Flood Prediction and Tracking Trapped

Manomy K V[1], Meghna C S[1], Renuka Jayan[1], Raghi R Menon[2]
[1] UG Student, Department of Computer Science and Engineering, Adi Shankara Institute of Engineering and Technology, Kalady, Ernakulam, Pin: 683574
[2] Assistant Professor, Department of Computer Science and Engineering, Adi Shankara Institute of Engineering and Technology, Kalady, Ernakulam, Pin: 683574

***Abstract*--Floods are the most common natural disasters that are highly complex to model. They can cause destruction and devastation of natural life, agriculture, infrastructure and properties, every year. There are many researches on the advancement of flood prediction models. These models have contributed well to risk reduction, minimisation of the loss of human life and reduction of property damage associated with flood. Here, we use Machine Learning models for prediction. ML methods provide better performance and cost effective solutions. Some of the ML algorithms which were reported as effective for both short-term and long-term flood forecasts are Artificial Neural Networks(ANN), Support Vector Machine (SVM), and Support Vector Regression(SVR). Here, in this paper we introduce the most promising prediction method for both short-term and long-term floods.**

Keywords: Artificial Neural Network, Flood Prediction, Support Vector Machine, Support Vector Regression

I. INTRODUCTION

Among the natural disasters, floods are the most destructive, causing massive damage to human life, infrastructure, agriculture, and the socioeconomic system. Governments, therefore, are under pressure to develop reliable and accurate maps of flood risk areas and further plan for sustainable flood risk management focusing on prevention, protection, and preparedness. Flood prediction models are of significant importance for hazard assessment and extreme event management. Robust and accurate prediction contribute highly to water resource management strategies, policy suggestions and analysis, and further evacuation modeling. Thus, the importance of advanced systems for short-term and long-term prediction for flood and other hydrological events is strongly emphasized to alleviate damage. However, the prediction of flood lead time and occurrence location is fundamentally complex due to the dynamic nature of climate condition. Therefore, today's major flood prediction models are mainly data-specific and involve various simplified assumptions. Physical models showed great capabilities for predicting a diverse range of flooding scenarios, they often require various types of hydro-geomorphological monitoring datasets, requiring intensive computation, which prohibits short-term prediction .

A reason for the popularity of ML models is that they can numerically formulate the flood nonlinearity, solely based on historical data without requiring knowledge about the underlying physical processes. Data-driven prediction models using ML are promising tools as they are quicker to develop with minimal inputs. ML is a field of artificial intelligence (AI) used to induce regularities and patterns, providing easier implementation with low computation cost, as well as fast training, validation, testing, and evaluation, with high performance compared to physical models, and relatively less complexity. The continuous advancement of ML methods over the last two decades demonstrated their suitability for flood forecasting with an acceptable rate of outperforming conventional approaches .

In comparison to traditional statistical models, ML models were used for prediction with greater accuracy. Many ML algorithms, e.g., artificial neural networks (ANNs) , neuro-fuzzy, support vector machine (SVM), and support vector regression (SVR), were reported as effective for both short-term and long-term flood forecasts. In addition, it was shown that the performance of ML could be improved through hybridization with other ML methods, soft computing techniques, numerical simulations, and/or physical models. Such applications provided more robust and efficient models that can effectively learn complex flood systems in an adaptive manner.

Nonetheless, ML algorithms have important characteristics that need to be carefully taken into consideration. The first is that they are as good as their training, whereby the system learns the target task based on past data. If the data is scarce or does not cover varieties of the task, their learning falls short, and hence, they cannot perform well when they are put into work. The second aspect is the capability of each ML algorithm, which may vary across different types of tasks. This can also be called a "generalization problem", which indicates how well the trained system can predict cases it was not trained for, i.e., whether it can predict beyond the range of the training dataset.For example, some algorithms may perform well for short-term predictions, but not for long-term predictions. These characteristics of the algorithms need to be clarified with respect to the type and amount of available training data, and the type of prediction task, e.g., water level and streamflow.

Here, we should note that this paper surveys ML models used for predictions of floods on sites where rain gauges or intelligent sensing systems are used. Our goal is to survey prediction models with various lead times to floods.. From this perspective, spatial flood prediction was not involved in this study, as we did not study prediction models used to estimate/identify the location of floods.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

II. RESEARCH METHOD

A. Flood

A flood is an overflow of water that submerges land that is usually dry[1]. Floods are an area of study of the discipline hydrology and are of significant concern in agriculture, civil engineering and public health e.t.c.

Flooding may occur as an overflow of water from water bodies, resulting in some of that water escaping its usual boundaries,or it may occur due to an accumulation of rainwater on saturated ground in an areal flood[9].

Some floods develop slowly, while others develop even without visible signs of rain.The primary effects[2] of flooding include loss of life and damage to buildings and other structures. Economic hardship due to a temporary decline in tourism, rebuilding costs, or food shortages leading to price increases is a common after-effect of severe flooding. The impact on those affected may cause psychological damage to those affected, in particular where deaths, serious injuries and loss of property occur.

B. Rainfall

Rain is liquid water in the form of droplets that have condensed from atmospheric water vapor and then become heavy enough to fall under gravity. Rain is a major component of the water cycle and is responsible for depositing most of the freshwater on the Earth. It provides suitable conditions for many types of ecosystem.

The major cause of rain production is moisture moving along three-dimensional zones of temperature and moisture contrasts known as weather fronts.

C. Water Level

A water level is the height of water in a standing area measured from a specific point from the ground. Water is easily procured for use, and easily discarded after use. The ends are held vertical, and the rest of the tubing lies on the ground or floor. The rise of water level can occur due to many reasons including improper canal system, or heavy rainfall e.t.c. The increase of water level above a specific point can cause loss of topsoil, fertility and even leads to drought or flood.

D. Prediction

A prediction or forecast, is a statement about a future event. A prediction [3] is often, but not always, based upon experience or knowledge.Although future events are necessarily uncertain, so guaranteed accurate information about the future is in many cases impossible, prediction can be useful to assist in making plans about possible developments[4].

The prediction algorithm used is naive bayes algorithm.Naive Bayes Algorithm is one of the popular classification machine learning algorithms that helps to classify the data based upon the conditional probability values computation. It implements the Bayes theorem for the computation and uses class levels represented as feature values or vectors of predictors for classification. Naive Bayes Algorithm is a fast algorithm for classification problems. This algorithm is a good fit for real-time prediction, multi-class prediction, recommendation system, text classification, and sentiment analysis use cases. Naive Bayes Algorithm can be built using Gaussian, Multinomial and Bernoulli distribution. This algorithm is scalable and easy to implement for the large data set[5].

It helps to calculate the posterior probability P(c|x) using the prior probability of class P(c), the prior probability of predictor P(x) and the probability of predictor given class, also called as likelihood P(x|c).

The formula or equation to calculate posterior probability is:

$$P(c|x) = (P(x|c) * P(c)) / P(x)$$

EQ 1: Naive bayes equation

E. Machine learning

Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a subset of artificial intelligence. Machine learning[6] algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so.

Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop conventional algorithms to perform the needed tasks.

Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning.In its application across business problems, machine learning is also referred to as predictive analytics

F. SVM

In machine learning, support-vector machines[7] are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. A SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible[8]. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

G. WIFI

Wi-Fi is the name of a wireless networking technology that uses radio waves to provide wireless high-speed Internet and network connections. A common misconception is that the term Wi-Fi is short for "wireless fidelity," however this is not the case. Wi-Fi is simply a trademarked phrase that means IEEE 802.11x.Wi-Fi networks have no physical wired connection between sender and receiver by using radio frequency (RF) technology -- a frequency within the electromagnetic spectrum associated with radio wave propagation.[9] When an RF current is supplied to an antenna, an electromagnetic field is created that then is able to propagate through space.The cornerstone of any wireless network is an access point (AP). The primary job of an access point is to broadcast a wireless signal that computers can detect and "tune" into. In order to connect to an access point and join a wireless network, computers and devices must be equipped with wireless network adapters.

H. GPS

The Global Positioning System, is a satellite-based radionavigation system owned by the United States government and operated by the United States Space Force. It is one of the global navigation satellite systems (GNSS) that provides geolocation and time information to a GPS receiver anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites. Obstacles such as mountains and buildings block the relatively weak GPS signals.

The GPS does not require the user to transmit any data, and it operates independently of any telephonic or internet reception, though these technologies can enhance the usefulness of the GPS positioning information. The GPS[10] provides critical positioning capabilities to military, civil, and commercial users around the world. The United States government created the system, maintains it, and makes it freely accessible to anyone with a GPS receiver.

III. SYSTEM DESIGN AND OVERVIEW

System Overview

The system is designed to predict future floods based on input data and previous water level data.Our system helps users to get flood related announcements including a warning message which gives an alert to the public **.** The list of trapped with their location is maintained as a table which includes details of requested person, trapped person, location, track and status.A login button is used to obtain the details of family.Details of family includes the number of members in each family, house number, house name, name of each members, unique ID, longitude, panchayat e.t.c

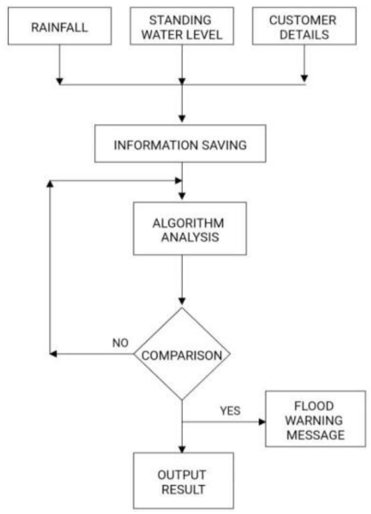

FIG 1: Basic working of the system

- **Prediction:**A dataset with the amount of rainfall and if a flood had occurred in a particular area/state/city, in the previous years, will be used. The dataset will have the rainfall data for a duration of 3 months approx.Using this dataset, we take average rainfall for every 10 days and plot it on a graph to visualize it. We take this average data of rainfall, as input to our machine learning model and if it causes a flood or not as the output labels. We train our model and save it.(depending on some threshold value of average rainfall in the dataset)Given the input data, for consecutive 10 days, we give this data as an input, and let the model predict, whether there is a possibility of flooding or not, by setting some threshold in the training data. Our basic approach for this problem is binary classification, using basic machine learning algorithms(linear regression or logistic regression).

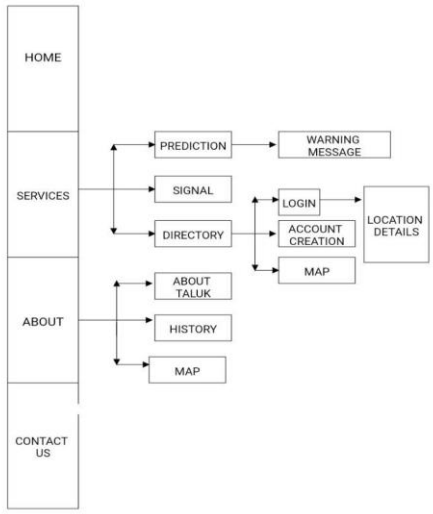

FIG 2: Flowchart represents the basic structure

- **Signal:** Flood related announcements include a warning message .System should retrieve details of trapped people and other notifications like hospital,relief centres A table of details of the family includes the number of members in each family, house number, house name, latitude, taluk / panchayat, name of each member, longitude, phone no., unique ID and type of message[11].

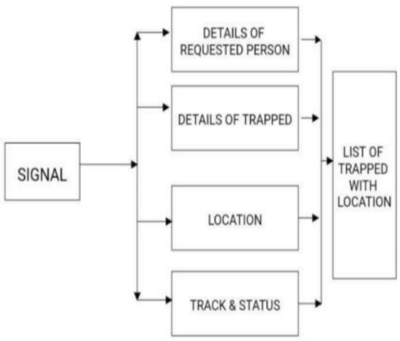

FIG 3: The flow diagram for counting the trapped

IV. FUTURE STUDY

The water level, location, month, climate is the parameter used in this method. So, for predicting more accurate results, more parameters can be added in the model inorder to produce more accurate and realistic predictions.
For future works important details such as route map to trapped people, nearby camps, requests of need, details of availability e.t.c can be added to make the method more easy and effective

V. RESULT

The expected results are accurate flood prediction and tracking of trapped. When inputs such as location, rainfall, climate are given it shows the chance of occurence of flood. And by client-server requests it gives an accurate number of trapped people.

VI. CONCLUSION

Flood prediction helps to inform the public about the possibility of an incoming flood disaster. Accurate prediction of flood is important to avoid or reduce the impact caused by the disaster. This paper proposes a machine learning algorithm to calculate the occurence of flood. In order to effectively obtain the required output, two inputs are used - standing water level of the selected area and the amount of rainfall. .A platform is used which helps the victims to send requests, which helps the authority to find the number of trapped.

VIII. REFERENCE

[1] K. Smith, Environmental hazards: assessing risk and reducing disaster. Routledge, 2003.
[2] Z. Sen, Flood Modeling, Prediction and Mitigation, 1st ed. Springer International Publishing, 2018.
[3] A. Y. Sun and B. R. Scanlon, ‘‘How can big data and machine learning benefit the environment and water management: A survey of methods, applications, and future directions,’’ Environ. Res. Lett., vol. 14, no. 7, p. 28, 2019.
[4] C. Cortes and V. Vapnik, ‘‘Support-vector networks,’’ Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.
[5] K. Kravari and N. Bassiliades, “A survey of agent platforms,” Journal of Artificial Societies and Social Simulation, vol. 18, no. 1, p. 11, 2015
[6] M. Westergaard and H. Verbeek, “Cpn tools.” [Online]. Available: http://cpntools.org/
[7] Ö. Kişi, ‘‘Streamflow forecasting using different artificial neural network algorithms,’’ J. Hydrol. Eng., vol. 12, no. 5, pp. 532–539, Sep. 2007
[8] Glossary of Meteorology (June 2000) Flood Archived 2007-08-24 at the Wayback Machine, Retrieved on 2009-01-09
[9] Brown, Richard; Chanson, Hubert; McIntosh, Dave; Madhani, Jay (2011). Turbulent Velocity and Suspended Sediment Concentration Measurements in an Urban Environment of the Brisbane River Flood Plain at Gardens Point on 12–13 January 2011. Hydraulic Model Report No. CH83/11. p. 120. ISBN 978-1-74272-027-2.
[10] Siegel, Eric (2013). Predictive Analysis: The Power to Predict Who Will Click, Buy, Lie, or Die. Hoboken, NJ: John Wiley & Sons. ISBN 978-1-118-35685-2.
[11] Naive Bayes classifier – Wikipedia en.wikipedia.org/wiki/Naive_Bayes_classifier