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

**Enrollment No.....**

Faculty of Engineering

End Sem Examination Dec 2024

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.

| | | | Marks | BL | PO | CO | PSO |
|---|---|---|---|---|---|---|---|
| Q.1 | i. | What is the primary goal of machine learning?<br>(a) To create static algorithms<br>(b) To visualize data<br>(c) To enable systems to learn from data<br>(d) To perform manual data entry | **1** | 1 | 1 | 1 | |
| | ii. | Which of the following is a type of supervised learning?<br>(a) K-Means clustering<br>(b) Hierarchical clustering<br>(c) Principal component analysis<br>(d) Decision trees | **1** | 1 | 1 | 1 | |
| | iii. | What does a confusion matrix provide?<br>(a) Summary of clustering performance<br>(b) Comparison of actual vs. predicted classifications<br>(c) Visualization of data distributions<br>(d) Evaluation of regression accuracy | **1** | 1 | 2 | 1 | |
| | iv. | In a binary classification problem, what does a false positive indicate?<br>(a) Correctly predicted positive case<br>(b) Incorrectly predicted positive case<br>(c) Incorrectly predicted negative case<br>(d) Correctly predicted negative case | **1** | 1 | 2 | 1,2 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| v. | Which algorithm is commonly used for clustering?<br>(a) K-Means<br>(b) Linear Regression<br>(c) Support Vector Machine<br>(d) None of these | **1** | 1 | 3 | 1 | | |
| vi. | What is the main goal of clustering?<br>(a) To predict an outcome<br>(b) To group similar data points together<br>(c) To reduce dimensionality<br>(d) To classify data into predefined categories | **1** | 1 | 3 | 1,2 | | |
| vii. | What does the term "R-squared" represent in regression analysis?<br>(a) The degree of correlation between variables<br>(b) The proportion of variance explained by the model<br>(c) The error rate of the predictions<br>(d) All of these | **1** | 1 | 4 | 1 | | |
| viii. | Which type of regression is used when the relationship between variables is not linear?<br>(a) Polynomial regression<br>(b) Multiple regression<br>(c) Simple linear regression<br>(d) None of these | **1** | 1 | 4 | 1 | | |
| ix. | Which of the following is a good practice when creating data visualizations?<br>(a) Using too many colors<br>(b) Overloading the chart with information<br>(c) Keeping the design simple and focused<br>(d) Using complex graphic | **1** | 1 | 5 | 1 | | |
| x. | What is the purpose of a scatter plot?<br>(a) To show distributions of categories<br>(b) To visualize the relationship between two continuous variables<br>(c) To summarize data using central tendency<br>(d) To represent hierarchical data | **1** | 1 | 5 | 1 | | |



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Q.2 | i. | Differentiate training data and test data. | **2** | 2 | 1 | 1 | |
| | ii. | What is reinforcement learning? | **3** | 1 | 1 | 1 | |
| | iii. | What is unsupervised learning? How does it differ from supervised learning? | **5** | 2 | 1 | 1,2 | |
| OR | iv. | Explain machine learning and enlist the application area of machine learning. | **5** | 3 | 1 | 1,2 | |
| Q.3 | i. | What is simple linear regression? | **3** | 1 | 2 | 1 | |
| | ii. | How can you evaluate the performance of a multiple regression model? What metrics would you use? | **7** | 2 | 2 | 1 | |
| OR | iii. | Explain the difference between overfitting and underfitting with example. | **7** | 1 | 2 | 1,2 | |
| Q.4 | i. | What is K-Means clustering? | **4** | 1 | 3 | 1 | |
| | ii. | What is the main difference between classification and clustering? Provide a brief explanation of each. | **6** | 2 | 3 | 1 | |
| OR | iii. | What is the role of the K value in K-Nearest Neighbors (KNN) classification? How does changing k affect the model's performance? | **6** | 2 | 3 | 1 | |
| Q.5 | i. | What is heatmap? How it can be useful in visualizing data? | **4** | 1 | 4 | 1 | |
| | ii. | What is a box plot? What information does it convey about a dataset? | **6** | 1 | 4 | 1 | |
| OR | iii. | Write short note on Q-Q Plot. | **6** | 1 | 4 | 1,2 | |
| Q.6 | | Write a short note on any two | | | | | |
| | i. | Whisker plot and bar plot | **5** | 1 | 5 | 1 | |
| | ii. | Networkx | **5** | 1 | 5 | 1 | |
| | iii. | Matplotlib and seaborn | **5** | 1 | 5 | 1 | |

******

## Marking scheme
## IT3ED06 Predictive Modeling & Data Visualization

| | | | Marks |
|---|---|---|---|
| Q.1 | i. | (c) To enable systems to learn from data | **1** |
| | ii. | (d) Decision trees | **1** |
| | iii. | (b) Comparison of actual vs. predicted classifications | **1** |
| | iv. | (b) Incorrectly predicted positive case | **1** |
| | v. | (a) K-Means | **1** |
| | vi. | (b) To group similar data points together | **1** |
| | vii. | (a) The proportion of variance explained by the model | **1** |
| | viii. | (a) Polynomial regression | **1** |
| | ix. | (c) Keeping the design simple and focused | **1** |
| | x. | (b) To visualize the relationship between two continuous variables | **1** |
| Q.2 | i. | Differentiate training data and test data. | **2** |
| | ii. | What is reinforcement learning? | **3** |
| | iii. | What is unsupervised learning 3M<br>How does it differ from supervised learning 2M | **5** |
| OR | iv. | Machine learning 3M<br>application area of machine learning.2M | **5** |
| Q.3 | i. | What is simple linear regression? | **3** |
| | ii. | How can you evaluate the performance of a multiple regression model? What metrics would you use? | **7** |
| OR | iii. | Explain the difference between overfitting and underfitting with example. | **7** |
| Q.4 | i. | What is K-Means clustering? | **4** |
| | ii. | difference between classification and clustering 3M<br>Provide a brief explanation of each. 3M | **6** |
| OR | iii. | K value in K-Nearest Neighbors (KNN) classification 3M<br>How does changing k affect the model's performance 3M | **6** |
| Q.5 | i. | What is heatmap? How it can be useful in visualizing data? | **4** |
| | ii. | box plot 3M<br>information does it convey about a dataset? 3M | **6** |
| OR | iii. | Write short note on Q-Q Plot. | **6** |
| Q.6 | | Write a short note on any two | |
| | i. | Whisker plot and bar plot | **5** |
| | ii. | Network x | **5** |
| | iii. | Matplotlib and seaborn | **5** |

******