LINIER
LITERATUR INFORMATIKA & KOMPUTER
e-ISSN: 3063-2218
DOI : 10.33096/linier.v2i2.3132
Literatur Informatika & Komputer
Vol 2, No 2 (2025), pp 234-245

Analisis Sentimen Pengguna Gojek Berdasarkan Ulasan pada App Store dengan Metode KNN, Naive Bayes, dan SVM

Arif Kurnia[a], Harlinda[b], Herdianti Darwis[c]
Universitas Muslim Indonesia, Makassar, Indonesia
[a]13120200019@umi.ac.id; [b]harlinda@umi.ac.id; [c]herdianti.darwis@umi.ac.id

Received: 17-01-2025 | Revised: 02-06-2025 | Accepted: 19-06-2025 | Published: 29-06-2025

Abstrak

Gojek adalah aplikasi layanan *on-demand* yang telah menjadi salah satu platform terbesar di Asia Tenggara dengan jutaan pengguna aktif dan 5 juta ulasan di *App Store*. Ulasan ini menjadi sumber informasi penting untuk mengevaluasi dan meningkatkan layanan. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna Gojek dengan mengelompokkan ulasan menjadi lima kelas sentimen: "Sangat Puas", "Puas", "Cukup", "Buruk", dan "Sangat Buruk". Metode yang digunakan meliputi *K-Nearest Neighbors* (KNN), *Naive Bayes*, dan *Support Vector Machine* (SVM). Setelah melakukan *text preprocessing*, ketiga metode tersebut dievaluasi berdasarkan *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model SVM dengan kernel Linear mencapai akurasi tertinggi sebesar 79.00%, diikuti kernel RBF dengan precision tertinggi sebesar 83.85%. Model *Naive Bayes* menunjukkan performa cukup baik dengan akurasi 78.00%, sementara KNN memiliki akurasi terendah sebesar 69.25%. Berdasarkan hasil ini, SVM, khususnya dengan kernel Linear dan RBF, terbukti menjadi metode paling efektif dalam menganalisis sentimen pengguna Gojek, memberikan wawasan yang lebih akurat untuk perbaikan layanan.

Kata kunci: *Sentiment Analysis*, *SVM, Naive Bayes, K-Nearest Neighbors, Gojek*.

Pendahuluan

Gojek adalah sebuah aplikasi layanan *on-demand* yang pertama kali diluncurkan di Indonesia pada tahun 2010. Aplikasi ini menawarkan berbagai layanan, mulai dari transportasi, pengiriman makanan, belanja kebutuhan sehari-hari, hingga pembayaran digital, yang semuanya diintegrasikan dalam satu platform. Sejak awal peluncurannya, Gojek telah berkembang pesat dan kini menjadi salah satu platform terkemuka di Asia Tenggara, dengan jutaan pengguna aktif harian. Hingga saat ini, aplikasi Gojek di *App Store* telah menerima sebanyak 5 juta ulasan dengan rata-rata *rating* sebesar 4.6 dari skala 5. Ulasan-ulasan ini mencerminkan berbagai pengalaman dan tingkat kepuasan pengguna terhadap layanan yang disediakan oleh Gojek.

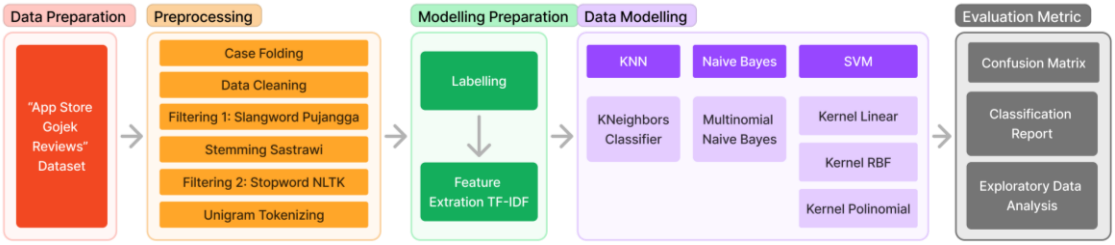
Ulasan pengguna merupakan elemen yang sangat penting dalam menjaga kualitas dan relevansi produk di pasar. Melalui ulasan, pengguna dapat mengekspresikan kepuasan maupun ketidakpuasan mereka terhadap layanan yang diterima, sekaligus memberikan masukan yang berharga untuk pengembang. Bagi perusahaan, ulasan ini menjadi sumber informasi kritis untuk mengevaluasi dan meningkatkan layanan secara berkesinambungan. Dengan demikian, ulasan pengguna tidak hanya menjadi indikator kinerja produk tetapi juga sebagai kompas untuk memastikan layanan yang diberikan tetap sesuai dengan harapan pasar dan berkembang seiring dengan kebutuhan pengguna.

Analisis sentimen adalah teknik yang digunakan untuk mengidentifikasi dan mengekstraksi opini, emosi, atau sentimen dari teks, seperti ulasan pengguna [1]. Dalam konteks aplikasi Gojek, analisis sentimen berperan penting dalam memahami bagaimana pengguna merespons berbagai aspek layanan yang disediakan [2]. Teknik ini memungkinkan pengelompokan ulasan berdasarkan sentimen positif, negatif, atau netral, akan tetapi pada penelitian ini, pengelompokan data dibagi menjadi 5 kelas, yaitu “Sangat Puas”, “Puas”, “Cukup”, “Buruk”, “Sangat Buruk” yang pada akhirnya dapat memberikan wawasan mendalam mengenai persepsi umum pengguna. Informasi ini sangat penting dalam membantu manajemen untuk merancang strategi peningkatan layanan yang berbasis data dan memfasilitasi pengambilan keputusan yang lebih tepat.

Penelitian ini akan memanfaatkan beberapa metode klasifikasi yang umum digunakan dalam analisis sentimen, yaitu *K-Nearest Neighbors* (KNN) [3], [4], [5], *Naive Bayes* [6], [7], dan *Support Vector Machine* (SVM) [8], [9]. Ketiga metode ini dipilih berdasarkan keunggulannya masing-masing dalam menangani data teks. KNN dikenal karena kesederhanaan dan kemampuannya dalam mengkategorikan data dengan presisi yang baik

dalam banyak kasus. *Naive Bayes* sering digunakan karena kecepatan dan akurasinya yang tinggi dalam klasifikasi teks. Sementara itu, SVM merupakan metode yang kuat dalam mengatasi data dengan dimensi tinggi dan sering kali memberikan hasil klasifikasi yang sangat akurat. Dengan menerapkan ketiga metode ini, penelitian ini bertujuan untuk mengevaluasi efektivitas dan akurasi dari setiap metode dalam menganalisis sentimen pengguna terhadap aplikasi Gojek.

Metode



Gambar 1. Desain Penelitian

Gambar 1 memperlihatkan alur desain penelitian yang digunakan dalam analisis sentimen pengguna Gojek berdasarkan ulasan pada *App Store*. Proses penelitian dimulai dengan *Data Preparation*, di mana data ulasan dari *App Store* diambil dan dikumpulkan. Tahap berikutnya adalah *Preprocessing* [10], [11], yang meliputi berbagai langkah untuk membersihkan dan mempersiapkan data teks, seperti *case folding*, *data cleaning, filtering slangword* menggunakan data dari Pujangga, serta *filtering stopwords* menggunakan NLTK dan *punctuations*. Setelah *preprocessing*, langkah sebelum pemodelan dilakukan, yang melibatkan pemberian label pada data serta ekstraksi fitur menggunakan teknik TF-IDF. Selanjutnya, data yang sudah diproses dimasukkan ke dalam tahap *Data Modeling*, di mana tiga metode klasifikasi, yaitu KNN, *Naive Bayes*, dan SVM, diterapkan untuk membangun model analisis sentimen. Akhirnya, hasil model dievaluasi melalui *Evaluation Metrics* yang mencakup *confusion matrix* dan *classification report* untuk menilai akurasi dan efektivitas model yang dibangun. Serta *exploratory data analysis* untuk melihat pola-pola ataupun *insight* pada *dataset* yang digunakan.

A. *Dataset* yang digunakan

Dataset yang digunakan dalam penelitian ini diperoleh dari *Kaggle* dengan judul "*App Store Gojek Reviews*" yang dirilis oleh Imam Satrio pada tanggal 27 Mei 2024. *Dataset* ini berisi 2000 ulasan pengguna mengenai aplikasi Gojek yang diunduh dari *App Store*. *Dataset* ini mencakup beberapa kolom penting yang digunakan dalam analisis, yaitu: date, yang menunjukkan tanggal ketika ulasan dikirimkan; *review*, yang berisi isi dari ulasan pengguna; *rating*, yang merupakan jumlah bintang yang diberikan oleh pengguna sebagai indikator kepuasan mereka; dan *title*, yang merangkum judul atau poin utama dari ulasan yang diberikan. Kolom-kolom ini menjadi dasar untuk melakukan analisis sentimen, di mana setiap ulasan akan dianalisis untuk memahami pola dan kecenderungan sentimen yang terkandung di dalamnya. Dan pada Tabel 1 telah termuat beberapa sampel dari dataset yang digunakan pada penelitian ini.

Tabel 1. Sampel *dataset*

#	Date	Review	Rating	Title
0	2017-09-11 12:10:16	Good	5	Awesome
1	2018-11-09 11:16:25	Kenapa tidak ada opsi login with email??? Tolo...	4	Login with Email
2	2019-05-18 09:11:13	Beberapa minggu ini jadi sering force close, g...	1	Force close di iPhone 6 Low Power Mode
3	2020-01-11 05:56:18	1. Go mart nya kok di hilangin kenapa sih, udh...	1	Gojek kamu telah berubah... :(
4	2019-12-24 03:36:02	Promo di hp tiap ornag beda, ada yg ada promon...	1	Brenti mainin harga

B. *Preprocessing*

Proses *preprocessing* merupakan langkah krusial dalam analisis sentimen yang bertujuan untuk mempersiapkan data teks agar dapat diolah dengan lebih efektif. Beberapa tahap *preprocessing* yang dilakukan dalam penelitian ini meliputi:

1. *Case Folding*
   Pada tahap ini, seluruh teks dalam ulasan dikonversi menjadi huruf kecil (*lowercase*). Tujuannya adalah untuk memastikan bahwa kata-kata yang seharusnya sama tidak dianggap berbeda oleh sistem hanya karena perbedaan huruf kapital, seperti "Gojek" dan "gojek".
2. *Data Cleaning*
   Langkah ini mencakup pembersihan teks dari karakter-karakter yang tidak relevan, seperti tanda baca, angka, atau simbol khusus. Hal ini dilakukan untuk menghilangkan elemen-elemen yang tidak memberikan nilai informasi yang signifikan dalam analisis sentimen.
3. *Slangword Pujangga*
   Proses *filtering* ini bertujuan untuk menangani kata-kata slang atau bahasa gaul yang sering muncul dalam ulasan pengguna. Dengan menggunakan kamus "*Slangword* Pujangga," kata-kata slang yang ada diubah menjadi bentuk standar agar lebih mudah dianalisis [12].
4. *Stemming Sastrawi*
   Pada tahap *stemming*, kata-kata dalam ulasan dikonversi ke bentuk dasar atau akarnya menggunakan *library* Sastrawi [13]. Misalnya, kata "berjalan" akan diubah menjadi "jalan". Ini penting untuk mengurangi variasi kata yang muncul akibat infleksi dan memastikan bahwa semua bentuk kata dihitung sebagai satu entitas.
5. *Stopword* NLTK
   Filtering kedua dilakukan untuk menghapus *stopwords*, yaitu kata-kata umum yang tidak memberikan kontribusi signifikan terhadap pemahaman makna teks, seperti "dan", "yang", "dari", dan sebagainya. Proses ini dilakukan menggunakan pustaka *Natural Language Toolkit* (NLTK) [14], yang memiliki daftar *stopwords* bahasa Indonesia yang komprehensif.
6. *Unigram Tokenizing*
   Tahap terakhir dalam *preprocessing* adalah tokenisasi *unigram*, di mana teks dibagi menjadi *token-token* kata tunggal (*unigram*) [15]. Setiap kata dalam teks diperlakukan sebagai unit analisis yang berdiri sendiri, yang kemudian dapat digunakan untuk ekstraksi fitur atau analisis lebih lanjut.

C. Pelabelan NLTK

Pada penelitian ini, pelabelan sentimen dilakukan dengan memanfaatkan NLTK, yang merupakan salah satu pustaka *Python* yang paling umum digunakan untuk pemrosesan bahasa alami. Langkah pertama dalam proses ini adalah menghitung jumlah kemunculan (*occurrences*) dari setiap kategori sentimen berdasarkan nilai rating yang diberikan oleh pengguna. Setiap ulasan diberi label sentimen berdasarkan skala rating, di mana rating yang lebih rendah (misalnya, 1 dan 2) dikategorikan sebagai sentimen negatif, rating sedang (misalnya, 3) sebagai netral, dan rating yang lebih tinggi (misalnya, 4 dan 5) dikategorikan sebagai sentimen positif. Pada penelitian ini, pengelompokan data dibagi menjadi 5 kelas, yaitu “Sangat Puas”, “Puas”, “Cukup”, “Buruk”, “Sangat Buruk”.

D. Fitur Ekstraksi TF-IDF

Dalam penelitian ini, ekstraksi fitur dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) [16]. Metode TF-IDF adalah teknik yang digunakan untuk mengevaluasi seberapa penting suatu kata dalam dokumen dibandingkan dengan seluruh koleksi dokumen. Skor TF-IDF untuk sebuah kata t dalam dokumen d dihitung dengan mengalikan *Term Frequency* (TF) dengan *Inverse Document Frequency* (IDF). Rumus untuk TF-IDF terlihat pada formula 1.

$$TF - IDF(t,d) = \left(\frac{f(t,d)}{|d|}\right) \times log\left(\frac{|D|}{|d_t|}\right) \quad (1)$$

Metode ini memungkinkan identifikasi kata-kata yang signifikan untuk analisis sentimen dengan memberikan bobot yang lebih tinggi pada kata-kata yang jarang tetapi relevan di seluruh koleksi dokumen.

E. Analisis Eksplorasi Data

Analisis Eksplorasi Data atau biasa dikenal dengan *Exploratory Data Analysis* (EDA) adalah tahap awal yang krusial dalam proses analisis data yang bertujuan untuk memahami karakteristik, pola, dan hubungan dalam *dataset* sebelum menerapkan model pembelajaran mesin. Dalam konteks penelitian ini, EDA dilakukan untuk mendapatkan wawasan mendalam tentang ulasan pengguna aplikasi Gojek yang tersedia. Proses ini melibatkan berbagai teknik visualisasi dan statistik untuk menganalisis distribusi kata, frekuensi kemunculan kata, dan pola umum dalam data.

F. Preferensi Pemodelan

Dalam penelitian ini, data dibagi menjadi dua bagian, yaitu data pelatihan (*training data*) dan data pengujian (*test data*), untuk memastikan bahwa model yang dikembangkan dapat dievaluasi dengan baik. Sebanyak 20% dari keseluruhan *dataset* dialokasikan sebagai data pengujian, sementara sisanya 80% digunakan sebagai data pelatihan. Pemisahan ini dilakukan secara acak dengan menggunakan *random state* yang diatur pada nilai 42. Pengaturan *random state* ini dimaksudkan untuk memastikan bahwa proses pembagian data dilakukan secara konsisten setiap kali model dijalankan, sehingga hasil evaluasi dapat direproduksi dengan hasil yang sama. Pemodelan dengan pendekatan ini memungkinkan untuk menguji keandalan dan generalisasi model pada data yang belum pernah dilihat sebelumnya, sehingga memberikan gambaran yang lebih akurat mengenai kinerja model dalam kondisi nyata.

G. Algoritma Pemodelan

Berikut adalah algoritma pemodelan yang digunakan dalam penelitian ini untuk analisis sentimen ulasan pengguna aplikasi Gojek:

1. *K-Nearest Neighbors*

K-Nearest Neighbors (KNN) adalah algoritma klasifikasi berbasis *instance* yang bekerja dengan mencari *k* tetangga terdekat dari titik data yang belum diklasifikasikan [3]. Algoritma ini menentukan label dari titik data dengan melihat *k* tetangga terdekat dan memilih label yang paling sering muncul di antara mereka. Jarak antar titik data biasanya diukur dengan metrik seperti jarak *Euclidean*, dan pemilihan nilai *k* yang tepat dapat mempengaruhi kinerja model. Jarak antar titik data biasanya diukur dengan metrik seperti jarak *Euclidean*, yang dirumuskan pada Formula 2 berikut:

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

di mana $d\ (x,\ y)$ adalah jarak antara dua titik x dan y, dan x_i serta y_i merupakan koordinat dari masing-masing titik pada dimensi ke- i. Semakin kecil nilai $d\ (x,\ y)$, semakin dekat kedua titik tersebut.

2. *Naive Bayes*

Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang didasarkan pada *Teorema Bayes* dengan asumsi independensi antar fitur [17]. Dalam penelitian ini, digunakan *Multinomial Naive Bayes* [18]. Algoritma ini menghitung probabilitas klasifikasi dengan asumsi bahwa fitur-fitur (kata-kata) mengikuti distribusi *multinomial*. Ini sangat efektif untuk data teks dan menghitung probabilitas dengan mempertimbangkan frekuensi kata dalam dokumen serta jumlah dokumen di seluruh korpus. Pada Multinomial Naive Bayes, kita mengasumsikan bahwa fitur-fitur berupa kata dalam teks mengikuti distribusi multinomial. Probabilitas setiap fitur x_i dalam dokumen dihitung sebagai:

$$P(x_i|C) = \frac{n_{i,C} + 1}{n_C + |V|} \quad (3)$$

di mana formula 3 mejelaskan bahwa $P(x_i|C)$ adalah probabilitas suatu kata x_i muncul dalam kelas $C, n_{i,C}$ adalah frekuensi kata x_i dalam dokumen yang termasuk kelas C, n_C adalah jumlah total kata dalam dokumen di kelas C, dan $|V|$ adalah ukuran dari kosakata (jumlah total kata unik dalam seluruh korpus). Penambahan 1 pada pembilang dikenal sebagai "Laplace Smoothing" yang berfungsi untuk mengatasi masalah ketika suatu kata tidak muncul dalam suatu kelas (probabilitas menjadi nol).

3. *Support Vector Machine*

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi [19]. SVM bekerja dengan mencari *hyperplane* terbaik yang memisahkan kelas-kelas data dalam ruang fitur. Untuk menentukan *hyperplane*, kita menggunakan fungsi *kernel*, yang memungkinkan SVM untuk bekerja baik pada data yang dapat dipisahkan secara *linear* maupun *non-linear*.

a. *Kernel Linear*

Digunakan ketika data dapat dipisahkan secara linear. Fungsi *kernel linear* dirumuskan sebagai:

$$K(x_i, x_j) = x_i \cdot x_j \quad (4)$$

di mana Formula 4 menjelaskan bahwa x_i dan x_j adalah vektor fitur, dan operasi ini menghasilkan *dot product* antara keduanya. Kernel ini digunakan untuk mencari *hyperplane linear* yang memisahkan data.

b. *Kernel Radial Basis Function*

Juga dikenal sebagai *Gaussian Kernel*, digunakan untuk data yang tidak dapat dipisahkan secara *linear* dengan memproyeksikannya ke ruang dimensi yang lebih tinggi. Rumus *kernel* RBF adalah:

$$K(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2) \quad (5)$$

di mana Formula 5 menjelaskan bahwa $|x_i - x_j|^2$ adalah jarak *Euclidean* antara dua vektor fitur, dan γ adalah parameter yang mengontrol lebar kernel. Kernel ini memungkinkan SVM untuk menangani hubungan non-linear dengan lebih baik.

c. *Kernel Polinomial*

Kernel ini mengubah ruang fitur dengan menggunakan *polinomial* dari derajat tertentu untuk menemukan hubungan yang lebih kompleks antara fitur. Rumus *kernel polinomial* adalah:

$$K(x_i, x_j) = (x_i \cdot x_j + c)^d \quad (6)$$

di mana Formula 6 menjelaskan bahwa d adalah derajat *polinomial*, c adalah konstanta yang memengaruhi pengaruh fitur tingkat rendah, dan $x_i \cdot x_j$ adalah *dot product* dari vektor fitur. Kernel ini sangat berguna untuk data dengan hubungan *non-linear* yang lebih kompleks.

H. Metrik Evaluasi

Dalam penelitian ini, evaluasi kinerja model dilakukan dengan menggunakan dua metrik utama: *confusion matrix* dan *classification report* [20]. *Confusion matrix* [21] adalah alat yang digunakan untuk menilai kinerja model klasifikasi dengan menunjukkan jumlah prediksi yang benar dan salah pada masing-masing kelas. Matriks ini menyediakan informasi detail mengenai *true positives* (TP), *false positives* (FP), *true negatives* (TN), dan *false negatives* (FN), yang memungkinkan analisis mendalam tentang kesalahan klasifikasi dan performa model secara keseluruhan. Selain itu, *classification report* memberikan metrik tambahan yang mencakup *precision*, *recall*, dan *F1-score* untuk setiap kelas:

a. *Accuracy*

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Accuracy mengukur proporsi prediksi yang benar (baik positif maupun negatif) dari total prediksi yang dilakukan oleh model.

b. Precision

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

Ini mengukur proporsi prediksi positif yang benar dari semua prediksi positif yang dilakukan oleh model.

c. Recall

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

Ini mengukur proporsi prediksi positif yang benar dari semua data positif yang sebenarnya.

d. F1-Score

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

F1-Score adalah ukuran harmonis dari *precision* dan *recall*, yang digunakan untuk memberikan gambaran yang lebih baik ketika ada ketidakseimbangan antara FP dan FN. Kombinasi dari *confusion matrix* dan *classification report* memberikan pandangan komprehensif mengenai efektivitas model dalam klasifikasi, terutama dalam menangani masalah data tidak seimbang.

Perancangan

A. Preprocessing

Tabel 2 menggambarkan langkah-langkah yang dilakukan dalam proses *text preprocessing* untuk ulasan pengguna aplikasi Gojek. Proses ini dimulai dengan *case folding* dan *data cleaning*, di mana teks asli diubah menjadi huruf kecil dan dibersihkan dari karakter yang tidak relevan, seperti *emoji*. Selanjutnya, tahap *filtering* dilakukan untuk menghilangkan kata-kata umum yang tidak berkontribusi signifikan terhadap analisis, sehingga hanya kata-kata kunci yang relevan yang tersisa. Setelah itu, proses *stemming* diterapkan untuk mengubah kata-kata menjadi bentuk dasarnya, mengurangi variasi kata dengan makna yang sama. Terakhir, teks yang telah diproses melalui tahap-tahap sebelumnya diubah menjadi bentuk *unigram tokenizing*, di mana setiap kata direpresentasikan sebagai *token* individu dalam bentuk daftar. Langkah-langkah ini penting untuk memastikan bahwa teks yang dianalisis lebih bersih, terstruktur, dan siap untuk dimasukkan ke dalam model pembelajaran mesin.

Tabel 2. *Text Preprocessing*

Original	***Case Folding & Data Cleaning***	***Filtering 1***	***Stemming***	***Unigram Tokenizing***
Makin kesini makin jelek banget yaa 😭😭	makin kesini makin jelek banget yaa	kesini jelek banget	kesini jelek banget	[kesini, jelek, banget]
Makin lama makin mahal, jarang bgt ada diskon wkwm	makin lama makin mahal jarang bgt ada diskon wkwm	mahal mahal jarang banget diskon	mahal mahal jarang banget diskon	[mahal, mahal, jarang, banget, diskon]
Chat di dalam aplikasi sering error. Nggak keluar fitur chat, sehingga harus sms/telpon	chat di dalam aplikasi sering error nggak keluar fitur chat sehingga harus sms telpon	chat error chat aplikasi error fitur chat sehingga sms telpon	chat error chat aplikasi error fitur chat hingga sms telpon	[chat, error, chat, aplikasi, error, fitur, chat, hingga, sms, telpon]

B. Analisis Eksplorasi Data

1. Word Cloud

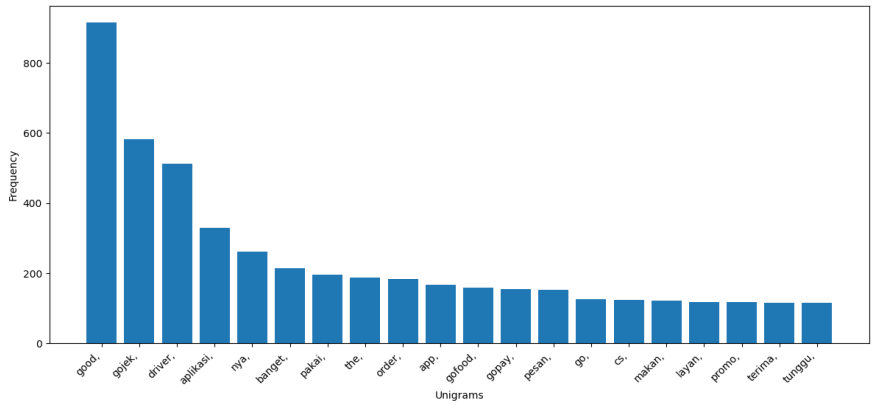
Gambar 2 menampilkan visualisasi *word cloud* yang dihasilkan dari kata-kata yang paling sering muncul dalam ulasan pengguna aplikasi Gojek yang diambil dari App Store . Semakin besar ukuran

kata dalam *word cloud*, semakin sering kata tersebut muncul dalam ulasan pengguna. Beberapa kata yang paling menonjol seperti "*good*", "*driver*", "gojek", dan "aplikasi" mengindikasikan bahwa pengguna sering kali membahas tentang kualitas aplikasi, layanan pengemudi, dan pengalaman mereka secara umum. Kata-kata seperti "*order*," "promo," dan "bantu" juga sering muncul, menunjukkan bahwa fitur pemesanan dan promosi merupakan aspek penting yang dibicarakan oleh pengguna. *Word cloud* ini membantu memberikan gambaran awal mengenai topik dan isu utama yang diangkat oleh pengguna dalam ulasan mereka, yang kemudian dapat dianalisis lebih lanjut untuk memahami sentimen dan persepsi mereka terhadap layanan Gojek.

Gambar 2. *Word Cloud*

2. *Unigram Frequency*

Gambar 3 menampilkan distribusi frekuensi *unigram*, yang merupakan kata-kata tunggal yang paling sering muncul dalam ulasan pengguna aplikasi Gojek di *App Store*. Pada grafik ini, sumbu horizontal (X) mewakili kata-kata unigram, sementara sumbu vertikal (Y) menunjukkan frekuensi kemunculan masing-masing kata. Kata "*good*" muncul dengan frekuensi tertinggi, menandakan bahwa banyak pengguna yang menggunakan kata ini dalam ulasan mereka, kemungkinan besar untuk memberikan penilaian positif terhadap layanan Gojek. Kata-kata lain seperti "*driver*", "aplikasi", dan "pesan", juga memiliki frekuensi tinggi, mencerminkan fokus pengguna pada aspek-aspek penting seperti kualitas layanan pengemudi, aplikasi itu sendiri, dan fitur pemesanan. Distribusi frekuensi *unigram* ini memberikan wawasan awal mengenai tema-tema utama yang disorot dalam ulasan pengguna, yang kemudian dapat dihubungkan dengan sentimen yang lebih mendalam dan analisis tekstual lainnya.

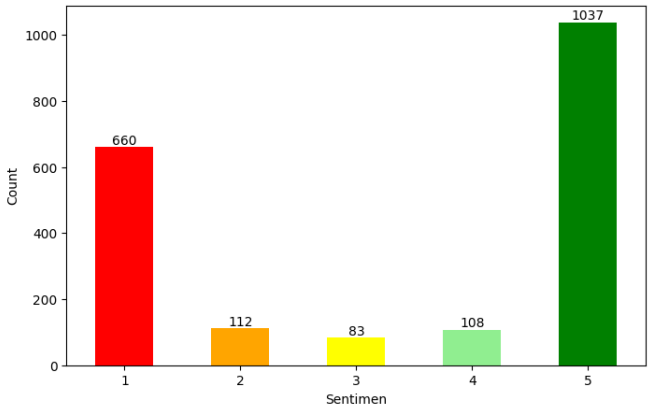


Gambar 3. *Unigram Frequency*

3. *Sentiment Polarity*

Gambar 4 menunjukkan distribusi polaritas sentimen berdasarkan pelabelan menggunakan NLTK. Grafik ini menggambarkan jumlah ulasan yang diklasifikasikan ke dalam lima kategori sentimen yang berbeda, mulai dari 1 hingga 5, di mana 1 mewakili sentimen paling negatif dan 5 mewakili sentimen

paling positif. Dari grafik ini, terlihat bahwa mayoritas ulasan memiliki sentimen yang sangat positif dengan *rating* 5, yang mencapai jumlah tertinggi sebanyak 1037 ulasan. Sebaliknya, ulasan dengan *rating* 1, yang menunjukkan sentimen sangat negatif, juga cukup signifikan dengan jumlah 660 ulasan. Kategori sentimen netral, yang direpresentasikan oleh *rating* 3, memiliki jumlah ulasan yang paling sedikit, yaitu 83 ulasan. Perbedaan distribusi ini mencerminkan bahwa ulasan pengguna cenderung berada di ujung spektrum sentimen, dengan lebih banyak ulasan yang sangat positif dan negatif, serta lebih sedikit ulasan yang bersifat netral.



Gambar 4. Polaritas Sentimen berdasarkan label NLTK

Pemodelan

A. Confusion Matrix

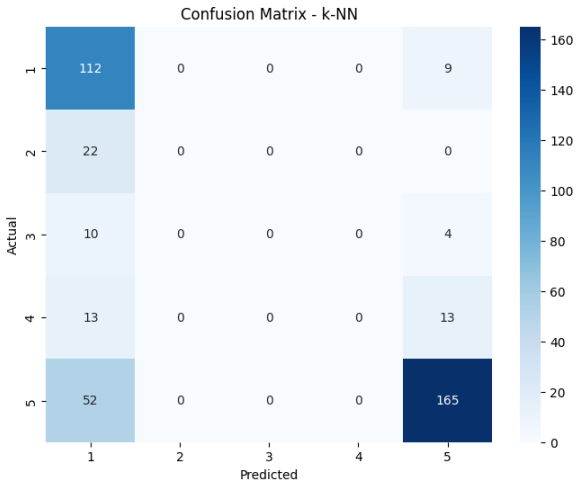
1. *K-Nearest Neighbors*

Confusion matrix untuk model KNN menunjukkan seberapa baik model KNN dalam memprediksi berbagai kategori sentimen, yaitu "Sangat Buruk", "Buruk", "Cukup", "Puas", dan "Sangat Puas". Model KNN mampu memprediksi ulasan "Sangat Buruk" dengan akurasi yang cukup baik, meskipun ada beberapa kesalahan dalam memprediksi kategori lain. Untuk kategori "Buruk", model juga menunjukkan performa yang cukup memadai, tetapi terdapat beberapa kesalahan prediksi, terutama pada kategori "Cukup". Pada kategori "Cukup", model ini menunjukkan kinerja terbaik dengan jumlah prediksi benar yang paling tinggi, meskipun masih terdapat beberapa kesalahan prediksi yang mengarah pada kategori "Puas". Prediksi untuk kategori "Puas" juga menunjukkan hasil yang cukup baik, meskipun terdapat beberapa kesalahan pada kategori lain. Terakhir, untuk kategori "Sangat Puas", model berhasil memprediksi dengan benar sebagian besar ulasan, namun masih ada beberapa kesalahan yang terjadi pada kategori-kategori yang lain. Secara keseluruhan, *confusion matrix* ini menunjukkan bahwa model KNN memiliki kinerja yang baik dalam mengklasifikasikan ulasan pengguna, meskipun masih ada beberapa kesalahan yang perlu diperhatikan, terutama pada kategori yang lebih dekat dengan sentimen netral atau puas. Gambar 5. Merupakan confusion matrix dari KNN

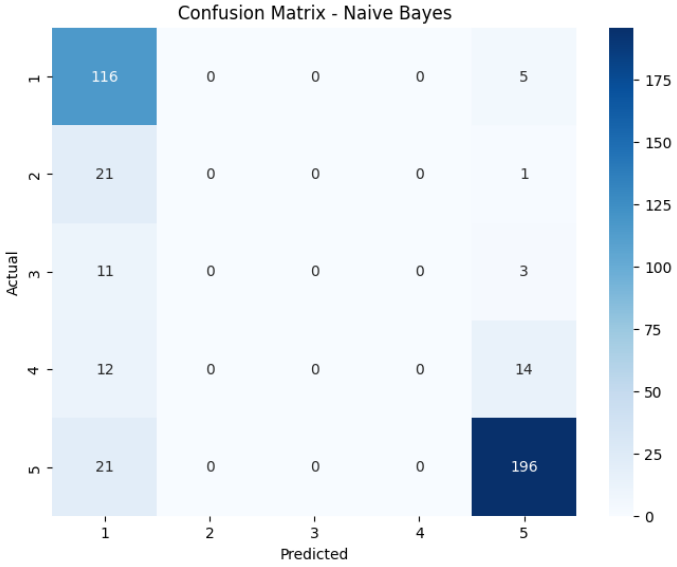
2. *Naive Bayes*

Confusion matrix yang ditunjukkan pada Gambar 6 memperlihatkan kinerja model *Naive Bayes* dalam memprediksi lima kategori sentimen, yaitu "Sangat Buruk", "Buruk", "Cukup", "Puas", dan "Sangat Puas". Pada kategori "Sangat Buruk", model *Naive Bayes* menunjukkan performa yang sangat baik dengan mayoritas prediksi yang benar, meskipun terdapat sedikit kesalahan prediksi ke kategori lain. Untuk kategori "Buruk", model ini berhasil memprediksi sebagian besar ulasan dengan benar, namun beberapa ulasan salah diklasifikasikan, terutama ke kategori "Cukup". Pada kategori "Cukup", model ini menunjukkan kinerja yang optimal dengan jumlah prediksi benar tertinggi, namun masih terdapat sejumlah kesalahan prediksi, terutama ke kategori "Puas". Prediksi untuk kategori "Puas" juga menunjukkan hasil yang memadai, meskipun terdapat beberapa ulasan yang salah diklasifikasikan ke kategori lain. Terakhir, untuk kategori "Sangat Puas," model berhasil memprediksi sebagian besar ulasan dengan benar, namun masih ada beberapa kesalahan prediksi yang terjadi pada kategori-kategori lain. Secara keseluruhan, *confusion matrix* ini menunjukkan bahwa model *Naive Bayes*

memiliki performa yang baik dalam mengklasifikasikan sentimen pengguna, dengan hasil yang mirip dengan model KNN, meskipun kesalahan prediksi masih terjadi pada beberapa kategori yang lebih dekat dengan sentimen netral atau puas. Gambar 6. Merupakan confusion matrix dari Naive Bayes



Gambar 5. *Confusion Matrix* model KNN



Gambar 6. *Confusion Matrix* model Naive Bayes

3. *Support Vector Machine*

Gambar 7 menampilkan tiga confusion matrix yang masing-masing mewakili performa model SVM dengan *kernel* yang berbeda: *Linear kernel,* RBF *kernel*, dan *Polynomial kernel*.

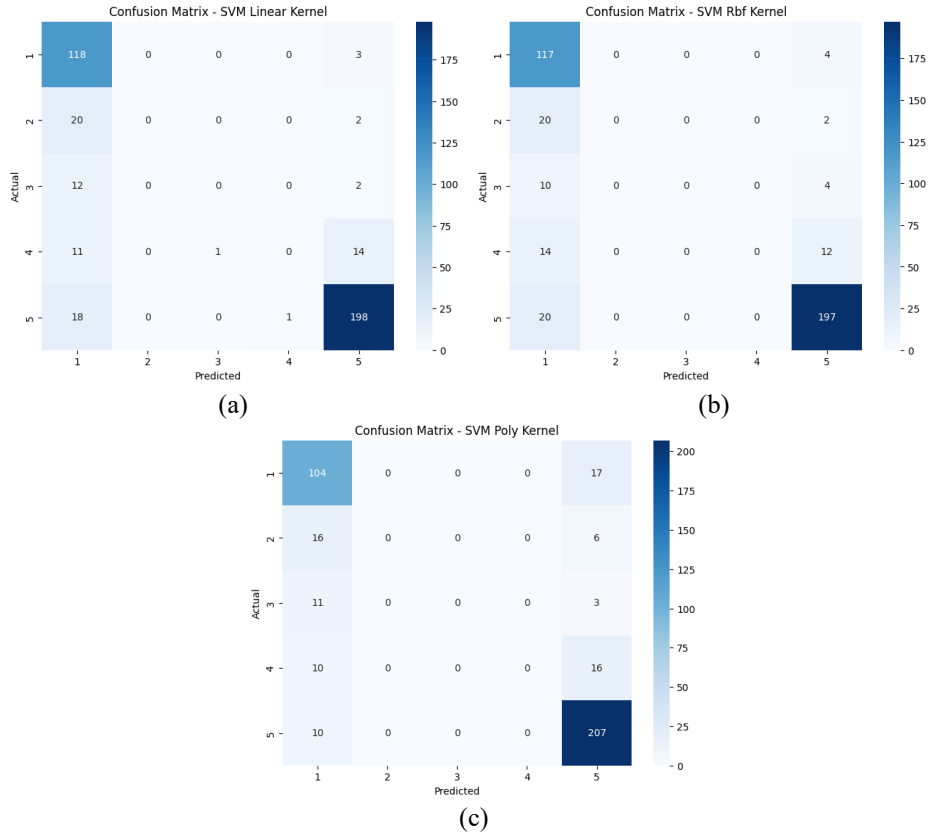
Pada SVM *Linear kernel* (Gambar 7a), model menunjukkan performa yang sangat baik dalam memprediksi kategori "Cukup" dengan 217 prediksi benar dan kesalahan prediksi yang sangat minim pada kategori lain. Kategori "Puas" juga diprediksi dengan baik dengan 121 prediksi benar, namun terdapat beberapa kesalahan kecil di kategori "Sangat Buruk" dan "Buruk". Untuk kategori "Sangat Puas", model berhasil melakukan 22 prediksi benar tanpa kesalahan pada kategori lainnya.

Pada SVM *RBF kernel* (Gambar 7b), hasil yang ditunjukkan sangat mirip dengan *Linear kernel*. Model ini juga memiliki performa terbaik pada kategori "Cukup" dengan 217 prediksi benar dan

sedikit kesalahan di kategori lain. Prediksi untuk kategori "Puas" dan "Sangat Puas" juga menunjukkan hasil yang memuaskan, tanpa kesalahan prediksi yang signifikan pada kategori lain.

Terakhir, SVM *Polynomial kernel* (Gambar 7c) messsnunjukkan performa yang identik dengan dua kernel sebelumnya. Model ini tetap menunjukkan kinerja terbaik dalam memprediksi kategori "Cukup" dan "Puas", dengan 217 dan 121 prediksi benar, masing-masing. Pada kategori "Sangat Puas," model juga berhasil memprediksi dengan benar 22 kali tanpa kesalahan pada kategori lain.

Secara keseluruhan, ketiga *kernel* pada model SVM ini menunjukkan kinerja yang sangat mirip, dengan keunggulan utama pada kategori "Cukup" dan "Puas." Tidak ada perbedaan signifikan dalam kesalahan prediksi di antara *kernel*, menunjukkan bahwa model SVM cukup konsisten dalam mengklasifikasikan sentimen pengguna terhadap aplikasi Gojek berdasarkan ulasan di *App Store*.



Gambar 7. *Confusion matrix* (a) model SVM *linear kernel*, (b) model SVM RBF *kernel*, (c) model *polynomial kernel*

B. Classification Report

Tabel 3. *Classification Report*

Model	***Accuracy(%)***	***Precision(%)***	***Recall(%)***	***F1-Score(%)***
KNN	69.25	27.99	33.71	29.75
Multinomial Naive Bayes	78.0	67.93	78.0	72.01
SVM-Linear Kernel	79.0	74.48	79.0	73.06
SVM-RBF	78.50	83.85	78.50	72.46
SVM-Polynomial	77.75	81.43	77.75	71.32

Hasil *classification report* pada Tabel 3 menunjukkan perbedaan kinerja antara berbagai model yang digunakan dalam analisis sentimen pengguna. Model *K-Nearest Neighbors* (KNN) hanya mencapai

akurasi sebesar 69.25% dengan nilai *precision* terendah, yaitu 27.99%. Model ini juga memiliki recall sebesar 33.71% dan *F1-score* 29.75%, yang menunjukkan performa kurang optimal. Sebaliknya, model *Multinomial Naive Bayes* mencatat akurasi lebih tinggi, yaitu 78.00%, dengan *precision* sebesar 67.93%, *recall* sebesar 78.00%, dan *F1-score* 72.01%. Model *Support Vector Machine* (SVM) dengan kernel Linear mencapai akurasi tertinggi sebesar 79.00%, dengan *precision* 74.48%, *recall* 79.00%, dan *F1-score* 73.06%. Sementara itu, SVM dengan kernel RBF memiliki kombinasi precision tertinggi, yaitu 83.85%, meskipun akurasinya sedikit lebih rendah, yaitu 78.50%, serta *recall* 78.50% dan *F1-score* 72.46%. Model SVM dengan kernel *Polynomial* menunjukkan kinerja yang sedikit lebih rendah dibandingkan kernel lainnya, dengan akurasi sebesar 77.75%, *precision* 81.43%, *recall* 77.75%, dan *F1-score* 71.32%. Dengan demikian, meskipun SVM secara umum memiliki kinerja lebih baik dibandingkan KNN dan *Multinomial Naive Bayes*, kinerja terbaik di antara semua model dapat bervariasi bergantung pada metrik evaluasi yang digunakan.

Kesimpulan

Berdasarkan hasil analisis eksplorasi data dan evaluasi model yang telah dilakukan dalam penelitian ini, beberapa temuan penting dapat disimpulkan. Visualisasi seperti *word cloud* dan distribusi frekuensi *unigram* menunjukkan bahwa pengguna aplikasi Gojek cenderung lebih sering memberikan ulasan mengenai kualitas layanan pengemudi, fitur aplikasi, dan pengalaman pemesanan. Kata-kata seperti "*good*", "*driver*", dan "aplikasi" muncul dengan frekuensi tinggi, menunjukkan aspek-aspek layanan yang paling diperhatikan oleh pengguna. Distribusi polaritas sentimen juga mengungkapkan bahwa ulasan pengguna cenderung ekstrem, dengan banyak ulasan yang sangat positif maupun sangat negatif, sementara ulasan netral relatif sedikit.

Dari hasil evaluasi model, ditemukan bahwa SVM, khususnya dengan kernel Linear dan RBF, menunjukkan kinerja terbaik dibandingkan dengan KNN dan *Naive Bayes* dalam analisis sentimen ulasan pengguna Gojek. SVM secara konsisten unggul dalam akurasi dan *precision* dengan kernel Linear mencapai akurasi tertinggi sebesar 79.00%, diikuti kernel RBF dengan precision tertinggi sebesar 83.85%, menjadikannya metode yang paling andal untuk tugas klasifikasi ini. Di sisi lain, KNN memiliki performa terendah sebesar 69.25%, sementara *Naive Bayes* menunjukkan kinerja yang cukup baik sebesar 78.00% tetapi masih kalah dibandingkan SVM. Hasil ini menegaskan bahwa pemilihan metode yang tepat, seperti SVM, dapat memberikan hasil yang lebih akurat dalam mendukung pengambilan keputusan berbasis sentimen pengguna.

Daftar Pustaka

[1] S. Padmaja and S. S. Fatima, “Opinion mining and sentiment analysis-an assessment of peoples’ belief: A survey,” *International Journal of Ad hoc, Sensor & Ubiquitous Computing*, vol. 4, no. 1, p. 21, 2013.

[2] G. Alexander, “Automating Large-scale Health Care Service Feedback Analysis: Sentiment Analysis and Topic Modeling Study,” *JMIR Med Inform*, vol. 10, no. 4, 2022, doi: 10.2196/29385.

[3] M. A. Ganaie, “KNN weighted reduced universum twin SVM for class imbalance learning,” *Knowl Based Syst*, vol. 245, 2022, doi: 10.1016/j.knosys.2022.108578.

[4] S. Mohsen, “Human Activity Recognition Using K-Nearest Neighbor Machine Learning Algorithm,” *Smart Innovation, Systems and Technologies*, vol. 262, pp. 304–313, 2022, doi: 10.1007/978-981-16-6128-0_29.

[5] K. Hulliyah, “Analysis of Public Sentiment Using The K-Nearest Neighbor (k-NN) Algorithm and Lexicon Based on Indonesian Television Shows on Social Media Twitter,” *2022 10th International Conference on Cyber and IT Service Management, CITSM 2022*, 2022, doi: 10.1109/CITSM56380.2022.9936011.

[6] P. Subarkah, W. R. Damayanti, and R. A. Permana, “Comparison of correlated algorithm accuracy Naive Bayes Classifier and Naive Bayes Classifier for heart failure classification,” *ILKOM Jurnal Ilmiah*, vol. 14, no. 2, pp. 120–125, 2022.

[7] I. Wickramasinghe and H. Kalutarage, “Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation,” *Soft comput*, vol. 25, no. 3, pp. 2277–2293, 2021.

[8] C. U. Kumari, "An automated detection of heart arrhythmias using machine learning technique: SVM," *Mater Today Proc*, vol. 45, pp. 1393–1398, 2021, doi: 10.1016/j.matpr.2020.07.088.

[9] N. P. Arthamevia, "Aspect-Based Sentiment Analysis in Beauty Product Reviews Using TF-IDF and SVM Algorithm," *2021 9th International Conference on Information and Communication Technology, ICoICT 2021*, pp. 197–201, 2021, doi: 10.1109/ICoICT52021.2021.9527489.

[10] A. P. Widyassari, "The 7-Phases Preprocessing Based On Extractive Text Summarization," *2022 7th International Conference on Informatics and Computing, ICIC 2022*, 2022, doi: 10.1109/ICIC56845.2022.10006998.

[11] I. Ho, "Preprocessing Impact on Sentiment Analysis Performance on Malay Social Media Text," *Journal of System and Management Sciences*, vol. 12, no. 5, pp. 73–90, 2022, doi: 10.33168/JSMS.2022.0505.

[12] H. Darwis, N. Wanaspati, and S. Anraeni, "Support Vector Machine untuk Analisis Sentimen Masyarakat Terhadap Penggunaan Antibiotik di Indonesia," *The Indonesian Journal of Computer Science*, vol. 12, no. 4, Aug. 2023, doi: 10.33022/ijcs.v12i4.3320.

[13] A. Amelia, L. N. Hayati, dan H. Darwis, "Analisis Sentimen Ulasan Pengguna terhadap Aplikasi MyPertamina Menggunakan Metode Random Forest, SVM, dan Naïve Bayes," *Universitas Muslim Indonesia*, Makassar, Indonesia, 2023.

[14] M. Potočár, "Comparison of Unigram, HMM, CRF and Brill's Part-of-Speech Taggers Available in NLTK Library," *Conference of Open Innovation Association, FRUCT*, vol. 2023, pp. 226–235, 2023, doi: 10.23919/FRUCT58615.2023.10143061.

[15] M. Garg, "UBIS: Unigram Bigram Importance Score for Feature Selection from Short Text," *Expert Syst Appl*, vol. 195, 2022, doi: 10.1016/j.eswa.2022.116563.

[16] N. P. Arthamevia, "Aspect-Based Sentiment Analysis in Beauty Product Reviews Using TF-IDF and SVM Algorithm," *2021 9th International Conference on Information and Communication Technology, ICoICT 2021*, pp. 197–201, 2021, doi: 10.1109/ICoICT52021.2021.9527489.

[17] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *ILKOM Jurnal Ilmiah,* vol. 10, no. 2, pp. 160-165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.

[18] E. Hossain, "Sentiment Polarity Detection on Bengali Book Reviews Using Multinomial Naïve Bayes," *Advances in Intelligent Systems and Computing*, vol. 1299, pp. 281–292, 2021, doi: 10.1007/978-981-33-4299-6_23.

[19] S. Chaudhury, "The Sentiment Analysis of Human Behavior on Products and Organizations using K-Means Clustering and SVM Classifier," *Proceedings of 3rd International Conference on Intelligent Engineering and Management, ICIEM 2022*, pp. 610–615, 2022, doi: 10.1109/ICIEM54221.2022.9853128.

[20] J. Polpinij, "A Comparative Study of Short Text Classification Methods for Bug Report Type Identification," *Proceedings - 2022 Research, Invention, and Innovation Congress: Innovative Electricals and Electronics, RI2C 2022*, pp. 27–33, 2022, doi: 10.1109/RI2C56397.2022.9910299.

[21] Nurul Ehsan Ramli, Zainor Ridzuan Yahya, and Nor Azinee Said, "Confusion Matrix as Performance Measure for Corner Detectors," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 29, no. 1, pp. 256–265, Dec. 2022, doi: 10.37934/araset.29.1.256265.