SOMAIYA VIDYAVIHAR UNIVERSITY

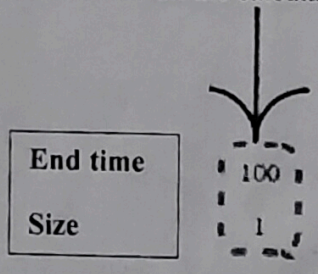
Semester: January 2025 – May 2025 **Examination: In-Semester Examination**		
Maximum Marks: 30		**Duration : 1 hour & 15 mins**
Programme code: 54 **Programme: B.Tech Computer Engineering**	**Class: TY**	**Semester: VI (SVU 2020)**
Name of the Constituent College: **K. J. Somaiya College of Engineering**	**Name of the department:** COMP & EXCP - Honours (DSA)	
Course Code: 116h54C601	**Name of the Course: Advanced Data Mining**	

Question No.		Max. Marks
Q1	Trace the results of using the FP-growth algorithm on the grocery store example with support count s = 2 and confidence threshold c = 50%. Show the candidate and frequent itemsets for each database scan. Enumerate all the final frequent itemsets. Also indicate the association rules that are generated and highlight the strong ones, sort them by confidence.	10

Transaction ID	Items Purchased
T1	I1,I2,I3
T2	I2,I5
T3	I4,I5
T4	I1,I2,I5
T5	I2,I3,I5

OR

Employ the DGIM algorithm. Shown below is a data stream with N=40 and current bucket configuration. Suppose that at times 101 through 105, 1's appear in the stream. Compute the set of buckets that would exist in the system at time 105. Also compute the number of 1's in latest k = 30 bits of the window. Show all the calculations with the explanation.



End time	100	98	95	92	87	80	65
Size	1	1	2	2	4	8	8

Q2	a) With suitable example, explain any 2 data mining tasks under descriptive data mining category? b) Provide example for characteristics of Big data- Velocity, Volume.	10

Q3	Perform single link and complete link hierarchical clustering. Show your results by drawing a dendogram. The dendogram should clearly show the order in which the points are merged.	10

Table 1 Distance matrix

	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00