

## OPTIMASI METODE LONG SHORT-TERM MEMORY MENGGUNAKAN SELEKSI FITUR BORUTA UNTUK MEMPREDIKSI KEDATANGAN WISATAWAN

Tumiyar Junaidi<sup>1\*</sup>, M. Amin Hariyadi<sup>2</sup>, Zainal Abidin<sup>3</sup>

<sup>1,2,3</sup>Magister Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim Malang

\*Email Korespondensi: [210605210008@student.uin-malang.ac.id](mailto:210605210008@student.uin-malang.ac.id)

Submitted : 21 April 2025; Revision : 23 April 2025; Accepted : 25 April 2025

### ABSTRAK

Peramalan jumlah kedatangan wisatawan mancanegara secara akurat sangat penting untuk mendukung perencanaan strategi dan kebijakan di sektor pariwisata, khususnya dalam masa pemulihan pasca pandemi. Penelitian ini membandingkan tiga pendekatan pemodelan Long Short-Term Memory (LSTM): univariate, multivariate dengan semua fitur, serta multivariate dengan seleksi fitur menggunakan algoritma Boruta. Data yang digunakan mencakup jumlah kedatangan wisatawan bulanan ke Indonesia dari tahun 2008 hingga 2025 dan indeks pencarian Google Trends yang relevan. Hasil penelitian menunjukkan bahwa model LSTM univariate memiliki performa terbaik dengan nilai  $R^2 = 0.552$ , sedangkan pendekatan multivariate dengan fitur hasil seleksi Boruta justru menunjukkan performa lebih buruk daripada model naïf dengan nilai  $R^2 = -0.913$ . Temuan ini menyoroti bahwa penambahan fitur tanpa mempertimbangkan sifat temporal dapat menurunkan akurasi prediksi, dan bahwa model sederhana dengan fokus tunggal lebih efektif untuk data deret waktu. Studi ini memberikan kontribusi praktis dan teoritis dalam pengembangan sistem prediksi wisatawan berbasis AI yang lebih kontekstual.

**Kata kunci :** LSTM, Boruta, kedatangan wisatawan, seleksi fitur, Google Trends.

### ABSTRACT

*Accurate forecasting of international tourist arrivals is crucial for strategic planning and policy-making, particularly during the post-pandemic recovery phase. This study compares three Long Short-Term Memory (LSTM) modeling approaches: univariate, multivariate with all features, and multivariate with feature selection using the Boruta algorithm. The dataset includes monthly tourist arrival records in Indonesia (2008–2025) and tourism-related search indices from Google Trends. The results show that the univariate LSTM model achieved the best performance with an  $R^2$  value of 0.552, while the multivariate approach with Boruta-selected features exhibited worse performance than a naïve model, with an  $R^2$  value of -0.913. These findings highlight that adding features without considering their temporal relevance may reduce predictive accuracy, and that simpler, single-focus models tend to be more effective for time series data. This study offers both practical and theoretical contributions to the development of more context-aware AI-based tourist forecasting systems.*

**Keywords :** LSTM, Boruta, tourist arrival, feature selection, Google Trends.

### PENDAHULUAN

Industri perjalanan lintas negara merupakan komponen vital dalam sektor pariwisata global dan berperan signifikan dalam mendukung perekonomian dunia. Pada tahun 2018, sektor ini menyumbang sekitar 10,4% dari Produk Domestik Bruto (GDP) global, namun pandemi COVID-19 menyebabkan penurunan drastis hingga hanya 3% pada tahun 2020, berdampak langsung terhadap pendapatan dan keberlangsungan para pelaku industri pariwisata (Kulshrestha, Krishnaswamy and Sharma, 2020; Subbiah and Kumar, 2022). Seiring dengan pemulihan pasca pandemi, kontribusi sektor ini mulai meningkat kembali mencapai 5,25%, menandakan pentingnya ketepatan dalam penentuan kebijakan

---

pemerintah dan strategi bisnis untuk mendukung pemulihan dan pertumbuhan sektor pariwisata secara berkelanjutan (Gunter, Smeral and Zekan, 2024).

Salah satu indikator kunci yang sangat mempengaruhi pendapatan sektor ini adalah jumlah kedatangan wisatawan. Oleh karena itu, kemampuan untuk meramalkan jumlah kunjungan secara akurat menjadi sangat penting bagi perencanaan kebijakan dan strategi yang efektif (Hsieh, 2021). Peramalan permintaan pariwisata biasanya dilakukan menggunakan tiga pendekatan utama: model deret waktu (time series), pendekatan ekonometrik, dan kecerdasan buatan. Meskipun model deret waktu banyak digunakan karena kesederhanaannya, pendekatan ini kurang mampu menangani pola non-linier yang sering ditemukan dalam data pariwisata. Di sinilah kecerdasan buatan, khususnya Long Short Term Memory (LSTM), menunjukkan keunggulannya dengan kemampuannya dalam mengidentifikasi pola kompleks dan dependensi jangka panjang secara lebih akurat (Hsieh, 2021).

LSTM sendiri dikembangkan oleh Hochreiter dan Schmidhuber (Hochreiter and Schmidhuber, 1997) sebagai solusi atas keterbatasan Recurrent Neural Network (RNN), dengan arsitektur yang memungkinkan pemrosesan data berurutan dalam jangka waktu panjang tanpa kehilangan informasi penting. Model ini terbukti efektif dalam berbagai aplikasi, termasuk peramalan musik, suara, dan data ekonomi. Dalam konteks pariwisata, (Hsieh, 2021) menunjukkan bahwa LSTM mampu memberikan prediksi yang lebih unggul dibandingkan model fuzzy time series, terutama dalam kondisi krisis seperti SARS dan COVID-19, serta menunjukkan kinerja konsisten di berbagai negara seperti Taiwan, Amerika Serikat, Australia, dan Singapura. Temuan ini semakin menguatkan bahwa penambahan variabel input seperti data ekonomi, iklim, dan perilaku digital dapat meningkatkan keakuratan model, sehingga pemilihan variabel menjadi aspek penting dalam peramalan.

Pemilihan variabel yang tepat tidak hanya meningkatkan akurasi, tetapi juga mengoptimalkan performa model LSTM, sebagaimana ditegaskan oleh (Kulshrestha, Krishnaswamy and Sharma, 2020). Salah satu metode yang relevan untuk mendukung seleksi variabel ini adalah algoritma Boruta, yang telah terbukti efektif dalam penelitian oleh (Ahmed *et al.*, 2021; Masrur Ahmed *et al.*, 2021) dalam konteks prediksi kelembaban tanah. Dengan memanfaatkan kombinasi LSTM dan Boruta, model prediktif menunjukkan hasil lebih akurat dibandingkan model-model lainnya seperti SVR dan MARS, menunjukkan potensi sinergis dalam penerapan metode ini pada domain lain termasuk peramalan pariwisata.

Boruta sendiri merupakan metode seleksi fitur yang dikembangkan oleh Kursa dan Rudnicki (Kursa and Rudnicki, 2010), yang bekerja berdasarkan prinsip Random Forest untuk mengidentifikasi fitur-fitur paling relevan dalam proses prediksi. Dengan menghasilkan berbagai pohon keputusan dan mengonsolidasikan hasilnya, Boruta mampu meningkatkan keandalan dan presisi model. Berdasarkan mekanisme dan performa yang ditunjukkan dalam studi sebelumnya, sangat mungkin metode ini diadaptasikan untuk meningkatkan performa peramalan kedatangan wisatawan dengan pendekatan LSTM, menjadikannya strategi yang menjanjikan dalam mendukung pengambilan keputusan berbasis data dalam sektor pariwisata. Dari penjelasan diatas bisa diambil hipotesis bahwa Boruta bisa digunakan sebagai seleksi fitur untuk meningkatkan performa prediksi kedatangan wisatawan menggunakan Long-Short-Term Memory.

## METODE

Penelitian ini diawali dengan tahap business understanding, pendekatan analitik, serta identifikasi kebutuhan data. Data utama yang digunakan berupa jumlah kedatangan wisatawan mancanegara ke Indonesia per bulan, yang diperoleh dari Badan Pusat Statistik (BPS) untuk periode Januari 2008 hingga Maret 2025. Sebagai variabel tambahan, digunakan data indeks kata kunci dari Google Trends yang relevan dengan sektor pariwisata, dan dikumpulkan untuk periode waktu yang sama berdasarkan penelitian terdahulu dapat dilihat pada tabel 1. Seluruh data dianalisis terlebih dahulu dalam tahap data understanding untuk memastikan kelayakan dan konsistensinya sebelum digunakan dalam pemodelan.

Pada tahap data preparation, dilakukan preprocessing data mencakup penanganan nilai hilang (*missing values*), normalisasi fitur, dan pemilihan fitur yang relevan menggunakan algoritma Boruta. Algoritma ini membantu dalam memilih fitur-fitur yang paling relevan terhadap variabel target, sehingga dapat meningkatkan akurasi model.

Tahap awal preprocessing dilakukan dengan mengidentifikasi seluruh fitur yang mengandung *missing value*. Hal ini dilakukan melalui fungsi eksplorasi data untuk menghitung jumlah nilai kosong pada setiap kolom. Fitur-fitur yang memiliki *missing value* kemudian dihapus dari dataset untuk menjaga integritas dan validitas model, khususnya karena model berbasis deep learning seperti LSTM sangat sensitif terhadap ketidakkonsistenan input.

Setelah data bersih dari *missing value*, seluruh fitur dengan data numerik dilakukan proses normalisasi dengan menggunakan metode Min-Max Scaling. Metode ini mengubah skala setiap fitur ke dalam rentang [0, 1], sehingga setiap variabel memiliki kontribusi yang setara dalam proses pelatihan model. Normalisasi tidak diterapkan pada kolom date karena merupakan tipe data temporal, serta kolom target visitors karena menjadi variabel yang akan diprediksi.

Setelah nomralisasi data fitur selesai maka fitur-fitur paling relevan dipilih oleh Boruta. Kemudian proses dilanjutkan ke tahap modeling, yaitu pembangunan model prediksi menggunakan metode Long Short-Term Memory (LSTM). LSTM dipilih karena kemampuannya dalam mempelajari pola sekuel dan menangani dependensi jangka panjang dalam data deret waktu. Arsitektur LSTM terdiri dari tiga komponen utama, yakni input gate, forget gate, dan output gate, yang secara sinergis mengatur aliran informasi pada setiap langkah waktu. Setelah model dilatih, dilakukan evaluasi performa berdasarkan metrik prediktif tertentu dan dilakukan analisis terhadap hasil yang diperoleh. Tahap akhir meliputi interpretasi hasil serta penarikan kesimpulan untuk menjawab tujuan penelitian.

Algoritma Boruta merupakan metode seleksi fitur berbasis permutasi yang dirancang untuk mengidentifikasi atribut yang paling relevan dalam prediksi. Algoritma ini bekerja dengan membuat shadow features—salinan acak dari fitur asli yang tidak memiliki korelasi dengan target—sebagai pembanding. Pentingnya suatu fitur diukur berdasarkan nilai feature importance yang diperoleh dari model seperti Random Forest. Nilai ini dibandingkan dengan fitur bayangan menggunakan uji statistik seperti z-score. Fitur dianggap signifikan jika memiliki skor kepentingan yang secara konsisten lebih tinggi dari fitur bayangan, dan proses ini diulang secara iteratif untuk menjamin stabilitas hasil.

Sementara itu, Google Trends dimanfaatkan sebagai sumber data sekunder untuk menangkap dinamika minat publik terhadap isu-isu pariwisata melalui analisis tren pencarian daring. Layanan ini mengolah data pencarian berdasarkan kata kunci yang dimasukkan oleh pengguna, dan menyajikannya dalam bentuk grafik deret waktu, diagram, maupun tabel. Visualisasi ini memungkinkan analisis temporal terhadap perilaku pencarian

masyarakat, yang kemudian dapat dikaitkan dengan pola kedatangan wisatawan sebagai variabel target dalam pemodelan.

Tabel 1. Kata kunci mesin pencari penelitian terdahulu

Penulis	Sumber	Topik	Kata Kunci
(Andariesta and Wasesa, 2022)	Google Trends	main entry point  international travel requirement  tourism planning	ngurah rai international airport soekarno-hatta international airport batam ferry terminal bali jakarta passport indonesia visa indonesia indonesia hotel indonesia resort indonesia restaurant indonesia travel
(Dinis <i>et al.</i> , 2019)	Google Trends	travel	travel destination country
(Nagy, Gabor and Bacoş, 2022)	Google Mobility	Retail and recreation  Grocery and Pharmacy  Parks  transit Stations  workplace  residence	restaurants cafes shoping centers theme parks museums libraries movie theaters grocery markets, food warehouses, farmers markets specialty food shops, drug stores, pharmachies 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 and Saayman, 2024)	Google Trends		flights, accomodation, travel, costs, holiday south african safaris, cape town, johansburg game reserves , the kruger national park
(Haqiq and Pharmasetiawan, 2019)	Google Trends	tourism  dining  traffic	bali indonesia, indonesia package, indonesia package tour indonesia currency, indonesia visa, visa for indonesia,

		lodging shopping weather	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 <i>et al.</i> , 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, Wu and Liang, 2024)	Google Trends	tourism	macao travel, macao hotel, macao, macao airport, macao acomodation
(Volchek <i>et al.</i> , 2019)	Google Trends	travel	british museum, national gallery, natural history museum, tate moder, science museum

## HASIL DAN PEMBAHASAN

Penelitian ini dilakukan untuk mengevaluasi dan mengoptimasi metode Long Short Term Memory (LSTM) dalam memprediksi jumlah kedatangan wisatawan bulanan, dengan membandingkan tiga pendekatan model: LSTM Univariate, LSTM Multivariate dengan seluruh fitur, serta LSTM Multivariate dengan seleksi fitur menggunakan algoritma Boruta. Evaluasi dilakukan menggunakan metrik koefisien determinasi ( $R^2$ ), Root Mean Squared Error (RMSE), dan Mean Absolute Error (MAE).

### Hasil Evaluasi Model

Model LSTM Univariate menunjukkan performa prediktif terbaik dengan nilai  $R^2$  sebesar 0.552, mengindikasikan bahwa lebih dari 55% variansi jumlah wisatawan berhasil dijelaskan oleh model. Nilai RMSE dan MAE yang relatif rendah menunjukkan tingkat galat yang dapat diterima dalam konteks prediksi berbasis waktu.

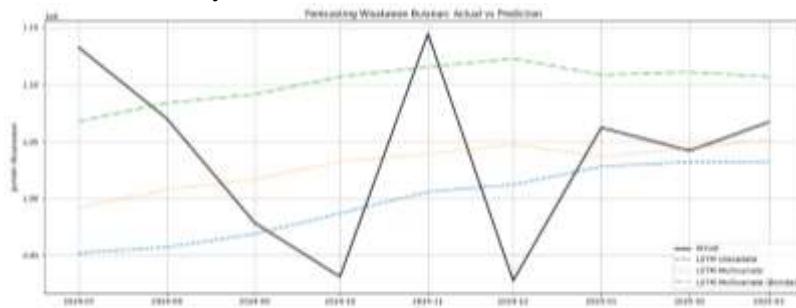
Tabel 2. Metrik evaluasi model

Model	R <sup>2</sup>	RMSE	MAE
LSTM Univariate	0.552	92,732.78	73,429.89
LSTM Multivariate (semua fitur)	0.237	82,789.53	67,932.69
LSTM Multivariate Boruta	0.913	102,949.75	82,932.22

Sebaliknya, model LSTM Multivariate dengan semua fitur memiliki nilai R<sup>2</sup> sebesar 0.237, menandakan bahwa meskipun memasukkan lebih banyak variabel, kompleksitas fitur justru menurunkan kemampuan model dalam menangkap pola utama data. Hal ini kemungkinan disebabkan oleh noise atau korelasi lemah antar fitur yang digunakan.

Model LSTM Multivariate dengan fitur hasil seleksi Boruta justru menunjukkan performa terburuk dengan nilai R<sup>2</sup> negatif (-0.913). Hal ini mengindikasikan bahwa model gagal dalam melakukan generalisasi dan prediksi cenderung menyimpang jauh dari nilai aktual. Meskipun Boruta dikenal andal dalam pemilihan fitur berbasis random forest, penerapannya pada data time-series tanpa mempertimbangkan keterkaitan temporal tampaknya kurang tepat. Nilai R<sup>2</sup> yang negatif menunjukkan bahwa kemampuan model dalam melakukan prediksi lebih rendah dibandingkan pendekatan prediksi sederhana berbasis rata-rata (naïf model). Hal ini mengindikasikan bahwa model gagal merepresentasikan variasi dalam data target secara lebih baik dibandingkan dengan hanya menggunakan nilai rata-rata sebagai acuan prediksi.

Gambar 1 memperlihatkan perbandingan visual antara data aktual dan hasil prediksi dari ketiga model. Model univariate paling mendekati tren aktual, sementara dua model multivariate cenderung menghasilkan prediksi yang terlalu halus atau menyimpang dari pola fluktuatif data sebenarnya.



Gambar 1. Perbandingan prediksi masing-masing model

### Pembahasan

Temuan ini memperkuat hipotesis bahwa model prediksi berbasis time-series seperti LSTM tidak selalu mendapatkan keuntungan dari penambahan banyak fitur tanpa seleksi dan validasi yang sesuai. Kompleksitas input dapat menyebabkan overfitting atau kesulitan model dalam mempelajari representasi data yang optimal. Dalam konteks ini, sederhana namun terfokus (seperti univariate) terbukti lebih efektif.

Adapun kegagalan Boruta dalam meningkatkan performa model menunjukkan bahwa metode seleksi fitur berbasis ensemble tree belum tentu cocok untuk pemrosesan data temporal, di mana urutan waktu dan dinamika musiman memiliki peran penting.

### DAMPAK DAN MANFAAT

Penelitian ini memberikan kontribusi signifikan dalam pengembangan metode prediksi kedatangan wisatawan berbasis deep learning, khususnya Long Short Term Memory (LSTM), dengan mengungkap bahwa model univariate lebih unggul

dibandingkan pendekatan multivariate dan seleksi fitur Boruta dalam konteks data deret waktu. Hal ini berdampak pada peningkatan akurasi model prediksi yang dapat diimplementasikan dalam sistem perencanaan pariwisata, serta mendorong pengembangan metodologi seleksi fitur yang lebih sesuai untuk time-series forecasting. Selain itu, temuan ini memberikan wawasan kritis bagi komunitas akademik terkait efektivitas pendekatan multivariat dalam domain peramalan temporal, yang selama ini sering dianggap lebih superior secara umum..

Penelitian ini bermanfaat secara praktis bagi pemerintah daerah dan pelaku industri pariwisata dalam merancang strategi promosi dan pelayanan berbasis prediksi jumlah wisatawan yang lebih akurat. Bagi akademisi dan peneliti, hasil penelitian ini memperkaya literatur dalam bidang machine learning dan forecasting dengan pendekatan LSTM, serta menjadi acuan penting dalam pemilihan fitur dan struktur data yang optimal. Selain itu, hasil penelitian ini dapat dimanfaatkan oleh pengembang sistem informasi untuk membangun dashboard prediksi wisatawan berbasis AI yang lebih andal, serta membantu pengambil kebijakan dalam menyusun langkah strategis dan responsif terhadap fluktuasi kunjungan wisatawan.

## KESIMPULAN

Penelitian ini menunjukkan bahwa model LSTM univariate menghasilkan performa prediksi terbaik dalam memodelkan jumlah kedatangan wisatawan bulanan, dengan nilai  $R^2$  tertinggi dan error terendah dibandingkan model LSTM multivariate dengan semua fitur maupun model dengan seleksi fitur Boruta. Hal ini mengindikasikan bahwa pola historis kunjungan wisatawan secara tunggal (univariate) sudah cukup kuat merepresentasikan dinamika temporalnya, sementara penambahan fitur-fitur lain justru menurunkan akurasi prediksi. Pendekatan seleksi fitur Boruta yang umumnya efektif dalam regresi klasik, ternyata kurang sesuai untuk data deret waktu, sehingga perlu disesuaikan dengan karakteristik temporal untuk mencapai hasil optimal.

Penelitian selanjutnya disarankan untuk mengeksplorasi metode seleksi fitur berbasis temporal yang lebih sesuai dengan karakteristik data time series, seperti attention mechanism atau temporal feature extraction berbasis domain knowledge. Selain itu, pengujian model pada data dengan skala harian atau mingguan dapat memberikan gambaran yang lebih granular dan realistik untuk kebutuhan operasional di sektor pariwisata. Pengembangan sistem prediksi terintegrasi dengan dashboard real-time juga dapat menjadi langkah strategis dalam penerapan hasil penelitian ini untuk mendukung pengambilan keputusan yang cepat dan berbasis data oleh pemerintah daerah maupun pelaku industri wisata.

## REFERENSI

- Ahmed, A.A.M. et al. (2021) *LSTM integrated with Boruta-random forest optimiser for soil moisture estimation under RCP4.5 and RCP8.5 global warming scenarios, Stochastic Environmental Research and Risk Assessment*. Springer Berlin Heidelberg. Available at: <https://doi.org/10.1007/s00477-021-01969-3>.
- Andariesta, D.T. and 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*, pp. 1–17. Available at: <https://doi.org/10.1108/JTF-10-2021-0239>.
- Botha, I. and Saayman, A. (2024) ‘Does Google Analytics Improve the Prediction of Tourism Demand Recovery?’, *Forecasting*, 6(4), pp. 908–924. Available at: <https://doi.org/10.3390/forecast6040045>.
- Dinis, G. et al. (2019) ‘Google Trends in tourism and hospitality research: a systematic

- literature review’, *Journal of Hospitality and Tourism Technology*, 10(4), pp. 747–763. Available at: <https://doi.org/10.1108/JHTT-08-2018-0086>.
- Gunter, U., Smeral, E. and Zekan, B. (2024) ‘Forecasting Tourism in the EU after the COVID-19 Crisis’, *Journal of Hospitality and Tourism Research*, 48(5), pp. 909–919. Available at: <https://doi.org/10.1177/10963480221125130>.
- Haqiq, A. and 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*, pp. 1–6. Available at: <https://doi.org/10.1109/ICISS48059.2019.8969795>.
- Hochreiter, S. and Schmidhuber, J. (1997) ‘LSTM paper’, *Neural Computation*, 9(8), pp. 1735–1780.
- Hsieh, S.C. (2021) ‘Tourism demand forecasting based on an lstm network and its variants’, *Algorithms*, 14(8). Available at: <https://doi.org/10.3390/a14080243>.
- Kulshrestha, A., Krishnaswamy, V. and Sharma, M. (2020) ‘Bayesian BiLSTM approach for tourism demand forecasting’, *Annals of Tourism Research*, 83(April), p. 102925. Available at: <https://doi.org/10.1016/j.annals.2020.102925>.
- Kursa, M.B. and Rudnicki, W.R. (2010) ‘Feature selection with the boruta package’, *Journal of Statistical Software*, 36(11), pp. 1–13. Available at: <https://doi.org/10.18637/jss.v036.i11>.
- Masrur Ahmed, A.A. et al. (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), p. 126350. Available at: <https://doi.org/10.1016/j.jhydrol.2021.126350>.
- Nagy, B., Gabor, M.R. and 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). Available at: <https://doi.org/10.3390/electronics11152317>.
- Nie, R.X., Wu, C. and 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). Available at: <https://doi.org/10.3390/su16166892>.
- Subbiah, S.S. and Kumar, S.P. (2022) ‘Deep Learning Based Load Forecasting with Decomposition and Feature Selection Techniques’, *Journal of Scientific and Industrial Research*, 81(5), pp. 505–517. Available at: <https://doi.org/10.56042/jsir.v81i05.56794>.
- Sun, S. et al. (2019) ‘Forecasting tourist arrivals with machine learning and internet search index’, *Tourism Management*, 70(February 2018), pp. 1–10. Available at: <https://doi.org/10.1016/j.tourman.2018.07.010>.
- Volchek, K. et al. (2019) ‘Forecasting tourist arrivals at attractions: Search engine empowered methodologies’, *Tourism Economics*, 25(3), pp. 425–447. Available at: <https://doi.org/10.1177/1354816618811558>.