



Avinashilingam Institute for Home Science and Higher Education for Women

(Deemed to be University under Category 'A' by MHRD, Estd,u/s 3 of UGC Act 1956)

Re-accredited with 'A' Grade by NAAC. Recognised by UGC Under Section 12B

Coimbatore- 641043, Tamil Nadu, India

**Bachelor of Vocation Degree Examination – January 2021
III Semester**

Class : II B.Voc

Time: 3 Hours

Major : Artificial Intelligence and Machine Learning

Max. Marks: 100

19VAIC09 Data Mining and Analytics

Part A

10 x 1 = 10

Choose the Correct Answer

1. What is true about data mining?
 - a. Data Mining is defined as the procedure of extracting information from huge sets of data
 - b. Data Mining also involves other processes such as Data Cleaning. Data Integration. Data Transformation.
 - c. Data Mining is the procedure of mining knowledge from data.
 - d. All the above
2. _____ is the essential process where intelligent methods are applied to extract data patterns.
 - a. Data warehousing
 - b. Data Mining
 - c. Text mining
 - d. Data selection
3. The mapping or classification of a class with some predefined group or class is known as?
 - a. Data Characterization
 - b. Data Discrimination
 - c. Data Set
 - d. Data Sub Structure
4. Cluster is _____.
 - a. Group on training data set to transform or simplify data in order to prepare it for a machine learning algorithm.
 - b. Group of similar objects that differ significantly from other objects.
 - c. Symbolic representation of facts or ideas from which information can potentially be extracted.
 - d. Both a and b.
5. _____ is a supervised learning algorithm which is used for both classifications as well as regressing.
 - a. Random forest
 - b. Decision Tree
 - c. K-Means
 - d. All the above
6. _____ are supervised learning methods used for classification and regression tasks that originated from statistical learning theory.
 - a. Decision Tree
 - b. Support Vector Machines
 - c. K-Means
 - d. All the above
7. _____ is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure.
 - a. K nearest neighbours
 - b. K-Means
 - c. SVM
 - d. All the above
8. _____ refers to the initial step in data analysis in which data analysts use data visualization and statistical techniques to describe dataset characterizations.
 - a. Data Management
 - b. Big Data
 - c. Data exploration
 - d. All the above
9. _____ are a data-visualization technique for large, hierarchical data sets.
 - a. Tree maps
 - b. Bar Chart
 - c. Box Plot
 - d. None of the above
10. Using _____ can highlight the most correlated variables in dataset.
 - a. corrgrams
 - b. line chart
 - c. pie chart
 - d. All the above

Part B

5 x 6 = 30

Answer ALL following

Each answer should not exceed 400 words or two pages

- 11.a. i) Differentiate between Analysis vs. Reporting.
ii) Give an account on Data Analysis.
(or)
- 11.b. Differentiate between Supervised and Unsupervised Learning.
- 12.a. Briefly outline the key features of various clustering methods with an example.
(or)
- 12.b. Explain constraint based association rule mining with an example.
- 13.a. Why use Random Forest? How does Random Forest algorithm work?
(or)
- 13.b. List the advantages of Support Vector Machines.
- 14.a. Explain the Data Reshaping through Melting and casting.
(or)
- 14.b. Write short notes on special functions across Data elements.
- 15.a. Give an account on Histogram visualization techniques.
(or)
- 15.b. Explain the Corrgrams visualization techniques with an example..

Part C

5 x 12 = 60

Answer ALL following

Each answer should not exceed 800 words or four pages

- 16.a. Explain the types of Data Analytics with an example.
(or)
- 16.b. Illustrate the Data Pre-Processing technique.
- 17.a. Consider the following dataset and find the frequent item sets and generate association rules for them.
- | TID | Items |
|------------|--------------|
| T1 | I1,I2,I5 |
| T2 | I2,I4 |
| T3 | I2,I3 |
| T4 | I1,I2,I4 |
| T5 | I1,I3 |
| T6 | I2,I3 |
| T7 | I1,I3 |
| T8 | I1,I2,I3,I5 |
| T9 | I1,I2,I3 |
- Minimum support count is 2
Minimum confidence is 60%
(or)
- 17.b. Why naïve Bayesian classification is called naïve? Briefly outline the major ideas of naïve Bayesian classification.
- 18.a. Discuss about the Support vector Machine classification algorithm with an example.
(or)
- 18.b. Explain in detail about K-Nearest Neighbors algorithm with an suitable example.
- 19.a. How to Import and Export of Data from database? Explain.
(or)
- 19.b. How to handle missing data management? Discuss.
- 20.a. Discuss about tree maps with an example.
(or)
- 20.b. We have a vector of maximum temperatures (in degree Celsius) for seven days as follows: c(22,27,26,24,23,26,28). Create any three types of charts Plot in R programming.

