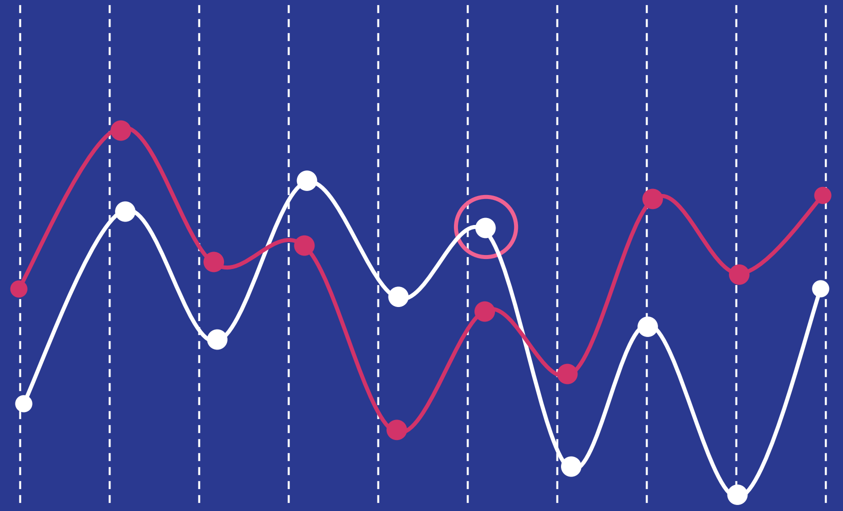
CEW613: Hydrological Modelling & Software Development

Survey : Flood Prediction Methodologies



Niharika Vadlamudi **2018122008**

Contents

Introduction

- The motivation of the project is to study various approaches for Flood prediction . Over the time , we have seen that these techniques have evolved from traditional runoff models to ML based models .

- **Tool Motivation** : The motivation of the tool is to model river discharge which is a high dimension problem with complex interactions . We are trying to model this coupled dynamic model with statistical data and inferences and using suggested Machine Learning methodologies .

Traditional Prediction Methods

- **Discharge Frequency Analysis** : This approach relies on the existence of long records of accurate river discharge measurements. After identification of annual discharge maxima and other data points,the analyst fits several statistical distributions and selects the distribution that most accurately describes the data .
- **Rainfall Runoff Model** : Hydrologic models can be used to convert estimates of extreme rainfall into design discharge estimates and design hydrographs.To do so, they must represent the movement of water across the landscape (a process known as runoff) and into the river channel.
    - Lumped Models
    - Distributed Models

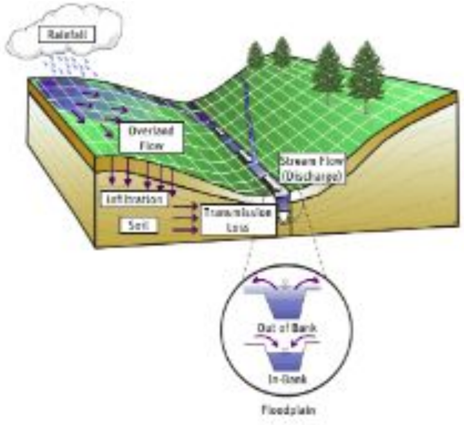


Figure 2: Rainfall-Runoff Diagram

ML Models

To mimic the complex mathematical expressions of physical processes of floods, during the past two decades, machine learning (ML) methods have highly contributed to the advancement of prediction systems providing better performance and cost effective solutions. Few popular techniques :

- Artificial Neural Networks
- Support Vector Machine
- Wavelet Neural Networks
- Decision Trees

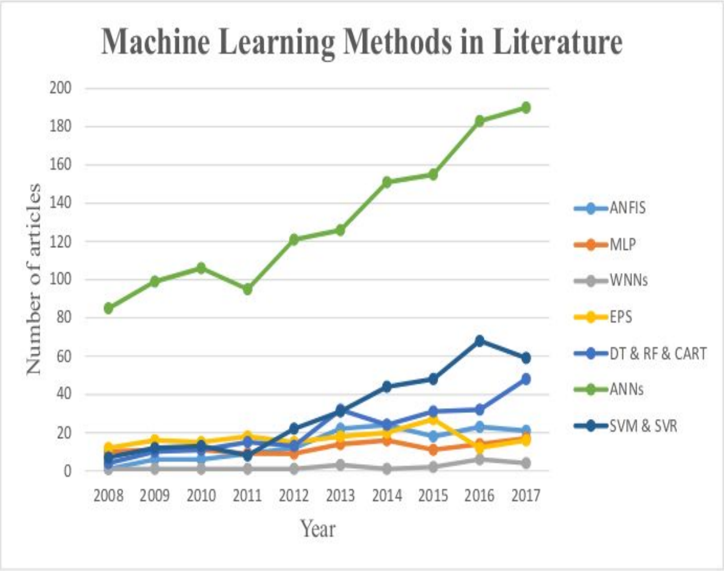


Figure 3. Major ML methods used for flood prediction in literature. Reference year 2008. Source: Scopus

GLOFAS - ERA5 Reanalysis Dataset

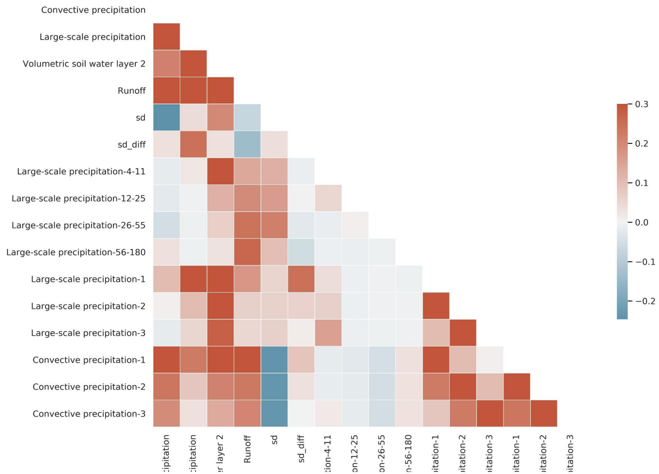
The entire dataset spans from 1981 - 2016 , containing daily dataset for a particular region . Total : 12490 x 13 data points.

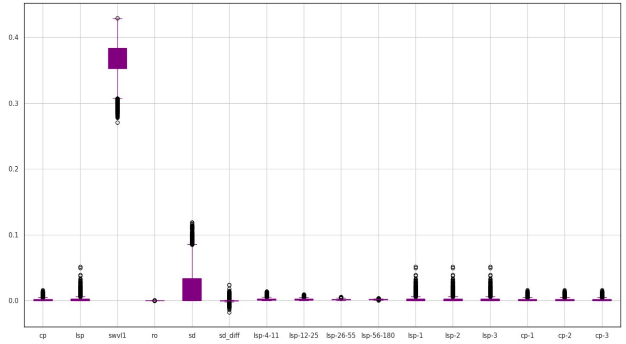
- Input : High dimensional spatial averages of features involving convectional precipitation , snow depth ,etc .
- Output: River Discharge at specific point(m3 s-1)
- River discharge, or river flow as it is also known, is defined as the amount of water that flows through a river section at a given time.The value is an average over a 24-hour period.
- Data Source : CDS Copernicus Dataset Website

URL:https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-historical?tab=overview

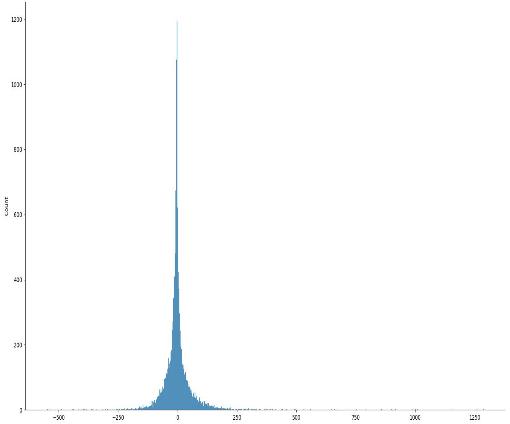
Feature Engineering

- To cut down on few features from the highly dimensional data , correlation analysis , box-plot analysis & quantile method was used .
- For a feature , we had multiple divisions.Eg : Convectional Precipitation - (Cp 1, Cp2 , Cp3) .
- Finally the model takes only 5 parameters - Snow Depth ,Convectional Precipitation , Soil Water Content Layer -1 , Runoff & Large Scale Precipitation .



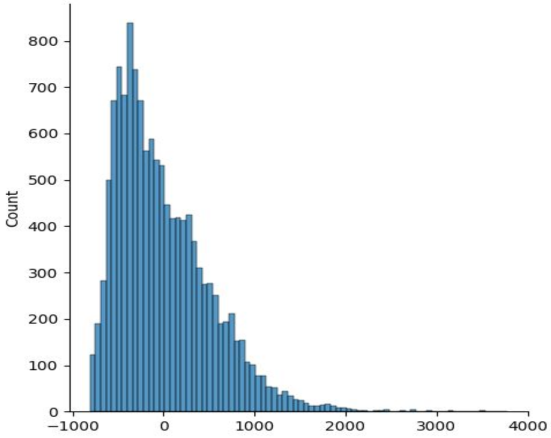


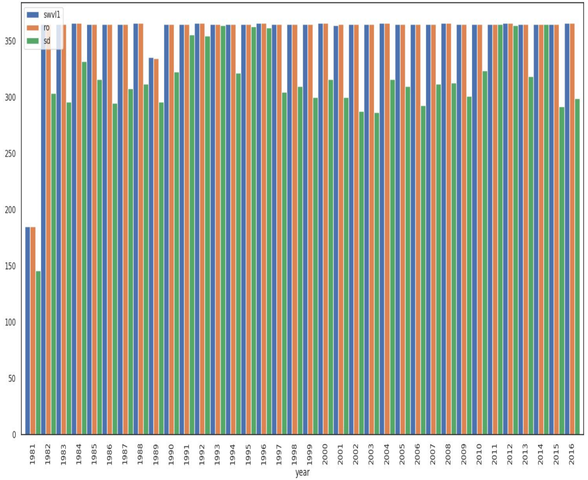
More Graphs on Data



Difference Discharge Plot

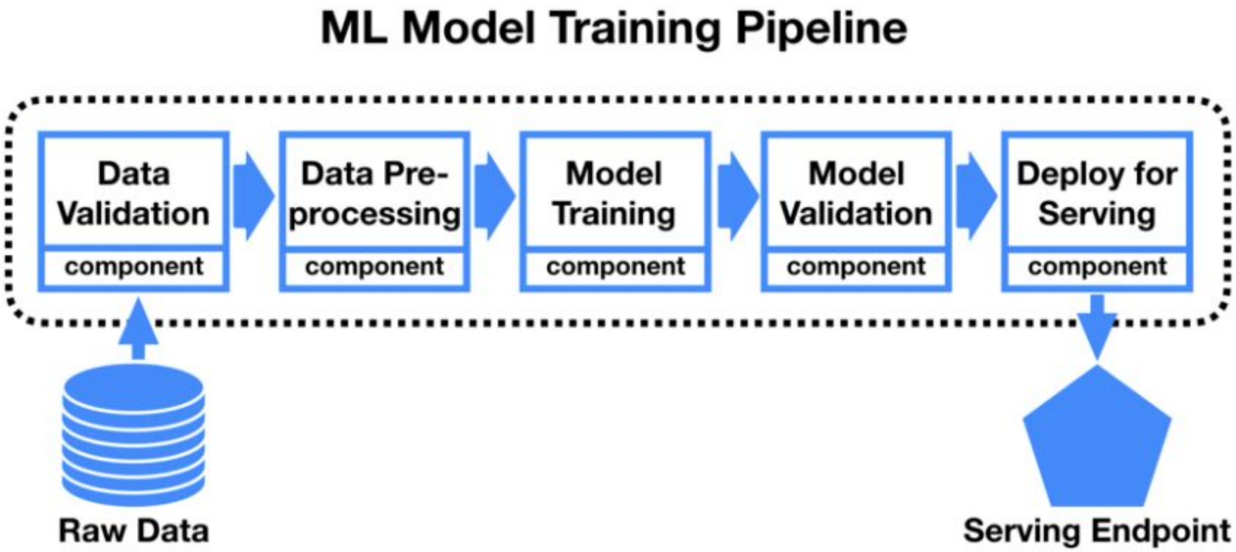
Normalised Discharge Plot





Bar Graph Analysis of Parameters

Model Pipeline Design

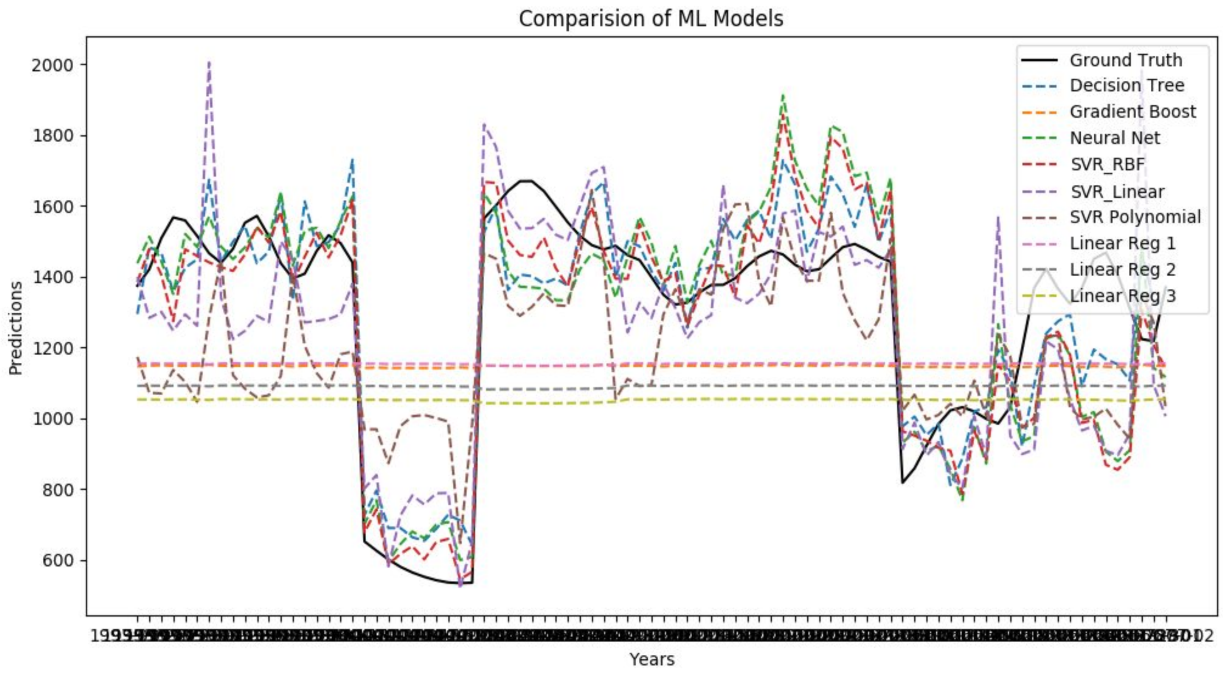


Models Experiments

- To analyse the ML models on the modelling of Flood Prediction , I’ve created a tool using the following ML models :

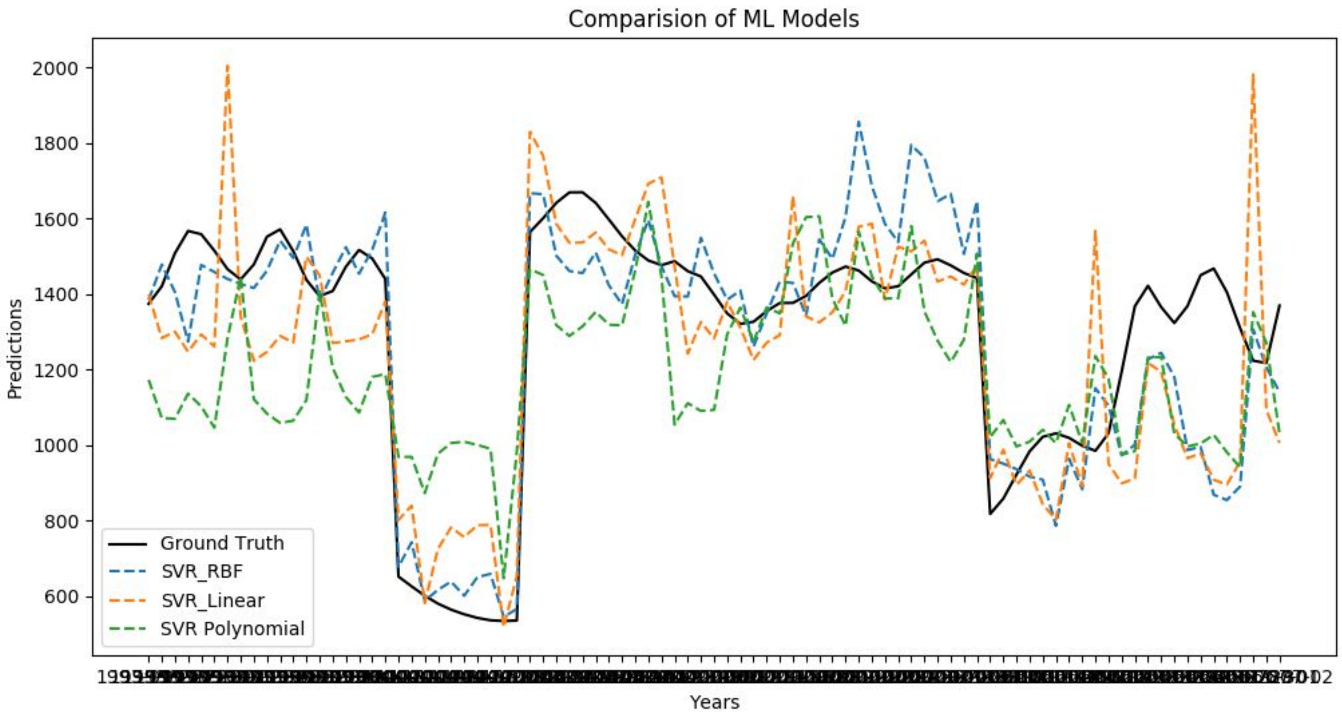
- Linear Regression (Various Degrees)
- Support Vector Regression (RBF , Linear , Polynomial)
- Neural Networks
- Decision Tree
- Random Forest
- Gradient Boost

Comparative Analysis

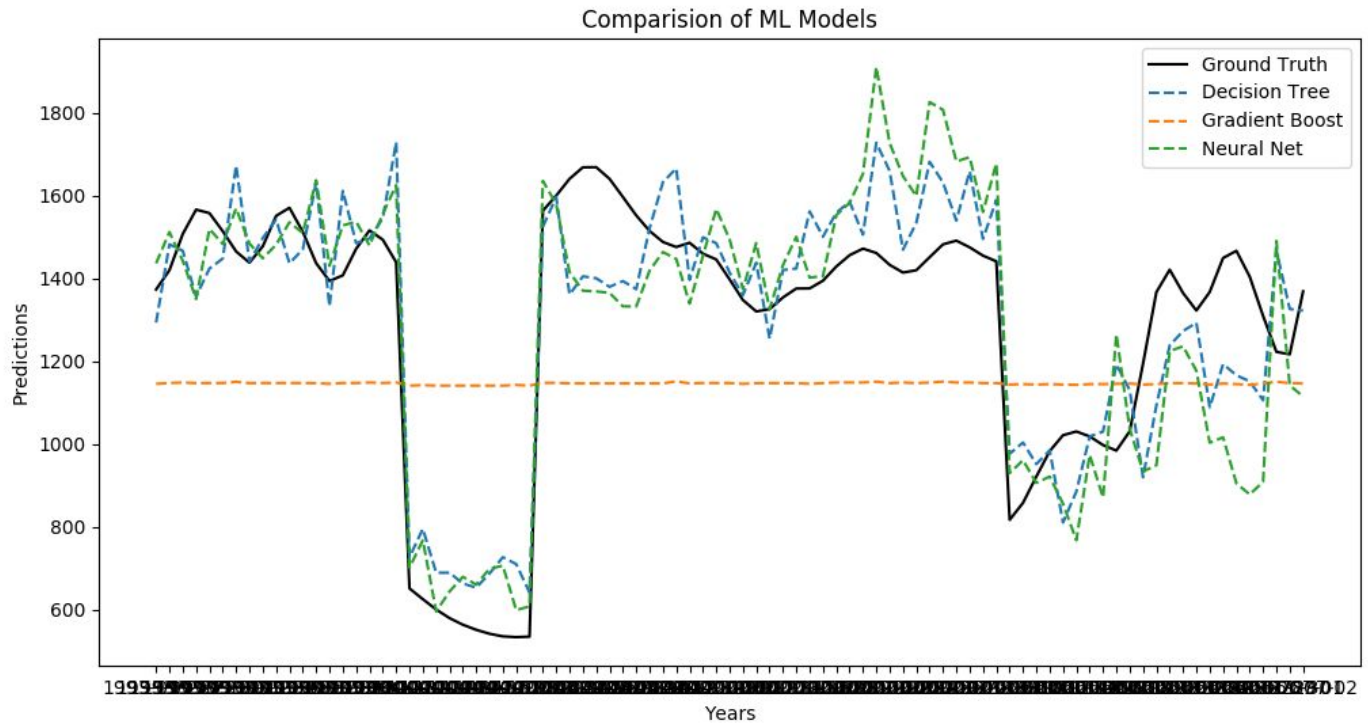


Support Vector Regressor with RBF outperforms the rest.

Comparative Analysis (Contd.)



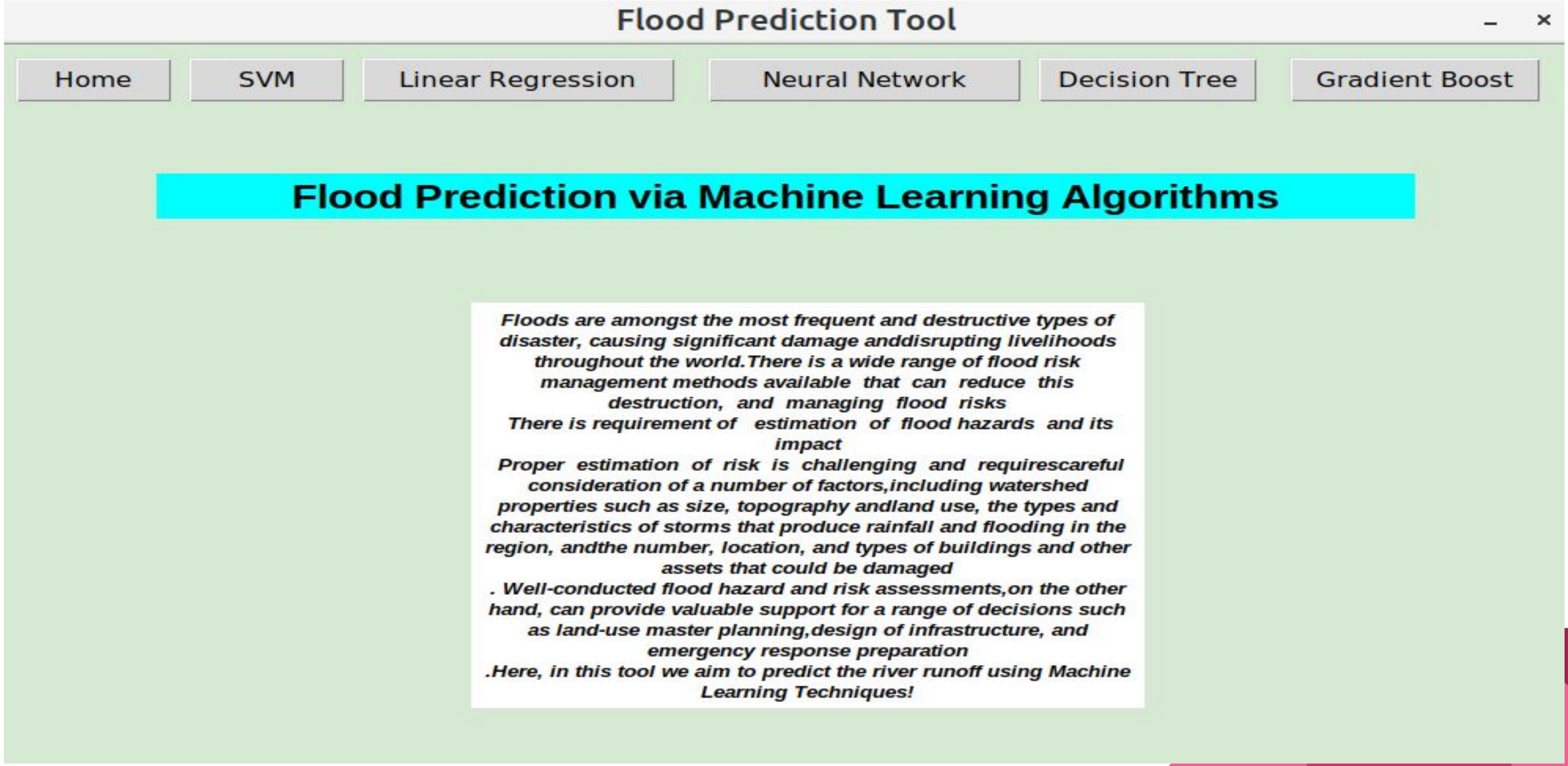
Comparative Analysis (Contd.)



Tabular Results

Machine Learning Model	MSE Loss (Test Dataset)
SVR-Radial Basis Function Kernel	21.25
SVR - Linear Kernel	62.34
SVR - Polynomial	106.25
Linear Regression (Degree 1)	158.19
Linear Regression (Degree 2)	131.23
Linear Regression (Degree 3)	127.93
Decision Tree	37.087
Random Forest	48.68
Neural Network	32.81
Gradient Boost	188.24

Tool Demo


Flood Prediction Tool
Home
SVM
Linear Regression
Neural Network
Decision Tree
Gradient Boost
Flood Prediction via Machine Learning Algorithms
Floods are amongst the most frequent and destructive types of disaster, causing significant damage anddisrupting livelihoods throughout the world.There is a wide range of flood risk management methods available that can reduce this destruction, and managing flood risks
There is requirement of estimation of flood hazards and its impact
Proper estimation of risk is challenging and requirescareful consideration of a number of factors,including watershed properties such as size, topography andland use, the types and characteristics of storms that produce rainfall and flooding in the region, andthe number, location, and types of buildings and other assets that could be damaged
. Well-conducted flood hazard and risk assessments,on the other hand, can provide valuable support for a range of decisions such as land-use master planning,design of infrastructure, and emergency response preparation
.Here, in this tool we aim to predict the river runoff using Machine Learning Techniques!

Flood Risk Prevention

- Once a flood hazard and vulnerability assessment has been completed, it is relatively simple to arrive at an estimate of flood risk. It is clear that completely “correct” risk calculations can only be accomplished when the probability and the magnitude of the loss can be estimated with complete accuracy. Making flood risk management decisions with no or poor risk information can result either in too little or too much protection.
- Too little protection means that citizens or economic assets face continued exposure to flood impacts, while too much protection means that money has been unnecessarily spent on unneeded protection.

Conclusion

- A major challenge in flood hazard and risk assessment is to understand the uncertainties that exist at every stage of the process, and to decide how to incorporate these uncertainties into subsequent risk management decisions
- For example, the estimation of design discharges, whether done using statistical methods or rainfall- runoff models, always depends on the use of multiple assumptions, incomplete datasets, and imperfect models. This will lead to errors in design discharges, which will in turn lead to errors in the water. So, its best to pair up with traditional models and keep them in check .
- The analysis was done on North American Dataset, we can apply the same for Indian Domain . (Few parameters like Snow Depth , will change depending on the region)

Thank You