Prediction Of Flash Flood Using Rainfall Data And Machine Learning

[1]P.Yogasree [2]N.Kesava Rao

[1]P.Yogasree, Department of CSE, Narayana Engineering College, Gudur

[2]N.Kesava Rao,Associate Professor, Department of CSE, Narayana Engineering College, Gudur

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 K-Nearest Neighbors.

1. 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, TamilNadu, 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.

Generally, the choice of features is crucial in image segmentation algorithms. Nath and Deb stated that one of the most promising features of digital images is color information. The other commonly used feature is texture or pattern information. Many researchers used these main features to identify flood events. For the detection of a flooding event, Lai and Chen employed threshold values to determine the potential foreground regions. Borges et al.introduced a probabilistic model for flood detection in videos. They combined the statistical characteristics of floods, such as color, texture, and saturation characteristics, using the Bayes classifier along with frame-to-frame changes to determine the flood presence.

2.LITERATURE SURVEY:

Numerous parts of southern India have recently encountered severe damage to lives and properties due to floods. Floods are one among the most destructive natural hazard and recovering to normal life takes ample time. During hazards, various technologies are in use for speeding up relief operations and to minimize the amount of damage, one such being the use of drones. Many algorithms are in need for automatic analysis of remote sensing and aerial images.

Nowadays, drones are being used for taking images from varied heights similar to aerial images, as they have cameras with exceptional features and effective sensors. This paper proposes a hybrid approach to classify whether a region in an aerial image is flood affected or not. A combination of Support Vector Machine (SVM) and k-means clustering proved capable of detecting flooded areas with good accuracy, classifying about 92% of flooded images correctly. Performance analysis is done by changing various kernel functions in SVM. The results show that there is a decrease in the prediction and training time when quadratic SVM is used.

3.METHODOLOGY:

3.1 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.

3.2 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.

3.3 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.

3.4 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.

3.5 Random Forest:

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction.

4. SYSTEM DESIGN:

4.1 Input Design:

In an information system, input is the raw data that is processed to produce output. During the input design, the developers must consider the input devices such as PC, MICR, OMR, etc.

4.2 Output Design:

The design of output is the most important task of any system. During output design, developers identify the type of outputs needed, and consider the necessary output controls and prototype report layouts.

5. MODULES:

5.1. User:

5.1.1 Upload:

User has ability to upload the dataset for the model building.

5.1.2 Model Selection:

User should selects the machine learning model for training.

5.1.3 prediction:

User needs to enter input in order to detect the desire output

5.1.4 View Results:

User has ability to view the results generated by the system.

5.2 System

5.2.1 Take the dataset:

System works with the dataset provided to it for model building.

5.2.2 Preprocessing:

In preprocessing step system works with to impute any disorders in the data set and extract the features.

5.2.3 Training:

In training phase system generates the model from the dataset by using python modules.

5.2.4 Generate Results:

System generates the detection results from the model whether the there is a chance of floods occurring or not.

6. UML DIAGRAM:

UML stands for Unified Modelling Language. UML is a standardized general-purpose modelling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object-oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modelling Language is a standard language for specifying, Visualization, Constructing and documenting the artefacts of software system, as well as for business modelling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modelling of large and complex systems.

The UML is a very important part of developing objects-oriented software and the software

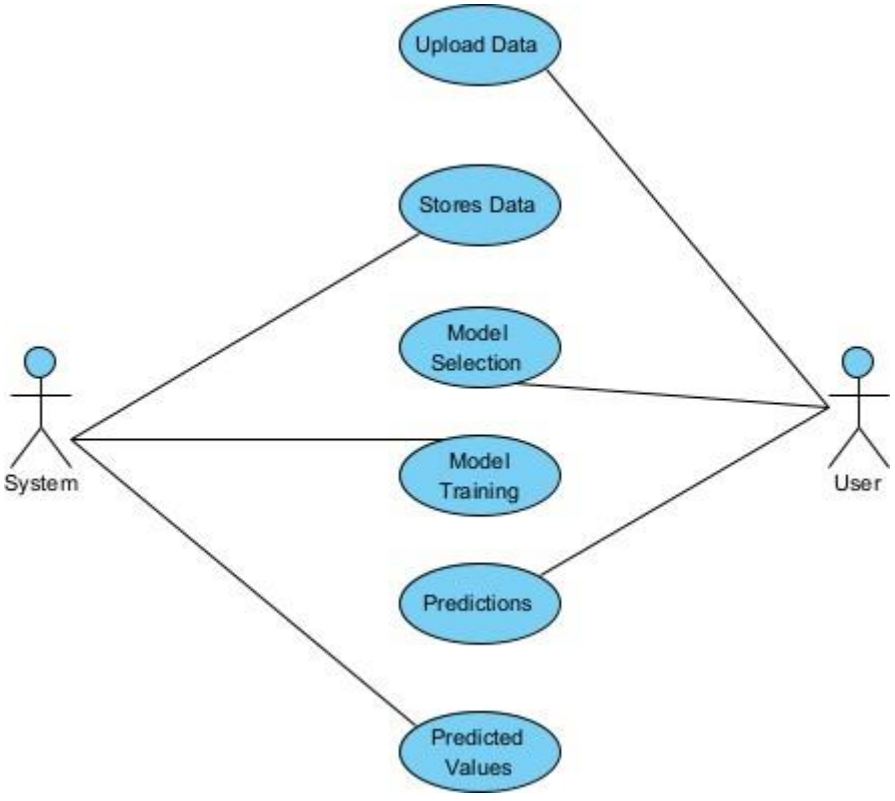
development process. The UML uses mostly graphical notations to express the design of software projects.

6.1 USE CASE DIAGRAM:

► A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis.

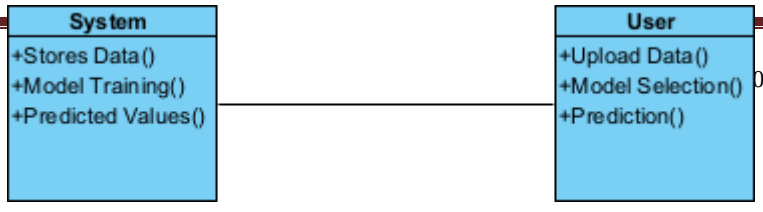
► Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases.

► The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



6.2 CLASS DIAGRAM:

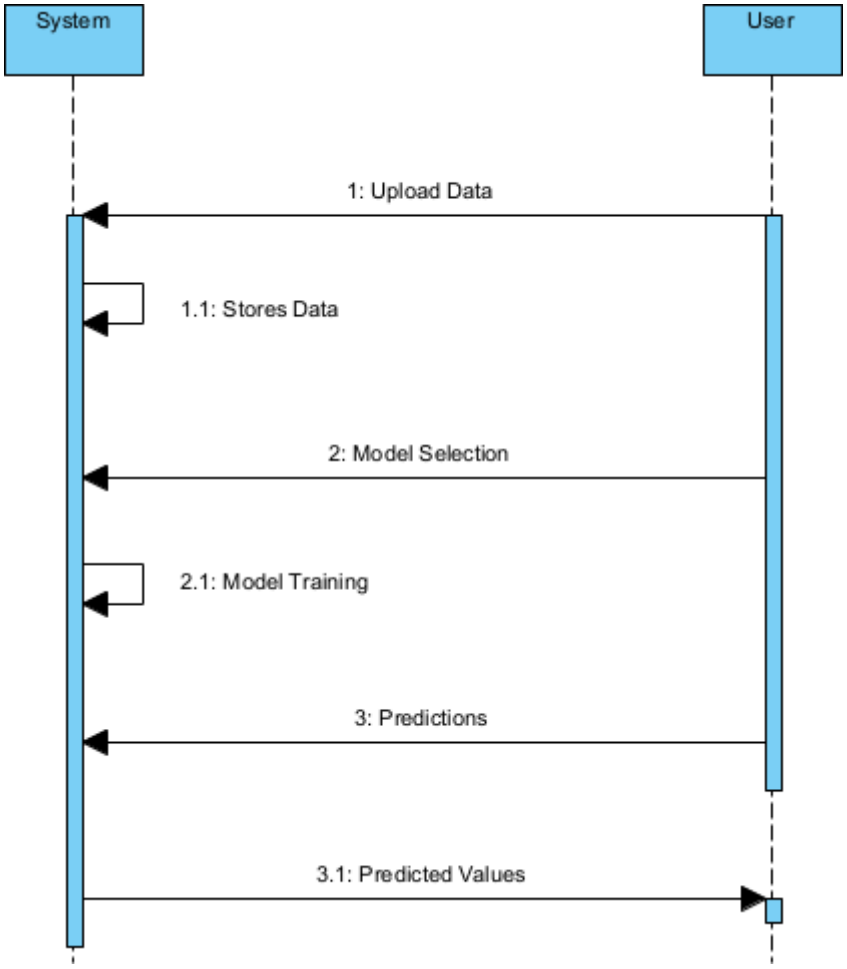
In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



6.3 SEQUENCE DIAGRAM:

- A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order.

- It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagram.

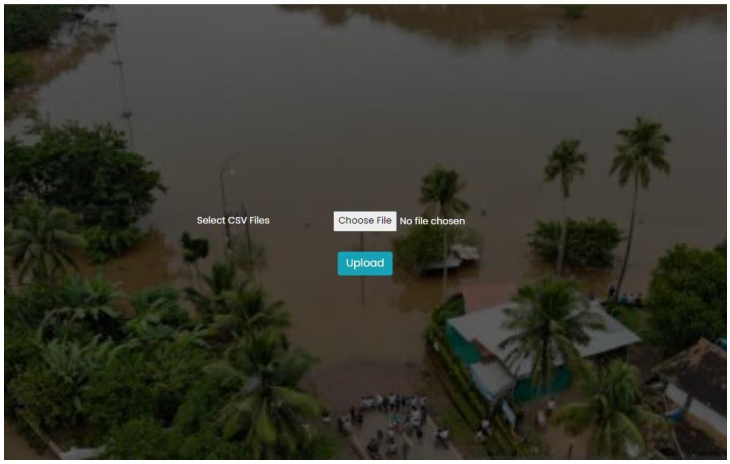


7. RESULTS AND ANALYSIS

7.1 Home:

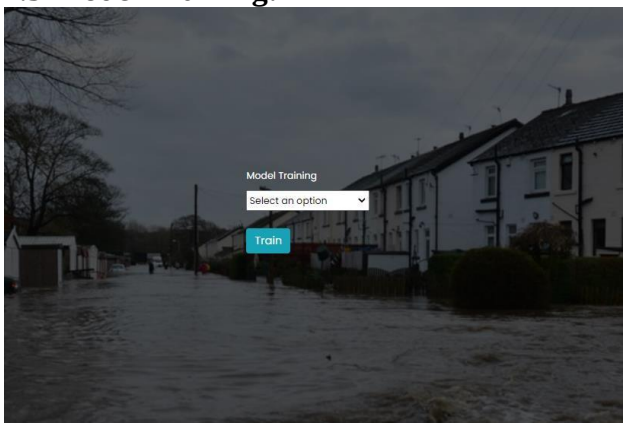
In our project, we are detecting whether the floods will occur or not.

7.2 Upload:

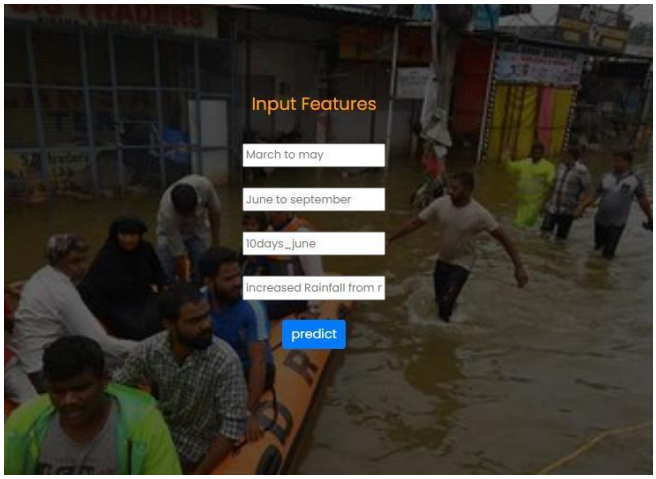


Here in this project we are uploading the dataset through which we are working.

7.3 Model Training:



7.4 Predictions:



8. CONCLUSION:

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.

9. REFERENCE

[1] J. Akshya and P. L. K. Priyadarsini, "A Hybrid Machine Learning Approach for Classifying Aerial Images of Flood-Hit Areas," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862138.

[2] A. B. Ranit and P. V. Durge, "Different Techniques of Flood Forecasting and Their Applications," 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), 2018, pp. 1-3, doi: 10.1109/RICE.2018.8509058.

[3] F. A. Ruslan, K. Haron, A. M. Samad and R. 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, pp. 179-184, doi: 10.1109/CSPA.2017.8064947.

[4] F. R. G. Cruz, M. G. Binag, M. R. G. Ga and F. A. A. Uy, "Flood Prediction Using Multi-Layer Artificial Neural Network in Monitoring System with Rain Gauge, Water Level, Soil Moisture Sensors," TENCON 2018 - 2018 IEEE Region 10 Conference, 2018, pp. 2499-2503, doi: 10.1109/TENCON.2018.8650387.

[5] G. Kaur and A. Bala, "An Efficient Automated Hybrid Algorithm to Predict Floods in Cloud Environment," 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), 2019, pp. 1-4, doi: 10.1109/CCECE.2019.8861897.

[6]. J. M. A. Opella and A. A. Hernandez, "Developing a Flood Risk Assessment Using Support Vector Machine and Convolutional Neural Network: A Conceptual Framework," 2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA), 2019, pp. 260-265, doi: 10.1109/CSPA.2019.8695980.

[7] Qianyu Zhang., Nttha Jindapetch.,Rakkrit Duangsoithong and Dujdow Buranapanichkit., "Investigation of Image Processing based Real-time Flood Monitoring," in Proc. of the 2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA 2018), Songkhla, Thailand, 2018.

[8]. Ryo Natsuaki and Akira Hirose, "L-Band Sar Interferometric Analysis for Flood Detection in Urban Area –A Case Study In 2015 Joso Flood, Japan," 2018.

[9]. Shijin LI., Kaikai MA., Zhou JIN., and Yuelong ZHU, "A new flood forecasting model based on SVM and boosting learning algorithms," 2016.

[10]. Sreehari E and Satyajee Srivastava, "Prediction of climate variable using Multiple Linear regression," in 2018 4th International Conference on Computing Communication and Automation (ICCCA, 2018.

[11]. Swapnil Bande and Virendra V. Shete, "Smart flood disaster prediction system using IOT & Neural Networks," in InternationalConference On Smart Technology for Smart Nation, 2017.

[12]. Tibin Mathew Thekkil and prabakaran N, "Real-time WSN Based Early Flood Detection and Control Monitoring System," in 2017 International Conference on Intelligent Computing,Instrumentation and Control Technologies (ICICICT),, 2017.

[13]. Vinothini A and Baghavathi priya A, "Survey of Machine Learning Methods for Big Data Applications," in International Conference on Computational Intelligence in Data Science, 2017.

[14]. Wahyu Sardjono and Widhilaga Gia Perdana, "The Application of Artificial Neural Network for Flood Systems Mitigation at Jakarta City," in International Conference on Information Management.