

Flood Prediction using Artificial Neural Network

Chaitanya Thekkunja¹, Dr. ShivaKumar G.S², Shubhang S Aroor³

Computer Science , Srinivas Institute of Technology, Mangalore, Karnataka, India

ABSTRACT

Floods are one of the foremost catastrophic natural disasters, and, thanks to their complex nature, it's tough to make a predictive model. The advanced research works on flood prediction models have contributed to risk reduction, policy suggestion, minimization of the loss of human life, and reduced property damage related to floods. In general, ML algorithms are utilized in the event of prediction systems, to mimic the complex mathematical expressions of the physical processes of floods providing better performance and cost-effective solutions. The MLP model is implemented in this system by calculating accuracy values with examining the confusion matrix parameters. The proposed system analyses the dataset using Multilayer Perceptron Classifier (MLP) algorithm to coach the predictive model, and floods are often predicted.

I. INTRODUCTION

Floods are amongst the foremost catastrophic natural disasters, causing massive damage to human life, infrastructure, agriculture, and therefore the socioeconomic system. Governments are under enormous pressure to develop reliable and accurate maps of flood risk areas and further plans for sustainable flood risk management, that consider prevention, protection, and preparedness. Many control measures are being implemented to reduce the losses incurred thanks to it. But these control measures haven't been that effective in areas that are vulnerable to flash floods. because of its capability to predict future events, machine learning has contributed tons in preventing many natural disasters like floods, earthquakes. Being a man-made Intelligence application, Machine Learning is where the systems can learn from past data, identify patterns, and learn to form predictions with minimal human interaction. It provides computers with the facility to seek out and make decisions without being programmed explicitly. It mainly focuses on developing computer programs which can change

when exposed to new data. the tactic of coaching and testing of a machine learning model involves the use of various specialized machine learning algorithms. The methodology follows the tactic of feeding the training data to an algorithm, and therefore the algorithm uses this training data to provide predictions on new test data. Machine learning are often roughly classified into three categories. they're Supervised Learning, Unsupervised Learning, and Reinforcement Learning. In Supervised learning, the prediction model is given a training dataset with both the input file and thus the corresponding output data. In Unsupervised Learning, the training set has no labels, and thus the algorithm tries to hunt out a structure or a pattern within the given un-labelled data set. The algorithms used for unsupervised learning are of two types called clustering algorithm and thus the expectation-maximization algorithm. within the case of reinforcement learning, the algorithm dynamically interacts with its environment, and it receives positive or feedback to reinforce its performance. Data scientists use many various sorts of machine learning algorithms to get patterns that cause actionable insights. At a high level, these

different algorithms are classified as supervised and unsupervised learning algorithms supported their way of coaching and therefore the sort of dataset they use to form predictions. Supervised learning are often further grouped into Classification and Regression problems. Both problems have a goal: constructing a succinct prediction model that predicts the worth of the dependent attribute from the input attribute variables. The difference between the 2 problems is that the dependent attribute is numerical for regression and categorical for classification. Classification may be a supervised learning approach that involves the method of predicting the label or classes of given data points. The task of predictive classification modelling is an approximation of a mapping function from input variables(X) to discrete output variables(Y). This data set could also be bi-class or multi-class containing discrete output values. Some well-known samples of classification problems might be document classification and speech recognition. during a regression problem, the results are predicted within endless output where the input variables are mapped to some continuous function. Regression problems are categorized into various types, like Simple rectilinear regression , Polynomial Regression, Random Forest Regression, Decision Tree Regression, Support Vector Regression

Unsupervised learning mainly deals with un-labelled data, and using the clustering algorithm; it finds structures or patterns within the given dataset. It allows you to perform more complex processing tasks compared to supervised learning. Unsupervised learning are often classified into Parametric unsupervised learning and Non-parametric unsupervised learning. The parametric unsupervised learning algorithm makes use of the Expectation-maximization algorithm to construct Gaussian Mixture Models to predict the category of the given un-labelled data. In Non-parametric unsupervised learning, the info is grouped into clusters by finding a pattern or a structure during a collection of un-labelled data. The clustering problem is taken into account because the most desirable unsupervised learning problem. Clustering algorithms are often classified as Exclusive Clustering, Overlapping Clustering, Hierarchical Clustering, and Probabilistic Clustering. The goal of this technique is to develop a machine learning

model for real-time flood forecasting. In recent times, sub-pass or low-lying areas in cities are most susceptible to waterlogging. In these flood-prone locations, water gets accumulated during a short period. Relative elevation, surface runoff, and insufficient passage of water to drainage are critical points within the development of waterlogging. Thus, flash flood forecasting at these places is important. This project's scope is to coach a prediction model with the dataset of rainfall records of all the states in India to predict flood warning prior. this is often to scale back the inefficiency in predicting the flood and to enhance the accuracy analysis in predicting floods prior, which might , in turn, save lives and also many to avoid wasting "many |to avoid wasting" to save many lots of lots of meteorological efforts and assets in predicting flood occurrence. To avoid wasting" to save lots of lots of meteorological efforts and assets in predicting flood occurrence.

II. EXISTING SYSTEM

The strategies of flood forecasting are expensive and highly complicated. Weather and rainfall forecasting may be a significant task behind the prediction of a flood. meteorology involves simulations supported by physics and differential equations. The rainfall forecast is completed using radars and satellite imaging. A Doppler radar is used to locate the precipitation and detect the motion of rain droplets. Dedicated weather satellites provide images using which information about rainfall is usually deduced. In short-term flash flood prediction in urban areas, a theoretical model incorporating the factors influencing flood is established using the facility of machine learning techniques, which is employed to estimate floods before time. Rapid urbanization, global global climate change , and extreme rainfall have resulted during a growing number of cases of urban flash floods in recent times. it's crucial to predict the occurrence of a flood in order that its aftermath are often minimized. As the name suggests, an urban flash flood occurs during an urban area during a short period. Hence, to scale back its impact, short-term forecasting is employed for the prediction of the very near future incident. In orthodox methods of flood forecasting, current weather are examining

using the conventional methods like radar and satellite imaging, and calculations involving complicated mathematical equations. The recent developments in Machine Learning (ML) and knowledge and Communication Technology (ICT) have helped us to review this hydrological problem from a special perspective. The aim is to style a theoretical model considering the parameters causing the urban flash flood and predict the event. The drawbacks of this technique include the subsequent , To implement a model for an populated area during which short-term forecasting of a flood may be a complex scenario, including multiple factors, that are imagined where a possible real-world scenario is undertaken. It can't be determine the regularity of rainfall data and achieve more accurate prediction results of a flash flood.

III.PROPOSED SYSTEM

This system aims to gather data from all the states of India and form a generalized dataset. A machine learning algorithm is applied to the labelled dataset, and patterns are extracted, which, in turn, obtain maximum accuracy with real-time input. In general, the dataset collected for predicting is split into a Training set and Test set. The Multilayer Perceptron Classifier algorithm is applied to train the predictive model and appropriate results are obtained with accuracy and precisely examining the confusion matrix parameters.

IV.OBJECTIVES

The objective of Flood Prediction using Machine Learning is to design a incremental model to predict floods based on the rainfall levels, channel flows, river levels etc. An approach of prediction is using Machine Learning that has very good working efficiency produces the accurate results. The system helps to improve the performance. This method can be used to predict both long term and short term floods in a given region.and maintaining the project is easy and manageable

V . SYSTEM IMPLEMENTATION

The evaluation of the prediction accuracy of the classifiers is administered using data validation, and therefore the results are compared to get the actual accuracy. It is used to seek out the accuracy of the training dataset, accuracy of the testing dataset, specification, false-positive rate, precision, and recall by comparing the algorithm models using python code. the subsequent Involvement steps are,

- Define a problem
- Preparing data
- Evaluating algorithms
- Predicting results

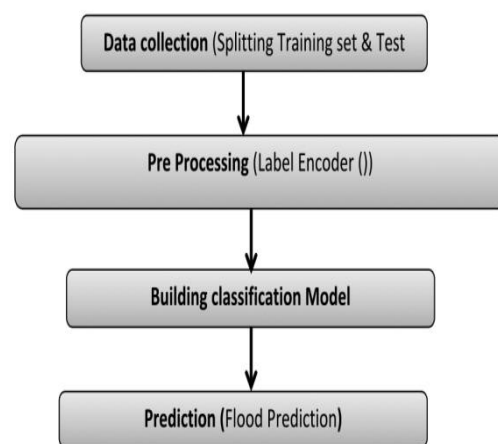


Fig 1: System architecture

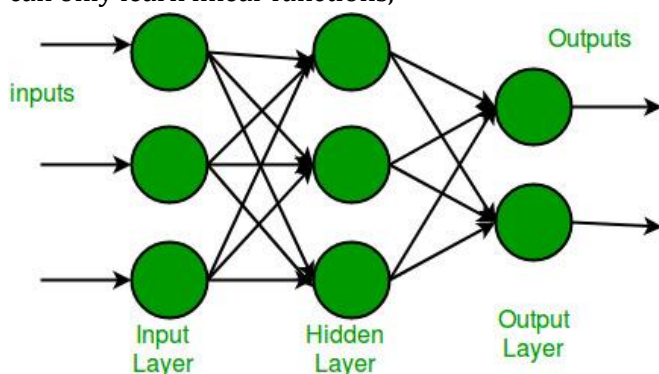
Data Validation Process is the Specific validation technique are used to determine the error rate of machine learning. To determine missing value, duplicate value, and descriptive data type, finding whether it is float or integer type. This process involves variable identification by data shape, data type, and evaluating the missing values. The goal of cleaning data is to detect and remove errors. A sample of data, known as the validation dataset, is held back from training the model, which can be used to make the best use of validation and test datasets when evaluation happens in your models, giving an estimate of model skill while tuning models and procedures. And one more process that is the Data Pre - Processing process of transforming data prior to feeding it to the algorithm is called data pre-processing. For better results, such a process takes place where the raw data is converted to clean data and fed to the algorithm. The data has to be in the proper manner to achieve good results. Some of the machine learning models need data in

the specified format, the random forest algorithm does not support null values. Therefore to execute a random forest algorithm the raw data has to be transformed to a null-free data set. And in the same data visualization that is considered a very important skill in applied statistics which focuses on quantitative descriptions and estimations of data and machine learning. Data visualization provides an important suite of tools for gaining a quantitative understanding which can be helpful when exploring and getting to know a dataset that can be helpful with identifying patterns, corrupt data, outliers, and more.

VI. SYSTEM IMPLEMENTATION

MULTI LAYER PERCEPTRON (MLP)

A Multi-Layer Perceptron (MLP) contains one or more hidden layers. While a single layer perceptron can only learn linear functions,



a multi-layer perceptron can also learn non-linear functions this neuron takes as input x_1, x_2, \dots, x_3 and outputs $f(\text{summed inputs} + \text{bias})$, where $f(\cdot)$ called the activation function. The main function of bias is to provide every node with a trainable constant value. Every activation function takes one number and performs a particular fixed mathematical process thereon. There are several activation functions listed below:

Sigmoid: takes the real-valued input and squashes it to range between 0 and 1.

Tanh: takes the real-valued input and squashes it to range $[-1, 1]$.

ReLU: ReLu stands for Rectified Linear Units. And it takes real-valued input and threshold it to 0 (replaces negative to 0).

PERFORMANCE ANALYSIS METRICS :

- **True Positive:** It is an outcome where the models correctly predicts the positive class. And the outcome is considered as one of the true positive when the system can correctly predict that an incident has indeed occurred.
- **True Negative:** It is an outcome where the models correctly predicts the negative class. And the outcome is considered as true negative when the system can correctly predict the particular incident has not occurred.
- **False Positive:** False Positive is an accuracy measure where models can mis-predicts the positive class. The outcome is considered as False Positive when the system cannot correctly predict the particular incident has occurred.
- **False Negative:** False Negative is an accuracy value where the models mis-predicts the negative class. And the outcome is considered as False Negative when the system cannot correctly predict the particular incident has not occurred.
- **Sensitivity:** Sensitivity is a measure of the proportion of true positive values, the actual number of positive cases that are correctly predicted as positive. It is also known as Recall value. There exists another proportion of actual positive cases that are mis - predicted, that can be represented in the form of a false negative rate. Therefore, the sum of sensitivity and false-negative rate value is represented as 1.

Mathematically sensitivity can be calculated as:

$$\text{Sensitivity/Recall} = (\text{True Positive}) / (\text{True Positive} + \text{False Negative})$$

The higher value of sensitivity would mean a higher value of the true positive and the lower value of false negative. And the lower value of sensitivity would mean a lower value of true positive and higher value of false negative.

Precision: The proportion of positive prediction that are actually correct.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

Precisions are calculated by dividing the number of correctly predicted positive observations by the total number of predicted positive observations. The High precision relates to the low false-positive rate. Recall: The proportion of positive observed values correctly predicted. (The proportion of actual defaulters that models correctly predicts) $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$, Recall (Sensitivity) - Recall is calculated by dividing the number of correctly predicted observation to the total numbers of observations in an actual class. F1 Score: F1 Scores are defined as the weighted average of Precision and Recall values. Hence the both false negative and false positives values are taken into account. When there's an uneven class distribution, the F1 Score value is generally more useful when compared to Accuracy value. And On the other hand, The accuracy values worksheets best when the values of false positive and false negatives have a similar cost. If both the values are different, then Precision and Recall values are taken into account.

General Formula:

$$\text{F-Measure} = 2\text{TP} / (2\text{TP} + \text{FP} + \text{FN})$$

F1-Score Formula:

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

VII. RESULTS AND DISCUSSION

	Precisi on	Reca ll	F1- Scor e	Sensitivi ty	Specifici ty	Accuracy(%)
ML P	0.87	0.82	0.84	0.82	0.98	0.98

Table 1: Accuracy Results

Table 1 shows the results of Multi layer perceptron with the precision, Recall, F1-score, Sensitivity, Specificity, and Accuracy.

Algorit hm	T P	T N	F P	F N	TP R	TN R	FP R	FN R	PP V	NP V
MLP	5 8	75 5	8	1 2	0. 82	0. 98	0. 01	0. 17	0. 87	0. 98

Table 2: Confusion Matrix Parameters of MLP

The above table shows the confusion matrix parameters of Multi Layer Perceptron.

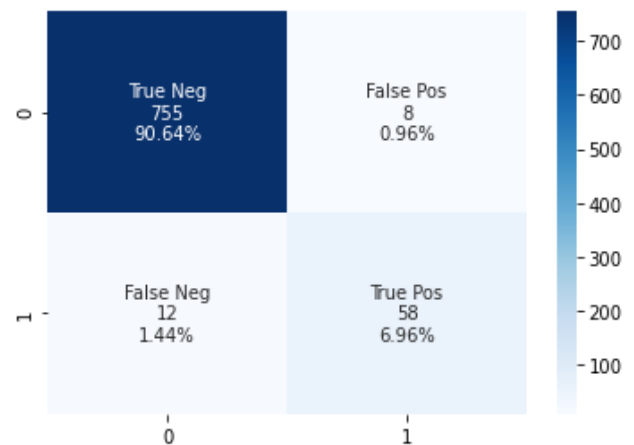


Fig 1: Confusion matrix visualization

VIII. CONCLUSIONS

The analytical process started with data cleaning and processing, And finding missing value, exploratory analysis, and finally, model building and evaluation. Finally, we predicted the flash flood using machine learning model Multilayer Perceptron, which gave accurate results. From the above observations and the analysis, it is been observed that MLP model gives the accuracy of 98%.

REFERENCES

- [1] An Innovative Flood Prediction System Using Improved Machine Learning Approach (csfjournal.com)
- [2] <https://encyclopedia.pub/73>
- [3] PyQt5 Reference Guide — PyQt v5.15 Reference Guide (riverbankcomputing.com)
- [4] Requests: HTTP for Humans™ — Requests 2.25.1 documentation (python-requests.org)
- [5] Tutorials — Matplotlib 3.4.2 documentation
- [6] <https://www.irjet.net/archives/V7/i5/IRJET-V7I51189.pdf>
- [7] Phyo Pa Pa Tun, Myint Myint Sein, "Flood Prediction System for Middle Region of Myanmar", 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE 2018).