Plagiarism Scan Report

18% Plagiarism	**6%** Exact Match
	12% Partial Match

82% Unique

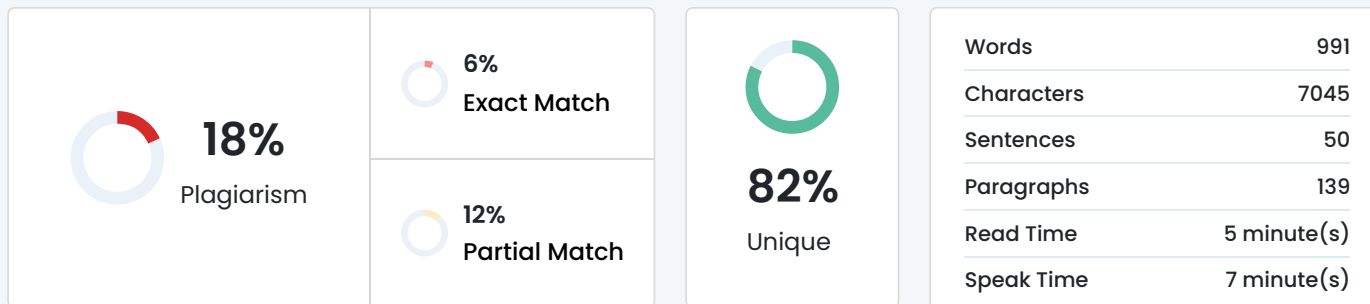|	
Words	991
Characters	7045
Sentences	50
Paragraphs	139
Read Time	5 minute(s)
Speak Time	7 minute(s)

Content Checked For Plagiarism

Water Quality Analysis using Machine
Learning

K Rajesh† , K Tharun Kumar Reddy† , Karthik Kemidi†
Chaitanya Bharathi Institute of Technology†
Department of Artificial Intelligence and Machine Learning†
Email: kurvarajesh72@gmail.com∗, tharunkumar5297@gmail.com†, kemidikarthik2004@gmail.com‡

Abstract

Water quality prediction was conducted to ascertain whether water is safe for consumption. This study also aimed to evaluate the performance of various machine learning models, including Decision Tree, Random Forest, XGBoost, KNN, SVM, and Gaussian Naive Bayes, in order to identify the most effective method for predicting water quality. Water is an essential resource for life, vital for the survival of all organisms, including humans. The sustainability of both business and agriculture relies heavily on the availability of freshwater. A critical aspect of managing freshwater resources is the assessment of water quality. Prior to utilizing water for any purpose—be it drinking, applying chemicals such as pesticides, or providing hydration for animals—it is imperative to evaluate its purity. Water quality has a direct influence on both the ecosystem and public health. Consequently, the analysis and prediction of water quality are essential for safeguarding environmental and human health. Machine learning techniques can be employed to analyze and predict water quality based on various parameters, including pH value, turbidity, hardness, conductivity, and dissolved solids. In this study, water quality is predicted by inputting the concentrations of these parameters into machine learning algorithms, classifying the water as either safe or unsafe for domestic use.

Keywords—Decision Tree, Random Forest, KNN, SVM, XGBoost, Gaussian Naive Bayes, Performance Metrics.

Introduction

The domain of machine learning examines the ways

in which computers acquire knowledge through experience. Given that the ability to learn is a core characteristic of what is considered intelligent, the terms "Machine Learning" and "Artificial Intelligence" are often used interchangeably by researchers. The primary objective of machine learning is to create systems that can adapt based on their experiences. Recent advancements in machine learning techniques have made it feasible to address this challenge. We have devised a method that employs data mining to determine the potability of water. The vast array of data concerning water quality can be analyzed to uncover valuable insights. Consequently, this area of study has gained increased importance. This research seeks to establish a system capable of predicting water quality with greater accuracy.

Literature Survey

[1] One comprehensive model analyzed water quality parameters using machine learning algorithms to predict conditions with high accuracy and reliability. This study highlighted the importance of data preprocessing and feature selection in improving prediction outcomes.[2] Another study compared multiple machine learning techniques to identify the most effective approach for water quality prediction. It was found that methods such as random forests and support vector machines demonstrated superior predictive performance due to their ability to handle non-linear relationships among water quality parameters.[3] Real-time monitoring systems integrated with machine learning algorithms have been developed to provide continuous water quality information. These systems enable prompt responses to potential health risks by detecting contaminants and other changes in water composition in real time. [4] Hybrid models, which combine different machine learning approaches, have shown improved performance in predicting specific water quality indicators such as salinity. These hybrid methods are particularly effective in capturing complex interactions between variables that traditional models may miss. [5] Some studies focus on analyzing and predicting water quality using datasets from field sensors and public records. By applying exploratory data analysis techniques, researchers have identified key patterns in the data, enabling them to predict long-Copyright Notice term trends and seasonal variations in water quality.[6] Recently, there has been a rise in research leveraging machine learning to assess and predict water potability, training models to evaluate water's suitability for drinking based on a variety of quality metrics. [7] Studies have evaluated machine learning algorithms for sustainable monitoring of drinking water quality, emphasizing these models' potential to make water management more efficient. This research often dis-

cusses the implications for sustainability and public health.[8] Comparative analyses of various machine learning algorithms on datasets with statistically imputed missing values have deepened the understanding of each model's strengths and limitations, highlighting how data preprocessing affects model accuracy. [9] The integration of machine learning with comprehensive weighting techniques has been explored to enhance prediction reliability by combining the outputs of multiple models. This method reduces errors from individual models, yielding more robust predictions.[10] Several studies have aimed to predict water potability by classifying water samples based on various quality indicators. This research is particularly valuable for public health monitoring.[11] Recent research has focused on optimizing machine learning models for water quality prediction by refining data preprocessing and tuning model parameters. These tailored models improve prediction accuracy significantly, especially when adapted to local environmental conditions and specific water quality attributes.

Problem Statement

Develop a machine learning model to check the water quality of a sample and to classify water samples as potable or non-potable based on physicochemical properties, including pH, hardness, and conductivity etc. to ensure safe drinking water availability. The Water Quality Analysis project aims to develop a machine learning model to classify water samples as either potable (safe for drinking) or non-potable (unsafe) based on various physicochemical attributes. The dataset includes features such as pH, hardness, total dissolved solids, chloramines, sulphate, conductivity, and nitrates.

Methodologies

The machine learning model is employed to ascertain whether water is safe for consumption or not. It is essential to import the necessary libraries to facilitate the testing and training of our dataset, along with installing specific packages related to nature-inspired algorithms. The dataset should be divided into training and testing subsets in an 80:20 ratio, followed by the Figure 1

selection of the appropriate model. The classifiers under consideration include the Support Vector Classifier (SVC), Decision Tree, GaussianNB, Random Forest, and XGBoost.

The dataset consists of observations of water quality for 3276 different sources of water:

Matched Source

Similarity 25%

Title:Depression detection for twitter users using sentiment analysis in ...

Jan 1, 2024 · In this section, we discuss the experimental results obtained from our research, which aimed to evaluate the performance of various machine learning classifiers on different datasets and feature representations.

https://www.sciencedirect.com/science/article/abs/pii/S0933365723002300

Similarity 15%

Title:Sustainability: the Triple Bottom Line – Profit, People, and ...

The sustainability approach seeks to consider all production and business decisions in light of how they affect economic, social, and environmental outcomes.

https://njaes.rutgers.edu/nursery/sustainable.html

Similarity 6%

Title:

medium.com · @a · machine-learning-vs-deepMachine Learning vs. Deep Learning: What's the Difference?

The primary objective of machine learning is to create predictive models capable of making informed decisions based on historical patterns and trends. Key Features of Machine Learning Structured ...

https://medium.com/@a.sale/machine-learning-vs-deep-learning-whats-the-difference-1bc289b04199

Similarity 5%

Title:"Machine-Learning-Based Approach to Decoding ...

by E Lashgari · 2021 — The aim of this dissertation is to develop computational tools, and especially machine learning techniques, to identify the most effective ...

https://digitalcommons.chapman.edu/cads_dissertations/27

Similarity 4%

Title:Do you believe health departments should be able to enter your...

I believe that health departments should be able to enter homes without a warrant in urgent situations to swiftly address potential health risks, ...

https://www.coursehero.com/tutors-problems/Law/57948138-Do-you-believe-health-departments-should-be-able-to-enter-your

---

Similarity 3%

Title:General Chemistry of Water - pH

7 days ago — We'll take a look at the following water properties: hardness, total dissolved solids (TDS), pH, and alkalinity. Read on to find out why ...

https://mytapscore.com/blogs/tips-for-taps/your-water-s-general-chemistry?srsltid=AfmBOorfwGpbXqTl5jd4psMYNSOKqRQ8WX9lhGmYjcHbiWanmDhvnSPC

Similarity 3%

Title:What Is a Machine Learning Model?

Aug 16, 2021 — A machine learning model is an expression of an algorithm that combs through mountains of data to find patterns or make predictions.

https://blogs.nvidia.com/blog/what-is-a-machine-learning-model

---

Similarity 3%

Title:Why is it wrong to train and test a model on the same ...

Dec 13, 2020 — The problem of training and testing on the same dataset is that you won't realize that your model is overfitting, because the performance of ...

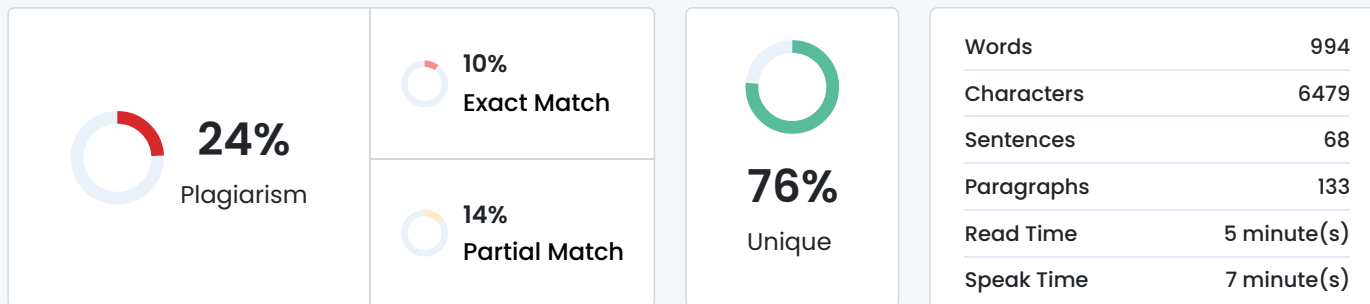https://datascience.stackexchange.com/questions/86632/why-is-it-wrong-to-train-and-test-a-model-on-the-same-dataset

---

Similarity 3%

Title:Water Quality Data - Dataset

The California Department of Water Resources (DWR) discrete (vs. continuous) water quality datasets contains DWR-collected, current and historical, chemical and ...

https://data.ca.gov/dataset/water-quality-data

Plagiarism Scan Report

24% Plagiarism	**10%** Exact Match / **14%** Partial Match	**76%** Unique

Words	994
Characters	6479
Sentences	68
Paragraphs	133
Read Time	5 minute(s)
Speak Time	7 minute(s)

Content Checked For Plagiarism

pH - This parameter measures the acidity or alkalinity of water on a scale from 0 to 14. According to the Environmental Protection Agency (EPA) guidelines, the pH of tap water should ideally fall between 6.5 and 8.5. The pH level is a significant determinant of the water's acid-base characteristics. The current study recorded pH values ranging from 6.52 to 6.83, which aligns with World Health Organization (WHO) standards.

Hardness - This refers to the quantity of soap that can dissolve in one liter of water. The primary contributors to water hardness are salts composed of calcium and magnesium. The duration of water's exposure to hardness-inducing substances affects its hardness in its natural state. Traditionally, hardness is defined by the water's capacity to form soap due to the precipitation of calcium and magnesium.

Total Dissolved Solids (TDS) - Water has the ability to dissolve a variety of chemicals and certain organic minerals or salts, including sodium, calcium, iron, zinc, bicarbonate ions, chloride ions, magnesium, and sulfates. These dissolved minerals can alter the water's appearance and contribute to unpleasant odors. A high TDS level indicates a significant presence of minerals in the water. For safe drinking, the recommended maximum TDS limit is 500 mg/l, with a desired limit of 100 mg/l.

Sulfates - Sulfates are organic compounds that occur naturally in various environments, including minerals, soil, rocks, air, groundwater, plants, and food within a given area. In the chemical industry, sulfates are primarily employed for commercial applications. Seawater typically contains approximately 2,700 mg/L Figure 2: data description of sulfate, while certain regions may exhibit much higher concentrations, reaching up to 1,000 mg/L. In contrast, most freshwater sources generally have sulfate levels ranging from 3 to 30 mg/L.

Conductivity - Pure water serves as an excellent insulator of electrical current. The level of dissolved particles within the liquid typically dictates its conductivity. Electrical conductivity (EC) assesses the efficiency with which these particles transmit electricity through their ionic interactions. According to the recommendations of the World Health Organization (WHO), the EC value should not exceed 400 S/cm.

Chloramines - The two main disinfectants employed in public water systems are chloride and chlorine. Ammonia is utilized alongside chlorine to purify drinking water. It is considered safe for drinking water to contain up to 4 mg/L of chlorine.

Potability - It is a metric for determining whether water is fit for human consumption. Unpotable equals zero (0), while potable is one (1).

A. Data Pre-processing

1. A. Data Pre-processing Enhancing data quality is essential during the processing phase of data analysis. In this stage, the Water Quality Index is established based on key parameters from the dataset. Data preparation refers to the transformation of collected data into a format suitable for machine learning algorithms. This step is crucial and serves as the foundational phase in the development of a machine learning algorithm. It is necessary to eliminate all instances where the value is zero, as zero is not a valid value; thus, such instances are discarded. The process of selecting feature subsets, which reduces data dimensionality and facilitates faster processing, involves the removal of irrelevant features and instances.

Figure 3: data pre-processing

Figure 4: Water Potability Distribution

B. Correlation Matrix

The correlation among all features is visualized using a heat map function. However, the heat map presented below indicates that there is no correlation among any of the features, suggesting that dimensionality reduction is not feasible.

C. Training and Testing of Data

In the field of machine learning, a model is directed to execute various tasks by utilizing a training dataset. This model undergoes training based on specific features extracted from the training data. For sentiment analysis, words or clusters of words are sourced from tweets. The model establishes relationships, comprehends concepts, formulates conclusions, and evaluates its confidence levels through the training data. The effectiveness of our data project is influenced by both

Figure 5: correlation matrix

the quality and quantity of the machine learning training data, as much as by the algorithms employed. Consequently, if the training set is accurately labeled, the model will be capable of learning the relevant

features.

D. Decision Tree

The decision tree is a machine learning algorithm primarily concerned with classification tasks. It features a structured classification system where the nodes represent various elements of a specific dataset. Within decision trees, there are two main types of nodes: decision nodes and leaf nodes.

E. Support Vector Machine

Support Vector Machine (SVM) is an algorithm employed in machine learning for the purpose of task categorization. It is commonly utilized in classification challenges. SVM distinguishes between two classes by transforming the data points into a high-dimensional space and subsequently identifying the optimal hyperplane.

F. Random Forest Classifier

Random Forest is an integral component of the supervised learning framework. It can be utilized for machine learning tasks that encompass both classification and regression. This technique is founded on the principle of ensemble learning, which involves the combination of multiple classifiers to tackle complex problems and improve the overall performance of the model. As suggested by its name, Random Forest operates as a classifier that employs numerous decision trees, each trained on different subsets of the input data, and aggregates their outputs to enhance the accuracy of predictions. Rather than relying on a single decision tree, the random forest aggregates the predictions from all trees and determines the final outcome based on the majority vote of these predictions.

G. XGBoost

Extreme Gradient Boosting is a framework that can run on multiple languages. It is popular supervised learning which works on large datasets. It is implemented on top of the gradient boost. The way the XGBoost algorithm is designed to work uses the parallelization concept. It uses sequentially generated shallow decision trees and a highly scalable training method to minimize overfitting to deliver accurate results.

Matched Source

Similarity 12%

Title: Water_Potability_Prediction

Solids (Total dissolved solids - TDS): Water has the ability to dissolve a wide range of inorganic and some organic minerals or salts such as potassium ...

https://www.kaggle.com/code/snehadridutta/water-potability-prediction

Similarity 5%

Title: The 6 Steps of a Data Analysis Process

In this article we focus on the 6 essential steps of a data analysis process with examples and addressing the core points of the process' methodology.

https://blog.bismart.com/en/data-analysis-steps-complete-guide

---

Similarity 5%

Title: Data transformation (statistics)

In statistics, data transformation is the application of a deterministic mathematical function to each point in a data set

https://en.wikipedia.org/wiki/Data_transformation_(statistics)

Similarity 4%

Title: Data Preparation for Machine Learning: The Ultimate Guide

https://www.pecan.ai/blog/data-preparation-for-machine-learning

---

Similarity 4%

Title: Machine learning

Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data

https://en.wikipedia.org/wiki/Machine_learning

Similarity 4%

Title: What is Training Data?

The algorithm develops confidence in its decisions by understanding the underlying patterns, relationships, and structures within a training dataset. The …

https://www.taus.net/resources/blog/what-is-training-data

Similarity 3%

Title:

towardsdatascience.com · a-complete-guide-toUnpacking the Power of Decision Trees: A Comprehensive Guide

Jan 5, 2022 · The Decision Tree is a machine learning algorithm that takes its name from its tree-like structure and is used to represent multiple decision stages and the possible response paths. The decision tree provides good results for classification tasks or regression analyses.

https://towardsdatascience.com/a-complete-guide-to-decision-trees-ac8656a0b4bb

---

Similarity 3%

Title: Support Vector Machine (SVM) Algorithm

Oct 10, 2024 — A Support Vector Machine (SVM) is a powerful machine learning algorithm widely used for both linear and nonlinear classification, as well as regression and …

https://www.geeksforgeeks.org/support-vector-machine-algorithm

---

Similarity 3%

Title:[Solved] The acronym DMAIC of Six Sigma method stands for

... problems, and improve the overall performance of the enterprise. The Six Sigma philosophy holds that every process can and should be repeatedly evaluated ...

https://testbook.com/question-answer/the-acronym-dmaic-of-six-sigma-method-stands-for-
-5f3b92cf9d5d6c0d1dda61f5

Similarity 3%

Title: Confusion matrix representation of the CNN model.

In addition to enhancing the predictive actuality of the dataset, 'Random Forest' employs multiple decision trees, each trained on different subsets of the ...

https://www.researchgate.net/figure/Confusion-matrix-representation-of-the-CNN-model_fig16_379574137

Similarity 3%

Title:INVESTIGATING THE EFFECT OF TRAFFIC SAMPLING ...

Jun 6, 2024 — based on the majority vote of these predictions. e) KNN: The K-Nearest Neighbour algorithm is a straightforward approach to Machine Learning ...

https://www.irjmets.com/uploadedfiles/paper/issue_6_june_2024/58611/final/fin_irjmets1717590700.pdf

---

Similarity 2%

Title:CIS feature selection based dynamic ensemble selection model for human ...

Apr 27, 2023 · The gradient descent approach was used to further improve the sequential model by reducing mistakes. Finally, by employing sequentially built shallow decision trees and a highly scalable training method that minimises overfitting, it produces dependable results .

https://link.springer.com/article/10.1007/s10586-023-04008-8

Similarity 2%

Title: Soap Bar Qualities

Hardness - This refers to the hardness of the soap bar. Higher is harder. A range of 29 to 54 is satisfactory for this soap quality.

http://www.soapcalc.net/info/soapqualities.asp

Similarity 2%

Title: What Causes High Conductivity In Water?

Sep 20, 2022 — According to the World Health Organization (WHO), electrical conductivity values should not exceed 400 microsiemens per centimeter (μS/cm). In ...

https://atlas-scientific.com/blog/what-causes-high-conductivity-in-water/?
srsltid=AfmBOoqeJ1GBHtswoTEe09C41HH7xF4L5BUJyKz93M0Ma-4Y6Flmj6AZ

Similarity 2%

Title: Data Quality and Quantity for Machine Learning

The efficacy of a machine learning (ML) algorithm's learning capabilities is subjective to the quality and quantity of the data it is fed.
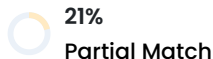
https://www.monolithai.com/blog/data-quality-and-quantity-for-machine-learning

---

Plagiarism Scan Report

23% Plagiarism

2% Exact Match

21% Partial Match

77% Unique

Words	883
Characters	5967
Sentences	69
Paragraphs	139
Read Time	5 minute(s)
Speak Time	6 minute(s)

Content Checked For Plagiarism

H. KNN
K-Nearest Neighbors (KNN) is a simple, non-parametric machine learning algorithm used for classification and regression tasks. It works by finding the "k" closest data points (neighbors) to a given input and making predictions based on the majority class or average value of these neighbors. KNN is easy to implement but can be computationally intensive, especially with large datasets, since it requires calculating distances between points.

I. Performance Metrics
Accuracy - Accuracy is measured as the total count of actual predictions to the available predictions and it is multiplied by 100.

Precision - The ratio of actual positives to the total available positives is known as precision.

Recall - It mainly focuses on type-2 errors the ratio of true positives to false negatives is called recall.

F1-score - The harmonic mean performance metric parameters precision with recall known as f1-score.

Algorithm
1. Import Libraries: Import necessary libraries like pandas, numpy, matplotlib, seaborn, and machine learning modules from sklearn.

2.Load Dataset: Load the datset from kaggle waterpotability.csv into a DataFrame.

3. Handle Missing Values: Impute missing values using the mean strategy with SimpleImputer.

4. Data Exploration and Visualization: Compute the correlation matrix and visualize it using a heatmap. Analyze the distribution of the target variable (Potability) with a bar chart.

5. Data Preprocessing: Separate features (X) and the target variable (y). Split the dataset into training and testing sets using an 80-20 split. Scale features using StandardScaler for uniform scaling.

6. Define and Train Models: Define multiple classifiers: Random Forest, Gradient Boosting, AdaBoost, and SVM. Train each model on the scaled training data.

7. Evaluate Models: Make predictions on the test data for each model. Calculate evaluation metrics: Accuracy, Classification Report, and Confusion Ma-

trix.Identify the model with the highest accuracy.

8. Cross-Validation: Perform 5-fold cross-validation on the best model and compute mean cross-validation scores.

9. Feature Importance (for Random Forest): If Random Forest is the best model, extract and visualize feature importances.

10. Save the Best Model: Save the trained best model to a .pkl file using joblib.

11. Load and Use the Model: Load the saved model for future predictions. Ensure input samples are scaled before making predictions.

12. Prediction Function: Define a function to predict water potability for a given sample. Use the best model to classify whether the water is potable or not.

13. Test Prediction: Provide sample inputs to the prediction function for testing. Interpret the model's output as "Potable" or "Not Potable".

Hardware Requirements

Hardware	Description
System	13th Gen Intel(R) Core(TM),1.90GHz
Hard Disk	512 GB
Monitor	HP P204v
Process	Intel i5-1340P

Table I: hardware requirements

Software Requirements

Software	Description
Operating System	Windows 11
Programming Language	Python 3.2
Database	Firebase
Tools	Jupyter Notebook, Google Colab, Python IDE

Table II: Software Requirements

Results and Discussion

• The figure 6 shows the accuracy of six different machine learning models: SVM, KNN, Decision Tree, Gaussian Naive Bayes, Random Forest, and XGBoost. The accuracy is fairly consistent with SVM, KNN, Random Forest, and XGBoost achieving higher values compared to Decision Tree.

Figure 6: Accuracy

Figure 7: Precision

• Figure 7 displays the precision for each model. Precision is highest for the Random Forest model, followed closely by KNN and XGBoost, while Gaussian Naive Bayes has slightly lower precision. Higher precision means fewer false positives for these models.

Figure 8: Recall

• Figure 8 illustrates recall values, where SVM and KNN show the highest recall, meaning these models are better at identifying true positives. Gaussian Naive Bayes has the lowest recall, indicating it misses more true positives.

• Figure 9 illustrates the F1-score, which is the harmonic mean of precision and recall. KNN has

Figure 9: F1-score

the highest F1-score, indicating a good balance between precision and recall. Gaussian Naive Bayes has the lowest F1-score, suggesting it is less effective in balancing precision and recall.

In summary, the Random Forest Classifier proved to be the most effective model, attaining an ideal equilibrium among accuracy, precision, recall, and F1-score. This model's exceptional balance suggests its suitability for applications that demand reliable accuracy across various evaluation metrics. The Random Forest Classifier demonstrated the highest performance in training the model, yielding an accuracy of approximately 70 percent when balanced with precision and recall.

Figure 10: Best Model

Figure 11: Sample prediction

Conclusions and Future Scope

Future urban environments stand to gain significantly from the implementation of real-time monitoring and evaluation of water quality, facilitated by advancements in machine learning techniques. This study presents the findings of our latest literature review and comparative analysis of recent research focused on assessing water quality through big data analytics and machine learning methodologies. Additionally, it provides insights into the challenges, requirements, and priorities for future research endeavors. The modeling and forecasting of water quality play a crucial role in environmental protection. The algorithm developed in this study enhances the efficacy of water quality classifiers. Our previous investigations into the performance metrics of machine learning algorithms revealed that the combination of Hyperparameter Tuning with the Random Forest Classifier yielded substantial improvements in various performance metrics of the models. This approach has resulted in enhanced performance metrics overall. Such a strategy can be further refined and applied to automated water quality monitoring systems.

Matched Source

Similarity 25%

Title:

[www.w3schools.com · python · python_ml_knnMachine Learning - K-nearest neighbors (KNN) - W3Schools](www.w3schools.com)

Using the input features and target class, we fit a KNN model on the model using 1 nearest neighbor: knn = KNeighborsClassifier(n_neighbors=1) knn.fit(data, classes) Then, we can use the same KNN object to predict the class of new, unforeseen data points. See more on w3schools

https://www.w3schools.com/python/python_ml_knn.asp

Similarity 25%

Title:How To Measure Accuracy and Precision in 5 Steps

Jul 30, 2024 — Accuracy is the degree of how close a calculated or measured value is to the actual value. It measures the statistical error, which is the ...

https://www.indeed.com/career-advice/career-development/how-to-measure-accuracy

Similarity 17%

Title:Understanding Model Performance Metrics: Precision, ...

Precision is the ratio of true positive predictions to the total predicted positives. It tells you how many of the positive predictions made by ...

https://medium.com/@jaberi.mohamedhabib/understanding-model-performance-metrics-precision-recall-f1-score-and-more-ad9ccdce7ff5

Similarity 10%

Title:Comparative Analysis of Encoding Methods in Regression ...

· Finally, we train each model on the scaled training data and evaluate it on the scaled test data. Scaling ensures that each feature contributes equally to the model fitting process, ...

https://link.springer.com/chapter/10.1007/978-3-031-68660-3_30

Similarity 6%

Title:Shriya Ellur - Consultant - Quation Solutions Private Limited

The Aim of this Machine Learning model is to classify whether the water is potable or not.(EDA, Logistic Regression, KNN, Naive Bayes, Decision Tree ...

https://in.linkedin.com/in/shriya-ellur-88b508191?trk=public_profile_browsemap_profile-result-card_result-card_full-click

---

Similarity 5%

Title:Model comparison based on performance Figure 7. Bar ...

Model comparison based on performance Figure 7. Bar graph shows the accuracy and precision values of all the algorithms used in the model-building process.

https://www.researchgate.net/figure/Model-comparison-based-on-performance-Figure-7-Bar-graph-shows-the-accuracy-and_fig5_369241086

Similarity 4%

Title:The F1 scores on four scenes. To clarify, we illustrate a ...

To further investigate their performance in different scenes, Figure 9 illustrates the comparison of F1 score in each scene. The traditional R-VLBC method ...

https://www.researchgate.net/figure/The-F1-scores-on-four-scenes-To-clarify-we-illustrate-a-sample-frame-for-each-scene-at_fig8_370018980

Similarity 4%

Title:Automated Water Quality Assessment Using Big Data ...

6 days ago — This shows that Big Data analytics and machine learning algorithms can interpret complex water quality d for public health and natural ...

https://www.researchgate.net/publication/385620035_Automated_Water_Quality_Assessment_Using_Big_Data_Ana

---

Similarity 4%

Title:Real-time water quality forecasting in rivers using satellite ...

by PG Whitehead · 2024 — Overall, integrating the biosensor data with real time water quality modelling and forecasting can play a critical role in monitoring and ...

https://www.frontiersin.org/journals/environmental-science/articles/10.3389/fenvs.2024.1331783/full

---

Similarity 3%

Title:Water quality classification using machine learning ...

by N Nasir · 2022 · Cited by 197 — Artificial Intelligence (AI) offers significant opportunities to help improve the classification and prediction of water quality (WQ). In this study, various AI ...