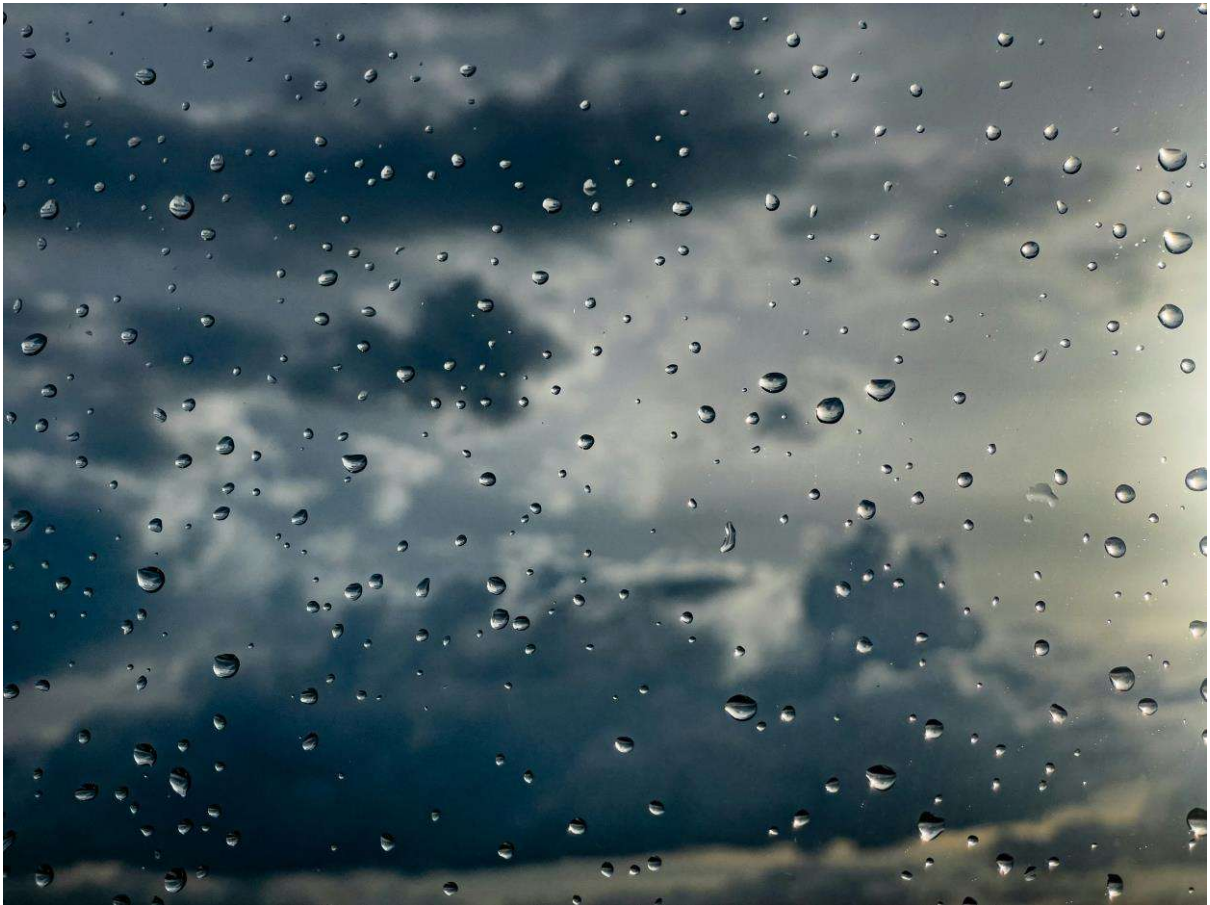


RainSight

(MLOps pipeline that predicts the next 21 days of rainfall probability)

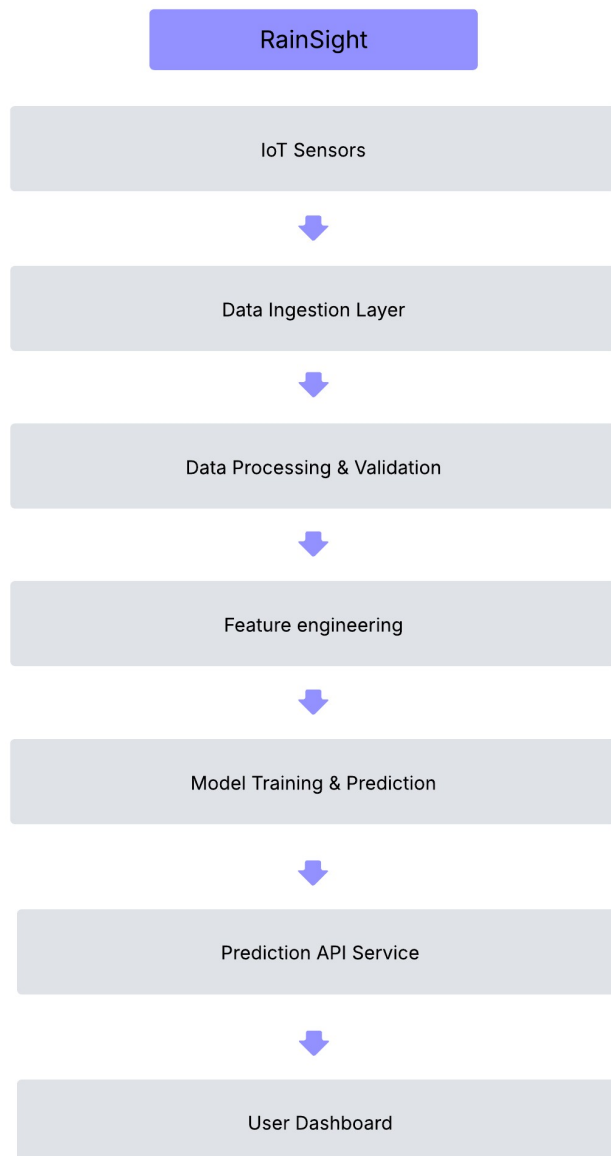


Tensor Team Presents

System Design Overview

The system ingests real-time weather data from IoT sensors, processes and cleans the data, and trains a machine learning model to predict daily rainfall probabilities for the next 21 days. It also includes mechanisms to handle faulty sensors and ensure data reliability.

System Diagram



Component Descriptions

01. IoT Sensors

- Collect real-time weather data (humidity, temperature, air pressure, etc.).
- Data is sent every 1 minute to the ingestion pipeline.

02. Data Ingestion Layer

- Gathers data from IoT devices via APIs.
- Stores raw data in a time-series database (e.g., InfluxDB, AWS Timestream).

03. Data Processing & Validation

- Cleans and preprocesses the data (removes duplicates, handles missing values).
- Implements anomaly detection to identify malfunctioning sensors (e.g., using statistical thresholds or ML-based anomaly detection).
- Uses fallback strategies like interpolation or redundant sensors to replace missing/corrupt data.

04. Feature Engineering

- Aggregates and transforms data for better model performance.
- Extracts features such as moving averages, trend analysis, and seasonality.

05. Model Training & Prediction

- Uses a machine learning model (e.g., LSTM, Random Forest, or XGBoost) to predict rainfall probability for the next 21 days.
- Trains the model daily using the latest validated data.
- Stores trained models in a model registry (e.g., MLflow).

06. Prediction API Service

- Deploys the trained model as an API (e.g., FastAPI, Flask, AWS Lambda).
- Provides predictions to consumers (e.g., mobile apps, dashboards).

07. User Dashboard / Consumer

- Displays rainfall forecasts for the next 21 days.
- Can be used by farmers, city planners, or other stakeholders.