

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Penurunan pencemaran udara di DKI Jakarta disebabkan oleh berkurangnya mobilisasi masyarakat dan transportasi kendaraan akibat pembatasan sosial berskala besar (PSBB) di wilayah Jabodetabek (Rushayati et al., 2020). Informasi yang didapatkan dari *website* resmi Berita Jakarta bahwa pencemaran udara di DKI Jakarta menurun pada saat masa pandemi COVID-19 (Awaluddin et al., 2020).

Perlu dilakukan klasifikasi untuk menghasilkan akurasi dengan proses perhitungan kualitas udara yang dihitung berdasarkan Indeks Standar Pencemaran Udara (ISPU). Terdapat lima parameter pencemar udara yang telah ditetapkan oleh ISPU, yaitu Tingkat Partikulat (PM10), Sulfur Dioksida (SO<sub>2</sub>), Karbon Monoksida (CO), Ozon (O<sub>3</sub>) dan Nitrogen Dioksida (NO<sub>2</sub>) (Lee et al., 2018).

Setelah proses perhitungan nilai ISPU, diperlukan perhitungan sistematis yang membutuhkan model *Machine Learning* atau algoritma yang dapat digunakan ke dalam data ISPU. *Machine Learning* diterapkan untuk memproses data ISPU. Proses klasifikasi ISPU pada penelitian ini membandingkan hasil akurasi algoritma *Decision Tree* dan *K-Nearest Neighbor*.

Algoritma *Machine Learning* telah banyak diterapkan untuk mengklasifikasi polusi udara, salah satunya pada penelitian (Nurjanah et al., 2020) dilakukan klasifikasi

kualitas udara di DKI Jakarta menggunakan algoritma *K-Nearest Neighbor*. Penerapan algoritma *K-Nearest Neighbor* pada klasifikasi ISPU Jakarta menggunakan data Indeks Standar Pencemar Udara (ISPU) tahun 2018, menunjukkan hasil klasifikasi dengan akurasi 95.78% dengan menunjukkan bahwa kualitas udara di bulan berikutnya adalah “Sedang”.

Pada tahun 2020, (Putra & Sitanggang, 2020) menerapkan algoritma *Decision Tree* C5.0 dan *Random Forest* pada prediksi dataset Indeks Standar Pencemar Udara (ISPU) di Jakarta pada tahun 2017. Hasil akurasi prediksi yaitu 99.22% dan 97.42% menggunakan *Decision Tree* C5.0 dan *Random Forest*. Kemudian dilakukan penelitian perbandingan algoritma oleh (Atmakuri & Prasad, 2020) tentang komparasi algoritma *Optimized Bayesian Network*, *Decision Tree* dan *Random Forest* untuk memprediksi perbedaan pada level *Air Quality Index*. Hasil tersebut membuktikan bahwa *Optimized Bayesian Network* memiliki akurasi terbaik yaitu 99.63% sementara *Decision Tree* dan *Random Forest* memiliki akurasi sebesar 81.95% dan 99.39%.

Algoritma *Decision tree* merupakan algoritma yang ada pada teknik klasifikasi dalam *Machine Learning*. Algoritma ini menciptakan pohon keputusan yang kemudian diterapkan ke set pengujian untuk menetapkan kelas (Yunita, 2017). Sementara algoritma *K-Nearest Neighbor* menentukan data latih dan kemudian memprediksi kelas dari nilai baru dengan mencari K observasi dalam set pelatihan yang paling dekat dengan nilai baru, metode ini populer dan paling banyak digunakan dalam penyelesaian berbagai kasus klasifikasi (Nurmahaludin & Cahyono, 2019). Karena mudah

dieksploitasi, generalisasi yang baik, mudah, kemampuan beradaptasi ke ruang fitur yang rumit, fleksibel, mudah diterapkan, sederhana dan memiliki hasil yang akurat (Yuliska & Syaliman, 2020).

Penelitian ini berfokus pada perbandingan hasil akurasi pada algoritma *Decision Tree* dan *K-Nearest Neighbor* masih harus dilakukan peningkatan dengan menambahkan lebih banyak data dan atribut yang digunakan hal tersebut berpengaruh pada hasil akurasi pada setiap algoritma. Perbandingan akurasi algoritma *Decision Tree* dan *K-Nearest Neighbor* untuk klasifikasi data ISPU. Hasil akurasi klasifikasi kedua metode tersebut dibandingkan sehingga dapat diketahui algoritma yang paling baik dalam melakukan proses klasifikasi.

Hasil akurasi dari proses klasifikasi kedua algoritma akan dilakukan validasi menggunakan *K-Fold Cross Validation*. Metode validasi *K-Fold Cross Validation* digunakan untuk mempekirakan model pembelajaran mesin pada data yang tidak terlihat. Sehingga algoritma *Decision Tree* dan *K-Nearest Neighbor* dapat mencapai hasil akurasi yang lebih maksimal.

## 1.2 Rumusan Masalah

Rumusan masalah dari latar belakang tersebut yaitu:

1. Bagaimana proses perbandingan algoritma klasifikasi *Decision Tree* dan *K-Nearest Neighbor* pada data ISPU di DKI Jakarta?

2. Sejauh mana perbandingan hasil akurasi yang dihasilkan algoritma *Decision Tree* dan *K-Nearest Neighbor* dalam klasifikasi kualitas ISPU di DKI Jakarta?

### 1.3 Batasan Masalah

Batasan masalah yang diambil yaitu:

1. Menggunakan data Indeks Standar Pencemar Udara (ISPU) di DKI Jakarta sebelum masa pandemi COVID-19 tahun 2017-2021.
2. Parameter data yang akan diproses untuk menghasilkan klasifikasi udara adalah Tingkat Partikulat (PM10), Sulfur Dioksida (SO<sub>2</sub>), Ozon (O<sub>3</sub>), Karbon Dioksida (CO<sub>2</sub>) dan Nitrogen Dioksida (NO<sub>2</sub>).
3. Menggunakan 10 atribut data pada klasifikasi Indeks Standar Pencemar Udara (ISPU).
4. Pembuatan model *Machine Learning* menggunakan bahasa pemrograman *python* dan beberapa *library* pada *python*.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Melakukan proses perbandingan hasil akurasi klasifikasi pada data Indeks Standar Pencemar Udara (ISPU) di DKI Jakarta tahun 2017-2021 menggunakan metode *Decision Tree* dan *K-Nearest Neighbor*.

2. Mendapatkan hasil perbandingan akurasi dari algoritma klasifikasi *Decision Tree* dan *K-Nearest Neighbor* pada data ISPU di DKI Jakarta berupa tingkat akurasi serta menentukan algoritma yang menghasilkan akurasi terbaik.

## **1.5 Manfaat Penelitian**

Manfaat penelitian ini adalah:

1. Mengetahui proses perbandingan algoritma klasifikasi *Decision Tree* dan *K-Nearest Neighbor* pada data ISPU di DKI Jakarta.
2. Mengetahui hasil perbandingan dari algoritma *Decision Tree* dan *K-Nearest Neighbor* dalam klasifikasi kualitas udara Indeks Standar Pencemar Udara (ISPU) berupa hasil akurasi.

## **1.6 Metodologi Penelitian**

Metodologi penelitian meliputi beberapa tahapan yang dilakukan dalam proses penelitian yaitu:

### **1.6.1 Pengumpulan Data**

Merupakan tahap pengumpulan data untuk mencari bahan dasar, yaitu mengumpulkan data untuk tahap *preprocessing*.

### **1.6.2 Preprocessing**

Tahap *preprocessing* dilakukan sebelum proses pengklasifikasian data. Proses ini dilakukan dengan beberapa tahap pengolahan awal data sehingga mendapatkan data yang berkualitas.

### **1.6.3 Proses Klasifikasi**

Merupakan tahapan klasifikasi data yang sudah dilakukan *preprocessing*. Kemudian mengasilkan nilai akurasi dari dua algoritma.

### **1.6.4 Evaluasi Model**

Merupakan tahapan evaluasi hasil akurasi klasifikasi data yang sudah diproses. Kemudian membandingkan nilai akurasi dari dua algoritma.

### **1.6.5 Kesimpulan dan Saran**

Dilakukan pengambilan kesimpulan yang didapat dari hasil pengujian dan adanya saran untuk pengembangan dari penelitian lebih lanjut.

## **1.7 Sistematika Penulisan**

Sistematika penulisan pada penyusunan penelitian ini dibagi menjadi beberapa bab, yaitu:

### **BAB I PENDAHULUAN**

Bab ini menjelaskan permasalahan umum yang meliputi latar belakang masalah, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metode penelitian dan sistematika penulisan.

## BAB II LANDASAN TEORI

Bab ini menjelaskan tentang pembahasan teoritis yang berhubungan dengan bahan penelitian. Meliputi teori pencemaran udara, teori kualitas udara ISPU, teori *Machine Learning*, teori *Decision Tree*, teori *K-Nearest Neighbor* serta teori penelitian terkait lainnya yang berhubungan dengan penelitian.

## BAB III METODOLOGI

Bab ini menjelaskan tentang metode yang dipilih dan digunakan yang terdiri dari beberapa tahapan yaitu pengumpulan data, *preprocessing*, proses klasifikasi, evaluasi dan menjelaskan langkah - langkah lain dari penelitian.

## BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas tentang hasil dari penelitian yang dilakukan. Meliputi hasil akurasi kasifikasi data ISPU menggunakan algortima *Decision Tree* dan *K-Nearest Neighbor*.

## BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bab terakhir dalam penulisan laporan penelitian yang berisi kesimpulan dan saran. Pada kesimpulan memuat rangkuman dari hasil pembahasan, sementara hal-hal yang perlu diperhatikan berdasarkan kekurangan untuk penelitian selanjutnya dituangkan ke dalam bentuk saran.

