Time Line of Project – Ryan Dearing

Step 1

* Load data - COMPLETE
* Format Data - COMPLETE
* Consolidate - COMPLETE
* Select Data - COMPLETE
* Filter Data - COMPLETE
* Explore - COMPLETE

Step 2

* Determine data bins - COMPLETE
* Bin data – COMPLETE

*Submit Initial Results* – COMPLETE

*Obtain Feedback -* November 9th

Step 3

* Implement word tokenization algorithms – November 12th
* Implement normalization algorithms - November 12th
* Remove numbers, stop words, punctuation, etc. November 13th
* Generate document term matrix - November 13th

Step 4

* Split into training and test sets – November 19th
* Build Naïve Bayes text classification - November 19th
* Use NB model to create predictive set - November 19th
* Evaluate NB model (confusion matrix, accuracy, recall etc.) - November 19th
* Build Support Vector Machines text classification - November 20th
* Use SVM model to create predictive set - November 20th
* Evaluate SVM model (confusion matrix, accuracy, recall etc.) - November 20th
* Select the most appropriate method - November 20th

Final Submission

* Prepare charts, graphs, etc. – November 21st
* Prepare formal submission – November 22-24th
* Proof read and submit – November 25th

Final Presentation

* Summarize overall report – November 26th
* Prepare slides for presentation – November 27th