

Sample Questions 1

1. There are many data mining tools available. Please do a research about such tools over the Internet and list five of them.
2. How is a data warehouse different from a database? How are they similar?
3. Outliers are often discarded as noise. However, they are useful and can be used for some purposes. How can they be used and in what applications are they important?
4. Suppose your task as a software engineer at Anadolu University is to design a data mining system to examine their university course database, which contains the following information: the name, address, and status of each student, the courses taken, and their cumulative grade point average (GPA). Describe the architecture you would choose. What is the purpose of each component of this architecture?
5. Define the following terms: Data, data mining, knowledge, outlier, and prediction.
6. List the major data mining functionalities.
7. List the major data preprocessing tasks. Explain two of them briefly.
8. Define collaborative filtering with a simple example. List its major problems.
9. List at least five binary similarity measures.
10. Explain cosine and Pearson correlation coefficient similarity measures.

P.S. The questions given above are sample questions related to the Introduction Chapter. You do not need to hand in their solutions. However, you will take a quiz including one or two of these questions in this week class. You should solve them in order to be prepared for the quiz.