Weather Trend Forecasting Project

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.

Abstract

This report presents a comprehensive analysis of global weather trends using the Global Weather Repository dataset. The project covers data cleaning and preprocessing, exploratory data analysis (EDA), advanced anomaly detection, and the development of forecasting models (ARIMA, Prophet, and an ensemble approach). In addition, advanced analyses including climate trends, environmental impact assessments, feature importance evaluation, spatial analysis, and examination of geographical patterns are provided. The study is aligned with the PM Accelerator mission to drive impactful data insights for effective project management.

1. Introduction

The purpose of this project is to forecast weather trends and extract actionable insights from the Global Weather Repository dataset. The analysis includes cleaning raw data, performing exploratory analysis, building forecasting models, and executing advanced analyses to study long-term climate patterns, environmental impact, and geographical variations. The work aligns with the PM Accelerator mission by delivering strategic insights to enhance project management practices.

2. Data Cleaning & Preprocessing

The dataset was loaded using Python's pandas library. The 'last_updated' column was parsed as a datetime type to facilitate time series analysis, and missing values were handled using a forward-fill method. This preprocessing step ensured data consistency and prepared the dataset for subsequent analysis.

3. Exploratory Data Analysis (EDA)

Initial exploratory analysis involved plotting daily trends for temperature and precipitation to uncover underlying patterns. The following visualizations were generated:

- **Daily Temperature Trends:** Line plots displaying variations in temperature (°C) over time.
- **Daily Precipitation Trends:** Line plots showing precipitation (mm) patterns.

These visualizations provided insights into seasonal variations and potential trends that warranted further investigation.

4. Advanced EDA – Anomaly Detection

Anomaly detection was performed using the Isolation Forest algorithm on temperature and precipitation data. This method helped identify unusual weather events that could influence the forecasting models. Anomalies were visualized by overlaying scatter plots on the temperature time series.

5. Forecasting Models

a. ARIMA Forecast

An ARIMA model was built to forecast temperature based on historical data. The dataset was resampled to daily means, split into training and testing sets (with a 90/10 split), and the ARIMA model (order (2, 1, 2)) was fitted. The forecast was then compared with actual values to assess performance.

b. Prophet Forecast

The Prophet model was used as an alternative forecasting approach. Data was reformatted (renaming date and target columns to 'ds' and 'y'), and the model was trained on the same

training data. Prophet's forecasts for the test period were visualized to compare with the ARIMA model.

c. Ensemble Forecast

To improve forecast accuracy, an ensemble approach was implemented by averaging the predictions from the ARIMA and Prophet models. The ensemble forecast was aligned with the test period dates and plotted against actual data, showing improved performance over individual models.

d. Model Evaluation

Forecast accuracy was evaluated using the Root Mean Squared Error (RMSE) metric. The RMSE for ARIMA, Prophet, and the ensemble forecast were calculated and compared, with the ensemble model showing a reduction in forecasting error.

6. Advanced Analyses

a. Climate Analysis

Long-term climate patterns were studied by grouping the data by country and resampling the temperature data on a monthly basis. The analysis highlighted seasonal trends and variations across selected countries, helping to understand regional climate dynamics.

b. Environmental Impact

The correlation between weather parameters and air quality metrics was investigated using a correlation matrix. Key variables such as temperature, humidity, precipitation, and various air pollutants (e.g., PM2.5, PM10, Ozone) were analyzed to assess environmental impacts on weather conditions.

c. Feature Importance

A Random Forest model was used to assess feature importance in predicting temperature. Features including humidity, wind speed, precipitation, pressure, and air quality measurements were ranked according to their influence on temperature predictions.

d. Spatial Analysis

Spatial analysis was conducted using an interactive map created with Folium. Weather data from unique locations (identified by latitude and longitude) were plotted to visualize geographical patterns, displaying key weather details (e.g., temperature and conditions) for each location.

e. Geographical Patterns

Weather conditions were further explored by aggregating data at the country level. Average statistics for temperature, humidity, and precipitation were computed, and visual comparisons (such as bar charts) were generated to highlight differences across countries and continents.

7. Conclusion and Insights

The project successfully cleaned and analyzed the global weather data, building robust forecasting models using both ARIMA and Prophet techniques. The ensemble forecast demonstrated improved accuracy over individual models. Advanced analyses revealed significant seasonal trends, environmental influences, and spatial patterns in the data. These insights provide valuable guidance for strategic planning and project management in line with the PM Accelerator mission.

8. Repository and Submission Details

The complete project, including code, visualizations, and this report, is available in the GitHub repository. The repository includes:

- A Jupyter Notebook (Advanced_Assessment.ipynb) with all code and visualizations.
- This report document.
- An interactive map (waether_map.html).
- A README.md with detailed project documentation and instructions.

GitHub Repository Link: https://github.com/avanthikarajesh30/PM_ACCELERATOR