Total No. of Questions: 20

Enrollment No : _____

MEDICAPS UNIVERSITY

Faculty of Management Studies

End Semester Examination May 2025

MS5EB02 Data Mining Techniques Using R

Programme	:	MBA	**Branch/Specialisation**	:	Business Analytics
Duration	:	3 hours	**Maximum Marks**	:	60

Note: *All questions are compulsory. Internal choices, if any, are indicated. Assume suitable data if necessary. Notations and symbols have their usual meaning.*

Section 1 (Answer all question(s))

		Marks	CO	BL
Q1.	What does CRISP-DM stand for?	1	1	1

- ○ Common Resource Interface Standard for Process
- ◉ Cross-Industry Standard Process for Data Mining
- ○ Customer Relationship and Sales Processing
- ○ None of the above

Q2.	Which data type is unstructured?	1	2	2

- ○ Sensor data
- ◉ Tweets
- ○ Relational tables
- ○ Temperature records

Q3.	What does EDA stand for?	1	2	1

- ○ Enhanced Data Analysis
- ◉ Exploratory Data Analysis
- ○ Efficient Data Aggregation
- ○ None of the above

Q4.	In k-means clustering, 'k' stands for-	1	2	2

- ○ Kernel value
- ◉ Number of centroids
- ○ Number of variables
- ○ Distance metric

Q5.	Which regression model is used for binary classification?	1	3	2

- ○ Linear Regression
- ◉ Logistic Regression
- ○ Multiple Regression
- ○ Polynomial Regression

Q6.	In R, the lm() function is used for-	1	3	2

- ○ Logistic regression
- ◉ Linear regression
- ○ Clustering
- ○ Decision trees

Q7.	Which of these is a dimensionality reduction technique?	1	3	1

- ○ SVM
- ◉ t-SNE
- ○ Decision Tree
- ○ K-means

Q8.	Which algorithm boosts weak learners into strong learners?	1	4	2

- ○ kNN
- ◉ Gradient Boosting
- ○ PCA
- ○ Naive Bayes

Q9.	Which R package is used for text mining?	1	5	1

- ◉ Tm
- ○ Arules
- ○ Lubridate
- ○ Forecast

Q10. ARIMA stands for- **1 5 1**

- (●) Auto-Regressive Integrated Moving Average
- ( ) Average Repeated Interval Moving Analysis
- ( ) Automated Regression Interpolation Model Analysis
- ( ) None of the above

Section 2 (Answer all question(s))

Marks CO BL

Q11. Define Data Mining. Discuss the CRISP-DM framework and explain its stages in detail. **4 1 1**

Rubric	Marks
Defination-2m Discuss crism-2m	4

Q12. (a) Explain the architecture of a data warehouse. What are fact and dimension tables? **6 1 1**

Rubric	Marks
architecture of data warehouse-3m Fact and Dimension Tables-3m	6

(OR)

(b) What is data preprocessing? Discuss data cleaning, integration, reduction, and transformation.

Rubric	Marks
data preprocessing-2m discuss data cleaning ,integration ,reduction and transfromation-4m	6

Section 3 (Answer all question(s))

Marks CO BL

Q13. What is Exploratory Data Analysis (EDA)? How is it performed in R? **4 2 3**

Rubric	Marks
Explain EDA -2m Perfromance in R -2m	4

Q14. (a) Describe the K-means clustering algorithm and its implementation in R. **6 2 1**

Rubric	Marks
kmean clustring with diagram-3m implimentation in R-3m	6

(OR)

(b) Explain association rule mining. How is Apriori algorithm implemented in R?

Rubric	Marks
Explain association rule mining-3m Apriori algorithm implemented in R-3m	6

Section 4 (Answer all question(s))

Marks CO BL

Q15. Explain linear regression and its implementation in R with an example. **4 3 4**

Rubric	Marks
Explain linear regression-2m implementation in R with an example-2m	4

Q16. (a) What is logistic regression? Explain how it is used for binary classification in R. **6 3 1**

Rubric	Marks
logistic regression-3m binary classification in R.-3m	6

(OR)

(b) Describe Random Forest and explain how it improves over a single decision tree.

Rubric	Marks
Random Forest-3m how it improves over a single decision tree-3m	6

Section 5 (Answer all question(s))

Marks CO BL

Q17. What is Support Vector Machine (SVM)? How is it implemented in R? **4 4 3**

Rubric	Marks
explain SVM-2m implemented in R-2m	4

Q18. (a) What is t-SNE? How does it help in visualizing high-dimensional data? **6 4 1**

Rubric	Marks
What is t-SNE-3m visualizing high-dimensional data-3m	6

(OR)

(b) Explain Principal Component Analysis and its implementation in R.

Rubric	Marks
Explain Principal Component Analysis (PCA)-3m implementation in R-3m	6

Section 6 (Answer all question(s))

Marks CO BL

Q19. Explain sentiment analysis in text mining. How is it implemented in R? **4 5 4**

Rubric	Marks
Explain sentiment analysis in text mining-2m implemented in R-2m	4

Q20. (a) Describe ARIMA model and its use in time series forecasting. **6 5 5**

Rubric	Marks
Describe ARIMA-3m use in time series forecasting-3m	6

(OR)

(b) Discuss ethical considerations in data mining.

Rubric	Marks
discuss consideration data mining.	6

******