

Series 4

1. Collecting data: Determine how you would collect the data (categorical, semantic differential, rank-order, multiple-choice) and give an example:
 - a. You want to evaluate the preferred characteristics of a new product (different forms, different colours, and so on)
 - b. You want to test the analytical thinking of job candidates by a test during the hiring process.
 - c. You want to rate the importance of the opening hours in the evening of a local store.
 - d. You want to find out how many foreign languages people are speaking.
 - e. You want to find out the personal personality traits.
 - f. You want to know the 5 most important benefits of a new mobile app.
 - g. You want to know the least useful and most useful features of a new vacuum cleaner.
 - h. You want to find out what people are thinking about the political agenda of a party.

2. A bank is asking you in supporting them in rogue trading analytics in the front office (i.e. the business department where the traders are working and performing). The management of the bank is expecting that you are performing data analytics and with advanced analytics procedures you would then detect patterns of a possible rogue trading. You know that there are several departments with traders. The back office does the work for all trading departments and are booking and reporting all trades.

Before you can start with applying methods, you need data. What are all the questions (a bit concretised to this example) you would ask for getting the required data?

3. The R-Script “TextMiner.R“ illustrates an example of text mining with R. The text files “1.txt” and “2.txt” contain two distinct speeches of American presidents (“State of the Union Address”). If you copy these two files into your current R work directory and let “TextMiner.R” run, then the 20 most frequent words of the speeches are depicted. This will enable you straightforwardly to allocate one speech to George W. Bush and the other one to Barack Obama.
 - a) Read through the code and grasp its functionality. Most code lines are explained. Further details can be found in Chapter 10 of the online script <http://www.rdatamining.com/docs/r-and-data-mining-examples-and-case-studies>.
 - b) Apply the script to a text or a series of texts of your choice. Update “TextMiner.R” accordingly, if required. Further fine tuning of the code may be reached by completing the list of “stopwords”.

What are your findings?