

BAB I

PENDAHULUAN

1.1 Latar Belakang

Cuaca adalah keadaan udara pada saat tertentu dan pada wilayah tertentu dalam jangka waktu singkat (Panggabea, 2021). Prediksi curah hujan yang akurat sangat penting untuk mendukung berbagai aspek perencanaan kota, termasuk pengelolaan sumber daya air dan mitigasi risiko bencana banjir (Farisi, 2024). Kemampuan memprediksi curah hujan secara akurat sangat dibutuhkan, terutama di negara beriklim tropis yang memiliki curah hujan tinggi. Di era *big data*, penggunaan algoritma *machine learning* dalam analisis data cuaca telah menjadi metode yang efisien untuk meningkatkan ketepatan prediksi. Kondisi cuaca di Indonesia saat ini cenderung tidak stabil, dimana daerah yang awalnya terlihat cerah dapat berubah menjadi hujan atau badai dalam waktu singkat (Tanjung, 2024).

Proses perencanaan dan pengambilan keputusan menjadi lebih efektif ketika informasi cuaca tersedia dengan lengkap, akurat, dan cepat (Rohana, 2024). Memperkirakan hujan sangat penting karena memungkinkan masyarakat untuk melakukan antisipasi terhadap dampak yang mungkin terjadi dikarenakan hujan (Nursobah, 2022). Namun perubahan iklim global menyebabkan pola cuaca yang tidak menentu, dengan pergantian musim kemarau dan hujan yang sulit diprediksi (Jones, 2024).

Generalized Linear Model (GLM) dan *Linear Regression* adalah dua algoritma umum digunakan dalam analisis data, generalisasi fleksibel dari regresi linear yang mampu menangani variabel respon dengan distribusi kesalahan normal (Siregar, 2020). Sementara itu, *Linear Regression* menawarkan pendekatan yang lebih sederhana namun sangat andal untuk analisis regresi linier. Penelitian ini bertujuan untuk mengevaluasi kinerja kedua algoritma dalam memprediksi curah hujan berdasarkan data cuaca, guna menentukan metode yang lebih optimal untuk menghasilkan prediksi akurat.

Penelitian sebelumnya menunjukkan bahwa metode berbasis data dapat meningkatkan ketepatan prediksi hujan. Misalnya, penelitian oleh (Mursianto, 2021) menunjukkan bahwa algoritma Random Forest dengan resampling

menggunakan SMOTE mencapai akurasi hingga 95.59% dalam prediksi hujan. Hasil ini menunjukkan potensi penggunaan algoritma machine learning dalam meningkatkan ketepatan prediksi cuaca. Lalu penelitian lain oleh (Jones, 2024) bahwa beberapa metode Machine Learning tunggal cukup baik dalam prakiraan cuaca menggunakan parameter default dengan akurasi tinggi sebesar 99,00% pada metode Naive Bayes dan akurasi terendah sebesar 70,85% pada metode Logistic Regression. Selain itu penelitian oleh (Ambarita, 2023) hasil estimasi menggunakan model Bayesian Vector Autoregressive (BVAR) Kota Medan didapatkan curah hujan tertinggi terjadi pada bulan September sebesar 571,87 mm dan terendah terjadi pada bulan Januari sebesar 54,59 mm dengan tingkat akurasi metode sebesar 4,75% yang menunjukkan bahwa penggunaan metode BVAR dalam estimasi sangat akurat digunakan.

Adapun penelitian sebelumnya oleh (Putra & Rani, 2020) Model terbaik yang diperoleh berdasarkan studi kasus penelitian ini adalah model ANN331, yaitu model ANN yang menggunakan parameter data input berupa cuaca permukaan dan arsitektur model menggunakan 3 neuron di lapisan tersembunyi. Tingkat kesalahan yang dihasilkan model tersebut adalah 9.7 mm dengan nilai korelasi 0.5. Selanjutnya, pada penelitian (Tanjung, 2024) Model LSTM telah di tuning hyperparameter dengan variasi pengisian data kosong untuk mencapai kinerja optimal, yaitu menggunakan hyperparameter 100 hidden neuron, jenis fungsi aktivasi tanh, 100 epoch, learning rate sebesar 0,005, dan teknik pengisian data kosong menggunakan nilai rata – rata menghasilkan prediksi cuaca pada data tes yang mendekati data asli dengan RMSE 0,0274 dan evaluasi parameter yang baik.

Dalam konteks ini, penelitian ini akan membandingkan GLM dan Linear Regression untuk mengetahui algoritma mana yang lebih efektif dalam memprediksi curah hujan. Penelitian fitur cuaca seperti suhu, kelembaban, kecepatan angin, tutupan awan, tekanan dan hujan. Diharapkan hasil penelitian ini dapat memberikan kontribusi signifikan dalam pengembangan sistem prediksi hujan yang lebih akurat dan efisien.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini antara lain sebagai berikut:

1. Bagaimana implementasi perbandingan kinerja dari algoritma *Generalized Linear Model* dengan algoritma *Linear Regression* dalam memprediksi hujan?
2. Algoritma mana yang memiliki akurasi lebih baik dalam prediksi hujan, *Generalized Linear Model* atau *Linear Regression*?

1.3 Tujuan Penelitian

Berikut Tujuan dari penelitian ini antara lain :

1. Membandingkan kinerja dari algoritma *Generalized Linear Model* dan *Linear Regression* dalam memprediksi hujan.
2. Menentukan algoritma yang memiliki akurasi lebih tinggi dan lebih optimal untuk prediksi hujan.

1.4 Manfaat

Manfaat penelitian ini antara lain sebagai berikut:

1. Memberikan wawasan baru tentang penggunaan algoritma GLM dan *Linear Regression* dalam prediksi cuaca.
2. Membantu pengambil keputusan, seperti instansi terkait, dalam memilih metode prediksi cuaca yang lebih akurat.

Mendukung pengembangan teknologi prediksi cuaca berbasis data di Indonesia.