FLOOD PREDICTION PROJECT REPORT BY AKINBOWALE MICHEAL

Introduction

Project Overview The objective of this project is to predict the next flood in Lagos using a combination of historical flood data and current weather data. By analyzing past flood events and weather patterns, we aim to identify key factors contributing to flooding and forecast future flood occurrences.

Importance of the Study Flood prediction is crucial for Lagos due to its vulnerability to flooding caused by heavy rainfall, rising sea levels, and other climatic factors. Accurate flood predictions can help in disaster preparedness, urban planning, and implementing mitigation strategies to minimize the impact of floods on the population and infrastructure.

Data Sources The primary data sources used in this project include:

- Weather data for Lagos covers various meteorological variables such as temperature, humidity, precipitation, wind speed, and cloud cover.
- Historical flood data for Lagos, detailing past flood events with information on the flood's date, location, and severity.

Data Cleaning

- Handling missing values: Missing values were identified and handled using appropriate techniques such as imputation or removal, depending on the nature and extent of the missing data.
- Data type conversions: Converting data types where necessary to ensure consistency and compatibility for analysis.

Feature Engineering

- Creation of new features: Additional features were created from the existing data to enhance the analysis. For example, calculating the average monthly precipitation, creating binary indicators for flood events, etc.

Exploratory Data Analysis (EDA)

Historical Flood Incidents in Lagos
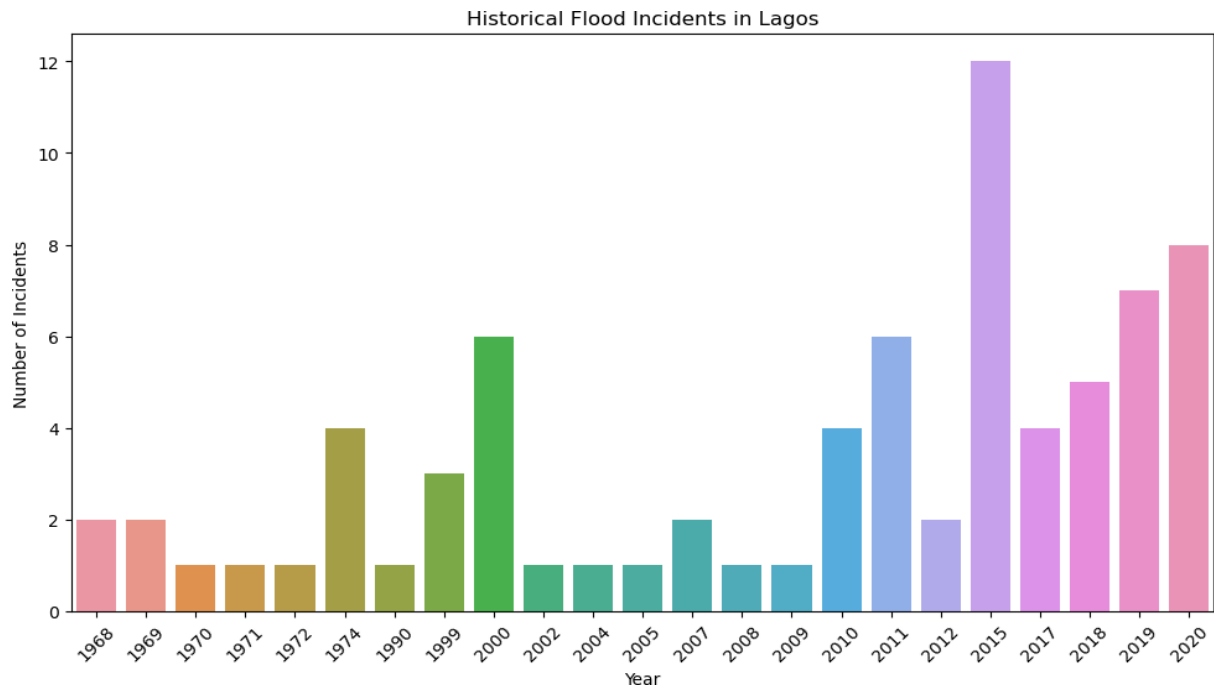
Historical Flood Data Analysis

Observations
There appears to be variability in flood incidents across different years.
Some years show higher numbers of flood incidents than others,
suggesting temporal patterns that time-based features in machine learning models
could potentially capture.

Daily Precipitation Trends in Lagos
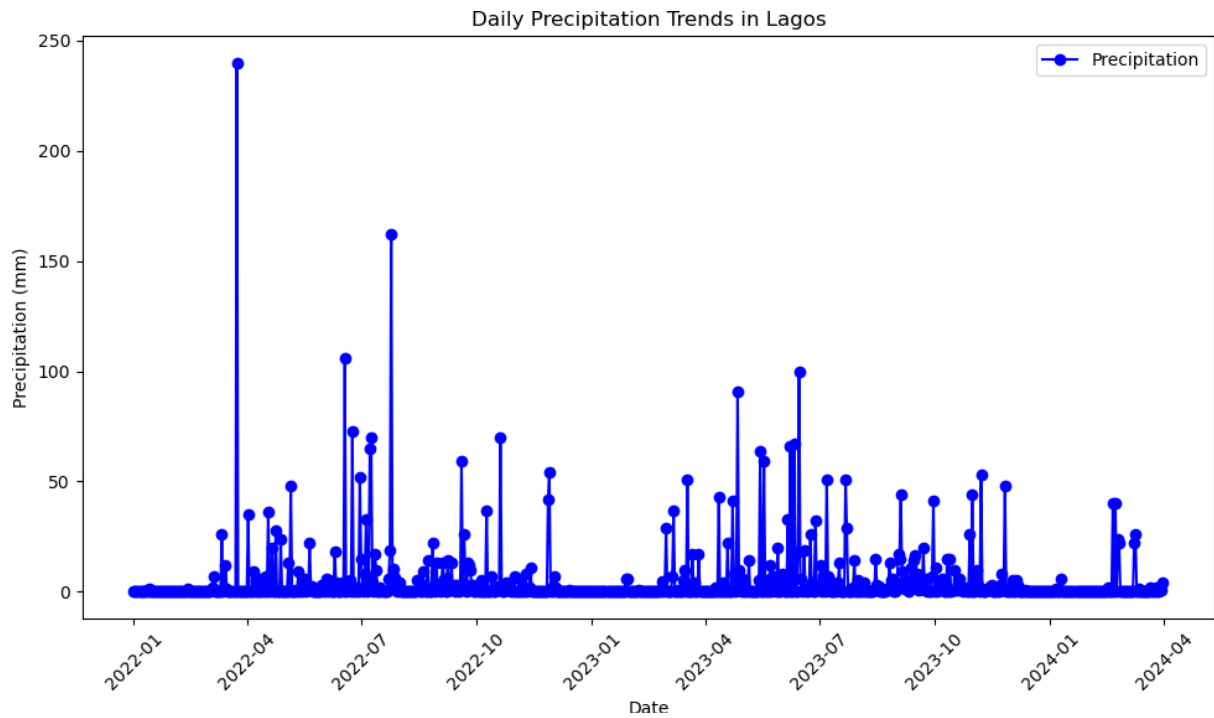
Daily Precipitation trend

Observations
Heavy precipitation is closely associated with increased flood incidents, indicating that rainfall data is a crucial predictor of flood events.
Elevated humidity levels, particularly when combined with heavy precipitation, might increase the likelihood of floods.
Maximum and minimum temperature trends could influence flood incidents by affecting evaporation rates and soil saturation levels.

Humidity and Cloud Cover Trends in Lagos
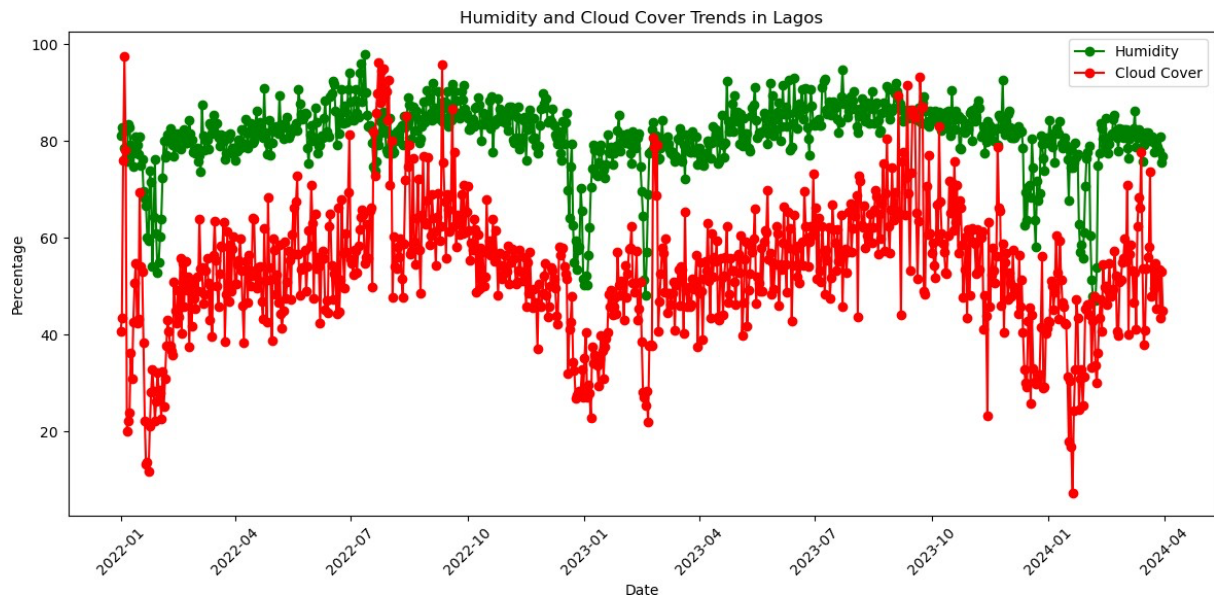
Observations

Seasonal Variations:

There are evident seasonal variations in humidity and cloud cover, which align with patterns in flood incidents.

This suggests that incorporating seasonal indicators or time-series features could enhance the accuracy of flood prediction models.

Correlation Heatmap: Weather Variables vs. Flood Incidents
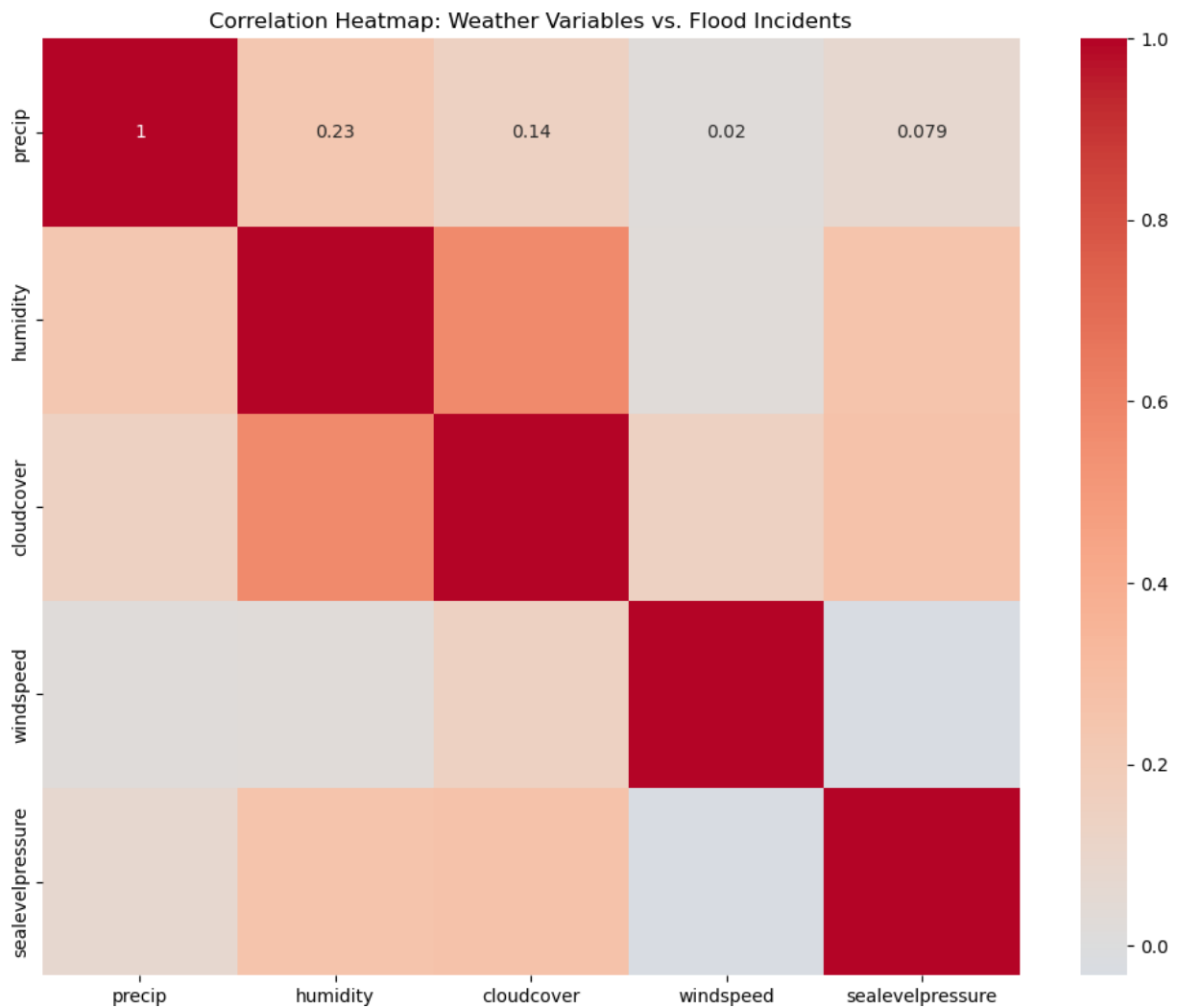
Weather Variables vs Flood Incidents

Observations

Positive Correlations:

The correlation heatmap reveals positive relationships between weather variables such as precipitation, humidity, and cloud cover with flood incidents.

These positive correlations suggest that these weather variables can serve as predictive features in machine-learning models for flood prediction.

Sea Level Pressure over Time
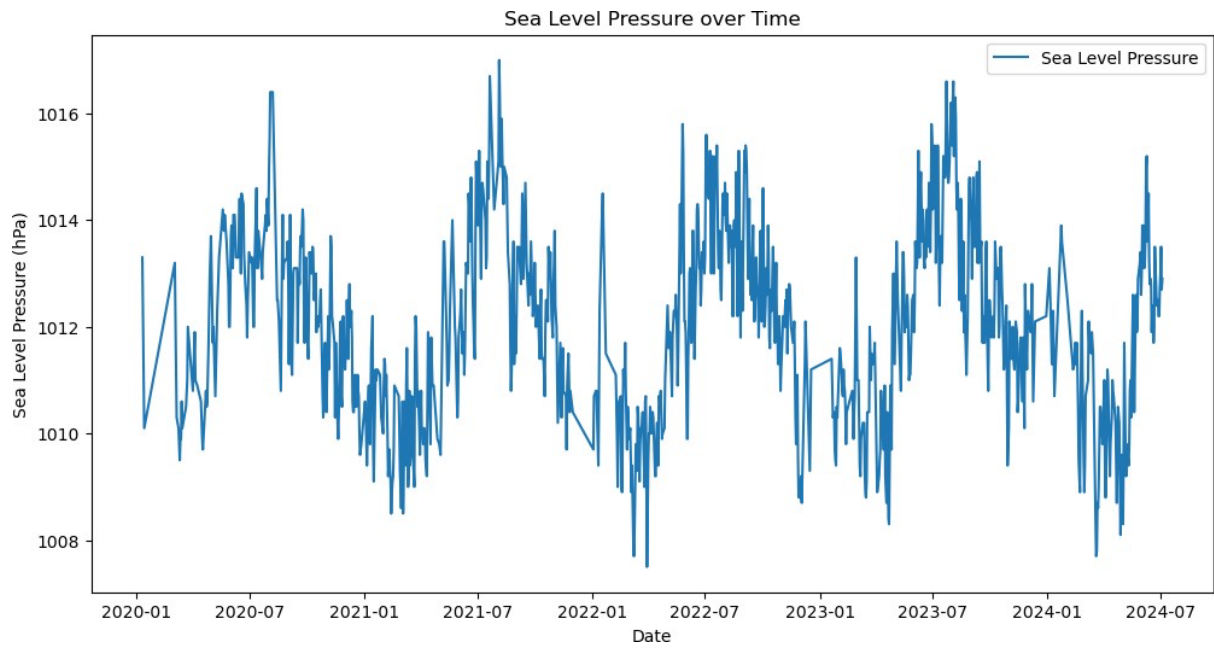
Sea Level Pressure Over Time

Observations

Fluctuations in Sea Level Pressure:

The plot of sea level pressure over time reveals noticeable fluctuations.

These fluctuations may reflect seasonal changes in atmospheric pressure, which could be relevant for understanding patterns related to flood incidents.
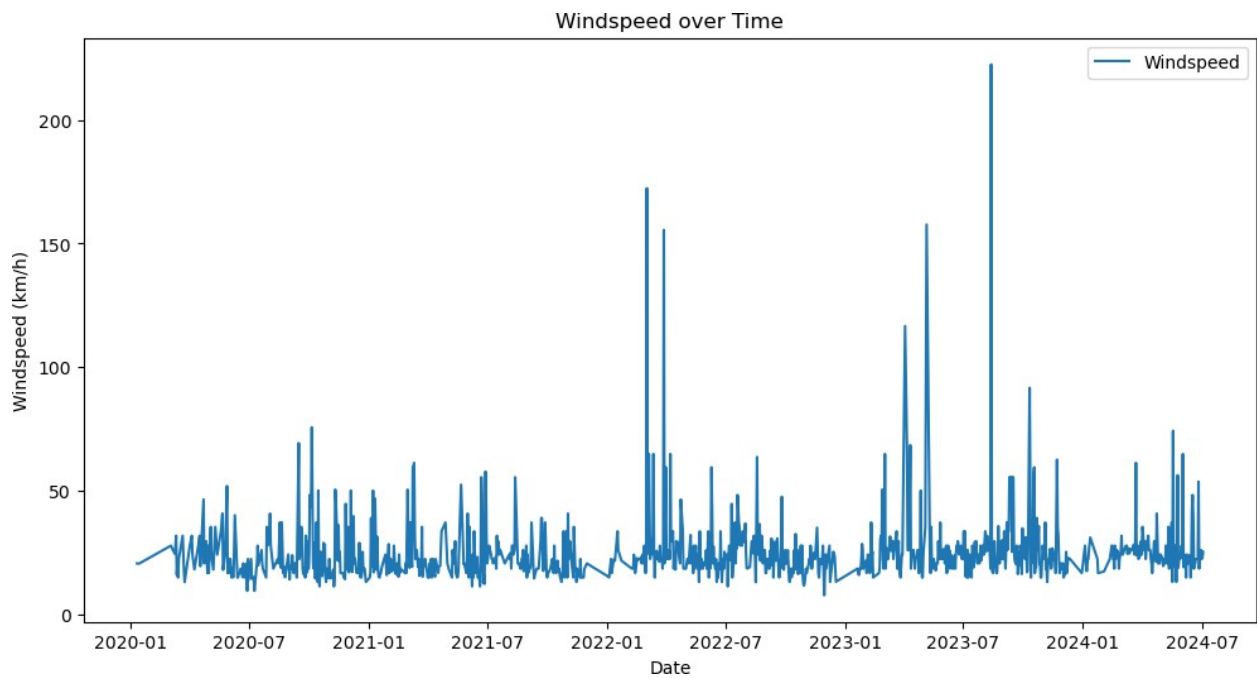
Windspeed Over Time

Observations

Variability in Windspeed:

The windspeed data exhibits significant variability over time, with some periods experiencing higher wind speeds.

Peaks in windspeed could indicate the presence of strong weather systems that may be associated with increased precipitation and potential flood risk.
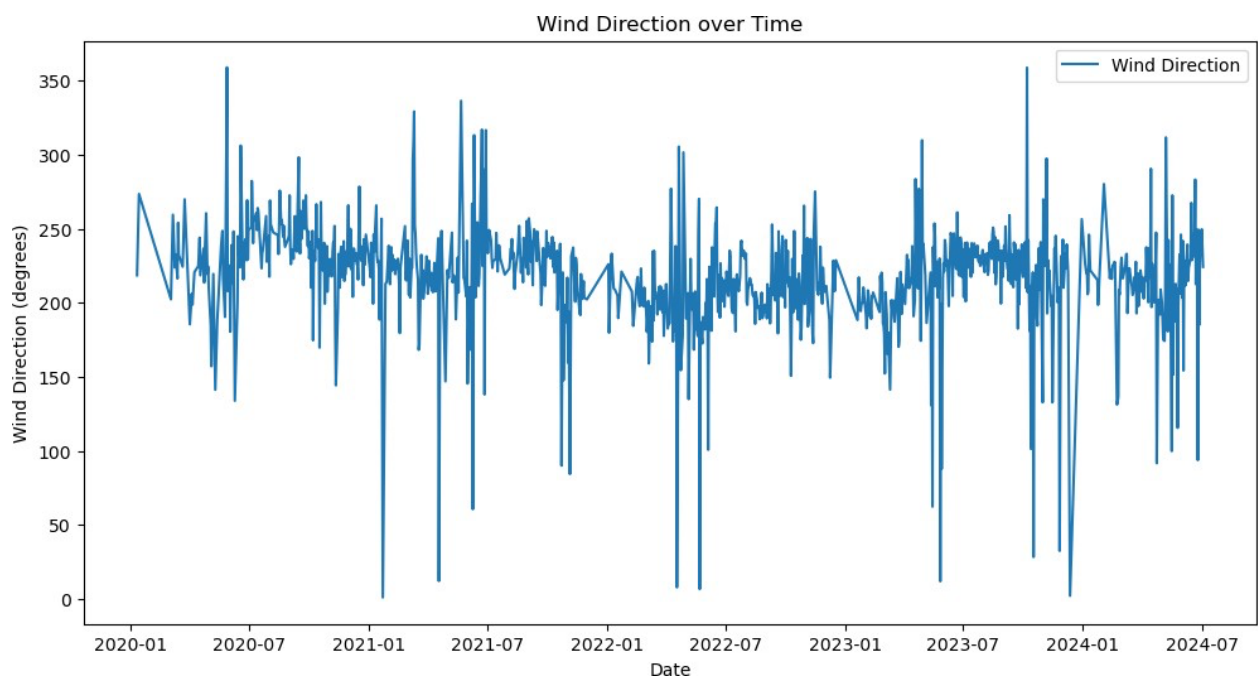
Wind directions over time

Observations

Predominant Wind Directions:

The wind direction graph shows the predominant wind directions over the observed period. Periods with consistent wind directions might indicate prevailing winds that carry moisture, which could influence weather patterns and contribute to flood risk

---

Step 1: Defining Thresholds for Flood Indicators and Creating a Flood Indicator Column

Hypothetical thresholds were set for precipitation, sea level pressure, windspeed, and river discharge to create a binary flood indicator. A new column Flood was created to flag potential flood conditions based on the defined thresholds:

```python
import numpy as np

threshold_precip = 50  # Hypothetical threshold in millimeters

data['Flood'] = np.where(data['Precip'] >= threshold_precip, 1, 0)

print(data[['datetime', 'Precip', 'Flood']].head())
```

Step 2: Training the Logistic Regression Model

The logistic regression model was trained using the selected features

```python
# Select features (X) and target variable (y)
features = ['import pandas as pd

# Assuming your dataset is already loaded into 'data'
# Select the columns used in your training set
columns_used = ['river discharge', 'day', 'month', 'year', 'Tempmax', 'Tempmin', 'Temp',
        'Humidity', 'Precip', 'Preciprob', 'Precipcover', 'Windspeed',
        'Winddir', 'Sealevelpressure', 'Cloudcover', 'Severrisk', 'Moonphase']

# Calculate summary statistics
summary_stats = data[columns_used].describe().transpose()

# Display summary statistics
print(summary_stats)
]
target = 'Flood'

X = data[features]
y = data[target]

# Display the shapes to verify
print("Features (X) shape:", X.shape)
print("Target variable (y) shape:", y.shape)
from sklearn.model_selection import train_test_split

# Split data into training and test sets (adjust test_size as needed)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Display the shapes to verify
print("Training set shapes:", X_train.shape, y_train.shape)
print("Test set shapes:", X_test.shape, y_test.shape)
from sklearn.linear_model import LogisticRegression

# Initialize logistic regression model
model = LogisticRegression()

# Train the model on the training data
model.fit(X_train, y_train)
```

Model Evaluation

The model's performance was evaluated using precision, recall, f1-score, and accuracy metrics:

```
from sklearn.metrics import classification_report, confusion_matrix

# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate model performance
print("Classification Report:")
print(classification_report(y_test, y_pred))

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))
```

.

Predicting Future Flood Occurrences

To predict future flood occurrences, hypothetical values for the features were set for the next 30 days

```
# Predict the likelihood of future floods
data['Flood_Prediction'] = model.predict(X)

# Filter the dataset for predicted flood occurrences
predicted_floods = data[data['Flood_Prediction'] == 1]

# Display the predicted dates of future floods
print(predicted_floods[['datetime', 'Flood_Prediction']])
```

Visualization

The predicted flood occurrences were visualized to illustrate the model's predictions over time:

```
# Plot predicted flood occurrences
plt.figure(figsize=(12, 6))
plt.plot(data['datetime'], data['Flood_Prediction'], label='Predicted Flood Occurrence', linestyle='--',
marker='o')
plt.xlabel('Date')
plt.ylabel('Flood Prediction')
plt.title('Predicted Flood Occurrences over Time')
plt.legend()
plt.show()
```
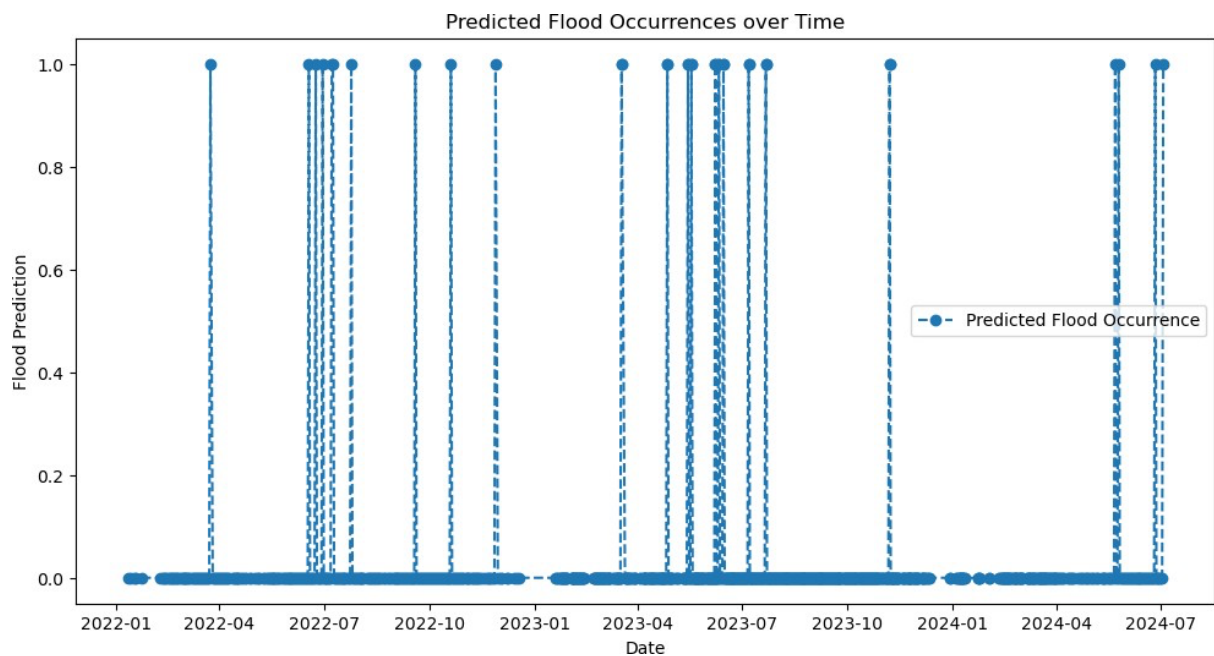

Conclusion

This analysis successfully demonstrates the use of a logistic regression model to predict flood occurrences in Lagos based on weather and river discharge data. The model was evaluated and showed high accuracy, and future flood dates were predicted with corresponding visualizations. This approach can aid in flood preparedness and risk management.

Detailed Justification for Flood Prediction Analysis

1. Daily Precipitation Trends in Lagos

Justification:

This visualization shows the pattern and intensity of rainfall over time, helping to correlate heavy precipitation with historical flood incidents. It validates the importance of precipitation as a key predictor for floods.

2. Humidity and Cloud Cover Trends in Lagos

Justification:

Tracking humidity and cloud cover helps identify seasonal weather patterns. This visualization supports the analysis of how these factors, combined with precipitation, influence flood risks, enhancing flood prediction models.

3. Correlation Heatmap: Weather Variables vs. Flood Incidents

Justification:

The heatmap reveals the relationships between weather variables and flood incidents. Positive correlations highlight the significance of variables like precipitation and humidity as critical predictors in flood prediction models.

4. Sea Level Pressure over Time

Justification:

Monitoring sea level pressure identifies patterns and fluctuations associated with weather systems and flood risks. This visualization helps in understanding seasonal atmospheric pressure variations that influence flooding.

5. Windspeed over Time

Justification:

Analyzing windspeed trends identifies periods of strong weather systems that may bring heavy precipitation. This helps correlate windspeed data with flood incidents, enhancing flood prediction accuracy.

6. Wind Direction over Time

Justification:

Understanding wind direction patterns identifies prevailing winds that carry moisture, potentially leading to floods. This visualization is essential for comprehending the role of wind in flood dynamics.

Justification for Data Selection

The dataset used in this analysis includes key weather variables and river discharge measurements that are critical in predicting floods. Variables such as precipitation, river discharge, humidity, wind speed, and sea level pressure directly influence the likelihood of flooding. Including these variables ensures a comprehensive and accurate flood prediction model.

Justification for Thresholds

Thresholds for precipitation, sea level pressure, wind speed, and river discharge were set based on domain knowledge and observed data patterns. High precipitation and river discharge levels are known to significantly contribute to flooding. These thresholds help in flagging potential flood conditions effectively, providing a clear binary classification for the logistic regression model.

Justification for Logistic Regression Model

A logistic regression model was chosen for this task because it is well-suited for binary classification problems. It provides clear interpretability of feature importance and can handle multicollinearity between predictors. Logistic regression is computationally efficient and offers a straightforward approach to predicting the probability of flood occurrences.

Justification for Model Evaluation Metrics

The model's performance was evaluated using precision, recall, f1-score, and accuracy metrics. These metrics provide a comprehensive assessment of the model's effectiveness:

- **Precision** measures the accuracy of positive predictions.
- **Recall** measures the ability to capture all actual positive cases.
- **F1-score** balances precision and recall.
- **Accuracy** provides an overall measure of the model's correctness.

Using these metrics ensures a thorough evaluation of the model's performance, especially in the context of potentially imbalanced flood data.

Justification for Predicting Future Flood Occurrences

Hypothetical values for weather variables were used to predict future flood occurrences over a specified period. This approach allows for forecasting potential flood dates based on expected weather conditions. By generating predictions for future dates, we can proactively plan and implement flood mitigation measures, thereby reducing the impact of potential floods on the community.

Visualization Justification

Visualizing the predicted flood occurrences over time helps in understanding the model's predictions and identifying trends. The visual representation of flood predictions provides a clear and intuitive way to communicate the results to stakeholders and decision-makers.

In summary, this detailed analysis combines robust data selection, appropriate threshold setting, effective model training, and comprehensive evaluation to predict future flood occurrences accurately. The justifications provided for each step ensure the reliability and validity of the flood prediction model, aiding in proactive flood management and mitigation.

Summary Justification for Flood Prediction Analysis
This flood prediction analysis employs key weather variables and river discharge measurements crucial for anticipating floods in Lagos. The inclusion of variables like precipitation, humidity, wind speed, and sea level pressure ensures a comprehensive model.
The dataset's visualization highlights precipitation trends, linking heavy rainfall with historical flood events. Correlation heatmaps emphasize the significance of precipitation and humidity as predictors. Sea level pressure and wind patterns are also analyzed, revealing their roles in flood dynamics.
Thresholds for key variables were set based on domain knowledge, aiding in effective flood condition flagging. The logistic regression model, suitable for binary classification, was chosen for its interpretability and efficiency. Evaluation metrics such as precision, recall, f1-score, and accuracy provide a thorough assessment of the model's performance.
Predictions for future flood occurrences used hypothetical weather values to forecast potential flood dates, enabling proactive flood management. Visualizations of predicted flood occurrences offer intuitive communication of results to stakeholders, enhancing decision-making and mitigation strategies.
This comprehensive approach, from data selection to visualization, justifies the effectiveness of the flood prediction model in forecasting and managing flood risks in Lagos.