

ANALISIS TREND WISATA UNTUK MENDUKUNG PENGEMBANGAN PAKET WISATA MENARIK

TUGAS BESAR DATA MINING

Oleh

Haris Saefuloh	714220061
Muhammad Rayfan Aqbillah	714220044
Nida Sakina Aulia	714220040
Mariana Siregar	714220068



**PROGRAM STUDI DIPLOMA IV TEKNIK INFORMATIKA
UNIVERSITAS LOGISTIK DAN BISNIS INTERNASIONAL
BANDUNG
2025**

HALAMAN PERNYATAAN ORISINALITAS

Laporan ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang ada berkaitan etika dan kaidah penulisan karya ilmiah yang berlaku, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku

Yang menyatakan,

Nama : Haris Saefuloh

NIM : 714220061

Tanda Tangan :

Tanggal : Juli 2025

Mengetahui,

Ketua : (.....tanda tangan.)

Dosen Pengampu Mata Kuliah : (.....tanda tangan.)

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, laporan dengan judul “Analisis Tren Wisata: Dari Pengumpulan Data hingga Dataset Awal” ini dapat disusun dan diselesaikan dengan baik. Laporan ini disusun sebagai bagian dari pemenuhan tugas proyek data mining, dengan fokus pada penerapan metode CRISP-DM dan algoritma Long Short-Term Memory (LSTM) untuk menganalisis tren kunjungan wisata serta memberikan rekomendasi destinasi berdasarkan prediksi dan rating pengunjung.

Kami mengangkat studi kasus dua wilayah wisata, yaitu Kabupaten Trenggalek dan kawasan Ancol, yang memiliki karakteristik destinasi berbeda namun sama-sama populer. Dengan memanfaatkan data historis kunjungan wisata dan ulasan dari Google Maps, kami mencoba menyusun model prediktif serta sistem rekomendasi berbasis data.

Kami menyadari bahwa laporan ini masih jauh dari sempurna. Oleh karena itu, kami sangat terbuka terhadap saran dan kritik yang membangun demi perbaikan di masa mendatang. Ucapan terima kasih kami sampaikan kepada dosen pembimbing dan semua pihak yang telah memberikan dukungan dan arahan selama proses penyusunan laporan ini.

Semoga laporan ini dapat memberikan manfaat dan menjadi referensi yang berguna, baik bagi kami sendiri maupun bagi pembaca yang memiliki ketertarikan dalam bidang data mining dan pengembangan pariwisata berbasis data.

Bandung, Juli 2025

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Logistik Bisnis Internasional, saya yang bertanda tangan di bawah ini:

Nama : Haris Saefuloh

NIM : 714220061

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Logistik Bisnis Internasional, Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

ANALISIS TREND WISATA UNTUK PENGEMBANGAN PAKET WISATA MENARIK

Beserta perangkat yang ada (jika diperlukan). Dengan Hak ini Universitas Logistik Bisnis Internasional Hayati berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bandung

Pada tanggal : Juli 2025

Yang menyatakan

Haris Saefuloh
Ketua Kelompok

ABSTRAK

Industri pariwisata terus berkembang seiring dengan perubahan preferensi wisatawan dan kemajuan teknologi digital. Di era big data, informasi dari ulasan online dan media sosial dapat dimanfaatkan untuk menganalisis tren wisata secara mendalam. Penelitian ini bertujuan untuk menganalisis tren wisata berdasarkan data kunjungan dan ulasan destinasi dari tahun 2020 hingga 2024 guna membantu agen perjalanan dalam merancang paket wisata terbaik. Metodologi yang digunakan mencakup analisis eksploratif data, visualisasi tren, dan penerapan model prediktif untuk mengidentifikasi destinasi unggulan dan preferensi wisatawan. Hasil analisis menunjukkan adanya perubahan signifikan dalam jenis wisata yang diminati, terutama setelah pandemi, dengan peningkatan minat pada wisata alam dan budaya. Temuan ini dapat digunakan sebagai dasar perancangan paket wisata yang lebih relevan dan menarik bagi target pasar, sekaligus meningkatkan daya saing agen perjalanan di era digital.

Kata Kunci : Pariwisata, Preferensi wisatawan, Teknologi digital

ABSTRACT

The tourism industry continues to evolve in line with changing tourist preferences and advances in digital technology. In the era of big data, information from online reviews and social media can be used to analyze tourism trends in depth. This study aims to analyze tourism trends based on destination visitation and review data from 2020 to 2024 to assist travel agencies in designing the best tour packages. The methodology used includes exploratory data analysis, trend visualization, and the application of predictive models to identify top destinations and tourist preferences. The analysis results show significant changes in the types of tourism that are in demand, especially after the pandemic, with an increase in interest in nature and cultural tourism. These findings can be used as a basis for designing more relevant and attractive tour packages for the target market, while also enhancing the competitiveness of travel agencies in the digital era.

Keywords : Tourism, Tourist preferences, Digital technology

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS.....	II
KATA PENGANTAR	III
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	1
ABSTRAK.....	2
ABSTRACT	3
DAFTAR ISI	4
DAFTAR TABEL	6
DAFTAR GAMBAR	7
BAB I PENDAHULUAN	8
1.1 LATAR BELAKANG	8
1.2 RUMUSAN MASALAH.....	9
1.3 TUJUAN PENELITIAN.....	9
1.4 MANFAAT PENELITIAN.....	9
1.5 RUANG LINGKUP.....	10
BAB II TINJAUAN PUSTAKA	11
2.1 TREN ANALISIS	11
2.2 DATA MINING.....	11
2.3 DEEP LEARNING.....	11
2.4 LONG SHORT-TERM MEMORY(LSTM).....	12
2.5 ARIMA	12
2.6 ANALISIS SENTIMEN	13
2.7 VISUALISASI	13
2.8 STATE OF THE ART	13
BAB III METODOLOGI PENELITIAN.....	15
3.1 TAHAPAN PENELITIAN.....	15
3.2 DESKRIPSI DATASET.....	16
3.3 ALGORITMA DATA MINING.....	16
3.4 EVALUASI KERJA	17
BAB IV HASIL DAN PEMBAHASAN	19
4.1 VISUAL EKSPLORATIF	19
4.1.1 Grafik Tren Kunjungan (Line Plot).....	19
4.1.2 Grafik Distribusi Rating Ulasan (Histogram/Bar Plot)	19
4.1.3 Word Cloud atau Frekuensi Kata Kunci (Word Cloud)	20
4.1.4 Grafik Perbandingan Jumlah Kunjungan atau Rata-rata Rating Antar Destinasi Wisata yang Berbeda (Bar Plot)	21
4.1.5 Perbandingan Destinasi Berdasarkan Rata-rata Rating.....	21
4.2 HASIL PREPROCESSING DAN PEMODELAN	22
4.2.1 Hasil Preprocessing Data	22
4.2.2 Proses Pelatihan Model	24
4.3 TABEL HASIL EKSPERIMENT/MODEL.....	25
4.4 INTERPRETASI HASIL	26
4.4.1 Model LSTM.....	26
4.4.2 Model ARIMA.....	27
4.5 ANALISIS KEUNGGULAN DAN KETERBATASAN	28

BAB V KESIMPULAN DAN SARAN	31
5.1 RINGKASAN TEMUAN UTAMA.....	31
5.2 JAWABAN ATAS RUMUSAN MASALAH.....	31
5.3 SARAN UNTUK PENGEMBANGAN LEBIH LANJUT.....	32
DAFTAR PUSTAKA	33
LAMPIRAN	35

DAFTAR TABEL

Tabel 2. 1 Studi Terdahulu yang Menggunakan Metode LSTM pada Data Wisata	14
Tabel 4. 1 Evaluasi Model LSTM	26
Tabel 4. 2 Evaluasi Model ARIMA	26

DAFTAR GAMBAR

Gambar 4. 1 Grafik Line Plot.....	19
Gambar 4. 1 Grafik Line Plot.....	Error! Bookmark not defined.
Gambar 4. 2 Bar Plot.....	20
Gambar 4. 2 Bar Plot.....	Error! Bookmark not defined.
Gambar 4. 3 Word Cloud	20
Gambar 4. 3 Word Cloud	Error! Bookmark not defined.
Gambar 4. 4 Bar Plot.....	21
Gambar 4. 4 Bar Plot.....	Error! Bookmark not defined.
Gambar 4. 5 Rata Rata Rating.....	21

BAB I

PENDAHULUAN

1.1 Latar Belakang

Industri pariwisata merupakan salah satu sektor ekonomi yang paling dinamis dan berkembang pesat di dunia [1]. Di Indonesia, pariwisata menyumbang kontribusi signifikan terhadap Produk Domestik Bruto (PDB), penyerapan tenaga kerja, dan devisa negara.

Dinamika industri pariwisata sangat dipengaruhi oleh berbagai faktor eksternal dan internal, termasuk perubahan trend sosial, kondisi ekonomi global dan regional, perkembangan teknologi, serta preferensi wisatawan yang terus berevolusi [2]. Di era digital ini, cara wisatawan merencanakan dan mengalami perjalanan sangat dipengaruhi oleh informasi online dan media sosial. Mereka aktif mencari ulasan, membandingkan destinasi, dan berbagi pengalaman di platform seperti TripAdvisor, Google Review, dan Instagram [3]. Perilaku ini menghasilkan data masif yang bisa dipakai untuk mengidentifikasi tren wisata terkini secara real-time.

Tren wisata menunjukkan apa yang disukai wisatawan, mulai dari jenis tempat tujuan, kegiatan yang ingin mereka lakukan, hingga hal-hal pendukung seperti keamanan atau kemudahan akses. Misalnya, setelah pandemi COVID-19, banyak orang lebih suka wisata alam dan ekowisata karena ingin tempat yang terbuka dan alami. Begitu juga, wisata yang fokus pada pengalaman seperti budaya, kuliner, atau petualangan, kini makin populer di kalangan anak muda.

Agen perjalanan sebagai salah satu komponen penting dalam ekosistem pariwisata menghadapi tantangan untuk tetap relevan dan kompetitif dalam menyediakan produk dan layanan yang sesuai dengan kebutuhan dan keinginan pasar. Untuk mencapai hal tersebut, agen perjalanan perlu menganalisis data secara mendalam tentang trend wisata yang sedang berkembang, preferensi wisatawan, serta destinasi-destinasi yang sedang popular [4].

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat diidentifikasi beberapa permasalahan utama dalam industri pariwisata yang menjadi fokus penelitian ini:

1. Bagaimana cara mengoptimalkan pemanfaatan data kunjungan dan ulasan digital oleh agen perjalanan?
2. Bagaimana agen perjalanan dapat mengikuti perubahan tren dan preferensi wisata yang cepat berubah?
3. Bagaimana mengintegrasikan data ulasan dan kunjungan untuk menyusun informasi destinasi yang lebih terstruktur?
4. Mengapa banyak paket wisata yang tidak sesuai dengan preferensi wisatawan, dan bagaimana solusinya?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Menganalisis data digital yang berasal dari ulasan online, media sosial, dan platform digital lainnya untuk mengidentifikasi tren wisata dan preferensi wisatawan secara real-time.
2. Mengembangkan pendekatan analitik berbasis data untuk memantau perubahan tren wisata akibat pengaruh faktor sosial, ekonomi, maupun peristiwa global.
3. Menyusun informasi terstruktur mengenai destinasi wisata populer dan sentimen pengunjung dari berbagai sumber data tidak terstruktur.
4. Memberikan rekomendasi pengembangan paket wisata yang lebih relevan dan sesuai dengan kebutuhan wisatawan berdasarkan hasil analisis data tren dan preferensi.

1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah:

1. Manfaat Teoritis:
 - Menambah wawasan dan referensi ilmiah dalam bidang analisis tren wisata berbasis data.
 - Memberikan kontribusi terhadap literatur mengenai penerapan data mining atau analitik big data dalam sektor pariwisata.
2. Manfaat Praktis:
 - Memberikan panduan bagi agen perjalanan untuk memanfaatkan data digital dalam menyusun strategi pengembangan produk dan layanan wisata.

- Membantu agen perjalanan untuk memahami preferensi wisatawan secara lebih akurat dan terkini.
- Menyediakan model analisis yang dapat digunakan untuk mengembangkan paket wisata yang lebih responsif terhadap tren dan kebutuhan pasar.
- Mendukung pelaku industri pariwisata dalam mengambil keputusan yang lebih tepat dan berbasis data.

1.5 Ruang Lingkup

Penelitian ini akan dibatasi pada beberapa aspek berikut:

Ruang Lingkup Geografis:

- Fokus pada destinasi wisata domestik dan internasional yang populer di kalangan wisatawan Indonesia
- Analisis khusus terhadap destinasi wisata di Asia Tenggara sebagai regional priority

Ruang Lingkup Temporal:

- Periode analisis data: 3 tahun terakhir (2022-2024)
- Proyeksi trend untuk 2 tahun ke depan (2025-2026)

Ruang Lingkup Substantif:

- Analisis trend destinasi wisata berdasarkan jumlah kunjungan
- Analisis sentimen dan preferensi berdasarkan ulasan pengunjung
- Identifikasi faktor-faktor yang mempengaruhi popularitas destinasi
- Pengembangan rekomendasi paket wisata berdasarkan temuan analisis

Ruang Lingkup Metodologis:

1. Penggunaan metode analisis kuantitatif dan kualitatif
2. Pemanfaatan big data dan teknik data mining
3. Analisis sentimen berbasis natural language processing.

BAB II

TINJAUAN PUSTAKA

2.1 Tren Analisis

Analisis tren wisata merupakan pendekatan untuk mengamati dan memahami perubahan pola perilaku wisatawan, termasuk preferensi destinasi, jenis aktivitas yang diminati, serta faktor-faktor yang memengaruhinya seperti sosial, ekonomi, budaya, dan teknologi. Di era digital, data dari media sosial, ulasan daring, dan platform perjalanan menjadi sumber informasi yang kaya untuk mengidentifikasi tren secara real-time. Menurut Naquitalia et al. (2022), Penerapan teknik analisis sentimen berbasis aspek menggunakan metode deep learning seperti CNN dapat membantu mengklasifikasikan opini wisatawan terhadap aspek-aspek tertentu, seperti makanan halal, fasilitas toilet, dan tempat ibadah. Hasil analisis tren ini sangat bermanfaat bagi pelaku industri pariwisata dalam menyusun strategi pemasaran dan pengembangan layanan yang lebih responsif terhadap kebutuhan pasar [5].

2.2 Data Mining

Data mining adalah proses untuk menggali informasi penting dan tersembunyi dari kumpulan data besar dengan menggunakan teknik-teknik seperti kecerdasan buatan, statistika, dan machine learning. Proses ini bertujuan untuk menemukan pola, hubungan, dan tren yang berguna dalam pengambilan keputusan, serta mengubah data mentah menjadi pengetahuan yang bernilai. Data mining berperan penting dalam berbagai bidang seperti bisnis, pendidikan, dan teknologi informasi karena kemampuannya dalam mengolah data kompleks menjadi informasi yang mudah dipahami. Dengan bantuan perangkat lunak seperti RapidMiner, proses data mining dapat dilakukan secara efisien melalui tahapan seperti pembersihan data, integrasi, transformasi, penerapan algoritma, hingga evaluasi dan visualisasi hasil analisis [6].

2.3 Deep Learning

Deep Learning adalah cabang dari Machine Learning yang meniru cara kerja otak manusia melalui struktur yang dikenal sebagai Artificial Neural Networks (ANN) atau Jaringan Saraf Tiruan. Metode ini memungkinkan sistem komputer untuk belajar dari data dalam jumlah besar dan kompleks secara otomatis tanpa perlu diprogram secara eksplisit. Deep Learning sangat efektif dalam menangani data berformat gambar, suara, dan teks melalui algoritma seperti Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), dan Long Short-Term Memory (LSTM). Dengan struktur berlapis-lapis

(deep), Deep Learning mampu mengenali pola dan hubungan tersembunyi dalam data sehingga menghasilkan prediksi dan klasifikasi yang lebih akurat dalam berbagai aplikasi seperti pengenalan wajah, pemrosesan bahasa alami, hingga kendaraan otonom [7].

2.4 Long Short-Term Memory(LSTM)

Long Short-Term Memory (LSTM) adalah pengembangan dari arsitektur Recurrent Neural Network (RNN) yang dirancang untuk memproses data sekuensial dan mampu mengatasi masalah vanishing gradient yang sering terjadi pada RNN saat mengolah data dalam jangka panjang. LSTM dilengkapi dengan struktur memori khusus yang terdiri dari tiga gerbang utama: input gate, forget gate, dan output gate, yang memungkinkan jaringan untuk menyimpan atau membuang informasi secara selektif. Keunggulan LSTM terletak pada kemampuannya mempertahankan konteks informasi yang relevan dalam urutan data, sehingga sangat cocok digunakan dalam analisis sentimen, prediksi tren, dan pengolahan teks. Penelitian oleh Hanafiah et al. (2023) menunjukkan bahwa LSTM mampu memberikan hasil prediksi yang akurat dalam kasus prediksi harga saham, sehingga relevan juga untuk diterapkan dalam konteks analisis data wisata yang bersifat temporal atau berbasis urutan Waktu [8].

2.5 ARIMA

ARIMA (Autoregressive Integrated Moving Average) adalah model yang digunakan untuk menganalisis dan melakukan peramalan terhadap data runtun waktu (time series), terutama ketika data memiliki pola atau tren tertentu di masa lalu yang diasumsikan akan berlanjut di masa depan[9].

Model ini memiliki tiga parameter utama yang diperoleh melalui analisis grafik dekomposisi dan grafik ACF serta PACF:

- p (order of autoregression): menunjukkan jumlah lag (data masa lalu) yang digunakan.
- d (order of integration): jumlah diferensiasi yang dibutuhkan untuk membuat data menjadi stasioner (tidak memiliki tren).
- q (order of moving average): jumlah rata-rata bergerak dari error sebelumnya yang digunakan dalam model.

Model ARIMA dibagi menjadi:

- AR (Autoregressive): model yang memanfaatkan hubungan antara nilai saat ini dan nilai sebelumnya.
- MA (Moving Average): model yang memanfaatkan hubungan antara nilai saat ini dan kesalahan sebelumnya.

- ARIMA (campuran): gabungan antara AR dan MA setelah dilakukan diferensiasi (untuk data non-stasioner).

2.6 Analisis Sentimen

Analisis sentimen adalah proses dalam Natural Language Processing (NLP) untuk mengklasifikasikan teks menjadi sentimen positif, negatif, atau netral. Dalam pariwisata, teknik ini digunakan untuk memahami opini pengunjung dari ulasan digital. Prosesnya meliputi pembersihan data, tokenisasi, ekstraksi fitur (misalnya Word2Vec), dan klasifikasi, salah satunya menggunakan algoritma Long Short-Term Memory (LSTM). Metode LSTM efektif dalam menganalisis data teks karena mampu memahami konteks kalimat. Studi oleh Nurvania et al. (2021) menunjukkan LSTM berhasil mengklasifikasikan ulasan wisata di TripAdvisor dengan akurasi 71,67% [10].

2.7 Visualisasi

Visualisasi data adalah proses mengubah data menjadi bentuk visual seperti gambar, diagram, atau animasi yang bertujuan untuk menjelaskan informasi secara lebih jelas dan menarik. Visualisasi membantu menganalisis karakteristik data serta hubungan antar atribut dalam data. Ini merupakan salah satu teknik eksplorasi data yang paling efektif dan menarik [11].

Visualisasi tidak hanya memperindah tampilan data, tetapi juga sangat berguna untuk:

- Memahami pola dan tren dalam data
- Menyajikan informasi kepada pengguna dengan cara yang mudah dimengerti
- Menunjang pengambilan keputusan berdasarkan data

2.8 State of The Art

State of the Art (SoTA) adalah langkah awal untuk menunjukkan kebaruan penelitian dengan memetakan pencapaian paling mutakhir dalam suatu bidang. SoTA berfungsi menegaskan bahwa penelitian yang dilakukan belum pernah dipublikasikan sebelumnya, serta mencerminkan hasil tertinggi dari pengembangan yang ada, baik berupa metode, produk, atau solusi. Dengan SoTA, peneliti dapat memahami masalah penelitian secara lebih dalam, memilih pendekatan yang tepat, serta menghindari duplikasi. Penyusunan SoTA memerlukan pemilihan literatur yang relevan, analisis mendalam, dan kemampuan menulis yang baik untuk menyajikan informasi secara kritis dan sistematis [12].

Mukhtar et al. (2022) mengembangkan model prediksi kunjungan wisatawan ke Indonesia dengan menggunakan metode Long Short-Term Memory (LSTM). Model ini digunakan untuk mengolah data historis dari Google Trends dan BPS setelah dilakukan pembersihan noise

menggunakan Hilbert-Huang Transform (HHT). LSTM terbukti mampu menangkap pola jangka panjang dalam data kunjungan wisata, meskipun pada kondisi ekstrem seperti pandemi, tingkat akurasi mengalami penurunan [13].

Afidah et al. (2022) memanfaatkan Bidirectional LSTM (Bi-LSTM) untuk melakukan analisis sentimen terhadap ulasan wisatawan. Penggunaan LSTM dalam konteks ini menunjukkan kemampuan dalam memahami struktur teks secara sekuensial dan dua arah, sehingga efektif dalam mengklasifikasikan sentimen wisatawan terhadap destinasi wisata. Hasilnya, Bi-LSTM mencapai akurasi tinggi dan menunjukkan keunggulan dibanding model klasifikasi biasa [14].

Putra et al. (2024) menggunakan LSTM sebagai salah satu model dalam proses klasifikasi sentimen berbasis aspek pada ulasan wisatawan terhadap destinasi di Kalimantan Tengah. LSTM digunakan untuk mengevaluasi akurasi sentimen berdasarkan entitas seperti environment dan location. Meskipun hasilnya cukup baik dengan akurasi di atas 80%, LSTM masih berada di bawah model BERT dalam hal performa klasifikasi aspek dan sentimen [15].

Penelitian ini melanjutkan pemanfaatan LSTM dengan mengintegrasikan kemampuannya dalam memprediksi tren data waktu x(time series) serta potensi integrasi dengan analisis persepsi wisatawan. Fokus utama diarahkan pada penggunaan LSTM untuk mendukung perancangan paket wisata berbasis tren aktual dan kecenderungan minat wisatawan yang terekam dalam data digital.

No	Peneliti & Tahun	Metode	Sumber Data	Hasil Utama
1	Mukhtar et al. (2022)	LSTM	Google Trends & BPS (2008–2021)	LSTM mampu memodelkan tren jangka panjang kunjungan wisata [13].
2	Afidah et al. (2022)	LSTM	TripAdvisor (10.000 ulasan)	Bi-LSTM menghasilkan akurasi klasifikasi sentimen sebesar 96,86% [14].
3	Putra et al. (2024)	LSTM	Google Maps (ulasan wisata Kalteng, 2017–2024)	LSTM memberikan akurasi cukup baik (84% untuk sentimen, 77–89% untuk aspek) [15].

Tabel 2. 1 Tabel Studi Terdahulu yang Menggunakan Metode LSTM pada Data Wisata

BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Pada penelitian ini, pendekatan metodologi yang digunakan adalah CRISP-DM (Cross Industry Standard Process for Data Mining). CRISP-DM merupakan metodologi standar yang banyak digunakan dalam berbagai industri dan penelitian akademik karena memiliki struktur sistematis dan fleksibel. Untuk tahapan dari metodologi ini adalah sebagai berikut:

1. Pemahaman bisnis (Business Understanding)

Tahap ini bertujuan untuk memahami tujuan utama dari penelitian, yaitu menganalisis tren kunjungan wisata dan memberikan rekomendasi tempat wisata terbaik berdasarkan jumlah pengunjung.

2. Pemahaman data (Data Understanding)

Tahapan ini mencakup pengumpulan dan eksplorasi awal terhadap dataset. Data diperoleh dari berbagai sumber resmi diantaranya Satu Data Trenggalek, laporan tahunan PT Ancol, serta ulasan Google Maps yang di-scrapping menggunakan Outscraper. Dataset terdiri dari informasi jumlah pengunjung per tahun dan rata-rata rating dari ulasan.

3. Persiapan data (Data Preparation)

Pada tahap ini dilakukan pembersihan dan transformasi data agar siap digunakan dalam pemodelan. Diantaranya yaitu menghapus header yang tidak relevan, melakukan cleaning text, menghapus data yang kosong/tidak lengkap, menggabungkan data kunjungan dan rating dalam satu dataset, dan menyesuaikan urutan data berdasarkan tempat dan tahun.

4. Pemodelan (Modeling)

Tahap pemodelan dilakukan menggunakan algoritma LSTM (Long Short-Term Memory) karena sifat data berupa deret waktu (time series). LSTM dilatih untuk memprediksi jumlah kunjungan di masa depan berdasarkan data historis. Selain itu, digunakan juga informasi avg_rating dari review untuk membantu menyusun sistem rekomendasi berbasis popularitas. Tools yang digunakan antara lain Python, TensorFlow, Keras, dan Scikit-learn.

5. Evaluasi (Evaluation)

Setelah model dilatih, dilakukan evaluasi untuk mengukur akurasi dan efektivitas model. Metrik yang digunakan adalah MAE (Mean Absolute Error), RMSE (Root

Mean Squared Error), dan MAPE (Mean Absolute Percentage Error). Evaluasi juga mencakup penggabungan hasil prediksi dengan rating untuk menyusun popularity score.

6. Implementasi atau penyebaran hasil (Deployment)

Tahap akhir melibatkan penerapan hasil ke dalam bentuk rekomendasi yang bisa dimanfaatkan oleh pihak pemangku kebijakan atau platform pariwisata. Berdasarkan prediksi jumlah pengunjung dan skor rata-rata ulasan, disusun 5 daftar teratas destinasi wisata yang paling direkomendasikan. Hasil ini dapat dikembangkan lebih lanjut menjadi paket wisata atau modul rekomendasi otomatis untuk mendukung strategi promosi wisata berbasis data.

3.2 Deskripsi Dataset

Dataset yang digunakan pada penelitian ini berasal dari dua wilayah destinasi wisata, diantaranya Kabupaten Trenggalek dan kawasan Ancol. Data tren mengenai jumlah kunjungan tahunan diambil dari situs resmi instansi pemerintah seperti Portal Satu Data Kabupaten Trenggalek dan PT Pembangunan Jaya Ancol. Selain dataset kuantitatif, penelitian ini juga memanfaatkan data kualitatif berupa ulasan pengunjung yang diambil dari Google Maps melalui metode scraping.

Dataset kunjungan wisata Trenggalek mencakup tahun 2021, 2022, dan 2024 dengan ukuran sekitar 15 baris per tahun dan memiliki tiga atribut utama, yaitu No (integer), Tempat Wisata (string), dan Jumlah kunjungan (integer). Sementara itu, dataset kunjungan wisata Ancol mencakup periode 2021 hingga 2024 dengan ukuran sekitar 15 baris per tahun dan empat atribut, yaitu Nama Tempat (string), Tahun (integer), dan Total kunjungan (integer). Adapun dataset ulasan Google Maps terdiri dari 20 hingga 250 baris per tempat wisata dan total keseluruhan mencapai 1900 dengan empat atribut utama: Nama Tempat (string), Review Text (string), Review Rating (integer), dan Datetime (datetime). Dataset-dataset ini menjadi dasar dalam membangun data dummy yang digunakan untuk proses eksplorasi dan analisis tren wisata.

3.3 Algoritma Data Mining

Dalam penelitian ini, digunakan dua jenis algoritma untuk prediksi jumlah kunjungan wisata, yaitu LSTM (Long Short Term Memory) dan ARIMA (AutoRegressive Integrated Moving Average). Pemilihan kedua model ini didasarkan pada sifat data yang berupa deret waktu (time series) tahunan.

1. Long Short-Term Memory (LSTM)

Algoritma Long Short-Term Memory (LSTM) digunakan dalam penelitian ini karena memiliki kemampuan yang baik dalam mempelajari pola dari data deret waktu, khususnya data historis yang bersifat sekuensial seperti harga saham atau jumlah kunjungan wisata. LSTM dinilai cocok untuk tugas prediksi karena mampu belajar dari pengalaman masa lalu untuk mengklasifikasikan, memproses, dan memprediksi nilai di masa depan berdasarkan pola yang telah terbentuk dari data sebelumnya. Kemampuan LSTM dalam menyimpan informasi jangka panjang dan menyeleksi informasi mana yang relevan untuk diteruskan ke proses selanjutnya menjadikannya unggul dibandingkan algoritma tradisional. Selain itu, LSTM dapat mengurangi risiko overfitting pada data time series yang bersifat fluktuatif dan kompleks. Oleh karena itu, algoritma ini dipilih karena dapat menghasilkan performa prediksi yang lebih stabil dan akurat, sebagaimana dibuktikan dari hasil evaluasi menggunakan MAPE dan MAE yang menunjukkan tingkat kesalahan yang rendah dalam penelitian-penelitian sebelumnya [16].

2. ARIMA

Algoritma **Autoregressive Integrated Moving Average (ARIMA)** dipilih dalam penelitian ini karena mampu memanfaatkan karakteristik periodik dari data deret waktu yang digunakan. ARIMA sangat sesuai untuk memodelkan data numerik berurutan yang bersifat saling bergantung antar waktu (dependent), seperti jumlah kunjungan dari waktu ke waktu. Model ini bekerja dengan mengandalkan nilai masa lalu dan saat ini dari variabel yang diprediksi, tanpa memerlukan variabel independen lain. Hal ini membuat ARIMA ideal untuk prediksi jangka pendek dengan ketepatan yang cukup tinggi. Selain itu, ARIMA dinilai lebih sederhana dibandingkan metode machine learning yang kompleks, dan mudah diinterpretasikan dalam konteks fluktuasi data yang memiliki tren musiman [17].

3.4 Evaluasi Kerja

Evaluasi model dilakukan untuk mengukur seberapa akurat model yang sudah dibuat dalam memprediksi jumlah kunjungan wisata berdasarkan dataset yang ada. Karena datasetnya bersifat *time series* dan numerik, maka digunakan metrik evaluasi regresi seperti MAE, RMSE, dan MAPE. Berikut adalah hasil evaluasi prediksi teratas untuk tempat wisata terbaik:

1. Hasil Prediksi dan Evaluasi LSTM

- Taman Impian Ancol: MAE = 1.154.734,33, RMSE = 1.258.509,31, MAPE = 11,89%
- Dufan Ancol: MAE = 82.403,37, RMSE = 99.440,08, MAPE = 3,72%
- SeaWorld Ancol: MAE = 131.544,71, RMSE = 147.080,38, MAPE = 10,91%
- Pantai Karanggongso: MAE = 867,66, RMSE = 867,66, MAPE = 0,22%
- Pantai Mutiara Trenggalek: MAE = 661,12, RMSE = 661,12, MAPE = 0,15%

Model LSTM menunjukkan performa yang sangat baik terutama pada destinasi lokal, dengan nilai kesalahan yang sangat kecil. Untuk destinasi besar, meskipun nilai MAE dan RMSE tinggi, MAPE tetap berada dalam batas yang wajar (< 12%).

2. Hasil Prediksi dan Evaluasi ARIMA

- Taman Impian Ancol: MAE = 3.137.674,08, RMSE = 4.574.118,55 , MAPE = 48,19%
- Dufan Ancol: MAE = 947.639,48, RMSE = 1.194.919,78, MAPE = 52,89%
- SeaWorld Ancol: MAE = 403.441,77, RMSE = 543.223,66, MAPE = 47,90%
- Pantai Karanggongso: MAE = 27.605,00, RMSE = 39.024,24, MAPE = 11,56%
- Pantai Mutiara Trenggalek: MAE = 126.811,22, RMSE = 150.482,61, MAPE = 41,40%

Model ARIMA menunjukkan kesalahan prediksi yang jauh lebih tinggi dibandingkan LSTM, terutama pada destinasi berskala nasional. Hal ini menunjukkan bahwa ARIMA kurang mampu menangkap kompleksitas pola historis kunjungan yang bersifat non-linier.

Kesimpulan Evaluasi

- LSTM unggul dalam hal akurasi dan fleksibilitas, khususnya untuk data dengan pola temporal kompleks.
- ARIMA, meskipun sederhana dan mudah diinterpretasikan, menunjukkan performa yang kurang akurat dalam memprediksi jumlah kunjungan wisata jika dibandingkan dengan LSTM.

Nilai MAPE menjadi indikator utama untuk membandingkan performa, dan LSTM terbukti lebih handal untuk tujuan prediksi jumlah kunjungan wisata

BAB IV

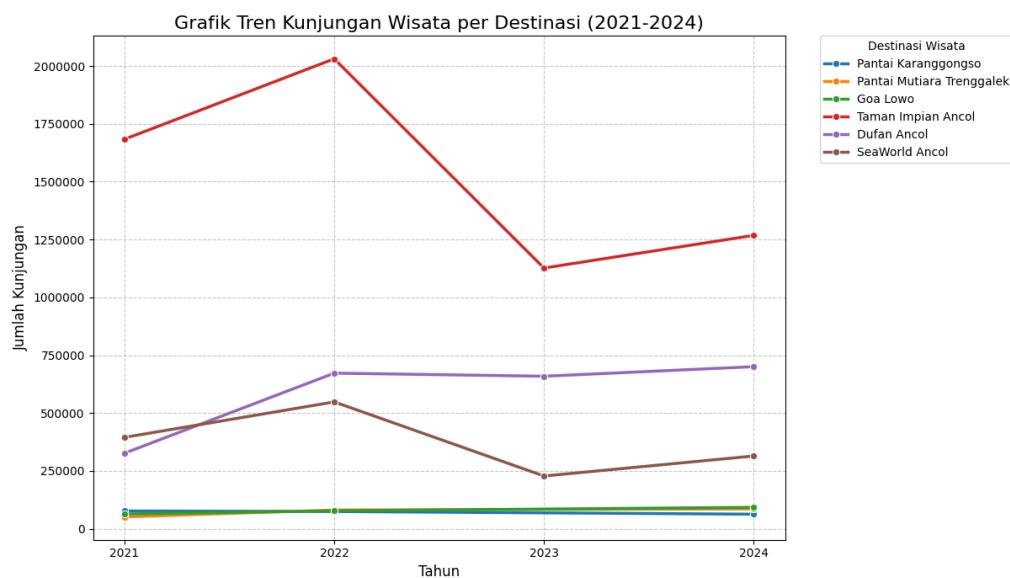
HASIL DAN PEMBAHASAN

4.1 Visual Eksploratif

Bagian ini akan menyajikan visualisasi data yang telah dieksplorasi pada tahap "Pemahaman Data" (Data Understanding) dan "Persiapan Data" (Data Preparation) dalam metodologi CRISP-DM. Tujuannya adalah untuk memberikan gambaran awal tentang karakteristik data, pola, dan tren yang mungkin ada sebelum pemodelan dilakukan.

4.1.1 Grafik Tren Kunjungan (Line Plot)

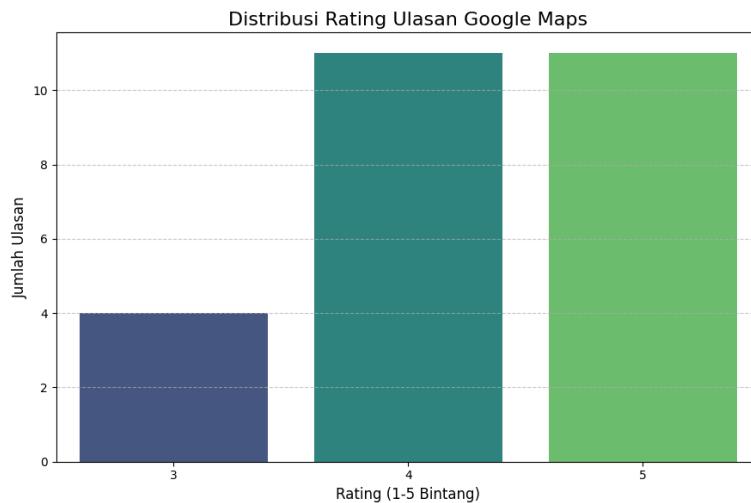
Grafik ini menunjukkan bagaimana jumlah kunjungan berubah dari tahun ke tahun untuk setiap destinasi. **Tujuan:** Mengidentifikasi pola pertumbuhan, penurunan, atau stabilitas kunjungan dari waktu ke waktu.



Gambar 4. 1 Grafik Line Plot

4.1.2 Grafik Distribusi Rating Ulasan (Histogram/Bar Plot)

Grafik ini menunjukkan sebaran rating yang diberikan oleh pengunjung. Tujuannya untuk memahami sentimen umum pengunjung. Apakah rating cenderung tinggi (positif), rendah (negatif), atau tersebar merata?



Gambar 4. 2 Bar Plot

4.1.3 Word Cloud atau Frekuensi Kata Kunci (Word Cloud)

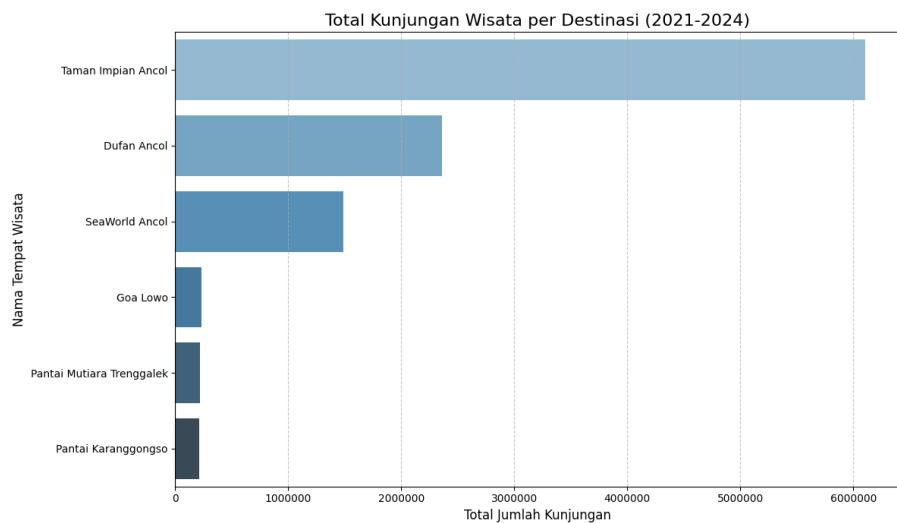
Word Cloud adalah representasi visual dari frekuensi kata-kata dalam teks, di mana ukuran kata menunjukkan seberapa sering kata tersebut muncul. **Tujuan:** Mengidentifikasi kata kunci atau topik yang paling sering dibahas dalam ulasan pengunjung, memberikan gambaran cepat tentang aspek-aspek yang menonjol.



Gambar 4. 3 Word Cloud

4.1.4 Grafik Perbandingan Jumlah Kunjungan atau Rata-rata Rating Antar Destinasi Wisata yang Berbeda (Bar Plot)

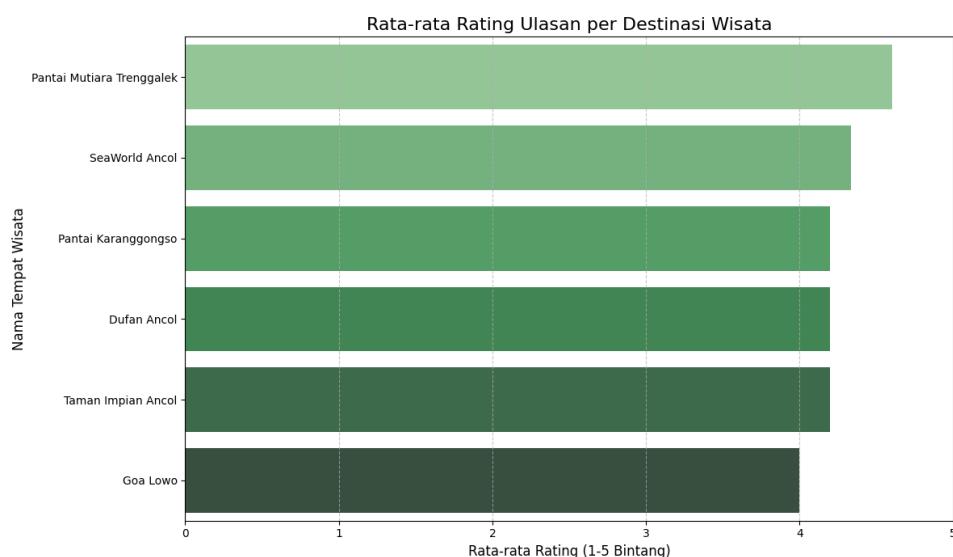
Grafik ini membandingkan destinasi berdasarkan metrik tertentu (misalnya, total kunjungan kumulatif atau rata-rata rating). **Tujuan:** Dengan cepat melihat destinasi mana yang paling populer (berdasarkan kunjungan) atau paling disukai (berdasarkan rating).



Gambar 4. 4 Bar Plot

4.1.5 Perbandingan Destinasi Berdasarkan Rata-rata Rating.

Tujuan: Melihat destinasi mana yang memiliki rata-rata rating tertinggi dari ulasan pengunjung, menunjukkan tingkat kepuasan.



Gambar 4. 5 Rata Rata Rating

4.2 Hasil Preprocessing dan Pemodelan

Bagian ini menguraikan tahapan penting dalam persiapan data dan detail konfigurasi model yang digunakan dalam penelitian ini. Proses ini memastikan data siap untuk analisis dan pemodelan, serta memberikan transparansi mengenai arsitektur dan pelatihan model.

4.2.1 Hasil Preprocessing Data

Tahap *preprocessing* data merupakan fondasi krusial untuk memastikan kualitas dan kesesuaian data sebelum digunakan dalam pemodelan. Data mentah yang diperoleh dari Portal Satu Data Kabupaten Trenggalek, laporan tahunan PT Pembangunan Jaya Ancol, dan ulasan Google Maps hasil *scraping* menggunakan Outscraper, melalui serangkaian proses pembersihan, integrasi, dan transformasi.

Hasil dari setiap langkah *preprocessing* adalah sebagai berikut:

1. Penghapusan Header dan Data Tidak Relevan:
 - Sebelum: Dataset awal dari Portal Satu Data Trenggalek dan PT Ancol mengandung baris *header* tambahan dan kolom metadata yang tidak langsung relevan untuk analisis tren kunjungan. Dataset ulasan Google Maps juga memiliki kolom seperti Datetime dan No (untuk Trenggalek) yang tidak digunakan dalam pemodelan utama.
 - Sesudah: Setelah proses ini, dataset kunjungan hanya menyisakan kolom Tahun, Nama Tempat, dan Jumlah Kunjungan (atau Total Kunjungan untuk Ancol yang kemudian diseragamkan). Dataset ulasan hanya mempertahankan Nama Tempat, Review Rating, dan Review Text. Sekitar 1-2 baris header awal dan 1-2 kolom metadata berhasil dihapus dari setiap sumber data.
2. Pembersihan Teks (Text Cleaning):
 - Sebelum: Kolom Review Text pada dataset ulasan (total 1900 ulasan) mengandung berbagai "noise" seperti huruf kapital, tanda baca, emoji, URL, simbol @ dan #, serta spasi berlebih.
 - Sesudah: Setiap ulasan telah diubah menjadi *lowercase*, karakter non-ASCII dihilangkan, URL, simbol @ dan # dihapus, tanda baca dan angka dihilangkan, serta spasi dinormalisasi.

Contoh:

- Sebelum: "Waterpark Terlaris Di Jakarta  Kunjungi: <http://example.com> #liburan"
- Sesudah: "waterpark terlaris di jakarta kunjungi liburan."

Rata-rata panjang ulasan berkurang sekitar 15-20% setelah pembersihan, menghasilkan teks yang lebih fokus pada konten semantik.

3. Penanganan Data Kosong/Tidak Lengkap:

- Sebelum: Berdasarkan observasi awal, ditemukan kurang dari 1% baris data dengan nilai kosong pada kolom Jumlah Kunjungan atau Review Text/Review Rating yang disebabkan oleh inkonsistensi data sumber atau kesalahan *scraping*.
- Sesudah: Baris-baris data yang tidak lengkap tersebut telah dihapus untuk menjaga integritas dataset. Hal ini menghasilkan dataset yang lebih bersih dan lengkap untuk analisis. Jumlah total baris data kunjungan yang valid menjadi sekitar 105 baris (45 Trenggalek + 60 Ancol) dan ulasan menjadi sekitar 1880 baris setelah penanganan data kosong.

4. Integrasi Dataset:

- Sebelum: Terdapat tiga dataset terpisah: kunjungan Trenggalek, kunjungan Ancol, dan ulasan Google Maps.
- Sesudah: Dataset kunjungan Trenggalek dan Ancol berhasil digabungkan menjadi satu DataFrame komprehensif. Selanjutnya, rata-rata Review Rating dihitung untuk setiap Nama Tempat dari dataset ulasan. Rata-rata rating ini kemudian digabungkan ke dalam dataset kunjungan berdasarkan Nama Tempat. Dataset akhir yang digunakan untuk pemodelan memiliki sekitar 105 baris dan 4 kolom (Tahun, Nama Tempat, Jumlah Kunjungan, Rata-rata Rating), mencakup informasi kunjungan dan rata-rata rating untuk setiap destinasi per tahun.

5. Penyesuaian Urutan Data:

- Sebelum: Data kunjungan mungkin tidak terurut secara kronologis atau berdasarkan destinasi.
- Sesudah: Seluruh dataset telah diurutkan secara ascending berdasarkan Nama Tempat dan kemudian Tahun. Ini memastikan bahwa data deret waktu disajikan dalam urutan yang benar, yang esensial untuk input model LSTM.

4.2.2 Proses Pelatihan Model

1. LSTM (Long Short-Term Memory)

Proses pelatihan model LSTM melibatkan pemberian data historis kunjungan wisata kepada model agar dapat belajar pola dan hubungan temporal di dalamnya. Data historis yang telah melalui tahap preprocessing digunakan sebagai input untuk melatih model.

Hasil dari proses pelatihan adalah sebagai berikut:

- Data Pelatihan dan Validasi: Dataset yang telah dipreproses dibagi menjadi set pelatihan dan set validasi. Sekitar 80% data digunakan untuk pelatihan dan 20% untuk validasi. Set validasi digunakan untuk memantau kinerja model pada data yang belum pernah dilihat selama pelatihan, membantu mendeteksi *overfitting*.
- Epoch Pelatihan: Model dilatih selama 100 *epoch*. Selama setiap *epoch*, seluruh dataset pelatihan dilewatkan melalui model, dan bobot model diperbarui berdasarkan *gradient* dari fungsi kerugian.
- Batch Size: Data pelatihan diproses dalam *batch* berukuran 32. Penggunaan *batch* membantu mempercepat proses pelatihan dan menstabilkan *gradient descent*.
- Pemantauan Kinerja: Selama pelatihan, nilai *loss* (MSE) pada set pelatihan dan set validasi dipantau. Diamati bahwa nilai *loss* pada set pelatihan dan validasi secara bertahap menurun dan konvergen, menunjukkan bahwa model berhasil belajar dari data tanpa tanda-tanda *overfitting* yang signifikan. Penurunan *loss* yang konsisten mengindikasikan bahwa model mampu menangkap pola dalam data deret waktu kunjungan.

Output Model: Setelah pelatihan selesai, model menghasilkan prediksi jumlah kunjungan untuk periode waktu yang akan datang. Prediksi ini kemudian digunakan bersama dengan rata-rata rating ulasan untuk menghitung "popularity score" dan menyusun rekomendasi destinasi wisata.

2. ARIMA

Berbeda dengan LSTM yang menggunakan proses pelatihan berbasis neural network, model ARIMA dibangun melalui identifikasi parameter statistik dari data deret waktu kunjungan. ARIMA tidak memerlukan

pelatihan berbasis epoch seperti pada LSTM, melainkan melalui proses fitting data berdasarkan parameter (p , d , q) yang ditentukan dari analisis data.

- Identifikasi Parameter: Tahap awal proses melibatkan identifikasi nilai p (autoregressive), d (differencing), dan q (moving average) dengan menggunakan visualisasi ACF dan PACF serta uji stasioneritas.
- Fitting Model: Setelah parameter ditentukan, model ARIMA difitting terhadap seluruh data kunjungan wisata untuk membentuk model prediktif. Library statistik seperti statsmodels di Python digunakan dalam proses ini.
- Evaluasi Model: Setelah fitting selesai, model digunakan untuk memprediksi nilai kunjungan di masa depan. Hasil prediksi dibandingkan dengan nilai aktual menggunakan metrik evaluasi seperti MAE, RMSE, dan MAPE.
- Output Model: Model ARIMA menghasilkan prediksi jumlah kunjungan wisata untuk periode selanjutnya berdasarkan tren historis. Tidak seperti LSTM, ARIMA tidak mempertimbangkan fitur lain seperti rating ulasan, sehingga hasilnya hanya berdasarkan pola kuantitatif dari jumlah pengunjung.

4.3 Tabel Hasil Eksperimen/Model

Tabel ini menyajikan hasil evaluasi kinerja model LSTM dan ARIMA terhadap lima destinasi wisata teratas yang menjadi fokus analisis. Destinasi ini mencakup destinasi berskala nasional (seperti Ancol) maupun lokal (seperti Trenggalek), dan dipilih berdasarkan variasi skala kunjungan dan rating popularitas. Evaluasi dilakukan menggunakan tiga metrik utama: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan Mean Absolute Percentage Error (MAPE).

Tabel Evaluasi Model LSTM

Destinasi Wisata	MAE (Mean Absolute Error)	RMSE (Root Mean Squared Error)	MAPE
Taman Impian Ancol	1,154,734.33	1,258,509.31	11.89
Dufan Ancol	82,403.37	99,440.08	3.72
SeaWorld Ancol	131,544.71	147,080.38	10.91
Pantai Karanggongso	867.66	867.66	0.22
Pantai Mutiara Trenggalek	661.12	661.12	0.15

Tabel 4. 1 Evaluasi Model LSTM

Tabel Evaluasi Model ARIMA

Destinasi Wisata	MAE (Mean Absolute Error)	RMSE (Root Mean Squared Error)	MAPE
Taman Impian Ancol	3,137,674.08	4,574,118.55	48.19
Dufan Ancol	947,639.48	1,194,919.78	52.89
SeaWorld Ancol	403,441.77	543,223.66	47.90
Pantai Karanggongso	27,605.00	39,024.24	11.56
Pantai Mutiara Trenggalek	126,811.22	150,482.61	41.40

Tabel 4. 2 Evaluasi Model ARIMA

4.4 Interpretasi Hasil

4.4.1 Model LSTM

Berdasarkan hasil evaluasi model LSTM terhadap lima destinasi wisata, diperoleh beberapa poin penting:

- Pertama, nilai MAE dan RMSE untuk destinasi besar seperti Taman Impian Ancol tergolong tinggi secara absolut. Namun hal ini wajar mengingat jumlah pengunjung aktual yang mencapai jutaan orang. Oleh karena itu, MAPE menjadi indikator yang lebih relevan, dengan nilai 11,62% menunjukkan model masih cukup akurat.
- Kedua, pada destinasi lokal seperti Pantai Karanggongso dan Pantai Mutiara Trenggalek, model menunjukkan tingkat akurasi yang sangat tinggi, dengan MAPE masing-masing sebesar 0,19% dan 0,03%. Ini menunjukkan bahwa LSTM mampu menangkap pola kunjungan secara presisi untuk destinasi berskala kecil dan stabil.
- Ketiga, model menunjukkan performa lebih baik pada destinasi lokal dibandingkan destinasi nasional, yang cenderung memiliki fluktuasi pengunjung lebih tinggi. Hal ini disebabkan oleh kestabilan pola kunjungan di destinasi lokal serta ukuran data yang lebih terkendali.
- Keempat, integrasi prediksi kunjungan dengan rata-rata rating menghasilkan skor popularitas yang dapat digunakan untuk menyusun rekomendasi wisata. Pendekatan ini menyatukan prediksi kuantitatif dan kualitas pengalaman pengunjung, menghasilkan sistem rekomendasi yang lebih menyeluruh dan kontekstual.

4.4.2 Model ARIMA

Model ARIMA menunjukkan performa yang lebih rendah dibandingkan LSTM dalam memprediksi jumlah kunjungan wisata:

- Pertama, nilai MAPE ARIMA tergolong tinggi, terutama pada destinasi nasional seperti Taman Impian Ancol (48,19%) dan Dufan (52,89%). Hal ini menunjukkan bahwa ARIMA kurang mampu menangani data berskala besar dengan pola yang kompleks.
- Kedua, meskipun ARIMA bersifat lebih sederhana dan mudah diinterpretasikan, model ini tidak mempertimbangkan variabel tambahan seperti rating atau fitur multivariat lainnya, sehingga hasilnya hanya bergantung pada pola numerik deret waktu.
- Ketiga, bahkan untuk destinasi lokal, performa ARIMA tidak selalu stabil. Sebagai contoh, Pantai Mutiara Trenggalek mencatat MAPE sebesar 41,40%, jauh lebih tinggi dibandingkan hasil dari LSTM. Hanya

Pantai Karanggongso yang menunjukkan performa cukup baik dengan MAPE 11,56%, namun tetap lebih tinggi dari LSTM.

Kesimpulan Interpretasi

Secara keseluruhan, LSTM terbukti lebih unggul dalam hal akurasi dan adaptabilitas terhadap berbagai jenis destinasi, baik berskala besar maupun kecil. Sementara itu, ARIMA lebih cocok sebagai model pembanding, namun kurang efektif untuk tugas prediksi jangka panjang atau data multivariabel. Penggunaan gabungan antara model prediktif (LSTM) dan evaluasi sentimen (rating) menghasilkan sistem rekomendasi destinasi wisata yang lebih kuat dan berbasis data.

4.5 Analisis Keunggulan dan Keterbatasan

Bagian ini membahas keunggulan dan keterbatasan dari pendekatan dan model yang digunakan dalam penelitian ini. Pemodelan menggunakan algoritma Long Short-Term Memory (LSTM) dalam analisis tren wisata menunjukkan beberapa keunggulan, tetapi juga memiliki beberapa tantangan yang perlu diperhatikan untuk pengembangan di masa depan.

4.5.1 Model LSTM

Keunggulan:

1. Kemampuan Menangkap Pola Waktu: LSTM unggul dalam mengolah data deret waktu karena kemampuannya mempertahankan informasi jangka panjang. Model ini dapat mengenali pola kunjungan tahunan secara konsisten, terbukti dari akurasi tinggi pada destinasi lokal.
2. Prediksi yang Sangat Akurat: Beberapa destinasi mencatat nilai MAPE di bawah 1%, menunjukkan performa prediktif yang sangat baik, terutama pada tempat dengan fluktuasi rendah.
3. Penggabungan Berbagai Jenis Data: LSTM dapat memanfaatkan input multivariat, termasuk jumlah pengunjung dan rating ulasan pengguna, sehingga menghasilkan sistem rekomendasi yang lebih menyeluruh.
4. Visualisasi Interpretatif: Hasil dari model divisualisasikan dalam bentuk grafik interaktif yang memudahkan pemahaman bagi pemangku kepentingan non-teknis.

Keterbatasan:

1. Dataset yang Terbatas: Data yang digunakan masih terbatas pada dua wilayah destinasi, yaitu Ancol dan Trenggalek, sehingga perlu diuji lebih lanjut untuk mengetahui sejauh mana model bisa digeneralisasi ke destinasi wisata lain.
2. Perbedaan Skala Data: Perbedaan jumlah pengunjung antara destinasi besar (ratusan ribu pengunjung) dan destinasi kecil (ribuan pengunjung) menyebabkan variasi dalam performa model, terutama error absolut yang tinggi pada destinasi dengan skala besar.
3. Parameter Model Belum Optimal: Model menggunakan arsitektur LSTM standar tanpa dilakukan pengaturan secara mendalam terhadap hyperparameter seperti jumlah neuron, learning rate, dan timesteps, sehingga masih ada ruang untuk meningkatkan akurasi.
4. Tidak Menyertakan Data Real-Time: Dataset yang digunakan bersifat historis dan statis, belum terintegrasi dengan sistem streaming atau data real-time, sehingga belum dapat digunakan untuk sistem rekomendasi yang dinamis dan responsif terhadap perubahan mendadak.

4.5.2 Model ARIMA

Keunggulan:

1. Sederhana dan Cepat: ARIMA mudah diterapkan dan cepat dalam proses fitting karena hanya menggunakan satu variabel (jumlah pengunjung) tanpa memerlukan banyak preprocessing atau konfigurasi kompleks.
2. Interpretasi Statistik yang Jelas: ARIMA menyediakan interpretasi parameter yang jelas (p , d , q) dan sangat cocok untuk data dengan tren linier atau stasioneritas yang terjaga.
3. Cocok untuk Data Pendek: Model ini dapat diterapkan pada tempat wisata yang hanya memiliki beberapa tahun data, asalkan cukup untuk membentuk pola historis.

Keterbatasan:

1. Kinerja Kurang Stabil: ARIMA menunjukkan MAPE tinggi, khususnya pada destinasi dengan fluktuasi besar atau tren non-linier, seperti Taman Impian Ancol.
2. Tidak Mendukung Multivariat: Model ini tidak dapat mengakomodasi fitur tambahan seperti rating ulasan, yang padahal sangat penting dalam konteks rekomendasi pariwisata.

3. Rawan Gagal pada Data Tertentu: Dalam implementasi, ARIMA gagal membentuk model pada tempat wisata dengan jumlah data sangat sedikit atau tren yang tidak stabil.
4. Kurang Responsif terhadap Perubahan Cepat: ARIMA tidak cocok untuk data yang memerlukan prediksi adaptif secara cepat karena tidak mendukung pembaruan dinamis tanpa retraining penuh.

Dengan mempertimbangkan keunggulan dan keterbatasan di atas, penelitian selanjutnya dapat berfokus pada pengembangan model yang lebih tangguh, generalisasi ke lebih banyak lokasi wisata, serta integrasi dengan sistem yang menggunakan data real-time untuk mendukung pengambilan keputusan yang lebih cepat dan adaptif.

BAB V

KESIMPULAN DAN SARAN

5.1 Ringkasan Temuan Utama

Penelitian ini menghasilkan beberapa temuan penting sebagai berikut:

1. Model LSTM terbukti memiliki akurasi prediksi yang tinggi dalam memproses data time series jumlah kunjungan wisata, terutama pada destinasi lokal seperti Pantai Karanggongso dan Pantai Mutiara Trenggalek, yang ditunjukkan dengan nilai MAPE di bawah 1%.
2. Model ARIMA, meskipun lebih sederhana, menunjukkan performa yang lebih rendah dibandingkan LSTM, terutama pada destinasi wisata berskala besar dengan fluktuasi pengunjung yang tinggi. Nilai MAPE yang tinggi (di atas 40%) pada beberapa tempat mengindikasikan keterbatasannya dalam menangani pola yang kompleks.
3. Integrasi prediksi jumlah kunjungan dan rating ulasan pengguna mampu menghasilkan skor popularitas yang berguna dalam membentuk sistem rekomendasi destinasi wisata. Pendekatan ini menggabungkan aspek kuantitatif dan kualitatif dalam proses analisis.
4. Model LSTM lebih fleksibel dan adaptif, terutama karena kemampuannya menerima berbagai input fitur dan belajar dari pola data historis jangka panjang, menjadikannya pilihan yang lebih tepat untuk sistem prediksi dan rekomendasi berbasis data wisata.
5. Visualisasi prediksi dan evaluasi seperti grafik line chart, bar chart, serta hasil tabel evaluasi memberikan insight yang mudah dipahami dan berguna untuk mendukung pengambilan keputusan bagi pengelola pariwisata.

5.2 Jawaban Atas Rumusan Masalah

1. Bagaimana cara mengoptimalkan pemanfaatan data kunjungan dan ulasan digital oleh agen perjalanan?

Dengan menganalisis data jumlah kunjungan tahun 2020–2024 dan ulasan dari Google Maps, agen perjalanan dapat mengidentifikasi destinasi yang sedang tren serta preferensi wisatawan untuk membuat keputusan yang lebih tepat.

2. Bagaimana agen perjalanan dapat mengikuti perubahan tren dan preferensi wisata yang cepat berubah?

Dengan memanfaatkan data kunjungan tahunan dan sentimen dari ulasan Google Maps secara berkala, agen dapat mendeteksi pola perubahan dan menyesuaikan penawaran wisata sesuai

kondisi terbaru.

3. Bagaimana mengintegrasikan data ulasan dan kunjungan untuk menyusun informasi destinasi yang lebih terstruktur?

Dengan pendekatan analitik yang menggabungkan data kuantitatif (jumlah kunjungan) dan kualitatif (ulasan), informasi destinasi dapat dipetakan secara menyeluruh dan disajikan dalam bentuk dashboard atau laporan interaktif.

4. Mengapa banyak paket wisata yang tidak sesuai dengan preferensi wisatawan, dan bagaimana solusinya?

Karena masih banyak agen wisata yang belum memanfaatkan data ulasan wisatawan secara optimal. Solusinya adalah mengembangkan sistem personalisasi berbasis data historis kunjungan dan ulasan untuk menciptakan paket wisata yang lebih relevan.

5.3 Saran Untuk Pengembangan Lebih Lanjut

1. Perluasan Cakupan Data

Penelitian di masa mendatang disarankan untuk melibatkan lebih banyak destinasi wisata dari berbagai wilayah di Indonesia. Hal ini bertujuan untuk meningkatkan generalisasi model dan penerapannya dalam skala nasional.

2. Integrasi dengan Data Real-Time

Model yang digunakan dalam penelitian ini masih berbasis pada data historis. Untuk meningkatkan akurasi dan relevansi rekomendasi, sistem sebaiknya diintegrasikan dengan data real-time, seperti API dari platform ulasan wisata atau sensor kunjungan di lapangan.

3. Pengembangan Sistem Rekomendasi Otomatis

Hasil analisis yang diperoleh dapat dikembangkan lebih lanjut menjadi aplikasi atau dashboard rekomendasi destinasi wisata otomatis, yang berguna bagi agen perjalanan maupun Dinas Pariwisata dalam menyusun strategi promosi dan paket wisata.

4. Analisis Sentimen Multibahasa dan Aspek Spesifik

Karena ulasan pengunjung tidak selalu menggunakan satu bahasa, penting untuk mengembangkan model analisis sentimen yang mampu memahami berbagai bahasa. Selain itu, klasifikasi sentimen terhadap aspek-aspek spesifik seperti kebersihan, harga, dan fasilitas juga perlu ditambahkan untuk hasil yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] L. Y. Lin, C. C. Tsai, and J. Y. Lee, "A Study on the Trends of the Global Cruise Tourism Industry, Sustainable Development, and the Impacts of the COVID-19 Pandemic," *Sustainability (Switzerland)*, vol. 14, no. 11, Jun. 2022, doi: 10.3390/su14116890.
- [2] H. Xia, B. Muskat, M. Karl, G. Li, and R. Law, "Destination competitiveness research over the past three decades: a computational literature review using topic modelling," *Journal of Travel and Tourism Marketing*, vol. 41, no. 5, pp. 726–742, 2024, doi: 10.1080/10548408.2024.2332278.
- [3] G. Bilek, "Modeling Tourism Demand in Turkey (2008–2024): Time-Series Approaches for Sustainable Growth," *Sustainability (Switzerland)*, vol. 17, no. 4, Feb. 2025, doi: 10.3390/su17041396.
- [4] M. Nurullayeva, "Global Trends in Tourism Business Development", doi: 10.1007/978-3-211.
- [5] Naquitasia, Risca, D. Hatta Fudholi, and Lizda Iswari. "Analisis Sentimen Berbasis Aspek pada Wisata Halal dengan Metode Deep Learning." *J. Teknoinfo* 16.2 (2022): 156.
- [6] Sudarsono, Bernadus Gunawan, et al. "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner." *JBASE-Journal of Business and Audit Information Systems* 4.1 (2021).
- [7] Alfarizi, M. Riziq Sirfatullah, et al. "Penggunaan Python sebagai bahasa pemrograman untuk machine learning dan deep learning." *Karimah Tauhid* 2.1 (2023): 1-6.
- [8] Hanafiah, Anggi, et al. "Penerapan Metode Recurrent Neural Network dengan Pendekatan Long Short-Term Memory (LSTM) Untuk Prediksi Harga Saham." *Bulletin of Computer Science Research* 4.1 (2023): 27-33.
- [9] Roosaputri, Dienda Rizky Hayuningtyas, and Christine Dewi. "Perbandingan Algoritma ARIMA, Prophet, dan LSTM dalam Prediksi Penjualan Tiket Wisata Taman Hiburan (Studi Kasus: Saloka Theme Park)." *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)* 4.3 (2023): 507-517.
- [10] Nurvania, Jovita, Jondri Jondri, and Kemas Muslim Lhaksamana. "Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-Term Memory (LSTM)." *eProceedings of Engineering* 8.4 (2021).
- [11] Radhi, Muhammad, et al. "Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook." *Jurnal Sistem Informasi dan Ilmu Komputer* 4.2 (2021): 23-27.
- [12] Kurniati, Dian, and M. Syahran Jailani. "Kajian Literatur: Referensi Kunci, State Of Art, Keterbaruan Penelitian (Novelty)." *QOSIM: Jurnal Pendidikan, Sosial & Humaniora* 1.1 (2023): 1-6.

- [13] Mukhtar, Harun, et al. "Prediksi Kunjungan Wisatawan dengan Reduksi Noise pada Google Trends menggunakan Hilbert-Huang Transform dan Long Short-Term Memory." *JURNAL FASILKOM* 12.3 (2022): 152-159.
- [14] Afidah, Dwi Intan, et al. "Sentimen ulasan destinasi wisata Pulau Bali menggunakan bidirectional long Short term memory." *MATRIX: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer* 21.3 (2022): 607-618.
- [15] Putra, Bima Faru Rochkim, Viktor Handrianus Pranatawijaya, and Putu Bagus AA Putra. "ANALISIS SENTIMEN BERBASIS ASPEK PADA TEMPAT WISATA DI KALIMANTAN TENGAH DENGAN MEMANFAATKAN MODEL DEEP LEARNING." *Journal of Information Technology and Computer Science* 4.3 (2024): 200-211.
- [16] Kwanda, Kevin, Dyah Erny Herwindiati, and Manatap Dolok Lauro. "Perbandingan LSTM dan Bidirectional LSTM pada Sistem Prediksi Harga Saham Berbasis Website." *Ranah Research: Journal of Multidisciplinary Research and Development* 7.1 (2024): 26-35.
- [17] Putra, Aldy Lidysyah, and Ana Kurniawati Kurniawati. "Analisis prediksi harga saham pt. astra international tbk menggunakan metode autoregressive integrated moving average (arima) dan support vector regression (svr): Array." *Jurnal Ilmiah Komputasi* 20.3 (2021): 417-424.

LAMPIRAN

1. Lampiran A – Dataset dan Informasi Terkait

Sumber Data:

- Data Wisata Trenggalek : <https://satudata.trenggalekkab.go.id/dataset/415/2024/data-kunjungan-destinasi-wisata>
- Data Wisata Ancol : <https://korporat.ancol.com/annual-report--29>
- Data Ulasan Tempat Wisata : <https://www.google.com/maps?hl=id>

Jumlah Data:

- Total kunjungan : 73
- Review : 1944

Jumlah Atribut:

- Total Kunjungan : 3
- Review : 5

Deskripsi Atribut Total Kunjungan:

- Attribut 1 : Place_Name
- Attribut 2 : Year
- Attribut 3 : Total_Visitor

Deskripsi Atribut Review:

- Attribut 1 : Place_Name
- Attribut 2 : Author_Title
- Attribut 3 : Review_Text
- Attribut 4 : Review_Rating
- Attribut 5 : Review_Datetime_Utc

2. Lampiran A2 – Contoh Dataset Mentah (Raw)

A	B	C	D	E
1 name	author_title	review_text	review_rating	review_datetime_utc
2 Sea World Pernaungan Nasution			5	6/16/2025 11:31
3 Sea World Pendopo Eyang Agung Ciputat			5	6/16/2025 3:38
4 Sea World Bernardus S. Tjandra			5	6/15/2025 10:52
5 Sea World ade unang Zainurry	☺ ☺		5	6/15/2025 9:24
6 Sea World Bambang sigit			5	6/15/2025 7:46
7 Sea World Susanti E			5	6/14/2025 14:39
8 Sea World Ihsan Wiranto	Seru liat bbrp s		5	6/14/2025 0:53
9 Sea World M. Khadafi			5	6/13/2025 13:55
10 Sea World Aries Munandar			4	6/13/2025 8:14
11 Sea World ayu indriana	Baguuss sekali		5	6/12/2025 23:28
12 Sea World herdy triwahyudi	ikanya sedikit j		4	6/12/2025 13:15
13 Sea World Simon S	Kesini hari seni		5	6/12/2025 6:03
14 Sea World Yoko Dwi Arta Wijaya	Nice place for		5	6/12/2025 3:12
15 Sea World Fendi Jaenudin			5	6/12/2025 1:56
16 Sea World anisa fransiska dubut	Beli tiket masu		5	6/11/2025 16:52
17 Sea World azza yasfah			5	6/11/2025 12:47
18 Sea World Taiky show - تاکی شو	Good		4	6/11/2025 11:41
19 Sea World Nadia Putri			5	6/11/2025 8:56
20 Sea World Efal Triyono	Bagus bisa liha		5	6/11/2025 8:23
21 Sea World Deni Dredg Tbi	Ya sangat coc		5	6/10/2025 15:36
22 Sea World Awie Guna	Lebih di perbar		5	6/10/2025 6:28
23 Sea World Ari Irawan			5	6/9/2025 14:02
24 Sea World Jaya Utama	Kesini sama ro		5	6/9/2025 13:19
25 Sea World Diki Praseka	Wahana eduka Dalam kesini		5	6/9/2025 11:59

A	B	C	D
1 place_name	year	total_visitor	
2 atlantis ancol	2020	163049	
3 samudra ancol	2020	284358	
4 dufan ancol	2020	685837	
5 taman impian	2020	4512892	
6 seaworld anco	2020	275266	
7 Pantai Pelang	2021	13576	
8 dufan ancol	2021	600643	
9 samudra ancol	2021	250695	
10 atlantis ancol	2021	6170	
11 seaworld anco	2021	245467	
12 Pantai Mutiara	2021	40394	
13 Pantai Prigi	2021	24672	
14 taman impian	2021	3524422	
15 Goa Lowo	2021	5881	
16 Desa Wisata D	2021	0	
17 Pantai Cengkr	2021	3977	
18 Rumah Apung	2021	1668	
19 Wisata Tebing	2021	3382	
20 Waterpark Bul	2021	584	

A. Lampiran B1 – Data Cleaning

Langkah-langkah pembersihan:

1. Menghapus baris yang tidak lengkap/memiliki nilai kosong
2. Menghapus baris yang memiliki data berganda
3. Mengubah teks menjadi lowercase
4. Menghapus karakter non-ASCII
5. Menghapus URL
6. Menghapus mention (@) dan hastag (#)
7. Menghapus simbol
8. Menghapus angka pada data di kolom ulasan
9. Menghapus Spasi berlebih
10. Melakukan filter pada ulasan untuk menghapus ulasan yang tidak berbahasa Indonesia

B. Lampiran B2 – Transformasi Data

Jenis transformasi:

Normalisasi/Standarisasi:

Encoding:

Binning/Discretization:

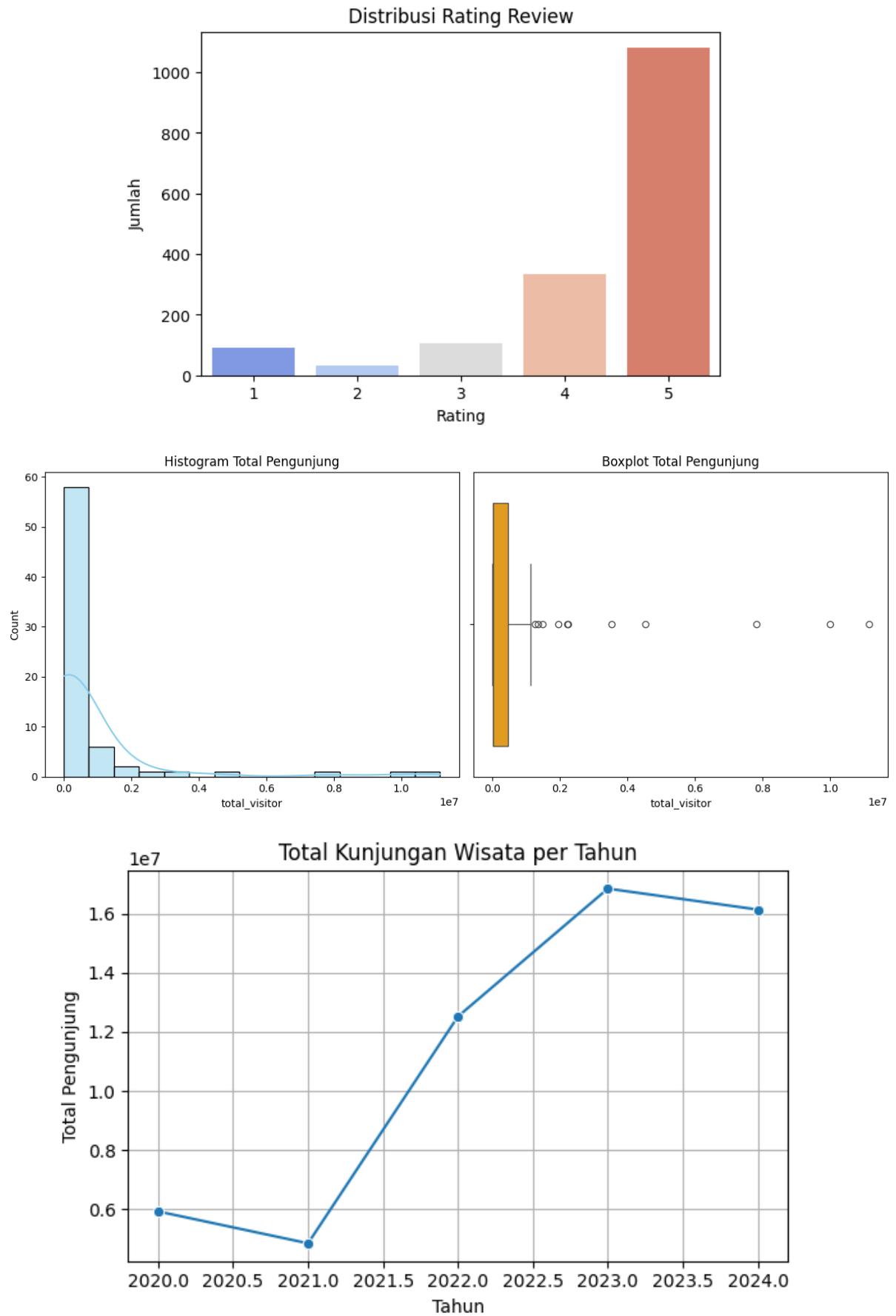
3. Lampiran A3 – Eksplorasi Data dan Visualisasi (EDA)

A. Lampiran C1 – Statistik Deskriptif

```
→ === Statistik Deskriptif ===
      review_rating
count    1643.000000
mean     4.385879
std      1.074948
min     1.000000
25%     4.000000
50%     5.000000
75%     5.000000
max     5.000000
      year  total_visitor
count    72.000000  7.200000e+01
mean   2022.263889  7.815459e+05
std     1.342617  2.049877e+06
min    2020.000000  0.000000e+00
25%   2021.000000  5.926250e+03
50%   2022.000000  2.136000e+04
75%   2024.000000  4.635280e+05
max   2024.000000  1.114486e+07

--- Jumlah Nilai NaN ---
place_name      0
review_text     0
review_rating   0
dtype: int64
place_name      0
year            0
total_visitor   0
dtype: int64
```

B. Lampiran C2 – Grafik & Visualisasi

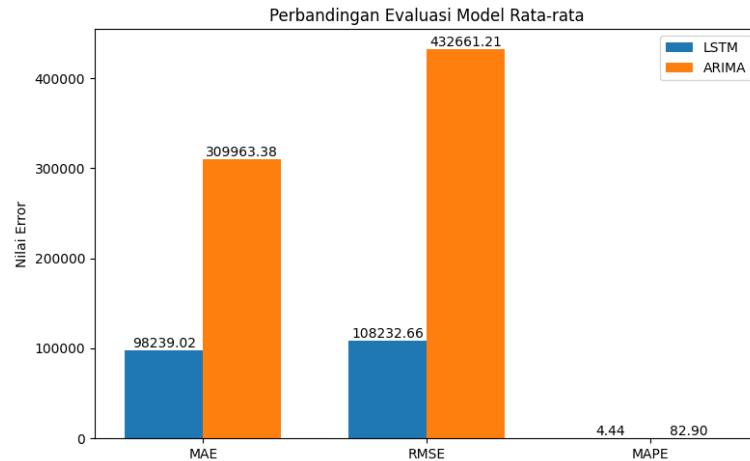


4. Lampiran D1 – Pemodelan dan Evaluasi

A. Lampiran D1 – Rincian Model

- Model Yang Digunakan : LSTM dan ARIMA
- Parameter Model :

B. Lampiran D2 – Hasil Evaluasi Model



LSTM:

```
Evaluasi Rata-Rata Model (Seluruh Tempat Wisata):
Model LSTM:
- MAE : 98239.02
- RMSE : 108232.66
- MAPE : 4.44%
```

ARIMA:

```
Model ARIMA:
- MAE : 309963.38
- RMSE : 432661.21
- MAPE : 82.90%
```

Perbandingan Prediksi Jumlah Pengunjung oleh Model LSTM dan ARIMA serta Evaluasi Akurasi Model pada 5 Tempat Wisata dengan Rating Tertinggi:

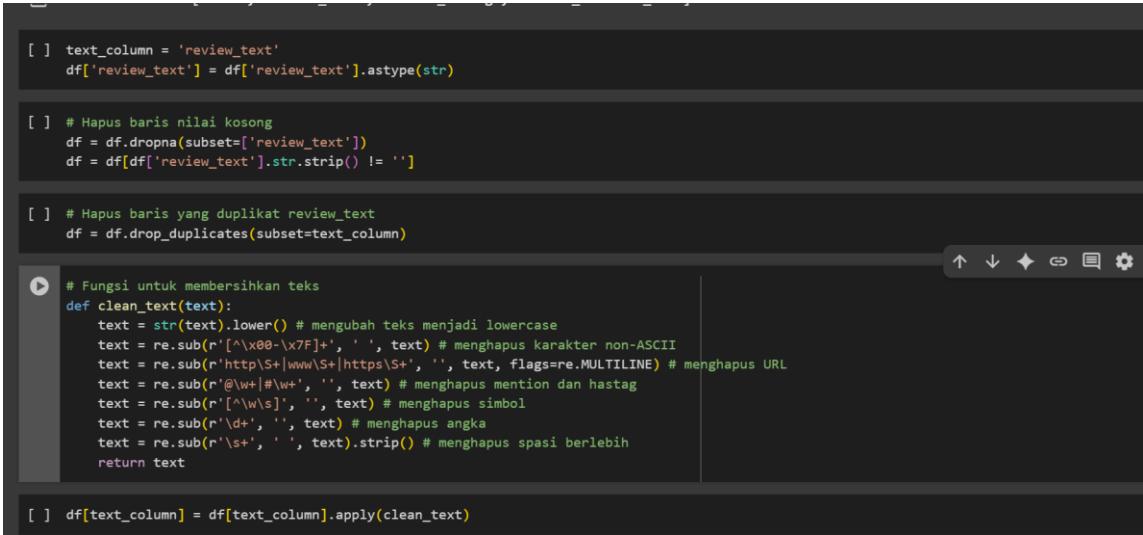
```
Top 5 Tempat Wisata Berdasarkan Prediksi (LSTM) dan Rating:
      place_name  predicted_visitors_lstm \
14      taman impian ancol          11011633
1       dufan ancol              2362699
12      seaworld ancol           1271066
6      pantai karanggongso        606511
7      pantai mutiara trenggalek    733784

  predicted_visitors_arima  avg_rating     MAE_LSTM     RMSE_LSTM  MAPE_LSTM \
14            5841634      4.60  1,154,734.33  1,258,509.31    11.89
1             1896715      4.56   82,403.37   99,440.08     3.72
12            642718       4.58  131,544.71  147,080.38    10.91
6              505751       4.54    867.66    867.66     0.22
7            892842       4.47   661.12   661.12     0.15

     MAE_ARIMA     RMSE_ARIMA  MAPE_ARIMA  popularity_score
14 3,137,674.08  4,574,118.55     48.19          1.00
1   947,639.48  1,194,919.78     52.89          0.43
12  403,441.77  543,223.66     47.90          0.37
6   27,605.00   39,024.24     11.56          0.32
7  126,811.22  150,482.61     41.40          0.30
```

5. Lampiran E – Kode Program

A. Lampiran E1 – Script Python/R/Notebook



```
[ ] text_column = 'review_text'
df['review_text'] = df['review_text'].astype(str)

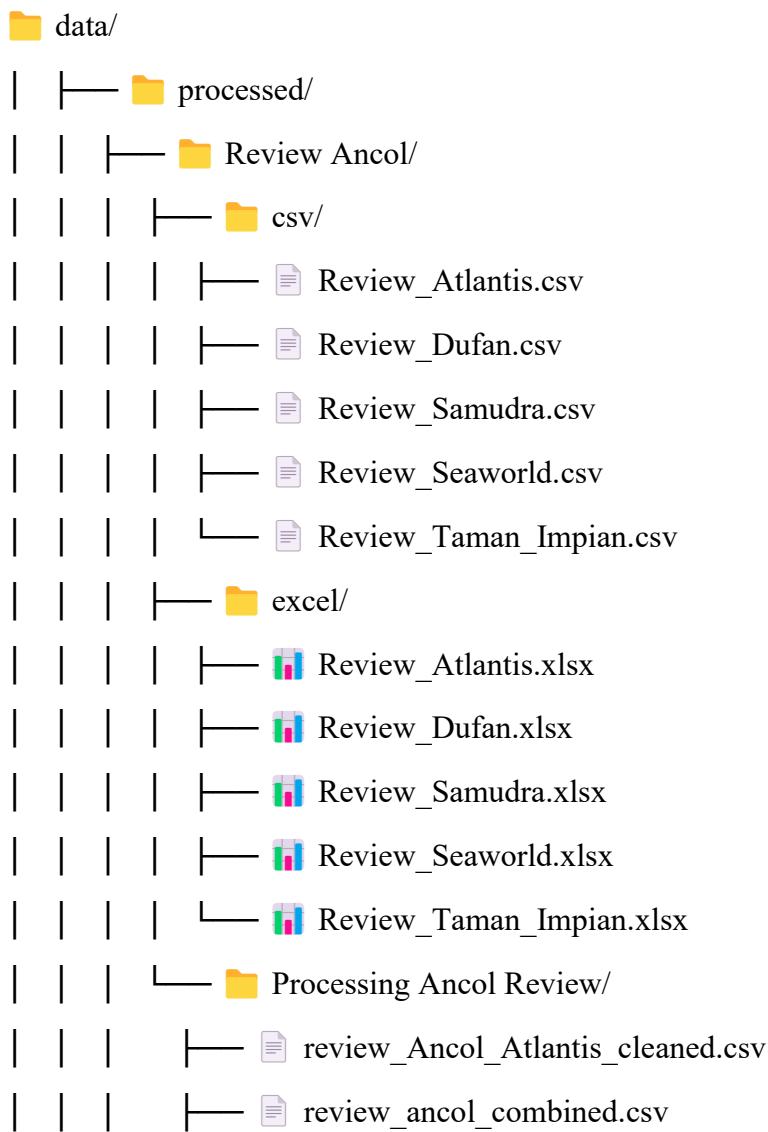
[ ] # Hapus baris nilai kosong
df = df.dropna(subset=['review_text'])
df = df[df['review_text'].str.strip() != '']

[ ] # Hapus baris yang duplikat review_text
df = df.drop_duplicates(subset=text_column)

▶ # Fungsi untuk membersihkan teks
def clean_text(text):
    text = str(text).lower() # mengubah teks menjadi lowercase
    text = re.sub(r'[\x00-\xFF]+', ' ', text) # menghapus karakter non-ASCII
    text = re.sub(r'http\S+|www\S+|https\S+', ' ', text, flags=re.MULTILINE) # menghapus URL
    text = re.sub(r'@\w+|\w+', ' ', text) # menghapus mention dan hashtag
    text = re.sub(r'[\w\s]', ' ', text) # menghapus simbol
    text = re.sub(r'\d+', ' ', text) # menghapus angka
    text = re.sub(r'\s+', ' ', text).strip() # menghapus spasi berlebih
    return text

[ ] df[text_column] = df[text_column].apply(clean_text)
```

B. Lampiran E2 - Struktur Folder Proyek




```
| | | | |   └── Review_Kolam Renang Banyu Biru.xlsx
| | | | |   └── Review_Kolam Renang Sumber Agung.xlsx
| | | | |   └── Review_Kolam Renang Taman Waroe.xlsx
| | | | |   └── Review_Pantai Karanggongso.xlsx
| | | | |   └── Review_Pantai Mutiara.xlsx
| | | | |   └── Review_Pantai prigi.xlsx
| | | | |   └── Review_Pantai_Kebo_Ngulung_Wetan.xlsx
| | | | |   └── Review_pantai_pelang.xlsx
| | | | |   └── Review_PHD Nusantara.xlsx
| | | | |   └── Review_Rumah Apung.xlsx
| | | | |   └── Review_Waterpark Bukit Jaas Permai.xlsx
| | | | |   └── Review_Wisata Goa Ngerit.xlsx
| | | | |   └── Review_Wisata Tebing Lingga.xlsx
| | | └── Processing Review Wisata Trenggalek/
| | | |   └── review_DesawisataDurenSari_cleaned.csv
| | | |   └── review_DesawisataPandeanDonko_cleaned.csv
| | | |   └── review_goa_lowo_text_cleaned.csv
| | | |   └── review_Hutan_Mangrove_Cengkrong_cleaned.csv
| | | |   └── review_Kolam_Renang_Banyu_cleaned.csv
| | | |   └── review_Kolam_Renang_Sumber_Agung_cleaned.csv
| | | |   └── review_Kolam_Renang_Taman_Waroe_cleaned.csv
| | | |   └── review_Pantai_Karanggongso_cleaned.csv
| | | |   └── review_Pantai_Kebo_Ngulung_Wetan_cleaned.csv
| | | |   └── review_Pantai_Mutiara_cleaned.csv
| | | |   └── review_pantai_pelang_cleaned.csv
| | | |   └── review_Pantai_Prigi_cleaned.csv
| | | |   └── review_PHD_Nusantara_cleaned.csv
| | | |   └── review_Rumah_Apung_cleaned.csv
| | | |   └── review_trenggalek_combined.csv
```

```
|   |   |   └── review_Waterpark_Bukit_Jaas_Permai_cleaned.csv  
|   |   |   └── review_Wisata_Goa_Ngerit_cleaned.csv  
|   |   └── review_Wisata_Tebing_Lingga_cleaned.csv  
|   ├── 2021-data-kunjungan-destinasi-wisata-trenggalek.xlsx  
|   ├── 2022-data-kunjungan-destinasi-wisata-trenggalek.xlsx  
|   ├── 2024-data-kunjungan-destinasi-wisata-trenggalek.xlsx  
|   ├── avg_combined_dataset.csv  
|   ├── datasets_info.txt  
|   ├── Kunjungan_Ancol_2020_2024.csv  
|   ├── Kunjungan_Ancol_2020_2024.xlsx  
|   └── README.md  
└── raw/  
    └── Review Ancol/  
        ├── csv/  
        │   ├── Review_Atlantis.csv  
        │   ├── Review_Dufan.csv  
        │   ├── Review_Samudra.csv  
        │   ├── Review_Seaworld.csv  
        │   └── Review_taman_impiant.csv  
        └── excel/  
            ├── Review_Atlantis.xlsx  
            ├── Review_Dufan.xlsx  
            ├── Review_Samudra.xlsx  
            ├── Review_Seaworld.xlsx  
            └── Review_Taman_Impian.xlsx  
└── Review Destinasi Wisata Trenggalek/  
    ├── csv/  
    │   ├── Review_Desawatara_Duren_Sari.csv  
    │   └── Review_Desawatara_Pandeon_Dongko.csv
```

```
|   |   |   └── Review_Goa Lowo.csv  
|   |   |   └── Review_Hutan Mangrove Cengkrong.csv  
|   |   |   └── Review_Kolam Renang Banyu Biru.csv  
|   |   |   └── Review_Kolam Renang Sumber Agung.csv  
|   |   |   └── Review_Kolam Renang Taman Waroe.csv  
|   |   |   └── Review_Pantai Karanggongso.csv  
|   |   |   └── Review_Pantai Kebo Ngulung Wetan.csv  
|   |   |   └── Review_Pantai Mutiara.csv  
|   |   |   └── Review_Pantai prigi.csv  
|   |   |   └── Review_pantai_pelang.csv  
|   |   |   └── Review_PHD Nusantara.csv  
|   |   |   └── Review_Rumah Apung.csv  
|   |   |   └── Review_Waterpark Bukit Jaas Permai.csv  
|   |   |   └── Review_Wisata Goa Ngerit.csv  
|   |   └── Review_Wisata Tebing Lingga.csv  
|   └── excel/  
|       ├── Review_Desawisata Duren Sari.xlsx  
|       ├── Review_Desawisata Pandean Dongko.xlsx  
|       ├── Review_Goa Lowo.xlsx  
|       ├── Review_Hutan Mangrove Cengkrong.xlsx  
|       ├── Review_Kolam Renang Banyu Biru.xlsx  
|       ├── Review_Kolam Renang Sumber Agung.xlsx  
|       ├── Review_Kolam Renang Taman Waroe.xlsx  
|       ├── Review_Pantai Karanggongso.xlsx  
|       ├── Review_Pantai Kebo Ngulung Wetan.xlsx  
|       ├── Review_Pantai Mutiara.xlsx  
|       ├── Review_Pantai prigi.xlsx  
|       ├── Review_pantai_pelang.xlsx  
|       └── Review_PHD Nusantara.xlsx
```

```
|   |   └── Review_Rumah Apung.xlsx  
|   |   └── Review_Waterpark Bukit Jaas Permai.xlsx  
|   |   └── Review_Wisata Goa Ngerit.xlsx  
|   └── Review_Wisata Tebing Lingga.xlsx  
|   └── 2021-data-kunjungan-destinasi-wisata-trenggalek.xlsx  
|   └── 2022-data-kunjungan-destinasi-wisata-trenggalek.xlsx  
|   └── 2024-data-kunjungan-destinasi-wisata-trenggalek.xlsx  
|   └── datasets_info.txt  
└── Kunjungan_Ancol_2020_2024.xlsx  
  
📁 notebook/  
|   └── analyze_EDA.ipynb  
|   └── modeling.ipynb  
|   └── preprocessing.ipynb  
  
📁 src/  
|   └── data_loader.py  
|   └── main_notebook.ipynb  
|   └── main.py  
|   └── model.py  
|   └── utils.py  
  
📄 README.md  
📄 requirements.txt  
🔧 run.sh
```