JOISIE (Journal Of Information Systems And Informatics Engineering)
Vol. 9, No.1, Juni 2025, Hlm 207-217
Received: 2 Mei 2025 Revised: 6 Mei 2025, Accepted: 30 Juni 2025

p- ISSN: 2503-5304
e- ISSN: 2527-3116

KLASIFIKASI DAN PREDIKSI ULASAN *E-COMMERCE* MENGGUNAKAN ALGORITMA *NAÏVE BAYES*

Adjeng Putri Nuriza[1)], Tukino[2)], Elfina Novalia[3)], Bayu Priyatna[4)]
[1-4)]Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang
email: [1]si21.adjengnuriza@mhs.ubpkarawang.ac.id

Abstract

This study aims to classify Tokopedia app user reviews into four main categories: app features, services, payments, and promotions. A total of 1,500 reviews were collected through web scraping techniques from the Google Play Store and processed using preprocessing stages such as tokenization, stopword removal, and stemming. Classification was carried out using the Multinomial Naive Bayes algorithm. From 458 test data, the model produced an accuracy of 83.41%, with the highest precision value in the app features category of 0.89 and the highest recall in the payment and promotions category of 0.97. These results indicate that the Naive Bayes algorithm is effective in automatically grouping reviews with an average macro performance of 0.84 (precision), 0.83 (recall), and 0.83 (f1-score). The main contribution of this study is the application of a text classification method that can help Tokopedia identify aspects of services that are most often discussed by users, so that it can support more targeted decision making.

Keywords: *Naive Bayes, text classification, user reviews, Tokopedia, e-commerce.*

Abstract

Penelitian ini bertujuan untuk mengklasifikasikan ulasan pengguna aplikasi Tokopedia ke dalam empat kategori utama: fitur aplikasi, layanan, pembayaran, dan promosi. Sebanyak 1.500 ulasan dikumpulkan melalui teknik web scraping dari Google Play Store dan diolah menggunakan tahapan preprocessing seperti tokenization, stopword removal, dan stemming. Klasifikasi dilakukan dengan menggunakan algoritma Multinomial Naive Bayes. Dari 458 data uji, model menghasilkan akurasi sebesar 83,41%, dengan nilai precision tertinggi pada kategori fitur aplikasi sebesar 0,89 dan recall tertinggi pada kategori pembayaran dan promosi sebesar 0,97. Hasil tersebut menunjukkan bahwa algoritma Naive Bayes efektif dalam mengelompokkan ulasan secara otomatis dengan rata-rata kinerja makro sebesar 0,84 (precision), 0,83 (recall), dan 0,83 (f1-score). Kontribusi utama dari penelitian ini adalah penerapan metode klasifikasi teks yang dapat membantu Tokopedia mengidentifikasi aspek layanan yang paling sering dibicarakan oleh pengguna, sehingga dapat mendukung pengambilan keputusan yang lebih terarah.

Keywords: *Naive Bayes, Klasifikasi Teks, Ulasan Pengguna, Tokopedia, E-Commerce.*

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong perubahan signifikan dalam perilaku konsumen, khususnya dalam berbelanja daring melalui platform *e-commerce* (Rachmawati, 2024). Teknologi informasi yang berkembang pesat memungkinkan data dapat diolah dengan cepat dan akurat sehingga dapat menghasilkan informasi yang sesuai dengan kebutuhan (Agneresa et al., 2022). Aktivitas jual beli daring di Indonesia menunjukkan peningkatan, didorong oleh kehadiran *marketplace* dengan *platform* yang mudah digunakan (Irawati & Prasetyo, 2021). Selain itu, belanja daring lebih hemat waktu dibandingkan berbelanja langsung di toko *offline* (I. Kurniawan et al., 2023). Tokopedia, sebagai salah satu *platform e-commerce* terbesar di Indonesia, menyediakan berbagai layanan yang memudahkan pengguna dalam bertransaksi. Seiring dengan tingginya jumlah pengguna, aplikasi ini menerima ribuan ulasan setiap harinya yang mencerminkan pengalaman dan kepuasan pengguna terhadap layanan yang diberikan (Sulistyawati & Munawir, 2024).

Kebanyakan orang lebih sering mengungkapkan pikiran dan pendapatnya melalui ulasan media sosial dibandingkan berkomunikasi secara tatap muka (Sanjaya et al., 2024). Ulasan-ulasan ini tidak hanya berfungsi sebagai masukan, tetapi juga dapat digunakan sebagai sumber data yang kaya untuk dijelaskan lebih lanjut. Dengan mengklasifikasikan ulasan pengguna, pengembang aplikasi dapat mengidentifikasi aspek-aspek yang

https://doi.org/10.35145/joisie.v9i1.4993

perlu ditingkatkan, seperti kualitas fitur, layanan pelanggan, sistem pembayaran, dan promosi yang ditawarkan (Seran, 2024). Namun, banyaknya ulasan data yang tidak terstruktur menjadi tantangan dalam proses analisis.

Untuk mengatasi tantangan tersebut, diperlukan pendekatan analisis data berbasis kecerdasan buatan, salah satunya dengan menggunakan algoritma *Naive Bayes. Naive Bayes* adalah salah satu metode klasifikasi paling sederhana dan paling banyak digunakan dalam data mining (Mardiani et al., 2023). Algoritma ini mampu mendeskripsikan probabilitas suatu kelas berdasarkan prediksi dengan asumsi independensi antar fitur (Hananto et al., 2024). Algoritma ini dikenal efektif dalam menangani data teks dan telah banyak digunakan dalam berbagai penelitian terkait klasifikasi teks, termasuk analisis sentimen dan pemrosesan bahasa alami (Kurniawan et al., 2023). Metode *Naïve Bayes* ini merupakan pendekatan statistik yang digunakan untuk melakukan inferensi induktif dalam permasalahan klasifikasi. Salah satu keunggulan dari metode ini adalah kemampuannya untuk bekerja dengan jumlah data pelatihan *training data* yang relatif sedikit dalam menghitung estimasi parameter yang dibutuhkan pada proses klasifikasi (Apriliyani & Salim, 2022).

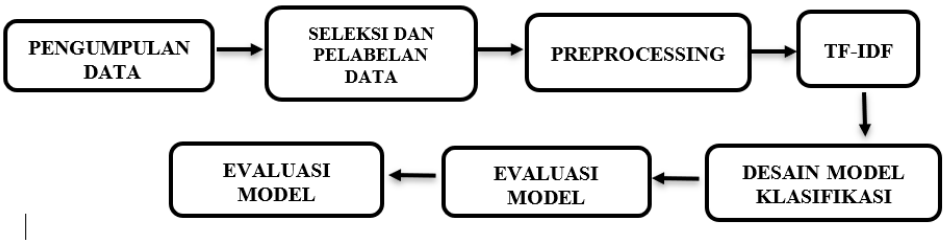
Penelitian sebelumnya yang dilakukan oleh Ernianti dan Elmo menemukan bahwa proses pengujian algoritma *Naïve Bayes* cocok untuk melakukan proses analisis sentimen pada aplikasi Amazon, dan menghasilkan akurasi rata-rata sebesar 82.15% (Hasibuan & Heriyanto, 2022). Penelitian internasional sebelumnya menerapkan algoritma *Naive Bayes* untuk menyarankan komentar YouTube berdasarkan domain tertentu. Dengan teknik penyempurnaan tertentu, metode ini mencapai akurasi 80%, menunjukkan kemampuan adaptasi yang baik untuk klasifikasi teks skala besar (Pirunthavi & Jayalath, 2021).

Namun, sebagian besar penelitian sebelumnya masih terbatas pada klasifikasi sentimen umum dan belum secara khusus meneliti klasifikasi ulasan pengguna aplikasi *e-commerc*e lokal seperti Tokopedia ke dalam kategori layanan tertentu. Penelitian ini memberikan kontribusi yang unik dengan menggunakan dataset ulasan asli dari Tokopedia yang dikumpulkan melalui teknik web *scraping* dan diberi label secara manual ke dalam empat kategori utama, yaitu fitur aplikasi, layanan, pembayaran, dan promosi. Selain itu, penelitian ini menerapkan tahap praproses teks yang lengkap dan sistematis, serta mengevaluasi kinerja model secara komprehensif menggunakan metrik *precision, recall,* dan F1-*score* untuk setiap kategori. Dengan pendekatan ini, penelitian dapat memberikan wawasan yang lebih terarah bagi pengembang aplikasi dalam memahami fokus utama keluhan atau kebutuhan pengguna.

Dalam penelitian ini, ulasan pengguna Tokopedia diklasifikasikan ke dalam empat kategori utama, yaitu fitur aplikasi, layanan, pembayaran, dan promosi menggunakan algoritma *Naive Bayes*. Hasil yang diperoleh menunjukkan bahwa metode ini mampu memberikan kinerja klasifikasi yang cukup baik dengan tingkat akurasi sebesar 83,41%. Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan sistem analisis ulasan otomatis yang membantu perusahaan memahami kebutuhan dan meningkatkan pengalaman pengguna secara keseluruhan.

2. METODE PENELITIAN

Penelitian ini bertujuan untuk mengklasifikasikan dan memprediksi ulasan pengguna aplikasi Tokopedia ke dalam beberapa kategori menggunakan algoritma *Naive Bayes.* Tahapan yang dilakukan dalam penelitian ini terdiri dari data *scraping,* data *labeling, preprocessing*, TF-IDF, *training models,* dan model *performance evaluation*. Tahapan tersebut dijelaskan secara rinci sebagai berikut:



Gambar 1. Tahapan Penelitian

https://doi.org/10.35145/joisie.v9i1.4993

2.1. Pengumpulan Data (Scraping)

Data ulasan pengguna diperoleh dari aplikasi Tokopedia dalam kurun waktu lima bulan terakhir terhitung sejak 16 Oktober 2024 hingga 27 Maret 2025. menggunakan teknik web *scraping*. Ulasan yang dikumpulkan merupakan tanggapan pengguna terhadap aplikasi yang tersedia di *platform* Google *Play Store* (Nugraha & Gustian, 2024).Setelah proses pengumpulan data, dilakukan pemilihan ulasan yang relevan dan kontekstual (Ruslan et al., 2023). Selain itu, data yang terpilih diklasifikasikan ke dalam beberapa label tertentu. Semua data disimpan dalam format teks dan disusun menjadi dataset yang digunakan pada tahap analisis selanjutnya.

2.2. Seleksi dan Pelabelan Data

Data ulasan pengguna diperoleh dari aplikasi Tokopedia menggunakan teknik web *scraping*, dengan total data ulasan yang diambil sebanyak 1.500 data. Ulasan yang dikumpulkan merupakan respon pengguna terhadap aplikasi yang tersedia pada platform Google *Play Store*. Setelah proses pengumpulan data dilakukan, dilakukan pemilihan ulasan yang relevan dan kontekstual. Selain itu, data yang terpilih dikelompokkan menjadi 4 kategori, yaitu: Fitur Aplikasi, Layanan, Promosi, dan Pembayaran. Seluruh data disimpan dalam format teks dan disusun menjadi dataset yang digunakan pada tahap analisis selanjutnya. Setelah data ulasan berhasil dikumpulkan, dilakukan proses pelabelan manual menjadi empat kategori utama. Dari total 1.500 ulasan, sebanyak 1.244 ulasan memenuhi kualifikasi, sedangkan sisanya tidak relevan atau tidak memiliki konteks yang jelas. Hasil pelabelan menunjukkan bahwa kategori Fitur Aplikasi mendominasi dengan jumlah 572 ulasan (45,98%), diikuti oleh Layanan sebanyak 394 ulasan (31,67%), Pembayaran sebanyak 79 ulasan (6,35%), dan Promo sebanyak 199 ulasan (16,00%). Distribusi ini menunjukkan bahwa sebagian besar ulasan pengguna berfokus pada pengalaman mereka dalam menggunakan fitur aplikasi Tokopedia.

Tabel 1. Pelabelan Data

No	***Content***	***Category***
0	sudah langganan plus, pakai vocer yg katanya b...	Promo
1	Gak jelas. Daftar baru kata nya dapat promo, u...	Promo
2	Tadinya saya dapat promo yang besar, tetapi se...	Promo
3	Untuk tokopedia hilangkan saja fitur kurir rek..	Fitur aplikasi
4	Tolong dikasih fitur ganti ekspedisi, saya mua...	Fitur aplikasi
…	…	…
1244	Pusing sama sistem kupon tokped, punya kupon t...	Fitur aplikasi

2.3. Preprocessing

Agar data teks dapat diproses oleh algoritma, maka dilakukan beberapa tahap preprocessing yaitu *lower case, cleaning, tokenizing, stopwords*, dan stemming sebagai berikut:

1. Case Folding

Case folding merupakan tahapan dalam pengolahan teks yang bertujuan untuk mengubah semua huruf kapital menjadi huruf kecil (Annisa & Ulama, 2023). Langkah ini berguna untuk menyamakan format penulisan, sehingga proses pencarian dan analisis teks menjadi lebih efektif. Hal

https://doi.org/10.35145/joisie.v9i1.4993

ini penting karena tidak semua dokumen menggunakan huruf kapital secara konsisten (Indriyani et al., 2023). Pada tabel 2 kolom sebelah kiri, kita dapat melihat huruf kapital seperti "Tadinya", "Gopay", dan "Ada". Setelah *case folding*, semua huruf kapital diubah menjadi huruf kecil, seperti "tadinya", "gopay", dan "ada".

Formulasi :

$$CF(d_i) = lower(d_i) \quad (1)$$

Keterangan:

1. Fungsi *lower*() mengubah semua huruf kapital menjadi huruf kecil.
2. Tujuannya adalah menyamakan bentuk kata, misalnya Ada dan ada menjadi sama.

Tabel 2. Hasil Case Folding

Text	*Case Folding*
Tadinya saya dapat promo yang besar, tetapi setelah saya top up saldo Gopay, promonya mendadak jadi kecil. Ada juga kupon lain yang tersedia tetapi tidak dapat digunakan padahal sudah memenuhi persyaratan.	tadinya saya dapat promo yang besar, tetapi setelah saya top up saldo gopay, promonya mendadak jadi kecil. ada juga kupon lain yang tersedia tetapi tidak dapat digunakan padahal sudah memenuhi persyaratan.

2. Cleaning

Tanda *Cleaning* merupakan langkah awal dalam pengolahan data teks yang bertujuan untuk membuang unsur-unsur yang tidak diperlukan atau dianggap mengganggu, sehingga data menjadi lebih bersih dan siap untuk dilakukan analisis lebih lanjut (Ravil et al., 2024). Pada tabel 3 kolom pertama (hasil dari lipatan huruf), teks masih mengandung tanda baca seperti koma (,), titik (.), dan penghubung lainnya. Setelah dibersihkan, tanda baca ini dihilangkan. Hasilnya adalah teks yang lebih bersih dan siap untuk diproses lebih lanjut tanpa gangguan dari simbol yang tidak perlu.

Formulasi :

$$CL(d_i) = d_i - \{c \in d_i \mid c \notin \text{alfabet dan spasi}\} \quad (2)$$

Keterangan:

1. Simbol, angka, dan karakter khusus dihapus dari teks.
2. Dapat menggunakan regex seperti [^a-zA-Z\s].

Tabel 3. Hasil *Cleaning*

Case Folding	*Cleaning*
tadinya saya dapat promo yang besar, tetapi setelah saya top up saldo gopay, promonya mendadak jadi kecil. ada juga kupon lain yang tersedia tetapi tidak dapat digunakan padahal sudah memenuhi persyaratan.	tadinya saya dapat promo yang besar tetapi setelah saya top up saldo gopay promonya mendadak jadi kecil ada juga kupon lain yang tersedia tetapi tidak dapat digunakan padahal sudah memenuhi persyaratan

3. Tokenizing

Proses ini dilakukan dengan memecah teks dari bentuk kalimat menjadi kata-kata terpisah. Tahap ini disebut tokenisasi dan bertujuan untuk memudahkan analisis dengan memecah kalimat menjadi unit kata (Prasetyo et al., 2023). Dari tabel 4, kita dapat melihat bahwa kalimat panjang dipecah menjadi 31 token yang mewakili setiap kata. Token-token ini kemudian akan digunakan dalam proses lebih lanjut seperti stemming, penghapusan stopword, atau klasifikasi teks.

Formulasi :

$$T(d_i) = \text{token}(CL(CF(d_i))) \quad (3)$$

Keterangan:

1. Fungsi token() memecah kalimat berdasarkan spasi.
2. Output berupa list kata: $T(d_i) = \{w_1, w_2, ..., w_m\}$

https://doi.org/10.35145/joisie.v9i1.4993

Tabel 4. Hasil *Tokenizing*

Cleaning	*Tokenizing*
tadinya saya dapat promo yang besar tetapi setelah saya top up saldo gopay promonya mendadak jadi kecil ada juga kupon lain yang tersedia tetapi tidak dapat digunakan padahal sudah memenuhi persyaratan	['tadinya', 'saya', 'dapat', 'promo', 'yang', 'besar', 'tetapi', 'setelah', 'saya', 'top', 'up', 'saldo', 'gopay', 'promonya', 'mendadak', 'jadi', 'kecil', 'ada', 'juga', 'kupon', 'lain', 'yang', 'tersedia', 'tetapi', 'tidak', 'dapat', 'digunakan', 'padahal', 'sudah', 'memenuhi', 'persyaratan']

4. Stopwords

Pada tahap *stop-word removal*, kata-kata umum yang tidak memiliki makna signifikan terhadap konteks teks, atau disebut *stop-word*, dihilangkan dari daftar token yang telah melalui proses stemming. Penghapusan ini bertujuan untuk meningkatkan efisiensi dalam proses pelatihan model atau pengklasifikasian data teks (Sabrani et al., 2020). Dari tabel 5 kata-kata seperti "yang", "dan", "saya", "tidak", atau "ada" sering muncul dalam teks tetapi tidak memberikan nilai informasi yang signifikan.

Formulasi :

$$SR(d_i) = \{w_j \in T(d_i) \mid w_j \notin SW\} \quad (4)$$

1. SW adalah himpunan stopword bahasa Indonesia.
2. Kata-kata seperti "yang", "ada", "saya", "dapat", akan dihapus.

Tabel 5. Hasil *Stopwords*

Tokenizing	*Stopwords*
['tadinya', 'saya', 'dapat', 'promo', 'yang', 'besar', 'tetapi', 'setelah', 'saya', 'top', 'up', 'saldo', 'gopay', 'promonya', 'mendadak', 'jadi', 'kecil', 'ada', 'juga', 'kupon', 'lain', 'yang', 'tersedia', 'tetapi', 'tidak', 'dapat', 'digunakan', 'padahal', 'sudah', 'memenuhi', 'persyaratan']	['tadinya', 'promo', 'besar', 'top', 'up', 'saldo', 'gopay', 'promonya', 'mendadak', 'jadi', 'kecil', 'kupon', 'tersedia', 'digunakan', 'padahal', 'memenuhi', 'persyaratan']

5. Stemming

Stemming merupakan proses untuk memetakan atau menyederhanakan sebuah kata ke bentuk dasarnya. Tujuan dari proses ini adalah untuk menghapus berbagai jenis imbuhan, seperti awalan (*prefix*), akhiran (*suffix*), maupun gabungan keduanya (*confix*), yang terdapat pada setiap (Sari & Hayuningtyas, 2019). Tujuannya adalah agar kata-kata yang memiliki makna sama tetapi berbeda bentuk (seperti “digunakan”, “menggunakan”, “penggunaan”) dianggap satu representasi: yaitu “guna”. Dengan begitu, hasil analisis akan lebih konsisten dan tidak membedakan kata hanya karena variasi bentuknya.

Formulasi :

$$S(d_i) = \{stem(w_j) \mid w_j \in SR(d_i)\} \quad (5)$$

Keterangan:

1. Fungsi stem() mengubah kata ke bentuk dasar.
2. Misalnya, "tadinya" → "tadi", "digunakan" → "guna".

Tabel 6. Hasil *Stemming*

Stopwords	*Stemming*
['tadinya', 'promo', 'besar', 'top', 'up', 'saldo', 'gopay', 'promonya', 'mendadak', 'jadi', 'kecil', 'kupon', 'tersedia', 'digunakan', 'padahal', 'memenuhi', 'persyaratan']	['tadi', 'promo', 'besar', 'top', 'up', 'saldo', 'gopay', 'promonya', 'dadak', 'jadi', 'kecil', 'kupon', 'sedia', 'guna', 'padahal', 'penuh', 'syarat']

https://doi.org/10.35145/joisie.v9i1.4993

2.4. TF-IDF

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk mengekstrak fitur-fitur penting dari teks jawaban kuesioner. TF-IDF mengukur tingkat kepentingan suatu kata dalam sebuah dokumen dibandingkan dengan keseluruhan kumpulan dokumen, dengan tujuan menghasilkan akurasi yang tinggi (Gautama, 2023).

1. *Term Frequency* (TF)

Mengukur seberapa sering sebuah term (kata) muncul dalam dokumen tertentu.
Formulasi :

$$TF(t, d) = f(t, d) / \max\{f(w, d) \mid w \in d\} \quad (6)$$

Keterangan:

1. f(t, d): jumlah kemunculan term t dalam dokumen d.
2. max{f(w, d)}: frekuensi maksimum dari semua term dalam dokumen d.
3. TF bernilai tinggi jika term sering muncul dalam dokumen.

2. *Inverse Document Frequency* (IDF)

Mengukur seberapa penting sebuah term dalam seluruh korpus.

Formulasi :

$$IDF(t) = \log(N / df(t)) \quad (7)$$

Keterangan:

1. N: jumlah total dokumen dalam korpus.
2. df(t): jumlah dokumen yang mengandung term t.
3. IDF tinggi jika term jarang muncul dalam dokumen.

3. Gabungan TF-IDF

TF-IDF memberikan bobot tinggi untuk kata yang sering muncul di dokumen tertentu namun jarang di seluruh korpus.

Formulasi :

$$TF\text{-}IDF(t, d) = TF(t, d) \times IDF(t) \quad (8)$$

Keterangan:

1. TF-IDF merupakan hasil perkalian antara nilai TF dan IDF dari sebuah term.
2. Makin tinggi nilai TF-IDF, makin penting kata tersebut untuk dokumen tersebut.

2.5. Desain Model Klasifikasi

Model dikembangkan menggunakan algoritma *Naive Bayes* yang sesuai untuk klasifikasi teks. Hasil klasifikasi yang diperoleh nantinya akan dievaluasi, dimana nilai evaluasi tersebut dapat digunakan untuk menilai sejauh mana keberhasilan metode yang diterapkan dalam pengujian. Salah satu cara untuk melakukan evaluasi dalam analisis sentimen adalah dengan menggunakan *Confusion Matrix* (Grandis et al., 2021). Evaluasi dilakukan menggunakan *recall* sejauh mana model dapat menemukan semua kasus positif yang ada. *Precision* berfokus pada keakuratan prediksi positif yang dibuat oleh model, mengurangi kesalahan dalam hasil positif. Akurasi memberikan gambaran umum tentang berapa banyak prediksi yang benar dibandingkan dengan total data, sementara *error rate* menunjukkan seberapa sering model membuat kesalahan dalam klasifikasi. Berikut merupakan formulasi Klasifikasi *Naïve Bayes.*

Formulasi :

$$P(w|c) = P(c) \times \prod_{(i=1)}^{(Q)} P(w_i|c) \quad (9)$$

https://doi.org/10.35145/joisie.v9i1.4993

1. P(w|c) : *Posterior probabili*ty (probabilitas hipotesis setelah melihat data)
2. P(w_i|c) : *Likelihood probability* (probabilitas fitur ke-i muncul dalam kelas c)
3. P(c) : *Prior probability* (probabilitas awal kelas c)
   Q : Jumlah fitur

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Ulasan yang diambil disusun berdasarkan urutan terbaru berdasarkan waktu publikasinya. Proses *scraping* dilakukan untuk mengumpulkan ulasan dalam kurun waktu lima bulan terakhir terhitung sejak 16 Oktober 2024 hingga 27 Maret 2025. Dari proses *scraping*, diperoleh sekitar 1.500 ulasan mentah. Selanjutnya, dilakukan proses pelabelan manual terhadap data ulasan ke dalam empat kategori utama, yaitu Fitur Aplikasi, Layanan, Pembayaran, dan Promo. Setelah melalui proses verifikasi dan penyaringan, diperoleh 1.244 data ulasan yang valid dan sesuai dengan kategori, sedangkan sisanya (256 data) tidak dimasukkan karena tidak relevan atau tidak sesuai konteks klasifikasi.

3.2 Preprocessing

Data ulasan akhir yaitu yang diklasifikasikan kemudian diproses melalui tahap praproses sehingga siap digunakan dalam model pelatihan. Pada tabel berikut, dapat dilihat bagaimana teks asli mengalami berbagai tahapan, dimulai dari *case folding* hingga *stemming.* Contoh teks yang awalnya berisi 31 kata (token) setelah *stemming* menjadi lebih ringkas dan terfokus, dengan hanya 17 token yang tersisa.

Tabel 8. Hasil Akhir *Preprocessing*

Text	*Learn*	*Preprocessing*	*Text*	*Learn*
tadinya saya dapat promo yang besar, tetapi setelah saya top up saldo gopay, promonya mendadak jadi kecil. ada juga kupon lain yang tersedia tetapi tidak dapat digunakan padahal sudah memenuhi persyaratan.	31	*Stemming*	['tadi', 'promo', 'besar', 'top', 'up', 'saldo', 'gopay', 'promonya', 'dadak', 'jadi', 'kecil', 'kupon', 'sedia', 'guna', 'padahal', 'penuh', 'syarat']	17

Berdasarkan hasil preprocessing pada data ulasan, dapat disimpulkan bahwa tahapan seperti *casefolding, cleaning, tokenizing, stopwords*, dan *stemming* mampu meningkatkan teks mentah menjadi data yang lebih bersih dan terstruktur secara signifikan. Dari contoh yang ditunjukkan tabel diatas, *case folding* yang terdiri dari 31 kata berhasil diringkas menjadi 17 kata inti setelah melalui proses ini. Proses ini membantu mengurangi kata-kata yang tidak relevan dan menyatukan berbagai bentuk kata menjadi bentuk dasarnya. Dengan demikian, hasil preprocessing ini sangat penting untuk meningkatkan efektivitas pada tahap analisis lanjutan, seperti klasifikasi sentimen atau pengelompokan topik ulasan.

3.3 TF-IDF

Secara umum, nilai TF-IDF yang tinggi menunjukkan bahwa kata-kata tersebut memiliki relevansi yang signifikan dalam konteks ulasan pengguna di platform Tokopedia. Kata-kata dengan nilai TF-IDF yang lebih tinggi berarti sering muncul dalam ulasan yang memiliki informasi penting, baik itu dalam hal produk, layanan, maupun proses transaksi. Berikut adalah 10 kata dengan rata-rata TF-IDF tertinggi, beserta penjelasan lebih lanjut:

https://doi.org/10.35145/joisie.v9i1.4993

Tabel 9. Hasil TF-IDF

Kata	Rata-rata TF-IDF
Aplikasi	0.0289876
Barang	0.0279853
Kirim	0.0263308
Belanja	0.022954
Bayar	0.0223628
Kurir	0.0204631
Lama	0.0201286
Promo	0.0200289
Guna	0.018697
Ongkir	0.0184766

Kata-kata seperti "aplikasi", "promo", "barang", dan "belanja" mencerminkan perhatian pengguna terhadap fitur aplikasi, penawaran khusus, dan produk. Sementara itu, kata-kata "kirim", "kurir", "ongkir" berfokus pada layanan pengiriman, dan "bayar" berkaitan dengan transaksi, dengan "lama" sering muncul dalam keluhan keterlambatan. Data ini memberi wawasan penting bagi Tokopedia untuk memperbaiki fitur aplikasi, kualitas produk, layanan pengiriman, dan promosi.

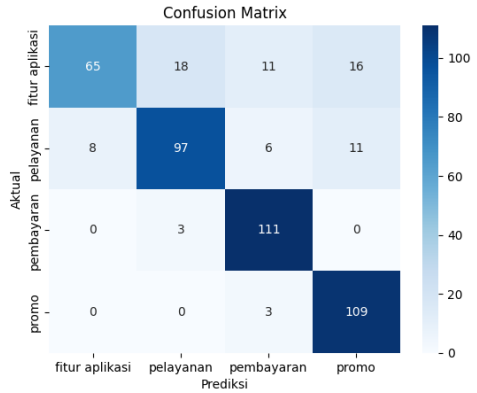
3.4 Desain Model Klasifikasi

Model klasifikasi yang digunakan dalam penelitian ini adalah *Multinomial Naive Bayes*. Evaluasi dilakukan terhadap data uji menggunakan metrik akurasi, presisi, *recall*, dan f1-*score*, serta didukung oleh hasil laporan klasifikasi. Model tersebut menghasilkan akurasi sebesar 83,41%, yang menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan ke dalam kategori yang sesuai.

Tabel 10. *Classificaton Report*

Kelas	***Precision***	***Recall***	**F1-*Score***	***Support***
Fitur Aplikasi	0.89	0.59	0.71	110
Pelayanan	0.82	0.80	0.81	122
Pembayaran	0.85	0.97	0.91	114
Promo	0.80	0.97	0.88	112
Accuracy			**0.83**	458
Macro Avg	0.84	0.83	0.83	458
Weighted Avg	0.84	0.83	0.83	458

Berdasarkan *classification report*, model *Naive Bayes* mampu mengklasifikasikan ulasan. Kategori pembayaran dan promo menunjukkan kinerja terbaik dengan *recall* tinggi, masing-masing sebesar 0,97, sedangkan fitur aplikasi memiliki *recall* rendah (0,59) meskipun presisinya cukup tinggi (0,89). Secara keseluruhan, nilai f1-*score* rata-rata sebesar 0,83 menunjukkan bahwa model cukup baik dalam membedakan keempat kategori ulasan.



Gambar. 2 *Confussion Matrix*

https://doi.org/10.35145/joisie.v9i1.4993

Berdasarkan *Confusion Matrix* di atas, terlihat bahwa model mampu mengklasifikasikan ulasan dengan cukup baik. Kategori pembayaran dan promo menunjukkan kinerja tertinggi dengan jumlah prediksi yang benar masing-masing 111 dan 109. Akan tetapi, terdapat beberapa kesalahan klasifikasi, terutama pada kategori fitur aplikasi dan layanan, yang masing-masing memiliki beberapa prediksi yang salah diarahkan ke kategori lainnya. Misalnya, 18 ulasan yang seharusnya masuk dalam fitur aplikasi malah diprediksi sebagai layanan. Meskipun demikian, secara umum model menunjukkan kinerja yang cukup baik dalam membedakan keempat kategori tersebut.

Analisis hasil klasifikasi menunjukkan bahwa kinerja model bervariasi antar kelas ulasan. Hal ini sebagian besar disebabkan oleh distribusi data yang tidak seimbang, di mana kelas seperti promo layanan atau memiliki data yang lebih sedikit daripada kelas fitur aplikasi. Ketidakseimbangan ini menyebabkan model cenderung bias terhadap kelas mayoritas. Selain itu, konten yang tumpang tindih antarkelas juga memengaruhi akurasi model. Beberapa ulasan berisi informasi yang terkait dengan lebih dari satu kategori, sehingga menyulitkan model untuk mengidentifikasi kelas yang paling relevan. Misalnya, ulasan tentang proses pembayaran yang lambat juga dapat berisi kritik terhadap layanan, yang menyebabkan ambiguitas dalam klasifikasi.

3.5 Evaluasi Model Prediksi Teks

Untuk melihat secara detail kemampuan model dalam mengklasifikasikan ulasan, berikut adalah 5 contoh hasil prediksi model *Naive Bayes*. Setiap ulasan dibandingkan antara label manual (label sebenarnya) dan label prediksi model untuk menentukan apakah model berhasil mengklasifikasikan dengan benar atau tidak.

Tabel 11. Hasil Prediksi Model *Naive Bayes*

Ulasan	Label manual	Prediksi
pertama kali kecewa tokped ngasih promo bayar paylater abis satu metode bayar dipake lapor cs kata direspon menit nyata lama banget ampe tengah jam	Pelayanan	Pembayaran
gratis ongkir naik minimal belanja, padahal sudah langganan. semoga kembali ke syarat yang dulu, soalnya sekarang jadi boros.	Fitur Aplikasi	Promo
bayar kartu tokped aplikasi tokped bayar cicil hampir jt lebih pas cek apk brimo ko bayar tagih muncul tgl hitung aneh banget cerah rugi dong kalo bayar g potong cicil	Fitur Aplikasi	Pembayaran
promo nihapaan semua voucer tdkbisa guna tiba gakbisa belanja pake vocer	Promo	Promo
metode bayar blokir alas jelas issu langgar tentu padahal murni transaksi	Pembayaran	Pembayaran

Dari 5 sampel ulasan yang diuji, terdapat 2 ulasan yang berhasil diklasifikasikan dengan benar oleh model Naive Bayes, yaitu pada ulasan ke-4 (promo) dan ke-5 (pembayaran). Sementara itu, 3 ulasan lainnya mengalami kesalahan klasifikasi. Ulasan ke-1 yang seharusnya termasuk dalam kategori pelayanan diprediksi sebagai pembayaran, ulasan ke-2 yang seharusnya fitur aplikasi diprediksi sebagai promo, dan ulasan ke-3 juga mengalami kekeliruan dengan label fitur aplikasi diprediksi sebagai pembayaran. Kesalahan ini menunjukkan bahwa model masih sulit membedakan ulasan yang konteksnya saling berkaitan, terutama antara fitur aplikasi, pembayaran, dan pelayanan.

4. SIMPULAN

Kesimpulan Model klasifikasi berhasil mencapai akurasi sebesar 83%, yang menunjukkan kinerja yang cukup baik dalam mengelompokkan ulasan pengguna ke dalam empat kategori. Kinerja

https://doi.org/10.35145/joisie.v9i1.4993

terbaik ditunjukkan pada kategori pembayaran (F1-*score*: 0.91) dan promo (F1-*score*: 0.88), di mana model mampu mengenali sebagian besar ulasan dengan tepat. Kategori pelayanan juga menunjukkan hasil yang stabil (F1-*score*: 0.81), sementara performa terendah ada pada kategori fitur aplikasi (F1-*score*: 0.71), menandakan masih adanya kebingungan model dalam mengenali ulasan terkait fitur. Secara keseluruhan, nilai *macro average* dan *weighted average* F1-*score* yang sama-sama 0.83 memperkuat bahwa model memiliki performa yang cukup merata, meskipun masih ada ruang perbaikan pada beberapa kelas.

Namun, penelitian ini memiliki beberapa keterbatasan, antara lain distribusi data yang tidak seimbang antarkategori, serta potensi hilangnya makna akibat tahapan praproses seperti penghilangan stopword atau stemming. Untuk pengembangan selanjutnya, model dapat disempurnakan dengan algoritma eksplorasi lain seperti SVM atau *deep learning*, penerapan teknik penyeimbangan dan penyetelan hiperparameter. Selain itu, representasi teks berbasis *word embedding* seperti Word2Vec atau BERT juga dapat dieksplorasi untuk meningkatkan akurasi.

Secara praktis, hasil penelitian ini dapat dimanfaatkan oleh platform *e-commerce* seperti Tokopedia untuk mengelompokkan ulasan pengguna secara otomatis, sehingga memudahkan analisis sentimen, meningkatkan layanan, dan membuat keputusan berbasis data secara lebih efisien.

5. DAFTAR PUSTAKA

Agneresa, A., Hananto, A. L., Hilabi, S. S., Hananto, A., & Tukino, T. (2022). Strategi Promosi Penerapan Data Mining Mahasiswa Baru Dengan Metode K-Means Clustering. *Dirgamaya: Jurnal Manajemen Dan Sistem Informasi*, *2*(2), 25–34. Https://Doi.Org/10.35969/Dirgamaya.V2i2.275

Annisa, Z., & Ulama, B. S. S. (2023). Analisis Sentimen Data Ulasan Pengguna Aplikasi "Pedulilindungi" Pada Google Play Store Menggunakan Metode Naïve Bayes Classifier Model Multinomial. *Jurnal Sains Dan Seni ITS*, *11*(6), D464--D471.

Apriliyani, E., & Salim, Y. (2022). Analisis Performa Metode Klasifikasi Naïve Bayes Classifier Pada Unbalanced Dataset. *Indonesian Journal Of Data And Science*, *3*(2), 47–54. Https://Doi.Org/10.56705/Ijodas.V3i2.45

Gautama, I. M. B. (2023). Klasifikasi Data Saran Pemustaka Di Perpustakaan STIKOM Bali Menggunakan TF-IDF Dan Multinomial Naive Bayes. *Jurnal Sistem Dan Informatika (JSI)*, *17*(2), 62–72. Https://Doi.Org/10.30864/Jsi.V17i2.490

Grandis, G. F., Sari, Y. A., & Indriati, I. (2021). Seleksi Fitur Gain Ratio Pada Analisis Sentimen Kebijakan Pemerintah Mengenai Pembelajaran Jarak Jauh Dengan K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *5*(8), 3507–3514.

Hananto, A. L., Hananto, A., Huda, B., Rahman, A. Y., Novalia, E., & Priyatna, B. (2024). Determination Of Training Participants In Community Work Training Centers Using The Naïve Bayes Classifier Algorithm. *JOIV: International Journal On Informatics Visualization*, *8*(3), 1162–1167. Https://Doi.Org/10.62527/Joiv.8.3.1995

Hasibuan, E., & Heriyanto, E. A. (2022). Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier. *Jurnal Teknik Dan Science*, *1*(3), 13–24. Https://Doi.Org/10.56127/Jts.V1i3.434

Indriyani, F. A., Fauzi, A., & Faisal, S. (2023). Analisis Sentimen Aplikasi Tiktok Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine. *TEKNOSAINS: Jurnal Sains, Teknologi Dan Informatika*, *10*(2), 176–184. Https://Doi.Org/10.37373/Tekno.V10i2.419

Irawati, R., & Prasetyo, I. B. (2021). Pemanfaatan Platform E-Commerce Melalui Marketplace Sebagai Upaya Peningkatan Penjualan Dan Mempertahankan Bisnis Di Masa Pandemi (Studi Pada UMKM Makanan Dan Minuman Di Malang). *Jurnal*

Penelitian Manajemen Terapan (PENATARAN), *6*(2), 114–133.

Kurniawan, B., Suwarisman, A., Afriyanti, I., Wahyudi, A., & Saputra, D. D. (2023). Analisis Sentimen Complain Dan Bukan Complain Pada Twitter Telkomsel Dengan SMOTE Dan Naïve Bayes. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *7*(1), 106–113. Https://Doi.Org/10.35870/Jtik.V7i1.691

Kurniawan, I., Hananto, A. L., Hilabi, S. S., Hananto, A., Priyatna, B., & Rahman, A. Y. (2023). Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *10*(1), 731–740. Https://Doi.Org/10.35957/Jatisi.V10i1.3582

Mardiani, E., Rahmansyah, N., Ningsih, S., Lantana, D. A., Wirawan, A. S. P., Wijaya, S. A., & Putri, D. N. (2023). Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *Innovative: Journal Of Social Science Research*, *3*(2), 13880–13892.

Nugraha, D., & Gustian, D. (2024). Analisis Sentimen Penggunaan Aplikasi Transportasi Online Pada Ulasan Google Play Store Dengan Metode Naive Bayes Classifier. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer Dan Manajemen)*, *5*(1), 326–335. Https://Doi.Org/10.30645/Kesatria.V5i1.341

Pirunthavi, S., & Jayalath, E. (2021). Naïve Bayes Algorithm For Large Scale Text Classification. *Instrumentation*, *8*(4), 55–62.

Prasetyo, S. D., Hilabi, S. S., & Nurapriani, F. (2023). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes Dan KNN. *Jurnal Komtekinfo*, *10*(1), 1–7. Https://Doi.Org/10.35134/Komtekinfo.V10i1.330

Rachmawati, M. (2024). Adopsi E-Commerce UMKM Sebagai Upaya Adaptasi Perubahan Perilaku Konsumen. *Jurnal EMT KITA*, *8*(2), 695–700. Https://Doi.Org/10.35870/Emt.V8i2.2377

Ravil, M., Agustian, S., Fikry, M., & Insani, F. (2024). Peningkatan Performa Klasifikasi Sentimen Tweet Kaesang Menggunakan Naïve Bayes Dengan PSO Pada Dataset Kecil. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(6), 2909–2917. Https://Doi.Org/10.30865/Klik.V4i6.1939

Ruslan, R., Khalifatun, U. N., & Rahman, U. (2023). Penelitian Grounded Theory: Pengertian, Prinsip-Prinsip, Metode Pengumpulan Dan Analisis Data. *Edu Sociata: Jurnal Pendidikan Sosiologi*, *6*(2), 699–708. Https://Doi.Org/10.33627/Es.V6i2.1483

Sabrani, A., Wedashwara, I. G. W., & Bimantoro, F. (2020). Multinomial Naïve Bayes Untuk Klasifikasi Artikel Online Tentang Gempa Di Indonesia. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTIKA)*, *2*(1), 89–100. Https://Doi.Org/10.29303/Jtika.V2i1.87

Sanjaya, J., Priyatna, B., Hilabi, S. S., & Others. (2024). Analisis Sentimen Terhadap Opini Proyek Kereta Cepat Menggunakan Metode Naïve Bayes Classifier. *JURNAL FASILKOM*, *14*(1), 263–270. Https://Doi.Org/10.37859/Jf.V14i1.6598

Sari, R., & Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website. *Indonesian Journal On Software Engineering (IJSE)*, *5*(2), 51–60. Https://Doi.Org/10.31294/Ijse.V5i2.6957

Seran, W. (2024). Analisis Kepuasan Pengguna Pada Fitur Marketplace Di Aplikasi Facebook Menggunakan Card Shorting. *CONTAR: Jurnal Ilmu Komputer*, *2*(1), 13–18.

Sulistyawati, U. S., & Munawir. (2024). Membangun Keunggulan Kompetitif Melalui Platform E-Commerce: Studi Kasus Tokopedia. *Jurnal Manajemen Dan Teknologi*, *1*(1), 43–56. Https://Doi.Org/10.63447/Jmt.V1i1.776