AI ENABLED GROUND WATER WELL PREDICTOR

Theertha Praveen N
TVE24MCA-2056

OVERVIEW

- **Purpose:** Predict optimal groundwater well locations using AI to reduce cost,time, and errors in traditional methods.
- **Data Used:** Geological (soil type, rock formation, aquifer depth, groundwater flow) and climatic (rainfall, temperature, humidity, evaporation) factors.
- **Dataset Source:** Kaggle (cleaned, normalized, and preprocessed).
- **Model:** Random Forest algorithm with 80% training / 20% testing split.
- **Accuracy:** Achieves above high prediction accuracy.
- **Frontend:** HTML, CSS, JavaScript for user input and results display.
- **Backend:** Python (Flask) to process data and run predictions.
- **Output:** Instant suitability prediction.
- **Impact**: Supports farmers, researchers, industries, and government in sustainable water resource management.

EXISTING SYSTEM

- Methods Used – Manual hydrogeological surveys, geophysical methods (e.g., electrical resistivity), GIS mapping, and climatic data analysis.
- Limitations:
    - Time-Consuming – Requires extensive field surveys and manual data collection.
    - High Cost – Involves expensive equipment and expert consultation.
    - Lower Accuracy – Relies on outdated records and ignores underground geological structures.
    - Requires Expertise – Dependent on specialized geologists and trained professionals.

PROPOSED SYSTEM

- High Accuracy – Predicts groundwater availability using Random Forest ML on geological & climatic data.
- Automation – Eliminates costly, time-consuming field surveys through AI-driven analysis.
- Data Sources – Uses only geological (soil type, rock formations, aquifer depth) & climatic (rainfall, temperature, humidity, evaporation) factors.
- Real-Time & Historical Analysis – Continuously learns from new data for updated predictions.
- Scalable & Fast – Processes large geographic regions quickly, suitable for agencies, researchers, and industries.
- User-Friendly Interface – Web/mobile app for instant predictions, accessible to non-experts.
- Sustainable Approach – Supports efficient, environmentally responsible groundwater management.

RANDOM FOREST ALGORITHM

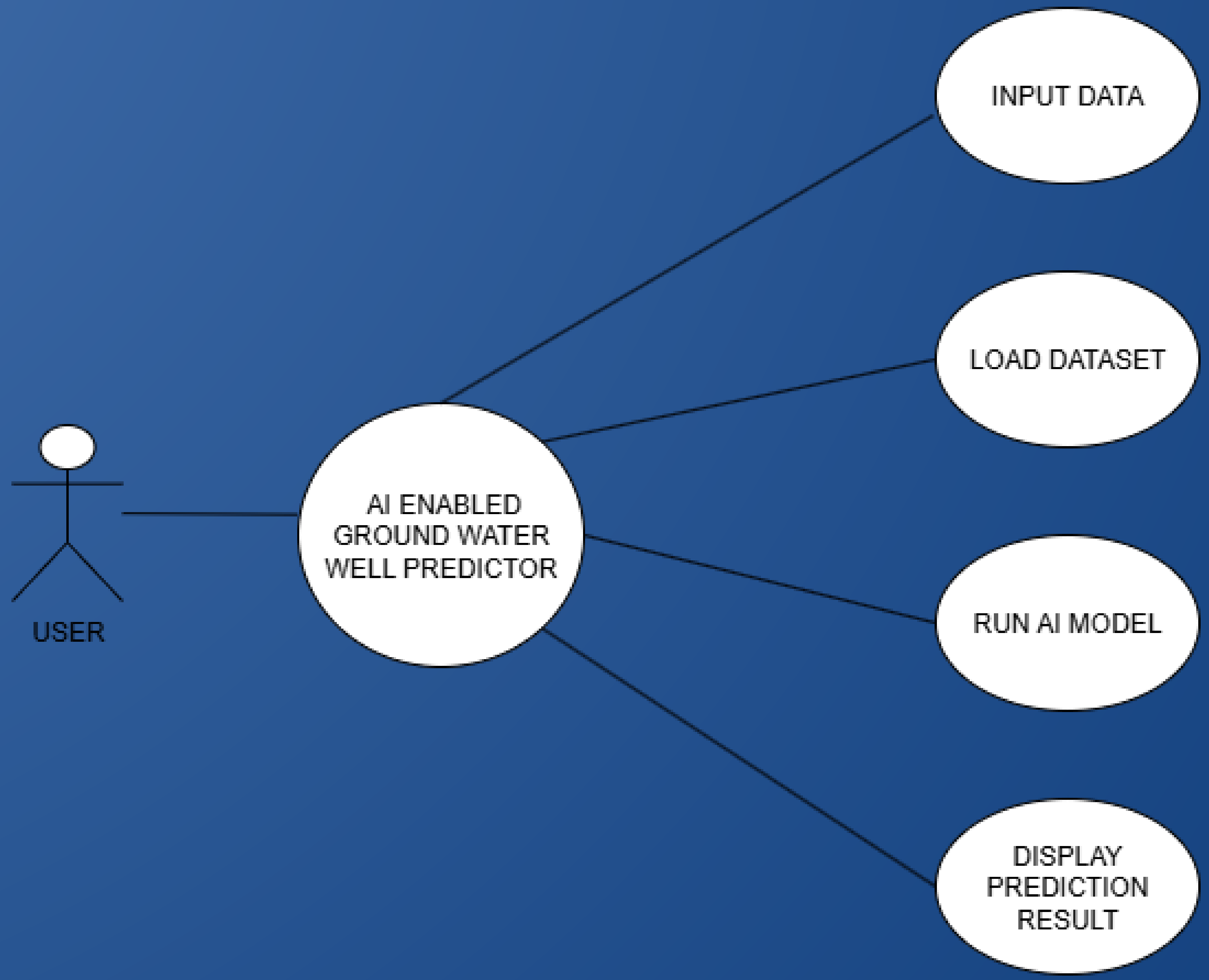
DEFINITION – A Random Forest is an ensemble learning algorithm that builds multiple decision trees and merges their results to improve accuracy and control overfitting.

TYPE – It's used for both classification and regression tasks.

WORKING PRINCIPLE

- Creates many decision trees using bootstrapped samples of the dataset.
- At each split, only a random subset of features is considered, adding diversity among trees.
- Final output is obtained by majority vote (classification) or average prediction (regression).
- High accuracy and robustness
- Handles large datasets well
- Reduces risk of overfitting compared to a single decision tree

UML DIAGRAM



INPUT DATA
LOAD DATASET
USER
AI ENABLED GROUND WATER WELL PREDICTOR
RUN AI MODEL
DISPLAY PREDICTION RESULT

- User inputs geological and climatic data.(Geological - Soil type, rock formations, aquifer depth, groundwater flow patterns .Climatic - Rainfall patterns, temperature, humidity, evaporation rates)

- Data preprocessing and feature selection. (Clean the data remove errors , Normalize values , Select the most important features and here the model is trained using the historical data ie 80% for training and the rest 20%for testing the data.)

- Model predicts well suitability using Random Forest.

- Results displayed in a user-friendly interface. (The system shows whether the location is good for drilling a well.

- Simple Input → Smart Prediction → Clear Output

MODULES

Module 1: Data Foundation & Processing

This is the foundational layer of our system. It is responsible for gathering and preparing the essential geological and climatic data that our AI model relies on. This module sources a comprehensive dataset from Kaggle, then cleans, normalizes, and processes the information to ensure it is accurate, consistent, and ready for analysis.

Module 2: Intelligent Feature Engineering

This module acts as the system's analytical core. It intelligently selects the most critical features from the dataset—such as soil type, rock formations, and rainfall patterns—that have the highest impact on groundwater availability. By focusing on these key indicators, we significantly enhance the model's predictive accuracy and efficiency.

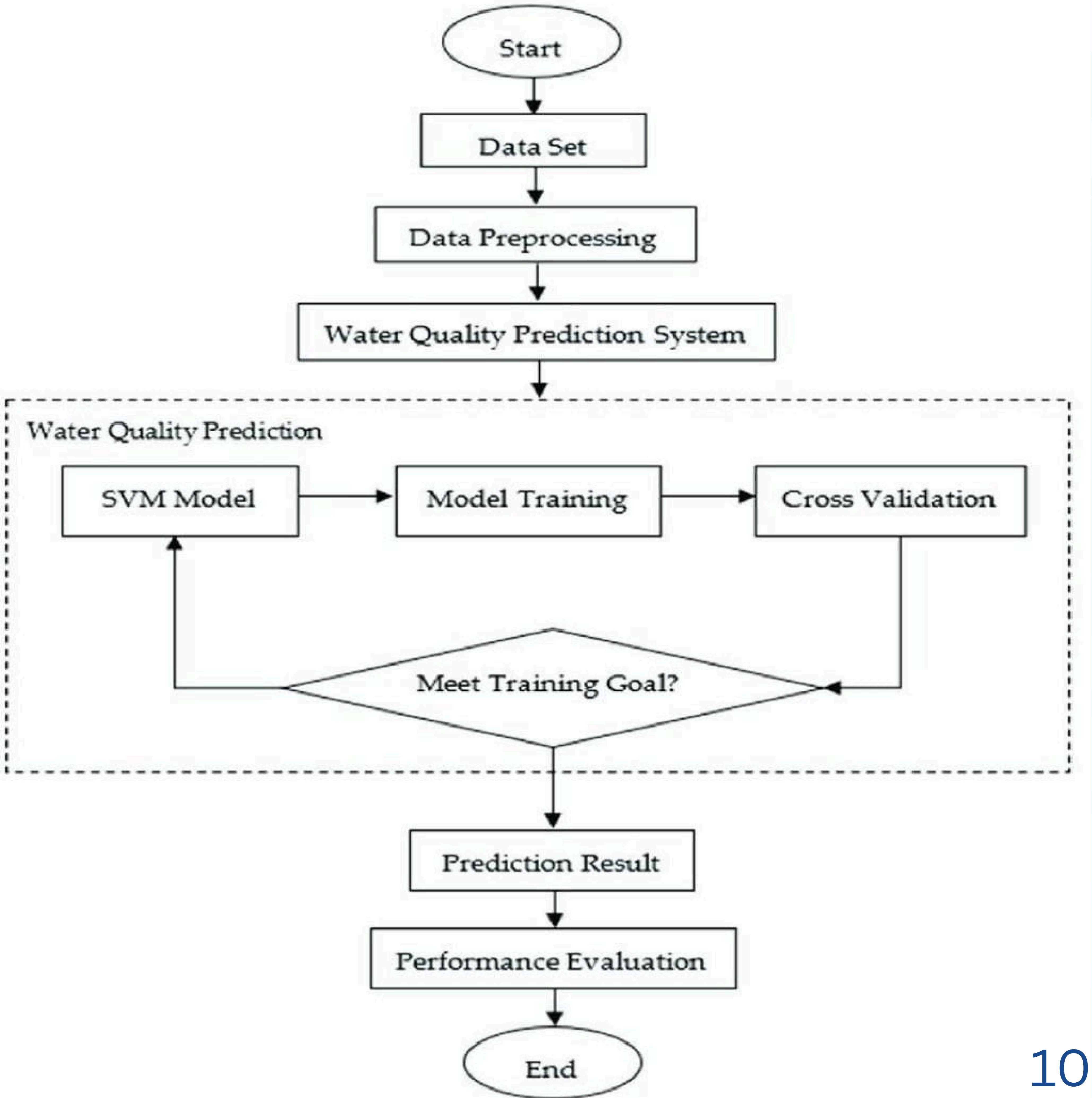
Module 3: AI Model Training Engine (Random Forest)

This is the heart of our project. Here, we employ the powerful Random Forest algorithm to train our predictive model. The engine processes the historical data, learning the complex relationships between geological conditions, climate, and the presence of groundwater. This rigorous training is what enables our system to achieve an most highest accuracy.

Module 4: Real-Time Prediction & Mapping

This is the operational engine that delivers results to the user. When a user inputs data for a new location, this module applies the trained Random Forest model to generate a precise prediction. The results, including well suitability, estimated depth, and potential discharge, are often visualized , providing a clear and actionable guide.

FLOW DIAGRAM



REFERENCE PAPERS

Research Papers & Articles

- Smith, J., & Brown, K. (2023). AI-Based Groundwater Prediction Using Machine Learning. International Journal of Environmental Sciences, 45(2), 120-135.

- Lee, P., & Wang, Y. (2022). Random Forest Approach for Predicting Groundwater Availability. Journal of Hydrology and Earth System Sciences, 78(4), 250-265.

- "Short-term Prediction of Groundwater Level Using Improved Random Forest Regression with a Combination of Random Features" Authors: Zhang, W., Zeng, S., & Wang, J. Published: 2023

Datasets

- Kaggle. (2024). Groundwater Level Prediction Dataset. Retrieved from www.kaggle.com

THANK YOU