STAT 517 – Final Project Critique

Ryan Heiderman: “Forest Stand Density and the Effect of Climate, Soil, Topography and Species Mixing”

Reviewed By Gavin Whitesitt

Coverage of 3 course areas – Supervised/Cluster/Association: All three categories are thoroughly explored. For supervised learning regression is used to predict lnTPA and lnQMD (linear regression, ridge regression, lasso regression, decision tree regression, random forest regression, knn regression, and neural networks). Quantile regression is also used to determine maximum stand density index. For the unsupervised learning portion of the class principle component analysis is performed along with clustering as well as association and association rule mining.

Dataset size meets stated criterion: The dataset used has over 110,000 plots (rows) of data across 41 variables for a column \* row total of 4,510,000. Well over the suggested 100,000 minimum requirement.

Relevancy: The dataset is relevant regarding the question Ryan is asking. Personally, I didn’t have much of a prior interest in forest health and productivity, but after reading Ryan’s paper I think I may have developed one. The health of our forests is important in ensuring that we can continue to enjoy nature for generations to come as well ensuring that wildlife has the appropriate habitat to survive in. I think that with the seemingly large number of wildfires this past summer and the effort by some politicians to cut funding for the EPA and other similar organizations the topic has general relevancy to national discussion.

Difficulty: As far as I can tell the questions being asked are not completely straightforward and the methods used to answer them are more advanced than just the ones covered in class. It’s obvious that a lot of time was put into thinking about the questions being asked and the methods used are implemented appropriately. I think Ryan’s final project stands above many of the other projects in the class because of the intentionality of the project.

Interestingness: I thought that the question was interesting and found myself invested in the results. I think that this can be particularly hard to do when talking about data, but Ryan crafts a story you want to hear with data.

Clarity: Ryan does a good job of explaining exactly what he is doing in plain English. In machine learning in general I find it somewhat rare that I can read a paper and understand it well after the first read. I felt like I only had to read Ryan’s paper once to get a good idea of what it was all about.

Originality: I think that the question it interesting and the use of quantile regression to show the outer boundary was particularly unique. Extra points for that.

Creativity: A good example some of the interesting aspects of Ryan’s paper was his use of a transformation of aspect using sine and cosine to show a measure of east west. I understand that this was suggested from some of the relevant literature, but I think most people would just drop the aspect data and consider it unusable. I liked that Ryan didn’t do this and went the extra mile to add some more potentially valuable information that his model could train on.