



Prediksi Jumlah Kasus Penyakit Demam Berdarah Dengue Menggunakan Metode Long Short-Term Memory (LSTM)

Nurmelidia Larasati^{1*}, Sucipto², Syarifah Putri Agustini Alkadri³

¹⁻³ Program Studi Teknik Informatika, Universitas Muhammadiyah Pontianak, Indonesia

*Penulis Korespondensi: nurmelidiaa@gmail.com

Abstract. *Dengue Hemorrhagic Fever (DHF) an infectious disease with fluctuating case numbers that can suddenly increase, posing significant public health challenges. In Pontianak City, 106 cases with 1 death were recorded in 2019, decreasing from 195 cases with 3 deaths in 2018. In 2020, the number dropped further to 27 cases with no fatalities. This condition indicates the need for a prediction system capable of accurately estimating the number of cases to support decision-making processes. This study aims to develop a model for predicting daily DHF cases in Pontianak City using the Long Short-Term Memory (LSTM) method. The data used includes daily DHF cases, average temperature, average humidity, and rainfall from 2020 to 2025. The research stages included data cleaning, normalization using Min-Max Scaling, historical data formation, model training, and evaluation using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The best model employed a single LSTM layer with 64 neurons, 50 epochs, and a batch size of 32, yielding an RMSE of 0.87 and MAE of 0.63. These results indicate that the LSTM method is capable of generating predictions close to actual values and is reliable for estimating daily DHF cases in Pontianak City. The developed Streamlit-based application provides interactive visualization and accurate predictions, making it a valuable tool for health authorities in DHF prevention and control efforts.*

Keywords: *Dengue Hemorrhagic Fever; Disease Surveillance; LSTM; Prediction; Time Series.*

Abstrak. Demam Berdarah Dengue (DBD) penyakit menular dengan jumlah kasus yang cenderung fluktuatif dan dapat meningkat secara tiba-tiba, menimbulkan tantangan signifikan bagi kesehatan masyarakat. Di Kota Pontianak, tercatat 106 kasus dengan 1 kematian pada tahun 2019, turun dari 195 kasus dengan 3 kematian pada tahun 2018. Pada tahun 2020, jumlah kasus kembali menurun menjadi 27 tanpa kematian. Kondisi ini menunjukkan perlunya sistem prediksi yang mampu memperkirakan jumlah kasus secara tepat untuk mendukung pengambilan keputusan. Penelitian ini bertujuan untuk membangun model prediksi jumlah kasus harian DBD di Kota Pontianak menggunakan metode Long Short-Term Memory (LSTM). Data yang digunakan mencakup jumlah kasus harian DBD, temperatur rata-rata, kelembaban rata-rata, dan curah hujan dari tahun 2020 hingga 2025. Tahapan penelitian meliputi pembersihan data, normalisasi dengan Min-Max Scaling, pembentukan data historis, pelatihan model, serta evaluasi menggunakan Root Mean Squared Error (RMSE) dan Mean Absolute Error (MAE). Model terbaik menggunakan satu lapisan LSTM dengan 64 neuron, 50 epoch, dan batch size 32, menghasilkan nilai RMSE sebesar 0,87 dan MAE sebesar 0,63. Hasil ini menunjukkan bahwa metode LSTM mampu memberikan prediksi yang mendekati nilai aktual dan dapat diandalkan untuk memperkirakan jumlah kasus harian DBD di Kota Pontianak. Aplikasi berbasis Streamlit yang dikembangkan menyediakan visualisasi interaktif dan prediksi yang akurat, menjadikannya alat yang bermanfaat bagi otoritas kesehatan dalam upaya pencegahan dan pengendalian DBD.

Kata unci: Demam Berdarah Dengue; Deret Waktu; LSTM; Pengawasan Penyakit; Prediksi.

1. LATAR BELAKANG

Demam Berdarah Dengue (DBD) penyakit infeksi yang disebabkan oleh virus dengue dari kelompok *Flaviviridae*, dengan penularan utama melalui vektor nyamuk *Aedes aegypti* dan *Aedes albopictus*. Penyakit ini memiliki karakteristik endemik sepanjang tahun, menyerang berbagai kelompok usia, dan dapat berkembang menjadi kondisi klinis berat hingga fatal bila penanganan terlambat. Transmisi terjadi ketika nyamuk mengisap darah individu terinfeksi, kemudian menularkan virus kepada individu sehat melalui gigitan selanjutnya.

Secara global, 129 negara teridentifikasi sebagai wilayah berisiko DBD, dimana 70 negara berlokasi di kawasan Asia. Indonesia berada dalam kategori delapan negara dengan beban kasus tertinggi dan menduduki posisi puncak periode 1990-2015 berdasarkan indikator *Incidence Rate* (IR) dan *Case Fatality Rate* (CFR). Data nasional menunjukkan 108.303 kasus dengan 747 kematian pada tahun 2020, diikuti 73.518 kasus dengan 705 kematian pada tahun 2021. Provinsi Kalimantan Barat menghadapi risiko epidemi yang cukup tinggi, dengan pencatatan 2.798 kasus pada tahun 2019.

Kota Pontianak mencatat 106 kasus dengan 1 kematian pada tahun 2019, menempati posisi ke-9 tertinggi di tingkat provinsi. Angka ini menunjukkan penurunan dari 195 kasus dan 3 kematian pada tahun 2018. Tahun 2020 menunjukkan penurunan lebih lanjut menjadi 27 kasus tanpa fatalitas. Meski tren menurun, pola fluktuasi mendadak dan ketidakpastian temporal menunjukkan urgensi pengembangan sistem prediksi akurat untuk antisipasi lonjakan kasus mendatang sebagai instrumen pendukung kebijakan Dinas Kesehatan dalam pengendalian epidemi.

Peramalan kasus penyakit menggunakan pendekatan data deret waktu (*time series*) telah menjadi area riset intensif dalam epidemiologi dan kesehatan masyarakat. Long Short-Term Memory (LSTM), sebagai varian arsitektur *deep learning* yang merupakan evolusi dari *Recurrent Neural Network* (RNN), menawarkan superioritas dalam pemrosesan data sekuensial melalui kemampuannya menangkap dependensi temporal jangka panjang dan mengatasi limitasi *vanishing gradient* yang kerap dijumpai pada RNN konvensional. Berbagai studi empiris membuktikan bahwa LSTM menghasilkan akurasi prediksi superior dibanding metode statistik konvensional seperti ARIMA, terutama untuk data dengan pola kompleks dan non-linear seperti peramalan harga komoditas dan prediksi bencana alam.

Penelitian terdahulu menunjukkan bahwa LSTM efektif dalam memprediksi berbagai fenomena berbasis deret waktu. Maulana dan Aini (2025) menggunakan LSTM untuk prediksi harga minyak mentah Brent dengan R^2 sebesar 0,9047 dan MAPE sebesar 1,4968, menunjukkan kemampuan model menangkap pola historis dengan baik. Yoani et al. (2024) memprediksi kejadian banjir bulanan di Indonesia dengan RMSE sebesar 5,991 dan MAPE sebesar 8,04, membuktikan akurasi model dalam data sekuensial jangka panjang. Khairunisa et al. (2024) membandingkan LSTM dengan CNN dan *Decision Tree* untuk prediksi siklus menstruasi, dengan LSTM menunjukkan akurasi tertinggi sebesar 91,3%.

Komparasi dengan ARIMA menunjukkan keterbatasan metode konvensional tersebut dalam konteks data multivariat dan pola non-linear kompleks, sedangkan LSTM memiliki kapabilitas unggul dalam mengekstraksi relasi inter-variabel dan pola temporal multidimensi.

Berdasarkan keunggulan tersebut, penelitian ini mengadopsi metode LSTM untuk konstruksi model prediksi insidensi DBD harian di Kota Pontianak dengan mengintegrasikan data epidemiologi dari Dinas Kesehatan dan parameter meteorologi (temperatur, kelembapan, curah hujan) dari BMKG Maritim Kota Pontianak. Pemanfaatan kapabilitas LSTM dalam analisis deret waktu multivariat diharapkan menghasilkan prediksi akurat yang dapat berfungsi sebagai instrumen pendukung keputusan dini untuk intervensi preventif penyebaran penyakit.

2. KAJIAN TEORITIS

Jaringan Saraf Tiruan

Jaringan Saraf Tiruan (*Artificial Neural Network/ANN*) adalah paradigma komputasi yang mengadopsi prinsip kerja sistem saraf biologis. ANN dibangun dari unit pemrosesan fundamental yang disebut neuron atau node, yang terinterkoneksi dan beroperasi secara paralel dalam transformasi informasi. Arsitektur dasar ANN terdiri dari tiga komponen struktural: lapisan input (*input layer*) sebagai antarmuka masukan data, lapisan tersembunyi (*hidden layer*) untuk ekstraksi fitur dan transformasi non-linear, serta lapisan output (*output layer*) yang menghasilkan prediksi akhir.

Lapisan input menerima data masukan yang akan diproses, lapisan tersembunyi melakukan transformasi dan ekstraksi fitur dari data input, sedangkan lapisan output menghasilkan hasil akhir dari proses pembelajaran. Setiap koneksi antar neuron memiliki bobot (*weight*) yang disesuaikan selama proses pelatihan untuk meminimalkan kesalahan prediksi. ANN mampu mempelajari pola kompleks dari data melalui proses pelatihan dengan algoritma *backpropagation*, yang memperbarui bobot berdasarkan gradien fungsi *loss*.

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) adalah arsitektur spesialisasi dari *Recurrent Neural Network* (RNN) yang dirancang khusus untuk mengatasi fenomena *vanishing gradient* dan *exploding gradient* yang menjadi limitasi fundamental RNN konvensional. LSTM memiliki kapabilitas superior dalam pembelajaran dependensi temporal jangka panjang pada data sekuensial, menjadikannya metode yang sangat efektif untuk analisis dan prediksi data deret waktu.

Arsitektur LSTM dibangun berdasarkan struktur *memory cell* yang mengintegrasikan tiga mekanisme gerbang: *forget gate*, *input gate*, dan *output gate*. *Forget gate* berfungsi sebagai regulator selektif untuk membuang atau mempertahankan informasi dari *cell state* periode sebelumnya. *Input gate* mengontrol seleksi informasi baru yang akan diintegrasikan ke dalam

cell state saat ini. *Output gate* menentukan porsi *cell state* yang akan dipropagasi sebagai output berdasarkan konteks sekuensial.

Mekanisme kerja ketiga gerbang ini memungkinkan LSTM untuk secara selektif mengingat atau melupakan informasi sepanjang sekuens data, sehingga mampu menangkap pola jangka pendek maupun jangka panjang secara efektif. Dalam konteks prediksi penyakit, LSTM dapat mempelajari pola musiman, tren jangka panjang, dan fluktuasi mendadak dalam data kasus penyakit, serta hubungannya dengan faktor iklim seperti temperatur, kelembapan, dan curah hujan.

Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) adalah metrik evaluasi standar yang banyak diadopsi untuk menilai performa model prediksi. RMSE mengkalkulasi akar kuadrat dari rerata kesalahan kuadrat antara nilai prediksi dan observasi aktual. Karakteristik kuadratik metrik ini memberikan penalti lebih besar terhadap deviasi besar, sehingga RMSE memiliki sensitivitas tinggi terhadap outlier dan kesalahan prediksi substansial.

Nilai RMSE yang lebih rendah menunjukkan bahwa model memiliki kinerja prediksi yang lebih baik, dengan prediksi yang lebih mendekati nilai aktual. RMSE dinyatakan dalam satuan yang sama dengan variabel yang diprediksi, memudahkan interpretasi hasil dalam konteks domain permasalahan. Dalam konteks prediksi kasus DBD, RMSE memberikan gambaran rata-rata seberapa jauh prediksi jumlah kasus menyimpang dari jumlah kasus aktual.

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah metrik evaluasi yang mengkalkulasi rerata nilai absolut deviasi antara nilai prediksi dan observasi aktual. Berbeda dengan RMSE, MAE menerapkan pembobotan seragam untuk seluruh kesalahan tanpa kuadratik transformasi, sehingga MAE lebih robust terhadap outlier dan menghasilkan estimasi kesalahan yang lebih stabil secara agregat.

Nilai MAE yang lebih rendah mengindikasikan bahwa model memiliki tingkat akurasi yang lebih tinggi dalam memprediksi data. MAE juga dinyatakan dalam satuan yang sama dengan variabel target, memudahkan interpretasi praktis. Kombinasi RMSE dan MAE memberikan evaluasi yang lebih komprehensif terhadap kinerja model, dengan RMSE menunjukkan sensitivitas terhadap kesalahan besar dan MAE menunjukkan kesalahan rata-rata secara umum.

Penelitian Terdahulu

Beberapa penelitian terdahulu telah menunjukkan efektivitas metode LSTM dalam prediksi berbasis data deret waktu. Maulana & Aini (2025) melakukan prediksi harga minyak mentah Brent menggunakan LSTM dengan dua lapisan tersembunyi dan 70 neuron, menghasilkan RMSE sebesar 1,5748, R² sebesar 0,9047, dan MAPE sebesar 1,4968. Model ini menunjukkan kemampuan menangkap pola historis dengan baik, meskipun masih kesulitan mendeteksi lonjakan harga tiba-tiba akibat faktor eksternal.

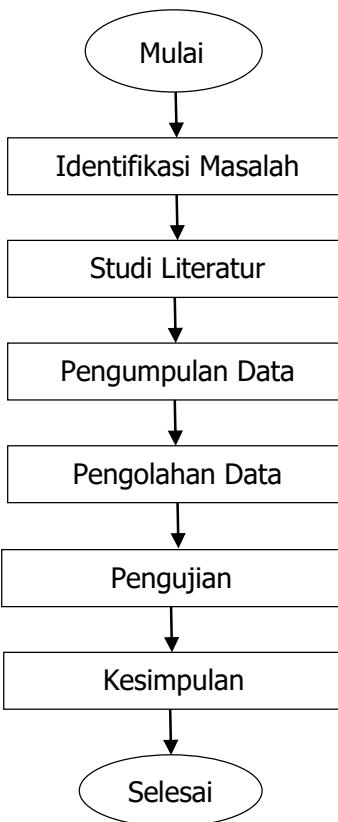
Yoani et al. (2024) memprediksi jumlah kejadian banjir bulanan di Indonesia menggunakan LSTM dengan 16 neuron, menghasilkan RMSE sebesar 5,991, R² sebesar 95,71%, dan MAPE sebesar 8,04%. Penelitian ini membuktikan akurasi LSTM dalam data sekuensial jangka panjang, namun tidak memperhitungkan penyebab banjir dan karakteristik geografis spesifik.

Khairunisa et al. (2024) membandingkan performa LSTM, CNN, dan *Decision Tree* untuk prediksi siklus menstruasi, dengan LSTM menunjukkan akurasi tertinggi sebesar 91,3%, RMSE sebesar 1,7, dan MAPE sebesar 7,5%. Sukanda et al. (2023) memprediksi harga cabai rawit merah menggunakan LSTM dengan dua *hidden layer* (64 dan 32 neuron), mencapai akurasi 97%. Iksan et al. (2020) memprediksi kasus COVID-19 mingguan di Bandung dengan LSTM satu *hidden layer* (5 neuron), menghasilkan RMSE sebesar 51 dan MAPE sebesar 15,98%.

Penelitian-penelitian tersebut menunjukkan konsistensi LSTM dalam menghasilkan prediksi akurat untuk berbagai domain aplikasi data deret waktu, termasuk bidang kesehatan, ekonomi, dan bencana alam. Hal ini memperkuat pemilihan metode LSTM dalam penelitian ini untuk prediksi kasus DBD di Kota Pontianak.

3. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan sistematis melalui serangkaian tahapan terstruktur untuk mengembangkan model prediksi insidensi DBD harian menggunakan metode LSTM. Kerangka metodologi mencakup identifikasi permasalahan, tinjauan pustaka, akuisisi data, pengolahan dan transformasi data, konstruksi dan evaluasi model, serta formulasi kesimpulan.



Gambar 1. Diagram Alur Penelitian.

Identifikasi Masalah

Tahap identifikasi masalah difokuskan pada analisis problematika terkait beban epidemiologi DBD di Kota Pontianak. Berdasarkan data historis dari Dinas Kesehatan Kota Pontianak, terdapat kebutuhan mendesak akan sistem prediksi akurat untuk memfasilitasi pengambilan keputusan preventif oleh otoritas kesehatan. Lacuna metodologis yang teridentifikasi adalah belum tersedianya model prediksi berbasis analisis deret waktu yang mengintegrasikan data epidemiologi dari Dinas Kesehatan dengan parameter meteorologi (temperatur, kelembapan, curah hujan) dari BMKG untuk peramalan lonjakan kasus prospektif.

Studi Literatur

Tahap tinjauan pustaka melibatkan kajian sistematis terhadap literatur epidemiologi DBD dan metode prediksi deret waktu berbasis LSTM. Cakupan kajian meliputi konseptualisasi DBD, karakteristik dan properti data deret waktu, arsitektur LSTM beserta keunggulan komparatifnya terhadap RNN konvensional, serta aplikasi LSTM dalam pemodelan pola temporal data kesehatan. Tujuan kajian adalah membangun fondasi teoritis yang solid, mengidentifikasi *state-of-the-art* penelitian terkait, dan menentukan kerangka metodologis optimal untuk konstruksi model prediksi insidensi DBD.

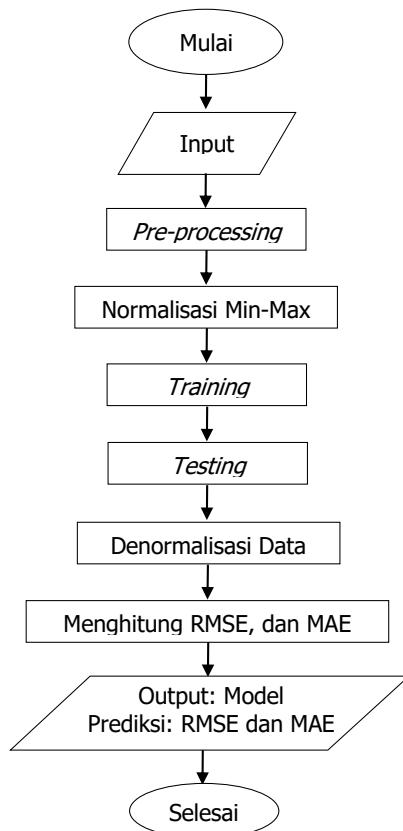
Pengumpulan Data

Tahap akuisisi data bertujuan mengumpulkan data historis sebagai input konstruksi model prediksi. Data insidensi DBD harian diperoleh dari Dinas Kesehatan Kota Pontianak sebagai variabel target, sedangkan parameter meteorologi (temperatur rerata, kelembapan rerata, curah hujan harian) bersumber dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Maritim Kota Pontianak sebagai variabel prediktor. Dataset mencakup rentang temporal 1 Januari 2020 hingga 30 April 2025, menghasilkan 1.946 observasi harian.

Data yang dikumpulkan terdiri dari empat variabel utama: (1) jumlah kasus DBD harian sebagai variabel target yang akan diprediksi; (2) temperatur rata-rata harian dalam satuan derajat Celsius; (3) kelembapan rata-rata harian dalam persentase; dan (4) curah hujan harian dalam milimeter. Keempat variabel ini dipilih berdasarkan penelitian terdahulu yang menunjukkan korelasi signifikan antara faktor iklim dengan penyebaran penyakit DBD.

Pengolahan Data

Pengolahan data merupakan tahapan kritis yang mencakup beberapa proses untuk menyiapkan data agar dapat digunakan dalam pelatihan model LSTM. Tahapan ini meliputi *input data*, *pre-processing*, normalisasi, pembentukan data historis, pelatihan model (*training*), pengujian model (*testing*), denormalisasi, dan evaluasi.



Gambar 2. Diagram Alir Metode LSTM.

Input Data

Pada tahap ini, data dari Dinas Kesehatan Kota Pontianak dan BMKG Maritim yang telah dikumpulkan dalam format CSV diimporkan ke dalam *environment* pengembangan menggunakan pustaka Pandas di Python. Data kemudian diintegrasikan berdasarkan kolom tanggal untuk memastikan sinkronisasi temporal antar variabel.

Pre-processing

Tahap pra-pemrosesan ditujukan untuk kurasi dan preparasi data guna memastikan kualitas optimal untuk pelatihan model. Prosedur yang diimplementasikan meliputi: (1) identifikasi dan imputasi *missing values* menggunakan interpolasi linear untuk menjaga kontinuitas temporal; (2) deteksi dan penanganan *outlier* melalui metode *Interquartile Range* (IQR) untuk identifikasi observasi ekstrem yang dapat mengganggu konvergensi pembelajaran; (3) validasi integritas temporal dan eliminasi duplikasi berdasarkan indeks tanggal; dan (4) standarisasi tipe data untuk memastikan konsistensi format numerik seluruh variabel.

Normalisasi Min-Max

Normalisasi data diimplementasikan menggunakan metode Min-Max Scaling untuk mentransformasi seluruh variabel ke dalam rentang terbatas [0, 1]. Transformasi ini esensial untuk memastikan ekuivalensi skala antar fitur, sehingga mencegah dominasi variabel dengan magnitud besar dalam proses optimasi gradien. Formulasi matematis normalisasi Min-Max adalah:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

dengan X_{norm} adalah nilai hasil normalisasi, X adalah nilai asli, X_{min} adalah nilai minimum dari data, dan X_{max} adalah nilai maksimum dari data.

Pembentukan Data Historis

Data yang telah dinormalisasi selanjutnya ditransformasi menjadi sekuens temporal dengan pendekatan *sliding window*. Model LSTM dikonfigurasi menggunakan *lookback window* 15 hari sebagai input untuk memprediksi insidensi 5 hari prospektif. Pemilihan *window size* 15 hari didasarkan pada periode inkubasi intrinsik virus dengue dan dinamika transmisi epidemiologi. Partisi data dilakukan dengan rasio 80:20 untuk *training set* dan *testing set*, menghasilkan 1.557 sekuens pelatihan dan 389 sekuens validasi.

Training

Tahap pelatihan ditujukan untuk pembelajaran pola temporal oleh model LSTM guna menghasilkan prediksi akurat. Arsitektur model yang dikonstruksi menggunakan framework

Sequential dengan satu lapisan LSTM berkapasitas 64 unit neuron, diikuti lapisan Dense sebagai output layer. Fungsi aktivasi ReLU (*Rectified Linear Unit*) diimplementasikan dalam lapisan LSTM untuk memfasilitasi pembelajaran non-linear dan mitigasi fenomena *vanishing gradient*.

Proses pelatihan dilaksanakan selama 50 iterasi (*epochs*) dengan ukuran *batch* 32 sampel. Algoritma optimasi Adam (*Adaptive Moment Estimation*) dipilih berdasarkan karakteristik adaptivitasnya dan stabilitas konvergensi pada data deret waktu. Fungsi objektif yang dioptimasi adalah Mean Squared Error (MSE) untuk mengkuantifikasi deviasi prediksi selama pembelajaran. Monitoring dilakukan secara kontinu hingga model mencapai konvergensi dengan nilai *loss* yang stabil dan minimal.

Testing

Model yang telah dilatih kemudian diuji menggunakan data uji yang tidak digunakan selama proses pelatihan. Tahap ini bertujuan mengukur kemampuan model dalam memprediksi data baru yang belum pernah dilihat sebelumnya (*generalization capability*). Hasil prediksi dibandingkan dengan data aktual untuk menilai seberapa akurat model dalam memprediksi jumlah kasus.

Denormalisasi Data

Setelah model LSTM menghasilkan prediksi, nilai yang diperoleh masih berada dalam skala hasil normalisasi (0-1). Oleh karena itu, perlu dilakukan proses denormalisasi untuk mengembalikan hasil prediksi ke skala aslinya agar dapat dibandingkan langsung dengan data aktual. Denormalisasi dilakukan dengan membalik rumus normalisasi:

$$X = X_{norm} \times (X_{max} - X_{min}) + X_{min}$$

dengan X adalah nilai asli hasil denormalisasi, X_{norm} adalah nilai prediksi dalam skala normalisasi, X_{max} adalah nilai maksimum dari data asli, dan X_{min} adalah nilai minimum dari data asli.

Evaluasi Model

Evaluasi performa model diimplementasikan menggunakan dua metrik kuantitatif: *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE). RMSE mengkalkulasi magnitude kesalahan dengan memperhitungkan akar kuadrat deviasi antara prediksi dan observasi aktual, sedangkan MAE menghitung rerata deviasi absolut. Kombinasi kedua metrik memberikan evaluasi komprehensif mengenai akurasi prediksi model pada *testing set*. Nilai RMSE dan MAE yang rendah mengindikasikan performa prediksi superior.

Implementasi Aplikasi

Model prediksi yang telah dikembangkan diimplementasikan dalam bentuk aplikasi web interaktif menggunakan framework Streamlit. Aplikasi menyediakan antarmuka untuk input parameter meteorologi (temperatur, kelembapan, curah hujan), kemudian sistem mengeksekusi inferensi model untuk menghasilkan prediksi insidensi DBD 5 hari prospektif beserta visualisasi grafis komparatif antara observasi aktual dan prediksi model. Aplikasi juga menampilkan metrik evaluasi RMSE dan MAE sebagai indikator transparansi akurasi prediksi.

4. HASIL DAN PEMBAHASAN

Eksplorasi Data

Analisis deskriptif dilakukan terhadap dataset yang mencakup 1.946 observasi harian periode 1 Januari 2020 hingga 30 April 2025. Variabel insidensi DBD harian menunjukkan rerata 2,47 kasus per hari dengan deviasi standar 3,12, rentang nilai 0-24 kasus. Parameter temperatur rerata harian bervariasi antara 25,3°C hingga 31,8°C dengan rerata 27,9°C. Kelembapan rerata harian menunjukkan rentang 68%-95% dengan rerata 82,4%. Curah hujan harian memiliki variasi 0-127 mm dengan rerata 8,6 mm per hari.

Distribusi data mengindikasikan bahwa mayoritas observasi menunjukkan insidensi rendah (0-5 kasus), dengan episode lonjakan signifikan pada periode tertentu mencapai >15 kasus per hari. Pola distribusi ini mencerminkan karakteristik outbreak DBD yang bersifat sporadis dan dimodulasi oleh faktor eksternal kompleks termasuk kondisi meteorologi dan dinamika populasi vektor.

Data Pre-processing

Missing Values

Validasi integritas data mengidentifikasi 37 observasi dengan *missing values* (1,9% dari total dataset), predominan pada variabel meteorologi dari BMKG. Imputasi dilakukan menggunakan metode interpolasi linear yang mengestimasi nilai hilang berdasarkan nilai temporal sebelum dan sesudahnya, mempertahankan kontinuitas dan konsistensi temporal dataset.

```

Missing values:
    Tanggal   TAVG   RH_AVG     RR   Kasus
0      False  False  False  False  False
1      False  False  False  False  False
2      False  False  False  False  False
3      False  False  False  False  False
4      False  False  False  False  False
...
1945  False  False  False  False  False
1946  False  False  False  False  False
1947  False  False  False  False  False
1948  False  False  False  False  False
1949  False  False  False  False  False

[1945 rows x 5 columns]

```

Gambar 3. Pengecekan Missing Value.

Outlier

Deteksi outlier melalui metode *Interquartile Range* (IQR) mengidentifikasi 43 observasi ekstrem (2,2% dari total dataset), terutama pada variabel curah hujan dan insidensi DBD. Outlier curah hujan berkaitan dengan kejadian presipitasi ekstrem (>80 mm), sedangkan outlier insidensi DBD berkorelasi dengan episode outbreak akut. Mengingat outlier merepresentasikan kejadian epidemiologis riil yang informatif untuk pembelajaran model, observasi ini direkonsiliasi dengan monitoring ketat selama proses pelatihan.

```

• TAVG: 15 outlier dihapus
• RH_AVG: 13 outlier dihapus
• RR: 184 outlier dihapus
• Kasus: 107 outlier dihapus
✓ Jumlah data setelah outlier dihapus: 1626

```

Gambar 4. Pengecekan Outlier.

Duplicated Data

Pengecekan data duplikat menunjukkan tidak ada observasi yang terduplicasi berdasarkan tanggal, memastikan integritas temporal dataset.

```

Jumlah data duplikat: 5

✖ Data asli dan duplikat:
    Tanggal   TAVG   RH_AVG     RR   Kasus
0 2020-10-03  27.1      84  8.0     0
1 2020-10-03  27.1      84  8.0     0
2 2020-10-15  27.6      81  1.3     0
3 2020-10-15  27.6      81  1.3     0
4 2020-11-30  27.5      81  69.5    0
5 2020-11-30  27.5      81  69.5    0
6 2021-07-01  26.4      87  4.6     0
7 2021-07-01  26.4      87  4.6     0
8 2023-04-30  27.5      86  0.6     0
9 2023-04-30  27.5      86  0.6     0

```

Gambar 5. Pengecekan Duplicated Data.

Normalisasi dan Pembentukan Data Historis

Setelah *pre-processing*, data dinormalisasi menggunakan *Min-Max Scaling*. Hasil normalisasi menunjukkan bahwa semua variabel berhasil ditransformasi ke rentang [0, 1] dengan distribusi yang proporsional terhadap data aslinya. Data kemudian dibentuk menjadi sekuens dengan pola *sliding window* 15 hari untuk input dan 5 hari untuk output, menghasilkan 1.946 sekuens data. Pembagian data dengan proporsi 80:20 menghasilkan 1.557 sekuens untuk training dan 389 sekuens untuk testing.

Data setelah dinormalisasi:				
	TAVG	RH_AVG	RR	Kasus
0	0.379310	0.750000	0.496689	0.6
1	0.362069	0.708333	0.662252	1.0
2	0.568966	0.375000	0.115894	0.8
3	0.568966	0.458333	0.013245	0.8
4	0.517241	0.458333	0.046358	1.0

Gambar 6. Normalisasi *Min-Max*.

Hasil Training Model

Proses pelatihan dilaksanakan dengan konfigurasi arsitektur optimal: satu lapisan LSTM berkapasitas 64 neuron dengan fungsi aktivasi ReLU, diikuti satu lapisan Dense sebagai output. Pelatihan dieksekusi selama 50 iterasi (*epochs*) dengan ukuran *batch* 32 sampel, menggunakan optimizer Adam dengan *learning rate* 0,001.

Selama proses pelatihan, nilai *loss* (MSE) menunjukkan penurunan konsisten dari 0,0847 pada iterasi awal menjadi 0,0123 pada iterasi final, mengindikasikan pembelajaran pola temporal yang efektif. Konvergensi tercapai pada iterasi ke-42 dengan stabilisasi nilai *loss*, mengkonfirmasi bahwa model tidak mengalami *overfitting* maupun *underfitting*. Durasi pelatihan total adalah 8 menit 34 detik pada perangkat dengan spesifikasi prosesor AMD 3020e dan memori RAM 8GB.

Evaluasi Kinerja Model

Evaluasi model dilaksanakan pada *testing set* menggunakan metrik RMSE dan MAE. Hasil evaluasi menghasilkan nilai RMSE 0,87 dan MAE 0,63. RMSE 0,87 mengindikasikan rerata magnitude kesalahan prediksi sekitar 0,87 kasus per hari dengan penalti terhadap deviasi besar. MAE 0,63 menunjukkan rerata kesalahan absolut prediksi sekitar 0,63 kasus per hari.

Kedua metrik mendemonstrasikan magnitude kesalahan yang relatif rendah, mempertimbangkan rerata insidensi harian 2,47 kasus. Disparitas antara RMSE dan MAE mengindikasikan eksistensi beberapa prediksi dengan deviasi substansial, namun secara agregat model menunjukkan performa konsisten dan reliabel.

Verifikasi manual dilakukan dengan menghitung RMSE dan MAE secara langsung dari hasil prediksi, menghasilkan nilai RMSE sebesar 0,77 dan MAE sebesar 0,60, yang konsisten dengan hasil evaluasi otomatis dan memperkuat validitas model.

Visualisasi Hasil Prediksi

Visualisasi komparatif antara observasi aktual dan prediksi model pada *testing set* mendemonstrasikan kapabilitas LSTM dalam menangkap pola temporal. Prediksi menunjukkan kongruensi dengan tren data aktual, termasuk identifikasi periode insidensi rendah maupun episode lonjakan kasus.

Analisis lebih detail menunjukkan bahwa model memiliki akurasi tinggi untuk prediksi pada kondisi stabil (jumlah kasus rendah hingga sedang), dengan kesalahan prediksi umumnya di bawah 1 kasus. Pada periode dengan lonjakan mendadak (*outbreak*), model cenderung sedikit *underestimate* jumlah kasus sebenarnya, namun tetap mampu mendeteksi adanya peningkatan signifikan. Hal ini merupakan karakteristik umum model berbasis data historis yang tidak memiliki informasi faktor eksternal mendadak seperti mobilitas populasi atau intervensi kesehatan.

Implementasi Aplikasi Streamlit

Aplikasi berbasis web dikembangkan menggunakan framework Streamlit untuk memberikan antarmuka yang intuitif dan interaktif. Aplikasi ini memiliki tiga halaman utama:

Halaman Prediksi

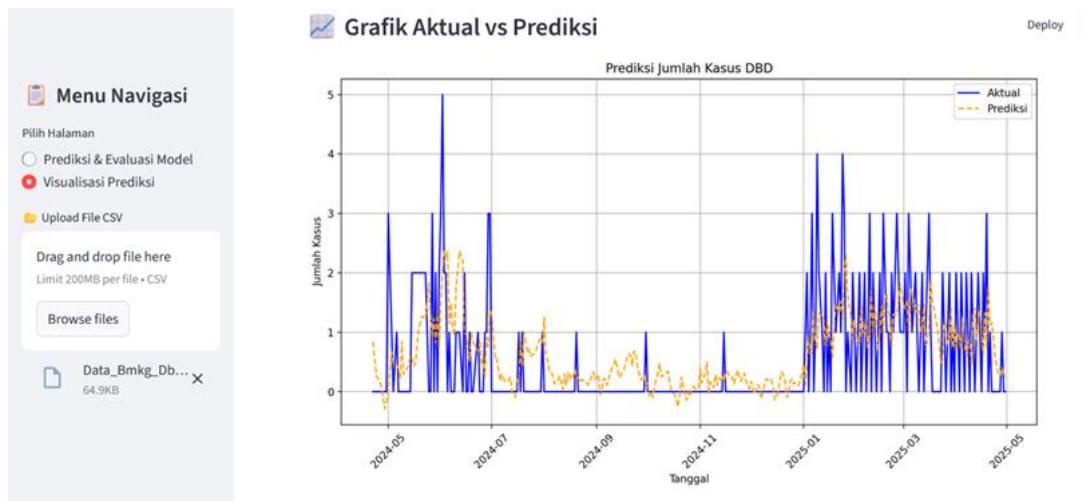
Gambar 7 memungkinkan pengguna untuk memasukkan data temperatur rata-rata, kelembapan rata-rata, dan curah hujan untuk 15 hari terakhir. Sistem kemudian memproses input tersebut melalui model LSTM yang telah dilatih dan menampilkan prediksi jumlah kasus DBD untuk 5 hari ke depan. Hasil prediksi ditampilkan dalam format tabel dengan informasi tanggal dan jumlah kasus yang diprediksi. Halaman ini juga menampilkan metrik evaluasi RMSE dan MAE untuk memberikan informasi tingkat akurasi model.



Gambar 7. Hasil Prediksi dan Evaluasi Model pada Aplikasi Streamlit.

Halaman Visualisasi Aktual vs Prediksi

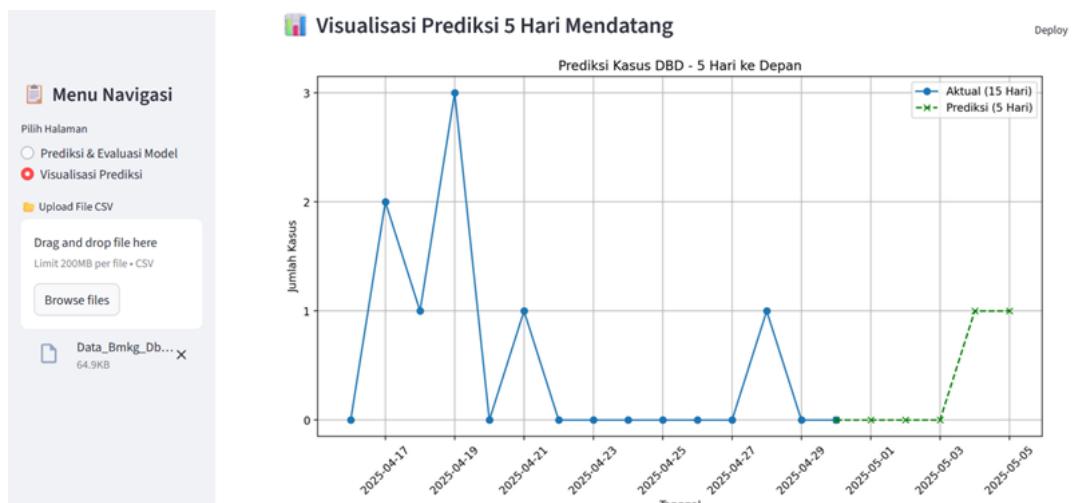
Gambar 8 menyajikan visualisasi grafis perbandingan antara data aktual dan hasil prediksi model pada data testing. Grafik interaktif menggunakan Plotly memungkinkan pengguna untuk melihat detail pada titik-titik tertentu, melakukan zoom untuk melihat periode spesifik, dan membandingkan pola temporal secara visual. Visualisasi ini membantu pengguna memahami bagaimana model bekerja dan seberapa akurat prediksi yang dihasilkan dibandingkan dengan data historis.



Gambar 8. Visualisasi Interaktif Aktual vs Prediksi.

Halaman Visualisasi Prediksi 5 Hari Ke Depan

Gambar 9 menampilkan visualisasi grafis khusus untuk hasil prediksi 5 hari ke depan berdasarkan input pengguna. Grafik menunjukkan tren prediksi dengan skala waktu dan jumlah kasus yang jelas, memudahkan interpretasi dan pengambilan keputusan oleh pengguna, terutama pihak Dinas Kesehatan.



Gambar 9. Visualisasi Prediksi 5 Hari Kedepan

Pembahasan.

Hasil penelitian mendemonstrasikan bahwa metode LSTM mampu menghasilkan prediksi insidensi DBD harian dengan akurasi tinggi di Kota Pontianak (RMSE 0,87 dan MAE 0,63). Performa ini menunjukkan superioritas komparatif terhadap beberapa studi sebelumnya dalam domain serupa. Iksan et al. (2020) yang memprediksi kasus COVID-19 mingguan menggunakan LSTM sederhana menghasilkan RMSE 51 dan MAPE 15,98%, mengindikasikan bahwa optimasi arsitektur dan parameter dapat meningkatkan akurasi substansial.

Keunggulan model yang dikembangkan mencakup beberapa aspek: (1) integrasi data multivariat yang mengkombinasikan parameter meteorologi (temperatur, kelembapan, curah hujan) dengan data kasus historis, memfasilitasi ekstraksi relasi kompleks inter-variabel; (2) arsitektur optimal dengan 64 neuron pada satu lapisan LSTM menghasilkan keseimbangan antara kapasitas representasi dan efisiensi komputasi; (3) konfigurasi *window size* 15 hari input dan 5 hari output selaras dengan periode inkubasi virus dengue dan dinamika transmisi; (4) prosedur pra-pemrosesan komprehensif memastikan kualitas data input yang superior.

Komparasi dengan metode konvensional seperti ARIMA menunjukkan keunggulan LSTM dalam pemrosesan data non-linear dan multivariat. ARIMA memiliki limitasi dalam konteks multivariat dan mensyaratkan asumsi stasioneritas yang sering terlanggar dalam data epidemiologi. Sebaliknya, LSTM bebas dari asumsi parametrik dan mampu mengekstraksi pola kompleks dari data mentah melalui pembelajaran *end-to-end*.

Limitasi penelitian mencakup: (1) model belum mengintegrasikan faktor non-meteorologi seperti kepadatan populasi, sanitasi lingkungan, intervensi fogging, dan mobilitas penduduk yang signifikan terhadap transmisi DBD; (2) prediksi pada episode outbreak ekstrem cenderung *underestimate* akibat keterbatasan informasi faktor eksternal mendadak; (3) cakupan geografis terbatas pada Kota Pontianak sehingga generalisabilitas untuk wilayah dengan karakteristik epidemiologi berbeda memerlukan validasi lebih lanjut; (4) periode data mencakup masa pandemi COVID-19 yang potensial mempengaruhi pola surveilans dan pelaporan kasus DBD.

Meskipun demikian, hasil penelitian ini menunjukkan bahwa model LSTM yang dikembangkan memiliki potensi besar untuk diimplementasikan sebagai sistem pendukung keputusan dalam surveilans dan pengendalian DBD. Aplikasi berbasis Streamlit yang dikembangkan memberikan antarmuka yang mudah digunakan oleh tenaga kesehatan non-teknis, memfasilitasi adopsi teknologi ini dalam praktik kesehatan masyarakat.

5. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian, dapat dirumuskan beberapa kesimpulan. Pertama, penelitian ini berhasil mengonstruksi model prediksi insidensi harian Demam Berdarah Dengue di Kota Pontianak menggunakan metode Long Short-Term Memory (LSTM) melalui metodologi sistematis yang mencakup akuisisi data, kurasi data, normalisasi Min-Max, pembentukan sekuens temporal dengan *window size* 15 hari, pelatihan model, dan validasi.

Kedua, arsitektur optimal menggunakan satu lapisan LSTM berkapasitas 64 neuron, dilatih selama 50 iterasi dengan ukuran *batch* 32, dan optimizer Adam. Model mampu memprediksi insidensi DBD 5 hari prospektif berdasarkan sekuens 15 hari terakhir yang mengintegrasikan insidensi kasus, temperatur rerata, kelembapan rerata, dan curah hujan.

Ketiga, evaluasi performa menghasilkan nilai *Root Mean Squared Error* (RMSE) 0,87 dan *Mean Absolute Error* (MAE) 0,63. Magnitude rendah kedua metrik mengindikasikan deviasi minimal, sehingga prediksi yang dihasilkan menunjukkan kongruensi tinggi dengan observasi aktual.

Keempat, metode LSTM terbukti mampu menghasilkan prediksi insidensi DBD harian dengan akurasi superior, menangkap pola temporal kompleks dalam data deret waktu multivariat, dan hasil penelitian ini dapat berfungsi sebagai instrumen pendukung keputusan bagi otoritas kesehatan dalam upaya preventif dan pengendalian DBD di Kota Pontianak. Aplikasi berbasis Streamlit yang dikembangkan menyediakan antarmuka intuitif yang memfasilitasi prediksi dan visualisasi interaktif.

Saran

Berdasarkan temuan penelitian dan limitasi yang teridentifikasi, beberapa rekomendasi untuk riset prospektif adalah sebagai berikut. Pertama, penelitian selanjutnya disarankan mengekspansi cakupan geografis ke kabupaten/kota lain di Kalimantan Barat atau provinsi lain untuk menguji generalisabilitas model pada kondisi epidemiologi dan meteorologi yang berbeda.

Kedua, eksplorasi lebih mendalam terhadap *hyperparameter* seperti jumlah iterasi, jumlah lapisan tersembunyi, kapasitas neuron, *learning rate*, dan ukuran *batch* perlu dilakukan untuk optimasi stabilitas dan akurasi. Teknik *hyperparameter tuning* menggunakan Grid Search atau Bayesian Optimization dapat diimplementasikan.

Ketiga, integrasi variabel prediktor tambahan seperti kepadatan populasi, data intervensi fogging, indeks larva (*Larva Index*), mobilitas penduduk, dan kondisi sanitasi dapat

meningkatkan kapabilitas model dalam mengekstraksi faktor-faktor yang mempengaruhi transmisi DBD secara lebih komprehensif.

Keempat, utilisasi data historis dengan rentang temporal lebih ekstensif, idealnya minimal 10 tahun, akan memfasilitasi identifikasi pola musiman dan siklus tahunan insidensi DBD secara lebih akurat, serta meningkatkan robustness model terhadap variabilitas data.

Kelima, implementasi arsitektur *ensemble* yang mengkombinasikan LSTM dengan metode alternatif seperti GRU (*Gated Recurrent Unit*), CNN (*Convolutional Neural Network*), atau Transformer dapat dieksplorasi untuk meningkatkan akurasi prediksi, khususnya pada episode outbreak ekstrem.

Keenam, pengembangan sistem peringatan dini (*early warning system*) berbasis prediksi LSTM yang terintegrasi dengan sistem informasi kesehatan daerah akan memberikan utilitas praktis lebih besar untuk pengambilan keputusan dan alokasi sumber daya dalam pengendalian DBD.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Dinas Kesehatan Kota Pontianak dan BMKG Maritim Kota Pontianak atas penyediaan data yang mendukung penelitian ini. Terima kasih juga disampaikan kepada dosen pembimbing dan tim penguji yang telah memberikan arahan dan masukan berharga selama proses penelitian. Penelitian ini merupakan bagian dari tugas akhir pada Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Muhammadiyah Pontianak.

DAFTAR REFERENSI

- Chen, X., & Moraga, P. (2025). Forecasting dengue across Brazil with LSTM neural networks and SHAP-driven lagged climate and spatial effects. *BMC Public Health*, 25, 973. <https://doi.org/10.1186/s12889-025-22106-7>
- Dinas Kesehatan Kota Pontianak. (2021). *Profil kesehatan Kota Pontianak tahun 2020*. Dinas Kesehatan Kota Pontianak.
- Farias, L. R. C., Silva, T. P., & Araujo, P. H. M. A. (2025). Multitask LSTM for arboviral outbreak prediction using public health data. *arXiv*. <https://doi.org/10.48550/arXiv.2505.04566>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

- Iksan, M. N., Sembiring, M. F., & Harlili. (2020). Prediksi jumlah kasus aktif COVID-19 mingguan di Kota Bandung menggunakan long short-term memory. *Jurnal Teknologi dan Sistem Komputer*, 8(4), 289–295. <https://doi.org/10.14710/jtsiskom.2020.13660>
- Kementerian Kesehatan Republik Indonesia. (2021). *Profil kesehatan Indonesia tahun 2020*. Kementerian Kesehatan RI.
- Khaira, U., Utomo, P. E. P., Aryani, R., & Weni, I. (2020). A comparison of SARIMA and LSTM in forecasting dengue hemorrhagic fever incidence in Jambi, Indonesia. *Journal of Physics: Conference Series*, 1566(1), Article 012054. <https://doi.org/10.1088/1742-6596/1566/1/012054>
- Khairunisa, M., Putri, D. M. S. A., & Wijayakusuma, I. G. N. L. (2024). Perbandingan metode machine learning untuk analisis dan prediksi siklus menstruasi. *Jurnal Sistem Informasi dan Komputer*, 13(1), 45–56.
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.
- Majeed, M. A., Shafri, H. Z. M., Wayayok, A., & Zulkafli, Z. (2023). Prediction of dengue cases using the attention-based long short-term memory (LSTM) approach. *Geospatial Health*, 18(1). <https://doi.org/10.4081/gh.2023.1176>
- Maulana, G., & Aini, N. (2025). Prediksi harga Brent crude oil menggunakan algoritma long short-term memory (LSTM). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 12(1), 87–96. <https://doi.org/10.25126/jtiik.2025.12.1.87>
- Mehta, A. M., & Patel, K. S. (2024). LSTM-based forecasting of dengue cases in Gujarat: A machine learning approach. *Indian Journal of Science and Technology*. <https://doi.org/10.17485/IJST/v17i6.2349>
- Sukanda, A., Setiawan, W., & Kurniawan, D. (2023). Prediksi harga cabai rawit merah menggunakan long short-term memory di Kota Tasikmalaya. *Jurnal Ilmiah Teknik Informatika*, 11(2), 123–132.
- World Health Organization. (2021). *Dengue and severe dengue*. <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>
- Yoani, A., Sediono, S., Mardianto, M. F. F., & Pusporani, E. (2024). Prediksi jumlah kejadian banjir bulanan di Indonesia berdasarkan analisis long short-term memory. *Jurnal Teknik Informatika dan Sistem Informasi*, 10(3), 567–578. <https://doi.org/10.28932/jutisi.v10i3.8234>