Exam Slot → **C**

APJ Abdul Kalam Technological University
First Semester M.Tech Degree Examination January 2016
Ernakulam II Cluster
COMPUTER SCIENCE AND ENGINEERING
Specialization: COMPUTER SCIENCE AND ENGINEERING

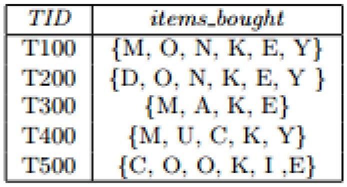
05CS 6005-DATA MINING AND WAREHOUSING

Time: 3 hrs Max. Marks: 60

I.

a) The age values for the data tuples are 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Use smoothing by bin means to smooth the above data, using a bin depth of 3. (4 Marks)

b) Find all frequent item sets using Apriori and FP-growth. Let min sup = 60% and min conf = 80%. Compare the efficiency of the two algorithms. (8 Marks)



TID	*items_bought*
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y }
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I ,E}

II.

a) Briefly outline the major steps of decision tree classification. (8 Marks)

b) Why is naive Bayesian classification called "naive"? (4 Marks)

III.

a) Discuss 3 tier data warehousing architecture? (10 Marks)

b) Consider a two dimensional database D with the records : R1(2, 2), R2(2, 4), R3(4, 2), R4(4, 4), R5(3, 6), R6(7, 6), R7(9, 6), R8(5, 10), R9(8, 10), R10(10, 10). The distance function is the L1 distance (Manhattan distance). Show the results of the k-means algorithm at each step, assuming that you start with two clusters (k = 2) with centers C1 = (6, 6) and C2 = (9, 7) ? (4 Marks)

c) Find the strength and weakness of k-means in comparison with the k-medoids algorithm? (4 Marks)

OR

IV.

a) Differentiate OLAP and OLTP? (5 Marks)

b) Give an example of how specific clustering methods may be integrated, for example, where one clustering algorithm is used as a preprocessing step for another. (5 Marks)

c) How do hierarchical clustering methods work? (8 Marks)

V.

a) Explain parametric statistical method for outlier detection (12 Marks)

b) Explain Outlier Analysis. Discuss the different applications of outlier analysis (6 Marks)

OR

VI.

a) Explain the proximity based outlier detection. (12 Marks)

b) Explain the non-parametric statistical method for outlier detection. (6 Marks)