26/11/24

C

Avinashilingam Institute for Home Science and Higher Education for Women
(Deemed to be University Estd. u/s 3 of UGC Act 1956, Category 'A' by MHRD)
Re-accredited with 'A++' Grade by NAAC. Recognised by UGC Under Section 12B
Coimbatore - 641 043, Tamil Nadu, India

Bachelor's Degree Examination – November 2024
V Semester

Class : III UG
Major : Computer Science

Time: 3 Hours
Max. Marks: 100

21BCSC21 Fundamentals of Data Science

Course Outcomes:

At the end of the course, students will
CO1.Students will be able to apply the basic Data science knowledge on the day problems they encounter.
CO2.Students will realize that there are various phases that contribute to the completion of a Data Science project and can select among the various modelling techniques.
CO3.Students will be able to apply Regression techniques for modelling a Data science project.
CO4.Students will be able to apply the Clustering and Association rule mining for modelling a Data science project.
CO5.Students can reproduce the knowledge gained and come with a sample case study which they come across

Part A

10 x 1 = 10

Choose the Correct Answer

1.Which of the following is the most important language for Data Science?
a. Java b. Ruby c. R d. PHP CO1 K1

2. Which of the following testing is concerned with making decisions using data?
a. Hypothesis b. Probability c. Causal d. Dynamic CO1 K1

3. Data that summarize all observations in a category are called _____ data.
a. frequency b. summarized c. raw d. processed CO2 K2

4. Which of the following object has a method cov to compute covariance between series?
a. Series b. DataFrame c. Panel d. Lists CO2 K1

5. How many variables are required to represent a linear regression model?
a. 3 b. 2 c. 1 d. 4 CO3 K2

6. What is the goal of gradient descent?
a. Reduce complexity b. Reduce overfitting
c. Maximize cost function d. Minimize cost function CO3 K1

7. Which of the following is not a supervised machine learning algorithm?
a. K-means b. Naive Bayes
c. SVM d. Decision tree CO4 K2

8. Which algorithm is best suited for a binary classification problem?
a. Decision Trees b. KNN
c. Random Forest d. Linear Regression CO4 K2

9. How do Quality Leaders interact with people rather than directing and supervising them?
a. Scold b. Train and coach
c. Fire d. Threaten CO5 K1

10. Which of the following is a dimension of 'product quality'?
a. Probability b. Machinability
c. Durability d. Satisfaction CO5 K1

Part B 5 x 6 = 30

Answer ALL questions

Each answer should not exceed 400 words or two pages

11.a.	Explain the data science processes.	CO1 K2
	(or)	
11.b.	Discuss about model evaluation with example.	CO1 K2
12.a.	Summarize the mapping problems.	CO2 K5
	(or)	
12.b.	Explain in general the two step classification model.	CO2 K4
13.a.	Illustrate about the linear regression model.	CO3 K4
	(or)	
13.b.	Explain with an example, the use of a prediction model.	CO3 K3
14.a.	Describe the cluster analysis in unsupervised method.	CO4 K2
	(or)	
14.b.	Categorize the association rules with example.	CO4 K4
15.a.	Analyse the documentation and deployment of a data science project.	CO5 K6
	(or)	
15.b.	Distinguish the version control for running documentation.	CO5 K4

Part C 5 x 12 = 60

Answer ALL questions

Each answer should not exceed 800 words or four pages

16.a.	Analyze about loading data into R. Explain with example.	CO1 K4
	(or)	
16.b.	Summarize graphics and visualization into R.	CO1 K2
17.a.	Classify the ranking models with example.	CO2 K2
	(or)	
17.b.	Explain about ensuring the model quality.	CO2 K4
18.a.	Outline about Evaluating scoring models.	CO3 K4
	(or)	
18.b.	Examine the finding relations and extracting advice with example.	CO3 K3
19.a.	Illustrate about K-Means algorithms with example.	CO4 K4
	(or)	
19.b.	Summarize the hierarchical clustering methods.	CO4 K5
20.a.	Describe about the buzz dataset with example.	CO5 K2
	(or)	
20.b.	Demonstrate about producing effective presentation.	CO5 K3

****