Intellihack 5.0

Task – 1

Weather Forecasting Challenge – Part II

Team	:	Outlier Rejects
Author	:	H. D. Rivindu Ashinsa Handuwala
Date	:	10-03-2025

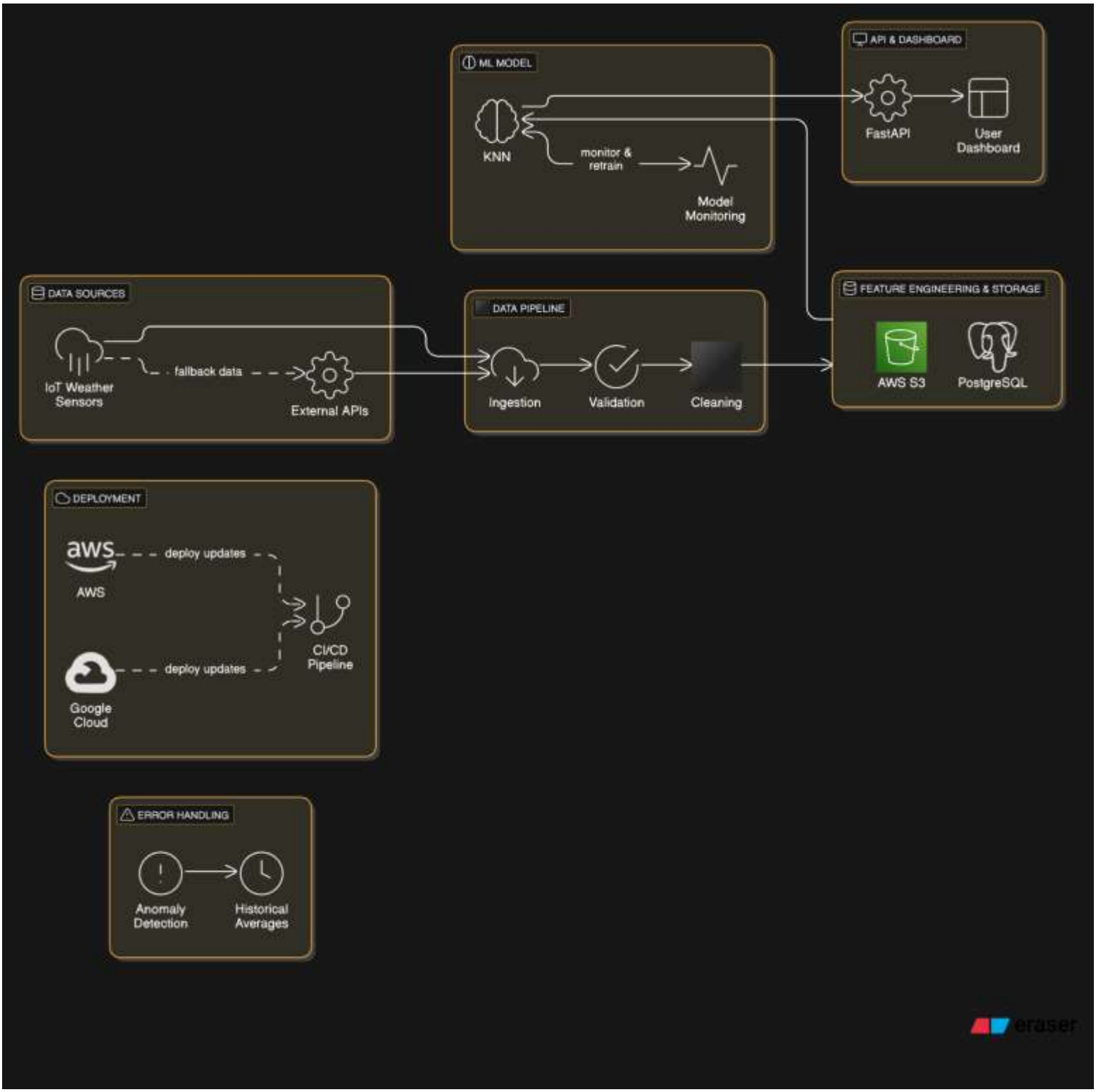
Introduction

This report presents a **real-time machine learning system** for predicting rainfall using IoT sensors. The system ensures **data reliability**, handles **sensor malfunctions**, and delivers **daily rain probabilities for the next 21 days**.

System Architecture and Data flow

- System Components:

1. **Data Sources:** IoT weather sensors & external APIs.
2. **Data Pipeline:** Real-time data ingestion, validation, and cleaning.
3. **Feature Engineering & Storage:** Stores data in **AWS S3 / PostgreSQL**.
4. **ML Model:** Uses **Random Forest** for predictions.
5. **Model Monitoring:** Tracks performance & retrains if drift detected.
6. **API & Dashboard:** Displays rain probability predictions.


ML MODEL
KNN
monitor & retrain
Model Monitoring
API & DASHBOARD
FastAPI
User Dashboard
DATA SOURCES
IoT Weather Sensors
fallback data
External APIs
DATA PIPELINE
Ingestion
Validation
Cleaning
FEATURE ENGINEERING & STORAGE
AWS S3
PostgreSQL
DEPLOYMENT
aws
AWS
deploy updates
CI/CD Pipeline
deploy updates
Google Cloud
ERROR HANDLING
Anomaly Detection
Historical Averages
eraser

Error Handling & Fault Tolerance

- **Handling Sensor Malfunctions:**
  - If IoT sensors fail, data is retrieved from external APIs.
  - Anomaly detection flags outliers & replaces them with historical averages.
- **Ensuring Reliable Predictions:**
  - **Automated retraining** every 30 days.
  - **Backup data storage** for continuous availability.

Conclusion & Deployment Strategy

- **Deployment:**
  - Model hosted on **AWS / Google Cloud** with **FastAPI for inference**.
  - **Continuous Integration (CI/CD) pipeline** for model updates.
- **Future Improvements:**
  - **Edge AI** for IoT sensors to process data locally.