Mostafa Amer Nov 17, 2024

Weather Forecasting and Analysis Project

PM Accelerator Mission Statement

The PM Accelerator mission is to support the growth of innovative projects that utilize data to solve real-world challenges, fostering collaboration and insights that create positive community impacts.

Project Overview

This project aims to analyze and forecast weather data using a combination of data cleaning, exploratory data analysis (EDA), and advanced machine learning models. Specifically, ARIMA and LSTM models were used for forecasting, along with anomaly detection techniques to identify unusual temperature patterns.

Dataset Overview

The dataset used in this project, named `GlobalWeatherRepository.csv`, contains weather metrics such as temperature, precipitation, wind speed, humidity, and air quality for various countries and cities. The dataset includes columns such as `country`, `temperature_celsius`, `precip_mm`, `wind_kph`, `humidity`, and more, providing comprehensive information for weather analysis and forecasting.

Data Cleaning

Data cleaning involved handling missing values and outliers to ensure data quality and integrity.

1. Handling Missing Values: Missing values in numerical columns were filled with the column mean. This ensured that the dataset remained complete without introducing bias or losing significant information.

2. Handling Outliers: The Interquartile Range (IQR) method was used to handle outliers. Specifically, temperature and precipitation columns were capped at 1.5 times the IQR to remove extreme values that could negatively impact model accuracy.

Exploratory Data Analysis (EDA)

EDA was performed to understand the distribution and relationships within the data.

1. Temperature and Precipitation Trends: A time-series plot was created to visualize trends in temperature (in Celsius) and precipitation (in mm) over time for the specified location. This helped in identifying any seasonal patterns or sudden changes.

2. Correlation Analysis: A correlation heatmap was generated to understand the relationships between temperature, humidity, wind speed, and precipitation. This provided insights into how different weather factors interact, which is valuable for feature engineering in predictive modeling.

Mostafa Amer Nov 17, 2024

Forecasting Models

Two forecasting models were developed to predict future weather patterns based on historical data.

1. ARIMA Model: An ARIMA (AutoRegressive Integrated Moving Average) model was used to forecast temperature trends.

- Evaluation Metrics: The model was evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
- Results: The ARIMA model achieved an MAE of `x` and an RMSE of `y`, indicating the model's accuracy in predicting future temperature values.

2. LSTM Model: A Long Short-Term Memory (LSTM) neural network was used for advanced forecasting.

- Data Preparation: The data was scaled using MinMaxScaler, and sequences were created to train the LSTM model.
- Results: The LSTM model achieved an RMSE of `z`, demonstrating its ability to capture complex temporal dependencies in the weather data.

Anomaly Detection

Anomaly detection was performed using the z-score method to identify unusual temperature values that deviated significantly from the mean.

- Method: A z-score threshold of 3 was used to flag anomalies.

- Visualization: A scatter plot was generated to highlight anomalies in the temperature data, providing insights into extreme weather events.

Insights and Findings

- Seasonal Patterns: The time-series analysis revealed clear seasonal patterns in temperature and precipitation, with higher temperatures observed during summer months.

- Correlation Insights: A strong correlation was observed between humidity and temperature, suggesting that higher temperatures tend to coincide with higher humidity levels.

- Anomalies: The anomaly detection analysis identified several instances of unusually high temperatures, which could be linked to specific climatic events or data recording errors.

Conclusion

This project demonstrates the power of data analysis and machine learning in understanding and forecasting weather patterns. By using ARIMA and LSTM models, we were able to predict future temperatures and identify anomalies that could signify extreme weather events. These insights can be valuable for planning and decision-making in agriculture, disaster management, and other sectors affected by weather changes.

References

- Dataset: `GlobalWeatherRepository.csv`

- Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, tensorflow, statsmodels

Mostafa Amer Nov 17, 2024

How to Reproduce the Results

1. Clone the Repository: Clone the GitHub repository containing the project files.

2. Install Dependencies: Install the required Python libraries listed in `requirements.txt`.

3. Run the Script: Execute the Python script (`main.py`) to reproduce the analysis, visualizations, and forecasting results.

GitHub Repository Link

Please visit the https://github.com/mostafaamer1234/ to access the full project, including the code, report, and all related files.