

UNIT TEST-2 (09th March 2022)

BSc. VIth Sem (Data Mining)

Marks : 30

Time : 1hours

All questions are compulsory.

Q. No.	Write all answers in one place.	Mark																																																
1	(i)Why association rule mining is used? (ii) Brief about k-fold cross validation. (iii)What is the limitation of k-mean clustering? (iv)How to measure distance between vector objects? (v) What is Ordinal variable? Give an example. (vi) Give a common example of nonlinear separable problem.	6 x 2																																																
2	Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8). (i)Compute the Euclidean distance and Manhattan distance between two objects . (ii)Compute the Minkowski distance between the two objects using p=3	06																																																
3	Using the below given table and Naïve Bayesian Classification algorithm find out the class of the new tuple, X={age=senior, income=medium, student=yes, credit rating =fair} <table><tr><th>RID</th><th>Age</th><th>Income</th><th>Student</th><th>Credit rating</th><th>Class : Buys computer</th></tr><tr><td>1</td><td>Youth</td><td>High</td><td>No</td><td>Fair</td><td>No</td></tr><tr><td>2</td><td>Youth</td><td>High</td><td>No</td><td>Excellent</td><td>No</td></tr><tr><td>3</td><td>Middle aged</td><td>High</td><td>No</td><td>Fair</td><td>Yes</td></tr><tr><td>4</td><td>Senior</td><td>Medium</td><td>No</td><td>Fair</td><td>Yes</td></tr><tr><td>5</td><td>Senior</td><td>Low</td><td>Yes</td><td>Fair</td><td>Yes</td></tr><tr><td>6</td><td>Senior</td><td>Low</td><td>Yes</td><td>Excellent</td><td>No</td></tr><tr><td>7</td><td>Middle aged</td><td>Low</td><td>Yes</td><td>Excellent</td><td>Yes</td></tr></table>	RID	Age	Income	Student	Credit rating	Class : Buys computer	1	Youth	High	No	Fair	No	2	Youth	High	No	Excellent	No	3	Middle aged	High	No	Fair	Yes	4	Senior	Medium	No	Fair	Yes	5	Senior	Low	Yes	Fair	Yes	6	Senior	Low	Yes	Excellent	No	7	Middle aged	Low	Yes	Excellent	Yes	06
RID	Age	Income	Student	Credit rating	Class : Buys computer																																													
1	Youth	High	No	Fair	No																																													
2	Youth	High	No	Excellent	No																																													
3	Middle aged	High	No	Fair	Yes																																													
4	Senior	Medium	No	Fair	Yes																																													
5	Senior	Low	Yes	Fair	Yes																																													
6	Senior	Low	Yes	Excellent	No																																													
7	Middle aged	Low	Yes	Excellent	Yes																																													
4	Suppose that the data mining task is to cluster the following nine points into three clusters. A ₁ (2, 10), A ₂ (2, 5), A ₃ (8,4), B ₁ (5, 8), B ₂ (7, 5), B ₃ (6, 4), C ₁ (1, 2), C ₂ (4, 9), C ₃ (3,4) The distance function is Euclidean distance. Suppose initially we assign A ₂ , B ₂ and C ₂ as the center of each cluster respectively. Use the k-means algorithm to show the three cluster centers after the first round of execution.	06																																																