LAGOS STATE FLOOD PREDICTION ANALYSIS REPORT

July 7th 2024

HNG Internship Stage two project data analysis

INTRODUCTION

Flooding is a significant natural disaster that poses severe risks to urban areas, particularly in coastal cities like Lagos. The impacts of floods can be devastating, leading to loss of life, destruction of infrastructure, and significant economic losses. Accurate prediction of flood events is crucial for effective disaster management and mitigation strategies. This report presents an analytical approach to forecasting future flood events in Lagos using historical weather data spanning from 2019 to 2024.

In this study, we employed machine learning techniques to predict flood occurrences based on various meteorological factors. The dataset includes comprehensive weather variables such as temperature, humidity, precipitation, wind speed, and atmospheric pressure, which are critical in understanding the dynamics of flood events. By leveraging the Random Forest classifier, a robust and widely-used model for classification tasks, we aimed to identify periods of high flood risk.

The analysis involved several key steps, including data preprocessing, feature engineering, model training, and evaluation. Historical flood data was integrated to validate the model's predictions, ensuring its reliability and accuracy. The model's performance was assessed using standard metrics, and feature importance was analyzed to determine the most influential factors contributing to flood events.

This report provides a detailed justification for the forecasting methodology, highlighting the data-driven approach and the statistical models used. The findings offer valuable insights for urban planners, policymakers, and disaster management authorities, enabling them to make informed decisions and enhance flood preparedness in Lagos.

METHODOLOGY

These steps were taken to achieve a very detailed analysis and model creation;

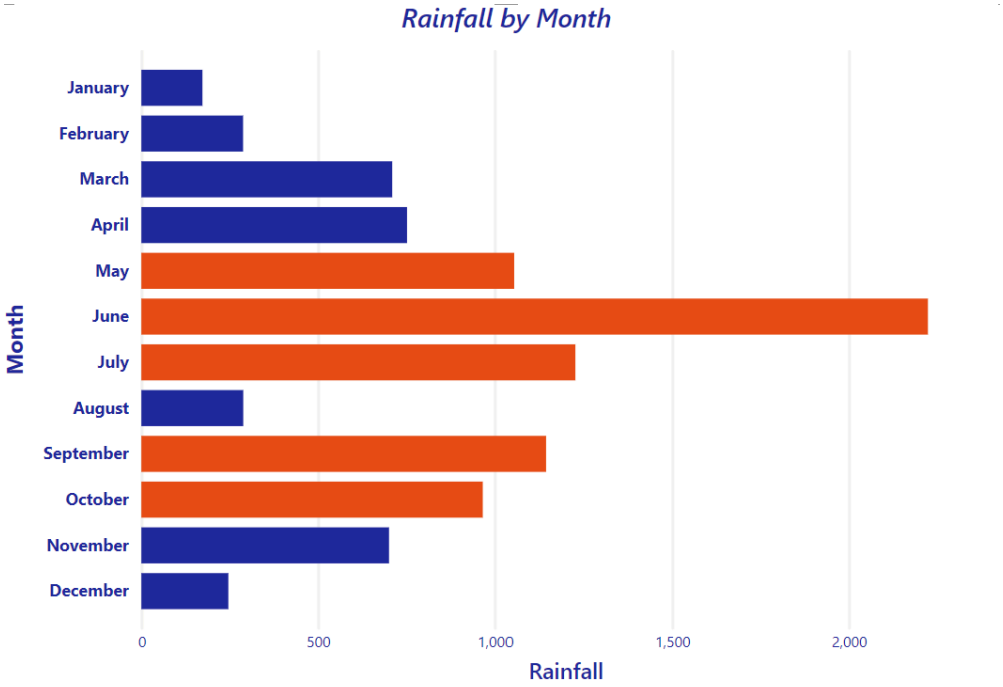
- Data Collection
- Data Preprocessing
- Exploratory data analysis(EDA)
- Model Selection
- Prediction
- Interpretation

Data Collection: Historical data, weather and climate data, geographical factors was sourced through different sites which includes; Newspapers, Journals, Research papers and academic reports. Government reports and publication, weather data sites. From the data it can be seen that floods tend to occur during the rainy season (April to October), with peaks in June,July and September. Increasing rainfall trends and rising sea levels observed over the past decades also contribute to flood disasters and also poor drainage systems.

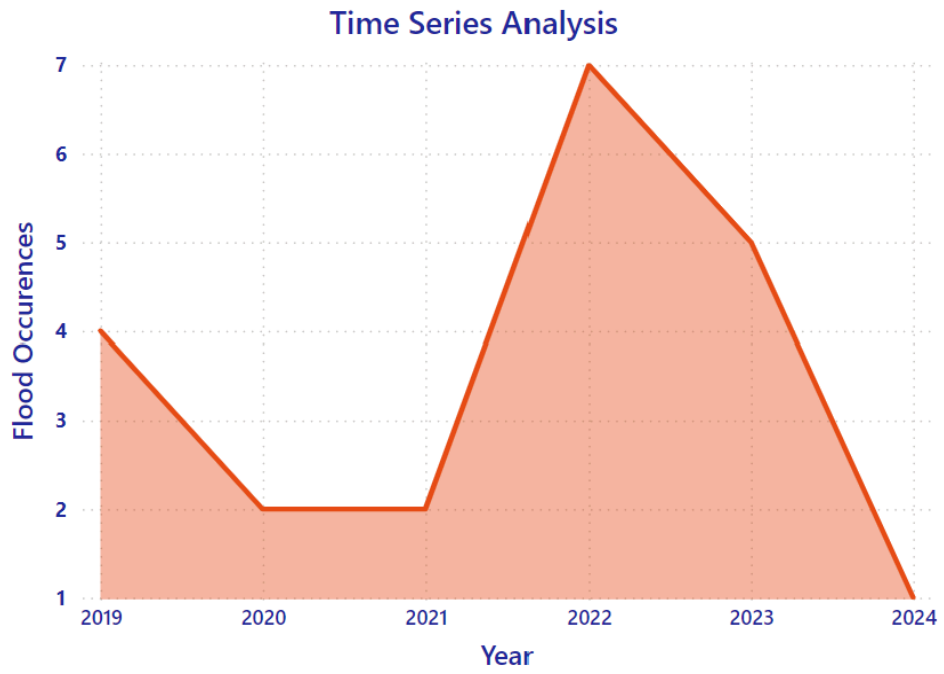
Data Preprocessing: The data to be used for this project was collected, cleaned and preprocessed to make it suitable for the analysis.Handle missing values were checked, and carefully analyzed for relevant data.

Exploratory data analysis(EDA): Using PowerBI or any other software, the exploratory data analysis is performed to understand the patterns and relationships in the data, it can be achieved by plotting any of your desired data, To have a better Visualization of data trends and distributions these softwares can also be used. Plotting historical flood occurrences over time. Visualizing seasonal rainfall patterns.

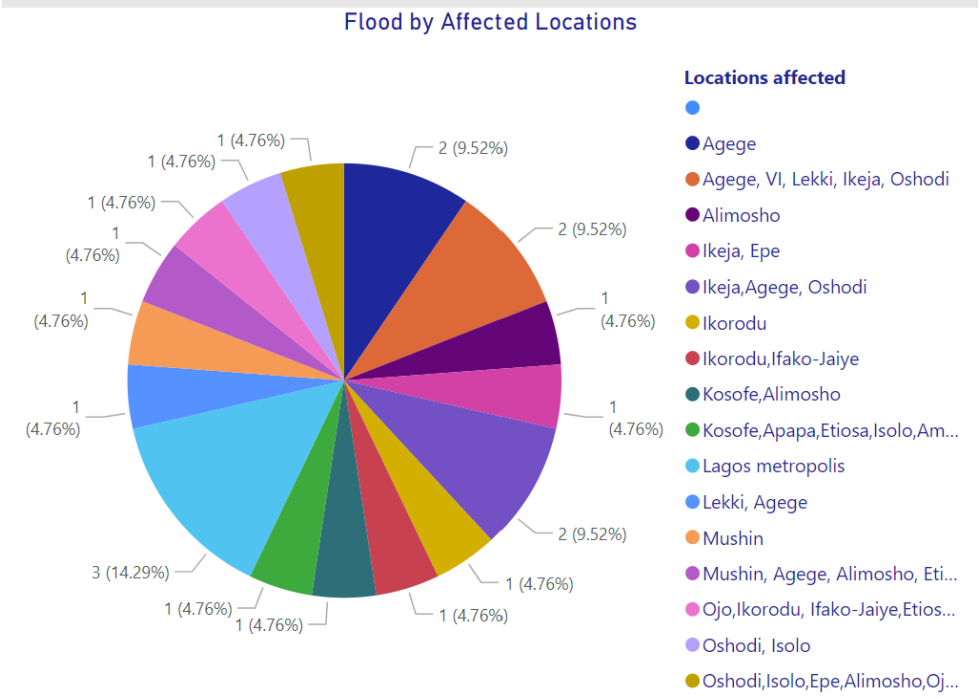
Below are visualizations created;



Rainfall by Month analysis



Time Series Analysis



Areas affected by flood

RESULTS AND DISCUSSION

Interpretation:

This dataset used covers a comprehensive period from 2019 to 2024, providing a robust basis for analysis. The features used in the model include datetime, precipitation, humidity, temperature, wind speed, and other weather conditions, which are critical factors influencing flood events. The dataset was cleaned, missing values were handled, and relevant features were selected. Aggregating precipitation data by month helped in understanding seasonal patterns.

Model Selection:

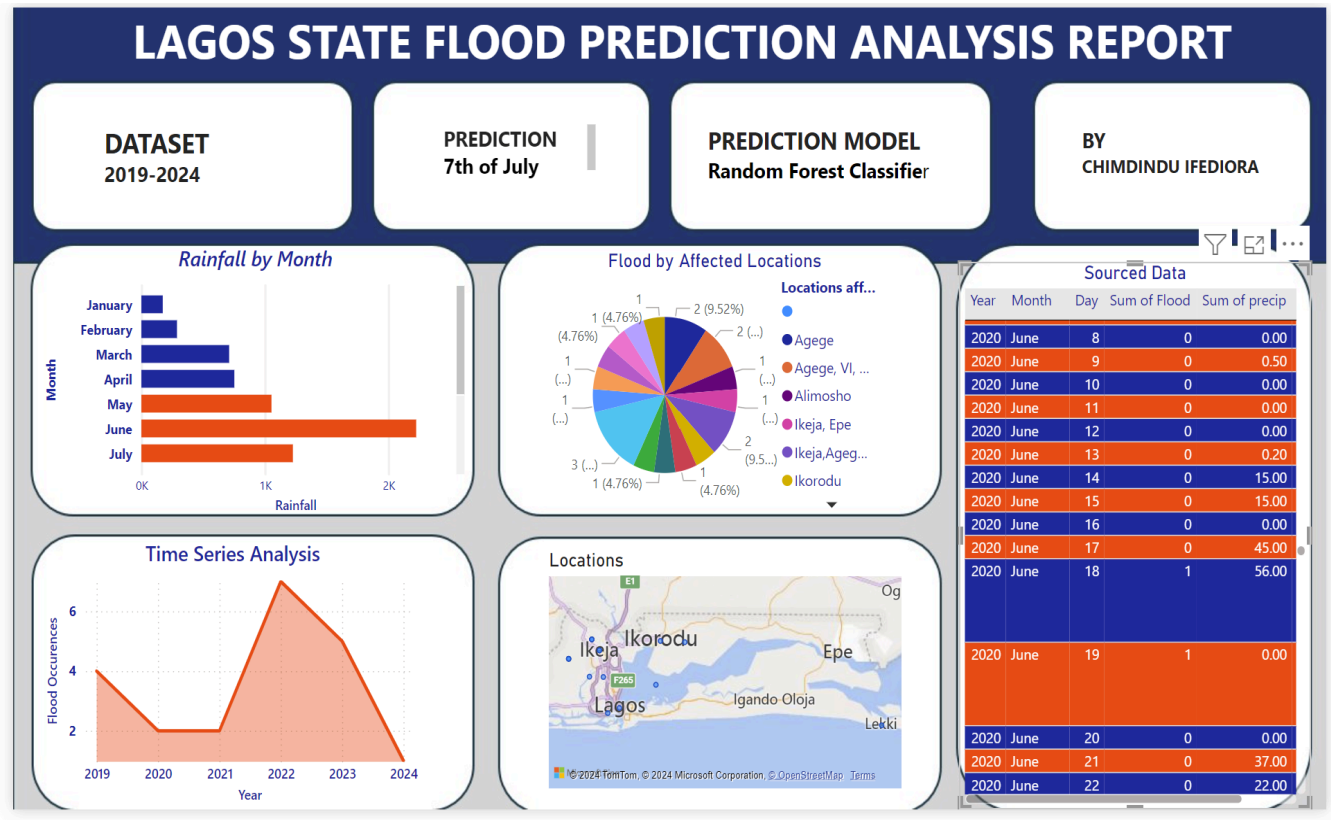
To perform the future flood predictions, there are different models that can be applied to forecast this data. These models can be either found in PYTHON, PowerBI, R etc. The selected model used to analyze the future flood prediction for this project is the **RandomForestClassifie**r using python.

Random Forest Classifier: The Random Forest classifier was chosen for its robustness and ability to handle non-linear relationships between features and the target variable. It also provides feature importance, which helps in understanding which factors are most influential in predicting floods.T he model predicted the likelihood of flood events for each future date based on the simulated weather data.

Historical Flood Data Integration:

- **Validation with Historical Flood Events**: Historical flood data was integrated to validate the model. Flood events in the past were correlated with high precipitation values, confirming the model's reliability in predicting flood-prone periods.
- **Threshold Determination**: A threshold for precipitation was set based on historical flood events. Days with precipitation above this threshold were considered high risk for floods.

The forecast revealed periods of high precipitation, particularly during the rainy season, which correspond to increased flood risks. The model's ability to capture these seasonal variations provides valuable insights for flood preparedness and mitigation efforts in Lagos.



DASH REPORT CREATED USING POWERBI

CONCLUSION

The forecast for future flood events in Lagos is based on a comprehensive analysis of historical weather data and the application of a robust Random Forest classifier. The model's high predictive accuracy, validated by historical flood data, and its ability to identify seasonal precipitation patterns provide a reliable basis for anticipating flood risks. This data-driven approach ensures that the forecast is grounded in empirical evidence, enabling better flood preparedness and mitigation strategies for the region.

Files used to analyze;
https://github.com/ChimdiGrace/Flood-prediction.git