

M. Tech I Year I Semester Regular Examinations, February 2018

**DATA MINING & KNOWLEDGE DISCOVERY
(CSE)**

Time : 3 hours

Max Marks : 60

Answer all **five** units. (5 x 12 = 60 Marks)**UNIT-I**

1. (a) What is Data Mining? Discuss the challenges that motivated the development of Data Mining
- (b) Discuss briefly the following data quality issues
- Outliers
 - Missing and inconsistent values
 - Duplicate data

OR

2. (a) Describe the various multidimensional analysis techniques with suitable example
- (b) Write the equation of Pearson's Correlation Coefficient between two data objects. Use the same to find the relationship between $x=(-3,6,0,3,-6)$ and $Y=(1,-2,0,-1,2)$

UNIT-II

3. Construct the decision tree classification model to classify the bank loan applications by assigning applications to one of the three risk classes using the Gini Index for selecting the attributes

Owns home?	Married	Gender	Employed	Credit rating	Risk class
YES	YES	MALE	YES	A	B
NO	NO	FEMALE	YES	A	A
YES	YES	FEMALE	YES	B	C
YES	NO	MALE	NO	B	B
NO	YES	FEMALE	YES	B	C
NO	NO	FEMALE	YES	B	A
NO	NO	MALE	NO	B	B
YES	NO	FEMALE	YES	A	A
NO	YES	FEMALE	YES	A	C
YES	YES	FEMALE	YES	A	C

OR

4. (a) Discuss Naïve Bayes classifier estimates the class conditional probabilities and hence predict the class label using suitable example
- (b) Discuss how Linear SVM is used in non separable classification problems.

Continued in page 2

UNIT-III

5. (a) Discuss Apriori algorithm for Frequent item set generation with suitable example
(b) Describe in brief various methodologies for applying association analysis for continuous attributes

OR

6. (a) Given $\text{min_supp}=3$. Construct FP-Tree for the given set of transactions. Find Frequent Patterns for item 'I5' using constructed FP Tree

TID	Items
T100.	{I1, I2, I5}
T200.	{I2, I4}
T300.	{I2, I3}
T400.	{I1, I2, I4}
T500.	{I1, I3}
T600.	{I2, I3}
T700.	{I1, I3}
T800.	{I1, I2, I3, I5}
T900.	{I1, I2, I3}

- (b) Describe the Apriori algorithm for mining Frequent sub graphs with suitable example

UNIT-IV

7. (a) Discuss the working of K-means clustering algorithm with suitable example
(b) Differentiate various types of clusters with neat diagrams.

OR

8. (a) Discuss the strength and weakness and working of DBSCAN to produce density-based clusters
(b) Examine the two approaches for unsupervised approaches for assessing cluster validity based on the proximity matrix

UNIT-V

9. Write short note on the following
i) Web content mining
ii) Web usage mining

OR

10. Discuss how page rank algorithm is used in ranking web pages with suitable example
