IJETAE
EXPLORING RESEARCH AND INNOVATIONS

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 7, July 2017)

Comparative Overview of the Advantages and Shortcomings of Popular Data Mining Techniques Used In Weather Prediction

Sameer kaul[i], Majid Zaman [ii], Muheet Ahmed Butt[iii]

[1]*Department of Computer Applications, Islamia College of Science & Commerce*
[2]*Department of Information Technology, University of Kashmir, India*
[3]*Department of Computer Sciences, University of Kashmir, India*

***Abstract*— Weather dynamics are always fluctuating that makes weather forecasting a very challenging task even after a rapid technological and scientific evolution in the last century. Data mining techniques and their applications have increasingly developed in the last decade. Many researchers have studied successful application of data mining tools in the prediction of weather condition and climatic change forecasting. This paper will provide an overview on the popularly used data mining techniques and gives an insight on the utility of data mining techniques in the weather prediction models. It will discuss the significant properties that are vital for data mining techniques to be incorporated in predictive model of weather forecasting. The main objective of this paper is to provide a comprehensive comparative analysis of strengths, weakness and outcomes of various data mining techniques used in weather prediction models.**

***Keywords:* Weather prediction, data mining, climate change, forecasting, predictive modelling, analytical techniques.**

I. INTRODUCTION

Data mining is a process to analyse from various perspectives and the extract useful information from it. The data mining tools employs various analytical techniques that allow analysis of data from different angles, to cognitive sciences, predictive models, social science methodologies and many more novel applications [1]Weather forecasting is amongst the most challenging scientific and technological domain of the present time owing to two major factors–first, large amount of data to be observed and analysed and second, chaotic data collection and inaccuracies in them [2]. Numerous scientists have used various techniques and analytical tools for forecasting meteorological characteristics, where some methods have been found to be more accurate than the other, one such novel technique is data mining[2],[3], [4]. The main objective of this paper is to perform a comprehensive comparative analysis of popular data mining techniques that are utilized in weather forecasting. This comparative analysis will be focused on benefits and shortcomings of each data mining techniques studied for weather prediction models along with its experimental outcomes.

A. Use of data mining in weather prediction

Weather forecasting remains a challenging and interesting field of study for scientists and lately it has been discovered that data mining techniques can be applied in this field. The advancement of technology in past few decades have developed efficient and precise methods to collect meteorological data, observational records, historical data, radar and satellite data and such more[4]. This makes it crucial to determine techniques to process such varied and huge amount of data and recognize patterns and correlation between different datasets to build an effective and accurate predictive classify it and identify patterns and correlation between them. Traditionally, the data mining techniques are employed in retail sector, financial sector, marketing to name a few. Nowadays data mining applications are increasing in problem-oriented domains, model for weather forecasting and climate change[2] prediction, that affects many areas like, agriculture, tourism to name a few[4].

A study was conducted by Olaiya and Adeyemo [2] to predict the weather and climate change– significant application in meteorology, by utilizing artificial intelligence along with the data mining techniques. Data mining utilizes many analytical tools, as mentioned earlier and aids in pattern recognition, relationship identification and valid predictions. Rainfall prediction model was developed by Joseph and Ratheesh [5] using data mining technique like classification and clustering techniques their rain forecasting model is depicted in figure 1.

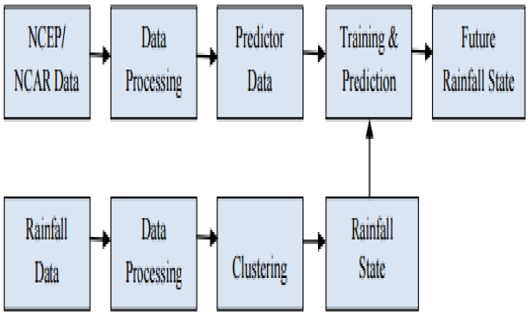

Figure 1 A Rainfall prediction model

Source: J. Joseph and R. T. K (2013) [5]

A weather forecasting model was proposed to analyse the ever fluctuating weather conditions by Yadav and Khatri [6] by using data mining techniques to recognize patterns from historical data collected as well as the current climate conditions in order to predict the upcoming weather conditions. Two main approaches to predict the weather are Empirical Method which is based on historical data analysis and analogues events, used in small scale prediction; and Dynamical Method is generally used for large scale forecasting and often computer modelling is used in this method [2], [6].

B. Introduction to popular data mining tools

Since 1960, data mining technique has been incorporated as a branch of applied artificial intelligence and has explosively grown with the developing technologies to use information intelligently. Data mining techniques used numerous types of databases, such as, relational, spatial, transactional and such. There are many data mining methods, such as characterization generalization, classification, clustering, data visualization association, pattern matching and more[1].

The data mining applications can be broadly classified be classified in two types, the first is descriptive data mining and analysis for analysing properties of existing data; and second application is predictive data mining that includes statistical analysis on available data to make predictions [2]. Data mining techniques that are most commonly used includes, Neural networks that are nonlinear predictive; genetic algorithm architecture that uses natural evolution concept in optimization techniques, decision trees for classification, dynamic prediction models, intelligence systems for modelling, knowledge-based systems, rule induction for extracting data with statistical properties, data visualizationfor correlating multi-dimensional data and many more[1], [2],[3],[6].

II. LITERATURE REVIEW

Change in weather is one of the crucial factors which influence day to day life, to an extent it influence the economy of any geographical area that relies on occupations such as agriculture [7]. Since last decade weather prediction has occurred as one of the most challenging job for meteorological departments of all around the world, even in the presence of advance scientific and technological tools [8]. In order to mitigate the damage and other challenges to economy there are some major data mining techniques to forecast weather conditions in future such as Decision Tree, Neural Networks, K-means clustering analysis, probability modelling and others [8], whose utility has been established through empirical reviewed f scholarly works.

G. C. Onwubolu, P. Buryan, S. Garimella, V. Ramachandran, V. Buadromo, and A. Abraham (2007) purported to understand the effectiveness of data mining techniques in weather forcasting has presented self-organizing modelling known as enhanced e-GMDH (Group Method Data Handling) to be an effective techniques to forecast weather condition [9]. It mines weather data using statistics and Neural Networks and provides accuracy in weather forecasting. Similarly, Olaiya and A. B. Adeyemo (2012) also investigated the use of Neural networks and Decision tree algorithm in predicting highest level of temperature, average rainfall, speed of the wind and evaporation. Their research found that Artificial Neural Networks and Decision Tree algorithms are efficient in detecting the relationship between weather parameters and predicting the future condition of weather even in the presence of complexity in the weather data [2]. Another study conducted by D. Chauhan, Shimla, and J. Thakur (2014) on weather attributes such as wind speed, humidity, temperature, rainfall and others has shown consistency with the results derived by F. Olaiya and A. B. Adeyemo, and further shown that CART using decision tree analysis, and K-means Clustering predict the possible weather condition with more accuracy as it solves the problem quickly by deleting inappropriate data [10]. Tending towards K-means clustering using Hidden Markov Model on JAVA Technology, K. Shrivastava, R. Wakle, and M. Nakade (2015) have shown that Hidden Markov Models provide highly reliable results than ID3 data mining technique, which by using probability modelling can estimate the weather condition for several days in advance through pattern recognition [12]. Another research based on Hidden Markov model , was performed by R. K. Yadav and R. Khatri (2016), where the authors determined that Hidden Markov model provides highly accurate and efficient prediction of the future condition of the weather and data modelling using K-means clustering technique for extraction of patterns and outliers [6].

Other tools and applications have also been used by researchers in the past to predict weather patterns. A research on data mining application in weather forecasting specifically rainfall using ***FPGA (Frequent Pattern Growth Algorithm)*** carried out by A. A. Taksande and P. S. Mohod (2015) has shown that FPGA predicts the weather condition better than traditional Neural Networks with an accuracy of more than 90% [11].

Therefore, difference in highlighting different data mining techniques over the period of years, in terms of accuracy and efficiency has been observed, although Neural Networks and Decision Tree Algorithm are clearly the most popular ones.

Based on the relevance and limitations of these data mining techniques used for weather forecast, defined broadly in the existing literature, a detailed discussion has been presented in the proceeding section.

III. RESULTS AND DISCUSSION

A. Features of Data Mining tools in Weather Prediction

Data Mining is a powerful technology that has great scope in scrutinizing and predicting significance data from databases. As Meteorological data collected for weather forecasting, it is voluminous, complex, dynamic and multi-dimensional [13]. Some of the desired features of the data mining tools used in weather prediction are discussed in this section. As the data mining process is dependent on the available data, it is vital that data collected should be complete and precise for accurate predictions, if the data available is incomplete then the results will be inaccurate. Thus, it is imperative that the data mining technique used must be intelligent enough to recognize incomplete data[1]. Such techniques could be visualization techniques for correlating the multi-dimensional data, Self-Organizing-Maps for complex data and suchothers [14]. An illustration of self-organizing map and multidimensional matrix is depicted in figure 2 and figure 3 respectively.

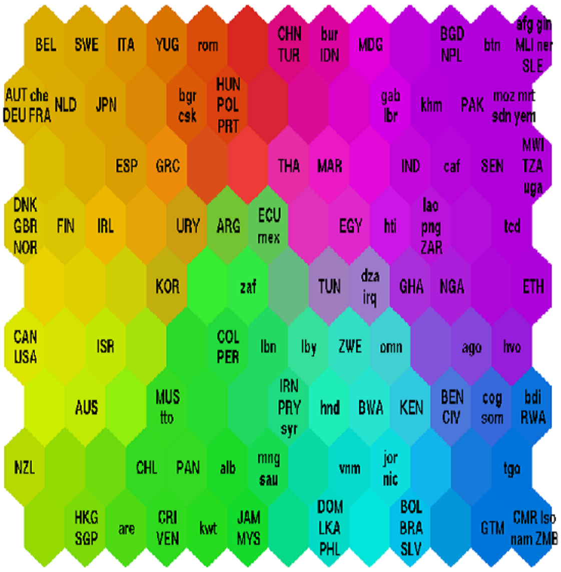

Figure 2: An illustration of self-organizing map

Source: Kevin Pang (2003) [15]

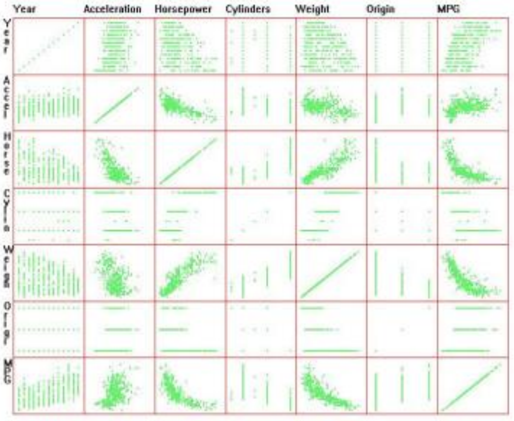

Figure 3: An illustration of multi-dimensional matrix

Source: W. Peng, M. O. Ward, and E. A. Rundensteiner (2004) [16]

The data mining technique have to be able to manage real time data and co-ordinate data from various databases as well as should have inbuilt algorithms for pattern recognition in available databases for accurate prediction[3]. The data mining technique must be able to handle text analysis in various formats available like, word files, pdf files, presentation files to name a few. It aids in recognizing redundancies in the data available[14]. The data mining techniques must also be interactive by using graphical analysis, image classification, multi-dimensional statistical assessment, and capable of classifying the outcomes as per the grouping i.e., in the case of weather forecasting it could be sunny, cloudy or rainy[14]. It is desirable that thetechniques used have to be scalable that can be used for short term predictions as well as long term predictions. Time efficiency, independence and cost effectiveness is also desirable[1].

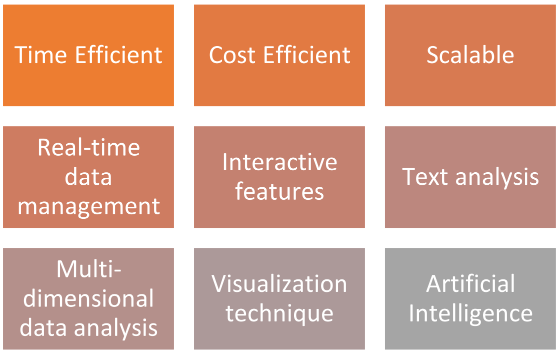
Time Efficient
Cost Efficient
Scalable
Real-time data management
Interactive features
Text analysis
Multi-dimensional data analysis
Visualization technique
Artificial Intelligence

Figure 4: Features of Data Mining tools in Weather Prediction

B. Comparison of data mining tools on weather forecasting

A decision tree is a decision support tool for data mining that is easy to interpret and understand. Classification and Regression Trees (CART) based on the concept of decision tree was used in weather forecasting by Petre[3]. This technique involved rules based collection of dataset and statistical information regarding the data which was generated during modelling. Using this technique, an average temperature of future months could be accurately predicted [3]. The main advantage of this technique is that it is a simple prediction model and has graphical interactive properties, while its main shortcoming lies in exhaustive search approach used [17].

Forecasting weather condition is challenging as it requires analysis of nonlinear and multidimensional data set, therefore a data mining approach using Frequent Pattern Growth Algorithm for weather prediction was performed by Taksande and Mohod[11]. These authors have combined this algorithm with five general algorithms, namely, neural network, random forest, support vector machine, classification and regression tree and k-nearest neighbour the work flow model is illustrated in figure d [11]. The On Frequent Pattern Growth Algorithm (FPGA) was used in this method for recognizing and deleting inappropriate data and the outcomes reflecting that combination of FPGA and general algorithm has 90% accuracy in the weather prediction [11]. The main advantage of this approach is that it is scalable i.e. can be used for different population sizes and cross probabilities. On the other hand, the shortcoming is that it is very time consuming due to tedious workload [18].

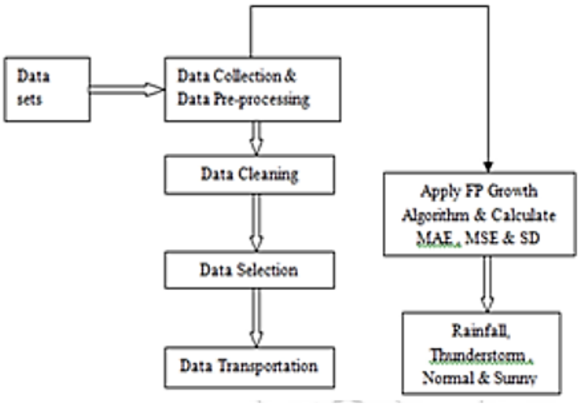
Data sets
Data Collection & Data Pre-processing
Data Cleaning
Data Selection
Data Transportation
Apply FP Growth Algorithm & Calculate MAE, MSE & SD
Rainfall, Thunderstorm, Normal & Sunny

Figure 5: A weather forecasting model using FP Growth Algorithm

Source: A. A. Taksande and P. S. Mohod (2013) [11]

Weather forecasting model was developed by Olaiya and Adeyemo[2] for forecasting rain, average temperature, evaporation rate and wind speed, using Decision Tree algorithms and Artificial Neural Network. The meteorological data collected and classified using classifier algorithm as well as was compared using standard performance measures [2]. The results obtained from this model and actual weather data was compared on a small scale and they were observed to be fairly accurate [2]. The main advantage of Artificial Neural Network is that it can process various data parameters, memory components and more to recognize hidden pattern in them, whilst the shortcomings are complexity of the technique and requirement of very accurate data for accurate prediction [19].

A self-organizing approach of data mining is enhanced Group Method of Data Handling (e-GMDH) that was used to predict daily temperature and pressure as well as monthly rainfall by Onwubolu et.al. [9]. The outcomes suggested that temperature predicted had absolute difference error of +/- 1.5 and the monthly rainfall prediction was similar to other researches [9]. The main advantage is that this technique can process large amount of data whereas over-fitting and poor generalization is the shortcoming of this technique [20].

Yadav and Khatri [6] proposed weather prediction model using the K-means clustering with the Hidden Markov Model for data extraction of the weather condition observations; and the proposed technique was performed on JAVA technology as shown in figure 6. When it was compared with traditionally used ID3 algorithm, it showed enhanced result in terms of accuracy, error rate and space-time complexity [6]. The main advantage of K-means clustering is that it uses hierarchical clustering that is time efficient whereas the shortcoming is that it is not much scalable [21].

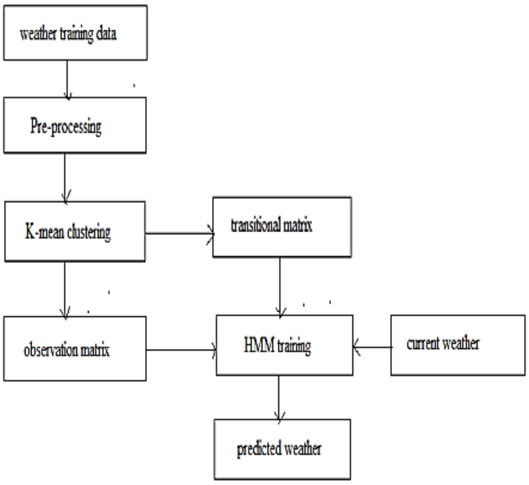

Figure 6: A weather forecasting technique using K-mean clustering

Source: R. K. Yadav and R. Khatri (2016) [6]

Shrivastava, Wakle and Nakade [22] had used Hidden Markov Model for weather prediction and observed that it could be used to predict the short term as well long term weather condition. The main advantage of Hidden Markov Model is very flexible model that can recognize hidden data and patterns whilst the shortcoming is it is very complex and has more time and memory requirement [23].

A comparative analysis of the aforementioned data mining techniques, their advantages, shortcoming and outcomes are presented in tabular form in table 1.

IV. CONCLUSION

Accurate weather data modelling and prediction model using numerous data mining techniques has been studied by several researchers, as data mining techniques can process large amount of multi-dimensional data to recognize link between them and recognize hidden pattern. This paper has given an overview on the data mining techniques and desirable attributes requisite for data mining tools used in weather forecasting.

This paper includes an empirical review on the previous studies conducted on application of various data mining tools in weather condition prediction. The author has performed a comparison between six data mining techniques, namely, Classification and Regression Trees (CART), Frequent Pattern Growth Algorithm (FPGA), Artificial Neural Network, enhanced Group Method of Data Handling (e-GMDH), K-mean clustering and Hidden Markov modelling. These techniques were assessed mainly, on the basis of their advantages and shortcomings. From the comparative analysis conducted taking the six data mining techniques, decision tree within the Artificial Neural Network can be considered to the optimal technique in terms of cost, time saver and prediction. Artificial Neural Network is known to exhibit 87% accuracy in prediction and 98% of precision, making it the most efficient among others. The author has also recognized that the weather prediction models can be further improved by using better classification and characterizing techniques, that could be done by using more accurate and large scale data collected over extended period of time.

Table 1:
A comparative analysis for various data mining techniques

Dimensions Tools	Objective based on the corresponding application	Attributes the tools forecast	Technique used	Advantages	Shortcomings	Outcome
CART [11]	Hourly rainfall study [24], Forecasting rainfall [25], Weather Prediction [26]	Temperature, Humidity, Wind Direction, Speed, Pressure	Decision tree	Simple and graphically interactive [3]	Exhaustive [3]	93% accuracy in prediction [24], less time consuming [25], Accurate average temperature prediction [11]
FPGA [13]	Weather forecasting [27]	Temperature, Humidity, Wind speed, Sea-level Pressure, Rainfall [27]	FPGA combined with five general algorithm,	Scalable [12]	Very time consuming [12]	90% accuracy in prediction [13], efficient in finding frequent pattern [27]
Artificial Neural Network [2]	Rainfall Prediction [28], Weather forecasting [29]	(Relative Humidity, Pressure, Temperature, Precipitable water, Wind speed) [28], (Monsoon Condition, Crop yield, and soil fertility) [29]	Decision Tree algorithms and Artificial intelligence	More processing capabilities [14]	Complexity [14], Time consuming [28]	(Accuracy 87%, Precision 98%,)[28], Applied on small scale [2], Identification of multiple layers of neurons [29]
e-GMDH [9]	Weather Forecasting [30], Weather forecasting [31]	Daily temperature, daily pressure and monthly rainfall,	Self-organizing data mining, Classification, clustering, time series forecasting, Sequential patterns,	Process large amount of data [20]	Over-fitting and poor generalization [20]	Slight difference errors [9], efficient in data cleansing and finding frequent pattern [30]
K-means clustering [6]	Weather Prediction [32], Weather Prediction [10]	Humidity, temperature, Rainfall	With HMM and based on JAVA technology	Hierarchical clustering [21]	Less scalability [21]	(Better Predictions, Outliers identification) [32], High accuracy in prediction[10]
Hidden Markov Model [22]	Daily rainfall [33], [34]	Daily Rainfall, Fog, Sun prediction	Probability modeling	Flexibility, recognition of data and patterns [23]	Complexity and time and memory constraints [23]	Applicable on Short term and long term prediction [33], [34][22]

REFERENCES

[1] Shu-Hsien Liao, Pei-Hui Chu, and Pei-Yuan Hsiao, 2012 Data mining techniques and applications – A decade review from 2000 to 2011..

[2] F. Olaiya and A. B. Adeyemo, 2012, Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, Inf. Eng. Electron. Bus., vol. 1, no. 1, pp. 51–59.

[3] E. G. Petre,2009.A Decision Tree for Weather Prediction, Bul. Univ. Pet. – Gaze din Ploiești, vol. LXI, no. 1.

[4] S. N. Kohail and A. M. El-Halees,2011. Journal of Information Technology Implementation of Data Mining Techniques for Meteorological Data Analysis," vol. 1, no. 3.

[5] J. Joseph and R. T. K,2013. Rainfall Prediction using Data Mining Techniques, Int. J. Comput. Appl., vol. 83, no. 8, pp. 975–8887.

[6] R. K. Yadav and R. Khatri, 2016. A Weather Forecasting Model using the Data Mining Technique, Int. J. Comput. Appl., vol. 139, no. 14, pp. 4–12.

[7] F. Sheikh, S. Karthick, D. Malathi, J. S. Sudarsan, and C. Arun, 2016. Analysis of Data Mining Techniques for Weather Prediction," Indian J. Sci. Technol., vol. 9, no. 38, pp. 1–9.

[8] M. A. Mandale, M. Jadhawar, and D. D. Aher,2015. Weather forecast prediction: a Data Mining application, Int. J. Eng. Res. Gen. Sci., vol. 3, no. 2, pp. 1279–1284.

[9] G. C. Onwubolu, P. Buryan, S. Garimella, V. Ramachandran, V. Buadromo, and A. Abraham, "Self-organizing data mining using Enhanced groupd method data handling approach,2007. in IADIS European Conference Data MIning, pp. 81–88.

[10] D. Chauhan, Shimla, and J. Thakur, "Data Mining Techniques for Weather Prediction: A Review | International Journal IJRITCC - Academia.edu," Int. J. Recent Innov. Trends Comput. Commun.,2014 vol. 2, no. 8, pp. 2184–2189.

[11] A. A. Taksande and P. S. Mohod, 2015. Applications of Data Mining in Weather Forecasting Using Frequent Pattern Growth Algorithm, Int. J. Sci. Res., vol. 4, no. 6, pp. 3048–3051.

[12] K. Shrivastava, R. Wakle, and M. Nakade,2015. Weather Prediction Using Hidden Markov Model," SSRG Int. J. Electron. Commun. Eng. MIT Aurangabad, India), vol. 2, no. 3, pp. 61–63.

[13] A. Gayathri, M. Revathi, and J. Velmurugan,2016.A survey on Weather forecasting by Data Mining, Int. J. Adv. Res. Comput. Commun. Eng., vol. 5, no. 2.

[14] Gopinadh Gulipalli, 2015. 12 Data Mining Tools and Techniques : Invensis Blog, .

[15] Kevin Pang, "Self-organizing Maps," 2003. .

[16] W. Peng, M. O. Ward, and E. A. Rundensteiner,2014.Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering Suggested Citation Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering Clutter Reduction in Multi-Dimensional Data Visualization Using Di,.

[17] W.-Y. Loh, 2011.Classification and regression trees.

[18] V. Kitowski and G. Gergi, 2016.E-Business Models - The Physical Touchpoint of Online Retailers in Business Model Frameworks,".

[19] C. Dumitru and V. Maria, 2013.Advantages and Disadvantages of Using Neural Networks for Predictions," Ovidius Univ. Ann. Econ. Sci. Ser., vol. XIII, no. 1, pp. 444–449.

[20] G. C. Onwubolu, P. Buryan, and A. Abraham,2007. SELF-ORGANIZING DATA MINING USING ENHANCED GROUP METHOD DATA HANDLING APPROACH.

[21] K. Singh, D. Malik, and N. Sharma, 2011.Evolving limitations in K-means algorithm in data mining and their removal. IJCEM Int. J. Comput. Eng. Manag. ISSN, vol. 12, pp. 2230–7893,

[22] K. Shrivastava, R. Wakle, and M. Nakade,2015. Weather Prediction Using Hidden Markov Model," SSRG Int. J. Electron. Commun. Eng. MIT Aurangabad, India), vol. 3.

[23] M. Khadr,2016. Forecasting of meteorological drought using Hidden Markov Model (case study: The upper Blue Nile river basin, Ethiopia), Ain Shams Eng. J., vol. 7, no. 1, pp. 47–56,.

[24] S.-Y. Ji, S. Sharma, B. Yu, and D. H. Jeong,2012.Designing a rule-based hourly rainfall prediction model," in Information Reuse & Integration (IRI), pp. 303–308.

[25] D. Gupta and U. Ghose, "A Comparative Study of Classification Algorithms for Forecasting Rainfall," IEEE. G.G.S Indraprastha University, 2015.

[26] S. S. Bhatkande and R. G. Hubballi, 2016.Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques, Int. J. Adv. Res. Comput. Commun. Eng., vol. 5, no. 5, pp. 483–487.

[27] A. A. Taksande and P. S. Mohod, 2013.Applications of Data Mining in Weather Forecasting Using Frequent Pattern Growth Algorithm," Int. J. Sci. Res. ISSN (Online Index Copernicus Value Impact Factor, vol. 14, no. 6, pp. 3048–3051.

[28] J. Joseph and R. T. K, Rainfall Prediction using Data Mining Techniques,2013. Int. J. Comput. Appl., vol. 83, no. 8, pp. 11–15.

[29] M. P. Shivaranjani and K. Karthikeyan, 2016.A Review of Weather Forecasting Using Data Mining Techniques, Int. J. Eng. Comput. Sci. Ms.P.Shivaranjani, vol. 5, no. 12, pp. 19784–19788.

[30] G. C. Onwubolu, P. Buryan, S. Garimella, V. Ramachandran, V. Buadromo, and A. Abraham,2007.Self-organizing data mining for weather forecasting, in Self-organizing data mining for weather forecasting, pp. 81–88.

[31] M. S. V. Shingne, P. A. D.Warbhe, and P. S. Dubey, 2015.International Journal on Recent and Innovation Trends in Computing and Communication Weather Forecasting using Adaptive technique in Data Mining, Int. J. Recent Innov. Trends Comput. Commun., vol. 3, no. 5, pp. 91–95.

[32] P. Kalaiselvi, D. Geetha, and M. P. Scholar,2016.Weather Prediction Using J48, EM And K-Means Clustering Algorithms, Int. J. Innov. Res. Comput. Commun. Eng., vol. 4, no. 12, pp. 20889–20895.

[33] A. W. Robertson, S. Kirshner, and P. Smyth, 2003.Hidden Markov models for modeling daily rainfall occurrence over Brazil, Irvine.

[34] E. Fosler-Lussier, 1998. Markov Models and Hidden Markov Models: A Brief Tutorial. International Computer Science Institute, Berkeley, pp. 1–7.

[i] Correspondence Address

[ii] Correspondence Address

[iii] Correspondence Address