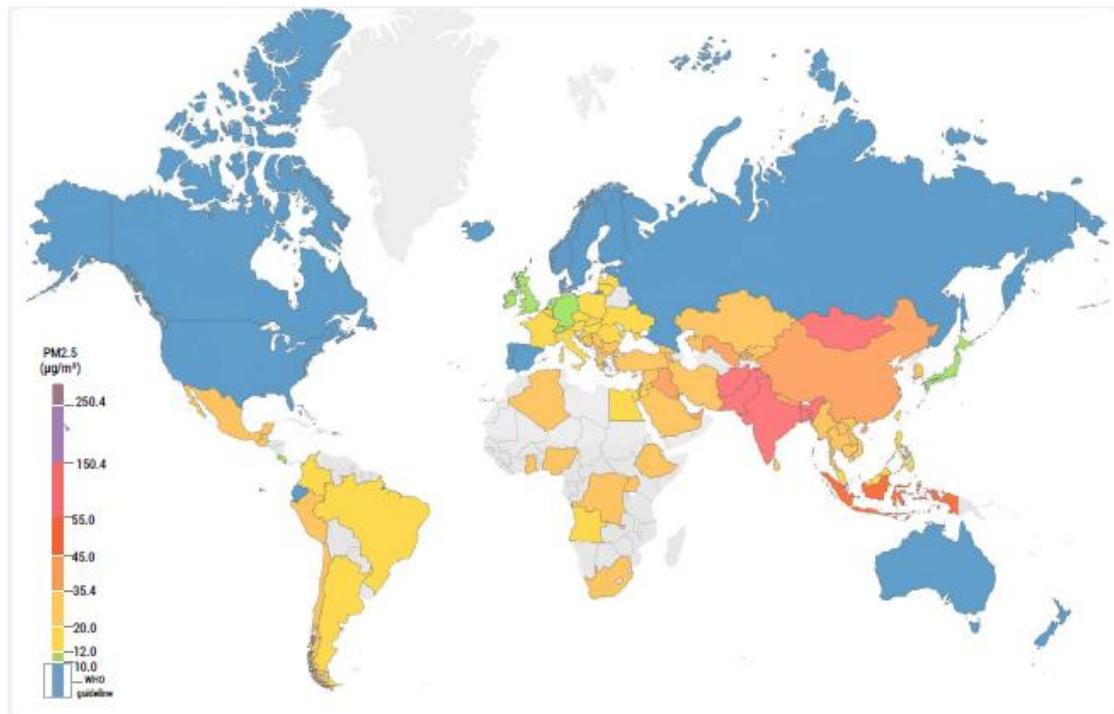


# BAB I PENDAHULUAN

## I.1 Latar Belakang

Udara penting bagi makhluk hidup terutama manusia, karena manusia bernapas dengan menghirup udara. Udara yang baik untuk makhluk hidup merupakan udara bersih yang tidak berwarna, tidak memiliki bau serta segar dan sejuk untuk dihirup. Manfaat menghirup udara bersih adalah dapat menghilangkan stres dan mencegah berbagai penyakit (Rizi et al., 2019).



Gambar I.1 Peta Kualitas Pencemaran Udara dunia tahun 2019

Sumber: (IQAir, 2019)

Seperti yang terlihat Gambar I.1 kualitas udara masih menjadi sebuah masalah di banyak negara karena polusi yang menyebabkan kualitas udara menjadi buruk. Bisa dilihat pada gambar I.1 yang daerah yang ditandai dengan warna merah menunjukkan tidak sehat, oranye menunjukkan hampir tidak sehat, kuning menunjukkan kualitas udara yang sedang, hijau menunjukkan kualitas udara baik, biru memiliki kualitas udara sangat sehat (IQAir, 2019).

Buruknya kualitas udara disebabkan oleh pencemaran udara contoh nya terdapat di negara Indonesia terutama pada provinsi DKI Jakarta. Pencemaran udara di DKI Jakarta disebabkan oleh beberapa faktor diantaranya kendaraan bermotor, Industri, Pembakaran batu bara, debu jalanan, dan debu aktivitas konstruksi. Namun yang menjadi penyebab utama adalah naiknya kebutuhan kendaraan yang menyumbang 31 sampai 40% pencemaran udara pada provinsi DKI Jakarta.

Tabel I-1 Pertumbuhan Kendaraan DKI Jakarta tahun 2019 sampai 2022

Sumber: (BPS Provinsi DKI Jakarta, n.d.)

Jenis Kendaraan	2019	2020	2021	2022
Mobil penumpang	3.310.426	3.365.467	3.544.491	3.776.059
Bus	34.905	35.266	36.339	37.180
Truk	669.724	679.708	713.059	748.395
Sepeda Motor	15.868.191	16.141.380	16.711.638	17.304.447
Jumlah	23.863.396	24.266.996	25.263.077	26.370.535

Terlihat pada tabel bahwa pada setiap tahun kendaraan di Provinsi DKI Jakarta meningkat dari tahun ke tahun, yang mana itu akan memicu kenaikan pencemaran udara bagi masyarakat Provinsi DKI Jakarta itu sendiri. Pencemaran udara dapat menyebabkan konsekuensi yang luas, termasuk dampak negatif pada kesehatan makhluk hidup, rusaknya lingkungan, terbentuknya hujan asam. Kualitas udara yang tercemar berpengaruh pada kesehatan masyarakat dengan meningkatkan risiko timbulnya penyakit, seperti masalah pernafasan, paru-paru, jantung, bahkan bisa menimbulkan jenis kanker yang sangat berbahaya (Abidin et al., 2019).

Dari permasalahan yang ada peneliti akan melakukan identifikasi untuk memantau kualitas udara yang terdapat pada data pencemaran kualitas udara Provinsi DKI Jakarta yang akan menjadi acuan kepada pemerintah untuk membuat kebijakan dan pengelolaan yang baik terhadap kualitas udara di Provinsi DKI Jakarta.

Untuk mengetahui hasil identifikasi tersebut maka diperlukan suatu metode yaitu *data mining*. Penggunaan *data mining* ini membantu untuk menemukan informasi yang memungkinkan untuk mendapatkan pola dan hubungan yang dapat digunakan untuk menentukan kondisi kualitas udara di Provinsi DKI Jakarta (Firdaus, 2017). Teknik yang digunakan dalam *data mining* ini merupakan teknik klasifikasi. Klasifikasi sendiri merupakan teknik yang mengelompokkan data pada kelas yang sebelumnya sudah dilabeli, salah satu algoritma klasifikasi yaitu *K-Nearest Neighbor*.

*K-Nearest Neighbor* adalah algoritma yang mengklasifikasikan dekatnya jarak dari satu objek dengan objek lainnya. pada penelitian sebelumnya terdapat perbandingan *K-Nearest Neighbor* dan *Naïve Bayes*, dimana *naïve bayes* menghasilkan akurasi 91% sedangkan untuk *K-Nearest Neighbor* menghasilkan akurasi 97% pada ketetanggaan  $K = 7$  (Sodiq & Sela, 2019). Pada penelitian sebelumnya terdapat perbandingan algoritma *K-Nearest Neighbor* dan *SVM* dalam klasifikasi jenis – jenis sapi dimana menghasilkan nilai akurasi *K-Nearest Neighbor* sebesar 100% dengan waktu pengerjaan selama 0,967 detik, sedangkan nilai akurasi algoritma *SVM* sebesar 80,3% dengan waktu pengerjaan komputasinya selama 1,645 detik (Wijaya et al., 2022). Pada penelitian sebelumnya tentang klasifikasi *K-Nearest Neighbor* terhadap masyarakat pra sejahtera Desa Sapikerep Kecamatan Sukapura mendapatkan nilai akurasi 97,36% pada ketetanggaan  $K = 3$ , nilai akurasi 98,68% didapat ketetanggaan  $K = 5$  dan 7 (Khairi, 2021).

## **I.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah disusun, rumusan masalah yang akan menjadi fokus penelitian dalam tugas akhir ini adalah sebagai berikut :

1. Bagaimana mengimplementasikan Algoritma *K-Nearest Neighbor* dalam klasifikasi Indeks Standar Pencemaran Udara DKI Jakarta
2. Bagaimana tingkat akurasi dari pengimplementasian *data mining* menggunakan algoritma *K-Nearest Neighbor* pada klasifikasi kualitas udara data ISPU di Provinsi DKI Jakarta?

### **I.3 Tujuan Tugas Akhir**

Berdasarkan rumusan masalah yang telah diajukan, penulis menetapkan tujuan penelitian yang akan dicapai sebagai berikut :

1. Dapat mengimplementasikan algoritma *K-Nearest Neighbor* pada analisis klasifikasi kualitas pencemaran udara Provinsi DKI Jakarta.
2. Dapat mengetahui tingkat akurasi algoritma *K-Nearest Neighbor* dalam melakukan klasifikasi data kualitas udara Provinsi DKI Jakarta.

### **I.4 Manfaat Tugas Akhir**

Adapun manfaat penelitian Tugas Akhir ini adalah sebagai berikut :

1. Bagi Penulis:
  - a. Dapat menerapkan Teknik *data mining* menggunakan algoritma *K-Nearest Neighbor* dalam klasifikasi kualitas udara pada provinsi DKI Jakarta.
  - b. Dapat mengetahui hasil dari implementasi algoritma *K-Nearest Neighbor* dalam klasifikasi kualitas udara Provinsi DKI Jakarta.
  - c. Dapat mengetahui hasil akurasi dari implementasi algoritma *K-Nearest Neighbor* pada klasifikasi kualitas udara Provinsi DKI Jakarta.
2. Pada Institusi :

Dapat menjadi bahan ajar untuk penelitian selanjutnya yang memiliki topik klasifikasi maupun algoritma yang serupa dengan penelitian ini.
3. Pada Masyarakat :

Dapat memberikan informasi kepada masyarakat tentang kualitas udara Provinsi DKI Jakarta sehingga masyarakat juga bisa membantu dalam penanggulangan pencemaran udara.

### **I.5 Batasan Masalah**

Penelitian Tugas Akhir memiliki batasan masalah yang disusun agar membatasi masalah apa saja yang akan dibahas agar tidak melebar dan keluar dari tujuan yang ada, batasan masalah penelitian ini mencakup:

1. Penelitian berfokus pada data kualitas udara yang terdapat pada Provinsi DKI Jakarta.
2. Data yang digunakan diambil dari sumber *website Jakarta Open Data (data.jakarta.go.id)*.
3. Data yang akan digunakan mengacu pada lima stasiun pencatatan kualitas udara di Provinsi DKI Jakarta, yaitu Stasiun 1 Bundaran HI, Stasiun 2 Kelapa Gading, Stasiun 3 Jagakarsa, Stasiun 4 Lubang Buaya, dan Stasiun 5 Kebon Jeruk.
4. Data yang akan digunakan dalam penelitian ini mencakup periode dari tahun 2019 hingga tahun 2022. Parameter yang akan dimanfaatkan dalam analisis meliputi PM10(Partikulat Matter), SO<sub>2</sub>(Sulfur Dioksida), CO(Karbon Monoksida), O<sub>3</sub>(Ozon), dan NO<sub>2</sub>(Nitrogen Oksida).

## **I.6 Sistematika Penulisan Tugas Akhir**

Penulisan penelitian ini diuraikan melalui sistematika penulisan tugas akhir sebagai berikut:

### **Bab I      Pendahuluan**

Bab ini berisi penjelasan mengenai latar belakang penelitian, perumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian, dan sistematika penulisan penelitian.

### **Bab II     Tinjauan Pustaka**

Bab ini berisi penjelasan mengenai teori yang sesuai dengan permasalahan yang dirumuskan seperti *data mining*, klasifikasi, *K-Nearest Neighbor* dan uji evaluasi algoritma menggunakan *confusion matrix*, *K fold Cross Validation*, dan *GridSearchCv*.

### **Bab III    Metodologi Penelitian**

Bab ini berisi model kerangka pemecahan masalah untuk merumuskan solusi dan tahapan penyelesaian dari permasalahan yang dibuat secara rinci.

**Bab IV Perancangan Sistem**

Bab ini berisi penjelasan tahapan pengerjaan studi kasus seperti pengumpulan data, *pre-processing* data, pemodelan *K-Nearest Neighbor*, evaluasi algoritma.

**Bab V Hasil dan Evaluasi**

Bab ini berisi penjelasan setiap temuan, hasil dan evaluasi yang diperoleh dari penelitian.

**Bab VI Kesimpulan dan Saran**

Bab ini berisi kesimpulan yang diperoleh ketika penelitian dan saran untuk penelitian selanjutnya.