

## MACHINE LEARNING

- 1) b
- 2) d
- 3) b
- 4) a
- 5) b
- 6) a
- 7) a
- 8) b
- 9) No answer
- 10) A
- 11) D
- 12) A
- 13) With Elbow Method Cluster Analysis is calculated
- 14) Cluster Quality is measured through Silhouette Technique.
- 15) Cluster analysis is grouping of different set of data points into similar groups of data points.  
There are three types
  - 1) K-Means
  - 2) Hierarchical
  - 3) DBSCAN

## WORKSHEET 1 SQL

- 1)acd
- 2)abd
- 3)b
- 4)b
- 5)a
- 6)c
- 7)b
- 8)b
- 9)b
- 10)c
- 11)A data Warehouse is where large amount of data are stored. It is a repository where anyone can access the required data, analyze them and use it for beneficial purpose.
- 12)OLTP works online where it captures and stores data whereas OLAP is a historical data where Queries helps to analyze the data.
- 13) The Characteristics of Data Warehouse are
  1. Subject Oriented
  2. Time Variant
  3. Integrated
  4. Non Volatile

14)The Star Schema is Multidimensional database where it organize data in database so that it is easy to understand and analyze.

15)NO Answer(SETL)

#### STATISTICS WORKSHEET 1

1)a

2)a

3)b

4)

5)c

6)b

7)b

8)a

9)c

10) A normal distribution is a continuous data where most are clustered around centre of mean.

11)Missing data can be replaced by mean or mode data.

Simple Imputer, KNN Imputer and Iterative Imputer to replace Null values

12)No Answer

13) mean Imputation is not a acceptable practice.

14) Linear Regression is a Supervised machine learning program to explains the relationship between the independent and dependent variables.

15) The Differential and Inferential statistics are two branches of Statistics.