# Flooding Project Team - Summary of Mathematical Modeling Flood Risks Report.

*COGS-269 Global Good Studio*
*Updated: 04/29/2020*

**Alex John Quijano**
*Applied Mathematics, University of California Merced*
**Madeline Brown**
*Environmental Systems, University of California Merced*
**Umesh Krishnamurthy**
*Cognitive and Information Sciences, University of California Merced*
**Ritesh Sharma**
*Electrical Engineering and Computer Science, University of California Merced*

---

**Terminology**
**IMERG** - Integrated Multi-satellitE Retrievals for GPM
**SMAP** - Soil Moisture Active Passive
**DEM** - Digital Elevation Model
**ML** - Machine Learning
**MCDM** - Multi-Criteria Decision Making
**VIKOR** - Vise kriterijumska optimizacijaik ompromisno Resenje
**SAW** - Simple Additive Weighting
**TOPSIS** - Technique for Order Preference by Similarity to Ideal Solution
**NBT** - Naive Bayes Tree
**NB** - Naive Bayes
**NDVI** - Normalized Difference Vegetation Index
**STI** - Stream Transport Index
**TWI** - Topographic Wetness Index
**SPI** - Stream Power Index
**GARP** - Genetic Algorithm Rule-Set Production
**QUEST** - Quick Unbiased Efficient Statistical Tree

---

**Goal:** To predict the flood risk of a particular area such that it can be used to create a flood alert map.

**Literature Review:**

1. Khosravi, Khabat, et al. **"A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods."** Journal of hydrology 573 (2019): 311-323.

***Methods.*** This paper discusses the methods of modeling flood susceptibility using Multi-Criteria Decision Making (MCDM) analysis techniques and Machine Learning (ML) techniques.

A. MCDM techniques
    a. Vise kriterijumska optimizacijaik ompromisno Resenje (VIKOR)
    b. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)
    c. Simple Additive Weighting (SAW)
B. ML techniques
    a. Naive Bayes tree (NBT)
    b. Naive Bayes (NB)

***Variables.*** There are multiple factors included in the models. The factors are selected using the Multicollinearity diagnosis test and Information Gain Ratio (IGR) test to choose which factors are most important to predicting flood susceptibility. The following is what they found to be significant factors in order of importance.

I. Altitude
II. Distance from river
III. Normalized Difference Vegetation Index (NDVI)
IV. Soil type
V. Ground slope
VI. Lithology (or characteristics of rocks)
VII. Land-use
VIII. Stream Transport Index (STI)
IX. Topographic Wetness Index (TWI)
X. Rainfall
XI. Stream Power Index (SPI)
XII. Curvature (no effect on flood occurrence)

***Flood Data.*** For the output, they used a flood inventory map which is a record or past flood occurrences. In their study, 166 flood events in Ningdu County, Jiangxi

Province, China were considered and were validated using field surveys and records through the use of Global Positioning System (GPS). The dataset was separated into a training set (70% of the data) and test set (30% of the data). This dataset was given by the Chinese Academy of Science.

***Results.*** The results from their paper indicate that the NBT method has the highest Area under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve which AUC=0.984. This means that the classification model using NBT is 98.4% accurate on predicting floods. Below is a list of methods in order of AUC values.

1. NBT (AUC=0.984)
2. NB (AUC=0.979)
3. SAW (AUC=0.97)
4. TOPSIS (AUC=0.986)
5. VIKOR (AUC=0.965)

2. Darabi, Hamid, et al. **"Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques."** Journal of hydrology 569 (2019): 142-154.

***Methods.*** This paper discusses the methods of mapping the urban flood risk using two machine learning models. The two methods are Genetic Algorithm Rule-Set Production (GARP) and Quick Unbiased Efficient Statistical Tree (QUEST). The main take point they do is to compute risk using two kinds of conditional factors.

*Risk = Hazard x Vulnerability*

Below is a short description of each machine learning method.

1. GARP - A machine learning algorithm based on a genetic algorithm that models genetic evolution. It analyzes the relationship between flood inundation data and the variables. It is an iterative process with conditional rules.
2. QUEST - A data-mining model that produces subsets of the data. This is a tree-structured classification algorithm. The structure has a growing binary-split decision tree that utilizes unbiased linear discriminant analysis method in splitting these trees. This method is similar to a decision tree algorithm but with an unbiased approach to selecting variables that best predict flood events.

**Variables.** The hazard factors considered in their models include the following.

I. Rainfall
II. Land-use
III. Elevation
IV. Slope percent
V. Curve number
VI. Distance to river
VII. Distance to channel
VIII. Depth groundwater

They also included vulnerability factors determined by socio-economic processes. These factors include the following.

I. Urban density (classified into “high”, “medium”, “low”, “very low”)
II. Building quality (classified into “very high”, “high”, “medium”, “low”, “very low”, “no building”)
III. Building age (classified into “very old”, “old”, “medium”, “new”, “newest”, “no building”)
IV. Population density (classified into “high”, “medium”, “low”, “very low”)
V. Socio-economic conditions (divided into five classes - “very good”, “good”, “moderate”, “weak”, “very weak”)

**Flood Data.** Flood inventory map in Sari City, Iran during 2015-2017 was used. The area was marked flooded (1) and non-flooded (0). Historical records on flood occurrence were also considered as essential information. There were 113 recorded flood areas while 76 were recorded non-flooded that were randomly chosen. In total there are 189 data points. Flooded areas were divided into training and test sets. 70% were used as training sets and 30% were used as test sets. Similarly with the non-flooded areas.

***Results.*** The results from their paper indicate that the GARP has AUC of 93.5% followed by QUEST with 89.2%.