

BAB I PENDAHULUAN

1.1 Latar Belakang

Kemunculnya *big data* pada awal abad ke-21 dan inovasi komputer super menggunakan *Graphics Processing Unit* (GPU) berperan penting terhadap perkembangan sejarah *machine learning*. Hal ini terbukti dari munculnya metode analisis data baru yang meningkatkan minat para ilmuwan terhadap penelitian [1]. Meskipun berbagai metode *machine learning* telah dikembangkan sejak tahun 1960-an, keterbatasan sumber daya komputasi dan data pada masa tersebut membatasi penerapannya. Perkembangan kemampuan komputasi, terutama melalui pemanfaatan *Graphics Processing Unit* (GPU), serta meningkatnya ketersediaan *big data* telah mendorong kemajuan signifikan dalam bidang *artificial intelligence* dan *machine learning* [2].

Prakiraan cuaca merupakan salah satu aspek penting yang mendukung berbagai sektor, seperti pertanian, transportasi, penerbangan, dan mitigasi bencana. Secara konvensional, prakiraan cuaca dilakukan menggunakan *Numerical Weather Prediction* (NWP), yaitu pendekatan berbasis model matematis yang telah dikembangkan sejak pertengahan abad ke-20 untuk memproyeksikan kondisi atmosfer [3]. Seiring meningkatnya kompleksitas dinamika atmosfer serta berkembangnya teknologi komputasi, pendekatan berbasis *machine learning* mulai banyak diterapkan sebagai pelengkap maupun alternatif terhadap NWP. Kemajuan kemampuan komputasi melalui *Graphics Processing Unit* (GPU), ketersediaan *big data*, dan perkembangan algoritma modern memungkinkan *machine learning* mempelajari pola kompleks dari data meteorologi dalam jumlah besar yang sulit diidentifikasi secara manual [4]. Meskipun demikian, penerapan *machine learning* dalam prakiraan cuaca masih menghadapi tantangan, terutama dalam menjaga stabilitas dan akurasi prediksi jangka panjang akibat akumulasi kesalahan selama proses prediksi [5].

Machine learning merupakan metode yang memungkinkan suatu sistem mempelajari pola dari data yang relevan dengan permasalahan tertentu, sehingga proses pembuatan model analisis dapat dilakukan secara otomatis dan digunakan untuk menyelesaikan tugas-tugas yang berhubungan [6]. Keputusan *machine*

learning dibuat berdasarkan data masukan menggunakan model algoritma tertentu. Data yang telah dimasukkan kemudian dibagi menjadi beberapa kumpulan data, yaitu data pelatihan, data evaluasi, dan data pengujian [7]. Prakiraan cuaca bertujuan untuk menemukan pola yang menjadikan dasar pengambilan keputusan dengan menggunakan data untuk diterapkan pada sebuah aplikasi dan model *machine learning*.

Pada penelitian ini, proses evaluasi dilakukan menggunakan model *machine learning extreme gradient boosting* (XGBoost) sebagai landasan dasar untuk memprakirakan cuaca, dengan menggunakan berbagai fitur seperti suhu, kelembaban, kecepatan angin dan fitur lainnya. XGBoost merupakan salah satu teknik *machine learning* yang bertujuan untuk meningkatkan model berbasis *tree boosting* (*decision tree*), yang merupakan metode pembelajaran mesin yang sangat efektif dan banyak digunakan [7], [8]. XGBoost merupakan algoritma *ensemble learning* yang dikenal memiliki fleksibilitas tinggi, akurasi prediksi yang kuat, kemampuan generalisasi yang baik, skalabilitas yang tinggi, efisiensi tinggi dalam pelatihan model, serta ketahanan yang kuat terhadap berbagai jenis data dan kondisi [9].

Hyperparameter tuning merupakan proses pencarian kombinasi *hyperparameter* yang optimal untuk menghasilkan performa terbaik pada model *machine learning*. Berbeda dengan parameter model yang dipelajari secara otomatis selama proses pelatihan, *hyperparameter* ditentukan sebelum pelatihan dan berperan dalam mengendalikan proses pembelajaran serta struktur model. Pemilihan *hyperparameter* yang tepat dapat meningkatkan akurasi, kemampuan generalisasi, dan efisiensi komputasi, sedangkan konfigurasi yang kurang sesuai berpotensi menurunkan performa model. Oleh karena itu, berbagai metode optimasi dikembangkan untuk menentukan kombinasi *hyperparameter* yang optimal secara lebih sistematis dan efisien dibandingkan penentuan secara manual, salah satunya melalui pendekatan *Grid Search* dan *Random Search* [10], [11].

Penelitian ini bertujuan menerapkan metode *extreme gradient boosting* ke dalam klasifikasi cuaca untuk memprakirakan cuaca berdasarkan dataset yang diuji. Dataset penelitian ini berasal dari situs web Kaggle.com yang menyediakan penjelasan mengenai berbagai parameter yang mempengaruhi kondisi cuaca di

suatu lokasi. Penelitian ini bertujuan memberikan panduan kepada pembaca dalam membangun model *machine learning* menggunakan metode *extreme gradient boosting* pada klasifikasi cuaca, untuk memberikan pemahaman umum tentang cara model memprakirakan cuaca.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana penerapan metode SMOTE (*Synthetic Minority Over-sampling Technique*) dalam mengatasi ketidakseimbangan data untuk meningkatkan kinerja model klasifikasi cuaca?
2. Bagaimana penerapan algoritma Extreme Gradient Boosting (XGBoost) dengan menggunakan teknik *hyperparameter tuning* dalam membangun model klasifikasi cuaca?
3. Bagaimana perbandingan kinerja model XGBoost tanpa *hyperparameter tuning* (model dasar) dengan model XGBoost menggunakan pendekatan *hyperparameter tuning* dalam klasifikasi cuaca?

1.3 Batasan Masalah

Agar penelitian lebih terfokus dan terarah, maka batasan masalah dalam penelitian ini ditentukan sebagai berikut:

1. Penelitian ini menggunakan dataset cuaca yang bersumber dari situs Kaggle dengan parameter – parameter cuaca yang tersedia dalam dataset tersebut.
2. Model yang dikembangkan menggunakan algoritma Extreme Gradient Boosting (XGBoost) sebagai metode utama tanpa melakukan perbandingan dengan algoritma lain.
3. Penelitian ini berfokus pada klasifikasi kondisi cuaca, bukan pada prediksi nilai numerik maupun *time series forecasting*.
4. Penanganan ketidakseimbangan data dilakukan menggunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*).
5. Optimasi model dilakukan menggunakan teknik *hyperparameter tuning* pada algoritma XGBoost.

6. Evaluasi model dilakukan menggunakan metode cross-validation dengan metrik akurasi, precision, recall, dan F1-score sebagai ukuran performa klasifikasi.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma Extreme Gradient Boosting (XGBoost) dalam membangun model klasifikasi kondisi cuaca berdasarkan dataset yang tersedia.
2. Menganalisis pengaruh penerapan hyperparameter tuning terhadap kinerja model XGBoost dalam klasifikasi kondisi cuaca.
3. Menganalisis pengaruh penerapan metode SMOTE (Synthetic Minority Over-sampling Technique) dalam mengatasi ketidakseimbangan data terhadap kinerja model klasifikasi.
4. Membandingkan kinerja model XGBoost tanpa hyperparameter tuning (model dasar) dengan model XGBoost yang telah dioptimasi dalam klasifikasi kondisi cuaca.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Secara Teoritis:

- a. Memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang *machine learning*, khususnya dalam penerapan XGBoost pada klasifikasi data cuaca.
- b. Menambah referensi dan literatur akademik terkait penggunaan algoritma XGBoost dalam konteks prakiraan cuaca.

2. Secara Praktis:

- a. Membantu lembaga atau institusi terkait dalam mengembangkan sistem prakiraan cuaca yang lebih akurat dan efisien menggunakan teknologi *machine learning*.
- b. Menyediakan panduan praktis dalam implementasi model XGBoost untuk klasifikasi data cuaca dengan parameter yang optimal.

1.6 Sistematika Penulisan

Berisi sistematika penulisan skripsi yang memuat uraian secara garis besar isi skripsi untuk tiap-tiap bab, Berisikan 5 Bab dengan rincian sebagai berikut:

BAB I PENDAHULUAN, Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA, Bab ini membahas teori-teori yang mendasari penelitian, seperti teori dasar tentang *machine learning*, algoritma XGBoost, serta studi terdahulu yang relevan dengan topik prakiraan cuaca.

BAB III METODE PENELITIAN, Bab ini menjelaskan metode yang digunakan dalam penelitian, meliputi tahapan pengumpulan data, praproses data, pemilihan fitur, pembangunan model XGBoost, dan evaluasi model.

BAB IV HASIL DAN PEMBAHASAN, Bab ini menyajikan hasil implementasi algoritma XGBoost pada dataset cuaca, serta analisis terhadap kinerja model berdasarkan metrik evaluasi yang telah ditentukan.

BAB V PENUTUP, Bab ini memuat kesimpulan dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.