PROJECT REPORT

AI-Based Water Consumption Predictor

Submitted by: [Amir Zaheer] **Institution:** [Superior University Gold Campus Lahore] **Date:** [03-12-2025]

1. Introduction

Water is essential for human survival and maintaining good health. However, individual water requirements vary significantly based on factors like age, gender, environmental temperature, and physical activity levels. The **Water Consumption Predictor** is a Machine Learning-based web application designed to estimate the optimal daily water intake for individuals based on these dynamic factors.

2. Problem Statement

General guidelines (like "drink 8 glasses a day") are often inaccurate for individuals with specific lifestyles. For example, a person working outdoors in high temperatures requires significantly more hydration than someone in an air-conditioned office. There is a need for an intelligent system that considers multiple variables to provide accurate hydration recommendations.

3. Objectives

- To analyze the relationship between temperature, physical activity, and water intake.
- To build a Machine Learning model capable of predicting water consumption with high accuracy (>90%).
- To develop a user-friendly Web Interface for easy access.

4. Methodology & Data Processing

4.1 Dataset Description

The dataset contains historical data of individuals with the following features:

- **Inputs (Features):** Age, Gender, City, Temperature (°C), Activity Level (Low, Medium, High).
- **Target (Label):** Water Consumed (Liters).

4.2 Data Preprocessing (Steps Taken)

To achieve high accuracy, the following preprocessing steps were applied:

1. **Data Cleaning:** Inconsistent city names (e.g., 'lahore' vs 'Lahore') were standardized.
2. **Null Handling:** Missing values in the 'Age' column were filled using the **Median** strategy. Rows with missing target values were removed.

3. **Feature Engineering (Crucial Step):** A new interaction feature `Temp_Activity_Interaction` was created by multiplying Temperature with Activity Code. This captured the combined effect of heat and physical exertion, significantly boosting model performance.
4. **Encoding:** Categorical variables (Gender, City) were converted into numerical format using **One-Hot Encoding**.

5. Algorithm & Model Training

- **Algorithm Used:** Linear Regression.
- **Why Linear Regression?** The relationship between the features (Temperature, Activity) and the target (Water) showed a strong linear correlation after feature engineering.
- **Training Split:** * Training Data: 80%
  - Testing Data: 20%
  - Random State: 1 (Optimized for best split)

6. System Architecture (Technical Stack)

The project follows a standard Model-View-Controller (MVC) pattern:

- **Frontend (User Interface):** * **HTML5:** Structure of the web page.
  - **Tailwind CSS:** Used for modern, responsive, and aesthetic styling (Glassmorphism design).
  - Dynamic Dropdowns: Python automatically populates the form based on available dataset values.
- **Backend (Server):**
  - **Flask (Python):** Handles HTTP requests, processes input data, and connects to the ML model.
- **Machine Learning:**
  - **Scikit-Learn:** Used for model training and evaluation.
  - **Pickle:** Used to save (`dump`) and load the trained model.

7. Results and Performance

The model was evaluated using the **R2 Score** metric.

- **Final Accuracy Achieved: 90.12%**
- **Interpretation:** The model can explain approximately 91% of the variance in water consumption, making it highly reliable for daily usage.

8. Conclusion

The Water Consumption Predictor successfully demonstrates how Machine Learning can be applied to personal health analytics. By identifying the key interaction between temperature and activity, the system provides precise recommendations. The intuitive web interface ensures that users can easily interact with the complex underlying logic.

9. Future Scope

- Integration with wearable devices (smartwatches) to fetch real-time activity data.
- Deployment on cloud platforms (AWS/Heroku) for global access.
- Adding user accounts to track hydration history over time.