Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893
p-ISSN: 2089-3787

Penerapan Algoritma *Naive Bayes* untuk Perbandingan Sentimen Ulasan *Lazada* dan *Tokopedia*

Rafli Prasetya[1*], April Lia Hananto[2], Elfina Novalia[3], Tukino [4]
Sistem Informasi, Universitas Buana Perjuangan, Karawang, Indonesia
*e-mail *Corresponding Author.* si21.rafliprasetya@mhs.ubpkarawang.ac.id

Abstract

The e-commerce sector is one of many aspects of life influenced by advancements in information and communication technology. Tokopedia and Lazada, as two popular platforms in Indonesia, are increasingly accessed through mobile applications. User reviews serve as valuable input for improving service quality and user satisfaction. This study aims to evaluate user sentiment toward the Lazada and Tokopedia applications on the Google Play Store using the Naïve Bayes algorithm. A total of 2,000 review data were collected using web scraping methods and underwent preprocessing, resulting in 1,864 data points ready for analysis. The Hold-Out technique was applied for data splitting to assess model performance. The results show an accuracy of 89% for both applications. The majority of user sentiment is positive, with Lazada achieving a precision of 94%, recall of 93%, and F1-score of 94%, while Tokopedia achieved a precision of 97%, recall of 86%, and F1-score of 91%. These findings demonstrate the effectiveness of combining the Naïve Bayes algorithm and the Hold-Out technique in sentiment classification.
Keywords: *E-commerce; Sentiment Analysis; Google Play Store; Naïve Bayes*

Abstrak

Sektor *e-commerce* adalah salah satu dari banyak aspek kehidupan yang dipengaruhi oleh kemajuan teknologi informasi dan komunikasi. Tokopedia dan Lazada sebagai dua *platform* populer di Indonesia, semakin sering digunakan melalui aplikasi mobile. Ulasan pengguna menjadi masukan penting dalam meningkatkan kualitas layanan dan kepuasan pengguna. Penelitian ini bertujuan untuk mengevaluasi sentimen terhadap pengguna aplikasi Lazada dan Tokopedia di *platform Google Play Store* melalui algoritma *Naïve Bayes*. Sebanyak 2000 data ulasan dikumpulkan menggunakan metode web scraping, kemudian dilakukan *preprocessing* sehingga diperoleh 1864 data siap analisis. Teknik *Hold-Out* digunakan dalam pembagian data untuk mengukur performa model. Hasil menunjukkan akurasi sebesar 89% untuk masing-masing aplikasi. Sentimen pengguna mayoritas bersifat positif, dengan Lazada memperoleh presisi 94%, *recall* 93%, dan skor F1 94%, sedangkan Tokopedia memperoleh presisi 97%, *recall* 86%, dan skor F1 91%. Temuan ini menunjukkan efektivitas kombinasi algoritma *Naïve Bayes* dan *Hold-Out* dalam klasifikasi sentimen.
Kata kunci: *E-commerce; Analisis Sentimen; Google Play Store; Naïve Bayes*

1. Pendahuluan

Berbagai aspek kehidupan manusia telah berubah oleh kemajuan teknologi informasi dan komunikasi, termasuk cara kita berbelanja. Perdagangan elektronik atau *e-commerce* telah menjadi salah satu sektor yang mengalami pertumbuhan pesat dalam beberapa tahun terakhir. Dua di antara *platform e-commerce* yang populer di Indonesia adalah Lazada dan Tokopedia. Dengan semakin meningkatnya pengguna *smartphone*, aplikasi mobile seperti Lazada dan Tokopedia semakin banyak diunduh dan digunakan oleh konsumen untuk melakukan berbagai transaksi belanja [1]. Aplikasi ini tersedia di banyak *platform*, salah satunya adalah *Google Play Store*, yang memungkinkan pengguna Android untuk mengunduh dan memberikan ulasan terhadap aplikasi-aplikasi tersebut [2].

E-commerce Tokopedia dan Lazada diambil dalam penelitian ini didasarkan pada posisi keduanya sebagai *platform e-commerce* terkemuka di Indonesia, dengan kualitas layanan dan jumlah pengguna yang sangat besar [3]. Namun, mengingat jumlah ulasan yang sangat banyak, diperlukan metode yang efisien guna mengidentifikasi tren opini yang tercermin dalam kumpulan

ulasan tersebut. Ini sangat berharga bagi pengembang aplikasi karena dapat memberikan masukan yang konstruktif terhadap aplikasi [4]. Maka permasalahan ini diambil karena dengan meningkatnya daya beli masyarakat yang cenderung beralih pada analisis belanja *online*, sentimen terhadap ulasan pengguna menjadi sangat penting untuk memahami preferensi konsumen dan mendukung pengembangan layanan aplikasi *e-commerce* yang lebih baik [5].

Analisis sentimen adalah teknik dalam menerjemahkan bahasa yang bertujuan mendeteksi serta mengelompokkan pendapat yang terdapat dalam teks, terutama untuk menilai apakah sentimen untuk menilai orientasi sikap yang diungkapkan dalam teks, apakah cenderung positif, negatif, atau bersifat netral [6]. Untuk kebutuhan analisis, penggunaan algoritma *Naïve Bayes* dinilai sangat tepat, berkat kemudahan dan efisiensinya dalam mengolah data teks, serta kemampuannya memberikan hasil yang cukup akurat [7]. Algoritma *Naive Bayes* dipilih karena memiliki kecepatan komputasi yang tinggi, mudah diimplementasikan, dan cocok untuk analisis teks seperti ulasan pengguna di *platform e-commerce* [8]. Hal ini menjadikannya metode yang relevan dan andal untuk penelitian ini. Selain itu, kemampuannya yang sederhana namun efektif dalam menangani teks data, termasuk ulasan pengguna. Algoritma ini bekerja berdasarkan prinsip probabilitas *Bayes* [9], yang memungkinkan pengklasifikasian sentimen dengan akurasi tinggi meskipun data bersifat besar dan kompleks [10]. Melalui pendekatan ini, dapat diperoleh Menghasilkan insight yang lebih luas mengenai bagaimana pengguna merasakan dan menilai kedua aplikasi tersebut, sekaligus mampu memberikan informasi yang berguna bagi penggunanya [11].

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi *e-commerce* Tokopedia dan Lazada yang tersedia di *Google Play Store* dengan memanfaatkan algoritma *Naïve Bayes*. Hasil dari penelitian ini diharapkan dapat memberikan gambaran yang komprehensif. Mengenai kecenderungan sentimen pengguna, baik positif maupun negatif, terhadap kedua aplikasi tersebut. Temuan ini diharapkan juga dapat menjadi acuan berharga bagi pengembang aplikasi dalam meningkatkan kualitas layanan serta pengalaman pengguna secara keseluruhan.

2. Tinjauan Pustaka

Kajian tentang analisis menggunakan *Naïve Bayes* dalam lulasan sentimen pada *market Place* memperlihatkan variasi dalam tingkat akurasi dan performa model. Penelitian ini menganalisis sentimen ulasan aplikasi Shopee di *Google Play Store*. Model klasifikasi dibangun menggunakan *Multinomial Naïve Bayes*, dengan data dibagi dengan *Hold-Out* (80% pelatihan, 20% pengujian) serta *K-Fold Cross Validation* (K=10). Hasilnya bahwa metode *Hold-Out* memberikan akurasi terbaik sebesar 83%, lebih unggul dibandingkan teknik *K-Fold* yang mencapai 82%. Ini cenderung pusitif dengan tingkat akurasi mencapai 82%, presisi 83%, *recall* 100%, serta nilai *f1* sebesar 91% [12].

Sementara itu, penelitian oleh Artanti Inez Tanggraeni yang meneliti analisis sentimen "Sentuh Tanahku" di *Google Play*. Proses penelitian diawali dengan pengambilan data ulasan diikuti dengan pelabelan manual untuk mengidentifikasi ulasan positif dan negatif, kemudian data diolah yang selanjutnya analisis dilakukan dengan mengklasifikasikan data menggunakan algoritma *Naïve Bayes* dan evaluasi model dilakukan dengan menghitung parameter dengan akurasi 89%, presisi 83%, dan *recall* 87% [13].

Penelitian oleh Bintang Zulfikar Ramadhan meneliti analisis sentimen ulasan pada aplikasi *e-commerce*. Langkah yang digunakan mengikuti kerangka *Knowledge Discovery in Database* (KDD) serta klasifikasi sentimen melalui pembagian data ke dalam tiga skenario (80:20, 70:30, dan 60:40). Parameter evaluasi yang dianalisis mencakup akurasi, presisi, *recall*, dan skor f1, dengan hasil rata-rata akurasi 85%, presisi 84%, *recall* 96%, dan f1 91% [14].

Penelitian oleh Siti Nur Fadhilah meneliti ulasan aplikasi Blibli. Tahapan yang digunakan meliputi pengumpulan data melalui *Python*, pelabelan data secara otomatis, serta rangkaian *pre-processing* diikuti dengan TF-IDF untuk mengonversi teks menjadi representasi numerik sebelum akhirnya dilakukan klasifikasi. Evaluasi kinerja model dilakukan dengan menghitung parameter akurasi keseluruhan serta metrik yang mampu mencapai akurasi 84%, presisi 79%, *recall* 95%, dan skor f1 86% [15].

Penelitian yang dilakukan oleh Muhammad Subhan Mahendrasyah berfokus pada sentimen pengguna aplikasi Bukalapak. Data dikumpulkan melalui teknik *web scraping* dengan total 5000 ulasan, kemudian melalui tahap *preprocessing* dan sebelum dilakukan pembobotan dengan TF-IDF. Model klasifikasi dibangun menggunakan *Naïve Bayes*, dengan data dibagi

untuk pelatihan dan pengujian guna mengevaluasi performa algoritma. Hasil penelitian menunjukkan bahwa akurasi mencapai sebesar 67% *recall* 67% dan presisi 69% [16].

Penelitian ini menghadirkan kebaruan dengan menghadirkan dua aplikasi *e-commerce* populer, Tokopedia dan Lazada, secara bersamaan, berbeda dari studi sebelumnya yang hanya fokus pada satu aplikasi. Selain itu, metode pengumpulan data dilakukan secara otomatis melalui *web scraping*, dilengkapi dengan *preprocessing* dan evaluasi menggunakan teknik *Hold-Out*. Pendekatan ini memberikan kontribusi baru dalam meningkatkan akurasi dan efektivitas klasifikasi sentimen, serta memperluas cakupan analisis terhadap persepsi pengguna aplikasi *e-commerce* di Indonesia.

3. Metodologi

3.1 Pemilihan Data

Data ulasan pengguna aplikasi *e-commerce* Lazada dan Tokopedia dipilih dari *platform Google Play Store*. Ini dilakukan dengan memastikan bahwa ulasan yang diambil relevan dengan tujuan penelitian yakni untuk analisis *sentiment* [17].

3.2 *Scrapping Data*

Scraping Data adalah proses otomatisasi pengambilan data dari halaman *web* [18]. Dalam konteks ini, Peneliti akan melakukan *web scraping* dengan *phyton*, sebagai Bahasa Pemrograman yang dipakai untuk mengumpulkan 2000 ulasan untuk aplikasi Lazada dan Tokopedia.

3.3 *Preprocessing Data*

Melakukan *Text Preprocessing* dengan tindakan ini dimaksudkan untuk meniadakan ketidakteraturan dalam data teks serta mengaturnya sedemikian rupa sehingga siap dipergunakan dalam proses analisis [19]. Setelah itu, sentiment analisis akan dilakukan untuk mengekstrak sentimen setiap ulasan pengguna. Kegiatan ini mencakup sebagai berikut.

1) *Case Folding*

   Proses *folding case* berfungsi mengonversi semua karakter alfabet ke dalam bentuk huruf kecil [20].

2) *Stopword*

   Proses *folding case* berfungsi mengonversi semua karakter alfabet ke dalam bentuk huruf kecil.

3) *Tokenization*

   Teks dihapus dari tanda baca dan dibagi menjadi bagian yang lebih kecil berdasarkan spasi karakter selama proses tokenisasi.

4) *Stemming*

   Proses *stemming* menemukan kata dasar dan menghilangkan semua imbuhan. Tujuan *stemming* adalah untuk menghilangkan kata-kata dengan pengejaan yang buruk. Algoritma *stemming* yang digunakan dalam setiap bahasa berbeda. Selain itu, tahapan ini digunakan untuk meningkatkan keseragaman data dan menurunkan jumlah kata yang ada [21].

3.4 Pembobotan Data TF-IDF

Pembobotan Nilai TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan indikator yang menentukan relevansi kata pada data dengan cara menilai seberapa sering kata tersebut muncul [22]. TF (*Term Frekuensi*) menghitung pada jumlah kali kata ditampilkan dalam dokumen, sedangkan IDF (*Inverse Document Frekuensi*) Kata-kata yang muncul secara terulang akan diberi bobot yang lebih rendah di berbagai dokumen [23].

3.5 Pembagian Data

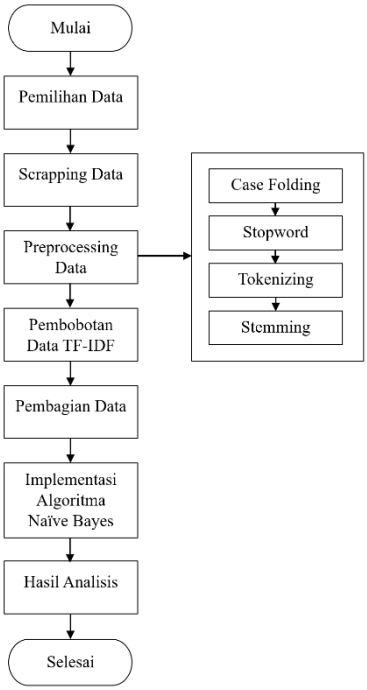
Pembagian data yang dilakukan dalam penelitian ini untuk memisahkan dataset menjadi dua bagian utama, yaitu pelatihan data dan data pengujian data. Data pelatihan untuk membangun dan melatih model, sedangkan data pengujian untuk memulai performa model dalam mengklasifikasi [24].

3.6 *Multinominal Naïve Bayes*

Salah satu bentuk turunan dari algoritma *Naïve Bayes* dimanfaatkan pengklasifikasian data berbasis teks adalah *Multinomial Naïve Bayes* [25]. Ini menggunakan probabilitas *Bayes*, yang menentukan kemungkinan bahwa data akan dimasukkan dalam kategori tertentu berdasarkan distribusi kata dalam dokumen. Dengan mengasumsikan bahwa kata-kata dalam dokumen bersifat independen, sehingga memungkinkan analisis data teks dengan efisien [26].

$$P(c|d) \propto P(c) \prod_{i=1}^{n_d} P(w_I |c) \quad \text{..... (1)}$$

Dimana:
P(c) = Kemungkinan awal dari kelas c
P(c|d) = Kemungkinan sebuah dokumen atau teks d termasuk ke dalam kelas c
$P(w_I|c)$ = kemungkinan munculnya kata wi dalam kelas c



Gambar 1. Pendekatan penelitian

4. Hasil dan Pembahasan

4.1 Pemilihan Data

Pada tahap ini, ulasannya data *e-commerce* Lazada dan Tokopedia mencangkup tanggal unggahan, peringkat, dan ulasan teks diambil melalui *platform Google Play Store*. Ini dilakukan dengan memastikan bahwa ulasan yang diambil relevan dengan tujuan penelitian yakni untuk analisis *sentiment*.

4.2 *Scrapping Data*

Untuk mengumpulkan 2000 ulasan terbaru dari aplikasi Lazada dan Tokopedia di *Google Play Store*, proses *web scraping* dilakukan dengan memanfaatkan bahasa pemrograman *Python* melalui *platform Google Colaboratory*. Untuk melakukan ini, *library* "*google-play-scraper*" digunakan, yang memungkinkan pengambilan data ulasan memasukkan ID aplikasi dari *link* aplikasi. Hasil *scraping* diekspor dalam format *Microsoft Excel* (.xlsx) setelah data dikumpulkan. Dokumen berikut adalah contoh yang dihasilkan dari proses *scraping web* tersebut.

Tabel 1. *Scrapping*

Dokumen	Ulasan Lazada	Ulasan Tokopedia
Dok. 1	Suka belanja di lazada karna mudah dan bayak diskon	Harga barangnya murah dan berkualitas premium
Dok. 2	sering belanja	Saya Pengguna baru, Semoga Tokped tidak mengecewakan
Dok. 3	bagus, sangat membantu konsumen	bagus
…	…	…
Dok. x	Bagus dan terpercaya	bagus banyak diskon

4.3 *Preprocessing Data*

Sebelum algoritma klasifikasi dapat diterapkan pada dokumen harus melakukan *Preprocessing* data terlebih dahulu, tahapan *preprocessing* sangat penting untuk masuk ke tahap selanjutnya. Tujuannya Untuk membersihkan dan menyusun ulang data teks agar dapat diterapkan dalam analisis, berikut adalah langkahnya.

1) *Case Folding*

Proses ini membuat huruf menjadi kecil, yang membuat teks menjadi sama. Pada proses ini, kata-kata juga dibersihkan dari tanda baca atau karakter-karakter tertentu yang tidak diperlukan, berikut Tabel 2.

Tabel 2. *Folding Case*

Lazada	Tokopedia
suka belanja di lazada karna mudah dan bayak diskon	harga barangnya murah dan berkualitas premium

2) *Stopword*

Kata-kata yang sangat umum dan sering kali tidak membawa arti signifikan itu akan dihapus dengan mengeliminasi kata-kata dari teks, kita bisa lebih memusatkan perhatian pada istilah yang benar bermakna. Hasilnya pada tabel di bawah ini.

Tabel 3. *Stopword*

Lazada	Tokopedia
suka belanja lazada karna mudah bayak diskon	harga barangnya murah berkualitas premium

3) *Tokenization*

Dengan *tokenizing*, Teks dihapus dari tanda baca dan dibagi menjadi bagian yang lebih kecil berdasarkan spasi karakter selama proses tokenisasi. Berikut.

Tabel 4. *Tokenization*

Lazada	Tokopedia
['suka', 'belanja', 'lazada', 'karna', 'mudah', 'bayak', 'diskon']	['harga', 'barangnya', 'murah', 'berkualitas', 'premium']

4) *Stemming*

Untuk menghilangkan kata-kata dengan pengejaan yang buruk, proses *stemming* menemukan kata dasar dan menghilangkan semua imbuhan. *stemming* yang digunakan dalam setiap bahasa berbeda. Selain itu, penggunaan ini juga meningkatkan keseragaman data dan menurunkan jumlah kata yang ada.

Tabel 5. *Stremming*

Lazada	Tokopedia
suka belanja lazada karna mudah bayak diskon	harga barang murah kualitas premium

4.4 Pembobotan Data TF-IDF

4.4.1 *Term Frequency* (TF)

Metode ini bertujuan untuk menghitung jumlah kemunculan kata dalam suatu *dataset*. Perhitungan bobot TF dapat dilakukan menggunakan rumus berik:

$$tf = 0.5 + 0.5 \frac{tf}{max\,(tf)} \quad \text{..... (2)}$$

Dimana:
tf = Banyaknya yang muncul
max(tf) = panjangnya kata

Sebagai gambaran, berikut perhitungan nilai TF yang didasarkan pada dokumen-1:

Tabel 6. TF

Lazada	Tokopedia
['suka', 'belanja', 'lazada', 'karna', 'mudah', 'bayak', 'diskon']	['harga', 'barangnya', 'murah', 'berkualitas', 'premium']

4.4.2 *Inverse Document Frequency* (IDF)

Dirancang mengurangi bobotnya pada kata yang memiliki frekuensi kemunculan tinggi, sehingga lebih menonjolkan kata-kata yang spesifik dan relevan dengan konteks dokumen tersebut. Berikut persamaannya.

$$idf = ln \frac{N}{df} + 1 \quad \text{..... (3)}$$

Dimana:
ln = logaritma natural
N = total keseluruhan
df = jumlah kemunculan kata

Penghitungan nilai untuk dokumen-1 dilakukan dengan mengumpulkan dan menggabungkan seluruh kata ke dalam satu *dataset*. Perhitungan bobot dilakukan pada langkah ini dengan mengombinasikan nilai TF dan IDF melalui operasi perkalian, sehingga ditampilkan.

Tabel 7. Perkalian

Aplikasi	Dokumen	TF-IDF	Label
Lazada	[‘suka’]	[3.2050971225120364]	Positif
Tokopedia	[‘harga’]	[3.2744456760230114]	Positif

4.5 Pembagian data

Metode *Hold-out* pembagian *dataset* dilakukan dengan komposisi 80% untuk pelatihan algoritma dan 20% untuk penguji performa prediksi. Pemrograman *Python* digunakan dalam pembagian data yang dilakukan dalam penelitian ini.

Evaluasi terhadap model yang diterapkan pada data pelatihan dan pengujian dilakukan dengan menggunakan *matrix confusion* berukuran 4 kali 3:

Tabel 8. Hasil *Confusion matrix*

Aplikasi	Data Prediksi	Data Aktual	
		Positif	Negatif
Lazada	Positif	26	9
	Negatif	10	143
Tokopedia	Positif	65	3
	Negatif	16	101

4.6 *Multinominal Naïve Bayes*

Performa *Multinominal Naïve Bayes* dengan metode pembagian data *Hold-Out* menghasilkan akurasi sebesar 89% untuk Lazada dan 89% Tokopedia, dengan nilai presisi, *recall*, dan skor f1 yang ditampilkan pada Gambar 2.

(Lazada)

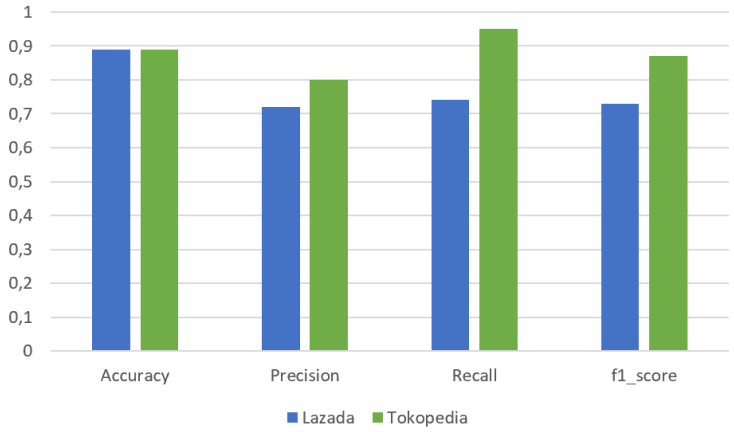
```
MultinomialNB Accuracy: 0.898936170212766
MultinomialNB Precision: 0.7222222222222222
MultinomialNB Recall: 0.7428571428571429
MultinomialNB f1_score: 0.7323943661971831
```

(Tokopedia)

```
MultinomialNB Accuracy: 0.8972972972972973
MultinomialNB Precision: 0.8024691358024691
MultinomialNB Recall: 0.9558823529411765
MultinomialNB f1_score: 0.87248322147651
```

Gambar 2. Hold-Out Lazada dan Tokopedia

Visualisasi pada Gambar 2 menyajikan hasil evaluasi performa model melalui empat utama yaitu akurasi, presisi, sensitivitas (*recall*), serta skor f1 kedua aplikasi tersebut.



Gambar 3. Akurasi

Dari gambar yang ditampilkan, diketahui bahwa metode *Hold-Out* memberikan hasil akurasi baik sebesar 89% untuk lazada, sementara Tokopedia juga menghasilkan 89%. Adapun kecenderungan sentimen ulasan kedua aplikasi melalui metode *Hold-Out* berikut.

(Lazada)

	precision	recall	f1-score	support
Negatif	0.72	0.74	0.73	35
Positif	0.94	0.93	0.94	153
accuracy			0.90	188
macro avg	0.83	0.84	0.84	188
weighted avg	0.90	0.90	0.90	188

(Tokopedia)

	precision	recall	f1-score	support
Negatif	0.80	0.96	0.87	68
Positif	0.97	0.86	0.91	117
accuracy			0.90	185
macro avg	0.89	0.91	0.89	185
weighted avg	0.91	0.90	0.90	185

Gambar 4. Hasil dari Sentimen Lazada dan Tokopedia

Seperti yang ditunjukkan oleh gambar, Lazada menghasilkan presisi 94%, *recall* sebesar 93%, yang juga berada di angka 94% f1 dan Tokopedia nilainya sebesar 97% presisi, diangka 86% *recall*, yang juga berada di angka 91% f1, ini menunjukkan bahwa pendapat pengguna kedua aplikasi *e-commerse* tersebut di *Google Play Store* cenderung positif.

4.7 Pembahasan

Hasil pengujian dalam penelitian ini menunjukkan bahwa algoritma *Naïve Bayes* yang diterapkan pada ulasan pengguna aplikasi *e-commerce* Tokopedia dan Lazada memberikan performa yang cukup baik, dengan akurasi masing-masing sebesar 89%. Nilai lainnya juga menunjukkan performa yang kuat, di mana Lazada menghasilkan presisi 94%, *recall* sebesar 93%, yang juga berada di angka 94% f1 dan Tokopedia nilainya sebesar 97% presisi, diangka 86% *recall*, yang juga berada di angka 91% f1. Temuan ini memperkuat hasil penelitian Siti Nur Fadhilah [15], sebelumnya yang menggunakan *Naïve Bayes* untuk aplikasi Blibli dan menghasilkan akurasi sebesar 84%, serta metrik yang mampu mencapai presisi 79%, *recall* 95%, dan skor f1 86%, menunjukkan bahwa *Naïve Bayes* tetap relevan dalam pengolahan teks ulasan pengguna.

Lebih lanjut, penelitian yang dilakukan dengan menerapkan pendekatan KDD dan skenario pembagian data berbeda, hasil penelitian ini memperlihatkan kinerja yang sebanding dalam hal akurasi dan nilai evaluasi lainnya. Dengan hanya menggunakan teknik *Hold-Out*, penelitian ini telah mencapai hasil yang baik bahkan lebih tinggi dalam beberapa aspek. Hal ini menunjukkan bahwa, algoritma *Naïve Bayes* tetap mampu memberikan performa yang stabil.

Dengan demikian, penelitian ini memberikan kontribusi nyata terhadap penguatan bukti bahwa *Naïve Bayes* merupakan algoritma klasifikasi teks yang efisien, khususnya dalam analisis pada *e-commerce*. Selain itu, dengan menampilkan dua *platform* sekaligus dalam satu model eksperimen, pendekatan ini dapat menjadi landasan bagi studi lanjutan yang ingin mengevaluasi performa model pada lebih dari satu aplikasi secara bersamaan, serta memperluas cakupan dan generalisasi model dalam klasifikasi sentimen pengguna.

5. Simpulan

Kesimpulan berdasarkan analisis yang telah dibahas, termasuk bahwa sebanyak 2000 ulasan dikumpulkan melalui metode *web scrapping* di *Google Play Store*. Selanjutnya, proses *preprocessing* yang digunakan menghasilkan total 1864 data bersih. Dengan penerapan teknik pembagian data *Hold-Out*, Algoritma *Naive Bayes* berhasil memberikan capaian akurasi yang mengesankan 89% untuk Lazada dan Tokopedia sebesar 89%. Hasil pengujian mengindikasikan bahwa metode *Hold-Out* memberikan tingkat bagus. Dari sini disimpulkan pengguna Lazada dan Tokopedia yang memberikan ulasan cenderung positif, pada Lazada nilai presisi mencapai 94%, *recall* sebesar 93%, skor F1 sebesar 94%, dan Tokopedia nilai presisi mencapai 97%, *recall* sebesar 86%, serta skor F1 sebesar 91%.

Daftar Referensi

[1] S. Hilabi, ‘Rancang Bangun Sistem Inventory Usaha (UMKM)“Karpet” Desa Kamurang Berbasis Web’, *Prosiding Konferensi Nasional Penelitian Dan Pengabdian Universitas Buana Perjuangan Karawang*, vol. 2, no. 1, pp. 1147–1155, 2022.

[2] N. A. Azizah and T. Ernawati, ‘Analisis Sentimen Ulasan Pelanggan Terhadap Produk Roti Macan Artisan Di E-Commerce Tokopedia Menggunakan Metode Clustering’, *Journal of Information System, Applied, Management, Accounting and Research*, vol. 8, no. 3, pp. 580–590, Aug. 2024, doi: 10.52362/jisamar.v8i3.1576.

[3] N. Lestari, ‘Upaya Perlindungan Hukum Bagi Konsumen Dalam Transaksi Jual Beli Online (E-Commerce)’, *Nusantara: Jurnal Ilmu Pengetahuan Sosial*, vol. 10, no. 6, pp. 2871–2878, Sep. 2023.

[4] A. Zumarniansyah, ‘Penerapan Sistem Pendukung Keputusan Penilaian Karyawan Terbaik Dengan Metode Simple Additive Weighting’, *Jurnal Sistem Informasi*, vol. 10, no. 2, pp. 75–81, Aug. 2021, doi: 10.51998/jsi.v10i2.419.

[5] B. Priyatna, A. L. Hananto, and F. Nurapriani, ‘Integrasi Payment Gateway Pada Website E-Commerce Hasanatul. Id Untuk Mitra UMKM’, *Prosiding Konferensi Nasional Penelitian Dan Pengabdian Universitas Buana Perjuangan Karawang*, vol. 3, no. 1, pp. 924–929, 2023.

[6] N. Suarna and W. Prihartono, 'Penerapan Nlp (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore', *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 2, pp. 1841–1851, 2024.
[7] A. Hananto, 'Pemanfaatan Sosial Media Dan E-Commerce Sebagai Media Promosi Umkm Di Desa Mekarpohaci', *Prosiding Konferensi Nasional Penelitian Dan Pengabdian Universitas Buana Perjuangan Karawang*, vol. 3, no. 1, pp. 1147–1152, 2023.
[8] B. Huda and A. Hananto, 'Sosialisasi Dan Workshop Multimedia Berbasis Aplikasi Digital Platform Sosial Media', *Prosiding Konferensi Nasional Penelitian Dan Pengabdian Universitas Buana Perjuangan Karawang*, vol. 3, no. 1, pp. 2178–2181, 2023.
[9] Q. Al Qorni, 'Pendekatan Algoritma Naïve Bayes dan Support Vector Machine dalam Menganalisis Tanggapan Terhadap Penutupan Tik Tok Shop', *Jurnal Ilmiah Matrik*, vol. 25, no. 3, pp. 282–290, 2023.
[10] S. M. Salsabila, A. Alim Murtopo, and N. Fadhilah, 'Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier', *Jurnal Minfo Polgan*, vol. 11, no. 2, pp. 30–35, Aug. 2022, doi: 10.33395/jmp.v11i2.11640.
[11] E. Novalia and A. Voutama, 'Prediction of Rice Field Planted Area with CRISP-DM Using Classification and Regression Tree (Cart) Algorithms', *Systematics*, vol. 5, no. 1, pp. 578–587, 2023.
[12] C.A. Nrhaliza Agustina, 'The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store', *Malcom: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 1, pp. 47–54, 2022.
[13] A. I. Tanggraeni and M. N. N. Sitokdana, 'Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes', *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 2, pp. 785–795, 2022.
[14] B. Z. Ramadhan, 'Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes', *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 2, pp. 220–225, 2022.
[15] S. N. Fadhilah and F. S. Utomo, 'Algoritma Naïve Bayes untuk Analisis Sentiment Review Blibli.com di Google Play Store Naïve Bayes Algorithm for Sentiment Analysis of Blibli.com Review on Google Play Store', *Sistemasi: Jurnal Sistem Informasi*, vol. 13, no. 2, pp. 831–840, Mar. 2024.
[16] M. S. Mahendrasyah and T. Hariguna, 'Analisis Sentimen Pengguna Aplikasi Bukalapak di Platform Playstore Menggunakan Metode Naïve Bayes', *Technology and Science (BITS)*, vol. 6, no. 2, pp. 733–745, 2024, doi: 10.47065/bits.v6i2.5528.
[17] A. Syaifuddin and M. Ningsih, 'Penerapan Metode Content-Based Filtering Dalam Strategi Komunikasi Pemasaran Pada Marketplace Tokopedia', *Jurnal Responsif*, vol. 5, no. 2, pp. 185–194, 2023.
[18] Tukino, A. Hananto, R. A. Nanda, E. Novalia, E. Sediyono, and J. Sanjaya, 'LSTM and Word Embedding: Classification and Prediction of Puskesmas Reviews Via Twitter', *E3S Web of Conferences*, vol. 500, no. 1, pp. 1–10, Mar. 2024, doi: 10.1051/e3sconf/202450001018.
[19] A. L. Hananto, A. Hananto, B. Huda, A. Y. Rahman, E. Novalia, and B. Priyatna, 'Determination of Training Participants in Community Work Training Centers Using the Naïve Bayes Classifier Algorithm', *JOIV : International Journal on Informatics Visualization*, vol. 8, no. 3, pp. 1162–1167, Sep. 2024, doi: 10.62527/joiv.8.3.1995.
[20] A. G. Pramita and F. Nugraha, 'Sistem Analisis Sentimen Produk Pada Aplikasi Lazada Menggunakan Metode Naive Bayes', *Jurnal Digit*, vol. 14, no. 1, pp. 23–30, 2024.
[21] Y. J. Mado, 'Desain Sistem E-Commerce Untuk Peningkatan Penjualan Produk Pertanian', *Jurnal In Create (Inovasi dan Kreasi dalam Teknologi Informasi)*, vol. 10, no. 1, pp. 1–4, Jan. 2024.
[22] A. Gerliandeva, Y. Chrisnanto, and H. Ashaury, 'Optimasi Klasifikasi Sentimen pada Komentar Online menggunakan Multinomial Naïve Bayes dan Ekstraksi Fitur TF-IDF serta N-grams', *Jurnal Pekommas*, vol. 9, no. 2, pp. 260–272, Dec. 2024, doi: 10.56873/jpkm.v9i2.5585.
[23] A. Lia Hananto *et al.*, 'Analysis Of Drug Data Mining With Clustering Technique Using K-Means Algorithm', *J Phys Conf Ser*, vol. 1908, no. 1, pp. 1–8, Jun. 2021, doi: 10.1088/1742-6596/1908/1/012024.

[24] R. N. Safitri and S. S. Wati, ‘Mengintegrasikan Teknologi Informasi Dalam Proses Supervisi Manajerial’, *Jurnal Media Akademik (JMA)*, vol. 2, no. 6, pp. 1–11, 2024.
[25] N. Al Maeni, ‘Analisis Sentimen Pengguna Shopee Berdasarkan Data Tweet Dari Twitter Menggunakan Metode Naive Bayes Classifier’, *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 2, pp. 1834–1840, 2024.
[26] L. M. A. Arrozi, ‘Klasifikasi Tingkat Tutur Bahasa Sasak Berbasis Teks Menggunakan Naive Bayes’, *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, pp. 3556–3565, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5149.