

BAB I

PENDAHULUAN

1.1 Latar Belakang

Negara Indonesia yang terletak di kawasan khatulistiwa menerima radiasi matahari sepanjang tahun. Letak geografi ini membentuk sistem iklim tropis yang secara signifikan memengaruhi dinamika cuaca di seluruh wilayahnya. Iklim tersebut ditandai oleh dua musim utama, yakni musim kemarau yang umumnya terjadi pada bulan April hingga Oktober, dan musim hujan pada November hingga Maret. Meskipun berpola musiman, intensitas curah hujan di Indonesia sangat bervariasi secara spasial, sehingga setiap wilayah memiliki karakteristik hidrometeorologi yang spesifik (Kasih et al., 2024). Indonesia dengan topografinya yang juga beragam menyebabkan intensitas curah hujan yang diterima dari masing-masing wilayahnya berbeda karena menyesuaikan dengan karakteristik topografi lokal (Ariska et al., 2024).

Dalam studi meteorologi atau ilmu cuaca, stasiun pengamatan cuaca permukaan seperti *Automatic rain gauges* (ARG) atau *Automatic Weather Station* (AWS) diakui secara universal sebagai standar kebenaran utama (*ground truth*) karena mengukur parameter cuaca curah hujan secara langsung di lokasi fisik (Drożdżioł & Absalon, 2023). Namun demikian, tantangan umum yang dihadapi para peneliti secara global maupun di Indonesia adalah fenomena kesenjangan jaringan stasiun (*station gaps*) (Tiara Putri et al., 2025). Distribusi instrumen fisik ini sangat tidak merata secara spasial; mayoritas terpusat di area perkotaan, dataran

rendah, dan wilayah daerah aliran sungai (DAS) namun sangat langka di kawasan topografi seperti pegunungan, hutan, pesisir, dan daerah terpencil. Selain itu, stasiun cuaca sering kali mengalami keterbatasan teknis akibat kegagalan alat, biaya pemeliharaan yang tinggi, dan pergantian instrumentasi, yang berdampak pada data yang terputus-putus atau terlalu pendek untuk analisis historis iklim (Kidd et al., 2017; Petrić et al., 2026).

Komunitas saintifik dan lembaga meteorologi internasional mengembangkan data reanalisis atmosfer yang mengombinasikan model komputasi numerik dengan asimilasi data historis yang masif untuk memecahkan masalah kekosongan data observasi (*observation gaps*) tersebut. Reanalisis adalah metode ilmiah yang dengan metode matematis memadukan (*assimilate*) jutaan observasi cuaca dari masa lalu dari stasiun pengamatan cuaca darat, kapal, buoy, pesawat, hingga satelit dengan model prakiraan cuaca numerik (*Numerical Weather Prediction*) modern. Hasilnya adalah representasi kondisi atmosfer historis yang sifatnya menyeluruh, konsisten secara fisik di seluruh grid lintang dan bujur bumi. Saat ini, data reanalisis paling mutakhir adalah ECMWF *Reanalysis* v5 (ERA5) yang dikembangkan oleh *European Centre for Medium-Range Weather Forecasts* (ECMWF). ERA5 merevolusi pemodelan iklim dengan menyediakan data cuaca resolusi temporal per jam (*hourly*) dengan resolusi spasial sekitar 31 kilometer sejak tahun 1940 hingga saat ini (Abdillah et al., 2025). Keandalan fisis ERA5 menjadikannya tulang punggung dalam penelitian fisika atmosfer (Hersbach et al., 2020). Di samping reanalisis, teknologi penginderaan jauh berupa pemindaian dengan satelit cuaca (seperti Himawari-9, GOES, Meteosat) dan data turunannya, umumnya disebut

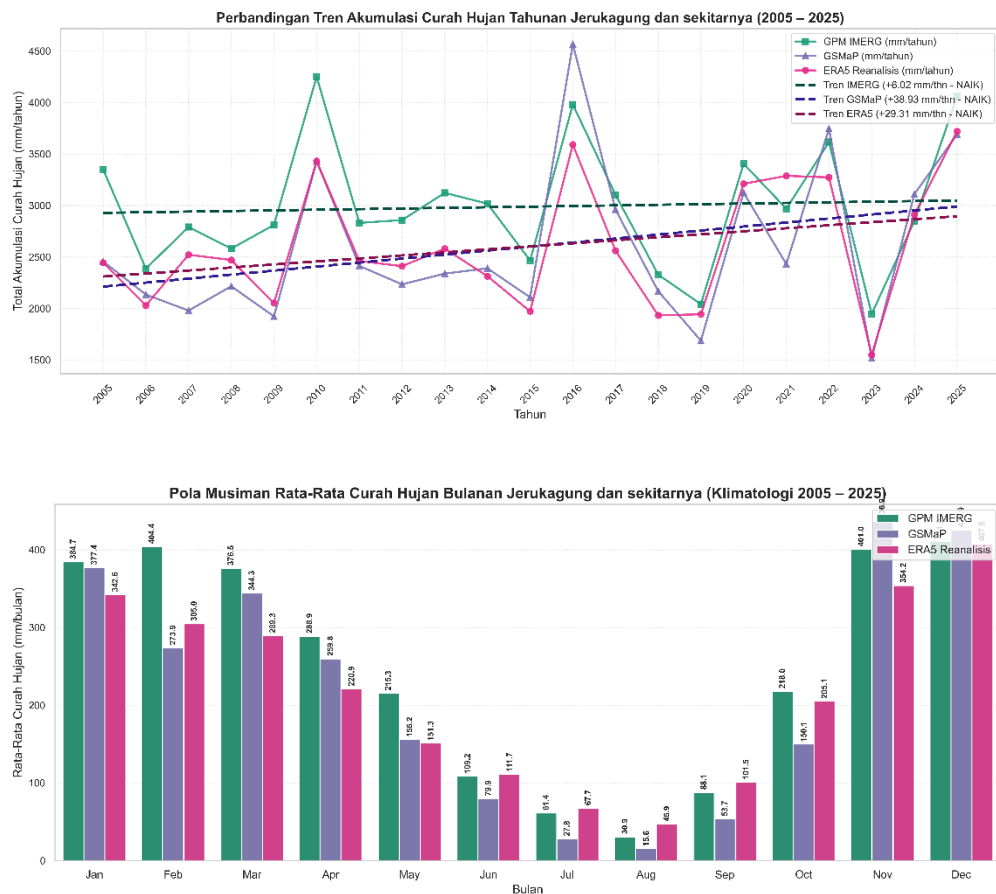
Satellite-based Precipitation Products (SPPs) data ini meliputi CHIRPS, GSMaP, IMERG, serta CMORPH yang digunakan untuk mengestimasi sebaran curah hujan dalam skala luas (Nguyen et al., 2025).

Berbeda dengan stasiun fisik, satelit tidak menakar air hujan, melainkan menggunakan sensor inframerah (*Infrared*) untuk mengukur suhu puncak awan dan gelombang mikro (*microwave*) dengan mendeteksi hamburan partikel es di dalam awan (Aulia Zikri et al., 2026). Keunggulan satelit terletak pada cakupan spasialnya yang luas dan resolusi yang sering kali lebih tinggi daripada data reanalisis. Namun demikian, estimasi satelit tetap memiliki kelemahan, sensor satelit kerap kesulitan membedakan antara awan tebal yang tidak menurunkan hujan dengan awan pembawa hujan, serta mengalami degradasi akurasi yang signifikan ketika berada di atas topografi pegunungan atau DAS (Hafyani et al., 2026; Ramadhan et al., 2023). Karena hal tersebut, satelit sering kali menunjukkan bias terlalu tinggi (*overestimation*) atau malah gagal mendeteksi hujan lebat (*underestimation*).

Berdasarkan observasi dan pengumpulan data cuaca, curah hujan merupakan parameter iklim yang sangat penting dalam menentukan dinamika perubahan iklim, memiliki dampak yang signifikan terhadap ketersediaan sumber daya air (Ariska et al., 2024), curah hujan merupakan komponen esensial dalam analisis cuaca, mitigasi bencana hidrometeorologi, menjaga stabilitas ekonomi dan ketahanan pangan (Suradhaniwar et al., 2021). Dalam sektor sumber daya air, prediksi curah hujan digunakan untuk pengelolaan sumber daya air, pencegahan banjir, sistem peringatan banjir, pemodelan aliran permukaan, kalibrasi pengukuran sebaran curah hujan terdistribusi, dan sistem peringatan dini aliran limpasan hujan

(Segovia-Cardozo et al., 2021). Dalam bidang pertanian, estimasi dan probabilitas curah hujan dijadikan sebagai salah satu pertimbangan dalam manajemen irigasi (Jamal et al., 2023), mencegah gagal panen dan merencanakan tanaman yang akan dibudidayakan (Apriyana et al., 2021). Namun demikian untuk memprediksi curah hujan yang akan datang merupakan masalah yang kompleks, karena dipengaruhi oleh banyak sifat dinamis variabel-variabel dalam cuaca yang diantaranya radiasi matahari, tekanan atmosfer, suhu, kelembapan, dan kecepatan angin, arah angin, lama penyinaran matahari (Safia et al., 2023).

Mendasari studi dari Hanafi et al., 2026 bahwa di Jawa Tengah menunjukkan adanya peningkatan variabilitas curah hujan sebesar 17% dan pergeseran musim yang signifikan. Hal ini menegaskan perlunya metode prediksi cuaca lokal yang lebih akurat untuk mitigasi bencana dan perencanaan dalam pertanian. Indratmoko et al., 2017, sebagaimana dikutip dalam Anandita Ayu Rochani et al., 2025 menyatakan bahwa Pulau Jawa, khususnya daerah pesisir Selatan Kebumen Jawa Tengah yang berbatasan langsung dengan samudera Hindia memiliki karakteristik curah hujan tahunan 4.548 mm/tahun (BPS, 2023; Sekaranom et al., 2021).



Gambar I-1. Tren dan Pola Musiman di Wilayah Jerukagung

Secara fisis dan geografis, tren curah hujan di wilayah Desa Jerukagung, Kecamatan Klirong, Kabupaten Kebumen dicirikan oleh tipe Sub-Monsunal karena variasi musimnya dipengaruhi angin monsun musiman, namun mendapat intervensi kuat dari dinamika lokal seperti massa udara basah dari Samudra Hindia, angin darat-laut pesisir, atau efek orografis pegunungan di dekatnya (Putra et al., 2024). Berdasarkan data historis 2005 hingga 2025, wilayah ini mendapatkan intensitas curah hujan yang masif pada periode musim hujan, di mana puncak curah hujan bulanan terkonsentrasi sangat tinggi pada rentang bulan November hingga Januari dengan total curah hujan melampaui angka sekitar 400 mm hingga mendekati 650

mm per bulan. Sebaliknya, wilayah ini juga dihadapkan pada periode kering yang ekstrem pada pertengahan tahun, ditandai dengan penurunan drastis volume curah hujan hingga berada di bawah ambang batas 50 mm pada bulan Juli hingga Agustus. Selain ketimpangan musiman yang tajam, grafik tersebut menangkap adanya dinamika variabilitas kuantitatif antar-dekade yang tidak menentu, terutama pada masa transisi (musim pancaroba). Karakteristik data yang memiliki pola sebaran timpang, lonjakan ekstrem mendadak, serta tingkat non-linearitas yang tinggi di kawasan ini menjadi justifikasi ilmiah utama mengapa model prakiraan konvensional linear cenderung masih kurang akurat.

Penggunaan metode observasi curah hujan dan probabilistik telah berevolusi dengan metode berbasis data historis yang panjang dan kecerdasan buatan untuk mengefisienkan proses prediksi dan mengurangi kesalahan prediksi (Sham et al., 2025). Ada beberapa langkah evaluasi untuk membandingkan kinerja model dalam memperkirakan curah hujan, selain metrik statistik yang umum pada machine learning seperti *Precision*, *Recall* dan *F1-Score* untuk model klasifikasi, selain itu juga dapat digunakan, *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE) untuk mengetahui besaran bias dari model jika menggunakan model regresi (Safia et al., 2023).

Berdasarkan permasalahan tersebut penulis berinisiatif untuk melakukan penelitian ini dengan mengembangkan model yang dirancang untuk memprediksi estimasi curah hujan menggunakan bantuan algoritma machine learning, *Extreme Gradient Boosting* (XGBoost) dan *deep learning*, *Long-Short Term Memory* (LSTM). Penggunaan algoritma XGBoost dan LSTM dipilih karena data dalam

penelitian ini masih berbentuk lokasi tunggal dalam format tabular deret waktu (*single-site point-scale multivariate time-series data*), kedua algoritma dapat digunakan untuk mendeteksi pola linear dan non-linear sepanjang data cuaca. Penelitian ini juga bertujuan mengidentifikasi algoritma machine learning maupun *deep learning* yang lebih sesuai untuk prediksi curah hujan.

Dalam upaya ini, penelitian mengenai prediksi curah hujan menjadi semakin penting, terutama di wilayah geografis dengan lahan pertanian atau sawah yang bergantung dengan ketersediaan air dari curah hujan seperti di Desa Jerukagung yang relatif dekat dengan pesisir dengan sistem cuacanya yang labil ditambah lebih rentan terhadap dinamisme perubahan iklim. Fenomena cuaca skala lokal yang sangat fluktuatif dan dinamis juga sangat sulit diprediksi secara akurat jika hanya mengandalkan permodelan atmosfer global berskala makro. Kesenjangan spasial pada data ERA5 dan data satelit menjadi kendala utama dalam melihat fenomena curah hujan lokal di Desa Jerukagung. Di sisi lain, penggunaan data in-situ AWS IoT terkendala oleh terbatasnya durasi data historis. Untuk menjembatani keterbatasan tersebut, diterapkan pendekatan melalui skema *Two-Stage Modeling*. Melalui tahap *Pre-Training*, model mempelajari pola atmosfer makro dari data model global, yang selanjutnya disesuaikan (*Fine-Tuning*) menggunakan data observasi AWS IoT agar presisi terhadap dinamika iklim lokal setempat.

1.2 Rumusan Penelitian

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana tingkat akurasi dan korelasi data satelit IMERG dan GSMAP terhadap data observasi AWS IoT di wilayah Desa Jerukagung berdasarkan nilai MAE, RMSE, dan Korelasi Pearson?
2. Bagaimana perbandingan performa model *Extreme Gradient Boosting* dan *Long Short-Term Memory* dalam memprediksi estimasi curah hujan jangka pendek dengan menggunakan metrik-metrik statistik?

1.3 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan untuk mengetahui korelasi antar variabel dan membandingkan perbandingan kedua model prediksi atau dalam beberapa poin di jelaskan di bawah ini

1. Mengetahui tingkat akurasi dan korelasi dari data satelit IMERG dan GSMAP terhadap data pengamatan sensor IoT di lokasi penelitian dengan nilai RMSE, MAE, dan Korelasi Pearson
2. Membandingkan performa model prediksi XGBoost dan LSTM untuk estimasi curah hujan jangka pendek dengan menggunakan metrik-metrik statistik.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi penelitian dan pengembangan yang dirinci sebagai berikut:

1. Manfaat Teoritis

- a. Memberikan kontribusi pada pengembangan literatur Sains Data dan pengembangan model prediksi curah hujan, khususnya dalam implementasi teknik *data fusion* (penggabungan data) antara data pengamatan berbasis pengamatan darat IoT dengan data hasil pemindaian citra satelit.
- b. Menjadi referensi ilmiah dalam perancangan algoritma *Machine Learning* (XGBoost) dan *Deep Learning* (LSTM) untuk pemodelan cuaca di wilayah tropis dan lokal.

2. Manfaat Praktis

- a. Bagi Masyarakat: Memberikan informasi estimasi curah hujan yang lebih akurat untuk membantu menentukan jadwal tanam, menentukan kebutuhan air dan meminimalkan risiko gagal panen akibat cuaca ekstrem.
- b. Bagi Pemerintah Daerah/Pemangku Kepentingan (*stakeholder*): Menyediakan basis data historis yang telah dikoreksi secara statistik sebagai landasan kebijakan mitigasi bencana banjir atau kekeringan di wilayah Jerukagung.

- c. Bagi Peneliti Selanjutnya: Memberikan kerangka kerja (*framework*) dan referensi penelitian dan pengembangan untuk pemrograman pemodelan dan analisis data cuaca yang dapat dikembangkan lebih lanjut untuk cakupan wilayah yang lebih luas atau prediksi variabel cuaca lainnya.

1.5 Batasan Penelitian

Pada penelitian ini ditetapkan batasan penelitian agar penelitian lebih fokus dan terukur, beberapa diantaranya:

1. Lingkup geografis sebagai obyek penelitian berada di wilayah Desa Jerukagung, Klirong, Kebumen.
2. Data observasi permukaan untuk validasi dan koreksi bias data berasal dari pengamatan in-situ mandiri milik penulis dengan variabel suhu, kelembapan udara, titik embun, tekanan udara, dan curah hujan dalam periode per menit dari tanggal 1 Januari 2025 hingga tanggal 31 Mei 2026.
3. Data sekunder yang digunakan dalam penelitian ini menggunakan data ERA5 dan data IMERG/GSMAP wilayah grid pengamatan Jerukagung, Klirong, Kebumen pada periode data dari tanggal 1 Januari 2005 hingga 31 Desember 2025 dalam interval waktu jam.
4. Penelitian ini dalam analisis dan pemodelannya menggunakan bahasa pemrograman Python dengan *framework* Tensorflow dan Keras untuk LSTM dan XGB *frameweork* untuk XGBoost.

5. Penelitian ini menggunakan metode *Two Stage Modeling* dimana jumlah kejadian curah hujan dan estimasinya memiliki model terpisah.
6. Penelitian ini berfokus pada perbandingan algoritma *Extreme Gradient Boosting* dan *Long Short Term Memory*.
7. Penelitian ini berfokus pada pengembangan, belum hingga ke tahap operasional prediksi
8. Penelitian ini hanya berfokus pada analisis algoritma *machine learning* dan *deep learning*, tidak mencakup bagian perancangan perangkat AWS IoT.

