International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Prediction of Flood by Rainfall using MLP Classifier of Neural Network Model

P.Vijayalakshmi M.C.A., M.Phil., M.Tech.,[1], M.Vandhana[2], K.Vasanthapriya[3], P.B.Thushikka[4]

[1]Associate professor, Dept. of Computer Science Engineering, Panimalar Engineering College, Tamilnadu, India
[2,3,4]Student, Dept. of Computer Science Engineering, Panimalar Engineering College, Tamilnadu, India

---

Abstract - Floods are among the most destructive natural disasters, which are highly complex to model. The research on the advancement of flood prediction models contributed to risk reduction, policy suggestion, minimization of the loss of human life, and reduction of property damage which are all associated with floods. To mimic the complex mathematical expressions of physical processes of floods, during the past two decades, neural network methods contributed highly in the advancement of prediction systems providing better performance and cost-effective solutions. To prevent this problem and to predict flood using rainfall dataset, the neural network based techniques are used. The analysis process started from variable identification, missing value treatments, data validation and data cleaning/preparing, exploratory analysis and evaluation. Various modules are built signifying the use of each of the analytical process made. Machine learning algorithms are made used for prediction. The Algorithms like, Logistic Regression, Support Vector Machines, K-Nearest Neighbor, Multi-Layer Perceptron Classifier. Among them, the one which provides the highest accuracy is made to predict the Flood. The performance of Flood prediction takes place by calculating accuracy along with the evaluation classification report, identifying the confusion matrix and the final result is shown as GUI based application by providing corresponding Input Parameters.

Key Words: Pre-processing, Multi Layer Perceptron, Confusion Matrix, Dataset, Neural network

1. INTRODUCTION

In the urban cities, sub-pass or a low-lying area is most vulnerable to water logging. On this flood-prone locations; water gets accumulated in a short period of time. Relative elevation, surface runoff and insufficient passage of water to drainage are key points in the development of water logging. Thus, flood forecasting at these places is essential.

1.1. OVERVIEW

Machine learning is to predict the future from past data. Machine learning (ML) is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a simple machine learning algorithm using python. Process of training and prediction involves use of specialized algorithms. It feed the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data. Machine learning can be roughly separated in to three categories. There are supervised learning, unsupervised learning and reinforcement learning.

Techniques of Supervised Machine Learning algorithms include logistic regression, multi-class classification, Decision Trees and support vector machines etc. Supervised learning problems can be further grouped into Classification problems. Unsupervised learning is a machine learning technique in which there is no supervision required instead the model is allowed to work on its own to discover information. Some applications of unsupervised machine learning techniques include: Clustering, Anomaly detection, Association mining, Latent variable models

1.2. SCOPE

The scope of this project is to investigate rainfall dataset of India to identify flash flood .Identifying flashflood is more difficult. In order to minimize the loss of human life and to reduce the property damage that are all associated with flood. This type of prediction will reduce the meteorological efforts and increase the probability of predicting the occurrence of flash flood.

1.3. LITERATURE SURVEY

The use of synthetic aperture radar (SAR) data is presently well established in operational services for flood management. However, some events might be missed because of the limited area that can be observed through a SAR image and the need of programming SAR acquisitions in advance. To tackle these problems, it is possible to setup a system that is able to trigger the SAR acquisitions based on flood forecasts and to take advantage of the various satellite SAR sensors that are presently operating. On behalf of the Italian Civil Protection Department (DPC), a proto- type of this kind of system has been setup and preliminary tested, using COSMO-Sky Med (CSK) and Sentinel-1 (S-1) data, to monitor the Po River (Northern Italy) flood occurred in November 2014. This paper presents the prototype system and describes in detail the near real-time flood mapping algorithm implemented in the system. The algorithm was previously developed to classify CSK images, and is modified here in order to be applied to S-1 data too. The major

outcomes of the monitoring of the Po River flood are also analyzed in this paper, highlighting the importance of the in advance programming of the radar acquisitions. Results demonstrate the reliability of the flood predictions provided by the model and the accuracy of the flood mapping algorithm. It is also shown that, when CSK and S-1 data are simultaneously acquired, their joint use allows for an interpretation of some ambiguous radar signatures in agricultural areas.

2. EXISTING SYSTEM

The conventional strategies of flood forecasting are expensive and highly complex. Weather and rainfall forecasting is a major task behind the prediction of a flood. Weather forecasting involves simulations based on physics and differential equations. The rainfall forecast is done using radars and satellite imaging. A Doppler weather radar is used to locate the precipitation and detect the motion of rain droplets. Dedicated weather satellite provides images using which information about rainfall can be deduced and it's for short-term flash flood prediction in urban areas is to establish a theoretical model incorporating the factors influencing flood and use the power of machine learning techniques to estimate flood ahead of time.

Rapid urbanization, climate change, and extreme rainfall have resulted in a growing number of cases of urban flash floods. It is important to predict the occurrence of a flood so that the aftermath of it can be minimized. As the name suggests, an urban flash flood occurs in an urban area in a very short span of time. To reduce the impact of these events, short-term forecasting or now casting is used for prediction of the very near future incident. In orthodox methods of flood forecasting, current weather conditions are examined using conventional methods such as the use of radar, satellite imaging and calculations involving complicated mathematical equations.

However, recent developments in Information and Communication Technology (ICT) and Machine Learning (ML) has helped us to study this hydrological problem from a different perspective. The aim is to design a theoretical model considering the parameters causing the urban flash flood and predict the event beforehand.

DRAWBACKS:

To implement a model for an urban area in which short-term forecasting of a flood and it is complex scenario including multiple cells is imagined where a possible real-world scenario is undertaken.

It can't thereby better determine the regularity of rainfall data and achieve more accurate prediction results of flash flood.

3. PROPOSED SYSTEM

Flood is one of the biggest natural disaster causing many lives as well as damages. Different types of floods like river flood, urban flood, coastal flood, and flash flood have been observed over the years. A flash flood is a direct response to a rainfall having very high intensity in small time. This kind of flood is seen typically in urban areas where the underlying ground cannot cope, or drain excess water away fast enough via the sewage system and draining canals in a short amount of time. In recent years, we have seen the impact of floods in cities such as mumbai, chennai, etc. Poor urban planning, inaccurate and delayed forecasting and inadequate flood mitigation system are the main reasons behind it. From the given dataset generalized dataset is formed and that is applied to extract patterns, results with maximum accuracy.

In this section three steps are done: report will load in data, checking the data, trimming the data clean given dataset for analysis. Make sure that the document steps carefully and justify for cleaning decisions. The data set collected for predicting given data is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using *Multi-Layer Perceptron Classifier* will be apply on the Training set and based on the test result accuracy, Test set prediction is done.

ANN models are often trained with a BPNN for flood prediction. The MLP is a class of FFNN which utilizes the supervised learning of BP for training the network of interconnected nodes of multiple layers. .The characteristic of MLP are nonlinear activation, and a high number of layers .These model was widely used in flood prediction and other complex hydrogeological models due to these characteristics,. MLP models were reported to be more efficient and found to be more difficult to optimize.

ADVANTAGES:

These reports are to the investigation of applicability of machine learning techniques for air quality forecasting in operational conditions. Finally, it highlights some observations on future research issues, challenges, and needs.

ML is a field of artificial intelligence (AI) used to induce regularities and patterns, providing easier implementation with low computation cost, as well as fast training, validation, testing, and evaluation, with high performance compared to physical models, and relatively less complexity.

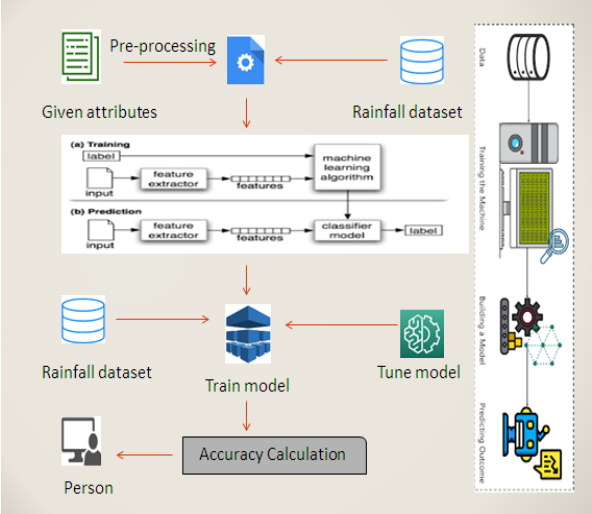


Fig.1

4. REQUIREMENT ANALYSIS AND SPECIFICATION

4.1. INPUT REQUIREMENTS:

The amount of Rainfall measured for the places where the occurrence of flood has to be predicted.

4.2OUTPUT REQUIREMENTS:

A Graphic User Interface(GUI) containing the column for the respective state and districts of India along with the amount threshold value .With all these data, the prediction can be successfully made.

4.3. FUNCTIONAL REQUIREMENTS:

The software requirements specification is a technical specification of requirements for the software product. It is the first step in the requirements analysis process. It lists requirements of a particular software system. The following details to follow the special libraries like sk-learn, pandas, numpy, matplotlib and seaborn.

4.4. HARDWARE ENVIRONMENT

Processor : i3/ i4
Hard disk : minimum 300 GB
RAM : minimum 4 GB

4.5. SOFTWARE ENVIRONMENT

Requirement

Operating System : Windows
Tool : Anaconda with Jupyter Notebook

5. MODULE DESIGN SPECIFICATION

5.1. DATA VALIDATION AND PRE-PROCESSING TECHNIQUE:

Data validation helps to get the error rate of Machine Learning(ML) model by finding the Missing values and duplicate values .It evaluates the given data set and helps to understand the data and it's properties. This knowledge would help us to identify the algorithm required to build the model.

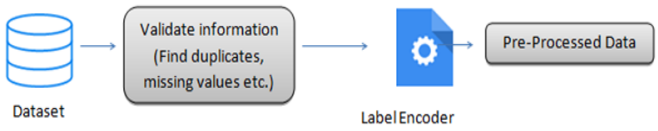


Fig.2

5.2. CREATE A PREDICTED VARIABLE BY RAINFALL RANGE:

The dataset is trained to give an estimated model skill and procedures that we can use to make the best use of validation and test datasets when evaluating our models .To understand our data in a better way, we need to look it in a visual form, like charting the time series data with line plots and categorical quantities with bar charts.

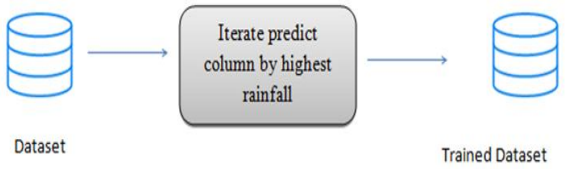


Fig.3

5.3. PERFORMANCE MEASUREMENTS OF ML ALGORITHMS:

To measure the performance, it is necessary that we must assure that our model got the correct patterns from the data, and it is not getting up too much noise. Outliers can result in misleading representations and can skew the summary distribution of attribute values in descriptive statistics. But in Cross validation, our model is trained using subset of the data-set and then evaluated using the complementary subset of the data-set.

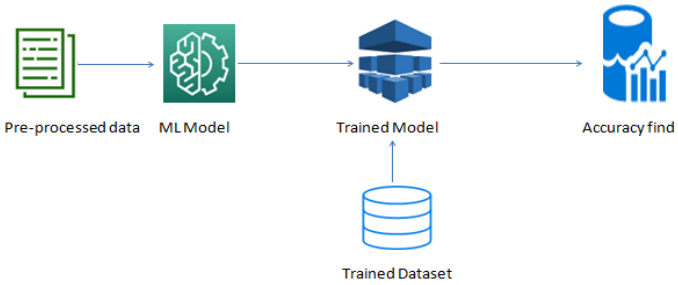


Fig.4

5.4.PERFORMANCE MEASUREMENTS OF MLP CLASSIFIER NEURAL NETWORK:

A Multi-Layer Perceptron or Multi-Layer Neural Network contains one or more hidden layers .An MLP

consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Each node is a neuron that uses a nonlinear activation function except for the input nodes .Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

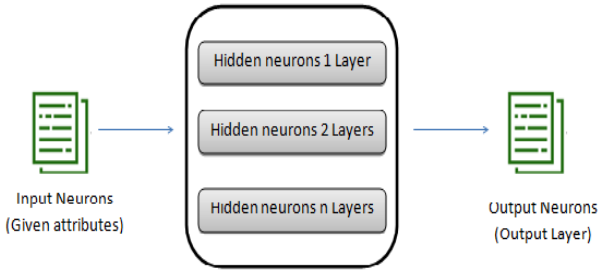


Fig.5

5.5. GUI BASED PREDICTION OF FLOOD BY RAINFALL:

The tkinter package ("Tk interface") is the standard Python interface to the Tk GUI toolkit. Python when combined with Tkinter provides a fast and easy way to create GUI applications. Tkinter can be used for security purpose of each users or accountants.

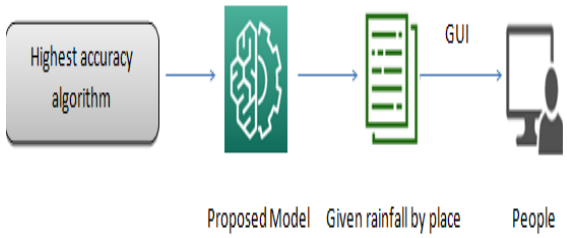


Fig.6

6. PROGRAM DESIGN LANGUAGE:

6.1. LOGISTIC REGRESSION

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes .Logistic Regression is used when the dependent variable (target) is categorical. For example, To predict whether an email is spam (1) or (0).

6.2. SUPPORT VECTOR MACHINES

Support vector machines are maximum-margin classifiers, which means they find the hyperplane that has the largest perpendicular distance between the hyperplane and the closest samples on either side. How to disentangle the many names used to refer to support vector machines. When a training set is given, an SVM training algorithm builds a model that assigns new examples to the categories, making it a non-probabilistic binary linear classifier.

6.3. K-*NEAREST NEIGHBOR*

The k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression. It is a classification algorithm, and it is supervised: it takes a bunch of labeled points and uses them to learn how to label other points. In k-NN classification, the output is a class membership. A peculiarity of the k-NN algorithm is that it is sensitive to the local structure of the data.

6.4. MULTILAYER PERCEPTRON:

A multilayer perceptron (MLP) is a class of feed forward artificial neural network (ANN). The term MLP is used ambiguously, sometimes loosely to refer to any feed forward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptron's (with threshold activation). Multilayer perceptron's are sometimes colloquially referred to neural networks, especially when they have a single hidden layer.

A Multi-Layer Perceptron or Multi-Layer Neural Network contains one or more hidden layers (apart from one input and one output layer). While a single layer perceptron can only learn linear functions, a multi-layer perceptron can also learn non – linear functions. An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

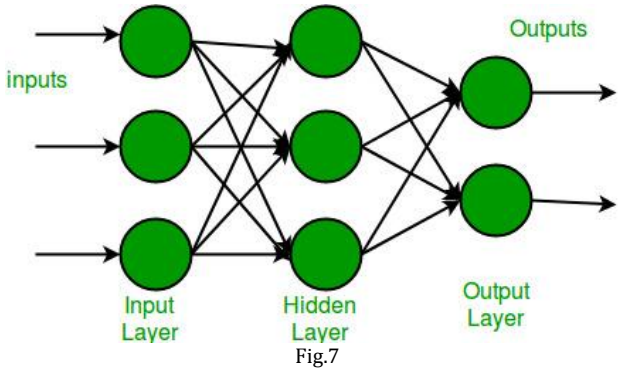


Fig.7

A perceptron is a very simple learning machine. It can take in a few inputs, each of which has a weight to signify how important it is, and generate an output decision of "0" or "1". However, when combined with many other perceptrons, it forms an artificial neural network. A neural network can, theoretically, answer any question, given enough training data and computing power.

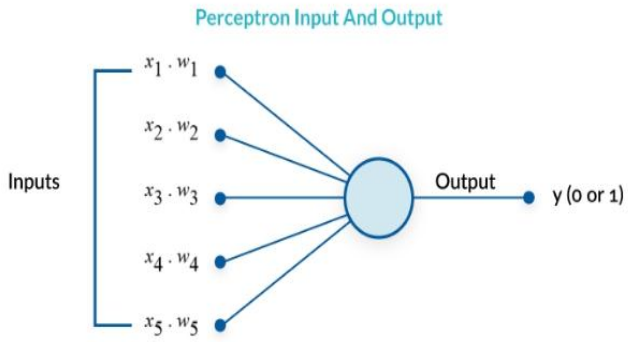


Fig.8

A multilayer perceptron (MLP) is a perceptron that teams up with additional perceptrons, stacked in several layers, to solve complex problems. The diagram below shows an MLP with three layers. Each perceptron in the first layer on the left (the input layer), sends outputs to all the perceptrons in the second layer (the hidden layer), and all perceptrons in the second layer send outputs to the final layer on the right (the output layer). Each perceptron sends multiple signals, one signal going to each perceptron in the next layer

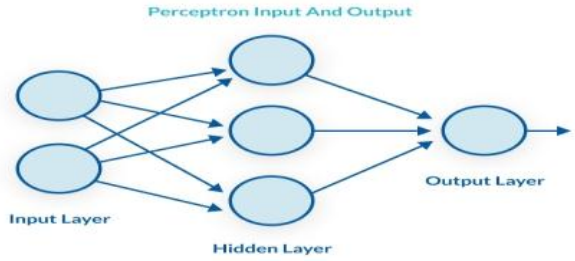


Fig.9

For each signal, the perceptron uses different weights. In the diagram above, every line going from a perceptron in one layer to the next layer represents a different output. Each layer can have a large number of perceptrons, and there can be multiple layers, so the multilayer perceptron can quickly become a very complex system.

The multilayer perceptron has another, more common name a neural network. A three-layer MLP, like the diagram above, is called a Non-Deep or Shallow Neural Network. An MLP with four or more layers is called a Deep Neural Network. One difference between an MLP and a neural network is that in the classic perceptron, the decision function is a step function and the output is binary.

In neural networks that evolved from MLPs, other activation function can be used which result in outputs of real values, usually between 0 and 1 or between -1 and 1. This allows for probability-based predictions or classification of items into multiple labels.

7. WORKING:

1. Takes the inputs, multiplies them by their weights, and computes their sum: The weights allow the perceptron to evaluate the relative importance of each of the outputs. Neural network algorithms learn by discovering better and better weights that result in a more accurate prediction. There are several algorithms used to fine tune the weights, the most common is called back propagation.

2. Adds a bias factor, the number 1 multiplied by a weight: This is a technical step that makes it possible to move the activation function curve up and down, or left and right on the number graph. It makes it possible to fine-tune the numeric output of the perceptron.

3. Feeds the sum through the activation function: The activation function maps the input values to the required output values. For example, input values could be between 1 and 100, and outputs can be 0 or 1. The activation function also helps the perceptron to learn, when it is part of a multilayer perceptron (MLP). Certain properties of the activation function, especially its non-linear nature, make it possible to train complex neural networks.

4. The result is the perceptron output: The perceptron output is a classification decision. In a multilayer perceptron, the output of one layer's perceptrons is the input of the next layer. The output of the final perceptrons, in the "output layer", is the final prediction of the perceptron learning model.

8. PERFORMANCE ANALYSIS:

PARAMETER AND ACCURACY CALCULATIONS:

I. True Positive: It is an outcome where the model *correctly* predicts the *positive* class.
II. True Negative: It is an outcome where the model *correctly* predicts the *negative* class.
III. False Positive: It is an outcome where the model incorrectly predicts the positive class.
IV. False Negative: It is an outcome where the model incorrectly predicts the negative class.

SENSITIVITY

Sensitivity is a measure of the proportion of actual positive cases that got predicted as positive (or true positive). Sensitivity is also termed as Recall. This implies that there will be another proportion of actual positive cases, which would get predicted incorrectly as negative (and, thus, could also be termed as the false negative). This can also be represented in the form of a false negative rate. The sum of sensitivity and false negative rate would be 1.

Mathematically, sensitivity can be calculated as the following:
Sensitivity = (True Positive) / (True Positive + False Negative)
The higher value of sensitivity would mean higher value of true positive and lower value of false negative. The lower value of sensitivity would mean lower value of true positive and higher value of false negative.
Specificity
Specificity is defined as the proportion of actual negatives, which got predicted as the negative (or true negative). This implies that there will be another proportion of actual negative, which got predicted as positive and could be termed as false positives. This proportion could also be called a false positive rate. The sum of specificity and false positive rate would always be 1.

Mathematically, specificity can be calculated as the following:
Specificity = (True Negative) / (True Negative + False Positive)
The higher value of specificity would mean higher value of true negative and lower false positive rate. The lower value of specificity would mean lower value of true negative and higher value of false positive.

Precision: The proportion of positive predictions that are actually correct. (When the model predicts default: how often is correct?)
Precision = TP / (TP + FP)
Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labeled as survived, how many actually survived? High precision relates to the low false positive rate. We have got 0.788 precision which is pretty good.

Recall: The proportion of positive observed values correctly predicted. (The proportion of actual defaulters that the model will correctly predict)
Recall = TP / (TP + FN)
Recall (Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.
F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

General Formula
F- Measure = 2TP / (2TP + FP + FN)
F1-Score Formula

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

Comparison of accuracy results:

Parameters	LR	SVC	KNN	MLP
PRECISION	0.57	0.75	0.84	0.93
Recall	1.00	0.87	0.87	0.93
F1-Score	0.73	0.80	0.85	0.93
Sensitivity	0.88	1.00	0.96	0.96
Specificity	1.00	0.00	0.25	0.75
Accuracy (%)	90.00	86.66	86.66	93.33

Table.1

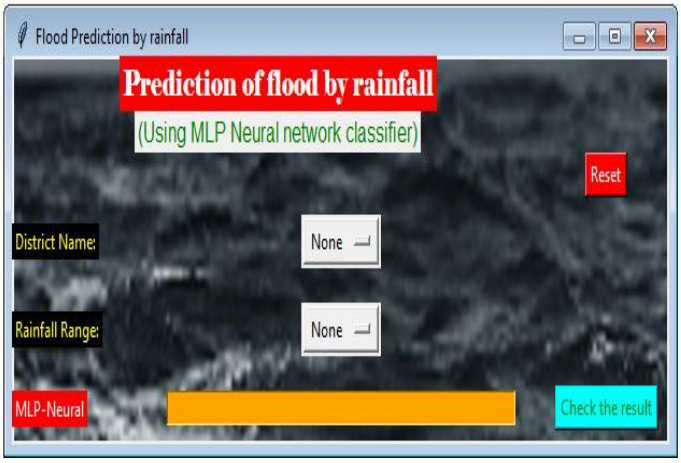
Comparison of confusion matrix parameters:

Algorithms	TP	TN	FP	FN	TPR	TNR	FPR	FNR	PPV	NPV
LR	4	23	3	0	1.00	0.88	0.11	0.00	0.57	1.00
SVC	0	26	0	4	0.00	1.00	0.00	1.00	Nan	0.86
KNN	1	25	1	3	0.25	0.96	0.03	0.75	0.50	0.89
MLPS	3	25	1	1	0.75	0.96	0.03	0.25	0.75	0.96

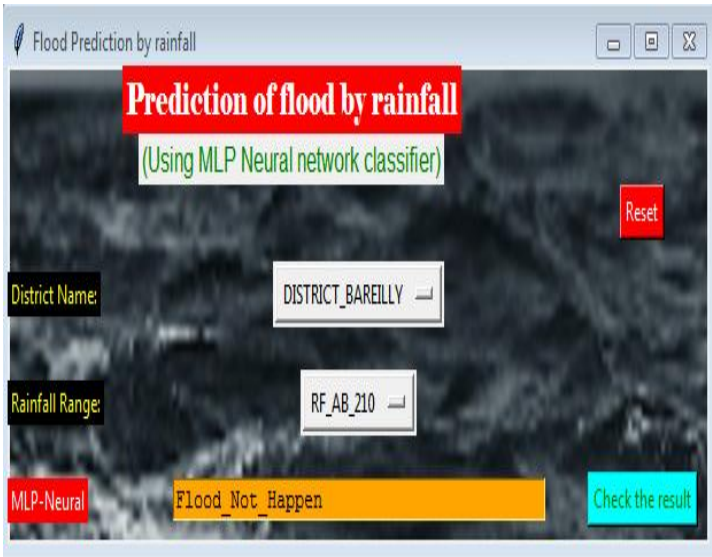
Table.2

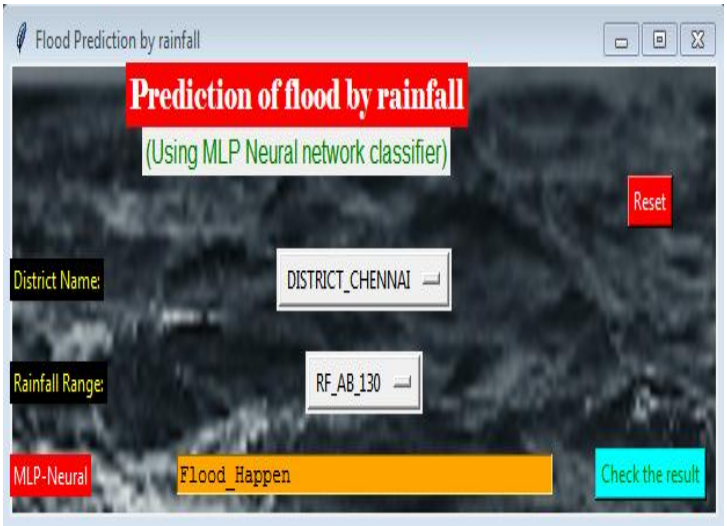
9. REPORT:

9.1.INPUT:



9.2.OUTPUT:





10.CONCLUSION:

There were limited attempts in the prediction of Flash Flood. Radar was the only option used for the prediction of Flood. So, the main idea of this project is to provide prediction using neural networks. For which the analytical process was used which started from data cleaning and processing, missing value, exploratory analysis, model building and evaluation. Various modules were built signifying the use of each of the analytical process made. Finally, we predict flash flood using different machine learning algorithm each proving distinct results. Logistic Regression, Support Vector Machines, K-Nearest Neighbor, Multi-Layer Perceptron Classifier are used .Among them, the best result was provided by Multi-Layer Perceptron Classifier algorithm with an accuracy of 93.33%. This brings some of the following insights about flood prediction .This signifies with an intensive model for the prediction of flash flood using Machine Learning.

11. FUTURE WORK

- Disaster management wants to automate the detecting the flash flood happened or not from eligibility process (real time).
- To automate this process by show the prediction result in web application or desktop application.
- To optimize the work to implement in Artificial Intelligence environment
- To enable automatic notification feature provided via Android Application.
- To include various other parameters to enhance the accuracy.

12. REFERENCES:

1. Giorgio Boni, Luca Ferraris, Luca Pulvirenti," A Prototype System for Flood Monitoring Based on Flood Forecast Combined With COSMO-SkyMed and Sentinel-1 Data "," IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 9, NO. 6, JUNE 2016".
2. Suraj Ogale, Sanjay Srivastava," Modelling and short term forecasting of flash floodsin an urban environment".
3. Wahyu Sardjono , Widhilaga Gia Perdana , "The Application of Artificial Neural Network for Flood Systems Mitigation at Jakarta City".
4. Swapnil Bande, Prof. Dr. Virendra V. Shete," Smart flood disaster prediction system using IoT & Neural Networks".
5. Pengzhan Cui, Yeqing Guan, Ying Zhu," Flood Loss Prediction of Coastal City Based on GM-ANN"
6. V Phyo Pa Pa Tun, Myint Myint Sein,," Flood Prediction System for Middle Region of Myanmar".
7. Ni Komang Ega Kartika, Muhammad Ary Murti, Casi Setianingsih," Floods Prediction Using Radial Basis Function(RBF) Based on Internet of Things (IoT)"," The 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)"
8. Ryo Natsuaki, Akira Hirose," L-BAND SAR INTERFEROMETRIC ANALYSIS FOR FLOOD DETECTION IN URBAN AREA – A CASE STUDY IN 2015 JOSO FLOOD, JAPAN"
9. Amitkumar B. Ranit, Dr. P.V.Durge," Different Techniques of Flood Forecasting and theirApplications".
10. Indrastanti R. Widiasari, Lukito Edi Nugoho2, Widyawan, Rissal Efendi," Context-based Hydrology Time Series Data for A Flood Prediction Model Using LSTM",
11. Fazlina Ahmat Ruslan, Khadijah Haron, Abd Manan Samad, Ramli Adnan," 4 Hours NNARX Flood Prediction Model Using "traingd" and "trainoss" Training Function: A Comparative Study","2018 IEEE 14th International Colloquium on Signal

Processing & its Applications (CSPA 2018), 9 -10 March 2018, Penang, Malaysia".

12. Akshya .J, P.L.K.Priyadarsini," A Hybrid Machine Learning Approach forClassifying Aerial Images of Flood-Hit Areas"," Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019)".
13. Febus Reidj G. Cruz , Matthew G. Binag, Marlou Ryan G. Ga , Francis Aldrine A. Uy „" Flood Prediction Using Multi-Layer Artificial Neural Network in Monitoring System with Rain Gauge, Water Level, Soil Moisture Sensors"," Proceedings of TENCON 2018 - 2018 IEEE Region 10 Conference (Jeju, Korea, 28-31 October 2018)"
14. Gurleen Kaur, Anju Bala," An Efficient Automated Hybrid Algorithm toPredict Floods in Cloud Environment"," 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)".
15. Fazlina Ahmat Ruslan, Khadijah Haron, Abd Manan Samad, Ramli Adnan," Multiple Input Single Output (MISO) ARX and ARMAX Model of Flood Prediction System:Case Study Pahang"," 2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA 2017), 10 - 12 March 2017, Penang, Malaysia 978-1-".
16. Kiruthika Devi B.S1, Aravindhan K2, P Srikanth Kini3 ,G Adarsh Reddy4,Subbulakshmi T5," A PREDICTION MODEL FOR FLOODED PACKETS IN SNMP NETWORKS"," Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3".