Groundwater Quality Assessment For Chattisgarh region Through Feature Selection and Predictive Models

immediate

Abstract- Groundwater is a vital source of freshwater in Chhattisgarh, underscoring the need to monitor its quality to ensure public health. This study employs multiple feature selection techniques to identify the most significant parameters influencing the Water Quality Index (WQI) and the potability of water. The feature selection methods applied include Correlation-Based Feature Selection (CFS), Recursive Feature Elimination (RFE), Entropy-Based Selection, and Principal Component Analysis (PCA). To evaluate the effectiveness of these techniques, two machine learning models, namely K-Nearest Neighbors (KNN) and Decision Tree, were trained using the selected features. A comparative performance analysis of these models was conducted, revealing that Recursive Feature Elimination (RFE) offered the best predictive performance, achieving an R^2 value of 0.92269. This demonstrates the efficiency of RFE in identifying the optimal set of features for assessing groundwater quality.

Keywords - Water Quality Index, Feature Selection, K-Nearest Neighbors, Decision Tree

Introduction

Assessing and ensuring water quality is a critical focus in environmental health research, particularly in regions dependent on groundwater as a primary water source [1] . Chhattisgarh is a mountainous, highly forested province in the Central Indian area. With an area of 135,192 sq. km, it is India's ninth-largest state. It has a population of over 30 million as of 2021, which makes it the nation's 17th most populous state. In the Samoda District, this issue is particularly pertinent due to the high dependency on groundwater. Effective and efficient methods to assess drinking water suitability are essential to protect public health and improve water management practices.

Traditional methods of assessing water quality in a specific region are labor-intensive and time-consuming. However, advancements in Machine Learning, Ensemble Learning, and Deep Learning offer innovative solutions that can minimize costs and effort in this area [2].

Groundwater plays a critical role in sustaining human life, serving as a vital source of drinking water for millions globally. Its quality is essential to human health, as it directly impacts the safety of the water consumed by populations. Several physical characteristics, such as pH, Total Dissolved Solids (TDS), Oxidation Reduction Potential (ORP), Dissolved Oxygen (DO), Temperature, and Salinity, along with major chemical components like cations and anions, act as indicators of groundwater quality, influencing its suitability for consumption[3][4].

Contaminated groundwater can be a significant pathway for the transmission of harmful pathogens, including bacteria, viruses, and protozoa, which may lead to serious waterborne diseases. These pathogens pose acute health risks, particularly when consumed through untreated drinking water, potentially triggering widespread disease outbreaks[5]. Ensuring the purity of groundwater, therefore, remains a public health priority, as the presence of contaminants can lead to long-term health implications for entire communities.

This study builds on such advancements, focusing on real-time data from Samoda District in Chhattisgarh. Our approach involved standardizing the data through normalization, followed by an analysis of missing values to ensure data quality. We then applied four feature extraction techniques—Correlation Analysis, Recursive Feature Elimination (RFE), Entropy-Based Selection, and Principal Component Analysis (PCA)—to identify the most informative features for predicting drinking water suitability. These extracted feature sets were evaluated using K-Nearest Neighbors (KNN) and Decision Tree models to determine predictive accuracy.

In this paper, we present a comparative analysis of various machine learning models to classify groundwater quality based on selected water quality parameters. The groundwater quality classes are assessed through predictive modeling techniques to determine drinking water suitability accurately. The paper is organized as follows: Section I provides an introduction to the study, Section II provides a Literature Review, Section Ill details the research methodology, Section III presents the results and discussion, and Section IV concludes the paper with insights and recommendations for future research.

Literature Review

Ahmadiyah et al. (2015) proposed a groundwater quality monitoring application aimed at increasing public awareness by providing real-time water quality data. The emphasis on accessible and accurate monitoring supports the value of your study's machine learning approach for improving predictive accuracy in groundwater quality assessments[6].

Giordano et al. (2010) designed an innovative monitoring system for sustainable groundwater management in coastal aquifers, emphasizing stakeholder conflict resolution to enhance groundwater monitoring acceptability and feasibility. This aligns with the collaborative and adaptive monitoring strategies needed in your study for effectively managing groundwater quality in Chhattisgarh[7].

R. V et al. (2023) proposed an optimized ensemble machine learning framework for water quality assessment, integrating the minimum redundancy maximum relevance technique to enhance model accuracy and reduce overfitting. Their approach to dimensionality reduction and model integration provides insights for your work on feature selection and ML model performance in WQI prediction[8].

Machine learning (ML) algorithms for both the classification and prediction of water quality are explored in [9]. Initially, the weighted arithmetic index method is employed to calculate the water quality index (WQI). Following this, Principal Component Analysis (PCA) is applied to the dataset to extract the most significant parameters influencing WQI. Various regression models are then used on the PCA-transformed data to predict WQI. For classification, the gradient boosting classifier is utilized to categorize the water quality status (WQS). The results demonstrate that the gradient boosting classifier performs effectively in accurately classifying WQS.

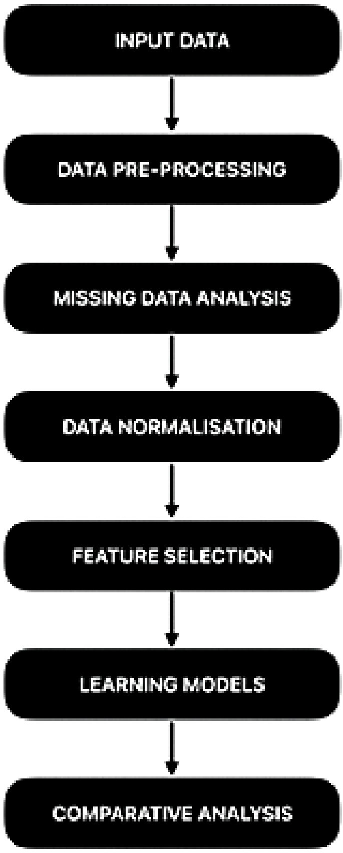

Fig. 1. Methodology

In [10], a comparison between traditional machine learning models and automated machine learning (AutoML) methods is presented for water quality assessment. The study considers nine key features: pH, turbidity, total dissolved solids (TDS), chemical oxygen demand (COD), sodium, phosphate, iron, nitrate, and chloride. The Synthetic Minority Oversampling Technique (SMOTE) is applied to address class imbalance by oversampling the minority class. Results are analyzed both before and after applying oversampling, highlighting the impact on model performance.

A. Dataset Description. The dataset used in this research was obtained from ___[11] and contains groundwater quality parameters. It includes measurements of calcium, magnesium, nitrate (NO_3), pH, sulfate (SO_4), total alkalinity (TA), total dissolved solids (TDS), and the water quality index (WQI). The data also assess the suitability of water for consumption across various locations. A total of 45 distinct areas within the Samoda district of Chhattisgarh were surveyed, with samples collected from each region for analysis.

The dataset includes eight key groundwater quality parameters:

- **Calcium (Ca)**: A measure of hardness in water, which impacts its potability.
- **Magnesium (Mg)**: Essential for water hardness and influencing taste.
- **Nitrate (NO_3)**: Excessive levels indicate contamination from fertilizers, posing health risks.
- **pH**: Indicates the acidity or alkalinity of the water, with potable water typically ranging between 6.5 and 8.5.
- **Sulfate (SO_4)**: High sulfate levels can cause gastrointestinal problems, affecting suitability for drinking.
- **Total Alkalinity (TA)**: Affects the water's buffering capacity, influencing its pH stability.
- **Total Dissolved Solids (TDS)**: Reflects the concentration of dissolved minerals, indicating the purity of the water.
- **Water Quality Index (WQI)**: A composite indicator summarizing the overall potability of the water based on the aforementioned parameters.

B. Data Preprocessing and Missing Value Analysis. Data preprocessing is a crucial step in data analysis and machine learning. This step involves transforming raw data into a suitable format for analysis, removing noise, and handling outliers to improve data quality. In particular, it offers benefits such as providing accurate analysis results, enhancing model performance, improving model generalization capabilities, and enabling faster model training[12].

The dataset obtained from [11] was initially in raw form. After filtering and transforming, the data was structured into a tabular format. However, some attributes contained missing values, and those missing places need to be filled in by suitable values to make the analysis fruitful. Missing data is common in real-world problems and can affect any statistical analysis significantly. A common way of dealing with this problem is to fill in the missing values. The technique used to impute data should be chosen very carefully to avoid incorrect inference about the data [13].

Several problems associated with missing values include loss of efficiency, bias, and reduction of statistical power. Different algorithms have a different impact on performance if the data has missing values [14]. The missing values are handled using K-Nearest Neighbors (KNN) Imputation. It works by identifying the k nearest neighbors to a data point with missing values and then imputing the missing value with the mean or median of the values of those nearest neighbors.

The data preprocessing phase also included outlier detection and removal to ensure data quality and reliability. Outliers are a kind of data which is dissimilar, inconsistent, irrelevant, or malicious compared to the rest of the data in a dataset. The process of identification and removal of such data from the dataset is called outlier detection [15]. Outliers were first identified through visual analysis using boxplots, followed by removal using the Interquartile Range (IQR) method.

The IQR method was implemented as follows:

Calculation of Quartiles

- $Q1$ (First Quartile): The 25th percentile of the data.
- $Q3$ (Third Quartile): The 75th percentile of the data.

Computation of IQR

$$\text{IQR} = Q3 - Q1 \tag{1}$$

Outlier Boundaries

- Lower Bound: $Q1 - 1.5 \times \text{IQR}$
- Upper Bound: $Q3 + 1.5 \times \text{IQR}$

Outlier Identification Any data point falling below the lower bound or above the upper bound was classified as an outlier and subsequently removed from the dataset.

The Python library used for this visualization is called Seaborn. Seaborn is a Python data visualization library built on top of Matplotlib. It provides a high-level interface for creating visually appealing statistical graphics.

C. Data Normalization. Data normalization[16] as one of the pre-processing strategies is utilized either to transform or scale the data in order to make an equal contribution of each attribute. For a given classification problem, the performance of any machine learning approach depends upon the quality of data in order to produce a generalized classification approach. Data normalization was performed to ensure that all features contribute equally to the analysis, as the collected parameters exhibit varying scales and units. Without normalization, features with larger ranges could disproportionately influence the outcomes of subsequent modeling and analysis.

To address this issue, the Min-Max Scaling technique was employed, which transforms each feature to a common scale between 0 and 1, thus preserving the relationships between values while ensuring that all parameters are treated equitably. The formula used is:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

where:

- X' is the normalized value,
- X is the original value,
- X_{min} and X_{max} are the minimum and maximum values of the feature.

D. Feature Selection. The primary objective of feature selection was to identify the most influential parameters for determining the potability of water. Feature selection is carried out to remove redundant and irrelevant features from error-prone data to improve predictive accuracy in software fault detection. Feature selection is also known as variable or attribute selection [17].

Feature selection [18] aims to select an optimal minimal feature subset from the original dataset and has become an indispensable preprocessing component in data mining and machine learning. Reducing the dataset to the most relevant features ensures improved model accuracy and efficiency while minimizing overfitting. Several feature selection techniques were employed to evaluate and retain the most significant attributes:

1. **Correlation-Based Feature Selection (CFS)**
   Correlation-based feature selection (CFS) is a filter method that aims to identify a subset of features that are highly correlated with the target variable while minimizing redundancy among themselves. This technique improves model performance by reducing noise and enhancing interpretability.

   A key tool used in CFS is the Pearson correlation coefficient. Pearson [19] proposed the Pearson correlation coefficient (PCC), which calculates linear correlation. A correlation coefficient of +1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no linear relationship.

   The primary Python package utilized for calculating the correlation matrix in this study is `pandas`, a widely used data analysis library. Specifically, the function `data.corr()` was employed to compute the Pearson correlation coefficient, returning a matrix of pairwise correlations between features. Features with high correlation with the target variable and low inter-correlation among themselves were selected.

2. **Recursive Feature Elimination (RFE)**
   The recursive feature elimination [20] method is effective at selecting and training features, providing high accuracy and ease of configuration. RFE is an iterative technique that builds models repeatedly, ranking features based on importance. At each iteration, the least significant features are removed until the optimal subset is identified, repeating until the best features for predicting water potability are selected./

   The RFE class, found within the `sklearn.feature_selection` module, was employed to automate the process of selecting the most important features for predicting water potability. Scikit-Learn is a comprehensive Python library for machine learning, offering various tools for data preprocessing, model building, and feature selection.

3. **Entropy-Based Feature Selection**
   Entropy-Based Feature Selection uses information gain (IG) or mutual information to measure the reduction in uncertainty about the target variable. This method selects a suitable subset of features based on entropy, where the subset with minimum entropy is chosen for data clustering. Subsequently, the dataset is clustered using two popular algorithms, k-Means and k-Medoids, to determine clustering accuracy [ref21].

   Entropy measures the uncertainty or randomness in data. If knowing a feature's value reduces uncertainty about the target variable, it is considered important. Information gain (IG) measures the reduction in entropy when a feature is known, while mutual information quantifies the amount of information shared between a feature and the target. IG is calculated as:

$$\text{IG}(Y, X) = H(Y) - H(Y|X) \tag{3}$$

   where $H(Y)$ is the entropy of the target and $H(Y|X)$ is the conditional entropy given the feature X.

   To implement entropy-based feature selection, the `mutual_info_regression` function from the `sklearn.feature_selection` module in Scikit-Learn was used. This function leverages mutual information (MI) to evaluate the dependency between input features and the target variable.

4. **Principal Component Analysis (PCA)**
   Principal Component Analysis (PCA) effectively transforms dependent variables in high-dimensional space into independent variables in a low-dimensional space while retaining maximum variability [22]. PCA-based algorithms enhance accuracy and efficiency, tailored to specific applications. PCA reduces the original features to a smaller set of principal components (orthogonal features), capturing maximum variance and uncovering underlying patterns.

   The data was normalized beforehand to ensure efficient PCA performance. The top components explaining the highest variance were selected for further model development. To implement PCA in this study, the `PCA` class from the `sklearn.decomposition` module in Scikit-Learn was used. The number of retained principal components was based on the cumulative variance explained by the components.

E. Learning Models. To evaluate the effectiveness of the feature selection methods, machine learning models were trained using the selected features. Machine learning can be used to analyze and predict the water quality based on the parameters like PH value, turbidity, hardness, conductivity, dissolved solids in water and other parameters[23]. Machine learning can be used for the analysis and prediction of water quality based on the water parameters that are provided as inputs for the particular machine learning model.

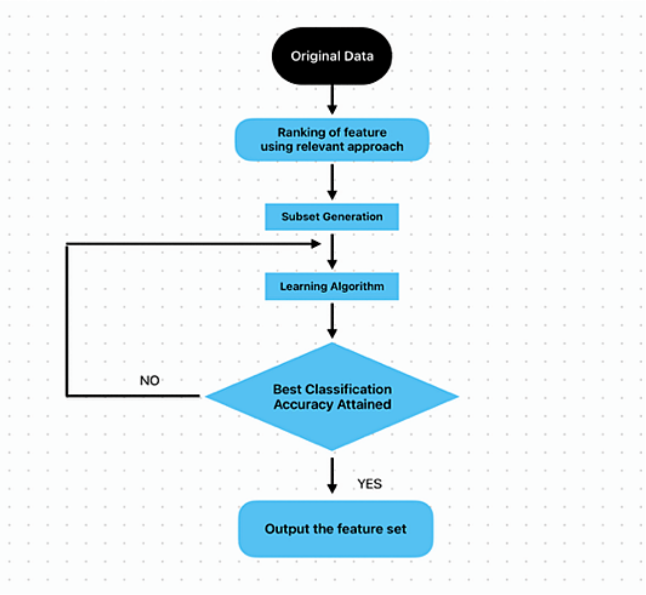
Original Data
Ranking of feature using relevant approach
Subset Generation
Learning Algorithm
NO
Best Classification Accuracy Attained
YES
Output the feature set

Fig. 2. Feature selection

The coefficient of determination (R^2 value) was employed as the primary metric to measure the models' predictive performance. By comparing the R^2 values across models trained with different feature sets, the optimal feature selection method and the most significant features for predicting water potability were identified.

The R^2 value, or coefficient of determination, measures how well the independent variables (selected features) explain the variance in the target variable (potability). It ranges between 0 and 1, where:

- $R^2 = 1$ indicates that the model perfectly predicts the target variable.
- $R^2 = 0$ implies that the model does not explain any variance in the target variable beyond the mean prediction.

The formula for R^2 is:

$$R^2 = 1 - \frac{RSS}{TSS} \tag{4}$$

where:

- RSS is the Residual Sum of Squares, which measures the error in the model's predictions,
- TSS is the Total Sum of Squares, which captures the variance in the target variable.

Several machine learning models were employed to evaluate the efficiency of feature selection:

1. **K-Nearest Neighbors (KNN)**
   K-Nearest Neighbors (KNN) is a non-parametric machine learning algorithm that makes predictions based

on the similarity between data points. KNN is a supervised learning algorithm that first appeared in 1968 by Hart as the k-nearest neighbor method. It typically uses the adjacent k sample categories to classify the current sample. By calculating the distance between the current sample and the k sample categories, the current sample can be classified into the nearest category [24].

For predicting water potability, the algorithm assigns a label—either suitable or unsuitable—by identifying the k-nearest neighbors of the test sample within the feature space. The Euclidean distance metric was used to measure the similarity between data points, and the optimal value of k was determined using cross-validation to minimize prediction error. The R^2 value was calculated to evaluate how well the KNN model could predict the potability of water based on the selected features.

In this study, the K-Nearest Neighbors (KNN) algorithm was implemented using the `KNeighborsRegressor` class from the Scikit-Learn library, which is well-suited for regression tasks. The model was trained using the `fit()` method, involving feeding the training dataset—comprising 80% of the total data—into the KNN regressor. The training dataset included both the input features (`X_train`) and the target variable (`Y_train`), which represented water potability as either "suitable" or "unsuitable." The remaining 20% of the data was reserved as the test set to evaluate the model's performance on unseen data.

2. **Decision Tree Model**

In this study, the Decision Tree algorithm was employed to classify water samples as suitable or unsuitable for consumption based on the selected groundwater parameters. The decision tree [25] is also a supervised learning method, mainly used for classification and regression problems. In classification, this method primarily focuses on data features. When represented visually, this classification process is called a decision tree.

The model works by recursively partitioning the dataset into subsets according to feature values, forming a hierarchical structure where each internal node represents a decision based on a feature threshold, and the leaf nodes correspond to class labels. Additionally, the Decision Tree model provides insights into the importance of individual features, helping identify the most influential parameters in predicting water potability. The coefficient of determination (R^2 value) was used to assess the model's effectiveness.

In this study, the Decision Tree algorithm was implemented using the `DecisionTreeRegressor` class from the Scikit-Learn library. The model was trained using the `fit()` method, which included the input features (`X_train`) and the target variable (`y_train`) from the training dataset, comprising 80% of the total data. The remaining 20% was designated as the test set, used to evaluate the model's performance on unseen data. The decision tree's hierarchical structure not only enabled accurate classification of water samples but also provided insights into the relative importance of each feature in the prediction process.

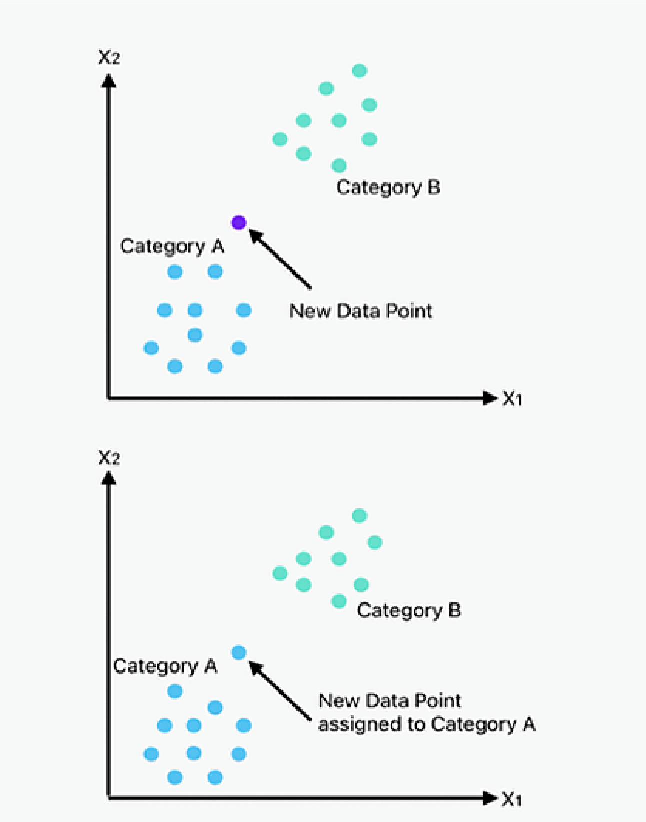

Fig. 3. K-Nearest Neighbor

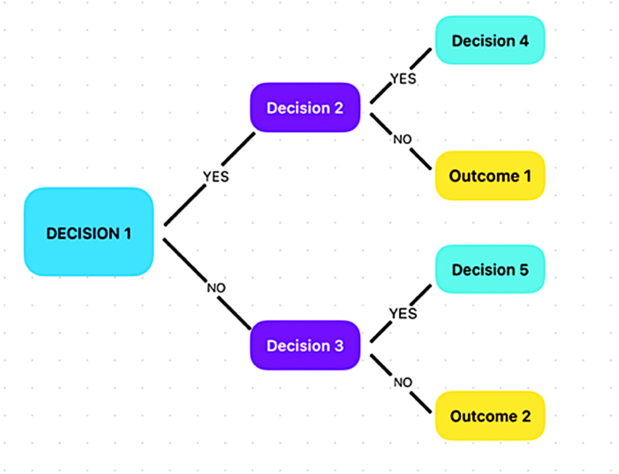

Fig. 4. Decision Tree model for classifying water samples based on selected groundwater parameters.

Results and Discussion

The table shows a comparative analysis of the results obtained when the effectiveness of Feature Selection techniques is evaluated by machine learning models. The results indicate that some techniques, such as Recursive Feature Elimination (RFE), outperform others, with Principal Component Analysis (PCA) following closely behind in terms of efficiency. This analysis highlights that certain features play a pivotal role in the accurate estimation of the Water Quality Index (WQI). Specifically, the most influential features identified are calcium (Ca), magnesium (Mg), nitrate (NO_3), Total Alkalinity (TA), and Total Dissolved Solids (TDS). These parameters are critical in determining the potability of groundwater, as their presence and concentration directly affect water quality.

Conclusion and Future Work

Groundwater quality analysis is a critical area of study, as water quality has direct and profound implications for public health. The study not only assessed water quality but also identified the most influential parameters that significantly affect water potability. These findings can support public health initiatives by enabling more targeted and efficient monitoring of groundwater sources, thus allowing for timely interventions and better management of water resources.

Future research could explore the use of ensemble models or deep learning techniques to improve the predictive accuracy of water potability models. Additionally, expanding the study to cover other geographic regions would provide a more comprehensive understanding of groundwater quality across different environments. Early detection of contaminants would not only protect communities from potential health crises but also support long-term strategies for maintaining safe drinking water supplies. As such, these advancements could significantly contribute to improving public health outcomes and safeguarding the region's water security.

Acknowledgement

We thank our institute NIT Raipur for helping us in this research and especially the Information Technology Department of NIT Raipur for providing constant support for the understanding of core Machine Learning concepts required for this Research. We also thank the creators of powerful languages like Python and concepts of Machine Learning which ease the research work and provide better results.

References

1. M. K. Jha, A. Shekhar, M. A. Jenifer, "Assessing groundwater quality for drinking water supply using hybrid fuzzy-GIS-based water quality index," *Water Research*, vol. 179, 2020.
2. I. Essamlali, H. Nhaila, M. El Khaili, "Advances in machine learning and IoT for water quality monitoring: A comprehensive review," *Heliyon*, vol. 10, no. 6, 2024.
3. G. Kanagaraj, L. Elango, S. G. D. Sridhar, G. Gowrisankar, "Hydrogeochemical processes and influence of seawater intrusion in coastal aquifers south of Chennai Tamil Nadu India," *Environ. Sci. Pollut. Res.*, vol. 25, no. 9, pp. 8989-9011, 2018.
4. T. V. Ramachandra, "Integrated Ecological Carrying Capacity of Uttara Kannada District Karnataka," *Sahyadri Conserv. Ser. 47 ENVIS Tech. Rep. 57. Cent. Ecol. Sci. IISc*, no. November, pp. 1-27, 2014.
5. A. H. Al-Fatlawi, "Microbial Risk Assessment to Estimate the Health Risk in Urban Drinking Water Systems of Al-Hilla City," 2018 11th International Conference on Developments in eSystems Engineering (DeSE), Cambridge, UK, 2018, pp. 225-230, doi: 10.1109/DeSE.2018.00034.
6. A. S. Ahmadiyah, "Analysis and design of groundwater quality monitoring application," 2015 International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation (ICAMIMIA), Surabaya, Indonesia, 2015, pp. 127-130.
7. R. Giordano et al., "An innovative monitoring system for sustainable management of groundwater resources: Objectives, stakeholder acceptability and implementation strategy," 2010 IEEE Workshop on Environmental Energy and Structural Monitoring Systems, Taranto, Italy, 2010, pp. 32-37.
8. M. R. V. V. R., S. N., S. S. Reddy, S. Bonthu, R. Rao Kurada, V. Vaishalini, "An Optimized Ensemble Machine Learning Framework for Water Quality Assessment System by Leveraging Forward Sequential Minimum Redundancy Maximum Relevance Feature Selection Method," 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2023, pp. 1-8.
9. I. Khan, d.S. Islam, N. Uddin, J. Islam, S. Nasir, M. K. Islam, "Water Quality Prediction and Classification Based on Principal Component Regression and Gradient Boosting Classifier Approach," *Journal of King Saud University - Computer and Information Sciences*, 2021.
10. D. V. V. Prasad, P. S. Kumar, "Automating water quality analysis using ML and auto ML techniques," 2024.
11. I. K., D. C. Jhariya, "Untitled work," 2024.
12. U.-J. Baek, M.-S. Lee, J.-T. Park, J.-W. Choi, C.-Y. Shin, M.-S. Kim, "Preprocessing and Analysis of an Open Dataset in Application Traffic Classification," 2023 24st Asia-Pacific Network Operations and Management Symposium (APNOMS), Sejong, Korea, Republic of, 2023, pp. 227-230.
13. A. Sadhu, R. Soni, M. Mishra, "Pattern-based Comparative Analysis of Techniques for Missing Value Imputation," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2020, pp. 513-518.
14. M. Brown, J. Kros, "Data Mining and Impact of Missing data," *Industrial Management and data systems*, 1999.
15. H. C. Mandhare, S. R. Idate, "A comparative study of cluster based outlier detection, distance based outlier detection and density based outlier detection techniques," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2017, pp. 931-935.
16. N. Singh, P. Singh, "Exploring the effect of normalization on medical data classification," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 2021, pp. 1-5.
17. S. Danasingh, A. Antony, G. Subramanian, A. A. Balamurugan, E. J. Leavline, "Towards higher accuracy in supervised learning and dimensionality reduction by attribute subset selection-A pragmatic analysis," Advanced Communication Control and Computing Technologies (ICACCCT) 2012 IEEE International Conference on., 2012.
18. P. Zhou, Y. Zhang, Z. Ling, Y. Yan, S. Zhao, X. Wu, "Online Heterogeneous Streaming Feature Selection Without Feature Type Information," *IEEE Transactions on Big Data*, vol. 10, no. 4, pp. 470-485, Aug. 2024.
19. E. Dominic, F. Móri, J. Székely, "On relationships between the Pearson and the distance correlation coefficients," *Statistics & Probability Letters*, vol. 169, 2021.
20. L. S. Matsa, G.-A. Zodi-Lusilao, F. B. Shava, "Recursive Feature Elimination for DDoS Detection on Software Define Network," 2021 IST-Africa Conference (IST-Africa), South Africa, South Africa, 2021, pp. 1-10.
21. M. K. Dhar, S. M. N. Hasan, T. R. Otushi, M. Khan, "Entropy-Based Feature Selection for Data Clustering Using k-Means and k-Medoids Algorithms," 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Bangalore, India, 2020, pp. 36-40.
22. A. Siddique, I. Hamid, W. Li, Q. Nawaz, S. M. Gilani, "Image Representation Using Variants of Principal Component Analysis: A Comparative Study," 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN), Chongqing, China, 2019, pp. 376-380.
23. N. S. Pagadala, M. Marri, A. Myla, B. Abburi, K. S. Ramtej, "Water Quality Prediction Using Machine Learning Techniques," 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2023, pp. 358-362.
24. B. A. Berhe, M. Elk, U. E. Dokuz, "INVESTIGATION OF IRRIGATION WATER QUALITY OF SURFACE AND GROUNDWATER IN THE KTAHYA PLAIN TURKEY," *Bulletin of the Mineral Research Exploration*, vol. 150, no. 150, pp. 145-162, 2015.
25. W. Lin, F. Weiwei, L. Haoran, X. Yongsheng, W. Jinzhuo, D. Kornack, et al., "Classification of Handheld Laser Scanning Tree Point Cloud Based on Different KNN Algorithms and Random Forest Algorithm," *Forests*, vol. 12, no. 3, pp. 2127-2130, Dec. 2021.