Most Likely Upcoming Questions

Here's a breakdown of likely future exam questions based on the provided past papers, considering frequency and recent appearance:

- **Unit - I**
    - What is data science? Explain the fundamental data mining tasks with suitable example for each. [8-2021, 8-2023]
    - With a neat diagram, explain how data science is placed in the context of various data related process. [8-2023, 10-2022]
    - Distinguish between Bigdata 1.0 and Bigdata 2.0. [6-2023, 4-2022]
    - Illustrate CRISP model for data mining and its stages. [6-2023]
    - Give the crisp model of data mining. Explain in detail its stages. [7-2021]
    - Explain the difference between Supervised and Unsupervised learning. [6-2021]
- **Unit - II**
    - What is probability estimation? Give all the steps to construct the probability estimation tree with smoothed probabilities. [10-2021, 8-2023]
    - Construct probability estimation tree with smoothed probabilities for the following dataset. [12-2023]
    - Construct a probability estimation tree with probability for the following data set. [12-2023]
    - Define entropy and information gain. Construct a decision tree for the following data set. [10-2022]
    - Demonstrate with example how attribute selection helps in constructing the decision tree. [8-2023]

- **Unit - III**
    - Write the equation of a linear model and generate a linear discriminant function for the following data and predict the target value for the sixth and seventh tuple? [8-2023]
    - For the given training set fit a linear discriminant model for classification and classify the last instance. [8-2023]
    - Illustrate support vector machines and objectives function with a neat diagram. [6-2023]
    - Explain with example how you can get class probability estimation using logistic repression. [5-2021, 7-2022]
    - List the objectives of support vector machine? How they are handled? Explain with suitable diagrams. [6-2022]
- **Unit - IV**
    - How K-means clustering is done? Explain with an example. [7-2021, 8-2023]
    - Find the distance between 1 and all others in the following table and identify the k-nearest neighbors. [12-2023]
    - Consider the following data set, classify Punit based on following methods. i) Majority voting ii) Similarity moderated voting [12-2023]
    - Construct a single link Dendrogram for the Euclidean distance matrix given below. Show all the intermediate distance matrices. [10-2022]
    - Discuss the issues with Nearest-Neighbour methods and explain how to overcome them. [6-2022]
- **Unit - V**
    - Outline a short note on: i) Link prediction and social recommendation ii) Evaluating classifiers [10-2023]
    - What is meant by association analysis? Compute, i) Support ii) Confidence iii) Lift iv) Leverage if the 10000 transactions analyzed, the data shows that 6000 of the customer transactions include computer games, while 7500 include videos and 4000 include both computer games and videos. For the rule, Boys (X, “Computer games”) ⇒ boys (X, “Video”). [10-2023]
    - How do you measure the term frequency and inverse document frequency with a suitable example. [6-2021, 10-2022]

    - Define Bag of Words. Give the steps to calculate the term frequency of a word document. [6-2021, 6-2023]
    - Analyze the basic measures for text retrieval and its methods in detail. [8-2023]

Key Observations & Reasoning:

- **Frequency Matters:** Questions appearing multiple times are highly probable.
- **Recent is Relevant:** Questions from 2023 are more indicative of future trends than those from 2021 alone.
- **Calculation/Application Questions:** Questions that involve calculations (e.g., building trees, discriminant functions, association rule metrics) are consistently popular. Expect variations on the data, but the *method* will likely be tested.
- **Definitions & Explanations:** Expect questions asking you to define key terms and explain concepts, often with examples.
- **Diagrams:** Questions that explicitly ask for diagrams (e.g., SVM objective function, data science context) should be prioritized in studying.
- **Comparison Questions:** Questions asking you to differentiate between two or more methods are highly likely to appear.