IJMRSETM

e-ISSN: 2395 - 7639

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

Volume 10, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.580

+91 99405 72462 +9163819 07438 ijmrsetm@gmail.com www.ijmrsetm.com

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 5, May 2023 |

Flood Prediction using Machine learning

D.Keerthi Reddy, B.Yashwanth Sai, S.Mrudula, P.Sripadh, K.Sushma, Dr.Jayanthi Kannan

UG Students, Dept. of I.S.E, Jain University, Bangalore, India

Head of the Department, Dept. of I.S.E, Jain University, Bangalore, India

ABSTRACT: Flooding is the most common natural disaster on the planet, affecting hundreds of millions of people and causing between 6,000 and 18,000 fatalities every year – of which 20 percent are in India. Reliable early warning systems have been shown to prevent a significant fraction of fatalities and economic damage, but many people don't have access to those types of warning systems. So, we're building Flood prediction system Based on ML or AI.This advancement of the prediction system provides cost-effective solutions and better performance. In this, a prediction model is constructed using rainfall data to predict the occurrence of floods due to rainfall. The model predicts whether "flood may happen or not" based on the rainfall range for particular locations. Indian district rainfall data is used to build the prediction model. The dataset is trained with various algorithms like K-Nearest Neighbors, XGBoost etc.,

KEYWORDS: Supervised learning, Machine Learning, Floods, XGBoost algorithm and KNearest Neighbours

I. INTRODUCTION

Every year, India is the topmost flood-prone disaster place in the world. Mostly water logging in urban cities occurs in low-lying areas. Moreover, the increase in water logging is due to some fundamental points such as surface runoff, relative altitude, and not enough path of the water to drainage So, flood forecasting is essential at these places.
In a recent year, there were many parts of countries which are prone to flood like Assam, Bihar, Goa, Odisha, Pune, Maharashtra, Tamil Nadu, Karnataka, Kerala, and Gujarat.

In the year 2015 rainfall, Chennai received 1049 millimeters (mm) of rainfall in November. Since 1918, 1088 mm of precipitation was the best recorded in November. Between October and December, the average rainfall in Kanchipuram district is 64 cm. It received the heaviest rainfall of 181.5 cm, which is 183% higher against average precipitation. In the Tiruvallur district, the average rainfall is 59 cm but recorded 146 cm of rain

There was much research for prediction of flood ahead, but not many methods give the estimate with high accuracy. The flood prediction analysis majorly uses Machine Learning (ML). There are many methods in machine learning to predict the problem with higher accuracy.

In this work, we have proposed to estimate the flash flood to prevent places that are prone to flood risk. The approach is to the establishment of the ML algorithm model. It incorporates the flood factor to estimate short term prediction in an urban area with higher accuracy

II. RELATED WORK

The Indian Meteorological Department data set consists of data from nine states from 1901 to 2018. There are 18 total features in the data set. For our work, the attributes used in the model are June, July, August, and September. As this is a classification problem, the output is in the form of 0 and 1, where "0" is when the average rainfall of monsoon seasonal month (June– September) is greater than the threshold rainfall level in June–September of the specific area . Similarly, "1" is when the average rainfall of monsoon seasonal month (June–September) is less than the threshold rainfall level in June–September. The threshold describes normal rainfall in the area, so, if supposedly the rainfall water level is higher in that area at that time, it will send an alert to all the residents in that area. Every area has their respective threshold water level, and based on that, the target class is determined. After defining all the inputs and output, one by one, all the supervised ML algorithms are tested on that. As all supervised ML algorithms are being applied, some of them show an overfitting issue and some of them score less accurately. So, to overcome this issue, GridSearchCV was applied. In the KNN algorithm, the param_grid 'n_neighbors,' 'weights,' and 'metric' was taken. The value that is considered under 'n_neighbors' is in the form of a list ranging from 1 to 31. Similarly, for 'weights,' the value considered is 'uniform' and 'distance.' For 'metric,' the values are 'euclidean' and 'manhattan.'.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 5, May 2023 |

Though flooding is seen as a common environmental threat globally, it has dramatically increased recently due to climate change, impacting underdeveloped and developing countries dangerously. For example, in most developing countries like Ghana, flooding has affected over four million people in terms of property damage, loss of lives, income and spread of diseases, resulting in economic harm beyond USD780 million. At least one major flood disaster does occur yearly. The recurring incidences of flooding and associated calamitous socio-economic risks and anticipated increase of its prevalence soon in cities of developing countries such as Ghana have necessitated an intelligence system to offer efficient and early warning of its occurrence. In this study, we explore the potential of the machine learning (ML) computing paradigm to propose a flooding prediction model. Specifically, four state-of-the-art ML algorithms, namely long short-term memory (LSTM), extreme gradient boosting (XGBoost), random forest (RF) and extremely randomised trees (Extra Trees), are used to implement four different flood prediction models. We measure the performance of our developed models with multiple statistical performance evaluators.

III. PROPOSED ALGORITHM

A. *Design Considerations:*

- We propose this application that can be considered a useful system since it helps to reduce the limitations obtained from traditional and other existing methods.
- In proposed system, we implement a Machine Learning algorithms for getting insights from the complex patterns in the data in order to predict the floods.
- This technique is computationally inexpensive because of its simple architecture.

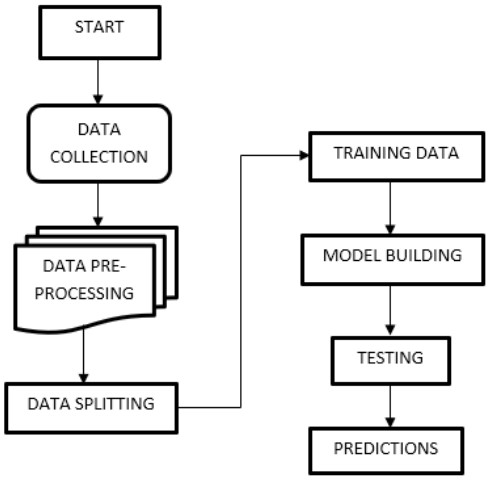

Fig 1. Block diagram

B. *Description of the Proposed Algorithm:*

XGBoost:

- XGBoost is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.
- XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class right now.
- Bagging: Now imagine instead of a single interviewer, now there is an interview panel where each interviewer has a vote. Bagging or bootstrap aggregating involves combining inputs from all interviewers for the final decision through a democratic voting process.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 5, May 2023 |

- XGBoost and Gradient Boosting Machines (GBMs) are both ensemble tree methods that apply the principle of boosting weak learners (CARTs generally) using the gradient descent architecture. However, XGBoost improves upon the base GBM framework through systems optimization and algorithmic enhancements.

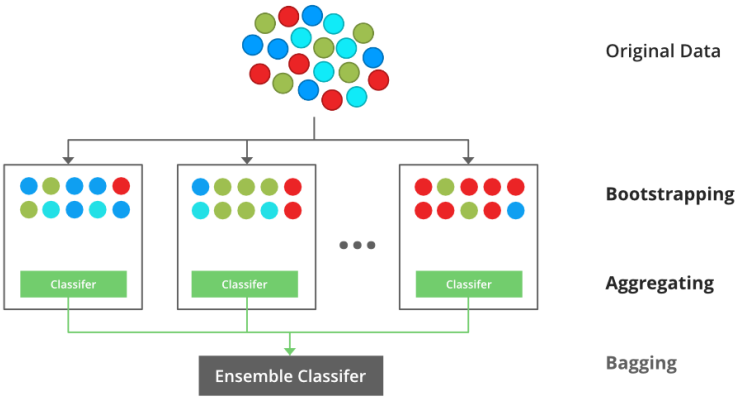

Fig 2. Working of XGBoost algorithm

K Nearest Neighbors:

- K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.
- It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.
- Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset.
- The K-NN working can be explained on the basis of the below algorithm:
  Step-1: Select the number K of the neighbors
  Step-2: Calculate the Euclidean distance of K number of neighbors
  Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
  Step-4: Among these k neighbors, count the number of the data points in each category.
  Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.
  Step-6: Our model is ready.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 5, May 2023 |

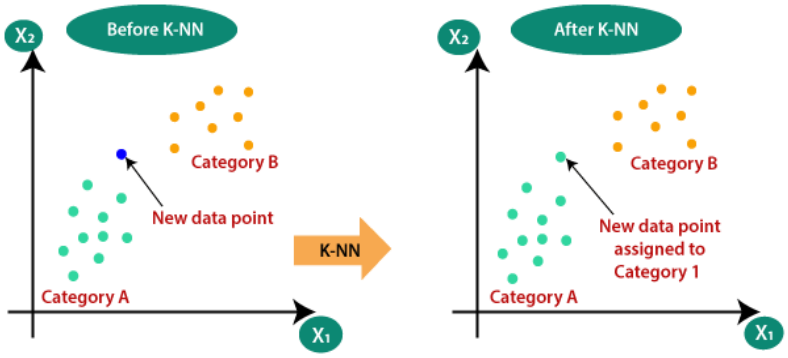

Fig 3. Working of KNN algorithm

Decision Trees:

- A tree has many analogies in real life, and turns out that it has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal.
- A decision tree is drawn upside down with its root at the top. In the image on the left, the bold text in black represents a condition/internal node, based on which the tree splits into branches/ edges. The end of the branch that doesn't split anymore is the decision/leaf, in this case, whether the passenger died or survived, represented as red and green text respectively.
- Although, a real dataset will have a lot more features and this will just be a branch in a much bigger tree, but you can't ignore the simplicity of this algorithm. The feature importance is clear and relations can be viewed easily. This methodology is more commonly known as learning decision tree from data and above tree is called Classification tree as the target is to classify passenger as survived or died. Regression trees are represented in the same manner, just they predict continuous values like price of a house. In general, Decision Tree algorithms are referred to as CART or Classification and Regression Trees.
- So, what is actually going on in the background? Growing a tree involves deciding on which features to choose and what conditions to use for splitting, along with knowing when to stop. As a tree generally grows arbitrarily, you will need to trim it down for it to look beautiful. Let's start with a common technique used for splitting.

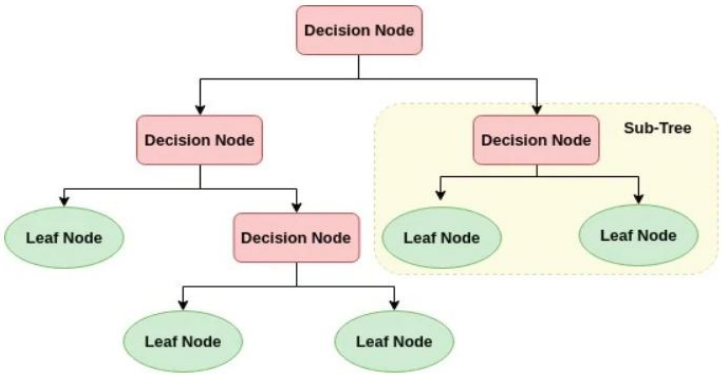

Fig 4. Working of Decision tree algorithm

Logistic Regression:

- Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable (target) is categorical.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 5, May 2023 |

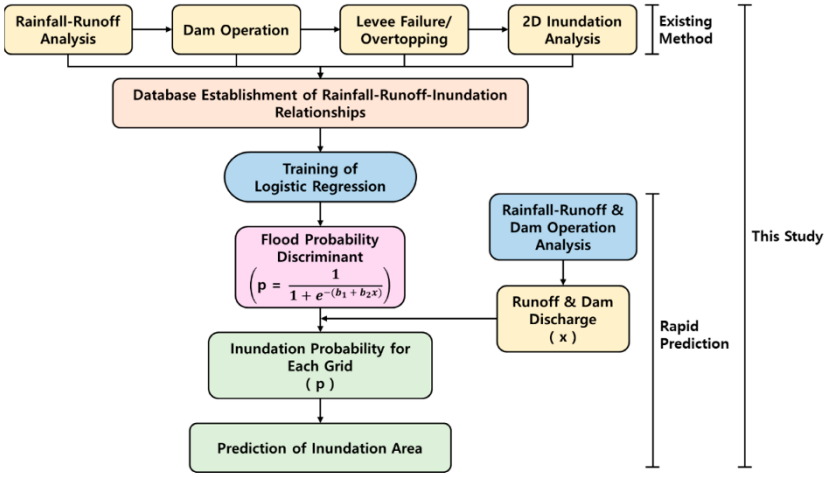

Fig 5. Working of Logistic regression algorithm

IV. SIMULATION RESULTS

Home:
In our project, we are detecting whether the floods will occur or not.

Fig 6. Home

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 5, May 2023 |

Upload:
Here in this project we are uploading the dataset through which we are working.

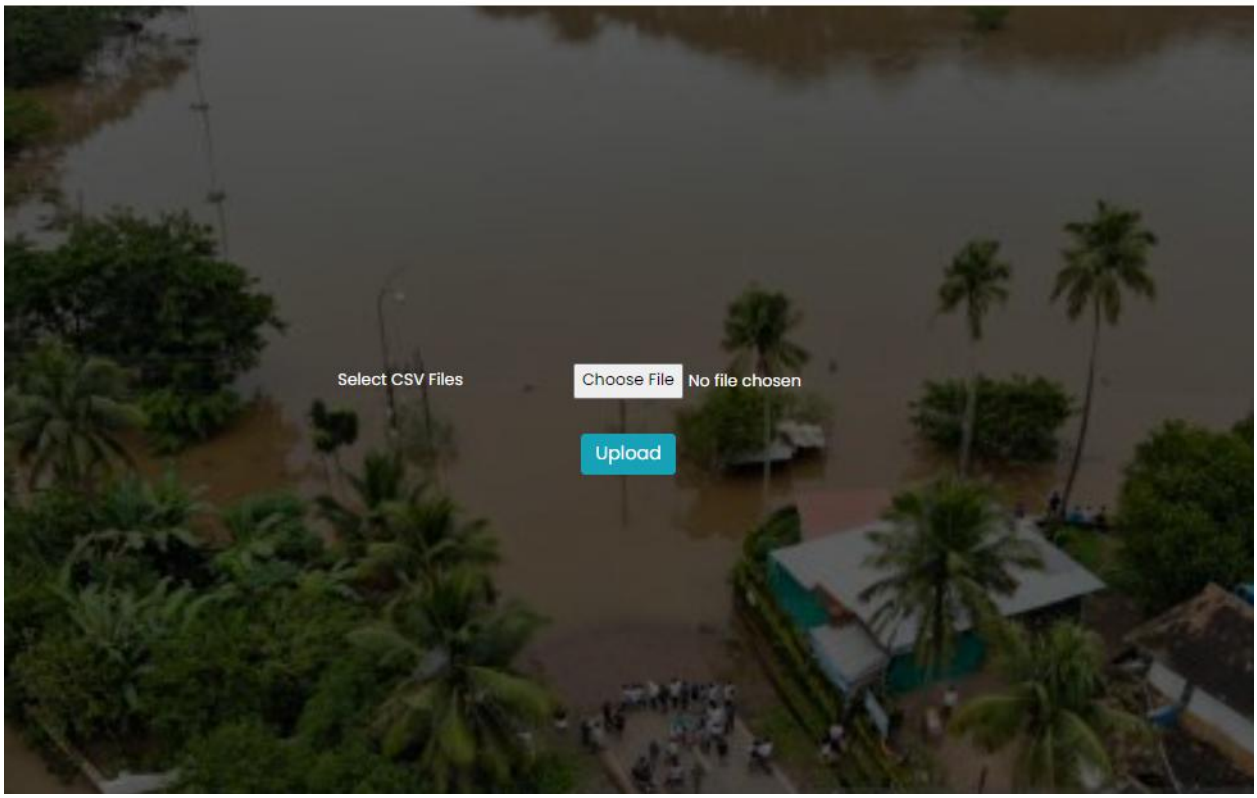

Fig 7. Upload file

Model Training:
We train the models with the algorithms.

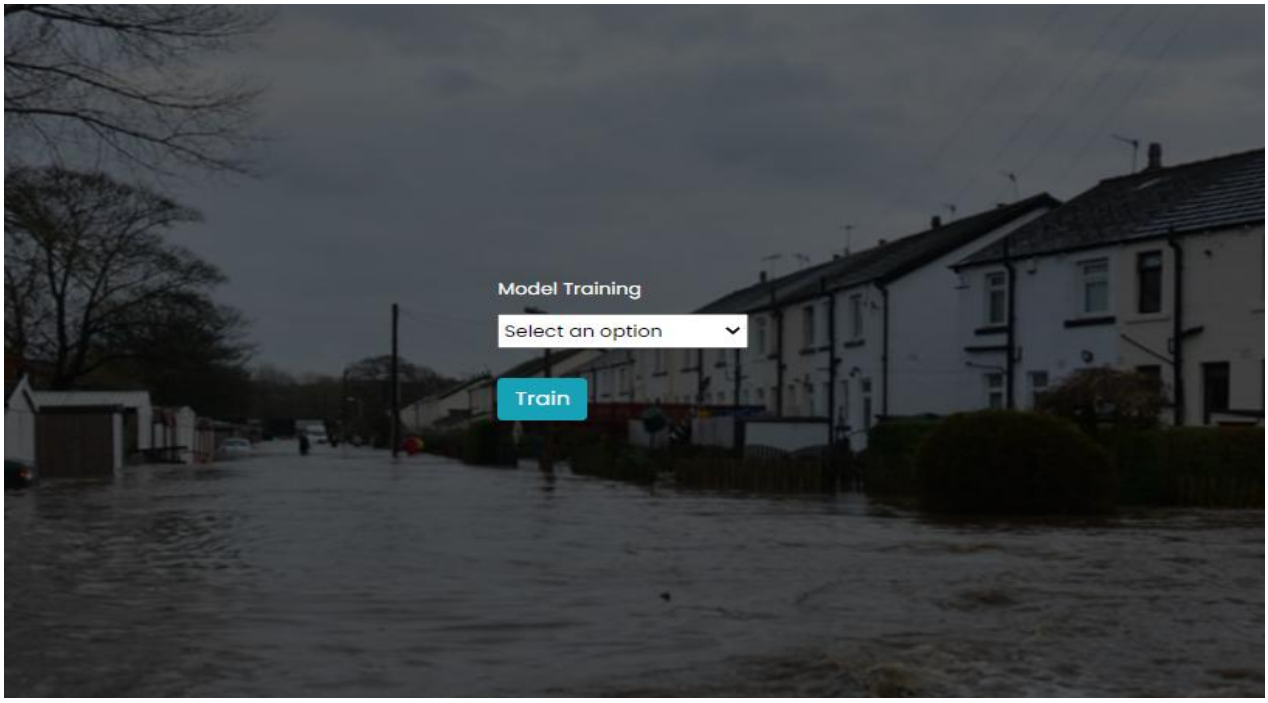

Fig 8. Training the model

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 5, May 2023 |

Predictions:
This is the final step where the occurrence of flood is predicted.

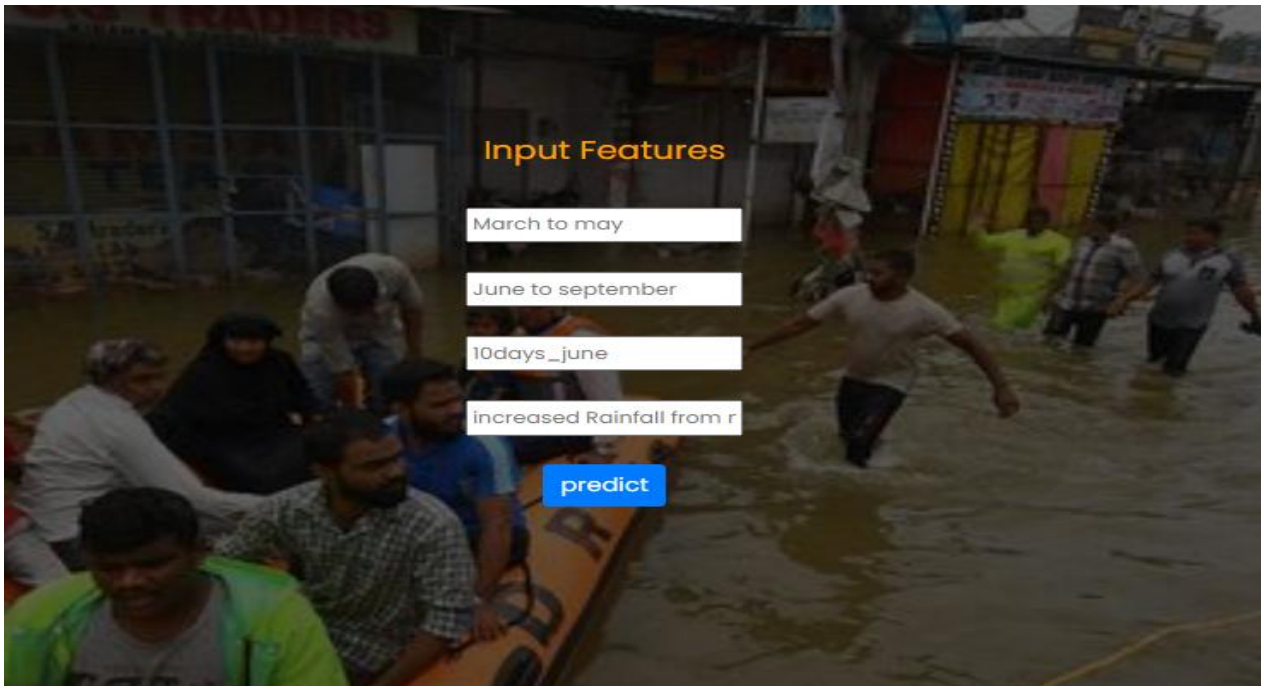

Fig 9. Prediction page with input features

S.NO	Test cases	I/O	Expected O/T	Actual O/T	P/F
1	Read the dataset.	Dataset path.	Dataset need to read successfully.	Dataset fetched successfully.	P
2	Performing pre-processing on the dataset	Pre-processing part takes place	Pre-processing should be performed on dataset	Pre-processing successfully completed.	P
3	Model Building	Model Building for the clean data	Need to create model using required algorithms	Model Created Successfully.	P

V. CONCLUSION AND FUTURE WORK

We have successfully developed a system to predict whether the floods will occur or not in this application. This is created in a user-friendly environment with Python programming and Flask. The system is likely to gather data from the user in order to predict whether there is a chance of flood occurring or not.We intend to investigate prediction approach with the revised data set and employ the most accurate and relevant machine learning algorithms for detection.

REFERENCES

[1] R. Balamurugan, K. Choudhary and S. P. Raja, "Prediction of Flooding Due to HeavyRainfall in India Using Machine Learning Algorithms: Providing Advanced Warning," inIEEE Systems, Man, and Cybernetics Magazine, vol. 8, no. 4, pp. 26-33, Oct. 2022, doi:10.1109/MSMC.2022.3183806.

[2] S. Abraham, J. V R, S. Thomas and B. Jose, "Comparative Analysis of Various Machine Learning Techniques for Flood Prediction," 2022 International Conference on Innovative Trends in Information Technology (ICITIIT), 2022, pp. 1-5, doi: 10.1109/ICITIIT54346.2022.9744177.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 5, May 2023 |

[3] I. K. Nti et al., "Enhancing Flood Prediction using Ensemble and Deep Learning Techniques," 2021 22nd International Arab Conferenceon Information Technology (ACIT),2021, pp. 1-9, doi: 10.1109/ACIT53391.2021.9677084.
[4] T. Sharma, A. Pal, A. Kaushik, A. Yadav and A. Chitragupta, "A Survey on Flood Prediction analysis based on ML Algorithm using Data Science Methodology," 2022 IEEE Delhi Section Conference (DELCON), 2022, pp. 1-8, doi: 10.1109/DELCON54057.2022.9753396.
[5] B. S. Prasad, S. Shreya, T. Sinha and S. Priya, "Ensemble Model-Based Prediction for River Management: A Case Study on River Kaveri and Coastal Karnataka," 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), 2021, pp. 1-6, doi: 10.1109/SMARTGENCON51891.2021.9645804.
[6] R. Thamilselvan, M. Prabhu, K. T. Selvi, V. Karthik and T. Bhuvaneshwaran, "An Extensive Review of Rainfall Prediction using Machine Learning and Deep Learning Techniques," 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2022, pp. 198-205, doi: 10.1109/ICAIS53314.2022.9742746
[7] M. Khalaf et al., "IoT-Enabled Flood Severity Prediction via Ensemble Machine Learning Models," in IEEE Access, vol. 8, pp. 70375-70386, 2020, doi:10.1109/ACCESS.2020.2986090
[8] Y. Zhu, J. Feng, L. Yan, T. Guo and X. Li, "Flood Prediction Using Rainfall-Flow Pattern in Data-Sparse Watersheds," in IEEE Access, vol. 8, pp. 39713-39724, 2020, doi: 10.1109/ACCESS.2020.2971264.
[9] I. K. Nti et al., "Enhancing Flood Prediction using Ensemble and Deep Learning Techniques," 2021 22nd International Arab Conference on Information Technology (ACIT), 2021, pp. 1-9, doi: 10.1109/ACIT53391.2021.9677084.
[10] N. Gauhar, S. Das and K. S. Moury, "Prediction of Flood in Bangladesh using k-Nearest Neighbors Algorithm," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021, pp. 357-361, doi: 10.1109/ICREST51555.2021.9331199..
[11] S. K. TK and M. Hanumanthappa, "Time series analysis for water inflow," International Journal of Engineering Research & Technology, vol. 6, p. 484, 2017
[12] R. K. Misra, "Rainfall prediction using machine learning approach: A case study for the state of odisha," Indian Journal of Natural Sciences, 2021.
[13] A. Katti, K. V. Ashish, A. Loke and K. Bade, "A Pluvial Flood Detection Model Using Machine Learning Techniques and Simulate The Flow of Water," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 1189-1195, doi: 10.1109/ICCES48766.2020.9137946.
[14] L. J. L. Canillo and A. A. Hernandez, "Flood Risk Visualization and Prediction Information System: Case of City Manila, Philippines," 2021 IEEE 17th International Colloquium on Signal Processing & Its Applications (CSPA), 2021, pp. 59-63, doi: 10.1109/CSPA52141.2021.9377276.
[15] M. M. A. Syeed, M. Farzana, I. Namir, I. Ishrar, M. H. Nushra and T. Rahman, "Flood Prediction Using Machine Learning Models," 2022 International Congress on HumanComputer Interaction, Optimization and Robotic Applications (HORA), 2022, pp. 1-6, doi: 10.1109/HORA55278.2022.9800023.
[16] S. V. Mahadevkar et al., "A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions," in IEEE Access, vol. 10, pp. 107293-107329, 2022, doi: 10.1109/ACCESS.2022.3209825.
[17] M. N and A. A U, "Cloud-Based Flood Prediction Using IoT Devices and Machine Learning Algorithms," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), 2021, pp. 754-762, doi: 10.1109/ICESC51422.2021.9532916.
[18] L. Dongming, L. Yan, Y. Chao, L. Chaoran, L. Huan and Z. Lijuan, "The application of decision tree C4.5 algorithm to soil quality grade forecasting model," 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI), 2016, pp. 552-555, doi: 10.1109/CCI.2016.7778985.
[19] X. Pan and J. Wu, "Bayesian Neural Network Ensemble Model Based on Partial Least Squares Regression and Its Application in Rainfall Forecasting," 2009 International Joint Conference on Computational Sciences and Optimization, 2009, pp. 49-52, doi: 10.1109/CSO.2009.311.
[20] S. Bande and V. V. Shete, "Smart flood disaster prediction system using IoT & neural networks," 2017 International Conference On Smart Technologies For Smart Nation(SmartTechCon), 2017, pp. 189-194, doi: 10.1109/SmartTechCon.2017.8358367

Impact Factor: 7.580

INTERNATIONAL JOURNAL
OF MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

+91 99405 72462 +91 63819 07438 ijmrsetm@gmail.com