School of Computing Science and Engineering
B.Tech CSE
ETE - Jun 2023

Time : 3 Hours

Marks : 100

Sem IV - E2UH401B - Data Analytics
Your answer should be specific to the question asked
*Draw neat labeled diagrams wherever necessary*

1.	What is Array Broadcasting? Explain iwith an example. Discuss ts utility in Data Analytics.	K1 CO1 (5)
2.	Write a Python Code to draw a histogram, bar chart, and pie chart, for the following dataset.	K2 CO2 (5)
3.	Discuss the Data Architecture Design through the help of a suitable diagram. Explain all the functionalities of three layers.	K3 CO3 (5)
4.	Explain in detail about Data Analytics and its 6 steps that are utilized for data analysis tasks. Your explanation should include a suitable example and the steps of data analysis process must be applied on the example.	K2 CO2 (10)
5.	Elaborate on unsupervised learning mechanisms through clustering and association rule mining. Discuss any one clustering algorithm with Python code and demonstrate it working with suitable example. Discuss any one association rule mining algorithm with Python code and demonstrate it working with suitable example.	K4 CO4 (10)

OR

	Discuss any five Python libraries that you have used for data analytics, Each library must include an explanation and code example of any two functions in the library.	K4 CO4 (10)
6.	Explain the working of K Nearest Neighbours (KNN) Classification. Write its algorithm. Write Python code for classifying the Iris dataset using KNN using sklearn module for K ranging from 3-10. Through an accuracy versus K graph, determine the best value of K.	K4 CO4 (10)
7.	There are four types of Data Analytics. Discuss each of them with the help of a supermarket example, where things are sold on a daily basis and the sales are recorded in a dataset.	K1 CO1 (10)
8.	What is density-based clustering? Explain its procedure and explain the DBSCAN algorithm. If Epsilon is 2 and minpoint is 2, what are the clusters that DBScan would discover with the following 8 examples: A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9). Compute the distance matrix and draw the 10 by 10 space to illustrate the discovered clusters. What clusters would be formed if Epsilon is increased to 3.3?	K4 CO4 (15)

OR

Explain clustering and its utility for data analytics. With suitable examples differentiate density-based, distance-based, and hierarchical clustering methods . Demonstrate the k-means algorithm, DBSCAN algorithm and Agglomerative clustering algorithm through figures and examples. Also draw the dendrograms drawn by Agglomerative clustering algorithm for the two cluster structures shown next:

K4 CO4 (15)
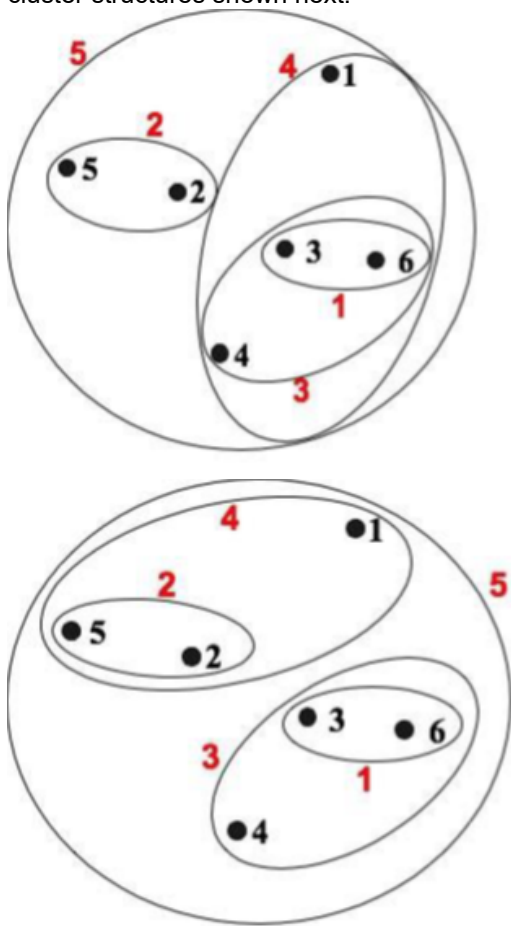
9. Using Python language, write a program that asks the user to enter the details, such as admission_number, name, marks_subject1, marks_subject2, marks_subject3, and final_grade of 10 students. Make a dataframe for the entries done and predict students' final grades using Gaussian Naive Bayes, Support Vector Machine, and Decision Tree classifiers, separately. Show a sample output for one execution of the code written.

K3 CO3 (15)

10. Consider the following dataset having two missing values.

K3 CO3 (15)

A	B	C
Job Position	Years of Experience	Salary (in USD per year)
CEO	5	100000
Senior Manager	4	80000
Junior Manager	3	
Employee		40000
Assistant Staff	1	20000

Using Pandas, write Python code to create dataframe for the above dataset and perform the following tasks:
a) Python code for removing the missing data rows from the dataset.
b) Python code for filling in missing values through statistical imputation.
c) Python code for filling in missing value through linear regression using sklearn.
Show sample output for one execution of each code.