Data Mining Assignment 1

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

**Churn prediction** is one of the most popular Big Data use cases in business. It consists of detecting customers who are likely to cancel a subscription to a service. This can be telecom companies, SaaS companies, and any other company that sells a service for a monthly fee.

1. What the data is.

The data is the records of previous customers of the service. The attributes include many factors like monthly fee, disinterest etc.

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

With the help of help of this, the organizations can find out exact reasons for the churn and work on the factors that are causing the churn to occur.

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

Classification is relevant data mining technique because the organization just needs to classify the users as churn predictors or not.

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

Regression is irrelevant for this prediction because there are no continuous values that needs to be predicted based on the user history.

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.