**MACHINE LEARNING**

1. (b) Market Trend Prediction
2. (d) None
3. (d) Reinforcement learning and unsupervised learning
4. (b) The tree representing how close the data points are to each other
5. (d) None
6. (c) k-nearest neighbour is same as k means
7. (d) 1,2 and 3
8. (a) 1 only
9. (a) 2
10. (c)
11. (a)
12. (b)

13.Importance of clustering:

Clustering is useful for exploring data. If there are many cases and no obvious groupings, clustering algorithms can be used to find natural groupings. Clustering can also serve as a useful data-preprocessing step to identify homogeneous groups on which to build supervised models.

14.Improve clustering performance:

 clustering algorithm can be significantly improved by using a better initialization technique, and by repeating (re-starting) the algorithm. When the data has overlapping clusters, k-means can improvethe results of the initialization technique.

**STATISTICS**

1.(b) Total variation =Residual variation+ Regression variation

2.(c) binary

3.(a) 2

4.(a) Type I error

5.(c) level of confidence

6.(b) increase

7.(b) Hypothesis

8.(d) All of the mentioned

9.(a) 0

10. Bayes’ Theorem:

Bayes Theorem provides a principled way for calculating a conditional probability. It is a deceptively simple calculation, although it can be used to easily calculate the conditional probability of events where intuition often fails .Although it is a powerful tool in the field of probability, Bayes Theorem is also widely used in the field of machine learning. Including its use in a probability framework for fitting a model to a training dataset, referred to as maximum a posteriori or MAP for short, and developing models for classification predictive modeling problems such as the Bayes Optimal Classifier and Naive Bayes.

11. Z- score :

A Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of [standard deviations](https://www.investopedia.com/terms/s/standarddeviation.asp) from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score. A Z-score of 1.0 would indicate a value that is one standard deviation from the mean. Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean.

12.T-test:

A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features. It is mostly used when the data sets, like the data set recorded as the outcome from flipping a coin 100 times, would follow a normal distribution and may have unknown variances. A t-test is used as a hypothesis testing tool, which allows testing of an assumption applicable to a population.

13.Percentile :

In  statistics, a percentile  is a score belowwhich a given percentage of scores in its frequency distribution falls or a score at or below which a given percentage falls. For example, the 50th percentile (the median) is the score below which or at or below which 50% of the scores in the distribution may be found.

14.ANOVA:

Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

15.How can ANOVA help?

An ANOVA test is a way to find out if survey or experiment results are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis. Basically, you're testing groups to see if there's a difference between them.