|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |



# PRESIDENCY UNIVERSITY BENGALURU

**School Of Computer Science and Engineering & Information Science**

**SUMMER TERM END TERM EXAMINATION AUGUST 2024**

|  |  |
| --- | --- |
| **Semester :** winter Semester | **Date :** 06-08-2024 |
| **Course Code :** CSA2007 | **Time :** 9:30 AM -12:30 PM |
| **Course Name :** Data Mining | **Max Marks :**100 |
| **Program** : BCA | **Weightage :** 50% |

Note: 1. Answer ALL 5 FULL Questions.

1. **Each Full Question carries 20 Marks**
2. **Scientific and non-programmable calculator are permitted.**
3. **Do not write any information on the question paper other than Roll Number.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.a. | Discuss how data mining aids in uncovering hidden patterns, trends, and relationships to support decision-making processes. **[Knowledge]** | | | | | | | **(CO1)** | **(04 Marks)** |
| 1.b. | Elucidate any five applications of example. **[Comprehension]** | | data | mining in | detail | with | suitable | **(CO1)** | **(06 Marks)** |
| 1.c. | How data mining is used in the healthcare industry. Explain with suitable example **[Application]** | | | | | | | **(CO1)** | **(10 Marks)** |
|  | **or** | | | | | | |  |  |
| 2.a. | Describe three challenges to data mining regarding data mining methodology and user interaction issues.**[Knowledge]** | | | | | | | **(CO1)** | **(04 Marks)** |
| 2.b. | How does classification  **[Comprehension]** | differ | from | clustering | in | data | mining? | **(CO1)** | **(06 Marks)** |
| 2.c. | Illustrate the steps involved in knowledge discovery process with suitable diagram. **[Application]** | | | | | | | **(CO1)** | **(10 Marks)** |
| 3.a. | How to convert the nominal attribute to Binary attribute with suitable example.  **[Knowledge]** | | | | | | | **(CO2)** | **(04 Marks)** |
| 3.b. | How dimensionality reduction is applied in pre-processing and list out its benefits **[Comprehension]** | | | | | | | **(CO2)** | **(06 Marks)** |
| 3.c. | Normalize the dataset using min-max and decimal point techniques with the custom range [-1,+1] for the following data set W=[150,250,350,450,550] **[Application]** | | | | | | | **(CO2)** | **(10 Marks)** |
|  | **Or** | | | | | | |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| 4.a. | Illustrate the term “market Basket analysis” with example **[Knowledge]** | **(CO2)** | **(04 Marks)** |
| 4.b. | Distinguish the Sampling with replacement and without replacement with example **[Comprehension]** | **(CO2)** | **(06 Marks)** |
| 4.c. | Apply smoothing by bin means, smoothing by bin medians , smoothing by bin boundaries for the following Data Points: 3, 6, 9, 12, 14, 17, 19, 22, 25,  28, 32, 35 [Bin Size: 4] **[Application]** | **(CO2)** | **(10 Marks)** |
| 5.a. | Explain about Nominal Attribute with example **[Knowledge]** | **(CO3)** | **(04 Marks)** |
| 5.b. | Compare the efficiency of the FP-Growth Algorithm with the Apriori Algorithm. **[Comprehension]** | **(CO3)** | **(06 Marks)** |
| 5.c. | A database consists of nine transactions taken from the Electronic store. Enumerate all the frequent itemset using Apriori algorithm with minimum support threshold S=2. **[Application]** | **(CO3)** | **(10 Marks)** |
|  | **or** |  |  |
| 6.a. | Explain the primary motivation for finding frequent patterns in data mining.**[Knowledge]** | **(CO3)** | **(04 Marks)** |
| 6.b. | Why is it important to generate frequent item sets and association rules efficiently in large datasets? **[Comprehension]** | **(CO3)** | **(06 Marks)** |
| 6.c. | Analyse transaction data to identify associations between different items purchased by customers by discovering frequent item sets and association rules on the supermarket dataset to given. [min\_support=4 confidence=70%]. Dataset: [**Application]** | **(CO3)** | **(10 Marks)** |

|  |  |  |  |
| --- | --- | --- | --- |
| 7.a | Describe the supervised and unsupervised learning with appropriate example **[Knowledge]** | **(CO4)** | **(04 Marks)** |
| 7.b. | Explain k-means clustering algorithm **[Comprehension]** | **(CO4)** | **(06 Marks)** |
| 7.c | Apply K-Means algorithm on the given datasets to form 2 clusters.  **[Application]** | **(CO4)** | **(10 Marks)** |
|  | **or** |  |  |
| 8.a | Write short note on Bayesian Classification? **[Knowledge]** | **(CO4)** | **(04 Marks)** |
| 8.b. | Write a brief note on Density based clustering method **[Comprehension]** | **(CO4)** | **(06 Marks)** |
| 8c. | Perform the KNN classification algorithm on the following dataset and predict the class for the given sample X = {height=170, weight =57} k=5. **[Application]**  Height (CM) Weight (KG) Class  167 51 underweight  182 62 Normal  176 69 Normal  173 64 Normal  172 65 Normal  174 56 Underweight  169 58 Normal  173 57 Normal  170 55 Normal  170 57 ? |  |  |
| 9.a | What is anomaly detection in data mining? why is it important?  **[Knowledge]** | **(CO5)** | **(04 Marks)** |
| 9.b | Outline the key steps involved in the web mining process. **[Comprehension]** | **(CO5)** | **(06 Marks)** |
| 9.c | Elaborate the main steps involved in the application of density-based outlier detection techniques? **[Application]** | **(CO5)** | **(10 Marks)** |
|  | **or** |  |  |
| 10.a | Identify the role of proximity-based methods in outlier detection and provide an example. **[Knowledge]** | **(CO5)** | **(04 Marks)** |
| 10.b | Explain the difference between global outliers and local outliers.  **[Comprehension]** | **(CO5)** | **(06 Marks)** |
| 10.c | Explain in detail Proximity-Based Approaches **[Application]** | **(CO5)** | **(10 Marks)** |