International Journal of Recent Advances in Multidisciplinary Research
Vol. 02, Issue 08, pp.0637-0642, August, 2015

Review Article

EFFICIENT ANALYSIS OF METEOROLOGICAL DATA BASED ON HUMIDITY LEVEL AND CYCLONE LEVEL USING NAÏVE BAYES CLASSIFICATION ALGORITHM

[1,*]Mayilvaganan, M. and [2]Vanitha, P.

[1]Department of Computer Science, PSG College of Arts and Science, Coimbatore, Tamil Nadu, India
[2]Department of Computer Science Karpagam University, Coimbatore, India

ARTICLE INFO

Article History:

Received 21st May 2015
Received in revised form
28th June, 2015
Accepted 13th July, 2015
Published online 31st August, 2015

Keywords:

Predict Meteorological Data,
Humidity,
Cyclone Level,
Rainfall Analysis,
Naïve Bayes Classification,
North East Monsoon,
South West Monsoon

ABSTRACT

The aim of this research work, focus meteorological data to analyses the monsoon seasons between twelve months to cluster the weather data based on Longitudinal and Latitudinal which can be used to analyze by the collection of Temperature, Humidity, Rainfall and Cyclone. In this research, Naïve Bayes Classification can be predicted the Meteorological condition of the region based on the cyclone and the humidity level in seasonally. It provides specific services for assessment of pollution impacts from various industries and thermal power plants. The atmospheric correlations play a significant role in determining the climate trends which are crucial in understanding the short and long-term trends in climate.

INTRODUCTION

Meteorological data analysis in the form of data mining is concerned to predict the knowledge of weather condition. To make an accurate prediction is one of the challenging of meteorologist to survey the weather condition efficiently. The atmospheric correlations play a significant role in determining the climate trends which are crucial in understanding the short and long-term trends in climate. The climate changes, so experienced today are mainly due to over ambitious strategies and actions of human beings on the eco-system. In this paper, the Naïve Bayes classification technique analyzes the meteorological data of five regions in various directions of Tamil Nadu.

Literature Survey

Folorunsho Olaiya and Adesesan Barnabas Adeyemo, reported the forecasting of maximum temperature, rainfall, evaporation and wind speed. This was carried out using Artificial Neural Network and Decision Tree algorithms and meteorological data collected between 2000 and 2009 from the city of Ibadan, Nigeria. A predictive Neural Network model was also discussed in the weather prediction program and the results compared with actual weather data for the predicted periods.

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, the domain structure of Seasonal Temperature, Humidity, Rainfall, cyclone occurring as shown in table II and Latitudinal and Longitudinal of several regions are taken as shown in Table 1.

Corresponding author: Mayilvaganan, M.,
Department of Computer Science, PSG College of Arts and Science, Coimbatore, Tamil Nadu, India.

Table 1. Latitudinal And Longitudinal of Region

Region	Latitudinal	Longitudinal
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

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

In the proposed methodology, from Fig 1, represents to collecting the meteorological weather analysis of temperature, humidity, rainfall and cyclone movements of the different region in Monsoon wise. The four divisions of monsoon seasonal data from the given dataset as Winter seasons in the month of January and February, Hot Summer Season in the month of March, April and May**,** South West Monsoon in the month of June, July, August and September and North East Monsoon of October, November and December month.

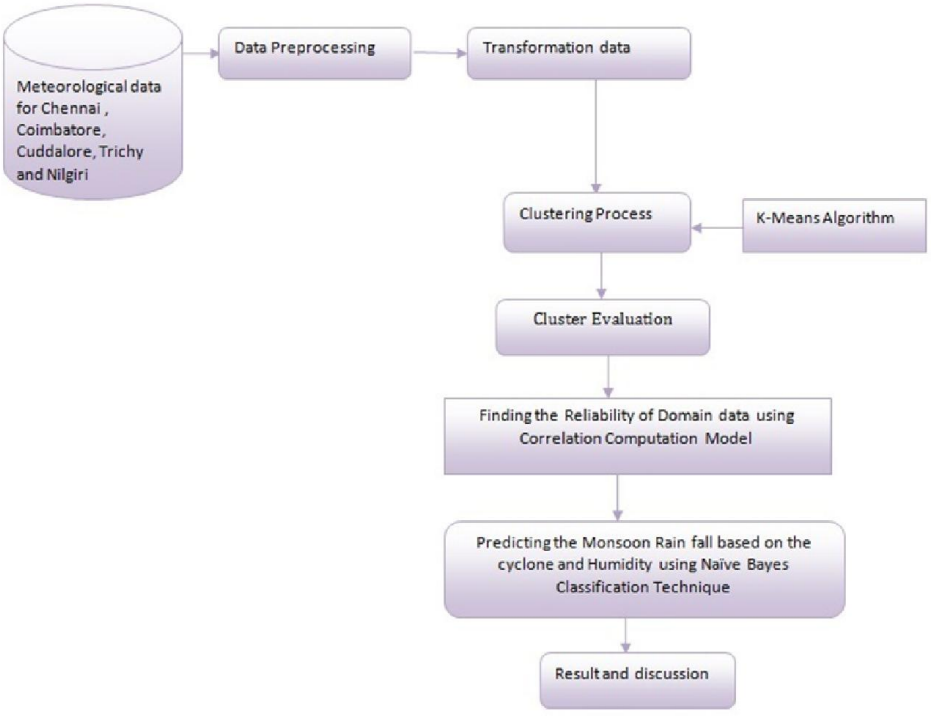
In this research, to Cluster the meteorological data based on the location of latitudinal and longitudinal, then to predict the rainfall level based on the level of changes in humidity and cyclone formation. The parameters along with Centroid Euclidean inter-cluster distance measure for the five clusters C0,C1,C2,C3 and C4 for grouping based on location using latitudinal and longitudinal to Cluster Centroid as given equation (2),

$$D_0(C_i,C_j)=\left[\sum_i (X^i_{centroid,i} - X^i_{centroid,j})\right]^{\frac{1}{2}} \quad (2)$$

Here Centroid based partition techniques are used the centroid of the cluster, Ci to represent that cluster distance measure based on longitudinal and latitudinal location Measured by dist (Ci, Cj) is Euclidean distance between two points. It can be clustered to different locations for analyzing temperature, humidity and rainfall level. From Fig2 represents, a longitudinal and Latitudinal data can be clustered into 5 clusters, which shows the distance to mean of inter cluster formation.

Table 2. Cyclone occurs in Monsoon

Year	W Cyclone (Jan- Feb)	Hot Sum Cyclone (Mar-May)	SW Cyclone (June-Sep)	NW Cyclone (Oct-Dec)
2000	0	1	1	0
2001	0	0	0	0
2002	0	0	0	3
2003	0	0	0	0
2004	0	0	0	0
2005	0	0	0	4
2006	0	0	0	0
2007	0	0	0	0
2008	0	1	4	5
2009	0	0	0	1
2010	0	0	0	1
2011	0	0	0	1
2012	0	0	0	1
2013	0	0	0	1
2014	0	0	0	0



1. Framework for Research Methodology

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 di stribution governing X is independent of the value of Y given Z; that is ($\forall i, j, k$) $P(X = x_i | Y = y_j, Z = z_k) = P(X = x_i | Z = z_k$
- 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 $X_1 \ldots X_n$ are any discrete or real valued attributes. Our goal is to train a classifier that will output the probability distribution over the 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 k^{th} possible value, according to Bayes rule (3), is

$$P\langle Y = y_k | X_1 \ldots X_n \rangle = \frac{P(Y = y_k) P\langle X_1 \ldots X_n | Y = y_k \rangle}{\sum_j P(Y = y_j) P\langle X_1 \ldots X_n | Y = y_j \rangle} \quad (3)$$

Computational Based on Training tuples such as
P(Wintemp=High| Winhumidity=high|Cyclone=yes)
P(WiPntemp=low| Winhumidity=high|Cyclone=no)
P(Wintemp=High| Winhumidity=medium|Cyclone=no)
P(Stemp=low|winhumdity=normal|cyclone=no)
P(Stemp=high|winhumdity=normal|cyclone=no)
P(Stemp=low|winhumdity=high|cyclone=no)
P(SWTemp=High|SWHumidity=high|Cyclone=Yes)
P(SWTemp=Low|SWHumidity=high|Cyclone=Yes)
P(SWTemp=High|SWHumidity=Medium|Cyclone=no)
P(NETemp=High|NEHumidity=high|Cyclone=Yes)
P(NETemp=Low|NEHumidity=high|Cyclone=Yes)
P(NETemp=High|NEHumidity=Medium|Cyclone=no)

RESULTS AND DISCUSSION

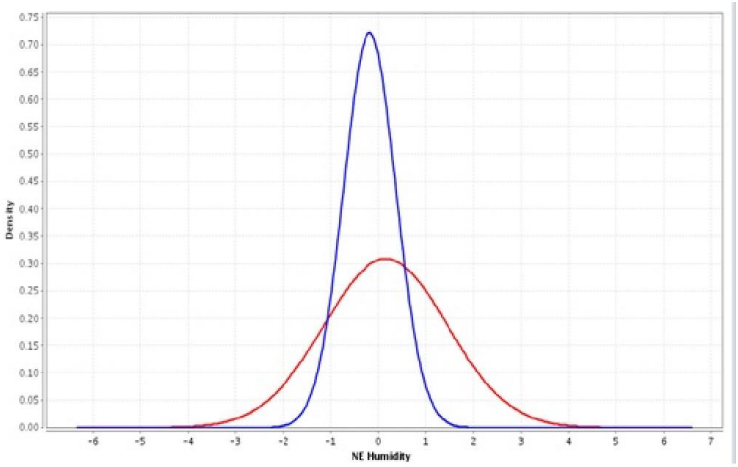
From the research, discussion can be analyzed the meteorological information can be experimented with classification techniques for predicting the level of rainfall based on the humidity level and cyclone levels from different

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
W Temp	30.818	31.175	30.667	30.167	19.800
W Cyclone	0	0	0	0	0
Win Humidity	77.545	69.475	50.333	83.333	70.147
W Normal	45.773	23.612	26.100	40.767	40.767
W Actual	32.291	15.962	3.633	100.783	27.227
S Temp	36.091	36.800	35.333	36.500	21.467
S Cyclone	0	0.050	0	0.167	0
S Humidity	76.727	66.750	71	80.667	73.367
S Normal	93.673	137.323	149.600	94.517	239.507
S Actual	158.127	154.672	120.100	61.900	254.493
SW Temp	36.182	34.425	31.667	36.500	17.533
SW Cyclone	0	0.125	0	0	0
SW Humidity	58	71.325	71.333	64.667	89
SW Normal	354.818	260.612	192.900	373.550	959.693
SW Actual	339.256	229.239	718.467	277.517	817.087
NE Temp	29.455	29.900	29	29.333	18.533
NW Cyclone	0.182	0.425	0	0.167	0.067
NE Humidity	83.727	80.350	80	84.500	78.133
NE Normal	650.882	330.200	327	726.483	387.373
NE Actual	936.375	314.505	332.100	447.167	458.920
NE Normal	650.882	330.200	327	726.483	387.373
NE Actual	936.375	314.505	332.100	447.167	458.920

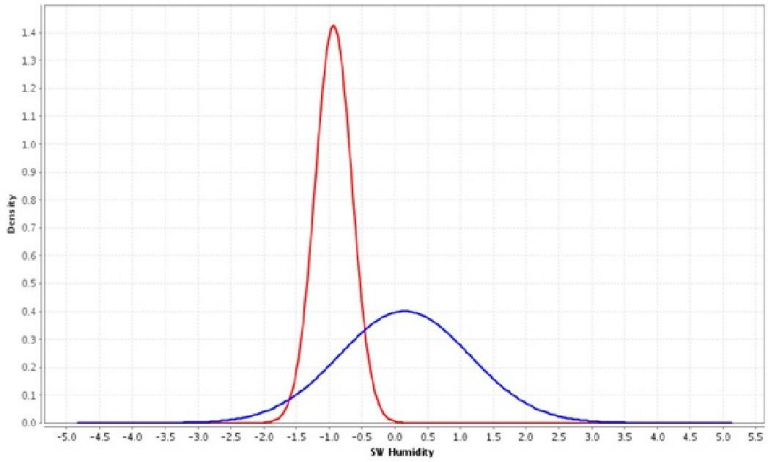
2. Distance to Mean of Inter Cluster Formation

regions in seasonally. From Fig. 3 represents the humidity level can increase in North East Monsoon based on cyclone occurrence in Chennai region. From Fig. 4 shows, when cyclone occurs in the density of 0.4 the humidity level of was increased in 1.4 of South West Monsoon in a region of Chennai. From Fig.5 shows the cuddalore region, when cyclone occurs in the density of 0.35 the humidity level of was increased in 0.65 in North East Monsoon rainfall range can be shown in the normal level during the cyclone forms in the year of 2000.

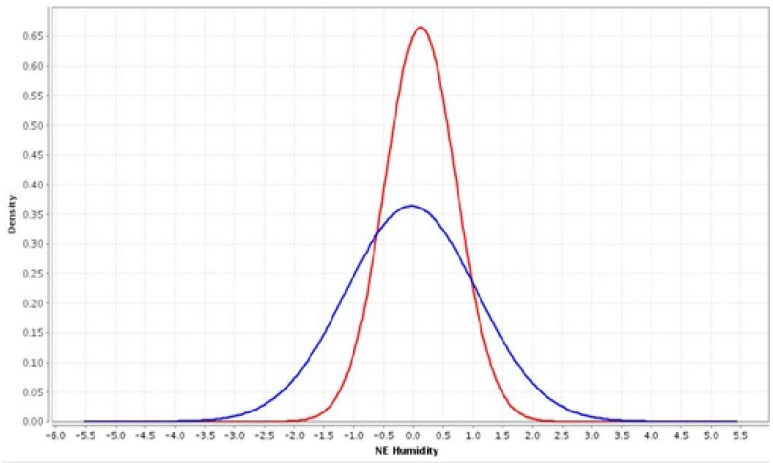
From Fig 6, represents the Coimbatore region, Based on humidity the North East Monsoon rainfall range can be shown in High level. From Fig 7, shows the Nilgiri region, Based on Humidity level of 0.5, North East Monsoon rainfall range can be shown in High level at 1.2. From Fig 8 represents the Nilgiri region, the humidity level of 0.40 occurs high rain at 0.75 density in South West Monsoon rainfall range can be shown in High level.



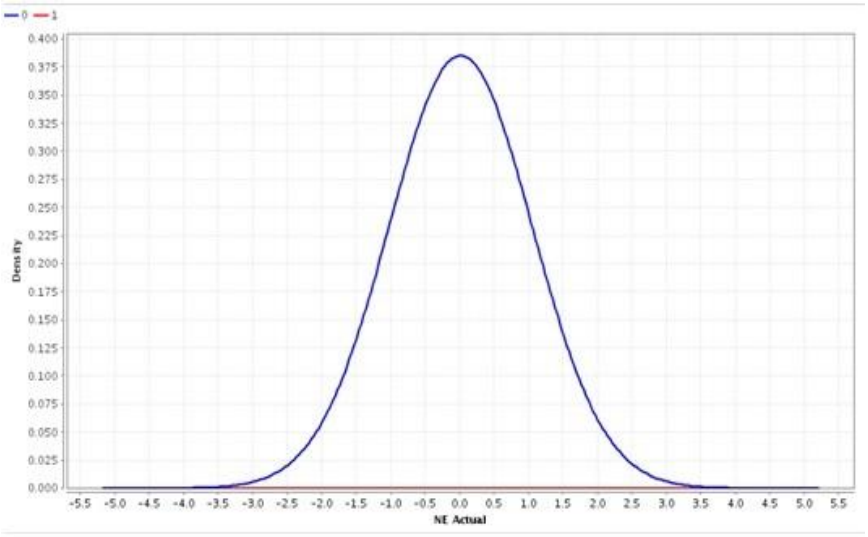
3. NORTH EAST MONSOON HUMIDITY LEVEL IN CHENNAI REGION



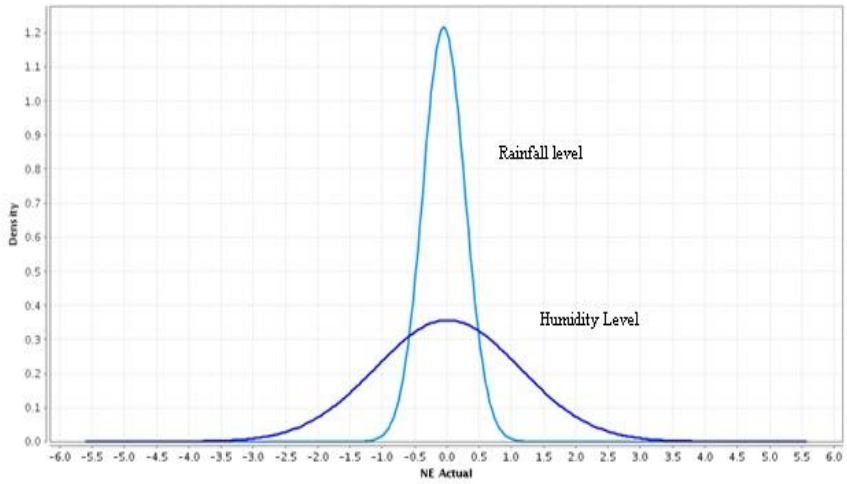
4. SOUTH WEST MONSOON HUMIDITY LEVE IN CHENNAI REGION



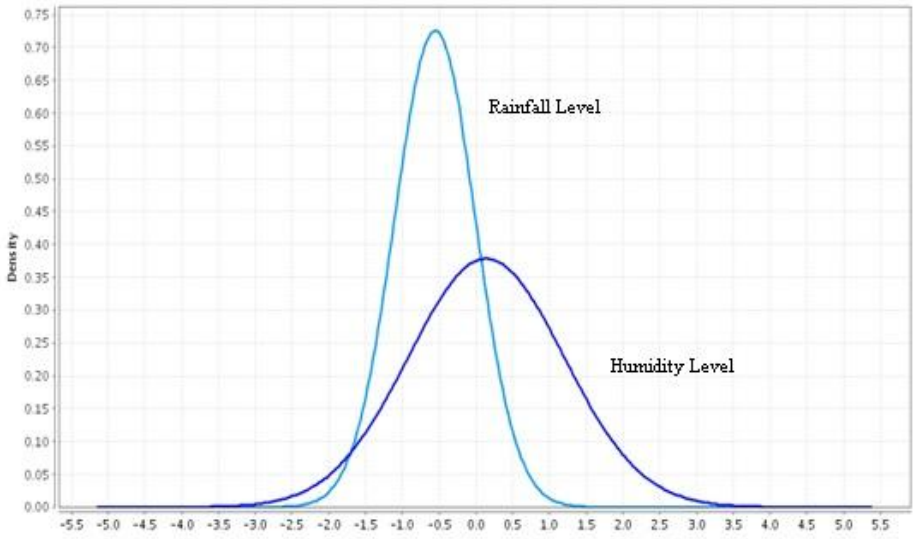
5. NORTH EAST MONSOON HUMIDITY LEVEL IN CUDALLORE REGION



6. NORTH EAST MONSOON HUMIDITY LEVEL FOR COIMBATORE REGION



7. NORTH EAST MONSOON OF RAIN FALL LEVEL ANALYSED BASED ON HUMIDITY LEVEL IN NILGIRI REGION



8. SOUTH WEST MONSOON OF RAINFALL LEVEL ANALYSED BASED ON HUMIDITY LEVEL IN NILGIRI REGION

Conclusion

In this research, it can be concluded that based on a 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. Thus the data mining provided a very accurate knowledge in a form of rules in classification techniques. This knowledge can be used to obtain useful prediction and support the decision making for support to the climatic environment in the ecosystem.

REFERENCES

Berkhin, Pavel. 2006. "A survey of clustering data mining techniques."Grouping multidimensional data. *Springer Berlin Heidelberg,* 25-71.

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

Kantardzic, 2011. Mehmed. Data mining: concepts, models, methods, and algorithms. John Wiley and Sons.

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

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

Manish Verma, MaulySrivastava, NehaChack, Atul Kumar Diswar and Nidhi Gupta, May-Jun 2012. "A Comparative Study of Various Clustering Algorithms in Data Mining", International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 3, pp.1379-

Nikhil Sethi, Dr.Kanwal Garg 2014. "Exploiting Data Mining Technique for Rainfall prediction", *International Journal of Computer Science and Information Technologies,* Vol. 5 (3), 3982-3984.

Nkrintra, S. et al., 2005. "Seasonal Forecasting of Thailand Summer Monsoon Rainfall", in *International Journal of Climatology*, Vol. 25, Issue 5, *American Meteorological Society*, pp. 649-664.

Olaiya, Folorunsho, and Adesesan Barnabas Adeyemo. 2012. "Application of data mining techniques in weather prediction and climate change studies."International Journal of Information Engineering and Electronic Business (IJIEEB) 4.151.

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

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

Thornton, Peter E., Steven W. Running, and Michael A. White. 1997. "Generating surfaces of daily meteorological variables over large regions of complex terrain."Journal of Hydrology 190.3 214-251.

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

****