


Course Code : 2101CS521

Date : 03-11-2023

Course Name : Data Mining

Duration : 150 Minutes

Total Marks : 70

Instructions:

1. Attempt all the questions.
2. Figures to the right indicates maximum marks.
3. Make suitable assumptions wherever necessary.

- Q.1 (A)** Define Data Mining? List out applications of data mining in real world scenarios. **4**
- (B)** Explain types of attributes in the context of dataset. **3**

OR

Justify this statement "All patterns are not interesting or useful".

- (C)** Explain issues and challenges with data mining. **7**

OR

Explain KDD Process with diagram.

- Q.2 (A)** What do you mean by missing values? Explain methods to fill/handle missing values. **4**
- (B)** What is Data Cleaning? Explain methods and significance of it. **3**

OR

What is binning? Explain it with example.

- (C)** Explain Data Integration with suitable example. **7**

OR

What do you mean by data pre-processing? Explain Data Transformation techniques in brief.

- Q.3 (A)** What is market basket analysis? Explain significance and industrial application of it. **4**
- (B)** What is confidence and support? Explain significance of it. **3**

OR

What is frequent itemset mining?

- (C)** Explain Apriori algorithm with suitable example. **7**

OR

Explain FP-TREE algorithm with suitable example.

- Q.4 (A)** Compare supervised and un-supervised learning strategies. **4**
- (B)** Define following in brief. **3**
1. Information Gain
 2. GINI Index
 3. Gain Ratio

OR

Discuss following performance evaluation criterions of the model.

1. Accuracy
 2. Precision
 3. Recall
- (C)** What is Cross Validation? Explain significance of it. Discuss different methods or strategies of it. **7**

OR

Write Bayes' theorem. Explain Bayesian classifier with suitable example.

- Q.5 (A)** Compare Hierarchical and Density based methods of clustering. **4**
- (B)** What is outlier in the context of clustering? **3**

OR

Explain in brief methods for "outlier detection".

- (C)** Explain "K-Means" clustering algorithm with suitable example. **7**

OR

Explain "K-medoid" clustering algorithm with suitable example.
