|  |  |  |
| --- | --- | --- |
| D:\NAAC (Anoop Tiwari)\IQAC\unnamed.png | | **SAGAR INSTITUTE OF SCIENCE & TECHNOLOGY**  **DEPARTMENT OF CSE-ARTIFICIAL INTELLIGENCE & DATA SCIENCE**  **QUESTION BANK** |
| **BRANCH** | **AIDS** |
| **SESSION** | **2023-24** |
| **SEMESTER** | **IV** |
| **SUBJECT/CODE: DATA SCIENCE/AD-404** | | |

|  |  |  |
| --- | --- | --- |
| **S.No** | **Question** | **Bloom’s Taxonomy** |
| **UNIT 1** | | |
| 1. | Why one should study data science? Discuss the application of data science in multiple fields. | 2(Understand) |
| 2. | Differentiate between structured and unstructured data with the help of suitable examples. | 2(Understand) |
| 3. | What do you mean by structured data? Give advantages, disadvantages of structured data and various tools used to process structured data. | 1(Remember) |
| 4. | Explain the concept of data wrangling. | 1(Remember) |
| 5. | Discuss the following: a) Pychart b) Bar chart | 2(Understand)) |
| 6. | Discuss the following: a) Pareto chart b) Histogram | 2(Understand) |
| 7. | Classify measures of central tendency of quantitative data. | 2(Understand) |
| 8. | Outline the measures of variability of data. | 2(Understand) |
| 9. | Demonstrate the use of conditional probability in the field of data science. | 3(Apply) |
| 10. | Discuss different types of probability distributions. | 2(Understand) |
| 11. | Explain exploratory data analysis. | 1(Remember) |
| 12. | Give a brief data science road map. | 2(Understand) |
| **UNIT 2** | | |
| 1. | What do you mean by structured data? Give advantages, disadvantages of structured data and various tools used to process structured data. | 1(Remember) |
| 2. | Compare descriptive, diagnostic, predictive and prescriptive data analytics in terms of their efficiency and complexity. | 2(Understand) |
| 3. | Why data visualization is considered as an important step in the process of getting insights for a business model? | 1(Remember) |
| 4. | Discuss feature map visualization. | 2(Understand) |
| 5. | Discuss the process adopted by t-SNE to reduce high dimensional data into low dimensional data. | 2(Understand) |
| 6. | Discuss various features of Microsoft excel that makes it a wonderful tool for data analysis. | 2(Understand) |
| 7. | Identify various type of charts available in Microsoft Excel. | 1(Remember) |
| 8. | How one can perform data validation using Microsoft Excel? | 1(Remember) |
| 9. | Discuss the role of scenario manager in Microsoft Excel. | 2(Understand) |
| 10. | Explain the protection mechanism used in Microsoft Excel to protect data. | 1(Remember) |
| **UNIT 3** | | |
| 1. | What is hypothesis testing ? Differentiate null and alternative Hypothesis. | 2(Understand) |
| 2. | In order to ensure safety, donated blood undergoes testing for infectious diseases and contaminants. Testing batches of donated blood instead of individual samples is a cost-effective and time-saving approach, as the majority of donated blood is typically deemed safe. As part of the testing process, a specific examination is conducted to detect the presence of a particular toxin. If the toxin is detected, the entire batch is discarded, analogous to utilizing null and alternative hypotheses to decide whether to retain or dispose of the batch.  The hypotheses being tested could be stated as:  H0: The batch does not contain the toxin. Ha: The batch contains the toxin.  Specify the consequence of Type I error here. | 3(Apply) |
| 3 | What is the process of parameter estimation using interval estimation, and how does it help in determining the range of possible values for a parameter with a certain level of confidence? | 3(Apply) |
| 4.. | What are shrinkage methods in statistics and how do they help in addressing the issue of overfitting in predictive models? | 3(Apply) |
| 5. | Illustrate the concept of Maximum Likelihood. Also derive the expression for Bayes theorem. | 4(Analyze) |
| 6. | Explain Bayesian Statistics with example. | 1(Remember) |
| 7. | Contrast Correlation and Regression Analysis. | 3(Apply) |
| 8. | What are the limitations of logistic regression? Also illustrate how to design its cost function. | 3(Apply) |
| 9. | "What insights can be gained from a comprehensive data analysis of customer behavior, market trends, and competitive intelligence to drive strategic decision-making and enhance business performance?" | 4(Analyze) |
| 10. | How can I use the INDEX and MATCH formula in Excel to find and retrieve specific data from a table? | 3(Apply) |
| 11. | What are dynamic array formulas and how can they be used in spreadsheet applications? | 2(Understand) |
| 12. | How can I effectively handle formula errors in my spreadsheet? | 3(Apply) |
| **UNIT 4** | | |
| 1. | What are the key features and functionalities of pandas that make it useful for data preprocessing in machine learning? | 2(Understand) |
| 2. | How can pandas be used for data cleaning and handling missing values in machine learning datasets? | 3(Apply) |
| 3. | How can regular expressions be applied in machine learning tasks? | 3(Apply) |
| 4. | Is there any specific Python package that provide a comprehensive set of tools for creating various types of charts, including pie charts, bar charts, and scatterplots? If so, write the code for the same. | 2(Understand) |
| 5. | What are the key steps involved in creating a histogram in Python, and how can I adjust parameters like bin size and range? | 2(Understand) |
| 6. | What are the common performance metrics used to evaluate the effectiveness of a machine learning model in regression and classification? | 2(Understand) |
| 7. | How does the concept of cross-validation help in evaluating the performance of a machine learning model? | 2(Understand) |
| 8. | Can you explain the concept of precision-recall curve and its significance in evaluating models with imbalanced datasets? | 3(Apply) |
| 9. | How are receiver operating characteristic (ROC) curves and area under the curve (AUC) used to assess the performance of binary classification models? | 3(Apply) |
| 10. | What is overfitting and under fitting in machine learning, and how does it affect the performance of a model? | 2(Understand) |
| 11. | Differentiate bagging and boosting techniques. Can bagging and boosting be combined together in ensemble learning? If so, how? | 3(Apply) |
| 12. | In what situations is gradient boosting a better choice than random forest? | 3(Apply) |
| **UNIT 5** | | |
| 1. | What is the importance of business intelligence and its role in decision-making? | 2(Understand) |
| 2. | Can you provide an overview of the different types of business intelligence? | 1(Remember) |
| 3. | What are some modern business intelligence tools that organizations use to analyze and visualize data? | 1(Remember) |  |  |
| 4. | How has business intelligence evolved in recent years with the advent of new technologies? | 3(Apply) |  |  |
| 5. | What are some potential ethical issues associated with the use of data science in business intelligence? | 2(Understand) |  |  |
| 6. | How does unfair discrimination arise in the context of business intelligence, and what are its implications? | 3(Apply) |  |  |
| 7. | In what ways does business intelligence reinforce human biases, and how can this impact decision-making? | 3(Apply) |  |  |
| 8. | What are the challenges and consequences of the lack of transparency in business intelligence processes? | 2(Understand) |  |  |
| 9. | How do privacy concerns come into play when it comes to collecting and analyzing data for business intelligence purposes? | 3(Apply) |  |  |
| 10. | What are the key security considerations organizations should keep in mind when working with business intelligence systems? | 2(Understand) |  |  |
| 11. | Discuss the ethical implications of using data science in business intelligence and decision-making. | 1(Remember) |  |  |
| 12. | What is the role of next-generation data scientists in addressing ethical issues related to business intelligence? | 2(Understand) |  |  |