

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Udara yang merupakan campuran gas tak berwarna yang meliputi permukaan Bumi dan tidak terlihat, tidak berbau, atau tidak memiliki rasa, merupakan sumber daya alam yang sangat penting bagi seluruh makhluk hidup, terutama manusia. *Oxygen* (O₂) sangat penting untuk respirasi manusia, karbon dioksida (CO₂) sangat penting untuk fotosintesis, dan ozon (O₃) membantu melindungi kulit dari sinar matahari[1]. Kondisi cuaca dapat memengaruhi kualitas udara dan menyebabkan pencemaran dan penurunan kualitas udara. Selain itu, kualitas udara menjadi lebih buruk karena pertumbuhan industri, peningkatan populasi manusia, dan peningkatan ruang terbuka hijau dan pelestarian yang masih rendah. Dengan peningkatan kualitas udara, efeknya terhadap kesehatan makhluk hidup meningkat, dan sebaliknya[2].

Peranan udara sangat vital bagi kehidupan makhluk hidup di bumi. Proses metabolisme yang terjadi di dalam tubuh makhluk hidup memerlukan keberadaan oksigen dari udara. Namun, selain mengandung oksigen, udara juga mengandung berbagai zat seperti karbon monoksida, karbon dioksida, sulfur dioksida, nitrogen oksida, ozon, dan zat lainnya. Beberapa zat tersebut masih dapat diterima oleh tubuh dalam konsentrasi yang rendah. Namun, jika melebihi batas normal, hal ini dapat menimbulkan masalah serius pada kesehatan manusia, seperti penyakit jantung, infeksi pernapasan, stroke, asma, hingga penyakit paru[3]. Mempertahankan kebersihan udara menjadi krusial guna mendukung kesehatan kita dan menjaga integritas lingkungan. Sayangnya, banyak orang yang kurang menyadari urgensi pemeliharaan kualitas udara. Akibatnya, baik dengan kesadaran maupun tanpa disadari, mereka dapat menyebabkan pencemaran udara[4].

Salah satu kota terburuk di dunia dalam hal pencemaran udara adalah Jakarta dan wilayah sekitarnya. Tingkat polusi di wilayah DKI Jakarta dan sekitarnya jauh melampaui ambang batas kesehatan Organisasi Kesehatan Dunia (WHO) dan standar baku kualitas udara ambien nasional, menurut

pemantauan kualitas udara yang dilakukan oleh Greenpeace pada tahun 2017. Sebaliknya, orang-orang di Jakarta dan sekitarnya masih kurang memahami polusi udara dan bahayanya. Hal ini terutama disebabkan oleh kekurangan informasi tentang kualitas udara[5]. Pada pemantauan tahun 2020, meskipun terjadi penurunan aktivitas manusia akibat pandemi COVID-19, kualitas udara dari bulan Maret hingga Mei masih menunjukkan tingkat pencemaran sedang hingga tidak sehat. Pemerintah provinsi DKI Jakarta telah melakukan pemantauan kualitas udara ambien dan telah menetapkan Indeks Standar Pencemaran Udara (ISPU).

Penentuan tingkat level ISPU dapat dipermudah menggunakan proses klasifikasi dari data mining. Data mining berfungsi untuk menggali informasi lebih dalam dengan jumlah data yang banyak. Implementasi data mining dapat menjadi solusi untuk mengetahui kualitas udara menggunakan teknik klasifikasi[6]. Klasifikasi adalah proses dengan menggunakan data pembelajaran untuk pembelajaran dan data pengujian untuk pengujian ini dikenal sebagai klasifikasi. Didasarkan pada karakteristik yang dimiliki oleh setiap objek, model ini mengelompokkan objek ke dalam kelas yang sesuai dengan aturan yang telah ditetapkan[7]. Metode untuk memetakan (mengklasifikasikan) data ke dalam kelas yang telah ditentukan sebelumnya dikenal sebagai klasifikasi sistem informasi. Ada banyak metode klasifikasi yang dapat digunakan, termasuk K-Nearest Neighbor (KNN) dan Modified K-Nearest Neighbor. Banyak penelitian sebelumnya telah mengaplikasikan konsep algoritma KNN, seperti dalam pengenalan pola, pengenalan teks, pengolahan objek, dan bidang lainnya. Algoritma KNN dianggap mudah untuk melatih dan memeriksa data, terutama dalam situasi data yang sangat besar[8].

Salah satu penelitian yang menggunakan metode K-Nearest Neighbor adalah penelitian yang dilakukan oleh Amalia pada tahun 2022 tentang Perkiraan Kualitas Udara menggunakan algoritma K-Nearest Neighbor (KNN) untuk memprediksi kualitas udara di Daerah Khusus Ibukota Jakarta menggunakan Indeks Standar Pencemaran Udara (ISPU). Hasil prediksi berbeda-beda tergantung pada nilai K hasil pengujian terbaik diperoleh pada

K = 7 dan memiliki akurasi 96%. Studi tambahan yang dilakukan oleh Siswanto dkk. (2018) mencoba mengkategorikan desa berdasarkan Indeks Pembangunan Desa (IPD) yang diterbitkan oleh Badan Pusat Statistik (BPS) pada tahun 2014. Studi ini menghasilkan tingkat ketepatan atau tingkat akurasi sebesar 92,22% dengan menggunakan algoritma KNN dengan nilai K = 3[9].

Pada penelitian menggunakan data Indeks Standar Pencemaran Udara (ISPU) Tahun 2023 dari DKI Jakarta dengan memanfaatkan metode K-Nearest Neighbor (KNN) untuk melakukan prediksi. Analisis klasifikasi akan dilakukan untuk memahami pola atau kategori kualitas udara berdasarkan ISPU, sehingga dapat memberikan wawasan lebih lanjut terkait kondisi udara di DKI Jakarta[6].

Untuk mengatasi masalah ketidakseimbangan kelas dalam dataset, teknik over-sampling sintetis minoritas (SMOTE) menggunakan sampel baru dari kelas minoritas untuk menyamakan distribusi kelas dalam dataset. Dengan cara ini, SMOTE menyintesis sampel tambahan dari kelas minoritas dengan melakukan oversampling pada sampel-sampel yang sudah ada. Dalam konteks masalah clustering menggunakan metode k-Means Clustering, penerapan SMOTE telah berhasil mengatasi gangguan (noise) dan menyelesaikan masalah ketidakseimbangan pada 71 dataset yang diuji. Penerapan SMOTE juga terbukti dapat meningkatkan kualitas hasil clustering[10].

Berdasarkan permasalahan pada wilayah DKI Jakarta penelitian ini mengimplementasikan algoritma KNN menggunakan SMOTE oversampling dalam melakukan klasifikasi terhadap ISPU di wilayah DKI Jakarta agar pihak dinas lingkungan hidup bisa mendapatkan hasil yang akurat terkait klasifikasi kualitas udara di wilayah DKI Jakarta untuk memberikan solusi maupun instruksi untuk pencegahan masalah kualitas udara kepada masyarakat di sekitar wilayah DKI Jakarta.

1.2. Rumusan Masalah

1.2.1. Permasalahan

Berdasarkan latar belakang masalah penelitian ini adalah data ISPU tidak memiliki kepastian yang akurat dikarenakan pengukuran dan pengumpulan data tidak selalu dilakukan dengan cara yang sama di semua lokasi sehingga menghasilkan data yang sulit untuk dibandingkan dan dianalisis secara konsisten. Oleh karena itu dibutuhkan klasifikasi kualitas udara untuk mendapatkan hasil yang akurat untuk mengelompokkan data multiclass.

1.2.2. Solusi Permasalahan

Metode yang akan digunakan dalam penelitian ini yaitu algoritma *K-Nearest Neighbor*.

1.2.3. Pertanyaan Penelitian

Berdasarkan masalah dan metode yang disajikan, pertanyaan penelitian sebagai berikut :

Berapa hasil akurasi dan k optimal pada klasifikasi menggunakan algoritma *K-Nearest Neighbor* terhadap kualitas udara dengan penyeimbangan data?

1.3. Batasan Masalah

Mengacu pada kesimpulan dari analisis permasalahan sebelumnya, riset ini perlu dirancang dengan lebih terstruktur agar batasan permasalahan dapat diuraikan sebagai berikut::

1. Data yang dipakai dalam penelitian ini berfokus pada dataset ISPU DKI Jakarta pada periode tahun 2023. Data diambil dari website (<https://bit.ly/data-ispu-2023>) dengan jumlah data 1825 data.
2. Dataset yang digunakan terdiri dari 6 atribut dan 3 kelas, Atribut-atribut yang akan digunakan terdapat dalam dataset ini antara lain PM₁₀, PM_{2.5}, SO₂, CO, O₃, NO₂, dan kelas Kategori terdiri dari 3 kelas yaitu BAIK, SEDANG, TIDAK SEHAT.
3. Sistem menggunakan bahasa pemrograman python.
4. Metode yang digunakan untuk penelitian ini yaitu *K-Nearest Neighbor*.
5. Penyeimbangan data menggunakan *SMOTE Oversampling*.
6. Pembagian *data training* dan *data testing* menggunakan *split data*.
7. Menghitung kinerja menggunakan *confusion matrix*.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Memperoleh hasil akurasi dan k optimal pada klasifikasi menggunakan algoritma *K-Nearest Neighbor* terhadap kualitas udara dengan menggunakan penyeimbangan data.

1.4.2 Manfaat

Adapun manfaat penelitian ini yaitu untuk mengetahui akurasi dan k optimal pada algoritma *K-Nearest Neighbor* terhadap kualitas udara menggunakan penyeimbangan data dan mengetahui perbandingan metode *split* data 80:20%, 70:30%, 60:40% pada klasifikasi menggunakan algoritma *K-Nearest Neighbor*.

1.5. Sistematika Penulisan

BAB I PENDAHULUAN

Latar belakang penelitian, masalah yang diangkat, perbaikan yang disarankan untuk masalah tersebut, batasan penelitian karena masalah yang diangkat, dan tujuan serta keuntungan dari penelitian yang disarankan semuanya tercakup dalam bab ini.

BAB II TINJAUAN PUSTAKA

Pada bab ini membahas teori yang dibutuhkan untuk penelitian, seperti algoritma *K-Nearest Neighbor* dalam klasifikasi kualitas udara, serta beberapa penelitian sebelumnya.

BAB III METODE PENELITIAN

Bab ini menjelaskan perancangan sistem, dimulai dengan perangkat lunak dan perangkat keras yang diperlukan serta rancangan modelnya.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil implementasi dan algoritma klasifikasi dan hasil evaluasi dari klasifikasi.

BAB V PENUTUP

Bab ini menjelaskan mengenai kesimpulan dari hasil klasifikasi beserta saran dalam penelitian ini.