MLOps System Design for Weather Forecasting

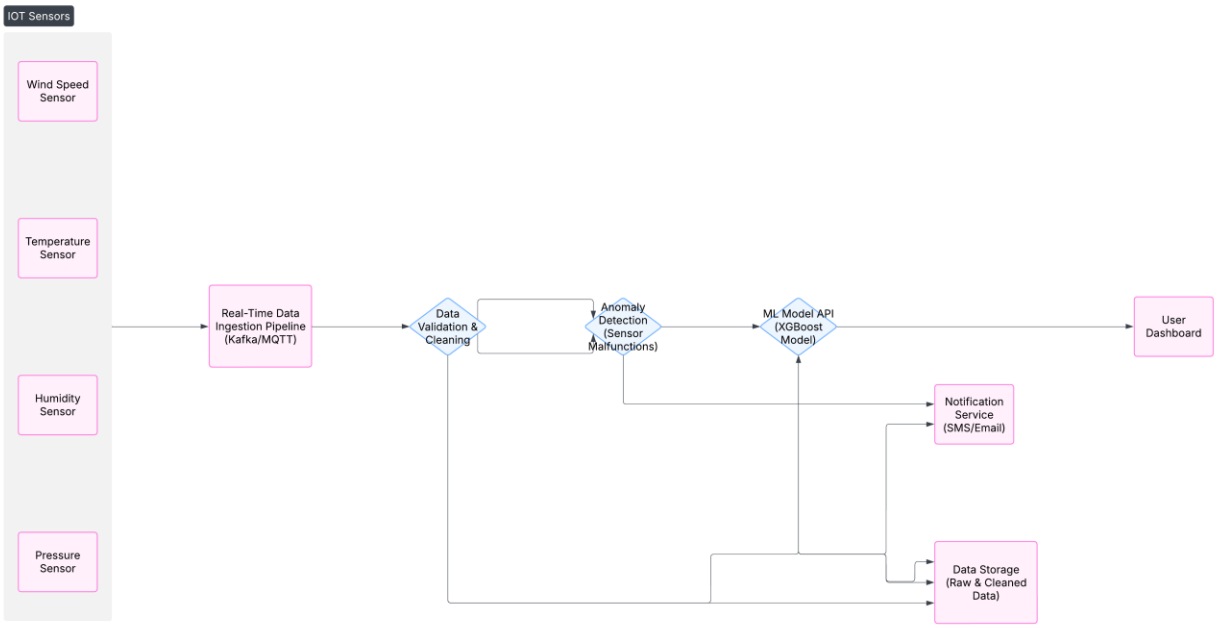
Team: TechSpark

IntelliHack 5.0

Intellihack_TechSpark_01

Submission Date: 3/8/2025

System Diagram:



Components Description:

1. IoT Sensors

IoT sensors collect real-time weather data, including:

- **Temperature** (°C)
- **Humidity** (%RH)
- **Wind Speed** (km/h)
- **Atmospheric Pressure** (hPa)
- **Cloud Cover** (%)

Possible Sensor Issues:

- **Inaccurate Readings:** Calibration errors or aging sensors.
- **Missing Data:** Power failures or connectivity issues.
- **Outliers:** Sudden, unrealistic changes in readings.

2. Data Ingestion Pipeline

The ingestion pipeline ensures real-time data flow using **Kafka/MQTT** to handle large sensor data streams efficiently. It supports:

- **Real-Time Data Collection:** Captures sensor readings every minute.
- **Message Queueing:** Manages high-frequency data traffic.
- **Loss Prevention:** Prevents data loss during transmission.

3. Data Validation & Cleaning

To ensure high-quality input data, the system:

- **Filters Outliers:** Removes unrealistic values (e.g., humidity > 100%).

- **Detects Anomalies:** Identifies faulty sensor readings.
- **Handles Missing Data:** Uses mean imputation or interpolation.
- **Normalizes Data:** Ensures consistency for ML model input.

4. Machine Learning Model API

The **ML Model API** provides real-time rainfall probability predictions using an optimized **XGBoost model**.

Key Functions:

- **Processes Cleaned Data:** Accepts validated sensor data.
- **Generates Predictions:** Computes rainfall probabilities.
- **Returns Outputs:** Provides structured API responses.
- **Ensures Fast Inference:** Designed for low-latency performance.

5. Database & Storage

The system maintains both **real-time and historical storage**:

- **Real-Time Database (Redis/InfluxDB):** Stores latest predictions for instant retrieval.
- **Historical Storage (SQL/Data Lake):** Logs past weather data for future analysis and model retraining.

6. User Application / Dashboard

The **User Dashboard** allows farmers and stakeholders to access real-time and historical rainfall predictions.

Features:

- **Live Weather Updates:** Displays real-time forecasts.

- **Historical Trends:** Enables comparison with past data.
- **Alerts & Notifications:** Sends SMS/email for critical conditions.
- **Interactive Charts:** Provides visual analytics for weather trends.

A well-structured UI ensures users can easily interpret forecasts and take necessary actions.