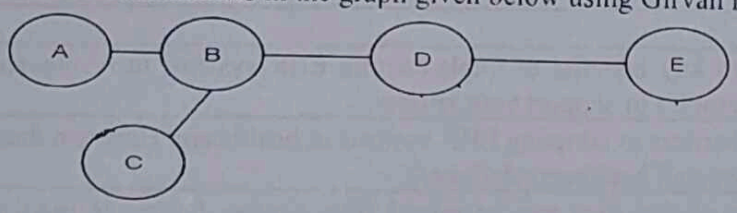
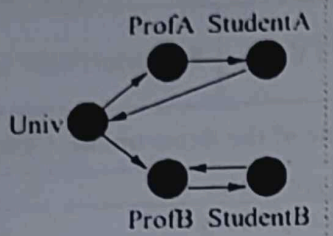
28·11·2024 (E)

SOMAIYA VIDYAVIHAR UNIVERSITY

Semester: July 2024 –November 2024		
Maximum Marks: 100 **Examination: ESE Examination** **Duration:3 Hrs.**		
Programme code: H54 **Programme: Honors in Data Science and Analytics**	**Class: TY**	**Semester:V(SVU 2020)**
Institute/School/Department: K. J. Somaiya School of Engineering	**Name of the department:** COMP//EXCP	
Course Code:116h54C501	**Name of the Course: Data Analytics**	
Instructions: 1)Draw neat diagrams 2) All questions are compulsory **3) Assume suitable data wherever necessary**		

Que. No.	Question	Max. Marks
Q1	Solve any **Four**	**20**
i)	Consider the following dataset representing monthly sales (in thousands): [220, 250, 280, 230, 270, 260, 240, 300] 1. Calculate the range of the dataset. 2. Compute the variance of the dataset. 3. Determine the standard deviation of the dataset.	5
ii)	What are the main differences between vector and raster data models in terms of storage, representation, and analysis?	5
iii)	What type of graph analysis can be done on social network like Twitter, list any 5 and its benefits	5
iv)	Give example for descriptive, diagnostic, predictive, and prescriptive analytics focusing on Air quality	5
v)	Describe the Random Walk Model for time series.	5
vi)	Examine the applications of NLP in healthcare, specifically focusing on its role in analyzing clinical text data.	5

Que. No.	Question	Max. Marks
Q2 A	Solve the following	**10**
i)	What is the role of GIS in sustainability planning and environmental conservation?	5
ii)	How do GIS tools help in visualizing, analyzing, and solving environmental and geographical problems in real-world scenarios?	5
	OR	
Q2 A	Explain DEM(Digital Elevation Model), DSM(Digital Surface Model) and DTM(Digital Terrain Model)	10
Q2 B	Solve any **One**	**10**
i)	Find the communities in the graph given below using Girvan newman algorithm 	10

ii)	Explain HUB node , AUTHORITY nodes and Simrank , For the following graph give the simplified node pair graph used to find simrank G  	10

Que. No.	Question	Max. Marks
Q3	Solve any **Two**	**20**
i)	Using the monthly visitor numbers for a tourist attraction in 2023: [210, 220, 240, 215, 225, 235, 220, 230, 250, 225, 235, 245] and assuming similar seasonal patterns continue in 2024, calculate a three-month moving average forecast to estimate visitor numbers for May 2024.	10
ii)	An event organizer wants to understand the entry pattern of attendees to an exhibition. She observes the first 25 people entering, with the following sequence of ticket holders (T) and non-ticket holders (N): T N T T N N T T N T N T N N T N T T N N N T N T N Using a significance level of $\alpha = 0.05$, test if the entry pattern appears random.	10
iii)	Find the exponential smoothing forecasts for periods 2-10 using the following data, whereα=0.10. Assume f1=a1 Week Demand 1 820 2 775 3 680 4 655 5 750 6 802 7 798 8 689 9 775 10 ?	10

Que. No.	Question	Max. Marks
Q4	Solve any **Two**	**20**
i)	Explain at least two key benefits of implementing EHR systems in healthcare. Provide detailed examples to support your answer.	10
ii)	Discuss two major barriers to adopting EHR systems in healthcare. How can these barriers impact the overall healthcare delivery?	10
iii)	What are the types of data that can be mined from sensor devices in medical informatics? Provide examples of sensors and the data they produce.	10

Que. No.	Question	Max. Marks
Q5	Solve any **four**	20
i)	The following dataset represents the daily temperatures (in degrees Celsius) over a 10-day period: [25, 27, 28, 22, 26, 29, 31, 35, 30, 24] 1. Calculate the interquartile range (IQR) of the dataset. 2. Identify any outliers in the dataset using the 1.5 * IQR rule. 3. Compute the mean and median with and without the outliers.	5
ii)	Explain Contours and Triangulated Irregular Network(TIN).	5
iii)	Clustering of Social-Network Graphs	5
iv)	Explain Locality w.r.t social media graph	5
v)	What are some EDA techniques commonly used for time series data?	5
vi)	What is meant by 'topology' in the context of the vector data model?	5

...ical values of r in the runs test*

...iven in the tables are various critical values of r for values of m and n less than or equal to 20. For the one-sample runs test, any observed value of r which is less than or equal to the smaller value, or is greater than or equal to the larger value in a pair is significant at the $\alpha = .05$ level.

m \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
											–	–	–	–	–	–	–	–	–
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
					–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
				9	9	–	–	–	–	–	–	–	–	–	–	–	–	–	–
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
			9	10	10	11	11	–	–	–	–	–	–	–	–	–	–	–	–
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
		–	9	10	11	12	12	13	13	13	13	–	–	–	–	–	–	–	–
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
		–	–	11	12	13	13	14	14	14	14	15	15	15	–	–	–	–	–
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
		–	–	11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
		–	–	–	13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
		–	–	–	13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
		–	–	–	13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
	–	–	–	–	13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
	–	–	–	–	–	15	16	17	18	19	19	20	20	21	21	22	22	23	23
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
	–	–	–	–	–	15	16	17	18	19	20	20	21	22	22	23	23	23	24
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
	–	–	–	–	–	15	16	18	18	19	20	21	22	22	23	23	24	24	25
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
	–	–	–	–	–	–	17	18	19	20	21	21	22	23	23	24	25	25	25
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
	–	–	–	–	–	–	17	18	19	20	21	22	23	23	24	25	25	26	26
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
	–	–	–	–	–	–	17	18	19	20	21	22	23	24	25	25	26	26	27
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
	–	–	–	–	–	–	17	18	20	21	22	23	23	24	25	26	26	27	27
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14
	–	–	–	–	–	–	17	18	20	21	22	23	24	25	25	26	27	27	28

* Adapted from Swed, and Eisenhart, C. (1943). Tables for testing randomness of grouping in a sequence of alternatives. *Annals of Mathematical Statistics*, 14, 83–86, with the kind permission of the authors and publisher.