KUMARAGURU
college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 2023

(Regulation 2018)

Fifth Semester

COMMON TO ALL BRANCHES

U18MAR0203: Data Analytics

COURSE OUTCOMES

CO1: Ability to use technical skills in predictive and prescriptive modeling to support business decision-making

CO2: Demonstrate the ability to think critically in making decisions based on data and deep analytics

Time: Three Hours   **Maximum Marks: 100**

Answer all the Questions: -

PART A (10 x 2 = 20 Marks)
(Answer not more than 40 words)

|			
1.	Compare Discrete and Continuous attributes.	CO2	[K_2]
2.	What is Correlation? What are its benefits?	CO2	[K_1]
3.	Differentiate between Supervised and Unsupervised Learning.	CO1	[K_2]
4.	What are symmetric and asymmetric binary variables?	CO2	[K_1]
5.	Give the significance of confusion matrix.	CO2	[K_2]
6.	How to measure the distance between clusters?	CO2	[K_2]
7.	What is Entropy in Data Mining?	CO2	[K_1]
8.	List out a few application areas in Data Mining.	CO1	[K_1]
9.	Discuss the purpose of visualization in Data Analytics.	CO1	[K_2]
10.	List the various kinds of outliers.	CO2	[K_1]

Answer any FIVE Questions: -
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11.	a)	Discuss the various stages of Data Mining Process.	8	CO2	[K_2]
	b)	What are the different ways to fill up the missing values in a dataset? Explain	8	CO2	[K_2]
12.	a)	Explain any 5 plots used for analyzing data.	10	CO2	[K_2]

b)   6   CO2   [K3]
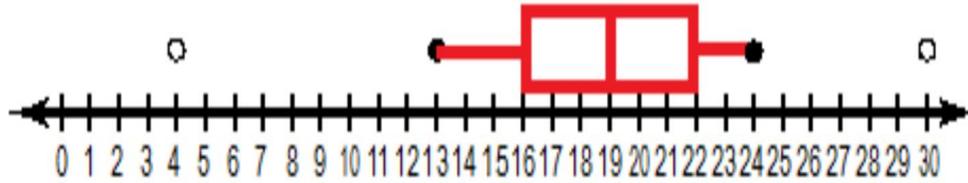
Identify the 5 number summary and any outliers depicted in the box plot above. Give the outliers from the above diagram.

13.

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

16   CO1   [K3]

Find the frequent item sets using Apriori algorithm and generate association rules on the above dataset. Assume that minimum support threshold (s = 33.33%) and minimum confident threshold (c = 60%).

14. a)   Given the Confusion Matrix for Cancer prediction,   10   CO1   [K3]

Actual Class\ Predicted Class	Cancer = Yes	Cancer = No	Total
Cancer = Yes	90	210	300
Cancer = No	140	9560	9700
Total	230	9770	10000

Calculate the following metrics:

    i.   Sensitivity

   ii.   Specificity

  iii.   Accuracy

  iv.   Precision

   v.   Recall

b)   Explain the Naïve Bayes Classifier model with an example.   6   CO1   [K2]

15. Consider the following dataset. Apply Decision Tree induction to determine whether to play Football or not?   16   CO1   [K3]

Outlook	Temperature	Humidity	Wind	Played football(yes/no)
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

16. Explain K-Means Clustering with suitable illustration. Discuss the drawbacks and the ways to overcome it.   16   CO1   [K2]

****