

Analisis Sentimen untuk Ulasan Produk E-Commerce Shopee Menggunakan BERT

Nadya Sikana¹, Sunaryo Winardi², Gunawan³, Gilbert Fernando Situmorang⁴, Rivaldi Lubis⁵

^{1,2,3,4,5}Universitas Mikroskil, Jl.Thamrin No. 124, 061-4573767

^{1,5}Fakultas Informatika, Teknologi Informasi, Universitas Mikroskil, Medan

^{2,3,4}Fakultas Informatika, Teknik Informatika, Universitas Mikroskil, Medan

e-mail: ¹nadya.sikana@mikroskil.ac.id, ²sunaryo.winardi@mikroskil.ac.id,

³gunawan@mikroskil.ac.id, ⁴gilbert.situmorang@mikroskil.ac.id, ⁵rivaldi.lubis@mikroskil.ac.id

Dikirim: 14-07-2025 | Diterima: 07-08-2025 | Diterbitkan: 31-10-2025

Abstrak

Analisis sentimen sangat penting untuk memahami opini konsumen dan menyempurnakan strategi *e-commerce*. Analisis ini menghadapi tantangan seperti bahasa informal, ambiguitas semantik, dan inkonsistensi antara sentimen tekstual dan peringkat bintang, yang memengaruhi akurasi klasifikasi. Penelitian ini menerapkan model BERT (*Bidirectional Encoder Representations from Transformers*) untuk mengklasifikasikan sentimen dalam ulasan pengguna Shopee. Data dikumpulkan dari penelitian sebelumnya dan menjalani praproses, termasuk tokenisasi, penghapusan *stopword*, dan normalisasi teks. Pendekatan analisis sentimen berbasis leksikon digunakan sebagai dasar perbandingan. Model BERT disempurnakan menggunakan optimasi hiperparameter, mencapai akurasi 83,08%, presisi 82,91%, recall 83,08%, dan *F1-score* 82,87%. Dibandingkan dengan studi sebelumnya yang menggunakan Naïve Bayes dengan N-Gram dan Information Gain, yang mencapai akurasi 92% tetapi presisi lebih rendah (56%), recall (65%), dan *F1-score* (60%), BERT mengungguli dengan metrik evaluasi yang lebih seimbang dan keandalan prediktif yang lebih besar. Hasil ini menunjukkan kemampuan BERT untuk menangkap konteks semantik dua arah, melampaui metode tradisional dalam menangani tugas analisis sentimen yang kompleks.

Kata kunci: Analisis Sentimen, BERT, *E-commerce*, Opini Konsumen

Abstract

Sentiment analysis is essential for understanding consumer opinions and refining e-commerce strategies. It faces challenges such as informal language, semantic ambiguity, and inconsistencies between textual sentiment and star ratings, which impact classification accuracy. This research applies BERT model (Bidirectional Encoder Representations from Transformers) to classify sentiment in Shopee user reviews. Data were collected from previous research and underwent preprocessing, including tokenization, stopword removal, and text normalization. A lexicon-based sentiment analysis approach was used as a baseline for comparison. The BERT model was fine-tuned using hyperparameter optimization, achieving accuracy of 83.08%, precision of 82.91%, recall of 83.08%, and F1-score of 82.87%. In comparison to a previous study using Naïve Bayes with N-Gram and Information Gain, which achieved 92% accuracy but lower precision (56%), recall (65%), and F1-score (60%), BERT outperformed with more balanced evaluation metrics and greater predictive reliability. These results demonstrate BERT's ability to capture bidirectional semantic context, surpassing traditional methods in handling complex sentiment analysis tasks.

Keywords: Sentiment Analysis, BERT, *E-commerce*, Consumer Opinions

1. PENDAHULUAN

Dalam era transformasi digital, platform *e-commerce* telah menjadi sarana penting dalam memfasilitasi transaksi dan membentuk perilaku konsumen. Volume besar konten buatan pengguna, khususnya ulasan produk, menyediakan wawasan penting mengenai kepuasan pelanggan, preferensi, dan ekspektasi mereka [1], [2]. Ulasan ini menjadi dasar bagi strategi berbasis data yang bertujuan meningkatkan kualitas layanan, mengoptimalkan penawaran produk, dan membangun hubungan pelanggan yang lebih kuat [3]. Analisis sentimen, cabang dari *Natural Language Processing* (NLP), memanfaatkan data tersebut untuk mengungkap sentimen, opini, dan persepsi konsumen [4]. Interpretasi yang akurat terhadap informasi ini memungkinkan bisnis mengidentifikasi area yang perlu diperbaiki, menyempurnakan produk, serta meningkatkan penjualan dan loyalitas pelanggan [5], [6].

Tantangan dalam analisis sentimen muncul akibat kompleksitas bahasa alami, khususnya dalam konteks informal seperti ulasan *e-commerce* [7], [8]. Konten buatan pengguna sering kali memuat bahasa gaul, singkatan, dan frasa ambigu yang dapat menyamarkan sentimen sebenarnya [9]. Ketidaksesuaian antara sentimen tekstual dan penilaian numerik, serta nuansa budaya dan konteks, turut memperumit klasifikasi sentimen secara akurat [10], [11]. Metode konvensional, seperti Naïve Bayes dan SVM, kerap gagal menangkap hubungan semantik mendalam dan ketergantungan kontekstual yang terdapat dalam data tersebut [12], [13]. Keterbatasan ini menekankan pentingnya pendekatan lanjutan yang lebih kontekstual.

Kemajuan dalam NLP telah memperkenalkan model berbasis transformer, seperti *Bidirectional Encoder Representations from Transformers* (BERT), yang mampu mengatasi banyak tantangan tersebut [14]. BERT menggunakan pendekatan dua arah untuk memahami konteks kata dalam kalimat, sehingga mampu menangkap nuansa semantik dan ketergantungan secara lebih akurat [15]. Penelitian ini mengeksplorasi potensi BERT dalam meningkatkan akurasi analisis sentimen terhadap ulasan pengguna di Shopee, salah satu platform *e-commerce* terkemuka di Asia Tenggara. Data ulasan yang melimpah di Shopee memiliki nilai strategis bagi intelijen bisnis. Penelitian sebelumnya yang menggunakan metode seperti Naïve Bayes, N-Gram, dan *Information Gain* menunjukkan berbagai keterbatasan. Meskipun mencapai akurasi 92 persen, metode tersebut menghasilkan presisi (56%), *recall* (65%), dan *F1-score* (60%) yang suboptimal, menandakan perlunya pendekatan yang lebih canggih dalam klasifikasi sentimen [16], [17].

Studi ini dibangun dari penelitian terdahulu dengan menerapkan model BERT yang telah disesuaikan untuk data ulasan Shopee. Tujuan utamanya adalah meningkatkan klasifikasi sentimen (positif, netral, atau negatif) melalui kemampuan BERT dalam memahami konteks dua arah dan menangkap kerumitan semantik. Penelitian ini mengatasi keterbatasan metode tradisional dan menggali potensi model berbasis transformer untuk menetapkan tolak ukur baru dalam kinerja analisis sentimen [18], [19].

Kerangka metodologis yang digunakan bertujuan menunjukkan efektivitas model BERT dalam mengatasi tantangan analisis sentimen pada ulasan *e-commerce*. Temuan studi ini memberikan kontribusi bagi bidang *Natural Language Processing* dan menawarkan implikasi praktis bagi dunia bisnis dalam memperoleh wawasan yang dapat ditindaklanjuti dari konten buatan pengguna. Klasifikasi sentimen yang ditingkatkan memungkinkan organisasi memahami pelanggan dengan lebih baik, mengoptimalkan penawaran, serta memperkuat posisi kompetitif di lanskap *e-commerce* yang dinamis [20], [21].

2. TINJAUAN PUSTAKA

2.1 *E-commerce*

E-commerce (perdagangan elektronik) merujuk pada aktivitas pertukaran barang dan jasa yang difasilitasi melalui media elektronik, khususnya jaringan internet dan berbagai platform digital [22]. Dalam dekade terakhir, *e-commerce* telah mengalami perkembangan pesat di kawasan Asia Tenggara, termasuk Indonesia. Hal ini ditandai oleh meningkatnya penetrasi internet, proliferasi perangkat seluler,

serta kemudahan akses terhadap layanan pembayaran digital [23]. Di Indonesia, ekosistem *e-commerce* berkembang menjadi salah satu sektor ekonomi digital yang strategis, dengan aktor-aktor utama seperti Shopee yang menyediakan tidak hanya sarana transaksi, tetapi juga fitur interaktif seperti ulasan konsumen (*product reviews*) yang memungkinkan komunikasi dua arah antara pembeli dan penjual. Ulasan ini menjadi sumber data yang potensial untuk dianalisis guna memahami persepsi konsumen terhadap produk atau layanan tertentu.

2.2 Opini Konsumen

Opini konsumen merupakan bentuk ekspresi individu atas pengalaman mereka dalam mengonsumsi suatu produk atau layanan, yang mencerminkan persepsi, kepuasan, maupun ketidakpuasan terhadap atribut-atribut tertentu dari produk atau layanan tersebut [24]. Dalam konteks *e-commerce*, opini ini umumnya terekam dalam bentuk ulasan teks yang bersifat tidak terstruktur, namun memiliki nilai strategis yang tinggi bagi perusahaan dalam merumuskan kebijakan pemasaran, perbaikan produk, dan peningkatan layanan. Ulasan konsumen tidak hanya berfungsi sebagai medium komunikasi horizontal antar pengguna, tetapi juga sebagai sumber data yang dapat dimanfaatkan melalui pendekatan analisis opini (*opinion mining*) untuk mengidentifikasi sentimen, mengevaluasi performa penjual, serta memengaruhi persepsi dan keputusan calon konsumen.

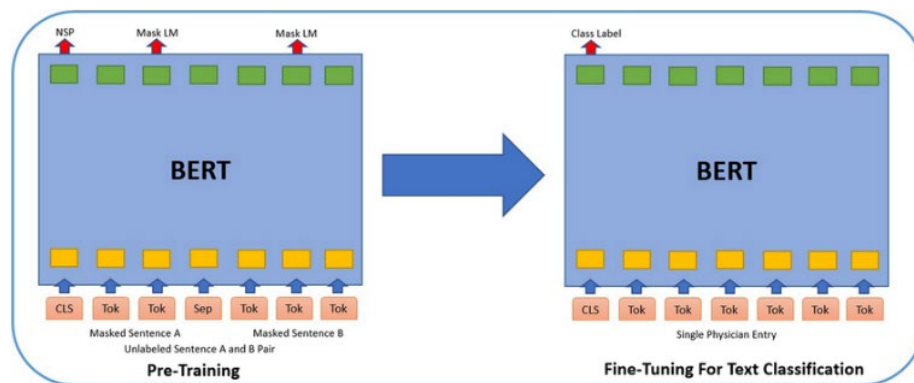
2.3 Analisis Sentimen

Analisis sentimen merupakan subbidang dari *text mining* dan *natural language processing* (NLP) yang berfokus pada identifikasi, ekstraksi, dan klasifikasi opini atau sikap subjektif yang diekspresikan dalam teks, dengan tujuan utama menentukan polaritas sentimen (positif, negatif, atau netral) [25]. Metodologi analisis sentimen berkembang dari pendekatan leksikal dan berbasis aturan hingga ke model berbasis *machine learning* dan *deep learning*, yang memungkinkan peningkatan akurasi dan skalabilitas dalam pemrosesan data teks yang besar. Dalam konteks *e-commerce*, analisis sentimen berperan penting dalam mengotomatisasi interpretasi ribuan ulasan konsumen secara efisien, sehingga dapat memberikan wawasan strategis bagi perusahaan dalam mengevaluasi kualitas produk, responsivitas layanan pelanggan, serta memahami persepsi pasar secara lebih mendalam.

2.4 BERT (*Bidirectional Encoder Representations from Transformers*)

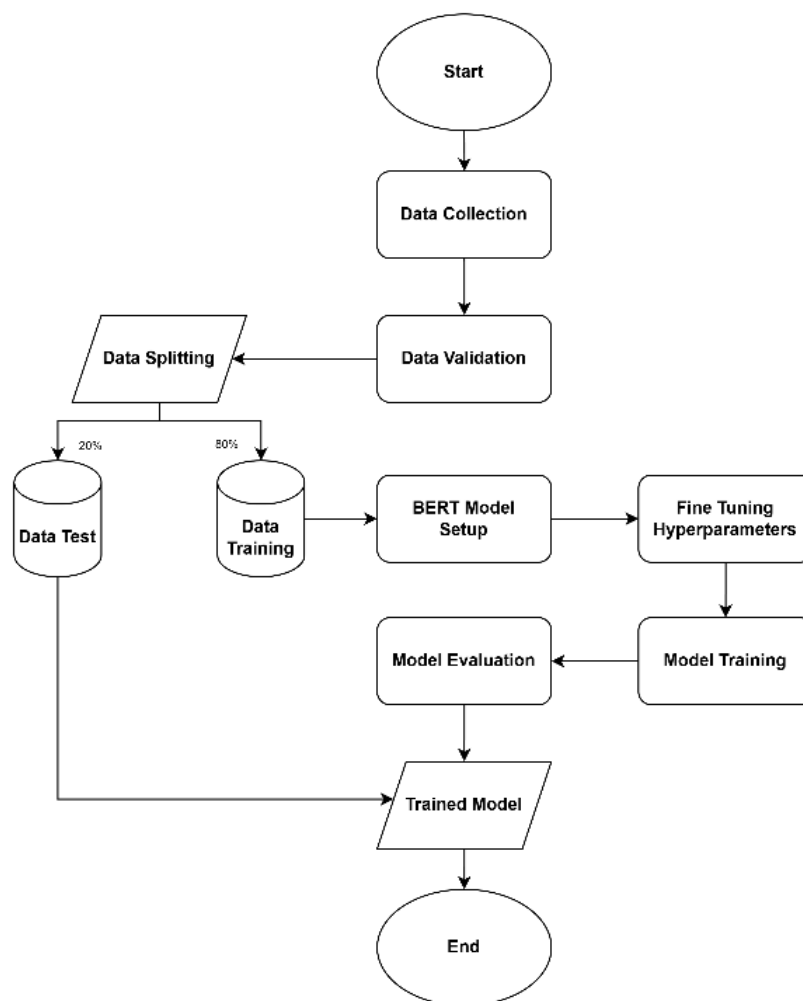
Bidirectional Encoder Representations from Transformers (BERT) adalah model representasi bahasa yang dibangun di atas arsitektur *Transformer encoder* yang sepenuhnya berbasis *self-attention* [26]. Secara teknis, BERT menggunakan pendekatan *bidirectional training* melalui dua skema pra-pelatihan utama: *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP). Dalam MLM, sejumlah token dalam input di-*masking* secara acak, dan model dilatih untuk memprediksi token asli berdasarkan konteks dua arah. Ini berbeda dengan pendekatan sebelumnya seperti GPT yang bersifat *unidirectional*. Dalam NSP, model dilatih untuk memahami hubungan antar kalimat, memungkinkan pemahaman konteks antarsegmen teks yang lebih luas.

Arsitektur BERT terdiri dari tumpukan *encoder transformer* yang masing-masing mengimplementasikan mekanisme *multi-head self-attention* dan *feed-forward network* yang distandarisasi. Representasi vektor hasil dari layer akhir dapat digunakan untuk berbagai tugas NLP dengan menambahkan lapisan output *task-specific*, seperti klasifikasi sentimen atau ekstraksi entitas. Keunggulan representasi kontekstual dua arah ini menjadikan BERT sangat efektif dalam menangkap dependensi panjang dalam teks, menyaring makna semantik dari konteks yang kompleks, dan memberikan performa superior pada *benchmark* NLP. Selain itu, karena sifat *pre-trained then fine-tuned*, BERT dapat dengan mudah disesuaikan (*fine-tuned*) untuk berbagai *downstream tasks* menggunakan sejumlah kecil data berlabel. Ilustrasi model BERT ditunjukkan pada Gambar 1.

Gambar 1. Ilustrasi Model BERT Pre-trained dan *Fine-Tuned* [27]

3. METODE PENELITIAN

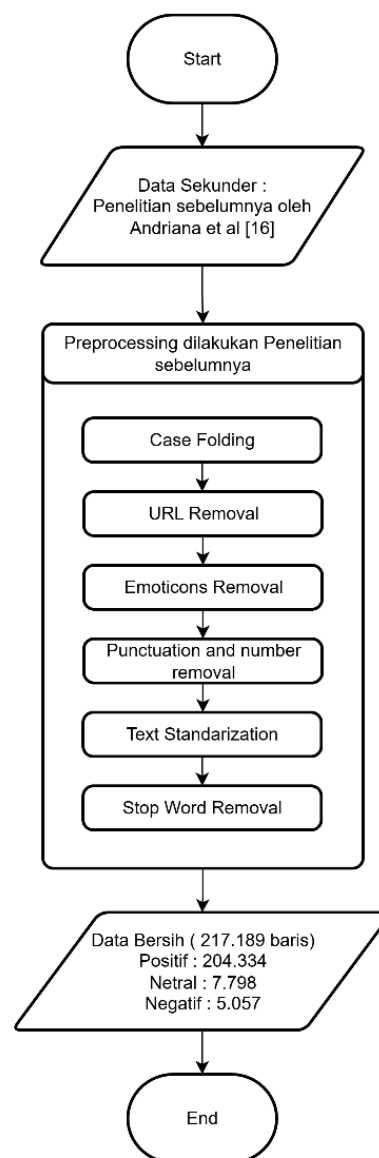
Penelitian ini terdiri dari beberapa tahap utama seperti ditunjukkan pada Gambar 2, yang meliputi pengumpulan data, pengolahan data, pelabelan data menggunakan pendekatan berbasis leksikon, klasifikasi data menggunakan model BERT (*Bidirectional Encoder Representations from Transformers*), dan evaluasi kinerja model.



Gambar 2. Metodologi Penelitian

3.1 Pengumpulan Data

Data ulasan pengguna merupakan data sekunder yang diperoleh dari penelitian sebelumnya yang dilakukan oleh Andriana et al., yang dikumpulkan melalui web scraping menggunakan teknik crawling. Data yang dikumpulkan sejumlah 220.337 baris yang mencakup berbagai kategori produk, *timestamp*, dan *rating* pengguna. Teks ulasan yang diperoleh telah mengalami praproses, termasuk *case folding*, pembersihan teks, normalisasi, penghapusan *stopword*, dan lemmatisasi. Praproses mengubah data yang tidak terstruktur atau *noise* menjadi data siap pakai [28]. Hasilnya, data yang digunakan dalam penelitian ini dalam format yang bersih dan siap untuk analisis lebih lanjut sejumlah 217.189 baris dengan 204.334 data berlabel positif, 7.798 berlabel netral, dan 5.057 data bernilai negatif.



Gambar 3. Pengumpulan Data

3.2 Validasi Data

Setelah data diperoleh, langkah validasi dilakukan untuk memastikan relevansi dan keselarasannya dengan tujuan penelitian. Proses validasi mencakup beberapa langkah penting: memastikan bahwa data ulasan sesuai untuk analisis dan relevan dengan studi yang dimaksud, memverifikasi bahwa data tetap terstruktur secara seragam setelah praproses, memeriksa entri yang hilang atau tidak lengkap, seperti tinjauan tanpa teks atau peringkat, dan mengonfirmasi bahwa teks tinjauan bebas dari elemen yang tidak relevan, seperti simbol, angka, atau karakter khusus yang mungkin terlewatkan selama praproses. Dengan melakukan validasi data, dipastikan bahwa kumpulan data yang digunakan dalam studi ini siap untuk diproses lebih lanjut dan memenuhi standar kualitas yang diperlukan untuk menghasilkan model yang akurat dan andal.

3.3 Pemisahan Data

Setelah diberi label, dataset dibagi menjadi dua bagian untuk memastikan model memiliki kemampuan generalisasi yang baik. Pembagiannya adalah sebagai berikut:

1. Data pelatihan: Bagian ini digunakan untuk melatih model. Dalam penelitian ini, 80% dari total data dialokasikan untuk pelatihan.
2. Data pengujian: Bagian ini digunakan untuk menguji kinerja model yang dilatih. Dalam penelitian ini, 20% dari total data dialokasikan untuk pengujian.

3.4 Implementasi Model BERT

Setelah memberi label dan membagi data, model BERT diterapkan untuk melakukan klasifikasi sentimen. Model dilatih menggunakan data pelatihan, dan pengaturan hiperparameter dilakukan untuk mencapai kinerja yang optimal. Parameter yang diuji dalam studi ini meliputi:

1. *Learning rate*: Mengontrol kecepatan pembaruan bobot selama proses pelatihan.
2. *Batch size*: Menentukan jumlah sampel yang digunakan untuk memperbarui bobot dalam satu iterasi.
3. *Epoch*: Menentukan jumlah siklus pelatihan lengkap pada data pelatihan.

3.5 Evaluasi Model

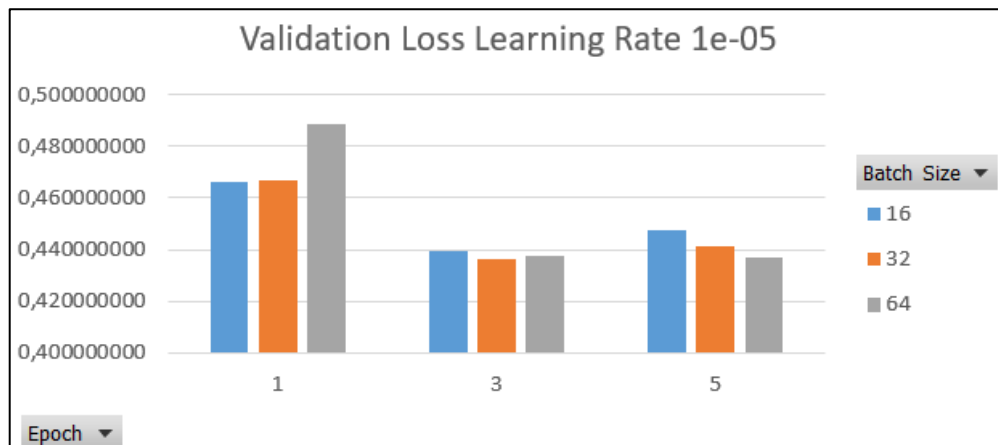
Model BERT akan dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil evaluasi akan dibandingkan dengan hasil analisis sentimen dari pendekatan berbasis leksikon untuk mengukur peningkatan kinerja yang dicapai.

4. HASIL DAN PEMBAHASAN

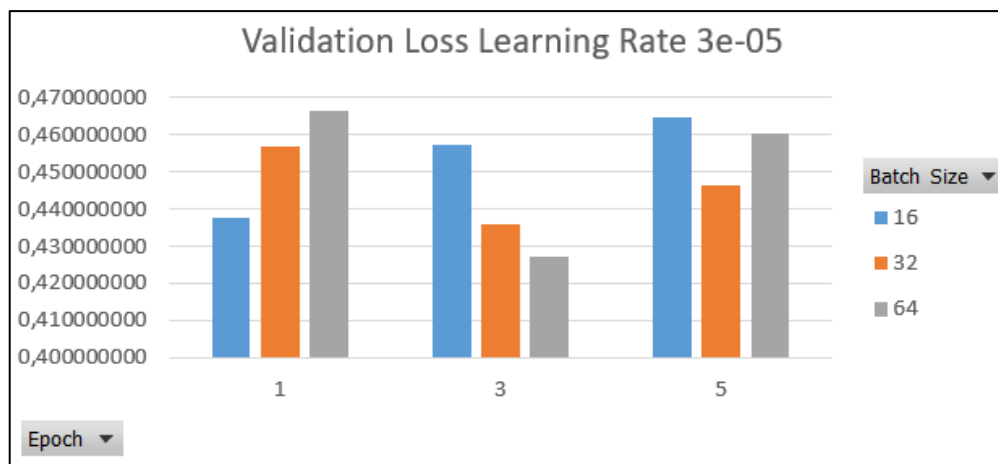
Hasil analisis diperoleh dengan menguji beberapa metrik evaluasi utama, yaitu *validation loss*, akurasi, presisi, *recall*, dan *F1-score*. Pengujian ini bertujuan untuk mengevaluasi kinerja model dari berbagai aspek, sehingga memberikan gambaran menyeluruh tentang efektivitasnya dalam klasifikasi sentimen.

4.1 Uji *Validation Loss*

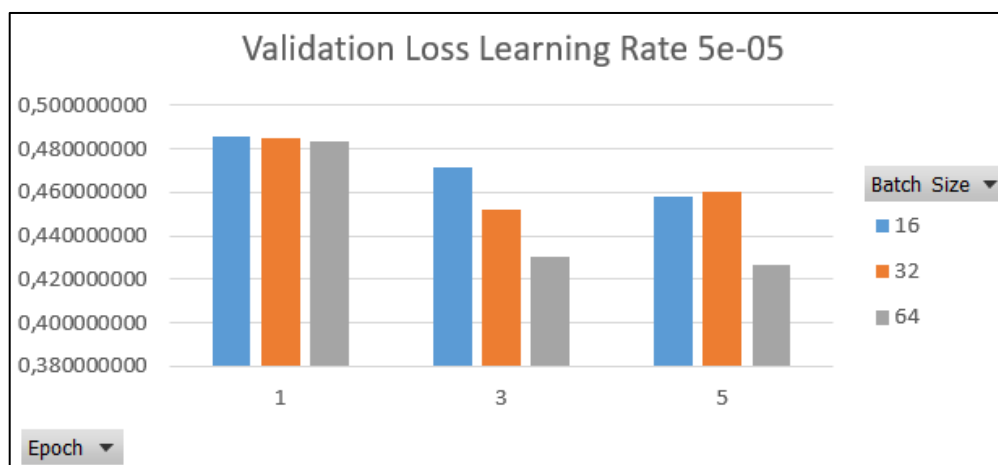
Uji pertama dalam eksperimen ini bertujuan untuk mengevaluasi *validation loss*, salah satu metrik utama untuk mengukur kinerja model selama pelatihan. *Validation loss* mengacu pada perbedaan antara prediksi model dan nilai aktual pada data validasi, yang digunakan untuk memantau apakah model mengalami *overfitting* atau *underfitting* [29]. *Validation loss* yang lebih rendah menunjukkan bahwa model lebih baik dalam memprediksi data yang tidak terlihat dan lebih mampu menghindari *overfitting*, di mana model terlalu beradaptasi dengan data pelatihan dan berkinerja buruk pada data baru. Hasil uji *validation loss* adalah sebagai berikut:



Gambar 4. Validation Loss pada Learning Rate 1e-05



Gambar 5. Validation Loss pada Learning Rate 3e-05



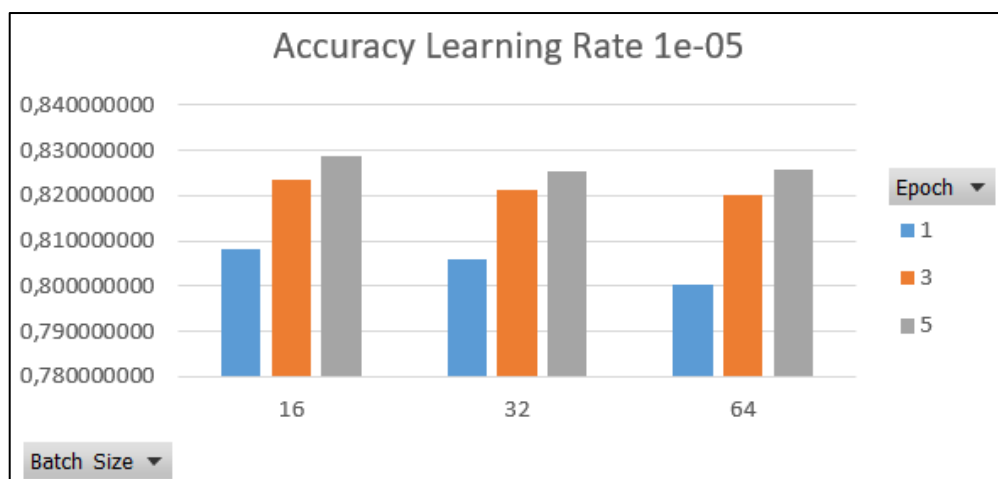
Gambar 6. Validation Loss pada Learning Rate 5e-05

Analisis *validation loss*:

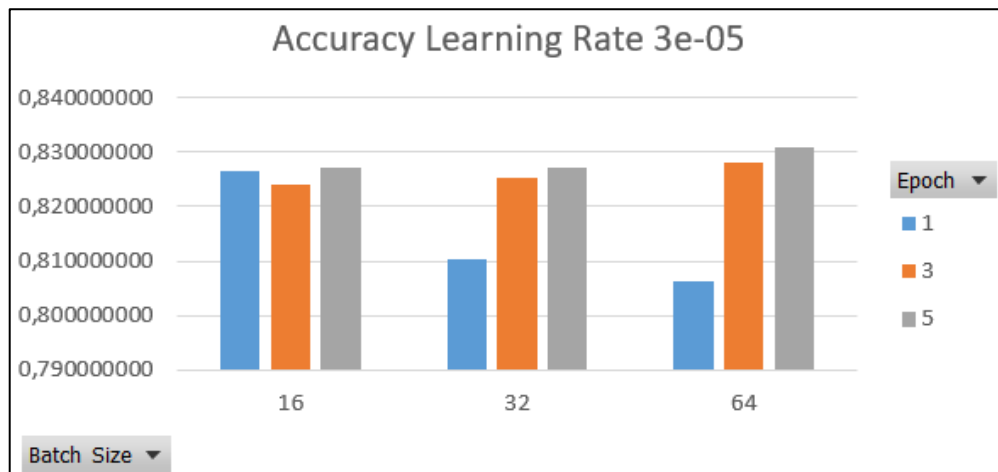
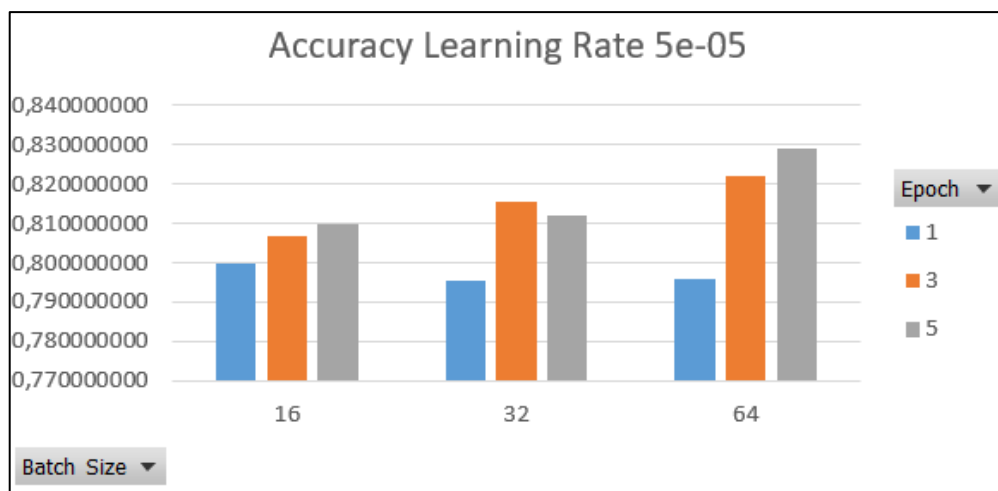
1. Efek *Learning Rate*: *Learning rate* yang terlalu tinggi dapat membuat model sulit untuk dikonvergensi dan menyebabkan nilai *loss* yang tidak stabil. Sebaliknya, *learning rate* yang terlalu rendah dapat memperlambat proses *learning*. Hasil eksperimen menunjukkan bahwa *learning rate* $1e-05$ menghasilkan hasil yang relatif stabil dan konsisten.
2. Efek *Batch Size*: Efek *batch size* pada nilai *loss* tidak konsisten. Terkadang *batch size* yang lebih besar menghasilkan hasil yang lebih baik, tetapi tidak selalu. Ini menunjukkan bahwa pemilihan *batch size* yang optimal sangat bergantung pada arsitektur model dan kumpulan data yang digunakan.
3. Efek *Epoch*: Umumnya, seiring bertambahnya jumlah *epoch*, *validation loss* cenderung menurun, yang menunjukkan bahwa model terus belajar dan meningkatkan akurasi. Namun, dalam beberapa kasus, *loss* mungkin stagnan atau bahkan meningkat setelah beberapa *epoch*, yang menunjukkan *overfitting*.

4.2 Uji Akurasi

Uji coba kedua dalam percobaan ini bertujuan untuk mengevaluasi akurasi, yang merupakan indikator penting dari kinerja model. Akurasi merupakan rasio antara jumlah kejadian yang diprediksi dengan benar dengan jumlah keseluruhan kejadian, yang sering dilaporkan dalam bentuk persentase [30]. Hasil uji akurasi adalah sebagai berikut:



Gambar 7. Akurasi pada *Learning Rate* $1e-05$

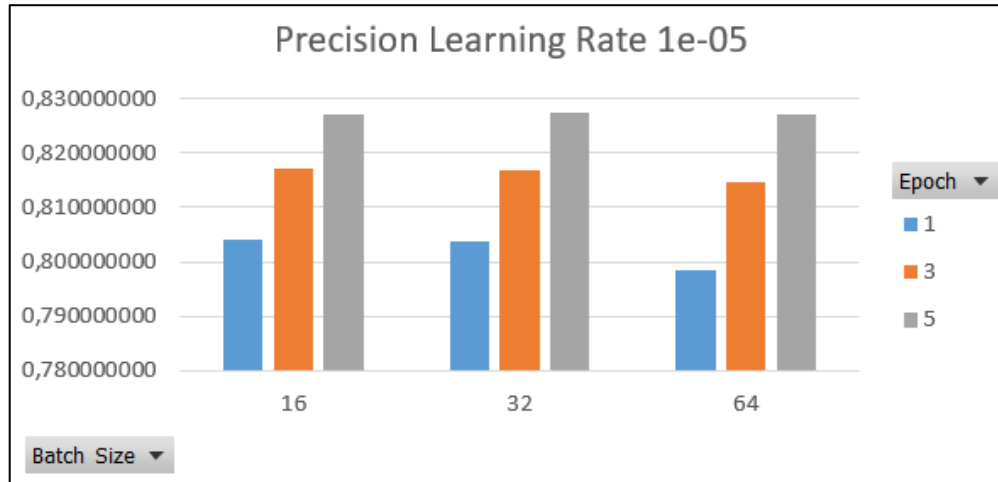
Gambar 8. Akurasi pada *Learning Rate* 3e-05Gambar 9. Akurasi pada *Learning Rate* 5e-05

Analisis akurasi:

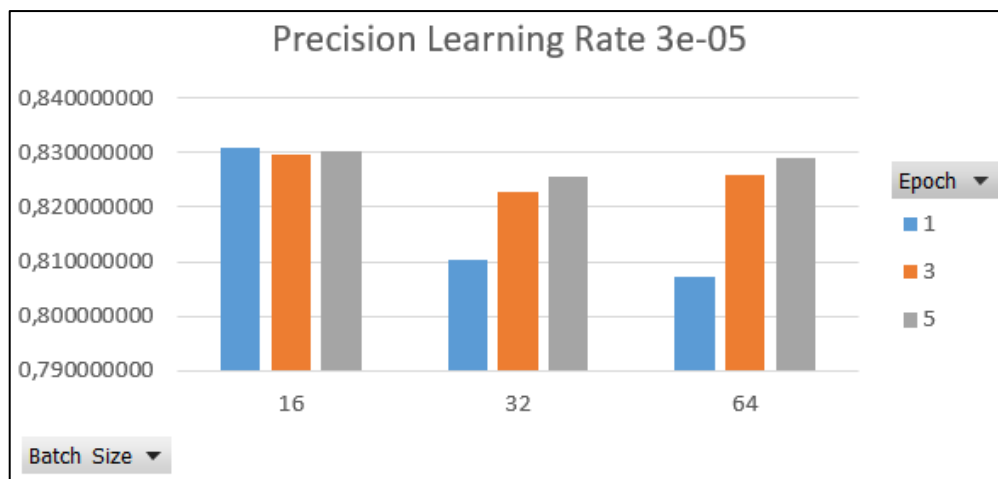
1. Efek *Learning Rate*: *Learning Rate* yang lebih rendah cenderung memberikan akurasi yang lebih stabil. Dalam percobaan ini, *learning Rate* 1e-05 menghasilkan akurasi yang relatif konsisten di semua kombinasi *batch size* dan *epoch*. *Learning Rate* 3e-05 menunjukkan sedikit peningkatan akurasi untuk beberapa kombinasi tetapi tidak secara konsisten mengungguli 1e-05. *Learning Rate* 5e-05 menghasilkan akurasi yang lebih rendah dibandingkan dengan dua lainnya.
2. Efek *Batch Size*: *Batch Size* yang lebih kecil (16) memberikan akurasi terbaik untuk *learning Rate* 1e-05 dan 3e-05, sementara *batch size* yang lebih besar (64) menghasilkan hasil yang kompetitif pada *learning Rate* 5e-05. *Batch size* yang lebih besar tidak selalu berkinerja lebih baik tetapi sering kali memberikan stabilitas yang lebih besar saat jumlah *epoch* meningkat.
3. Efek *Epoch*: Peningkatan jumlah *epoch* secara umum meningkatkan akurasi untuk sebagian besar kombinasi, dengan hasil tertinggi terlihat pada epoch 5. Namun, untuk beberapa kombinasi *batch size* dan *learning Rate*, akurasi stabil bahkan pada *epoch* yang lebih rendah, yang menunjukkan bahwa tidak semua konfigurasi memerlukan banyak *epoch* untuk mencapai akurasi optimal.

4.3 Uji Presisi

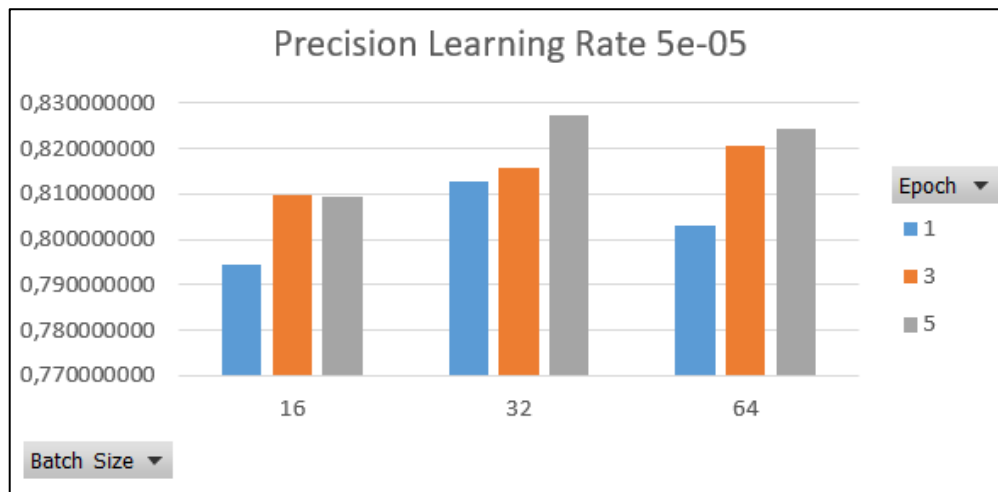
Uji ketiga dalam percobaan ini mengevaluasi presisi. Presisi mengukur proporsi prediksi positif yang benar terhadap semua prediksi positif yang dibuat oleh model. Presisi yang tinggi menunjukkan lebih sedikit prediksi positif yang salah [31]. Hasil uji presisi adalah sebagai berikut:



Gambar 10. Presisi pada *Learning Rate* 1e-05



Gambar 11. Presisi pada *Learning Rate* 3e-05

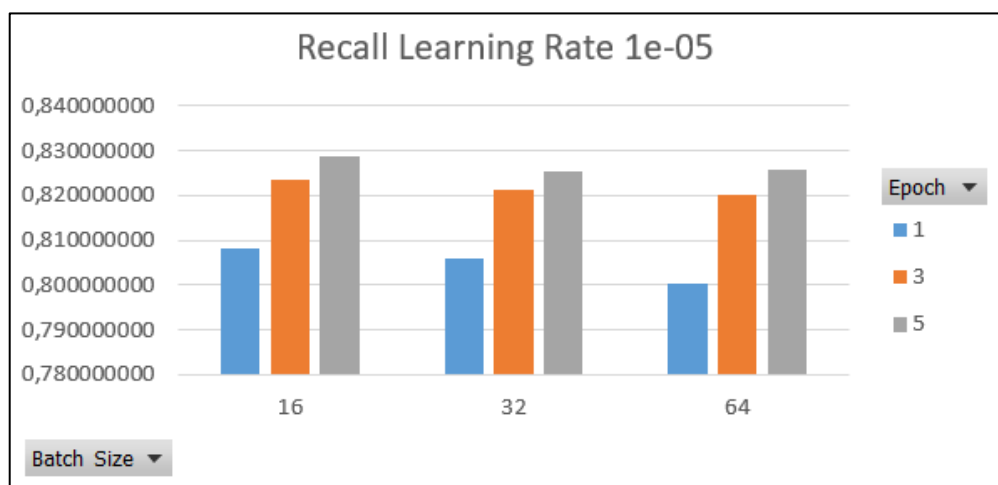
Gambar 12. Presisi pada *Learning Rate* 5e-05

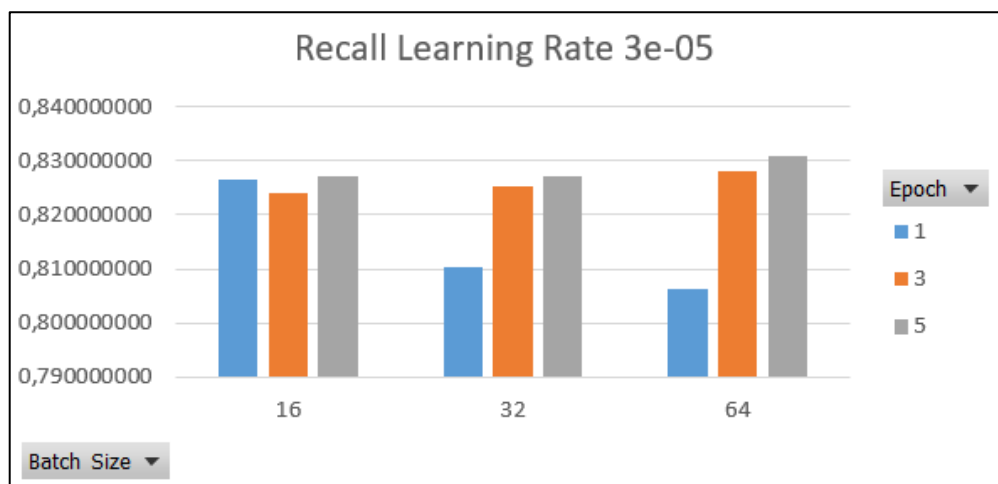
Analisis presisi:

1. Efek *Learning Rate*: Presisi tertinggi dicapai dengan *learning rate* 3e-05 pada *batch size* 16 dan *epoch* 1, yang menunjukkan kemampuan model untuk membuat prediksi positif yang akurat. *Learning rate* 1e-05 memberikan presisi yang stabil, dengan nilai terbaik 0,827 pada *batch size* 16 dan *epoch* 5. Sementara itu, 5e-05 menunjukkan fluktuasi presisi yang lebih besar.
2. Efek *Batch Size*: *Batch size* 16 memberikan presisi terbaik pada *learning rate* 1e-05 dan 3e-05. Namun, *batch size* yang lebih besar (64) menghasilkan hasil yang kompetitif pada *learning rate* 5e-05, meskipun dengan kinerja yang kurang konsisten.
3. Efek *Epoch*: *Epoch* yang lebih tinggi (5) menunjukkan peningkatan presisi untuk sebagian besar kombinasi, meskipun dalam beberapa kasus, presisi tidak meningkat secara signifikan setelah *epoch* 3. Hal ini menunjukkan bahwa pada kombinasi tertentu, model mencapai titik optimalnya lebih cepat.

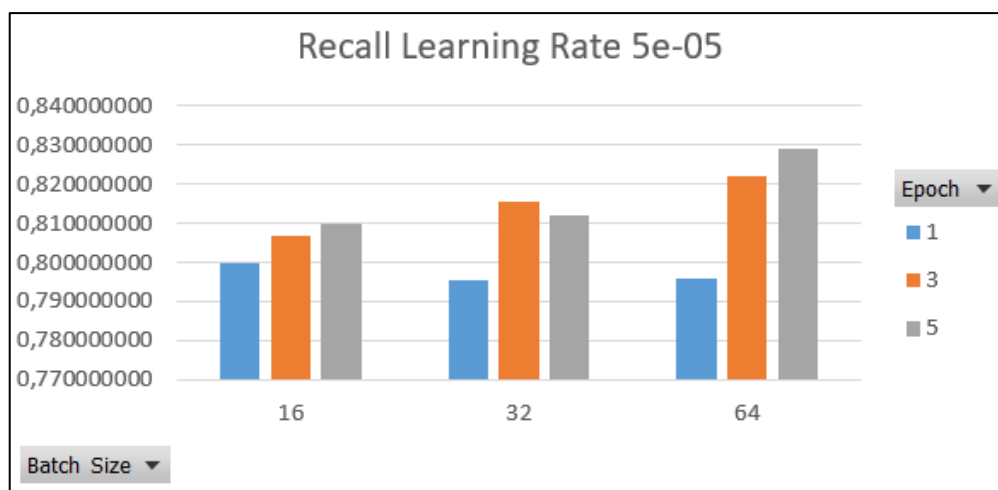
4.4 Uji Recall

Uji keempat dalam percobaan ini mengevaluasi *recall*. *Recall* mencerminkan kemampuan model untuk mengidentifikasi semua contoh yang relevan. Metrik ini dihitung sebagai rasio positif sejati terhadap jumlah total positif aktual. *Recall* yang tinggi menunjukkan lebih sedikit prediksi negatif palsu [31]. Hasil uji *recall* adalah sebagai berikut:

Gambar 13. Recall pada *Learning Rate* 1e-05



Gambar 14. Recall pada Learning Rate 3e-05



Gambar 15. Recall pada Learning Rate 5e-05

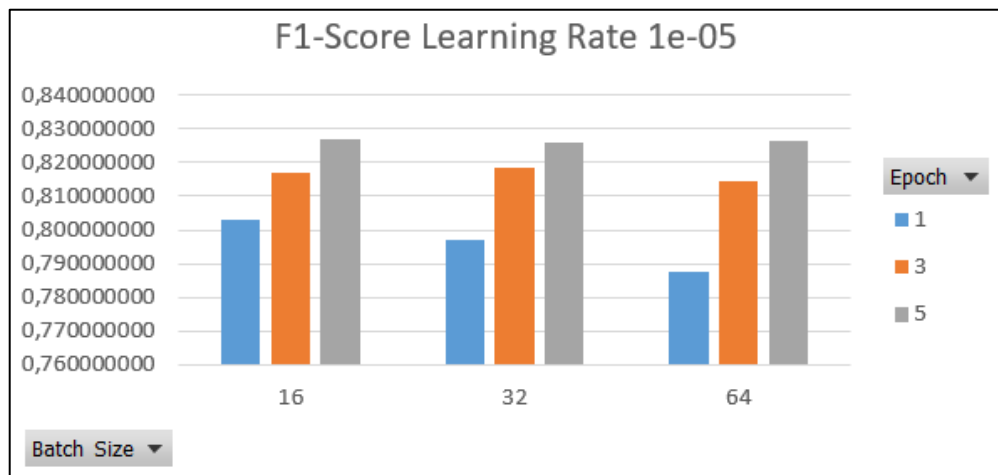
Analisis recall:

1. Efek *Learning Rate*: Recall terbaik dicapai pada *learning rate* 3e-05 dengan nilai 83,08% pada *batch size* 64 dan *epoch* 5, meskipun dengan fluktuasi yang lebih besar pada *epoch* yang lebih rendah. *Learning rate* 1e-05 memberikan *recall* yang stabil, dengan nilai tertinggi 82,89%. *Learning rate* 5e-05 menunjukkan kinerja yang lebih bervariasi dan secara keseluruhan lebih rendah dibandingkan dengan 1e-05 dan 3e-05.
2. Efek *Batch Size*: Recall yang lebih tinggi dicapai dengan *batch size* 32 dan 64 pada *learning rate* rendah hingga sedang. *Batch size* yang lebih kecil (16) menghasilkan *recall* yang lebih rendah, terutama pada *learning rate* yang lebih tinggi (5e-05).
3. Efek *Epoch*: Recall meningkat dengan lebih banyak *epoch*, mencapai nilai tertingginya pada *epoch* 5 untuk sebagian besar kombinasi. Ini menunjukkan bahwa model menjadi lebih baik dalam mengenali data positif setelah lebih banyak iterasi pelatihan.

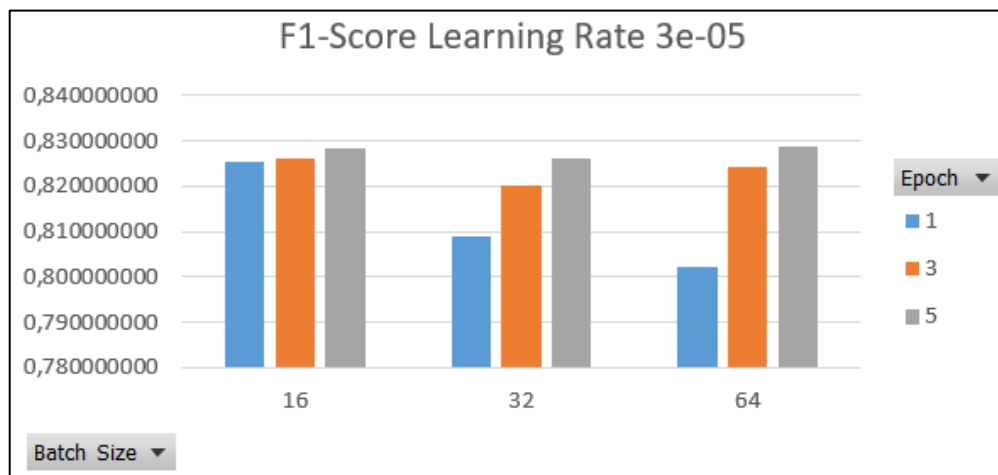
4.5 Uji F1-Score

Uji kelima dalam percobaan ini mengevaluasi *F1-score*. *F1-score* adalah rata-rata harmonik presisi dan perolehan kembali, yang memberikan keseimbangan antara kedua metrik tersebut. *F1-score*

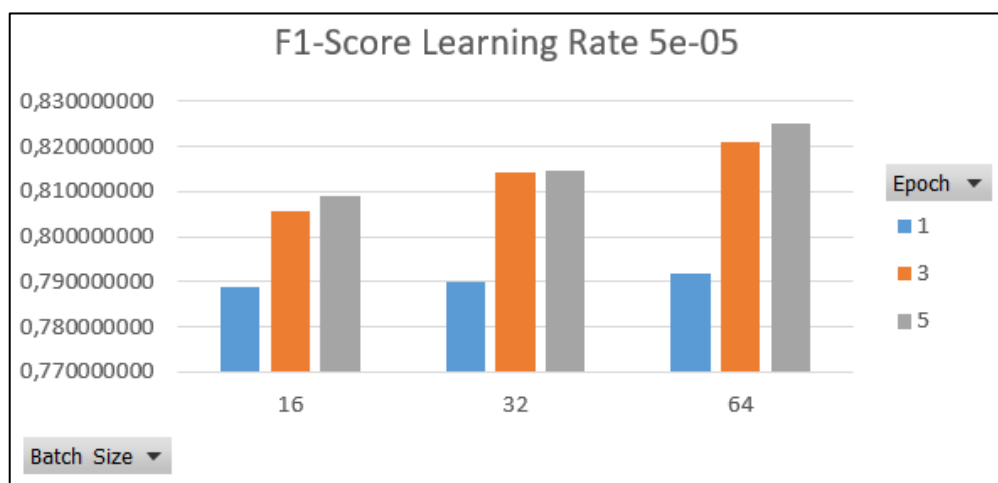
sangat berguna untuk kumpulan data yang tidak seimbang [31]. Hasil uji *F1-score* adalah sebagai berikut:



Gambar 16. *F1-Score* pada *Learning Rate* 1e-05



Gambar 17. *F1-Score* pada *Learning Rate* 3e-05



Gambar 18. *F1-Score* pada *Learning Rate* 5e-05

Analisis *F1-score*:

1. Efek *Learning Rate*: *F1-score* terbaik dicapai dengan *learning rate* $3e-05$ pada *batch size* 64 dan *epoch* 5 (0,829). *Learning rate* $1e-05$ juga memberikan *F1-score* yang sangat baik (0,827 pada *batch size* 16 dan *epoch* 5), yang menunjukkan keseimbangan yang baik antara presisi dan *recall*. Sementara itu, $5e-05$ menunjukkan fluktuasi yang lebih besar, dengan hasil terbaik 0,828 pada *batch size* 64 dan *epoch* 5.
2. Efek *Batch Size*: *Batch size* yang lebih kecil (16) menghasilkan *F1-score* yang baik pada *learning rate* yang lebih rendah ($1e-05$), sementara *batch size* yang lebih besar (64) berkinerja terbaik pada *learning rate* $3e-05$ dan $5e-05$.
3. Efek *Epoch*: Seperti metrik lainnya, *epoch* 5 memberikan *F1-score* terbaik, yang menunjukkan bahwa model memerlukan lebih banyak iterasi untuk mencapai keseimbangan optimal antara presisi dan *recall*.

5. KESIMPULAN

Studi ini menunjukkan bahwa kombinasi parameter terbaik, yaitu *learning rate* sebesar $3e-05$, *batch size* 64, dan 5 *epoch*, menghasilkan kinerja model yang optimal dengan akurasi 83,08%, presisi 82,91%, *recall* 83,08%, dan *F1-score* 82,87%. Jika dibandingkan dengan penelitian sebelumnya yang menggunakan metode Naïve Bayes dengan pendekatan N-Gram dan Information Gain pada data ulasan Shopee E-Commerce, memang terdapat keunggulan pada aspek akurasi yang mencapai 92%. Namun, dalam penelitian ini, model menunjukkan performa yang secara signifikan lebih baik dalam hal presisi, *recall*, dan *F1-score*. Presisi yang tinggi mengindikasikan ketepatan model dalam mengenali sentimen positif, sedangkan *recall* yang tinggi mencerminkan kemampuannya dalam menangkap keseluruhan data positif yang relevan. *F1-score* yang tinggi menunjukkan keseimbangan antara presisi dan *recall*, yang penting dalam konteks klasifikasi tiga kelas. Meskipun terdapat perbedaan akurasi, keunggulan dalam tiga metrik evaluasi utama ini menegaskan bahwa model BERT yang digunakan lebih andal dan representatif dalam menangani prediksi sentimen. Perbedaan hasil antara kedua pendekatan tersebut kemungkinan dipengaruhi oleh perbedaan karakteristik data, teknik praproses, dan algoritma yang diterapkan. Temuan ini memiliki implikasi yang lebih luas, di mana model berbasis *transformer* seperti BERT dapat dimanfaatkan untuk membangun sistem analisis sentimen otomatis yang lebih akurat dan adaptif, yang pada gilirannya dapat mendukung pengambilan keputusan strategis oleh penyedia platform *e-commerce* serta meningkatkan efektivitas sistem rekomendasi berbasis persepsi konsumen. Selain itu, pendekatan serupa berpotensi diterapkan di sektor lain, seperti pelayanan publik atau analisis kebijakan, di mana klasifikasi opini secara tepat sangat penting dalam merumuskan kebijakan berbasis data.

6. SARAN

Penelitian di masa mendatang dapat mengeksplorasi beberapa cara untuk memajukan analisis sentimen dalam *e-commerce*. Salah satu arah yang menjanjikan adalah integrasi data multimoda, seperti menggabungkan ulasan tekstual dengan gambar atau video, untuk menyediakan konteks yang lebih kaya untuk interpretasi sentimen. Menggabungkan teknik seperti transfer learning dengan prapelatihan khusus domain pada kumpulan data *e-commerce* dapat lebih meningkatkan kemampuan model untuk memahami sentimen pelanggan yang bernuansa. Selain itu, mengeksplorasi model berbasis *transformer* alternatif seperti varian RoBERTa, XLNet atau GPT dapat menawarkan wawasan komparatif dan berpotensi meningkatkan kinerja. Menyelidiki dampak teknik praproses yang lebih canggih, seperti menggunakan penyematan kontekstual untuk variasi bahasa gaul dan bahasa daerah, dapat mengatasi tantangan khusus untuk teks informal. Terakhir, analisis sentimen *real-time* dan penerapan model ini dalam lingkungan produksi dapat diperiksa, dengan fokus pada skalabilitas, latensi, dan efisiensi biaya untuk memastikan penerapan praktis dalam skenario *e-commerce* yang dinamis. Kemajuan ini akan berkontribusi pada sistem analisis sentimen yang lebih tangguh, yang memungkinkan bisnis untuk memperoleh wawasan yang lebih mendalam dan membuat keputusan yang lebih tepat.

DAFTAR PUSTAKA

- [1] K. Selvasheela, A. M. Abirami, and A. K. Askarunisa, "Effective Customer Review Analysis Using Combined Capsule Networks with Matrix Factorization Filtering," 2023, *Tech Science Press*. doi: 10.32604/csse.2023.029148.
- [2] A. S. Otto, D. M. Szymanski, and R. Varadarajan, "Customer satisfaction and firm performance: insights from over a quarter century of empirical research," *J Acad Mark Sci*, vol. 48, no. 3, pp. 543–564, 2020, doi: 10.1007/s11747-019-00657-7.
- [3] F. M. Dewi, L. Sulivyo, and Listiawati, "Influence of Consumer Behavior and Marketing Mix on Product Purchasing Decisions," *APTISI Transactions on Management*, vol. 6, no. 2, pp. 151–157, Feb. 2022, doi: 10.33050/atm.v6i2.1737.
- [4] R. Lourdusamy, P. Thangavel, and J. S., "Sentiments Unleashed : Pioneering the Frontier of Sentiment Analysis through Cutting-Edge Applications and Methodologies," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 10, pp. 205–220, Feb. 2024, doi: 10.32628/CSEIT24105105.
- [5] Nikhil Sanjay Suryawanshi, "Sentiment analysis with machine learning and deep learning: A survey of techniques and applications," *International Journal of Science and Research Archive*, vol. 12, no. 2, pp. 005–015, Jul. 2024, doi: 10.30574/ijrsra.2024.12.2.1205.
- [6] M. Rizwan Rashid, A. Nawaz, A. Raza, A. Alahmadi, and T. Alsaedi, "Sentiment Mining in E-Commerce: The Transformer-based Deep Learning Model," *International journal of electrical and computer engineering systems*, vol. 15, pp. 641–650, Feb. 2024, doi: 10.32985/ijeces.15.8.2.
- [7] pir bakhsh, M. I. Mangrio, M. khan, M. shaikh, and R. Memon, "Optimisation of Sentiment Analysis for E-Commerce," *VFAST Transactions on Software Engineering*, vol. 12, pp. 243–262, Feb. 2024, doi: 10.21015/vtse.v12i3.1907.
- [8] Y. Liu, "Sentiment Analysis for E-commerce Product Reviews by Deep Learning Model of Bert-BiGRU-Softmax," *Journal of Corrosion and Materials*, vol. 48, pp. 42–57, Feb. 2024, doi: 10.61336/jcm2023-5.
- [9] J. Saura, D. Palacios-Marqués, and D. Ribeiro-Soriano, "Privacy concerns in social media UGC communities: Understanding user behavior sentiments in complex networks," *Information Systems and e-Business Management*, pp. 1–21, Feb. 2023, doi: 10.1007/s10257-023-00631-5.
- [10] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl Based Syst*, vol. 226, p. 107134, 2021, doi: <https://doi.org/10.1016/j.knosys.2021.107134>.
- [11] P. Yadav, I. Kashyap, and B. S. Bhati, "Contextual Ambiguity Framework for Enhanced Sentiment Analysis," *Tehnicki Glasnik*, vol. 18, no. 3, pp. 385–393, 2024, doi: 10.31803/tg-20231227064230.
- [12] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, no. 7, pp. 5731–5780, 2022, doi: 10.1007/s10462-022-10144-1.
- [13] P. Dmytro and B. Oleh, "Review of methods for semantic text classificationОГЛЯД МЕТОДІВ СЕМАНТИЧНОЇ КЛАСИФІКАЦІЇ ТЕКСТУ," *System technologies*, vol. 5, pp. 120–136, Feb. 2024, doi: 10.34185/1562-9945-5-154-2024-13.
- [14] M. Salıcı and Ü. E. Ölçer, "Impact of Transformer-Based Models in NLP: An In-Depth Study on BERT and GPT," in *2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2024, pp. 1–6. doi: 10.1109/IDAP64064.2024.10710796.
- [15] F. M. Sinaga, R. Purba, S. J. Pipin, W. S. Lestari, and S. Winardi, "Optimization of Sentiment Analysis Classification of ChatGPT on Big Data Twitter in Indonesia using BERT," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 3, p. 1665, Jul. 2024, doi: 10.30865/mib.v8i3.7861.

- [16] A. Sunjaya, N. Wijaya, P. W. Ng, and S. Winardi, *Implementation of Sentiment Analysis of Shopee E-Commerce Reviews using Naïve Bayes, N-Gram, and Information Gain*. 2023. doi: 10.1109/ICIC60109.2023.10381959.
- [17] P. Shah, H. Patel, and P. Swaminarayan, "Multitask Sentiment Analysis and Topic Classification Using BERT," *ICST Transactions on Scalable Information Systems*, vol. 11, Jul. 2024, doi: 10.4108/eetsis.5287.
- [18] K. Clark, U. Khandelwal, O. Levy, and C. D. Manning, "What Does BERT Look At? An Analysis of BERT's Attention," Jun. 2019, [Online]. Available: <http://arxiv.org/abs/1906.04341>
- [19] S. Ruder, M. E. Peters, S. Swayamdipta, and T. Wolf, "Transfer Learning in Natural Language Processing," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*, A. Sarkar and M. Strube, Eds., Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 15–18. doi: 10.18653/v1/N19-5004.
- [20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Oct. 2018, [Online]. Available: <http://arxiv.org/abs/1810.04805>
- [21] A. Vaswani *et al.*, "Attention Is All You Need," Jun. 2017, [Online]. Available: <http://arxiv.org/abs/1706.03762>
- [22] T. Wolfe, "E-commerce," *Springer Handbooks*, pp. 1411–1430, Jan. 2023, doi: 10.1007/978-3-030-96729-1_67.
- [23] A. A. Tanjung and M. E. Syafii, "Analisis Pengaruh Ekonomi Digital Terhadap Pertumbuhan Ekonomi di Indonesia: Model Data Panel," *Ekonomi, Keuangan, Investasi dan Syariah (EKUITAS)*, vol. 4, no. 2, pp. 567–575, Dec. 2022, doi: 10.47065/ekuitas.v4i2.2223.
- [24] H. Liao, "Learning consumer preferences from online textual reviews and ratings based on the aggregation-disaggregation paradigm with attitudinal Choquet integral," *Economic Research-Ekonomska Istraživanja*, vol. 36, no. 2, Aug. 2022, doi: 10.1080/1331677x.2022.2106282.
- [25] "Sentiment Analysis Using Text Mining: A Survey," *International Research Journal of Modernization in Engineering Technology and Science*, Aug. 2023, doi: 10.56726/irjmets43989.
- [26] L. M. Pham and H. L. The, "LNLf-BERT: Transformer for Long Document Classification with Multiple Attention Levels," *IEEE Access*, p. 1, Jan. 2024, doi: 10.1109/access.2024.3492102.
- [27] S. Talebi, E. Tong, A. Li, G. Yamin, G. Zaharchuk, and M. Mofrad, "Exploring the performance and explainability of fine-tuned BERT models for neuroradiology protocol assignment," *BMC Med Inform Decis Mak*, vol. 24, May 2024, doi: 10.1186/s12911-024-02444-z.
- [28] S. Daisy, F. Mary, and G. Mageswary, "An Efficient Data Preprocessing Techniques for Sentiment Analysis Using MOOC Dataset in Machine Learning," *Conference Proceeding Issue Published in International Journal of Trend in Research and Development*.
- [29] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [30] J. Hartmann, M. Heitmann, and C. Schamp, "More than a Feeling: Accuracy and Application of Sentiment Analysis Christian Siebert 2." [Online]. Available: <https://ssrn.com/abstract=3489963>
- [31] H. Sharma, R. Ajmera, and D. Dharamdasani, "Statistical Analysis and Accuracy Assessment of Improved Machine Learning Based Opinion Mining Framework," 2024. [Online]. Available: <https://internationalpubs.com>