DATA SCIENCE

SAMPLE QUESTION BANK

T.Y.B.Sc(CS) SEM-VI

1.Data Science is a field that ____
   A. extracts meaningful insights from data
   B. used to make decisions and predictions
   C. uses analytics and machine learning algorithms
   D. All of the above
2. What are the two types of data, with regard to Data Science?
   A. Traditional and Modern
   B. Small and Big data
   C. Qualitative and Quantitative
   D. None of the above
3. Qualitative data is also called as_____data.
   A. Continuous
   B. Numerical
   C. Categorical
   D. Discrete
4. Quantitative data is also called as_____data.
   A. Ordinal
   B. Numerical
   C. Categorical
   D. Nominal
5. Student name is_____type data.
   A. Ordinal
   B. Nominal
   C. Binary
   D. Numerical
6. Result of a coin toss is_____type data.
   A. Numerical
   B. Ordinal
   C. Nominal
   D. Binary
7. _____is the organization, publication and presentation of data for reuse.
   A. Data Curation
   B. Data smoothing
   C. Data transformation
   D. Data wrangling
8. _____is a web-based version-control and collaboration platform for software developers.
   A. XML
   B. GitHub
   C. AWS
   D. HBase
9. In MongoDB, data is written in_____like format.
   A. JSON
   B. XML

C. HTML
D. C

10. _____data has each variable saved in its own column.
    A. Structured
    B. Unstructured
    C. Web
    D. Unorganized
11. _____is the on-going management of data through its lifecycle.
    A. Data transformation
    B. Data Curation
    C. Data visualization
    D. Data cleaning
12. _____is an example of a NoSQL.
    A. MySQL
    B. ORACLE
    C. MongoDB
    D. Excel
13. learning does not have a response variable.
    A. Supervised
    B. Unsupervised
    C. Classification
    D. Regression
14. In_____learning, for each observation of predictor variables, there is an associated response variable.
    A. Supervised
    B. Unsupervised
    C. Reinforced
    D. None of the above
15. _____is an unsupervised learning method.
    A. Linear regression
    B. Logistic regression
    C. Cluster Analysis
    D. SVM
16. _____is a supervised learning method.
    A. K-means clustering
    B. Principal Component Analysis
    C. Hierarchical clustering
    D. SVM
17. Problems with a quantitative response are called ____ problems.
    A. Classification
    B. Regression
    C. Clustering
    D. Dimension reduction
18. Problems with a qualitative response are often called ____ problems.
    A. Classification
    B. Regression
    C. Time series
    D. None of the above
19. _____are techniques used to reduce error by avoiding overfitting.
    A. Smoothing
    B. Aggregation
    C. Cross validation

D. Regularization

20. RSS stands for _____
    A. Repeated sum of scores
    B. Residual sum of squares
    C. Retrieved sum of scales
    D. None of the above
21. When the model tries to learn or fit every noise data point in the data set, it results in_____.
    A. Overfitting
    B. Underfitting
    C. Smoothing
    D. Aggregation
22. Overfitting can be controlled by _____
    A. Regularization
    B. Increasing the size of training dataset
    C. Both A and B
    D. Decreasing the size of training dataset
23. _____causes underfitting.
    A. Bias
    B. Variance
    C. Regularization
    D. None of the above