# Effects of rainfall on groundwater level based on Usage, Population and Climate conditions.

Akshat Saihgal (2110110933) & Pranav Uppal (2110110382)

Under Dr. Sonia Khetarpaul & Dr. Suchi Kumari

Major Project

# ABSTRACT

The project investigates the impact of rainfall on groundwater levels, considering factors such as population, water usage, and climate conditions. Rainfall significantly influences groundwater levels, particularly in karst areas, where the rapid response of groundwater to rainfall is observable. The study explores the fractal behavior of rainfall and its subsequent effects on groundwater recharge and level fluctuations. By analyzing the relationship between rainfall patterns, groundwater exploitation, and climate variables, this project aims to understand the processes driving groundwater dynamics. Here, rainfall changes account for 22.08%, while groundwater exploitation contributes 77.92%. The project employs advanced machine learning algorithms, including Random Forest, Gradient Boost Regressor, and Linear Regression, to model and predict groundwater availability. Additionally, techniques like Explainable AI (e.g., SHAP and LIME) will be used to interpret the model predictions and offer insights into the key factors influencing groundwater levels. Exploratory Data Analysis (EDA) will further inform the modeling process, uncovering patterns and relationships between rainfall, groundwater usage, and other climatic variables. The expected output is a predictive tool that estimates future groundwater availability, aiding in effective water resource management. This system will be developed with a Python backend, while the frontend will utilize HTML, CSS, and JavaScript for user interaction and visualization.

**Keywords -**

Rainfall, Groundwater levels, Machine learning, Random Forest, Gradient Boost Regressor, Linear Regression, Explainable AI, SHAP, LIME, EDA, Water management, Hydrology.

# CONTENT

# 1. DECLARATION

I hereby declare that the project work entitled "Effects of rainfall on groundwater level based on Usage, Population and Climate conditions" submitted to the SNU Delhi-NCR, is a record of an original work done by me under the guidance of Dr. Md. Sonia Khetarpaul & Dr. Suchi Kumari, Dept Of Computer Science & Engineering, Shiv Nadar University, and this project work is submitted in the partial fulfillment of the requirements for the award of the degree of Bachelors of Technology in Computer Science & Engineering.

(Sign)

Akshat Saihgal
Pranav Uppal

# 2. ACKNOWLEDGMENTS

I take this opportunity to thank Dr. Sonia Khetarpaul & Dr. Suchi Kumari , Head of Dept – Dr. Sweta Mishra, Dean -Students Activities, and other faculty members who helped in preparing the guidelines.

I extend my sincere thanks to one and all of SNU Computer Science Department for the completion of this document on the project report format guidelines.

# 3. Introduction

**Problem Statement-**

Rainfall plays a vital role in influencing groundwater levels, particularly in regions where water demand, population growth, and climatic conditions exert pressure on groundwater resources. In India, where groundwater is a critical source of water supply for both agricultural and domestic use, understanding the relationship between rainfall and groundwater dynamics is essential. Rainfall variability can directly impact groundwater recharge, especially in regions with high water extraction rates. In areas with karst formations, groundwater responds rapidly to rainfall, and the interaction between rainfall and groundwater chemistry offers insights into hydrogeochemical processes. Furthermore, fluctuations in groundwater levels due to rainfall can affect water flow within the soil, potentially influencing slope stability. In regions like northern India, rainfall and groundwater over-extraction together contribute to significant changes in groundwater levels. For example, studies indicate that rainfall accounts for a smaller proportion of groundwater level changes compared to excessive groundwater exploitation. However, technologies such as rainwater harvesting and artificial recharge systems have the potential to mitigate groundwater depletion by capturing rainfall and stabilizing groundwater reserves. This interplay between rainfall, population-driven water use, and climate variability underscores the need to manage groundwater resources effectively in India.

Climate change affects groundwater in several direct and indirect ways. Major direct drivers are changes in precipitation, snowmelt, and evapotranspiration. It is known that the groundwater storage reduction caused by pumping could easily far exceed natural recharge. Groundwater in India is a vital resource crucial for sustaining various sectors, particularly agriculture. India heavily relies on groundwater to bridge seasonal gaps in surface water availability, especially during dry spells and erratic monsoon seasons. Relying significantly on groundwater, it contributes approximately 62% to irrigation, about 85% to rural water supply, and roughly 45% to urban water consumption.

**Objective of the Project -**

Investigate the relationship between rainfall patterns and groundwater level fluctuations, particularly in karst regions, to understand how rainfall contributes to changes in groundwater availability. The goal of this project is to develop a predictive system that integrates machine learning (ML) techniques with real-time data inputs to forecast groundwater availability in response to changing environmental and usage patterns. Specifically, the project will focus on leveraging advanced ML algorithms, including Random Forest, Gradient Boosting Regressor, and Linear Regression, to predict fluctuations in groundwater levels based on variables such as rainfall, groundwater. exploitation, population growth, and climate conditions. These machine learning models will be trained using historical data from the region, providing insights into the relationships between different variables and their collective influence on groundwater levels.
One of the key innovations of this project is the use of Explainable AI (XAI) methods, such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations), which will be integrated into the system to make the machine learning models more interpretable and transparent. These techniques will help stakeholders—such as policymakers, water resource managers, and local communities—understand the factors driving groundwater availability and make informed decisions based on the model's predictions. Explainability is a crucial feature, as it allows users to trust the system and gain insights into how specific factors, such as rainfall patterns or water usage, influence future groundwater levels.

## 4. Literature

- Several studies have examined the impact of rainfall on groundwater levels in India, particularly in the context of population growth, water usage, and climate conditions. Research indicates that excessive groundwater extraction for agriculture and domestic use, driven by a growing population, significantly limits the capacity of rainfall to replenish groundwater reserves. Studies such as those by Rodell et al.

(2009) and Singh et al. (2010) highlight the severe groundwater depletion in northern India and urban areas like Delhi, where natural recharge from rainfall is insufficient due to unsustainable extraction rates. Climate change further exacerbates this issue by altering monsoon patterns, leading to erratic rainfall that often increases surface runoff rather than groundwater recharge, as observed by Shah (2014). Regional studies, such as those by Pandey et al. (2011), emphasize that groundwater dynamics vary across India, with population density and agricultural practices heavily influencing groundwater levels. Solutions like rainwater harvesting and artificial recharge, as explored by Agarwal et al. (2015), offer potential mitigation strategies to stabilize groundwater levels, especially in areas experiencing significant water stress.

- Impact of Intensive Agriculture on Groundwater Depletion
  Tiwari et al. (2011) focused on how intensive agricultural practices in northern India, particularly in states like Punjab and Haryana, affect groundwater levels. With high water demands for irrigation, particularly for water-intensive crops like rice and wheat, groundwater levels have been rapidly declining despite seasonal rainfall. Their study emphasizes that over-reliance on groundwater, paired with inefficient irrigation methods, has drastically reduced the aquifer recharge capacity, particularly during years with below-average rain.

# 5. Chapters

## 5.1 About –

- India is an enormous country with very diverse lifestyles. India is a country of many famous rivers flowing across the country and the world's highest rainfall zone is in India. Though, due to unplanned usage and wastage of water, many of the states in India is f facing a severe shortage of water during summer.

- This dataset contains 689 rows for each district in India and 16 columns for different statistical data related to water extraction and recharge. The dataset contains data for the year 2017.
- Inspiration- In recent times, India is going through a severe water crisis in many of the regions during the summer. Proper planning and forecasting can save wastage of water. Distribution of water and regulating the extraction of groundwater can be done more efficiently by analyzing the dataset.
- The dataset is provided by https://data.gov.in/. The dataset is open for analysis and research.
- Groundwater reserves are crucial to global agriculture but are being depleted at rates far in excess of recharge. This depletion in groundwater poses a direct risk to the livelihood and food security of several hundred millions of people. Few nations face more acute challenges than India, where irrigation-dependent agriculture depends on groundwater resources that have been steeply declining. Understanding the drivers of depletion is clearly necessary, but inherently difficult: decline is the result of complex interactions between natural and human variables and policy inputs, and exhibits temporal and spatial heterogeneity in the timing and incidence of groundwater depletion

## 5.2 Existing System -

The existing systems for managing and predicting groundwater availability primarily rely on traditional hydrological models, manual data collection, and statistical approaches. These systems typically analyze factors like rainfall, groundwater levels, and usage patterns through basic linear models or empirical relationships. However, many of these systems are limited in their ability to account for the complex, non-linear interactions between multiple factors such as fluctuating rainfall patterns, groundwater exploitation, and climate variability. Traditional methods also often lack the capability to provide real-time predictions or the flexibility to incorporate large datasets, making them less effective for dynamic and rapidly changing conditions.

In some regions, groundwater models have been developed to simulate the flow and storage of groundwater, using techniques like numerical modeling or hydrological simulations. While these models can provide valuable insights, they often require extensive data and computational resources, and they may not account for all variables influencing groundwater levels, such as population growth or local climate shifts. Moreover, these systems usually operate in isolation from real-time data inputs, making them less adaptable to immediate changes in water usage or rainfall patterns.

### 5.3 Proposed system -

The proposed system aims to overcome the limitations of existing groundwater prediction models by integrating advanced machine learning algorithms with Explainable AI techniques, creating a more accurate, adaptable, and transparent solution for groundwater management. The system will utilize data such as rainfall, groundwater levels, water usage, and climate conditions to forecast future groundwater availability. Machine learning models like Random Forest, Gradient Boost Regressor, and Linear Regression will be used to predict groundwater levels based on these input factors. Additionally, the system will incorporate Explainable AI methods such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) to provide transparency in the predictions and to help stakeholders understand the factors driving the model's outputs.

## 5.4 Data Collection –

- The Jupyter notebook titled "Effects of Rainfall in India" undertakes an exploratory analysis focusing on the impact of rainfall and groundwater extraction across various states in India. It appears to combine data related to groundwater resources, rainfall patterns, and possibly other environmental metrics like water availability and usage. The primary goal of this analysis is to investigate the relationship between these factors and the distribution of water resources across the country, particularly in the context of groundwater extraction, which is a critical issue in India due to the country's reliance on this resource for agricultural, domestic, and industrial use.

- Structure – The data in the notebook is being used to focus on multiple states and their groundwater extraction levels. Dataset is from government and public databases related to water resources, rainfall statistics, and state-wise environmental data. Groundwater extraction rates are represented as percentages, and the data includes the names of the states being analyzed. The data is relatively comprehensive, containing data for at least 36 Indian states and union territories.
- Libraries being used in the project in jupyter notebook (as given in Figure 5.2.1) –
    - **Pandas**: This library is primarily used for data manipulation and analysis. It allows for loading, cleaning, and organizing the dataset, making it easy to work with structured data (e.g., data in tables). Pandas provides functionality to filter, group, and summarize the data, which is crucial for analyzing the groundwater extraction across different states.
    - **Matplotlib**: Matplotlib is used for data visualization. In this notebook, it helps to create histograms and other plots to visually represent the distribution of groundwater extraction percentages across various states. It enables the generation of a grid of histograms, offering a state-wise comparison of data in a clear, visual format. Seaborn is built on top of Matplotlib and provides a high-level interface for drawing attractive and informative statistical graphics.
    - **NumPy** (most likely used): Often used alongside Pandas, NumPy provides support for numerical operations, especially for handling arrays and matrices. Although it isn't explicitly mentioned, NumPy usually plays a role behind the scenes in mathematical computations and data handling within Pandas and Matplotlib.

```
[4]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns

[5]: pip install seaborn --upgrade

     Requirement already satisfied: seaborn in c:\users\saihg\anaconda3\lib\site-packages (0.13.2)
     Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\saihg\anaconda3\lib\site-packages (from seaborn) (1.26.4)
     Requirement already satisfied: pandas>=1.2 in c:\users\saihg\anaconda3\lib\site-packages (from seaborn) (2.1.4)
     Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\saihg\anaconda3\lib\site-packages (from seaborn) (3.8.0)
     Requirement already satisfied: contourpy>=1.0.1 in c:\users\saihg\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.2.0)
     Requirement already satisfied: cycler>=0.10 in c:\users\saihg\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.11.0)
     Requirement already satisfied: fonttools>=4.22.0 in c:\users\saihg\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.25.0)
     Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\saihg\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.4)
     Requirement already satisfied: packaging>=20.0 in c:\users\saihg\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (23.1)
     Requirement already satisfied: pillow>=6.2.0 in c:\users\saihg\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (10.2.0)
     Requirement already satisfied: pyparsing>=2.3.1 in c:\users\saihg\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.0.9)
     Requirement already satisfied: python-dateutil>=2.7 in c:\users\saihg\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.8.2)
     Requirement already satisfied: pytz>=2020.1 in c:\users\saihg\anaconda3\lib\site-packages (from pandas>=1.2->seaborn) (2023.3.post1)
     Requirement already satisfied: tzdata>=2022.1 in c:\users\saihg\anaconda3\lib\site-packages (from pandas>=1.2->seaborn) (2023.3)
     Requirement already satisfied: six>=1.5 in c:\users\saihg\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.1
     6.0)
     Note: you may need to restart the kernel to use updated packages.
```

*Figure 5.2.2 Libraries installed*

## 5.5 Data Preprocessing –

### Data Loading and Importing

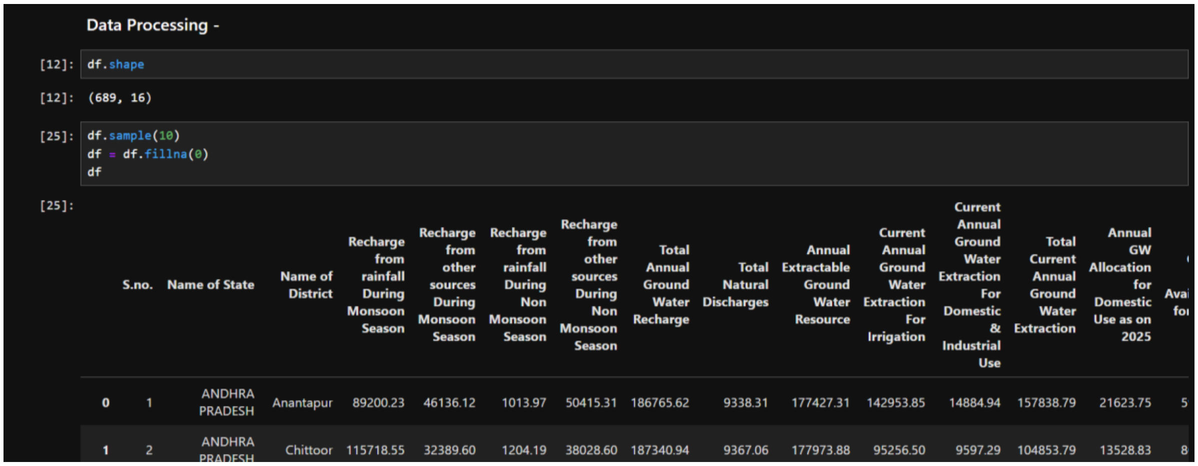
- **Purpose**: This step involves reading the dataset into the project, usually in a structured format such as CSV or Excel.
- **Implementation**: The dataset is loaded into a Pandas DataFrame, which allows for easier manipulation and analysis.
- **Visual Representation**: Raw dataset or the result of df.head() would illustrate with showing a table with columns like "State Name," "Ground Water Extraction," and other variables being analyzed.

### Handling Missing Values

- **Purpose**: Missing or incomplete data can distort the analysis, so this step ensures data integrity by either removing or filling missing values (as done in figure 5.3.1 and figure 5.3.2).
- **Implementation**: Depending on the missing values' nature, rows with missing values could be dropped, or values might be filled using methods such as mean, median, or mode. We use df = df.fillna(0), df, df.isna().sum() which tells us that how

many Null values are present in this notebook, According to that we use df.fillna(0) which fills up all the na values with 0.

- **Visual Representation**:

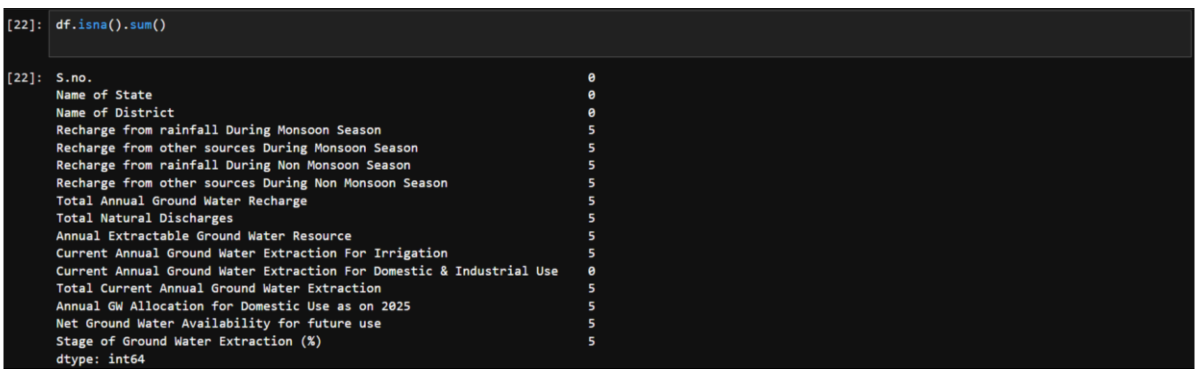


Data Processing -

```
[12]: df.shape
[12]: (689, 16)
[25]: df.sample(10)
      df = df.fillna(0)
      df
[25]:
```

| | S.no. | Name of State | Name of District | Recharge from rainfall During Monsoon Season | Recharge from other sources During Monsoon Season | Recharge from rainfall During Non Monsoon Season | Recharge from other sources During Non Monsoon Season | Total Annual Ground Water Recharge | Total Natural Discharges | Annual Extractable Ground Water Resource | Current Annual Ground Water Extraction For Irrigation | Current Annual Ground Water Extraction For Domestic & Industrial Use | Total Current Annual Ground Water Extraction | Annual GW Allocation for Domestic Use as on 2025 | Avai for |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | ANDHRA PRADESH | Anantapur | 89200.23 | 46136.12 | 1013.97 | 50415.31 | 186765.62 | 9338.31 | 177427.31 | 142953.85 | 14884.94 | 157838.79 | 21623.75 | 5 |
| 1 | 2 | ANDHRA PRADESH | Chittoor | 115718.55 | 32389.60 | 1204.19 | 38028.60 | 187340.94 | 9367.06 | 177973.88 | 95256.50 | 9597.29 | 104853.79 | 13528.83 | 8 |

*Figure 5.3.1 Initial processing of Data*



```
[22]: df.isna().sum()
[22]: S.no.                                                          0
      Name of State                                                  0
      Name of District                                               0
      Recharge from rainfall During Monsoon Season                   5
      Recharge from other sources During Monsoon Season              5
      Recharge from rainfall During Non Monsoon Season               5
      Recharge from other sources During Non Monsoon Season          5
      Total Annual Ground Water Recharge                             5
      Total Natural Discharges                                       5
      Annual Extractable Ground Water Resource                       5
      Current Annual Ground Water Extraction For Irrigation          5
      Current Annual Ground Water Extraction For Domestic & Industrial Use   0
      Total Current Annual Ground Water Extraction                   5
      Annual GW Allocation for Domestic Use as on 2025               5
      Net Ground Water Availability for future use                   5
      Stage of Ground Water Extraction (%)                           5
      dtype: int64
```

*Figure 5.3.2 Find Null Values*

## Data Cleaning

- **Purpose**: Cleaning the data ensures that it's consistent, accurate, and free from errors or redundancies. It can include correcting data types, removing duplicates, or formatting the columns appropriately.
- **Implementation**: The notebook might involve renaming columns, converting data types (e.g., turning strings into numerical data), or removing irrelevant information. Drop rows where 'State' is missing - df = df.dropna(subset=['State']).

**Filtering Data by Criteria**

- **Purpose**: This step involves focusing on specific data that is relevant to the analysis. In this case, filtering the dataset to focus on the top 36 states in terms of groundwater extraction data.
- **Implementation**: The filtering happens using Pandas by selecting only the required states or regions and focusing on their corresponding data. For instance, df['Name of State'].value_counts() might be used to sort the states by the number of data points (as done in figure 5.3.3).
- **Visual Representation**:

```
[27]: #statewise groundwater reserve
state_list = []
total_frnd_water_recharge = []
curr_gw_extr_list = []
future_available_GW_list = []
#Net Ground Water Availability for future use

for state, subset in df.groupby('Name of State'):
    #print(state, sum(subset['Net Ground Water Availability for future use']))
    state_list.append(state)
    total_frnd_water_recharge.append(sum(subset['Total Annual Ground Water Recharge']))
    curr_gw_extr_list.append(sum(subset['Total Current Annual Ground Water Extraction']))
    future_available_GW_list.append(sum(subset['Net Ground Water Availability for future use']))
dfnew = pd.DataFrame({"State":state_list, "GW_Recharge":total_frnd_water_recharge, "GW_Extraction": curr_gw_extr_list, "Future_GW_Available": future_avail
```

*Figure 5.3.3 Creating DataFrame for Statewise groundwater reserve*

## 5.6 Exploratory Data Analysis (Citation -*"Causes and implications of groundwater depletion in India: A review")* –

### 1. Correlation Heatmap of Numeric Data:

**Purpose**: This code generates a heatmap showing the correlation between the numeric columns in the dataset (Given in figure 5.4.1).

**Process**:

- df.select_dtypes(include=['number']): Selects only numeric columns from the dataset.

- The heatmap visualizes how each numeric variable relates to others in terms of correlation, with annotations showing the exact correlation coefficients.

**Visualization**: A color-coded map where closer to 1 means a strong positive correlation, and closer to -1 indicates a negative correlation.

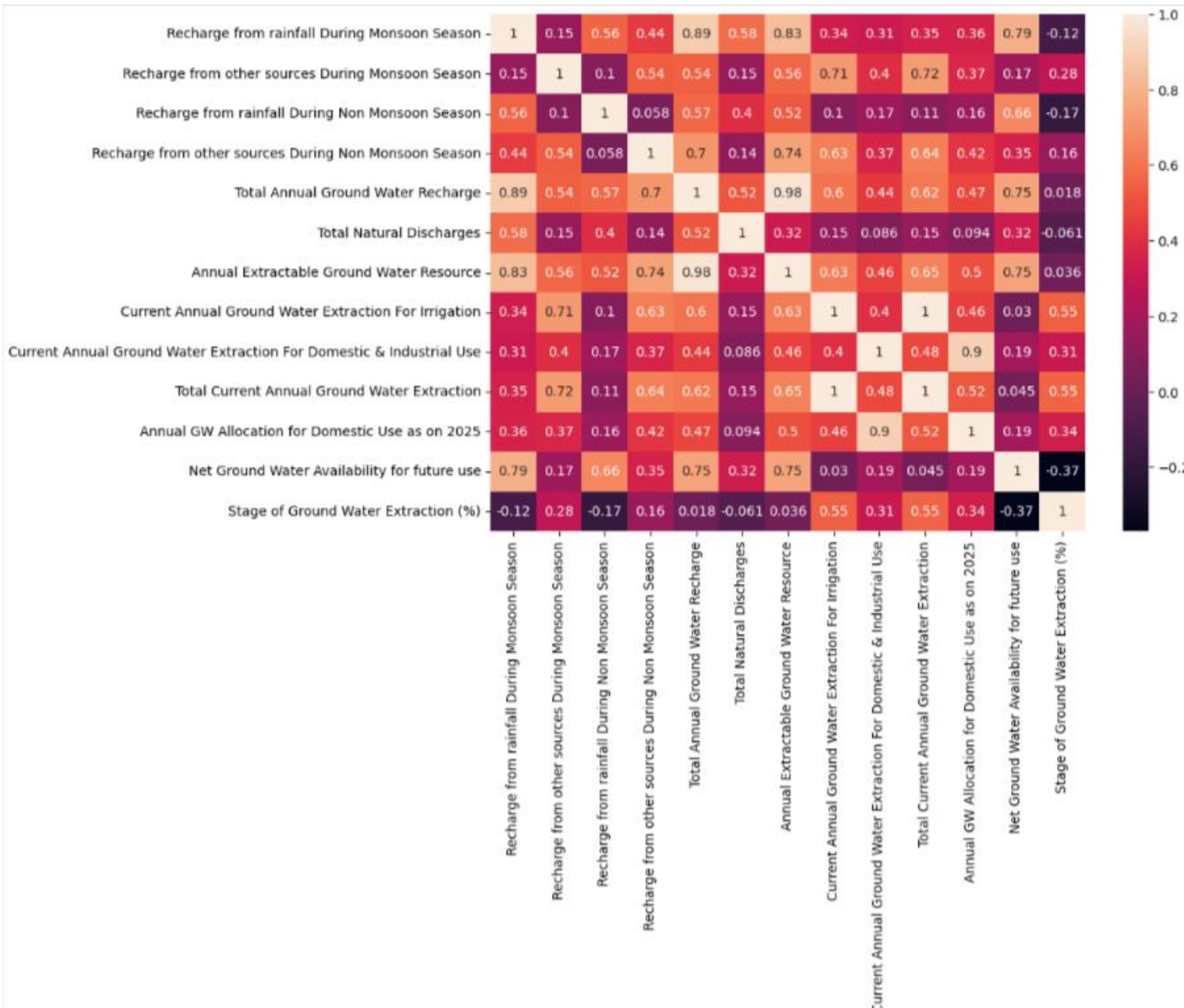



*Figure 5.4.1 Heatmap showing the correlation matrix of groundwater recharge, extraction and availability*

2. **Bar Plot of Annual Groundwater Reserve by State**:

**Purpose**: Displays the annual groundwater reserve (difference between recharge and extraction) by state (as given in figure 5.4.2).

**Process**:

a) The code calculates `annual_reserve` by subtracting groundwater extraction from recharge.
b) A horizontal bar plot is used, where green bars represent positive reserves, and red bars indicate negative reserves (states where more groundwater is extracted than recharged).

**Visualization**: A bar graph with different states, showing whether they have positive(sustainable) or negative (unsustainable) groundwater reserves.

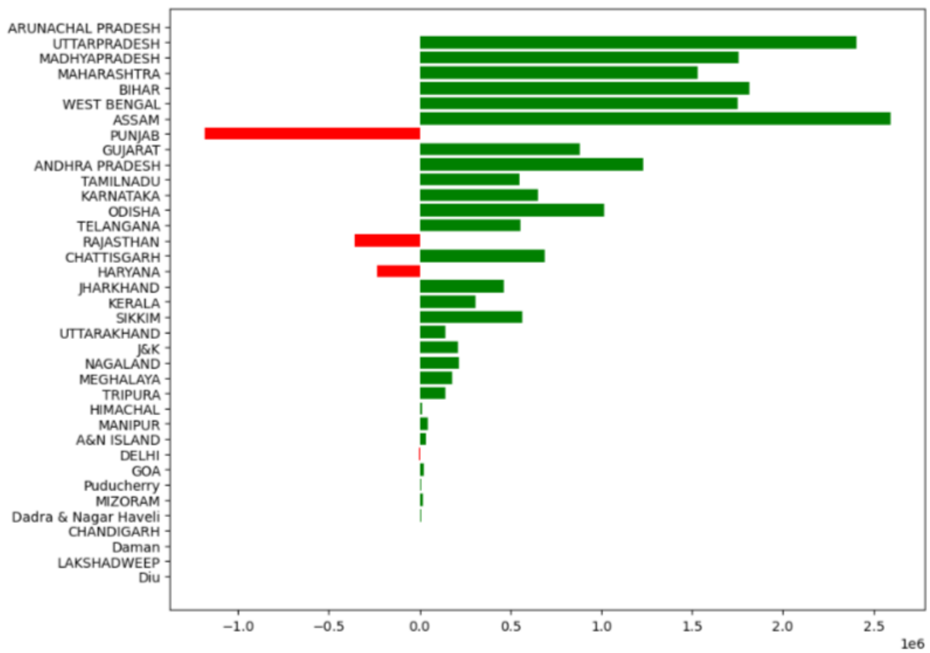



*Figure 5.4.2 Bar chart showing the state-wise distribution of groundwater recharge or extraction*

3. **Groundwater Extraction vs Availability**:

**Purpose**: This compares the available groundwater recharge with groundwater extraction by state (as given in figure 5.4.3).
**Process**:

- The blue bars represent available groundwater (GW_Recharge), while the red bars represent the amount extracted (GW_Extraction).

- The plot shows a side-by-side comparison for each state, helping to highlight which states are in a critical zone (where extraction equals or exceeds recharge).

**Visualization**: States with higher extraction than recharge are likely in critical zones for groundwater sustainability.

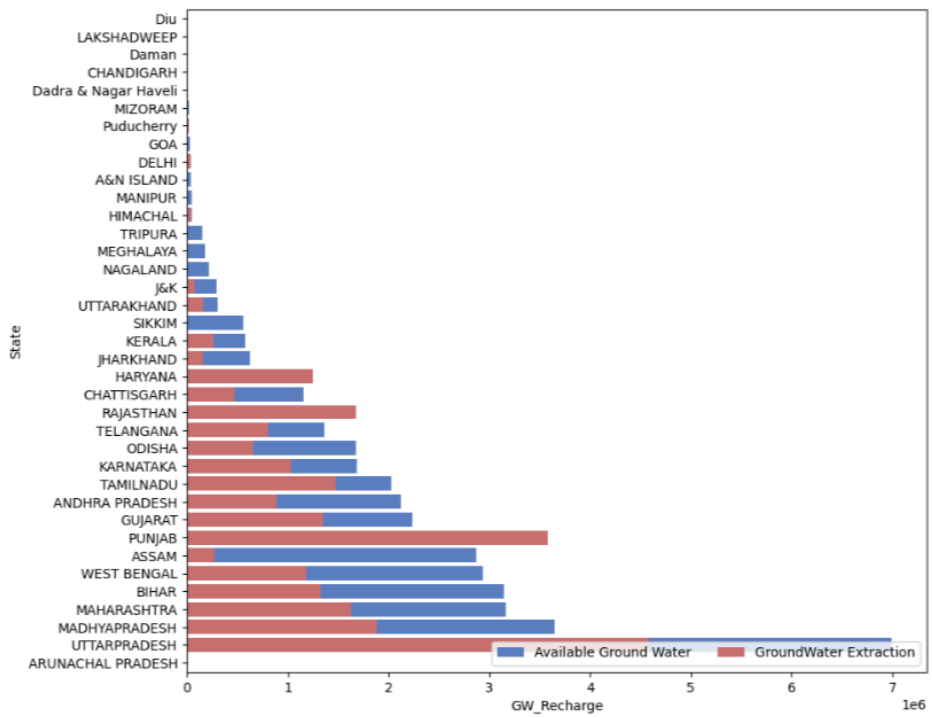



*Figure 5.4.3 Groundwater Availability Vs Extraction*

4. **Future Groundwater Availability:**

**Purpose**: Visualizes the future groundwater availability by state (as given in figure 5.4.4).

**Process**:

- The data is sorted by Future_GW_Available, and a bar plot is created to display the future reserves for each state.

**Visualization**: The graph shows which states are projected to have higher future groundwater reserves, with missing data for certain states like Arunachal Pradesh. States like Assam, Uttar Pradesh, and Bihar have higher future reserves compared to others.

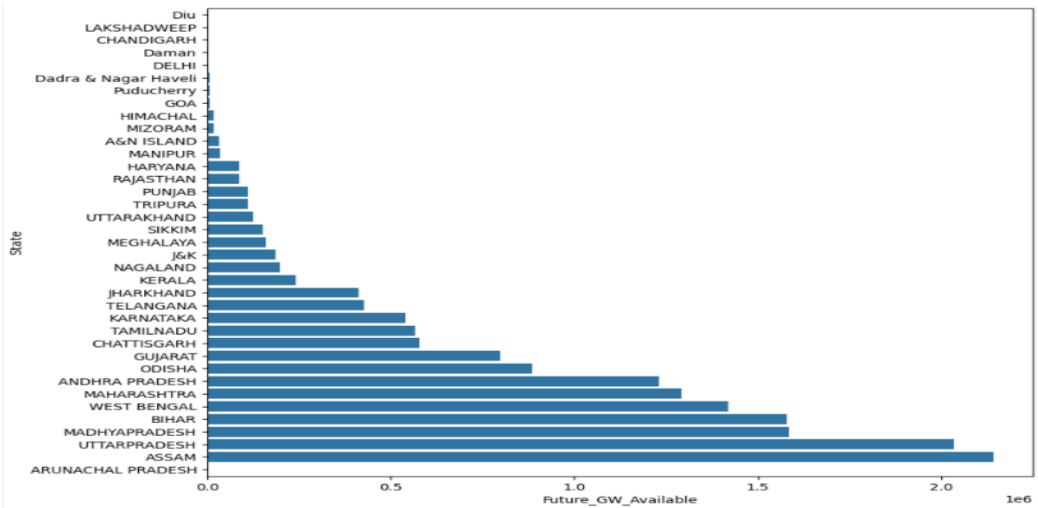



*Figure 5.4.4 Future Availability of Groundwater*

## 5.7 Case Study: Scenario in Punjab –

**Code -** df_pun = df[df['Name of State']=='PUNJAB'], df_pun.head(),

**Purpose**: This filters the dataset to show only the data for Punjab by selecting rows where the state is 'PUNJAB'.

**Output**: Displays the first few rows of data specific to Punjab.

- **Bar Plot: Groundwater Allocation in Punjab's Districts:**

  **Purpose**: Visualizes the annual groundwater allocation for domestic use by district in Punjab, projected for 2025.

  **Process**:

  The sns.barplot creates a bar chart that shows the top districts in terms of groundwater allocation for domestic use (as given in figure 5.5.1).

  **Visualization**:

  Ludhiana, Jalandhar, and Amritsar are the top three districts in terms of groundwater allocation, indicating that these areas are projected to consume the highest amounts of groundwater.

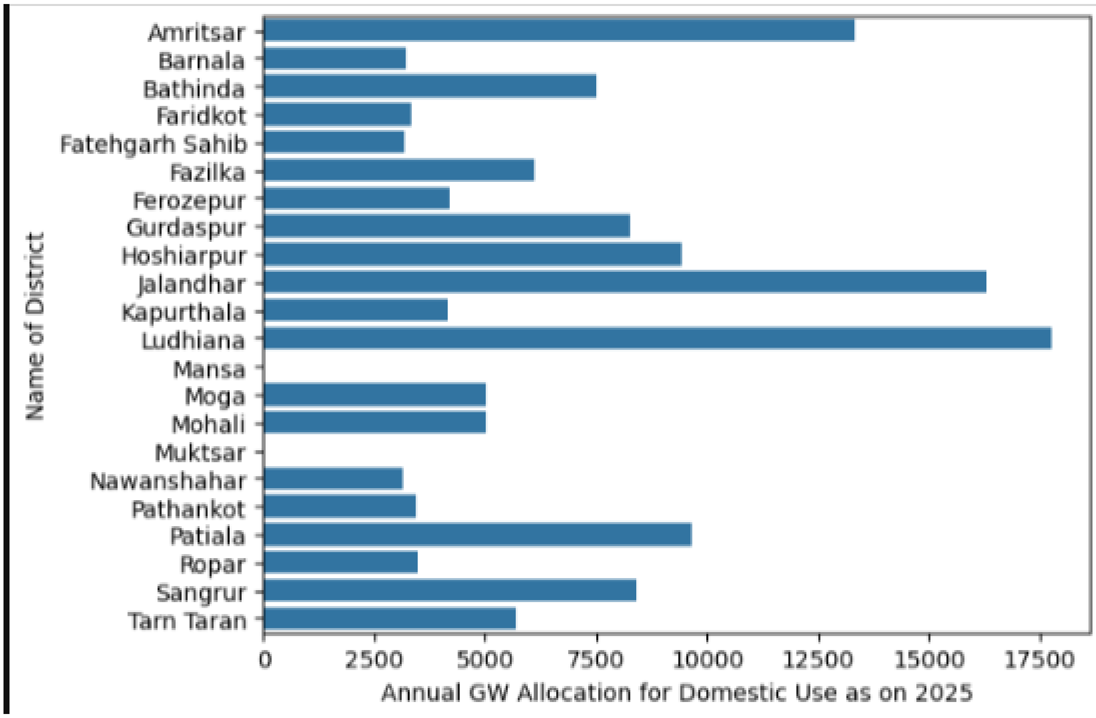



*Figure 5.5.1 Groundwater in Punjab's Districts*

- **Top Districts in Groundwater Extraction in Punjab**:

  **Code**- df_pun.sort_values('Stage of Ground Water Extraction (%)', ascending=False, inplace=True), df_pun[['Name of State', 'Name of District', 'Stage of Ground Water Extraction (%)']].head(5)

  **Purpose**: Sorts the districts in Punjab by the percentage of groundwater extraction and displays the top five(as given in table 5.5.1).

  **Process**:

  The code sorts the Punjab data based on the "Stage of Ground Water Extraction (%)", indicating how much groundwater is extracted compared to what is available.

  **Output**: Shows the top five districts with the highest percentage of groundwater extraction. These districts are at the most risk of over-extraction and depletion.

| | Name of State | Name of District | Stage of Ground Water Extraction (%) |
|---|---|---|---|
| 464 | PUNJAB | Sangrur | 260.00 |
| 453 | PUNJAB | Jalandhar | 239.16 |
| 457 | PUNJAB | Moga | 229.47 |
| 454 | PUNJAB | Kapurthala | 223.75 |
| 462 | PUNJAB | Patiala | 216.82 |

*Table 5.5.1 Top Districts in Groundwater Extraction in Punjab*

**Conclusion –**

- Punjab is facing a critical water challenge, particularly in districts like Sangrur, Jalandhar, and Moga. These regions not only have the highest projected groundwater allocation for domestic use in 2025 but are also at the highest risk of groundwater depletion due to over-extraction. The high "Stage of Ground Water Extraction" percentage in these districts suggests that groundwater is being extracted faster than it can be replenished. If groundwater extraction continues at this rate without proper management, Punjab, particularly its northern districts, could face severe water shortages in the near future

## 5.8 State-wise Ground Water Recharge and Extraction Analysis –

- Code - plt.figure(figsize=(10,12)), sns.barplot(x='Recharge from rainfall During Monsoon Season', y='Name of State', data=df), plt.show()
- This code generates a bar plot using Seaborn (sns.barplot). It shows the **state-wise groundwater recharge** during the monsoon season. Here's how it works:
  **x-axis**: Represents the recharge from rainfall.
  **y-axis**: Lists the names of Indian states.
  The bar lengths correspond to the amount of groundwater recharge due to rainfall in each state (as given in figure 5.5.2).

- **Visualization Insight**:

  The graph allows us to easily compare how much groundwater different states receive during the monsoon, highlighting regional differences in water replenishment from rainfall.

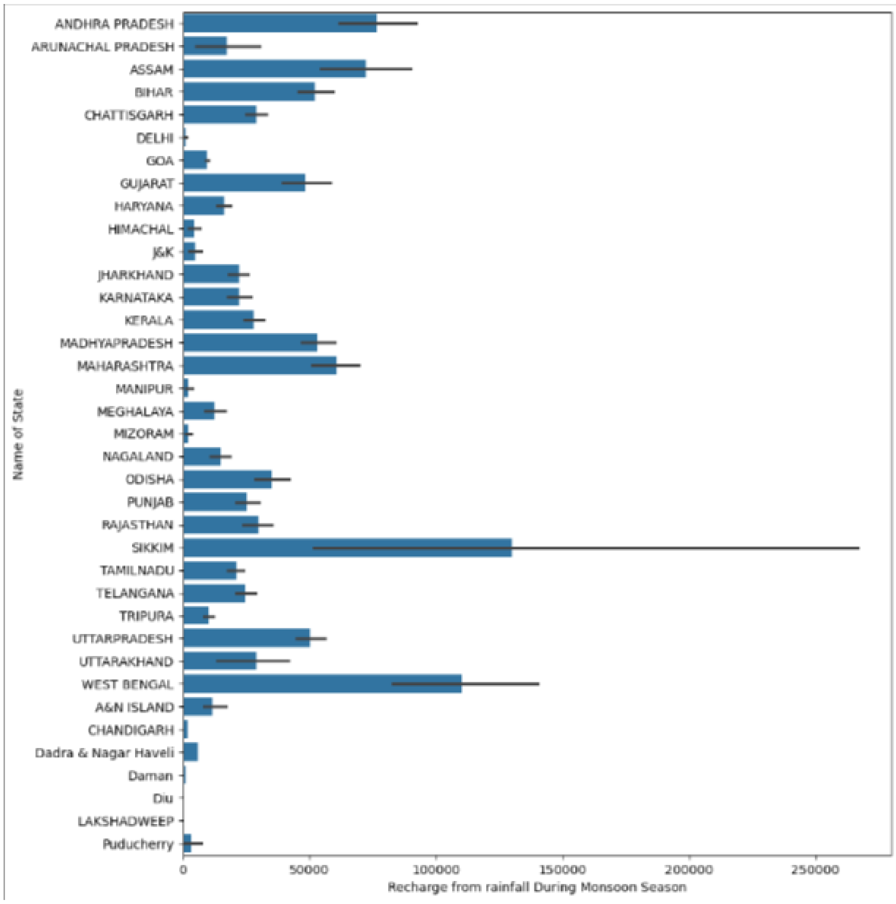



*Figure 5.5.2 Recharge of Water in Monsoon Season*

```
fig,axs = plt.subplots(6,6,figsize=(12,8))

for idx,state in enumerate(df['Name of State'].value_counts().sort_values(ascending=False)[0:36].index):
    print(idx,state)
    axs[idx//6,idx%6].hist(x = df[df['Name of State']==state]['Stage of Ground Water Extraction (%)'], color='b')
    axs[idx//6,idx%6].set_title(state)
plt.suptitle("State wise GW Extraction distribution")
plt.tight_layout()
fig.subplots_adjust(top=0.88)
plt.show()
```

- This code creates a 6x6 grid of histograms to visualize the **distribution of groundwater extraction** across 36 states (as given in figure 5.5.3).
  Each subplot represents a state's groundwater extraction percentages.
  Histograms show how much groundwater has been extracted as a percentage of the total available.
- **Visualization Insight**:
  This provides an understanding of how different states utilize their groundwater resources, highlighting states with potential over-extraction risks

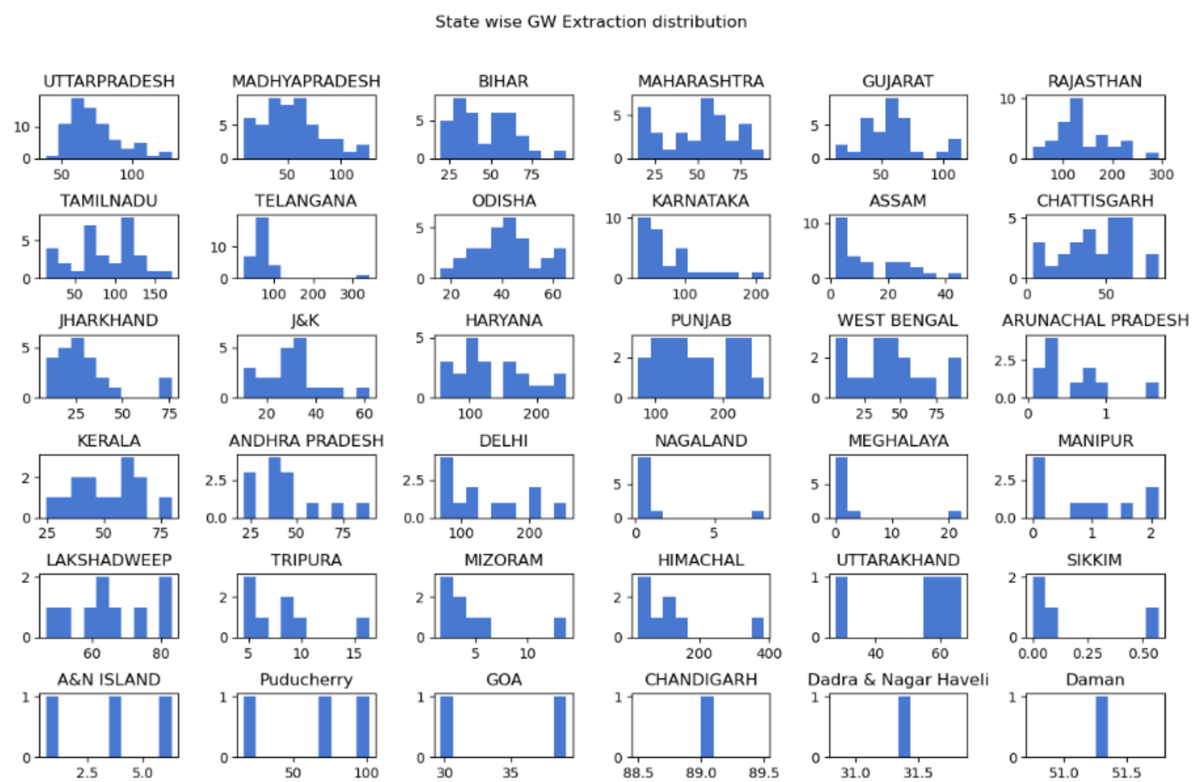



*Figure 5.5.3 Statewise Groundwater Extraction Distribution*

## 5.9 Algorithms–

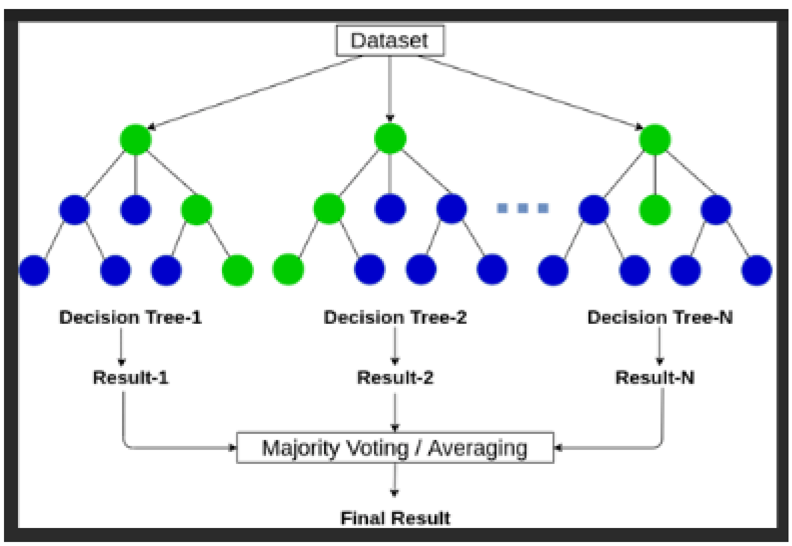
### 5.9.1 Random Forest

## Definition:

Random Forest is an ensemble learning method that combines multiple decision trees to improve classification accuracy and reduce overfitting. Each tree in the forest is trained on a random subset of the data and features, making the model robust and effective for large datasets.

**Internal Working:**

1. Bootstrap Sampling: Random Forest uses bootstrap sampling to generate different subsets of the training data. Each tree is trained on a unique subset.
2. Random Feature Selection: For each tree, only a random subset of features is considered when splitting nodes, reducing correlation between trees and increasing model diversity.
3. Decision Trees: Each tree independently produces a classification result. Trees are typically grown without pruning, meaning they go as deep as possible.
4. Ensemble Voting: For classification, the forest aggregates the predictions from each tree and selects the class with the majority vote.




## 5.9.2 Linear Regression –

**Linear Regression** is one of the simplest and most widely used algorithms in machine learning for predictive modeling. It is a statistical method used to understand the relationship between one dependent variable (target) and one or more independent variables (predictors). The goal of linear regression is to model the relationship between these variables by fitting a linear equation to the observed data.

**Internal Working**:

In **Simple Linear Regression**, there is a single independent variable (XXX) and a dependent variable (YYY):

Y=β0+β1X+ϵY = \beta_0 + \beta_1 X + \epsilonY=β0+β1X+ϵ

Where:

- YYY is the dependent variable (the target you want to predict).
- XXX is the independent variable (the feature or input).
- β0\beta_0β0 is the **intercept** of the regression line.
- β1\beta_1β1 is the **slope** of the regression line, representing how much YYY changes with a one-unit change in XXX.

ϵ\epsilonϵ is the **error term** (random noise or residuals), which accounts for the difference between the predicted value and the actual value.

### 5.9.3 Gradient Boosting Regressor–

**Gradient Boosting Regressor (GBR)** is an ensemble machine learning technique used for both regression and classification tasks. It is based on the concept of boosting, where weak learners (typically decision trees) are trained sequentially, with each new model focusing on the errors made by the previous models. The goal of Gradient Boosting is to create a strong model by combining multiple weak models in a way that improves predictive accuracy. The first step in GBR is to initialize the model with a base prediction. For regression, this could simply be the mean of the target values in the dataset. Let’s say the initial prediction for each data point is:

f0(x)=mean(y)f_0(x) = \text{mean}(y)f0(x)=mean(y)

Where yyy is the target variable (groundwater level, for example).

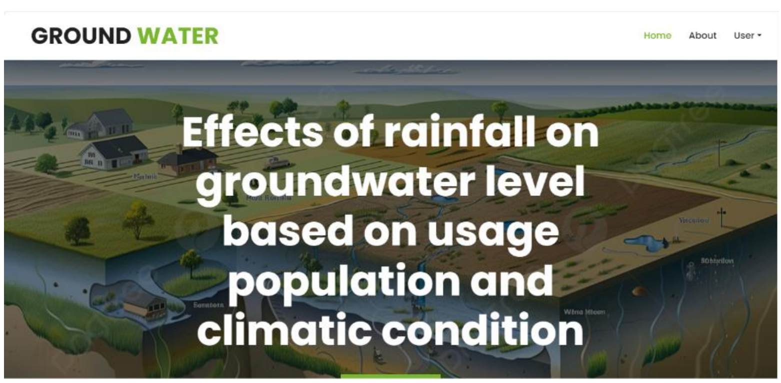
In regression tasks, Gradient Boosting is used to predict continuous values. The key idea is to reduce errors by iteratively correcting the mistakes made by prior models using gradient-based optimization.

### 5.9.4 Machine Learning and Deep Learning Models-

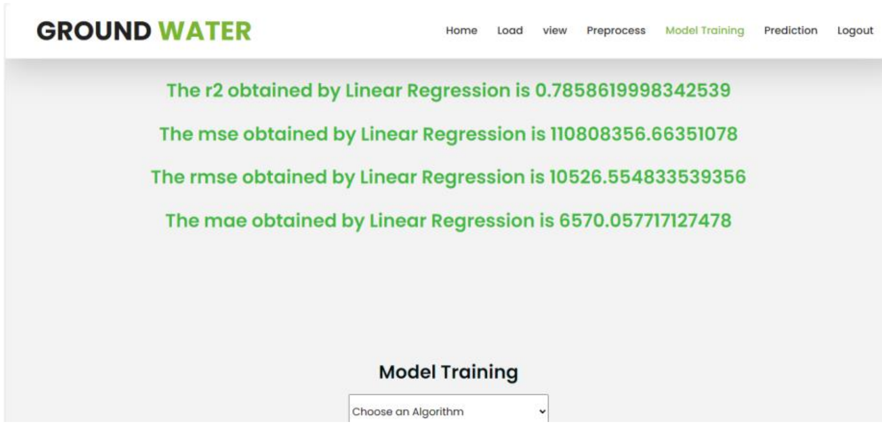
- Ensemble Learning: This module employs ensemble models (e.g., Random Forest, Gradient Boosting) to detect patterns that resemble cyber terrorist activity based on historical data.
- Deep Learning Models: Neural networks, such as CNNs and RNNs, process high-dimensional data (e.g., sequences and logs) for pattern recognition. Additionally, a hybrid model could integrate BiLSTM with CNN layers to capture spatial and sequential data correlations.
- X-AI (Explainable AI) Framework: The X-AI layer integrates interpretability models (e.g., LIME, SHAP) to explain model predictions, helping analysts understand and trust the AI system's decision-making process.

## 5.10 Output –

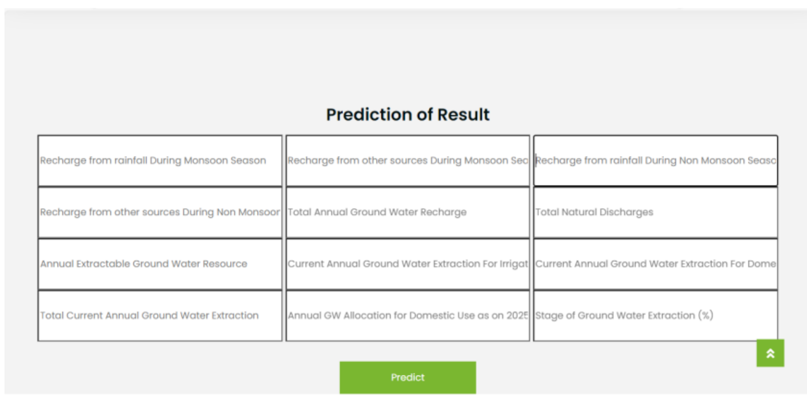
**Home Page:** The Home Page serves as the landing page of your application. It provides an overview of the project's features, objectives, and benefits. Users can navigate to other sections of the application from this page.




**Algorithms:** User can select the algorithms




**Prediction Page: :** The Prediction Page allows users to input data and receive predictions based on the trained machine learning models. This page typically includes a form or interface for uploading or entering data




## 6. Summary & Conclusions

This project aims to develop a predictive system for groundwater level forecasting, integrating rainfall, groundwater usage, and climate conditions using advanced machine learning techniques. By leveraging algorithms such as Random Forest, Gradient Boosting Regressor, and Linear Regression, the project models the complex interactions between environmental factors and groundwater availability, providing a more accurate and adaptive approach compared to traditional methods the project ensures that predictions remain up-to-date and reflect current conditions, providing timely insights for sustainable water management. This real-time adaptability will be particularly valuable in regions experiencing rapid changes in rainfall patterns, population growth, and groundwater extraction.

The system's ultimate goal is to assist policymakers, water resource managers, and local communities in effectively managing groundwater resources. By predicting future groundwater availability with higher accuracy and offering explainable insights into the underlying factors, this tool will enable better decision-making for water conservation, extraction management, and long-term sustainability.

In conclusion, the successful implementation of this predictive system will not only improve the management of groundwater resources but also contribute to more sustainable water practices, ensuring the long-term availability of this critical resource in the face of climate change, population growth, and over-exploitation.

**Key Observations**:

- Some states exhibit significantly higher levels of groundwater extraction, likely due to high population densities and intensive agricultural or industrial usage.
- There is a direct correlation between regions with lower rainfall and higher groundwater extraction rates, especially in states where surface water resources are limited.
- State-Wise Analysis: The analysis identifies the states where groundwater extraction exceeds sustainable limits, presenting a clear picture of over-extraction in certain areas. States like Punjab, Haryana, and Rajasthan, known for intensive agricultural activity, show higher groundwater depletion compared to others.
- Climate Impact: The study highlights that states with lower annual rainfall rely more heavily on groundwater resources, leading to over-extraction and subsequent depletion. States with moderate to high rainfall exhibit better groundwater recharge rates, reducing dependency.

The analysis effectively demonstrates that groundwater levels are significantly influenced by rainfall patterns, usage intensity, and population density. States with higher groundwater extraction percentages tend to experience more severe depletion, especially in regions with low rainfall and high population growth.

## 6.1 Future Scope –

- While this project provides a robust foundation for predicting groundwater availability, there are several opportunities for future enhancement to further improve the system's accuracy, scalability, and overall effectiveness. One key area of improvement lies in the incorporation of additional environmental variables. In its current form, the system primarily relies on rainfall, groundwater usage, and climate conditions. However, the inclusion of additional factors such as soil moisture, land use patterns, vegetation cover, and topography could significantly enhance prediction accuracy. These factors influence groundwater recharge and discharge processes, especially in areas with diverse geographical features, thereby providing a more comprehensive understanding of groundwater dynamics.
- Another important enhancement is the exploration of advanced machine learning techniques. While Random Forest, Gradient Boosting, and Linear Regression have proven effective, more complex models such as Deep Learning (e.g., Neural Networks) and Long Short-Term Memory (LSTM) networks could improve the system's ability to capture non-linear relationships and temporal dependencies in groundwater data. These advanced models are particularly well-suited for time-series forecasting and could provide deeper insights into the long-term trends and patterns in groundwater fluctuations, leading to more accurate predictions.
- The integration of satellite data and remote sensing technologies would be another valuable enhancement. By incorporating data from satellites—such as soil moisture levels, vegetation indices, and evapotranspiration rates—the system could provide more accurate, region-specific insights into groundwater recharge and availability. Remote sensing would also enable monitoring of groundwater levels in remote or inaccessible regions, increasing the spatial coverage and precision of the model.

# 7. References


***Book:***

**1.** Hong Wang, Jian En Gao , Meng-jie Zhang , Xing-hua Li , Shao-long Zhang , Li-Zhi Jia, CATENA, vol. 127, P.no. 80-91, University of Chinese Academy of Sciences, Beijing, China, April 2015.

***Journal:***

**1.** Swarup Dangar, Akarsh Asoka, Vimal Mishra, "Causes and implications of groundwater depletion in India: A review", *Journal of Hydrology*, vol. 596, p. 126103, Mar. 2021.

***Web Links:***

**1.** Causes and implications of groundwater depletion in India: A review - ScienceDirect

**2.** Effects of rainfall intensity on groundwater recharge based on simulated rainfall experiments and a groundwater flow model - ScienceDirect

**3.** Potential impact of rainfall variability on groundwater resources: a case study in Uttar Pradesh, India, (PDF) Potential impact of rainfall variability on groundwater resources: a case study in Uttar Pradesh, India (researchgate.net)

**4.** Code to the project Effects of Rainfall github

## 7. References

***Book:***

1. Hong Wang, Jian En Gao , Meng-jie Zhang , Xing-hua Li , Shao-long Zhang , Li-Zhi Jia, CATENA, vol. 127, P.no. 80-91, University of Chinese Academy of Sciences, Beijing, China, April 2015.

***Journal:***

1. Swarup Dangar, Akarsh Asoka, Vimal Mishra, "Causes and implications of groundwater depletion in India: A review", *Journal of Hydrology*, vol. 596, p. 126103, Mar. 2021.

***Web Links:***

1. Causes and implications of groundwater depletion in India: A review - ScienceDirect
2. Effects of rainfall intensity on groundwater recharge based on simulated rainfall experiments and a groundwater flow model - ScienceDirect
3. Potential impact of rainfall variability on groundwater resources: a case study in Uttar Pradesh. India, (PDF) Potential impact of rainfall variability on groundwater resources: a case study in Uttar Pradesh, India (researchgate net)
4. Code to the project Effects of Rainfall github

11/10/24

Signature of Group Member(s)

Signature of Project Advisor