

BAB I

PENDAHULUAN

1.1. Latar Belakang

Air merupakan sebuah sumber kehidupan bagi setiap makhluk hidup dan menjadi kebutuhan penting bagi setiap manusia, khususnya masyarakat di wilayah dengan tingkat penduduk yang tinggi (perkotaan). Kebutuhan air bersih setiap tahunnya mengalami peningkatan namun ketersediaan air bersih malah terbatas, hal ini disebabkan pembangunan yang semakin banyak tanpa memperhatikan keseimbangan lingkungan sekitar dan sempitnya daerah resapan terutama di daerah perkotaan, hal ini menimbulkan masalah yang serius yaitu kurangnya ketersediaan sumber air bersih.

Di Indonesia sendiri pencemaran air menjadi masalah utama, sumber utama dari pencemaran air yang banyak dihasilkan dari limbah rumah tangga serta domestik, hal ini menyebabkan menurunnya kualitas air dan ketersediaan air bersih di setiap daerah. Akibat dari kualitas air yang buruk dapat menimbulkan penyakit yang cukup serius salah satunya ialah penyakit diare, menurut (Rieke Saraswati, 2020) di setiap tahunnya diperkirakan kurang lebih ada 800 ribu orang yang meninggal akibat diare, hal ini ditimbulkan akibat dari konsumsi air yang kurang bersih atau terkontaminasi, kurangnya kebersihan tangan saat mengkonsumsi air, dan masalah sanitasi yang ada pada air yang diminum.

Menurut (Littia Obeng Soemarwoto, 1991) yang merupakan anggota dari organisasi internasional Program Lingkungan Perserikatan Bangsa-Bangsa (*UN Environment Programme*) mengutarakan opininya mengenai pentingnya peranan air yaitu pasokan air bersih yang memadai merupakan faktor yang sangat penting dalam rangka meningkatkan kesejahteraan umat manusia secara bersama-sama. Mengingat peranannya yang sangat penting, maka upaya peningkatan kualitas hidup harus mencakup upaya pemeliharaan, pengaturan, pemanfaatan dan

pengembangan sarana penyediaan air minum, khususnya untuk kebutuhan manusia. Kehidupan dan derajat kesehatan manusia dan lingkungan.

Untuk mengetahui kualitas air yang layak dikonsumsi bagi masyarakat luas, perlu adanya identifikasi dini yang dapat dijadikan sebagai informasi awal. Peneliti menggunakan teknik *data mining* untuk mengidentifikasi dini masalah ini, yang mana teknik *data mining* banyak digunakan dalam melakukan analisis data dengan jumlah besar yang tersimpan pada *database*, potensinya sendiri sudah banyak diketahui secara luas, *data mining* banyak membantu permasalahan secara umum dengan memanfaatkan data yang ada serta relevan dan menghasilkan informasi baru yang mungkin belum pernah terpikirkan berdasarkan identifikasi pola-pola di atribut yang ada di dalam dataset.

Pada *data mining* sendiri mempunyai cukup banyak teknik dalam mengolah data yang mana salah satunya teknik klasifikasi. Klasifikasi adalah proses dalam menemukan model atau fungsi yang bisa menjelaskan serta mencirikan kelas data yang dapat digunakan untuk kepentingan tertentu. Dalam penelitian ini teknik klasifikasi yang digunakan bertugas untuk melakukan prediksi suatu kategori label yang sebelumnya tidak diketahui serta membedakan objek-objek yang satu dengan lainnya, pada teknik ini juga terdapat algoritma klasifikasi yang dapat digunakan seperti prediksi dan lainnya, yang mana algoritma ini salah satunya adalah *K-Nearest Neighbor*.

Peneliti memilih teknik klasifikasi *data mining* dengan menggunakan Algoritma *K-Nearest Neighbor*, dikarenakan metode klasifikasi dengan Algoritma *K-Nearest Neighbor* ini dianggap lebih tepat dengan perhitungan akurasi tinggi yaitu sebesar sehingga cukup akurat dalam melakukan prediksi, selain itu Algoritma *K-Nearest Neighbor* juga lebih tepat dalam menghitung atribut data yang bersifat numerik secara dominan. Kontribusi dari penerapan algoritma *K-Nearest Neighbor* terhadap lingkungan itu sendiri cukup besar hal ini termasuk untuk peramalan curah hujan, banjir atau bahkan perubahan iklim.

Kelebihan dari metode klasifikasi Algoritma *K-Nearest Neighbor* sendiri adalah mudah dipahami dan mudah untuk diimplementasikan, hal ini dikarenakan dalam mengklasifikasinya cukup menentukan fungsi untuk menghitung jarak antar-*instance* yang mana dalam menentukan kelas yang tepat dengan hanya melihat kelas label mana yang paling banyak muncul dalam *k instance* terdekat, selain itu algoritma ini juga kuat atau tangguh dalam sehubungan dengan data sampel pelatihan yang *noisy*, dan cukup efektif ketika data training yang dimiliki sampelnya besar.

Berdasarkan pembahasan pada latar belakang diatas maka peneliti sangat tertarik untuk dapat merancang sistem prediksi kualitas air berdasarkan kumpulan data yang merupakan data sekunder bersumber dari *website kaggle (www.kaggle.com)* yang diupload oleh MsSmartyPants pada tahun 2021 dengan judul *Water Quality (Dataset for water quality classification)*. Di mana dataset tersebut dapat digunakan untuk memberikan informasi mengenai kualitas air yang aman dikonsumsi(dipakai) dan tidaknya yang diambil dari lingkungan perkotaan. Total data yang terdapat didataset berjumlah 7999 record dan dilakukan data balancing sebanyak 1824 data, sampling dilakukan berdasarkan nilai akurasi yang dapat dihasilkan cukup tinggi dengan presentase 85.24% dan perbandingan nilai *recall* dan *precision* yang lebih stabil pada tabel *confusion matrix* serta menghindari *overfitting*. Data yang dipakai merupakan dataset baru dengan akurasi yang lebih baik dan rekam data serta memiliki nilai yang lebih konsisten dibandingkan dengan dataset yang lama.

Setelah model dari KNN didapatkan, selanjutnya sistem prediksi dengan berbasis website akan dibuat dengan melalui proses SDLC (*System Development Life Cycle*) dengan metode *waterfall*. Perancangan sistem dimulai dari analisis kebutuhan menggunakan PIECES (*Performance, Information, Economy, Control, Efficiency, dan Service*), dengan adanya sistem usulan dari analisis kebutuhan dengan metode analisis PIECES rancangan sistem dan pengimplementasian sistem dapat dilakukan hingga pada proses *testing* sistem. Semua ini dilakukan dengan harapan peneliti lain

atau masyarakat umum yang memiliki alat-alat test dapat melakukan prediksi kualitas air secara mudah dan cepat sehingga mempermudah dalam melakukan prediksi kualitas air yang aman dikonsumsi atau tidaknya.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah di atas peneliti dapat merumuskan masalah sebagai berikut:

- a. Bagaimana merancang sistem untuk memprediksi kualitas air yang baik untuk dikonsumsi atau tidak dengan menerapkan Algoritma *K-Nearest Neighbor* berdasarkan data?
- b. Bagaimana performa/kinerja model dari Algoritma *K-Nearest Neighbor* dalam memprediksi kualitas air yang baik untuk dikonsumsi?

1.3. Ruang Lingkup Penelitian

Untuk mendapatkan hasil yang terbaik dalam penelitian ini, peneliti membatasi ruang lingkup dalam penelitian ini sebagai berikut:

1. Data yang digunakan bersumber dari website *www.kaggle.com* yang diupload oleh MsSmartyPants pada tahun 2021 dengan judul *Water Quality (Dataset for water quality classification)*. Variabel yang digunakan pada penelitian ini untuk memprediksi kualitas air yaitu sejumlah 21 atribut dan sudah termasuk kelas/label yaitu atribut kimia *aluminium* (mg/L), *ammonia* (mg/L), *arsenic* (mg/L), *barium*(mg/L), *cadmium*(mg/L), *chloramine*(mg/L), *chromium* (mg/L), *copper* (mg/L), *flouride* (mg/L), *lead* (mg/L), *nitrate* (mg/L), *nitrite* (mg/L), *mercury* (mg/L), *perchlorate* (mg/L), *radium* (mg/L), *selenium* (mg/L), *silver* (mg/L), *uranium* (mg/L), atribut biologis *bacteria* dan *viruses* (mg/L), serta *is_safe* (kelas/label dari variabel yang ada).
2. Pengukuran performa Algoritma *K-Nearest Neighbor* menggunakan tabel *confusion matrix*.

3. Membangun sebuah sistem untuk memprediksi kualitas air dengan menerapkan Algoritma *K-Nearest Neighbor*.
4. Perancangan sistem yang akan dibangun dengan pemodelan visual prosedural dari metode DFD (*Data Flow Diagram*).
5. Sistem yang akan dibangun berbasis bahasa pemrograman *Python* dan menggunakan *framework Flask*.

1.4. Tujuan Penelitian

Dengan didapatkan suatu permasalahan sesuai pembahasan di atas maka tujuan dari penelitian ini adalah:

1. Menghasilkan sistem untuk memprediksi kualitas air sehingga mempermudah peneliti lainnya untuk mendapatkan hasil prediksi.
2. Diterapkannya Algoritma *K-Nearest Neighbor* untuk melakukan suatu prediksi kualitas air yang baik atau tidaknya dengan hasil berupa akurasi dan mengetahui prediksi kualitas air berdasarkan dataset.

1.5. Manfaat Penelitian

Dengan begitu dapat diuraikan berdasarkan rumusan masalah dan ruang lingkup yang ada, penelitian ini diharapkan dapat memberikan sebuah manfaat sebagai berikut:

1. Bagi Program S1 Sistem Informasi
Penelitian ini dapat dijadikan sebuah referensi informasi khususnya untuk mahasiswa S1 Sistem Informasi yang kemungkinan sedang melakukan penyusunan tugas akhir.
2. Bagi Peneliti
Penelitian ini dapat dijadikan sebuah sarana dalam mengukur kemampuan dari peneliti dalam hal wawasan dan ilmu yang dimiliki dan didapatkannya selama melakukan proses pembelajaran dalam perkuliahan yang dilewati dengan menerapkannya dalam penelitian tugas akhir ini.

3. Bagi Pembaca

Dengan harapan penelitian ini dapat menyediakan informasi khusus kepada pembaca dalam hal untuk menentukan apakah kualitas air di daerah anda atau sekitar anda baik untuk dikonsumsi.

1.6. Luaran yang diharapkan

Dengan harapan luaran yang diberikan dapat menghasilkan suatu sistem untuk memprediksi kualitas air yang baik dikonsumsi dan mengetahui tingkat akurasi dari prediksi tersebut dengan menerapkan metode klasifikasi dengan Algoritma *K-Nearest Neighbor*. Sistem ini ditujukan untuk mempermudah peneliti lain yang ingin melakukan uji coba kualitas air di beberapa tempat berdasarkan parameter kimia dan biologis dari dataset yang bersifat universal.

1.7. Sistematika Penulisan

Peneliti dalam melakukan penulisan tugas akhir ini menggunakan sistematika penulisan sebagai berikut.

BAB I PENDAHULUAN

Pada bab ini berisikan tentang latar belakang, rumusan masalah, ruang lingkup penelitian, tujuan penelitian, manfaat penelitian, dan luaran yang diharapkan, serta sistematika penulisan dari penelitian ini.

BAB II TINJAUAN PUSTAKA

Bab ini berisikan uraian dari teori yang digunakan sebagai acuan dasar penyusunan tugas akhir ini yang mendukung judul dari kegiatan penelitian ini yang bersumber dari jurnal, buku, website resmi atau prosiding.

BAB III METODOLOGI PENELITIAN

Bab ini berisikan tahap penelitian, uraian penelitian, kerangka berpikir, alat dan bahan bantu penelitian serta tahapan kegiatan yang meliputi waktu dan tempat pelaksanaan, dan jadwal kegiatan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini dilakukannya pembahasan mengenai penerapan algoritma *K-Nearest Neighbor* dari untuk memprediksi kualitas air. Hasil analisis yang telah diolah oleh peneliti akan menghasilkan sistem prediksi yang disesuaikan dengan data yang telah diolah pada proses *data mining* dan sistem dapat menghasilkan prediksi dengan harapan cukup akurat dalam memprediksi data baru.

BAB V PENUTUP

Pada bab ini berisikan uraian kesimpulan dari kegiatan penelitian yang telah dilakukan, serta informasi yang memuat saran penulis untuk memberikan informasi ke penelitian lainnya atau selanjutnya.

DAFTAR PUSTAKA

Sumber-sumber dari penelitian terkait yang telah dilakukan.

RIWAYAT HIDUP

LAMPIRAN