**TECHNICAL REPORT ON FLOOD PREDICTION**

**IN LAGOS STATE, NIGERIA**

**BY**

**OPEYEMI FAYIPE**

**(**[**opeyemifayipe@gmail.com**](mailto:opeyemifayipe@gmail.com)**)**

**July, 2024**

#### 

#### 

#### 

#### 

#### 

#### 

#### **Introduction**

Flood prediction is a crucial aspect of disaster management and urban planning. In this project, I aimed to predict the occurrence of floods in Lagos, Nigeria, using historical weather data.

Flooding typically occurs when prolonged rain falls over several days, when intense rain falls over a short period of time, or when an ice or debris jam causes a river or stream to overflow onto the surrounding area.

Also human activities, which damage the environment, for example, sand mining, deforestation and poor garbage disposal, increase the risk of flooding. Areas most likely to be worst affected by flooding are: Low-lying coastal areas. Areas near gully banks

The city of Lagos is bordered by several freshwater streams, a lagoon and the Atlantic Ocean. Its geographical location combined with its relatively flat topography and an average elevation of only 1.5 m above sea level make this megacity vulnerable to coastal flooding, which will be exacerbated by the rising sea level.

Knowledge about the characteristics of a river's drainage basin, such as soil-moisture conditions, ground temperature, snowpack, topography, vegetation cover, and impermeable land area, can help to predict how extensive and damaging a flood might become.

The main tools used to detect heavy rainfall associated with flash floods are satellite, lightning observing systems, radar, and rain gauges.

Flood prediction is crucial for mitigating the adverse effects of floods, especially in vulnerable regions like Lagos State. This project aims to predict flood occurrences using historical weather data. The prediction model leverages machine learning techniques to analyze weather patterns and forecast potential flood events. This report provides a comprehensive overview of the methodology, data processing steps, challenges faced, and recommendations for future work.

### 

### 

### **Methodology**

The methodology for this flood prediction project involves the following steps:

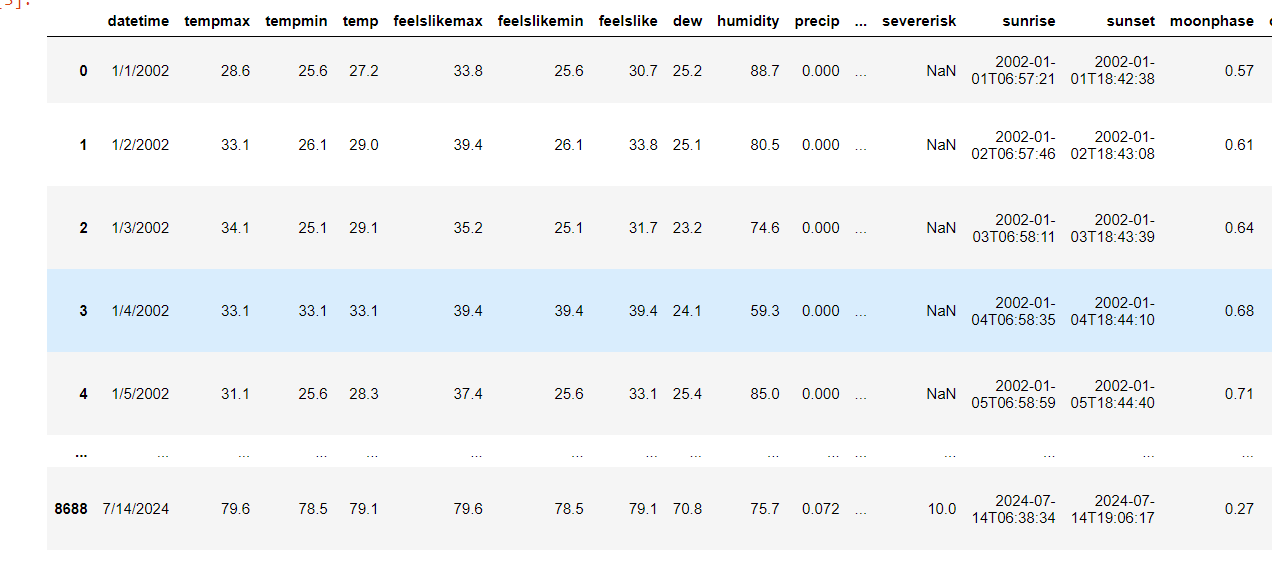
1. **Data Collection**: Historical weather data and flood occurrence data were collected. The weather data includes various meteorological parameters, while the flood data records the dates of past flood events.
2. **Data Processing**: The data was preprocessed to handle missing values, normalize features, and create new features relevant to flood prediction.
3. **Exploratory Data Analysis (EDA)**: Various visualizations and statistical analyses were performed to understand the relationships between weather parameters and flood occurrences.
4. **Feature Engineering**: New features such as monthly rainfall totals were created to enhance the predictive power of the model.
5. **Model Selection and Training**: A machine learning model was selected and trained using the processed data. Logistic Regression was initially chosen for its simplicity and interpretability.
6. **Evaluation**: The model's performance was evaluated using metrics like accuracy, precision, recall, and the confusion matrix.

#### 

#### 

#### **Data Description**

The dataset used for this project consists of weather data from 01/01/2002 to 07/3/2024, with the following columns:



* Datetime: Date of observation
* Temperature related: tempmax, tempmax, tempmin, temp, feels\_like\_max, feels\_like\_min, feels\_like
* Dew
* Humidity
* Precipitation: precipprob, precipcover, preciptype
* Wind Gust
* Wind speed
* Wind direction
* Sea level pressure
* Cloudcover
* Visibility
* Solar radiation
* Solar energy
* Uv index
* Severe risk
* Moonphase

Additional lagged and rolling features for precipitation were created for better temporal analysis. A binary target variable Flood\_dataset was derived from separate flood occurrence data.

I had a separate dataset with dates of flood occurrences in Lagos from 2002 to 2024.



#### 

#### 

#### 

### **Data Processing**

#### **Initial Data Preparation**

* **Loading Data**: Weather data and flood data were loaded from CSV files.
* **Datetime Conversion**: The date columns were converted to datetime format for easier manipulation and merging.
* **Merging Data**: Weather and flood datasets were merged on the date column to align weather conditions with flood events.

#### **Handling Missing Values**

* Missing values in the weather data were filled using the mean of the respective columns.

#### **Feature Engineering**

* **Monthly Rainfall Totals**: New features representing the total rainfall for each month were created to capture seasonal trends.
* **Normalization**: Features were normalized using StandardScaler to ensure they are on a similar scale.
* Preciptype: This column register if there were rain or not, it was in string format and converted to integer (0 - for no rain, 1 - for rain)
  + Rolling averages for precipitation (7-day, 14-day, and 30-day sums).
  + Lagged features for precipitation to capture temporal dependencies.

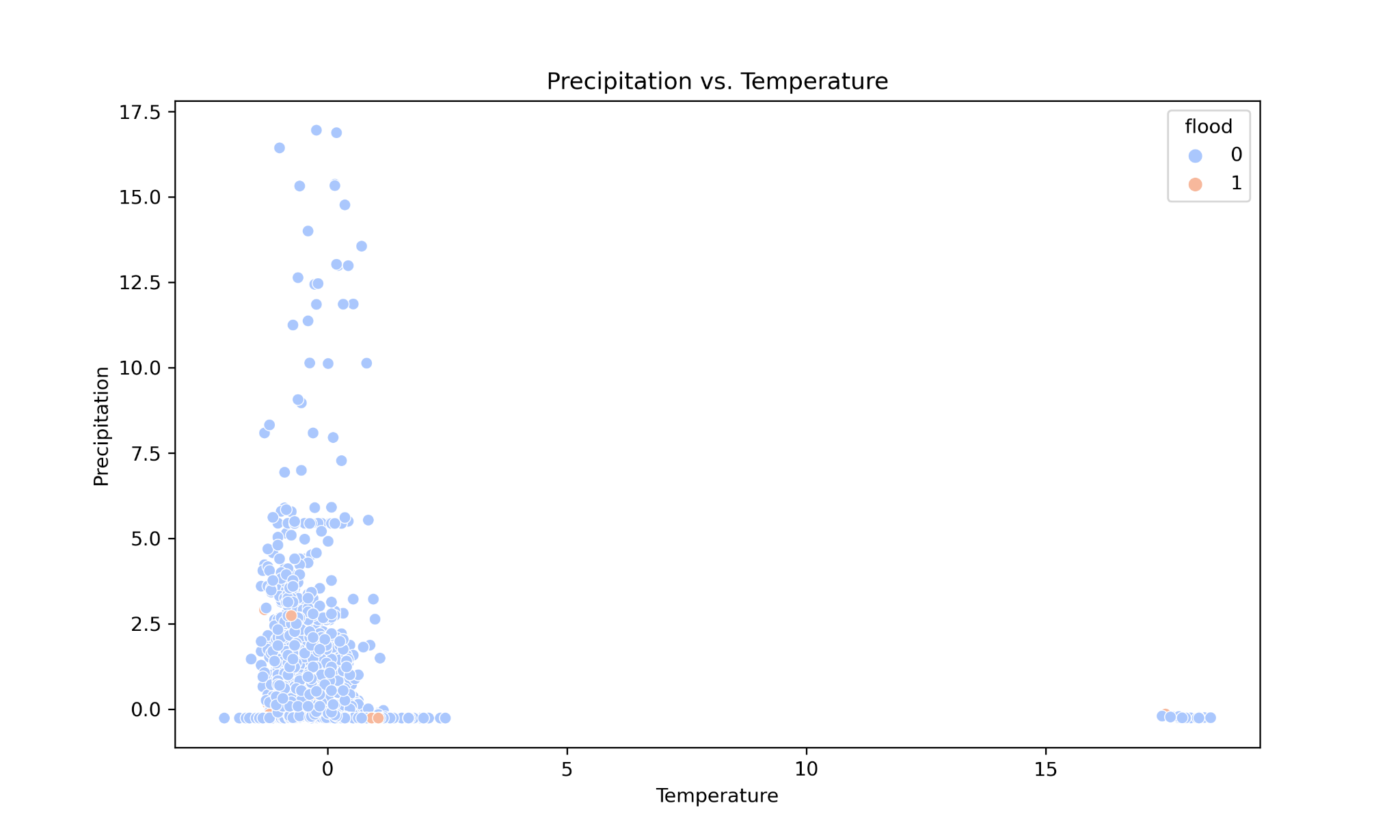
### **Exploratory Data Analysis (EDA)**

EDA was conducted to visualize and understand the data. Key visualizations included:

1. **Scatter Plot of Precipitation vs. Temperature**:
   * Showed the relationship between precipitation and temperature, with flood events highlighted.
2. **Correlation Heatmap**:
   * Displayed the correlations between different weather parameters and flood occurrences.
   * High correlations indicate strong relationships that could be important for prediction.

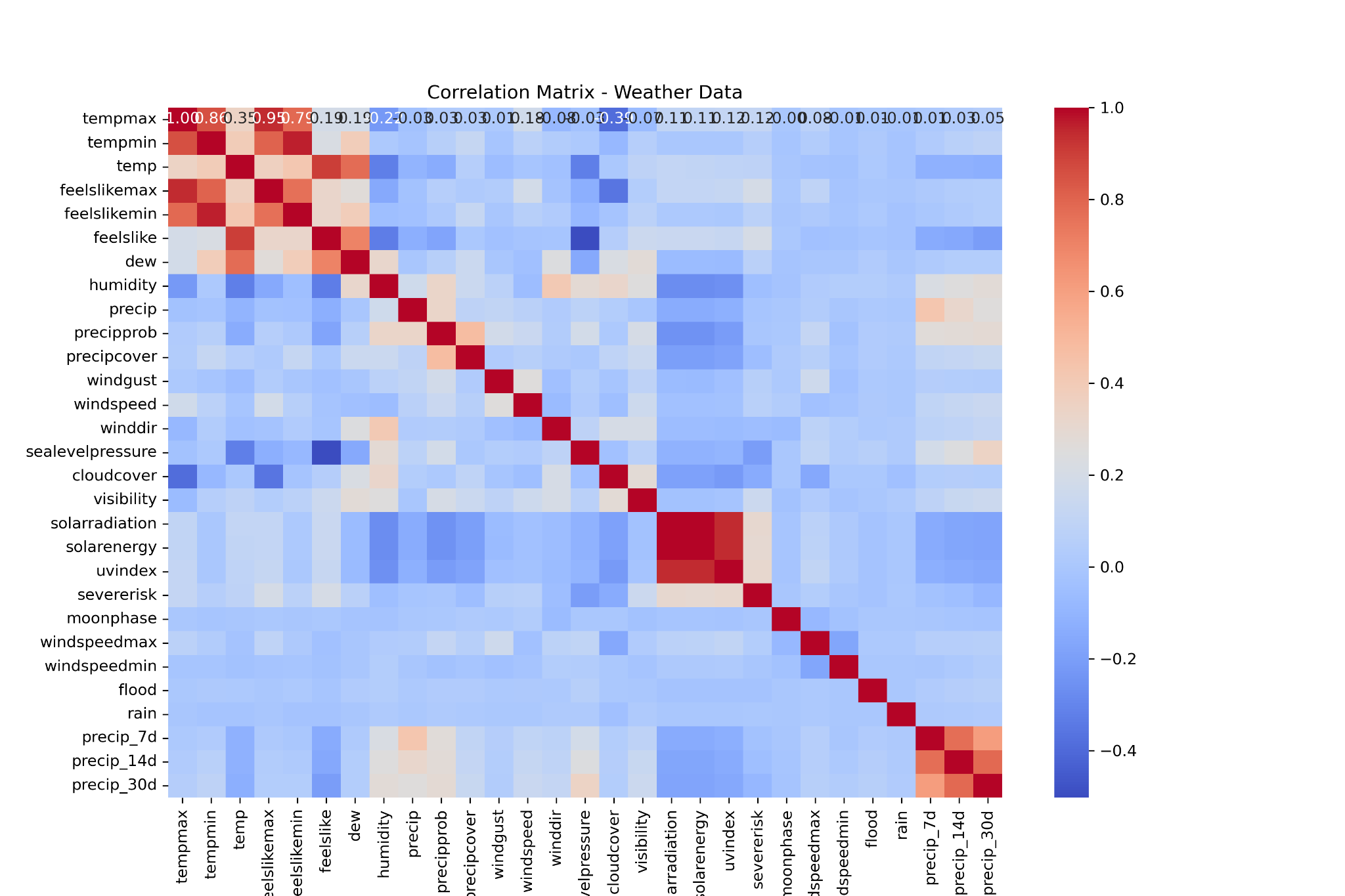
**Scatter Plot of Precipitation vs. Temperature**:

*Precipitation is the part of the water cycle that delivers water from the atmosphere to the Earth's surface. When and where precipitation falls is determined by the climate system, especially by the patterns of atmospheric and ocean circulation, and how much water returns in the atmosphere.*



Correlation Matrix

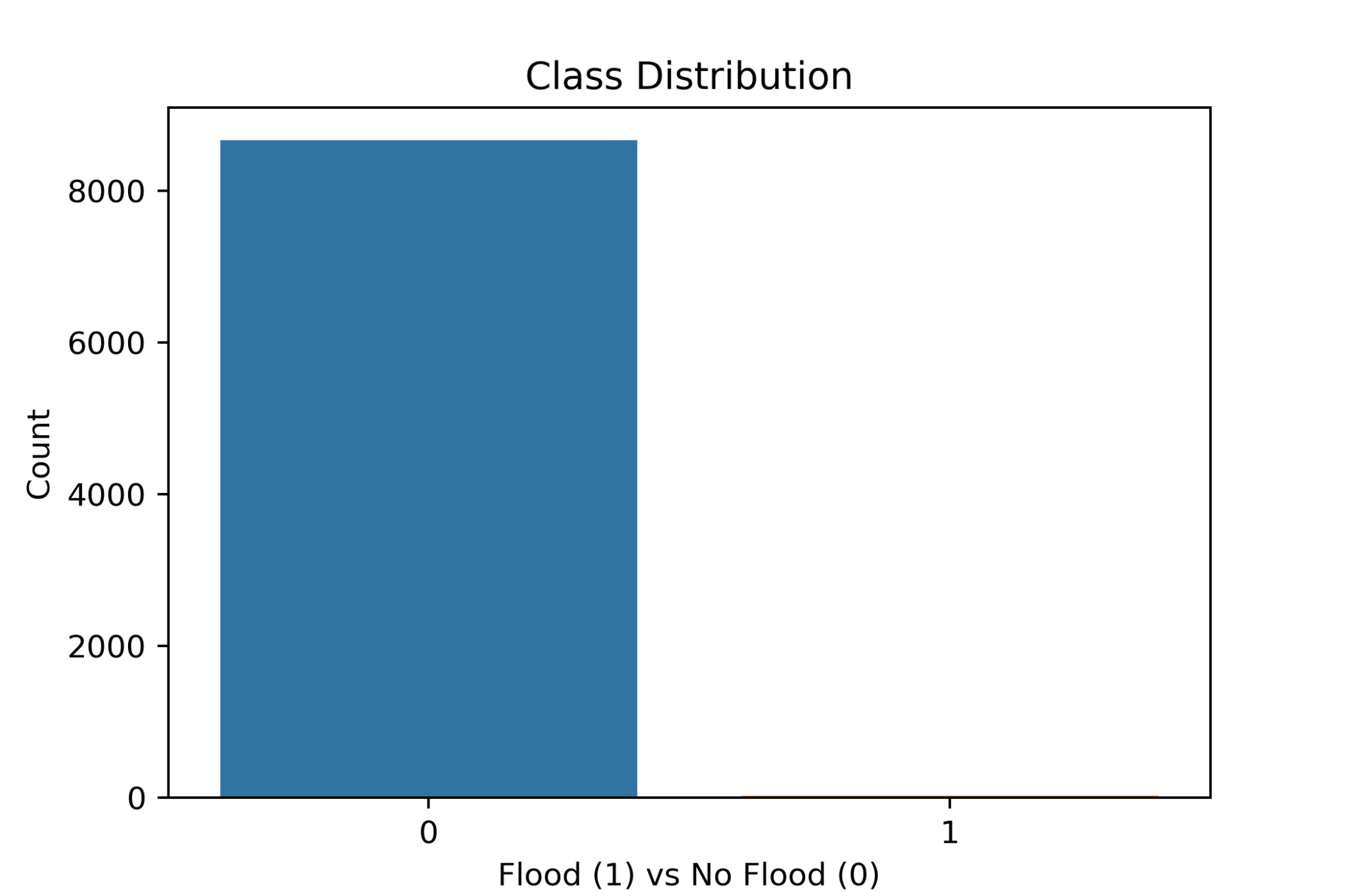
*The diagram below implies that the relationship between the variables is negative.*



#### 

#### **Flood Distribution**

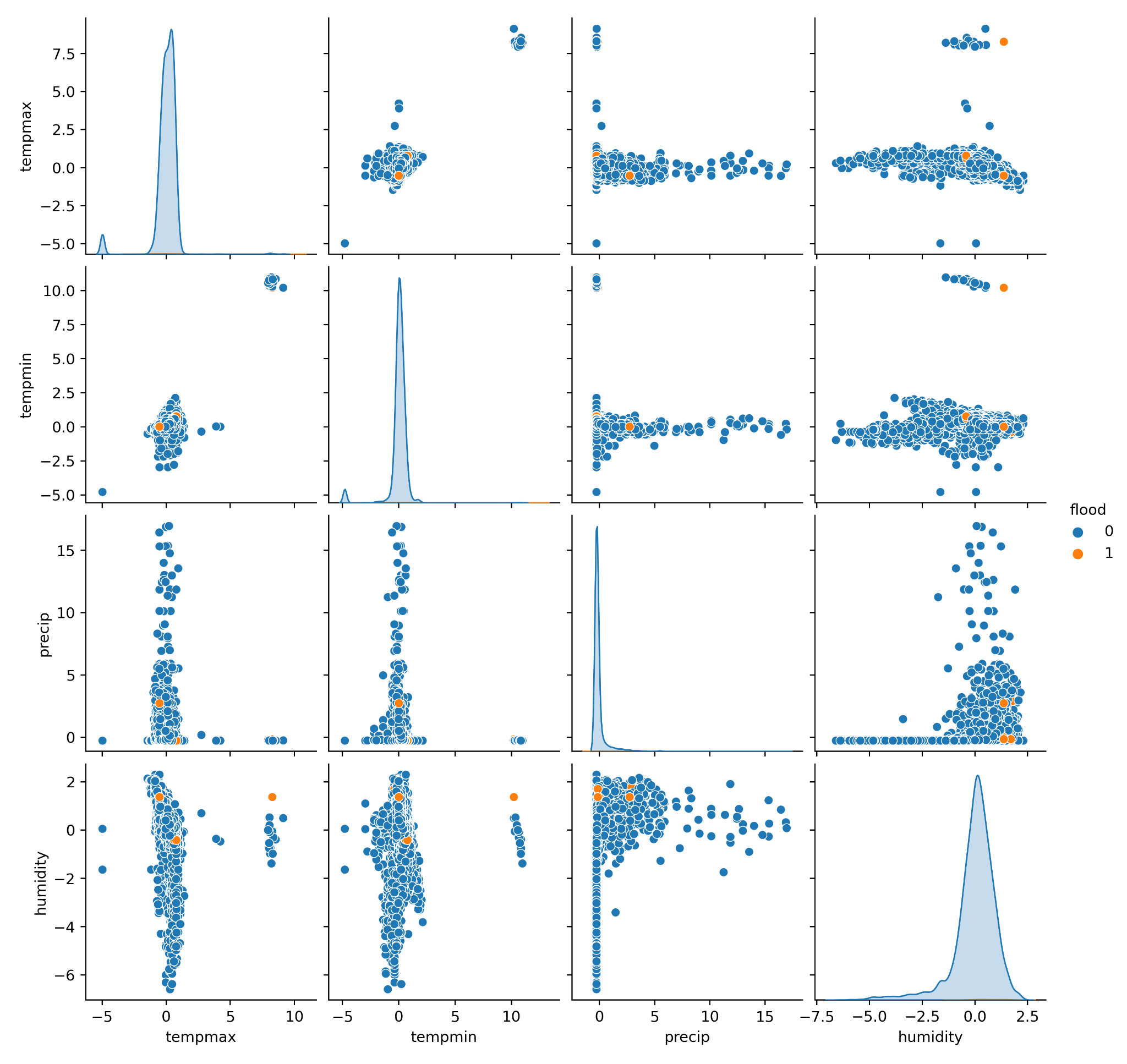
*The diagram below shows from the dataset that there are higher counts of no flood than flood.*



#### 

#### **Pairplot**

*Relationships of variables between flood and no flood.*



#### 

#### **Histogram of Numerical features**

*Histogram of each column in weather\_data*

#### 

#### **Model Development**

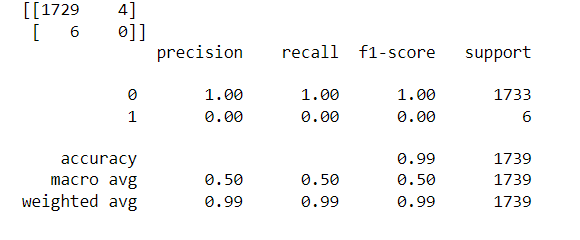
I used a Random Forest classifier to predict flood occurrences. Given the severe class imbalance (floods are rare events), I explored several strategies:

1. **Class Weight Adjustment**: Adjusted the class weights to give higher importance to the minority class (floods).
2. **SMOTE (Synthetic Minority Over-sampling Technique)**: Applied SMOTE to create a balanced training dataset by generating synthetic samples for the minority class.

#### 

#### **Results**

Despite my efforts, the model struggled to predict floods accurately. The confusion matrix and classification report showed that the model had high accuracy for the majority class (no floods) but failed to predict the minority class (floods).

****

### **Additional Data for Improved Prediction**

To enhance the prediction accuracy, the following data would be beneficial but were not available:

1. **Topography**: Detailed elevation and slope information to understand water flow and accumulation patterns.
2. **Urbanization**: Data on land use, population density, and impervious surface coverage to assess the impact of urban development on flood risks.
3. **Drainage Systems**: Information on drainage capacity, locations, and maintenance status to evaluate the effectiveness of water removal during heavy rains.
4. **Historical Flood Events**: Comprehensive records including the severity and exact locations of past floods to better train the model on flood-prone areas.
5. **Hydrological Data**: River levels, streamflow data, and soil moisture content to predict flooding from upstream water sources and ground saturation levels.

Incorporating these additional data sources would likely improve the model's ability to predict floods more accurately by providing a more complete picture of the factors contributing to flood risks.

**Challenges**

Imbalanced Dataset: The dataset was highly imbalanced with very few flood occurrences compared to non-flood events. This made it difficult for the model to learn and predict floods accurately.

Missing Data: Handling missing values appropriately was challenging and required careful consideration to avoid introducing biases.

Feature Engineering: Creating meaningful features that capture the underlying patterns in the data was crucial but complex.

Model Performance: Despite various efforts, the model struggled to predict flood events due to the imbalance and potential lack of relevant features.

### **Recommendations**

1. Address Imbalance: Use techniques like SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset and improve model performance.
2. Incorporate Additional Features: Include more features such as topography, urbanization, and drainage system data, which are critical for accurate flood prediction.
3. Advanced Models: Experiment with more sophisticated models like Random Forest, Gradient Boosting, or XGBoost that can handle complex patterns and imbalanced data better.
4. Domain Expertise: Collaborate with hydrologists and urban planners to incorporate domain knowledge into the model development process.

### **Conclusion**

This project provided valuable insights into the challenges and potential strategies for flood prediction in Lagos State. While the initial model faced difficulties due to data imbalance and feature limitations, the methodology and findings lay the groundwork for future improvements. Addressing the identified challenges and incorporating additional data can significantly enhance the accuracy and reliability of flood predictions, ultimately aiding in better flood preparedness and mitigation efforts.

**References:**

**visualcrossing.com\weather\data**

**ResearchGate**