International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 05 | May -2020 ISSN: 2582-3930

Data Analytics and Visualization

Anurag Kumar[1], Anuj Singh[2]

[1,2]*Student, Dept. of Computer Science Engineering, RKGIT Ghaziabad, Uttar Pradesh, India*

Abstract -Data analytics means to research large volumes of data or big data. The info is collected from an oversized assortment of sources, like websites, videos, images, and documents. The most important aim of data Analytics is to find new models and relationships that are invisible, and it can give a new understanding of the users who created it. Data could be a new era of the world's economic and global changes. Data Analytics provides a grand challenge on the way of highly performable systems and algorithms to integrate the data and provide largely hidden knowledge from datasets that are diverse, complex, and of an enormous scale. Possible future advancements include optimization algorithms, methods, systems, and data analytics software that effectively and successfully discover useful and hidden knowledge from the data. The development of knowledge analysis will significantly change infertility, predict, and ensure the tasks performed by accountants and auditors. However, applying data analytics to those tasks can generate concerns about organizations maintaining data confidentiality and integrity, the results of knowledge analytics on data bias, and whether data analytics will increase the legal liability of auditors. These concerns motivate research into how data analytics is and can be, employed in accounting. because it is completed for decision making, it's important to grasp the only purpose of knowledge analysis. The main purpose of data analysis is that the interpretation, evaluation & organization of knowledge and to form the data presentable.

1. INTRODUCTION

Data analysis is the technique by which most information sets are decomposed to reveal valuable information. These data situations include business models, patterns towards customers, filled instances, and loose communication systems. Examination findings often open up new possibilities for income, improve efficiency and productivity, and promote greater economic benefits. The analysis basically classifies data into things that the organization knows or doesn't know, and can be used to measure any current, past, or even future events. They transform ideas into influence, and link trends and patterns to the real objectives of the corporation. Analysis of the data is more specific and directed than data science. It concentrates on accomplishing specific objectives through the categorization of large data sets. The analysis is more automated because it can help provide better insights in specific areas. Data analysis also includes the analysis of large data sets to find smaller and more useful pieces of information to achieve organizational goals. Organizations often rely on information to help them resolve basic business deals. Useful data analysts allow information researchers, leading biologists, and a variety of investigative experts to bridge the gap in the exchange of information. They can also benefit from large-scale information surveys to analyze information that does not exist in traditional business plans. The most important issue in the minds of data analysts has always been recruiting specialists and the risks of internal evaluation and data breaches. The amount of data to be analyzed and changed is also an important element of control.

Data Analytics classified in 4 types:

- Descriptive Analytics
- Diagnostic Analytics
- Predictive Analytics
- Prescriptive Analytics

In Descriptive analytics, we summarize data and convert it into a type that may be simply understood by humans. they will describe well a happening that has occurred within the past. this type of data analysis is useful in deriving any sequence if any of previous events or understandings are viewed in such a manner that higher long-term methods are drawn from them. This is a mostly used analytics method in the organization.

In Diagnostic analytics, Historical data is measured against alternative data to answer the question of why one thing has happened. Diagnostic Analytics provides an in-depth insight into the drawback. At the same time, the organization should have careful information at its disposal, otherwise, the knowledge assortment could prove to be individual and long for each issue.

Predictive analytics is all about prediction. Whether or not there is a chance that something will happen in the future, predict a quantitative amount, or estimate a certain amount of time that one thing may happen – these square measurements are all done through prognostic models.

Predictive models usually use the spread of variable information to make predictions.

In Prescriptive analytics, we explain the step-by-step process in a situation. the aim of prescriptive analytics is to virtually order what action to require to eliminate a future downside or take full advantage of a promising trend. Prescriptive analytics uses advanced tools and technologies, like machine learning, business rules, and algorithms, that make it subtle to implement and manage.

1.1. Data Analytics – why we need?

Data Analytics is needed in all organizations. Organizations gather data from customers, companies, the industry, and by doing the survey. After collection, data is then extracted and classified according to the requirement and analysis is done to study patterns of transaction and so on.

The idea is to form a sense of the info you've got, to research it, and share better business prospects within the near future and the way you're getting to roll in the hay is with the concepts of analytics. Data analytics involves generating trends, patterns and valuable information from a group of available data that, unless examined, can be of no ill effect. It's a kind of business analysis now used to make profits and make better use of the resources. If this data is not analyzed, it is an urge to waste, whereas if properly analyzed, this data can help us to find information that is powerful to show the changes in the pattern of how an organization working or going on. Just imagine a source of unleashed information exists, and you haven't dived in yet to urge the grip of it. Your business can take competitive advantage of it and do wonders with the info. It will obtain perspectives that can provide an economic advantage for an industry or an organization.

Data and information are growing rapidly, the data growth rates are so high that the data available to us is going to become volatile throughout the near future. Data is obtained through a whole hundred users, industries, and companies. Try to amalgamate if this data, not the large data but the info you've got gathered from your business if wasted what you'll be losing on. Modeling and visualizing are among the main aspects of data analysis and then you really need to understand the complexities of it as a whole in the desire to get an up gear from this. Earlier data required a number of qualified experts to process information while we now have tools to handle high-speed data analysis on large amounts of data, and this provides a chance for the entrepreneurs to include data analytics when making decisions.

2. LITERATURE SURVEY

Data analytics require more resources and energy to analyze, process, and manage the data. But sometimes these resources may not be available to the machines which may be due to the availability and budget of inexpensive machines. In the area of information many organizations want a single and effective solution to process the data which may be available or not but if it is available than its cost must be much higher. A solution can be a single system having a number of processors and memories but it will be very expensive to build such a system. Another solution can be the clustering that is to build a system with high availability clusters. This type of clustering system looks like a single system and also it requires proper installation, management, and administration services. Organizations were looking for a solution that can be cost-effective. Another solution cloud computing comes into picture which is not so expensive i.e. very economical. It also contains the required resources which are necessary to perform computations. This solution is based on the single instruction multiple data algorithm. In which a large volume of data is transformed. In this processing of each and every data item happen independently that is the processing of each data item doesn't depend on another one and also the processing of one data item doesn't have any impact on the other data item.

P.N. Tan, V. Kumar, provides many important properties that should be studied so that we can get the right and effective measure for a data analytics application. These properties have been studied in comparison using 21 measures which have initially been developed in various areas such as Science, Machine learning, data mining, and statistics. Finally, He represents an algorithm for choosing a small set of patterns in a way that data analysts can easily find a way that suits their necessities in the best way by grading or positioning the small set of patterns. He also proposed a summary of few prevailing "measures of interestingness" and also comments on properties of the same. Overall, interestingness measures are divided into objective and subjective measures. This paper emphasis an objective measure of interestingness only.

According to Dhruva M.S, data analytics researchers have found the info collected is split into numerous data Analytics application like, Structured Analysis, Text Analysis, etc. There are many techniques that will be accustomed to method datasets. Some techniques are machine learning, A/B testing. These techniques, analyze the new combination of datasets and supply the United States the higher result for the business and organizations. Reports of International information Corporation predicts that world information from 2005 to 2020 can grow by the factor of ten. large data examination venture is promptly rising because the honorable declare perceiving business and innovation slants that are irritating typical data administration ways.

Data analysis involves complex processes that extract useful knowledge from large data repositories. Compared with traditional approaches, data analytics provides benefits in terms of cost-effectiveness, usability, and ability to detect new patterns in real-time. A number of challenges and risks may arise with data analytics. Visual representation technology shows the level of participation of the consumer or user in learning and implementation. It easily identifies users with a low level of participation. It generates reports based on interactions between users. AWE is an alert engine and evidence-based system that helps identify learners to improve analytical performance. Data analysis can be performed in a variety of Data Analytics applications such as, Structured Analysis, Text Analysis, etc.

3. PROPOSED WORK AND PROCESS

To perform the data analytics, we have to plan a proper methodology and rules so that the system can perform perfectly as needed.

Below are some important instructions for the data analytics system that provide us a better understanding to plan a system design:

- Create the best and the lowest complex algorithm.
- Store the data in the correct manner as required by the algorithm.
- Develop the best and most understandable User interface.
- Show all the output as per the datasets and need from the user.

Process of Data Analytics are following:

Data requirements

Data are needed as inputs to the analysis. The choice of analytical approach shall determine the data requirements for the analytical methods to be used to require specific data content, formats, and representations, guided by domain knowledge. These data may be separated by age, demographic, income, or gender. Data values may be numeric or may be divided by category.

Data Collection

Data can be collected from a variety of sources. The data are identified and gathers in a structured, unstructured, and semi-structured manner, that is relevant to the problem domain. If we encounter gaps in data collection and terms, we may need to revise the data requirements and collect more data. It can be collected from the interviews, online documents, and from the free online sources.

Data understanding

We need to perform a variety of techniques referred to begin understanding the messages contained in the data. Visualization techniques and descriptive statistics can help us understand the content of the data, assess the quality of the data, and discover final insights into the data. The process of understanding may result in complex data cleaning or additional requests for data, so these types of activities may iterative in nature.

Modeling and algorithms

Starting with the first version of the ready data set, we need to use a training set, historical data in which the result of interest is known to develop probabilistic or accurate models using the analytical approach already described. In general, models may be developed to determine a specific variable in the data based on other variables in the data, with some residual error based on the accuracy of the model.

Data Presentation:

Once the data is analyzed, it may be presented in a number of formats to the users of the analysis to support their specifications. As such, a large part of the analytical cycle is iterative. When determining how to present the results, we may need to consider data visualization techniques to help the public communicate the message clearly and effectively. Data visualization uses displays of information (such as tables and charts) to help express key messages found in the data. Tables and graphs are helpful for a user who may look up numerical statistics.

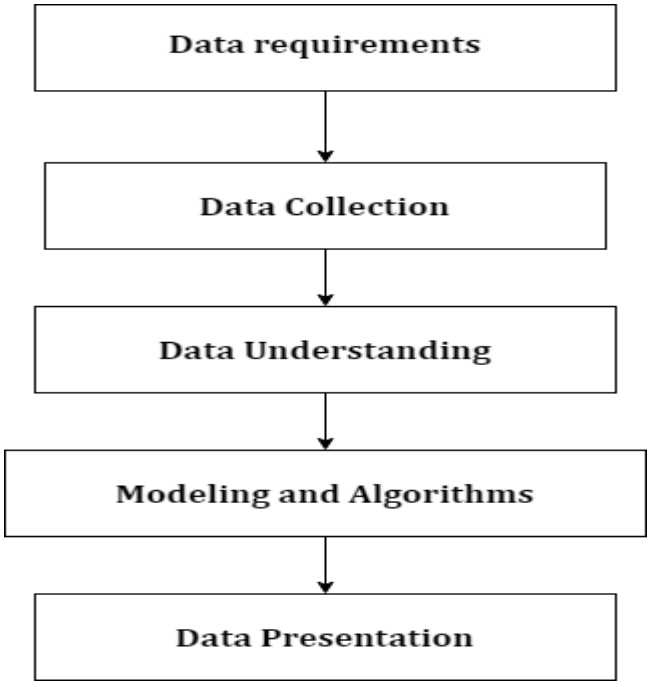

Fig. 1. Data Analytics Process

4. DATA VISUALIZATION

Visualization of data is nothing but a visual representation of data. This visual form may be a diagram, graphs, lists, maps, etc. This depiction helps people understand the data magnitude. Data visualization provide us to make the decision by seeing graphically presented analyzed data, allowing them to identify new patterns and knowledge in data. Data visualization can help businesses explain trends and stats much easier with the use of charts, graphs, and design elements. Visualization of data also shows patterns, trends, and correlations that might otherwise go undetected. we can use plots, statistical graphics information graphics, and many other techniques for the simple and reliable communication of information. Mathematical data may be embedded using dots, lines, or bars to express a quantitative message visually.

To understand, let's take one example, Data Visualization very clearly. In order to help the business person understand the transaction history of the items respectively, a pie chart or graph will help him/her to understand the purchase history numbers better. So, business owners will be able to plan their business according to trend.

Below, there are various visualization techniques:

Charts

A chart is the easiest way to be able to show the development of one or more data sets. Charts vary from bar and line charts showing the relationship between elements over time to pie charts showing the components or proportions of the elements of an entire.

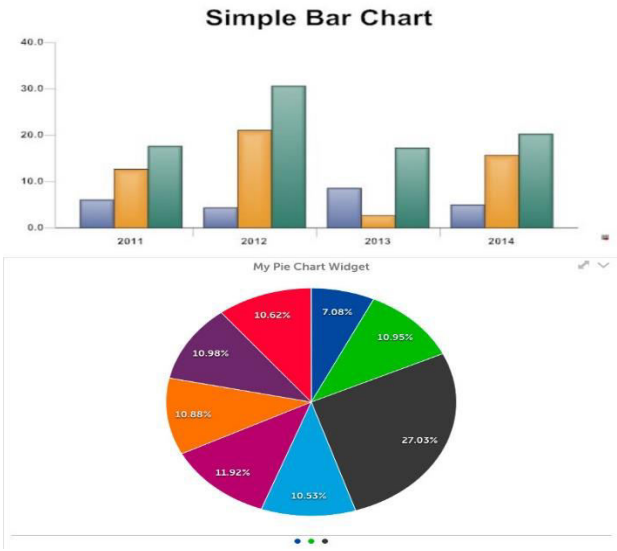

Fig. 2. Bar & Pie Chart

Plots

Plots allow two or more data sets to be distributed over a 2D or even 3D space to show the relationship between those sets and the parameters on the plot. Plots vary too: the most traditional are scatter and bubble plots. Although when it comes to big data, analysts use box plots that allow the relationship between large volumes of different data to be visualized.

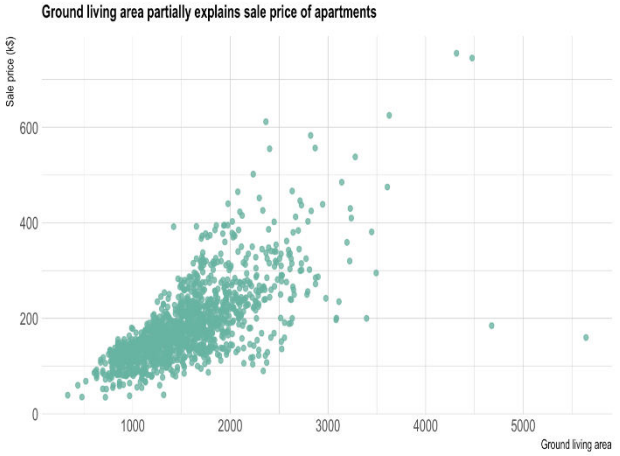

Fig. 3. Scatter plot

Maps

Maps are widely used in various branches of industry. They allow elements to be positioned on relevant objects and areas-geographical maps, construction plans, website layouts, etc. Heat maps, dot-distribution maps, cartograms are among the most popular map visualizations.

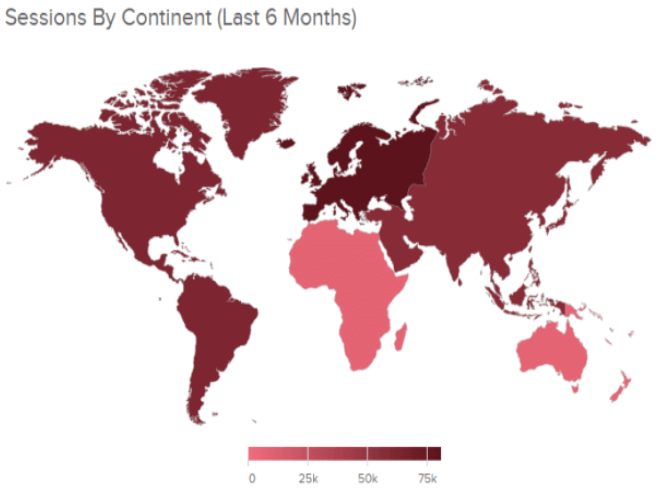

Fig. 4. Heat Map

Histograms

It is a precise representation of the numerical data distribution; it only relates to one variable. Includes bin or bucket- the range of values that divides the entire range of values into a series of intervals and then counts how many values fall in at each interval. Bins are consecutive, non-overlapping variable intervals. When the adjacent bins left no gaps, the rectangles of the histogram touch each other to suggest continuity of the initial value.

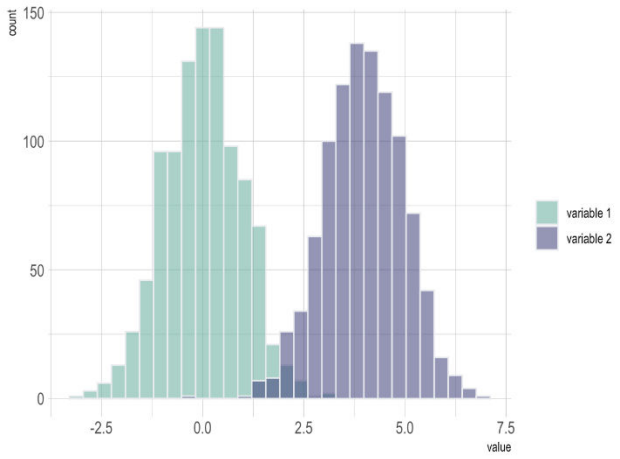

Fig. 5. Histogram

5. CONCULSION

We see information overloading practically everywhere today. In the Era of data, we tend to live in voluminous sorts of high-velocity knowledge are being created daily, and among them are intrinsic details and patterns of hidden data that should be extracted and used. Data analytics is therefore applied to leverage business modification and improve the process by implementing efficient analytical techniques to the vast knowledge and revealing unseen insights and important data. The benefits are many and varied, varying from high-quality education to state-of-the-art medical research, and while more study is needed to make sure that information is protected from exploitation,there are many exciting discoveries waiting to be discovered through data analytics.

REFERENCES

1. Xu R, Wunsch D. Clustering. Hoboken: Wiley-IEEE Press; 2009.
2. Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao & Athanasios V. Vasilakos, Big data analytics: a survey Journal of Big Data
3. P.N. Tan, V. Kumar, J. Srivastava, Selecting the right objective measure for association analysis, Inf. Science
4. Dhruva M.S, Shashikala M.K, Survey on Big Data Analytics IJEACS UK
5. Applications of Big Data analytics and Related Technologies in Maintenance by Jens Baum, Christoph Laroque