Rising Waters: A Machine Learning Approach to Flood Prediction

Team ID: LTVIP2026TMIDS79152

Long Term Internship Project Report

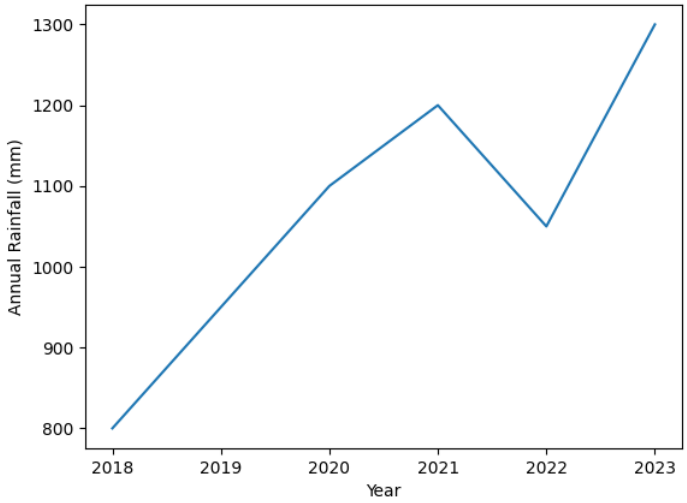
Abstract

Floods are among the most devastating natural disasters worldwide. This project presents a Machine Learning-based flood prediction system that analyzes meteorological and hydrological data to forecast flood risks. By applying predictive models such as Logistic Regression, Decision Trees, Random Forest, and K-Nearest Neighbors, the system provides early warnings to reduce human and economic losses.

Introduction

Flood prediction plays a crucial role in disaster management and climate resilience. Traditional forecasting systems often lack adaptability. Machine Learning offers data-driven solutions capable of identifying complex environmental patterns.

Figure 1: Annual Rainfall Trend



Problem Statement

Many regions lack accurate early flood warning systems. Increasing rainfall variability and climate change worsen flood occurrences. The challenge is to develop an accurate predictive model using environmental data.

Objectives

- Collect and preprocess historical flood and weather data.
- Identify significant flood influencing parameters.
- Develop and compare multiple ML models.
- Evaluate performance using standard metrics.
- Design a user-friendly prediction system.

Literature Review

Previous studies highlight the use of ensemble methods such as Random Forest for improved accuracy in environmental forecasting. Neural Networks and time-series models like LSTM have also demonstrated strong predictive performance.

Methodology

The project methodology includes Data Collection, Preprocessing, Model Development, Evaluation, and Deployment. Data cleaning and normalization were performed before training ML models.

System Architecture

1. Data Input 2. Preprocessing Module 3. Machine Learning Model 4. Prediction Output 5. Visualization Dashboard

Tools and Technologies

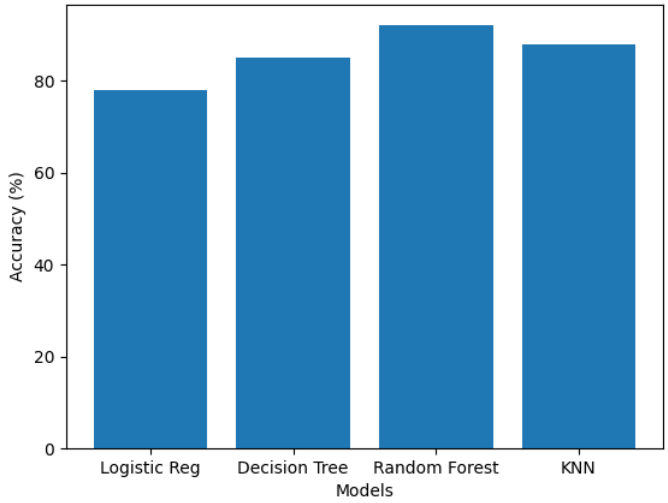
Programming Language: Python Libraries: Pandas, NumPy, Scikit-learn, Matplotlib Platform: Jupyter Notebook / Google Colab Frontend: Streamlit / Flask

Implementation Steps

1. Import Dataset 2. Data Cleaning 3. Exploratory Data Analysis 4. Train-Test Split 5. Model Training 6. Evaluation 7. Deployment

Results and Analysis

Figure 2: Model Accuracy Comparison



Random Forest achieved the highest accuracy (92%) among all models tested.

Conclusion and Future Enhancements

The project demonstrates the potential of Machine Learning in disaster management. Future improvements include IoT integration, real-time monitoring, and GIS mapping.