**Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Roll No: \_\_\_\_\_\_\_\_\_\_\_\_\_**

**Note: Circle the correct option, avoid cutting and over writing Marks: 20**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Identify correct period for “empirical science” in the evolution of sciences | | | | | | |
| a | Before 1600 | b | 1600 – 1950 | c | 1950 – 1990 | d | 1990 – now |
| 2 | Identify correct period for “theoretical science” in the evolution of sciences | | | | | | |
| a | Before 1600 | b | 1600 – 1950 | c | 1950 – 1990 | d | 1990 – now |
| 3 | Identify correct period for “computational science” in the evolution of sciences | | | | | | |
| a | Before 1600 | b | 1600 – 1950 | c | 1950 – 1990 | d | 1990 – now |
| 4 | Identify correct period for “data science” in the evolution of sciences | | | | | | |
| a | Before 1600 | b | 1600 – 1950 | c | 1950 – 1990 | d | 1990 – now |
| 5 | Which of the following science traditionally meant simulation | | | | | | |
| a | Computational | b | Theoretical | c | Empirical | d | Data |
| 6 | In the evolution of database technology when RDBMS and advanced data models emerged | | | | | | |
| a | 1970s | b | 1980s | c | 1990s | d | 2000s |
| 7 | In the evolution of database technology when application oriented DBMS emerged | | | | | | |
| a | 1960s | b | 1970s | c | 1980s | d | 1990s |
| 8 | In the evolution of database technology multimedia databases belong to which period | | | | | | |
| a | 1960s | b | 1970s | c | 1980s | d | 1990s |
| 9 | Which of the following describe interesting patterns | | | | | | |
| a | Trivial | b | Explicit | c | Potentially useful | d | a and b |
| 10 | Which of the following does not describe interesting patterns | | | | | | |
| a | Trivial | b | Explicit | c | Potentially useful | d | a and b |
| 11 | Which of the following are alternative names for data mining | | | | | | |
| a | KDD | b | Data archaeology | c | Business intelligence | d | All of them |
| 12 | Which of the following does not fall in data preprocessing step in KDD process | | | | | | |
| a | Data integration | b | Normalization | c | Feature selection | d | Classification |
| 13 | Which of the following is part of data mining step in KDD process | | | | | | |
| a | Data integration | b | Normalization | c | Feature selection | d | Classification |
| 14 | Which of the following refer to operations on data cube | | | | | | |
| a | Roll-down | b | Roll-up | c | Drill-down | d | b & c |
| 15 | A frequent itemset of length 100 has how many frequent subset of length 2 | | | | | | |
| a | 50 | b | 200 | c | 4950 | d | 10000 |
| 16 | Which of the following metrics are used to evaluate classifier performance | | | | | | |
| a | Recognition rate | b | Sensitivity | c | Specificity | d | F-Score |
| 17 | If lift(A, B) is equal to one, then A and B are | | | | | | |
| a | Dependent | b | +vely correlated | c | -vely correlated | d | Independent |
| 18 | Mining frequent patterns leads to the discovery of interesting | | | | | | |
| a | Association | b | Correlation | c | Affiliation | d | Regression |
| 19 | The process of finding a model that describes and distinguishes concepts | | | | | | |
| a | Classification | b | Regression | c | Association | d | Clustering |
| 20 | Which of the following are methods for constructing classification models | | | | | | |
| a | SVM | b | KNN | c | Bayesian classification | d | All of them |

**Subjective Time Allowed: 50 minutes**

**Note: Attempt all questions. Each question is of 10 marks.**

**Question No 1:** Convert the data given below in vertical data format and find frequent 2-item-sets (**L2**) without candidate generation. Assume minimum support count of 2.

|  |  |
| --- | --- |
| **Trans ID** | **Itemsets** |
| T001 | {Egg, Bread, Jam} |
| T002 | {Egg, Bread, Butter, Milk} |
| T003 | {Egg, Bread, Milk, Jam} |
| T004 | {Bread, Jam, Milk} |
| T005 | {Butter, Jam, Milk} |

**Question No 2**: Confidence of an association rule is given by the formula:

**confidence( A => B) = support\_count(A U B) / support\_count(A)**

where **A => B** is the rule with **left hand side = A** and **right hand side = B**

calculate confidence for the following rules:

i) {Milk} => {Bread}

ii) {Bread, Jam} => {Milk}

iii) {Egg, Jam} => {Bread}

##### **Question No 3**: Fill the shaded cells of the given table by selected the correct items from the options {True Positive=9, False Positive=0, True Negative=4, False Negative=1}

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Predicted class** | | **Total** |
|  |  | **Yes** | **No** |  |
| **Actual class** | **Yes** |  |  |  |
| **No** |  |  |  |
| **Total** |  |  |  |  |

##### **Question No 4:** Using your answer to Question No 2 calculate the following:

##### 1. accuracy = (TP + TN) / (P + N )

##### 2. precision = TP/(TP + FP)

##### **Question No 5:** What is the difference between Classification and Clustering. Name at least 3 algorithms for classification and clustering.

**Key**

|  |  |  |  |
| --- | --- | --- | --- |
| **1** | **a** | **16** | **New question** |
| **2** | **b** | **17** | **New question** |
| **3** | **c** | **18** | **a** |
| **4** | **d** | **19** | **a** |
| **5** | **a** | **20** | **d** |
| **6** | **b** | **21** |  |
| **7** | **c** | **22** |  |
| **8** | **d** | **23** |  |
| **9** | **c** | **24** |  |
| **10** | **d** | **25** |  |
| **11** | **d** | **26** |  |
| **12** | **d** | **27** |  |
| **13** | **d** | **28** |  |
| **14** | **d** | **29** |  |
| **15** | **New question** | **30** |  |