International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 09 | Sep 2020 www.irjet.net p-ISSN: 2395-0072

BALANCING BETWEEN 'BIG DATA ANALYTICS TOOLS'

Shilpa Pandey[1] and Prof. Meenakshi Garg[2]

[1]*Student, Department of MCA, Vivekanand Education Society's Institute of Technology (VESIT), University of Mumbai, India*
[2]*Assistant Professor, Department of MCA, Vivekanand Education Society's Institute of Technology (VESIT), University of Mumbai, India*

Abstract: Large data set and huge collection of information from various data bases and sources. It can be of any type and tough to be interpreted hence we need some tool or software that can easily analyses the data and give us some insight out of it. Among various interesting tools R, python and tableau are the biggest tools which deals with the big data analytics also it generates the output in visualization technique that are More understandable and presentable. In this paper we are comparing and contrasting the working of both the tools with some big dataset and sources along with the importance and need functionality of the tool in the field of data analytics. This study gives the clear picture of growing data and the tools and software which can help more effectively and efficiently.

Keyword Used: Tableau, Tableau server, R programming and R stream, Python

1. Introduction: Big data is a field that helps you to understand the business objective and visualize it simultaneously according to your business perspective the different way to systematically extract information or visualization that are too complex to be dealt with by the general data processing application tools. Large amount of Data is always seen in terms of their 3Vs volume, velocity, variety and added value to the business process. It is mostly oriented in terms of large processing, systems such as grids and easy to understand but majority of its aspect is shifted to cloud area. Foundation of this technology which is the backbone gets upgraded with each requirement will be needed data is a term that describes the large volume and data both structured and unstructured that gives our business day to day basis. What organizations do with the data that matters changes according to their perspective. Big data can be analysed for insights that should be better decisions and strategic business moves of ach and every in individuals.

These are few analytics tools or software that are used in the industry:

- Hadoop,
- Apache Hadoop: This is the most promising and used tool in data industry with its enormous capability of large-scale processing and visualizing data
- Apache Spark
- Apache Storm
- Rapid Miner
- Mongo DB
- Tableau: Currently, the most used tool visualizations, data discovery and visualization is R, Python and Tableau. Tableau is one of the fastest busines intelligence (BI) tool. It is fast to connect deploy easy to learn and very useful for customer. Tableau has five main products facilitate to needs for professionals and organizations. They are: tableau desktop: for individual use and licensed person. tableau desktop: for individual use and licensed person tableau server: collaboration for any organization

Tableau desktop has both a professional and personal use. Tableau online is available to users in yearly subscription and manual and enlarges to support thousands of users. Whereas, the "R" is statistical version of what is used to handle the code big data. R is a scripting language like a programming language hence it is better tool for users. It integrates with the complex strategy publishing system or in other words. Easily possible that the statistical output and graphics generated by R and python can be merged with publication-quality documents. It is easy to operate and use including an economical tool.

Methodologies Used: In this work we have followed three methodology. It includes three steps like: *Data collection:* we are using data sets in this work. The data set are collected from some companies or ERP system. *Analyzing the data and information:* this is the second step of our methodology. Here the datasets are analyzed using the tools R, python and tableau server. *Comparing the performance out of them:* This is the final step of the proposed technique. Here the tools are compared on the basis of their past performance.

Problem Statement: The aim of this work is to analyze and undertand the dataset using data tools analysis will be done using tools like R, Python and Tableau. A comparative study will be done on the basis of the past and current performance of the tools and software.

Dataset Description:

This dataset is collected from the UCI Repository platforms. This is a multivariate and descriptive dataset. Number of instances in this dataset is 748. It has five attributes named

Recency, Frequency, Monetary, whether it has donated blood. It says about months since last donation Frequency gives the information about total number of donation. Monetary attribute value holds the total blood donated in c. It time is the attribute which shows months since the first data set. The last attribute holds binary values stating whether a users has donated blood. The dataset does not have any missing value. We have used the dataset in CSV and text file for R and .xls format in Tableau.

Forest Fires Data Set*:* This dataset is also collected from UCI Repository values. The characteristics of this dataset is multivariate which have numbers of instances and 16 attributes text format of this. Dataset is used in R where the .xls format is used in Tableau. The dataset used in the present research work was downloaded from the Integrated Network for Social Research website.

Results: The result shows us that the tools worked on the big data set. It has generated different kinds of pictorial representation and visualization of the analyzed objective. While working with both the tools, we found some advantages and disadvantages of their software. They are listed as follows:

R Scripts Consideration: R is the most accurate measurable examination package available for users

The graphical abilities of R scripts are extraordinary, giving a completely programmable design platform dialect that outperforms most other measurable and graphical package

R scripts is free and open source programming, anybody to utilize and, imperatively, to chamge it. R is authorized under the GNU General Public License, copyright assisted by The R Foundation for Statistical Computing in industry

Tableau software desktop/Server: Tableau has an excellent user interface and secure.

Its integration feature is also very attractive and efficient. It can integrate with other big data platforms like Hadoop And big data mining process

Tableau supports in mobile devices, desktop. The report on the tableau dashboard is automatically optimized in mobile and server.

Disadvantages: R Scripts Consideration: To use R script for learners one needs to learn it very wel006C. Otherwise it cannot be used effectively And programming too. All the packages used in R is not always give perfect result initially Tableau: Initial data processing is needed here for users in Tableau and this should be done by professional kit expert and daily users.

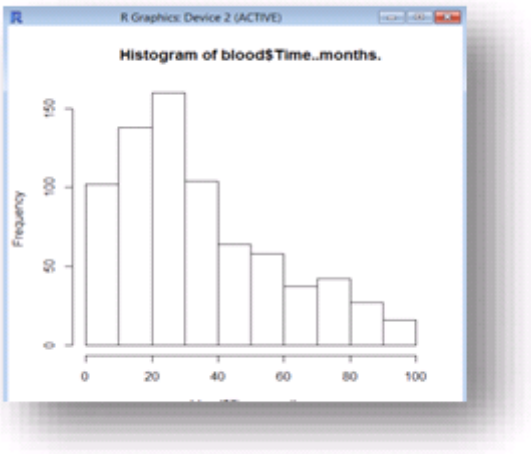

Chart -1: Blood Transfusion Service

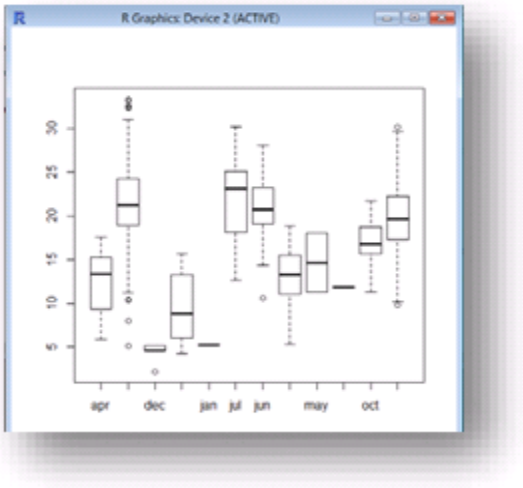

Chart -2: Forest Fire Dataset (Scatter plot)

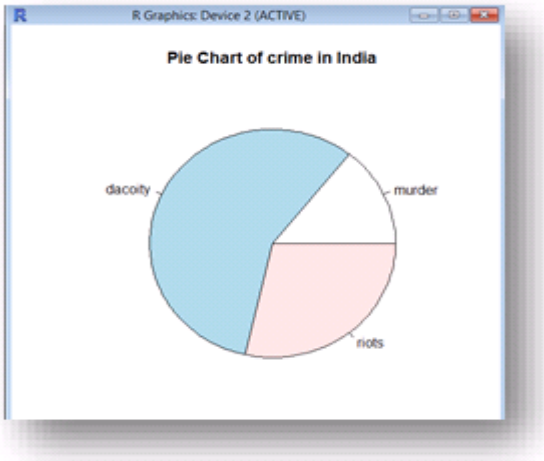

Chart -3: Crime dataset (Pie chart)

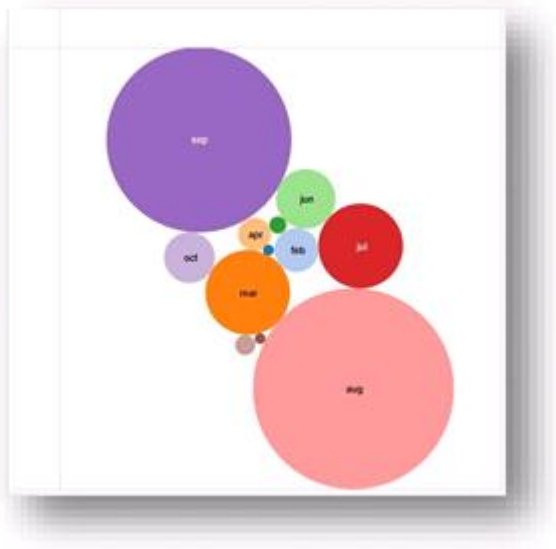

Chart -4: Forest Fire dataset

Conclusion: Data analytics using the tools is very effective, less time consuming, innovative and interesting. Our study we conclude that Tableau is the more efficient tool more powerful than R script in big data analytics. The usefulness of Tableau in data analytics can be measured by its performance, user friendly environment, easy and speed. There are more features of Tableau which was not taken into account of study in this study as the time was limited.

References:

1) Keim, Danil A,. "Visual analytics: Scope and hallenges", Visual Data Mining, Springer Berlin, 2016, 76-90.2) Fan wei, and Albert Bitet. "Mining big data: current status, and forecast to the future." ACM SIAKDD Explorations Newsletter, **14**, (2015), 1-5.3) Katal, Avita, Mohammad Wazid, and R. H. Goudar, "Big data: issues, challenges, tools and good practices", Contemporary Computing (IC3), Sixth International Conference on IEEE, 2015.4) Kalambe, S.D. Pratiba, and Pritam Shah. "Big Data Mining Tools for Unstructured Data: Review, IJITR, **3**, (2015).