

NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI-15
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
B. Tech DEGREE VI SEMESTER, II CYCLE TEST, APRIL-2021
CSPE14–Data Warehousing and Data Mining

DATE: 22-04-2021 TIME: 03.30p.m. -04.30 p.m. MAX.MARKS:20 marks

Answer all Questions

5 x 4 =20 marks

1. The following table consists of training data from an employee database. The data have been generalized. For example, “31... 35” for *age* represents the age range of 31 to 35. For a given row entry, *count* represents the number of data tuples having the values for *department*, *status*, *age*, and *salary* given in that row.

<i>department</i>	<i>status</i>	<i>age</i>	<i>salary</i>	<i>count</i>
sales	senior	31...35	46K...50K	30
sales	junior	26...30	26K...30K	40
sales	junior	31...35	31K...35K	40
systems	junior	21...25	46K...50K	20
systems	senior	31...35	66K...70K	5
systems	junior	26...30	46K...50K	3
systems	senior	41...45	66K...70K	3
marketing	senior	36...40	46K...50K	10
marketing	junior	31...35	41K...45K	4
secretary	senior	46...50	36K...40K	4
secretary	junior	26...30	26K...30K	6

Let status be the class label attribute.

Given a data tuple having the values “systems,” “26. . . 30” and “46–50K” for the attributes department, age, and salary, respectively, what would a naive Bayesian classification of the status for the tuple be?

2. A database has five transactions. Let minimum support count more than 50%

TID Items bought

- 1 Milk, Tea, Cake
- 2 Eggs, Tea, Cold Drink
- 3 Milk, Eggs, Tea, Cold Drink
- 4 Eggs, Cold Drink
- 5 Juice

Find all frequent item sets using Apriori algorithm.

3. Consider a database, D , consisting of 7 transactions. Use this table to show the implementation of k-means algorithm together with Euclidean distance function. Use $K=2$ and suppose individual 1 and 4 are selected as the initial means.

Individual	Variable 1	Variable 2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

4. Explain the different ensemble-based classification techniques using examples.
5. Write short notes on categorization of major Clustering Methods

-----BestWishes-----