Enrollment No......................................

Faculty of Engineering
End Sem (Odd) Examination Dec-2022
IT3ED06 Predictive Modeling & Data Visualization
Programme: B.Tech.   Branch/Specialisation: IT

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.

Q.1 i.   Forming definitions from examples of concepts is the process of-   **1**
          (a) Deduction   (b) Induction
          (c) Conjunction   (d) Generation

     ii. Labelled data is used to train machine in which of the following   **1**
          Learning technique.
          (a) Supervised learning   (b) Unsupervised learning
          (c) Reinforcement learning   (d) None of these

     iii. The best fit line in Linear regression for the given dataset can be **1**
          found using which of the following methods.
          (a) Logarithmic methods   (b) Maximum likelihood method
          (c) Exponential method   (d) Least square Error

     iv. The optimization problem in case of simple linear regression is   **1**
          basically a-
          (a) Maximization problem   (b) Minimization problem
          (c) Overfitting problem   (d) Underfitting problem

     v.   Which is needed by K-means clustering?   **1**
          (a) Defined distance metric
          (b) Number of clusters
          (c) Initial guess as to cluster centroids
          (d) All of these

     vi. Which clustering technique requires a merging approach?   **1**
          (a) Partitional   (b) Hierarchical
          (c) Naive bayes   (d) None of these

P.T.O.

vii. The ___ function applies one or more conditions in the specified range **1** of data and returns only those which fulfill all of the conditions.
(a) COUNTIFS   (b) COUNT
(c) IF   (d) None of these

viii. Which of the intricate techniques is not used for data visualization? **1**
(a) Bullet graphs   (b) Bubble clouds
(c) Fever maps   (d) Heat maps

ix. Which of the following plots are used to check if a data set or time **1** series is random?
(a) Lag   (b) Random   (c) Lead   (d) None of these

x. Which Library of python is use to plot charts such as line, bar, scatter, **1** histograms?
(a) Pandas   (b) Matplotlib
(c) Numpy   (d) None of these

Q.2   i. What is predictive modelling? Explain the with an example. **2**
    ii. Write a difference between supervised and unsupervised learning. **3**
    iii. What is reinforcement? Explain the characteristics of reinforcement **5** learning.
OR   iv. What is semi-supervised learning? Explain the benefits and limitations **5** of semi-supervised learning.

Q.3   i. Explain cross-validation with suitable example. **2**
    ii. Explain regression with an example. What are the five assumptions of **8** linear multiple regression?
OR   iii. Explain the following: **8**
(a) Loss function   (b) Cost function
(c) Gradient descent   (d) Response variable

Q.4   i. Write a comparison between classification and clustering. **3**
    ii. What is K-Means clustering? Explain its algorithm and also write its **7** advantages.
OR   iii. How many binary classifiers would you need to implement one-vs-all **7** for three classes? How does it work?

Q.5   i. How can we control the height or width of a data visualization in a 2D **4** cartesian coordinate system? Explain with an example.
    ii. What is the difference between count histogram, relative frequency **6** histogram, cumulative frequency histogram and density histogram?
OR   iii. Defined following question with suitable example: **6**
(a) Boxplot   (b) Heat map
(c) Scatter plot

Q.6   Attempt any two:
    i. Which different colour schemes are used in data visualization? **5** Explain anyone with an example.
    ii. How can we set the attributes of a plot that are not semantically **5** mapped using scatterplot function explain with an example?
    iii. Write a short notes on: **5**
(a) Histogram   (b) Whisker plot

*****

Marking Scheme
IT3ED06 Predictive Modelling & Data Visualization

Q.1	i)	B		1
	ii)	A		1
	iii)	D		1
	iv)	B		1
	v)	D		1
	vi)	B		1
	vii)	A		1
	viii)	C		1
	ix)	A		1
	x)	B		1
Q.2	i.	Predictive modelling		1.5
		example		.5
	ii.	Difference between supervised and unsupervised learning.		.5 mark for each
	iii.	What is reinforcement		2
		characteristics of reinforcement learning		3
OR	iv.	Semi-supervised Learning?		2
		Benefits		1.5
		limitations of Semi-supervised Learning.		1.5
Q.3	i.	Explain cross-validation with suitable example.		2
	ii.	Explain regression with an example.		3
		What are the five assumptions of linear multiple regression?		5
OR	iii.	Explain Following:		2
		loss function		Marks for each
		cost function		
		gradient descent		
		response variable		
Q.4	i.	Write a Comparison between Classification and Clustering.		.5 for each

	ii.	What is K-Means Clustering?	2
		algorithm	3
		advantages.	2
OR	iii.	How many binary classifiers would you need to implement one-vs-all for three classes?	5
		How does it work?	2
Q.5	i.	How can we control the height or width of a data visualization in a 2D Cartesian coordinate system explain with an example?	4
	ii.	What is the difference between count histogram, relative frequency histogram, cumulative frequency histogram and density histogram?	1 marks for each
OR	iii.	Defined Following Question with suitable example:	2 marks for each
		a) Boxplot	
		b) heat map	
		c) Scatter plot	
Q.6		Attempt any two:	
	i.	Which different colour schemes are used in data visualization? Explain anyone with an example.	3
			2
	ii.	How can we set the attributes of a plot that are not semantically mapped using scatterplot function explain with an example?	3
			2
	iii.	Short notes on:	2.5 For each
		Histogram	
		Whisker plot	

******