Data Mining Assignment 1

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

Problem: Electronic shopping

1. What the data is.

The website holds every information of customers, the demographics details, transactional data, type of products bought and so on.

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

Using data mining, retailers can fully leverage the buyer's data to generate new insights that could enable them to make more profits and improve competitiveness. It can optimize the marketing campaigns by identifying the probability of the purchase of an individual products and can develop effective methods for cross-selling and up-selling. Also, knowing customers' buying behaviors can further quantifies the degree to which a customer is satisfied with a product and services.

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

Association analysis algorithm would be well suited for this area. Association rule mining is typically used to discover unknown relations that exist within data.

It can be used in many aspects and one of the most common use can be to find the products which are complementary to each other and are frequently consumed together.

For example, if a customer purchase product A, you can display other products like B, C, D and E which are usually purchased along with A.

Other types of association rules you can use is to discover links between the products added and not added to the basket together. Based on this, you can find significant customer buying behavior like product type preferences and favorite brands.

Also, by taking into account of the repeatability nature of purchases over time, you can further customize the appearance of the website to the customer.

4. Name one type of data mining that you think would not be relevant, and describe briefly why not.

Regression analysis may not be relevant to this specific area. As Association analysis focus on understanding a phenomena. They look for relationships between variables and outcomes, but they might not have predictive power. Prediction studies use many variables to create predictors. They learn patterns in the training data to make predictions on new data.

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.

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.