SMART WATER USAGE SPIKE DETECTION

Project Type: Machine Learning Project
Student Name: Muthu Rakshana.R
Roll Number: E24AI031

1. Introduction

Water is a critical natural resource, and efficient water management is essential, especially during peak usage periods such as festivals and summer seasons. Households often experience sudden increases in water consumption due to higher temperatures, increased occupancy, or special occasions. Detecting such abnormal spikes early can help in planning, conservation, and awareness.

This project focuses on predicting household water usage and identifying abnormal spikes using Machine Learning regression techniques. A user-friendly Streamlit web application is developed to allow real-time prediction based on user inputs.

2. Problem Statement

Sudden and unexpected spikes in household water consumption can lead to water shortages and inefficient resource management. Traditional monitoring systems often fail to predict these spikes in advance.

Objective: To design a machine learning-based system that predicts daily household water usage and detects abnormal spikes using historical and contextual data.

3. Dataset Description

A synthetic dataset with 100 records is used for this project. The dataset represents realistic household water usage patterns.

Input Features:

- **Temperature (°C):** Daily average temperature
- **DayType:** 0 – Weekday, 1 – Weekend
- **HouseholdSize:** Number of people in the house
- **Season:** 0 – Summer, 1 – Winter, 2 – Monsoon
- **Festival:** 0 – No, 1 – Yes
- **PreviousUsage:** Water usage of the previous day (litres)

Output Variable:

- **WaterUsage:** Predicted daily water consumption (litres)

4. Methodology

4.1 Data Preprocessing

- Categorical values were encoded numerically
- Feature scaling was applied using StandardScaler
- Dataset was split into training (80%) and testing (20%) sets

4.2 Machine Learning Models Used

a) Ridge Regression

Ridge Regression is a linear regression technique with L2 regularization. It reduces overfitting by penalizing large coefficients.

b) Bagging Regressor (Ensemble Method)

Bagging (Bootstrap Aggregating) combines multiple models to improve prediction accuracy and stability.

5. Spike Detection Logic

The predicted water usage is compared with the average historical usage.

- **High Usage (Spike Detected):** > 20% above average
- **Medium Usage (Possible Spike):** 5% – 20% above average
- **Low Usage (Normal):** 5% above average

This rule-based logic helps in identifying abnormal consumption patterns.

6. System Architecture

1. User enters input values through Streamlit UI
2. Inputs are scaled using the trained scaler
3. Ridge and Bagging models generate predictions
4. Spike detection logic classifies the usage
5. Results and alerts are displayed to the user

7. Streamlit Application

The Streamlit web application provides:

- Interactive input controls (sliders, dropdowns)
- Predict button to trigger model execution
- Display of predicted usage values
- Visual alerts for spike detection

The app ensures ease of use and real-time interaction.

8. Results and Discussion

The system successfully predicts daily water usage based on user inputs. Ensemble predictions were found to be more stable compared to individual regression outputs. The spike detection mechanism effectively categorizes usage into Normal, Possible Spike, and Spike Detected.

9. Applications

- Household water management
- Smart city water monitoring systems
- Festival and seasonal demand forecasting
- Awareness and conservation planning

10. Conclusion

This project demonstrates the effective use of regression and ensemble learning techniques for predicting water usage and detecting abnormal spikes. The Streamlit interface enhances usability and makes the system accessible to non-technical users. The approach can be extended to real-world IoT-based water monitoring systems.