SRM University, Kattankulathur
Department of Information Technology
IT1101/ DATA WAREHOUSING AND DATA MINING
Cycle Test II

Class: III/VI Sem/ BTech **Date: 27-04-2017**
Duration: 3 hrs ` **Max. Marks: 100 marks**

Reg No.

List of Instructional Objectives covered in this Test:

IO2 - Find the unseen pattern in large volumes of historical data that helps to manage an organization efficiently.
IO3. Understand the concepts of various data mining Techniques

Outcomes covered in this test:

i) - An ability to use current techniques, skills and tools necessary for computing practice.

i1 :An ability to understand current techniques and Skills

i2 :An ability to understand tools necessary for computing practice

j) - An ability to use and apply current technical concepts and practices in the core information technologies.

j1: An ability to use and apply current technical concepts in the core information technologies

j2: An ability to use and apply current practices in the core information technologies

Question No	Instructional Objective	Course Outcome	Marks Scored	Outcome Met Yes/No
1	Io2	i1		
2	Io2	i1		
3	Io2	i1		
4	Io2	i1		
5	Io2	i1		
6	Io3	j1		
7	Io3	j1		
8	Io3	j1		
9	Io3	j1		
10	Io3	j1		
11	Io2	i2		
12	Io2	i2		
13	Io2	i2		
14	Io2	i2		

15	Io2	i2		
16	Io3	i2		
17	Io3	i2		
18	Io3	j2		
19	Io3	j2		
20	Io3	j2		
21	Io3	j1		
22	Io3	j1		
23	Io3	j1		
24	Io3	j1		
25	Io3	j2		
26	Io3	j2		
27	Io3	j2		
28	Io3	j1		
29	Io3	j1		
30	Io3	j2		
31	Io3	j2		
32	Io3	j2		

Total Marks: /100

Signature:

Part-B (Answer any 5 questions) (5x4=20 marks)

21. Briefly discuss the interestingness measures – support and the confidence used in Pattern evaluation?
22. Explain the various forms of integrating a Data Mining system with a Data warehouse.
23. With a neat sketch explain the Data Mining System architecture.
24. Differentiate Classification from Clustering with examples.
25. Write short notes on Support Vector Machines?
26. Define an outlier. List out and explain the different types of outliers.
27. Write short notes on Linkage measures (Distance measures).

Part-C (Answer all the following) (5x12=60 marks)

28. a. Explain the steps involved in data mining and knowledge discovery and discuss the major issues in data mining.

OR

b.Write any three Data pre-processing techniques.

29. a. Write the procedure for ID3 and apply the Decision tree Induction for the following dataset:

Model	Engine	SC/Turbo	Weight	Fuel Eco	Fast
Prius	small	no	average	good	no
Civic	small	no	light	average	no
WRX STI	small	yes	average	bad	yes
M3	medium	no	heavy	bad	yes
RS4	large	no	average	bad	yes
GTI	medium	no	light	bad	no
XJR	large	yes	heavy	bad	no
S500	large	no	heavy	bad	no
911	medium	yes	light	bad	yes
Corvette	large	no	average	bad	yes
Insight	small	no	light	good	no
RSX	small	no	average	average	no
IS350	medium	no	heavy	bad	no
MR2	small	yes	average	average	no
E320	medium	no	heavy	bad	no

In the above table – Model,Engine,SC/Turbo,Weight,Fuel Eco represents input attributes of vehicle dataset. Based on these the class attribute “Fast” is identified.

Or

b. What is a Naïve bayes theorm ? Explain the Naïve Baye’s classification for the above Training data set in predicting a class label. Classify the following test data - X= (Model=IS350 , Engine =large, SC/Turbo = Yes, Weight =medium, Fuel Eco=good)

30.a.Write the steps in Apriori Technique? Apply the apriori algorithm for the following dataset. Minimum support count = 3 and minimum confidence = 70%

Transaction No	Item_IDs.
T0	1,2,3,4,5
T1	7,2,3,4,5,6

T2	1,8,4,5
T3	1,9,0,4,6
T4	0,2,2,4,5

(1) Find the frequent itemsets.
(2) Generate association rules

Or

b. Write short notes on :
(i) "Strong rules are not Necessarily Interesting"
(ii) From Association Analysis to Correlation Analysis.

31.a. Explain the K-Means partitions method and write the algorithm for the same.

Or

b.Write short notes on (1) AGNES (2) DIANA

32. Explain the Data Mining applications with respect to Science and Engineering.

Or

Discuss the Data Mining for Intrusion Detection and Prevention.
