www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 1 Jan 2016, Page No. 15432-15440

SURVEY ON METEOROLOGICAL WEATHER ANALYSIS BASED ON NAÏVE BAYES CLASSIFICATION ALGORITHM

P.Vanitha[1], Dr M.Mayilvaganan[2]

1. Research Scholar, Department of Computer Science, Karpagam University, Coimbatore. Assistant Professor, Department of Information & Computer Technology, Hindusthan College of Arts & Science, Coimbatore.

2. Associate Professor, Department of Computer Science, PSG College of Arts and Science, Coimbatore.

ABSTRACT

The aim of this research work, to focuses meteorological data to predict the monsoon seasons to separate the weather data based on Longitudinal and Latitudinal which can be used to analyse the reliability factor of Temperature, Humidity, Rainfall and Cyclone.

The present paper analyses the monthly weather data and rainfall data from the Indian monsoon months between the year 2000 to 2014.

To find the relationship between the rainfall details, humidity, temperature, After finding the reliability, Naïve Bayes Classification can be predict the Meteorological condition of region based on the cyclone and humidity level in seasonally.

It provides specific services to ass essment of pollution impacts from different companies and thermal power plants. The environment correlations play a significant role in determining the climate trends which are crucial in understanding the short and long-term trends in climate.

Keywords: Longitudinal, Latitudinal Temperature, Humidity, Rainfall and Cyclone, monsoon

I.INTRODUCTION

Rainfall prediction modeling involves a combination of computer simulations, and findings the trends and patterns.

The weather data used for the research include daily temperature, daily pressure and monthly rainfall.,. Data Mining for meteorological Data and applied knowledge discovery process is used to extract knowledge from Gaza city weather dataset.

A hybrid data mining technique that can be used to predict more

accurately the mean daily temperature values.

II. RELATED WORK

The aim of this study, the variation in humidity and Temperature and direction affects rainfall and analyse the cyclone movements. In this research, to group the meteorological data in location based on the latitudinal and Longitudinal using K- Means clustering technique.

Finds the relationships between the weather variables like Wintemp, Stemp, SWtemp and NEtemp, winhumidity, Sum humidity, SWhumidity and NEhumidity, WinActual, Sactual, SWActual, NEActual using Karl Pearson correlation coefficient computation.

Using Naïve Bayes classification Technique, can be predict the level of meteorological weather condition of region in seasonally which based on cyclone form and level of humidity.

III.DATA ANALYSIS

In this research work, the datasets are taken in the real time weather and rainfall dataset under five regions such as Chennai, Coimbatore, Cuddalore, Trichy and Nilgiri during the period of 2000-2014 in Tamilnadu district.

The present work analyses the rainfall information, temperature and humidity data during summer, and winter, northeast, southwest periods. Rainfall is measured by millimeter (mm), temperature is measured by Celsius and humidity is measured by percentage.

The data sets are collected from the India Meteorological Department section websites, here the sample training data in the region of Chennai can be represent in Table I, further data also taken in same type of domain structure of Seasonal Temperature Humidity, Rainfall, cyclone occuring and Latitudinal and Longitudinal of several region and

Table II, represent the abbreviation of the domain values which can be used to analyse in this research work.

Table I: Training Data Collection for weather and Rainfall data Chennai Region

Year	W Temp	W Cyclone	Win Humidity	W Normal	W Actual	S Temp	S Cyclone	S Humidity
2000	31	0	75	20	5.2	36	1	68
2001	32	0	74	18.5	0	37	0	65
2002	32	0	75	18.5	2.6	37	0	52
2003	32	0	76	28.5	3.5	40	0	76
2004	32	0	80	17.5	2	37	0	80
2005	33	0	74	18.1	2.8	38	0	79
2006	32	0	68	28.5	8.3	35	0	76
2007	31	0	70	28.5	0	36	0	64
2008	31	0	78	28.5	49.7	36	1	78
2009	32	0	80	13.1	0.5	35	0	76
2010	31	0	78	4.7	0	37	0	77
2011	31	0	84	11.8	2	36	0	72
2012	31	0	85	17	2	38	0	82
2013	30	0	86	16.9	2.2	38	0	79
2014	31	0	77	28.5	49.7	38	0	65

NE Temp	NW Cyclone	NE Humidity	NE Normal	NE Actual
29	0	78	286.3	418
29	0	79	286.3	218.6
30	3	82	290.7	164.7
30	0	80	290.7	210.8
31	0	82	293.7	223.7
29	4	79	291	246.5
29	0	78	291	589.4
29	0	82	291	358.3
30	5	73	291	324.3
30	1	80	291	363.4
30	1	81	280	394
29	1	80	287	241.5
30	1	85	283	195.4
30	1	86	292	174.4
30	0	79	285	134.4

S Normal	S Actual	SW Temp	SW Cyclone	SW Humidity	SW Normal	SW Actual
161.5	139	34	1	72	322.1	230
157.4	145.9	35	0	78	316.2	393.1
154.3	62.8	34	0	78	316.2	259.1
154.3	97	35	0	70	316.2	167.9
151.3	269.9	33	0	79	342.3	232.8
150.4	228.7	34	0	70	317	211.7
130.2	289.7	33	0	80	316.2	308.1
130.2	68.8	34	0	71	316.2	339.2
130.2	216.35	32	4	70	316.2	236.4
130.2	152.1	31	0	82	316.2	349.1
153.3	212	35	0	70	330	165.8
132.2	259.2	36	0	82	325	326.5
150	268.5	36	0	80	335	256.9
134	227.4	35	0	82	228	286.5
150	268.5	35	0	72	332	275.6

Table I. Training Data Collection for weather and Rainfall data Chennai Region

Winter Rain Fall		Summer Humidity %	Summer Rain Fall	
Normal (mm)	Actual (mm)		Normal (mm)	Actual (mm)
20	5.2	68	161.5	139
18.5	0	65	157.4	145.9
18.5	2.6	52	154.3	62.8
28.5	3.5	76	154.3	97
17.5	2	80	151.3	269.9
18.1	2.8	79	150.4	228.7
28.5	8.3	76	130.2	289.7
28.5	0	64	130.2	68.8
28.5	49.7	78	130.2	216.35
13.1	0.5	76	130.2	152.1
4.7	0	77	153.3	212
11.8	2	72	132.2	259.2
17	2	82	150	268.5
16.9	2.2	79	134	227.4
28.5	49.7	65	150	268.5

Year	SW Humidity %	SW Rainfall		NE Humidity %	NE Rainfall	
		Normal (mm)	Actual (mm)		Normal (mm)	Actual (mm)
2000	72	322.1	198.9	78	286.3	418
2001	78	316.2	393.1	79	286.3	218.6
2002	78	316.2	259.1	82	290.7	164.7
2003	70	316.2	167.9	80	290.7	2108
2004	79	342.3	232.8	82	293.7	223.7
2005	70	317	211.7	79	291	246.5
2006	80	316.2	308.1	78	291	589.4
2007	71	316.2	339.2	82	291	358.3
2008	70	316.2	236.4	73	291	324.3
2009	82	316.2	349.1	80	291	363.4
2010	70	330	165.8	81	280	394
2011	82	325	326.5	80	287	241.5
2012	80	335	256.9	85	283	195.4
2013	82	228	286.5	86	292	174.4
2014	72	332	275.6	79	285	134.4

Table I. Latitudinal And Longitudinal Of Region (contd.)

Region	Latitudinal	Longitudinal
Madras (Chennai)	13° 04' N	80° 17' E
Tiruchchirappalli	10° 50' N	78° 46' E
Cuddalore	11° 43' N	79° 49' E
Coimbatore	11° 00' N	77° 00' E
Nilgiri	11° 24' N	76° 44' E

Latitudinal And Longitudinal Of Region

Region	Latitudinal	Longitudinal
Madras (Chennai)	13° 04' N	80° 17' E
Tiruchchirappalli	10° 50' N	78° 46' E
Cuddalore	11° 43' N	79° 49' E
Coimbatore	11° 00' N	77° 00' E
Nilgiri	11° 24' N	76° 44' E

Domain Variables:

Domain Variables	Abbreviation
W in Temp	Temperature of winter seasons in January and February Month
W in Humidity	Humidity of winter seasons
S Temp	Temperature of Hot Summer Season in March, April and May month
Sum Humidity	Humidity of Hot Summer Season
SW Temp	Temperature of South west Monsoon in June, July, august and September month
SW Humidity	Humidity of South west Monsoon
NE Temp	Temperature of North east Monsoon in October, November and December month
NE Humidity	Humidity of North east Monsoon

IV.PROPOSED RESEARCH METHODOLOGY – To Cluster the seasonal Monsoon Data Using K- Means Algorithm

In the proposed methodology, to focus the meteorological weather analysis of temperature, humidity, rainfall and cyclone movements of the different region in Monsoon wise.

The four division of monsoon seasonal data from the given dataset as Winter seasons in the month of January and February. The parameters along with Centroid Euclidean inter-cluster distance measure for the five clusters C0,C1,C2,C3 and C4 for grouping based on location using latitudinal and longitudinal to Cluster Centroid as given equation ,

Here Centroid based partition techniques uses the centroid of cluster, Ci to represent

that cluster distance measures based on longitudinal and latitudinal location

Measured by dist(Ci, Cj) is Euclidean distance between to point. It can be cluster to

different location for analyzing temperature, humidity and rainfall level. K-Means Algorithm

Select the k-points it is treated as centroid for grouping the location based on latitudinal

and Longitudinal

1. Choose k number of clusters to be determined.

2. Choose k objects from dataset as the initial cluster centers of latitudinal and Longitudinal.

3. Repeat

3.1. Assign each object to their closest cluster to which object is the most similar based on mean value of the object in the cluster.

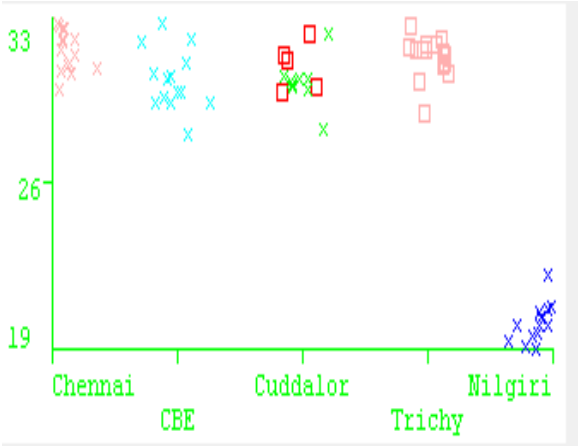
3.2. Compute new clusters, i.e. Calculate mean points of object in each cluster

4. until

4.1. No changes on cluster centers (i.e. Centroids do not change location any more)

V.Result in Clustering of monsoon data from various regions

Temperature and Humidity for Winter monsoon



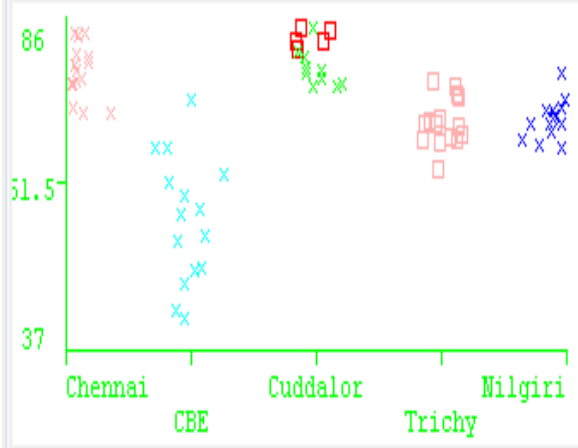
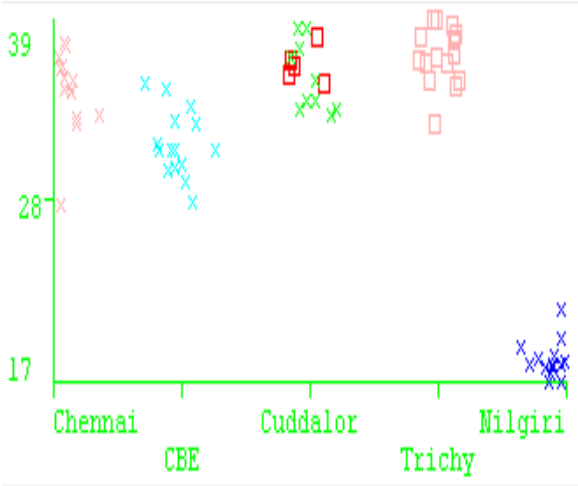
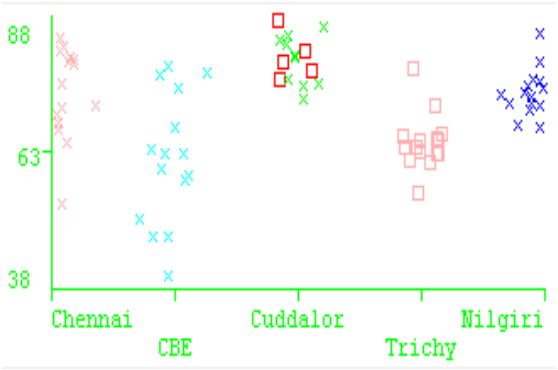


Fig.1 Winter Temperature
Fig. 2 Winter Humidity

The K –Means clustering algorithm well suited for this research analysis for grouping the temperature, Humidity and rainfall details in various stations. Here Fig. 1, represent the winter temperature data can be grouped in various region of Chennai Coimbatore, Cuddalore, Trichy and Nilgiri. Fig. 2, represents the Humidity denoted in percentage for Winter Season.

Temperature and Humidity for Hot Summer monsoon





From fig. 3, represent the Hot Summer temperature data can be grouped in various regions of Chennai Coimbatore, Cuddalore,

Trichy and Nilgiri. Here the temperature denotes in different range of degree in celcius. Fig. 4, represents the Humidity denotes in percentage for Hot Summer Season.

Reliability factor of temperature and Humidity for various regions in Tamil Nadu District

In this research work, the evaluation can be carried out on three stages. In input stage, the collection of data can be import to the data repository file for analyzing the coefficient factor for pair variable. In next stage, by using Karl Pearson's coefficient method the analysis can be evaluated based on the r equation .

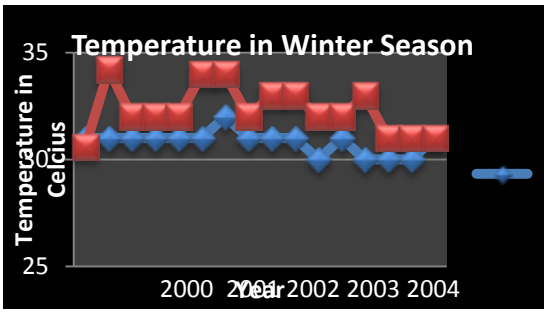
$$r = \frac{N\sum XY - (\sum X \sum Y)}{\sqrt{N\sum X^2 - (\sum X)^2}\sqrt{N\sum Y^2 - (\sum Y)^2}}$$

Here Wintemp, Stemp, SWtemp and NEtemp denote as X, domain variable of temperature and winhumidity, Sum humidity, swhumidity and NEhumidity denotes as Y, domain variable of Humidity.

Analysis of Correlation between Temperature and Humidity

Region	Relation Ship Between		Correlation
Chennai	Win temp	Win Humidity	0.068
	S Temp	Sum Humidity	0.225
	SW temp	SW Humidity	-0.274
	NE Temp	NE Humidity	0.201

Chennai Statistical Analysis in Temperature in Winter Season (January and February)



Predicting the Monsoon Rainfall using naïve bayes classification

Given random variables X,Y and Z, where X is conditionally independent of Y given Z, if and only if the probability distribution governing X is independent of the value of Y given Z; that is (∀ i, j, k)P(X = xi |Y = y j ,Z = zk) = P(X = xi |Z = zk

Consider three Boolean random variables to describe the current weather: Rain, Cyclone and Humidity. It might reasonably assert that cyclone is independent of Rain given Humidity. Because Cyclone causes Rainfall, once it know whether or not there is humidity, no additional information about cyclone is provided by the value of Rain.

Let us now derive the Naive Bayes algorithm, assuming in general that Y is any discrete-valued variable, and the attributes X1 ...Xn are any discrete or real valued attributes. Our goal is to train a classifier that will output the probability distribution over possible values of Y, for each new instance (Class label)X that it to classify. The expression for the probability that Y will take on its kith possible value, according to Bayes rule, is

$$P\left\langle Y = y_k \middle| X_{1.....} X_n \right\rangle = \frac{P(Y = y_k)P\left\langle X_{1.....} X_n \middle| Y = y_k \right\rangle}{\sum_j P(Y = y_j)P\left\langle X_{1.....} X_n \middle| Y = y_j \right\rangle}$$

VI.CONCLUSION

The K –Means clustering algorithm well suited for the research analysis of grouping the meteorological data which relates with different region based on its latitudinal and Longitudinal location.

Karl Pearson Correlation method is used, to analyze the relationships between the Rainfall, Humidity and the temperature using mathematical model.

When we determine correlation in Chennai region, southwest monsoon climate denotes negative correlation between the temperature and Humidity data and others are shows as positive correlated.

In Coimbatore region, hot summer, Northeast monsoon, Southwest monsoon are positively correlated between temperature and Humidity. In Cuddalore district, hot summer and southwest monsoon are positive correlated. In Trichy region, hot summer is positive correlated. Finally, in Nilgiri region all relations are positive correlated.

Based on Naïve Bayes Classification algorithm to identify the level of meteorological data of rainfall with respect to Cyclone and Humidity in seasonally to various region.

From this analysis, the rainfall occurs in mostly Southwest and North east Monsoon in Coimbatore, Chennai, Trichy and

Cuddalore. In Nilgiri, the most rainfall occurs in South West Monsoon.

VII.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. Lawrence, Mark G. "The relationship between relative humidity and the dew point temperature in moist air: A simple conversion and applications." Bulletin of the American Meteorological Society 86.2 (2005): 225-233.

5. Pasanen, A-L., et al. "Laboratory studies on the relationship between fungal growth and atmospheric temperature and humidity." Environment International17.4 (1991): 225-228. Swinbank, W. CQJR. "Long-wave radiation from clear skies." Quarterly Journal of the Royal Meteorological Society 89.381 (1963): 339-348.

6. 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.

7. Solomon, M. E. "Control of humidity with potassium hydroxide, sulphuric acid, or other solutions." Bulletin of Entomological Research 42.03 (1951): 543-554

8. Zhu, Xingquan, and Ian Davidson, eds. Knowledge Discovery and Data Mining: Challenges and Realities. Igi Global, 2007.

9. 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-

10. T.F. Gonzales. Clustering to minimize the maximum inter cluster distance. Theoretical Computer Science,1985,38(2-3):293-306.

11. S. Kotsiantis and et. al., "Using Data Mining Techniques for Estimating Minimum, Maximum and Average Daily Temperature Values", World Academy of Science, Engineering and Technology 2007 pp. 450-454