

Perbandingan Analisis Sentimen Ulasan Pengguna Aplikasi ShopeePay pada Google Play Store Menggunakan Metode *Support Vector Machine (SVM)* dan *Decision Tree Classification*

Chellcia Mutiara Iwfanka ^{1*}, Agung Riyadi ^{2**}

* Teknik Informatika, Politeknik Negeri Batam

** Teknik Informatika, Politeknik Negeri Batam

chellciamutiara3017@gmail.com ¹, agung@polibatam.ac.id ²

Article Info

Article history:

Received 2025-12-16

Revised 2025-12-16

Accepted 2025-12-16

Keyword:

Analisis sentimen, Decision Tree Classification, ShopeePay, Support Vector Machine

ABSTRACT

Technological developments, especially in finance, have led to the rise of financial technology (fintech), including digital wallets like ShopeePay, which facilitate electronic transactions via smartphones. This study compares Support Vector Machine (SVM) and Decision Tree classification in classifying sentiments of ShopeePay user reviews from the Google Play Store using Support Vector Machine (SVM) achieved 95% accuracy, with 95% precision, 94% recall, and 94% F1-score for negative reviews, and 95% precision, 96% recall, and 96% F1-score for positive reviews. Decision Tree Classification obtained 91% accuracy, with lower recall for negative reviews 82% due to overfitting on the majority class and sensitivity to data imbalance. While Decision Tree Classification offers high interpretability, Support Vector Machine (SVM) is more reliable and consistent for sentiment classification of ShopeePay user reviews.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Terobosan teknologi terkini telah berdampak pada berbagai industri, termasuk keuangan. Teknologi keuangan (*fintech*) adalah penggunaan teknologi untuk mempercepat dan meningkatkan efisiensi layanan keuangan. Dompet digital (*e-wallet*) adalah salah satu jenis teknologi keuangan yang umum digunakan di Indonesia. Dompet digital/*e-wallet* adalah aplikasi yang memfasilitasi pengguna untuk menyimpan uang secara elektronik dan melakukan transaksi online menggunakan ponsel pintar mereka [1].

Penggunaan *E-Wallet* memberikan dampak positif karena lebih efisien dan membantu mengurangi peredaran uang tunai. *E-wallet* juga sering memberikan banyak diskon dan cashback yang mendorong masyarakat untuk beralih menggunakan metode pembayaran ini [2]. ShopeePay adalah layanan dompet digital populer yang memungkinkan berbagai macam transaksi. Menurut survei Lembaga Riset IDN yang berjudul "Laporan Generasi Milenial dan Gen Z Indonesia 2025," 77% dari 1.500 responden menggunakan ShopeePay sebagai pilihan pembayaran.

ShopeePay merupakan layanan pembayaran elektronik yang digunakan untuk melakukan transaksi di aplikasi Shopee atau merchant *offline* yang bekerja sama dengan Shopee. ShopeePay pertama kali disahkan pada tahun 2018 oleh Bank Indonesia [3]. Awalnya, ShopeePay merupakan opsi pembayaran dalam aplikasi Shopee dan pada tahun 2024, program ini dirilis sebagai dompet digital. Selanjutnya, ShopeePay terhubung dengan layanan ShopeePayLater dan hal ini yang membuat banyak masyarakat menggunakan layanan dompet digital pada Shopee. Meskipun merupakan aplikasi baru, ShopeePay telah diunduh lebih dari 10 juta kali di Google Play Store dan memiliki rating 4,8. Terlepas dari rating aplikasi yang baik, beberapa pengguna menyatakan ketidakpuasan terhadap fungsinya. Beberapa pengguna menganggap kemampuan ShopeePay bermanfaat, sementara yang lain mengalami masalah seperti kesulitan teknis atau kerusakan sistem. Biasanya, pengguna menyampaikan ulasan mereka melalui google play store. Ulasan-ulasan tersebut menggambarkan pandangan dan pengalaman pengguna terhadap aplikasi shopeepay, sekaligus memberikan informasi tentang tingkat kepuasan maupun ketidakpuasan pengguna [4].

Untuk mengetahui persepsi pengguna terhadap aplikasi, ulasan pengguna dapat dimanfaatkan untuk mengevaluasi kinerja layanan dan menetapkan tujuan perbaikan oleh perusahaan. Untuk menangani ulasan dengan baik, diperlukan analisis sentimen yang bertujuan untuk mengidentifikasi dan mengklasifikasikan sentimen atau perasaan yang terkandung dalam teks, apakah itu positif, atau negatif [5]. Pendekatan ini melibatkan keluarnya fitur dari teks ulasan dan penggunaan model klasifikasi. *Support Vector Machine* (SVM) dipilih karena dapat memisahkan dua kelas dengan akurasi tinggi, yang sangat berguna dalam analisis sentimen. Sementara itu, *Decision Tree Classification* dipilih karena mudah dipahami, memberikan gambaran yang jelas tentang bagaimana keputusan dibuat saat mengklasifikasikan sentimen dalam setiap ulasan. Tujuan penelitian ini adalah untuk mengevaluasi dan membandingkan kinerja dan efektivitas metode *Support Vector Machine* (SVM) dan *Decision Tree Classification* untuk mengklasifikasikan sentimen dalam evaluasi pengguna aplikasi ShopeePay di Google Play Store.

II. LANDASAN TEORI

A. Penelitian Terdahulu

Penelitian yang dilakukan oleh Irma Surya Kumala Idris, Yasin Aril Mustofa, dan Irvan Abraham Salihi dengan judul “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma *Support Vector Machine* (SVM)”, yang bertujuan untuk mengklasifikasikan komentar pengguna aplikasi Shopee menjadi komentar positif dan negatif, dengan mempelajari pendapat pengguna melalui ulasan yang diberikan. Selain itu, penelitian ini juga bertujuan untuk menilai kinerja metode yang digunakan. Hasil penelitian tersebut menghasilkan kinerja yang cukup baik dengan hasil akurasi sebesar 98%, dan f1-score sebesar 98% [6].

Penelitian yang dilakukan oleh Khoiril Abbi Rokhman, Berlilana, dan Primdani Arsi dengan judul “Perbandingan Metode *Support Vector Machine* dan *Decision Tree* untuk Analisis Sentimen review komentar pada aplikasi Transportasi Online”, yang bertujuan untuk membandingkan akurasi metode *Support Vector Machine* dan *Decision Tree* menggunakan data ulasan yang tersedia di Google Play Store. Hasil akurasi yang diperoleh dari metode *Support Vector Machine* sebesar 90.20% sedangkan untuk metode *Decision Tree* menghasilkan akurasi sebesar 89.80%. Dapat disimpulkan bahwa metode *Support Vector Machine* memiliki nilai akurasi yang lebih tinggi dibandingkan dengan metode *Decision Tree* [7].

Penelitian yang dilakukan oleh Indra Kurniawan, April Lia Hananto, Shofa Shofia Hilabi, Agustia Hananto, Bayu Priyatna, dan Aviv Yuniar Rahman dengan judul “Perbandingan Algoritma *Naive Bayes* dan *SVM* dalam Sentimen Analisis Marketplace pada Twitter”, yang bertujuan untuk mengetahui algoritma klasifikasi paling baik berdasarkan nilai akurasinya. Hasil akurasi yang diperoleh algoritma *Naive Bayes* untuk aplikasi Tokopedia sebesar

85.34%, Shopee 80.04% dan Lazada 83.52%. Sedangkan hasil akurasi dengan menggunakan algoritma *SVM* untuk aplikasi Shopee sebesar 86.82%, Shopee 80.91% dan Lazada 88.93%. Kesimpulan yang didapatkan dari perbandingan tersebut adalah algoritma *SVM* merupakan algoritma yang memiliki nilai akurasi terbaik [8].

Penelitian yang dilakukan oleh Nevita Cahaya Ramadani, Imam Tahyudin, Azhari Shouni Barkah dengan judul “Perbandingan Algoritma *Support Vector Machine*, *Decision Tree*, dan *Logistic Regresion* Pada Analisis Sentimen Ulasan Aplikasi Netflix”, yang bertujuan untuk menganalisis sentiment terhadap ulasan pengguna netflix. Hasil penelitian tersebut *Support Vector Machine* mencapai akurasi sebesar 88.18% [9].

Penelitian yang dilakukan oleh Maharani dan Fathoni dengan judul “Analisis Sentimen Pengguna Terhadap Faktor Penggunaan PayPal Menggunakan Metode *Decision Tree*”, yang bertujuan untuk mengetahui sentimen masyarakat terhadap layanan PayPal berdasarkan teori bauran pemasaran. Hasil dari pengujian tersebut menunjukkan bahwa metode *Decision Tree* memiliki performa yang baik dengan akurasi sebesar 97,33% untuk aspek produk, 99,68% untuk aspek harga dan 96,72% untuk aspek tempat pada pengujian menggunakan *K-Fold Cross Validation* [10].

Tabel 1 Perbandingan Klasifikasi

No.	Judul	Algoritma	Peneliti
1	<u>Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma <i>Support Vector Machine</i> (SVM)</u> ,	Support Vector Machine (SVM)	Irma Surya Kumala Idris, Yasin Aril Mustofa, dan Irvan Abraham Salihi
2	<u><i>Support Vector Machine</i> dan <i>Decision Tree</i> untuk Analisis Sentimen review komentar pada aplikasi Transportasi Online</u>	Support Vector Machine (SVM) dan Decision Tree	<u>Khoiril Abbi Rokhman, Berlilana, dan Primdani Arsi</u>
3	<u>Perbandingan Algoritma <i>Naive Bayes</i> dan <i>SVM</i> dalam Sentimen Analisis Marketplace pada Twitter</u>	Naive Bayes dan Support Vector Machine (SVM)	<u>Indra Kurniawan, April Lia Hananto, Shofa Shofia Hilabi, Agustia Hananto, Bayu Priyatna, dan Aviv Yuniar Rahman</u>
4	<u>Perbandingan Algoritma <i>Support Vector Machine</i>, <i>Decision Tree</i>, dan <i>Logistic Regresion</i> Pada Analisis Sentimen Ulasan Aplikasi Netflix</u>	Support Vector Machine, Decision Tree, dan Logistic Regresion	<u>Nevita Cahaya Ramadani, Imam Tahyudin, Azhari Shouni Barkah</u>
5	<u>Analisis Sentimen Pengguna Terhadap Faktor Penggunaan PayPal Menggunakan Metode <i>Decision Tree</i></u>	Decision Tree	Maharani, Fathoni

B. Dasar Teori

1. Data Mining

Data Mining adalah proses penggalian data untuk menemukan pola yang tersembunyi dari suatu data. Hasil dari proses penggalian data dapat digunakan untuk membuat suatu keputusan bisnis di masa depan [11].

2. Analisis Sentimen

Analisis Sentimen atau *opinion mining* adalah studi yang menganalisis data dengan banyak sudut pandang, sentimen, dan emosi yang disampaikan untuk menentukan nilai positif dan negatif [8].

3. E-Wallet

Dompot digital atau yang bisa disebut dengan *E - Wallet* adalah sebuah layanan keuangan yang memungkinkan penggunaanya dapat menyimpan uang dalam bentuk digital dan bertransaksi secara *online* [1].

4. Ulasan

Ulasan merupakan penilaian atau *review* terhadap suatu karya. Ketika sedang mengulas suatu karya, pengulas harus bersikap kritis agar hasil ulasannya memberikan kontribusi bagi kemajuan karya tersebut [12].

5. Web Scraping

Web scraping adalah kegiatan mengumpulkan suatu data yang ada pada halaman *web* di internet dengan tujuan untuk mendapatkan data dengan cara mengekstraksi informasi yang ada [13].

6. Support Vector Machine (SVM)

Support Vector Machine (SVM) yaitu sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah fitur yang berdimensi tinggi dan dilatih dengan menggunakan algoritma pembelajaran yang didasarkan pada teori optimasi. *Support Vector Machine (SVM)* diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian dari beberapa konsep unggulan pada bidang *pattern recognition*. *Support Vector Machine (SVM)* adalah pendekatan klasifikasi yang menggunakan garis untuk menetapkan batas antara dua kelas. *Support Vector Machine (SVM)* berupaya mengidentifikasi fungsi pemisah (*hyperplane*) optimal antara dua kelas. Fungsi pemisah dikatakan optimal jika menghasilkan nilai margin tertinggi antara dua vektor dari kelas yang berbeda sambil tetap berada di tengah-tengah kedua vektor tersebut. Algoritma *Support Vector Machine (SVM)* terkenal karena kemampuannya untuk mengklasifikasikan dan menghasilkan solusi optimal. *Support Vector Machine (SVM)* adalah model pembelajaran mesin berbasis kernel yang dapat diterapkan pada klasifikasi dan regresi. Bahkan dengan data pelatihan minimal, *Support Vector Machine (SVM)* berkinerja baik dalam generalisasi, terutama untuk data linier. Namun, jika data tidak linier, *Support Vector Machine (SVM)* membutuhkan fungsi kernel untuk memprosesnya.

Support Vector Machine (SVM) bekerja dengan mengubah ruang input asli menjadi ruang fitur berdimensi tinggi, yang memaksimalkan pemisahan antar kelas. Fungsi kernel digunakan untuk mentransformasi data ke

ruang dimensi yang lebih tinggi agar data dapat dipisahkan secara linear. *Support Vector Machine (SVM)* juga sangat efektif untuk data berdimensi tinggi karena hanya menggunakan data yang relevan untuk membentuk model klasifikasi. Meskipun demikian, belum ada kesimpulan pasti tentang jenis kernel yang paling efektif untuk setiap aplikasi. *Support Vector Machine* memiliki empat kernel yang umum digunakan, yaitu [14] :

1. Linear Kernel
 $K = (X_i, X_j) = (X_i \cdot X_j)$
2. Polynomial Kernel
 $K = (X_i, X_j) = (X_i \cdot X_j + 1)^p$
3. RBF Kernel
 $K = (X_i, X_j) = e^{-\gamma(X_i, X_j)^2}$
4. Sigmoid Kernel
 $K = (X_i, X_j) = \text{tanh}(\eta X_i \cdot X_j + v)$

7. Decision Tree Classification

Decision Tree Classification adalah cara untuk mengklasifikasikan sesuatu menggunakan struktur pohon. Setiap bagian dari pohon disebut simpul (*node*). Simpul mewakili atribut, cabang mewakili nilai dari atribut tersebut, dan daun menunjukkan kelas. *Node* paling atas disebut *root*. Ada tiga jenis *node* dalam *Decision Tree Classification*, yaitu :

1. Simpul Akar / *Root Node*
Simpul paling atas yang tidak memiliki input, tetapi bisa memiliki lebih dari satu output.
2. Simpul Internal / *Internal Node*
Simpul percabangan yang hanya memiliki satu input dan mempunyai output minimal dua.
3. Simpul Daun / *Leaf Node* (terminal node)
Simpul akhir dengan satu input dan tidak memiliki output.

Pohon keputusan adalah metode klasifikasi yang kuat dan populer. Metode pohon keputusan ini mengubah data yang kompleks menjadi pohon yang merepresentasikan aturan yang mudah dipahami dengan bahasa alami, dan juga dapat diubah menjadi bahasa basis data seperti SQL untuk mencari data dalam kategori tertentu. Pohon keputusan berguna untuk mengeksplorasi data dan menemukan hubungan tersembunyi antara variabel input dan variabel target.

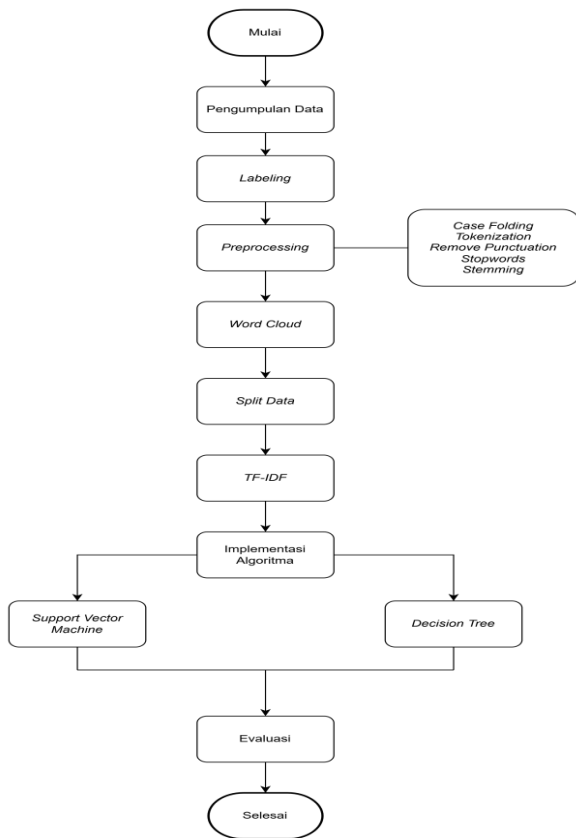
Model pohon keputusan dibentuk oleh serangkaian aturan yang membagi data yang beragam menjadi kelompok-kelompok yang lebih seragam, berdasarkan variabel targetnya. Pohon keputusan dapat dibuat secara manual atau dibentuk secara otomatis dengan menggunakan algoritma pohon keputusan untuk memodelkan kumpulan data yang belum terklasifikasi [15].

8. *TF-IDF (Term Frequency Inverse Document Frequency)*
TF-IDF (Term Frequency Inverse Document Frequency) merupakan proses pembobotan kata. Jumlah kemunculan suatu kata dalam sebuah dokumen ditentukan oleh *Term Frequency (TF)*. Sementara itu, *Inverse Document Frequency (IDF)* digunakan untuk menilai kemiripan kata dan kata kunci yang dicari. Istilah yang sering muncul memiliki sedikit pengaruh terhadap relevansi kata kunci di dalam sebuah halaman. TF-IDF dapat dihitung menggunakan persamaan di bawah ini:

$$W_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$

9. *Confusion Matrix*
Confusion matrix adalah sebuah tabel yang digunakan untuk mengevaluasi kinerja algoritma pada model yang sudah dibuat [16].

III. METODOLOGI PENELITIAN



Gambar 1. Alur Tahapan Penelitian

Berikut merupakan uraian dari alur penelitian yang akan dilakukan :

A. Pengumpulan Data

Data dikumpulkan dengan mengambil ulasan pengguna dari aplikasi ShopeePay di Google Play Store menggunakan Google Play Scraper.

B. Labeling

Setelah mengumpulkan data, langkah selanjutnya adalah memberi label pada data tersebut. Data dilabeli secara manual menjadi kategori positif atau negatif berdasarkan sentimen yang terkandung dalam ulasan.

C. Preprocessing

1. Case Folding

Case folding adalah proses mengubah data menjadi huruf kecil.

Tabel 2
 Contoh Tahap Case Folding

Sebelum	Sesudah
Sangat membantu disaat butuh banget Terimakasih shopeepay	sangat membantu disaat butuh banget terimakasih shopeepay

2. Remove Punctuation

Remove punctuation merupakan proses hapus semua karakter yang berlebihan, termasuk tanda baca, angka, dan simbol.

Tabel 3
 Contoh Tahap Remove Punctuation

Sebelum	Sesudah
apl,sangat bermanfaat,at mudah di pahami,,,,,	apl sangat bermanfaat mudah di pahami

3. Tokenization

Tokenization merupakan proses memecah teks kalimat menjadi satuan kata.

Tabel 4
 Contoh Tahap Tokenization

Sebelum	Sesudah
sangat membantu disaat butuh banget terimakasih shopeepay	["sangat", "membantu", "disaat", "butuh", "banget", "terimakasih", "shopeepay"]

4. Stopwords

Stopwords merupakan proses menghapus kata-kata yang umum digunakan tetapi tidak memiliki makna penting dalam sebuah kalimat.

Tabel 5
 Contoh Tahap Stopwords

Sebelum	Sesudah
apl sangat bermanfaat mudah di pahami	apl bermanfaat mudah pahami

5. Stemming

Stemming adalah proses mengubah kata-kata yang terhubung menjadi kata-kata dasar dengan menghilangkan imbuhan.

Tabel 6

Contoh Tahap Stemming

Sebelum	Sesudah
aku suka dengan cara mudah pelayanan sopeepay serba mudah, jaya lah terus sopeepay	aku suka dengan cara mudah layan sopeepay serba mudah jaya lah terus sopeepay

C. Word Cloud

Data yang sudah melalui preprocessing selanjutnya akan divisualisasikan dengan menggunakan word cloud. Word Cloud bertujuan untuk menampilkan kata yang sering muncul pada ulasan pengguna.

D. Split Data

Data yang telah diproses akan dipisahkan menjadi set pelatihan dan pengujian. Data pelatihan digunakan untuk melatih model mengenali pola yang ada, sedangkan data pengujian digunakan untuk mengevaluasi kinerja algoritma. Dalam penelitian ini, data pelatihan dan pengujian dibagi dengan perbandingan 80:20.

E. TF-IDF

Setelah dataset melalui tahap data split langkah selanjutnya adalah melakukan pembobotan kata pada data training. Pembobotan kata dilakukan dengan menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*) untuk menentukan pentingnya kata dalam ulasan. Kata-kata yang lebih sering muncul dalam ulasan tertentu dan jarang muncul di keseluruhan dataset mendapatkan bobot yang lebih tinggi.

F. Implementasi Algoritma

Proses implementasi ini menggunakan algoritma *Support Vector Machine* (SVM) dan *Decision Tree Classification* untuk mengukur seberapa baik performa algoritma dalam mengklasifikasikan data tersebut.

- *Support Vector Machine*

Metode *Support Vector Machine* (SVM) digunakan untuk melakukan klasifikasi dengan cara menemukan *hyperplane* terbaik yang memisahkan dua kelas data. *Hyperplane* ini dipilih agar memiliki jarak maksimum terhadap data terdekat dari masing-masing kelas, yang dikenal sebagai *support vectors*. Proses ini dimulai dengan melatih model menggunakan data latih, kemudian model yang telah terbentuk digunakan untuk mengklasifikasikan data uji. Pada penelitian ini, kernel linear dipilih karena cocok untuk teks data ulasan karena data yang digunakan dapat dipisahkan secara linear, sehingga penggunaan kernel ini lebih sederhana, cepat, dan efisien dibandingkan dengan kernel lainnya.

- *Decision Tree Classification*

Klasifikasi menggunakan metode *Decision Tree Classification* dibuat dengan memanfaatkan data latih, di mana fitur-fitur utama dari ulasan digunakan sebagai dasar pembentukan percabangan pohon.

Setiap cabang pada pohon mencerminkan keputusan berdasarkan fitur tertentu, sehingga data uji yang dimasukkan ke model akan mengikuti jalur percabangan hingga mencapai kelas akhirnya.

Metode	Kelebihan	Kekurangan
Support Vector Machine	SVM efektif untuk dataset dengan dimensi tinggi dan mampu memisahkan data yang tidak linear dengan bantuan kernel trick	SVM bisa memakan waktu komputasi yang lama terutama jika dataset sangat besar karena perhitungan margin dan kernel
Decision Tree Classification	Mudah dipahami dan diinterpretasikan, karena menghasilkan diagram pohon yang jelas (visualisasi aturan keputusan)	Rentan terhadap overfitting, apalagi jika pohon terlalu dalam atau dataset kecil, sehingga performa di data baru bisa menurun

G.

Evaluasi

Evaluasi dilakukan terhadap model dengan tujuan melihat seberapa baik kinerja algoritma klasifikasi yang telah diterapkan dengan menggunakan *Confussion Matrix*. *Confussion Matrix* adalah metode yang digunakan untuk menghitung tingkat akurasi pada algoritma klasifikasi. Tolak ukur yang digunakan untuk mengukur kinerja nilai *accuracy*, *precision*, *recall*, dan *f-1 score*

Tabel 7 Confusion Matrix

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP	FP
	FALSE	FN	TN

Keterangan:

TP (True Positive) : Jumlah data positif yang diprediksi positif

FP (False Positive) : Jumlah data positif yang salah diprediksi sebagai negatif

TN (True Negative) : Jumlah data negatif yang diprediksi negatif

FN (False Negative) : Jumlah data negatif yang salah diprediksi sebagai positif

1. *Accuracy*
Untuk mengukur seberapa akurat algoritma dapat memprediksi data dengan benar. *Accuracy* dapat dihitung dengan persamaan berikut.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. *Precision*
Untuk mengukur seberapa tepat algoritma dalam mengidentifikasi data. *Precision* dapat dihitung menggunakan persamaan berikut.

$$precision = \frac{TP}{TP + FP}$$

3. *Recall*
Untuk mengukur seberapa baik algoritma dalam mengidentifikasi data. *Recall* dapat dihitung menggunakan persamaan berikut.

$$recall = \frac{TP}{TP + FN}$$

4. *F1-Score*
Nilai yang menggabungkan *precision* dan *recall* untuk memberi gambaran keseimbangan kinerja algoritma. *F1-score* dapat dihitung menggunakan persamaan berikut

$$f1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang dikumpulkan melalui proses *scraping* dilakukan dari 31 Januari 2024 hingga 16 September 2024. Informasi yang diperoleh meliputi *username*, *score*, *at*, dan *content*, yang kemudian disimpan dalam format *xlsx*. Data yang diperoleh sebanyak 3.968 ulasan pengguna.

B. Labeling

Setelah pengumpulan data, ulasan pengguna akan diberi label positif dan negatif sesuai dengan isi ulasan pengguna. Berikut distribusi label pada tabel berikut:

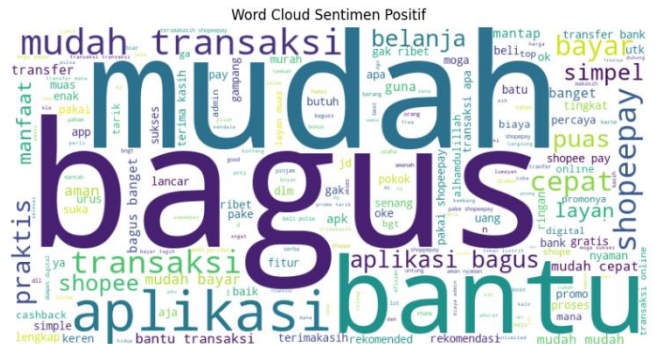
Tabel 8
Distribusi Label

Label	Jumlah
Positif	2.246
Negatif	1.722

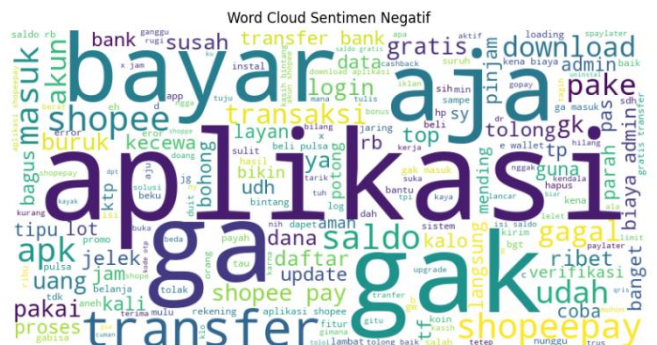
C. Wordcloud

Setelah melalui tahap *preprocessing* data akan divisualisasikan dengan *wordcloud*. *Wordcloud* digunakan untuk memvisualisasikan kata-kata yang paling sering muncul dalam teks sehingga memudahkan identifikasi kata kunci atau tema utama. Visualisasi ini membantu memahami pola umum dalam data teks, termasuk sentimen atau topik yang dominan, secara cepat dan intuitif. Selain itu, *word cloud*

membuat penyajian data teks menjadi lebih menarik dan mudah dipahami. Berikut gambar *wordcloud* label positif dan negatif:



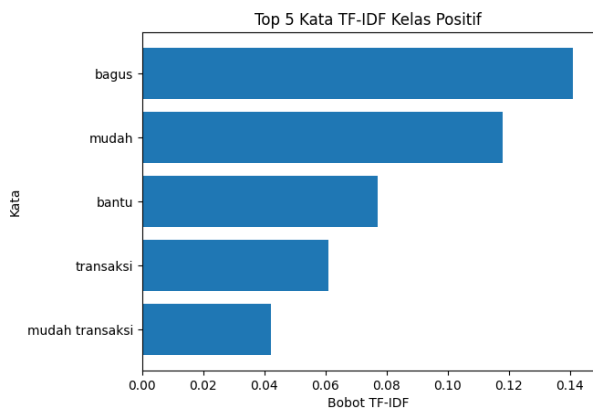
Gambar 2
Wordcloud label positif



Gambar 3
Wordcloud label negatif

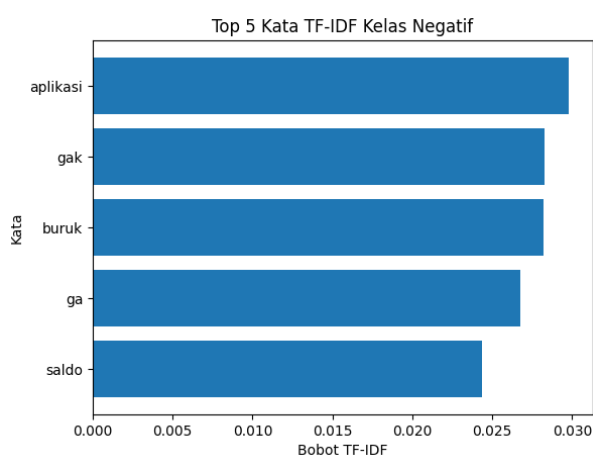
D. Hasil Analisis TF-IDF

Hasil analisis TF-IDF pada data training dilakukan untuk mengidentifikasi kata-kata yang paling relevan bagi masing-masing kelas sentimen. Pada kelas positif, kata-kata dengan bobot tertinggi menunjukkan pengalaman menyenangkan atau kepuasan pengguna, sedangkan pada kelas negatif, kata-kata teratas mencerminkan ketidakpuasan atau masalah yang dialami. Hasil ini membantu model lebih fokus pada kata-kata penting untuk membedakan sentimen positif dan negatif. Berikut gambar 5 kata pada ekstraksi TF-IDF dengan bobot tertinggi.



Gambar 4

5 Kata Teratas TF-IDF Label Positif

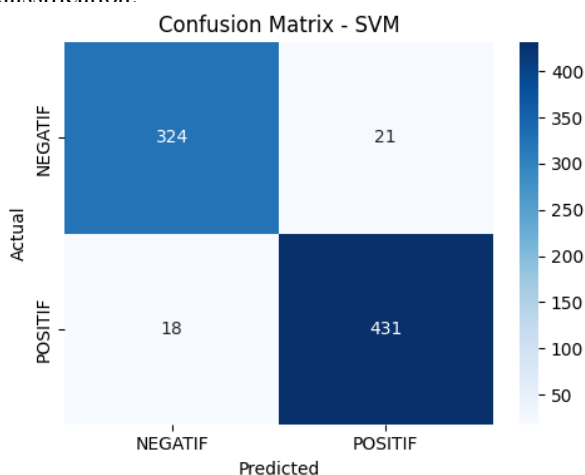


Gambar 5

5 Kata Teratas TF-IDF Label Negatif

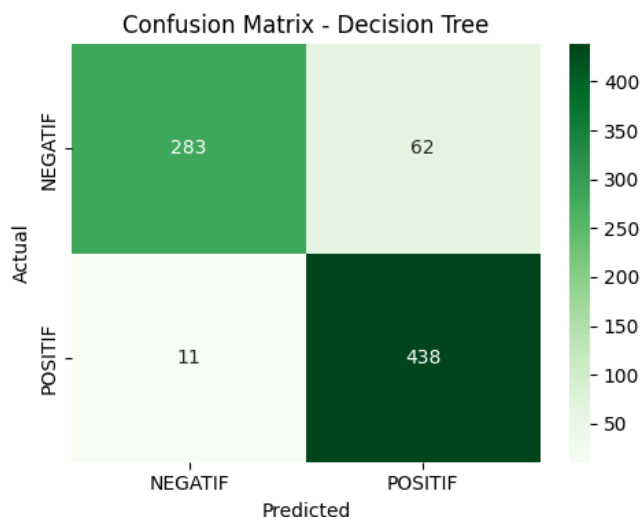
E. Hasil Klasifikasi dan Evaluasi

Klasifikasi pada Penelitian ini menggunakan metode Support Vector Machine (SVM) dan Decision Tree Classification.



Gambar 6 Confusion Matrix Support Vector Machine (SVM)

Berdasarkan confusion matrix di atas, *Support Vector Machine (SVM)* dapat memprediksi 324 data ulasan negatif dan 431 data ulasan positif dengan benar dari 20% data uji. Sementara itu, ada 21 data ulasan negatif yang terprediksi sebagai ulasan positif, dan 18 data ulasan positif yang terprediksi sebagai ulasan negatif.



Gambar 7

Confusion Matrix Decision Tree Classification

Berdasarkan confusion matrix di atas, *Decision Tree Classification* dapat memprediksi 283 data ulasan negatif dan 438 data ulasan positif dengan benar dari 20% data uji. Sementara itu, ada 62 data ulasan negatif yang terprediksi sebagai ulasan positif, dan 11 data ulasan positif yang terprediksi sebagai ulasan negatif. Berdasarkan hasil confusion matrix sebelumnya, dapat diperoleh metrik evaluasi dari kedua model berikut:

Tabel 9
Metrik Evaluasi

Metode	Akurasi	Presisi	Recall	F1-Score
Support Vector Machine (SVM)	95%	95%	95%	95%
Decision Tree Classification	91%	92%	90%	90%

Berdasarkan tabel di atas *Support Vector Machine (SVM)* menunjukkan performa lebih konsisten dibanding *Decision Tree Classification*. *Support Vector Machine (SVM)* mencapai akurasi 95%, dengan precision, recall, dan F1-score yang seimbang pada kedua kelas (positif dan negatif). *Decision Tree Classification* memiliki akurasi 91%, namun performanya menurun pada kelas negatif karena *overfitting* pada kelas mayoritas dan sensitivitas terhadap ketidakseimbangan data, seperti ulasan negatif yang mengandung kata positif.

V. KESIMPULAN

Berdasarkan hasil evaluasi, *Support Vector Machine (SVM)* mengungguli *Decision Tree* dalam mengklasifikasikan sentimen ulasan pengguna ShopeePay. *Support Vector Machine (SVM)* mampu mempertahankan performa tinggi di kedua kelas dengan akurasi 95%, sedangkan *Decision Tree Classification* menurun performanya pada kelas minoritas dengan akurasi 91%. *Overfitting* dan ketidakseimbangan data menyebabkan *Decision Tree* kurang konsisten, meskipun model ini tetap berguna untuk analisis visual dan interpretasi aturan keputusan. Dengan demikian, *Support Vector Machine (SVM)* direkomendasikan sebagai metode utama untuk klasifikasi sentimen pada ulasan aplikasi ShopeePay.

DAFTAR PUSTAKA

- [1] D. Mertiwi *et al.*, “ANALISIS SENTIMEN APLIKASI DANA PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES DAN K- NEAREAST,” no. August 2023, pp. 1063–1072, 2024.
- [2] A. O. Permana and S. Saepudin, “Jurnal Computer Science and Information Technology (CoSciTech) Comparison of the K-Nearst Neighbor and Naive Bayes Algorithms on the Sentiment Analysis Shopee Application,” *J. Comput. Sci. Inf. Technol.*, vol. 4, no. 1, pp. 25–32, 2023.
- [3] N. Habibah, E. Budianita, M. Fikry, and I. Iskandar, “Analisis Sentimen Mengenai Penggunaan E-Wallet Pada Google Play Menggunakan Lexicon Based dan K-Nearest Neighbor,” vol. 10, no. 1, pp. 192–200, 2023, doi: 10.30865/jurikom.v10i1.5429.
- [4] I. S. Widiyanto, Y. R. Ramadhan, and M. A. Komara, “Analisis sentimen e-wallet gopay, shopeepay, dan ovo menggunakan algoritma naive bayes,” *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 12, no. 3, 2024.
- [5] M. Xanderina *et al.*, “J-ENSISTEC (Journal of Engineering and Sustainable Technology) Vol. 10|No. 02, June 2024 ANALISIS SENTIMEN ULASAN E-COMMERCE SHOPEE PADA GOOGLE PLAY STORE MENGGUNAKAN MACHINE LEARNING,” vol. 10, no. 02, pp. 990–998, 2024.
- [6] A. Saepudin, A. Faqih, and G. Dwilestari, “Perbandingan Algoritma Klasifikasi Support Vector Machine , Random Forest dan Logistic Regression Pada Ulasan Shopee,” vol. 18, no. 1, pp. 178–192.
- [7] K. A. Rokhman, Berlilana, and P. Arsi, “PERBANDINGAN METODE SUPPORT VECTOR MACHINE DAN DECISION TREE UNTUK ANALISIS SENTIMEN REVIEW KOMENTAR PADA APLIKASI TRANSPORTASI ONLINE Abstraksi Keywords : Pendahuluan Tinjauan Pustaka,” *JOISM J. Inf. Syst. Manag.*, vol. 2, no. 2, 2021.
- [8] R. T. Aldisa and P. Maulana, “Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID- 19 Dengan Perbandingan Metode Naive Bayes , Decision Tree dan SVM,” vol. 4, no. 1, pp. 106–109, 2022, doi: 10.47065/bits.v4i1.1581.
- [9] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023.
- [10] H. Sulastri and A. I. Gufroni, “PENERAPAN DATA MINING DALAM PENGELOMPOKAN PENDERITA THALASSAEMIA,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 02, pp. 299–305, 2017.
- [11] I. Kurniawan, A. L. Hananto, S. S. Hilabi, and A. Hananto, “Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter,” vol. 10, no. 1, pp. 731–740, 2023.
- [12] V. Putri, D. Waruwu, and T. Harefa, “Peningkatan Kemampuan Menceritakan Kembali Teks Ulasan dengan Model Pembelajaran Problem Based Learning,” vol. 1, no. 1, pp. 69–74, 2022.
- [13] D. Oktavia and Y. R. Ramadhan, “Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM),” vol. 4, no. 1, pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.
- [14] Y. Partogi, A. Pasaribu, and Sutrisno, “PERANCANGAN METODE DECISION TREE TERHADAP SISTEM PERPUSTAKAAN STMIK KUWERA,” *J. Sist. Inf. DAN Teknol. (S I N T E K)*, vol. 1, no. 2, pp. 20–26, 2021.
- [15] T. Meisya *et al.*, “PERBANDINGAN KERNEL SUPPORT VECTOR MACHINE (SVM) DALAM PENERAPAN ANALISIS SENTIMEN VAKSINISASI COVID-19,” vol. 4, no. 2, pp. 139–145, 2021.
- [16] Maharani and Fathoni, “Analisis Sentimen Pengguna Terhadap Faktor Penggunaan PayPal Menggunakan Metode Decision Tree,” *J. Ilm. Teknol. Inf. Asia*, vol. 18, no. 1, pp. 71–83, 2024.