Report on Predicting the Next Flood in Lagos State

Objective

The primary goal of this project is to predict the next occurrence of floods in Lagos based on weather data and historical flood records. This predictive model will help in early warning systems and disaster management planning. By leveraging the insights gained from the Exploratory Data Analysis (EDA), including the observed humidity and rainfall patterns, the model aims to provide accurate and timely predictions that can aid in effective response strategies to minimize the impact of floods on the community.

Data Description

Weather Data

The weather data is collected from various sources and includes daily records from 2002 to 2024. The key features include:

- datetime: Date of the record
- tempmax: Maximum temperature
- tempmin: Minimum temperature
- temp: Average temperature
- dew: Dew point
- humidity: Humidity level
- precip: Precipitation level
- precipprob: Probability of precipitation
- precipcover: Precipitation cover
- windspeed: Wind speed
- winddir: Wind direction
- sealevelpressure: Sea level pressure
- cloudcover: Cloud cover
- visibility: Visibility
- solarradiation: Solar radiation

- solarenergy: Solar energy
- uvindex: UV index
- moonphase: Moon phase

link: https://shorturl.at/kgOra

Flood History Data

link: https://shorturl.at/CJ3QE

Exploratory Data Analysis (EDA)

EDA is a critical step in understanding the dataset and identifying patterns that can help in building a predictive model. Several plots were created to visualize the datasets and gain insights. The following steps were carried out:

1. Data Cleaning and Preprocessing

Missing Values: Missing values in the dataset were identified and handled appropriately. Imputation techniques were used for continuous variables, while categorical variables were filled using the most frequent value.

Date Parsing: The 'Date' column was parsed to datetime format to facilitate time series analysis.

Data Merging: Weather data and flood records were merged on the 'Date' column to create a unified dataset for analysis. The flood column was created as an Indicator of flood occurrence (1 for flood, 0 for no flood) from the flood history dataset.

2. Rainfall Analysis

Descriptive Statistics: The average, median, and maximum rainfall values were calculated to understand the overall distribution.

Trend Analysis: This chart provides a visual representation of the variability and intensity of rainfall in Lagos, Nigeria. Understanding these trends is crucial for flood prediction and disaster

management. By analyzing the patterns and extreme events, stakeholders can make informed decisions to mitigate the impacts of heavy rainfall and prepare for potential flood risks.

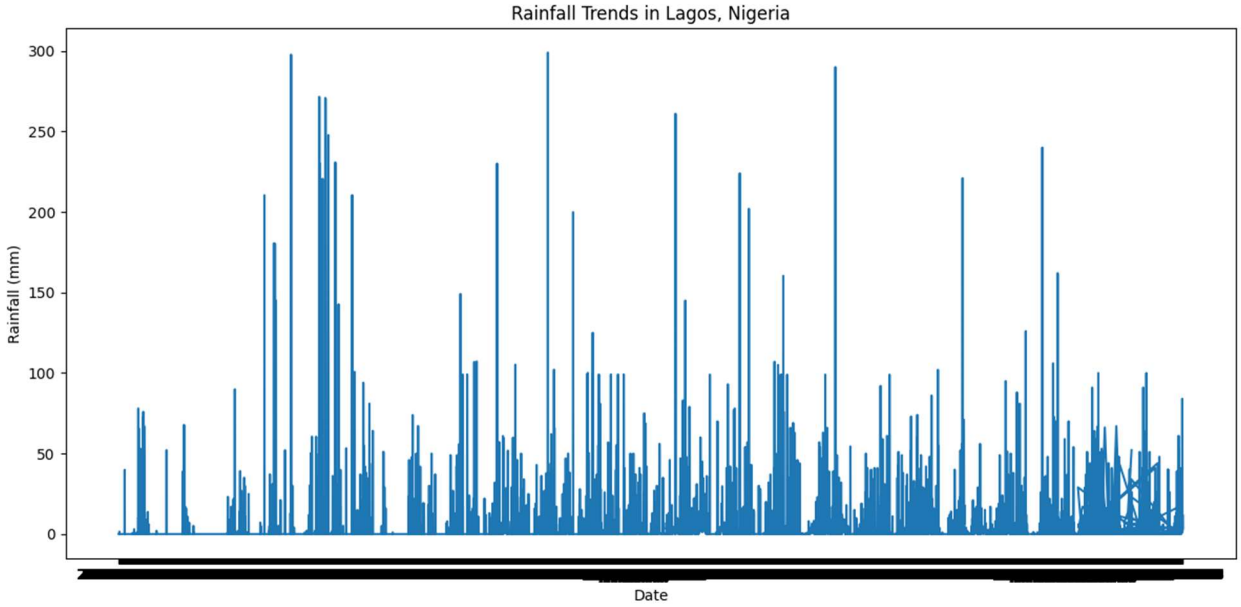


Figure 1 plot for rainfall in lagos

Monthly rainfall analysis: The chart provides valuable insights into the seasonal rainfall patterns in Lagos, Nigeria. Understanding these patterns is essential for flood prediction, water resource management, agriculture planning, and preparing for weather-related challenges. The clear identification of wet and dry seasons helps in making informed decisions and implementing effective measures to mitigate the adverse effects of extreme weather conditions.

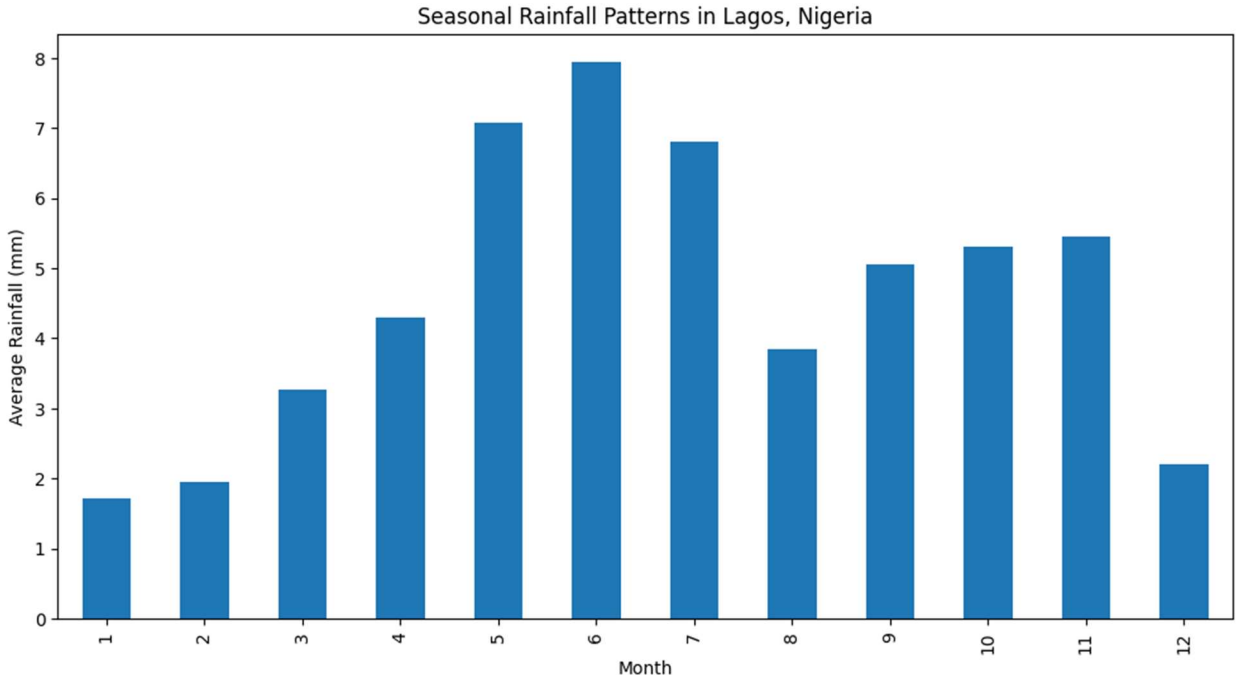


Figure 2 monthly plot for average rainfall

Rainfall Distribution: A histogram of daily rainfall was created to visualize the distribution and identify any skewness.

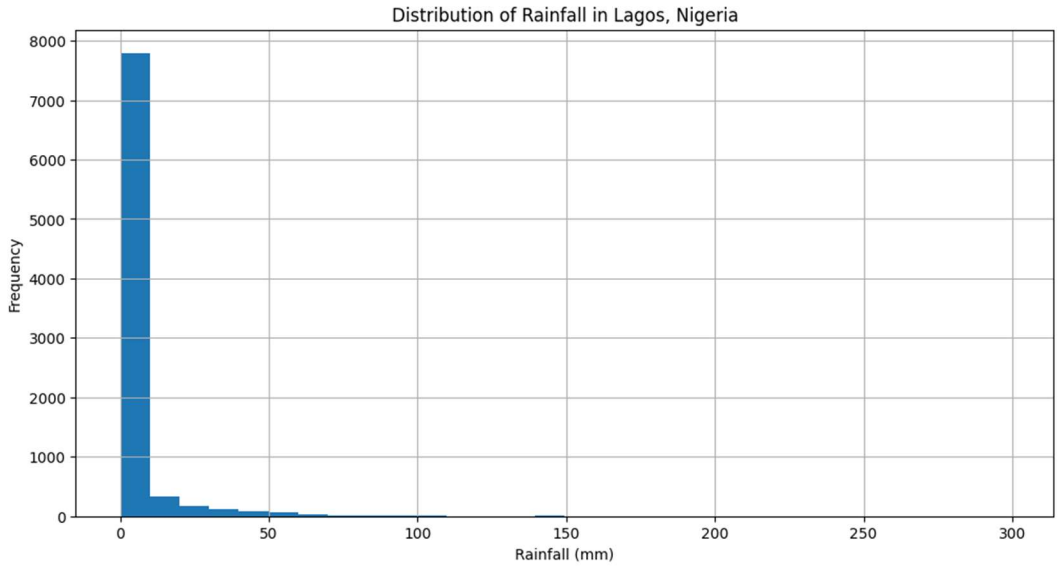


Figure 3 histogram plot for rainfall distibrution.

3. Temperature and Humidity Analysis

Temperature: The daily temperature was analyzed to identify any correlation with rainfall and flood occurrence. A scatter plot was used to visualize the relationship between temperature and rainfall.

Humidity: Humidity levels were studied to understand their impact on flood prediction. Higher humidity levels often accompany heavy rainfall, increasing the likelihood of flooding. A plot of humidity for the days was also created.

The chart provides a visual representation of the humidity levels in Lagos, Nigeria, over time. The consistently high humidity, coupled with noticeable daily fluctuations, suggests a humid climate. Understanding these trends is crucial for flood prediction. By analyzing the humidity patterns alongside other weather variables like rainfall enables prediction for potential flood risks.

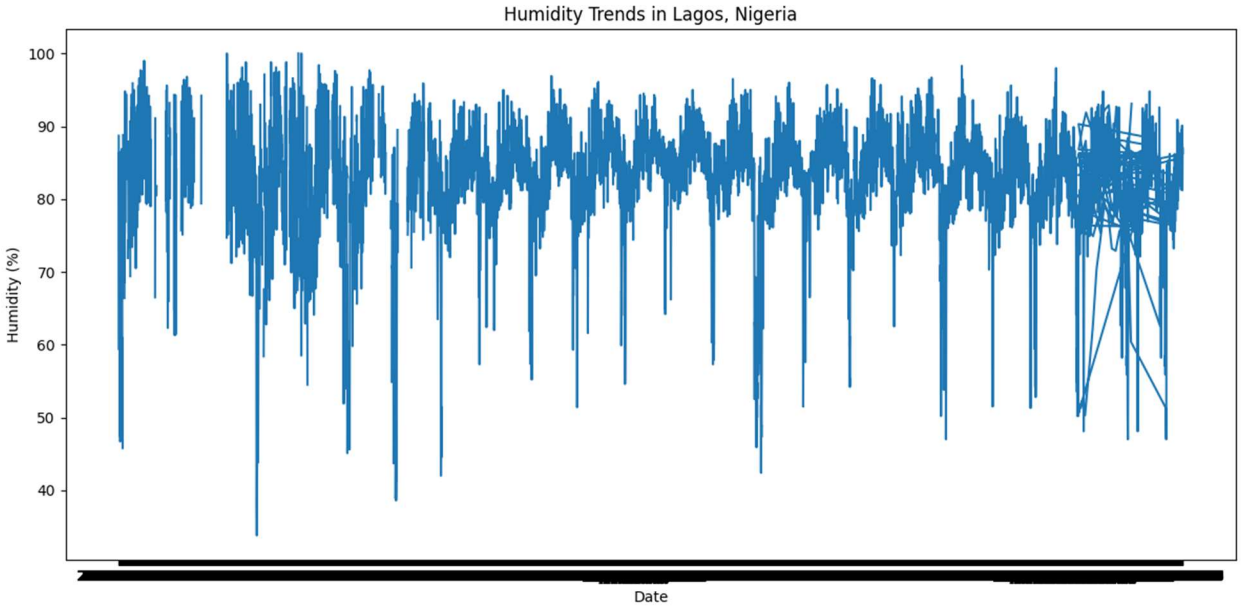
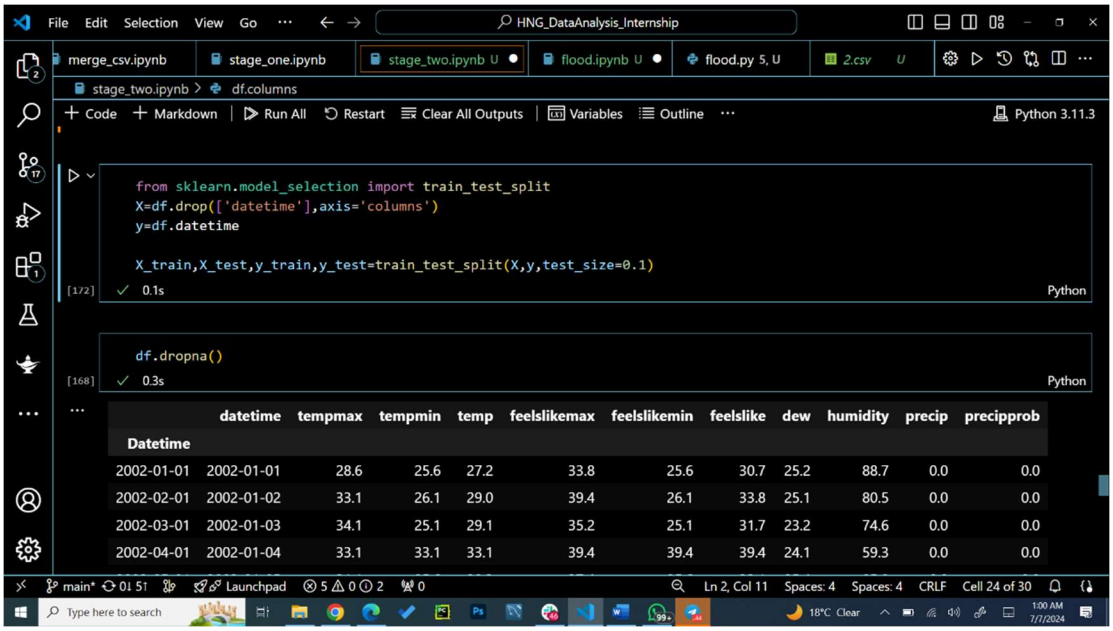


Figure 5 Plot for Humidity trend in the state

Flood Risk Prediction

Predictive Modeling

The next step involved building a predictive model to forecast flood occurrence based on the weather data.



Datetime	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity	precip	precipprob
2002-01-01	2002-01-01	28.6	25.6	27.2	33.8	25.6	30.7	25.2	88.7	0.0	0.0
2002-02-01	2002-01-02	33.1	26.1	29.0	39.4	26.1	33.8	25.1	80.5	0.0	0.0
2002-03-01	2002-01-03	34.1	25.1	29.1	35.2	25.1	31.7	23.2	74.6	0.0	0.0
2002-04-01	2002-01-04	33.1	33.1	33.1	39.4	39.4	39.4	24.1	59.3	0.0	0.0

1. Feature Selection

Key Features: The most relevant features for flood prediction were identified. These included rainfall, temperature and humidity.

2. Model Selection and Training

Model Choice: A Random Forest Classifier was selected due to its robustness and ability to handle complex interactions between features.

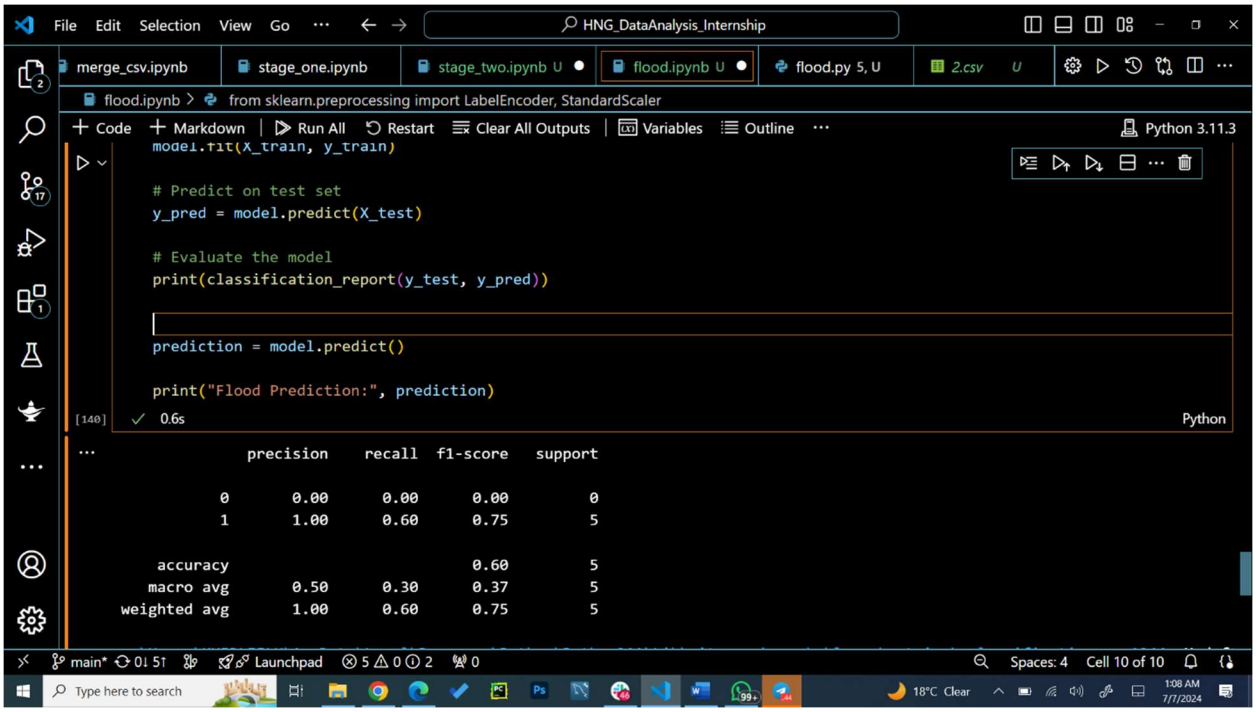
Train-Test Split: The dataset was split into training and testing sets to evaluate the model's performance.

Model Training: The model was trained on the training set using the selected features.

3. Model Evaluation

Prediction Accuracy: The model's accuracy in predicting floods was assessed using the test set.

Classification Report: Precision, recall, and F1-score metrics were calculated to evaluate the model's performance.



Conclusion

The flood prediction model indicates potential flooding on July 8-10, based on forecast data for the next two weeks. You can view the forecast at this link:https://shorturl.at/Rea4k. The model shows reasonable accuracy in identifying days with high flood potential. Further improvements can be made by incorporating additional weather variables and using more advanced modeling techniques. Continuous monitoring and updating of the model with new data will also enhance its predictive capability. The model would provide more precise predictions with the availability of the right dataset.

Recommendations

- Preparedness: Lagos residents and authorities in charge should be prepared for potential flooding during the identified high-risk periods, especially July 8-10 as predicted by the model.
- Continuous Monitoring: Continuous monitoring of weather conditions and updating the flood prediction model will be crucial to refine the forecasts and improve preparedness efforts.
- Data Quality: The provision of the right dataset for a project like this is crucial as the accuracy of the prediction heavily relies on the data.

By conducting a thorough EDA and developing a robust flood prediction model, this report provides a comprehensive analysis of the weather conditions in Lagos, Nigeria, and identifies the key factors contributing to flood risk.