Jurnal Informatika: Jurnal pengembangan IT

Vol. 10, No. 4, 2024
ISSN : 2477-5126, DOI:10.30591/jpit.v10i4.8796

Perbandingan *Logistic Regression*, *SVM*, dan *Random Forest* untuk Analisis Sentimen Ulasan Aplikasi Gopay

Ridho Agung Prasetyo [1], Wahju Tjahjo Saputro [2], Dewi Chirzah [3]
[1,2,3]Teknologi Informasi, Fakultas Teknik, Universitas Muhammadiyah Purworejo
Jl. KH. Ahmad Dahlan No.3, Purworejo, Jawa Tengah, Indonesia

Info Artikel

Riwayat Artikel:

Received 2025-05-21
Revised 2025-11-27
Accepted 2025-12-04

Abstract – *The expansion of Indonesia's digital financial landscape has triggered a surge in the adoption of e-wallets, most notably GoPay. Within this context, feedback available on application platforms such as the Google Play Store serves as a crucial metric for assessing user sentiment and service quality. Sentiment analysis based on machine learning algorithms allows for systematic and objective identification of public opinion. This study used 3,000 user reviews collected through web scraping from the Google Play Store, received up to April 21, 2025, with initial labeling based on a lexicon approach. Although many studies have compared sentiment classification algorithms, there has been no research specifically comparing the performance of Logistic Regression, Support Vector Machine (SVM), and Random Forest in the context of GoPay user reviews with lexicon-based labeling. This paper aims to fill the existing void by evaluating the comparative performance of three algorithms based on sentiment classification metrics. Preprocessing procedures encompassed cleaning, case-folding, stemming, slang normalization, tokenizing, filtering, and labeling to ensure data quality. The models, built within the Scikit-learn environment, were tested for accuracy, precision, recall, and F1-score. Empirical results confirm that Logistic Regression outperformed the alternatives, securing 88.16% accuracy while maintaining stability across all sentiment categories. SVM recorded 87.5% accuracy but was weak in detecting negative sentiment. Random Forest showed the lowest performance with 79.33% accuracy and less consistent classification results. Thus, Logistic Regression is recommended as the most effective algorithm for GoPay user sentiment analysis. Future research can explore deep learning-based approaches to handle higher sentiment complexity.*

Keywords: *FinTech, Logistic Regression, Random Forest, Support Vector Machine, Sentiment Analysis.*

Corresponding Author:

Ridho Agung Prasetyo

Email:

ridhoagungprasetyo13@gmail.com

Abstrak – *Pertumbuhan yang cepat dalam layanan keuangan digital di Indonesia telah mendorong peningkatan pemakaian dompet digital seperti GoPay. Ulasan dari pengguna di platform seperti Google Play Store telah menjadi sumber penting untuk menilai kepuasan dan pengalaman pengguna. Analisis sentimen yang didasarkan pada algoritma pembelajaran mesin dapat membantu dalam mengidentifikasi pendapat publik secara sistematis dan objektif. Penelitian ini menggunakan 3. 000 ulasan pengguna yang dikumpulkan dengan teknik penggalian data dari Google Play Store, diterima hingga 21 April 2025, dengan pelabelan awal berdasarkan metode leksikon. Meskipun terdapat banyak penelitian yang membandingkan algoritma klasifikasi sentimen, belum ada yang secara khusus menilai kinerja Regresi Logistik, Support Vector Machine (SVM), dan Random Forest dalam konteks ulasan pengguna GoPay dengan pelabelan berbasis leksikon. Penelitian ini bertujuan untuk mengisi kekurangan ini dengan melakukan evaluasi perbandingan terhadap ketiga algoritma tersebut berdasarkan metrik akurasi dan efektivitas klasifikasi sentimen. Tahap prapemrosesan data meliputi pembersihan, pengaturan huruf besar-kecil, stemming, normalisasi bahasa gaul, tokenisasi, penyaringan, dan pelabelan. Model ini dibangun menggunakan pustaka Scikit-learn dan dievaluasi dengan metrik akurasi, presisi, recall, dan f1-score. Hasil dari penelitian tersebut mengindikasikan bahwa Regresi Logistik memberikan performa terbaik dengan akurasi sebesar 88,16% dan menunjukkan performa yang stabil di semua kategori sentimen. SVM memperoleh akurasi 87,5%, tetapi kurang efektif dalam mendeteksi sentimen negatif. Sementara itu, Random Forest menunjukkan hasil terendah dengan akurasi 79,33% dan hasil klasifikasi yang tidak konsisten. Oleh sebab itu, Regresi Logistik direkomendasikan sebagai algoritma paling efektif untuk analisis sentimen pengguna GoPay. Penelitian di masa mendatang dapat mempertimbangkan pendekatan berbasis pembelajaran mendalam untuk menangani kompleksitas sentimen yang lebih tinggi.*

Kata Kunci*: Analisis Sentimen, FinTech, Logistic Regression, Random Forest, Support Vector Machine.*

I. PENDAHULUAN

Perkembangan teknologi digital di Indonesia mengalami pertumbuhan yang pesat, didukung oleh meningkatnya penetrasi internet dan penggunaan *smartphone*. Pada periode awal tahun 2024, penetrasi internet di Indonesia tercatat telah menyentuh angka 79,5% dari total penduduk. Hal ini setara dengan 221,56 juta jiwa, sebagaimana dilaporkan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII).[1]. Sejalan dengan itu,

laporan dari Otoritas Jasa Keuangan (OJK) menunjukkan bahwa terdapat sekitar 233 juta pengguna *smartphone* di Indonesia, mencerminkan tingginya adopsi teknologi digital di berbagai lapisan masyarakat[2]. Perkembangan ini turut mendorong pertumbuhan industri teknologi finansial *(fintech)*, yang memainkan peran penting dalam ekosistem ekonomi digital. OJK mencatat bahwa sektor fintech di Indonesia berkontribusi signifikan terhadap pertumbuhan ekonomi digital, dengan *Gross Merchandise Value* (GMV) yang diperkirakan mencapai USD 90 miliar pada tahun 2024, meningkat 13% dibandingkan tahun sebelumnya[3].

Penggunaan *mobile fintech* mampu menjangkau segmen masyarakat bawah piramida *(Bottom of the Pyramid/BOP)* yang sebelumnya sulit dijangkau oleh layanan keuangan formal [4]. Pertumbuhan industri *fintech* di Indonesia telah mendorong maraknya penggunaan layanan *e-wallet*, salah satunya GoPay [5]. *E-wallet* menjadi bagian penting dalam transformasi sistem transaksi keuangan di Indonesia, didorong oleh preferensi masyarakat terhadap solusi yang praktis, cepat, dan minim kontak fisik [6]. Keputusan penggunaan *e-wallet* sangat dipengaruhi oleh faktor-faktor seperti kemudahan penggunaan, persepsi manfaat, keamanan, promosi, serta kepercayaan pengguna ([7], [8]).

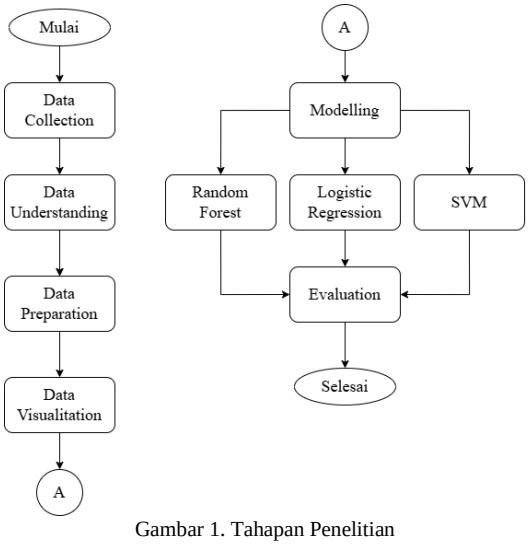
Kesiapan teknologi pengguna berpengaruh signifikan terhadap faktor kepercayaan, kemudahan penggunaan, dan persepsi kegunaan berperan dalam membentuk niat perilaku untuk menggunakan GoPay ([9], [10], [11], [12]). Kepercayaan pengguna terhadap aplikasi dipengaruhi oleh ulasan dan komentar di Google Play Store [13]. Analisis sentimen terhadap ulasan dan komentar pengguna di Google Play Store menjadi metode efektif untuk mengevaluasi opini publik terhadap aplikasi [14]. Hal ini diperkuat oleh penelitian ([15], [16], [17]) yang menggunakan algoritma *machine learning* untuk mengklasifikasikan kepuasan pengguna terhadap layanan keuangan digital melalui ulasan *online*.

Dalam praktiknya, analisis sentimen bergantung pada kemampuan algoritma *machine learning* dalam mengklasifikasikan data teks secara akurat. Berbagai studi menunjukkan variasi hasil performa *Logistic Regression*, SVM, dan *Random Forest* dalam tugas klasifikasi sentimen. *Logistic Regression* umumnya unggul dalam data teks linear separable dengan dimensi tinggi [18]. Sebaliknya, SVM menunjukkan performa baik pada dataset besar dengan *boundary* yang kompleks [19]. *Random Forest* cenderung kompetitif pada dataset dengan fitur numerik namun tidak selalu optimal untuk data *sparsed text* seperti TF–IDF [20].

Namun, penelitian sebelumnya menunjukkan adanya variasi dalam performa berbagai algoritma, terutama saat diterapkan pada jenis dan karakteristik data teks yang berbeda. Selain itu, belum ada penelitian yang secara khusus menganalisis perbandingan ketiga algoritma dalam konteks ulasan GoPay dengan pelabelan yang berdasarkan lexicon. Kekurangan ini menjadi perhatian dalam penelitian ini, yang memiliki tujuan untuk menemukan algoritma paling efektif dalam memahami pandangan masyarakat terhadap layanan GoPay melalui ulasan pengguna. Penelitian ini menganalisis seberapa akurat, presisi, recall, dan f1-score dari algoritma Logistic Regression, SVM, dan Random Forest dalam menilai sentimen terhadap aplikasi GoPay. Metode ini dipilih karena memungkinkan peneliti untuk secara objektif mengukur dan membandingkan kinerja setiap model dengan menggunakan data yang sama dalam analisis sentimen.

II. METODE

Penelitian menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional. Pendekatan ini dipilih karena mampu menyajikan analisis yang objektif melalui data numerik yang dapat diukur dan diuji secara statistik [21]. Tahapan yang dilakukan peneliti dalam penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

A. Data Collection

Pengumpulan data dilakukan dengan mengambil ulasan pengguna aplikasi Gopay yang tersedia secara publik di platform Google Playstore. Proses ini dilakukan menggunakan teknik *web scraping* untuk mempermudah pengambilan data dalam jumlah besar [22]. Meskipun setiap entri memiliki beberapa atribut, penelitian ini hanya memfokuskan pada kolom *content*, yaitu bagian yang memuat teks ulasan dari pengguna.

B. Data Understanding

Data understanding merupakan tahapan yang krusial dalam proses *data mining*, karena memungkinkan pemahaman mendalam terhadap struktur, kualitas, dan distribusi data sebelum dilakukan analisis [23]. Pemahaman yang mendalam terhadap data dapat meningkatkan interpretabilitas dan keandalan model *machine learning* [24]. Pada tahapan ini dilakukan eksplorasi awal untuk memahami struktur data, serta mengecek adanya data duplikat atau *noise*.

C. Data Preparation

Data preparation adalah langkah krusial yang diambil untuk mengolah data mentah sehingga siap dipakai dalam tahap analisis berikutnya. [25]. Tahap ini bertujuan untuk membersihkan dan menyiapkan data teks agar layak digunakan dalam proses pelatihan model. Langkah-langkah yang dilakukan meliputi:

1. *Cleaning*

   Fungsi *cleaning* didasarkan untuk menghapus bagian-bagian yang tidak penting dari teks yang dapat menghambat proses analisis [26]. Beberapa proses *cleaning* dilakukan, seperti menghapus tanda baca, menghapus angka, menghapus karakter khusus seperti simbol @, #, &, dan sejenisnya.

2. *Casefolding*

   Fungsi *casefolding* dipakai untuk mengubah semua huruf dalam tulisan menjadi huruf kecil [22]. Tujuannya adalah untuk menyamakan penyajian kata yang seharusnya sama, tetapi ditulis dengan kapital atau huruf kecil yang berbeda.

3. *Stemming*

   Proses *stemming* berfungsi untuk mengubah kata ke bentuk dasar ([20], [27]). Proses ini penting dalam Bahasa Indonesia karena banyak kata turunan seperti membayar, pembayaran, dibayar yang memiliki akar kata bayar [28]. Algoritma *stemming* yang digunakan dalam penelitian ini berasal dari pustaka NLP Bahasa Indonesia seperti *Sastrawi*.

4. *Slangword*

   Fungsi ini digunakan untuk Menerjemahkan istilah non-formal atau bahasa sehari-hari ke dalam bentuk kata yang baku. Hal ini penting karena banyak ulasan yang menggunakan bahasa tidak formal atau percakapan sehari-hari.

5. *Tokenizing*

   Tahap *tokenizing* bertujuan untuk untuk membagi teks menjadi bagian-bagian kata *(tokens)* [29]. Ini adalah tahap awal sebelum melakukan analisis lebih lanjut.

6. *Filtering*

   Proses *filtering* dilakukan untuk menghilangkan *stopwords*, yang merupakan kata-kata yang tidak memiliki arti signifikan dalam kegiatan klasifikasi, seperti yang, dan, di, ke, dari, ini, itu. Penghapusan *stopword* membantu mengurangi ukuran data dan meningkatkan perhatian pada kata-kata yang penting [30].

7. *Labelling*

   Labelling merupakan proses pemberian kategori pada data berdasarkan karakteristik yang dimilikinya. Pada penelitian ini pengelompokkan teks atau pernyataan dalam analisis sentimen ke dalam kelas positif memiliki nilai minimal nol atau lebih, dan negatif kurang dari nol dengan menggunakan kamus *lexicon* bahasa Indonesia yang tersedia dalam Github.

D. Visualization

Untuk mendukung pemahaman lebih lanjut terhadap data dan pola distribusi sentimen, dilakukan *visualization* menggunakan *bar chart, wordcloud, dan line chart. Visualization* ini membantu menilai jumlah data di setiap kelas dan kata-kata yang paling sering muncul dalam setiap jenis sentimen.

E. Modelling

Data terbagi menjadi dua kategori utama, yakni data latih dan data uji. Proses pembagian ini dilakukan dengan memanfaatkan fungsi *train_test_split* yang ada dalam pustaka *Scikit-learn*, dengan perbandingan 80% untuk data pelatihan dan 20%. Pembagian ini bertujuan untuk menyediakan data yang cukup bagi model untuk proses pembelajaran, sekaligus memastikan tersedianya data validasi yang memadai untuk menguji performa model secara objektif.

Representasi fitur dilakukan menggunakan TF–IDF. Metode ini mentransformasikan data tekstual menjadi representasi numerik melalui pemberian bobot pada tiap istilah. Perhitungan bobot tersebut didasarkan pada intensitas kemunculan kata dalam sebuah dokumen spesifik dibandingkan dengan distribusinya di seluruh kumpulan korpus. Transformasi dilakukan dengan melatih TF–IDF pada data latih menggunakan fungsi *fit_transform*, sementara data uji diubah dengan menggunakan *transform* agar bobot tetap sama. Matriks hasil ekstraksi fitur tersebut kemudian digunakan sebagai *input* pada tahap pemodelan.

Tahap pemodelan melibatkan pengembangan sistem klasifikasi yang mengevaluasi tiga algoritma pembelajaran mesin: Logistic Regression, Support Vector Machine (SVM), dan Random Forest. Tahap awal dimulai dengan melatih setiap model pada dataset latih menggunakan konfigurasi parameter default, yang bertujuan untuk menetapkan standar kinerja (baseline) awal sebelum dilakukan pengembangan lebih lanjut.

1. *Logistic Regression*

Logistic Regression merupakan pendekatan statistik yang digunakan untuk menganalisis asosiasi antara variabel independen dan variabel dependen bertipe kategori. Metode ini berfokus pada bagaimana sekumpulan prediktor dapat menjelaskan variasi dalam kelompok atau kelas yang menjadi target klasifikasi. *Logistic Regression* menghitung probabilitas bahwa suatu teks mengandung sentimen positif atau negatif.

$$p(y = 1|x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_n x_n)}} \quad (1)$$

Di mana:
$y \in \{0,1\}$ = 0 untuk sentimen negatif, 1untuk sentimen positif,
$x_1 \dots x_n$ = fitur teks,
β_0 = bias/*intercept*.

Keputusan kalsifikasi:
Jika $p(y = 1|x) \geq 0.5$, maka prediksi sentimen positif,
Jika $p(y = 1|x) < 0.5$, maka prediksi sentimen negatif.

2. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) berupaya untuk menemukan *hyperplane* yang paling optimal yang memisahkan dua kategori data dengan jarak maksimum.

$$f(x) = w^T x + b \quad (2)$$

$$y = sign(f(x)) \quad (3)$$

Di mana:
w = vektor bobot fitur,
x = vektor fitur dari teks,
b = bias/*intercept*,
sign = fungsi tanda yang menentukan kelas *output* (+1 atau -1).

3. *Random Forest*

Random Forest adalah metode yang menggunakan teknik *ensemble learning* dengan mengintegrasikan sejumlah pohon keputusan untuk menghasilkan prediksi yang lebih tepat dan konsisten.

$$\hat{y} = mode(T_1(x), T_2(x), \dots, T_1(x)) \quad (4)$$

Di mana:
$T_i(x)$ = hasil prediksi dari pohon keputusan ke-i,
$\hat{y}$ = label sentimen hasil akhir (positif/negatif),
x = fitur teks.

Implementasi model dilakukan menggunakan fitur hasil ekstraksi TF–IDF sebagai *input*, sehingga setiap teks direpresentasikan dalam bentuk matriks numerik yang mencerminkan relevansi kata terhadap dokumen dan korpus ulasan secara keseluruhan. Selanjutnya dilakukan proses *tuning hiperparameter* menggunakan

GridSearchCV untuk memperoleh konfigurasi terbaik dari masing-masing algoritma. Pada *Logistic Regression*, parameter regularisasi C diuji pada tiga nilai, yaitu 0.1, 1, dan 10. Pada SVM, dua kombinasi *hiperparameter* dievaluasi, yaitu nilai C (0.1, 1, dan 10) serta jenis *kernel* (*linear* dan RBF). Sementara itu, pada Random Forest, jumlah *estimators* dicoba pada tiga konfigurasi, yaitu 50, 100, dan 200. Setiap model dilatih menggunakan skema lima lipat *cross-validation* untuk memastikan stabilitas dan replikabilitas hasil, sebelum kemudian digunakan untuk memprediksi label sentimen pada data uji. Validasi dilaksanakan dengan metode Stratified K-Fold Cross Validation menggunakan k=5.. Hasil evaluasi dilaporkan dalam bentuk rata-rata dan standar deviasi untuk menggambarkan stabilitas performa model.

F. Evaluation

Pengukuran kinerja model dilakukan melalui pengujian terhadap data uji dengan mengacu pada matriks evaluasi standar. Parameter yang digunakan untuk menilai kualitas hasil klasifikasi ini mencakup nilai akurasi, presisi, recall, dan F1-score.

1. *Accuracy*

   Menilai seberapa banyak prediksi yang akurat jika dibandingkan dengan jumlah keseluruhan prediksi:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

2. *Precision*

   Mengukur *accuracy* prediksi positif:

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

3. *Recall*

   Mengukur kemampuan model dalam menangkap semua data positif yang sebenarnya:

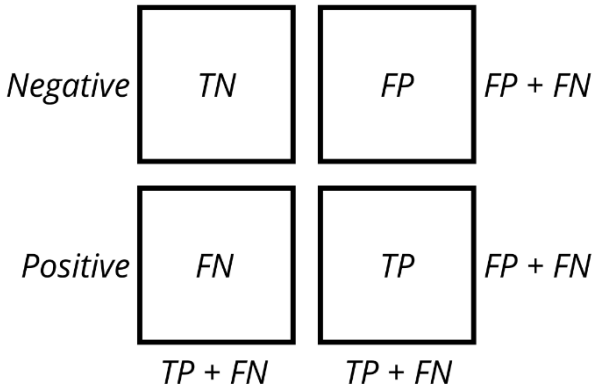
$$Recall = \frac{TP}{TP+FN} \quad (7)$$

4. *F1-score*

   Merupakan rata-rata harmonis dari *Precision* dan *Recall*:

$$F1 - score = 2 . \frac{Precision \, . \, Recall}{Precision+Recall} \quad (8)$$

Selain itu, dilakukan penilaian menggunakan *Confusion Matrix*, merupakan tahap penting dalam mengukur kinerja algoritma. *Confusion Matrix* berfungsi untuk mengukur sejauh mana algoritma mampu melakukan prediksi dengan akurat [31]. Dalam proses klasifikasi, digunakan empat indikator utama seperti yang ditunjukkan pada Gambar 2, menggambarkan hasil dari klasifikasi. Berdasarkan *evaluation* ini, dihitung nilai *recall, precision,* dan *accuracy,* yang semuanya termassuk indikator penting untuk menilai kinerja algoritma secara keseluruhan [32].



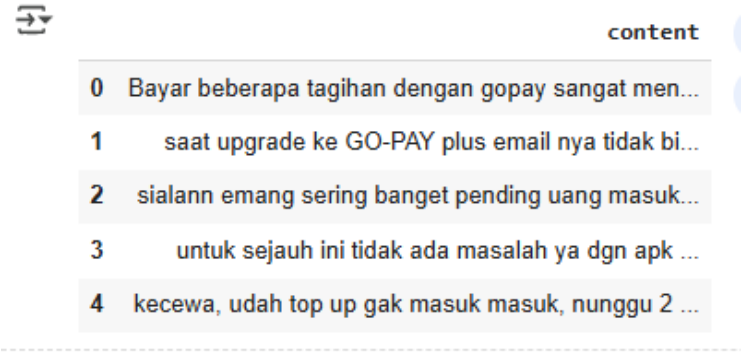
Gambar 2. *Confusion Matrix*

Keterangan:
TP = Data positif diidentifikasi dengan tepat
TN= Data negatif diidentifikasi dengan akurat
FP = Data negatif diidentifikasi sebagai positif
FN= Data positif diidentifikasi sebagai positif

III. HASIL DAN PEMBAHASAN

A. *Data Collection*

Data Collection dilakukan melalui metode *web scraping* dari Google Play Store dengan memanfaatkan *library* Python *google-play-scraper* di google colab, dilakukan pada 3.000 baris data yang masuk hingga 21 April 2025. Pengambilan data mengikuti prinsip etika pengumpulan data publik dan hasilnya disimpan dalam bentuk *file* gopay.csv. Gambar 3 merupakan cuplikan isi *dataset* yang akan digunakan.

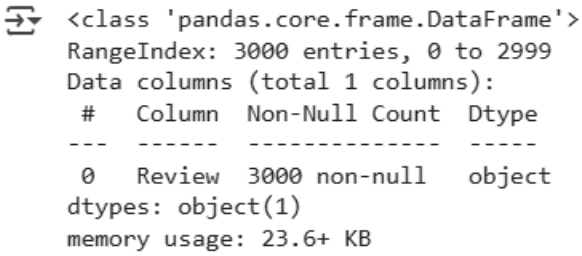


	content
0	Bayar beberapa tagihan dengan gopay sangat men...
1	saat upgrade ke GO-PAY plus email nya tidak bi...
2	sialann emang sering banget pending uang masuk...
3	untuk sejauh ini tidak ada masalah ya dgn apk ...
4	kecewa, udah top up gak masuk masuk, nunggu 2 ...

Gambar 3. Isi *Dataset*

B. *Data Understanding*

Setelah data terkumpul, dilakukan eksplorasi awal untuk memahami struktur data, serta mengecek adanya data duplikat atau *noise*. Gambar 4 menunjukkan informasi *dataset*.



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3000 entries, 0 to 2999
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Review  3000 non-null   object
dtypes: object(1)
memory usage: 23.6+ KB
```

Gambar 4. Informasi *Dataset*

C. *Data Preparation*

Tahap ini berfokus pada pengolahan dan persiapan data teks untuk memastikan bahwa data tersebut siap dipakai dalam pelatihan model. Dari keseluruhan proses yang telah dilakukan, mulai dari proses *cleaning*, *casefolding*, *stemming*, *slangword*, *tokenizing*, *filtering*, dan *labelling* menghasilkan *data preparation* yang terdapat pada Tabel 1.

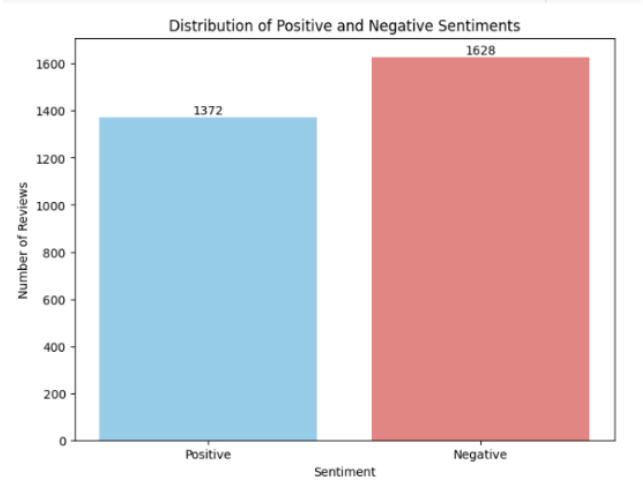
TABEL 1
HASIL *DATA PREPARATION*

Proses	Hasil
Dataset Awal	Setiap ingin transaksi dengan Qris pasti eror tidak bisa, padahal saya sudah memasukkan pin yg benar. ingin top up dari store ig pun tidak bisa.
Cleaning	Setiap ingin transaksi dengan Qris pasti eror tidak bisa padahal saya sudah memasukkan pin yg benar ingin top up dari store ig pun tidak bisa
Casefolding	setiap ingin transaksi dengan qris pasti eror tidak bisa padahal saya sudah memasukkan pin yg benar ingin top up dari store ig pun tidak bisa
Stemming	tiap ingin transaksi dengan qris pasti eror tidak bisa padahal saya sudah masuk pin yg benar ingin top up dari store ig pun tidak bisa
Slangword	tiap ingin transaksi dengan qris pasti eror tidak bisa padahal saya sudah masuk pin yang benar ingin top up dari store ig pun tidak bisa

Tokenizing	['tiap', 'ingin', 'transaksi', 'dengan', 'qris', 'pasti', 'eror', 'tidak', 'bisa', 'padahal', 'saya', 'sudah', 'masuk', 'pin', 'yang', 'benar', 'ingin', 'top', 'up', 'dari', 'store', 'ig', 'pun', 'tidak', 'bisa']
Filtering	['transaksi', 'qris', 'eror', 'masuk', 'pin', 'top', 'store', 'ig']
Labelling	[(transaksi, positive, 2), (qris, positive, 1), (eror, negative, -5), (masuk, negative, -3), (pin, positive, 1), (top, positive, 4), (store, positive, 2), (ig, negative, -1)]

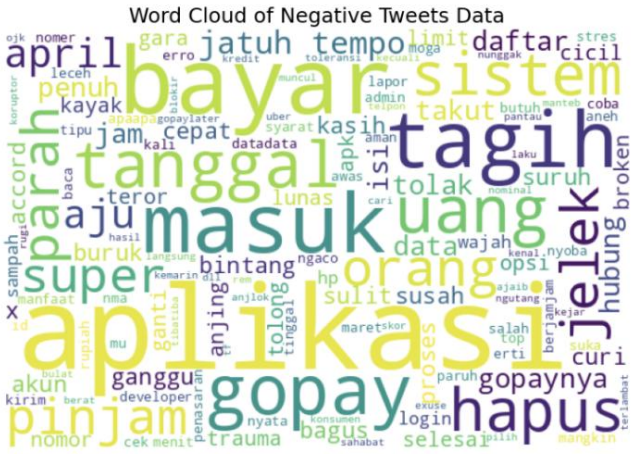
D. Visualization

Visualization dilakukan untuk menggambarkan sebaran dan kecenderungan sentimen pengguna terhadap aplikasi Gopay. Gambar 5 memvisualisasikan sebaran sentimen yaevalutng ditampilkan dalam bentuk *bar chart* yang menunjukkan jumlah ulasan positif dan negatif. Berdasarkan *visualization* tersebut, terlihat bahwa ulasan dengan sentimen negatif sedikit lebih banyak dibandingkan ulasan positif, yaitu masing-masing sebanyak 1.628 dan 1.372 ulasan, sehingga rasio kelas relatif seimbang. Oleh karena itu, penelitian ini tidak menerapkan penanganan khusus untuk ketidakseimbangan kelas, dan *class_weight* dipertahankan dalam nilai *default*. Namun, untuk menghindari bias dalam proses pelatihan, *Stratified K-Fold* digunakan agar proporsi kelas tetap konsisten pada setiap *fold*.



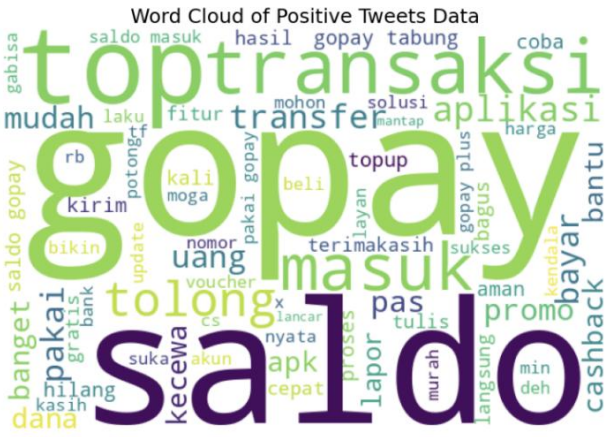
Gambar 5. Sebaran Sentimen

Selanjutnya, untuk menemukan kata-kata dominan dalam setiap kategori sentimen, dilakukan *visualization* dalam bentuk *wordcloud*. *Wordcloud* untuk ulasan dengan sentimen negatif, Gambar 6 menunjukkan dominasi kata-kata seperti "bayar", "tagih", "hapus", dan "gagal", yang mencerminkan adanya keluhan terkait sistem pembayaran dan penggunaan aplikasi.



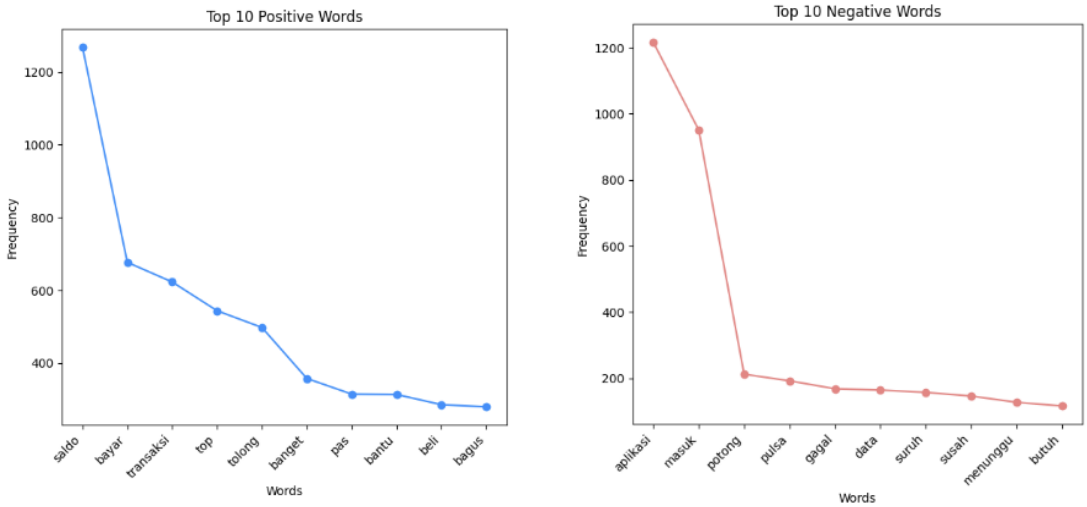
Gambar 6. *Wordcloud Negative*

Sebaliknya, Gambar 7 menampilkan *wordcloud* dari ulasan positif. Dalam *visualization* ini, kata-kata seperti “saldo”, “top”, “transaksi”, dan “tolong” mendominasi, menunjukkan aspek-aspek layanan yang dirasakan bermanfaat dan memuaskan oleh pengguna.



Gambar 7. *Wordcloud Positive*

Untuk memperkuat temuan dari *wordcloud*, ditampilkan pula grafik frekuensi kata pada masing-masing kategori sentimen. Gambar 8(a) menunjukkan sepuluh kata dengan frekuensi tertinggi pada ulasan positif, di mana kata "saldo", "bayar", dan "transaksi" muncul paling sering. Sementara itu, Gambar 8(b) menampilkan sepuluh kata yang dominan dalam ulasan negatif, seperti "aplikasi", "masuk", dan "potong", yang mengindikasikan permasalahan atau ketidakpuasan pengguna.



(a) (b)

Gambar 8. Frekuensi (a) sentimen positif dan (b) sentimen negatif

E. Evaluation

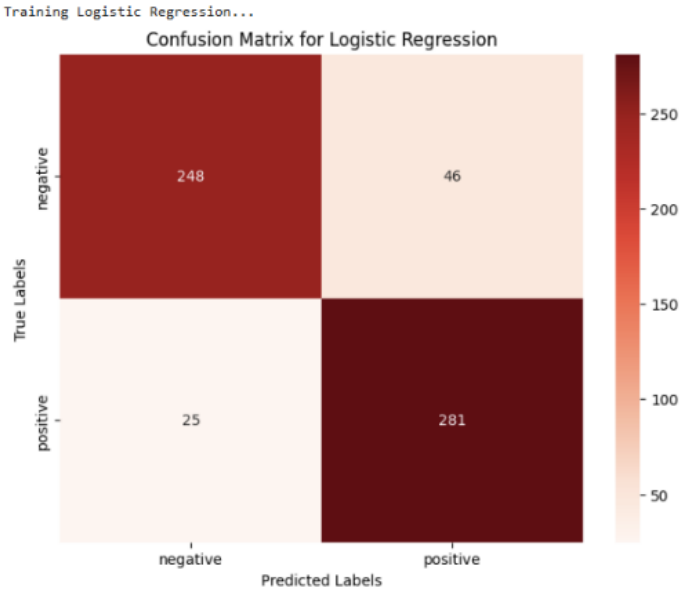
Evaluation terhadap performa model dilakukan setelah proses pelatihan menggunakan data latih selesai. Tiga algoritma klasifikasi yang dipakai dalam penelitian ini adalah *Logistic Regression, Support Vector Machine (SVM),* dan *Random Forest. Evaluation* performa dilakukan berdasarkan beberapa matriks utama, yaitu *accuracy, precision, recall,* dan *f1-score,* serta didukung oleh analisis *Confusion Matrix*.

```
Training Logistic Regression...
Training SVM...
Training Random Forest...
                 Model  Test Accuracy
0  Logistic Regression       0.881667
1                  SVM       0.875000
2        Random Forest       0.793333
```

Gambar 9. *Accuracy Model*

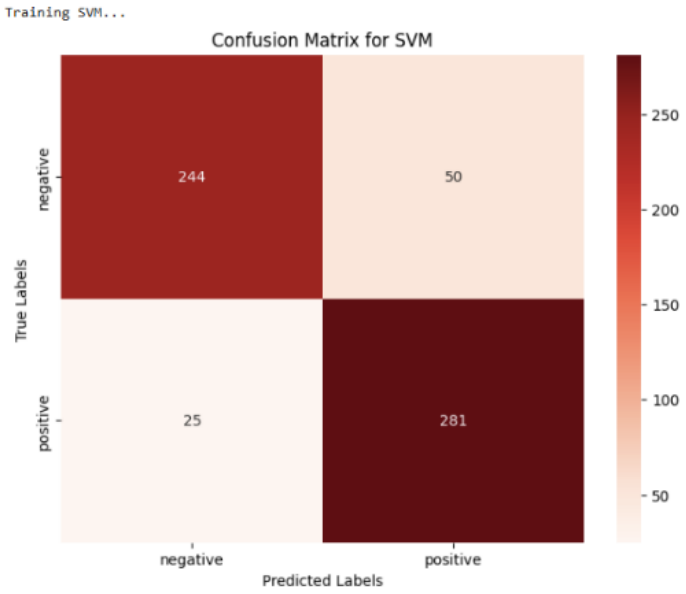
Berdasarkan hasil *evaluation* yang ditampilkan Gambar 9, model *Logistic Regression* menunjukkan *accuracy* tertinggi sebesar 88,16%, diikuti oleh SVM dengan *accuracy* sebesar 87,50%, dan *Random Forest* dengan accuracy sebesar 79,33%. Hasil ini mengindikasikan bahwa *Logistic Regression* memiliki kinerja paling optimal dalam melakukan klasifikasi sentimen ulasan pengguna Gopay. Hasil *evaluation* menunjukkan bahwa performa model sangat dipengaruhi oleh kualitas tahapan preprocessing teks. *Slangword* dan *stopword removal* berperan penting dalam meningkatkan kejelasan makna setiap kalimat, sehingga model dapat mempelajari token yang benar-benar relevan terhadap sentimen. *Stemming* juga berkontribusi dengan mengurangi variasi bentuk kata, sehingga membantu TF–IDF menangkap informasi berbasis makna, bukan permukaan kata. Di sisi lain, pendekatan berbasis leksikon digunakan pada tahap pelabelan, bukan pada tahap ekstraksi fitur, sehingga model tetap belajar berdasarkan representasi TF–IDF secara statistikal, bukan sekadar pencocokan kata positif dan negatif. Kombinasi *preprocessing* tersebut menghasilkan fitur yang lebih bersih dan stabil, yang terbukti mendukung performa model terutama pada Logistic Regression dan SVM.

Lebih lanjut, analisis *Confusion Matrix* memperkuat temuan ini. *Confusion Matrix* untuk model *Logistic Regression* ditampilkan pada Gambar 10. Hasil pengujian menunjukkan bahwa model mampu mengidentifikasi secara akurat 248 ulasan bersentimen negatif dan 281 ulasan positif. Meskipun demikian, analisis kesalahan mencatat adanya 25 data false negative (positif yang terdeteksi negatif) serta 46 data false positive (negatif yang terdeteksi positif). Hal ini menegaskan bahwa Logistic Regression memiliki stabilitas prediksi yang baik dengan margin kesalahan yang minimal. Performa model tampak sangat efektif pada kategori negatif, yang dibuktikan dengan capaian nilai recall yang signifikan.



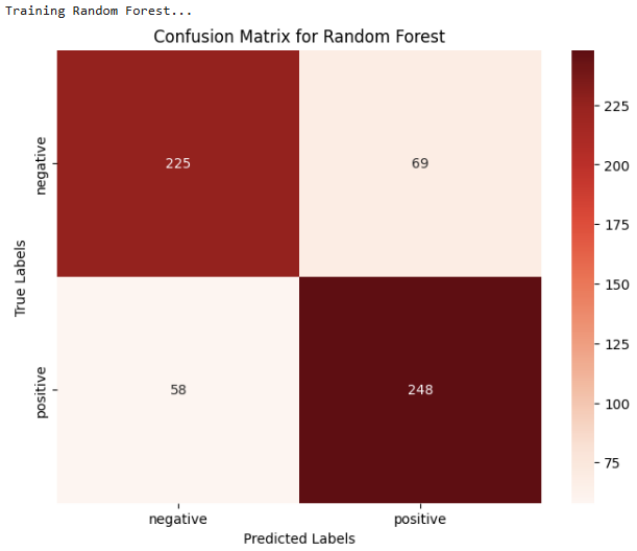
Gambar 10. *Confusion Matrix Logistic Regression*

Confusion Matrix untuk model SVM disajikan pada Gambar 11. Model ini menghasilkan klasifikasi yang hampir serupa dengan *Logistic Regression*, di mana 244 ulasan negatif dan 281 ulasan positif berhasil diklasifikasikan dengan benar. Kesalahan klasifikasi yang tercatat adalah 25 data yang sebenarnya posistif diklasifikasikan sebagai negatif dan 50 data sebenarnya negatif salah diklasifikasikan sebagai positif. Dibandingkan *Logistic Regression*, SVM menunjukkan performa yang sedikit lebih rendah terutama pada kemampuan mendeteksi sentimen positif, sebagaimana ditunjukkan dari nilai *recall* yang rendah pada kelas tersebut.



Gambar 11. *Confusion Matrix SVM*

Gambar 12 menampilkan *Confusion Matrix* untuk model *Random Forest*. Model ini menunjukkan jumlah kesalahan klasifikasi yang lebih tinggi dibandingkan dua model sebelumnya. Sebanyak 225 ulasan negatif dan 248 ulasan positif berhasil diklasifikasikan dengan benar. Namun, terdapat 58 data yang sebenarnya positif telah salah diidentifikasi sebagai negatif dan 69 data yang sebenarnya negatif diperkirakan sebagai positif. Tingkat kesalahan ini berdampak pada menurunnya nilai precision dan *recall* secara keseluruhan. Hal ini menunjukkan bahwa model *Random Forest* tidak cukup efisien dalam mengenali pola pada data teks yang diterapkann dalam analisis sentimen.



Gambar 12. *Confusion Matrix Random Forest*

Jika dibandingkan secara keseluruhan, model *Logistic Regression* menunjukkan tingkat kesalahan klasifikasi terendah dengan distribusi prediksi yang paling seimbang antara sentimen positif dan negatif. SVM memiliki performa yang hampir setara, meskipun terjadi sedikit peningkatan jumlah kesalahan pada kategori sentimen negatif. Sementara itu, *Random Forest* menampilkan hasil terburuk di antara ketiga model, dengan jumlah kesalahan klasifikasi yang cukup signifikan pada kedua kategori sentimen. Oleh karena itu, dapat disimpulkan bahwa *Logistic Regression* merupakan model dengan kinerja paling stabil dan akurat dalam tugas klasifikasi sentimen berbasis teks pada dataset ini.

```
Training Logistic Regression...
Best parameters: {'C': 10}
Accuracy on testing set: 0.88
              precision    recall  f1-score   support

    negative       0.86      0.92      0.89       306
    positive       0.91      0.84      0.87       294

    accuracy                           0.88       600
   macro avg       0.88      0.88      0.88       600
weighted avg       0.88      0.88      0.88       600
```

(a)

```
Training SVM...
Best parameters: {'C': 10, 'kernel': 'rbf'}
Accuracy on testing set: 0.88
              precision    recall  f1-score   support

    negative       0.85      0.92      0.88       306
    positive       0.91      0.83      0.87       294

    accuracy                           0.88       600
   macro avg       0.88      0.87      0.87       600
weighted avg       0.88      0.88      0.87       600
```

(b)

```
Training Random Forest...
Best parameters: {'n_estimators': 200}
Accuracy on testing set: 0.79
              precision    recall  f1-score   support

    negative       0.78      0.82      0.80       306
    positive       0.80      0.77      0.78       294

    accuracy                           0.79       600
   macro avg       0.79      0.79      0.79       600
weighted avg       0.79      0.79      0.79       600
```

(c)

Gambar 13. Hasil *evaluation precisison, recall,* dan *f1-score* pada (a) *Logistic Regression,* (b) *SVM,* dan (c) *Random Forest*

Selain *accuracy*, hasil *evaluation precision, recall,* dan *f1-score* yang ditunjukkan oleh gambar 13(a), 13(b), 13(c), menunjukkan bahwa *Logistic Regression* dan SVM memiliki nilai matriks yang seimbang antara kedua kelas sentimen, sedangkan *Random Forest* cenderung menghasilkan nilai yang lebih rendah pada seluruh *evaluation matrix*. Tabel 2 merangkum perbandingan performa ketiga model berdasarkan matriks *accuracy, precision, recall,* dan *f1-score.*

TABEL 2
PERBANDINGAN PERFORMA MODEL

Model	***Accuracy***	***Precision*** **(Positif)**	***Recall*** **(Positif)**	***F1-Score*** **(Positif)**	***Precision*** **(Negatif)**	***Recall*** **(Negatif)**	***F1-Score*** **(Negatif)**
Logistic Regression	0.8816	0.91	0.84	0.87	0.86	0.92	0.89
SVM	0.8750	0.91	0.83	0.87	0.85	0.92	0.88
Random Forest	0.7933	0.80	0.77	0.78	0.78	0.82	0.80

Logistic Regression menunjukkan *accuracy* tertinggi serta keseimbangan matriks yang baik antara kelas sentimen positif dan negatif. SVM juga menunjukkan kinerja yang bersaing dengan perbedaan *accuracy* yang kecil. Sebaliknya, algoritma Random Forest menunjukkan kinerja yang kurang kompetitif, ditandai dengan capaian nilai precision, recall, dan F1-score yang berada di bawah performa dua model lainnya. Berdasarkan perbandingan komprehensif tersebut, Logistic Regression ditetapkan sebagai arsitektur yang paling efektif dan relevan untuk menjalankan tugas klasifikasi sentimen dalam studi ini.

IV. SIMPULAN

Penelitian ini dilakukan untuk menjawab pertanyaan penelitian mengenai model klasifikasi sentimen yang paling optimal dalam menganalisis ulasan pengguna terhadap layanan GoPay. Dengan membandingkan tiga algoritma yaitu *Logistic Regression, Support Vector Machine* (SVM), dan *Random Forest* menggunakan representasi fitur TF–IDF, diperoleh temuan bahwa *Logistic Regression* memberikan performa terbaik, ditunjukkan melalui nilai akurasi rata-rata tertinggi pada skema *Stratified 5-Fold Cross Validation*. Dengan demikian, pertanyaan penelitian terjawab, *Logistic Regression* merupakan model paling efektif untuk analisis

sentimen pada data ulasan GoPay. Secara praktis, temuan ini memberikan implikasi bahwa *Logistic Regression* dapat direkomendasikan untuk implementasi monitoring sentimen ulasan aplikasi GoPay di lingkungan produksi, karena stabil, efisien secara komputasi, dan tidak memerlukan konfigurasi *hiperparameter* yang kompleks. Hal ini memungkinkan perusahaan melakukan analisis sentimen secara otomatis, *real-time*, dan berkelanjutan guna mendukung pengambilan keputusan dalam peningkatan layanan. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Proses pelabelan menggunakan pendekatan *lexicon* berpotensi menghasilkan bias semantik pada kalimat dengan ambiguitas konteks. Selain itu, data yang digunakan berasal dari satu domain platform (Google Play Store), sehingga kemungkinan terjadi *domain-shift* ketika model diterapkan pada platform lain seperti Twitter atau TikTok dengan karakteristik bahasa yang berbeda. Penelitian ini juga belum melakukan analisis dampak masing-masing tahapan preprocessing *(ablation study)*, sehingga pengaruh spesifik dari *data preparation* belum dapat disimpulkan secara *independen*. Untuk penelitian lanjutan, beberapa arah pengembangan potensial meliputi penerapan *large pre-trained language models* seperti BERT atau IndoBERT, arsitektur LSTM/GRU, domain *adaptation* untuk analisis lintas platform, serta *cost-sensitive learning* untuk menangani kemungkinan ketidakseimbangan kelas di *dataset real-world*. Selain itu, dataset berlabel manual atau skema pelabelan *semi-supervised* dapat dipertimbangkan untuk mengurangi bias akibat pelabelan berbasis *lexicon*.

DAFTAR PUSTAKA

[1] M. Arif, "Jumlah Pengguna Internet Indonesia Tembus 221 Juta Orang," APJII. Diakses: 1 Mei 2025. [Daring]. Tersedia pada: https://apjii.or.id/berita/d/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang?

[2] M. I. Riyadi, "The 6th Indonesia Fintech Summit & Expo (IFSE) & Bulan Fintech Nasional (BFN) 2024," OJK. [Daring]. Tersedia pada: https://www.ojk.go.id/id/berita-dan-kegiatan/siaran-pers/Pages/Dorong-Literasi-dan-Inklusi-Keuangan-Digital-Serta-Perkuat-Ekosistem-Fintech-BFN-IFSE-2024.aspx?

[3] Hallokalsel, "Perkembangan Teknologi Keuangan (Fintech) di Indonesia: Tren dan Inovasi 2024," HALLOKALSEL.COM. [Daring]. Tersedia pada: https://hallokalsel.com/perkembangan-teknologi-keuangan-fintech-di-indonesia-tren-dan-inovasi-2024

[4] M. Ashoer, C. Jebarajakirthy, X. J. Lim, M. Mas'ud, dan Z. A. Sahabuddin, "Mobile fintech, digital financial inclusion, and gender gap at the bottom of the pyramid: An extension of mobile technology acceptance model," *Procedia Comput. Sci.*, vol. 234, no. 2023, hal. 1253–1260, 2024, doi: 10.1016/j.procs.2024.03.122.

[5] A. P. Rabbani, A. Alamsyah, dan S. Widiyanesti, "An Effort to Measure Customer Relationship Performance in Indonesia's Fintech Industry," *arXiv*, 2021.

[6] M. Diva dan M. I. Anshori, "Penggunaan E-Wallet Sebagai Inovasi Transaksi Digital: Literatur Review," *Mult. J. Glob. Multidiscip.*, vol. 2, no. 6, hal. 1991–2002, 2024, [Daring]. Tersedia pada: https://journal.institercom-edu.org/index.php/multiple

[7] B. Banutama, "Keputusan Penggunaan E-wallet Sebagai Alat Transaksi Digital : Sebuah Kajian Literatur 2012-2023," *JIAKu*, hal. 301–318, 2023.

[8] F. R. Prawira, N. T. Prakoso, P. W. Handayani, dan N. C. Harahap, "The influence of information security factors on the continuance use of electronic wallet," *Procedia Comput. Sci.*, vol. 234, no. 2023, hal. 1467–1475, 2024, doi: 10.1016/j.procs.2024.03.147.

[9] U. Rahardja, I. D. Hapsari, P. O. H. A. D. I. Putra, dan A. N. Hidayanto, "Technological readiness and its impact on mobile payment usage: A case study of go-pay," *Cogent Eng.*, vol. 10, no. 1, 2023, doi: 10.1080/23311916.2023.2171566.

[10] A. Nurherwening, A. W. Dari, D. Urumsah, dan H. T. Wibowo, "The success of go-pay financial technology service adoption," *J. Contemp. Account.*, vol. 3, no. 2, hal. 98–111, 2021, doi: 10.20885/jca.vol3.iss2.art5.

[11] N. A. Syabila dan I. Khasanah, "Analisis Pengaruh Persepsi Kemudahan Penggunaan, Manfaat, Dan Risiko Terhadap Minat Berkelanjutan Dengan Kepercayaan Sebagai Variabel," *Diponegoro J. Manag.*, vol. 12, hal. 1–15, 2023.

[12] A. Azmy, P. Subakrie, dan M. Z. Azhari, "Bisma: Jurnal Bisnis dan Manajemen THE FACTORS THAT INFLUENCE CONSUMER SATISFACTION ON GOPAY," vol. 14, no. 1, hal. 10–18, 2020, [Daring]. Tersedia pada: https://jurnal.unej.ac.id/index.php/BISMA

[13] D. A. Fitri dan Damayanti, "Komparasi Algoritma Random Forest Classifier Dan Support Vector Machine Untuk Sentimen Masyarakat Terhadap Pinjaman Online Di Media Sosial," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 4, hal. 2018–2029, 2024, [Daring]. Tersedia pada: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/5608

[14] K. P. J. Sitompul, A. R. Pratama, dan K. A. Baihaqi, "Komparasi Algoritma Naïve Bayes, Support Vector Machine, Dan Logistic Regression Pada Analisis Sentimen Pengguna Aplikasi Transportasi Online," *Kumpul. J. Ilmu Komput.*, vol. 10, no. 01, hal. 27–38, 2023.

[15] H. M. Putri, M. Faisal, dan M. Fachrul Kurniawan, "Multi Stage Analisis Sentimen Berbasis Aspek Pada Ulasan Pengguna Aplikasi Dompet Digital Menggunakan Metode Multinomial Naïve Bayes," *Indones. J. Comput. Sci.*, vol. 13, no. 5, hal. 8018–8028, 2024.

[16] H. Wisnu, M. Afif, dan Y. Ruldevyani, "Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes," *J. Phys. Conf. Ser.*, vol. 1444, no. 1, 2020, doi: 10.1088/1742-6596/1444/1/012034.

[17] H. Adiningtyas dan A. S. Auliani, "Sentiment analysis for mobile banking service quality measurement," *Procedia Comput. Sci.*, vol. 234, hal. 40–50, 2024, doi: 10.1016/j.procs.2024.02.150.

[18] A. G. Budianto, A. Trisno, E. Suryo, dan G. Rudi, "Perbandingan Performa Algoritma Support Vector Machine (SVM) dan Logistic Regression untuk Analisis Sentimen Pengguna Aplikasi Retail di Android," vol. 10, no. November, hal. 1–10, 2024, doi: 10.34128/jsi.v10i2.911.

[19] I. Septiana dan D. Alita, "Perbandingan Random Forest dan SVM dalam Analisis Sentimen Quick Count Pemilu 2024," *JPIT*, vol. 9, no. 3, hal. 224–233, 2024, doi: 10.30591/jpit.v9i3.6640.

[20] Indriani dan A. Davy Wiranata, "Comparison Of Accuracy Levels Of SVM, Decision Tree And Random Forest Algorithms In Sentiment Analysis Of User Responses Of The Gopay Application," *J. Tek. Inform.*, vol. 5, no. 3, hal. 777–787, 2024, [Daring]. Tersedia pada: https://doi.org/10.52436/1.jutif.2024.5.3.1885

[21] J. Mackiewicz, *A Mixed-Method Approach*. 2018. doi: 10.4324/9780429469237-3.

[22] D. Cahyo Ramadhan dan F. Irwiensyah, "Analisis Sentimen Pengguna Terhadap Aplikasi Bing Chat di Google Play Store dengan Metode Naïve Bayes," *Media Online*, vol. 4, no. 5, hal. 2410–2418, 2024, doi: 10.30865/klik.v4i5.1769.

[23] Sreejit Ramakrishnan, "The Importance of Data Mining & Predictive Analysis," *Int. J. Eng. Technol. Manag. Sci.*, vol. 7, no. 4, hal. 593–598, 2023, doi: 10.46647/ijetms.2023.v07i04.081.

[24] N. Mehdiyev, M. Majlatow, dan P. Fettke, "Interpretable and Explainable Machine Learning Methods for Predictive Process Monitoring: A Systematic Literature Review," *arXiv*, hal. 1–73, 2023, [Daring]. Tersedia pada: http://arxiv.org/abs/2312.17584

[25] N. Wahyuningsih dan H. Hendry, "Perbandingan Metode Klasifikasi Dalam Analisis Sentimen Masyarakat Terhadap Identitas Kependudukan Digital (Ikd)," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 8, no. 4, hal. 1218–1227, 2023, doi: 10.29100/jipi.v8i4.4155.

[26] O. N. Mbadiwe dan A. I. Otuonye, "Challenges of Data Collection and Preprocessing for Phishing Email Detection," *Int. J. Comput. Sci. Softw. Eng.*, vol. 11, no. 2, 2024, doi: 10.5281/zenodo.12651047.

[27] Pande sindu, Agus Aan Jiwa Permana, dan I Nyoman Saputra Wahyu Wijaya, "Identifikasi Dan Normalisasi Teks Slang Dengan Fasttext Pada Twitter Dalam Bahasa Indonesia," *J. Pendidik. Teknol. dan Kejuru.*, vol. 21, no. 1, hal. 33–44, 2024, doi: 10.23887/jptkundiksha.v21i1.66381.

[28] D. S. Dewi, "Penerapan Algoritma Stemming Sastrawi Dan Cosine Similarity Pada Information Retrieval System Al-Quran Dengan Query Tren Twitter," *Braz Dent J.*, vol. 33, no. 1, hal. 1–12, 2022.

[29] S. Wulandari dan F. N. Hasan, "Analisis Sentimen Masyarakat Indonesia Terhadap Pengalaman Belanja Thrifting Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *J. Media Inform. Budidarma*, vol. 8, no. 2, hal. 768, 2024, doi: 10.30865/mib.v8i2.7520.

[30] N. Ambika Hapsari dan A. Dwi Indriyanti, "Analisis Sentimen pada Aplikasi Dompet Digital Menggunakan Algoritma Random Forest," *J. Emerg. Inf. Syst. Bus. Intell.*, vol. 04, no. 03, hal. 186–192, 2023.

[31] Jan Melvin Ayu Soraya Dachi dan Pardomuan Sitompul, "Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit," *J. Ris. Rumpun Mat. Dan Ilmu Pengetah. Alam*, vol. 2, no. 2, hal. 87–103, 2023, doi: 10.55606/jurrimipa.v2i2.1470.

[32] M. R. Sholahuddin, F. Atqiya, S. R. Wulan, M. Harika, S. Fitriani, dan Y. Sofyan, "Implementasi Sistem Identifikasi Senjata Real Time Menggunakan YOLOv7 dan Notifikasi Chat Telegram," *J. Inf. Syst. Res.*, vol. 4, no. 2, hal. 598–606, 2023, doi: 10.47065/josh.v4i2.2774.