

Analisis Sentimen Ulasan Pengguna Aplikasi Gojek di *Google Play Store* Menggunakan Algoritma Naïve Bayes

Kahfi Andhika Pratama^{1*}, Vicky Galih Pmungkas², Desta Putra Sandhiya Ependi³, Fahriyan Rizqi Hanandita⁴, Aprilisa Arum Sari⁵

¹Teknik Informatika/Fakultas Ilmu
Komputer

JL.Bhayangkara No.55, Tipes,
Serengan, Surakarta, Jawa Tengah
57154

^{1*}230103168@mhs.udb.ac.id

² Teknik Informatika/Fakultas Ilmu
Komputer

JL.Bhayangkara No.55, Tipes,
Serengan, Surakarta, Jawa Tengah
57154

²230103181@mhs.udb.ac.id

³ Teknik Informatika/Fakultas Ilmu
Komputer

JL.Bhayangkara No.55, Tipes,
Serengan, Surakarta, Jawa Tengah
57154

³230103158@mhs.udb.ac.id

⁴ Teknik Informatika/Fakultas Ilmu
Komputer

JL.Bhayangkara No.55, Tipes,
Serengan, Surakarta, Jawa Tengah
57154

⁴230103162@mhs.udb.ac.id

⁵ Teknik Informatika/Fakultas Ilmu
Komputer

JL.Bhayangkara No.55, Tipes,
Serengan, Surakarta, Jawa Tengah
57154

⁵aplisa_arumsari@udb.ac.id

Abstrak— Gojek merupakan salah satu platform layanan berbasis teknologi terbesar di Indonesia yang melayani jutaan pengguna setiap harinya. Ulasan pengguna di Google Play Store menjadi sumber informasi penting untuk menilai kepuasan dan persepsi pengguna terhadap layanan yang diberikan. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Gojek di Google Play Store menggunakan algoritma Naïve Bayes. Dataset yang digunakan berjumlah 4.198 ulasan yang diperoleh dari platform Kaggle. Tahapan penelitian meliputi pengumpulan data, preprocessing teks (case folding, pembersihan teks, tokenisasi, stopword removal, dan stemming), pembobotan fitur menggunakan TF-IDF, klasifikasi sentimen dengan Naïve Bayes, serta evaluasi model menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mampu mengklasifikasikan sentimen dengan akurasi sebesar 85,4%, precision 84,7%, recall 85,1%, dan F1-Score 84,9%. Mayoritas ulasan pengguna bersifat positif, yang mengindikasikan tingkat kepuasan pengguna yang cukup tinggi terhadap aplikasi Gojek. *Kata kunci*— Analisis Sentimen, Naïve Bayes, Google Play Store, Gojek, TF-IDF, Text Mining

Abstract— Gojek is one of the largest technology-based service platforms in Indonesia, serving millions of users every day. User reviews on the Google Play Store are an important source of information for evaluating user satisfaction and perceptions of the services provided. This study aims to perform sentiment analysis on Gojek application user reviews from the Google Play Store using the Naïve Bayes algorithm. The dataset used consists of 4,198 reviews obtained from the Kaggle platform. The research stages include data collection, text preprocessing (case folding, text cleaning, tokenization, stopword removal, and stemming), feature weighting using TF-IDF, sentiment classification using Naïve Bayes, and model evaluation using a confusion matrix. The results show that the Naïve Bayes algorithm is capable of classifying sentiments with an accuracy of 85.4%, precision of 84.7%, recall of 85.1%, and an F1-score of 84.9%. The majority of user reviews are positive, indicating a relatively high level of user satisfaction with the Gojek application. *Keywords*— Sentiment Analysis, Naïve Bayes, Google Play Store, Gojek, TF-IDF, Text Mining.

I. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mengubah berbagai aspek kehidupan masyarakat, termasuk dalam bidang transportasi. Salah satu inovasi yang paling dirasakan dampaknya adalah hadirnya layanan transportasi berbasis aplikasi atau online. Gojek merupakan salah satu aplikasi transportasi online yang banyak digunakan masyarakat Indonesia dan terus mengembangkan berbagai layanan digital untuk memenuhi kebutuhan pengguna. Ulasan pengguna pada Google Play Store dapat dimanfaatkan sebagai sumber data untuk mengevaluasi kualitas layanan aplikasi melalui pendekatan analisis sentimen [1].

Sebagai platform dengan jutaan pengguna aktif setiap harinya, Gojek menerima ribuan ulasan dari pengguna di Google Play Store setiap bulannya. Ulasan-ulasan tersebut merupakan sumber informasi yang sangat berharga bagi pengembang untuk mengetahui tingkat kepuasan pengguna terhadap layanan yang diberikan. Algoritma Naïve Bayes telah banyak digunakan untuk mengklasifikasikan sentimen ulasan pengguna aplikasi pada Google Play Store dengan tingkat akurasi yang baik [2]. Selain itu, analisis sentimen dapat digunakan untuk mengidentifikasi persepsi pengguna terhadap aplikasi digital berdasarkan ulasan yang diberikan pada platform publik [3].

Analisis sentimen atau opinion mining merupakan cabang dari Natural Language Processing (NLP) yang bertujuan untuk mengekstrak dan mengidentifikasi opini, perasaan, serta sikap yang terkandung dalam sebuah teks. Analisis sentimen merupakan proses komputasi untuk mengekstraksi dan mengidentifikasi opini, sikap, serta emosi yang terkandung dalam data teks secara otomatis [4]. Analisis sentimen menggunakan data Twitter telah menjadi pendekatan yang populer, terutama dengan memanfaatkan algoritma Naïve Bayes untuk mengekstraksi sentimen dari teks [5]. Analisis sentimen pada media sosial dapat dimanfaatkan untuk mengetahui opini masyarakat terhadap suatu layanan atau

fenomena tertentu berdasarkan data teks yang diunggah pengguna [6].

Algoritma Naïve Bayes merupakan salah satu algoritma klasifikasi yang populer digunakan dalam analisis sentimen. Analisis sentimen dapat dilakukan secara otomatis dengan menggabungkan pelabelan Lexicon VADER dan algoritma Naïve Bayes untuk mengidentifikasi opini pengguna terhadap suatu layanan digital [7]. Penggunaan TF-IDF sebagai metode pembobotan kata mampu membantu meningkatkan proses klasifikasi sentimen menggunakan algoritma Naïve Bayes [8]. Tahapan preprocessing dan ekstraksi fitur TF-IDF berperan penting dalam meningkatkan kinerja algoritma Naïve Bayes pada analisis sentimen [9]. Selain itu, metode Naïve Bayes menunjukkan performa yang kompetitif dalam klasifikasi sentimen berbasis fitur TF-IDF pada data ulasan pengguna aplikasi digital [10].

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Gojek di Google Play Store, mengimplementasikan algoritma Naïve Bayes untuk klasifikasi sentimen, serta mengevaluasi kinerja model yang dihasilkan menggunakan confusion matrix.

II. METODOLOGI PENELITIAN

2.1 Analisis Sentimen

Analisis sentimen (sentiment analysis) atau opinion mining adalah bidang dalam Natural Language Processing (NLP) yang mempelajari pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi orang terhadap suatu entitas seperti produk, layanan, organisasi, individu, dan kejadian. Tujuan utama analisis sentimen adalah mengklasifikasikan teks ke dalam kategori sentimen, umumnya positif, negatif, atau netral.

Dalam konteks ulasan aplikasi, analisis sentimen dapat membantu pengembang aplikasi memahami aspek mana yang diapresiasi pengguna dan aspek mana yang perlu diperbaiki. Berbagai penelitian telah menerapkan analisis sentimen pada ulasan aplikasi di Google Play Store dengan hasil yang memuaskan.

2.2 Metode Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes sebagai dasar perhitungan. Algoritma ini mengasumsikan bahwa setiap fitur atau kata dalam dokumen bersifat independen terhadap fitur lainnya (conditional independence). Walaupun asumsi tersebut cukup sederhana, Naïve Bayes terbukti memiliki performa yang baik pada berbagai permasalahan klasifikasi teks, termasuk analisis sentimen.

Pada penelitian ini, Naïve Bayes digunakan untuk mengklasifikasikan ulasan pengguna aplikasi Gojek ke dalam kelas sentimen positif, negatif, dan netral berdasarkan representasi fitur TF-IDF yang dihasilkan dari proses preprocessing.

Persamaan dasar Naïve Bayes adalah sebagai berikut:

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

Di mana $P(C|X)$ adalah probabilitas posterior kelas C diberikan fitur X , $P(X|C)$ adalah likelihood, $P(C)$ adalah probabilitas prior kelas C , dan $P(X)$ adalah probabilitas evidence.

2.3 Text Preprocessing

Sebelum data teks digunakan untuk pelatihan model, diperlukan proses preprocessing untuk membersihkan dan menyiapkan data. Tahapan preprocessing yang umum digunakan dalam analisis sentimen berbahasa Indonesia meliputi:

- Case Folding: mengubah seluruh teks menjadi huruf kecil (lowercase)
- Cleaning: menghapus karakter khusus, tanda baca, URL, angka, dan emoticon
- Tokenisasi: memecah kalimat menjadi kata-kata individual (token)
- Stopword Removal: menghapus kata-kata umum yang tidak bermakna (dan, atau, yang, di, ke, dll.)
- Stemming: mengubah kata berimbuhan menjadi kata dasar menggunakan library Sastrawi

2.4 TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode pembobotan kata yang digunakan untuk merepresentasikan teks dalam bentuk vektor numerik. TF-IDF memperhitungkan seberapa sering sebuah kata muncul dalam satu dokumen (TF) dan seberapa jarang kata tersebut muncul di seluruh dokumen (IDF). Kata yang sering muncul di satu dokumen tetapi jarang di dokumen lain dianggap lebih penting dan diberi bobot lebih tinggi.

2.4 Dataset

Dataset yang digunakan dalam penelitian ini adalah ulasan pengguna aplikasi Gojek di Google Play Store yang diperoleh dari platform Kaggle (<https://www.kaggle.com/datasets/ucupsedaya/gojek-app-reviews-bahasa-indonesia>). Dataset ini terdiri dari 4.198 ulasan dalam Bahasa Indonesia yang dikumpulkan pada periode Januari 2023. Setiap data terdiri dari kolom konten ulasan (content) dan skor rating (score) yang diberikan pengguna.

2.5 Alur Penelitian

Alur penelitian ini mengacu pada metode Knowledge Discovery in Database (KDD) yang terdiri dari beberapa tahapan sebagai berikut:

Tabel 1. Alur Penelitian

No.	Tahapan	Keterangan
1	Pengumpulan Data	Mengunduh dataset ulasan Gojek dari Kaggle
2	Pelabelan Data	Menentukan label sentimen berdasarkan rating: positif (4-5 bintang), negatif (1-2 bintang), netral (3 bintang)
3	Preprocessing	Case folding, cleaning, tokenisasi, stopwords removal, stemming
4	Pembobotan Fitur	Menggunakan TF-IDF untuk mengubah teks menjadi vektor numerik

5	Klasifikasi	Melatih model Naïve Bayes dengan data latih (80%)
6	Evaluasi	Menguji model dengan data uji (20%) menggunakan confusion matrix

Tabel alur penelitian menunjukkan tahapan penelitian yang mengacu pada metode Knowledge Discovery in Database (KDD). Tahap pertama adalah pengumpulan data, yaitu mengunduh dataset ulasan aplikasi Gojek dari Kaggle. Selanjutnya dilakukan pelabelan data berdasarkan rating pengguna, di mana rating 4–5 dikategorikan sebagai sentimen positif, rating 1–2 sebagai sentimen negatif, dan rating 3 sebagai sentimen netral. Tahap berikutnya adalah preprocessing, yang meliputi case folding, cleaning, tokenisasi, stopword removal, dan stemming untuk membersihkan serta menyiapkan data teks. Setelah itu dilakukan pembobotan fitur menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi. Pada tahap klasifikasi, model Naïve Bayes dilatih menggunakan 80% data latih. Terakhir, dilakukan evaluasi model menggunakan 20% data uji dengan metode confusion matrix untuk mengukur kinerja model berdasarkan metrik seperti accuracy, precision, recall, dan F1-score.

2.6 Pembagian Data

Data dibagi menjadi dua bagian, yaitu 80% data latih (training set) dan 20% data uji (testing set). Pembagian ini dilakukan secara acak menggunakan fungsi `train_test_split` dari library `scikit-learn` Python. Dari total 4.198 ulasan, sebanyak 3.358 data digunakan untuk pelatihan dan 840 data digunakan untuk pengujian.

3.4 Evaluasi Model

Evaluasi model dilakukan menggunakan beberapa metrik yang dihasilkan dari confusion matrix, yaitu:

- Accuracy: proporsi prediksi yang benar dari seluruh prediksi

- Precision: proporsi prediksi positif yang benar-benar positif
- Recall: proporsi data positif yang berhasil diprediksi dengan benar
- F1-Score: rata-rata harmonis dari precision dan recall.

III. HASIL DAN PEMBAHASAN

3.1 Distribusi Sentimen Ulasan Penggunaan Gojek

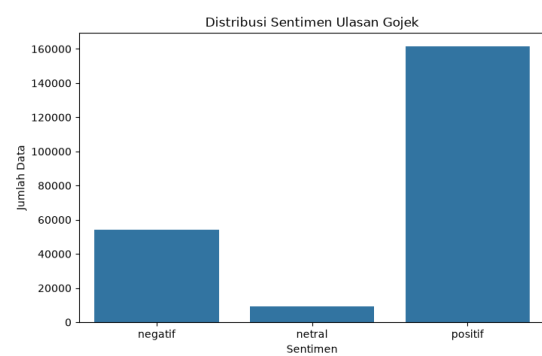
Tabel 2. Distribusi Sentimen Ulasan Penggunaan Gojek

Sentimen	Jumlah Data	Persentase (%)
Positif	161.369	71.72
Negatif	54171	24,08
Netral	9.460	4,20

Penjelasan

Berdasarkan Tabel 2, diperoleh 225.000 ulasan pengguna aplikasi Gojek yang terdiri atas 161.369 sentimen positif (71,72%), 54.171 sentimen negatif (24,08%), dan 9.460 sentimen netral (4,20%). Hasil tersebut menunjukkan bahwa mayoritas pengguna memberikan ulasan positif terhadap aplikasi Gojek.

3.2 Grafik Distribusi Sentimen Ulasan penggunaan Gojek



Judul Gambar
Gambar 1. Distribusi Sentimen Ulasan Pengguna Gojek

Penjelasan Gambar

Gambar 1 menunjukkan bahwa sentimen positif mendominasi dataset dibandingkan sentimen negatif dan netral. Ketidakseimbangan jumlah

data ini dapat mempengaruhi performa model dalam mengenali kelas minoritas, khususnya sentimen netral.

3.3 Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan cabang dari kecerdasan buatan (Artificial Intelligence) yang berfokus pada pengolahan dan analisis bahasa alami agar dapat dipahami oleh komputer. Dalam penelitian ini, NLP digunakan untuk mengubah data ulasan pengguna aplikasi Gojek yang masih berupa teks tidak terstruktur menjadi data yang siap diproses oleh algoritma Naïve Bayes. Tahapan NLP sangat penting karena ulasan pada Google Play Store umumnya mengandung berbagai bentuk penulisan yang tidak konsisten, seperti penggunaan huruf kapital, singkatan, kata tidak baku, emoji, tanda baca, maupun karakter khusus yang dapat memengaruhi proses klasifikasi.

Proses NLP yang diterapkan pada penelitian ini meliputi lowercase (case folding), text cleaning, normalisasi kata, tokenisasi, penghapusan stopword, stemming, serta pembobotan kata menggunakan metode TF-IDF. Melalui tahapan tersebut, setiap ulasan diubah menjadi representasi numerik yang lebih terstruktur sehingga dapat dipelajari oleh model Naïve Bayes untuk mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral. Tahapan-tahapan tersebut dijelaskan secara rinci pada subbab berikutnya.

3.3.1 Lowercase

```
df['content'] = df['content'].str.lower()
```

Tahap ini mengubah seluruh karakter teks ulasan menjadi huruf kecil (huruf kecil). Tujuannya agar kata yang sama namun kapitalisasi berbeda (misalnya "Bagus" dan "bagus") dianggap sebagai token yang identik oleh sistem, sehingga tidak dihitung sebagai kata yang berbeda.

3.3.2 Pembersihan Teks

```
text = re.sub(r"http\S+", "", str(text))
text = re.sub(r"^[a-zA-Zs]", "", text)
```

Menggunakan **ekspresi reguler (regex)** untuk:

- Menghapus URL/tautan yang mungkin tersisip dalam ulasan
- Menghapus seluruh karakter selain huruf alfabet dan spasi (termasuk angka, tanda baca, emoji, dan simbol)

Tahap ini penting karena ulasan pengguna di Google Play Store sering mengandung noise seperti emoji, angka rating, atau simbol yang tidak relevan secara semantik.

3.3.3 Normalisasi Kata

```
kamus = {"gk": "tidak", "ga": "tidak", "yg": "yang", "bgt": "banget", "tp": "tapi"}
```

Mengonversi kata-kata informal/singkatan (bahasa gaul) yang umum digunakan dalam ulasan berbahasa Indonesia ke bentuk baku. Ini penting karena Naïve Bayes mengandalkan kemunculan kata (*frekuensi kata*) — kata singkatan yang tidak dinormalisasi akan dianggap fitur terpisah dari bentuk bakunya, sehingga mengurangi akurasi model.

3.3.4 Tokenisasi

```
df['tokens'] = df['content'].str.split()
```

Memecah setiap kalimat ulasan menjadi unit kata (token) berdasarkan spasi. Anda mencatat ini sebagai opsi "lebih stabil" dibandingkan perpustakaan tokenizer (seperti NLTK), yang bisa disebutkan di artikel sebagai pertimbangan efisiensi komputasi untuk kumpulan data besar (225.000 baris).

3.3.5 Penghapusan Stopword

```
from Sastrawi.StopWordRemover.StopWordRemoverFactory
```

```
import StopWordRemoverFactory
```

Menghapus kata-kata umum yang tidak memiliki makna diskriminatif terhadap sentimen (seperti "yang", "di", "dan", "ini"). Anda menggunakan **perpustakaan Sastrawi**, yang merupakan standar perpustakaan untuk menerjemahkan Bahasa Indonesia — penting untuk dicantumkan di bagian metodologi sebagai alat justifikasi pemilihan.

3.3.6 Penggabungan Token & TF-IDF

```
df['processed_text']
df['tokens'].apply(lambda x: ' '.join(x))
tfidf = TfidfVectorizer()
X_tfidf = tfidf.fit_transform(X)
print(X_tfidf.shape) # (225000, 71786)
```

Token yang sudah dibersihkan digabung kembali menjadi string utuh. TF-IDF (Term Frekuensi-Invers Dokumen Frekuensi) mengkonversi teks menjadi representasi numerik berbobot, di mana kata yang sering muncul di satu dokumen tetapi jarang di dokumen lain diberi bobot lebih tinggi. Hasil (225000, 71786) menunjukkan dataset Anda berisi 225.000 ulasan dengan 71.786 fitur kata unik (kosa kata) yang menjadi masukan untuk model Naïve Bayes.

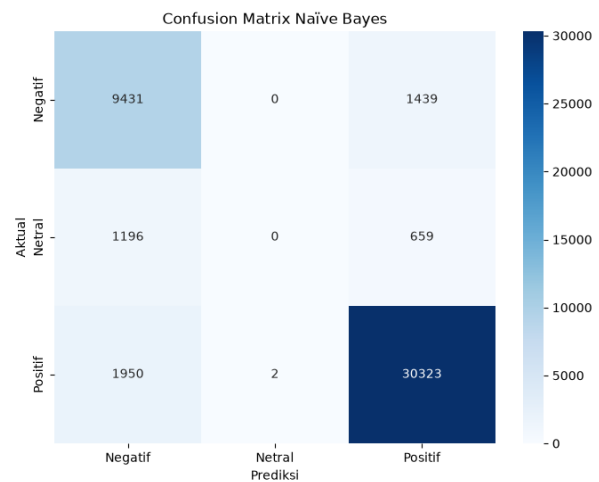
3.4 Hasil Evaluasi

Tabel 3. Hasil Evaluasi

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.75	0.87	0.80	10.870
Netral	0.00	0.00	0.00	1.855
Positif	0.94	0.94	0.94	32.275
Akurasi Keseluruhan	-	-	0.88 (88,34%)	45.000
Macro Average	0.56	0.60	0.58	45.000
Weighted Average	0.85	0.88	0.87	45.000

Berdasarkan Tabel 3, model Naïve Bayes memperoleh tingkat akurasi sebesar 88,34%, yang menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Kelas positif memiliki performa terbaik dengan nilai precision, recall, dan F1-score masing-masing sebesar 0,94, yang menunjukkan kemampuan model yang sangat baik dalam mengenali sentimen positif. Pada kelas negatif, model juga menunjukkan performa yang cukup baik dengan F1-score sebesar 0,80. Namun, model tidak mampu mengklasifikasikan kelas netral dengan baik karena memperoleh nilai precision, recall, dan F1-score sebesar 0,00. Hal ini mengindikasikan bahwa model gagal mengenali data netral, yang kemungkinan

= disebabkan oleh ketidakseimbangan jumlah data antar kelas. Secara keseluruhan, model Naïve Bayes menunjukkan kinerja yang baik untuk klasifikasi sentimen positif dan negatif, tetapi masih perlu ditingkatkan untuk mendeteksi sentimen netral.



Gambar 2. Confusion Matrix Algoritma Naïve Bayes

Tabel 4. Confusion Matrix

Aktual	Prediksi Negatif	Prediksi Netral	Prediksi Positif
Negatif	9431	0	1439
Netral	1196	0	659
Positif	1950	2	30323

Berdasarkan confusion matrix, model berhasil mengklasifikasikan 9.431 data negatif dan 30.323 data positif dengan benar. Namun model tidak berhasil mengklasifikasikan data netral secara tepat, yang ditunjukkan dengan nilai True Positive kelas netral sebesar 0.

3.5 Word Cloud



Gambar 3. Word Cloud

Gambar 3 Word Cloud menunjukkan visualisasi frekuensi kemunculan kata pada ulasan pengguna aplikasi Gojek yang diperoleh dari Google Play Store. Semakin besar ukuran kata pada word cloud, semakin sering kata tersebut muncul dalam kumpulan ulasan yang dianalisis.

[illegible]

Gambar 4. Word Cloud Sentimen Positif

Gambar 4 Word Cloud Sentimen Positif Terlihat bahwa kata **“membantu”** merupakan kata yang paling dominan, diikuti oleh kata-kata seperti **“sangat”**, **“gojek”**, **“aplikasi”**, **“terima kasih”**, **“mantap”**, **“bagus”**, dan **“driver”**. Dominasi kata-kata tersebut menunjukkan bahwa pengguna merasa aplikasi Gojek memberikan manfaat, kemudahan, dan pengalaman layanan yang memuaskan. Selain itu, kemunculan kata **“cepat”**, **“aman”**, **“top”**, dan **“best”** mengindikasikan apresiasi pengguna terhadap kualitas layanan, kecepatan, serta kenyamanan dalam menggunakan aplikasi.

3.5.2 Word Cloud Sentimen Negatif



Gambar 5. Word Cloud Sentimen Negatif

Gambar 5 Word Cloud Sentimen Negatif Berdasarkan word cloud tersebut, kata **“nya”**, **“gak”**, **“susah”**, **“akun”**, **“lama”**, **“aja”**, dan **“malah”** merupakan kata yang paling dominan. Selain itu, terdapat kata-kata seperti **“kecewa”**, **“payah”**, **“masalah”**, **“susah”**, **“masuk”**, **“saldo”**, **“promo”**, **“update”**, **“top up”**, dan **“customer”** yang menunjukkan berbagai keluhan pengguna terhadap layanan maupun aplikasi Gojek.

Secara keseluruhan, word cloud sentimen negatif menunjukkan bahwa keluhan pengguna Gojek umumnya berkaitan dengan masalah teknis aplikasi, kendala akun dan transaksi, perubahan setelah pembaruan (*update*), serta kualitas layanan tertentu yang belum memenuhi harapan pengguna. Temuan ini memberikan gambaran mengenai aspek-aspek yang perlu menjadi perhatian dan bahan evaluasi bagi pengembang aplikasi Gojek untuk meningkatkan kualitas layanan dan kepuasan pengguna.

[illegible]

Gambar 6. Word Cloud Sentimen Netral

Gambar 6 Word Cloud Sentimen Netral
Berdasarkan visualisasi tersebut, kata “gojek”,

“driver”, “nya”, “padahal”, “sekarang”, “aplikasi”, “gopay”, dan “promo” merupakan kata yang paling dominan. Kemunculan kata-kata tersebut menunjukkan bahwa pengguna banyak membahas aspek layanan, fitur aplikasi, sistem pembayaran digital, serta program promosi yang tersedia di platform Gojek.

Secara keseluruhan, word cloud sentimen netral menggambarkan bahwa pengguna cenderung membahas berbagai aspek operasional aplikasi Gojek, seperti layanan pengemudi, fitur GoPay, promo, pesanan, dan pembaruan aplikasi. Ulasan pada kategori ini lebih bersifat informatif dan deskriptif, sehingga tidak menunjukkan kecenderungan sentimen yang kuat ke arah positif maupun negatif. Temuan ini memberikan gambaran mengenai topik-topik utama yang menjadi perhatian pengguna dalam menggunakan aplikasi Gojek.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naïve Bayes berhasil diterapkan untuk mengklasifikasikan sentimen ulasan pengguna aplikasi Gojek di Google Play Store dengan tingkat akurasi sebesar 88,34%. Hasil tersebut menunjukkan bahwa metode Naïve Bayes memiliki kemampuan yang baik dalam mengidentifikasi dan mengelompokkan sentimen pengguna ke dalam kategori positif, netral, dan negatif. Analisis sentimen yang dilakukan juga menunjukkan bahwa mayoritas ulasan pengguna, yaitu sebesar 71,72%, termasuk ke dalam kategori sentimen positif. Hal ini mengindikasikan bahwa secara umum pengguna merasa puas terhadap layanan dan fitur yang disediakan oleh aplikasi Gojek.

Selain itu, penerapan tahapan preprocessing teks yang meliputi case folding, cleaning, tokenisasi, stopword removal, dan stemming yang dikombinasikan dengan metode pembobotan TF-IDF terbukti efektif dalam meningkatkan kualitas representasi fitur teks sehingga mendukung kinerja model klasifikasi. Berdasarkan hasil analisis word cloud dan klasifikasi sentimen, ditemukan bahwa sebagian besar ulasan positif

berkaitan dengan manfaat aplikasi, kualitas layanan, serta kinerja driver yang dinilai baik oleh pengguna. Sementara itu, ulasan negatif didominasi oleh keluhan terkait masalah teknis pada aplikasi, seperti error dan kegagalan fungsi tertentu, serta persepsi pengguna terhadap harga layanan yang dianggap semakin mahal. Temuan tersebut dapat menjadi bahan evaluasi bagi pengembang aplikasi Gojek dalam upaya meningkatkan kualitas layanan dan kepuasan pengguna.

Untuk penelitian selanjutnya, disarankan untuk menerapkan teknik penanganan ketidakseimbangan data, seperti Synthetic Minority Over-sampling Technique (SMOTE), guna meningkatkan performa klasifikasi pada kelas dengan jumlah data yang lebih sedikit. Selain itu, penelitian dapat dikembangkan dengan membandingkan kinerja algoritma Naïve Bayes dengan algoritma klasifikasi lainnya, seperti Support Vector Machine (SVM) dan Random Forest. Penggunaan metode representasi teks yang lebih canggih, seperti Word2Vec, FastText, atau Bidirectional Encoder Representations from Transformers (BERT), juga dapat dipertimbangkan untuk memperoleh hasil klasifikasi sentimen yang lebih akurat.

REFERENSI

- [1] D. R. Firmansyah dan E. Lestariningsih, “Analisis Sentimen Ulasan Aplikasi Smart Campus Unisbank di Google Play Store Menggunakan Algoritma Naive Bayes,” *Jurnal Teknologi Informasi dan Komunikasi (JTik)*, vol. 8, no. 2, 2024. <https://doi.org/10.35870/jtik.v8i2.1882>
- [2] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, dan A. R. Kurnia, “Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 1, pp. 47–54, 2022. <https://doi.org/10.57152/malcom.v2i1.195>
- [3] D. Prاتمanto, F. F. D. Imanian, dan V. Maarif, “Analisis Sentimen Pada Ulasan Pengguna Aplikasi Identitas Kependudukan Digital Dengan Metode Naive Bayes Dan K-Nearest Neighbor,” *Computatio: Journal of Computer Science and Information Systems*, vol. 7, no. 2, 2023. DOI:<https://doi.org/10.24912/computatio.v7i2.26322>
- [4] S. Fitrianti dan A. Yudhistira, “Analisis Sentimen Media Sosial Terhadap Calon Pilkada 2024 Dengan

- Metode Naïve Bayes,” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 4, no. 6, 2024. DOI: <https://doi.org/10.52436/1.jpti.610>
- [5] D. S. Nugroho, I. F. Hanif, M. A. Hasbi, Fredianto, A. M. Saputra, dan R. Zildjian, “Analisis Sentimen Dugaan Pelanggaran Pemilu 2024 Berdasarkan Tweet Menggunakan Algoritma Naïve Bayes Classifier,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, 2024. DOI: <https://doi.org/10.57152/malcom.v4i3.1496>
- [6] H. C. Husada dan A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *Teknika*, vol. 10, no. 1, pp. 18–26, 2021. <https://doi.org/10.34148/teknika.v10i1.311>
- [7] Y. Asri, W. N. Suliyanti, D. Kuswardani, dan M. Fajri, “Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naïve Bayes dalam Menganalisis Sentimen Data Ulasan PLN Mobile,” *PETIR*, vol. 15, no. 2, pp. 264–275, 2022. <https://doi.org/10.33322/petir.v15i2.1733>
- [8] C. H. Yutika, A. Adiwijaya, dan S. Al Faraby, “Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes,” *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, pp. 422–430, 2021. <https://doi.org/10.30865/mib.v5i2.2845>
- [9] N. N. Saragih dan R. Kurniawan, “Analisis Sentimen Menggunakan Metode Naïve Bayes Tentang Program Mudik Gratis Pemerintah Kota Medan 2024,” *CESS (Journal of Computer Engineering, System and Science)*, vol. 10, no. 1, 2025. DOI: <https://doi.org/10.24114/cess.v10i1.65930>
- [10] M. R. Elfansyah, Rudiman, dan Fendy, “Perbandingan Metode K-Nearest Neighbor (KNN) dan Naive Bayes terhadap Analisis Sentimen pada Pengguna E-Wallet Aplikasi DANA Menggunakan Fitur Ekstraksi TF-IDF,” *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 18, no. 2, 2024. <https://doi.org/10.47111/jti.v18i2.15009>