Introduction In this report, we explore the development of a predictive model for water levels, addressing challenges posed by unorganized and sporadic data. Leveraging Random Forest models and time series analysis, we aim to forecast water levels for the next three years.

Data Preprocessing and Exploratory Data Analysis (EDA)

- The dataset initially exhibited irregularities, with significant missing data and non-continuous entries.
- Columns with predominantly zero values were identified, leading to the removal of Volume_CSAL, Volume_CSA, and Depth_to_Groundwater_DIEC.
- High correlation among certain columns was observed, prompting the removal of redundant features.

Model 1: Random Forest for Target Columns

- A Random Forest model (Model 1) was constructed to predict three key water level columns.
- This model was trained on available data, addressing the challenge of missing and sporadic entries.

Model 2: Time Series Analysis for Other Columns

- Another Random Forest model (Model 2) was developed for time series analysis, focusing on columns beyond the three target water level variables.
- Extended data for the next three years was generated, and additional features were derived by breaking down date columns into day, month, and year components, including their powers up to the fifth degree.

Predictions and Smoothing

- Model 1 was applied to the extended dataset, predicting the three water level columns.
- The results were further smoothed using a moving average, enhancing the robustness of the predictions.

Conclusion This predictive approach integrates Random Forest models for both target and extended columns, addressing the complexities of unorganized and sporadic water level data. By combining machine learning and time series analysis, our methodology provides a reliable framework for forecasting water levels over an extended period.

Recommendation

- Consider further feature engineering and model refinement for continuous improvement.
- Periodically update the model as more data becomes available.

This approach ensures a comprehensive understanding of water level dynamics, enabling proactive planning and resource management.

Correlation between parameters