

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur - 603 203

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

QUESTION BANK



VII SEMESTER

1922701 – STATISTICAL APPROACHES FOR DATA SCIENCE

Regulation – 2019

Academic Year 2025 – 2026 Odd Semester

Prepared by

R. Danu, Assistant Professor (O.G)

UNIT 1 INTRODUCTION TO DATA SCIENCE

Introduction to Data Science - Components of Data Science- Challenges in Data Science technology - History - Applications of Data Science - Data Science Process: Discovery – Preparation – Model Planning - Model Building - Operationalize - Communicate Results - Difference between Data Science with Business Intelligence.

PART – A

| Q.No | Question | Competence | Level |
|------|--|------------|-------|
| 1 | What is Data Science? | Remember | BTL 1 |
| 2 | Difference between data science and data analytics. | Understand | BTL 2 |
| 3 | What are the conditions for Overfitting and Underfitting? | Remember | BTL 1 |
| 4 | What is Ensemble Learning? Give an important example of Ensemble Learning? | Remember | BTL 1 |
| 5 | Summarize the Data Science Components. | Evaluate | BTL 5 |
| 6 | Illustrate the Data Analytics Lifecycle. | Apply | BTL 3 |
| 7 | Generalize the goals of data science. | Create | BTL 6 |
| 8 | Can you List out few applications of Data Science. | Create | BTL 6 |
| 9 | List out few challenges of Data Science Technology. | Remember | BTL 1 |
| 10 | Give any two differences between Data Science and Business Intelligence. | Understand | BTL 2 |
| 11 | Give reason Why we need the machine learning in Data Science? | Understand | BTL 2 |
| 12 | Define Data Handling. | Create | BTL 6 |
| 13 | Justify the importance of Data Engineering. | Evaluate | BTL 5 |
| 14 | What is data visualization? | Remember | BTL 1 |
| 15 | Generalize some general techniques for handling volumes of data. | Create | BTL 6 |
| 16 | How would you handle imbalanced data? | Understand | BTL 2 |
| 17 | Can you define the importance of Data Wrangling. | Create | BTL 6 |
| 18 | Summarize any three libraries commonly used in python for data science projects. | Evaluate | BTL 5 |
| 19 | Analyze the tools used for Data Science Process. | Analyze | BTL 4 |
| 20 | What is Data Cleaning? | Remember | BTL 1 |
| 21 | List out the components of data science process. | Remember | BTL 1 |
| 22 | List out the features of Business Intelligence. | Remember | BTL 1 |
| 23 | Define Business Intelligence. | Create | BTL 6 |
| 24 | Summarize the features of Data Science. | Evaluate | BTL 5 |

PART-B

| Q.No. | Question | Competence | Level |
|-------|--|------------|-------|
| 1 | Compare data analytics and data science. (13) | Analyze | BTL 4 |
| 2 | What are some core steps to take for data pre-processing? (13) | Remember | BTL 1 |
| 3 | What do you understand by the term Data Science? Explain the importance of Data Science? List out advantages and disadvantages. (13) | Remember | BTL 1 |

| | | | | |
|----|---|------|------------|-------|
| 4 | Analyze the various lifecycle phases of data analytics. | (13) | Analyze | BTL 4 |
| 5 | Summarize about common tools for the Model Planning Phase. | (13) | Evaluate | BTL 5 |
| 6 | Discuss the steps involved in Data Discovery. | (13) | Understand | BTL 2 |
| 7 | Illustrate the evolution of Data Science. | (13) | Apply | BTL 3 |
| 8 | Briefly explain common data science problem faced by Data Scientists. | (13) | Evaluate | BTL 5 |
| 9 | Point out any five business applications for data science. | (13) | Apply | BTL 3 |
| 10 | List out the categories that may be used to classify the problems that can be solved with the assistance of data science. | (13) | Remember | BTL 1 |
| 11 | Summarize the major tasks involved in data preparation. | (13) | Evaluate | BTL 5 |
| 12 | Formulate Data Science Challenges in Healthcare applications. | (13) | Apply | BTL 3 |
| 13 | Explain in detail about Data Science Components. | (13) | Analyze | BTL 4 |
| 14 | Illustrate few Use cases for Data science and Artificial Intelligence. | (13) | Apply | BTL 3 |
| 15 | Briefly explain the issues of Data Science Process. | (13) | Evaluate | BTL 5 |
| 16 | Describe the Usage of Data Science Process. | (13) | Analyze | BTL 4 |
| 17 | Point out the features of Data Science and Business Intelligence. | (13) | Apply | BTL 3 |

PART - C

| | | | | |
|----|---|------|----------|-------|
| 1. | State and explain the difference between data analytics and data science? | (15) | Create | BTL 6 |
| 2. | Summarize how data science differ from traditional application programming? | (15) | Evaluate | BTL 5 |
| 3. | Evaluate the Steps how to approach and address a solution to Data Science Problems. | (15) | Evaluate | BTL 5 |
| 4. | Briefly explain in detail about Data Science Process Life Cycle. | (15) | Evaluate | BTL 5 |
| 5. | Justify how Data Science differ with Business Intelligence. | (15) | Evaluate | BTL 5 |

UNIT II STATISTICS

Statistics Fundamentals - Descriptive Statistics - Exploratory Data Analysis- Probability - Skewness -Percentiles and Outliers - Random Variables - Cumulative Distributive Function - Bayes Theorem - Two area of Statistics in Data Science.

PART – A

| Q.No | Question | Competence | Level |
|------|---|------------|-------|
| 1 | Define Statistics. | Remember | BTL 1 |
| 2 | List out the five basic terminologies used in Statistics. | Remember | BTL 1 |
| 3 | What are the types of Statistics? | Remember | BTL 1 |
| 4 | Compare Descriptive and Inferential Statistics. | Analyze | BTL 4 |
| 5 | Define Mean, Median and Mode in Statistics. | Remember | BTL 1 |
| 6 | List out the Variability Measurements in Statistics. | Remember | BTL 1 |

| | | | |
|----|---|------------|-------|
| 7 | Difference Between Categorical and Numerical Variables. | Understand | BTL 2 |
| 8 | Give an Example where the median is a better measure than the mean. | Understand | BTL 2 |
| 9 | How do you calculate the needed sample size? | Analyze | BTL 4 |
| 10 | Generalize Skewness and Kurtosis. | Create | BTL 6 |
| 11 | Difference between a population and a sample. | Understand | BTL 2 |
| 12 | Compare Quantitative and Qualitative data. | Analyze | BTL 4 |
| 13 | Discuss exploratory data analysis. | Create | BTL 6 |
| 14 | What is an Outlier? How can outliers be determined in a dataset? | Remember | BTL 1 |
| 15 | How is missing data handled in statistics? | Evaluate | BTL 5 |
| 16 | The random variables X and Y have variances 0.2 and 0.5 respectively. Let $Z=5X-2Y$. The variance of Z is? | Evaluate | BTL 5 |
| 17 | Find median and mode of the messages received on 9 consecutive days 15,11,9,5,18,4,15,13,17. | Evaluate | BTL 5 |
| 18 | Compare left-skewed and right-skewed distribution. | Analyze | BTL 4 |
| 19 | Define Central Limit Theorem and its significance in distribution functions? | Remember | BTL 1 |
| 20 | Identify Probability Density Functions (PDF). | Create | BTL 6 |
| 21 | Identify Cumulative Distribution Functions (CDF). | Create | BTL 6 |
| 22 | Compare the relationship between PDF and CDF for a Continuous Random Variable. | Analyze | BTL 4 |
| 23 | Explain Bayes Theorem. | Analyze | BTL 4 |
| 24 | List out some unique applications of Bayesian Statistics and Bayes Theorem. | Remember | BTL 1 |

PART-B

| Q.No. | Question | Competence | Level | | | | | | | | | | | | |
|----------------|---|----------------|-------|-----|------|-----|------|----------------|---|---|---|---|---|----------|-------|
| 1 | Consider a small unit of a factory where there are 5 employees: a supervisor and four labourers. The workers earn a salary of Rs. 5,000 per month each while the supervisor gets Rs. 15,000 per month. Calculate the mean, median and mode of the salaries. (13) | Evaluate | BTL 5 | | | | | | | | | | | | |
| 2 | A survey conducted on 20 houses in an area by a group of people resulted in the subsequent frequency table for the number of family members in a house: <table border="1" style="margin-left: 20px; margin-top: 10px;"> <tr> <td>Size of Family</td> <td>1-3</td> <td>3-5</td> <td>5-7</td> <td>7-9</td> <td>9-11</td> </tr> <tr> <td>No.of Families</td> <td>7</td> <td>8</td> <td>2</td> <td>2</td> <td>1</td> </tr> </table> Find the mode of this data. (13) | Size of Family | 1-3 | 3-5 | 5-7 | 7-9 | 9-11 | No.of Families | 7 | 8 | 2 | 2 | 1 | Evaluate | BTL 5 |
| Size of Family | 1-3 | 3-5 | 5-7 | 7-9 | 9-11 | | | | | | | | | | |
| No.of Families | 7 | 8 | 2 | 2 | 1 | | | | | | | | | | |
| 3 | Illustrate the types of descriptive statistics. (13) | Apply | BTL 3 | | | | | | | | | | | | |
| 4 | Discuss Skewness in descriptive statistics. (13) | Create | BTL 6 | | | | | | | | | | | | |
| 5 | Describe Kurtosis in descriptive statistics. (13) | Analyze | BTL 4 | | | | | | | | | | | | |
| 6 | Explain is correlation in descriptive statistics. (13) | Analyze | BTL 4 | | | | | | | | | | | | |
| 7 | Summarize the methods to detect outliers. (13) | Evaluate | BTL 5 | | | | | | | | | | | | |

| 8 | Write a note on dealing with Outliers. (13) | Remember | BTL 1 | | | | | | | | | | | | | | | | | | | | | | |
|---|---|------------|-------|-------------------------|---|----|----|----|----|------|-----|------|------|-----|----|----|----|----|----|----|----|----|----|----|----|
| 9 | Describe Measures of Skewness and Kurtosis. (13) | Create | BTL 6 | | | | | | | | | | | | | | | | | | | | | | |
| 10 | What is a Percentile in Statistics and How to calculate 90 Percentile if you have mark of 20 students. (13) | Create | BTL 6 | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <thead> <tr> <th colspan="2">Marks Scored Out of 100</th> </tr> </thead> <tbody> <tr> <td>89</td> <td>97</td> </tr> <tr> <td>78</td> <td>45</td> </tr> <tr> <td>94</td> <td>50</td> </tr> <tr> <td>66</td> <td>69</td> </tr> <tr> <td>50</td> <td>73</td> </tr> <tr> <td>43</td> <td>94</td> </tr> <tr> <td>92</td> <td>58</td> </tr> <tr> <td>75</td> <td>87</td> </tr> <tr> <td>81</td> <td>77</td> </tr> <tr> <td>53</td> <td>45</td> </tr> </tbody> </table> | | | | Marks Scored Out of 100 | | 89 | 97 | 78 | 45 | 94 | 50 | 66 | 69 | 50 | 73 | 43 | 94 | 92 | 58 | 75 | 87 | 81 | 77 | 53 | 45 |
| Marks Scored Out of 100 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 89 | 97 | | | | | | | | | | | | | | | | | | | | | | | | |
| 78 | 45 | | | | | | | | | | | | | | | | | | | | | | | | |
| 94 | 50 | | | | | | | | | | | | | | | | | | | | | | | | |
| 66 | 69 | | | | | | | | | | | | | | | | | | | | | | | | |
| 50 | 73 | | | | | | | | | | | | | | | | | | | | | | | | |
| 43 | 94 | | | | | | | | | | | | | | | | | | | | | | | | |
| 92 | 58 | | | | | | | | | | | | | | | | | | | | | | | | |
| 75 | 87 | | | | | | | | | | | | | | | | | | | | | | | | |
| 81 | 77 | | | | | | | | | | | | | | | | | | | | | | | | |
| 53 | 45 | | | | | | | | | | | | | | | | | | | | | | | | |
| 11 | Discuss on the following: (6) a. Quartiles (7) b. Percentiles (7) | Create | BTL 6 | | | | | | | | | | | | | | | | | | | | | | |
| 12 | Imagine that Jeremy took part in an examination. The test is having a mean score of 160, and it has a standard deviation of 15. If Jeremy's z-score is 1.20, what would be his score on the test? (13) | Create | BTL 6 | | | | | | | | | | | | | | | | | | | | | | |
| 13 | Generalize the concept of Skewness and Kurtosis in distribution functions. (13) | Create | BTL 6 | | | | | | | | | | | | | | | | | | | | | | |
| 14 | i. Write a short note on Percentiles of a Distribution and list of the Special Cases. (6) ii. List out the Properties of CDF. (7) | Analyze | BTL 4 | | | | | | | | | | | | | | | | | | | | | | |
| 15 | i. Explain the role of Bayes Theorem in Machine Learning and provide an example of a practical application of Bayes Theorem. (6) ii. Explain the difference between Bayes Theorem and Conditional Probability. (7) | Understand | BTL 2 | | | | | | | | | | | | | | | | | | | | | | |
| 16 | i. Explain Navie Bayes Classifier. (6) ii. Compare Maximum Likelihood Estimation (MLE) and Bayesian Statistics. (7) | Understand | BTL 2 | | | | | | | | | | | | | | | | | | | | | | |
| 17 | Illustrate Fundamental statistics concepts for data science. (13) | Apply | BTL 3 | | | | | | | | | | | | | | | | | | | | | | |
| PART - C | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | Demonstrate Exploratory Data Analysis for Bank Marketing data set. (15) | Create | BTL 6 | | | | | | | | | | | | | | | | | | | | | | |
| 2 | (i) What is the cumulative distribution function formula? (6) Find: a. $P(x=3)$ b. $P(x>2)$ (ii) What is the CDF of normal distribution in R? (7) Find: a. $P(x\leq 4.5)$ b. $P(x>4.5)$ | Evaluate | BTL 5 | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <thead> <tr> <th>x</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> </tr> </thead> <tbody> <tr> <td>F(x)</td> <td>0.2</td> <td>0.32</td> <td>0.67</td> <td>0.9</td> <td>1</td> </tr> </tbody> </table> | | | | x | 1 | 2 | 3 | 4 | 5 | F(x) | 0.2 | 0.32 | 0.67 | 0.9 | 1 | | | | | | | | | | |
| x | 1 | 2 | 3 | 4 | 5 | | | | | | | | | | | | | | | | | | | | |
| F(x) | 0.2 | 0.32 | 0.67 | 0.9 | 1 | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | |
|----|---|----------|-------|----|----|----|----|----|----|----|----|----|----------|-------|
| 3 | (i) Explain the Measures of Skewness and Kurtosis. (5) (ii) Using the interquartile range to find outliers. Dataset has 11 values check whether they are outliers. <table border="1" style="margin: 10px auto;"><tr><td>25</td><td>37</td><td>24</td><td>28</td><td>35</td><td>22</td><td>31</td><td>53</td><td>41</td><td>64</td><td>29</td></tr></table> (10) | 25 | 37 | 24 | 28 | 35 | 22 | 31 | 53 | 41 | 64 | 29 | Evaluate | BTL 5 |
| 25 | 37 | 24 | 28 | 35 | 22 | 31 | 53 | 41 | 64 | 29 | | | | |
| 4 | Briefly explain in detail about Cumulative Distributive Function with an example. (13) | Evaluate | BTL 5 | | | | | | | | | | | |
| 5 | (i) What is Bayes Theorem? Provide with Formula. (6) (ii) Explain the difference between Bayesian and Frequentist Statistics. (7) | Create | BTL 6 | | | | | | | | | | | |

UNIT III - VISUALIZATION

Context of data visualization – Definition, Methodology, Visualization design objectives- Mapping - Time series - Connections and correlations - Scatterplot maps -Data representation, Data Presentation, Seven stages of data visualization, widgets, data visualization tools-Port scan visualization - Vulnerability assessment and exploitation - Firewall log visualization

PART – A

| Q.No. | Question | Competence | Level |
|-------|---|------------|-------|
| 1 | Define Data Visualization. | Remember | BTL 1 |
| 2 | Why is Data Visualization important? | Remember | BTL 1 |
| 3 | List out some popular techniques used in data visualization. | Remember | BTL 1 |
| 4 | Illustrate Common Data Visualization use cases. | Apply | BTL 3 |
| 5 | List out top Data visualization tools. | Remember | BTL 1 |
| 6 | Give few examples of interactive map data visualizations. | Understand | BTL 2 |
| 7 | Define Time Series Data. | Create | BTL 6 |
| 8 | Give some common Data Preparation Operations to use for Time Series Data? | Understand | BTL 2 |
| 9 | What is Scatter plot? | Remember | BTL 1 |
| 10 | Define Correlations. | Remember | BTL 1 |
| 11 | Write any two advantages and disadvantages in data visualization. | Understand | BTL 2 |
| 12 | Discuss Data Representation. | Understand | BTL 2 |
| 13 | Can you List out few graphical representations of data. | Evaluate | BTL 5 |
| 14 | Illustrate data visualization process flow. | Apply | BTL 3 |
| 15 | Can you Define Widgets. | Analyze | BTL 4 |
| 16 | Point out five best widget types. | Analyze | BTL 4 |
| 17 | Illustrate tools for visualizing data. | Apply | BTL 3 |
| 18 | Analyze firewall analyser? | Apply | BTL 3 |
| 19 | Discuss zero day vulnerability. | Analyze | BTL 4 |
| 20 | What is a vulnerability assessment? | Create | BTL 6 |

| 21 | Discuss about Exploit and its types. | Analyze | BTL 4 |
|-----------------|--|------------|-------|
| 22 | Give out few examples of vulnerabilities and exploits. | Create | BTL 6 |
| 23 | Analyze the two types of Exploit? | Apply | BTL 3 |
| 24 | Where firewall logs are visible in GUI? | Understand | BTL 2 |
| PART-B | | | |
| Q.No. | Question | Competence | Level |
| 1 | What is Data Visualization? What makes Data Visualization Effective? (13) | Remember | BTL 1 |
| 2 | Briefly explain visualization design objectives. (13) | Create | BTL 6 |
| 3 | Explain advantages and disadvantages of Data Visualization. (13) | Analyze | BTL 4 |
| 4 | Describe why data visualization is important? (13) | Analyze | BTL 4 |
| 5 | Summarize the categories of Data Visualization. (13) | Evaluate | BTL 5 |
| 6 | Give reason how the IQR (Interquartile Range) used in Time Series Forecasting? (13) | Understand | BTL 2 |
| 7 | Illustrate some real-world applications of Time-series Forecasting? (6) Why does a Time Series have to be Stationary? (7) | Remember | BTL 1 |
| 8 | Write a note on importing the libraries for timeseries data. (13) | Create | BTL 6 |
| 9 | Illustrate Scatter plot with an example code. (13) | Apply | BTL 3 |
| 10 | Explain Data representation with an example. (13) | Create | BTL 6 |
| 11 | Explain how to improve data visualization strategy. (13) | Evaluate | BTL 5 |
| 12 | Summarize any three widget types for data visualization. (13) | Evaluate | BTL 5 |
| 13 | Illustrate any 5 data visualization tools. (13) | Apply | BTL 3 |
| 14 | Discuss about the types of Vulnerability. (13) | Understand | BTL 2 |
| 15 | Compare Vulnerability and Exploit. (13) | Understand | BTL 2 |
| 16 | Point out the methods of analyzing firewall logs. (13) | Create | BTL 6 |
| 17 | Illustrate some firewall events. (13) | Apply | BTL 3 |
| PART – C | | | |
| 1 | Explain in detail about Data Visualization and its importance. Why to use Data Visualization? (15) | Evaluate | BTL 5 |
| 2 | Compare some Forecasting Techniques for Stationary and Non-stationary Time-series with examples. (15) | Create | BTL 6 |
| 3 | Illustrate the types of Graphical Data Representation with example. (15) | Create | BTL 6 |
| 4 | Summarize the seven stages of data visualization. (15) | Evaluate | BTL 5 |
| 5 | Discuss on the following: a. Vulnerability assessment and Exploitation. (5) b. Firewall log visualization. (5) c. Advantages and disadvantages in vulnerability assessment. (5) | Create | BTL 6 |

UNIT IV INTRODUCTION TO MACHINE LEARNING

Introduction - Inductive Classification - Decision Tree Learning - Ensemble Learning - Experimental-Evaluation of Learning Algorithms - Computational Learning Theory - Rule Learning : Propositional and First Order - Artificial Neural Networks - Support Vector Machines

PART – A

| Q.No. | Question | Competence | Level |
|-------|--|------------|-------|
| 1 | What is Machine Learning? | Remember | BTL 1 |
| 2 | List out the features of Machine Learning. | Remember | BTL 1 |
| 3 | Why there is the need for Machine Learning? | Remember | BTL 1 |
| 4 | List out the classification of Machine Learning. | Remember | BTL 1 |
| 5 | Point out few applications of Machine Learning. | Analyze | BTL 4 |
| 6 | Discuss Inductive Classification. | Analyze | BTL 4 |
| 7 | What is Hypothesis Selection. | Remember | BTL 1 |
| 8 | Define Inductive bias and its two forms. | Create | BTL 6 |
| 9 | Define Decision Tree. | Remember | BTL 1 |
| 10 | List out the decision tree terminologies. | Create | BTL 6 |
| 11 | Define Ensemble Learning with an example. | Remember | BTL 1 |
| 12 | List out few common performance evaluation Metrics. | Create | BTL 6 |
| 13 | Why it is important to evaluate the performance of learning systems? | Remember | BTL 1 |
| 14 | Define k-Fold Cross Validation. | Create | BTL 6 |
| 15 | Generalize Computational Learning Theory. | Evaluate | BTL 5 |
| 16 | Summarize the main goals of computational learning Theory. | Evaluate | BTL 5 |
| 17 | Justify the use of first order logic in AI. | Analyze | BTL 4 |
| 18 | Compare Propositional Logic and First-order Logic. | Analyze | BTL 4 |
| 19 | Point out the Features of Propositional Logic. | Apply | BTL 3 |
| 20 | Give reason the importance of Artificial Neural Network? | Understand | BTL 2 |
| 21 | List out few applications of Artificial Neural Networks. | Remember | BTL 1 |
| 22 | Can a support vector be used in a nonlinear data set? | Evaluate | BTL 5 |
| 23 | Justify the purpose of Support Vector Machine. | Evaluate | BTL 5 |
| 24 | Illustrate the steps involved in SVM. | Apply | BTL 3 |

PART-B

| Q.No. | Question | Competence | Level |
|-------|---|------------|-------|
| 1 | What is Machine Learning? Explain the classification of Machine Learning. (13) | Remember | BTL 1 |
| 2 | Explain the necessary steps for implementing Inductive learning algorithm. (13) | Understand | BTL 2 |
| 3 | How to choose the best attribute at each node in a decision tree? (13) | Understand | BTL 2 |
| 4 | Explain the basic steps of the CART algorithm. (13) | Evaluate | BTL 5 |

| | | | |
|----|---|----------|-------|
| 5 | Discuss what are appropriate problems for Decision tree learning? (13) | Analyze | BTL 4 |
| 6 | List out the advantages and disadvantages of decision tree. (13) | Remember | BTL 1 |
| 7 | Summarize in detail about Attribute Selection Measures. (13) | Create | BTL 6 |
| 8 | Explain algorithms based on Bagging and Boosting. (13) | Analyze | BTL 4 |
| 9 | Discuss about simple ensemble techniques (13) | Analyze | BTL 4 |
| 10 | Write a short notes on the following: a. Bagging (7) b. Boosting (6) | Create | BTL 6 |
| 11 | Identify some of the main challenges faced by computational learning theory? (13) | Analyze | BTL 4 |
| 12 | Point out some of the main methods used in Computational Learning Theory? (13) | Apply | BTL 3 |
| 13 | Briefly explain in detail about First-order Logic. (13) | Evaluate | BTL 5 |
| 14 | List out advantages and disadvantages of Artificial Neural Networks. (13) | Remember | BTL 1 |
| 15 | Describe the architecture of an Artificial Neural Network. (13) | Create | BTL 6 |
| 16 | Explain in detail how SVM works? (13) | Evaluate | BTL 5 |
| 17 | Summarize in detail about SVM and its types. (13) | Evaluate | BTL 5 |

PART – C

| | | | |
|---|---|----------|-------|
| 1 | Summarize the Evolution of Machine Learning. (15) | Evaluate | BTL 5 |
| 2 | Illustrate how does the decision tree algorithm works with an example. (15) | Create | BTL 6 |
| 3 | Describe about Advanced ensemble techniques. (15) | Create | BTL 6 |
| 4 | Formulate the Classifier Evaluation Procedure. (15) | Create | BTL 6 |
| 5 | Illustrate how do Artificial Neural Networks work and its types. (15) | Create | BTL 6 |

UNIT V PREDICTIVE MODELING

Regression, Classification, Data Preprocessing, Model Evaluation and Ensembles – Data Mining Dimensionality Reduction, Clustering, Association Rules, Anomaly Detection – Network Analysis and Web Applications.

PART – A

| Q.No. | Question | Competence | Level |
|-------|---|------------|-------|
| 1 | Define Linear Regression Algorithm. | Remember | BTL 1 |
| 2 | What are the basic assumptions of the Linear Regression Algorithm? | Remember | BTL 1 |
| 3 | How do you interpret a linear regression model? | Evaluate | BTL 5 |
| 4 | List down some of the metrics used to evaluate a regression model? | Remember | BTL 1 |
| 5 | What is MAE and MAPE | Remember | BTL 1 |
| 6 | Can you choose a classifier based on the size of the training set? | Evaluate | BTL 5 |
| 7 | How does ROC curve and AUC value help to measure how good a model is? | Evaluate | BTL 5 |
| 8 | What is Data Pre-processing? What pre-processing steps do you need? | Analyze | BTL 4 |
| 9 | Compare the difference between Data Processing and Data Mining? | Understand | BTL 2 |

| | | | |
|----|---|------------|-------|
| 10 | Summarize some common problems that occur during data processing? How can they be fixed? | Evaluate | BTL 5 |
| 11 | How does discretization help us deal with outliers? | Evaluate | BTL 5 |
| 12 | What is dimensionality reduction? | Analyze | BTL 4 |
| 13 | What is Clustering? | Remember | BTL 1 |
| 14 | List out some common applications of clustering. | Create | BTL 6 |
| 15 | Difference between hard and soft clustering. | Understand | BTL 2 |
| 16 | Point out the challenges associated with clustering? | Apply | BTL 3 |
| 17 | List out the Clustering Methods. | Create | BTL 6 |
| 18 | Analyze the applications of clustering in different fields. | Apply | BTL 3 |
| 19 | Justify the use of Anomaly Detection and its categories. | Apply | BTL 3 |
| 20 | Justify the need of anomaly detection. | Apply | BTL 3 |
| 21 | Compare different types of association learning algorithms and why it is important in analyzing data? | Analyze | BTL 4 |
| 22 | Analyze recommender system and how do you build a recommender system? | Apply | BTL 3 |
| 23 | List out the components of NLP and the order of steps in NLP. | | |
| 24 | Define Web Services. | Remember | BTL 1 |

| PART-B | | | |
|---------------|--|-------------------|--------------|
| Q.No. | Question | Competence | Level |
| 1 | Generalize the Gradient Descent algorithm with respect to linear regression. (13) | Create | BTL 6 |
| 2 | In linear regression, what is the value of the sum of the residuals for a given dataset? Explain with proper justification. (13) | Apply | BTL 3 |
| 3 | i. What are the disadvantages of the linear regression algorithm? (3) ii. Difference between classification and regression in Machine Learning. (5) iii. Explain the types of regression techniques. (5) | Remember | BTL 1 |
| 4 | Illustrate the following: a. Difference between Bagging and Boosting Algorithms. (6) b. Difference between Generative classifiers and Discriminative classifiers with example. (7) | Analyze | BTL 4 |
| 5 | Discuss how to handle missing values? (13) | Analyze | BTL 4 |
| 6 | i. What is the best way to deal with skewed datasets? (6) ii. What is the difference between supervised and unsupervised learning? (7) | Remember | BTL 1 |
| 7 | Examine why Dimensionality Reduction important in Machine Learning and Predictive Modelling? (13) | Evaluate | BTL 5 |
| 8 | List out the advantages and dis-advantages of Dimensionality reduction. (13) | Remember | BTL 1 |
| 9 | Discuss about Principal Component Analysis and its importance and limitations. (13) | Evaluate | BTL 5 |

| | | | |
|-----------------|---|------------|-------|
| 10 | Illustrate in detail about k-means clustering and the steps involved. (6) What are some stopping criteria for k-Means Clustering? (7) | Analyze | BTL 4 |
| 11 | i. Compare Hierarchical and k-Means Clustering. (3) ii. Explain some cases where k-Means clustering fails to give good results. (5) iii. How is Entropy used as a clustering validation Measure? (5) | Understand | BTL 2 |
| 12 | Explain in detail about Anomaly Detection techniques. (13) | Evaluate | BTL 5 |
| 13 | Describe in detail about the three types of algorithms used in Association rule learning and its rule evaluation metrics. (13) | Create | BTL 6 |
| 14 | List out the advantages and disadvantages of NLP. (13) | Create | BTL 6 |
| 15 | What is NLP? Illustrate the different approaches to deal NLP tasks. (13) | Analyze | BTL 4 |
| 16 | Write a short note on the following: a. Applications of NLP (6) b. Components of NLP (7) | Create | BTL 6 |
| 17 | Examine why Dimensionality Reduction important in Machine Learning and Predictive Modelling? (13) | Apply | BTL 3 |
| PART - C | | | |
| 1 | Explain the normal form equation of the linear regression. Also explain when should it be preferred to the Gradient Descent method instead of the Normal Equation in Linear Regression Algorithm? (15) | Evaluate | BTL 5 |
| 2 | Explain in detail about the steps involved in Data Pre-processing. (15) | Evaluate | BTL 5 |
| 3 | Evaluate in detail about Dimensionality reduction and its approaches. (15) | Evaluate | BTL 5 |
| 4 | Describe Recommender system in detail. (15) | Create | BTL 6 |
| 5 | Summarize in detail about Natural Language Processing. (15) | Evaluate | BTL 5 |

