

**INTERNSHIP REPORT APPROVAL FORM**

July 1, 2019

With immense pleasure, this is to approve that the students of Sona College of Technology, Salem i.,e

**Kalkeseetharaman PK,**

**Harshaa A,**

**Indhumathi MP** and

**Aruna V**

Successfully completed their Project and Project Report on **“Flood Prediction”** under our guidance.

We are highly impressed with the work that they have done and commend them on their quick grasping skills. They have shown good intent to learn and have put the knowledge gained into application in the from of this project. We appreciate the hard work and commitment shown by them.

We, hereby approve that this document is completely checked and accepted by SmartBridge Technical Team. It’s been an absolute pleasure to educate and mentor these students. We hope that this document will also serve as a Letter of Recommendation, to whomsoever applied.

We wish them success in all future endeavors and a great career ahead.

**GD Abhishek**

AI Developer

**1.FLOOD FORECASTING USING MACHINE LEARNING METHODS**

**1.1 INTRODUCTION**

Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions) and self-correction. Particular applications of AI include [expert systems](https://searchenterpriseai.techtarget.com/definition/expert-system), [speech recognition](https://searchcrm.techtarget.com/definition/voice-recognition) and [machine vision](https://searchenterpriseai.techtarget.com/definition/machine-vision-computer-vision).

AI can be categorized as either [weak](https://searchenterpriseai.techtarget.com/definition/narrow-AI-weak-AI) or [strong](https://searchenterpriseai.techtarget.com/definition/artificial-general-intelligence-AGI). Weak AI, also known as narrow AI, is an AI system that is designed and trained for a particular task. Virtual personal assistants, such as Apple's Siri, are a form of weak AI. Strong AI, also known as artificial general intelligence, is an AI system with generalized human cognitive abilities. When presented with an unfamiliar task, a strong AI system is able to find a solution without human intervention.

Because hardware, software and staffing costs for AI can be expensive, many vendors are including AI components in their standard offerings, as well as access to Artificial Intelligence as a Service ([AIaaS](https://searchenterpriseai.techtarget.com/definition/Artificial-Intelligence-as-a-Service-AIaaS)) platforms. AI as a Service allows individuals and companies to experiment with AI for various business purposes and sample multiple platforms before making a commitment. Popular AI cloud offerings include [Amazon AI](https://searchaws.techtarget.com/definition/Amazon-AI)services, [IBM Watson Assistant](https://whatis.techtarget.com/definition/IBM-Watson-Assistant), [Microsoft Cognitive Services](https://whatis.techtarget.com/definition/Microsoft-Cognitive-Services) and [Google AI](https://whatis.techtarget.com/definition/Google-AI) services.

**Examples of AI technology**

1. Automation
2. Machine learning
3. Machine vision
4. Natural Language Processing
5. Robotics
6. Self driving cars

Getting into Machine Learning and AI is not an easy task, but is a critical part of data science programs. Many aspiring professionals and enthusiasts find it hard to establish a proper path into the field, given the enormous amount of resources available today. The field is evolving constantly and it is crucial that we keep up with the pace of this rapid development. In order to cope with this overwhelming speed of evolution and innovation, a good way to stay updated and knowledgeable on the advances of ML, is to engage with the community using Python is the best way because of the following reasons.The programmers of big companies use Python as it has created a mark for itself in the software development with characteristic features like-

* Interactive
* Interpreted
* Modular
* Dynamic
* Object-oriented
* Portable
* High level
* Extensible in C++ & C

**1.2 OBJECTIVES OF RESEARCH**

This is a Machine learning project approach of flood prediction depending on previous year rainfall data and other environmental conditions. Till now an accuracy of 70.08 % is obtained just by taking 4 features. If more features like temperature, cloud, humidity are added then the accuracy can be improved. We use classification based machine learning algorithm to train the model. We have predicted floods just for the state of Kerala depending on the rainfall data and other environmental conditions for one century. However this method can be used for prediction for any state in India, with the given data. The input features used are the annual rainfall of previous year, the average rainfall in the month of January to February, the average rainfall in the month of March to May , the average rainfall in first 10 days of June and the increase in rainfall from the month of May to June, for all the years from 1901 to 2015.We used the months Jan to Feb and March to May, to predict flood in the later days of the month of June and July, as in Kerala the rainy season starts from the month of June. Depending on other states, other months will be used as the features.

The **specific objectives of FLOOD-serv** are:

1. to make use of the best available data in order to identify the location and potential impacts that natural hazards as floods can have on people, property and natural environment

2. to improve the systems of warning and emergency communications

3. to provide support for the public authorities and government institutions’ hazard mitigation efforts, including planning and action coordination

4. to inform the public on the risk exposure to natural hazards and how they can get prepared, respond, recover and mitigate the impacts of such events

So given the above mentioned objectives, the project will offer an opportunity for collective problem solving, knowledge sharing, social exchange and community-wide participation at local and global scale. This will lead to an insight into the information and preparedness requirements of local communities and the development of solutions adapted to the social realities.

Secondly, it will lead to a closer cooperation and coordination for flood forecasting and warning services of public institutions based on user needs.

Thirdly, based on the flood event studies, and including consultations with affected communities and other recipients of flood warnings, improved technical means of detecting the areas at imminent risk and warning more effectively, will be developed.

Technically the project will focus on developing a collaborative platform that will link citizen, public authorities and other stakeholders and on enabling the public to be warned en masse so that actions can be taken to reduce the adverse effects of the flood.

**1.3 PROBLEM STATEMENT**

Disaster prevention and prediction Flood prediction using machine learning approach. A dataset with the amount of rainfall and if a flood had occured in a particular area/state/city, in the previous years, will be used. The dataset will have the rainfall data for a duration of 3 months approx.Using this dataset, we take average rainfall for every 10 days.We take this average data of rainfall, as input to our machine learning model and if it causes a flood or not as the output labels. We train our model and save it.(depending on some threshold value of average rainfall in the dataset).Given the input data, for consecutive 10 days, we give this data as an input, and let the model predict, if whether there is a possibility of flooding or not. This approach can be made real time prediction and accuracy can be improved with adding more features such as the type of land in that area, the location of the area etc.

**2.REVIEW OF LITERATURE**

**Richard M. Vogel, Sankarasubramanian Arumugam, in** [**Flood Forecasting**](https://www.sciencedirect.com/book/9780128018842)**, 2016-**Currently, operational flood forecasts tend to target large [watersheds](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/watershed). Most run-of-the-river hydropower systems are on [upstream](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/upstream) reaches; thus such systems are not commonly included in operational flood forecasts.

**Thomas E. AdamsIII, in** [**Indus River Basin**](https://www.sciencedirect.com/book/9780128127827)**, 2019-**The Central Water Commission (CWC) has national responsibility for monitoring flood water levels and discharges along the major rivers in India during the annual monsoon period, flood forecasts are issued to local administrative bodies, project authorities, state governments, and the Home Ministry of India.

### Forecast Dissemination Issues-Relatively long-lead time forecasts are needed in order to have sufficient time to mobilize rural, poor populations to evacuate to areas of safety. This points to general problems human societies have in relation to our complex interrelationship with our environment, both in a dependent way, in terms of fishing, agriculture, transportation, water supply, etc., and in ways that we directly influence our environment, such as using [waterways](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/waterway) for [waste disposal](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/waste-disposal), withdrawing water for [irrigation](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/irrigation), constructing large multipurpose [reservoirs](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/reservoir), and so on.

Based on the above three proposed models we developed a flood prediction model including rainfall and environmental conditions of a century as a real time data.

**3.DATA COLLECTION**

**`** The dataset are collected from the Regional Meteorological Centre, Chennai and India Meteorological Department through data request form. Through the link ([**https://www.imdchennai.gov.in/Drf.pdf**](https://www.imdchennai.gov.in/Drf.pdf)**)** we got dataset of rainfall for all the states in India with the data length of a century.From this dataset we considered the datasets of Kerala.

For other environmental parameters like temperature,humidity,cloud conditions we got data from [**http://www.imd.gov.in**](http://www.imd.gov.in/), In this dataset all the temperature, humidity and cloud were annually calculated for the state of Kerala.

From both the datasets we appended them to form a new dataset as we required. With the column names temperature,humidity,cloud,annual rainfall,average rainfall of Jan to Feb, March to May, Average of june, Rate of raise in May to June.

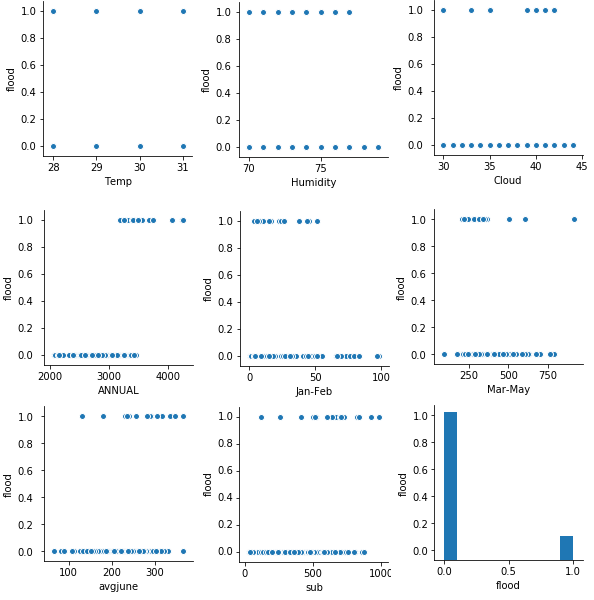
From <https://www.kerala.gov.in> we got the Dataset which tells us whether flood occurred or not for a century till 2015.We appended this to the dataset.

Hence this dataset can be used for further prediction using machine learning model.

**4.METHODOLOGY**

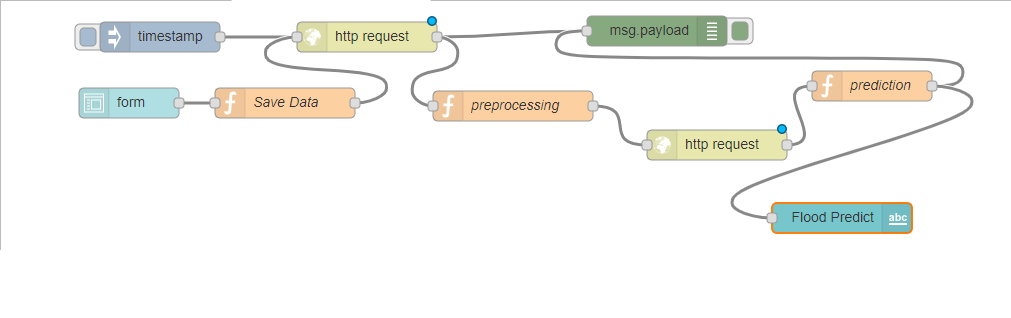
Initially we followed all the basic machine learning model steps like data reading and cleansing, seperated dependent and independent variables namely humidity, temperature, cloud, annual, Jan-Feb, March-May, Average June, Sub as independent and flood as dependent.

To know the type of problem and understand the problem more deeper we plotted graphs for all the dependent and independent variables as follows.

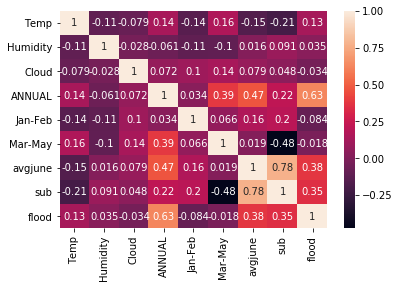


**Fig 4: Pair plot for all the variables**

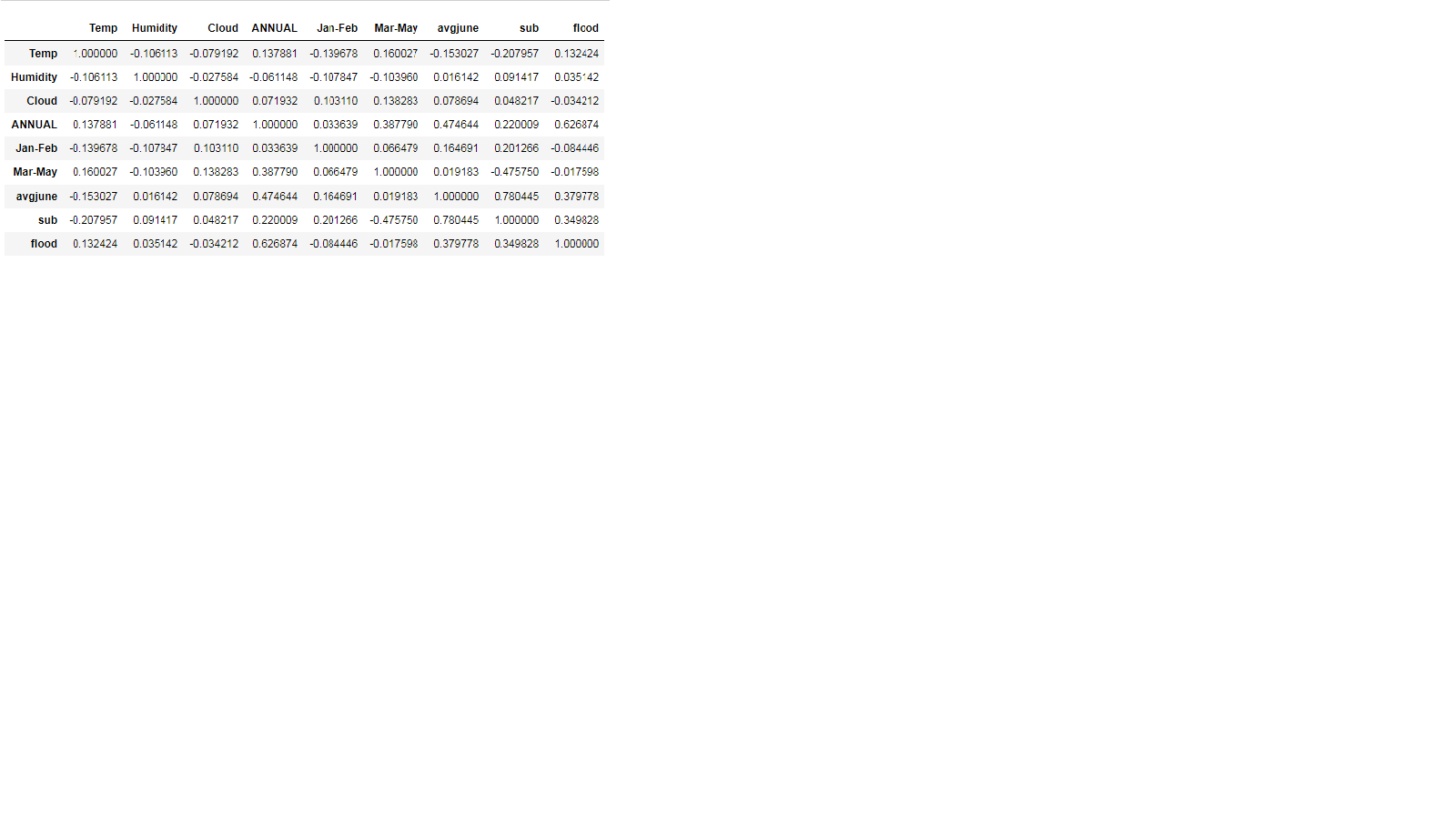
**4.1 Exploratory data analysis**

**4.1.1 Figures**

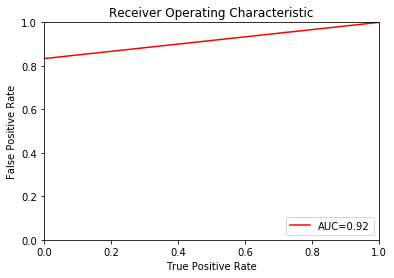
**Fig 4.1.1 Node-RED flow**

****

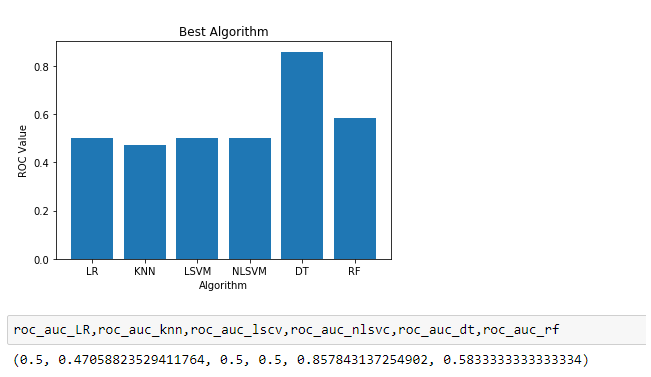
**Fig 4.1.2 Correlation Heat map**

****

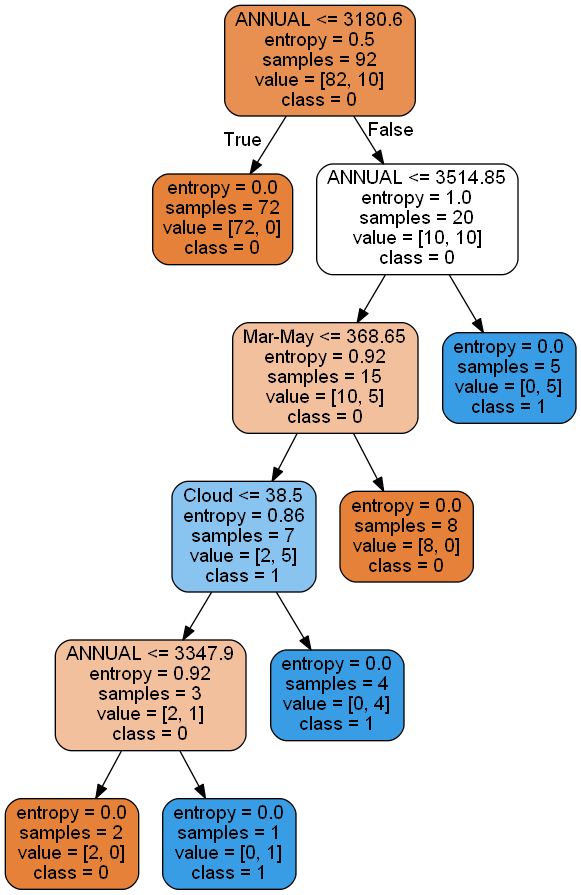
**Table 4.1:Correlation table**

****

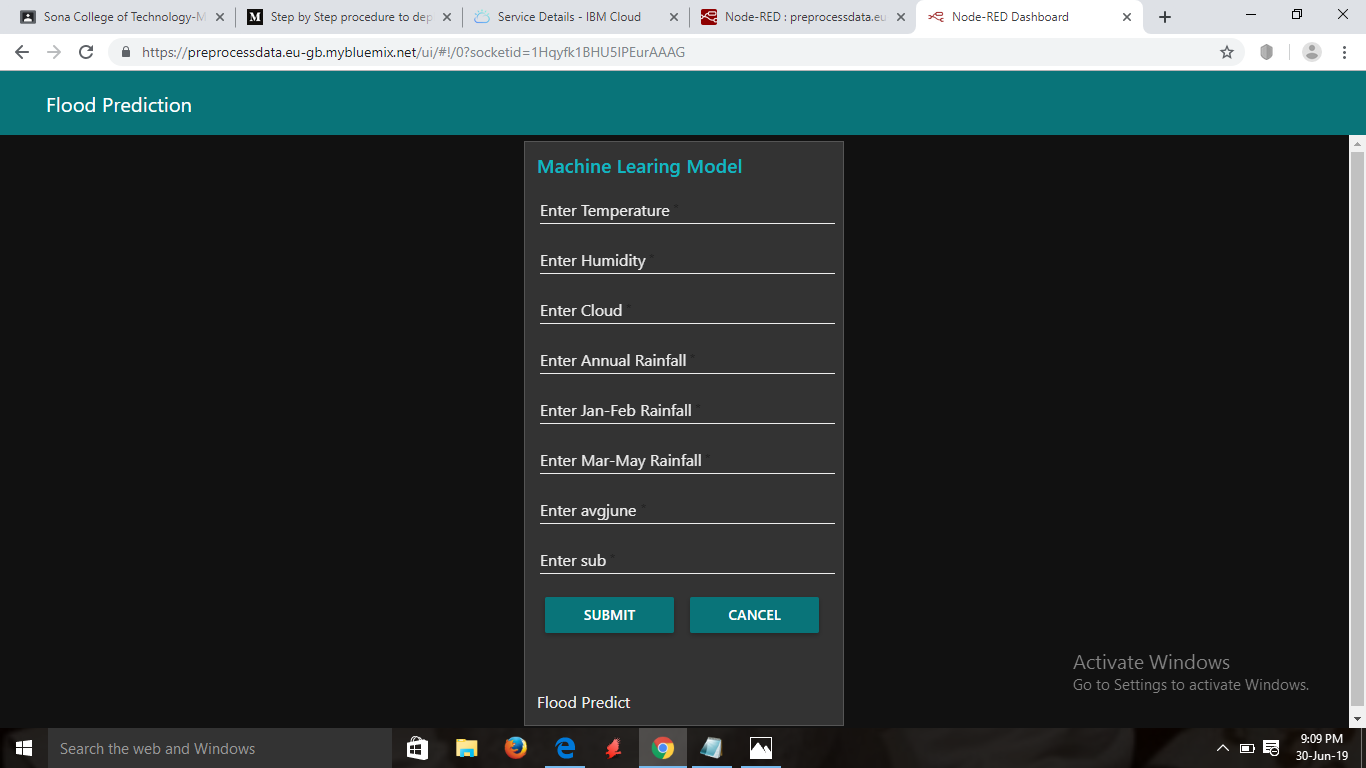
**Fig 4.1.3 roc\_auc\_dt**

****

**Fig 4.1.4 Best fit graph**

****

**Fig 4.1.5 Flow of decision tree**

****

**Fig 4.1.6 User Interface**

**4.2 Data modelling**

Initially we followed all the basic machine learning model steps like data reading and cleansing, separated dependent and independent variables namely humidity, temperature, cloud, annual, Jan-Feb, March-May, Average June, Sub as independent and flood as dependent. We plotted graphs for all the dependent and independent variables as in Fig 4. From the graphs we found that it is classification based problems. Now we found correlation table for all the dependent and independent variables to know the dependency of the dependent variable with the independent variable as in Table 4.1, to more clearly visualize the correlation of the variables we plotted heat map (Fig, which helped to easily remove the uncorrelated variable. When the data are subjected to best fit algorithm the best accuracy was found with Decision tree classifier algorithm with better ROC (Fig 4.1.4). By applying Decision tree classifier algorithm the accuracy score of 86% and ROC was about 85%.We plotted the ROC curve of decision tree classifier (Fig 4.1.3).

To understand the backend flow, on what basis decisions have taken in the algorithm. We visualize the flow of decision tree for flood prediction based on entropy ( Fig 4.1.5).

We then deployed it in IBM Watson studios, and made a data flow UI creation using NODE RED, Fig 4.1.1., then using created web UI Fig:4.1.6 User Interface, enter the inputs and get the predicted outputs.

**5. REFERENCE**

**1.**<https://www.researchgate.net/publication/328562202_Flood_Prediction_Using_Machine_Learning_Models_Literature_Review>

2. <https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050917X00185/1-s2.0-S1877050917323979/main.pdf>

3. <http://machine-learning-project-website.s3-website-us-east-1.amazonaws.com/>

4. <http://imdchennai.gov.in/csframe.htm>

5. <https://github.com/rajatkeshri/Flood-prediction>

**6. CONCLUSION**

Flood have huge impact towards the human existency even in the modern era. Prediction of flood helps the survival to defend themselves in more fine way and economy loss can be reduced. This flood prediction is done based on rainfall and other environmental parameters. On real time you can make a IoT bot to get the details of temperature, humidity, cloud and rainfall.

Hence using this machine learning model for flood prediction, flood can be predicted priorly and easily which helps the public and government to take necessary steps to define them form the vast disaster.