

Q.6	Attempt any two:				
i.	Explain the steps involved in sentiment analysis using text mining and its implementation in R.	5	3	11	5
ii.	Discuss the application of data mining techniques in the healthcare industry with an example.	5	5	12	5
iii.	What is exponential smoothing? How is it used for time series forecasting? Provide examples of its implementation in R.	5	1	7	5

\*\*\*\*\*

Total No. of Questions: 6

Total No. of Printed Pages:4

Enrollment No.....



Faculty of Management Studies

End Sem Examination Dec 2024

MS5EB02 Data Mining Techniques Using R

Programme: MBA

Branch/Specialisation: Business Analytics

Duration: 3 Hrs.

Maximum Marks: 60

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

		Marks	BL	PO	CO	PSO
Q.1	i. Which of the following is an example of qualitative data? (a) Ratings (b) Quantity (c) Feedback (d) Revenue	1	1	2	1	
	ii. Which of the following data mining techniques is used to group similar data points into clusters? (a) Classification (b) Regression (c) Clustering (d) Association	1	1	4	1	
	iii. Which R function is commonly used to summarize numeric data in a data frame? (a) str() (b) summary() (c) head() (d) plot()	1	1	4	2	
	iv. After performing hierarchical clustering in R, which function is typically used to extract clusters? (a) hclust() (b) kmeans() (c) cutree() (d) dist()	1	3	5	2	
	v. Which function in R is primarily used to build a linear regression model? (a) lm() (b) glm() (c) plot() (d) summary()	1	4	2	3	
	vi. Which R package is commonly used for building decision trees using the CART algorithm?	1	3	1	3	

[2]

		(a) ggplot2	(b) rpart				
		(c) dplyr	(d) Caret				
vii.	Which kernel function is commonly used in Support Vector Machines (SVMs) for handling non-linear data?	<b>1</b>	2	6	4		
	(a) Linear						
	(b) Polynomial						
	(c) Radial Basis Function						
	(d) Sigmoid						
viii.	Which R package is widely used for implementing neural networks?	<b>1</b>	1	8	4		
	(a) nnet						
	(b) rpart						
	(c) Caret						
	(d) ggplot2						
ix.	Which R package is commonly used for text preprocessing and sentiment analysis	<b>1</b>	2	9	5		
	(a) tm						
	(b) Forecast						
	(c) Caret						
	(d) Zoo						
x.	In time series analysis, which function in R is used to fit an ARIMA model?	<b>1</b>	3	7	5		
	(a) auto.arima()						
	(b) arima()						
	(c) forecast()						
	(d) ts()						
Q.2	i. Differentiate between structured and unstructured data.	<b>2</b>	4	8	1		
	ii. What is the purpose of the knowledge discovery process? Define classification in data mining.	<b>3</b>	1	6	1		
	iii. Explain the concept of clustering in data mining and its significance. Also describe the architecture of a data warehouse with a focus on fact tables and dimension tables.	<b>5</b>	3		1		
OR	iv. Explain the process of data pre-processing in detail, highlighting data cleaning, integration, and reduction. What are the different types of data mining systems?	<b>5</b>	3	10	1		

[3]

Q.3	i. What is a data frame in R, and why is it important in data analysis?	<b>2</b>	1	13	2		
	ii. Explain the steps of the K-Means Clustering algorithm and demonstrate its implementation in R with an example.	<b>8</b>	3	9	2		
OR	iii. Explain hierarchical clustering in detail, including the agglomerative and divisive approaches, with a discussion on how dendrograms are used to interpret results.	<b>8</b>	4	11	2		
Q.4	i. Explain the key assumptions of linear regression. How these assumptions are validated in R.	<b>3</b>	3	10	3		
	ii. Describe the process of building a decision tree using the CART algorithm. Include an explanation of pruning and how the model is evaluated in R.	<b>7</b>	2	5	3		
OR	iii. Compare and contrast Random Forest and Decision Trees. Discuss how feature importance is evaluated in Random Forest using R.	<b>7</b>	4	2	3		
Q.5	i. Explain the role of kernel functions in Support Vector Machines (SVM). Provide examples of commonly used kernel functions.	<b>4</b>	5	8	4		
	ii. Explain the back propagation algorithm used in neural networks. How is it implemented in R for training a model?	<b>6</b>	3	10	4		
OR	iii. Describe the concept of boosting in Gradient Boosting Machines (GBM). Also give its advantages over traditional decision tree methods.	<b>6</b>	5	12	4		

**Marking Scheme****MS5EB02 Data Mining Techniques Using R-(T)**

Q.1	i)	c) <b>Feedback</b>	<b>1</b>	OR	Q.4	i.	Explain the key assumptions of linear regression. – 1.5 marks How these assumptions are validated in R– 1.5 marks	<b>3</b>			
	ii)	c) <b>Clustering</b>	<b>1</b>			ii.	Describe the process of building a decision tree using the CART algorithm. – 3.5 marks Include an explanation of pruning and how the model is evaluated in R– 3.5 marks	<b>7</b>			
	iii)	b) <b>summary()</b>	<b>1</b>			OR	iii.	Compare and contrast Random Forest and Decision Trees.- 4 marks Discuss how feature importance is evaluated in Random Forest using R. – 3 marks	<b>7</b>		
	iv)	c) <b>cutree()</b>	<b>1</b>								
	v)	a) <b>lm()</b>	<b>1</b>			Q.5	i.	Explain the role of kernel functions in Support Vector Machines (SVM). - 2 marks Provide examples of commonly used kernel functions. – 2 marks	<b>4</b>		
	vi)	b) <b>rpart</b>	<b>1</b>					ii.	Explain the back propagation algorithm used in neural networks. – 4 Marks How is it implemented in R for training a model – 2 marks	<b>6</b>	
	vii)	c) <b>Radial Basis Function</b>	<b>1</b>						OR	iii.	Discuss the concept of boosting in Gradient Boosting Machines (GBM). – 4 marks Also give its advantages over traditional decision tree methods – 2 marks
	viii)	a) <b>nnet</b>	<b>1</b>								
	ix)	a) <b>tm</b>	<b>1</b>			Q.6	Attempt any two: (5 marks each)				
	x)	b) <b>arima()</b>	<b>1</b>				i.	Explain the steps involved in sentiment analysis using text mining and its implementation in R. Steps- 3 marks, implementation- 2 marks	<b>5</b>		
Q.2	i.	Structured data – 2 marks Unstructured data – 2 marks	<b>2</b>	OR	iii.	Discuss the application of data mining techniques in the healthcare industry with an example Application- 2.5 marks, example- 2.5 marks	<b>5</b>				
	ii.	What is the purpose of the Knowledge Discovery Process. – 1.5 mrks Define classification in data mining. – 1.5 marks	<b>3</b>			iii.	What is exponential smoothing, and how is it used for time series forecasting. Provide examples of its implementation in R.	<b>5</b>			
Q.3	iii.	Explain the concept of clustering in data mining and its significance. – 2.5 marks Also describe the architecture of a data warehouse with a focus on fact tables and dimension tables.- 2.5 marks	<b>5</b>	OR	iv.	Explain the process of data pre-processing in detail, data cleaning, integration, and reduction. – 3 marks What are the different types of data mining systems . – 2 marks	<b>5</b>				
	iv.	Explain the process of data pre-processing in detail, data cleaning, integration, and reduction. – 3 marks What are the different types of data mining systems . – 2 marks	<b>5</b>								
Q.4	i.	What is a data frame in R – 1 mark why is it important in data analysis – 1 mark	<b>2</b>	OR	iii.	What is exponential smoothing, and how is it used for time series forecasting. Provide examples of its implementation in R.	<b>5</b>				

[2]

exponential smoothing- 1.5 mark  
how time series- 1.5 mark  
implementation- 2 mark

\*\*\*\*\*

[3]