

Prediksi Pola Kedatangan Turis Mancanegara dan Menganalisis Ulasan Tripadvisor dengan LSTM dan LDA

FAHIRA, CAHYO PRIANTO

Universitas Logistik Bisnis Internasional, Indonesia
Email: 1204044@std.ulbi.ac.id

ABSTRAK

Penelitian ini menggunakan model LSTM dan LDA untuk meramalkan pola kedatangan wisatawan mancanegara dan menganalisis ulasan Tripadvisor di Jakarta. LSTM memproyeksikan pola kedatangan berdasarkan data historis dan menunjukkan peningkatan jumlah wisatawan dalam periode satu tahun mendatang, sedangkan LDA mengidentifikasi topik utama dalam ulasan dengan tujuan rekomendasi spesifik untuk kota Jakarta. Rekomendasi penelitian meliputi peningkatan pelayanan, kebersihan, infrastruktur, promosi tempat wisata alternatif, dan komunikasi yang jelas kepada wisatawan. Evaluasi menunjukkan performa yang baik, dengan MAPE 2,69% dalam memprediksi kedatangan wisatawan. Penelitian ini menjadi dasar untuk pengambilan keputusan dan perencanaan industri pariwisata Jakarta. Secara keseluruhan, penelitian ini memberikan wawasan berharga untuk pengembangan industri pariwisata Jakarta dengan prediksi akurat dan analisis ulasan wisatawan.

Kata kunci: LSTM, LDA, Wisatawan, Ulasan Tripadvisor, Industri pariwisata

ABSTRACT

This study uses the LSTM and LDA models to predict arrival patterns of foreign tourists and analyze Tripadvisor reviews in Jakarta. The LSTM projects arrival patterns based on historical data and shows an increase in the number of tourists in the coming one year period, while the LDA identifies the main topics in the review with specific recommendation objectives for the city of Jakarta. Research recommendations include service improvement, cleanliness, infrastructure, promotion of alternative tourist attractions, and clear communication to tourists. Evaluation shows good performance, with MAPE 2.69% in predicting tourist arrivals. This research forms the basis for decision making and planning for the Jakarta tourism industry. Overall, this research provides valuable insights for the development of Jakarta's tourism industry with accurate predictions and analysis of tourist reviews.

Keywords: LSTM, LDA, Tourist, Tripadvisor reviews, Tourism industry

1. PENDAHULUAN

Pariwisata merupakan sektor ekonomi yang penting bagi banyak negara di dunia. Keberhasilan suatu destinasi pariwisata bergantung pada jumlah kedatangan wisatawan yang tinggi serta pengalaman positif yang diberikan kepada mereka (**Yakup, 2019**). Oleh karena itu, penting untuk memiliki pemahaman yang mendalam tentang pola kedatangan turis dan ulasan yang diberikan oleh wisatawan terkait destinasi pariwisata (**Simanjuntak et al., 2017**). Dalam era digital saat ini, ulasan Tripadvisor telah menjadi sumber informasi yang sangat berharga bagi para calon wisatawan. Informasi ini mencakup pengalaman wisatawan, rekomendasi, serta penilaian terhadap aspek-aspek tertentu dalam suatu destinasi pariwisata. Namun, jumlah ulasan yang sangat besar dan beragam mempersulit analisis manual terhadap ulasan tersebut. Oleh karena itu, diperlukan pendekatan komputasional yang dapat mengidentifikasi topik utama dalam ulasan serta memprediksi pola kedatangan turis secara efisien (**Asnawi, 2023**).

Penelitian ini mengusulkan penggunaan dua metode yaitu *Long Short-Term Memory* (LSTM) dan *Latent Dirichlet Allocation* (LDA) untuk menganalisis pola kedatangan turis dan ulasan Tripadvisor terkait destinasi pariwisata (**Puteri, 2023**). LSTM adalah jenis arsitektur jaringan saraf rekurensi yang mampu mengolah data dengan urutan dan mempelajari pola jangka panjang dari data historis (**Bogaerts et al., 2020**). Sedangkan LDA adalah metode analisis topik yang dapat mengidentifikasi topik utama yang muncul dalam koleksi dokumen (**Farkhod et al., 2021**). Model LSTM digunakan untuk memprediksi pola kedatangan turis berdasarkan data historis jumlah kedatangan turis sebelumnya. Dengan memanfaatkan pola yang telah dipelajari, model ini dapat memberikan prediksi yang akurat tentang jumlah turis yang akan datang di masa depan (**Adil et al., 2021**). Sementara itu, LDA digunakan untuk mengidentifikasi topik utama yang sering dibicarakan dalam ulasan Tripadvisor. Dengan mengetahui topik yang paling relevan dalam ulasan, pihak berkepentingan dapat memahami kekuatan dan kelemahan suatu destinasi pariwisata, serta mengambil langkah-langkah yang tepat untuk meningkatkan pengalaman wisatawan (**Putri & Kusumaningrum, 2017**).

Hasil penelitian ini diharapkan dapat memberikan wawasan berharga bagi industri pariwisata dalam meningkatkan pengalaman wisatawan dan mengoptimalkan potensi destinasi pariwisata. Dengan memahami pola kedatangan turis dan topik yang sering dibicarakan dalam ulasan, pengambilan keputusan dan perencanaan di industri pariwisata dapat dilakukan secara lebih efektif dan efisien.

2. METODE PENELITIAN

2.1 CRISP-DM

CRISP-DM adalah suatu metode yang umum digunakan dalam proses data mining yang melibatkan serangkaian langkah terstruktur, mulai dari pemahaman bisnis hingga evaluasi hasil. Berikut CRISP-DM dapat diterapkan dalam penelitian ini:

- a. *Business Understanding*
Tahap ini mencakup pemahaman terhadap tujuan penelitian dan kebutuhan industri pariwisata. Tujuannya adalah untuk memahami aspek bisnis yang terkait dengan penelitian dan mengidentifikasi tujuan yang ingin dicapai (**Martínez-Plumed et al., 2019**). Penelitian ini bertujuan untuk memprediksi pola kedatangan turis mancanegara dan menganalisis ulasan Tripadvisor untuk meningkatkan pengalaman wisatawan dan mengoptimalkan potensi destinasi pariwisata .
- b. *Data Understanding*

Tahapan ini memberikan informasi tentang data yang digunakan dalam penelitian. Hal ini meliputi mempelajari sumber data, mengeksplorasi struktur data, dan memahami karakteristik serta kualitas data yang ada. Memahami data menjadi penting untuk memastikan data yang digunakan dalam penelitian sesuai dengan tujuan yang ingin dicapai **(Feblian & Daihani, 2016)**.

Terdapat 2 sumber data yang digunakan dalam penelitian ini. Pertama, data diperoleh dari Badan Pusat Statistik (BPS) dengan kategori pariwisata. Data ini memiliki judul "Kunjungan Wisatawan Mancanegara per bulan Menurut Kebangsaan (Kunjungan), 2023" dan rentang data yang digunakan mencakup periode tahun 2017 hingga 2023. Sumber data kedua berasal dari TripAdvisor melalui proses scraping, yaitu pengambilan ulasan wisatawan dari situs web TripAdvisor.

c. *Data Preparation*

Tahapan ini mencakup persiapan dan pemrosesan data sebelum melakukan analisis. Langkah-langkah dalam tahap ini meliputi pembersihan data, transformasi data, integrasi data, dan pemilihan atribut yang relevan. Persiapan data yang baik sangat penting untuk memastikan data siap digunakan dalam proses analisis yang akan dilakukan **(Urva et al., 2023)**.

d. *Modeling*

Pada tahapan ini dilakukan pengembangan model prediktif atau model analisis yang sesuai dengan tujuan penelitian. Pada tahap ini, teknik-teknik pemodelan seperti algoritma *machine learning*, statistik, atau teknik lainnya digunakan untuk membangun model yang dapat menganalisis atau memprediksi data yang ada **(Wiemer, Drowatzky, & Ihlenfeldt, 2019)**.

e. *Evaluation*

Tahapan ini dilakukan untuk mengevaluasi kinerja model LSTM dalam memprediksi pola kedatangan turis serta hasil analisis topik menggunakan model LDA dalam mengidentifikasi topik utama dalam ulasan Tripadvisor **(Clark, 2018)**.

f. *Deployment / Kesimpulan*

Tahap ini melibatkan penerapan hasil penelitian untuk pengambilan keputusan dan perencanaan di industri pariwisata. Hasil penelitian ini dapat digunakan sebagai dasar untuk pengambilan keputusan dan perencanaan yang lebih baik dalam industri pariwisata **(Septiani & Budi, 2022)**.

2.2 Data Scraping Selenium

Data *Scraping* menggunakan *Python Selenium* merupakan sebuah metode yang digunakan untuk mengambil ulasan turis dari website TripAdvisor secara otomatis. TripAdvisor adalah salah satu sumber informasi yang sangat berharga bagi para calon wisatawan yang mencari ulasan dan rekomendasi tentang destinasi pariwisata. Namun, mengumpulkan ulasan secara manual dari website TripAdvisor yang memiliki jumlah ulasan yang sangat besar dapat menjadi tugas yang melelahkan dan memakan waktu. Dengan menggunakan *scraping*, peneliti dapat mengambil ulasan turis secara otomatis dari website TripAdvisor **(Han, S., & Anderson, C. K., 2021)**.

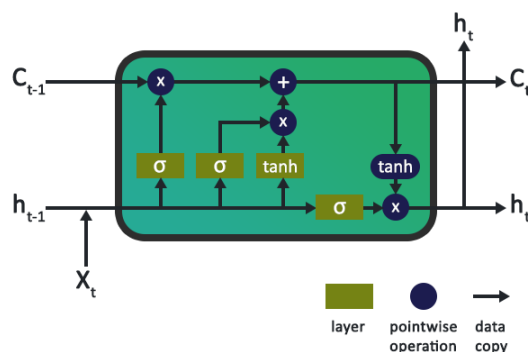
2.3 Interpolasi Linear

Metode interpolasi linear adalah metode yang memanfaatkan hubungan linier antara titik data yang diketahui untuk mengisi nilai-nilai yang hilang atau mengubah interval waktu data. Dalam penelitian ini, data kunjungan wisatawan per bulan dari tahun 2017 hingga 2022 digunakan sebagai titik data yang diketahui. Dalam proses interpolasi, setiap bulan dipecah menjadi jumlah hari yang sesuai dengan bulan tersebut, dan nilai per hari diisi berdasarkan interpolasi linear antara nilai kunjungan wisatawan per bulan yang diketahui **(Andiojaya & Demirhan, 2019)**. Hasil dari proses interpolasi linear adalah data per hari yang mencerminkan tren dan

fluktuasi data asli dengan lebih rinci. Dengan menggunakan data per hari yang dihasilkan, analisis yang lebih detail tentang pola kunjungan wisatawan harian dapat dilakukan (Jermsittiparsert & Chankoson, 2019).

2.4 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) adalah jenis arsitektur jaringan saraf rekurensi yang digunakan untuk memprediksi pola kedatangan turis berdasarkan data historis. Dalam penelitian ini, model LSTM digunakan untuk mengidentifikasi pola dan tren dari jumlah kedatangan turis mancanegara pada periode sebelumnya (Yang et al., 2020).



Gambar 1. Arsitektur LSTM

Formula umum yang digunakan dalam LSTM adalah sebagai berikut:

$$\begin{aligned}
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
 c_t &= f_t \odot c_{t-1} + i_t \odot g(W_c \cdot [h_{t-1}, x_t] + b_c) \\
 h_t &= o_t \odot \tanh(c_t)
 \end{aligned}$$

Keterangan:

- f_t adalah *gate "forget"*
- i_t adalah *gate "input"*
- o_t adalah *gate "output"*
- c_t adalah *cell state*
- h_t adalah *output*

Dalam penelitian ini, model LSTM dilatih menggunakan data historis jumlah kedatangan turis. Dengan memanfaatkan pola-pola yang telah dipelajari, model LSTM dapat memberikan prediksi akurat tentang jumlah turis yang kemungkinan akan datang di masa depan. Prediksi ini dapat digunakan oleh industri pariwisata dalam perencanaan dan pengambilan keputusan terkait pengelolaan destinasi pariwisata (Wu et al., 2021).

2.5 Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) adalah sebuah model probabilistik yang digunakan untuk mengidentifikasi topik utama dalam koleksi dokumen. Tujuan dari LDA adalah untuk mengklasifikasikan setiap dokumen ke dalam beberapa topik berdasarkan distribusi probabilitas kata-kata yang muncul dalam dokumen tersebut (Jelodar et al., 2019).

Rumus LDA :

Inisialisasi variabel:

D: Jumlah dokumen dalam koleksi.

K: Jumlah topik yang diinginkan.

V: Jumlah kata yang ada dalam kamus.

w: Matriks kata dengan ukuran $D \times N$, yang berisi kata-kata dalam setiap dokumen.

Langkah-langkah dalam algoritma LDA:

a. Inisialisasi:

Inisialisasi distribusi probabilitas topik dalam dokumen (θ) secara acak dan Inisialisasi distribusi probabilitas kata dalam topik (ϕ) secara acak.

b. Iterasi

- Untuk setiap dokumen d : Hitung probabilitas topik dalam dokumen (θ).
- Untuk setiap kata w dalam dokumen d : Hitung probabilitas topik dalam kata (ϕ).
- Gunakan probabilitas topik dalam dokumen dan probabilitas topik dalam kata untuk mengestimasi probabilitas topik dalam kata-kata di dokumen tersebut.
- Gunakan hasil iterasi untuk memperbarui distribusi probabilitas topik dalam dokumen (θ) dan distribusi probabilitas kata dalam topik (ϕ).

c. Lakukan iterasi hingga konvergensi atau mencapai jumlah iterasi maksimum.

Penelitian ini melibatkan pengumpulan dan pra-pemrosesan data dokumen, pembentukan kamus kata yang unik, pembentukan representasi vektor dokumen, inisialisasi parameter LDA, estimasi probabilitas menggunakan metode iteratif, evaluasi model menggunakan metrik seperti *perplexity* dan *coherence*, interpretasi hasil yang melibatkan identifikasi kata-kata kunci, dan penggunaan model LDA dalam pengklasifikasian dokumen dan pengambilan keputusan di industri pariwisata.

2.6 Model Evaluasi

Setelah melakukan pemodelan, evaluasi model diperlukan untuk menentukan sejauh mana model tersebut cocok dibandingkan dengan kondisi aktual atau referensi. Evaluasi model dapat dilakukan dengan beberapa metode, termasuk menghitung nilai *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), dan *Mean Absolute Error* (MAE) (AI Muluk, 2021).

a. MAPE

Mean Absolute Percentage Error (MAPE) merupakan metrik evaluasi yang umum digunakan untuk mengukur tingkat kesalahan pada model prediksi, khususnya dalam kasus *time series* seperti LSTM (*Long Short-Term Memory*). MAPE menggambarkan kesalahan rata-rata dalam bentuk persentase antara nilai aktual dan nilai prediksi. Dengan menggunakan rumus MAPE, setiap nilai prediksi dibandingkan dengan nilai aktual yang sebenarnya (Meliana, 2021). Rumus MAPE adalah:

$$MAPE = (1/n) * \sum(|(Actual - Forecast)/Actual|) * 100\%$$

Keterangan:

n adalah jumlah observasi dalam dataset

Actual adalah nilai aktual

Forecast adalah nilai prediksi

Tabel 1. Range MAPE

Range MAPE	Arti Nilai
<10%	Kemampuan model peramalan sangat baik
10 - 20%	Kemampuan model peramalan baik
20 - 50%	kemampuan model peramalan layak
>50%	kemampuan model peramalan buruk

b. Interpretasi *Topic Clustering*

Interpretasi *topic clustering* sangat penting dalam proses evaluasi hasil clustering yang dilakukan. Dalam konteks evaluasi, interpretasi *topic clustering* membantu dalam mengevaluasi kualitas dan interpretabilitas kelompok topik yang terbentuk. Dengan melakukan interpretasi *topic clustering*, kita dapat menganalisis apakah kelompok topik yang terbentuk secara intuitif dan konsisten dengan apa yang diharapkan. Kita dapat memeriksa apakah kata-kata kunci yang muncul dalam setiap kelompok topik benar-benar mencerminkan tema atau konsep yang relevan. Jika kelompok topik tidak menghasilkan interpretasi yang jelas dan bermakna, maka dapat menunjukkan bahwa algoritma clustering atau parameter yang digunakan mungkin perlu disesuaikan (Quatrini et al., 2022).

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Dalam tahap *Business Understanding*, tujuan penelitian dan kebutuhan industri pariwisata dipahami. Penelitian ini bertujuan untuk memprediksi pola kedatangan turis mancanegara dan menganalisis ulasan Tripadvisor untuk meningkatkan pengalaman wisatawan dan mengoptimalkan potensi destinasi pariwisata.

3.2 Data Understanding

Pada tahap *Data Understanding*, data yang digunakan dalam penelitian dijelaskan. Terdapat dua sumber data yang digunakan: data kunjungan wisatawan dari Badan Pusat Statistik (BPS) dan data ulasan wisatawan dari Tripadvisor yang diambil menggunakan teknik scraping data dengan *library selenium python*. Data kunjungan wisatawan dari BPS mencakup informasi jumlah kunjungan wisatawan dari Indonesia per bulan dari tahun 2017 hingga 2023 dimana kolom Indonesia sebagai jumlah kunjungan wisatawan ke Indonesia. Data ulasan wisatawan dari Tripadvisor mencakup ulasan wisatawan mengenai destinasi pariwisata di Jakarta, Bali, dan Lombok. Pada penelitian akan dianalisa terutama pada wisata di kota Jakarta.

Tabel 2. Contoh data mentah kunjungan wisatawan 2017

tahun	bulan	Indonesia
Januari	2017	22339
Februari	2017	16982

Tabel 3. Contoh data mentah kunjungan wisatawan 2023

tahun	bulan	Indonesia
Januari	2017	40141
Februari	2017	27226

Tabel 4. Contoh data mentah ulasan wisatawan TripAdvisor

<i>review</i>	<i>rating</i>
<i>a big square full of people having picnic taking photos. some stalls at the entrance. we did not go up as we were feeling the night before and it was a Sunday and super busy. we will add it to the list to go again when we come again.</i>	5

3.3 Data Preparation

Pada tahap *Data Preparation*, langkah-langkah yang dilakukan untuk mempersiapkan data sebelum analisis dilakukan adalah sebagai berikut:

- a. **Pembersihan Data:** Data yang diperoleh dari sumber seperti BPS dan TripAdvisor sering kali mengandung *noise*, duplikasi, atau informasi yang tidak relevan. Oleh karena itu, langkah pertama dalam *Data Preparation* adalah melakukan pembersihan data. Ini melibatkan penghapusan data yang tidak lengkap atau tidak valid, penanganan nilai yang hilang, dan penghapusan duplikasi data.
- b. **Transformasi Data:** Pada tahap ini, data yang telah dibersihkan dan valid perlu diubah atau disesuaikan agar sesuai dengan kebutuhan analisis selanjutnya. Dalam penelitian ini, data kunjungan wisatawan dari BPS telah mengalami transformasi melalui interpolasi linear. Interpolasi linear digunakan untuk mengubah interval waktu data dari per bulan menjadi per hari. Dengan menggunakan interpolasi linear, nilai kunjungan wisatawan per hari diisi berdasarkan hubungan linier antara nilai kunjungan wisatawan per bulan yang diketahui.
- c. **Integrasi Data:** Jika penelitian menggunakan lebih dari satu sumber data, langkah selanjutnya adalah mengintegrasikan data dari berbagai sumber tersebut. Dalam penelitian ini, data kunjungan wisatawan dari BPS dan data ulasan wisatawan dari TripAdvisor merupakan dua sumber data yang digunakan.
- d. **Pemilihan Atribut yang Relevan:** Dalam proses *Data Preparation*, pemilihan atribut yang relevan juga menjadi langkah penting. Pemilihan atribut yang tepat dapat mempengaruhi kualitas analisis yang dilakukan. Dalam penelitian ini, atribut yang relevan dari data kunjungan wisatawan dari BPS adalah jumlah kunjungan wisatawan per hari. Sedangkan dari data ulasan wisatawan dari TripAdvisor, atribut yang relevan adalah ulasan wisatawan mengenai destinasi pariwisata di Jakarta, Bali, dan Lombok.

Setelah langkah-langkah *Data Preparation* selesai dilakukan, data telah siap untuk digunakan dalam tahap analisis berikutnya. Persiapan data yang baik menjadi penting untuk memastikan kualitas dan kebersihan data serta relevansi dengan tujuan penelitian.

Tabel 5. Contoh data bersih kunjungan wisatawan

tahun	bulan	Indonesia
Januari	2017	22339
Februari	2018	49318
Juni	2019	75120
Agustus	2020	571
September	2021	571
Oktober	2022	27316
Maret	2023	31521

Pada Tabel 5 di atas, keseluruhan data berjumlah 76 baris dan 3 kolom

Tabel 6. Contoh data bersih ulasan wisatawan TripAdvisor

<i>review</i>	<i>rating</i>
<i>big square full people having picnic taking photos. stalls entrance. not go up feeling night before Sunday super busy. will add list go again when come again.</i>	5
<i>lot of good national information presented inside MONAS take time read chronological order. there scheduled short moment where protected door opened to display independent declaration.</i>	4

Pada Tabel 6 di atas, untuk data ulasan TripAdvisor setelah data dibersihkan menggunakan library NLTK, Lemmatization serta penghapusan kata-kata yang tidak memiliki makna secara topik. Total data ada 3 file csv mengenai destinasi pariwisata di Jakarta, Bali, dan Lombok yang secara keseluruhan berjumlah 3060 baris dan 2 kolom.

3.4 Modeling

Pada tahap Modeling, dua model utama digunakan dalam penelitian ini, yaitu model LSTM (*Long Short-Term Memory*) dan model LDA (*Latent Dirichlet Allocation*). Kedua model ini bertujuan untuk menganalisis dan memahami data kunjungan wisatawan serta ulasan wisatawan terkait destinasi pariwisata.

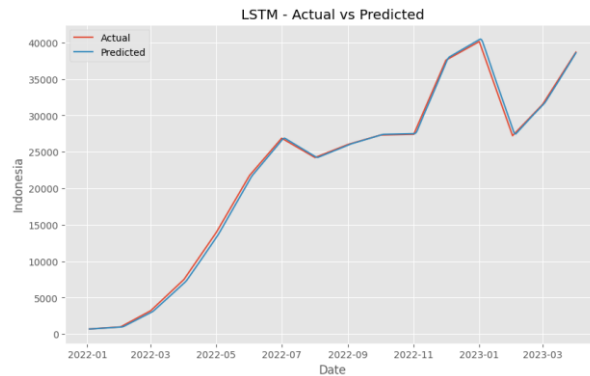
3.4.1 Model LSTM (*Long Short-Term Memory*)

Model LSTM digunakan untuk memprediksi pola kedatangan turis berdasarkan data historis kunjungan wisatawan dari BPS. Dalam model LSTM, data kunjungan wisatawan telah melalui proses interpolasi linear untuk mengubah interval waktu data menjadi per hari. Berikut adalah langkah-langkah dalam penerapan model LSTM:

- a. **Membangun model LSTM:** Menggunakan `Sequential()` untuk membuat model sequential, Menambahkan lapisan LSTM dengan jumlah unit yang ditentukan (misalnya, 50) dan fungsi aktivasi ReLU, Menambahkan lapisan Dense dengan satu neuron dan Mengompilasi model dengan optimizer Adam dan loss function `mean_squared_error`.
- b. **Definisikan *Early Stopping callback*:** Menggunakan `EarlyStopping()` untuk menghentikan pelatihan jika terjadi overfitting atau tidak ada peningkatan yang signifikan dalam loss function pada tahap validasi.
- c. **Melatih model dengan menggunakan callback *Early Stopping*:** Menggunakan fungsi `fit()` untuk melatih model pada data latihan dengan jumlah epoch, batch size, dan parameter lain yang ditentukan.

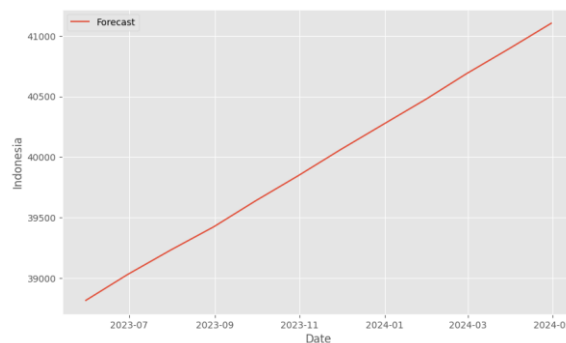
Dengan langkah-langkah ini, model LSTM akan dilatih menggunakan data time series untuk memprediksi pola kedatangan turis. *Callback early stopping* akan memantau performa model pada data validasi dan menghentikan pelatihan jika diperlukan. Data hasil interpolasi bulanan ke harian tahun 2017 - 2021 digunakan sebagai data training dan sisanya 2021 - 2023 bulan April digunakan sebagai data testing untuk perbandingan *score MAPE*. Hasil *time series* yang dihasilkan sebagai berikut :

Memprediksi Pola Kedatangan Turis Mancanegara dan Menganalisis Ulasan Tripadvisor dengan LSTM dan LDA



Gambar 2. Perbandingan nilai *actual* dan *predicted*

Dari Gambar 2, dapat dilihat bahwa model LSTM yang dihasilkan dapat mengikuti pola dari nilai aktual atau nilai sebenarnya. Dengan ini maka model memiliki nilai *error* yang begitu kecil dan bisa untuk memprediksi masa depan di tahun-tahun berikutnya.



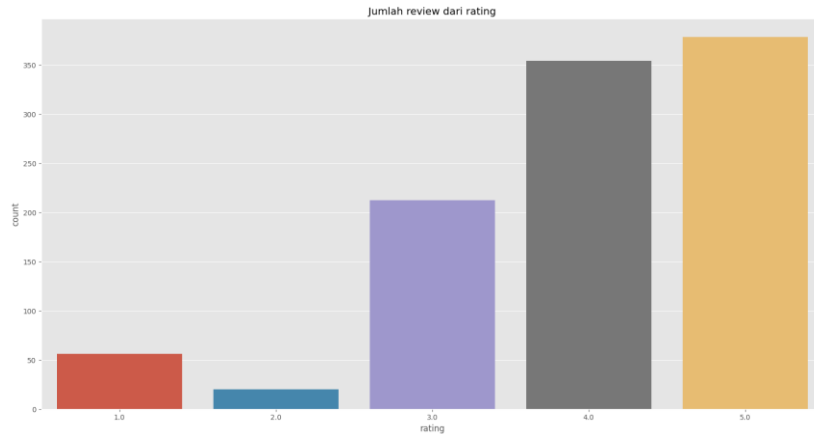
Gambar 3. Prediksi 2023 bulan juni - 2024 bulan Mei

Berdasarkan Gambar 3, model LSTM memberikan prediksi jumlah kunjungan wisatawan mancanegara ke Indonesia dari bulan Juni 2023 hingga Mei 2024. Dari hasil prediksi tersebut, terlihat bahwa terjadi kecenderungan peningkatan kunjungan wisatawan mancanegara selama periode tersebut. Hal ini menunjukkan adanya potensi peningkatan kunjungan wisatawan mancanegara ke Indonesia dalam waktu yang akan datang. Pola kenaikan yang terlihat pada hasil prediksi memberikan indikasi bahwa industri pariwisata Indonesia dapat mengalami pertumbuhan yang signifikan dalam periode tersebut. Kenaikan kunjungan wisatawan mancanegara dapat memberikan dampak positif bagi sektor pariwisata, seperti peningkatan pendapatan dari sektor perhotelan, restoran, objek wisata, dan sektor terkait lainnya.

3.4.2 Model LDA (*Latent Dirichlet Allocation*)

model LDA dilakukan untuk menganalisis terhadap ulasan wisatawan yang berasal dari TripAdvisor untuk mendapatkan wawasan tentang pengalaman wisatawan dan mengidentifikasi topik-topik utama yang muncul dalam ulasan tersebut. Langkah-langkah yang dilakukan meliputi:

- a. **Persebaran Rating dan Review:** Menunjukkan jumlah rating yang diberikan oleh wisatawan dan melakukan visualisasi grafik batang untuk melihat persebaran rating.



Gambar 4. Persebaran Rating dan Review

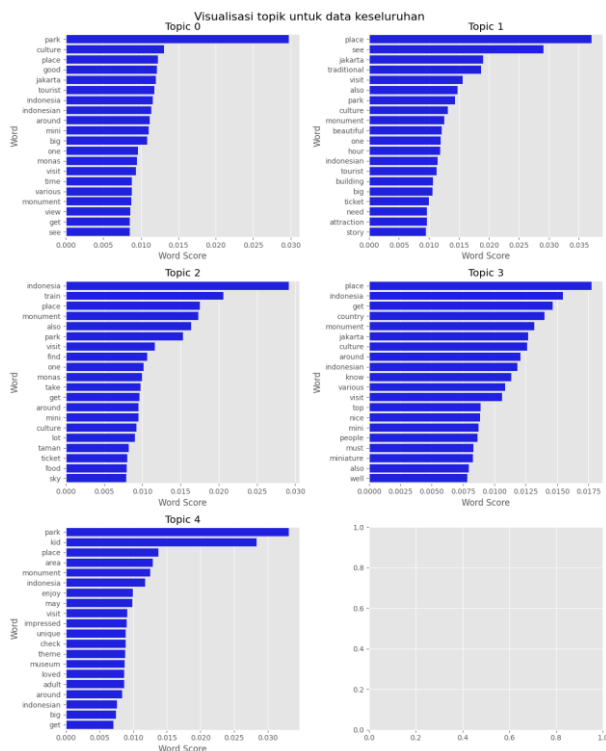
- b. **Visualisasi Word Cloud:** Membuat word cloud untuk memvisualisasikan kata-kata yang paling umum dalam ulasan. Word cloud ini menyoroti kata-kata yang paling sering muncul dalam ulasan dengan ukuran kata yang berbeda-beda.



Gambar 5. Word Cloud

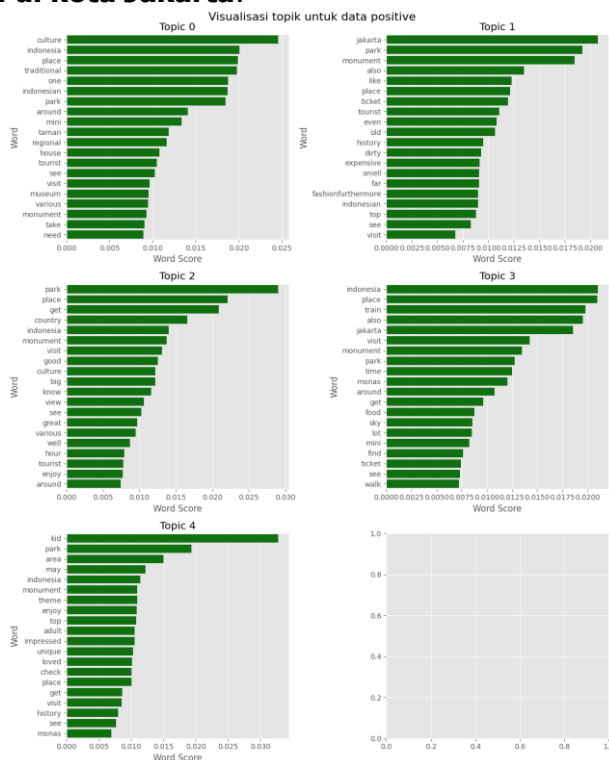
- c. **Topic Modeling dan Analisis Rating Sentimen pada Setiap Review:** Rating wisatawan diubah menjadi sentimen yang sesuai (4-5 maka positif, 3 maka netral, dan 1-2 maka negatif). Selanjutnya, model LDA diterapkan pada setiap sentimen untuk mengidentifikasi topik-topik utama yang muncul dalam ulasan. Langkah-langkahnya adalah sebagai berikut:
- Menggunakan representasi vektor dokumen yang telah dibuat sebelumnya.
 - Menginisialisasi model LDA dengan jumlah topik yang diinginkan.
 - Mengestimasi distribusi probabilitas topik dalam kata-kata dan probabilitas topik dalam dokumen menggunakan metode Gibbs Sampling atau metode iteratif lainnya.
 - Mengidentifikasi topik utama yang muncul dalam ulasan wisatawan berdasarkan distribusi probabilitas kata-kata dalam topik.
 - Membuat daftar kata-kata yang paling relevan untuk setiap topik yang diidentifikasi.
- d. **Topic Modelling untuk keseluruhan rating pada ulasan wisata di kota Jakarta:** Model LDA diterapkan untuk mengidentifikasi topik-topik utama yang muncul dalam ulasan TripAdvisor di kota Jakarta.

Memprediksi Pola Kedatangan Turis Mancanegara dan Menganalisis Ulasan Tripadvisor dengan LSTM dan LDA



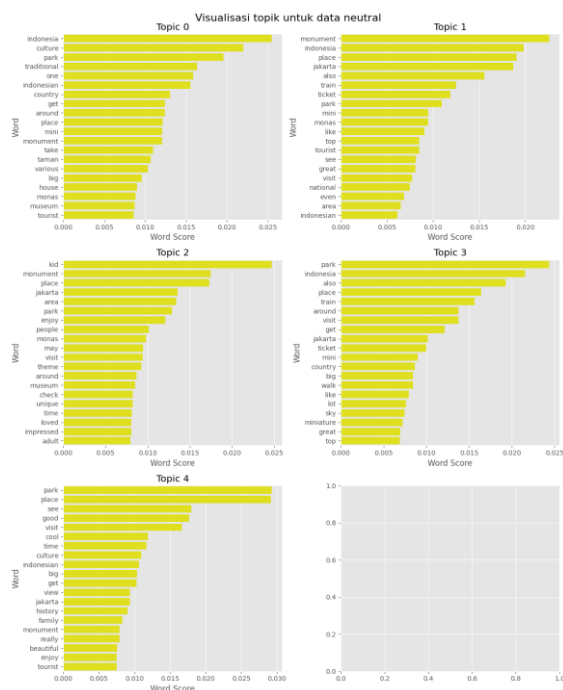
Gambar 5. Topic Modeling untuk keseluruhan rating di kota Jakarta

e. Topic Modelling untuk rating 4-5 yang dikategorikan sentimen positif pada ulasan wisata di kota Jakarta:



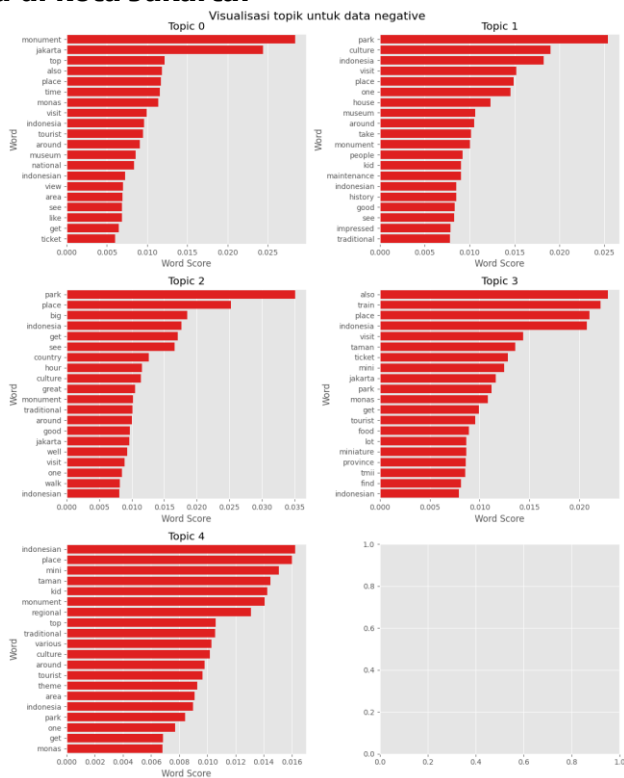
Gambar 6. Topic Modeling untuk sentimen positif di kota Jakarta

f. Topic Modelling untuk rating 3 yang dikategorikan sentimen netral pada ulasan wisata di kota Jakarta:



Gambar 7. Topic Modeling untuk sentimen netral di kota Jakarta

g. Topic Modelling untuk rating 1-2 yang dikategorikan sentimen negatif pada ulasan wisata di kota Jakarta:



Gambar 8. Topic Modeling untuk sentimen negatif di kota Jakarta

3.5 Evaluation

Hasil pelatihan model LSTM terlihat dalam beberapa epoch dengan mencatat loss dan val_loss. Val_loss yang lebih kecil menunjukkan performa yang lebih baik pada data validasi. Selain itu, dilakukan perhitungan MAPE pada data validasi yang mengukur tingkat kesalahan rata-rata dalam persentase. Contohnya, MAPE yang diberikan sekitar 2.69%. Topic Modeling LDA digunakan untuk evaluasi hasil clustering topik dan mengevaluasi koherensi antara topik-topik dalam kelompok. Menganalisis serta mengambil *insight* dari masing-masing kelompok topik dan memastikan di masing-masing kelompok topik tidak ada yang serupa atau saling tumpang tindih, hal tersebut dapat mengindikasikan adanya masalah dalam pengelompokan atau representasi topik.

3.6 Deployment / Kesimpulan

Setelah melalui pelatihan dan evaluasi, model LSTM siap digunakan dalam praktik untuk memprediksi nilai berdasarkan data masukan baru. Model ini berhasil dilatih dengan baik dan memiliki performa yang tinggi dalam memprediksi nilai. Pelatihan melibatkan data latihan dan validasi, di mana model belajar pola dan hubungan dalam data untuk melakukan prediksi akurat. Evaluasi menggunakan data validasi menunjukkan val_loss dan MAPE yang rendah, menunjukkan performa yang baik dalam memprediksi nilai pada data baru. Analisis topic modeling LDA dalam ulasan wisatawan Tripadvisor juga penting untuk industri pariwisata, memahami pandangan dan pengalaman wisatawan, serta mengidentifikasi aspek yang perlu diperbaiki atau ditingkatkan.

4. KESIMPULAN

Penelitian ini menggunakan model LSTM untuk memprediksi jumlah wisatawan di Indonesia dan memberikan rekomendasi kepada industri pariwisata terkait destinasi wisata di Jakarta. Prediksi peramalan menunjukkan peningkatan jumlah wisatawan dalam periode satu tahun mendatang. Selain itu, penelitian ini juga menggunakan metode cluster sentimen dan topic modeling LDA untuk memberikan rekomendasi spesifik untuk kota Jakarta. Rekomendasi tersebut termasuk peningkatan pelayanan, perbaikan kebersihan, infrastruktur, promosi tempat wisata alternatif, dan komunikasi informasi yang jelas kepada wisatawan. Penelitian ini memberikan wawasan berharga bagi industri pariwisata di Jakarta dalam meningkatkan pengalaman wisatawan dan mengoptimalkan potensi destinasi wisata. Rekomendasi yang dihasilkan dapat menjadi panduan dalam pengambilan keputusan dan perencanaan yang lebih baik untuk memajukan industri pariwisata di Jakarta.

UCAPAN TERIMA KASIH

Penulis ingin mengungkapkan rasa terima kasih kepada seluruh civitas akademik di Universitas Logistik dan Bisnis Internasional, khususnya program studi D4 Teknik Informatika, serta kepada teman dan partner yang telah memberikan bantuan dan berdiskusi selama proses penelitian ini.

DAFTAR RUJUKAN

- Adil, M., Wu, J. Z., Chakraborty, R. K., Alahmadi, A., Ansari, M. F., & Ryan, M. J. (2021). Attention-based STL-BiLSTM network to forecast tourist arrival. *Processes*, 9(10), 1759.

- Andiojaya, A., & Demirhan, H. (2019). A bagging algorithm for the imputation of missing values in time series. *Expert Systems with Applications*, *129*, 10-26.
- Asnawi, M. A. (2023). Pengaruh Media Sosial, Citra Destinasi, Dan Produk Destinasi Melalui Keputusan Berwisata Terhadap Sustainability Tingkat Kunjungan Wisatawan Di Daerah Tujuan Wisata (DTW) Provinsi Sulawesi Utara (Studi Empiris Pada Empat Kabupaten Di Provinsi Sulawesi Utara)= The Effect Of Social Media, Destination Image, And Destination Products Through Travel Decisions On The Level Of Tourist Visits In The Tourist Destination Area (DTW) Of North Sulawesi Province (An Empirical Study In Four Districts In North Sulawesi) (Doctoral dissertation, Universitas Hasanuddin).
- Bogaerts, T., Masegosa, A. D., Angarita-Zapata, J. S., Onieva, E., & Hellinckx, P. (2020). A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. *Transportation Research Part C: Emerging Technologies*, *112*, 62-77.
- Clark, A. (2018). The machine learning audit—crisp-dm framework. *Isaca Journal*, *1*, 42-47.
- Farkhod, A., Abdusalomov, A., Makhmudov, F., & Cho, Y. I. (2021). LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS) model. *Applied Sciences*, *11*(23), 11091.
- Feblian, D., & Daihani, D. U. (2016). Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X. *Jurnal Teknik Industri*, *6*(1).
- Han, S., & Anderson, C. K. (2021). Web scraping for hospitality research: Overview, opportunities, and implications. *Cornell Hospitality Quarterly*, *62*(1), 89-104.
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, *78*, 15169-15211.
- Jermsittiparsert, K., & Chankoson, T. (2019). Behavior of tourism industry under the situation of environmental threats and carbon emission: Time series analysis from Thailand. *International Journal of Energy Economics and Policy*, *9*(6), 366.
- Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., ... & Flach, P. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, *33*(8), 3048-3061.
- Meliana, C. (2021). Perbandingan Metode Long Short Term Memory (LSTM) DAN Genetic Algorithm-Long Short Term Memory (GA-LSTM) Pada Peramalan Polutan Udara (Doctoral dissertation, Universitas Muhammadiyah Semarang).

- Puteri, D. I. (2023). Implementasi Long Short Term Memory (LSTM) dan Bidirectional Long Short Term Memory (BiLSTM) Dalam Prediksi Harga Saham Syariah. *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, 11(1), 35-43.
- Putri, I. R., & Kusumaningrum, R. (2017). Latent Dirichlet allocation (LDA) for sentiment analysis toward tourism review in Indonesia. *Journal of Physics: Conference Series*, 801(1), 012073). IOP Publishing.
- Quatrini, E., Colabianchi, S., Costantino, F., & Tronci, M. (2022). Clustering application for condition-based maintenance in time-varying processes: A review using latent dirichlet allocation. *Applied Sciences*, 12(2), 814.
- Septiani, A., & Budi, I. (2022). Klasifikasi Ulasan Pengguna Aplikasi: Studi Kasus Aplikasi Ipusnas Perpustakaan Nasional Republik Indonesia (PNRI). *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 7(4), 1110-1120.
- Simanjuntak, B. A., Tanjung, F., & Nasution, R. (2017). Sejarah pariwisata: menuju perkembangan pariwisata Indonesia. Yayasan Pustaka Obor Indonesia.
- Urva, G., Albanna, I., Sungkar, M. S., Gunawan, I. M. A. O., Adhicandra, I., Ramadhan, S., ... & Junaidi, S. (2023). *Penerapan Data Mining Di Berbagai Bidang: Konsep, Metode, dan Studi Kasus*. PT. Sonpedia Publishing Indonesia.
- Wiemer, H., Drowatzky, L., & Ihlenfeldt, S. (2019). Data mining methodology for engineering applications (DMME)—A holistic extension to the CRISP-DM model. *Applied Sciences*, 9(12), 2407.
- Wu, D. C. W., Ji, L., He, K., & Tso, K. F. G. (2021). Forecasting tourist daily arrivals with a hybrid Sarima–Lstm approach. *Journal of hospitality & tourism research*, 45(1), 52-67.
- Yakup, A. P. (2019). Pengaruh Sektor Pariwisata terhadap Pertumbuhan Ekonomi di Indonesia (Doctoral dissertation, Universitas Airlangga).
- Yang, C. H., Wu, C. H., & Hsieh, C. M. (2020). Long short-term memory recurrent neural network for tidal level forecasting. *IEEE Access*, 8, 159389-159401.