

**OPTIMASI METODE *LONG SHORT-TERM MEMORY*
MENGUNAKAN SELEKSI FITUR *BORUTA*
UNTUK MEMPREDIKSI JUMLAH
KEDATANGAN WISATAWAN
BERBASIS GOOGLE TRENDS**

THESIS

**Oleh :
TUMIYAR JUNAIDI
NIM. 210605210008**



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**OPTIMASI METODE *LONG SHORT-TERM MEMORY* MENGGUNAKAN
SELEKSI FITUR *BORUTA* UNTUK MEMPREDIKSI
JUMLAH KEDATANGAN WISATAWAN
BERBASIS GOOGLE TRENDS**

THESIS

**Diajukan Kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh :
TUMIYAR JUNAIDI
NIM. 210605210008**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**OPTIMASI METODE *LONG SHORT-TERM MEMORY* MENGGUNAKAN
SELEKSI FITUR *BORUTA* UNTUK MEMPREDIKSI
JUMLAH KEDATANGAN WISATAWAN
BERBASIS GOOGLE TRENDS**

THESIS

**Diajukan Kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh :
TUMIYAR JUNAIDI
NIM. 210605210008**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**OPTIMASI METODE *LONG SHORT-TERM MEMORY* MENGGUNAKAN
SELEKSI FITUR *BORUTA* UNTUK MEMPREDIKSI
JUMLAH KEDATANGAN WISATAWAN
BERBASIS GOOGLE TRENDS**

THESIS

Oleh :
TUMIYAR JUNAIDI
NIM. 210605210008

Telah diperiksa dan disetujui untuk diuji:
Tanggal 4 Juni 2025

Pembimbing I,



Dr. M. Amin Hariyadi, M.T
NIP. 19670118 200501 1 001

Pembimbing II,



Dr. Zainal Abidin, M.Kom
NIP. 19760613 200501 1 004

Mengetahui,

Ketua Program Studi Magister Informatika

Fakultas Sains Dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Cahyo Crysdian

NIP. 19740424 200901 1 008

**OPTIMASI METODE *LONG SHORT-TERM MEMORY* MENGGUNAKAN
SELEKSI FITUR *BORUTA* UNTUK MEMPREDIKSI
JUMLAH KEDATANGAN WISATAWAN
BERBASIS GOOGLE TRENDS**

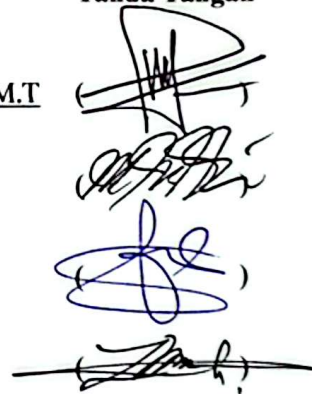
THESIS

Telah Dipertahankan di Depan Dewan Penguji Thesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal 4 Juni 2025

Susunan Dewan Penguji

Penguji I	:	<u>Dr. Ir. Yunifa Miftachul Arif, S.ST., M.T</u> NIP. 19830616 201101 1 004
Penguji II	:	<u>Dr. M. Imamudin Lc, MA</u> NIP. 19740602 200901 1 010
Pembimbing I	:	<u>Dr. M. Amin Hariyadi, M.T</u> NIP. 19670118 200501 1 001
Pembimbing II	:	<u>Dr. Zainal Abidin, M.Kom</u> NIP. 19760613 200501 1 004

Tanda Tangan



Mengetahui dan Mengesahkan,
Ketua Program Studi Magister Informatika
Fakultas Sains Dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Tumiyar Junaidi

NIM : 219695210008

Program Studi : Magister Informatika

Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri kecuali dengan mencantumkan sumber cuplikan pada daftar Pustaka. Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

ang, 2 Juni 2025
buat pernyataan,

Tumiyar Junaidi
NIM. 210605210008



MOTTO

*“Allah sudah mengatur timeline kita masing-masing dengan sempurna.
Tidak perlu risau,
nikmati saja prosesnya sambil diusahakan seoptimal mungkin
apa-apa yang harus diusahakan.”*

PERSEMBAHAN

Dengan mengucapkan syukur *Alhamdulillah*, Allah Tuhan semesta alam. Thesis ini saya persembahkan untuk :

1. Diri saya sendiri atas kerja kerasnya selama ini dan sampai nanti.
2. Seluruh keluarga dan orang-orang tercinta yang selalu memberikan do'a dan support terbaiknya.
3. Seluruh civitas akademika Universitas Islam Negeri Maulana Malik Ibrahim Malang atas bantuan dan kerjasamanya.
4. Seluruh teman-teman mahasiswa terutama Angkatan 4 Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim, kalian memang terbaik.
5. Terkhusus admin Prodi Magister Informatika, Mbak Citra, makasih banget bantuan dan supportnya. *Jazakillaahu Lak*, semoga Allah balas semua kebbaikannya mbak berlipat-lipat. Kalau capek dan ujing-uringan boleh tapi jangan lama-lama, kasihan teman-teman mahasiswanya.

KATA PENGANTAR

Assalamu 'alaikum Warrahmatullahi Wabarakatuh.

Alhamdulillah penulis haturkan ke hadirat Allah SWT yang telah memberikan limpahan Hidayah dan Rahmat-Nya, sehingga penulis dapat menyelesaikan Thesis untuk syarat memperoleh gelar Magister Komputer di Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang ini dengan baik.

Selanjutnya penulis sampaikan terimakasih dan do'a semoga Allah balas semua kebaikan kepada semua pihak yang telah mendukung dan membantu terselesaikannya Thesis ini. Ucapan ini penulis sampaikan kepada :

1. Keluarga dan orang tercinta yang senantiasa mendoakan dan memberi dukungan.
2. Bapak Dr. Ir. Mokhammad Amin Hariyadi, M.T dan Bapak Dr. Zainal Abidin, M.Kom selaku dosen pembimbing yang telah memberikan banyak pengarahan dan pengalaman berharga.
3. Mbak Citra Fidya Atmalia selaku admin prodi yang sudah sabar membantu dan mengarahkan banyak hal.
4. Segenap civitas akademika dan Bapak/Ibu dosen Program Studi Magister Informatika atas ilmu dan waktunya.
5. Semua teman-teman Angkatan dan seperjuangan yang ikut membantu dan mendukung.

Penulis menyadari dari penulisan Thesis ini masih banyak kekurangan, semoga kekurangannya bisa menjadi bahan pengembangan untuk kedepannya dan semoga ini bisa bermanfaat untuk banyak orang.

Wassalamu 'alaikum Warahmtullahi Wabarakatuh.

Malang, 20 Maret 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN.....	v
PERNYATAAN KEASLIAN TULISAN	vi
MOTTO	vii
PERSEMBAHAN.....	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
DAFTAR RUMUS	xv
ABSTRAK	xvi
ABSTRACT	xvii
ملخص.....	xviii
BAB I.....	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah.....	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	6
BAB II	7
2.1 Prediksi Kedatangan Wisatawan	7
2.2 Kerangka Teori.....	11

2.3 Long Short-Term Memory	13
2.4 Optimasi	14
2.5 Google Trend.....	14
2.6 Boruta	16
2.7 Penelitian Terdahulu	18
BAB III.....	22
3.1 Desain Penelitian.....	22
3.3.1 <i>Business Understanding</i>	23
3.1.1 <i>Analytical Approach</i>	24
3.1.2 <i>Data Requirement</i>	24
3.1.3 <i>Data collection</i>	24
3.1.4 <i>Data understanding</i>	25
3.1.5 <i>Data preparation</i>	30
3.2 Desain Model	32
3.3 Eksperimen.....	33
3.3.2 Boruta	33
3.3.3 Long Short-Term Memory	36
3.4 Evaluasi	39
3.5 Instrumen Penelitian.....	41
BAB IV	43
4.1 <i>Business Understanding</i>	43
4.1.1 <i>Problem statements and Goals</i>	44
4.1.2 Metodologi	45
4.1.3 Matrik	45
4.2 <i>Analytic Approach</i>	46
4.2.1 <i>Data Requirement</i>	47

4.3	<i>Data Collection</i>	48
4.4	<i>Data Understanding</i>	49
4.4.1	<i>Exploratory Data Analysis</i> – Deskripsi Fitur	49
4.5	<i>Data Preparation</i>	53
4.5.1	<i>Data Preprocessing</i> – Pemilihan Fitur dengan Boruta	53
4.5.2	<i>Data Preprocessing</i> – Train Test Split	57
4.5.3	<i>Data Preprocessing</i> – Normalisasi	58
4.6	<i>Modeling</i>	59
4.7	Evaluasi	61
4.8	Integrasi Islam	64
4.8.1	<i>Habluminallah</i>	64
4.8.2	<i>Habluminannas</i>	65
4.8.3	<i>Habluminal'alam / (mu'amalah)</i>	66
BAB V	69
5.1	Kesimpulan.....	69
5.2	Saran.....	70
DAFTAR PUSTAKA	72
LAMPIRAN	76

DAFTAR GAMBAR

Gambar 3. 1 Alur Proses Penelitian	23
Gambar 3. 2 Contoh data set	28
Gambar 3. 3 Flowchart Alur Penelitian dengan Boruta dan LSTM	32
Gambar 3. 4 Desain Model	33
Gambar 4. 1 Contoh normalisasi data fitur	59
Gambar 4. 2 Grafik hasil prediksi model.....	62

DAFTAR TABEL

Tabel 2. 1 Theoritical Framework for Tourist Arrival forecasting	11
Tabel 2. 2 Best performing baseline methode.....	12
Tabel 2. 3 Daftar Kata Kunci Pencarian penelitian terdahulu	14
Tabel 2. 4 Rangkuman Literature Review	21
Tabel 3. 1 Variabel dataset dan deskripsi.....	25
Tabel 4. 1 Fitur yang digunakan	49
Tabel 4. 2 daftar fitur terpilih Boruta	54
Tabel 4. 3 data split	58
Tabel 4. 4 Hasil prediksi jumlah wisatawan	60
Tabel 4. 5 Matrik hasil model	61

DAFTAR RUMUS

(3.1) Fitur bayangan	34
(3.2) Importansi	34
(3.3) Z-score	35
(3.4) Klasifikasi hasil Boruta	35
(3.5) <i>Looping</i> Boruta.....	36
(3.6) <i>Forget gate</i>	36
(3.7) <i>Input gate</i>	37
(3.8) Kandidat <i>Cell State</i>	37
(3.9) <i>Update Cell State</i>	38
(3.10) <i>Hidden state</i>	38
(3.11) <i>Output</i> LSTM	39
(3.12) RMSE	40

ABSTRAK

Junaidi, Tumiyar. 2025. **Optimasi Metode *Long Short-Term Memory* Menggunakan Seleksi Fitur Boruta untuk Memprediksi Jumlah Kedatangan Wisatawan Berbasis Google Trends**. Thesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. M. Amin Hariyadi, M.T (II) Dr. Zainal Abidin, M.Kom

Kata kunci: LSTM, Boruta, prediksi jumlah wisatawan, seleksi fitur, Google Trends.

Prediksi jumlah kedatangan wisatawan secara akurat sangat penting dalam mendukung perencanaan strategis dan pengambilan kebijakan di sektor pariwisata. Penelitian ini bertujuan untuk mengoptimalkan performa model prediksi berbasis Long Short-Term Memory (LSTM) dengan menerapkan seleksi fitur menggunakan algoritma Boruta. Tiga pendekatan dibandingkan, yaitu LSTM univariate, LSTM multivariate dengan semua fitur, dan LSTM multivariate dengan fitur hasil seleksi Boruta. Dataset yang digunakan terdiri dari data bulanan kedatangan wisatawan ke Jawa Timur Indonesia (2008–2025) dan indeks pencarian dari Google Trends. Hasil evaluasi menunjukkan bahwa model LSTM univariate menghasilkan performa terbaik dengan nilai R^2 sebesar 0.732, RMSE sebesar 3238.38, dan MAE sebesar 2737.36. Meskipun demikian, model LSTM multivariate dengan fitur Boruta menunjukkan peningkatan performa yang cukup signifikan dibandingkan pendekatan multivariate tanpa seleksi, yang berarti algoritma Boruta tetap memiliki kontribusi positif dalam konteks pengurangan fitur yang tidak relevan. Temuan ini menekankan pentingnya pemilihan fitur yang mempertimbangkan karakteristik temporal dalam peramalan deret waktu.

ABSTRACT

Junaidi, Tumiya. 2025. **Optimizing Long Short-Term Memory Forecasting with Boruta Feature Selection: A Case Study on Tourist Arrival Prediction using Google Trends**. Thesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Promotor: (I) Dr. M. Amin Hariyadi, M.T (II) Dr. Zainal Abidin, M.Kom

Keyword: LSTM, Boruta, tourist arrival forecasting, feature selection, Google Trends.

Accurate forecasting of international tourist arrivals is essential for strategic planning and decision-making in the tourism sector. This study aims to optimize the predictive performance of Long Short-Term Memory (LSTM) models by applying feature selection using the Boruta algorithm. Three approaches were compared: univariate LSTM, multivariate LSTM using all features, and multivariate LSTM with Boruta-selected features. The dataset consisted of monthly tourist arrivals to East Java Indonesia (2008–2025) and Google Trends search index data. The evaluation results show that the univariate LSTM model achieved the best performance with an R^2 of 0.732, RMSE of 3238.38, and MAE of 2737.36. However, the Boruta-based multivariate model significantly improved performance compared to the multivariate model without feature selection, indicating that Boruta remains useful for eliminating irrelevant features. These findings highlight the importance of feature selection that accounts for temporal characteristics in time-series forecasting.

ملخص

جندي، تومبار.2025. تحسين طريقة الذاكرة طويلة المدى قصيرة المدى (LSTM) باستخدام اختيار سمات بوروتا (Boruta) للتنبؤ بعدد الوافدين السياحيين بناءً على اتجاهات جوجل (Google Trends). رسالة الماجستير. برنامج ماجستير المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرفون: (١) الدكتور محمد أمين هريدي، ماجستير في الهندسة (٢) الدكتور زين العابدين، ماجستير في الحاسوب.

الكلمات المفتاحية: LSTM، Boruta، التنبؤ بعدد السياح، اختيار الميزات، اتجاهات جوجل.

يُعدّ التنبؤ الدقيق بعدد السياح الأجانب الوافدين بالغ الأهمية لدعم التخطيط الاستراتيجي وصنع السياسات في قطاع السياحة. تهدف هذه الدراسة إلى تحسين أداء نموذج التنبؤ القائم على الذاكرة طويلة المدى قصيرة المدى (LSTM) من خلال تطبيق اختيار الميزات باستخدام خوارزمية بوروتا. تُقارن ثلاثة مناهج، وهي: LSTM أحادي المتغير، و LSTM متعدد المتغيرات مع جميع الميزات، و LSTM متعدد المتغيرات مع ميزات مختارة من بوروتا. تتكون مجموعة البيانات المستخدمة من بيانات شهرية عن السياح الوافدين إلى جاوة الشرقية إندونيسيا من عام 2008 إلى عام 2025 ومؤشرات البحث من جوجل تريندز. تظهر نتائج التقييم أن نموذج LSTM أحادي المتغير يحقق أفضل أداء بقيمة R^2 تبلغ 0.732، و RMSE يبلغ 3238.38، و MAE يبلغ 2737.36. ومع ذلك، يظهر نموذج LSTM متعدد المتغيرات مع خصائص بوروتا زيادة ملحوظة في الأداء مقارنةً بالنهج متعدد المتغيرات دون اختيار، مما يعني أن خوارزمية بوروتا لا تزال تقدم مساهمة إيجابية في تقليل الخصائص غير ذات الصلة. تؤكد هذه النتيجة على أهمية اختيار الخصائص الذي يراعي الخصائص الزمنية في التنبؤ بالسلاسل الزمنية.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bepergian atau safar dalam Bahasa arab, menjadi sebuah kebutuhan untuk setiap manusia dengan berbagai macam tujuannya. Lebih dari itu safar juga dianjurkan oleh Allah di dalam Alqur'an surah Al-Mulk ayat 15 yang berbunyi:

هُوَ الَّذِي جَعَلَ لَكُمُ الْأَرْضَ ذُلُولًا فَامْشُوا فِي مَنَاكِبِهَا وَكُلُوا مِن رِّزْقِهِ ۚ وَإِلَيْهِ النُّشُورُ

Artinya : *Dialah yang menjadikan bumi untuk kamu yang mudah dijelajahi, maka jelajahilah di segala penjurunya dan makanlah sebagian dari rezeki-Nya. Dan hanya kepada-Nyalah kamu (kembali setelah) dibangkitkan.* (Q.S Al Mulk : 15)

Dalam tafsir ibnu katsir (Ishaq, 2013) bahwa maksud dari ayat tersebut adalah berjalanlah kalian ke mana pun yang kamu kehendaki di berbagai kawasannya, serta lakukanlah perjalanan mengelilingi semua daerah dan kawasannya untuk keperluan mata pencaharian dan perniagaan. Dan ketahuilah bahwa upaya kalian tidak dapat memberi manfaat sesuatu pun bagi kalian kecuali bila Allah memudahkannya bagi kalian.

Industri perjalanan terutama lintas negara menjadi bagian dari utama dari industri pariwisata terlepas apa tujuan dari kedatangan turis menuju negara lain, akan terdata sebagai wisatawan. Menariknya, industri pariwisata memiliki kontribusi penting dalam perekonomian dunia. Terhitung ikut andil 10,4% GDP atau 1,7 triliun dolar di tahun 2018 (Kulshrestha et al., 2020) yang kemudian turun hingga menjadi 3% GDP atau 74% pada tahun 2020 dibandingkan tahun sebelumnya akibat dari pandemi COVID-19 yang berarti para pihak terkait industry

pariwisata mengalami penurunan pendapatan bahkan kerugian (Subbiah & Kumar, 2022).

Meskipun GDP dalam sektor pariwisata naik menjadi 5,25% dan terus meningkat sampai hari ini, ketepatan penentuan kebijakan pemerintah dan pengembangan dalam sektor pariwisata sangatlah penting. Begitu pun penentuan strategi oleh para pelaku industri juga tak kalah penting untuk menghindari kerugian (Gunter et al., 2024).

Pendapatan sektor pariwisata sangat dipengaruhi oleh kedatangan wisatawan. Maka dari itu, menjadi sangat penting mengetahui lebih lebih awal jumlah kedatangan wisatawan bagi pemerintah dan pelaku industri untuk menentukan kebijakan dan strategi yang lebih tepat. Dan disinilah peran dari peramalan / forecasting jumlah kedatangan wisatawan atau permintaan pariwisata dibutuhkan. (Hsieh, 2021)

Memperkirakan atau memprediksi kejadian yang akan datang berdasarkan kejadian masa lalu juga dijelaskan oleh Allah di dalam Al-Qur'an yang berbunyi :

يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَلْتَنْظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ

Artinya : *Wahai orang-orang yang beriman! Bertakwalah kepada Allah dan hendaklah setiap orang memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat), dan bertakwalah kepada Allah. Sungguh, Allah Mahateliti terhadap apa yang kamu kerjakan.* (Q.S Al Hasyr : 18)

Menurut tafsir ibnu katsir (Katsir, 2003) maksud dari ayat tersebut adalah seruan untuk melaksanakan perintah dan menjauhi larangan Allah. Kemudian hisablah (hitunglah) amal perbuatan kalian yang sudah kalian perbuat sebelum dihisab oleh Allah nanti di hari perhitungan agar kalian bersiap diri dan berorientasi hari akhir karena apapun yang kalian lakukan hari ini akan kalian bawa di hari akhir

nanti. Hukum tuai-tanam berlaku dan hari akhir dipengaruhi oleh apa yang terjadi atau di lakukan di masa lampau.

Permintaan pariwisata secara garis lurus ditentukan oleh jumlah kedatangan wisatawan. Peramalan permintaan pariwisata dibedakan menjadi 3 yaitu model time series, pendekatan ekonometrik, dan kecerdasan buatan. Time series merupakan model paling populer tetapi memiliki kekurangan dalam peramalan non-linier. AI (Artificial Intelligence) / kecerdasan buatan menawarkan potensi pemecahan masalah peramalan non-linier salah satunya adalah LSTM. (Hsieh, 2021) melakukan penelitian terkait permintaan wisata dan menemukan bahwa metode terbaik yang bisa digunakan untuk peramalan permintaan wisata adalah LSTM dibandingkan dengan beberapa metode lain. LSTM tidak hanya mampu menangani perubahan tren data, tetapi juga menjelaskan dependensi data time series.

metode ini ditemukan oleh (Hochreiter & Schmidhuber, 1997). Arsitektur LSTM dirancang untuk mengatasi masalah kesalahan yang terjadi pada Recurrent Neural Network (RNN). Dengan kemampuan untuk mengolah data dalam interval waktu yang panjang dan kompleks, LSTM menawarkan solusi yang efektif untuk mengatasi keterbatasan RNN konvensional. Keefektifan LSTM ini dicapai melalui algoritma yang efisien, yang memungkinkan jaringan untuk mempelajari pola-pola kompleks dalam data dengan lebih akurat. Pada awal ditemukan, LSTM diuji untuk memprediksi data time-series, komposisi music, dan pemrosesan suara. Namun pada perkembangannya LSTM mampu digunakan untuk berbagai macam kebutuhan.

(Hsieh, 2021) menguji coba LSTM untuk peramalan kedatangan wisatawan di Taiwan dengan data non-linier atau dengan sock-data yaitu saat terjadinya SARS dan COVID-19. Hasil penelitiannya menunjukkan bahwa LSTM mampu memprediksi lebih baik dibandingkan dengan fuzzy time-series lainnya. LSTM juga mampu digunakan untuk memprediksi kedatangan wisata disaat terjadi bencana-bencana lain yang berdampak pada kepariwisataan dan bahkan LSTM tidak hanya secara spesifik untuk sebuah negara saja karena sudah dilakukan data sejenis di Amerika Serikat, Australia, Singapura dan hasilnya tetap lebih baik juga.

(Hsieh, 2021) menggunakan data ekonomi untuk pemodelan LSTM permintaan wisatawan dan menyarankan bahwa penggunaan variable tambahan seperti data cuaca dan data search engine akan mampu meningkatkan akurasi peramalan. pemilihan variable yang tepat akan menjadi arah penelitian selanjutnya.

(Kulshrestha et al., 2020) menemukan bahwa LSTM ini cocok untuk data yang memiliki ketergantungan jangka Panjang seperti permintaan wisata. Penelitian menunjukkan bahwa pemilihan parameter model atau variable model sangat mempengaruhi hasil dari peramalan. Maka dari itu diperlukanlah pemilihan parameter atau pemilihan variable agar hasil peramalan lebih optimal.

Ada beberapa algoritma seleksi variable atau lebih dikenal seleksi fitur yang bisa digunakan untuk mengoptimalkan LSTM.(Ahmed et al., 2021) melakukan studi tentang peramalan kelembaban tanah menggunakan model LSTM yang dikombinasikan dengan Boruta sebagai seleksi fitur. Dataset yang digunakan adalah dataset Global Climate Models (GCMs) RCP4.5 dan RCP8.5. hasil penelitian menunjukkan bahwa penggunaan Boruta sebagai seleksi fitur memiliki keunggulan dibandingkan dengan model tunggal LSTM, SVR , dan MARS.

Boruta merupakan metode untuk menentukan atribut yang paling berpengaruh dalam suatu model prediksi. Metode ini dikembangkan oleh (Kursa & Rudnicki, 2010) dan merupakan salah satu metode yang efektif dalam menentukan atribut yang signifikan. Boruta menggunakan prinsip kerja yang sama dengan Random Forest, yaitu membuat banyak pohon keputusan yang berbeda dan menggabungkan hasilnya untuk mendapatkan prediksi yang lebih akurat.

Dari penjelasan diatas bisa diambil hipotesa bahwa Boruta bisa digunakan sebagai variable selection untuk meningkatkan performa forecasting kedatangan wisatawan menggunakan Long-Short-Term Memory.

1.2 Pernyataan Masalah

1. Seberapa tingkat akurasi dan hasil Analisa pendekatan LSTM menggunakan optimasi Boruta untuk melakukan prediksi jumlah kedatangan wisatawan berbasis google trends?

1.3 Tujuan Penelitian

1. Mengukur akurasi dan hasil analisa pendekatan LSTM menggunakan optimasi Boruta untuk melakukan prediksi jumlah kedatangan wisatawan berbasis google trends.

1.4 Manfaat Penelitian

1. Membantu pemerintah dan pihak terkait dalam menentukan kebijakan daerah dalam pembangunan atau pengembangan industry wisata.
2. Membantu para pelaku industry wisata dalam mengoptimalkan kebutuhan industri.

1.5 Batasan Masalah

1. Dikhususkan untuk pemodelan regresi tanpa cakupan klasifikasi.
2. Akurasi dijadikan indikator dalam mengevaluasi performa model.
3. Sumber utama dalam menentukan jumlah wisatawan berasal dari data resmi BPSBadan Pusat Statistik
4. Data yang digunakan adalah data kedatangan wisatawan dari web Badan Pusat Statistik
5. Hanya sampai tahap evaluasi dan analisa.

BAB II

LITERATURE REVIEW

2.1 Prediksi Kedatangan Wisatawan

(Hsieh, 2021) melakukan studi untuk meningkatkan prediksi Tourism Demand (TD) dengan melakukan pendekatan deep learning menggunakan 3 metode yaitu Long Short-Term Memory (LSTM) , Bidirectional LSTM (Bi-LSTM), dan Gated Recurrent Unit networks (GRU). Dari 3 metode basic tersebut dioptimasi menggunakan Adam Optimization. Kebutuhan akan prediksi TD masih sangat luas. Ketidak-reliabilitas metode-metode tradisional membuat prediksi TD masih menjadi sebuah tantangan. Dataset yang digunakan adalah jumlah kedatangan turis bulanan yang mengunjungi Taiwan. Berdasarkan hasil pengujian, metode LSTM menghasilkan nilai RMSE sebesar 31.182, yang merupakan nilai terendah dibandingkan metode Bi-LSTM (34.872) dan GRU (34.770). Dengan demikian, LSTM menunjukkan kinerja prediksi yang paling akurat di antara ketiga pendekatan yang diuji dalam penelitian ini.

(Peng et al., 2021) melakukan studi tentang prediksi kedatangan wisatawan menggunakan pendekatan deep learning hybrid Random Forest-Differential Evolution-Long Short Term Memory (RF-DE-LSTM). Konsep yang digunakan pada dasarnya menggunakan indeks mesin pencari untuk dihubungkan dengan data kedatangan wisatawan bulanan sebagai tujuan peramalan. Random forest digunakan untuk menemukan menyeleksi kata kunci paling relevan, Differential evolution digunakan untuk memilih lag length dari masing-masing query dan histori kedatangan wisatawan untuk merekonstruksi input peramalan,

dan Long Short-Term Memory digunakan untuk pemodelan hubungan non-linier antara kedatangan wisatawan dengan data indeks query pencarian. Berdasarkan hasil evaluasi model, metode ARIMA memperoleh nilai MAPE sebesar 38,92%, MSE sebesar 884.860, MAE sebesar 668,51, dan RSMPE sebesar 63,56%. Sementara itu, metode VAR menunjukkan kinerja yang lebih baik dengan MAPE sebesar 15,86%, MSE 147.520, MAE 341,81, dan RSMPE 19,17%. Adapun model BPNN mencatat nilai MAPE 3,54%, MSE 27.458,85, MAE 108, dan RSMPE 4,93%. Di antara seluruh metode yang diuji, model hibrida RF-DE-LSTM menghasilkan performa terbaik dengan MAPE terendah sebesar 2,14%, MSE 3.108,95, MAE 45,23, dan RSMPE 2,80%. Hal ini menunjukkan bahwa pendekatan RF-DE-LSTM mampu memberikan akurasi prediksi tertinggi dibandingkan metode lainnya.

(Law et al., 2019) melakukan penelitian tentang kedatangan wisatawan menggunakan pendekatan Deep Learning Method (DLM). Dataset yang digunakan adalah data kedatangan wisatawan bulanan Macau dan Search Intensity Indices (SII). Hasil dari penelitian menunjukkan hasil DLM memiliki nilai MAPE 1,467, MAE 37397, RMSE 48500, ANN memiliki nilai MAPE 14,307, MAE 315487, RMSE 360245, SVR memiliki nilai MAPE 6,482, MAE 162121, RMSE 211713, ARIMA memiliki nilai MAPE 6,468, MAE 161316, RMSE 200187, Naïve memiliki nilai MAPE 5,156, MAE 134313, RMSE 220436. Metode unggulan dengan nilai MAPE, MAE, dan RMSE terendah adalah DLM.

(Yu & Chen, 2022) melakukan studi tentang prediksi permintaan wisata dengan membandingkan model dasar Long Short-Term Memory (LSTM) dengan optimasi Stacked AutoEncoder LSTM (SAE-LSTM). Dataset yang digunakan

adalah data jumlah wisatawan dan data pencarian bulanan pada mesin pencari. Hasil dari penelitian menunjukkan hasil dari LSTM memiliki nilai MAPE 3,82%, MAE 110311, RMSE 152685, dan SAE-LSTM memiliki nilai MAPE 2,91%, MAE 82523, RMSE 127551. Metode unggulan dengan nilai MAPE, MAE, dan RMSE terendah adalah SAE-LSTM.

(Gunter et al., 2024) melakukan studi tentang prediksi Tourism Imports (IMP) dan GDP pasca krisis pandemic COVID-19 menggunakan Full Modified Ordinary Least Squares (FMOLS). Dataset yang digunakan adalah data Tourism Imports dan GDP dari 10 negara Eropa yang memegang peran penting dalam traveling internasional di Eropa (Austria, Belgia, Republik Ceko, Prancis, Jerman, Italia, Belanda, Polandia, Spanyol, dan Swedia).

(Sun et al., 2019) melakukan penelitian tentang peramalan kedatangan wisatawan menggunakan model Kernel Extreme Learning Machine (KELM). Dataset yang digunakan adalah data time series kedatangan wisatawan dan indeks mesin pencari. Ada dua mesin yang digunakan yaitu Baidu dan Google. Baidu memegang dominasi market share mesin pencari sebesar 80,5% di Tiongkok sehingga digunakan untuk menangani data domestic sedangkan Indeks Google digunakan untuk menangani data internasional karena memegang market share mesin pencari global sebesar 92,5%. Hasil penelitian ditinjau dari performa menunjukkan hasil KELM memiliki nilai MAPE 0,814% NRMSE 0,973, LSSVR memiliki nilai MAPE 1,426 NRMSE 1,678, SVR memiliki nilai MAPE 1,732 NRMSE 1,914, ANN memiliki nilai MAPE 1,698 NRMSE 1,783. Ditinjau dari robustness KELM memiliki nilai Std. of MAPE 0,0016 Std. of NRMSE 0,0035, LSSVR memiliki nilai Std. of MAPE 0,0019 Std. of NRMSE 0,0067, SVR

memiliki nilai Std. Metode dengan performa terbaik yang memiliki nilai MAPE dan NRMSE terendah adalah KELM. Metode dengan robustness terbaik yang memiliki Standard of MAPE dan Standard of NRMSE terendah adalah KELM.

(Xie et al., 2021) melakukan studi tentang peramalan permintaan wisata di bidang kapal pesiar menggunakan Least Square Support Vector Regression-Gravitational Search Algoritm (LSSVR-GSA). Penelitian ini menjadi sangat penting karena lebih dari 10 tahun terakhir pengembangan secara eksponensial tetapi tingkat permintaan wisata kapal pesiar malah menurun. Ada peningkatan resiko keuangan untuk inverstasi Pelabuhan sandar, kapal pesiar, serta promosi jika tidak dilakukan pengambilan kebijakan secara tepat. Dataset yang dipakai adalah data time series permintaan wisata kapal pesiar, indeks ekonomi (Purchasing Manager Index, Consumer Confidence Index, Real Effective Exchange Rate Index), dan Big Data (Search Query Data Google dan Baidu). Hasil penelitian ditinjau dari performa menunjukkan hasil ARIMAX memiliki nilai RMSE 30493,220 MAPE 5,596 WIA 0,9571 , BPNN memiliki nilai RMSE 26661,749 MAPE 4,783 WIA 0,9673, RBF memiliki nilai RMSE 25437,250, MAPE 4,593, WIA 0,9697, LSSVR memiliki nilai RMSE 20004,678, MAPE 4,376 WIA 0,9819, LSSVR-GSA memiliki nilai RMSE 14958,891, MAPE 3,311, WIA 0,9896. Sedangkan ditinjau dari robustness ARIMAX memiliki nilai Std. of RMSE 8,0510 Std. of MAPE 0,0026 Std. of WIA 0,0005, BPNN memiliki nilai Std. of RMSE 7,6006 Std. of MAPE 0,0016 Std. of WIA 0,0003, RBF memiliki nilai Std. of RMSE 7,4205 Std. of MAPE 0,0016 Std. of WIA 0,0003, LSSVR memiliki nilai Std. of RMSE 4,1820 Std. of MAPE 0,0010 Std. of WIA 0,0002, LSSVR-GSA memiliki nilai Std. of RMSE 2,3867 Std. of MAPE 0,0001 Std. of

WIA 0,0000. Metode unggulan ditinjau dari performa dengan RMSE dan MAPE terendah adalah LSSVR-GSA. Metode unggulan ditinjau dari robustness dengan Standard of RMSE, Standard of MAPE, dan Standard of WIA terendah adalah LSSVR-GSA.

2.2 Kerangka Teori

Bagian ini menguraikan sejumlah landasan teori yang menjadi pijakan dalam penelitian terkait prediksi jumlah kunjungan wisatawan. Berbagai metode yang dinilai relevan dan memiliki performa terbaik akan dibahas sebagai pertimbangan dalam pemilihan pendekatan prediktif yang digunakan. Adapun teori-teori yang dijelaskan dalam sub-bab ini bertujuan untuk memberikan pemahaman konseptual dalam merancang dan mengimplementasikan model prediksi kunjungan wisatawan.

Tabel 2. 1 Kerangka teori prediksi jumlah kedatangan wisatawan

Input	Proses			Output
<ul style="list-style-type: none"> Kedatangan wisatawan Fitur kedatangan wisatawan 	Peneliti	Model	Dataset	Prediksi kedatangan wisatawan
	Kulshrestha et al, 2020	LSTM	tourist arrival from five major country (Australia, France,	
	Hsieh, 2021	LSTM	tourists volume	
	Gunter et al., 2024	FMOLS	Real Import Tourism (IMP) , Real GDP (RGDP)	
	Bouhaddour et al., 2024	LSTM	tourist arrival in Singapore	
	Jiang et al., 2021	ECM	tourist arrival, GDP Index, Consumer Price Index, exchange rate.	
	Sun et al., 2019	KELM	monthly Beijing tourists	
	Bufalo dan Orlando, 2024	CIR#	nights spent tourist accomodation establishments	
	Xie et al., 2021	LSSVR	Monthly Chinese cruise tourist volumes, Purchasing Manager Index, Consumer Confidence Index, Real Effective Exchange Rate Index, Baidu Search Query	
	Peng et al, 2021	LSTM	tourists volume, Baidu search index	

Variabel input memiliki peran penting dalam menentukan performa algoritma *machine learning*. Dalam konteks ini, data yang digunakan mencakup jumlah kedatangan wisatawan beserta faktor-faktor yang memengaruhinya. Terdapat beragam algoritma yang dapat diterapkan untuk membangun model prediktif yang optimal. Oleh karena itu, penulis melakukan studi literatur dari sejumlah publikasi yang membahas pemodelan prediksi kedatangan wisatawan. Beberapa algoritma yang diperoleh kemudian direview dan dianalisis berdasarkan tingkat akurasi dan besarnya nilai error yang dihasilkan. Hasil tinjauan tersebut disajikan dalam bentuk tabel sebagai berikut.

Tabel 2. 2 Performa terbaik metode *baseline*

Penulis	Metode	Matrik	
(Kulshrestha et al., 2020)	LSTM	Error	0,0523%
(Hsieh, 2021)	LSTM	Error	0,0584%
(Gunter et al., 2024)	FMOLS	Error	0,0682%
(Bouhaddour et al., 2024)	LSTM	Error	0,7900%
(Jiang et al., 2021)	ECM	Error	0,8340%
(Sun et al., 2019)	KELM	Error	0,9730%
(Bufalo & Orlando, 2024)	CIR#	Error	2,2100%
(Xie et al., 2021)	LSSVR	Error	2,3867%
(Peng et al., 2021)	LSTM	Error	2,8000%

Berdasarkan data pada tabel, model LSTM sebagai baseline menunjukkan tingkat kesalahan terendah, yaitu sebesar 0,0523%. Hasil ini sejalan dengan berbagai studi sebelumnya yang menunjukkan bahwa LSTM merupakan salah satu metode yang paling banyak digunakan dalam prediksi deret waktu, serta memiliki performa yang relatif lebih unggul dibandingkan metode lainnya..

2.3 Long Short-Term Memory

Long Short-Term Memory (LSTM) adalah suatu jenis jaringan saraf tiruan yang dirancang untuk mengolah data time series dengan lebih efektif. LSTM memiliki kemampuan untuk mengingat informasi dalam jangka panjang dan memperbarui informasi tersebut secara dinamis. Hal ini memungkinkan LSTM untuk mengatasi masalah vanishing gradient yang sering terjadi dalam jaringan saraf tiruan tradisional.

Pendekatan pemodelan berbasis deep learning dalam penelitian ini juga didasarkan pada efektivitas metode yang telah diujicobakan (Arif, Putra, et al., 2023) yang mengembangkan sistem rekomendasi multi-kriteria berbasis *6As Tourism Destination Framework* untuk destinasi wisata. Model ini berhasil mengintegrasikan berbagai aspek destinasi dan memandu wisatawan secara cerdas. Sedangkan dalam konteks input minim, (Arif, Wardani, et al., 2023) menggunakan model ANN untuk membangun sistem rekomendasi wisata tanpa data rating pengguna dan tetap mampu memberikan prediksi akurat. Kedua studi ini memperkuat alasan pemilihan arsitektur LSTM dalam penelitian ini sebagai algoritma yang fleksibel baik dalam skenario data multivariat maupun univariat

LSTM bekerja dengan menggunakan tiga jenis gerbang yang digunakan untuk mengontrol aliran informasi dalam jaringan. Gerbang-gerbang tersebut adalah gerbang input, gerbang keluaran, dan gerbang lupa. Dengan menggunakan gerbang-gerbang tersebut, LSTM dapat memilih informasi mana yang akan diingat dan informasi mana yang akan dilupakan. Dengan demikian, LSTM dapat meningkatkan kemampuan jaringan saraf tiruan dalam mengolah data time series dan membuat prediksi yang lebih akurat.

2.4 Optimasi

Optimasi merupakan suatu proses sistematis untuk memperoleh solusi terbaik atau kondisi yang paling optimal dalam suatu konteks permasalahan tertentu. Proses ini umumnya melibatkan penerapan algoritma dan teknik matematis guna meningkatkan efisiensi, efektivitas, maupun kinerja suatu sistem. Konsep optimasi digunakan secara luas dalam berbagai bidang, termasuk di antaranya optimasi matematis, optimasi numerik, optimasi kombinatorial, hingga optimasi stokastik, tergantung pada karakteristik masalah yang dihadapi dan pendekatan yang digunakan.

2.5 Google Trend

Google trends adalah data dan alat analisis yang disediakan oleh Google untuk memantau dan menganalisis tren pencarian online. Daftar kata kunci pencarian yang dimasukkan oleh para pengguna dikumpulkan, dianalisis, kemudian disajikan kembali oleh google dalam bentuk grafik time series, chart, dan table sehingga memudahkan pengguna untuk memahami tren dan pola pencarian. Penulis merangkum beberapa penelitian yang menggunakan google trends sebagai acuan untuk pemilihan kata kunci indeks pencarian google. Dan berikut table hasil rangkuman penelitian tersebut.

Tabel 2. 3 Daftar Kata Kunci Pencarian penelitian terdahulu

Penulis	Sumber	Topik	Kata Kunci
(Andariesta & Wasesa, 2022)	Google Trends	main entry point international travel requirement	ngurah rai international airport soekarno-hatta international airport batam ferry terminal bali jakarta passport indonesia

		tourism planning	visa indonesia indonesia hotel indonesia resort indonesia restaurant indonesia travel
(Dinis et al., 2019)	Google trends	travel	travel destination country
(Nagy et al., 2022)	Google Mobility	Retail and recreation Grocery and Pharmacy Parks transit Stations workplace residence	restaurants cafes shopping centers theme parks museums libraries movie theaters grocery markets, food warehouses, farmers markets specialty food shops, drug stores, pharmacies local parks, national parks, public beaches, marinas dog parks, plazas, public gardens public transport hubs, subway, bus, train stations places for work places of residence
(Botha & Saayman, 2024)	google trends		flights, accomodation, travel, costs, holiday south african safaris, cape town, johansburg game reserves , the kruger national park
(Haqiq & Pharmasetiawan, 2019)	Google Trends	tourism dining traffic lodging shopping weather	bali indonesia, indonesia package, indonesia package tour indonesia currency, indonesia visa, visa for indonesia, visa to indonesia, indonesia time indonesia food, time in indonesia, fried rice indonesia flight, flights to indonesia, indonesia temple, indonesia beach, bali beach, borobudur, travel to indonesia, travel to bali hotel jakarta kuta shopping, lotte shopping weather in indonesia, bali weather, indonesia weather weather in bali
(Sun et al., 2019)	Google Trends	tourism	china travel, beijing weather, peking duck, duck recipes, beijing hotels, beijing restaurants, beijing shopping,

			zhongguancun, beijing travel, great wall, beijing flights, beijing airports, beijing railways, beijing maps, beijing bars, beijing shows
(Nie et al., 2024)	Google Trends	tourism	macao travel, macao hotel, macao, macao airport, macao acomodation
(Volchek et al., 2019)	google trends	travel	british museum, national gallery, natural history museum, tate moder, science museum

2.6 Boruta

Boruta merupakan metode untuk menentukan atribut yang paling berpengaruh dalam suatu model prediksi. Metode ini dikembangkan oleh (Kursa & Rudnicki, 2010) dan merupakan salah satu metode yang efektif dalam menentukan atribut yang signifikan. Boruta menggunakan prinsip kerja yang sama dengan Random Forest, yaitu membuat banyak pohon keputusan yang berbeda dan menggabungkan hasilnya untuk mendapatkan prediksi yang lebih akurat.

Dalam metode Boruta, atribut yang dianggap signifikan adalah atribut yang memiliki nilai impor yang lebih tinggi daripada atribut bayangan. Atribut bayangan adalah atribut yang dibuat secara acak dan tidak memiliki hubungan dengan variabel target. Dengan membandingkan nilai impor atribut asli dengan atribut bayangan, Boruta dapat menentukan atribut mana yang paling berpengaruh dalam suatu model prediksi. Metode ini telah banyak digunakan dalam berbagai bidang, termasuk biologi, ekonomi, dan psikologi, untuk menentukan atribut yang signifikan dan meningkatkan akurasi model prediksi.

Boruta merupakan pengembangan dari random forest. Hal utama yang membedakan antara Boruta dan random forest adalah pembuatan atribut atau variable bayangan.

Berikut Langkah-langkah yang dilakukan pada proses Boruta :

1. Pembuatan Atribut Bayangan

Proses Boruta dimulai dengan pembuatan atribut bayangan yang dibuat secara acak dari atribut asli. Atribut bayangan ini memiliki distribusi yang sama dengan atribut asli, tetapi tidak memiliki hubungan dengan variabel target. Tujuan pembuatan atribut bayangan ini adalah untuk membandingkan nilai impor atribut asli dengan nilai impor atribut bayangan.

2. Penggabungan Atribut Asli dan Atribut Bayangan

Atribut asli dan atribut bayangan kemudian digabungkan dalam satu dataset. Dataset ini digunakan sebagai input untuk proses Random Forest.

3. Proses Random Forest

Random Forest digunakan untuk memprediksi variabel target menggunakan dataset yang telah digabungkan. Proses ini melibatkan pembuatan banyak pohon keputusan yang berbeda dan menggabungkan hasilnya untuk memperoleh prediksi yang lebih akurat.

4. Penghitungan Nilai Import

Setelah proses Random Forest selesai, nilai impor dihitung untuk setiap atribut. Nilai impor ini menunjukkan seberapa besar kontribusi atribut tersebut dalam memprediksi variabel target.

5. Perbandingan Nilai Impor Atribut Asli dan Atribut Bayangan

Nilai impor atribut asli kemudian dibandingkan dengan nilai impor atribut bayangan. Atribut yang memiliki nilai impor yang lebih tinggi daripada atribut bayangan dianggap signifikan dan dipertahankan.

6. Pengulangan Proses

Proses Boruta diulangi beberapa kali untuk memperoleh nilai impor yang stabil untuk setiap atribut. Hal ini bertujuan untuk memastikan bahwa hasil yang diperoleh tidak dipengaruhi oleh kebetulan atau kesalahan acak.

2.7 Penelitian Terdahulu

Penelitian yang dilakukan (Ahmed et al., 2021) membahas penerapan metode Long Short Term Memory (LSTM) dalam memprediksi kelembaban tanah, yang dikombinasikan dengan algoritma seleksi fitur Boruta untuk meningkatkan akurasi model dan efisiensi pemrosesan data. Boruta diadaptasi untuk mengekstraksi fitur-fitur yang relevan dalam variable hidro-meteorologi yang secara intrinsik terkait dengan kelembaban tanah yang nantinya akan dimasukkan kedalam pemodelan LSTM untuk memprediksi kelembaban tanah dalam scenario perubahan iklim global. Model ini kemudian digunakan untuk memperkirakan kelembaban tanah di masa depan menggunakan data dari Coupled Model Intercomparison Phase-5 (CMIP5). Hasil penelitian menunjukkan bahwa model Boruta-LSTM dapat secara signifikan mengungguli model basic lainnya dalam memperkirakan kelembaban tanah di masa depan. Model ini juga menunjukkan superioritas dalam estimasi kelembaban tanah dengan lebih dari

95% dari semua kesalahan prediksi yang berada di bawah 0,02 mm dan RMSE 1,06%.

(Karbasi et al., 2024) melakukan penelitian dengan dua pendekatan deep learning yaitu Long Short Term Memory (LSTM) dan Bidirectional Long Short-Term Memory (BiLSTM), dibandingkan dengan Adaptive Boosting dan General Neural Network untuk memprediksi penguapan panik (pan evaporation) di dua stasiun iklim kering di Iran (Ahvaz dan Yazd). Data time series dari data meteorologi dan data penguapan panci digunakan sebagai input untuk pemodelan machine learning. Dua metode seleksi fitur digunakan dalam penelitian ini yaitu Boruta dan XGBoost, digunakan untuk memilih input paling relevan dan mengurangi kompleksitas model. Penelitian ini menunjukkan bahwa kombinasi Boruta dan BiLSTM prediksi yang lebih akurat di kedua stasiun yaitu RMSE 1,6857 untuk stasiun Ahvaz, dan RMSE 1,3996 untuk stasiun Yazd.

(Thu et al., 2023) melakukan studi dengan mengusulkan model jaringan saraf tiruan hybrid yang menggabungkan Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM) dengan seleksi fitur Boruta untuk memprediksi kecepatan angin di Swiss. Pertama, fitur-fitur penting diantara parameter meteorologi yang sangat mempengaruhi kecepatan angin ditentukan menggunakan Boruta. Kemudian, fitur-fitur tersebut dimasukkan ke dalam model CNN-LSTM untuk memprediksi kecepatan angin. Akhirnya, kinerja model hybrid ini dibandingkan dengan 5 model lain, yaitu model tunggal CNN, LSTM dengan atau tanpa Boruta, dan model hybrid CNN-LSTM tanpa Boruta. Hasil penelitian ini menunjukkan bahwa model yang diusulkan memiliki akurasi dan

efektifitas tertinggi disbanding model lain dengan MAPE sebesar 10,01% dan RMSE sebesar 1,23 km/jam.

(Ahmed et al., 2021) menerapkan metode seleksi fitur berbasis dua machine learning yaitu Long Short-Term Memory (LSTM) dan Gated Recurrent Unit (GRU) untuk meningkatkan kemampuan peramalan tingkat aliran air sungai di enam stasiun pengamatan di Murray Darling Basin, Australia. Penelitian ini menggabungkan memori lag yang memiliki pengaruh signifikan dari indeks mode iklim, curah hujan, dan beberapa factor lain dalam dataset bulanan. Algoritma seleksi fitur Boruta kemudian diterapkan dalam dua fase sebelum dan sesudah memperoleh input lag yang signifikan untuk menyaring variable predictor yang optimum. Kemudian hasil dari masing-masing penggunaan model dibandingkan dimulai dari model tunggal LSTM, model tunggal GRU, model LSTM menggunakan Boruta, dan model GRU menggunakan Boruta. Hasil penelitian menunjukkan bahwa model Boruta-LSTM merupakan yang paling unggul dengan mencapai akurasi lebih dari 98% dengan RRMSE 1,3% dan RMAE 0,882% mengungguli semua model lain.

(Subbiah et al., 2023) melakukan studi terkait peningkatan kinerja peramalan kecepatan angin dengan pemodelan Bi-directional Long Short-Term Memory (BiLSTM) menggunakan seleksi fitur Boruta. BiLSTM digunakan untuk menangani ketidakpastian dan non-linieritas menggunakan data set time series kecepatan angin, sedangkan seleksi fitur Boruta digunakan untuk menangani over-dimensi dan overfitting dengan mengidentifikasi fitur yang relevan untuk peramalan kecepatan angin dari fitur meteorologi. Kemudian pemodelan hybrid tersebut dibandingkan dengan model tunggal Multilayer Perceptron (MLP), MLP

dengan Boruta, dan model tunggal LSTM. Hasil penelitian ini menunjukkan bahwa Boruta-BiLSTM mengungguli model lain dengan menghasilkan RMSE sebesar 0,784, MAE sebesar 0,530, MSE sebesar 0,615, dan R2 sebesar 0,8766.

Berdasarkan sejumlah studi terdahulu yang mengimplementasikan metode Long Short-Term Memory (LSTM) dan algoritma Boruta, berikut disajikan tabel yang merangkum hasil kajian literatur tersebut sebagai referensi untuk mendukung penelitian ini..

Tabel 2. 4 Literature Review

Penulis	Dataset	Matrik Penilaian	Nilai
(Ahmed et al., 2021)	Coupled Model Inter-comparison Project 5 (CMIP5)	RMSE	1,06
(Karbasi et al., 2024)	Temperatur min-max, kecepatan angin, Realtive Humidity (RH), Solar Radiation, Pan Evaporation	RMSE	1,3996
(Thu et al., 2023)	Kecepatan angin, temperature	MAPE RMSE	10,01% 1,23 km/jam
(Masrur Ahmed et al., 2021)	Streamflow Water Level (SWL), curah hujan, periodicity, synoptic-scale climate mode indices.	RRMSE RMAE	1,30% 0,882%
(Subbiah et al., 2023)	Data time series kecepatan angin, temperature, radiasi, presiptasi, geospasial, dan agrokultural.	RMSE MAE MSE R2	0,784 0,530 0,615 0,8766

BAB III

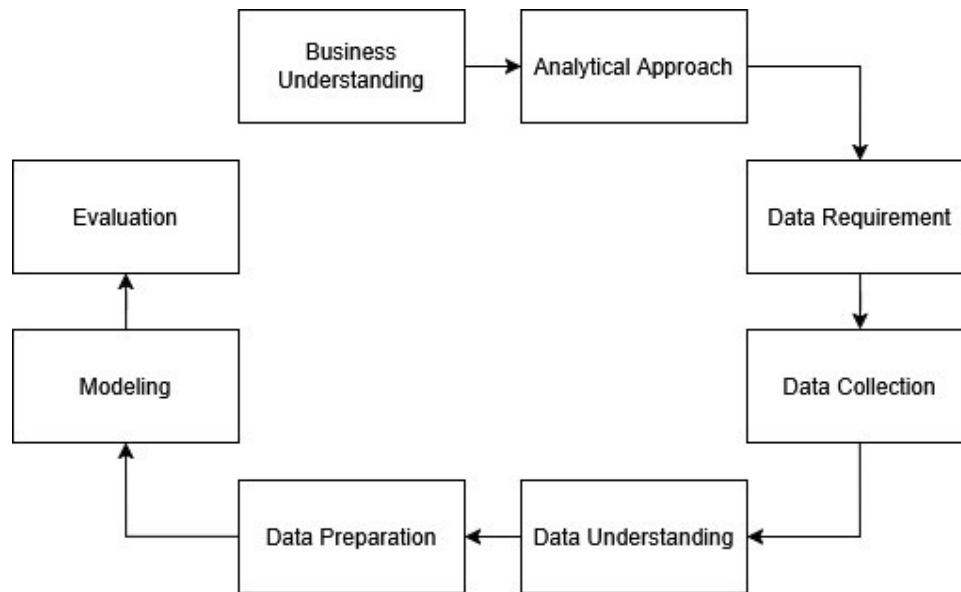
METODOLOGI PENELITIAN

3.1 Desain Penelitian

Desain penelitian ini menjelaskan tahapan-tahapan kerja sistem yang dikembangkan dalam studi ini. Proses dimulai dari tahap *business understanding*, diikuti dengan *analytic approach*, identifikasi kebutuhan data (*data requirements*), dan pengumpulan data (*data collection*). Pada tahap ini, data jumlah kedatangan wisatawan Indonesia diperoleh dari sumber resmi yaitu Badan Pusat Statistik (BPS), yang kemudian dilanjutkan dengan proses pemahaman data (*data understanding*).

Tahap selanjutnya adalah *data preparation*, di mana teknik seleksi fitur menggunakan algoritma Boruta diterapkan untuk meningkatkan efisiensi dan relevansi input terhadap model. Setelah data siap, tahap modeling dilakukan dengan membangun model prediksi menggunakan algoritma Long Short-Term Memory (LSTM). Model yang dibangun kemudian dievaluasi pada tahap evaluation melalui pengukuran performa dan analisis hasil.

Penelitian ini dibatasi hanya sampai pada tahap evaluasi model dan tidak mencakup tahap implementasi penuh. Alur keseluruhan proses penelitian ini disajikan pada Gambar 3.1.



Gambar 3. 1 Alur Penelitian

3.3.1) *Business Understanding*

Tahap *business understanding* bertujuan untuk mendefinisikan secara jelas permasalahan utama yang hendak diselesaikan melalui proses data mining. Pada fase ini, perlu diidentifikasi terlebih dahulu permasalahan utama dalam konteks prediksi jumlah kedatangan wisatawan, serta menentukan tujuan spesifik yang ingin dicapai dari pemodelan yang dikembangkan.

Selain pemahaman terhadap masalah dan tujuan, tahapan ini juga mencakup penentuan metode yang akan digunakan dalam proses analisis data. Pemilihan metode dilakukan berdasarkan relevansi dengan karakteristik data dan tujuan prediksi. Selanjutnya, dilakukan pula identifikasi terhadap metrik evaluasi yang akan digunakan untuk mengukur kinerja model, seperti seberapa akurat model dalam memprediksi jumlah kunjungan wisatawan berdasarkan data historis. Metrik-metrik tersebut nantinya akan menjadi dasar dalam menilai efektivitas dan keandalan model yang dibangun.

3.1.1) *Analytical Approach*

Setelah perumusan masalah bisnis dilakukan secara komprehensif, tahap selanjutnya adalah menentukan pendekatan analitis yang sesuai untuk menyelesaikan permasalahan tersebut. Langkah ini mencakup penerjemahan permasalahan ke dalam kerangka statistik dan pembelajaran mesin (*machine learning*), sehingga dapat dirumuskan secara teknis dan terukur.

Pendekatan analitis ini berperan penting dalam mengarahkan pemilihan model yang tepat, serta menentukan strategi prediktif yang paling efektif dalam menjawab pertanyaan penelitian. Dengan menggunakan pendekatan yang tepat, proses pemodelan akan lebih terfokus dan relevan terhadap tujuan akhir yang ingin dicapai.

3.1.2) *Data Requirement*

Tahap ini bertujuan untuk mengidentifikasi kebutuhan data yang akan digunakan dalam penelitian, baik dari segi konten, format, maupun sumbernya. Informasi yang dikumpulkan pada tahap ini mencakup jenis data yang relevan, struktur data yang dibutuhkan, serta asal data yang akan dijadikan dasar dalam proses analisis dan pemodelan. Identifikasi ini dilakukan untuk memastikan bahwa data yang diperoleh memenuhi syarat dalam menjawab permasalahan penelitian secara komprehensif dan dapat diolah secara efisien dalam tahapan selanjutnya.

3.1.3) *Data collection*

ada tahap ini, dilakukan identifikasi terhadap sumber data yang relevan dan tersedia untuk mendukung penyelesaian permasalahan penelitian. Proses pengumpulan data dapat dilakukan melalui pengambilan data dari situs resmi atau

melalui pemanfaatan repositori yang telah menyediakan kumpulan data siap pakai. Umumnya, data yang diperoleh tersedia dalam format terstruktur seperti CSV atau Excel, sehingga memudahkan dalam proses integrasi dan pengolahan pada tahap analisis selanjutnya.

3.1.4) *Data understanding*

Tahap ini bertujuan untuk memperoleh pemahaman yang lebih mendalam terhadap karakteristik data yang akan digunakan dalam proses pelatihan model. Melalui tahap ini, peneliti dapat mengevaluasi struktur informasi dalam dataset, mengidentifikasi potensi masalah seperti ketidakseimbangan data, data hilang, atau inkonsistensi, serta menilai kualitas data secara keseluruhan.

Dalam penelitian ini, data yang digunakan diperoleh dari sumber publik, yaitu Google Trends dan Badan Pusat Statistik (BPS). Dataset yang dikumpulkan terdiri dari 74 entri (baris) dengan 4 atribut (kolom). Penjelasan mengenai setiap fitur, jenis, dan tipe data disajikan pada tabel berikut.

Tabel 3. 1 rencana fitur dataset

No	fitur	Jenis	Tipe Data
1	juanda international airport	Time series google trends	Numerik
2	surabaya international airport	Time series google trends	Numerik
3	surabaya ferry terminal	Time series google trends	Numerik
4	surabaya	Time series google trends	Numerik
5	bromo	Time series google trends	Numerik
6	passport indonesia	Time series google trends	Numerik
7	visa indonesia	Time series google trends	Numerik
8	bromo hotel	Time series google trends	Numerik
9	ijen resort	Time series google trends	Numerik
10	ijen restaurant	Time series google trends	Numerik

11	ijen travel	Time series google trends	Numerik
12	surabaya travel destination	Time series google trends	Numerik
13	bromo cafes	Time series google trends	Numerik
14	surabaya shopping centers	Time series google trends	Numerik
15	surabaya theme parks	Time series google trends	Numerik
16	surabaya museums	Time series google trends	Numerik
17	surabaya libraries	Time series google trends	Numerik
18	movie theaters	Time series google trends	Numerik
19	ijen package	Time series google trends	Numerik
20	ijen package tour	Time series google trends	Numerik
21	ijen acomodation	Time series google trends	Numerik
22	banyuwangi travel	Time series google trends	Numerik
23	banyuwangi hotel	Time series google trends	Numerik
24	banyuwangi airport	Time series google trends	Numerik
25	indonesia currency	Time series google trends	Numerik
26	indonesia visa	Time series google trends	Numerik
27	visa for indonesia	Time series google trends	Numerik
28	visa to indonesia	Time series google trends	Numerik
29	indonesia time	Time series google trends	Numerik
30	surabaya food	Time series google trends	Numerik
31	fried rice	Time series google trends	Numerik
32	surabaya flight	Time series google trends	Numerik
33	malang	Time series google trends	Numerik
34	banyuwangi beach	Time series google trends	Numerik
35	ijen bluefire	Time series google trends	Numerik
36	bromo	Time series google trends	Numerik
37	travel to surabaya	Time series google trends	Numerik
38	travel to bromo	Time series google trends	Numerik

39	surabaya jakarta	Time series google trends	Numerik
40	surabaya shopping	Time series google trends	Numerik
41	bromo shopping	Time series google trends	Numerik
42	weather in surabaya	Time series google trends	Numerik
43	bromo weather	Time series google trends	Numerik
44	weather in bromo	Time series google trends	Numerik
45	weather in bromo	Time series google trends	Numerik
46	bromo travel	Time series google trends	Numerik
47	surabaya travel	Time series google trends	Numerik
48	surabaya restaurants	Time series google trends	Numerik
49	surabaya shopping	Time series google trends	Numerik
50	surabaya flights	Time series google trends	Numerik
51	surabaya airports	Time series google trends	Numerik
52	surabaya railways	Time series google trends	Numerik
53	surabaya maps	Time series google trends	Numerik
54	surabaya bars	Time series google trends	Numerik
55	surabaya shows	Time series google trends	Numerik
56	surabaya grocery markets	Time series google trends	Numerik
57	surabaya specialty food shops	Time series google trends	Numerik
58	surabaya local parks	Time series google trends	Numerik
59	surabaya plazas	Time series google trends	Numerik
60	surabaya public transport hubs	Time series google trends	Numerik
61	surabaya farmers markets	Time series google trends	Numerik
62	surabaya drug stores	Time series google trends	Numerik
63	bromo accomodation	Time series google trends	Numerik
64	surabaya safaris	Time series google trends	Numerik
65	game reserves	Time series google trends	Numerik
66	surabaya pharmachies	Time series google trends	Numerik

67	surabaya national parks	Time series google trends	Numerik
68	surabaya public beaches	Time series google trends	Numerik
69	surabaya subway	Time series google trends	Numerik
70	surabaya bus	Time series google trends	Numerik
71	surabaya train stations	Time series google trends	Numerik
72	bromo costs	Time series google trends	Numerik
73	indonesia holiday	Time series google trends	Numerik
74	Jumlah Kedatangan Wisatawan	Time series	Numerik

Untuk memberikan gambaran umum mengenai format dan isi data, berikut ditampilkan beberapa baris contoh dari dataset yang digunakan. Data ini mencakup variabel-variabel utama yang menjadi input dalam pemodelan prediksi jumlah kunjungan wisatawan

Date,Kedatangan,indonesia holiday,bromo cost,surabaya train station,surabaya bus,surabaya subway,surabaya public beach,surabaya national park	
2008-01,927,11,0,0,0,0,0,0,89,0,0,0,0,0,0,0,1,4,0,4,6,0,0,0,0,0,11,0,0,0,0,0,0,0,0,0,8,0,9,0,0,0,0,0,10,0,0,1,0,9,0,4,0,13,24,4,4,0,0,0,3,0,46,0,0,0,0	
2008-02,1328,10,0,0,0,0,0,0,83,0,0,0,0,0,0,0,2,4,0,5,5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,8,0,9,0,0,0,0,0,12,0,0,1,0,8,0,4,0,11,22,4,4,0,0,0,2,0,40,0,0,0,0	
2008-03,1686,12,0,0,2,0,0,0,0,74,0,0,0,0,0,0,0,2,5,0,5,5,0,0,0,0,0,9,9,0,0,0,0,0,0,0,0,7,0,10,0,0,0,0,0,13,0,0,1,0,8,0,5,0,15,23,4,4,0,0,0,3,0,37,0,0,0,0	
2008-04,1669,8,0,0,0,0,0,0,73,0,0,0,0,0,0,0,2,5,0,5,7,0,0,0,0,0,10,10,0,0,0,0,0,0,0,0,8,0,12,0,0,0,0,0,16,0,0,1,0,7,0,5,0,13,21,5,5,0,0,0,3,0,31,0,0,0,0	
2008-05,1697,10,0,0,2,0,0,0,0,71,0,0,0,0,0,4,0,2,6,0,5,6,0,0,0,0,9,10,0,0,0,0,0,0,0,0,9,0,12,0,0,0,0,0,16,0,0,1,0,7,0,6,0,13,27,5,5,0,0,0,3,0,44,0,0,0,0	
2008-06,1607,7,0,0,3,0,0,0,0,68,0,0,0,0,0,0,0,2,6,0,5,6,0,0,0,0,9,0,0,0,0,0,0,0,0,0,11,0,13,0,0,0,0,0,14,0,0,1,0,8,0,6,0,12,27,5,5,0,0,0,4,0,50,0,0,0,0	
2008-07,1952,13,0,0,0,0,0,0,75,0,0,0,0,0,4,0,0,1,7,0,6,8,0,0,0,0,8,9,0,0,0,0,0,0,0,0,12,4,14,0,0,0,0,0,11,0,0,1,0,8,0,7,0,12,28,5,4,0,0,0,4,0,62,0,0,0,0	
2008-08,2048,9,0,0,4,0,0,0,0,78,0,0,0,0,0,0,0,2,7,0,5,8,0,0,0,0,7,12,0,0,0,0,0,0,0,0,12,0,14,0,0,0,0,0,14,0,0,1,0,8,0,7,0,15,24,5,5,0,0,0,4,0,47,0,0,0,0	
2008-09,2109,11,0,0,2,0,0,0,0,70,0,0,0,0,0,4,0,0,2,6,0,6,6,0,0,0,0,10,0,0,0,0,0,0,0,0,3,14,3,12,0,0,0,0,0,16,0,0,1,0,9,0,6,0,13,24,5,6,1,0,0,4,0,33,0,0,0,0	
2008-10,2432,9,0,0,2,0,0,0,0,75,0,0,0,0,0,5,0,0,2,5,0,7,6,0,0,0,0,13,8,0,0,0,0,0,0,3,0,0,10,0,13,0,0,0,0,0,16,0,0,1,0,8,0,5,0,16,24,3,5,0,0,0,3,0,37,0,0,0,0	
2008-11,2032,10,0,0,2,0,0,0,0,64,0,0,0,0,0,0,0,2,6,0,7,7,0,0,0,0,11,0,0,0,42,0,0,0,0,0,15,3,14,0,0,0,0,0,16,0,0,1,0,8,0,6,0,14,23,4,4,0,0,0,5,0,55,0,0,0,0	
2008-12,2308,12,0,0,2,0,0,0,0,58,0,0,0,0,0,5,0,0,2,6,0,6,6,0,0,0,0,10,0,0,0,0,0,0,0,14,0,14,0,0,0,0,0,13,0,0,1,0,8,0,6,0,12,21,5,4,0,0,0,6,0,47,0,0,0,0	
2009-01,1981,12,0,0,1,0,0,0,0,76,0,0,0,0,0,0,0,2,6,0,5,5,0,0,0,0,12,7,0,0,0,0,0,0,0,3,12,3,17,0,0,0,0,0,13,0,0,1,0,11,0,6,0,12,24,5,4,0,0,0,4,0,60,0,0,0,0	
2009-02,1956,10,0,0,2,0,0,0,0,79,0,3,0,0,0,0,0,0,2,7,0,7,7,0,0,2,0,8,11,0,0,0,0,0,0,0,0,12,4,16,0,0,0,0,0,16,0,0,1,0,10,0,7,0,13,23,5,4,0,0,0,5,0,51,0,0,0,0	
2009-03,2274,13,0,0,2,0,0,0,0,70,0,4,0,0,0,3,0,0,2,7,0,6,5,0,0,0,0,10,7,0,0,0,0,0,0,0,11,3,18,0,0,0,0,0,17,0,0,1,0,10,0,7,0,13,24,6,6,1,0,0,4,0,39,0,0,0,0	
2009-04,2538,13,0,0,2,0,0,0,0,79,0,4,0,0,0,5,0,0,2,7,0,7,6,0,0,0,0,10,12,0,0,0,0,0,0,0,0,12,0,14,0,0,0,0,0,17,0,0,1,0,9,0,7,0,14,26,5,4,0,0,0,5,0,42,0,0,0,0	
2009-05,2543,12,0,0,1,0,0,0,0,79,0,4,0,0,4,0,0,2,8,0,8,8,0,0,0,0,8,11,0,7,0,0,0,0,0,0,3,16,4,19,0,0,0,0,0,18,0,0,1,0,9,0,8,0,14,27,6,5,0,0,0,5,0,51,0,0,0,0	
2009-06,2393,13,0,0,3,0,0,0,0,82,0,4,0,0,5,0,0,2,10,0,9,12,0,0,0,0,11,11,0,5,0,0,0,0,0,3,0,0,19,4,21,0,0,0,0,0,17,0,0,1,0,9,0,10,0,16,28,6,5,0,0,0,8,0,60,0,0,0,0	
2009-07,2970,13,0,0,4,0,0,0,0,86,0,4,0,0,8,0,0,2,9,0,9,12,0,0,0,0,11,10,0,9,0,0,0,0,0,3,18,0,19,0,0,0,0,0,16,0,0,1,0,10,0,9,0,18,27,7,5,0,0,0,7,0,70,0,0,0,0	
2009-08,3559,12,0,0,4,0,0,0,0,92,0,0,0,0,0,7,0,0,2,8,0,8,10,0,0,0,0,14,17,0,11,0,0,0,0,0,3,0,3,18,4,19,0,0,1,0,0,0,15,0,0,1,0,10,0,8,0,15,27,6,5,0,0,0,5,0,60,0,0,0,0	

Gambar 3. 2 Contoh data set

Tahap *data understanding* bertujuan untuk memperoleh pemahaman menyeluruh terhadap struktur, kualitas, serta karakteristik data yang akan digunakan dalam penelitian. Beberapa langkah yang dilakukan pada tahap ini antara lain:

a. *Exploratory Data Analysis – Data Loading*

Proses ini mencakup pemuatan data dari sumber eksternal ke dalam lingkungan pemrosesan yang digunakan, seperti perangkat lunak analisis data atau sistem basis data. Data dapat diperoleh dari berbagai sumber, seperti file lokal, server eksternal, atau repositori daring.

b. *Exploratory Data Analysis – Deskripsi Variabel*

Langkah ini dilakukan untuk mengidentifikasi dan mendeskripsikan setiap variabel dalam dataset, termasuk tipe data, rentang nilai, distribusi, serta tingkat variasinya. Deskripsi ini penting untuk mengetahui bagaimana setiap variabel berkontribusi terhadap proses analisis.

c. *Exploratory Data Analysis – Penanganan Outlier dan Missing Value*

Outlier merupakan nilai-nilai ekstrem yang berbeda jauh dari pola umum data dan dapat memengaruhi hasil analisis jika tidak ditangani dengan tepat. Sementara itu, *missing values* atau nilai kosong perlu dikenali dan diatasi untuk menjaga kualitas data. Salah satu teknik yang digunakan untuk mendeteksi outlier adalah metode *Interquartile Range* (IQR), yaitu dengan menetapkan batas bawah dan atas berdasarkan kuartil pertama (Q1) dan kuartil ketiga (Q3). Nilai yang berada di luar rentang $Q1 - 1.5 \times IQR$ dan $Q3 + 1.5 \times IQR$ dikategorikan sebagai outlier.

d. *Exploratory Data Analysis – Univariate Analysis*

Analisis univariat bertujuan untuk memahami karakteristik dasar dari masing-masing variabel secara individual. Fokus analisis ini mencakup

penyebaran data (*distribution*), kecenderungan pusat (*central tendency*), serta tingkat variabilitas. Analisis ini juga menjadi tahap awal untuk mengidentifikasi pola yang mungkin berkaitan dengan target prediksi.

e. *Exploratory data analysis – Multivariate Analysis*

Analisis multivariat dilakukan untuk mengeksplorasi hubungan antar variabel. Teknik ini berguna untuk mengetahui seberapa besar pengaruh satu variabel terhadap variabel lain, serta bagaimana interaksi antar variabel tersebut membentuk pola yang kompleks dalam data.

3.1.5) *Data preparation*

Tahap *data preparation* merupakan proses prapemrosesan yang penting sebelum data digunakan dalam proses pelatihan model. Tujuan dari tahap ini adalah untuk memastikan bahwa data dalam kondisi optimal dan sesuai dengan kebutuhan algoritma pembelajaran mesin yang akan digunakan. Adapun langkah-langkah dalam tahap ini antara lain:

1. *Data preparation – Data Encoding*

Data kategorik perlu diubah ke dalam format numerik agar dapat diproses oleh algoritma pembelajaran mesin. Salah satu teknik yang digunakan adalah *one-hot encoding*, yaitu metode yang mengubah setiap kategori menjadi fitur biner (*dummy variable*) dengan nilai 0 atau 1.

2. *Data Preparation – Reduksi Dimensi dengan Boruta*

Seleksi fitur menggunakan algoritma Boruta dilakukan untuk mengurangi dimensi data sekaligus mempertahankan fitur-fitur

yang paling relevan terhadap target prediksi. Karena pentingnya tahapan ini dalam penelitian, pembahasannya akan disampaikan secara khusus pada sub-bab tersendiri.

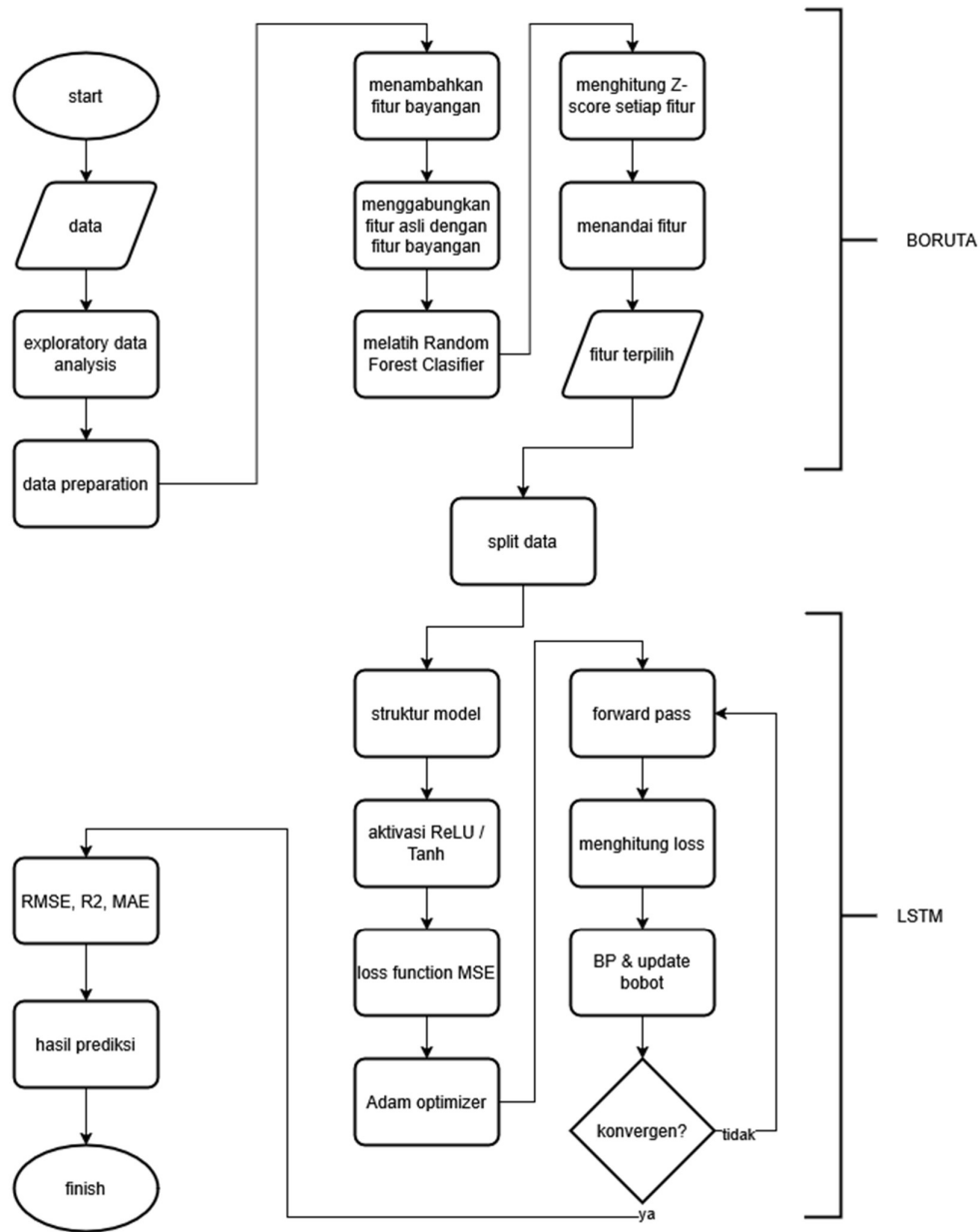
3. *Data Preparation – Split Data Training dan Testing*

Dataset dibagi menjadi dua bagian, yaitu data pelatihan (training data) dan data pengujian (testing data). Data pelatihan digunakan untuk membangun dan melatih model prediksi, sedangkan data pengujian digunakan untuk mengevaluasi kinerja model terhadap data yang belum pernah dilihat sebelumnya.

4. *Data Preparation - Normalisasi*

Proses normalisasi bertujuan untuk menyamakan skala nilai pada setiap fitur agar memiliki distribusi yang seragam, biasanya dalam rentang 0 hingga 1. Normalisasi penting dilakukan karena perbedaan skala antar fitur dapat memengaruhi kinerja algoritma pembelajaran mesin, terutama yang sensitif terhadap nilai absolut seperti LSTM.

Rangkaian langkah dari pemahaman data hingga persiapan akhir untuk proses modeling disusun secara sistematis dalam gambar 3.3 berikut sebagai gambaran umum proses yang dilakukan dalam penelitian ini.



Gambar 3. 3 Flowchart Alur Penelitian dengan Boruta dan LSTM

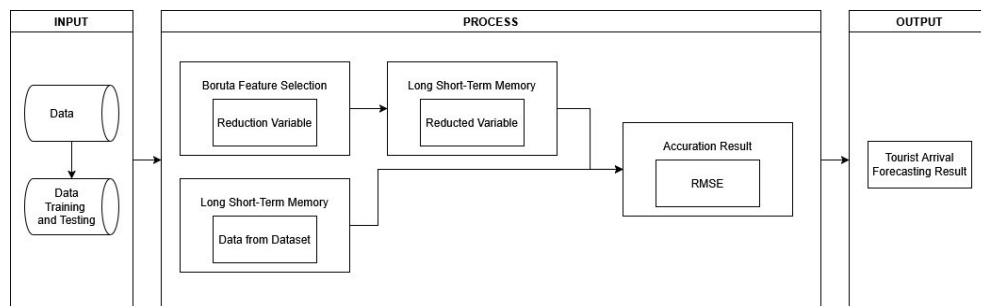
3.2 Desain Model

Desain sistem ini menjelaskan alur kerja dari model prediksi yang dikembangkan dalam penelitian. Input utama dalam sistem ini berasal dari data indeks pencarian Google, yang memuat berbagai kata kunci (keywords) sebagai

fitur, serta data jumlah kedatangan wisatawan yang digunakan sebagai target prediksi.

Algoritma utama yang digunakan untuk membangun model prediktif adalah Long Short-Term Memory (LSTM), yaitu salah satu varian jaringan saraf tiruan yang dirancang khusus untuk menangani data deret waktu. Untuk meningkatkan efektivitas dan efisiensi model, dilakukan proses seleksi fitur menggunakan algoritma Boruta, yang berfungsi untuk mengidentifikasi kata kunci paling relevan dari sekian banyak fitur dalam indeks pencarian Google.

Dengan pendekatan ini, sistem yang dirancang diharapkan mampu menghasilkan keluaran berupa prediksi jumlah kunjungan wisatawan bulanan, berdasarkan tren penelusuran digital yang relevan. Desain ini memungkinkan pemanfaatan big data dari mesin pencari sebagai indikator dalam pemodelan permintaan wisata yang lebih adaptif dan akurat.



Gambar 3. 4 Desain Model

3.3 Eksperimen

3.3.2) Boruta

Boruta adalah sebuah algoritma seleksi fitur yang digunakan untuk memilih fitur-fitur yang relevan dan berguna dalam data. Algoritma Boruta bekerja dengan membandingkan importance score fitur asli

terhadap *shadow features*, yaitu salinan dari fitur asli yang telah diacak (permutasi nilai). *Shadow features* ini tidak memiliki hubungan langsung dengan variabel target, sehingga dapat digunakan sebagai batas acuan untuk menentukan pentingnya fitur asli..

Berikut Langkah-langkah yang dilakukan pada proses Boruta :

1. Inisialisasi

Data pelatihan dan target variable diinisialisasi.

2. Pembuatan Fitur Bayangan (duplikasi dan permutasi fitur)

Untuk setiap fitur asli $Permutation(X_j)$ dalam dataset, dibuat satu salinan acak $Shadow(X_j)$ dengan mengacak urutan nilainya.

Berikut adalah persamaannya :

$$Shadow(X_j) = Permutation(X_j) \quad (3.1)$$

3. Training Random Forest dan Perhitungan Importansi

Dataset diperluas dengan menambahkan semua *shadow features*, lalu dilakukan pelatihan model Random Forest. Dari model ini dihitung skor pentingnya fitur menggunakan metrik Gini importance, yaitu pengurangan rata-rata impuritas (Gini impurity) akibat pemisahan pada setiap node pohon untuk setiap fitur j . Berikut rumusnya :

$$Importance(X_j) = \sum_{t \in T_j} \Delta i(t) \cdot \frac{N_t}{N} \quad (3.2)$$

keterangan:

X_j : fitur ke j

T_j : himpunan node dalam pohon keputusan yang membagi
pada fitur X_j

$\Delta i(t)$: pengurangan impurity Gini pada node t

N_t : jumlah sampel pada node t

N : jumlah total sampel

4. Perhitungan Z-Score

Menghitung Z-score untuk setiap fitur asli berdasarkan perbandingan importansi dengan *shadow features* seperti rumus berikut:

$$Z_j = \frac{I_j - \mu_{shadow}}{\sigma_{shadow}} \quad (3.3)$$

keterangan:

Z_j : nilai z-score fitur asli ke- j

I_j : importansi dari fitur asli

μ_{shad} : rata-rata importansi dari fitur bayangan

σ_{sha} : standar deviasi importansi dari fitur bayangan

5. Pengambilan Keputusan

Jika z-score fitur asli lebih besar dari z-score fitur bayangan, maka fitur asli tersebut dianggap relevan dan dipilih. Jika tidak, maka fitur asli tersebut dianggap tidak relevan dan tidak dipilih.

$$\text{If } Z_j > \max(Z_{shadow}) \Rightarrow X_j \text{ is confirmed} \quad (3.4)$$

6. Pengulangan Proses

Boruta melakukan proses ini secara berulang hingga semua fitur diklasifikasikan.

Repeat until all $X_j \in \{Confirmed, Rejected\}$ (3.5)

3.3.3) Long Short-Term Memory

Long Short-Term Memory (LSTM) merupakan salah satu arsitektur Recurrent Neural Network (RNN) yang dirancang khusus untuk mengatasi masalah *vanishing gradient* yang umum terjadi pada RNN konvensional, terutama saat memproses data sekuensial dengan dependensi jangka panjang. LSTM memiliki struktur sel memori yang memungkinkan jaringan untuk mempertahankan informasi dalam jangka waktu yang panjang, serta menggunakan mekanisme gerbang atau *gate* untuk mengontrol aliran informasi.

Setiap unit LSTM memiliki tiga gerbang utama, yaitu *forget gate*, *input gate*, dan *output gate*, yang masing-masing berfungsi untuk memutus, memperbarui, dan menghasilkan informasi dari memori jangka panjang. Proses komputasi dalam satu unit LSTM dijelaskan melalui langkah-langkah berikut :

1. Inisialisasi

Inisialisasi sel LSTM dengan nilai awal yang acak.

2. Proses pada Forget Gate

Gerbang ini menentukan informasi mana dari cell state sebelumnya yang akan dilupakan. Dihitung seperti rumus berikut:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.6)$$

keterangan:

f_t : output dari *forget gate*

σ : fungsi aktivasi sigmoid

W_f : matriks bobot dari *forget gate*

h_{t-1} : hidden state pada waktu sebelumnya

x_t : input pada waktu ke- t

b_f : vektor bias dari *forget gate*

3. Input gate dan kandidat *cell state*

Gerbang ini memutuskan informasi baru mana yang akan ditambahkan ke cell state. Terdiri dari dua komponen yaitu gerbang input i_t dan kandidat cell state \tilde{C}_t . dan berikut persamaannya:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.7)$$

keterangan:

i_t : gate input pada waktu t

σ : fungsi aktivasi sigmoid

W_i : matriks bobot untuk gate input

x_t : input pada waktu t

h_{t-1} : hidden state pada waktu sebelumnya

b_i : bias untuk gate input

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3.8)$$

keterangan:

\tilde{C}_t : kandidat cell state baru

\tanh : fungsi aktivasi hiperbolik tangent

W_C : matriks bobot cell state

h_{t-1} : hidden state pada waktu sebelumnya

x_t : input pada waktu t

b_C : bias untuk cell state

4. Update cell state

Memperbarui nilai dari cell state berdasarkan hasil dari dua proses sebelumnya. Dan berikut persamaannya:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (3.9)$$

keterangan:

C_t : nilai sel pada waktu t

f_t : gate forget pada waktu t

C_{t-1} : nilai sel pada waktu sebelumnya

i_t : gate input pada waktu t

\tilde{C}_t : kandidat cell state baru

\odot : operasi perkalian elemen-per-elemen (*Hadamard product*)

5. Penghitungan Hidden State

Penghitungan hidden state LSTM berdasarkan nilai sel dan gate output. Berikut persamaannya:

$$h_t = o_t \odot \tanh(C_t) \quad (3.10)$$

keterangan:

h_t : hidden state pada waktu t

o_t : gate output pada waktu t

\odot : operasi perkalian elemen-per-elemen (*Hadamard product*)

\tanh : fungsi aktivasi tangens hiperbolikus

C_t : nilai sel pada waktu t

6. Penghitungan Output

Penghitungan output LSTM berdasarkan hidden state berikut ini persamaannya:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.11)$$

keterangan:

o_t : output pada waktu t

σ : fungsi aktivasi sigmoid

W_o : matriks bobot untuk output

h_{t-1} : hidden state pada waktu sebelumnya

b_o : bias untuk output

3.4 Evaluasi

Setelah model selesai dibangun dan dilatih, tahap selanjutnya adalah melakukan evaluasi untuk mengukur performa model dalam melakukan prediksi. Evaluasi bertujuan untuk menilai sejauh mana model mampu menghasilkan output yang mendekati nilai aktual. Pada dasarnya, proses ini dilakukan dengan membandingkan antara nilai prediksi dan nilai aktual, di mana perbedaan antara keduanya disebut sebagai error.

Sebelum melakukan evaluasi, diperlukan proses scaling ulang terhadap data uji. Hal ini penting dilakukan agar data numerik pada tahap pengujian memiliki skala yang konsisten dengan data pelatihan, sehingga mencegah terjadinya data leakage atau kebocoran data yang dapat memengaruhi hasil evaluasi.

Dalam penelitian ini, kinerja model dievaluasi menggunakan tiga metrik regresi utama, yaitu:

- R^2 (R-Squared / Koefisien Determinasi), yang mengukur proporsi variasi data target yang dapat dijelaskan oleh model,
- RMSE (Root Mean Squared Error), yang mengukur seberapa besar rata-rata kesalahan prediksi dalam satuan aslinya, dan
- MAE (Mean Absolute Error), yang menghitung rata-rata dari nilai absolut selisih antara hasil prediksi dan nilai aktual.

Adapun rumus dari masing-masing metrik tersebut disajikan pada bagian berikut ini:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.12)$$

keterangan :

y_i : nilai aktual ke-i

\hat{y}_i : nilai prediksi ke-i

\bar{y} : rata-rata dari nilai aktual

n : jumlah total data

$\sum_{i=1}^n (y_i - \hat{y}_i)^2$: jumlah kuadrat error prediksi (Residual Sum of Squares – RSS)

$\sum_{i=1}^n (y_i - \bar{y})^2$: jumlah kuadrat total (Total Sum of Squares – TSS)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.13)$$

keterangan :

n : jumlah data

y_i : nilai aktual

\hat{y}_i : nilai prediksi

Σ : penjumlahan total

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.14)$$

keterangan :

n : jumlah data

y_i : nilai aktual

\hat{y}_i : nilai prediksi

$|y_i - \hat{y}_i|$: nilai absolut dari selisih antara aktual dan prediksi

3.5 Instrumen Penelitian

Instrumen penelitian merupakan komponen penting yang menentukan kualitas dan validitas dari hasil yang diperoleh. Ketepatan dalam pemilihan dan penggunaan instrumen akan sangat memengaruhi keberhasilan proses pengujian dan analisis model.

Dalam penelitian ini, variabel bebas (independen) terdiri atas data indeks pencarian Google serta data jumlah kedatangan wisatawan. Data indeks pencarian diperoleh melalui Google Trends, sedangkan data kunjungan wisatawan bersumber dari Badan Pusat Statistik (BPS).

Adapun variabel terikat (dependen) adalah kinerja model prediksi, yang diukur berdasarkan tingkat kesalahan (error) menggunakan metrik evaluasi seperti MAE, RMSE, dan R^2 . Selain itu, terdapat variabel antara (intervening), yaitu hasil

prediksi jumlah kedatangan wisatawan yang dihasilkan dari model pembelajaran mesin yang dibangun.

Dengan instrumen tersebut, penelitian ini bertujuan untuk menghasilkan model prediksi yang akurat dan dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan strategis di sektor pariwisata.

BAB IV

PEMBAHASAN

Pada bagian ini dipaparkan hasil penelitian sekaligus pembahasan yang komprehensif. Penjelasan langkah-langkah dan hasil analisisnya. Berikut penjelasannya.

4.1 Business Understanding

Industri pariwisata merupakan sektor yang memiliki potensi ekonomi tinggi, karena mampu mendorong pertumbuhan berbagai bidang terkait seperti perhotelan, transportasi, kuliner, hingga ekonomi kreatif. Potensi inilah yang menjadikan sektor pariwisata menarik bagi para pelaku usaha. Namun demikian, karena sifatnya yang sangat dipengaruhi oleh berbagai faktor seperti musim, tren sosial, kondisi ekonomi, dan kebijakan pemerintah, maka pengelolaan bisnis di sektor ini menuntut kehati-hatian dan perencanaan yang matang (Becken, 2013).

Para pelaku bisnis pariwisata dituntut untuk mampu mengantisipasi dan merespons fluktuasi dalam permintaan wisatawan. Salah dalam memperkirakan jumlah kedatangan wisatawan dapat mengakibatkan kelebihan atau kekurangan dalam penyediaan layanan dan fasilitas, yang pada akhirnya dapat menyebabkan kerugian finansial. Ketidakakuratan ini tentu merugikan karena setiap bisnis berorientasi pada profit. Oleh karena itu, memiliki kemampuan untuk memprediksi jumlah kedatangan wisatawan secara akurat merupakan keunggulan kompetitif yang penting bagi pelaku usaha pariwisata (Gholipour et al., 2016).

Hal yang sama juga berlaku bagi para pembuat kebijakan. Prediksi jumlah wisatawan yang akurat dapat membantu perencanaan pembangunan destinasi

wisata, alokasi anggaran, dan penyusunan kebijakan yang lebih adaptif. Kebijakan yang selaras dengan proyeksi jangka pendek, menengah, dan panjang akan mampu memberikan dampak positif terhadap pengembangan pariwisata yang berkelanjutan. Oleh karena itu, kemampuan untuk memprediksi kedatangan wisatawan secara sistematis dan *data-driven* sangat krusial dalam mendukung keputusan bisnis maupun kebijakan publik..

4.1.1 Problem statements and Goals

Berdasarkan pemahaman terhadap konteks bisnis dan kebutuhan prediksi di sektor pariwisata, penelitian ini bertujuan untuk merancang dan membangun sistem prediksi jumlah kedatangan wisatawan. Fokus utama penelitian diarahkan pada pemanfaatan data pencarian dari Google Trends dan penerapan algoritma Long Short-Term Memory (LSTM) yang dioptimasi melalui seleksi fitur menggunakan metode Boruta.

Adapun rumusan masalah yang hendak dijawab dalam penelitian ini adalah sebagai berikut:

1. Bagaimana tingkat akurasi dan performa model LSTM yang telah dioptimasi menggunakan algoritma seleksi fitur Boruta dalam memprediksi jumlah kunjungan wisatawan berdasarkan data Google Trends?

Untuk menjawab pertanyaan masalah tersebut, maka tujuan dari penelitian ini adalah sebagai berikut :

1. Mengukur akurasi dan menganalisa hasil model LSTM yang dioptimasi dengan seleksi fitur Boruta dalam memprediksi jumlah kedatangan wisatawan berbasis google trends.

4.1.2 Metodologi

Sebagaimana dijelaskan dalam Bab 3, penelitian ini bertujuan untuk memprediksi jumlah kedatangan wisatawan, yang merupakan variabel kontinu. Oleh karena itu, pendekatan yang digunakan adalah membangun model regresi dengan target variabel jumlah kedatangan wisatawan. Model yang digunakan adalah Long Short-Term Memory (LSTM) yang dioptimasi dengan seleksi fitur Boruta.

4.1.3 Matrik

Untuk mengevaluasi kinerja model prediksi, digunakan metrik evaluasi yang umum dalam masalah regresi, yaitu koefisien determinasi (R^2), Mean Absolute Error (MAE), dan Root Mean Square Error (RMSE).

R^2 menunjukkan proporsi variansi data yang dapat dijelaskan oleh model, dengan nilai mendekati 1 menandakan bahwa model mampu menjelaskan sebagian besar variasi data target. MAE mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual, yang memberikan gambaran kesalahan model dalam satuan aslinya tanpa mempertimbangkan arah selisih. Sementara itu, RMSE mengukur seberapa jauh hasil prediksi dari nilai aktual dengan memberikan penalti lebih besar pada kesalahan yang besar, sehingga sangat sensitif terhadap outlier.

Dengan menggunakan ketiga metrik ini secara bersamaan, diperoleh gambaran yang lebih menyeluruh tentang tingkat akurasi, stabilitas, dan keandalan dari model prediktif yang dibangun.

4.2 *Analytic Approach*

Untuk memprediksi jumlah kedatangan wisatawan, penelitian ini mengadopsi metode Long Short-Term Memory (LSTM), sebuah arsitektur jaringan saraf tiruan yang dirancang khusus untuk menangani data deret waktu. LSTM memiliki kemampuan untuk menangkap pola temporal dan hubungan jangka panjang dalam data, menjadikannya pilihan yang tepat untuk peramalan dalam konteks pariwisata. Studi oleh (Sugiartawan et al., 2017) menunjukkan bahwa LSTM unggul dalam memodelkan data deret waktu yang kompleks dan non-linear, seperti data kedatangan wisatawan.

Namun, meskipun LSTM memiliki keunggulan dalam menangani data deret waktu, model ini juga memiliki keterbatasan, terutama ketika dihadapkan pada data dengan fluktuasi tinggi atau jumlah fitur yang besar. Kelebihan fitur yang tidak relevan dapat menyebabkan overfitting dan meningkatkan kompleksitas model, yang pada akhirnya menurunkan akurasi prediksi. Untuk mengatasi masalah ini, seleksi fitur menjadi langkah penting dalam proses pra-pemrosesan data. Salah satu metode seleksi fitur yang efektif adalah algoritma Boruta, yang bekerja dengan membandingkan pentingnya setiap fitur asli dengan versi acaknya (*shadow features*) menggunakan model Random Forest. Fitur yang secara signifikan lebih penting daripada shadow features dipertahankan, sementara yang lainnya dihapus. Pendekatan ini tidak hanya meningkatkan akurasi model tetapi juga mengurangi waktu pelatihan dan kompleksitas komputasi.

Integrasi antara LSTM dan seleksi fitur Boruta telah terbukti efektif dalam berbagai studi. Misalnya, dalam penelitian oleh (Hsieh, 2021), kombinasi LSTM

dengan Boruta berhasil meningkatkan akurasi prediksi kedatangan wisatawan di Taiwan. Demikian pula, studi oleh (Pabuccu & Barbu, 2023) menunjukkan bahwa penggunaan Boruta dalam seleksi fitur dapat meningkatkan kinerja model prediksi deret waktu di bidang keuangan.

Dengan demikian, pendekatan yang menggabungkan LSTM dan seleksi fitur Boruta diharapkan dapat memberikan hasil prediksi yang lebih akurat dan efisien dalam konteks prediksi kedatangan wisatawan.

4.2.1 Data Requirement

Dalam penelitian ini, data deret waktu (time series) digunakan sebagai dasar untuk memprediksi jumlah kedatangan wisatawan. Pendekatan ini telah banyak diterapkan dalam berbagai studi karena kemampuannya dalam menangkap pola musiman dan tren jangka panjang dalam data pariwisata. Misalnya, (Xu et al., 2024) mengembangkan kerangka kerja untuk memprediksi kedatangan pengunjung di atraksi wisata menggunakan data deret waktu, menunjukkan efektivitas pendekatan ini dalam konteks pasca-COVID-19.

Selain data historis kedatangan wisatawan, beberapa penelitian juga memanfaatkan indeks ekonomi sebagai variabel prediktor dalam model peramalan pariwisata. Indeks seperti Produk Domestik Bruto (PDB), tingkat inflasi, dan nilai tukar mata uang sering digunakan untuk menangkap faktor-faktor ekonomi yang mempengaruhi permintaan wisata. Sebagai contoh, penelitian oleh (Andariesta & Wasesa, 2022) menunjukkan bahwa faktor-faktor ekonomi makro memiliki pengaruh signifikan terhadap permintaan pariwisata internasional.

Dalam beberapa tahun terakhir, data dari mesin pencari internet, seperti Google Trends, telah digunakan sebagai indikator untuk memprediksi kedatangan

wisatawan. Data ini mencerminkan minat dan keinginan perjalanan calon wisatawan berdasarkan perilaku pencarian mereka di internet. Studi (Gholipour et al., 2016) menunjukkan bahwa data pencarian internet dapat meningkatkan akurasi model peramalan pariwisata, terutama dalam jangka pendek

Berdasarkan penelitian-penelitian tersebut, penelitian ini akan menggabungkan data deret waktu kedatangan wisatawan dan data indeks pencarian internet dari Google Trends untuk membangun model prediksi yang lebih responsif terhadap perubahan perilaku wisatawan. Pendekatan ini diharapkan dapat memberikan wawasan yang lebih mendalam bagi pemangku kepentingan dalam industri pariwisata untuk perencanaan dan pengambilan keputusan yang lebih efektif.

4.3 Data Collection

Data yang digunakan dalam penelitian ini diperoleh dari sumber data publik. Data historis kedatangan wisatawan bulanan dikumpulkan dari Badan Pusat Statistik (BPS) Indonesia, mencakup periode Januari 2008 hingga Maret 2025. Data ini digunakan sebagai variabel target dalam model prediksi.

Selain itu, data dari Google Trends dikumpulkan untuk berbagai kata kunci yang relevan dengan pariwisata berdasarkan saran pencarian Google Trends. Data ini mencerminkan volume pencarian bulanan untuk setiap kata kunci, yang digunakan sebagai variabel prediktor dalam model. Dataset yang dihasilkan terdiri dari 207 baris (jumlah bulan dalam periode tersebut) dan 100 kolom (jumlah kata kunci ditambah satu kolom untuk data target dan satu kolom tanggal). Semua data disimpan dalam format CSV untuk memudahkan proses analisis dan pemodelan.

4.4 Data Understanding

4.4.1 Exploratory Data Analysis – Deskripsi Fitur

Pada tahap ini, dilakukan identifikasi dan deskripsi terhadap setiap variabel atau fitur yang terdapat dalam dataset yang digunakan dalam proses pemodelan. Tujuan dari tahap ini adalah untuk memahami karakteristik masing-masing variabel yang akan digunakan sebagai input maupun target dalam proses prediksi jumlah kedatangan wisatawan. Deskripsi mencakup fitur, jenis data, serta tipe data dari masing-masing fitur dalam struktur dataset..

Tabel 4. 1 Fitur yang digunakan

No	Fitur (kata kunci)	Jenis	Tipe data
1	Candi jawa tengah	time series google trends	numerik
2	East java	time series google trends	numerik
3	Magetan jawa timur	time series google trends	numerik
4	Pantai pacitan	time series google trends	numerik
5	Tiket pesawat	time series google trends	numerik
6	West java	time series google trends	numerik
7	East java.1	time series google trends	numerik
8	Central java	time series google trends	numerik
9	Pulau jawa	time series google trends	numerik
10	Pesawat	time series google trends	numerik
11	Kota malang jawa timur	time series google trends	numerik
12	Kota di jawa timur	time series google trends	numerik
13	Jawa timur indonesia	time series google trends	numerik
14	Jawa timur	time series google trends	numerik
15	Jawa tengah	time series google trends	numerik
16	Jawa barat	time series google trends	numerik
17	Ijen	time series google trends	numerik

18	Garuda indonesia	time series google trends	numerik
19	Citilink	time series google trends	numerik
20	Borneo	time series google trends	numerik
21	Bali	time series google trends	numerik
22	Bandung	time series google trends	numerik
23	Batu	time series google trends	numerik
24	Indonesian	time series google trends	numerik
25	Harga	time series google trends	numerik
26	Port	time series google trends	numerik
27	Indonesia	time series google trends	numerik
28	Medan	time series google trends	numerik
29	Batu city	time series google trends	numerik
30	Surabaya	time series google trends	numerik
31	Candi jawa timur	time series google trends	numerik
32	Kota malang	time series google trends	numerik
33	Sarangan	time series google trends	numerik
34	Baluran national park	time series google trends	numerik
35	Tiket kereta api	time series google trends	numerik
36	Bromo	time series google trends	numerik
37	Bandara di jawa timur	time series google trends	numerik
38	Turen	time series google trends	numerik
39	Wisata jawa timur	time series google trends	numerik
40	Train ticket	time series google trends	numerik
41	Pantai	time series google trends	numerik
42	Kereta api	time series google trends	numerik
43	Harga tiket pesawat	time series google trends	numerik
44	Harga tiket	time series google trends	numerik
45	Java	time series google trends	numerik

46	Jatim	time series google trends	numerik
47	Palembang	time series google trends	numerik
48	Gunung jawa timur	time series google trends	numerik
49	Kalimantan	time series google trends	numerik
50	Hotel jatim park	time series google trends	numerik
51	Central java	time series google trends	numerik
52	Yogyakarta	time series google trends	numerik
53	Blitar	time series google trends	numerik
54	Bandara jawa timur	time series google trends	numerik
55	Pelabuhan	time series google trends	numerik
56	Kediri	time series google trends	numerik
57	Mosque	time series google trends	numerik
58	Lampung	time series google trends	numerik
59	Bromo jawa timur	time series google trends	numerik
60	Pacitan	time series google trends	numerik
61	Hotel di malang	time series google trends	numerik
62	Hotel batu malang	time series google trends	numerik
63	Tiket bus	time series google trends	numerik
64	Jatim park 3	time series google trends	numerik
65	Batu malang	time series google trends	numerik
66	Hotel malang	time series google trends	numerik
67	Bandara	time series google trends	numerik
68	Candi of indonesia	time series google trends	numerik
69	Mountain	time series google trends	numerik
70	Jakarta	time series google trends	numerik
71	Jatim park	time series google trends	numerik
72	Juanda	time series google trends	numerik
73	Pantai jawa timur	time series google trends	numerik

74	Jatim park 3.1	time series google trends	numerik
75	Pantai pacitan jawa timur	time series google trends	numerik
76	Batu night spectacular (BNS)	time series google trends	numerik
77	Pacitan jawa timur	time series google trends	numerik
78	Sarangan jawa timur	time series google trends	numerik
79	Jawa timur park 3	time series google trends	numerik
80	Bandar udara internasional juanda	time series google trends	numerik
81	Lamongan regency	time series google trends	numerik
82	Blitar regency	time series google trends	numerik
83	Nama bandara	time series google trends	numerik
84	Tiket jatim park	time series google trends	numerik
85	Gunung tertinggi jawa timur	time series google trends	numerik
86	Hotel dekat jatim park	time series google trends	numerik
87	Abdul rachman saleh airport	time series google trends	numerik
88	Jatim park malang	time series google trends	numerik
89	Pricing strategy	time series google trends	numerik
90	Bandar udara abdul rachman saleh	time series google trends	numerik
91	Magetan regency	time series google trends	numerik
92	Pasuruan regency	time series google trends	numerik
93	Indonesia safari park II prigen	time series google trends	numerik
94	Mount bromo	time series google trends	numerik
95	Nama bandara jawa timur	time series google trends	numerik
96	Pacitan regency	time series google trends	numerik
97	Tourist attractiveness	time series google trends	numerik
98	Juanda international airport	time series google trends	numerik

4.5 Data Preparation

4.5.1 Data Preprocessing – Pemilihan Fitur dengan Boruta

Untuk mengurangi jumlah fitur yang kurang relevan dan mempertahankan hanya fitur yang paling informatif bagi model prediksi, penelitian ini menggunakan algoritma Boruta. Boruta merupakan metode seleksi fitur berbasis wrapper yang memanfaatkan algoritma pohon keputusan, dalam hal ini Random Forest, untuk mengevaluasi pentingnya setiap fitur dibandingkan dengan shadow features yang bersifat acak (Kursa & Rudnicki, 2010).

Proses seleksi dilakukan selama 100 iterasi, dengan jumlah total fitur awal sebanyak 99. Pada iterasi awal (1–7), tidak ada fitur yang dapat dikonfirmasi sebagai relevan, seluruh fitur dikategorikan sebagai tentatif. Hal ini wajar terjadi karena model masih melakukan bootstrap sampling untuk menghitung distribusi pentingnya setiap fitur.

Mulai dari iterasi ke-8, terlihat adanya fitur yang mulai dikonfirmasi. Tercatat 7 fitur dikonfirmasi sebagai penting, 11 masih tentatif, dan sisanya ditolak. Seiring berjalannya iterasi, jumlah fitur yang dikonfirmasi terus bertambah, menunjukkan stabilisasi dan konvergensi dari proses seleksi. Hingga iterasi ke-67, jumlah fitur yang dikonfirmasi mencapai 11, dan kemudian bertambah satu lagi menjadi 12 pada iterasi ke-72 dan bertahan stabil hingga akhir proses pada iterasi ke-100.

Setelah 100 iterasi, Boruta berhasil mengidentifikasi:

- 27 fitur yang dikonfirmasi penting (accepted)
- 1 fitur yang masih tentatif (tentative)
- 71 fitur yang ditolak (rejected).

Berikut daftar fitur hasil seleksi menurut Boruta beserta nilai importansi dari masing-masing fitur:

Tabel 4. 2 daftar fitur terpilih Boruta

No	Fitur (kata kunci)	Importansi (%)	Status
1	Candi jawa tengah	3,21	Accepted
2	East java	3,21	Accepted
3	Magetan jawa timur	3,21	Accepted
4	Pantai pacitan	3,21	Accepted
5	Tiket pesawat	3,21	Accepted
6	West java	3,21	Accepted
7	East java.1	3,21	Accepted
8	Central java	3,21	Accepted
9	Pulau jawa	3,21	Accepted
10	Pesawat	3,21	Accepted
11	Kota malang jawa timur	3,21	Accepted
12	Kota di jawa timur	3,21	Accepted
13	Jawa timur indonesia	3,21	Accepted
14	Jawa timur	3,21	Accepted
15	Jawa tengah	3,21	Accepted
16	Jawa barat	3,21	Accepted
17	Ijen	3,21	Accepted
18	Garuda indonesia	3,21	Accepted
19	Citilink	3,21	Accepted
20	Borneo	3,21	Accepted
21	Bali	3,21	Accepted
22	Bandung	3,21	Accepted
23	Batu	3,21	Accepted
24	Indonesian	3,21	Accepted
25	Harga	3,21	Accepted
26	Port	3,21	Accepted
27	Indonesia	3,21	Accepted
28	Medan	1,61	Tentative
29	Batu city	1,07	Rejected
30	Surabaya	1,07	Rejected
31	Candi jawa timur	6,43	Rejected
32	Kota malang	6,43	Rejected
33	Sarangan	5,36	Rejected
34	Baluran national park	4,59	Rejected
35	Tiket kereta api	4,02	Rejected
36	Bromo	3,57	Rejected
37	Bandara di jawa timur	2,92	Rejected
38	Turen	2,92	Rejected
39	Wisata jawa timur	2,92	Rejected
40	Train ticket	2,30	Rejected
41	Pantai	2,30	Rejected
42	Kereta api	2,14	Rejected
43	Harga tiket pesawat	1,89	Rejected
44	Harga tiket	1,89	Rejected
45	Java	1,79	Rejected
46	Jatim	1,69	Rejected
47	Palembang	1,61	Rejected
48	Gunung jawa timur	1,46	Rejected

49	Kalimantan	1,46	Rejected
50	Hotel jatim park	1,40	Rejected
51	Central java	1,29	Rejected
52	Yogyakarta	1,29	Rejected
53	Blitar	1,24	Rejected
54	Bandara jawa timur	1,19	Rejected
55	Pelabuhan	1,15	Rejected
56	Kediri	1,11	Rejected
57	Mosque	1,04	Rejected
58	Lampung	1,04	Rejected
59	Bromo jawa timur	9,74	Rejected
60	Pacitan	9,74	Rejected
61	Hotel di malang	9,45	Rejected
62	Hotel batu malang	9,18	Rejected
63	Tiket bus	8,93	Rejected
64	Jatim park 3	8,46	Rejected
65	Batu malang	8,46	Rejected
66	Hotel malang	8,24	Rejected
67	Bandara	8,04	Rejected
68	Candi of indonesia	7,65	Rejected
69	Mountain	7,65	Rejected
70	Jakarta	7,48	Rejected
71	Jatim park	7,14	Rejected
72	Juanda	7,14	Rejected
73	Pantai jawa timur	6,99	Rejected
74	Jatim park 3.1	6,84	Rejected
75	Pantai pacitan jawa timur	6,70	Rejected
76	Batu night spectacular (BNS)	6,43	Rejected
77	Pacitan jawa timur	6,43	Rejected
78	Sarangan jawa timur	6,30	Rejected
79	Jawa timur park 3	6,18	Rejected
80	Bandar udara internasional juanda	5,95	Rejected
81	Lamongan regency	5,95	Rejected
82	Blitar regency	5,74	Rejected
83	Nama bandara	5,74	Rejected
84	Tiket jatim park	5,64	Rejected
85	Gunung tertinggi jawa timur	5,45	Rejected
86	Hotel dekat jatim park	5,45	Rejected
87	Abdul rachman saleh airport	5,36	Rejected
88	Jatim park malang	5,27	Rejected
89	Pricing strategy	5,18	Rejected
90	Bandar udara abdul rachman saleh	5,02	Rejected
91	Magetan regency	5,02	Rejected
92	Pasuruan regency	4,95	Rejected
93	Indonesia safari park II prigen	4,87	Rejected
94	Mount bromo	4,66	Rejected
95	Nama bandara jawa timur	4,66	Rejected
96	Pacitan regency	4,66	Rejected
97	Tourist attractiveness	4,66	Rejected
98	Juanda international airport	4,66	Rejected

Hasil seleksi fitur yang ditampilkan dalam Tabel 4.2 menunjukkan daftar kata kunci pencarian Google Trends beserta nilai importansi yang diperoleh melalui

algoritma Boruta. Setiap fitur diklasifikasikan ke dalam tiga kategori: *Accepted*, *Tentative*, dan *Rejected*, berdasarkan kontribusinya terhadap proses prediksi jumlah kedatangan wisatawan mancanegara.

Sebanyak 27 fitur dikategorikan sebagai *Accepted*, yang berarti fitur-fitur tersebut dianggap secara signifikan memberikan kontribusi yang stabil dan konsisten terhadap target variabel selama proses iteratif Boruta. Semua fitur yang diterima memiliki nilai importansi yang identik sebesar 3,21%, yang menunjukkan bahwa kontribusi relatif mereka terhadap model prediksi berada di atas ambang batas yang ditetapkan berdasarkan distribusi *shadow features* atau fitur bayangan. *Shadow features* merupakan salinan acak dari fitur asli yang digunakan sebagai *baseline* acak untuk pengujian signifikansi dalam algoritma Boruta (Kursa & Rudnicki, 2010).

Sementara itu, terdapat satu fitur dengan status *Tentative*, yakni “Medan”, yang menunjukkan bahwa fitur tersebut memiliki tingkat kontribusi yang tidak dapat dipastikan secara signifikan lebih baik atau lebih buruk dibanding *shadow features*. Dalam praktiknya, fitur *Tentative* memerlukan peninjauan atau pengujian lanjutan untuk menentukan status akhir, terutama dalam kasus di mana distribusi importance berada di sekitar batas keputusan algoritma.

Hal yang menarik ditemukan pada kategori *Rejected*, di mana terdapat sejumlah fitur yang justru memiliki nilai importansi yang tinggi, bahkan melebihi fitur-fitur dalam kategori *Accepted*. Contohnya, fitur “Bromo Jawa Timur” dan “Pacitan” memiliki nilai importance sebesar 9,74%, serta fitur-fitur seperti “Hotel di Malang” dan “Tiket Bus” memiliki nilai masing-masing sebesar 9,45% dan

8,93%. Meski demikian, Boruta tetap mengelompokkan fitur-fitur ini ke dalam kategori *Rejected*.

Kondisi tersebut terjadi karena algoritma Boruta tidak hanya mempertimbangkan besarnya nilai importansi dalam satu atau dua iterasi, tetapi lebih mengutamakan stabilitas kontribusi fitur dibandingkan *shadow features* selama seluruh proses iterasi. Artinya, fitur yang sesekali menunjukkan importansi tinggi namun tidak secara konsisten lebih informatif dari *shadow features* akan tetap ditolak. Hal ini merupakan mekanisme Boruta dalam mencegah *overfitting* dan menjaga kualitas generalisasi model, terutama pada data yang bersifat kompleks dan temporal seperti data deret waktu (*time-series*). Dengan kata lain, nilai importansi yang tinggi namun tidak stabil tidak cukup untuk menjamin sebuah fitur diterima atau dinilai penting oleh Boruta dalam seleksi akhir.

Temuan ini mengindikasikan bahwa nilai importansi absolut bukanlah satu-satunya indikator relevansi fitur dalam model prediktif. Sebaliknya, aspek konsistensi, kestabilan kontribusi, dan hubungan multivariat dengan fitur lain menjadi faktor yang tidak kalah penting dalam menentukan kelayakan fitur.

4.5.2 Data Preprocessing – Train Test Split

Tahapan ini merupakan proses di mana dataset dibagi menjadi dua bagian utama, yaitu data pelatihan (*training data*) dan data pengujian (*testing data*). Data pelatihan digunakan untuk membangun dan melatih model prediksi, sedangkan data pengujian digunakan untuk menguji sejauh mana model yang telah dilatih mampu melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

Proses split data ini dilakukan sebelum tahap standardisasi atau normalisasi, dengan tujuan untuk menjaga kemurnian data uji dari informasi yang mungkin bocor selama proses pelatihan. Hal ini penting agar evaluasi terhadap model benar-benar mencerminkan kemampuannya dalam menghadapi data baru secara independen.

Dalam penelitian ini, digunakan rasio pembagian 80% untuk data pelatihan dan 20% untuk data pengujian. Setelah melalui tahapan preprocessing, jumlah total data yang tersedia adalah sebanyak 207 baris, sehingga data dibagi menjadi 166 baris untuk pelatihan dan 41 baris untuk pengujian. Perincian pembagian ini disajikan pada tabel berikut.

Tabel 4. 3 pemisahan data

Perbandingan	Data Pelatihan	Data Pengujian
8:2	166	41

Setelah proses pembagian data dilakukan, tahap selanjutnya adalah normalisasi data.

4.5.3 *Data Preprocessing* – Normalisasi

Proses ini bertujuan untuk menyamakan skala nilai pada setiap fitur agar model pembelajaran mesin dapat bekerja secara optimal. Normalisasi sangat penting karena algoritma machine learning umumnya lebih sensitif terhadap perbedaan skala antar atribut, sehingga perbedaan yang terlalu besar dapat menyebabkan model bias terhadap fitur tertentu.

Melalui proses normalisasi, nilai-nilai numerik dalam dataset akan diubah ke dalam rentang tertentu — dalam hal ini, rentang 0 hingga 1. Penyesuaian ini dilakukan agar seluruh fitur memiliki kontribusi yang seimbang dalam proses

pelatihan model. Normalisasi juga dapat mempercepat proses konvergensi dan meningkatkan stabilitas pelatihan.

Berikut ini disajikan contoh hasil data yang telah melalui proses normalisasi.

Date	Visitors	surabaya airport	bromo travel	surabaya jakarta	bromo.1	fried rice	indonesia currency	indonesia visa	visa for indonesia	visa to indonesia	malang hotel
01/01/2008	927	0	0	0	0	5,55556E+15	3,33333E+15	1,26582E+16	4,54545E+15	4,7619E+16	1,02041E+16
01/02/2008	1328	0	0	0	2,22222E+16	2,77778E+15	1,11111E+16	1,01266E+16	4,54545E+15	4,7619E+16	0
01/03/2008	1686	9,89011E+15	0	1,0989E+16	3,33333E+15	2,77778E+15	5,55556E+16	1,13924E+16	4,54545E+15	4,7619E+16	1,02041E+16
01/04/2008	1669	1,0989E+15	0	3,2967E+15	6,66667E+15	0	3,33333E+15	8,86076E+15	9,09091E+15	9,52381E+15	1,02041E+16
01/05/2008	1697	1,0989E+15	0	3,2967E+15	6,66667E+15	0	3,33333E+15	1,64557E+16	9,09091E+15	9,52381E+15	1,02041E+16
01/06/2008	1607	0	0	4,3956E+16	4,44444E+15	2,77778E+15	2,22222E+16	1,64557E+16	9,09091E+15	9,52381E+15	2,04082E+15
01/07/2008	1952	9,89011E+15	0	5,49451E+16	1,11111E+16	2,77778E+15	2,22222E+16	1,77215E+16	9,09091E+15	4,7619E+16	2,04082E+15
01/08/2008	2048	1,31868E+16	0	5,49451E+16	4,44444E+15	2,77778E+15	5,55556E+16	1,26582E+16	9,09091E+15	9,52381E+15	2,04082E+15
01/09/2008	2109	0	3E+16	3,2967E+15	6,66667E+15	5,55556E+15	3,33333E+15	1,26582E+16	9,09091E+15	1,42857E+16	2,04082E+15
01/10/2008	2432	8,79121E+15	0	4,3956E+16	6,66667E+15	2,77778E+15	6,66667E+15	1,26582E+16	0	9,52381E+15	1,02041E+16
01/11/2008	2032	1,20879E+15	0	5,49451E+16	6,66667E+15	2,77778E+15	4,44444E+15	1,13924E+16	4,54545E+15	4,7619E+16	3,06122E+15
01/12/2008	2308	0	0	5,49451E+16	3,33333E+15	2,77778E+15	2,22222E+16	8,86076E+15	9,09091E+15	4,7619E+16	4,08163E+15
01/01/2009	1981	7,69231E+15	3E+16	8,79121E+16	3,33333E+15	1,11111E+15	2,22222E+16	1,26582E+16	9,09091E+15	4,7619E+16	2,04082E+15
01/02/2009	1956	1,20879E+15	0	7,69231E+15	6,66667E+15	8,33333E+15	3,33333E+15	1,13924E+16	9,09091E+15	4,7619E+16	3,06122E+15
01/03/2009	2274	7,69231E+15	0	9,89011E+15	7,77778E+15	8,33333E+15	3,33333E+15	1,26582E+16	1,36364E+16	1,42857E+16	2,04082E+15
01/04/2009	2538	1,31868E+16	0	5,49451E+16	7,77778E+15	5,55556E+15	4,44444E+15	1,51899E+16	9,09091E+15	4,7619E+16	3,06122E+15
01/05/2009	2543	1,20879E+15	3E+16	1,0989E+16	8,88889E+15	5,55556E+15	4,44444E+15	1,64557E+16	1,36364E+16	9,52381E+15	3,06122E+15
01/06/2009	2393	1,20879E+15	0	1,31868E+16	7,77778E+15	5,55556E+15	6,66667E+15	1,77215E+16	1,36364E+16	9,52381E+15	6,12245E+16
01/07/2009	2970	1,0989E+15	3E+16	1,0989E+16	6,66667E+15	8,33333E+15	8,88889E+15	1,64557E+16	1,81818E+16	9,52381E+15	5,10204E+15
01/08/2009	3559	1,86813E+16	3E+16	1,0989E+16	5,55556E+16	8,33333E+15	5,55556E+16	1,64557E+16	1,36364E+16	9,52381E+15	3,06122E+15
01/09/2009	2633	1,64835E+16	0	1,0989E+16	6,66667E+15	5,55556E+15	3,33333E+15	1,39241E+16	1,81818E+16	9,52381E+15	3,06122E+15

Gambar 4. 1 Contoh normalisasi data fitur

4.6 Modeling

Algoritma Long Short-Term Memory (LSTM) merupakan salah satu metode pembelajaran terawasi (supervised learning) yang secara khusus dirancang untuk mengolah data deret waktu (time series). Dalam penelitian ini, dilakukan proses pemodelan dengan tiga pendekatan berbeda sesuai dengan desain sistem yang telah dirumuskan sebelumnya, yaitu:

- Model prediksi berbasis LSTM dengan pendekatan univariate, yang hanya menggunakan satu variabel sebagai input,
- Model LSTM multivariate yang menggunakan seluruh fitur atau variabel yang tersedia, serta

- Model LSTM dengan seleksi fitur Boruta, yaitu pendekatan yang mengombinasikan algoritma Boruta untuk pemilihan fitur paling relevan sebelum data dimasukkan ke dalam model LSTM.

Perbandingan antara model LSTM dengan dan tanpa penerapan Boruta dilakukan untuk mengevaluasi sejauh mana penggunaan metode seleksi fitur dapat meningkatkan performa prediksi, khususnya dalam konteks peramalan jumlah kedatangan wisatawan. Semua tahapan praproses data (data preparation) dilakukan secara seragam untuk seluruh pendekatan, dengan satu-satunya perbedaan terletak pada penerapan algoritma Boruta yang dilakukan pada tahap seleksi fitur.

Adapun hasil pemodelan dalam bentuk prediksi jumlah kedatangan wisatawan ditampilkan dalam bentuk numerik untuk sembilan bulan ke depan, dimulai dari bulan setelah data aktual terakhir hingga akhir tahun 2025. Hasil ini akan digunakan untuk menilai kemampuan model dalam memproyeksikan tren kunjungan wisata secara jangka pendek-menengah..

Tabel 4. 4 Hasil prediksi jumlah wisatawan

Apr'25	Mei'25	Jun'25	Jul '25	Agust'25	Sep '25	Okt'25	Nov'25	Des '25
25,144	25,231	24,923	25,114	24,083	21,906	20,230	18,994	19,279

Dalam pemodelan saat ini yang bisa diproses oleh model hanya model LSTM univariate dikarenakan LSTM univariate bekerja berdasarkan variabel tunggal yaitu data time-series jumlah kedatangan wisatawan di tahun atau bulan-bulan sebelumnya. Sedangkan LSTM multivariate menggunakan variabel-variabel dalam bulan berjalan dalam hal ini fitur kata kunci yang tentunya data tersebut tidak mungkin ada.

4.7 Evaluasi

Evaluasi model dilakukan untuk mengukur kinerja dari ketiga pendekatan algoritma LSTM yang digunakan dalam penelitian ini, yaitu: LSTM Univariate, LSTM Multivariate menggunakan seluruh fitur, dan LSTM Multivariate dengan seleksi fitur menggunakan algoritma Boruta.

Proses evaluasi bertujuan untuk menilai sejauh mana model mampu menghasilkan prediksi yang mendekati nilai aktual, serta untuk membandingkan efektivitas masing-masing pendekatan dalam konteks peramalan jumlah kedatangan wisatawan.

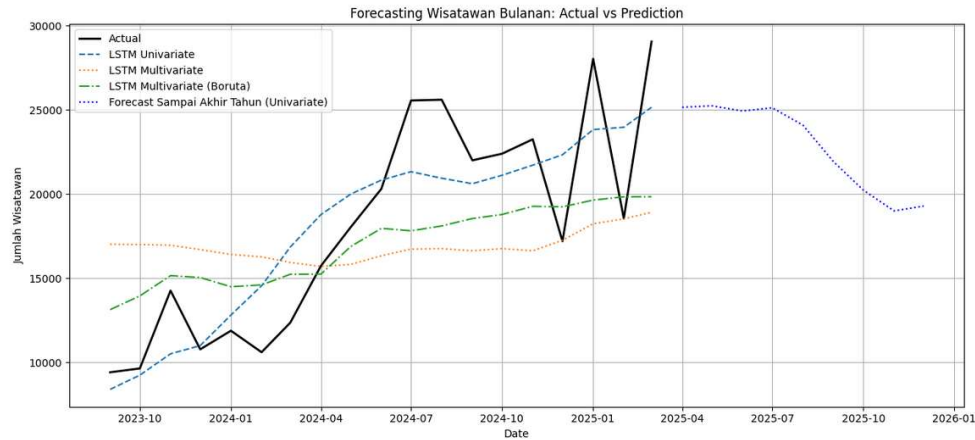
Penilaian performa dilakukan dengan menggunakan tiga metrik evaluasi utama, yaitu:

- Koefisien Determinasi (R^2): Mengukur proporsi variasi dalam data target yang dapat dijelaskan oleh model.
- Mean Absolute Error (MAE): Menunjukkan rata-rata selisih absolut antara nilai aktual dan prediksi.
- Root Mean Squared Error (RMSE): Mengukur rata-rata kesalahan prediksi dalam satuan aslinya, dengan penalti yang lebih besar terhadap kesalahan yang ekstrem.

Hasil dari evaluasi ketiga model disajikan pada Tabel 4.4, serta divisualisasikan dalam bentuk grafik perbandingan pada Gambar 4.3.

Tabel 4. 5 Matrik hasil model

Matrik	LTMS Univariate	LSTM Semua Fitur	LSTM + Boruta
R2	0.732	0.061	0.453
MAE	2737.36	5200.75	3885.49
RMSE	3238.38	5441.76	4628.79



Gambar 4. 2 Grafik hasil prediksi model

Penelitian ini bertujuan untuk mengoptimalkan kinerja model LSTM untuk memprediksi jumlah kedatangan wisatawan menggunakan pendekatan seleksi fitur Boruta. Hipotesis awal yang diajukan adalah bahwa model LSTM Multivariate dengan fitur hasil seleksi Boruta akan memberikan performa prediksi terbaik, karena hanya menggunakan fitur-fitur yang dianggap paling relevan terhadap target, sehingga diharapkan dapat mengurangi kompleksitas dan meningkatkan akurasi.

Berdasarkan hasil pemodelan pada tabel 4.4 dan gambar 4.3, hasil evaluasi model tiga metrik regresi — yakni koefisien determinasi (R^2), Root Mean Squared Error (RMSE), dan Mean Absolute Error (MAE) — menunjukkan bahwa model LSTM Univariate memberikan performa terbaik secara keseluruhan. Model ini mencatatkan R^2 sebesar 0.732, yang berarti sekitar 73,2% variasi jumlah kedatangan wisatawan berhasil dijelaskan oleh model hanya dengan menggunakan data historis sebagai satu-satunya input. Selain itu, nilai RMSE sebesar 3238.38 dan MAE sebesar 2737.36 menunjukkan tingkat galat yang rendah, menandakan bahwa model ini mampu mengikuti pola data secara efektif.

Hasil ini sejalan dengan penelitian oleh (Hsieh, 2021) yang menyatakan bahwa LSTM univariate sangat efektif untuk data deret waktu yang memiliki pola musiman dan tren jangka panjang, karena kemampuan LSTM dalam mempelajari dependensi jangka panjang antar waktu.

Sebaliknya, model LSTM Multivariate dengan seluruh fitur Google Trends menunjukkan performa yang jauh lebih rendah dengan R^2 hanya 0.061, serta galat prediksi yang jauh lebih tinggi (RMSE: 6064.82, MAE: 5200.75). Hal ini mengindikasikan bahwa penambahan banyak fitur tanpa seleksi tidak menjamin peningkatan performa, bahkan berpotensi menyebabkan overfitting atau learning noise yang tidak relevan terhadap target. Hal ini diperkuat oleh temuan (Zhao et al., 2017) dan (Ahmed et al., 2021) yang menjelaskan bahwa model deep learning seperti LSTM sangat sensitif terhadap kualitas dan relevansi fitur, sehingga pemilihan fitur menjadi aspek krusial terutama pada data multivariat.

Sementara itu, model LSTM Multivariate dengan fitur hasil seleksi Boruta menunjukkan hasil yang lebih baik daripada multivariate tanpa seleksi, dengan R^2 sebesar 0.453, RMSE sebesar 4628.79, dan MAE sebesar 3885.49. Algoritma Boruta berhasil mengurangi jumlah fitur dan menyaring fitur yang dinilai penting berdasarkan pendekatan Random Forest. Namun, performanya tetap belum mampu melampaui model univariate. Salah satu kemungkinan penyebabnya adalah Boruta bekerja berdasarkan evaluasi importance statis, tanpa mempertimbangkan hubungan temporal antar fitur. Dalam konteks data deret waktu, pengabaian terhadap struktur urutan waktu bisa berdampak pada penurunan akurasi, sebagaimana dinyatakan oleh (Lim & Zohren, 2021) yang

menekankan pentingnya penggunaan temporal-aware feature selection pada *task time series forecasting*.

Perbedaan antara hipotesis awal dan hasil aktual ini mengindikasikan bahwa meskipun algoritma Boruta dapat meningkatkan efisiensi dengan menyaring fitur yang relevan, pola historis jumlah wisatawan itu sendiri tetap menjadi faktor yang paling dominan dalam membentuk prediksi. Hal ini mendukung pandangan bahwa model sederhana namun fokus (*univariate*) dapat bekerja lebih optimal dibandingkan model kompleks dengan banyak input, terutama jika tidak ditopang oleh pemilihan fitur berbasis dinamika temporal..

4.8 Integrasi Islam

Dalam menilai hasil ini, penting kiranya tidak hanya dilihat dari sisi teknis, tetapi juga melalui pendekatan nilai-nilai Islam yang menekankan keseimbangan hubungan manusia dengan Tuhan, sesama manusia, dan alam sekitar.

4.8.1 *Habluminallah*

Dalam proses membangun dan mengevaluasi model prediksi, penting bagi seorang peneliti Muslim untuk meniatkan aktivitasnya sebagai bentuk ibadah dan pengabdian kepada Allah SWT. Setiap ilmu yang digali dan diterapkan harus membawa seseorang lebih dekat kepada Allah dan berkontribusi pada kemaslahatan umat

Allah SWT berfirman:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا قِيلَ لَكُمْ تَفَسَّحُوا فِي الْمَجَالِسِ فَافْسَحُوا يَفْسَحِ اللَّهُ لَكُمْ وَإِذَا قِيلَ
انْشُرُوا فَانْشُرُوا يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ

Artinya : *Wahai orang-orang yang beriman! Apabila dikatakan kepadamu, "Berilah kelapangan di dalam majelis-majelis," maka lapangkanlah, niscaya Allah akan memberi kelapangan untukmu. Dan apabila dikatakan, "Berdirilah kamu," maka berdirilah, niscaya Allah akan mengangkat (derajat) orang-orang yang*

beriman di antaramu dan orang-orang yang diberi ilmu beberapa derajat. Dan Allah Mahateliti apa yang kamu kerjakan.(Q.S Al Mujadilah : 11)

Menurut Tafsir Ibnu Katsir (Katsir, 2003), ayat ini menunjukkan kemuliaan ilmu dan keutamaan orang yang memilikinya, karena ilmu adalah cahaya petunjuk yang menyampaikan seseorang kepada ridha Allah. Al-Qurṭubi (Syaiikh Imam Al Qurthubi, 2015) menambahkan bahwa ayat ini menunjukkan keutamaan mencari ilmu dan pengaruhnya terhadap kedudukan seseorang di sisi Allah. Ayat ini bersesuaian dengan QS. Adz-Dzariyat: 56.

وَمَا خَلَقْتُ الْجِنَّ وَالْإِنْسَ إِلَّا لِيَعْبُدُونِ

Artinya : *Aku tidak menciptakan jin dan manusia melainkan agar mereka beribadah kepada-Ku.*(Q.S Adz-Dzariyat : 56)

Di mana dalam proses pengolahan data dan pembuatan model prediksi, niat yang ikhlas dan orientasi kepada Allah SWT menjadi fondasi utama. Penelitian ini harus menjadi bagian dari ibadah jika ditujukan untuk kebaikan umat dan dilakukan dengan niat *lillah*. Menurut Tafsir Ibnu Katsir (Katsir, 2003), makna “*liya’budun*” ialah *liyuwahhiduni* (untuk mentauhidkan-Ku), artinya bahwa semua aktivitas manusia, termasuk intelektual dan akademik, harus diarahkan untuk menguatkan nilai *tauhid* dan ketaatan kepada Allah. Imam al-Qurṭubi (Syaiikh Imam Al Qurthubi, 2015) menambahkan bahwa segala amal yang diniatkan untuk mendekatkan diri kepada Allah, meskipun bersifat duniawi, termasuk dalam kategori ibadah.

4.8.2 Habluminannas

Model prediksi yang baik bertujuan memberi kemaslahatan bagi banyak pihak, seperti pemerintah, pelaku wisata, dan masyarakat lokal. Islam sangat

menekankan pentingnya tolong-menolong dalam hal yang membawa manfaat.

Allah SWT berfirman :

وَتَعَاوَنُوا عَلَى الْبِرِّ وَالتَّقْوَىٰ وَلَا تَعَاوَنُوا عَلَى الْإِثْمِ وَالْعُدْوَانِ ۚ وَاتَّقُوا اللَّهَ ۚ إِنَّ اللَّهَ شَدِيدُ الْعِقَابِ

Artinya : *Dan tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan permusuhan. Bertakwalah kepada Allah, sungguh, Allah sangat berat siksaan-Nya.*(Q.S Al Maidah: 2)

Menurut Tafsir Ibnu Katsir (Katsir, 2003), ayat ini juga menunjukkan bahwa segala bentuk kerja sama yang membawa manfaat dan tidak melanggar syariat adalah bentuk kebaikan yang diperintahkan Allah. Maka penelitian ini selayaknya diarahkan untuk mempermudah pengambilan kebijakan di sektor pariwisata yang memberi dampak ekonomi dan sosial kepada masyarakat.

Hadits Nabi SAW juga mendukung pentingnya kebermanfaatan ilmu:

حَيْرُ النَّاسِ أَنْفَعُهُمْ لِلنَّاسِ

Artinya : *“Sebaik-baik manusia adalah yang paling bermanfaat bagi manusia lain.”*.(HR. Ahmad, no. 23408)

Memberikan hasil penelitian yang dapat digunakan untuk meningkatkan kesejahteraan masyarakat termasuk bagian dari amal saleh dalam konteks ini, hasilnya bisa menjadi dasar kebijakan strategis yang berdampak langsung bagi masyarakat, terutama pelaku UMKM dan pengelola destinasi wisata

4.8.3 *Habluminal’alam / (mu’amalah)*

Aspek *mu’amalah* atau *habluminal-‘alam* menyoroti hubungan manusia dengan alam semesta sebagai bagian dari tanggung jawab ilahiyah. Allah SWT menciptakan bumi dan segala isinya sebagai amanah yang harus dijaga, dimanfaatkan secara bijak, dan dilestarikan. Islam tidak memisahkan aspek ibadah vertikal dari tanggung jawab horizontal terhadap makhluk dan lingkungan. Oleh

karena itu, pelestarian lingkungan adalah bagian dari ibadah dan etika kekhalifahan.

Prediksi jumlah wisatawan yang akurat tidak hanya memberikan manfaat praktis dalam pengelolaan destinasi, tetapi juga berdampak pada keberlangsungan lingkungan. Ketika kunjungan wisata tidak terkendali, potensi kerusakan terhadap alam, budaya lokal, dan ekosistem akan meningkat. Model prediktif berbasis data seperti yang dikembangkan dalam penelitian ini membantu mengantisipasi hal tersebut, sehingga keputusan strategis dapat diambil secara lebih proporsional dan berkelanjutan. Allah SWT berfirman:

وَلَا تُفْسِدُوا فِي الْأَرْضِ بَعْدَ إِصْلَاحِهَا وَادْعُوهُ خَوْفًا وَطَمَعًا^{٥٦}

Artinya: *“Dan janganlah kamu berbuat kerusakan di bumi setelah (Allah) memperbaikinya. Berdoalah kepada-Nya dengan rasa takut dan penuh harap.”* (Q.S. Al-A'raf: 56)

Menurut Tafsir Al-Qurtubi (Syaiikh Imam Al Qurthubi, 2015), ayat ini merupakan larangan keras untuk melakukan kerusakan baik secara fisik maupun moral setelah Allah menjadikan bumi dalam kondisi baik dan layak huni. (Imamudin, 2012) mempertegas bahwa ayat ini menunjukkan perhatian Islam terhadap pelestarian lingkungan jauh sebelum munculnya kesadaran ekologi dalam sains modern. Ia menyatakan bahwa Allah SWT telah memperbaiki bumi dari kondisi rusak, sehingga manusia sebagai khalifah tidak boleh merusaknya kembali dengan keserakahan dan ketidakseimbangan.

Islam memberikan perhatian khusus terhadap keseimbangan (*mizan*), baik dalam aspek sosial, ekonomi, maupun ekologi. (Imamudin, 2012) menyebut bahwa ketidakseimbangan lingkungan seperti polusi, eksploitasi sumber daya, dan kerusakan ekosistem adalah akibat dari kesalahan perilaku manusia. Oleh

karena itu, pelestarian lingkungan dalam Islam tidak sekadar etika, tetapi bagian dari *maqāṣid al-syari'ah*, yakni menjaga keberlangsungan hidup (*ḥifẓ al-bi'ah*).

Sebagai bentuk solusi, penelitian ini sejalan dengan prinsip tersebut, karena menyediakan alat bantu analitik bagi para pengambil kebijakan untuk merumuskan strategi pengelolaan wisata yang tidak merusak lingkungan. Dengan demikian, pendekatan *ḥabluminal 'alam* dalam penelitian ini bukan hanya bersifat teknis, tetapi juga membawa nilai moral dan spiritual, yaitu menjaga amanah Allah atas bumi dan seluruh isinya demi kebaikan generasi kini dan masa depan.

BAB V

PENUTUP

5.1 Kesimpulan

Penelitian ini bertujuan untuk memprediksi jumlah wisatawan mancanegara ke Indonesia khususnya ke Jawa Timur menggunakan tiga pendekatan model Long Short-Term Memory (LSTM) yaitu LSTM Univariate, LSTM Multivariate (semua fitur), dan LSTM Multivariate dengan seleksi fitur Boruta. Hasil evaluasi menunjukkan bahwa :

1. LSTM Univariate menghasilkan performa prediksi terbaik dengan R^2 sebesar 0.732, RMSE sebesar 3238.38, dan MAE sebesar 2737.36, mengindikasikan bahwa data historis jumlah wisatawan memiliki informasi temporal yang cukup kuat untuk melakukan prediksi tanpa tambahan fitur eksternal..
2. Model LSTM Multivariate dengan semua fitur menunjukkan performa terendah ($R^2 = 0.061$), yang menunjukkan bahwa penambahan fitur yang tidak diseleksi secara optimal dapat menyebabkan noise dan menurunkan akurasi model.
3. Model LSTM Multivariate dengan seleksi fitur Boruta memberikan hasil menengah ($R^2 = 0.453$), menunjukkan bahwa Boruta dapat meningkatkan akurasi dibandingkan multivariate biasa, tetapi masih belum melampaui performa model univariate. Hal ini disebabkan oleh keterbatasan Boruta dalam mempertimbangkan hubungan temporal antar fitur.

Dengan demikian dapat disimpulkan bahwa model LSTM univariate memberikan performa prediksi terbaik dalam memodelkan jumlah kedatangan wisatawan bulanan, dengan nilai R^2 tertinggi dan tingkat kesalahan yang paling rendah. Hal ini menunjukkan bahwa pola historis tunggal memiliki daya prediksi yang sangat kuat dalam konteks deret waktu.

Namun demikian, pendekatan LSTM multivariate dengan fitur hasil seleksi Boruta menunjukkan perbaikan performa yang cukup signifikan dibandingkan model multivariate tanpa seleksi fitur. Ini menunjukkan bahwa algoritma Boruta masih memiliki peran penting dalam meningkatkan akurasi model dengan menyaring fitur yang tidak relevan, meskipun belum mampu mengungguli model univariate. Oleh karena itu, pendekatan Boruta dapat tetap dipertimbangkan dalam skenario multivariat yang lebih kompleks.

5.2 Saran

Untuk penelitian selanjutnya, disarankan agar proses seleksi fitur dilakukan dengan mempertimbangkan sifat temporal dari data deret waktu, misalnya dengan mengintegrasikan pendekatan *attention mechanism* atau *domain-based temporal filtering*. Selain itu, pengujian model pada skala waktu yang lebih granular (harian atau mingguan) dapat memberikan perspektif baru dalam evaluasi performa.

Penggunaan algoritma Boruta sebaiknya tidak sepenuhnya diabaikan, mengingat kemampuannya dalam meningkatkan model multivariate secara signifikan. Penelitian lanjutan dapat mengeksplorasi kombinasi Boruta dengan teknik deep learning lain seperti Temporal Convolutional Networks (TCN) atau

Transformer, untuk mengevaluasi apakah seleksi fitur tetap relevan pada arsitektur yang lebih kompleks.

DAFTAR PUSTAKA

- Ahmed, A. A. M., Deo, R. C., Ghahramani, A., Raj, N., Feng, Q., Yin, Z., & Yang, L. (2021). LSTM integrated with Boruta-random forest optimiser for soil moisture estimation under RCP4.5 and RCP8.5 global warming scenarios. In *Stochastic Environmental Research and Risk Assessment* (Vol. 35, Issue 9). Springer Berlin Heidelberg. <https://doi.org/10.1007/s00477-021-01969-3>
- Andariesta, D. T., & Wasesa, M. (2022). Machine learning models for predicting international tourist arrivals in Indonesia during the COVID-19 pandemic: a multisource Internet data approach. *Journal of Tourism Futures*, 1–17. <https://doi.org/10.1108/JTF-10-2021-0239>
- Arif, Y. M., Putra, D. D., & Khan, N. (2023). Selecting Tourism Site Using 6 As Tourism Destinations Framework Based Multi-Criteria Recommender System. *Applied Information System and Management (AISM)*, 6(1), 7–12. <https://doi.org/10.15408/aism.v6i1.25140>
- Arif, Y. M., Wardani, D., Nurhayati, H., & Diah, N. M. (2023). Non-Rating Recommender System for Choosing Tourist Destinations Using Artificial Neural Network. *Applied Information System and Management (AISM)*, 6(2), 61–68. <https://doi.org/10.15408/aism.v6i2.26741>
- Becken, S. (2013). Developing a framework for assessing resilience of tourism sub-systems to climatic factors. *Annals of Tourism Research*, 43, 506–528. <https://doi.org/10.1016/j.annals.2013.06.002>
- Botha, I., & Saayman, A. (2024). Does Google Analytics Improve the Prediction of Tourism Demand Recovery? *Forecasting*, 6(4), 908–924. <https://doi.org/10.3390/forecast6040045>
- Bouhaddour, S., Sbihi, M., Guerouate, F., & Saadi, C. (2024). Assessing Tourism Prediction Models: A Comparative Study of SARIMA, Random Forest, and LSTM, Considering Nonlinear Trends and the Influence of COVID-19. *Ingenierie Des Systemes d'Information*, 29(6), 2443–2454. <https://doi.org/10.18280/isi.290631>
- Bufalo, M., & Orlando, G. (2024). Improved tourism demand forecasting with CIR# model: a case study of disrupted data patterns in Italy. *Tourism Review*, 79(2), 445–464. <https://doi.org/10.1108/TR-04-2023-0230>
- Dinis, G., Breda, Z., Costa, C., & Pacheco, O. (2019). Google Trends in tourism and hospitality research: a systematic literature review. *Journal of Hospitality and Tourism Technology*, 10(4), 747–763. <https://doi.org/10.1108/JHTT-08-2018-0086>
- Gholipour, H. F., Tajaddini, R., & Nguyen, J. (2016). Happiness and inbound tourism. *Annals of Tourism Research*, 57, 251–253.

<https://doi.org/10.1016/j.annals.2015.12.003>

- Gunter, U., Smeral, E., & Zekan, B. (2024). Forecasting Tourism in the EU after the COVID-19 Crisis. *Journal of Hospitality and Tourism Research*, 48(5), 909–919. <https://doi.org/10.1177/10963480221125130>
- Haqiq, A., & Pharmasetiawan, B. (2019). Data Analytics for Forecasting Arrival of Tourism Visit in Indonesia. *Proceeding - 2019 International Conference on ICT for Smart Society: Innovation and Transformation Toward Smart Region, ICISS 2019*, 1–6. <https://doi.org/10.1109/ICISS48059.2019.8969795>
- Hochreiter, S., & Schmidhuber, J. (1997). LSTM paper. *Neural Computation*, 9(8), 1735–1780.
- Hsieh, S. C. (2021). Tourism demand forecasting based on an lstm network and its variants. *Algorithms*, 14(8). <https://doi.org/10.3390/a14080243>
- Imamudin, M. (2012). Lingkungan Dan Pelestariannya Dalam Prespektif Islam. *El-Hayah*, 2(1), 21–30. <https://doi.org/10.18860/elha.v2i1.1790>
- Ishaq, A. bin M. bin A. bin. (2013). *Tafsir Ibnu Katsir 8.1.pdf* (pp. 1–113).
- Jiang, D., Liu, J., Sirisrisakulchai, J., & Sriboonchitta, S. (2021). Forecasting Thailand inbound tourist flow association for tourism demand. *Journal of Physics: Conference Series*, 1978(1). <https://doi.org/10.1088/1742-6596/1978/1/012063>
- Karbasi, M., Ali, M., Bateni, S. M., Jun, C., Jamei, M., & Yaseen, Z. M. (2024). Boruta extra tree-bidirectional long short-term memory model development for Pan evaporation forecasting: Investigation of arid climate condition. *Alexandria Engineering Journal*, 86(July 2023), 425–442. <https://doi.org/10.1016/j.aej.2023.11.061>
- Katsir, I. (2003). *Tafsir Ibnu Katsir 3.1.pdf*. In *Pustaka Imam AS'syafi'i*.
- Kulshrestha, A., Krishnaswamy, V., & Sharma, M. (2020). Bayesian BILSTM approach for tourism demand forecasting. *Annals of Tourism Research*, 83(April), 102925. <https://doi.org/10.1016/j.annals.2020.102925>
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the boruta package. *Journal of Statistical Software*, 36(11), 1–13. <https://doi.org/10.18637/jss.v036.i11>
- Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research*, 75(January), 410–423. <https://doi.org/10.1016/j.annals.2019.01.014>
- Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey.

Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379(2194). <https://doi.org/10.1098/rsta.2020.0209>

- Masrur Ahmed, A. A., Deo, R. C., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., & Yang, L. (2021). Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. *Journal of Hydrology*, 599(April), 126350. <https://doi.org/10.1016/j.jhydrol.2021.126350>
- Nagy, B., Gabor, M. R., & Bacoş, I. B. (2022). Google Mobility Data as a Predictor for Tourism in Romania during the COVID-19 Pandemic—A Structural Equation Modeling Approach for Big Data. *Electronics (Switzerland)*, 11(15). <https://doi.org/10.3390/electronics11152317>
- Nie, R. X., Wu, C., & Liang, H. M. (2024). Exploring Appropriate Search Engine Data for Interval Tourism Demand Forecasting Responding a Public Crisis in Macao: A Combined Bayesian Model. *Sustainability (Switzerland)*, 16(16). <https://doi.org/10.3390/su16166892>
- Pabuccu, H., & Barbu, A. (2023). Feature Selection with Annealing for Forecasting Financial Time Series. *Financial Innovation*. <https://doi.org/10.48550/arXiv.2303.02223>
- 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. <https://doi.org/10.1007/s12559-020-09747-z>
- Subbiah, S. S., & Kumar, S. P. (2022). Deep Learning Based Load Forecasting with Decomposition and Feature Selection Techniques. *Journal of Scientific and Industrial Research*, 81(5), 505–517. <https://doi.org/10.56042/jsir.v81i05.56794>
- Subbiah, S. S., Paramasivan, S. K., Arockiasamy, K., Senthivel, S., & Thangavel, M. (2023). Deep Learning for Wind Speed Forecasting Using Bi-LSTM with Selected Features. *Intelligent Automation and Soft Computing*, 35(3), 3829–3844. <https://doi.org/10.32604/iasc.2023.030480>
- Sugiartawan, P., Pulungan, R., & Kartika, A. (2017). Prediction by a Hybrid of Wavelet Transform and Long-Short-Term-Memory Neural Network. *International Journal of Advanced Computer Science and Applications*, 8(2). <https://doi.org/10.14569/ijacsa.2017.080243>
- Sun, S., Wei, Y., Tsui, K. L., & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management*, 70(February 2018), 1–10. <https://doi.org/10.1016/j.tourman.2018.07.010>
- Syaikh Imam Al Qurthubi. (2015). *Tafsir Al-Qurtubi Jilid 17*. 507.
- Thu, N. T. H., Van, P. N., Nam, N. V. N., Minh, P. H., & Bao, P. Q. (2023).

- Forecasting Wind Speed Using A Hybrid Model Of Convolutional Neural Network And Long-Short Term Memory With Boruta Algorithm-Based Feature Selection. *Journal of Applied Science and Engineering*, 26(8), 1055–1062. [https://doi.org/10.6180/jase.202308_26\(8\).0001](https://doi.org/10.6180/jase.202308_26(8).0001)
- Volchek, K., Liu, A., Song, H., & Buhalis, D. (2019). Forecasting tourist arrivals at attractions: Search engine empowered methodologies. *Tourism Economics*, 25(3), 425–447. <https://doi.org/10.1177/1354816618811558>
- Xie, G., Qian, Y., & Wang, S. (2021). Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach. *Tourism Management*, 82(October 2019), 104208. <https://doi.org/10.1016/j.tourman.2020.104208>
- Xu, K., Zhang, J., Huang, J., Tan, H., Jing, X., & Zheng, T. (2024). Forecasting Visitor Arrivals at Tourist Attractions: A Time Series Framework with the N-BEATS for Sustainable Tourism. *Sustainability (Switzerland)*, 16(18). <https://doi.org/10.3390/su16188227>
- Yu, N., & Chen, J. (2022). Design of Machine Learning Algorithm for Tourism Demand Prediction. *Computational and Mathematical Methods in Medicine*, 2022. <https://doi.org/10.1155/2022/6352381>
- Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). LSTM network: A deep learning approach for Short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2), 68–75. <https://doi.org/10.1049/iet-its.2016.0208>

LAMPIRAN