International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Flood forecasting by using Machine Learning

Dr. C. Murugamani*[1], K Lakshmi Prasanna[2], R. Mamatha[3]
[1]Associate Professor, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India
[2,3]Student, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

Article Info

Publication Issue :
Volume 8, Issue 5
September-October-2022

Page Number : 329-333

Article History
Accepted: 01 Oct 2022
Published: 18 Oct 2022

ABSTRACT

Floods have always been one of the worst disasters in the world. It not only affects living creatures but also impact the surrounding. The prediction of floods well before its arrival can provide a larger safety measures and can help to protect the habitat. The collection of large data for prediction and analyzing it approximately is always being the point of concern for the increasing technology. Various Artificial Intelligence model and machine learning algorithm software's are developed to predict the flood. The previous developed system still needs modification. Our proposed work is based on Apache System ML machine learning software. Since this platform supports Python programming, the optimization and collection of large number of flood related data can be analyzed. The algorithm codes or programming codes are writable, reduced in error and readable. The proposed system will be more efficient and scalable and is expected to give better results and predictions.

Keywords: Machine Learning, Artificial Intelligence, Apache SystemML, Python

I. INTRODUCTION

There are certain natural calamities that cause great damage and impact. One of such natural calamity is flood. Alertness for safety is the important part concerning with flood. The sudden arrival of flood can cause a huge amount of loss of life. Hence prediction of flood is necessary for providing safety measures. There are lots of technique for flood prediction. Now-a-days trend is of artificial brain that can mimic human brain so it should adapt with changing parameters and work fast. Machine Learning is the major application of artificial intelligence. Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. The use of machine learning in flood prediction has sorted out many problems. Many open source software's uses C or C++ programming which is less efficient than python programming. Apache SystemML has R-like and python language support. The optimization of collecting flood related data will become easy. The proposed Apache SystemML based

flood prediction is expected to be more efficient and scalable than previously developed system.

II. LITERATURE REVIEW

Anil Kumar Lohani et al. [1] proposed the system to improve real time flood forecasting using fuzzy inference system. The system uses Takagi Sugeno (T–S) fuzzy inference system. The proposed modified fuzzy inference systems provide an option of analyzing and computing cluster centers and membership functions for two different hydrological situations, i.e. low to medium flows (frequent events) as well as high to very high flows (rare events) generally encountered in real time flood forecasting. TSC-T–S fuzzy model provide reasonably accurate forecast with sufficient lead-time. The system lacks the proper prediction. RC Deo and Mehmat Sahin [2] proposed a system incorporating data driven models for drought prediction by Artificial Neural Network algorithm. An evaluation of the model performance based on statistical rules yielded time-averaged Coefficient of Determination. The complex optimum model, the Levenberg–Marquardt and Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasiNewton back propagation algorithms were utilized to train the network, tangent and logarithmic sigmoid equations. Turgay Partal [3] proposed the system that uses wavelet–neural network structure that combines wavelet transform and artificial neural networks to forecast the river flows of Turkey.The feed-forward back-propagation method was studied with respect to artificial neural network applications to water resources data. The wavelet and feed-forward back-propagation model was only superior to the other models in terms of selected performance criteria.

Biswajit Pradhan et al. [4] presented a system to access and evaluate the prediction capability of SVM technique with different kernel functions for spatial predictions of flood occurrence. The flood inventory was partitioned into training and testing data setsthrough random selection.SVM technique is an efficient and powerful technique for flood susceptibility but somewhere it lacks the exact accuracy.

G.B Sahooo et al.[5] presented the system which uses application of artificial neural networks (ANNs) to assess flash floods and their attendant water quality parameters using measured data of a Hawaii stream. The paper illustrates that ANNs predict stream flow with a correlation coefficient (R)and turbidity and specific conductance with R-values. The system demonstrates that the upstream water quality parameters depend on weather forces and land use of the watershed and the downstream water quality parameters additionally influenced by oceanic tides. The system can be improved if the quality of data collection parameter is increased and developed.

C. Shu et al. [6] presented a form of methodology using adaptive neuro-fuzzy inference systems for flood estimation at ungauged. The approach has the system identification and interpretability of fuzzy models and the learning capability of artificial neural networks. The system uses subtractive clustering algorithm and hybrid learning algorithm. The software used here for training are not python based and are not that much efficient.

Mehdi Razaeian et al.[7] presented the system for prediction of monthly discharge volume for reservoir management and evaluation of drinking water supplies. The work shows the study of different algorithms including resilient back propagation, scaled conjugate gradient, variable learning rate and Levenberg-Marquardt. The algorithms were trained and tested on previously recorded data. The system is area specific i.e. only for semi-arid regions.

Gokmen Tayfur et al. [8] presented the system for flood hydrograph prediction. The prediction was performed using ANN, GA and fuzzy logic methods. Tools used here were powerful. There was no significant estimation problem. The approximation obtained from this system was not perfect and was only satisfactory. Hence efficient algorithms are needed which are less complex an less bulky.

Fan Liu et al. [9] proposed an approach for flood forecasting based on deep learning via integrated stacked auto encoders with BP neural network. In this approach, multiple SAE-BP modules are adopted to simulate their corresponding categories of data. The issues in this approach is imbalance of data distribution.

Wenrui Huang et al. [10] presented a model that uses a feed forward, back-propagation network structure with an optimized conjugated training algorithm. The model has better accuracy than ARIMA model. The main drawback of this system is that it does not go beyond one or two hidden layer for problematic non-convex application.

III. PROPOSED SYSTEM

The aim of this project is to get all the rainfall data of India and from a dataset containing yearly rainfall data. By providing real time input to different models of machine learning, those are Logistic Regression, Support Vector Machine, K-Nearest Neighbors and Decision Tree Classifier. The inputs provided to models are pre-processed and patterns are extracted by getting maximum accuracy. The data provided is split into a Training set and Test set. It is split in the ratio of 7:3. The all four models are used to predict and by comparing all the results of model and considering the confusion matrix of all the models the accuracy is determined. The best model is chosen by comparing the accuracy of each model.

The objective of Flood Prediction using Machine Learning is to design a model to predict the flood using the rainfall data. The prediction of different models is taken and compared within each other to find the best model that has high accuracy. The flood can be predicted in different states of India in different months. The confusion matrix of different models in Machine learning is considered to evaluate the accuracy and precision of the system.

The prediction accuracy of the different models is evaluated using data validation, and the results are compared to get accuracy. The accuracy of the training dataset, accuracy of the testing dataset, false-positive rate, specification, precision, and recall are calculated by comparing algorithms using python code.

The steps involved are:

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

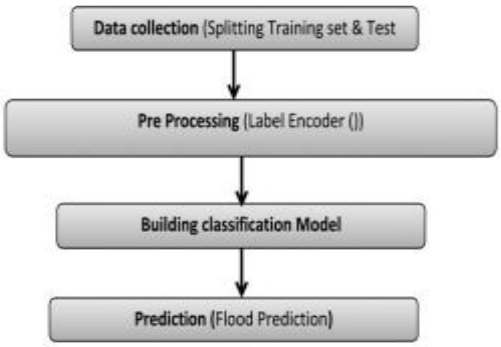

Fig 1: System architecture

IV. RESULTS AND DISCUSSION

ML Mode-ls	Precision	Recall	F1-Score	Sensitivity	Specificity	Accuracy (%)
Logistic Regression	0.95	0.98	0.96	0.96	0.99	99.39
Support Vector Machine	0.93	0.88	0.90	0.88	0.99	98.37
K-Nearest Neighbours	0.9	0.81	0.85	0.81	0.99	97.47
Decision Tree Classifier	0.72	0.75	0.73	0.75	0.97	95.07

Table 1: Comparison of Accuracy Results

Table 1 compares the accuracy and precision results of the Logistic Regression, Support Vector Machine, K-

Nearest Neighbors, Decision Tree Classifier and Random Forest Classifier algorithms. From the above calculated values we can observe that Support Vector Machine and Logistic Regression has comparatively better results.

Algorithm	TP	TN	FP	FN	TPR	TNR	FPR	FNR	PPV	NPV
LR	76	752	4	1	0.98	0.99	0.01	0.02	0.95	0.99
SVC	68	751	5	9	0.88	0.99	0.01	0.12	0.93	0.98
KN-N	63	749	7	14	0.81	0.99	0.01	0.19	0.9	0.98
DTC	58	734	22	19	0.75	0.97	0.03	0.25	0.72	0.97

Table 2: Comparison of Confusion Matrix Parameters

The above table shows the comparison of confusion matrixparameters of Logistic Regression, Support Vector Classifier, K-Nearest Neighbors, Decision Tree Classifier, Random Forest Classifier algorithms. From the Calculated result we can conclude that Logistic Regression has produced best results.

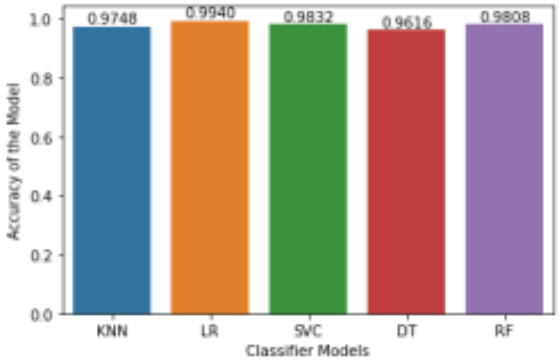

Fig 2: Accuracy vs Algorithms

V. CONCLUSION

The machine learning system ignited by data cleaning and processing, replacing or removing the null values, model building and evaluation. At the end the flood prediction model has given different accuracy results from four different models. From the above results and analysis, the best algorithm for flood prediction is Logistic Regression with (99%).

VI. REFERENCES

[1]. A.K Lohani, N.K Goel , Bhatia K., "Improving real time flood forecasting using fuzzy inference system. "Journal of hydrology, 2014, 509, 25-41.

[2]. Ozgar Kişi, "Streamflow forecasting using different artificial neural network algorithms," Journal of Hydrologic Engineering, 2007, 12, 532-539.

[3]. R.C. Deo, Mehmat Şahin, "Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydro meteorological parameters and climate indices in eastern Australia," Atmospheric research 2015, 161, 65-81.

[4]. Biswajit Pradhan, M.S. Tehrany, Noordin Ahmad, "Flood susceptibility assessment using GIS-based support vector machine model with different kernel types," Catena 2015, 125, 91-101.

[5]. G.B. Sahoo,C. Ray, E.H De Carlo, "Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii," Journal of Hydrology 2006, 327, 525-538

[6]. C. Shu, Taha BMJ Ouarda, "Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. ," Journal of Hydrology 2008, 349, 31-43.

[7]. Mehdi Rezaeian, Hossein Tabary, Hirad Abghari, "Prediction of monthly discharge volume by different artificial neural network algorithms in semi-arid regions," Arabian Journal of Geosciences 2013, 6, 2529- 2537.

[8]. Gokmen Tayfur, Vijay P. Singh, Tommaso Moramarco and Silvia Barbetta, "Flood Hydrograph Prediction Using Machine

Learning Methods," MDPI, Water Journal, 2018.

[9]. Fan Liu, Feng Xu, Sai Yang," A Flood Forecasting Model based on Deep Learning Algorithm via Integrating Stacked Auto encoders with BP Neural Network," IEEE Third International Conference on Multimedia Big Data,2017.

[10]. Wenrui Huang, Bing Xu, Amy Chan Hilton, "Forecasting flows in Apalachicola river using neural networks," Hydrological Processes,30 June 2004.

Cite this article as :

Dr. C. Murugamani, K Lakshmi Prasanna, R. Mamatha, "Flood forecasting by using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8, Issue 5, pp.329-333, September-October-2022.