**Dataset Description**For this project, a dataset which includes information on all Kickstarter Projects in 2018 will be used in tandem with the Random Forest algorithm to predict a project’s success, failure, or cancellation. The dataset has around 379k observations with 15 attributes to work with. The data was made available on Kaggle and obtained from www.Kickstarter.com.

**Variables for Analysis**The target variable for this analysis will be “state” which is simply the state the project is in: success, failure, or cancellation. Success means the project met its monetary goal by the deadline, failure implies it did not, and cancellation refers to the Kickstarter campaign being scrapped prematurely. The attributes are a mixture of qualitative and quantitative variables. The quantitative variables include the funding goal (in project currency and in USD), the number of backers, and the amount of money pledged (in project currency and in USD). The qualitative variables include the project category (games,film,music…), the country the project is in, and the currency of the project. Launch Date and Deadlines will be used to examine how the length of the campaign affects the state of a project.

**Problem Statement**Modern technology and social media provide more and more opportunities for entrepreneurs and creative people to see their ideas and ambitions through. Kickstarter.com is a platform which allows such individuals or organizations to obtain financial support for their concepts by sharing them with the public over the web. The project owner creates a funding deadline and monetary goal to reach. If they do not reach this goal by the deadline, no money changes hands and the campaign is over. However, if they do manage to achieve their funding goal, they will be able to enter the development phase. The goal of this analysis is to use data obtained from Kickstarter to predict if a campaign will fail, succeed, or get cancelled prematurely. Once the significant factors for determining project success rates are realized, the analysis can potentially be used to increase the number of successful Kickstarter campaigns, up from 75%.

**Methodology**First, the data will be preprocessed to handle any missing values or invalid data. Then a simple random forest algorithm will be run on the data with default parameters without feature selection to come up with a base model for this analysis to build off of. From here, parameters will be altered, and feature selection processes will be explored to come up with a model that performs and generalizes well. To evaluate the model, cross-validation as well as the train/test split will be used to come up with an accuracy and AUC score. The out of bag error will also be discussed in the evaluation.

**References**Mitra, Tanushree, and Eric Gilbert. "The language that gets people to give: Phrases that predict success on kickstarter." Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. ACM, 2014.

Etter, Vincent, Matthias Grossglauser, and Patrick Thiran. "Launch hard or go home!: predicting the success of kickstarter campaigns." Proceedings of the first ACM conference on Online social networks. ACM, 2013.

Marom, Dan, Alicia Robb, and Orly Sade. "Gender dynamics in crowdfunding (Kickstarter): Evidence on entrepreneurs, investors, deals and taste-based discrimination." (2016).