Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/15243
DOI URL: http://dx.doi.org/10.21474/IJAR01/15243

RESEARCH ARTICLE

PREDICTING RAINFALL BASED ON MACHINE LEARNING ALGORITHM: AN EVIDENCE FROM BOGURA DISTRICT, BANGLADESH

Md. Mostafizur Rahman[1] and M. Sayedur Rahman[2]

1. Associate Professor, Data Mining Research Group, Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh.
2. Professor, Environmental Science Research Group, Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh.

Manuscript Info

Manuscript History
Received: 23 June 2022
Final Accepted: 25 July 2022
Published: August 2022

Key words:-
K-Nearest Neighbors, Classification and Regression Tree, Logistic Regression, Support Vector Machine, Random Forest and Bogura

Abstract

Accurately and timely predicting climatic variables are most challenging task for the researchers. Scientists have been trying numerous methods for forecasting environmental data with different methods and found confusing performance of different methods. Recently machine learning tools are considering as a robust technique for predicting climatic variables because these tools extracted hidden relationship from the data and can predict more correctly than existing methods. In this paperwe compare the forecasting performance of various machine learning algorithms such as Classification and Regression Trees (CART), Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (K-NN) and Random Forest (RF) in case of Bogura district in Bangladesh. The weekly rainfall related time series data such as temperature, humidity, wind speed, sunshine, minimum temperature and maximum temperature for the time period January, 1971 to December, 2015 were considered. The model evaluation criteria precision, recall and f-measure and overall accuracy confirms that Random Forest algorithm give best forecasting performance and cross validation approach which produce some graphical view model comparison also confirm that the Random Forest algorithm is the most suitable algorithm for predicting rainfall in case of Bogura district, Bangladesh during this study period.

Introduction:-

The world is suffering the impact of climate changing because the weather conditions always changes and become extreme day by day. There are number of new techniques and technologies are appearing to modeling and forecasting the climatic variables. Among these new techniques machine learning algorithms are proven robust method for this purpose (Gupta and Ghose 2015). The data mining as well as machine learning algorithm analyze historical data and extract hidden information which is used for recognizing similar pattern of data for classification task or making prediction (Yadav and Khatri, 2016). Ahamad et. al. (2017) used data mining technique for predicting upcoming situation in various domains such as finance, stock market and climate change etc.The time series data is collected over specific period of time such as hourly, daily, weekly, monthly, quarterly or yearly (Mishra et al., 2018). The occurrence of rainfall depends on the interaction between several complex atmospheric

Corresponding Author:- Md. Mostafizur Rahman
Address:- Associate Professor, Data Mining Research Group, Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh.

processes. The humidity, temperature, wind speed & direction and cloud coverage are the most important factors that influence the occurrence of rainfall. A large uncertainty involved in determining the contribution of the atmospheric processes is one of the biggest challenges to face in developing rainfall prediction models. Pre-awareness about the rainfall status helps in planning the water management, warning for early flood, transport and construction work and flight operation etc. (Chau and Wu 2010, Wu et al. 2015). Rahman and Rahman (2020a) monitored drought vulnerability at Different Time Scales in case of Rajshahi district. Rainfall based drought also investigated by Rahman and Rahman (2020b & 2021) in case of northwestern part of Bangladesh.

To extract information from the data is useful to transform the information into knowledge which plays vital role in different sectors. It is always challenging for the scientist to predict climatic variables with new techniques. Badhiye et al. (2012) used K-nn method to discover knowledge from data for classification and prediction purpose. Zaw and Naing (2008) investigated the performance of different statistical models for predicting rainfall data of Myanmar. Olaiya and Adeyemo (2012) compared the performance of ANN model and decision tree algorithm for predicting different climatic variables and found ANN model performed better than Decision tree algorithm for all of these variables. Cramer et al (2017) compared the performance of Markov Chain model with Genetic algorithm, Support Vector Regression, Radial Basis Neural Networks, M5 Rules, M5 Model trees, and k-Nearest Neighbors for predicting rainfall for 42 cities and concluded that all of these machine learning techniques perform better than the existing Markov Chain model for all of these cities.Zainudin et al (2016) examined the performance of Naïve Bayes, Support Vector Machine, Decision Tree, Neural Network and Random Forest algorithm for predicting Malaysian rainfall data found the better performance of Random Forest algorithm. Aftab et al (2018) examined the performance of Support Vector Machine (SVM), Naïve Bayes (NB), k Nearest Neighbor (kNN), Decision Tree (J48) and Multilayer Perceptron (MLP) algorithm for rainfall prediction in case of Lahore city and found that these techniques performed well for no-rain class however for rain class, the techniques did not perform well. Rahman et al. (2021) compared the performance of different data mining technique for predicting rainfall data from Rajshahi district, Bangladesh and found the preference of K-Nearest Neighbors algorithm. Beside these different authors such as Sivapragasam et al (2001), Monira et al (2010), Kannan et al. (2010), Sethi and Garg (2014), Talib et al. (2017) andTharun et al. (2018) investigated the performance of different data mining models and techniques for forecasting the rainfall for specific cities or regions.

From the above discussion we found that predicting climatic variables with new algorithm is always interesting for researchers. Comparing the performance of different machine learning algorithms for forecasting the rainfall data in case of Bogura district in Bangladesh is rare. So, the aim of this paper is to compare the forecasting performance of different machine learning algorithms in case of Bogura district, Bangladesh. This study will be helpful to aware in advance for rainfall condition in coming year which will save lives and properties of the people and contribute for the development of the economy in case of study area. The rest of the paper is organized as follows: Section 2 describes machine learning algorithms, Section 3 present the study area, result and discussion are given in section 4 and finally section 5 concludes the paper.

Machine Learning Algorithms

Machine learning algorithms provide accurate knowledge in the form of useful rules, techniques, visual graphs and models for the weather parameters over the datasets. This knowledge can be used to support the decision-making for various sectors. The goals for data analysis are those which involve weather variations that affect our daily runtime changes in min and max temperature, humidity level, rainfall chances and speed of wind. The flowchart of machine learning algorithm is given below:

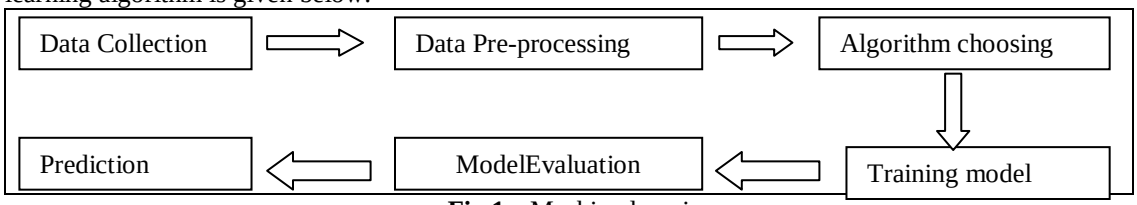


Fig 1:- Machine learning process.

Data Collection

The data used for this work was collected from Bangladesh Meteorological Department (BMD). The case weekly data covered the period from January, 1971 to December, 2015. The input dataset for rainfall prediction is obtained

consists of several atmospheric attributes such as temperature, humidity, wind speed, sunshine, minimum temperature and maximum temperature its name, type and measurement unit are given in Table 1.

Table 1:- Data sets attributes with their measurements scale.

Variables	Type	Measurement	Missing Values
Rainfall status	Categorical	(yes/no)	00
Temperature	Continuous	Degrees Celsius	05
Humidity	Continuous	%	40
Wind Speed	Continuous	Meters per second	10
Sunshine	Continuous	Hour	19
Max.Temp	Continuous	Degrees Celsius	30
Min.Temp	Continuous	Degrees Celsius	15

Data Pre-processing

Data Pre-processing is a crucial stage in classification framework which ensures the high accuracy of mining results. This stage consists of two activities: cleaning and data transformation. The data from Table 1 indicate that all of the variables are numerical and has different measurement scale and also some missing value of each attributes. Beside the missing values, dataset also contained noise where value resides below or exceeds from a certain limits. The missing value was filled up by the smoothing technique. Since our data sets contain different measuring scale so we transform all of the variables by Minimum Maximum Normalization method which confirm the entire attribute converted to the same range.

Selected Models

In this paper we consider different machine learning techniques for rainfall prediction these are describe below:

Classification and Regression Trees (CART)

Breiman et. al. (1984) introduced Classification & Regression Trees (CART) which is strictly binary, containing exactly two branches for each decision node. Each root node represents a single input variable (x) and a split point on that variable which is numeric. The leaf nodes of the tree contain an output variable (y) which is used to make a prediction. For example, the data from Afonso et. al. (2012), determine whether a person is infected by influenza or not. The CART decision tree is grown up as:

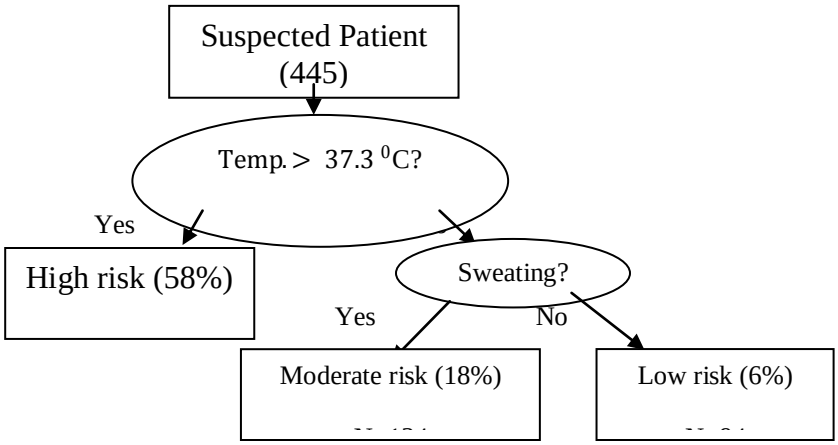


Fig 2:- CART decision tree (Source: Afonso et. al. (2012)).

Logistic Regression (LR)

Logistic regression accomplishes binary classification tasks by predicting the probability of an outcome, event, or observation. It belongs to supervised machine learning algorithm. The logistic regression model produces a binary outcome which contain two possible classes yes or no. or true/false or 0/1 (Kanade, 2022). The data generated from this hypothesis can fit into the log function that creates an S-shaped curve known as sigmoid. Using this log function, we can further predict the category of class. Mathematically, the binary Logistic Regression model can be written as:

$$y = \frac{e^{(b_0+b_1x)}}{1+e^{(b_0+b_1x)}}$$

where x is input value, y is predicted value, b_0 is intercept and $b_1$is co-efficient for input (x).

Support Vector Machine (SVM)

The Support Vector Machines known as SVM introduced by Vapnikwhich enables us to model higher dimensional, non-linear models (Vapnik, 1963). The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points.

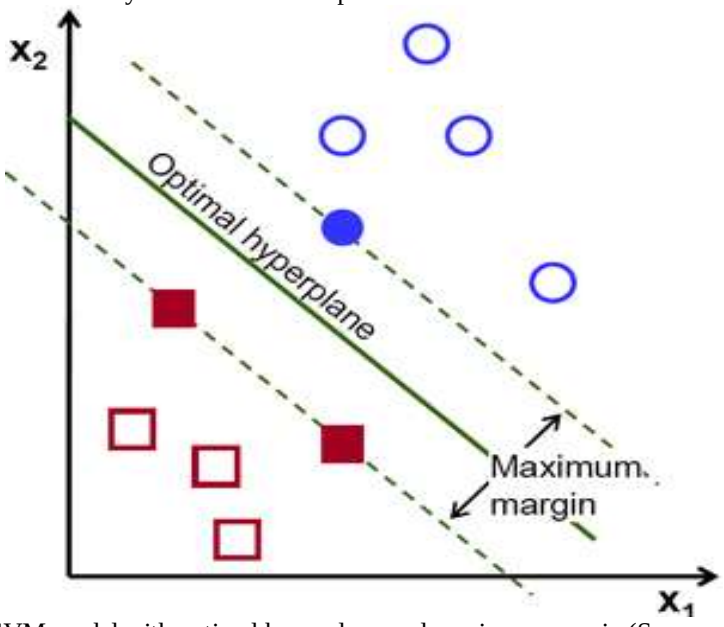


Fig 3:- Linear SVM model with optimal hyperplane and maximum margin (Source: Gandhi, 2018).

In the above figure there are two classes (red and blue color). In order to separate these two classes there are many possible hyperplane which distinguish these two classes. The aim is to find a plane which has maximum margin which indicates the maximum distance between these two data points. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence (Gandhi, 2018).

K-Nearest Neighbors (K-NN)

The k-nearest neighbors (K-NN) algorithm is a classifier belongs to supervised learning algorithm, which uses proximity to make classifications or predictions problem. It performs classification task based on the assumption that similar points can be found near one another. Predictions are made for a new instance by searching through the entire training set for the K most similar instances and summarizing the output variable for those K instances (Harrison, 2018). The following algorithm described the calculation method for K-NN algorithms.

1. Decide on the number of neighbors (K). It converts any real value between 0 and 1 into another value.
2. Determine the Euclidean distance between K neighbors.
3. Using the obtained Euclidean distance, find the K closest neighbors.
4. Count the number of data points in each category among these k neighbors.
5. Assign the new data points to the category with the greatest number of neighbors.

In this way the algorithm performs the classification task.

Random Forest

The random forest algorithm is more commonly applied to classification problems. The forest generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms (Mbaabu, 2020). Classification in random forests using an ensemble technique to obtain the desired result. This dataset consists of observations and features that will be selected randomly during the splitting of nodes.

A random forest system relies on different decision trees. The decision nodes, leaf nodes and root nodes are the basic component of decision tree. The leaf node of each tree is the final output produced by that specific decision

tree. The selection of the final output follows the majority-voting system. In this case, the output chosen by the majority of the decision trees becomes the final output of the random forest system. The diagram below shows a simple random forest classifier.

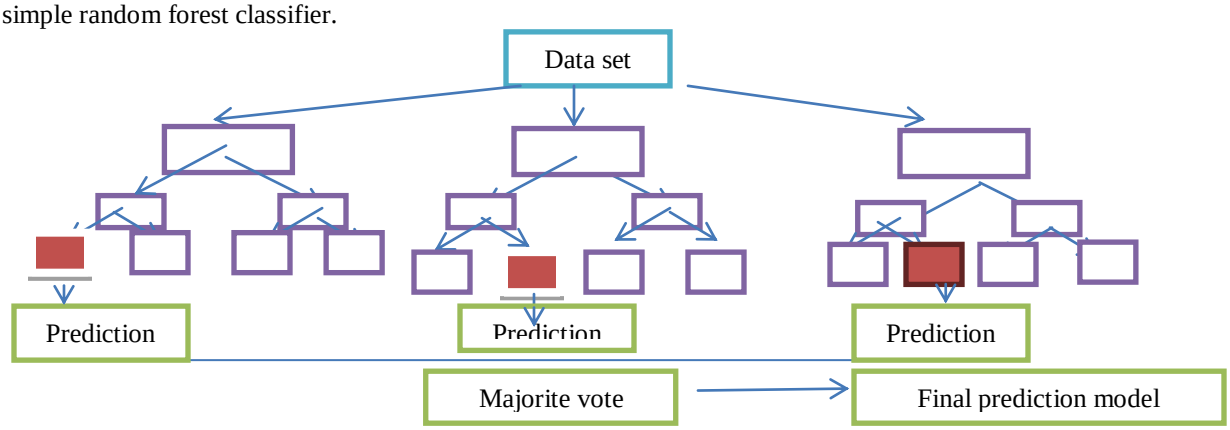


Fig 4:- Random Forest classifier (Image Source: Medium, https://miro.medium.com/max/5752/1*5dq_1hnqkboZTcKFfwbO9A.png)

Model Evaluation

The confusion matrix classifies the predictions against the actual values. It produces the 2×2 matrix which contains True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) that produce Precision, Recall, F-measure and Overall Accuracy.

Precision

Precision is the ratio of True Positive (TP) entities against all positive entities. The calculating formula for precision is given by:

$$Precision = \frac{TP}{(TP + FP)}$$

Recall

Recall is the ratio of True Positive (TP) entities against all positive entities in reality. In case of false negative priority recall is used. The calculated formula for recall is:

$$Recall = \frac{TP}{(TP + FN)}$$

F-Measure

Sometime precision and recall can not identify model accurately if there is diverse value precision and recall. In this situation F-measure used to choose best performing model. The mathematical formula for F-measure is given by:

$$F - measure = (2 * Precision * Recall)/(Precision + Recall)$$

Overall Accuracy

The overall accuracy which is the proportion of the total number of predictions that was correct and this measure can be calculated as:

$$Accuracy = (TP + TN)/(TP + FP + FN + TN)$$

Study Area

Bogra district officially known as Bogura district is situated in the northern part of Bangladesh. Its total area is 2898.25 sq km and located in between 24°32' and 25°07' north latitudes and in between 88°58' and 89°45' east longitudes. This district is surrounding by Joupurhat and Gaibandhadistrict on the north, Natore and Sirajganjon the south, Jamunariver andJamalpurdistrict on the east, Naogaondistrict on the west. Karatoya, Jamuna and Ichamati are the major rivers in this district. The remarkable beels are Betgari, Koigari, Subeel, NurulerBeel, PoradahaBeel, Keshpathar, Kalidaha, Garai, Erulia, Dasukdaha, Sara, Gobarchapa, Ramchandrapur, Kokira, BahuarBeel and KachiarBeel (Banglapedia, 2021). The map of the Bogura district is given below.

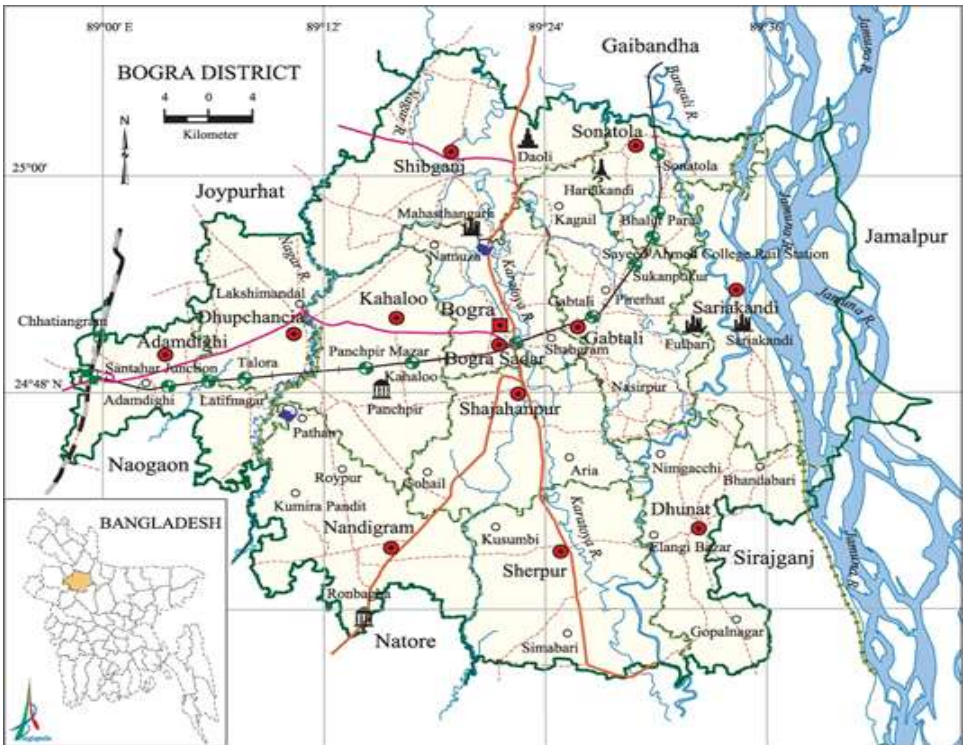


Fig 5:- Map of Bogura district (Source: Banglapedia).

Results and Discussions:-

A confusion matrix is used to measure the performance of classification in case of Machine learning algorithm. It usually helps us to know the performance of the classification model on a set of test data for that the true values are known.In our study we compare the forecasting performance of various machine learning algorithms such as Classification and Regression Trees (CART), Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (K-NN) and Random Forest (RF) algorithm in case of Bogura district, Bangladesh. In our study we consider 75% as training data and 25% as test data and applied these models in the training data and predict based on test data. The performance of different machine learning algorithms is verified by Precision, Recall, F-measure and Overall Accuracy with 95% Confidence Interval (CI) for both rain (Y) class and no-rain (N) class. The estimated result of the confusion matrix is reported in Table 2.

Table 2:- Model evaluation criteria.

Model	Class	Precision	Recall	F-measure	Overall accuracy (95 % CI)
CART	Y	0.7979	0.6455	0.5635	0.7867
	N	0.7634	0.8750	0.6364	(0.750, 0.819)
LR	Y	0.8324	0.7306	0.5937	0.7920
	N	0.7273	0.8300	0.6241	(0.756, 0.824)
SVM	Y	0.8426	0.7545	0.6014	0.7955
	N	0.7249	0.8210	0.6215	(0.760, 0.827)
KNN	Y	0.8892	0.8125	0.6130	0.8427
	N	0.7682	0.8599	0.6323	(0.810, 0.871)
RF	Y	0.8750	0.7955	0.6140	0.8566
	N	0.8254	0.8949	0.6416	(0.825, 0.884)

The empirical results from Table 2 confirmed that the Random Forest algorithm performed better based on precision, recall and F-measure for both rain and no-rain class where the Classification and Regression Trees (CART) algorithm performed worst. The overall accuracy with 95% confidence interval indicates that the Random Forest algorithm showed highest classification performance than all of the other models used in this study in case of

Bogura district for both rain and no-rain class for this time interval. So, Table 2 concluded that RF algorithm is the most suitable classification algorithm for predicting rainfall status in case of Bogura district, Bangladesh. For more confirmation we use cross validation approach. We repeated the procedure 10 times and found the accuracy measure. The summaries such as minimum, first quartile, median, mean, third quartile and maximum value of accuracy is presented in Figure 5.

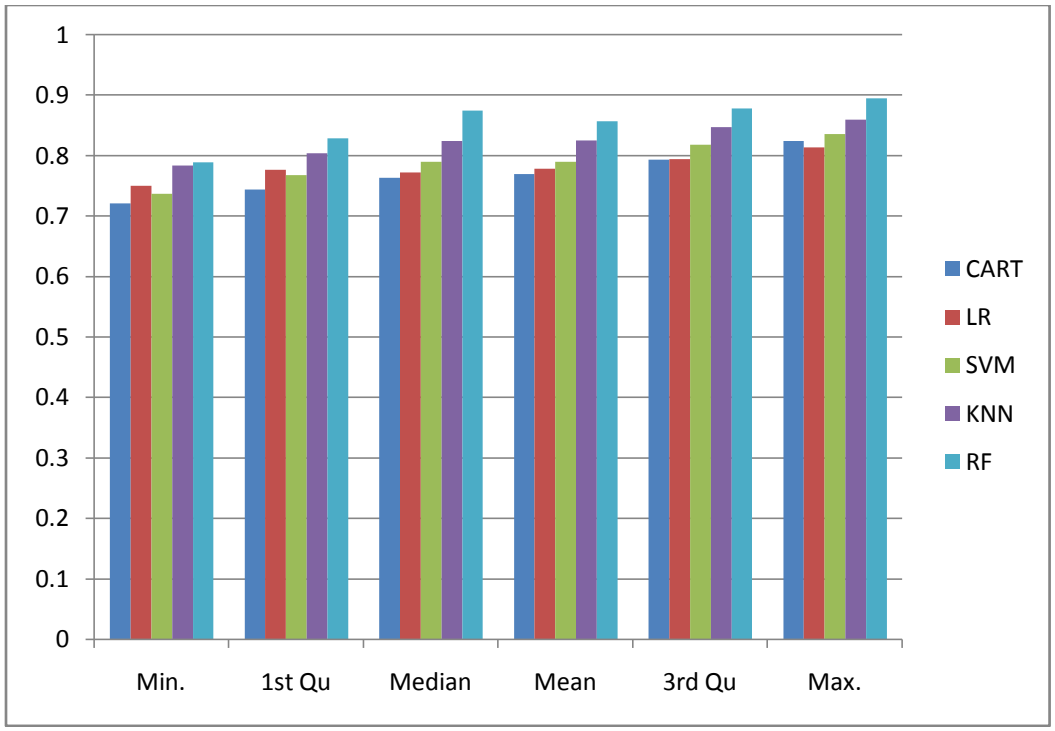
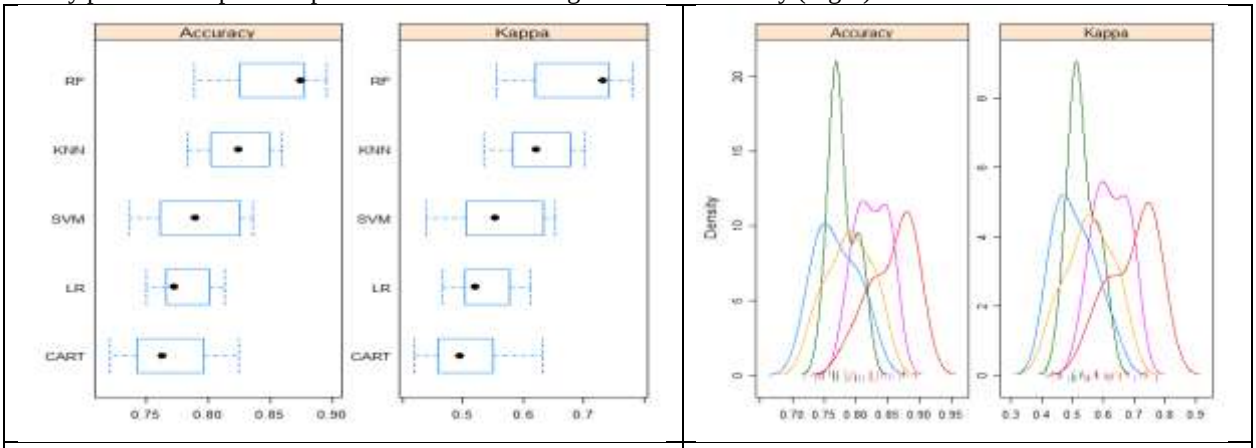


Fig 6:- Performance based on summary statistics.

The Figure 5 showed that the mean of accuracy in case of Random Forest algorithm is higher in case of predicting rainfall data where the mean accuracy in case of CART is lower. We also consider Box and Whisker Plots and density plots to compare the performance of our using model in this study (Fig 6).



Box and Whisker plots | Density plots

Fig 7:- Box and whisker and Density plots.

From the Fig 7, the boxes from Box and Whisker plot the box are ordered from highest to lowest mean accuracy which indicates that Random Forest algorithm shows highest predicting performance in case of rainfall data. So, if we order the models according to the predicting performance we can show RF is the first performing model, KNN is the second performing model, SVM is the third performing model, LR is the fourth performing model and CART is

the last performing model for predicting rainfall status in case of Boguradistric, Bangladesh. The density plot also shows the similar performance.

Conclusions:-

Rainfall is one of the most important factors for agriculture production which directly contribute the economic development of a country and this rainfall is also influenced by different factors such as temperature, humidity, wind speed, sunshine, minimum temperature and maximum temperature.So, it is always challenging task for researchers to predict rainfall data efficiently. There are several methods and techniques for rainfall predicting but still it is hard to find unique model for all of the geographic location. Recently machine learning algorithm successfully predict the rainfall data by extracting information and using the hidden knowledge from past weather data. In this paper the performance of Classification and Regression Trees (CART), Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (K-NN) and Random Forest (RF) for predicting rainfall data in case of Bogura districts Bangladesh for the time period January, 1971 to December, 2015 on weekly basis.The empirical results based on confusion matrix as well as cross validation approach suggest that the Random Forest (RF) algorithm is the most suitable algorithm for predicting rainfall in case of Bogura district, Bangladesh for the following time period. This study will give prior awareness to the policy maker to tackle the water management system of a particular region which ensures robust agriculture production.

Acknowledgement:-

This research is funded by Faculty of Science, University of Rajshahi research project (No. A 163/5/52/RU/science-21/2021-2022)

Reference:-

1. Afonso R.F., Hachul H., Kozasa E.H., Oliveira D.D.S., Goto V., Rodrigues D., Tufik S., Leite J.R.(2012): Yoga decreases insomnia in postmenopausal women. Menopaus,19:186–193. doi: 10.1097/gme.0b013e318228225f
2. Aftab, S., Ahmad, M., Hameed, N., Bashir, M. S., Ali, I., and Nawaz, Z. (2018): Rainfall Prediction in Lahore City using Data Mining Techniques.International Journal of Advanced Computer Science and Applications, Vol. 9, No. 4, 254-260.
3. Aftab, S., Ahmad, M., Hameed, N., Bashir, M.S., Ali, I. and Nawaz, Z. (2018): Rainfall Prediction using Data Mining Techniques: A Systematic Literature Review. International Journal of Advanced Computer Science and Applications,Vol. 9, No. 5, 143-150.
4. Ahmad, M., Aftab, S. and Muhammad, S. S. (2017): Machine Learning Techniques for Sentiment Analysis: A Review. Int. J. Multidiscip. Sci. Eng., vol. 8, no. 3, pp. 27-32.
5. Badhiye, S.S., Wakode, B.V. and Chatur, P.N. (2012): Analysis of temperature and humidity data for future value prediction.International Journal of Computer Science and information Technologies, Vol. 3(1), 3012-3014.
6. Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984): Classification and Regression Trees, Chapman & Hall/CRC Press, Boca Raton, FL, 1984.
7. Chau, K. W. and Wu, C. L.(2010): A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. J. Hydroinformatics, vol. 12, no. 4, p. 458.
8. Cramer, S. Kampouridis, M., Freitas, A. A. andAlexandridis, A. K. (2017): An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. Expert Syst. Appl., vol. 85, pp. 169–181.
9. Gupta, D and Ghose, U. (2015): A Comparative Study of Classification Algorithms for Forecasting Rainfall. Environmental Science, Computer Science, 4th International Conference on Reliability, Infocom Technologies and Optimization, Conference Proceeding, pp. 0–5, 2015.
10. Harrison, O. (2018): Machine learning basics wih k-Nearest Neighbors algorithm. Towards Data Science (https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761)
11. Kannan, M., Prabhakaran, S. and Ramachandran, P. (2010): Rainfall Forecasting Using Data Mining Technique. International Journal of Engineering and Technology, Vol. 2 (6), 397-401.
12. Mbaabu, O. (2020): Introduction to Random Forest in Machine Learning. Section's Engineering Education (EngEd) Community. https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/

13. Mishra, N., Soni, H. K., Sharma, S. and Upadhyay, A. K. (2018): Development and Analysis of Artificial Neural Network Models for Rainfall Prediction by Using Time-Series Data.Int. J. Intell. Syst. Appl., vol. 10, no. 1, pp. 16–23.
14. Monira, S. S., Faisal, Z. M. and Hirose, H. (2010): Comparison of artificially intelligent methods in short term rainfall forecast, Proc. 2010 13th Int. Conf. Comput. Inf. Technol. ICCIT 2010, no.Iccit, pp. 39–44, 2010.
15. Olaiya, F. and Adeyemo, A.B. (2012): Application of Data Mining Techniques in Weather Prediction and Climate Change Studies.I.J. Information Engineering and Electronic Business, 2012, 1, 51-59.
16. Rahman, M.M., Khalek, M.A. and Rahman, M.S. (2021): Performance of different data mining methods for predicting rainfall of Rajshahi district, Bangladesh. Data Science and SDGs, Springer Nature Singapore Pte Ltd.
17. Rahman, M.M. and Rahman, M.S. (2020a): Monitoring Drought Vulnerability at Different Time Scales: A Case Study from Rajshahi District, Bangladesh. International Journal of Advanced Research, 8(11), 404-414.
18. Rahman, M.M. and Rahman, M.S. (2020b): Comparison the drought Probability at Different Decade from Northwestern area of Bangladesh. International Journal of Science and Research. Vol. 9, Issue 11, 1252-1259.
19. Rahman, M.M. and Rahman, M.S. (2021): Impact of threshold value for detecting Drought index: A case study from Pabna district of Bangladesh. International Journal of Science and Research. Vol. 10, Issue 2, 706-717.
20. Sivapragasam, C., Liong, S. and Pasha, M.(2001): Rainfall and runoff forecasting with SSA-SVM approach, J. Hydroinformatics, no. April 2016, pp. 141–152.
21. Sethi, N. and Garg, D.K. (2014): Exploiting Data Mining Technique for Rainfall Prediction. International Journal of Computer Science and Information Technologies, Vol. 5 (3), 3982-3984.
22. Talib, M.R., Ullah, T., Sarwar, M.U., Hanif, M.K. and Ayub, N. (2017): Application of Data Mining Techniques in Weather Data Analysis. International Journal of Computer Science and Network Security, Vol. 17, No. 6, 22-28.
23. Tharun, V.P., Prakash, P. and Devi, S.R. (2018): Prediction of Rainfall Using Data Mining Techniques. Proceeding of the 2 nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Eplore Compliant-Part Number: CFP18BAC-ART: ISBN: 978-1-5386-1974-2.
24. Wu, J., Long, J. and Liu, M. (2015): Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm. Neurocomputing, vol. 148, pp. 136–142.
25. Yadav, R.K. and Khatri, R. (2016): A Weather Forecasting Model using the Data Mining Technique. International Journal of Computer Applications, 139(14):4-12.
26. Zainudin, S., Jasim, D. S. and Bakar, A. A. (2016): Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction, Int. J. Adv. Sci. Eng. Inf. Technol., vol. 6, no. 6, pp. 1148–1153.
27. Zaw, W.T. and Naing, T.T.(2008): Empirical Statistical Modeling ofRainfallPrediction over Myanmar.International Journal of Computer and Information Engineering, Vol.2, N0. 10. 3418-3421.
28. Banglapedia (2021): Online linkhttps://en.banglapedia.org/index.php/Bogra_District#:~:text= Bogra%20District%28rajshahi%20division%29%20area%202898.25%20sq%20km%2C%20located,and%20in %20between%2088%C2%B058%27%20and%2089%C2%B045%27%20east%20longitudes.
29. Kanade, V. (2022): What is Logistic Regression? Equation, Assumptions, Types and Best Practices, TOOLBOX/Tech, https://www.toolbox.com/tech/artificial-intelligence/articles/what-is-logistic-regression/
30. Gandhi, R. (2018): Support Vector Machine-Introduction to Machine Learning Algorithms. Towards Data Science.https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.
31. Vapnik, V. (1963): Pattern recognition using generalized portrait method. Autom Remote Control.24:774–780.