International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 03 | Mar 2021 www.irjet.net p-ISSN: 2395-0072

A Review on Rainfall Prediction using Machine Learning and Neural Network

Vamshi K[1], Sachin Kumar S[2], Muralidhar B R[3], Manjunath N[4], Mrs. Savitha P[5]

[1]*B.E. Student, Department of CSE, Sir M Visvesvaraya Institute of Technology, Bengaluru, India*
[2]*B.E. Student, Department of CSE, Sir M Visvesvaraya Institute of Technology, Bengaluru, India*
[3]*B.E. Student, Department of CSE, Sir M Visvesvaraya Institute of Technology, Bengaluru, India*
[4]*B.E. Student, Department of CSE, Sir M Visvesvaraya Institute of Technology, Bengaluru, India*
[5]*Assistant Professor, Department of CSE, Sir M Visvesvaraya Institute of Technology, Bengaluru, India*

Abstract - *In India, agriculture is the most important factor for survival of human being. For agriculture, the most important thing is water. i.e., rainfall. Nowadays rainfall prediction is a major problem. Predicting the amount of rainfall gives alertness to farmers by knowing early so that they can protect their crops and properties from rain. There are more techniques to predict the rainfall. The ML algorithms are best suited for prediction of rainfall. Here are some of the major ML algorithms used rapidly which are Auto regressive integrated moving average Model (ARIMA), Artificial neural network (ANN), Support Vector Machine, Logistic regression, and Self organizing map. And there are two models commonly used to predict periodic rainfall such as Linear and Non-linear models. ARIMA Model is the first used model. Although using ANN (Artificial neural network) the prediction of rain can easily completed by using Cascade NN, Layer recurrent network, or Back propagation NN. Artificial NN is similar as Biological neural networks.*

***Key Words*: Machine Learning, LASSO regression, ANN(Artificial Neural Network) approach, ARIMA MODEL(Auto Regressive Integrated Moving Average), Naive Bayes.**

1. INTRODUCTION

Nowadays, rainfall is considered to be one of the most liable factor for most of the significant things in the world. In India, agriculture is one of the most important factor in deciding the economy of the country and agriculture is totally dependent on the rainfall. Apart from agriculture, rainfall is also more important in coastal areas around the world by getting to know the rainfall is very much necessary to protect their life's from the floods and heavy rainfall. In some of the areas which are having drought, to establish an rainfall harvester, proper prediction of rainfall is necessary. In this project we are dealing with predicting of rainfall using Machine learning and Neural networks. In this, we are executing an comparative study of machine learning approaches and neural network approaches then accordingly selects the perfect approach for prediction of rainfall. First the preprocess is performed, i.e. representing the input dataset in graph form. Such as histogram, bar graph etc. In ML techniques, LASSO (Least absolute shrinkage and selection operator) regression is used and ANN (Artificial Neural Network) approach is used for neural network. After the calculation, accuracy and errors of the LASSO and ANN is been compared and accordingly the conclusion is made. In this, the prediction of rainfall is made with the approach which has highest accuracy in the outcomes. The prediction of rainfall is done with the dataset which has the information related to rainfall.

1.1 MOTIVATION

Predicting rainfall is an application of science and technology for predicting the amount of rain over an area. The most important thing is to accurately determine the rainfall for active use of rainfall for water resources, crops, pre-planning of water resources and for agricultural purposes. In earlier rainfall information benefits the farmers for better managing their crops and properties from heavy rainfall. The farmers better manage to increase the economic growth of the country by efficient rainfall information. Prediction of precipitation is necessary to save the life of people's and properties from flooding. Prediction of rainfall helps people in coastal areas by preventing the floods.

2. RELATED WORK

Flood prediction using machine learning models: Literature review. Water (Switzerland),

Floods are among the most dangerous natural disasters, which are complex to model. The advanced flood prediction models research contributed to policy suggestion, risk reduction, minimization of the loss of human life in floods, and reduction in property damage with respect to floods due to heavy rainfall. To imitate the complex mathematical expressions of physical processes of floods, for last two decades, machine learning methods are best in advancement of prediction systems which are provided with better performance and very reasonable solutions. due to more benefits and potential of ML, the popularity of ML is highly increased among hydrologists. Researchers introducing novel ML methods and hybridizing of the existing ones aim at discovering more effective and accurate prediction models. The contribution of this paper is to demonstrate the state of the art of ML models in flood prediction and to give awareness into the most suitable models. In this paper, the literature where ML models were set an standard through a qualitative analysis of robustness, accuracy, effectiveness, and speed are particularly investigated to provide an extensive overview on the various ML algorithms used in the field. The performance comparison of ML models gives an in-detail understanding of various ML techniques within a framework of a evaluation and discussion. As a result in this paper, it introduces the most effective and accurate methods for short and long term floods. Further, the flood prediction methods are investigated for quality improvement. Among them, data decomposition, algorithm ensemble, hybridization, and optimization model are reported as effective strategies for improvement in ML methods. This survey can be used for hydrologists and climate scientists for choosing a better ML method for their projects according to their needs for the task. The methods used are ANN-Artificial Neural Network, SVM-Support vector Machine, MLP- Multi layered Perception, DT-Decision Tree, ANFIS- Adaptive Neuro-Fuzzy Inference System, WNN- Wavelet Neural Network, and EPS- Ensemble Prediction Systems. Comparing from the single algorithm hybrid algorithm improves the accuracy of flood prediction. It utilizes large time and huge resources.

Analysis on the weather forecasting and techniques.A brief summary of Daily weather forecasting is employed for multiple reasons in many areas like agriculture, transportations, etc. In forecast reports, the accuracy of weather is shown and it is very necessary. In this paper, the review is conducted to analyze a higher approach for prediction that is used for various sorts of prediction. Among which ARIMA model with time series performs prediction with no errors and better accuracy in outputs. For seasonal statistic prediction, Box and Jenkins had projected a comparatively successful variation of ARIMA model, viz. the seasonal ARIMA (SARIMA) .The popularity for the ARIMA model is increasing especially as a result in its flexibility to represent more sorts of statistic with simplicity still because of there lated Bx-Jenkins methodology for some optimal model an ARIMA model is build that is eco-friendly in nature which gives more efficiency and reliability. It studied itself more in training data and builds a lot of related techniques that are very helpful for weather forecasting. This paper gone through many techniques and focuses mainly on ARIMA MODEL technique for daily meteorology. To forecast the daily weather this technique uses many different parameters in terms of rainfall, humidity, temperature, cloud condition, and weather of the day. The main goal of this paper is to compare the present meteorology model and to select the precise model to support their predictive ability. ARIMA model is used. It is supporting for forecasting the weather for future based on the data. Accuracy is less than 60%.

Open Access Study of Various Rainfall Estimation & Prediction Techniques Using Data Mining.It is important to guess the accurate rainfall for the effective use of water resources and optimal planning of water resources and availability. For this purpose, the various techniques and models are developed to estimate rainfall in various researches using data mining techniques. The accurate and exact estimation of rainfall prediction and estimation of precipitation is possible through there are many available techniques. The usage of data mining techniques to predict rainfall and its solutions may prove significance in the prediction of efficient rainfall that will help the farmers to help with their decisions on their crops and it also improves the growth of the agriculture sector. This paper studies various techniques of rainfall prediction and estimation and their results with the actual rainfall value. In this paper, PCA, K-Means methods are used. It is simple and consumes less time and less resource. Accuracy is less than 65%.

Machine Learning Techniques for rainfall prediction: A Review.Predicting rainfall is more important thing as heavy rainfall lead to many disasters. the predicting of rainfall helps people to take preventive measures on their crops and properties and moreover the prediction should be more accurate and efficient. There are two types of predictions, short term rainfall prediction and long-term rainfall prediction. In Predicting mostly short-term prediction can gives us the accurate result. The main challenge to us is to build a model for long term rainfall prediction. Heavy precipitation prediction could be a major drawback for weather forecasting department because it is closely associated with the economy and lifetime of human beings. natural disasters like flood and drought are a cause that square measure encountered by individuals across the world every year. Predicting of rainfall with more accuracy has a very importance for countries like India because whose main economy is totally dependent on agriculture is India. The dynamic nature of atmosphere, applied mathematics techniques are failed to provide sensible accurate result for precipitation statement. The prediction of precipitation using machine learning techniques may use regression technique. The Intention of this project is to offer non-expert's easy access to the techniques, and also approaches utilized in the sector of precipitation prediction and provide a comparative study among the various machine learning techniques. In this method's used are SVM, Logistic Regression. It gives the high accuracy up to 89%. We can use only for the prediction we can't use for forecasting.

Rainfall prediction using data mining techniques: A systematic literature review.Rainfall prediction is one of the challenging tasks in weather forecasting. Accurate and timely rainfall prediction can be very helpful to take effective security measures in advance regarding: ongoing construction projects, transportation activities, agricultural tasks, flight operations and flood situation, etc. Data mining techniques can effectively predict the rainfall by extracting the hidden patterns among available features of past weather data. This research contributes by providing a critical analysis and review of latest data mining techniques, used for rainfall prediction. Published papers from year 2013 to 2017 from renowned online search libraries are considered for this research. This review will serve the researchers to analyze the latest work on rainfall prediction with the focus on data mining techniques and also will provide a baseline for future directions and comparisons. We have find the difficulties in rainfall prediction. There is no real time implementation.

Rainfall prediction using Machine Learning TechniquesPrediction of rainfall is very important aspect in the country that leads to economy of the country too and can help the people by preventing from some serious natural disasters. some areas in India are economically dependent on the rainfall as agriculture is dependent on the rainfall and agriculture is primary occupation of many states. This prediction helps to identify the crops patterns and proper management of water resources for crops. For this, linear and non- linear models are very commonly used for seasonal rainfall prediction. Some of few algorithms used for rainfall prediction are CART, Genetic Algorithms and SVM, these are computer aided rule-based algorithms. In this paper, we have performed qualitative and accurate analysis using some classification algorithms like Support vector machines(SVM), Artificial Neural Networks, Logistic regression. Dataset used for this classification application is taken from hydrological department of Rajasthan. Overall, we analyze that algorithm which are efficient are to be used in order to qualitatively predict rainfall. Support vector machines (SVM), Artificial Neural Networks, Logistic regression methods are used. ANN was able to yield an accuracy of 87%. It requires the huge volume of data to train the model and it consumes more time.

A review of unsupervised feature learning and deep learning for time-series modeling.In this paper, a review of the recent improvements in deep learning and unsupervised feature learning for problems like time series. These techniques have shown some promise for modeling of static data, such as computer vision, applying them to time-series data is the process of gaining increasing attention. In this paper, it overviews some particular challenges which are present in the time-series data and a review is provided which consists of works that have either applied time-series data to unsupervised feature learning algorithms or alternatively have committed to changes of feature learning algorithms to take through into account the challenges present in the time-series data. Recurrent Neural Network (RNN) method is used. Accuracy is high for time series modeling. There is a need to focus less on the pre- processing pipeline

for a specific time-series problem and focus more on learning better feature representations for a general-purpose algorithm for structured data, regardless of the application.

Prediction of rainfall rate based on weather radar measurements.In this paper, weather radars are used to measure the electromagnetic radiation which is backscattered by cloud raindrops. Clouds produce more rain because clouds that backscatter more electromagnetic radiation have larger droplets of rain and therefore the clouds will produce more rain. The idea of this paper is to predict the rainfall rate by using the weather radar rather then the rain-gauges which is used to measure rainfall on the ground. In an experiment during the days June and august 1997 over the Italian Swiss Alps, the accurate data is collected from weather radar and surrounding rain-gauges at the same time. The statistical KNN classifier and neural SOM were executed for the grouping task using the radar data as input and the rain-gauges measurements as output. The rainfall rate on the ground was predicted by including the radar reflections with an average error rate of 23%. The results in this work shows that the prediction of rainfall rate based on weather radar measurements is possible**.** KNN is used. The system yielded a satisfactory success rate, which outperformed the traditional power-law relationship. It does not support for time series data.

Forecasting with Artificial Neural Networks: The state of the art. The interest in using artificial neural networks (ANNs) for forecasting has led to a massive rush in research activities in the past decade. While ANNs provides a great deal of promise, they also represent much uncertainty. Researchers till date are still not certain about the effect of ANNs key factors on forecasting performance. This paper presents a state-of-the- art survey of ANN applications in forecasting. Our purpose is to provide (1) a synthesis of published research in this area,(2) insights on ANN modeling issues, and (3) the future research directions. ANN is used. Overall, ANNs give satisfactory performance in forecasting. There may be a limit on what ANNs can learn from the data and make predictions

Forecasting the NN5 time series with hybrid models, International Journal of Forecasting In this paper, we suggest a simple way of predicting time series with the recurring seasonal periods. The values which are missing of the time series are estimated and interpolated in a preprocessing step. Here, we combine the several forecasting methods by accepting the weighted mean of forecasts that were generated with time-domain models which were authenticated on left-out parts of the time series. The hybrid model is a combination of a neural network group, an group of nearest trajectory models and a model for the 7- day cycle. This approach is applied to the NN% time series competition data set. A simple seven day cycle model, Nearest neighbor models are used. Accuracy is almost 70%.

2.1 PROPOSED SYSTEM

Existing System

The prestd function is used for preprocessing the NCEP/NCAR datasets. It is fed as inputs for training. Using the subtractive clustering the rainfall values are clustered and the rainfall states are identified as low, medium, and heavy. And it gives as outputs for training. The main important task of analyzing the data mining models are separating the input data into training set and testing set. when we are done with separating the data into testing set and training set, major portion of the data is used for the training set and minor portion of the data is used for the testing set. Here, the dataset used for training set is around 80% and the remaining 20% of dataset is used for the testing.

Proposed System

We prefer the deep learning approach for predicting the rainfall by using the Auto-encoder Neural network and multilayer perceptron. We are comparing the other state approaches with the present architecture. The results have an opinion that in terms of RMSE and MSE, our recommended planning performs better than the other remaining methods. The correctness of this can measured by RMSE and MSE, comparing with the other methods. With the conditions of the management and water resources, people's lives and the weather they own, predicting the precipitation is very important. We could face the wrong or unfinished estimation because of the precipitation measurement is effected by local and spatial climate changes, and some properties. In this project we are going to provide an study on the various types of methodologies which are used for forecasting and predicting of rain and there might be many issues founded when we applied various

methods for rainfall forecasting.The following algorithms are used in this project

1.ANN
2.ARIMA

1.ANN:

ANN (Artificial neural network) uses 2 deep learning methods which are the Auto-Encoders and Multilayer Perceptron. The Auto-Encoder extracts the all non-linear features, and then send it to Multilayer Perceptron (MLP), this helps the prediction much better.

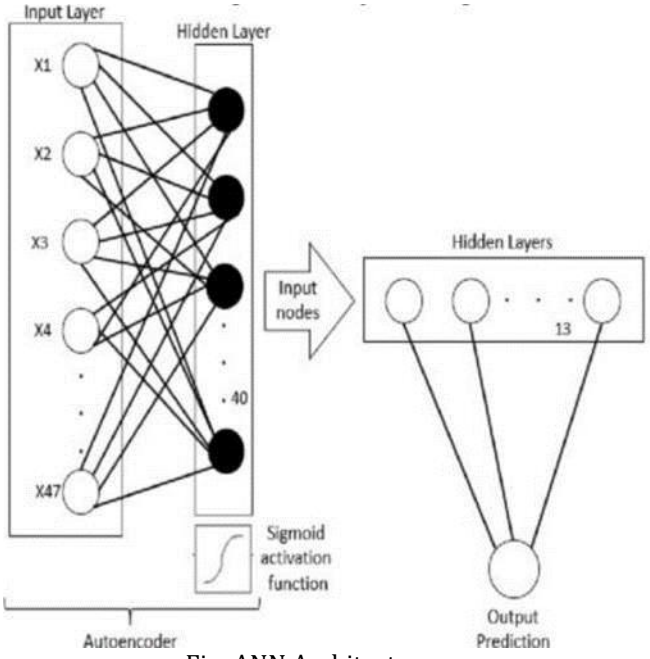


Fig: ANN Architecture

2.ARIMA MODEL(auto-regressive integrated moving average):

The ARIMA model is mainly used for time series prediction, forecasting and analysis. It has 4 methods and is suggested by the jenkins and Box, ARIMA model has 4 steps which are as follows.

Stage 1: In this stage recognizing series of the responses is been done. which are used for time series calculation and the autocorrelations by using the statement called IDENTIFY.

Stage 2: In this step, the variables which are already identified are estimated and also the evaluation of parameters are done by using statement called ESTIMATE.

Stage 3: In this stage the Diagnostics checking for above already collected parameters and variables is being done.

Stage 4: In this step, prediction of the time series values are being forecasted. which are the future values, by using the ARIMA model and also used the statement called FORECAST.

In this model the parameters used are m,n,o which defines 'm' as number of the lag observations, 'n' as the differencing degree and 'o' as moving average order.

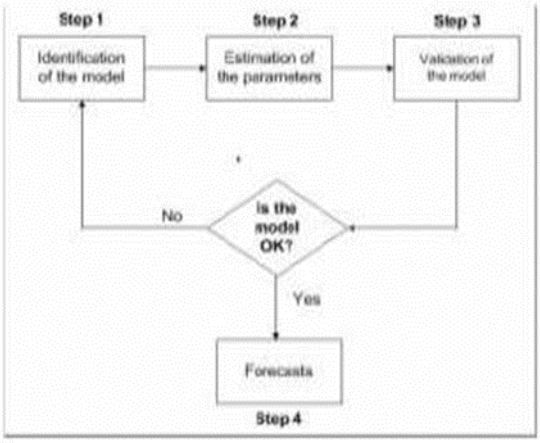
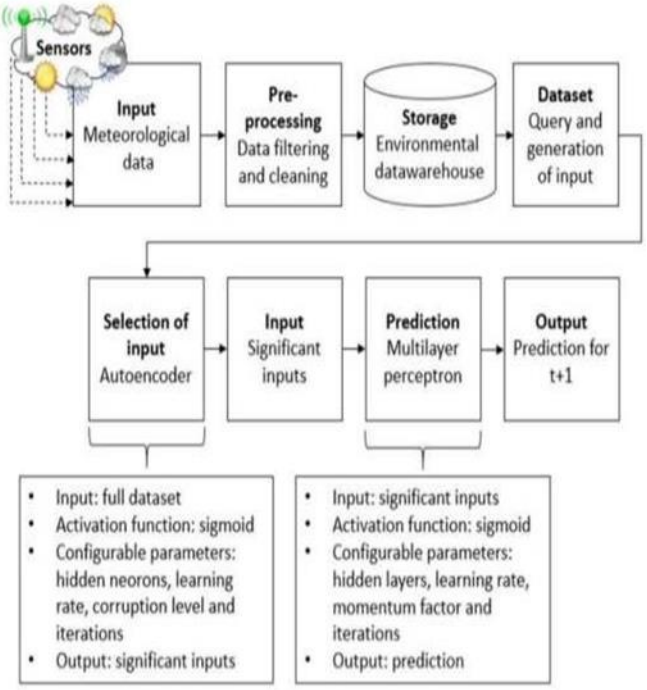


Fig: ARIMA Steps

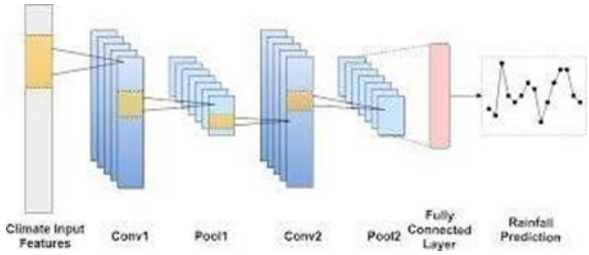
System Architecture:



METHODOLOGIES

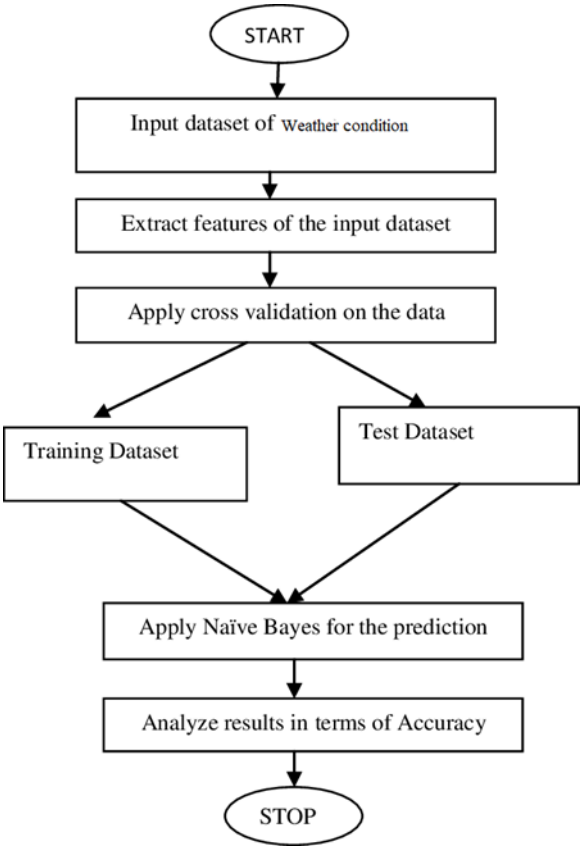
An ANN model (Artificial Neural Network) is the adaptive system in which that changes its formation based upon the internal or external data which rushes

through the network, during it's learning phase. The collection of connected Neurons together is called neural network. The all connected neurons are connected to the output from the one neuron is fetching as input to the others till the final output is being reached. In this network it learns when example of some set of the input data and the with the already recognized output for the inputs are present to it, and adjusted the weighting factors (whichever by programmed algorithm or the human interruption) and the connection weights will store important knowledge essential to bring the concluding result very nearer to the identified outputs/results. In this learning, Artificial Neural Network models are developed with 3 training algorithms for rainfall forecasting daily. By consuming the data which is available of study zone, error approach and the trail is employed to finalize present ANN (Artificial Neural Network) structure. In this model development the Neuro solution version 5 has been used. The first ANN (Artificial Neural Network) model 'A' was capable by using the MLP Back Propagation algorithm system with easy structure, 4 nodes are in the input layer and the single hidden layer with the seven nodes and the one node is in the output layer. The model's is the rainfall data (t) of present day and a 3 day lagged rainfall [(t 1) (t 2) (t 3)], and the output for this is the next day rainfall (t þ 1). The sigmoid function with 400 numbers of epochs are used for the transfer function. In the second ANN (Artificial Neural Network) model 'B', to train the network the Radial Basis Function (RBF) is being used. For this, the input and output of the training dataset was saved same as in MLP network. But, the transfer function, TanhAxon, was being used. The third ANN (Artificial Neural Network) model 'C' were skilled by using the Time Lagged Recurrent Networks (TLRNs). The data used to train the model for previous 2 models (A & B) was same for this. In this TLRN algorithm, number of nodes increased in hidden layer has been reduced the performance and so number of nodes which are in the hidden layer are reduced by two. For this all networks, the total number of hidden layers and the number of Neurons or nodes in the each layer was discovered by trail and the error. In this, out of '47' years of the rainfall information, 12 years of rainfall information is used as testing set and the remaining 35 years of rainfall information is used for the training set.



Naive Bayes

The Naïve Bayesian classifier was first reported in the year 1973 and also in 1992. The Naïve Bayes classifiers are the collection of the classification algorithms based on the Bayes Theorem. Bayesian classifiers are statistical classifiers. Naive Bayes classifier is based on Bayes rule of the conditional probability. The Naïve Bayes classifier analysis the each of its Attributes individually and it assumes that all of the attributes are important and autonomous. In fault-proneness prediction the Naïve Bayes Classifiers are used extensively. It needs a minor amount of training data set to estimate the parameters which are essential for the classification, it is one of the advantages of the Naïve Bayes classifiers.



3. CONCLUSION

We are developing an desktop application which uses the python language where system automatically predicts the rainfall information in the given region. In this project we represented an deep learning approach to predict the rainfall by using Auto encoder neural network and Multilayer perceptron. By comparing other state approaches with current architecture. The outcomes for this project aim that in the terms of RMSE and MSE, our planned architecture better performs the remaining methods. The accurateness of the project outcomes can measured by RMSE and MSE by relating the results with other methods. By including the conditions of the water resources and the management, people's valuable lives and climate they own, to predict the precipitation is of more important. Incorrect or the aborted precipitation is been influenced by the spatial, the local change and the property. In this paper, we have covered an study on various kinds of methods used for the forecasting and to predicting the rainfall and the problems that are found while applying various methods for predicting the rain. Because of the non-linear relations in the rainfall datasets and the capability for learning from past approaches, a better solution is made by Artificial Neural Network to all approaches available.

REFERENCES

[1] Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. Water (Switzerland), 10(11). https://doi.org/10.3390/w10111536

[2] Janani, B; Sebastian, P. (2014). Analysis on the weather forecasting and techniques. International Journal of Advanced Research in Computer Engineering & Technology, 3(1), 59–61. http://ijarcet.org/wp-content/uploads/IJARCET-VOL-3-ISSUE-1-59- 61.pdf

[3] Chaudhari, M. S., & Choudhari, N. K. (2017). Open Access Study of Various Rainfall Estimation & Prediction Techniques Using Data Mining. American Journal of Engineering Research (AJER), 7, 137– 139. http://www.ajer.org/papers/v6(07)/Q0607137139.pdf

[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). https://www.researchgate.net/profile/Aakash_Parmar 4/publication/319503839_Machine_Learning_Techniques_For_Rainfall_Prediction_A_Review/links/59afb922458515150e4cc2e4/Machine-LearningTechniques-For-Rainfall-Prediction-A-Review.pdf

[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. https://doi.org/10.14569/IJACSA.2018.090518

[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. https://doi.org/10.1038/s41598-019-49242-6

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

[8] 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). https://doi.org/10.1088/1742- 6596/983/1/012059

[9] Zeyi Chao, Fangling Pu, Yuke YinLing, B. and X. (2018). Research on real-time local rainfall prediction based on MEMS sensors. Journal of Sensors, 2018. https://doi.org/10.1155/2018/6184713

[10] Kaushik Dutta, Gouthaman. P. (2020). Rainfall Prediction using Machine Learning and Neural Network. https://www.ijrte.org/wp-content/uploads/papers/v9i1/A2747059120.

[11] J.Refonaa, M. Lakshmi, Raza Abbas, Mohammad Raziullha. (2019). Rainfall Prediction using Regression Model. https://www.ijrte.org/wp-content/uploads/papers/v8i2S3/B10980782S319.