GROUND WATER QUALITY PREDICTION IN GUJARAT PROVINCE

Under Dr. Rajeev Kumar Gupta mentorship.

PROJECT

ABHIJEETSINGH MEENA (20BCP241)
SMIT RAVAL (20BCP259)

CONTENT

ABSTRACT

- This study investigates groundwater quality prediction using Entropy based Water Quality Index (EBWQI) and Weighted Arithmetic Water Quality Index (WAWQI) formulas. It deals with various regression models and classification models to predict the ground water quality indices.

- After downloading the PDF file from the government website, we extracted the data from it.

- The analysis of the results shows the best performance in the CatBoost Model with an accuracy of approximately 96 % followed by LightGBM and XGBoost model with the accuracy of 95% accuracy and then followed by RandomForest Classifier with the value of 93% accuracy.

INTRODUCTION

- Water is an essential resource crucial for sustaining life on Earth, requiring a certain quality to support various species, including humans[1].
- While covering a significant portion of the planet's surface, only a small fraction is freshwater, with groundwater being vital for activities like irrigation and drinking but susceptible to contamination from human activities.
- Researchers assess groundwater suitability using parameters like pH and Electrical Conductivity (EC)[5], but this can be challenging and costly[6].
- As a cost-effective alternative, water quality indices like WAWQI and EBWQI[7] have emerged. Machine learning techniques have become prominent in predicting groundwater quality for drinking purposes[8], outperforming traditional statistical methods[9].

LITERATURE SURVEY

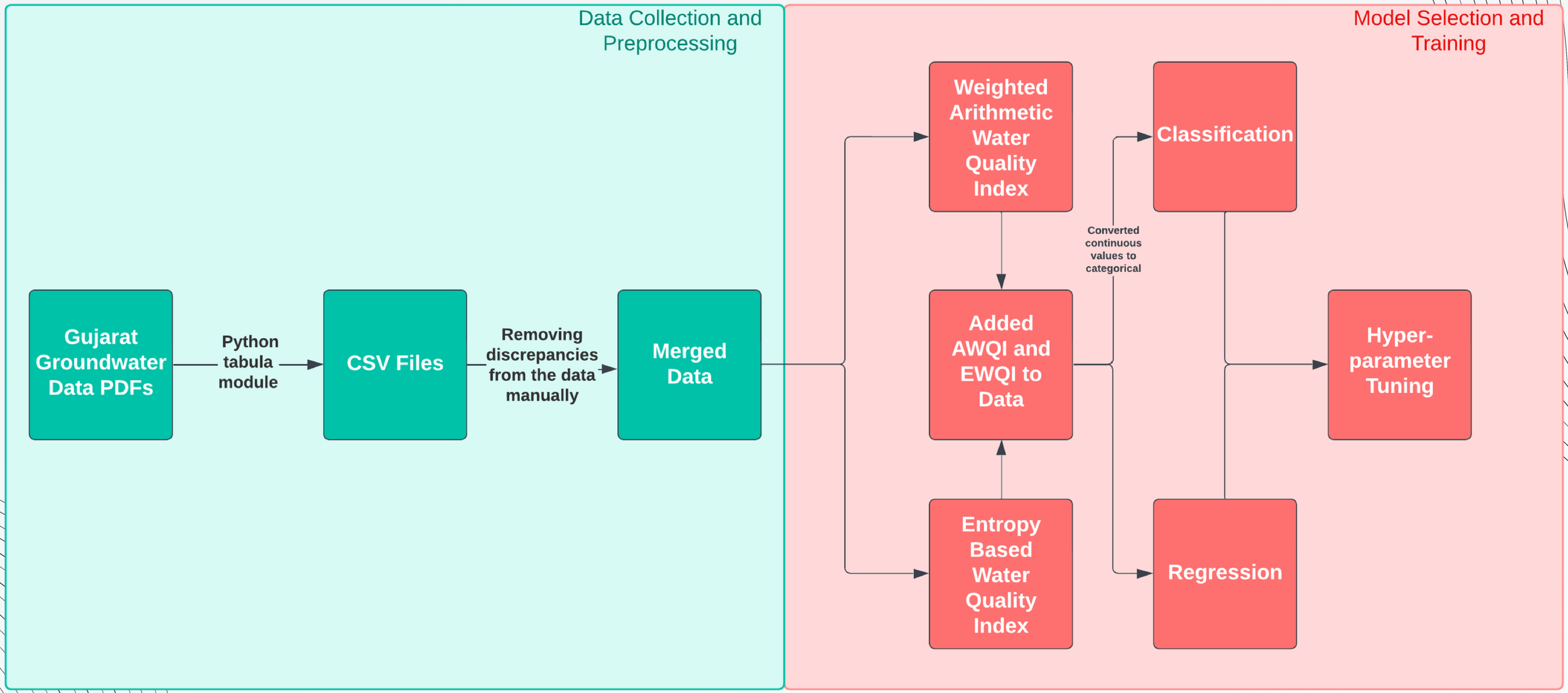
Paper Name	Paper's Insights
Prediction of groundwater quality indices using machine learning algorithms	The methods in our project were taken from a paper that compares three machine learning algorithms (Deep Neural Network, Gradient Boosting Machine, Extreme Gradient Boosting) for predicting groundwater quality indices (Entropy Water Quality Index and Water Quality Index) in Haryana, India. Results indicate that the Deep Neural Network outperforms the other models, showing lower error values and higher accuracy for both indices. Electrical Conductivity (EC) is identified as the most significant input parameter, while pH is the least significant. The study suggests using DNN models to predict groundwater quality for potability assessment, emphasizing the importance of machine learning in this context [2].
Groundwater Quality Evaluation for Potable Use and Associated Human Health Risk in Gaobeidian City, North China Plain	The second paper assesses groundwater quality in Gaobeidian City and its potential health risks for children and adults. The entropy weight method determines parameter weights for groundwater quality evaluation. Parameters including NO2-, Fe, As, Cr6, and NO2-N are key influencers. The study classifies groundwater quality as Excellent Water (class 1) based on entropy weighted water quality index. Groundwater quality deterioration is observed in the southwest. Natural factors are presumed to impact groundwater chemistry more than human activities. Certain water samples pose noncarcinogenic health risks to both adults and children, primarily due to F-, NO2-, NO3-, and Cr6 concentrations [4].
Groundwater Quality: The Application of Artificial Intelligence	Another paper we read the study develops a hybrid model using single exponential smoothing (SES) with bidirectional long short-term memory (BiLSTM) and adaptive neuro-fuzzy inference system (ANFIS) to predict water quality (WQ) in Al-Baha, Saudi Arabia. Data were split into training and testing phases, and efficiency statistics assessed the SES-BiLSTM and SES-ANFIS models' prediction abilities. Both models performed well in predicting water quality index (WQI), but SES-BiLSTM had higher accuracy (R 99.95, RMSE 0.00910) during testing compared to SES-ANFIS (R 99.95, RMSE 2.2941 x 100-07). Results suggest both models can predict WQI accurately, potentially enhancing water quality [11].

PROBLEM DEFINITION

- The project aims to address the challenge of predicting groundwater quality in the Gujarat province. With increasing industrialization, agricultural activities, and urbanization, groundwater contamination has become a significant concern in the region.

- The lack of accurate and timely assessment tools hampers effective water resource management efforts. Therefore, the project seeks to develop a robust predictive model that can reliably forecast groundwater quality, aiding in informed decision-making for sustainable water management practices in Gujarat.

PROPOSED METHODOLOGY



PROPOSED METHODOLOGY

Weighted Arithmetic Water Quality Index

1. In the first instance, 'assigning weight': twelve parameters are weighted according to its relative importance. The weight 5 is assigned to the most significant parameters and 1 is for the least significant. The relative weight (W_i) is obtained from Equation (15) and has been shown in Table 3.

(15) $$W_i = \frac{wi}{\sum_{i=0}^{n} wi}$$

where, wi is the weight of each parameter and *n* is the number of parameters.

2. In second instance, 'quality rating (q_i) scale calculation' firstly, the water sample concentration is multiplied by 100 and then the results are divided by its limits given by the BIS (2015) in Equation (16):

(16) $$q_i = \frac{C_i}{S_i} \times 100$$

where, C_i is the concentration of chemicals in the water sample in (mg/l), and S_i is the drinking water standard for each chemical parameter BIS (2015).

3. In the third instance, 'calculation of water quality index (WQI)', firstly SI_i (water quality index of the i^{th} parameter) value is calculated by Equation (17) and WQI equals the sum of all the values of SI_i of each parameter as by Equation (18):

(17) $$SI_i = W_i \times q_i$$

(18) $$\text{WQI} = \sum_{i=1}^{n} SI_i$$

Note : From "Prediction of groundwater quality indices using machine learning algorithms" by IWA Publishing, 2021, (https://iwaponline.com/wpt/article/17/1/336/85564/Prediction-of-groundwater-quality-indices-using)

PROPOSED METHODOLOGY

Entropy Based Water Quality Index

In first instance, normalization is carried out on initial matrix 'X'. After this, the standard matrix as (7) & (8) is stated as Y =(y_{ij}) * (m × n) where:

(7) $$X = \begin{bmatrix} x_{11} & x_{1n} \\ x_{m1} & x_{nm} \end{bmatrix}$$

(8) $$Y = \begin{bmatrix} y_{11} & y_{1n} \\ y_{m1} & y_{nm} \end{bmatrix}$$

In Equations (7) and (8) *m* and *n* are the number of groundwater samples and parameter for a sample respectively.

Then the value of y_{ij} is obtained by Equation (9):

(9) $$y_{ij} = \frac{x_{ij} - (x_{ij})_{\min}}{(x_{ij})_{\max} - (x_{ij})_{\min}}$$

where, x_{ij} is the j^{th} evaluation index of the i^{th} groundwater sample.

After calculating the standardized value, the ratio of index value of the *j* index using *i* sample is calculated by Equation (10):

(10) $$P_{ij} = \frac{y_{ij}}{\sum_{i=1}^{m} y_{ij}}$$

Next step is to calculate entropy weight w_j and information entropy e_j by Equations (11) and (12) as:

(11) $$e_j = -\frac{1}{\ln m}\sum_{i=1}^{m} y_{ij} \ln y_{ij}$$

Note : From "Prediction of groundwater quality indices using machine learning algorithms" by IWA Publishing, 2021, (https://iwaponline.com/wpt/article/17/1/336/85564/Prediction-of-groundwater-quality-indices-using)

PROPOSED METHODOLOGY

Entropy Based Water Quality Index

(12) $$w_j = \frac{1 - e_j}{\sum_{j=1}^{n} (1 - e_j)}$$

To calculate $EWQI$, a quality rating scale q_j for each parameter is assigned where q_j is obtained from the following Equation (13)

(13) $$q_j = \frac{c_j}{s_j} \times 100$$

where, c_j represents the concentration of the parameter (mg/l), and s_j denotes the water standards of groundwater for each drinking parameter according to Bureau of Indian standards (BIS). The EWQI can then be calculated by the following Equation (14):

(14) $$EWQI = \sum_{j=1}^{n} w_j \, q_j$$

Note : From "Prediction of groundwater quality indices using machine learning algorithms" by IWA Publishing, 2021, (https://iwaponline.com/wpt/article/17/1/336/85564/Prediction-of-groundwater-quality-indices-using)

CURRENT STATUS

- The initial phase of our study involved a comprehensive analysis of the topic, which included a thorough review of relevant research papers to establish a foundational understanding. Subsequently, data acquisition was undertaken by retrieving pertinent information from government websites in PDF format. We then meticulously extracted and compiled this data into a well-structured CSV file encompassing multiple years for comprehensive analysis. Following this, rigorous data cleaning procedures were implemented to ensure the integrity and quality of the dataset.

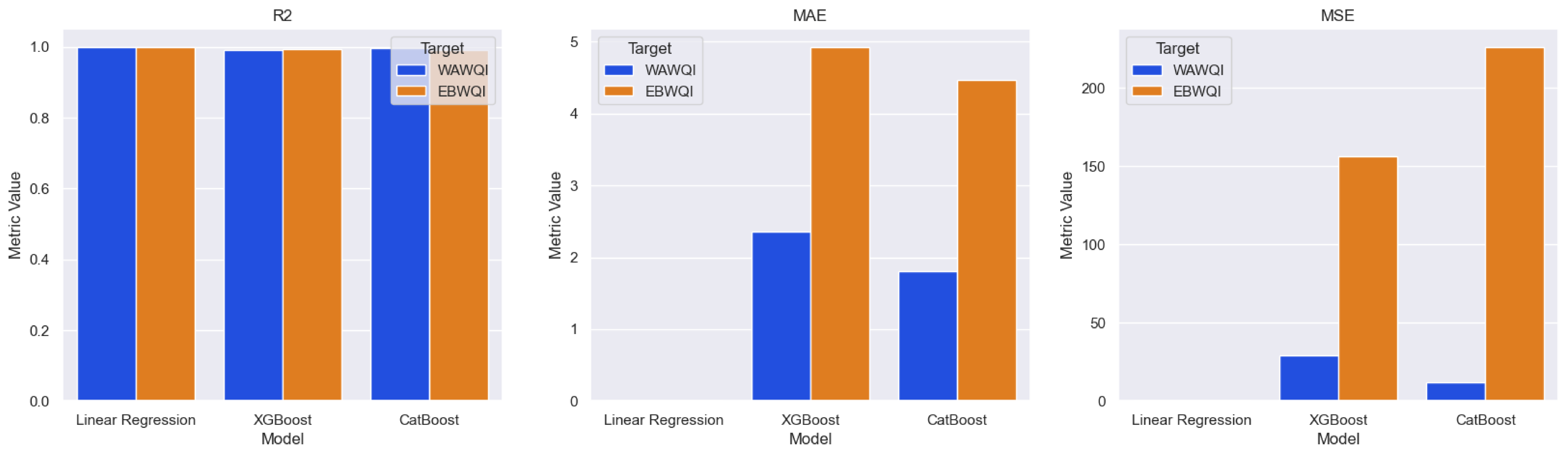
- Furthermore, in our pursuit to explore alternative methodologies beyond the WAWQI, we delved into researching various other methods. Upon establishing a robust dataset, we applied both EBWQI and WAWQI formulas to evaluate groundwater quality.

- To assess predictive performance, an array of regression and classification models was employed, enabling a comparative analysis between methodologies and determining the superior approach among EBWQI and WAWQI. Hyperparameter tuning techniques were then utilized to optimize model performance.

- Finally, the ensemble method of Voting Classifier was applied to further enhance predictive accuracy by aggregating results from multiple classification models.

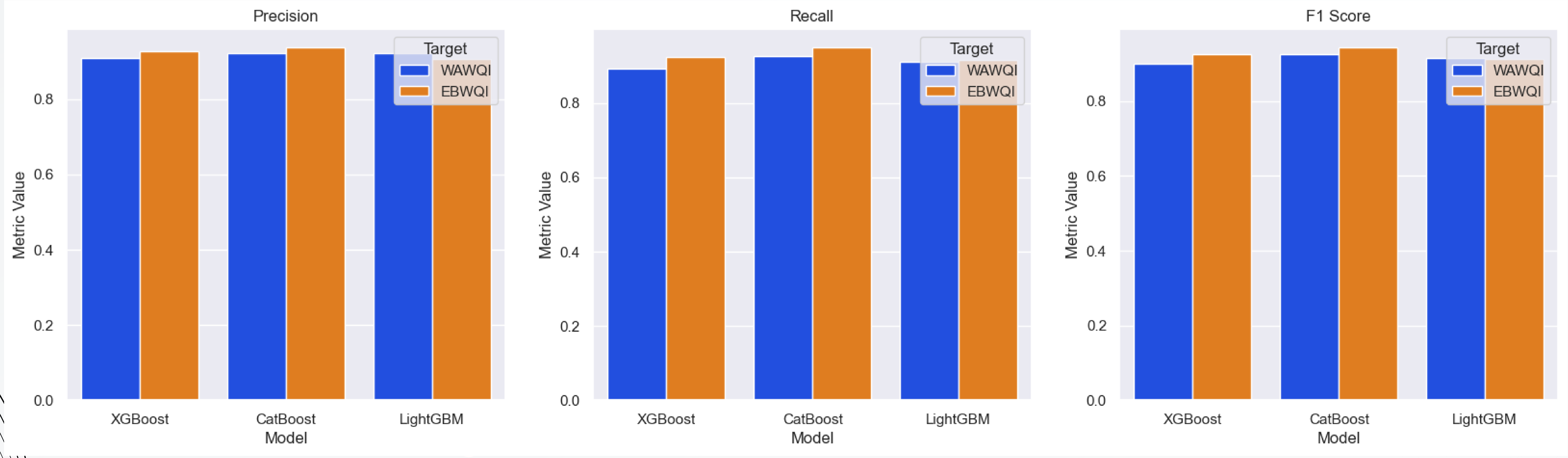
RESULTS

REGRESSION



RESULTS

CLASSIFICATION



GANTT CHART

Task	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9
Researched about the topic and collected the papers and realated documents to collect information about the domain knowledge									
Dataset Exploration and topic related materials collection									
Learn about the topic and understand the dataset. Also, read research papers to find the best ways to solve the problem. Collected data on Gujarat's groundwater from 2013 to 2022.									
Scraped data from a PDF for all years and made one dataset. Implemented WAWQI and gave weights to features. Sorted water into 5 groups based on WQI. Used regression and classification models like Linear Regression and RandomForest to predict water quality for test data. Fine-tuned RandomForest for better results.									
Applied other Gradient Boosting Algorithms like Catboost, LightGBM and XGBoost. Performed Hyperparameter Tuning for all the gradient boosting algorithms using Optuna Library									
Collected Data PDFs for years from 2014-2021 and extracted data from them. Researched and found EBWQI. Applied gradient boosting algorithms like on 5 years dataset (2017-2021) and did in-depth data visualization for 2021. Found worst water quality in patan and best water found in gir somnath.									
Removed small data discrepancies. Implemented EBWQI on the data. Applied various regression and classification models to draw a comparison between 2 methods EWQI and WQI. Applied and tuned Gradient Boosting Algorithms and used KFold CV for it.									
Read research papers for unearth possibilities of further experimentations.									
Merged different notebooks into a single master notebook and learned and applied Voting Classifier.									

CONCLUSION

- In conclusion, our study focused on predicting groundwater quality by applying the Weighted Arithmetic Water Quality Index (WAWQI) and Entropy based Water Quality Index (EBWQI) formulas to classify water into five distinct categories based on various parameters. By calculating WAWQI and EBWQI values, we were able to assess the overall quality of groundwater samples.

- Furthermore, we utilized regression and classification techniques to analyze the relationship between input parameters and WAWQI/EBWQI values, as well as to predict the category of groundwater quality and to compare different model's performance. Among the models evaluated, the CatBoost model demonstrated the highest results in classification tasks.

- Overall, our findings suggest that the combination of WAWQI and EBWQI formulas, coupled with advanced machine learning techniques such as CatBoost, can effectively predict groundwater quality and aid in making informed decisions regarding water resource management and environmental conservation efforts.

- The analysis of the results shows the best performance in the CatBoost Model with an accuracy of approximately 96 % followed by LightGBM and XGBoost model with the accuracy of 95% accuracy and then followed by RandomForest Classifier with the value of 93% accuracy.

REFRENCES

[1]Mohamed, I., Othman, F., Ibrahim, A. I. N., Alaa-Eldin, M. E., & Yunus, R. M. (2014, November 30). Assessment of water quality parameters using multivariate analysis for Klang River basin, Malaysia. Environmental Monitoring and Assessment. https://doi.org/10.1007/s10661-014-4182-y

[2]Raheja, H., Goel, A., & Pal, M. (2021, December 1). Prediction of groundwater quality indices using machine learning algorithms. Water Practice & Technology. https://doi.org/10.2166/wpt.2021.120

[3]Chaudhari, A. N., Mehta, D., & Sharma, N. (2021, March 24). An assessment of groundwater quality in South-West zone of Surat city. Water Science & Technology: Water Supply. https://doi.org/10.2166/ws.2021.083

[4]Fu, X., Dong, Z., Gan, S., Wang, Z., & Wei, A. (2021, November 22). Groundwater Quality Evaluation for Potable Use and Associated Human Health Risk in Gaobeidian City, North China Plain. Journal of Chemistry. https://doi.org/10.1155/2021/3008567

[5]Vasanthavigar, M., Srinivasamoorthy, K., Vijayaragavan, K., Ganthi, R. R., Chidambaram, S., Anandhan, P., Manivannan, R., & Vasudevan, S. (2010, January 21). Application of water quality index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamilnadu, India. Environmental Monitoring and Assessment. https://doi.org/10.1007/s10661-009-1302-1

[6]Kumar, V., Amarender, B., Dhakate, R., Sankaran, S., & Kumar, K. (2014, June 15). Assessment of groundwater quality for drinking and irrigation use in shallow hard rock aquifer of Pudunagaram, Palakkad District Kerala. Applied Water Science. https://doi.org/10.1007/s13201-014-0214-6

[7]Kumar, P. J. S., & Augustine, C. M. (2021, February 23). Entropy-weighted water quality index (EWQI) modeling of groundwater quality and spatial mapping in Uppar Odai Sub-Basin, South India. Modeling Earth Systems and Environment. https://doi.org/10.1007/s40808-021-01132-5

[8]Wang, D., Wu, J., Wang, Y., & Ji, Y. (2019, June 10). Finding High-Quality Groundwater Resources to Reduce the Hydatidosis Incidence in the Shiqu County of Sichuan Province, China: Analysis, Assessment, and Management. Exposure and Health. https://doi.org/10.1007/s12403-019-00314-y

[9]Adimalla, N., Qian, H., & Li, P. (2020, December 1). Entropy water quality index and probabilistic health risk assessment from geochemistry of groundwaters in hard rock terrain of Nanganur County, South India. Geochemistry. https://doi.org/10.1016/j.chemer.2019.125544

[10]Perumal, B., Rajarethinam, N., Velusamy, A. I. S., & Sundramurthy, V. P. (2023, November 7). Water Quality Prediction Based on Hybrid Deep Learning Algorithm. Advances in Civil Engineering. https://doi.org/10.1155/2023/6644681

[11]Al-Adhaileh, M. H., Aldhyani, T. H. H., Alsaade, F. W., Al-Yaari, M., & Albaggar, A. K. A. (2022, August 24). Groundwater Quality: The Application of Artificial Intelligence. Journal of Environmental and Public Health. https://doi.org/10.1155/2022/8425798

THANK YOU