Predicting Harris County Flooding Levels Using Machine Learning Models

Serena Shah & Osvaldo Salinas

April 16, 2024

Problem Statement

Flooding is one of the most common natural disasters, causing widespread damage to infrastructure, displacement of populations, and significant economic losses annually. The ability to predict the severity of potential flood events accurately is crucial for effective planning, timely responses, and mitigating these impacts. This project aims to develop a machine learning-based system to predict the highest levels of flooding, leveraging historical flood data in the Harris county area. The goal is to provide actionable insights that can assist in flood risk management and emergency preparedness.

Data Sources

1. **Flood Data:** WebMap showing all of the High Water Marks taken by Harris County Flood Control District's Hydrologic Operation team after a flooding event.The maps are feature layers for ArcGIS pro but can be exported to tables in excel. The tables will have data like water level, latitude, longitude, and dates. We have data from 2006 to 2019 for specific flood days.
   https://www.arcgis.com/home/item.html?id=bd266d344dab42b993c5d4f0b4599282
2. **USGS Elevation Point Query Service:** We are going to attempt to add more data to the flood tables by appending an elevation measurement to each coordinate using this API: https://apps.nationalmap.gov/epqs/

Methods, Techniques, and Technologies

The list of high-level methods, techniques, and technologies being considered encompasses several key components. Firstly, the use of ArcGIS Pro tools will be used to tabularize the data

for processing. Next, data preprocessing techniques are crucial for cleaning, integrating, and transforming raw data into a suitable format for analysis. We will use regression models employed to predict quantitative flood levels and classification models utilized to identify high-risk flood areas. The models under consideration include Random Forest and K-Nearest Neighbors. Implementation is facilitated by the Python programming language, supported by essential libraries such as scikit-learn, pandas, and NumPy for machine learning and data manipulation tasks. Furthermore, visualization tools like Matplotlib and Folium are being considered to effectively present the analyzed data, enhancing comprehension and decision-making processes.Packaging the model inference server in a Docker container image allows for easy deployment and scalability. Docker Hub serves as a repository for storing and sharing container images, ensuring accessibility across different platforms.

Products to be Delivered

The key deliverables for this project encompass several components. Firstly, a predictive model utilizing machine learning techniques to forecast flood levels and identify high-risk areas accurately. Additionally, a comprehensive project report will be compiled, documenting the methodology, analysis, model development, and evaluation process, references including discussions on the model's efficacy and constraints. Furthermore, a video presentation will be prepared, presenting the project findings in a manner suitable for both academic and professional audiences. Access to the complete project code and datasets will be provided through a Git repository, ensuring transparency and reproducibility. Lastly, a user guide or documentation will be crafted to facilitate interaction with the predictive system, enabling users to effectively utilize its capabilities.

Data Sample for One Flood Point:

High Water Mark ID:	P-0002
Watershed:	GREENS BAYOU
Benchmark Elevation:	27.50
ADJUST	0.00
SE10YR	15.70
SE50YR	18.90
SE100YR	20.40
SE500YR	23.30
Flooding Date:	6/20/2006, 12:00 AM
Field Check Date:	6/20/2006, 9:27 AM
High Water Mark Elevation:	5.00
Longitude	-95.21
Latitude	29.78
Top of Pipe Elevation:	27.15