Reddit calls itself “The Front Page of the Internet”, and there has never been as more appropriate slogan. The website is a mix of news, social commentary, cool links, and pictures of cats. It boasts over 250 million users and 542 million monthly users. These users drive posts into the spot light using “upvotes” or make them disappear by “downvoting” them. It’s clear to see why this is a wonderful platform to expose your idea, campaign, or website.

But understanding the way posts become popular can be a bit tricky. The website is made up of numerous sub-pages called “sub-reddits”. There are 853,824 currently and that number is ever growing. This is merely one in many variables that need to be accounted for in order to make sure your product is out there for all to see.

Looking at comments on a post give us a clear indication of how popular a post is. Simply put, the more comments the post has, the more popular/controversial the post is. Luckily the website makes it incredibly easy to access the information we need by simply scraping the html elements through a python script. Using the aggregated data, we can build classification models and predict what objects influence a post having a high-comment threshold.

Our baseline accuracy was 50.3%. Using a random forest classifier model and setting our independent variable as the subreddit, we were able to produce a model that was 10.3% more accurate. Through this model, we were able to see that /rFacepalm was the most important feature in the model, with a coefficient of 0.0245.