­­Class Project – Apply Data Science Principals to a Dataset of your Choosing

Find a dataset of your own choosing to perform some analysis on. The dataset must be:

1. Of an appropriate size. Needs to be large enough for the algorithm to learn some patterns that solve a particular problem. However, for grading purposes, not too big – can be emailed
2. Must not be proprietary data, either publically available or private data shared with the permission of the owner for use in this class.
3. Allow you to solve one or more prediction problems using machine learning algorithms

Once the data has been curated, determine 1-2 problems you can use the data to solve using a machine-learning algorithm. Using R, Python, Java or Weka, perform:

1. Feature selection – this can be very simple, using correlation

**Deadlines**

10.16.2014 – Submit