

ASSIGNMENT 1

1.WHAT IS THE DIFFERENCE BETWEEN DATA ANALYSIS AND MACHINE LEARNING?

DATA ANALYSIS:

In Data Analysis we use the basic statistical calculations, database manipulation & transformation, visualize data and conclude the data points by providing meaningful insights from provided dataset. Business takes significant decisions based on these insights.

APPLICATIONS:

- Maintaining database systems
- Predicting sales, customer behavior, Insurance claims and loss amount using basic linear & logistic regression
- Creating homogeneous groups using classification and regression Trees(CART).
- Market Basket Analysis for any Grocery store like Big Bazaar.

MACHINE LEARNING:

Machine Learning is the advanced level of Data Analysis, here it comprises of study and application of algorithm to extract data and provide statistical inferences. Here the algorithms improve themselves from data experience without rely on specific programming. When data changes, the algorithm are designed in such a way appropriate action is enabled.

APPLICATIONS:

- Best route estimation on Uber.
- Movies & TV series recommended to customer based on customer behavior like on Netflix.
- Recommender system predicts relevant product to customer baed on historical experience and other relevant customer information like on Amazon.

2.WHAT IS BIG DATA?

Big data is a combination of structured, semistructured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications with reduced cost, enhanced efficiency, enhanced sales.

Systems that process and store big data have become a

common component of data management architectures in organizations, combined with tools that support big data analytics uses.

Big data is often characterized by the 5 V's

- Variety
- Volume
- Values
- Veracity
- Velocity

3. What the four main things we should know before studying data analysis?

- Statistical and Analytical Skills
- Programming Knowledge
- Communication skills
- Creative Thinking and Visualisation

4. Most common characteristics used in descriptive statistics?

Descriptive statistics can be useful for two purposes:

1) To provide basic information about variables in a

dataset

2) To highlight potential relationships between variables.

Common Characteristics of Descriptive Statistics are:

- Central Tendency
 - > Mean
 - > Median
 - > Mode
- Measure of Statistics
 - > Variance
 - > Standard Deviation
 - > Range
 - > Skew

5. What is Quantitative Data and Qualitative Data?

Quantitative Data:

Quantitative data are the result of counting or measuring attributes of a population. Quantitative data are always numbers.

Example:

- Amount of money you have
- Height
- Weight
- Number of people living in your town
- Number of students who take statistics

Qualitative Data:

Qualitative data are the result of categorizing or describing attributes of a population. Qualitative data are generally described by words or letters.

Examples:

- Hair color
- Blood type
- Ethnic group
- The car a person drives
- The street a person lives on