

Peran Media Sosial Instagram dalam Memprediksi Jumlah Wisatawan Nusantara (Studi Kasus: Labuan Bajo, Nusa Tenggara Timur)

Navika Artiari¹, Erna Nurmawati^{2*}, Adelina Rahmawati³, Teguh Sugiyarto⁴

¹BPS Kota Sabang, Aceh, Indonesia

²Program Studi Komputasi Statistik, Politeknik Statistika STIS, Jakarta, Indonesia

³BPS Kabupaten Buol, Sulawesi Tengah, Indonesia

⁴Direktorat Statistik Keuangan, Informasi Teknologi, dan Pariwisata, Badan Pusat Statistik, Jakarta, Indonesia

e-mail: ^{2*}erna.nurmawati@stis.ac.id

ABSTRAK

Pemanfaatan big data dari media sosial seperti Instagram banyak digunakan sebagai sumber data alternatif dalam analisis pariwisata, khususnya untuk memahami pola promosi destinasi dan keterlibatan audiens secara daring. Penelitian ini bertujuan untuk mengidentifikasi topik utama konten promosi pariwisata pada media sosial serta menganalisis kontribusinya dalam memprediksi jumlah perjalanan wisatawan nusantara (wisnus) ke Provinsi Nusa Tenggara Timur. Pemodelan topik dilakukan menggunakan metode *Latent Dirichlet Allocation* (LDA) terhadap caption unggahan Instagram para pemangku kepentingan pariwisata di Nusa Tenggara Timur. Hasil pemodelan menghasilkan tiga topik utama, yaitu keindahan wisata bahari, penawaran paket trip dan aktivitas berlayar, serta penawaran trip dan sewa kapal. Analisis *engagement rate* (ER) menunjukkan bahwa rata-rata ER bulanan akun pariwisata di Nusa Tenggara Timur mencapai 28,07. Di antara ketiga topik yang terbentuk, topik keindahan wisata bahari memiliki rata-rata ER tertinggi, yaitu sebesar 42,24, yang mengindikasikan tingginya ketertarikan audiens terhadap konten visual dan narasi terkait daya tarik alam bahari. Selanjutnya, prediksi jumlah perjalanan wisnus dilakukan dengan mengombinasikan variabel ekonomi makro, indikator pariwisata, dan variabel berbasis media sosial. Hasil evaluasi menunjukkan bahwa model *Long Short-Term Memory* (LSTM) dengan seluruh variabel bebas, yaitu tingkat penghunian kamar, inflasi, indeks harga transportasi, variabel pandemi COVID-19, ER akun, serta ER pada masing-masing topik, memberikan kinerja terbaik. Model tersebut menghasilkan nilai Mean Absolute Percentage Error (MAPE) sebesar 18,75 persen, Mean Absolute Error (MAE) sebesar 83.577,40, dan Root Mean Square Error (RMSE) sebesar 111.630,97.

Kata kunci :

Engagement Rate Instagram; LDA; LSTM; Prediksi; Wisatawan Nusantara

ABSTRACT

The use of big data derived from social media platforms such as Instagram has increasingly been adopted as an alternative data source in tourism analysis, particularly to understand destination promotion patterns and online audience engagement. This study aims to identify the main topics of tourism promotional content on social media and to examine their contribution to predicting the number of domestic tourist trips to the Province of East Nusa Tenggara. Topic modeling was conducted using the Latent Dirichlet Allocation (LDA) method on Instagram captions posted by tourism stakeholders in East Nusa Tenggara. The results of the topic modeling revealed three main topics, namely marine tourism attractiveness, tour package and sailing activity offerings, and trip and boat rental promotions. The analysis of engagement rates (ER) indicates that the average monthly ER of tourism-related accounts in East Nusa Tenggara reached 28.07. Among the identified topics, the marine tourism attractiveness topic recorded the highest average ER, at 42.24, indicating strong audience interest in visual content and narratives related to marine natural attractions. Furthermore, the prediction of domestic tourist trips was carried out by integrating macroeconomic variables, tourism indicators, and social media-based variables. The evaluation results demonstrate that the Long Short-Term Memory (LSTM) model incorporating all explanatory variables—including hotel occupancy rates, inflation, transportation price indices, the COVID-19 pandemic variable, account-level ER, and topic-specific ER—achieved the best performance. This model produced a Mean Absolute Percentage Error (MAPE) of 18.75 percent, a Mean Absolute Error (MAE) of 83,577.40, and a Root Mean Square Error (RMSE) of 111,630.97.

Keywords :

Instagram Engagement Rate; LDA; LSTM; Prediction; Domestic Tourist

A. PENDAHULUAN

Nusa Tenggara Timur merupakan salah satu provinsi di Indonesia yang memiliki potensi sektor pariwisata tinggi. Pada masa pasca pandemik pemerintah berupaya memulihkan sektor pariwisata dengan pengembangan desa wisata dan pembangunan Destinasi Super Prioritas (DSP) (Bahri & Abdillah, 2022; Sugiyarto & Mahagangga, 2020). Destinasi Prioritas (DPP) yang berjumlah sepuluh kembali difokuskan menjadi lima DSP dengan harapan

mampu meningkatkan kuantitas dan kualitas kunjungan wisatawan domestik dan internasional. DPSP tersebar di lima provinsi yaitu Danau Toba di Sumatera Utara, Borobudur di Jawa Tengah, Mandalika di Nusa Tenggara Barat (NTB), Labuan Bajo di Nusa Tenggara Timur (NTT), dan Likupang di Sulawesi Utara (Rochmah & Lentera, 2023).

Sektor pariwisata telah mendukung terbukanya lapangan pekerjaan baru, peningkatan taraf hidup, dan pendapatan masyarakat, serta peningkatan nilai

tukar mata uang (Anggara et al., 2025). Beberapa dampak yang diciptakan sektor pariwisata dilihat dari aspek ekonomi tersebut tentunya sangat berperan dalam pertumbuhan ekonomi suatu negara (Sun et al., 2025). Di Indonesia, sektor pariwisata telah menyumbang lima persen pada tahun 2019 menurut laporan *e-Trends and Policies* tahun 2022 (OECD, 2022). Namun, angka tersebut mengalami penurunan yang besar pada tahun 2020 yaitu menjadi 2,2% akibat adanya pandemi Covid-19.

Di Nusa Tenggara Timur, pada tahun 2023 jumlah kunjungan wisatawan provinsi Nusa Tenggara Timur mencapai 824.428 orang, yang terdiri dari 744.428 wisatawan domestik dan 80.000 wisatawan mancanegara (BPS Provinsi NTT, 2023). Jumlah wisatawan ke destinasi wisata Provinsi Nusa Tenggara Timur didominasi oleh Labuan Bajo, yaitu sebesar 400.000 orang. Hal ini selaras dengan ditetapkan Labuan Bajo sebagai Destinasi Super Prioritas (DSP). Dari jumlah kunjungan tersebut, sektor pariwisata menyumbang pendapatan daerah Provinsi Nusa Tenggara Timur sebesar Rp1,2 triliun. Dominasi wisatawan nusantara menunjukkan bahwa pergerakan wisata domestik memiliki peran strategis dalam menjaga keberlanjutan sektor pariwisata daerah. Oleh karena itu, prediksi jumlah wisatawan nusantara di NTT menjadi penting sebagai dasar perencanaan kebijakan, pengelolaan destinasi, serta penyediaan infrastruktur dan layanan pariwisata (Kariyamin et al., 2025; Prayoga & Wijayanto, 2024).

Seiring dengan perkembangan teknologi digital, media sosial, khususnya Instagram, telah menjadi salah satu sumber informasi utama bagi wisatawan dalam mencari referensi destinasi (Dewi, 2022; Dianafakroh et al., 2025; Ismarizal & Kusumah, 2023; Maulana et al., 2024). Pemilihan Instagram dalam penelitian ini didasarkan pada karakteristik platform yang paling sesuai dengan promosi destinasi berbasis visual seperti Labuan Bajo. Berbagai studi menunjukkan bahwa Instagram memiliki pengaruh signifikan terhadap penyebaran konten visual dan narasi pengalaman wisata secara cepat dan luas (Andzani et al., 2024; Ganur & Putra, 2023; Havianto & Artiningrum, 2022; Qian, 2022), sehingga berpengaruh terhadap minat dan keputusan berkunjung wisatawan (Joseph et al., 2024). Dibandingkan Twitter (X) yang lebih berbasis teks dan diskusi opini, serta TikTok yang sangat bergantung pada algoritma For You Page yang fluktuatif, Instagram menawarkan struktur konten visual dan narasi yang lebih stabil serta terkurasi untuk membangun citra destinasi secara konsisten.

Dari sisi metodologis, Instagram juga lebih relevan untuk integrasi *engagement rate* dan *topic modeling*. Metrik keterlibatan seperti *likes*, *comments*, dan *shares* pada Instagram relatif terstandarisasi dan telah banyak digunakan sebagai indikator efektivitas promosi destinasi. Secara teoretis, hubungan antara *engagement rate* (ER) dan niat berkunjung dapat dijelaskan melalui kerangka *Electronic Word of*

Mouth (eWOM) dan *destination image theory*. Dalam perspektif eWOM, interaksi pengguna berupa *likes*, komentar, dan *shares* merepresentasikan respons kognitif dan afektif terhadap informasi destinasi yang tersebar secara digital, yang selanjutnya memengaruhi persepsi, kepercayaan, dan niat perilaku wisatawan (Joseph et al., 2024; Ramdan et al., 2025). ER yang tinggi menunjukkan intensitas eksposur dan partisipasi audiens terhadap konten promosi, yang dalam teori pembentukan citra destinasi berkontribusi pada pembentukan *affective image* dan *conative image*, yaitu tahap yang secara langsung berkaitan dengan niat berkunjung (Blanco-Moreno et al., 2024; Dianafakroh et al., 2025). Studi empiris terbaru juga menegaskan bahwa interaksi media sosial berperan sebagai sinyal sosial (*social proof*) yang memperkuat persepsi kualitas dan daya tarik destinasi, sehingga meningkatkan *travel intention* secara signifikan (Wu et al., 2024). Dengan demikian, ER dapat diposisikan sebagai indikator kuantitatif dari dinamika eWOM yang secara konseptual dan empiris berkorelasi dengan niat berkunjung wisatawan.

Keefektifan promosi pariwisata melalui Instagram sangat dipengaruhi oleh konten dan performa postingan. Postingan yang menampilkan keindahan destinasi, pengalaman wisata, serta informasi perjalanan terbukti mampu meningkatkan minat kunjungan wisatawan domestik (Blanco-Moreno et al., 2024). Dalam konteks NTT, berbagai akun Instagram stakeholder pariwisata meliputi Dinas Pariwisata Provinsi, pengelola objek wisata, penyedia jasa dan promosi pariwisata, serta penyedia jasa akomodasi memainkan peran penting dalam membentuk citra destinasi dan menarik wisatawan nusantara. Aktivitas posting dari akun-akun tersebut berpotensi memengaruhi tingkat keterlibatan pengguna dan pada akhirnya berdampak pada jumlah kunjungan wisata.

Untuk memahami isi dan pesan yang disampaikan dalam konten Instagram secara sistematis, diperlukan pendekatan analitis yang mampu mengekstraksi pola tematik dari data teks (Recuero-Virto, 2022). *Topic modeling*, khususnya Latent Dirichlet Allocation (LDA), telah banyak digunakan dalam penelitian pariwisata untuk mengidentifikasi tema dominan dalam ulasan wisatawan maupun konten media sosial (Saura et al., 2020; Shang & Luo, 2023). Penerapan *topic modeling* pada *caption* Instagram akun stakeholder pariwisata memungkinkan peneliti mengidentifikasi tema promosi utama, seperti keindahan alam, paket wisata, atau fasilitas pendukung, yang relevan dengan minat wisatawan nusantara.

Selain analisis tema, *engagement rate* (ER) Instagram yang mencerminkan tingkat interaksi pengguna melalui *likes*, *comments*, dan *shares* merupakan indikator penting untuk menilai keefektifan postingan akun media sosial (Kusuma et al., 2023). ER yang tinggi menunjukkan bahwa konten yang disajikan mampu menarik perhatian dan

keterlibatan audiens secara signifikan (Shahbaznezhad et al., 2021). Dalam konteks pariwisata, *engagement rate* telah digunakan sebagai proksi minat dan respons calon wisatawan terhadap konten promosi destinasi (Mariani & Borghi, 2021; Swandeni et al., 2024).

Dengan demikian, integrasi *engagement rate* akun Instagram stakeholder pariwisata dan *engagement rate* berdasarkan hasil topic modeling caption menjadi pendekatan yang menjanjikan untuk meningkatkan akurasi prediksi jumlah wisatawan nusantara. Pendekatan ini memungkinkan pemanfaatan data media sosial tidak hanya dari sisi kuantitas interaksi, tetapi juga dari sisi kualitas dan relevansi tema konten. Oleh karena itu, penelitian yang mengombinasikan analisis topic modeling, *engagement rate* Instagram, dan model prediksi jumlah wisatawan nusantara di Nusa Tenggara Timur menjadi penting untuk menghasilkan informasi yang lebih komprehensif dan mendukung pengambilan kebijakan pariwisata berbasis data.

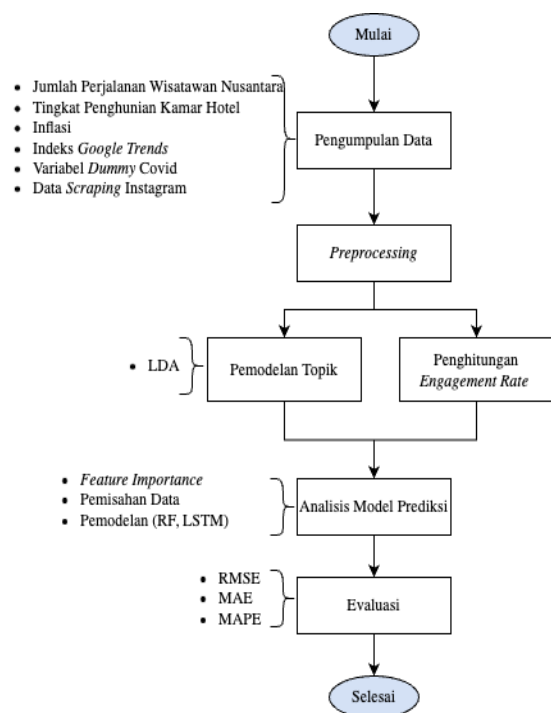
Memprediksi jumlah perjalanan wisatawan nusantara dapat dilakukan dengan pendekatan machine learning dan deep learning (Wu et al., 2024), di antaranya Random Forest Regression (RFR) (Peng et al., 2021) dan Long Short-Term Memory (LSTM) (Salamanis et al., 2022), yang telah terbukti efektif dalam menangkap pola nonlinier dan dinamika temporal data pariwisata. Akurasi model prediksi dapat ditingkatkan dengan memasukkan variabel-variabel yang berkaitan langsung dengan aktivitas dan permintaan wisatawan (Zhang et al., 2025), seperti Tingkat Penghunian Kamar (TPK) hotel, inflasi, Indeks Google Trends (IGT) berdasarkan kombinasi kata kunci pencarian terkait perjalanan wisata, serta variabel dummy Covid-19 untuk merepresentasikan dampak gangguan eksternal terhadap mobilitas wisata.

Meskipun berbagai penelitian telah mengkaji faktor-faktor yang memengaruhi jumlah perjalanan wisatawan nusantara, sebagian besar studi masih berfokus pada variabel makroekonomi (seperti PDB, pendapatan per kapita, harga, inflasi), faktor musiman, serta indeks Google Trends. Hingga saat ini, belum ditemukan penelitian yang secara eksplisit mengintegrasikan *Engagement Rate* media sosial sebagai variabel eksogen dalam model peramalan jumlah perjalanan wisatawan domestik. Padahal, *Engagement Rate* mencerminkan tingkat interaksi, perhatian, dan respons audiens terhadap konten pariwisata di platform digital, yang secara teoritis dapat merepresentasikan minat aktual dan niat bepergian secara lebih dinamis dibandingkan indikator konvensional. Oleh karena itu, penggunaan *Engagement Rate* sebagai variabel independen dalam pemodelan prediktif jumlah perjalanan wisatawan nusantara merupakan pendekatan baru yang mengisi *research gap* pada integrasi indikator interaksi media sosial dengan model peramalan permintaan pariwisata.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengkaji konten dan efektivitas unggahan akun Instagram pariwisata, dengan mengukur nilai *engagement rate* (ER) berdasarkan jenis akun stakeholder meliputi Dinas Pariwisata Provinsi, pengelola objek wisata, penyedia jasa dan promosi pariwisata, serta penyedia jasa akomodasi dan hasil pemodelan topik pada caption Instagram. Selanjutnya, nilai ER akun dan ER berbasis topik dikombinasikan dengan variabel TPK, inflasi, IGT, dan dummy Covid-19 sebagai variabel pendukung data historis dalam memprediksi jumlah perjalanan wisatawan nusantara di Provinsi Nusa Tenggara Timur, dengan membangun dan membandingkan kinerja model RFR dan LSTM menggunakan berbagai kombinasi variabel untuk memperoleh hasil prediksi yang paling optimal.

B. METODE PENELITIAN

Alur metode penelitian secara keseluruhan yang dilakukan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Metode Penelitian

Pengumpulan Data

Penelitian ini menggunakan data sekunder yang diperoleh beberapa sumber. Data jumlah perjalanan wisnus, TPK (Prayoga & Wijayanto, 2024), dan Inflasi (Raifu & Afolabi, 2026), dikumpulkan dari website BPS. Data IGT (Llewellyn et al., 2023) dikumpulkan dari Trends *Explore tool* yang disediakan oleh Google. IGT dikumpulkan berdasarkan tiga keyword pencarian yang merujuk dari penelitian (Rödel, 2017) dan disesuaikan dengan kebutuhan penelitian ini. Daftar keyword yang digunakan pada penelitian ini yaitu “wisata (diikuti provinsi)”, “provinsi” dan “hotel (diikuti provinsi)”.

Keyword pencarian tersebut difilter berdasarkan lokasi yaitu “Indonesia”, kategori yaitu “Travel”, dan jenis pencarian berupa “Web Search”. Sedangkan data covid merupakan data dummy yang bernilai 0 atau 1. Nilai 0 berarti tidak ada pandemi covid sedangkan 1 menunjukkan adanya pandemi covid. Pembentukan data didasarkan keadaan pandemi sesuai Keputusan Presiden No 11 tahun 2020 dan Keputusan Presiden No. 17 tahun 2023.

Data Instagram terkait pariwisata dikumpulkan dengan *scraping* menggunakan library Instaloader di python pada post dari akun Instagram stakeholder pariwisata. Terdapat tiga jenis stakeholder pariwisata yaitu pemerintah, masyarakat setempat, dan pelaku usaha/swasta (Murphy & Murphy, 2004). Pemerintah diwakilkan oleh akun Instagram dinas pariwisata provinsi. Masyarakat setempat yang berperan sebagai host dalam menciptakan lingkungan wisata yang kondusif (Rahim, 2012) diwakilkan oleh akun pengelola objek wisata DSP. Sedangkan untuk pelaku usaha/swasta diwakilkan oleh akun penyedia jasa dan promosi wisata serta akun penyedia jasa akomodasi.

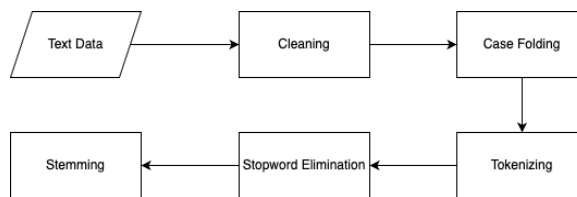
Akun Instagram dinas pariwisata dan pengelola objek wisata DSP dipilih berdasarkan kriteria keresmian akun. Sedangkan kedua akun pelaku usaha/swasta dipilih berdasarkan kecakupan waktu sejak akun dibuat dan ER rata-rata yang tertinggi. Daftar akun Instagram stakeholder pariwisata di Provinsi Nusa Tenggara Timur yang digunakan pada penelitian ini dapat dilihat pada tabel 1. Pengumpulan data dari seluruh post Instagram mencakup atribut ‘Post URL’, ‘Date’, ‘Time’, ‘Likes’, ‘Comments’, dan ‘Caption’. Data tersebut kemudian disimpan dalam format excel.

Tabel 1. Akun Stakeholder Pariwisata

Stakeholder Pariwisata	Akun Instagram Stakeholder Pariwisata	Nama Akun
Pemerintah	Dinas pariwisata provinsi	disparekraf.ntt
Masyarakat setempat	Pengelola objek wisata	bpolbf
Pelaku usaha/swasta	Penyedia jasa dan promosi pariwisata	labuanbajoku_trip
	Penyedia jasa akomodasi	plataran.komodo

Preprocessing

Data yang telah dikumpulkan perlu melalui tahap preprocessing sebelum digunakan dalam pemodelan untuk mempersiapkan data menjadi data yang terstruktur dan dapat diproses pada tahapan berikutnya (Han et al., 2022). Data post Instagram yang telah dikumpulkan diperiksa untuk menyesuaikan periode waktu penelitian dan menghapus baris data jika merupakan duplikasi. Tahap preprocessing pada *caption* dilakukan untuk menyiapkan data *caption* sebelum dilakukan pemodelan topik. Selanjutnya dilakukan tahap preprocessing pada data *caption* dari post Instagram dengan tahapan pada gambar 2.



Gambar 2. Preprocessing Caption

a. Cleaning

Cleaning bertujuan untuk menghilangkan berbagai karakter simbol berupa tanda baca, emoji, dan angka. Selain itu, tahapan ini juga akan menghilangkan komponen teks seperti username, hashtag, dan lain sebagainya. Tahap *cleaning* dilakukan menggunakan library *re.sub()* yang berfungsi untuk menangani penghapusan dan penggantian pada data string.

b. Case Folding

Pada tahap ini seluruh huruf dalam data teks akan diubah menjadi huruf kecil dengan memanfaatkan library *lower()*.

c. Tokenizing

Tokenizing adalah proses pemisahan kalimat menjadi token-token yang berupa kata yang saling independen. Tahap ini dilakukan menggunakan library *word_tokenize* dari *nlTK* dan akan menghasilkan token atau potongan kata dari data teks.

d. Stopword Elimination

Stopword merupakan kata-kata yang tidak informatif karena tidak memberi pengaruh namun sering muncul dalam dokumen. *Stopword* yang paling sering ditemui misalnya kata hubung dan kata ganti orang. Pada tahap ini akan dilakukan penghapusan *stopword* untuk mengurangi kata dalam corpus menggunakan library *stopwords()*.

e. Stemming

Stemming adalah tahap terakhir dari preprocessing data teks, dimana seluruh kata diubah menjadi kata dasarnya dengan menghilangkan semua imbuhan. Tahap ini bertujuan untuk menghilangkan noise dengan mengurangi variasi kata yang sebenarnya berasal dari kata dasar sama. *Stemming* dilakukan dengan library *StemmerFactory()* dari Sastrawi.

Pemodelan Topik

Data *caption* hasil *preprocessing* kemudian akan digabungkan berdasarkan provinsi. Kemudian data tersebut akan digunakan pada tahap pemodelan topik yang bertujuan untuk mengetahui topik apa saja yang ada pada post Instagram akun pariwisata di provinsi Nusa Tenggara Timur. Pemodelan topik dilakukan menggunakan metode *Latent Dirichlet Allocation (LDA)* (Uthirapathy & Sandanam, 2023) menggunakan library *Gensim* di python.

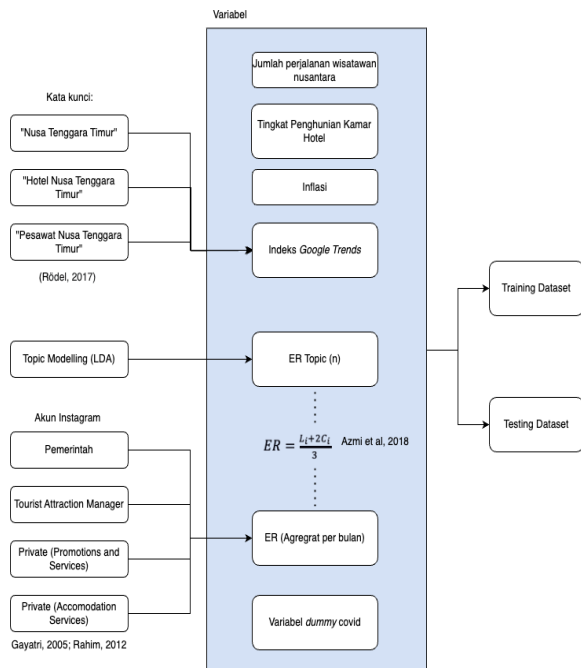
Penghitungan Engagement Rate (ER)

Tahap selanjutnya setelah dilakukan pemodelan topik pada *caption* adalah perhitungan ER. Nilai ER merupakan salah satu nilai yang digunakan untuk mengetahui efektivitas penyebaran informasi

dari suatu konten. Perhitungan engagement rate dilakukan pada setiap post Instagram berdasarkan jumlah likes dan comments (Azmi & Budi, 2018) dengan menggunakan persamaan (1).

$$ER_i = \frac{L_i + 2C_i}{3} \quad (1)$$

- ER_i = Engagement rate dari post i
- L_i = Likes dari post i
- C_i = Comments dari post i



Gambar 3. Alur Pembentukan Variabel

Setelah melalui tahap pemodelan topik dan perhitungan ER tersebut, akan dibentuk variabel yang akan digunakan pada pemodelan. Variabel ER akun merupakan rata-rata ER seluruh post dari akun stakeholder di provinsi Nusa Tenggara Timur. Sedangkan variabel ER topik sejumlah tertentu dihasilkan dengan merata-ratakan ER post berdasarkan topik yang terbentuk. Sehingga, bagan alur pembentukan variabel tersebut hingga persiapan data untuk pemodelan dalam penelitian ini dapat dilihat pada gambar 3.

Analisis Model Prediksi

a. Feature Importance dan kombinasi variabel

Teknik feature importance dilakukan untuk mengetahui variabel apa saja yang paling berpengaruh (Molnar, 2020) terhadap jumlah perjalanan wisnu di setiap provinsi DSP. Feature importance dilakukan dengan menerapkan metode random forest menggunakan library RandomForestRegressor dari sklearn di python.

b. Data Splitting

Sebelum dilakukan pemodelan jumlah perjalanan wisnu, data terlebih dahulu dibagi

menjadi data pelatihan (data training) dan data pengujian (data testing). Data pelatihan dan data pengujian dibagi dengan perbandingan 85% untuk data pelatihan dan sisanya 15% untuk data pengujian (Brownlee, 2016).

c. Pemodelan

Data yang telah dipisahkan menjadi data training dan data testing kemudian dimasukkan ke dalam model. Model tersebut kemudian diuji dan dibandingkan untuk mendapatkan model terbaik dari model machine learning menggunakan Random Forest Regression dan LSTM.

Pada saat pembangunan model, akan diatur hyperparameter untuk model Random Forest Regression dan LSTM dengan menggunakan 5-fold cross validation grid search (Andariesta and Wasesa 2022; Salamanis et al. 2022; Swetha, Rasheed, and Harigovindan 2023). Pada RFR akan diatur jumlah pohon, kedalaman, min sample split, dan min sample leaf.

Tabel 2. Hyperparameter RFR

Hyperparameter	Ukuran
Max depth	128, 192, 256
Min samples leaf	2, 5, 10
Min samples split	1, 2, 4
N estimators	50, 100, 150

Selain itu, pada pemodelan juga akan dilakukan dengan mempertimbangkan feature importance dari RFR dan kombinasi beberapa variabel statistik.

Selanjutnya, pembangunan metode LSTM dilakukan dengan menerapkan hyperparameter tuning dengan metode grid search (Aggarwal, 2018). Kombinasi ukuran hyperparameter yang digunakan pada penelitian ini dapat dilihat pada Tabel 3.

Tabel 3. Hyperparameter LSTM

Hyperparameter	Ukuran
Batch size	32
Dropout rate	0,05; 0,15; 0,25
Epoch	50, 100, 150
Learning rate	0,0001; 0,001; 0,01
LSTM units	128, 192, 256
Optimizer	adam, sigmoid

Selain itu, terdapat hyperparameter n_past yang merupakan lag pada model LSTM yang ditetapkan sebelum pemodelan (Brownlee, 2016). Semakin tinggi nilai n_past semakin kompleks model yang terbentuk tetapi waktu yang dibutuhkan juga semakin lama. Pada penelitian ini nilai n_past ditentukan dengan mempertimbangkan panjang data yang terbatas.

Evaluasi Model

Penelitian ini menerapkan tiga matriks akurasi pada model yaitu MAPE, MAE, dan RMSE. Perhitungan ketiga matriks akurasi dilakukan dengan mengimport fungsi dari library sklearn.metrics.

C. HASIL DAN PEMBAHASAN

Pemodelan Topik

Pemodelan topik dilakukan untuk seluruh unggahan dari akun stakeholder provinsi DSP Nusa Tenggara Timur. Percobaan pemodelan topik dilakukan dengan mengatur parameter jumlah topik dengan rentang 3 sampai 10 dan jumlah passes yaitu 1 sampai 20. Pemodelan dipilih berdasarkan nilai coherence score tertinggi. Sehingga hasil pemodelan topik setiap provinsi menghasilkan jumlah kategori topik dan bahasan yang berbeda. Hasil dan pembahasan diperbolehkan menunjukkan narasi yang dilengkapi dengan data-data berupa tabel maupun gambar. Hasil harus didukung oleh referensi terkait ataupun dapat membandingkan dengan penelitian sebelumnya.

Tabel 4. Hasil Pemodelan Topik Nusa Tenggara Timur

Jenis Topik	Kata kunci	Bahasan
0	temu, pariwisata, pulau, Indonesia, alam, indah, nikmat, taut, laut, pantai	wisata bahari dan keindahannya
1	pax, trip, com, berlayar, www, pantai, info, pelayaran, kabin, pulau	penawaran paket trip dan berlayar
2	trip, perjalanan, boat, wa, pulau, biaya, call, pribadi, island, info	penawaran trip dan sewa boat dengan biaya tertentu

Sumber: Data primer yang diolah, 2024

Tabel 4 menunjukkan hasil pemodelan topik untuk Nusa Tenggara Timur yang memiliki nilai coherence score 0,770 pada jumlah topik 3 dan jumlah *passes* 5. Berdasarkan kata kunci penyusunnya topik 0 merupakan topik yang memiliki bahasan mengenai wisata bahari dan keindahannya. Topik 1 dengan kata penyusun ‘pax’, ‘trip’, ‘com’, ‘berlayar’, dst, memiliki bahasan mengenai penawaran paket trip dan berlayar dengan menyertakan informasi lebih lanjut. Sedangkan topik 2 menunjukkan bahasan mengenai penawaran trip pribadi dan sewa *boat* dengan biaya tertentu.

Hasil analisis konten media sosial menunjukkan bahwa akun-akun yang dikelola oleh stakeholder pariwisata di Nusa Tenggara Timur (NTT) secara konsisten memposting destinasi wisata berbasis pantai. Beberapa pantai yang paling sering muncul dalam unggahan antara lain Pantai Lasiana, Pantai Kolbano, Pantai Koka, Pantai Walakiri, Pantai Nihiwatu (Nih), Pantai Oetune, Pantai Oesapa, Pantai Tenau, Pantai Nemberala (Nembrala), Pantai Ratenggaro, Pantai Tarimbang, Pantai Wini, Pantai Manikin, Pantai Cepiwatu, Pantai Pede, Pantai Oebali, Pantai Oeseli, Pantai Bolok, Pantai Mali, Pantai Kobano, Pantai Liangbala, Pantai Napae, Pantai Watotena, Pantai Weri, Pantai Mbawana (Bawana), Pantai Panmuti, dan Pantai Kadelang. Unggahan mengenai pantai-pantai tersebut umumnya memperoleh jumlah like dan komentar yang tinggi,

sehingga menghasilkan tingkat engagement rate (ER) yang relatif besar dibandingkan konten non-pantai. Tingginya interaksi ini mencerminkan daya tarik wisata bahari sebagai ikon utama pariwisata NTT. Temuan tersebut juga selaras dengan karakteristik geografis NTT sebagai provinsi kepulauan yang memiliki garis pantai panjang dan keragaman objek wisata pesisir, sehingga konten pantai secara alami menjadi fokus promosi sekaligus sumber respons positif dari audiens media sosial.

Perhitungan Engagement Rate

Perhitungan ER dilakukan berdasarkan persamaan (1). Nilai ER dari setiap unggahan akun dihitung kemudian dirata-ratakan berdasarkan bulan untuk akhirnya diperoleh rata-rata ER akun bulanan.

Hasil penghitungan ER di provinsi Nusa Tenggara Timur dapat dilihat pada Tabel 5.

Tabel 5. ER Nusa Tenggara Timur

Jumlah Post	Jumlah Likes	Jumlah Comments	Rata-rata ER
5.625	558.938	12.404	28.074

Sumber: Data primer yang diolah, 2024

Sedangkan rincian rata-rata ER dari setiap akun stakeholder pariwisata di Nusa Tenggara Timur dapat diamati pada tabel 6.

Tabel 6. Rata-rata ER Berdasarkan Jenis Stakeholder

Jenis Stakeholder	Rata-rata ER
Pengelola Objek Wisata DSP	66.261
Dinas Pariwisata	12.932
Penyedia Jasa dan Promosi	18.448
Pariwisata	
Penyedia akomodasi	31.446

Analisis Deskriptif

Berdasarkan tahapan yang telah dilakukan, dataset yang akan digunakan pada pemodelan terdiri dari 59 baris. Variabel bebas yang digunakan untuk pemodelan prediksi jumlah perjalanan wisnus ke Nusa Tenggara Timur adalah TPK, Inflasi, IGT, Covid, ER akun, ER_topik0, ER_topik1, dan ER_topik2. Tabel 7 merupakan statistik deskriptif dari data yang digunakan untuk pemodelan prediksi jumlah perjalanan wisnus ke Nusa Tenggara Timur. Variabel wisnus memiliki rata-rata 352.866,10 dengan standar deviasi sebesar 184.513,42. Variabel TPK memiliki rata-rata sebesar 39,00 dengan standar deviasi sebesar 11,25. Sedangkan inflasi di Nusa Tenggara Timur memiliki rata-rata sebesar 0,19 dengan standar deviasi sebesar 0,58.

Tabel 7. Statistik Deskriptif Data Nusa Tenggara Timur

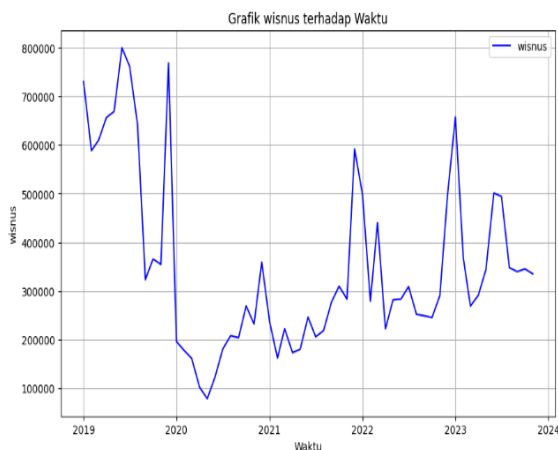
Variabel	Mean	Standar Deviasi	Min.	Median	Max.
Wisnus	352.866,10	184.513,42	78.698,00	290.876,00	800.071,00
TPK	39,00	11,25	11,57	41,39	58,72
Inflasi	0,19	0,58	-0,94	0,23	1,64
IGT	36,30	8,78	17,93	37,35	62,68

Covid	0,67	0,47	0,00	1,00	1,00
ER akun	28,73	16,35	10,28	24,32	88,87
ER_topik0	42,24	32,15	14,43	33,26	187,95
ER_topik1	23,97	37,51	8,86	18,19	297,93
ER_topik2	17,39	6,11	0,00	17,06	31,13

Sumber: Data primer yang diolah, 2024

Variabel bebas yang diperoleh dari *big data* juga memiliki statistik deskriptif yang beragam. Rata-rata IGT Sumatera Utara adalah sebesar 36,30 dengan standar deviasi sebesar 8,78. ER akun stakeholder di Sumatera Utara memiliki rata-rata sebesar 28,73 dengan standar deviasi 16,35. ER topik 0 yang berisi bahasan mengenai wisata bahari dan keindahannya memiliki rata-rata sebesar 42,24 dan standar deviasi 32,15. ER topik 1 merupakan topik dengan bahasan penawaran paket trip dan berlayar memiliki rata-rata 23,97 dan standar deviasi 37,51. Sedangkan ER topik 2 yang berisi bahasan penawaran trip dan sewa boat dengan biaya tertentu memiliki rata-rata 17,39 dan standar deviasi mencapai 6,11.

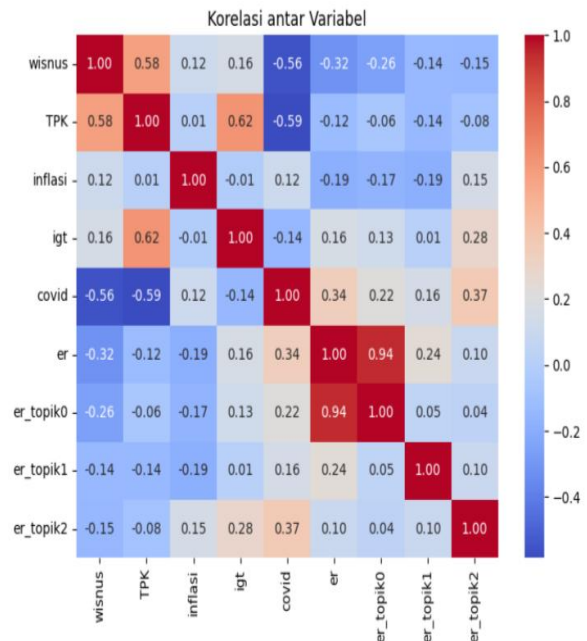
Berdasarkan grafik pada Gambar 4, dapat disimpulkan bahwa jumlah perjalanan wisnus ke Nusa Tenggara Timur memiliki pola yang fluktuatif. Jumlah perjalanan wisnus ke Nusa Tenggara Timur mengalami penurunan paling tinggi pada Januari 2020 dan cenderung terus mengalami penurunan hingga November 2022. Hal ini bersamaan dengan adanya pandemi covid-19 di Indonesia yang ditetapkan resmi pertama kali pada Maret 2020.



Gambar 4. Jumlah perjalanan wisnus ke Nusa Tenggara Timur, Januari 2019-November 2023

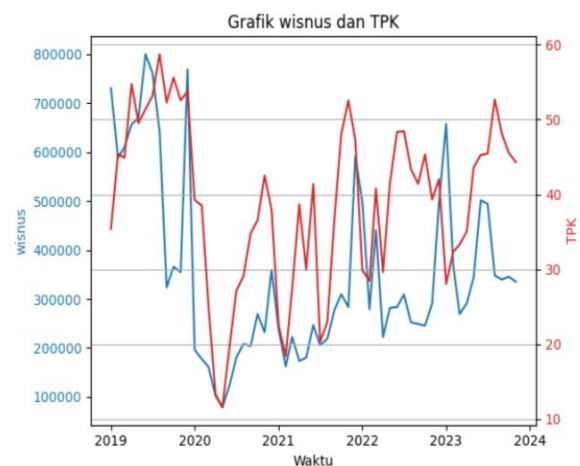
Gambar 5 menunjukkan plot korelasi dari seluruh variabel di Nusa Tenggara Timur yang dihitung menggunakan koefisien korelasi pearson. Nilai korelasi antara jumlah perjalanan wisnus dan TPK di Nusa Tenggara Timur memiliki nilai yang positif dan cukup kuat yaitu sebesar 0,58. Nilai korelasi yang positif dengan jumlah perjalanan wisnus juga dimiliki oleh IGT yaitu sebesar 0,16 yang berarti hubungan keduanya positif namun cukup lemah. Sedangkan variabel dummy covid-19 merupakan variabel yang memiliki korelasi yang

negatif dan cukup kuat dengan jumlah perjalanan wisnus yaitu sebesar -0,56.



Gambar 5. Plot korelasi antara jumlah perjalanan wisnus ke Nusa Tenggara Timur dan variabel bebas yang digunakan

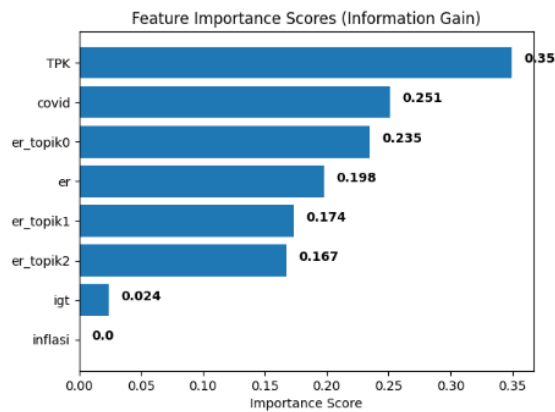
Berdasarkan grafik pada gambar 6, terlihat bahwa TPK di Nusa Tenggara Timur juga memiliki pola yang fluktuatif. TPK mengalami penurunan sejak awal 2020 hingga menjelang akhir 2022. Hal ini menunjukkan bahwa penurunan TPK terjadi bersama dengan adanya pandemi covid-19. Sehingga, TPK dan jumlah perjalanan wisnus yang juga mengalami penurunan selama covid-19 cenderung berbanding lurus. Hal ini sesuai dengan nilai korelasi keduanya yang positif yaitu sebesar 0,58.



Gambar 6. Grafik jumlah perjalanan wisnus dan TPK Nusa Tenggara Timur pada Januari 2019-November 2023

Pemodelan Prediksi Wisatawan Nusantara

Sebelum memasuki tahapan pemodelan, variabel bebas dari data dihitung skor importance terlebih dahulu. Hasil perhitungan tersebut terlihat pada Gambar 7.



Gambar 7. Grafik Hasil Feature Importance

Gambar 7 merupakan hasil perhitungan skor feature importance untuk Nusa Tenggara Timur. TPK memiliki skor feature importance tertinggi yaitu 0,35; diikuti oleh covid, ER_topik0, dan ER akun. dengan masing-masing memiliki skor feature importance sebesar 0,251; 0,235; dan 0,198. Sehingga variabel bebas TPK, covid, ER_topik0, dan ER akun akan digunakan dalam pemodelan jumlah perjalanan wisnus ke Nusa Tenggara Timur berdasarkan kombinasi variabel hasil skor *feature importance*.

Tabel 8. Hyperparameter terbaik RFR

Kombinasi Variabel	Hyperparameter	Ukuran
RFR	Max depth	128
(wisnus + TPK + Inflasi	Min samples leaf	4
+ IGT + covid + ER akun	Min samples split	2
+ ER_topik0 +	N estimators	150
ER_topik1 + ER_topik2)		
RFR	Max depth	128
(wisnus + TPK + covid +	Min samples leaf	4
ER akun + ER_topik0)	Min samples split	10
	N estimators	100
RFR	Max depth	128
(wisnus + TPK+inflasi)	Min samples leaf	4
	Min samples split	10
	N estimators	100

Sumber: Data primer yang diolah, 2024

Hyperparameter tuning pada model RFR dengan kombinasi variabel secara berurutan dapat diamati pada tabel 8. Berdasarkan hasil tersebut, kemudian akan dilakukan pemodelan RFR menggunakan hyperparameter tersebut dengan $n_past=3$. Nilai tersebut dipilih dengan mempertimbangkan bahwa pola data dalam tiga periode sebelumnya dapat memberikan kontribusi signifikan terhadap variabel target yang akan diprediksi. Evaluasi dari model RFR yang dibangun berdasarkan kombinasi variabel dan hyperparameter terbaik tersebut ditampilkan pada tabel 9.

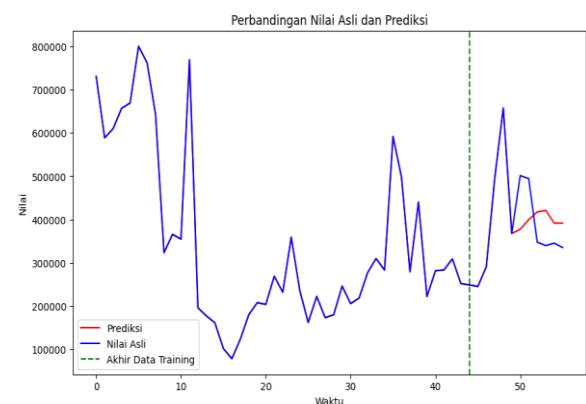
Tabel 9. Model Prediksi RFR

Model	MAPE	MAE	RMSE
RFR			
(wisnus + TPK + Inflasi + IGT + covid + ER akun + ER_topik0 + ER_topik1 + ER_topik2)	25,12	97020,46	99832,13

RFR			
(wisnus + TPK + covid + ER akun + ER_topik0)	19,68	78738,19	82820,82
RFR			
(wisnus +TPK+inflasi)	29,19	106459,82	111489,79

Sumber: Data primer yang diolah, 2024

Model RFR dengan kombinasi variabel bebas yaitu TPK, covid, ER akun, dan ER_topik0 merupakan model prediksi jumlah perjalanan wisatawan nusantara ke Nusa Tenggara Timur yang memiliki nilai MAPE, MAE, dan RMSE terkecil. Kombinasi variabel ini memiliki nilai MAPE 19,68 persen; nilai MAE 78738,19; dan nilai RMSE 82820,82. Berdasarkan nilai MAPE tersebut maka model ini termasuk dalam kategori baik dalam melakukan prediksi.



Gambar 8. Grafik Hasil Prediksi Terbaik RFR

Gambar 8 merupakan grafik model prediksi terbaik jumlah perjalanan wisnus ke Nusa Tenggara Timur menggunakan Random Forest Regression dengan kombinasi variabel bebas yaitu TPK, ER akun, dan ER_topik0. Pada grafik dapat dilihat bahwa data prediksi sudah mengikuti pola data uji dan sudah cukup berhimpitan di beberapa titik.

Tabel 10. Hyperparameter terbaik LSTM

Kombinasi Variabel	Hyperparameter	Ukuran
LSTM	Batch size	32
(wisnus + TPK +	Dropout rate	0,15
Inflasi + IGT + covid	Epoch	150
+ ER akun +	Learning rate	0,0001
ER_topik0 +	LSTM units	128
ER_topik1 +	Optimizer	adam
ER_topik2)		
LSTM	Batch size	32
(wisnus + TPK +	Dropout rate	0,25
covid + ER akun +	Epoch	150
ER_topik0)	Learning rate	0,01
	LSTM units	128
	Optimizer	adam
LSTM	Batch size	32
(wisnus + TPK +	Dropout rate	0,25
inflasi)	Epoch	150
	Learning rate	0,01
	LSTM units	256
	Optimizer	adam

Sumber: Data primer yang diolah, 2024

Hyperparameter tuning pada model LSTM dengan kombinasi variabel secara berurutan dapat diamati pada tabel 10. Berdasarkan hasil tersebut, kemudian akan dilakukan pemodelan LSTM menggunakan hyperparameter tersebut dengan $n_past=2$. Evaluasi dari model LSTM yang dibangun berdasarkan kombinasi variabel dan hyperparameter terbaik tersebut ditampilkan pada tabel 11.

Tabel 11. Model Prediksi LSTM Nusa Tenggara Timur

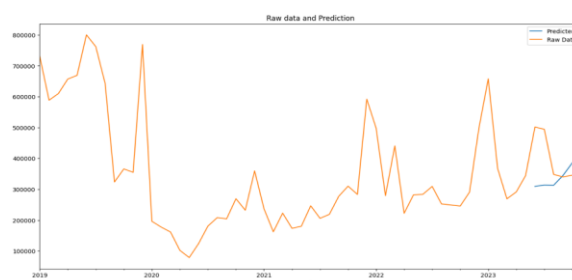
Model	MAPE	MAE	RMSE
LSTM (wisnus + TPK + Inflasi + IGT + covid + ER akun + ER_topik0 + ER_topik1 + ER_topik2)	18,75	83577,40	111630,97
LSTM (wisnus + TPK + covid + ER akun + ER_topik0)	23,02	100535,18	127757,51
LSTM (wisnus + TPK + inflasi)	29,99	116472,03	125846,09

Sumber: Data primer yang diolah, 2024

Tabel 11 merupakan tabel evaluasi setiap model LSTM dengan kombinasi variabel dan hyperparameter terbaik hasil grid search. Model LSTM dengan semua variabel bebas (TPK + Inflasi + IGT + covid + ER akun + ER_topik0 + ER_topik1 + ER_topik2) memiliki nilai evaluasi MAPE sebesar 18,75 persen; MAE sebesar 83577,40; dan RMSE sebesar 111630,97. Ketiga nilai evaluasi ini merupakan nilai terkecil jika dibandingkan dengan kelima model LSTM lainnya. Sehingga model LSTM dengan semua variabel bebas (TPK + Inflasi + IGT + covid + ER akun + ER_topik0 + ER_topik1 + ER_topik2) merupakan model LSTM terbaik dalam memprediksi jumlah perjalanan wisnus ke Nusa Tenggara Timur dan berdasarkan nilai MAPE yang dihasilkan termasuk ke dalam kategori baik.

Temuan prediktif dalam penelitian ini, yang menunjukkan kinerja model LSTM dengan nilai MAPE sebesar 18,75 persen, sejalan dengan hasil penelitian terdahulu yang membandingkan performa berbagai model pembelajaran mesin dalam konteks data runtun waktu. Dalam studi komparatif terhadap model Prophet, Random Forest, dan Long Short-Term Memory (LSTM) pada peramalan beban listrik, menemukan bahwa model LSTM secara konsisten menunjukkan close fit terhadap data aktual dan mampu menangkap fluktuasi permintaan dengan lebih akurat dibandingkan Prophet maupun Random Forest (Manandhar et al., 2024). Temuan tersebut menegaskan bahwa keunggulan LSTM terletak pada kemampuannya memodelkan dinamika temporal yang kompleks dan nonlinier, khususnya ketika pola musiman dan fluktuasi jangka pendek saling berinteraksi. Dalam konteks pariwisata Nusa Tenggara Timur, karakteristik data jumlah perjalanan wisatawan nusantara yang bersifat musiman, dipengaruhi oleh faktor ekonomi (TPK dan inflasi),

faktor digital (Indeks Google Trends dan engagement rate media sosial), serta gangguan struktural seperti pandemi COVID-19, menuntut model yang memiliki memori jangka panjang dan fleksibilitas temporal. LSTM, melalui mekanisme gating pada sel memorinya, mampu mempertahankan informasi relevan dari periode sebelumnya dan menangkap pola musiman berulang secara lebih efektif dibandingkan Random Forest yang pada dasarnya memperlakukan setiap observasi sebagai independen kecuali melalui rekayasa fitur lag. Oleh karena itu, hasil penelitian ini tidak hanya menunjukkan kinerja empiris yang baik, tetapi juga konsisten secara metodologis dengan literatur terbaru yang menempatkan LSTM sebagai model unggul dalam peramalan data runtun waktu dengan fluktuasi musiman yang kuat dan pengaruh multivariat yang kompleks.



Gambar 9. Grafik Model Prediksi Terbaik LSTM

Gambar 9 merupakan grafik hasil prediksi jumlah perjalanan wisnus ke Nusa Tenggara Timur menggunakan model terbaik dari LSTM. Dapat dilihat bahwa grafik dari data prediksi sudah mengikuti pola data uji dan sudah cukup berhimpitan.

E. SIMPULAN

Hasil pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA) terhadap caption unggahan Instagram dari akun *stakeholder* pariwisata di Provinsi Nusa Tenggara Timur mengungkap adanya tiga tema utama yang secara konsisten muncul dalam strategi komunikasi digital, yaitu keindahan wisata bahari, penawaran paket trip dan berlayar, serta penawaran trip dan sewa boat. Temuan ini menunjukkan bahwa promosi pariwisata Nusa Tenggara Timur di media sosial masih sangat didominasi oleh narasi visual dan pengalaman wisata berbasis laut, yang memang menjadi daya tarik utama kawasan ini. Dari sisi keterlibatan audiens, analisis *engagement rate* (ER) memperlihatkan bahwa nilai rata-rata ER bulanan unggahan Instagram *stakeholder* pariwisata di Nusa Tenggara Timur mencapai 28,074, mencerminkan tingkat interaksi pengguna yang relatif tinggi terhadap konten promosi pariwisata. Lebih lanjut, topik yang mengangkat keindahan wisata bahari terbukti paling efektif dalam menarik perhatian pengguna, dengan nilai rata-rata ER tertinggi yang mencapai 42,24, jauh melampaui dua topik lainnya. Hal ini mengindikasikan bahwa konten yang menonjolkan estetika alam dan pengalaman visual memiliki daya tarik emosional yang lebih kuat bagi

calon wisatawan nusantara dibandingkan konten promosi berbasis penawaran komersial semata.

Temuan deskriptif tersebut kemudian diperkuat melalui analisis prediktif jumlah perjalanan wisatawan nusantara ke Nusa Tenggara Timur. Hasil pengujian menunjukkan bahwa model *Long Short-Term Memory* (LSTM) dengan penyertaan seluruh variabel bebas meliputi Tingkat Penghunian Kamar (TPK), inflasi, Indeks Google Trends (IGT), variabel dummy Covid-19, engagement rate akun, serta engagement rate berbasis tiga topik hasil LDA merupakan model terbaik dalam memprediksi jumlah perjalanan wisatawan nusantara. Model ini mampu menangkap dinamika temporal data pariwisata serta pengaruh variabel ekonomi dan digital secara simultan, dengan kinerja yang ditunjukkan oleh nilai MAPE sebesar 18,75 persen, MAE sebesar 83.577,40, dan RMSE sebesar 111.630,97.

Hasil ini menegaskan bahwa integrasi data media sosial berbasis engagement dan topik konten dengan indikator ekonomi dan pariwisata konvensional secara signifikan meningkatkan akurasi prediksi jumlah wisatawan nusantara. Dengan demikian, penelitian ini tidak hanya menyoroti peran strategis Instagram sebagai sarana promosi pariwisata, tetapi juga menunjukkan potensinya sebagai sumber data alternatif yang relevan dan informatif dalam mendukung perencanaan dan pengambilan kebijakan pariwisata berbasis data di Provinsi Nusa Tenggara Timur.

DAFTAR PUSTAKA

- Aggarwal, C. C. (2018). *Neural networks and deep learning* (Vol. 10, Issue 978). Springer.
- Andzani, D., Virgin, D., & Setijadi, N. (2024). Peran media sosial dalam membangun citra destinasi pariwisata yang menarik. *JMBI UNSRAT (Jurnal Ilmiah Manajemen Bisnis Dan Inovasi Universitas Sam Ratulangi)*, 11(1), 188–195.
- Anggara, B., Budiman, Sumardani, R., & Martina, C. D. (2025). A Systematic Literature Review on The Economic Impact of Local Communities on High Priority Destinations in the Indonesian area. *Sadar Wisata: Jurnal Pariwisata*, 8(2 SE-Articles), 156–168. <https://doi.org/10.32528/sw.v8i2.4290>
- Azmi, A. F., & Budi, I. (2018). Exploring practices and engagement of Instagram by Indonesia Government Ministries. 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), 18–21. <https://doi.org/10.1109/ICITEED.2018.8534799>
- Bahri, A. S., & Abdilah, F. (2022). Potensi sumber daya manusia bidang pariwisata di kawasan strategis pariwisata nasional Labuan Bajo, Nusa Tenggara Timur. *Destinesia: Jurnal Hospitaliti Dan Pariwisata*, 3(2), 84–92.
- Blanco-Moreno, S., González-Fernández, A. M., Muñoz-Gallego, P. A., & Casaló, L. V. (2024). Understanding engagement with Instagram posts about tourism destinations. *Journal of Destination Marketing & Management*, 34, 100948.
- BPS Provinsi NTT. (2023). *Statistik Pariwisata Provinsi Nusa Tenggara Timur 2023*.
- Brownlee, J. (2016). *Machine learning mastery with Python: understand your data, create accurate models, and work projects end-to-end*. Machine Learning Mastery.
- Dewi, N. P. A. L. (2022). Pengaruh media sosial terhadap minat berkunjung wisatawan ke daya tarik wisata sukawana sunrise: The effect of social media on the interest of tourists visiting the attraction of sunrise variety tourism. *Jurnal Ilmiah Pariwisata Dan Bisnis*, 1(2), 276–285.
- Dianafakroh, D., Santi, I. N., Syarifuddin, U., & Evrianti, H. (2025). Peran Promosi melalui Instagram, Kualitas Layanan dan Fasilitas Wisata terhadap Minat Berkunjung Kembali Wisatawan. *Jurnal Samudra Ekonomi Dan Bisnis*, 16(3), 525–537.
- Ganur, Y., & Putra, P. G. P. (2023). Peran Media Sosial Instagram Sebagai Media Promosi Pariwisata Di Pantai Pandawa Ungasan Bali. *Jurnal Daya Tarik Wisata*, 5(2), 1–11.
- Han, J., Pei, J., & Tong, H. (2022). Data mining: concepts and techniques. Morgan kaufmann.
- Havianto, C. A., & Artiningrum, T. (2022). Pemanfaatan Media Sosial sebagai Sarana Promosi Objek Wisata Bumi Almira. *Geoplanart*, 4(2), 91–103.
- Ismarizal, B., & Kusumah, A. H. G. (2023). The Instagram effect on tourist destination choices: Unveiling key attraction elements. *Journal of Consumer Sciences*, 8(2), 124–137.
- Joseph, A., Vasundhara, T. P., & Thomas, T. (2024). How Social Media Influences Travel Decisions: The Effect of User-Generated Content, Visual Appeal, And Storytelling on Destination Intentions. *South India Journal of Social Sciences*, 22(4), 349–359.
- Kariyamin, Hamka, S. F., Sudewo, E. D. B., & Alyandi, L. O. (2025). Prediksi Jumlah Kunjungan Wisatawan Nusantara ke Sulawesi Tenggara Menggunakan Pendekatan Machine Learning. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(3 SE-Articles), 1031–1043. <https://doi.org/10.51454/decode.v5i3.1352>
- Kusuma, A., Darma, D., & Sari, S. R. (2023). Analisis pemanfaatan media sosial Instagram sebagai media promosi di perpustakaan Universitas Bangka Belitung berdasarkan engagement rate, impression dan reach. *Lentera Pustaka: Jurnal Kajian Ilmu Perpustakaan, Informasi Dan Kearsipan*, 9(2), 105–116.

- Llewellyn, M., Ross, G., & Ryan-Saha, J. (2023). COVID-era forecasting: Google trends and window and model averaging. *Annals of Tourism Research*, 103, 103660. <https://doi.org/https://doi.org/10.1016/j.annals.2023.103660>
- Manandhar, P., Rafiq, H., Rodriguez-Ubinas, E., & Palpanas, T. (2024). New Forecasting Metrics Evaluated in Prophet, Random Forest, and Long Short-Term Memory Models for Load Forecasting. In *Energies* (Vol. 17, Issue 23, p. 6131). <https://doi.org/10.3390/en17236131>
- Mariani, M., & Borghi, M. (2021). Are environmental-related online reviews more helpful? A big data analytics approach. *International Journal of Contemporary Hospitality Management*, 33(6), 2065–2090.
- Maulana, R., Dewi, S. F., & Syaifulloh, M. (2024). Efektivitas promosi melalui media sosial instagram terhadap minat berkunjung wisatawan di Agrowisata Kebun Teh Pagilaran Batang. *Journal of Management and Digital Business*, 4(1), 132–142.
- Molnar, C. (2020). Interpretable machine learning. Lulu.com.
- Murphy, P., & Murphy, A. (2004). Strategic Management for Tourism Communities: Bridging the Gaps. In *Strategic Management for Tourism Communities*. <https://doi.org/10.21832/9781873150856>
- OECD. (2022). *OECD Tourism Trends and Policies 2022*.
- Peng, L., Wang, L., Ai, X.-Y., & Zeng, Y.-R. (2021). Forecasting tourist arrivals via random forest and long short-term memory. *Cognitive Computation*, 13(1), 125–138.
- Prayoga, S. W., & Wijayanto, A. W. (2024). Optimasi Prediksi Jumlah Wisatawan Nusantara ke Provinsi Bali Melalui Big Data Analytics dengan Integrasi Google Trends dan Tingkat Penghunian Kamar Hotel. *Seminar Nasional Official Statistics*, 2024(1), 571–580.
- Qian, X. (2022). Three narrative patterns of the city image visually presented on Instagram under the influence of self-presentation. *Media, Culture & Society*, 44(6), 1149–1165.
- Rahim, F. (2012). *Pedoman Kelompok Sadar Wisata | i* (1st ed.). Kemenparekraf.
- Raifu, I. A., & Afolabi, J. A. (2026). Does rising inflation affect the tourism industry? Evidence from Nigeria. *Journal of Policy Research in Tourism, Leisure and Events*, 18(1), 167–183. <https://doi.org/10.1080/19407963.2024.2314495>
- Ramdan, M., Musadieg, M., Yulianto, E., & Pangestuti, E. (2025). The Impact of E-WOM Social Media in Influencing Destination Image and Visit Intention in the Context of Tourism Industry: An Empirical Investigation of Social Media Enterprise Instagram Followers in Malang Raya, East Java. *Formosa Journal of Multidisciplinary Research*, 4, 1257–1272. <https://doi.org/10.55927/fjmr.v4i3.117>
- Recuero-Virto, N. (2022). The Role of User-Generated Content Data for Collaborative Learning: Identifying Tourism Hot Topics During the Pandemic. In *COVID-19 and a World of Ad Hoc Geographies* (pp. 683–699). Springer.
- Rochmah, W. Y., & Lentera, C. (2023). Analisis eWOM terhadap Destinasi Wisata Super Prioritas Menggunakan Text Mining. *Indonesian Journal of Multidisciplinary on Social and Technology*, 1(3 SE-Articles), 270–276. <https://doi.org/10.31004/ijmst.v1i3.227>
- Rödel, E. (2017). Forecasting tourism demand in Amsterdam with Google Trends (A research into the forecasting potential of Google Trends for tourism demand in Amsterdam 2) (pp. 6–54).
- Salamanis, A., Xanthopoulou, G., Kehagias, D., & Tzovaras, D. (2022). LSTM-Based Deep Learning Models for Long-Term Tourism Demand Forecasting. *Electronics (Switzerland)*, 11(22). <https://doi.org/10.3390/electronics11223681>
- Saura, J. R., Reyes-Menendez, A., & Palos-Sanchez, P. R. (2020). The digital tourism business: A systematic review of essential digital marketing strategies and trends. *Digital Marketing Strategies for Tourism, Hospitality, and Airline Industries*, 1–22.
- Shahbaznezhad, H., Dolan, R., & Rashidirad, M. (2021). The Role of Social Media Content Format and Platform in Users' Engagement Behavior. *Journal of Interactive Marketing*, 53, 47–65. <https://doi.org/https://doi.org/10.1016/j.intmar.2020.05.001>
- Shang, Z., & Luo, J. M. (2023). Topic modelling for wildlife tourism online reviews: Analysis of quality factors. *Current Issues in Tourism*, 26(14), 2317–2331.
- Sugiyarto, A., & Mahagangga, I. G. A. O. (2020). Kendala Pengembangan Pariwisata di Destinasi Pariwisata Labuan Bajo Nusa Tenggara Timur. *Jurnal Destinasi Pariwisata; Vol 8 No 1 (2020): VOL 8, NO 1 (2020): (JANUARY-JUNE) Jurnal Destinasi Pariwisata*doi10.24843/JDEPAR.2020.V08.I01.P03.<https://ojs.unud.ac.id/index.php/destinasi-par/article/view/61842>
- Sun, Z., Liu, L., Pan, R., Wang, Y., & Zhang, B. (2025). Tourism and economic growth: The role of institutional quality. *International Review of Economics & Finance*, 98, 103913. <https://doi.org/https://doi.org/10.1016/j.iref.2025.103913>
- Swandeni, N. K., Indrayani, I. G. A. P. W., Trisdayanti, N. P. E., Darmaputra, P. G. E., &

- Tirtawati, N. M. (2024). Lensa Netnografi: Analisis Sentimen Untuk Peningkatan Engagement Konten Promosi Pariwisata Melalui Film “The Last Of Us 2. *Jurnal Ilmiah Pariwisata*; Vol 29 No 2 (2024): *Jurnal Ilmiah Pariwisata*.
<https://doi.org/10.30647/jip.v29i2.1798>
- Uthirapathy, S. E., & Sandanam, D. (2023). Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model. *Procedia Computer Science*, 218, 908–917.
- Wu, Doris Chenguang, Zhong, Shiteng, Wu, Ji, & Song, Haiyan. (2024). Tourism and Hospitality Forecasting With Big Data: A Systematic Review of the Literature. *Journal of Hospitality & Tourism Research*, 49(3), 615–634.
<https://doi.org/10.1177/10963480231223151>
- Zhang, Z., Xu, Z., & Yu, W. (2025). Tourism demand forecasting with multi-source data: a hybrid model incorporating seasonal-trend decomposition. *Annals of Operations Research*, 1–31.