



MUTHAYAMMAL ENGINEERING COLLEGE

(An Autonomous Institution)

(Approved by AICTE, New Delhi, Accredited by NAAC & Affiliated to Anna University)
Rasipuram - 637 408, Namakkal Dist., Tamil Nadu.

Department of Artificial Intelligence And Data Science Question Bank - Academic Year (2021-22)

Course Code & Course Name : 19ADC05/Introduction to data science
Name of the Faculty : Dr.P.Srinivasan
Year/Sem/Sec : 11/111

UNIT-1 IMPLEMENT DATA SCIENCE AND ITS APPLICATIONS

2MARKS

1. Define Data science.
2. Summarize the supervised and unsupervised learning.
3. Define Data cleaning.
4. Categorize Supervised and Unsupervised learning.
5. Relate Interpersonal Skill.
6. What do data Scientist do?
7. List the applications of Data Science.
8. Fundamentals steps to complete a Data Analytics Project.
9. Define Data Analyst.
10. Define Data Scientist.
11. What are the essential skills for Data Science.
12. Define Database Administrator
13. Define the Data and Analysis Manager.
14. Define Closing Remarks.
15. What is the role of healthcare in data science in various field.

16Marks

1. Discover Essential of data Science Skills.
2. Mention the Dta Scientist roles and Responsibilities.
3. Stages in a Data Science Project.

4. Identify evaluation of data science in various fields.
5. Mention evaluation of Data Science.
6. Discuss the Data Science roles.

UNIT 2- APPLY RESULTS ON DATA COLLECTION AND DATA PRE- PROCESSING

2MARKS

1. Apply stages in data science project.
2. Differentiate the data warehousing and data Integration.
3. Order steps involved Data Transformation.
4. Why data discretization.
5. Classify data pre processing
6. Organise feature Selection Technique.
7. Define Data Collection.
8. What is data pre processing in ML?
9. What is One –Hot Encoding.
10. What are types of Data reduction?
11. What is Dimensionality Reduction?
12. What is the importance of dimensionality reduction?
13. What are Dimensionality reduction Techniques?
14. List the some advantages and disadvantages in machine learning?

16MARKS

1. Classify the Components of Data Integration.
2. Utilize the Dimensionality Reduction- Techniques, Methods, Components.
3. Analyse the necessary data Pre- Processing steps required for data scientist.
4. Interpret Data Pre-Processing in machine learning.
5. Classify the data collection Strategies.
6. Write advantages and disadvantages in machine learning
7. Detail about the Data Discretization.
8. Discuss about the Data Transformation.

UNIT-3 IMPLEMENT THE GRAPH IN STATISTICS

2MARKS

- 1. Define mean with example.**
- 2. Compare skewness and Kurtosis.**
- 3. Build sample ANOVA**
- 4. Build the terminologies used in ANOVA Table**
- 5. Assess Pivot Table.**
- 6. Define range and variance .**
- 7. Define variability.**
- 8. What is Graphical Representation?**
- 9. What is Boxplots?**
- 10. What does the ANOVA test mean?**
- 11. What is meant by Degrees of Freedom?**
- 12. What are types of ANOVA test?**
- 13. Compute the standard deviation of the sample data : 3,4,7 with the sample mean of 5.**
- 14. Find the range of the given set .
12,29,32,34,38,49,57.**

16MARKS

- 1. Classify the mean, median, mode , variance , and standard deviation with example.**
- 2. Organise ANOVA hypothesis towards one way and two way .**
- 3. Demonstrate descriptive statistics.**
- 4. Classify the Skewness and Kurtosis**
- 5. Discuss the real world applications of ANOVA test.**
- 6. Explain the terminologies in ANOVA test.**
- 7. Build a Highlights Table or Heat Map using Tableau.**
- 8. Calculate the Skewness coefficient of the sample.**

UNIT4- ANALYSE MODEL DEVELOPMENT AND EVALUATION

2MARKS

- 1. Construct the Residual Analysis**

2. **Lable confusion matrix.**
3. **Show visualization data.**
4. **Recall terms used in confusion matrix.**
5. **What is Regression.**
6. **What is error function.**
7. **What is root mean squared error?**
8. **Differntiate the forward and backward selection.**
9. **What is representation learning?**
10. **What are imporatant terms of confusion matrix.**
11. **What is mean absolute error?**

16MARKS

1. **Asses the visualization of regression towards Data Science.**
2. **Compare model evaluation towards data science and machine learning.**
3. **Build a confusion matrix towards the data science.**
4. **Categorize the various error function**
5. **Conclude the benefits of data visualization**
6. **Discuss the correlation Heatmap for advertisting data.**

UNIT5-ANALYSE MODEL EVALUATION METRICS AND VALIDATION

2MARKS

1. **Build precdssion recall curve.**
2. **Compare over fitting and under fitting**
3. **Define Regularization.**
4. **Define Evaluation metrics**
5. **What is meant by Accuracy in data science.**
6. **What is meant by thresholding.**
7. **Compare overfitting and underfitting.**
8. **What is Ridge Regression?**

16 MARKS

- 1. Survey the Generalization , Regularization, over fitting, Bias and Variance in machine learning.**
- 2. Develop confusion matrix**
- 3. Discuss the evaluation metrics**
- 4. Explain the underfitting and overfitting in machine learning.**
- 5. How to prevent the overfitting and underfitting**
- 6. Discuss testing multiple parameters by using Grid search.**