

1. PENDAHULUAN

1.1. Latar Belakang

Udara adalah suatu campuran gas alam yang mengelilingi permukaan bumi. Udara tersusun dari banyaknya gas, diantaranya nitrogen 78%, oksigen 20%, argon 0,93% dan karbon dioksida 0,30%, dan gas gas lainnya [1]. Udara memiliki peranan penting untuk segala aspek makhluk hidup di bumi, baik itu manusia, hewan atau tumbuhan. Tanpa adanya udara, makhluk hidup tidak ada yang bisa bertahan hidup. Udara berbentuk transparan, tidak bisa dilihat tapi bisa kita rasa. Banyak manfaat udara bersih bagi manusia untuk kelangsungan hidupnya, diantaranya: menyehatkan saluran pernapasan, menurunkan risiko penyakit kronis, memperpanjang usia, meningkatkan stamina dan fokus dan memperbaiki *mood*.

Pencemaran udara adalah kehadiran suatu atau lebih substansi fisik, kimia, atau biologi di atmosfer dalam jumlah yang dapat membahayakan kesehatan manusia, hewan, dan tumbuhan, mengganggu estetika dan kenyamanan, atau merusak properti [2]. Pencemaran udara membuat kualitas udara disuatu daerah menjadi turun, dan kurang bagus untuk makhluk hidup, bahkan bisa membahayakan. Banyak faktor yang memicu penurunan kualitas udara disuatu daerah diantaranya aktivitas transportasi, industri, jasa, kegiatan lainnya yang meningkat, dan buangan sisa-sisa kegiatan ke udara. Di kota besar seperti DKI Jakarta, pencemaran udara merupakan salah satu masalah yang cukup serius yang perlu dihadapi oleh masyarakat DKI Jakarta. Menurut data *AirVisual* yang ditampilkan oleh AQI, Indonesia menjadi negara ke-9 sebagai negara yang paling berpolusi di dunia pada tahun 2020, dan Jakarta menjadi kota ke-202 sebagai kota paling berpolusi di dunia tahun 2020 [3]. Dari data *AirVisual* yang berikan oleh AQI, Jakarta bisa dikategorikan menjadi kota dengan udara yang sedang – tidak sehat.

Berdasarkan permasalahan udara yang sedang terjadi pada DKI Jakarta, dapat kita menerapkan suatu proses olah data untuk menjadi suatu nilai informasi. *Data mining* adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar. Klasifikasi adalah model fungsi yang menggambarkan dan membedakan kelas atau konsep untuk prediksi masa depan, dari data yang diperoleh sebelumnya. *Naive bayes* merupakan suatu algoritma klasifikasi yang berakar pada probabilitas dan statistik. Data yang digunakan bersumber dari *Open Jakarta* yang dimana *dataset* membahas Indeks Standar Pencemaran Udara (ISPU) yang diukur dari 5 Stasiun Pemantau Kualitas Udara (SPKU) pada Provinsi DKI

Jakarta. Datanya bersifat *kontinu*, yang berbentuk *numerical* dan *categorycal* yang cocok digunakan pada *naive bayes*.

1.2. Perumusan Masalah

Berdasarkan latar belakang, terdapat rumusan masalah sebagai berikut:

1. Bagaimana penerapan *data mining* menggunakan algoritma *naive bayes* untuk klasifikasi tingkat kualitas udara di Provinsi DKI Jakarta?
2. Bagaimana hasil dari penerapan *data mining* menggunakan algoritma *naive bayes* untuk klasifikasi tingkat kualitas udara di Provinsi DKI Jakarta?
3. Bagaimana akurasi dari penerapan *data mining* menggunakan algoritma *naive bayes* untuk klasifikasi tingkat kualitas udara di Provinsi DKI Jakarta?

1.3. Tujuan

Menerangkan kondisi apa yang hendak dicapai adalah sebagai berikut:

1. Menerapkan algoritma *naive bayes* sebagai bahan analisis terhadap kualitas udara DKI Jakarta
2. Mengetahui hasil penerapan algoritma *naive bayes* dan memberi informasi terhadap kualitas udara DKI Jakarta

1.4. Batasan Masalah Penelitian

Batasan masalah penelitian bertujuan untuk membatasi penelitian agar tidak terlalu luas dan tidak melebar dari tujuan yang ada. Maka batasan masalah penelitian tugas akhir ini sebagai berikut:

1. Menggunakan data yang disediakan dari *website Open Jakarta* (<https://data.jakarta.go.id/>)
2. Data yang digunakan dari tahun februari 2021 hingga bulan oktober 2021
3. Penggunaan algoritma *naive bayes*
4. Tools yang digunakan adalah *Jupyter Notebook* dengan menggunakan bahasa pemrograman *python 3.8*
5. Parameter yang digunakan hanya karbon monoksida (CO), sulfur dioksida (SO₂), nitrogen dioksida (NO₂), ozon permukaan (O₃), partikel debu (PM₁₀), partikel debu (PM₂₅), dan *location* (lokasi)

1.5. Manfaat Penelitian

Adapun manfaat dari Tugas Akhir ini, sebagai berikut:

1. Manfaat untuk penulis:

1.1 Penulis dapat menerapkan dan memahami klasifikasi *Naive Bayes*

1.2 Penerapan dari pembelajaran pengetahuan dari dalam perkuliahan maun pun diluar perkuliahan

2. Manfaat untuk Institusi (Telkom University):

2.1 Diharapkan sebagai bentuk pembelajaran untuk penelitian selanjutnya yang topik atau algoritma yang berhubungan dengan Tugas Akhir ini

3. Manfaat untuk Organisasi

3.1 Membantu masyarakat DKI Jakarta agar menambah informasi mengenai kualitas udara di DKI Jakarta

1.6. Sistematika Penulisan

BAB I Pendahuluan, pada bagian ini berisikan latar belakang, penjelasan masalah, perumusan masalah, tujuan masalah, batasan masalah, manfaat penelitian dan sistematika penulisan.

BAB II Tinjauan Pustaka, pada bagian ini berisikan teori/studi/literatur yang mendukung (terkait erat) dengan topik Tugas Akhir yang dikerjakan

BAB III Metodologi Penelitian, pada bagian ini berisikan penjelasan rancangan dan sistem yang dihasilkan

BAB IV Analisa Hasil, pada bagian ini berisikan dua sub-bagian, yaitu skenario uji dan analisis hasil pengujian.

BAB V Kesimpulan, pada bagian ini berisikan hasil pengujian yang telah dirangkum pada bab sebelumnya.