

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Air minum yang bersih dan aman adalah kebutuhan sehari-hari mendasar bagi setiap manusia maupun makhluk hidup lainnya. Namun, banyak orang diberbagai belahan dunia masih kesulitan untuk mendapatkan akses terhadap air minum yang layak. Menurut laporan WHO tahun 2020, lebih dari satu miliar orang yang tidak memiliki akses ke air minum bersih. Situasi ini dapat memicu berbagai masalah kesehatan yang serius. Di negara berkembang, penyakit seperti diare yang disebabkan oleh buruknya kualitas air minum menjadi salah satu penyebab utama kematian pada anak-anak [1].

Dataset kualitas air minum mencakup berbagai macam parameter fisik dan kimia yang digunakan untuk menentukan kelayakan air minum. Beberapa parameter penting meliputi pH, total padatan terlarut(TDS), kekerasan air, kadar kloramin, sulfat, daya hantaran listrik dan tingkat kekeruhan. Parameter-parameter ini memberikan informasi tentang mengenai kontaminasi organik maupun nonorganik dalam air. Selain itu, data tersebut juga mencerminkan kualitas air yang berpengaruh terhadap kelayakan konsumsi. Hal ini penting untuk memastikan air aman dan sehat untuk diminum [2].

Masalah *missing value* sering di temui pada berbagai Kumpulan data yang tersedia untuk umum atau public, termasuk dalam penelitian kualitas air minum. Banyak dataset yang dikumpulkan melalui survei, pengukuran otomatis, serta metode lain yang cenderung memiliki data yang tidak lengkap. Ketidaklengkapan data ini dapat disebabkan oleh kesalahan sistem, pengukuran yang tidak akurat, atau gangguan teknis. Kehadiran dari *missing value* ini dapat mengurangi akurasi analisis yang dilakukan. Selain itu, hal ini tidak hanya menurunkan kualitas hasil analisis, tetapi juga dapat menyebabkan model prediksi menjadi kurang optimal ditangani dengan baik[3].

Untuk menangani masalah *missing value*, Pendekatan *Center of Tendency* merupakan salah satu metode umum yang digunakan untuk mengatasi masalah *missing value* dari berbagai macam dataset. Metode ini memanfaatkan mean, median, atau modus dari data yang tersedia untuk mengisi data yang hilang. Teknik ini sederhana dan efektif dalam menjaga struktur data tanpa menyebabkan perubahan besar pada pola data. Dengan cara ini, nilai yang hilang dapat diperkirakan berdasarkan data yang sudah tersedia, sehingga analisis lebih lanjut dapat dilakukan dengan lebih akurat. Metode ini memperkirakan nilai yang hilang berdasarkan data yang ada, sehingga analisis lebih lanjut dapat dilakukan dengan baik dan tepat[4].

Penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam penanganan *missing value* pada dataset kualitas air minum. Dengan menggunakan metode *Center of Tendency*, diharapkan kualitas data akan menjadi lebih baik, sehingga hasil analisis akan menjadi lebih akurat. Selain itu, penerapan metode ini juga dapat dijadikan sebagai referensi untuk pengolahan data pada bidang lain yang menghadapi permasalahan yang serupa. Metode ini menawarkan pendekatan yang sederhana namun efektif untuk meningkatkan integritas data. Harapannya, penelitian ini dapat memberikan manfaat yang luas dalam berbagai konteks pengolahan data.

## 1.2 Rumusan Masalah

Berdasar latar belakang masalah diatas, maka rumusan masalah pada penelitian ini sebagai berikut:

1. Bagaimana pengaruh metode imputasi *missing value* menggunakan pendekatan *Center of Tendency* terhadap performa model klasifikasi potabilitas air minum?
2. Bagaimana perbandingan akurasi model klasifikasi (SVM, *Naïve Bayes*, dan *Decision Tree*) setelah diterapkan metode imputasi dengan pendekatan *Center of Tendency*?
3. Metode imputasi *Center of Tendency* (Mean, Median, dan Modus) manakah yang memberikan hasil terbaik dalam meningkatkan akurasi model

klasifikasi *potabilitas* air minum?

### 1.3 Batasan Masalah

Dari rumusan masalah yang di tetapkan dalam penelitian ini, maka penulis membuat batasan masalah yang diteliti sebagai berikut :

1. Penggunaan algoritma menggunakan parameter default.
2. Menggunakan dataset public yang di dapatkan dari situs Kaggle.
3. Proses Penanganan *missing value* *Center of Tendency* (mean, median, dan modus).
4. Penelitian ini menggunakan *hyper parameter* secara default.

### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan penilitan ini yang hendak dicapai sebagi berikut:

1. Mengatasi permasalahan *missing value* pada dataset menggunakan pendekatan *Center of Tendency*(mean, median, dan modus).
2. Mengetahui tingkat akurasi model yang dihasilkan setelah dilakukan penanganan *missing value* dengan metode *Center of Tendency*.
3. Mengevaluasi bagaimana kinerja metode *Center of Tendency* dalam meningkatkan kualitas data, sehingga dari kualitas datanya dapat menghasilkan model prediksi yang optimal.

### 1.5 Manfaat Penelitian

Melalui penelitian ini, manfaat yang dapat di ambil adalah:

1. Menyediakan referensi untuk menangani *missing value* pada dataset kualitas air minum, yang dapat digunakan dalam berbagai penilitan tentang kualitas air dan *missing value*.
2. Memberikan wawasan tentang seberapa efektif metode *Center of Tendency* dalam mempertahankan pola data aslinya tanpa ada perubahan signifikan dalam pengolahan data.
3. Memudahkan analisis data yang memiliki *missing value* sehingga dapat

menghasilkan akurasi yang lebih akurat.

4. Medukung penggunaan Teknik penangan data yang sederhana dan efektif dalam penanganan *missing value* tanpa menggunakan metode atau teknik yang kompleks.
5. Memberikan panduan praktis yang dapat di terapkan pada dataset di bidang lainya yang menghadapi masalah yang serupa dalam menangani *missing value*.
6. Memberitahukan orang lain tentang betapa pentingnya menangani *missing value* untuk kemajuan teknologi analisis data.

