

[Dashboard](#) ▶ [My courses](#) ▶ [Prof. Anshu Chaturvedi](#) ▶ [Data Mining and Warehousing 680313](#) ▶

[Mid Sem Examination II](#)

Started on Wednesday, 8 December 2021, 5:35 PM

State Finished

Completed on Wednesday, 8 December 2021, 5:41 PM

Time taken 5 mins 52 secs

Question 1

Complete

Marked out of 1.00

Which among the following is not an advantage of PAM? CO6

- ☒ a. Does scale well for large data sets
- ☐ b. Ordering of input does not impact results
- ☐ c. Handles outliers well
- ☐ d. Works efficiently for small data sets

Your answer is correct.

Question 2

Complete

Marked out of 1.00

BIRCH stands for----- CO1

- ☐ a. Balanced Iterative Reducing and Clustering using Hierarchies
- ☐ b. Balancing the Iterative Reducing and Clustering using Hierarchies
- ☒ c. Balanced Iterative Reduction and Clustering using Hierarchies
- ☐ d. Balanced Iteration Reducing and Clustering using Hierarchies

Your answer is incorrect.

Question 3

Complete

Marked out of 1.00

Which among the following statements is correct with respect to Data selection? CO2

- ☒ a. The stage of selecting the right data for a KDD process
- ☐ b. The actual discovery phase of a knowledge discovery process
- ☐ c. Record oriented classes finding
- ☐ d. A subject-oriented integrated time-variant non-volatile collection of data in support of management

Your answer is correct.

Question 4

Complete

Marked out of 1.00

Classification Problems can be solved by _____ CO2

- ☒ a. All of them
- ☐ b. specifying boundaries
- ☐ c. using posterior probabilities
- ☐ d. using probability distribution

Your answer is correct.

Question 5

Complete

Marked out of 1.00

Among the given statements which is correct about data mining? CO3

- ☐ a. It can be referred to as the procedure of mining knowledge from data
- ☐ b. Data mining can be defined as the procedure of extracting information from a set of the data
- ☒ c. All of the Others
- ☐ d. The procedure of data mining also involves several other processes like data cleaning, data transformation, and data integration

Your answer is correct.

Question 6

Complete

Marked out of 1.00

Among the given following what are the functions of Data Mining? CO3

- ☐ a. Cluster analysis and Evolution analysis
- ☐ b. Association and correctional analysis classification
- ☒ c. All of the Others
- ☐ d. Prediction and characterization

Your answer is correct.

Question 7

Complete

Marked out of 1.00

DMQL stands for_____ CO3

- ☐ a. Data Marts Query Language
- ☐ b. Database Miner Query Language
- ☒ c. Data Mining Query Language
- ☐ d. Dataset Mining Query Language

Your answer is correct.

Question 8

Complete

Marked out of 1.00

Among the given sequences below which one is the right approach of Data Mining? CO3

- ☒ a. Infrastructure, exploration, analysis, exploitation, interpretation
- ☐ b. Infrastructure, analysis, exploration, exploitation, interpretation
- ☐ c. Infrastructure, exploration, analysis, interpretation, exploitation
- ☐ d. Infrastructure, analysis, exploration, interpretation, exploitation

Your answer is incorrect.

Question 9

Complete

Marked out of 1.00

State True or False

Bayes Theorem attempts to find a model that explains the observed data by first creating a hypothesis and then testing hypothesis against the data. CO3

Select one:

- ☐ True
- ☒ False

Question 10

Complete

Marked out of 1.00

Which of the following is not an application of Classification? CO4

- ☒ a. Loan Retrieval
- ☐ b. Classifying financial market trends
- ☐ c. Loan Approval
- ☐ d. Medical Diagnosis

Your answer is correct.

Question 11

Complete

Marked out of 1.00

_____ link is largest distance between an element in one cluster and an element in the other cluster is known CO4

- ☐ a. Diameter
- ☒ b. Average
- ☐ c. Complete
- ☐ d. Single

Your answer is incorrect.

Question 12

Complete

Marked out of 1.00

Which among the following is not a Partitional Algorithm? CO4

- ☒ a. BA
- ☐ b. Nearest Neighbor
- ☐ c. PAM
- ☐ d. MST

Your answer is correct.

Question 13

Complete

Marked out of 1.00

Which among the following is a disadvantage of Partitioning Algorithm? CO4

- ☐ a. May have many candidates during second scan.
- ☐ b. Adapts to available main memory
- ☐ c. Easily parallelized
- ☒ d. Maximum number of database scans is two.

Your answer is incorrect.

Question 14

Complete

Marked out of 1.00

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

With which of the following is this equation associated with? CO5

- ☒ a. Chi Squared Statistic
- ☐ b. None of the other
- ☐ c. Hypothesis Testing
- ☐ d. Bayes Theorem

Your answer is correct.

Question 15

Complete

Marked out of 1.00

Which one of the following is not a disadvantage of Bayesian Classification? CO5

- ☐ a. Does not handle continuous data
- ☐ b. It does not always yield good results
- ☐ c. Attributes usually are not independent
- ☒ d. Cannot handle missing values easily.

Your answer is correct.

Question 16

Complete

Marked out of 1.00

Which among the following is not a step in K-means? CO5

- ☐ a. High degree of similarity among elements in a cluster is obtained.
- ☒ b. Initial set of clusters purposely chosen.
- ☐ c. Iteratively, items are moved among sets of clusters until the desired set is reached.
- ☐ d. High degree of dissimilarity among elements in different clusters is also achieved.

Your answer is correct.

Question 17

Complete

Marked out of 1.00

In Nearest Neighbour, complexity depends on the number of _____ CO5

- ☐ a. clusters
- ☐ b. links
- ☒ c. items
- ☐ d. iterations

Your answer is correct.

Question 18

Complete

Marked out of 1.00

Is it possible that between successive iterations in K-Means, the assignment of observations to clusters does not change? CO5

- ☐ a. None of these
- ☒ b. Yes
- ☐ c. Can't Say
- ☐ d. No

Your answer is correct.

Question 19

Complete

Marked out of 1.00

Neural Network is actually an information processing system that consist of a graph representing the processing system as well as various algorithms that access that graph. CO6

Select one:

- ☐ True
- ☒ False

Question 20

Complete

Marked out of 1.00

In order to improve the accuracy of Linear Regression model (Supervised Learning), we can use Clustering (Unsupervised Learning) by creating----

1. different models for different cluster groups.
2. an input feature for cluster ids as an ordinal variable.
3. an input feature for cluster centroids as a continuous variable.
4. an input feature for cluster size as a continuous variable. CO6

☒ a. 1, 2, 3 and 4

☐ b. 1 and 4

☐ c. 3 only

☐ d. 1 only

☐ e. 1 and 2

☐ f. 2 and 4

Your answer is correct.

Previous activity

◀ Assignment VI

Jump to...

Next activity

|| Faculty Feedback ▶

Stay in touch

Madhav institute of Technology & Science , Gwalior, M.P.

🌐 <http://mitsgwalior.in/>

☎ [Phone :0751-2409300](tel:0751-2409300)

✉ pro@mitsgwalior.in



📱 Get the mobile app