

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Letak geografis Indonesia sangatlah strategis, terletak di kawasan tropis di antara dua benua besar, Asia dan Australia, serta di antara dua samudera penting, Pasifik dan Hindia. Selain itu, Indonesia dilintasi oleh garis khatulistiwa dan terdiri dari sejumlah pulau yang membentang dari barat ke timur. Dikelilingi oleh lautan yang luas, kondisi ini memberikan Indonesia keanekaragaman dalam hal cuaca dan iklim (Ridwan, 2019). Salah satu keuntungan dari wilayah yang dilintasi oleh garis khatulistiwa adalah memiliki 2 musim, yaitu musim hujan dan musim kemarau (Arpan et al., 2004).

Cuaca adalah fenomena atmosfer yang diamati dalam rentang waktu yang singkat dan di lokasi tertentu. Menurut Konferensi Iklim Dunia, cuaca adalah hasil dari interaksi kompleks di atmosfer, melibatkan perubahan, evolusi, serta kemunculan atau lenyapnya berbagai fenomena udara (Fadholi, 2013). Banyak faktor yang dapat mempengaruhi perubahan cuaca dan iklim, termasuk suhu, kelembapan, tekanan udara, kecepatan angin, arah angin pada kecepatan maksimum, durasi sinar matahari, dan tingkat curah hujan (Mujiasih, 2011). Hujan merupakan salah satu elemen cuaca yang penting karena mempengaruhi kondisi sumber air di daerah yang terkena hujan. Hujan menunjukkan variasi ruang dan waktu, yang berarti jumlah hujan berbeda-beda antara satu tempat dengan tempat lain dan antara satu waktu dengan waktu lainnya. Proses pembentukan hujan memerlukan tiga hal: jumlah uap air yang cukup, proses pendinginan untuk terjadinya kondensasi, dan jumlah inti kondensasi yang memadai (Sumner, 1988). Proses pembentukan hujan membutuhkan tiga hal: jumlah uap air yang memadai, adanya proses pendinginan untuk memungkinkan kondensasi, dan cukupnya inti kondensasi. Kelembapan udara, yang merepresentasikan jumlah uap air, memainkan peran penting semakin tinggi kelembapan, semakin besar kemungkinan terjadinya hujan. Ada beberapa mekanisme untuk menciptakan kondensasi, seperti pengangkatan massa udara, kontak dengan massa udara yang lebih dingin, dan pencampuran dengan massa udara yang lebih dingin (Linacre & Hobbs, 1987).

Perkembangan teknologi saat ini sangat mendukung manusia dalam melakukan

pengamatan cuaca, seperti pengklasifikasian curah hujan harian. Salah satu metode yang digunakan adalah K-NN dan algoritma genetika. Algoritma K-Nearest Neighbors (K-NN) digunakan dalam *machine learning* untuk tugas klasifikasi dan regresi. Algoritma ini bekerja dengan mengklasifikasikan data baru berdasarkan mayoritas label dari k tetangga terdekat dalam ruang fitur. Dalam klasifikasi, K-NN menghitung jarak antara data yang akan diklasifikasikan dengan data pelatihan berlabel. Semakin dekat jaraknya, semakin besar kemungkinan data tersebut memiliki label yang sama. Meskipun K-NN tidak memerlukan proses pelatihan yang rumit, algoritma ini sering kali membutuhkan komputasi yang intensif saat melakukan prediksi pada dataset yang besar (Han et al., 2011).

Namun, Metode K-Nearest Neighbors (K-NN) memiliki kelemahan yaitu penentuan nilai k, yaitu jumlah nilai tetangga terdekat yang akan dipertimbangkan dalam pengambilan keputusan. Pengambilan nilai k yang optimal sangatlah krusial, namun sering kali bersifat bias dan abstrak, bergantung pada berbagai percobaan yang memakan waktu dan sumber daya. Untuk mengatasi kelemahan ini memerlukan pendekatan yang lebih sistematis dan adaptif dalam menentukan nilai k yang optimal. Salah satu solusi yang diusulkan adalah penggunaan algoritma genetika.

Algoritma genetika adalah metode optimasi yang didasarkan pada prinsip evolusi alami. Dalam algoritma ini, sekelompok solusi (populasi) dievaluasi, dan solusi terbaik digunakan sebagai dasar untuk menghasilkan solusi baru melalui proses seleksi, rekombinasi, dan mutasi. Proses ini diulang secara iteratif hingga solusi yang memuaskan ditemukan. Algoritma genetika sering diterapkan untuk memecahkan masalah optimasi yang kompleks di berbagai bidang, termasuk pengembangan model prediksi dan optimalisasi struktur jaringan saraf tiruan (Goldberg, 1989).

Dengan menggunakan algoritma genetika, nilai k dapat dioptimalkan secara lebih efisien dan akurat, mengurangi bias dan ketidakpastian yang biasanya muncul dalam penentuan nilai k secara manual. Pendekatan ini memungkinkan peneliti untuk menemukan nilai k yang paling sesuai dengan karakteristik data yang dihadapi, sehingga meningkatkan performa dan akurasi model K-NN secara keseluruhan.

Berdasarkan penelitian yang pernah dilakukan sebelumnya yang merupakan klasifikasi curah hujan harian dengan dataset yang diperoleh dari Stasiun Pengamatan Ngurah Rai, Bali. Yang menggunakan metode Learning Vector Quantization berhasil

diimplementasikan dengan hasil yang berupa Tingkat akurasi sebesar 49,6%. Hal ini dapat terjadi karena data yang digunakan untuk pelatihan dan pengujian cenderung sedikit dan terdapat *overfitting* didalamnya yang sekaligus menandakan bahwa metode Learning Vector Quantization memiliki performa yang buruk apabila digunakan untuk melakukan klasifikasi pada dataset yang sedikit dan *overfitting* (Gunadi, 2022).

Kemudian pada penelitian selanjutnya yang melakukan klasifikasi curah hujan di Kota Semarang yang menggunakan dataset yang diperoleh dari BMKG Semarang yang terdiri atas 9 variabel dan 1096 record data. Penelitian tersebut menggunakan metode Regresi Logistik, Random Forest, serta Gradient Boosting. Pada penelitian ini, metode Regresi Logistik menghasilkan Tingkat akurasi sebesar 69,8%, lalu metode Random Forest menghasilkan Tingkat akurasi sebesar 72%, serta metode Gradient Boosting menghasilkan Tingkat akurasi sebesar 73,1% (Usman & Sudiby, 2022). Ketiga metode tersebut memiliki kelemahan seperti regresi logistic yang tidak cukup fleksibel untuk menangkap hubungan non-linear antara variabel input dan output. Dalam kasus data meteorologi yang sering kali non-linear, metode ini bisa kurang efektif. Kemudian random forest yang dapat terpengaruh oleh variabel yang kurang penting dan noise dalam data dan cenderung *overfitting* pada data pelatihan jika tidak diatur dengan baik. Serta gradient boosting yang memiliki kecenderungan untuk *overfitting* jika parameter tuning tidak dilakukan dengan hati-hati. Selain itu, metode ini juga lebih sensitif terhadap data yang memiliki noise, dan performanya bisa sangat bergantung pada kualitas praproses data. Oleh karena itu, ketiga metode tersebut kurang optimal jika diterapkan untuk klasifikasi curah hujan harian.

Sedangkan pada penelitian yang menggunakan metode K-Nearest Neighbor dengan optimasi algoritma genetika yang digunakan untuk klasifikasi pada 5 dataset medis menghasilkan Tingkat akurasi sebesar 99,2%, 86,44%, 71,69%, 98,59%, dan 87,5% untuk setiap datasetnya. Hal tersebut menunjukkan bahwasanya metode K-Nearest Neighbor dengan optimasi algoritma genetika merupakan metode yang efektif dalam melakukan klasifikasi terhadap berbagai dataset yang dibuktikan dengan keluaran Tingkat akurasi yang tinggi pada setiap dataset yang digunakan untuk proses klasifikasi. Hal inilah yang menyebabkan penulis memilih untuk menggunakan metode tersebut dalam penelitian ini.

Dari latar belakang tersebut, penulis melakukan penelitian dengan judul

“KLASIFIKASI CURAH HUJAN HARIAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR DENGAN OPTIMASI ALGORITMA GENETIKA” dengan *output* yang diharapkan yaitu tingkat akurasi dari penerapan metode yang digunakan.

### **1.2. Rumusan Masalah**

Dengan merujuk pada konteks yang telah diuraikan sebelumnya dalam penelitian “Klasifikasi Curah Hujan Harian Menggunakan Metode K-Nearest Neighbor Dengan Optimasi Algoritma Genetika”, maka dapat di buat beberapa rumusan masalah sebagai berikut:

1. Bagaimana penerapan Metode K-Nearest Neighbor Dengan Optimasi Algoritma Genetika untuk klasifikasi curah hujan harian?
2. Bagaimana tingkat akurasi dari penerapan Metode K-Nearest Neighbor Dengan Optimasi Algoritma Genetika untuk klasifikasi curah hujan harian?

### **1.3. Tujuan**

Dari analisis masalah yang telah diuraikan sebelumnya, penelitian ini bertujuan untuk mengatasi tantangan yang dihadapi dengan mengidentifikasi beberapa tujuan khusus yang akan dikejar dalam penelitian ini:

1. Mengimplementasikan Metode K-Nearest Neighbor dan Algoritma Genetika untuk klasifikasi curah hujan harian.
2. Mengetahui tingkat akurasi dari penerapan Metode K-Nearest Neighbor dan Algoritma Genetika untuk klasifikasi curah hujan harian.

### **1.4. Manfaat**

Setelah tujuan penelitian telah diungkapkan sebelumnya, studi ini juga memiliki sejumlah manfaat yang dapat diperoleh, antara lain:

1. Mengetahui implementasi Metode K-Nearest Neighbor dan Algoritma Genetika ketika digunakan untuk proses data mining.
2. Menyampaikan potensi kontribusi penelitian terhadap pengembangan sistem kecerdasan buatan untuk klasifikasi curah hujan.

### 1.5. Batasan Masalah

Berdasarkan pembahasan mengenai permasalahan sebelumnya, terdapat sejumlah batasan yang telah ditetapkan untuk mengarahkan lingkup penelitian “Klasifikasi Curah Hujan Harian Menggunakan Metode K-Nearest Neighbor Dengan Optimasi Algoritma Genetika” diantaranya:

1. Sumber data yang digunakan dalam penelitian ini adalah data primer yang diperoleh langsung dari website resmi Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Indonesia. Dataset tersebut mencakup pengamatan cuaca harian dari Stasiun Meteorologi Kelas III Sultan Muhammad Salahuddin - Bima selama lima tahun terakhir, dengan total 1826 rekaman data.
2. Dataset ini dibagi menjadi enam kategori cuaca, yaitu cerah, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem.
3. Pendekatan yang digunakan dalam analisis data adalah metode K-Nearest Neighbor yang telah dioptimalkan menggunakan Algoritma Genetika.
4. Pelaksanaan penelitian dilakukan menggunakan bahasa pemrograman Python 3 dengan memanfaatkan Google Collaboratory sebagai editor kode.
5. Penelitian ini memanfaatkan beberapa library Python, termasuk DEAP, Numpy, Pandas, Matplotlib, Sklearn, Random, dan Time.
6. Dataset yang digunakan terdiri dari sembilan atribut yang relevan untuk mengevaluasi kinerja algoritma, termasuk Tn, Tx, Tavg, RH\_avg, RR, ss, ff\_x, ddd\_x, dan ff\_avg.
7. Keluaran yang dihasilkan adalah nilai k yang ditemukan, tingkat keakurasian dari klasifikasi curah hujan harian, pada generasi beberapa nilai k tersebut ditemukan, serta waktu eksekusi program.