Data Mining Assignment 1

Identify a problem from your own experience that you think would be amenable to data mining. For that problem describe:

- Data mining is a process of discovering patterns in a large set of data and data warehouses. One of my friend's mobile application about supermarket requires data mining. For this, data of the previous purchases of customers need to be discovered and stored to guess the choices of customers accurately.

1. What the data is.?

- In computing, data is information that has been translated into a form that is efficient for movement or processing. This information can be in the form of audio clips, images and programs. Based on the requirement of the problem, the data can be modified into the required format.
- In this problem, the data is the previous purchases of various customers using the application.

2. What type of benefit you might hope to get from data mining.

- From data mining we can understand the structure of the data and that helps in identifying the type of data that needed to be used to provide the solution to the problem we have. Before collecting, the processing of data is a greater step and one which is difficult. We have to clean the data and make it ready to make predictions on the data.
- By retrieving the data for this application, sales can be improvised as after collecting the data from the stored data base, as we can clean the data and then the similarity patterns can be identified.

3. What type of data mining (classification, clustering, etc.) you think would be relevant.

- Based on the problem we have to see which technique we want to use. If a value is needed to be predicted, classification but when need to identify a pattern from the data we have to use clustering method on the data.
- So, for the problem I considered, I would use clustering because I need to identify the pattern by using which I can predict a value. Instead of just saying Yes or No for every single product offered in the store.
- For example, if a customer 'x' buy chocolates twice a week, then I can predict that he would buy chocolates the next week or not.
- Hence, I think clustering is relevant for this problem.

4. Name one type of data mining that you think would not be relevant, and describe briefly why not.
For each, illustrate with an example, e.g., if you think clustering is relevant, describe what you think a likely cluster might contain and what the real-world meaning would be.

- I think classification is not relevant for the problem I considered. Classification is used where I use the trained data set to classify the values I have.
  In this problem, if I need to follow classification, based on the previous products a customer bought, I should be able to say whether he'll buy a product I have or not.

- For suppose, a customer ‘x’ bought milk and jam. Then, I can say whether he would buy the bread or not in classification.

- So, I think classification is not relevant for this problem.

Write one to two pages of 11 point single-spaced typeset text - you aren't writing a paper, but it isn't short answer either.