International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

IMPORTANCE OF DATA ANALYSIS

Author: Jayshree R Kumbhare

Jayshree R Kumbhare M.C.A. Tilak Mahavidhyalay Pune Maharashtra

Ex prof.: Mrs.Shweta Nigam

Abstract - Statistics is used to study the collection, analysis and interpretation of data. The process can be illustrated as the collection of data, exploratory data analysis, the application of probability and random numbers to the data, followed by inference which relates the analysis of the subset of data collected to the entire set of data.

Data analysis is a new driver of the world economic and societal changes. The world's data collection is reaching a tipping point for major technological changes that can bring new ways in decision making, managing our health, cities, finance and education. While the data complexities are increasing including data's volume, variety, velocity and veracity, the real impact hinges on our ability to uncover the `value' in the data through Data Analytics technologies. Data Analytics poses a grand challenge on the design of highly scalable algorithms and systems to integrate the data and uncover large hidden values from datasets that are diverse, complex, and of a massive scale. Potential breakthroughs include new algorithms, methodologies, systems and applications in Data Analytics that discover useful and hidden knowledge from the Data efficiently and effectively.

Data analytics must also be team effort cutting across academic institutions, government and society and industry, and by researchers from multiple disciplines including computer science and engineering, health, data science and social and policy areas.

Data analyst analyzes and interprets data. They convert those data to useful information related to improving business. Thus, the data analyst's job is very crucial in the policy-making and decision-making level of an organization or business.

Key Words: Data Extraction, Data Collection, Data Cleaning, Data Processing.

1.INTRODUCTION

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

[1] *Analysis*, refers to dividing a whole into its separate components for individual examination. *Data analysis*, is a process for obtaining raw data, and subsequently converting it into information useful for decision-making by users. *Data*, is collected and analyzed to answer questions, test hypotheses, or disprove theories.

[2] Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data.

2. METHODOLOGY

The best methodology for Data analysis are Statistical analysis and text analysis. There are various examples where data analysis is used, ranging from transportation, risk and fraud detection, web search, digital advertisement and many more.

A Statistical analysis: Statistics involves data collection, interpretation, and validation. Statistical analysis is the technique of performing several statistical operations to quantify the data and apply statistical analysis. Quantitative data involves descriptive data like surveys and observational data. It is also called a descriptive analysis. It includes various tools to perform statistical data analysis such as SAS (Statistical Analysis System), SPSS (Statistical Package for the Social Sciences), Stat soft, and more. After collecting data you can **analyze** it to: Summarize the data.

Text analysis : Text analysis is a technique to analyze texts to extract machine-readable facts. It aims to create structured data out of free and unstructured content. The process consists of slicing and dicing heaps of unstructured, heterogeneous files into easy-to-read, manage and interpret data pieces. It is also known as text mining, text analytics, and information extraction.

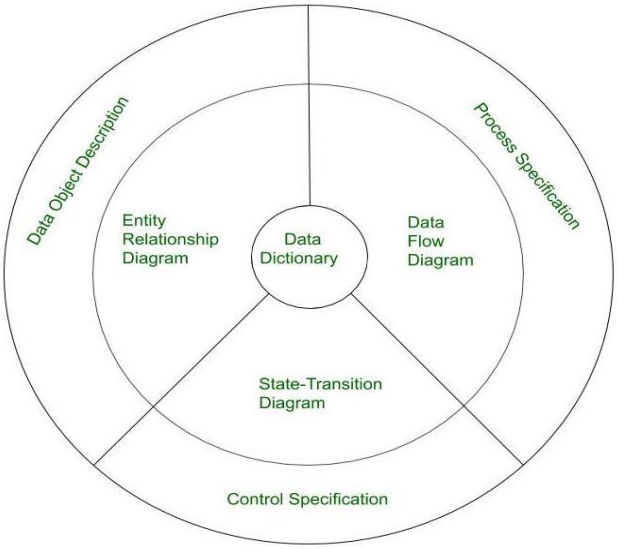
The ambiguity of human languages is the biggest challenge of text analysis. For example, humans know that "Red Sox Tames Bull" refers to a baseball match. Still, if this text is fed to a computer without background knowledge, it would generate several linguistically valid interpretations. Sometimes people who are not interested in baseball might have trouble understanding it too.

3. MODELING AND ANALYSIS

Analysis Model is a technical representation of the system. It acts as a link between system description and design model. In Analysis Modelling, information, behavior and functions of the system is defined and translated into the architecture, component and interface level design in the design modeling.

Objectives of Analysis Modelling:

1. It must establish a way of creation of software design.
2. It must describe requirements of customer.
3. It must define set of requirements which can be validated, once the software is built.



Data Dictionary:

It is a repository that consists of description of all data objects used or produced by software. It stores the collection of data present in the software. It is a very crucial element of the analysis model. It acts as a centralized repository and also helps in modelling of data objects defined during software requirements.

Entity Relationship Diagram (ERD):

It depicts relationship between data objects and used in conducting of data modelling activity. The attributes of each object in the Entity Relationship Diagram can be described using Data object description. It provides the basis for activity related to data design.

Data Flow Diagram (DFD):

It depicts the functions that transform data flow and it also shows how data is transformed when moving from input to output. It provides the additional information which is used during the analysis of information domain and serves as a basis for the modeling of function. It also enables the engineer to develop models of functional and information domain at the same time.

State Transition Diagram:

It shows various modes of behavior (states) of the system and also shows the transitions from one state to other state in the system. It also provides the details of how system behaves due to the consequences of external events. It represents the behavior of a system by presenting its states and the events that cause the system to change state. It also describes what actions are taken due to the occurrence of a particular event.

Process Specification:

It stores the description of each functions present in the data flow diagram. It describes the input to a function, the algorithm that is applied for transformation of input, and the output that is produced. It also shows regulations and barriers imposed on the performance characteristics that are applicable to the process, and layout constraints that could influence the way in which the process will be implemented.

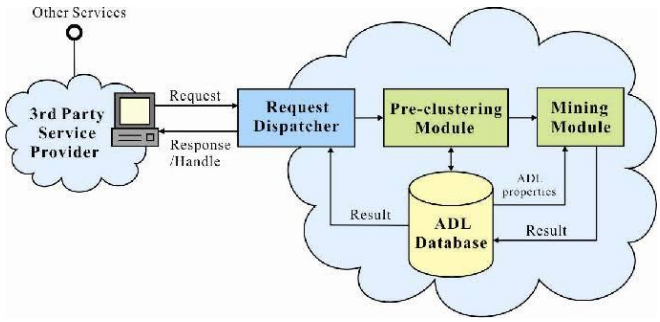


Fig: block representation of data analysis

Control Specification:

It stores the additional information about the control aspects of the software. It is used to indicate how the software behaves when an event occurs and which processes are invoked due to the occurrence of the event. It also provides the details of the processes which are executed to manage events.

Data Object Description:

It stores and provides the complete knowledge about a data object present and used in the software. It also gives us the details of attributes of the data object present in Entity Relationship Diagram. Hence, it incorporates all the data objects and its attributes.

Data Analysis Tools:

There are many data analysis tools you can get started with, depending on your technical skills, budget, and type of data you want to analyze. They're also a great way of speeding up manual and time-consuming data analysis, which is key if you need up-to-the-minute insights to make fast decisions.

Here's a quick rundown of the top data analysis tools that can help you perform everything from text analysis to data visualization

- **MonkeyLearn**: No-code machine learning platform that provides a full suite of text analysis tools and a robust API. Easily build custom machine learning models in a point and click interface.
- **KNIME**: Open-source platform for building advanced machine learning solutions, and visualizing data.
- **RapidMiner** For data analytics teams that want to tackle challenging tasks and handle large amounts of data.
- **Microsoft Excel**: Filter, organize, and visualize quantitative data. The perfect tool for performing simple data analysis. Explore common functions and formulas for data analysis in Excel.
- **Tableau**: A powerful analytics and data visualization platform. Connect all your data and create interactive dashboards that update in real-time.
- **R**: A free software environment for statistical computing and graphics. Learning R is relatively easy, even if you don't have a programming background.
- **Python**: The preferred programming language for machine learning. Use it to build data analysis solutions for various use cases.

Areas where data analysis application have been employeed:

1. Policing/ security.
2. Transportation.
3. Fraud and risk detection.
4. Delivery logistic.
5. City planning.
6. Health care.
7. Travel.
8. Internet/web search.
9. Energy management.
10. Digital advertisement.

This are the live application of data analysis which use in day to day life. Almost everybody within an organization can make use of data for increase of productivity and make very important decision. Use of data would have a positive impact on business and even the society in general.

CONCLUSIONS:

In this paper we proposed the data analysis and its classification system. We have to conclude that the data and conclusions are both key elements of a scientific research process. In carrying out a study or experiment, data is the result collected from testing. Conclusions are your interpretation of data.

ACKNOWLEDGEMENT:

I would like to express my sincere gratitude to my advisor Prof & H.O.D. of M.C.A. Department Mrs. Shweta Nigam for the continuous support of my MCA study & research for her patience, motivation , enthusiasm , and immense knowledge .Her guidance helped me in all the time of research and writing of this paper. I could not have imagined having a better advisor and mentor for my MCA study.

REFERENCES

1. **https://en.wikipedia.org/wiki/data analysis**
2. Data Analytics Made Accessible: 2020 Edition
   a. By Anil K. Maheshwari, Ph.D.
3. **Data Coding and Exploratory Analysis (EDA)**
   Rules for Data Coding Exploratory Data Analysis
   (EDA) Statistical Assumptions
   ByNancy Leech, Karen Barrett, George A Morgan
4. John Tukey-The Future of Data Analysis-July 1961
5. *"Data.Gov:Long-Term Pavement Performance (LTPP)"*.

   May 26, 2016. Retrieved November 10, 2017.