

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kehidupan sehari – hari tidak lepas dari konsumsi air di karenakan air merupakan kebutuhan bagi kehidupan manusia karena memastikan keberlangsungan hidup manusia dan makhluk hidup lainnya di Bumi. Namun, beberapa air tidak aman untuk dikonsumsi, sehingga diperlukan identifikasi jenis air mana yang aman untuk dikonsumsi (Nazar Yuniar 2023). Pembangunan yang semakin hari makin gencar di pedesaan maupun di lingkup kota dapat mempengaruhi kualitas dan kelayakan, Air limbah domestik limbah yang berasal dari usaha dan atau kegiatan di permukiman, seperti rumah makan, perkantoran, bisnis, apartemen juga ikut andil dalam kontribusi untuk pencemaran (Sulistia and Septisya 2020), yang mana dapat mempengaruhi kualitas dan kelayakan konsumsi air.

Tubuh menikmati banyak manfaat dari minum air putih, tetapi air yang di minum harus sesuai dengan peraturan. Beberapa dia antaranya ialah terkait dengan *pH*, Di angka 7 *pH* di anggap netral lebih rendah dari netral air di anggap lebih asam sedangkan jika lebih dari 7 di anggap basa, Air yang layak untuk di konsumsi tidak memiliki bau, rasa ataupun warna untuk jenis air putih atau air mineral, suhu juga mempengaruhi dapat atau tidaknya air untuk di konsumsi. Dalam pendekatan dalam uji kelayakan air dalam penelitian ini di gunkan parameter-parameter antara lain yaitu *pH*, *Hardness*, *Solids*, *Chloramines*, *Sulfate*, *Conductivity*, *Organi Carbon*, *Trihalomethanes*, *Turbidity*, parameter tersebut dapat di gunakan untuk menentukan apakah air layak atau tidak untuk konsumsi, pada era digital ini *machine learning* adalah metode pendekatan yang efektif untuk memecahkan berbagai macam masalah atantara lain yaitu mengklasifikasikan air.

Pendekatan pengenalan kelayakan air dengan *Machine Learning* melibatkan beberapa masalah yang perlu diperhatikan. Berikut adalah beberapa masalah umum yang mungkin dihadapi dalam pengenalan kelayakan air menggunakan *Machine Learning* antara lain kualitas data, Data yang tidak lengkap,

tidak akurat, atau tidak representatif dapat menghasilkan model yang tidak dapat diandalkan. Pemilihan fitur yang relevan dan informatif sangat penting dalam membangun model yang efektif untuk pengenalan kelayakan air. Pemilihan fitur yang kurang tepat dapat mengurangi kinerja model. *Machine learning* terbagi menjadi beberapa macam salah satunya adalah *Supervised Learning* dimana pembelajaran modelnya sudah berlabel sebelumnya dan *Unsupervised learning* merupakan pembelajaran model yang di gunakan dalam pengelompokan berdasarkan kemiripan, dalam melakukan klasifikasi menggunakan pendekatan *Supervised Learning* dengan menggunakan algoritma k-NN memberikan kemudahan lebih efisien dalam melakukan pengelompokan.

Algoritma k-NN (*k-Nearest Neighbors*) merupakan salah satu metode klasifikasi yang sangat populer (Siringoringo 2018) yang sudah umum di gunakan dalam melakukan klasifikasi, Selain populer k-NN juga sederhana sehingga mudah untuk di aplikasikan kemudahana ini merupakan salah satu dari keunggulanya, Algoritman k-NN ini sendiri merupakan *intance based learning* atau data tanpa memerlukan data latih terlebih dahulu. Selain itu, Algoritma ini merupakan salah satu metode *lazy learning* (Cahyanti, Rahmayani, and Ainy 2020), yang mana pendekatan dalam *machine learning* model tidak melakukan proses pelatihan secara *eksplisit* pada data pelatihan yang diberikan. Sebaliknya, Model hanya menyimpan data pelatihan dan melakukan komputasi saat ada permintaan untuk membuat prediksi atau melakukan klasifikasi terhadap data baru hal ini memberikan keuntungan untuk sebuah uji penelitian air dimana sudah terdapat atau memiliki data yang sudah terlabel atau terkelompokan sebelumnya.

Pemilihan algoritma k-NN pada penelitian ini di landasi dengan kemudahan dalam pengaplikasiannya selain itu algoritma k-NN biasanya menghasilkan hasil yang baik dalam kasus klasifikasi yang sederhana dan ketika batas keputusan antar kelasnya relatif jelas, Walaupun terdapat juga beberapa kelemahan yang terdapat pada algoritma k-NN salah satunya kurang efisien untuk data berdimensi tinggi akan tetapi penelitian yang di lakukan oleh (Syarifuddin 2023) dengan menggunakan nilai $k = 3$. mendapatkan tingkat akurasi 80% berdasarkan perhitungan Confusion Matrix. Dan (Hardiana Said, Nur Hafifah Matondang 2022) mendapat akurasi tertinggi 85,24% dengan nilai $k = 3$ dari beberapa penelitian

terdahulu algoritma k-NN mampu melakukan klasifikasi untuk prediksi kualitas kelayakan air yang layak di minum.

1.2. Rumusan Masalah

Dari pemaparan latar belakang di atas dapat ditarik beberapa kesimpulan untuk rumusan masalah, adapun rumusan masalah sebagai berikut :

1. Bagaimana penerapan algoritma k-Nearest Neighbors (k-NN) dapat membantu dalam mengklasifikasikan kualitas air berdasarkan parameter-parameter kelayakan konsumsi?
2. Bagaimana performa algoritma k-NN dalam mengidentifikasi jenis air yang aman untuk dikonsumsi?

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma k-Nearest Neighbors (k-NN) dalam mengklasifikasikan kualitas air berdasarkan parameter-parameter kelayakan konsumsi
2. Mengetahui performa algoritma k-NN dalam mengidentifikasi jenis air yang aman untuk dikonsumsi.

1.4. Manfaat Penelitian

Dari penelitian ini dapat ditarik kesimpulan terkait dengan manfaat yang dapat sebagai berikut:

1.4.1. Manfaat Praktis

Adapun manfaat praktis sebagai berikut:

1. Menyediakan informasi yang berguna bagi lembaga riset, dan masyarakat dalam mengambil keputusan terkait pengelolaan sumber daya air.

1.4.2. Manfaat Teoritis

Adapun manfaat teoritis sebagai berikut:

1. Penggunaan Algoritma Machine Learning: Teknologi machine learning, seperti algoritma k-Nearest Neighbors (k-NN), dapat digunakan untuk mengklasifikasikan kualitas air berdasarkan data yang telah terlabel

1.5. Batasan Masalah

Adapun batasan – batasan masalah pada penelitaian ini adalah sebagai berikut:

1. Penelitian ini membatasi fokus pada penerapan algoritma k-NN dalam mengklasifikasikan kualitas air berdasarkan kelayakan konsumsi.
2. Data yang digunakan adalah data kualitas air yang sudah terlabel atau terkelompokkan sebelumnya yang di dapatkan dari *kaggle.com*
3. Evaluasi terhadap akurasi algoritma k-NN dibatasi pada klasifikasi jenis air yang layak dan tidak layak untuk dikonsumsi.
4. Tidak membahas terkait keamanan sistem yang di rancang

UNUGIRI