

Enrollment No.....



Faculty of Engineering
End Sem Examination Dec-2023

ME3CO31 Data Science for Mechanical Engineers

Programme: B.Tech.

Branch/Specialisation: ME

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 purpose of data pre-processing in data analysis? **1**
 (a) To draw conclusions from the data
 (b) To visualize data effectively
 (c) To clean and prepare data for analysis
 (d) To create data visualizations
- ii. Which EDA technique is used to examine the relationships between two continuous variables? **1**
 (a) Histogram (b) Scatter plot
 (c) Bar chart (d) Pie chart
- iii. In a one-way ANOVA, how many groups are typically compared? **1**
 (a) Two (b) Three (c) Four (d) More than three
- iv. What is the purpose of hypothesis testing in statistics? **1**
 (a) To make inferences about population parameters
 (b) To prove that the null hypothesis is true
 (c) To confirm the alternative hypothesis
 (d) To summarize the data
- v. What does the seaborn library in Python provide in addition to Matplotlib? **1**
 (a) Machine learning algorithms
 (b) Statistical data analysis functions
 (c) Data preprocessing capabilities
 (d) High-level, aesthetically pleasing statistical data visualization
- vi. What is the range of possible values for the Pearson correlation coefficient? **1**
 (a) -1 to 1 (b) 0 to 1 (c) $-\infty$ to ∞ (d) -1 to 0

[2]

- vii. Which evaluation metric is used to assess the performance of a binary classification model when the goal is to minimize false negatives? **1**
(a) Accuracy (b) Precision (c) Recall (d) F1-score
- viii. Which of the following is an example of a clustering algorithm? **1**
(a) Linear Regression (b) K-Means
(c) Naive Bayes (d) Random Forest
- ix. Which machine learning technique is commonly used for predictive maintenance to identify patterns in historical data? **1**
(a) Linear regression
(b) Decision trees
(c) K-Means clustering
(d) Artificial neural networks
- x. What is the primary advantage of using data-driven techniques for optimization over traditional methods? **1**
(a) Data-driven techniques are faster
(b) Data-driven techniques can handle complex, high-dimensional data
(c) Data-driven techniques do not require data
(d) Data-driven techniques are less accurate
- Q.2 i. What is data science? Give its application in the field of mechanical engineering. **2**
ii. What is Exploratory Data Analysis (EDA)? Explain EDA in brief with an example. **3**
iii. How do missing values impact data analysis? What are some strategies for handling missing data during the data cleaning process? **5**
- OR iv. Describe the types of data that are best represented using different types of charts or graphs. Provide examples of situations where each type is appropriate. **5**
- Q.3 i. What are descriptive statistics? **2**
ii. Explain the principles of Analysis of Variance (ANOVA) and its significance in experimental research and provide detailed steps for conducting a one-way ANOVA. **8**
- OR iii. Discuss the factors that affect the width of a confidence interval. How does variability in the data and confidence level influence the width? **8**
- Q.4 i. Give difference between Matplotlib and Seaborn Libraries of Python. **3**
ii. Explain the concept of correlation analysis. How does it help in understanding the relationship between variables? **7**

[3]

- OR iii. Describe the concept of dimensionality reduction and its role in visualizing high-dimensional data. Explain PCA, t-SNE and provide examples of how they can be used to visualize multidimensional data. **7**
- Q.5 i. What is machine learning? Write difference between artificial intelligence, machine learning and deep learning. **4**
ii. Discuss the logistic regression model. Explain how it differs from linear regression and its use in binary and multiclass classification. Provide an example of a classification problem where logistic regression could be applied. **6**
- OR iii. Explore the real-world applications of unsupervised learning, combining both clustering and dimensionality reduction techniques. Provide examples of scenarios where these techniques can be used together to gain insights from data. **6**
- Q.6 Attempt any two:
i. What are the key data sources and sensors used in condition monitoring for predictive maintenance? Also explain common challenges in collecting and managing data for predictive maintenance and condition monitoring. **5**
ii. Explain the role of machine learning and data analytics in the context of fault diagnosis and failure prediction. Provide examples of industries where these techniques have been successfully applied. **5**
iii. Explain the importance of model interpretability and transparency in data-driven design optimization. How can stakeholders make informed decisions when dealing with complex, black-box optimization models? What techniques can enhance model explainability? **5**

Marking Scheme
ME3CO31 (T)- Data Science for Mechanical Engineers

- Q.1 i) c. To clean and prepare data for analysis **1**
 ii) b. Scatter plot **1**
 iii) d. More than three **1**
 iv) a. To make inferences about population parameters **1**
 v) d. High-level, aesthetically pleasing statistical data visualization **1**
 vi) a. -1 to 1 **1**
 vii) c. Recall **1**
 viii) b. K-Means **1**
 ix) d. Artificial neural networks **1**
 x) b. Data-driven techniques can handle complex, high-dimensional data. **1**
- Q.2 i. What is data science? **1mark**
 Give its application in the field of Mechanical Engineering. **1mark**
 ii. What is Exploratory Data Analysis? **2marks**
 Explain EDA brief with an example. **1mark**
 iii. How do missing values impact data analysis, what are some strategies for handling missing data during the data cleaning process? **3marks**
2marks
 OR iv. Describe the types of data that are best represented using different types of charts or graphs. **3marks**
 Provide examples of situations where each type is appropriate. **2marks**
- Q.3 i. What are Descriptive Statistics? **2marks**
 ii. Explain the principles of Analysis of Variance (ANOVA) **2marks**
 significance in experimental research **2marks**,
 detailed steps for conducting a one-way ANOVA. **4marks**
 OR iii. Discuss the factors that affect the width of a confidence interval. **4 marks**
 How does variability in the data and confidence level influence the width? **4 marks**
- Q.4 i. Give difference between Matplotlib and Seaborn Libraries of Python: give 3 differences then each difference carries **1mark**

- ii. Explain the concept of correlation analysis. **4marks**
 How does it help in understanding the relationship between variables? **3marks**
 OR iii. Describe the concept of dimensionality reduction and its role in visualizing high-dimensional data. **2marks**
 Explain PCA and t-SNE **3marks**
 examples of how they can be used to visualize multidimensional data. **1mark**
- Q.5 i. What is machine learning? **2marks**
 Write difference between Artificial Intelligence, Machine Learning and Deep Learning. **2marks**
 ii. Discuss the logistic regression model. **2marks**
 Explain how it differs from linear regression and its use in binary and multiclass classification. **2marks**
 Provide an example of a classification problem where logistic regression could be applied. **2marks**
 OR iii. Explore the real-world applications of unsupervised learning, combining both clustering and dimensionality reduction techniques. **4marks**
 Provide examples of scenarios where these techniques can be used together to gain insights from data. **2marks**
- Q.6 i. What are the key data sources and sensors used in condition monitoring for predictive maintenance? **3marks**
 Also explain common challenges in collecting and managing data for predictive maintenance and condition monitoring? **2marks**
 ii. Explain the role of machine learning and data analytics in the context of fault diagnosis and failure prediction. **3marks**
 Provide examples of industries where these techniques have been successfully applied. **2marks**
 iii. Explain the importance of model interpretability and transparency in data-driven design optimization.
 How can stakeholders make informed decisions when dealing with complex, black-box optimization models, and what techniques can enhance model explainability? **2.5marks**

[2]

[3]