

BAB I

PENDAHULUAN

1.1 Latar Belakang

Banjir didefinisikan sebagai luapan air sementara ke daratan (Pradhan dkk., 2023). Banjir bandang disebabkan oleh hujan deras dan hutan gundul pada suatu daerah (Narulita & Ningrum, 2018). Dampaknya mencakup kerusakan rumah, bisnis, infrastruktur, hingga menyebabkan korban jiwa (Marzukhi dkk., 2018). Fluktuasi curah hujan yang lebih intens dan frekuensi curah hujan yang tidak terduga dapat meningkatkan risiko banjir dan mengganggu distribusi sumber daya air yang efisien (Miller & Hutchins, 2017). Sehubungan dengan catatan bencana BNPB di Indonesia, banjir merupakan bencana dengan kejadian terbanyak pada tahun 2021 tercatat sebanyak 1.298 kejadian menurut Lia Agustina, (2021) dan tahun 2022 tercatat sebanyak 1.531 kejadian menurut Adel Andila Putri, (2023) dibandingkan bencana lainnya dan masih belum mengalami penurunan yang signifikan pada tahun 2023 tercatat pada 1255 kejadian (Admin BNPB, 2024).

Data curah hujan yang telah dimodelkan untuk menghitung pola kejadian berulang, menghadapi banyak tantangan seperti perdebatan model terbaik, keterbatasan aplikasi, dan perbedaan karakteristik hujan antar wilayah (Ximenes dkk., 2021). Berbagai metode prediksi curah hujan telah dikembangkan, masih terdapat kesenjangan signifikan dalam akurasi prediksi, terutama untuk kejadian curah hujan ekstrem. Metode konvensional, seperti MLE, MoM, dan POME, memiliki keterbatasan dalam menangkap kompleksitas pola curah hujan yang semakin tidak menentu. Pendekatan lain mencakup penggunaan distribusi *gamma*, *log-normal*, eksponensial, *Weibull*, dan distribusi probabilitas hibrida (Shen & Xiang, 2023).

Penelitian ini mengintegrasikan algoritma *Long Short Term Memory* (LSTM) dan *Isolation Forest* serta penambahan teknik augmentasi data menggunakan *Gaussian Noise*. LSTM, dengan fitur *Recurrent Neural Network*, memiliki kemampuan unggul dalam memodelkan data sekuensial seperti curah hujan (Hu dkk., 2018). Sementara itu, *Isolation Forest* menawarkan metode efektif untuk mencari anomali pada data (Hu dkk., 2018), sehingga cukup relevan untuk

diterapkan untuk mengidentifikasi potensi banjir melalui periode curah hujan tinggi.

Inovasi utama penelitian ini terletak pada integrasi algoritma *Long Short Term Memory* (LSTM) dan *Isolation Forest* dan penerapan teknik augmentasi data menggunakan *Gaussian noise*. Pendekatan ini diharapkan dapat meningkatkan akurasi prediksi curah hujan ekstrem dan identifikasi dini potensi banjir. Berbagai masalah LSTM menjadi alasan dalam penelitian ini salah satunya adalah kebutuhan jumlah data yang besar dalam melakukan pelatihan untuk menghasilkan prediksi yang akurat (Arifin dkk., 2023). Kemudian masalah ketidakseimbangan dalam distribusi data curah hujan yang berfluktuasi tinggi yang berakibat *overfitting* pada model (Pal dkk., 2020). Perbedaan signifikan dalam jumlah sampel antar kategori dapat menyebabkan model LSTM cenderung memprediksi kategori dominan secara tidak proporsional, mengakibatkan kinerja buruk dalam memprediksi kejadian langka (Wen & Li, 2023).

Urgensi masalah ketidakseimbangan distribusi data curah hujan, serta ketepatan perhitungan konvensional yang selama ini dilakukan untuk menentukan peluang curah hujan tinggi menjadi pokok pikiran pengembangan model LSTM. Dengan menambahkan *Gaussian noise* pada *dataset* dapat meningkatkan ketahanan model terhadap variasi data yang mungkin tidak tercakup dalam *dataset* asli (Arslan dkk., 2019). Kemudian Septian & Kartini, (2023) menyebutkan LSTM efektif untuk memprediksi curah hujan jangka pendek menghasilkan *forecasting* dalam waktu 1 hari dengan hasil *validation* cukup baik dibanding metode lain seperti *Convolutional Neural Network* (CNN), meskipun hasil *Deep Neural Investigation Network* (DNIN) lebih baik, namun masih bisa mengimbangi untuk hasilnya. Pada penelitian Carnegie & Chairani, (2023) menemukan LSTM lebih akurat dibandingkan GRU dalam memprediksi curah hujan. Prediksi jangka panjang masih menjadi tantangan untuk penelitian mendatang. Hal ini menjadi dasar pengembangan LSTM dalam penelitian.

Permasalahan metode statistik konvensional untuk menentukan probabilitas curah hujan puncak masih menghadapi ketidaksesuaian dan belum mampu mengatasi variasi regional serta karakteristik hujan yang berbeda (Shen & Xiang, 2023). *Isolation Forest* yang digunakan dalam penelitian ini diharapkan mampu

menyelesaikan masalah dalam menentukan curah hujan tinggi pada periode tertentu. Curah hujan puncak merupakan parameter penting dalam menghitung struktur bangunan air, karena hal ini menentukan ketahanan dan desain perencanaan terkait struktur bangunan air. Periode waktu yang dihasilkan dapat digunakan sebagai prakiraan potensi banjir.

Isolation Forest merupakan salah satu metode deteksi *outlier* yang unggul karena tidak bergantung pada asumsi distribusi data dan dapat mendeteksi *outlier* secara efisien bahkan dalam *dataset* yang besar dan kompleks (Tan dkk., 2022). Metode ini telah banyak digunakan dalam berbagai aplikasi, seperti deteksi fraud, pemantauan sistem, dan analisis anomali dalam data *time series* (Anand dkk., 2023). Beberapa studi sebelumnya telah menunjukkan keunggulan *Isolation Forest* dalam mendeteksi *outlier* dibandingkan dengan metode tradisional lainnya, seperti LOF (*Local Outlier Factor*) dan One-Class SVM. *Isolation Forest* memiliki kompleksitas komputasi yang lebih rendah dibandingkan dengan metode deteksi *outlier* berbasis jarak atau kepadatan, sehingga lebih efisien untuk diterapkan pada *dataset* yang besar (Domingues dkk., 2018). Dalam konteks penelitian ini, penggunaan *Isolation Forest* untuk mengidentifikasi *outlier* dalam data *time series* yang dimodelkan dengan LSTM akan membantu memahami pola data secara lebih baik dan meningkatkan pencarian probabilitas curah hujan puncak untuk pengambilan keputusan.

Perumusan masalah dalam penelitian ini adalah mengembangkan model prediksi curah hujan menggunakan LSTM yang dioptimalkan dengan augmentasi data *Gaussian noise* serta dikombinasikan dengan algoritma *Isolation Forest* untuk mengidentifikasi curah hujan tinggi berpotensi banjir. Masalah ini mencakup prediksi curah hujan dan identifikasi curah hujan tinggi sebagai potensi banjir. Kemudian alasan penggunaan model yang diusulkan adalah sebagai berikut :

- a) Penggunaan model LSTM : LSTM fokus dalam pemecahan masalah distribusi curah hujan pada masa depan. Hasil *forecasting* LSTM akan menjadi *input* untuk *Isolation Forest*.
- b) Penggunaan model *Isolation Forest* : *Isolation Forest* akan menganalisis hasil prediksi LSTM untuk mengidentifikasi nilai-nilai anomali atau *outlier*. Nilai-nilai yang teridentifikasi sebagai anomali ini kemudian akan dievaluasi

menggunakan *threshold* curah hujan dari BMKG untuk menentukan potensi banjir.

Penelitian ini memanfaatkan kekuatan prediktif LSTM untuk data *time series*, sementara *Isolation Forest* memberikan kemampuan untuk mendeteksi kejadian ekstrem yang mungkin tidak teridentifikasi oleh model prediksi konvensional.

Batasan penelitian ini adalah penelitian hanya dilakukan menggunakan kombinasi model LSTM dan *Isolation Forest*. Penelitian ini fokus menghasilkan model prediksi dan *forecasting* curah hujan berpotensi banjir. Hasil tidak mencakup aspek lain seperti analisis drainase, kapasitas bendungan, serta sistem informasi mitigasi banjir. Keterbatasan pada data curah hujan digunakan secara maksimal dalam penentuan model *forecasting* curah hujan. Model yang dihasilkan diinterpretasikan dengan hati-hati dan dipertimbangkan dalam konteks perencanaan dan pengambilan keputusan yang lebih luas. Penelitian ini memiliki urgensi dengan parameter dampak serius banjir dalam kurun waktu beberapa tahun terakhir. Dengan memanfaatkan masing-masing fungsi algoritma, dapat membantu dalam pengambilan keputusan yang lebih baik dalam perencanaan tata ruang dan pembangunan infrastruktur.

1.2 Tujuan

Penelitian ini bertujuan untuk menghasilkan model prediksi dan *forecasting* curah hujan berpotensi banjir dengan mengintegrasikan LSTM dan *Isolation Forest*.

1.3 Manfaat

Penelitian ini diharapkan mampu memberikan kontribusi dalam pemanfaatan model dalam prediksi curah hujan berpotensi banjir, membantu perencanaan dan pengelolaan sumber daya yang lebih baik, serta peningkatan responsibilitas pemerintah dalam menghadapi ancaman banjir.