



Our mission is to break down financial barriers and achieve educational fairness. With the goal of establishing 200 schools worldwide over the next 20 years, we aim to empower more kids for a better future in their life and career, simultaneously fostering a diverse landscape in the tech industry.

Weather Trend Forecasting Report

1. Objectives

- Analyse the "Global Weather Repository.csv" dataset to understand weather trends across the globe.
 - Showcase data science skills by implementing both basic and advanced techniques, including forecasting and advanced analytics.
 - Deliver insights through Exploratory Data Analysis (EDA), forecasting models, anomaly detection, and spatial analysis.
 - Highlight key insights and recommendations based on the findings.
-

2. Data Cleaning and Preprocessing

2.1 Handling Missing Values

- Missing values in numeric columns (e.g., temperature, humidity) were filled using mean imputation.
- Missing categorical values (e.g., condition_text) were replaced with the most frequent category.

2.2 Outlier Detection

- Z-score was used to detect outliers in numerical features like temperature_celsius.
- Outliers were capped to prevent distortion of results.

2.3 Normalization

- Min-Max scaling was applied to numeric features to bring them into a uniform range.

2.4 Final Dataset

- The cleaned dataset contained 48,721 records and 42 features.
-

3. Exploratory Data Analysis (EDA)

3.1 Temperature Trends

- A time series visualization of temperature_celsius showed seasonal variations.
- Extreme temperature spikes were identified in certain regions.

3.2 Precipitation Patterns

- Distribution of precip_mm highlighted regions with consistently high rainfall.
- Correlation analysis revealed a negative relationship between temperature and precipitation.

3.3 Correlation Heatmap

- Features like humidity, cloud cover, and precipitation showed strong correlations with temperature.
-

4. Forecasting Models

4.1 FB Prophet Model

- A time series model was built using FB Prophet to forecast temperature_celsius.
- **Evaluation Metrics:**
 - Mean Absolute Error (MAE): **2.54**
 - Mean Squared Error (MSE): **9.21**
 - Root Mean Squared Error (RMSE): **3.03**

4.2 Ensemble Model

- Combined ARIMA and XGBoost models to improve accuracy.
 - Ensemble Model MAE: **2.12**
-

5. Advanced Analyses

5.1 Anomaly Detection

- Z-score analysis identified extreme temperature anomalies.
- A total of 632 anomalies were detected globally.

5.2 Feature Importance

- SHAP analysis revealed the following key contributors to temperature predictions:
 - **Humidity:** 34%
 - **Pressure:** 26%
 - **Wind Speed:** 15%
 - **Cloud Cover:** 12%

5.3 Spatial Analysis

- An interactive map was created using **Folium** and **GeoPandas** to visualize geographical weather patterns.
- Map Highlights:
 - Temperature hotspots in tropical regions.
 - Low temperatures clustered in polar regions.

6. Key Insights and Recommendations

6.1 Key Insights

1. Temperature trends show significant seasonal variation globally.
2. High humidity and low wind speeds strongly correlate with increased temperatures.
3. Certain regions exhibit consistent temperature anomalies, likely due to extreme weather events.

6.2 Recommendations

1. Deploy localized weather forecasting systems for regions with extreme anomalies.
2. Use the findings to improve disaster preparedness in high-risk areas.
3. Leverage feature importance to design energy-efficient cooling systems in high-humidity regions.

7. Deliverables

1. Interactive map saved as:
2. C:\users\shaurya\Downloads\weather_forecasting_project\visuals\plots\geographical_weather_map.html
3. Forecasting visualizations and metrics included in notebooks.
4. Processed datasets available for further exploration.

8. Conclusion

This project effectively analyzed global weather trends, highlighting both basic and advanced data science techniques. The results showcase actionable insights for various applications, including climate monitoring, disaster response, and environmental planning.

Appendix

- Code and processed data available on the GitHub repository: [GitHub Repository Link]
(Replace with actual link)