

# BAB I

## PENDAHULUAN

### A. Latar Belakang

Udara memainkan peran yang sangat penting dalam kehidupan manusia dan makhluk hidup lainnya. Semua makhluk hidup membutuhkan udara untuk dapat hidup secara optimal, sehingga menjadikannya sebagai sumber daya alam yang harus dilindungi bagi manusia dan makhluk hidup lainnya. Hal itu yang menyebabkan udara menjadi bagian tidak terpisahkan dari lingkungan hidup sehingga harus dilestarikan dan diperhatikan kualitasnya. Maka dari itu, pemantauan kualitas udara sangat penting mengingat tingginya tingkat polusi udara pada saat ini [1].

Polusi udara adalah keadaan udara yang tercampur oleh substansi kimiawi, fisik, maupun biologis yang berpotensi membahayakan makhluk hidup dan lingkungan sekitarnya. Bahaya yang ditimbulkan dari polusi tersebut antara lain seperti masalah kesehatan maupun peningkatan suhu udara secara signifikan sehingga menyebabkan pemanasan global. Hal ini sering terjadi di perkotaan besar ataupun kawasan industri yang sering membuang gas-gas sisa hasil produksi. Selain itu, peningkatan jumlah kendaraan bermotor di area perkotaan.

Peningkatan jumlah kendaraan bermotor di Indonesia memiliki peran dalam peningkatan emisi gas CO yang dihasilkan sehingga pencemaran udara yang semula 70% menjadi 80% [2]. Polusi udara yang tinggi menyebabkan penurunan kualitas udara. Sehubungan dengan masalah tersebut, pengukuran dan pengamatan kualitas udara harus selalu dilakukan. Indonesia menerapkan standar pengukuran kualitas udara yang resmi, yaitu Indeks Standar Pencemaran Udara (ISPU), sesuai dengan Keputusan Menteri Negara Lingkungan Hidup Nomor: KE-45/MENLH/1997 tentang Standar Indeks Pencemaran Udara (ISPU). Parameter ISPU meliputi sulfur dioksida (SO<sub>2</sub>), partikel (PM<sub>10</sub>), karbon monoksida (CO), ozon (O<sub>3</sub>) dan nitrogen dioksida (NO<sub>2</sub>) [3]. Pendataan ISPU harian dapat digunakan untuk penelitian, seperti memprediksi parameter ISPU di masa

mendatang. Data-data ISPU harian digunakan untuk penelitian, seperti memprediksi parameter ISPU di masa mendatang.

Analisis data tersebut dapat dioptimalkan dengan pemanfaatan teknologi machine learning. Machine learning mampu mempelajari secara cepat dan otomatis data yang diberikan, baik itu data yang bersifat nominal maupun numerik. Kumpulan algoritme machine learning dapat dimanfaatkan untuk melakukan *classification*, termasuk algoritma *Naïve-bayes*, *Decision Tree*, *K-Means*, *K-Nearest Neighbors*, dan lain sebagainya.

Penelitian mengenai klasifikasi kualitas udara pada Daerah Istimewa Yogyakarta dilakukan dengan menerapkan algoritme *Random Forest* dan *K-Nearest Neighbors*. Algoritme *Random Forest* (RF) merupakan metode untuk menghasilkan child node pada setiap node secara acak sehingga meningkatkan keakuratan hasil. *Random Forest* ini digunakan untuk membuat pohon keputusan yang terdiri dari root node, internal node, dan leaf node dengan mengambil data atribut secara acak dan sesuai dengan peraturan yang berlaku. Penggunaan algoritme *Random Forest* memiliki kinerja klasifikasi yang baik dengan tingkat *error* atau kesalahan relatif kecil, kemampuan menangani *dataset* dengan jumlah data yang banyak secara efisien, sekaligus mampu menawarkan metode yang lebih unggul untuk bekerja dengan data yang *noisy* atau hilang. Nilai yang hilang tersebut digantikan oleh variabel yang paling sering muncul di node tertentu.

Algoritme kedua yang digunakan pada penelitian adalah *K-Nearest Neighbors*. KNN adalah metode untuk mengklasifikasikan sampel berdasarkan berdasarkan kedekatan atau jaraknya dengan pelatihan yang disimpan dataset yang tersimpan. Ini adalah metode yang sederhana tetapi efektif untuk klasifikasi. KNN merupakan jenis algoritma *supervised learning* yang di mana kelas dari sampel ditugaskan ke kelas pelatihan terdekat kelompok. Salah satu keuntungan dari metode KNN adalah tidak ada pelatihan yang terlibat dalam KNN. *Dataset* pelatihan disimpan tanpa diproses dan digunakan secara langsung sebagai basis pengetahuan atau *Knowledge Bases*. Untuk menggunakan metode KNN, sebuah nilai K yang tepat harus dipilih karena keberhasilan klasifikasi sangat bergantung pada nilai K ini.

Berdasarkan permasalahan diatas penulis hendak melakukan penelitian tugas akhir terhadap dataset kualitas udara Yogyakarta tahun 2020 hingga 2021 yang didapatkan melalui kaggle [4] [5]. Penelitian dilakukan dengan memanfaatkan model klasifikasi yang bertugas menganalisa dan menentukan keadaan udara ke dalam kategori tertentu sesuai dengan ketentuan dari ISPU. Algoritme atau metode yang hendak diterapkan dalam penelitian ini yaitu Random Forest dan K-Nearest Neighbor.

## **B. Rumusan Masalah**

Berdasarkan latar belakang yang di atas, dapat dijabarkan permasalahan sebagai berikut:

1. Bagaimana membuat model algoritme *Random Forest* dan *K-Nearest Neighbors* yang dapat mengklasifikasi kualitas udara di Daerah Istimewa Yogyakarta kedepannya.
2. Bagaimana tingkat akurasi yang dihasilkan oleh algoritma random forest dan k-nearest neighbors dalam mengklasifikasi data ISPU kota Yogyakarta.

## **C. Batasan Masalah**

Pada penelitian ini, ditetapkanlah pembatasan masalah yang menjadi fokus pada penelitian ini:

1. Data yang digunakan merupakan dataset berisi pengukuran polusi udara, seperti *Particulate Matter* (PM10), Sulfur Dioksida (SO<sub>2</sub>), Karbon Monoksida (CO), Ozon (O<sub>3</sub>), dan Natrium Dioksida (NO<sub>2</sub>) yang berada di Daerah Istimewa Yogyakarta.
2. Data yang digunakan adalah data ISPU Kota Yogyakarta pada tahun 2020 hingga 2021.
3. Analisis diterapkan dengan pemanfaatan aplikasi *data mining* WEKA.
4. Metode analisis menggunakan pemodelan algoritme *Random Forest* dan *K-Nearest Neighbors*.

## **D. Tujuan Penelitian**

Berdasarkan latar belakang dan rumusan masalah yang telah disampaikan di atas, adapun tujuan dari penelitian ini adalah:

1. Dapat membuat model dan pengklasifikasi kualitas udara di Daerah Istimewa Yogyakarta dengan algoritme *Random Forest* dan *K-Nearest Neighbors*.
2. Dapat mengukur performa model berdasarkan tingkat akurasi algoritme *Random Forest* dan *K-Nearest Neighbors*.

## **E. Metode Penelitian**

Pada penelitian ini metode yang diterapkan yaitu studi pustaka. Studi pustaka dilakukan dengan pengumpulan data dan sumber literatur yang mendukung topik penelitian. Sumber data diperoleh dari *website* bernama kaggle, sebuah platform *website* yang membantu *data scientist*, ilmuwan, dan peminat *machine learning* untuk mempublikasikan *dataset* mereka, mengasah skill analisis *data science* melalui kompetisi, sekaligus membentuk komunitas belajar.

## **F. Sistematika Penulisan**

Sistematika penulisan tugas akhir dapat diuraikan menjadi beberapa bagian yang tertulis secara sistemik, meliputi:

### **1. BAB I PENDAHULUAN**

Penulisan bab terdiri atas subbab yang ditulis secara urut berupa latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan. Latar belakang penelitian menjabarkan mengenai konteks dan kondisi yang mendorong dilakukannya penelitian. Rumusan masalah dan batasan masalah menjelaskan tentang masalah yang hendak diangkat dalam penelitian beserta batasan-batasannya dalam ruang lingkup penelitian. Tujuan penelitian menjabarkan tentang tujuan yang hendak dicapai dalam penelitian dan metode penelitian menjabarkan tentang metode yang digunakan pada penelitian.

### **2. BAB II TINJAUAN PUSTAKA**

Penulisan bab ini berisi penelitian-penelitian terdahulu yang pernah dilakukan dan menyerupai topik penelitian yang dilakukan. Selain itu, terdapat

tabel perbandingan antara penelitian yang dilakukan sekarang dengan penelitian terdahulu.

### 3. BAB III LANDASAN TEORI

Penulisan bab berisi mengenai berbagai teori yang berhubungan dengan topik penelitian yang dibahas. Teori tersebut memuat pengertian beserta definisi konsep yang berkaitan dengan penelitian yang dilakukan. Seluruh teori yang berkaitan dengan aplikasi pengembangan dan implementasi tidak dimasukkan pada bagian ini.

### 4. BAB IV DATASET DAN PENGEMBANGAN MODEL

Penulisan bab ini menjelaskan gambaran umum tentang masalah yang hendak diselesaikan dengan Machine Learning. Bab ini terdiri atas subbab deskripsi masalah, dataset, dan pengembangan model.

### 5. BAB V HASIL PENELITIAN DAN PEMBAHASAN

Penulisan bab ini menjabarkan tentang deskripsi penelitian, hasil penelitian *Random Forest* dan KNN beserta pembahasan hasil penelitian. Pembahasan juga mencakup penjelasan mengenai kesesuaian antara hasil percobaan dengan tujuan percobaan.

### 6. BAB VI PENUTUP

Penulisan bab ini merupakan kesimpulan dari hasil penelitian yang telah dilakukan. Kesimpulan pun diambil berdasarkan hasil yang diperoleh dan disesuaikan dengan tujuan penelitian. Selain itu, bab ini juga berisi berbagai saran yang berguna untuk penelitian kedepannya.