Forrest Fallon  
IST 707  
10/6/21  
  
1a. Dividing the customers of a company according to their gender.

No, this task is simply separating data into already known qualities. This is done via one query.

1b. Dividing the customers of a company according to their profitability.

No, same as #1, this task is simply sorting out data that is already quantified.

1c. Computing the total sales of a company.

No, this is a simple summation of already known data.

1d. Sorting a student database based on student identification numbers.

No, this is a simple organizational query.

1e. Predicting the outcomes of tossing a (fair) pair of dice.

No, the statistical values of “fair” dice are already known.

1f. Predicting the future stock price of a company using historical records.

Yes, since we are predicting the future value of a complex system, we would need to use data mining techniques in order to do so properly.

1g. Monitoring the heart rate of a patient for abnormalities.

Yes, this data is not currently known or easily predicted based on previous results, therefore we would need to use data mining techniques in order to properly study the heart patterns.

1h. Monitoring seismic waves for earthquake activities.

Yes, we would be classifying the seismic activity into “earthquake” and “non-earthquake.” We would then use those classifications to model the activity currently being registered.

1i. Extracting the frequencies of a sound wave.

No, we are not predicting or classifying here, just monitoring and storing information.

2. We will never be able to predict with 100% certainty where a person will go on our search engine, however, the closer we get to 100% certainty, the better user experience overall. With association rule mining we can give the user related suggestions as to where to go next based off of their first search; that is only possible with classification and clustering. Anomaly detection would allow us to notice when the previous method is not working with certain groups of people, and we can then adjust the algorithm accordingly.

3a. Census data collected from 1900–1950.  
 No, this data is already widely known and used in studies both previous and ongoing.

3b. IP addresses and visit times of Web users who visit your Website.  
 Yes, user data is under heavy protections and cannot be mistreated.

3c. Images from Earth-orbiting satellites.  
 No, although this information is useful, it is often not detailed enough to put anyone’s privacy at risk.

3d. Names and addresses of people from the telephone book.  
 No, this is already known information available to the public.

3e. Names and email addresses collected from the Web.  
 No, similar to the phone book question, any info that is publicly available is already non-private.