Progress Made Thus Far:

We have started building a foundation for an application that will accurately classify sarcastic tweets. We started by brainstorming different classification modeling techniques, features to engineer, and researching existing white papers on more advanced applications. Through our initial research, we decided to implement a series of models ranging from naive to advanced including Decision Tree Classifier, Logistic Regression, Linear SVC, SVC with radial-basis kernel, K-Nearest Neighbors Classifier, Gaussian Naive Bayes, Random Forest Classifier, AdaBoost Classifier and Gradient Boosting Classifier. We first spot-checked these models (using default parameters) via 10-fold cross-validation on the entire training set. The results showed that linear models such as logistic regression and linear SVC worked as well as did the more complex nonlinear models. We then proceeded to perform hyper-parameter tuning (using 10-fold cross-validation and F1 score as the metric) to optimize the models. Equally as important as the modeling techniques, we needed to engineer features to capture the nuances that determine whether a tweet is sarcastic. So far, we have implemented features to represent the number of users tagged in the tweet, number of hashtags, length of tweet, and number of characters in the tweet. We also have tried to capture the sentiment of the tweet by creating features to represent ellipses, laughter, affirmations, negations, interjections, intensifiers, punctuation and emojis. After creating these features, we also used TSNE to visualize our features to get an idea of how they cluster and represent the data thus far. Overall, we have seen the best performance with the linear/regression models compared to more complex models such as AdaBoost or the Gradient Boosted Classifier. Our best metrics thus far are as follows: Precision = 0.6006, Recall = 0.8355 and F1 = 0.6988.

Remaining Tasks:

Our largest area for improvement of our current work is to engineer additional features to capture more of the nuances of sarcastic tweets. Since our complex models have not performed well thus far, we think this indicates that our features are not complex enough to capture characteristics of sarcastic tweets. We think our biggest wins will come from creating features based on the context of each tweet. Since sarcasm is often a response as part of a conversation, we are hoping this will improve model performance. We also plan on implementing features to represent text patterns in sarcastic tweets like parts of speech, n-grams, and topics represented in the tweet using LDA or a similar technique. We also think it would be interesting to implement a deep learning model after the additional feature engineering and evaluate how this performs against the other models. Additionally, we plan to explore an Ensemble model that combines the models that have performed well to see if we can optimize performance that way. In the end, we plan to have a robust suite of models that are trained on data with features that capture the nuances of sarcastic tweets.

Challenges/Issues:

Our main obstacle we are currently facing is that we are not reaching baseline performance. With the remaining tasks that we have detailed above, we believe we will be able to reach baseline accuracy with our final implementation. Our other challenge has been time. Between finishing up content for the course and preparing for the second exam, we have not been able to prioritize working on this project. Half of our team has now completed the exam so we are not concerned with time being a challenge as we complete this project over the next couple of weeks.