IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 56-59

Salient Schemes for Data Analysis

Chandana Nighut[1], Samira Nigrel[2], Sushma Amudalapally[3], Nida Parkar[4]
[1](Department of Computer Engineering, Atharva College of Engineering, Mumbai, India)
[2](Department of Information & Technology, Atharva College of Engineering, Mumbai, India)
[3](Department of Electronics & Telecommunication, Atharva College of Engineering, Mumbai, India)
[4](Department of Computer Engineering, Atharva College of Engineering, Mumbai, India)

Abstract: *In the digitalisation era, the data is no longer scanty, its robust. Hence to enable an Organisation or a firm to access the data skillfully and easily, the data needs to besifted and sorted. Data analysis is the process of evaluating data usinglogical & analytical reasoning to examine each component of the data provided. This paper presents various schemes/ methods to analyze the widespread range of data.*
Keywords: *Data analysis, logical & analytical reasoning*

I. 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. It is a method to overcome huge load of powerful data for on Organization by using easy handy techniques. This paper makes an attempt to depict the importance of analyzing data. It also sparks a light on the various types of data analysis that help us sort and scrutinize data; and also the techniques to go through the process.

II. Importance

Every organization makes attempts to gather data, for instance, by monitoring its competitors' performance, sales figures, and buying trends etc. in an effort to be more competitive. However, nobody can understand customers' behaviors and competitors' performance without the skills to analyze all that data.

Data analysis, therefore, is a necessity for making well-informed and efficient decisions. Data analysis is what helps organizations determine their positions in the market relative to competitors. It is what helps us identify the potential risks that need to be avoided and the opportunities that must be grabbed in order to grow.
It is, in fact, data analysis that enables us to gauge the satisfaction level of the customers and their needs in order to come up with new products and services that provide greater satisfaction to them. Therefore, it is an understatement to say that data analysis is important for the success of businesses.

III. Types

1. Descriptive (least amount of effort): The discipline of quantitatively describing the main features of a collection of data. It describes a set of data.
It is typically the first kind of data analysis performed on a data set. Commonly it is applied to large volumes of data, such as census data.

Figure No. 01: Example of Descriptive Data Analysis

2. Exploratory: An approach to analyzing data sets to find previously unknown relationships. Exploratory models are good for discovering new connections. They are also useful for defining future studies/questions. Exploratory analyses are usually not the definitive answer to the question at hand, but only the start, which can be applied to Census and Convenience Sample Data Set.

Figure No. 02: Example of Exploratory Data Analysis

3. Inferential: Aims to test theories about the nature of the world in general (or some part of it) based on samples of "subjects" taken from the world (or some part of it). That is, use a relatively small sample of data to say something about a bigger population. Inference is commonly the goal of statistical models
Inference involves estimating both the quantity you care about and your uncertainty about your estimate
Inference depends heavily on both the population and the sampling scheme
It can be applied to Observational, Cross Sectional Time Study, and Retrospective Data Set – the right, randomly sampled population.

Figure No. 03: Example of Inferential Data Analysis

4. Predictive: The various types of methods that analyze current and historical facts to make predictions about future events. The models predicts, but it does not mean that the independent variables cause. Accurate prediction depends heavily on measuring the right variables. It can be applied to Prediction Study Data Set – a training and test data set from the same population.

Figure No. 04: Example of Predictive Data Analysis

5. Causal: It is used to find out what happens to one variable when you change another.
Implementation usually requires randomized studies. There are approaches to inferring causation in non-randomized studies
Causal models are said to be the "gold standard" for data analysis. It can be applied to Randomized Trial Data Set – data from a randomized study.

Figure No. 05: Example of Casual Data Analysis

6. Mechanistic (most amount of effort): It understands the exact changes in variables that lead to changes in other variables for individual objects. It is usually modeled by a deterministic set of equations (physical/engineering science).Generally, the random component of the data is measurement error.
If the equations are known but the parameters are not, they may be inferred with data analysis
This type of data set applied to Randomized Trial Data Set – data about all components of the system.

Figure No. 06: Example of Mechanistic Data Analysis

IV. Applications

1. Smart Buildings- A data analysis approach can be used in order to predict energy consumption in buildings. The different steps of the data analysis process are carried out in order to realize smart buildings, where the building management and control operations including heating, ventilation, air conditioning, lighting and security are realized automatically by miming the needs of the building users and optimizing resources like energy and time.

2. Analytics & Business Intelligence- Analytics is the "extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions." It is a subset of business intelligence, which is a set of technologies and processes that use data to understand and analyze business performance.

3. Education- In education, most educators have access to a data system for the purpose of analyzing student data. These data systems present data to educators in an over-the-counter data format (embedding labels, supplemental documentation, and a help system and making key package/display and content decisions) to improve the accuracy of educators' data analyses.

4. Statistical Modeling- Statistical modeling involves building predictive models based on pattern recognition and knowledge discovery. It is used in environmental and geographical studies, predicting election outcomes, survival analysis of populations, and more. Meteorologists use statistical tools to help them predict the weather. The line between statistical modeling and machine learning is becoming increasingly blurry — Robert Tibshirani, a statistician at Stanford called machine learning "glorified statistics".

5. Clinical Trial Design- One of the most important applications of statistical analysis is in designing clinical trials. When a new drug or treatment is discovered, it has to first be tested on group/groups of people to understand its efficacy and safety. A clinical trial involves selecting a population/sample size, defining the time range over which to monitor the treatment, designing the phases, and selecting parameters that will help decide how effective the treatment is and if it is better than an existing one. Biostatisticians can take on the task of performing a statistical analysis of the study, helping not only to design it but also analyze and determine the outcomes.

V. Conclusion

For any firm or an organization to function smoothly, the large amount of data that they have amassed needs to be analyzed skillfully; for the process of which the modernized technologies offer a varied range of methods and techniques for analyzing the data sets promptly. As presented in this paper, there are many types to analyze data, even on large scale. And the data analysis plays major role in diverse applications such as medical, educational and technological advancements.

References

[1]. Pierre Perron The great crash, the oil price shock, and the unit root hypothesis Econometrica, 57 (6) (1989), pp. 1361-1401
[2]. G.S. Maddala, I.M. Kim Unit Roots, Co integration, and Structural Change Cambridge University Press, Cambridge (2003)
[3]. R. Lumsidaine, D.H. Papel Multiple trend breaks and the unit root hypothesis Rev Econ Stat, 79 (1997), pp. 212-218
[4]. S. Johansen, K. Juselius Maximum likelihood estimation and inference on co integration- with applications to the demand for money Oxf Bull Econ Stat, 52 (1990), pp. 169-210
[5]. Irving Fisher The Purchasing Power of Money-its Determination and Relation to Credit, Interest, and Crises The Macmillan Co, New York (1922)