PROPOSAL MATA KULIAH PEMBELAJARAN MESIN DENGAN JUDUL Klasifikasi Kualitas Air Minum dengan Metode Machine Learning: Analisis Performa Algoritma KNN dan Logistic Regression

Nama : Muhammad Naffa Rizky Fissabilli

NIM : A11.2022.14755

Kelompok : A11.4412

**Abstrak**

Air adalah sebuah kebutuhan untuk menunjang hidup manusia. Air yangbersih juga berkualitas merupakan sebuah prioritas agar manfaatnya pun berguna bagi pengonsumsi. pH air merupakan faktor penentu bahwa kualitas air tersebut baik untuk dikonsumsi ataupun tidak. Dengan adanya penelitian ini membantu agar pembaca atau para konsumen air mineral kemasan, dapat menentukan mana merk air mineral yang berkualitas dan bermanfaat bagi tubuh serta tidak menimbulkan efek samping di masa depan. Studi ini bertujuan untuk mengklasifikasikan kualitas air mineral berdasarkan pada pengukuran atau perbandingan variabel data dengan menggunakan pendekatan machine learning. Air yang dikonsumsi oleh masyarakat harus memenuhi standar mutu yang ditetapkan agar aman dan bermanfaat. Dalam penelitian ini, saya melakukan klasifikasi untuk menentukan kualitas air l berdasarkan data yang tersedia serta keakuratan dari masing-masing metode machine learning. Beberapa metode machine learning seperti K-Nearest Neighbors (KNN) dan Logistic Regression diterapkan untuk melakukan klasifikasi.

**Kata Kunci**: Klasifikasi, Kualitas Air mineral, Machine Learning, pH, Akurasi.

**PENDAHULUAN**

1. **Latar Belakang**

Tujuan dari penelitian ini adalah untuk mengembangkan model klasifikasi kualitas air mineral kemasan yang lebih akurat dan efektif menggunakan K-Nearest Neighbors (KNN), dan Logistic Regression diterapkan untuk melakukan klasifikasi serta perbandingan antar model. Kualitas air bersih memiliki persyaratan agar layak untuk diminum dan harus sesuai dengan Standar Nasional Indonesia. Parameter yang harus dijadikan acuan adalah pH(tingkat keasaman) serta kekeruhan air. PH air minum yang layak untuk dikonsumsi harus berada pada rentang 6,0 hingga 8,5. pH air dibawah 7 termasuk dalam kategori air asam, sedangkan diatas 7 termasuk dalam kategori air basa. Selain daripada itu sudah termasuk air yang tidak layak untuk dikonsumsi, jika dikonsumsi akan menimbulkan masalah kesehatan, gangguan pencernaan, serta sakit kepala. Dengan masalah pemilihan air mineral yang baik untuk dikonsumsi juga identifikasi air yang bersih. Pengukuran kualitas air bersih menggunakan parameter dan variabel, dengan pengambilan dataset pada Kaggle. Pada penelitian ini penulis melakukan klasifikasi kualitas air bersih serta perbandingan model dengan menggunakan penerapan Machine Learning, yang mencakup pemodelan dengan K-Nearest Neighbor(KNN), dan Logistic Regression. Dalam penelitian ini menggunakan pendekatan supervised learning menggunakan library phyton dalam mengelola data. Metode ini digunakan untuk menganalisis data dalam melakukan klasifikasi serta membandingkan kedua model.

**II. Tujuan**

1. Membuat model klasifikasi yang dapat memprediksi kualitas air dengan tingkat akurasi yang tinggi.

2. Membandingkan performa algoritma KNN dan Logistic Regression dalam memprediksi kualitas air.

1. **Metode**

1. Pengumpulan data: Data "water potability" dari Kaggle.com. Penelitian ini menggunakan dataset dari Kaggle, dengan menggunakan sekitar 10 parameter yang nantinya akan dimodelkan dalam metode machine learning yang telah dipilih. Tahapan penelitian meliputi pengumpulan data, preprocessing citra, ekstraksi fitur, pembagian data latih dan data uji, serta implementasi metode machine learning.

Penelitian yang dilakukan menggunakan data yang diambil dari Kaggle yaitu water\_potability berformat csv. Data tersebut memiliki sepuluh parameter, yaitu.

a. pH (Power of Hydrogen) adalah skala yang digunakan untuk mengukur tingkat keasaman atau kebasaan suatu larutan. pH adalah ukuran konsentrasi ion hidrogen (H+) dalam suatu larutan. pH skala berjalan dari 0 hingga 14, dengan 7 sebagai titik netral. pH yang lebih rendah dari 7 menunjukkan keasaman, sedangkan pH yang lebih tinggi dari 7 menunjukkan kebasaan.

b. Hardness adalah parameter yang digunakan untuk mengukur kandungan mineral keras dalam air. Kerasan dapat berupa kalsium, magnesium, atau lain-lain. Kerasan dapat mempengaruhi kualitas air dan dapat berbahaya jika terlalu tinggi[4].

c. Solids adalah parameter yang digunakan untuk mengukur kandungan material padat dalam air. Padatan dapat berupa partikel, debu, atau lain-lain. Padatan dapat mempengaruhi kualitas air dan dapat berbahaya jika terlalu tinggi[5].

d. Chloramines adalah senyawa yang terbentuk dari klor dan amonia. Kloramin dapat berbahaya jika terlalu tinggi dalam air.

e. Sulfate adalah ion yang terbentuk dari sulfur dan oksigen. Sulfat dapat berbahaya jika terlalu tinggi dalam air.

f. Conductivity adalah parameter yang digunakan untuk mengukur kemampuan air untuk mengalirkan listrik. Konduktivitas dapat dipengaruhi oleh kandungan ion dalam air.

g. Organic Carbon adalah parameter yang digunakan untuk mengukur kandungan karbon dalam bentuk organik dalam air. Karbon organik dapat berbahaya jika terlalu tinggi dalam air.

h. Trihalomethanes adalah senyawa yang terbentuk dari klorinasi air. Trihalomethane dapat berbahaya jika terlalu tinggi dalam air.

i. Turbidity adalah parameter yang digunakan untuk mengukur kejernihan air. Turbiditas dapat dipengaruhi oleh kandungan partikel dalam air.

j. Potability adalah parameter yang digunakan untuk mengukur kesesuaian air untuk dikonsumsi. Potabilitas dapat dipengaruhi oleh kandungan ion, partikel, dan lain-lain dalam air.

2. Pre-processing Data

Normalisasi dan scaling data untuk meningkatkan kualitas data. Melakukan proses data scaling dengan mentransformasi data yang bertujuan untuk memperbaiki isi data dengan mengatur rentang nilai data agar sesuai dengan persyaratan yang ditentukan. Data scaling dilakukan ketika proses pengukuran data tidak sesuai dengan persyaratan penggunaan, seperti rentang waktu yang berbeda atau skala yang tidak sesuai. Penelitian ini menggunakan metode supervised learning. Pada tahapan awal pengolahan data ini adalah melakukan Exploratory Data Analysis (EDA) pada tahapan ini berfungsi untuk menganalisis data, pada penelitian ini pada EDA menganalisis tipe data, missing value, serta menganalis banyaknya data yang layak dikonsumsi dan tidak untuk dikonsumsi.

3. Pembuatan model: Membuat model KNN dan SVM untuk memprediksi kualitas air minum.

4. Evaluasi model: Menghitung tingkat akurasi, precision, recall, dan F1-score untuk mengetahui performa model.

**Model yang digunakan :**

K-Nearest Neighbors

K-Nearest Neighbor (KNN) adalah salah satu algoritma machine learning yang sederhana dan efektif dalam menangani masalah klasifikasi dan regresi. Algoritma ini termasuk dalam domain supervised learning dan memiliki aplikasi yang luas dalam pengenalan pola, data mining, dan deteksi intrusi. KNN bekerja dengan prinsip bahwa data yang mirip cenderung berada dekat satu sama lain, sehingga algoritma ini membandingkan satu titik data dengan sekelompok data yang telah dilatih untuk membuat prediksi.

Logistic Regression

Logistic regression adalah suatu model statistik yang memodelkan log-odds dari suatu event sebagai kombinasi linear dari satu atau lebih variabel independen. Tujuan utama logistic regression adalah untuk memprediksi probabilitas suatu event terjadi atau tidak berdasarkan data sebelumnya. Model ini sangat efektif dalam menyelesaikan tantangan kelasifikasi biner dan digunakan secara luas dalam berbagai industri seperti pemasaran, keuangan, dan medis.

1. **Hasil**

Hasil yang didapat adalah mana metode machine learning yang akurat dan presisi juga mengetahui kualitas air minum yang baik juga perbandingan antar model dengan f1-score terbaik.

Sitasi Daftar Pustaka:

1. Lidya Savitri, Rahmat Nursalim. Klasifikasi Kualitas Air Minum menggunakan Penerapan Algoritma Machine Learning dengan Pendekatan Supervised Learning. Department Of Mathematics : Universitas Bengkulu. 2023. https://ejournal.unib.ac.id/diophantine/article/download/28260/12639

2. Faza Gustaf Marrera, Barlian Henryranu Prasetio, Hurriyatul Fitriyah. Sistem Klasifikasi Air Mineral Layak Minum berdasarkan Nilai PH dan Kekeruhan Menggunakan Metode Naïve Bayes berbasis Arduino Uno. Universitas Brawijaya. Vol. 7, 2023.

https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/12139/5532/86875

3. https://www.kaggle.com/code/nimapourmoradi/water-potability