Week 10, Project4, Business Requirements

1. Determine data set to be used
   1. Suggested datasets of spam and ham (non-spam) are available at:  
      <https://spamassassin.apache.org/old/publiccorpus/>
   2. Notes on these files are available:
      1. The first video shows how to unzip the provided files  
         <https://youtu.be/6IzhRaSePKU>
      2. The second video provides a short overview of predictive classifiers.  
         <https://youtu.be/5ikDo4SrLNQ>
2. Download and store the dataset(s)
3. Unzip the files (these files need to be unzipped twice)
4. Use a portion of the dataset to create/train a model to classify the documents
5. Use the model to classify a portion or all of the records not used in training
6. Deliverables:
   1. .rmd file
   2. Rpubs URL
   3. GitHub URL
7. Adjusted due date: **2018-11-06 11:59pm EST**