|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Velegapudi Ramakrishna Siddhartha Engineering College::Vijayawada**  **(Autonomous)**  III /IV B Tech Degree Examinations(May/2023)  Sixth Semester  **Department of Computer Science and Engineering**  **20CS6404B:PROGRAMMING FOR DATA SCIENCE** | | | | | | | | | |
| Time:3Hrs | | | **MODEL QUESTION PAPER** | | Max Marks:70 | | | | |
| Part – A is Compulsory  Answer one (01) question from each unit of Part – B  Answers to any single question or its part shall be written at one place only | | | | | | | | | |
| ***Cognitive Levels(K): K1-Remember;K2-Understand; K3-Apply; K4-Analyze; K5-Evaluate; K6-Create*** | | | | | | | | |  |
| **Q. No** | | **Question** | | **Marks** | | **Course Outcome** | **Cog. Level** | | **POI** |
| **Part - A** | | | | **10X1=10M** | | | | |  |
| 1 | a | Define Data Science? | | 1 | CO1 | | | K2 | 1.2.1 |
|  | b | List Associated fields of Data Science that heavily relies on it | | 1 | CO1 | | | K2 | 1.7.1 |
|  | c | Outline Data object and attribute? List various types of attributes with examples. | | 1 | CO2 | | | K2 | 1.7.1 |
|  | d | Justify the following statement  **“Data science borrows computational techniques from the disciplines of statistics, machine learning, experimentation, and database theories”.** | | 1 | CO1 | | | K2 | 1.7.1 |
|  | e | Describe Discrete Wavelet Transform | | 1 | CO2 | | | K2 | 1.2.1 |
|  | f | Write a short note on data cube**.** | | 1 | CO3 | | | K2 | 1.2.1 |
|  | g | Illustrate support and confidence measures | | 1 | CO4 | | | K2 | 2.8.1 |
|  | h | Explain Discretization by binning | | 1 | CO2 | | | K2 | 1.2.2 |
|  | i | List typical requirements of clustering in data mining | | 1 | CO4 | | | K2 | 1.7.1 |
|  | j | Write a python code for summarizing the data using Histogram for the following and plot the result | | 1 | CO1 | | | K2 | 1.2.1 |
| **Part - B** | | | | **4X15 =60M** | | | | | |
| **UNIT - I** | | | | | | | | |  |
| 2 | a | Compare and Contrast various Data Science tasks with examples | | 7 | CO1 | | K2 | | 1.7.1 |
|  | b | How to measure the Dispersion of data using range, Quartile, variance, Standard deviation and Interquartile range? Explain with an example. | | 8 | CO1 | | K2 | | 1.2.1 |
| **(OR)** | | | | | | | | |  |
| 3 | a | Explain about Measuring data similarity and dissimilarity. List and explain proximity measures for the following with an examples  i) Nominal Attributes  ii) Binary Attributes  iii) Ordinal Attributes  iv) Numeric Attributes | | 8 | CO1 | | | K2 | 1.7.1 |
|  | b | Write a python code for measuring central tendency (Mean, Median and Mode) for the following values of salary (in thousands of dollars), shown in increasing order: 30, 36, 47, 50, 52, 52, 56, 60, 63, 70, 70, 110 | | 7 | CO1 | | | K2 | 1.2.1 |
| **UNIT - II** | | | | | | | | |  |
| 4 | a | Elaborate the approaches for cleaning the data | | 7 | CO2 | | K2 | | 2.5.2 |
|  | b | Explainthe following  i) Principal Component Analysis  ii) Data Transformation and Data Discretization | | 8 | CO2 | | K2 | | 2.8.1 |
| **(OR)** | | | | | | | | |  |
| 5 | a | Describe Redundancy and Correlation Analysis in data integration | | 8 | CO2 | | K2 | | 2.6.4 |
|  | b | Suppose that a group of 1500 people was surveyed. The gender of each person was noted. Each person was polled as to whether his or her preferred type of reading material was fiction or nonfiction. Thus, we have two attributes, gender and preferred reading. The observed frequency (or count) of each possible joint event is summarized in the contingency table shown below. Calculate the correlation and expected frequencies on the data distribution for both attributes. | | 7 | CO2 | | K3 | | 2.8.1 |
| **UNIT - III** | | | | | | | | |  |
| 6 | a | Differentiate OLAP and OLTP systems | | 7 | CO3 | | K2 | | 2.6.4 |
|  | b | Discuss three-tier data warehousing architecture with a neat sketch | | 8 | CO3 | | K2 | | 1.7.1 |
| **(OR)** | | | | | | | | |  |
| 7 | a | Design a data warehouse for a regional weather bureau. The weather bureau has about 1000 probes, which are scattered throughout various land and ocean locations in the region to collect basic weather data, including air pressure, temperature, and precipitation at each hour. All data are sent to the central station, which has collected such data for more than 10 years. Your design should facilitate efficient querying and online analytical processing. | | 7 | CO3 | | K3 | | 3.6.1 |
|  | b | Illustrate Data cube? List and explain star, snow flake and fact constellation schemas. | | 8 | CO3 | | K2 | | 2.6.4 |
| **UNIT - IV** | | | | | | | | |  |
| 8 | a | Distinguish supervised learning from unsupervised learning | | 7 | CO4 | | K4 | | 2.6.4 |
|  | b | State K-means algorithm. Apply k-means algorithm with two iterations to form two clusters by taking the initial cluster centres as subjects 1 and 4 | | 8 | CO4 | | K3 | | 2.8.1 |
| **(OR)** | | | | | | | | |  |
| 9 | a | A database has five transactions. Let min sup = 60% and min conf = 80%.    (a) Find all frequent itemsets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes.  (b) List all the strong association rules (with support s and confidence c) matching the following meta rule, where X is a variable representing customers, and itemi denotes variables representing items (e.g., “A,” “B,”): | | 8 | CO4 | | K3 | | 3.5.1 |
|  | b | What is prediction? Explain about Decision tree Induction classification technique. | | 7 | CO4 | | K2 | | 2.5.2 |

|  |  |  |
| --- | --- | --- |
| **Designation** | **Name in Capitals** | **Signature with Date** |
| **Course Coordinator** | Dr J.KARTHIK |  |
| **Program Coordinator** | Dr K.SRINIVAS |  |
| **Head of the Department** | Dr D.RAJESWARA RAO |  |