Rising Waters: A Machine Learning Approach To Flood Prediction

Solution Architecture

1) Data Sources Layer (Input)

Inputs include river and stream water level sensors, rainfall gauges, satellite and radar feeds, soil moisture and terrain data, historical flood records, and dam discharge information. Data may arrive in real time or as batch datasets.

2) Data Ingestion Layer

This layer collects and transports data through APIs and streaming pipelines. It validates incoming information, handles missing values, and aligns timestamps.

3) Data Storage Layer

Raw and processed data is stored in data lakes, time-series databases, and relational systems designed for high volume, fast retrieval, and reliability.

4) Data Processing & Feature Engineering

Here raw inputs are transformed into model-ready features such as rainfall intensity over time windows, river rise rate, upstream accumulation, and soil saturation indices.

5) Machine Learning Layer

Models perform regression for water level prediction, classification for flood detection, and time-series forecasting to estimate when flooding may occur.

6) Prediction & Decision Engine

Model outputs are translated into risk categories such as green, yellow, orange, and red, based on expert-defined thresholds.

7) Alerting & Visualization Layer

Results are delivered via dashboards, SMS alerts, GIS maps, and reporting tools for authorities and communities.

8) Feedback & Model Improvement

After events, predictions are compared with actual outcomes. Models and thresholds are refined, enabling continuous learning.

Flow Summary: Sensors → Ingestion → Storage → Processing → ML → Risk Score → Alerts → Learning Loop