|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Reg. No.** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# SRM Institute of Science and Technology SRM Institute of Science and Technology Vector Logo - (.SVG + .PNG) - VectorLogoSeek.Com College of Engineering and Technology

**SET - C**

**School of Computing**

# DEPARTMENT OF COMPUTING TECHNOLOGIES

SRM Nagar, Kattankulathur – 603203, Chengalpattu District, Tamilnadu

# Academic Year: 2023-2024 (ODD)

**Test:** CLAT-2 **Date:** 07.10.2023

**Course Code & Title:** 18CSE355T - Data Mining And Analytics **Duration:** 2 Periods

**Year & Sem:** III & IV Year & 05th & 07th Semester **Max. Marks:** 50 Marks

**Course Articulation Matrix: *(to be placed)***

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S.**  **No.** | **Course Outcome** | **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** |
| 1 | CO2 | 3 |  |  |  |  |  |  | 3 |  |  |  |  |
| 2 | CO3 |  | 3 |  |  |  |  |  | 3 |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Part – A**  **(10 x 1 = 10 Marks)**  **Answer all questions. The duration for answering the part A is 15 minutes (MCQ Answer sheet will be collected after 15 minutes)** | | | | | | |
| **Q.**  **No** | **Question** | **Marks** | **BL** | **CO** | **PO** | **PI**  **Code** |
| **1** | When do you consider an association rule interesting?   1. If it only satisfies min\_support 2. If it only satisfies min\_confidence 3. **If it satisfies both min\_support and min\_confidence** 4. There are other measures to check so | **1** | **1** | **2** | **1** | **1.7.1** |
| **2** | What is the relation between candidate and frequent itemsets?   1. A candidate itemset is always a frequent itemset 2. **A frequent itemset must be a candidate itemset.** 3. No relation between the two 4. Both are same | **1** | **1** | **2** | **1** | **1.7.1** |
| **3** | Decision tree is a \_\_\_\_\_\_ algorithm.  a) supervised learning  **b) unsupervised learning**  c) Both  d) None of these | **1** | **3** | **3** | **1** | **1.7.1** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **4** | Spam Classification is an example for ?   1. **Naive Bayes** 2. Probabilistic condition 3. Random Forest 4. All the Above | **1** | **2** | **3** | **1** | **1.7.1** |
| 5 | Ensemble methods seek to \_\_\_\_\_\_\_\_\_\_\_.  a) Reduce variance of individual weak learners by aggregating their predictions.  b) Improve performance by exploiting prediction diversity.  **c) Both (a) and (b).**  d) None of the above. | **1** | **1** | **3** | **1** | **1.7.1** |
| **6** | Training in parallel that occurs in bagging aims to capitalize on the \_\_\_\_\_\_\_\_\_\_ of each base learner, while the sequential training in boosting capitalizes on the \_\_\_\_\_\_\_\_ of the learners.  **a) Independence , dependence.**  b) Dependence, Independence.  c) Dependence , Dependence.  d) Independence, Independence. | **1** | **1** | **3** | **2** | **2.5.2** |
| **7** | In Random Forest the Memory requirement for the storage process?  **a) High Memory**  b) Low Memory  c)No Memory  d) None of the Above | **1** | **1** | **3** | **2** | **2.5.2** |
| **8** | 5. Which of the following is an example of a classification problem?  a) Predicting the price of a house based on its features  b) Predicting the weight of a person based on their height  **c) Predicting whether a customer will churn or not**  d) Predicting the age of a person based on their income | **1** | **1** | **2** | **2** | **2.5.2** |
| **9** | Which one of the following correctly refers to the task of the classification?   1. A measure of the accuracy, of the classification of a concept that is given by a certain theory 2. **The task of assigning a classification to a set of examples** 3. A subdivision of a set of examples into a number of classes 4. None of the above | **1** | **1** | **2** | **2** | **2.5.2** |
| **10** | Which of the following is the direct application of frequent itemset mining?  a) Social Network Analysis  **b) Market Basket Analysis**  c) Outlier Detection  d) Intrusion Detection | **1** | **1** | **2** | **2** | **2.5.2** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Part – B**  **(4 x 5 = 20 Marks) Answer any 4 Questions** | | | | | | |
| **11** | How to generate candidate itemset in Apriori algorithm? Explain in detail  A screenshot of a computer  Description automatically generated  A diagram of a diagram  Description automatically generated | **5** | **2** | **2** | **8** | **8.4.1** |
| **12** | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | The following table contains training dataset of grade database.  Use information gain decision tree algorithm to construct a  decision tree from the given dataset.   |  |  |  |  | | --- | --- | --- | --- | | subject | standard | marks | Grade | | Science | 10 | 75-100 | A | | Science | 10 | 55-74 | B | | Maths | 9 | 90-100 | A | | Maths | 9 | 55-89 | B | | Maths | 10 | 80-100 | A | | Maths | 10 | 55-79 | B | | Science | 9 | 80-100 | A | | Science | 9 | 55-79 | B | | | **5** | **2** | **3** | **1** | **1.7.1** |
| **13** | How ensemble method improves the classification accuracy? Justify with suitable example.  A diagram of a method  Description automatically generated  A diagram of a method  Description automatically generated | **5** | **2** | **3** | **1** | **1.7.1** |
| **14** | Give comparison between the strong and the weak association rules  A clipboard with a list of characteristics  Description automatically generated  A white background with black text  Description automatically generated | **5** | **2** | **3** | **2** | **2.6.4** |
| **15** | Describe the following   1. What are attributes and how they influence a learning model. Give Examples      1. Define a stump.   A decision stump is a machine learning model consisting of a one-level decision tree. That is, it is a decision tree with one internal node (the root) which is immediately connected to the terminal nodes (its leaves). A decision stump makes a prediction based on the value of just a single input feature.  A diagram of a computer  Description automatically generated with medium confidence | **5** | **2** | **2** | **2** | **2.6.4** |
| **Part – B**  **(2 x 10 = 20 Marks)** | | | | | | |
| **16** | |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | **Brand** | **Fuel type** | | **Mileage** | | **Model** | | **Buy** | | A1 | petrol | low | | SUV | | Not recommended | | | A1 | petrol | medium | | Sports | | recommended | | | A2 | diesel | high | | SUV | | recommended | | | A1 | diesel | medium | | SUV | | Not recommended | | | A3 | diesel | low | | Sports | | Not recommended | | | A2 | CNG | medium | | SUV | | recommended | | | A4 | petrol | low | | SUV | | recommended | | | A3 | petrol | medium | | Sports | | Not recommended | | | A1 | diesel | high | | SUV | | recommended | | | A2 | diesel | medium | | SUV | | Not recommended | | | A3 | diesel | low | | Sports | | Not recommended | | | A4 | CNG | medium | | SUV | | Not recommended | | | A1 | diesel | low | | Sports | | Not recommended | | | A1 | diesel | medium | | Sports | | recommended | | | A2 | petrol | medium | | Sports | | recommended | |   Consider the following dataset that contains four attributes  and one class, use Naïve base classifier to determine the class  of Buy for A1 diesel low SUV. | 10 | L3 | 2 | 4 | 2.7.1 |
| **[OR]** | | | | | | |
| **17** | What is Closed frequent itemset and maximal frequent item set?  Explain in detail about how frequent item sets are identified with any example using Apriori Algorithm.  A diagram of a closed item  Description automatically generated  A diagram of a computer  Description automatically generated | **10** | **L3** | **2** | **4** | **2.7.1** |
| **18** | Explain about Naive Bayes Classification algorithm with example?  A diagram of a number of statistical data  Description automatically generated with medium confidenceA graph and diagram of a person's equation  Description automatically generated | **10** | **L2** | **3** | **2** | **2.5.2** |
| **[OR]** | | | | | | |
| **19** | Write in detail about Random Forests and write how they  fit well with the ensembling, bagging and bootstrapping concept.  A diagram of a tree  Description automatically generated  A diagram of a method  Description automatically generated | **10** | **L2** | **3** | **4** | **2.6.4** |

\*Performance Indicators are available separately for Computer Science and Engineering in AICTE examination reforms policy.

Course Outcome (CO) and Bloom’s level (BL) Coverage in Questions

**CO Coverage (%)**

A blue pie chart with white text

Description automatically generated**50**

**40**

**30**

**20**

**10**

**0**

**CO1 CO2 CO3 CO4 CO5**

Approved by the Audit Professor/Course Coordinator