Hall Ticket No: | | | | | | | | | | |

Question Paper Code: 20MBAP435

# MADANAPALLE INSTITUTE OF TECHNOLOGY & SCIENCE, MADANAPALLE

(UGC-AUTONOMOUS)

**MBA II Year II Semester (R20) Supplementary End Semester Examinations, July - 2025**

## DATA MINING FOR BUSINESS DECISIONS

**Time: 3Hrs** **Max Marks: 60**

Attempt all the questions. All parts of the question must be answered in one place only.
**In Q.no 1 to 5 answer either Part A or Part B only. Q.no 6 which is a case study is compulsory**

| Q.No | Question | Marks | CO | BL |
|---|---|---|---|---|
| Q.1(A) | Write 5 examples of business tasks where data mining tools can be used effectively and explain how it can be done. (Hints: for example, detection of credit card fraud) | 10M | 1 | 4 |
| | **OR** | | | |
| Q.1(B) | Why you need to analyse business data? Explain how data mining techniques help in analysing business data to extract important information. | 10M | 1 | 3 |
| Q.2(A) | Explain step by step the method of developing decision tree classifier from a set of training data with an example. | 10M | 2 | 3 |
| | **OR** | | | |
| Q.2(B) | Why random choice of an attribute for splitting as the root node is not good? What method do we adopt for selection of root node to obtain an optimal decision tree? Define Gini Index, Entropy and Information Gain. | 10M | 2 | 3 |
| Q.3(A) | In the following samples, X1 and X2 are the attributes and Y is the class. Use the first 6 samples as training samples for a kNN classifier with k = 3. Decide the class (C1/C2) of the 7th test sample. | 10M | 3 | 3 |

| ID | X1 | X2 | Y |
|---|---|---|---|
| 1 | 1 | 0 | C1 |
| 2 | 2 | 1 | C1 |
| 3 | 1 | 2 | C1 |
| 4 | 5 | 2 | C2 |
| 5 | 6 | 3 | C2 |
| 6 | 5 | 4 | C2 |
| 7 | 2 | 2 | ? |

| | | | | |
|---|---|---|---|---|
| | **OR** | | | |
| Q.3(B) | State and explain Bayes' Theorem. What is the difference between Bayes' classifier and Naïve Bayes Classifier? Explain the working of Bayes' classifier. | 10M | 3 | 3 |
| Q.4(A) | What do you mean by dimensionality reduction? Write in details about any one method of dimensionality reduction. | 10M | 4 | 3 |
| | **OR** | | | |

| | | | | |
|---|---|---|---|---|
| **Q.4(B)** | Define cluster cohesion and cluster separation? Give the formula of within cluster sum of squares and between cluster sum of squares. Define MSE (Mean square error) and explain how the proper value of k is selected in k-Means algorithm | 10M | 4 | 3 |
| **Q.5(A)** | The following table shows the sales and advertisement expenditure of a business firm. Coefficient of correlation r= 0.9. Estimate the sales for a proposed advertisement expenditure of Rs. 10 crores using regression analysis. | 10M | 5 | 3 |

| | Sales in crores | Advertisement Expenditure(crores) |
|---|---|---|
| Mean | 40 | 6 |
| Standard deviation | 10 | 1.5 |

**OR**

| | | | | |
|---|---|---|---|---|
| **Q.5(B)** | For 5 pairs of observations the following results are obtained $\sum X=15$, $\sum Y=25$, $\sum X2 =55$, $\sum XY=83$ Find the equation of the lines of regression of Y on X and estimate the value of Y if X=5. | 10M | 5 | 3 |
| **Q.6** | **Case Study**<br>You are the owner of a shop selling health and beauty products. Explain how you use regression analysis to predict the profit of your shop from past data. What will be your independent variables and the dependent variable. | 10M | 5 | 4 |

*****END*****