Alex Buslawski

SpringBoard Capstone Project Proposal

For my Springboard Project, I’ll be taking a dataset from Twitter containing variables including but not limited to account creation date, number of followers, what type of account this is, and actual tweets from these accounts. The goal of this project is to use this data to create and fit a logistic regression model to accurately determine whether an account is either real or a bot. At this point of time in history where nearly a quarter of the world population is connected through social media sites, and yellow journalism (or ‘fake news’) seems to be more of a problem now than it was even twenty or thirty years ago, being able to distinguish between genuine social media accounts and bots is important in being able to curb or, in the best case scenario, eliminate entirely, the spread of yellow journalism. The ability of Twitter accounts to retweet other accounts and use hashtags to get topics trending is a big deal because it can be abused by bot accounts run by a single organization to retweet news articles by biased sources and get topics trending that normally wouldn’t be trending. In short, social media bots (socbots) are a valuable tool in spreading propaganda. This dataset contains labeled data of both genuine Twitter accounts and Twitter socbots. As for now, the only deliverables for this project would be the deliverables assigned to me by the course. I’ll have to wrangle the data, apply data storytelling and statistics to the data, and finally fit the regression model.