

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Analisis Sentimen

Istilah "analisis" berasal dari bahasa Inggris *analysis*, yang secara etimologis berasal dari bahasa Yunani Kuno *analusis*. Kata ini tersusun dari dua bagian, yaitu *ana*, yang berarti "kembali", dan *luein*, yang berarti "melepaskan" atau "menguraikan". Jika digabungkan, maknanya menjadi "menguraikan kembali". Secara umum, analisis merujuk pada suatu proses yang melibatkan berbagai tahapan, seperti memisahkan, membedakan, dan mengelompokkan suatu objek atau konsep berdasarkan kriteria tertentu, kemudian menghubungkan elemen-elemen tersebut untuk memahami serta menafsirkan maknanya (Khofifah, dkk., 2022). Secara umum, analisis sentimen terbagi ke dalam dua kategori utama (Santoso, dkk., 2017), yaitu:

1. *Coarse – grained sentiment analysis* merupakan proses menganalisis dan menentukan kecenderungan sentimen dalam suatu dokumen secara keseluruhan. Sentimen ini biasanya dikategorikan ke dalam tiga jenis utama, yaitu positif, netral, dan negatif. Namun, dalam beberapa pendekatan, nilai sentimen tidak bersifat diskrit, melainkan kontinu, sehingga memungkinkan adanya tingkat sentimen yang lebih bervariasi.
2. *Fined – grained sentiment analysis* adalah proses klasifikasi sentimen yang tidak dilakukan pada tingkat dokumen secara keseluruhan, melainkan pada tingkat kalimat di dalam dokumen tersebut. Analisis sentimen ini melibatkan tiga subproses utama (Santoso, dkk., 2017), yaitu:
 - a. *Subjectivity Classification* adalah proses untuk mengidentifikasi apakah suatu kalimat mengandung opini atau tidak.

- b. *Orientation Detection* adalah proses mengkategorikan opini ke dalam tiga jenis sentimen, yaitu positif, negatif, atau netral.
- c. *Opinion Holder and Target Detection* adalah proses mengidentifikasi siapa yang memberikan opini (*opinion holder*) serta objek atau subjek yang menjadi sasaran opini tersebut (*target*).

2.1.2 Ulasan Google Maps Pantai Pudak

Google Maps merupakan layanan peta online gratis yang disediakan oleh *Google*. Aplikasi ini dapat digunakan melalui peramban web maupun perangkat seluler. Pengguna dapat mengaksesnya secara langsung melalui situs web *maps.google.com* (Khofifah, dkk., 2022). *Google Maps* bukan sekadar aplikasi penunjuk jalan, *Google Maps* juga menyediakan informasi komprehensif tentang berbagai tempat, termasuk destinasi wisata dan restoran, serta kondisi dan jalur alternatif jalan. *Google Review* adalah fitur ulasan yang terhubung dengan layanan *Google My Business* (GMB) dan *Google Maps*. Dengan fitur ini, pelanggan dapat membagikan pengalaman mereka secara publik mengenai suatu bisnis, baik dalam hal layanan maupun produk yang ditawarkan (Tukino & Rahman Hakim, 2024).

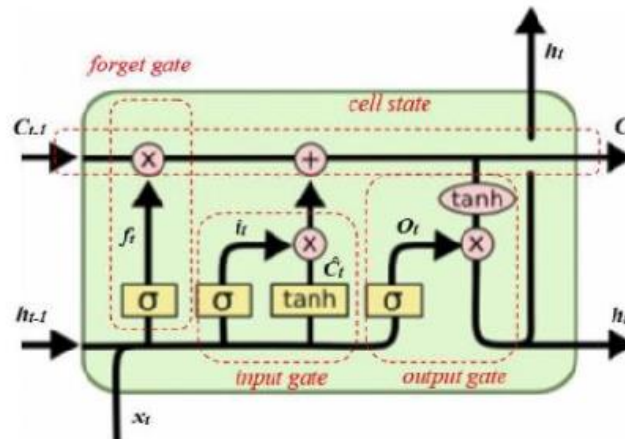
Pantai Pudak, yang terletak di Dusun Banyuurip, Desa Ngadipuro, Kecamatan Wonotirto, Kabupaten Blitar, Jawa Timur, adalah salah satu destinasi wisata yang mulai populer dalam beberapa tahun terakhir. Pantai ini dikenal dengan pasir putihnya yang lembut, air laut yang jernih, serta dikelilingi oleh perbukitan hijau dan tebing *karst* yang menambah keindahan alamnya. Keberadaan pulau karang di lepas pantainya juga menjadi daya tarik tersendiri bagi pengunjung. Pengalaman positif yang dibagikan oleh pengunjung di *Google Maps* dan platform ulasan lainnya dapat meningkatkan minat wisatawan untuk mengunjungi Pantai Pudak. Ulasan-ulasan tersebut memberikan informasi mengenai keindahan pantai, fasilitas

yang tersedia, serta tips bagi calon pengunjung, sehingga membantu mereka merencanakan kunjungan dengan lebih baik.

2.1.3 Long Short-Term Memory

LSTM (*Long Short-Term Memory*) adalah bagian dari keluarga RNN (*Recurrent Neural Networks*), yang merupakan jaringan saraf yang dibangun untuk menangani data sekuensial dengan berbagi bobot internal di seluruh urutan (Setyanto, dkk., 2022). RNN menangkap hubungan temporal antara kata-kata dalam sebuah kalimat. Informasi tekstual adalah data deret waktu dimana urutan kata memainkan peran penting dalam menentukan arti kata dan kalimat. LSTM adalah implementasi dari RNN dengan tautan khusus antar node. Komponen khusus di dalam unit LSTM meliputi *input*, *output*, dan *forget gate* (Elfaik & Nfaoui, 2021).

Gambar 2.1. Arsitektur LSTM



(sumber: Ikhsan, 2023)

Keterangan:

χ_t : Vector input saat ini

h_t : Hidden state saat ini

C_t : Cell state saat ini

Gambar 2.1 adalah arsitektur dari LSTM. Pada LSTM terdapat empat komponen yang berfungsi untuk memproses data. Komponen-komponen tersebut adalah *forget gate*, *input gate*, *update gate*, dan *output gate*. Proses pertama yaitu *forget gate*. *Forget gate* dihitung dari gabungan *hidden state* sebelumnya (h_t) dan *input gate* sekarang (χ_t). *Forget gate* akan dikalikan dengan *cell state timestep* sebelumnya (C_{t-1}). Persamaan 1 merupakan operasi matematika *forget gate*.

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \dots \dots \dots (1)$$

Kemudian selanjutnya adalah proses pada *input gate*. Pada *input gate* ini, terdapat opsi atau pemilihan informasi yang akan diperbaharui untuk ke bagian *cell state*. Proses pemilihan informasi ini menggunakan fungsi *sigmoid* juga. Berikut adalah persamaan dari proses *input gate*.

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \dots \dots \dots (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, X_t]) + b_c \dots \dots \dots (3)$$

Proses yang ketiga adalah *update gate*. Pada tahap ini akan menggunakan *output* dari *input gate*. *Output* dari *update gate* akan menghasilkan *cell state* terbaru (C_t). Berikut adalah cara menghitung proses *update gate*.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \dots\dots\dots(4)$$

Proses yang terakhir yaitu *output gate* (o_t). Pada proses ini akan menghasilkan nilai *output* dan nilai *hidden state* (h_t). Setelah mendapat hasil dari proses *output gate*, maka hasil tersebut akan dilanjutkan ke *timestep* berikutnya ($t+1$).

$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \dots\dots\dots(5)$$

$$h_t = \tanh(C_t) \dots\dots\dots(6)$$

2.1.4 Google Colab

Menurut Masanori Kuroki menulis penelitian berjudul "*Using Python and Google Colab to teach undergraduate microeconomic theory*" dalam *International Review of Economics Education* (Kuroki, 2021). *Google Colab* adalah Lingkungan Pengembangan Terpadu (*Integrated Development Environment*) berbasis web untuk *Python* yang memungkinkan penulisan dan eksekusi kode langsung di peramban web. Alat ini gratis, sumber terbuka, dan populer untuk pengajaran pemrograman, khususnya dalam ilmu data dan pembelajaran mesin. Keuntungan utamanya termasuk eliminasi kebutuhan instalasi *Python* di mesin lokal, memberikan bantuan kepada siswa dengan akses terbatas. *Colab* juga menyediakan lingkungan kolaboratif untuk proyek kelompok, dengan fitur seperti penyorotan kode, otomatisasi penyelesaian, dan pemeriksaan kesalahan untuk meningkatkan efisiensi penulisan dan debugging kode. Keseluruhannya, *Google Colab* adalah alat yang kuat dan ramah pengguna untuk belajar dan mengajar *Python*, cocok bagi siapa saja yang ingin memulai pemrograman atau meningkatkan keterampilan dalam ilmu data dan pembelajaran mesin.

2.1.5 Confusion Matrix

Confusion Matrix adalah salah satu teknik untuk mengevaluasi kinerja sebuah model yang disajikan dalam bentuk tabel. Tabel ini menunjukkan hasil prediksi terhadap dua kelas, seperti yang ditampilkan pada Tabel 2.1 (Ipmawati, dkk., 2024).

Tabel 2.1. Confusion Matrix

Kelas Prediksi	Kelas aktual	
	Positif (+)	Negatif (-)
Positif (+)	TP	FP
Negatif (-)	FN	TN

Keterangan :

- True Positive (TP)*: Ulasan yang diklasifikasikan benar sebagai positif (misalnya, ulasan yang memiliki sentimen positif).
- True Negative (TN)*: Ulasan yang diklasifikasikan benar sebagai negatif (misalnya, ulasan yang memiliki sentimen negatif).
- False Positive (FP)*: Ulasan yang salah diklasifikasikan sebagai positif (misalnya, ulasan yang sebenarnya netral atau negatif tetapi diprediksi sebagai positif).
- False Negative (FN)*: Ulasan yang salah diklasifikasikan sebagai negatif (misalnya, ulasan yang sebenarnya positif atau netral tetapi diprediksi sebagai negatif).

2.2 Kajian Penelitian

Tabel 1. Kajian Penelitian

No.	Judul Penelitian	Penulis (tahun) Nama Jurnal	Hasil/Isi Penelitian
1.	Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps	Wawah Khofifah, Dhian Nur Rahayu, Arif Maulana Yusuf (2022) Jurnal: Jurnal Interkom: Jurnal Pblikasi Ilmiah Bidang Teknologi Informasi	Berdasarkan analisis sentimen terhadap review masyarakat di <i>Google Maps</i> untuk lima pantai di Kabupaten Karawang, ditemukan bahwa Pantai Cibendo dan Pantai Tanjung Baru memiliki nilai negatif, masing-masing 0,550 dan 0,650. Sebaliknya, Pantai Samudera Baru, Pantai Sedari, dan Pantai Pakis mendapatkan nilai positif, yaitu 0,850, 0,700, dan 0,650. Akurasi analisis menggunakan <i>RapidMiner</i> menunjukkan Pantai Samudera Baru memiliki akurasi tertinggi sebesar 85%. Hasil analisis kata-kata yang sering muncul menunjukkan bahwa Pantai Cibendo diidentifikasi dengan kata-kata negatif seperti "kurang" dan "tidak terawat", sedangkan Pantai Samudera Baru dikenal dengan kata-kata positif seperti "bersih"

			dan "bagus". Temuan ini memberikan wawasan penting untuk perbaikan dan pengembangan tempat wisata di daerah tersebut.(Khofifah, dkk., 2022)
2.	Recurrent Neural Network (RNN) Dengan Long Short-Term Memory (LSTM) Untuk Analisis Sentimen Data Instagram	Rudy Cahyadi, Ariesta Damayanti, Dede Aryandi (2020) Jurnal: Jurnal Informatika dan Komputer (JIKO)	Penelitian ini menunjukkan bahwa penggunaan metode <i>Recurrent Neural Network</i> (RNN) dengan <i>Long Short Term Memory</i> (LSTM) efektif dalam melakukan analisis sentimen terhadap komentar di media sosial, khususnya <i>Instagram</i> . Dengan menerapkan teknik <i>preprocessing</i> yang mencakup <i>case folding</i> , penghapusan URL, dan penghilangan tanda baca, sistem berhasil mengklasifikasikan sentimen komentar menjadi positif, netral, atau negatif. Hasil penelitian ini menunjukkan potensi yang besar dalam memahami persepsi masyarakat terhadap konten yang dipublikasikan, serta memberikan kontribusi signifikan dalam pengembangan teknologi analisis sentimen di Indonesia.(Cahyadi, dkk., 2020)
3.	Analisis Sentimen Menggunakan Algoritma LSTM Pada Media Sosial	Andi Aljabar, Arifin A Abd Karim (2022) Jurnal: JUPIKOM	Penelitian ini menunjukkan bahwa penggunaan metode LSTM dalam analisis sentimen terhadap komentar masyarakat mengenai acara televisi sangat efektif, dengan tingkat akurasi mencapai 97% dan loss sebesar 12%. Penelitian ini menegaskan pentingnya pemanfaatan teknologi machine learning untuk memahami opini publik, yang dapat membantu stasiun televisi dalam meningkatkan kualitas tayangan mereka. Dengan demikian, LSTM terbukti sebagai alat yang handal dalam mengolah data teks dan memberikan wawasan yang berharga mengenai persepsi masyarakat terhadap konten yang disajikan. (Aljabar & Karim, 2022)
4.	Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM	Mohammed Hafizh Al-Areef, Kana Saputra S.(2023) Jurnal: Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)	Dalam penelitian ini, Mohammed Hafizh Al-Areef menggunakan pendekatan analisis sentimen untuk mengevaluasi opini publik terhadap calon presiden Indonesia tahun 2024 melalui data Twitter. Metodologi yang diterapkan mencakup tahap pre-processing data, termasuk text cleansing, case folding, stopword removal, tokenization, normalisasi, dan stemming, sebelum menerapkan teknik word embedding dengan fastText. Model Long Short Term Memory (LSTM) digunakan untuk mengklasifikasikan sentimen menjadi positif dan negatif, dengan hasil evaluasi menunjukkan akurasi yang baik untuk masing-masing model calon presiden. Temuan ini memberikan wawasan berharga mengenai preferensi masyarakat dan potensi perubahan yang diharapkan dari para calon pemimpin.(Al-Areef & Saputra S, 2023)

5.	<i>Sentiment Analysis of Tourist Attractions Based on Reviews on Google Maps Using the Support Vector Machine Algorithm</i>	Joang Ipma ^w ati, Saifulloh, Kusnawi (2024) Jurnal: <i>MALCOME: Indonesian Journal of Machine Learning and Computer Science</i>	Penelitian ini menerapkan metode Support Vector Machine (SVM) untuk menganalisis sentimen ulasan pengunjung terhadap tempat wisata di Yogyakarta yang dikumpulkan melalui Google Maps. Hasil analisis menunjukkan tingkat akurasi sebesar 87%, dengan presisi mencapai 90%, recall sebesar 65%, dan skor F1 sebesar 75%. Pengujian menggunakan kata kunci "Tempat Bersih dan Nyaman" menunjukkan bahwa ulasan tersebut tergolong dalam kategori "Positif" dengan probabilitas 86%. Temuan ini mengonfirmasi efektivitas SVM dalam analisis sentimen serta memberikan wawasan berharga bagi pengelola tempat wisata dalam memahami persepsi pengunjung dan meningkatkan kualitas layanan mereka.(Ipma ^w ati, dkk., 2024)
6.	Analisis Sentimen Komentar Youtube Terhadap Ceramah Ning Umi Laila Sindir Rhoma Irama Menggunakan Algoritma LSTM	Septian Ade Putra, Windu Gata (2024) Jurnal : Jurnal Ilmiah Komputer	Hasil penelitian ini mengungkap bahwa algoritma Long Short-Term Memory (LSTM) terbukti sangat efektif dalam menganalisis sentimen komentar di YouTube, khususnya terhadap ceramah Neng Umi Laila yang mengkritik Rhoma Irama. Dengan menerapkan teknik pemrosesan bahasa alami dan mengumpulkan data dari 1.108 komentar, penelitian ini berhasil mencapai tingkat akurasi tinggi sebesar 96%. Analisis yang dilakukan memberikan wawasan mendalam mengenai respons masyarakat terhadap konten tersebut, termasuk pola serta tren sentimen publik di media sosial. Temuan ini diharapkan dapat memberikan manfaat bagi pembuat kebijakan dan peneliti dalam memahami opini publik serta mendukung pengambilan keputusan yang lebih tepat.(Putra & Gata, 2024)
7.	Penerapan Algoritma Sentiment Analysis Dan Naïve Bayes Terhadap Opini Pengunjung Di Tempat Wisata Pantai Pintu Kota, Kota Ambon	Julius Chrisostomus Aponno (2022) Jurnal : Jurnal Teknik Informatika dan Sistem Informasi	Penelitian ini menunjukkan bahwa analisis sentimen terhadap ulasan pengunjung di Pantai Pintu Kota, Ambon, menggunakan metode Naïve Bayes menghasilkan akurasi yang tinggi, yaitu 90.65%. Hasil analisis ini mengindikasikan bahwa pengunjung memberikan umpan balik yang beragam mengenai pengalaman mereka, dengan beberapa komentar positif mengenai keindahan pantai dan fasilitas yang ada, sementara ada juga kritik terkait kebersihan dan infrastruktur. Temuan ini dapat menjadi acuan bagi pengelola wisata dan pemerintah untuk meningkatkan kualitas dan infrastruktur tempat wisata, sehingga Pantai Pintu Kota dapat menjadi destinasi unggulan yang lebih menarik bagi wisatawan. Selain itu, disarankan untuk mengembangkan sistem online yang dapat mengumpulkan data ulasan secara otomatis dan menciptakan kamus khusus untuk

			analisis bahasa Indonesia yang tidak baku.(Aponno, 2022)
8.	Analisis Sentimen Terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Long Short-Term Memory	Muhammad Ikhsan dan Sri Ngudi Wahyuni (2023) Jurnal : IJCSR: The Indonesian Journal of Computer Science Research	Hasil evaluasi menunjukkan akurasi tertinggi mencapai 90% pada beberapa percobaan, dengan pola prediksi yang cenderung lebih tinggi untuk sentimen positif dibandingkan negatif. Temuan ini mengindikasikan bahwa model LSTM dapat diandalkan dalam menganalisis sentimen pengguna Twitter terkait isu kenaikan harga bahan bakar minyak. Penelitian selanjutnya disarankan untuk mengeksplorasi metode pelabelan data yang berbeda dan menggunakan word embedding GloVe dengan dimensi lebih tinggi untuk meningkatkan akurasi dan pemahaman pola sentimen.(Ikhsan & Wahyuni, 2023)
9.	Analisis Sentimen Wisatawan terhadap Taman Nasional Bunaken dan Top 10 Hotel Rekomendasi Tripadvisor Menggunakan Algoritma SVM dan DT berbasis CRISP-DM	Yerik Afrianto Singgalen (2023) Jurnal : Journal of Computer System and Informatics (JoSYC)	Penelitian ini berhasil mengidentifikasi dan menganalisis sentimen wisatawan terhadap Taman Nasional Bunaken serta rekomendasi 10 hotel terbaik berdasarkan ulasan di Tripadvisor. Dengan menggunakan kerangka kerja Cross Industry Standard Process for Data Mining (CRISP-DM), penelitian ini menerapkan algoritma Decision Tree (DT) dan Support Vector Machine (SVM) untuk menganalisis data ulasan. Hasilnya menunjukkan bahwa algoritma SVM, terutama dengan penerapan operator SMOTE, memberikan performa yang lebih baik dalam hal akurasi, presisi, recall, dan nilai AUC. Temuan ini menegaskan pentingnya analisis sentimen dalam memahami kepuasan wisatawan dan dapat menjadi dasar untuk pengoptimalan manajemen destinasi wisata di Taman Nasional Bunaken.(Singgalen, 2023)
10.	Nonlinear dynamic soft sensor modeling with supervised long short-term memory network	Xiaofeng Yuan, Lin Li, Yalin Wang (2020) Jurnal : IEEE TRANSACTION ON INDUSTRIAL INFORMATICS	In conclusion, the study demonstrates the effectiveness of the supervised long short-term memory (SLSTM) network in enhancing the predictive modeling of complex industrial processes, such as penicillin fermentation. By addressing the challenges associated with traditional recurrent neural networks (RNNs) and standard LSTMs, the SLSTM model successfully captures the nonlinear dynamic behaviors of the process, leading to improved accuracy in predicting key variables. The results indicate that the SLSTM outperforms its counterparts, particularly in terms of reducing prediction errors, thereby offering a robust solution for process optimization and control in industrial applications. This advancement not only contributes to the field of industrial informatics but also paves the way for more sophisticated modeling techniques in

			various biochemical processes.(Yuan, dkk., 2020)
11.	Deep Learning with Long Short-Term Memory for Time Series Prediction	Yuxiu Hua, Zhifeng Zhao, Rongpeng Li, Xianfu Chen, Zhiming Liu, Honggang Zhang (2019) Jurnal : IEEE International Conference on Computer Communication and the Internet	In conclusion, the RCLSTM model presents a significant advancement in the field of time series prediction, particularly within telecommunications. By integrating stochastic connectivity into the conventional LSTM architecture, RCLSTM not only reduces computational complexity but also maintains robust performance in traffic and user mobility predictions. This innovative approach addresses the challenges posed by the dynamic nature of network traffic and user behavior, making it a valuable tool for ensuring quality of service in increasingly complex mobile communication environments. As the demand for efficient data processing continues to grow, the RCLSTM model stands out as a promising solution that balances accuracy and computational efficiency.(Hua, dkk., 2019)

Dari 11 penelitian yang paling mendekati penelitian ini adalah penelitian yang dilakukan oleh Septiawan Ade Putra dan Windu Gata pada tahun (2024) yang berjudul “Analisis Sentimen Komentar Youtube Terhadap Ceramah Ning Umi Laila Sindir Rhoma Irama Menggunakan Algoritma LSTM”. Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada objek yang dikaji, sementara keunggulan penelitian ini adalah penggunaan algoritma LSTM untuk menganalisis sentimen pengunjung terhadap fasilitas dan layanan di destinasi wisata, yakni Pantai Pudak. Selain itu, penelitian ini menitikberatkan pada pengalaman langsung pengguna, berbeda dari penelitian sebelumnya yang lebih berfokus pada analisis komentar di platform digital. Penelitian ini diharapkan dapat berkontribusi dalam meningkatkan kualitas fasilitas dan layanan di bidang pariwisata dengan memanfaatkan analisis sentimen berbasis kecerdasan