International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 11, Issue 10, October 2020, pp. 1582-1591, Article ID: IJARET_11_10_151
Available online at
http://iaeme.com/Home/issue/IJARET?Volume=11&Issue=10
ISSN Print: 0976-6480 and ISSN Online: 0976-6499
DOI: 10.34218/IJARET.11.10.2020.151

COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR RAINFALL PREDICTION – A CASE STUDY IN NEPAL

Nawaraj Paudel and Tekendra Nath Yogi
Central Department of Computer Science and Information Technology,
Tribhuvan University, Kathmandu, Nepal

ABSTRACT

One bitter truth associated with Nepal is it one of the richest country in fresh water resources in the world and also an agriculture dominant country. However, it largely depends on monsoon for irrigation purpose. Its economy relies on farming of the nation. Due to the fact that its farming is based on natural monsoon, it is necessary for the farmers to know about when a rainfall occurs. In addition, rainfall information also helps people to become aware of any likely disaster which can cause damage to lives and property. This research carried the comparative study on the performance of Random Forest, Decision Tree, and Support Vector Machine classification algorithms for the prediction of rainfall on the dataset recorded by various weather stations of Nepal. The performance of these three classification algorithms were assessed using accuracy, precision, recall and F-measure. The obtained prediction result based on the performance measures showed that the Random Forest outperformed all the competing algorithms as it had highest accuracy value of 80.56 %, precision value of 74.50%, recall value of 76.50% and F- measure value of 75.50%.

Key words: Classification, Decision Tree, Rainfall prediction, Random Forest, Support Vector Machine.

Cite this Article: Nawaraj Paudel and Tekendra Nath Yogi, Comparative Study of Machine Learning Algorithms for Rainfall Prediction – A Case Study in Nepal, *International Journal of Advanced Research in Engineering and Technology,* 11(10), 2020, pp. 1582-1591.
http://iaeme.com/Home/issue/IJARET?Volume=11&Issue=10

1. INTRODUCTION

Weather forecasting is the task of predicting the state of the atmosphere at a future time and a specified location [1]. Rainfall prediction is a type of weather forecasting. Rainfall patterns around the world change rapidly and continuously in the recent years. Correct rainfall prediction is necessary in today's daily life. It has huge influence in the agriculture sector as well as in natural disaster management. Cultivation of crops mainly depends on rainfall and it is important to predict whether it will rain or not. A good rainfall prediction can help farmers to make decision about what to cultivate and what to not. Prediction of rainfall helps not only the farmers it also helps almost all peoples, being alert toward any disaster which can cause damage to life and property. Along with the agriculture and disaster management, rainfall also sometimes affects many other sectors such as tourism, transportation, construction, air traffic services, hydropower etc. so that a good rainfall prediction can be helpful in each and every aspect of the daily life.

From ancient time people are trying to find the pattern of rainfall and predict rainfall for their well-being. From the very beginning of science and technology, rainfall prediction is a very interesting field of study. Predicting rainfall is a tough and challenging task due to the complication of the physics and different parameters which cause rainfall. It is actually a very noisy and deterministically disordered natural event. Some very important factors such as cyclone, monsoon, wind, moisture, temperature, rotation of earth etc. plays a vital role on the rainfall [2].

Traditionally, rainfall predictions are performed with the help of large complex models of physics, which utilize different atmospheric conditions such as atmospheric pressure, temperature, wind direction, precipitation, humidity etc. over a long period of time. These conditions are often unsteady because of perturbations of the rainfall system, causing the models to provide not exact predictions. The models are generally run on hundreds of nodes in a large high-performance computing environment which consumes a large amount of electrical energy. Despite using these costly and complex resource intensive devices, there are often inaccurate rainfall predictions because of incorrect initial measurements of the conditions or an incomplete understanding of atmospheric processes. Moreover, it generally takes a long time to solve composite models like these [1] [2].

We can use different machine learning techniques to produce accurate results for rainfall prediction. We can just do it by having the historical data analysis of rainfall and can predict the rainfall for future seasons. For this, we can apply many techniques like classification, regression according to the requirements and also we can calculate the error between the actual and prediction and also the accuracy. Different techniques produce different accuracies so it is important to choose the right algorithm and model it according to the requirements.

This study focuses on rainfall prediction based on weather datasets recorded by various weather stations of Nepal using three classification algorithms Decision Tree, Random Forest, and Support Vector Machine. To improve the processing efficiency and rainfall predication performance the data has been preprocessed, such as handling missing data before fed as input to the classification algorithms. The classification algorithms have been evaluated based four performance evaluation parameters accuracy, precision, recall and F-measure. Finally, the comparison of the three classification algorithms has been made in order to decide the better algorithm for rainfall prediction purpose.

2. RELATED WORKS

Rainfall prediction is one of the interesting and challenging fields in the science and technology. It affects the day to day activity of each and every people. Even a single miss prediction of rainfall may cause loss of life and property. Therefore, this field attracts researchers to investigate the facts related to the rainfall prediction to make the prediction accurate and more reliable.

In recent years, various machine learning techniques have been applied in rainfall prediction around the world because the modern machine learning based forecasting is more accurate and faster on large data set. A linear regression model and a variation on a functional regression model were used [1], with the latter able to capture trends in the weather using maximum and the minimum temperature for seven days, given weather data for the past two days. Both of our models were outperformed by professional weather forecasting services.

Based on the historical data from surrounding area of a particular area Nashville in Tennessee, USA, the researchers found that machine learning techniques are simpler and efficient to predict weather [2]. They also showed that it is more effective than considering only the area for which rainfall prediction is done.

Three machine learning algorithms Linear Regression, Support Vector Regression and Multilayer Perceptron were compared using mean squared error, correlation coefficient, coefficient of efficiency and mean absolute error, on rainfall dataset in Odisha state of India. This study showed that the Support Vector Regression outperformed than two other compared algorithms [3]. Classification and Regression Tree algorithm, Naïve Bayes approach, K nearest Neighbor, and 5-10-1 Pattern Recognition Neural Network classification algorithms were compared using dataset of 2245 samples of New Delhi from June to September from 1996 to 2014 collected from Weather Underground [4]. This study showed that 5-10-1 Pattern Recognition Neural Network provided better result than any of the other discussed algorithms.

On meteorological data collected between 2000 and 2009 from the city of Ibadan, Nigeria, a Comparative analysis of Decision Tree variants and Artificial Neural Network variants were performed by using Mean Squared Error and Correlation Coefficient as an evaluation parameter [5]. It was found that the C5 decision tree outperformed than other techniques. It was also concluded that, on large data set, data mining techniques are better for weather prediction.

To be familiar with appropriateness of Neural Network in climate prediction and spatial interpolation, a comprehensive literature review of past 50 years was done and it was established that Neural Network such as Backpropagation Network, Radial Basis Function were best appropriate to predict chaotic behavior of climate variables like rainfall, rainfall runoff, and have efficient enough for prediction in long period [6].

Monthly meteorological data by Central Bureau of Statistics Sudan from 2000 to 2012 for 24 meteorological stations has been used to find the relationship of rainfall in Sudan with important parameters such as Station, Wind Direction, Date, Humidity, Min-Temperature, Max-Temperature and Wind Speed with the help of correlation and best parameters were also selected by using various feature selection methods. Finally, 14 machine learning algorithms were compared and evaluated for the model construction to perform the prediction in the future [7].

The Prediction models of heavy rain damage was developed by applying machine learning techniques such as decision trees, bagging, random forests, and boosting. As a result of evaluating the prediction performance of each model, the Area under the Curve value of the boosting model using meteorological data from the past 1 to 4 days was the highest at 95.87% and was selected as the final model. Prediction models of heavy rain damage were developed for the Seoul Capital Area in the Republic of Korea from 1994 to 2015 [8].

The applicability of random forest, stochastic gradient boosted model and extreme learning machine methods to quantitative precipitation estimation models was investigated using case studies with polarization radar data from Gwangdeoksan radar station. Various combinations of input variable sets were tested, and results showed that machine learning algorithms can be applied to build the quantitative precipitation estimation model of the polarization radar data in South Korea. The machine learning-based quantitative precipitation estimation models led to better performances than Reflectivity and rainfall rate relationship-based models, particularly for heavy rainfall events. The extreme learning machine is considered the best of the algorithms used based on Root Mead Square Error (RMSE) evaluation criteria [9].

To construct the different classification model, the five classification algorithms Naïve Bayes, Decision Tree, Support Vector Machine, Neural Network and Random Forest was applied on Malaysian data [10]. The dataset was collected from multiple stations in Selangor, Malaysia. The models constructed were evaluated based on different evaluation criteria such as accuracy, precision, recall and F-measure, in order to determine the most appropriate technique. The experimental results showed that for Rainfall prediction, Decision Tree, and Random Forest performed well.

Weather prediction is complex due to noise and missing values dataset. A noise tolerant, efficient algorithm WP-MKNN (Weather Prediction using Modified kNN) had been proposed and found that the proposed algorithms is better than KNN and MKNN on the data set collected from National Climatic Data Center of Pakistan. The experimental results showed that the proposed technique is more accurate even in noisy dataset [11].

Accuracy of prediction is important in rainfall prediction. A Backpropagation Neural Network (BPNN) algorithm has been used to model and predict rainfall in Tenggarong, East Kalimantan, Indonesia. After testing the three architectures with different epochs 500, 1000 and 1500, then the best MSE value obtained was 0.00096341, with 2-50-20-1 architecture and epochs 1000. The results of this study have showed that BPNN models was suggested as a predictive algorithm that provides a good predictive accuracy [12].

Two hybrid neural models Multilayer Perceptron (MLP) and Radial Basis Function (RBF) were proposed to enhance the accuracy of weather forecasting based on the weather data of Saudi Arabia. Correlation coefficient, RMSE and scatter index are the standard yard sticks adopted for forecast accuracy measurement. On individual standing MLP forecasting results were better than RBF, however, the proposed simplified hybrid neural model came out with better forecasting accuracy as compared to both individual networks [13].

To perform the accurate and reliable rainfall prediction, a new methodology was proposed and compared with the existing Support Vector Regression. The result of the experiment done with the data of Chittagong, Bangladesh showed that the proposed model outperformed than regular technique used. The proposed model predicted better than any model before one day. The 7 days ahead model also performed very well than conventional processes [14].

Data mining techniques were investigated and compared in forecasting different atmospheric phenomena especially atmospheric dust using Decision Tree, k-NN and Naïve biased algorithms [15]. Each model was evaluating based on the dataset from Cairo Airport, Egypt. The experimental results showed that Decision tree outperformed better than others.

A statistical downscaling approach for improving extreme rainfall simulation was proposed to predict the daily rainfalls at Shih-Men Reservoir catchment in northern Taiwan [16]. The structure of the proposed downscaling approach is composed of two parts: the rainfall-state classification and the regression for rainfall-amount prediction. Predictors of classification and regression methods were selected from the large-scale climate variables of the NCEP reanalysis data based on statistical tests. The data during 1964–1999 and 2000–2013 were used for calibration and validation, respectively. Three classification methods, including linear discriminant analysis (LDA), random forest (RF), and support vector classification (SVC), were adopted for rainfall-state classification and their performances were compared. After rainfall-state classification, the least square support vector regression (LS-SVR) was used for rainfall-amount prediction for different rainfall states. Two rainfall states (i.e., dry day and wet day) and three rainfall states (dry day, non-extreme-rainfall day, and extreme-rainfall day) were defined and compared for judging their downscaling performances. The results showed that RF outperformed LDA and SVC for rainfall-state classification. Using RF for three-rainfall-states classification and LS-SVR for rainfall-amount prediction can improve the extreme rainfall downscaling.

In this research, we have compared three machine learning classification algorithms including Decision tree, Random Forest, and support Vector Machine to predict rainfall on rainfall data collected by various weather stations of Nepal. These three algorithms have been compared by using the methodology mentioned in the subsequent section.

3. METHODOLOGY

The rainfall data from different areas of Nepal has been used to test train and test three different algorithms including Support Vector Machine, Decision Tree and Random Forest after preprocessing the collected data from http://data.opennepal.net. These three algorithms have also been compared using the evaluation matrices accuracy, precision, recall, and F-measure.

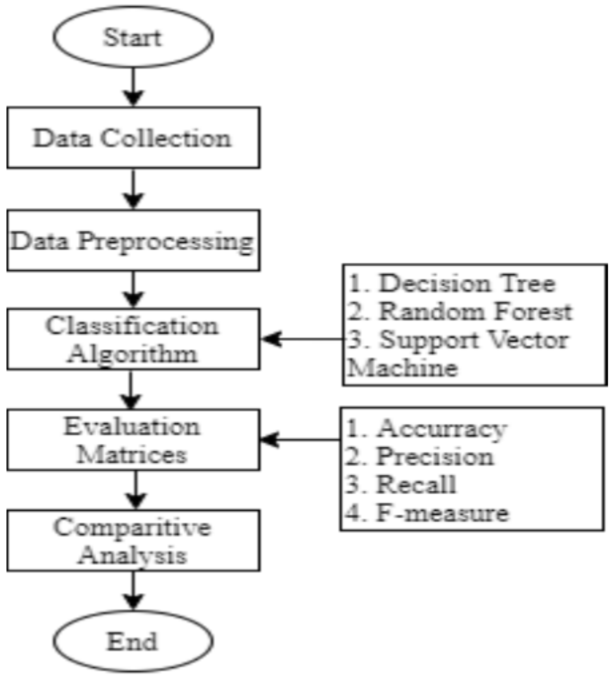

Figure 1 Methodology

3.1. Data Collection

The data set used in this research was collected from the http://data.opennepal.net/, a web site with a collection of data in the various sectors in Nepal. This dataset contains about 2 years, for the years 2012-2013, of daily rainfall observations from numerous weather stations of Nepal including Dadeldhura, Dipayal, Dhangadi, Birendranagar, Nepalganj, Jumla, Dang, Pokhara, Bhairahawa, Simara, Kathmandu, Okhaldhunga, Taplejung, Dhankuta, Biratnagar, and Jiri. The dataset contains 5 attributes date, stations, Minimum temperature, Maximum temperature and Rainfall, the first four attributes measures the daily Rainfall associated attribute data and last 5th attribute Rainfall is the target variable with many possible values measures the amount of rainfall on a particular day. The value of Rainfall greater than zero represents it will rain tomorrow and the zero value represents it will not rain tomorrow. The temperature data are in degree Celsius and rainfall is in mm. This data set has 10756 weather observations with their corresponding Rainfall amount. In the dataset the rainfall data are as, recording ended at 05:45 PM on a particular day and * means ending at 08:45 AM that day.

3.2. Tools and Languages used

The experiment conducted in this research has been done in python programming language by using Anaconda IDE.

3.3. Data Preprocessing

The rainfall dataset collected for this research was noisy, dirty and not ready for direct use. In order to make the noisy and dirty data clean and ready for the use some preprocessing has been done. First of all, in the rainfall dataset some attribute values are missing, such as the Maximum temperature 354 attribute has total missing values, Minimum temperature attribute has total 311 missing values and Rainfall attribute has total 814 missing values, the observations with missing data were eliminated from the data set. After removing observations with missing data only 9497 observations remained in the dataset. The date and stations attributes present in the dataset weren't considered in this research so, these two attribute columns has been dropped from the data set. The rainfall attribute has '*' symbol in some measured values, this special symbol was also removed from the dataset. Finally, the rainfall attribute has been converted into categorical attribute with yes and no values from the numeric attribute as, if the amount of rainfall is greater than zero then the rainfall attribute value is yes otherwise no. The value yes of Rainfall attribute represents it will rain tomorrow and the no value represents it will not rain tomorrow. After preprocessing, the dataset contains only 9497 total observations, with 2847 observations labeled yes category and 6650 observations labeled no category. This distribution of the data is as shown in figure below. This dataset is now ready to be used with above mentioned algorithms.

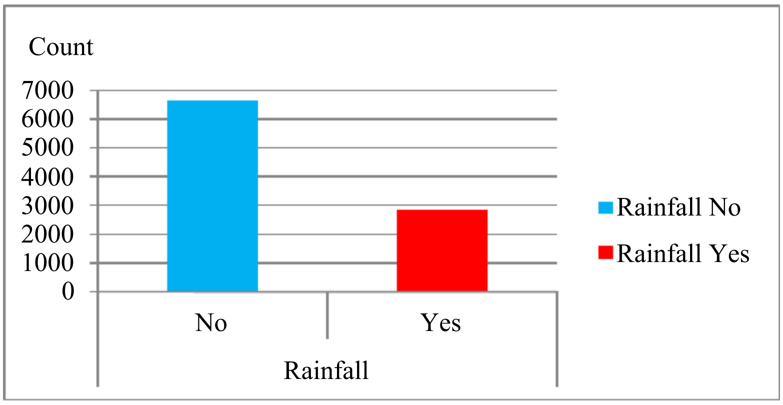

Figure 2 Distribution of data in each category of Rainfall

3.4. Algorithms Used

In this research, three classification algorithms mentioned before have been used to train and test based on the dataset.

3.4.1. Decision Tree

From the training data set a decision tree has been created as model by using the steps mentioned below.

Generate_Decision_Tree(Sample S, Attribut_list A)

1. create a node N
2. If all samples are of the same class C then label N with C; terminate;
3. If A is empty then label N with the most common class C in S (majority voting); terminate;

4. Select $a \in A$, with the highest information gain [17]; Label N with a;
5. For each value v of a:
   a. Grow a branch from N with condition a=v;
   b. Let S_v be the subset of samples in S with a=v;
   c. If S_v is empty then attach a leaf labeled with the most common class in S;
   d. Else attach the node generated by Generate_Decision_Tree (S_v, A-a)

To classify a new data sample based on the decision tree created by using above steps, the attributes of a new data sample are tested against the decision tree. A path is traced from the root to a leaf node which holds the prediction for that new sample to be classified [17].

3.4.2. Random Forest

Random Forest (RF) is an ensemble learning method used mainly for classification. In Ensemble learning multiple models are used to solve the same problem. In ensemble classification, multiple classifiers are used and are more accurate than the individual classifiers in the ensemble. A voting scheme is then used to determine the class label for unlabeled instances. A simple and yet effective voting scheme is majority voting. In majority voting, each classifier in the ensemble is asked to predict the class label of the instance being considered. Once all the classifiers have been queried, the class that receives the greatest number of votes is returned as the final decision of the ensemble. RF constructs a collection of decision trees with controlled variation. Using bagging, each decision tree in the ensemble is constructed using a sample with replacement from the training data. Each tree in the ensemble acts as a base classifier to determine the class label of an unlabeled instance. This is done via majority voting where each classifier casts one vote for its predicted class label, and then the class label with the most votes is used to classify the instance [18] [19].

3.4.3. Support Vector Machines

A Support Vector Machine (SVM) performs classification by finding the hyper plane (classifier) that maximizes the margin between the two classes subject to the constraint that all the training observations should be correctly classified. Hyper plane is defined by using the data points that are closest to the boundary. These points are called support vectors and the decision boundary itself is called support vector machine. The main advantage of SVM classifier is that it minimizes the training set error and the test set error. To obtain a SVM classifier with the best generalization performance, appropriate training is required. The most commonly used and popular algorithm for training SVM is the SMO algorithm. The main advantage of SMO algorithm is that it works analytically on a fixed size working set by decomposing the large training data set. So, that it can works fine even for large data sets and it also gives superb performances in almost all kinds of training data sets [20] [21].

3.5. Evaluation matrices

The comparative analysis of three classification algorithms for rainfall prediction has been made by measuring the performance of each algorithm with the help of following parameters accuracy, precession, recall, and F-measure [22].

Accuracy of classification algorithm for rainfall prediction on given data dataset is the percentage of observations in a rainfall dataset that are correctly classified as rainfall yes or rainfall no.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$$

Where, True Positive (TP) is the number of observations with label yes that are correctly labeled by the algorithm. True Negative (TN) is the numbers of observations with label no, that are correctly labeled by algorithm. False Positive (FP) is the number of observations with label no that are incorrectly labeled as yes. False Negative (FN) is the number of observations with label yes that are mislabeled as no.

Precision refers to the measure of exactness that means what percentage of observations labeled as positive sentiment category are actually such.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall refers to the true positive or rainfall label yes that means the proportion of Rainfall label yes observations that are correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN}$$

The F-measure combines both measures precision and recall as the harmonic mean.

$$\text{F-measure} = \frac{2 \times precision \times recall}{pression + recall}$$

4. EXPERIMENTS AND RESULTS

The three classification algorithms have been analyzed for rainfall prediction based on the rainfall dataset that is recorded by various weather stations of Nepal. The rainfall dataset had 10756 observations in total, but after preprocessing only 9497 observations among them 2847 belonged to yes rainfall category and 6650 to no rainfall category as shown in figure 2. For training only 7597 (5299 no category and 2298 yes category) observations were taken and remaining 1900 (1351 no category and 549 yes category) were taken for testing purpose. Table 1 below shows confusion matrix of the classification report that has been obtained after testing three algorithms on test dataset.

Table 1 Confusion Matrix

Predicted/ Actual	Decision Tree		Random Forest		Support Vector Machine	
	Yes	No	Yes	No	Yes	No
Yes	323	226	257	292	328	221
No	165	1186	125	1226	148	1203

Based on the classification report shown in Table 1 the calculated summary performance result for the comparison of all three algorithms applied on weather dataset is shown in Table 2. The accuracy, precision, recall and F-measure value are shown in Table 2 is the average of precision, recall and F-measure for both Rainfall categories.

Table 2 Performance Result

Algorithms	Accuracy	Precision	Recall	F-Measure
Random Forest	80.58%	74.50%	76.50%	75.50%
Decision Tree	79.42%	73.5%	75.0%	74.0%
SVM	78.05%	69.0%	74.0%	70.0%

It is clearly seen that the accuracy value of Random forest is got high level of 80.58% and SVM got less accuracy of level 78.05%. In case of precision Random forest had also got high precision level of 74.50% and SVM got less precision level of 69.0%. The Recall value of Random Forest was high with 76.50%. Whereas, SVM got less recall level of 74.0 % respectively. Finally, the F-measure of Random Forest had outperformed two other compared algorithms with value of 75.50% and SVM had got minimum value of 70.0%.

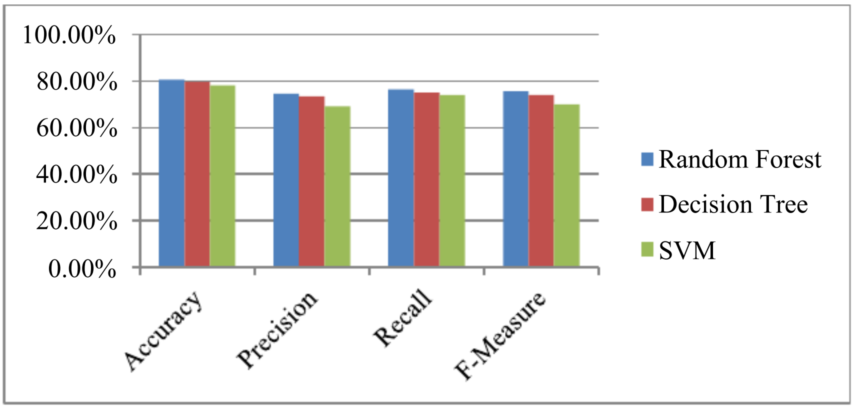

Figure 3 Graph of table 2

5. CONCLUSION AND RECOMMENDATIONS

In this study, Random Forest, Decision Tree, and Support Vector Machine classification algorithms have been compared on the rainfall data recorded by the various weather stations of Nepal. The final result showed that Random Forest algorithm was superior and more accurate in predicting rainfall than other competing models. This algorithm had highest accuracy, precision, recall and F-Measure score of 80.58 %, 74.50%, 76.50%, and 75.50% respectively whereas, SVM has low accuracy, precision, recall and F-Measure score of 78.05%, 69.0%, 74.0%, 70.0% respectively. With the method demonstrating good prediction accuracy, it was concluded that the Random Forest technique has high potential to be used to predict rainfall. On label unbalanced, small sized, low dimensional, numeric rainfall dataset Random Forest has better performance among other two.

This work can be enhanced for large, high dimension, numeric as well as image data associated with weather in order to predict the various weather associated parameter such as temperature, rainfall, snowfall, cyclone, etc. Different machine learning classification algorithm as well as for more efficiency on large data set deep learning approach can also be used.

REFERENCES

[1] M. Holmstrom, D. Liu and C. Vo, Machine Learning Applied to Weather Forecasting, Stanford University, 2016.

[2] A. H. M. Jakaria, M. M. Hossain and M. A. Rahman, Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee, ACM Mid-Southeast Conference, Gatlinburg, Tennessee, USA, 2018.

[3] X. Zhang, S. N. Mohanty, A. K. Parida, S. K. Pani, B. Dong and X. Cheng, Annual and Non-Monsoon Rainfall Prediction Modelling Using SVR-MLP: An Empirical Study From Odisha, IEEE Access, 8, 2020, 30223-30233.

[4] D. Gupta and U. Ghose, A comparative study of classification algorithms for forecasting rainfall Noida, 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, 2015.

[5] F. Olaiya and A. B. Adeyemo, Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, International Journal of Information Engineering and Electronic Business(IJIEEB), 4(1), 2012, 51-59.

[6] S. Karmakar, S. Choubey and P. Mishra, Appropriateness of Neural Networks in Climate Prediction and Interpolations: A Comprehensive Literature Review, International Journal of Applied Information Systems, 10(10), 2016, 33-54.

[7] N. O. Bushara and A. Abraham, "Weather Forecasting in Sudan Using Machine Learning Scheme, Journal of Networking and Innovative Computing, 2, 2014, 309-317.

[8] C. Choi, J. Kim, J. Kim, D. Kim, Y. Bae and H. S. Kim, Development of Heavy Rain Damage Prediction Model Using Machine Learning Based on Big Data, Advances in Meteorology, 2018.

[9] J.-Y. Shin, Y. Ro, J.-W. Cha, K.-R. Kim and J.-C. Ha, Assessing the Applicability of Random Forest, Stochastic Gradient Boosted Model, and Extreme Learning Machine Methods to the Quantitative Precipitation Estimation of the Radar Data: A Case Study to Gwangdeoksan Radar, South Korea, in 2018, Advances in Meteorology, 2019.

[10] S. Zainudin, D. S. Jasim and A. A. Bakar, Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction, International Journal on Advanced Science, Engineering and Information Technology, 6(6), 2016, 1148-1153.

[11] M. Abrar, A. T. H. Sim, D. Shah, S. Khusro and Abdusalam, WEATHER PREDICTION USING CLASSIFICATION, Science International (Lahore), 26(5), 2014, 2217-2223.

[12] Mislan, Haviluddin, S. Hardwinarto, Sumaryono and M. Aipassa, Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan - Indonesia, International Conference on Computer Science and Computational Intelligence, Procedia Computer Science , 2015.

[13] T. Saba, A. Rehman and J. S. AlGhamdi, Weather forecasting based on hybrid neural model, Appl Water Sci, 7, 2017, 3869–3874.

[14] N. Hasan, N. C. Nath and R. Islam , A support vector regression model for forecasting rainfall, 2nd International Conference on Electrical Information and Communication Technologies (EICT), Khulna, 2015.

[15] M. F. M.Ali, S. A. Asklany, M. A. El-wahab and M. A. Hassan, Data Mining Algorithms for Weather Forecast Phenomena: Comparative Study, International Journal of Computer Science and Network Security, 19(9), 2019, 76-81.

[16] Q. B. Pham, T.-C. Yang, C.-M. Kuo, H.-W. Tseng and P.-S. Yu, Combing Random Forest and Least Square Support Vector Regression for Improving Extreme Rainfall Downscaling, Water, 11(3), 2019.

[17] J. Han, M. Kamber and J. Pei, Classification: Basic Concepts and Advanced Methods," Data Mining: Concepts and Techniques, 3rd ed., Waltham, Massachusetts: Morgan Kaufmann Publishers, 2012, 327-442.

[18] K. Kirasich, T. Smith and B. Sadler, Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets, SMU Data Science Review, 1(3), 2018.

[19] K. Fawagreh, M. M. Gaber and E. Elyan, Random forests: from early developments to recent advancements, Systems Science & Control Engineering: An Open Access Journal, 2(1), 2014, 602-609.

[20] J. Platt, Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, Technical Report MSR-TR-98-14, Microsoft Research, 1998.

[21] K. P. Bennett and C. Campbell, Support vector machines: hype or hallelujah?, SIGKDD Explorations, 2(2), 2000, 1-13.

[22] G. KOU, Y. LU, Y. PENG and Y. SHI, EVALUATION OF CLASSIFICATION ALGORITHMS USING MCDM AND RANK CORRELATION, International Journal of Information Technology & Decision Making, 11(1), 2012, 197-225.