G-Tech : Jurnal Teknologi Terapan
Volume 7, No. 4, Oktober 2023, hal. 1663-1672
E-ISSN: 2623-064x | P-ISSN: 2580-8737

Prediksi Jumlah Kejadian Banjir Bulanan di Indonesia Berdasarkan Analisis *Long Short Term Memory*

Alfredi Yoani[1], Sediono[2]✉, M. Fariz Fadillah Mardianto[3], Elly Pusporani[4]

[1, 2, 3, 4] Statistika, Fakultas Sains dan Teknologi, Universitas Airlangga, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 08-10-2023
Direvisi : 12-10-2023
Diterima : 15-10-2023

ABSTRAK

Banjir merupakan salah satu bencana alam yang paling umum dan berbahaya di dunia. Terjadinya banjir dapat menyebabkan kehilangan nyawa, hingga ketidakstabilannya perekonomian negara. Di Indonesia sendiri, banjir merupakan bencana alam yang paling sering terjadi sejak tahun 2009. Tingginya frekuensi tersebut menambah urgensi untuk memprediksi jumlah kejadian bencana alam guna membantu pemerintah dan masyarakat dalam mengambil langkah-langkah mitigasi yang tepat, serta membantu percepatan tercapainya *Sustainable Development Goals* ke-15 mengenai Ekosistem Darat. Metode yang digunakan untuk memprediksi jumlah kejadian banjir bulanan di Indonesia adalah *Long Short Term Memory* (LSTM). Metode LSTM dipilih karena memiliki kemampuan untuk memproses data sekuensial dalam jangka waktu yang panjang. Setelah dianalisis, diperoleh hasil peramalan yang sangat akurat, dengan nilai *Mean Absolute Persentage Error* (MAPE) sebesar 8,04% dan *Root Mean Square Error* (RMSE) sebesar 5,991. Model juga dapat mengestimasi data *training* dengan sangat baik, yaitu dengan nilai R^2 sebesar 95,71%.

Kata Kunci:

Banjir, LSTM, MAPE, Prediksi, SDGs

Keywords **:**

Flood, LSTM, MAPE, Forecasting, SDGs

ABSTRACT

Floods are among the most common and dangerous natural disasters worldwide, leading to loss of life and economic instability. In Indonesia, floods have been the most frequently occurring natural disaster since 2009. The high frequency underscores the urgency of predicting the number of natural disaster events to assist the government and the public in taking appropriate mitigation measures, as well as contributing to the achievement of Sustainable Development Goal 15 regarding Terrestrial Ecosystems. The method used to predict the monthly occurrence of floods in Indonesia is Long Short Term Memory (LSTM). LSTM was chosen for its ability to process sequential data over a long period of time. Upon analysis, highly accurate forecasting results were obtained, with a Mean Absolute Percentage Error (MAPE) of 8.04%, a Root Mean Square Error (RMSE) of 5.991. The model is also proficient at estimating training data, with an R^2 value of 95.71%.

Corresponding Author **:**
Sediono
Statistika, Fakultas Sains dan Teknologi, Universitas Airlangga, Indonesia
Jl. Dr. Ir. H. Soekarno, Mulyorejo, Kec. Mulyorejo, Surabaya, Jawa Timur 60115
Email: sediono101@gmail.com

PENDAHULUAN

Banjir merupakan salah satu bencana alam yang paling umum dan berbahaya di dunia (Liu *et al.*, 2019). Banjir seringkali menyebabkan kerusakan yang tidak terkendali, terutama di negara-negara berpenghasilan rendah yang sistem infrastruktur ataupun drainasenya tidak memadai (Makbul *et al.*, 2023). Terjadinya banjir dapat menyebabkan kerusakan masif, kehilangan nyawa, kerusakan infrastruktur, hingga ketidakstabilannya perekonomian negara (Lakawa, 2020). Selain itu, air banjir yang menggenang juga dapat menyebarkan penyakit menular karena mengandung mikroorganisme penyebab penyakit dari kotoran seperti sampah dan limbah septic tank (Nurullita *et al.*, 2021). Dengan demikian, terjadinya bencana alam dapat mengganggu tercapainya poin *Sustainable Development Goals* ke-15 mengenai Ekosistem Darat karena kehidupan tanaman dan hewan serta keanekaragaman hayati di daratan menjadi terancam (United Nations, 2022).

Berdasarkan studi yang dilakukan oleh World Bank, diperkirakan bahwa 1,47 miliar orang di seluruh dunia secara langsung terpapar risiko banjir yang intens dan hampir 600 juta orang diantaranya adalah orang miskin. Selain itu, sekitar 2,2 miliar orang atau 29% dari populasi dunia juga tinggal di lokasi yang rawan terjadi banjir. Banjir merupakan ancaman yang bersifat universal karena terdapat 189 negara di seluruh dunia yang penduduknya dianalisis tidak aman dari banjir dan mayoritas dari negara tersebut berasal dari negara di Asia (Rentschler & Salhab, 2020).

Di Indonesia sendiri, menurut data dari Badan Nasional Penanggulangan Bencana (BNPB), banjir merupakan bencana alam yang paling sering terjadi sejak tahun 2009 hingga bulan Juli 2023, yaitu sebanyak 11.679 kejadian. Hal ini tentu mengkhawatirkan mengingat dari tiga tahun terakhir saja (2020 - 2022), jumlah fasilitas pendidikan dan kesehatan yang rusak berturut-turut 1.457 dan 386 fasilitas. Tak hanya itu, jumlah rumah yang rusak juga diperkirakan mencapai 131.458 rumah (Badan Nasional Penanggulangan Bencana, 2023).

Tingginya frekuensi bencana alam banjir yang terjadi di Indonesia tentu saja menambah urgensi dari penelitian ini. Prediksi jumlah kejadian bencana alam yang tepat dapat membantu pemerintah dan masyarakat dalam mengambil langkah-langkah mitigasi yang tepat. Jika pemerintah ataupun masyarakat tidak siap dengan adanya kejadian banjir, hal ini dapat menyebabkan pada terlambatnya mitigasi bencana alam, sehingga dapat berdampak pada kerugian-kerugian yang lebih masif.

Terdapat beberapa penelitian terdahulu mengenai prediksi kejadian banjir, salah satunya berjudul Prediksi Kejadian Banjir dengan *Ensemble Machine Learning* Menggunakan *Back Propagation Neural Network* (BP-NN) dan *Support Vector Machine* (SVM) (Fitriyaningsih & Basani, 2019). Pada penelitian tersebut, kejadian banjir dalam rentang satu minggu ke depan telah berhasil diprediksi dengan akurat. Akan tetapi, hingga saat ini, belum terdapat jurnal yang memprediksi jumlah kejadian banjir dalam tiap bulannya. Dengan demikian, kebaruan penelitian ini adalah dengan membangun model yang mampu untuk memprediksi jumlah banjir tiap bulannya dengan akurat.

Metode *Long Short Term Memory* (LSTM) juga telah pernah digunakan untuk memprediksi berbagai data, salah satunya berjudul Implementasi Metode *Long Short Term Memory* untuk Memprediksi Kualitas Udara di DKI Jakarta (Arkadia, 2022). Dengan metode LSTM, didapatkan hasil prediksi kadar konsentrasi partikulat (PM10) yang sangat memuaskan, yaitu MAPE sebesar 4,37%. Metode LSTM sendiri memiliki berbagai keuntungan, salah satunya adalah kemampuan memori jangka pendek dan panjang yang memungkinkan LSTM untuk mengingat informasi penting dalam berbagai jangka waktu (Pamungkas, 2023). Selain itu, LSTM juga memiliki kemampuan yang baik dalam penanganan data berurutan karena dapat memahami pola sekuensial dan melibatkan konteks historis dalam prediksi (Pipin *et al.*, 2023). Dengan segala kelebihan dari LSTM, metode tersebutlah digunakan pada penelitian ini untuk memprediksi jumlah kejadian banjir.

METODE PENELITIAN

Langkah-Langkah Analisis Data

Data yang digunakan dalam penelitian ini merupakan data bulanan dari jumlah kejadian banjir di Indonesia yang diperoleh dari situs resmi Data Informasi Bencana Indonesia (DIBI) di bawah naungan Badan Nasional Penanggulangan Bencana (BNPB) mulai dari bulan Januari 2009 hingga bulan Juli 2023 (Badan Nasional Penanggulangan Bencana, 2023).

Berdasarkan percobaan yang dilakukan oleh Montesinos-López (2023), pembagian dataset yang optimal adalah 85% data *training* dan 15% data *testing*, sehingga proporsi pembagian tersebut diterapkan pada penelitian ini (Montesinos-López *et al*., 2023). Data *training* merupakan data yang dimanfaatkan dalam proses pembangunan model, sementara data *testing* merupakan data yang digunakan untuk membandingkan hasil peramalan dengan data aktual guna mengevaluasi kualitas model tersebut. Berikut merupakan langkah-langkah analisis dari penelitian ini:

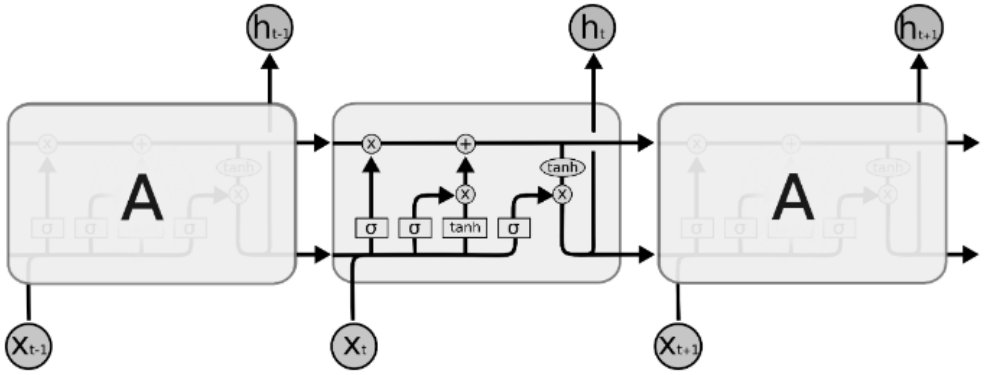
1. Mengimpor data yang telah dikumpulkan.
2. Melakukan preprocessing dengan menormalisasi data menjadi normal baku.
3. Memisahkan rangkaian data menjadi data *training* dan data *testing*.
4. Melakukan proses *training* data dengan memasukkan data *training* ke dalam arsitektur LSTM.
5. Mengevaluasi ukuran kebaikan model.
6. Ulangi langkah 4 dengan mengubah *hyperparameter tuning* hingga mendapatkan model yang optimal.
7. Menyimpan model yang diperoleh dari proses *training* dengan *hyperparameter tuning* optimal.
8. Melakukan prediksi.
9. Melakukan transformasi data kembali dari hasil prediksi yang didapatkan.
10. Memvisualisasikan hasil prediksi.

Long Short Term Memory (LSTM)

LSTM pertama kali diperkenalkan oleh Sepp Hochreiter dan Jurgen Schmidhuber pada tahun 1997. LSTM adalah sebuah perkembangan dari model Jaringan Neural Rekuren (RNN) yang diciptakan untuk mengatasi masalah yang timbul saat mengolah data berurutan dalam jangka waktu yang panjang, yang dikenal sebagai masalah gradien menghilang (Husada & Toba, 2020). Arsitektur RNN mengalami kendala dalam mengatasi ketergantungan jangka panjang karena kurang efisien dalam mempertahankan informasi dari periode sebelumnya (Pradana *et al*., 2023). Dengan kata lain, informasi lama yang tersimpan dalam RNN cenderung menjadi kurang relevan dan tergantikan oleh informasi yang baru.

LSTM mengatasi masalah gradient menghilang dengan cara menggunakan sel memori (*memory cell*) untuk menyimpan informasi selama periode waktu tertentu, dan juga memanfaatkan berbagai unit pintu (*gate units*) untuk mengendalikan aliran informasi yang masuk dan keluar dari sel memori ini (Alghifari *et al*., 2022). Dengan pendekatan ini, LSTM mampu lebih efektif dalam mengatasi ketergantungan jangka panjang dalam data sekuensial.

Pada LSTM terdapat beberapa gate unit yang berinteraksi dalam cara yang sangat khusus, antara lain: *forget gate*, *input gate*, dan *output gate* (Rafael & Adikara, 2023). Gambaran dari arsitektur LSTM ditunjukkan pada Gambar 1.



Gambar 1. Arsitektur Umum LSTM

Lapisan pertama adalah *forget gate* (f_t). *Forget gate* adalah sebuah komponen pada model LSTM yang berfungsi untuk menentukan apakah informasi yang ada dalam cell state akan dihapus atau dipertahankan. *Forget gate* mengambil *output* pada waktu (t – 1) dan input pada waktu t untuk diterapkan pada fungsi aktivasi sigmoid. Karena memanfaatkan fungsi aktivasi sigmoid, maka *output* dari *forget* gate akan bernilai 0 atau 1. Jika $f_t = 0$ maka cell state sebelumnya akan dilupakan, sementara jika $f_t = 1$ cell state sebelumnya akan dipertahankan. Persamaan dari *forget gate* akan dirumuskan sebagai berikut:

$$f_t = \sigma\left[W_{fh} \times h_{t-1} + W_{fx} \times x_t + b_f\right] \quad (1)$$

dengan:
f_t : *Forget gate*
σ : Fungsi aktivasi sigmoid
W_{fh} : Nilai pembobot *forget gate* untuk nilai h
h_{t-1} : Nilai *output* pada saat $(t-1)$
W_{fx} : Nilai pembobot *forget gate* untuk nilai x
x_t : Nilai *input* pada saat t
b_f : Nilai galat pada *forget gate*

Lapisan kedua adalah *input gate*. *Input gate* sendiri berisi penjumlahan dari *input gate* (i_t) dan input *node* (g_t). Sama seperti *forget gate*, *input gate* berperan untuk mengambil *output* pada waktu (t – 1) dan input pada waktu t serta diterapkan pada fungsi aktivasi sigmoid. Selanjutnya, perhitungan dari input *node* juga mirip seperti *input gate*, tetapi dengan nilai pembobot yang berbeda dan dilewatkan melalui fungsi aktivasi tangen hiperbolik (tanh). Persamaan dari *input gate* dan input *node* akan dirumuskan sebagai berikut:

$$i_t = \sigma[W_{ih} \times h_{t-1} + W_{ix} \times x_t + b_i] \quad (2)$$

dengan:
i_t : Nilai *input gate*
W_{ih} : Nilai pembobot *input gate* untuk nilai h
W_{ix} : Nilai pembobot *input gate* untuk nilai x
b_i : Nilai galat pada *input gate*

$$g_t = tanh\left[W_{gh} \times h_{t-1} + W_{gx} \times x_t + b_g\right] \quad (3)$$

dengan:
g_t : Nilai i*nput node*
$tanh$: Fungsi aktivasi tangen hiperbolik
W_{gh} : Nilai pembobot *input node* untuk nilai h
W_{gx} : Nilai pembobot *input node* untuk nilai x
b_g : Nilai galat pada *input node*

Setelah mendapatkan hasil dari *forget gate*, *input gate* dan *input node*, maka informasi pada *cell state* dapat diperbaharui dengan perhitungan sebagai berikut:

$$C_t = f_t \times C_{t-1} + i_t \times g_t \quad (4)$$

dengan:
C_t : *Cell state* yang baru pada saat t
f_t : Nilai *forget gate*
C_{t-1} : *Cell state* pada saat $(t-1)$

Lapisan terakhir adalah *output gate*. Lapisan ini memiliki peran penting dalam menentukan jenis informasi yang akan menjadi keluaran berdasarkan input dan *cell state*. Persamaan untuk *output gate* ini dirumuskan sebagai berikut:

$$O_t = \sigma[W_{oh} \times h_{t-1} + W_{ox} \times x_t + b_o] \quad (5)$$

dengan:
O_t : Nilai *output gate*
W_{oh} : Nilai pembobot *output gate* untuk nilai h
W_{ox} : Nilai pembobot *output gate* untuk nilai x
b_o : Nilai galat pada *output gate*

Selanjutnya, *cell state* diproses melalui fungsi aktivasi tanh terlebih dahulu, kemudian dikalikan dengan hasil *output gate* yang melalui fungsi aktivasi sigmoid agar *output* yang dihasilkan sesuai dengan keputusan sebelumnya. Perhitungan akhir dari *output* yang dihasilkan adalah sebagai berikut:

$$h_t = \tanh(C_t) \times O_t \quad (6)$$

Metriks Evaluasi

RMSE merupakan besarnya tingkat kesalahan hasil prediksi, dimana semakin kecil (mendekati 0) nilai RMSE maka hasil prediksi akan semakin akurat. RMSE memberikan bobot yang lebih besar pada kesalahan yang lebih signifikan karena menghitung akar kuadrat dari nilai kesalahan (Anwar, 2023). Hal ini membuat RMSE lebih sensitif terhadap nilai-nilai yang jauh dari nilai aktual. Perhitungan matematis dari RMSE dapat dituliskan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \times \sum(x_i - \hat{x}_i)^2} \quad (7)$$

dengan:
x_i : Nilai data aktual
$\hat{x}_i$: Nilai hasil prediksi
n : Banyaknya data

R^2 adalah metriks evaluasi yang umum digunakan dalam mengukur tingkat kecocokan model dengan data. R^2 memberikan informasi tentang sejauh mana model cocok dengan data yang diamati atau diprediksi. Semakin tinggi nilai R^2, semakin baik model dianggap cocok dengan data. Tidak hanya itu, nilai R^2 juga dapat mendeteksi apabila terjadinya *overfitting* karena R^2 dapat meningkat jika model terlalu kompleks dan "menghafal" data historis, sehingga menghasilkan peramalan yang buruk untuk data baru. Perhitungan matematis dari R^2 dapat dituliskan sebagai berikut:

$$R^2 = 1 - \frac{Sum\ Square\ Error}{Sum\ Square\ Total} = 1 - \frac{\sum(x_i - \hat{x}_i)^2}{\sum(x_i - \bar{x}_i)^2} \quad (8)$$

MAPE adalah ukuran ketepan relatif yang digunakan untuk mengetahui besar persentase penyimpangan hasil estimasi. MAPE seringkali digunakan untuk mengindikasi seberapa besar

kesalahan dalam meramal yang dibandingkan dengan nilai nyata. Perhitungan matematis dari MAPE dapat dituliskan sebagai berikut:

$$MAPE = \frac{1}{n} \times \sum_{0}^{n} \left| \frac{\bar{x}_i - x_i}{x_i} \right| \times 100\% \quad (9)$$

Interpretasi dari nilai MAPE dirangkum pada Tabel 1 (Lewis, 1982).

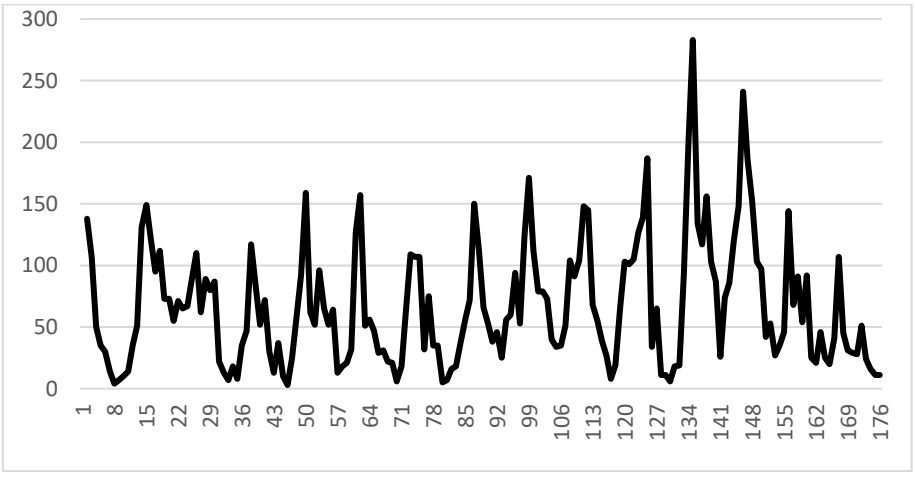
Tabel 1. Kategori Nilai MAPE

Range MAPE	Deskripsi
> 50%	Model memiliki kemampuan peramalan yang tidak akurat.
20% - 50%	Model memiliki kemampuan peramalan yang cukup akurat.
10% - 20%	Model memiliki kemampuan peramalan yang akurat.
<10%	Model memiliki kemampuan peramalan yang sangat akurat.

HASIL DAN PEMBAHASAN

Statistik Deskriptif

Visualisasi dari data yang telah dihimpun akan ditampilkan pada Gambar 2.



Gambar 2. Grafik Runtun Waktu Jumlah Kejadian Banjir Bulanan di Indonesia

Berdasarkan grafik runtun waktu tersebut, dapat terlihat dengan jelas bahwa jumlah kejadian dari banjir di Indonesia selalu mencapai titik maksimum lokal pada awal tahun dan kebalikannya mencapai titik minimum lokal pada pertengahan tahun.

Dengan demikian, dapat menyimpulkan bahwa jumlah kejadian banjir terbanyak terdapat pada bulan februari 2020 yang lalu dengan total banjir mencapai 283 kejadian di seluruh indonesia. Kebalikannya, jumlah kejadian banjir paling sedikit terdapat pada bulan september 2021 dengan total banjir hanya sebanyak 3 kejadian di seluruh indonesia.

Hyperparameter Tuning Terbaik

Pertama-tama, agar mendapatkan hasil yang terbaik, harus dinentukan *hyperparameter tuning* yang terbaik terlebih dahulu. pada penelitian ini, akan dicoba berbagai kemungkinan *number of LSTM unit*, *batch size*, dan *learning rate*. Perlu menjadi catatan bahwa setiap percobaan menggunakan *epoch* sebanyak 650 dengan pengambilan model dari *epoch* yang memiliki *validation loss* terendah. Langkah ini diambil agar mengurangi risiko *overfitting*. Metriks evaluasi dari masing-masing percobaan akan ditampilkan pada Tabel 2.

Tabel 2. Hasil Percobaan *Hyperparameter tuning*

Number of LSTM unit	*Batch size*	*Learning rate*	MAPE	RMSE	R^2
16	**16**	0,01	11,35%	5,797	94,79%
		0,001	**8,04%**	**5,991**	**95,71%**
		0,0001	16,87%	8,075	21,21%
	32	0,01	11,32%	5,907	95,13%
		0,001	12,24%	6,261	94,85%
		0,0001	21,48%	9,954	71,71%
	64	0,01	10,70%	5,581	95,25%
		0,001	11,50%	5,951	93,86%
		0,0001	28,71%	13,099	52,57%
32	16	0,01	13,56%	7,049	96,19%
		0,001	12,40%	6,306	94,66%
		0,0001	12,42%	6,101	84,43%
	32	0,01	11,35%	6,310	96,51%
		0,001	11,68%	5,977	94,87%
		0,0001	16,58%	8,008	77,88%
	64	0,01	11,53%	5,988	95,09%
		0,001	11,82%	6,095	94,31%
		0,0001	28,08%	12,846	56,90%
64	16	0,01	13,62%	7,208	97,15%
		0,001	13,35%	6,735	94,22%
		0,0001	10,71%	5,551	90,34%
	32	0,01	11,77%	6,380	96,26%
		0,001	11,04%	5,866	95,25%
		0,0001	15,74%	7,451	77,20%
	64	0,01	11,41%	5,947	94,89%
		0,001	11,32%	5,843	94,92%
		0,0001	16,52%	7,822	74,52%

Berdasarkan percobaan tersebut, didapatkan nilai R^2 terbaik dengan penggunaan *number of LSTM unit* sebanyak 64, *batch size* berukuran 16, dan *learning rate* sebesar 0,01. hal ini menandakan bahwa model yang diperoleh dapat mempelajari data *training* dengan sangat baik. Selain itu, perlu diperhatikan juga bahwa nilai MAPE dan RMSE yang didapatkan berturut-turut 13,62% dan 7,208. angka tersebut menunjukkan bahwa model yang diperoleh berhasil untuk memprediksi jumlah kejadian banjir dengan baik.

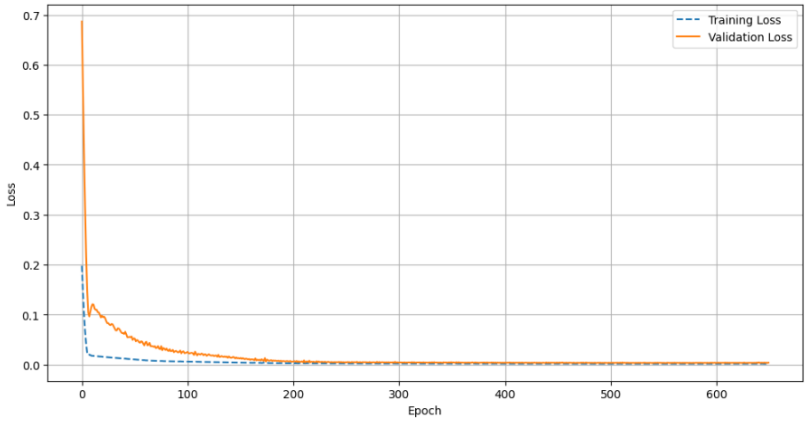
Bila ditinjau dari nilai MAPE yang terendah, model yang terbaik didapatkan melalui penggunaan *number of LSTM unit* sebanyak 16, *batch size* berukuran 16, dan *learning rate* sebesar 0,001. hal ini menandakan bahwa model dapat menghasilkan nilai prediksi yang sangat akurat. Selain itu, model ini juga memiliki nilai RMSE terendah, yaitu sebesar 5,991. Artinya, tingkat kesalahan prediksi juga tergolong relatif rendah. Untuk model ini, didapatkan nilai R^2 Sebesar 95,71% yang tidak terpaut jauh dari model dengan nilai R^2 tertinggi (97,15%).

Dengan demikian, setelah mempertimbangkan metriks evaluasi dari percobaan-percobaan tersebut, dapat ditarik sebuah kesimpulan bahwa model terbaik dalam memprediksi jumlah kejadian banjir adalah dengan menerapkan *hyperparameter tuning number of LSTM unit* sebanyak

16, *batch size* berukuran 16, dan *learning rate* sebesar 0,001. salah satu pertimbangan terbesar dalam memilih model tersebut adalah karena menghasilkan nilai prediksi dengan MAPE terbaik dengan nilai dibawah 10%, yaitu sebesar 9,55% dan nilai RMSE terendah, yaitu 5,991. Walaupun nilai R^2 sebesar 95,71% bukan yang tertinggi dibanding model lainnya, namun angka tersebut sudah cukup memuaskan karena sudah berhasil mengestimasi data *training* dengan sangat baik.

Hasil Analisis LSTM

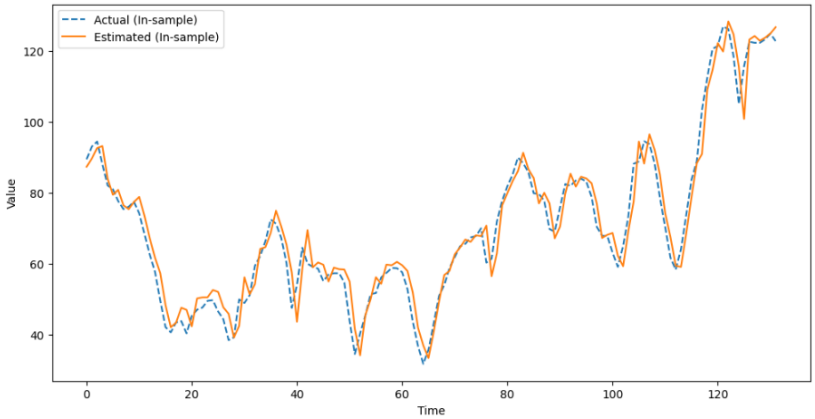
Pertama-tama, untuk memastikan bahwa model tersebut telah diberikan jumlah *epoch* yang terbaik, akan ditunjukkan grafik dari *training* loss dan *validation loss* untuk tiap iterasi *epoch* pada Gambar 3.



Gambar 3. Grafik *Training Loss* dan *Validation Loss* Tiap *Epoch*

Setelah melalui *epoch* sejumlah 650, model berhasil menyentuh nilai *validation loss* terendah, yaitu senilai 0,00325 pada *epoch* ke 431. Hal ini memungkinkan karena model akan disimpan sesuai dengan *epoch* yang nilai *validation loss*-nya terendah. Jumlah *epoch* dibatasi pada 650 untuk menghindari terjadinya *overfitting*, karena ketika jumlah *epoch* di atas 650, model cenderung memiliki nilai R^2 mendekati 99% namun dengan nilai MAPE di atas 20% yang tentunya bukan merupakan hasil yang diinginkan.

Berikutnya, peforma model tersebut dalam mengestimasi data *training* ditampilkan melalui *time series* plot gabungan antara data aktual dan data hasil estimasi pada Gambar 4.



Gambar 4. Grafik Runtun Waktu Data Aktual dan Hasil Estimasi Data *Training*

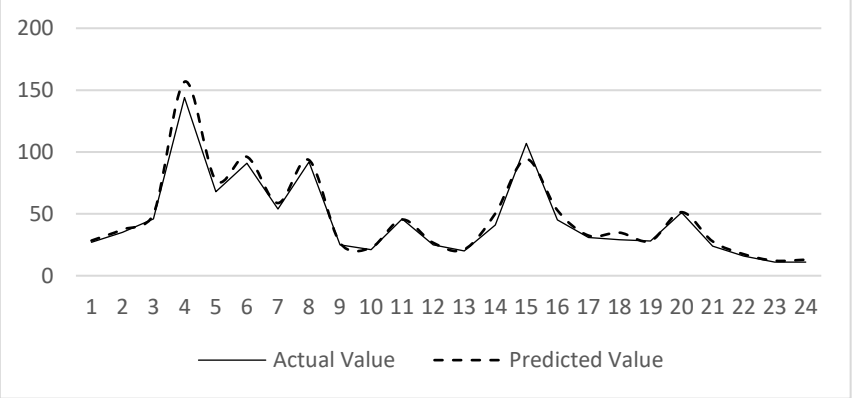
Melihat plot tersebut, dapat dikatakan bahwa model memiliki peforma yang sangat baik dalam mengestimasi data in sample, terlihat secara visual bahwa garis data estimasi hampir mendekati data aktualnya. Secara objektif, pernyataan tersebut dapat diperkuat dengan melihat nilai dari R^2 *training* atau *in-sample*, yaitu sebesar 95,71%. Angka tersebut merupakan angka yang sangat memuaskan bagi sebuah nilai R^2 karena telah mendekati sempurna dalam pengestimasian.

Berikutnya, hasil prediksi dari model tersebut untuk beberapa jangka waktu ke depan akan ditampilkan pada Tabel 3.

Tabel 3. Hasil Prediksi Jumlah Kejadian Banjir

Tahun	Bulan	Nilai Aktual	Nilai Prediksi	Tahun	Bulan	Nilai Aktual	Nilai Prediksi
2021	Agustus	27	28,38	2022	Agustus	20	21,24
2021	September	35	37,42	2022	September	41	50,15
2021	Oktober	46	50,02	2022	Oktober	107	94,22
2021	November	144	156,78	2022	November	45	52,98
2021	Desember	68	77,31	2022	Desember	31	32,50
2022	Januari	91	96,14	2023	Januari	29	34,83
2022	Februari	54	58,72	2023	Februari	28	28,11
2022	Maret	92	93,53	2023	Maret	51	51,36
2022	April	25	26,73	2023	April	24	27,80
2022	Mei	21	22,64	2023	Mei	16	17,41
2022	Juni	46	45,39	2023	Juni	11	12,16
2022	Juli	25	26,67	2023	Juli	11	12,96

Setelah mendapatkan hasil prediksi tersebut, dapat dibuatkan *time series* plot gabungan antara data aktual dan data hasil estimasi untuk data *testing*.



Gambar 5. Plot Runtun Waktu Data Aktual dan Hasil Estimasi Data *Testing*

Secara visual, dapat terlihat bahwa hasil prediksi kejadian banjir sudah sangat baik karena sangat mendekati data aktualnya. agar dapat mengetahui hasil prediksi dengan lebih objektif, dapat dilihat dari nilai MAPE, yaitu sebesar 8,04%. Artinya model memiliki kemampuan kemampuan peramalan yang sangat akurat karena memiliki nilai MAPE dibawah 10%.

KESIMPULAN DAN SARAN

Kesimpulan

Dalam memprediksi jumlah kejadian bencana alam banjir, model yang didapatkan melalui metode LSTM memiliki kemampuan peramalan yang sangat akurat. Dengan pemilihan *hyperparameter number of LSTM unit* sebanyak 16, *batch size* berukuran 16, dan *learning rate* sebesar 0,001, model berhasil mendapatkan nilai MAPE sebesar 8,04%, RMSE sebesar 5,991, dan R^2 sebesar 95,71%. Harapannya, hasil penelitian ini dapat menjadi acuan bagi pemerintah untuk meningkatkan kualitas mitigasi bencana alam di Indonesia, serta membantu percepatan tercapainya poin *Sustainable Development Goals* ke-15 mengenai Ekosistem Darat.

Saran

Untuk penelitian selanjutnya, disarankan untuk dapat mencari pengaruh jumlah bencana alam banjir dari faktor-faktor lainnya. Selain itu, dapat juga membandingkan hasil penelitian ini dengan pengembangan model LSTM lainnya.

REFERENSI

Alghifari, D. R., Edi, M., & Firmansyah, L. (2022). Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia. *Jurnal Manajemen Informatika (JAMIKA)*, 12(2), 89-99.

Anwar, G. S. (2023). Implementasi dan Optimalisasi Metode Single Moving Average untuk Memprediksi Stok Barang Percetakan dan ATK. *(Doctoral dissertation, Universitas Siliwangi).*

Arkadia, A. H. (2022). Optimasi Long Short Term Memory Dengan Adam Menggunakan Data Udara Kota DKI Jakarta. *In Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, 3(2), 599-609.

Badan Nasional Penanggulangan Bencana. (2023). Statistik Bencana Menurut Waktu. Retrieved from BNPB: *dibi.bnpb.go.id*

Fitriyaningsih, I., & Basani, Y. (2019). Prediksi Kejadian Banjir dengan Ensemble Machine Learning Menggunakan BP-NN dan SVM. *Jurnal Teknologi dan Sistem Komputer*, 7(3), 93-97.

Husada, I., & Toba, H. (2020). Pengaruh Metode Penyeimbang Kelas Terhadap Tingkat Akurasi Analisis Sentimen pada Tweets Berbahasa Indonesia. *Jurnal Teknik Informatika dan Sistem Informasi*, 6(2,: 400-413.

Lakawa, I. (2020). Analisis Banjir Faktor Penyebab Dan Prioritas Penanganan Sungai Anduonuhu. *Sultra Civil Engineering Journal*, 1(2), 54-71.

Lewis, C. D. (1982). *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. London; Boston: Butterworth Scientific.

Liu, Y., You, M., Zhu, J., Wang, F., & Ran, R. (2019). Integrated Risk Assessment for Agricultural Drought and Flood Disasters Based on Entropy Information Diffusion Theory in the Middle and Lower Reaches of the Yangtze River, China. *International Journal of Disaster Risk Reduction*, 38, 101194.

Makbul, R., Zulhamah, H. R., Tanje, H. R., Sugiarto, Djufri, H., Bungin, E.R., Faisal, Z., Wijaya, Y., Johra, Firdaus, M., & Sudirman. (2023). *Pengembangan Sumber Daya Air*. Makassar: CV Tohar Media.

Montesinos-López, O. A., Saint, P. C., Gezan, S. A., Bentley, A. R., Mosqueda-González, B. A., Montesinos-López, A., & Crossa, J. (2023). Optimizing Sparse *Testing* for Genomic Prediction of Plant Breeding Crops. *Genes*, 14(4), 927.

Nurullita, U., Ritonga, G. M., & Mifbakhuddin, M. (2021). Pengetahuan Warga Tentang Bahaya Keselamatan dan Bahaya Kesehatan yang Terjadi pada Banjir (Studi di Daerah Rawan Banjir di Bandarharjo Semarang). *Jurnal Kesehatan Masyarakat Indonesia*, 16(3), 154-159.

Pamungkas, T. H. (2023). Perbandingan Prediksi Pergerakan Data IHSG Menggunakan Feed Forward Neural Network dan Long Short Term Memory. *(Doctoral dissertation, Institut Teknologi Sepuluh Nopember)*.

Pipin, S. J., Purba, R., & Kurniawan, H. (2023). Prediksi Saham Menggunakan Recurrent Neural Network (RNN-LSTM) dengan Optimasi Adaptive Moment Estimation. *Journal of Computer System and Informatics*, 4(4), 806-815.

Pradana, Y. A., Cholissodin, I., & Kurnianingtyas, D. (2023). Analisis Sentimen Pemindahan Ibu Kota Indonesia pada Media Sosial Twitter menggunakan Metode LSTM dan Word2Vec. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(5), 2389-2397.

Rentschler, J., & Salhab, M. (2020). People in Harm's Way: Flood Exposure and Poverty in 189 Countries. *The World Bank*.

United Nations. (2022). *The Sustainable Development Goals Report 2022*. New York: Department of Economic and Social Affairs (DESA).