

## **BAB I** **PENDAHULUAN**

### **1.1 Latar Belakang**

Perubahan iklim global telah menyebabkan ketidakstabilan cuaca yang signifikan di berbagai wilayah, termasuk Kota Semarang. Ketidakpastian cuaca ini berdampak pada berbagai sektor diantaranya pariwisata, transportasi bahkan sektor pertanian, yang sangat bergantung pada kondisi cuaca yang dapat diprediksi[1]. Peneliti dalam kegiatan penelitian ini menemukan bahwa ketidakmampuan untuk mengklasifikasi cuaca secara akurat dapat mengakibatkan kerugian ekonomi dan operasional yang besar bagi masyarakat dan industri setempat.

Dalam beberapa tahun terakhir, hujan lebat dan panas sering terjadi di Kota Semarang, yang sering kali tidak terprediksi dengan baik oleh sistem prediksi cuaca konvensional. Hal ini menimbulkan pertanyaan mengenai efektivitas metode klasifikasi cuaca yang saat ini digunakan. Peneliti mengamati bahwa metode tradisional sering kali tidak mampu menangkap kompleksitas pola cuaca yang dinamis dan berubah-ubah[2].

Seiring dengan perkembangan teknologi dan data science, muncul berbagai metode analisis data yang lebih canggih, seperti regresi linier, regresi logistik, decision tree, dan random forest. Metode-metode ini menawarkan potensi untuk meningkatkan akurasi prediksi cuaca dengan memanfaatkan data historis dan variabel cuaca yang lebih beragam. Peneliti melihat adanya peluang untuk mengaplikasikan metode ini dalam konteks klasifikasi cuaca di Kota Semarang, guna mengatasi permasalahan ketidakakuratan klasifikasi yang ada[3].

Penelitian ini bertujuan untuk membandingkan performa dari keempat metode analisis data tersebut dalam mengklasifikasi cuaca di Kota Semarang. Dengan demikian, diharapkan dapat ditemukan metode terbaik untuk digunakan dalam sistem prediksi klasifikasi cuaca di masa depan, sehingga dapat meminimalisir dampak negatif dari cuaca ekstrem sehingga perencanaan di berbagai sektor menjadi lebih efektif dan efisien.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah disusun untuk menjabarkan permasalahan yang menjadi fokus dalam penelitian ini. Adapun rumusan masalah yang diajukan dalam penelitian ini meliputi:

1. Bagaimana performa Regresi Linier dalam mengklasifikasikan cuaca di Kota Semarang?
2. Bagaimana performa Regresi Logistik dalam mengklasifikasikan cuaca di Kota Semarang?
3. Bagaimana performa Decision Tree dalam mengklasifikasikan cuaca di Kota Semarang?
4. Bagaimana performa Random Forest dalam mengklasifikasikan cuaca di Kota Semarang?
5. Model algoritma mana yang memiliki performa terbaik dalam klasifikasi cuaca di Kota Semarang?

## 1.3 Batasan Masalah

Agar penelitian ini dapat dilaksanakan secara efektif dan mencapai tujuan yang telah ditetapkan, diperlukan adanya batasan yang memperjelas ruang lingkup penelitian. Adapun batasan-batasan dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan hanya berasal dari dataset cuaca historis dari Kota Semarang yang tersedia dalam format CSV, dengan fitur-fitur terbatas seperti suhu, kelembaban, durasi sinar matahari, kecepatan angin, dan arah angin.
2. Prediksi yang dilakukan hanya bersifat klasifikasi biner, yaitu memprediksi apakah akan terjadi hujan (label 1) atau tidak hujan (label 0).
3. Model machine learning yang digunakan terbatas pada Linear Regression, Logistic Regression, Decision Tree, dan Random Forest.
4. Teknik balancing data yang diterapkan hanya menggunakan undersampling pada kelas mayoritas untuk mengatasi ketidakseimbangan data.

5. Evaluasi performa model hanya menggunakan metrik AUC-ROC, akurasi, presisi, recall, atau F1-score.
6. Prediksi hanya dilakukan pada data dengan fitur numerik, tanpa mempertimbangkan fitur kategori yang mungkin ada.
7. Prediksi terhadap data baru dilakukan dengan menggunakan model Machine Learning paling optimal yang telah diperoleh melalui proses pelatihan dan penyetelan parameter.

#### 1.4 Tujuan Penelitian

Penelitian ini bertujuan membandingkan performa empat algoritma, yaitu Regresi Linier, Regresi Logistik, Decision Tree, dan Random Forest, dalam mengklasifikasikan cuaca di Kota Semarang.

#### 1.5 Manfaat Penelitian

Manfaat penelitian yang ingin dicapai oleh penulis dalam penelitian ini sebagai berikut :

##### a. Manfaat secara teoritis

Penelitian ini diharapkan bisa memberikan kontribusi dalam pengembangan ilmu pengetahuan, terutama di bidang data science dan machine learning, khususnya dalam penerapan algoritma klasifikasi untuk data cuaca. Dengan membandingkan beberapa algoritma seperti Regresi Linier, Regresi Logistik, Decision Tree, dan Random Forest, hasil dari penelitian ini dapat menjadi bahan referensi untuk pengembangan model klasifikasi cuaca berbasis data historis. Selain itu, penelitian ini juga menambah wawasan terkait cara menangani data yang tidak seimbang serta bagaimana melakukan penyetelan model menggunakan hyperparameter tuning dalam konteks klasifikasi cuaca.

##### b. Manfaat secara praktis

Secara praktis, hasil dari penelitian ini bisa dimanfaatkan oleh instansi terkait seperti Badan Meteorologi, Klimatologi, dan Geofisika (BMKG)

maupun pemerintah daerah untuk membantu meningkatkan ketepatan prediksi cuaca, terutama di wilayah Kota Semarang. Model klasifikasi cuaca yang akurat dapat membantu masyarakat dalam merencanakan kegiatan sehari-hari serta mengurangi risiko dari cuaca ekstrem. Selain itu, model ini juga berpotensi untuk diintegrasikan ke dalam sistem peringatan dini berbasis teknologi informasi, sehingga proses pengambilan keputusan bisa dilakukan lebih cepat dan didukung oleh data yang akurat.

## **1.6 Sistematika Penulisan**

**BAB I PENDAHULUAN**, Bab ini membahas latar belakang masalah yang mendasari penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, ruang lingkup penelitian, serta sistematika penulisan skripsi secara keseluruhan.

**BAB II TINJAUAN PUSTAKA**, Bab ini memuat teori-teori yang relevan dengan topik penelitian, termasuk penjelasan mengenai klasifikasi cuaca, machine learning, algoritma Regresi Linier, Regresi Logistik, Decision Tree, dan Random Forest. Juga mencakup pembahasan mengenai teknik balancing data, evaluasi model klasifikasi (AUC-ROC, akurasi, presisi, recall, atau F1-score), serta penelitian-penelitian terdahulu yang relevan sebagai referensi dan perbandingan.

**BAB III METODOLOGI PENELITIAN**, Bab ini menjelaskan metode penelitian yang digunakan, mulai dari jenis dan pendekatan penelitian, sumber dan jenis data, tahapan preprocessing data, metode undersampling, pembagian data, implementasi algoritma klasifikasi, evaluasi model, serta tahapan hyperparameter tuning. Juga dijelaskan tools dan perangkat lunak yang digunakan dalam eksperimen.

**BAB IV HASIL DAN PEMBAHASAN**, Bab ini menyajikan hasil implementasi dan evaluasi dari masing-masing model klasifikasi (Regresi Linier, Regresi Logistik, Decision Tree, dan Random Forest), baik sebelum maupun sesudah balancing data. Juga dilakukan analisis dan perbandingan performa model berdasarkan metrik evaluasi. Disertai dengan visualisasi hasil dan interpretasi

terhadap pengaruh fitur-fitur cuaca dalam proses klasifikasi.

**BAB V PENUTUP**, Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran-saran yang dapat diberikan untuk penelitian selanjutnya atau pengembangan lebih lanjut dari model klasifikasi cuaca.

