

PENERAPAN ALGORITMA C4.5 PADA ULASAN APLIKASI SHOPEE DI GOOGLE PLAY STORE

Indra Fitri Andriasih¹⁾, Alexius Endy Budianto²⁾, Ainia Walidaroyani³⁾

Universitas PGRI Kanjuruhan Malang^{1,2,3)}

email : indriafitri2000@gmail.com

Abstrak

Aplikasi Shopee sebagai salah satu platform e-commerce terpopuler di Indonesia menerima banyak ulasan dari penggunanya di Google Play Store. Ulasan-ulasan ini mengandung informasi penting yang dapat dimanfaatkan untuk mengevaluasi kualitas layanan aplikasi. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pengguna menggunakan algoritma C4.5 agar dapat membantu pihak pengembang dalam memahami persepsi pengguna. Data yang digunakan diperoleh dari Google Play Store dan melalui beberapa tahapan pengolahan, yaitu preprocessing (case folding, stopword removal, stemming, normalisasi kata, dan pelabelan sentimen), transformasi data menggunakan metode TF-IDF, serta pembagian data menjadi data latih dan data uji. Algoritma C4.5 diimplementasikan menggunakan model DecisionTreeClassifier dengan kriteria entropy. Hasil penelitian menunjukkan bahwa model klasifikasi mampu menghasilkan akurasi sebesar 83,25% pada data uji. Klasifikasi terhadap sentimen positif menunjukkan performa yang tinggi, sedangkan sentimen negatif dan netral masih kurang optimal akibat ketidakseimbangan data antar kelas. Dengan demikian, algoritma C4.5 cukup efektif untuk mengklasifikasikan sentimen ulasan pengguna, terutama dalam mengenali sentimen positif. Hasil ini dapat menjadi bahan evaluasi bagi pengembang aplikasi Shopee dalam meningkatkan layanan mereka berdasarkan masukan dari pengguna.

Kata Kunci : *Shopee*; klasifikasi sentiment; algoritma C4.5; *TF-IDF*; *Decision Tree*; *Google Play Store*.

Abstract

Shopee, as one of the most popular e-commerce platforms in Indonesia, receives numerous user reviews on the Google Play Store. These reviews contain valuable information that can be leveraged to evaluate the quality of the application's services. This study aims to classify user sentiment using the C4.5 algorithm to assist developers in better understanding user perceptions. The data were collected from the Google Play Store and processed through several stages, including preprocessing (case folding, stopword removal, stemming, word normalization, and sentiment labeling), data transformation using the TF-IDF method, and splitting the dataset into training and testing sets. The C4.5 algorithm was implemented using the DecisionTreeClassifier model with entropy as the criterion. The results indicate that the classification model achieved an accuracy of 83.25% on the test data. The model demonstrated strong performance in classifying positive sentiment, while the classification of negative and neutral sentiments was less optimal due to class imbalance. Therefore, the C4.5 algorithm proves to be effective in classifying user review sentiment, particularly in identifying positive sentiment. These findings can serve as valuable input for Shopee's developers to improve their services based on user feedback.

Keywords : *Shopee*; sentiment classification; C4.5 algorithm; *TF-IDF*; *Decision Tree*; *Google Play Store*.

1. PENDAHULUAN

Persaingan bisnis di Indonesia saat ini berkembang dengan sangat pesat, tercermin dari ramainya kemunculan usaha baru dan meningkatnya intensitas kompetisi di pasar. Perkembangan teknologi digital dan akses internet yang semakin luas telah memungkinkan banyak pelaku usaha beralih dari operasional manual ke platform online [1]. Internet saat ini bukan hanya menjadi media komunikasi, melainkan juga sarana untuk melakukan transaksi jual

beli, yang secara signifikan meningkatkan efektivitas dan efisiensi dalam menjalankan aktivitas bisnis [2]

Sejalan dengan data resmi dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII, 2024), jumlah pengguna internet di Tanah Air terus meningkat secara signifikan mencapai lebih dari 200 juta orang pada tahun 2023, yang berarti lebih dari 77% penduduk Indonesia aktif menggunakan internet [3]. Para pengguna yang berusia 7 hingga 64 tahun umumnya mengakses internet melalui berbagai perangkat elektronik, seperti smartphone, komputer, atau tablet (Muticara, A. C. 2023).

Meningkatnya jumlah pengguna internet di Indonesia, disertai dengan harga layanan internet yang semakin terjangkau, telah mendorong masyarakat untuk memanfaatkan teknologi digital dalam mendukung berbagai aspek kehidupan sehari-hari. Fenomena ini menunjukkan bahwa internet kini tidak hanya digunakan untuk keperluan komunikasi semata, tetapi juga telah menjadi sarana utama dalam aktivitas ekonomi dan sosial masyarakat modern [5]. Di sisi lain, pesatnya pertumbuhan e-commerce turut dipengaruhi oleh beragamnya pilihan produk serta layanan yang ditawarkan secara inovatif dan menarik. Keberadaan e-commerce memberikan kemudahan bagi konsumen dalam melakukan transaksi jual beli secara daring. Dengan hanya mengakses platform belanja online, masyarakat dapat mencari dan memperoleh barang yang dibutuhkan tanpa harus menghabiskan waktu dan tenaga untuk mengunjungi toko fisik (Sifa et al., 2024).

Perkembangan e-commerce telah memberi dampak signifikan bagi para pelaku bisnis dengan memperluas jangkauan pasar mereka secara drastis. Informasi produk kini tersebar lebih cepat dan meluas, mengubah pola konsumsi masyarakat hingga menjadi bagian dari gaya hidup modern. Hadirnya berbagai platform belanja online menawarkan kemudahan transaksi, sehingga konsumen tidak perlu lagi menghabiskan banyak waktu dan tenaga untuk mencari produk. Fenomena ini menjadikan persuasi digital sebagai faktor utama dalam menarik minat beli melihat perilaku konsumen yang semakin sering terdorong oleh iklan, diskon flash sale, dan fitur gratis ongkir [7], [8]

Meskipun e-commerce memberikan banyak keuntungan, dampaknya tidak selalu positif bagi konsumen. Salah satu masalah yang sering terjadi adalah ketidaksesuaian produk yang diterima dengan ekspektasi atau deskripsi di situs belanja online. Kondisi ini menyebabkan ketidakpuasan konsumen, dan semakin rendah kepercayaan, semakin kecil kemungkinan konsumen melakukan pembelian kembali [9]

Platform Shopee menawarkan keunggulan berupa pemeliharaan kualitas produk yang ditawarkan oleh para produsen kepada konsumen. Shopee sebagai marketplace menyediakan berbagai jenis produk seperti elektronik, pakaian, kebutuhan rumah tangga, dan lainnya, sehingga mempermudah konsumen dalam memilih barang sesuai kebutuhan mereka [10]. Berdasarkan studi yang mendukung, keandalan kualitas produk ini penting karena sangat memengaruhi keputusan pembelian ulang konsumen dan persepsi mereka terhadap situs belanja online tersebut. Meskipun e-commerce memberikan banyak manfaat, fitur ulasan produk juga dapat membawa risiko tertentu bagi konsumen dan platform. Fitur ini memungkinkan pelanggan memberikan feedback terkait produk yang mereka beli, yang mencerminkan sejauh mana ekspektasi produk terpenuhi. Ulasan yang negatif atau tidak sesuai harapan konsumen dapat menurunkan kepercayaan terhadap toko online dan secara langsung memengaruhi keputusan pembelian ulang oleh calon konsumen atau pembeli sebelumnya (Pramudita, 2024).

2. ALGORITMA

Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 dan digunakan untuk membangun model pohon keputusan (decision tree) dalam data mining dan machine learning. Algoritma ini memilih atribut terbaik untuk membagi data berdasarkan entropy dan gain ratio, serta mendukung fitur seperti penanganan atribut kontinu, nilai yang hilang, dan pemangkasan pohon (pruning) guna meningkatkan akurasi serta menghindari overfitting (Bachtiar & Mahradianur, 2023; Lu & Chen, 2016).

- 1) Struktur Pohon Keputusan
 - a) Setiap node: merepresentasikan atribut.
 - b) Cabang: nilai dari atribut tersebut.
 - c) Daun (leaf): hasil atau kelas dari prediksi.
- 2) Proses Kerja Algoritma C4.5
 - a) Pemilihan Atribut: menggunakan metrik Gain Ratio untuk menentukan atribut terbaik.
 - b) Pembagian Data: dataset dibagi berdasarkan nilai atribut yang dipilih.
 - c) Rekursi: proses pemilihan atribut dan pembagian data diulang sampai pohon terbentuk sempurna.
 - d) Pruning: pemangkasan cabang yang tidak signifikan untuk mencegah overfitting.
 - e) Penanganan Data Hilang dan Atribut Kontinu: C4.5 dapat mengabaikan atau mengatur nilai yang hilang dan mengubah atribut numerik menjadi kategorikal.
 - f) Leaf Decision: keputusan akhir ditentukan berdasarkan mayoritas kelas pada leaf node.
- 3) Komponen Matematis C4.5
 - a) Entropy
Mengukur tingkat ketidakpastian data dalam satu himpunan.

$$\text{Entropy}(S) = - \sum_{i=1}^n p_i \cdot \log_2(p_i)$$

Keterangan:

S : Himpunan data

p_i : Proporsi kelas ke- i dalam S

- b) Information Gain

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot \text{Entropy}(S_i)$$

Keterangan:

A : Atribut

S_i : Subset data setelah dibagi berdasarkan A

- c) Split Info

$$\text{SplitInfo}(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \left(\frac{|S_i|}{|S|} \right)$$

Keterangan:

S : Himpunan kasus

A : Atribut

Si : Jumlah sampel untuk atribut A

d) Gain Ratio

$$\text{GainRatio}(S, A) = \frac{\text{Gain}(S, A)}{\text{SplitInfo}(S, A)}$$

Keterangan:

S: Himpunan kasus

A: Atribut

Gain (S,A): Information Gain pada atribut A

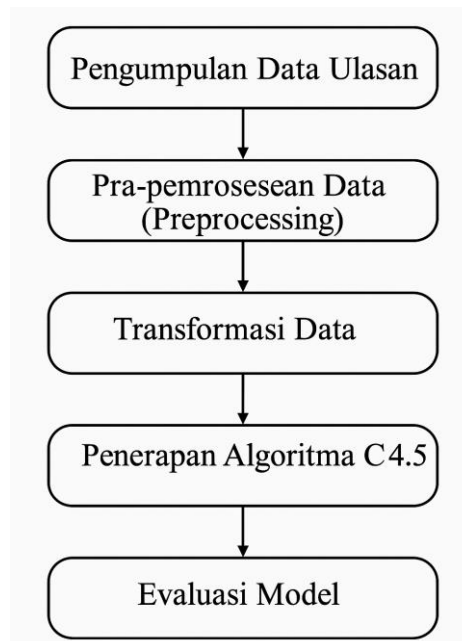
SplitInfo(S,A): Split information pada atribut A

3. METODOLOGI

Penelitian ini merupakan studi kuantitatif dengan pendekatan eksperimen, yang bertujuan untuk menguji performa algoritma klasifikasi C4.5 dalam mengelompokkan sentimen pada ulasan pengguna aplikasi Shopee. Penelitian berada dalam lingkup data mining dengan fokus pada klasifikasi sentimen ke dalam tiga kelas, yaitu positif, negatif, dan netral. Metode yang digunakan adalah supervised learning, di mana data latih diperoleh melalui proses pelabelan otomatis berdasarkan skor ulasan pengguna. Data yang telah dilabeli digunakan untuk melatih model agar mampu mengenali pola klasifikasi dan melakukan prediksi pada data baru.

Tahapan Penelitian

Penelitian ini mengikuti metode ilmiah dengan tahapan sistematis, dimulai dari pengumpulan data, preprocessing, pembangunan model, hingga evaluasi klasifikasi. Pendekatan yang digunakan memungkinkan pelaksanaan setiap langkah secara logis dan terukur, sehingga hasilnya dapat diuji secara empiris dan direplikasi di masa mendatang. Tahapan penelitian disusun dalam struktur yang runtut:



Gambar 1. Tahapan Penelitian

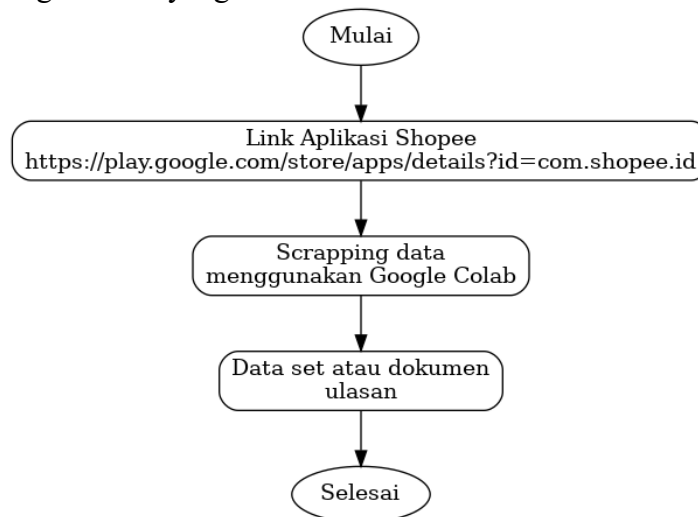
Adapun tahapan-tahapan dalam penelitian ini dapat dijelaskan sebagai berikut:

- 1) Pengumpulan Data Ulasan: Mengambil data ulasan aplikasi Shopee dari Google Play

- Store menggunakan teknik scraping.
- 2) Pra-pemrosesan Data (Preprocessing): Melakukan pembersihan dan normalisasi teks seperti case folding, word normalization, stopword removal, stemming, dan label sentimen.
 - 3) Pembagian data pelatihan dan pengujian menggunakan rasio 80:20.
 - 4) Transformasi Data: Mengubah data menjadi format yang dapat dianalisis, termasuk encoding atribut.
 - 5) Penerapan Algoritma C4.5: Membangun model pohon keputusan berdasarkan data yang telah diproses.
 - 6) Evaluasi Model: Menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score.

Sumber dan Teknik Pengambilan Data

Penelitian ini memanfaatkan data ulasan pengguna aplikasi Shopee dari Google Play Store, yang dikumpulkan melalui teknik scraping menggunakan Google Colab dan library Python. Hanya ulasan berbahasa Indonesia yang diambil untuk menjaga konsistensi analisis. Data yang telah dikumpulkan disimpan dalam format CSV dan diolah di Google Colab untuk keperluan klasifikasi dan pelabelan sentimen (positif, negatif, netral). Proses pengumpulan data dijelaskan melalui diagram alur yang disertakan:



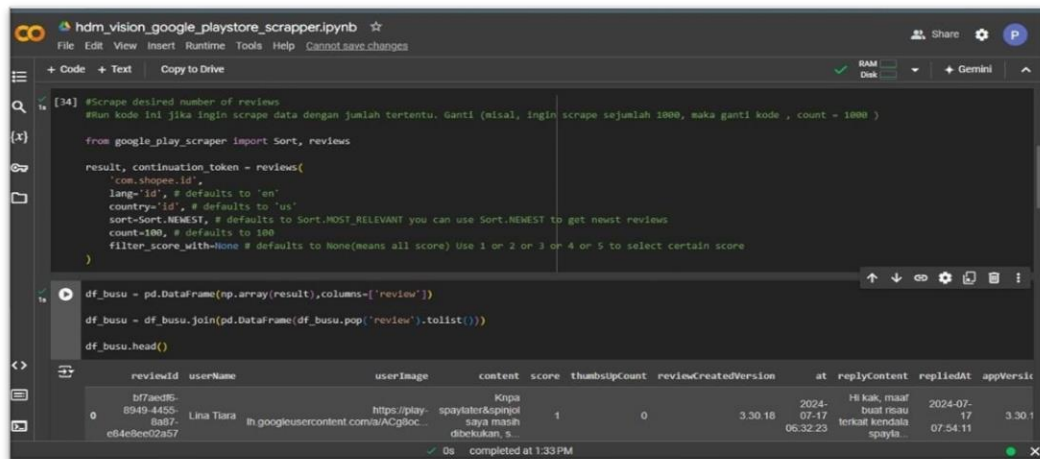
Gambar 2. Alur Scrapping

Diagram alur tersebut menunjukkan bahwa proses pengumpulan data dilakukan secara terstruktur mulai dari inisialisasi scraping hingga penyimpanan data dalam format yang siap diolah. Data yang telah terkumpul kemudian menjadi dasar untuk tahapan preprocessing dan analisis lebih lanjut.

4. HASIL DAN PEMBAHASAN

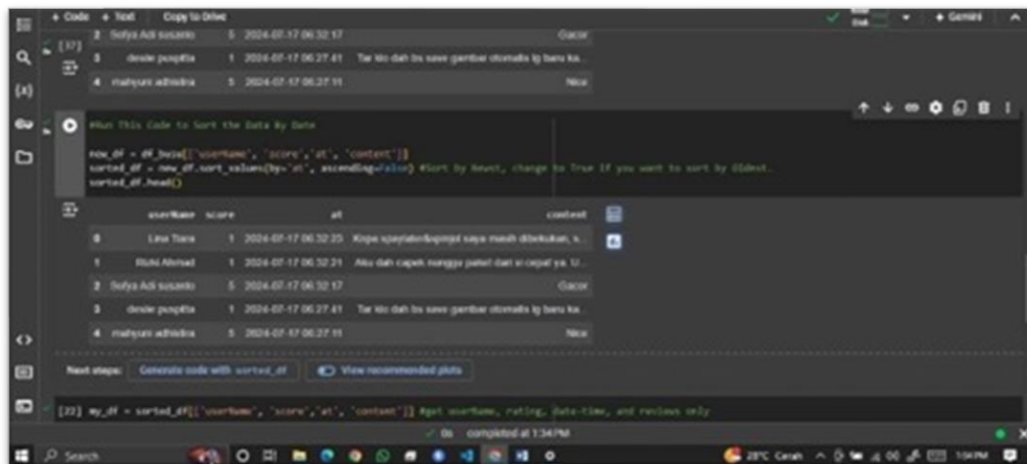
Proses Pengambilan Data

Langkah pertama yang dilakukan adalah membuka google collab, lalu masukkan link aplikasi *Shopee* ke dalam google collab.



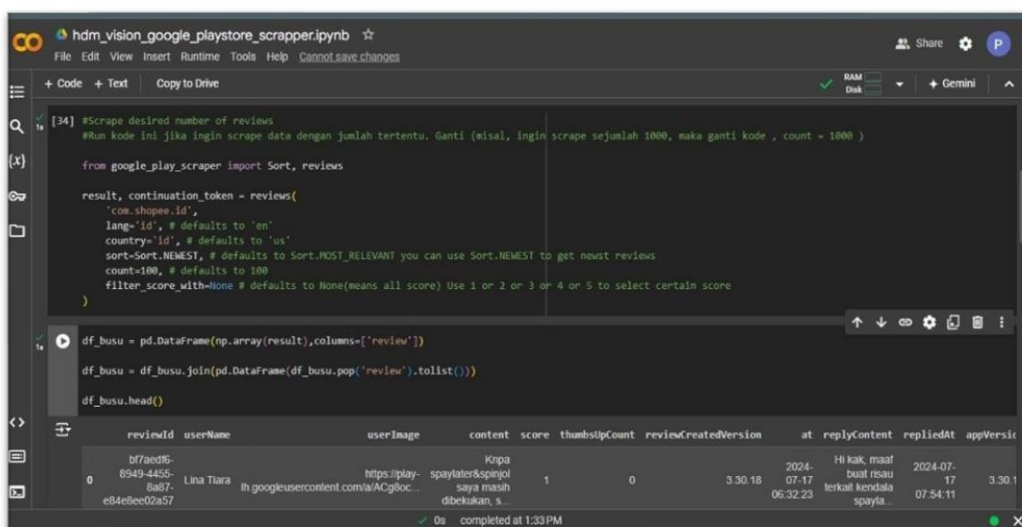
Gambar 3. Alur Awal Pengambilan Data

Jika tampilannya sudah muncul, maka akan muncul data ulasan aplikasi *Shopee* yang mau di ambil.



Gambar 4. Data Ulasan

Jika tampilannya sudah muncul, maka akan muncul data ulasan aplikasi *Shopee* yang mau di ambil.



Gambar 5. Tampilan Jumlah Data Dan Hasil Scrapping

Data ulasan pengguna Shopee dikumpulkan melalui proses scraping di Google Colab menggunakan pustaka Python seperti requests, pandas, dan BeautifulSoup4, lalu disimpan dalam format .csv untuk pengolahan lebih lanjut. Penelitian ini memfokuskan pada tiga atribut utama: score, at, dan content, yang dianggap paling relevan untuk klasifikasi sentimen. Setelah proses impor data ke Google Colab menggunakan pustaka pandas, struktur data ditampilkan dalam bentuk tabel untuk dianalisis lebih lanjut pada tahap preprocessing.

Tabel 1. Data Hasil Scrapping

No	Username	Score	At	Content
0	Mariani Ketut	1	2025-07-20 04:15:47	gak bisa di pencett
1	Warni Warni	5	2025-07-20 04:15:38	bagus sangat adil
2	Heni Hendrawati	5	2025-07-20 04:14:40	good
3	rama rahmatul h	1	2025-07-20 04:14:15	pelayan costumer service kurang baik
4	Icah setiawati	5	2025-07-20 04:14:00	😊
...	...	...	...	...
995	Aldo	2	2025-07-19 10:20:44	walaupun aplikasinya berat, namun pesanan data...
996	Nata Suary	5	2025-07-19 10:20:21	the best
997	Anto	5	2025-07-19 10:19:56	bgus
998	Rioprata	5	2025-07-19 10:19:44	sangat membantu dalam belanja terimakasih Shop...
999	silna lutfiana	5	2025-07-19 10:19:24	olshop terbaaaaaaaik

Tabel menampilkan sebagian data ulasan pengguna Shopee yang telah di-scrape dan dimuat ke Google Colab, terdiri dari atribut score, at, dan content. Data ini menjadi dasar untuk proses preprocessing, transformasi data, dan pembentukan model klasifikasi sentimen menggunakan algoritma C4.5.

Preprocessing Data

Pra-pemrosesan dilakukan untuk membersihkan dan menyiapkan data teks sebelum digunakan dalam model klasifikasi, yang mencakup beberapa langkah utama.

1) Case Folding

Case folding merupakan tahap awal preprocessing yang mengubah semua huruf menjadi huruf kecil untuk menghindari perbedaan makna akibat kapitalisasi. Proses ini dilakukan menggunakan Python di Google Colab.

Tabel 2. Hasil Case Folding

content	content_casefolded
gak bisa di pencett	gak bisa di pencett
bagus sangat adil	bagus sangat adil
good	good
pelayan costumer service kurang baik	pelayan costumer service kurang baik
😊	😊
pesanan sesuai	pesanan sesuai
pelayanan baik, kualitas barangnya bagus	pelayanan baik, kualitas barangnya bagus
mksh shoope	mksh shoope
SANGAT MEMBANTU BUAT SEMUA KEPERLUAN APAPUN.....	sangat membantu buat semua keperluan apapun.....
update 20 juli 25 aplikasi begitu di buka,langsung force close	update 20 juli 25 aplikasi begitu di buka,langsung force close

Tabel diatas menunjukkan bahwa huruf kapital pada ulasan berhasil diubah menjadi huruf kecil, sementara karakter non-alfabet masih dipertahankan. Hasilnya disimpan dalam file dataset_casefolding.csv untuk tahapan preprocessing berikutnya.

2) Stopword Removal

Setelah case folding, tahap preprocessing dilanjutkan dengan stopwords removal, yaitu penghapusan kata-kata umum seperti "dan", "yang", dan "untuk" agar teks lebih

ringkas dan algoritma lebih efisien. Proses ini dilakukan menggunakan pustaka NLTK di Python.

Tabel 3. Hasil Stopword Removal

content_casfolded	content_stopword_removed
gak bisa di pencett	gak pencett
bagus sangat adil	bagus adil
good	good
pelayan costumer service kurang baik	pelayan costumer service
😊	😊
pesanan sesuai	pesanan sesuai
pelayanan baik, kualitas barangnya	pelayanan baik, kualitas barangnya bagus
bagus	
mksh shoope	mksh shoope
sangat membantu buat semua keperluan	membantu keperluan apapun... trims shoope...
apapun.....	
update 20 juli 25 aplikasi begitu di	update 20 juli 25 aplikasi buka,langsung masuk...
buka,langsung force close	

Dari hasil tersebut, terlihat bahwa kata-kata umum yang tidak terlalu memberikan informasi penting telah berhasil dihapus dari teks ulasan. Hasil akhir disimpan ke dalam file `dataset_stopword_removal.csv` untuk proses lanjutan seperti stemming dan klasifikasi.

3) Stemming

Setelah stopwords removal, tahap selanjutnya adalah stemming, yaitu mengubah kata menjadi bentuk dasarnya untuk menyamakan variasi kata dan meningkatkan akurasi klasifikasi. Proses ini dilakukan menggunakan pustaka Sastrawi yang khusus untuk Bahasa Indonesia.

Tabel 4. Hasil Stemming

content_stopword_removed	content_stemmed
gak pencett	gak pencett
bagus adil	bagus adil
good	good
pelayan costumer service	layan costumer service
😊	
pesanan sesuai	pesan sesuai
pelayanan baik, kualitas barangnya	layan baik kualitas barang bagus
bagus	
mksh shoope	mksh shoope
membantu keperluan apapun... trims	bantu perlu apa trims shoope
shoope...	
update 20 juli 25 aplikasi	update 20 juli 25 aplikasi buka langsung masuk
buka,langsung masuk...	

Hasil stemming menunjukkan kata-kata berhasil diubah ke bentuk dasar, seperti "pelayanan" menjadi "layan" dan "membantu" menjadi "bantu". Dataset hasilnya disimpan dalam file `dataset_stemming.csv` untuk proses transformasi data selanjutnya.

4) Word Normalization

Setelah stemming, dilakukan word normalization untuk menyamakan kata tidak baku atau slang menjadi kata formal, seperti "mksh" menjadi "terima kasih". Proses ini penting agar algoritma dapat memahami makna kata dengan lebih akurat, dan dilakukan menggunakan pendekatan berbasis kamus (*dictionary-based*).

Tabel 5. Hasil Word Normalization

content_stemmed	content_normalized
gak pencettt	gak pencettt
bagus adil	bagus adil
good	good
layan costumer service	layan costumer service
pesan sesuai	pesan sesuai
layan baik kualitas barang bagus	layan baik kualitas barang bagus
mksh shoope	terima kasih shoope
bantu perlu apa trims shoope	bantu perlu apa terima kasih shoope
update 20 juli 25 aplikasi buka	update 20 juli 25 aplikasi buka langsung masuk
langsung masuk	

Hasil normalisasi berhasil menyatukan kata informal ke bentuk standar, seperti “mksh” dan “trims” menjadi “terima kasih”, sehingga meningkatkan akurasi analisis sentimen. Dataset hasilnya disimpan dalam file dataset_normalized.csv untuk digunakan pada tahap klasifikasi dengan algoritma C4.5.

5) Labeling Sentimen

Langkah akhir preprocessing adalah labeling, yaitu mengklasifikasikan ulasan berdasarkan skor: skor 1–2 sebagai sentimen negatif, skor 3 netral, dan skor 4–5 positif. Sebelum labeling, data dibersihkan dari konten kosong dan indeks disusun ulang.

Tabel 6. Hasil Labeling Sentimen

userName	score	at	content	sentiment
Mariani Ketut	1	2025-07-20 04:15:47	gak bisa di pencettt	Negatif
Warni Warni	5	2025-07-20 04:15:38	bagus sangat adil	Positif
Heni Hendrawati	5	2025-07-20 04:14:40	good	Positif
Rama Rahmatul H	1	2025-07-20 04:14:15	layan costumer service	Negatif
Edah Jubaedah	5	2025-07-20 04:13:40	pesan sesuai	Positif
Kandaya Mana	5	2025-07-20 04:12:52	layan baik kualitas barang bagus	Positif
Iwan Gunawan	5	2025-07-20 04:12:10	terima kasih shoope	Positif
Bagoez HD	5	2025-07-20 04:12:03	bantu perlu apa terima kasih	Positif
Roni Jodicraft	1	2025-07-20 04:12:02	update 20 juli 25 aplikasi buka	Negatif
Agus Sulaiman	4	2025-07-20 04:11:34	besar nominal koin biar kaya	Positif

Setelah menyelesaikan seluruh tahapan preprocessing, data ulasan Shopee kini siap digunakan untuk proses selanjutnya. Data bersih dan terstruktur ini disimpan dalam file dataset_preprocessed.csv dan menjadi dasar penting dalam transformasi data serta penerapan algoritma C4.5.

Pembagian Data (Train-Test Split)

Setelah preprocessing dan labeling selesai, data dibagi menjadi data latih (80%) dan data uji (20%) menggunakan fungsi train_test_split dari Scikit-learn. Tujuan pemisahan ini adalah untuk melatih dan mengevaluasi algoritma C4.5, dengan pembagian data yang acak namun konsisten melalui pengaturan random_state. Adapun hasil dari proses pemisahan data adalah sebagai berikut:

Tabel 7. Hasil Pemisahan Data Latih dan Uji

No	Keterangan	Jumlah Data
1	Jumlah total data setelah preprocessing	953 data
2	Jumlah data latih (training)	762 data
3	Jumlah data uji (testing)	191 data

File hasil pemisahan ini disimpan dalam dua file terpisah:

- 1) train_data.csv untuk data latih

- 2) test_data.csv untuk data uji

Transformasi Data

Setelah preprocessing dan pembagian data, dilakukan transformasi data menggunakan metode TF-IDF untuk mengubah teks ulasan Shopee menjadi representasi numerik. TF-IDF menghitung pentingnya suatu kata dalam sebuah dokumen relatif terhadap seluruh dokumen. Tahapan ini meliputi pembagian data, penerapan TF-IDF, pembentukan matriks fitur, dan verifikasi dimensi hasil transformasi.

Hasil dari proses transformasi data ditunjukkan pada output berikut:

- 1) Bentuk data latih setelah TF-IDF: (762, 1212)
- 2) Bentuk data uji setelah TF-IDF: (191, 1212)

Dari hasil di atas diketahui bahwa:

- 1) Data latih terdiri dari 762 data, sedangkan data uji sebanyak 191 data.
- 2) Terdapat 1.212 fitur unik yang dihasilkan dari seluruh korpus dokumen setelah dilakukan proses tokenisasi dan transformasi TF-IDF.

Oleh karena itu, dilakukan transformasi data menggunakan metode TF-IDF. Hasil transformasi tersebut ditampilkan pada tabel berikut:

Tabel 8. Cuplikan Hasil Transformasi TF-IDF (5 Data Pertama)

10	100	12	12gb	15	...	ya	yahh	yg	zonk
0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.47	0.0
0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00	0.0
0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00	0.0
0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00	0.0
0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00	0.0

Tabel di atas hanya menampilkan sebagian kecil fitur untuk keperluan visualisasi. Seluruh data direpresentasikan dalam bentuk matriks berukuran 762×1212 untuk data latih dan 191×1212 untuk data uji.

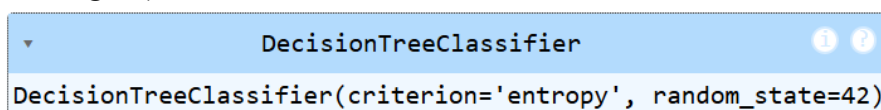
Dengan demikian, data teks telah berhasil diubah menjadi representasi numerik dan siap untuk digunakan dalam proses klasifikasi pada tahap selanjutnya.

Penerapan Algoritma C4.5

Setelah data ulasan Shopee ditransformasikan dengan metode TF-IDF, proses klasifikasi dilakukan menggunakan algoritma C4.5 yang diimplementasikan melalui library DecisionTreeClassifier dari scikit-learn. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi performa klasifikasi.

- 1) Pelatihan Model

Tahap pelatihan model merupakan proses inti dalam membangun model klasifikasi menggunakan algoritma C4.5. Data latih hasil transformasi TF-IDF digunakan untuk membentuk pohon keputusan berdasarkan nilai entropi dan information gain. Implementasi dilakukan menggunakan pustaka DecisionTreeClassifier dari scikit-learn dengan parameter criterion='entropy' untuk menyesuaikan prinsip kerja algoritma C4.5. Model dilatih menggunakan data latih X_train_tfidf sebagai representasi numerik hasil transformasi dari teks ulasan, dan y_train sebagai label sentimen yang telah dilabeli sebelumnya (Positif, Netral, atau Negatif).



Gambar 6. Pelatihan Model

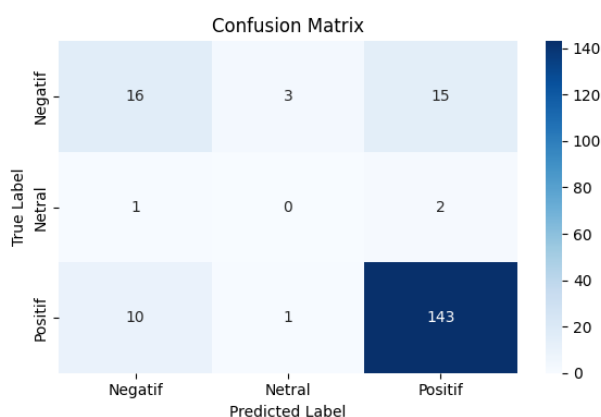
Tahap pelatihan model merupakan proses inti dalam membangun model klasifikasi menggunakan algoritma C4.5. Data latih hasil transformasi TF-IDF digunakan untuk membentuk pohon keputusan berdasarkan nilai entropi dan information gain. Implementasi dilakukan menggunakan pustaka DecisionTreeClassifier dari scikit-learn dengan parameter `criterion='entropy'` untuk menyesuaikan prinsip kerja algoritma C4.5.

2) Pengujian Model

Pengujian model bertujuan untuk mengevaluasi performa model yang telah dibentuk pada tahap pelatihan. Data uji digunakan sebagai input model, dan hasil prediksi dibandingkan dengan label sebenarnya untuk menilai akurasi serta kemampuan klasifikasi model. Hasil pengujian menunjukkan seberapa baik model dalam mengklasifikasikan data baru. Evaluasi dilakukan menggunakan metrik klasifikasi yang umum digunakan. Berikut adalah hasil evaluasi model berdasarkan pengujian terhadap 191 data uji:

Tabel 9. Hasil Evaluasi Model Klasifikasi C4.5

Kelas Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.59	0.47	0.52	34
Netral	0.00	0.00	0.00	3
Positif	0.89	0.93	0.91	154



Gambar 7. Matriks Kebingungan (Confusion Matrix)

Model klasifikasi sentimen dengan algoritma C4.5 menghasilkan akurasi sebesar 83,25%, yang menunjukkan bahwa sekitar 83 dari 100 data uji diklasifikasikan dengan benar. Meskipun akurasi cukup tinggi, evaluasi tambahan seperti precision, recall, dan f1-score tetap diperlukan karena distribusi kelas tidak seimbang, dengan performa terbaik pada kelas positif dan lebih rendah pada kelas negatif dan netral.

Analisis dan Interpretasi Hasil

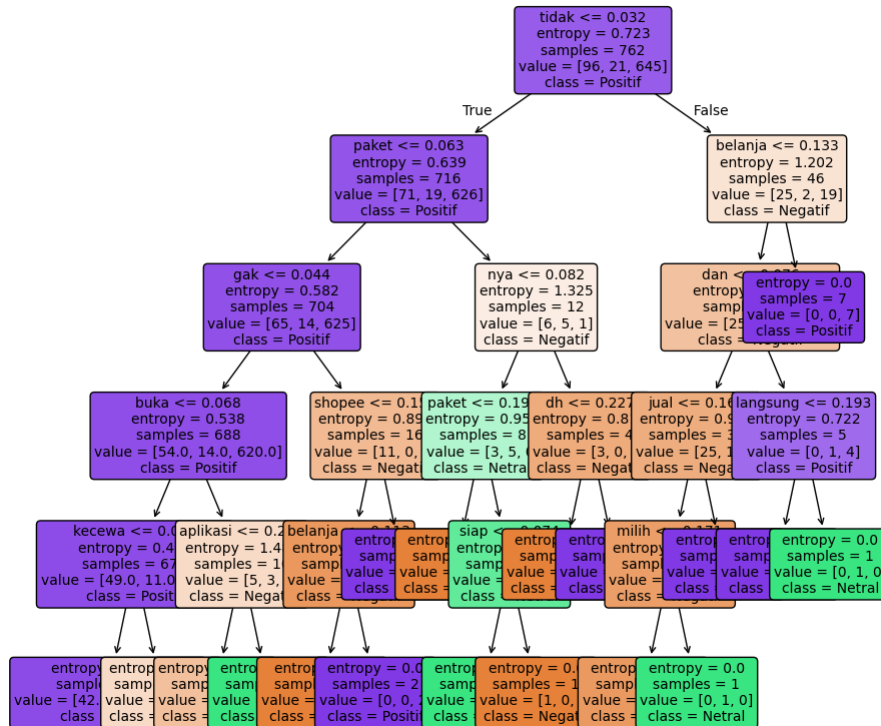
Model klasifikasi C4.5 mencapai akurasi sebesar 83,25% dari 191 data uji, menunjukkan performa yang cukup baik, terutama dalam mengenali sentimen positif (f1-score 0,91). Namun, model gagal mengklasifikasikan sentimen netral (f1-score 0,00) dan kurang optimal pada sentimen negatif (f1-score 0,52), akibat distribusi data yang tidak seimbang. Sebagian besar data netral dan negatif diklasifikasikan sebagai positif. Oleh karena itu, meskipun akurasinya tinggi, hasil harus ditafsirkan dengan hati-hati. Penelitian selanjutnya disarankan menerapkan teknik penyeimbangan data atau algoritma lain yang lebih tangguh terhadap data tidak seimbang.

Visualisasi Decision Tree (C4.5)

Visualisasi pohon keputusan yang dihasilkan bertujuan untuk memberikan gambaran mengenai bagaimana model klasifikasi bekerja, dengan menampilkan kata-kata kunci yang

Penerapan Algoritma C4.5 pada Ulasan Aplikasi Shopee di Google Play Store

Visualisasi Decision Tree (C4.5)



Visualisasi pohon keputusan menunjukkan bahwa kata-kata seperti “tidak”, “paket”, “gak”, “shopee”, dan “buka” menjadi node penting dalam proses klasifikasi sentimen. Struktur pohon menggambarkan bagaimana model memisahkan data berdasarkan nilai entropy dan distribusi kelas. Warna node menunjukkan prediksi kelas mayoritas, serta informasi jumlah data dan distribusi label. Visualisasi ini memudahkan pemahaman terhadap logika klasifikasi dan interpretasi hasil model.

Berdasarkan hasil penelitian, mayoritas ulasan pengguna aplikasi Shopee di Google Play Store tergolong dalam sentimen positif, yang menunjukkan tingkat kepuasan pengguna terhadap layanan yang diberikan. Penerapan algoritma C4.5 mampu mengklasifikasikan sentimen dengan akurasi sebesar 83,25%, terutama efektif dalam mengenali ulasan positif. Namun demikian, performa terhadap sentimen netral dan negatif masih belum optimal karena adanya ketidakseimbangan data antar kelas. Oleh karena itu, disarankan agar penelitian selanjutnya mempertimbangkan penerapan teknik penyeimbangan data seperti oversampling atau undersampling, serta membandingkan performa dengan algoritma lain seperti Naive Bayes, Random Forest, atau SVM. Penggunaan dataset yang lebih besar dan bervariasi juga akan meningkatkan generalisasi model. Selain itu, hasil klasifikasi sentimen ini dapat dimanfaatkan oleh pengembang Shopee sebagai dasar evaluasi dan peningkatan kualitas layanan serta strategi bisnis secara menyeluruh.

REFERENSI

- [1] S. N. Putri, G. M. Fahrurrozi, M. I. Fahlevi, and M. A. Septiadi, “Dinamika Persaingan Bisnis di Era Digital yang Berpengaruh Terhadap Pedagang Offline,” *Jurnal Strategi Bisnis (JSB)*, 2024.
- [2] S. Nurhayati, R. Aulia, E. S. Irawan, and D. R. W. Utami, “Peran Platform Digital Terhadap Pengembangan Umkm Di Indonesia,” 2024.
- [3] S. Mufti Prasetyo, R. Gustiawan, and F. Rizzel Albani, “BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia Analisis Pertumbuhan Pengguna Internet Di Indonesia,” vol. 2, no. 1, 2024, [Online]. Available: <https://jurnalmahasiswa.com/index.php/biikma>
- [4] Angelique Cahya Mutiara A, “Pengaruh Perceived Usefulness, Perceived Ease of Use, Perceived Cost, Customer Trust dan Customer Satisfaction terhadap Repurchase Intention pada Website Gerai,” *Universitas Multimedia Nusantara*, 2023.
- [5] L. Grace and R. Pardede, “The Influence Of Sales Promotion And Customer Experience On Purchase Intention Mediated By Customer Satisfaction,” *International Journal of Social Service and Research*, vol. 3, no. 10, pp. 2692–2700, Oct. 2023, doi: 10.46799/ijssr.v3i10.569.
- [6] Abdan Sifa, Isti Masruroh, Muhamad Abdan Zulfa, Sirfi Nur Fitriani, and Naerul Edwin Kiky Aprianto, “Transformasi Digital E-Commerce Dalam Menguasai Kosentrasi Pasar di Indonesia,” *Jurnal Ilmiah Ekonomi dan Manajemen*, vol. 2, no. 12, pp. 405–413, Dec. 2024, doi: 10.61722/jiem.v2i12.3239.
- [7] S. Silaban and N. Kusnadi, “AGRISOCIONOMICS The Impact of E-Commerce on the Performnce of Micro and Small Industries (Silaban et al The Impact Of E-Commerce Implementation On The Performance Of Micro And Small Industries In Indonesia,” *Jurnal Sosial Ekonomi dan Kebijakan Pertanian*, vol. 8, no. 1, p. 126, 2024, [Online]. Available: <http://ejournal2.undip.ac.id/index.php/agrisocionomics>
- [8] S. Ekasari, Yusi Faizathul Octavia, Y. R. Satato, Ratnawita, and Putri Ekaresty Haes, “The Impact Analysis of E-Commerce Usage on Behavior Changes of Shopee Customers in Indonesia,” *JEMSI (Jurnal Ekonomi, Manajemen, dan Akuntansi)*, vol. 9, no. 5, pp. 1781–1786, Oct. 2023, doi: 10.35870/jemsi.v9i5.1476.
- [9] N. Cahayani, S. V. Riorini, M. Maythili, and D. V. Santoso, “Consequences of Online Shopping Attributes on E-Commerce Customers,” *Jurnal Ilmiah Manajemen Kesatuan*, vol. 13, no. 1, pp. 65–72, Jan. 2025, doi: 10.37641/jimkes.v13i1.3031.
- [10] L. Maryamis, I. Iswati, A. Fitriyasari, and A. Widiyatmoko, “Pengaruh Kualitas Produk dan Harga terhadap Keputusan Pembelian Mie Gacoan Merr Surabaya,” *Global Leadership Organizational Research in Management*, vol. 2, no. 3, pp. 328–336, 2024, doi: 10.59841/glory.v2i3.1601.
- [11] A. Sovia Pramudita, “How Reviews and Ratings Influence Consumer Purchase Intentions in Indonesian E-Commerce,” *Jurnal Logistik Bisnis*, vol. 14, no. 2, 2024, [Online]. Available: <https://ejurnal.poltekpos.ac.id/index.php/logistik/>
- [12] L. Bachtiar and M. Mahradianur, “Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai,” *Jurnal Informatika*, vol. 10, no. 1, pp. 28–36, Mar. 2023, doi: 10.31294/inf.v10i1.15115.
- [13] K. Lu and M. Chen, “Research on Application of C4.5 Algorithm in Performance Analysis,” 2016.