

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pencemaran udara merupakan suatu masalah yang berdampak buruk bagi makhluk hidup. Pencemaran udara menimbulkan berbagai macam penyakit dari yang tidak menular hingga penyakit yang menular. Melalui pengamatan yang dilakukan oleh *Green Peace*, beberapa kota-kota besar di negara China terkenal memiliki tingkat kualitas yang sangat tinggi. Berdasarkan hasil pengamatan yang dilakukan oleh *Green Peace* lebih dari 3000 kota menunjukkan tingginya tingkat kualitas udara di kota-kota seluruh dunia. Lebih dari 50% diantaranya melebihi pedoman paparan tahunan *World Healthy Organization (WHO)* yaitu $10\mu\text{g} / \text{m}^3$ untuk partikel halus atau biasa dikenal PM2.5.

Kualitas udara di beberapa kota Indonesia saat ini juga tidak terlalu baik atau tergolong tidak sehat. Menurut *Green Peace* pada tahun 2018 kota Jakarta mendapat peringkat sepuluh besar sebagai ibu kota negara dengan kualitas udara terburuk di dunia. Pada tahun 2020 Dinas Lingkungan Hidup juga melakukan pemantauan melalui beberapa titik stasiun yang ada di Jakarta dengan hasil pola perubahan tingkat udara setiap harinya.

Berdasarkan hasil data pemantauan Dinas Lingkungan Hidup selama tahun 2020 melalui beberapa titik stasiun yang ada di kota Jakarta, diperoleh pola data tingkat kualitas udara di Jakarta yang sangat banyak.

Untuk mengatasi masalah tersebut, salah satu solusi untuk mengolah data adalah menggunakan teknik *Data Mining* dan *Google Colab* untuk merancang program yang bertujuan mengetahui klasifikasi pola kualitas tingkat pencemaran udara.

Data Mining merupakan sebuah proses menemukan hubungan atau pola yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang mudah dipahami sehingga menjadi pengetahuan (Larose, D.T, 2005). Salah satu fungsi utama *Data Mining* adalah klasifikasi. Untuk proses klasifikasi dalam *Data Mining* terdapat

algoritma yang cocok untuk digunakan. Salah satunya menggunakan algoritma *Naïve Bayes*. Pengklasifikasi Bayesian merupakan pengklasifikasi statistic untuk memprediksi probabilitas keanggotaan tertentu (Hang., 2006), menghitung peluang untuk suatu hipotesis, menghitung peluang dari suatu kelas dari masing-masing kelompok atribut yang ada dan menentukan kelas mana yang paling optimal (Hajjaratie, 2011). *Naïve Bayes* merupakan suatu kelas, dengan menggunakan perhitungan probabilitas matematika dengan syarat nilai bahwa nilai keputusan adalah benar, berdasarkan informasi dari objek. *Google Colab* adalah *Executeable Document* yang dapat digunakan untuk menulis, menyimpan dan membagikan program. *Colab* merupakan *cloud-based runtime*, yang dijalankan menggunakan *browser* (*Chrome*, *Firefox*, dan *Safari*). *Colab* memungkinkan menjalankan kode *python* tanpa memerlukan proses instalasi dan semua fungsionalitasnya yang dimiliki oleh *python* serta memanfaatkan *built-in library* yang disediakan oleh *Google Colab*.

Terdapat beberapa penelitian yang dilakukan berkaitan dengan *Naïve Bayes*. Penelitian berjudul Penerapan *Data Mining* Untuk Mengklasifikasi Bahaya Polutan PM10 Menggunakan Algoritma *Naïve Bayes* Di Kota Banjarbaru yang dilakukan oleh (Nur Amiramah, Mirza Yogi Kurniawan, dan Al Fath Riza Kholdani, 2020) telah berhasil melakukan proses klasifikasi dengan algoritma *Naïve Bayes* dengan akurasi sebesar 60%.

Selanjutnya, penelitian juga pernah dilakukan oleh (Nurul Aini, Rima Ruktiari, M Riyaldi Pratama, dan A. Fitrah Buana, 2019) yang berjudul Sistem Prediksi Tingkat Pencemaran Polusi Udara dengan Algoritma *Naïve Bayes* di Kota Makasar dengan akurasi sebesar 0, 60 atau 60%.

Selanjutnya, penelitian lain juga dilakukan oleh (Reza Taquuddin, 2017) yang berjudul Klasifikasi Kualitas Udara Menggunakan *Naïve Bayes Classifier* Pada Sistem Terdistribusi *Raspberry PI Cluster Server* dengan akurasi 98%.

Berdasarkan latar belakang permasalahan diatas, maka penelitian yang akan dilakukan akan mengambil judul **“Penerapan *Data Mining* Untuk Mengetahui Pola**

Klasifikasi Tingkat Kualitas Pencemaran Udara Di Kota Jakarta Menggunakan Algoritma *Naïve Bayes*".

1.2 Identifikasi Masalah

Berdasarkan permasalahan diatas, maka dapat disimpulkan identifikasi masalah tersebut adalah sebagai berikut:

1. Tingkat kualitas udara di Jakarta sangat buruk.
2. Belum adanya pola klasifikasi tingkat pencemaran udara di kota Jakarta selama bulan Januari tahun 2020.

1.3 Rumusan Masalah

Berdasarkan latar belakang diatas maka rumusan masalahnya adalah Bagaimana mengetahui pola klasifikasi tingkat pencemaran udara di kota Jakarta menggunakan algoritma *Naïve Bayes*?

1.4 Batasan Masalah

Dalam penelitian ini hanya dibatasi pada:

1. Stasiun sekitar Bunderan HI, Kelapa Gading, Jagakarsa, Lubang Buaya dan Kebon Jeruk di Jakarta selama tahun 2020.
2. Indeks standar yang digunakan adalah PM10, SO₂, CO, O₃, dan NO₂
3. Penulis belum bisa menghitung hanya dengan menggunakan 1 data.

1.5 Tujuan Penelitian

Adapun tujuan dari penelitian ini, yaitu sebagai berikut:

1. Mengetahui tingkat pencemaran udara di kota Jakarta selama tahun 2020.
2. Mengetahui tingkat akurasi yang didapatkan dalam penerapan algoritma *Naïve Bayes*.

1.6 Manfaat Peneliti

Manfaat dari penelitian yang dilakukan ini adalah semoga dapat meningkatkan kesadaran masyarakat kota Jakarta akan bahayanya pencemaran udara.

1.7 Waktu Penelitian

Penelitian dilakukan pada tanggal 15 April 2021 – 15 Mei 2021

1.8 Sistematika Penulisan

Penelitian ini akan dibagi dalam lima bab gambaran masing-masing bab akan dijelaskan di bawah ini:

BAB I PENDAHULUAN

Dalam bab ini berisi penjelasan tentang latar belakang masalah, maksud dan tujuan penelitian, rumusan masalah, pembahasan masalah, metode pengumpulan data dan sistematika penulisan.

BAB II LANDASAN TEORI

Dalam Bab ini menjelaskan tentang Konsep dasar dari sebuah system informasi yang akan di bahas dalam ruang lingkup penelitian terutama yang berhubungan dengan penerapan konsep dalam system tersebut.

BAB III METODOLOGI PENELITIAN

Pada Bab ini membahas tentang pengumpulan data, metode yang diimplementasikan pada penelitian yaitu metode pengembangan sistem baik perangkat lunak maupun perangkat keras.

BAB IV PERANCANGAN SISTEM DAN IMPLEMENTASI

Bab ini membahas mengenai perancangan sistem terhadap identifikasi masalah dan menghasilkan serangkaian instruksi yang membangun sebuah program komputer untuk menghasilkan output.

BAB V PENUTUP

Berisi tentang beberapa kesimpulan yang didapatkan dari hasil pembahasan bab-bab sebelumnya, serta saran-saran yang dapat di kembangkan atau dilakukan sebagai penerapan untuk perusahaan kedepannya.