**Title:** Exploratory Data Analysis and Machine Learning Modeling for Groundwater Situation Prediction

**Introduction:**

Groundwater is a vital natural resource essential for various human activities, including agriculture, industry, and domestic use. Understanding the factors influencing groundwater availability and predicting groundwater situations accurately are crucial for sustainable water resource management. In this project, we conducted exploratory data analysis (EDA) and developed machine learning models to predict groundwater situations based on a dataset of groundwater-related variables.

**Dataset Description:**

The dataset used in this project contains information on groundwater availability based out of different states in India, rainfall, usage, and other relevant factors across different regions. It includes both numerical and categorical variables, such as groundwater availability for future irrigation use, recharge from rainfall during monsoon and non-monsoon seasons, and the current groundwater situation categorized as 'EXCESS,' 'MODERATED,' 'SEMI-CRITICAL,' or 'CRITICAL.'

**Methodology:**

Exploratory Data Analysis (EDA): I started by exploring the dataset through descriptive statistics, data visualization, and correlation analysis. Visualizations such as count plots, scatter plots, and heatmaps were used to understand the distribution of variables, identify patterns, and analyze the relationships between them.

Preprocessing and Feature Engineering: To prepare the data for machine learning modeling, I performed preprocessing steps such as handling missing values, encoding categorical variables using one-hot encoding, and scaling numerical features to ensure uniformity across the dataset.

**Model Development:**

**Logistic Regression:** Initially, I employed logistic regression to predict groundwater situations. The model was trained on the preprocessed data and evaluated using accuracy metrics and confusion matrices.

**Decision Tree Classifier:** Additionally, a decision tree classifier was trained to predict groundwater situations. I assessed the model's performance and analyzed feature importances to understand the key factors influencing groundwater situations.

**Random Forest Classifier:** Finally, I utilized a random forest classifier to further improve prediction accuracy. Different configurations of the random forest model were experimented with, and their performance was compared.

**Results and Discussion:**

The exploratory data analysis revealed insights into the distribution of groundwater-related variables and their correlations. For instance, I observed strong correlations between rainfall, recharge, and groundwater availability.

Machine learning models, including logistic regression, decision tree classifier, and random forest classifier, were developed to predict groundwater situations. The models achieved varying levels of accuracy, with random forest showing the highest performance.

Feature importance analysis provided valuable insights into the factors contributing most to groundwater situations, such as rainfall patterns, recharge rates, and usage levels.

**Conclusion:**

In conclusion, this project demonstrated the effectiveness of exploratory data analysis and machine learning modeling in understanding groundwater dynamics and predicting groundwater situations. By leveraging these techniques, policymakers and stakeholders can make informed decisions to manage and conserve groundwater resources effectively, ensuring their sustainable use for future generations.

**Future Work:**

Future work could involve refining the machine learning models, incorporating additional features or datasets, and exploring advanced techniques such as deep learning for groundwater prediction. Additionally, efforts to collect more comprehensive and real-time data would enhance the accuracy and reliability of predictive models.

**References**

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0276155>

[Welcome to Bhuvan | ISRO's Geoportal | Gateway to Indian Earth Observation (nrsc.gov.in)](https://bhuvan-app1.nrsc.gov.in/gwis/)

[NWIC](https://nwic.gov.in/data)

[seaborn: statistical data visualization — seaborn 0.13.2 documentation (pydata.org)](https://seaborn.pydata.org/)