



Perbandingan Akurasi BERT dan RNN pada Analisis Sentimen Komentar Hotel

Mahruzar^{1*}, Setiawan Assegaff², Jasmir³, Yosefina Venus⁴

¹⁻⁴Program Studi Magister Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Indonesia

Email: ruzar_27@yahoo.com^{1*}, setiawanassegaff@unama.ac.id², jasmir@unama.ac.id³, ydifera@gmail.com⁴

Alamat: Jl. Jend. Sudirman, The Hok, Kec. Jambi Sel., Kota Jambi, Jambi 36138

Abstract. *The increasing volume of online hotel reviews provides valuable insights into customer perceptions but poses challenges for manual analysis due to its unstructured nature. This study aims to compare the performance of Recurrent Neural Network (RNN) and Bidirectional Encoder Representations from Transformers (BERT) in hotel review sentiment analysis. A total of 20,491 TripAdvisor hotel reviews were classified into three sentiment categories: negative, neutral, and positive. The research methodology includes text preprocessing, stratified data splitting, class imbalance handling using Random Over-Sampling, tokenization, and supervised model training. Model performance was evaluated using a confusion matrix and classification metrics. The results indicate that BERT outperforms RNN, achieving an accuracy of 80.54%, while RNN reached 62.21%. BERT demonstrated superior capability in capturing contextual and semantic information in hotel reviews. These findings suggest that transformer-based models are more effective for sentiment analysis of complex textual data in the hospitality domain and can support data-driven service improvement strategies.*

Keywords: BERT; hotel reviews; RNN; sentiment analysis.

Abstrak. Meningkatnya jumlah ulasan hotel pada platform digital menyimpan informasi penting mengenai persepsi dan kepuasan pelanggan, namun sulit dianalisis secara manual karena sifatnya yang tidak terstruktur. Penelitian ini bertujuan membandingkan kinerja model *Recurrent Neural Network* (RNN) dan *Bidirectional Encoder Representations from Transformers* (BERT) dalam analisis sentimen komentar hotel. Dataset yang digunakan terdiri atas 20.491 ulasan hotel dari TripAdvisor yang diklasifikasikan ke dalam tiga kelas sentimen, yaitu negatif, netral, dan positif. Metode penelitian meliputi praproses teks, pembagian data secara *stratified*, penanganan ketidakseimbangan data menggunakan *Random Over-Sampling*, tokenisasi, dan pelatihan model terawasi. Evaluasi kinerja dilakukan menggunakan confusion matrix dengan metrik klasifikasi standar. Hasil penelitian menunjukkan bahwa model BERT memperoleh akurasi sebesar 80,54%, lebih tinggi dibandingkan model RNN dengan akurasi 62,21%. Hal ini menunjukkan bahwa BERT lebih efektif dalam menangkap konteks semantik pada ulasan hotel. Temuan penelitian ini memberikan dasar empiris dalam pemilihan model analisis sentimen yang tepat untuk mendukung pengambilan keputusan berbasis data di industri perhotelan.

Kata kunci: analisis sentimen; BERT; RNN; ulasan hotel.

1. LATAR BELAKANG

Perkembangan teknologi digital telah membawa perubahan signifikan dalam industri perhotelan, khususnya dalam cara konsumen menyampaikan pengalaman dan mengevaluasi kualitas layanan. Platform reservasi daring dan situs ulasan memungkinkan pelanggan untuk memberikan komentar secara terbuka terkait fasilitas, pelayanan, dan

kenyamanan hotel. Ulasan daring tersebut terbukti mampu merepresentasikan pengalaman pelanggan dan memengaruhi persepsi serta keputusan calon tamu (Phan et al., 2025). Oleh karena itu, data komentar pelanggan memiliki nilai strategis bagi pengelola hotel sebagai dasar evaluasi dan peningkatan kualitas layanan. Namun, tingginya volume komentar serta keberagaman gaya bahasa yang digunakan menyebabkan analisis manual menjadi tidak efisien dan berpotensi menimbulkan bias subjektif, sehingga diperlukan pendekatan komputasional untuk mengolah data teks secara sistematis dan akurat.

Analisis sentimen merupakan pendekatan yang umum digunakan untuk mengidentifikasi polaritas opini pelanggan berdasarkan data tekstual (Amira & Irawan, 2020). Pada tahap awal, penelitian analisis sentimen ulasan hotel banyak mengadopsi metode *machine learning* konvensional. (Nengsih et al., 2021) menerapkan *coarse-grained sentiment analysis* pada ulasan hotel menggunakan algoritma Random Forest dengan representasi fitur TF-IDF untuk mengklasifikasikan sentimen pada level dokumen. Hasil penelitian menunjukkan performa klasifikasi yang cukup baik. Akan tetapi, pendekatan ini masih bergantung pada rekayasa fitur manual dan belum mampu menangkap konteks semantik yang kompleks serta hubungan antar kata dalam komentar pelanggan yang cenderung panjang dan bervariasi.

Seiring berkembangnya pembelajaran mendalam, model berbasis neural network mulai digunakan dalam analisis sentimen ulasan hotel. (D. Apriliani et al., 2021) mengembangkan aplikasi SentiHotel dengan pendekatan *supervised learning* berbasis neural network untuk menganalisis sentimen layanan hotel menggunakan data ulasan berbahasa Indonesia. Penelitian tersebut berfokus pada pengembangan sistem dan evaluasi kinerja satu model tertentu sebagai pendukung pengambilan keputusan manajemen hotel. Meskipun menunjukkan hasil yang baik, penelitian ini tidak melakukan perbandingan dengan arsitektur *deep learning* lain yang memiliki mekanisme pemodelan konteks berbeda, sehingga keunggulan relatif antar model belum dapat dievaluasi secara komprehensif.

Perkembangan berikutnya ditandai dengan hadirnya model bahasa berbasis transformer, khususnya Bidirectional Encoder Representations from Transformers (BERT), yang mampu memahami konteks kata secara dua arah melalui proses *pre-training* pada korpus teks berskala besar. (Bello et al., 2023) mengusulkan kerangka kerja

BERT untuk analisis sentimen dan menunjukkan bahwa representasi kontekstual dua arah mampu meningkatkan akurasi klasifikasi sentimen dibandingkan pendekatan sebelumnya. Selanjutnya, (Kelvin et al., 2025) mengombinasikan BERT dengan metode optimasi tambahan untuk meningkatkan performa klasifikasi sentimen dan melaporkan akurasi yang tinggi pada dataset berskala besar. Namun, kedua penelitian tersebut lebih berfokus pada evaluasi performa BERT secara individual dan tidak secara khusus diterapkan pada domain ulasan hotel maupun dibandingkan dengan model sekuensial seperti Recurrent Neural Network (RNN).

Selain itu, (S. Apriliani et al., 2025) menganalisis ulasan pelanggan hotel melalui pendekatan *opinion mining* dengan metode konvensional untuk mengaitkan sentimen dengan pengalaman pelanggan. Penelitian tersebut memberikan gambaran umum pola sentimen pelanggan, namun belum mengevaluasi perbedaan karakteristik kinerja antarsitektur *deep learning*. Secara keseluruhan, penelitian-penelitian terdahulu cenderung menggunakan satu pendekatan tertentu baik metode konvensional, neural network, maupun BERT atau membandingkan *deep learning* dengan algoritma klasik menggunakan dataset, skema pelatihan, dan metrik evaluasi yang berbeda-beda.

Kondisi tersebut menunjukkan bahwa hingga saat ini masih terbatas penelitian yang melakukan evaluasi komparatif secara langsung antara model berbasis sekuensial seperti RNN dan model berbasis transformer seperti BERT pada analisis sentimen komentar hotel dengan dataset dan metrik evaluasi yang seragam. Akibatnya, keunggulan relatif kedua model dalam menangani karakteristik khas komentar hotel seperti bahasa yang bersifat informal, panjang teks yang bervariasi, serta konteks layanan yang kompleks belum dapat dinilai secara objektif dan menyeluruh. Celah penelitian ini menegaskan urgensi dilakukannya studi komparatif yang sistematis untuk memberikan dasar empiris dalam pemilihan model analisis sentimen yang tepat di sektor perhotelan.

Berdasarkan celah penelitian tersebut, penelitian ini bertujuan untuk membandingkan tingkat akurasi model BERT dan RNN dalam analisis sentimen komentar hotel menggunakan dataset dan metrik evaluasi yang sama. Hasil penelitian diharapkan dapat memberikan kontribusi empiris berupa evaluasi komparatif yang objektif terhadap kedua model, serta menjadi rujukan dalam penerapan metode *deep learning* yang efektif untuk analisis sentimen di bidang perhotelan.

2. KAJIAN TEORITIS

A. Analisis Sentimen

Analisis sentimen merupakan subbidang dalam *natural language processing* yang bertujuan untuk mengidentifikasi serta mengklasifikasikan opini, sikap, atau emosi pengguna terhadap suatu entitas berdasarkan data teks (Wen et al., 2023). Klasifikasi sentimen umumnya dibedakan ke dalam kategori positif, negatif, dan netral (Etaiwi et al., 2021). Dalam konteks industri perhotelan, analisis sentimen dimanfaatkan untuk menangkap persepsi pelanggan terhadap kualitas layanan, fasilitas, serta pengalaman menginap melalui ulasan yang disampaikan pada platform digital. Karakteristik komentar hotel yang bersifat subjektif, menggunakan bahasa informal, serta memiliki variasi panjang teks menjadikan analisis sentimen sebagai pendekatan yang efektif untuk mengekstraksi informasi secara sistematis dan objektif.

B. Text Mining

Text mining merujuk pada proses pengolahan data teks tidak terstruktur guna menghasilkan informasi dan pola yang bermakna. Tahapan text mining meliputi pengumpulan data, prapemrosesan teks, ekstraksi fitur, serta pemodelan dan evaluasi (Laili Yuhana et al., 2022). Dalam konteks analisis sentimen, text mining berperan dalam mengonversi teks mentah menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning* maupun *deep learning*. Proses prapemrosesan seperti *case folding*, *tokenization*, penghapusan *stopword*, dan *stemming* bertujuan untuk meminimalkan kebisingan linguistik serta meningkatkan kualitas data, khususnya pada komentar hotel yang kerap mengandung bahasa tidak baku, singkatan, dan ekspresi emosional (Hanum et al., 2024).

C. Recurrent Neural Network

Recurrent Neural Network (RNN) merupakan arsitektur *deep learning* yang dirancang untuk memproses data sekuensial dengan mempertahankan informasi dari langkah sebelumnya melalui mekanisme *hidden state*. RNN banyak diterapkan dalam tugas pemrosesan bahasa alami karena kemampuannya memodelkan urutan kata dalam teks. Dalam analisis sentimen, RNN mempelajari konteks berdasarkan urutan kata pada tingkat kalimat maupun dokumen (Amira & Irawan, 2020). Meskipun demikian, RNN memiliki keterbatasan dalam menangani dependensi jangka panjang akibat permasalahan

vanishing gradient, sehingga kinerjanya dapat menurun ketika diterapkan pada teks ulasan yang panjang dan kompleks, seperti komentar pelanggan hotel.

D. Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Encoder Representations from Transformers (BERT) merupakan model bahasa berbasis *transformer* yang memanfaatkan mekanisme *self-attention* untuk menangkap relasi antar kata secara dua arah (*bidirectional*) (Wen et al., 2023). Berbeda dengan model sekuensial, BERT mengintegrasikan konteks kata dari arah sebelum dan sesudah secara simultan. Model ini dilatih terlebih dahulu (*pre-trained*) pada korpus teks berskala besar dan selanjutnya dapat disesuaikan (*fine-tuned*) untuk tugas spesifik, termasuk analisis sentimen (Gutierrez-Choque et al., n.d.). Kemampuan BERT dalam menghasilkan representasi teks yang kontekstual menjadikannya unggul dalam memahami nuansa bahasa dan implikasi semantik yang umum ditemukan pada komentar pelanggan hotel.

E. Evaluasi Kinerja Model

Evaluasi kinerja model klasifikasi bertujuan untuk mengukur tingkat ketepatan prediksi model terhadap label actual (Gutierrez-Choque et al., n.d.). Salah satu metode evaluasi yang umum digunakan dalam analisis sentimen adalah *confusion matrix*, yang menyajikan ringkasan jumlah prediksi benar dan salah dengan membandingkan hasil prediksi model terhadap label *ground truth*.

Pada klasifikasi sentimen biner, *confusion matrix* terdiri atas empat komponen utama, yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). *True positive* menunjukkan data sentimen positif yang diprediksi dengan benar, sedangkan *true negative* menunjukkan data sentimen negatif yang diprediksi dengan benar. *False positive* merupakan data sentimen negatif yang keliru diprediksi sebagai positif, sementara *false negative* merupakan data sentimen positif yang keliru diprediksi sebagai negatif.

Berdasarkan *confusion matrix* tersebut, kinerja model dapat diukur menggunakan beberapa metrik evaluasi, yaitu akurasi, presisi, *recall*, dan *F1-score*. Akurasi mengukur tingkat ketepatan model secara keseluruhan dan dirumuskan sebagai (Jayanto et al., 2022):

$$A = \frac{T + T}{T + F + T + F}$$

Presisi mengukur ketepatan prediksi kelas positif dan dirumuskan sebagai :

$$P = \frac{T}{T + F}$$

Recall mengukur kemampuan model dalam mendeteksi seluruh data sentimen positif yang ada dan dirumuskan sebagai :

$$R = \frac{T}{T + F}$$

Sementara itu, *F1-score* merupakan nilai harmonisasi antara presisi dan *recall*, yang dirumuskan sebagai :

$$F1 - S = 2 \times \frac{P \times R}{P + R}$$

Dalam penelitian ini, *confusion matrix* digunakan sebagai dasar perhitungan metrik evaluasi untuk membandingkan kinerja model Recurrent Neural Network (RNN) dan Bidirectional Encoder Representations from Transformers (BERT) dalam mengklasifikasikan sentimen komentar hotel. Penggunaan metrik evaluasi yang sama pada kedua model memungkinkan proses perbandingan dilakukan secara objektif dan setara, sehingga hasil evaluasi mencerminkan keunggulan relatif masing-masing arsitektur dalam menangani karakteristik data ulasan hotel.

3. METODE PENELITIAN

Penelitian ini bertujuan membandingkan kinerja model Recurrent Neural Network (RNN) dan Bidirectional Encoder Representations from Transformers (BERT) dalam analisis sentimen komentar hotel. Metode penelitian disusun melalui tahapan pengumpulan data, praproses teks, pembagian dan penyeimbangan data, tokenisasi, pelabelan data, serta evaluasi kinerja model. Seluruh tahapan dilakukan secara sistematis untuk memastikan hasil analisis yang objektif dan dapat dipertanggungjawabkan.

A. Sumber dan Akuisisi Data

Data yang digunakan dalam penelitian ini berasal dari dataset TripAdvisor Hotel Reviews yang diperoleh melalui platform Kaggle. Dataset tersebut berisi 20.491 ulasan pelanggan hotel dalam bentuk teks beserta rating numerik pada rentang 1–5 yang merepresentasikan tingkat kepuasan pengguna terhadap layanan hotel. Dataset diakses melalui laman : <https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews>

Pada tahap awal, dilakukan pengolahan data awal yang meliputi penghapusan data duplikat berdasarkan kesamaan isi ulasan, penghapusan data yang memiliki nilai kosong (*missing values*), serta seleksi atribut yang relevan, yaitu teks ulasan dan rating pelanggan. Tahapan ini bertujuan untuk memastikan kualitas data sebelum dilakukan pengolahan lebih lanjut.

B. Praproses Teks (Text Preprocessing)

Untuk meningkatkan kualitas data dan mengurangi *noise*, dilakukan beberapa tahapan praproses teks sebagai berikut :

- 1) Case Folding : Seluruh teks ulasan diubah menjadi huruf kecil untuk menghindari perbedaan makna yang disebabkan oleh penggunaan huruf kapital.
- 2) Penghapusan URL, Mention, dan Hashtag : Elemen non-linguistik seperti URL, mention, dan hashtag dihapus karena tidak berkontribusi terhadap proses analisis sentimen.
- 3) Penghapusan Tanda Baca : Tanda baca dihilangkan guna menyederhanakan struktur teks dan mengurangi kompleksitas data.
- 4) Penghapusan Angka : Angka dihapus karena tidak memiliki kontribusi signifikan terhadap penentuan polaritas sentimen dalam konteks ulasan hotel.
- 5) Normalisasi Spasi : Spasi berlebih dihapus untuk menjaga konsistensi format teks.

Tahapan praproses ini bertujuan untuk menghasilkan teks yang lebih bersih dan representatif sebelum dilakukan proses tokenisasi oleh model berbasis *transformer*.

C. Pembagian Data (Data Splitting)

Dataset yang telah melalui tahap praproses dibagi menjadi tiga bagian menggunakan metode *stratified sampling* untuk menjaga proporsi kelas sentimen pada setiap subset data. Pembagian data dilakukan dengan komposisi yaitu 80% data latih, 10% data validasi, dan 10% data uji. Pembagian ini dilakukan untuk memastikan distribusi kelas sentimen tetap seimbang pada setiap subset data yang digunakan dalam proses pelatihan dan evaluasi model.

D. Penanganan Ketidakseimbangan Data (Imbalanced Data Handling)

Distribusi kelas sentimen pada dataset bersifat tidak seimbang. Oleh karena itu, penelitian ini menerapkan teknik Random Over-Sampling (ROS) pada data latih. Teknik

ini bekerja dengan menambahkan sampel secara acak pada kelas minoritas hingga jumlahnya seimbang dengan kelas mayoritas. Penerapan oversampling hanya dilakukan pada data latih untuk menghindari terjadinya *data leakage* yang dapat memengaruhi hasil evaluasi model.

E. Tokenisasi

Proses tokenisasi dilakukan menggunakan RoBERTa Tokenizer dengan konfigurasi sebagai berikut :

- 1) Panjang maksimum token sebesar 256
- 2) Mekanisme *padding* dan *truncation* otomatis
- 3) Representasi token dalam bentuk tensor PyTorch

Proses tokenisasi ini memungkinkan teks ulasan diproses secara efektif dan konsisten oleh model berbasis RoBERTa.

F. Metode Pelabelan Data

- 1) Label Awal

Dataset asli menyediakan label berupa rating numerik pada rentang 1–5 yang merepresentasikan tingkat kepuasan pengguna terhadap layanan hotel.

- 2) Transformasi Label

Untuk keperluan analisis sentimen, rating numerik dikonversi menjadi tiga kelas sentimen, yaitu negatif, netral, dan positif. Skema transformasi label ditetapkan sebagai berikut :

Tabel 1 Rating Label Sentimen

Rating Asli	Label Sentimen	Kode Label
1 – 2	Negatif	0
3	Netral	1
4 – 5	Positif	2

Transformasi ini dilakukan untuk menyederhanakan proses klasifikasi serta meningkatkan stabilitas pelatihan model.

- 3) Justifikasi Skema Pelabelan

Pembagian label menjadi tiga kelas dilakukan dengan pertimbangan bahwa rating 1–2 mencerminkan ketidakpuasan pelanggan, rating 3 bersifat ambigu dan merepresentasikan sentimen netral, serta rating 4–5 menunjukkan tingkat kepuasan

yang tinggi terhadap layanan hotel. Pendekatan ini umum digunakan dalam penelitian analisis sentimen berbasis ulasan pelanggan dan memungkinkan perbandingan kinerja model secara lebih objektif.

4) Representasi Label

Label sentimen direpresentasikan dalam bentuk numerik (0, 1, dan 2) untuk mendukung proses klasifikasi multi-kelas menggunakan Cross Entropy Loss.

G. Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan untuk mengukur kemampuan RNN dan BERT dalam mengklasifikasikan sentimen komentar hotel. Metode evaluasi yang digunakan adalah *confusion matrix*, yang membandingkan hasil prediksi model dengan label aktual. Berdasarkan *confusion matrix*, dihitung metrik evaluasi berupa akurasi, presisi, *recall*, dan *F1-score*. Penggunaan metrik evaluasi yang sama pada kedua model bertujuan untuk memastikan bahwa perbandingan kinerja dilakukan secara objektif dan setara.

4. HASIL DAN PEMBAHASAN

A. Kinerja Model BERT dalam Analisis Sentimen

Evaluasi kinerja model BERT dilakukan untuk mengukur kemampuannya dalam mengklasifikasikan sentimen ulasan hotel ke dalam tiga kelas, yaitu negatif, netral, dan positif. Pengujian dilakukan pada data uji sebanyak 2.050 data menggunakan metrik *precision*, *recall*, *F1-score*, dan akurasi. Hasil evaluasi kinerja model BERT disajikan pada Tabel 2.

Tabel 2 Hasil Evaluasi Kinerja Model BERT

Kelas Sentimen	Precision	Recall	F1-Score	Jumlah Data
Negatif (0)	0.8786	0.7640	0.8173	322
Netral (1)	0.3347	0.7752	0.4675	218
Positif (2)	0.9771	0.8185	0.8908	1510
Akurasi			0.8054	2050
Rata-rata Makro	0.7301	0.7859	0.7252	2050
Rata-rata Tertimbang	0.8933	0.8054	0.8342	2050

Berdasarkan Tabel 2, model BERT memperoleh nilai akurasi sebesar 80,54%, yang menunjukkan kinerja yang baik dalam mengklasifikasikan sentimen ulasan hotel. Kinerja terbaik ditunjukkan oleh kelas sentimen positif dengan nilai *F1-score* sebesar

0,8908, yang mengindikasikan bahwa model mampu mengenali pola bahasa positif secara sangat efektif. Hal ini dipengaruhi oleh dominasi jumlah data pada kelas positif sehingga model memiliki representasi konteks yang lebih kuat.

Pada kelas sentimen negatif, model BERT juga menunjukkan performa yang stabil dengan *F1-score* sebesar 0,8173. Sementara itu, kelas sentimen netral memiliki nilai *precision* yang relatif rendah meskipun nilai *recall* cukup tinggi. Kondisi ini menunjukkan bahwa model cenderung kesulitan membedakan sentimen netral secara tegas, karena karakteristik bahasanya sering tumpang tindih dengan sentimen positif maupun negatif.

B. Confusion Matrix Model BERT

Untuk memberikan gambaran yang lebih rinci mengenai distribusi hasil prediksi model BERT, digunakan *confusion matrix*. Hasil *confusion matrix* model BERT disajikan pada Tabel 3.

Tabel 3 Confusion Matrix Model BERT

Aktual \ Prediksi	Negatif (0)	Netral (1)	Positif (2)
Negatif (0)	246	73	3
Netral (1)	23	169	26
Positif (2)	11	263	1236

Berdasarkan Tabel 3, sebagian besar data sentimen positif berhasil diprediksi dengan benar, yaitu sebanyak 1.236 data. Kesalahan prediksi pada kelas ini relatif kecil dibandingkan kelas lainnya. Pada kelas sentimen negatif, sebagian besar data juga terklasifikasi dengan tepat, meskipun masih terdapat kesalahan prediksi ke kelas netral. Adapun pada kelas sentimen netral, tingkat kesalahan prediksi relatif lebih tinggi, terutama ke kelas positif, yang mengindikasikan adanya kemiripan pola bahasa antara kedua kelas tersebut.

C. Kinerja Model RNN dalam Analisis Sentimen

Sebagai pembanding, dilakukan evaluasi kinerja model Recurrent Neural Network (RNN) menggunakan metrik yang sama. Pengujian dilakukan pada data uji sebanyak 4.099 data. Hasil evaluasi kinerja model RNN disajikan pada Tabel 4.

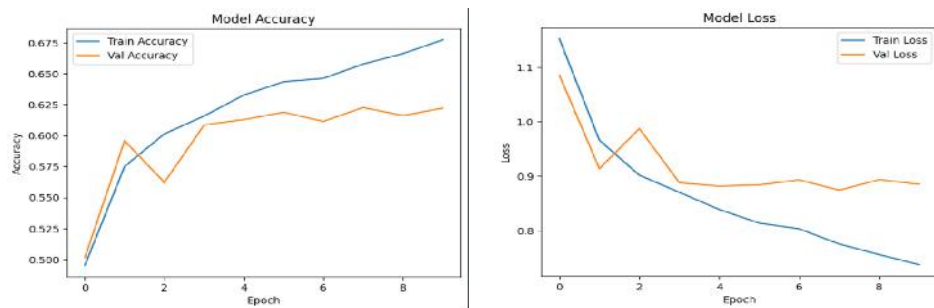
Tabel 4 Hasil Evaluasi Kinerja Model RNN

Kelas Rating	Precision	Recall	F1-Score	Jumlah Data
Rating 1 (0)	0.66	0.65	0.65	284
Rating 2 (1)	0.41	0.52	0.46	359
Rating 3 (2)	0.44	0.18	0.26	437
Akurasi			0.6221	4099
Rata-rata Makro	0.56	0.54	0.53	4099
Rata-rata Tertimbang	0.61	0.62	0.61	4099

Berdasarkan Tabel 3, model RNN memperoleh nilai akurasi sebesar 62,21%, yang menunjukkan kinerja lebih rendah dibandingkan model BERT. Kinerja RNN relatif lebih baik pada kelas rating terendah, namun mengalami penurunan signifikan pada kelas lainnya. Nilai *recall* yang rendah pada salah satu kelas menunjukkan keterbatasan RNN dalam menangkap konteks sentimen secara menyeluruh, khususnya pada ulasan hotel yang memiliki struktur kalimat panjang dan kompleks.

E. Visualisasi Kinerja Model dan Analisis Data

1) Grafik Akurasi dan Loss Model RNN

**Gambar 1** Grafik Akurasi dan Loss Model RNN

Gambar 1 menampilkan perkembangan akurasi dan loss model RNN selama proses pelatihan. Terlihat bahwa akurasi data pelatihan mengalami peningkatan secara bertahap seiring bertambahnya epoch. Namun, akurasi validasi cenderung stagnan dan tidak mengalami peningkatan yang signifikan, yang mengindikasikan keterbatasan model dalam melakukan generalisasi terhadap data baru.

Pada grafik loss, nilai training loss menurun secara konsisten, sedangkan validation loss relatif stabil dan berfluktuasi kecil. Perbedaan tren ini menunjukkan adanya

Penelitian ini membandingkan kinerja model Recurrent Neural Network (RNN) dan Bidirectional Encoder Representations from Transformers (BERT) dalam analisis sentimen komentar hotel. Hasil pengujian menunjukkan bahwa model BERT memiliki kinerja yang lebih baik dengan akurasi sebesar 80,54%, dibandingkan RNN yang hanya mencapai akurasi 62,21%. Selain itu, BERT menghasilkan nilai F1-score tertimbang yang lebih tinggi, terutama pada kelas sentimen positif, yang menunjukkan kemampuannya dalam menangkap konteks semantik ulasan hotel secara lebih efektif. Sebaliknya, RNN menunjukkan keterbatasan dalam memodelkan dependensi jangka panjang, sehingga performanya menurun pada data ulasan yang panjang dan kompleks. Meskipun demikian, penelitian ini masih terbatas pada satu dataset dan konfigurasi model tertentu, sehingga hasilnya belum dapat digeneralisasi secara luas. Penelitian selanjutnya disarankan untuk menggunakan variasi arsitektur sekuensial yang lebih lanjut, melakukan optimasi

hiperparameter, serta menguji model pada dataset dan domain yang berbeda guna memperoleh hasil yang lebih komprehensif.

DAFTAR REFERENSI

- Amira, S. A. & Irawan, M. I. (2020). Opinion Analysis of Traveler Based on Tourism Site Review Using Sentiment Analysis. *IPTEK The Journal for Technology and Science*, 31(2), 223. <https://doi.org/10.12962/j20882033.v31i2.6338>
- Apriliani, D., Abidin, T., Sutanta, E., Hamzah, A. & Somantri, O. (2021). Sentihotel: A sentiment analysis application of hotel services using an optimized neural network. *Bulletin of Electrical Engineering and Informatics*, 10(3), 1536–1543. <https://doi.org/10.11591/eei.v10i3.3040>
- Apriliani, S., Erfina, A. & Warman, C. (2025). Fine-Tuned IndoBERT for Aspect-Based Sentiment Analysis of Indonesian Five-Star Hotel Reviews. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 14(4), 437–445. <https://doi.org/10.32736/sisfokom.v14i4.2491>
- Bello, A., Ng, S. C. & Leung, M. F. (2023). A BERT Framework to Sentiment Analysis of Tweets. *Sensors*, 23(1). <https://doi.org/10.3390/s23010506>
- Etaiwi, W., Suleiman, D. & Awajan, A. (2021). Deep Learning Based Techniques for Sentiment Analysis: A Survey. In *Informatica (Slovenia)* (Vol. 45, Issue 7, pp. 89–95). Slovene Society Informatika. <https://doi.org/10.31449/inf.v45i7.3674>
- Gutierrez-Choque, A.-C., Medina-Mamani, V., Castro-Gutierrez, E., Núñez-Pacheco, R. & Aguaded, I. (n.d.). Transformer based Model for Coherence Evaluation of Scientific Abstracts: Second Fine-tuned BERT. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 13, Issue 5). www.ijacsa.thesai.org
- Hanum, A. R., Zetha, I. A., Fajrina, J. N., Wulandari, R. A., Putri, S. C., Andina, S. P. & Yudistira, N. (2024). ANALISIS KINERJA ALGORITMA KLASIFIKASI TEKS BERT DALAM MENDETEKSI BERITA HOAKS. 11(3), 537–546. <https://doi.org/10.25126/jtiik2024118093>
- Jayanto, R., Kusumaningrum, R. & Wibowo, A. (2022). Aspect-based sentiment analysis for hotel reviews using an improved model of long short-term memory. *International Journal of Advances in Intelligent Informatics*, 8(3), 391–403. <https://doi.org/10.26555/ijain.v8i3.691>
- Kelvin, K., Sinaga, F. M., Wulan Sri Lestari, Sunaryo Winardi, Hawani Rambe, K. & Purba, R. (2025). ENHANCING SENTIMENT ANALYSIS ACCURACY WITH BERT AND SILHOUETTE METHOD OPTIMIZATION. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 11(1). <https://doi.org/10.33480/jitk.v11i1.6392>

- Laili Yuhana, U., Fatichah, C. & Jati Santoso, B. (2022). *EFFECTIVENESS OF DEEP LEARNING APPROACH FOR TEXT CLASSIFICATION IN ADAPTIVE LEARNING*. 11(3).
- Nengsih, W., Mahrus Zein, M. & Hayati, N. (2021). Coarse-Grained Sentiment Analysis Berbasis Natural Language Processing-Ulasan Hotel (Coarse-Grained Sentiment Analysis Based on Natural Language Processing-Hotel Review). In *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* | (Vol. 10, Issue 1).
- Phan, Q. D., Nguyen, L. P. & Nguyen, H. T. T. (2025). ONLINE REVIEWS CAN TELL US HOW THE EXPERIENCE OF THE GUEST WITH THE VIETNAMESE HOTEL. *Geojournal of Tourism and Geosites*, 60, 1290–1301. <https://doi.org/10.30892/gtg.602spl25-1501>
- Wen, Y., Liang, Y. & Zhu, X. (2023). Sentiment analysis of hotel online reviews using the BERT model and ERNIE model—Data from China. *PLoS ONE*, 18(3 March). <https://doi.org/10.1371/journal.pone.0275382>