

# Sentiment Analysis of Zalora Products on Google Play Store Using Random Forest Method: Analisis Sentimen terhadap Produk Zalora di Google Play Store Menggunakan Metode Random Forest

Rizky Saputra Cendikiawan  
Ali Ibrahim  
Mira Afrina  
Rizka Dhini Kurnia

Program Studi Sistem Informasi, Universitas Sriwijaya  
Program Studi Sistem Informasi, Universitas Sriwijaya  
Program Studi Sistem Informasi, Universitas Sriwijaya  
Program Studi Sistem Informasi, Universitas Sriwijaya

**General Background:** The rapid growth of e-commerce platforms has intensified the need to understand consumer sentiment to improve service quality and competitiveness. **Specific Background:** ZALORA, as a leading online fashion retailer in Southeast Asia, has accumulated vast user-generated feedback, particularly on platforms like the Google Play Store. **Knowledge Gap:** Despite the availability of such data, limited studies have analyzed consumer sentiment using machine learning methods specifically tailored to ZALORA's mobile platform. **Aims:** This study aims to examine consumer sentiment toward ZALORA products and assess the effectiveness of the Random Forest algorithm in classifying sentiment. **Results:** Utilizing a quantitative approach, 1,200 user reviews were analyzed, with 63.5% expressing positive sentiment. Word cloud visualization supported this finding, revealing frequently mentioned terms such as "product," "goods," and "shopping." The Random Forest model achieved an accuracy of 80%, with precision, recall, and F1-score values for positive sentiment all exceeding 0.80. **Novelty:** This research integrates TF-IDF-based preprocessing with Random Forest classification to enhance sentiment analysis performance specifically for mobile commerce reviews. **Implications:** The findings highlight the potential of machine learning in extracting actionable insights from user reviews, offering practical implications for improving customer experience and guiding strategic development in digital retail platforms.

## Highlights:

- High accuracy (80%) achieved using Random Forest for sentiment classification.
- TF-IDF preprocessing significantly improved model performance.
- Word cloud analysis revealed key satisfaction indicators from users.

**Keywords:** Sentiment Analysis, Random Forest Method, Zalora Consumers

## Pendahuluan

Kemajuan teknologi informasi dan komunikasi telah membawa perubahan besar dalam kehidupan masyarakat, termasuk dalam aktivitas berbelanja secara daring. Peningkatan signifikan dalam nilai transaksi online di seluruh dunia menunjukkan bahwa e-commerce telah menjadi pilihan utama

konsumen untuk memenuhi kebutuhan mereka [1]. Survei terbaru menunjukkan bahwa sekitar 2,71 miliar orang di seluruh dunia melakukan pembelian secara online melalui platform e-commerce dan media sosial, dengan 41% di antaranya melalui toko online atau situs web bisnis [2].

Aplikasi e-commerce berperan penting dalam meningkatkan kualitas layanan dengan menyediakan pengalaman yang lebih personal kepada pengguna. Hal ini dapat dilakukan dengan memanfaatkan preferensi pelanggan dan riwayat transaksi sebelumnya untuk memberikan rekomendasi produk yang relevan [3]. Dengan kemudahan dan jaminan kualitas yang ditawarkan, masyarakat semakin memilih toko online sebagai sarana berbelanja [4].

Analisis sentimen merupakan proses yang digunakan untuk mengevaluasi opini atau perasaan pengguna terhadap produk atau layanan tertentu. Teknik ini menjadi penting seiring meningkatnya kebutuhan perusahaan untuk memahami opini publik dan memperbaiki layanan berdasarkan masukan tersebut [5]. Analisis ini menggunakan pendekatan pemrosesan bahasa alami (Natural Language Processing) dan pembelajaran mesin (Machine Learning) untuk mengklasifikasikan opini sebagai positif, negatif, atau netral [6].

Salah satu tahapan penting dalam analisis sentimen adalah preprocessing data, seperti proses pembersihan data, stemming, dan penghapusan kata umum (stopword) yang tidak bermakna. Tahapan ini bertujuan untuk menyederhanakan teks dan meningkatkan akurasi model klasifikasi [7].

ZALORA merupakan salah satu e-commerce yang populer di Indonesia, khususnya dalam kategori fashion. Sebagai salah satu platform belanja yang banyak digunakan, persepsi konsumen terhadap produk dan layanan ZALORA sangat penting untuk dianalisis. Persepsi konsumen merupakan proses bagaimana individu menafsirkan dan memahami informasi yang diterimanya, sehingga dapat mempengaruhi keputusan pembelian [8].

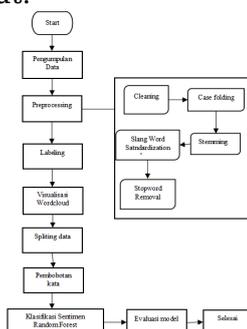
Untuk memahami persepsi tersebut, diperlukan metode yang mampu mengolah data ulasan konsumen secara efektif. Salah satu metode yang populer adalah Random Forest, yaitu algoritma pembelajaran mesin berbasis ensemble yang mampu menangani data dalam jumlah besar dan menghasilkan klasifikasi yang akurat [9]. Beberapa penelitian terdahulu menunjukkan bahwa Random Forest memiliki performa yang baik dalam mengklasifikasikan opini atau sentimen dalam data teks [10].

Dengan mempertimbangkan pentingnya analisis sentimen terhadap ulasan konsumen, serta efektivitas metode Random Forest dalam klasifikasi data, maka penelitian ini dilakukan untuk menganalisis sentimen konsumen terhadap produk ZALORA di Google Play Store

## Metode

Penelitian ini menggunakan pendekatan kuantitatif untuk melakukan analisis sentimen terhadap ulasan konsumen mengenai produk ZALORA di Google Play Store. Metode yang digunakan adalah algoritma klasifikasi Random Forest, dengan proses pengolahan data yang terdiri dari beberapa tahapan sebagaimana ditunjukkan pada Gambar 1.

Adapun tahapan penelitian terdiri dari beberapa langkah sistematis, dimulai dari pengumpulan data hingga evaluasi model, sebagai berikut:



**Figure 1.** Alur Penelitian

### **1. Pengumpulan Data**

Data dikumpulkan melalui scraping ulasan konsumen terhadap aplikasi ZALORA di Google Play Store. Data ini berisi komentar pengguna yang selanjutnya dianalisis untuk mengidentifikasi sentimen.

### **2. Preprocessing**

Proses ini bertujuan untuk membersihkan dan menyiapkan data teks agar siap dianalisis oleh model. Tahapan preprocessing meliputi:

Clearing: Menghapus karakter tidak penting seperti angka, simbol, dan tanda baca.

Case Folding: Mengubah seluruh teks menjadi huruf kecil untuk menyeragamkan format data.

Slang Word Standardization: Mengubah kata-kata tidak baku menjadi bentuk standarnya.

Stemming: Mengembalikan kata ke bentuk dasarnya menggunakan algoritma seperti Nazief-Adriani.

Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna penting seperti “yang”, “dan”, “di”, dan sebagainya [11].

### **3. Labeling**

Tahap ini bertujuan untuk memberi label sentimen pada setiap data ulasan berdasarkan polaritasnya, yakni positif atau negatif, secara manual maupun otomatis menggunakan kata kunci [12].

### **4. Visualisasi Wordcloud**

Data yang telah diproses kemudian divisualisasikan dalam bentuk word cloud untuk menampilkan frekuensi kata-kata yang paling sering muncul. Ini membantu menggambarkan topik dominan dalam ulasan [13].

### **5. Splitting Data**

Data dibagi menjadi dua bagian, yaitu data latih (training) dan data uji (testing). Umumnya pembagian dilakukan dengan perbandingan 80:20 atau 70:30 [14].

### **6. Pembobotan Kata**

Data teks dikonversi menjadi bentuk numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Rumus TF-IDF adalah sebagai berikut:

$$TFIDF(t,d) = TF(t,d) \times \log(N / DF(t))$$

Keterangan:

TF(t, d) = Frekuensi kemunculan term t dalam dokumen d

IDF(t) = Inverse Document Frequency, yaitu:  $\log(N / DF(t))$

N = Jumlah total dokumen

DF(t) = Jumlah dokumen yang mengandung term t [15]

### 7. Klasifikasi Sentimen dengan Random Forest

Metode Random Forest digunakan untuk melakukan klasifikasi sentimen. Algoritma ini membentuk beberapa pohon keputusan (decision tree) dan hasil akhirnya dipilih berdasarkan voting mayoritas dari semua pohon. Rumus prediksi klasifikasi menggunakan Random Forest adalah:

Keterangan:

$$H(x) = \text{majorityvote } h_1(x), h_2(x), \dots, h_n(x)$$

H(x) = Hasil prediksi akhir (dari Random Forest)

$h_1(x), h_2(x), \dots, h_n(x)$  = Prediksi dari masing-masing pohon keputusan

$h_1(x), h_2(x), \dots, h_n(x)$  = Prediksi dari masing-masing pohon keputusan

majority\_vote = Pengambilan hasil berdasarkan mayoritas voting dari semua pohon [16]

### 8. Evaluasi Model

Model dievaluasi menggunakan Confusion Matrix untuk menghitung metrik akurasi, presisi, recall, dan F1-score guna mengukur performa klasifikasi model [17].

## Hasil dan Pembahasan

### A. Hasil Scraping Data

Berdasarkan proses pengumpulan data melalui teknik scraping yaitu dengan mengambil komentar ulasan pengguna Zalora pada playstore secara otomatis, berhasil mengumpulkan sebanyak 1633 data ulasan yang mencakup kolom rating dan review, yang berisi teks ulasan dari pengguna produk Zalora. Teks ulasan ini mengandung berbagai simbol, angka, emoji, serta singkatan dalam bahasa yang tidak formal, sehingga memerlukan analisis lebih lanjut. Gambar 2 menunjukkan beberapa ulasan yang berhasil dikumpulkan melalui teknik scraping.

diskon sangat menarik buat belanja							
Zalora tempat belanja sangat bagus barang2 berkualitas & dapat discond sangat besar .							
Belanja di zalora menyenangkan, lebih pede, karna barangnya selalu original, terus sering kasih discount,							
Barang kwalitas bagus ukuran sesuai pesanan....MANTAP !							
Pengembalian barang ribet, gajelas. UI LO BENERIN!							
zalora terbaik							
lengkap & banyak promo. mantap							
belum di terima barang nya							
aplikasi nya bagus , balanja di Zalora menyenangkan							
kapok belanja zalora, baru sekali belanja langsung menyesal							

**Figure 2.** Hasil Scraping Data

## B. Hasil Preprocessing

Tahap pra-pemrosesan teks bertujuan untuk menyiapkan data komentar sebelum diterapkan pada model. Proses ini mencakup beberapa langkah penting, antara lain pembersihan data (cleaning), penyamaan huruf menjadi huruf kecil (case folding), normalisasi, normalisasi kata tidak baku atau bahasa gaul (slang normalization), penghapusan kata-kata umum yang tidak memiliki makna signifikan (stopword removal), dan stemming untuk mengubah kata ke bentuk dasarnya. Seluruh tahapan ini dilakukan menggunakan platform Google Colaboratory. Hasilnya dapat dilihat pada Tabel 1 berikut :

Sebelum Proses Processing	
Original	Zalora tempat belanja sangat bagus barang2 berkualitas & dapat discond sangat besar
Sesudah Proses Processing	
Cleaning Data	Zalora tempat belanja sangat bagus barang berkualitas dapat discond sangat besar
Case Folding	zalora tempat belanja sangat bagus barang berkualitas dapat discond sangat besar
Slang Normailized	zalora tempat belanja sangat bagus barang berkualitas dapat discond sangat besar
Stop Removal	['zalora', 'belanja', 'bagus', 'barang', 'berkualitas', 'discond']
Stemming	['zalora', 'belanja', 'bagus', 'barang', 'berkualitas', 'discond']

**Table 1.** Hasil Proses Preprocessing

## C. Hasil Labeling

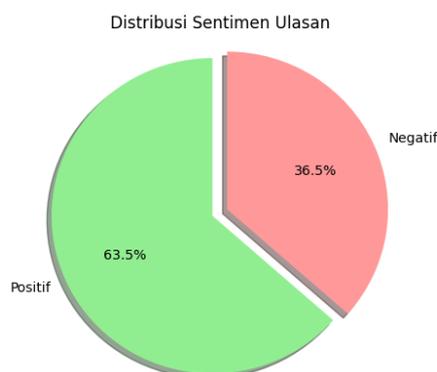
Pada data yang di dapat melalui proses scraping sebelumnya belum terdapat label positif, negatif dan netral. Sehingga penulis melakukan pelabelan berdasarkan dictionary kosa kata yang berkonotasi positif dan negatif. Dictionary kosa kata yang berkonotasi positif dan negatif di peroleh melalui link berikut :

Dictionary kosa kata positif :

<https://raw.githubusercontent.com/masdevid/ID-OpinionWords/master/positive.txt>

Dictionary kosa kata negatif : <https://raw.githubusercontent.com/masdevid/ID-OpinionWords/master/negative.txt>

Berdasarkan dictionary kosa kata diatas, komentar yang tidak terdapat pada dictionary tersebut di anggap netral dan akan di hapuskan. Sehingga ulasan yang di dapat adalah sebanyak 1200 ulasan yang berisi 762 komentar positif dan 438 komentar negatif. Hasil dari labeling data di visualisasikan dalam bentuk pie chart yang dapat dilihat pada Gambar 2 berikut:



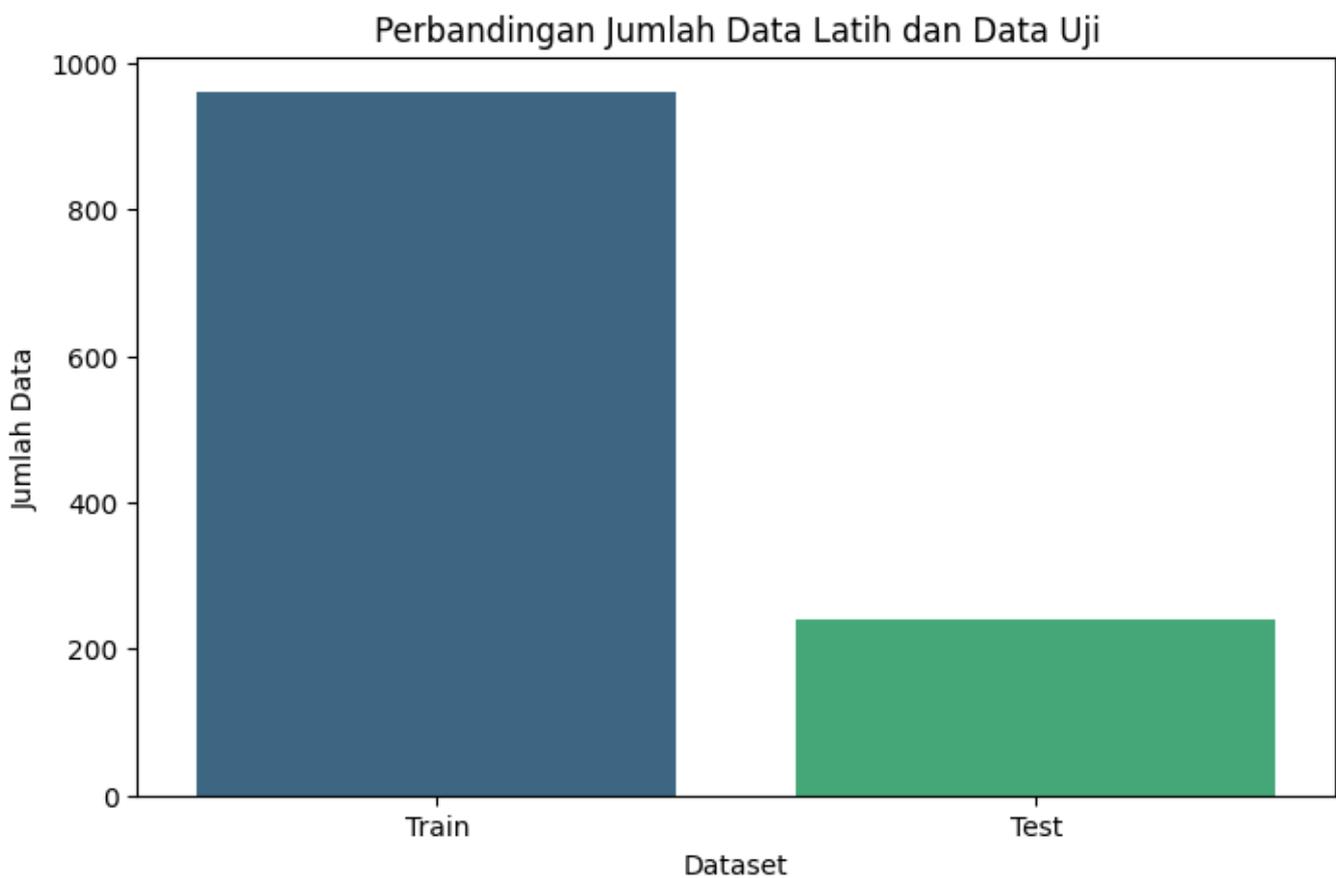
**Figure 3.** Visualisasi Pie Chart Labeling Data



Melalui visualisasi WordCloud untuk komentar negatif di atas, terlihat bahwa istilah-istilah seperti "zalora", "barang", "kirim", "belanja", dan "pesan" menjadi kata yang paling menonjol dalam ulasan pengguna. Besarnya kata "barang" dan "kirim" mengindikasikan bahwa banyak keluhan pelanggan berhubungan dengan permasalahan pengiriman, seperti keterlambatan atau ketidaksesuaian produk yang diterima. Dominasi kata "zalora" dan "belanja" menunjukkan bahwa pengalaman berbelanja di platform ini sering menjadi sumber ketidakpuasan pengguna. Selain itu, kata-kata seperti "gak", "batal", "salah", "buruk", dan "kecewa" juga kerap muncul, memperkuat adanya sentimen negatif yang berkaitan dengan layanan, kualitas produk, maupun proses pembelian. Secara keseluruhan, WordCloud ini memperlihatkan bahwa mayoritas keluhan pelanggan Zalora berkisar pada isu ketidaksesuaian produk dan masalah dalam pengiriman.

### **E. Hasil Splitting Data**

Dataset ulasan produk Zalora dibagi menjadi data latih dan data uji guna memastikan proses pelatihan dan evaluasi model berjalan secara objektif. Pembagian dengan rasio 80:20 menghasilkan sebanyak 960 data uji dan 240 latih, dapat dilihat pada Gambar 6 berikut:



**Figure 6.** Hasil Visualisasi Splitting Data

Melalui visualisasi barplot, dapat diketahui bahwa pembagian dataset ulasan produk Zalora telah dilakukan sesuai dengan proporsi yang direncanakan, yakni 80% untuk data latih dan 20% untuk data uji. Sebagian besar data dialokasikan untuk proses pelatihan guna memungkinkan model belajar dari berbagai variasi data, sementara sebagian kecil sisanya digunakan untuk menguji performa model. Pembagian ini bertujuan agar model yang dihasilkan memiliki akurasi yang tinggi serta mampu melakukan generalisasi dengan baik terhadap data baru. Selain itu, distribusi yang seimbang ini juga berfungsi untuk meminimalkan risiko overfitting selama proses pelatihan

### F. Hasil Pembobotan Kata TF-IDF

Metode TF-IDF digunakan untuk menentukan bobot pada setiap kata dalam teks ulasan, sehingga membantu dalam mengenali istilah-istilah penting secara lebih akurat. Dengan pendekatan ini, TF-IDF mampu menyoroti kata-kata yang sering muncul dalam satu ulasan tetapi jarang ditemukan di seluruh kumpulan data, sehingga memberikan wawasan yang lebih mendalam terkait opini pengguna terhadap produk Zalora. Hasil dari pembobotan kata TF-IDF dapat dilihat pada Gambar 7 berikut:

	feature	tfidf	document
0	diskon	0.526571	0
1	tarik	0.769895	0
2	belanja	0.360533	0
3	belanja	0.243423	1
4	zalora	0.179994	1
5	bagus	0.327594	1
6	barang	0.193864	1
7	kualitas	0.395385	1
8	discond	0.779177	1
9	belanja	0.120619	2

Figure 7. Hasil TF-IDF

### G. Hasil klasifikasi Random Forest

Model klasifikasi sentimen berbasis Random Forest yang digunakan berhasil mencapai akurasi sekitar 80%, menandakan performa yang cukup baik dalam mengenali pola data. Hasil evaluasi menunjukkan bahwa model lebih unggul dalam mengidentifikasi ulasan positif, ditunjukkan oleh nilai precision, recall, dan f1-score yang masing-masing berada di atas 0.80. Hasilnya dapat dilihat pada Gambar 8. berikut :

```
Akurasi: 0.7958333333333333

Laporan Klasifikasi:
      precision    recall  f1-score   support

 Negatif      0.71      0.74      0.72         87
  Positif      0.85      0.83      0.84        153

 accuracy                0.80         240
 macro avg              0.78      0.78      0.78         240
 weighted avg           0.80      0.80      0.80         240
```

Figure 8. Hasil Klasifikasi Random Forest

## H. Evaluasi Model

Dari confusion matrix yang diperoleh, terlihat bahwa model lebih akurat dalam mengenali sentimen positif dibandingkan sentimen negatif. Sebanyak 127 ulasan positif berhasil diklasifikasikan dengan benar, sedangkan hanya 64 ulasan negatif yang terdeteksi secara tepat. Masih terdapat kekeliruan dalam prediksi, di mana 23 ulasan negatif salah diprediksi sebagai positif dan 26 ulasan positif diklasifikasikan sebagai negatif. Temuan ini mengindikasikan bahwa model memiliki kecenderungan untuk lebih mudah mengenali ulasan yang bernada positif, sementara kemampuan mendeteksi sentimen negatif masih perlu ditingkatkan.

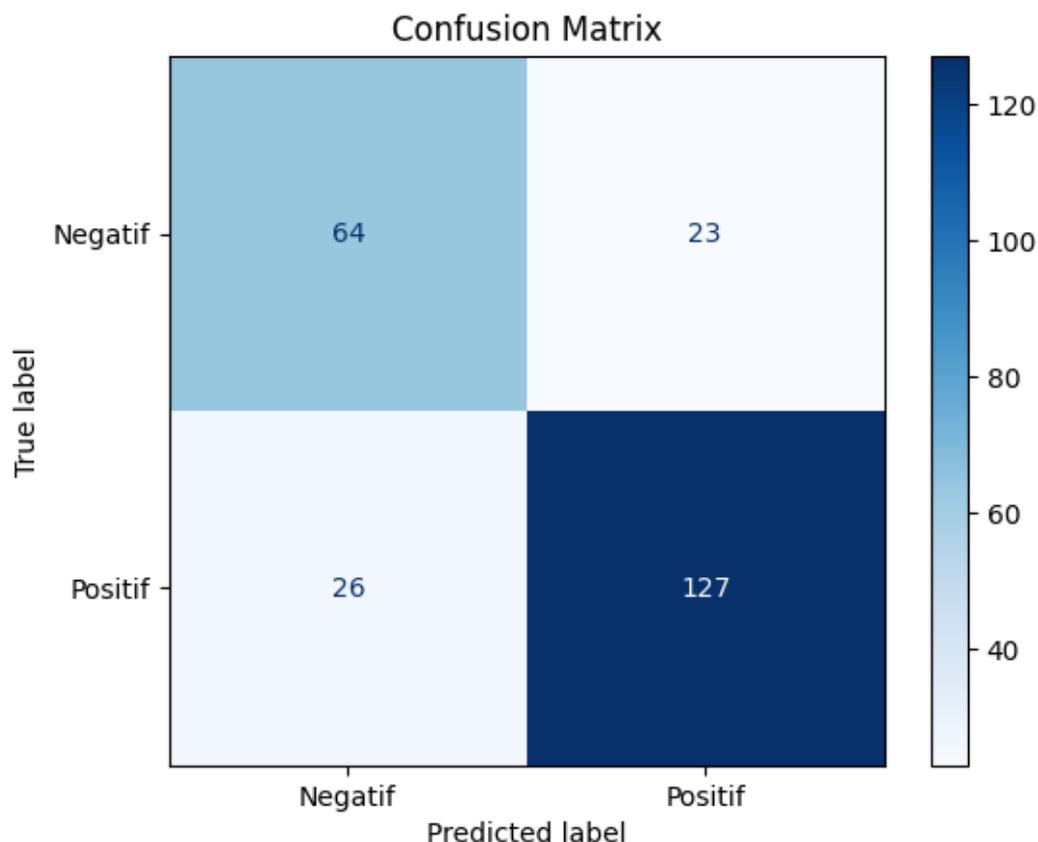


Figure 9. Hasil Visualisasi Confusion Matrix

## Simpulan

Berdasarkan hasil penelitian mengenai analisis sentimen konsumen terhadap produk ZALORA di Google Play Store menggunakan metode Random Forest, dapat disimpulkan beberapa poin utama sebagai berikut: (1) Sentimen konsumen terhadap produk ZALORA cenderung positif, sebagaimana ditunjukkan oleh hasil labeling data yang menunjukkan bahwa dari 1.200 ulasan yang dianalisis, sebanyak 762 ulasan (63,5%) merupakan ulasan positif. Visualisasi word cloud juga memperkuat kesimpulan ini, dengan dominasi kata-kata seperti "produk", "barang", dan "belanja" yang mencerminkan kepuasan konsumen terhadap layanan dan kualitas produk dari ZALORA. (2) Metode Random Forest menunjukkan performa yang baik dalam mengklasifikasikan sentimen ulasan konsumen. Berdasarkan evaluasi menggunakan confusion matrix, model ini berhasil mencapai akurasi sebesar 80%. Nilai precision, recall, dan F1-score untuk kategori sentimen positif masing-masing berada di atas angka 0,80, menunjukkan kemampuan klasifikasi yang cukup andal. Namun, performa model dalam mendeteksi ulasan negatif masih perlu ditingkatkan, mengingat adanya ketidakseimbangan klasifikasi pada beberapa data uji. (3) Proses preprocessing dan pembobotan kata menggunakan TF-IDF berperan penting dalam meningkatkan kualitas input untuk

model klasifikasi. Tahapan seperti cleaning, stemming, dan stopwords removal berhasil menyederhanakan data ulasan, sehingga membantu dalam pengenalan pola sentimen oleh algoritma

## References

- [1] S. Ahadiyah, "Tren Perkembangan E-Commerce di Era Digital," *Jurnal Teknologi dan Bisnis Digital*, vol. 8, no. 1, pp. 12-20, 2024.
- [2] SellersCommerce, "E-Commerce Statistics and Trends," [Online]. Available: <https://sellerscommerce.com> [Accessed: May 14, 2025].
- [3] S. Asyifah, A. D. Pratama, and M. R. Hanif, "Personalisasi Layanan dalam Aplikasi E-Commerce," *Jurnal Sistem Informasi Bisnis*, vol. 10, no. 2, pp. 75-83, 2023.
- [4] M. Setiawan, L. S. Putri, and A. M. Fadilah, "Pengaruh Kualitas Produk dan Kemudahan Belanja terhadap Minat Beli Konsumen Online," *Jurnal Ekonomi dan Bisnis*, vol. 7, no. 1, pp. 45-52, 2024.
- [5] M. Aminullah, Y. D. Ramadhani, and R. Prasetyo, "Analisis Sentimen Menggunakan Metode NLP terhadap Opini Konsumen," *Jurnal Informatika dan Sistem Informasi*, vol. 12, no. 1, pp. 30-38, 2024.
- [6] A. Jennifer, I. R. Lubis, and H. Z. Pratama, "Penerapan Sentiment Analysis dalam Dunia Bisnis," *Jurnal Ilmu Komputer dan Aplikasi*, vol. 15, no. 2, pp. 90-97, 2024.
- [7] N. Sagita, H. D. Rahmat, and Y. Anggraini, "Optimasi Preprocessing Data Teks untuk Analisis Sentimen," *Jurnal Sains Data*, vol. 5, no. 1, pp. 22-29, 2023.
- [8] Y. Yudawisastro, F. P. Nugraha, and S. Kurniawan, *Manajemen Pemasaran Strategik*, Bandung: Alfabeta, 2024.
- [9] E. Nursiyono, *Pengantar Pembelajaran Mesin untuk Analisis Data*, Yogyakarta: Deepublish, 2023.
- [10] R. Rania and I. Syah, "Penerapan Random Forest untuk Klasifikasi Sentimen Review Konsumen," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 9, no. 2, pp. 60-68, 2024.
- [11] F. Saputri, D. K. Anjani, and A. R. Maulana, "Penerapan Preprocessing dalam Pengolahan Teks Bahasa Indonesia," *Jurnal Teknologi Informasi*, vol. 7, no. 1, pp. 45-52, 2023.
- [12] A. D. Nugroho and T. P. Putri, "Sistem Labeling Otomatis Sentimen Menggunakan Lexicon Bahasa Indonesia," in *Proc. Seminar Nasional Teknologi Informasi*, 2022.
- [13] R. Irawan, "Visualisasi Word Cloud untuk Analisis Sentimen Media Sosial," *Jurnal Media Informatika*, vol. 5, no. 2, pp. 61-69, 2021.
- [14] A. Hidayat and B. Kurniawan, "Pembagian Data Latih dan Uji yang Optimal pada Algoritma Machine Learning," *Jurnal Ilmiah Komputer dan Informatika*, vol. 9, no. 1, pp. 23-30, 2022.
- [15] K. Sari and F. Hidayat, "Implementasi TF-IDF untuk Analisis Ulasan Produk," *Jurnal Teknologi dan Informasi*, vol. 11, no. 2, pp. 101-110, 2023.
- [16] B. Nugroho and R. A. Putri, "Studi Perbandingan Metode Random Forest dan Decision Tree," *Jurnal Informatika Universitas ABC*, vol. 8, no. 1, pp. 55-64, 2024.
- [17] D. S. Lestari, A. R. Mahendra, and M. P. Rachman, "Evaluasi Model Klasifikasi Sentimen Menggunakan Confusion Matrix," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 6, no. 3, pp. 88-97, 2023.