

BAB I

PENDHULUAN

1.1 Latar Belakang

Indonesia memiliki posisi geografis yang sangat strategis, terletak di kawasan tropis antara Benua Asia dan Australia, serta diapit oleh Samudra Pasifik dan Samudra Hindia. Wilayahnya yang dilintasi garis khatulistiwa dan terdiri dari ribuan pulau yang membentang dari barat ke timur, serta dikelilingi perairan luas, menjadikan kondisi iklim dan cuaca di Indonesia sangat bervariasi. Karakteristik geografis ini mendukung terbentuknya iklim hutan hujan tropis, yang ditandai oleh suhu udara yang cenderung stabil dan curah hujan tinggi yang tersebar relatif merata sepanjang tahun [1].

DKI Jakarta sebagai salah satu provinsi di Indonesia, juga turut merasakan dampak dari perubahan iklim tersebut. Perubahan iklim global berdampak pada kondisi atmosfer di berbagai wilayah, termasuk Jakarta. Salah satu parameter yang paling terdampak adalah curah hujan, yang distribusinya menunjukkan perubahan signifikan dan berpotensi menimbulkan bencana seperti banjir. Penelitian mengenai data curah hujan dasarian periode 1990–2020 mengungkap adanya pergeseran awal dan panjang musim hujan di seluruh Zona Musim (ZOM) Jakarta. Selain itu, intensitas curah hujan di atas 200 mm/dasarian tercatat lebih sering terjadi pada dekade 2001–2020 dibandingkan dekade sebelumnya[2].

Sebagai bagian dari wilayah pesisir yang mengalami perkembangan pesat, Jakarta Utara, khususnya melalui Stasiun Meteorologi Maritim Tanjung Priok telah menjadi sumber data historis curah hujan yang penting dalam analisis iklim. Data

ini sangat berharga untuk mendukung strategi pembangunan yang adaptif terhadap perubahan iklim, sebagaimana dianjurkan dalam berbagai studi dan proyeksi iklim jangka panjang yang disediakan oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dan *Global Circulation Models* (GCMs)[3].

Algoritma *K-Nearest Neighbors* (KNN) merupakan salah satu metode klasifikasi yang paling sederhana dan banyak digunakan dalam pembelajaran mesin. Prinsip kerjanya adalah dengan menentukan kelas suatu data berdasarkan kategori dari data-data terdekat di sekitarnya dalam ruang fitur. Konsep ini bersifat intuitif, di mana objek yang memiliki karakteristik serupa dengan kelompok tertentu kemungkinan besar juga termasuk dalam kelompok tersebut[4].

Penelitian yang secara khusus mengeksplorasi penerapan algoritma *K-Nearest Neighbors* (KNN) dalam klasifikasi tingkat curah hujan pada stasiun Meteorologi Maritim Tanjung Priok masih terbatas. Pada penelitian ini, algoritma *K-Nearest Neighbors* (KNN) diterapkan untuk mengklasifikasikan curah hujan harian dengan menggunakan *Euclidean Distance* sebagai metrik jarak. Pemilihan *Euclidean Distance* didasarkan pada kelebihanannya dibandingkan *Manhattan Distance*. Metrik ini menghitung jarak lurus antara dua titik dalam ruang multidimensi, sehingga lebih cocok untuk data kontinu seperti suhu, kelembaban, dan curah hujan. Selain itu, *Euclidean Distance* mampu menangkap hubungan geometris antar variabel yang relevan dengan pola cuaca. Kemudahan implementasi dan akurasinya menjadikan *Euclidean Distance* sebagai pilihan yang tepat untuk mendukung performa algoritma KNN dalam klasifikasi curah hujan[5].

Penelitian yang menjelaskan tentang curah hujan merupakan komponen penting dalam iklim dan klasifikasinya menjadi topik yang menarik. Tingkat curah hujan sulit ditentukan secara pasti, namun dapat diprediksi. Metode *K-Nearest Neighbors* (KNN) adalah pendekatan non-parametrik yang efisien untuk klasifikasi, dengan teknik *Ensemble Subset K-Nearest Neighbor* digunakan untuk meningkatkan akurasi dalam data curah hujan. Penelitian ini menggunakan data BMKG 2014-2018. Hasilnya menunjukkan bahwa pada nilai $k = 3$ dan $k = 6$, tingkat kesalahan mencapai 0,260223 dengan akurasi klasifikasi 0,739777 [6].

Penelitian yang membahas tentang Ketidakpastian curah hujan dapat menyebabkan bencana seperti banjir, mengganggu aktivitas masyarakat. Untuk mengantisipasi hal ini, prediksi cuaca yang akurat diperlukan. Penelitian ini menggunakan dataset BMKG dari 1 Januari 2019 hingga 31 Oktober 2023 dan menerapkan algoritma *K-Nearest Neighbors* (KNN) untuk mengklasifikasikan curah hujan ke dalam lima kategori. Penelitian tersebut memiliki tujuan untuk memprediksi curah hujan di BMKG Sulawesi Tengah. Hasil evaluasi menunjukkan nilai $K = 23$ dengan akurasi 83,0%, menandakan kinerja KNN yang baik dalam klasifikasi cuaca [7].

Penelitian ini juga menerapkan metode *Cross Validation* dan *Grid Search* untuk mengoptimalkan performa model. *Cross Validation* dilakukan misalnya dengan teknik *k-fold* sebagai upaya validasi menyeluruh untuk memastikan model mampu digeneralisasi ke data baru. *Grid Search* kemudian digunakan untuk menelusuri kombinasi parameter terbaik, di mana setiap konfigurasi diuji

menggunakan *Cross Validation* agar pemilihan hyperparameter didasarkan pada evaluasi yang komprehensif dan independen[8].

Dengan memanfaatkan teknologi dan metodologi yang tepat, serta data yang relevan dari Stasiun Meteorologi Maritim Tanjung Priok, diharapkan penelitian ini dapat memberikan kontribusi dalam pemahaman yang lebih baik tentang pola curah hujan di Kota Jakarta Utara. Penelitian ini juga dapat membantu dalam perencanaan adaptasi dan mitigasi terkait perubahan iklim. Studi sebelumnya menunjukkan bahwa wilayah pesisir Jakarta Utara rentan terhadap perubahan pola curah hujan dan kenaikan muka air laut, sehingga analisis data historis sangat penting untuk mendukung kebijakan adaptasi dan pengelolaan sumber daya wilayah[3].

Penelitian ini bertujuan untuk mengembangkan model klasifikasi yang efektif berdasarkan data curah hujan harian historis pada Stasiun Meteorologi Maritim Tanjung Priok dengan judul penelitian **“Klasifikasi Tingkat Curah Hujan Harian Di Stasiun Meteorologi Maritim Tanjung Priok Dengan Algoritma K-Nearest Neighbors”**

1.2 Identifikasi Masalah

Berdasarkan latar belakang tersebut, dapat diidentifikasi masalah sebagai berikut:

1. Tidak adanya model klasifikasi curah hujan harian yang memadai.
2. Belum optimalnya pengolahan data curah hujan historis di Stasiun Meteorologi Maritim Tanjung Priok, sehingga informasi berharga belum dimanfaatkan sepenuhnya.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah tersebut, rumusan masalah yang dapat diambil adalah:

1. Bagaimana mengembangkan model klasifikasi curah hujan dengan algoritma *K-Nearest Neighbors*?
2. Bagaimana mengoptimalkan pengolahan data curah hujan historis di Stasiun Meteorologi Maritim Tanjung Priok?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, tujuan yang dapat diambil dari penelitian kali ini adalah:

1. Mengembangkan model klasifikasi curah hujan dengan menggunakan algoritma *K-Nearest Neighbors* (KNN) yang dapat mengelompokkan data curah hujan berdasarkan tingkat intensitasnya.
2. Mengoptimalkan pengolahan data curah hujan historis di Stasiun Meteorologi Maritim Tanjung Priok untuk meningkatkan akurasi dan efesiensi analisis data cuaca.

1.5 Manfaat Penelitian

Berdasarkan tujuan penelitian tersebut, manfaat yang dapat di ambil dari penelitian ini adalah manfaat teoritis dan manfaat praktis dari penelitian klasifikasi curah hujan menggunakan algoritma *K-Nearest Neighbor* (KNN):

1. Manfaat Teoritis

Penelitian ini bisa menambah pengetahuan di bidang ilmu komputer, terutama dalam hal pengolahan data menggunakan metode KNN. Secara teori, penelitian ini menunjukkan bahwa algoritma KNN juga bisa diterapkan untuk memproses data cuaca harian yang sifatnya berubah-ubah dan punya pola musiman. Selain itu, penelitian ini bisa menjadi bahan pertimbangan bagi orang lain yang ingin melakukan penelitian serupa, misalnya dalam memilih metode klasifikasi yang cocok atau cara pengolahan data agar hasilnya lebih akurat.

2. Manfaat Praktis

Hasil dari penelitian ini bisa bermanfaat langsung bagi instansi seperti BMKG, para petani, ataupun pemerintah daerah. Dengan adanya sistem yang bisa mengelompokkan curah hujan ke dalam kategori tertentu, mereka bisa lebih mudah mengambil keputusan yang berkaitan dengan cuaca. Contohnya, petani bisa tahu kapan waktu yang tepat untuk mulai menanam atau memanen, sementara pemerintah bisa lebih siap dalam menghadapi risiko banjir atau kekeringan. Sistem ini juga bisa dikembangkan menjadi aplikasi yang bisa diakses langsung oleh masyarakat untuk melihat informasi cuaca harian secara cepat dan mudah.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Pada penelitian ini menggunakan data yang diambil dari website BMKG dari tahun 2015 sampai tahun 2023 dari Stasiun Meteorologi Maritim Tanjung Priok yang ada di Kota Jakarta Utara sebanyak 3287 data, sehingga hasil analisis hanya merepresentasikan pola cuaca di wilayah ini.

2. Hasil klasifikasi cuaca akan dibagi ke dalam tiga kategori, yaitu berawan, hujan ringan, hujan sedang, hujan sangat lebat, yang diwakili oleh variabel rangkuman cuaca sebagai data uji. Sementara itu, variabel lainnya seperti temperatur minimal, temperatur maksimal, temperatur rata-rata, kelembaban rata-rata, penyinaran matahari, kecepatan angin, arah angin, kecepatan angin rata-rata akan digunakan sebagai data latih. Total dataset yang digunakan berjumlah 3287 data untuk setiap variabel.
3. Metode klasifikasi yang digunakan adalah algoritma *K-Nearest Neighbors* (KNN) sebagai algoritma pengujian klasifikasi cuaca dengan nilai k yang digunakan adalah 3,5,7, dan 9. Penelitian tidak mencakup perbandingan mendalam dengan algoritma lain.
4. Dataset yang digunakan dalam penelitian ini adalah berbentuk *xlsx* atau *excel*. Dataset awal tidak memiliki kategori curah hujan, sehingga kategori akan dibuat berdasarkan standar BMKG atau literatur yang relevan.
5. Dampak dari penelitian menghasilkan model klasifikasi yang efektif. Namun, implemetasi langsung dari hasil penelitian ini untuk perencanaan mitigasi atau adaptasi perubahan iklim tidak menjadi fokus dalam penelitian.

1.7 Sistematika Penulisan

Penjelasan ini tentang subjek atau elemen dalam topik skripsi ini berbentuk dari lima bab, sebagai berikut inti dari setiap babnya:

BAB I PENDAHULUAN

Bagian ini diuraikan latar belakang, identifikasi masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bagian ini menguraikan secara singkat tentang teori-teori yang berhubungan dengan judul.

BAB III METODOLOGI PENELITIAN

Bagian ini berisi mengenai objek penelitian, kerangka penelitian yang langsung berkaitan dengan metode pengumpulan data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dari proses pengkodean yang telah dilakukan, dengan fokus pada tampilan antarmuka sistem saat dijalankan. Selain itu, bab ini juga menjelaskan tahapan pengujian yang dilakukan terhadap sistem tersebut.

BAB V PENUTUP

Bagian ini berisi mengenai kesimpulan dari penelitian yang telah dilakukan dan juga saran untuk penelitian selanjutnya yang akan dilakukan.