

BAB 1

PENDAHULUAN

1.1 LATAR BELAKANG

Perkembangan Teknologi Informasi (TI) di era digital telah membawa transformasi besar di berbagai sektor, termasuk kesehatan, pendidikan, ekonomi, dan lingkungan. Salah satu kontribusi utama TI adalah kemampuannya dalam mengelola serta menganalisis data berskala besar secara cepat dan akurat. Dalam bidang lingkungan, TI memiliki peran penting dalam memantau, mengolah, dan menganalisis data lingkungan, khususnya kualitas udara yang menjadi indikator krusial dalam menentukan kesehatan suatu wilayah.

Polusi udara merupakan isu global yang memberikan dampak serius terhadap kesehatan dan kualitas hidup masyarakat. Di Indonesia, polusi udara menjadi masalah yang semakin mendesak, terutama di kota-kota besar seperti Jakarta. Menurut laporan terbaru, Indonesia menempati peringkat terburuk dalam kualitas udara di Asia Tenggara, yang berkontribusi terhadap meningkatnya kasus penyakit pernapasan serta kematian yang terkait dengan polusi [1].

Dampak polusi udara tidak hanya memengaruhi aspek kesehatan, tetapi juga memberikan konsekuensi ekonomi yang signifikan. Penurunan kualitas udara dapat mengurangi angka harapan hidup masyarakat serta meningkatkan biaya perawatan kesehatan [1]. Oleh karena itu, diperlukan solusi yang efektif dalam mengatasi permasalahan ini, salah satunya melalui penerapan teknologi dan metode analisis data.

Dalam konteks ini, machine learning hadir sebagai alat yang menjanjikan untuk menganalisis dan memprediksi kualitas udara. Berbagai algoritma telah diterapkan dalam klasifikasi data kualitas udara, termasuk K-Nearest Neighbors (KNN) dan Naïve Bayes. Namun, meskipun banyak penelitian telah dilakukan, masih terdapat kekurangan studi yang secara khusus membandingkan kedua algoritma ini dalam konteks klasifikasi kualitas udara [2].

Pemilihan algoritma K-Nearest Neighbors (KNN) dan Naïve Bayes sebagai fokus penelitian didasarkan pada beberapa pertimbangan. Kedua algoritma ini dikenal mudah dipahami serta diimplementasikan. KNN hanya membutuhkan parameter k dan metrik jarak, sedangkan Naïve Bayes menawarkan kecepatan serta akurasi tinggi dalam mengklasifikasikan data berukuran besar [3].

Berdasarkan hal tersebut, penelitian ini bertujuan untuk melakukan perbandingan performa K-Nearest Neighbors dan Naïve Bayes dalam klasifikasi kualitas udara. Melalui analisis ini, diharapkan dapat diperoleh metode terbaik yang dapat diterapkan untuk pemantauan kualitas udara, sehingga berkontribusi pada upaya pengelolaan lingkungan yang lebih baik. Selain itu, hasil penelitian ini juga berpotensi memberikan panduan bagi pengembang sistem pemantauan lingkungan dalam memilih algoritma yang paling sesuai untuk aplikasi mereka.

Dengan latar belakang tersebut, penulis tertarik untuk melakukan penelitian berjudul “**Perbandingan Analisis Algoritma K-Nearest Neighbors dan Naïve Bayes dalam Klasifikasi Kualitas Udara**” guna memberikan kontribusi terhadap solusi permasalahan polusi udara yang semakin meningkat di Indonesia.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa model *K-Nearest Neighbor (KNN)* dan *Naïve Bayes* dalam melakukan klasifikasi kualitas udara?
2. Model manakah di antara *K-Nearest Neighbor (KNN)* dan *Naïve Bayes* yang memberikan hasil terbaik dalam klasifikasi kualitas udara?

Rumusan masalah ini bertujuan untuk mengeksplorasi efektivitas kedua algoritma dalam konteks klasifikasi kualitas udara serta memberikan rekomendasi berdasarkan hasil analisis tersebut.

1.3 BATASAN MASALAH

Dalam penelitian ini, terdapat beberapa batasan masalah yang perlu diperhatikan untuk menjaga fokus dan kejelasan tujuan penelitian:

1. **Algoritma yang Digunakan:** Penelitian ini hanya akan menggunakan dua algoritma, yaitu K-Nearest Neighbors (KNN) dan Naïve Bayes, sebagai metode

2. klasifikasi kualitas udara. Algoritma lain tidak akan dipertimbangkan dalam analisis ini.
3. **Sumber Data:** Data yang digunakan dalam penelitian ini merupakan dataset publik yang diambil dari situs web KAGGLE. Penelitian ini tidak akan mencakup pengumpulan data primer atau data dari sumber lain.
4. **Tools yang Digunakan:** Analisis dan implementasi algoritma akan dilakukan menggunakan Jupyter Notebook.
5. **Fokus pada Klasifikasi Kualitas Udara:** Penelitian ini akan berfokus pada klasifikasi kualitas udara berdasarkan parameter yang tersedia dalam dataset, tanpa membahas faktor-faktor eksternal lainnya yang mungkin mempengaruhi kualitas udara.

1.4 TUJUAN DAN MANFAAT PENELITIAN

Untuk memberikan pemahaman yang jelas, tujuan dan manfaat dari penelitian ini akan dipaparkan sebagai berikut:

1.4.1 TUJUAN PENELITIAN

1. Memberikan informasi mengenai hasil performa dari model *K-Nearest Neighbors* dan *Naïve Bayes*.
2. Menemukan model terbaik dalam klasifikasi udara.

1.4.2 MANFAAT PENELITIAN

Menghasilkan model machine learning terbaik yang dapat diimplementasikan untuk melakukan klasifikasi udara.

1.5 SISTEMATIKA PENULISAN

Berikut penyajian sistematika penulisan dari penelitian ini agar mempermudah dalam memahami penulisan laporan:

BAB 1 : PENDAHULUAN

Bab ini menguraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II : LANDASAN TEORI

Dalam landasan teori ini, bab memuat konsep konsep teori yang digunakan sebagai kerangka atau dasar yang digunakan untuk mendukung pemahaman penelitian yang penulis lakukan.

BAB III : METODOLOGI PENELITIAN

Dalam bab ini berisi metode pengumpulan data, prosedur penelitian, dan metode analisis berupa pendekatan penyelesaian masalah yang dilakukan untuk mendukung penelitian.

BAB IV : ANALISIS dan HASIL

Pada bab ini berisi tentang gambaran umum tentang data yang akan di analisis dan bab ini membahas hasil analisis dari pengujian data yang dilakukan dan hasil yang dicapai dari pengujian data yang telah dites

BAB V : KESIMPULAN dan SARAN

Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian mengenai perbandingan algoritma *K-Nearest Neighbors (KNN)* dan *Naïve Bayes* dalam klasifikasi kualitas udara. Selain itu, bab ini juga memuat saran-saran yang dapat menjadi bahan pertimbangan untuk penelitian selanjutnya agar dapat menghasilkan hasil yang lebih baik dan optimal.