Data Mining Assignment 1

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

1. What the data is.
2. What type of benefit you might hope to get from data mining.
3. What type of data mining (classification, clustering, etc.) you think would be relevant.
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.

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.

Cancer are a common cause of death now-a-days. Many people are dying due to lack of knowledge on their medical and health conditions. So when a model is predicted where people can enter their basic health details like weight loss, blood in urine, night sweats..etc. Then based on the attribute values entered the severity of occurrence of cancer can be predicted earlier.

1. What the data is.

Count of patients, Types of Cancer, Count of people who have symptoms, count of people who are getting medication and count of people who died due to cancer.

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

- Data mining technology is something that helps one person in their decision making and that decision making is a process wherein which all the factors of mining is involved precisely
- Predictions for recovery based on type of cancer or type of medication or based on other health conditions.

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

Based on type of cancer, regression helps to deal with the numerical data which helps to analyse the number of people getting affected and also recovery and death rate.

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

Classification helps deal with different kinds of cancer like one type of cancer can have different numerical data like death rate, recovery rate and other types of cancer can have other data. So, these types can be classified for easy analysis. But it won't be suitable for my dataset because I have data only for one type of Cancer. So, which is why classification can't be done.