

BAB I

PENDAHULUAN

1.1 Latar Belakang

Istilah cuaca dan iklim sering dianggap sama, padahal keduanya memiliki perbedaan mendasar terutama dalam dimensi temporal. Cuaca merujuk pada keadaan fisik lapisan udara di lokasi dan momen spesifik yang bersifat jangka pendek, sementara iklim merepresentasikan pola yang terbentuk dari kumulatif fenomena cuaca yang telah melalui proses pengumpulan data dan perhitungan rata-rata dalam jangka waktu yang panjang (Fadlan et al., 2022). Cuaca didefinisikan sebagai kondisi udara yang terjadi pada suatu wilayah tertentu dalam cakupan area yang relatif terbatas dan berlangsung dalam jangka waktu yang singkat. Hal ini sejalan dengan pernyataan dari *World Climate Conference* yang mendefinisikan cuaca sebagai keadaan atmosfer secara keseluruhan, mencakup proses perubahan, perkembangan, hingga berakhirnya suatu fenomena atmosfer.

Banyaknya variabel dan parameter atmosferik yang berpengaruh terhadap cuaca menghadirkan tantangan dalam mewujudkan prediksi cuaca dengan tingkat akurasi tinggi dan proses yang efisien. Peramalan cuaca, sebagai suatu bidang ilmu terapan yang mengintegrasikan pengetahuan meteorologi dan sistem teknologi canggih, didefinisikan sebagai usaha sistematis untuk memproyeksikan kondisi dan karakteristik atmosfer bumi pada waktu-waktu mendatang dengan referensi lokasi geografis yang tertentu.

Menurut (Pabontongan et al., 2024), Cuaca merupakan suatu kondisi udara yang berlangsung dalam rentang waktu singkat dan hanya mencakup wilayah dengan area yang relatif sempit. Definisi ini diperkuat oleh *World Climate Conference* yang menyatakan bahwa cuaca menggambarkan keseluruhan keadaan atmosfer, termasuk di dalamnya segala bentuk perubahan, dinamika perkembangan, serta proses berakhirnya suatu fenomena yang terjadi di atmosfer. Hal ini disebut dengan *curse of dimensionality*.

(Fadlan et al., 2022) Dalam penelitiannya menjelaskan bahwa perubahan kondisi atmosfer dipengaruhi oleh sejumlah parameter yang saling berkaitan dalam mendukung proses tersebut. Namun demikian, setiap parameter tidak

selalu memberikan pengaruh yang sama terhadap perubahan kondisi atmosfer, sehingga diperlukan suatu pendekatan atau metode tertentu untuk mengidentifikasi parameter-parameter mana saja yang secara signifikan berkontribusi terhadap perubahan tersebut.

Pada penelitian yang dilakukan oleh (Setiawan et al., 2025), Perubahan cuaca yang semakin tidak menentu dan cenderung ekstrem kini menjadi isu yang mendapat perhatian serius di tingkat global. Di Indonesia, kondisi geografis yang sangat beragam menjadikan pola cuaca lokal semakin kompleks dan sulit untuk diprediksi secara akurat. Oleh karena itu, ketersediaan prakiraan cuaca yang tepat menjadi sangat penting mengingat ketidakakuratan dalam prediksi cuaca dapat memberikan dampak yang merugikan bagi berbagai sektor kehidupan, sehingga dibutuhkan suatu metode yang andal dan sesuai dalam menganalisis serta memprediksi kondisi cuaca. Hal tersebut menjadi faktor bahwa diperlukannya sebuah teknologi yang dapat secara tepat dalam proses prakiraan cuaca.

Beberapa penelitian terhadap algoritma klasifikasi untuk permasalahan prakiraan cuaca dan diantaranya beberapa melakukan komparasi algoritma untuk mendapatkan hasil yang terbaik. Pada penelitian sebelumnya beberapa teknik data mining yaitu *Association rule*, *decision tree C4.5*, *Classification* dan *Random Forest* (Naldy & Andri, 2021). Hasil yang didapat yaitu algoritma *decision tree C4.5* memiliki tingkat akurasi tertinggi. Penelitian lainnya dilakukan oleh (Informasi, 2023) menggunakan *algoritma decision tree See5 (C5.0)* untuk menemukan parameter-parameter yang memiliki pengaruh besar terhadap perubahan cuaca. Mereka mendapatkan bahwa temperatur maksimum, temperatur minimum, curah hujan, penguapan, dan kecepatan angin terindikasi sebagai parameter-parameter yang paling signifikan mempengaruhi perubahan cuaca. Penelitian selanjutnya dilakukan oleh (Husriansyah & Anshari, 2025) menggunakan algoritma klasifikasi *decision tree C5.0* untuk menghasilkan *decision tree* dan aturan klasifikasi parameter cuaca. Penelitian juga menggunakan algoritma *Artificial Neural Networks (ANN)* yang dapat mendeteksi hubungan antara variabel input dan menghasilkan output berdasarkan pola observasi data. Penelitian berikutnya

dilakukan oleh (Sari et al., 2024) menggunakan 5 algoritma data mining untuk membentuk model prediksi yaitu *Artificial Neural Networks (ANN)*, *Random Forest*, *Classification and Regression Tree (C&RT)*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbour (k-NN)*. Algoritma *FP growth* dengan evaluasi *Mean Absolute Error (MAE)*, *Mean Square Error (MSE)*, dan *Standard Deviation (SD)* menampilkan hasil yang lebih akurat dibandingkan dengan algoritma *Artificial Neural Network (ANN)*.

Berdasarkan paparan tersebut, permasalahan yang terdapat pada prakiraan cuaca yaitu banyaknya atribut yang digunakan dalam datasets. Pada umumnya, dengan atribut yang jumlahnya banyak dan data yang memiliki jumlah yang banyak, jika semua atribut tersebut digunakan, maka akan mengurangi kinerja dari algoritma menyebabkan akurasi menjadi rendah. Hal lain yang dapat menyebabkan akurasi sebuah algoritma tidak maksimal yaitu adanya ketidakseimbangan kelas antara kelas mayoritas dan kelas minoritas. Hal ini disebut dengan *imbalanced class*. Penelitian yang dilakukan menggunakan algoritma *decision tree C4.5* untuk menemukan pola klasifikasi cuaca dan berfungsi untuk mengeliminasi perhitungan yang tidak diperlukan karena sample yang diuji hanya berdasarkan pada kelas tertentu sehingga kinerja yang dihasilkan dapat lebih baik dan akurat.

Salah satu tantangan utama dalam klasifikasi data cuaca adalah permasalahan yang dikenal dengan istilah *curse of dimensionality*, yaitu kondisi di mana penggunaan seluruh atribut yang tersedia dalam dataset justru menurunkan kinerja algoritma klasifikasi. Hal ini disebabkan oleh keberadaan atribut-atribut yang tidak relevan atau redundan, yang menambah kompleksitas komputasi tanpa memberikan kontribusi berarti terhadap peningkatan akurasi. Pada dataset cuaca yang memiliki 18 atribut seperti yang digunakan dalam penelitian ini, tidak semua atribut memiliki tingkat signifikansi yang sama terhadap penentuan kondisi cuaca. Atribut-atribut tertentu mungkin saling berkorelasi tinggi satu sama lain, sementara atribut lainnya mungkin mengandung noise yang justru mempersulit proses pembentukan pola klasifikasi.

Untuk mengatasi permasalahan tersebut, diperlukan suatu metode seleksi fitur yang mampu mengidentifikasi dan memilih subset atribut yang paling optimal dari keseluruhan atribut yang tersedia. Salah satu metode yang terbukti efektif dalam melakukan seleksi fitur adalah Particle Swarm Optimization (PSO). PSO merupakan algoritma metaheuristik yang terinspirasi dari perilaku kolektif kawanan burung atau ikan dalam mencari sumber makanan. Dalam konteks seleksi fitur, setiap partikel dalam PSO merepresentasikan satu kandidat solusi berupa kombinasi subset atribut yang dipilih, dan melalui mekanisme iteratif yang mempertimbangkan pengalaman terbaik masing-masing partikel (personal best) serta pengalaman terbaik seluruh kawanan (global best), PSO secara bertahap mengarahkan pencarian menuju kombinasi atribut yang menghasilkan akurasi klasifikasi tertinggi. Keunggulan PSO dibandingkan metode seleksi fitur lainnya terletak pada kemampuannya menjelajahi ruang solusi yang besar secara efisien tanpa memerlukan evaluasi ekshaustif terhadap seluruh kemungkinan kombinasi subset atribut, sehingga proses optimasi dapat diselesaikan dalam waktu komputasi yang lebih singkat.

Beberapa penelitian terdahulu telah membuktikan efektivitas kombinasi PSO dengan algoritma Decision Tree dalam berbagai domain klasifikasi. Penerapan PSO sebagai metode seleksi fitur terbukti mampu meningkatkan akurasi algoritma klasifikasi secara signifikan sekaligus mengurangi jumlah atribut yang digunakan, sehingga model yang dihasilkan menjadi lebih efisien dan lebih mudah diinterpretasikan. Oleh karena itu, penelitian ini mengusulkan penggabungan algoritma Decision Tree C4.5 dengan metode optimasi PSO sebagai pendekatan terpadu dalam menyelesaikan permasalahan klasifikasi data cuaca wilayah Surabaya.

Berdasarkan permasalahan yang ada, penelitian ini dibuat untuk tujuan mengklasifikasi cuaca menggunakan metode *Decision Tree 4.5*. Datasets yang digunakan adalah datasest cuaca Surabaya tahun Januari 2021 – 2024 dari situs *Iowa Environmental Mesonet* dengan interval waktu Harian.

1.2 Rumusan Masalah

Kendala yang terdapat dalam proses prakiraan cuaca yaitu banyaknya atribut yang digunakan, sehingga dengan banyaknya jumlah atribut dan jumlah

data, maka jika semua atribut tersebut digunakan akan mengurangi kinerja dari algoritma dalam melakukan proses prakiraan cuaca. Berdasarkan hal tersebut, maka yang menjadi rumusan masalah pada penelitian sebagai berikut:

- a. Parameter apa saja yang mempengaruhi pola klasifikasi cuaca yang dihasilkan oleh algoritma data mining *decision tree C4.5*?
- b. Bagaimana mengoptimalkan fitur dengan menggunakan PSO serta bagaimana akurasi setelah dioptimalkan?
- c. Apakah benar terdapat ketidak seimbangan kelas pada data yang digunakan?

1.3 Tujuan dan Manfaat

A. Tujuan

Tujuan dari penelitian yang dilakukan bertujuan untuk mencari pola klasifikasi cuaca dengan algoritma klasifikasi data mining *decision tree C4.5* dalam proses prediksi kondisi cuaca dan optimasi dengan algoritma *Particle Swarm Optimization (PSO)* terhadap hasil klasifikasi cuaca di wilayah Surabaya.

B. Manfaat

Manfaat dari penelitian yang dilakukan diharapkan mengetahui kinerja algoritma klasifikasi *decision tree C4.5* yang telah dioptimalkan dengan algoritma *Particle Swarm Optimization (PSO)* dalam mengoptimalkan fitur dalam mengatasi ketidak seimbangan kelas yang terjadi sehingga melakukan proses klasifikasi terhadap data historis cuaca dapat menemukan pola klasifikasi cuaca yang cukup akurat.

1.4 Ruang Lingkup Penelitian

Batasan pada penelitian ini terletak pada jumlah data yang digunakan dalam pengujian algoritma yang menggunakan software RapidMiner. Data yang digunakan merupakan data historis cuaca wilayah Surabaya tahun 2021 sampai tahun 2024 yaitu sebanyak 1388 record data dengan atribut dan 1 atribut kelas serta 17 jenis kelas didalamnya dengan interval selang waktu setiap 24 Jam setiap harinya.