Jack Bonacci

Capstone Project Proposal

13 July 2018

**What is the problem you want to solve?**

The goal of this project is to implement machine learning algorithms to predict with a high level of confidence whether a potential interviewee will attend his or her interview.

**Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?**

The client for this project is in general any firm that interviews a high volume of potential job candidates throughout a hiring season.  More specifically, individuals who work in human resources departments or whose commission based pay depends on interviewing a high number of candidates. For example, a friend of mine is on a path to being a branch manager for a mortgage finance firm called Supreme Lending.  The first part of his path to being a manager is a year of working at headquarters to gain a better understanding of how the business works as well as what kind of people fit the mold for a Supreme Lending employee. He is paid commission based on scheduled interviews, but earns more if those interview actually take place.  Someone in this position would be able to decide whether scheduling an interview with a potential candidate is worth the time of interviewers within the company.

**What data are you going to use for this? How will you acquire this data?**

The dataset for this project comes directly from Kaggle.  I have already downloaded the data in a .csv format. There are 1233 observations and 28 variables (1234x28).  The specific measure that I am attempting to predict is the column titled “Observed Attendance”. A logistic regression model can be implemented to predict the probability that a candidate will attend the interview.

**In brief, outline your approach to solving this problem (knowing that this might change later).**

A logistic regression model can be implemented to predict the probability that a candidate will attend the interview.

**What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.**

Deliverables for this project include the code used to implement an algorithm, a training dataset, a testing dataset, and a paper explaining the entire process both through writing and through visuals. An ambitious goal of mine for this project is to host the model on a front-end application so that firms can easily use it without necessarily understanding what goes on behind the scenes.