KEVIN PHAN

500558667

CKME 136 DATA ANALYTICS: CAPSTONE COURSE ABSTRACT

FALL 2018

The ensuing capstone project will explore the world of entrepreneurship as we look at data surrounding the online ventures. This project makes use of three data sets with approximately 100,000 rows; containing descriptions and metrics of over 300,000 Kickstarter projects. We will identify the ideal attributes of a Kickstarter project resulting in goal completions. The research question is as follows:

What attributes of a Kickstarter Projects are the most profound predictors of whether the goal would be reached or not? What is the driving factor to a successful Kickstarter? Using these findings, the project aims to develop a guideline for future entrepreneurs to follow to reach their financial goals

The Data used consists of three CSV formatted datasets, each containing their own unique attributes relating to the Kickstarter project. The attributes, a combination of nominal, categorical and ordinal will be outer-joined into one data frame using the Kickstarter ID as a primary key. Each dataset contains approximately 35,000 rows and about 15 attributes. Techniques used will be KNN regression, logistic regression. and a statistical A-B test (confusion matrix and comparison of RMSE and R Squared values) to determine which model better classifies Kickstarter projects. The classifications would be Success or Failure as depicted in the datasets. Once the models have been created, we will separate each into a testing and training set and graph our predictions to errors on a graph that will later be determined. Data techniques such as cleansing, outlier-detection and normalizations will take place.