

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Cuaca dan iklim merupakan elemen fundamental dari sistem tata bumi yang memiliki intervensi krusial terhadap hampir seluruh sektor kehidupan manusia. Dalam konteks tata ruang kota, dinamika cuaca khususnya curah hujan memegang peranan yang sangat vital, terutama bagi wilayah perkotaan padat penduduk yang bertindak sebagai pusat perputaran ekonomi dan pemerintahan seperti Daerah Khusus Ibukota (DKI) Jakarta. Secara geografis dan hidrologis, letak DKI Jakarta sangat rentan terhadap bencana hidrometeorologi. Kota metropolitan ini merupakan kawasan dataran rendah dengan sebagian wilayah pesisir utaranya berada di bawah permukaan laut, serta dilintasi oleh tiga belas aliran sungai besar yang bermuara ke Teluk Jakarta [2]. Kombinasi antara topografi yang cekung, masifnya alih fungsi lahan hijau menjadi kawasan beton komersial, dan sistem drainase perkotaan yang kerap kelebihan beban, menjadikan curah hujan dengan intensitas tinggi sebagai ancaman langsung yang berujung pada bencana banjir tahunan. Oleh karena itu, ketersediaan informasi mengenai prediksi cuaca dan presipitasi curah hujan yang cepat, presisi, dan dapat diandalkan menjadi prasyarat mutlak dalam merancang sistem mitigasi bencana perkotaan yang efektif.

Dalam beberapa dekade terakhir, fenomena perubahan iklim global telah memicu anomali cuaca yang sangat ekstrem. Dinamika atmosfer tidak lagi mengikuti siklus musim yang statis. Anomali iklim global seperti El Niño dan La Niña terbukti telah mendisrupsi pola curah hujan secara signifikan di berbagai wilayah Indonesia, termasuk memicu intensitas hujan di luar batas normal [6]. Di Jakarta, tingginya curah hujan ini secara persisten menjadi katalisator utama meluapnya kapasitas sungai dan memicu genangan berskala luas yang melumpuhkan urat nadi aktivitas harian masyarakat. Bencana banjir di Jakarta tidak sekadar berdampak pada kerusakan infrastruktur fisik, melainkan meluas pada kerugian sosial-ekonomi yang masif, terhentinya jalur logistik, hingga ancaman terhadap kesehatan dan keselamatan jiwa warga. Sebagai representasi dari fenomena tersebut, laporan aktual kebencanaan pada awal tahun 2026 mencatat bahwa hujan ekstrem yang turun tanpa henti telah memicu banjir parah yang merendam ratusan Rukun Tetangga (RT) di berbagai sudut ibu kota. Bencana ini

memaksa ribuan jiwa untuk mengevakuasi diri ke tempat pengungsian dan memicu kelumpuhan lalu lintas yang parah, dengan total estimasi kerugian ekonomi yang ditaksir mencapai angka triliunan rupiah (sumber: Ribuan Warga Harus Ngungsi, Jakarta Rugi Triliunan Efek Dihajar Banjir - CNBC Indonesia, 2026). Rekam jejak kerugian yang terus berulang dan bereskalasi ini menegaskan bahwa banjir di Jakarta adalah krisis hidrometeorologis yang membutuhkan penanganan preventif tingkat tinggi, bukan sekadar respons reaktif pasca-bencana.

Akar dari keterlambatan respons mitigasi sering kali bersumber pada kelemahan sistem peramalan cuaca yang ada. Secara tradisional, peramalan cuaca dilakukan menggunakan metode statistik konvensional atau model numerik berbasis fisika atmosfer yang membutuhkan daya komputasi superkomputer yang sangat besar dan waktu pemrosesan yang lama. Pendekatan konvensional ini sering kali dihadapkan pada tantangan akurasi ketika harus memodelkan data meteorologi yang memiliki sifat sangat dinamis, non-linear, dan penuh dengan fluktuasi *noise* [3]. Lebih jauh, studi lintas iklim yang dilakukan oleh Ejike dkk. (2025) secara empiris membuktikan bahwa kompleksitas variabilitas atmosfer di kawasan tropis seperti DKI Jakarta secara inheren jauh lebih tinggi dibandingkan kawasan beriklim sedang, sehingga model prediksi curah hujan konvensional yang dirancang untuk iklim temperate tidak dapat langsung ditransfer ke konteks tropis tanpa adaptasi metodologis yang mendasar [17]. Kurangnya sistem prediksi lokal yang mampu memberikan tingkat akurasi tinggi secara *real-time* membuat masyarakat dan pemangku kebijakan (seperti Badan Penanggulangan Bencana Daerah) sering kali tidak memiliki waktu respons (*lead time*) yang cukup untuk melakukan evakuasi dini atau rekayasa hidrologis seperti pengaktifan pompa air. Oleh karena itu, modernisasi menuju sebuah *Early Warning System* (Sistem Peringatan Dini/EWS) yang cerdas, proaktif, dan adaptif terhadap pola cuaca historis menjadi sebuah urgensi riset yang kritis dan tidak dapat ditunda [1].

Menjawab kompleksitas dinamika atmosfer tersebut, pemanfaatan ilmu Kecerdasan Buatan (*Artificial Intelligence/AI*), secara khusus pada cabang *Machine Learning* (ML), telah mendapatkan atensi akademik yang sangat luas dalam domain klimatologi komputasi. Algoritma *Machine Learning* terbukti sangat mumpuni dalam mengekstraksi informasi dan mengenali pola matematis tersembunyi dari sekumpulan data meteorologi multivariat seperti suhu, kelembapan, kecepatan angin, dan tekanan udara tanpa memerlukan asumsi linearitas yang kaku [3]. Validitas global dari pendekatan ini telah dikonfirmasi oleh berbagai riset internasional. Appiah-Badu dkk. (2022) membuktikan superioritas algoritma

berbasis pohon seperti *Random Forest* (RF) dan *Extreme Gradient Boosting* (XGBoost) dalam mengklasifikasikan curah hujan harian baik hujan maupun tidak hujan di berbagai zona ekologi Ghana menggunakan data iklim harian dari stasiun meteorologi resmi, dengan ROC-AUC yang konsisten melampaui model konvensional [10]. Sejalan dengan hal tersebut, penelitian terkini oleh Wu dkk. (2024) menegaskan keunggulan komparatif XGBoost dan RF dibandingkan regresi logistik dalam memodelkan sensitivitas wilayah terhadap banjir, sekaligus membuktikan bahwa analisis *feature importance* dari kedua algoritma mampu mengungkap variabel meteorologi yang paling deterministik dalam pemicu kejadian banjir [11]. Sejumlah studi referensial di dalam negeri pun mengonfirmasi tren tersebut. Pringandana dan Kusnawi (2025) membuktikan bahwa penerapan XGBoost yang dioptimasi mampu menghasilkan klasifikasi curah hujan *multiclass* dengan tingkat akurasi yang sangat tinggi, menjadikannya tulang punggung yang ideal bagi sistem peringatan dini [1]. Sejalan dengan hal tersebut, Achmad dkk. (2025) berhasil mendemonstrasikan bahwa algoritma RF memiliki ketangguhan luar biasa dalam memetakan tingkat keparahan banjir di DKI Jakarta berbasis data historis kebencanaan [2].

Di luar konteks Jakarta, eksistensi algoritma berbasis pohon (*tree-based ensemble learning*) seperti RF dan XGBoost juga secara konsisten menunjukkan performa superior di berbagai topologi wilayah. Penelitian Imanullah dan Fanani (2026) di area Gunung Ungaran menemukan bahwa XGBoost merupakan model paling reliabel untuk memprediksi hujan harian [4]. Di Sumatera Selatan, komparasi RF dan XGBoost untuk mitigasi Kebakaran Hutan dan Lahan (Karhutla) menunjukkan bahwa model *ensemble* dari keduanya memberikan tingkat galat terkecil [3]. Secara internasional, Ren dkk. (2024) menegaskan keunggulan metode *random sampling* dalam melatih RF dan XGBoost untuk penilaian kerentanan banjir, dengan XGBoost mencapai ROC-AUC tertinggi dalam skenario data yang tidak seimbang [14]. Dalam konteks kota-kota tropis yang secara klimatologis paling dekat dengan profil Jakarta, Al-Aizari dkk. (2024) berhasil mendemonstrasikan bahwa kombinasi RF dan XGBoost secara efektif mereduksi ketidakpastian dalam pemetaan kerentanan banjir di lingkungan perkotaan, dengan analisis fitur yang mengonfirmasi dominasi variabel curah hujan historis [15]. Lebih lanjut, Mdegela dkk. (2023) secara khusus mengembangkan kerangka klasifikasi kejadian hujan ekstrem (*extreme rainfall event classification*) menggunakan beragam algoritma ML sebagai fondasi teknis EWS banjir sungai, dengan temuan bahwa ketidakseimbangan kelas (*class imbalance*) antara hari hujan dan tidak

hujan adalah tantangan metodologis utama yang wajib ditangani secara eksplisit [12]. Sementara itu, meskipun pendekatan *Deep Learning* seperti *Long Short-Term Memory* (LSTM) di Bandung [5] dan model *Fuzzy Time Series* [6] juga mampu memprediksi curah hujan, algoritma *ensemble* berbasis pohon (RF dan XGBoost) tetap menjadi primadona karena keseimbangannya antara efisiensi komputasi dan akurasi tinggi. Hal ini didukung oleh temuan Imani dkk. (2025) yang secara komprehensif membuktikan bahwa RF dan XGBoost yang dikombinasikan dengan teknik *Synthetic Minority Oversampling Technique* (SMOTE) secara konsisten menghasilkan performa terbaik dalam menangani ketidakseimbangan kelas yang sangat umum terjadi pada data cuaca di mana jumlah hari “tidak hujan” sering kali jauh lebih banyak daripada hari “hujan” [13]. Konsistensi superioritas algoritma berbasis pohon ini juga dikonfirmasi oleh analisis performa model *ensemble* untuk prakiraan banjir di kawasan tropis, yang secara empiris membuktikan bahwa arsitektur pohon keputusan bertingkat tetap memberikan akurasi terdepan bahkan pada distribusi data hidrometeorologi yang sangat skewed [18].

Meskipun adopsi *Machine Learning* dalam prediksi cuaca telah dilakukan secara masif, analisis mendalam terhadap literatur yang ada mengungkap sebuah celah penelitian (*research gap*) yang krusial. Mayoritas studi klimatologi komputasi saat ini masih terperangkap pada paradigma “kotak hitam” (*black-box model*). Penelitian terdahulu cenderung memusatkan seluruh fokus pada pencapaian probabilitas metrik evaluasi tertinggi seperti *Accuracy*, *Precision*, dan *Recall* namun sering kali abai dalam menjelaskan aspek interpretabilitas atau kemampuan model untuk dipahami oleh manusia [13], [14]. Belum banyak penelitian komparatif yang meletakkan akurasi dan interpretabilitas pada kedudukan yang setara. Padahal, dalam konteks mitigasi bencana publik, para meteorolog dan perancang kebijakan sangat membutuhkan transparansi model (sifat *explainable*). Mereka perlu mengetahui secara spesifik fitur cuaca mana yang paling deterministik (analisis *feature importance*); misalnya, seberapa besar anjloknya suhu udara atau meroketnya persentase kelembapan relatif bertindak sebagai pemicu utama model dalam meyakini akan turunnya hujan lebat [3]. Selain itu, sebagian besar penelitian terdahulu masih mengevaluasi model menggunakan metode pemisahan data acak bersilang (*random cross-validation*), yang berisiko menyebabkan kebocoran data (*data leakage*) masa depan ke dalam proses pelatihan masa lalu. Komparasi kinerja model yang secara ketat menggunakan skenario pemisahan data berdasarkan kronologi urutan waktu (*time series splitting*) masih sangat minim, padahal metode inilah yang paling merepresentasikan kondisi riil implementasi EWS di lapangan

[18].

Untuk menjembatani celah penelitian tersebut, penelitian ini mengajukan kerangka solusi terstruktur dengan melakukan analisis komparatif yang komprehensif terhadap dua algoritma klasifikasi *ensemble* terkuat saat ini: *Random Forest* (yang menggunakan pendekatan *Bagging*) dan *XGBoost* (yang menggunakan pendekatan *Boosting*). Pemilihan kedua algoritma ini didasari oleh arsitekturnya yang tangguh dalam menangani data *outlier*, kemampuannya memproses relasi antar-variabel cuaca yang kompleks, ketersediaan modul bawaan untuk mengekstraksi nilai *feature importance* guna menjawab tantangan interpretabilitas, serta rekam jejaknya yang unggul secara konsisten dalam berbagai konteks perkotaan dan iklim tropis sebagaimana ditunjukkan oleh Alaizari dkk. (2024) dan Wu dkk. (2024) [15], [11], [1], [4]. Eksperimen akan dilakukan menggunakan kumpulan data historis iklim di wilayah DKI Jakarta, dengan menerapkan metode validasi *time series splitting* untuk menjamin keabsahan evaluasi model. Melalui penelitian ini, diharapkan dapat dilahirkan model klasifikasi kejadian hujan yang tidak hanya superior secara angka (akurat), tetapi juga transparan secara logika (dapat diinterpretasikan). Pada muaranya, luaran penelitian ini dirancang sebagai fondasi teknis yang konkret untuk mendukung pemangku kepentingan dalam mengorkestrasi EWS dan manajemen mitigasi banjir DKI Jakarta yang lebih antisipatif, tangkas, dan presisi. Berdasarkan karakteristik fundamental yang membedakan kedua paradigma tersebut *Bagging* yang mengutamakan stabilitas dan *Boosting* yang mengutamakan koreksi kesalahan bertahap penelitian ini menduga akan ditemukan pola trade-off performa yang konsisten dengan literatur sebelumnya, sekaligus membuka ruang analisis interpretabilitas yang belum banyak dieksplorasi pada konteks DKI Jakarta.

1.2 Rumusan Masalah

Berdasarkan latar belakang, fenomena lapangan, studi literatur terdahulu, serta kebutuhan atas model data cuaca historis yang komprehensif, maka rumusan masalah dalam penelitian ini didefinisikan ke dalam poin-poin berikut:

- a. Bagaimana analisis komparatif akurasi dan interpretabilitas algoritma *Random Forest* dan *XGBoost* dalam klasifikasi kejadian hujan di DKI Jakarta?
- b. Berapa tingkat akurasi dan metrik performa model *Random Forest* dan

XGBoost dalam melakukan klasifikasi kejadian hujan apabila dievaluasi menggunakan metode *time series splitting*?

1.2.1 Hipotesis Penelitian

Berdasarkan kajian latar belakang dan tinjauan penelitian terdahulu yang menunjukkan karakteristik berbeda antara paradigma Bagging dan Boosting, penelitian ini mengajukan hipotesis sebagai berikut:

- a. Terdapat perbedaan performa yang signifikan antara algoritma Random Forest dan XGBoost dalam klasifikasi kejadian hujan di DKI Jakarta, di mana XGBoost diduga unggul pada metrik Recall karena karakteristik Boosting yang berfokus mengoreksi kesalahan klasifikasi kelas minoritas, sedangkan Random Forest diduga unggul pada Accuracy dan Precision karena karakteristik Bagging yang menghasilkan prediksi lebih stabil dan konservatif.
- b. Analisis interpretabilitas menggunakan Mean Decrease in Impurity (MDI) dan SHapley Additive exPlanations (SHAP) diduga akan menunjukkan konsistensi dalam mengidentifikasi variabel meteorologis yang paling berpengaruh terhadap kejadian hujan, sejalan dengan prinsip-prinsip klimatologi yang telah mapan.

1.3 Tujuan Penelitian

Sejalan dengan pertanyaan riset pada rumusan masalah di atas, penelitian ini secara spesifik bertujuan untuk:

- a. Menganalisis serta membandingkan secara mendalam akurasi prediktif dan kemampuan interpretabilitas (*feature importance*) antara algoritma *Random Forest* dan XGBoost dalam kasus klasifikasi kejadian hujan di wilayah DKI Jakarta.
- b. Mengukur secara kuantitatif tingkat performa (*metrics evaluation*) dari kedua model *machine learning* tersebut dalam memetakan pola hujan dengan mengadopsi pendekatan uji evaluasi runtut waktu (*time series splitting*).

1.4 Batasan Masalah

Untuk menjaga arah, fokus, dan kedalaman pembahasan agar tidak menyimpang dari tujuan awal, maka ruang lingkup penelitian ini dibatasi pada poin-poin berikut:

- a. Data observasi yang digunakan bersumber secara resmi dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), yang secara spesifik difokuskan pada cakupan tiga titik pengamatan di wilayah geografis DKI Jakarta, yaitu:
 1. Jakarta Pusat: Stasiun Meteorologi Kemayoran
 2. Jakarta Timur: Stasiun Meteorologi Halim Perdanakusuma
 3. Jakarta Utara: Stasiun Meteorologi Maritim Tanjung Priok
- b. Dataset historis cuaca (variabel iklim) yang diolah terbatas pada rentang waktu perekaman harian terhitung mulai tanggal 6 Maret 2024 hingga 4 Maret 2026.
- c. Algoritma *Machine Learning* yang diimplementasikan dan dikomparasikan performanya dalam penelitian ini hanya dibatasi pada *Random Forest* dan *XGBoost*.
- d. Referensi metodologis dan kaidah teori secara eksklusif berakar pada landasan pustaka dan jurnal yang telah ditetapkan.
- e. Metode evaluasi efikasi algoritma diukur menggunakan matriks kebingungan (*confusion matrix*) yang meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-score*.
- f. Penelitian ini bertitik berat pada fase klasifikasi kejadian hujan dengan luaran kelas diskrit/kategorikal (hujan atau tidak hujan), bukan berfokus pada prediksi estimasi kuantitatif debit air atau besaran curah hujan (regresi).
- g. Model yang dikembangkan dibatasi pada lingkungan simulasi eksperimental dan tidak membahas terkait tahapan implementasi sistem (*deployment*) secara *real-time* berbasis perangkat lunak produksi.

1.5 Manfaat Penelitian

Penelitian komparatif ini diharapkan mampu menghadirkan implikasi positif secara multidimensi, yang diklasifikasikan menjadi dua aspek:

a. Manfaat Akademis

1. Memberikan kontribusi perluasan literatur pada cabang keilmuan *Machine Learning*, khususnya terkait eksplorasi komparasi algoritma *ensemble learning*.
2. Mengembangkan diskursus pada area kajian klimatologi komputasi mengenai seberapa jauh parameter cuaca dapat dimodelkan serta diinterpretasikan melalui pendekatan kecerdasan buatan.

b. Manfaat Praktis

1. Menyediakan sumber literatur dan referensi kuantitatif bagi pihak pemangku kebijakan publik (Pemerintah Daerah DKI Jakarta) dan otoritas meteorologi seperti BMKG.
2. Mengakselerasi terwujudnya desain fondasi arsitektur Sistem Peringatan Dini (EWS) yang lebih cermat guna menekan dampak bencana serta mendukung penyempurnaan manajemen mitigasi banjir Jakarta.

1.6 Sistematika Penulisan

Sistematika penulisan memuat rancangan alur elaborasi skripsi agar pembahasan menjadi sistematis dan berurutan. Skripsi ini dibagi menjadi lima bab utama dengan susunan sebagai berikut:

a. BAB 1 PENDAHULUAN

Bab ini memuat uraian latar belakang masalah, fenomena aktual, identifikasi celah penelitian, rumusan masalah, ruang lingkup batasan masalah, tujuan dan manfaat dari penyusunan skripsi, serta diakhiri dengan sistematika penulisan.

b. BAB 2 LANDASAN TEORI

Bab ini memaparkan landasan kajian pustaka, penelitian terdahulu yang relevan, serta penjabaran teori mendasar mengenai klimatologi, curah hujan,

hingga tinjauan matematis mengenai *Machine Learning*, algoritma *Random Forest*, XGBoost, dan teknik evaluasi model.

c. **BAB 3 METODOLOGI PENELITIAN**

Bab ini menjelaskan *blueprint* dari tata cara penelitian, mendeskripsikan sumber dataset historis cuaca, arsitektur skenario eksperimen, tahapan prapemrosesan data (*preprocessing*), skema *time series splitting*, hingga mekanisme evaluasi performa algoritma.

d. **BAB 4 HASIL DAN DISKUSI**

Bab ini berisi representasi komprehensif mengenai temuan eksperimental. Membahas penjabaran hasil komparasi uji akurasi model dan analisis pembedahan tingkat interpretabilitas kedua algoritma dalam menyeleksi fitur utama penentu kejadian hujan di Jakarta.

e. **BAB 5 SIMPULAN DAN SARAN**

Bab terakhir yang merangkum konklusi inti penelitian untuk menjawab rumusan dan tujuan masalah, sekaligus menjabarkan saran aplikatif bagi pengembangan riset-riset prediktif sejenis di masa yang akan datang.

