# A Machine Learning Approach to Fundraising Success in Higher Education

Tiehu Jiang

August 21, 2019

## 1. Introduction

### 1.1 Background

I am a senior application developer from the Florida State University Foundation. The mission of our organization is to enhance the academic vision and priorities of FSU through its organized fundraising activities and funds management. Established in 1960, the FSU Foundation today manages an endowment of more than $509 million (as of June 30, 2019).

Like other nonprofits. we have collected and are managing over a million of records of constituent data. We implemented the web based Blackbaud CRM system over 5 years ago. which is an expansive, multifaceted constituent relationship management (CRM) software designed for nonprofits. Blackbaud CRM is a highly inclusive constituent database with a variety of features to boost the fundraising and donor stewardship efforts long-term.

With the big database in place, there's always a gold mine of insights that add great value to organization's business decision making. Data science has many tools in its bag such as Analytics, Deep Learning, Machine Learning, etc. to dig this mine.

But according to a survey carried out by Every Action, 90% of nonprofits indicate that they're collecting data, but "almost half say they aren't fully aware of the ways data can (and does) impact their work."

Our IS management has been aware of this trend and started the Data Analytics Workgroup early this year. Our first step is to train ourselves by enrolling the IBM Data Science Professional Certificate program. The Foundation is going to reimburse us up to $300 for completion. Since we brought problems to learn, I would like to utilize our own dataset and use the skills and the tools learnt from Coursera to solve the problems. This is the reason why I choose my Capstone project as a machine learning approach to fundraising success in higher education.

## 1.2 Problem

New donor acquisition and current donor promotion are the two major programs in fundraising for higher education, and developing proper targeting strategies plays an important role in the both programs.

The size of the donor group determines the scope of fundraising. Acquiring new donors is always important for fundraisers. However, contacting randomly without a clear targeting strategy can be inefficient and may be disturbing for those who do not wish to be contacted. Keeping asking the wrong people will annoy them and give them a bad feeling about the institution. As a result, a proper targeting strategy that helps fundraisers to locate the potential donors is important not only for boosting fundraising efficiency but also for protecting the university's reputation. An efficient fundraising program should always start with developing a targeting strategy to accurately identify the pool of prospects.

The problem of targeting potential donors can be modeled as a supervised learning problem in machine learning, with the goal of identifying potential donors from all the candidates according to their personal and affiliation factors. There are two types of records in the data set: donors and rejectors. In particular:

- Donors: People who donated in the last 10 years and whose personal information is in database.
- Rejectors: People who rejected to donate in the last 5 years and never donated before and whose personal information is in database.

Since personal information is required in our model, samples used from the data set include alumni or alumni families or relatives. To properly train and test the model and to avoid random errors, the training and testing process is conducted using K-fold cross validation. The results are the average values from the cross validation.

## 1.3 Interest

Despite readily available alumni data warehoused at our server, the potential of alumni data has not been fully explored by fundraisers to enhance fundraising outcome. In this Capstone project, using machine learning methods, we will address two important problems in the higher education fundraising industry, identifying prospective donors and “promising donors", a term referring to donors who will upgrade their pledge. The first problem is to look for new donors from alumni, alumni families or relatives. The second problem is to look for existing donors who have a high potential to increase their donation.

# 2. Data acquisition and feature selection

## 2.1 Data sources

Dataset used in this project will be pulled out from our CRM report data warehouse, which is a Microsoft SQL-Server database. T-SQL will be used to select all the features we interest and make sure all record values have a good distribution to the features. The dataset that includes randomly selected donor information then will be exported to an .csv file for using in Jupyter notebook.

## 2.2 Feature selection

There are two type of features can be selected from our data warehouse: personal feature and affiliation feature.

### 2.2.1 Personal feature

- *Age*: Many experienced fundraisers believe that age can be one of the most important personal features that make a huge difference to their donation behavior. This is reasonable because different ages mean different life stages and imply different financial situation, child situation and life goals. And these factors together will make a difference in their donation behavior.
- *Income*: Wealth information is the most important information since it is believed to be highly correlated to the donation behavior. Even though the personal income is private information and we don't have it in our database, but for Florida state employees, it is open. We can get this information from "Florida Has Right to Know" web page. https://www.floridahasarighttoknow.myflorida.com/search_state_payroll
- *Sex*: Gender may be another important factor impacting the donating behavior.
- *Marital status*: Marital status might indicate the person's living condition, which may make a difference on their donating behavior. Marital status is available in our dataset.

### 2.2.2 Affiliation feature

- *Number of degrees achieved*: Because the number of degrees achieved may indicate the number of years that the donor has spent in a university, the number of degrees a donor received from the university should be a strong indicator that the donor will donate to the university.
- *Number of events registered*: The number of events registered is another important a

liation information, because it indicates whether the person is willing to keep in touch and maintain the connection with the university.

- *Family relation*: People cherish family relation. Most people are generous to their loved ones, and such favor is expanded to their loved ones' favorite things.

In our database, we can get some other features for each donor. But for simplicity we only choose above features for this project. we also have donor's address information and geocode information such as latitude and longitude in our database. This information could be useful for clustering donors with similar attributes and display them on Foursquare map for visualization. This is also helpful for fundraiser to identify potential donors. This can be a future work.