Total No. of Questions: 6 *Total No. of Printed Pages:3*

Enrollment No.....

Faculty of Engineering
End Sem Examination Dec-2023
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. Assume suitable data if necessary. Notations and symbols have their usual meaning.

Q.1 i. What is the primary objective of predictive modeling? **1**

(a) To analyze historical data
(b) To make predictions about future outcomes
(c) To create interactive visualizations
(d) To develop machine learning algorithms

ii. In which of the following applications is machine learning commonly used for natural language processing? **1**

(a) Weather forecasting (b) Inventory management
(c) Sentiment analysis (d) Bridge construction

iii. In the context of machine learning, what does the cost function measure? **1**

(a) Model accuracy
(b) Model complexity
(c) The error between predicted and actual values
(d) The number of features in the model

iv. Which of the following best describes underfitting in a machine learning model? **1**

(a) The model fits the training data well but doesn't generalize to new data
(b) The model fits the training data and generalizes well to new data
(c) The model is overly complex and fits the training data too closely
(d) The model is too simple and doesn't capture the underlying patterns in the data

v. Which model evaluation technique assesses the performance of a classification model by considering true positives, true negatives, false positives, and false negatives? **1**
(a) Mean Absolute Error (MAE)
(b) Root Mean Square Error (RMSE)
(c) Confusion Matrix
(d) F1 Score

vi. In logistic regression, what type of output does the model produce for binary classification tasks? **1**
(a) Continuous numeric values
(b) Probabilities between 0 and 1
(c) Discrete class labels
(d) Cluster assignments

vii. Which type of data visualization is most suitable for comparing the total sales of different products in a store? **1**
(a) Bar plot (b) Dot plot (c) Heat map (d) Histogram

viii. A Quantile-Quantile (Q-Q) plot is used primarily for: **1**
(a) Comparing the distributions of two variables
(b) Visualizing hierarchical data structures
(c) Representing categorical data
(d) Creating line charts

ix. When working with network graphs and graphs in Python, which library is commonly used? **1**
(a) Seaborn (b) Matplotlib
(c) Bokeh (d) Networkx

x. Which Python library is built on top of Matplotlib and provides a high-level interface for creating statistical data visualizations? **1**
(a) Seaborn (b) Matplotlib (c) Bokeh (d) Folium

Q.2 i. Describe the fundamental principles of a machine learning approach. **2**

ii. What is the primary difference between semi-supervised learning and reinforcement learning, and in what contexts are they typically applied? **3**

iii. Explain the key steps involved in the process of predictive modelling and how it is applied in solving real-world problems. Provide examples to illustrate your answer. **5**

OR iv. Explain the role of training data and test data in the development of machine learning models. How does the appropriate splitting of data into these sets impact the model's performance. **5**

Q.3 i. What is simple linear regression, and what is its primary goal in data analysis? **2**

ii. Explain the concept of gradient descent and its role in optimizing machine learning models. **8**

OR iii. Define overfitting and underfitting. Discuss why finding the right balance is critical in machine learning. **8**

Q.4 i. How does a decision tree work in a classification task? **3**

ii. Compare and contrast hierarchical clustering and K-means clustering in detail. **7**

OR iii. Discuss the various model evaluation techniques used in classification tasks. **7**

Q.5 i. What is data visualization, and why is it important in conveying information effectively? **4**

ii. Compare and contrast bar plots, dot plots, and heat maps as visualization methods for representing quantities. **6**

OR iii. Detail the role of histograms and density plots in visualizing data distributions with example. **6**

Q.6 Attempt any two:

i. Explain the concept of multiple plots and subplots in Python. How can they be used to create complex and informative visualizations? **5**

ii. Describe the primary purpose of Networkx in Python. How it aids in visualizing network graphs. **5**

iii. Write a short note on: **5**
(a) Box plot (b) Scatter plot

******

Marking Scheme
IT3ED06- Predictive Modeling & Data Visualization

Q.1	i)	b) To make predictions about future outcomes		**1**
	ii)	c) Sentiment analysis		**1**
	iii)	c) The error between predicted and actual values		**1**
	iv)	d) The model is too simple and doesn't capture the underlying patterns in the data.		**1**
	v)	c) Confusion Matrix		**1**
	vi)	b) Probabilities between 0 and 1		**1**
	vii)	a) Bar plot		**1**
	viii)	a) Comparing the distributions of two variables		**1**
	ix)	d) Networkx		**1**
	x)	a) Seaborn		**1**
Q.2	i.	Principles of a machine learning approach.	-2 mark	**2**
	ii.	Definition semi-supervised learning	-1 mark	**3**
		Definition reinforcement learning	-1 mark	
		Example	-1 mark	
	iii.	key steps	-2 mark	**5**
		real-world solution example	-3 mark	
OR	iv.	Training data and test data	-2 mark	**5**
		Working with example	-3 mark	
Q.3	i.	Definition	-1 mark	**2**
		Goal	-1 mark	
	ii.	Definition	-2 mark	**8**
		Concept	-3 mark	
		Defining Role	-3 mark	
OR	iii.	Overfitting Definition	-3 mark	**8**
		Underfitting Definition	-3 mark	
		Explanation	-2 mark	
Q.4	i.	Decision Tree	-1 mark	**3**
		Working with classification	-2 mark	
	ii.	hierarchical clustering	-2 mark	**7**
		K-means clustering	-2 mark	
		Comparison	-3 mark	
OR	iii.	Definition	-3 mark	**7**
		Explanation	-4 mark	
Q.5	i.	Data Visualization	-2 mark	**4**
		Importance	-2 mark	
	ii.	bar plots	-2 mark	**6**
		dot plots	-2 mark	
		heat maps	-2 mark	
OR	iii.	Histograms with example	-3 mark	**6**
		density plots with example	-3 mark	
Q.6		Attempt any two:		
	i.	Multiple plots and subplots	-2 mark	**5**
		Example	-3 mark	
	ii.	Define Networks	-2 mark	**5**
		Graph representation	-3 mark	
	iii.	Write a short note on: (a) Box plot	-2.5 mark	**5**
		(b) Scatter plot	-2.5 mark	

******