International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Association Rule Mining for South West Monsoon Rainfall Prediction and Estimation over Mumbai Station

R. Varahasamy, S. Meganathan, Durga Karthik

Abstract— Rainfall is important for agricultural yield and hence early prediction is required. It has a vital role in the improving the economy of a country. Accurate and timely weather prediction for rainfall forecasting has been one of the most challenging problems around the world as it changes the physical characteristics of the hydrologic system. Rainfall prediction model involves observation of weather data, deriving knowledge from it and implementing using computer models. The proposed work observed rainfall during south-west monsoon months of Mumbai (Latitude 19.0760°N / Longitude 72.8777°E) city. Predictive Apriori Algorithm was used to derive association rules for spot prediction, 24 hours ahead prediction and 48 hours ahead prediction, also to estimate a no rain day, moderate rain day and heavy rain day.

Keywords— Southwest Monsoon, Rainfall, prediction, estimation, Predictive Apriori Algorithm

I. INTRODUCTION

Rainfall forecasting has been one of the most challenging problems around the world as it changes the physical characteristics of the hydrological system. To make an accurate prediction is one of the major challenges, all over the world. Data Mining [1] provides various tools and techniques for weather prediction and estimation.

Association rule [2] can be used on weather data for quantitative prediction of rainfall. Forecasting [3] rainfall can be done for spot,24 hours ahead or 48 hours ahead. Weather day or temperature patterns[4,5] are used for accurate prediction of rainfall that can be used for cultivation of crops. Classification [6,7] algorithms like Bayesian and Naïve Bayesian classifiers were used for predicting rainfall for a region.

Predictive Apriori algorithm [8] can be used to model the given dataset for both prediction and estimation of rainfall for a region. Rule based [9] classification methods provide better accuracy with lesser false positive results. The work is implemented for forecasting and estimating rainfall during southwest monsoon for Mumbai City.

A. Materials and Methods

Mumbai (Latitude 19.0760°N / Longitude 72.8777°E) is located in Maharashtra state of India, which is a metropolitan area with an airport. Rainy days during southwest (SW) monsoon months June, July, August and September has been taken as the parameters for the forecast due to its operational importance. Data used are those recorded by India Meteorological Department (IMD) that is included in World Meteorological Organization (WMO) data bank and managed by the National Oceanic and Atmospheric Administration (NOAA)[10], which is a federal agency focusing on the condition of the oceans and the atmosphere worldwide. Daily rainfall data for the 7 years 2012-2017 were considered for developing the model. Data preprocessing techniques were applied to incorrect, incomplete and duplicated data by removing them from the dataset. After preprocessing the raw data was filtered using discretization approach to provide a range. The filtered data was given as input to Predictive Apriori Algorithm to find interesting (strong) rules for rainfall prediction. Parameters such as mean temperature, humidity, wind speed, visibility, and precipitation were used. Based on the parameters, various predictions that are possible are heavy rainy day, rainy day, and no rainy day. The Predictive Apriori Algorithm was used to develop a weather program and results were compared with the actual weather data.

II. RESULTS AND DISCUSSION

After preprocessing data, 445 records were extracted from the data set. The discretization algorithm formed various best-fit ranges for the five weather circumstances. Based on the discretization algorithm the weather variables were ranged into the nominal values such as low, medium and high with its best ranges using machine learning tool Weka 3.6. The nominal values for the weather predicates are shown in Table 1. A use case diagram depicting the above model is shown in figure 1.

Manuscript published on 30 September 2019
* Correspondence Author
R. Varahasamy*, Department of Computer Science, Srinivasa Ramanujan Centre, SASTRA, Kumbakonam, Tamil Nadu, India. (Email: rvsamy@src.sastra.edu)
S. Meganathan Department of Computer Science, Srinivasa Ramanujan Centre, SASTRA, Kumbakonam, Tamil Nadu, India. (Email: meganathan@src.sastra.edu)
Durga Karthik, Department of Computer Science, Srinivasa Ramanujan Centre, SASTRA, Kumbakonam, Tamil Nadu, India. (Email: durgakathik@src.sastra.edu)

Table 1. Nominal values for predicate data & Results

Predicate Data	LOW	MEDIUM	HIGH
Temperature (F)	<82	82 - 87	>87
Dew Point (F)	<74.5	74.5 – 77.3	>77.3
Visibility (M)	<1.76	1.76-2.63	>2.63
Wind Speed (Knots)	<5.63	5.63 – 10.26	>10.26
Precipitation for prediction(mm)	NO RAIN = 0 RAIN > 0		
Precipitation for estimation(mm)	NO RAIN =0 MODERATE RAIN (0 – 20) HEAVY RAIN >20		

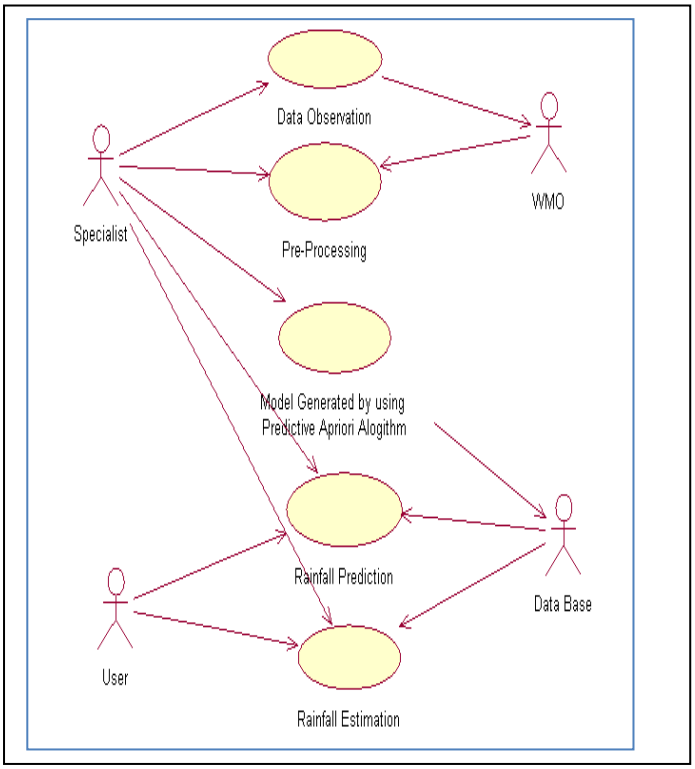

Figure 1: Use case diagram depicting the scenario

The predictive algorithm finds the association and relationships among the weather data. The proposed predictive algorithm for the prediction of rainfall on the mentioned meteorological station works is given below:

Step 1: Scan all weather transactions and find all frequent items towards the class label that have support above required minimum support value.

Step 2: Build potential weather data sets of k weather items from Lk-1 by using pairs of weather item sets in Lk-1 such that each weather data pair has the first k-2 items in common.

Step 3; Scan all transactions and find all k-weather item sets in candidate weather item Ck that are frequent.

Step 4: Terminate when no further weather frequent item sets are found, otherwise continue with Step 2.

The five parameters were subjected to Predictive Apriori Algorithm that yielded the rules for various prediction and estimation of rainfall for spot, 2 hours and 48 hours respectively is shown in Table 2.

The above rules were implemented as a rainfall prediction and estimation system in .Net and a sample screen shot is shown in Figure 2.

The results from the above rainfall prediction system reveals that from 445 instances more than 60% of data were predicted correctly for prediction and estimation of rainfall. Similarly the error is very small and hence negligible. The statistical summary of rainfall prediction result is shown in Table 3.

A detailed summary of prediction and estimation model with respect to various data mining accuracy measures such as true positive rate, false positive rate, precision for all the rules are shown in Table 4.

III. CONCLUSION

The model yielded 4 class labels in estimation and 3 class labels in prediction of rainfall. The weather association rules of machine learning technique has extracted the rainfall patterns over Mumbai station on Arabian sea region. Analyzing the accuracy measure on the class labels shows that precision is nearly 0.7 except for few cases (estimation – 24 hours, 48 hours estimation of heavy, moderate and weighted average) The prediction rules have approximately 0.7 for most of the cases. True positive rate is higher, with false positive rate very low for all rules. Hence the above rules can be used for prediction of rainfall for a region. Accuracy of rules for spot estimation of rainfall is higher hence can be adopted for the day. As future work the rules will be tuned to better estimation of rainfall in 24 and 48 hours category.

IV. ACKNOWLEDGMENT

The authors thank Hon'ble Vice-Chancellor, SASTRA Deemed to be University. Also the authors thank the Department of Science and Technology – Fund for improvement of S & T Infrastructure in Universities and Higher Educational Institutions, Government of India (SR/FST/MSI-107/2015), for Discrete Mathematics Laboratory.

Table 2. Predicate values of weather association rules for rainfall predication and estimation.

Prediction Type	Temperature	Dew Point	Visibility	Wind Speed	Precipitate Prediction	Precipitate Estimation
To day	Medium	Medium	High	Medium	No Rain	No Rain
	Medium	High	Medium	High	Rain	Moderate Rain
	Medium	Medium	Medium	Medium	No Rain	No Rain
	Medium	High	Low	Medium	Rain	Heavy Rain
24 Hours ahead	Medium	Medium	High	Medium	No Rain	No Rain
	Medium	Medium	Medium	Low	Rain	Moderate Rain
	Medium	High	Medium	High	Rain	Moderate Rain
	Medium	Medium	Medium	Medium	No Rain	No Rain
	Medium	High	Medium	Medium	Rain	Heavy Rain
48 Hours ahead	Medium	Medium	High	Medium	No Rain	No Rain
	Medium	Medium	Medium	Low	Rain	Moderate Rain
	Medium	High	Medium	Medium	Rain	Heavy Rain

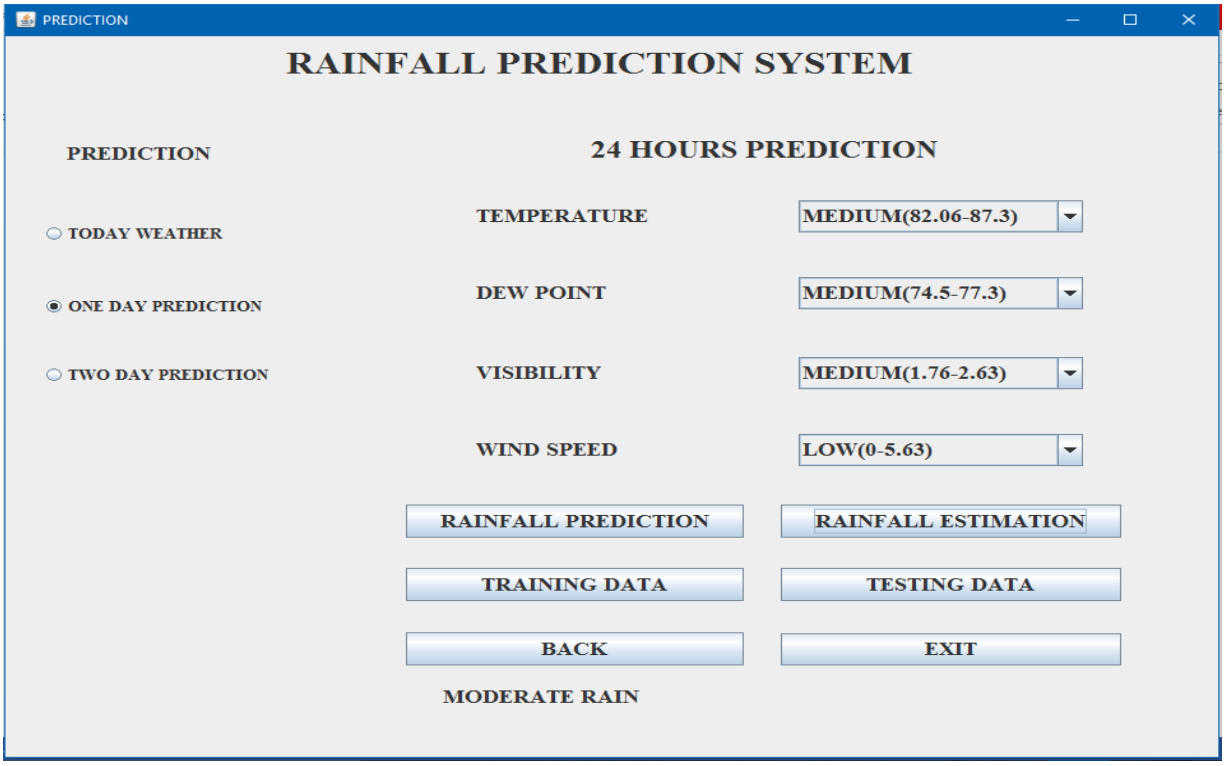

Figure 2: Snapshot of rainfall prediction system

Table 3. Statistical summary of the rainfall prediction.

	Spot prediction	24 Hours Prediction	24 Hours Estimation	48 Hours Prediction	48 Hours Estimation
Total number of instances	445	444	444	443	443
Correctly classified instances in %	63	77	63	75	58
Incorrectly classified instances in %	37	23	37	25	42
Mean absolute error	0.3021	0.2854	0.3142	0.323	0.3463
Root mean squared error	0.3943	0.3918	0.4006	0.4105	0.4198

Table 4. Detailed summary of the classifier model.

Prediction Time	Prediction Type	Class Name	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Today	Estimation	No Rain	0.473	0.06	0.793	0.473	0.592	0.832
		Moderate	0.773	0.484	0.609	0.773	0.681	0.717
		Heavy	0.634	0.277	0.658	0.634	0.628	0.778
		Weighted Average	0.634	0.277	0.658	0.634	0.628	0.778
Today	Prediction	No Rain	0.5	0.074	0.768	0.5	0.606	0.832
		Rain	0.926	0.5	0.791	0.926	0.854	0.832
		Weighted Average	0.787	0.36	0.784	0.787	0.772	0.832
24 Hours	Estimation	No Rain	0.497	0.08	0.75	0.497	0.598	0.812
		Moderate	0.764	0.464	0.618	0.764	0.683	0.724
		Heavy	0.494	0.101	0.513	0.494	0.503	0.793
		Weighted Average	0.628	0.274	0.642	0.628	0.623	0.765
48 Hours	Prediction	No Rain	0.451	0.104	0.677	0.451	0.542	0.761
		Rain	0.896	0.549	0.772	0.896	0.83	0.757
		Weighted Average	0.752	0.404	0.741	0.752	0.736	0.758
48 Hours	Estimation	No Rain	0.451	0.104	0.677	0.451	0.542	0.763
		Moderate	0.823	0.637	0.56	0.823	0.667	0.671
		Heavy	0.114	0.041	0.375	0.114	0.175	0.713
		Weighted Average	0.576	0.357	0.565	0.576	0.538	0.708

REFERENCES

1 B.Prashanthi, S.Meganathan, R.Bala Krishnan, R. Varahasamy and S.Swaminathan, "Data mining as a tool for hot day prediction during summer monsoon", IEEE Xplore, Int. Conference on Circuit, Power and Computing Technologies 2016, pages 1-3, DOI: 10.1109/ICCPCT.2016.7530229

2 T. R. Sivaramakrishnan and S. Meganathan, "Association Rule Mining and Classifier Approach for Quantitative Spot Rainfall Prediction", Journal of Theoretical and Applied Information Technology, Vol. 34, No.2, 173-177, 2011.

3 S. Meganathan and T. R. Sivaramakrishnan, "A Technique for Spot Forecasting", MAUSAM, 66(1), 33-42, 2015.

4 S.Meganathan and T.R.Sivaramakrishnan, "Winter Day Patterns and Weather Forecasting for Agricultural Crops over Cuddalore Region", Research Journal of Pharmaceutical Biological and Chemical Sciences, 7(2), 299-302, 2016.

5 S.Meganathan, T.R.Sivaramakrishnan, M. Poornima and A.Sumathi, "Extraction of Winter Temperature Patterns for Agricultural Operations", Research Journal of Pharmaceutical Biological and Chemical Sciences, 6(3), 1439-1442, 2015.

6 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', Research Journal of Applied Sciences, Engineering and Technology, Vol. 5(14), 3692-3696, 2013.

7 R.Varahasamy, S.Meganathan, Durga Karthik and K. Vijayarekha, "Quantitative Rainfall Prediction for Arid Regions in India using Bayesian Classifier", Pollution Research, 37(1), 208-211, 2018.

8 Kumar, Dr Mukesh, "Evaluating the performance of apriori and predictive apriori algorithm to find new association rules based on the statistical measures of datasets " International Journal of Engineering Research and Technology, 1, 1-5, 2012.

9 Aditi Mahajan, Anita Ganpati, "Performance Evaluation of Rule Based Classification Algorithms", International Journal of Advanced Research in Computer Engineering & Technology ,3 (10), 3545-3550, 2014.

10 *www.noaa.gov*