

BAB I PENDAHULUAN

1.1 Latar Belakang

Pencemaran udara adalah masalah yang memiliki dampak negatif terhadap kehidupan makhluk hidup. Kondisi ini dapat menyebabkan penurunan kualitas udara, terutama di wilayah perkotaan. Faktor yang menyebabkan terjadinya pencemaran udara yaitu aktivitas transportasi, sektor industri, serta kegiatan lainnya yang membuang sisa limbah ke udara (Handayani, Silva et al., 2020). Udara yang tercemar dapat berdampak buruk pada Kesehatan serta kerusakan pada ekosistem dan bangunan, bahkan bisa berakibat fatal. Pencemaran udara merupakan masalah global yang sangat memprihatinkan di banyak negara, termasuk Indonesia. Jakarta, sebagai kota metropolitan terbesar, terkenal memiliki tingkat pencemaran udara yang cukup tinggi.

Pada hari pertama dibulan Oktober 2024 menunjukkan tingkat kualitas udara di DKI Jakarta dalam kategori ‘tidak sehat’. Berdasarkan data Indeks Kualitas Udara atau *Air Quality Index* (AQI) tercatat diangka 160 dengan konsentrasi partikel halus (*Particulate Matter/PM*) 2,5 menembus 68,7 *microgram* per meter kubik. Konsentrasi PM2.5 di Jakarta 13,7 kali lebih banyak dari standar WHO (Tempo, 202428). Standar yang direkomendasikan oleh Organisasi Kesehatan Dunia (WHO) sebesar 5 $\mu\text{g}/\text{m}^3$ (*IQAir Staff Writers*, 2024). Pada akhir bulan lalu, tepatnya pada 28 – 30 September 2024, kualitas udara di Jakarta masih berada dalam kategori ‘tidak sehat bagi kelompok sensitif. Meskipun masih terpengaruh oleh polusi, kategori ini sedikit lebih baik dibandingkan dengan ‘tidak sehat’. Indeks kualitas udara di Jakarta sempat turun dari 147 menjadi 109, yang sebelumnya naik Kembali ke 140 dan akhirnya mencapai 160 (Tempo, 202428). *Greenpeace* Indonesia memperkirakan 7.390 penduduk Jakarta meninggal lebih awal karena polusi, sedangkan 2000 bayi lahir dengan berat badan rendah karena penyebab yang sama. Populasi yang paling terdampak oleh polusi udara antaranya adalah anak – anak, lansia dan individu dengan penyakit komorbid (Health, 2024). Masalah pencemaran udara di DKI Jakarta bisa dibilang merupakan tantangan terbesar dan paling kompleks yang dihadapi oleh pemerintah (Bernadet et al., 2023).

Pemerintah Indonesia, melalui Keputusan Badan Pengendalian Dampak Lingkungan (Bapedal) Nomor KEP-107/Kabapedal/11/1997, menetapkan Indeks Standar Pencemaran Udara (ISPU) sebagai alat untuk menilai kualitas udara disuatu wilayah serta dampaknya terhadap kesehatan manusia setelah terpapar udara selama beberapa jam hingga beberapa hari. Penetapan Indeks Standar Pencemaran Udara mempertimbangkan kualitas udara berdasarkan pengaruhnya terhadap kesehatan manusia, hewan, tumbuhan serta nilai estetika.

Penentu level ISPU dapat dilakukan dengan menerapkan teknik klasifikasi dalam data mining sebagai solusi yang efektif (Firdaus et al., 2021). Data mining adalah teknik yang digunakan untuk menggali informasi baru dari suatu data dengan mengidentifikasi pola atau aturan tertentu dari kumpulan data yang besar (Ramadhan et al., 2023). Penerapan data mining menjadi solusi untuk menganalisis kualitas udara di DKI Jakarta dengan menggunakan pendekatan teknik klasifikasi.

Menurut penelitian yang di lakukan Lasulika (2019). mengenai Komparasi *Naive Bayes*, *Support Vector Machine* dan *K-Nearest Neighbor* untuk Mengetahui Akurasi Tertinggi pada Kelancaran Pembayaran TV Kabel. Berdasarkan hasil pengujian bahwa algoritma *naive bayes* unggul dari algoritma K-NN dan SVM untuk tipe data *multivariant* dengan nilai *accuracy* sebesar 96%, sementara algoritma K-NN diuji coba dengan beberapa nilai K yaitu 3,5,7 dan 9, didapatkan *accuracy* tertinggi yaitu pada nilai $K = 3$ sebesar 92%. Sementara SVM hanya mendapatkan *accuracy* sebesar 66%. Pada uji coba *ROC Curve* menunjukkan bahwa *naive bayes* mencapai nilai AUC yang terbaik yaitu 0.99, K-NN 0.971 dan SVM 0.786.

Kemudian oleh Arnap (2024) melakukan penelitian tentang Meningkatkan Deteksi Serangan Injeksi SQL Menggunakan Metode *Naive Bayes* dan SMOTE pada Dataset yang Tidak Seimbang. Penelitian ini memanfaatkan SMOTE untuk data yang tidak seimbang yang dapat mempengaruhi efektivitasnya. Penelitian ini membagi data menjadi 80% data latih serta 20% data uji, dataset terdiri dari 4.200, dengan 1.128 diberi label 1 dan 3.072 diberi label 0. Setelah penerapan teknik SMOTE dataset memiliki 2.978 untuk label 1 dan 2.978 untuk label 0. Menghasilkan akurasi 99.50%, f1-score 99.50%, presisi 99.50% dan recall 99.50%.

menunjukkan bahwa teknik SMOTE berhasil menyeimbangkan kelas dalam dataset dan meningkatkan deteksi serangan injeksi SQL.

Berdasarkan latar belakang diatas, tujuan penelitian yang dilakukan adalah untuk mengukur kinerja algoritma *naive bayes* menggunakan teknik SMOTE dalam melakukan klasifikasi pencemaran udara di wilayah DKI Jakarta.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana penerapan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) dapat meningkatkan kinerja klasifikasi kualitas udara pada data ISPU yang tidak seimbang?
2. Bagaimana hasil nilai evaluasi yang diperoleh dalam klasifikasi data ISPU menggunakan algoritma *naive bayes* dengan teknik SMOTE dibandingkan dengan klasifikasi tanpa menggunakan teknik SMOTE?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis penerapan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) dalam meningkatkan kinerja klasifikasi kualitas udara pada data ISPU yang tidak seimbang.
2. Mengukur hasil nilai evaluasi klasifikasi data ISPU menggunakan algoritma *naive bayes* dengan teknik SMOTE dan membandingkan hasilnya dengan klasifikasi tanpa menggunakan SMOTE.

1.4 Manfaat

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Memberikan data klasifikasi kualitas udara yang lebih akurat dengan menerapkan teknik SMOTE untuk mengatasi ketidakseimbangan data ISPU, sehingga dapat mendukung perumusan kebijakan mitigasi polusi dan mempermudah pemantauan kualitas udara.
2. Mengonfirmasi keakuratan algoritma *naive bayes* dengan dan tanpa SMOTE dalam mengklasifikasikan data ISPU, serta memberikan rekomendasi berbasis data untuk meningkatkan pengelolaan dan pemantauan kualitas udara secara lebih optimal.