

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Polusi udara adalah masuknya berbagai zat secara berbahaya ke dalam atmosfer kita, baik yang disebabkan oleh tindakan manusia yang disengaja maupun tidak, atau yang timbul dari kejadian alam (Aryanta and Maharani, 2023). Menurut Organisasi Kesehatan Dunia (WHO), dampak buruk dari polusi udara mengakibatkan kematian dini pada 2 juta orang setiap tahunnya (Riyanto et al., 2023). Hanya 5% negara yang mampu mematuhi pedoman polusi udara yang ditetapkan oleh WHO (Iqair, 2023). Indonesia menempati peringkat ke-17 secara global dalam hal tingkat polusi udara dan peringkat ke-1 secara wilayah Asia Tenggara (Kompas, 2022). Berdasarkan *Air Quality Live Index* (AQLI) pada bulan April 2021, DKI Jakarta sebagai ibukota negara menempati posisi kota ke-6 di dunia dengan kualitas udara paling buruk (Amalia et al., 2022). Meningkatnya aktivitas manusia telah memicu kekhawatiran yang mendesak terhadap masalah polusi udara (Decy Arwini, 2020). Untuk menanggulangi masalah polusi udara yang hari demi hari terus memburuk, diperlukan adanya tindakan seperti menggunakan kendaraan umum sebagai alat transportasi sehari-hari, menerapkan konsep 3R yaitu *reduce, reuse, recycle*, memanfaatkan sumber energi yang berkelanjutan dan ramah lingkungan dan rutin mengecek kualitas standar udara (Therin and Santoso, 2021). Di Indonesia pengukuran standar kualitas udara yang resmi yaitu Indeks Standar Pencemar Udara (ISPU), hal ini sesuai dengan Keputusan Menteri Negara Lingkungan Hidup Nomor : KEP 45/MENLH/1997 Tentang Indeks Standar Pencemar Udara (ISPU) (Insani and Darlianti, 2019).

ISPU berfungsi sebagai tolok ukur utama untuk mengkategorikan dan menjelaskan kualitas udara yang didasarkan pada dampak besar terhadap kesejahteraan individu dan vitalitas semua makhluk hidup. Terdapat lima parameter dalam ISPU sebagai pengamatan kualitas udara seperti : Tingkat Partikulat (PM10), Oksida Nitrogen (NO<sub>2</sub>), Sulfur Dioksida (SO<sub>2</sub>), Karbon Monoksida (CO), dan ozon permukaan (O<sub>3</sub>) (Hidayatullah et al., 2022). Jika

senyawa-senyawa tersebut melampaui ambang batas yang dapat diterima, maka berpotensi membahayakan kesejahteraan seseorang, bahkan menyebabkan kematian, terutama pada sistem pernafasan (Fath, 2021). ISPU dibagi menjadi lima kategori berbeda, yaitu baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya (Aljuaid & Alwabel, 2019).

Penerapan klasifikasi ISPU memerlukan metode yang mampu menganalisis dan menguraikan pola data yang berasal dari sensor yang mengukur tingkat polutan udara. Metode umum yang sering digunakan adalah metode pembelajaran mesin (*machine learning*), sebuah bidang kecerdasan buatan yang memungkinkan komputer memperoleh pengetahuan dari data secara mandiri, sehingga tidak memerlukan pemrograman eksplisit (Sya'ban et al., 2022).

*Machine learning* yang digunakan pada penelitian ini adalah *eXtreme Gradient Boosting* (XGBoost). XGBoost adalah algoritma pembelajaran mesin yang terkenal digunakan dalam klasifikasi dan regresi. Algoritma ini mengadopsi pendekatan ensemble dari pohon keputusan, di mana sejumlah pohon keputusan dibangun secara berurutan untuk meningkatkan kinerja keseluruhan model. Algoritma pembelajaran mesin sering memerlukan metode atau teknik tambahan untuk meningkatkan performa secara keseluruhan. Teknik-teknik seperti penanganan data tidak seimbang dan pemilihan fitur dapat membantu mengatasi berbagai tantangan data dan memastikan bahwa model menghasilkan prediksi yang lebih tepat dan dapat diandalkan. Teknik yang digunakan pada penelitian tersebut adalah *Synthetic Minority Over-sampling Technique* (SMOTE).

SMOTE adalah teknik yang umum digunakan dalam pemrosesan data yang tidak seimbang (*imbalanced data*). Teknik ini bekerja dengan membuat sampel sintetis baru dalam kelas minoritas. Metode dan teknik tersebut perlu dilakukan perbandingan dengan metode dan teknik lainnya untuk mengetahui mana yang lebih baik dalam klasifikasi ISPU.

Pada tahun 2023 (Sapari et al., 2023) menerapkan algoritma XGBoost untuk klasifikasi kualitas udara. Parameter yang digunakan dalam klasifikasi tersebut antara lain *Particulate Matter 10* (PM10), *Particulate Matter 2.5* (PM2.5), *Sulfur Dioksida* (SO<sub>2</sub>), *Karbon Monoksida* (CO), *Ozon* (O<sub>3</sub>) dan *Nitrogen Dioksida* (NO<sub>2</sub>). Dalam penelitian tersebut didapatkan nilai *precision* sebesar 97%, nilai *recall* sebesar 100%, nilai *F-1 Score* sebesar 98% dan nilai

*accuracy* sebesar 98.61%.

Penelitian selanjutnya tentang model prediksi kepadatan lalu lintas : Perbandingan antara algoritma Random Forest dan XGBoost pernah dilakukan oleh (Lisanthoni et al., 2023). Data yang digunakan dari Metro Interstate Traffic Volume, yang disediakan oleh repositori data UCI. Hasil *accuracy* yang didapatkan untuk algoritma XGBoost sebesar 95.92%, sedangkan algoritma Random Forest sebesar 95.53% dengan algoritma XGBoost memiliki waktu pemrosesan lebih cepat 532% dibanding algoritma Random Forest.

Pada tahun 2023 (Rachmi, 2023) melakukan klasifikasi pencemaran udara di DKI Jakarta dengan algoritma Naïve Bayes. *Variable* yang digunakan adalah PM10, PM25, SO2, CO, O3, dan NO2 dengan data sebanyak 1826 dalam kurun waktu 12 bulan. Penelitian ini didapatkan hasil bahwa algoritma Naïve Baiyes untuk klasifikasi udara dengan nilai *accuracy* sebesar 91.96%.

Selanjutnya penelitian tentang penerapan teknik SMOTE pada klasifikasi objektivitas berita online dengan algoritma KNN (Kasanah et al., 2019). Nilai k tetangga yang digunakan bervariasi yaitu 1, 3, 5, 7 dan 9. Dari hasil pengujian teknik SMOTE dapat meningkatkan nilai *accuracy*. Dengan nilai *accuracy* k=1 meningkat 5,00 dan untuk nilai k=3 meningkat 1,72. SMOTE meningkatkan nilai *precision* untuk semua nilai k tetangga. Untuk nilai *recall* dan *F-measure* SMOTE dapat menurunkan nilai keduanya untuk semua nilai k tetangga.

Berdasarkan referensi penelitian-penelitian sebelumnya, menunjukkan bahwa algoritma XGBoost dapat melakukan klasifikasi dengan baik dan teknik SMOTE bisa meningkatkan nilai dengan signifikan. Algoritma XGBoost memang membutuhkan *preprocessing* yang tepat guna meningkatkan kualitas hasilnya. Selain itu, algoritma XGBoost merupakan metode klasifikasi *supervised*, yang berarti mereka digunakan untuk membangun model yang dapat memprediksi kelas atau label dari data baru berdasarkan contoh-contoh data dengan label yang diberikan dalam dataset pelatihan.

Berdasarkan penelitian-penelitian sebelumnya, peneliti ingin melakukan perbandingan algoritma *eXtreme Gradient Boosting* dengan teknik imbalance data pada klasifikasi indeks standar pencemaran udara (ISPU) dengan *Synthetic Minority Over-sampling Technique*. Data yang digunakan adalah ISPU DKI Jakarta tahun 2022-2023 yang berasal dari website Satu Data Jakarta:

<https://satudata.jakarta.go.id/home>. Perbandingan yang digunakan mempertimbangkan faktor-faktor penting seperti *accuracy*, *recall*, *f1-score*, *precision* dan *confusion matrix*. Pada tahap akhir dapat diketahui algoritma dan teknik *imbalance* data terbaik yang dapat mengklasifikasikan ISPU dengan baik. Penelitian ini dilakukan karena belum ada penelitian sebelumnya yang membandingkan teknik *imbalance* data dengan algoritma XGBoost menggunakan teknik SMOTE untuk klasifikasi indeks standar pencemaran udara. Oleh karena itu peneliti mengangkat penelitian ini dengan judul “Implementasi *Synthetic Minority Over-sampling Technique* (SMOTE) Pada Algoritma *eXtreme Gradient Boosting* (XGBoost) Untuk Klasifikasi Indeks Standar Pencemaran Udara (ISPU)”.

## 1.2. Rumusan Masalah

Dari uraian latar belakang yang telah disajikan, dapat diidentifikasi rumusan masalah yang menjadi fokus pembahasan sebagai berikut:

1. Bagaimana cara algoritma *eXtreme Gradient Boosting* dapat mengklasifikasikan indeks standar pencemaran udara?
2. Bagaimana implementasi *Synthetic Minority Over-sampling Technique* untuk meningkatkan hasil klasifikasi indeks standar pencemaran udara?
3. Bagaimana hasil klasifikasi yang dilakukan pada indeks standar pencemaran udara menggunakan algoritma XGBoost dengan SMOTE?

## 1.3. Tujuan Penelitian

Merujuk dari rumusan masalah diatas, adapun tujuan penelitian yang ingin dicapai peneliti dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui cara algoritma *eXtreme Gradient Boosting* dapat mengklasifikasikan indeks standar pencemaran udara.
2. Untuk mengetahui implementasi *Synthetic Minority Over-sampling Technique* untuk meningkatkan hasil klasifikasi indeks standar pencemaran udara.
3. Untuk mengetahui hasil klasifikasi yang dilakukan pada indeks standar pencemaran udara menggunakan algoritma XGBoost dengan SMOTE.

#### 1.4. Manfaat Penelitian

Manfaat yang bisa dihasilkan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan akurasi dari algoritma dan teknik *imbalance* data terbaik yang digunakan dalam penelitian ini.
2. Memberikan gambaran implementasi *Synthetic Minority Over-sampling Technique* untuk klasifikasi indeks standar pencemaran udara.
3. Penelitian ini bisa dijadikan rujukan untuk penelitian-penelitian selanjutnya dalam hal klasifikasi indeks standar pencemaran udara.

#### 1.5. Batasan Masalah

Dalam penelitian ini penting untuk menetapkan batasan yang jelas untuk memastikan hanya fokus pada isu-isu yang diteliti. Oleh karena itu, berikut batasan masalah yang diterapkan dalam penelitian ini.:

1. Data yang digunakan dalam penelitian ini merupakan data sekunder (data yang diambil secara tidak langsung oleh peneliti) yang diambil dari website Satu Data Jakarta.
2. Data yang digunakan adalah indeks standar pencemaran udara di DKI Jakarta tahun 2021-2023 dengan jumlah 4015 data.
3. Peneliti menggunakan parameter pencemaran udara *Particulate Matter 10* (PM10), *Particulate Matter 2.5* (PM2.5), *Sulfur Dioksida* (SO<sub>2</sub>), *Karbon Monoksida* (CO) dan *Nitrogen Dioksida* (NO<sub>2</sub>).
4. Pada penelitian klasifikasi indeks standar pencemaran udara hanya mempunyai 3 jenis kategori yaitu kategori tidak sehat, kategori sedang dan kategori baik.
5. Program ini menggunakan Bahasa Python dengan bantuan *library*.
6. Keluaran yang dihasilkan adalah keakurasian dari klasifikasi indeks standar pencemaran udara.