Rainfall Prediction and Flood Warning using Machine Learning

Prof. A. D. Wankhade[1], Bhagyashri Jaiswal[2], Divya Gupta[3], Mahima Gadodiya[4], Sanket Raut[5]

[1] Assistant Professor, Dept. Of IT, Government College of Engineering, Amravati
[2],[3],[4],[5] B. Tech students, Dept. Of IT, Government College of Engineering, Amravati – Maharashtra

ABSTRACT
Rainfall prediction has become one of the challenging and major tasks which has serious impact on society. If the predictions are made accurately and in time it can prevent huge loss of lives and financial loss. The aim of this paper is to help civilization by providing on-the-spot assistance at the time of natural disaster. This chapter deals with the comparative study of prediction of rainfall using machine learning and neural networks. It calculates the value of predicted flood in upcoming year and also contain feature to analyze the previous year flood via statistical analysis and mathematical parameters. Rainfall Prediction and Flood Warning Software helps society by providing immediate assistance in time of natural disaster. It gives necessary alert warnings, required guidelines for evacuation, emergency help, assessing damage, continuing assistance and the immediate restoration so human lives will safeguard themselves and damage will minimize.
Keywords —Neural Network, Recurrent neural network (RNN), Machine learning, Rainfall prediction, Flood Warning, Data Analysis, Linear Discriminant Analysis.

I.   INTRODUCTION

In today's situation, rainfall is considered to be one of the sole responsible factors for most of the very important and significant things across the world. In India, agriculture is considered to be one of the important factors for deciding the economy of the country and agriculture is solely dependent on rainfall. Apart From that in the coastal areas across the world, getting to know the amount of rainfall is very much necessary. In some of the areas where there is shortage of water in such places there is need to establish projects like rain water harvester and prior prediction of the rainfall should be done. This project performs the through study of machine learning approaches and neural network approaches then accordingly portrays the efficient approach for rainfall prediction. First requirement is to Ingest data related to flood and rainfall. Then analyses of data is done by using data preparation–analysis, exploration, cleaning, feature extraction, etc. Then Machine Learning model training, fine tuning and evaluation of model is done. At the end will also provide user interface for real time graphs and data analysis.

To work on warning system for floods will also provide a user interface to the common public to check the level of water flow in rivers for future and will provide a mechanism for notifications if there is any possibility of flood due to any river in nearby future. This study predict future prior to twelve months. Along with this it also provides the historical trends of river flows and rainfall in the subdivision. So, by having this information in advance one can prepare and alert the citizens so that huge loss can be prevented.

II. RELATED WORKS

Machine learning approach deals with predicting rainfall using machine learning approach. There are two types of errors to finds the accuracy of the machine learning approach i.e. RE and RMSE. In these four major trends of machine learning are being used. The first one is called hybridization, which means multiple machine learning approaches are being used together and accordingly prediction is being done. The second one deals with improving the quality of dataset which is being used. The technique to improve the quality of dataset used is called decomposition technique. The third one is to use of ensemble of method for increasing the ability of the algorithm which is being used and the final ne is using add on optimizer for increasing the accuracy of the algorithm. One of the major advantages of this system is its ability to increase the quality of algorithm and dataset. The more the data's used for prediction will be quality and efficient, the more will be accuracy of the prediction. Same goes with the accuracy of the algorithm. One of the biggest disadvantages of this system is its error finding technique. There are many more errors which can have negative effect of algorithm's accuracy such as MSME (Mean squared error), MAE (Mean absolute error) etc. which are not being calculated in this system. Only R2 (R-Squared) and RMSE (Root mean square error) are being evaluated to finding the accuracy of the algorithm. So, the accuracy of any algorithm can more be tested if the number of errors affecting it can be increased. By testing few levels of accuracies, it is observed that neural network approach is better and more effective than all the machine learning approaches because it is capable of finding the entire

nonlinear pattern present inside any system [4].

Data mining approach helps to find the hidden pattern, which will help to predict the rainfall correctly. This approach takes all the parameters, which affect the rainfall such as climate, wind speed etc. and predict the rainfall in future. Customized, integrated and improved data mining technique is used to predict rainfall. Many climate variables are being taken to predict rainfall. Data is such as polarity, climate, wind, maximum temperature, minimum temperature etc. are being taken. It says that using maximum parameters does not mean that the prediction will be more accurate. Both supervised and unsupervised techniques are used for prediction. According to this paper, the prediction was tried in some of the countries like India, Australia, Columbia, Indonesia, Malaysia etc. The key factors, which affected the result, are past weather data, which are used to train the algorithm; the climate attributes which was taken as predictors and the location in which prediction is to be done. Some of the latest research trend on this domain are finding the correlation between the weather features whereas some of the researchers giving more Importance to the data which is used to train the algorithm. There are many hidden patterns presents, which can affect the rainfall prediction, and data mining technique have the ability to extract all the hidden patterns. This technique needs to be integrated and optimized in a way that all the prediction should be more error free. So, in the future work this data mining technique should be enhanced, optimized and integrated in a way so that all the present problems related to finding the hidden patterns should be resolved and along with that a proper correlation should be find out between the weather factors [5]. Deep learning method deals in three stages. First stage deals with finding an algorithm which suits the best. The algorithm is selected on the basis of the dataset present i.e., which algorithm works well with the given dataset. The second stage deals with finding the model which best suits with the algorithm. In the final stage, some Metrics items are being adopted for quantitative precipitation forecasting, false alarm ratio and threat score. The calculation done in first two stages, are done on original predicted time series, so it also includes negative values. Advantage of this method is that all the nonlinear pattern is being identified and the correlation coefficient are being calculated in a proper way but because it uses original predicted data for calculation, so some of the times the calculation takes some time and doesn't give accurate result [6]. A new extension of LSTM (Long short-term memory) called ConvLSTM is being proposed. LSTM is a type of RNN (Recurrent neural network). One of the main advantages of RNN is, it can predict accurately if the data is stored in short term memory but on the other hand if the data is stored in long term memory, it near LSTM. As the gap length increases, the effectiveness of RNN gets reduced. The major advantage of this research work is its ability to store and retrieve the data's which are stored for longer period of time but along with that it also shows the inability of neural network to do prediction for the same. RNN requires LSTM or ConvLSTM to retrieve the data which is being stored for longer period of time. This process deals with getting the data from the data source and to how to use it effectively [7]

III. METHODOLOGY

A. Dataset

A dataset in machine learning is a collection of data pieces that can be treated by a computer as a single unit for analytic and prediction purposes. This means that the data collected should be made uniform and understandable for a machine that sees data in different way from humans. In Machine Learning projects, we need a training data set. It is the data set that is used to train the model for performing various actions.

Every machine learning model requires a dataset to train it. This project use dataset provided by the Kaggle, it's a community of data scientist and machine learning practitioners. Dataset size is one of the parameters on which accuracy of the ML model depends. Make sure the size of dataset should be as big as possible. Two datasets are required to perform this project. One is state wise rainfall data and second, yearly rainfall data.

A. Machine learning model

For predicting the value of flood, the following machine learning model is used: Linear discriminant analysis (lda). Predicting Future: linear discriminant analysis and prophet tool (For time series). Analysis of past and historical trends can be view with the help of linear discriminant analysis model.

A. Linear discriminant analysis

Linear discriminant analysis (LDA) or discriminant function analysis is a generalize form of Fisher's linear discriminant, a method used in statistics, pattern

recognition and machine learning to find a linear combination of features that characterizes or separates two or more than two classes of objects or events. The resulting combination may be used as a linear classifier, or, commonly, for dimensionality reduction before and after classification. Linear discriminant analysis primarily used to minimize the number of features to more manageable number before classification. LDA is also quite similar to principal component analysis. It works when the measurements made on independent variables for each observation are continuous quantities. It is type of linear combination, a mathematical process using various data items which is set to separately analyze

multiple classes of objects. While dealing with above algorithms. From now on, the main algorithm for testing on daily basis dataset instead of accumulated dataset categorically independent variables, an equivalent technique is discriminant correspondence analysis.
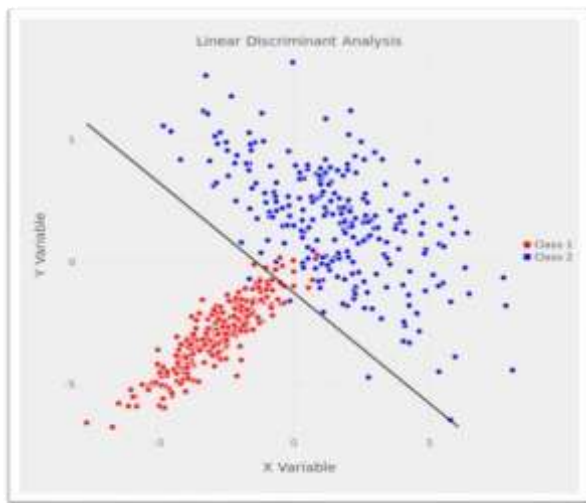

Figure. 1 Red Dot: No flood, Blue Dot: Flood

B. FBPROPHET (FACEBOOK FBPROPHET)

It is basically a library to build forecasting models for time series data, but instead of using the traditional way of building the model such as using ARIMA, etc., it is fitting additive regression models or known as 'curve fitting'. FbProphet is a powerful time series analysis package released by Core Data Science Team at Facebook. It is simple and easy to go package for performing time series analytics and forecasting at scale. They have implemented the core part of the procedure in Stan's probabilistic programming language. Because of this, "Stan performs the MAP optimization for parameters extremely quickly (<1 second), gives us the option to estimate parameter uncertainty using the Hamiltonian Monte Carlo algorithm, and allows us to re-use the fitting procedure across multiple interface languages.", according to the authors.

- It is designed to handles typical data challenges likethe followings by default.

- A reasonable number of missing observations or large outliers.

- Historical trend changes, for instance due to productlaunches or logging changes.

Trends that are non-linear growth curves, where a trend hits a natural limit or saturates "We have found Prophet's default settings to produce forecasts that are often accurate as those produced by skilled forecasters, with much less effort."

"With Prophet, you are not stuck with the results of a completely automatic procedure if the forecast is not satisfactory—an analyst with no training in time series methods can improve or tweak forecasts using a variety of easily-interpretable parameters."
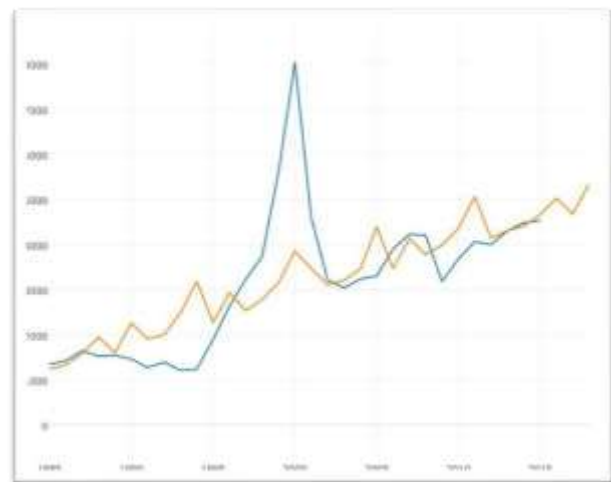

Figure. 2 Orange Line: Actual, Blue Line: Predicted.

IV. CONCLUSION

Indian economy impacts vulnerably due to heavy rainfall or sudden variation in rainfall due to climate change and various other reasons. Rainfall causes heavy floods. Study finds that better achievement in Human Development Index has significantly minimized flood related mortalities. In India, flood causes heavy damages. So far on average, 7.2 million hectors of agriculture and non-agriculture land has gotten affected, crop damage around 1119 crores rupees, many human lives were lost. So, Rainfall should be considered the primary and major concern for all of us. The current approach for rainfall prediction fails because it unable to achieve precise prediction in most of the complex cases. To achieve an effective way to predict rainfall, this study based on machine learning approach and Recurrent neural networks approach. Which calculates the value of predicted flood in upcoming year and also contain feature to analyze the previous year flood via statistical analysis and mathematical parameters.

Dataset is being used that includes previous rainfall amount in various regions in and around the country. Dataset is divided into two parts one is train data and test data. Train data is for training the algorithm and test data is for doing the prediction. Both of these processes were compared based on their accuracy and along with that, error types such as MSE, MAE, R-SQUARED and RSME were considered. The one with more accurate was considered and prediction was performed with that approach itself. The rainfall was predicted from that data used for testing as part of the data being used totrain the algorithm.

The future enhancement of this project can be an approach towards about how to reduce the percentage of errors present and to increase efficiency and Accuracy of model. In future, by building disaster system with extra features like rescue operation and donation funding, immediate help assistance, to perform search and rescue operations etc. so the model can be made more efficient. Along with that one of the major enhancements will be to decrease the ratio for train data to test data, so that it will assist in improving the level of prediction within the available time and complexity.

V. ACKKNOWLEDGEMENT

VI. REFERENCE

[1] Kar, K., Thakur, N., & Sanghvi, P. (2019). Prediction of Rainfall Using Fuzzy Dataset. International Journal of Computer Science and Mobile Computing, 8(4), 182–186.

[2] Etuk, E. H., & Mohamed, T. M. (2014). Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by Sarima Methods. International Journal of Scientific Research in Knowledge, July, 320–327.

[3] Wahyuni, E. G.Fauzan, L. M.F.Abriyani, F.Muchlis, N. F., & Ulfa, M. (2018). Rainfall prediction with backpropagation method. Journal of Physics: Conference Series, 983(1).

[4] Aakash Parmar, Kinjal Mistree, M. S. (2017). Machine Learning Techniques for rainfall prediction: A Review. International Conference on Innovations in Information Embedded and Communication Systems (ICIIECS).

[5] Aftab, S., Ahmad, M., Hameed, N., Bashir, M. S., Ali, I., & Nawaz, Z. (2018). Rainfall prediction using data mining techniques: A systematic literature review. International Journal of Advanced Computer Science and Applications, 9(5), 143–150.

[6] Meng – Hua Yen , Ding – Wei Liu , Yi – Chia Hsin, C. – E. L. and C. – C. C. (2019). Application of the deep learning for the prediction of rainfall in Southern Taiwan. Scientific Reports, 9(1), 1–9.

[7] Shi, X., Chen, Z., & Wang, H. (2015). Convolutional LSTM Network. Nips, 2–3.

[8] Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. A Sherstinsky - Physica D: Nonlinear Phenomena, 2020 – Elsevier.

[9] Real-time flood monitoring and warning system. Jirapon Sunkpho , Chaiwat Oottamakorn , June 2011.

[10] https://machinelearningmastery.com/linear-discriminant-analysis-for-machine-learning.