



Manajerial: Jurnal Manajemen dan Sistem Informasi

Journal homepage: <https://ejournal.upi.edu/index.php/manajerial>



Analisis Sentimen Aplikasi Pemesanan Tiket Bus di Google Play Menggunakan Support Vector Machine

Muhamad Kurniawan, Hendri Mahmud Nawawi

Universitas Nusa Mandiri

*Correspondence: E-mail: muhamadkurniawan.dev@gmail.com, hendri.hiw@nusamandiri.ac.id

ABSTRAK

RedBus merupakan salah satu aplikasi pemesanan tiket bus secara daring yang banyak digunakan di Indonesia. Namun, seiring tingginya volume pengguna, muncul berbagai ulasan dengan sentimen beragam yang belum dianalisis secara sistematis. Penelitian ini mempunyai tujuan untuk menganalisis sentimen pengguna terhadap aplikasi RedBus dengan menerapkan metode *Support Vector Machine* (SVM) pada kumpulan data berisi 5.000 ulasan. Proses dimulai dengan pengambilan data menggunakan teknik *web scraping*, kemudian dilanjutkan dengan tahapan *preprocessing* yang mencakup *cleaning*, tokenisasi, *case folding*, *stemming*, penghapusan kata tidak bermakna (*stopword removal*), dan normalisasi. Data diberi label secara otomatis menggunakan pendekatan berbasis leksikon (*lexicon-based*), lalu dibagi menjadi data latih sebesar 80% dan data uji sebesar 20%. Penelitian ini juga membandingkan performa tiga algoritma klasifikasi—SVM, *Naïve Bayes*, dan *K-Nearest Neighbor* (KNN)—dengan bantuan *library* PyCaret. Hasil *cross-validation* menampilkan bahwa model SVM memperoleh akurasi tertinggi sebesar 99,13%, disertai oleh *Naïve Bayes* sebesar 98,68% dan KNN sebesar 98,26%. Model terbaik, yaitu SVM dengan *kernel* linear, kemudian diterapkan pada data uji dan menghasilkan akurasi akhir sebesar 94,16%. Nilai metrik evaluasi seperti *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,94, yang memperlihatkan bahwa model mempunyai performa klasifikasi yang cukup optimal untuk mengkaji sentimen pengguna aplikasi RedBus.

ARTICLE INFO

Article History:

Submitted/Received 25 Jan 2023

First Revised 15 Mar 2023

Accepted 19 May 2023

First Available online 20 May 2023

Publication Date 01 Jun 2023

Kata Kunci:

Analisis Sentimen, Data Mining,
KNN, Naïve Bayes, SVM, redBus.

1. PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi kini menghadirkan pengaruh signifikan terhadap berbagai aspek kehidupan, termasuk transportasi. Digitalisasi layanan publik memungkinkan masyarakat untuk mengakses berbagai fasilitas secara daring, termasuk pemesanan tiket transportasi. Salah satu moda transportasi yang masih menjadi andalan masyarakat Indonesia, terutama di daerah dengan mobilitas tinggi seperti Kabupaten Wonogiri, adalah bus. Kabupaten Wonogiri memiliki populasi lebih dari satu juta jiwa, dengan sekitar 25% di antaranya merupakan perantau (Riyadi et al., 2021). Tingginya arus keluar-masuk penduduk menciptakan kebutuhan yang tinggi terhadap sarana transportasi darat yang andal, nyaman, dan terjangkau.

Bus merupakan salah satu pilihan moda transportasi yang populer karena menawarkan harga yang kompetitif dan kenyamanan yang semakin meningkat berkat fasilitas yang terus dikembangkan. Selain digunakan saat musim mudik, layanan bus juga dimanfaatkan untuk berbagai kebutuhan seperti wisata, perjalanan dinas, dan aktivitas antarkota lainnya. Beberapa penelitian terdahulu menunjukkan bahwa aspek harga, fasilitas, dan kualitas pelayanan menjadi faktor penting dalam menentukan kepuasan pengguna, di mana harga cenderung memiliki pengaruh yang paling dominan (Sinaga et al., 2020).

Seiring dengan meningkatnya kebutuhan akan layanan transportasi yang praktis, aplikasi berbasis mobile seperti redBus hadir untuk memudahkan proses pemesanan tiket bus secara daring. Aplikasi ini memungkinkan pengguna untuk memilih tanggal keberangkatan, operator bus, kelas, tempat duduk, serta titik naik dan turun (Pramudibyo et al., 2024). Kehadiran redBus bertujuan merevolusi sistem transportasi darat di Indonesia dengan memperkenalkan konsep pemesanan tiket secara digital yang efisien dan user-friendly. Aplikasi redBus dapat diunduh melalui Google Play Store dan meninggalkan ulasan terkait pengalaman penggunaan (Husnina et al., 2023).

Meskipun redBus menawarkan berbagai kemudahan, penilaian yang dituliskan oleh pengguna yang tercatat di platform khususnya Google Play Store menunjukkan tanggapan yang bermacam-macam. Di satu sisi, banyak ulasan positif mengapresiasi kemudahan dalam pemesanan dan variasi pilihan armada; namun di sisi lain, tidak sedikit juga ditemukan komentar negatif terkait gangguan teknis, kesulitan dalam menghubungi layanan pelanggan, error pada sistem pemilihan kursi, serta masalah refund. Hal ini menunjukkan pentingnya analisis lebih lanjut terhadap opini dan sentimen pengguna agar pengembang aplikasi dapat melakukan evaluasi dan perbaikan layanan secara tepat sasaran.

Salah satu pendekatan yang bisa diterapkan untuk mengetahui ulasan pengguna secara sistematis adalah analisis sentimen. Analisis sentimen ialah teknik pada *text mining* yang dimaksudkan untuk mengetahui dan mengelompokkan pendapat pengguna pada sebuah kategori positif atau negatif (Saputro et al., 2023). Dengan memanfaatkan teknik ini, perusahaan dapat memperoleh wawasan berharga mengenai kepuasan dan keluhan pelanggan, serta melakukan penyesuaian layanan berdasarkan data yang bersifat real-time.

Dalam riset ini, algoritma Support Vector Machine (SVM) dipilih menjadi metode pokok guna klasifikasi sentimen. SVM ialah salah satu algoritma *machine learning* yang sering diterapkan pada pengolahan teks sebab keunggulannya dalam mengelola data berdimensi tinggi, toleransi terhadap noise, serta performa yang stabil dalam berbagai kondisi dataset (Nadhifah et al., 2024).

Sejumlah penelitian sebelumnya sudah membuktikan keunggulan algoritma SVM dalam tugas analisis sentimen terhadap ulasan aplikasi. Penelitian terhadap 520 ulasan aplikasi MOLA menggunakan SVM dengan pendekatan *Knowledge Discovery in Databases* (KDD) menunjukkan performa terbaik pada skenario 90:10 dengan kernel RBF, memperoleh

ketepatan senilai 92,31%, precision 96,3%, F1-score 92,86% dan recall 89,66% (Hendriyanto et al., 2022). Penelitian lain terhadap 10.000 ulasan Google Meet menunjukkan akurasi 94%, dengan precision 98%, recall 95%, dan F1-score 96% (Fitri & Putri, 2022). Studi terhadap aplikasi BRImo juga menampilkan bahwa SVM lebih baik dibandingkan Naive Bayes, memperoleh akurasi masing-masing 97,69% dan 96,53% (Astuti et al., 2022).

Oleh sebab itu, pada studi ini memiliki urgensi untuk mengeksplorasi kinerja algoritma SVM dalam menganalisis sentimen pengguna aplikasi redBus, serta membandingkannya dengan pendekatan lain seperti *K-Nearest Neighbor* (KNN) dan *Naive Bayes* guna memperoleh model klasifikasi terbaik. Penulis mengharapkan temuan dari penelitian ini mampu menghadirkan kontribusi nyata untuk mengembangkan layanan transportasi digital serta mendukung pengembangan aplikasi sehingga dapat dipahami kebutuhan dan ekspektasi pengguna secara lebih baik.

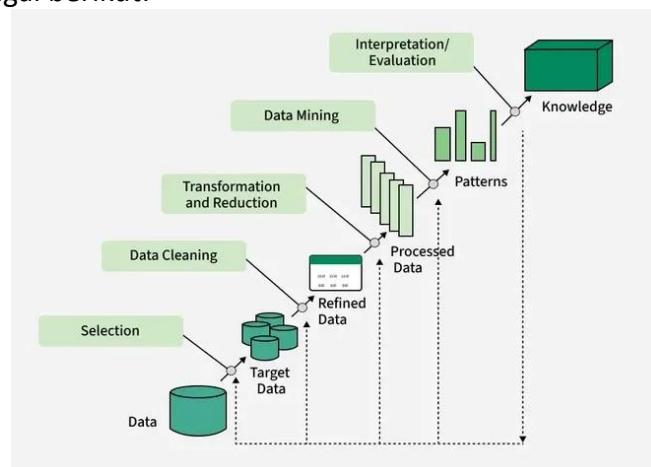
2. KAJIAN PUSTAKA

2.1. Machine Learning

Pembelajaran mesin (*machine learning*) termasuk elemen atas kecerdasan buatan dengan fokus utama untuk mengembangkan sistem yang mampu belajar dari data dan meningkatkan kinerja seiring bertambahnya pengalaman (Bell, 2015). Dalam konteks analisis sentimen, *machine learning* membuat sistem dapat memahami pola sentimen dari data ulasan yang telah diberi label, dan selanjutnya memperkirakan sentimen pada data baru menurut pola tersebut. Pendekatan yang sering diterapkan yaitu *supervised learning*, dengan pelatihan model dilakukan dengan data yang sudah dilabeli sentimen, misalnya negatif, positif, atau netral, baik melalui kosakata maupun anotasi manual. (Savitri et al., 2021). Melalui proses pelatihan ini, model dapat membangun fungsi klasifikasi yang akurat untuk tugas analisis opini.

2.2. Data Mining

Dalam mendukung analisis data, proses *data mining* memiliki peran penting dalam menggali pola, tren, atau hubungan tersembunyi dari kumpulan data berukuran besar (T & Larose, 2015). Tujuan utamanya adalah menghasilkan informasi yang berguna untuk pengambilan keputusan. *Data mining* termasuk elemen utama dari tahap *Knowledge Discovery in Databases* (KDD), yaitu rangkaian langkah sistematis untuk mengidentifikasi informasi bermakna dari data (Nuraliza et al., 2022). Adapun proses *Knowledge Discovery in Databases* (KDD) sebagai berikut:



Gambar 1. Proses KDD

Tahapan KDD dimulai dari seleksi data, yaitu pemilihan data relevan untuk dianalisis. Selanjutnya dilakukan praproses atau pembersihan data, guna mengatasi duplikasi, kesalahan penulisan, dan ketidakkonsistenan. Setelah data bersih, tahap transformasi dilaksanakan guna merubah format data supaya selaras dengan kepentingan analisis. Proses inti berupa penggalian data memanfaatkan algoritma tertentu untuk menemukan pola atau tren tersembunyi. Terakhir, interpretasi dan evaluasi dilakukan untuk memastikan hasil analisis dapat dipahami dan digunakan secara efektif dalam pengambilan keputusan (Sumantri & Utami, 2020).

2.3. Text Mining

Text mining berperan penting dalam mengekstrak data berkualitas tinggi dari dokumen dalam jumlah besar. Proses ini mencakup identifikasi pola dan tren menggunakan pendekatan statistik dan komputasional, serta mencakup teknik seperti klasifikasi, pengelompokan, ekstraksi entitas, dan analisis sentimen (Purbo, 2019). Sebelum data dianalisis, diperlukan tahapan praproses teks untuk membersihkan dan menstandarkan format data. Tahap ini mencakup *case folding* (merubah segala huruf besar menjadi huruf kecil), *filtering* (menghilangkan simbol atau karakter tidak sesuai), *tokenizing* (mengurai kalimat menjadi per kata), *stopword removal* (mengeliminasi kata yang terlalu banyak muncul tetapi tidak bermakna signifikan), serta *stemming* (merubah kata menjadi bentuk awal) (Ridwansyah, 2022).

2.4. Analisis Sentimen

Analisis sentimen, atau yang juga dikenal sebagai *opinion mining*, ialah elemen dari studi komputasi yang berfokus pada identifikasi, interpretasi, dan klasifikasi opini atau emosi yang diungkapkan dalam bentuk teks (Alhaq et al., 2021). Tujuan utamanya adalah untuk mengevaluasi persepsi publik terhadap suatu entitas, layanan, atau peristiwa, dan mengklasifikasikan ekspresi sentimen tersebut ke dalam kelompok negatif, positif, atau netral (Manullang et al., 2023). Proses analisis sentimen menggunakan sejumlah tahapan utama, yaitu dari pengumpulan data ulasan, diikuti oleh pra-pemrosesan teks untuk membersihkan dan menyiapkan data. Selanjutnya dilakukan ekstraksi fitur menggunakan teknik seperti *TF-IDF*, yang kemudian diolah oleh model klasifikasi seperti SVM, Naive Bayes, atau KNN dalam tahap klasifikasi sentimen.

2.5. Support Vector Machine

SVM termasuk dalam pendekatan *supervised learning* yang beroperasi dengan membangun sebuah *hyperplane* sebagai batas pemisah optimal antara dua kelas data. *Hyperplane* ini dirancang untuk memaksimalkan margin, yakni jarak terjauh antara data dari masing-masing kelas terhadap batas tersebut (Fide et al., 2021). Data point yang posisinya paling dekat *hyperplane*, dikenal sebagai *support vector*, memiliki peran krusial dalam menentukan posisi dan arah pemisah. Dengan memanfaatkan pola dari data latih, SVM mampu mengklasifikasikan data baru secara akurat berdasarkan jarak dan posisi relatif terhadap bidang pemisah. Rumus SVM disajikan dalam persamaan di bawah ini:

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right)$$

Keterangan; (x): klasifikasi data, x_i : data latih, y_i : label kelas data latih, α_i : koefisien bobot setiap data latih, $K(x_i, x)$: jarak antara data latih x_i dan data uji x , b : nilai bias (Sari & Suryono, 2024).

2.6. Naive Bayes

Naive Bayes yaitu algoritma klasifikasi berbasis probabilistik yang termasuk dalam metode *supervised learning*. Algoritma ini bekerja dengan menerapkan *Teorema Bayes* dan menganggap jika setiap fitur memiliki sifat saling independen—sebuah asumsi yang menjadikannya “naive”. *Naive Bayes* menjelaskan kinerja yang optimal dalam berbagai tugas pengelompokan, termasuk analisis sentimen (Kusnawi & Pratama, 2024). Keunggulan utamanya terletak pada kesederhanaan, efisiensi komputasi, dan keahliannya dalam mengatasi data berukuran besar. Rumus *Naive Bayes* dijabarkan pada persamaan berikut:

$$P(H | X) = \frac{P(X | C_i) \cdot P(C_i)}{P(X)}$$

Keterangan; X: data belum kelasnya, C_i : hipotesis data ke dalam kelas tertentu, $P(C_i|X)$: probabilitas hipotesis C_i menurut data X (*posterior probability*), $P(C_i)$: probabilitas awal hipotesis C_i sebelum memperhatikan data X (*prior probability*), $P(X|C_i)$: Kemungkinan munculnya data X asumsi hipotesis C_i benar (*likelihood*), $P(X)$: probabilitas total dari data X (*evidence*) (Martantoh & Yanih, 2022).

2.7. K-Nearest Neighbor

K-Nearest Neighbor (KNN) merupakan algoritma klasifikasi dalam *supervised learning* yang bekerja berdasarkan prinsip *learning by analogy*, yakni objek-objek yang serupa diasumsikan memiliki kelas yang sama. Tahap pengelompokan dilaksanakan dengan mengukur jarak diantara data baru dan data latih menggunakan metrik seperti *Euclidean distance*, menentukan kelas menurut sebagian besar dari K tetangga terdekat (Huda et al., 2024). KNN dikenal karena kesederhanaannya dan kemampuannya dalam menangani data non-linear, meskipun kinerjanya sangat dipengaruhi oleh pemilihan nilai K dan ukuran data latih. Berikut ini rumus dari KNN:

$$d(x, y) = \sqrt{\left(\sum_{i=1}^n (x_i - y_i)^2 \right)}$$

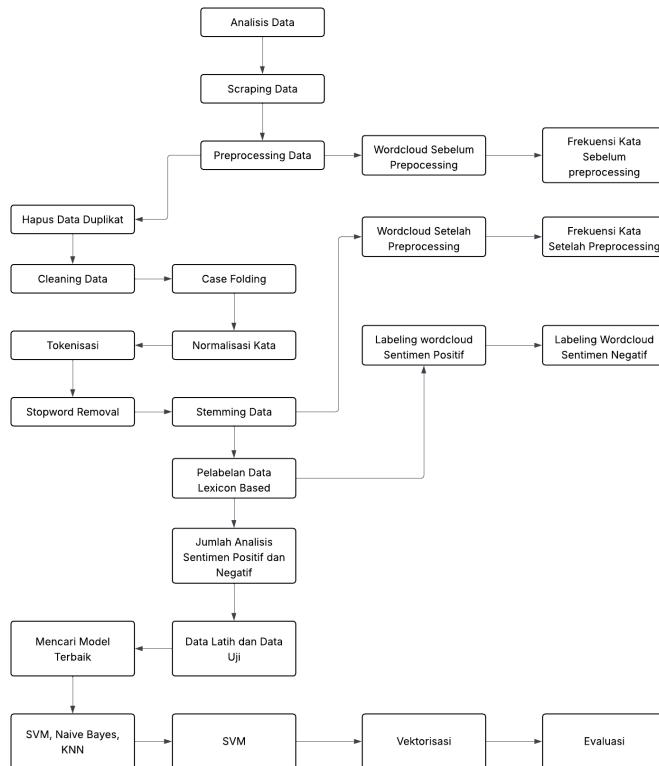
Keterangan; d: nilai jarak antara dua titik data, x: data uji yang diklasifikasikan, y: data pelatihan, i: indeks setiap fitur, n: jumlah total fitur data (Widya Utami & Artana, 2022).

3. METODE PENELITIAN

Studi ini menerapkan dua metode utama, yaitu pengumpulan dan analisis data. Dalam tahap mengumpulkan data, diterapkannya tiga teknik, yakni observasi, telaah literatur (studi pustaka), serta pengumpulan data melalui penilaian pengguna yang didapatkan dari *platform Google Play Store*. Sementara itu, analisis data dilakukan melalui serangkaian tahapan, meliputi pemrosesan teks, penerapan model berbasis *Lexicon*, mencari model algoritma terbaik seperti *Naive Bayes*, *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM) sebagai pembanding, serta penyusunan matriks konfusi untuk evaluasi performa klasifikasi, dan penyajian hasil dalam bentuk visualisasi

Alur metodologi penelitian ini dimulai dari proses analisis dan pengumpulan data melalui *scraping*. Data yang diperoleh kemudian diproses melalui tahap *preprocessing* untuk membersihkan dan menyiapkannya, mencakup penghapusan duplikat, tokenisasi, *stopword removal*, *case folding*, normalisasi kata, dan *stemming*. Setelah *preprocessing*, dilakukan visualisasi dalam bentuk *wordcloud* dan perhitungan frekuensi kata, baik sebelum maupun sesudah pembersihan data. Selanjutnya, data diberi label sentimen positif atau negatif menggunakan pendekatan *lexicon-based*, kemudian dilakukan analisis jumlah sentimen. Data

yang sudah dilabel dikelompokkan menjadi data uji dan latih, lalu diubah ke bentuk vektor untuk keperluan pemodelan. Algoritma machine learning seperti SVM, KKN, dan *Naive Bayes* diterapkan untuk membangun model klasifikasi sentimen, dan hasilnya dievaluasi untuk menentukan model terbaik.



Gambar 2. Tahapan Penelitian

4. HASIL PENELITIAN DAN PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data dilaksanakan dengan bantuan *library Google Play Scraper* guna mengekstraksi ulasan pengguna aplikasi redBus dari *Google Play Store*. Seluruh proses dijalankan memanfaatkan *platform Google Colaboratory* sebagai lingkungan pemrosesan data.

```

 5] import pandas as pd
      data = pd.read_csv("hasil_scraper_ulasan_app_redbus.csv")
      data.info()
      <class 'pandas.core.frame.DataFrame'>
      RangeIndex: 5000 entries, 0 to 4999
      Data columns (total 5 columns):
      #   Column           Non-Null Count  Dtype  
      --  Review ID        5000 non-null   object 
      1   Username        5000 non-null   object 
      2   Rating          5000 non-null   int64  
      3   Review Text     5000 non-null   object 
      4   Date            5000 non-null   object 
      dtypes: int64(1), object(4)
      memory usage: 195.4+ KB

 6] data.head(10)
      Review ID      Username  Rating  Review Text          Date
      0  b414eb88-5ba3-481b-b47c-c40ac12d0ccb  Niisu Lala Oktafiani  5  sangat membantu sekali 2025-06-16 16:51:10
      1  95957251-8a8f-4791-b443-135302ddc877  Ali kurina  5  proses cepat dan sangat puas 2025-06-16 16:16:17
      2  7644e359-bea1-4265-b206-82c36a2538e7  Ranendra Sagara  3  baik 2025-06-16 13:25:49
      3  31260dc-9ac2-4047-93ec-e457c7c9e9eb  Satri Risoni  5  saya selalu pakai redbus.. tapi kali ini saya ...
      4  3b4bd546-a0b8-4274-ae94-9e4b5729d0e5  Theresia Tri Tumi  1  suka, keluar sendiri 2025-06-16 09:21:40
      5  b0d37a45-ad98-488a-b445-d919cd8c037  nyoman santawan  5  terimakasih RedBus 2025-06-16 08:54:36
      6  4b4a599a-73ba-4725-bd76-0300c8b80f03d  fahrezy  4  mantap 2025-06-16 05:54:16
      7  d4b070fe-7bcd-4e63-a0c9-f54a67c27c5  Shang ale  5  apikasi yg sangat di perlukan dan mempermudah...
      8  675be509-311a-4c19-bb1f-e6455dce7455  Ajil Nurkolla  3  bisa cuma kok agak eror, di buka kadang GK bisa ...
      9  a85aae83-429b-48ea-bdbb-428a1c7c4697  FATHUR ROHMAN  5  bagus 2025-06-16 01:52:28
  
```

Gambar 3. Data Hasil Scraper

4.2 Preprocessing Data

4.2.1 Hapus Data Duplikat

Dari data sebanyak 5.000 ulasan yang didapatkan melalui *Google Play Store*, dilakukan tahapan penyaringan untuk menghapus data yang bersifat duplikat. Tersisa sebanyak 4.071 ulasan. Data yang diterapkan pada studi ini menggunakan dasar urutan waktu, dimulai dari ulasan yang paling awal.

```
✓ [12] df.info()
Df
> <class 'pandas.core.frame.DataFrame'>
Index: 4071 entries, 0 to 4997
Data columns (total 11 columns):
 #   Column           Non-Null Count Dtype  
--- 
 0   Review ID        4071 non-null   object  
 1   Username         4071 non-null   object  
 2   userImage        4071 non-null   object  
 3   Review Text      4071 non-null   object  
 4   Rating           4071 non-null   int64  
 5   thumbsUpCount    4071 non-null   int64  
 6   reviewCreatedVersion 3325 non-null   object  
 7   Date             4071 non-null   datetime64[ns]
 8   replyContent     4050 non-null   object  
 9   repliedAt        4050 non-null   datetime64[ns]
 10  appVersion       3325 non-null   object  
dtypes: datetime64[ns](2), int64(2), object(7)
memory usage: 381.7+ KB
```

Gambar 4. Hasil Hapus Data Duplikat

4.2.2 Wordcloud Sebelum Preprocessing

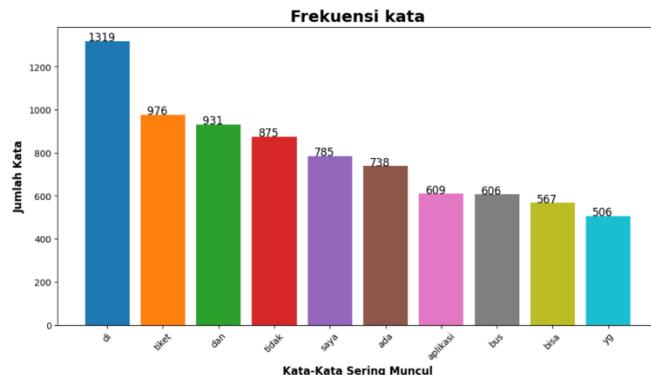
Tahapan ini dimaksudkan guna memperlihatkan kata-kata yang umum dalam data ulasan sebelum dilakukan proses pembersihan data.



Gambar 5. Hasil Wordcloud Sebelum Preprocessing

4.2.3 Frekuensi Kata Sebelum *Preprocessing*

Frekuensi kemunculan kata sebelum preprocessing dapat diamati melalui visualisasi grafik, menunjukkan jumlah kata yang sering muncul sebelum data dibersihkan.

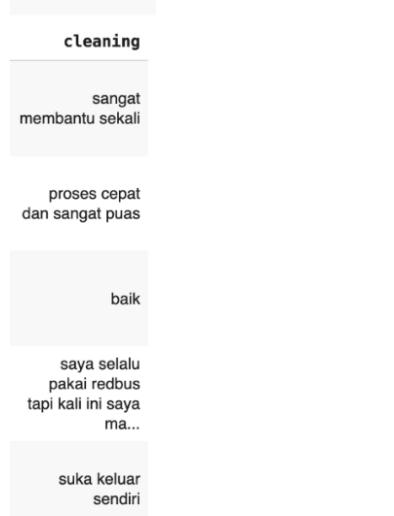


Gambar 6. Hasil Frekuensi Kata Sebelum Preprocessing

4.2.4 *Cleaning Data*

Tahap *cleaning* dilaksanakan guna mengeliminasi bagian-bagian yang tidak sesuai atau tidak memiliki kontribusi terhadap proses klasifikasi data ulasan. Seperti hashtag, tautan

(URL), tanda baca, angka, serta simbol emoji yang dapat menimbulkan gangguan dalam analisis data.



Gambar 7. Hasil Cleaning Data

4.2.5 Case Folding

Adalah tahapan merubah keseluruhan huruf kapital pada data ulasan menjadi huruf kecil (*lowercase*). Dalam proses ini, seluruh data ulasan sudah dikonversi ke dalam format huruf kecil untuk menyamakan representasi kata yang semula berbeda karena perbedaan kapitalisasi.



Gambar 8. Hasil Case Folding

4.2.6 Normalisasi Kata

Normalisasi dilakukan dengan mengganti kata yang belum baku yang ada pada ulasan menjadi bentuk baku menurut Kamus Besar Bahasa Indonesia (KBBI). Dalam tahap ini, peneliti memanfaatkan sumber data kamus Bahasa Indonesia dari *kaggle.com*.

Review Text	cleaning	case_folding	normalisasi
sangat membantu sekali	sangat membantu sekali	sangat membantu sekali	sangat membantu sekali
proses cepat dan sangat puas			
baik	baik	baik	baik
saya selalu pakai redbus.. tapi kali ini saya ...	saya selalu pakai redbus tapi kali ini saya ma...	saya selalu pakai redbus tapi kali ini saya ma...	saya selalu pakai redbus tapi kali ini saya ma...
suka. keluar sendiri	suka keluar sendiri	suka keluar sendiri	suka keluar sendiri

Gambar 9. Hasil Normalisasi

4.2.7 Tokenisasi

Tokenisasi tahap memilah teks dalam ulasan menjadi unit kata yang lebih kecil, yang dikenal sebagai token. Setiap potongan kata hasil pemisahan tersebut akan digunakan sebagai satuan analisis dalam tahap pemrosesan data selanjutnya.

Review Text	cleaning	case_folding	normalisasi	tokenize
sangat membantu sekali	sangat membantu sekali	sangat membantu sekali	sangat membantu sekali	[sangat, membantu, sekali]
proses cepat dan sangat puas	[proses, cepat, dan, sangat, puas]			
baik	baik	baik	baik	[baik]
saya selalu pakai redbus.. tapi kali ini saya ...	saya selalu pakai redbus tapi kali ini saya ma...	saya selalu pakai redbus tapi kali ini saya ma...	saya selalu pakai redbus tapi kali ini saya ma...	[saya, selalu, pakai, redbus, tapi, kali, ini,...]
suka. keluar sendiri	suka keluar sendiri	suka keluar sendiri	suka keluar sendiri	[suka, keluar, sendiri]

Gambar 10. Hasil Tokenisasi

4.2.8 Stopword Removal

Stopword removal guna menghapus kata yang tidak memberikan kontribusi arti signifikan pada isi ulasan, seperti kata hubung atau kata bantu.

cleaning	case_folding	normalisasi	tokenize	stopword removal
sangat membantu sekali	sangat membantu sekali	sangat membantu sekali	[sangat, membantu, sekali]	[membantu]
proses cepat dan sangat puas	proses cepat dan sangat puas	proses cepat dan sangat puas	[proses, cepat, dan, sangat, puas]	[proses, cepat, puas]
baik	baik	baik	[baik]	[]
saya selalu pakai redbus tapi kali ini saya ma...	saya selalu pakai redbus tapi kali ini saya ma...	saya selalu pakai redbus tapi kali ini saya ma...	[saya, selalu, pakai, redbus, tapi, kali, ini,...]	[pakai, redbus, kali, buka, aplikasi, redbus, ya]
suka keluar sendiri	suka keluar sendiri	suka keluar sendiri	[suka, keluar, sendiri]	[suka]

Gambar 11. Hasil Stopword Removal

4.2.9 Stemming Data

Stemming bertujuan untuk mengonversi setiap kata pada ulasan menjadi bentuk kata dasar berdasarkan kaidah Bahasa Indonesia. Setelah tahap ini, seluruh data ulasan telah melalui normalisasi dan pembersihan.

case_folding	normalisasi	tokenize	stopword removal	stemming_data
sangat membantu sekali	sangat membantu sekali	[sangat, membantu, sekali]	[membantu]	bantu
proses cepat dan sangat puas	proses cepat dan sangat puas	[proses, cepat, dan, sangat, puas]	[proses, cepat, puas]	proses cepat puas
baik	baik	[baik]	[]	
saya selalu pakai redbus tapi kali ini saya ma...	saya selalu pakai redbus tapi kali ini saya ma...	[saya, selalu, pakai, redbus, tapi, kali, ini,...]	[pakai, redbus, kali, buka, aplikasi, redbus, ya]	pakai redbus kali buka aplikasi redbus ya
suka keluar sendiri	suka keluar sendiri	[suka, keluar, sendiri]	[suka]	suka

Gambar 12. Hasil Stemming Data

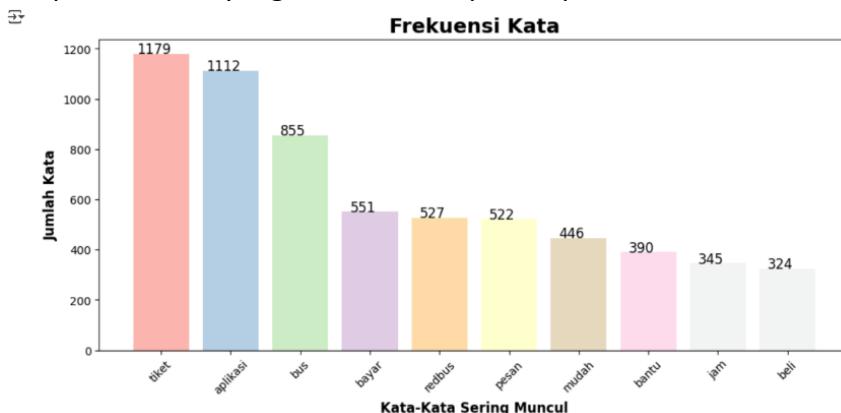
4.2.10 Wordcloud Setelah Preprocessing

Wordcloud pascaproses pembersihan data diterapkan guna memvisualisasikan kata-kata yang paling banyak muncul setelah data ulasan telah melalui tahapan *preprocessing*.

**Gambar 13.** Hasil Wordcloud Setelah Preprocessing

4.2.11 Frekuensi Kata Setelah Preprocessing

Visualisasi frekuensi kata setelah preprocessing menunjukkan kata-kata bermakna yang paling banyak ada pada ulasan, yang telah melalui proses pembersihan.

**Gambar 14.** Hasil Frekuensi Kata Setelah Preprocessing

4.3 Pelabelan Data Metode Lexicon Based

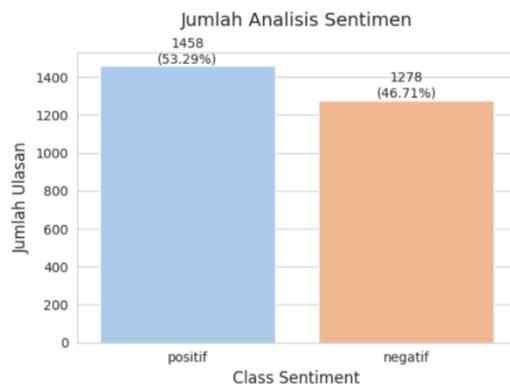
Klasifikasi berbasis *lexicon* bergantung pada kelengkapan kamus *lexicon* Bahasa Indonesia, di mana setiap kata diberi skor berdasarkan kategorinya sebagai positif atau negatif. Nilai sentimen ditentukan dari selisih jumlah skor kata positif dan negatif. Jika nilai ≥ 0 , ulasan dikategorikan sebagai sentimen positif, sedangkan jika nilai < 0 , diklasifikasikan sebagai sentimen negatif.

		stemming_data	Score	Sentiment	
1		proses cepat puas	2.0	positif	
5		terimakasih redbus	1.0	positif	
6		mantap	1.0	positif	
10	spekfasiilitas bis toll toiletsmoking room wipe...		-1.0	negatif	
11		loding ya	1.0	positif	
12	saldo potong bayar gagal complaint ribet admin...		-6.0	negatif	
13	salah rute aju koreksi fasilitas ya aju batal ...		-4.0	negatif	
14		diskon harga top	3.0	positif	
16	aplikasi aneh mesen tiketd gagal bayar aneh k...		-2.0	negatif	
17	bayar pakai briva tiket gagal pesan uang kembali		-1.0	negatif	

Gambar 15. Hasil Labeling Data Lexicon Based

4.3.1 Jumlah Analisis Sentimen

Melalui proses pelabelan data menggunakan pendekatan *Lexicon Based*, dapat diketahui jumlah ulasan yang tergolong ke dalam sentimen positif maupun sentimen negatif terhadap aplikasi redBus.



Gambar 16. Jumlah Analisis Sentimen Positif dan Negatif

4.3.2 Labeling *Wordcloud* Sentimen Positif dan Negatif

Wordcloud hasil pelabelan merupakan visualisasi dari tahap *labeling* menggunakan pendekatan *Lexicon Based*, yang memanfaatkan jumlah analisis sentimen negatif dan positif untuk menampilkan kata-kata yang banyak muncul



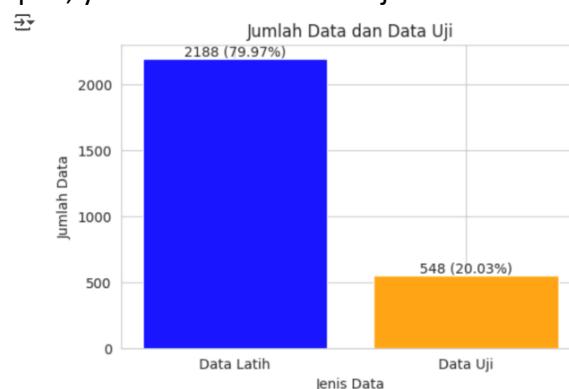
Gambar 17. Wordcloud Sentimen Negatif



Gambar 18. Wordcloud Sentimen Negatif

4.4 Data Latih dan Data Uji

Untuk mengevaluasi tingkat ketepatan algoritma klasifikasi dalam melakukan pengelompokan data dengan benar, digunakan data uji sebagai acuan. Pada studi ini, data terbagi dalam dua kelompok, yaitu 20% untuk data uji dan 80% untuk data latih.



Gambar 19. Hasil Data Uji dan data Latih

4.5 Mencari Model Terbaik

Untuk menentukan model terbaik dalam analisis sentimen aplikasi redBus, peneliti menggunakan *PyCaret* untuk membandingkan performa beberapa algoritma klasifikasi dan mengidentifikasi model yang paling optimal.

Model		Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
svm	SVM - Linear Kernel	0.9913	0.3705	0.9913	0.9830	0.9871	nan	0.2000	0.0790
nb	Naive Bayes	0.9868	0.3977	0.9868	0.9978	0.9918	nan	0.3691	0.0480
knn	K Neighbors Classifier	0.9826	0.3000	0.9826	0.9660	0.9741	nan	0.0000	0.0500

Gambar 20. Model SVM, NB, dan KNN

Dari hasil evaluasi pada gambar 20, tiga algoritma dipilih, yakni SVM, *Naive Bayes*, dan KNN. SVM menunjukkan akurasi tertinggi sebesar 99,13% dan menjadi kandidat utama. *Naive Bayes* unggul dalam *precision* (99,78%) dan *F1-score* (99,18%) dengan nilai MCC tertinggi (0,3691), menunjukkan performa seimbang. KNN dipertahankan sebagai pembanding dengan akurasi kompetitif (98,26%), meskipun memiliki MCC lebih rendah.

4.6 Model Support Vector Machine

Dalam penerapan SVM, vektorisasi berperan penting guna merubah informasi teks menjadi bentuk angka, sehingga mudah dikenali dan diolah oleh model. Setiap vektor merepresentasikan fitur-fitur dari teks yang dianalisis.

```

    # Menampilkan hasil vektorisasi
    print("Matriks Vektorisasi untuk Data Latih:")
    print(x_train_vectorized.toarray())

    # Menampilkan sebagian kecil matriks
    print("\nSebagian kecil Matriks Vektorisasi untuk data Latih:")
    print(x_train_vectorized[:5, :].toarray())

```

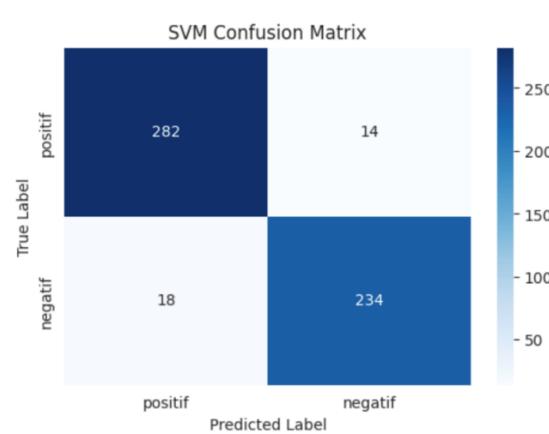
→ Matriks Vektorisasi untuk Data Latih:
[[0. 0. 0. ... 0. 0. 0.]
[0. 0. 0. ... 0. 0. 0.]
[0. 0. 0. ... 0. 0. 0.]
...
[0. 0. 0. ... 0. 0. 0.]
[0. 0. 0. ... 0. 0. 0.]
[0. 0. 0. ... 0. 0. 0.]]

Sebagian kecil Matriks Vektorisasi untuk data Latih:
[[0. 0. 0. ... 0. 0. 0.]
[0. 0. 0. ... 0. 0. 0.]
[0. 0. 0. ... 0. 0. 0.]
[0. 0. 0. ... 0. 0. 0.]
[0. 0. 0. ... 0. 0. 0.]]

Gambar 21. Hasil Vektorisasi

4.6 Evaluasi

Proses evaluasi bertujuan menilai performa model SVM dalam klasifikasi sentimen, dengan menerapkan *confusion matrix* dalam menilai akurasi untuk membedakan sentimen negatif dan positif dari data ulasan.

**Gambar 22.** Confusion Matrix

Berdasarkan Gambar 22 confusion matrix di atas, model *Support Vector Machine* (SVM) dengan *kernel linear* menunjukkan performa klasifikasi yang optimal dalam membedakan sentimen negatif dan positif. Dari total 296 data berlabel positif, sebanyak 282 data dikelompokkan dengan benar, dan hanya 14 data salah dikelompokkan sebagai negatif. Dari 252 data berlabel negatif, sebanyak 234 data dikelompokkan dengan benar, dan hanya 18 data salah dikelompokkan sebagai positif.

```

→ SVM Accuracy: 0.9416058394160584
SVM Accuracy: 94.16%
SVM Classification Report:
precision      recall   f1-score  support
negatif        0.94     0.93     0.94     252
positif        0.94     0.95     0.95     296
accuracy       0.94     0.94     0.94     548
macro avg      0.94     0.94     0.94     548
weighted avg   0.94     0.94     0.94     548

```

Gambar 23. Akurasi Model SVM

Model SVM mencapai akurasi 94,16% pada 548 data uji, dengan prediksi benar terhadap 282 data positif dan 234 data negatif. Evaluasi menunjukkan performa kuat dan seimbang, dengan precision, recall, dan F1-score sebesar 0.94–0.95 untuk kedua kelas, serta nilai macro

dan weighted average yang konsisten (0.94). Hal ini membuktikan bahwa model dapat diandalkan dalam analisis sentimen aplikasi *redBus*.

5. KESIMPULAN DAN SARAN

Berdasarkan temuan penelitian, maka kesimpulan yang didapatkan yaitu algoritma SVM menunjukkan performa terbaik dibandingkan Naive Bayes dan K-Nearest Neighbor (KNN) dengan akurasi akhir 94,16%, serta nilai precision, recall, dan F1-score yang masing-masing mencapai 0,94–0,95. SVM terbukti efektif dalam klasifikasi sentimen ulasan pengguna aplikasi RedBus, mampu memprediksi sentimen negatif dan positif secara tepat dan konsisten pada 548 data uji. Penelitian ini memberikan kontribusi bagi RedBus dalam memahami persepsi pengguna dan sebagai dasar perbaikan layanan berbasis data sentimen pengguna. Untuk penelitian selanjutnya, disarankan untuk menggunakan pendekatan deep learning seperti LSTM atau BERT guna mengoptimalkan ketepatan dan menangani konteks bahasa yang lebih kompleks, serta memperluas data dari berbagai sumber ulasan agar model lebih general dan representatif terhadap opini pengguna secara menyeluruh.

6. REFERENCES

- Alhaq, Z., Mustopa, A., Mulyatun, S., & Santoso, J. D. (2021). PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER. *JOISM : JURNAL OF INFORMATION SYSTEM MANAGEMENT*, 3(1), 16–21. <https://doi.org/https://doi.org/10.24076/joism.2021v3i2.558>
- Astuti, A. P., Alam, S., & Jaelani, I. (2022). Komparasi Algoritma Support Vector Machine dengan Naive Bayes Untuk Analisis Sentimen Pada Aplikasi BRImo. *Jurnal Bangkit Indonesia*, 11(2), 1–7. <https://doi.org/10.52771/bangkitindonesia.v11i2.196>
- Bell, J. (2015). *Machine Learning HANdS-ON FOR DEVELOPERS AND TECHNICAL PROFESSIONALS* (C. Long (ed.)). Jhon Wiley & Sons.
- Fide, S., Suparti, & Sudarno. (2021). ANALISIS SENTIMEN ULASAN APLIKASI TIKTOK DI GOOGLE PLAY MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN ASOSIASI. *JURNAL GAUSSIAN*, 10(3), 346–358. <https://doi.org/https://doi.org/10.14710/j.gauss.10.3.346-358>
- Fitri, D. A., & Putri, A. (2022). Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma Support Vector Machine. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(3), 472–478. <https://doi.org/10.37859/coscitech.v3i3.4260>
- Hendriyanto, M. D., Ridha, A. A., & Enri, U. (2022). Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine. *INTECOMS: Journal of Information Technology and Computer Science*, 5(1), 1–7. <https://doi.org/10.31539/intecoms.v5i1.3708>
- Huda, K., Pohan, S. D., & Herlina, Y. (2024). PENERAPAN PEMBOBOTAN TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY DAN ALGORITMA K- NEAREST NEIGHBOR UNTUK ANALISIS ULASAN HOTEL DI SITUS TRIPADVISOR. *JiTET (Jurnal Informatika Dan Teknik Elektro Terapan)*, 12(3), 2536–2546. <https://doi.org/http://dx.doi.org/10.23960/jitet.v12i3.4800>
- Husnina, D. N. N., Ratnawati, D. E., & Rahayudi, B. (2023). Analisis Sentimen Pengguna Aplikasi RedBus berdasarkan Ulasan di Google Play Store menggunakan Metode Naïve Bayes.

Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 7(2), 737–743.

Kusnawi, & Pratama, A. H. (2024). *Belajar Mudah dan Singkat MACHINE LEARNING* (L. M. W (ed.)). ANDI.

Manullang, O., Prianto, C., & Harani, N. H. (2023). Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest. *Jurnal Ilmiah Informatika (JIF)*, 11(2), 159–169. <https://doi.org/https://doi.org/10.33884/jif.v11i02.7987>

Martantoh, E., & Yanih, N. (2022). Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan Php Mysql. *Jurnal Teknologi Sistem Informasi*, 3(2), 166–175. <https://doi.org/10.35957/jtsi.v3i2.2896>

Nadhifah, S., Aini, F. N., Kusumawardhani, H. H., & Febrianto, M. Y. (2024). Analisis Sentiment Ulasan Aplikasi Gopay Pada Google Play Store Menggunakan Algoritma Support Vector Machine. *SURYA INFORMATIKA*, 14(1), 1–6.

Nuraliza, H., Pratiwi, O. N., & Hamami, F. (2022). Analisis Sentimen IMBD Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance. *Jurnal Mirai Manajemen*, 7(1), 1–17.

Pramudibyo, N., Muttaqin, M. R., & Sunandar, M. A. (2024). Analisis Kepuasan Pengguna Terhadap Aplikasi Redbus Dengan Menggunakan Metode End User Computing Satisfaction (Eucs). *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(2), 1099–1107. <https://doi.org/10.23960/jitet.v12i2.4133>

Purbo, O. W. (2019). *Text Mining Analisis Medsos, Kekuatan Brand & Intelejen di Internet* (A. A. Christian (ed.)). ANDI.

Ridwansyah, T. (2022). Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 2(5), 178–185. <https://doi.org/10.30865/klik.v2i5.362>

Riyadi, A., Andryana, S., & Winarsih, W. (2021). Pemilihan Transportasi Bus Antar Kota Antar Provinsi (AKAP) Dengan Metode Simple Additive Weighting (SAW), Weighted Product (WP), dan Promethee Berbasis Android. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 5(3), 247. <https://doi.org/10.35870/jtik.v5i3.177>

Saputro, W. E., Yuana, H., & Puspitasari, W. D. (2023). ANALISIS SENTIMEN PENGGUNA DOMPET DIGITAL DANA PADA KOLOM KOMENTAR GOOGLE PLAY STORE DENGAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(2), 1151–1156.

Sari, P. K., & Suryono, R. R. (2024). KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN METAVERSE. *Jurnal MNEMONIC*, 7(1), 31–39. <https://doi.org/https://doi.org/10.36040/mnemonic.v7i1.8977>

Savitri, N. L. P. C., Rahman, R. A., Venyutzky, R., & Rakhmawati, N. A. (2021). Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(1), 47–58. <https://doi.org/10.28932/jutisi.v7i1.3216>

- Sinaga, L. R., Efendi, N., & Harori, M. I. (2020). Pengaruh Kualitas Pelayanan, Fasilitas, Dan Harga Terhadap Kepuasan Konsumen Pengguna Jasa Transportasi Bus Damri. *Jurnal Perspektif Bisnis*, 3(2), 89–96. <https://doi.org/10.23960/jpb.v3i2.33>
- Sumantri, R. B. B., & Utami, E. (2020). Penentuan Status Tahapan Keluarga Sejahtera Kecamatan Sidareja Menggunakan Teknik Data Mining. *Jurnal Teknologi Informasi*, 15(3), 71–82. <https://doi.org/10.35842/jtir.v15i3.375>
- T, L. D., & Larose, C. D. (2015). *Data Mining and Predictive Analytics* (Second Edi). Simultaneously.
- Widya Utami, N., & Artana, M. (2022). Text Mining Dalam Analisis Sentimen Pembelajaran Daring Di Masa Pandemi Covid 19 Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Informatika Teknologi Dan Sains*, 4(2), 140–148. <https://doi.org/10.51401/jinteks.v4i2.2034>