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.
Facebook linking you to certain political news pages that you open. Then my feed is flooded with that news subject which doesn't spark my interest all the time.

2. What type of benefit you might hope to get from data mining.
Knowledge and patterns of certain subjects which can affect decision making and operation.

3. What type of data mining (classification, clustering, etc.) you think would be relevant.
Clustering which divides data into similar patterns. If someone links to one page on Facebook it is clustered into pages of similar content or affiliation.

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.
Regression since it is numerical and usually used to measure future data and make certain predictions.

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.

From my own experience, I had a faced a problem in Facebook during the election period. The problem is about some political news pages came in my news feed without liking those pages. Then my news feed was flooded with that uninteresting political news and made me to lost my interest to see Facebook. By learning some concepts in data mining, I got to know that it is because those political news pages made their promotion in Facebook to make know about them. By this I got to know that this is because of Clustering concept in data mining. This Clustering concept is processed as if someone links to one page on Facebook, it is clustered into pages of similar content. And it will be shown in news feed as a recommendation to some people. I think one of the types of data mining, the concept of Regression would not be relevant to my problem as this type of data mining deals with numeric data and usually used to measure future data and make predictions.