IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

ENHANCING THE ACCURACY OF MUMBAI RAINFALL PREDICTION PROCESS THROUGH MINING TECHNIQUES

Ms. B. MADHUVANTHI
Research Scholar,
PG & Research Department of Computer Science,
A. Veeriya Vandayar Memorial Sri Pushpam College (Autonomous), Poondi - 613503, Thanjavur,

Dr. T.S. BASKARAN
Associate Professor& Research Supervisor,
PG & Research Department of Computer Science,
A Veeriya Vandayar Memorial Sri Pushpam College (Autonomous), Poondi - 613503, Thanjavur,

"Affiliated to Bharathidasan University, Tiruchirappalli-620024", TamilNadu, India.

Abstract— The changing of physical characteristics of the hydrological system has cause natural phenomenon which leads to heavy rainfall. It is one of the problems of economic damages and affects people. Rainfall is one of the greatest challenges in the meteorological department. Rainfall prediction is necessary to inform people and prepare them in advance about the upcoming weather condition. Rainfall prediction involves recording the various parameters of weather like temperature, dew point, wind speed, visibility, and precipitation. It has been seen that data mining techniques have achieved good performance and accuracy in rainfall prediction. This research work aims to compare the performance of a few data mining algorithms for predicting rainfall using historical weather data of Mumbai, India, which is collected from the World Metrological Organization (WMO). From the collected weather data which comprises 11 attributes, only 5 attributes that are most relevant to rainfall prediction are considered. The data mining process model is followed to obtain accurate and correct prediction results. In this thesis, various data mining algorithms were explored which include decision tree-based J48, Random Forest, Logistic Regression, Naive Bayes and Support Vector Machine (SVM). The experimental results show that the Random Forest algorithm has a good level of accuracy than other algorithms.

Keywords— Rainfall Prediction, Data Mining, Naive Bayesian, Support Vector Machine,Logistic Regression, J48, Random Forest.

I. INTRODUCTION

Rainfall prediction has been a challenging problem in the meteorological department every year. Even after the technological and scientific advancement, the accuracy in the prediction of weather has never been sufficient. Even in the current date, this domain remains as a thesis topic in which scientists and mathematicians are working to produce a model or an algorithm that will accurately predict the weather [12]. There have been immense improvements in the sensors that are responsible for recording the data from the environment and cancel the noise present in them; along with these new models have been proposed which include different attributes related to

weather to make an accurate prediction.

Currently one of the most widely used techniques for rainfall prediction is data mining. Data mining offers a way of analyzing data numerically and extract rules that can be used for predictions. Presently it is being used in many domains such as stock marketing organizations, retail, financial, etc. Scientists have now realized that data mining can be used as a tool for rainfall prediction as well. The basic object of data mining is the data itself. It is defined as a raw set of information that can be used to extract meaningful information depending upon the requirements of the application. Data can be stored in an organized manner [5].

The term data mining refers to the techniques that are used to extract the required information from the given set of data that might be useful for statistical purposes or making predictions by learning patterns in data and correlation between different parameters. Data mining has now been adopted by many domains such as sports, banking, meteorological department, etc., and because of this, scientists, mathematicians, and researchers have come up with a wide range of algorithms for finding the solution.

There is a huge amount of weather data available that is rich in information and can be used for rainfall prediction. Various data mining techniques are applied to the weather data to predict atmospheric parameters like temperature, wind speed, rainfall, meteorological pollution, etc. which tend to change from time to time and weather calculation varies with the geographical location along with its atmospheric parameters. Some commonly used data mining techniques for weather prediction are Naïve Bayes Algorithm, Support Vector Machine, Logistic Regression, C4.5 (J48), and Random Forest [15].

The prediction of correct weather condition especially the prediction of rainfall is very important. Rainfall is important for crop production, water resource management, humidifying the atmosphere, producing streams and rivers, replenishing the water and redistribution of freshwater in the water cycle. The occurrence of a prolonged dry period of heavy rain at the critical stages of the crop growth and development may lead to significantly reduce crop yield. The prolonged and incessant rainfall can also lead to floods that destroy crops, livestock, orchards, etc. and can also decimate housing and business infrastructure.

II. LITERATURE SURVEY

In this section, we provide a critical analysis of the existing prediction models. The researchers study different data mining and Artificial Neural Network algorithms for Rainfall Prediction and their survey analysis report in Table I. The analysis is divided into two subdivisions where one explains the models used in traditional weather prediction i.e. numerical and statistical models and the second subdivision deals with the data mining efforts made in weather prediction [3], [9].

TABLE I Survey analysis report

S. No	Year	Author	Title	Methodology	Result
1.	2013	S. Meganathan and T.R.Sivaramakrishnan	Association Rule Mining and Classifier Approach for 48-Hour Rainfall Prediction Over Cuddalore Station of East Coast of India	Apriori algorithm, association mining, classification, K* algorithm	48 hours prediction systemusing K* classification with cross-validation method 75.06% accuracy and split method 75.70% accuracy.
2.	2013	Deepak Ranjan Nayak Amitav Mahapatra Pranati Mishra	A Survey on Rainfall Prediction usingArtificial NeuralNetwork	Back Propagation Network, Radial BasisFunction Networks (RBFN), Support Vector Machine (SVM), Self- Organizing Map (SOM),	Prediction and analysis of the occurrence of Rainfall Prediction using Artificial Neural Network. It gives an MLP, BPN, RBFN, SOM and SVM are suitable to predict rainfall than other forecasting techniques such as statistical and numerical methods.
3.	2014	Mohammad Abrar, Alex Tze Hiang Sim, Dilawar Shah, Shah Khusro, Abdusalam	Weather prediction usingclassification	Linear Regression, SMOReg, MultiLayer Perceptron, RepTree, Knn, MkNN	Weather Prediction system using classification algorithms with MkNN very good results as compareto other data mining prediction techniques.
4.	2015	A.H.M. Rahmatullah Imon, Manos C Roy, S. K. Bhattacharjee	Prediction of Rainfall Using Logistic Regression	Logistic Regression	Prediction of Rainfall Using the Logistic model can correctly

					predict 95.25% when it rained and 84.48% when not rained.
5.	2015	Pijush Samui1, Venkata Ravibabu Mandla, Arun Krishna, and TarunTeja	Prediction of Rainfall Using Support Vector Machine and Relevance Vector Machine	support vector machine (SVM) and relevance vector machine (RVM)	This study of the rainfall prediction system with RVMis a more robust model than the SVM.
6.	2015	Chi-shing Calvin Cheung, Melissa Anne Hart, Mervyn R. Peart	Projection of future rainfall in hong kong using logistic regression and Generalized linear model	logistic regression, generalized linear model	Future Rainfall prediction in Hong kong in monthly rain occurrence (days) is > 0.62 and monthly rainfall volume (mm) is > 0.73 for model validation method.
7.	2016	Fahad Sheikh , S. Karthick, D. Malathi , J. S. Sudarsan and C. Arun	Analysis of Data Mining Techniques for Weather Prediction	C4.5, Naive Bayes	Weather Prediction systemwith accuracy forC4.5 was 88.2% accuracy and Naïve Bayes classifier poor performance of 54.8% accuracy.
8.	2017	R. Sukanya, K. Prabha	Comparative Analysis for Prediction of Rainfall using Data Mining Techniques withArtificial NeuralNetwork	CART, C4.5, Artificial Neural Network, Back Propagation Neural Network, Support Vector Machine	ANN improves the efficiency of Rainfall prediction by combining two or more classification algorithms that can improve the accuracy of rainfall prediction.
9.	2017	Aakash Parmar, Kinjal Mistree, Mithila Sompura	Machine Learning Techniques For	ARIMA Model, Artificial	Review work and comparison of different

			Rainfall Prediction: A Review	Neural Network, Support Vector Machine, Self-Organizing Map (SOM).	approaches and algorithms used by researchers for rainfall prediction.
10.	2017	KolluruVenkata Nagendra, Dr. Maligela Ussenaiah Assistant	A survey classification techniques usedfor rainfall forecasting	Bayesian Data mining, Decision Tree,SVM, Fuzzy logic and Genetic Algorithm	Feed-Forward Neural Network gives good performance on monthly rainfall prediction. Support Vector Machines gives better performance on yearly rainfall data. In weekly rainfall, the Naïve Bayesian gives better performance for forecasting.
11.	2017	Anik Pait, VenkatesanM	The Open Environment System: Rainfall Prediction Using Naïve Bayesian Algorithm	Naïve Bayesian Algorithm	Naïve Bayesian Classifier will predict the rainfallor no rainfall on that particular day.
12.	2018	S. Karthick, D. Malathi, C.Arun	Weather prediction analysis using the random forest algorithm	RandomForest, C4.5	The comparative analysis was made between the C4.5 Decision Tree (J48) and the Random Forest algorithm. C4.5 achieved an accuracy of 82.4% and Random Forest achieved an 87.1%accuracy proving it to be better.
13.	2018	Shabib Aftab, Munir Ahmad, Noureen	Rainfall Prediction in	Support Vector Machine,	The rainfall prediction system

		Hameed, Muhammad Salman Bashir, Iftikhar Ali, Zahid Nawaz	Lahore Cityusing Data Mining Techniques	Naïve Bayes, k Nearest Neighbor, Decision Tree, and Multilayer Perceptron	to measure the performance of used data mining techniques is analyzed in termsof precision, recall, and f- measure with various ratios of training and test data.
14.	2018	Tharun V.P, Ramya Prakash S, Renuga Devi	Prediction of Rainfall UsingData Mining Techniques	Random Forest; Regression; Statistical model; Support Vector Regression.	The prediction model developed by the RF regression technique was found out to be a better and efficientmodel compared to SVR and DT models. StatisticalModelling fails to provide good accuracy of prediction.
15.	2018	Razeef Mohd, Muheet Ahmed Butt and MajidZaman Baba	Comparative Study of Rainfall Prediction Modelling Techniques (A Case Study on Srinagar, J&K, India)	J48, Random Forest, IBk, Naive Bayesian, Bagging.	Random Forest produces best rainfall prediction results with an accuracy of 87.76% in the cross-validation method and also exhibits the highest values in Recall, F- Measure, and ROC as comparedto other classification algorithms.

III. DATA MINING PROCESS

Data mining is the process of extracting useful information from a large collection of data that was previously unknown. For extracting useful information we need to follow the data mining process model that will give us a clean valuable dataset for model computation and better prediction. Very rarely data are available in the form required by the data mining algorithms. Most of the data mining algorithms would require data to be structured in a tabular format with records in rows and attributes in columns. The methodological discovery of useful relationships and patterns in data is enabled by a set of iterative activities known as the data mining process. Not all discovered patterns lead to knowledge. It is up to the practitioner to invalidate irrelevant patterns and identify meaningful information.

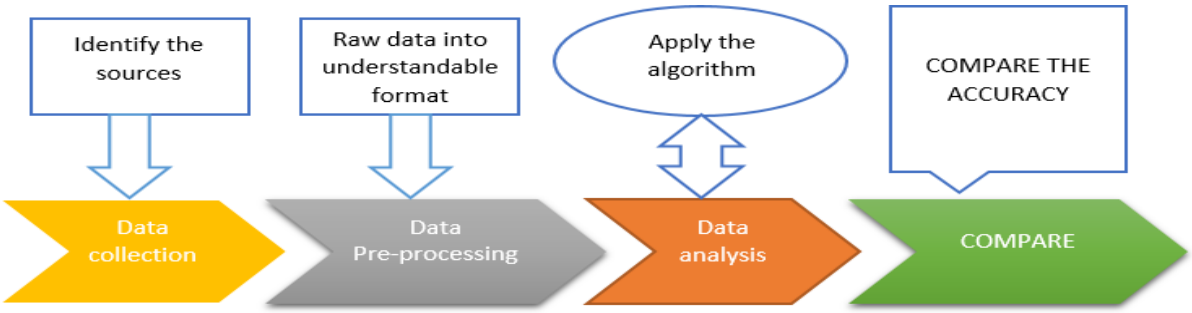

Figure 1 Data mining process

a. Data collection

Mumbai (Latitude 19°4'22.19" N / Longitude 72°52'57" E) is a coastal station in Maharashtra located on the western coast of India. This is taken as a test site. This observatory is maintained by World Metrological Organization since long and data pertaining to 2012-2018 were used for analysis. For the weather parameters temperature, dew point, wind speed, visibility and precipitation (rainfall) were considered for analysis. Mumbai (Latitude 19°4'22.19" N / Longitude 72°52'57" E) is a coastal station in Maharashtra located on the western coast of India. This is taken as a test site. This observatory is maintained by World Metrological Organization since long and data pertaining to 2012-2018 were used for analysis. For the weather parameters temperature, dew point, wind speed, visibility andprecipitation (rainfall) were considered for analysis [5].

Figure 2 Study area

b. Pre – processing

Data pre-processing steps were applied to the new set of seasonal data and they were converted into raw data into an understandable format. The main challenge in weather prediction is poor data quality and selection [14]. For this reason, we try to pre-process data carefully to obtain accurate and correct prediction results [7].

Data Cleaning

In this phase, unwanted data or noise is removed from the collected data setwhich is done by removing the unwanted attributes and keeping the most relevant attributes that help in better prediction. Another major issue that is to be rectified isthe missing values in the collected data set.

Missing values in the data set are filled by using various techniques. In this work, the missing values for attributes in the dataset are replaced with the modes and means based on existing data. Adding the missing values provides a more complete dataset for the classifiers to be trained on.

Data Integration

This phase involves the integration of multiple databases or files. To combine separate year files to one file.

Data Transformation

In this phase converting data from one format to another. To convert excel to CSV and CSV to .arff file format.

Data Reduction

In this phased process of reducing the amount of capacity required to store data. Data reduce can increase storage efficiency and reduce costs.

Feature Extraction

It is the technique of selecting a subset of relevant features for building Strong learning models. Many features like Temperature, Humidity, Wind Speed and visibility affect the precipitation (rainfall) [1]. Out of it, the most relevant four features are considered in this paper. The following are the features selected.

Temperature

Earth's atmosphere varies with the distance from the equator (latitude) and height above the surface (altitude). Air temperature is the intensity aspect of the sun's energy that strikes the earth's surface. Because energy from the sun reaching the earth varies from day to day, from season to season, and from latitude to latitude, temperatures also vary. The earth as an entire receives a constant flow of radiant short-wave energy from the sun. The earth also radiates long-wave energy to space. During the day, the movement of short-wave radiation absorbed exceeds long-wave energy emitted, and the surface temperature increases.

Humidity

The relative humidity is a term used to describe the amount of water vapor in a combination of air and water vapor. It is defined as the ratio of the partial pressure of water vapor in the air-water combination to the saturated vapor pressure of water at the prescribed temperature. The relative humidity of air is determined by not only on temperature but also on the pressure of the system of interest. The relative humidity is often used instead of complete humidity in situations where the rate of water evaporation is essential, as it takings into account the difference in saturated vapor pressure.

Visibility

Visibility is a measure of the distance at which an object or sunlit can be recognized. Visibility is reported within surface weather observation and METAR either in meters or miles, dependent upon the country.

Wind Speed

The wind is the movement of gases on a large scale. On Earth, wind contains the bulk movement of air. The wind is caused by differences in pressure. When a difference in pressure exists, the air is enhanced from higher to lower pressure. Wind speed is affected by several factors and situations, operating on varying scales. These contain the pressure gradient, Ross by waves, and jet streams, and local weather situations. There are also links to be found between wind speed and wind direction, especially with the pressure gradient and surfaces over which the air is found.

Precipitation

Precipitation is the compressed water vapor that falls due to gravity, the valuegreater than zero indicates rain.

TABLE II Weather parameters

Attribute	Type	Description
Temperature	Numerical	Temperature inFahrenheit
Humidity	Numerical	Humidity in Fahrenheit
Wind Speed	Numerical	Wind Speed in Knot
Visibility	Numerical	Visibility in Miles
Precipitation (Rainfall)	Nominal	Precipitation in mm

Data Discretization

Convert numerical data into nominal data in Precipitation. Precipitation value is 0denote the "NO RAIN" otherwise "RAIN".

IV. CLASSIFICATION ALGORITHM FOR RAINFALL PREDICTION

This research aims to analyze the performance of data mining techniques on rainfallprediction in Mumbai city using a classification framework Figure 3 [13].

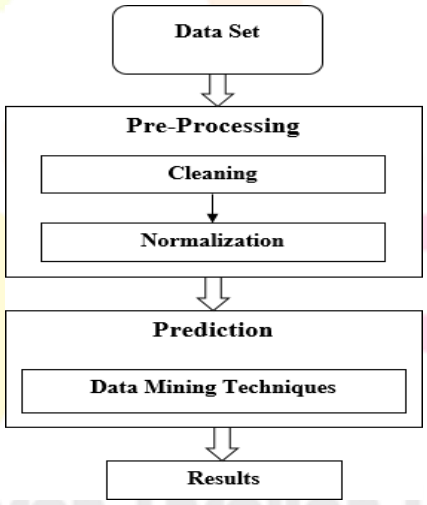

Figure 3 Classification framework

a. Naive Bayes

Naive Bayes classifiers are a group of classification algorithms based on Bayes'Theorem [10]. It is not a single algorithm but a group of algorithms where all of them share a

common principle, i.e. every pair of features being classified is autonomous of each other. It builds the own model very easily without complex parameter estimation. For this reason, only these classification methods used for large data sets. While dealing with the real-world problem; this method gives an obvious model for its deployment [11]. It uses very fewer training data to predict classification parameters.

b. Support Vector Machine

Support Vector Machine is supervised learning classification techniques for both linear and non-linear data [2]. Support Vector Machine is the most robust and accurate classification technique. Support Vector Machine is normally used for classification, regression and ranking functions. Support Vector Machine creates one or a set of hyperplanes in an infinite-dimensional space. The data points in the variable space are mapped into the hyperplane by selecting kernel functions. Support Vector Machine provides better performance than other neural network models. It is an important technique to solve many classification problems in recent years. It is relevant for non-vector data such as graphs and maps. Support Vector Machine provides a unique and best solution. The best separating hyperplane can be written as,

$$\text{Weight}.x + b = 0$$

Where Weight is a weight vector, x refers to the values of attributes and b is scalar frequently referred to as bias [5].

c. C4.5 Decision Tree-Based (J48)

Unlike Naïve Bayes, C4.5 is a classification algorithm used to create a decision tree for the given dataset. It is constructed on the information entropy concept. Construction of the decision tree is done by selecting the best probable attribute that will be able to split a set of samples most effectively. The attribute having the utmost entropy difference or normalized data gain is selected as the splitting criteria for that specific node. A similar manner to follow and nodes are added to the decision tree [12]. Each penultimate node carries the last attribute or several attributes for creating the last decision of the problem.

d. Logistic regression

Logistic regression is a classification system that uses mathematical logistic regression functions. The most popular models are comprehensive linear models [4]. It measures the relationship between the categorical dependent variable and one or more independent variables by approximating probabilities using a logistic function, which is the cumulative logistic distribution. Thus, it treats the same set of problems as probit regression using similar techniques, with the latter using a cumulative normal distribution curve instead. Equally, in the latent variable interpretations of these two methods, the first assumes a standard logistic distribution of errors and the second a standard normal distribution of errors. The regression coefficients are usually assessed using maximum likelihood assessment.

Researchers have applied statistical methods such as univariate or multivariate binary logisticregression to predict rainfall [6].

e. Random Forest

Random Forest is also another approach under the ensemble classifier. Random Forest is a classifier based on decision trees that exhibits great performance in computer engineering. The random forest has one important benefit that it is fast and handles a large number of input attributes. It includes tens or hundreds of trees [12]. In the structure of the decision tree, a random choice of attributes is involved. The trees are created using the following strategy:

1. Each tree's root node has a sample bootstrap data which is equal to the actual data. There is a different bootstrap sample for each tree.

2. Using the best split method subset of variables is randomly selected from input variables.

3. Each tree is then grown to the maximum extent possible without pruning.

4. When all trees are built in the forest, new instances are attached to all the trees then the voting process takes place to select the classification with maximum votes as the newinstance(s) prediction.

V. RESEARCH METHODOLOGY

a. Data retrieval

The Mumbai rainfall dataset is taken from the WMO website for a daily basis for months June, July, August, and September. The south-west monsoon rainfall data is highly complex. The data can be separated in the year 2012 – 2018 contain 861 data.

The raw weather data collected, it consists of eleven measured attributes. This data contains noisy and irrelevant data.

b. Data pre– processingData Cleaning

The data obtained is noisy because there are some missing values (Precipitation = 99.9) and some unwanted data. Clean the data manually. After cleaning contain the 566 data.

Data Integration

They separated seven years of data set that can combine the single data set.

Data Reduction

Reduce the number of attributes to increase accuracy.

Data Discretization

Discretization is the conversion of numerical data into nominal data. Theattribute class for rainfall is based on the distribution listed in Table III:

TABLE III Discretization

Rainfall	Category
Rainfall ≤ 0.1 mm	No Rainfall
0.1 mm < Rainfall	Rainfall

Data Transformation

The CSV file was converted to .arff format to feed it into the Data Mining tool – WEKA.

c. Comparison of different classification algorithms for rainfall prediction

The thesis is conducted on weather data of Mumbai, Maharashtra, India from 2012 to 2018 South-West (SW) monsoon. Which is first pre-possessed and cleaned by implementing the data mining process model [10]. The experiments are conducted to compare various data mining algorithms for rainfall prediction. In our collected weather data set, the predicted variable which tells whether it will rain on the next day or not. WEKA tool is used for the implementation of the thesis. The Training set is chosen for the thesis. By applying various algorithms on the cleaned data set models are generated which are also known as classifiers [8]. The percentage of correctly classified instances by the classifier (model) known as classification accuracy gives us the performance measure of the classifier (model). There is a total of 566 instants in a dataset. Each record has 5 attributes including the last attribute defines the class label of the record, whether it will rain or not.

The results of various machine learning algorithms are compared on the Accuracy, Precision, Recall, F-measure, Kappa statistic, Receiver Operating Curve (ROC), mean absolute error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE) and Accuracy. Prediction accuracy and performance measures of applied prediction models based on the rainfall dataset are shown in table IV and are also graphically observed in Figure 4. Each value presented in table IV is the result of the training set run.

TABLE IV Performance measure of algorithms using rainfall data

Algorithm	Precision	Recall	F-measure	Kappa Statistic	ROC	MAE	RMSE	RAE	Accuracy
NaiveBayes	0.765	0.772	0.767	0.4501	0.82	0.2811	0.395	65.0847 %	77.1681 %
SVM	0.79	0.786	0.763	0.4303	0.688	0.2142	0.4628	49.5941 %	78.5841 %
C4.5	0.856	0.853	0.845	0.6299	0.842	0.2365	0.3439	54.7722 %	85.3097 %
Random Forest	0.986	0.986	0.986	0.9672	0.999	0.1099	0.177	25.4528 %	98.5841 %
Logistic regression	0.796	0.802	0.792	0.5046	0.837	0.2967	0.3811	68.7179 %	80.177 %

It can be seen from the results that Random Forest has the best prediction model as compared to other algorithms. The graph given in Figure 4 shows the result of various classification algorithms and their performance measures [15].

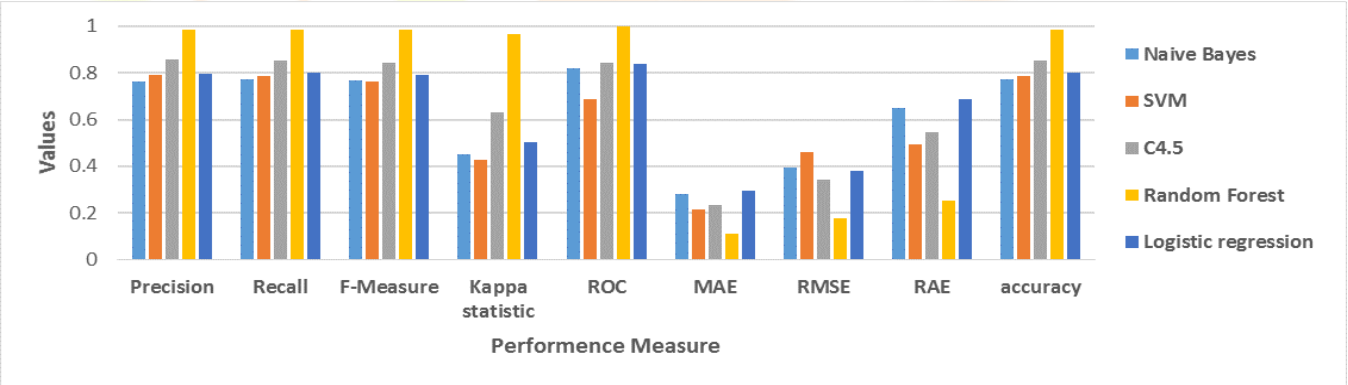

Figure 4 Different Performance Measures

From figure 5, we can observe that the accuracy value of Random Forest is highest as compared to other data mining algorithms. The different values of Precision, Recall, F- Measure, ROC, MAE, RMS, RAE and Accuracy for a given rainfall dataset are shown in table IV. It can be observed that out of seven classification algorithms, Random Forest exhibits the highest values of Precision, Recall, F-measure, ROC, and Accuracy. Random Forest also produces a minimum amount of Root Mean Square Error (RMSE) among all the eight algorithms used.

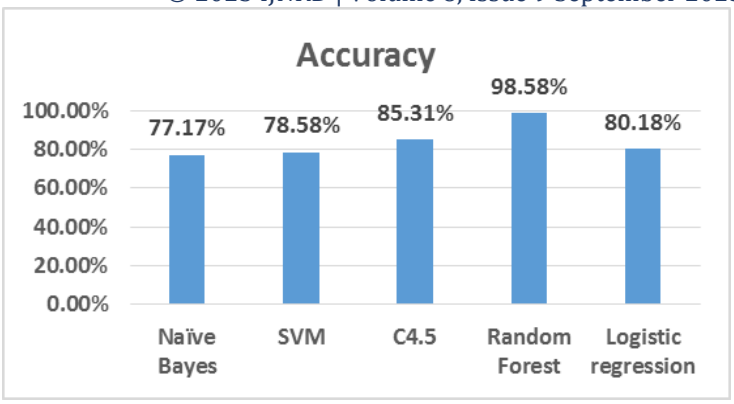

Figure 5 Accuracy of Various Algorithms

VI. CONCLUSIONS

In this study, we compared popular data mining algorithms for rainfall prediction using various performance measures over weather data of Mumbai, Maharashtra, India. The different measuring attributes play a pivotal role in giving precise rainfall prediction. We have observed that Random Forest produces best rainfall prediction results with an accuracy of 98.58% in the training set method and also exhibits the highest values in Recall, F-Measure, and ROC as compared to other classification algorithms. In our case, the Random Forest approach proves to be an efficient and acceptable method for rainfall prediction. The level of accuracy and prediction highly depends on the data being used as input for classification and prediction. Every method has its advantages and limitations; choosing the best algorithm is difficult. The prediction accuracy of the model can be increased by developing a hybrid prediction model where multiple machine learning algorithms are put to work together [15]. For our weather dataset, it was concluded after analyzing various models of supervised learning that the Random Forest classification algorithm has an appreciable level of accuracy and acceptance.

ACKNOWLEDGMENT

The authors wish to thank World Metrological Organization (WMO) Mumbai, Maharashtrafor applying data.

REFERENCES

[1] S. Meganathan and T. R. Sivaramakrishnan, “Association Rule Mining and Classifier Approach for 48-Hour Rainfall Prediction Over Cuddalore Station of East Coast of India,” vol. 5, no. 14, pp. 3692–3696, 2013.
[2] D. R. Nayak, “A Survey on Rainfall Prediction using Artificial Neural Network,” vol. 72, no. 16, 2013.
[3] M. Abrar, D. Shah, and S. Khusro, “Weather Prediction using Classification,” no. February 2015, 2014.
[4] A. H. M. R. Imon, “Prediction of Rainfall Using Logistic Regression,” no. May, 2015.
[5] V. R. Mandla and T. Teja, “Prediction of Rainfall Using Support Vector Machine and Relevance Vector Machine,” 2015.
[6] C. Cheung, M. A. Hart, and M. R. Peart, "projection of future rainfall in hong kongusing logistic regression and generalized linear model."
[7] F. Sheikh, S. Karthick, D. Malathi, J. S. Sudarsan, and C. Arun, “Analysis of DataMining Techniques for Weather Prediction,” vol. 9, no. October, 2016.
[8] R. Sukanya and K. Prabha, “Comparative Analysis for Prediction of Rainfall usingData Mining Techniques with Artificial Neural Network,” no. 6, pp. 288–292, 2017.
[9] A. Parmar, K. Mistree, and M. Sompura, “Machine Learning Techniques For RainfallPrediction : A Review,” no. March, 2017.
[10] I. No, “Available Online at www.ijarcs.info A SURVEY ON CLASSIFICATIONTECHNIQUES USED FOR RAINFALL,” vol. 8, no. 8, pp. 226–229, 2017.
[11] A. Pait and M. Venkatesan, “The Open Environment System : Rainfall Prediction Using Naïve Bayesian Algorithm,” vol. 19, no. 5, pp. 1–10, 2017.

[12] S. Karthick, D. Malathi, C.Arun, "Weather prediction analysis using random forest algorithm," vol. 118, no. 20, pp. 255–262, 2018.
[13] S. Aftab, M. Ahmad, N. Hameed, M. S. Bashir, I. Ali, and Z. Nawaz, "Rainfall Prediction in Lahore City using Data Mining Techniques," vol. 9, no. 4, pp. 254–260, 2018.
[14] S. R. Devi, "Prediction of Rainfall Using Data Mining Techniques School of Electronics Engineering," *2018 Second Int. Conf. Inven. Commun. Comput. Technol.*, no. Icicct, pp. 1507–1512, 2018.
[15] R. Mohd, M. A. Butt, and M. Baba, "Comparative Study of Rainfall Prediction Modeling Techniques (A Case Study on Srinagar , J & K , India)," vol. 7, no. 3, pp.13–19, 2018.