**Question 2.1**

Describe a situation or problem from your job, everyday life, current events, etc., for which a classification model would be appropriate. List some (up to 5) predictors that you might use.

# An everyday situation for which a classification model would be appropriate would be the classification of email as spam or not. In this particular case, an algorithm is used to automatically sort out electronic messages in an effort to streamline our email inboxes and generally improve the user experience while using specific email services. There are multiple predictors such an algorithm might use to classify incoming emails as spam or not – these may include:

# Recognition of the person/entity sending the incoming email (whether said entity is a known contact/well-known organization)

1. Commonly recognized spelling or grammatical errors in the text of the received email

# The inclusion of links to external webpages in the email (especially unsecured or potentially malicious links)

1. The presence of attached documents in said email (as well as the format of said documents, e.g. .txt, .doc, .exe)
2. The inclusion and amount of numbers (pricing, phone numbers, etc.) present in the email text

# There are many more subtle factors that could be used in parsing through the text of the email (for instance, the frequency of word repetition in the email, the complexity of the language used, the presence of significant words that could be pre-marked as “red flags,” etc.). A classification model would be appropriate in this instance because we are trying to categorize the emails into two distinct qualitative groups (“spam” or “not spam”) based on a given number of predictors.