



IMPLEMENTING DATA SCIENCE FOR RAINFALL PREDICTION WITH VARIABLE PARAMETERS THROUGH MACHINE LEARNING AND ADVANCED BIG DATA TOOLS.

V.Rahamathulla
PP Comp Sci-209
Computer Science Department
Rayalaseema University,
Kurnool,India

Prof. S. Ramakrishna
CCMIS,
S.V.University
Tirupati,India

Abstract: India is an agriculture based country. Most of the population is living in villages and their occupation is agriculture. As Mahathma Gandhiji said “The Development of the nation is depends on the development of the villages” [1]. Rainfall is most required factor for agriculture. The complete agriculture depends on Rainfall [2]. Due to abnormal conditions of the weather the formers are forgoing their crops and indulging as debtors. Most of the farming areas in the country are depends on the Rainfall, In India only 34% of the fertile land getting water through the rivers. So there is a huge necessity for Prediction of Rainfall. The Rainfall is one the most important Phenomena of the weather. The weather may be described as change-in-climate. Rapid changes occurred in the weather. It is the major task to understand and predict the weather.

The weather parameters are

- Temperature
- Humidity
- Wind speed
- Wind direction
- Rainfall
- Pressure

The major occupation of the Indian people is Agriculture. Rainfall plays a major role in agriculture. The prediction of rainfall helps the farmer to decide the type of a crop. The crops broadly categorized in to 2 types.

- 1) Short Term crops
- 2) Long Term crops

Short Term Crops: The short term crops required less water and the crop completes with in the short period. Eg: Leafs,

Long Term Crops: The long term crops required more water and the crop takes long period such as 3 months, 6 months, 1 year etc.

Most of the crops are depends on Rainfall. So the prediction of rainfall helps the farmers to decide the type of crop.

Rainfall Impact on Human Life

Day to day life of a human being contains a huge number of tasks. The human life is completely depends on weather. The water is one of the most required and necessary element for human life. Most of the human needs may be full filled by the water. The major resource for getting the water is rainfall. Another source of water is Ground water. The Ground water recharge is based on the rainfall. So rainfall plays a major role in all the aspects of the human life.

The rainfall prediction is a very complex task. In the early 80's there is no particular methodology for rainfall prediction [3]. People are facing lot of problems for making decisions about crops, preserving water and implementation of usage plans. The town planning should also be considering the rainfall prediction report to avoid the loss due to floods. The major causes of rainfall is Tsunami, Cyclone, Decrease of Temperature are the unexpected natural climatic conditions. So it is essential to have a predicted rainfall report with all the departments and general bodies. Based on our research one can take the decisions in advance.

The new research methodology implementing the rainfall forecasting through the data science may lead to produce the accurate results. The Data Science is a combination of various tools such as Machine Learning, Statistical Implementations, Neural Network Approach and other tools. The data science consider the huge data input values, process the input values and generates the reports, this approach produces the prediction result with very less error factor (approximately less than 10).

Research approach: This research was started by observing the weather parameters such as

- Temperature
- Humidity
- Cloud cover
- Wind speed
- Wind direction
- Rainfall
- Pressure

The weather is a complex phenomena and changes due to the geographical objectives. So it becomes a highly complex task for analyzing the weather. Now a day's most of the researchers started research by implementing various technologies for weather analysis.

Keywords: Weather parameters, Rainfall, Data Science, Machine Learning, Clustering.

INTRODUCTION

Most of the days we cannot get rainfall, the rainfall occurred very rare cases. In the weather phenomena the very rare occurrence is rainfall. It becomes a very challenging issue to predict the rainfall. The long range of statistical approach and integration techniques are required to predict the rainfall. This approach will not be providing maximum accuracy. The implementation contains lot of complex calculations and graphs are needed for prediction. Most of the times we can get failure results with this method. The traditional method accepts very less number of inputs and considers very little factors. Because of these reasons, it is essential to consider a lot of inputs with multi parameters. The increased, number of inputs and parameters lead to more accuracy in prediction. The data science is a new technology of organizing data with various patterns. The data science is a collection of various technologies that supports various methods for data implementation and execution. The implementation contains more than 100000 data tuples for prediction process. The final results are synchronized with 95%.

RESEARCH PLAN

In the early days people attempted for weather forecasting with the help of meteorological department. The study of meteorology defines the atmosphere state at future. The study, predict the atmosphere condition in an advance. There future prediction is possible for a short period. Weather forecasting became one of the most important practical experiments of meteorology. The rainfall prediction supports the farmers, water supply department, town planning commission and airport authorities. More than 200 ways have been defined for forecasting the weather with the known experiences by the Greek Philosopher. The book was written about 300 b.c.

Weather forecasting is a cross platform research environment. The scientific researcher needs to form a team with meteorology department to gather the weather parameters. The parameters should be collected at relative positions. The relative data collection system should be observe and the weather centers. The data should be collected based on time slots. The data should be collected by observing the parameters at relatively spread weather centers [4]. The first weather center was developed in the year 1930 by Vilhelm and Bjerknes. The network successfully collected the data. The observed transferred by the telegraph. In india central government was established a IMD (Indian Meteorology Department) for weather forecasting. The IMD Pune established a national regional center for weather forecasting “National Climate Center “in the year of 1995.About Data ScienceIn the present days the data is generating with huge growth rate. It becomes tedious task to maintain the data or organize the data [5]. The data contents are huge volumes, millions and billions of data is generating due to maximum usage,

- what’s app
 - Face book
 - Twitter
-
- Attachments through mails

- Data forwarding etc

Due to good communication a single data file is rotating with much number of users. The data is generating in huge volumes. Every transaction records bytes of data; it becomes a huge task for the data organizers to maintain the huge volumes of data.

The researcher started the research to find out contemporary methods to organize the data. The first data analysis came into the picture (paper 8) in the year 1962 by the new term ‘bit’. The further implementation was started in the year 1974. The major data analysis research was undertaken in the year 1989 with the implementation of Knowledge Discovery Databases.

DATA SCIENCE ARCHITECTURE:

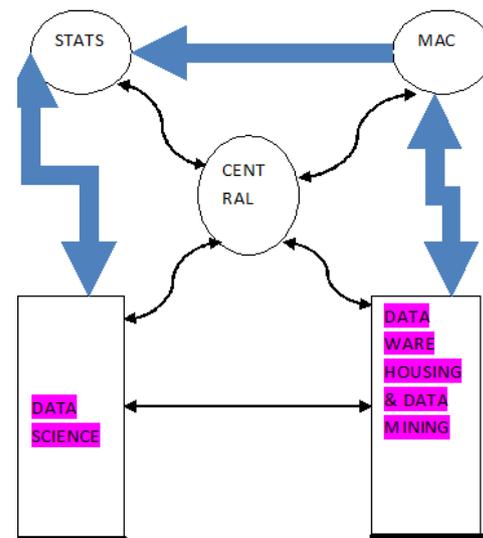


Fig: Data Science Architecture

Commercial usage of Data Science

Data science is a very powerful and secret tool for commercial. The Data Science applications may be implemented in any of the areas, such as Business, Engineering, Medical, construction and all other fields. The Data Science helps to solve many more problems in the commercial applications. The Data Science contains rich set of tools, that can be implemented for getting the solutions of many number of real world problems.

Implementation of Data Science for Rainfall Prediction:

Prediction is a plan of action to derive the results about an uncertain event. The prediction may also be derived based on the activity. The activity should be conducted on a specific set of conditions.

The prediction process includes collection of large amounts of data, alignment of large amounts of data and analyzing the data in order to identify patterns in the data. The identified patterns may be implemented to build the predictive models.

The rainfall prediction is a rarely occurred event. It needs specific observations. The experimental model collects the data in the form of

- Unstructured data
- Semi structured data
- Structured data

The data contains huge volumes of information. The prediction process should identify the patterns based on conditions [6]. The Data Science provides a specific support for Data analysis and prediction of rainfall with maximum accuracy.

Machine Learning

Machine Learning is a combination of different tools and contains various stages of implementation. The main purpose of machine learning is to train the network so that it inculcates the habit of learning. Machine learning is a combination of following steps

- Collection of data
 - Unstructured data
 - Semi structured data
 - Structured data
- Data Splitting
- Implementing supervised learning
- Implementing Unsupervised learning
- Applying Linear Regression

EXPERIMENT SETUP

The weather parameters are observed by the help of various weather monitoring devices such as

Weather parameter	Measuring Device
Temperature	Digital Thermometer & Temperature sensors
Humidity	Hygrometer
Wind speed	Anemometer
Wind Direction	Wind wane
Cloud cover	Ceilometer
Rainfall	Rainguage Station

Images of Weather Parameters measurement devices[7]:



DATA COLLECTION FOR EXPERIMENT

The data was collected from 40 local weather stations and downloaded from IMD Pune website and Data.gov Indian website.

The experiment was carried out with more than 100000 data tuples. The experiment was executed successfully on the High Power Computing and acquired results.

The Experiment carried out with the implementation of

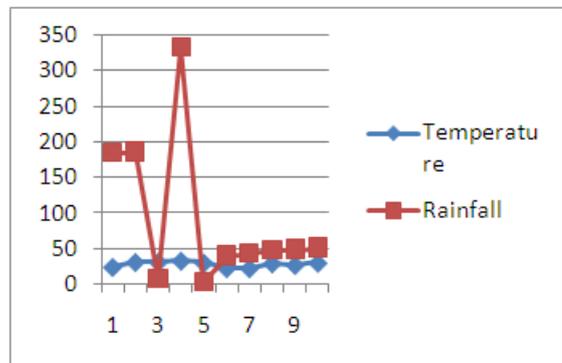
- Time Series
- Prediction Series

The table displays the results. It contains weather parameters, predicted rainfall.

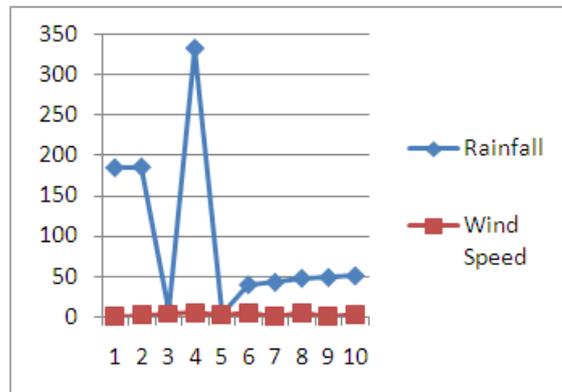
S No	Temperature	Wind Speed	Wind Direction	Pressure	Humidity	Ranfall (predicted)
1	24.51	0.54	225.07	960.04	75.02	185m
2	31.3	1.52	255.87	961.51	50	186
3	31.3	3.37	158.11	969.13	44.04	8
4	33.11	4.55	132.21	968.26	37	333.11
5	30.38	1.52	54.99	961.02	59.97	3
6	23	4.45	322.83	957.89	99.95	40
7	22.8	0.05	90.18	959.65	99.95	43
8	29.5	4.15	331.13	968.35	63.98	48
9	27.39	0.54	298.88	959.46	77.96	49
10	30.28	2.49	213.83	959.85	50	51

Graphical Representation of weather phenomena [8].

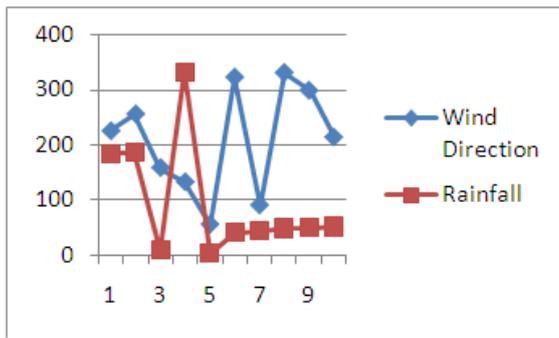
1. Graphical Representation of Temperature and Rainfall.



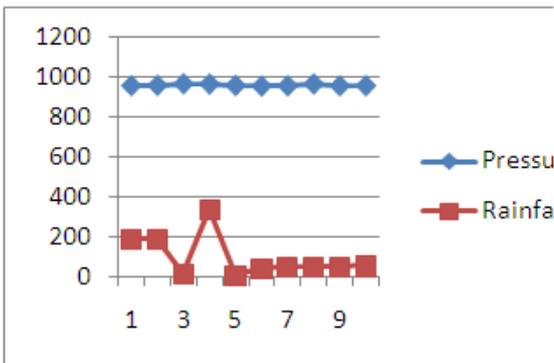
2. Graphical Representation of windspeed and Rainfall.



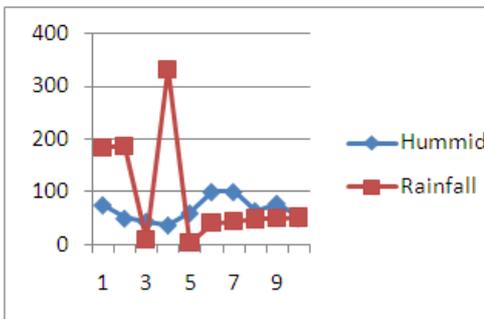
3. Graphical Representation of Wind direction and Rainfall.



4. Graphical Representation of Pressure and Rainfall.



5. Graphical Representation of Humidity and Rainfall.



CONCLUSION

The rainfall is depends on various geographical parameters, the experiment carried out with the various devices to collect the real time weather parameter data [9]. The huge

set of data sets will provide accurate results with less error factor. The implementation of data science supports to accept more than one lakh tuples. Rainfall was predicted as hourly, daily, weekly, fortnight, and monthly. The results was compared with the actual results and conclude that 95% of accuracy in the results. The data science is upcoming and powerful tool for prediction analysis[10].

REFERENCES

- [1]. Gandhi Mahatma, Mahadevan, Desai, "Gandhi, An Auto Biography" Edition1, Volume1, 1953, Beacon Press, Page numbers 6 to 10.
- [2]. 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 sebiba, volume 1, ssubotica, seriba-2010, page number 465 to 469.
- [3]. 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 128 to 133
- [4]. Zaman M. Faisal, H.Hirose, Sumi S. Monira " Comparision of Artificially Intelligent Methods in Short Term Rainfall Forecast" the 13th International Conference on Computer And Information Technology, Vol. PID728, 2010 page number 106 to 109.
- [5]. Foster provost, Tom Fawcett, " Data Science for Business" Edition 1, volume 1, 2016, O Relly Media, page number 105.
- [6]. Rachel Schutt, "Doing Data Science" Orelly Meida, Edition 1, Volume 1, 2013, page number 203
- [7]. Peter A Flach "Machine Learning", volume 10994, issue 1, 2018, springer Publications, page numbers 1573-0565 (online).
- [8]. Pranjali Awasthi, Maria Florina Balcan, Konstantin Voevodski " Local algorithms for interactive clustering" volume 107 Feb-2018, Journal of Machine Learning Research" page number 481.
- [9]. Gowthami, "The Prediction of Indian Monsoon Rainfall-A Regression Approach", volume-6, issue -2, January-2007, Solstice-An Electronic Journal of Geography and mathematics, page number 206.
- [10]. Dr. Vincent Granville, "Developing Analytic Talent", volume 1, Edition 1, , August 2015, Wiley publications, pagenumber 306.