International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-1, May 2020

Rainfall Prediction using Machine Learning and Neural Network

Kaushik Dutta, Gouthaman. P

Abstract—Rainfall prediction model mainly based on artificial neural networks have been proposed in India until now. This research work does a comparative study of two rainfall prediction approaches and finds the more accurate one. The present technique to predict rainfall doesn't work well with the complex data present. The approaches which are being used now-a-days are statistical methods and numerical methods, which don't work accurately when there is any non-linear pattern. Existing system fails whenever the complexity of the datasets which contains past rainfall increases. Henceforth, to find the best way to predict rainfall, study of both machine learning and neural networks is performed and the algorithm which gives more accuracy is further used in prediction. Recently, rainfall is considered the primary source of most of the economy of our country. Agriculture is considered the main economy driven source. To do a proper investment on agriculture, a proper estimation of rainfall is needed. Along with agriculture, rainfall prediction is needed for the people in coastal areas. People in coastal areas are in high risk of heavy rainfall and floods, so they should be aware of the rainfall much earlier so that they can plan their stay accordingly. For areas which have less rainfall and faces water scarcity should have rainwater harvesters, which can collect the rainwater. To establish a proper rainwater harvester, rainfall estimation is required. Weather forecasting is the easiest and fastest way to get a greater outreach. This research work can be used by all the weather forecasting channels, so that the prediction news can be more accurate and can spread to all parts of the country.

Keywords – Artificial neural network, Machine learning, Rainfall prediction, Tensor flow, Visualization

I. INTRODUCTION

In today's situation, rainfall is considered to be one of the sole responsible factors for most of the 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.

Manuscript received on April 02, 2020.
Revised Manuscript received on April 15, 2020.
Manuscript published on May 30, 2020.
* Correspondence Author
Kaushik Dutta, Information Technology, S.R.M Institute of Science and Technology, Chennai, India. Email: kaushikdutta700@gmail.com
Gouthaman.P*, Information Technology, S.R.M Institute of Science and Technology, Chennai, India. Email: gouthamanps@gmail.com

In some of the areas which have water scarcity, to establish rain water harvester, prior prediction of the rainfall should be done. This project deals with the prediction of rainfall using machine learning & neural networks. The project performs the comparative study of machine learning approaches and neural network approaches then accordingly portrays the efficient approach for rainfall prediction. First of all, preprocess is performed. Preprocess is the process of representing the dataset in the form of several graphs such as bar graph, histogram etc. When it comes to machine learning, LASSO regression is being used and for neural network, ANN (Artificial neural network) approach is being used. After calculation, types of errors, accuracy of both LASSO and ANN has been compared and accordingly conclusion has been made. To reduce the systems complexity, the prediction has been done with the approach that has better accuracy. The prediction has been done using the dataset which contains rainfall data from year 1901 to 2015 for different regions across the country. It contains month wise data as well as annual rainfall data for the same.

Currently, rainfall prediction has become one of the key factors for most of the water conservation systems in and across country. One of the biggest challenges is the complexity present in rainfall data. Most of the rainfall prediction system, nowadays are unable to find the hidden layers or any non-linear patterns present in the system. This project will assist to find all the hidden layers as well as non-linear patterns, which is useful for performing the precise prediction of rainfall [1]. Rainfall prediction is the application to predict the rainfall in a given region. It can be done in two types. The first is to analyze the physical law that affects rainfall and the second one is to make a system which will discover hidden patterns or the features that affects the physical factors and the process involved in achieving it. The second one is better because it doesn't include any type of mathematical calculations and can be useful for complex and non-linear data [2]. Due to presence of the system which doesn't find the hidden layers and nonlinear patterns accurately, the prediction results to be wrong for most of the times and that may lead to huge losses. So, the main objective for this research work is to find a system that can resolve both the issues i.e. able to find complexity as well as hidden layers present, which will give proper and accurate prediction thereby assisting the country to develop when it comes to agriculture and economy [3].

II. RELATED WORKS

Machine learning approach deals with predicting rainfall using machine learning approach. It finds the accuracy of the machine learning approach using two types of errors 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 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 stated 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 future rainfall. Customized, integrated and modified 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 taken 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 find 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, the future work left is that 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, which was 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 major advantages of RNN is, it can predict correctly 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 needs LSTM or ConvLSTM to retrieve the data which is being stored for long period of time. This process deals with getting the data from the data source and to how to use it effectively [7].

The back-propagation technique works well with less complex system, but as the complexity of the system increases back propagation method's accuracy decreases. This process deals with four types of inputs and three types of outputs layers.
Following are the four-input layer used:

1. Air temperature
2. Air humidity
3. Wind speed
4. Sunshine duration

Following are the output layers used:

1. Rainfall
2. Medium rainfall
3. High rainfall

[8].

In MEMS sensor, we will take the help of LSTM and ARMA model to see the prediction Firstly STL algorithm is applied to predict rainfall. STL stands for seasonal trend decomposition using Loess. In this, the observed data is being divided into different components such as trend, seasonal and reminder component and then the prediction is being done. Then the real time local rainfall prediction is being found out. Compared both the second approach was found more successful. Real time prediction means predicting data from taking the previous one- and seasonal-time data means predicting data based on the rules of LSTM. Now going detail on STL algorithm, it divides the data into trend, seasonal and reminder. Now talking about finding the real time local rainfall prediction this type of prediction is most useful because based on this we can plan our own things. This algorithm takes observed time series as input and based on that the future prediction is being done.

LSTM deals with finding two types of rainfall, one is real time and the second one is seasonal rainfall.
ARMA model does not work accurate with season wise real time rain prediction. The reason for it is ARMA helps to predict rainfall systematically which works well with local prediction not seasonal prediction. The advantage of these methods is its ability to predict real time rain locally and disadvantage is season wish rain prediction [9].

SARIMA (Seasonal auto regression integrated moving average) model was experimented in Sudan. Several calculations are made and on the basis of that graphs are being plotted and results are being found. Three kinds of graphs are plotted as a result. The resultant graphs show that stationary hypothesis is not true. It cannot be stationary but can be seasonal of 12. It exhibits a horizontal secular trend. The biggest disadvantage of this model is that, it is limited to very less area. The future enhancement of this approach can be to broaden its field [10].

When it is difficult to find whether the state is true or false, fuzzy logic provides a logic for it. The rainfall prediction with fuzzy logic takes place in three stages. Stages are fuzzification, inference engine and defuzzification. Fuzzification means converting of dataset into fuzzy dataset. Inference engine means setting up rules for the dataset and defuzzification means giving result as non-fuzzy value. First the linguistic variables are initialized, then member functions and rules are initialized. Then, with the help of a member function, the crisp data is converted into fuzzy variables. Then base rules are evaluated and results are calculated and at the end the results are changed into non fuzzy words. The advantage of using this logic is its easy of using but this logic in not accurate [11].

Hybrid classifier comprises of three segments. The segments are simulation segment, training segment and then at the last, testing segment. Simulation segment deals with data processing. The training segment deals with finding the best algorithm with the dataset and the data are being tested to get the predicted value. Post testing, it is found that this approach gave a better result. The future enhancement of this approach will be improving the time series rainfall prediction and effectively hybridizing the support vector regression model [12].

Linear regression approach deals with finding a relation between dependent and nondependent variables. This approach can give a good estimation of rainfall for a certain period of time. It deals with collection of data set and set it for further processing. The data's collected are further processed and result is predicted. The primary advantage of using this approach is that it is better than data mining approaches. Data mining approach gives generalized value unlike linear regression which gives estimated value. The biggest disadvantage of this approach is that it fails when it comes for long term estimation. The future enhancement of this process will be using multiple regressions for finding rainfall [13].

III. PROPOSED METHODOLOGY

The proposed system predicts rainfall for the approach which is more accurate. The data set is collected. There are two techniques to predict rainfall. The first one is machine-learning approach, which includes LASSO regression. The second one is neural network approach. This system first compares both the process and then accordingly gives result with the best algorithm. Steps associated with the proposed system are input of data, preprocess of data, splitting of data, training of the algorithm, testing of the dataset, comparing both the algorithm, giving the best algorithm, prediction with the more accurate algorithm and result at the end.

The main reason for not doing prediction with both the algorithm is to reduce the complexities of the whole system, so the system first finds the most accurate algorithm between machine learning and neural network and accordingly does prediction with the better one. The result will be received in the form of graphs and excel sheets. For preprocess , all the result will be received in the form of different graphs and for machine learning and neural network , the accuracy will be received in the form of Metrics as well as excel sheet and accordingly the predicted value will be received in the form of excel sheet which will contain two columns ID and predicted value. IDs will be same as that of in the datasheet. To get for which region prediction is being done, IDs should be matched with the IDs present in dataset.

A. Modules

The first module is UI (User Interface). UI is coded in python. It contains five buttons in it which are Browse, Clear, Preprocess, LASSO, Neural network and Quit. Browse button is used to select dataset from the system. Clear button is to clear the dataset selected. Preprocess is for the visualization, similarly LASSO and neural network is for the prediction purpose and quit for quitting the application. After the end of any process, a dialog box comes which displays the message that the process is successful. The UI is completely coded in python. It has several code snippets which connect to all other python coded files. This module is connected with all the rest modules present. After clicking the preprocess button, it gets connected to the visualization module, after clicking to LASSO button it gets connects to LASSO module and after clicking to Neural network button it gets connected to neural network module. So, in this way all the other modules are connected with this module. Second module is Visualization. Before going for the prediction, this process can be done. It is representation of the dataset in form of graph. It eases the process of comparison. After clicking on preprocess button this process is being performed and we receive several graphs.
Even before execution of both machine learning and neural network algorithm, it first comes to this module and then it goes to its own module, but in that case no graphs are being formed because that will make the whole process more complex by creating duplicity.
The graphs received after clicking on preprocess button are Correlation Metrics, Scatter Metrics, Max value for month by month, Subdivision mean value for month by month, Sum of every quarter of subdivision, Sum of every quarterly, Sum of year by year and Sum of month year by year. The third module is LASSO.

LASSO deals with structuring data at a single point. It means to bring everything at a central point.
It performs variable selection to give accurate prediction. Now the final module i.e. Neural network. Deep learning algorithm is being used. It is a form of algorithm, which mimics learning process as human being. It has three layers i.e. Input, Hidden and Output.

B. Dataset

The dataset used in this system contains the rainfall of several regions in and across the country. It contains rainfall from 1901 – 2015 for the same. Along with that annual rainfall is also been used and the rainfall between the transition of two months. There are in total 4116 rows present in the dataset. The dataset is been collected from data.gov.in.
Category – Rainfall in India
Released under – NDSAP
Contributor – Ministry of Earth Sciences, IMD
Group – Rainfall
Sectors – Atmosphere science, earth sciences, science & technology

Source: OGD
Below is the snapshot for the same.

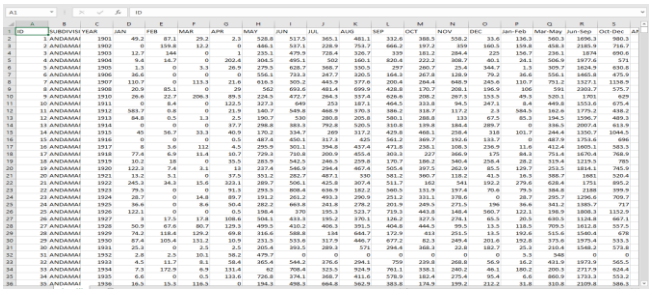

Fig. 1. Glimpse of the dataset

C. Architecture

Entities associated with the architecture are user input data, preprocess, LASSO, neural network, splitting of data, training of the algorithm, testing of the data and result at the end. User gives data to the system from the local system. After that User does the preprocess step and then compares the LASSO regression and neural network approach. User finds the accuracy of both the algorithm and accordingly does prediction with the most accurate one. Preprocess is the step to represent dataset in the form of graphs. Following are the types of graphs extracted from the data. The graphs are line graph, bar graph and histogram. Line graph is the simple plotting of points in the graph and connecting two points with the help of a straight line. To identify better, all the line graphs for different regions are represented with different colors. Bar graph is the plotting of data in the form of bar. Mainly stacked bar graph is used in this project. Histogram is the representing of data with the help of bars of different height. For both LASSO and neural network, accuracy is calculated after algorithm is being trained. Both the algorithms are trained and then the accuracy is received for both. The accuracy is being calculated after calculating errors. The errors calculated are MAE, MSE, R-SQUARED and RMSE. To reduce the complexity of the system, prediction will only be done for the most accurate algorithm because if both the algorithms are giving predicted value, then that can lead to unnecessary usage of data and that will increase the complexity of the system and make it slow. The result will be received in the form of graphs, Metrics and excel sheet. The graphs received are Correlation Metrics, Scatter Metrics, Max value for month by month, Subdivision mean value for month by month, Sum of every quarter of subdivision, Sum of every quarterly, Sum of year by year and Sum of month year by year. Along with this two Metrics, which will be show the accuracy of both the algorithms will be received. Three excel sheet will be received at the end of all process. Two will have the accuracy for both LASSO and neural network in tabular form and one excel sheet will have the predicted value which will be received from the most accurate algorithm.

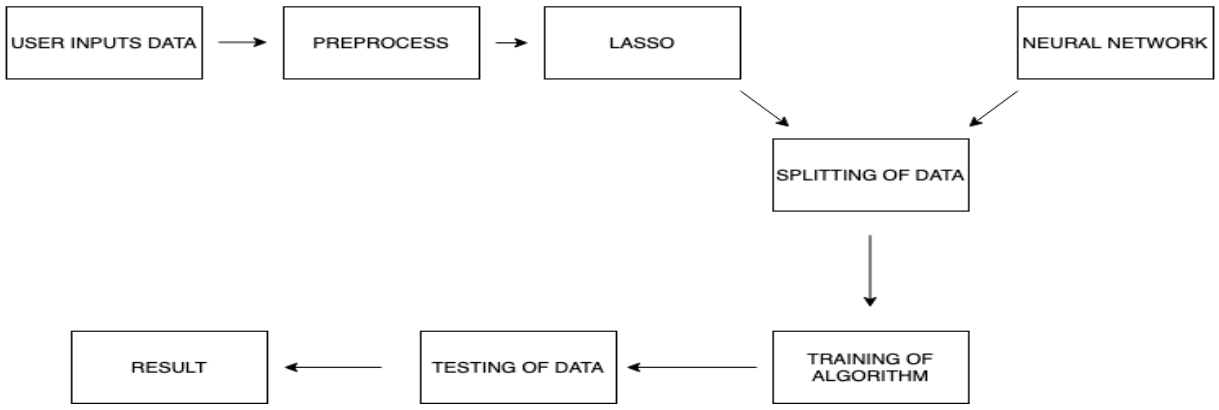

Fig. 2. Architecture

IV. IMPLEMENTATION AND RESULT

User gives the dataset as input in the system. Along with that, there will be three more buttons present. First one is for preprocess which represent the dataset in the form of graphs, the second one is LASSO which gives the accuracy of LASSO regression and the third one is neural network which gives the neural network 's accuracy.

So, to have a better understanding of the dataset and for better comparison, first preprocess should be done. Before going for the prediction, preprocess can be done. It is representation of the dataset in form of graph. It eases the process of comparison and along with that it also gives a better understanding of the dataset present. Dataset should be split in two parts, the first part deals with training the algorithm used and the rest part used to predict the amount of rainfall. Rainfall is predicted only with the algorithm with more accuracy. The algorithm used should undergo training before it does prediction. So, in this part of the system, the training is been done. This is done in both the approaches i.e. on LASSO and ANN. This step gives a proper idea of which algorithm is more accurate among the two. Then the remaining dataset (which is not used in training) is being used and rainfall prediction is been done. This part is also done in both the approaches. Finally, after the all the process is completed, the result is received in form of graph and table which shows the future rainfall and the accuracy of the algorithm. After preprocess the graphs which are received are Correlation Metrics, Scatter Metrics, Max value for month by month, Subdivision mean value for month by month, Sum of every quarter of subdivision, Sum of every quarterly, Sum of year by year and Sum of month year by year. After that for both LASSO and neural network, the accuracy is received in the form of Metrics and excel sheet. In Metrics along with the accuracy different types of errors are also shown and the same is represented in the excel sheet. After all, at last the predicted value is stored in excel sheet and is received.

A. Errors calculated

The accuracy of the approaches is being calculated against the types of errors that can produce negative effect on the algorithm. These errors can affect the algorithm's accuracy and hence are being calculated. The types of errors that is being calculated are MAE, MSE, RMSE and R-SQUARED.

MAE calculates all the absolute errors and then finds the mean value for all. It first calculated the mean of all the dataset present, then subtract the mean value with each data individually and add all the resultant value and finally divides it with the total number of dataset present.

$$MAE = \frac{1}{n}\sum|xi - x| \quad (1)$$

Next error is MSE. It is almost similar to mean absolute error.

$$MSE = \frac{1}{n}\sum(xi - x)^2 \quad (2)$$

The only difference is, instead of adding the resultant (subtracted value of mean with each dataset), it finds the square of it and add them. RMSE error is being calculated by subtracting all the predicted and actual values with each other, finding all the squares of it and adding all the squared value.

$$RMSE = \frac{\sqrt{\sum(xi-yi)^2}}{N} \quad (3)$$

(**x**i = Predicted value yi = actual value)

The total value that we will receive is stored. The stored value then further divided by total values present. The resultant value is squared rooted.

$$R^2 = 1 - \frac{x}{y} \quad (4)$$

(X= regression error (sum squared) Y = total error (sum squared))

The above errors are being calculated by subtracting the division value of sum squared regression error and sum squared total error with value one.

B. Front end

This screen will pop up in the screen as soon as the project starts. The user needs to give dataset in the column given. After dataset is being inserted in the project, the directory of dataset is shown in the column. For every process, at the end a dialogue box is popped up with a message that the following process is completed. If dataset is not given in the system and if user wants to execute any process, then a dialog box will pop up which will ask the user to select dataset

C. Anaconda

Anaconda is an open source software used to run python codes.

Tenser Flow name: tf

The errors values and the accuracy value are popped up in anaconda for LASSO and NEURAL NETWORK. In anaconda, while the project is being executed, the dataset which is being used is shown there. So, while execution of the code, the user can cross verify it and can stop the processing instant on finding that the wrong dataset is being used.

D. Result folder

All the graphs, Metrics and tables are stored in a specific folder named "Result".

E. Visualization

This option deals with representing the dataset in form of graphs. Different types of graphs are being produced after execution of this process. Below are the different Metrics and graphs:

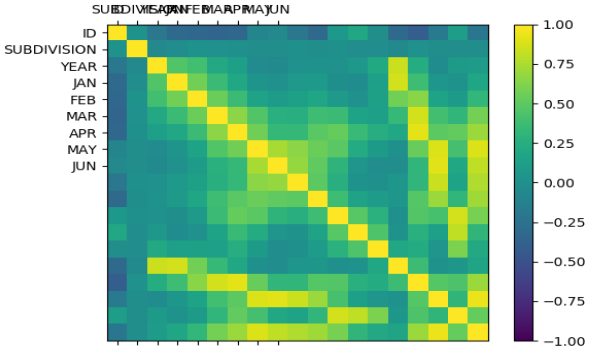

Fig. 3. Correlation/ Scatter metrix

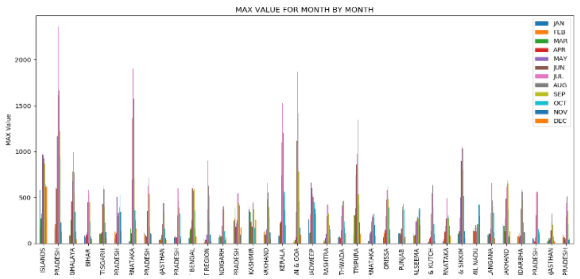

Fig. 4. Max value for month by month

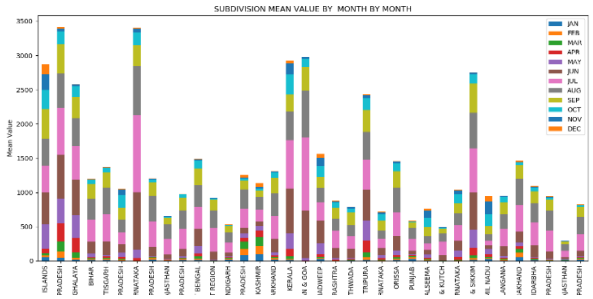

Fig. 5. Subdivision mean value by month by month

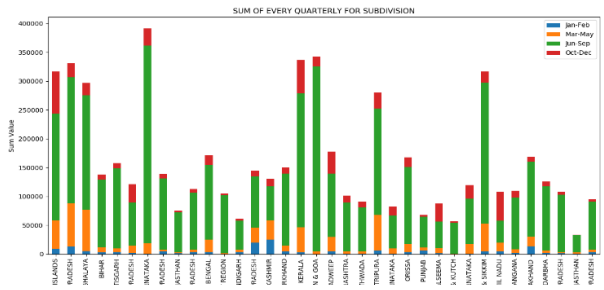

Fig. 6. Sum of every quaterly for subdivision

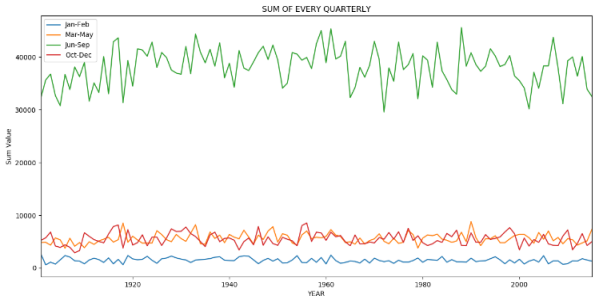

Fig. 7. Sum of every quarterly

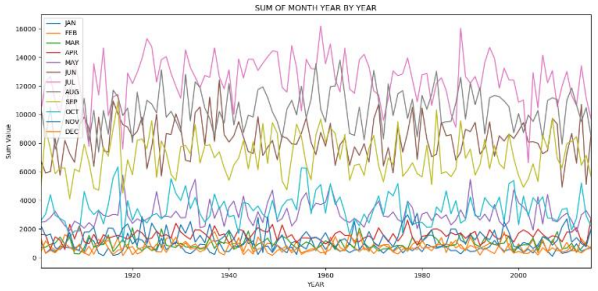

Fig. 8. Sum of month year by year

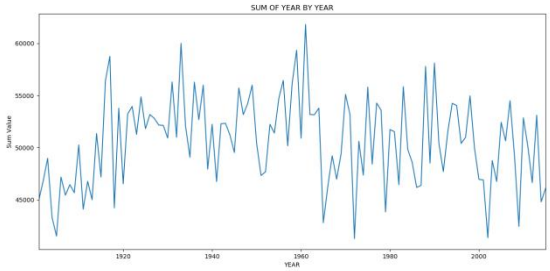

Fig. 9. Sum of year by year

F. Lasso

This process gives the accuracy in form of Metrics as well as table. This algorithm being the most accurate one also gives prediction in form of tables. The table below provides insights about error table for LASSO.

Table-I: Error Table for LASSO

Error	Percentage
MSE	0.08663
MAE	0.249793
R- SQUARED	1
RMSE	0.29433

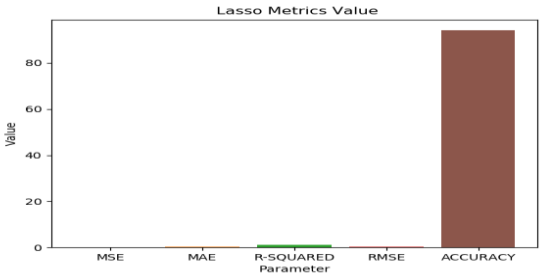

Fig. 10. Accuracy metrics of lasso

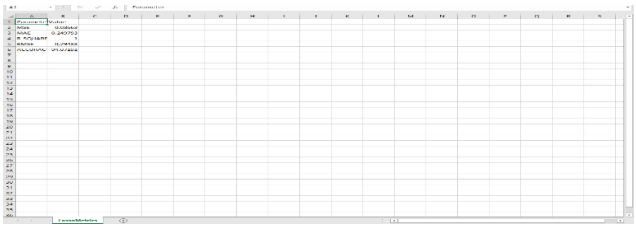

Fig. 11. Accuracy of lasso in form of excel sheet

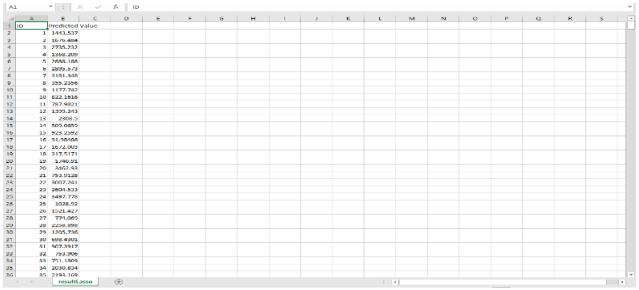

Fig. 12. Predicted value in form of excel sheet

G. Neural network

This process gives accuracy in form of Metrics and table. The table below provides insights about error table for Neural network.

Table-II: Error Table for Neural Network

Error	Percentage
MSE	29.30957
MAE	14.50885
R- SQUARED	2.548836
RMSE	17.12004

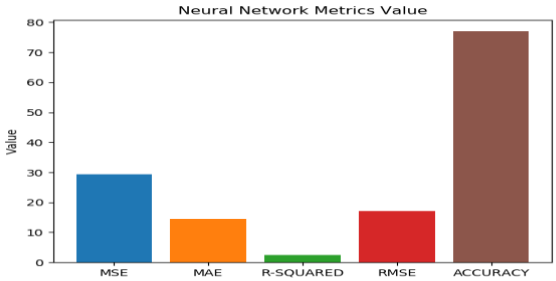

Fig. 13. Accuracy metrics of neural network

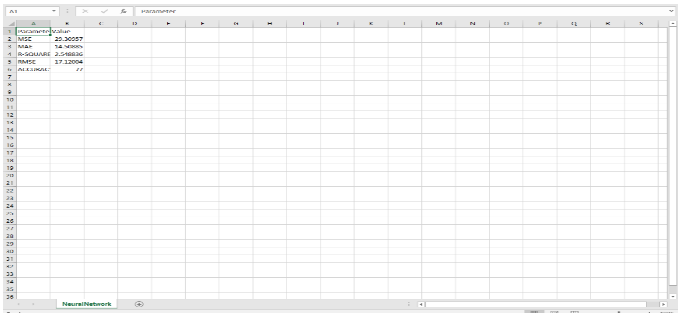

Fig. 14. Accuracy of neural network in form of excel sheet

V. CONCLUSION

Rainfall being one of the sole responsibilities for maximum economy of India, it should be considered the primary concern for most of us. The current approach for rainfall prediction fails in most of the complex cases, as it is unable to predict the hidden layers present, which is yet to be recognized for performing the precise prediction. To achieve an effective way to predict rainfall, two ways are being compared. One is machine-learning approach and the second one is artificial neural networks approach. In the first one, LASSO regression approach is being taken. Before the comparison is being done, visualization of the dataset is performed which is beneficial for effective comparison. Dataset of 4116 rows is being used that includes previous rainfall amount in various regions in and around the country. Dataset is divided into two parts i.e. 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 to train the algorithm. After performing the comparison, the conclusion of the system is that LASSO regression process is more accurate than the artificial neural network process. After comparison, we got to understand that the accuracy for LASSO is around 94% whereas ANN is 77%. Therefore, LASSO is the best analytical algorithm for predicting the rainfall in any given region.

The future enhancement of this project can be an approach towards about how to reduce the percentage of errors present. 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. The accuracy of the algorithm can be additionally tested on increase in the complexity. Many other types of errors can be calculated in order to test the accuracy of any of the above algorithms. Henceforth, algorithm for testing daily basis dataset instead of accumulated dataset could be of paramount Importance for further research.

More the accuracy of the system used for rainfall prediction, smarter will be the agriculture. Along with that, this will be an efficient tool for people in coastal areas of the country thereby making them well aware of the situation in advance.

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_Parmar4/publication/319503839_Machine_Learning_Techniques_For_Rainfall_Prediction_A_Review/links/59afb922458515150e4cc2e4/Machine-Learning-Techniques-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. 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. https://doi.org/10.12983/ijsrk-2014-p0320-0327
11. 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. https://ijcsmc.com/docs/papers/April2019/V8I4201937.pdf
12. KavithaRani, B., & Govardhan, A. (2014). Effective Features and Hybrid Classifier for Rainfall Prediction. *International Journal of Computational Intelligence Systems*, *7*(5), 937–951. https://doi.org/10.1080/18756891.2014.960234
13. Prabakaran, S., Naveen Kumar, P., & Sai Mani Tarun, P. (2017). Rainfall prediction using modified linear regression. *ARPN Journal of Engineering and Applied Sciences*, *12*(12), 3715–3718. http://www.arpnjournals.org/jeas/research_papers/rp_2017/jeas_0617_6115.pdf

AUTHORS PROFILE

Kaushik Dutta is currently pursuing his final year of B.Tech in the Department of Information Technology from S.R.M Institute of Science and Technology, Chennai. He completed his 10th standard in the year 2014 and 12th standard in the year 2016. He has been honored with a certificate from CBSE for his outstanding performance and for being amongst the top 0.1 percent of successful candidates in physical education during his 12th standard board exam. His area of interest includes machine learning, artificial neural network and artificial intelligence. In addition, he has completed an online course in NPTEL within the Cloud computing domain and also learning online courses related to his interests. He is currently working as an Intern in an IT company.

Gouthaman.P received his Bachelor Degree in Computer Science Engineering from Anna University, Chennai in the year 2005 and Master of Information Systems in the year 2010 from the University of Ballarat, Australia. From 2015, he has been working as Assistant Professor with the School of Computing, SRMIST, Chennai and teaches courses, such as, Software Engineering and Project Management, System Integration and Architecture and IT Infrastructure Management. Prior to that, he was working in the IT industry as a Project Lead for 7 years where he was appreciated for his outstanding contributions in various projects. He is a member of IAENG, IRED and CSTA. His research interests include Software Engineering, Internet of Things and Machine Learning.