Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, APRIL 2018

Course Code: IT304

Course Name: DATA WAREHOUSING AND MINING (IT)

Max. Marks: 100 Duration: 3 Hours

PART A

Answer any two full questions, each carries 15 marks. Marks

1 a) Which are the methods to handle missing values during data mining? (5)

b) What is Data Cleaning? How can we use binning to handle noisy data? (5)

c) What is data visualization? (2)

d) Why data visualization is important in data mining? List out the softwares used for data visualization? (3)

2 a) Write any 5 applications where data mining can be directly applied. (10)

b) Explain the data transformation and discretization methods. (5)

3 a) Explain how huge data are stored and processed using Data Warehouses and OLAP. (15)

PART B

Answer any two full questions, each carries 15 marks.

4 a) Build a decision tree by calculating the information gain of each attribute for the following training data (10)

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

	b)	Explain the classifier that uses probability for classification process.	(5)
5	a)	Explain the concept of Neural Networks	(10)
	b)	What is a prediction model? List any four prediction models.	(5)
6	a)	Explain how Support Vector Machines are used for classification.	(15)

PART C

Answer any two full questions, each carries 20 marks.

7	a)	Write the step by step procedure of Apriori algorithm with the help of an example.	(10)
	b)	Use the k-means algorithm and Euclidean distance to cluster the following 8 samples into 3 clusters: A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9). Suppose that the initial seeds (centers of each cluster) are A1, A4 and A7. Execute the k-means algorithm for 1 epoch. At the end of this epoch show the following: a. The new clusters (i.e. the samples belonging to each cluster); b. The centers of the new clusters; c. Draw a 10 by 10 space with all the 8 points and show the clusters after the first epoch and the new centroids. d. How many more iterations are needed to converge?	(10)
8	a)	Identify the role of R package in data mining and demonstrate five salient features of R software.	(10)
	b)	Describe the steps involved in building classifier models in Weka.	(7)
	c)	What is market basket analysis?	(3)
9	a)	Explain CRM lifecycle and architecture.	(15)
	b)	Explain about weighted graph partitioning.	(5)

****