

Pengembangan Model Long Short-Term Memory Pada Analisis Sentimen Ulasan Hotel

Anita Carolina Wibowo¹, Vihi Atina², Hanifah Permatasari³

^{1,2,3} Sistem Informasi, Universitas Duta Bangsa Surakarta

Email: ¹210101006@mhs.udb.ac.id, ²vihi_atina@udb.ac.id, ³hanifah_permatasari@udb.ac.id

Email Penulis Korespondensi: 210101006@mhs.udb.ac.id

Abstrak

Sektor pariwisata di Indonesia memiliki potensi pengembangan yang signifikan karena kemampuannya untuk menciptakan peluang kerja dan meningkatkan pendapatan. Dalam industri perhotelan, persaingan yang ketat berarti bisnis harus terus mengevaluasi kinerja mereka, termasuk melalui ulasan tamu hotel. Analisis sentimen termasuk salah satu teknik *Natural Language Processing* (NLP) yang menganalisis opini, sikap, dan emosi yang diekspresikan dalam teks. Teknik ini membantu mengidentifikasi kepuasan pelanggan dan mendukung pengambilan keputusan bisnis. Beberapa penelitian telah menggunakan berbagai metode untuk analisis sentimen ulasan hotel, termasuk *Support Vector Machine* (SVM), BERT dan *Long Short-Term Memory* (LSTM), dengan tingkat akurasi yang berbeda-beda. Meskipun BERT mencapai akurasi tertinggi, metode ini juga membutuhkan sumber daya komputasi yang besar. LSTM, di sisi lain, dapat menangkap pola temporal dan ketergantungan jangka panjang dengan biaya komputasi yang lebih rendah. Penelitian yang dilakukan di PT Global Data Inspirasi ini berfokus pada analisis sentimen ulasan hotel dari Y Group, yang mengoperasikan hotel-hotel di berbagai wilayah di Indonesia. LSTM dipilih karena kemampuannya untuk memproses data berurutan secara efisien. Model yang dikembangkan mencapai akurasi 86,79%, presisi 71,63%, recall 76,25%, dan F1-Score 72,24%. Temuan ini menunjukkan bahwa performa model lebih unggul dibandingkan dengan penelitian sebelumnya yang memanfaatkan LSTM. Hasil ini menunjukkan potensi model ini dalam mendukung strategi bisnis melalui analisis sentimen yang efektif.

Kata Kunci: Analisis, Hotel, LSTM, Sentimen, Ulasan

Abstract

The tourism sector in Indonesia has significant development potential due to its ability to create job opportunities and increase income. In the hospitality industry, intense competition means businesses must constantly evaluate their performance, including through hotel guest reviews. Sentiment analysis is a *Natural Language Processing* (NLP) technique that analyses opinions, attitudes and emotions expressed in text. This technique helps to identify customer satisfaction and supports business decision-making. Several studies have employed various methods for sentiment analysis of hotel reviews, including *Support Vector Machine* (SVM), BERT and *Long Short-Term Memory* (LSTM), achieving varying degrees of accuracy. While BERT achieves the highest accuracy, it also requires substantial computational resources. LSTM, on the other hand, can capture temporal patterns and long-term dependencies at a lower computational cost. This study, conducted at PT Global Data Inspirasi, focuses on sentiment analysis of hotel reviews from the Y Group, which operates hotels across various regions in Indonesia. LSTM was chosen for its ability to efficiently process sequential data. The developed model achieved an accuracy of 86.79%, a precision of 71.63%, a recall of 76.25% and an F1 score of 72.24%. These results indicate that the model's performance is better than that of previous studies using LSTM variants. These results demonstrate the potential of the model in supporting business strategies through effective sentiment analysis.

Keywords: Analysis, Hotel, LSTM, Sentiment, Review

1. PENDAHULUAN

Tahun 2011, Badan Perencanaan Pembangunan Nasional Republik Indonesia (Bappenas) menuturkan bahwa sektor pariwisata mempunyai potensi untuk dikembangkan karena menciptakan lapangan pekerjaan dan meningkatkan pendapatan[1]. Persaingan ketat dalam bisnis perhotelan membuat pelaku bisnis mau tidak mau melakukan berbagai strategi untuk dapat bertahan serta mampu bersaing dengan pesaing. Evaluasi berdasarkan ulasan pengunjung hotel menjadi salah satu upaya memahami keunggulan dan kekurangan hotel dari sudut pandang pengunjung hotel[2]. Pengambilan keputusan pemilihan hotel oleh calon pengunjung juga dipengaruhi ulasan pengunjung sebelumnya[3]. Analisis sentimen adalah salah satu cara yang dapat diterapkan untuk mengolah ulasan tamu hotel[1].

PT Global Data Inspirasi (Datains) merupakan anak perusahaan PT Gamatechno Indonesia yang menyediakan solusi *Big Data* dan *Artificial Intelligence* yang berpengalaman dalam penerapan teknologi di segmen pendidikan, pemerintahan, hingga korporasi. Datains memiliki proyek analisis sentimen pada ulasan hotel Y Group. Banyaknya lokasi hotel di berbagai daerah di Indonesia menjadi tantangan dalam memantau seluruh cabang secara efektif dan efisien.

Analisis sentimen termasuk salah satu teknik *Natural Language Processing* (NLP). Analisis sentimen melakukan analisis pendapat, sikap, dan emosi suatu objek berupa teks. Hasil analisis sentiment dijadikan bahan evaluasi yang digunakan sebagai pendukung pengambilan keputusan[4]. Metode *Long Short-Term Memory* (LSTM) adalah salah satu metode *deep learning* yang bisa diaplikasikan dalam *Natural Language Processing* (NLP) salah satunya analisis sentimen[5].

Penelitian mengenai analisis sentimen terhadap ulasan hotel pernah dilakukan oleh Vincentius dan Fitrah [6] pada tahun 2022 menggunakan menggunakan *Support Vector Machine* (SVM) dan menghasilkan model dengan akurasi 87%. Penelitian berikutnya dengan objek yang sama dilakukan Vidya dkk [7] menggunakan BERT dan menghasilkan model dengan akurasi 91.40%. Penelitian berikutnya menggunakan metode berbeda yaitu LSTM pernah dilakukan oleh Windi dan Dhomas [8] pada tahun 2023 dan diperoleh hasil akurasi model sebesar 76% dan masih ditemukan ketidaksesuaian prediksi label dalam pengujian sentimen. Penelitian lain yang menggunakan LSTM dengan objek berbeda yaitu ulasan pelanggan Grab pernah dilakukan Gunawan dkk [9] pada tahun 2025 menghasilkan model dengan akurasi 84% untuk variasi Stacked LSTM dan 87% untuk variasi Bi-Directional LSTM. Berdasarkan penelitian yang telah dilakukan, BERT memiliki akurasi model paling tinggi dari ketiga metode. BERT memiliki keunggulan dalam akurasi dan pemahaman konteks namun LSTM juga menunjukkan kemampuan dalam menangkap pola temporal dan dependensi jangka panjang dalam teks [10]. Walaupun mempunyai pemahaman konteks dua arah yang baik, BERT memiliki kompleksitas yang tinggi dan memerlukan sumber daya komputasi yang besar [11]. LSTM memperlihatkan kemampuan dalam menangkap pola temporal dan dependensi jangka panjang dalam teks [10].

Pada penelitian ini akan dilakukan pengembangan model analisis sentimen yang dapat diterapkan dalam identifikasi kepuasan tamu, pendapat tamu tentang pelayanan, hingga pengambilan keputusan bisnis yang lebih optimal. LSTM dipilih pada penelitian ini karena memiliki kemampuan yang baik dalam memproses dan memahami teks berurutan tanpa membutuhkan sumber daya komputasi yang besar.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menerapkan metode *Cross Industry Standard for Data Mining* (CRISP-DM) untuk melakukan pengembangan model analisis sentimen. Berikut adalah tahapannya.

a. *Business Understanding*

Tahap awal meliputi pemahaman masalah dan penentuan tujuan penelitian [12]. Studi literatur dilakukan untuk memahami analisis sentimen dalam industri perhotelan yang bergantung pada pengalaman pelanggan. Ulasan dari platform daring dianalisis guna memperoleh wawasan mendalam dan meningkatkan kualitas layanan berdasarkan masukan pelanggan.

b. *Data Understanding*

Tahap berikutnya yaitu mengumpulkan data yang akan diperlukan dalam penelitian untuk lebih memahami tentang karakteristik data, mengidentifikasi potensi masalah kualitas, memperoleh wawasan awal, dan menemukan segmen data yang menarik sebagai dasar pembentukan hipotesis informasi tersembunyi [13]. Pengumpulan data ulasan pengunjung hotel dari sumber-sumber terkait, seperti Traveloka, Google Review, TripAdvisor, dan sebagainya. Data yang dikumpulkan mencakup isi ulasan, bahasa, *rating*, serta informasi pendukung lain.

c. *Data Preparation*

Pada tahap ini data dilakukan memproses data mentah yang sudah dikumpulkan menjadi data yang siap untuk diolah. Prosesnya mencakup pemilihan atribut yang relevan, transformasi data, pembersihan *noise*, serta penyaringan informasi untuk memastikan bahwa hanya data yang berkualitas dan relevan yang akan digunakan [12].

d. *Modelling*

Tahap berikutnya adalah pembangunan model analisis sentimen dengan algoritma *Long Short-Term Memory* (LSTM).

e. *Evaluation*

Tahap berikutnya adalah mengevaluasi kinerja model dengan mengukur akurasi dan efektivitas model yang telah dibangun. Evaluasi model akan menggunakan *confusion matrix* untuk melakukan perhitungan perhitungan performa model [14].

f. *Deployment*

Tahap terakhir adalah melakukan dokumentasi keseluruhan tahap penelitian. Hasil dari pengembangan model dan evaluasi performa model disusun dalam bentuk laporan atau jurnal publikasi [12].

3. HASIL DAN PEMBAHASAN

3.1. *Business Understanding*

Permasalahan yang melatarbelakangi penelitian ini adalah sulitnya memantau dan melakukan evaluasi kepuasan tamu di seluruh cabang hotel Y Group secara efektif dan efisien karena masih dilakukan dengan cara melakukan pemantauan rating pada aplikasi ulasan daring saja. Perlunya metode analisis yang dapat mengolah data teks ulasan tamu untuk mendapatkan wawasan yang berguna untuk pengambilan keputusan bisnis.

‘pilihan’, ‘masukan’, ‘tolong’, ‘sediakan’, ‘sendal’]	‘pilihan’, ‘masukan’, ‘tolong’, ‘sediakan’, ‘sendal’, ‘pad’, ‘pad’]
--	---

Proses selanjutnya adalah menghapus stopwords. Pada proses ini dilakukan penghapusan kata-kata yang tidak dapat diolah dengan algoritma komputer seperti kata ganti atau kata imbuhan. Hasil menghapus stopwords ditunjukkan pada tabel 5.

Tabel 5. Mennghapus *Stopword*

Ulasan Awal	Hasil
[‘pelayanan’, ‘ramah’, ‘kebersihan’, ‘kamar’, ‘baik’, ‘menu’, ‘sarapan’, ‘banyak’, ‘pilihan’, ‘masukan’, ‘tolong’, ‘sediakan’, ‘sendal’, ‘pad’, ‘pad’]	[‘pelayanan’, ‘ramah’, ‘kebersihan’, ‘kamar’, ‘baik’, ‘menu’, ‘sarapan’, ‘banyak’, ‘pilihan’, ‘masukan’, ‘sediakan’, ‘sendal’, ‘pad’, ‘pad’]

Proses berikutnya adalah label encoding, yaitu proses mengubah label yang berupa teks menjadi numerik. Pada penelitian ini, data ulasan memiliki 3 label yaitu positif, negatif, dan netral. Hasil label encoding ditunjukkan pada tabel 6.

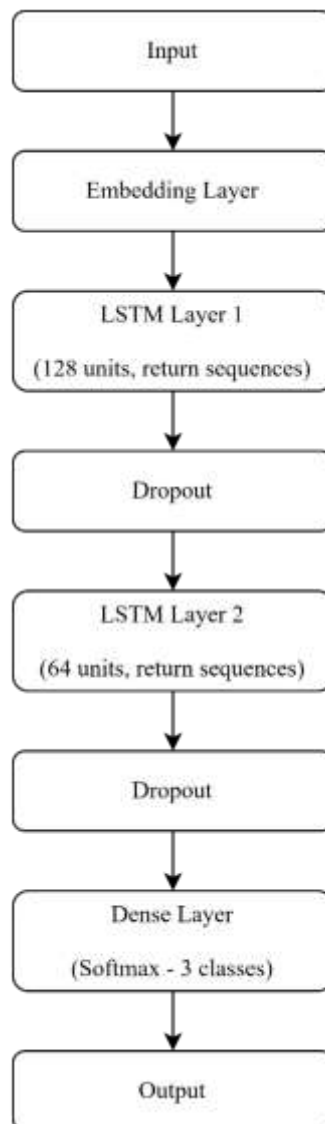
Tabel 6. *Label Encoding*

Label Ulasan Awal	Hasil
Negatif	0
Netral	1
Positif	2

Setelah dilakukan proses-proses tersebut, jumlah dataset menjadi 8.829 baris data. Proses berikutnya adalah splitting data. Proses ini bertujuan memisahkan data ke dalam dua bagian: data pelatihan dan data pengujian. Splitting data dilakukan dengan mengambil 1000 data tiap sentimen dan menggunakan sisanya sebagai data uji, sehingga sebanyak 3000 baris data digunakan sebagai data latih dan 5.829 sisanya digunakan sebagai data uji.

3.4. *Modelling*

Tahap ini dilakukan *modelling* dengan LSTM. Pada penelitian ini, variasi LSTM yang digunakan adalah *Stacked LSTM*. Model dilatih selama 3 iterasi penuh. Proses LSTM ditunjukkan pada gambar 1.

Gambar 1. Proses *Long Short-Term Memory* (LSTM)

Model menggunakan 2 lapisan LSTM berurutan dan berakhir dengan *Dense Layer* beraktivasi *Softmax* klasifikasi 3 kelas. Hasil pelatihan model ditunjukkan pada tabel 6.

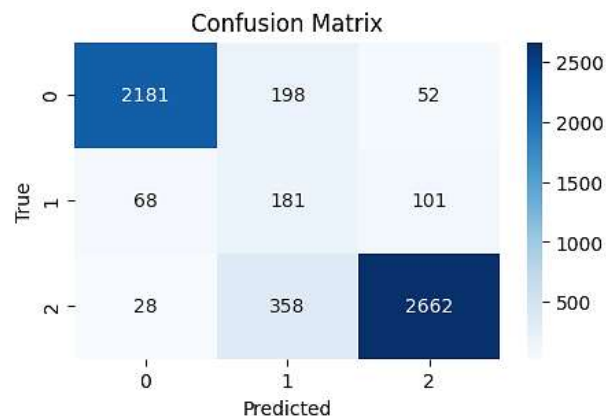
Tabel 7. Hasil Pelatihan Model

Epoch	Accuracy (train)	Loss (train)	Accuracy (val)	Loss (train)
1	55.60 %	0.9117	80.22 %	0.4426
2	80.38 %	0.4862	74.87 %	0.5033
3	87.92 %	0.3150	86.19 %	0.3458

Model menunjukkan peningkatan performa dari *epoch* ke *epoch*. Pada epoch ke-3 validasi dan akurasi pelatihan model menunjukkan performa yang baik dengan overfitting yang relatif kecil.

3.5. Evaluation

Tahap berikutnya adalah mengukur performa model menggunakan confusion matrix yang ditunjukkan gambar 2.



Gambar 2. Confusion Matrix

Berdasarkan *confusion matrix* pada gambar 1, model melakukan 5.024 prediksi dengan benar dari 5829 data uji. Namun pada kelas 1 atau netral, model melakukan 169 kesalahan dari 350 data uji. Jumlah data uji yang tidak seimbang juga mempengaruhi hasil tersebut. Untuk performa keseluruhan, model yang telah dibuat memiliki akurasi sebesar 86.79%, presisi sebesar 71.63%, *recall* 76.25%, dan F1-Score 72.24%.

4. KESIMPULAN

Penelitian ini telah memanfaatkan algoritma *Long Short-Term Memory* (LSTM) dalam penerapan analisis sentimen. Proses *modelling* perlu ditopang dengan data yang seimbang tiap kelasnya agar pelatihan dan pengujian model lebih maksimal. Hasil evaluasi kinerja model menunjukkan bahwa model analisis sentimen yang dibuat memperoleh akurasi sebesar 86.79%, presisi sebesar 71.63%, *recall* 76.25%, dan F1-Score 72.24%. Hasil ini mengindikasikan bahwa model yang telah dibuat memiliki tingkat akurasi yang lebih tinggi dibanding penelitian sebelumnya yang menggunakan algoritma LSTM variasi *Stacked LSTM*.

UCAPAN TERIMAKASIH

Dengan tulus, penulis menyampaikan terima kasih kepada seluruh pihak yang telah membantu, baik secara langsung maupun tidak langsung, dalam proses penyusunan artikel ini.

DAFTAR PUSTAKA

- [1] H. C. Morama, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentreng 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>
- [2] J. Miharja, J. L. Putra, and N. Hadianito, "Comparison of Machine Learning Classification Algorithm on Hotel Review Sentiment Analysis (Case Study: Luminor Hotel Pecenongan)," *J. Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 59–64, 2020, doi: 10.33480/pilar.v16i1.1131.
- [3] S. A. Amira, S. Utama, and M. H. Fahmi, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pelanggan Hotel Di Tripadvisor," *J. Ilmu Komput. dan Sist. Inf.*, vol. 10, no. 2, pp. 40–48, 2022, doi: 10.24912/jiksi.v10i2.22538.
- [4] Z. Ke, J. Sheng, Z. Li, W. Silamu, and Q. Guo, "Knowledge-Guided Sentiment Analysis Via Learning From Natural Language Explanations," *IEEE Access*, vol. 9, pp. 3570–3578, 2021, doi: 10.1109/ACCESS.2020.3048088.
- [5] M. A. Nurrohmah and A. SN, "Sentiment Analysis of Novel Review Using Long Short-Term Memory Method," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 3, p. 209, 2019, doi: 10.22146/ijccs.41236.
- [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] V. Chandradev, I. M. A. D. Suarjaya, and I. P. A. Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *J. Buana Inform.*, vol. 14, no. 02, pp. 107–116, 2023, doi: 10.24002/jbi.v14i02.7244.
- [8] W. Astriningsih and D. H. Fudholi, "Identifikasi Multi Aspek Sentimen Analisis pada Review Hotel Menggunakan Deep learning," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 3, p. 433, 2023, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/5321>
- [9] A. R. Gunawan and R. F. A. Aziza, "Sentiment Analysis Using LSTM Algorithm Regarding Grab Application

- Services in Indonesia,” *J. Appl. Informatics Comput.*, vol. 9, pp. 322–332, 2025, [Online]. Available: <https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/8696/2656>
- [10] M. Khadapi and V. M. Pakpahan, “Analisis Sentimen Berbasis Jaringan LSTM dan BERT terhadap Diskusi,” vol. 6, pp. 130–137, 2024.
- [11] 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, [Online]. Available: <https://aclanthology.org/N19-1423.pdf>
- [12] H. Hidayatullah *et al.*, “PENERAPAN NAÏVE BAYES DENGAN OPTIMASI INFORMATION GAIN DAN SMOTE UNTUK ANALISIS SENTIMEN PENGGUNA APLIKASI CHATGPT,” vol. 7, no. 3, 2023.
- [13] Febriansyah, Efan, and S. Arinda Purnama, “Implementasi Algoritma K-Means Untuk Klasterisasi Besarnya Penggunaan Listrik Rumah Tangga Di Pagar Alam,” *J. Sist. Inf. TGD*, vol. 4, pp. 437–445, 2025.
- [14] D. R. P. Jaya and S. Lestari, “Analisis Sentimen Naturalisasi Tim Nasional Indonesia U-23 di Era Shin Tae-yong Menggunakan Algoritma Naïve Bayes,” vol. 5, no. 3, pp. 3262–3277, 2024.