Team-SCOPE

Intellihack 5.0 Task – 1

Weather Forecasting Challenge

Part II

Real-Time Rain Prediction System with Advanced Tech Stack

1. Introduction

This report presents a **high-performance, scalable, and fault-tolerant** system to predict daily rain probabilities for the next **21 days** using **real-time IoT sensors**. The system leverages modern **AI/ML models, cloud computing, and DevOps best practices** to ensure reliability and accuracy.

---

2. System Architecture & Tech Stack

◆ System Components & Technologies

Component	Technology Stack
IoT Sensors	Raspberry Pi + DHT22 (Temperature, Humidity) + Anemometer (Wind Speed)
Data Ingestion (API Layer)	**FastAPI (Python) + Kafka (Real-time Streaming) + WebSockets**
Preprocessing & Storage	**Apache Spark + PostgreSQL + AWS S3 (Data Lake)**
Machine Learning Model	**XGBoost / Random Forest + TensorFlow/Scikit-Learn (ML Model)**
Model Deployment	**AWS SageMaker / Google Vertex AI / ONNX (Optimized ML Model Serving)**
Monitoring & Retraining	**MLflow (Model Monitoring) + Airflow (Retraining Pipeline)**
API Gateway & Microservices	**GraphQL + AWS Lambda / Kubernetes (Scalability)**
Frontend (Dashboard)	**Next.js + Tailwind CSS + Recharts (Visualizations)**
CI/CD & Deployment	**Docker + Kubernetes + GitHub Actions + Terraform (Infrastructure as Code)**

---

3. System Diagram

4. Fault Tolerance & Error Handling

✅ **Real-Time Data Validation:** Kafka + Apache Spark detects missing/corrupt data instantly.
✅ **Backup Sensors:** If one sensor fails, the system switches to backup IoT devices.
✅ **Anomaly Detection:** Auto-flags extreme readings and replaces them with historical rolling averages.
✅ **Failover Prediction Model:** If the ML model fails, a secondary statistical model takes over.
✅ **Redundant Cloud Backup:** PostgreSQL & AWS S3 store all historical and real-time data.

---

5. Deployment & Scalability Strategy

◆ Cloud Infrastructure

- **Compute:** AWS Lambda / Kubernetes (Auto-Scalable Model Serving)
- **Storage:** AWS S3 (Data Lake), PostgreSQL (Structured Data)
- **ML Model Deployment:** AWS SageMaker (Optimized AI Inference)
- **CI/CD:** Docker + Kubernetes + GitHub Actions (Automated Deployment)
- **Monitoring:** Prometheus + Grafana (System Performance), MLflow (Model Drift)

---

6. Future Enhancements

📌 **Edge AI:** IoT sensors run lightweight ML models locally (faster predictions).
📌 **Blockchain Integration:** Secure data storage and prevent tampering.
📌 **Federated Learning:** Improve model accuracy across multiple locations without data sharing.

---

7. Conclusion

By integrating **real-time IoT data, scalable cloud computing, and optimized ML models**, this system ensures **high-accuracy rain predictions** while being **fault-tolerant and scalable**. This solution can greatly benefit **precision agriculture, flood prediction, and weather monitoring systems**.