A Study on Prediction of Rainfall Using Datamining Technique

R.Senthil Kumar
Research Scholar
Department of Computer Science and Engineering
Satyabama University
Chennai.
email: phdsenthilscholar@gmail.com

Dr. C.Ramesh
Research Supervisor
Department of Computer Science and Engineering
Satyabama University
Chennai.
email: crameshmail@gmail.com

Abstract— India is an agricultural country and its economy is largely based upon crop productivity and rainfall. For analyzing the crop productivity, rainfall prediction is require and necessary to all farmers. Rainfall Prediction is the application of science and technology to predict the state of the atmosphere. It is important to exactly determine the rainfall for effective use of water resources, crop productivity and pre planning of water structures. Using different data mining techniques it can predict rainfall. Data mining techniques are used to estimate the rainfall numerically. This paper focuses some of the popular data mining algorithms for rainfall prediction. Naive Bayes, K- Nearest Neighbour algorithm, Decision Tree, Neural Network and fuzzy logic are some of the algorithms compared in this paper. From that comparison, it can analyze which method gives better accuracy for rainfall prediction.

Keywords:- Naive Bayes, K- Nearest Neighbour algorithm, Decision Tree, Neural Network, Fuzzy Logic.

I. INTRODUCTION

A wide variety of rainfall forecast methods are available in India, because India is an agricultural country and the success of agriculture depends of rainfall and humidity. There are mainly two approaches to predict rainfall in India. They are Empirical method and dynamical method. The empirical approach is based on analysis of historical data available from http://www.imdchennai.gov.in/ of the rainfall and its relationship to a variety of atmospheric variables. The most widely used empirical approaches used for climate prediction are regression, Artificial Neural Network(ANN), Decision Tree algorithm, fuzzy logic and group method of data handling. In dynamical approach, predictions are generated by physical models based on systems of equations that predict the evolution of the global climate system in response to initial atmospheric conditions[1].

In this paper, various algorithms have been analyzed. Data mining techniques are efficiently used in rainfall prediction.

2. LITERATURE REVIEW

2.1 Works Related to Classification Techniques

F.Dell presented a more general approach for identifying the varying the wind speed. The metrological data is divided into two groups and performed the classification on these groups to identify the hurricanes and non-hurricanes. The above authors analyzed and predicted the rainfall.

Valmik.B et al. proposed a model for predicting the weather data based on classification technique and considered several important attributes such as wind pressure, humidity, vapor, wind speed and pessimistic results obtained from the experimental result. The results showcased good accuracy by correlating the above parameters.

Prediction of data using PCA is proposed by Evangelos.T et al. the works are carried out ERA-40 data set and the results are showcased using high relevance accuracy. Meteorology's are presented using image acquisition , wind pressure, temperature and future extraction models.

Improved Nave Bayesian classification is considered by James-N.K.Liu by varying the weather state data. The Bayesian classification is used for identifying the weather prediction. The algorithm is compared with different models based on genetic algorithms and results showed that the considerable amount of accuracy is obtained.

Works on k-nearest neighbor classification are utilized to evaluate the wind speed and using the Manhattan and miniskowski distance the speed of wind is estimated. The authors Menhet.w et al. have also analyzed these works using three parameters like wind direction, air temperature and atmospheric pressure.

The fluctuating wind series are modeled by P.J Trombe et al. using the time series analysis together with Markovian switching model and there by a methodology is carried out for estimating the weather changes in the sea using radars. The classification is used to find the relationship between precipitation and wind attributes.

A renewable energy source optimization model is proposed by M.Iqbal et al. for classifying the wind sources using optimization classification algorithm. The works are further extended by Vibav.M by integrating clustering and

classification by which wind speed evaporation is calculated and minimum and maximum temperatures can be comfort obtainable. k-means algorithm is utilized for clustering the data.

2.2 Works Related To Clustering Techniques

Clustering is considered as grouping the homogenous data from the heterogeneous group of data available in the networks. Many clustering models are available in the literature for effective analysis of weather data. Among these models, in the literature works are highlighted using both partition clustering and aggloromative clustering techniques. This section of the paper highlights a brief review of the works carried out in the area of weather prediction in the recent past.

TV Rajinikanth et al. have proposed a methodology for analyzing the metrological data that is very much suited in particular to the Indian weather environments using different data mining techniques. The related data is grouped basing on k-means algorithm, a decision tree is used for predicting the observations. J.48 classification technique is used for the data classification purpose. The experimentations are considered based on time series analysis on the data available during 1955 to 1965.

Kavita.P et al. have presented a model for interpreting the weather using k-means clustering an effort is made to analyze the temperature at atmospheric pressures with 400hpa, 500hpa,700hpa. The experimentation is also followed by varying the humidity and for presenting the model the weather data set of Bangladesh of Pakistan is considered.

Badhiye S.S. et al. presented an approach to analyze and predict attribute of temperature, humidity values for future using clustering technique. The outlier analysis is considered to detect the outliers with respect to the data and clustering analysis is used to partitioning the data based on the similarity of the objects. K-nearest neighbor algorithm is used to predict the values of temperature and humidity parameters of weather condition. To find the distances between datasets in the KNN algorithm absolute distance and Euclidean distance measuring are used.

Zahoor Jan M. Abrar et al. developed a system for seasonal to inter-annual weather prediction by using data mining KNN technique. The main purpose is how to use a data mining technique K-NN and how to develop a system that uses numeric historical data to forecast the climate of specific region, city or country in the months in advance. Sea surface temperature is considered as a main factor and regression tree technique is used to find the predication of climate. Euclidean distance metric is used to find the distance between the samples. The data set for the system was collected from the Pakisthan meteorological department and the Capital and National Climate Data (NCDC) which consists of 10 years of historic data and stored in MS ACCESS format as backend.

Sarah N.Kohail and Alaa M.El-Halees presented a model for analyzing meteorological data using different data mining techniques. The authors tried to extract useful knowledge from weather daily historical data collected at Gaza strip city during the duration of 9 years period (1977-1985). Daily average relative humidity, average temperature, wind speed, wind direction, the time of the highest wind speed and rainfall are the parameters considered. Time series analysis is used to identifying temperature and linear interpolation method is used to fill missing values. Outliers are removed by using outlier analysis. The authors use cluster analysis to partitioning data by using K-means algorithm at K=4. ANN and least median squares linear regression method are the prediction methods. Classification and association rules are also used.

Meghali A.Kalyankar and S.J.Alaspurkar presented a data mining technique to analyze the meteorological data. The weather data is extracted to find hidden patterns inside the large dataset by using clustering technique i.e., K-means partitioning algorithm. The authors used 10 channel midi-data logger system that provide weather data in form of excel sheets and can automatically collect data on a 24-hour(day) basis. Data preprocessing technique is used applied to the weather data which consists of various parameters like as temperature, humidity, rain, wind speed etc. Data preprocessing involves Data cleaning, Data transformation is used to reduce noisy and to fill the missing values. Data transformation is a phase in which the selected data is transformed into appropriate forms.

Sanjay Chakraborty et al. presented a methodology for weather forecasting using incremental K-means clustering. The weather forecasting is done based on the incremental air pollution data base of west Bengal in the years of 2009 and 2010 available from the website. In this approach, the authors proposed a generic methodology for time series forecasting through clustering. The air pollution data is collected from "West Bengal Air Pollution Control Board". This data base consists of 4 air pollution elements or attributes. They are Carbon Dioxide (Co2), Respirable Particulate Matter (RPM), Sulphur Dioxide (So2) and Oxides of Nitrogen (NoX). K-means is applied for initial data and incremental K-means is applied to the new coming data by using Manhattan metric measure.

Marwa F.Al-Roby and Alaa M.El-Halees presented a data mining technique for wind speed analysis. The data recorded between 2004 to November 2006 daily historical data by meteorological station of Gaza is considered. After preprocessing the data, they applied data mining techniques, Association rules, Classification Rule, Cluster Analysis, outlier analysis are used. The observed data of wind speed contain 4 years from January 2003 to November 2006 daily historical data. To normalize the data, Normalization is applied.

Abay Kumar et al. presented a model using K-means clustering algorithm. The authors have used Probability Density Function Algorithm to generate numerical results in

the k-means clustering for weather –related predictions. In this work, they constructed a model for predicting the probability of play class for YES and NO category through k-means clustering technique. The original data is collected from Quinlan in the year 1986. They considered five attributes namely outlook, temperature, humidity, windy and play. The attribute outlook has three possibilities sunny, overcast and rainy. The attribute windy can be either true or false. The outcome class play is either YES or NO in this the authors obtained the results 77.36% of overall accuracy, 85.635 of precision, 91.63% of recall.

Zohreh Nazeri and Jainping Zhang presented a methodology to analyze severe weather impacts on National Airspace System (NAS) performance by using data mining applications. The frame work of this approach consists of three phases: data preparation, feature extraction and data mining. For this experiment, the authors used three sources: Airline Service Quality Performance (ASQP), Enhanced Traffic Management System (ETMS) and National Convective Weather Forecast (NCWF) supplied by National Center for Atmosphere Research from April to September in 2000 to represent the severe weather season. 152 days of complete weather data were used for this technique. The feature extraction phase consists of four steps in this work: image segmentation, weather feature extraction, and air traffic feature extraction and representation conversion. The authors used segmentation algorithm to identify the severe weather areas. They considered correlation analysis to calculate the eight weather and air traffic features with the performance. They applied clustering to group the days with similar weather impacts on NAS performance. K-means clustering algorithms was also applied for this data. The authors used classification method to partition the performance measure into three classes: bad, medium and good. They applied decision tree algorithm C5.0 to learn rules for these classes. They achieve 86% classification accuracy using 100-fold cross validation on the two-class problem and 77% on the three-class problem.

M.Mayilvahanan et al, presented an approach for estimating the availability of sunshine using data mining techniques. This work deals with estimating the temperature values of four cities in Tamil Nadu, South India namely Chennai, Coimbatore, Madurai and Kanyakumari. The dataset was collected from the India Meteorological Department portal like http://www.imdchennai.gov.in/. The dataset consists of monthly mean maximum, minimum temperature and total rainfall based upon 1901-2000 data of various cities in India. The attributes present in the data are station name, month, period(years), number of years, maximum mean temperature in degrees Celsius, minimum mean temperature Celsius and mean rainfall in millimeters. Data preprocessing was applied to the data to clean and improve the quality of collected data. K-Means clustering algorithm and Expectation Maximization algorithms were used to compare the availability of sunshine in the cities. The authors claimed that the maximum amount of sunshine is recorded in the city of Chennai when compared to the other cities.

Stefnao Serafin et al presented an application of cluster analyzes technique to the verification of quantitative precipitation forecasts. Clustering analyzes is adopted as an objective method to create groups of rain gauges displaying interrelated measurements. In this research work, most of the events are considered with an upper level throw and a surface level low approaching northern Italy from the west. The authors detected the cluster by using the average linkage algorithm. The results of the verification of the precipitation forecasts are highly affected by distribution of rain gauges.

J.Mgutierrez et al. presented an application of clustering methods for statistical down scaling in short range weather forecasts. In this research work, the clustering technique compared with the standard nearest neighbors analogue methods. Some validation results daily gives precipitation and maximum wind speed operative downscaling (lead time 1 to 5 days) on a network of hundred stations in the iberian peninsula for the period 1998 to 1999 are collected as data. In this work, Self-Organizing Maps (SOMS) to define weather classes and study the associated distributions of local precipitation. The authors claimed the outputs of an operative circulation model on different local area or large-scale greets are considered to characterize the atmospheric circulation patterns.

Vipul Kedia et al., presented a methodology for time series forecasting through clustering technique. In this research paper, the authors used subset of the data set to build up the system model by compressing the information through clustering and coming up with inherent patterns for data. The pattern was represented as the curves such that any time series from the given set is expected to follows. The authors used linear regression method by matching to the closest pattern to each time series that has to be predicted. In this research paper, the authors classified the research paper in to two partitions these are dead or alive papers.

N.Raja shekar, T.V.Rajinikanth presented a weather analyzers of Guntur district of Andhra region using SVM data mining techniques. The authors developed a machine learning algorithm that is hybrid SVM (Support Vector Machine) model. The model is used for effective weather prediction by analyzing the given weather data and to recognize the patterns exiting it. SVM comes under the set of supervised learning methods for classification and regression in this research paper, Guntur district weather data sets were considered for analyzers using the hybrid SVM data mining techniques. The authors applied K-means clustering technique over the clustered data set they considered the data set for about 102 years namely, monthly mean for each year average temperature (1901- 2002).

The authors yielded good result in predicting the weather that the existing machine learning programming techniques Ankita Singh et al. presented an efficient clustering methods

for atmospheric conditions prediction using art algorithm. The aim of this research is to develop artificial neural networks based on clustering method for ambient atmospheric conditions prediction in an Indian city. In this research work they presented a clustering method that classifies cities based on atmospheric conditions like temperature, pressure, wind speed and humidity. The data represents the month wise atmospheric conditions under two parameters namely temperature and pressure. And collected from the ten different cities namely Delhi, Kolkata, Bhopal, Mumbai, Jaipur, Amritsar, Cochin, Lucknow, Bhubaneshwar and Guwahati.

Ngugen Dinh Hoa et al. presented a methodology for weather now casting from satellite image sequences using picture fuzzy clustering and spatiotemporal. Weather now casting comprises the detail description of the current weather along with forecasts obtained by the extrapolation for every short range period of zero to six hours ahead. In this research paper, firstly the satellite image pixels are partitioned into clusters by using fuzzy clustering-a fuzzy clustering method based on the theory of picture fuzzy sets secondly, the fast Fourier transform method is used to filter out non predictable scales leading to the increasing time ranges of predictability. Finally the spatiotemporal regression method is used to forecast the predicted sequences of images the authors claimed the better results than the relevant once in weather now casting.

S.Gokila et al. presented an approach based on clustering and classification rule in support of climatology to mine weather data. The aim of this paper is to provide a review report on various popular data mining techniques applied on weather data set in support of weather prediction and climate analyzers. In this research work, clustering techniques applied on the climate data helps to produce similar patterns of climate with the consideration of spatial nature. The classification techniques are used to relate the attributes of weather data to predict the future climate.

Dingsheng, Yaming, Nan and Yufeng [16] proposed the annual average exciting rainfall prediction model based on BP network combined with stepwise discriminant method and use Bayesian statistical method to further improve the network's generalization ability and model prediction accuracy, but the overall performance of the technique can be further improved, such as to further improve the correct ratio on discriminant analysis. The experimental results validate the method and the prediction accuracy is satisfactory.

Aastha Sharma et al. presented a semi supervised technique for weather condition prediction using DBSCAN and KNN. In this research paper, certain atmospheric parameters are taken for four years on a day wise basis in a certain city the weather attributes like snow, rain, fog, are considered by authors in this paper. The authors proposed a new approach that combines data mining technologies like classification and clustering techniques. The semi supervised clustering algorithm to identify clusters in large spatial data sets. The clusters are formed they are given for training to KNN. K-Nearest Neighbor Rule (KNN) has been one of the most known supervised learning an algorithm in pattern classification. In this research work, the authors considered overall accuracy, class wise accuracy, class wise precision, class wise recall, class wise f-measure as elevation parameters in performance analyzers they considered three class fog, rain and snow for the conditions accuracy, precision, recall.

Hans Henrick Benzon and Thomas Bovith et al. presented a methodology for simulation and prediction weather radar cluster using a wave propagator on high resolution NWP data. Cluster is a common problem caused by non-standard wave propagation in weather radar application. In this paper, the novel proposed method uses a wave propagator to identify areas. The wave propagator uses a three dimensional refractivity field like temperature, humidity and pressure. These methods are obtained from high resolution. The wave propagator is based on the parabolic equation. The authors obtained the results by removing of cluster echoes while preserving precipitation echoes in this worked paper.

Subartono , Dwi , Bambang, Sutikno and Heri [25] proposed an ensemble method based on ANFIS and Autoregressive Integrated Moving Average (ARIMA) for forecasting monthly rainfall data at certain area of Indonesia, namely pujon and Wagis. Two empirical rainfall datasets were used to compare the forecasting accuracy between individual ARIMA, ANFIS and ensemble methods. The results showed that individual method is more accurate than ensemble methods. For wagis accuracy is predicted using ARIMA and for the area of pujon is better accuracy predicted using ANFIS. Two rainfall data from Jan 1975 – Dec 2010, for testing Jan – Dec 2010 were used. In general, these results in line with M3 competition results that more complicated model not always yields better forecast than simpler one.

Narasimha, Prudhvi and Naidu [26] proposed a decision tree method using SLIQ to implement the precipitation model. The authors have observed that decision tree method achieves closer agreement between actual and estimated rainfall. SLIQ method gives high accuracy rate when compared to other prediction model like fuzzy logic, NN etc. The use of Gini index for rainfall analysis is quite apt because of the irregularities present in the statistical data of precipitation. It gives accuracy of nearly 72.3% and completely based on historical data. The decision tree constructed and the classification rule are generated as per rules.

Mark, Bobby, Yung and Beth [27] proposed time series analysis is used as prediction algorithm. There are two components like rainfall/evaporation and crop management. Decision support system for Agriculture management using prediction algorithm aimed to develop a system that will determine the trend of rainfall and evaporation using time series analysis as its prediction algorithm, to develop web-based application that displays graphs and tables according to

the result of the prediction algorithm, and to utilize a classification of crops that aids farmers as basis for recommendation according to the predicted amount of rainfall per quarter. The system is found useful in terms of efficiency and reliability. It shows interface the quarter of the year labeled Q1, Q2,Q3, Q4, prediction of average amount of rainfall and evaporation, the trends, and the seasonal effects in its provided field in the table.

3. Materials and Methods

Data Collection -The data used for this work was collected from specific region. Following stages of the research applied on collected data: Data Cleaning, Data Selection, Data Transformation and Data Mining. Data Collected from the web like: http://www.imdchennai.gov.in/statistics.htm

Data Cleaning- In this stage, a consistent format for the data model was developed which took care of missing data, finding duplicated data, and weeding out of bad data. Finally, the cleaned data were transformed into a structured format suitable for data mining. A very low-quality information is available in various data sources and on the Web; many organizations are interested in how to transform the data into cleaned forms which can be used for high-profit purposes. This goal generates an urgent need for data analysis aimed at cleaning the raw data.

Data Selection - At this stage, data relevant to the analysis was decided on and retrieved from the dataset. The meteorological dataset had ten (10) attributes, their type and description is presented in Table 1. Due to the nature of the Cloud Form data where all the values are the same and the high percentage of missing values in the sunshine data both were not used in the analysis.

TABLE 1: ATTRIBUTES OF METEOROLOGICAL DATASET

Attribute	Type	Description
Year	Numerical	Year Considered
Month	Numerical	Month Considered
Wind Speed	Numerical	Wind run in Kilometers
Evaporation	Numerical	Evaporation
Cloud Form	Numerical	The Mean Cloud Amount
Radiation	Numerical	The amount of Radiation
Sunshine	Numerical	The amount of Sunshine
MinTemp	Numerical	The Monthly Minimum Temperature
Rainfall	Numerical	Total monthly Rainfall
MaxTemp	Numerical	Maximum Temperature

Data Transformation-This is also known as data consolidation. It is the stage in which the selected data is transformed into forms appropriate for data mining. The data file was saved in Commas Separated Value (CVS) file format and the datasets were normalized to reduce the effect of scaling on the data.

Data Mining Stage -The data mining stage was divided into three phases. At each phase all the algorithms were used to analyze the meteorological datasets. The testing method adopted for this research was percentage split that train on a percentage of the dataset, cross validate on it and test on the remaining percentage. Thereafter interesting patterns representing knowledge were identified.

4. Comparison of Data Mining Techniques

According to the previous work done by researchers presented in the literature review, a comparison can be done. Various data mining techniques are used to predict different parameters of weather like humidity, temperature and wind gust. Various attributed used for the comparison are applications, authors, data mining techniques, algorithms, attributes, time period, dataset size, accuracy percentage, advantages and disadvantages. They yield different results with their cons and pros. For weather prediction, decision tree and k-mean clustering proves to be good with higher prediction accuracy than other techniques of data mining and it is proved by using of MATLAB Software Tool. Regression technique could not find accurate value of prediction. However, approximate value could be retrieved. It is also observed that with the increase in dataset size, the accuracy is first increases but then decreases after a certain extent. One of the reasons may be due to over fitting of training dataset. The work done by different researchers and their comparison is jotted down in Table.2.

TABLE 2: COMPARISON OF A DATA MINING TECHNIQUES FOR WEATHER PREDICTION

Authors	Applications	Techniques	Algorithms	Attributes	Time Period	Dataset Size	Accuracy	Advantages	Disadvantages
P.Hemalatha[27]	Weather prediction for ship navigation	Decision Tree	C4.5, ID3	Climate,Humidity, Stormy, Temperature	4-5 location	20- 30 instances	-	Verifiable performance	Do not handle continuous range data directly.
E. G. Petre[10]	Weather prediction	Decision tree	CART	Pressure, clouds quantity, humidity, precipitation, temperature	4 years	48 instances	83%	Good prediction accuracy	Data transformation is required. Extra computation required.
S Yeon etal.[33]	Hourly rainfall prediction	Decision tree	C4.5,CART	Temperature,wind direction, speed, gust, humidity, pressure	3 years	26280 instances	99%, 93%	High prediction accuracy	Small data is leftfor prediction.

S Kannan ,S Ghosh[29]	Daily rainfall prediction in river basin	Decision tree, Clustering	CART, k-Mean clustering	Temperature,MSLP, pressure, wind, rainfall	50 years	432000 instances	-	Grouping of multisite rainfall datain clusters	Small data is leftfor prediction. No verification isdone.
F Oliya,AB Adeyemo [17]	Weather Prediction and Climate Change Studies	Decision tree, ANN	C4.5,CART, TLFN	temperature,rainfall, evaporation, wind speed	10 years	36000 instances	82%	Best networkis selected for prediction	Accuracy varies highly with size of training dataset
P Sallis, S Shanmuganathan [34]	Wind gust prediction	Decision tree, ANN	C5.0, CRT, QUEST, CHAID, SOM	Dew point, humidity, temperature, wind direction, wind speed	4 years	86418 instances	99%, 85%	Good for analyzing ad hoc dataset	Data recorded at irregular intervals. Do not handle continuous data.
GJ Sawale [12]	Weather prediction general	ANN	BPN,Hopfield networks	Temperature,humidity, wind speed	3 years	15000 instances	-	Combining both gives better prediction curacy	Attribute normalization is required
Amarakoon[1]	Climate prediction in Sri Lanka	ANN	KNN	Temperature, humidity, precipitation, wind speed	1 year	365 instances	-	Beneficial for dynamic data.	Need to integrate feature selection techniques.
S Badhiye et al [4].	humidity and temperature prediction	Lazy learning, clustering	KNN, K-mean clustering	Temperature, humidity	-	-	100% approx	Suitable for multi-modal classes.	Cannot predict data in remote areas
Z Jan et al.[38]	Inter annual climate prediction	Lazy learning	KNN	Wind speed, dew point, seal level, snow depth, rain	10 years	40000 instances	96%	Long term accurate results with large set of attributes.	Cannot incorporate to reflect global changes.
M. A. Kalyankar, S. J. Alaspurkar [23]	Meteorological data analysis	Clustering	K- mean clustering	Temperature, humidity, rain, wind speed	4years	8660 instances	-	Good prediction accuracy	Dynamic data mining methods required.
K Pabreja [16]	Cloud burst predicion	Clustering	K- mean clustering	Temperature, humidity	2 days		100% clustering	Supplement with NWP models.	Not good for long term predictions.
PS Dutta, H Tahbilder[28]	Rainfall prediction	Regression	MLR	Min and max temperature, wind direction, humidity, rainfall	6 years	72 instances	63%	Acceptable accuracy.	Attribute elimination required for better accuracy
M Kannan et al.[32]	Short Term Rainfall prediction	Regression	MLR	Min and max temperature, wind direction, humidity, rainfall	3 months for 5 years	450 instances	52%	Can work even with small dataset	Instead of accurate, an approximated value is retrieved.
N Khandelwal R Davey [25]	Drought prediction	Regression	MLR	Rainfall, sea level, humidity, temperature	1 year	365 instances	-	Correlation and statistical analysis is also applied.	Verification is not done.

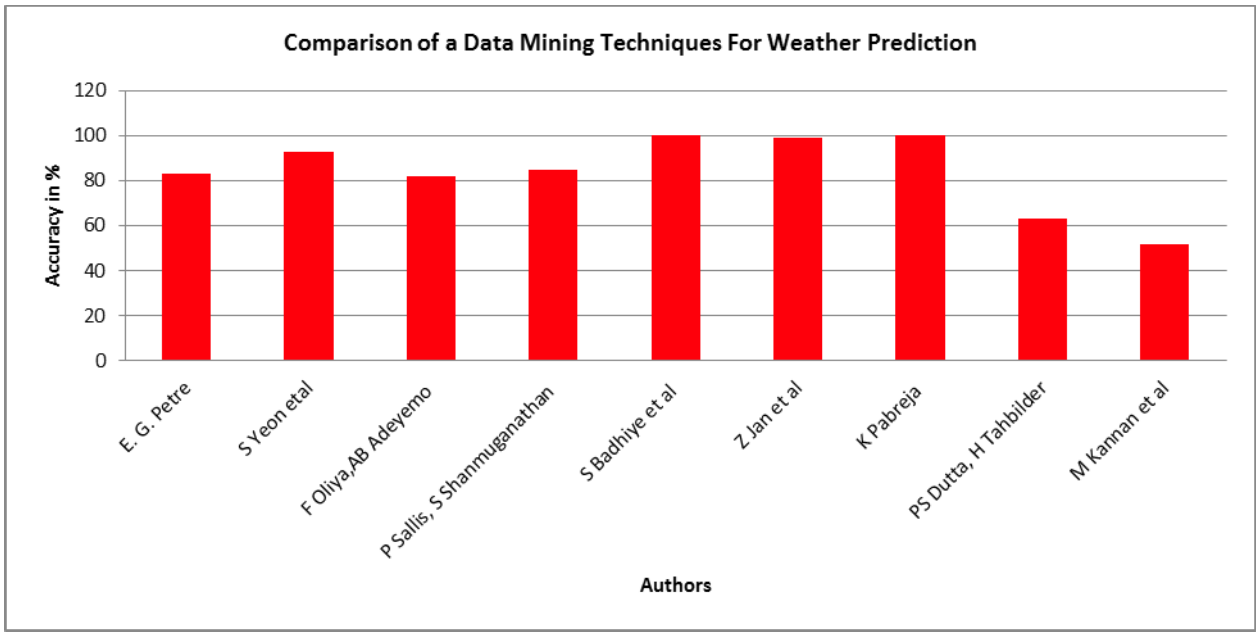

Figure 1: Comparison of a Data Mining Techniques for Weather Prediction

In the above comparison, K.Pabreja et al., is better performance to compare with the other authors. M.Kannan et al., is very low accuracy compare with the other techniques.

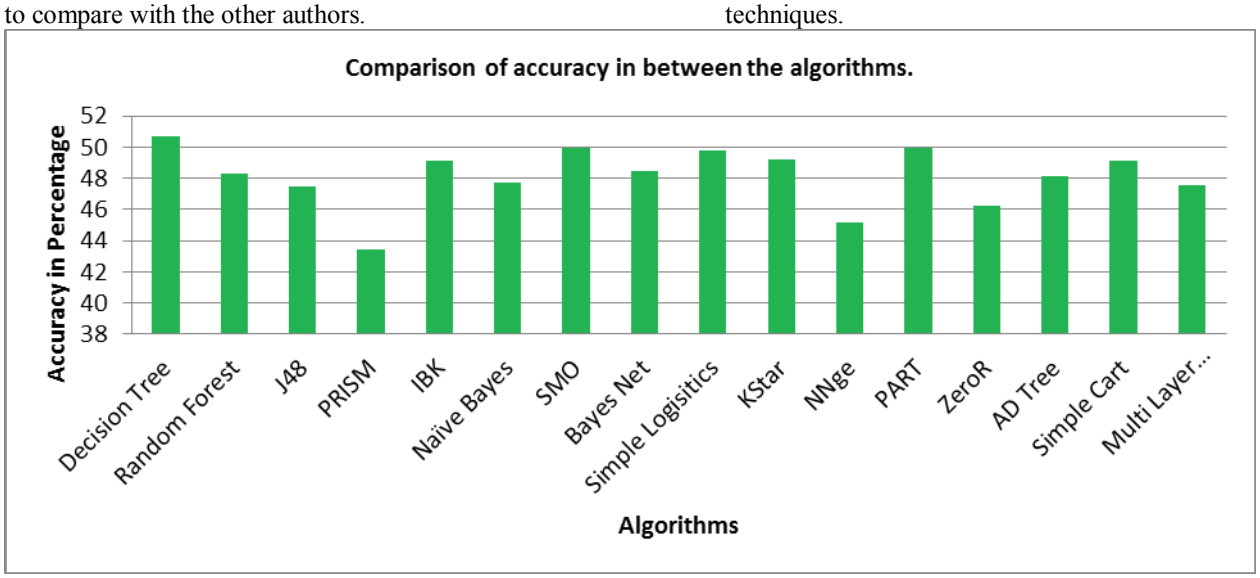

Figure 2: Comparison of accuracy in between the algorithms.

The above figure shows the accuracy of different data mining algorithms. Here, Decision Tree is only the better accuracy to compare with the other algorithms.

5. CONCLUSION

In this research paper, analysis of various popular data mining algorithms is presented for rainfall prediction. Data Mining deploys techniques based on machine learning, alongside the conventional methods. More importantly, these techniques can generate decision or prediction models, based on historical data get from Tamilnadu government website. Based on this analysis BP is combined with various other algorithms. Recent algorithms analyzed in this work are Naive Bayes, K- Nearest Neighbour algorithm, Decision Tree, ANFIS, ARIMA, SLIQ,Neural Network and fuzzy logic are some of the algorithms compared in this paper. A comparison is made in this paper, which shows that decision trees and k-mean clustering are best suited data mining technique for this application. With the increase in size of training set, the accuracy is first increased but then decreased after a certain limit.

REFERENCES

[1]. Valmik B Nikam and B.B. Meshram, "Modeling Rainfall Prediction using Data Mining Method", Fifth International Conference on Computational Intelligence, Modeling and Simulation, Issue No: 2166-8531, PP:132-136, 2013.

[2]. Evangelos Tsagalidis and Georgios Evangelidis, "The effect of Training Set Selection in Meteorological Datamining", 14th Panhellenic Conference on Informatics, Volume: 3, Issue No: 7695-4172, PP: 61-65, 2010.

[3]. Mehmet Yesillbudak, Seref Sagiroglu and Ilhami Colak, "A new approach to very short term wind speed prediction using K-nearest neighbor classification", energy conversion and management, volume- 69, PP: 77-86, 2013.

[4]. Pierre Julien Trombe and Pierre Pinson, "Automatic classification of offshore wind regimes with weather radar observations", IEEE journal of selected topics in applied earth observations and remote sensing, volume: 07, No: 1, PP: 116-125, January 2014.

[5]. M.Iqbal, M.Azam, M.Naeem, A.S.Khwaja and Anpalagan, "Optimization classification algorithms and tools for renewable energy: a review", renewable and sustainable energy reviews, volume: 39, PP: 640-654, 2014.

[6. T.V. Rajini Kanth, Balaram, N.Rajashekar, "Analysis of Data Sets using Data Mining Techniques", Computer science and technology, DOI: 10.5121/esit.2014.4510, PP- No 89-94, 2014.

[7]. Kavita Pabreja , Research scholar, "Clustering Technique to Interpret Numerical Weather Prediction Output Products for Forecast of Cloudburst", International Journal of Computer Science and Information Technologies (IJCSIT), Vol-3(1), ISSN:0975-9646, PP. 2996-2999, 2012.

[8]. Badhiye S.S, Dr.Chatur P.N, Wakode B.V, "Temperature and Humidity Data Analysis for Future Value Prediction using clustering Technique: An Approach", International Journal of Emerging Technology and Advanced Engineering, ISSN:2250-2459, volume-2, Issue-1, PP- 88-91, January 2012.

[9]. Sarah N.Kohail, Alaa M.EL-Halees, "Implementation of Data Mining Techniques for Mateorological Data Analysis", International Journal of Information and Communication Technology Research, ISSN:2223-4985, Volume-1, No:3, PP:96-100, July 2011.

[10]. Meghali A.Kalyankar, Prof.S.J. Alaspurkar, "Data Mining Technique to Analyze the Meteorological Data", International Journal of Advanced Research in Computer science and Software Engineering, ISSN:2277-128X, Volume-3,Issue:2,PP:114-118, February 2013.

[11]. Sanjay Chakraborty, Prof.N.K. Nagwani, Lopamudra dey, "Weather Forecasting using Incremental K-means Clustering", International Conference in High Performance Architecture and Grid Computing, Vol.169, Part-2, PP:338-341,2011.

[12]. Marwa F.AI-Roby, Alaa M.El-Halees, Data Mining Techniques for Wind Speed Analysis", Journal of Computer Engineering, ISSN:2010-1619, Vol-2 ,No:1, PP:1-5, 2011.

[13]. Abhay Kumar, Ramnish Sinha, Daya Shankar Verma, Vandhana Bhattacherjee, Satendra Singh (2012), "Modeling using K-Means Clustering Algorithm", First International Conference on Recent Advances in Information Technology, Vol:4, Issue-1, Issue No: 4577-0697, PP:1-5.

[14]. M.Mayilvahanan and M. Sabitha, "Estimating the Availability of Sunshine using Data Mining Techniques", 2013 International Conference on Computer Communication and Informatics (ICCCI-2013), PP:1-4,2013

[15]. N.Rajashekar and T.V.Rajinikanth, "Weather analysis of Guntur district of Andhra region hybrid SVM data mining techniques", International journal of engineering and advanced technology (IJEAT), ISSN: 2449-8958, Volume-3, issue-4, PP: 133-136, April 2014.

[16]. Ankith Singh, Dr. Bhupesh Gour, Anshul Khandelwal and Harsha Lackhwani, "An efficient clustering method for atmospheric conditions prediction using art algorithm", international journal of advanced research in computer engineering and technology, Volume-1, Issue-1, PP: 12-17, March 2012.

[17]. Nguyen Dinh Hoa, Pham Huy Thong, Le Hoang Son, "Weather now casting from satellite image sequences using picture fuzzy clustering and Spatio-Temporal regression", international symposium on geo informatics for spatial infrastructure development for earth and allied sciences, PP: 1-6, 2014.

[18]. S.Gokila, K.Anand Kumar, A.Bharathi, "Clustering and classification in support of climatology to mine weather data-a review", International conference computing and intelligence systems, ISSN: 2278-2397, Volume: 04, Pages: 1336-1340, March 2015.

[19]. Aastha Sharma, Setu Chaturvedi and Bhupesh Gour, "A semi-supervised technique for weather condition prediction using DBSCAN and KNN", international journal of computer applications (0975-8887), volume 95, No: 10, PP-21-26, June 2014.

[20]. Gurbrinder Kaur, "Meteorological Data Mining Techniques: A Survey", International Journal of Emerging Technology and Advanced Engineering, Volume-2, ISSN-2250-2459, Issue-8, PP-325-327, August 2012.

[21]. Folorunsho olaiya, Adesesan Barnabas Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies", I.J. Information Engineering and Electronic Business, DOI:10.5815/ijieeb.2012.01.07, pp-51-59, February 2012.

[22]. Gaurav J.Sawale, Dr.Sunil R.Gupta,"Use of Artificial Neural Network in Data Mining for Weather Forecasting", International Journal of Computer Science and Applications, ISSN:0974-1011, vol.6, No.2, PP:383-387,April 2013.

[23]. Mehrnoosh Torabi and Sattar Hashemi, "A DataMining Paradigm to Forecast Weather Sensitive Short-Term Energy Consumption", The 16th CSI International Symposium on Artificial Intelligence and Signal Processing, Volume: 4, Issue No: 4673-1479, PP: 579-584, 2012.

[24]. Soumadip Ghosh, Amitava Nag, Debashish Biswas and Jyoti Prakash Singh, "Weather Data Mining using Artificial Neural Network", PP:192-195, 2011.

[25]. L.A1-Matarnch, A.Sheta, S.Bani-Ahmad, J.Alshaer and La1-Oquily,"Development of temperature based forecasting models using neural networks and fuzzy logic", International journal of multimedia and ubiquitous engineering, Vol: 9, No:12, PP:343-366,2014.

[26]. Eiman Tamah A1-Shammari, Mohsen Amirmojahedi, Sahaboddin Shamshirband, Dalibor Petkovic Nenad T.Pavlovic, Hossein Bonakdari, "Estimation of wind turbine wake effect by adaptive neuro fuzzy approach", Flow measurement and instrumentation, volume: 45, PP: 1-6, 2015.

[27] Valmik B Nikam and B.B.Meshram, "Modeling Rainfall Prediction Using Data Mining Method", Fifth International Conference on Computational Intelligence, Modeling and Simulation, 2013.

[28] N.A.Charaniya and S.V.Dudul, "Focused Time delay neural network model for Rainfall Prediction Using Indian Ocean Dipole Index", Fourth International Conference on Computational Intelligence and Communication Networks, 2012.

[29] Soo-Yeon Ji, Sharad Sharma , Byunggu Yu and Dong Hyun Jeong, "Designing a Rule-Based Hourly Rainfall Prediction Model, IEEE IRI, August 8-19, 2012.

[30] Liu Cheng, Tian Yi-mei and Wang Xiao-hua, "Study of Rainfall Prediction Model Based on GM (1,1) – Markov Chain" , National Water Pollution Control and Management of Science and Technology Project of China and the Science and Technology Innovation Special Foundation of Tianjin, 2011.

[31] Nizar Ali Charaniya and Dr.Sanjay.V.Dudul, "Committee of Artificial Neural Networks for Monthly Rainfall Prediction using Wavelet transform", International Conference on Business, Engineering and Industrial Applications, 2011.

[32] Dingsheng Wan, Yaming Wang, Nan Gu and Yufeng Yu, "A Novel Approach to Extreme Rainfall Prediction Based on Data

Mining", 2nd International Conference on Computer Science and Network Technology, 2012.

[33] Jiaxing Du, Bin Zhao and Shaohui Miao, "An application on the Immune Evolutionary Algorithm based on Back Propagation in the Rainfall Prediction", International Conference on Computer Science and Electronics Engineering,2012.

[34] Long Jin, Ying Huang and Hua-sheng Zhao, "Ensemble Prediction of Monthly Mean Rainfall with a Particle Swarm Optimization – Neural Network Model, IEEE IRI, August 8-10, 2012.

[35] Scott Hellman, Amy McGovern and Ming Xue, "Learning Ensembles of Continuous Bayesian Networks : An Application to Rainfall Prediction", Conference on Intelligent Data Understanding, 2012.

[36] Suhartono, Ria Faulina, Dwi Ayu Lusia, Bambang W.Otok, Sutikno and Heri Kuswanto, "Ensemble Method based on ANFIS-ARIMA for Rainfall Prediction",2013.

[37] Narasimha Prasad, Prudhvi Kumar and Naidu MM, "An Approach to Prediction of Precipitation Using Gini Index in SLIQ Decision Tree", 4th International Conference on Intelligent Systems, Modeling and Simulation, 2013.

[38] Mark Ian Animas, Yung-Cheol Byun, Ma.Beth Concepcion and Bobby D. Gerardo, "Decision Support System for Agricultural Management Using Prediction Algorithm", 2013.