

## ANALISA SENTIMEN ULASAN APLIKASI TRANSPORTASI MENGUNAKAN METODE SVM DENGAN PENDEKATAN INSET LEXICON-BASED

Eryca Dhamma Shanty<sup>1</sup>, Brandon Alexander Jayadi<sup>2</sup>, Marco<sup>3</sup>, Viny Christanti Mawardi<sup>4</sup>

<sup>1</sup>Teknik Informatika, Universitas Tarumanagara Jakarta

Email: [eryca.535230071@stu.untar.ac.id](mailto:eryca.535230071@stu.untar.ac.id)

<sup>2</sup> Teknik Informatika, Universitas Tarumanagara Jakarta

Email: [brandon.535230088@stu.untar.ac.id](mailto:brandon.535230088@stu.untar.ac.id)

<sup>3</sup> Teknik Informatika, Universitas Tarumanagara Jakarta

Email: [marco.535230070@stu.untar.ac.id](mailto:marco.535230070@stu.untar.ac.id)

<sup>4</sup> Teknik Informatika, Universitas Tarumanagara Jakarta

Email: [viny@fti.untar.ac.id](mailto:viny@fti.untar.ac.id)\*

---

Masuk: 18-05-2025, revisi: 05-06-2025, diterima untuk diterbitkan: 07-06-2025

---

### ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap tiga aplikasi transportasi publik digital di Indonesia, yaitu Access by KAI, MyMRTJ, dan MitraDarat. Pendekatan yang digunakan dalam penelitian ini adalah kombinasi metode klasifikasi *Support Vector Machine* (SVM) dan pendekatan leksikal menggunakan kamus *INSET Lexicon-Based*. Data ulasan dikumpulkan melalui proses scraping dari *Google Play Store* dan kemudian melalui tahapan pre-processing, pelabelan sentimen, serta proses klasifikasi. Hasil penelitian menunjukkan bahwa mayoritas ulasan pengguna didominasi oleh sentimen positif. Access by KAI memperoleh ulasan positif terbanyak (446 ulasan), diikuti oleh MitraDarat (322 ulasan), dan MyMRTJ (286 ulasan). Model SVM memberikan hasil klasifikasi yang baik dengan akurasi tertinggi dicapai oleh MyMRTJ (87%), diikuti Access by KAI (84%), dan MitraDarat (82%). Selain itu, visualisasi word cloud berhasil menampilkan kata-kata dominan yang sering muncul dalam ulasan seperti "bagus", "mudah", dan "jalan", yang menunjukkan kepuasan pengguna terhadap layanan aplikasi. Pendekatan *INSET Lexicon-Based* terbukti efektif dalam mengenali polaritas kata dalam konteks lokal berbahasa Indonesia, serta meningkatkan akurasi pelabelan sebelum proses klasifikasi dilakukan. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan aplikasi transportasi digital berbasis data ulasan pengguna.

**Kata Kunci:** Analisis Sentimen; Aplikasi Transportasi; INSET Lexicon-Based; SVM; Word Cloud

### ABSTRACT

This study aims to analyze user sentiment in reviews of three public transportation applications in Indonesia: Access by KAI, MyMRTJ, and MitraDarat. The research approach combines the Support Vector Machine (SVM) classification method with a lexical-based approach using the INSET sentiment lexicon. User reviews were collected through a scraping process from the Google Play Store and then processed through pre-processing, sentiment labeling, and classification stages. The results show that most user reviews carry positive sentiment. Access by KAI received the highest number of positive reviews (446), followed by MitraDarat (322), and MyMRTJ (286). The SVM model yielded good classification performance, with the highest accuracy achieved by MyMRTJ (87%), followed by Access by KAI (84%), and MitraDarat (82%). Furthermore, the word cloud visualization highlighted frequently used words in the reviews, such as "bagus" (good), "mudah" (easy), and "jalan" (travel), indicating user satisfaction. The INSET Lexicon-Based approach proved effective in identifying word polarity within the local Indonesian language context and helped improve labeling accuracy before classification. This research is expected to contribute to the development of user-driven digital transportation applications.

**Keywords:** Sentiment Analysis; INSET Lexicon-Based; SVM; Transportation App; Word Cloud

## 1. PENDAHULUAN

### Latar Belakang

Kemajuan teknologi informasi di era digital telah memicu transformasi besar dalam berbagai aspek kehidupan masyarakat, termasuk dalam sektor transportasi. Layanan transportasi publik kini semakin mengintegrasikan teknologi digital guna meningkatkan efisiensi pelayanan dan kenyamanan pengguna (Sahara & Zetalia, 2024). Di Indonesia, sejumlah aplikasi transportasi seperti Access by KAI, MyMRTJ, dan MitraDarat dikembangkan sebagai bagian dari digitalisasi sistem transportasi nasional untuk merespons kebutuhan mobilitas masyarakat yang terus berkembang (Culio & Astuti, 2024).

Access by KAI merupakan aplikasi resmi dari PT Kereta Api Indonesia yang menyediakan layanan *e-ticketing* dari informasi perjalanan kereta api. MyMRTJ, yang dikembangkan oleh MRT Jakarta, menawarkan berbagai fitur pendukung perjalanan transportasi massal berbasis rel di wilayah ibu kota. Sementara itu, Mitra Darat merupakan aplikasi yang dirancang oleh Direktorat Jendral Perhubungan Darat untuk memberikan informasi mengenai angkutan umum darat serta layanan pendukung lainnya. Ketiga aplikasi ini menjadi representasi dari upaya pemerintah dan penyedia layanan transportasi untuk menghadirkan sistem transportasi yang adaptif dan berbasis teknologi (Iqrom, *et al.*, 2025).

Seiring dengan meningkatnya penggunaan aplikasi transportasi, pengguna secara aktif memberikan ulasan dan penilaian terhadap kualitas layanan yang diterima. Ulasan yang tersedia pada platform seperti Google Play Store mengandung informasi berharga mengenai pengalaman dan persepsi pengguna. Data ini, apabila dianalisis dengan tepat, dapat dimanfaatkan sebagai dasar evaluasi dan peningkatan kualitas layanan oleh pengembang maupun pemangku kepentingan (Satria, *et al.*, 2023). Dalam hal ini, analisis sentimen menjadi metode yang banyak digunakan untuk mengidentifikasi opini publik secara otomatis melalui pemrosesan teks. Analisis ini bertujuan untuk mengklasifikasi opini ke dalam kategori sentimen positif, negatif, atau netral, serta membantu penyedia layanan dalam memahami persepsi pelanggan dengan lebih terstruktur (Kholifah, *et al.*, 2024).

Untuk meningkatkan akurasi dalam proses klasifikasi sentimen, kombinasi antara pendekatan berbasis leksikon dan algoritma pembelajaran mesin telah banyak diteliti. Salah satu algoritma yang terbukti unggul dalam klasifikasi teks adalah Support Vector Machine (SVM), yang memiliki performa baik dalam menangani data dengan dimensi tinggi dan jumlah besar (Nugroho, *et al.*, 2024). Di sisi lain, pendekatan INSET *Lexicon-Based*, yaitu kamus sentimen Bahasa Indonesia, memungkinkan identifikasi polaritas kata dalam konteks lokal secara lebih tepat (Asiyah, *et al.*, 2025). Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna dari tiga aplikasi transportasi digital tersebut, dengan memanfaatkan pendekatan kombinatorik antara metode SVM dan leksikon INSET. Melalui pendekatan ini, diharapkan dapat diperoleh gambaran yang lebih objektif mengenai opini publik terhadap performa layanan aplikasi transportasi di Indonesia.

### Rumusan Masalah

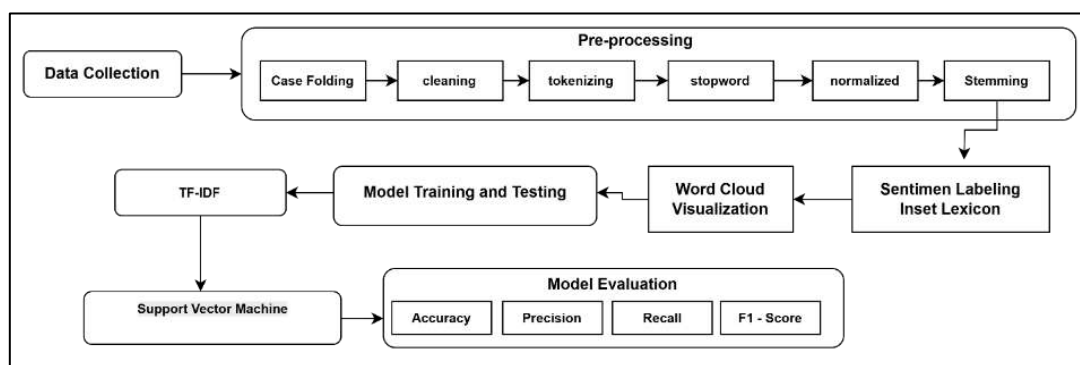
Berdasarkan latar belakang yang telah di kemukakan, maka penelitian ini berupaya menjawab beberapa permasalahan pokok sebagai berikut:

1. Bagaimana tanggapan dan persepsi pengguna terhadap tiga aplikasi transportasi digital, yaitu Access by KAI, MyMRTJ, dan MitraDarat, berdasarkan data ulasan yang dihimpun melalui platform Google Play Store?

2. Bagaimana pendekatan berbasis kamus INSET mampu mengidentifikasi dan mengklasifikasikan sentimen dalam ulasan berbahasa Indonesia secara akurat?
3. Bagaimana karakteristik dan distribusi sentimen yang muncul pada masing-masing aplikasi, serta bagaimana perbandingan kecenderungan sentimen positif dan negatif di antara ketiganya?
4. Bagaimana karakteristik dan distribusi sentimen yang muncul pada masing-masing aplikasi, serta bagaimana perbandingan kecenderungan sentimen positif dan negatif di antara ketiganya?

## 2. METODE PENELITIAN

Metode penelitian yang digunakan dilihat pada Gambar 1, yang dimulai dengan pengumpulan data, dilanjutkan dengan pra-proses data, pemberian label sentimen, visualisasi menggunakan *word cloud*, pelatihan dan pengujian model, perhitungan frekuensi kemunculan data, hingga evaluasi model setelah klasifikasi menggunakan SVM.



Gambar 1. Alur Penelitian

### Pengumpulan Data

Pada tahap pengumpulan data, data ulasan dikumpulkan dari beberapa aplikasi transportasi, yaitu Access by KAI, MyMRTJ, dan MitraDarat. Proses ini dimulai dengan mengakses halaman aplikasi di Google Play Store dan memperoleh data menggunakan ID aplikasi masing-masing. Selanjutnya, dilakukan proses *scraping* untuk mengambil ulasan yang diberikan oleh pengguna di platform tersebut. Setelah data terkumpul, langkah berikutnya adalah seleksi dan pemrosesan data menggunakan metode konkatenasi dengan memanfaatkan *library* Python Pandas untuk menggabungkan data yang relevan, khususnya yang berkaitan dengan atribut skor 1 dan 5.

### Pra-proses

Tahap awal pra-pemrosesan dimulai dengan pengumpulan data ulasan pengguna dari Google Play Store untuk tiga aplikasi transportasi, yaitu Access by KAI, MyMRTJ, dan MitraDarat. Proses pengumpulan dilakukan melalui teknik *scraping*, dilanjutkan dengan seleksi manual untuk menyaring hanya ulasan yang relevan. Seleksi ini bertujuan agar data yang dianalisis benar-benar mencerminkan opini pengguna terhadap layanan masing-masing aplikasi.

Setelah data terkumpul, dilakukan *case folding* untuk menyeragamkan seluruh huruf menjadi huruf kecil. Langkah ini penting agar variasi penulisan huruf kapital dan kecil tidak menimbulkan duplikasi makna (Putri & Cahyono, 2024). Selanjutnya, data dibersihkan dari unsur-unsur yang tidak relevan seperti karakter khusus, angka, *URL*, dan simbol menggunakan *library* regex serta

fungsi *str.maketrans*. Proses ini juga menghapus simbol seperti *mention* (@), *hashtag* (#), dan kata sambung yang tidak memiliki kontribusi signifikan terhadap analisis (Putri & Cahyono, 2024). Pembersihan ini dilakukan agar hasil analisis tidak dipengaruhi oleh elemen-elemen yang bersifat noise.

Langkah berikutnya adalah tokenisasi, yaitu memecah teks menjadi unit-unit kata atau token, yang mempermudah pengolahan data secara komputasional (Fathoni *et al.*, 2024). Setelah itu dilakukan penghapusan *stopword* kata-kata umum seperti "yang", "dan", dan "di" yang dianggap tidak membawa makna penting dalam konteks analisis sentimen. Proses ini dilanjutkan dengan normalisasi, yakni menyamakan kata tidak baku menjadi bentuk baku atau standar untuk menghindari kesalahan interpretasi akibat perbedaan ejaan atau singkatan (Fathoni *et al.*, 2024).

Tahapan selanjutnya adalah *stemming*, yaitu mengubah kata ke bentuk kata dasar dengan menghapus imbuhan atau akhiran. Tujuannya adalah menyederhanakan variasi kata yang memiliki makna serupa agar dapat diklasifikasikan sebagai satu entitas yang sama. Proses pra-pemrosesan ini diakhiri dengan analisis sentimen menggunakan TextBlob. *Tools* ini memanfaatkan pendekatan berbasis leksikon untuk menilai polaritas (positif atau negatif) dan subjektivitas suatu teks. Nilai polaritas menunjukkan kecenderungan emosi dalam ulasan, sedangkan subjektivitas menunjukkan sejauh mana ulasan tersebut bersifat opini atau fakta (Suanpang, *et al.*, 2021).

### ***Labeling Inset Lexicon***

Pada tahapan ini, pelabelan *dataset* dilakukan dengan mengimpor file yang berisi kumpulan dari kata yang telah memiliki bobot sentimen positif atau negatif. Proses ini menggunakan *lexicon* atau kamus sentimen, yang berfungsi untuk mengidentifikasi kata-kata dalam teks beserta nilai sentimen yang relevan. Setiap kata dalam dataset dianalisis dan diberi label sentimen berdasarkan kata-kata yang terdapat dalam kamus.

Dalam analisis sentimen, terutama untuk ulasan pengguna aplikasi Access by KAI, pelabelan dilakukan secara otomatis dengan menggunakan metode *auto modeling*, di mana kamus INSET leksikon yang diterapkan akan menentukan label sentimen untuk setiap ulasan. Pada pendekatan ini, *Supervised Learning* digunakan, yang membutuhkan kelas atau target untuk setiap *record* dalam *dataset*, sehingga model dapat mempelajari pola dan mengklasifikasikan ulasan ke dalam kategori sentimen yang sesuai (Fathoni, *et al.*, 2024).

### ***Word Cloud***

*Word Cloud* adalah visualisasi data menggambarkan frekuensi kemunculan kata-kata dalam sebuah kumpulan teks. Pada data visualisasi yang akan ditampilkan data kata yang paling besar akan ditampilkan dengan ukuran font yang lebih besar, sementara untuk kata-kata yang jarang muncul akan divisualisasikan dengan ukuran font yang lebih kecil.

### **TF – IDF**

TF-IDF merupakan metode yang digunakan untuk menilai seberapa penting sebuah kata dalam suatu dokumen ataupun korpus teks. Pendekatan ini menggabungkan dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). TF menghitung seberapa sering sebuah kata muncul dalam satu dokumen, yang dapat dinyatakan dengan rumus:

$$tf_{ij} = \frac{f_d(i)}{\max f_d(i)} \quad (1)$$

Di mana  $f_i$  menyatakan frekuensi kemunculan kata ke- $i$  dan  $\max f_i$  adalah jumlah kemunculan tertinggi dari suatu kata dalam dokumen tersebut.

Sementara itu, IDF mengukur tingkat kelangkaan sebuah kata diseluruh dokumen dalam korpus, yang diformulasikan sebagai:

$$idf(t, D) = \log\left(\frac{N}{df(t) + 1}\right) \quad (2)$$

Dengan  $N$  adalah jumlah total dokumen dalam korpus dan  $df(t)$  adalah banyaknya dokumen yang mengandung kata  $t$ .

Dengan mengkombinasikan kedua komponen tersebut, TF-IDF menghasilkan bobot yang lebih tinggi untuk kata-kata yang sering muncul di satu dokumen namun jarang ditemukan di dokumen lain. Pendekatan ini sangat efektif untuk mengekstraksi fitur-fitur penting dalam data teks, serta membantu mengurangi dominasi kata-kata yang bersifat umum dan kurang relevan dalam analisis. Pada penelitian ini, TF-IDF digunakan untuk mengonversi ulasan pengguna berbasis teks menjadi representasi numerik, sehingga dapat diproses lebih lanjut menggunakan algoritma klasifikasi seperti SVM (Yutika & Al Faraby, 2021) (Mahendra, *et al.*, 2023).

## SVM

Pada tahapan ini, algoritma Support Vector Machine (SVM) digunakan untuk membangun model klasifikasi sentimen berdasarkan data ulasan pengguna. SVM bertujuan untuk mencari sebuah *hyperplane* optimal yang mampu memisahkan data dari dua kelas berbeda dengan margin selebar mungkin. Prediksi kelas dilakukan dengan menggunakan fungsi Keputusan sebagai berikut:

$$Prediksi = \text{sign}\left(\sum_{i=1}^m \alpha_i y_i K(x_i, x) + b\right) \quad (3)$$

Dalam rumus (3) tersebut  $\alpha_i$  merupakan bobot hasil dari proses pelatihan.  $y_i$  adalah tabel data latih.  $K(x_i, x)$  adalah fungsi kernel yang menghitung tingkat kesamaan antara data latih  $x_i$  dengan data uji  $x$ , dan  $b$  adalah nilai bias.

Melalui penggunaan fungsi kernel, SVM mampu membentuk pemisahan data tidak hanya dalam ruang berdimensi dua, tetapi juga dalam ruang berdimensi lebih tinggi, memungkinkan klasifikasi terhadap data yang tidak dapat dipisahkan secara linear (Madjid, *et al.*, 2023).

## Evaluasi Model

Setelah proses pelatihan model selesai, tahap berikutnya adalah mengevaluasi performa model terhadap data uji guna menilai kemampuannya dalam mengenali pola pada data yang belum pernah dianalisis sebelumnya. Evaluasi dilakukan dengan menggunakan metrik umum seperti akurasi,

presisi, *recall*, dan *f1-score*. Akurasi klasifikasi positif, *recall* mengukur keberhasilan model dalam mendeteksi kelas positif, dan *f1-score* memberikan nilai rata-rata harmonis dari presisi dan *recall*. Untuk memperkuat analisis, digunakan juga *confusion matrix* yang menyajikan visualisasi prediksi benar dan salah untuk setiap kelas. Evaluasi ini krusial dalam menilai kemampuan generalisasi model terhadap data yang belum dikenali sebelumnya (Thomas & Rumaisa, 2022).

### 3. HASIL DAN PEMBAHASAN

#### *Scraping Data*

Proses scraping data ulasan pengguna dari Google Play Store menghasilkan 783 baris data untuk Access by KAI, serta masing-masing 500 baris data untuk MyMRTJ dan MitraDarat. Tahapan pre-processing seperti case folding, cleaning, dan stemming dilakukan untuk memastikan data siap dianalisis. Hasil ini sejalan dengan penelitian oleh Nugroho, *et al.* (2024) yang juga menggunakan teknik scraping untuk mengumpulkan data ulasan aplikasi transportasi. Namun, penelitian ini melibatkan lebih banyak data dibandingkan dengan studi sebelumnya, sehingga memberikan cakupan yang lebih luas dalam analisis sentimen.

Tabel 1. Tahapan *Pre-processing* (Sampel *Dataset* dari Access by KAI)

Tahapan	Hasil
<i>content</i>	Aplikasi sangat bagus, memudahkan saya melakukan perjalanan ke Jakarta.
<i>case folding</i>	aplikasi sangat bagus, memudahkan saya melakukan perjalanan ke jakarta. 🤔 🤔
<i>remove punctuation</i>	aplikasi sangat bagus memudahkan saya melakukan perjalanan jakarta
<i>stopword</i>	['aplikasi', 'bagus', 'memudahkan', 'perjalanan', 'jakarta']
<i>normalized</i>	['aplikasi', 'bagus', 'memudahkan', 'perjalanan', 'jakarta']
<i>stemming teks</i>	['aplikasi', 'bagus', 'mudah', 'jalan', 'jakarta']

#### Hasil Implementasi TF-IDF

Setelah melalui tahapan praproses, data diberi bobot menggunakan metode TF-IDF. Proses TF-IDF dilakukan terhadap ulasan pengguna dari ketiga aplikasi setelah tahap pre-processing selesai. Nilai TF-IDF ini mencerminkan seberapa penting kata-kata tersebut dalam konteks masing-masing aplikasi. Kata-kata tersebut sejalan dengan hasil analisis sentimen, di mana ketiganya didominasi oleh ulasan bernada positif, dan menjadi indikator kuat dalam proses klasifikasi menggunakan model SVM. Hasilnya dapat dilihat pada Tabel 2.

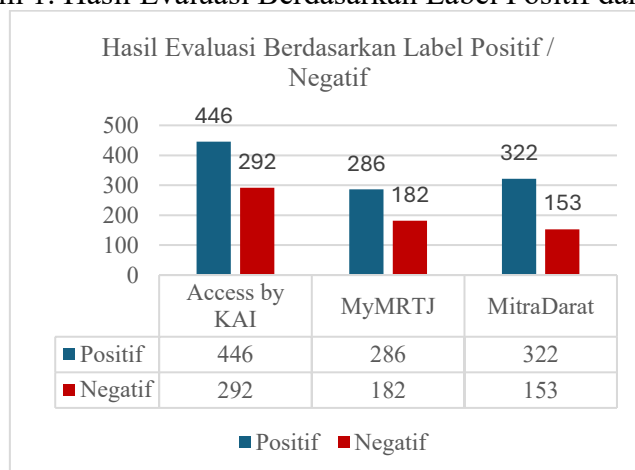
Tabel 2. Hasil TF-IDF (Sampel *Dataset* dari Access by KAI)

No.	aplikasi	fasilitas	kampung	layan	mudah	pulang	sulit
1	1.405465	0.000000	1.405465	1.693147	1.405465	1.405465	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.098612
3	0.000000	0.000000	1.405465	0.000000	0.000000	1.405465	0.000000
4	1.405465	2.098612	0.000000	1.693147	1.405465	0.000000	0.000000
5	2.810930	0.000000	1.405465	0.000000	1.405465	1.405465	0.000000

#### Pelabelan

Diagram 1 menunjukkan hasil evaluasi sentimen berdasarkan label positif dan negatif terhadap tiga aplikasi layanan transportasi publik. Access by KAI memperoleh sentimen positif tertinggi dengan 446 ulasan, diikuti oleh MitraDarat sebanyak 322 ulasan positif, dan MyMRTJ sebanyak 286 ulasan positif. Sementara itu, jumlah ulasan negatif tertinggi juga tercatat pada Access by KAI (292 ulasan), kemudian MyMRTJ (182 ulasan), dan MitraDarat (153 ulasan).

Diagram 1. Hasil Evaluasi Berdasarkan Label Positif dan Negatif



Temuan ini menunjukkan bahwa secara umum, persepsi pengguna terhadap ketiga aplikasi cenderung lebih dominan positif, meskipun masih terdapat sejumlah keluhan atau kritik yang tercermin dalam ulasan negatif.

### Hasil Evaluasi Support Vector Machine (SVM)

Berikut ini adalah hasil pemodelan SVM yang diterapkan pada tiga aplikasi transportasi, yaitu Access by KAI, MitraDarat, dan MyMRTJ.

Diagram 2. Hasil Pemodelan SVM Positif

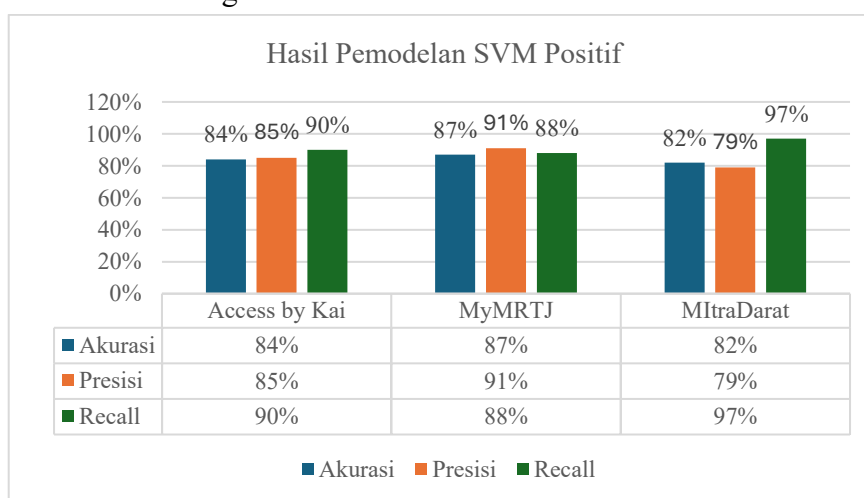


Diagram 2 menunjukkan hasil pemodelan algoritma SVM untuk tiga aplikasi, yaitu Access by KAI, MyMRTJ, dan MitraDarat, dengan memperlihatkan variasi performa berdasarkan metrik akurasi, presisi, dan recall. MyMRTJ menunjukkan performa terbaik dengan presisi tertinggi sebesar 91%, diikuti dengan akurasi sebesar 87% dan *recall* sebesar 88%. Access by KAI berada di posisi berikutnya, dengan *recall* tertinggi (90%) di antara ketiganya, serta akurasi 84% dan presisi 85%. Sementara itu, MitraDarat mencatat akurasi sebesar 82% dan presisi terendah (79%), tetapi memiliki *recall* tertinggi secara keseluruhan, yaitu 97%. Perbedaan nilai metrik ini menggambarkan variasi efektivitas model SVM dalam mengklasifikasikan sentimen positif berdasarkan data dari masing-masing aplikasi.

### Visualisasi Data (*Word Cloud*)

*Word Cloud* dibuat berdasarkan semua kata yang muncul dalam sentimen ketiga aplikasi. Hasilnya dapat dilihat pada gambar 2, gambar 3, dan gambar 4, yang masing-masing mencerminkan *Word Cloud* dari setiap aplikasi.



Gambar 2.  
*Word Cloud* Aplikasi  
Access by KAI



Gambar 3.  
*Word Cloud* Aplikasi  
MitraDarat



Gambar 4.  
*Word Cloud* Aplikasi  
MyMRTJ

Sebagian besar sentimen yang disampaikan oleh pengguna berkaitan dengan aplikasi, tiket, dan kegiatan pulang kampung, yang menunjukkan bahwa banyak pengguna aplikasi ini menggunakannya saat mudik atau Lebaran. Kata-kata dengan sentimen positif yang sering muncul, terutama dalam huruf kapital, adalah "bagus", "mudah", dan "baik". Sementara itu, kata-kata dengan sentimen negatif yang sering muncul dalam huruf kecil adalah "gagal", "sulit", dan "lambat". Dari sini dapat disimpulkan bahwa sebagian besar sentimen terhadap ketiga aplikasi ini bersifat positif, meskipun masih terdapat beberapa keluhan yang perlu diperbaiki, yang terlihat dari adanya sentimen negatif yang muncul pada setiap aplikasi.

Beberapa penelitian dalam lima tahun terakhir menunjukkan bagaimana algoritma SVM dan pendekatan leksikal bisa efektif dalam menganalisis sentimen terkait layanan transportasi. Misalnya, penelitian yang dilakukan oleh Nugraha & Gustian, (2023) menggunakan algoritma SVM untuk mengklasifikasi sentimen terhadap ulasan aplikasi Grab, Gojek, dan Maxim. Mereka menemukan bahwa SVM terbukti bekerja dengan baik dalam konteks transportasi. Di sisi lain, penelitian Devi, *et al.* (2024) menggabungkan SVM dengan metode Word2Vec untuk menganalisis sentimen terhadap layanan Access by KAI, dan hasilnya menunjukkan bahwa kombinasi kernel RBF dan teknik *oversampling* SMOTE dapat meningkatkan akurasi klasifikasi hingga 81%, dengan AUC 0,81. Penelitian lainnya oleh Anggoro, *et al.* (2025) yang menganalisis *tweet* tentang subsidi mobil listrik menggunakan SVM, menunjukkan akurasi 86,43% dengan mayoritas sentimen bersifat netral. Hal ini menegaskan bahwa SVM bisa menjadi alat yang efektif untuk menganalisis opini publik di media sosial.

## 4. KESIMPULAN DAN SARAN

Penelitian ini mengkaji bagaimana perasaan pengguna terhadap tiga aplikasi transportasi publik di Indonesia seperti Access by KAI, MyMRTJ, dan MitraDarat, dengan menggunakan metode Support Vector Machine (SVM) dan kamus *INSET Lexicon-Based*. Hasilnya menunjukkan bahwa mayoritas ulasan yang diberikan positif, dengan Access by KAI mendapatkan ulasan positif terbanyak. Ketepatan klasifikasi yang dilakukan oleh SVM juga menunjukkan hasil yang baik, terutama dengan MyMRTJ yang mencapai akurasi 87%. Visualisasi dalam bentuk *Word Cloud* menampilkan kata-kata positif yang sering muncul, seperti "baik", "mudah", dan "jalan".



Pendekatan INSET *Lexicon-Based* terbukti efektif dalam memahami sentimen dalam bahasa Indonesia. Berdasarkan temuan ini, disarankan agar pengembang aplikasi transportasi memanfaatkan analisis sentimen untuk lebih memahami kebutuhan dan keluhan pengguna, sehingga dapat meningkatkan pengalaman mereka. Penelitian lebih lanjut dengan data yang lebih banyak juga dapat membantu meningkatkan akurasi model. Selain itu, memperkenalkan analisis sentimen secara *real-time* akan sangat membantu agar aplikasi bisa merespons dengan cepat perubahan atau masalah yang muncul dari pengguna.

### Ucapan Terima Kasih

Kami ingin mengucapkan terima kasih kepada Fakultas Teknologi Informasi Universitas Tarumanagara atas dukungan fasilitas dan sumber daya yang telah diberikan. Tak lupa, kami juga mengucapkan terima kasih kepada Ibu Viny Christanti Mawardi selaku dosen pembimbing yang memberikan pengarahan, saran, dan bimbingan yang sangat berarti selama proses penyusunan artikel ini. Kami berharap hasil penelitian ini dapat memberikan manfaat positif bagi pengembangan aplikasi, serta mendorong inovasi dan perbaikan fitur yang masih perlu dibenahi.

### REFERENSI

- Asiyah, N. L., Mursityo, Y. T. & Setiawan, N. Y., 2025. *Analisis Sentimen Ulasan Pengguna Aplikasi C-Access Menggunakan Support Vector Machine*. Tersedia pada: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/14529> [Diakses 9 Mei 2025].
- Culio, S. P. & Astuti, L. G., 2024. Analisis Sentimen Ulasan Aplikasi RedBus Menggunakan Metode SVM dan AdaBoost. *JNATIA (Jurnal Nasional Teknologi Informasi dan Aplikasinya)*, 2(3), pp. 17-24.
- Fathoni, M. F. N., Puspaningrum, E. Y. & Sihananto, A. N., 2024. Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM. *Jurnal Informatika dan Sains Teknologi*, 2(3), pp. 62-76.
- Iqrom, M. et al., 2025. ANALISIS SENTIMEN APLIKASI GOJEK, GRAB, MAXIM MENGGUNAKAN ALGORITMA SVM. *JURNAL INOVTEK POLBENG - SERI INFORMATIKA*, 10(1), pp. 237-249.
- Kholifah, B., Thoib, I., Sururi, N. & Kurnia, N. D., 2024. Analisis Sentimen Warganet terhadap Isu Layanan Transportasi Online Berbasis InSet Lexicon menggunakan Logistic Regression. *Jurnal Ilmiah KLIK (Kumpulan Jurnal Ilmu Komputer)*, 11(1), pp. 14-25.
- Madjid, M. F., Ratnawati, D. E. & Rahayudi, B., 2023. Sentiment Analysis on App Reviews Using Support Vector Machine and Naive Bayes Classification. *Sinkron : Jurnal dan Penelitian Teknik Informatika*, 7(1), pp. 556-562.
- Mahendra, M. H., Murdiansyah, D. T. & Lhaksmana, K. M., 2023. Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer. *Dike: Jurnal Ilmu Multidisiplin*, 1(2), pp. 37-43.
- Nugroho, M. A. S., Susilo, D. & Retnoningsih, D., 2024. ANALISIS SENTIMEN ULASAN APLIKASI "ACCESS BY KAI" MENGGUNAKAN ALGORITMA MACHINE LEARNING. *JURNAL TEKINKOM (TEKNIK INFORMASI DAN KOMPUTER)*, 7(2), pp. 820-827.
- Putri, R. R. & Cahyono, N., 2024. ANALISIS SENTIMEN KOMENTAR MASYARAKAT TERHADAP PELAYANAN PUBLIK PEMERINTAH DKI JAKARTA DENGAN ALGORITMA SUPER VECTOR MACHINE DAN NAIVE BAYES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), pp. 2363-2371.

- Sahara, S. & Zetalia, S., 2024. Analisis Efektivitas Aplikasi KAI Acces Dalam Meningkatkan Efisiensi Perjalanan Penumpang KAI Commuter Line. *Journal Of Social Science Research*, pp. 11904-11910.
- Satria, F., Putra, W. H. N. & Hanggara, B. T., 2023. Analisis Sentimen Mengenai Kepuasan Pengguna terhadap Pelayanan Transportasi Minibus PT Mahkota Tria Wisata dengan menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(8), pp. 3577-3583.
- Suanpang, P., Jamjuntr, P. & Kaewyong, P., 2021. SENTIMENT ANALYSIS WITH A TEXTBLOB PACKAGE. *Journal of Management Information and Decision Sciences*, 24(6), pp. 1-9.
- Thomas, V. W. D. & Rumaisa, F., 2022. Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan TF-IDF. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(3), pp. 1767-1774.
- Yutika, C. H. & Al Faraby, A. S., 2021. Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), pp. 442-430.