I.J. Mathematical Sciences and Computing, **2015, 1, 11-20**
Published Online July 2015 in MECS (http://www.mecs-press.net)
DOI: 10.5815/ijmsc.2015.01.02

Available online at http://www.mecs-press.net/ijmsc

Prediction of Rainfall Using Unsupervised Model based Approach Using K-Means Algorithm

G.Vamsi Krishna

Research scholar, Department of CSE, GITAM University

Abstract

Prediction of rainfall has gained a significant importance because of many associated factors like cultivating, aqua-culture and other indirect parameters allied with the rainfall like global heat. Therefore it is necessary to predict the rainfall from the satellite images effectively. In this article, a segmentation algorithm is developed based on Gaussian mixture models. The initial parameters are estimated using k-means algorithm. The process is presented by using an 2-fold architecture, where in the first stage database creation is considered and the second stage talks about the prediction. The performance analysis is carried out using metrics like PSNR, IF and MSE. The developed model analyzes the satellite images and predicts the Rainfall efficiently.

Index Terms: Rainfall prediction, Gaussian mixture model, K-Means algorithm, rainfall estimation, PSNR, MSE.

1. Introduction

Weather forecasting is considered to be one of the most recurrent challenging issues. The prosperity and economic scenario of a country completely depends on the rainfall. Due to the increase in global heats, it is inevitable to predict the rainfall accurately. The prediction of rainfall is generally based on the cloud formulation. The clouds are classified into three groups, namely; nimbostratus, cumulonimbus and cumulus. Among these clouds, the first two types of clouds produce rainfall and the rainfall associated with the third type of cloud is rare. Therefore predictive methodologies are to be implemented for the identification of cloud and thereby estimating the probability of rainfall. Many models have been discussed in the literature about the cloud formation and prediction of the rainfall. In general, the onsite prediction of rainfall is carried out in two stages, stage-1 discusses about the dataset and extraction of the features from the dataset. In the stage-2, detection of the cloud is estimated. In order to portray the present methodology, the dataset is considered from

* Corresponding author. Tel.:
E-mail address: vamsikrishna527@gmail.com

satellite images from the meteorological department, Visakhapatnam. The study is based on the infrared visible regions and water vapor regions. These images are extracted from INSAT-3A and KALPANA-1. For the cloud detection, the satellite images are preprocessed and basing on the high reflectance of the cloud in the visible spectrum and low temperatures in infrared spectrum, the analysis is carried out. From the satellite images, the cloud top temperature is extracted using the infrared rays and basing on the precipitation, one can analyze, whether the cloud is a cold -cloud or a hot- cloud. In the absence of cloud, the satellite images process the data and record the data as infrared intensity. Cloud density helps to understand the thickness of the cloud which indirectly decides in the formation of the rain.

It is generally considered as a Boolean which returns a zero, when cloud is not present and intensity value which is equal to one, if cloud is present. Water vapor is another significant parameter that decides the status of the rainfall.

The architecture involved is presented below in Fig.1

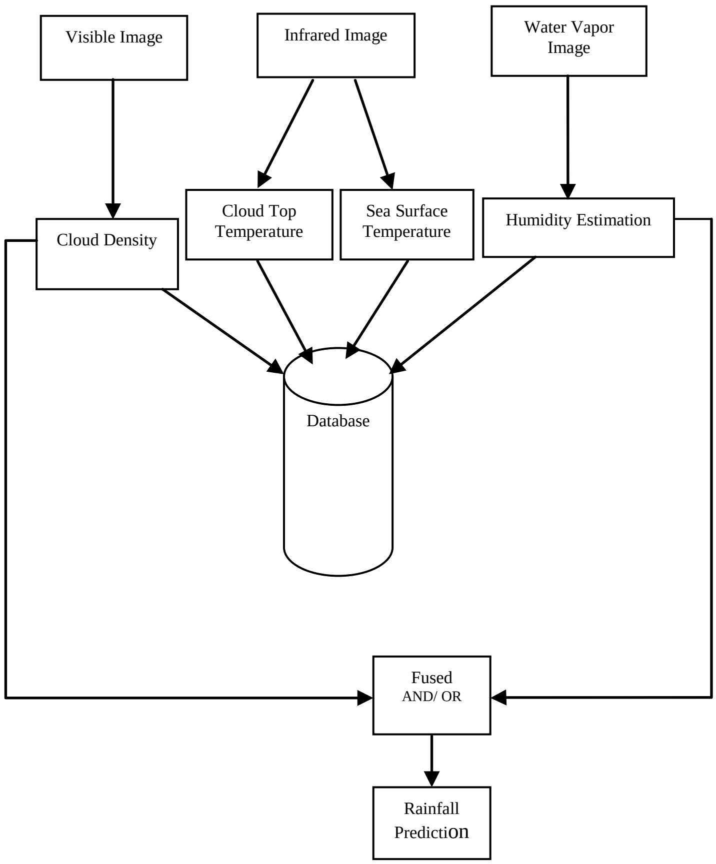

Fig. 1. Architecture

The research work in this area has been carried out by many meteorologists and most of the models developed by the eminent researchers could not able to predict the rainfall efficiently, because of the compression ratio which is associated with the retrieved images. The other factors associated with the satellite images include; snow, sand and icy regions, these associated factors affect the retrieval accuracy. Therefore effective methodologies are still to be involved for effective identification of cloud and formation of cloud.

Therefore, this paper highlights a model based on feature extraction together with clustering techniques based on k-means algorithm. The classification of clouds is based on the parameters which include water vapor, humidity and the other parameters estimated from Gaussian mixture model.

The rest of the paper is organized as follows section-2 of the paper deals with related work in the area; section-3 presents the feature extraction methodology. The clustering technique for segmenting the images based on K-Means algorithm is presented in section-4, in the section-5 of the paper highlights about the classification algorithm based on Gaussian mixture model, section-6 of the paper discusses about the fusion and the prediction of the rainfall is carried out by using KL-divergence which is highlights in section-7 of the paper. In section-8 the design and implementation is shown. In section-9 evaluation is highlighted and section-10 of the paper shows experiments and results and then the paper concludes.

2. Related Work

Many authors have presented there view of the prediction of rainfall. D.K.Richards and G.D.Sullivan has presented a methodology in which the prediction of the rainfall is based on the discrimination of the cloud for which the classification is done basing on the cloud formation [1]. M.K.Kundu and D.Priyank proposed a multiband wavelet transform for extracting the features from the cloud [2]. K.L.Hsu etal proposed a method for identifying the brightness of temperature in the infrared regions there by used these regions for classification of the rainfall [3]. Ian.H et al has proposed a model to printout the various shapes of cloud and proposed a model for identifying the type of cloud which results into rainfall [4]. Yanhong Tang et al in the paper proposed a methodology to extract the features for the cloud and also proposed that early methods have some disadvantages and they have modified the approaches to present a novel methodology and for identifying the features based on CVIR [5]. V.G.kin et al proposed a 3-D point estimation for identifying the cloudy regions [6]. Valmik.B et al proposed a model for the weather data prediction in which the classification methodology is considered the results of the experimentations are presented by correlating the various patterns [11]. F.Dell has proposed a framework for identifying the variations in the wind speed for there experimentations the database is classified into two groups namely hurricane group and non-hurricane group. In this process the weather data was extracted on a 24-hour basis and data preprocess techniques together with data cleaning is used to minimize the noise in the data and also the missing values are eliminated [12]. Classification techniques also help in forecasting the weather effectively and hence methodologies were developed by Sanjay Chakrabotry et al using time series analysis together with incremental k-means algorithms [13]. Marwa.F.Robe, Ala.M et al have utilized these data mining techniques for identification of effective rainfalls in Gaza district [14]. Abay Kumar et al, Zoher.N et al, M.Sabhita et al S.Sereften et al, Y.Mutirez et al have considered the data mining techniques for analysis of the weather data based on future extraction techniques like k-nerest neighborhood and k-means algorithms [15][16].

Other works in this area are presented on support vector machine(SVM) techniques by N.Rajashekar et al, Ankitha singh et al utilized artificial neural networks and Anitha et al, Gokila et al, Horis.H et al have presented models using support vector machines and neural networks as classifiers together with fuzzy c-means algorithm, k-means algorithm, hierarchal clustering algorithm for clustering the weather data[17][18].

However in most of these methodologies, the analysis for identifying the features that an associated for the prediction of rainfall could not be able to meet the requirement since they have eliminated the impact of the other features like water vapor, humidity which are essential for the prediction also these methodologies failed because of the other factors like sand, ice etc associated with the cloud. Therefore in the present methodology we use the concepts of data mining together with the concepts of fusion for effective interpretation of the rainfall.

3. Future Extraction

Future extraction plays a vital role in identifying the significant data based on the patterns. This data helps in better classification or clustering many clustering algorithms are used. The main focus of the clustering algorithm is that they help in better identification of the data by labeling the data from which the data can be portioned. The proposed model is based on 2-tier architecture where in the first stage the satellite image extracted is converted into grayscale. The segmentation is carried out using k-means algorithm to categorize different types of clouds and also identify the initial parameters of each of the cloud regions. These initial parameters are given as input to Gaussian mixture model to estimate the probability density functions.

4. K-Means Algorithm

For effective segmentation of the images into different regions is carried out using k-means algorithm. However to initialize the value of k the histogram of each image extracted for satellite and considered and basing on the number of peaks and value of k is decided. This k is given as input to the k-means algorithm for effective analysis

K-Means algorithm is as follows:

Step 1: Begin with the initial value of k.

Step 2: Select number of clusters.

Step 3: Partition the input pixels into k clusters by assigning each pixel xi to the closest centroid v using any of the distance measure.

Step 4: Compute the cluster assignment metrics using

$$\sum_{i=1}^{k} u_{ij} = 1 \forall i \,\&\, j$$

Where u is the cluster assignment matrix.

Step 5: If the cluster centroid or assigned matrix changes, repeat the above steps.

5. Gaussian Mixture Model

In order to classify the data many models are presented in the literature among which the models are being classified as parametric based which involved the shape, size and other parameters of the data into consideration while segmenting the data and the second methodology deals with non-parametric models such as SVM’s artificial neural networks and other models without considering the attributes of the model. Among these models parametric models are rated to be the best for classifying the data under severe involvement of image parameters. Among this parametric models Gaussian mixture models are mostly focused. The main advantage is that its ability to scatter for long data sets having infinite peaks. In this chapter the Gaussian mixture model to classify the weather data. The probability density function of Gaussian mixture model is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$-\infty \leq x \leq \infty, x \geq 0 \quad (1)$$

Using the initial estimates from the k-means algorithm the pdf for each segment is calculated

6. Fussion

In order to have better classification accuracy the satellite images are processed and the water vapour region and cloud density region are calculated. These regions are fussed along with the pdf values obtained in earlier section to have a effective future vector. In order to fuse the data we are logical fusion which consists AND/OR operations.

7. KL-Divergence

To test the type of the cloud, we consider the query image and extract the features to identify the type of cloud. This query image is tested for relevance against the satellite images to identify the type of the cloud. For the comparison, we need to make a relevance match and for which KL-Divergence is used. The input image extracted from the weather data from the satellite images are considered as the testing data to find the relevance we use the KL-divergence which is given by

$$KL(p_1, p_2) = \int p_1(x) \log\left(\frac{p_1(x)}{p_2(x)}\right) dx \tag{2}$$

Where p1 is the image under consideration which is extracted from the satellite images and p2 is the image already available in the data base. If the value is minimum we say that the test data is relevant to the training data for which it matches.

8. Design and Implementation

8.1. Implementation Methods

The execution of the proposed methodology consists of six states. In the first state, the data is collected and the status of the cloud is identified in the second phase. In the third phase, the status of the cloud is verified, thereby the type of the cloud is identified in the next phase, and basing on these stages, the last stage decides the possibility of the rain fall.

Table 1. Attributes of Meteorological Dataset

Attribute	Type	Description
Year	Numerical	Year considered
Month	Numerical	Month considered
Wind speed	Numerical	Wind run in km
Evaporation	Numerical	Evaporation
CloudForm	Numerical	The mean cloud amount
Radiation	Numerical	The amount of radiation
Sunshine	Numerical	The amount of sunshine
MinTemp	Numerical	The monthly Minimum Temperature
Rainfall	Numerical	Total monthly rainfall
MaxTemp	Numerical	Maximum Temperature

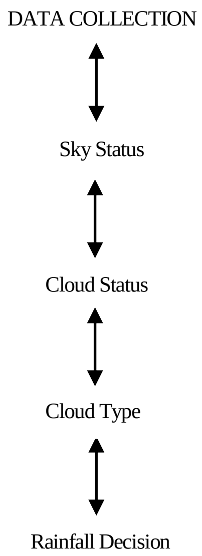

Fig. 2. Implementation Methods

Fig. 3. a) Sample Input 1 b) Sample input 2

8.2. Sky Status

To underline the sky status, in this paper we have considered the Gaussian Mixture model.

Fig. 4. a) Sky status of Input 1 b) Sky status of Input 2

The main intension is to identify the cloud, sot the cloud data is to be separated such that the cloud points are isolated. A threshold is chosen for this purpose, the image region is considered to be a cloud region if it is in the range > 50 and <200. While identifying the cloud regions, the sky status obtained using GMM is separated.

8.3. Cloud Status

In the identification process of the cloud status, the cloud region varies in thickness, so if the thickness is identified, then the cloud status can be modeled, the high density portions in the images are generally considered as the cloud regions. The central core part in the satellite images are considered to be the sky regions. In mild time or moody time sky is considered as cloud. To find the mask the histogram equalization is used the value with the highest

8.4. Cloud Type

The Cloud type is established by using clustering perception. In particular, in this paper, we have considered the K-Means clustering algorithm in order to find the type of cloud.

9. Evaluation Metrics

In choosing the most relevant cloud that matches the against the cloud images in the data set, evaluation metrics are used. In this paper, we have considered two evaluation parameters, Peak Signal to Noise Ratio (PSNR) and Mean Squared Error (MSE)

9.1. Mean Squared Error

Mean squared error is one of the most commonly used measures of success for numeric prediction. This value is computed by taking the average of the squared differences between each computed value and its corresponding correct value.

9.2. Peak Signal to Noise Ratio

It is defined as 1/MSE and more the output value, the better accuracy.

10. Experiments and Results

The experimentation is carried out in MATLAB environment where the extracted images from the database are stored along with the demarcation of the type of clouds each image from the satellite are considered for testing and preprocessed using the k-means algorithm present in section-4 of the paper, the segmentation is carried out and the segmentation regions are identified. The various cloud images obtained are presented in figures-5(a) to 5 (d)

Using the infrared frequencies the water vapor and cloud density regions are estimated from regions of the pixels and extracted and fussed with that of the pixels in each cloud. Using the logical fusion technique the results are obtained are presented below. For the analysis of the work metrics like PSNR and MSE are used.

The results arrived from the above metrics in table 2 of the data.

Table 2. Results obtained after the Performance Evaluation

Image	Performance Measure			
	PSNR		MSE	
	Without K-means	With K-means	Without K-means	With K-means
Im-1	32.5	39.98	0.54	0.59
Im-2	34.56	41.32	0.44	0.51
Im-3	33.76	38.55	0.32	0.45
Im-4	23.65	29.54	0.29	0.34
Im-5	29.65	35.34	0.38	0.51
Im-6	41.34	47.77	0.38	0.45
Im-7	29.90	39.19	0.35	0.55
Im-8	33.87	43.39	0.31	0.43
Im-9	41.34	44.66	0.22	0.39
Im-10	22.44	27.72	0.43	0.67
Im-11	32.85	39.98	0.37	0.55
Im-12	28.91	43.19	0.45	0.61
Im-13	22.56	31.34	0.49	0.64
Im-14	32.54	39.98	0.51	0.69

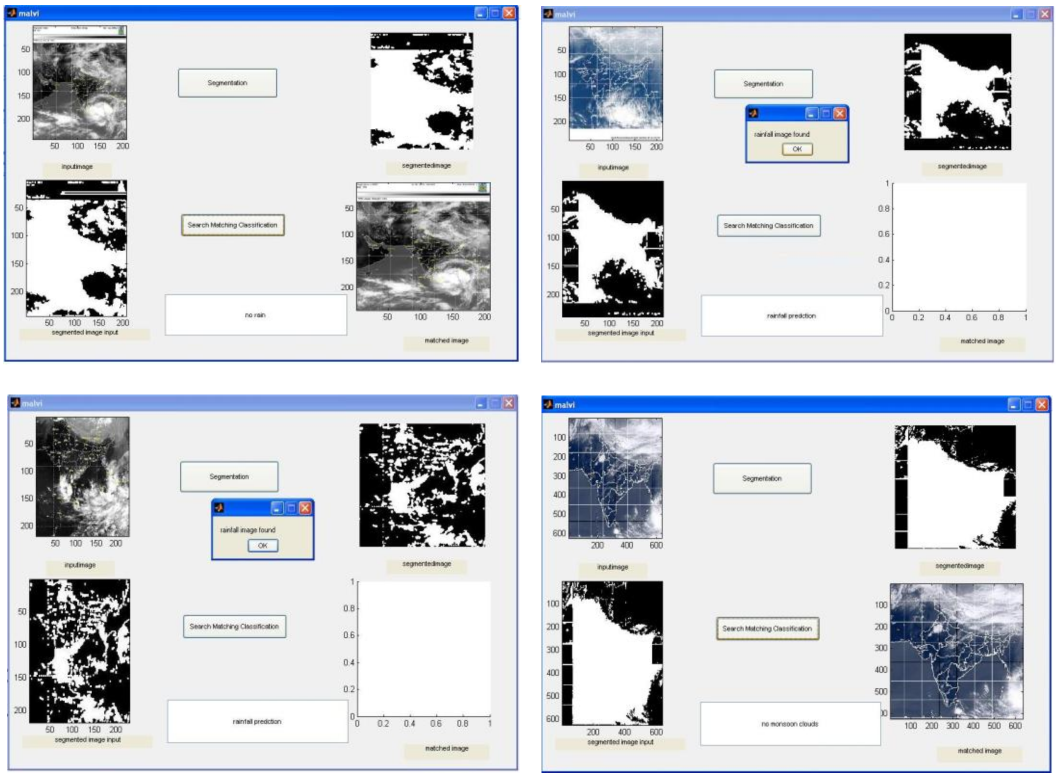

Fig. 5. (a) to (d) : Cloud images

11. Conclusions

The category of cloud is predictable and the rainfall is anticipated by using the proposed methodology based on Gaussian Mixture model together with K- means Clustering. The performance is tested in the presence of the clustering algorithm and also discarding the clustering algorithm. In order to test the performance, performance metrics such as Peak Signal to Noise Ratio and Mean squared error are considered. The output is presented in Table-1. From the above table-1, it could be clearly understood that the model performs better in the presence of the clustering algorithm. The PSNR Values are high in case, when the clustering algorithm, which signifies that the input images are more closely in accordance with the database images, and also the MSE values of the proposed method in case of K-Means clustering is less The status of sky is found the Gaussian Mixture model. The class of cloud is established using the Cloud Mask Algorithm together with the histogram equalization. The nature of cloud can be identified by using the K- Means Clustering technique. The category of rainfall cloud is predicted by analyzing the color and density of the cloud images. The cloud images considered are stored in the database in JPEG format. Analysis was done over several images. The result predicts the type of cloud together with the additional information such as, type of the cloud, the cloud position, altitude and appearance. It also provides the information about the possibility of the rainfall.

References

[1] K. Richards and G.D. Sullivan," Estimation of Cloud Cover using Colour and Texture" Intelligent Systems Group, University of Reading, RG6 2AY,2006.

[2] Malay K. Kundu and Priyank Bagrecha,"Color Image Retrieval Using M-Band Wavelet Transform Based Color-Texture Feature".

[3] Kuo-Lin Hsu, X. Gao, and Soroosh Sorooshian,"Rainfall Estimation Using Cloud Texture Classification Mapping".

[4] Liu Jian and Xu Jianmin," An Automated, Dynamic Threshold Cloud Detection Algorithm for FY-2C Images", National Satellite Meteorological Center, Beijing, 100081, China.

[5] Anuj Srivastava and Ian H. Jermyn, "Looking for Shapes in Two-Dimensional, Cluttered Point Clouds", IEEE Transaction.

[6] Yanling Hao, Wei ShangGuan,Yi Zhu, and YanHong Tang," Contented-Based Satellite Cloud Image Processing and Information Retrieval".

[7] Aleksey Golovinskiy, Vladimir G. Kim and Thomas Funkhouser,"Shape-based Recognition of 3D Point Clouds in Urban Environments".

[8] Peter S. Masika, "Cloud height determination and comparison with observed rainfall by using meteosat second generation (msg) imageries",Kenya Meteorological Department, 2006.

[9] Shou Yixuan, Li Shenshen, Shou Shaowen and Zhao Zhongming, "Application of a cloud-texture analysis scheme to the cloud cluster structure recognition and rainfall estimation in a mesoscale rainstorm process", Advances in Atmospheric Sciences, Science Press, co-published with Springer-Verlag GmbH, 0256-1530 (Print) 1861-9533 (Online), Volume 23, Number 5 / October, 2006, 767-774.

[10] Wei Shangguan; Yanling Hao; Zhizhong Lu; Peng Wu, "The Research of Satellite Cloud Image Recognition Base on Variational Method and Texture Feature Analysis", Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on Volume, Issue , 23-25 May 2007.

[11] Valmik B Nikam and B.B. Meshram, "Modeling Rainfall Prediction using DataMining Method", Fifth International Conference on Computational Intelligence, Modeling and Simulation, Issue No: 2166-8531, PP:132-136, 2013.

[12] F.Dell Acqua and P.Gamba, "A simple Model Approach to the problem of meteorological object Tracking", Volume: 1, Issue No: 7803-6359, PP: 2152-2154, Italy, 2000.

[13] Sanjay Chakraborty, Prof.N.K. Nagwani, Lopamudra dey, “Weather Forecasting using Incremental K-means Clustering”, International Conference in High Performance Architechture and Grid Computing, Vol.169, Part-2, PP:338-341,2011.
[14] Marwa F.AI-Roby, Alaa M.El-Halees, Data Mining Techniques for Wind Speed Analysis”,Journal of Computer Engineering, ISSN:2010-1619, Vol-2 ,No:1, PP:1-5, 2011.
[15] Abhay Kumar, Ramnish Sinha, Daya Shankar Verma, Vandhana Bhattacherjee, Satendra Singh (2012), “Modeling using K-Means Clustering Algorithm”, First International Conference on Recent Advances in Information Technology, Vol:4, Issue-1, Issue No: 4577-0697, PP:1-5.
[16] Zohreh Nazeri and Jianping Zhang “Mining Aviation Data to Understand Impacts of Severe Weather on Airspace Syatem Performance”, Proceedings of the International Conference on Information Technology: Coding and Computing [ITCC’OL], Volume-1, ISSN: 7695-1506, PP: 1-6, 2002.
[17] N.RAJASHEKAR and T.V.RAJINIKANTH, “weather analysis of Guntur district of Andhra region hybrid SVM data mining techniques”, international journal of engineering and advanced technology (IJEAT), ISSN: 2449-8958, Volume-3, issue-4, PP: 133-136, April 2014.
[18] ANKITH SINGH, Dr. BHUPESH GOUR, ANSHUL KHANDELWAL and HARSHA LACKHWANI, “an efficient clustering method for atmospheric conditions prediction using art algorithm”, international journal of advanced research in computer engineering and technology, volume-1, issue-1, PP: 12-17, march 2012.
[19] HANS-HENRIK BENZON and THOMAS BOVITH, “simulation and prediction of weather radar clutter using a wave propagator on high resolution NWP data”, IEEE transactions an antennas and propagation, volume: 56, No: 12, PP: 3885-3890, December 2008.
[20] Shaminder Singh and Jasmeen Gill,”Temporal Weather Prediction using Back Propagation based Genetic Algorithm Technique”, DOI: 10.5815/ijisa.2014.12.08,PP:55-61.
[21] Folorunsho Olaiya and Adesesan Barnabas Adeyemo “Application of Data Mining Techniques in Weather Prediction and Climate Change Studies”, DOI: 10.5815/ijieeb.2012.01.07,PP:51-59.
[22] Shaminder Singh” Temporal Weather Prediction using Back Propagation based Genetic Algorithm Technique” .J. Intelligent Systems and Applications, 2014, 12, DOI: 10.5815/ijisa.2014.12.08 PP 55-61.

Author’s Profile

G.Vamsi Krishna (born April 9, 1989) is a Research Scholar at the Computer Science and Engineering Department of GITAM University, Visakhapatnam, Andhra Pradesh, India. He is currently pursuing his PhD, in the area of data mining and in particular on Weather Forecasting Models. He holds a Masters Degree in Computer Science and Technology, from Andhra University; his areas of interest include programming languages, Data Mining

How to cite this paper: G.Vamsi Krishna,"Prediction of Rainfall Using Unsupervised Model based Approach Using K-Means Algorithm", International Journal of Mathematical Sciences and Computing(IJMSC), Vol.1, No.1, pp.11-20, 2015.DOI: 10.5815/ijmsc.2015.01.02