



PERBANDINGAN MODEL SVR DAN LSTM BERBASIS SENTIMEN PADA PERAMALAN JUMLAH WISATAWAN DOMESTIK DI PROVINSI BALI

FAUZAN FAJARI



PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan Model SVR dan LSTM Berbasis Sentimen Pada Peramalan Jumlah Wisatawan Domestik di Provinsi Bali” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2026

Fauzan Fajari
G1401221055



ABSTRAK

FAUZAN FAJARI. Perbandingan Model SVR dan LSTM Berbasis Sentimen Pada Peramalan Jumlah Wisatawan Domestik di Provinsi Bali. Dibimbing oleh AKBAR RIZKI dan LA ODE ABDUL RAHMAN.

Peramalan pariwisata merupakan kebutuhan strategis bagi pemerintah daerah dan pemangku kepentingan industri dalam mengelola sumber daya dan merumuskan kebijakan. Penelitian ini membandingkan performa model *Support Vector Regression* (SVR) dan *Long Short-Term Memory* (LSTM) yang diperkaya dengan fitur sentimen yang diekstraksi dari berita daring untuk meramalkan jumlah kunjungan wisatawan domestik bulanan di Provinsi Bali pada periode Januari 2014 hingga Desember 2025. Analisis sentimen dilakukan terhadap 3.569 artikel berita pariwisata menggunakan model bahasa *pre-trained* IndoBERT, menghasilkan rasio positif bulanan sebagai fitur eksogen bersama data historis kunjungan wisatawan dan variabel *dummy* COVID-19. Kedua model dioptimalkan menggunakan GridSearchCV dengan *time series k-fold cross-validation* ($k=5$) dalam skema *sliding window*. Model SVR dengan *hyperparameter* optimal ($C=100$, $\gamma=0,01$, $\epsilon=0,001$) menghasilkan MAPE sebesar 10,52% dan WMAPE sebesar 10,46% pada data uji. Model LSTM dengan konfigurasi optimal ($units=16$, $dropout_rate=0,2$, $batch_size=8$, $epochs=100$) menghasilkan MAPE sebesar 11,00% dan WMAPE sebesar 11,19%. SVR secara konsisten mengungguli LSTM pada data latih maupun data uji, meskipun selisih pada data uji kurang dari 1%. Kedua model menghasilkan MAPE di bawah 15%, ambang batas yang umumnya dianggap memadai untuk keperluan perencanaan pariwisata. Peramalan untuk tahun 2026 mencerminkan pola musiman historis dengan puncak kunjungan diprediksi terjadi pada bulan April dan lembah terendah pada bulan Maret. Temuan ini menunjukkan bahwa SVR dengan fitur yang diperkaya sentimen menawarkan pendekatan yang praktis dan akurat untuk peramalan permintaan pariwisata jangka pendek.

Kata kunci: IndoBERT, Analisis Sentimen, Support Vector Regression, Long-Short Term Memory, Peramalan Pariwisata



ABSTRACT

FAUZAN FAJARI. Comparison of SVR and LSTM Models Based on Sentiment in Forecasting the Number of Domestic Tourists in Bali Province. Supervised by AKBAR RIZKI and LA ODE ABDUL RAHMAN.

Tourism forecasting is a strategic need for regional governments and industry stakeholders in managing resources and formulating policies. This study compares the performance of Support Vector Regression (SVR) and Long Short-Term Memory (LSTM) models augmented with sentiment features derived from online news to forecast monthly domestic tourist arrivals in Bali Province over the period January 2014 to December 2025. Sentiment analysis was conducted on 3,569 tourism-related news articles using the IndoBERT pre-trained language model, generating a monthly positive ratio as an exogenous feature alongside historical visitor data and a COVID-19 dummy variable. Both models were tuned using GridSearchCV with time series k -fold cross-validation ($k=5$) under a sliding window scheme. The SVR model with optimal hyperparameters ($C=100$, $\gamma=0.01$, $\epsilon=0.001$) achieved a MAPE of 10.52% and WMAPE of 10.46% on the test set. The LSTM model with optimal configuration (units=16, dropout_rate=0.2, batch_size=8, epochs=100) yielded a MAPE of 11.00% and WMAPE of 11.19%. SVR consistently outperformed LSTM across both training and test sets, though the margin on test data was less than 1%. Both models produced MAPE below 15%, a threshold generally considered adequate for tourism planning purposes. Forecasts for 2026 reflect historical seasonal patterns, with predicted peaks in April and troughs in March. These findings suggest that SVR with sentiment-augmented features offers a practical and accurate approach to short-term tourism demand forecasting.

Keywords: IndoBERT, Sentiment Analysis, Support Vector Regression, Long-Short Term Memory, Tourism Forecasting



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University



PERBANDINGAN MODEL SVR DAN LSTM BERBASIS SENTIMEN PADA PERAMALAN JUMLAH WISATAWAN DOMESTIK DI PROVINSI BALI

FAUZAN FAJARI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Sarjana Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Perbandingan Model SVR dan LSTM Berbasis Sentimen pada
Peramalan Jumlah Wisatawan Domestik di Provinsi Bali

Nama : Fauzan Fajari
NIM : G1401221055

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Akbar Rizki, S.Stat., M.Si.

Pembimbing 2:
La Ode Abdul Rahman, S.Si., M.Si

Diketahui oleh

Ketua Program Studi:
Dr. Bagus Sartono, S.Si., M.Si.
NIP 197804112005011002

Tanggal Ujian:
24 Juni 2026

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah Subhanaahu wa Ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Februari 2026 sampai bulan Juli 2026 ini ialah perbandingan performa model, dengan judul “Perbandingan Model SVR dan LSTM Berbasis Sentimen pada Peramalan Jumlah Wisatawan Domestik di Provinsi Bali”.

Penulisan karya ilmiah ini tidak lepas dari dukungan dan bimbingan berbagai pihak sehingga karya ilmiah dapat terselesaikan dengan baik. Oleh karenanya, izinkan penulis menyampaikan terima kasih melalui prakata ini kepada:

1. Bapak Edy Sunarso, Ibu Rita Lely, dan juga Fikri Edriansyah selaku orang tua dan abang yang tidak pernah lelah mengajarkan, mendukung, juga memberikan do'a serta menjadi motivasi bagi penulis;
2. Ibu Akbar Rizki, S.Stat., M.Si. dan Bapak La Ode Abdul Rahman, S.Si., M.Si. selaku komisi pembimbing atas do'a dan kesabarannya dalam mengarahkan dan membimbing penulis selama proses penyusunan karya ilmiah ini;
3. Bapak Dr. Agus Mohamad Soleh, S.Si., M.T; Ibu Eka Putri Nur Utami, S.Stat., M.Si; dan Ibu Laily Nissa Atul Mualifah, S.Si., M.Si; selaku moderator kolokium, seminar hasil, juga penguji luar komisi, atas arahan, saran, dan masukannya bagi penulis dalam menyelesaikan karya ilmiah;
4. Seluruh dosen dan tenaga pendidik Program Studi Statistika dan Sains Data yang telah memberi ilmu bermanfaat dan menunjang segala kebutuhan penulis selama perkuliahan dan penyusunan karya ilmiah ini;
5. Bapak, Ibu, Abo, Mbah dan seluruh keluarga yang telah memberikan dukungan, baik moral maupun moril bagi penulis sehingga dapat menyelesaikan karya ilmiah ini sebagaimana mestinya;
6. Mahib, Eji, Abyan, Aji, Rehan, Yapi, Hapis, Patin, Jigan, Razu, Jefri, Muaz, Syahrul, Padil, Topik, Ajam, Haris, Wildan, Ucup, Rino, Nasrul, Faqih, Arip, Nara, Abdul, Abdil, Ajril, Iki, Omen, Panjul, Abud, Ajis dan Pais yang sama-sama berjuang, membantu, menemani dan mendukung penulis selama proses penyusunan karya ilmiah ini hingga akhir;
7. Teman-teman, abang-abang, adik-adik, dan anak-anak Marinestic 59, As-Sadaad, Gaskeun, Meluncur, Inference, Alpha, Leverage, Outsco B7, yang telah memberikan banyak sekali warna dan cerita bagi penulis;
8. Serta seluruh rekan dan pihak lain yang tidak bisa penulis sebutkan satu per satu karena telah mendukung penyusunan karya ilmiah ini.

Pada akhirnya, penulis menyadari bahwa karya ilmiah ini masih sangat jauh dari kata sempurna. Meski demikian, penulis berharap bahwa karya ilmiah ini dapat bermanfaat bagi pihak yang membutuhkan serta kemajuan ilmu pengetahuan.

Bogor, Juli 2026

Fauzan Fajari



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



DAFTAR ISI

DAFTAR TABEL	xii
DAFTAR GAMBAR	xii
I PENDAHULUAN	13
1.1 Latar Belakang	13
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Analisis Sentimen	3
2.2 IndoBERT	3
2.3 Support Vector Regression (SVR)	4
2.4 Long-Short Term Memory (LSTM)	6
2.5 <i>Time Series K-Fold Cross Validation</i> dengan <i>Sliding Window</i>	8
2.6 Metrik Evaluasi	9
III METODE	10
3.1 Data	10
3.2 Prosedur Analisis	10
IV HASIL DAN PEMBAHASAN	14
4.1 Pelabelan data artikel dengan IndoBERT	14
4.2 Eksplorasi Data	15
4.3 Kestasioneran Data dan <i>Feature Engineering</i>	16
4.4 Pemodelan <i>Support Vector Regression</i>	18
4.5 Pemodelan <i>Long-Short Term Memory</i>	20
4.6 Perbandingan Performa Model	21
4.7 Prediksi Jumlah Wisatawan	22
V SIMPULAN	24
DAFTAR PUSTAKA	25
LAMPIRAN	27



DAFTAR TABEL

1	Ilustrasi model <i>Support Vector Regression</i> kernel RBF	6
2	Ilustrasi arsitektur <i>long-short term memory</i>	6
3	Ilustrasi <i>time series k-fold cross validation</i> dengan skema <i>sliding window</i> dan $k=5$	9
4	Grafik distribusi sentimen berita hasil pelabelan menggunakan IndoBERT	14
5	Grafik deret waktu jumlah wisatawan domestik Bali Januari 2014 – Desember 2025	15
6	Grafik deret waktu nilai rasio positif sentimen berita Januari 2014 – Desember 2025	16
7	Plot ACF data (a) jumlah wisatawan, (b) nilai rasio positif sentimen berita	16
8	Plot ACF data jumlah wisatawan domestik setelah differencing	17
9	Plot PACF data jumlah wisatawan domestik setelah differencing	17
10	Plot CCF data jumlah wisatawan dan rasio positif <i>sentimen</i> berita	18
11	Hasil prediksi (a) data latih dan (b) data uji model SVR	20
12	Hasil prediksi (a) data latih dan (b) data uji model LSTM	21
13	Prediksi model SVR dan LSTM pada tahun 2026	23

DAFTAR GAMBAR

1	Ilustrasi model <i>Support Vector Regression</i> kernel RBF	6
2	Ilustrasi arsitektur <i>long-short term memory</i>	6
3	Ilustrasi <i>time series k-fold cross validation</i> dengan skema <i>sliding window</i> dan $k=5$	9
4	Grafik distribusi sentimen berita hasil pelabelan menggunakan IndoBERT	14
5	Grafik deret waktu jumlah wisatawan domestik Bali Januari 2014 – Desember 2025	15
6	Grafik deret waktu nilai rasio positif sentimen berita Januari 2014 – Desember 2025	16
7	Plot ACF data (a) jumlah wisatawan, (b) nilai rasio positif sentimen berita	16
8	Plot ACF data jumlah wisatawan domestik setelah differencing	17
9	Plot PACF data jumlah wisatawan domestik setelah differencing	17
10	Plot CCF data jumlah wisatawan dan rasio positif <i>sentimen</i> berita	18
11	Hasil prediksi (a) data latih dan (b) data uji model SVR	20
12	Hasil prediksi (a) data latih dan (b) data uji model LSTM	21
13	Prediksi model SVR dan LSTM pada tahun 2026	23