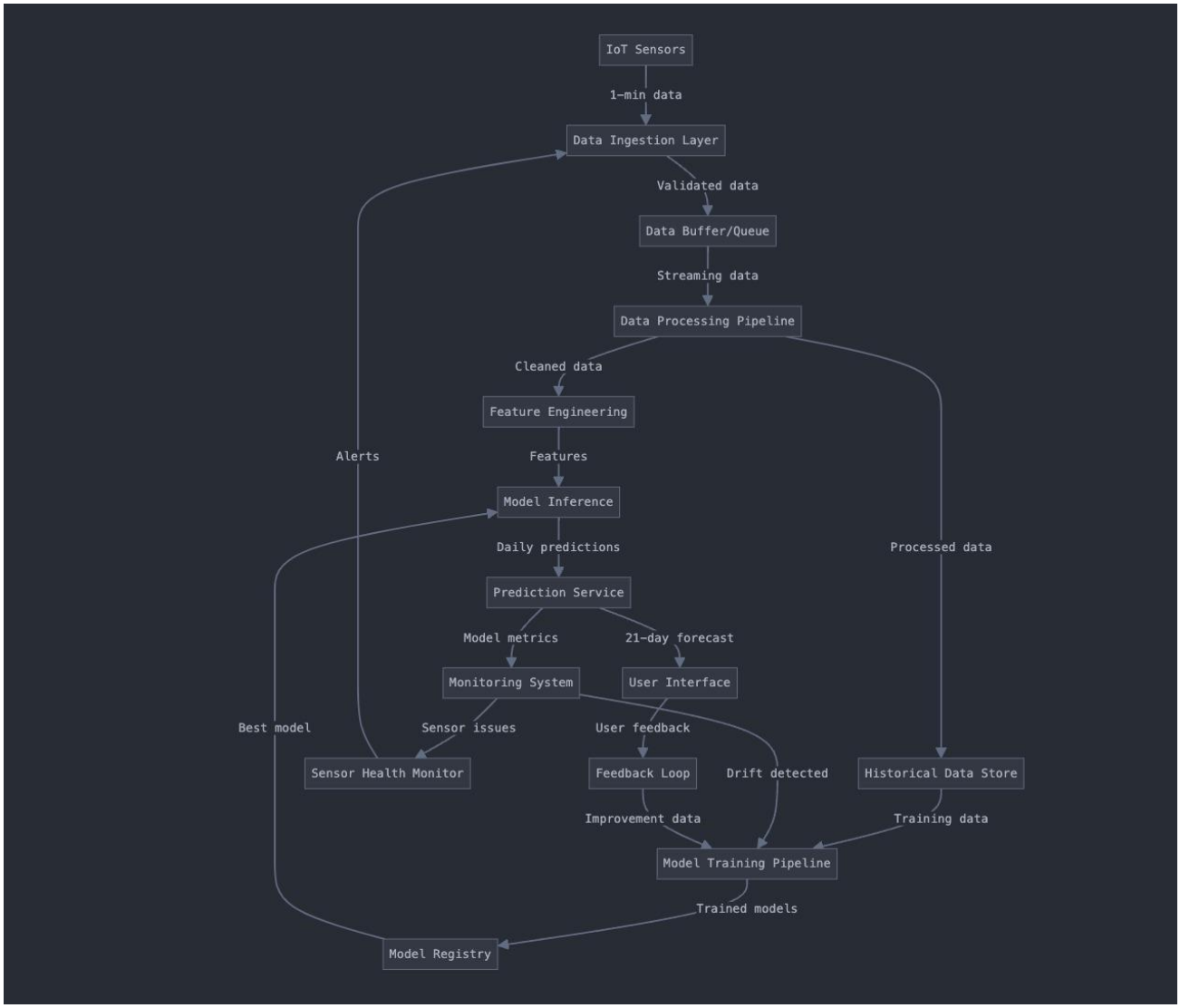
Weather Prediction for 21 days

Report Part 2:

Predicting Rainfall Probability Using IoT

Sensor Data



Introduction

This report outlines the design of a robust machine learning system to predict daily rainfall probabilities for the next 21 days using real-time data gathered from IoT sensors. The system is designed to handle challenges such as sensor malfunctions and ensure highquality predictions through continuous improvement mechanisms. Below, we describe the system's components and their roles in achieving the project goals.

System Components and Workflow IoT Sensors

IoT sensors are deployed to collect real-time environmental data at 1-minute intervals. These sensors measure variables such as temperature, humidity, and pressure, which are critical for weather prediction.

Data Ingestion Layer
The data ingestion layer is responsible for collecting raw data from the IoT sensors. It validates the data to filter out incomplete or corrupted entries, ensuring that only highquality data proceeds further in the pipeline.

Data Buffer/Queue

Validated data is temporarily stored in a buffer or queue. This component ensures smooth streaming of data into subsequent layers and prevents loss during peak loads or network interruptions.

Data Processing Pipeline

The data processing pipeline cleans the incoming streaming data by removing noise, handling missing values, and preparing it for analysis. This step ensures that the data is reliable and usable for feature extraction.

Feature Engineering

Feature engineering extracts meaningful features from the cleaned data. These features include derived metrics (e.g., temperature trends or humidity changes) that improve the accuracy of rainfall predictions.

Model Inference

A pre-trained machine learning model uses these features to infer daily rainfall probabilities. The model generates predictions based on historical patterns and real-time environmental factors.

Prediction Service

The prediction service provides users with daily predictions and a 21-day rainfall forecast. This service acts as an interface between the model and end-users, delivering actionable insights.

Monitoring System

The monitoring system tracks key metrics such as model accuracy and detects drift in predictions over time. It also oversees sensor health to identify malfunctioning devices.

Sensor Health Monitor

This component specifically monitors IoT sensors for issues such as hardware failures or irregular readings. Alerts are generated when sensor malfunctions are detected, enabling timely maintenance or replacement.

User Interface (UI)

The user interface displays the 21-day rainfall forecast in an accessible format. It also allows users to provide feedback on prediction accuracy, contributing to system improvement.

Feedback Loop
User feedback is incorporated into the system via a feedback loop. This mechanism identifies areas for improvement and provides additional training data to refine the model.

Historical Data Store
Processed data is stored in a historical repository, creating a dataset for retraining models when performance degrades or new patterns emerge.

Model Training Pipeline
Using historical data and improvement feedback, this pipeline retrains models periodically to ensure optimal performance under changing conditions.

Model Registry
The model registry maintains a catalog of trained models, ensuring that only the bestperforming model is deployed in production.

System Features and Strengths
1. Real-Time Predictions: The system processes high-frequency sensor data (1-minute intervals) to provide accurate daily forecasts.

2.Fault Tolerance: Sensor malfunctions are handled via monitoring systems and alerts.

3.Continuous Improvement: User feedback and drift detection enable ongoing refinement of predictions.

4.Scalability: The modular architecture supports scalability for larger datasets or additional sensors.

5.User Accessibility: Predictions are made available through an intuitive user interface.

Conclusion
The proposed system integrates IoT technology with advanced machine learning techniques to deliver reliable rainfall predictions over a 21-day period. By addressing challenges like sensor malfunctions and leveraging feedback loops, this design ensures robustness, accuracy, and continuous improvement.