Office copy (F)

Roll No: _____

Thapar Institute of Engineering and Technology, Patiala

Computer Science and Engineering Department

BE (CSE) Third Year

(Dera Bassi Campus)

Course: Machine Learning **Code: UML501**

Time: 2 Hrs; Date: 25.1.2021 **MM: 50**

Note: Attempt any Five questions.

1 a)	To perform the data analysis of COVID cases and to make its predictions, which machine learning technique you will use and why? Justify your answer.	5
b)	Why Naïve Bayes Classification is considered as Naïve? In case of zero probability for an event, how data scientist handle this situation in Naïve Bayes Classification?	5
2)	Let's suppose as a data scientist you have to predict the price of house by considering following dataset. How will you perform predictions over this dataset? Explain your approach by following all the steps like data loading, data pre-processing, model building and testing etc. by giving corresponding Python code.	10

Number of Bedrooms	Covered Area in yards	Granite Flooring	Upgraded Kitchen	Distance from City in Km	Age of flat in years	Price
3	150	Y	Y	2	2	70
2	250	Y	Y	1	1	140
4	320	N	Y	2	1	200
1	130	Y	Y	3	2	60
4	500	Y	N	5	3	200
3	160	N	N	1	2	60

3 a)	How a data scientist decides about the number of nodes in hidden layer, input layer and output layer? For database given in Q 2, What will number of these nodes in each of these layers.	5
b)	What are the limitations of k-means algorithm?	5
4 a)	How Gini Index can be used to build the decision tree? Explain the intitition behing building decision tree by using Gini Index?	5
b)	Calculate the following for given Confusion Matrix. <table><tr><td></td><td></td><td colspan="2">Predicted</td></tr><tr><td rowspan="3">Actual</td><td></td><td>0</td><td>1</td></tr><tr><td>0</td><td>28</td><td>3</td></tr><tr><td>1</td><td>2</td><td>23</td></tr></table> • True Positive • False Negative • Precision • Recall • F-Measure	5
5 a)	Fit the straight line for simple linear regression with 0_0=2 and θ_1=2 as given below. $Y = 0_0 + \theta_1 X$	5
b)	What will happen if no elitism is applied in Genetic Algorithm?	5
6)	Find the following for the threshold support of 50% and threshold confidence of 70% over the given database. <table><tr><td>T1</td><td>1,3,4</td></tr><tr><td>T2</td><td>2,3,5</td></tr><tr><td>T3</td><td>1,2,3,5</td></tr><tr><td>T4</td><td>2,5</td></tr></table> C1, L1, C2, L2, C3, L3, Possible Rules, Qualified Rules, Identify Improvement due to pruning. Explain all the steps with their corresponding rule.	10

7 a).	Compute the support for item sets *{b, d}*, and *{b, d, e}* by treating each transaction ID as a market basket. Also compute the confidence and lift for the association rules *{b, d}* → *{e}* and *{e}* → *{b, d}*.	5
b)	Explain the working principle of agglomerative and divisive clustering algorithm?	5

Customer ID	Transaction ID	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}