



Analisis Sentimen Kepuasan Pengguna Lintas Rel Terpadu (LRT) menggunakan Metode Support Vector Machine

Rangga Febri Kasih^{1*}, Rendra Gustriansyah², Zaid Romegar Mair³

¹⁻³Program Studi Teknik Informatika, Universitas Indo Global Mandiri, Indonesia

Email: 2021110014@students.uigm.ac.id^{1*}, rendra@uigm.ac.id², zaidromegar@uigm.ac.id³

Alamat: Jl. Jend. Sudirman Km.4 No. 62, 20 Ilir D. IV, Kec. Ilir Tim. I, Kota Palembang, Sumatera Selatan 30129

**Korespondensi Penulis*

Abstract. This study aims to analyze public sentiment toward the Palembang LRT service by utilizing user reviews available on the Google Maps platform. Sentiment analysis was conducted to understand public perceptions of service quality, which can serve as a basis for decision-making in improving public transportation services. The method employed in this research is the Support Vector Machine (SVM) algorithm combined with Term Frequency-Inverse Document Frequency (TF-IDF) for word weighting, which classifies reviews into two sentiment categories: positive and negative. A total of 500 reviews were randomly selected as the dataset and processed through a text preprocessing stage, including data cleaning, tokenization, and stopword removal to enhance data quality. The SVM model was then evaluated using an 80:20 split for training and testing, achieving an accuracy of 91%, which indicates excellent performance in identifying sentiment patterns in the Indonesian language. The findings of this study confirm that SVM-based approaches are effective and reliable for sentiment analysis in the context of public transportation. These results provide practical contributions for Palembang LRT management, as insights into public sentiment can be used as a strategic reference for decision-making, reputation management, and improving service quality based on user needs. Future research is recommended to expand the dataset, include neutral sentiment categories, and compare SVM performance with other machine learning algorithms to achieve more comprehensive and robust results.

Keywords: Confusion Matrix, LRT, Sentiment Classification, Support Vector Machine, TF-IDF.

Abstrak. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap layanan LRT Palembang dengan memanfaatkan data ulasan publik yang tersedia pada platform Google Maps. Analisis sentimen dilakukan untuk mengetahui persepsi masyarakat terhadap kualitas layanan transportasi publik ini, sehingga dapat digunakan sebagai dasar pengambilan keputusan dalam peningkatan pelayanan. Metode yang digunakan adalah algoritma Support Vector Machine (SVM) yang dikombinasikan dengan teknik pembobotan kata Term Frequency-Inverse Document Frequency (TF-IDF) untuk mengklasifikasikan ulasan ke dalam dua kategori sentimen, yaitu positif dan negatif. Sebanyak 500 ulasan dipilih secara acak sebagai data penelitian, kemudian melalui tahapan praproses teks, yang meliputi pembersihan data, tokenisasi, dan penghapusan kata-kata umum (stopwords) untuk meningkatkan kualitas data. Model SVM kemudian diuji menggunakan pembagian data 80% untuk pelatihan dan 20% untuk pengujian, menghasilkan akurasi sebesar 91%, yang menunjukkan performa sangat baik dalam mengenali pola sentimen pada teks berbahasa Indonesia. Hasil penelitian ini membuktikan bahwa pendekatan berbasis SVM efektif dan dapat diandalkan untuk analisis sentimen dalam konteks transportasi publik. Temuan ini memberikan kontribusi praktis bagi pengelola LRT Palembang, karena informasi sentimen masyarakat dapat menjadi acuan strategis dalam pengambilan keputusan, pengelolaan reputasi, dan peningkatan kualitas layanan berbasis kebutuhan pengguna. Ke depan, penelitian ini direkomendasikan untuk dikembangkan dengan menambahkan lebih banyak data, mempertimbangkan aspek sentimen netral, dan membandingkan kinerja SVM dengan algoritma pembelajaran mesin lainnya untuk memperoleh hasil yang lebih komprehensif.

Kata kunci: Confusion Matrix, Klasifikasi Sentimen, LRT, Support vector machine, TF-IDF.

1. LATAR BELAKANG

Di era teknologi informasi yang semakin maju, aktivitas masyarakat dalam mengakses dan membagikan informasi mengalami peningkatan yang sangat signifikan. Perkembangan media digital telah membuka ruang bagi publik untuk menyampaikan pendapat, pengalaman, maupun keluhan secara terbuka melalui berbagai platform daring. Data digital berupa opini dan

komentar ini menjadi salah satu sumber informasi penting yang dapat diolah untuk menggambarkan persepsi masyarakat terhadap layanan publik maupun sektor lainnya(Pameka et al., 2024).

Seiring meningkatnya jumlah data opini yang tersebar secara daring, diperlukan pendekatan yang tepat untuk mengidentifikasi pola sentimen yang terkandung di dalamnya. Analisis sentimen menjadi salah satu metode yang relevan digunakan untuk memahami kecenderungan opini masyarakat, baik dalam bentuk penilaian positif maupun negatif(Rohim et al., 2023). Dengan adanya analisis sentimen, data teks yang awalnya bersifat tidak terstruktur dapat diubah menjadi informasi yang lebih terarah, sehingga dapat memberikan gambaran mengenai tingkat kepuasan, kebutuhan, serta potensi perbaikan layanan di masa mendatang(R.D. Shafitri, 2025). Proses ini pada akhirnya diharapkan mampu mendukung pengambilan keputusan berbasis data, khususnya dalam meningkatkan mutu pelayanan publik di berbagai sektor.

LRT Palembang merupakan sistem transportasi massal berbasis rel yang beroperasi di Kota Palembang, Sumatera Selatan. LRT ini resmi beroperasi pada 1 Agustus 2018 dan menjadi moda transportasi modern pertama di Indonesia yang menghubungkan pusat kota dengan Bandara Internasional Sultan Mahmud Badaruddin II(Magdalena & Akustia, 2021). Dibangun sebagai bagian dari persiapan Asian Games 2018, LRT Palembang bertujuan untuk mengurangi kemacetan serta menyediakan transportasi yang lebih efisien dan ramah lingkungan bagi masyarakat. Dengan jalur sepanjang 23,4 km dan 13 stasiun yang tersebar di berbagai titik strategis. Meskipun demikian, Minat Masyarakat menggunakan LRT masih terbilang rendah(Putri & Sahara, 2023). Selain itu, faktor tarif, kenyamanan, serta frekuensi perjalanan juga menjadi aspek yang berpengaruh terhadap minat masyarakat dalam menggunakan LRT sebagai pilihan utama transportasi harian(Ade Rizki Ananda, 2022).

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi berbasis pembelajaran mesin untuk menganalisis sentimen kepuasan pengguna LRT Palembang menggunakan metode *Support vector machine* (SVM). Metode SVM dipilih karena kemampuannya dalam mengolah dataset berukuran besar secara efektif, termasuk ketika menghadapi jumlah data yang banyak, yaitu 523 data (Dwinanto et al., 2024). SVM mampu mengatasi masalah data yang tidak seimbang dan bekerja dengan baik dalam menentukan pola sentimen dari berbagai opini (Akmal & Kurniasih, 2023). Untuk meningkatkan performa klasifikasi, penelitian ini menerapkan teknik pemrosesan teks, seperti pembersihan data, tokenisasi, dan reduksi dimensi untuk menghilangkan informasi yang kurang relevan dalam dataset ulasan pengguna(Lambang et al., 2024). Dengan penerapan metode yang tepat, diharapkan sistem ini dapat memberikan

wawasan yang lebih akurat mengenai kepuasan pengguna terhadap layanan LRT Palembang serta membantu pengelola dalam meningkatkan kualitas layanan berdasarkan opini masyarakat.

Penelitian sebelumnya oleh Nur Arafah, menggunakan *Improved K-Nearest Neighbor* (K-NN) sebagai metode klasifikasi sentimen, sedangkan penelitian ini akan menerapkan SVM untuk mengeksplorasi apakah SVM lebih unggul dalam analisis sentimen kepuasan pengguna LRT Palembang. Selain perbedaan metode, penelitian sebelumnya mengumpulkan data dari Twitter dan kuesioner, sementara penelitian ini akan menggunakan ulasan *Google maps* sebagai sumber data utama. Hasil penelitian sebelumnya menunjukkan bahwa metode *Improved K-NN* mencapai akurasi tertinggi sebesar 74,07% dengan konfigurasi 90% data latih dan 10% data uji, serta akurasi terendah 63,04% pada konfigurasi 50% data latih dan 50% data uji(Ridho & Buchari, 2023).

Penelitian sebelumnya oleh Sahfitri, menggunakan metode *ServQual* dan *Customer Satisfaction Index (CSI)* untuk mengukur tingkat kepuasan pengguna terhadap layanan LRT Palembang(Ainul Wildan et al., 2021). Hasil penelitian menunjukkan bahwa meskipun pengguna merasa puas secara keseluruhan, masih terdapat beberapa atribut pelayanan yang memiliki nilai gap negatif, seperti kehandalan dan ketanggapan petugas, yang perlu diperbaiki untuk meningkatkan kualitas layanan. Berbeda dengan penelitian tersebut yang bersifat kuantitatif dan menggunakan kuesioner langsung sebagai sumber data, penelitian ini akan menerapkan metode analisis sentimen berbasis machine learning dengan algoritma SVM, serta memanfaatkan ulasan publik dari *Google maps* sebagai data utama.

2. KAJIAN TEORITIS

A. Analisis Sentimen

Analisis sentimen adalah cabang dari *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi dan mengkategorikan opini atau emosi dalam teks, seperti ulasan atau komentar, menjadi sentimen positif, atau negatif. Teknik ini memanfaatkan algoritma pembelajaran mesin dan pembelajaran mendalam untuk mengolah data teks yang tidak terstruktur. Analisis sentimen telah mengalami perkembangan signifikan dalam NLP, namun masih menghadapi tantangan, terutama dalam memahami konteks dan ironi dalam teks(Jim et al., 2024). Dalam konteks transportasi publik, menerapkan analisis sentimen pada data media sosial untuk mengevaluasi persepsi masyarakat terhadap transportasi multimoda selama pandemi COVID-19, yang memberikan wawasan berharga bagi

perencana transportasi(Chen et al., 2023). Selain itu, penelitian oleh (Pagliara & Cutolo, 2025) menggunakan teknik analisis sentimen lanjutan untuk menilai persepsi pengguna terhadap layanan kereta api di beberapa negara Eropa, membantu perusahaan kereta api dalam meningkatkan kualitas layanan . Dengan demikian, analisis sentimen menjadi alat penting dalam memahami opini publik dan mendukung pengambilan keputusan di berbagai sektor.

B. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF adalah salah satu algoritma representasi teks yang banyak digunakan dalam pemrosesan bahasa alami (*Natural Language Processing / NLP*). Tujuannya adalah untuk mengubah dokumen teks ke dalam bentuk numerik atau vektor agar dapat diproses lebih lanjut oleh algoritma pembelajaran mesin. Metode ini bekerja dengan menilai pentingnya suatu kata dalam sebuah dokumen tertentu yang berada dalam kumpulan dokumen atau korpus. TF-IDF terdiri dari dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF), yang bekerja secara bersamaan untuk memberikan bobot terhadap setiap kata dalam dokumen.

C. Machine Learning

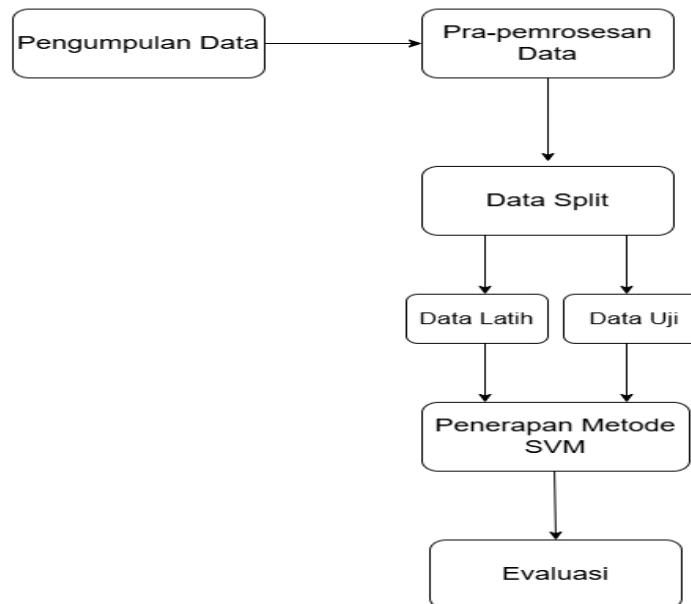
Berdasarkan penjelasan IBM, *machine learning* adalah bagian dari kecerdasan buatan (*Artificial Intelligence*) serta ilmu komputer yang memanfaatkan data dan algoritma untuk meniru proses pembelajaran pada manusia. Metode ini mampu meningkatkan akurasi seiring dengan berjalannya waktu. Kualitas keputusan yang dihasilkan akan semakin baik apabila algoritma *machine learning* yang diterapkan semakin optimal. Terdapat beberapa jenis algoritma *machine learning* dengan fungsi dan tujuan yang berbeda-beda, di antaranya yang paling umum digunakan adalah *supervised learning*, *unsupervised learning*, dan *semi-supervised learning*.(Puspasari et al., 2022).

D. Klasifikasi

Klasifikasi adalah proses dalam pembelajaran mesin (*machine learning*) yang digunakan untuk mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan kesamaan fitur atau karakteristik antar data(Gustriansyah et al., 2024). Proses ini dilakukan dengan membangun model klasifikasi dari data latih yang telah diberi label, kemudian digunakan untuk memprediksi kelas dari data baru yang belum diketahui labelnya (Alnuaimi & Albaldawi, 2024). Teknik klasifikasi memiliki peranan penting dalam berbagai aplikasi seperti diagnosis penyakit, deteksi email spam, pengenalan tulisan tangan, dan analisis opini publik (Fahtu Rahman et al., 2024).

3. METODE PENELITIAN

Metode yang diterapkan dalam penelitian ini mencakup beberapa langkah sistematis untuk memastikan hasil yang diperoleh akurat. Proses penelitian dimulai dari tahap pengumpulan data sampai pada tahap pengujian klasifikasi. Gambaran lengkap tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian.

A. Pengumpulan Data

Dalam proses analisis sentimen ini, metode SVM digunakan sebagai alat utama untuk mengklasifikasikan opini-opini yang telah dikumpulkan. Dalam penelitian ini, data yang dimanfaatkan berupa kumpulan ulasan sentimen publik mengenai layanan LRT Palembang, yang diperoleh melalui platform *Google maps*. Sebanyak 500 komentar berhasil dikumpulkan secara acak dengan bantuan perangkat *Instant Data Scraper*.

B. Pra-Pemrosesan Data

Untuk mempermudah pengelolaan data, ulasan yang telah dikumpulkan akan melalui tahap pra-pemrosesan(Nazori Suhandi et al., 2025).

1) *Cleaning*

Proses *cleaning* dilakukan dengan menghapus karakter-karakter yang dianggap tidak penting, seperti simbol dan huruf di luar alfabet a-z, termasuk tanda baca, URL, *hashtag*, serta *username* yang terdapat pada data(Khoirunnisa & Topiq, 2024).

2) Case Folding

Pada langkah ini, teks dengan huruf kapital diubah menjadi huruf kecil. *Case folding* merupakan bagian dari tahap pra-pemrosesan yang berfungsi untuk menyamakan bentuk penulisan karakter dalam data(Nurrochmah et al., 2025).

3) Tokenize

Pada tahap ini, setiap kata dalam sebuah kalimat pada dokumen dipisahkan satu per satu. Pemisahan kata umumnya dilakukan dengan menggunakan spasi sebagai pemisah. Proses tokenisasi memecah kalimat menjadi unit kata secara terstruktur(Dahlia Winingsih, 2023).

4) Stemming

Stemming merupakan proses yang mengubah setiap kata dalam dokumen menjadi bentuk dasar atau akar katanya (*root word*). Proses *Stemming* dilakukan untuk mengembalikan kata ke bentuk dasar agar variasi kata yang memiliki akar kata sama dapat dikelompokkan. Penggunaan algoritma *Stemming* yang tepat sangat berpengaruh pada keakuratan pembobotan kata(Albab et al., 2023).

5) Stopword Removal

Pada tahap ini, dilakukan proses penyaringan terhadap kata-kata yang memiliki frekuensi kemunculan yang sangat tinggi maupun sangat rendah, yang dikenal dengan istilah *Stopword Removal*. Proses ini bertujuan untuk menghapus kata-kata yang dianggap kurang memberikan makna penting terhadap analisis, seperti kata hubung atau kata umum lainnya menggunakan library Stopword (*Indonesian*)(Angelina et al., 2023).

6) Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF merupakan metode yang sering digunakan untuk mengukur bobot kata dalam suatu teks. Pendekatan ini dikenal karena efisiensinya, kesederhanaannya, serta akurasinya dalam menghitung *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*. Tahap pembobotan TF-IDF menghasilkan nilai bobot yang menunjukkan seberapa penting suatu kata dalam sebuah dokumen(Nurhaliza Agustina et al., 2024).

C. Data Split

Setelah proses pengolahan data, tahapan berikutnya adalah Pembagian Data. Pembagian data ini bertujuan untuk memisahkan dataset menjadi dua bagian, yaitu data latih dan data uji dengan perbandingan 80:20. Dimana 80% di gunakan sebagai data latih dan 20% digunakan sebagai data uji

D. Penerapan Metode *Support Vector Machine*

SVM termasuk dalam metode *supervised learning* yang sering dimanfaatkan baik untuk klasifikasi maupun regresi. Cara kerja SVM yaitu dengan menentukan *hyperplane* terbaik yang berfungsi sebagai garis atau bidang pemisah antara dua kelas data (Naa & Yuniar Rahman, 2024) dengan margin maksimum. Margin yang lebih luas menunjukkan potensi model dalam mengklasifikasikan data baru secara lebih akurat. Pada klasifikasi linier, tujuan SVM adalah menemukan *hyperplane* yang dapat memisahkan dua kelas dengan margin optimal. Adapun persamaan *hyperplane* tersebut adalah:

$$\mathcal{F}(x) = \sum(w_i * x_i) + b = 0 \quad (2.1)$$

Keterangan :

$\mathcal{F}(x)$: Output atau hasil prediksi dari model

w_i : Bobot, diperoleh dari proses optimasi model SVM

x_i : Fitur dari data latih (biasanya merupakan *support vector*)

b : Bias atau intercept, nilai koreksi agar model lebih akurat

E. Evaluasi

Evaluasi terhadap performa model SVM dilakukan menggunakan *Confusion Matrix* setelah proses klasifikasi selesai. *Confusion Matrix* digunakan untuk menghitung berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score, yang secara keseluruhan memberikan gambaran mengenai kemampuan model dalam mengelompokkan data secara tepat (Gustriansyah et al., 2024).

True Positive (TP) menggambarkan jumlah data yang diklasifikasikan dengan benar ke dalam kelas positif. Sebaliknya, *False Positive* (FP) terjadi ketika model secara keliru mengklasifikasikan data negatif sebagai positif, dikenal juga sebagai kesalahan Tipe I. Sementara itu, *False Negative* (FN) menunjukkan kasus di mana data positif diklasifikasikan sebagai negatif oleh model, yang merupakan kesalahan Tipe II dan berpotensi mengakibatkan terlewatnya deteksi penting, seperti pada diagnosis penyakit. Adapun *True Negative* (TN) adalah jumlah data negatif yang berhasil dikenali secara akurat oleh model. Dari *Confusion Matrix* inilah, sejumlah indikator performa model dapat dihitung

untuk mengetahui sejauh mana efektivitas model dalam klasifikasi. Tabel *Confusion Matrix* dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix.*

<i>Confusion Matrix</i>	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	TP	FN
<i>Actual Negative</i>	FP	TN

dimana :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (2.4)$$

$$Precision = \frac{TP}{TP+FP} \quad (2.5)$$

$$recall = \frac{TP}{TP+FN} \quad (2.6)$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = \frac{2 \cdot TP}{2 \cdot TP + FN + FP} \quad (2.7)$$

4. HASIL DAN PEMBAHASAN

Pada bagian ini dipaparkan hasil penerapan algoritma SVM dalam proses klasifikasi data. Tahapan dimulai dengan pra-pemrosesan untuk memudahkan pengolahan data, kemudian dilanjutkan pada proses klasifikasi menggunakan SVM. Kinerja hasil klasifikasi selanjutnya dievaluasi melalui *Confusion Matrix* dengan mengacu pada metrik akurasi, presisi, recall, serta F1-Score.

A. Pra-pemrosesan Data

Ulasan yang terkumpul akan diolah melalui tahap *preprocessing* untuk mempermudah pengelolaan data. Tahap ini melibatkan serangkaian proses, seperti pembersihan simbol atau karakter yang tidak diperlukan (*cleaning*), penyeragaman teks dengan mengubah semua huruf menjadi kecil (*case folding*), pemecahan kalimat ke dalam unit kata (*tokenizing*), pengembalian kata ke bentuk asalnya (*stemming*), serta penghapusan kata-kata umum yang tidak relevan bagi analisis (*stopword removal*).

1) *Cleaning*

Berdasarkan jumlah opini yang telah dikumpulkan sebanyak 500 ulasan, tahap selanjutnya dalam pengolahan opini terhadap LRT Palembang di internet adalah proses

pembersihan data atau data *Cleaning*. Tahapan ini bertujuan untuk menghapus karakter-karakter yang tidak diperlukan seperti tanda baca dan simbol khusus. Tabel 2 dan Tabel 3 menggambarkan kondisi data sebelum dan sesudah pembersihan.

Tabel 1. Sebelum di *Cleaning*.

No	Ulasan
1	Rapi dan pelayanannya oke..
2	Pelayanan customer servicenya jg oke 🌟
3	Tidak ada petugas yang bantu lansia 😊
...	...
498	Sering telat jadwalnya, kesel
499	Petugasnya ramah dan sangat membantu..stasiunnya bersih
500	Nyaman dan petugas cukup komunikatif, ramah dan sopan..

Tabel 2. Sesudah di *Cleaning*.

No	Ulasan
1	Rapi dan pelayanannya oke
2	Pelayanan customer servicenya jg oke
3	Tidak ada petugas yang bantu lansia
...	...
498	Sering telat jadwalnya kesel
499	Petugasnya ramah dan sangat membantu stasiunnya bersih
500	Nyaman dan petugas cukup komunikatif ramah dan sopan

2) *Case Folding*

Tahap *Case Folding* merupakan proses mengubah seluruh huruf dalam teks menjadi huruf kecil dengan tujuan untuk mengurangi variasi penulisan dari kata yang sama, baik yang ditulis dengan huruf kapital maupun huruf kecil. Tabel 4 menunjukkan kondisi data sesudah dilakukan proses ini.

Tabel 4. Sesudah di *Case Folding*.

No	Ulasan
1	rapi dan pelayanannya oke
2	pelayanan customer servicenya jg oke
3	tidak ada petugas yang bantu lansia
...	...
498	sering telat jadwalnya kesel
499	petugasnya ramah dan sangat membantu stasiunnya bersih
500	nyaman dan petugas cukup komunikatif ramah dan sopan

3) *Tokenize*

Pada tahap ini, dilakukan proses pemisahan setiap kata dalam sebuah kalimat pada dokumen. Pemisahan kata umumnya menggunakan spasi sebagai pemisah antar kata. Meskipun dalam praktiknya penulisan bisa bervariasi, tujuan utamanya adalah untuk memecah kalimat menjadi unit-unit kata penyusun. Tabel 5 menampilkan kondisi data sesudah dilakukan tokenisasi.

Tabel 5. Sesudah di *Tokenize*.

No	Ulasan
1	[‘rapi’, ‘dan’, ‘pelayanannya’. ‘oke’]
2	[‘pelayanan’, ‘customer’, ‘servicenya’, ‘jg’, ‘oke’]
3	[‘tidak’, ‘ada’, ‘petugas’, ‘yang’, ‘bantu’, ‘lansia’]
...	...
498	[‘sering’, ‘telat’, ‘jadwalnya’, ‘kesel’]
499	[‘petugasnya’, ‘ramah’, ‘dan’, ‘sangat’, ‘membantu’, ‘stasiunnya’, ‘bersih’]
500	[‘nyaman’, ‘dan’, ‘petugas’, ‘cukup’, ‘komunikatif’, ‘ramah’, ‘dan’, ‘sopan’]

4) *Stemming*

Pada tahap selanjutnya, dilakukan proses *Stemming*, yaitu mengubah setiap kata dalam kalimat menjadi bentuk kata dasar atau kata baku. Untuk tahap *Stemming*, di mana digunakan pustaka Python bernama Sastrawi yang berfungsi untuk pemrosesan bahasa alami dalam bahasa Indonesia. Tabel 6 menyajikan kondisi data sesudah dilakukan *Stemming*.

Table 6. Sesudah di *Stemming*.

No	Ulasan
1	['rapi', 'dan', 'layan', 'oke']
2	['layan', 'customer', 'service', 'jg', 'oke']
3	['tidak', 'ada', 'tugas', 'yang', 'bantu', 'lansia']
...	...
498	['sering', 'telat', 'jadwal', 'kesel']
499	['tugas', 'ramah', 'dan', 'sangat', 'bantu', 'stasiun', 'bersih']
500	['nyaman', 'dan', 'tugas', 'cukup', 'komunikatif', 'ramah', 'dan', 'sopan']

5) *Stopword Removal*

Pada tahap ini, dilakukan proses penyaringan kata-kata yang terlalu sering muncul maupun yang jarang digunakan, yang dikenal dengan istilah “*Stopword Removal*”. Tabel 7 menunjukkan data sesudah dilakukan proses ini.

Tabel 7. Sesudah di *Stopword Removal*.

No	Ulasan
1	rapi layan oke
2	layan customer service jg oke
3	tidak tugas bantu lansia
...	...
498	telat jadwal kesel
499	tugas ramah bantu stasiun bersih

500 nyaman tugas komunikatif ramah sopan

6) **Term Frequency-Inverse Document Frequency (TF-IDF)**

Dengan metode ini, kata-kata yang sering muncul tetapi hanya pada satu ulasan tidak akan mendapat bobot yang sama dengan kata-kata yang signifikan di berbagai ulasan. Hasil pembobotan kemudian diubah ke bentuk vektor numerik agar dapat diolah oleh algoritma klasifikasi SVM. Proses ini memastikan data teks dapat diinterpretasikan menjadi data angka tanpa kehilangan makna konteksnya. Penerapan Metode SVM. Pada Tabel 8 menunjukkan hasil pembobotan.

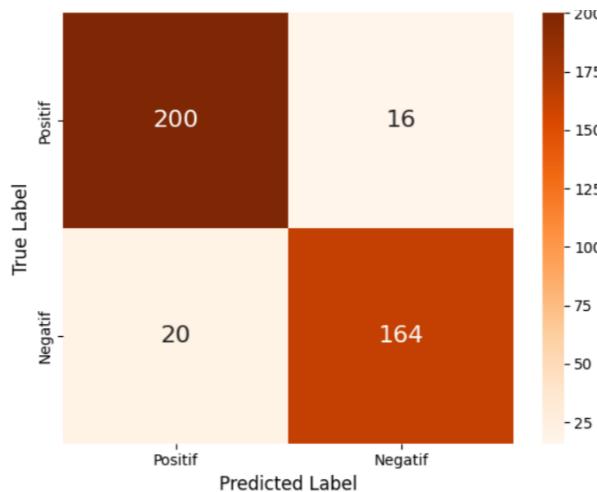
Tabel 8. Hasil Pembobotan.

No	sentimen	Label
1	{‘rapi’: 0.5234, ‘layan’: 0.3425, ‘oke’: 0.612}	Positif
2	{‘layan’: 0.3425, ‘customer’: 0.311, ‘service’: 0.103, ‘jg’: 0.021, ‘oke’: 0.612}	Positif
3	{‘tidak’:-0.4621, ‘tugas’: 0.2113, ‘bantu’: 0.2111, ‘lansia’: 0.01}	Negatif
...	...	
498	{‘telat’:-0.5461, ‘jadwal’: 0.0568, ‘kesel’:-0.3721}	Negatif
499	{‘tugas’: 0.2113, ‘ramah’: 0.5098, ‘bantu’: 0.2111, ‘stasiun’: 0.011, ‘bersih’: 0.2761}	Positif
500	{‘nyaman’: 0.4361, ‘tugas’: 0.2113, ‘komunikatif’: 0.145, ‘ramah’: 0.5098, ‘sopan’: 0.3251}	Positif

B. Penerapan Metode SVM

Dalam analisis sentimen, algoritma *Support vector machine* (SVM) dimanfaatkan untuk mengelompokkan opini atau teks ke dalam dua kategori sentimen, yaitu positif dan negatif. Pendekatan ini bekerja dengan cara memisahkan data menggunakan *hyperplane* yang paling optimal agar dapat membedakan setiap kelas sentimen secara jelas. SVM mulai diterapkan setelah data melalui serangkaian tahap pra-pemrosesan, seperti pembersihan data dan konversi teks ke bentuk numerik yang sesuai dengan kebutuhan pembelajaran mesin. Model SVM dilatih menggunakan data yang sudah diproses agar pembelajaran dapat berjalan lebih

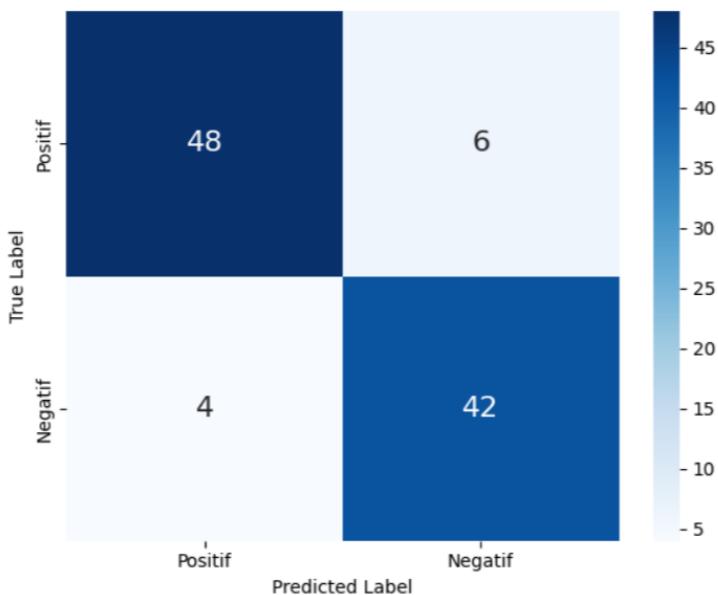
efektif. Tahap pelatihan dilakukan secara bertahap dan dapat diulang dengan penyesuaian parameter kernel agar sesuai dengan karakteristik data teks yang dianalisis. Pada Gambar 2 menunjukan *Confusion Matrix* dari hasil pelatihan menggunakan data latih.



Gambar 1. *Confusion Mattrix* Data Latih.

C. Evaluasi

Pada Gambar 3, *Confusion Matrix* untuk data latih terdiri dari 400 data sampel dengan hasil 42 True Negative (TN), 4 False positive(FP), 48 True positive(TP), dan 6 False Negative (FN). Hasil ini menunjukkan bahwa model SVM cukup mampu mengenali data yang benar-benar positif maupun negatif meskipun masih terdapat kesalahan klasifikasi. Nilai True positivesebanyak 48 berarti model berhasil mengklasifikasikan banyak data positif dengan benar. Namun, adanya False positivesebanyak 4 dan False Negative sebanyak 6 mengindikasikan masih terdapat kesalahan prediksi, baik dalam mendekripsi data positif maupun negatif. Sementara itu, jumlah True Negative yang cukup tinggi, yaitu 42, menunjukkan bahwa model memiliki kinerja yang baik dalam mengenali data negatif. Hasil ini membuktikan bahwa model SVM sudah memiliki kemampuan klasifikasi yang baik, namun tetap diperlukan optimasi lebih lanjut untuk mengurangi kesalahan klasifikasi dan meningkatkan akurasi secara keseluruhan.



Gambar 2. Confusion Matrix skema 80:20.

$$\text{Accuracy} = \frac{48 + 42}{48 + 6 + 4 + 42} = \frac{90}{100} = 0.9 = 90\%$$

$$\text{Precision} = \frac{48}{48 + 4} = \frac{48}{52} = 0.9231 = 92.31\%$$

$$\text{Recall} = \frac{48}{48 + 6} = \frac{48}{54} = 0.8889 = 88.9\%$$

$$F1 = 2 \times \frac{(0.9231) \times (0.8889)}{0.9231 + 0.8889} = 2 \times \frac{0.8206}{1.812} = 0.905 = 90.5\%$$

5. KESIMPULAN DAN SARAN

Dalam penelitian ini, metode *Support Vector Machine* (SVM) berhasil diimplementasikan untuk klasifikasi sentimen ulasan pengguna LRT Palembang yang diambil dari Google Maps. Dari total 500 ulasan yang digunakan, tahap pra-pemrosesan data seperti *cleaning*, *case folding*, *tokenizing*, *stemming*, *stopword removal*, serta pembobotan menggunakan TF-IDF mampu meningkatkan mutu data sebelum dilakukan klasifikasi.

Berdasarkan hasil evaluasi, model SVM memperoleh akurasi sebesar 91% pada pembagian data 80:20. Capaian ini menegaskan bahwa SVM cukup efektif dalam mengidentifikasi pola sentimen pada teks bahasa Indonesia. Nilai *True Positive* dan *True Negative* yang relatif tinggi juga menunjukkan kemampuan model dalam mendekripsi sentimen positif dan negatif, meski masih terdapat kesalahan klasifikasi pada *False Positive* dan *False Negative*.

DAFTAR REFERENSI

- Ade Rizki Ananda. (2022). *Kajian persepsi masyarakat terhadap rencana pembangunan LRT (Light Rail Transit) di Kota Medan.*
- Ainul Wildan, R. S., Adam Rajagede, R., & Rahmadi, R. (2021). Analisis sentimen politik berdasarkan big data dari media sosial YouTube: Sebuah tinjauan literatur. *Prosiding Automata*, 2(1).
- Akmal, R. A., & Kurniasih, A. (2023). Penerapan algoritma klasifikasi untuk menangani data tidak seimbang pada peningkatan kualitas siswa. *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 563.
- Albab, M. U., P., Y. K., & Fawaiq, M. N. (2023). Optimization of the stemming technique on text preprocessing President 3 periods topic. *Jurnal Transformatika*, 20(2), 1–12. <https://doi.org/10.26623/transformatika.v20i2.5374>
- Alnuaimi, A. F. A. H., & Albaldawi, T. H. K. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, 97, 1–24. <https://doi.org/10.1051/bioconf/20249700133>
- Angelina, S. J., Bijaksana, A., Negara, P., & Muhardi, H. (2023). Analisis pengaruh penerapan stopword removal pada performa klasifikasi sentimen tweet bahasa Indonesia. *JUARA (Jurnal Aplikasi dan Riset Informatika)*, 2(1), 165–173. <https://doi.org/10.26418/juara.v2i1.69680>
- Chen, X., Wang, Z., & Di, X. (2023). Sentiment analysis on multimodal transportation during the COVID-19 using social media data. *Information (Switzerland)*, 14(2), 1–12. <https://doi.org/10.3390/info14020113>
- Dahlia Winingsih, O. (2023). *Ekstraksi informasi metadata statistik pada artikel penelitian ilmiah menggunakan algoritma machine learning* (Program Studi Magister Teknik Elektro).
- Dwinanto, R. W., A, A. S. S., & Ardianto, R. (2024). Klasifikasi berisiko stunting pada balita: Perbandingan K-Nearest Neighbor, Naïve Bayes, Support Vector Machine. *Jurnal Manajemen Informatika dan Komputerisasi Akuntansi (JMIKA)*, 8(2), 264–273. <https://doi.org/10.46880/jmika.Vol8No2.pp264-273>
- Fahru Rahman, M. A., Mair, Z. R., & Sartika, D. (2024). Klasifikasi ulasan pelanggan Shopee Mall terhadap e-commerce penjualan baju batik metode Naïve Bayes. *IDEALIS: Indonesian Journal of Information System*, 7(2), 164–177. <https://doi.org/10.36080/idealis.v7i2.3178>
- Gustriansyah, R., Suhandi, N., Puspasari, S., & Sanmorino, A. (2024). Machine learning method to predict the toddlers' nutritional status. *Jurnal Infotel*, 16(1), 32–43. <https://doi.org/10.20895/infotel.v15i4.988>
- Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, 6(January), 100059. <https://doi.org/10.1016/j.nlp.2024.100059>
- Khoirunnisa, F., & Topiq, S. (2024). Masyarakat pada proses penegak hukum di Indonesia dengan menggunakan algoritma. *Jurnal Teknologi Informasi dan Terapan*, 12(3), 2128–2139. <https://doi.org/10.23960/jitet.v12i3.4683>

- Lambang, S., Pradana, S., Sains, F., & Teknologi, D. (2024). Analisis sentimen masyarakat media sosial Twitter terhadap kinerja penjabat gubernur DKI Jakarta menggunakan model IndoBERT. <https://repository.uinjkt.ac.id/dspace/handle/123456789/77071>
- Magdalena, M., & Akustia, W. (2021). Keterpaduan antarmoda transportasi untuk mendukung operasional LRT Kota Palembang. *Jurnal Transportasi Multimoda*, 19(1), 32–47. <https://doi.org/10.25104/mtm.v19i1.1858>
- Naa, R., & Yuniar Rahman, A. (2024). Klasifikasi motif kain batik Papua menggunakan metode multiclass support vector machine (M-SVM). *Jurnal Informatika dan Sistem Informasi*, 11(3), 421–429.
- Nazori Suhandi, R., Gustriansyah, R., & Destria, A. (2025). Klasifikasi penyakit TBC menggunakan metode UMAP dan K-NN. *Bit-Tech*, 7(3), 843–852. <https://doi.org/10.32877/bt.v7i3.2227>
- Nurhaliza Agustina, C. A., Novita, R., Mustakim, & Rozanda, N. E. (2024). The implementation of TF-IDF and Word2Vec on booster vaccine sentiment analysis using support vector machine algorithm. *Procedia Computer Science*, 234, 156–163. <https://doi.org/10.1016/j.procs.2024.02.162>
- Nurrochmah, D. S., Rahaningsih, N., Dana, R. D., & Rohmat, C. L. (2025). Penerapan algoritma Naive Bayes dalam analisis sentimen ulasan aplikasi KitaLulus di Google Play Store. *Jurnal Informatika Terpadu*, 11(1), 1–11. <https://doi.org/10.54914/jit.v11i1.1544>
- Pagliara, F., & Cutolo, G. (2025). Advanced sentiment analysis techniques to detect users' perceptions related to the "health" of a railway company: Some evidence from European countries. *Journal of Intelligent and Public Data*, 9(1), 1–16. <https://doi.org/10.24294/jipd9242>
- Pameka, A., Heriansyah, R., & Astuti, L. W. (2024). Optimalisasi feature selection untuk mendeteksi penyakit diabetes mellitus menggunakan metode decision tree. *JUPITER: Jurnal Penelitian dan Kajian Teknik Informatika*, 589–599.
- Puspasari, S., Ermatita, & Zulkardi. (2022). Machine learning for exhibition recommendation in a museum's virtual tour application. *International Journal of Advanced Computer Science and Applications*, 13(4), 404–412. <https://doi.org/10.14569/IJACSA.2022.0130448>
- Putri, N. H. A., & Sahara, S. (2023). Analisis penambahan sarana penunjang kegiatan LRT untuk kemudahan mobilitas masyarakat di wilayah Palembang. *Advanced in Social Humanities Research*, 1(2), 31–37. <https://doi.org/10.46799/adv.v1i12.147>
- R.D. Shafitri. (2025). *MyPertamina untuk produk bersubsidi berbasis*.
- Ridho, M. F., & Buchari, E. (2023). Transportasi Light Rail Transit (LRT) Palembang Sumatera Selatan berdampak lingkungan dan pengembangan usaha perkotaan sektor non fare box. *Bearing: Jurnal Penelitian dan Kajian Teknik Sipil*, 8(1), 39. <https://doi.org/10.32502/jbearing.v8i1.6268>
- Rohim, A., Haviz Irfani, M., Ramadhan, M., & Ubaidillah, U. (2023). Penerapan metode text mining dengan chatbot questions and answer pada PT PLN (Persero) Sumatera Selatan. *Klik - Jurnal Ilmu Komputer*, 4(2), 59–67. <https://doi.org/10.56869/klik.v4i2.551>