

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pencemaran udara saat ini menjadi salah satu masalah yang besar di Indonesia. Perlu disadari pencemaran udara itu juga disebabkan oleh 2 faktor, yaitu alam dan manusia. Tanpa disadari, manusia kurang memperhatikan tentang bagaimana kegiatan yang mereka lakukan itu akan berimbas ke lingkungan sekitar bahkan akan meramban ke diri mereka sendiri nantinya[1] . Pencemaran udara adalah masuknya zat-zat yang berbahaya ke dalam atmosfer dimana bisa berpengaruh pada kualitas lingkungan bahkan bisa menimbulkan banyak penyakit berbahaya [2]. Udara mempunyai peran yang sangat penting dalam kehidupan manusia, terutama kandungan oksigen yang dibutuhkan oleh manusia. Selain oksigen udara juga mempunyai polutan lain yaitu *partikulat (PM10)*, *sulfida (dalam bentuk SO₂)*, *karbon monoksida (CO)*, *ozon (O₃)*, *nitrit(NO₂)*, dan polutan lainnya[3].

Kota Surabaya merupakan salah satu kota yang memiliki jumlah penduduk yang relatif banyak dengan beragam aktifitas yang berlangsung didalamnya [4] . Oleh karena itu, banyak sekali tantangan terhadap lingkungan, salah satunya pencemaran udara. Masalah pencemaran udara di kota ini disebabkan oleh banyak faktor, misalnya asap kendaraan bermotor, aktivitas industri pabrik, dan juga proyek-proyek Pembangunan[3]. Tidak hanya masalah dalam kegiatan manusia, tapi juga letak geografis dan kondisi cuaca yang tergolong kering, apalagi saat musim kemarau, menyebabkan polutan udara menjadi bertahan lama di atmosfer, yang makin meningkatkan resiko pencemaran udara. Masalah ini harus segera diatasi oleh pemerintah kota Surabaya, dengan mencari Solusi terbaik untuk penanganannya. Tidak hanya pemerintah saja yang perlu bertindak, juga perlu kesadaran dari masyarakat juga, karena dampak pencemaran ini sangat berbahaya bagi kesehatan dan bisa mengganggu kegiatan dan aktivitas Masyarakat [5]. Penelitian ini akan dilakukan proses klasifikasi dengan menggunakan 5 kategori ISPU yang menjadi patokan utama dalam standart kualitas udara di Indonesia [6]. Proses klasifikasi merupakan proses dalam machine learning dimana melakukan identifikasi dan memberikan kategori kelas dalam data yang telah dilakukan selama pemrosesan[7]. Tujuan utama dari proses ini yaitu untuk mengembangkan model dengan melakukan pemberian kelas dari

setiap data secara akurat [6]. Dengan melakukan proses ini menggunakan metode yang mampu dan mendukung proses yang akurat diharapkan mampu memberikan hasil yang sesuai dengan alur penelitian. *XGBoost*, *CatBoost*, dan juga LGBM. Dalam melakukan penelitian terkait pencemaran udara di kota Surabaya metode ini bisa menjadi metode yang bisa menganalisis dan dimanfaatkan guna memahami data masalah pencemaran udara di kota Surabaya. Algoritma *XGBoost*, *Catboost*, dan juga LGBM, ini merupakan Teknik *ensemble learning* dimana bisa melakukan pengelolaan data yang rumit dan dengan volume banyak guna mempunyai output yang akurat untuk menentukan hasil nantinya . Dengan implementasi Algoritma Machine Learning yaitu *XGBoost*, *Catboost*, dan juga LGBM diharapkan bisa memberikan hasil yang akurat dan membantu memberikan penyelesaian masalah dari pencemaran udara di Kota Surabaya.

XGBoost merupakan salah satu metode yang bagian dari pengembangan *gradient tree boosting* yang basisnya *ensemble*, dimana bisa melakukan pemrosesan data dengan nilai besar. Metode ini memiliki kelebihan dalam proses yang cepat dalam perhitungan dan bisa melakukan pencegahan *overfitting*. *XGBoost* juga bisa menyelesaikan beberapa kasus terkait klasifikasi, regresi dan ranking [8]. Konsep dasar dari metode ini yaitu dengan adanya penyesuaian dari parameter yang akan digunakan dilakukan secara berulang dimana nantinya akan menurunkan *loss function*(mekanisme evaluasi atas model). *XGBoost* ini karakteristiknya lebih teratur dalam menentukan pohon keputusannya, proses penentuan pohon Keputusan ini dapat memberikan kinerja yang baik untuk mengurai kompleksitas dan menghindari *overfitting*. Hasil yang diperoleh setelah melakukan tahapan yang ada di *XGBoost* yaitu penjumlahan hasil prediksi atau klasifikasi di setiap pohon keputusan. Dalam penelitian sebelumnya mengenai prediksi Indeks Pencemaran Udara (ISPU) di kota Pekanbaru [3], yang menggunakan metode *XGBoost* membuktikan bahwa metode tersebut mampu melakukan klasifikasi terhadap polutan udara seperti PM10, SO2, CO, O3, dan NO2. Dalam Penelitian tersebut nilai akurasi dari *XGBoost* mencapai 98%, yang tergolong sangat baik. Hal ini menunjukkan bahwa *XGBoost* mampu dalam menangani data yang tergolong kompleks dan mampu melakukan klasifikasi yang efektif. Namun, kekurangannya dalam penelitian ini adalah tidak ada perbandingan dengan metode yang lainnya sehingga tidak mampu memberikan pandangan yang lebih luas dalam hasil yang akan digunakan nantinya.

Selanjutnya, dalam penelitian ini dilakukan analisis dengan menggunakan Algoritma *Catboost*. Algoritma *CatBoost* merupakan salah satu bagian dari metode Gradient Boosting Decision Tree (GBDT). *CatBoost* memiliki kemampuan yang berfokus pada karakteristik dari kategori dalam data yang digunakan [9]. Selain itu, *CatBoost* mampu dalam melakukan penanganan nilai yang hilang dan data dengan jumlah yang besar [9]. Dalam cara kerjanya *CatBoost* mampu dalam menganalisis data tanpa perlu melakukan proses encoding secara manual misalnya dengan one-hot atau label encoding. *CatBoost* melakukan proses Ordered Target Encoding dimana hal ini mampu dalam mengurangi target leakage dan juga overfitting. Dengan kelebihan dari *CatBoost* diharapkan mampu memberikan hasil klasifikasi yang mampu dalam menganalisis kualitas pencemaran udara di Kota Surabaya. Sama halnya dengan penelitian sebelumnya yaitu mengenai Prediksi Indeks Kualitas Udara [10]. Penelitian ini dilakukan untuk memprediksi kualitas udara di kota Jakarta dengan menggunakan Algoritma *CatBoost*. Dimana data yang diambil berasal dari SPKU atau Sistem Pemantauan Kualitas Udara. Dimana dalam penelitian ini *CatBoost* dilatih dengan beberapa scenario kerja pembagian data yaitu 90:10, 80:20 dan 70:30. Hasil dalam penelitian ini yaitu nilai akurasi pada pembagian data 90:10 yaitu 97, lalu untuk data 80:20 menghasilkan akurasi senilai 96% dan data dengan 70:30 menghasilkan akurasi sebesar 95%. Hal ini membuktikan bahwa *CatBoost* merupakan salah satu Algoritma yang efektif dalam menangani kasus prediksi kualitas udara dalam penelitian tersebut. Dalam penelitian tersebut hanya menerapkan satu Algoritma yang dimana berfokus dalam *Catboost* dan tidak melakukan perbandingan dengan metode lainnya. Selain itu juga, dalam penelitian tersebut berfokus pada ISPU guna proses dalam penelitian tersebut. Oleh karena itu, dalam penelitian ini nantinya akan dilakukan Penerapan dari tiga metode yang berbeda, guna melihat kinerja dari setiap metode agar memperoleh hasil yang maksimal.

Tidak hanya algoritma *XGBoost* dan *CatBoost* yang digunakan dalam penelitian ini, tetapi juga algoritma Light Gradient Boosting Machine (LGBM) yang turut diterapkan sebagai metode perbandingan dalam melakukan analisis kualitas pencemaran udara di kota Surabaya. LGBM adalah suatu algoritma pohon keputusan, yang melakukan pembagian parameter pada input menjadi bagian yang berbeda dan membangun hubungan antara input dan output [11]. LGBM membangun pohon secara vertical, sedangkan umumnya Algoritma yang lain membangun pohon secara vertikal

[12]. Proses dalam pembentukan pohon ini akan digunakan leaf dengan nilai yang besar akan dilakukan pembaruan proses dan pengembangan[12]. Keunggulan Algoritma ini adalah teknik leaf-wise growth dimana melakukan perbaikan nilai error hingga mengurangi nilai error dari klasifikasi yang digunakan. Penelitian sebelumnya yaitu “*Air Quality Prediction System using LightGBM*” [12], yang memanfaatkan metode LGBM sebagai metode utama. LGBM menghasilkan nilai akurasi yang tinggi yaitu 98,38% dan dapat mengkategorikan seluruh data menjadi beberapa kategori. Dalam penelitian ini akan dilakukan penggunaan LGBM guna melakukan analisis berdasarkan data pencemaran udara. Selain itu juga LGBM akan dilihat dan dilakukan perbandingan dengan 2 metode lainnya yaitu *XGBoost* dan juga *CatBoost*.

Dengan menggunakan ketiga algoritma, yaitu *XGBoost*, *CatBoost*, dan LGBM, diharapkan penelitian ini mampu melakukan analisis pencemaran udara di Kota Surabaya secara optimal dan menghasilkan klasifikasi kelas yang akurat berdasarkan parameter-parameter polutan seperti PM10, SO2, CO, O3, dan NO2. Ketiga algoritma ini dipilih karena kemampuannya dalam menangani data kompleks serta memberikan hasil klasifikasi yang efisien dan presisi, sesuai dengan tujuan utama penelitian. Melalui penerapan algoritma-algoritma tersebut, diharapkan penelitian ini dapat memberikan kontribusi nyata dalam upaya memahami dan mengatasi pencemaran udara di Kota Surabaya. Hasil dari analisis ini diharapkan mampu memberikan insight mendalam bagi Pemerintah Kota Surabaya dalam mengidentifikasi faktor-faktor utama penyebab pencemaran udara, sehingga dapat dirancang program atau kebijakan strategis sebagai solusi konkret. Namun, penanggulangan pencemaran udara tidak hanya menjadi tanggung jawab pemerintah semata, tetapi juga memerlukan kesadaran aktif dari masyarakat Surabaya. Sebagai contoh, tindakan sederhana seperti tidak melakukan pembakaran sampah sembarangan dapat membantu mengurangi frekuensi dan tingkat pencemaran udara di kota ini.

1.2 Rumusan Masalah

- 1 Bagaimana penerapan algoritma *XGBoost*, *CatBoost*, dan LGBM untuk menganalisis pencemaran udara di Kota Surabaya?
- 2 Bagaimana hasil dari proses klasifikasi pencemaran udara berdasarkan variabel polutan dengan algoritma *XGBoost*, *CatBoost*, dan LGBM?
- 3 Seberapa efisien kinerja model *XGBoost*, *CatBoost*, dan LGBM dalam analisis kasus pencemaran udara di kota Surabaya?

1.3 Tujuan Penelitian

- 1 Melakukan Penerapan algoritma *XGBoost*, *CatBoost*, dan LGBM untuk melakukan analisis pencemaran udara di Kota Surabaya.
- 2 Melakukan analisis dengan menggunakan klasifikasi udara dari Penerapan Algoritma *XGBoost*, *CatBoost*, dan LGBM untuk memproses data pencemaran udara berdasarkan parameter polutan seperti PM10, SO2, CO, O3, dan NO2.
- 3 Memberikan gambaran dari hasil analisis yang dapat menjadi informasi guna memahami kondisi pencemaran udara di Kota Surabaya.

1.4 Manfaat Penelitian

- 1 Bagi Pemerintahan: Penelitian ini dapat memberikan analisis dari data yang telah dilakukan pengolahan dan bisa digunakan untuk menemukan Solusi untuk permasalahan pencemaran udara di Kota Surabaya.
- 2 Bagi Mahasiswa: Dalam Pengerjaan penelitian ini mahasiswa dapat mengembangkan kemampuannya dalam pengolahan data dan *Machine Learning*. Selain itu juga, mahasiswa dapat mengetahui bagaimana mengurus surat menyurat untuk keperluan kepada instansi yang dibutuhkan.
- 3 Bagi Institusi : Dengan adanya ini kampus bisa menambah relasi untuk kerja sama oleh instansi guna keperluan akademik kedepannya.

1.5 Batasan Masalah

- 1 Hanya meneliti terkait dengan pencemaran udara di Kota Surabaya dan tidak membahas hal diluar topik utama (seperti pencemaran air dan tanah)
- 2 Menggunakan variabel yang ada pada konsentrasi polutan (PM10, SO2, CO, NO2, dan O3.) dan yang berkaitan untuk menyelesaikan masalah yang ada.
- 3 Data yang digunakan hanya data tahun 2023 yang diambil dari Dinas Lingkungan Hidup (DLH) kota Surabaya.
- 4 Penelitian ini hanya berfokus pada analisis hasil klasifikasi dari model *XGBoost*, *CatBoost*, dan LGBM tanpa membahas integrasi sistem atau penerapan model secara real-time.

BAB II

TINJAUN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian-penelitian yang sudah dilakukan peneliti sebelumnya yang terkait dengan penelitian yang akan dilakukan adalah sebagai berikut.

1. Analisis dan Komparasi Algoritma Klasifikasi dalam Indeks Pencemaran Udara di DKI Jakarta [1]. Pada penelitian ini melakukan pengamatan terhadap Indeks pencemaran udara di DKI Jakarta. Dimana dengan melakukan perbandingan analisis dan komparasi dari beberapa algoritma yang ada yaitu *Neural Network* , *SVM (Support Vector Machine)*, *K-Nearest Neighbors*, *Naive Bayes* dan *Decission Tree* dengan mengambil *T-Test* yang digunakan untuk bahan pengujian. Dimana dalam penelitian ini digunakan proses pengolahan data dimana akan mencari data yang masih belum tepat dan salah. Setelah itu data akan dilakukan penerapan ke 5 metode yang akan digunakan. Setelah dilakukan semua penerapan metode, *Decission Tree* menghasilkan sebesar 99,80% pada akurasi, lalu pada nilai Kappa sebesar 0,996, serta dalam waktu pengolahan hanya membutuhkan waktu 0,8 detik. Namun, metode lainnya tetap memiliki kualitas yang masih bagus dimana akurasi diatas 90% yang bisa dibilang masih tinggi.
2. Prediksi Kualitas Udara Menggunakan *XGBoost* dengan *Synthetic Minority Oversampling Technique (SMOTE)* Berdasarkan Indeks Standar Pencemaran Udara (ISPU)[3]. Penelitian dilakukan dengan menggunakan algoritma *XGBoost* dan menggunakan teknik SMOTE dimana berdasarkan kategori ISPU yang menjadi patokan utama. Data yang digunakan yaitu diambil dari Kementerian Lingkungan Hidup rentang tahun 2017 sampai tahun 2021. Sebelum melakukan proses *XGBoost* dilakukan proses SMOTE dimana melakukan inisiasi dimana misalnya terjadi ketidakseimbangannya data yang digunakan. Setelah proses penelitian terjadi hasil menunjukkan bahwa dengan melakukan kombinasi 2 proses tersebut menghasilkan kinerja yang sangat baik dimana nilai rata-rata akurasi sebesar 98%, *Precision* sebesar 79%, *Recall* sebesar 79%, nilai *F1-score* 98% dan ROC AUC sebesar 99%
3. Dalam penelitian “Prediksi Kualitas Udara Menggunakan Metode *CatBoost*” [10], Dimana yang mengambil studi kasus udara Jakarta. Dalam studi kasus tersebut pentingnya kualitas udara menjadi salah satu fokus utama dalam permasalahan