

**Q.6**      Attempt any two:

- i. Explain the steps involved in sentiment analysis using text mining and its implementation in R.
  - ii. Discuss the application of data mining techniques in the healthcare industry with an example.
  - iii. What is exponential smoothing? How is it used for time series forecasting? Provide examples of its implementation in R.

5            3            11            5

5 5 12 5

5 1 7 5

\* \* \* \*



Faculty of Management Studies

End Sem Examination Dec 2024

MS5EB02 Data Mining Techniques Using R

Programme: MB

Branch/Specialisation: Business Analytics

Maximum Marks: 60

**Duration: 3 Hrs.**

---

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.

*Total No. of Questions: 6*

*Total No. of Printed Pages:4*

**Enrollment No.....**



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

Q.1	i) c) <b>Feedback</b>	1	ii. Explain the steps of the K-Means Clustering algorithm – 4 marks demonstrate its implementation in R with an example – 4 marks	8
	ii) c) <b>Clustering</b>	1	OR     iii. Explain hierarchical clustering in detail, including the agglomerative and divisive approaches, with a discussion on how dendograms are used to interpret results. – 8 marks	8
	iii) b) <b>summary()</b>	1		
	iv) c) <b>cutree()</b>	1		
	v) a) <b>lm()</b>	1		
	vi) b) <b>rpart</b>	1		
	vii) c) <b>Radial Basis Function</b>	1		
	viii) a) <b>nnet</b>	1		
	ix) a) <b>tm</b>	1		
	x) b) <b>arima()</b>	1		
Q.2	i. Structured data – 2 marks Unstructured data – 2 marks	2		
	ii. What is the purpose of the Knowledge Discovery Process. – 1.5 mrks	3		
	Define classification in data mining. – 1.5 marks			
	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	5		
	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	5		
Q.3	i. What is a data frame in R – 1 mark why is it important in data analysis – 1 mark	2	Q.4     i. Explain the key assumptions of linear regression. – 1.5 marks How these assumptions are validated in R– 1.5 marks	3
			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	7
			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	7
			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	4
			ii. Explain the back propagation algorithm used in neural networks. – 4 Marks How is it implemented in R for training a model – 2 marks	6
			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	6
			Q.6     Attempt any two: (5 marks each)	
			i. Explain the steps involved in sentiment analysis using text mining and its implementation in R. Steps- 3 marks, implementation- 2 marks	5
			ii. Discuss the application of data mining techniques in the healthcare industry with an example Application- 2.5 marks, example- 2.5 marks	5
			iii. What is exponential smoothing, and how is it used for time series forecasting. Provide examples of its implementation in R.	5

[2]

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

\*\*\*\*\*

[3]

P.T.O.