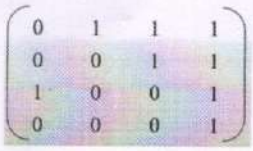
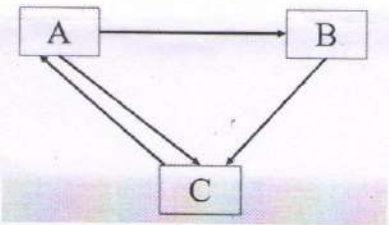
K. J. Somaiya Institute of Engineering and Information Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)
End Semester Exam
Nov – Dec 2021
(B.Tech.) Program: **Computer Engineering**
Examination: TY Semester: V
Course Code: **1UCEC504** and Course Name: **Data Warehousing and Mining**
Duration: 03 Hours Max. Marks: 60

Instructions:
(1)All questions are compulsory.
(2)Draw neat diagrams wherever applicable.
(3)Assume suitable data, if necessary.

		Max. Marks	CO	BT level
Q 1	**Solve any six questions out of eight:**	**12**		
i)	Define Market basket analysis. What do you mean by Frequent Pattern Mining?	**2**	CO3, CO4	U
ii)	Compare Star Schema and Snowflake Schema.	**2**	CO1	U
iii)	Draw OLAP cube **Slice** operation with suitable example.	**2**	CO1	U
iv)	Differentiate Web content mining and Web structure mining (4 points)	**2**	CO5	U
v)	What are the issues in data mining? (any 4 issues)	**2**	CO2	U
vi)	Define Clustering and Classification. List two algorithms for each.	**2**	CO3,CO4	U
vii)	What do you mean by Data Mining task primitives?	**2**	CO2	U
viii)	What are the types of distances uses in Clustering? Write formula with examples.	**2**	CO3,CO4	U
Q.2	**Solve any four questions out of six.**	**16**		

i)	Calculate the Hubs and Authority score using HITS algorithm for k=3. The adjacency matrix is 	4	CO5	Ap
ii)	Suppose the data for clustering is {2, 4, 10, 12, 3, 20, 30, 11, 25,5,36,41,14}. Assuming number of clusters to be 2 i.e. K = 2, cluster the given data using k-means clustering algorithm.	4	CO3,CO4	U
iii)	Explain slowly changing dimensions with example.	4	CO1	U
iv)	Draw KDD process and explain pre-processing steps.	4	CO2	U
v)	Explain classification using Decision tree induction considering Entropy & Information Gain.	4	CO3,CO4	U
vi)	With suitable example illustrate the Apriori principle.	4	CO3,CO4	U
Q.3	**Solve any two questions out of three.**	**16**		
i)	Consider a Data warehouse for a sport manufacturing company storing sales details of various sports equipment sold and the time of the sale. Create a cube and describe following OLAP operations i) Rollup ii)Drill down iii) Slice iv)Dice and v) Pivot	8	CO1	Ap
ii)	Compute the page rank for the given graph by using page rank algorithm upto iteration 2. 	8	CO5	Ap
iii)	Draw and explain in detail Data Mining Architecture.	8	CO2	U
Q.4	**Solve any two questions out of three.**	**16**		

i)		8	CO3, CO4	Ap
ii)		8	CO3, CO4	Ap

PATIENT	DISEASE	SUGAR LEVEL	SURVIVAL CHANCES
Small	Serious	High	Yes
Medium	Normal	Low	Yes
Senior	Lifetime	Normal	Yes
Small	Lifetime	High	No
Small	Normal	High	Yes
Senior	Serious	Normal	No
Medium	Serious	Low	Yes
Senior	Normal	Low	No
Medium	Lifetime	Normal	Yes
Medium	Serious	High	No
Senior	Normal	Low	No

Using the given training dataset classify the above tuples using Naïve Bayes algorithm:
<PATIENT: Senior, DISEASE: Normal, SUGAR LEVEL: Normal

Consider a database, D, consisting of 8 transactions. Use this table to show the implementation of k-means algorithm together with Euclidean distance function. Use K=3 and suppose A1, B1 and C1 are selected as the center of each cluster. Show three cluster centers after first and second round of execution only.

Individual	Variable1	Variable2
A1	2	10
A2	2	5
A3	8	4
B1	5	8
B2	7	5
B3	6	4
C1	1	2
C2	4	9

iii)	Use the Apriori to algorithm to identify the frequent item-sets in the following database. Then extract the strong association rules from these sets. Min. Support = 30% Min. Confidence=75%	**8**	CO3, CO4	3

TID	Items
01	A, B, D, E, F
02	B, C, E
03	A, B, D, E
04	A, B, D, E
05	A, B, C, D, E, F
06	B, C, D
07	A, B, D, E