Title: Flood Risk Prediction

Streamlit Deployment Link:

https://it-vedant-project-zarknshuahd3pgguikeqnd.streamlit.app/

Problem Statement:

Unpredictable rainfall patterns pose significant challenges to agriculture, urban planning, and disaster management. Farmers face difficulties in crop planning, cities struggle with waterlogging and floods, and governments lack accurate insights for resource allocation. Traditional methods of weather forecasting often fail to provide localized, timely, and percentage-based rainfall predictions that communities can rely on.

There is a clear need for a data-driven, machine learning based rainfall prediction system that can deliver more accurate, interpretable, and actionable rainfall forecasts.

Purpose solution:

We propose a machine learning based rainfall prediction system that leverages historical weather data (rainfall, temperature, humidity, pressure, etc.) to predict the probability and percentage of rainfall for a given region.

Our solution uses data preprocessing, feature scaling, and an optimized XGBoost model to ensure accuracy. The model is deployed using Streamlit, providing a simple and interactive web app where users can input weather parameters and instantly receive rainfall predictions in percentage form.

This system empowers farmers, policymakers, and urban planners with actionable insights for better decision-making in agriculture, water management, and flood prevention.

Methodology:

Tech Stack

- **Programming Language:** Python
- **Libraries & Frameworks:**
    - **Pandas, NumPy** → Data preprocessing & feature engineering
    - **Scikit-learn** → Scaling & evaluation metrics
    - **XGBoost** → Machine learning model for rainfall prediction
    - **Streamlit** → Interactive web app deployment
- **Tools:** Jupyter Notebook, GitHub, Streamlit Cloud

Problem Understanding – Identified the need for accurate rainfall percentage prediction.

Data Collection & Cleaning – Historical weather data (rainfall, temperature, humidity, etc.) processed and cleaned.

Feature Engineering & Scaling – Applied preprocessing and scaling using StandardScaler.

Model Development – Trained an XGBoost classifier/regressor for accurate rainfall percentage prediction.

Evaluation – Validated using accuracy, precision, recall, and other performance metrics.

Deployment – Built a Streamlit app for real-time user interaction and predictions.

Results / Demo

Model Used: XGBoost Regressor

- **Train R^2 Score:** 0.819 → Model explains about 82% of the variance in training data
- **Test R^2 Score:** 0.810 → Strong generalization on unseen data
- **Test MAE:** 0.0175 → Very small average prediction error
- **Test MSE:** 0.00049 → Low overall error
- **Test RMSE:** 0.0221 → Predictions are very close to actual rainfall value

Interpretation

- The model shows high accuracy and reliability in predicting rainfall percentage.
- The small error values (MAE & RMSE) confirm that predictions are consistent and precise.
- Demonstrates the potential of machine learning in weather forecasting and risk management.

Slide-Friendly Version

Results (XGBoost Regressor):

- Train R^2: 0.819
- Test R^2: 0.810
- MAE: 0.0175
- RMSE: 0.0221
- Accurate and reliable rainfall prediction system

Future Scope of this project:

In the future, this rainfall prediction system can be enhanced in several ways for rainfall and flood predictions.

Conclusion:

In this project, we successfully developed a machine learning-based regression model to predict flood levels using historical and environmental data. By analyzing key factors such as rainfall, river flow, soil moisture, and other meteorological parameters, our model was able to learn complex patterns and provide accurate predictions. The evaluation metrics, including R^2, MAE, and RMSE, indicate that the model performs well in capturing trends and minimizing prediction errors. These predictions can serve as a valuable tool for early warning systems, disaster management authorities, and policymakers, enabling proactive measures to mitigate the impact of floods on communities.

The success of this project underscores the potential of machine learning in environmental forecasting and disaster risk reduction. Future improvements could involve integrating real-time data, using ensemble modeling techniques, and incorporating geospatial information to further enhance prediction accuracy and reliability.