Rainfall Prediction Using Data Visualisation Techniques

Yogesh Kumar Joshi
Department of CSE
Amity University Uttar Pradesh,
Noida, UP, India
Email: yogesh22217@gmail.com

Udit Chawla
Department of CSE
Amity University Uttar Pradesh,
Noida, UP, India
Email: udchawla02@gmail.com

Shipra Shukla
Department of CSE
Amity University Uttar Pradesh,
Noida, UP, India
Email: ershiprashukla88@gmail.com

***Abstract*—** The volume of big data has opened up great opportunities for prediction and analysis of different aspects of weather. Data Visualisation is common in day to day life. Various charts and graphs are used to illustrate the practical approach towards the classification of rainfall with the help of data visualisation methods. Since it was impossible to analyze the large datasets earlier, the data visualisation techniques has made easier to plot the graphs for the better understanding of the weather. With the help of data visualisation patterns such as the highest, lowest and average rainfall in the States/Union Territories the weather of India has been visualised. In this paper, the rainfall pattern in the States/Union Territories of India was successfully visualised. The pattern identifies drought prone region in India, decrease in the annual rainfall over the century and heavy rainfall in the coastal regions of India.

Keywords

Data Visualisation, Weather Analysis, Big Data, Data Analytics, Rainfall Pattern

I. INTRODUCTION

Data visualisation is the representation of the data in a graphical form. It converts the raw data present into a much more representable form, as a result this data can be communicated to the user more efficiently. Data Visualisation makes the data more understandable and user-friendly.

In the age of big data and data analytics the incomprehensible amount of data present can be put to a great use with the help of data visualisation. Data visualisation opens up an ocean of opportunity in field of big data and data analytics [1]. Data visualisation brings us a step closer to advance technologies such as Artificial Intelligence and Big Data.

Due to the complexity in analysing rainfall and the ability to learn from the past datasets Data visualisation plays a vital role in the analysis of the large amount of data and it could prove a great aid to Artificial Intelligence.

Data visualisation is divided into five categories. Temporal is the first category and it is used when the data is linear and one dimensional in nature. It uses lines which are either stand alone or overlapping and are having a start and finish time. These graphs are very basic and easy to understand.

Temporal data visualisation consists of line graph, scatter plot and time series sequences. Hierarchical is the second category of graphs. They are used when a cluster of information flowing from a single point is present. These graphs are usually very complex and difficult to understand with the exception of tree diagrams. Examples of hierarchical data visualisation are tree diagrams, ring chart, sunburst diagram. Third category is Networks. Networks are used when there is a relation person between datasets. Node-link diagram, matrix diagram and word cloud are the examples of network data visualisation. Multidimensional is the fourth category, as the name suggests multidimensional data visualisation have multiple dimensions. Examples of multidimensional data visualisations are Scatter plot, pie charts and Venn diagram. The fifth category is geospatial. Geospatial data visualisation overlays the real life physical location with different data points. For example flow map, density map and heat map.

The data can be easily used with the help of tools and libraries such as seaborn, numpy and pandas present in the Python language. These libraries consists of various tools which can be used convert the raw data into a more understandable form. In this paper bar and line graphs are used to represent the data. Visualising the data has great benefits. It gives a clearer view of thing by converting the raw data into a much more understandable and user-friendly one.

By visualising the data can make sense of large data sets. Visualising also encourage drawing out patterns trends and comparing various data. Drawing out these patterns has great use in Search Engine Optimisation (SEO). Visualising the data shows all the intricacies of perplexing subjects which is of great use in the field of Artificial Intelligence.

The datasets used in this paper is shared by the India Meteorological Department (IMD) Govt. of India under Govt. Open Data License - India. It contains the monthly rainfall detail of 36 meteorological sub-divisions of India. The datasets consists of monthly as well as annual rainfall of every state and union territory from 1901 to 2015[2].

This paper has been divided into five sections. Sections 2 provide us with the works which has already been done related to the project. The method used for the visualisation of the data is explained in the Section 3. Section 4 consists of the results obtained from the dataset. Section 5 concludes the paper.

978-1-7281-2791-0/20/$31.00 ©2020 IEEE

II. RELATED WORK

Many researchers had already worked on the rainfall prediction and their connection with climate change with different methods of machine learning. Several studies are being discussed in India to predict the climate and, researchers used different lengths of information and now studies have been reported using information over a long time. Various researches have explained that predicting and analysis of monthly rainfall was suggesting different algorithms and methods. Some of them are:

Tripathi et al [3] ,have discussed rainfall prediction with the use of Artificial neural network by the Cascade-forward Back propagation Neural Network (CFBPNN) technique.[4] In this paper, to study the data set of the rainfall information they used 2 hidden layers of CFBPNN technique, with three different epochs.[5] To measure the performance of the developed research, the mean square error (MSE) algorithm is employed using CFBPNN. The results have shown that architecture with epoch fix to 500 and learning rate 0.1, produced a good result. Eventually, CFBPNN algorithm has provided a best accuracy model to predict monthly rainfall.

Chakraborty S. et al [6] have discussed about the Forecasting of Weather with the help of Incremental K-means Clustering. In this paper, a general method is proposed for weather forecasting by the help of clustering algorithm. In incremental K-means the new data is grouped into clusters whose weather category has already been defined. By doing the algorithm makes a strategy to predict the weather.[7] This forecasting algorithm was successfully implemented for the weather of West Bengal. This forecasting method was developed to reduce the impact of air pollution and for predicting future weather events.

Different researches by Indians for the meteorological department, there are also researches that are proposed by foreigners.

Trombe P. et al [8] have discussed about the Automatic Classification of Offshore Wind Regimes with Weather Radar Observations. In this paper, the automatic classification of offshore and wind regimes (i.e., wind fluctuations with specific frequency and amplitude) with the help of reflectivity observations from single weather radar system has been addressed. Spatial continuity, motion of precipitation echoes on the images and global intensity are described from the above attributes. Finally, classification tree was used to find the relationship between wind regimes and precipitation attributes.[9]

Chawla G. et al [10] have discussed about the big data analysis which will help in getting a view of the given dataset . In this paper the researchers focuses on Big Data visualisation, its challenges, various tools. Researchers have explored a new way to visualise and analyse complex and dynamic datasets using virtual reality. They have also founded that how virtual reality has extremely changed the world of Big Data Visualisation.

Manoj et al[11] have discussed about the data visualisation in the data mining technique visualisation is important because of the changes in patterns, behaviour which helps to study the data set more carefully so, in this paper they discuss about the visualisation and the study of the dataset for rainfall[]. It is presented on analysing the time series dataset of Monthly Rainfall dataset of India. The visualisation was made possible with the help of the ARIMA Model and suitable model is developed for future forecasting of weather.[13]

Various Machine learning models can also help to predict the useful insights from the large datasets [14, 15].

III. METHODOLOGY

There are several steps involved in the visualisation of data. From the studying of the datasets to plotting it in graphs, each step taken for the visualisation of the data is explained in this section.

A. Analysing the Problem

Analysing the problem can be done by studying the data sets. By studying the datasets NaN values present in the datasets can be found. In this stage the type of visualisation which can be used for the dataset is selected. Studying the dataset gives a clear view and better understanding of the problem statement.

B. Cleaning the Datasets

Incorrect data leads to false and inaccurate conclusions. These inconsistencies in the datasets can be due to user error. To correct these inconsistencies data cleaning is used to identify and correcting(or removing) all the inaccurate records present in the dataset. Data cleaning can find out the inaccurate, incomplete, irrelevant and incorrect parts of the data and remove or replace them with median values. There are many tools available for data cleaning, in this paper pandas library is used for the cleaning of data. After removing these irregularities in the datasets, Data Preprocessing is done.

C. Data Preprocessing

Good quality of data needs to be accurate, precise, consistent and uniform. Data Processing is used to achieve the quality standards of a dataset. Data Preprocessing is used to convert the raw data into a more refined and useful one. Data preprocessing is very important as it improves the quality of the raw data and helps in achieving better results. In this report, data preprocessing includes the replacement of NaN values with median values and then grouping of data in ascending order as well as grouping of data with respect to their respective states. These changes increase the reliability of the result and improve the efficiency of the program.

D. Algorithm and Methods

The data present in the datasets is visualised in this stage. In this paper, bar graph is used to visualise the highs and lows of rainfall in each state and maximum and minimum rainfall of all the states is compared. Line graph is used for visualising the annual rainfall of each state from 1901-2015. Both bar and line graph, been very simple and familiar, is best suited for the dataset used.

Line graph are plotted using matplotlib.pyplot library and bar graphs are plotted using the seaborn library.

The data previously grouped in the third stage are plotted in axes and the axes and are labelled to incresae the readability and understandability of the graph. In line graph, legends are used to explain what each line represent.

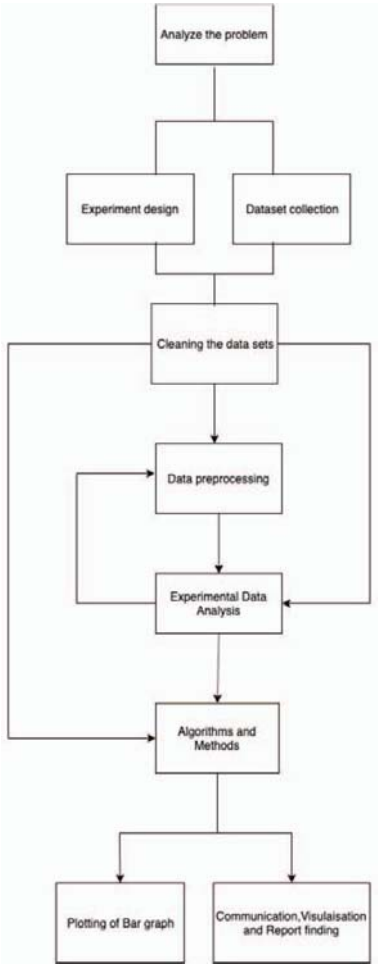

Figure 1: Flow Chart for Data Visualisation

IV. RESULT & DISCUSSIONS

A. Minimum Rainfall State/Union Territories

Minimum rainfall graph (Figure 2) shows the top ten States/Union Territories off India having the least rainfall in the period 1901 to 2015. The graph is plotted between rainfall in millimeter (x- axis) and names of States/Union Territories (y-axis).

As seen in figure 2, West Rajasthan has the least rainfall in India of a little over 20000mm in over a century followed by Saurashtra & Kutch and Haryana Delhi & Chandigarh. States/Union Territories in figure 2 has always been subject of scarcity of water and depleting underground water.

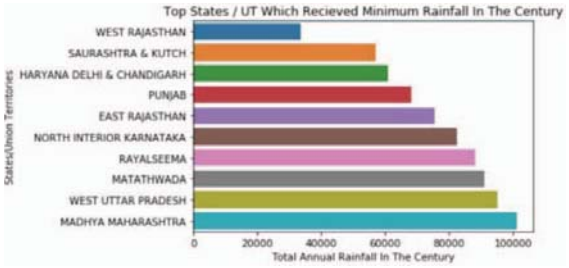

Figure 2: Minimum Rainfall State/Union Territories

B. Rainfall in West Rajasthan

Figure 3 shows the annual rainfall in the region of West Rajasthan. As it can be seen, rainfall in Rajasthan is very scarce. The average annual rainfall in West Rajasthan is around 300mm.

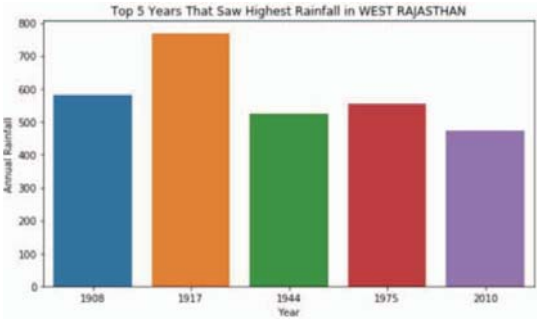

Figure3: Annual Rainfall In West Rajasthan

This graph also shows the drought prone areas in the states/Union Territories of India. As per the rainfall pattern seen by the India Meteorological Department, areas receiving less than 750mm of annual rainfall are considered as drought prone areas. West Rajasthan, having an average rainfall of 300m, comes under the chronic drought prone area.

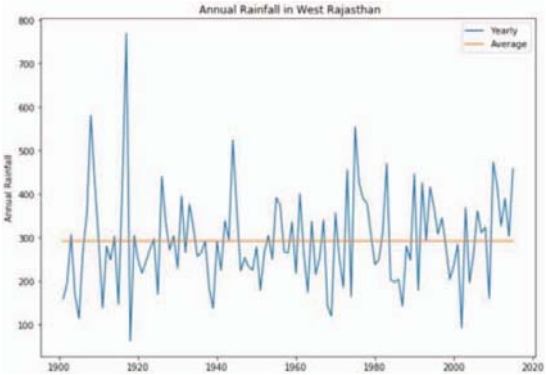

Figure 4: Highest Rainfall in West Rajasthan

Figure 4 shows the top five years which saw the highest annual rainfall in the region of West Rajasthan. The highest being around 700mm in the year 1971. The lowest being a little less than 60mm in the year 1918.

Figure 5 shows the top five years which saw the lowest annual rainfall in the region of West Rajasthan. The lowest being a little less than 60mm in the year 1918.

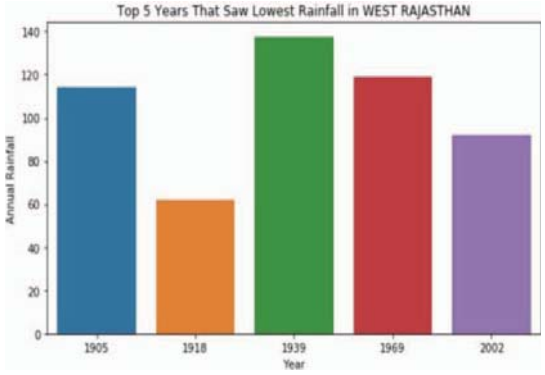

Figure 5: Lowest Rainfall in West Rajasthan

C. Maximum Rainfall State/Union Territories

Maximum rainfall graph (Figure 6) shows the top 10 States/Union Territories of India having the highest rainfall in the period 1901 to 2015. The graph is plotted between rainfall in millimetre (x-axis) and names of States/Union Territories (y-axis). As seen in figure 6, Costal Karnataka has the highest rainfall in India of about 400000mm in over a century followed by Konkan & Goa and Kerala.

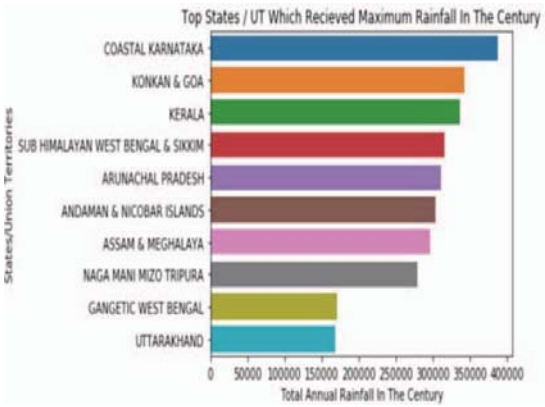

Figure 6: Maximum Rainfall State/Union Territories

D. Rainfall in Coastal Karnataka

Figure 7 shows the annual rainfall in the region of Coastal Karnataka. Coastal Karnataka has the highest rainfall in India. The average annual rainfall is around 3400mm. Despite of having the highest rainfall in the India, Karnataka comes second to Rajasthan in the extent of arid land.

As it can be seen, rainfall in Coastal Karnataka is very high and best suited for rainwater harvesting. As a result, Karnataka has implemented one of the largest rainwater harvesting project to tackle the existing problem of water scarcity.

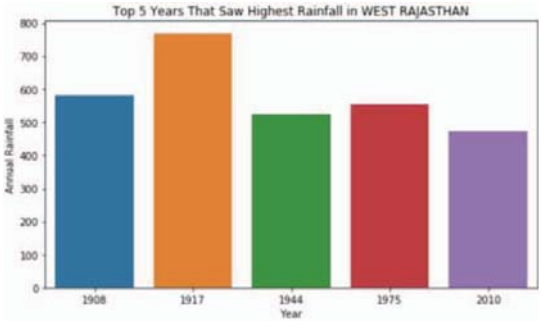

Figure 7: Annual Rainfall in Coastal Karnataka

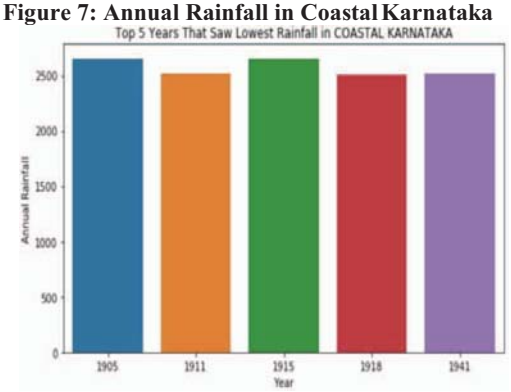

Figure 8: Lowest Rainfall in Coastal Karnataka

Figure 8 shows the top five years which saw the lowest annual rainfall in the region of Coastal Karnataka.The lowest being around 2500mm in the year 1915.

Figure 9 shows the top five years which saw the lowest annual rainfall in the region of Coastal Karnataka.The highest being over 5000mm in the year 1961.

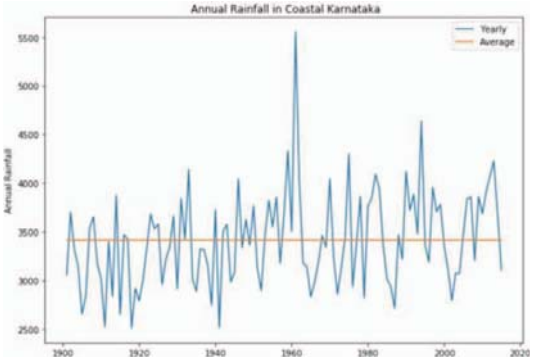

Figure 9: Highest Rainfall in Coastal Karnataka

V. CONCLUSIONS

In this paper, all the various trends in rainfall in the States/Union Territories over the past century has been shown with the help of data

visualisation using bar as well as line graphs. Top ten highest and lowest rainfall in States/Union Territories of India, Highest and Lowest rainfall in the region of West Rajasthan and Coastal Karnataka along with the annual rainfall has been shown in this paper. The regions having highest and lowest rainfall in India has been discussed.From the results of data visualisation an insignificant decreasing trend in annual average rainfall in the region of West Rajasthan and Coastal Karnataka has been seen. The chronic drought state of the region of West Rajasthan and the destructively heavy rainfall in the region of Coastal Karnataka is talked about. Visualising the data has greatly simplified the data and improved the understandability. Future studies can be carried out to establish other characteristics of rainfall. The results obtained in this paper can be used for the prediction of rainfall with the help of regression which can be of great benefit in the field of agriculture.

REFERENCES

[1]. Compos J., Sharma P., Gabiria U., Jantunen E., Baglee D., "A big data analytical architecture for the Asset Management", The 9th CIRP IPSS Conference, PP: 369-374.

[2]. Datasets for Rainfall available in https://www.kaggle.com/datasets

[3]. Tripathi G., Gupta R., "A New Approach for Rainfall Prediction using Artificial Neural Network", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 7 Issue VII, July 2019

[4]. Bengio Y., Courville A., Vincent P., "Representation learning: A review and new perspectives," Institute of Electrical and Electronics Engineering, vol. 35, 2013.

[5]. Jordan M., Mitchell T., "Machine learning: Trends, perspectives, and prospects," Science, vol. 349, PP. 255-260, 2015.

[6]. Chakraborty S., Nagwani N., Dey L., "Weather Forecasting using Incremental K-means Clustering", International Conference in High Performance Architecture and Grid Computing, Vol.169, Part-2, PP:338- 341,2011.

[7]. Fuchs R., Waser J., Groller M., "Visual human with machine learning", IEEE Transactions on Visualization and Computer Graphics, 15(6):1327–1334, 2009.

[8]. Trombe P., and Pinson P., "Automatic classification of offshore wind regimes with weather radar observations", IEEE journal of selected topics in applied earth and remote sensing, volume: 07, No: 1, PP: 116-125, January 2014.

[9]. Tino P., Nabney I., "Hierarchical GTM: Constructing localized non-linear projection manifolds in a principled way", IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5):639–656, 2002.

[10]. Chawla G., Bamal S., Khatana R., "Big Data Analytics for Data Visualisation : Review of Techniques", International Journal of Computer Application (0975 - 8887), Volume 182 - No. 21.

[11]. Manoj K., Madhu A., "An Application of Time Series ARIMA Forecasting Model for Predicting Sugarcane Projection in India".

[12]. Hannan E., "The estimation of the Order of ARMA process, Annals of statistics. Vol. 8, PP. 1071-1081.

[13]. Raymond Y., "An Application of ARIMA model to real-estate prices in Hong Kong", Jouranal of Property Finance, Vol 8, PP. 152-163.

[14]. Majumdar N., Bhatnagar A., and Shukla S., "Using Linear Discriminant Analysis for Dimensionality Reduction for Predicting Anomalies of BGP data", International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019 Volume 8-11, pp. 1989-1995.

[15]. Bhatnagar A. Majumdar N., and Shukla S., "BGP Anomaly Detection using Decision Tree Based Machine Learning Classifiers", International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019 Volume 8-12, pp. 4015-4020.