

# Optimal Prediction of Weather Condition Based on C4.5 Classification Technique

**M. Manikandan<sup>1\*</sup>, R. Mala<sup>2</sup>**

<sup>1</sup>Dept. of computer science, Marudupandiyar college of arts and science, Thanjavur, Tamil Nadu, India

<sup>2</sup>Research Advisor and Head Incharge, Alagappa University College, Paramakudi, Tamil Nadu, India.

\*Corresponding Author: rmgvm2007@yahoo.com Mobile: +91 9488424357

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 12/Oct/2018, Published: 31/Oct/2018

**Abstract**— In this world many of task is very challenged for researchers. By the way the accurate weather prediction is one of the disputes for the meteorologist. So this paper focuses the weather prediction for an implementing the classification technique of C4.5 Classification technique. This technique can be analyzed for the performance and accuracy of weather condition. Also this decision tree algorithm can be applied in weather prediction parameter of training data under the various regions. Such as, Tamil Nadu, Andra Pradesh, Gujarat and Odhisa states are taken from India for this research work. These states are mainly focus for the purpose of different monsoon seasons and climates vary from actual period of time. Finally, weather condition can be predicted on various monsoons seasonally on the respective class label of climate range.

**Keywords**—C4.5; Temperature; Cloudcover; Vapor pressure; Relative humidity; Confusion matrix

## I. INTRODUCTION

Classification is a task in Data mining. Data mining, as indicated before, is a machine learning discipline, and is inspired by pattern recognitions, which is a branch of science, of which one of its goals is to classify objects into a number of categories referred to as classes. Data classification is a two step process.

- In the first step, a model is built by analyzing the data tuples from training data having a set of attributes. For each tuple in the training data, the value of class label attribute is known. Classification algorithm is applied on training data to create the model.
- In the second step of classification, test data is used to check the accuracy of the model. If the accuracy of the model is acceptable then the model can be used to classify the unknown data tuples.

A decision tree is a classifier expressed as a recursive partition of the instance space. The decision tree consists of nodes that form a rooted tree, meaning it is a directed tree with a node called “root” that has no incoming edges [23]. All other nodes have exactly one incoming edge. A node with outgoing edges is called an internal or test node. All other nodes are called leaves (also known as terminal or decision nodes). In a decision tree, each internal node splits the instance space into two or more subspaces according to a certain discrete function of the input attributes values.

## II. RELATED WORKS

Kotsiantis et al discussed [2,15] a prediction of rainfall with the domain variable of daily average, maximum and minimum temperature for Patras city in Greek analyzing based on six different data mining methods such as Feed-Forward Back Propagation (BP), k-Nearest Neighbor (KNN), M5rules algorithm, linear least-squares regression (LR), Decision tree and instance based learning (IB3). They use four years period data of temperature, relative humidity and rainfall. The results they obtained in this study were accurate in terms of Correlation Coefficient and Root Mean Square. Data mining have been employed successfully to build a very important application in the field of meteorology like predicting abnormal events like hurricanes, storms and river flood prediction. These applications can maintain public safety and welfare.

Accurate and timely weather forecasting is a major challenge for the scientific community. Rainfall prediction modeling involves a combination of computer models, observation and knowledge of trends and patterns. Using these methods, reasonably accurate forecasts [3,16] can be made up. Several recent research studies have developed rainfall prediction using different weather and climate forecasting methods[5,6]. Regression is a statistical empirical technique and is widely used in business, the social and behavioral sciences, the biological sciences, climate prediction, and many other areas.

Ewona, presented the paper belongs to regression constants  $a$  and  $b$  were therefore extracted from the equations. This is known as the deterministic model

$$Y = A + BX \quad (1)$$

Here  $Y$  =Dependent variable  $X$ =independent variable  $A$ ,  $B$ =Regression parameter which is reported in the form of constant parameter  $a$ , which is a reflection of the trend of the parameter lies between two variable. Monthly mean daily total rainfall shows marked latitudinal dependence as can be seen in the positive slopes of the graph of  $b$  against latitude. Rainfall data collected by the Nigerian Meteorological Agency. It shows consistent increase during the thirty years of this study. Constant  $b$  which is an indication of the volume of rainfall shows strong latitudinal dependence.

The weather data used for the research include daily temperature [14], daily pressure and monthly rainfall. Sarah N. Kohail, Alaa M. El-Halees, described Data Mining for meteorological Data and applied knowledge discovery process to extract knowledge from Gaza city weather dataset[7,17].

Nizar and Sanjay proposed an artificial neural network based model[4,8]with wavelet decomposition for prediction of monthly rainfall on account of the preceding events of rainfall data. Wavelet transform an extraction of approximate and detail coefficient of the rainfall data series. The coefficients obtained from wavelet decomposition are used along with ANN for learning and knowledge extraction processes[13]. After wavelet decomposition of rainfall time series, a multilayer perception with two hidden layer is found optimal for approximate coefficient prediction. Further focused time lag recurrent network with gamma memory is found optimal for prediction of detail coefficients. Thus a committee of two different ANN configurations is proposed for reliable rainfall prediction. The accuracy predicted for the rainfall model is reasonable.

Sohn, [3,10,12] has developed a prediction model for the occurrence of heavy rain in South Korea using multiple linear and logistics regression, decision tree and artificial neural network. M. T. Mebrhatu [4,9,11]modeled for prediction categories of rainfall (below, above, normal) in the highlands of Eritrea.

Table 1. Domain variable of collecting data

Domain Variables	Abbreviation
W_Temp	Temperature of winter seasons in January and February Month
W_Cloud cover	Cloud cover of winter seasons
W_Vapor pressure	Vapor pressure in winter season
S_Temp	Temperature of Hot Summer Season in March, April and May month
S_Cloud cover	Cloud cover of Hot Summer Season

S_Vapor pressure	Vapor pressure of summer season
SW_Temp	Temperature of South west Monsoon in June, July, august and September month
SW_Cloud cover	Cloud cover of South west Monsoon
SW_Vapor pressure	Vapor pressure of South west Monsoon
NE_Temp	Temperature of North east Monsoon in October, November and December month
NE_Cloud cover	Cloud cover of North east Monsoon
NE_Vapor pressure	Vapor pressure of North east Monsoon

The most influential predictor of rainfall amount was the southern Indian Ocean SST. Experimental results showed that the hit rate for the model was 70%.

### III. DATA FOR RESEARCH

In this research work, the datasets are taken in the real time weather data under 16 regions from the state of TamilNadu, Andhra Pradesh, Gujarat, and Odhisa with different regions on each state during the period from 1990 to 2016 taken from meteorological department. For these stations the data can be collected by qualitative data represents temperature, Vapor pressure, cloud cover for analyzing the weather prediction during the season of summer, winter, northeast, southwest monsoon. Temperature is measured by Celsius, Vapor pressure is measured by millibar (mb), and cloud cover is measured by %. In this paper, relative humidity can be calculated by Vapor pressure and saturation of vapor pressure which is measured by percentage.

Table 2. Sample Training data set for predicting weather Condition

Location	Year	W_Temp	W Cloud cover	Vapour Pressure
Kanchipuram	1990	28	28.06	22.4
	1991	28	28.05	23.6
	1992	28	28.05	22.4
	1993	28	28.05	22.4
	1994	28	28.05	22.4
	1995	28	28.05	22.4
	1996	28	15.8	22.4
	1997	28	23.4	22.3
	1998	29	21.0	23.9
	1999	28	27.2	22.4
	2000	28	23.8	23.5
	2001	28	35.7	23.3
	2002	28	37.6	23.7
	2003	28	28	22
	2004	28	28	24
	2005	28	28	22
2006	28	28	22	
2007	30	28	22	
2008	29	28	22	
2009	29	16	22	
2010	29	23	22	
2011	30	21	24	
2012	29	27	22	
2013	30	24	24	

	2014	30	36	23
	2015	30	38	24
	2016	30	17	22

The winter seasons are comes January and February, the summer seasons are comes from March to May Southwest periods follows June, July, August, and September and a Northeast monsoon period follows on October November, December. The data sets are collected from the India Meteorological Department section websites, here the sample training data can be represent in Table 1, further data also taken in same type of domain structure and Table 2, represent the domain values which can be used to classify for predicting the weather condition from meteorological data in different season.

#### A. Scope Of The Research

The methodology used for present work to implementing the classification technique of C4.5 and analyzing the performance, accuracy of these algorithms. Finally, weather condition can be predicted on various monsoon time and then find the relationships between the weather variables using correlation coefficient computation.

#### IV. IMPLEMENTATION OF C4.5 TREE ALGORITHM FOR WEATHER PREDICTION

The decision tree built using the training set, because of the way it was built, deals correctly with most of the records in the training set. Pruning of the decision tree is done by replacing a whole sub tree by a leaf node. The replacement takes place if a decision rule establishes that the expected error rate in the sub tree is greater than in the single leaf [18].

The decision tree that can be constructed from the given set of attributes. While some of the trees are more accurate than others, finding the optimal tree is computationally infeasible because of the exponential size of the search space [19]. These algorithms usually employ a greedy strategy that grows a decision tree by making a series of locally optimum decision about which attributes to use for partitioning the data. Each path from the root of a decision tree to one of its leaves can be transformed into a rule simply by conjoining the tests along the path to form the antecedent part, and taking the leaf's class prediction as the class values.

Steps of the System:

1. Selecting dataset as an input to the algorithm for processing.
2. Selecting the classifiers.
3. Calculate entropy, information gain, and gain ratio of attributes.
4. Processing the given input dataset according to the defined algorithm of C4.5 data mining.
5. According to the defined algorithm of improved C4.5 data mining processing the given input dataset.

6. The data which should be inputted to the tree generation mechanism is given by the C4.5 and improved C4.5 processors. Tree generator generates the tree for C4.5 and improved C4.5 decision tree algorithm.

Table 3. Ranges to classify the weather condition

Monsoon Seasonal	Ranges to classify the weather condition
Winter	cold and cloudy
	Moderate and cloudy
Summer	Moderate and cloudy
	Hot and dry
	cold and cloudy
	Moderate hot
South West	Moderate and cloudy
	Hot and dry
	cold and cloudy
	Moderate hot
	Warm
North East	cold and cloudy
	Moderate hot
	Moderate and cloudy

C4.5 is a decision tree technique which is enhanced by ID3 algorithm. It is one of the most popular algorithms for rule base classification [19]. Here an attributes can be split into two partition based on the selected threshold value, all the value satisfied by the constraint it will be assigned in one child and remaining values can be store in another child respectively. It also handles missing values. Here it can be gather of all binary tests through entropy gain and the values are sorted based on the values in continuous attribute values which are calculated in one scan. This process is repeated for each continuous attributes when the process is terminated.

The rule set is formed from the initial state of decision tree [18]. Each path from the initial state, the condition will be evaluate and simplified by the effect of rule and an outcomes will put on the required leaf, the step will continuous when it comes discarding the condition. Let freq (C<sub>i</sub>, S) stand for the number of samples in S that belong to class C<sub>i</sub> (out of k possible classes), and |S| denotes the number of samples in the set S. Then the entropy of the set of equation (1) such as

$$\text{Info}(s) = \sum_{i=1}^k ((\text{freq}(c_i, s) / |s|) \cdot \log_2 (\text{freq}(c_i, s) / |s|)) \quad (1)$$

After set T has been partitioned in accordance with n outcomes of one attribute test X:

$$\text{Info}_x(S) = \sum_{j=1}^n \frac{|S_j|}{|S|} \cdot \text{Info}(S_j) \quad (2)$$

$$\text{Gain}(x) = \text{Info}(S) - \text{Info}_x(S) \quad (3)$$

Table 4. Saturation (Sat) vapor pressure of Temperature in Celsius

Temp in Celcius	(Sat) Vapor Pressure (mb)	Temp in Celcius	(Sat) Vapor Pressure (mb)
-18	1.5	18	21.0
-15	1.9	21	25.0
-12	2.4	24	29.6
-09	3.0	27	35.0
-07	3.7	29	41.0
-04	4.6	32	48.1
-01	5.6	35	56.2
02	6.9	38	65.6
04	8.4	41	76.2
07	10.2	43	87.8
10	12.3	46	101.4
13	14.8	49	116.8
16	17.7	52	134.2

The gain ratio “normalizes” the information gain as follows

$$GainRatio(a_i, S) = \frac{InformationGain(a_i, S)}{Entropy(a_i, S)} \tag{4}$$

**Pseudo-code**

```

ComputeClassFrequency (T);
if OneClass or FewCases return a leaf; create a decision node N;
For Each Attribute A ComputeGain(A);
N.test = AttributeWithBestGain;
if N.test is continuous and Threshold;
For Each T' in the splitting of T
if T' is Empty Child of N is a leaf else
Child of N = FormTree( T' );
Compute Errors of N;
return N.
    
```

**A. Rule Based Classifier**

- Each classification rule is of form  $r : (Condition) \rightarrow y$
- LHS of the rule (Condition), called rule antecedent or precondition, is a conjunction of attribute tests [15].
- RHS, also called the rule consequent, is the class label , Rule set:  $R = \{ r_1, r_2, \dots, r_n \}$

Before construction of the rule base to find the relative humidity with respect to vapor pressure. Relative humidity may be defined as the ratio of the actual to the saturation vapor pressure.

$$RH = \frac{(Actual\ Vapor\ Pressure)}{(Saturation\ Vapor\ Pressure)} \times 100\%$$

Actual vapour pressure is a measurement of the amount of water vapour in a volume of air and increases as the amount of water vapour increases. Air that attains its saturation vapour pressure has established equilibrium with a flat surface of water.

Saturation (Sat) vapour pressure is a unique function of temperature as given in the Table 4 as above. Each temperature in the table may be interpreted as a dew point temperature, because as the ground cools, dew will begin to form at the temperature corresponding to the vapour pressure in this table.

For example:

$$Relative\ Humidity\ (RH) = 100\% \times (10.21 / 25.0) = 41\%$$

**V. CONSTRUCTION OF A RULE BASED CLASSIFIER FROM DATA**

**Indirect Method:**

- Rules are mutually exclusive and exhaustive.
- Rule set contains as much information as the tree.
- Rules can be simplified as follows,

```

Winter Climate = cold and cloudy: TN_Nilgiri (27.0)
Winter Climate = Moderate and cloudy
| North East Climate = cold and cloudy:
TN_Thiruannamalai (27.0)
| North East Climate = Moderate hot: TN_Kanchipuram
(27.0)
| North East Climate = Moderate and cloudy: TN_Karaikal
(27.0)
Winter Climate = Cold and cloudy
| South west Temp <= 27.08675
| | Summer climate = Moderate and cloudy: AP_Anantapur
(0.0)
| | Summer climate = Hot and dry
| | | South west Climate = Moderate and cloudy:
AP_Anantapur (0.0)
| | | South west Climate = cold and cloudy:
AP_Anantapur (0.0)
| | | South west Climate = warm: AP_Anantapur (0.0)
| | | South west Climate = Cold and cloudy:
AP_Anantapur (27.0)
| | | South west Climate = Hot and dry: AP_Anantapur
(0.0)
| | | South west Climate = Warm
| | | | Winter Temp <= 21.2: AP_Visakapatnam (27.0)
| | | | Winter Temp > 21.2: AP_Srikakulam (27.0)
| | | | South west Climate = Moderate hot: AP_Anantapur
(0.0)
| | | Summer climate = Cold and cloudy
| | | Sum_cloud COVer <= 22.047
| | | | SWCloud Cover <= 45.583: G_Katchch (27.0)
| | | | SWCloud Cover > 45.583: G_Surat (27.0)
| | | Sum_cloud COVer > 22.047: OD_Puri (27.0)
    
```

```

| | Summer climate = Moderate hot: AP_Anantapur (0.0)
| | South west Temp > 27.08675
| | | NWEcloud cover <= 18.151
| | | | South West Hum <= 80.890421: G_Gandhinagar (27.0)
| | | | | South West Hum > 80.890421: G_Narmada (27.0)
| | | | | | NWEcloud cover > 18.151
| | | | | | | NorthEast Temp <= 24.265833
| | | | | | | | Summer Hum <= 87.516477
| | | | | | | | | Winter Hum <= 93.423661
| | | | | | | | | | South West Hum <= 83.808182
| | | | | | | | | | | South west Temp <= 27.605375: OD_Bargarh (2.0)
| | | | | | | | | | | | South west Temp > 27.605375: OD_Sambalpur (28.0/1.0)
| | | | | | | | | | | | | South West Hum > 83.808182: OD_Bargarh (3.0)
| | | | | | | | | | | | | | Winter Hum > 93.423661: OD_Bargarh (5.0)
| | | | | | | | | | | | | | | Summer Hum > 87.516477: OD_Bargarh (16.0)
| | | | | | | | | | | | | | | NorthEast Temp > 24.265833
| | | | | | | | | | | | | | | Summer climate = Moderate and cloudy: AP_Krishna (0.0)
| | | | | | | | | | | | | | | Summer climate = Hot and dry: AP_Krishna (27.0)
| | | | | | | | | | | | | | | Summer climate = Cold and cloudy: AP_Krishna (0.0)
| | | | | | | | | | | | | | | Summer climate = Moderate hot: OD_Jagatsingpur (27.0)
Winter Temp <= 24.25275: cold and cloudy (351.0)
Winter Temp > 24.25275: Moderate and cloudy (81.0)
NorthEast Temp <= 22.1305
| | Sum_cloud COVer <= 35.609333
| | | South west Climate = Moderate and cloudy: cold and cloudy (0.0)
| | | | South west Climate = cold and cloudy: cold and cloudy (81.0)
| | | | South west Climate = warm: cold and cloudy (0.0)
| | | | South west Climate = Hot and dry: cold and cloudy (0.0)
| | | | South west Climate = Warm: Hot and dry (27.0)
| | | | South west Climate = Moderate hot: cold and cloudy (0.0)
| | | | Sum_cloud COVer > 35.609333: Moderate and cloudy (27.0)
NorthEast Temp > 22.1305
| | SWCloud Cover <= 70.644
| | | North East Hum <= 85.896587: Hot and dry (183.0)
| | | | North East Hum > 85.896587
| | | | | SWCloud Cover <= 69.64425
| | | | | | North East Hum <= 85.910296: Moderate hot (2.0)
| | | | | | | North East Hum > 85.910296: Hot and dry (54.0/1.0)
| | | | | | | | SWCloud Cover > 69.64425: Moderate hot (7.0/1.0)
| | | | | | | | SWCloud Cover > 70.644
| | | | | | | | Summer <= 29.427333
| | | | | | | | | Location = TN_Nilgiri: Moderate hot (0.0)
| | | | | | | | | Location = TN_Kanchipuram: Moderate hot (0.0)
| | | | | | | | | Location = TN_Thiruanmalai: Moderate hot (0.0)

```

```

| | | Location = TN_Karaikal: Moderate hot (0.0)
| | | Location = AP_Anantapur: Moderate hot (0.0)
| | | Location = AP_Krishna: Moderate hot (0.0)
| | | Location = AP_Srikakulam: Moderate hot (0.0)
| | | Location = AP_Visakapatinam: Moderate hot (0.0)
| | | Location = G_Gandhinagar: Moderate hot (0.0)
| | | Location = G_Katchch: Moderate hot (0.0)
| | | Location = G_Narmada: Moderate hot (0.0)
| | | Location = G_Surat: Moderate hot (0.0)
| | | Location = OD_Bargarh: Hot and dry (2.0)
| | | Location = OD_Jagatsingpur: Moderate hot (23.0)
| | | Location = OD_Puri: Moderate hot (0.0)
| | | Location = OD_Sambalpur: Hot and dry (2.0)
| | Summer > 29.427333: Hot and dry (24.0)

```

From the above construction of rule base, the weather condition can be analysed under each monsoon season. From Figure.1, represents the tree view of the classification of weather prediction based on the class label of monsoon climate in each location.

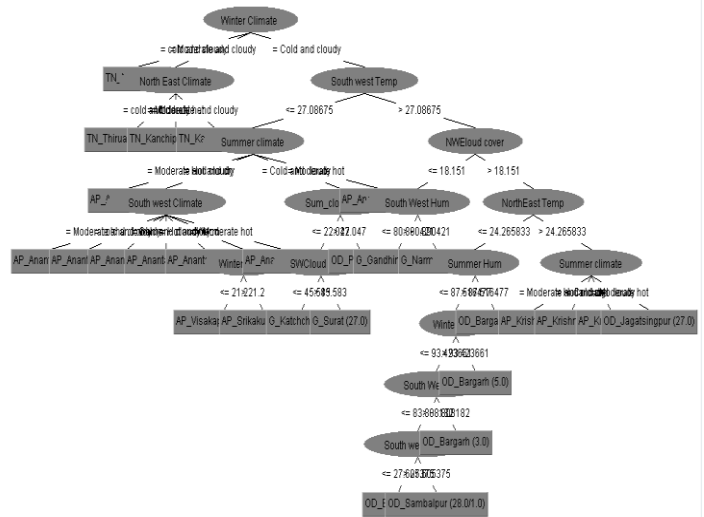


Figure.1 Tree View of the C4.5 Classification of Weather Prediction

From the classification analysis, the weather condition of temperature and humidity can be predicted under the seasonal monsoon of winter, summer, Southwest and north east monsoon, which can be predicted 99.9% from the probability distribution of actual prediction.

From the confusion matrix, the winter season predicted weather condition accuracy can be analyzed as follows

```

a b <-- classified as
350 1 | a = cold and cloudy
0 81 | b = Moderate and cloudy

```

Here 350 data of cold and cloudy and 81 item of moderate and cloudy can be predicted in winter season out of 432 instance of data. By the same process, Summer monsoon weather condition can be predicted as

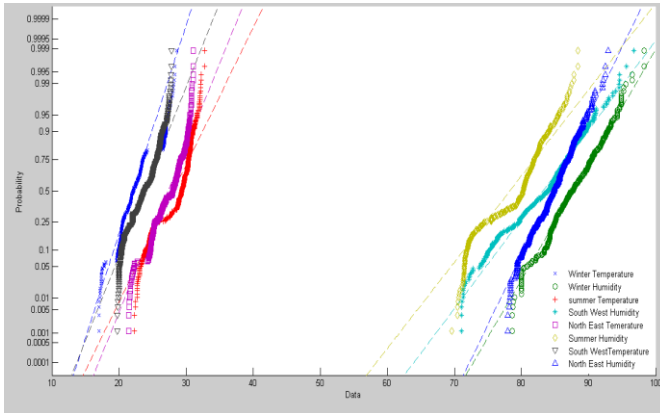


Figure.2 Temperature and Humidity level on monsoon Season wise

```
a b c d <-- classified as
27 0 0 0 | a = Moderate and cloudy
0 288 0 5 | b = Hot and dry
0 0 82 0 | c = cold and cloudy
0 3 0 29 | d = Moderate hot
```

The diagonal of classifier can be indicates the weather condition of monsoon seasons respectively. Southwest Monsoon Weather condition

```
a b c d e f <-- classified as
80 2 1 0 2 0 | a = Moderate and cloudy
1 121 0 3 0 0 | b = cold and cloudy
1 0 15 0 0 0 | c = warm
0 1 0 33 0 3 | d = Hot and dry
1 0 0 0 74 1 | e = Warm
0 0 0 3 0 98 | f = Moderate hot
```

Finally, North East Monsoon Weather condition

```
a b c <-- classified as
377 0 2 | a = cold and cloudy
1 26 0 | b = Moderate hot
0 1 26 | c = Moderate and cloudy
```

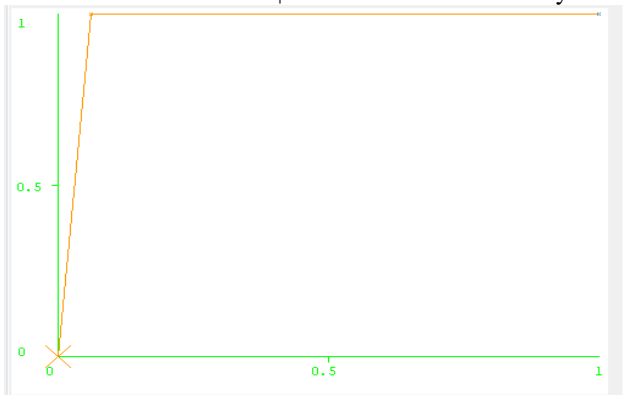


Figure.3 True positive rate of weather prediction

From the Figure.3 represents the 99.7% of data can be predicted true positive rate, which can be exposed the optimal solution given by the C4.5 classifier in efficient manner.

### V. CONCLUSION

In this research work, it can be concluded that the data mining techniques to support decision making for analyzing the weather condition of each monsoon wise in seasonally. Also finding these decisions required this research on data mining techniques can be predicted the climate with respect of rule base classifier. From the result analysis, the C4.5 algorithm can be 99.7 in 0.01 seconds successfully.

### REFERENCES

- [1] Olaiya, Folorunsho, and Adesesan Barnabas Adeyemo. "Application of data mining techniques in weather prediction and climate change studies." *International Journal of Information Engineering and Electronic Business (IJIEEB)* 4.1 (2012): 51.
- [2] Kantardzic, Mehmed. "Data mining: concepts, models, methods, and algorithms." John Wiley & Sons, 2011.
- [3] Berkhin, Pavel. "A survey of clustering data mining techniques." *Grouping multidimensional data*. Springer Berlin Heidelberg, 2006.25-71.
- [4] Allen, Richard G. "Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56." *FAO, Rome* 300 (1998): 6541.
- [5] Lawrence, Mark G. "The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications." *Bulletin of the American Meteorological Society* 86.2 (2005): 225-233.
- [6] Howarth, Edgar, and Michael S. Hoffman. "A multidimensional approach to the relationship between mood and weather." *British Journal of Psychology* 75.1 (1984): 15-23.
- [7] Pasanen, A-L., et al. "Laboratory studies on the relationship between fungal growth and atmospheric temperature and humidity." *Environment International* 17.4 (1991): 225-228.
- [8] Burk, R. L., and M. Stuiver. "Oxygen isotope ratios in trees reflect mean annual temperature and humidity." *Science* 211.4489 (1981): 1417-1419.
- [9] Zhang, Guang Jun, and Michael J. Mcphaden. "The relationship between sea surface temperature and latent heat flux in the equatorial Pacific." *Journal of climate* 8.3 (1995): 589-605.
- [10] Campbell, A., et al. "Temperature requirements of some aphids and their parasites." *Journal of applied ecology* (1974): 431-438.
- [11] Thornton, Peter E., Steven W. Running, and Michael A. White. "Generating surfaces of daily meteorological variables over large regions of complex terrain." *Journal of Hydrology* 190.3 (1997): 214-251.
- [12] Solomon, M. E. "Control of humidity with potassium hydroxide, sulphuric acid, or other solutions." *Bulletin of Entomological Research* 42.03 (1951): 543-554
- [13] Zhu, Xingquan, and Ian Davidson, eds. "Knowledge Discovery and Data Mining". Challenges and Realities. Igi Global, 2007.
- [14] Manish Verma, MaulySrivastava, NehaChack, Atul Kumar Diswar and Nidhi Gupta, "A Comparative Study of Various Clustering Algorithms in Data Mining", *International Journal of Engineering Research and Applications (IJERA)* Vol. 2, Issue 3, May-Jun 2012, pp.1379
- [15] Hall, Mark, et al. "The WEKA data mining software: an update." *ACM SIGKDD Explorations Newsletter* 11.1 (2009): 10-18.

- [16] T.F. Gonzales. "Clustering to minimize the maximum inter cluster distance". Theoretical Computer Science,1985,38(2-3):293-306.
- [17] Kannan, M., S. Prabhakaran, and P. Ramachandran. "Rainfall forecasting using data mining technique."(2010)
- [18] Arun K Pujari, "Data mining techniques", University Press (India).2013
- [19] Jiawei Han Mischeline Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publisher an imprint of Elsevier. 2006

### **Aurthor Profile**

---

Mr. M. Manikandan, pursuing P.hD in Computer Science in PG and Research department of Computer Science, Marudupandiyasr college of arts and Science, Thanjavur is Under affiliated by Bharathidasan University, Trichy. He complete M.Phil in Computer Sscience in SASTRA univerisyt at 2010. Also he had 6 years experience in teaching field.

Dr. R. Mala, pursed M.C.A, M.Phil., Ph.D in Computer Science. She complete her Ph.D at Mother Tteresa University. She has 20-yerars experience in Teaching field and 6- years experience in research field. Also she published more then 40 research paper in National and International journals and 3 books. She is working as Assistant Professor and Head incharge of Departatment of Computer science, Alagappa University college, Paramakudi. She is an efficent and motivated person for the research scholars under her guidance.

---