

KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI SHOPEE DENGAN ALGORITMA NAÏVE BAYES DAN ID3

Sidik Mahesa Putra, Deny Sundari

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Darwan Ali
Jalan Batu Berlian No. 10 Sampit
sidikmp288@gmail.com

ABSTRAK

Pesatnya perkembangan *e-commerce* di Indonesia telah meningkatkan penggunaan aplikasi Shopee sebagai salah satu *platform* belanja daring terpopuler. Fitur ulasan pengguna pada aplikasi ini menghasilkan data teks dalam jumlah besar yang berpotensi menjadi sumber informasi penting terkait kepuasan konsumen. Namun, tingginya volume ulasan membuat analisis manual menjadi tidak efisien. Penelitian ini bertujuan mengklasifikasikan sentimen ulasan pengguna Shopee secara otomatis serta membandingkan kinerja algoritma Naïve Bayes dan ID3. Data penelitian menggunakan 1.950 ulasan dari platform *Kaggle*, yang diproses melalui metode *Knowledge Discovery in Databases* (KDD) meliputi seleksi data, pra-pemrosesan teks, transformasi fitur menggunakan TF-IDF, pembagian data dengan skema 80:20 (1.560 data pelatihan, 390 data pengujian), penerapan algoritma klasifikasi, dan evaluasi menggunakan metrik akurasi, presisi, *recall*, serta *F1-score*. Hasil pengujian menunjukkan bahwa *Naïve Bayes* memiliki kinerja lebih baik dengan akurasi 83,07%, presisi positif 87%, presisi negatif 80%, *recall* positif 77%, *recall* negatif 88%, *F1-score* positif 82%, dan *F1-score* negatif 84%. Sementara itu, ID3 mencatat akurasi 73,85% dengan keunggulan pada *recall* kelas positif sebesar 79%. Temuan ini mengindikasikan bahwa *Naïve Bayes* lebih efektif dan konsisten untuk klasifikasi sentimen ulasan pengguna Shopee dibandingkan ID3.

Kata kunci : Klasifikasi Sentimen, Shopee, Naïve Bayes, ID3, TF-IDF

1. PENDAHULUAN

Perkembangan industri *e-commerce* di Indonesia dalam beberapa tahun terakhir menunjukkan pertumbuhan yang pesat serta memberikan kontribusi signifikan terhadap perekonomian digital [1]. Salah satu *platform e-commerce* yang mendominasi pasar adalah Shopee, yang mencatatkan jumlah kunjungan pengguna hingga 277,5 juta dalam satu periode [2]. Shopee tidak hanya menawarkan layanan belanja daring, tetapi juga fitur tambahan seperti ShopeePay, ShopeeFood, dan sistem ulasan pengguna. Fitur ulasan ini menjadi sumber informasi berharga untuk menilai tingkat kepuasan konsumen dan memberikan masukan bagi pengembangan layanan. Namun, meningkatnya volume ulasan membuat proses analisis manual menjadi tidak efisien, sehingga dibutuhkan sistem otomatis yang mampu mengklasifikasikan sentimen secara cepat dan akurat [3][4].

Klasifikasi sentimen adalah metode yang digunakan untuk mengidentifikasi pendapat individu maupun kelompok terhadap suatu isu, produk, layanan, atau kelompok tertentu. Metode ini dapat membantu perusahaan *e-commerce* dalam memahami persepsi pengguna dan meningkatkan kualitas layanan. Dalam penelitian ini digunakan dua algoritma machine learning, yaitu Naïve Bayes yang berbasis probabilistik dan dikenal efisien dalam pengolahan teks, serta ID3 Decision Tree yang membentuk model pohon keputusan berdasarkan atribut dengan informasi tertinggi [5][6][7]. Kedua algoritma tersebut dipilih karena keunggulannya masing-masing dalam klasifikasi berbasis teks.

Penelitian ini menggunakan dataset ulasan pengguna Shopee yang diperoleh dari *Kaggle*. Data ulasan kemudian melalui tahap pelabelan berdasarkan rating dan diproses menggunakan kerangka kerja *Knowledge Discovery in Databases* (KDD). KDD merupakan rangkaian proses terintegrasi yang dilakukan secara berurutan, di mana data mining menjadi salah satu tahap yang terdapat di dalamnya [8]. KDD mencakup tahap seleksi data, preprocessing, transformasi, klasifikasi, dan evaluasi. Evaluasi kinerja model dilakukan menggunakan empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, yang memberikan gambaran lebih komprehensif dibanding hanya menggunakan akurasi semata.

Beberapa penelitian terdahulu, seperti penelitian oleh Utami dan Erfina yang menganalisis sentimen pinjaman online di Twitter menggunakan algoritma *Support Vector Machine* (SVM), hanya berfokus pada metrik akurasi dengan nilai 62%. Penelitian tersebut menunjukkan dominasi sentimen negatif sebesar 59% dibandingkan sentimen positif sebesar 41%. Berbeda dengan penelitian tersebut, studi ini membandingkan dua algoritma sekaligus, yaitu *Naïve Bayes* dan ID3, serta menilai kinerjanya berdasarkan empat metrik evaluasi utama untuk mendapatkan gambaran yang lebih menyeluruh.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan dan membandingkan kinerja algoritma *Naïve Bayes* dan ID3 dalam klasifikasi sentimen ulasan pengguna aplikasi Shopee. Hasil penelitian diharapkan dapat memberikan kontribusi pada pengembangan studi *machine learning* berbasis teks, sekaligus menjadi referensi praktis bagi

pengembang sistem dalam merancang mekanisme klasifikasi ulasan secara otomatis pada *platform e-commerce*.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian oleh Utami dan Erfina [9], melakukan analisis sentimen terhadap pinjaman online menggunakan algoritma *Support Vector Machine* (SVM) dengan data dari Twitter. Hasil pengujian menunjukkan akurasi sebesar 62,00%, di mana sentimen negatif mendominasi sebesar 59% dibandingkan sentimen positif sebesar 41%. Penelitian ini mengindikasikan bahwa persepsi negatif terhadap pinjaman online masih tinggi akibat risiko dan kerugian yang dialami pengguna.

Penelitian oleh Pasaribu dan Sriani [10], melakukan analisis sentimen terhadap ulasan pengguna aplikasi Shopee menggunakan algoritma *Naïve Bayes*. Dataset diperoleh dari Google Play Store sebanyak 5.000 data, kemudian setelah proses pembersihan diperoleh 3.946 data yang dibagi menjadi 3.141 data latih dan 805 data uji dengan rasio 8:2. Hasil pengujian menunjukkan akurasi sebesar 86,00%, presisi 80,74%, *recall* 78,13%, dan *F1-score* 79,00%, dengan sentimen positif lebih dominan sebanyak 2.050 ulasan dibandingkan sentimen negatif sebanyak 1.199 ulasan. Penelitian ini menegaskan bahwa jumlah dan kualitas data latih berpengaruh terhadap ketepatan prediksi sentimen.

Penelitian oleh Marcillia *et al.* [11], melakukan analisis sentimen terhadap ulasan penyedia layanan menggunakan algoritma C4.5. Proses penelitian meliputi tahap *text preprocessing*, pembobotan TF-IDF, dan klasifikasi dengan membagi data menjadi data latih dan data uji. Hasil pengujian dengan 98 data latih dan 42 data uji menghasilkan akurasi sebesar 64,29%, *precision* 62,16%, dan *recall* 95,83%. Dengan 112 data latih dan 28 data uji, diperoleh akurasi 71,43%, *precision* 70,83%, dan *recall* 94,44%. Sementara itu, penggunaan 126 data latih dan 14 data uji memberikan akurasi tertinggi sebesar 78,57%, *precision* 83,33%, dan *recall* 90,91%. Penelitian ini menunjukkan bahwa peningkatan jumlah data latih berdampak positif terhadap akurasi klasifikasi sentimen.

2.2. Klasifikasi

Klasifikasi merupakan proses untuk mengelompokkan data masukan ke dalam kelas keluaran yang telah didefinisikan sebelumnya, dengan tujuan agar sistem dapat melakukan prediksi terhadap data baru yang belum diketahui kategorinya [12]. Proses ini umumnya dilakukan menggunakan metode *text mining* dalam kerangka kerja (KDD), yang meliputi tahapan seleksi data, pembersihan dan pra-proses, transformasi data, penerapan algoritma, serta evaluasi hasil. Tahap pra-proses teks mencakup pembersihan karakter tidak relevan, *case folding*, normalisasi, *tokenizing* penghapusan *stopwords*, dan

stemming untuk mengubah kata menjadi bentuk dasarnya.

2.3. TF-IDF

TF-IDF merupakan metode pemberian bobot kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dan kelangkaannya di seluruh dokumen. Teknik ini menurunkan bobot kata umum dan meningkatkan kata yang lebih khas dari dokumen tertentu. Adapun bobot setiap kata dihitung menggunakan rumus TF-IDF sebagai berikut.

a. Term Frequency (TF)

Mengukur seberapa sering suatu kata muncul dalam sebuah dokumen.

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \quad (1)$$

Keterangan:

t : Term atau kata yang dihitung frekuensinya

d : Dokumen tempat term muncul

$f_{t,d}$: Jumlah kemunculan term t dalam dokumen d

$\sum_k f_{k,d}$: Jumlah total seluruh term dalam dokumen d

b. Inverse Document Frequency (IDF)

Menilai tingkat kepentingan suatu term dengan menurunkan bobotnya jika term tersebut sering muncul di berbagai dokumen.

$$IDF(t) = \log \left(\frac{N}{DF(t)} \right) \quad (2)$$

Keterangan:

t : Term atau kata yang dihitung IDF-nya

N : Jumlah keseluruhan dokumen dalam korpus

df_t : Jumlah dokumen yang mengandung term t

$\log$: Fungsi logaritma basis 10 (atau natural log)

c. TF-IDF

Nilai akhir merupakan hasil perkalian antara TF-IDF:

$$TF-IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

2.4. Naïve Bayes

Naïve Bayes adalah metode pembelajaran statistik yang menggunakan teorema Bayes dengan asumsi independensi antar fitur untuk memprediksi probabilitas [13]. Pendekatan ini sangat efektif ketika diaplikasikan pada data teks yang telah direpresentasikan dalam bentuk numerik, seperti TF-IDF.

a. Rumus Naïve Bayes

Rumus dasar dari *Naïve Bayes* berdasarkan teorema Bayes.

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (4)$$

Keterangan:

$P(C_i | X)$: Probabilitas data X termasuk ke dalam kelas C_i

$P(X | C_i)$: Probabilitas data X diberikan kelas C_i

$P(C_i)$: Probabilitas apriori kelas C_i

$P(X)$: Probabilitas data X secara keseluruhan

2.5. ID3

Algoritma ID3 merupakan algoritma pembelajaran yang digunakan untuk membentuk pohon keputusan [14]. ID3 bekerja dengan menyusun pohon keputusan dengan memilih atribut yang memaksimalkan *Information Gain* pada setiap langkah iteratif, hingga seluruh data diklasifikasikan secara sempurna atau atribut telah habis. Adapun rumus yang digunakan berdasarkan *Entropy* dan *Information Gain* sebagai berikut:

a. Entropy

Entropy digunakan untuk mengukur ketidakpastian dalam dataset. Semakin tinggi *entropy*, semakin besar ketidakaturan data.

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2 p_i \quad (5)$$

Keterangan:

S : Sekumpulan data atau himpunan contoh pelatihan

c : Jumlah kelas dalam dataset

p_i : Probabilitas kemunculan kelas ke- i dalam data S

b. Information Gain

Information Gain mengukur pengurangan *entropy* yang dicapai dengan memisahkan data berdasarkan atribut tertentu.

$$Gain(S, A) = Entropy(S) - Entropy_A(S) \quad (6)$$

Keterangan:

$Gain(S, A)$ = nilai information gain atribut A

Atribut dengan nilai gain tertinggi dipilih untuk node pohon Keputusan.

2.6. Evaluasi Model

Evaluasi efektivitas model dilaksanakan dengan memakai ukuran akurasi, presisi, *recall*, dan *F1-score* yang diambil dari *confusion matrix*. Keempat ukuran ini memberikan pandangan yang lebih menyeluruh tentang kinerja model dibandingkan hanya bergantung pada satu ukuran evaluasi.

a. Akurasi

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Akurasi menilai itu seberapa banyak prediksi yang benar dibandingkan dengan keseluruhan data.

b. Presisi

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

Mengukur ketepatan prediksi positif.

c. Recall

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

Mengukur kemampuan model menemukan semua data positif.

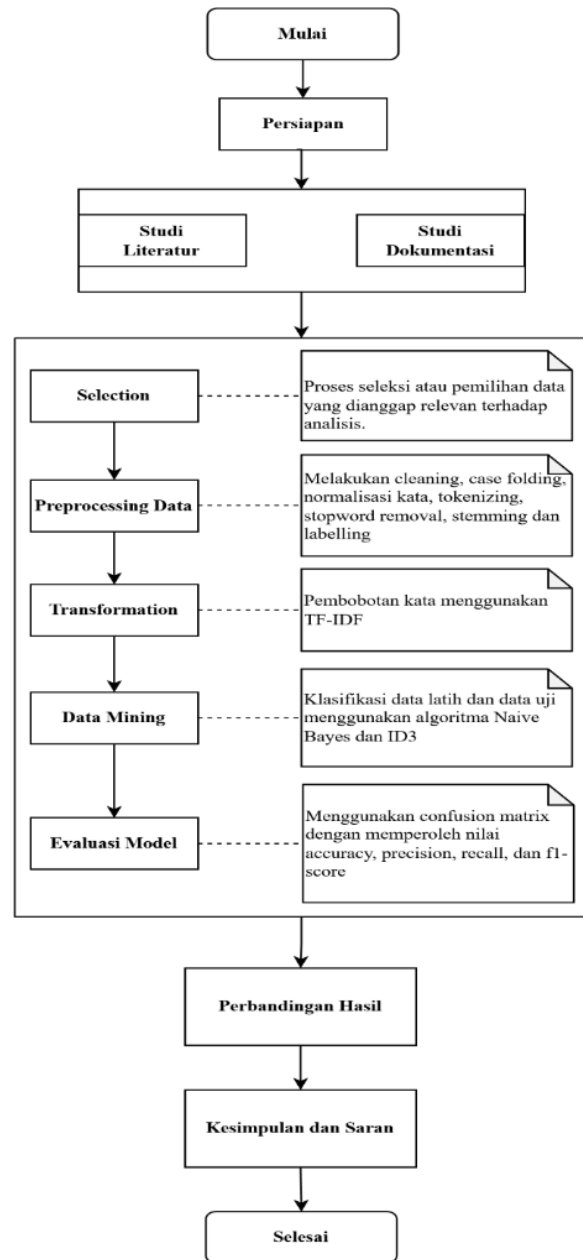
d. F1-Score

$$F1-Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (10)$$

Merupakan rata-rata harmonis antara Precision dan Recall.

3. METODE PENELITIAN

3.1. Tahapan Penelitian



Gambar 1. Tahapan penelitian

Penjelasan:

Gambar 1. Tahapan penelitian, menggambarkan alur sistematis dari perumusan masalah hingga evaluasi hasil. Penelitian dimulai dengan penetapan masalah dan tujuan, dilanjutkan dengan persiapan berupa identifikasi kebutuhan data dan metode. Studi literatur dan dokumentasi dilakukan untuk mengkaji teori serta mengumpulkan data ulasan dari *Kaggle*. Metode KDD diterapkan melalui tahapan seleksi atribut (ulasan, rating, tanggal, user), selanjutnya pra-prosesan teks, serta pelabelan sentimen, transformasi data dari TF-IDF, dan klasifikasi dengan algoritma *Naïve Bayes* serta ID3. Evaluasi dilakukan menggunakan *confusion matrix* dengan metrik

akurasi, presisi, *recall*, dan *F1-score*, lalu dibandingkan untuk menentukan algoritma yang paling optimal.

3.2. Sumber Data

Penelitian ini menggunakan dataset publik dari platform *Kaggle* yang berisi ulasan pengguna aplikasi Shopee di Google Play Store. Dataset mencakup atribut seperti tanggal, nama pengguna, rating, dan isi ulasan. Data dikumpulkan pada 15 Juni 2025 dengan periode ulasan antara 11 Juli 2022 hingga 13 April 2023. Jumlah data awal sebanyak 2.500 entri, yang selanjutnya digunakan sebagai dasar analisis dan klasifikasi sentimen.

3.2. Teknik Analisis Data

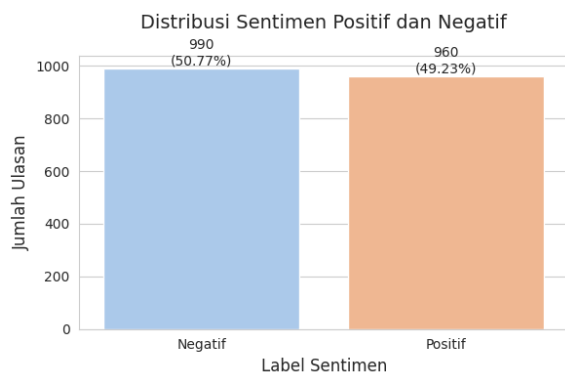
Analisis data pada penelitian ini dilakukan mengikuti tahapan proses data mining yang meliputi *preprocessing*, pelabelan, ekstraksi fitur, klasifikasi, dan evaluasi. Tahap *preprocessing* meliputi *cleaning*, *case folding*, *tokenizing*, *stopword removal*, normalisasi, dan *stemming* untuk menghasilkan teks yang bersih dan seragam.

Pelabelan sentimen dilakukan dengan pendekatan *rating-based*, yaitu ulasan dengan rating lebih dari 3 dikategorikan positif dan rating kurang dari 3 dikategorikan negatif, sedangkan rating 3 dikategorikan netral dan dikeluarkan dari analisis. Dari 2.500 data awal, setelah pembersihan dan penghapusan data kosong diperoleh 1.950 ulasan, terdiri atas 990 ulasan negatif dan 960 ulasan positif.

Representasi teks dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengubah teks menjadi vektor numerik yang dapat diproses oleh algoritma klasifikasi. Setelah itu, dataset dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Dataset



Gambar 2. Distribusi pembagian dataset

Dataset ulasan pengguna aplikasi Shopee yang diperoleh dari platform *Kaggle* telah melalui tahapan pra-pemrosesan. Setelah proses tersebut, data yang awalnya berjumlah 2.500 ulasan menyusut menjadi 1.950 ulasan bersih yang siap dianalisis. Proses pelabelan dilakukan dengan pendekatan *rating-based*,

di mana ulasan dengan rating lebih dari tiga dikategorikan sebagai positif, sedangkan ulasan dengan rating kurang dari tiga dikategorikan sebagai negatif. Berdasarkan hasil pelabelan, diperoleh 990 ulasan negatif dan 960 ulasan positif. Distribusi sentimen ini ditunjukkan pada Gambar 2.

4.2. Hasil Ekstraksi Fitur TF-IDF

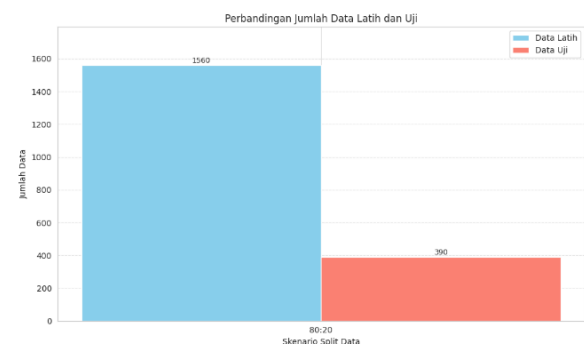
Tahap ekstraksi fitur dilakukan dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk merepresentasikan teks ulasan menjadi bentuk vektor numerik. Analisis bobot rata-rata menunjukkan sepuluh kata dengan nilai tertinggi yang menggambarkan kata-kata paling dominan dalam ulasan pengguna Shopee (Tabel 1).

Tabel 1. Beberapa hasil ekstraksi fitur

Rank	Term	Average TF-IDF
1	bagus	0.048627
2	shopee	0.040500
3	lambat	0.037534
4	aplikasi	0.032372
5	mantap	0.027159
6	ok	0.022736
7	banget	0.021419
8	belanja	0.016261
9	bantu	0.014971
10	baik	0.014185

4.3. Hasil Klasifikasi

Tahap klasifikasi dilakukan dengan membangun model prediktif menggunakan representasi fitur dari TF-IDF dan label sentimen positif-negatif yang diperoleh melalui pendekatan *rating-based*. Dataset dibagi menjadi dua bagian dengan skema 80:20, yaitu 1.560 data untuk pelatihan dan 390 data untuk pengujian dari total 1.950 ulasan. Skema ini bertujuan untuk memastikan model dapat dilatih secara optimal sekaligus dievaluasi menggunakan data yang tidak terlibat dalam proses pelatihan.



Gambar 3. Visualisasi pembagian dataset

4.4. Hasil Evaluasi Model

Kinerja model *Naïve Bayes* dan *ID3* diukur menggunakan *accuracy*, *precision*, *recall*, dan *F1-score* pada masing-masing kelas sentimen. Hasil evaluasi ditunjukkan pada Tabel 2 berikut.

Tabel 2. Hasil evaluasi performa secara keseluruhan

Matrix Evaluasi	Naïve Bayes	ID3
Akurasi	83.07%	73.85%
Presisi Positif	87%	71%
Presisi Negatif	80%	77%
Recall Positif	77%	79%
Recall Negatif	88%	69%
F1-Score Positif	82%	75%
F1-Score Negatif	84%	73%

Berdasarkan hasil pada Tabel 2, algoritma *Naïve Bayes* menunjukkan kinerja yang lebih unggul secara keseluruhan dibandingkan ID3 dengan akurasi 83,07% dan *F1-score* yang konsisten tinggi pada kedua kelas sentimen, yaitu 82% untuk kelas positif dan 84% untuk kelas negatif. *Naïve Bayes* juga memiliki presisi positif (87%) dan presisi negatif (80%) yang lebih tinggi dibandingkan ID3, yang masing-masing hanya mencapai 71% dan 77%, menunjukkan kemampuannya dalam menghasilkan prediksi yang lebih tepat. Meskipun demikian, ID3 mencatat *recall* positif yang lebih tinggi (79%) dibandingkan *Naïve Bayes* (77%), yang berarti ID3 lebih sensitif dalam mendeteksi ulasan positif, namun performanya pada *recall* negatif (69%) tertinggal jauh dibandingkan *Naïve Bayes* (88%). Hasil ini mengindikasikan bahwa *Naïve Bayes* lebih efektif dalam menjaga keseimbangan antara presisi dan *recall* pada kedua kelas, sehingga lebih andal dalam mengklasifikasikan sentimen secara menyeluruh.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang dilakukan terhadap 1.950 ulasan pengguna aplikasi Shopee, proses klasifikasi sentimen dilakukan menggunakan metode *Knowledge Discovery in Databases* (KDD) yang meliputi tahapan pemahaman data, pembersihan data, transformasi data, penerapan model, dan evaluasi hasil. Fitur TF-IDF digunakan untuk merepresentasikan teks menjadi vektor numerik, dengan pembagian data 80:20 yaitu 1.560 data untuk pelatihan dan 390 data untuk pengujian. Hasil evaluasi menunjukkan bahwa algoritma *Naïve Bayes* memiliki kinerja yang lebih baik dibandingkan ID3 dengan nilai akurasi 83,07%, presisi positif 87%, presisi negatif 80%, *recall* positif 77%, *recall* negatif 88%, *F1-score* positif 82%, dan *F1-score* negatif 84%. Sementara itu, algoritma ID3 memperoleh akurasi 73,85%, presisi positif 71%, presisi negatif 77%, *recall* positif 79%, *recall* negatif 69%, *F1-score* positif 75%, dan *F1-score* negatif 73%. Temuan ini menunjukkan bahwa *Naïve Bayes* lebih efektif dalam klasifikasi sentimen ulasan pengguna aplikasi Shopee dibandingkan ID3 pada dataset yang digunakan.

Penelitian selanjutnya disarankan untuk menguji algoritma atau metode *ensemble* sebagai pembanding, menambahkan kelas sentimen netral agar analisis lebih kaya, serta menguji model menggunakan data real-time untuk mengukur stabilitas performa. Penggunaan uji statistik juga direkomendasikan guna memastikan signifikansi perbedaan hasil antar algoritma. Selain itu,

pengembangan model yang mempertimbangkan manfaat aplikatif, seperti integrasi ke dalam sistem *monitoring* ulasan atau sistem rekomendasi *e-commerce*, dapat meningkatkan nilai praktis dari penelitian ini.

DAFTAR PUSTAKA

- [1] M. Lanteng, "Dampak E-commerce bagi Perekonomian Indonesia," *Kementerian Keuangan Direktorat Jenderal Keuangan*, 2024. <https://www.djkn.kemenkeu.go.id/kpknl-parepare/baca-artikel/17140/Dampak-E-commerce-bagi-Perekonomian-Indonesia.html> (accessed Jul. 02, 2025).
- [2] Tempo.co, "Jejak Persaingan Shopee dengan Tokopedia, Siapa Penguasa Pasar E-Commerce RI Saat Ini?," 2024. <https://www.tempo.co/ekonomi/jejak-persaingan-shopee-dengan-tokopedia-siapa-penguasa-pasar-e-commerce-ri-saat-ini--14288> (accessed May 07, 2025).
- [3] E. Fenton, "Online Reputation Monitoring: How to Monitor Customer Reviews," *Visualping*, 2022. <https://visualping.io/blog/online-reputation-monitoring> (accessed Apr. 27, 2025).
- [4] R. Karisma, S. Lestanti, and M. T. Chulkamdi, "Aplikasi Klasifikasi Sentimen Pada Ulasan Smartphone Di Situs Jual Beli Online Berbasis Web Menggunakan Naive Bayes Dengan Tf-Idf," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 1, pp. 31–37, 2021, doi: 10.36040/jati.v6i1.4365.
- [5] L. Abdillah Fudholi, N. Rahaningsih, and R. Danar Dana, "Sentimen Analisis Perilaku Penggemar Coldplay Di Media Sosial Twitter Menggunakan Metode Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 4150–4159, 2024, doi: 10.36040/jati.v8i3.9827.
- [6] M. Raffi, A. Suharso, and I. Maulana, "Analisis Sentimen Ulasan Aplikasi Binar Pada Google Play Store Menggunakan Algoritma Naïve Bayes Sentiment Analysis of Binar Application Reviews on Google Play Store Using Naïve Bayes Algorithm," *J. Inf. Technol. Comput. Sci.*, vol. 6, no. 1, pp. 1–7, 2023.
- [7] S. Widaningsih, W. Muhamad, R. Hendriyanto, and H. Nugroho, "An ID3 Decision Tree Algorithm-Based Model for Predicting Student Performance Using Comprehensive Student Selection Data at Telkom University," *Ing. des Syst. d'Information*, vol. 28, no. 5, pp. 1205–1212, 2023, doi: 10.18280/isi.280508.
- [8] S. Yunita and Nurahman, "METODE KLASIFIKASI DATA MINING ALGORITMA C4.5 DAN PART UNTUK PREDIKSI WAKTU KELULUSAN MAHASISWA DI UNIVERSITAS DARWAN ALI," *J. Inform. dan Teknol. Inf.*, vol. 1, no. 2, pp. 98–103, 2022.
- [9] D. S. Utami and A. Erfina, "Analisis Sentimen Pinjaman Online Di Twitter Menggunakan Algoritma Support Vector Machine (Svm),"

- SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform., vol. 1, pp. 299–305, 2021.
- [10] N. A. Pasaribu and Sriani, “The Shopee Application User Reviews Sentiment Analysis Employing Naïve Bayes Algorithm,” *Int. J. Softw. Eng. Comput. Sci.*, vol. 3, no. 3, pp. 194–204, 2023, doi: 10.35870/ijsecs.v3i3.1699.
- [11] Miske Marcillia, Muhammad Fadlan, and L. Hartono, “Analisis Sentimen pada Ulasan Penyedia Layanan Menggunakan Algoritma C4.5,” *J. Big Data Anal. Artif. Intell.*, vol. 6, no. 2, pp. 1–5, 2023, doi: 10.71302/jbidai.v6i2.45.
- [12] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to Data Mining: Instructor’s Solution Manual*, 2nd ed. Boston: Pearson Education, Inc., 2023.
- [13] Heliyanti Susana, “Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet,” *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.
- [14] F. Maisa Hana, “Penerapan Algoritma ID3 Decision Tree Pada Klasifikasi Penyakit Diabetes,” *J. Din. Inform.*, vol. 12, no. 2, pp. 1–11, 2023.