

GIS-MAP based Spatial Analysis of Rainfall Data of Andhra Pradesh and Telangana States using R

Y. Jeevan Nagendra Kumar¹, T. V. Rajini Kanth²

¹Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India

²Department of Computer Science Engineering, SNIST, Hyderabad

Article Info

Article history:

Received Jun 5, 2016

Revised Aug 31, 2016

Accepted Sep 15, 2016

Keyword:

Hybrid machine learning techniques

Map

Rainfall patterns

Spatial analysis

Visual analytics

ABSTRACT

The rainfall conditions across wide geographical location and varied topographic conditions of India throw challenge to researchers and scientists in predicting rainfall effectively. India is Agriculture based country and it mainly depends on rainfall. Seasons in India are divided into four, which is winter in January and February, summer is from March to May, monsoon is from June to September and post monsoon is from October to December. India is Agriculture based country and it mainly depends on rainfall. It is very difficult to develop suitable rainfall patterns from the highly volatile weather conditions. In this Paper, it is proposed that Map based Spatial Analysis of rainfall data of Andhra Pradesh and Telangana states is made using R software apart from Hybrid Machine learning techniques. A Study will be made on rainfall patterns based on spatial locations. The Visual analytics were also made for effective study using statistical methods and Data Mining Techniques. This paper also introduced Spatial mining for effective retrieval of Remote sensed Data to deal with retrieval of information from the database and presents them in the form of map using R software.

Copyright © 2017 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Yalla Jeevan Nagendra Kumar,

Departement of Information Technology,

Gokaraju Rangaraju Institute of Engineering and Technology,

Flat No: 303, Datta Sree Avenue, Annapurna Enclave, Chanda nagar, Hyderabad 500050, Telangana, India.

Email: jeevannagendra@gmail.com

1. INTRODUCTION

India has an ample variety of climatic circumstances across geological range and diverse geography which is very difficult for generalization. India has six weather sub categories, ranging from alpine tundra, glaciers in the north, humid tropical regions supporting rainforests in the southwest, the island territories and arid desert in the west. Many areas have different micro climates. Its geology and topography are climatically crucial. In North West there is a Thar Desert and in North side there are Himalayas. They work in cycle to affect ethnically and cost-effectively important for monsoonal system. India is considered as climatically unstable because of monsoonal, tropics, and other conditions. The summer, monsoon, thunderstorms and wet seasons are controls the climate in India [13]. Southeast storms are originating from a high-stress mass centered Indian Ocean and the monsoonal torrents provide over 80% of India's yearly rainfall.

Rainfall data [1] is one of the meteorological parameters that have a greater bearing on the livelihood of individuals' world over. The main objective was to develop a spatial framework for flood and drought disasters in Zambia. Precise inference of the temporal and spatial allocation of rainfall [2] is a vital input parameter for hydrologic reproduction and substantiation.

The rain measures used to monitor rainfall is insufficient to settle the spatial and chronological allocations of rainfall. New techniques were developed to increase the precision of radar rainfall

approximations, and construct high spatial and chronological rainfall data in standardization and rationalization. India is an agricultural country and most of economy of India depends upon the agriculture. Rainfall plays an important role in agriculture so premature prediction of rainfall is necessary for the better economic growth of our country. Rainfall prediction [3] has been one of the most challenging concerns around the world in last year. Regression analysis, classification, clustering, and Artificial Neural Networks (ANN) etc. techniques are using widely in prediction process.

2. LITERATURE SURVEY

K.S. Reddy et al [4] evaluated LARS-WG model for southern Telangana and Andhra Pradesh region, used thirty year climate statistics from 1980 to 2010 to produce the enduring climate series for 2011 - 2060. The version forecasted the rise in standard yearly rainfall in 2030 is 5.16% and in 2060 is 9.5% for Yacharam related to Hayath nagar. It was ranked as best model in terms of effectiveness in all selected Rangareddy mandals of Telangana and is applied to all other mandals because the climatic conditions are similar in those regions.

Ishappa Muniyappa Rathod et al [5] identified that the rainfall is the vital aspect which decides the crop and yield model of a region. The success and failure of the crop depends upon the climatic circumstances. They studied the precipitation traits like spatial allocation, seasonal variability of the Coimbatore district. The research is built on Forty Nine years of rainfall data for thirty three rain scale places. There is a heavy precipitation in north, south parts and less precipitation in east part of the district.

Rajinikanth TV et al [6] stated that there are a quite a lot of climatic conditions in various time periods that are varied geologically. It has substantial precipitation in Chirapunji, high warmth at Rajasthan and cold environment at Himalayas. These extremes make us uncomfortable and predictions of climate requires systematic approaches like machine learning procedures, K-means algorithm, J48 classification methods for efficient study and extrapolations of climatic conditions.

Kusre B.C. et al [7] analyzed the spatiotemporal disparity of the precipitation in Nagaland. The study illustrates that there is a huge dissimilarity in the precipitation with disparity from 859 mm to 2123 mm. Yearly rainfall model indicates the northern part has high rainfall as related to east, west of Nagaland. In the same way the north part receives more rainfall in monsoon season and less rainfall in winter season as related to east and west part of Nagaland.

Marc G. Genton et al [8] stated that the use of vigorous geo-statistical techniques on the statistics of rainfall dimensions for Switzerland. They are de-trended through non parametric approximation with leveling factor. The finest trend is calculated with a flattening factor based on cross validation. The variogram is then calculated by a vigorous evaluator. The parametric variogram prototype is comprehended by considering variance – covariance composition of variogram approximates. Fascinating outcomes are yielded by comparing kriging with initial quantities. All of these estimates are done with new methods in ‘S+ SPATIALSTATS’ software.

C.Sarala et al [9] stated that rainfall is irregular in India. It presents rainfall analysis by taking geological method in preparation of maps in geographical systems and characterizing spatial, temporal dissemination of monthly and yearly rainfall in Telangana with the help of trend exploration. The initial study is built on the information from 10 districts and 457 mandals. In this analysis, numerous GIS remote sensor practices were used by incorporating various geo reference data sets in the generation of maps of rainfall in Telangana.

S. Nagini, Rajinikanth T.V. et al [10] stated in their paper titled “Effective Analysis of Land Surface Water Resources of Andhra Pradesh with Rough Set based Hybrid Data Mining Techniques Using R”, that Agriculture plays an important role in economy of India. More than half of the population in India depends on Agriculture. It provides raw material for many Industries. In early days, more than half of the land mass is used for Agriculture and over the years there is decline in agriculture land. Various factors like the urbanization and development results in the growth of Non-Agriculture land year by year. Agriculture is the largest abstractor and prime consumer of ground water resources across the globe and hence studies of agro-economies that are ground water dependent became widely popular. Agriculture Irrigation, Surface water and Ground water resources are interlinked to each other. Water Usage and Food Production are dependent on each other extensively. Water is the major parameter that controls the crop yield. In many countries, the agriculture yield depends on the rain fall. Many times, the rainfall is not sufficient to crop yields. It made researchers to do rigorous analysis on water resource availability and suggest farmers for its effective utilization. This paper aims at, development and application of new Hybrid Data Mining (HDM) Techniques for effective analysis of Land Surface Water Resources (LSWR) like Canals, Tube wells, Tanks and other water resources. Apart from that analysis is also made on various Agriculture yield’s i.e., both for Cereals and Millets namely Kharif, Rabi, Sugarcane, Maize, Ragi, Wheat, Barley, etc., using new Hybrid Data

Mining (HDM) techniques. To model the complex logic, Decision Tables (DT) is used. The results were proved to be good when new Rough Set Based Hybrid Data Mining (RSBHDM) Techniques are applied over the refined data sets.

N. Rajasekhar, RajiniKanth T.V. [11] stated that Weather Prediction is the application of science and technology to estimate the state of the atmosphere at the particular spatial location. Due to the availability of huge data researchers, got interest to analyze and forecast the weather. Accurate prediction helps the human being existence and prosperity. Forecasting techniques are helpful in predicting natural disasters, crop and jungle growth, nautical routing, air craft scheming and armed functions. The Data Mining techniques are better than the obtainable methodologies or conventional methods. They were projected hybrid support vector machine replica to predict, analyze the climatic data and to discover the prototypes exist in it. They considered Krishna district climate data for the case study and it produced high quality results rather than machine learning methods in the process of prediction.

Ananthoju Vijay Kumar, RajiniKanth T.V. [12] stated that the rainfall has intense consequence on agriculture. A standard rainfall is crucial for vegetation. Excessive or diminutive rainfall can damage cultivation. Diminutive can abolish cultivation and excessive can help to grow dangerous fungus. Cultivation in India largely depends on rainfall, so an effort is made to forecast the stimulus of rainfall on harvest of groundnut. For this the data set is constructed with yearly capacities of crop and rainfall for sixty two years. The data was gathered from various Government sectors. The investigation exposed that the crop is destructively prejudiced by rainfall

3. PROPOSED APPROACH

In the Proposed approach initially the various years rainfall data of Andhra Pradesh, Telangana was taken and preprocessed for cleaning, removal of redundancy, filling the missing values with suitable mean values and molded into required format. Then apply hybridization of Data Mining (HDM) Techniques on the preprocessed Rainfall dataset. The results thus obtained were analyzed effectively by constructing various GIS Maps using the Rainfall data set with the help of R software [14]. It has proved that there is a substantial progress in performance.

4. IMPLEMENTATION OF PROPOSED METHODOLOGY

Initially the raw spatial data set is Pre-processed and converted in to the required format thus obtained is called refined spatial data set, suitable for further processing. Info-Gain Attribute Evaluation procedure along with Ranker Algorithm is applied and attributes selection was done. This concept finds the value of an attribute by measuring information gain for a given class. The optimized spatial data set is divided into Train data set and Test data set. It is then subjected to Machine learning Algorithm namely Classification algorithm of Data mining technique called J48 tree classification. The performance is calculated and the resultant decision Tree J48 classifier with refined data set is subjected to Association Rule Mining Algorithm namely Apriori Algorithm. Then the generated Association Rules will be analyzed for the patterns. The refined spatial data set is used to construct customized maps [17] using required R software code. The visual analytics were used for spatial analysis of the rainfall data sets of Andhra Pradesh and Telangana states.

5. RESULTS AND ANALYSIS

The attributes of the Rainfall data set are namely State, District, Latitude, Longitude, Year, January, March, April, May, June, July, August, September, October, November, December and Annual Total. The Info-Gain of the attributes was calculated and it was found that, except the attributes namely State, District, Longitude, Latitude, May, July, August, October, November, December, other attributes has zero Info-Gain values. After that the Classification Algorithm Class for generating a pruned or un-pruned C4.5 decision tree known as J48 [15] is applied and the resultant Classifier Decision Tree represented by Figure 1. This Decision tree says that the Andhra Pradesh and Telangana lies in between Longitude boundaries are 77.601, 83.897 where as the Latitude boundaries are 19.664, 13.217.

The resultant decision tree classifier data set is subjected to again Info Gain and removed the attributes whose Gain value is zero or near to it. The final attributes are State, District, Longitude, Latitude, May, July, August, October, November and December. Over this data set, Predictive Apriori Algorithm was applied. The Predictive Apriori Algorithm [16] is to extract association rules in the given Class. The algorithm discovers the best rules by considering the threshold and support based confidence rate. It adds

those rules which are on far with the anticipated accuracy. The Association Rules were found and they are listed Figure 2.

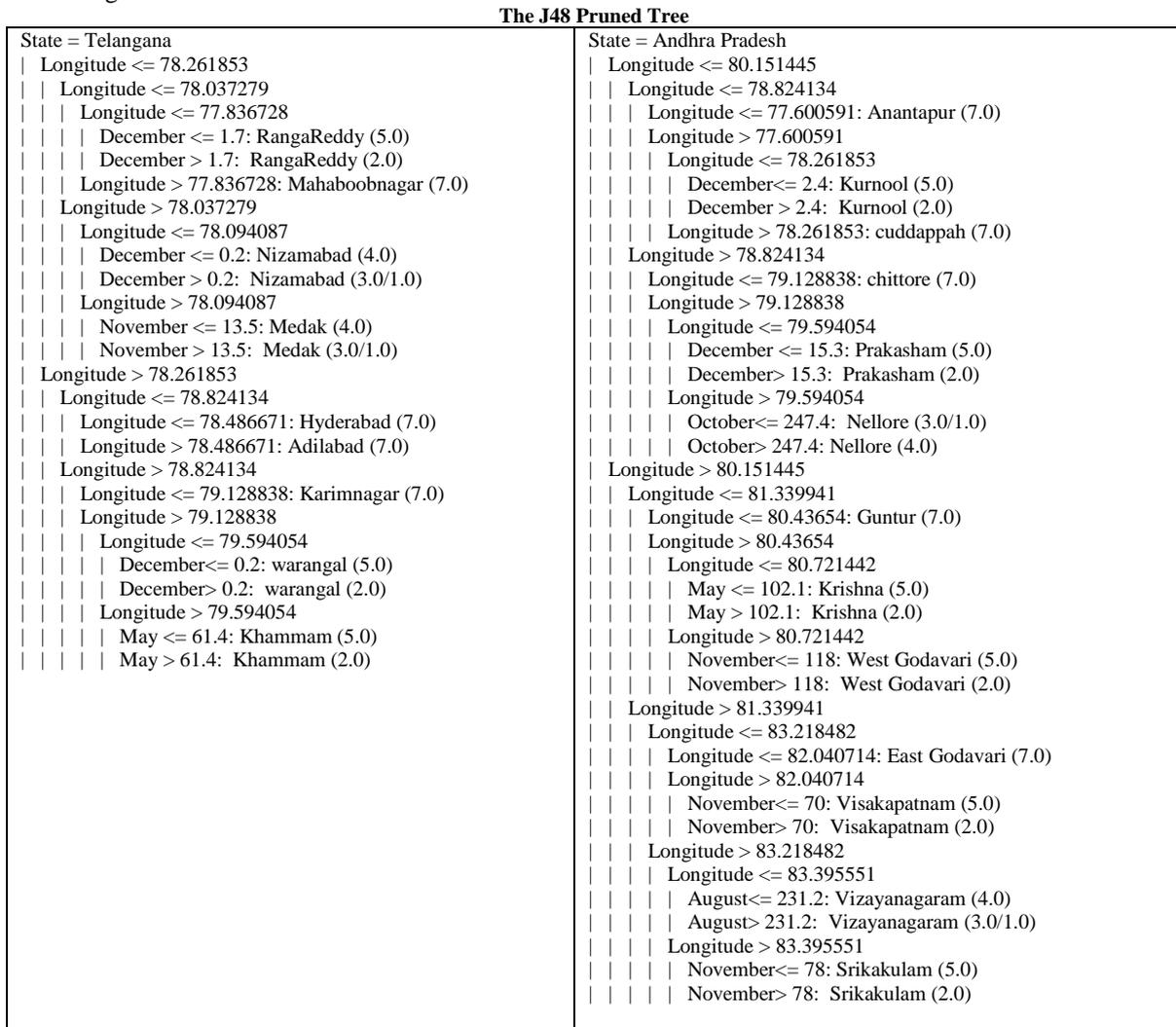


Figure 1. The Classifier Tree

Association Rules : Predictive Apriori

Best rules found:

<ol style="list-style-type: none"> 1. May=0,N=0, D=4 ==>Adlabad, State=Telangana 2. Jul=170.3,D=1.7 ==>Krishna, State=Andhra Pradesh 3. N=8.7, D=0.3 ==>East Godavari, State=Andhra Pradesh 4. May=2 ==>Nizamabad, State=Telangana 5. May=21 ==>Medak, State=Telangana 6. May=40.6 ==>Prakasam, State=Andhra Pradesh 7. May=45.5, D=0 ==>Srikakulam, State=Andhra Pradesh 	<ol style="list-style-type: none"> 8. May=63.3 ==>Guntur, State=Andhra Pradesh 9. May=67.5, Jul=24.7, N=3.9 ==>Kurnool, State=Andhra Pradesh 10. Jul=41.6 ==>Mahaboobnagar, State=Telangana 11. A=225.9 ==>Hyderabad, State=Telangana 12. O=36, N=33.6 ==>Khammam, State=Telangana 13. O=74.9 ==>Cuddapa, State=Andhra Pradesh 14. N=12.4, D=0, D=1, D=3.3 ==>Anantapur, State=Andhra Pradesh
---	---

15. Longitude=77.836728, Latitude=17.389138, Longitude=78.008122, Latitude=16.737509 Longitude=78.094087, Latitude=18.672505, Longitude=78.261853, Latitude=18.052936, Longitude=78.486671, Latitude=17.385044, Longitude=78.532011, Latitude=19.664062, Longitude=79.594054, Latitude=17.968901, Longitude=79.128838, Latitude=18.438555, Longitude=80.151445, Latitude=17.247253 ==> State=Telangana

17. Longitude=78.037279, Latitude=15.828126, Longitude=78.824134, Latitude=14.467354, Longitude=79.100329, Latitude=13.217176, Longitude=79.560344, Latitude=15.348463, Longitude=79.986456, Latitude=14.442599, Longitude=80.43654, Latitude=16.306652, Longitude=80.721442, Latitude=16.609991, Longitude=81.339941, Latitude=16.917418, Longitude=82.040714, Latitude=17.32125, Longitude=83.218482, Latitude=17.686816, Longitude=83.395551, Latitude=18.106658, Longitude=83.896781, Latitude=18.296975, Latitude=13.217176, Longitude=79.100329, Latitude=14.442599, Longitude=79.986456, Latitude=14.681888, Longitude=77.600591, ==> State=Andhra Pradesh

Figure 2. Association rules generated using Predictive Apriori Algorithm

The Latitude and longitude values corresponding to the States Andhra Pradesh and Telangana are associated using Predictive Apriori Algorithm and in the same way the Rain fall values corresponding to the Districts were also perfectly associated by this Predictive Apriori Algorithm. For example May=0,N=0, D=4 values are associated with the district Adilabad, State=Telangana and May=67.5, Jul=24.7, N=3.9 values are associated with the district Kurnool, State=Andhra Pradesh. The Annual rain fall for the various years from 2004 to 2010 is shown in Figure 3. In this graph, the Districts were taken along X-axis and the Annual Rain fall was taken along Y-axis. From the Graph, it is inferred that in the year 2005 Karimnagar has received high rainfall followed by Nellore, Chittore and Visakapatnam. In the year 2006 Nellore received high annual Rainfall followed by Vizayanagaram, Srikakulum and Visakapatnam. In the year 2010 East Godavari received highest rainfall followed by West Godavari, Krishna and Khammam. The Figure 4 shows map [17] of the Rainfall data of Andhra Pradesh along with Telangana States across districts and years. In Figure 5, the Rainfall data of 2009 across the districts of Andhra Pradesh and Telangana were shown in constructed customized map – Prakasam district rain fall data is shown in Pop up window. In Figure6, the Rainfall data of 2009 across districts of Andhra Pradesh and Telangana were shown in constructed customized map – west Godavari district rainfall shown in pop up window. In Figure 7, the Rainfall data of 2009 across districts of Andhra Pradesh and Telangana were shown in constructed customized map with Annual rain fall data values. In Figure 8, the Rainfall data of 2010 across districts of Andhra Pradesh and Telangana were shown in constructed customized map with Annual Rainfall data values. In Figure 9, the Rainfall data of 2004 across districts of Andhra Pradesh and Telangana were shown in constructed customized map with Annual Rainfall Data Values. In Figure 10, the Rainfall data of 2005 across districts of Andhra Pradesh and Telangana were shown in constructed customized map with Annual Rainfall data values.

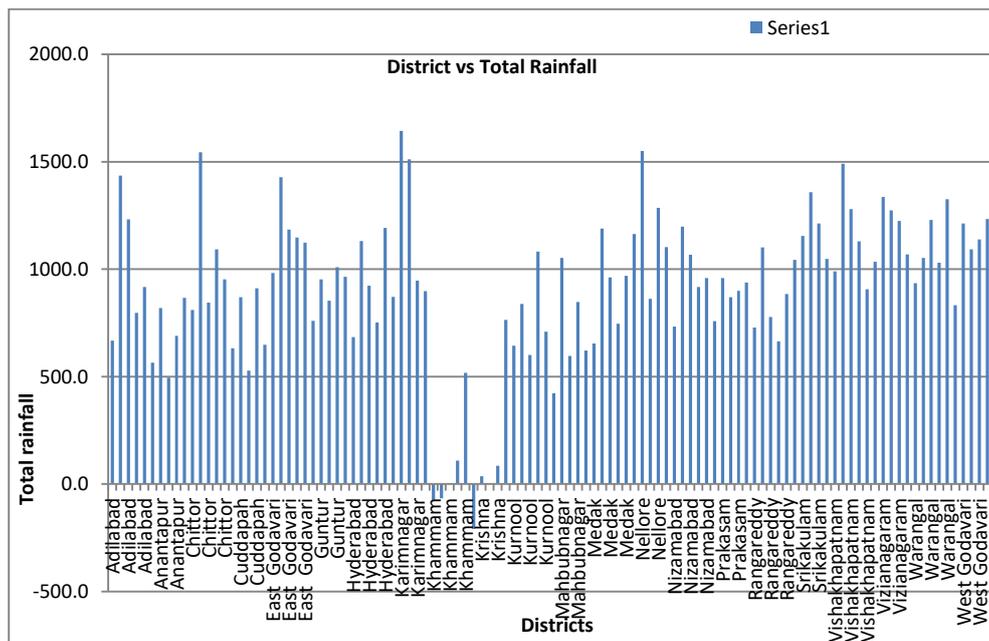


Figure 3. Districts Vs years Total Rain fall Data

6. CONCLUSION

The Annual Rainfall Data set was refined with Pre-processing Techniques and tested with Hybrid Data Mining Techniques namely Classification i.e. Decision Tree Classifier Algorithm namely J48. Apriori algorithm was applied on the resultant data set for finding Association Rules. The results of Latitude and Longitude values are perfectly associated with the respective states namely Andhra Pradesh and Telangana. The Highest Rainfall has taken in the year 2005 across the districts. The Maps shows, the spread of Rainfall data across districts of Andhra Pradesh and Telangana. From these maps it is evident that the coastal area districts namely Nellore, Visakapatnam, Vizayanagaram, East Godavari and West Godavari etc. received

more rainfall than compared to other Areas of the states of Andhra Pradesh and Telangana. There is an exceptional case with karimnagarin 2005 year.

7. FIGURES AND TABLES

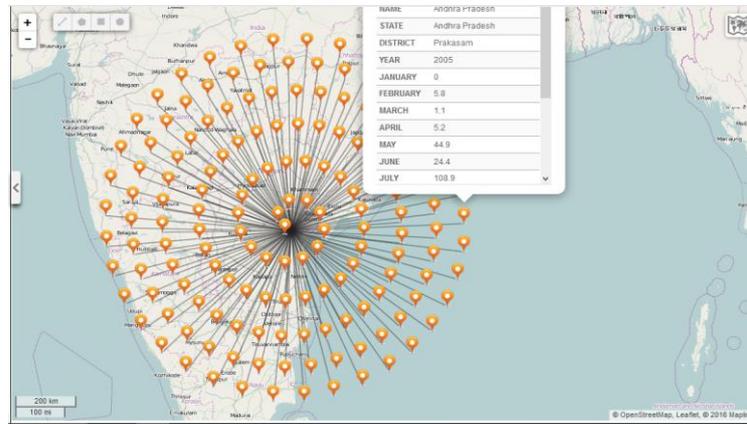


Figure 4. The Rainfall data of Andhra Pradesh along with Telangana across districts and across years were shown in the Map

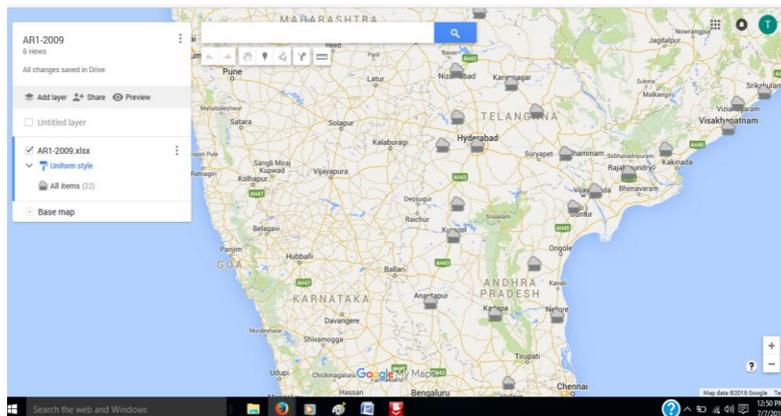


Figure 5. The Rainfall data of 2009 across districts of Andhra Pradesh and Telangana were shown in constructed customized map – Prakasham district rain fall data is shown in Pop up window

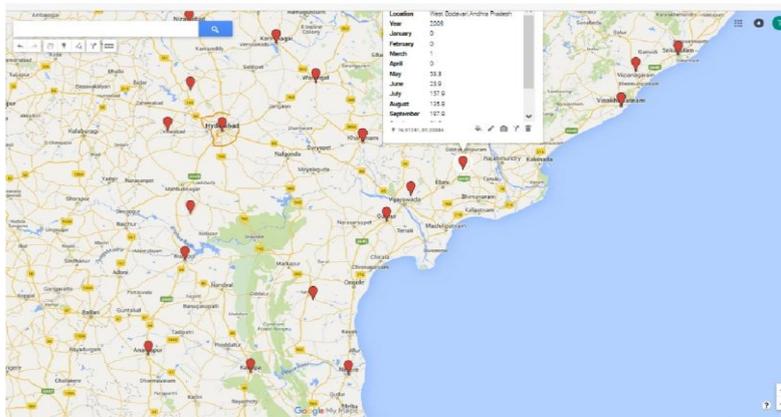


Figure 6. The Rainfall data of 2009 across districts of Andhra Pradesh and Telangana were shown in constructed customized map – west Godavari district rainfall shown in pop up window

- [17] Andy South, R worldmap: A New R package for Mapping Global Data
[18] Doantam Phan, Ling Xiao, Ron Yeh, Pat Hanrahan, and Terry Winograd, Flow Map Layout

BIOGRAPHIES OF AUTHORS



Mr. Y. Jeevan Nagendra Kumar, obtained his M.Tech Computer Science Technology from Andhra University in 2005. He is pursuing Ph.D from Acharya Nagarjuna University. He is working as Associate Professor in GRIET since 2005. He has about 6 Research Papers in International / National Conferences and Journals and also attended many FDP Programs to enhance his knowledge. With his technical knowledge he guided the students in developing the useful Web applications and data mining related products. As B O S member was able to introduce new subjects, topics in UG / PG Courses. Students are encouraged to work on research projects, engineering projects as well as for industrial training.



Dr. T.V. Rajini Kanth has obtained his Ph.D. degree in C.S.E. branch from Osmania university, Hyderabad in July, 2008 and M. Tech.(C.S.E.) degree from Osmania University, Hyderabad in January, 2001. His specialization area in research is "Spatial Data mining". He obtained his PGDCS degree from HCU, Hyderabad in 1996. He received his M. Sc. (Applied mathematics) degree in the year 1989 from S.V. University, Tirupati as University Ranker. He is currently working as professor in CSE, at SNIST, Hyderabad. He worked as a Professor and Head, Department of IT, GRIET, Hyderabad from 2008 June to Oct 2013. Before that he worked as a Professor in CSE department, GRIET, Hyderabad since 2007 November. Prior to that, he worked as Asso. Prof. in VNRVJIET, Hyderabad where he joined in 1996. His total teaching experience is 23 years. His writings have appeared in numerous Professional conferences and Journals (International journals-29, national Journals-3, International conferences- 23, national conferences-1, Total =56). Under his guide ship 5 Ph.D. Research Scholars got awarded (2-JNTUH & 3-ANU). He has received seminar grants from AICTE and DST organizations. Reviewer for many International Conferences like ICACM-2011, ICACM-2013, ICRSKT, ITQM, ICCCT etc., He has also received UGC Project Grant as a Principle Investigator and AICTE project grant as Co-Principle Investigator. He is a Review Committee Member \ Editorial Board member for 18 International Journals namely IJAEGT, IJAC, JDECS, IJEAT, IJAENT, etc., He was an author for few books i.e Artificial Intelligence etc. His current research area interests include Image processing, Data Warehousing & Mining, Spatial data mining, web mining, Text mining and Robotic area etc. Presently guiding research students in the research areas like Spatial data mining, Web mining, image Processing and Text mining. He has conducted two International conferences namely ICACM-11, and ICACM-13 at GRIET, Hyderabad as Convener and also acted as session chair for many conferences like ICACT-08, ICRSKT-2014 etc. He is presently guiding (supervisor & co-supervisor level) many Ph.D. scholars at various universities namely JNTUH, JNTUK, JNTUA and ANU. He was called for around 55 AICTE sponsored \ TEQIP workshops as resource person. He is Life Member in ISTE, CSI and a member in IEEE.