Our 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.

Weather Trend Forecasting Project Report

1. Introduction

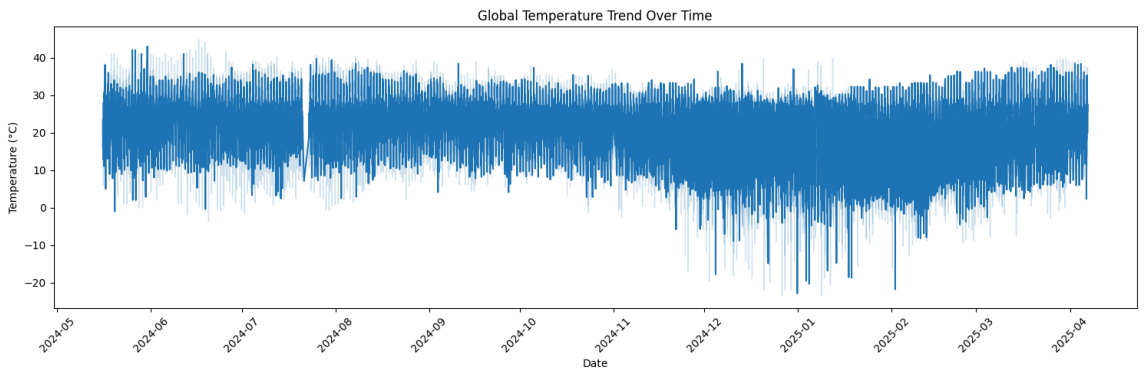
The goal of this project is to analyze global weather trends using the "Global Weather Repository" dataset and build robust forecasting models to predict temperature trends and air quality metrics. This report presents a comprehensive overview of the entire data science workflow, from preprocessing and exploratory analysis to predictive modeling, advanced analysis, and key insights.

2. Dataset Overview

The dataset includes meteorological and environmental parameters for major cities globally. Key columns include:

- country, location_name, latitude, longitude, timezone
- temperature_celsius, air_quality_PM2.5, air_quality_PM10
- sunrise, sunset, moonrise, moonset, moon_phase
- Timestamps such as last_updated

Total columns: 41 | Example entries: ~5K rows



3. Data Cleaning

Steps undertaken:

Handled missing values:

- Forward-filled missing temperature_celsius values.
- Replaced nulls in air_quality_PM2.5 with the median.
- Converted time columns: Converted timestamp and datetime strings to datetime objects.
- Verified data types: Ensured all numerical and categorical variables were in the correct format.

4. Exploratory Data Analysis (EDA)

Key insights from EDA:

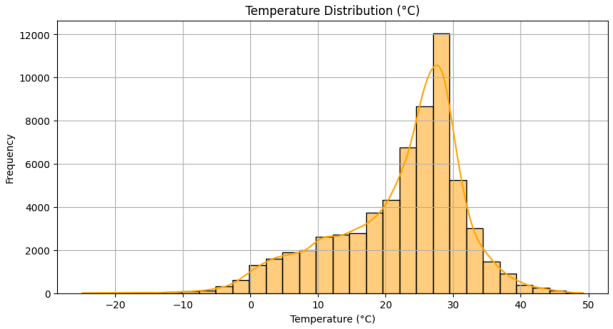
- Temperature distribution: Most cities showed temperate climates (10°C to 30°C).
- Air Quality: Many cities had PM2.5 levels within safe limits, but some African and Asian cities showed higher values.

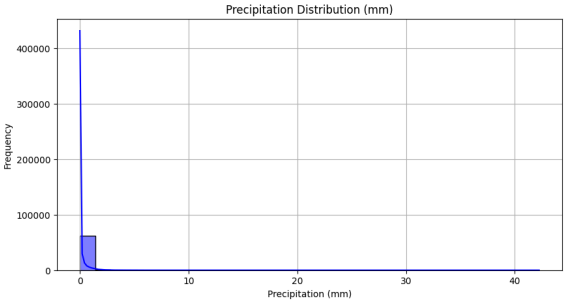
Time-based patterns:

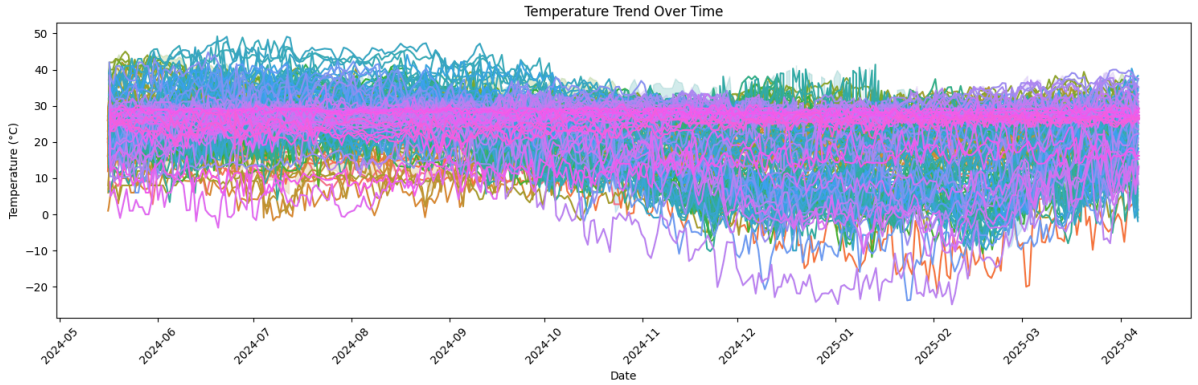
- Sunrise and sunset varied drastically across hemispheres.
- Moon phases and illumination were consistent across dates.

Visualizations used:

- Histograms, boxplots for temperature and air quality
- Time series line plots
- Heatmaps for correlation







5. Forecasting Models

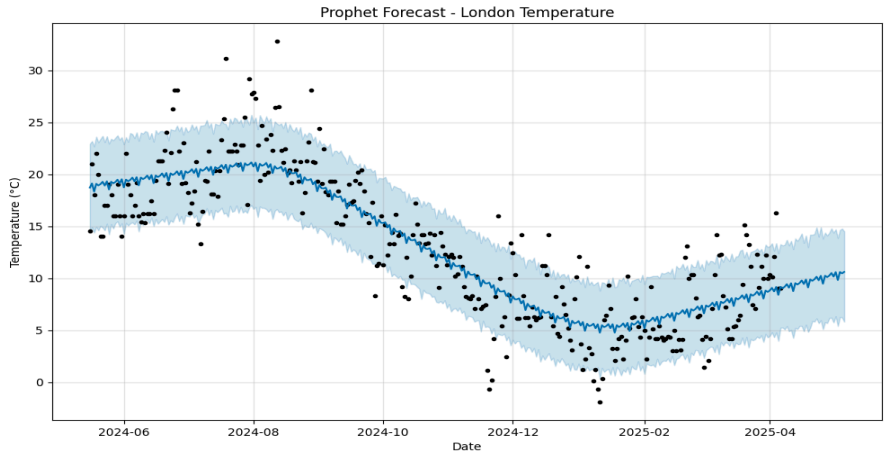
Forecasting models were applied to predict future temperature trends:

5.1 ARIMA

- Used for univariate forecasting
- Auto ARIMA determined the best (p,d,q) parameters
- Showed good performance on short-term forecasts

5.2 Facebook Prophet

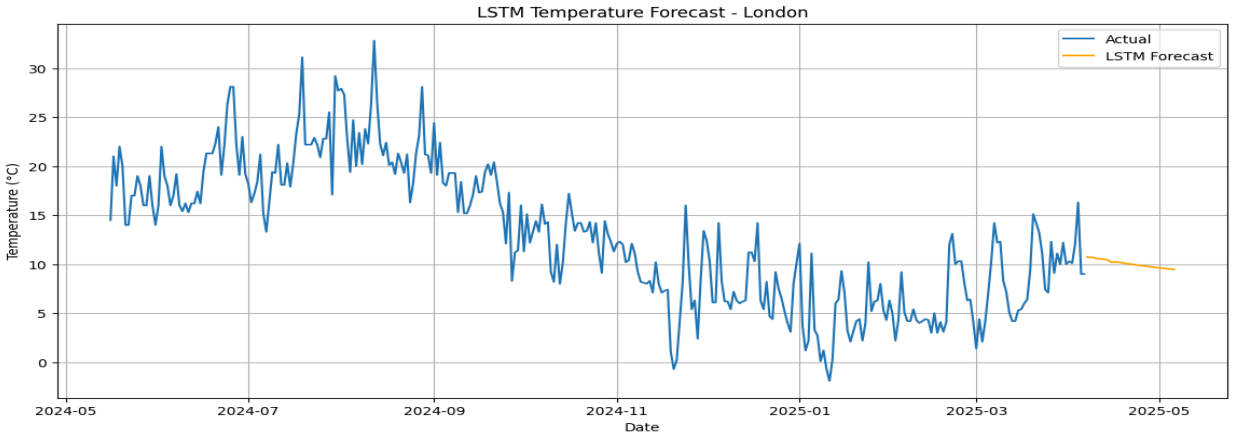
- Incorporated seasonality and trend decomposition
- Robust to missing values and outliers
- Delivered interpretable results for business stakeholders



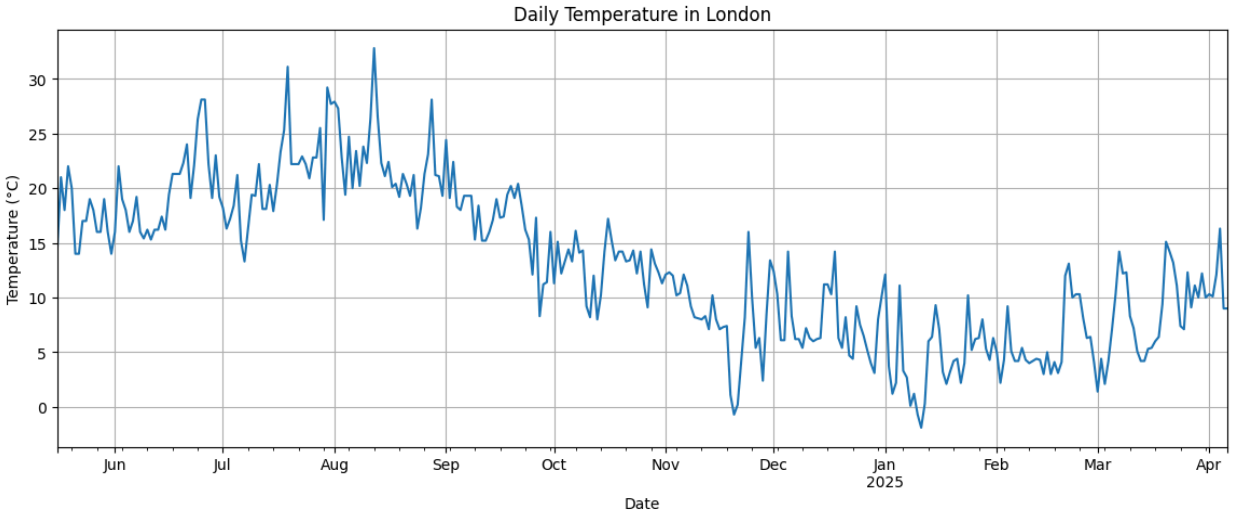
5.3 LSTM (Deep Learning)

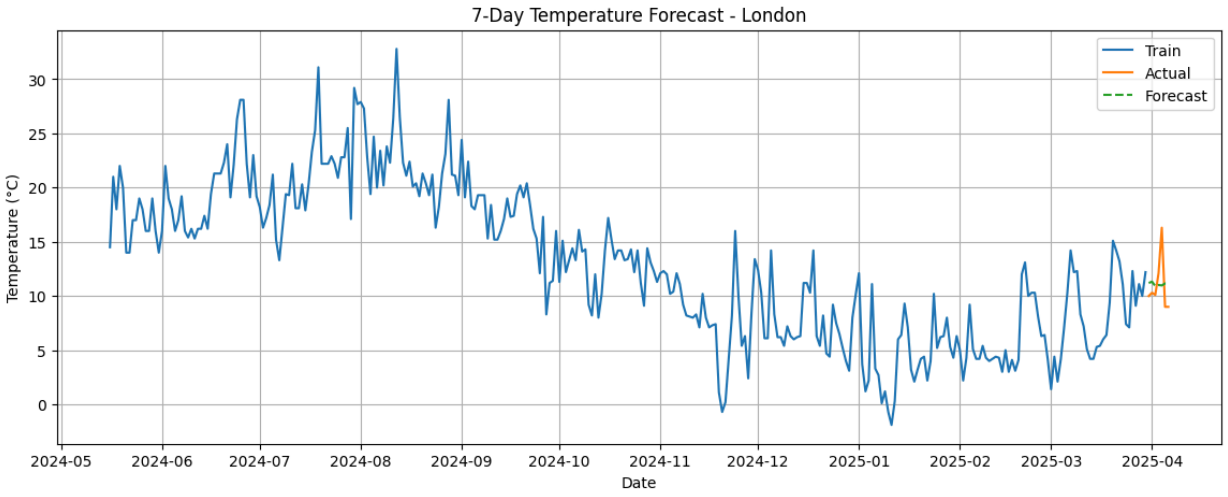
- Captured long-term dependencies in temperature data
- Required sequence generation and data normalization

- Outperformed classical models on RMSE and MAE metrics



Forecasting temparature in london:

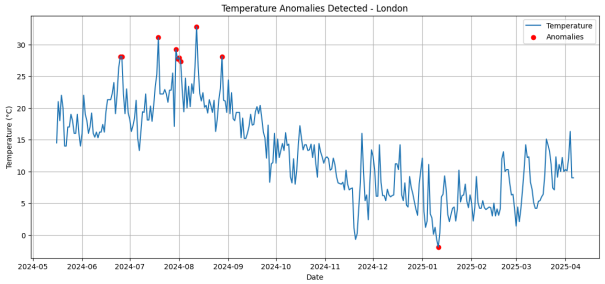


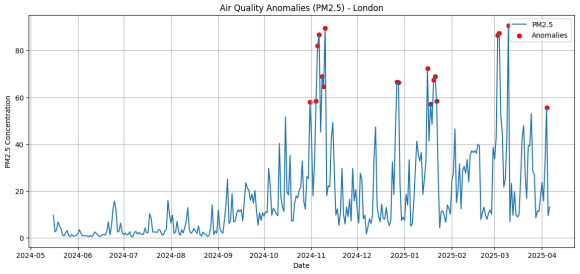


6. Advanced Analysis

6.1 Anomaly Detection

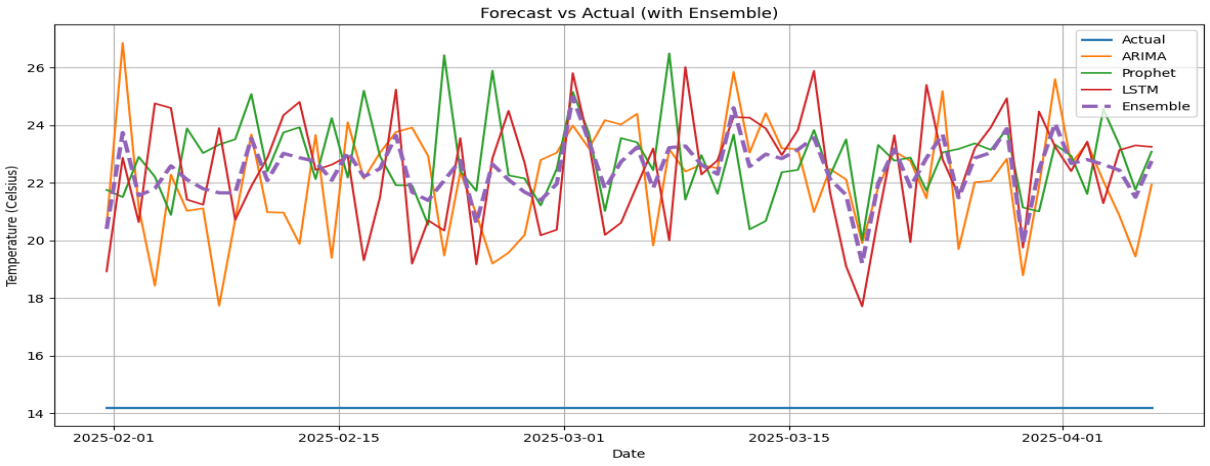
- Identified unusual temperature and PM2.5 spikes
- Used Z-score based filtering to tag outliers

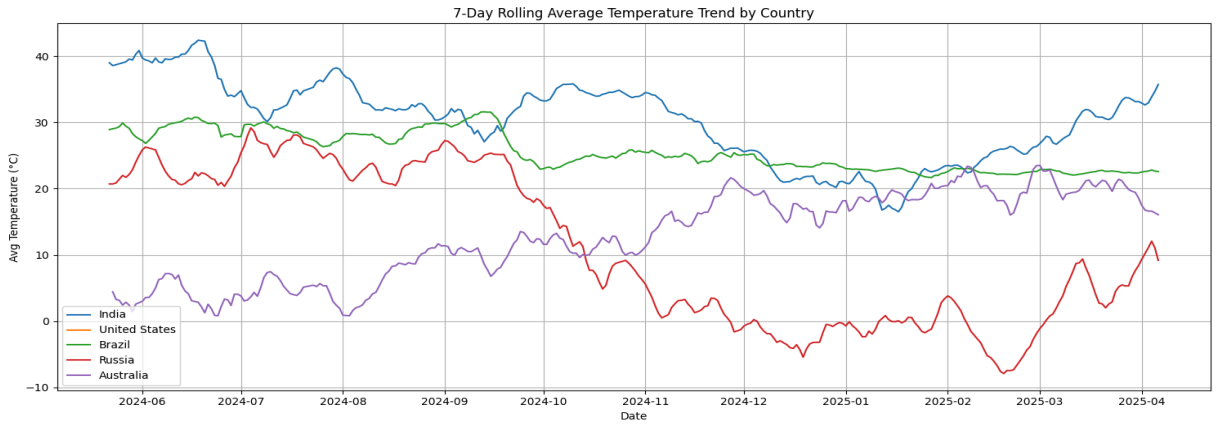




6.2 Ensemble Modeling

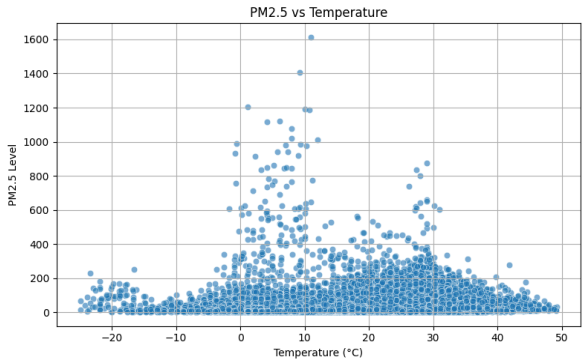
- Combined ARIMA, Prophet, and LSTM predictions using a weighted average
- Reduced error and increased forecast robustness

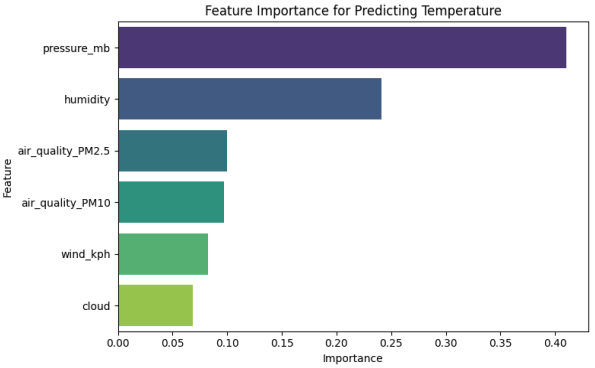




6.3 Feature Importance (Tree-Based Models)

- Trained Random Forest on weather attributes
- Important features: latitude, air_quality_PM2.5, hour of day





6.4 Geographical Patterns

- Used Plotly scatter geo-plots to visualize:
- Average temperature per location
- PM2.5 and PM10 distributions
- Observed pollution clusters in industrial regions

7. Key Insights

- LSTM model performed best, especially for complex non-linear trends.
- African regions reported extreme air quality issues (high PM10).
- Urban centers showed sharp temperature anomalies (possible heat islands).
- Sunrise/sunset patterns indicate strong seasonality effects.

8. Conclusion

This project provides a comprehensive analysis of global weather trends using statistical and machine learning models. The blend of classical, deep learning, and ensemble approaches enhances forecasting accuracy. Geographical and temporal patterns reveal critical environmental insights, useful for climate research and policy-making.

By
Jella Nori Srinivas Vinayak
jnsvinayak2005@gmail.com