UN CLIMATE CHALLENGE: PREDICTING & ALLEVIATING ROAD FLOODING IN SENEGAL

Keyan Halperin
Harvard

Lingzi Hong
UMD

Brendon Machado
Georgia Tech

Ricardo Macias
Georgia Tech

Mentors: Dr. Bistra Dilkina, Caleb Robinson, Amrita Gupta
UN Challenge

• Data for Climate aims to leverage big data to identify revolutionary new approaches to climate mitigation and adaptation.

Data Sources Provided by UN

- BBVA DATA & ANALYTICS
- Crimson Hexagon
- EARTH NETWORKS
- Nielsen
- Orange
- Planet
- PlumeLabs
- Schneider Electric
- Waze

Data Scientists

Data Analysts

Researchers

Innovators
Motivation: Climate Mitigation & Transportation

- Climate change has the potential to raise the risk of flood for coastal countries, which can in turn affect the living environment and threaten the success of crucial development schemes.
Research Questions

Our goal is to determine both **which segments of the roads are under risk of flooding** and **which roads should be targeted for repairs**.

- Develop a model to determine the **flooding risk of each road segment**, based on historical weather and topographical data.

- Model the volume of traffic of each road to quantify its **contribution to accessibility** between different parts of the country.
Flood Risk Prediction - Target

- To predict road flood risk, we need to evaluate the flooding possibility of the areas where each road goes through.

- We created a grid of 1km*1km cells as units of analysis to build the flood prediction model. The target will be the occurrence of flooding every biweekly period in each grid cell.
Flood Risk Prediction - Framework

**Features**

- **Physical Terrain:** statistics of slopes and elevations in the area
- **Land Cover:** percentage of each land cover type defined by LCCS
- **Hydrology:** distance to waterways intersection with water areas
- **Urban or Rural:** population
- **Weather History:** statistics of temperature, precipitation, dew point etc..

**Model**

- **ML Model:** Random Forest, XGBoost, Support Vector Machines
  - training: 2012-07 ~ 2016-04
  - testing: 2016-04 ~ 2017-04
  - 5-fold cross validation for tuning parameters

**Evaluation**

- classification: ROC, F1, recall, precision

**Predict**

- **Predict:**
  - The probability of each grid area being flooded given certain weather conditions at biweekly time point
  - Understand the relation between weather and flooding
  - Evaluate the long term flooding risk based on dynamic weather condition
Flood Risk Prediction - Data

- distance to the nearest intersection area
- water ways
- water area
- land cover
- average by distance weighted
- weather
- elevation
- slope
- population
- raster zonal statistics
Flood Risk Prediction - Evaluation

- **Recall**: Given that the cell is flooded, what is the probability we will predict it as flooded?

- **Precision**: Given that the cell is predicted as flooded, what is the probability that it is actually flooded?

<table>
<thead>
<tr>
<th>Season</th>
<th>#Samples (flooded/not)</th>
<th>ROC</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>#Samples (flooded/not)</th>
<th>Accuracy</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry</td>
<td>6483 / 61946</td>
<td>0.9145</td>
<td>0.8019</td>
<td>0.8743</td>
<td>1622 / 24413</td>
<td>0.7767</td>
<td>0.4969</td>
<td>0.1712</td>
</tr>
<tr>
<td>Rainy</td>
<td>7613 / 67492</td>
<td>0.9358</td>
<td>0.8956</td>
<td>0.8186</td>
<td>1616 / 20081</td>
<td>0.9160</td>
<td>0.7085</td>
<td>0.4587</td>
</tr>
</tbody>
</table>

**Rainy Season**: Jun - Nov  
**Dry Season**: Dec - May
# Flood Risk Prediction – Weather

- **Most Important Features** based on the Random Forest model

<table>
<thead>
<tr>
<th>Rainy Season Model</th>
<th>Dry Season Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Feature</strong></td>
<td><strong>Meaning</strong></td>
</tr>
<tr>
<td>e_mean</td>
<td>mean elevation</td>
</tr>
<tr>
<td>SD.TEMP.14</td>
<td>variance of temperature in the previous 14 days</td>
</tr>
<tr>
<td>sumarea</td>
<td>sum intersection area with water body</td>
</tr>
<tr>
<td>s_stdev</td>
<td>variance of slopes</td>
</tr>
<tr>
<td>MAX.TEMP.30</td>
<td>max temperature of the last month</td>
</tr>
<tr>
<td>SD.PRCP.14</td>
<td>variance of precipitation in the last 14 days</td>
</tr>
<tr>
<td>TEMP.14</td>
<td>mean temperature of the last 14 days</td>
</tr>
<tr>
<td>cd170</td>
<td>percentage of wetland</td>
</tr>
<tr>
<td>PRCP.14</td>
<td>mean precipitation in the last 14 days</td>
</tr>
<tr>
<td>s_mean</td>
<td>mean slope</td>
</tr>
<tr>
<td>DEWP.14</td>
<td>mean dew point in the last 14 days</td>
</tr>
<tr>
<td>TOT.PRCP.30</td>
<td>total precipitation in the last month</td>
</tr>
<tr>
<td></td>
<td>s_stdev</td>
</tr>
<tr>
<td></td>
<td>variance of slopes</td>
</tr>
<tr>
<td></td>
<td>pop</td>
</tr>
<tr>
<td></td>
<td>population</td>
</tr>
<tr>
<td></td>
<td>s_mean</td>
</tr>
<tr>
<td></td>
<td>mean of the slopes</td>
</tr>
<tr>
<td></td>
<td>SD.DEWP.30</td>
</tr>
<tr>
<td></td>
<td>variance of dew point in the last month</td>
</tr>
<tr>
<td></td>
<td>e_mean</td>
</tr>
<tr>
<td></td>
<td>mean of the elevation</td>
</tr>
<tr>
<td></td>
<td>cd170</td>
</tr>
<tr>
<td></td>
<td>percentage of the wetland</td>
</tr>
<tr>
<td></td>
<td>DEWP.30</td>
</tr>
<tr>
<td></td>
<td>mean dew point in the last month</td>
</tr>
<tr>
<td></td>
<td>distancekm</td>
</tr>
<tr>
<td></td>
<td>distance to water ways</td>
</tr>
<tr>
<td></td>
<td>SD.TEMP.14</td>
</tr>
<tr>
<td></td>
<td>variance of temperature in the last 14 days</td>
</tr>
<tr>
<td></td>
<td>s_max</td>
</tr>
<tr>
<td></td>
<td>max slope</td>
</tr>
<tr>
<td></td>
<td>sumarea</td>
</tr>
<tr>
<td></td>
<td>sum intersection area with water body</td>
</tr>
<tr>
<td></td>
<td>s_range</td>
</tr>
<tr>
<td></td>
<td>the difference between max slope and min slope</td>
</tr>
</tbody>
</table>
Flood Risk Prediction - Visualization

- Flood Risk Map for the last year
  - Flood risk: frequency of biweekly flooding
  - Correlation = 0.69

Historical Map

Predicted Map
Road Importance - Dataset

- Locations of 1623 cell phone towers, scattered throughout Senegal
- Time and date of use of each tower for each customer
- 300,000 customers, 50 weeks => 105,000,000 samples of customer movements between cell towers
Road Importance - Method

• Road Traffic
  • For each customer movement record between two cell towers, we find the shortest path between two locations within the two cell towers coverage areas.
  • Repeat for all 105 million movement records

• Road Importance
  • We figure out how much worse an alternate route would be for each movement if one of the roads are flooded.
Road Importance – Senegal Road Traffic
Road Importance

Traffic

Flood Risk
Road Importance - Optimization

- Due to our limited budget, we must find which roads should be pre-emptively fortified or repaired first.
- The idea is to maximize accessibility while minimizing costs.
Outcomes

• A model that predicts biweekly flood risk
  • Based on historical weather data and topographic features

• An algorithm that can suggest roads to repair
  • Makes suggestions based on flood risk, budget, and importance of the road in relation to accessibility of surrounding regions

• These models can help decision makers to make more efficient strategies regarding climate mitigation for transportation.
Thank You!