

# Evaluasi Performa Aplikasi Gojek Melalui Klasifikasi Kata Ulasan Pengguna Dengan Metode SVM

*Evaluation Of Gojek Application Performance Through User Review Word Classification Using The SVM Method*

Febrianda Putra<sup>1</sup>, Raja Muhammad Ihsan\*<sup>2</sup>, Hafsa Fulaila Tahiyat<sup>3</sup>, Lusiana Efrizoni<sup>4</sup>,  
Rahmadden<sup>5</sup>

<sup>1,2,3,4,5</sup>Teknik Informatika, Universitas Sains dan Teknologi Indonesia

E-mail: <sup>1</sup>febriandaputra11@gmail.com, <sup>2</sup>rajamuhammadihsan02@gmail.com,  
<sup>3</sup>hafsaht1@gmail.com, <sup>4</sup>lusiana@stmik-amik-riau.ac.id, <sup>5</sup>rahmadden@sar.ac.id

**\*Corresponding author**

## Abstrak

Aplikasi Gojek, sebagai salah satu aplikasi ride-hailing terkemuka di Indonesia, menghadapi tantangan berkelanjutan dalam mempertahankan dan meningkatkan kualitas layanan serta kepuasan pengguna. Penelitian ini bertujuan mengevaluasi performa aplikasi Gojek melalui analisis sentimen dari 200.000 ulasan pengguna yang dikumpulkan dari platform aplikasi selama periode 2021-2024, menggunakan metode Support Vector Machine (SVM). Fokus penelitian meliputi aspek layanan, keamanan, promosi, dan masalah teknis. Metode SVM dipilih karena kemampuannya dalam menangani klasifikasi dengan dataset kompleks. Data ulasan pengguna, yang ditulis dalam bahasa Indonesia, diolah dengan teknik pemrosesan teks dan fitur ekstraksi FastText, yang dikenal cepat dan akurat. Praproses data melibatkan tokenisasi, penghapusan stopword, dan teknik preprocessing lainnya. Kinerja model diukur dengan akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa FastText menawarkan waktu pelatihan cepat dengan akurasi kompetitif, sementara SVM unggul dalam presisi dan recall. Penerapan SVM dengan pembagian data 80:20 mencapai akurasi 91% dalam mengklasifikasikan sentimen pengguna. Temuan ini memberikan wawasan penting untuk pengembangan dan evaluasi aplikasi berbasis ulasan pengguna.

Kata kunci: Analisis Sentimen, Aplikasi Gojek, Evaluasi performa, FastText, SVM

## Abstract

*The Gojek application, as one of the leading ride-hailing applications in Indonesia, faces ongoing challenges in maintaining and improving service quality and user satisfaction. This study aims to evaluate the performance of the Gojek application through sentiment analysis from 200,000 user reviews collected from the application platform during the 2021-2024 period, using the Support Vector Machine (SVM) method. The focus of the research includes aspects of service, security, promotion, and technical issues. The SVM method was chosen because of its ability to handle classifications with complex datasets. User review data, written in Indonesian, is processed with text processing techniques and FastText extraction features, which are known for being fast and accurate. Data preprocessing involves tokenization, stopword removal, and other preprocessing techniques. Model performance is measured by accuracy, precision, recall, and F1-score. The results show that FastText offers fast training times with competitive accuracy, while SVMs excel in precision and recall. The implementation of SVM with 80:20 data sharing achieved 91% accuracy in classifying user sentiment. These findings provide important insights for app development and evaluation based on user reviews.*

Keywords: Sentiment Analyst, Gojek App, Performance Evaluation, FastText, SVM

## 1. PENDAHULUAN

Di era digital, berbagai teknologi mendampingi kehidupan kita sehari-hari dengan fungsi masing-masing, seperti telepon genggam yang kini bisa digunakan untuk memesan taksi online. Gojek adalah perusahaan teknologi asal Indonesia yang didirikan pada tahun 2010 oleh Nadiem Makarim, Michaelangelo Moran, dan Kevin Aluwi. Gojek ialah platform yang paling diminati di Asia Tenggara yang menggabungkan beberapa layanan ke dalam satu platform. Di Indonesia, Gojek adalah aplikasi on-demand yang paling populer karena telah meningkatkan kualitas hidup orang-orang di Indonesia. Saat ini, pertumbuhan Gojek juga juga cepat. Dengan bantuan aplikasi Gojek, pengemudi bisa mendapat akses ke lebih banyak pelanggan[1]. Dengan jumlah pengguna yang terus meningkat, ulasan pengguna memainkan peran penting dalam mengevaluasi performa aplikasi.

Klasifikasi adalah langkah menggabungkan serangkaian model untuk mengidentifikasi label kelas, tujuannya adalah untuk memperkirakan kelas objek yang benar. Data yang harus diidentifikasi harus berupa data supervised learning, yakni data yang telah diberi label dan dimasukkan ke dalam model prediktif (dengan hasil berupa nilai variabel target)[2]. Klasifikasi diproses dengan memanfaatkan teknik SVM guna meraih output akhir dalam pengembangan sistem[3]. Algoritma ini merupakan salah satu model dalam machine learning yang dapat mengklasifikasikan teks[4]. SVM ialah suatu teknik pembelajaran mesin yang berusaha mencari hiperplan terbaik untuk memisahkan dua kelas dalam ruang input (input space)[5]. Penggunaan SVM dengan tepat memungkinkan untuk dapat melakukan klasifikasi terhadap banyak kelas[6]. SVM adalah cara terbaik untuk belajar matematika menggunakan hyper-planes[7]. Dikarenakan SVM merupakan sebuah metode klasifikasi, maka diberikan sekumpulan data latihan yang telah diberi label sebagai salah satu dari kategori kelas, algoritma pelatihan SVM akan membuat sebuah model yang dapat mengidentifikasi apakah data baru yang dimasukkan termasuk dalam kategori tertentu atau tidak[8]. Metode SVM digunakan untuk melakukan klasifikasi opini karena hasilnya lebih akurat daripada algoritma lain[9].

Pada penelitian[10], model BiLSTM digunakan untuk analisis sentimen pada ulasan aplikasi Grab, menunjukkan bahwa BiLSTM dapat menghasilkan akurasi yang lebih baik dibandingkan dengan LSTM dalam kasus analisis sentimen. Namun, penelitian ini masih terbatas pada penggunaan BiLSTM dan belum mengeksplorasi metode lain seperti Support Vector Machine (SVM) yang memiliki karakteristik berbeda dalam hal generalisasi dan kemampuan klasifikasi pada dataset yang bervariasi.

Penelitian lain[11], menggunakan SVM dan seleksi fitur chi-square untuk menganalisis ulasan aplikasi dompet digital DANA. Meskipun hasilnya menunjukkan akurasi yang tinggi, pendekatan ini tidak memanfaatkan teknik embedding modern seperti Word2Vec atau FastText, yang dapat memberikan representasi kata yang lebih kaya dan kontekstual.

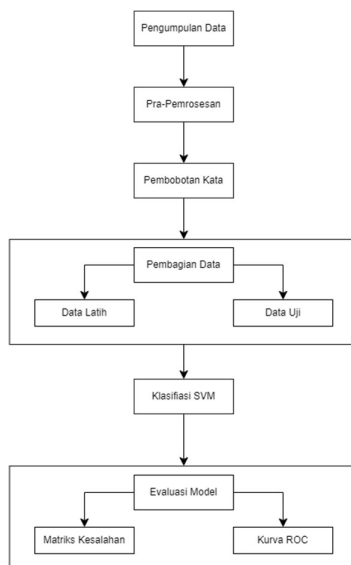
Dalam Penelitian [5], analisis sentimen dilakukan terhadap aplikasi Gojek dan Grab menggunakan SVM dengan ekstraksi fitur Word2Vec. Hasilnya cukup baik untuk kedua aplikasi, namun penggunaan Word2Vec memiliki keterbatasan dalam menangkap nuansa kontekstual kata, terutama dalam bahasa yang kompleks seperti bahasa Indonesia. Oleh karena itu, penelitian ini memperkenalkan FastText sebagai alternatif yang lebih canggih untuk ekstraksi fitur, dengan harapan dapat meningkatkan akurasi dan relevansi hasil klasifikasi.

Tujuan penelitian ini adalah membangun model klasifikasi SVM untuk mengelompokkan ulasan pengguna menjadi kategori positif dan negatif, menggunakan ekstraksi fitur FastText. Penelitian ini bertujuan untuk menganalisa aplikasi gojek dalam meningkatkan kualitas pelayanan gojek serta mengetahui apakah gojek dapat menjadi jasa transportasi online terpercaya[12].

Penelitian ini memperkenalkan penggunaan metode SVM dengan ekstraksi fitur FastText untuk klasifikasi ulasan pengguna Gojek, berbeda dari studi sebelumnya yang umumnya menggunakan teknik sentiment analysis tradisional. Dengan dataset 200.000 ulasan dari 2021 hingga 2024, penelitian ini meningkatkan akurasi klasifikasi dan memberikan wawasan lebih

mendalam tentang aspek-aspek ulasan pengguna yang memengaruhi persepsi kualitas layanan Gojek.

## 2. METODE PENELITIAN



Gambar 1 Metode Penelitian

Penelitian ini menggunakan dataset ulasan pengguna aplikasi Gojek dari kaggle, diproses dengan FastText untuk menghasilkan representasi kata, lalu digunakan sebagai input untuk model Support Vector Machine (SVM). Evaluasi model melibatkan metrik akurasi, presisi, recall, dan F1-score ukuran.

### 2.1 Pengumpulan Data

Pengumpulan data adalah proses sistematis untuk mengumpulkan informasi yang diperlukan guna menjawab pertanyaan atau mencapai tujuan penelitian. Dalam penelitian ini, data dikumpulkan dari platform aplikasi Gojek selama periode 2021-2024 untuk menganalisis ulasan pengguna.

### 2.2 Pra Pemrosesan

Pra-Pemrosesan (Pre-processing) adalah metode data mining yang melibatkan mengkonversi data mentah menjadi format terstruktur yang dapat dipahami. Proses pre-processing adalah langkah awal di mana data yang semula tidak berstruktur diubah menjadi berstruktur melalui 4 tahapan, yaitu, case holding, tokenizing, filtering, dan stemming[6].

#### a. Case Folding

Case folding adalah mengubah karakter huruf besar menjadi huruf kecil, dan menghapus kata-kata yang tidak perlu untuk mengurangi noise[7]. Hal ini dilakukan untuk memastikan bahwa perbedaan dalam huruf besar dan kecil tidak mempengaruhi analisis teks.

#### b. Tokenizing

Tokenizing bertujuan untuk memecah karakter dalam teks menjadi kata-kata untuk mendapatkan fitur dari setiap dokumen setelah dinilai[8].

#### c. Filtering

Dalam pra-proses data teks, filtering adalah langkah untuk menghapus kata atau karakter yang tidak diperlukan dari teks. Ini penting untuk membersihkan data sebelum analisis atau pemodelan, dan biasanya dilakukan setelah tokenisasi untuk meningkatkan kualitas dan relevansi data.

d. *Stemming*

Stemming adalah proses mengubah kata-kata ke bentuk dasarnya atau "stem". Dalam bahasa Indonesia, ini berarti mengubah kata berimbuhan atau turunan menjadi bentuk dasarnya, seperti "berlari" menjadi "lari" dan "makanan" menjadi "makan". Stemming penting dalam pemrosesan bahasa alami karena mengurangi variasi kata dan meningkatkan efisiensi analisis teks.

### 2.3 Pembobotan Kata

Pada fase ketiga, dilakukan pembagian dataset untuk membagi data menjadi dua atau lebih subset data. Biasanya data splitting membagi data menjadi dua bagian, satu bagian untuk evaluasi atau uji data dan bagian lainnya untuk melatih model.[2]. Ekstraksi fitur berguna untuk mengeksplorasi informasi yang mungkin ada dan mengubah kata-kata menjadi vektor fitur. Vektor ini akan dipakai sebagai masukan untuk teknik klasifikasi pada langkah berikutnya[3]. Ekstraksi fitur dilakukan untuk mengambil fitur pada himpunan data[13]. Menggunakan FastText untuk mendapatkan representasi vektor dari kata-kata dalam ulasan.

### 2.4 Pembagian Data

Pembagian dataset merupakan tahapan krusial dalam studi machine learning untuk memastikan bahwa model yang dikembangkan mampu generalisasi dengan baik pada data baru. Dataset dipisah menjadi dua bagian: data training dan data testing.

a. *Pelatihan Data*

Pelatihan Data (Data Training), merupakan tahapan pelatihan model menggunakan data training yang disebut X train dan y train[4]. Sebagian kecil data yang dipilih untuk digunakan dalam pelatihan model machine learning. Informasi latihan ini mencakup contoh-contoh yang sudah diberi label atau output yang diharapkan, yang akan digunakan oleh model untuk belajar dan menghasilkan prediksi pada data yang belum pernah dilihat sebelumnya[14].

b. *Pengujian Data*

Bagian kecil dari kumpulan data digunakan untuk menguji seberapa baik model machine learning bekerja setelah dipelajari dengan data latihannya. Pengujian data berguna untuk menilai kemampuan model untuk menggeneralisasi ke data baru yang belum pernah dilihat sebelumnya. Data tersebut dapat disebut sebagai data yang belum terlihat.[14].

### 2.5 Klasifikasi SVM

Penggunaan pertama Metode Support Vector Machine(SVM) oleh Vapnik pada tahun 1992 membawa inovasi penting dalam pengenalan pola[7]. Metode SVM dipilih untuk klasifikasi opini karena memiliki kinerja yang lebih baik dibandingkan dengan algoritma lain berdasarkan penelitian sebelumnya[9]. SVM memiliki kinerja yang lebih unggul jika dibandingkan dengan metode machine learning lainnya[15].

Garis hyperplane pada SVM memisahkan data menjadi kelas yang berbeda, dengan jumlah titik di setiap sisi menunjukkan distribusi data positif dan negatif. Hyperplane adalah garis (di dua dimensi) atau bidang (di tiga dimensi) yang membagi dua kelas data. Persamaan hyperplane dapat ditulis sebagai:

$$w \times b + x = 0 \quad (1)$$

Keterangan:

- w = vector bobot
- x = vector input
- b = bias atau intercept

### 2.6 Evaluasi Model

Evaluasi pada penelitian ini mengandalkan pada Confusion Matrix (alat ukur performa untuk kasus klasifikasi Machine Learning dengan dua kelas atau lebih) yang digunakan untuk menilai kinerja sistem yang telah dikembangkan berdasarkan hasil klasifikasi. Evaluasi dilakukan di python dengan SVM model digunakan[15].

### 2.7 Analisis Hasil

Analisis hasil mengevaluasi data penelitian untuk menarik kesimpulan bermakna. Penelitian ini menunjukkan bahwa pre-processing efektif dan penggunaan model SVM signifikan dalam klasifikasi ulasan pengguna Gojek, membantu pengembang meningkatkan kepuasan pengguna.

### 2.8 Evaluasi

Evaluasi hasil penelitian ini melibatkan analisis performa model Support Vector Machine (SVM) dalam mengklasifikasikan ulasan aplikasi Gojek berdasarkan data yang telah diproses dengan teknik pre-processing.

## 3. HASIL DAN PEMBAHASAN

Berdasarkan dari metodologi penelitian, maka proses Evaluasi Performa Aplikasi Gojek Melalui Klasifikasi Kata Ulasan Pengguna Dengan Metode Support Vector Machine menghasilkan beberapa hasil.

### 3.1 Pengumpulan Data

Data ulasan pengguna diperoleh dari Kaggle, mencakup komentar dan penilaian terkait performa aplikasi Gojek selama periode 2021-2024. Dataset terdiri dari 200.000 ulasan lebih yang mencakup informasi relevan mengenai pengalaman pengguna.

### 3.2 Pra Pemrosesan

Proses pre-processing yang terdiri dari case folding, tokenizing, filtering, dan stemming berhasil diterapkan pada dataset ulasan pengguna aplikasi Gojek. Hasil dari setiap tahapan pre-processing ditunjukkan pada Tabel 1.

Tabel 1 Proses Pra Pemrosesan

Input	Case Folding	Tokenizing	Filtering	Stemming	Output
Akun GOJEK saya ngebug	akun gojek saya ngebug	[ 'akun', 'gojek', 'saya', 'ngebug' ]	akun gojek ngebug	akun gojek bug	akun gojek bug
SANGAT ramah	sangat ramah	[ 'sangat', 'ramah' ]	ramah	ramah	ramah

Contoh tabel tersebut menunjukkan proses preprocessing teks ulasan pengguna yang terdiri dari beberapa tahap. Pada kalimat "Akun GOJEK saya ngebug", setelah case folding diubah menjadi "akun gojek saya ngebug". Kemudian, tokenizing memecah kalimat tersebut menjadi kata-kata individu: ['akun', 'gojek', 'saya', 'ngebug']. Selanjutnya, filtering menghapus kata-kata yang tidak relevan sehingga menjadi "akun gojek ngebug". Terakhir, stemming mengubah kata-kata menjadi bentuk dasarnya, menghasilkan "akun gojek bug".

### 3.3 Pembobotan Kata

Setelah dilakukan pra-pemrosesan data, dilakukan proses pembobotan kata dengan menggunakan FastText untuk mendapatkan representasi vektor dari kata-kata dalam ulasan.

FastText memungkinkan pembelajaran kata dalam konteks yang lebih luas, termasuk menangani kata yang jarang muncul dan kombinasi kata. Tabel II menunjukkan hasil pembobotan kata.

Tabel 2 Pembobotan kata menggunakan FastText

Kata	Keterangan
Akun GOJEK saya ngebug	Negative
SANGAT ramah	Positive

### 3.4 Pembagian Data

Ada 4 percobaan pembagian data untuk menguji kinerja model SVM dengan berbagai proporsi data training dan testing. Percobaan ini bertujuan menemukan pembagian data yang optimal untuk kinerja model terbaik. Data training yang digunakan yaitu 90%, 80%, 70%, dan 60%. Data testing yang digunakan yaitu 10%, 20%, 30%, dan 40%.pengguna.

### 3.5 Klasifikasi SVM

Penelitian ini menggunakan algoritma SVM, teknik pengklasifikasian yang mencari hyperplane terbaik untuk memisahkan dua label data. Secara sederhana, SVM berusaha menentukan margin maksimal antara data label. Bagian ini membahas tentang model klasifikasi menggunakan Support Vector Machine (SVM).

#### a. Konsep SVM dan Hyperplane

SVM berusaha untuk menemukan hyperplane yang maksimal memisahkan kelas-kelas tersebut.

```
# Plotting the hyperplane
plt.figure(figsize=(10, 7))

# Plotting data points
plt.scatter(X_train_pca[:, 0], X_train_pca[:, 1], c=y_train, cmap='coolwarm', s=30, edgecolors='k')

# Getting the coefficients and intercept of the hyperplane
w = svm_model.coef_[0]
b = svm_model.intercept_[0]

# Calculating the slope and intercept for the hyperplane
slope = -w[0] / w[1]
intercept = -b / w[1]

# Defining the range for the hyperplane
x_vals = np.linspace(min(X_train_pca[:, 0]), max(X_train_pca[:, 0]))

# Calculating y values for the hyperplane
y_vals = slope * x_vals + intercept

# Plotting the hyperplane
plt.plot(x_vals, y_vals, 'k-', label='Hyperplane')

# Adding labels and title
plt.xlabel('PCA Component 1')
plt.ylabel('PCA Component 2')
plt.title('SVM Hyperplane with PCA-reduced Features')
plt.legend()

# Menghitung jumlah misclassification
misclassified = (y_pred != y_test).sum()
total_data = len(y_test)

# Persentase misclassification
misclassification_rate = misclassified / total_data * 100

print(f"Jumlah data yang salah kelas: {misclassified} dari {total_data}")
print(f"Persentase misclassification: {misclassification_rate:.2f}%")

# Displaying the plot
plt.show()
```

Gambar 2 Kode SVM menampilkan Hyperplane

Kode ini menghitung dan mencetak jumlah data positif dan negatif di setiap sisi hyperplane.

#### b. Hasil Evaluasi Model

Evaluasi model dilakukan dengan mengukur metrik kinerja utama seperti akurasi, presisi, recall, dan F1-score pada setiap percobaan pembagian data. Hasil evaluasi untuk berbagai rasio pembagian data disajikan dalam tabel berikut.

Tabel 3 Splitting Data dengan ratio 90 : 10

	<b>Precision (%)</b>	<b>Recall (%)</b>	<b>F1-Score</b>	<b>Support</b>
<b>Positive</b>	0.80	0.85	0.82	5293
<b>Negative</b>	0.95	0.93	0.94	16227
<b>Accuracy</b>			0.91	21520

Untuk pembagian data 90:10, ulasan positif memiliki precision 80%, recall 85%, dan F1-Score 0.82, dengan jumlah ulasan 5.293. Ulasan negatif memiliki precision 95%, recall 93%, dan F1-Score 0.94, dengan jumlah ulasan 16.227

Tabel 4 Splitting Data dengan ratio 80 : 20

	<b>Precision (%)</b>	<b>Recall (%)</b>	<b>F1-Score</b>	<b>Support</b>
<b>Positive</b>	0.81	0.85	0.83	10770
<b>Negative</b>	0.95	0.93	0.94	32270
<b>Accuracy</b>			0.91	43040

Pada pembagian data 80:20, ulasan positif menunjukkan sedikit peningkatan dengan precision 81%, recall tetap 85%, dan F1-Score 0.83, serta jumlah ulasan meningkat menjadi 10.770. Ulasan negatif tetap stabil dengan precision 95%, recall 93%, dan F1-Score 0.94, dengan jumlah ulasan meningkat menjadi 32.270.

Tabel 5 Splitting Data dengan ratio 70 : 30

	<b>Precision (%)</b>	<b>Recall (%)</b>	<b>F1-Score</b>	<b>Support</b>
<b>Positive</b>	0.81	0.85	0.83	16189
<b>Negative</b>	0.95	0.93	0.94	48371
<b>Accuracy</b>			0.91	64560

Dalam pembagian data 70:30, ulasan positif terus menunjukkan precision 81%, recall 85%, dan F1-Score 0.83, dengan jumlah ulasan lebih banyak yaitu 16189. Ulasan negatif tetap menunjukkan precision 95%, recall 93%, dan F1-Score 0.94, dengan jumlah ulasan meningkat menjadi 48371.

Tabel 6 Splitting Data dengan ratio 60 : 40

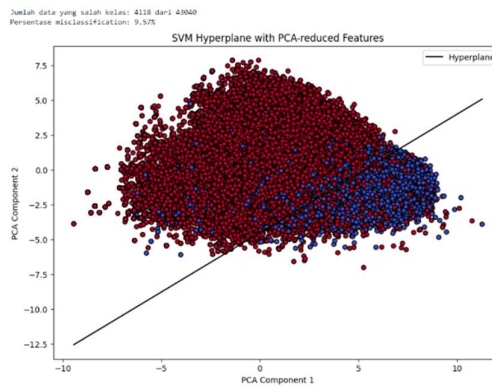
	<b>Precision (%)</b>	<b>Recall (%)</b>	<b>F1-Score</b>	<b>Support</b>
<b>Positive</b>	0.81	0.85	0.83	21604
<b>Negative</b>	0.95	0.93	0.94	64476
<b>Accuracy</b>			0.91	86080

Pada pembagian data 60:40, ulasan positif tetap dengan precision 81%, recall 85%, dan F1-Score 0.83, dengan jumlah ulasan meningkat menjadi 21604. Ulasan negatif juga tetap stabil dengan precision 95%, recall 93%, dan F1-Score 0.94, dengan jumlah ulasan meningkat menjadi 64476.

Berdasarkan hasil pengujian, pembagian data dengan rasio 80:20 memiliki akurasi tertinggi dibandingkan rasio lainnya, yaitu 0.91145. Pembagian data lainnya menggunakan rasio 90:10, 70:30, dan 60:40.

c. Visualisasi Hyperplane

Untuk lebih memahami bagaimana SVM memisahkan data menjadi berbagai kelas, garis hyperplane divisualisasikan. Berikut adalah visualisasi hyperplane bersama dengan jumlah data positif dan negatif di setiap sisi garis tersebut.



Gambar 3 Hyperplane

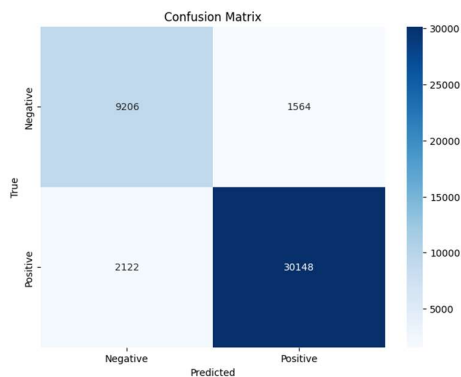
Visualisasi hyperplane menunjukkan pemisahan yang baik antara data positif dan negatif, dengan margin yang jelas antara kedua kelas.

### 3.6 Evaluasi Model

Dalam fase evaluasi ini, kita memanfaatkan sejumlah metrik kinerja utama seperti akurasi, keketatan, pengecualian, dan F1-Score. Selain itu, kita juga menggunakan visualisasi seperti Confusion Matrix dan Kurva ROC untuk memahami kinerja model secara menyeluruh.

a. Confusion Matrix

Confusion Matrix memberikan gambaran yang mendetail tentang cara model mengklasifikasikan data uji. Misalkan kita mendapatkan Confusion Matrix berikut setelah menguji model SVM:



Gambar 4 Heatmap Confusion Matrix

Dari Confusion Matrix ini, kita dapat menghitung beberapa metrik penting:

- Akurasi



$$= \frac{30148 + 9206}{30148 + 9206 + 1564 + 2122} = \frac{39354}{43040} = 0.91 \quad (2)$$

- Presisi
 
$$= \frac{30148}{30148 + 1564} = \frac{30148}{31712} = 0.95 \quad (3)$$

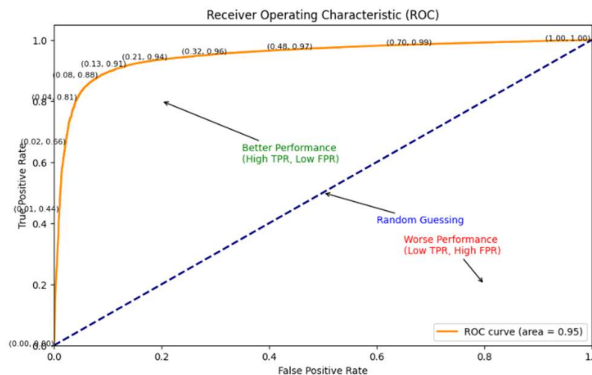
- Recall
 
$$= \frac{30148}{30148 + 9206} = \frac{30148}{39354} = 0.76 \quad (4)$$

- F1-Score
 
$$= \frac{2(0.95 \times 0.76)}{0.95 + 0.76} = \frac{2(0.72)}{1.71} = \frac{1.44}{1.71} = 0.84 \quad (5)$$

F1 Score memberikan keseimbangan antara presisi dan recall, dengan nilai yang lebih tinggi menunjukkan kinerja model yang baik.

b. Kurva ROC

Gambar di bawah ini menunjukkan kurva ROC dari hasil klasifikasi ulasan pengguna aplikasi Gojek menggunakan model SVM. Setiap kurva ROC mewakili performa model dalam membedakan antara kelas positif dan negatif pada berbagai threshold.



Gambar 5 Analisis Kurva ROC

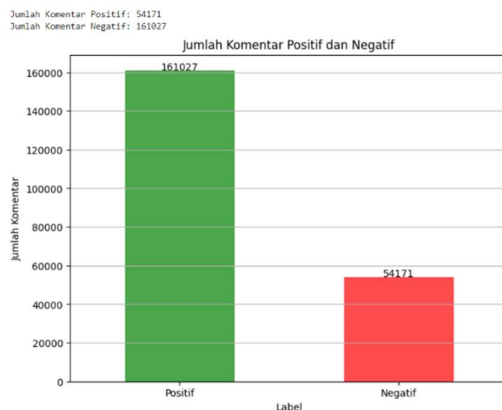
Pada gambar 5 Dengan menggunakan kurva ROC, kita dapat menilai performa model dalam membedakan antara kelas-kelas yang berbeda pada berbagai threshold. Kurva ROC yang mendekati sudut kiri atas menunjukkan performa yang baik, sementara kurva yang mendekati garis diagonal (AUC = 0.5) menunjukkan performa yang buruk atau acak.

- True Positive Rate (TPR)
 
$$= \frac{TP}{TP + FN} = \frac{30148}{30148 + 9206} = \frac{30148}{39354} = 0.76 \quad (6)$$

- False Positive Rate (FPR)
 
$$= \frac{FP}{FP + FN} = \frac{1564}{1564 + 9206} = \frac{1564}{10770} = 0.145 \quad (7)$$

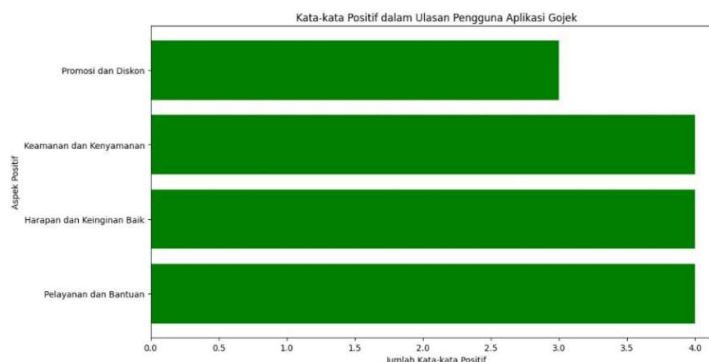
### 3.7 Hasil Analisis

Algoritma SVM dengan pembagian data 80:20 menunjukkan akurasi 91%, dengan 161.027 ulasan positif dan 54.171 ulasan negatif. Model ini efektif dalam mengklasifikasikan ulasan pengguna aplikasi Gojek.



Gambar 6 Jumlah Komentar Positif dan Negatif

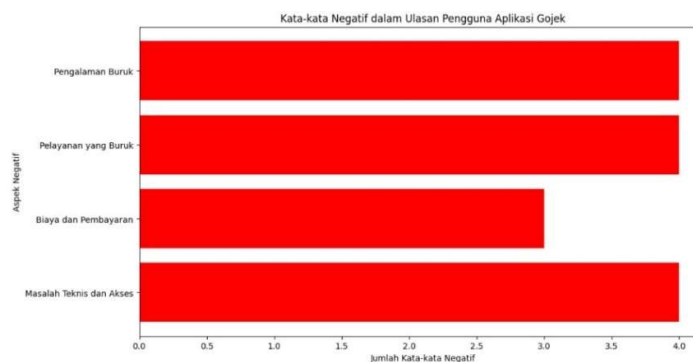
Dari ulasan positif/negative, dan menggunakan metode SVM untuk mengekstraksi kata-kata yang berkontribusi pada fitur-fitur FastText tertentu, didapatkan fitur FastText yang berkontribusi:



Gambar 7 Fitur Positif

Pada Gambar 7, kata-kata positifnya yaitu:

1. Pelayanan dan Bantuan: Kata-kata seperti 'mmbantu', 'terbantu', dan 'membantu' menunjukkan apresiasi terhadap pelayanan dan bantuan yang diberikan oleh aplikasi.
2. Harapan dan Keinginan Baik: Kata-kata seperti 'sukses', 'semoga', dan 'jaya' mencerminkan harapan baik terhadap aplikasi.
3. Keamanan dan Kenyamanan: Kata-kata seperti 'harga', 'denda', dan 'hari' menunjukkan kepuasan terhadap aspek keamanan dan kenyamanan aplikasi.
4. Promosi dan Diskon: Kata-kata seperti 'promonyamahal', 'murah', dan 'kasih' menunjukkan ulasan positif terkait dengan promosi dan diskon.



Gambar 8 Fitur Negatif

Pada Gambar 8, kata-kata negatifnya yaitu:

1. Masalah Teknis dan Akses: Kata-kata seperti 'hp', 'akun', dan 'blokir' menunjukkan masalah teknis dan akses yang sering menjadi alasan ulasan negatif.
2. Biaya dan Pembayaran: Kata-kata seperti 'biaya', 'bayar', dan 'gopay' menunjukkan ketidakpuasan terhadap biaya dan proses pembayaran.
3. Pelayanan yang Buruk: Kata-kata seperti 'kok', 'kenapa', dan 'kecewa' mencerminkan pengalaman buruk terkait dengan pelayanan yang dianggap tidak memuaskan.

#### 4. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengevaluasi performa aplikasi Gojek melalui klasifikasi ulasan pengguna menggunakan metode Support Vector Machine (SVM). Dengan memanfaatkan data ulasan dari Kaggle, serta melaksanakan pra-pemrosesan teks dan pembobotan kata dengan FastText, model SVM berhasil mengidentifikasi ulasan sebagai positif atau negatif dengan tingkat keberhasilan yang tinggi. Evaluasi menunjukkan bahwa model ini berhasil dalam mengolah sentimen, dengan hasil akurasi, presisi, recall, dan F1-Score yang memuaskan.

Berdasarkan hasil penelitian, disarankan kepada Gojek untuk memperhatikan umpan balik dari pengguna dalam rangka meningkatkan kualitas layanan dan pengalaman pengguna. Model ini menunjukkan potensi dalam mengidentifikasi kekuatan dan kelemahan aplikasi melalui analisis sentimen, dan dapat digunakan sebagai alat untuk memantau dan mengevaluasi perubahan layanan berdasarkan ulasan pengguna.

Meskipun hasil penelitian ini memuaskan, terdapat beberapa batasan yang perlu diperhatikan. Batasan tersebut meliputi:

1. Bias Data: Data ulasan mungkin memiliki bias, seperti ulasan yang tidak mewakili seluruh pengguna atau cenderung fokus pada aspek tertentu dari aplikasi.
2. Pra-pemrosesan: Meskipun metode pra-pemrosesan teks digunakan, ada kemungkinan informasi penting hilang selama proses ini.
3. Kualitas Model: Hasil model sangat bergantung pada parameter dan teknik yang digunakan, dan mungkin ada potensi untuk perbaikan lebih lanjut dengan eksperimen tambahan.

Penelitian lebih lanjut dapat mencakup eksplorasi teknik lain seperti deep learning atau pembelajaran transfer, serta penggunaan data yang lebih bervariasi untuk meningkatkan akurasi dan generalisasi model. Dengan demikian, Gojek dapat lebih baik dalam merespons kebutuhan dan harapan pengguna mereka.

#### DAFTAR PUSTAKA

- [1] A. Fauzi *et al.*, "Transformasi Gojek untuk Keunggulan Kompetitif Dalam Perkembangan Ekonomi (Literature Review Metodologi Riset Bisnis)," *J. Ilmu Multidisiplin*, vol. 1, no. 3, pp. 705–716, 2022.

- [2] I. S. Milal, M. H. M. Hasanudin, M. A. Nur Azhari, R. A. Nugraha, N. Agustina, and S. E. Damayanti, "Klasifikasi Teks Review Pada E-Commerce Tokopedia Menggunakan Algoritma Svm," *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 5, no. 1, pp. 34–45, 2023, doi: 10.53580/naratif.v5i1.191.
- [3] H. N. Irmanda and Ria Astriratma, "Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 915–922, 2020, doi: 10.29207/resti.v4i5.2313.
- [4] M. F. Naufal, T. Arifin, and H. Wirjawan, "Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan Naïve Bayes untuk Klasifikasi Cyberbullying pada Media Sosial," *Jurasik*, vol. 8, p. 82, 2023, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- [5] E. Suryati, Styawati, and A. A. Aldino, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," *J. Teknol. Dan Sist. Inf.*, vol. 4, no. 1, pp. 96–106, 2023, [Online]. Available: <https://doi.org/10.33365/jtsi.v4i1.2445>
- [6] S. N. Cahyani and G. W. Saraswati, "Penerapan Metode Support Vector Machine Pada Klasifikasi Buku Perpustakaan Sekolah Dengan Kombinasi TF-IDF Dan Word2Vec," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1555–1566, 2023.
- [7] Y. El Saputra, S. Agustian, and S. Ramadhani, "Klasifikasi Sentimen SVM Dengan Dataset yang Kecil Pada Kasus Kaesang Sebagai Ketua Umum PSI," vol. 4, no. 6, pp. 2902–2908, 2024, doi: 10.30865/klik.v4i6.1944.
- [8] Oryza Habibie Rahman, Gunawan Abdillah, and Agus Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 17–23, 2021, doi: 10.29207/resti.v5i1.2700.
- [9] T. P. Lestari, "Analisis Text Mining pada Sosial Media Twitter Menggunakan Metode Support Vector Machine (SVM) dan Social Network Analysis (SNA)," *J. Inform. Ekon. Bisnis*, vol. 4, no. 3, pp. 65–71, 2022, doi: 10.37034/infv.v4i3.146.
- [10] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia," *J. Manaj. Inform.*, vol. 12, no. 2, pp. 89–99, 2022, doi: 10.34010/jamika.v12i2.7764.
- [11] A. A. Muhammad, Ermatita, and D. S. Prasvita, "ANALISIS SENTIMEN PENGGUNA APLIKASI DANA BERDASARKAN ULASAN PADA GOOGLE PLAY MENGGUNAKAN METODE SUPPORT VECTOR MACHINE Prodi S1 Informatika / Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 194–204, 2022.
- [12] K. D. Indarwati and H. Februariyanti, "Analisis Sentimen Terhadap Kualitas Pelayanan Aplikasi Gojek Menggunakan Metode Naive Bayes Classifier," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 1, 2023, doi: 10.35957/jatisi.v10i1.2643.
- [13] S. D. Lestari and E. B. Setiawan, "Sentiment Analysis Based on Aspects Using FastText Feature Expansion and NBSVM Classification Method," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 469–477, 2022, doi: 10.47065/josyc.v3i4.2202.
- [14] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, and L. Efrizoni, "Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 273–281, 2024, doi: 10.57152/malcom.v4i1.1085.
- [15] D. Oktavia and Y. R. Ramadahan, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Media Online*, vol. 4, no. 1, pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.