



SOMAIYA
VIDYAVIHAR UNIVERSITY

Semester: January 2023 – May 2023		
Maximum Marks: 30	Examination: In-Semester Examination	Duration :1:15
Programme code:	Class: TY	Semester: V (SVU 2020)
Programme: Minor Programme in Computer Engineering		
Name of the Constituent College: K. J. Somaiya College of Engineering		Name of the Department: COMP
Course Code: 116m75E605	Name of the Course: Data Mining	

Question No.		Max. Marks	CO Mapped	BT Level																																																								
Q1	a. Explain KDD process with diagram (5m) b. Find the Jaccard similarity for sets $A = \{1, 2, 3, 7, 8\}$ and $B = \{1, 2, 4, 6, 8, 10\}$ (5m)	10	CO1,CO2	AP																																																								
Q2	Answer any 2 with example a. Types of Attributes; b. Statistical Description of Data c. Why Preprocessing is important	10	CO2	UN																																																								
Q3	A. Explain Linear and nonlinear regression, Logistic Regression with example OR B. Using the following data set, predict whether a young person with low income, has own house or rented house by building decision tree. (Assume required data) <table><tr><th>No</th><th>Income</th><th>Age</th><th>Own house</th></tr><tr><td>1</td><td>Very high</td><td>Young</td><td>Yes</td></tr><tr><td>2</td><td>High</td><td>Middle</td><td>Yes</td></tr><tr><td>3</td><td>Low</td><td>Young</td><td>Rented</td></tr><tr><td>4</td><td>High</td><td>Middle</td><td>Yes</td></tr><tr><td>5</td><td>Very high</td><td>Middle</td><td>Yes</td></tr><tr><td>6</td><td>Medium</td><td>Young</td><td>Yes</td></tr><tr><td>7</td><td>High</td><td>Old</td><td>Yes</td></tr><tr><td>8</td><td>Medium</td><td>Middle</td><td>Rented</td></tr><tr><td>9</td><td>Low</td><td>Middle</td><td>Rented</td></tr><tr><td>10</td><td>Low</td><td>Old</td><td>Rented</td></tr><tr><td>11</td><td>High</td><td>Young</td><td>Yes</td></tr><tr><td>12</td><td>Medium</td><td>Old</td><td>Rented</td></tr><tr><td>13</td><td>Medium</td><td>Young</td><td>Rented</td></tr></table>	No	Income	Age	Own house	1	Very high	Young	Yes	2	High	Middle	Yes	3	Low	Young	Rented	4	High	Middle	Yes	5	Very high	Middle	Yes	6	Medium	Young	Yes	7	High	Old	Yes	8	Medium	Middle	Rented	9	Low	Middle	Rented	10	Low	Old	Rented	11	High	Young	Yes	12	Medium	Old	Rented	13	Medium	Young	Rented	10	CO3	AN
No	Income	Age	Own house																																																									
1	Very high	Young	Yes																																																									
2	High	Middle	Yes																																																									
3	Low	Young	Rented																																																									
4	High	Middle	Yes																																																									
5	Very high	Middle	Yes																																																									
6	Medium	Young	Yes																																																									
7	High	Old	Yes																																																									
8	Medium	Middle	Rented																																																									
9	Low	Middle	Rented																																																									
10	Low	Old	Rented																																																									
11	High	Young	Yes																																																									
12	Medium	Old	Rented																																																									
13	Medium	Young	Rented																																																									