**Weather Trend Forecasting - Technical Assessment Report**

**Author:** Prajwal Kumar  
**Date:** April 2025

**1. Introduction**

**PM Accelerator Mission**

By making industry-leading tools and education available to individuals from all backgrounds, we level the playing field for future PM leaders. This is the PM Accelerator motto, as we grant aspiring and experienced PMs what they need most – Access. We introduce you to industry leaders, surround you with the right PM ecosystem, and discover the new world of AI product management skills.

**Objective**

The purpose of this project is to analyze and forecast global weather trends using the "Global Weather Repository" dataset. The assessment includes data preprocessing, exploratory data analysis (EDA), multiple forecasting models, and advanced analyses such as anomaly detection, climate pattern study, and spatial analysis.

**Dataset Overview**

The dataset, sourced from Kaggle, contains daily weather data from cities worldwide with over 40 features, including temperature, precipitation, air quality, wind speed, and humidity.

**2. Data Cleaning and Preprocessing**

**Handling Missing Values**

* The dataset was examined for missing values using pandas functions.
* Missing values in numerical features were handled using **median imputation**.
* Missing categorical values were filled using **mode imputation**.
* Timestamp column was interpolated to fill in missing time-based values.

**Outlier Detection and Treatment**

* Used **Isolation Forest** to identify extreme anomalies in temperature, wind speed, and air pressure.
* Capped extreme values based on statistical thresholds like **1.5x IQR method**.

**Feature Engineering**

* Created additional features such as:
  + **Temperature Difference**: Convert Celsius to Fahrenheit.
  + **Relative Humidity Index**: Derived from humidity and temperature.
  + **Wind Chill Factor**: Factored using wind speed and temperature.
  + **Seasonal Indicators**: Derived from the date to label weather seasons.
* Applied **MinMaxScaler** to normalize numerical features for better model performance.**A group of graphs showing different weather parameters

  AI-generated content may be incorrect.A screenshot of a graph

  AI-generated content may be incorrect.A diagram of a global temperature distribution

  AI-generated content may be incorrect.**

**3. Exploratory Data Analysis (EDA)**

**Key Findings:**

* **Temperature Trends:** Warmer temperatures near the equator, colder in polar regions.
* **Precipitation Patterns:** Seasonal rainfall distribution varies across continents.
* **Correlation Analysis:**
  + Temperature and humidity showed strong negative correlation.
  + Wind speed and air pressure had notable positive correlation.

**Visualizations:**

* **Time Series Analysis**: Used seaborn and matplotlib to visualize seasonal variations.
* **Geospatial Plots**: Implemented Folium maps for spatial visualization.
* **Histograms & Boxplots**: Identified data distributions and outliers.

A graph with blue squares

AI-generated content may be incorrect.A comparison of a graph

AI-generated content may be incorrect.A blue and red dotted graph

AI-generated content may be incorrect.A graph of a graph showing the temperature of the earth

AI-generated content may be incorrect.A graph of a temperature

AI-generated content may be incorrect.A graph showing the temperature distribution

AI-generated content may be incorrect.A graph with a red line

AI-generated content may be incorrect.

**4. Forecasting Models**

**Model Selection & Evaluation**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **MAE (°C)** | **RMSE (°C)** | **R² Score** |
| ARIMA | 1.89 | 2.45 | 0.78 |
| Prophet | 1.67 | 2.18 | 0.82 |
| Random Forest | 1.45 | 1.95 | 0.87 |
| Ensemble (Weighted Avg) | **1.38** | **1.89** | **0.89** |

* **ARIMA:** Traditional time series model used with differencing to remove trends.
* **Prophet:** Developed by Facebook, used for handling seasonality effectively.
* **Random Forest:** A non-linear model trained on historical data for regression.
* **Ensemble Model:** Combined outputs of ARIMA, Prophet, and Random Forest using weighted averaging.

A graph showing a line

AI-generated content may be incorrect.A graph of a graph showing the number of data

AI-generated content may be incorrect.A graph with numbers and a line

AI-generated content may be incorrect.A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

AI-generated content may be incorrect.A graph with blue bars

AI-generated content may be incorrect.A graph with a bar graph

AI-generated content may be incorrect.A graph showing a line graph

AI-generated content may be incorrect.

**5. Advanced Analysis**

**Anomaly Detection**

* Used **Isolation Forest** to flag extreme temperature fluctuations and sudden weather shifts.
* Visualized anomalies using scatter plots.

**Climate Pattern Analysis**

* Long-term temperature trends were analyzed using **rolling averages**.
* Urban regions showed a steady increase in temperatures, indicating possible urban heat effects.

**Air Quality & Weather Correlation**

* Pearson correlation showed **PM2.5 levels** were inversely related to wind speed.
* High humidity contributed to **lower air quality** due to pollutant condensation.

**Geospatial & Spatial Analysis**

* Created **heatmaps** of extreme weather events using Folium.
* Mapped precipitation density and drought-prone regions globally.

A graph with blue squares

AI-generated content may be incorrect.A comparison of a graph

AI-generated content may be incorrect.A graph showing the temperature trends

AI-generated content may be incorrect.A graph with blue bars

AI-generated content may be incorrect.A graph of red bars

AI-generated content may be incorrect.A graph of green bars

AI-generated content may be incorrect.A screenshot of a graph

AI-generated content may be incorrect.A comparison of a graph

AI-generated content may be incorrect.

**6. Conclusions and Recommendations**

**Key Insights:**

* Global warming indicators were evident in long-term temperature analysis.
* The combination of multiple models improves forecasting accuracy.
* Wind speed plays a critical role in air quality variations.

**Future Work:**

* Integrate **deep learning models** such as LSTM and GRU for more advanced forecasting.
* Implement **real-time weather data streaming** via APIs.
* Improve **anomaly detection using reinforcement learning** techniques.

**7. Project Repository & Submission**

* **GitHub Repository:** [Insert Link]
* **Project Folder Includes:**
  + Jupyter Notebook with complete code.
  + Processed dataset.
  + Model evaluation results and visualizations.
  + **README.md** detailing methodology and insights.

**End of Report**