Climate Intelligence System: Real-Time Weather Prediction and Analysis

**Group Members :** Waqar Nadeem (24K-2524) Nafees (24K-2526)

Azfar Abbasi (24K-2576)

**1. Problem Statement**

Current weather forecasting lacks comprehensive trend analysis and long-term insights, severely impacting **agriculture, disaster management, and public safety**. This project develops an intelligent system to process real-time data, identify complex patterns, and deliver **accurate, actionable forecasts** using advanced statistical analysis.

**2. Real-Life Importance**

Advanced climate intelligence offers critical value: **reducing crop losses (15-20%)** by optimizing agriculture; **saving lives** through early detection of extreme weather; supporting long-term **urban planning**; and enabling power companies to **optimize energy distribution** and renewable generation.

**3. Role of Data Science and Advanced Analysis**

The system relies on advanced data science techniques: **Pattern Recognition** identifies complex, non-linear climate relationships; **Statistical Modeling** (time-series regression and classification) generates accurate forecasts; **Automated Data Processing** ensures efficient real-time handling; and **Anomaly Detection** enables robust early warning systems for extreme events.

**4. Dataset Description**

We use **OpenWeatherMap Historical Climate Data (2020-2025, hourly)**, consisting of 10k-50k observations of temperature, pressure, wind, and precipitation. Target predictions are **Next-Day Temperature, Precipitation Probability, and Weather Category**. Key challenges include **5-10% missing values** and quality issues that require robust preprocessing via statistical imputation and outlier removal.

**5. Methodology**

The pipeline includes **Automated Data Collection** (OpenWeatherMap API), rigorous **Preprocessing** (imputation, outlier removal via IQR, feature engineering using **pandas/NumPy**), and **Exploratory Data Analysis (EDA)**. We will develop and evaluate **Statistical Time-Series Models**, optimizing model parameters with techniques like cross-validation, and evaluating performance using **RMSE/MAE** (Regression) and **Accuracy/Precision/Recall** (Classification).

**6. Expected Deliverables**

Key outputs include the **Python Codebase** (containing scripts for data handling and modeling), **Optimized Statistical Models** (targeting >85% classification accuracy and <2∘C temperature error), **Technical Documentation**, and a final **Presentation** of findings.

**7. Project Significance**

This project provides an **end-to-end data science demonstration**, creating a **scalable, deployable** prediction system that solves a tangible, high-impact problem across agriculture, public safety, and planning sectors.