

# Data Visualization Tools and Techniques For Datasets In Big Data

# INTRODUCTION

- The term data analytics is also known as Data Analysis, which is used to study raw data to gain knowledge about the information. The data analytics is used by the company can gain the knowledge about the needs of the customers and to improve the business strategies of the company. The company can categorize the collected data from different sources and they can be separated according to the patterns and techniques for proper data analysis. The classifications of data analysis are a) Social Network Analytics b) Business Analytics c) Business Impact Analytics d) Big data Analytics are some of the important analysis which is carried in the present world for analyzing the data.

# TYPES OF ANALYTICS

- **Social Network Analysis (SNA)**

The qualitative and quantitative analysis process of the social network such as (Facebook, Twitter, Research Gate, LinkedIn etc) is carried in SNA. Such kind of method is done to measure the activities of the customers. This process of analyzing is done by monitoring their interaction through chat, post, comments etc[1]. This type of social network analysis can be carried in groups, community

It helps to provide the information in both analysis methods like mathematical or visual so that they can understand the needs of the customer and it also helps to improve their business strategies.

# Business Analytics (BA)

- This type of analysis is used by the organization to measure the performance of their company. It also carried for evaluating their position in the market and to find where they should improve their strategies. This type of analytics uses statistical methods that can be applied for specific product or process by the company. The main goal of company to run the business analytics is to monitor their flow of business and to identify the disadvantages of the existing processes and highlight meaningful data[2]. This helps the company to know the area of improving in their business for future growth and to handle the challenges. The business analytics plays a vital role to make decisions, improves the business strategies

# Social Media Analytics (SMA)

- The term Social media analytics is used to collect and analyze the data from different social Medias and blogs etc. The monitoring of data is done after the data is collected and this strategy is used for improving the process of the company and makes them to produce a better and quality product. This analytics helps to understand the concept carried in the business functions such as marketing, customer service etc. It also helps to improve the customer experience[2]. The Social Media Analytics is the best way to understand real-time choices and behaviour of the customers. The tactics followed by the organization or company behind social media analytics is they can get detailed information about customer base on a more emotional level. The main task during the social media analytics is to find out the which business objective can gain profit to the company

# DATA VISUALIZATION

- The definition of data visualization explains the importance of the data by placing the data in terms of visual context [5]. It involves the creation and study of the visual representation of data which is known as information. The data visualization provides the user to acquire more knowledge about the raw data which is collected from the variety of sources. The visualization can be done by using the dashboards, where the undetected text, patterns and correlations can be easily visualized

## . Improve In Decision Making

- The data visualization helps the organization to view their position and the process carried by the organization. According to the visualization and analyzes of data the company can take better decision and they can also change their business flow according to it[7]

## . Improvement In ROI

They can also increase the company ROI by knowing their plus and minus of their business flow. Once analyzing process is done the visualization gives the company a clear idea about their mistake. By rectifying it the company can have their increase in return on investment.

- **Information Sharing**

- The data visualization helps the company to gain knowledge about their previous and present business flow. The plays a vital role in improving the process to gain information from the raw data and they can share to avoid misconceptions

- **Time Saving**

- It is one of the main advantages of using data visualization in company[8]. Instead of going for trail error method, the company can identified the problem and the immediate countermeasures is taken, so that the saving of time is done by retrieving accurate insights of data in short span of time



# BIG DATA VISUALIZATION

- The Big data visualization explains about the description of data by providing the user with effective visualization techniques. The real time data changes, complex data processing results can be easily shown by using the big data visualization. It can displayed in the form of charts, graphs etc. The business people are forced to know about the every piece of information about their data[9]. Some of the important features of data visualization and the role of data visualizations in big data environment
- **Real-time Data Analysis**
- The most important feature of data visualization is they play a vital role on real time data by providing deep insights to the user about every piece of information about the data[10]. So that the business people can make a better decision by seeing the result displayed by the dashboards

- **Dynamic Nature**

- The massive data collected from variety of sources can be dynamic by changing the view in different types such as graph (bar, line) charts (pie, bar) scatter plots

- **Interactive Presentations**

- The usage of the data visualization tools by the company provides them an interactive presentation of data with reports. It displays all details about the data to the user, so that the user can makes some effective changes according to the result.

- **In-Memory**

- The data can be visualized by multiple users, so that each user will produce different opinion inorder to improve their business strategies[9]. Many types of data visualizations result are stored in memory for easy access.

- **Secured**

- Most of the data displayed in big data environment will be secured by putting some right access to the user. Some data can be viewed by certain users to have security and the data can be compressed to have low memory space.

# ERRORS TO AVOID IN DATA VISUALIZATIONS

- The data visualization in big data environment helps the user to know about the value of data. But they also face some complex problems while handling large volume of data in bigdata environment[1]. Some of the error carried by the business people while displaying makes the user or data analyst in trouble.
- 
- **Exposing All Data**
- It is one of the errors occurred while carrying the data visualization. The importance of data visualization is to give a visual treat for the users and business people. It will give the Knowledge about insights of the data[12]. The dashboard is used to give clear cut information, but some companies will give a clumsy view of data. This kind of approach makes the user in a confused state, where they can't get a proper conclusion on the data.

- **Displaying Errors**

- It is another factor to be noted while visualizing data. The company should know about the wrong side of their business strategies. These methods help the company to know the unnecessary data for making decision. The problem facing while displaying the wrong data is they makes the user to take incorrect decision

- **Lack Of Planning**

- Before displaying the data, the organization should select the proper dashboard for displaying the data. This is because, certain data can be displayed in the form of graphs and it will look worse if the user use some kind of inappropriate dashboards. So they need a proper planning by selecting effective and proper dashboards for displaying

# DATA VISUALIZATION TOOLS IN BIG DATA

- **Data Wrapper**

- It is used to create data visualization and make it very easy for the end user to grasp the knowledge from the raw data[9]. By using the data wrapper tool the user can easily generate graph and it can be done by simple steps with good web based GUI (Graphics User Interface).The user can save time for creating visualizations. The user should upload the data and they should choose which kind of visualization they need for analyzing.

- **Dygraphs**

- It is one of the main data visualization tool used for representing large volume of data. They use java script based charting library. Even though they use some scripting language they are user friendly with an effective output interface. The user can able to get knowledge due to its flexibility interface. The user should have prior knowledge about web programming to get started with a chart.

- **D3**

- The term D3 refers to the data driven documents. It contains JavaScript library to help user bind random data to the DOM (document object model). They have the ability to apply data-driven transformations to the document. As you know, DOM is a programming API that allows programmers to access documents as objects. These objects indicate the structure of document they need to model. The user should know about programming knowledge for creating graphs

- **Raw**

- It is a web-based tool that allows user to simply paste the data. It has very simple steps for creating graphs. The Raw is completely based on the D3.js library, which makes the user to easily accessible, so that the user can have all components in packed form of D3 into a format that is ready to be used by non-programmers.

- **Leaflet**

- It is a lightweight mobile friendly data visualization tool used to study the data generated by high traffic and good conversion rates. It has JavaScript library to help user for developing interactive maps [12]. It is simple to use and works on both desktop and mobile with good performance.

- **Tableau**

- It is one of the important and most usable data visualization tool in big data environment [8]. The user can gain knowledge with their data, by creating charts, graphs, maps and many other graphics [10]. It helps non-programmers and business types to have deep insights and perfect data ingestion; they can also have fast exploration with interactivity, animation etc.



- **Ember Charts**

- The name indicates that this tool is completely based on the Ember.js framework and uses D3.js under the hood. They are mostly used for visualize the data in the form time series, bar, pie and scatter charts. The user can provide more amounts of data and the availability is high, where they won't crash when they have bad data on it. N.

### **NVD3**

- The NVD3 runs top of D3.js, where they have re-usable charts and components. The main of using NVD3 is to provide the user an easily understandable chart which can be reusable and customize according to the need of the user. O

# TECHNIQUES IN BIG DATA VISUALIZATION

- The existing visualization methods for analyzing data can be categorized based on several factors. According to user task or their requirements the visualization techniques are decided [13]. The visualization techniques include 1D, 2D, 3D, multidimensional, temporal, tree, and network. The user tasks comprises of history, detailed representation of data, general overview of data. Sometimes interaction/distortion
- **One Dimensional (1-D)**
  - The data set which comes under the 1 d consists of one variable and it has only a value per each data item. The histograms are used for carrying data visualizations for one dimensional data B.
- **Two Dimensional (2-D)**
  - Mostly two dimensional is used for visualize the data set, which contains two variables. It can be done easily by knowing the relationship between two variables. The 2D visualizations can be represented in the form of line graphs, by comparing the relationship between two variables ad plotting can be done according to it. The 2d can also be represented in form of bar charts, area charts, pie charts, maps, scatter plots and stream line and arrow visualizations. C

- **Three Dimensional (3-D)**

- The 3 d representation of data will give more knowledge to the user, where they can easily find the merits and demerits of their business flow, study etc. It contains values in three dimensional spaces. it gives information in the form of slicing techniques, 3D bar charts, Iso-surface and realistic renderings.

- **Multi-Dimensional**

- The multi dimensional visualization gives the user a clear idea in different perspective. The different techniques used such as parallel coordinates, maps, scatterplot matrices, auto-glyphs.

- **Temporal Technique**

- It is a technique, where most of the data can be easily displayed and the temporal technique has the ability to display the data in many views such as timeline, time series and scatter plot.

- **Tree Map**

- It is also known as hierarchical model, where the data is nested in form of rectangle and it represents each branch of the tree[14]. The sub branch is represented as in form of smaller rectangles and leaf node is used for describing the specified dimension on the data. Sometimes the coloured leaf nodes are used to display a separate dimension of data. It also provides the user a proper display of data in a hierarchical manner.

# Business Impact Analysis (BIA)

- The analytics carried by the business impact helps to identify impact of critical and non-critical systems. The company can gain knowledge by processing the business analytics and they can take some precautionary steps at the time of disaster. This type analysis includes information about the estimated recovery times and recovery requirement, it also used for measuring the risks of failure against the costs of upgrading a particular system, find out the major risks and their fallout like losing all the data on the company main servers

# Big Data Analytics (BDA)

- The very important type of analytic in the present world, is known as big data analytics, which has the ability to analyze massive volume of data. It refers to the strategy of analyzing massive volumes of data, which is known big data. The origin of big data can be collected anywhere, such as social networking sites, the sensors used to collect climate data, videos and digital images, sale transaction records and cell phone GPS signals[6]. It helps to explore concealed patterns, unidentified connections and provide accurate and valuable information about the data. The main task of big data analytics is to assist organizations in making better business decisions. It also helps data scientists and other users to evaluate or process massive volumes of transaction data and other data sources, which might be not, exploited by traditional Business Intelligence (BI) systems
- This is because traditional systems often fail to analyze large data sources, which include Web server logs, social networking activity reports, Internet click-stream data and sensorcaptured data. The regular data warehouse cannot be suited to process big data. So they use some new technologies such as Hadoop, MapReduce and NoSQL databases. This technology is used for managing large volume of data in an efficient way for analytics purpose. The big data analytics helps to extract significant value from big data