

Instructions: Show work or attach R code used to perform calculations (or any other technology used). Be sure to answer all parts of each problem as completely as possible, and attach work to this cover sheet with a staple.

1. For each of the following machine learning model types, describe the kind of data they can be used on, one specific algorithm that is in the category, and at least one advantage and one disadvantage.
 - a. Natural Language Processing
 - b. Sentiment Analysis
 - c. Clustering
 - d. Classification
 - e. Regression
 - f. Map Reduce/Parallel Processing
 - g. Graph Theory
2. What is the purpose of splitting data into test and training sets? K-fold cross-validation?
3. Using the data set provided with either Homework #4 or Homework #5, experiment with an appropriate machine learning model in R. (You may need to download an additional package to do this.) Compare your outcomes to the outcomes from the corresponding homework question. How accurate is the model? Were you able to improve it? How easy is the model to interpret?