Flash Flood Prediction Using Machine Learning

Tanmay Damle- Dept of computer Engineering VESIT, 2020.tanmay.damle@ves.ac.in
Harsh Deshmukh- Dept of computer Engineering VESIT, 2020.harsh.deshmukh@ves.ac.in
Anish Nair- Dept of computer Engineering VESIT, 2020.anish.nair@ves.ac.in
Digvijay Kocharekar- Dept of computer Engineering VESIT, 2020.digvijay.kocharekar@ves.ac.in
Dr. Mrs.Sujata Khedkar- Dept of computer Engineering, VESIT, sujata.khedkar@ves.ac.in

Abstract–

The societal impacts of flash floods are more significant than any other weather-related hazard. Flash floods are often manifested in the form of infrastructure damage, flooding roadways and bridges, creating deadly hazards to motorists and inundation of crops and pasture. For instance, the flash floods in Mumbai that occurred on 26th July 2005 caused a heavy amount of damage to human life and property.This system aims to predict flash floods in various parts of our country by learning from past data. The user will input parameters such as location, expected rainfall, water levels in nearby reservoirs etc. and the system will give the probability of flash floods in the region in the upcoming few days. A number of various ML models namely LogisticRegression,RandomForestClassifier,KNeighborsClassifier,VotingClassifier, GradientBoost are used and their performance is evaluated. Evaluation measures such as accuracy, recall, f1 score are used to validate the predictions. Based on these evaluation metrics the proposed system works most efficiently for GradientBoostingClassifier Algorithm with an accuracy of about 91%.

Introduction–

India, a tropical nation with around 400 rivers and high rainfall throughout much of the year in diverse places, is vulnerable to flooding. Flooding, particularly flash floods in particular, which happen so quickly—within 6 to 8 hours—cause a lot of damage.Flash floods are a frequent natural disaster that can result in substantial destruction and fatalities. Although we are aware of the devastation that flash floods can inflict, no reliable system exists to predict catastrophic floods. The current systems are inefficient to a great extent, which lowers forecast accuracy. Traditional techniques of forecasting flash floods frequently rely on labour-intensive and erroneous manual observation and analysis. Accurate flood predictions could be made using machine learning techniques, helping to lessen the effects of floods. The economic burden of flash floods can also be enormous. The average cost of a flash flood in India, according to the National Oceanic and Atmospheric Administration (NOAA), is $8.2 million. Data collection for

a machine learning project that predicts flash floods would include information on the numerous elements that cause them, such as terrain and rainfall amounts. A machine learning algorithm can be trained using this data to forecast the likelihood of a flash flood occurring in a specific area. The system has been developed and evaluated using historical information gathered from different parts of India, and the findings significantly outperform conventional models. The suggested technique avoids false warnings and pointless evacuation orders, which can frighten residents of the impacted areas and not only offers correct predictions but can also generate fear. In general, the suggested flash flood prediction system has the ability to give inhabitants of the affected areas early and precise warnings, enabling them to take the appropriate preparations and lessen the damage caused by flash floods.

Literature Survey–

Amir Mosavi etal[1] Flood prediction using Machine Learning Models: This study elaborates on the many ML algorithms that are most effective for forecasting flash floods such as ANNs.

Avish Jain etal[2] India flood inventory: Creation of a multi-source national geospatial database to facilitate comprehensive flood research. This paper is about the Indian Flood Inventory dataset used for the proposed system. It shows past flood data bases on various parameters.

Joel T. de Castro etal[3] Dissecting the flash flood forecasting problem. This paper proposed to develop a technology on Flash Flood Warning System Using SMS with advanced warning information based on prediction algorithm devised by the researcher regarding increasing water level and water speed

Daniele Norbiato etal[4] Flash flood warning based on rainfall thresholds and soil moisture condition. This paper discusses the working and implementation of KNN ML model. It gives the mathematical representation of the same.

Jonathan J. Gourley etal[5] Evaluation of past, present and future tools for radar-based flash-flood prediction in the USA. This study describes the tools proposed over recent decades in the`1 USA to predict flash flooding and evaluates them using a common observational data set.

Theresa M Carpenter etal[6]. Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model.In this paper. We diagnose the manner with which rainfall input and parametric uncertainty influence the character of the flow simulation uncertainty in a validated distributed hydrologic model.

Gourley, J. J. etal[7]. Evaluation of Tools Used for Monitoring and Forecasting Flash Floods in the United States, Weather and Forecasting. This paper evaluates flash-flood guidance (FFG) values and recently developed gridded FFG (GFFG) used by the National Weather Service to monitor and predict imminent flash flood..

D. Gaur etal[8] "Machine Learning-Based Flash Flood Prediction Using Wireless Sensor Networks. The proposed model collects data from multiple sensors, including rainfall, temperature, and humidity, to predict flood events. The model uses machine learning algorithms, such as decision trees and support vector machines, to predict the likelihood of flood events based on the collected sensor data.

Y. Zhang etal[9], "A Novel Method for Flash Flood Prediction Based on LSTM Networks and Gray Model,". This paper proposes a novel method for flash flood prediction based on long short-term memory (LSTM) networks and gray model. The proposed method uses historical rainfall and river level data to predict flood events.

M. M. Islam etal[10], "A Machine Learning-Based Framework for Flood Prediction,". This paper presents a machine learning-based framework for flash flood prediction. The proposed framework uses multiple models, including decision trees, artificial neural networks, and support vector machines, to predict flood events.

A. M. Azad etal[11],"Flash Flood Prediction using Machine Learning Techniques: A Comparative Study,". This paper presents a comparative study of various machine learning techniques for flash flood prediction. The study compares the performance of decision trees, random forests, and support vector machines on a real-world dataset. The results show that random forests outperform the other two models in terms of accuracy and computational efficiency.

S. S. Saha etal[12] "A machine learning approach to flash flood prediction using hydro-meteorological data,". This paper proposes a machine learning approach for flash flood prediction using hydro-meteorological data. The proposed approach uses artificial neural networks to predict flood events based on rainfall, temperature, and river level data. The results show that the proposed approach outperforms traditional flood prediction methods in terms of accuracy and computational efficiency.

H. R. Rahman etal[13] "Flash flood prediction using artificial neural network based on improved particle swarm optimization,". This paper proposes a flash flood prediction model based on artificial neural networks and improved particle swarm optimization. The proposed model uses historical rainfall and river level data to predict flood events.

A. Farid etal[14], "Flash Flood Prediction Using Machine Learning Techniques: A Review. This paper provides a comprehensive review of the state-of-the-art machine learning techniques for flash flood prediction. The review covers various types of machine learning models, datasets, and evaluation metrics used in previous studies.

L. Ma etal[15] "Flash flood prediction using a hybrid model combining Bayesian regularization and generalized regression neural network,". This paper proposes a hybrid model combining Bayesian regularization and generalized regression neural network for flash flood prediction. The model uses rainfall and river level data to predict flood events.

Overall all the systems studied so far have been very region specific and have not been able to provide predictions with reliable accuracy. Moreover data collection is a big challenge for such systems as the factors determining the occurrence of flash floods are very large in number and complex.

Methodology–

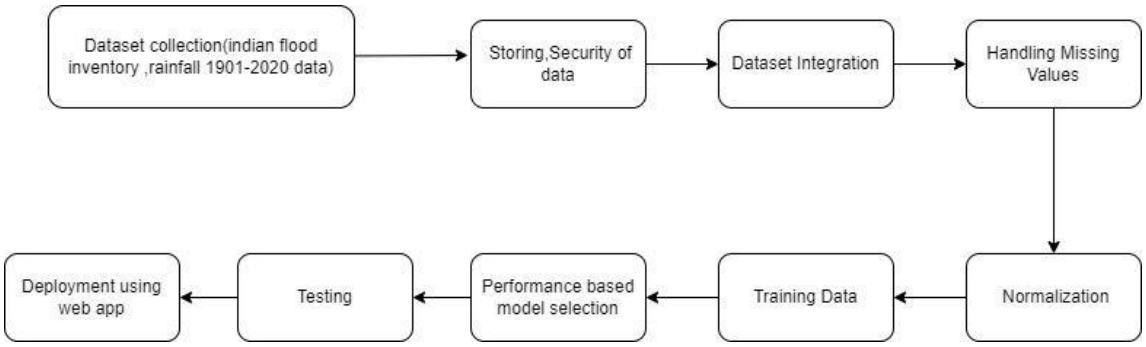

Fig 1.0- Methodology Block Diagram

Fig 1.0 shows the the block diagram of the methodolog implemented for the proposed system.

I. Data Extraction/Collection:-

Data extraction is the process of collecting data needed for the proposed system. It includes sourcing data from multiple sources and applying various techniques for creating a unified dataset. The dataset used for this system is the IFI(Indian Flood Inventory) rainfall data. The IFI Dataset contains rainfall data from 3 different sources where each source is specified by the 'UID' column which is an unique id from where the data is sourced. The data is stored in a CSV(Comma separated Values) format. For a more accurate and small scale demonstration data of Kerala state has also been used which is much more cleaner and organized.

II. Data Preprocessing:-

It is the process of filtering and cleaning the dataset before it is fed into the Machine Learning Model. Preprocessing is done on the dataset by removing NaN values, correcting the date formats, and removing inconsistent entries in the dataset. Another aspect of preprocessing is that the data should not be biassed. To solve this problem, we introduce synthetic data by using k nearest neighbors algorithm. This balances the classes. Figure 1.1 and 1.2 show the class distribution before and after applying the process. Most Machine Learning algorithms cannot work with categorical data and needs to be converted into numerical data. One approach to solve this problem can be label encoding where we will assign a numerical value to these labels, but this can add bias in our model as it will start giving higher preference to values encoded as higher values. The solution to this problem is using 'One hot encoding'. The 'Main Cause' attribute of the dataset is a string, each of which represents the cause for the particular flood event. Main cause is one of the prime features involved in the prediction. This string value is converted into numeric value using 'One Hot Encoding'.

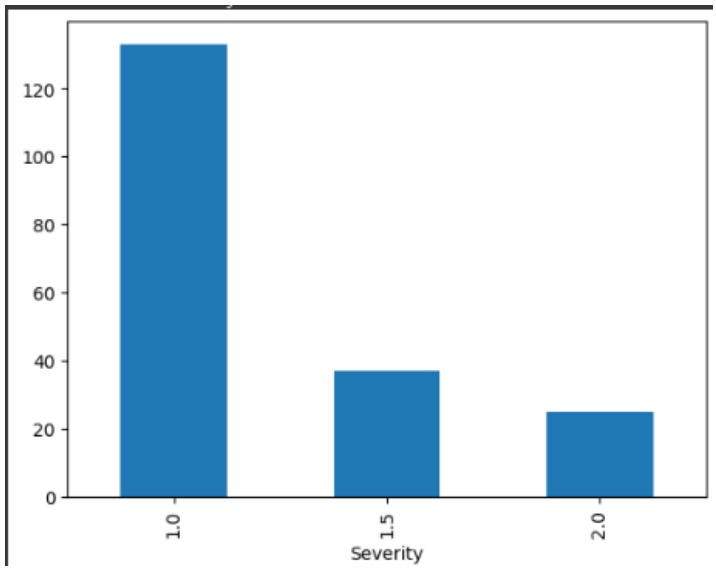

Fig 1.1- Class Distribution of the Dataset

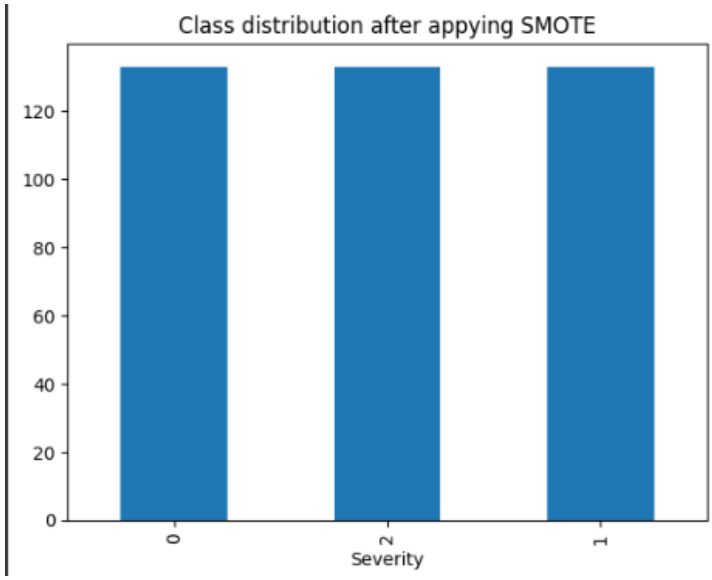

Fig 1.2- Class Distribution after Preprocessing

III. Dividing dataset into train and test sets:-

After Preprocessing is applied the dataset is then divided into two sets namely the 'Training Set' and the 'Test Set' in the ratio 80:20. The training set is passed to the ML model and the test set is used to verify the accuracy of the model.

IV. Applying ML Models and Testing:-

Classification and Regression models are applied to the training set. The models implemented are LogisticRegression, KNeighbours Classifier, GradientBoostingRegressor, GradientBoostingClassifier, Ensemble models. The outcome of these models is evaluated primarily using the accuracy score of the model. GradientBoostingClassifier yields the highest accuracy and is thus selected for the same. The result of the implementation is then documented in a tabular format. The proposed system is then deployed as a web application.

Results–

Table 1.0 shows the accuracy of the various models employed in the project. It is evident that GradientBoostingClassifier is the most appropriate model for the same.

Model Name	Accuracy
Logistic Regression	40%
K-Neighbours Classifier	75%
Gradient Boosting Regressor	57.19%
Gradient Boosting Classifier	91.25%
Ensemble Models	62.5%

Table 1.0- Various Models and their accuracy

From the above table it is evident that Gradient Boosting Classifier gives the best results. Table 1.1 shows the detailed evaluation metrics of Gradient Boosting Classifier algorithm.

Severity level	Precision	Recall	F1-Score	Support
0	0.89	0.84	0.86	19
1	0.88	1.0	0.94	29
2	0.97	0.88	0.92	32
Accuracy			0.91	80
Macro avg	0.91	0.91	0.91	80
Weighted avg	0.92	0.91	0.91	80

Table 1.1- Classification report of Gradient Boosting Classifier

Conclusion–

Flash Floods have wide ranging socio-economic impacts wherever they occur. This paper studies one of the approaches to predict the flash floods by using machine learning by analyzing the existing systems. The paper outlines the methodology adopted for the same and compares various algorithms based on their performance metrics. The best performing model is then selected and further detailed analysis of its performance metrics like ‘F1-score’, ‘Recall’ is carried out.

References–

[1] A. Mosavi, P. Ozturk, and K. Chau, “Flood Prediction Using Machine Learning Models: Literature Review,” Water, vol. 10, no. 11, p. 1536, Oct. 2018, doi: 10.3390/w10111536

[2] A. Jain, R. R. Baishya, and S. Haobam, "India flood inventory: Creation of a multi-source national geospatial database to facilitate comprehensive flood research," in 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Yokohama, Japan, 2019, pp. 3720-3723. doi: 10.1109/IGARSS.2019.8898996.

[3] J. T. de Castro, G. M. Salistre, Y.-C. Byun, et al., "Dissecting the flash flood forecasting problem," in 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 2017, pp. 4255-4258. doi: 10.1109/IGARSS.2017.8128102.

[4] D. Norbiato, M. Borga, S. Degli Esposti, E. Gaume, et al., "Flash flood warning based on rainfall thresholds and soil moisture condition," in 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 2015, pp. 3148-3151. doi: 10.1109/IGARSS.2015.7326589.

[5] J. J. Gourley, Z. L. Flamig, Y. Hong, et al., "Evaluation of past, present and future tools for radar-based flash-flood prediction in the USA," in 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, Australia, 2013, pp. 2985-2988. doi: 10.1109/IGARSS.2013.6723422

[6] T. M. Carpenter, K. P. Georgakakos, et al., "Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model," in 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 2012, pp. 1556-1559. doi: 10.1109/IGARSS.2012.6351111.

[7] J. J. Gourley, J. M. Erlingis, Y. Hong, and E. B. Wells, "Evaluation of Tools Used for Monitoring and Forecasting Flash Floods in the United States," Weather and Forecasting, vol. 27, no. 4, pp. 879-897, 2012. doi: 10.1175/WAF-D-11-00114.1.

[8] D. Gaur, N. Singh, and A. Kumar, "Machine Learning-Based Flash Flood Prediction Using Wireless Sensor Networks," in IEEE Sensors Journal, vol. 21, no. 8, pp. 9644-9654, 15 April 2021

[9] Y. Zhang, X. Song, Y. Liu and Z. Yan, "A Novel Method for Flash Flood Prediction Based on LSTM Networks and Gray Model," in IEEE Access, vol. 8, pp. 39059-39070, 2020, doi: 10.1109/ACCESS.2020.2974238.

[10] M. M. Islam, R. Siddique, and M. A. Hossain, "A Machine Learning-Based Framework for Flood Prediction," in IEEE Access, vol. 5, pp. 1813-1822, 2017, doi: 10.1109/ACCESS.2017.2677644.

[11] A. M. Azad, Y. B. Kim and H. S. Kim, "Flash Flood Prediction using Machine Learning Techniques: A Comparative Study," 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China, 2018, pp. 325-329, doi: 10.1109/BIGCOMP.2018.00056.

[12] S. S. Saha, S. K. Saha, M. A. Rahman and M. H. Rahman, "A machine learning approach to flash flood prediction using hydro-meteorological data," 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), Tokyo, Japan, 2019, pp. 119-124, doi: 10.1109/ICIEA.2019.8833903.

[13] H. R. Rahman, M. B. I. Reaz, M. A. M. Ali, and M. N. Mohd Salleh, "Flash flood prediction using artificial neural network based on improved particle swarm optimization," in Measurement, vol. 135, pp. 470-480, 2019, doi: 10.1016/j.measurement.2018.11.057.

[14] A. Farid et al., "Flash Flood Prediction Using Machine Learning Techniques: A Review," in Water, vol. 11, no. 10, 2019, doi: 10.3390/w11102109.

[15] L. Ma, C. Liao and Z. Guan, "Flash flood prediction using a hybrid model combining Bayesian regularization and generalized regression neural network," Journal of Hydrology, vol. 537, pp. 221-232, 2016, doi: 10.1016/j.jhydrol.2016.03.044.