George Smith

IST 707

HW 1

**Task 1**

1. Dividing the customers of a company according to their gender
   * Dividing the customers of a company according to their gender would not considered a data mining task as it would not require data science skills to complete this simple grouping exercise .
2. Dividing the customers of a company according to their profitability
   * Dividing the customers of a company according to their profitability would not be considered a data mining task problem as it would not require data science skills to complete this simple grouping exercise.
3. Computing the total sales of a company
   * Computing the total sales of a company would not be considered a data mining task, a company should have their sales information readily availability.
4. Sorting a student database based on student identification numbers
   * Soring a student databased based on student ID numbers would not be considered a data mining task as this is a simple filter on available data.
5. Predicting the outcomes of tossing a pair of dice
   * Predicting the outcomes of tossing a pair of dice would not be considered a data mining task as the outcomes are easily calculated using basic statistics.
6. Predicting the future stock price of a company using historical records
   * Predicting the future stock price of a company using historical records would be considered a data mining task as this is a predictive task which would require accurate data.
7. Monitoring the heart rate of a patient for abnormalities
   * Monitoring the heart rate of a patient could be considered a data mining task as advanced data and modeling techniques are necessary to detect anomalies.
8. Monitoring seismic waves for earthquake activities
   * Monitoring seismic waves for earthquake activities could be considered a data mining task if we are trying to predict the likelihood and strength of potential earthquakes.
9. Extracting the frequencies of a sound wave
   * Extracting the frequencies of a sound wave would not be a data mining task as there is no decision making output

**Task 2**

In the article Google Flu Trend: The Limits of Big Data social scientists are crucial of Google’s flu-tracking service, saying that it wildly overestimated the number of flu cases in the United States. Google has been accused of “ big data hubris”, or the assumption that big data sets trump traditional data collection and analysis. The author’s claimed that using the recent trend of C.D.C reports from doctors on influenza like illness would have been a more accurate predictor than Google Flu Trends. The technical criticism of Google Flu trends is that it is not using a broader array of data analysis tools. May respected authors and academics point to google flu trends as a proof of the triumph of the big data approach, Tracking 45 flu – related search terms over billions of searches, monitoring trends and making correlation would win out . Google Flu trends could be better suited to give early warning signals of flu outbreaks one to two weeks ahead of the C.D.C surveillance reports. Google continues to review the service every year and make improvements.

In the Article In Defense of Google Flu Trends the author begins by summarizing much of the criticism Google has faced regarding its big data efforts to predict flu trends. It soon mentioned that much of this criticism is not just, the CDC was aware that by combining Googles Flu Trends data with the CDCs own data they would have a better epidemiological understanding of the country. Google never meant for their system to be a replacement for traditional surveillance networks. There is great value that comes with triangulating these flu tracking systems.

My thoughts on Googles Flu trends experiment is that it should be considered a success. Although it may not accurately predict the amount of flu cases, it has proven to be useful for generating signals related to the flu. I do not believe the criticism that Google has experience related to their Flu trends is just. Google has created a data set that can be combined with traditional prediction methodologies to get more accurate results, because of this Google Flu trends should be considered a success.