# **FLOOD PREDICTION**

Name : Y. Pavan Ganesh Reddy

Reg No : 39110758

College : Sathyabama Institute of Science and Technology

Branch: B.E, CSE.

Project Title: Flood prediction using Machine Learning.

# **INDEX**

- 1. INTRODUCTION
  - a. Overview
  - b. Purpose
- 2. LITERATURE SURVEY
  - a. Existing Solution
  - b. Proposed solution
- 3. THEORETICAL ANALYSIS
- 4. HARDWARE / SOFTWARE REQUIREMENTS
- 5. FLOW CHART
- 6. RESULT
- 7. ADVANTAGES AND DISADVANTAGES
- 8. APPLICATIONS
- 9. CONCLUSION
- 10. FEATURE SCOPE
- 11. BIBILOGRAPHY

### 12. APPENDIX

- a. Source code
- b. UI output Screen-shots

### INTRODUCTION

Floods are inevitable, but with timely alerts, their effects can be minimized. There are many people who die every year due to devastating floods, the number of people become homeless and many people die due to lack of proper help after a flood. The lack of timely alerts has always been an issue concerning it.

Delay in alerts in flood-prone areas is the biggest loophole of an economy. Conventional systems run a little low in forecasting floods at the right time so that proper actions could be taken before any disaster.

By using machine learning we can predict floods or forecast floods with better accuracy. This project aims at building predictive modeling based on the historical weather data of particular areas in order to predict the occurrence of floods.

The predictive model is built on different machine learning algorithms. The concerned authority monitor this flood prediction system through a web application.

# LITERATURE SURVEY

For this project, we need to use classification algorithms as the output we will get is categorical values which is either floods will occur or not so, we need to build the most accurate classification algorithm to get the desired output

In this project,we will be using classification algorithms such as Decision tree, Random forest, KNN, and xgboost. We will train and test the data with these algorithms. From this best model is selected and saved in save format. We will be doing flask integration and cloud deployment.

# THEORETICAL ANALYSIS

- 1. Install Required Libraries.
  - . Collect the dataset or Create the dataset
- 2. Data collections
  - Import the Libraries.
- 3. DataPreprocessing
  - · Importing the dataset.
- Understanding Data Type and Summary of features.
- · Take care of missing data
- · Data Visualization.
- · Drop the column from DataFrame & replace the missing value.
- Splitting the Dataset into Dependent and Independent variables
- Splitting Data into Train and Test.
- 4. Model Building
- · Training and testing the model
- · Evaluation of Model
- · Saving the Model
- 5. Application Building
  - · Create an HTML file
- · Build a Python Code
- 6. Final UI
  - · Dashboard Of the flask app.

# HARDWARE / SOFTWARE REQUIREMENTS

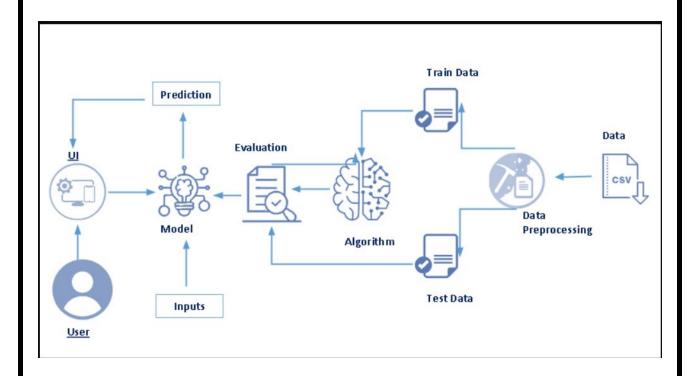
Category: Machine Learning

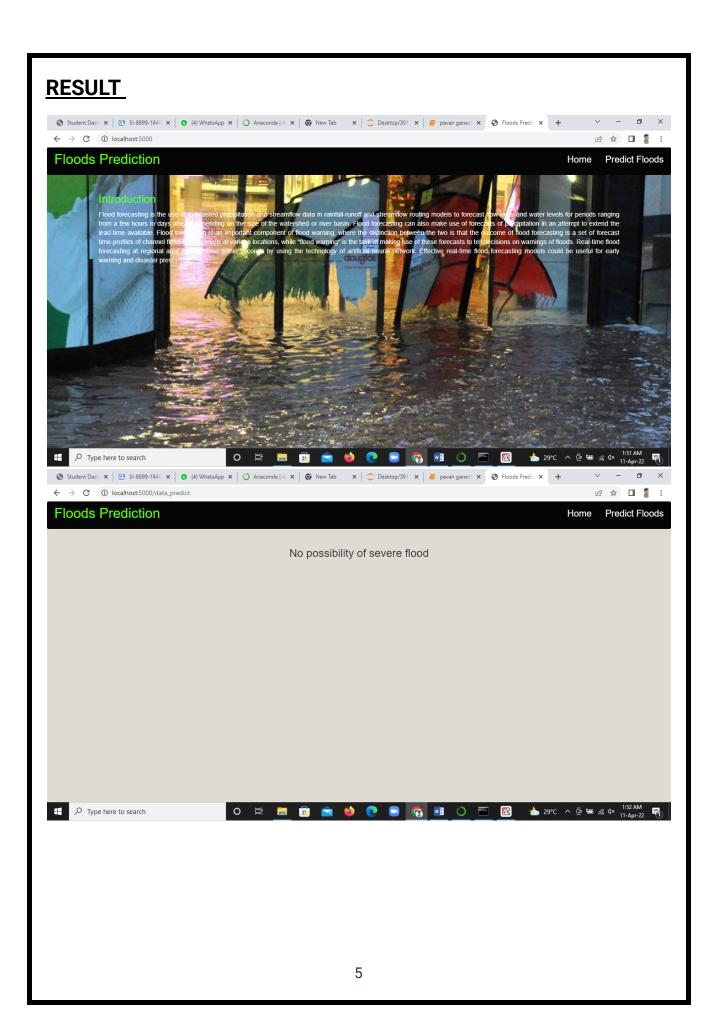
Skills Required:

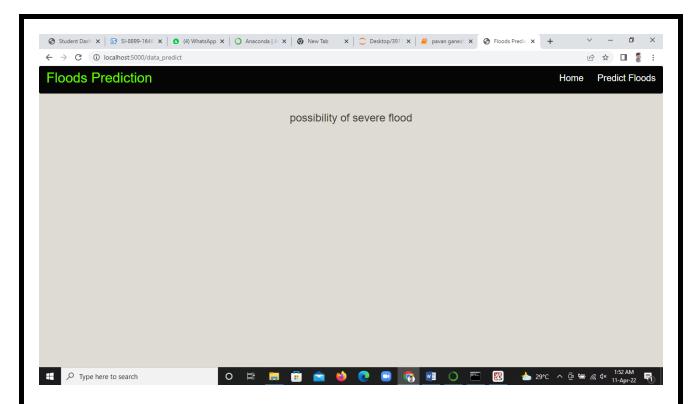
Python,Python Web Frame Works,Python For Data Analysis,Exploratory Data Analysis,Data Preprocessing Techniques,Machine Learning

While working on the solution, I have seen the dataset of the floods and its quite frequent to see the whether forecast change drastically and cloud cover increases, rainfalls in the months April to June and lastly floods occur. There is interconnection of all the factors which leads to floods.

# **FLOW CHART**







# ADVANTAGES AND DISADVANTAGES

### **ADVANTAGES:**

- ->The main purpose of flood warning is to save life by allowing people, support and emergency services time to prepare for flooding.
- ->The model which I built is 96 % accurate so one can rely on the prediction done by the model.
- ->With the help of this model , you can determine and manage environmental and water resource systems.
- ->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, which can help to predict how extensive and damaging a flood might become.

#### DISADVANTAGES:

->One of the algorithm used in this model is Random Forest, where it can become a plot as it gets prediction accuracy on complex

problems is usually inferior to Gradient - Boosted Trees.

# **APPLICATIONS**

The applications in flood prediction can be classified according to flood resource variables, i.e., water level, river flood, soil moisture, rainfall—discharge, precipitation, river inflow, peak flow, river flow, rainfall—runoff, flash flood, rainfall, streamflow, seasonal stream flow.

# **CONCLUSION**

The summary of the project is that even though Floods are inevitable, they can be prevented with accurate measures like predicting the outcomes, changes to be taken to be on the safe side like building dams, reservoirs etc..and especially the technology has taken huge turn so, it will be matter of time before we gain control on what we are challenged.

## **FUTURE SCOPE**

There can be future changes can be made to my model by adding the data like water level, soil moisture, average temperature of the area, rainfall - discharge, precipitation, drainage systems in the area etc.... in the dataset makes even more accurate and since the data will be enhanced, the analysis becomes the key part in the prediction.

## **BIBILOGRAPHY**

There are some references which are to be considered while making the project.

URL - 1 : Flood prediction - Oxford Reference

URL - 2 : Flood Forecasting - an overview | ScienceDirect Topics

#### APPENDIX - SOURCE CODE 😚 Student x 🔯 SI-8899 x | 🐧 (4) What x | 🐧 Anaconc x | 🚱 New Tab x | 💍 Desktop x | 🤌 Pavan g x 🚱 Flood p x x | 💽 Flood p x x | 🗜 Flood p x x | + V - G X 🗧 🗦 🖰 🕦 localhost:888/notebooks/Desktop/39110758-Y%20PAVAN%20GANESH%20REDDY-FLOOD%20PREDICTION%20USING%20MACHINE%20LEARNING/Training/pavan%2... 🖻 🛣 🚦 🚦 Jupyter pavan ganesh 7889 Last Checkpoint: 8 hours ago (autosaved) Logout File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel) O operable program or batch file. In [1]: #import required libraries #Import requires thoruntes import numpy as np #for dealing high demensional data import pandas as pd #to do statistical data analysis import matpollib.pyplot as plt #for 2D visualization import seaborn as sns #High end data visualization In [2]: #read the dataset dataset=pd.read\_excel(r'C:\Users\Pavan\Desktop\Y PAVAN GANESH-FLOOD PREDICTION\DATASET\flood dataset7889.xlsx') In [3]: #check the first 5 observations dataset.head() Out[3]: Temp Humidity Cloud Cover ANNUAL Jan-Feb Mar-May Jun-Sep Oct-Dec **0** 29 70 30 3248.6 73.4 386.2 2122.8 666.1 274.866667 649.9 75 40 3326.6 9.3 275.7 2403.4 638.2 130.300000 256.4 **2** 28 75 42 3271.2 21.7 336.3 2343.0 570.1 186.200000 308.9 3 29 71 44 3129.7 26.7 339.4 2398.2 365.3 366.066667 862.5 0 4 31 74 40 2741.6 23.4 378.5 1881.5 458.1 283.400000 586.9 0 In [4]: print(dataset.shape) (115, 11) In [5]: print(dataset.info()) Type here to search 29°C ∧ @ / 4× 11-Apr-22 O 🛱 🤚 **E**S 🔞 Student x | 💽 SI-8899- X | 🐧 (4) What X | 🐧 Anacons x | 🐧 New Tab X | 🔘 Desktop X | 🤌 pavan g x X (4) Tool Fix X | Tool Fix X ← → C ① localhost:8888/notebooks/Desktop/39110758-Y%20PAVAN%20GANESH%20REDDY-FLOOD%20PREDICTION%20USING%20MACHINE%20LEARNING/Training/pavan%2... 🖻 🖈 🔲 🥻 : Jupyter pavan ganesh 7889 Last Checkpoint: 8 hours ago (autosaved) Logout File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel) O + Run ■ C >> Code In [5]: print(dataset.info()) <class 'pandas.core.frame.DataFrame'> RangeIndex: 115 entries, 0 to 114 Data columns (total 11 columns): Non-Null Count Dtype # Column 115 non-null 115 non-null Humidity 115 non-null Cloud Cover 115 non-null ANNUAL 115 non-null int64 int64 ANNUAL float64 115 non-null 115 non-null 115 non-null 115 non-null 115 non-null 115 non-null Jan-Feb float64 Mar-May Jun-Sep Oct-Dec float64 float64 float64 avgjune sub 9 sub 115 non-null 10 flood 115 non-null dtypes: float64(7), int64(4) memory usage: 10.0 KB float64 In [7]: dataset.describe().T Out[7]: 8

