

COMPARISON OF RAINFALL PREDICTION BASED ON ALMANAC USING DATA MINING TECHNIQUES

K.S. Mohanasundaram

Research Scholar, Department of Computer Science,
Chikkanna Government Arts College, Tiruppur, Tamil Nadu, India.
Assistant Professor in Computer Science and Application, SRM Institute of Science and Technology,
Chennai, Tamil Nadu, India.

Dr. G. M. Nasira

Assistant Professor & Head, Department of Computer Applications,
Chikkanna Government Arts College, Tiruppur, Tamil Nadu, India.

Abstract- Rainfall becomes a major factor in agricultural based country like India. Rainfall prediction has become one of the most systematically and technically taxing problems in around world. Farmers in Tamil Nadu are still following the agronomic activities based on astrological facts of Panchangam (Almanac). Yet there is very few ever attempted to see the rationality of the ancient knowledge system. Almanac also has a mathematical base for predicting the meteorological occurrences. During the study, the rainfall prediction by one of the traditional Almanac is studied in concentration for one cycle of 60 Tamil year's corresponding to the Gregorian Year from 1957 to 2017. We have a lot of data mining techniques to extract information. In this work, we applied various classification algorithms such as SMO, Random Forest and REPTree on the almanac rainfall dataset in WEKA tool. This paper shows REPTree is best for prediction of rainfall using data mining techniques based on almanac.

Keywords-Rainfall Prediction, Almanac, Prediction, Classification.

1. INTRODUCTION

Data mining is the search and analysis of large data sets, in order to discover meaningful patterns and rules. The key idea is to find effective ways to combine the computer's power to process data with the human eye's ability to detect patterns. Data mining techniques have been broadly applied almost in all fields to analysis the data for pattern the rules, classification, prediction, decision trees, fuzzy rules and so on.

Rainfall is important for planning the activities of agriculturists, builders, water supply engineers, and all activity plans in the nature. India is an agricultural country and its economy is largely based upon crop productivity. Thus rainfall prediction becomes a significant factor in agricultural based countries like India. Rainfall Prediction is one of the most challenging tasks. Though already many algorithm have being proposed but still accurate prediction of rainfall is very

difficult. In an agricultural country like India, the success or failure of the crops and water scarcity in any year is always viewed with greatest concern.

Astronomy is an area where Data Mining has been playing a big role. Several techniques of Data mining have been used to solve tasks in Astrology. There has been increasing research interest in use of data mining techniques to scrutinize in the Astrology area.

At present the Meteorology Department is informing only short term forecasting about weather but long term forecasting is needed for planning. This can be achieved by two methods namely traditional forecasting and scientific weather forecasting. Traditional forecasting is based on observations and experience using combinations of plants, animals, insects, meteorological and astronomical indicators, and almanacs or panchangs over a period of time. The scientific weather forecasting is based on past records of climate prevailed in the area using mathematical models.

2. EXISTING APPROACH

Kolluru Venkata Nagendra [1] surveys a range of classification techniques used by various researchers. Artificial Neural Network is applied for Rainfall Forecasting on various parameters are analyzed. They found that MLP method, Naïve Bayesian classifiers and Support Vector Machines are best to predict rainfall compare to other techniques (Numerical & Statistical). They identify that for weekly, monthly and yearly rainfall forecasting Naïve Bayesian, Feed Forward Neural Network and SVM gives best performance respectively.

Dhawal Hirani [2] reports a detailed survey on rainfall prediction using different rainfall prediction methods extensively survey lasted 20 years. From the survey it has been found that most of the researchers used artificial neural networks for rainfall prediction and got significant results. They found that MLP, BPN, RBFN, SOM & SUM are suitable for predict rainfall forecasting techniques.

Seema Mahajan [3] examined the relationship of Gujarat rainfall with significant universal parameters such as SLP, SST, U – Wind & Windstress and V- Wind & Windstress. They taken one month lagged (June – July) for 40 years (From 1960 to 1999) data from National Oceanic and Atmospheric Administration (NOAA) and perform multilinear

regression on the generated and measured rainfall series. They found 0.8377 is the correlation coefficient between generated and measured rainfall series.

B. Kavitha Rani [4] applied ANN to predict the summary rainfall data in Thailand. And found that back propagation gives accuracy result. Valmik B Nikam [5] extract the IMD (Indian Meteorological Department) Pune, weather data comprising of 36 attributes, only 7 attributes are relevant to rainfall prediction is taken and used Bayesian approach and got accuracy result. Jyothis Joseph [6] used clustering and classification techniques for prediction of rainfall and got the accuracy 87%. M. Kannan [7] used Multiple Linear Regression Model for rainfall prediction. They got approximate value not accurate value. Parneet Kaur [8] & M. Sivasakthi [9] found that Multi Layer Perception gives the best accuracy in EDM (Educational Data Mining) among Naive Bayes, SMO, J48 and REPTree classifications.

N. Vivekanandan [10] [22] applies ANN based MLP and RBF for AER of Joshimath and Tohana rain-gauge stations. And found MLP for Joshimath and RBF for Tohana is suitable for AER. A. Subasini [11] explore the applicability of data mining technique to predict the breast cancer. And analyzes the performance of C5.0, ID3, APRIORI, C4.5 and Naïve Bayes algorithms. Experimental found C5.0 gives highest accuracy. Ozlem Terzi [12] used to estimate monthly rainfall values of Isparta. The monthly rainfall data of Senirkent, Uluborlu, Egirdir and Yalvac stations are taken. The best appropriate algorithm is multilinear regression & it gives relative error is 0.7%. M Ramzan Talib [13] collected weather data for 10 years from 2007 to 2016 at Faisalabab city, Pakistan and applied K-means clustering algorithm and Decision Tree algorithm for these data. Sarah N. Kohail [14] applied knowledge discovery process to take out knowledge from Gaza city weather data for 9 years from 1977 to 1985. Outlier analysis, prediction, classification, association and clustering data mining techniques applied. Harneet Kaur [15] studied an overview of different techniques and tools of data mining such as KNIME, Orange, RapidMiner, Tanagra and WEKA. And identified the challenges in health care domain. An implementation was shown in Tanagra tools for Classification and Visualization methods. Godfrey C. Onwubolu [16] used enhanced Group Method of Data Handling (e-GMDH) which uses the daily pressure & temperature and monthly rainfall and gives the good experimental results. Sweta [17] suggested to improving quality of service by properly managed security concerns.

Divya Chauhan [18] reviewed the different algorithms and techniques in predicting various weather phenomena like rainfall, thunderstorms and temperature. Then comparison is done between the techniques and found that decision trees and k-mean clustering gives the best results. Dhananjay P. Atole [19] computed the rainfall prediction for 5 India cities (Chennai, Delhi, Mumbai, Nagpur and Pune) using 7 years data of rainfalls daily, weekly, monthly. Accuracy result is got by using Multi variable polynomial regression (MPR) technique. Siddharth S. Bhatkande [20] used meteorological data from 2012 to 2015 for various cities and Decision tree algorithm for classification of weather parameters such as minimum & maximum

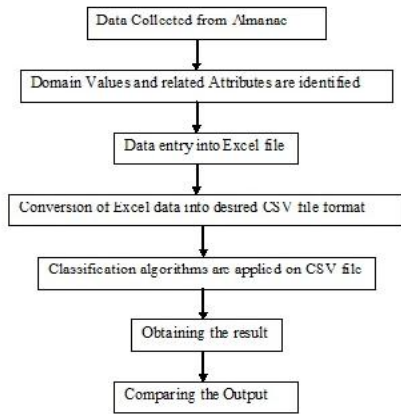
temperature of the data. They proved decision tree is best for weather prediction.

Norraseth Chantasut [21] computed the rainfall prediction for monthly from historical rainfall data from 1941 to 1999 from 245 rainfall monitor stations in Thailand around Chao Phraya River using ANN in which the number of training pattern is 372 and testing pattern is 96. The Neural Network gives 99.6% and 96.9% of accuracy of training and testing data respectively. R. Sukanya [23] compared several classification algorithms CART, C4.5, ID3, Back propagation and SVM. Finally concludes that nowadays almost researches using hybrid method for getting more accuracy results.

D Angchok [24] predict rainfall by Tibetan astrological theories with meteorological predictions was accepted. They suggested as very few scientific studies have ever been conducted in ancient Astro-science and almost all of them have reported encouraging and positive outputs, there seems to have enormous scope lying in studying ancient sciences, especially Astro - disciplinary approaches. S Sivaprakasam [25] suggested the traditional methods of forecasting rainfall may be riddled with in accuracies but they cannot be ignored altogether. R. Raja [26] analysis 90 years (1909-1999) historical annual rainfall data of Coimbatore correlation with a particular Tamil year cycle with fourth coming Tamil cycle years. Pankaj S. Kulkarni [27] deals with converting ancient principles related to astrology into predictions using data mining techniques. Neelam Chaplot [28] taken total 102 records, an half of the records were of persons are doctor and other half records of are not doctor by Profession,. They compared various Supervised classification techniques such as Logistic, Naïve Bayes, Simple Cart, Decision Stump, Decision Table and DTNB algorithm. The better results were produced by simple logistic with 12 fold cross validation with an accuracy of 54.902%. Decision Stump algorithm with 14 fold classification gave results with an accuracy of 50%. S. R. Gedam [29] analyzed five data mining algorithms such as Bayesian, Decision table, Multilayer Perception, Random Decision Tree, and Random Forest. And got the accuracy result 84.7458%, 98.3051%, 94.9153,99.0202 and 100% respectively. Concluded that Random Forest Method is the best classification method. Rahul Shajan [30] analyzed the several data mining algorithms to correctly classify and predict health of a human being. They got the results 81.25 % accuracy for J48, J48 graft and Naïve Bayes algorithms and 93.75% accuracy for Random forest algorithm.

3. PROPOSED APPROACH

Experimental research methodology has been adopted for this work. Through the extensive search of literature and discussion with experts, the number of attributes which influencing the rainfall has been finalized. For this work, data are collected from particular almanac or panchang. This data is then filtered out using manual techniques. Then data is transformed into a standard format required by the WEKA tool.



A record of 60 years data from year 1957 to year 2017 from almanac has taken for analysis. From an panchang or almanac we have considered five influencing attributes for rainfall are King, Minister, Megathipathi, Megam (Cloud Type), Rainfall value (Marakkal) each of which has sub item sets, which as shown in Table I.

TABLE I. ATTRIBUTES INFLUENCING FOR RAINFALL

S. No.	Attribute	Description	Domain Value
1	King	Ruling Planet of the year	{Sun, Moon, Mars, Mercury, Jupiter, Venus, Saturn}
2	Minister	Minister Planet of the year	{Sun, Moon, Mars, Mercury, Jupiter, Venus, Saturn}
3	Megathipathi	Planet supporting rainfall for the year	{Sun, Moon, Mars, Mercury, Jupiter, Venus, Saturn}
4	Megam(Cloud Type)	Type/Formation of the Cloud	{Aavarta, Samvarta, Pushkara, Drona, Kaala, Neela, Varuna, Vayu, Dharmo}
5	Almanac Rainfall	Rainfall of the year as per Almanac	{ Kuruni, Pathaku, Mukkuruni, Thooni }

In this work various data mining techniques are used to predict rainfall. WEKA is used to apply the classification techniques and for predictions. The output has been analyzed with three classification algorithms such as SMO, RandomForest and REPTree.

4. RESULT & DISCUSSION

TABLE II. PERFORMANCE OF VARIOUS CLASSIFICATION ALGORITHMS

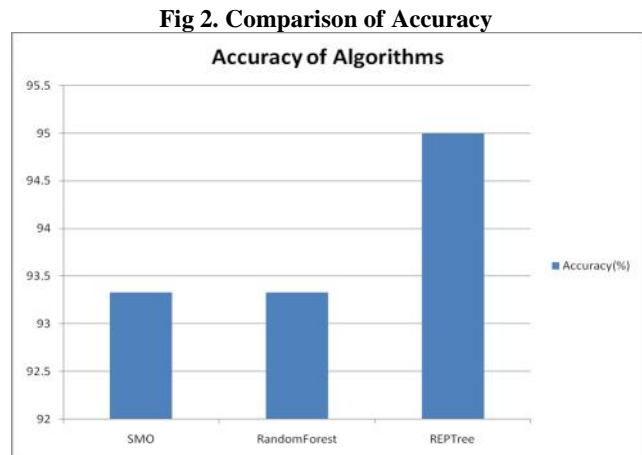
	SMO		Random Forest		REPTree	
	Precision	Recall	Precision	Recall	Precision	Recall
Kuruni	0.875	0.933	0.929	0.867	0.933	0.933
Pathaku	1	0.857	1	0.857	1	0.857
Mukkuruni	0.947	0.947	0.900	0.947	0.947	0.947
Thooni	0.947	0.947	0.950	1	0.950	1
Weighted Average	0.935	0.933	0.935	0.933	0.951	0.950

We got the results by tested and analyze with three data mining classification algorithms such as SMO, RandomForest and REPTree that shows in above Table II. The correct accuracy of all the algorithms is given below in Table III.

TABLE III. ACCURACY OF CLASSIFICATION ALGORITHMS

Data Mining Algorithm	Accuracy (%)
SMO	93.33
RandomForest	93.33
REPTree	95

Both SMO and RandomForest algorithm got accuracy 93.33%. The best accuracy is 95% performed by REPTree algorithms. The following chart shows the performance accuracy of algorithm.



5. CONCLUSION

In this paper, Rainfall predicting attitudes and data sets are taken for cycle of 60 Tamil year's related to the Gregorian Year from 1957 to 2017 from Almanac. For this data set three data mining classification algorithms such as SMO, RandomForest and REPTree was applied using in WEKA tool. We get more accuracy result 95% for REPTree classification data mining algorithm. So, the existing REPTree algorithm is sufficient to find the similar patterns in the almanac rainfall predictions.

REFERENCES

- [1] Kolluru Venkata Nagendra, Dr. Maligela Ussenaiah "A Survey On Classification Techniques Used For Rainfall Forecasting", International Journal of Advanced Research in Computer Science, Volume 8, No. 8, September-October 2017, PP-226-229. ISSN No. : 0976-5697.
- [2] Mr. Dhawal Hirani, Dr. Nitin Mishra "A survey on rainfall prediction techniques", Internal Journal of Computer Application, Volume 6- No.2, March-April 2016 (2250-1797)
- [3] Seema Mahajan, Dr. S. K. Vij "Modeling and Prediction of Rainfall Data using Data Mining", International Journal of Engineering Science and Technology, Vol. 3 No.7 July 2011, PP-6799-6804, ISSN : 0975-5462.
- [4] B. Kavitha Rani and A. Govardhan " Rainfall Prediction using Data Mining Techniques – A Survey" CS & IT-CSCP 2013, PP. 23-30.
- [5] Valmik B Nikam, B. B. Meshram "Modeling Rainfall Prediction Using Data Mining Method A Bayesian Approach" 2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation, PP-132-136.
- [6] Jyothis Joseph, Ratheesh T K "Rainfall Prediction using Data Mining Techniques", International Journal of Computer Applications (0975-8887), Volume 83 – No. 8, December 2013, PP-11-15.
- [7] M. Kannan, S. Prabhakaran, P. Ramachandran "Rainfall forecasting using data mining Technique", International Journal of Engineering and Technology Vol.2(6), 2010, 397-401
- [8] Parneet Kaur, Manpreet Singh, Gurpreet Singh Josan "Classification and prediction based data mining algorithms to predict slow learners in education sector" Procedia Computer Science 57 , 2015, PP-500-508.
- [9] M. Sivasakthi " Classification and Prediction based Data Mining Algorithms to Predict Students' Introductory programming Performance" IEEE Xplore Compliant – Part Number : CFP17L34-ART, ISBN : 978-1-5386-4031-9.
- [10] N. Vivekanandan "Modelling of Annual Extreme Rainfall Using MLP and RBF Networks", BEST: Journal of Management, Information Technology and Engineering, Vol.2, Issue 2, Dec 2016, 13-22.
- [11] A. Subasini, Nirase Fathima abubacker, Dr. Rekha "Analysis of classifier to improve Medical diagnosis for Breast Cancer Detection using Data Mining Techniques", International Journal Advanced Networking and Applications, Volume:5 Issue:6 Pages : 2117-2122 (2014) ISSN : 0975-0290.
- [12] Ozlem Terzi "Monthly Rainfall Estimation Using Data-Mining Process", Applied Computational Intelligence and Soft Computing, Volume 2012, Article ID 698071, Pages: 1- 6.
- [13] M Ramzan Talib, Toseef Ullah, M Umer Sarwar, M Kashif Hanif and Nafees Ayub "Application of Data Mining Techniques in Weather Data Analysis, International Journal of Computer Science and Network Security, VOL.17 No.6, June 2017. PP- 22-28.
- [14] Sarah N. Kohail, Alaa M. El-Halees "Implementation of Data Mining Techniques for Meteorological Data Analysis (A case Study for Gaza Strip), Volume 1 No.3, July-2011, PP-96-100, ISSN-2223-4985.
- [15] Harneet Kaur "A comprehensive study of Data Mining techniques in prediction of Cancer", Journal of Advanced Computing and Communication Technologies, Volume No. 5 Issue No. 3, June 2017, PP-81-88, ISSN : 2347-2804.
- [16] Godfrey C. Onwubolu, Petr Buryan, Sitaram Garimella, Visagaperuman Ramachandran, Viti Buadromo and Ajith Abraham "Self-Organizing Data Mining For Weather Forecasting", IADIS European Conference Data Ming 2007, PP-81-88.
- [17] Sweta "Need and Applications of Data Mining", International Journal of Novel Research in Computer Science and Software Engineering, Vol. 2, Issue 2, PP:39-45, Month: May-August 2015, ISSN 2394-7314.
- [18] Divya Chauhan, Jawahar Thakur "Data Mining Techniques for Weather Prediction: A Review", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2, Issue : 8, PP-2184-2189, ISSN : 2321-8169.
- [19] Dhananjay P. Atole, Deepak Kapgate, Rushi Longadge "Statistical Modeling for Rainfall Prediction using Data Mining Technique", ABHIYANTRIKI An International Journal of Engineering & Technology, Vol.2, No. 1, January 2015, PP-35-39, eISSN : 2394-627X.
- [20] Siddharth S. Bhatkande, Roopa G. Hubballi "Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2016, PP-483-487, ISSN (Online) 2278-1021, ISSN (Print) 2319 5940.
- [21] Norraseth Chantasut, Charoen Charoenjit and Chularat Tanprasert "Predictive Mining of Rainfall Predictions Using Artificial Neural Networks for Chao Phraya River", The 2nd World Congress on Computers in Agriculture and Natural Resources, August 9-12, 2004, Bangkok, Thailand, PP-117-122.
- [22] N. Vivekanandan "Prediction of Rainfall Using MLP and RBF Networks", International Journal Advanced

- Networking and Applications, Volume : 05, Issue : 04, Pages : 1974-1979, 2014, ISSN : 0975-0290.
- [23] R. Sukanya, K. Prabha “ Comparative Analysis for Prediction of Rainfall using Data Mining Techniques with Artificial Neural Network”, International Journal of Computer Sciences and Engineering, Volume – 5, Issue – 6, 2017, PP-288-292, E-ISSN: 2347-2693.
- [24] D Angchok & V K Dubey “Traditional method of rainfall prediction through Almanacs in Ladakh”, Indian Journal of Traditional Knowledge, Vol. 5(1), January 2006, pp. 145-150
- [25] S Sivaprakasam & V Kanakasabai “Traditional almanac predicted rainfall – A case study”, Indian Journal of Traditional Knowledge, Vol. 8(4), October 2009, pp. 621-625.
- [26] R. Raja, T. N. Balasubramanian and R. Karthikeyan “Almanac study on forecasting annual rainfall”, The Madras Agricultural Journal 90(10-12) : 596-600 October-December 2003.
- [27] Mr. Pankaj S. Kulkarni, Dr. S. S. Sane, Prof. N. L. Bhale “Use of Neural Networks in Horoscope prediction”, ResearchGate, Conference Paper March 2012.
- [28] Neelam Chaplot, Praveen Dhyani, O. P. Rishi “Astrological prediction for Profession Doctor using Classification techniques of Artificial Intelligence” International Journal of Computer Applications, Volume 122 – No.15, July 2015 (0975-8887).
- [29] S. R. Gedam, Dr. R. A. Ingolikar “ Application of Mining Techniques to Classification of Star”, International Journal of Advance Research in Computer Science and Management Studies, Volume 3, Issue 9, September 2015, PP-153-157, ISSN : 2321-7782 (Online).
- [30] Rahul Shajan and Dr. Gladston Raj “Health Prediction Astrology using Data Mining Techniques”, International Journal of Advanced Research, Volume 4, Issue 4, 2016, PP-680-683, ISSN: 2320-5407.