

LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034



B.Sc. DEGREE EXAMINATION – STATISTICS

FOURTH SEMESTER – APRIL 2022

UST 4602 – DATA MINING

Date: 23-06-2022

Dept. No.

Max. : 100 Marks

Time: 09:00 AM - 12:00 NOON

SECTION – A

Answer ALL questions. Each carries two marks.

(10 X 2 = 20)

1. Define data mining.
2. What are the tasks associated with data mining?
3. Define a relational database.
4. What are outliers and how you detect them?
5. What are Numeric attributes? Give an example.
6. Why do we use Exploratory Data Analysis?
7. Define ROC.
8. What is the purpose of splitting data as train and test data?
9. Define Euclidean distance function.
10. Define frequent Itemset.

SECTION – B

Answer any FIVE questions. Each carries eight marks.

(5 X 8 = 40)

11. Explain the objective and different methods of binning data.
12. Elaborate the steps in cleaning data and handling missing information.
13. Use the following stock price data (in rupees) 10, 7, 20, 12, 75, 15, 9, 18, 4, 12, 8, 14.
 - (i) Calculate the mean, median, and mode stock price.
 - (ii) Compute the SD of the stock price. Interpret what this number means.
 - (iii) Find the min–max normalized stock price for the stock price 20 rupees. .
 - (iv) Identify the outlier.
 - (v) Verify that this value is an outlier, using the Z-score method.
 - (vi) Verify that this value is an outlier, using the IQR method.
14. Explain the difference between supervised and unsupervised methods. Which data mining tasks are associated with supervised methods? Elaborate.
15. Consider the problem of playing golf, if the humidity is 74, will you decide to play golf or not?

		Humidity								Mean	StDev	
Play	yes	86	96	80	65	70	80	70	90	75	79.1	10.2
Golf	no	85	90	70	95	91					86.2	9.7

16. Elaborate all the steps in (i) Bagging algorithm and (ii) Boosting algorithm.
17. What are Artificial Neural Networks? Explain the imitation of the human brain in the construction of ANN.
18. Explain the usage of Regression Analysis in Data mining and mention the various types.

SECTION – C

Answer any TWO Questions. Each carries twenty marks.

(2 X 20 = 40)

19. Explain in detail the CRISP for Data mining.
20. Explain classification, and the test metrics used to access the model.
21. Explain how the Bayes theorem and Naïve assumption is used in Naïve Bayes algorithm.
22. Following are the transactions made at the roadside vegetable stand

Transaction	Items Purchased
1	Broccoli, green peppers, corn
2	Asparagus, squash, corn
3	Corn, tomatoes, beans, squash
4	Green peppers, corn, tomatoes, beans
5	Beans, asparagus, broccoli
6	Squash, asparagus, beans, tomatoes
7	Tomatoes, corn
8	Broccoli, tomatoes, green peppers
9	Squash, asparagus, beans
10	Beans, corn
11	Green peppers, broccoli, beans, squash
12	Asparagus, beans, squash
13	Squash, corn, asparagus, beans
14	Corn, green peppers, tomatoes, beans, broccoli

Generate the frequent itemsets and find the support and confidence.

@@@@@@@