MACHINE LEARNING

1) b. Market trend prediction
2) d. None
3) a. Supervised 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) d. None of them
9) b. 4
10) b. Given a database of information about your users, automatically group them into different market segments.
11) a.
12) b.
13) Clustering is the task of grouping a set of objects so that objects in the same group are more similar to each other than to those in other groups.
14) A good clustering method will produce high quality clusters in which: – the intra-class (that is, intra intra-cluster) similarity is high - the inter-class similarity is low. The quality of a clustering result also depends on both the similarity measure used by the method and its implementation.

STATISTICS

1) b) Total Variation = Residual Variation + Regression Variation
2) c) binomial
3) a) 2
4) a) Type-I error
5) a) Power of the test
6) a) Decrease
7) b) Hypothesis
8) d) All of the mentioned
9) a) 0
10) Bayes' Theorem states that the conditional probability of an event, based on the occurrence of another event, is equal to the likelihood of the second event given the first event multiplied by the probability of the first event.
11) 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 from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score.

12) A t-test is a statistical test that compares the means of two samples. It is used in hypothesis testing, with a null hypothesis that the difference in group means is zero and an alternate hypothesis that the difference in group means is different from zero.
13) A percentile is a term that describes how a score compares to other scores from the same set.
14) A statistical method in which the variation in a set of observations is divided into distinct components.
15) ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources.