

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Udara memiliki peran yang sangat penting dalam mendukung kehidupan di bumi. Selain menyediakan oksigen, udara juga berfungsi sebagai media untuk menyampaikan suara, sebagai pendingin bagi objek yang panas, bahkan dapat menjadi penyebab penyebaran penyakit pada manusia, hewan, dan tumbuhan [1]. Dalam menjaga kesehatan makhluk hidup khususnya manusia, Kualitas udara memiliki peran penting untuk menjadi perhatian jangka panjang [2]. Menurut data dari WHO, lebih dari 90% penduduk dunia tinggal dan menghirup udara yang berbahaya. Hal ini menjadi penting ketika melihat statistik yang menunjukkan adanya 4,2 juta kematian dini di dunia pada tahun 2016. Sebanyak 91% dari jumlah tersebut terjadi di negara-negara berpendapatan menengah ke bawah, termasuk kawasan Asia Tenggara dan Pasifik Barat, Salah satunya Indonesia yang memiliki tingkat pencemaran udara yang tinggi dibandingkan dengan negara-negara lain di kawasan Asia Tenggara [3].

Pencemaran udara merupakan isu sosial yang sering terjadi, baik di sektor industri, kendaraan bermotor, maupun kejadian seperti kebakaran hutan. Situasi ini dapat menimbulkan sejumlah masalah bagi makhluk hidup di sekitarnya, karena polusi tersebut mengandung zat-zat berbahaya. Dampak dari polusi udara termasuk timbulnya Infeksi Saluran Pernapasan Akut (ISPA), bahkan dapat berujung pada kematian pada manusia [4]. Pada tahun 2019, Provinsi DKI Jakarta berada di urutan ketiga dari sepuluh kota paling terpapar polusi di Indonesia [5]. Oleh karena itu, Pemerintah Provinsi DKI Jakarta secara terus-menerus melakukan pemantauan kualitas udara di lingkungan sekitar dan telah menetapkan Indeks Standar Pencemar Udara (ISPU) sesuai dengan ketentuan yang diatur dalam Peraturan Pemerintah (PP) Nomor 41 Tahun 1999.

Indeks Standar Pencemar Udara (ISPU) merupakan ukuran yang menilai tingkat pencemaran udara akibat adanya bahan kimia dan partikel di dalamnya. Parameter yang digunakan untuk mengevaluasi kualitas udara termasuk PM10, SO<sub>2</sub>, CO, O<sub>3</sub>,

dan NO<sub>2</sub> [6]. Dengan merujuk pada data Indeks Standar Pencemar Udara (ISPU) yang diperoleh dari Stasiun Pemantau Kualitas Udara (SPKU) Dinas Lingkungan Hidup DKI Jakarta pada tahun 2019, dapat diamati bahwa DKI Jakarta mengalami 2 hari (1%) dengan kualitas udara baik, 172 hari (48%) dengan kualitas udara sedang, 183 hari (50%) dengan kualitas udara tidak sehat, dan 8 hari (2%) dengan kualitas udara sangat tidak sehat [7].

Berdasarkan data Indeks Standar Pencemar Udara (ISPU), Faktor utama peningkatan kualitas udara disebabkan oleh pencemaran udara. Pencemaran udara di Jakarta disebabkan oleh beberapa faktor, termasuk polusi dari kendaraan bermotor, pertumbuhan jumlah pabrik di kawasan industri yang mengeluarkan polusi udara ke lingkungan sekitarnya, dan pembakaran sampah yang tidak terkendali sehingga menyebabkan udara tercemar. Dalam mengatasi masalah polusi udara, pentingnya kerja sama antara pemerintah dan masyarakat. Program-program pencegahan polusi udara harus dirancang dengan memanfaatkan teknologi yang sudah tersedia, seperti penggunaan algoritma *machine learning*. Diharapkan *machine learning* dapat memberikan informasi tentang tingkat kualitas udara di Jakarta [8].

Dalam mengetahui tingkat kualitas udara di Jakarta dilakukan dengan klasifikasi menggunakan 6 algoritma yaitu KNN, *Decision Tree*, *Naïve Bayes*, SVM, *Random Forest* dan XGBOOST tanpa dan dengan menggunakan optimasi, dan 5 algoritma hibrida yaitu KNN-XGBoost, *Decision Tree*-XGBoost, *Naïve Bayes*-XGBoost, SVM-XGBoost, dan *Random Forest*-XGBoost. Dengan menggunakan klasifikasi yang berarti suatu metode di mana data atau objek baru dikelompokkan ke dalam kelas atau label tertentu berdasarkan karakteristik atau variabel khusus yang dimiliki [9].

Penelitian sebelumnya telah dilakukan oleh Umri, Syekh S A Firdaus, Muhammad S Firdaus, Primajaya Aji pada tahun 2021 yang melakukan perbandingan kinerja lima algoritma, termasuk *Naïve Bayes*, SVM, *Decision Tree*, KNN, dan *Neural Network Backpropagation*. Metode validasi yang digunakan adalah 10-fold *cross-validation*, dan seleksi fitur dilakukan dengan *backward elimination*. Hasil penelitian menunjukkan bahwa algoritma dengan kinerja terbaik adalah *Decision Tree*, mencapai tingkat akurasi sebesar 99,80% [10].

Penelitian selanjutnya dilakukan oleh Suci Cahaya Hati Nasution, Fibri Rakhmawati, Riri Syafitri Lubis pada tahun 2021 mengenai klasifikasi tingkat pencemaran udara pada sektor industri menggunakan algoritma *Random Forest*. Klasifikasi dalam penelitian ini dikelompokkan menjadi tiga tingkatan, yaitu rendah, menengah, dan tinggi, dengan total 500 pohon yang diamati. Hasil akurasi yang diperoleh oleh algoritma *Random Forest* sebesar 95% dan prediksi model dinyatakan benar[11].

Penelitian selanjutnya juga dilakukan oleh Sri Erlina Yulianti, Oni Soesanto, dan Yuana Sukmawaty ,pada tahun 2022 mengenai klasifikasi nasabah kartu kredit dengan menggunakan algoritma XGBoost. Pada penelitian ini menggunakan *hyperparameter* yang kemudian divalidasi dengan menggunakan *10-Fold Cross Validation*. Hasil akurasi yang diperoleh oleh algoritma XGBoost sebelum menggunakan *hyperparameter* adalah 80,02%, untuk presisi 85,32%, dan *recall* 94,86%. Sedangkan sesudah menggunakan *hyperparameter* hasil akurasi nya mencapai 80,039%, presisi mencapai 81,338%, dan *recall* mencapai 96,854% [12].

Penelitian ini menerapkan *machine learning* dengan metode KDD dan menggunakan klasifikasi untuk klasifikasi tingkat kualitas udara di Jakarta. Pada penelitian ini *machine learning* dipilih karena *machine learning* diperkenalkan untuk meningkatkan kemampuan deteksi otomatis. Metode ini bekerja secara otomatis dan menawarkan model klasifikasi yang efisien dan efektif dengan menggabungkan metode matematika dan teknik pencarian dari ilmu komputer [13]. Sedangkan untuk penggunaan metodologi KDD, dalam penelitian ini KDD dipilih untuk menyajikan kerangka kerja yang terstruktur untuk memperoleh pengetahuan baru dari data kualitas udara[14]. Urgensi melakukan klasifikasi terletak pada proses untuk menemukan kesamaan karakteristik dalam suatu kelompok atau kelas[15]. Dalam penelitian ini , terdapat 6 algoritma yang digunakan, seperti KNN, *Naïve Bayes*, SVM, *Decision Tree*, *Random Forest*, dan XGBoost. Penelitian ini dilakukan dengan tujuan untuk mengembangkan metode klasifikasi tingkat kualitas udara di Jakarta dengan menggunakan algoritma *machine learning* berdasarkan fenomena yang telah diuraikan.

Mengacu pada penelitian [10], perbandingan dilakukan dengan menggunakan algoritma *Naïve Bayes*, SVM, *Decision Tree*, KNN, dan *Neural*

*Network Backpropagation*. Hasil penelitian menunjukkan bahwa *Decision Tree*, memiliki akurasi terbaik sebesar 99,80%. Mengacu pada penelitian [11], mengenai klasifikasi tingkat pencemaran udara pada sektor industri menggunakan algoritma *Random Forest*. Hasil menunjukkan bahwa algoritma *Random Forest* dapat memprediksi dengan benar karena memiliki akurasi sebesar 95%, dan mengacu pada penelitian [12] mengenai klasifikasi nasabah kartu kredit dengan menggunakan algoritma *XGBoost* dan penggunaan *hyperparameter*, bahwa penggunaan algoritma *XGBoost* memiliki akurasi baik, dan dengan menggunakan optimasi *hyperparameter* dapat meningkatkan performa *XGBoost* dengan akurasi sebesar 80,039%. Oleh karena itu, pemilihan 6 algoritma tersebut bertujuan untuk pelengkapan eksplorasi penelitian dan menjadi sebagai pembeda dari penelitian terdahulu.

Pembeda dari penelitian sebelumnya, penelitian ini menggunakan dataset Indeks Standar Pencemar Udara di SPKU DKI Jakarta pada tahun 2013-2023. Dataset diambil dari *website* Satu Data Jakarta, kemudian dilakukan dengan membandingkan kinerja 22 model, 22 model tersebut terdiri dari 6 algoritma *machine learning* yaitu *KNN*, *Decision Tree*, *Naïve Bayes*, *SVM*, *Random Forest* dan *XGBoost*, yang diuji dengan menggunakan optimasi dan tanpa optimasi. Sementara 5 algoritma hibrida *machine learning* yaitu *KNN-XGBoost*, *Decision Tree-XGBoost*, *Naïve Bayes-XGBoost*, *SVM-XGBoost*, dan *Random Forest-XGBoost* juga diuji menggunakan optimasi dan tanpa optimasi. Optimasi yang digunakan pada penelitian ini adalah penggunaan *hyperparameter tuning* dengan *library Gridsearch*, dan penggunaan *feature selection* dilakukan menggunakan *SelectKBest*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, dapat disimpulkan suatu rumusan masalah. Rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana hasil performa 6 algoritma *Machine Learning* tanpa menggunakan optimasi dalam klasifikasi tingkat kualitas udara di Jakarta?

2. Bagaimana hasil performa 6 algoritma *Machine Learning* menggunakan optimasi dalam klasifikasi tingkat kualitas udara di Jakarta?
3. Bagaimana hasil performa 5 algoritma Hibrida *Machine Learning* tanpa menggunakan optimasi dalam klasifikasi tingkat kualitas udara di Jakarta?
4. Bagaimana hasil performa 5 algoritma Hibrida *Machine Learning* menggunakan optimasi dalam klasifikasi tingkat kualitas udara di Jakarta?
5. Bagaimana tingkat kualitas udara di Jakarta berdasarkan data Indeks Standar Pencemaran udara Jakarta?

### 1.3 Batasan Masalah

Pembatasan masalah yang dikaji sangat diperlukan agar penelitian ini tidak terlalu luas ulasannya dan tidak menyimpang dari rumusan masalah di atas. Berikut adalah batasan masalah dalam penelitian ini :

1. Penelitian ini menggunakan dataset Indeks Standar Pencemaran Udara di Jakarta yang mencakup periode dari bulan Januari tahun 2013 sampai dengan bulan Desember tahun 2023.
2. Penelitian ini hanya berfokus pada klasifikasi tingkat kualitas udara di Jakarta menggunakan *Knowledge Discovery in Database (KDD)* sebagai metodologi penelitian.
3. Penelitian ini menggunakan 6 algoritma KNN, *Decision Tree*, *Naïve Bayes*, SVM, *Random Forest* dan XGBoost dengan optimasi dan tanpa optimasi, dan juga melibatkan 5 algoritma hibrida seperti KNN-XGB, *Naïve Bayes-XGBoost*, SVM-XGBoost, *Decision Tree-XGBoost*, *Random Forest-XGBoost* .
4. Penelitian ini hanya melakukan perbandingan berdasarkan tingkat akurasi, presisi, *recall* dan *f1- score* pada algoritma KNN, *Decision Tree*, *Naïve Bayes*, SVM , *Random Forest* dan XGBoost. dengan optimasi dan tanpa optimasi, dan juga melakukan perbandingan untuk algoritma hibrida seperti KNN-XGB, *Naïve Bayes-XGBoost*, SVM-XGBoost, *Decision Tree-XGBoost*, *Random Forest-XGBoost* dengan optimasi dan tanpa optimasi.

## 1.4 Tujuan dan Manfaat Penelitian

### 1.4.1 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Melakukan evaluasi terhadap tingkat akurasi, presisi, *recall*, dan *f1-score* pada algoritma KNN, *Naïve Bayes*, SVM, *Decision Tree*, *Random Forest*, dan XGBoost tanpa menggunakan optimasi.
2. Melakukan evaluasi terhadap tingkat akurasi, presisi, *recall*, dan *f1-score* pada algoritma KNN, *Naïve Bayes*, SVM, *Decision Tree*, *Random Forest*, dan XGBoost, dengan menggunakan optimasi.
3. Melakukan evaluasi terhadap tingkat akurasi, presisi, *recall*, dan *f1-score* pada algoritma hibrida KNN-XGB, *Naïve Bayes*-XGBoost, SVM-XGBoost, *Decision Tree*-XGBoost, dan *Random Forest*-XGBoost, tanpa menggunakan optimasi.
4. Melakukan evaluasi terhadap tingkat akurasi, presisi, *recall*, dan *f1-score* pada algoritma hibrida KNN-XGB, *Naïve Bayes*-XGBoost, SVM-XGBoost, *Decision Tree*-XGBoost, dan *Random Forest*-XGBoost, dengan menggunakan optimasi.
5. Menentukan tingkat kualitas udara di Jakarta berdasarkan data Indeks Standar Pencemaran udara Jakarta

### 1.4.2 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah :

1. Memberikan wawasan tentang keakuratan algoritma KNN, *Decision Tree*, *Naïve Bayes*, SVM, *Random Forest* dan XGBoost dengan optimasi dan tanpa optimasi, dan juga melibatkan algoritma hibrida seperti KNN-XGB, *Naïve Bayes*-XGBoost, SVM-XGBoost, *Decision Tree*-XGBoost, *Random Forest*-XGBoost dalam memodelkan klasifikasi tingkat kualitas udara, yang dapat membantu dalam pengembangan penelitian di masa yang akan datang.

2. Pemerintah dapat merencanakan langkah-langkah pencegahan sebelum kualitas udara memburuk, sehingga dapat mengurangi dampak negatif dari pencemaran udara.
3. Memberikan kesadaran bagi masyarakat Jakarta untuk lebih bijak dalam mengurangi polusi udara.

### **1.5 Sistematika Penulisan**

Penelitian ini disusun dengan pembagian ke dalam beberapa bab sehingga memungkinkan pembaca memahami kontennya secara terstruktur dan jelas. Sistematika penulisan yang diterapkan dalam penelitian ini melibatkan langkah-langkah berikut :

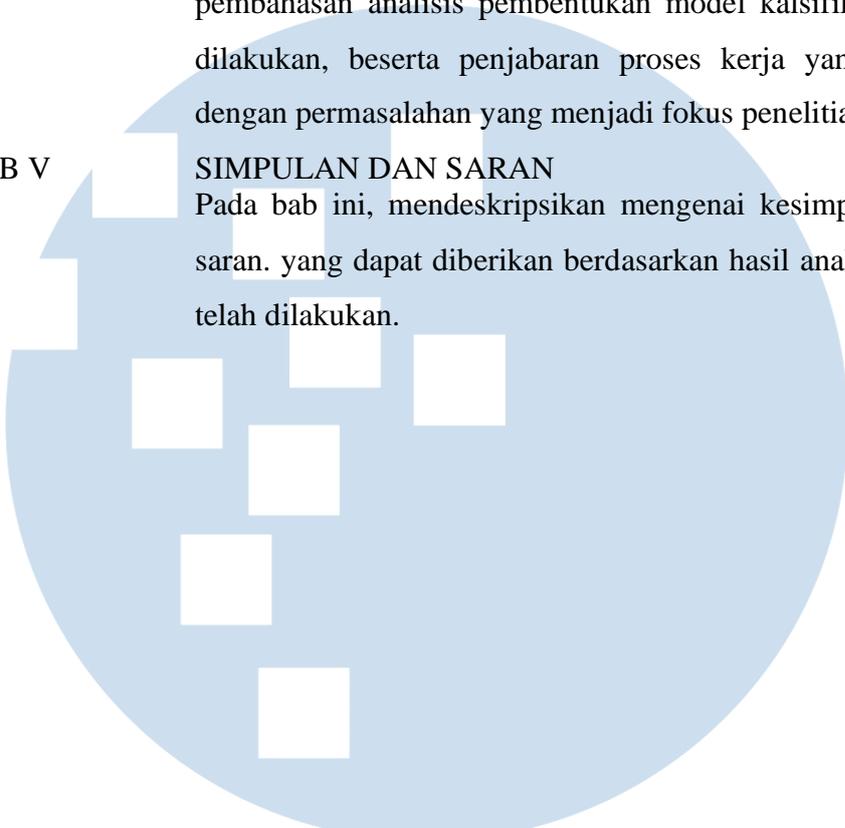
- BAB I                    PENDAHULUAN**  
Pada bab ini, mendeskripsikan uraian mengenai konteks permasalahan yang menjadi fokus penelitian, termasuk latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan penelitian.
- BAB II                    LANDASAN TEORI**  
Pada bab ini, mendeskripsikan mengenai rangkuman teori pendukung yang terkait dengan penelitian, meliputi konsep-konsep yang terkait dengan topik penelitian, kerangka kerja, algoritma, perangkat yang digunakan, serta tinjauan literatur mengenai penelitian sebelumnya. Menggunakan teori yang berasal dari kutipan artikel jurnal yang relevan dengan fokus penelitian.
- BAB III                    METODELOGI PENELITIAN**  
Pada bab ini, mendeskripsikan mengenai obyek penelitian, pendekatan metodologi yang diterapkan, variabel yang menjadi fokus penelitian, metode pengumpulan data, teknik pengambilan sampel, dan teknik analisis data yang diaplikasikan terhadap temuan hasil penelitian
- BAB IV                    ANALISIS DAN HASIL PENELITIAN**

Pada bab ini, mendeskripsikan mengenai hasil dan pembahasan analisis pembentukan model kalsifikasi yang dilakukan, beserta penjabaran proses kerja yang terkait dengan permasalahan yang menjadi fokus penelitian ini.

## BAB V

### SIMPULAN DAN SARAN

Pada bab ini, mendeskripsikan mengenai kesimpulan dan saran. yang dapat diberikan berdasarkan hasil analisis yang telah dilakukan.



# UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA