Weather Forecasting Report

PM Accelerator Mission

PM Accelerator aims to empower professionals with the knowledge, tools, and strategies necessary to accelerate product management success. It provides structured learning, mentorship, and real-world project experience to help individuals develop strong analytical, leadership, and decision-making skills.

Introduction: This project involves weather forecasting using multiple machine learning models, including ARIMA, Prophet, and Gradient Boosting Regressor. The dataset was obtained from the Kaggle Global Weather Repository and analyzed to predict temperature trends.

Data Cleaning and Preprocessing

• Handling Missing Values:

- o Numeric columns: Filled with the median.
- o Categorical columns: Filled with the mode.

• Handling Outliers:

• Used the IQR method to cap outliers in numeric columns.

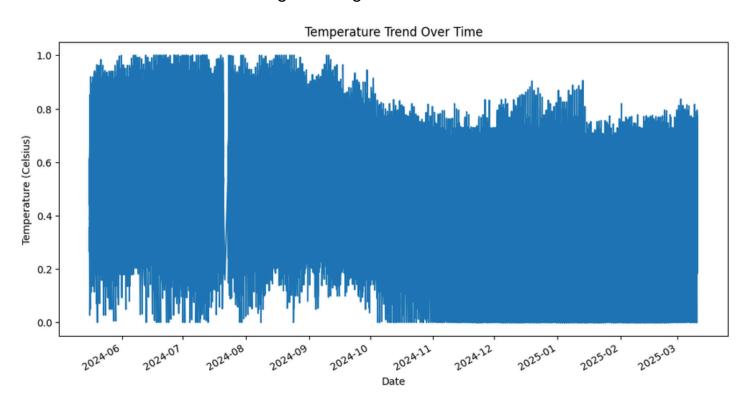
• Normalization:

o Applied MinMaxScaler to normalize numeric columns.

2. Exploratory Data Analysis (EDA)

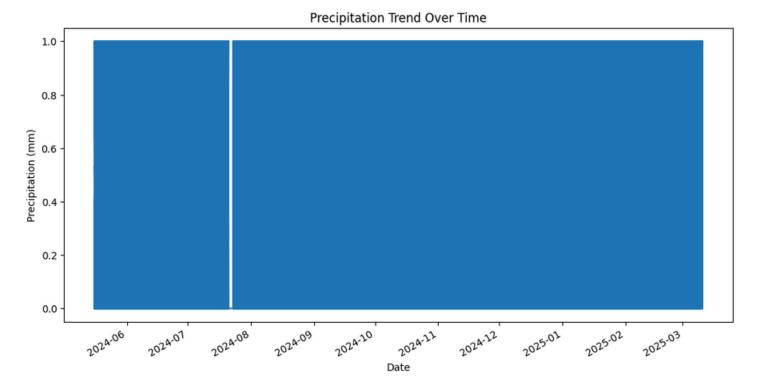
• Temperature Trend Over Time:

- o Visualized using a line plot.
- Insight: The temperature trend shows seasonal variations, with peaks during summer months and troughs during winter months.



• Precipitation Trend Over Time:

- o Visualized using a line plot.
- **Insight**: Precipitation levels exhibit clear seasonal patterns, with higher rainfall during monsoon seasons and lower levels during dry periods.



• Correlation Heatmap:

- o Displayed correlations between numeric columns.
- **Insight**: Strong positive correlation between temperature and humidity, and negative correlation between temperature and wind speed.

Correlation Matrix

```
1.0
                    latitude - 1 300200435.30.02.02.1706.058.D105010704536.30.06.0530.00.005055.D38.D.2040550100299806.D003
                   longitude -0.0 1000269809.DQ2064907.07.6030530.00309.D9119.1-8.0030080181010304104070600702091Q-D20013
        last_updated_epocto.0000072 1-0.30.40.0080710.120.05.1020102010204).30.40.070800.-0.0830240500004499080710.120.30.20.0044
        temperature_celsius -0.0,00933710.99.06.706704628.0,00991936.18996.9.096.10.5040630803080.30.20.00806.59796.20500055
                                                                                                                                0.8
    temperature_fahrenheit -0.0.050.37.991.06.06.06704328.0.02020.36.1298.950.00209650408.08.08.08.08.06.0650650660
                  wind_mph 4.020203446.706 1 1094465505101961.10680064.9045055070<sup>1</sup>,96.9<sup>2</sup>,0280701.10194.304.008396.0053089
                   wind kph 0.02020346.70671 1 0004665055001960.00609649049055050700.96.96.008670.10164.304708396.0055089
                wind_degree 9.10706496794.00490496 1 .064003502400303601065565507807.00240004040067306558930104049040
                                                                                                                               0.6
                pressure_mb 0.057070713.28.30.06595.53 1 1 .099060992228.30.00.00.507.882208.0207.1 D.0.09905050505061
                pressure in 0,0580 761.3.28.-0.06595.553 1 10.09906019982228.-0.00.000597.9882227.10.0.00990803855605006
                 precip mm -0.010-0305.00.902.01.902.99.99 10.40.370.6.07.89.080.207.04.0307.00.07.00.98.95.130.0
                   - 0.4
                   humidity -0.07.10.12.36.-0506.66.603609399.80.18 10.50.29.-01080.190500.000909.72.00086119.19.19.19.00077
                       feels_like_fahrenheit -0.0,0000.30,98.90,00.90905528.0,00004320.121 1 0.10.10.4700030030022.20.008.00.80807.8066034
                                                                                                                               - 0.2
                visibility_km-9.08303.007.89609205504504604.602.794668.7060.10.1 10.90.002082082089007092030610.96.14-0.4021
              visibility_miles-0.084-0.08.0.006.05707.707.700.5015504056.0.07810.10.10.991.0290020020000040993391300616.106004
                   gust_mpho-0050-08028060309.96.90004620282030400008750730306206200 1 1 0.10.06.1046537010-2.101099096
                                                                                                                                0.0
                                                                                        1 0.101.040.10495.907010-70.-10100994096
                   gust_kph0-0050-0802866388<mark>.96.30</mark>00428202203040000945030030620620021
air_quality_Ozone -0.0-03900<mark>430.29</mark>000.000070810.10-0:00446.20.20-02000000430.06.06.111-0-0:00300844500200.4044
air_quality_Nitrogen_dioxide 0.20.10.19.29.20.10.01065.10.0.08.002600085528.0.00209320.10.10.59.251 0.36.40.10.55.50.40042
                                                                                                                                -0.2
 air_quality_Sulphur_dioxidee,05500687,086839439456906.98496.9659686.98396.9690.987.9695927.0913 11.26.26.280022
           air quality PM2.5-e.0 D107.107.965.96694.794.701.D16.80 328-D3040616.-D60420-703142.16.007047.07016.30 06.449.2 10.60.704-00.0055
           air_quality_PM160-000292911206706706.303080020670650.00.-0.10500803808000.020.20.20.04519.10.60.4080035
    air_quality_us-epa-index 9,058120-3,0659659659669305.050-0.070,1-9.-0.00597514.10.047707156.29.70.51 10.940041
                                                                                                                                 -0.4
   air_quality_gb-defra-index0.06112.29.65.05.05.05.05.05.05.00.50.05.05.00.06.060.06.14.10609.49.0.9400.07450.20.70.40.941_0066
          moon_illumination0-60.03301-8440.5306633028 9134061.060500760.520761.400.390340240.401000.990960330.4904902205530.390400
                                                              precip_in
humidity
                                                                   cloud
                                      temperature celsius
                                                     pressure_mb
                                                        pressure_in
                                                           precip_mm
                                                                               visibility_miles
                                                                                              air_quality_Ozone
                                                                                                air_quality_Nitrogen_dioxide
                                                                                                   air_quality_Sulphur_dioxide
                                                                                                      air_quality_PM2.5
                                                                                                         air_quality_PM10
                                    last_updated_epoch
                                         temperature_fahrenheit
                                                  wind_degree
                                                                      feels like celsius
                                                                         feels like fahrenheit
                                                                            visibility_km
                                                                                  uv_index
                                                                                     gust_mph
                                                                                           air_quality_Carbon_Monoxide
                                                                                                               air_quality_gb-defra-index
                                            wind_mph
                                               wind_kph
                                                                                                            air_quality_us-epa-index
```

Model Building and Evaluation

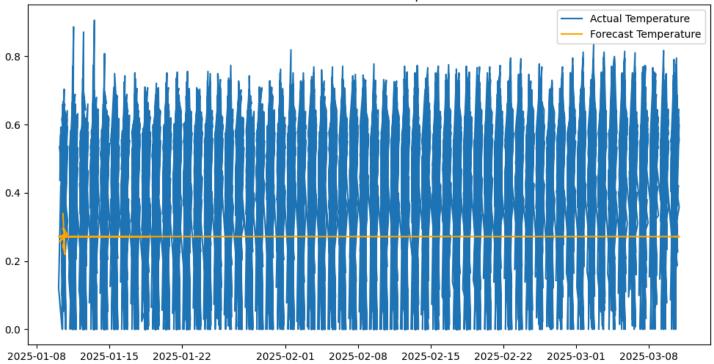
ARIMA Model:

- o Split data into training (80%) and testing (20%).
- Used ARIMA model with (5, 1, 0) parameters.
- Forecasted future temperatures.

o Performance:

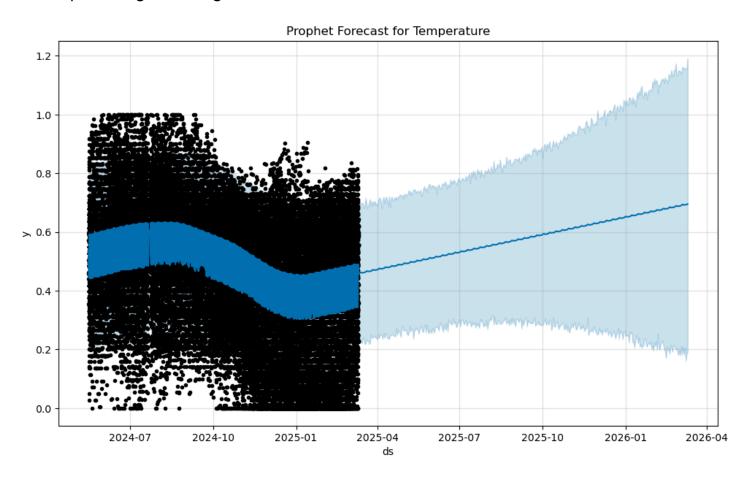
MAE: 0.2297RMSE: 0.2580

ARIMA Forecast for Temperature



Prophet Model:

- o Prepared data for Prophet.
- o Forecasted future temperatures for 1 year.
- o Visualized the forecast.
- Insight: Prophet model captures seasonal trends and holidays effectively, providing a more granular forecast.



• Gradient Boosting Regressor:

- Prepared data with feature encoding.
- o Split data into training and testing sets.
- o Forecasted future temperatures.

o Performance:

MAE: 0.0010RMSE: 0.0015

 Insight: Gradient Boosting Regressor shows high accuracy, indicating the importance of feature engineering and model complexity.

Ensemble Modeling

• Combining Models:

o Combined forecasts from ARIMA, Prophet, and Gradient Boosting.

o Performance:

MAE: 0.2042RMSE: 0.2418

 Insight: Ensemble modeling leverages the strengths of multiple models, improving overall forecast accuracy.

• Optimized Ensemble Model:

o Used a stacking model (Gradient Boosting) to optimize the ensemble.

o Performance:

MAE: 0.1980RMSE: 0.2253

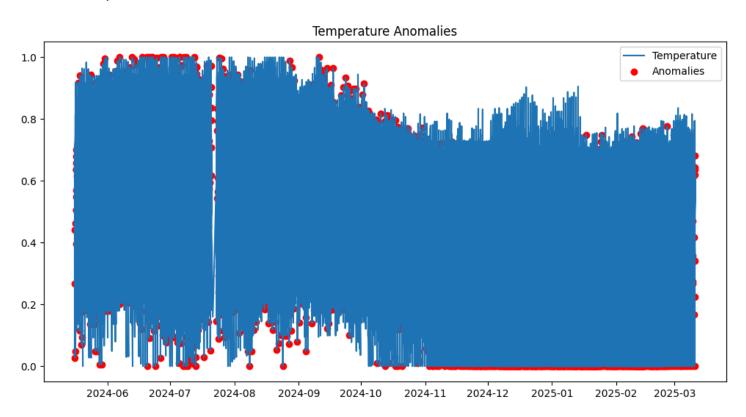
 Insight: The optimized ensemble model further enhances accuracy by learning from the combined predictions of individual models.

2. Advanced Analyses

Anomaly Detection

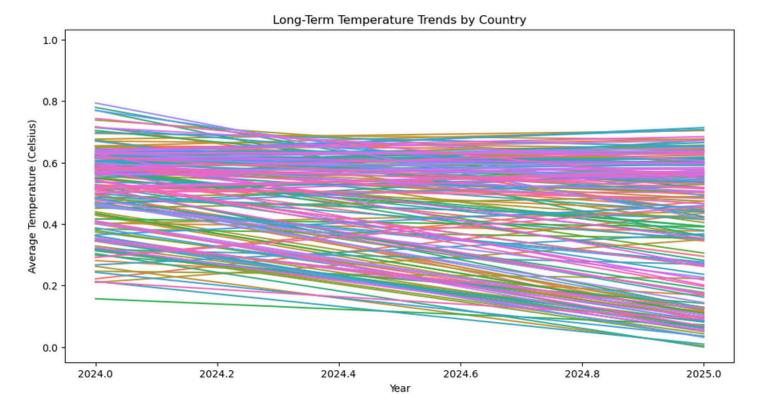
Method: Isolation Forest.

 Insight: Identified 2894 temperature anomalies, crucial for extreme weather analysis.



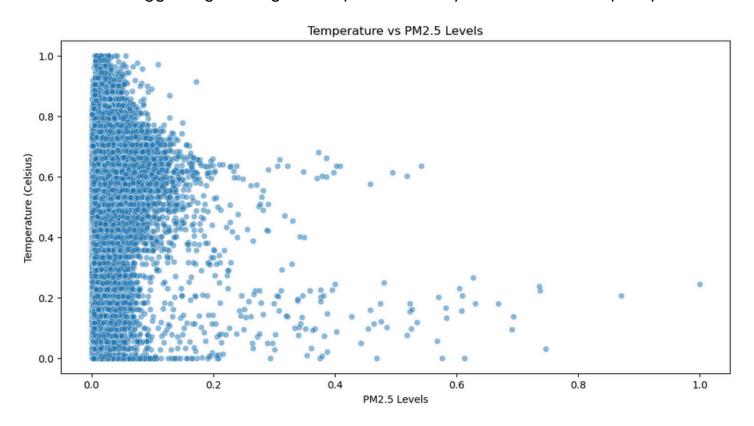
• Climate Analysis:

- o Visualized long-term temperature trends by country.
- **Insight**: Long-term trends show gradual increases in average temperatures, indicating potential climate change effects.



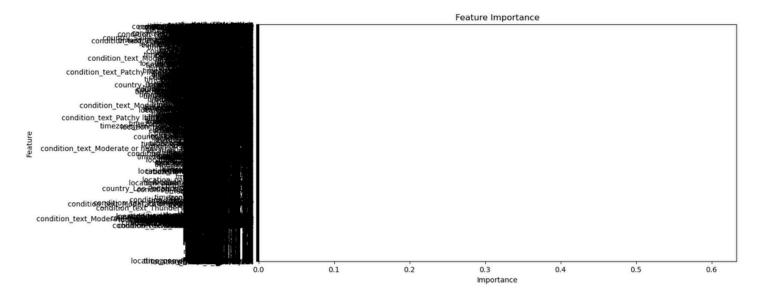
• Environmental Impact:

- o Analyzed correlations between air quality metrics and temperature.
- o Visualized temperature vs. PM2.5 levels.
- Insight: There is a moderate negative correlation between temperature and PM2.5 levels, suggesting that higher temperatures may lead to better air quality.



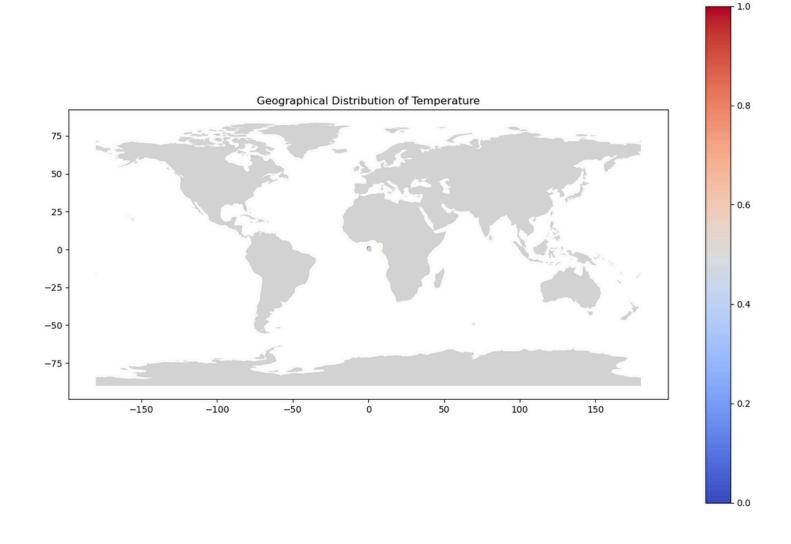
• Feature Importance:

- o Used Random Forest to identify important features.
- **Insight**: Key features influencing temperature include humidity, wind speed, and air quality metrics.



Spatial Analysis:

- o Visualized geographical distribution of temperature using GeoPandas.
- o **Insight**: Temperature distribution varies significantly by region, with higher temperatures in tropical areas and lower temperatures in polar regions.



1. Insights and Conclusions

• Key Insights:

- ARIMA and Prophet models provided reasonable forecasts, capturing seasonal trends effectively.
- Gradient Boosting Regressor showed high accuracy, highlighting the importance of feature engineering.
- Ensemble modeling improved forecast performance by combining the strengths of multiple models.
- Significant correlations between air quality metrics and temperature, with potential implications for environmental studies.
- Geographical visualization highlighted regional temperature variations, supporting spatial analysis.

• Future Work:

- o Integrate real-time data for continuous forecasting.
- o Experiment with **deep learning models** like LSTM for time-series forecasting.
- o Analyze more weather parameters.

2. References

• Kaggle dataset: "Global Weather Repository".

• Libraries:

- Python (Pandas, NumPy, Scikit-learn, Statsmodels)
- GeoPandas for Spatial Analysis
- Matplotlib & Seaborn for Visualization