

Optimisasi Klasifikasi Sentimen Pada Review Hotel Bahasa Inggris Dengan Model Roberta Twitter

Bagas Setiadi¹, Eko Purwanto², Hanifah Permatasari³

^{1,2,3}Sistem Informasi, Universitas Duta Bangsa Surakarta
Jln. Bhayangkara No. 55-57 Tipe, Kec. Serengan, Kota Surakarta, Indonesia

e-mail: 202020472@mhs.udb.ac.id¹, eko_purwanto@udb.ac.id², hanifah_permatasari@udb.ac.id³

Received : March, 2024

Accepted : April, 2024

Published : August, 2024

Abstract

In the digital era, travelers often use the internet to search for and book hotels. Online hotel reviews are crucial sources of information for prospective guests, yet the sentiment in these reviews is difficult to interpret due to complex and contextual language. By employing a more comprehensive model training approach, this research contributes to improving the accuracy of multi-aspect sentiment analysis of English hotel review data using the RoBERTa Twitter model. This research involves stages of problem identification, data selection, preprocessing, modeling, and hyperparameter testing. The RoBERTa Twitter model demonstrates good performance in classifying sentiment and aspects of hotel review data. Evaluation results show satisfactory model performance with a sentiment accuracy of 88%, but accuracy for specific aspects is lower, with facilities at 75%, service at 78%, rooms at 81%, and location and price at 84%. The RoBERTa Twitter model shows promising performance in classifying sentiment and aspects of hotel review data, supporting more informed decisions in hotel selection for online platform users.

Keywords: BERT, roberta, hotel, sentiment

Abstrak

Dalam era digital, wisatawan sering menggunakan internet untuk mencari dan memesan hotel. Ulasan hotel online adalah sumber informasi penting bagi calon tamu, namun sentimen dalam ulasan ini sulit diinterpretasikan karena bahasa yang kompleks dan kontekstual. Dengan pendekatan pelatihan model yang lebih komprehensif, penelitian ini memberikan kontribusi dalam meningkatkan akurasi analisis sentimen multi-aspek data review hotel berbahasa Inggris menggunakan model RoBERTa Twitter. Penelitian ini melibatkan tahapan identifikasi masalah, seleksi data, preprocessing, pemodelan, dan pengujian hyperparameter. Model RoBERTa Twitter menunjukkan kinerja yang bagus dalam mengklasifikasikan sentimen dan aspek review data hotel. Hasil evaluasi menunjukkan kinerja model yang memuaskan dengan akurasi sentimen 88%, namun akurasi untuk aspek tertentu lebih rendah, dengan fasilitas 75%, pelayanan 78%, kamar 81%, serta lokasi dan harga 84%. Model RoBERTa Twitter menunjukkan kinerja yang menjanjikan dalam mengklasifikasikan sentimen dan aspek review data hotel, mendukung keputusan yang lebih tepat dalam pemilihan hotel hotel bagi pengguna platform daring.

Kata Kunci: BERT, roberta, hotel, sentimen

1. PENDAHULUAN

Dalam era digital saat ini, platform daring telah menjadi sarana utama bagi wisatawan dalam mencari dan memesan akomodasi hotel. Ulasan hotel yang diposting oleh pengguna menjadi sumber informasi yang penting bagi calon pelanggan dalam membuat keputusan. Namun, memahami sentimen yang terkandung dalam ulasan tersebut menjadi sebuah tantangan. Penggunaan bahasa yang kompleks, variasi nuansa, serta aspek kontekstual dalam ulasan seringkali menjadi hambatan dalam mengklasifikasikan sentimen dengan tepat. Oleh karena itu, diperlukan pendekatan yang cermat dan terkini untuk meningkatkan akurasi klasifikasi sentimen pada ulasan hotel dalam bahasa Inggris.

Klasifikasi sentimen merupakan salah satu teknik dalam analisis teks yang memungkinkan untuk mengidentifikasi apakah suatu teks (dalam konteks ini, ulasan hotel) menyatakan sentimen positif, negatif, atau netral. Beberapa studi sebelumnya telah menunjukkan bahwa model berbasis *BiDirectional Encoder Representations from Transformers* (BERT) termasuk yang paling canggih dan akurat dalam analisis sentimen [1]. Namun, dalam konteks klasifikasi sentimen pada ulasan hotel, terdapat kompleksitas yang beragam yang perlu diatasi. Studi penelitian yang diterbitkan dalam jurnal "Jurnal Elektronik Ilmu Komputer Udayana" pada tahun 2022 menunjukkan bahwa tantangan ini meliputi pemahaman terhadap variasi bahasa, nuansa yang berbeda-beda, dan aspek-aspek kontekstual yang mempengaruhi interpretasi sentimen.[2]. Analisis sentimen hotel ini diperlukan bagi manajer hotel untuk menganalisis tren alam sentimen pelanggan seiring waktu, yang bisa menunjukkan apa yang disukai dan tidak disukai pelanggan tentang layanan mereka dan memberikan umpan balik untuk meningkatkan kualitas layanan dan fasilitas sehingga reputasi memperkuat aspek positif yang dihargai oleh pelanggan, bagi pelanggan dapat membuat keputusan yang lebih tepat tentang hotel mana yang akan dipesan berdasarkan sentimen umum dari ulasan lain. Penggunaan model bahasa yang tidak dapat menangkap nuansa, kosa kata yang beragam, serta pemrosesan teks yang kurang memadai dapat mengurangi akurasi klasifikasi sentimen. Oleh karena itu, diperlukan pendekatan yang lebih canggih dan terfokus

guna meningkatkan keakuratan analisis sentimen pada ulasan hotel dalam bahasa Inggris.

Model RoBERTa Twitter, sebagai salah satu jenis model dalam ranah *deep learning* yang termasuk ke dalam kategori model bahasa, menjadi salah satu fokus utama dalam penelitian ini. *Robustly optimized BERT Pretraining Approach* (RoBERTa) adalah penyempurnaan dari model *Bidirectional Encoder Representations from Transformers* (BERT) yang dikembangkan oleh Google. RoBERTa mengadopsi arsitektur dasar BERT namun melalui sejumlah modifikasi dalam proses pelatihan yang lebih cermat dan intensif, yang bertujuan untuk meningkatkan kinerja serta kemampuan pemahaman bahasa dalam konteksnya [3].

Berdasarkan paparan di atas, maka penelitian ini berupaya untuk melakukan studi mendalam dan terarah tentang pemanfaatan potensi model RoBERTa Twitter serta teknik penyetulan hyperparameter guna mengoptimalkan klasifikasi sentimen pada ulasan hotel bahasa Inggris. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam meningkatkan akurasi analisis sentimen untuk mendukung pengambilan keputusan yang lebih tepat dalam memilih akomodasi hotel bagi pengguna platform daring.

2. TINJAUAN PUSTAKA

2.1 Landasan Penelitian Terkait

Bagian ini menjelaskan secara umum apa yang menjadi hasil dari penelitian-penelitian sebelumnya untuk dijadikan sebagai sumber literasi pada penelitian ini. Di dalam Tabel 1 dipaparkan beberapa penelitian terkait yang ada sebelumnya.

Tabel 1: Penelitian Terkait

No	Nama Peneliti (Tahun)	Judul	Metode	Hasil Penelitian	Perbedaan
1	[4] Chandradev, 2022	Analisis Sentimen Review	Deep Learning BERT	Menggunkan model SmallB	Menggunkan model SmallBERT yang

		w Hotel Meng gunak an Meto de Deep Learn ing BERT		ERT yang dilat ih pada datase t 515k review hotel di Indone sia dan menca pai akurasi 91,40 %	lebih ringan dan hanya mengkl asifikasi kan sentime n menjadi positif dan negatif daripad a model RoBERT a Twitter yang digunak an dapat mengkl arifikasi kan sentime n menjadi positif ,negatif, dan netral				omial Naive Bayes denga n TfidfVe ctorize r menca pai akurasi rata- rata terting gi sebesa r 74,48 % dan f- measu re rata- rata terting gi sebesa r 84,58 %.	model RoBERT a Twitter dan mengkl asifikasi kan sentime n pada lima aspek yaitu kamar, pelayan an, lokasi, fasilitas, dan harga	
2	[5] Mah enda, 2020	Analisis Sentimen Berda sarkan Aspek pada Ulasan Hotel Berba hasa Indon esia	Naive Bayes dan Multi nomi nal Naive Bayes	Dalam proses klasifikasi sentimen, menganalisis representasi fitur Count Vectorizer dan TfidfVec torize r, algoritma Multin	Menggunakan model BERT mengklasifikasi sentimen pada empat aspek yaitu lokasi, pelayanan, kebersihan, dan harga yang berbeda dari						
3	[6] Tho mas, 2022	Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan TF-IDF								Akurasi 87% untuk SVM dan TF-IDF	Penelitian tersebut menggunakan sentimen general dan berfokus pada bahasa indonesia sedangkan penelitian yang dilakukan menggunakan bahasa

					inggris dan menggunakan aspek-aspek penting
--	--	--	--	--	---

2.2 Analisis Sentimen

Analisis sentimen merupakan subbidang pemrosesan bahasa alami (NLP) yang berfokus pada interpretasi dan pemrosesan teks sesuai dengan sentimen, pandangan, sikap, dan penilaiannya [7]. Analisis sentimen dapat diterapkan pada berbagai konten tekstual, termasuk berita, evaluasi produk, komentar di media sosial, dan banyak lagi. Tujuan dari analisis sentimen adalah untuk mengekstrak apa yang bermanfaat dan berharga dari teks, seperti kepuasan pelanggan, preferensi pasar, tren masyarakat, dan lain-lain.

Ada beberapa teknik komputasional yang dapat digunakan untuk melakukan analisis sentimen, seperti metode berbasis leksikal, metode berbasis pembelajaran mesin, dan metode berbasis *deep learning*. Metode berbasis leksikal mengandalkan ketersediaan kamus atau leksikon yang berisi kata-kata atau frasa beserta nilai sentimennya, seperti positif, negatif, atau netral [8]. Metode ini sederhana dan mudah diimplementasikan, namun memiliki keterbatasan dalam menangani konteks, ironi, dan makna implisit [9]. Analisis sentimen dapat dilakukan dengan menggunakan berbagai teknik komputasi, termasuk pendekatan berbasis leksikal, berbasis machine learning, dan berbasis *deep learning*. Ketersediaan kamus atau leksikon yang memuat kata atau frasa beserta nilai sentimennya, baik positif, negatif, atau netral, diperlukan untuk pendekatan berbasis leksikal [8]. Meskipun pendekatan ini lugas dan mudah digunakan, pendekatan ini tidak begitu efektif ketika mengelola ironi, konteks, atau makna tersirat [9]. Algoritma seperti *Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbor*, dan lainnya yang dapat belajar dari data yang diberi label sentimental digunakan dalam teknik berbasis machine learning. Meskipun dibutuhkan banyak data berkualitas tinggi untuk melatih model, pendekatan ini dapat mengungguli pendekatan berbasis leksikal dalam hal performa [10]. Topologi jaringan

saraf tiruan, seperti jaringan saraf berulang dan jaringan saraf konvolusional, yang dapat mensimulasikan sifat tekstual kompleks digunakan dalam teknik berbasis *deep learning* [8]. Pendekatan ini dapat mengatasi beberapa kekurangan pendekatan berbasis leksikal dan machine learning, termasuk ketidakmampuan menangkap makna laten, keterkaitan semantik, dan konteks jangka panjang.

2.3 Model Transformer dan RoBERTa

Model transformer adalah desain jaringan saraf yang mempelajari representasi rangkaian data, khususnya data teks, melalui mekanisme perhatian. Pendekatan ini pertama kali diperkenalkan oleh [11] dengan judul "Attention is All You Need". Beberapa kelemahan model sebelumnya, termasuk jaringan saraf konvolusional (CNN) dan jaringan saraf berulang (RNN), yang bergantung pada urutan data temporal atau spasial, diatasi dengan pendekatan ini. Terlepas dari lokasi suatu elemen atau jarak dari elemen lainnya, hubungan antara masing-masing elemen dalam suatu rangkaian dapat dimodelkan menggunakan model transformator. Selain itu, model ini mudah diparalelkan, sehingga meningkatkan efisiensi komputasi.

Model transformer yang dirancang untuk terjemahan bahasa alami (NLP) disebut RoBERTa. Pengembangan model Bidirect Encoder Representations from Transformers (BERT) [1] menjadi dasar model ini. Melalui metode pra-pelatihan dan penyesuaian, model BERT—model transformer dapat memperoleh representasi teks yang kontekstual. Banyak tugas NLP, termasuk analisis sentimen, klasifikasi teks, pemahaman bacaan, dan banyak lagi, dapat dilakukan dengan model ini. Model RoBERTa dibuat oleh [12] dengan tujuan untuk meningkatkan kinerja model BERT dengan menggunakan lebih banyak data, lebih banyak komputasi, dan beberapa penyesuaian pada proses pra-pelatihan dan penyesuaian. Model RoBERTa menunjukkan hasil yang lebih baik daripada model BERT dalam beberapa tugas NLP, seperti GLUE (General Language Understanding Evaluation), SQuAD (Stanford Question Answering Dataset), dan RACE (Reading Comprehension from Examinations).

Dalam model transformer, penyesuaian hyperparameter sangat penting karena

berdampak besar pada efektivitas dan kualitas model. Berdasarkan beberapa penelitian, akurasi model dapat ditingkatkan hingga 5% dengan menyempurnakan hyperparameter dibandingkan dengan konfigurasi default benchmark GLUE. Selain itu, penyesuaian hyperparameter juga dapat mengurangi biaya komputasi dan waktu pelatihan dengan memilih ukuran batch dan kecepatan pembelajaran yang optimal [13]. Mengingat beragamnya model transformator yang tersedia, seperti BERT, RoBERTa, GPT-3, dan lainnya, penyesuaian hyperparameter juga dapat membantu dalam memilih model terbaik untuk tugas dan kumpulan data tertentu.

Beberapa contoh hyperparameter yang umum digunakan adalah learning rate, batch size, epoch, jumlah tree, kedalaman tree, atau ukuran embedding. Pengaturan hyperparameter yang tepat dapat meningkatkan akurasi dan efektivitas model klasifikasi sentimen.

2.4 Pentingnya Pengolahan Data yang Berkualitas

Analisis teks adalah proses penggunaan teknologi komputer untuk membaca dan menganalisis konten tertulis manusia untuk memperoleh pengetahuan. Salah satu penerapan analisis teks adalah pengolahan data review hotel. Manajemen hotel dapat memperoleh manfaat dari informasi yang diberikan melalui ulasan perusahaan mereka untuk meningkatkan kepuasan pelanggan dan kualitas layanan. Namun, pemrosesan data ulasan hotel merupakan tantangan karena fitur

3. METODE PENELITIAN

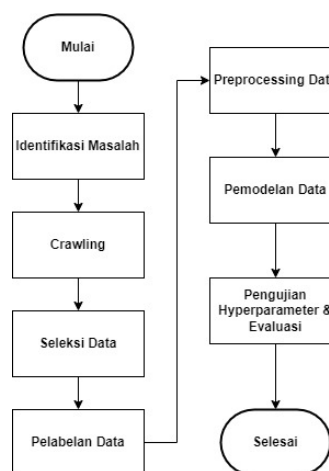
Penelitian ini akan dilakukan dalam beberapa tahap. Tahapan-tahapan tersebut antara lain identifikasi masalah, *crawling*, seleksi data, *preprocessing* data, pemodelan data, dan pengujian hyperparameter dan evaluasi. Tahapan-tahapan tersebut dilihat pada Gambar 1.

data yang kompleks, yang mencakup ketidakterstruktur, keragaman, kebisingan, dan ketidakseimbangan [14].

Adanya data yang bermasalah, atau data yang memiliki kesalahan, inkonsistensi, atau tidak relevan, menjadi penghalang bagi analisis sentimen karena dapat menghambat efektivitas dan akurasi analisis. Ada berbagai jenis gangguan data, termasuk kesalahan ketik, bahasa gaul, emoji, dan spam. Langkah-langkah pra-pemrosesan data seperti formalisasi dan penerjemahan, pelipatan huruf, penghapusan nomor, penghapusan tanda baca, penghapusan spasi, penghapusan kata-kata yang tidak penting, dan stemming diperlukan untuk mengurangi noise data.

Tantangan lain dalam analisis sentimen adalah adanya ketidakseimbangan kelas sentimen, yaitu ketika jumlah data untuk satu kelas sentimen jauh lebih banyak daripada kelas sentimen lainnya. Ketidakseimbangan kelas sentimen dapat menyebabkan bias pada hasil analisis, yang cenderung mengabaikan kelas sentimen minoritas. Untuk mengatasi ketidakseimbangan kelas sentimen, diperlukan teknik penyeimbangan data, seperti oversampling, undersampling, atau kombinasi keduanya.

Dengan demikian, pengolahan data ulasan hotel yang berkualitas memerlukan metode-metode yang dapat menangani noise data dan ketidakseimbangan kelas sentimen, sehingga dapat menghasilkan informasi yang akurat dan relevan bagi manajemen hotel.



Gambar 1. Metode Penelitian

3.1 Identifikasi Masalah

Penelitian ini memfokuskan pada penggunaan model RoBERTa Twitter, yang merupakan model dalam ranah *deep learning* dan merupakan penyempurnaan dari model BERT. Meskipun RoBERTa memiliki potensi, namun penelitian yang lebih mendalam dan terarah diperlukan untuk memanfaatkan sepenuhnya potensinya. Penelitian ini juga mencakup penyesuaian hyperparameter untuk mengoptimalkan klasifikasi sentimen pada ulasan hotel bahasa Inggris.

3.2 Crawling

Penelitian ini menggunakan data ulasan hotel dalam bahasa Inggris, data tersebut dikumpulkan dengan teknik crawling. Crawling merupakan teknik untuk menjelajahi, membaca dan untuk melakukan penarikan atau pengambilan informasi dari web kemudian menyimpan informasi tersebut untuk digunakan [15]. Crawling data menggunakan bahasa pemrograman Python melalui emulator Google Colab [16]. Sumber data dapat berasal dari platform review hotel online seperti Agoda, TripAdvisor, Booking.com, atau sumber data lainnya yang menyediakan ulasan hotel dalam bahasa Inggris.

3.3 Seleksi Data

Seleksi data ini dilakukan pada dataset. Seleksi data adalah proses penting dalam pembangunan model prediktif yang efektif, pendekatan ini melibatkan pemilihan data terbaik yang dapat mengoptimalkan informasi dan memaksimalkan akurasi model [17]. Data yang dikumpulkan yakni ulasan review hotel yang berbahasa Inggris. Dari hasil seleksi data diperoleh 10.000 ulasan review hotel. Berikut tabel beberapa ulasan review hotel bahasa Inggris.

Tabel 2: Review Hotel

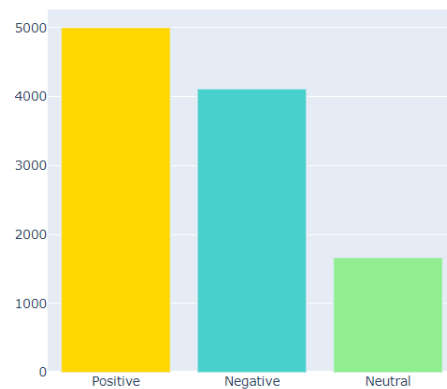
Review
First i got treatment is fruit basket and welcome drink from mr yogi, great service . And second i took room service,i took hawaian pizza and kwetiaw goreng ,good recommend from the staff.and so many thanks for mr yogi and aston solo for nice treatment and food recommend
Aisha kimura from guest relations was very kind to assist me to the horizon lounge and escort me with the facilities in the hotel. I will come to the hotel soon again to see Aisha. She

is very nice. Thank you

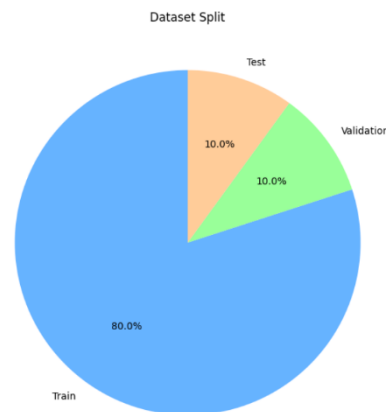
Room clean, simple but modern.

3.4 Preprocessing Data

Data perlu dilakukan preprocessing terlebih dahulu agar mendapat data yang bersih sebelum dilakukan proses training [18]. Tahapan-tahapan yang dilakukan yakni penghilangan noise, memberikan label sentimen terhadap aspek-aspek secara manual yaitu kamar, pelayanan, lokasi, fasilitas, dan harga sentimen berupa positif, negatif dan netral , text cleaning untuk menghilangkan karakter-karakter yang tidak dibutuhkan, split dataset membagi dataset menjadi train data train yang digunakan 80% dari data keseluruhan, validation 10% dan data test 10%, tokenization serta encoding label urutan [19]. Berikut grafik gambar hasil pelabelan data dan dataset split.



Gambar 2. Data Sentimen Review Hotel



Gambar 3. Dataset Split

4. PENGUJIAN DAN HASIL

4.1 Hyperparameter

Hyperparameter adalah proses menentukan kombinasi nilai-nilai optimal untuk parameter-

parameter yang memengaruhi kinerja model dalam pembelajaran mesin [20]. Peneliti melakukan tuning parameter berikut dalam pengaturannya, output_dir ditetapkan sebagai direktori sementara yang telah dibuat sebelumnya. Parameter lainnya memberikan arahan untuk pelatihan model, seperti jumlah epochs (3), ukuran batch pelatihan per perangkat (16), langkah pemanasan (500), regularisasi berat (0.01), dan strategi evaluasi dan logging.

Tabel 3: Perbandingan Hyperparameter

Epochs	3	3	3
Ukuran ukuran batch pelatihan per perangkat	16	32	32
langkah pemanasan	500	500	500
regularisasi berat	0.01	0.01	0.02

4.2 Training

Training merupakan tahapan krusial dalam pengembangan model pembelajaran mesin, di mana model mengadaptasi diri terhadap pola-pola yang terdapat dalam data latihan [21]. Pada tahap awal dalam pengembangan sentimen multi-aspek, enam model dibangun. Lima model difokuskan pada aspek yang berbeda dalam ulasan hotel, sedangkan satu model ditujukan untuk sentimen secara umum. Proses pelatihan dilakukan untuk beberapa sentimen, seperti kamar, pelayanan, lokasi, fasilitas, dan harga. Berikut ini adalah hasil dari proses pelatihan yang telah dilakukan.

Tabel 4: Training Kamar

Epoch	Training Loss	Validation Loss
1	0.494700	0.487136
2	0.454400	0.485303
3	0.346600	0.463576

Tabel 5: Training Pelayanan

Epoch	Training Loss	Validation Loss
1	0.511500	0.517882
2	0.482800	0.493624
3	0.348300	0.533074

Tabel 6: Training Lokasi

Epoch	Training Loss	Validation Loss
-------	---------------	-----------------

1	0.421300	0.435854
2	0.375600	0.407090
3	0.269300	0.414987

Tabel 7: Training Fasilitas

Epoch	Training Loss	Validation Loss
1	0.599400	0.606527
2	0.532500	0.594660
3	0.419100	0.627957

Tabel 8: Training Harga

Epoch	Training Loss	Validation Loss
1	0.450900	0.531306
2	0.446600	0.463913
3	0.325700	0.515378

Tabel 9: Training Sentimen

Epoch	Training Loss	Validation Loss
1	0.383400	0.327503
2	0.288900	0.353486
3	0.178000	0.428709

Setelah training dilakukan kemudian akan menyimpan model terbaik. Tabel 10 menunjukkan Hasil akurasi model pada tiap aspek.

Tabel 10: Akurasi Setiap Aspek Model

kamar	Pelayanan	Lokasi	Fasilitas	harga	sentimen
0.81	0.78	0.84	0.75	0.84	0.88

Dari hasil akurasi yang tertera dalam Tabel 10, terbukti bahwa model RoBERTa Twitter yang telah dibuat berhasil dalam mengklasifikasikan ulasan hotel berdasarkan kriteria yang diinginkan. Dengan tingkat akurasi yang tinggi, model tersebut menjadi standar dalam klasifikasi aspek dan sentimen. Oleh karena itu, model ini mampu memahami dan menganalisis setiap penilaian yang terdapat dalam kalimat review dengan baik.

Validitas model yang dikembangkan dipengaruhi oleh volume data teks yang tersedia untuk setiap aspek. Variasi ini dapat memengaruhi kemampuan model untuk menangkap pola-pola yang lebih kompleks dan menghasilkan prediksi yang lebih tepat. Model pembelajaran yang dilatih dengan dataset yang lebih besar cenderung memiliki tingkat akurasi yang lebih tinggi dan secara konsisten

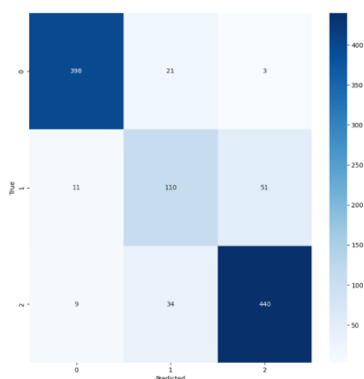
memberikan hasil yang lebih baik serta stabil [22].

4.3 Evaluasi

Dari model terbaik yang disimpan menghasilkan *clarification report* dan *confusion matrix* pada gambar berikut.

	precision	recall	f1-score	support
Negative	0.95	0.94	0.95	422
Neutral	0.67	0.64	0.65	172
Positive	0.89	0.91	0.90	483
accuracy			0.88	1077
macro avg	0.84	0.83	0.83	1077
weighted avg	0.88	0.88	0.88	1077

Gambar 4. Clarification Report



Gambar 5. Confusion Matrix

Evaluasi kinerja model klasifikasi menunjukkan hasil yang memuaskan, dengan akurasi keseluruhan sebesar 88%. Model menunjukkan tingkat ketepatan yang baik dalam mengidentifikasi sentimen negatif, dengan nilai precision sebesar 95% dan recall sebesar 94%. Hal ini menandakan bahwa model cenderung memberikan prediksi yang benar dan mampu menemukan sebagian besar instance sentimen negatif pada dataset.

Namun, performa model sedikit menurun ketika mengidentifikasi sentimen netral, dengan nilai precision sebesar 67% dan recall sebesar 64%. Ini menunjukkan bahwa model memiliki kesulitan dalam membedakan sentimen netral, yang mungkin memiliki pola yang lebih subyektif atau sulit diidentifikasi.

Sementara itu, model menunjukkan hasil yang baik dalam mengklasifikasikan sentimen positif, dengan nilai precision sebesar 89% dan recall sebesar 91%.

Precision mengindikasikan seberapa tepat model dalam mengklasifikasikan instance positif dari kelas tertentu. Recall, di sisi lain, mengukur seberapa baik model dalam menangkap instance positif dari kelas tersebut. Dalam konteks analisis sentimen, precision memberi kita informasi tentang seberapa sering model benar-benar mengenali sentimen tertentu ketika memprediksi sentimen tersebut, sedangkan recall memberi kita gambaran tentang seberapa banyak dari instance sentimen tersebut yang berhasil ditangkap oleh model dari seluruh instance yang sebenarnya memiliki sentimen tersebut.

Sebagai contoh, dalam kasus nilai precision sentimen negatif sebesar 95%, ini berarti 95% dari teks yang diprediksi sebagai negatif oleh model memang memiliki sentimen negatif. Di sisi lain, recall sentimen negatif sebesar 94% menunjukkan bahwa model berhasil mengidentifikasi 94% dari semua teks yang sebenarnya memiliki sentimen negatif.

Kombinasi dari precision dan recall memberikan informasi yang komprehensif tentang kinerja model dalam mengklasifikasikan sentimen. Semakin tinggi nilai keduanya, semakin baik kinerja model. Namun, terkadang ada trade-off antara precision dan recall meningkatkan satu dapat menyebabkan penurunan yang lain. Oleh karena itu, penting untuk mempertimbangkan keduanya secara bersamaan dalam mengevaluasi kinerja model.

5. KESIMPULAN

Dalam penelitian ini, berhasil dikembangkan model sentimen multi-aspek untuk data review hotel menggunakan model RoBERTa Twitter. Terdapat enam model yang dirancang, termasuk lima model untuk masing-masing aspek dan satu model untuk sentimen secara keseluruhan, yang mampu merepresentasikan nilai-nilai dalam ulasan dengan baik. RoBERTa Twitter menunjukkan kinerja yang baik dalam mengklasifikasikan sentimen data review hotel dalam bahasa Inggris.

Kesimpulan dari evaluasi ini adalah bahwa model tersebut mampu mengklasifikasikan sentimen pada dataset uji dengan akurasi keseluruhan sebesar 88%. Namun, terdapat

variasi dalam akurasi untuk masing-masing aspek, dengan nilai 75% untuk fasilitas, 78% untuk pelayanan, 81% untuk kamar, dan 84% untuk lokasi dan harga. Meskipun demikian, hasil ini menunjukkan kemampuan model dalam mengidentifikasi aspek-aspek dalam ulasan dengan baik, menunjukkan kehandalan dalam mengenali dan memprediksi sentimen yang terdapat dalam data.

Pentingnya penelitian ini terletak pada kontribusinya terhadap pemahaman dan aplikasi analisis sentimen pada ulasan hotel dalam bahasa Inggris. Meskipun demikian, masih terdapat ruang untuk peningkatan, terutama dalam meningkatkan akurasi klasifikasi untuk setiap aspek secara individual. Sebagai langkah lanjutan, penelitian mendatang dapat mempertimbangkan pendekatan yang lebih canggih atau teknik pemrosesan yang lebih lanjut untuk meningkatkan kinerja model. Selain itu, perbandingan lebih lanjut dengan model-model lain atau metode analisis sentimen alternatif dapat memberikan wawasan tambahan tentang keunggulan dan batasan dari pendekatan yang digunakan dalam penelitian ini.

Dengan demikian, penelitian ini memberikan landasan yang kuat untuk pengembangan lebih lanjut dalam bidang analisis sentimen, terutama dalam konteks ulasan hotel, serta mengilhami penelitian-penelitian selanjutnya untuk mengembangkan pendekatan yang lebih baik dan lebih akurat dalam menganalisis sentimen dalam bahasa Inggris.

PERNYATAAN PENGHARGAAN

Penulis ingin menyampaikan rasa terima kasih kepada layanan hotel yang telah memberikan akses publik terhadap data penelitian, memungkinkan penulis untuk melakukan penelitian dengan menggunakan data yang tersedia. Penulis juga ingin mengucapkan terima kasih kepada Program Studi S1 Sistem Informasi Universitas Duta Bangsa Surakarta atas kesempatan yang diberikan dalam pelaksanaan penelitian mengenai optimisasi klasifikasi sentimen pada review hotel berbahasa Inggris dengan menggunakan model RoBERTa Twitter. Kehadiran dan dukungan dari Program Studi ini telah menjadi dasar yang kuat bagi kemajuan dan kesuksesan penelitian ini.

Semoga kerja sama yang baik ini dapat berlanjut dan memberikan kontribusi yang berarti bagi perkembangan ilmu pengetahuan di masa yang akan datang.

DAFTAR PUSTAKA

- [1] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Naacl-Hlt 2019*, no. Mlm, pp. 4171–4186, 2019.
- [2] N. P. A. Dewi, N. A. Sanjaya ER, A. E. Karyawati, I. B. M. Mahendra, I. B. G. Dwidasmara, and I. G. A. Wibawa, "Analisis Sentimen Berbasis Aspek Ulasan Pelanggan Hotel di Bali Menggunakan Metode Decision Tree," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 11, no. 3, p. 625, 2022, doi: 10.24843/jlk.2023.v11.i03.p19.
- [3] C. H. Lin and U. Nuha, "Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00782-9.
- [4] V. Chandradev, I. M. Agus, D. Suarjaya, and I. P. A. Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," pp. 107–116, 2022.
- [5] M. D. MAHENDA, "Analisis Sentimen Berdasarkan Aspek Pada Ulasan Hotel Berbahasa Indonesia," pp. 78–79, 2020, [Online]. Available: <https://etd.repository.ugm.ac.id/penelitian/detail/191509>
- [6] V. W. D. Thomas and F. Rumaisa, "Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan TF-IDF," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1767, 2022, doi: 10.30865/mib.v6i3.4218.
- [7] M. Priandi, "Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor," no. September, pp. 311–319, 2021.
- [8] H. Utami, "Analisis Sentimen dari Aplikasi Shopee Indonesia

- Menggunakan Metode Recurrent Neural Network,” *Indones. J. Appl. Stat.*, vol. 5, no. 1, p. 31, 2022, doi: 10.13057/ijas.v5i1.56825.
- [9] E. Budianita, E. P. Cynthia, A. Pranata, and D. Abimanyu, “Pendekatan berbasis Machine Learning dan Leksikal Pada Analisis Sentimen,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, pp. 99–104, 2022, [Online]. Available: <https://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/19137>
- [10] N. E. Oktaviana, Y. A. Sari, and I. Indriati, “Analisis Sentimen terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features dan Support Vector Machine,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 357–362, 2022, doi: 10.25126/jtiik.2022925625.
- [11] A. Vaswani *et al.*, “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
- [12] Y. Liu *et al.*, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” no. 1, 2019, [Online]. Available: <http://arxiv.org/abs/1907.11692>
- [13] L. H. P. de Chavannes, M. Kongsbak, T. M. R. Lagermann, and L. Derczynski, “Hyperparameter Power Impact in Transformer Language Model Training,” *Sustain. 2021 - 2nd Work. Simple Effic. Nat. Lang. Process. Proc. Sustain.*, pp. 96–118, 2021, doi: 10.18653/v1/2021.sustainlp-1.12.
- [14] M. H. Chyntia, D. E. Ratnawati, and I. Arwani, “Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 1702–1708, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [15] V. A. Flores, P. A. Permatasari, and L. Jasa, “Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword,” *Maj. Ilm. Teknol. Elektro*, vol. 19, no. 2, p. 157, 2020, doi: 10.24843/mite.2020.v19i02.p06.
- [16] A. P. Putra, A. N. Alivia, M. Meilani, N. J. Azzahra, and N. A. Rakhmawati, “Analisis Sentimen Warga Twitter Terhadap Game Shopee Cocoki Dengan Metode Naive Bayes Classifier,” *J. Ilm. Inform. Komput.*, vol. 28, no. 2, pp. 137–148, 2023, doi: 10.35760/ik.2023.v28i2.9494.
- [17] S. T. Sipil, F. Teknik, U. Sebelas, and M. Surakarta, “Dan Penurunan Fondasi Tiang Menggunakan Correlation Based Feature Selection (Cfs),” vol. 11, no. 3, pp. 237–244, 2023.
- [18] B. Hakim, “Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning,” *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 2, pp. 16–22, 2021, doi: 10.30813/jbase.v4i2.3000.
- [19] H. Najjichah, A. Sukur, and H. Subagyo, “Pengaruh Text Preprocessing dan Kombinasinya,” *J. Teknol. Inf.*, vol. 15, no. 1, pp. 1–11, 2019.
- [20] T. A. E. Putri, T. Widiari, and R. Santoso, “Penerapan Tuning Hyperparameter Randomsearchcv Pada Adaptive Boosting Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung,” *J. Gaussian*, vol. 11, no. 3, pp. 397–406, 2023, doi: 10.14710/j.gauss.11.3.397-406.
- [21] N. Khamidah, K. Sadik, A. M. Soleh, and G. A. Dito, “Regularisasi model pembelajaran mesin dengan regresi terpenalti pada data yang mengandung multikolinearitas (Studi kasus prediksi Indeks Pembangunan Manusia di 34 provinsi di Indonesia),” vol. 24, no. 1, pp. 12–26, 2024.
- [22] W. Astriningsih and D. Hatta Fudholi, “Identifikasi Multi Aspek Dan Sentimen Analisis Pada Review Hotel Menggunakan Deep Learning,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 3, p. 433, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>