Prediction of rainfall through Data Science using Time Series analysis and Prediction Analysis

V.Rahamathulla[1], S.Ramakrishna[2]
[1]Research Scholar, PP comp Sci-209, Computer Science Department, Rayalaseema University, Kurnool, Andhra Pradesh.
[2]Professor, CCMIS, S V University, Tiruapti, Andhra Pradesh

Abstract: *In the present days it becomes a tedious scarcity of water, due to increase of population. The heavy usage of water leads to decrement of natural resources such as water. The huge increment of pollution is also effects the natural resources. The water is considered as an important natural resource. The main source of water for all the areas is rainfall. The rainfall is purely depends on natural climatic conditions [1]. The geographical objectives play a main role for the rainfall. One of the most tedious tasks is prediction of rainfall. The rainfall is a natural water resource and linkup with many geographical objectives such as temperature, wind direction, wind speed, humidity, and cloud coverage [2]. The prediction of rainfall contains a lot of process such as collection of huge datasets, processing the data, cleaning the data sets, identifying the accurate objectives, implementing various techniques such as machine learning, Bigdata tools, for processing huge collection of data sets.*

Introduction:

The Indian government is mostly concentrate on protecting the natural resources such as water, rosewood, sandal wood, minerals and raw materials. One of the most precious natural climatic resources is water. It is the responsibility of each Indian citizen to preserve water for future. It is the prime duty of every citizen to preserve the water, because the water is completely depends on rainfall. The rainfall is very occurrence event. The rayalaseema region is most draught hit region in the terms of rainfall. So people were facing a lot of problems due to water scarcity. So it motivates me to do research on natural climatic conditions such as rainfall prediction.

The research motivation:

According to the geographical objectives, the Indian country recognizes very less rainfall. The average rainfall in the entire country is 20cm to 40 cm [3]. The Andhra Pradesh is purely depends on rainfall for water. The Andhra Pradesh is having fertile lands, but only the scarcity is water. As the Andhra Pradesh are not having any perennial rivers. Most of the people of Andhra Pradesh are majorly depends on agriculture. The agriculture needs water. So it becomes very essential to predict the rainfall. The rainfall depends on various factors such as

- Geographical location of the region,
- Temperature
- The wind speed related to the geographical location.
- The wind direction related to climatic conditions
- Humidity
- Cloud coverage

Research Motivation

The weather parameters are highly inconsistent [4]. It is a combination of many more meteorological factors. The weather parameters are highly required input source for the rainfall. The role of parameters are highly inconstant for rainfall prediction, so rainfall becomes a tedious task and forced me to develop a relevant approach that can identify and recognize the weather parameters pattern. The identified data models can be used for rainfall prediction with a very least error factor and feasible with the help of recent techniques

- ➢ Artificial Neural Networks,
- ➢ Data Science
- ➢ Big data, instead of traditional methods such as
- ➢ Probability distribution
- ➢ Curve fitting
- ➢ Normalization
- ➢ Back ground study.

The objective of this research is

This proposed research helps the farmer for agriculture to decide the crop.
Helping the water shed department for water preservation.
Our research helps to analyze the ground water level.
Analyze the reports and helps for future predictions.

Back ground study:

- Allen and Vermon defined Objective forecasting (only one forecast), implemented in 1951 [5].
- Glahn et. Al. implemented the statistical approach for weather forecasting in 1972 [6].
- Ozelkan and Duckstein developed a regression analysis model in 1996 [7].
- Liu and Chandrasekhar developed a fuzzy logic in 2000 [8].
- Bae et al. Implemented ANFI (Adaptive Neuro fuzzy interference) for prediction in 2007 [9].

The traditional models produce more error factor, so it is necessary to develop a new model for rainfall forecasting. The soft computing techniques are capable to process huge amount of data, accurate prediction results, very less error factor.

Research Implementation:

Step 1:
Collection of data: This research was carried out the 53510 datasets. The data was collected from various rain gauge stations in and around of Tiruapti. The past data was also downloaded from the web such as data.gov.in, iim-pune etc.

Step 2:
Data cleaning: The raw data contains many irrelevant factors, Machine Learning techniques was implemented, to clean the data. We have separated the relevant data and irrelevant data into two parts.

Step 3:
Data Conversion: The raw data contains various formats such as
- Un Structure
- Semi Structured
- Structured

The data refining techniques are used to convert the data into structured, from unstructured, semi structured.

Step 4:
Rainfall prediction process: The objectives of this research are to predict the rainfall with an interval time slots.
- Hourly Rainfall
- Weekly Rainfall
- By Monthly
- Monthly

Step 5:

Implementing Time Series:
The time series analysis and Prediction analysis of Data Science techniques are used to predict the rainfall. The prediction process produced the results up to 96% accurate.

Step 6:
Results Analysis: The predicted results are compared with the actual results and identified that more accuracy in the results. The results which are predicted by this process are accurate to 96%.

Step 7:

Pictorial Representation of Data and Results:
The relationship between the parameters and the difference in the weather parameters is represented through graphs.

About time series Analysis:
The weather forecasting data contains various formats called as unstructured data. So analysis of unstructured data is not possible with the traditional methods such as statistical analysis. The time series provide the best solution for these types of problems [10].

The solution process considers a set of series data points as an index. The listed graphical values are also taken as the input, so the input may be a collection of different values.

Time series analysis is a collection of various methods that can be used for analyzing series of data. The user can build a model for prediction. Such types of models help the user to predict the future. The future corresponding consequences may be predicted with the help of time series. The input may be the observed values in the past. The input time series may not be continuous, it may interrupt due to time [11]. Time series data can be applied for real time analysis, the data may not be continuous it may be
- Interrupt data
- Capable to accept different data items
- Continuity of data not required.
- The different symbols may be supported
- The combination of numeric, semi numeric, characters, strings and collection of the given data types will be processed.
- Past values to future values prediction.

Time series analysis also supports data visualization. The visualization supports GUI (Graphical User Interface) based data Representation. Various forms of data may be represented as
- Graphs
- Charts
- Gap Chart
- Graphs with slopes
- Line charts with reduced data
- Circular graphs etc.

About the predictive analysis:
Extracting information is the key method of predictive analysis. The extracted data is used to predict the current trends and the system behavior models, patterns. The predictive analysis can be implemented for the known and unknown type of models. It supports present, past and future extractions [12].

Example Crime Analysis, credit card frauds, Seibel score and organization performance throughout the year.

The main principle of predictive analytics is capturing the relationships between important variables and predicting variables from the past experiences. Based on the previous history we can explore the variables. The quality of implementation and assumptions leads to accuracy of results [13].

Figure 1: The process of Predictive Analytics

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 7, Issue 1, January - February 2018   **ISSN 2278-6856**

1. Defining a Project: The user need to define the project scope, outcomes, expectations and identifying the data sets.

2. Collection of Data: Collection of data by implementing various methods, identifying the useful information, defining the complete view of customer.

3. Analysis of Data: The major role of predictive analytics is analysis of data; it contains data analysis, data cleaning, and data modeling with the objective of identifying the required data.

4. Implementing Statistical Analysis: Building the hypothetical assumptions, testing and defining the final user outcomes.

5. Building Data Models: The predictive analysis supports to build accurate predictive models, the user can choose one of the best among the constructed data models.

6. Machine Deployment: The built-up model may be deploying on the analytical results in our daily life. The process may be implemented to get results, reports and outputs by implementing the model.

7. Inspecting the Model: It is essential to have periodical observations on the defined model. The model should be inspecting to ensure the results.
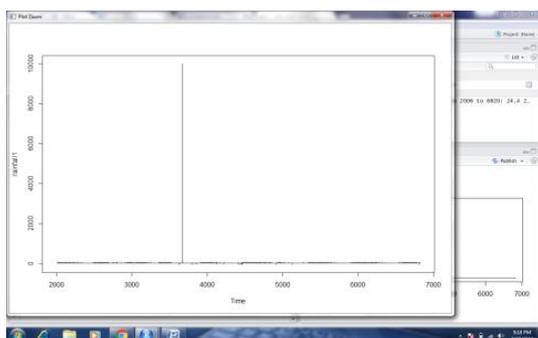
Implementation & Results
The experiment was carried out with the huge datasets and implemented based on time. Specific intervals have been measured and implemented, the interval time considered as
1. 4 Hours Frequency Duration with every 10 minutes of observations.
2. 6 Hours Frequency Duration with every 10 minutes of observations.
3. 12 Hours Frequency Duration with every 10 minutes of observations.
4. 24 Hours Frequency Duration with every 10 minutes of observations.
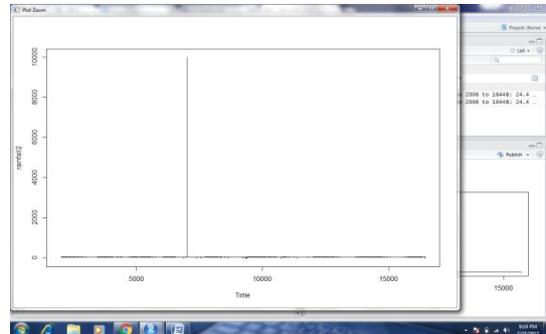
The results as follows:
1. The 4 Hours Frequency Duration with every minute of observations was implemented to derive the short term rainfall. The accuracy of results is 98%.
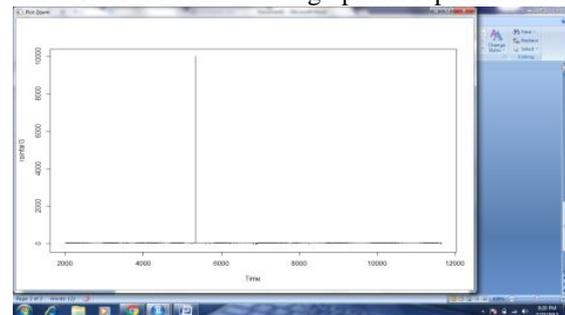The results are defined in a graphical representation.

Figure 2: 4 Hours Frequency Duration

2. 6 Hours Frequency Duration with every 10 minutes of observations was implemented to derive the quarter of a day rainfall. The accuracy of results is 97%.
The results are defined in a graphical representation.

Figure 3: 6 Hours Frequency Duration

3. Step3. 12 Hours Frequency Duration with every 10 minutes of observations was implemented to derive the quarter of a day rainfall. The accuracy of results is 95%.
The results are defined in a graphical representation.

Figure 4: 12 Hours Frequency Duration

Step 4: 24 Hours Frequency Duration with every 10 minutes of observations was implemented to derive the quarter of a day rainfall. The accuracy of results is 96%.The results are defined in a graphical representation.
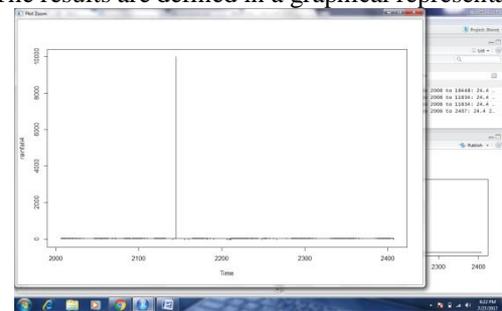
Figure 5: 24 Hours Frequency Duration

Conclusion: Now a day, it becomes a very tedious task to estimate the climate. Due to pollution and other unwanted chemicals causes the effect of climate. The change of climate is abuse. The change of climate directly affects the daily life of a human being. Due to this reason it becomes highly complex for the social survive of human beings, all other mammal on the earth. The Data Science and Big Data

are the much more suitable tools for the prediction of rainfall. The rainfall prediction helps the people in all aspects. The major problem of water scarcity may be prevented with the help of rainfall prediction. The agriculturist may overcome the problem of heavy loss due to natural phenomena. The implemented technologies provide much more accuracy in the results. The approximate error free solution can be build for the rainfall prediction.

References:
1. David Johan, jean Palutikof, Sarah Boulter, "Natural disasters and adoption to Climate Change", Cambridge University Press, Online Publication data: October 2013, ISBN: 9780511845710.
2. Rahamathulla Vempalli, Ramakrishna s, Latheefa Vempalli, "Day wise Rainfall Prediction System Using Artificial Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, issue 3, March 2014, ISSN 2277 128X , Page No: 858 to 866.
3. Indian Metrological Department, pune, www.impdpune.gov.in, news letter, January 2017.
4. http://www.indiawaterportal.org/articles/measurement-weather-parameters-data-collection-and-analysis-presentation-acwadam.
5. Shaminder singh, Baljeet singh, shaminder singh, "Training Back Propagation Neural Networks with genetic algorithm for weather forecasting" IEEE 8th International Symposium on intelligent systems and informatics sebia, volume 1, ssubotica, seriba-2010, page number 463 to 469.
6. Saeed Reza Khodashenas et al, Najmeh Khalili "Daily Rainfall Forecasting for Mashhad Synoptic Station using Artificial Neural networks", https://www.researchgate.net/publication/228450101, International Conference on Environmental and Computer Science IPCBEE , 2011. VOL.19 (2011), page numbers 124 to 128.
7. Březková L, Šálek M, Soukalová EC (2010) Predictability of flood events in view of current meteorology and hydrology in the conditions of the Czech Republic. Soil Water Res 2 (4):154–166.
8. Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach. Sanjay D. Sawaitul1, Prof. K. P. Wagh2, Dr. P. N. Chatur3, www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 1, January 2012)
9. Satellite Rainfall Applications for Surface Hydrology Author: Mekonnen (EDT) /ain Gebremichael, Editor: Mekonnen gebremichael Faisal Hossain.
10. Gralel GloriMould, Hadley Wikham, "Hands On programming with R", Orelley publishers, Volume 1, Edition 1, 2011.
11. James MCCafree, "R Programming", Syncfusion, volume1, Edition 1, 2014.
12. Radian Johnson, "Fundamentals of R programming Language and Statistical Analysis", Packtub publishers, volume 1, Edition 1, 2015.
13. Kuan Chingli, Hai Jiang, T. Yang "Big Data Analytics, Algorithms, and Applications", CRC Press, Volume 1, Edition 1, 2015.

Authors:

Mr V. Rahamathulla is a research scholar, having 6+ years of research experience. The research domains are artificial neural networks, Bigdata analysis. The author published 5 international journal papers and also published 6 national papers.