Remik: Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 10, Nomor 2, Agustus 2026
http://doi.org/10.33395/remik.v10i3.16503
e-ISSN : 2541-1330
p-ISSN : 2541-1332

Analisis Sentimen Aplikasi *E-Commerce* Blibli pada Google *Playstore* Menggunakan Model RoBERTa

[1,*]Wisnu Sri Utomo, [2]Nani Mintarsih, [3]Yuli Maharetta Arianti, [4]Risma Rahmalia Fitriani

[1*,4] Prodi Informatika, Fakultas Teknologi Industri
[2] Prodi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi
[3] Prodi Program Diploma Manajemen Informatika
Universitas Gunadarma, Jakarta, Indonesia

wisnusri31@gmail.com

*Penulis Korespondensi
Diajukan : 30/06/2026
Diterima : 15/07/2026
Dipublikasi : 16/07/2026

ABSTRAK

Perkembangan *e-commerce* di Indonesia telah mengubah perilaku masyarakat dalam berbelanja secara daring, sehingga ulasan pengguna pada *Google Play Store* menjadi sumber informasi penting untuk mengevaluasi kualitas layanan aplikasi. Salah satu aplikasi *e-commerce* yang banyak digunakan adalah Blibli, sehingga diperlukan analisis sentimen untuk memahami persepsi pengguna secara lebih komprehensif. Penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi Blibli dengan mengklasifikasikan opini ke dalam kategori positif, negatif, dan netral menggunakan model *RoBERTa* (*Robustly Optimized BERT Pretraining Approach*). Penelitian menerapkan metode *Knowledge Discovery in Database* (*KDD*) yang meliputi tahapan *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Dataset yang digunakan terdiri atas 7.998 ulasan berbahasa Indonesia yang diperoleh melalui teknik *web scraping* pada *Google Play Store*. Evaluasi model dilakukan menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa sebelum dilakukan *data augmentation*, model memperoleh akurasi sebesar 81% dengan *F1-score* 0,85 pada kelas positif dan 0,83 pada kelas negatif, namun belum mampu mengidentifikasi kelas netral. Setelah diterapkan *data augmentation*, akurasi menjadi 78%, tetapi kemampuan klasifikasi menjadi lebih seimbang dengan *F1-score* 0,84 untuk kelas positif, 0,77 untuk kelas negatif, dan 0,68 untuk kelas netral. Dengan demikian, model *RoBERTa* terbukti efektif untuk analisis sentimen ulasan aplikasi Blibli, sedangkan penerapan *data augmentation* mampu meningkatkan keseimbangan performa klasifikasi pada seluruh kategori sentimen meskipun terjadi sedikit penurunan akurasi.

Kata Kunci: Analisis Sentimen, *RoBERTa*, Blibli, *Google Play Store*, KDD

I. PENDAHULUAN

Perkembangan teknologi digital telah mendorong pertumbuhan aplikasi *e-commerce* sebagai media utama dalam aktivitas transaksi jual beli secara daring. Tingkat keberhasilan suatu aplikasi tidak hanya ditentukan oleh kualitas fitur yang ditawarkan, tetapi juga oleh pengalaman pengguna selama memanfaatkan layanan tersebut. Dalam konteks ini, ulasan pengguna pada *Google Play Store* menjadi sumber informasi yang penting untuk mengidentifikasi tingkat kepuasan sekaligus menggambarkan kecenderungan sentimen terhadap suatu aplikasi. Oleh karena itu, analisis terhadap ulasan pengguna memiliki peran strategis sebagai dasar evaluasi dan pengembangan kualitas layanan di tengah persaingan industri *e-commerce* yang semakin kompetitif.

Remik: Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 10, Nomor 2, Agustus 2026
http://doi.org/10.33395/remik.v10i3.16503
e-ISSN : 2541-1330
p-ISSN : 2541-1332

Blibli merupakan salah satu platform *e-commerce* yang banyak digunakan di Indonesia dan menawarkan berbagai produk dari beragam *brand* serta *merchant* terpercaya. Meskipun aplikasi ini telah memperoleh jumlah unduhan yang tinggi dan nilai *rating* yang baik di *Google Play Store*, masih terdapat perbedaan antara jumlah unduhan dan pengguna aktif. Kondisi tersebut menunjukkan perlunya kajian lebih lanjut untuk memahami persepsi pengguna melalui analisis sentimen terhadap ulasan yang diberikan.

Analisis sentimen merupakan bagian dari *Natural Language Processing* (*NLP*) yang digunakan untuk mengklasifikasikan opini pengguna ke dalam kategori positif, negatif, atau netral. Seiring berkembangnya teknologi *deep learning*, model berbasis *Transformer* semakin banyak dimanfaatkan karena mampu memahami konteks bahasa secara lebih efektif dibandingkan metode klasifikasi konvensional. Salah satu model yang memiliki performa tinggi adalah *RoBERTa* (*Robustly Optimized BERT Pretraining Approach*), yang dikembangkan melalui optimalisasi proses pelatihan *BERT* sehingga mampu menghasilkan representasi semantik teks yang lebih baik. Dengan karakteristik tersebut, *RoBERTa* dinilai sesuai untuk meningkatkan akurasi klasifikasi sentimen pada ulasan pengguna aplikasi Blibli.

II. STUDI LITERATUR

E-Commerce

E-commerce merupakan sistem perdagangan elektronik yang memungkinkan transaksi jual beli barang maupun jasa melalui jaringan internet. Perkembangan *e-commerce* telah mengubah perilaku konsumen dengan memberikan kemudahan dalam proses pencarian produk, transaksi pembayaran, hingga layanan purna jual (Fauzan, 2025). Tingginya persaingan antarplatform mendorong perusahaan untuk terus meningkatkan kualitas layanan berdasarkan masukan dan pengalaman pengguna yang tercermin dalam ulasan pelanggan. Oleh karena itu, analisis terhadap ulasan pengguna menjadi salah satu pendekatan yang efektif untuk mengevaluasi kualitas layanan dan meningkatkan kepuasan pelanggan.

Blibli

Blibli merupakan salah satu platform *e-commerce* terbesar di Indonesia yang menyediakan berbagai kategori produk dari *merchant* resmi maupun mitra usaha. Sebagai aplikasi yang memiliki jutaan pengguna, Blibli menerima ribuan ulasan pada Google Play Store yang berisi pengalaman pengguna terhadap fitur, kualitas layanan, kecepatan pengiriman, maupun sistem transaksi. Ulasan tersebut menjadi sumber informasi yang penting dalam mengidentifikasi persepsi pengguna sehingga dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam pengembangan layanan berbasis data (Rahman et al., 2025).

Data mining

Data mining merupakan proses menemukan pola, hubungan, maupun informasi yang bernilai dari sekumpulan data berukuran besar menggunakan berbagai teknik statistik, *machine learning*, dan kecerdasan buatan. Dalam penelitian analisis sentimen, *data mining* dimanfaatkan untuk mengekstraksi pengetahuan dari data teks melalui tahapan *Knowledge Discovery in Database* (*KDD*) yang meliputi *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Pendekatan tersebut memungkinkan proses klasifikasi dilakukan secara sistematis sehingga menghasilkan informasi yang dapat mendukung pengambilan keputusan (Buulolo, 2021).

Analisis Sentimen

Analisis sentimen merupakan salah satu bidang dalam *Natural Language Processing* (*NLP*) yang bertujuan mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau emosi yang terkandung dalam data teks ke dalam kategori positif, negatif, maupun netral. Penerapan analisis sentimen semakin luas pada berbagai platform digital karena mampu memberikan informasi mengenai persepsi pengguna terhadap suatu produk atau layanan berdasarkan ulasan yang diberikan (Sulistiowati & Santoso, 2025). Pada lingkungan *e-commerce*, analisis sentimen berperan

Remik: Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 10, Nomor 2, Agustus 2026
http://doi.org/10.33395/remik.v10i3.16503
e-ISSN : 2541-1330
p-ISSN : 2541-1332

penting sebagai dasar evaluasi kualitas layanan, pengalaman pengguna, serta pengambilan keputusan strategis oleh perusahaan.

RoBERTa

RoBERTa merupakan pengembangan dari model *Bidirectional Encoder Representations from Transformers* (*BERT*) yang dioptimalkan melalui strategi *pre-training* yang lebih efektif sehingga mampu menghasilkan representasi kontekstual yang lebih baik. Model ini menghilangkan tugas *Next Sentence Prediction*, menggunakan ukuran *mini-batch* yang lebih besar, serta memanfaatkan data pelatihan yang lebih banyak sehingga meningkatkan kemampuan klasifikasi teks. Dalam analisis sentimen berbahasa Indonesia, *RoBERTa* menunjukkan performa yang tinggi karena mampu memahami hubungan semantik antar kata secara kontekstual (Hidayat et al., 2025).

***Knowledge Discovery in Database* (KDD)**

Metode *Knowledge Discovery in Database* (*KDD*) merupakan pendekatan sistematis dalam proses penemuan pengetahuan dari sekumpulan data. Tahapan *KDD* meliputi *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Pendekatan ini banyak digunakan dalam penelitian *text mining* karena mampu menghasilkan proses pengolahan data yang terstruktur sebelum dilakukan klasifikasi menggunakan algoritma *machine learning* maupun *deep learning* (Widaningsih, 2022).

Deep Learning

Deep learning merupakan cabang dari *machine learning* yang menggunakan jaringan saraf tiruan berlapis (*deep neural network*) untuk mempelajari representasi data secara otomatis. Dalam bidang *Natural Language Processing* (*NLP*), *deep learning* mampu memahami hubungan kontekstual antar kata sehingga menghasilkan performa yang lebih baik dibandingkan metode klasifikasi konvensional (Rachman & Santoso, 2021). Model berbasis *Transformer*, seperti *BERT* dan *RoBERTa*, menjadi salah satu pendekatan yang banyak digunakan dalam analisis sentimen karena memiliki kemampuan memahami makna kalimat secara dua arah (*bidirectional contextual representation*) sehingga meningkatkan akurasi klasifikasi teks (Lee et al., 2021).

Confusion Matrix

Confusion Matrix merupakan metode evaluasi yang digunakan untuk mengukur performa model klasifikasi dengan membandingkan hasil prediksi terhadap label sebenarnya. Matriks ini menghasilkan empat komponen utama, yaitu *True Positive* (*TP*), *True Negative* (*TN*), *False Positive* (*FP*), dan *False Negative* (*FN*), yang selanjutnya digunakan untuk menghitung metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* (Chicco & Jurman, 2022). Penggunaan *Confusion Matrix* memberikan gambaran yang lebih komprehensif mengenai kemampuan model dalam mengklasifikasikan setiap kelas sentimen, khususnya pada dataset yang memiliki distribusi kelas tidak seimbang (Ida et al., 2023).

Google Play Store

Google Play Store menyediakan ulasan pengguna yang menggambarkan pengalaman nyata selama menggunakan aplikasi. Informasi tersebut menjadi sumber data yang kaya untuk mengukur tingkat kepuasan pengguna, mengidentifikasi permasalahan layanan, serta mengetahui kebutuhan pengguna secara lebih objektif (Andrian & Giap, 2026). Oleh karena itu, ulasan pada Google Play Store banyak dimanfaatkan sebagai dataset dalam penelitian analisis sentimen karena bersifat dinamis dan mencerminkan kondisi aktual penggunaan aplikasi.

WordCloud

WordCloud merupakan teknik visualisasi data teks yang menampilkan kata berdasarkan frekuensi kemunculannya dalam suatu dokumen atau kumpulan dokumen. Semakin sering suatu kata muncul, maka ukuran kata pada visualisasi akan semakin besar. Dalam penelitian analisis sentimen, *WordCloud* digunakan untuk membantu mengidentifikasi kata-kata dominan pada

Remik: Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 10, Nomor 2, Agustus 2026
http://doi.org/10.33395/remik.v10i3.16503
e-ISSN : 2541-1330
p-ISSN : 2541-1332

masing-masing kategori sentimen sehingga memudahkan interpretasi karakteristik opini pengguna (Tambunan & Hapsari, 2021). Visualisasi ini tidak digunakan sebagai metode klasifikasi, melainkan sebagai pendukung analisis eksploratif terhadap data teks.

III. METODE

Metode penelitian yang digunakan dalam studi ini adalah *Knowledge Discovery in Database* (KDD). Pendekatan ini terdiri dari lima tahapan utama, yaitu *Data Selection, Preprocessing Data, Transformation, Data Mining*, dan *Evaluation*.

Gambar 1. Alur Penelitian

1. *Data Selection*

Tahap awal dilakukan dengan mengumpulkan data ulasan pengguna aplikasi Blibli melalui teknik *web scraping* menggunakan pustaka *google-play-scraper* pada lingkungan *Google Colab*. Sebanyak 7.998 ulasan berbahasa Indonesia berhasil dikumpulkan dan disimpan dalam format *CSV* sebagai dataset penelitian.

2. *Preprocessing Data*

Tahap *preprocessing* bertujuan meningkatkan kualitas data sebelum proses klasifikasi. Tahapan yang dilakukan meliputi:

- *Cleaning*, yaitu menghapus karakter khusus, tanda baca, angka, *URL*, dan simbol yang tidak diperlukan.
- *Case folding*, yaitu mengubah seluruh huruf menjadi huruf kecil.
- *Tokenizing*, yaitu memisahkan kalimat menjadi kumpulan kata (*token*).
- *Filtering*, yaitu menghilangkan *stopwords* yang tidak memiliki kontribusi terhadap analisis sentimen.
- *Stemming*, yaitu mengubah setiap kata menjadi bentuk dasarnya.
- Penyusunan kembali hasil *stemming* menjadi kalimat yang siap diproses oleh model.

3. *Data Transformation*

Pada tahap ini dilakukan transformasi data teks menjadi representasi numerik yang dapat diproses oleh model *RoBERTa*. Label sentimen dikodekan menggunakan *Label Encoder* dengan kategori negatif = 0, netral = 1, dan positif = 2. Dataset kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan metode *train-test split*. Selanjutnya, seluruh teks dikonversi menjadi *input_ids* dan *attention_mask* melalui *tokenizer RoBERTa*.

4. *Data Mining*

Proses klasifikasi dilakukan menggunakan model *RoBERTa* (*Robustly Optimized BERT Pretraining Approach*). Model dilatih menggunakan parameter *learning rate* sebesar 2×10^{-5}, *batch size* 16, dan empat *epoch*. Untuk mengatasi ketidakseimbangan distribusi kelas, diterapkan teknik *data augmentation* pada kelas sentimen netral sehingga model mampu mengenali seluruh kategori sentimen secara lebih proporsional.

5. *Evaluation*

Kinerja model dievaluasi menggunakan *Confusion Matrix* dan *Classification Report*. Metrik evaluasi yang digunakan meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Hasil evaluasi sebelum dan sesudah penerapan *data augmentation* dibandingkan untuk mengetahui pengaruh penyeimbangan data terhadap performa model klasifikasi sentimen.

IV. HASIL DAN PEMBAHASAN

Penelitian ini menerapkan metode *Knowledge Discovery in Database* (*KDD*) yang terdiri atas

Remik: Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 10, Nomor 2, Agustus 2026
http://doi.org/10.33395/remik.v10i3.16503
e-ISSN : 2541-1330
p-ISSN : 2541-1332

tahapan *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Setiap tahapan menghasilkan keluaran yang saling berkaitan hingga diperoleh model klasifikasi sentimen menggunakan *RoBERTa*. Model ini digunakan untuk mengklasifikasikan ulasan ke dalam tiga kategori sentimen: positif, negatif, dan netral.

Hasil *Data Selection*

Tahap *data selection* dilakukan melalui proses *web scraping* terhadap ulasan pengguna aplikasi Blibli pada Google Play Store menggunakan pustaka *google-play-scraper* berbasis Python di Google Colab. Dari proses tersebut diperoleh sebanyak 7.998 ulasan berbahasa Indonesia yang kemudian disimpan dalam format *.csv* sebagai dataset penelitian. Dataset ini menjadi dasar dalam proses klasifikasi sentimen ke dalam tiga kelas, yaitu positif, negatif, dan netral. Jumlah data yang diperoleh dinilai cukup representatif untuk menggambarkan persepsi pengguna terhadap aplikasi Blibli. Penggunaan teknik *web scraping* memungkinkan pengumpulan data secara otomatis sehingga proses akuisisi data menjadi lebih efisien dan mampu memperoleh ulasan dalam jumlah besar.

```python
from google_play_scraper import reviews, Sort
import csv
import pandas as pd

scrapreview, _ = reviews(
    'blibli.mobile.commerce',
    lang='id',
    country='id',
    sort=Sort.MOST_RELEVANT,
    count=8000
)

# Simpan ke CSV
with open('ReviewBlibli.csv', mode='w', newline='', encoding='utf-8') as file:
    writer = csv.writer(file)
    writer.writerow(['Review'])
    for review in scrapreview:
        writer.writerow([review['content']])

# Simpan ke DataFrame dan CSV
app_reviews_df = pd.DataFrame(scrapreview)
app_reviews_df.to_csv('ulasan_aplikasi.csv', index=False)

print(f"Scrapping berhasil! Jumlah ulasan: {len(scrapreview)} disimpan ke 'ReviewBlibli.csv'")
```

Gambar 2. Pengambilan *dataset*

Preprocessing Data

Tahap *preprocessing* bertujuan meningkatkan kualitas data dengan menghilangkan komponen yang tidak relevan serta mengubah teks mentah menjadi data yang siap diproses oleh model klasifikasi. Proses ini dilakukan melalui beberapa tahapan, yaitu *cleaning text*, *case folding*, *tokenizing*, *filtering*, *stemming*, dan *to sentence*. Pada tahap *cleaning text*, karakter khusus, tanda baca, angka, tautan, *hashtag*, serta karakter baris baru dihapus untuk mengurangi *noise* pada data. Selanjutnya, *case folding* diterapkan dengan mengubah seluruh huruf menjadi huruf kecil (*lowercase*) sehingga format penulisan menjadi seragam. Tahap *tokenizing* kemudian memisahkan kalimat menjadi kumpulan kata (*token*) agar setiap kata dapat diproses secara individual. Proses berikutnya adalah *filtering*, yaitu menghilangkan *stopwords* yang tidak memberikan kontribusi signifikan terhadap penentuan sentimen. Setelah itu, dilakukan *stemming* untuk mengembalikan setiap kata ke bentuk dasarnya sehingga variasi kata memiliki representasi yang konsisten. Tahap terakhir, *to sentence*, menyusun kembali hasil *stemming* menjadi kalimat utuh yang siap digunakan sebagai masukan bagi model *RoBERTa* pada proses klasifikasi sentimen.

Hasil *Transformation*

Pada tahap *transformation*, setiap label sentimen diubah menjadi representasi numerik menggunakan *LabelEncoder* dengan ketentuan negatif = 0, netral = 1, dan positif = 2. Selanjutnya dataset dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan *train_test_split*. Proses transformasi juga melibatkan *tokenizer RoBERTa* untuk menghasilkan input_ids dan attention_mask sebagai masukan utama model.

Remik: Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 10, Nomor 2, Agustus 2026
http://doi.org/10.33395/remik.v10i3.16503
e-ISSN : 2541-1330
p-ISSN : 2541-1332

```
[ ] df_utama[['Review', 'text_stemming', 'Rating', 'Sentimen']].head(10)
```

	Review	text_stemming	Rating	Sentimen
0	Proses dan pengiriman barangnya lama, sudah gi...	proses kirim barang gitu batal kembali dana ma...	1	negatif
1	Ternyata pengembalian dana disini itu prosesny...	kembali dana proses aplikasi belah aplikasi be...	1	negatif
2	Coba Blibli punya Expedisi sendiri dong sepert...	coba blibli expedisi aplikasi orange biar gk p...	5	positif
3	aplikasi tampilan sudah user friendly. barang ...	aplikasi tampil user friendly barang lengakap ...	4	positif
4	saya sangat KECEWA dengan pelayanan Blibli yan...	kecewa layan blibli batal pesan qris dsb hak b...	1	negatif
5	Min coba perbaiki sistem checkoutnya, sekali c...	min coba baik sistem checkoutnya checkout bara...	3	netral
6	untuk pertama kali coba langsung di kasih peng...	kali coba langsung kasih alam kecewa isi pulsa...	1	negatif
7	apk apaan' bisa bobol saldo saya, saya mempuny...	apk bobol saldo jt tarik apk blibi blibi saldo...	1	negatif
8	Pengiriman lama, lemot, ga ada tindakan yang c...	kirim lambat tindak cepat laku blibli inisemua...	1	negatif
9	Jangan ada yang beli lewat bli bli. system nya...	beli bli bli system kampung beli shopee dana p...	1	negatif

Gambar 3. *Output* Tranformasi Data

Transformasi data menghasilkan representasi numerik yang mampu mempertahankan konteks kalimat sehingga model *Transformer* dapat memahami hubungan antar kata secara lebih efektif dibandingkan representasi teks konvensional.

Hasil *Data Mining*

Tahap *data mining* dilakukan dengan menerapkan model *RoBERTa* (*Robustly Optimized BERT Pretraining Approach*) untuk mengklasifikasikan ulasan pengguna aplikasi Blibli ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Sebelum proses pelatihan dimulai, setiap label sentimen dikonversi ke dalam bentuk numerik menggunakan teknik *label encoding*, dengan ketentuan negatif = 0, netral = 1, dan positif = 2. Selanjutnya, dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 untuk memperoleh model yang mampu melakukan generalisasi secara optimal. Pada tahap berikutnya, setiap ulasan diproses menggunakan *tokenizer RoBERTa* sehingga menghasilkan representasi numerik berupa *input_ids* dan *attention_mask* sebagai masukan bagi model. Model *RoBERTa* yang telah melalui proses *pre-trained* kemudian diimplementasikan menggunakan pustaka *Transformers* dengan parameter *learning rate* sebesar 2×10^{-5}, *batch size* 16, dan empat *epoch*. Seluruh proses pelatihan dijalankan pada lingkungan Google Colab yang memanfaatkan akselerasi GPU guna meningkatkan efisiensi komputasi. Kinerja model dievaluasi menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur kemampuan klasifikasi pada setiap kategori sentimen. Mengingat distribusi data yang tidak seimbang, terutama pada kelas netral yang memiliki jumlah sampel paling sedikit, diterapkan teknik *data augmentation* untuk meningkatkan proporsi data pada kelas tersebut. Penerapan teknik ini menghasilkan performa klasifikasi yang lebih merata, ditunjukkan oleh meningkatnya kemampuan model dalam mengenali seluruh kategori sentimen, khususnya kelas netral, meskipun terjadi sedikit penurunan pada nilai akurasi keseluruhan.

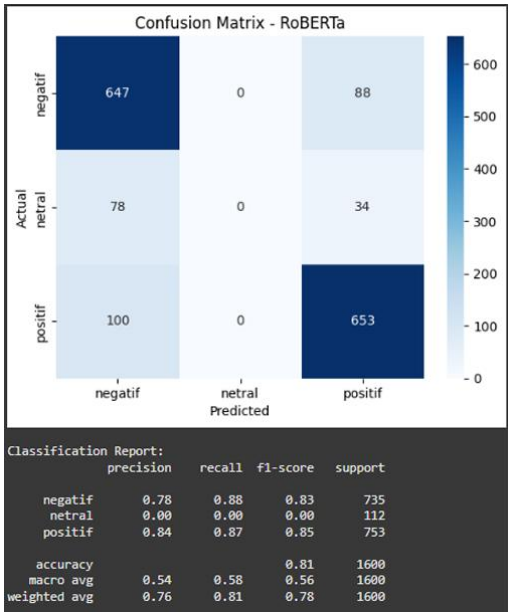
Hasil Evaluasi

Tahap evaluasi bertujuan untuk menilai kemampuan model *RoBERTa* dalam mengklasifikasikan sentimen ulasan setelah proses pelatihan selesai. Pengukuran kinerja dilakukan menggunakan metrik *precision*, *recall*, dan *F1-score* yang diperoleh melalui *classification report*, serta didukung oleh analisis *confusion matrix* untuk menggambarkan hasil klasifikasi pada setiap kategori sentimen.

1. *Data Augmentation*

Hasil evaluasi model sebelum *Data Augmentation* menunjukkan kinerja yang baik dalam mengidentifikasi ulasan dengan sentimen positif dan negatif. Hal tersebut ditunjukkan oleh nilai akurasi sebesar 81%, dengan *F1-score* masing-masing 0,85 pada kelas positif dan 0,83 pada kelas negatif. Sebaliknya, performa model pada kelas netral masih sangat rendah dengan *F1-score* sebesar 0,00, yang mengindikasikan bahwa model belum mampu mengenali ulasan netral secara memadai. Kondisi tersebut dipengaruhi oleh distribusi data yang tidak seimbang, di mana jumlah data pada kelas netral jauh lebih sedikit dibandingkan kelas positif dan negatif. Akibatnya, model lebih banyak mempelajari karakteristik dua kelas mayoritas sehingga kemampuan generalisasi terhadap kelas netral menjadi terbatas.

Remik: Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 10, Nomor 2, Agustus 2026
http://doi.org/10.33395/remik.v10i3.16503
e-ISSN : 2541-1330
p-ISSN : 2541-1332



Classification Report:

	precision	recall	f1-score	support
negatif	0.78	0.88	0.83	735
netral	0.00	0.00	0.00	112
positif	0.84	0.87	0.85	753
accuracy			0.81	1600
macro avg	0.54	0.58	0.56	1600
weighted avg	0.76	0.81	0.78	1600

Gambar 4. Data Sebelum Augmentasi

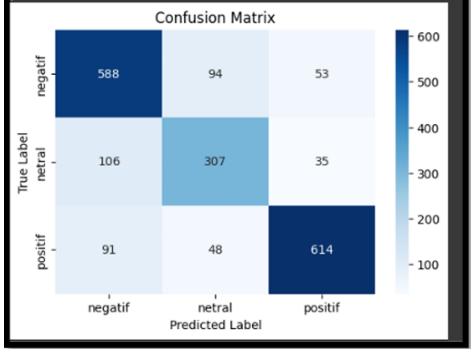
Data yang telah di augmentasi menunjukkan, penerapan *data augmentation* pada kelas netral memberikan perubahan terhadap performa model. Meskipun akurasi keseluruhan menurun menjadi 78%, hasil evaluasi menunjukkan distribusi kinerja yang lebih merata pada seluruh kategori sentimen. Peningkatan paling signifikan terjadi pada kelas netral, dengan nilai *F1-score* yang meningkat dari 0,00 menjadi 0,68. Hasil tersebut menunjukkan bahwa model telah mampu mengenali sebagian besar ulasan netral yang sebelumnya sulit diklasifikasikan. Pada kelas negatif diperoleh *F1-score* sebesar 0,77 dengan nilai *recall* 0,80, sedangkan kelas positif tetap mempertahankan performa terbaik dengan *F1-score* sebesar 0,84 dan *recall* 0,82. Meskipun terjadi sedikit penurunan *recall* pada kelas positif dibandingkan sebelum augmentasi, peningkatan nilai *macro average F1-score* dari 0,56 menjadi 0,77 menunjukkan bahwa kemampuan klasifikasi model menjadi lebih seimbang pada ketiga kelas sentimen. Temuan ini mengindikasikan bahwa penerapan *data augmentation* efektif dalam mengurangi dampak *class imbalance*, sehingga model tidak hanya mempertahankan performa pada kelas mayoritas, tetapi juga meningkatkan kemampuan dalam mengenali kelas minoritas secara lebih konsisten.

Classification Report:

	precision	recall	f1-score	support
negatif	0.75	0.80	0.77	735
netral	0.68	0.69	0.68	448
positif	0.87	0.82	0.84	753
accuracy			0.78	1936
macro avg	0.77	0.77	0.77	1936
weighted avg	0.78	0.78	0.78	1936

Confusion Matrix:

	negatif	netral	positif
negatif	588	94	53
netral	106	307	35
positif	91	48	614

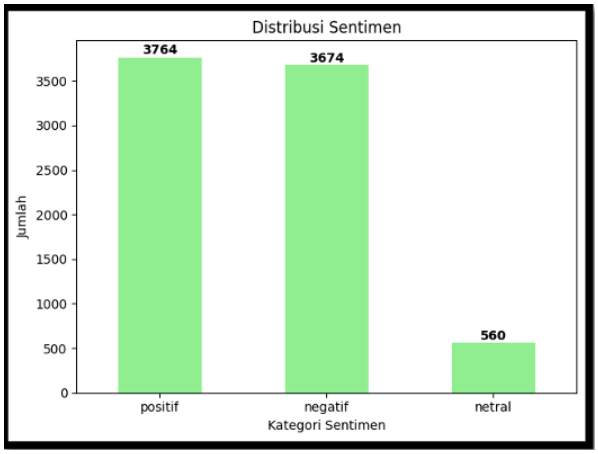


Gambar 5. Data Sesudah Augmentasi

Remik: Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 10, Nomor 2, Agustus 2026
http://doi.org/10.33395/remik.v10i3.16503
e-ISSN : 2541-1330
p-ISSN : 2541-1332

2. Visualisasi Distribusi Sentimen

Visualisasi distribusi sentimen menunjukkan bahwa ulasan pengguna didominasi oleh sentimen positif sebanyak 3.764 data, diikuti sentimen negatif sebanyak 3.674 data, sedangkan sentimen netral hanya berjumlah 560 data. Distribusi tersebut memperlihatkan bahwa persepsi pengguna terhadap aplikasi Blibli cenderung terbagi antara pengalaman positif dan negatif, sedangkan ulasan yang bersifat netral relatif sedikit. Ketimpangan distribusi ini menjadi salah satu alasan utama diterapkannya teknik *data augmentation*.



Gambar 6. Visualisasi Diagram

3. *WorldCloud*

Visualisasi *WordCloud* digunakan untuk mengidentifikasi kata-kata yang paling sering muncul pada masing-masing kategori sentimen. Pada gambar 7 menunjukkan *WordCloud* sentimen positif memperlihatkan dominasi kata-kata yang berkaitan dengan kepuasan pengguna terhadap kualitas layanan, kemudahan penggunaan aplikasi, dan pengalaman berbelanja.

Gambar 7. *Wordcloud* Positif

Gambar 8. *Wordcloud* Negatif

Pada gambar 8 menunjukkan sentimen negatif, kata-kata yang sering muncul berkaitan dengan kendala transaksi, proses pengiriman, maupun performa aplikasi. Hal tersebut menunjukkan bahwa aspek pelayanan dan stabilitas sistem masih menjadi perhatian utama pengguna.

Sementara itu, *wordcloud* sentimen netral didominasi oleh kata-kata yang bersifat informatif tanpa menunjukkan kecenderungan emosi tertentu. Visualisasi ini memperkuat hasil klasifikasi bahwa setiap kategori memiliki karakteristik kosakata yang berbeda sehingga dapat dimanfaatkan sebagai dasar evaluasi layanan aplikasi.

Remik: Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 10, Nomor 2, Agustus 2026
http://doi.org/10.33395/remik.v10i3.16503
e-ISSN : 2541-1330
p-ISSN : 2541-1332

Gambar 9. *Wordcloud* Netral

V. KESIMPULAN

Penelitian ini menganalisis sentimen ulasan pengguna aplikasi *e-commerce* Blibli yang diperoleh dari Google Play Store dengan memanfaatkan model *RoBERTa* (*Robustly Optimized BERT Pretraining Approach*). Sebanyak 7.998 ulasan berbahasa Indonesia dikumpulkan melalui teknik *web scraping* dan diproses menggunakan tahapan *Knowledge Discovery in Database* (*KDD*), yang meliputi *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Hasil pengujian menunjukkan bahwa model *RoBERTa* mampu mengelompokkan ulasan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Sebelum penerapan *data augmentation*, model menghasilkan akurasi sebesar 81% dengan nilai *F1-score* masing-masing 0,85 untuk sentimen positif dan 0,83 untuk sentimen negatif, sedangkan kelas netral belum dapat diidentifikasi secara optimal (*F1-score* 0,00). Setelah dilakukan *data augmentation* pada kelas netral, akurasi model menjadi 78%. Meskipun mengalami sedikit penurunan, kemampuan klasifikasi antar kelas menjadi lebih seimbang, yang ditunjukkan oleh nilai *F1-score* sebesar 0,84 pada kelas positif, 0,77 pada kelas negatif, dan 0,68 pada kelas netral. Secara keseluruhan, hasil penelitian mengindikasikan bahwa model *RoBERTa* memiliki kinerja yang baik dalam analisis sentimen ulasan aplikasi *e-commerce*. Penerapan *data augmentation* terbukti meningkatkan kemampuan model dalam mengenali kelas minoritas sehingga menghasilkan performa klasifikasi yang lebih seimbang. Temuan ini menegaskan bahwa penanganan ketidakseimbangan distribusi data merupakan faktor penting untuk meningkatkan kemampuan generalisasi model dalam mengklasifikasikan seluruh kategori sentimen secara lebih konsisten.

VI. REFERENSI

Andrian, C., & Giap, Y. C. (2026). Sentiment Analisis Review Aplikasi Ruang Guru Pada Google Playstore Dengan Metode Naïve Bayes. *Jati (Jurnal Mahasiswa Teknik Informatika)*, *10*(1), 236-241. https://doi.org/10.36040/jati.v10i1.16638

Buulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi*. Deepublish.

Chicco, D., & Jurman, G. (2022). An invitation to greater use of Matthews correlation coefficient in robotics and artificial intelligence. *Frontiers in Robotics and AI*, *9*, 876814. https://doi.org/10.3389/frobt.2022.876814

Dewi, Ida Ayu Murah Cahya, I. Komang Dharmendra, and Ni Wayan Setiasih. 2023. "Analisis Sentimen Review Aplikasi Satu Sehat Mobile Menggunakan Model Sampling Tomek Links." Jurnal Teknologi Informasi Dan Komputer 9(5):497–504.

Fauzan Mahmudi. (2025). Transformasi Kebiasaan Konsumen Dalam Era Digital: Studi Perubahan Pola Belanja Akibat E-Commerce. *Iqtishodiah: Jurnal Ekonomi Dan Perbankan Syariah*, *7*(2), 96–100. https://doi.org/10.62490/iqtishodiah.v7i2.1367

Hidayat, J., Setyowati, C., Amin, M., Bimasakti, K., & Werdana, A. (2025). Deep Learning-based Sentiment Analysis of Public Comments on Military Education Using RoBERTa Algorithm and Rule-Based Hybird Parameters. *Jurnal Media Computer Science*, *4*(2), 277-292. https://doi.org/10.37676/jmcs.v4i2.8769

Lee, J., Cheon, S. U., & Yang, J. (2021). Connectivity-based convolutional neural network for classifying point clouds. *Pattern Recognition*, *112*, 107708. https://doi.org/10.1016/j.patcog.2020.107708

Rachman, F. P., & Santoso, H. (2021). Perbandingan Model Deep Learning untuk Klasifikasi Sentiment Analysis dengan Teknik Natural Language Processing. *Jurnal Teknologi dan Manajemen Informatika*, *7*(2), 103-112. https://doi.org/10.26905/jtmi.v7i2.6506

Rahman, R., Setiawan, N., & Bachtiar, F. (2025). Analisis Sentimen Pengguna Aplikasi Mobile Berbasis Review Pada Platform Blibli Menggunakan Metode Bidirectional Encoder Representations from Transformers (BERT). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *9*(4). https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/14641

Sulistiowati, Y., & Santoso, B. J. (2025). Analisis Sentimen Ulasan Pengguna Aplikasi Mobile SP4N-LAPOR! dengan Pendekatan Machine Learning. *Jurnal Informatika Polinema*, *11*(3), 283–290. https://doi.org/10.33795/jip.v11i3.7189

Tambunan, H. B., & Hapsari, T. W. D. (2021). Analisis Opini Pengguna Aplikasi New PLN Mobile Menggunakan Text Mining. *PETIR*, *15*(1), 121–134. https://doi.org/10.33322/petir.v15i1.1352

Widaningsih, S. (2022, September 14). Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor. *JATISI*, *9*(3), 2598-2611. https://doi.org/10.35957/jatisi.v9i3.859