Q.4	i.	What is Euclidean Distance and why is it important in clustering algorithms?	**3**	2	10	3
	ii.	Explain the K-Means clustering algorithm in detail.	**7**	2	10	3
OR	iii.	What are scaling and standardization in the context of clustering, and why are they necessary?	**7**	2	10	3
Q.5	i.	What is the role of the Sigmoid curve in binary classification? How does it affect the model's decision-making process?	**4**	2	10	4
	ii.	Explain the concept of Decision Trees in classification. Discuss the advantages and limitations of Decision Trees compared to Rule-Based Classification models.	**6**	2	10	4
OR	iii.	What is Multivariate Logistic Regression? Explain its application and how it differs from simple logistic regression.	**6**	2	10	4
Q.6		Attempt any two:				
	i.	Explain the concept of blockchain technology. What are the main advantages of using a decentralized ledger system over a centralized one?	**5**	2	10	5
	ii.	Describe the process of a blockchain transaction from initiation to confirmation. How is transaction management handled within a blockchain network?	**5**	2	10	5
	iii.	Explain the role of public and private keys, digital signatures, and hash values in ensuring security and data integrity.	**5**	2	10	5

******

Total No. of Questions: 6 *Total No. of Printed Pages:4*

Enrollment No.....

Faculty of Management Studies
End Sem Examination Dec 2024
MS5EB05
Introduction to ML & Blockchain Technology

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 methods is used to determine the line of best fit in simple linear regression? (a) Least squares method (b) Maximum likelihood estimation (c) Gradient descent (d) Principal component analysis	**1**	1	10	1	
	ii.	Which method is commonly used to handle multicollinearity in multivariate linear regression models? (a) Principal Component Analysis (PCA) (b) Ordinary Least Squares (OLS) (c) Stepwise Regression (d) Ridge Regression	**1**	1	10	1	
	iii.	What is the primary purpose of time series forecasting? (a) To predict future values based on past data patterns (b) To find correlations between multiple variables (c) To create data for model training (d) To classify data into categories	**1**	1	10	2	

iv.	In lag analysis, what does a lag value represent? (a) The amount of time between observations in a time series (b) The number of periods for which past data is used to predict future values (c) The trend component of the time series (d) The seasonal pattern in the data	**1**	1	10	2
v.	Which of the following is a common method used to measure the distance between data points in clustering? (a) Manhattan distance (b) Euclidean distance (c) Jaccard index (d) Pearson correlation	**1**	1	10	3
vi.	Which of the following steps is involved in the K-Means clustering algorithm? (a) Selecting an optimal number of clusters (b) Assigning each data point to the nearest centroid (c) Re-calculating the centroid based on the data points in the cluster (d) All of these	**1**	1	10	3
vii.	What is the key difference between regression and classification? (a) Regression predicts categorical outcomes, while classification predicts continuous outcomes (b) Regression predicts continuous outcomes, while classification predicts categorical outcomes (c) Both predict continuous outcomes (d) Both predict categorical outcomes	**1**	1	10	4
viii.	In the context of binary classification, what does the sigmoid curve represent? (a) The probability of the class belonging to class 0 (b) The probability of the class belonging to class 1	**1**	1	10	4

	(c) The linear decision boundary between classes (d) The decision tree for classification					
ix.	What is a digital signature in the context of blockchain? (a) A private key that encrypts transaction data (b) A hash value that represents the integrity of a block (c) A mechanism to validate the authenticity of a transaction or message using the sender's private key (d) A public key used for data verification	**1**	1	10	5
x.	Which of the following consensus mechanisms is commonly used in Blockchain 1.0 (Bitcoin)? (a) Proof-of-Stake (b) Proof-of-Work (c) Byzantine Fault Tolerance (d) Delegated Proof-of-Stake	**1**	1	10	5
Q.2 i.	Write the equation for simple linear regression and explain the meaning of each term.	**2**	2	10	1
ii.	Why is multivariate regression used over simple linear regression? Illustrate with an example.	**3**	2	10	1
iii.	Discuss the steps involved in building a Multivariate Linear Regression (MLR) model and the importance of feature selection.	**5**	2	10	1
OR iv.	How do you interpret the regression coefficients in a simple linear regression model? Provide an example using a real-world business scenario.	**5**	2	10	1
Q.3 i.	Define the term overfitting in the context of time series forecasting.	**2**	1	10	2
ii.	Provide a detailed explanation of the ARIMA Model (Auto Regressive Integrated Moving Average) and its components.	**8**	2	10	2
OR iii.	Explain Lag Analysis in time series forecasting and describe the steps involved in conducting it using XLMiner.	**8**	2	10	2

Scheme of Marking

Introduction to ML and Blockchain Technology-(T)-(MS5EB05)

Q.1	i)	a). Least Squares Method	**1**
	ii)	d) Ridge Regression	**1**
	iii)	a) To predict future values based on past data patterns	**1**
	iv)	b) The number of periods for which past data is used to predict future values	**1**
	v)	b) Euclidean Distance	**1**
	vi)	d) All of the above	**1**
	vii)	b) Regression predicts continuous outcomes, while classification predicts categorical outcomes	**1**
	viii)	b) The probability of the class belonging to class 1	**1**
	ix)	c) A mechanism to validate the authenticity of a transaction or message using the sender's private key	**1**
	x)	b) Proof-of-Work	**1**
Q.2	i.	Equation-1 mark, explanation-1 mark	**2**
	ii.	Reason-2 marks, example-1 marks	**3**
	iii.	Steps – 3 marks, importance 2 marks	**5**
OR	iv.	Interpretation- 3 marks, example- 2 marks	**5**
Q.3	i.	Definition- 2 marks	**2**
	ii.	Explanation of ARIMA components – 5 marks, Example of ARIMA application – 3 marks	**8**
OR	iii.	Explanation of Lag Analysis (3 Marks), **Steps in Conducting Lag Analysis using XLMiner (5 Marks):**	**8**
Q.4	i.	Definition- 2 marks, importance-1 marks	**3**
	ii.	**K-Means Algorithm (5 marks):** Explanation of Convergence (2 marks)	**7**
OR	iii.	Scaling and Standardization (3 marks) Application to clustering (4 marks)	**7**
Q.5	i.	Role- 2 marks, affect-2 marks	**4**
	ii.	Explanation-2 marks, advantages- 2 marks, disadvantage 2 marks	**6**
OR	iii.	Explanation-2 marks, application 2 marks, comparision- 2 marks	**6**
Q.6			
	i.	Concept-3 marks, advantage-2 marks	**5**
	ii.	Process-3 marks, handling transcation-2 marks	**5**
	iii.	Public and private key-2 marks Digital signature-2 marks Hash value -1 marks	**5**

******