

I. PENDAHULUAN

1.1 Latar Belakang

Indonesia adalah negara beriklim tropis karena terletak di garis khatulistiwa (Susilo, 2021). Iklim tropis ini membuat Indonesia rentan terhadap perubahan pola cuaca terutama dalam hal curah hujan (Toyib et al., 2024). Cuaca adalah kualitas udara yang diamati dalam waktu yang relatif singkat atau di daerah yang kecil (BMKG, 2021). Cuaca menurut WCC (*World Climate Conference*) adalah keadaan atmosfer yang diukur secara menyeluruh dengan memperhitungkan perubahan, perkembangan, dan datang atau lenyapnya suatu fenomena udara (Luthfiarta et al., 2020). Unsur-unsur yang memengaruhi cuaca yaitu suhu, udara, tekanan udara, kelembaban udara, laju uap air, awan, curah hujan dan angin (Farikhul Firdaus & Papatungan, 2022). Curah hujan sebagai salah satu unsur yang memengaruhi cuaca didefinisikan sebagai ketinggian air hujan yang terkumpul di alat penakar hujan pada permukaan datar yang tidak menyerap, meresap, mengalir, ataupun menguap. Semakin tinggi ketinggian air hujan yang terukur menunjukkan semakin banyaknya frekuensi hujan atau intensitas hujan yang terjadi di wilayah tersebut (Prayoga Dhenanta & Kholifah, 2022).

Menurut Putratama (2020), pemanasan global telah menyebabkan pola cuaca menjadi semakin tidak terduga. Peningkatan suhu udara global memengaruhi kelembaban, pola angin, dan intensitas curah hujan di seluruh dunia, termasuk Indonesia. Curah hujan adalah elemen penting dalam cuaca yang sangat memengaruhi kehidupan sehari-hari, mulai dari sektor pertanian hingga perencanaan infrastruktur (Agungnoe, 2024). Namun, sifat curah hujan yang fluktuatif membuatnya sulit diprediksi secara akurat (Wicaksono, 2022). Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), sebagai lembaga resmi di Indonesia yang bertugas memberikan informasi prakiraan cuaca, telah menyediakan data historis dan model prediksi untuk berbagai wilayah, termasuk Kota Jambi. Namun, data curah hujan sering kali mengandung nilai yang tidak terukur atau rusak (*missing value*). Nilai 8888 menunjukkan curah hujan yang tidak terukur dengan intensitas <0,5 mm, yang dapat digantikan dengan simbol TTU atau nilai 0. Sementara itu, nilai 9999 menunjukkan data kosong akibat alat pengukur yang rusak atau data yang belum tercatat (BMKG, 2024). Keberadaan nilai-nilai ini dapat mengganggu analisis data, karena dalam suatu penelitian, kualitas data sangat mempengaruhi hasil dari suatu penelitian (Ericko et al., 2023). *Missing value* dapat menyebabkan tingkat keakuratan suatu data menjadi

berkurang dan menurunnya kualitas data pada saat akan dilakukan pengolahan data lanjut, seperti prediksi (Lutfi & Hasyim, 2019).

Penanganan *missing value* menjadi tahap yang sangat penting dalam *preprocessing* data, terutama untuk data *time series* seperti curah hujan (Riko Anshori Prasetya & Mudi Priyatno, 2023). Salah satu teknik yang dapat digunakan untuk menangani *missing value* adalah interpolasi (Widianti & Pratama, 2024). Teknik interpolasi digunakan untuk memperkirakan nilai yang hilang dengan cara mengisi data berdasarkan pola di antara titik data yang diketahui (Febrianti, 2019). Dalam penelitian ini, tiga metode interpolasi akan digunakan untuk menangani *missing value* pada data curah hujan harian Kota Jambi, yaitu interpolasi linier, kuadrat, dan spline kuadratik. Teknik interpolasi linear adalah *polynomial* tingkat pertama yang digunakan untuk mengetahui nilai suatu titik dalam garis lurus (Rizky Ismail et al., 2023). Teknik interpolasi kuadrat adalah *polynomial* berderajat dua yang digunakan untuk mencari titik-titik diantara 3 buah titik. Sedangkan teknik interpolasi spline kuadratik adalah metode aproksimasi dengan melakukan interpolasi dari titik-titik x yang terletak diantara dua titik x_n dan x_{n+1} dengan mengasumsikan fungsi berbentuk *polynomial* berpangkat dua (Sofiyani & Permanasari, 2023). Penanganan *missing value* melalui teknik interpolasi ini bertujuan untuk menghasilkan dataset yang bersih dan berkualitas, yang akan digunakan dalam proses prediksi curah hujan menggunakan metode *Long Short Term Memory* (LSTM).

Penelitian mengenai prediksi yang telah dilakukan akhir-akhir ini menunjukkan bahwa prediksi menggunakan *deep learning* telah mencapai akurasi lebih dari 85% (Sulistyo Budi et al., 2021). CNN, RNN, dan *Long Short Term Memory* (LSTM) adalah model *deep learning* yang sering digunakan untuk prediksi *time series* (Tian et al., 2018). Menurut Arrofiqoh & Harintaka (2018) CNN merupakan salah satu metode *deep learning* yang mampu melakukan proses pembelajaran mandiri untuk pengenalan objek, ekstraksi objek dan klasifikasi. Dalam penelitian Nurhikmat (2018) yang berjudul Implementasi *Convolutional Neural Network* untuk *Image Classification* pada citra wayang golek memperoleh hasil Tingkat keakurasian *training* dan *testing* dalam pengklasifikasian gambar wayang golek sebesar 95% *training* dan 90% *testing*. Sementara itu RNN cocok untuk menangani data sekuensial namun RNN tradisional menghadapi masalah *vanishing gradient*, yang menyebabkan kesulitan dalam mempelajari ketergantungan jangka panjang dalam data *time series* (Arwansyah et al., 2022). Masalah *vanishing gradient* terjadi ketika nilai gradien menjadi sangat kecil seiring waktu, sehingga mengurangi kontribusinya dalam proses pembelajaran (Carnegie & Chairani, 2023). Dalam penelitian yang dilakukan oleh Irawan (2024)

menunjukkan bahwa LSTM memiliki rata-rata nilai RMSE sebesar 4.21%, lebih rendah dibandingkan RNN dengan RMSE sebesar 4.26%, menunjukkan akurasi prediksi yang lebih tinggi. Oleh karena itu, LSTM dipilih dalam penelitian ini karena kemampuannya yang unggul dalam menangkap pola data *time series* yang kompleks dan ketergantungan jangka panjang. Dibandingkan dengan metode lain, seperti *Recurrent Neural Network* (RNN), LSTM memiliki mekanisme *forget gate*, *input gate*, dan *output gate* yang dapat memitigasi masalah *vanishing gradient*, sehingga lebih efektif dalam memprediksi data dengan pola musiman atau fluktuasi tinggi seperti curah hujan (Tarkus et al., 2020).

Meskipun LSTM telah terbukti efektif dalam berbagai aplikasi prediksi, banyak penelitian sebelumnya yang tidak memberikan perhatian yang cukup terhadap penanganan *missing value* dalam dataset. Penelitian Terdahulu, seperti yang dilakukan oleh Rizki et al. (2020) dan Farikhul Firdaus & Papatungan (2022), lebih fokus pada pengembangan arsitektur model tanpa mempertimbangkan teknik yang digunakan untuk menangani *missing value*. Kekurangan ini menunjukkan adanya kebutuhan untuk mengintegrasikan metode penanganan *missing value* yang efektif sebelum model prediksi seperti LSTM diterapkan, guna meningkatkan kualitas dan akurasi prediksi.

Berdasarkan uraian di atas, penelitian ini akan berfokus pada tahap *preprocessing data*, terutama dalam menangani *missing value* pada data curah hujan. Teknik interpolasi linear, kuadrat, dan spline kuadratik dipilih untuk mengevaluasi metode yang paling efektif dalam mengisi data yang hilang. Pendekatan ini penting karena kualitas data masukan sangat memengaruhi akurasi model prediksi, sehingga diperlukan metode pengisian *missing value* yang tidak hanya sederhana tetapi juga mampu menangkap pola data secara akurat. Dataset hasil interpolasi kemudian akan digunakan untuk melatih model prediksi *Long Short Term Memory* (LSTM) dan di evaluasi menggunakan metrik *Root Mean Square Error* (RMSE) untuk menentukan metode interpolasi yang paling optimal.

Penelitian ini bertujuan untuk memberikan solusi yang lebih efektif dalam menangani *missing value* pada data curah hujan, sekaligus mengevaluasi teknik interpolasi mana dari ketiga teknik interpolasi yang paling optimal dalam mendukung prediksi curah hujan. Dengan hasil yang diperoleh, diharapkan penelitian ini dapat memberikan kontribusi signifikan dalam pengolahan data *time series*, khususnya dalam meningkatkan akurasi prediksi curah hujan di Kota Jambi.

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian maka dapat dirumuskan permasalahannya sebagai berikut:

1. Bagaimana hasil interpolasi untuk menangani *missing value* pada data curah hujan dalam penelitian ini?
2. Bagaimana mengimplementasikan metode *Long Short Term Memory* (LSTM) dalam memprediksi curah hujan di Kota Jambi?
3. Bagaimana evaluasi performa dan akurasi metode *Long Short Term Memory* dalam memprediksi curah hujan di Kota Jambi?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahui hasil interpolasi untuk menangani *missing value* pada data curah hujan dalam penelitian ini.
2. Mengimplementasikan metode *Long Short Term Memory* (LSTM) dalam memprediksi curah hujan di Kota Jambi.
3. Mengevaluasi performa dan akurasi metode *Long Short Term Memory* (LSTM) dalam memprediksi curah hujan di Kota Jambi.

1.4 Batasan Penelitian

Batasan masalah dari penelitian yang dibahas yaitu sebagai berikut:

1. Data yang digunakan didapatkan dari website Data Online BMKG (Badan Meteorologi, Klimatologi dan Geofisika) dan ukuran curah hujan menggunakan *milimeter*.
2. Penelitian difokuskan pada prediksi curah hujan di Kota Jambi.
3. Data yang digunakan adalah data curah hujan harian Kota Jambi dari tahun 2016-2024.
4. Penelitian ini menggunakan metode interpolasi linear, kuadrat, dan spline kuadrat.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Manfaat Teoritis
 - a. Penelitian ini memberikan pemahaman lebih terhadap proses interpolasi dalam menangani *missing value* pada data *time series* dengan menggunakan metode *Long Short Term Memory* (LSTM).
 - b. Sebagai bahan referensi untuk penelitian selanjutnya yang berkaitan dengan prediksi curah hujan.

2. Manfaat Praktis

- a. Hasil penelitian ini dapat digunakan sebagai alternatif metode prediksi curah hujan yang melengkapi model yang telah digunakan BMKG, khususnya Kota Jambi.
- b. Penelitian ini dapat menjadi bahan rujukan untuk pengembangan algoritma serupa dan studi terkait prediksi cuaca yang membutuhkan pendekatan yang lebih adaptif terhadap pola data yang kompleks.