oleh Fikri dkk (2020) yang berjudul “Perbandingan Metode Klasifikasi *Support Vector Machine* dan *Naïve Bayes* Pada Analisis Sentimen Twitter” Analisis dilakukan dengan mengklasifikasikan tweets yang berisi sentimen masyarakat mengenai *UMM*. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Metode Support Vector Machine* dan *Naïve Bayes* dengan pembobotan menggunakan *TF-IDF*. Hasil komparasi kedua metode menunjukkan bahwa *Naïve Bayes* mendapatkan hasil akurasi yang lebih baik dari *Metode Support Vector Machine* dengan akurasi sebesar 73,65% [21].

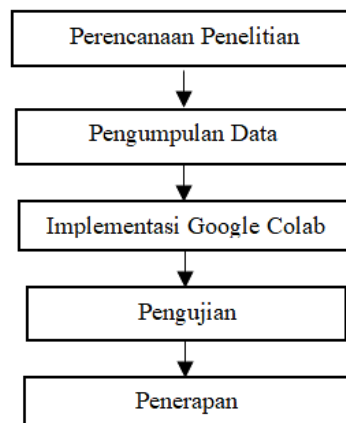
Sedangkan penelitian yang dilakukan oleh Yunita & Kamayani (2023) yang berjudul “Perbandingan Algoritma *Metode Support Vector Machine* dan *Naïve Bayes* Pada Analisis Sentimen Kebijakan Penghapusan Kewajiban Skripsi” Penelitian ini berasal dari data Twiter. Penelitian ini membandingkan hasil evaluasi algoritma *Metode Support Vector Machine* dengan *Naïve Bayes* menggunakan 80% data latih dan 20% data uji. Berdasarkan hasil yang diperoleh dari penelitian ini didapatkan 331 sentimen positif serta 369 sentimen negatif dan ditarik kesimpulan bahwa *Support Vector Machine* menjadi algoritma yang terbaik dengan *accuracy* 80%, *recall* 83%, *precision* 76%, dan *F1-Score* 79% [22].

Dalam upaya untuk memahami dan meningkatkan kepuasan pengguna bus listrik di Kota Medan, diperlukan pendekatan analisis prediktif yang memadai. Dengan menggunakan metode-metode seperti *Support Vector Machine* (SVM) dan *Naïve Bayes*. Dalam konteks ini, perbandingan antara metode SVM dan *Naïve Bayes* akan memberikan wawasan yang berharga tentang keefektifan masing-masing metode dalam memprediksi kepuasan pengguna bus listrik di Kota Medan.

## 2. METODOLOGI PENELITIAN

### 2.1 Kerangka Penelitian

Kerangka penelitian adalah desain atau gambaran tahapan yang digunakan dalam penelitian. Pada penelitian ini menggunakan pendekatan kuantitatif karena akan membandingkan kinerja dua metode *Support Vector Machine* dan *Naïve Bayes* dalam menganalisis data mengenai kepuasan pengguna bus listrik.



Gambar 1. Kerangka Penelitian

Tahap awal dalam penelitian ini melibatkan perencanaan untuk memastikan kelancaran dan validitas proses penelitian. *Metode Support Vector Machine* dan *Naïve Bayes* akan diterapkan pada data yang dikumpulkan melalui teknik *scraping* komentar di beberapa postingan Instagram yang terkait dengan layanan bus listrik di Kota Medan. Dengan menggunakan kedua metode ini, penelitian bertujuan untuk menghasilkan analisis yang akurat dan mendalam tentang kepuasan pengguna, serta memperoleh perbandingan kedua algoritma dalam hal tingkat akurasi yang tinggi dan kualitas yang baik dalam penentuan tingkat kepuasan pengguna layanan bus listrik.

Mengumpulkan komentar dari beberapa postingan Instagram yang terkait dengan layanan bus listrik di Kota Medan. Menggunakan teknik *scraping* untuk mengumpulkan komentar dari pengguna Instagram yang memberikan tanggapan tentang layanan bus listrik. Komentar yang telah dikumpulkan akan disimpan dalam

format yang sesuai untuk analisis lebih lanjut, seperti format Excel. Setelah periode pengumpulan selesai, data dari komentar tersebut akan diproses dan dibersihkan untuk memastikan kualitas dan relevansi data.

Google Colab akan digunakan sebagai platform untuk mengimplementasikan algoritma *Metode Support Vector Machine* dan *Naïve Bayes* dalam menganalisis data kepuasan pengguna. Proses ini melibatkan persiapan lingkungan kerja di Google Colab, pengimporan data yang telah dikumpulkan, serta penerapan preprocessing data untuk mempersiapkan data untuk analisis. Setelah data siap, algoritma *Metode Support Vector Machine* dan *Naïve Bayes* akan diimplementasikan untuk menganalisis sentimen dan memprediksi tingkat kepuasan pengguna. Hasil analisis ini kemudian akan disimpan dan disiapkan untuk tahap pengujian dan evaluasi.

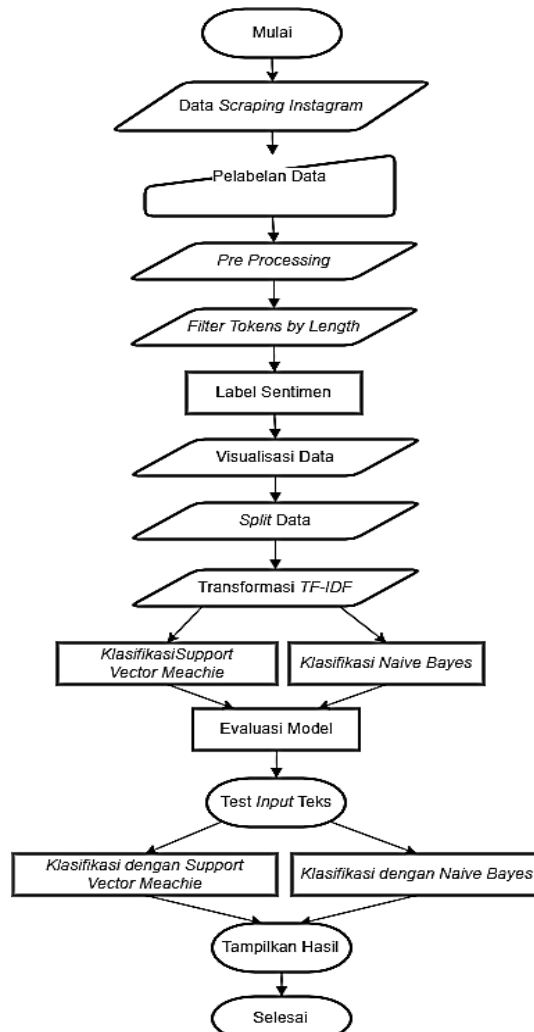
Mengacu pada sejauh mana data yang dikumpulkan konsisten dan dapat diandalkan. Memeriksa apakah ada data yang tidak valid atau tidak sesuai dengan kriteria inklusi. Semua data yang diperlukan telah terkumpul secara lengkap dan konsisten. Ini termasuk memeriksa apakah ada nilai yang hilang atau tidak lengkap, serta apakah data-data tersebut konsisten dengan format yang ditetapkan [23].

Setelah mendapatkan hasil dari proses pengolahan data dengan menggunakan metode *Support Vecthor Meachie* dan *Naïve Bayes*, dilakukan pemrosesan kembali dengan menggunakan *confusion matrix* untuk mendapatkan hasil dari *akurasi, presisi, recall dan f1-score* [24].

## 2.2 Perancangan Support Vector Machine dan Naive Bayes

Perancangan yaitu tahap dari perencanaan, penggambaran dan pembuatan skesta atau pengetahuan dari beberapa elemen yang terpisah ke dalam satu kesatuan yang utuh. Perancangan model terdiri dari metode *Support Vector Machine* dan *naïve bayes*.

Google Colab akan digunakan sebagai platform untuk mengimplementasikan algoritma *Metode Support Vector Machine* dan *Naïve Bayes* dalam menganalisis data kepuasan pengguna. Proses ini melibatkan persiapan lingkungan kerja di Google Colab, pengimporan data yang telah dikumpulkan, serta penerapan preprocessing data untuk mempersiapkan data untuk analisis. Setelah data siap, algoritma *Metode Support Vector Machine* dan *Naïve Bayes* akan diimplementasikan untuk menganalisis sentimen dan memprediksi tingkat kepuasan pengguna. Hasil analisis ini kemudian akan disimpan dan disiapkan untuk tahap pengujian dan evaluasi.



Gambar 2. Flowchart klasifikasi Support Vector Machine dan Naive Bayes

Model ini akan memprediksi sentimen teks berdasarkan pola yang ditemukan dalam data latih. Menilai kinerja model *Support Vector Machine* dan *Naïve Bayes* dengan menggunakan metrik seperti akurasi, presisi, *recall*, dan *f1-score* untuk memahami seberapa baik model dalam mengklasifikasikan sentimen. Mengambil *input* teks dari pengguna untuk diklasifikasikan oleh model yang telah dilatih. Model *Support Vector Machine* yang telah dilatih mengklasifikasikan *input* teks pengguna. Model *Naïve Bayes* yang telah dilatih mengklasifikasikan *input* teks pengguna. Menampilkan hasil klasifikasi dari kedua model, memberikan pengguna informasi tentang sentimen teks yang dimasukkan. Proses selesai, memberikan hasil akhir dari analisis sentiment.

Setelah mendapatkan hasil dari proses pengolahan data dengan menggunakan metode *Support Vector Machine* dan *Naïve Bayes*, dilakukan pemrosesan kembali dengan menggunakan *confusion matrix* untuk mendapatkan hasil dari akurasi, presisi, *recall* dan *f1-score*.

Data dikumpulkan dari postingan *Instagram* yang terkait dengan topik bus listrik. Ini melibatkan penggunaan teknik *web scraping* untuk mendapatkan teks komentar dari berbagai postingan yang relevan. Setelah data dikumpulkan, setiap komentar atau postingan diberi label secara manual berdasarkan sentimen yang terkandung di dalamnya, yaitu apakah sentimen tersebut positif atau negatif. Teks yang terkumpul diproses lebih lanjut untuk membersihkan dan menormalkan data. Ini mencakup menghilangkan mention (@username), tautan (URL), dan karakter atau emotikon yang tidak diinginkan, menghapus karakter non-alfanumerik dan mengurangi spasi berlebih, normalisasi teks dengan mengubah kata-kata slang atau singkatan menjadi bentuk standar yang lebih formal, teks di filter berdasarkan jumlah kata untuk memastikan setiap entri memiliki panjang yang sesuai (misalnya, antara 3 hingga 50 kata) guna meningkatkan kualitas analisis, setiap teks diberi label sentimen berdasarkan nilai skor yang telah ditentukan. Misalnya, skor  $\leq 3$  dikategorikan sebagai negatif, dan skor  $> 3$  sebagai positif. Menampilkan distribusi sentimen dengan grafik batang untuk melihat proporsi data positif dan negatif dan membuat *word cloud* untuk visualisasi kata-kata yang sering muncul dalam teks positif dan negatif. Data dibagi menjadi dua set data latih (80% dari total data) untuk melatih model dan data uji (20% dari total data) untuk menguji kinerja model. Mengubah teks menjadi fitur numerik menggunakan teknik *TF-IDF*, yang memberikan bobot lebih besar pada kata-kata yang unik dan penting dalam konteks dokumen tertentu. Menggunakan fitur *TF-IDF* untuk melatih model *Support Vector Machine* dan *Naïve Bayes* dan menguji kinerjanya.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Pada penelitian ini, peneliti menganalisis data sentimen terhadap layanan bus listrik di Kota Medan berdasarkan komentar pengguna di *Instagram*. Tahap pertama dalam penelitian adalah melakukan pengumpulan komentar *instagram*. Proses pengumpulan data komentar *instagram* mengenai layanan bus listrik dilakukan pada beberapa postingan *Instagram* terkait dengan menggunakan *library Python* sebagai *scraping* data.

*Scraping* adalah proses otomatis untuk mengekstrak data spesifik dari halaman web. Ini melibatkan mengakses halaman web, mengambil konten yang diinginkan, dan menyimpan data tersebut dalam format yang lebih terstruktur misalnya, *CSV*, *database*.

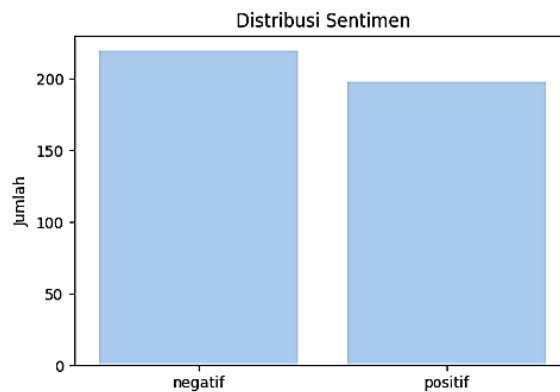
Proses *scraping* data dalam penelitian ini tidak bergantung pada rentang waktu tertentu. Data hasil *scraping* merupakan komentar pada postingan yang terkait dengan layanan bus listrik di Kota Medan. Hasil dari proses *scraping* akan diubah ke dalam format *.CSV* untuk kemudian diolah menggunakan bahasa pemrograman *Python*. Jumlah data yang diperoleh dari hasil *scraping* adalah 417 komentar terbaru yang terkait dengan layanan bus listrik. Proses *scraping* dilakukan dengan teknik yang memanfaatkan *library Selenium* dan *BeautifulSoup* untuk mengekstraksi data dari halaman *web Instagram*. Data yang terkumpul kemudian disimpan dalam file *.CSV* untuk memudahkan proses analisis lebih lanjut. Hasil data komentar yang berhasil dikumpulkan ditunjukkan pada Tabel 1 di bawah ini.

Tabel 1. Hasil Data *scraping Instagram*

No	Data <i>scraping</i>
1	Saran :1. Ada web atau apps supaya bs tracking posisi bus, soalnya pengalaman naik kmrn nunggu di halte sampe sejam lebih, ga sesuai dgn jadwal yg di halte.Kasihannya yg lansia atau yg bawa balita.2. Setiap halte kl bisa ada kursi dan atap, jd penumpang yg lg menunggu ga capek dan ga kepanasan, aplg jam tiba di halte yg kdg delay ga menentu, again butuh apps utk bs tracking posisi bus.3. Pintu bus yg tengah apa emg blm d fungsikan ya? Soalnya cm 1 pintu di depan utk penumpang keluar masuk nya, jd desak2an dan berjubel d pintu.Thx
2	Udah sangat membantu sekali, cuma saran kalau bisa ada web untuk tracking keberadaan busnya jadi tau estimasi tiba di tiap halte
3	Ga guna ini busnya, supir ga mau nepi ke halte, ga mau nunggu ntah 10-15 detik main gas aja□. Mentang mentang masih gratis ya jd sukaa sukaa□
4	Enak, cm kmren sial aja. Penumpang naik posisi berdiri, ketek nya bau□. Padahal masih pagi loh□

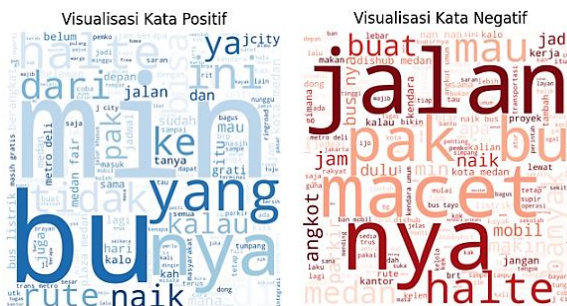
Tahap selanjutnya adalah dengan melakukan *preprocessing* data, yakni tahap penting dalam analisis data, terutama untuk data yang diperoleh dari media sosial seperti *Instagram*. Proses ini bertujuan untuk membersihkan dan menyiapkan data agar siap untuk analisis lebih lanjut. Data dikumpulkan melalui teknik *web scraping* pada beberapa postingan *Instagram* yang relevan dengan topik penelitian. Metode ini digunakan untuk mengotomatisasi proses pengambilan data dari *internet*. Tahapan sistem pada saat proses *preprocessing* data yang meliputi *Cleanig Data, Normalisasi, Stopword, Tokenized, Stemming*.

Memahami sebaran data dengan memvisualisasikan distribusi sentimen, kita dapat dengan mudah melihat bagaimana sentimen positif dan negatif tersebar dalam dataset. Hal ini membantu dalam memahami keseimbangan atau ketidakseimbangan antara dua kategori sentimen tersebut. Mengidentifikasi pola visualisasi memungkinkan kita untuk mengidentifikasi pola tertentu dalam data, seperti tren kenaikan atau penurunan sentimen positif atau negatif dari waktu ke waktu atau dalam konteks tertentu. Perbandingan yang mudah grafik memberikan cara yang cepat dan efisien untuk membandingkan jumlah komentar positif dan negatif. Ini sangat berguna dalam situasi di mana kita perlu membuat keputusan berdasarkan perbandingan ini, seperti dalam analisis kepuasan pelanggan. Komunikasi yang efektif visualisasi data memudahkan komunikasi hasil analisis kepada berbagai pemangku kepentingan, termasuk yang mungkin tidak memiliki latar belakang teknis. Grafik dan diagram membuat informasi lebih mudah diinterpretasikan dibandingkan dengan tabel angka yang kompleks.



Gambar 3. Visualisasi Sentimen

*Word Cloud* adalah representasi visual dari teks di mana ukuran kata menunjukkan frekuensi atau pentingnya dalam teks dan membantu kita melihat kata-kata yang paling sering muncul dalam teks berdasarkan frekuensi. Kata-kata yang sering muncul dalam teks ditampilkan dengan ukuran yang lebih besar, sehingga memudahkan identifikasi kata-kata yang paling penting atau paling sering digunakan. Dengan melihat kata-kata yang paling sering muncul, kita dapat dengan cepat memahami tema atau topik utama dalam teks. *Word Cloud* juga dapat digunakan untuk visualisasi sentimen dengan memanfaatkan palet warna. Misalnya, kata-kata yang berhubungan dengan sentimen positif dapat diberi warna biru, sementara kata-kata yang berhubungan dengan sentimen negatif diberi warna merah. *Word Cloud* membuat informasi yang kompleks menjadi lebih mudah dipahami oleh audiens yang lebih luas, termasuk mereka yang mungkin tidak memiliki latar belakang teknis.



Gambar 4. Wordcloud Sentimen

*Splitting* data (membagi data) bertujuan untuk memisahkan dataset menjadi dua bagian utama data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk melatih model, sementara data uji digunakan untuk mengevaluasi kinerja model. Dengan membagi data, kita dapat memastikan bahwa model kita diuji pada data yang tidak pernah dilihat sebelumnya, memberikan gambaran yang lebih akurat tentang bagaimana model akan berkinerja pada data baru.

*Dataset* berisi 417 sampel:

- a. Menghitung 20% dari dataset untuk data uji:

Jumlah data uji:  $417 \times 0.2 = 83.4417 \times 0.2 = 83.4417 \times 0.2 = 83.4$  (Dibulatkan menjadi 83 sampel).

b. Menghitung 80% dari dataset untuk data latih:

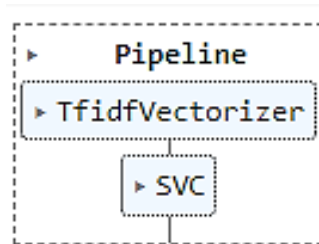
Jumlah data latih:  $417 \times 0.8 = 333.6417 \times 0.8 = 333.6417 \times 0.8 = 333.6$  (Dibulatkan menjadi 334 sampel).

Menggunakan `random_state = 42`: `random_state` memastikan bahwa kita mendapatkan pembagian yang sama setiap kali kode dijalankan. Tanpa *random\_state*, hasil pembagian dapat bervariasi setiap kali kita menjalankan kode.

### 3.2 Pelatihan Model Support Vector Machine

*Pipeline* yang digunakan untuk melatih model *machine learning*.

1. *Input Data*, yakni dengan melakukan penginputan data teks yang tidak terstruktur diberikan.
2. *TfidfVectorizer*, yang kemudian diubah menjadi fitur numerik menggunakan *TF-IDF*.
3. *SVC*, kemudian digunakan untuk melatih model *Support Vector Machine*.
4. *Output*, berupa model yang terlatih dapat digunakan untuk membuat prediksi berdasarkan fitur yang dihasilkan oleh *TfidfVectorizer*.



**Gambar 5.** Melatih model *Support Vector Machine*

Setelah melakukan pelatihan menggunakan algoritma *Support Vector Machine* selanjutnya menguji model dengan mencetak laporan klasifikasi dan menampilkan matriks maka akan diperoleh hasil klasifikasi sentimen yang berupa label sentimen. Label hasil klasifikasi yang dihasilkan akan dibandingkan dengan label sebenarnya sehingga akan diketahui nilai *accuracy*, *precision*, *recall* dan *f1-score* dari model yang digunakan terhadap dataset

Secara keseluruhan nilai-nilai diatas dapat disajikan dalam *classification report*. Berikut disajikan *classification report* dari proses analisis dengan menggunakan metode *Support Vector Machine*.

Classification report (SVM):				
	precision	recall	f1-score	support
negatif	0.58	0.76	0.66	46
positif	0.54	0.34	0.42	38
accuracy			0.57	84
macro avg	0.56	0.55	0.54	84
weighted avg	0.56	0.57	0.55	84

**Gambar 6.** *Classification Report Support Vector Machine*

Pada Gambar maka dapat dihitung seberapa besar nilai *accuracy*, *precision*, *recall*, *f1-score*:

$$Accuracy = \frac{35+13}{35+13+25+11} \times 100\% = 57\%$$

$$Precision = \frac{35}{35+25} \times 100\% = 58\%$$

$$Recall = \frac{35}{35+11} \times 100\% = 76\%$$

$$F1-score = \frac{2 \times 0.58 \times 0.76}{0.58 + 0.76} \times 100\% = 66\%$$

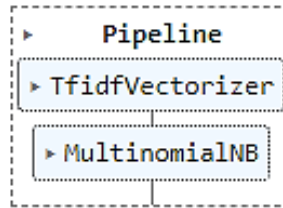
Dari hasil perhitungan diatas dapat dilihat bahwa jumlah data yang di dapat nilai *accuracy* sebesar 57%, *precision* sebesar 58%, *recall* sebesar 76%, *f1-score* sebesar 66%

### 3.3 Pelatihan Model Naïve Bayes

*Pipeline* yang digunakan untuk melatih model *machine learning*.

1. *Input Data*, yakni dengan melakukan penginputan data teks yang tidak terstruktur diberikan.
2. *TfidfVectorizer*, yang kemudian diubah menjadi fitur numerik menggunakan *TF-IDF*.
3. *SVC*, kemudian digunakan untuk melatih model *Naive Bayes*

4. *Output*, berupa model yang terlatih dapat digunakan untuk membuat prediksi berdasarkan fitur yang dihasilkan oleh *TfidfVectorizer*.



Gambar 7. Melatih model *Naïve Bayes*

Setelah melakukan pelatihan menggunakan algoritma *Naïve Bayes* selanjutnya menguji model dengan mencetak laporan klasifikasi dan menampilkan matriks maka akan diperoleh hasil klasifikasi sentimen yang berupa label sentimen. Label hasil klasifikasi yang dihasilkan akan dibandingkan dengan label sebenarnya sehingga akan diketahui nilai *accuracy*, *precision*, *recall* dan *f1-score* dari model yang digunakan terhadap dataset.

Classification report (Naive Bayes):				
	precision	recall	f1-score	support
negatif	0.59	0.76	0.67	46
positif	0.56	0.37	0.44	38
accuracy			0.58	84
macro avg	0.58	0.56	0.56	84
weighted avg	0.58	0.58	0.57	84

Gambar 8. *Classification Report Naïve Bayes*

Pada Gambar maka dapat dihitung seberapa besar nilai *accuracy*, *precision*, *recall*, *f1-score*:

$$Accuracy = \frac{35+14}{35+14+24+11} \times 100\% = 58\%$$

$$Precision = \frac{35}{35+24} \times 100\% = 59\%$$

$$Recall = \frac{35}{35+11} \times 100\% = 76\%$$

$$F1-score = \frac{2 \times 0,59 \times 0,76}{0,59+0,76} \times 100\% = 67\%$$

Dari hasil perhitungan diatas dapat dilihat bahwa jumlah data yang di didapat nilai *accuracy* sebesar 58%, *precision* sebesar 59% , *recall* sebesar 76%, *f1-score* sebesar 67%.

### 3.4 Hasil Kedua Model *Support Vector Machine* dan *Naïve Bayes*

*Naïve Bayes* lebih unggul dari pada *Support Vector Machine* bisa di lihat hasil dari melatih kedua model *Support Vector Machine* dan *Naïve Bayes* menggunakan *Pipeline* yang melibatkan *TfidfVectorizer*, kita dapat mengukur kinerja model dengan menggunakan berbagai *metrik* evaluasi seperti *accuracy*, *presisi*, *recall*, dan *F1-score*.

	Model	AccuracyScore	Precision	Recall	F1-Score
0	Support Vector Machine	0.571429	0.564484	0.571429	0.551343
1	Multinomial Naive Bayes	0.583333	0.578192	0.583333	0.566138

Gambar 9. Hasil *accuracy*, *presisi*, *recall*, dan *F1-score* untuk kedua model

Mengklasifikasikan teks *input* menggunakan dua model *machine learning* yang berbeda *Support Vector Machine* dan *Naïve Bayes*.

- Melatih dua model klasifikasi teks yang berbeda *Support Vector Machine* dan *Naïve Bayes*
- Mengklasifikasikan teks input menggunakan kedua model tersebut.
- Menampilkan hasil klasifikasi dari kedua model untuk perbandingan.
  - Klasifikasi positif  
Input Teks Program  
Contohnya : “busnya keren full ac nya dingin”.

Masukkan teks yang ingin diklasifikasi: busnya keren full ac nya dingin  
Input kata: busnya keren full ac nya dingin

Hasil Klasifikasi:  
Support Vector Machine: positif  
Multinomial Naive Bayes: positif

Gambar 10. Hasil Klasifikasi Teks Positif

b. Klasifikasi Negatif

Input Teks Program

Contohnya : “sebelum membeli busnya buat dulu jalurnya udah banyak kecelakaan”.

Masukkan teks yang ingin diklasifikasi: sebelum membeli busnya buat dulu jalurnya udah banyak kecelakaan  
Input kata: sebelum membeli busnya buat dulu jalurnya udah banyak kecelakaan

Hasil Klasifikasi:  
Support Vector Machine: negatif  
Multinomial Naive Bayes: negatif

Gambar 11. Hasil Klasifikasi Teks Negatif

## 4 KESIMPULAN

Berdasarkan hasil dari analisis sentimen pada ulasan pengguna terhadap Bus Listrik di kota Medan menggunakan algoritma *Support Vector Machine* dan *Naive Bayes*, dapat ditarik kesimpulan bahwa dengan menggunakan dataset yang diperoleh dengan mengumpulkan komentar dari *Instagram* terkait Bus Listrik di kota Medan. Proses pengumpulan data dilakukan dengan teknik *scraping* menggunakan *library Python*, menghasilkan total 417 data. Setelah dilakukan *preprocessing*, dataset dibagi menjadi 338 data latih dan 83 data uji. Algoritma *Support Vector Machine* juga digunakan untuk analisis sentimen dengan hasil evaluasi menunjukkan nilai akurasi sebesar 57%, *precision* sebesar 58%, *recall* sebesar 76%, dan *f1-score* sebesar 66%. Sedangkan algoritma *Naive Bayes* juga digunakan untuk analisis sentimen dengan hasil evaluasi menunjukkan nilai akurasi sebesar 58%, *precision* sebesar 59%, *recall* sebesar 76%, dan *f1-score* sebesar 67%. Berdasarkan hasil perbandingan, algoritma *Naive Bayes* menunjukkan kinerja yang sedikit lebih baik dibandingkan dengan *Support Vector Machine* dalam analisis sentimen ini. Meskipun demikian, kedua algoritma memberikan *insight* yang berharga tentang sentimen masyarakat Medan terhadap Bus Listrik. Hasil analisis menunjukkan bahwa sentimen positif dan negatif dari masyarakat terhadap Bus Listrik cukup beragam. Jumlah data latih yang digunakan dalam sistem berpengaruh signifikan terhadap akurasi prediksi. Selain itu, kualitas data latih juga berperan penting karena semakin tinggi kualitas data, sistem akan mendapatkan kosakata yang lebih kaya sehingga lebih akurat dalam memprediksi kelas sentimen.

## REFERENCES

- [1] M. R. Pratama, “Tinjauan Lokasi Halte Bus Trans Metro Deli Di Koridor 5 Medan Lapangan Merdeka – Tembung Terhadap Naik Turun Penumpang Bus Trans Metro Deli,” vol. 1, 2021.
- [2] I. W. A. Agustin, V. Gloria, and I. R. D. Ari, “Prioritas Peningkatan Kinerja Bus Listrik Metrotrans Rute 7A Kampung Rambutan – Lebak Bulus Jakarta,” *STTI*, vol. 1, no. 1, pp. 1–17, May 2024, doi: 10.47134/stti.v1i1.2418.
- [3] A. Novantirani, M. Sabariah Kania, and V. Effendi, “Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine,” *e-Proceeding of Engineering*, vol. 2, 2018.
- [4] M. Rifadh and R. S. M. Sihombing, “KAPASITAS DINAS PERHUBUNGAN KOTA MEDAN DALAM PENGENDALIAN SARANA DAN PRASARANA BUS TRANS METRO DELI DI KOTA MEDAN,” *JSSR*, vol. 6, no. 1, p. 174, Feb. 2023, doi: 10.54314/jssr.v6i1.1187.
- [5] M. Furqan, M. Ikhsan, and R. Aini, “Algoritma Support Vector Machine Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Pandemi Virus Corona Di Media Sosial,” vol. 4, no. 4, 2023.
- [6] A. H. Hasugian, I. Rusydi, and M. Ramadhani, “Penerapan Algoritma Naive Bayes Classifier Untuk Mengukur Tingkat Kepuasan Pasien,” vol. 6, 2023.
- [7] A. A. Arif, M. Firdaus, and Y. Maruhawa, “Comparison of Data Mining Methods for Prediction of Rainfall with C4.5, Naive Bayes, and KNN Algorithm,” 2022.
- [8] D. S. O. Panggabean, E. Buulolo, and N. Silalahi, “Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda,” *Jur. Ris. Kom.*, vol. 7, no. 1, p. 56, Feb. 2020, doi: 10.30865/jurikom.v7i1.1947.
- [9] R. Bahtiar, “Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor,” 2023.
- [10] A. Firdaus and W. I. Firdaus, “Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan),” vol. 13, no. 1, 2021.

- [11] A. H. Lubis, L. P. A. Lubis, and Sriani, "Sentiment analysis on twitter about the death penalty using the support vector machine method," *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, vol. 11, no. 2, pp. 312–321, 2024, doi: 10.37373.
- [12] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *JTIK*, vol. 8, no. 1, p. 147, Feb. 2021, doi: 10.25126/jtiik.0813944.
- [13] S. Rahayu and J. J. Purnama, "KLASIFIKASI KONSUMSI ENERGI INDUSTRI BAJA MENGGUNAKAN TEKNIK DATA MINING," *JTI*, vol. 16, no. 2, p. 395, Jul. 2022, doi: 10.33365/jti.v16i2.1984.
- [14] R. Husna El, R. Wasono, and M. Al Haris, "Analisis Sentimen Pada Twitter Mengenai Netflix Diblokir Telkom Menggunakan Metode Support Vector Machine," *Seminar Nasional Variansi*, pp. 214–222, 2020.
- [15] S. Hilda Kusumahadi, H. Junaedi, and J. Santoso, "Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine," *jpit*, vol. 4, no. 1, pp. 54–60, Jan. 2019, doi: 10.30591/jpit.v4i1.1125.
- [16] A. M. Zaki, M. S. Asih, and K. Chiuloto, "SISTEM PENDUKUNG KEPUTUSAN UNTUK PENERIMAAN TENAGA KERJA MENGGUNAKAN METODE NAÏVE BAYES STUDI KASUS PT. ALFATHI BERKAH MULIA," 2020.
- [17] S. D. Pramukti, A. Nugroho, and A. S. Sunge, "Analisis Sentimen Masyarakat Dengan Metode Naïve Bayes dan Particle Swarm Optimization," *tc*, vol. 21, no. 1, pp. 61–74, Feb. 2022, doi: 10.33633/tc.v21i1.5332.
- [18] A. Guswandri and L. Andraini, "Menerapkan Analisis Sentimen Metode Naïve Bayes dan svm," vol. 2, 2021.
- [19] Suparyanto, "Klasifikasi Kepuasan Layanan Akademik Di STMIK El Rahma Menggunakan Metode Algoritma Naive Bayes," *FAHMA*, vol. 20, no. 2, pp. 100–111, May 2022, doi: 10.61805/fahma.v20i2.37.
- [20] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging," *RESTI*, vol. 5, no. 3, pp. 504–510, Jun. 2021, doi: 10.29207/resti.v5i3.3067.
- [21] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *tekno*, vol. 10, no. 2, pp. 176–184, Jul. 2023, doi: 10.37373/tekno.v10i2.419.
- [22] H. Tuhuteru, "Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine," 2020.
- [23] T. Hutapea Ediyantama and K. Dewi, "Penerapan Metode Support Vector Machine untuk Sistem Tanya Jawab Pada Kasus Front Office," *UNIKOM Press*, 2022.
- [24] M. Ikhsan and R. Kurniawan, "Penerapan Text Mining pada Sistem Rekomendasi Pembimbing Skripsi Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier," *Indonesian Journal of Computer Science*, vol. 12, no. 6, 2023.