

# ANALISIS SENTIMEN APLIKASI SHOPEE, TOKOPEDIA, LAZADA DAN BLIBLI MENGGUNAKAN LEKSIKON DAN RANDOM FOREST

Adryan Syah<sup>1\*</sup>, Firman Nurdyansyah<sup>2</sup>, Aviv Yuniar Rahman<sup>3</sup>

<sup>1,2,3</sup>Universitas Widyagama Malang, Jl. Borobudur No.35, Mojolangu, Kec. Lowokwaru, Kota Malang (0341) 411291

---

Received: 5 Agustus 2024

Accepted: 5 Oktober 2024

Published: 12 Oktober 2024

**Keywords:**

Sentiment Analysis;  
E-Commerce;  
Lexicon-Based Approach;  
Random Forest.

**Corespondent Email:**

adryan.shinta.16@gmail.com

**Abstrak.** Dalam era digital, aplikasi *e-commerce* telah menjadi sarana utama bagi masyarakat untuk berbelanja. Keberhasilan aplikasi *e-commerce* tidak hanya bergantung pada fungsionalitasnya tetapi juga pada pengalaman pengguna. Ulasan pengguna di *Play Store* menjadi indikator penting dalam mengevaluasi kepuasan dan sentimen pengguna terhadap aplikasi tertentu. Penelitian ini bertujuan untuk menganalisis sentimen pada ulasan aplikasi *Shopee*, *Tokopedia*, *Lazada*, dan *Blibli* di *Play Store* menggunakan pendekatan *Lexicon-based* dan algoritma *Random Forest*. Metode ini dipilih untuk memberikan interpretasi yang jelas terhadap sentimen teks dan meningkatkan akurasi analisis sentimen. Hasil penelitian menunjukkan bahwa aplikasi *Lazada* memiliki kinerja terbaik dengan akurasi 88,33%, presisi 88,88%, *recall* 88,33%, dan *F1 score* 88,34%. Aplikasi *Blibli* berada di posisi kedua dengan akurasi 85,66%, presisi 85,82%, *recall* 85,66%, dan *F1 score* 85,60%. *Shopee* memiliki akurasi 85,16%, presisi 85,62%, *recall* 85,16%, dan *F1 score* 85,26%. *Tokopedia* menunjukkan performa terendah dengan akurasi 80,33%, presisi 80,96%, *recall* 80,33%, dan *F1 score* 80,12%. Penelitian ini menunjukkan bahwa rasio pembagian data latih dan data uji mempengaruhi kinerja model, dengan model bekerja lebih efektif ketika jumlah data latih lebih besar dari data uji.

**Abstract.** In the digital era, *e-commerce* applications have become the primary means for people to shop. The success of *e-commerce* applications depends not only on their functionality but also on user experience. User reviews on the *Play Store* are an important indicator for evaluating user satisfaction and sentiment towards a particular application. This study aims to analyze sentiment in reviews of the *Shopee*, *Tokopedia*, *Lazada*, and *Blibli* applications on the *Play Store* using a *Lexicon-based* approach and *Random Forest* algorithm. This method was chosen to provide a clear interpretation of text sentiment and improve sentiment analysis accuracy. The results show that the *Lazada* application has the best performance with an accuracy of 88.33%, precision of 88.88%, *recall* of 88.33%, and an *F1 score* of 88.34%. The *Blibli* application is in second place with an accuracy of 85.66%, precision of 85.82%, *recall* of 85.66%, and an *F1 score* of 85.60%. *Shopee* has an accuracy of 85.16%, precision of 85.62%, *recall* of 85.16%, and an *F1 score* of 85.26%. *Tokopedia* shows the lowest performance with an accuracy of 80.33%, precision of 80.96%, *recall* of 80.33%, and an *F1 score* of 80.12%. This study shows that the ratio of training data to test data affects the model's performance, with the model working more effectively when the amount of training data is larger than the test data.

---

## 1. PENDAHULUAN

Dalam era digital dan kemajuan teknologi, aplikasi *e-commerce* telah menjadi sarana utama bagi masyarakat untuk berbelanja [1]. Keberhasilan dan penerimaan aplikasi *e-commerce* tidak hanya bergantung pada fungsionalitasnya, tetapi juga pada pengalaman pengguna [2]. Ulasan pengguna di *Play store* menjadi indikator penting dalam mengevaluasi kepuasan dan sentimen pengguna terhadap aplikasi *e-commerce* tertentu. Dengan meningkatnya persaingan di pasar *e-commerce*, pemahaman terhadap sentimen pengguna menjadi kunci strategis untuk pengembangan dan perbaikan aplikasi. Oleh karena itu, analisis sentimen pada ulasan aplikasi *e-commerce* sangat relevan untuk memahami persepsi pengguna dan mendukung pengambilan keputusan bagi penyedia layanan *e-commerce*.

Dalam beberapa tahun terakhir, *e-commerce* populer seperti *Shopee*, *Tokopedia*, *Lazada*, *Blibli*, dan *Bukalapak* telah menjadi pilihan utama konsumen di Indonesia [3]. Data yang dirilis oleh *databoks.katadata.co.id* mengungkapkan statistik kunjungan ke *platform e-commerce* terkemuka di Indonesia selama tahun 2023. Dalam periode Januari hingga Desember 2023, tercatat bahwa *Shopee* memimpin dengan total 2,3 miliar kunjungan. *Tokopedia* menempati posisi kedua dengan 1,2 miliar kunjungan, diikuti oleh *Lazada* yang mencapai 762,4 juta kunjungan. Sementara itu, *Blibli* dan *Bukalapak* masing-masing mencatatkan 337,4 juta dan 168,2 juta kunjungan selama periode tersebut.

*E-commerce* merujuk pada kegiatan jual beli barang dan jasa secara elektronik melalui internet, memungkinkan transaksi bisnis antara penjual dan pembeli tanpa pertemuan fisik [4]. Beberapa *platform e-commerce* populer di Indonesia seperti *Shopee*, *Tokopedia*, *Lazada*, dan *Blibli* menjadi objek penelitian ini. *Shopee*, didirikan oleh Forrest Li pada 2009 dan bagian dari *Sea Limited*, awalnya berfokus pada model *C2C* dan kemudian berkembang menjadi *hybrid C2C* dan *B2C* dengan *Shopee Mall* pada 2015, mencatat lebih dari 80 juta unduhan pada 2017 [5]. Pada 17 Agustus 2009, William Tanuwijaya dan Leontinus Alpha Edison mendirikan *Tokopedia* dengan visi mewujudkan pemerataan ekonomi di era digital Indonesia. *Platform* ini dirancang untuk memungkinkan transaksi jual beli yang

menjangkau seluruh pelosok nusantara, menghubungkan penjual dan pembeli dari berbagai daerah [6]. *Lazada*, diluncurkan oleh *Rocket Internet* pada Maret 2012, menawarkan berbagai produk mulai dari *fashion* hingga elektronik dan kesehatan [7]. *Blibli* dikenal karena reputasi tinggi dan komitmennya terhadap kualitas layanan dan ketepatan waktu pengiriman, memprioritaskan kepuasan pelanggan [8] [9]. Sementara analisis sentimen sendiri merupakan suatu metode dalam domain pembelajaran mesin dan pemrosesan bahasa alami [10]. Teknik ini digunakan untuk mengenali serta mengekstrak opini, sentimen, ulasan, sikap, dan emosi yang terkandung dalam teks, memberikan pemahaman yang lebih mendalam terhadap respon dan pandangan dalam konteks tertentu [11].

Penelitian mengenai tanggapan publik terhadap produk atau layanan dalam dunia belanja online menunjukkan berbagai metode analisis sentimen yang telah diterapkan. [12] menggunakan *Support Vector Machine* untuk mengklasifikasi ulasan aplikasi belanja online di *Google Play Store*, mencapai akurasi tertinggi 90,67% pada aplikasi *Tokopedia*. [13] membandingkan algoritma *PNN* dan *KNN* dalam analisis ulasan pengguna *Shopee*, dengan *KNN* menunjukkan akurasi lebih tinggi 77,85% untuk aplikasi dan 91,43% untuk produk. [14] menerapkan *Convolutional Neural Network (CNN)* untuk ulasan *e-commerce* dengan akurasi 86,6%. [15] menggunakan algoritma *Decision Tree (DT)* dan *Random Forest (RF)* dalam analisis ulasan *marketplace*, dengan akurasi tertinggi 96,55% pada *Shopee*. [11] menyoroti pentingnya analisis sentimen dalam memahami opini dan emosi dalam teks. [16] menggunakan *Naïve Bayes* dan Regresi Logistik untuk ulasan aplikasi di *Google Play Store*, dengan Regresi Logistik mencapai akurasi 84,58% pada dataset *Shopee*. [17] melakukan analisis sentimen ulasan *Shopee* menggunakan *Naïve Bayes*, dengan teknik *Hold-Out* menghasilkan akurasi 83%. [18] membandingkan metode *Lexicon-Based* dan *Naive Bayes Classifier* untuk ulasan aplikasi investasi, dengan *Naive Bayes* menunjukkan akurasi lebih tinggi 78%. Penelitian-penelitian ini memberikan wawasan berharga bagi pengembang aplikasi dan pelaku bisnis dalam memahami respon pengguna serta meningkatkan layanan dan strategi bisnis.

Berdasarkan penelitian terdahulu, penting untuk memahami bagaimana pandangan pengguna terhadap aplikasi *e-commerce* sebagai bahan acuan kepada penyedia layanan untuk memperbaiki layanannya dan untuk memberikan rekomendasi yang nyata untuk pengguna aplikasi *e-commerce*. Namun hingga saat penelitian ini dilakukan, masih sedikit penelitian yang melakukan analisis sentimen terhadap aplikasi *e-commerce* yang sedang populer atau sedang banyak digunakan pengguna. Selain itu, masih sedikit penelitian yang menerapkan penggunaan leksikon atau *lexicon based approach* untuk melakukan *labeling* yang lebih cepat, efisien dan sesuai dibanding *labeling* manual atau *labeling* menggunakan rating untuk penelitian dengan *dataset* yang diperoleh dari *google play store*. Oleh karena itu, peneliti melakukan pendekatan yang berbeda yaitu menggunakan *lexicon based approach* untuk melakukan *auto labeling* pada *dataset* hasil *scraping* untuk meningkatkan efisiensi dan ketepatan atau kesesuaian kata dengan makna yang sebenarnya. Hal ini peneliti lakukan dengan harapan dapat memberikan gambaran yang komprehensif terkait analisis sentimen pada aplikasi *e-commerce* populer di Indonesia.

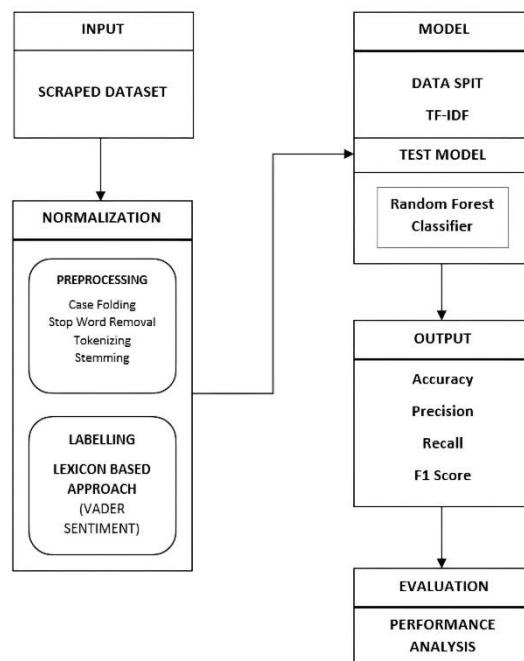
Dalam penelitian ini, pendekatan *Lexicon Based* dan *Random Forest* dipilih sebagai metode analisis sentimen yang saling melengkapi. Pendekatan berbasis *lexicon* digunakan untuk memberikan interpretasi yang jelas terhadap sentimen teks, memanfaatkan daftar kata kunci dan nilai sentimen yang telah terdefinisi [19]. Metode ini sangat efektif untuk menganalisis sentimen pada teks pendek, seperti ulasan atau komentar di media sosial, dan memiliki keunggulan dalam keefisienan komputasi. Di sisi lain, penggunaan *Random Forest* sebagai model *machine learning* memberikan keuntungan akurasi yang tinggi dalam mengatasi kompleksitas data teks. *Random Forest* dikenal memiliki kinerja yang baik dalam menangani *overfitting* dan ketangguhan terhadap *noise* atau *outlier* dalam *dataset* [20]. Kemampuan model ini untuk memberikan informasi tentang kepentingan setiap fitur juga merupakan nilai tambah, membantu pemahaman terhadap kontribusi kata-kata atau fitur dalam klasifikasi sentimen.

Penelitian ini bertujuan untuk menganalisis sentimen pada ulasan aplikasi *Shopee*,

*Tokopedia*, *Lazada*, dan *Blibli* di *Play Store* dengan menggunakan pendekatan *Lexicon-based* dan algoritma *Random Forest*. Penelitian ini ingin menjawab bagaimana penerapan kedua metode tersebut dalam analisis sentimen dan bagaimana rasio pembagian data latih dan data uji mempengaruhi performa model *Random Forest*. Ruang lingkup penelitian mencakup ulasan pengguna dari *platform Play Store* dengan data yang dikumpulkan dari Mei hingga Juni 2024. Penelitian ini diharapkan bermanfaat bagi penyedia layanan *e-commerce* dalam memahami sentimen pengguna serta bagi pengembangan metode analisis sentimen yang lebih efektif di masa mendatang.

## 2. METODE PENELITIAN

Penelitian ini menyajikan model analisis sentimen yang dirancang untuk mengolah dan menganalisis ulasan pengguna dari aplikasi *e-commerce*. Model ini menggabungkan pendekatan berbasis leksikon dengan algoritma pembelajaran mesin untuk mengklasifikasikan sentimen ulasan pengguna secara efektif. Gambar 1 memvisualisasikan alur kerja model analisis sentimen yang digunakan dalam penelitian ini, mulai dari pengumpulan data hingga evaluasi kinerja model.



Gambar 1. Model Analisis Sentimen *Lexicon Based* dan *Random Forest*

## 2.1. Input

Dalam proses pengumpulan data, peneliti memanfaatkan *library Google Play Scraper* untuk mengekstraksi informasi dari platform *Google Play Store*. Melalui metode ini, berhasil dihimpun sebanyak 40.596 entri data. Selanjutnya, *dataset* yang terkumpul akan diproses melalui tahap normalisasi untuk mempersiapkan analisis lebih lanjut.

## 2.2. Normalization

Tahap normalisasi merupakan proses penyesuaian data agar kompatibel dengan model *Random Forest Classifier* yang akan diimplementasikan untuk klasifikasi dalam studi ini. Proses ini terdiri dari dua komponen utama yaitu *preprocessing* dan *labeling*.

### 2.2.1. Preprocessing

Pada tahap ini, *dataset* akan menjalani proses pembersihan untuk mempersiapkan data agar dapat digunakan dalam model. Dalam *preprocessing*, beberapa tahapan penting dilakukan untuk mempersiapkan data agar dapat dianalisis lebih lanjut. Proses ini meliputi *Case Folding*, yang mengubah semua huruf dalam teks menjadi huruf kecil untuk menyeragamkan format. *Stopword removal*, yaitu menghapus kata-kata umum yang tidak memiliki makna signifikan. *Tokenizing*, yang memecah teks menjadi unit-unit individual atau token, biasanya berupa kata-kata. Dan terakhir *Stemming*, proses mengubah kata-kata menjadi bentuk dasarnya dengan menghilangkan imbuhan.

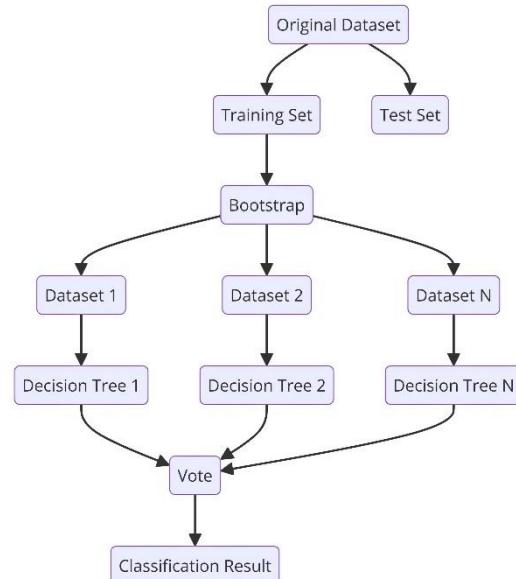
### 2.2.2. Labeling

Pasca tahap *preprocessing*, *dataset* akan dialihbahasakan ke dalam bahasa Inggris. Translasi ini memfasilitasi proses *auto labeling* dengan memanfaatkan *library VADER Sentiment*. Melalui proses *labeling* ini, sentimen akan diklasifikasikan ke dalam tiga kategori: positif, negatif, dan netral.

## 2.3. Model

Pada tahap *modelling* ini, Pemodelan dilakukan dengan membagi *dataset* menjadi beberapa rasio pengujian, yaitu 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20, dan 90:10, untuk mendapatkan hasil maksimal. Kemudian diubah menjadi representasi numerik menggunakan teknik *TF-IDF* (*Term Frequency-Inverse Document Frequency*). Setelah data disiapkan, model klasifikasi dilatih

menggunakan algoritma *Random Forest Classifier* untuk mencapai hasil optimal. Adapun arsitektur *random forest* ditampilkan pada Gambar 2.



Gambar 2. Arsitektur *Random Forest*

Dalam arsitektur *Random Forest* pada Gambar 2, kumpulan data asli dibagi menjadi satu set pelatihan dan satu set pengujian. Set pelatihan mengalami *bootstrapping* untuk membuat beberapa set data. Setiap set data digunakan untuk melatih pohon keputusan individual. Keluaran dari pohon keputusan ini kemudian diagregasi melalui mekanisme pemungutan suara untuk menghasilkan hasil klasifikasi akhir [21].

## 2.4. Output

*Output* model akan digunakan sebagai evaluasi kinerja model yang dilakukan dengan menggunakan beragam metrik, mencakup akurasi, presisi, *recall*, dan *F1 score*. Metrik-metrik ini menyajikan penilaian menyeluruh terhadap efektivitas klasifikasi yang dihasilkan oleh model. Representasi matematis dari masing-masing metrik tersebut disajikan dalam persamaan (1), (2), (3), dan (4).

$$\text{Akurasi} = \frac{\text{Jumlah prediksi benar}}{\text{Total prediksi}} \quad (1)$$

$$\text{Presisi} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \quad (3)$$

$$\text{F1 Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

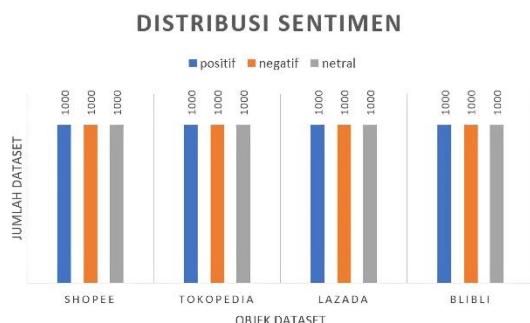
### 2.5. Evaluation

Untuk mengevaluasi kinerja model secara komprehensif, kami akan melakukan analisis performa tiap objek menggunakan rasio uji dan *confusion matrix*. Analisis ini bertujuan mengidentifikasi kekuatan dan kelemahan model secara rinci. Kami akan menghitung berbagai metrik evaluasi, termasuk akurasi, presisi, *recall*, dan *F1 score* untuk setiap objek. Selanjutnya, kami akan memeriksa *confusion matrix* masing-masing objek guna memperoleh pemahaman mendalam tentang kemampuan model dalam mengklasifikasikan kasus positif, negatif, dan netral.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Scraping

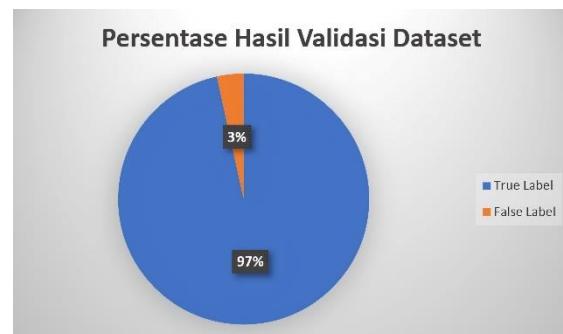
Pada saat pengumpulan data peneliti mendapatkan total 40.596 data dari *platform Google Playstore*. Namun pada penelitian ini kami ingin melakukan analisis sentimen dengan *dataset* yang seimbang untuk meningkatkan performa model, oleh karena itu peneliti melakukan *filtering* data dengan mengambil masing-masing 3000 data per objek. Hasilnya didapatkan *dataset* untuk dilanjutkan ketahap pengujian, 1000 data positif, 1000 data negatif dan 1000 data netral pada masing-masing objek seperti yang ditampilkan pada Gambar 3.



Gambar 3. Distribusi Sentimen

Selanjutnya validasi data manual juga dilakukan untuk memastikan hasil *auto labeling* sesuai dengan data yang sebenarnya. Berdasarkan validasi manual yang dilakukan, peneliti mengidentifikasi bahwa kesalahan *auto labeling* sebagian besar disebabkan oleh ketidakakuratan hasil terjemahan. Ketidaktepatan terjemahan ini mengakibatkan model kesulitan dalam mengidentifikasi kata-

kata dengan benar. Visualisasi persentase hasil validasi *dataset* terlihat pada Gambar 4.



Gambar 4. Persentase Hasil Validasi Dataset

Gambar 4 menunjukkan bahwa meskipun terdapat kendala, tingkat akurasi pelabelan tetap menunjukkan hasil yang sangat baik, yakni mencapai 97%, sebagaimana terlihat pada Gambar 4, *True Label* merupakan *dataset* dengan label yang sudah tepat dan *False Label* merupakan *dataset* dengan label yang salah.

### 3.2. Normalization

Normalisasi meliputi tahap *preprocessing* dan tahap *labeling* untuk menyiapkan *dataset* bagi pemodelan. Hasil dari *preprocessing* ditampilkan pada Tabel 1 dan tahap *labeling* pada Tabel 2.

Tabel 1. Hasil Preprocessing

Proses	Teks
Teks Asli	Sangat membantu dan tidak mengecewakan
Case Folding	sangat membantu dan tidak mengecewakan
Stopword Removal	membantu mengecewakan
Tokenizing	['membantu', 'mengecewakan']
Stemming	bantu kecewa

Hasil *preprocessing* yang ditampilkan dalam Tabel 1 menggambarkan transformasi yang efektif dari teks asli ke bentuk yang lebih sederhana dan standar untuk analisis lebih lanjut. Proses *case folding* berhasil menstandarkan kapitalisasi, sementara *stopword removal* efisien mengeliminasi kata-kata tidak signifikan, meningkatkan fokus pada kata-kata kunci yang lebih informatif. Tahap *tokenizing* memisahkan teks menjadi token-token individual yang siap untuk analisis lanjutan dan *stemming* mengurangi kata-kata ke bentuk dasar mereka.

Tabel 2. Hasil *Labeling*

Ulasan	Sentimen
the application is ugly, expensive, not like an orange store, not like a discount like the store next to the tree, it's complicated	Negatif
Just hold on to the Blibli application	Netral
Original items. Satisfied with shopping at Blibli	Positif

Tabel 2 memperlihatkan hasil *labeling* sentimen dari ulasan aplikasi yang telah diterjemahkan dan diproses menggunakan *library Vader Sentiment*. Analisis ini mencerminkan keakuratan *Vader* dalam membedakan nuansa emosional dari teks yang bervariasi. Ulasan pertama, yang menyatakan aplikasi sebagai "*ugly, expensive, not like an orange store, not like a discount like the store next to the tree, it's complicated*", secara tepat diberi label sebagai negatif. Kata-kata kunci seperti "*ugly*" dan "*expensive*" langsung menunjukkan konotasi negatif yang kuat, yang menunjukkan ketidakpuasan pengguna. Ulasan kedua, "*Just hold on to the Blibli application*", diberikan label netral, menunjukkan bahwa teks tersebut tidak secara eksplisit mengungkapkan sentimen positif atau negatif, melainkan sebuah saran atau komentar yang lebih netral. Akhirnya, ulasan ketiga, "*Original items. Satisfied with shopping at Blibli*", dengan jelas menunjukkan kepuasan, tercermin dari penggunaan kata "*satisfied*" yang menandai sentimen positif. Penggunaan *Vader Sentiment* dalam penelitian ini membuktikan efektivitasnya dalam menginterpretasi dan mengklasifikasikan sentimen berdasarkan kata kunci yang ada dalam konteksnya, memberikan gambaran yang jelas tentang persepsi pengguna terhadap produk atau layanan.

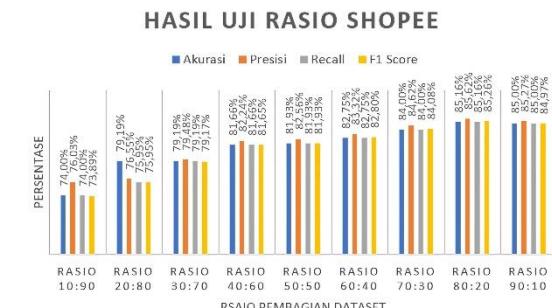
### 3.3. Evaluasi Model

Proses evaluasi diawali dengan menganalisis aplikasi *Shopee*, dilanjutkan dengan *Tokopedia*, *Lazada* dan *Blibli*. Pada tahap ini, kami akan menyajikan analisis dari hasil evaluasi pengujian rasio dan *confusion matrix*.

#### 3.3.1. Shopee

Untuk memfasilitasi analisis komparatif performa model pada berbagai proporsi pembagian data, Gambar 5 menyajikan

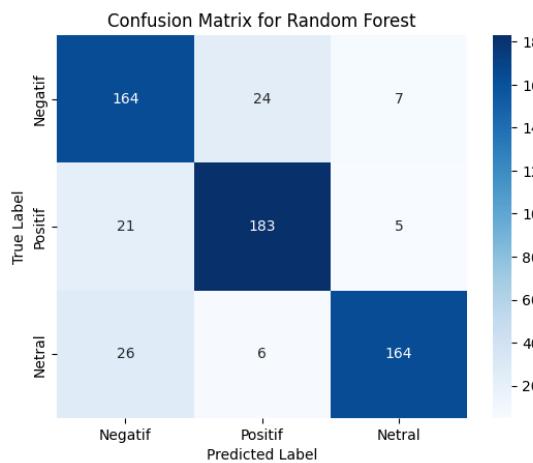
kompilasi hasil pengujian dari setiap rasio yang telah dievaluasi dalam konteks aplikasi *Shopee*.



Gambar 5. Hasil Uji Rasio Shopee

Gambar 5 menyajikan hasil pengujian analisis sentimen pada aplikasi *Shopee* menggunakan algoritma *Random Forest* dengan berbagai rasio pembagian data latih dan data uji. Hasil tersebut mengungkapkan bahwa rasio pembagian 80:20 memberikan performa terbaik dalam hal akurasi, presisi, *recall*, dan *F1 Score*. Akurasi tertinggi sebesar 85,16%, presisi tertinggi sebesar 85,62%, *recall* tertinggi sebesar 85,16% dan *F1 Score* tertinggi sebesar 85,26 dicapai pada rasio 80:20, sedangkan nilai terendah ditemukan pada rasio 10:90 dengan akurasi 74,00%, presisi 76,03%, *recall* 74,00% dan *F1 Score* 73,89%. Hal ini mengindikasikan bahwa model bekerja lebih efektif ketika jumlah data latih lebih besar dibandingkan data uji. Oleh karena itu, rasio pembagian 80:20 dapat dianggap sebagai pilihan optimal untuk melakukan analisis sentimen aplikasi *Shopee* pada penelitian ini.

Hasil evaluasi kinerja model klasifikasi pada rasio 80:20 ditunjukkan pada *confusion matrix* yang divisualisasikan pada Gambar 8. Metrik ini memberikan gambaran rinci tentang prediksi yang benar dan salah untuk setiap kelas, memungkinkan kita untuk menganalisis akurasi model secara lebih mendalam.

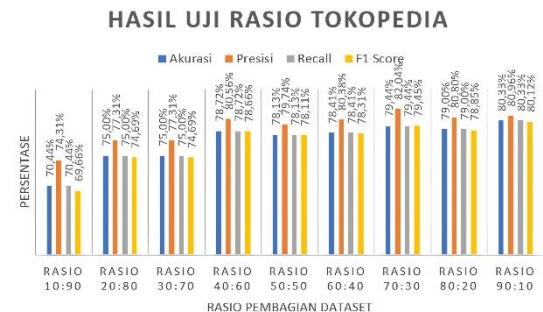


Gambar 6. Confusion Matrix Shopee

Berdasarkan *confusion matrix* yang disajikan pada Gambar 6, kita dapat menganalisis performa model dalam mengklasifikasikan sentimen ke dalam tiga kelas: Negatif, Netral, dan Positif. Untuk kelas Negatif, model berhasil mengklasifikasikan 164 sampel dengan benar sebagai Negatif, 24 sampel Negatif salah diklasifikasikan sebagai Positif, dan 7 sampel Negatif salah diklasifikasikan sebagai Netral. Untuk kelas Positif, model berhasil mengklasifikasikan 183 sampel dengan benar sebagai Positif, 5 sampel Positif salah diklasifikasikan sebagai Netral, dan 21 sampel Positif salah diklasifikasikan sebagai Negatif. Untuk kelas Netral, model berhasil mengklasifikasikan 164 sampel dengan benar sebagai Netral, 26 sampel Netral salah diklasifikasikan sebagai Negatif, dan 6 sampel Netral salah diklasifikasikan sebagai Positif. Hasil analisis ini mengindikasikan bahwa model menunjukkan performa yang memuaskan dalam mengklasifikasikan ketiga kelas sentimen tersebut.

### 3.3.2. Tokopedia

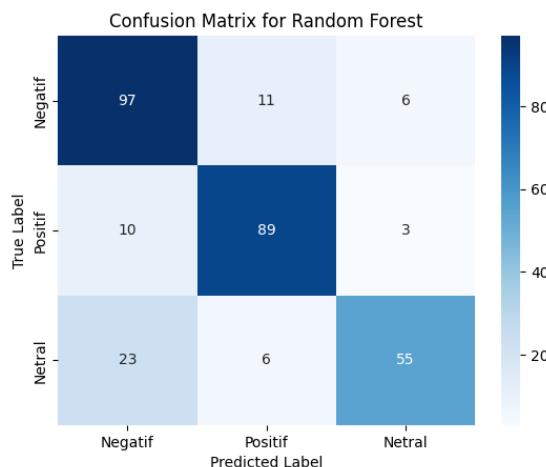
Agar analisis perbandingan kinerja model pada berbagai rasio pembagian data dapat dilakukan dengan mudah, Gambar 7 menyajikan ringkasan hasil pengujian untuk setiap rasio yang telah dievaluasi pada aplikasi *Tokopedia*.



Gambar 7. Hasil Uji Rasio Tokopedia

Gambar 7 menyajikan hasil pengujian analisis sentimen pada aplikasi *Tokopedia* menggunakan algoritma *Random Forest* dengan berbagai rasio pembagian data latih dan data uji. Hasil tersebut mengungkapkan bahwa rasio pembagian 90:10 memberikan performa terbaik dalam hal akurasi, presisi, *recall*, dan *F1 Score*. Akurasi tertinggi sebesar 80,33%, presisi tertinggi sebesar 80,96%, *recall* tertinggi sebesar 80,33% dan *F1 Score* tertinggi sebesar 80,12% dicapai pada rasio 90:10, sedangkan nilai terendah ditemukan pada rasio 10:90 dengan akurasi 70,44%, presisi 74,31%, *recall* 70,44% dan *F1 Score* 69,66%. Hal ini mengindikasikan bahwa model bekerja lebih efektif ketika jumlah data latih lebih besar dibandingkan data uji. Oleh karena itu, rasio pembagian 90:10 dapat dianggap sebagai pilihan optimal untuk melakukan analisis sentimen aplikasi *Tokopedia* pada penelitian ini.

Gambar 8 menampilkan *confusion matrix* yang memvisualisasikan hasil evaluasi performa model klasifikasi dengan rasio data 90:10. Metrik ini menyajikan informasi detail mengenai ketepatan dan kesalahan prediksi untuk tiap kategori, sehingga memungkinkan analisis yang lebih komprehensif terhadap tingkat akurasi model.

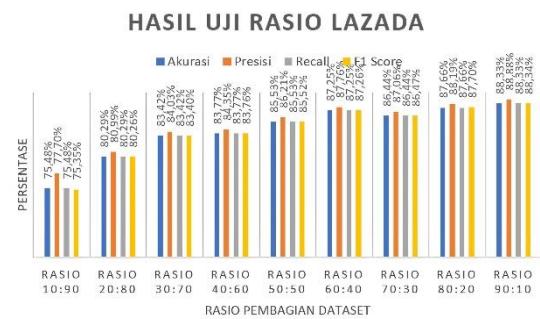


Gambar 8. Confusion Matrix Tokopedia

Berdasarkan *confusion matrix* yang ditampilkan pada Gambar 8, kita dapat mengevaluasi kinerja model dalam mengklasifikasikan sentimen menjadi tiga kategori: Negatif, Netral, dan Positif. Untuk kategori Negatif, model berhasil mengidentifikasi dengan benar sebanyak 97 sampel, meskipun terdapat 11 sampel yang salah terkласifikasi sebagai Positif dan 6 sampel yang salah terklasifikasi sebagai Netral. Pada kategori Positif, model menunjukkan kemampuan yang baik dengan mengklasifikasikan 89 sampel secara akurat, meskipun 3 sampel Positif salah terklasifikasi sebagai Netral dan 10 sampel lainnya salah terklasifikasi sebagai Negatif. Sedangkan untuk kategori Netral, model mampu mengklasifikasikan 55 sampel dengan benar, namun terdapat 23 sampel Netral yang salah diklasifikasikan sebagai Negatif dan 6 sampel yang salah diklasifikasikan sebagai Positif. Dari evaluasi ini, terlihat bahwa model memiliki performa yang cukup baik dalam klasifikasi kategori Negatif, Positif, dan Netral, meskipun beberapa kesalahan klasifikasi masih memerlukan perhatian lebih untuk peningkatan di masa mendatang.

### 3.3.3. Lazada

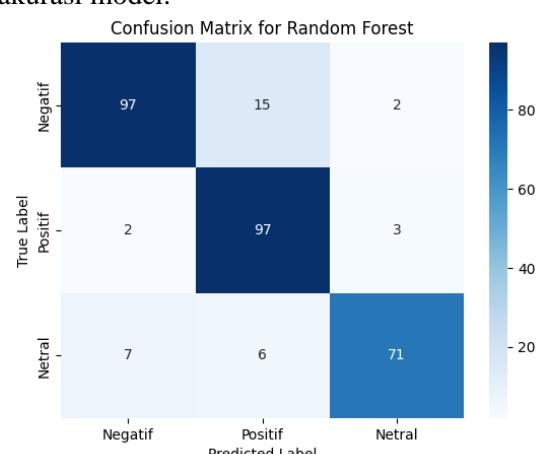
Untuk memfasilitasi analisis komparatif kinerja model pada berbagai rasio pembagian data, Gambar 9 menyajikan ringkasan hasil pengujian untuk setiap rasio yang dievaluasi pada aplikasi *Lazada*.



Gambar 9. Hasil Uji Lazada

Gambar 9 menunjukkan hasil pengujian analisis sentimen pada aplikasi *Lazada* menggunakan algoritma *Random Forest* dengan berbagai rasio pembagian data pelatihan dan pengujian. Hasil tersebut mengungkapkan bahwa rasio pembagian 90:10 memberikan kinerja terbaik dalam hal akurasi, presisi, *recall*, dan *F1 Score*. Akurasi tertinggi yang dicapai adalah 88,88%, dengan presisi 88,88%, *recall* 88,88%, dan *F1 Score* 88,88% pada rasio 90:10. Sebaliknya, performa terendah terjadi pada rasio 10:90 dengan akurasi 75,48%, presisi 77,70%, *recall* 75,48%, dan *F1 Score* 75,35%. Hal ini menunjukkan bahwa model lebih efektif ketika jumlah data pelatihan lebih besar dibandingkan dengan data pengujian. Oleh karena itu, rasio 90:10 dapat dianggap sebagai pilihan optimal untuk analisis sentimen aplikasi *Lazada* dalam penelitian ini.

Gambar 10 memvisualisasikan *confusion matrix* yang menggambarkan kinerja model klasifikasi dengan rasio data 90:10. Metrik ini menyajikan informasi detail mengenai ketepatan dan kesalahan prediksi untuk tiap kategori, sehingga memungkinkan analisis yang lebih komprehensif terhadap tingkat akurasi model.

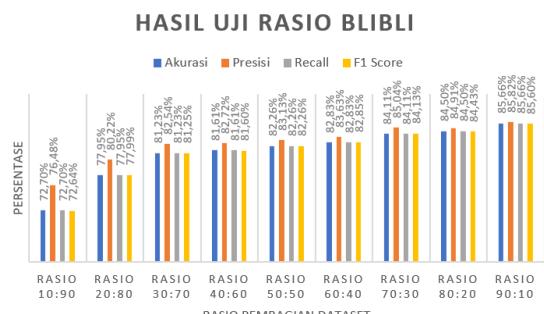


Gambar 10. Confusion Matrix Lazada

Berdasarkan *confusion matrix* yang ditampilkan pada Gambar 10, kita dapat mengevaluasi kinerja model dalam mengklasifikasikan sentimen ke dalam tiga kategori: Negatif, Netral, dan Positif. Untuk kategori Negatif, model berhasil mengidentifikasi dengan benar 97 sampel sebagai Negatif, meskipun terdapat 15 sampel Negatif yang salah diklasifikasikan sebagai Positif dan 2 sampel yang salah diklasifikasikan sebagai Netral. Dalam kategori Positif, model juga mengklasifikasikan 97 sampel dengan benar sebagai Positif, dengan kesalahan klasifikasi sebanyak 3 sampel Positif yang dianggap sebagai Netral dan 2 sampel Positif yang salah diklasifikasikan sebagai Negatif. Sedangkan untuk kategori Netral, model mampu mengklasifikasikan 71 sampel dengan benar, namun masih terdapat 7 sampel Netral yang salah diklasifikasikan sebagai Negatif dan 6 sampel yang salah diklasifikasikan sebagai Positif. Analisis ini menunjukkan bahwa model memiliki kinerja yang baik dalam mengklasifikasikan sentimen pada ketiga kategori tersebut, yaitu Negatif, Positif, dan Netral.

### 3.3.4. Blibli

Gambar 11 memvisualisasikan hasil pengujian dari berbagai rasio pembagian data yang telah dievaluasi dalam konteks aplikasi *Blibli*.

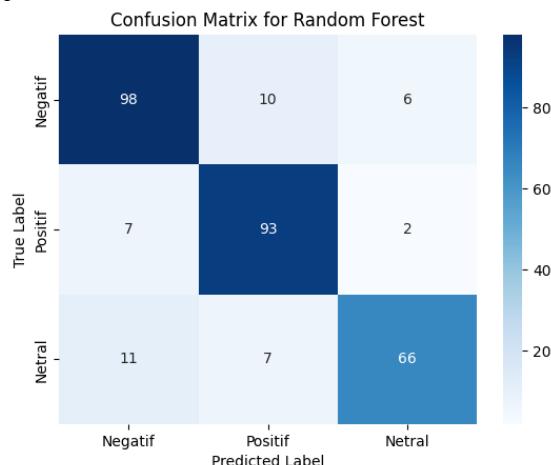


Gambar 11. Hasil Uji Rasio *Blibli*

Gambar 11 menunjukkan hasil pengujian analisis sentimen pada aplikasi *Blibli* menggunakan algoritma *Random Forest* dengan berbagai rasio pembagian data pelatihan dan pengujian. Hasilnya menunjukkan bahwa rasio 90:10 memberikan kinerja terbaik dalam hal akurasi, presisi, *recall*, dan *F1 Score*. Akurasi tertinggi dicapai sebesar 85,66%, dengan presisi 85,82%, *recall* 85,66%, dan *F1 Score* 85,60 pada rasio 90:10. Sebaliknya,

performa terendah ditemukan pada rasio 10:90 dengan akurasi 72,70%, presisi 76,48%, *recall* 72,70%, dan *F1 Score* 72,64%. Ini menunjukkan bahwa model lebih efektif ketika jumlah data pelatihan lebih besar dibandingkan data pengujian. Oleh karena itu, rasio 90:10 dapat dianggap sebagai pilihan optimal untuk analisis sentimen aplikasi *Blibli* dalam penelitian ini.

*Confusion matrix* yang ditampilkan pada Gambar 12 memberikan gambaran visual tentang performa model klasifikasi menggunakan pembagian data 90:10. Metrik ini menguraikan secara rinci keakuratan dan kesalahan prediksi untuk setiap kelompok, memfasilitasi evaluasi mendalam terhadap presisi model secara keseluruhan.



Gambar 12. *Confusion Matrix Blibli*

Berdasarkan *confusion matrix* yang ditampilkan pada Gambar 12, kita dapat mengevaluasi kinerja model dalam mengklasifikasikan sentimen ke dalam tiga kategori: Negatif, Netral, dan Positif. Untuk kategori Negatif, model berhasil mengidentifikasi dengan benar 98 sampel sebagai Negatif, meskipun ada 10 sampel yang salah diklasifikasikan sebagai Positif dan 6 sampel yang salah diklasifikasikan sebagai Netral. Pada kategori Positif, model dengan tepat mengklasifikasikan 93 sampel, sementara 2 sampel Positif salah diklasifikasikan sebagai Netral dan 7 sampel lainnya salah diklasifikasikan sebagai Negatif. Sementara itu, untuk kategori Netral, model berhasil mengklasifikasikan 66 sampel dengan benar, tetapi 11 sampel salah diklasifikasikan sebagai Negatif dan 7 sampel salah diklasifikasikan sebagai Positif. Analisis ini menunjukkan

bawa model memiliki kinerja yang baik dalam mengklasifikasikan sentimen ke dalam kategori Negatif, Positif, dan Netral.

### 3.4. Perbandingan Hasil Pengujian

Berdasarkan hasil pengujian yang dilakukan, peneliti menyajikan visualisasi data dalam Tabel 3 sebagai berikut.

Tabel 3. Perbandingan Hasil Pengujian

Objek	Evaluasi			
	Akurasi	Presisi	Recall	F1 Score
<i>Shopee</i>	85,16%	85,62%	85,16%	85,26%
<i>Tokopedia</i>	80,33%	80,96%	80,33%	80,12%
<i>Lazada</i>	<b>88,33%</b>	<b>88,88%</b>	<b>88,33%</b>	<b>88,34%</b>
<i>Blibli</i>	85,66%	85,82%	85,66%	85,60%

Berdasarkan Tabel 7, aplikasi *Lazada* menunjukkan kinerja pengujian terbaik dibandingkan aplikasi lainnya dengan akurasi sebesar 88,33%, presisi 88,88%, recall 88,33%, dan F1 score 88,34%. Aplikasi *Blibli* berada di posisi kedua dengan akurasi 85,66%, presisi 85,82%, recall 85,66%, dan F1 score 85,60%. *Shopee* mengikuti dengan akurasi 85,16%, presisi 85,62%, recall 85,16%, dan F1 score 85,26%. Sementara itu, *Tokopedia* berada di urutan terakhir dalam evaluasi performa dengan akurasi 80,33%, presisi 80,96%, recall 80,33%, dan F1 score 80,12%.

### 3.5. Diskusi Perbandingan Penelitian Terdahulu

Setelah menyajikan hasil penelitian, perbandingan dengan studi-studi sebelumnya akan disajikan dalam Tabel 4. Tabel ini akan menampilkan temuan utama dan membandingkan berbagai aspek yang menjadi fokus analisis pada penelitian-penelitian terdahulu. Untuk memberikan evaluasi yang komprehensif terhadap kinerja model, objek, dan metodologi yang digunakan, tabel tersebut akan mencakup matriks evaluasi seperti akurasi (A), presisi (P), recall (R), dan F1-score (F).

Tabel 4. Diskusi Penelitian Terdahulu

Penelitian	Objek	Metode	Evaluasi			
			A	P	R	F
[22]	<i>Shop ee</i>	<i>Naive Bayes Classifier</i>	79 %	-	-	-
[14]	<i>E-com</i>	<i>Conv olutio</i>	86 %	-	-	-

	<i>merc e</i>	<i>nal Neura l Network</i>					
[16]	<i>Marketplace</i>	<i>Naïve Bayes</i>	84,3 3%	84,3 2%	84,3 4%	-	
		<i>Logistic Regression</i>	84,5 8%	84,6 6%	84,6 3%	-	
[17]	<i>Shop ee</i>	<i>Naïve Bayes</i>	82,1 0%	69,4 0%	51,9 9%	51,9 9%	

Berdasarkan Tabel 4 Perbandingan Penelitian Terdahulu menunjukkan bahwa penelitian oleh [22] menganalisis sentimen ulasan aplikasi *Shopee* di *Google Play Store* menggunakan algoritma *Naïve Bayes* dengan 4849 dataset yang terbagi dalam dataset positif, netral, dan negatif, dan menghasilkan akurasi 79%. Sebagai perbandingan, penelitian yang diajukan menggunakan dataset seimbang dengan total 3000 dataset per objek dan memperoleh akurasi tertinggi 88,33%. Penelitian oleh [14] menggunakan algoritma *Convolutional Neural Network (CNN)* untuk menganalisis sentimen review aplikasi *e-commerce* di *Google Play Store*, menghasilkan akurasi 86%, sedikit lebih rendah dari penelitian yang diajukan. [16] membandingkan algoritma *Naïve Bayes* dan *Logistic Regression* dalam menganalisis sentimen aplikasi *marketplace* di *Google Play Store*, menunjukkan *Logistic Regression* lebih unggul dengan akurasi 84,58%, namun masih lebih rendah dari penelitian yang diajukan dengan akurasi 88,33%. [17] menganalisis sentimen ulasan pengguna aplikasi *Shopee* di *Google Play Store* menggunakan algoritma *Naïve Bayes*, menghasilkan akurasi 82,10%, presisi 69,40%, recall 51,99%, dan F1 Score 51,99%, yang juga lebih rendah dibandingkan dengan penelitian yang diajukan. Penelitian yang diajukan menganalisis sentimen ulasan aplikasi *e-commerce* populer di Indonesia seperti *Shopee*, *Tokopedia*, *Lazada*, dan *Blibli* di *Google Play Store* menggunakan pendekatan berbasis leksikon untuk *auto labeling* dan algoritma *Random Forest* untuk klasifikasi, menunjukkan kinerja terbaik dengan akurasi

tertinggi 88,33%, presisi 88,88%, *recall* 88,33%, dan *F1 Score* 88,34%.

#### 4. KESIMPULAN

- a. Penelitian ini menggunakan pendekatan berbasis leksikon (*Lexicon-based Approach*) untuk melakukan pelabelan otomatis pada dataset yang diperoleh dari hasil scraping di platform *Google Play Store*. Dengan menggunakan algoritma *Random Forest*, peneliti melakukan analisis sentimen dan membandingkan empat *e-commerce* terkemuka di Indonesia: *Shopee*, *Tokopedia*, *Lazada*, dan *Blibli*. Evaluasi performa model menunjukkan akurasi, presisi, *recall*, dan *F1 Score* yang terukur dengan baik.
- b. Pada aplikasi *Shopee*, performa terbaik diperoleh dengan rasio pembagian data latih dan data uji 80:20, mencapai akurasi 85,16%, presisi 85,62%, *recall* 85,16%, dan *F1 Score* 85,26%. Untuk *Tokopedia*, performa terbaik dicapai pada rasio 90:10 dengan akurasi 80,33%, presisi 80,96%, *recall* 80,33%, dan *F1 Score* 80,12%. Sementara itu, *Lazada* menunjukkan performa terbaik dari semua *e-commerce* dengan akurasi tertinggi 88,33%, presisi 88,88%, *recall* 88,33%, dan *F1 Score* 88,34%. Terakhir, *Blibli* mencapai performa terbaik pada rasio 90:10 dengan akurasi 85,66%, presisi 85,82%, *recall* 85,66%, dan *F1 Score* 85,60%.
- c. Hasil penelitian menunjukkan bahwa rasio pembagian data latih dan data uji (10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20, dan 90:10) memiliki pola yang konsisten dalam memengaruhi kinerja model. Model bekerja lebih efektif ketika jumlah data latih lebih besar dibandingkan data uji, yang terlihat dari performa model pada berbagai rasio.
- d. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan.

Mengingat performa model belum mencapai akurasi 90%, disarankan agar penelitian selanjutnya mempertimbangkan penggunaan jenis leksikon lain, seperti *InSet* yang menggunakan Bahasa Indonesia, atau mengeksplorasi algoritma *machine learning* lain seperti *Naïve Bayes*, serta algoritma *deep learning* seperti *LSTM* atau *BERT* untuk meningkatkan performa model. Kompleksitas dan ambiguitas bahasa yang digunakan di Indonesia juga perlu diperhatikan untuk mendapatkan hasil yang lebih optimal.

#### DAFTAR PUSTAKA

- [1] M. Azizah and M. Aswad, "Pengaruh Belanja Online Pada E-Commerce Shopee Terhadap Perilaku Konsumtif Generasi Millennial di Blitar," *Jurnal Cendekia Ilmiah*, vol. 1, no. 4, 2022.
- [2] A. Asyifah, A. Syafi'i, Hanipah, and S. Ispiyani, "PENGEMBANGAN APLIKASI E-COMMERCE UNTUK PENINGKATAN PENJUALAN ONLINE," *Action Research Literate*, vol. 7, no. 1, pp. 1–6, 2023.
- [3] D. Witro, M. Izazi Nurjaman, D. Ayu, and C. Sofwan Al-Rasyid, "Kontestasi Marketplace di Indonesia pada Era Pandemi: Analisis Strategi Promosi Tokopedia, Shopee, Bukalapak, Lazada, dan Blibli dalam Pemulihan Ekonomi Nasional," *Jurnal Kajian Ekonomi dan Keuangan Syariah*, vol. 2, no. 2, pp. 33–42, 2021.
- [4] A. E. Saragih, M. F. Bagaskara, and Mulyadi, "PERLINDUNGAN HUKUM TERHADAP KONSUMEN DALAM TRANSAKSI E-COMMERCE," *Jurnal Kajian Hukum dan Pendidikan Kewarganegaraan*, vol. 2, no. 2, 2023, [Online]. Available: <http://jurnal.anfa.co.id>
- [5] R. A. Simbolon, R. Aryansyah, and Nurbaiti, "Pengaruh analisis swot dalam e-business (studi kasus e-business shopee di kalangan mahasiswa Universitas Islam Negeri Sumatera Utara)," *Insight Management Journal*, vol. 2, no. 2, pp. 54–61, 2022.
- [6] Mardiana, "PERGESERAN INVESTASI PERUSAHAAN TELEKOMUNIKASI JEPANG DI E-COMMERCE INDONESIA (2013-2019)," 2022.
- [7] D. A. C. R. Devi and N. L. P. Indiani, "PENGARUH E-SERVICE QUALITY, PROMOSI DAN KEPERCAYAAN

- TERHADAP KEPUTUSAN PEMBELIAN PELANGGAN PADA LAZADA DI ERA COVID-19,” Jurnal Ekonomi Syariah, vol. 6, no. 1, pp. 53–67, 2023.
- [8] A. Hadidu, E. Horas, and S. Ypup, “Analysis Of The Influence Of Service Quality And Brand Reputation On Purchase Intentions Of Blibli.Com Customers Analisis Pengaruh Service Quality Dan Reputasi Merek Terhadap Purchase Intention Pelanggan Blibli.Com,” 2022. [Online]. Available: <http://journal.yrpipku.com/index.php/msej>
- [9] M. Adhitya Wardhana, Y. Mustaqim, and N. A. Firmansyah, “ANALISIS MANAJEMEN KETEPATAN WAKTU(JIT), KUALITAS LAYANAN ELEKTRONIK, TERHADAP KEPUASAN PELANGGAN ELEKTRONIK DI PADA E-COMMERCE BLIBLI,” 2023.
- [10] F. Khoirunnisa and S. Topiq, “ANALISIS SENTIMEN TERHDAP KEPERCAYAAN MASYARAKAT PADA PROSES PENEGAK HUKUM DI INDONESIA DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES,” Jurnal Informatika dan Teknik Elektro Terapan, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4683.
- [11] I. Zulfahmi, J. Williem Iskandar, P. V Medan, S. Tuan, and D. Serdang, “Analisis Sentimen Aplikasi PLN Mobile Menggunakan Metode Decission Tree,” Jurnal Penelitian Rumpun Ilmu Teknik (JUPRIT), vol. 3, no. 1, pp. 11–21, 2024, doi: 10.55606/juprit.v3i1.3096.
- [12] M. Iqbal Ahmadi et al., “SENTIMENT ANALYSIS ONLINE SHOP ON THE PLAY STORE USING METHOD SUPPORT VECTOR MACHINE (SVM),” Seminar Nasional Informatika, 2020.
- [13] M. Afdal and L. Waroka, “Shopee Application Review Classification Using Probabilistic Neural Network Algorithm And K-Nearest Neighbor,” Indonesian Journal of Informatic Research and Software Engineering, vol. 2, no. 1, pp. 49–58, 2022.
- [14] F. A. Khatami, B. Irawan, and C. Setianingsih, “ANALISIS SENTIMEN TERHADAP REVIEW APLIKASI LAYANAN E-COMMERCE MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK,” e-Proceeding of Engineering, vol. 7, no. 2, pp. 4559–4566, Aug. 2020.
- [15] S. W. Iriananda, R. P. Putra, and K. S. Nugroho, “ANALISIS SENTIMEN DAN ANALISIS DATA EKSPLORATIF ULASAN APLIKASI MARKETPLACE GOOGLE PLAYSTORE,” 2021.
- [16] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, “Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling,” Journal of Information Systems and Informatics, vol. 5, no. 3, pp. 915–927, Aug. 2023, doi: 10.51519/journalisi.v5i3.539.
- [17] N. C. Agustina, D. Herlina Citra, W. Purnama, C. Nisa, and A. Rozi Kurnia, “The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store,” MALCOM : Indonesian Journal of Machine Learning and Computer Science, vol. 2, no. 1, pp. 47–54, Apr. 2022.
- [18] F. Amaliah, I. Kadek, and D. Nuryana, “Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter,” Journal of Informatics and Computer Science, vol. 03, 2022.
- [19] M. R. A. Giffari, “ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI TANGERANG LIVE MENGGUNAKAN LATENT DIRICHLET ALLOCATION DAN NAÏVE BAYES,” Universitas Islam Negeri Syarif Hidayatullah Jakarta, 2022. Accessed: Jan. 12, 2024. [Online]. Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/65391>
- [20] A. Khairunnisa, “PERBANDINGAN MODEL RANDOM FOREST DAN XGBOOST UNTUK PREDIKSI KEJAHATAN KESUSILAAN DI PROVINSI JAWA BARAT,” JIKO (Jurnal Informatika dan Komputer), vol. 7, no. 2, p. 202, Sep. 2023, doi: 10.26798/jiko.v7i2.799.
- [21] Chen et al., “Pre-evacuation Time Estimation Based Emergency Evacuation Simulation in Urban Residential Communities,” Int J Environ Res Public Health, vol. 16, p. 4599, Jul. 2019, doi: 10.3390/ijerph16234599.
- [22] A. Tri Rizky, Rianto, and A. Irham Gufroni, “Implementation of the Naive Bayes Classifier for Sentiment Analysis of Shopee E-Commerce Application Review Data on the Google Play Store,” 2023.