GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES
BIG DATA ANALYSIS THROUGH R FOR WEATHER MONITORING

Vandana B[*1] & Dr. S Sathish Kumar[2]
[*1]Research Scholar, Dept. of CSE, RNSIT, Visvesvaraya Technological University Belagavi, India
[2]Associate Professor, Dept. of CSE, RNS Institute of Technology Bengaluru, India

ABSTRACT

Growth of technology has resulted in a large amount of data. Data is collected from various sources and it is in the form of both structured and unstructured. This variety of data is referred as Big Data. Traditional data analytic technology is inadequate to process this huge amount of data. This paper describes importance of big data analytics and visualization. It gives an overview about various Big Data analytics and visualization tools. The proposed analytics is carried out using temperature data through R tool. It enables to develop a decision making system.

Keywords: *Big Data, Visualization, R Tool.*

I.   INTRODUCTION

Big Data has gained great importance in recent years because of huge amount of data gathered from various devices like satellite images, sensors, mobile data, historical data etc. It covers various domains of data like ecommerce, agriculture, quantum computing, social media, transaction records etc. It constitutes variety of data generated from different kinds of devices. Big data refers to large volume, variety and velocity of data. Traditional data processing techniques are not efficient to process the big data. New tools are needed to process this large amount of data. Data visualization tools are required to analyses the large amount of data generated by various sources. In this paper we discuss various data visualization tools used to analyses the big data. Bar charts, pie charts, maps are used to visualize the data effectively. Data visualization tools are used to extract the information from the data more efficiently. It helps the business people to generate a decision making system.

It is used to analyses huge volume of data. It helps to identify the patterns, similarities and frequency of occurrence of data and helps to retrieve accurate results. Large organizations can take better decisions using big data analytical tools. Hadoop, Map reduce techniques are used to process large volume of data which cannot be analysed effectively by traditional data processing techniques.

Big data analytics helps extract useful information from the gathered data. Various analytics and visualization tools are used to process this large volume of data.

II.   R-TOOL

R is an environment for data analysis. R has built in libraries which supports time-series analysis, statistical analysis etc. It supports various types of graphical and clustering techniques. It is an open source platform which can be used effectively for data handling. It supports object oriented concepts. It consist various tools for data analysis. It runs on different types of operating system Windows, Linux, and Mac OS [14].

III.   RELATED WORK

Majority of the data generated today will be in unstructured format. Extracting the information from this unstructured data is a challenging task. The value of information is high when it is analysed in a proper format.

Big Data analysis is carried out in different domains to develop a decision making system. It is performed on various types of data like streaming data, media data, medical data, agriculture data etc. Various data mining algorithms like LDA, Random forest, K-Means etc. are integrated with big data to retrieve the data efficiently.

Before visualizing the data cleaning and pre-processing techniques are necessary to retrieve the accurate information.

Different types of data visualization tools are used to generate interactive maps, graphs, charts etc. to help the organization to take timely decisions. It is used effectively in the field of medicine, agriculture, marketing, trading etc. to derive predictive models.

Tools like R are used to generate an accurate predictive model for big data in various domains.

In the proposed work various Big data analytical methods are discussed. Characteristics of Big Data are explained in detail. Big data stack is proposed. It also focuses on key issues in big data processing. [1] The work proposed by R.S. Raghav et al gives an overview about data analytical techniques. Various data visualization tools are discussed. It focuses on importance of data visualization and various tools associated with data visualization. [2] Importance of pre-processing techniques in data mining is discussed in detail. Various data cleaning approaches and methods used to remove the noise from collected data. It explains the importance of data cleaning method and need of a hybrid data cleaning algorithm. [3] Proposed work describes density based clustering algorithm. It deals with the challenging issues associated with clustering algorithm. It explains the density based clustering algorithm for data streams. Proposed work discusses density micro clustering algorithm and grid based clustering algorithm. [4] Proposed work describes classification methodology for big data analysis. It highlights the necessity of data analysis technique in large volume of data. Decision tree technique using keel software is implemented. [5] Application of Big data techniques in portable network is discussed. It gives a solution for self-healing property of portable networks using big data analytic technique. [6] It uses huge amount of data generated from social network and applies K-Means and support vector machines to predict the possibility of dengue fever. It uses huge amount of data generated from social network and applies K-Means and support vector machines to predict the possibility of dengue fever. [7] Work describes the usage of big data analytics in health care industry. It focuses on tobacco smoking youth percentage's tool and Tableau is used perform data analysis. [8] Vatsal Jatakia discusses various search algorithms that can be used extract patterns in Big data. It helps to improve the customer retention rate in large organization. Exhaustive search, Beacon guided search, nearest neighbour search and Genetic algorithms are discussed with respect to big data [9]. K. L. Ponce-Guevara et. al. develops a tool to analyses the vegetable crops data from the green house. Discuss about the integration of data mining and big data techniques. Develops a predictive model for crop growth in green house environment is developed using big data tools. [10] Priyanka P. Shinde et.al describes R tool. Random Forest and LDA algorithms are applied to develop a model. Predictive model is developed using R analytical tool. [11]. Proposed work describes fast data retrieval approach using R analytical tool. Usage of R tool in manufacturing environment [12]. A smart phone based knowledge sharing system is developed to address the needs of small scale farmers. [17]

IV.   DATA VISUALIZATION

Data visualization techniques help to extract the information from raw data. It gives more knowledge about the required pattern. Advantages of data visualization are

A. Visualization of Relationships between the Activities.
Relationship between various business activities is easily understood. Patterns can be analysed in the business. In the competitive business environment it helps to generate a decision making system.

B. Time Saving
Since user will get a clear idea about patterns analyzation time will be reduced. Organization can focus on other work.

C. Trend Analysis

In the current business scenario trend analysis is very much essential. Market trend is changing based on geographical location, customer type, Customer interest, average income etc. Big data analysis helps to identify the current market trends. It can be used to promote the new products.

D. Information Sharring

Information can be easily gathered from the raw data. It is shared among the needed users in timely manner. Since Big data supports large volume of data like trade analysis in time data analysis is needed to generate decision making system.

E. User can Interact with Data Directly

Charts and traditional tools allows the user to view the data but Big data analytical tools provision the user to directly interact with the data.

F. Faster Processing of Data

Visual data is processed fast than text data. Patterns can be easily identified by visualization tools. Online data visualization techniques are more effective when compared to traditional spreadsheets, charts and graphs. Proposed Architecture.

V.   METHODOLOGY

Temperature data is collected using DH22 sensors in the farmland. Data cleaning procedures are applied to remove the noise. Temperature data is segregated on hourly basis. R Tool is used to analyse the temperature values. It is used to develop a predictive model of temperature analysis. Fig. 1. Describes the architecture of the proposed work.
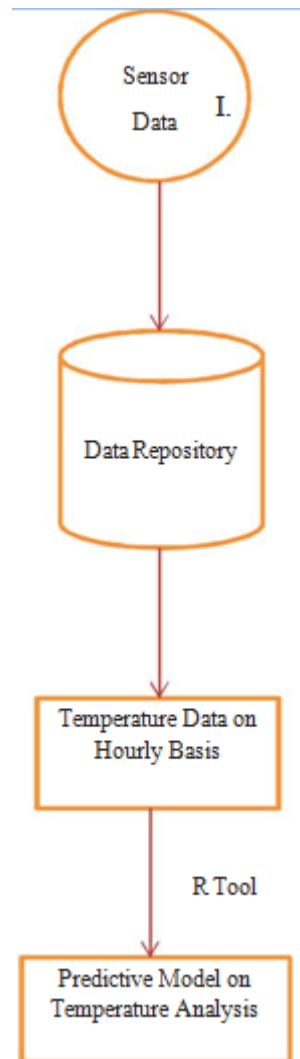
Fig.1. System Architecture

VI.   BIG DATA VISUALIZATION TOOLS

A. Tableau
Tableau is used for manipulating big data. Tool provides drag and drop facility. Coding is not needed for this tool. It is compatible with different types of data generated from various sources.

B. Fusion Chart
It is JavaScript based charting library used for web and mobile applications. It is compatible with XML and JSON data sources.

C. Qlikview
It is used in business intelligence and data analytics field. It has customizable features which can be used to create in a faster and efficient way. It is supported by larger user community.

D. High Charts
It contains charting functions written in java script language. It supports various browsers and mobile devices. It efficiently processes the real time data from the JSON data source. It also supports various types of browsers. It has compatibility with android and iOS operating systems.

E. Datawrapper
It is used analyses statistical data. It is used in media application. It easily creates charts from the csv files. It is used to develop interactive applications.

F. Infogram
It is used to retrieve the information from large datasets. Various templates are available to customize the visualization according to user requirements. It is majorly used in the field of journalism and accounting.

G. Leaflet
It is open source software compatible with JavaScript, html and cuss. User can create interactive maps. It is used in data visualization application development for mobile devices.

H. ChartBlocks
It is an online tool helps build the visualization from spreadsheets and live field. It is supported by HTML5 and Java script library. It is suitable for any screen size.

K. Google Charts
It is supported by HTML 5. It provides multiple browser compatibility. It is designed to work on various kinds of mobile devices like Android and iOS. It has user friendly interface.

L. Polymaps
It represents geographical data. CSS and Java script supports the polymaps. User can create interactive maps using this.

M. Silk
It generates interactive charts and maps. Less coding is sufficient to use this tool. It also supports collaboration with user groups.

N. CartoDB
It is used to represent location data. It is used in creation of maps. It also provides features like collaboration and controlled access among user groups.

O. Plotly
It creates graphs, 2D and 3D charts without much programming. It also supports custom themes. It is handy tool beginners.

P. Qubole
It performs the data analytics on Google, AWS, and Azure Cloud systems. It is compatible with hive and spark.

Q. BigML
It supports machine learning concepts for big data analytics. It easily generates a predictive model from huge data.

VII. RESULTS

Temperature data is collected using DH22 sensor on hourly basis. Data is collected for one month and testing is carried out on temperature data values using R studio. Temperature is read in an experimental environment and R tool is used to analyse the temperature.

Fig. 1. Displays loading of a csv file to R studio which consists of collected temperature values.
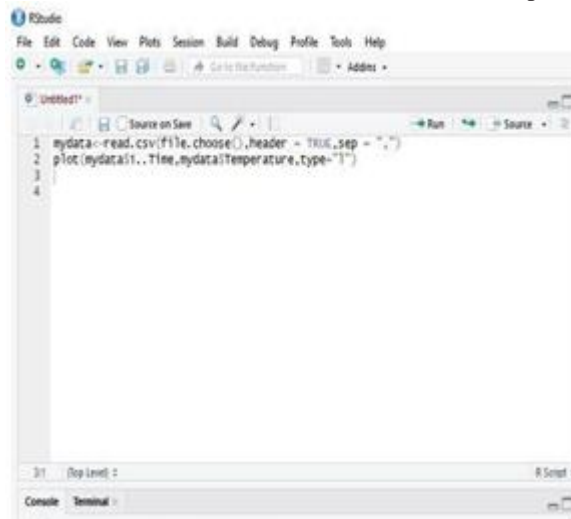

Fig.2 Loading the Temperature Data

Fig.2. displays the analysis of temperature data using R studio. Temperature values are collected on hourly basis.
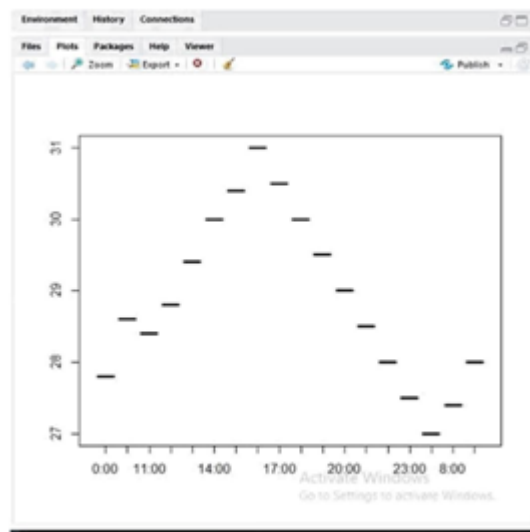
R tool is used to analyse the temperature variation

Fig. 3. Temperature Data Analysis

VIII.   CONCLUSION

Data analytics and visualization is important to retrieve the information from huge data sets. It is important in various fields to develop a decision making system. Proper data visualization tool is important for the large organizations to improve the business value. Big Data visualization tool results in time saving and reduced human errors while analyzing the data. In the proposed work R tool is used to analyses the temperature data.

IX.   FUTURE WORK

Weather monitor plays an important role in the field of precision agriculture. It can be used for automation of irrigation activities which reduces the water consumption in the field land. R tool can be used to develop a predictive model for weather monitoring.

X.   ACKNOWLEDGMENT

The authors would like to thank the H.o.D, Principal and Management of RajaRajeswari College of Engineering and RNS Institute of Technology Bangalore for their generosity towards supporting the research work.

REFERENCES

1. Yojna Arora, Dr. Dinesh Goyal, "BigData: A Review of Analytics Methods & Techniques,"2nd1nternationalConference on Contemporary Computing and1nformatics, pp. 225-230, in IEEE 2016.
2. JayaramHariharakrishnan, Mohanavalli.S, Srividya, Sundhara Kumar K.B,"Survey of Pre-processing Techniques for Mining Big data,"IEEE International Conference on Computer, Communication, and Signal Processing (ICCCSP-2017) in IEEE 2017.
3. K. Shyam Sunder Reddy , C. Shoba Bindu ," A Review on Density-
4. Based Clusteringi Algorithms for Big Data Analysis," International conference gon I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017). in IEEE 2017
5. Shilpa Patil , Ashok kumar P S, Prasadgouda Patil, Puneet Palagi, "Data Analysis & Classification Methodology for Knowledge Discovery in Big Data ," International Conference on Inventive Systems and Control(ICISC-2017) in IEEE 2017
6. Shilpa Balan, Nishant Shristiraj, Vrunda Shah, Anusha Manjappa, "Big Data Analysis of Youth Tobacco Smoking Trends in the United States,",2017 IEEE International Conference on Big Data (BIGDATA), IEEE 2017.
7. Vatsal Jatakia, Sameer Korlahalli, Khushali,"A Survey of Different Search Techniques for Big Data,"2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), in IEEE 2017.
8. K. L. Ponce-Guevara et. al. ," GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial)", in IEEE 2017.
9. Priyanka P. Shinde, Kavita S. Oza, R. K. Kamat, "Big Data Predictive Analysis:Using R Analytical Tool ," International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017), IEEE 2017.
10. Nitin Naik1, Paul Jenkins, Nick Savage, Vasilios Katos, "Big Data Security Analysis Approach Using Computational Intelligence Techniques in R for Desktop Users," in IEEE 2016
11. Komal Verma Rajiv Pandey, SM IEEE Arpit Gupta, "Optimizing Geographical Clusters through R for Job Allocation in Make in India," 8th International Conference on Computational Intelligence and Communication Networks, in IEEE 2016.
12. R.S. Raghav, Sujatha Pothula, T. Vengattaraman, Dhavachelvan Ponnurangam, "A Survey Of Data Visualization Tools For Analyzing Large Volume Of Data In Big Data Platform," 2016.
13. P .S Sangeetha, Sweety Francia, Mistica Dhas.Y, "Comparative Study of Big Data Techniques Curative in Portable Networks," 2016

14. *Tran Duc Chung, Rosdiazli Ibrahim, Sabo Miya Hassan, Nurfatihah Syalwiah Rosli, "Fast Approach for Automatic Data Retrieval using R Programming Language," 2016.*
15. *www.r-project.org*
16. *Giuseppe Bruno, Demetrio Condello, Alberto Falzone, Andrea Luciani, "Big Data processing: Is there a framework suitable for Economists and Statisticians,", BIG DATA CONFERENCE, VOL. 14, NO. 8, DECEMBER 2017,"IEEE 2017*
17. *Ricardo J. Machado, Marcelo Aparecido Carlos, Marcelo Nogueira, "Analysis of Dengue Outbreaks Using Big Data Analytics and Social Networks," The 2017 4th International Conference on Systems and Informatics (ICSAI 2017), in IEEE 2017.*
18. *Vandana B, S Sathish Kumar, "A Smart Phone Based Information Sharing System in Precision Agriculture", Ist International Conference on Sustainable Engineering and technology, Bengaluru, 2018.*

.