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

Precipitation Prediction for South West Monsoon Over Karnataka using Supervised Learning Technique

S. Meganathan, Michael Raj.T.F., RajaKumar.B, Raghuraman.K, N. Rajesh Kumar

Abstract*— Weather forecasting is a major field of study in the area of Meteorology. Data Scientists, meteorologists and weather forecasters are implementing the experimentation of weather forecasting base on numerical and statistical methods. Traditional models used the fluid and thermal dynamic strategies for grid-point time series prediction based on few inherited constraints, such as the adoption of incomplete boundary rules, model assumptions and numerical instabilities. The nominated work is focused on finding the south west monsoon months' precipitation patterns over the specific stations of Karnataka State. A multi-dimensional data framework for climate database with implementation online based data analysis has been developed. This works is carried out on the basis of monsoons that have prevailed during a year for the past 10 years. The proposed model emphasis the implementation of the association rules which has been extracted by the supervised classifier approach of data mining algorithms. The data mining technique of association rules emphasis the occurrence of the precipitation and will be helpful to take decisions in advance to the day to day operations in business, agriculture, water management and etc.

Keywords: Weather forecasting, Naïve Bayes classification, Data mining, supervised learning

I. INTRODUCTION

India is an agrarian nation and the achievement of farming depends of precipitation and stickiness. Forecasting rainfall and humidity is very essential process for the develop agriculture and increase the productivity. There are several methods to forecast the rainfall. It can be classified into two types, namely factual methods and dynamic methods. The factual methods used the historical data to analyze and predict the rainfall whereas dynamical approach uses physical model to simulate the environment and generate the solutions based on the formula [1]. In this paper, supervised learning algorithmic approach has been analyzed for the precipitation prediction of south west monsoon months over Karnataka coastal region.

Manuscript published on 30 September 2019
* Correspondence Author
S. Meganathan*, Faculty of Computer Science and Engineering, SASTRA Deemed University, Kumbakonam, Tamil Nadu, India. (Email: meganathan@src.sastra.edu)
Michael Raj.T.F, , Faculty of Computer Science and Engineering, SASTRA Deemed University, Kumbakonam, Tamil Nadu, India. (Email: tfmichaelraj23@src.sastra.edu)
RajaKumar.B, Faculty of Computer Science and Engineering, SASTRA Deemed University, Kumbakonam, Tamil Nadu, India. (Email: brk@src.sastra.edu)
Raghuraman.K, Faculty of Computer Science and Engineering, SASTRA Deemed University, Kumbakonam, Tamil Nadu, India. (Email: raghuraman@src.sastra.edu)
N. Rajesh Kumar, Faculty of Computer Science and Engineering, SASTRA Deemed University, Kumbakonam, Tamil Nadu, India. (Email: rajeshkumar.rb@src.sastra.edu)

II. LITERATURE REVIEW

Weather forecasting is more challenging task in meteorology, usually done for a specific graphical area. In most cases, rain prediction is required for individual city, harbors and geographic locations with essential installation.

Recently, the computer engineers have attempted to extract the behavioral patterns of occurrence of rainfall based on the temperature, due point, visibility and wind speed for costal station Cuddalore in East Coast of India using data mining approach. This paper reveals that the prediction of weather parameter 24 hours ahead and this process has been implemented a candidate generation of frequent pattern item set using Apriori algorithm. The experimental result of the data mining approach is satisfactory [2].

The influence of El Nino – Southern Oscillation (ENSO) is determined over the Northeast Monsoon Rainfall of East Coast of India. It is observed that monsoon period is affected due to the southern oscillation over the south peninsular of India. This monsoon months covers October, November, December for North-East rainfall and June, July, August and September for South-West rainfall [3, 5, 6]. The inspection of huge amount of data extracts the behavioural patterns and their relationship is represented in the form of association rules [4]. This approach is being emerged trend in climatology and meteorological problems.

III. METHODOLOGY

The proposed approach uses the following architecture depicted in Fig. 1 to obtain the association rule. It has various modules to manipulate the metrological data. By uploading the rainfall datasets with attribute and uploading the data as in the form of EXCEL file. For example, in original dataset, there is some attributes without data so we are removing the empty rows and columns. And implementing the preprocessing steps to convert the unstructured data into structured format. Eliminating the irrelevant and missing data from datasets. Finally classifying and clustering each attributes with constructed rules. Making decision to analyze the rain fall for particular area.

Data Inspection: The dataset analyzed for the proposed investigation was gathered from Mangalore, Karnataka state. Various phases are applied on data collection are data cleaning, data Selection, data transformation and data mining process.

Data extracted from online portal of world meteorological organization. Filtering of dataset- A consistent framework was developed to identify the missing data, finding the duplicated information and filtered out of bad data. After preprocessing, the dataset were normalized into a structured format suitable for data analysis. These online dataset contains collection of meteorological data. There is a need to do the preprocessing to filter out the data for effective data analysis for getting the accurate result. This approach provides the solution for effective data cleaning process. Data Choosing – This phase describes that the relevant data are selected from the extracted data for the analysis period which is taken from World Meteorological organization website. This dataset consist of weather parameters such as, daily mean temperature, humidity, sea level pressure, surface temperature, visibility, maximum wind speed, maximum temperature observed, minimum temperature observed and rainfall. The parameters are presented in Table 1. The sample records are shown in this table.

Data Transformation:

The preprocessed data are normalized based on the discretized values such as low, medium, high. For the class label precipitation, the occurrence of rainfall is mentioned as "YES" and the non-occurrence of rainfall is noted as "NO". For the advanced prediction of rainfall, the attribute values of the class label are described as "NO", "NORMAL", "HIGH". For the class label "NO" its threshold is 0 millimeter, the class label "NORAL" describes, about the threshold values between10 mm to 50 mm. The class label "HIGH" means that the threshold values is above 50 millimeter. The extracted file is normalized into nominal values and then converted into comma, separated, values (CSV) in the Microsoft Excel format. This file is uploaded into the machine learning tool WEKA. This tool visualizes the rainfall patterns based on the weather parameters specified in our analysis over the period for above specific station. After finding the discretized values of the input parameters as per the Table 2 which is transformed into nominal values as shows in Table 3.

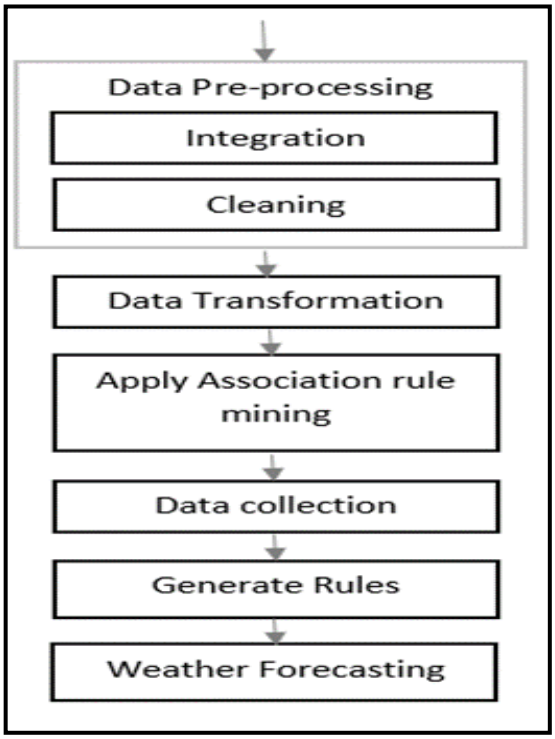


Fig 1. Proposed architecture

Table 1. Rainfall Dataset for Mangalore city

TEMP	DEWP	SLP	STP	VISIB	MXSPD	MAX	MIN	PRCP
28	75.2	1007.8	996.4	2.5	8	91	76.1	0
28.33333	76	1008.8	997.4	2.5	8.9	90.3	74.5	0
29.27778	75.7	1009.4	998	2.5	9.9	91	77	0.508
28.94444	76.6	1007.9	996.5	2.5	9.9	94.1	76.8	0.762
25	74.8	1006.5	995	2.5	9.9	83.1	74.5	19.05
24.11111	72.8	1006.2	994.6	1.8	15	81	72	134.874
24	73.4	1006.9	995.3	1.6	8	81.7	70.9	46.99
26.55556	73.3	1008.1	996.6	3	8.9	87.3	72	66.04
27.77778	74.6	1008.5	997.1	2.5	8	88.9	72.7	12.954
26.83333	75	1008.7	997.2	2.5	8	88	74.1	0
27.88889	75.9	1008.8	997.5	2.5	6	88.2	75.2	0
28.11111	74.8	1009.2	997.8	2.5	8.9	93.7	75	0
27.88889	75.7	1010.1	998.6	2.5	9.9	89.4	75.4	0
27.16667	75.3	1010.1	998.6	2.5	8.9	88.2	75.2	0

Table 2. Range value of Rainfall Dataset

PARA METER	NOMINAL VALUE	RANGE	COUNT
TEMP	LOW MEDIUM HIGH	0<25.44 25.44-27.72 >=27.72	366 758 90
DEWP	LOW MEDIUM HIGH	0<73.4 73.4<76.1 >=76.1	114 935 165
SLP	LOW MEDIUM HIGH	0<4002.1 4002.1-7001 >=7001	788 0 426
STP	LOW MEDIUM HIGH	0<3994.566 3994.566-6997.233 >=6997.233	788 0 426
VISIB	LOW MEDIUM HIGH	0<2.166 2.166-3.533 >=3.533	242 862 110
MXSPD	LOW MEDIUM HIGH	0<335.233 335.233-667.566 >667.566	1212 0 2
MAX	LOW MEDIUM HIGH	0<82.033 82.033-88.166 >88.166	148 894 172
MIN	LOW MEDIUM HIGH	0<73.033 73.033-77.166 >77.16	346 835 33
PRCP	No Rainfall Normal Rainfall High Rainfall	0 mm <=40 mm >40 mm	130 792 292

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

Red and Blue color refers positive and negative classes respectively in diagram, the colors are assigned automatically to each categorical value. There is useful to get a quick idea of whether the problem is easily separable for a given attribute. For example, all the red and blue are cleanly separated for a single attribute.The Fig.2 shows discretised values of weather parameters

TEMPERATURE:

In High, there is 94 count
Low, there is 350 count
Medium, there is 770 count

DEWPOINT:

In Medium, there is 905 count
High, there is 195 count
Low, there is 114 coun

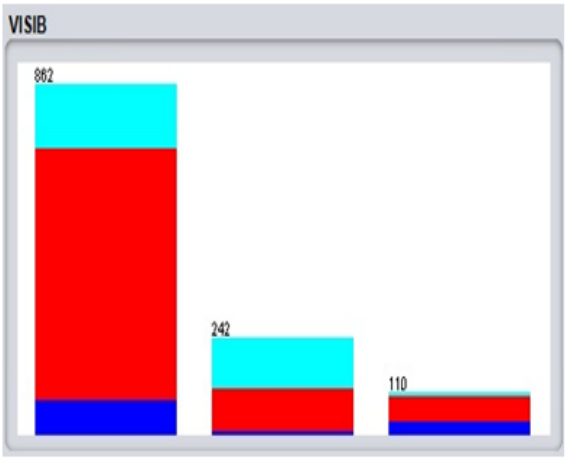
SLP:

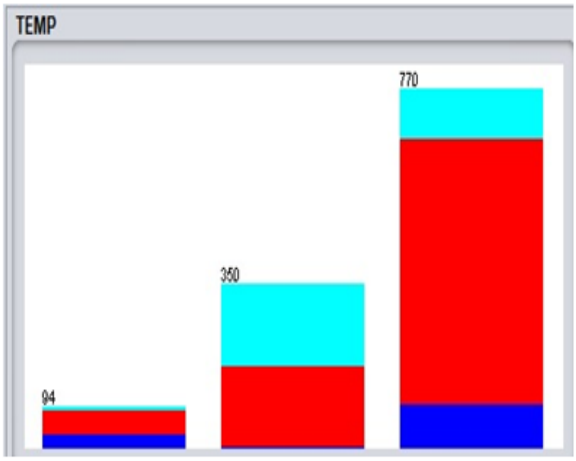
Low, there is 788 count
High, there is 426 coun

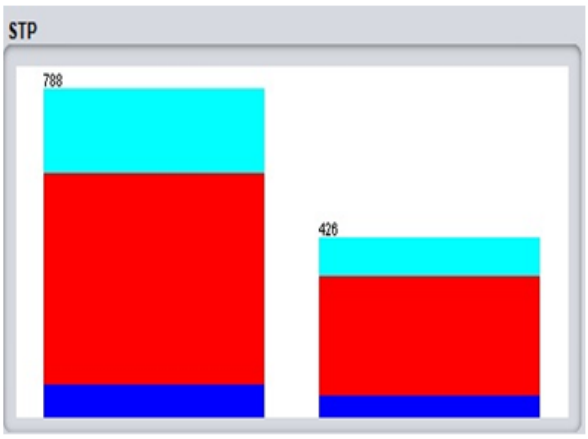
STP:

Low, there is 788 count
High, there is 426 coun

The extracted association rules for the occurrence of precipitation on the specified location is shown in Table 4.







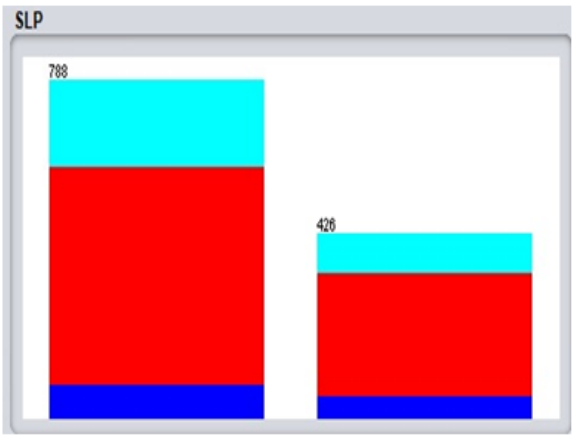


Fig2. Discretization ranges of weather parameters.

Table 3. Nominal value of Rainfall Dataset for Mangalore city

TEMP	DEWP	SLP	STP	VISIB	MXSPD	MAX	MIN	PRCP
High	Medium	Low	Low	Medium	Low	High	Medium	No
High	Medium	Low	Low	Medium	Low	High	Medium	No
High	Medium	Low	Low	Medium	Low	High	Medium	Normal
High	High	Low	Low	Medium	Low	High	Medium	Normal
Low	Medium	Low	Low	Medium	Low	Medium	Medium	Normal
Low	Low	Low	Low	Low	Low	Low	Low	High
Low	Low	Low	Low	Low	Low	Low	Low	High
Medium	Low	Low	Low	Medium	Low	Medium	Low	High
High	Medium	Low	Low	Medium	Low	High	Low	Normal
Medium	Medium	Low	Low	Medium	Low	Medium	Medium	No
High	Medium	Low	Low	Medium	Low	High	Medium	No
High	Medium	Low	Low	Medium	Low	High	Medium	No
High	Medium	Low	Low	Medium	Low	High	Medium	No
Medium	Medium	Low	Low	Medium	Low	High	Medium	No

Table 4. Association Rules

Rules	Condition	Result
Rule 1:	If TEMPERATURE = "Low" or "Medium" and DEWPOINT="Medium" and SLP="Low" or "High" and STP="Low" or "High" and VISIBILITY="Medium" and MAXIMUMSPEED="Low" and MaxiumTemperature="Medium" and MinimumTemperature="Low" or "Medium" then	PRECIPITATION="NO RAINY DAY"
Rule 2:	If TEMPERATURE = "Low" and DEWPOINT="Medium" and SLP="High" and STP="High" and VISIBILITY="Low" and MAXIMUMSPEED="Low" and MaxiumTemperature="Low" and MinimumTemperature="Medium" then	PRECIPITATION="NORMAL RAINY DAY"
Rule 3:	If TEMPERATURE = "Medium" and DEWPOINT="Medium" and SLP="Low" and STP="Low" and VISIBILITY="Medium" and PRECIPITATION="NO RAINY DAY"MAXIMUMSPEED="Low" and MaxiumTemperature="Medium" and MinimumTemperature="Medium" then	PRECIPITATION="HIGH RAINY DAY"

V. CONCLUSION

This Research focus on predicting the rainfall using historical of the 10 years data (2009-2018). This consists of 3620 records with the various attributes such as Temperature, Dewpoint, Visibility, etc., After preprocessing the number of records are reduced as 1214 datasets for rainfall prediction. This research uses Naive Bayes approach to classify and Simple K means algorithm is applied for clustering. Based on the identified clustering, the association rules are generated. It is used to predict the rainfall. The result shows that there is a improvement in proposed approach.

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. Sivaramakrishnan, T. R., and S. Meganathan. "Association rule mining and classifier approach for quantitative spot rainfall prediction." Journal of Theoretical and Applied Information Technology 34.2 (2011): 173-177.
3. Balachandran, S., R. Asokan, and S. Sridharan. "Global surface temperature in relation to northeast monsoon rainfall over Tamil Nadu." *Journal of earth system science* 115.3 (2006): 349-362.
4. Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami. "Mining association rules between sets of items in large databases." *Acm sigmod record*. Vol. 22. No. 2. ACM, 1993.
5. Zubair, Lareef, and Chester F. Ropelewski. "The strengthening relationship between ENSO and northeast monsoon rainfall over Sri Lanka and southern India." *Journal of Climate* 19.8 (2006): 1567-1575.
6. Chew, F. H. S., Piechota, T. C., Dracup, J. A. and Mcmohan, T. A., 1998, “Elnino/ Southern Oscillation and Rainfall Variations”, Journal of Hydrology, p138.
7. Lau, K. M., 1992, “East Asian Summer Monsoon Rainfall Variability and Climate teleconnections”, Journal of Meteorological Society of Japan, 70, p211.