
Analisis Sentimen Multi-aspek untuk Ulasan Hotel dengan Metode Hybrid RNN-Long Short-Term Memory (LSTM)

Bella Hutauruk¹, Yuliant Sibaroni²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹bellahtk@students.telkomuniversity.ac.id, ²yuliant@telkomuniversity.ac.id

Abstrak

Pertumbuhan pesat sektor pariwisata di Indonesia telah menyebabkan peningkatan jumlah ulasan hotel secara online, sehingga menciptakan kebutuhan bagi industri perhotelan untuk menganalisis ulasan tersebut secara efisien guna mendukung pengambilan keputusan. Penelitian ini melakukan analisis sentimen multi-aspek pada ulasan hotel menggunakan pendekatan LSTM dan BERT untuk memahami pola sentimen pada enam aspek utama: kebersihan, lokasi, harga, makanan, fasilitas, dan layanan. Analisis ini mengatasi tantangan seperti penggunaan bahasa informal, sentimen yang ambigu, dan ulasan yang mencakup beberapa aspek secara bersamaan. Data dikumpulkan dari platform Tiket.com dan melalui tahapan pra-pemrosesan meliputi pembersihan data, case folding, tokenisasi, normalisasi, penghapusan kata berhenti, stemming, serta penyeimbangan data menggunakan teknik SMOTE. Dataset kemudian dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian. Model LSTM mencapai akurasi tertinggi dalam mengidentifikasi sentimen untuk aspek kebersihan 94,93%, lokasi 89,98%, dan harga 89,95%. Sebaliknya, model BERT menunjukkan performa yang lebih unggul dalam menangani sentimen yang lebih kompleks dan ambigu. Temuan dari penelitian ini memiliki implikasi signifikan baik bagi wisatawan maupun manajer hotel. Penelitian ini membantu wisatawan memilih hotel yang sesuai dengan preferensi mereka, sekaligus membantu manajer hotel meningkatkan kualitas layanan berdasarkan umpan balik pelanggan, yang pada akhirnya meningkatkan pengalaman pelanggan secara keseluruhan dalam industri perhotelan serta membantu efisiensi operasional.

Kata kunci: Sentimen Analisis, LSTM, BERT, Ulasan Hotel, Multi-Aspek

Abstract

The rapid growth of the tourism sector in Indonesia has led to an increase in the number of online hotel reviews, creating a need for the hospitality industry to analyze these reviews efficiently to support decision-making. This study performs multi-aspect sentiment analysis on hotel reviews using LSTM and BERT approaches to understand sentiment patterns across six key aspects: cleanliness, location, price, food, facilities, and service. The analysis addresses challenges such as informal language, ambiguous sentiments, and reviews covering multiple aspects simultaneously. Data was collected from the Tiket.com platform and underwent preprocessing steps including data cleaning, case folding, tokenization, normalization, stopword removal, stemming, and balancing using the SMOTE technique. The dataset was split into 70% for training and 30% for testing. The LSTM model achieved the highest accuracy in identifying sentiments for cleanliness 94.93%, location 89.98%, and price 89.95%. In contrast, the BERT model showed superior performance in handling more complex and ambiguous sentiments. The findings of this study have significant implications for both travelers and hotel managers. It helps travelers select hotels that match their preferences, while also assisting hotel managers in improving service quality based on customer feedback, ultimately enhancing the overall customer experience in the hospitality industry, helping improve operational efficiency.

Keywords: Sentiment Analysis, LSTM, BERT, Hotel Reviews, Multi-Aspect

1. Pendahuluan

Latar Belakang

Sektor pariwisata menjadi pendorong utama pertumbuhan ekonomi Indonesia, menyumbang 4,3% atau Rp842.303,16 miliar terhadap Produk Domestik Bruto (PDB) pada tahun 2022 [1]. Pertumbuhan ini tercermin dalam lonjakan jumlah wisatawan domestik dan mancanegara setiap tahunnya. Peningkatan signifikan terlihat dalam jumlah perjalanan wisatawan domestik pada semester pertama tahun 2023, mencapai 433,57 juta perjalanan, yang menunjukkan kenaikan sebesar 12,57% dari tahun sebelumnya. Tidak hanya itu, tren positif juga terjadi pada kunjungan wisatawan mancanegara. Pada bulan Juli 2023, jumlah kunjungan mereka mencapai 6,31 juta, melonjak sebesar 196,85% dibandingkan tahun sebelumnya [2].

Dalam menghadapi informasi yang semakin kompleks, penting bagi industri perhotelan untuk dapat menganalisis ulasan tersebut dengan cepat dan efisien. Salah satu metode yang dapat digunakan adalah

sentiment analysis. Analisis ini juga bertujuan untuk mengotomatisasi pemrosesan data teks guna menghasilkan informasi mengenai sentimen yang terdapat dalam suatu opini [3]. *Sentiment analysis* berbasis aspek melibatkan pengambilan aspek dan sentimen yang relevan, serta menentukan orientasi dari setiap aspek tersebut dengan bentuk orientasi yang dapat berupa positif dan negatif [4].

Namun, *sentiment analysis* terhadap ulasan *online* hotel memiliki beberapa tantangan. Pertama, ulasan *online* umumnya ditulis dalam bahasa yang tidak formal dan mengandung banyak *slang*, singkatan, dan kesalahan ejaan. Kedua, ulasan *online* seringkali mengandung sentimen yang ambigu atau tersirat. Ketiga, ulasan *online* dapat memuat beberapa aspek sekaligus dalam satu kalimat, sehingga diperlukan *multi-aspect sentiment analysis* untuk memahami sentimen terhadap setiap aspek secara terpisah. Dengan menggunakan *multi-aspect sentiment analysis*, wisatawan dapat mempertimbangkan aspek kategori pada pemilihan hotel berdasarkan ulasan wisatawan lain. Akan tetapi, sulit untuk mendefinisikan aspek kategori dari ulasan pengunjung. *Sentiment analysis* akan mengekspresikan ulasan yang positif dan negatif sehingga wisatawan dengan mudah dapat menentukan pilihan hotel.

Saat ini, terdapat banyak penelitian mengenai analisis sentimen, khususnya terkait beberapa ulasan hotel yang menggunakan metode *machine learning* seperti RF, Naive Bayes dan SVM. Pada penelitian analisis sentimen dengan *machine learning* metode SVM memiliki akurasi yang lebih baik [5]. Selain penelitian pada *machine learning*, terdapat penelitian analisis sentimen menggunakan *deep learning* seperti LSTM [6] dan BERT [7]. Metode *Long-Short Term Memory* menghasilkan nilai akurasi yang lebih baik dibanding metode naïve bayes [8].

Dalam analisis sentimen, penerapan kerangka kerja hybrid Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM) diusulkan untuk melakukan analisis sentimen multi-aspek. Recurrent Neural Networks (RNN), sebagai kategori model pembelajaran mendalam, dikenal mampu memproses data berurutan secara efektif [9]. Dalam konteks analisis sentimen, RNN unggul dibandingkan model tradisional dengan kemampuannya menangkap konteks dan detail sentimen dalam teks. Di sisi lain, jaringan Long Short-Term Memory (LSTM) menghadirkan inovasi dengan mekanisme pengaturan aliran informasi dalam jaringan, memungkinkan mereka untuk mempertahankan dan memperbarui status internal dalam jangka waktu yang panjang [10].

Penggunaan teknologi RNN dan LSTM menawarkan peluang besar untuk peningkatan sinergis. Penggabungan sensitivitas konteks RNN dengan fungsi memori jangka panjang LSTM memungkinkan analisis sentimen multi-aspek yang lebih tangguh dan akurat. Model hybrid ini telah menunjukkan hasil yang menjanjikan, namun masih ada potensi untuk peningkatan lebih lanjut. Model ini dapat memberikan kinerja yang lebih baik dalam mengenali sentimen yang rumit pada dataset kompleks, sehingga menghasilkan hasil yang unggul untuk aplikasi analisis sentimen yang lebih luas.

Oleh karena itu, penelitian ini mengusulkan penggunaan kerangka kerja hybrid Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM) untuk analisis sentimen multi-aspek. LSTM, sebagai model pembelajaran mendalam yang merupakan pengembangan dari RNN, efektif untuk tugas yang melibatkan input berurutan dengan mengatasi masalah terkait ketergantungan jangka panjang [11]. LSTM dipilih karena kemampuannya untuk mencapai akurasi yang lebih tinggi dalam prediksi sentimen dibandingkan beberapa model pembelajaran mesin [12].

Penelitian ini bertujuan untuk berkontribusi dalam pengembangan metode yang lebih akurat dan efisien untuk analisis sentimen multi-aspek. Selain itu, temuan dari penelitian ini diharapkan dapat memberikan wawasan berharga bagi wisatawan dan manajer hotel. Bagi wisatawan, analisis sentimen dapat membantu dalam memilih hotel yang sesuai dengan kebutuhan dan preferensi mereka. Bagi manajer hotel, analisis sentimen dapat membantu memahami kepuasan pelanggan, mengidentifikasi area yang perlu diperbaiki, dan mengembangkan strategi pemasaran yang lebih efektif.

Topik dan Batasannya

Penelitian ini memiliki batasan untuk menyederhanakan masalah dan memastikan kelayakan dalam waktu yang tersedia. Data yang digunakan dibatasi pada ulasan hotel dalam bahasa Indonesia untuk fokus pada konteks lokal dan memanfaatkan alat pemrosesan bahasa alami yang mendukung bahasa Indonesia, sehingga ulasan dalam bahasa lain tidak dianalisis. Selain itu, penelitian hanya mencakup enam aspek sentimen utama, yaitu kamar, lokasi, kebersihan, pelayanan, harga, dan fasilitas, yang dianggap paling relevan bagi pengguna hotel. Pembatasan ini dilakukan untuk mempermudah proses analisis dan memungkinkan penelitian lebih fokus mendalami performa model *multi-aspect sentiment analysis*.

Tujuan

Berikut tujuan dari penelitian ini adalah:

1. Mengetahui aspek-aspek yang paling dominan dalam setiap sentimen pada *multi aspect sentiment analysis*.
2. Mengetahui hasil performa dari setiap aspek dengan metode *deep learning* menggunakan model RNN-Long Short Term Memory (LSTM) untuk ulasan hotel.

Organisasi Tulisan

Penelitian ini disusun dalam beberapa bagian. Setelah Pendahuluan, bagian Studi Terkait membahas berbagai penelitian sebelumnya yang relevan dengan analisis sentimen multi-aspek pada ulasan hotel. Selanjutnya, bagian Sistem yang Dibangun menjelaskan langkah-langkah yang diambil dalam

penelitian ini, mulai dari pengumpulan dataset ulasan hotel, proses preprocessing data hingga implementasi model Long Short-Term Memory (LSTM). Bagian Evaluasi memaparkan hasil eksperimen dengan analisis performa model termasuk evaluasi performa pada berbagai aspek seperti harga, lokasi, layanan, makanan, fasilitas, dan kamar. Terakhir, bagian Kesimpulan merangkum hasil dari penelitian.

2. Studi Terkait

Pesatnya pertumbuhan sektor pariwisata telah mendorong peningkatan jumlah wisatawan yang memesan hotel dan berbagi pengalaman mereka dalam bentuk ulasan tentang penginapan di hotel tertentu. Ulasan-ulasan ini menjadi informasi berharga bagi calon tamu yang mencari akomodasi yang sesuai. Ulasan positif memainkan peran penting dalam mempromosikan produk atau layanan secara tidak langsung, terutama jika atribut positif yang dijelaskan dalam ulasan tersebut disampaikan dengan jelas [13].

Sentimen yang terkandung dalam ulasan-ulasan ini dapat dianalisis menggunakan analisis sentimen, khususnya analisis sentimen multi-aspek, yang memungkinkan pemahaman lebih mendalam terhadap berbagai aspek yang dievaluasi oleh pengguna. Dengan meningkatnya jumlah ulasan online, kebutuhan untuk menganalisis sentimen yang terkandung di dalamnya menjadi semakin penting. Analisis sentimen multi-aspek memberikan cara untuk mendapatkan wawasan terkait berbagai aspek yang dinilai oleh pengguna.

Pada tahap awal pengembangan ini, berbagai penelitian telah dilakukan untuk mengembangkan metode dan pendekatan dalam mengklasifikasikan dan memahami sentimen pada ulasan hotel. Beberapa penelitian tentang analisis sentimen multi-aspek secara khusus berfokus pada ulasan hotel.

Salah satu penelitian dilakukan oleh Utama et al. [14], yang menganalisis ulasan hotel berbasis aspek. Penelitian ini menggunakan data ulasan hotel dari TripAdvisor yang tersedia di platform Kaggle. Penelitian tersebut mengkategorikan ulasan ke dalam lima aspek: lokasi, kamar, makanan, harga, dan layanan. Dengan menggunakan pendekatan klasifikasi hybrid yang menggabungkan Random Forest (RF), Support Vector Machine (SVM), dan Naïve Bayes, hasilnya menunjukkan bahwa kombinasi metode ini mencapai akurasi tinggi dalam klasifikasi sentimen pada aspek-aspek tersebut.

Analisis sentimen berbasis aspek pada ulasan hotel juga telah dieksplorasi menggunakan berbagai model canggih. Sebuah penelitian mengusulkan metode berbasis LSTM yang dioptimalkan untuk rangkuman tingkat aspek, termasuk makanan, kamar, layanan, dan lokasi, dengan mengotomatisasi jumlah neuron tersembunyi. Model terbaiknya mencapai skor F1 sebesar 75,28%, dengan lapisan tersembunyi pertama berisi 1.200 neuron dan fungsi aktivasi tanh, serta lapisan tersembunyi kedua berisi 600 neuron dengan fungsi aktivasi ReLU. Model ini melampaui baseline LSTM dengan peningkatan 10,16% [15].

Pendekatan lain menggunakan RNN untuk klasifikasi sentimen, mencapai skor F1 sebesar 91,0%, menunjukkan keandalannya dalam tugas analisis sentimen. Teknik deep learning lainnya, seperti AraBERT, LSTM, CNN, dan GRU, juga telah diuji pada ulasan hotel berbahasa Arab, dengan AraBERT mencapai akurasi tertinggi (96,44%) karena pretraining pada data Wikipedia Arab. Model lain, termasuk LSTM dan BiLSTM, juga menunjukkan akurasi yang kompetitif sekitar 95%, sementara metode berbasis CNN memiliki kinerja yang sedikit lebih rendah (93,78%)[16].

Eksplorasi lebih lanjut menggunakan model LSTM dan GRU untuk menganalisis ulasan hotel dari Lombok, Indonesia, dengan akurasi keseluruhan sebesar 91%. Model tersebut berkinerja baik dalam mengidentifikasi sentimen positif (presisi 0,94, recall 0,98, F1-Score 0,96), tetapi kesulitan dalam menangani sentimen negatif dan netral karena ketidakseimbangan data, di mana ulasan positif mendominasi dataset [17]. Untuk mengatasi masalah ini, model hybrid yang mengintegrasikan BiLSTM dan LSTM dengan embedding GloVe pra-latih dan teknik SMOTE-ENN untuk penyeimbangan data berhasil mencapai akurasi sebesar 89,47% dan skor AUC sebesar 0,9607. Model ini menunjukkan kinerja yang kuat dalam menangkap hubungan semantik pada ulasan pelanggan dan mengatasi tantangan dalam analisis sentimen di sektor perhotelan[11].

Penelitian lain tentang analisis sentimen multi-aspek dilakukan oleh Astriningsih dan Fudhol [4], yang menggunakan data ulasan hotel dari berbagai platform online seperti Tiket.com dan Traveloka. Penelitian ini mengkategorikan ulasan ke dalam enam aspek: harga, lokasi, layanan, makanan, fasilitas, dan kamar. Dalam penelitian ini, digunakan model Long Short-Term Memory (LSTM) berbasis deep learning. Hasilnya menunjukkan bahwa metode tersebut mencapai akurasi tinggi dalam mengklasifikasikan sentimen pada aspek-aspek yang telah diidentifikasi.

3. Sistem yang Dibangun

3.1 Rancangan Sistem

Perancangan sistem bertujuan untuk menjelaskan secara rinci tahapan-tahapan yang dilakukan dalam penelitian analisis sentimen multi-aspek pada ulasan hotel menggunakan metode Recurrent Neural Network (RNN) dengan Long Short-Term Memory (LSTM) dan implementasi BERT (*Bidirectional Encoder Representations from Transformers*). Penjelasan perancangan sistem ini dapat dilihat pada Gambar 1.