

Metode KNN (*K-Nearest Neighbor*) untuk Menentukan Kualitas Air

Fahmi Malik Namus Akbar

Teknik Industri, Informatika, Universitas Pembangunan Nasional Yogyakarta, Indonesia
Email: fahmimna@gmail.com

Abstrak—Penelitian ini bertujuan untuk membuat aplikasi yang lebih inovatif dari *metode K-Nearest Neighbor* (KNN) dalam menentukan kualitas air dengan mempertimbangkan 20 unsur kimia yang relevan. Kualitas air adalah aspek penting dalam kehidupan sehari-hari, dan upaya untuk memastikan air yang aman untuk dikonsumsi merupakan prioritas. Penelitian ini mengajukan pertanyaan utama: "Apakah metode KNN dapat memberikan pendekatan yang efektif dalam mengklasifikasikan kualitas air?" Metodologi penelitian ini melibatkan penggunaan dataset sekunder mengenai kualitas air dan menerapkan algoritma KNN untuk mengkategorikan data tersebut. Hasil dari metode KNN dibandingkan dengan metode klasifikasi lain yang umum digunakan dalam analisis kualitas air. Penelitian ini bertujuan untuk memberikan pandangan yang lebih dalam dan pemahaman yang lebih baik tentang potensi metode KNN. Berbeda dengan penelitian sebelumnya, penelitian ini membawa kebaruan dengan pendekatan yang lebih canggih dan mempertimbangkan lebih banyak unsur kimia daripada yang umumnya digunakan dalam penelitian sebelumnya. Hasil penelitian ini diharapkan dapat memberikan wawasan baru tentang efektivitas metode KNN dalam mengklasifikasikan kualitas air yang dapat membantu masyarakat dan pemerintah dalam mengambil tindakan respons yang lebih efisien terkait dengan potensi kontaminasi dan masalah kualitas air lainnya. Dengan demikian, penelitian ini memiliki potensi untuk memberikan kontribusi berharga pada pemahaman kita tentang kualitas air dan pemanfaatan metode KNN dalam analisisnya.

Kata Kunci: KNN, Air, Kualitas, Klasifikasi, Algoritma

Abstract— This research aims to make a more innovative application of the K-Nearest Neighbor (KNN) method in determining water quality by considering 20 relevant chemical elements. Water quality is a crucial aspect of daily life, and ensuring safe drinking water is a top priority. This study poses the main question: "Can the KNN method provide an effective approach to classifying water quality?" The research methodology involves using secondary datasets on water quality and implementing the KNN algorithm to categorize this data. The results obtained using the KNN method are compared with other commonly used classification methods in water quality analysis. The objective of this research is to provide a deeper insight and better understanding of the potential of the KNN method. In contrast to previous studies, this research introduces novelty through a more sophisticated approach and considers a broader range of chemical elements than typically utilized in earlier research. The findings of this study are expected to offer fresh insights into the effectiveness of the KNN method in classifying water quality, which can aid both the public and government authorities in taking more efficient response actions related to potential contamination and other water quality issues. Thus, this research has the potential to make a valuable contribution to our understanding of water quality and the utilization of the KNN method in its analysis.

Keywords: KNN, Water, Quality, Classification, Algorithm

1. PENDAHULUAN

Air adalah unsur yang mendominasi wilayah terluas di planet bumi, mencakup hampir 71% dari seluruh permukaan dan ada di berbagai bentuk, termasuk air laut, air sungai, air permukaan, air atmosfer, air rawa, dan air tanah [1]. Selain memenuhi kebutuhan dasar manusia seperti konsumsi dan memasak, air juga memiliki peran penting dalam industri, pertanian, dan pembangunan infrastruktur.

Kualitas air memiliki dampak yang sangat signifikan pada kehidupan kita. Faktor-faktor lingkungan fisik, seperti pola curah hujan, topografi, dan jenis batuan, berperan dalam menentukan kuantitas air di suatu wilayah. Sementara itu, kualitas air sangat dipengaruhi oleh aspek-aspek sosial seperti kepadatan penduduk dan tingkat interaksi sosial. Di tengah kompleksitas ini, perhatian terhadap kualitas air menjadi semakin penting. Pembangunan yang pesat seringkali mengubah aliran sungai menjadi area permukiman dan industri di kota-kota, yang pada gilirannya dapat mengancam kualitas air yang mengalir di sepanjang sungai tersebut [2].

Di Indonesia, negara kepulauan yang kaya akan sumber daya alam, air memiliki peran kunci dalam mendukung berbagai sektor kehidupan. Namun, sayangnya, situasi kualitas air di Indonesia menghadapi tantangan yang serius. Menurut peneliti Senior di Southeast Asian Food and Agriculture and Technology (Seafast) Center IPB University, Ratih Dewanti Hariyadi, dari 34 provinsi yang ada di Indonesia, 10 di antaranya masih memiliki indeks kualitas air yang buruk akibat berbagai kontaminasi. Masalah ini tidak hanya mempengaruhi sektor pertanian dan industri, tetapi juga mengancam kesehatan masyarakat secara keseluruhan.

Salah satu tantangan utama dalam mengatasi masalah kualitas air adalah keterbatasan akses terhadap air bersih dan aman. Data menunjukkan bahwa sebagian besar rumah tangga di Indonesia, yakni 7 dari 10 rumah tangga, mengkonsumsi air yang terkontaminasi oleh bakteri *E. coli*. Hanya sekitar 11,9 persen rumah tangga yang

memiliki akses terhadap air yang aman untuk dikonsumsi. Hal ini menekankan pentingnya memiliki sistem yang dapat memantau dan menilai kualitas air dengan akurat untuk mengidentifikasi masalah sejak dini.

Oleh karena itu, untuk mengatasi masalah ini, diperlukan pendekatan ilmiah yang efisien dan inovatif dalam mengolah data terkait kualitas air. Salah satu metode alternatif yang muncul adalah data mining, suatu teknik pengolahan data yang memungkinkan pengurangan data melalui teknik katalogisasi, klasifikasi, dan segmentasi. Data mining telah digunakan dalam berbagai bidang dan menawarkan potensi besar dalam memahami serta mengekstraksi informasi yang relevan dari data.

Penelitian tentang klasifikasi air minum telah banyak dilakukan, terutama mengenai klasifikasi kualitas air minum menggunakan machine learning. Penelitian oleh Aldi dkk. (2023) menggunakan metode Naive Bayes, Decision Tree, dan K-Nearest Neighbors untuk menemukan metode yang paling akurat. Hasil penelitian menunjukkan bahwa metode Decision Tree memiliki akurasi tertinggi [3]. Penelitian oleh Prismahardi dkk. (2023) menggunakan metode Support Vector Machine, Decision Tree, Naïve Bayes, dan Artificial Neural Network. Hasil penelitian menunjukkan bahwa metode Random Forest Classifier memiliki akurasi tertinggi [4]. Kemudian penelitian yang dilakukan oleh Ilyas dkk yang bertujuan untuk mengklasifikasikan kualitas air minum menggunakan algoritma Random Forest. Data yang digunakan adalah data kualitas air minum dari 267 sampel air sumur di Jakarta. Hasil penelitian menunjukkan bahwa algoritma Random Forest dapat mengklasifikasikan kualitas air minum dengan akurasi sebesar 82,3% [5]. Terakhir adalah penelitian yang dilakukan oleh Armono dkk yang bertujuan untuk mengidentifikasi kualitas air menggunakan algoritma K-Nearest Neighbor. Data yang digunakan adalah data kualitas air minum dari 200 sampel air sumur di Kota Surakarta. Hasil penelitian menunjukkan bahwa algoritma K-Nearest Neighbor dapat mengidentifikasi kualitas air dengan akurasi sebesar 82,6% [6]. Dari sumber penelitian sebelumnya tersebut, peneliti tertarik menggunakan metode KNN karena dapat disimpulkan bahwa metode KNN dapat digunakan untuk mengklasifikasikan kualitas air minum dengan akurasi mencapai hingga 82%.

Dalam konteks ini, penelitian ini bertujuan untuk membuat aplikasi yang lebih inovatif dari metode K-Nearest Neighbor (KNN) menggunakan MATLAB dalam menentukan kualitas air dengan mempertimbangkan 20 unsur kimia yang relevan. Efektivitas metode KNN akan dievaluasi dalam mengklasifikasikan data kualitas air yang diharapkan dapat memberikan wawasan yang lebih baik dan membantu dalam mengatasi tantangan kualitas air yang dihadapi oleh masyarakat Indonesia dan negara-negara lainnya.

KNN adalah algoritma pembelajaran mesin terawasi yang digunakan untuk klasifikasi dan regresi, memanipulasi data training serta mengklasifikasikan data testing berdasarkan metrik jarak. Dengan mencari K tetangga terdekat dari data uji, algoritme ini melakukan klasifikasi berdasarkan mayoritas label kelas, dan termasuk dalam kelas *instance based learning* serta *lazy learning* [7]. Dengan penerapan teknik *K-Nearest Neighbor* (KNN), merupakan salah satu pendekatan klasifikasi terhadap kumpulan data yang berfokus pada mayoritas kategori. Tujuannya adalah untuk mengategorikan objek baru dengan merujuk pada atribut dan contoh-contoh sampel dari data pelatihan. Dengan demikian, upaya ini bertujuan untuk mendekati akurasi yang lebih tinggi saat melakukan evaluasi pembelajaran [8].

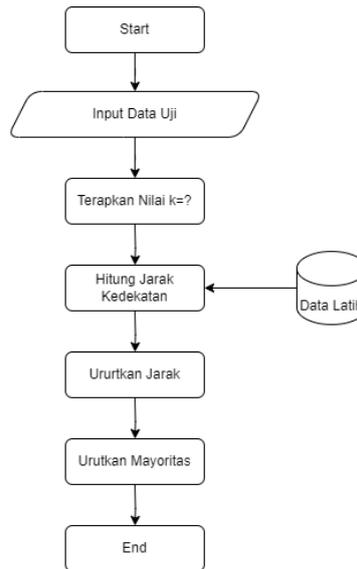
Penelitian ini diharapkan dapat memberikan manfaat yang signifikan bagi masyarakat luas serta pihak-pihak yang memerlukan informasi mengenai tingkat keamanan air. Dataset yang digunakan dalam penelitian ini berasal dari Kaggle, yang merupakan sumber data yang terpercaya dan terkemuka dalam berbagai bidang penelitian. Penelitian ini akan mempertimbangkan berbagai parameter penting dalam menentukan kualitas air, seperti *Aluminium, Ammonia, Arsenic, Barium, Cadmium, Chloramine, Chromium, Copper, Fluoride, Bakteri, Virus, Timbal, Nitrat, Nitrit, Merkuri, Perchlorate, Radium, Selenium, Perak, dan Uranium* Informasi yang diperoleh dari penelitian ini diharapkan akan memberikan pemahaman yang lebih baik tentang kualitas air dan membantu dalam mengidentifikasi potensi masalah terkait dengan kandungan-kandungan tersebut. Hal ini dapat berkontribusi positif dalam upaya menjaga keamanan dan kualitas air yang kita konsumsi.

Manfaat dari penelitian ini adalah memberikan panduan dan acuan bagi kebijakan pengelolaan sumber daya air yang lebih efektif. Dengan memanfaatkan data mining dan algoritma KNN, penelitian ini berupaya menjadi solusi konkret dalam menghadapi tantangan kualitas air yang ada. Hasil dari penelitian ini dapat digunakan sebagai acuan untuk mengambil keputusan yang tepat dalam mengelola dan menjaga kualitas air di berbagai wilayah di Indonesia.

Melalui pendekatan ilmiah yang sistematis dan analisis yang mendalam, diharapkan penelitian ini dapat memberikan kontribusi positif dalam meningkatkan kesadaran masyarakat tentang pentingnya menjaga kualitas air. Selain itu, hasil penelitian ini juga diharapkan dapat digunakan oleh pemerintah, lembaga nirlaba, dan pihak berkepentingan lainnya dalam merumuskan kebijakan dan strategi yang bertujuan untuk menjaga kualitas air yang lebih baik di masa depan. Diharapkan penelitian ini dapat menjadi langkah awal yang berarti dalam menghadapi tantangan kualitas air di masa depan, dan menjadi landasan untuk penelitian lebih lanjut yang berfokus pada upaya pelestarian dan pengelolaan sumber daya air secara berkelanjutan.

2. METODE PENELITIAN

Penelitian ini dimulai dengan mengumpulkan dataset yang diperoleh dari situs Kaggle.com, yang merupakan salah satu basis data dalam kompetisi di bidang penglihatan komputer dan dapat diakses oleh publik dengan judul *qualityWatter1* dalam bentuk ekstensi .csv [9]. Data tersebut kemudian diproses secara *pre-processing* untuk memastikan bahwa data siap digunakan sebagai data target. Total data target yang terkumpul adalah 800 record. Proses klasifikasi melibatkan pencarian model atau fungsi yang dapat menggambarkan serta membedakan antara konsep atau kelas-kelas data. Tujuannya adalah untuk memprediksi kelas dari objek yang tidak memiliki label, sehingga memberikan perkiraan mengenai kelasnya [10]. Data latih selanjutnya dimasukkan ke dalam algoritma K-Nearest Neighbor untuk mengembangkan model klasifikasi. Setelah model klasifikasi terbentuk, dilakukan uji akurasi menggunakan data uji. Gambar 1 menunjukkan alur dari model klasifikasi yang dibangun menggunakan K-Nearest Neighbor. Metode ini digunakan untuk mencari tetangga terdekat dari data uji dan melakukan klasifikasi berdasarkan mayoritas label kelas dari tetangga terdekat tersebut [11].



Gambar 1. Tahapan Model K-NN

2.1 Data Mining

Data mining adalah metode yang digunakan untuk mengidentifikasi pola-pola tertentu dari sejumlah besar data, termasuk berbagai teknik seperti teknik klasifikasi [12]. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery in Databases* (KDD). Proses pencarian pengetahuan ini menggunakan berbagai teknik-teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksikannya. Proses pencarian bersifat iteratif dan interaktif untuk menemukan pola atau model yang benar, baru, bermanfaat, dan dapat dimengerti. Dalam penerapannya, data mining memerlukan berbagai perangkat lunak analisis data untuk menemukan pola dan relasi data agar dapat digunakan untuk membuat prediksi dengan akurat. Data Mining dengan K-Nearest Neighbor biasanya digunakan untuk klasifikasi data yang didasarkan pada tetangga terdekat dari data yang diinginkan (*data Testing*). Beberapa aturan yang ada pada data mining menggunakan *K-Nearest Neighbor* adalah sebagai berikut :

- Tersedia beberapa data (x,y) yang terklasifikasi minimal 2 kelas
- Diberikan *query point* yang akan menentukan kelas dari data yang akan di uji
- Menghitung sejumlah k tetangga terdekat untuk menentukan kelas dari query tersebut berdasarkan *majority vote* (hasil terbanyak)

2.2 Metode KNN (K-Nearest Neighbor)

K-Nearest Neighbor (KNN) adalah suatu algoritma yang digunakan untuk mengklasifikasikan data dengan memanfaatkan data latih (*training record*) dari tetangga terdekat, di mana k merupakan jumlah tetangga terdekat yang digunakan dalam proses klasifikasi [13]. KNN adalah jenis algoritma yang sederhana dibandingkan algoritma lain dalam *machine learning*. Prinsip kerjanya adalah dengan menghitung jarak perbedaan antara data uji dan data latih yang ada dalam sistem, kemudian mencari nilai terkecil atau paling mirip untuk mengklasifikasikan data tersebut [14]. Suatu Objek di klasifikasikan dengan class mayoritas dari class tetangga,

dimana diambil class yang paling banyak muncul dalam batasan K tertentu dari tetangganya, sehingga diperoleh class baru sesuai dengan tetangga yang paling umum, jika $K=1$, maka objek ditetapkan sesuai dengan class tetangga terdekatnya. Pada fase klasifikasi, K adalah sebuah konstanta yang ditetapkan pengguna. Class baru dari suatu data test diklasifikasikan dengan menetapkan class yang paling sering muncul diantara sampel ke titik data training yang ada. Metode *K-Nearest Neighbor* pada prinsip kerjanya mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (Neighbor) terdekatnya dalam data sampel.

Euclidean distance merupakan jarak antara dua titik atau koordinat yang dihitung menggunakan rumus Pythagoras. Ini adalah panjang garis lurus yang menghubungkan dua titik dalam ruang, yang disebut titik a dan titik b. Garis ini, juga dikenal sebagai garis miring, terbentang di antara sumbu x dan sumbu y dengan koordinat yang diberikan untuk titik a dan titik b [15]. Gambar berikut menunjukkan rumus perhitungan untuk mencari jarak antara dua titik yaitu titik pada data training (x) dan titik pada data testing (y) maka digunakan rumus *Euclidean*, seperti pada persamaan berikut.

$$D(x,y) = \sqrt{\sum_{k=1}^n (p_k - q_k)^2} \tag{1}$$

Dengan D adalah jarak antara titik pada data training x dan titik data testing yang akan diklasifikasi, dimana $x=x_1, x_2, \dots, x_i$ dan y_1, y_2, \dots, y_i dan i mempresentasikan nilai atribut serta n merupakan dimensi atribut.

Langkah-langkah untuk menghitung metode Algoritma KNN :

- a. Menentukan parameter K (jumlah tetangga paling dekat).
- b. Menghitung kuadrat jarak Euclid (*query instance*) masing-masing objek terhadap data sampel yang diberikan
- c. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak Euclid terkecil.
- d. Mengumpulkan kategori Y (klasifikasi *nearest neighbor*)
- e. Dengan menggunakan kategori Nearest Neighbor yang paling mayoritas maka dapat diprediksi nilai *query instance* yang telah dihitung.

2.3 Dataset

Dataset adalah kumpulan data yang terdiri dari beberapa entitas atau sampel yang diambil dari berbagai sumber. Dalam konteks penelitian ilmiah, dataset biasanya berisi sekumpulan variabel atau atribut yang diukur atau diamati pada setiap entitas, serta informasi tambahan yang relevan untuk analisis. Dataset merupakan komponen penting dalam penelitian karena menjadi dasar untuk melakukan analisis, pembuatan model, dan menyimpulkan temuan penelitian.

Dalam penelitian mengenai metode KNN untuk menentukan kualitas air, dataset berisi data pengukuran atau observasi berbagai parameter kualitas air, seperti konsentrasi logam berat, kandungan bakteri, dan kualitas kimia lainnya. Setiap baris dalam dataset mewakili satu entitas atau contoh, seperti sampel air yang diambil dari berbagai lokasi atau sumber. Sedangkan, setiap kolom merepresentasikan atribut atau parameter yang diamati pada setiap entitas.

Contoh dataset yang digunakan dalam penelitian ini diambil dari website *Kaggle.com* yang berjudul *waterQuality1*. Data ini telah dikumpulkan dari berbagai lokasi dan telah melalui proses validasi dan pengolahan sebelum digunakan dalam klasifikasi. Sedangkan klasifikasi sendiri merupakan langkah dimana data dikelompokkan ke dalam kelas-kelas yang telah terdefinisi sebelumnya, berdasarkan atribut atau fitur-fitur tertentu yang dimiliki oleh data tersebut [16]. Proses ini bertujuan untuk mengidentifikasi pola atau karakteristik yang membedakan setiap kelas dan memungkinkan untuk pengambilan keputusan yang lebih baik berdasarkan pada klasifikasi tersebut.

Klasifikasi adalah tahap fundamental dalam analisis data dan pembelajaran mesin, di mana model yang telah dihasilkan dari data latih digunakan untuk memprediksi kelas dari data uji yang belum pernah dilihat sebelumnya. Dalam konteks penelitian dan aplikasi praktis, klasifikasi dapat digunakan dalam berbagai bidang seperti pengenalan pola, diagnosis medis, analisis risiko keuangan, dan banyak lagi. Proses ini melibatkan penggunaan algoritma pembelajaran mesin, seperti *K-Nearest Neighbor* untuk memetakan data ke dalam kelas-kelas yang telah ditentukan sebelumnya, sehingga memberikan pemahaman yang lebih dalam tentang karakteristik dan pola yang ada dalam data tersebut [17].

Berikut adalah penjelasan singkat dari 20 unsur yang digunakan untuk mendukung metodologi penelitian KNN dalam menentukan kualitas air. Unsur-unsur tersebut dijadikan sebagai fitur dalam model KNN. Nilai fitur dari setiap sampel air kemudian dapat digunakan untuk memprediksi kualitas air tersebut.

Tabel 1. Tabel Sifat Unsur

Nama Unsur	Sifat
Aluminium	Logam ringan dan tahan karat
Ammonia	Gas yang tidak berwarna, berbau busuk, dan mudah terbakar
Arsenik	Logam yang berwarna putih keperakan, bersifat beracun, dan dapat menyebabkan berbagai penyakit

Barium	Logam yang berwarna putih keperakan, bersifat radioaktif, dan dapat menyebabkan kanker
Cadmium	Logam yang berwarna putih keperakan, bersifat beracun, dan dapat menyebabkan berbagai penyakit
Chloramine	Senyawa kimia yang mengandung klorin dan amonia, digunakan dalam pengolahan air untuk membunuh bakteri dan virus
Chromium	Logam yang berwarna putih keperakan, bersifat beracun, dan dapat menyebabkan kanker
Tembaga	Logam yang berwarna merah muda, bersifat penting bagi tubuh manusia
Fluor	Unsur kimia yang tidak berwarna, tidak berbau, dan tidak berasa, penting bagi kesehatan gigi dan tulang
Bakteri	Organisme hidup yang sangat kecil, dapat menguntungkan atau merugikan manusia
Virus	Organisme hidup yang sangat kecil, tidak dapat hidup sendiri, dan dapat menyebabkan berbagai penyakit
Timbal	Logam yang berwarna putih keperakan, bersifat beracun, dan dapat menyebabkan berbagai penyakit
Nitrat	Senyawa kimia yang mengandung nitrogen dan oksigen, digunakan dalam pupuk untuk menyuburkan tanaman
Nitrit	Senyawa kimia yang mengandung nitrogen dan oksigen, digunakan dalam pembuatan makanan dan obat-obatan
Merkuri	Logam yang berwarna perak putih, bersifat beracun, dan dapat menyebabkan berbagai penyakit
Perchlorate	Senyawa kimia yang mengandung klorin dan oksigen, digunakan dalam bahan peledak dan roket
Radium	Unsur radioaktif yang berwarna putih keperakan, bersifat beracun, dan dapat menyebabkan kanker
Selenium	Unsur kimia yang berwarna putih keperakan, bersifat penting bagi tubuh manusia
Perak	Logam yang berwarna putih keperakan, bersifat penting bagi tubuh manusia
Uranium	Unsur radioaktif yang berwarna putih keperakan, bersifat beracun, dan dapat menyebabkan kanker

Beberapa pertimbangan dalam memilih unsur-unsur tersebut sehingga digunakan sebagai fitur dalam model KNN karena relevansi unsur dengan kualitas air, ketersediaan data untuk unsur tersebut, keakuratan dan kompleksitas perhitungan unsur tersebut [18].

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengembangkan sebuah program yang dapat melakukan penentuan kualitas air dengan menggunakan Metode K-Nearest Neighbor (KNN). Program ini menggunakan dataset yang diperoleh dari Kaggle.com, dengan judul "waterQuality1". Dataset ini mengandung informasi mengenai 20 unsur yang digunakan sebagai fitur-fitur klasifikasi untuk menentukan apakah suatu sampel air layak konsumsi atau tidak. Setiap unsur memiliki nilai-nilai yang memiliki batasan yang menandakan apakah nilainya berada dalam kisaran yang aman untuk dikonsumsi. Dalam konteks ini, batasan-batasan tersebut dapat diartikan sebagai nilai-nilai ambang yang harus dipatuhi oleh setiap unsur agar air dianggap aman untuk dikonsumsi. Adapun batas dari masing-masing unsur yaitu:

Tabel 2. Tabel Batasan Aman Unsur

Nama Unsur	Batasan Aman Unsur
Aluminium	$\leq 2,8$
Ammonia	$\leq 32,5$
Arsenic	$\leq 0,01$
Barium	≤ 2
Cadmium	$\leq 0,0005$
Chloramine	≤ 4
Chromium	$\leq 0,1$
Copper	$\leq 1,3$
Flouride	$\leq 1,5$
Bacteria	≤ 0
Viruses	≤ 0

Lead	≤ 0,015
Nitrates	≤ 10
Nitrites	≤ 1
Mercury	≤ 0,002
Perchlorate	≤ 56
Radium	≤ 5
Selenium	≤ 0,5
Silver	≤ 0,1
Uranium	≤ 0,3

Berikut ini merupakan gambaran sebagian dari tampilan dataset yang diperoleh dari sumbernya, yaitu situs web Kaggle.com, yang berjudul "waterQuality1". Dataset ini mengandung sebanyak 8.000 sampel yang beragam untuk tujuan penulisan artikel ini, tidak semua sampel akan ditampilkan secara lengkap. Meskipun demikian, bagian dari dataset yang disajikan akan memberikan pandangan awal mengenai isi dan komposisi data yang digunakan dalam penelitian ini.

aluminium	ammonia	arsenic	barium	cadmium	chloramin	chromium	copper	fluoride	bacteria	viruses	lead	nitrate	nitrite	mercury	perchlorate	radium	selenium	silver	uranium	is_safe
1.65	9.08	0.04	2.85	0.007	0.35	0.83	0.17	0.05	0.2	0.054	16.08	1.13	0.007	37.75	6.78	0.08	0.34	0.02	1	
2.32	21.16	0.01	3.31	0.002	5.28	0.68	0.66	0.9	0.65	0.65	0.1	2.01	1.93	0.003	32.26	3.21	0.08	0.27	0.05	1
1.01	14.02	0.04	0.58	0.008	4.24	0.53	0.02	0.99	0.05	0.003	0.078	14.16	1.11	0.006	50.28	7.07	0.07	0.44	0.01	0
1.36	11.33	0.04	2.96	0.001	7.23	0.03	1.66	1.08	0.71	0.71	0.016	1.41	1.29	0.004	9.12	1.72	0.02	0.45	0.05	1
0.92	24.33	0.03	0.2	0.006	2.67	0.69	0.57	0.61	0.13	0.001	0.117	6.74	1.11	0.003	16.9	2.41	0.02	0.06	0.02	1
0.94	14.47	0.03	2.88	0.003	0.8	0.43	1.38	0.11	0.67	0.67	0.135	9.75	1.89	0.006	27.17	5.42	0.08	0.19	0.02	1
2.36	5.6	0.01	1.35	0.004	1.28	0.62	1.88	0.33	0.13	0.007	0.021	18.6	1.78	0.007	45.34	2.84	0.1	0.24	0.08	0
3.93	19.87	0.04	0.66	0.001	6.22	0.1	1.86	0.86	0.16	0.005	0.197	13.65	1.81	0.001	53.35	7.24	0.08	0.08	0.07	0
0.6	24.58	0.01	0.71	0.005	3.14	0.77	1.45	0.98	0.35	0.002	0.167	14.66	1.84	0.004	23.43	4.99	0.08	0.25	0.08	1
0.22	16.76	0.02	1.37	0.007	6.4	0.49	0.82	1.24	0.83	0.83	0.109	4.79	1.46	0.01	30.42	0.08	0.03	0.31	0.01	1
3.27	3.6	0.001	2.69	0.005	5.75	0.15	0.6	1.29	0.04	0.008	0.145	8.47	1.25	0.006	55.4	7.8	0.05	0.33	0.06	0
1.35	21.96	0.04	0.84	0.002	0.1	0.76	0.17	0.58	0.52	0.52	0.011	18.4	1.49	0.009	21.52	1.3	0.08	0.48	0.08	1
1.88	19.26	0.02	2.78	0.008	0.05	0.42		1.009	0.91	0.91	0.103	4.37	1.95	0.006	22.12	1.97	0.03	0.06	0.05	1
4.93	23.98	0.04	3.05	0.008	0.7	0.51	1.35	1.07	0.7	0.7	0.101	1.16	1.11	0.008	26.8	5.58	0.09	0.38	0.03	1
2.89	18.82	0.05	3.77	0.008	5.99	0.54	0.79	0.54	0.2	0.009	0.126	17.56	1.82	0.009	17.54	4.33	0.1	0.05	0.02	1
0.61	2.41	0.03	0.59	0.002	1.94	0.77	1.54	0.62	0.23	0.001	0.017	1.99	1.08	0.007	11.16	0.98	0.01	0.47	0.03	1
3.47	15.84	0.02	0.06	0.001	5.29	0.47	1.08	1.43	0.89	0.89	0.08	1.91	1.2	0.008	19.8	6.89	0.06	0.12	0.08	1
2.11	17.03	0.02	0.88	0.009	7.78	0.88	1.15	0.34	0.85	0.85	0.065	17.86	1.53	0.003	19.4	1.14	0.1	0.4	0.01	1
4.88	26.94	0.02	0.36	0.001	1.21	0.68	0.71	0.99	0.75	0.75	0.071	0.31	1.22	0.002	56.7		1	0.41	0.05	0
4.12	17.99	0.02	3.43	0.006	0.01	0.41	1.82	0.22	0.99	0.99	0.108	8.06	1.76	0.005	24.29	0.88	0.1	0.1	0.07	1
0.68	18.99	0.001	0.04	0.006	4.57	0.2	1.18		1.92	0.92	0.086	9.46	1.41	0.007	21.79	3.05	0.03	0.13	0.08	1
1.15	8.12	0.02	0.97	0.007	3.47	0.65	1.51	1.46	0.58	0.58	0.061	8.96	1.5	0.004	14.6	1.74	0.03	0.01	0.06	1
0.27	10.67	0.02	0.55	0.001	3.74	0.12	1.77	0.43	0.8	0.8	0.114	12.69	1.18	0.008	34.64	0.9	0.02	0.16	0.06	1
4.32	20.64	0.03	2.6	0.008	7.24	0.61	1.23	1.44	0.56	0.56	0.012	9.42	1.74	0.004	36.23	3.22	0.07	0.18	0.08	0
2.36	27.05	0.01	0.68	0.003	4.07	0.13	1.34	0.29	0.96	0.96	0.167	15.05	1.92	0.002	56.32	7.99	0.06	0.5	0.06	0
3.31	22.07	0.03	0.46	0.001	7.22	0.73	1.05		1.025	0.007	0.109	1.92	1.07	0.001	39.4	0.49	0.04	0.47	0.05	1
1.82	6.81	0.01	0.85	0.006	2.55	0.25	1.09	1.35	0.16	0.002	0.031	16.99	1.7	0.007	44.76	1.15	0.08	0.26	0.08	1
3.42	2.4	0.001	2.8	0.003	2.87	0.73	0.27	0.53	0.44	0.002	0.11	13.73	1.69	0.009	55.27	3.29	0.04	0.17		0
4.41	15.14	0.03	1.76	0.007	6.63	0.62	1.57	0.26	0.69	0.69	0.182	1.49	1.81	0.008	24.91	2.39	0.01	0.2	0.05	1
4.57	25.84	0.01	3.04	0.002	2.78	0.72	0.46	1.41	0.08	0.006	0.049	9.64	1.43	0.006	36.57	4.55	0.03	0.19		0
1.69	22.8	0.001	2.01	0.002	5.87	0.78	1.53	0.87	0.13	0.008	0.079	8.64	1.19	0.007	24.61	7.77	0.06	0.11	0.05	1
1.87	6.1	0.05	3.74	0.008	2.86	0.52	1.03	0.51	0.77	0.77	0.012	1.11	1.55	0.008	20.32	2.21	0.02	0.48	0.04	1
2.71	26.19	0.04	3.02	0.006		4.017	1.56	0.91	0.38	0.005	0.013		11	1.006	14.16	6.16	0.05	0.05	0.08	1
1.63	15.75	0.03	2.54	0.008	4.25	0.74	1.35	1.24	0.12	0.006	0.193	17.29	1.27	0.001	58.15	1.71	0.09	0.44	0.05	0
0.01	29.29	0.001	2.93	0.007	7.75	0.68	1.09		0.8	0.8	0.091	17.33	1.86	0.001	43.09	3.76	0.09	0.3		0
4.49	4.07	0.05	1.87	0.005	4.55	0.32	0.9	0.55	0.3	0.002	0.135	5.23	1.43	0.004	36.52	6.15	0.04	0.16	0.04	1
3.52	23.19	0.03	2.69	0.006	4.94	0.76	1.85	1.24	0.75	0.75	0.124	1.91		1.008	24.84	3.19	0.08	0.37	0.06	1
4.35	26.85	0.02	0.88	0.004	0.64	0.61	0.91	0.42	0.44	0.006	0.107	13.24	1.19	0.009	0.73	0.5	0.04	0.35		0
2.51	13.08	0.04	1.58	0.002	2.86	0.87	1.88	0.4	0.13	0.004	0.085	11.96	1.35	0.008	46.3	4.7	0.1	0.04	0.05	0
2.34	19.9	0.03	2.67	0.006	4.3	0.26	0.08	0.34	0.26	0.003	0.146	7.96	1.56	0.006	51.39	2.41	0.09	0.12	0.05	0
1.71	9.49	0.05	3.63	0.006	2.24	0.9	0.22	0.68	0.63	0.63	0.162	11.4	1.83	0.005	42.64	6.14	0.09	0.44	0.06	0
3.45	22.32	0.01	0.55	0.004	2.4	0.36	0.29	0.51	0.86	0.86	0.048	1.79	1.68	0.005	43.51	5.45	0.07	0.04	0.07	0
3.37	1.07	0.04	1.86	0.003	2.9	0.84	1.92	1.2	0.32	0.003	0.082	18.91	1.73	0.008	23.6	0.37	0.02	0.46	0.08	1
3.13	28.19	0.001	0.25	0.001	3.58	0.68	1.29	1.04	0.49	0.001	0.079	15.05	1.09	0.005	23.68	3.39	0.01	0.2	0.03	0
0.29	13.74	0.02	0.52	0.007	2.99	0.11	1.14	0.36	0.13	0.008	0.071	18.7	1.25	0.004	39.31	2.18	0.04	0.01	0.02	0
1.51	0.14	0.02	0.06	0.003	5.71	0.88	1.71	1.42	0.27	0.002	0.103	16.16	1.25	0.006	17.24	2.3	0.07			0
1.09	28.84	0.01	2.16	0.001	1.26	0.85	1.59	1.04	0.08	0.007	0.141	6.48	1.43	0.006	44.66	2.67	0.05	0.35	0.08	0
0.94	10.83	0.01	3.94	0.008	5.38	0.72	0.3	0.52	0.38	0.002	0.131	8.21	1.28	0.005	43.79	4.27	0.06	0.25	0.05	0
2.68	9.02	0.04	3.98	0.002	0.22	0.09	0.54	0.14	0.27	0.004	0.069	8.43	1.84	0.007	21.79	2.91	0.08	0.43	0.09	1

Gambar 2. Dataset

Dari data tersebut, tergolong klasifikasi manakah air dengan aluminium :2, ammonia : 20, arsenic : 0.02, barium : 3, cadmium : 0.005, chloramine : 4, chromium : 0.5, copper : 0.5, fluoride : 0.3, bacteria : 0.4, viruses : 0.3, lead : 0, nitrates : 2, nitrites : 1, mercury : 0.1, perchlorate : 35, radium : 5, selenium : 0.07, silver : 0.3, uranium. : 0.02.

$d(data1, data\ baru) =$

$$\sqrt{(1.65 - 2)^2 + (9.08 - 20)^2 + (0.04 - 0.02)^2 + (2.85 - 3)^2 + (0.007 - 0.005)^2 + (0.35 - 4)^2 + (0.83 - 0.5)^2 + (0.17 - 0.5)^2 + (0.05 - 0.3)^2 + (0.2 - 0.4)^2 + (0 - 0.3)^2 + (0.054 - 0)^2 + (16.08 - 2)^2 + (1.13 - 1)^2 + (0.007 - 0.1)^2 + (37.75 - 35)^2 + (6.78 - 5)^2 + (0.08 - 0.07)^2 + (0.34 - 0.3)^2 + (0.02 - 0.02)^2}$$

$$= 31,97646 (1)$$

$$d(\text{data1}, \text{data baru}) =$$

$$\sqrt{(2.32 - 2)^2 + (21.16 - 20)^2 + (0.01 - 0.02)^2 + (1.31 - 3)^2 + (0.002 - 0.005)^2 + (5.28 - 4)^2 + (0.68 - 0.5)^2 + (0.66 - 0.5)^2 + (0.9 - 0.3)^2 + (0.65 - 0.4)^2 + (0.65 - 0.3)^2 + (0.1 - 0)^2 + (2.01 - 2)^2 + (1.93 - 1)^2 + (0.003 - 0.1)^2 + (32.26 - 35)^2 + (3.21 - 5)^2 + (0.08 - 0.07)^2 + (0.27 - 0.3)^2 + (0.05 - 0.02)^2}$$

$$= 7,81583 (1)$$

$$d(\text{data1}, \text{data baru}) =$$

$$\sqrt{(1.01 - 2)^2 + (14.02 - 20)^2 + (0.04 - 0.02)^2 + (0.58 - 3)^2 + (0.008 - 0.005)^2 + (4.24 - 4)^2 + (0.53 - 0.5)^2 + (0.02 - 0.5)^2 + (0.99 - 0.3)^2 + (0.05 - 0.4)^2 + (0.003 - 0.3)^2 + (0.078 - 0)^2 + (14.16 - 2)^2 + (1.11 - 1)^2 + (0.006 - 0.1)^2 + (50.28 - 35)^2 + (7.07 - 5)^2 + (0.07 - 0.07)^2 + (0.44 - 0.3)^2 + (0.01 - 0.02)^2}$$

$$= 34,99128 (0)$$

$$d(\text{data1}, \text{data baru}) =$$

$$\sqrt{(2.32 - 2)^2 + (21.16 - 20)^2 + (0.01 - 0.02)^2 + (1.31 - 3)^2 + (0.002 - 0.005)^2 + (5.28 - 4)^2 + (0.68 - 0.5)^2 + (0.66 - 0.5)^2 + (0.9 - 0.3)^2 + (0.65 - 0.4)^2 + (0.65 - 0.3)^2 + (0.1 - 0)^2 + (2.01 - 2)^2 + (1.93 - 1)^2 + (0.003 - 0.1)^2 + (32.26 - 35)^2 + (3.21 - 5)^2 + (0.08 - 0.07)^2 + (0.27 - 0.3)^2 + (0.05 - 0.02)^2}$$

$$= 7,81583 (1)$$

Data tersebut dikelompokkan ke dalam kategori klasifikasi 1, yang mengindikasikan bahwa data tersebut dianggap aman atau memenuhi standar yang diperlukan untuk dapat dikonsumsi.

3.1 Implementasi

Berikut implementasi klasifikasi *K-Nearest Neighbors* (KNN) menggunakan MATLAB yang melibatkan beberapa langkah penting, seperti yang dijelaskan berikut:

```
>> training = [1.45 9.00 0.04 2.55 0.007 0.35 0.95 0.17 0.05 0.2 0 0.054 16.00 1.13 0.007 37.75 4.78 0.08 0.34 0.021 2.32 21.16 0.01 3.31]
>> training

training =

Columns 1 through 6
    1.4500    9.0000    0.0400    2.5500    0.0070    0.3500
    2.3200   21.1600    0.0100    3.3100    0.0020    5.2800
    1.0100   14.0200    0.0400    0.5800    0.0080    4.2400

Columns 7 through 12
    0.9900    0.1700    0.0500    0.2000         0    0.0540
    0.4800    0.6600    0.9000    0.6500    0.4550    0.1000
    0.5300    0.0200    0.9900    0.0500    0.0030    0.0780

Columns 13 through 19
   16.0800    1.1300    0.0070   37.7500    4.7800    0.0800
    2.0100    1.9300    0.0030   32.2600    3.2100    0.0800
   14.1600    1.1100    0.0060   50.2800    7.0700    0.0700

Columns 19 through 20
    0.3400    0.0200
    0.2700    0.0500
```

Gambar 3. Data Training

Langkah pertama adalah menyiapkan data *training* atau data latih yang akan digunakan untuk melatih model KNN. Data training terdiri dari sejumlah contoh yang sudah memiliki label atau kelas yang ditentukan. Pada penelitian klasifikasi kualitas air kali ini, data training berisi informasi tentang komposisi kimia air (seperti Aluminium, Ammonia, Arsenic, Barium, Cadmium, Chloramine, Chromium, Copper, Fluoride, Bakteri, Virus, Timbal, Nitrat, Nitrit, Merkuri, Perchlorate, Radium, Selenium, Perak, dan Uranium) serta label yang menunjukkan apakah air tersebut aman atau tidak. Kemudian berikut adalah gambar data sample yang digunakan pada penelitian ini yang ditampilkan dalam MATLAB.

```
>> sampel = [2 20 0.02 3 0.005 4 0.5 0.5 0.3 0.4 0.3 0 2 1 0.1 35 5 0.07 0.3 0.02]

sampel =

Columns 1 through 6
    2.0000    20.0000    0.0200    3.0000    0.0050    4.0000

Columns 7 through 12
    0.5000    0.5000    0.3000    0.4000    0.3000         0

Columns 13 through 18
    2.0000    1.0000    0.1000   35.0000    5.0000    0.0700

Columns 19 through 20
    0.3000    0.0200
```

Gambar 4. Data Sample

Data sample adalah data yang ingin kita klasifikasikan menggunakan model KNN setelah dilatih dengan data training. Ini adalah data yang belum memiliki label atau kelas, dan akan memprediksi labelnya berdasarkan tetangga terdekatnya dalam data training. Dalam penelitian kali ini, berupa komposisi kimia air yang akan ditentukan apakah aman untuk diminum atau tidak. Berikutnya adalah Group atau kelompok yang bisa dilihat pada gambar berikut.

```
>> group = [1; 1; 0]

group =

     1
     1
     0
```

Gambar 5. Group

Grup atau kelompok adalah label atau kelas yang sudah diketahui dari data training. Dalam implementasi KNN, akan dicari tetangga terdekat dari data *sample* dalam data *training*, dan kemudian memilih label mayoritas dari tetangga-tetangga ini sebagai prediksi kelas untuk data *sample*. Dalam contoh kualitas air, ini bisa menjadi kelas "aman" dengan tanda 1 atau "tidak aman" dengan tanda 0. Berikut adalah hasil dari klasifikasi yang dilakukan menggunakan MATLAB dengan mengimplementasikan metode KNN.

```
>> class = fitcknn(training, group, 'NumNeighbors', 1)

class =

ClassificationKNN
    ResponseName: 'Y'
    CategoricalPredictors: []
    ClassNames: [0 1]
    ScoreTransform: 'none'
    NumObservations: 3
    Distance: 'euclidean'
    NumNeighbors: 1

Properties, Methods

>> klasifikasi = predict(class, sampel)

klasifikasi =

     1
```

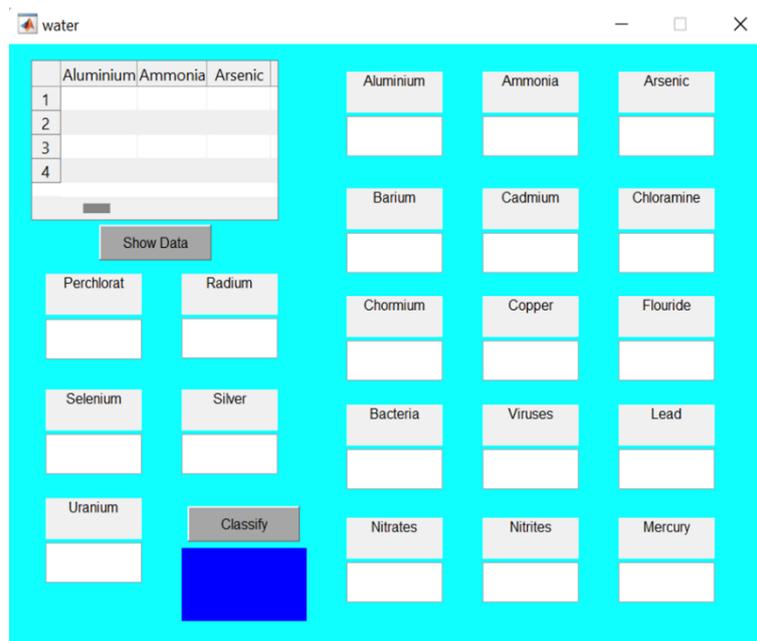
Gambar 6. Klasifikasi

Langkah terakhir adalah melakukan proses klasifikasi sebenarnya. Proses ini melibatkan beberapa Langkah. Pertama, menghitung jarak antara data sample dan setiap contoh dalam data training. Ini dilakukan menggunakan metrik jarak seperti *Euclidean distance* atau metrik lainnya. Kedua, memilih K tetangga terdekat dengan jarak terpendek dari data sample. Nilai K adalah parameter yang harus ditentukan sebelumnya. Misalnya, jika K = 3,

maka kita akan memilih tiga tetangga terdekat. Ketiga, menentukan label mayoritas dari tetangga-tetangga ini. Dalam contoh kualitas air, jika dua dari tiga tetangga terdekat diklasifikasikan sebagai "aman", maka data sample juga akan diklasifikasikan sebagai "aman". Keempat, mengembalikan hasil klasifikasi, yaitu label atau kelas yang diprediksi untuk data sample berdasarkan mayoritas label dari tetangga terdekat.

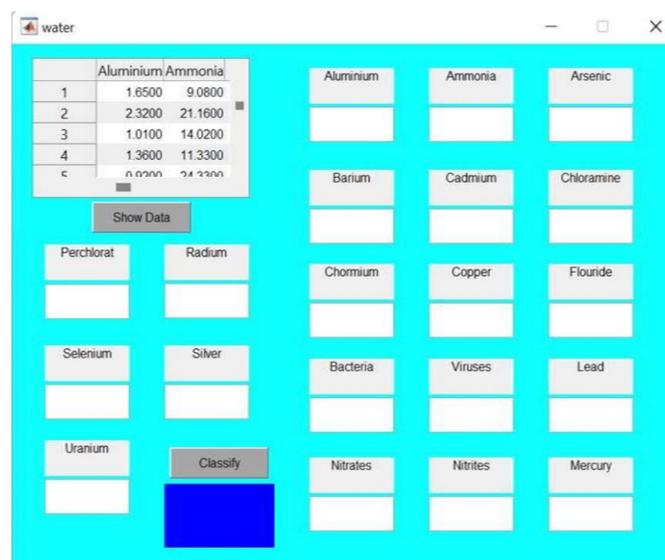
Dalam implementasi MATLAB, langkah-langkah di atas biasanya melibatkan penggunaan fungsi-fungsi dan algoritma yang telah disediakan oleh MATLAB, seperti perhitungan jarak, pemilihan tetangga terdekat, dan klasifikasi berdasarkan mayoritas. MATLAB memiliki pustaka yang memungkinkan pengguna untuk dengan mudah menerapkan algoritma KNN dengan memberikan data training, data sample, jumlah tetangga (K), dan fungsi jarak yang sesuai.

3.2 Tampilan GUI MATLAB



Gambar 7. Tampilan Awal GUI

Saat pertama kali program dijalankan maka semua table dan unsur akan kosong seperti pada Gambar 2.2.2 Tampilan Awal GUI. Untuk menampilkan data dari dataset dapat menggunakan tombol *Show Data*. Dengan begitu isi dari dataset waterQuality1.xlsx akan muncul pada tampilan GUI program. Pengguna dapat mengulirkan layar untuk melihat keseluruhan isi tabel.



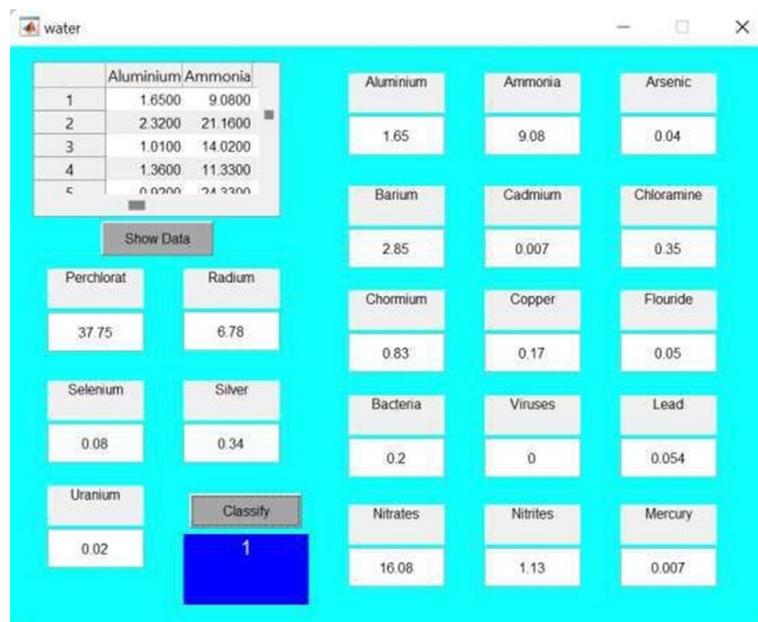
Gambar 8. Tampilan Show Data pada GUI

Setelah dataset muncul dari tabel data waterQuality1 user dapat menginputkan nilai unsur yang akan diinputkan sesuai dengan unsur yang telah tersedia.

Adapun percobaan yang dilakukan pada program Penentuan Kualitas Air ini yaitu percobaan dengan menginput nilai dari tiap-tiap unsur yang akan menghasilkan nilai 1 dan percobaan menginputkan nilai dari tiap-tiap unsur yang akan menghasilkan nilai 0. Nilai 1 pada program Penentuan Kualitas Air ini adalah nilai yang menjelaskan bahwa kandungan unsur dalam air tersebut aman untuk diminum dan nilai 0 adalah nilai yang menjelaskan bahwa kandungan unsur dalam air tersebut tidak aman untuk diminum. Untuk mengetahui kualitas layak minum dari suatu air dapat diketahui dengan cara menginputkan nilai dari masing-masing unsur yang terkandung dalam air tersebut. Adapun percobaan yang dilakukan yaitu:

1. Percobaan Pertama

Tahap awal dari percobaan ini melibatkan proses penginputan data yang sesuai dengan tampilan yang tergambar pada Gambar 9. Langkah ini diambil untuk memulai eksperimen dan mengumpulkan informasi yang dibutuhkan. Dengan mengikuti panduan visual yang disajikan dalam Gambar 9, kita dapat menyusun dataset yang diperlukan untuk analisis lebih lanjut.

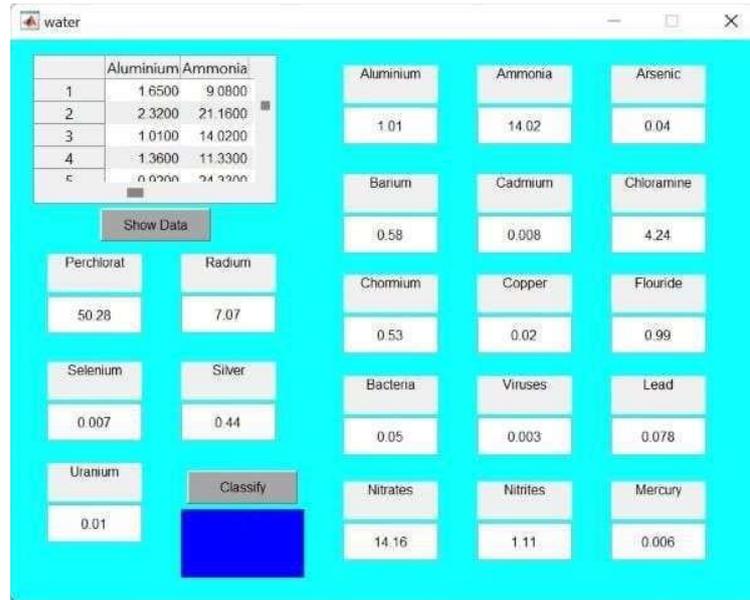


Gambar 9. Tampilan Hasil Percobaan Pertama

Setelah melakukan klik pada tombol "Classify", hasil nilai dari percobaan awal terungkap sebagai 1. Melalui hasil ini, dapatlah diambil kesimpulan bahwa berdasarkan nilai dari setiap unsur yang telah dimasukkan dalam percobaan pertama, kualitas air dinilai aman dan layak untuk dikonsumsi. Informasi ini memberikan keyakinan bahwa komposisi unsur pada sampel air percobaan pertama memenuhi standar keselamatan yang ditetapkan.

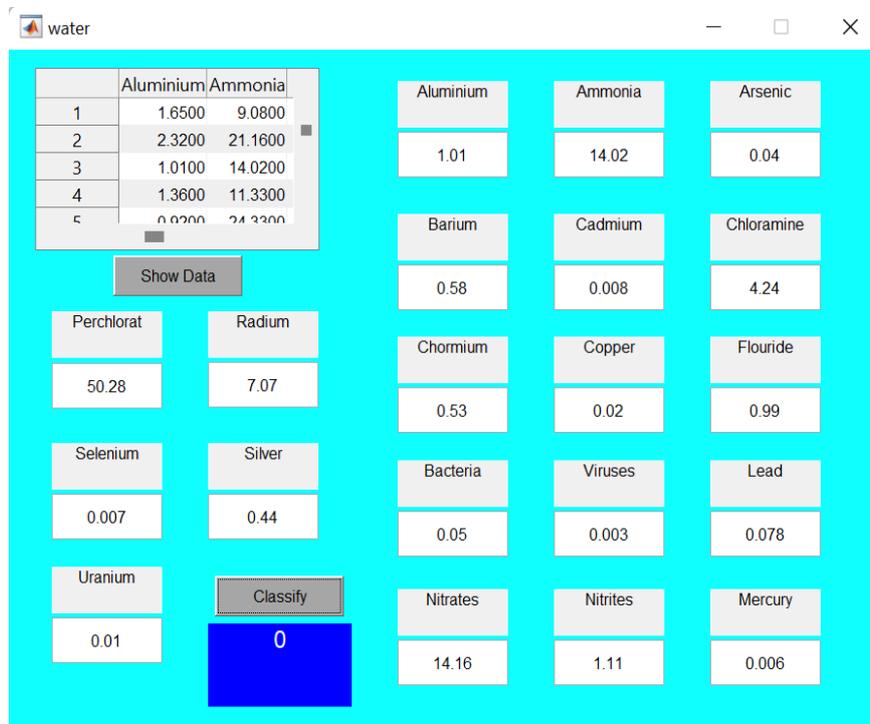
2. Percobaan Kedua

Dalam tahap percobaan kedua, pendekatan yang diambil melibatkan pengisian nilai-nilai unsur sesuai dengan yang ditampilkan dalam Gambar 10. Proses ini dilakukan dengan tujuan untuk mengevaluasi respons sistem terhadap komposisi kimia yang berbeda. Dengan mengikuti pola yang terlihat dalam Gambar 10, percobaan ini memberikan wawasan tentang bagaimana perubahan dalam nilai unsur tersebut dapat memengaruhi penentuan kualitas air seperti pada Gambar berikut



Gambar 10. Tampilan Hasil Percobaan Kedua

Setelah seluruh nilai unsur berhasil diinputkan, langkah selanjutnya adalah melakukan penilaian terhadap apakah komposisi nilai unsur dalam sampel air pada percobaan kedua ini masih memenuhi standar keamanan seperti yang ditemukan pada percobaan pertama. Untuk melakukannya, kita dapat mengikuti prosedur yang sama seperti pada percobaan sebelumnya, yaitu dengan mengklik tombol "Classify". Dengan mengaplikasikan metode ini, kita dapat menilai apakah nilai-nilai unsur dalam sampel air percobaan kedua ini mengindikasikan keamanan atau keberisikan air, berdasarkan hasil klasifikasi yang dihasilkan dari algoritma *K-Nearest Neighbor*.



Gambar 11. Tampilan Hasil Nilai Percobaan Kedua

Setelah tombol "Classify" ditekan, terlihat bahwa hasil nilai dari percobaan kedua berbeda dengan hasil dari percobaan pertama, yang pada percobaan pertama bernilai 1. Dari perbandingan ini, dapat disimpulkan bahwa komposisi nilai unsur dalam sampel air pada percobaan kedua ini menghasilkan penilaian yang berbeda dan

menunjukkan bahwa kualitas air, berdasarkan nilai tiap unsur yang diinputkan pada percobaan kedua, dianggap tidak aman untuk dikonsumsi.

KESIMPULAN

Dari pelaksanaan penelitian ini, setelah dilakukan percobaan dengan memasukkan input nilai pada sistem GUI MATLAB, hasilnya sesuai dengan perhitungan manual. Maka dapat disimpulkan bahwa pengklasifikasian data kualitas air dari 20 unsur yang terdapat dalam 8000 dataset menggunakan metode KNN yang diimplementasikan pada GUI MATLAB ini memiliki potensi besar dalam membantu masyarakat luas dan pemerintah dalam menilai kualitas air dengan mempertimbangkan 20 unsur yang terkandung di dalamnya. Metode KNN mampu menjadi alat yang efektif untuk mengklasifikasikan data kualitas air dan memberikan informasi yang berharga mengenai tingkat keamanannya. Dengan metode ini, potensi kontaminasi atau masalah lain yang berkaitan dengan kualitas air dapat diidentifikasi lebih awal, memungkinkan upaya penanganan yang lebih cepat dan efisien.

Untuk penyempurnaan dan pengembangan proyek Penentuan Kualitas Air menggunakan metode KNN, beberapa saran dapat dipertimbangkan guna meningkatkan kualitas dan utilitas sistem ini.

- a. **Optimalisasi Performa Metode KNN:**
Perlu dilakukan upaya untuk mengoptimalkan performa metode KNN dalam pengklasifikasian data kualitas air. Peningkatan performa ini dapat dicapai dengan mempertimbangkan pemilihan nilai K yang optimal, memperbaiki skema validasi silang, atau mengimplementasikan teknik lain yang dapat meningkatkan akurasi dan efisiensi model.
- b. **Pengembangan Fitur dan Menu yang Lebih Kompleks:**
Sistem Penentuan Kualitas Air menggunakan metode KNN diharapkan dapat mengembangkan fitur-fitur dan menu-menu yang lebih lengkap dan kompleks. Fitur-fitur ini dapat mencakup kemampuan untuk melakukan analisis lebih mendalam terhadap parameter kualitas air tertentu, serta menawarkan pilihan visualisasi data yang lebih variatif dan informatif. Dengan menyediakan fitur-fitur yang lebih kaya, sistem akan lebih mudah digunakan oleh masyarakat luas dan membantu mereka dalam memahami hasil penilaian kualitas air.
- c. **Penggunaan Data Terbaru dan Representatif:**
Untuk menjaga relevansi dan keakuratan analisis, penting untuk menggunakan data terbaru dan representatif dalam proyek ini. Pembaruan data secara berkala akan memastikan bahwa sistem memiliki informasi yang tepat dan akurat dalam menilai kualitas air di berbagai lokasi.
- d. **Kolaborasi dengan Instansi Terkait:**
Sistem Penentuan Kualitas Air menggunakan metode KNN dapat memperoleh manfaat lebih lanjut dengan melakukan kolaborasi dengan instansi terkait, seperti lembaga lingkungan, universitas, atau badan pemerintah. Kolaborasi ini dapat meningkatkan akses ke data berkualitas tinggi, memperkuat analisis, dan memberikan informasi yang lebih komprehensif kepada masyarakat dan pemerintah.

Dengan mengimplementasikan saran-saran ini, sistem Penentuan Kualitas Air menggunakan metode KNN akan semakin berkembang dan menjadi alat yang lebih kuat dalam mengatasi isu kualitas air. Diharapkan, proyek ini dapat berkontribusi secara signifikan dalam upaya menjaga dan meningkatkan kualitas air, serta memberikan manfaat nyata bagi masyarakat dan lingkungan secara keseluruhan.

UCAPAN TERIMAKASIH

Terima kasih yang sebesar-besarnya kami sampaikan kepada seluruh pihak yang telah memberikan dukungan penuh dalam pelaksanaan penelitian ini. Khususnya, kami ingin mengucapkan terima kasih kepada Civitas Akademika Universitas Pembangunan Nasional "Veteran" Yogyakarta, terutama kepada jurusan Informatika di Fakultas Teknik Industri. Dukungan berharga dari tim akademis, dosen, staf, dan mahasiswa merupakan fondasi yang memungkinkan penelitian ini berjalan dengan lancar dan sukses. Kerjasama dan kontribusi dari pihak universitas telah memberikan dorongan yang tak ternilai dalam menghasilkan temuan yang berarti dalam bidang penelitian ini.

REFERENCES

- [1] M. A. Rahman, N. Hidayat dan A. A. Supianto, "Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, pp. 6346-6353, Desember 2018.

- [2] F. Y. Rahman, I. I. Purnomo dan N. Hijriana, "PENERAPAN ALGORITMA DATA MINING UNTUK KLASIFIKASI KUALITAS AIR," *Technologia*, vol. 3, pp. 28-232, 2022.
- [3] Aldi, Saputra dan Rahmat, "Klasifikasi Kualitas Air Minum Menggunakan Metode Machine Learning," *Jurnal Ilmiah Komputer dan Teknologi Informasi*, vol. 2, pp. 1-8, 2023.
- [4] Prismahardi dan Saputra, "Analisis Kualitas Air Minum dengan Menggunakan Teknik Lucifer Machine Learning," *Jurnal Ilmiah Komputer dan Teknologi Informasi*, vol. 2, pp. 1-9, 2023.
- [5] Ilyas, Irsyad dan Syamsul, "Klasifikasi Kualitas Air Minum dengan Menggunakan Algoritma Random Forest," *Naratif*, vol. 2, 2023.
- [6] Armono, Saptomo dan Harsadi, "Implementasi Algoritma K-Nearest Neighbor Untuk Identifikasi Kualitas Air (Studi Kasus: PDAM Kota Surakarta)," *Jurnal TIKomSIN*, vol. 6, 2018.
- [7] B. Prihambodo, A. W. F. Y, E. Prayoga dan A. Jaffar, "Klasifikasi Kualitas Air Sungai Berbasis Teknik Data Mining Dengan Metode K-Nearest Neighbor (K-NN)," *MITOR: Jurnal Teknik Elektro*, vol. 23, pp. 31-36, Maret 2023.
- [8] P. Putra, A. M. H. Pardede dan S. Syahputra, "ANALISIS METODE K-NEAREST NEIGHBOUR (KNN) DALAM KLASIFIKASI DATA IRIS BUNGA," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 6, pp. 297-286, 2022.
- [9] L. Farokhah, "IMPLEMENTASI K-NEAREST NEIGHBOR UNTUK KLASIFIKASI BUNGA DENGAN EKSTRAKSI FITUR WARNA RGB," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 7, pp. 1129-1136, 2020.
- [10] N. Krisandi, Helmi dan B. Prihandono, "ALGORITMA k-NEAREST NEIGHBOR DALAM KLASIFIKASI DATA HASIL PRODUKSI SAWIT PADA PT.MINAMAS KECAMATAN PARINDU," *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, vol. 02, pp. 33-39, 2013.
- [11] T. A. Munandar dan A. Q. Munir, "Implementasi K-Nearest Neighbor Untuk Prototype Sistem Pakar Identifikasi Penyakit Jantung," *Jurnal Teknologi Informasi*, vol. XVII, pp. 44-50, Juni 2022.
- [12] A. Tangkelayuk dan E. Mailoa, "Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes Dan Decision Tree," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, pp. 109-1119, 2022.
- [13] H. Said, N. Matondang dan H. N. Irmada, "Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi," *Techno.COM*, vol. 2, pp. 256-267, Mei 2022.
- [14] C. Muhammad, R. Maulana dan M. H. H. Ichsan, "Purwarupa Perahu untuk Monitoring dan Klasifikasi Kualitas Air Bendungan dengan Metode K-Nearest Neighbor (KNN)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, pp. 651-659, Februari 2020.
- [15] S. M. Prisca, B. Hidayat dan S. Darana, "KLASIFIKASI UNTUK DETEKSI KUALITAS KEJU CHEDDAR MENGGUNAKAN PENGOLAHAN CITRA DIGITAL DENGAN METODE CONTENT BASED IMAGE RETRIEVAL DAN K-NEAREST NEIGHBOR BERBASIS ANDROID," *e-Proceeding of Engineering*, vol. 5, pp. 2138-2146, Agustus 2018.
- [16] V. R. Atfira, N. Ibrahim dan E. Wulandari, "DETEKSI KUALITAS KEMURNIAN SUSU SAPI MELALUI PENGOLAHAN CITRA DIGITAL MENGGUNAKAN METODE ACTIVE CONTOUR DENGAN KLASIFIKASI K-NEAREST NEIGHBOR," *e-Proceeding of Engineering*, vol. 6, pp. 3830-3838, Agustus 2019.
- [17] F. Liantoni, "Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor," *ULTIMATICS*, vol. VII, pp. 98-105, 2015.
- [18] z. Wang, J. Wang, J. Liu dan B. Li, "Application of K-nearest neighbors algorithm for water quality assessment," *Science of The Total Environment*, vol. 671, pp. 203-211, 2019.