MACHINE LEARNING
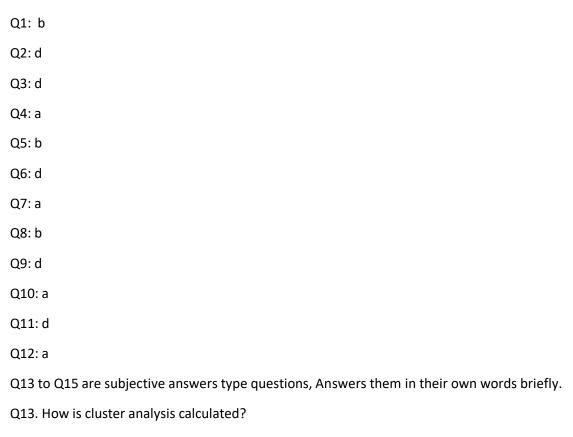
Q1 to Q12 have only one correct answer. Choose the correct option to answer your question

Answers:

Q1: b

Q2: d

Q3: d

Q4: a

Q5: b

Q6: d

Q7: a

Q8: b

Q9: d

Q10: a

Q11: d

Q12: a

Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly.

Q13. How is cluster analysis calculated?

Ans: *calculate the distance

   *Link the clusters

   *Choose a solution by selecting the right number of clusters

Q14. How is cluster quality measured?

Ans: Use the average Silhouette coefficient value of all objects in the dataset.

Q15. What is cluster analysis and its types?

Ans: The process to find similar groups of objects in order to form clusters. It is an unsupervised machine learning, based algorithm that acts on unlabed data .A group of data points would comprise together to form a cluster in which all objects would belong to the same group.
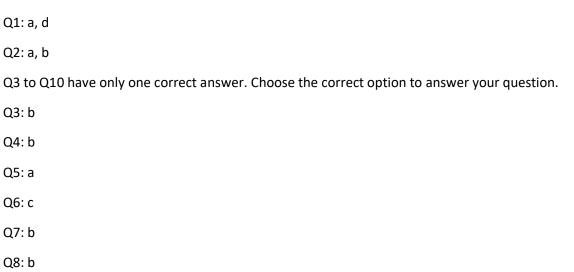
STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question

Q1: a

Q2: a

Q3: b

Q4: d

Q5: c

Q6: b

Q7: b

Q8: a

Q9: c

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

Q10. What do you understand by the term Normal Distribution?

Ans: Its a probability Continuous distribution. Means major data points gather to middle range.

Q11. How do you handle missing data? What imputation techniques do you recommend?

Ans: In cases of missing data, there is a missing. Observations, a Data scientist Can Calculate the mean (or) median.

* I recommend: Mean, Median & Mode method

Q12. What is A/B testing?

Ans: It is a way to compare two version of something. In that a version which gives better performance.

Q13. Is mean imputation of missing data acceptable practice?

Ans: Yes, Mean imputation of missing data acceptable practice.

Q14. What is linear regression in statistics?

Ans: Practice Linear regression is a Linear approach of modelling between Variable and scalar response in statistic.

Q15. What are the various branches of statistics?

Ans: Descriptive statistics, inferential statistics, Data collection are the three branches of statistics

WORKSHEET 1 SQL

Q1 and Q2 have one or more correct answer. Choose all the correct option to answer your question.

Q1: a, d

Q2: a, b

Q3 to Q10 have only one correct answer. Choose the correct option to answer your question.

Q3: b

Q4: b

Q5: a

Q6: c

Q7: b

Q8: b

Q9: d

Q10: b

Q11 to Q15 are subjective answer type questions, Answer them briefly

Q11. What is data-warehouse?

Ans: An insight engine for customers to store search and analyze large volume of data.

Q12. What is the difference between OLTP VS OLAP?

Ans: Online transaction processing (OLTP) capture, storeand processing data from transactions in real time. Online analytical processing (OLAP) uses complex quarries to analyse aggregated historical data from OLTP system.

Q13. What are the various characteristics of data-warehouse?

Ans: They are also called features of a data ware house, include SUBJECT ORIENTED, TIME VARIENT , INTEGRATED and NON VOLATILE

Q14. What is Star-Schema??

Ans: A star scheme is a data base organizational structure optimized for use in data work house (or) business intelligence that uses in a single large fact table to store translation (or) measured data, and one (or) more smaller dimensional tables that store attributes about the data.

Q15. What do you mean by SETL?

Ans: SET   language.   It   is   a   very   high   level   programming   language   based   on   the mathematical theory of sets.