Emerging Markets: Comparing University Poverty Alleviation with Charitable Education Donations

A Model for Post-Secondary Investments

February 1, 2016

Summary

Contents

1 Introduction

1.1 The Problem

The Goodgrant Foundation is a charitable organization with a mission to improve educational performance of students attending colleges and universities in the United States. The foundation intends to donate a total of 100,000,000 USD to a group of schools each year, for five years. They also wish to avoid investing in universities that have already received grants from major charitable organizations such as the Gates and Lumina Foundation.

We present a mathematical model to outline an ranking and investment strategy to provide the maximal return on investment(ROI) appropriate for an education charity. The model is split into two sections and is applicable in choosing $1 \cdots N$ institutions.

The first section outlines determining bench mark points and comparing the rest of the institutions to them via a k-Nearest Neighbours(k-NN) Algorithm. In the second section, we derive a ROI measuring the effect the grant would have on improving student performance. In particular, our ROI reflects changes in terms of incoming poverty and outgoing poverty. In the Model Testing section, we input a subset of the U.S. National Center on Education Statistics data set and College Scorecard data set and output a list of $1, \cdots N$ schools, the investment amount, and number of students affected. In Sensitivity Analysis, we restrict the distance in the kNN algorithm to obtain a subset of our prioritized schools and assess the model's stability. Lastly, we summarize the findings of the paper and suggest future additions.

1.2 Assumptions and Rationale/Justification

- **The Gates Foundation and Lumina Foundation will continue to donate to the same school for at least two years:** Typically large charities often donate sums over the course of number years to allow highest impact. [3] We are able to remove those particular schools from our candidate list to avoid duplicating the two foundation efforts.

- **Students who receive a grant will stay in the program and successfully graduate:** There are many factors that could cause students to not be able to graduate, but we will focus on financial support because it is the root of other factors..[1]

1.3 Summary of our Approach

- Create a benchmark setting: top 10 schools for each degree type school, that we consider to have good performance. Each school represent a class.

- Identify universities that have similar characteristics to the benchmark list.

- Prioritize universities to be invested in by the amount of grants they received from other sources.

- Determine the amount of investment and allocation to maximize the return on investment.

2 Model Design and Approach

2.1 Notations and Definitions

1. $U_j^{[i]}$ = university j of class i
2. $U_o^{[i]}$ = benchmark university for class i
3. $A_j^{[i]}$ = amount invested in university j for class i
4. $T_j^{[i]}$ = Tuition in university j for class i
5. $n_j^{[i]}$ = number of students receving our grant in university j for class i
6. $IP_j^{[i]}$ = poverty rate of students entering university j for class i
7. $OP_j^{[i]}$ = poverty rate of students exiting university j for class i
8. $\phi_j^{[i]}$ = number of students in university j for class i
9. Φ = Total number of universities
10. $g_j^{[i]}$ = maximum amout of grants to be given to university j for class i
11. Group, refers to the Carnegie Classification of the university
12. Class, refers to which benchmark within the group, the university is closest to.

2.2 Benchmark Selection

First we omitted any below medium size universities or communal universities such as racial or religious schools. Then we partition all universities by number of years required for graduation. Lastly, we rank the universities by the amount of their poverty elevation standards (PES), this is done based on the result and recommendation of Trombitas [1] and College-Score Card [10].

$$R_j^{[i]} = \frac{IP_j^{[i]} - OP_j^{[i]}}{IP_j^{[i]}} \tag{1}$$

Next we choose the top, highest ranking universities per 'years of graduation'-group. We set these chosen universities as our benchmarks for each class in each group.
i.e.

$$o^{[i]} = \underset{j}{\operatorname{argmax}}\, R_j^{[i]}$$

Where o, is to represent the benchmark index of the class i.

2.3 Categorizing Institutions into Classes

We aim to classify each university into some class i. This is done via the kNN algorithm: a non-parametric method used in classification supervised learning problems. The output is determined by a majority vote, where the object is assigned to the class most common with it's neighbors. [6] We set up the following model: Assuming $j \in [1, n]$,where $n =$ amount of variables. For a detailed selection list please see Appendix 1. Let $y_j^{[i]}$ be the j-th variable for the i-th class benchmark. Let x_j be the j-th variable of a particular university. A university u, is considered a member of the k-th class if the variables x_i of that university is 'close' to a particular variable of a given class in euclidean-space.

That is, if u belongs to the k-th class then

$$k = \operatorname*{argmin}_i \sqrt{\sum_{j=1}^{n} (x_j - y_j^{[i]})^2}$$

2.4 Investment Strategy

2.4.1 Number of students-Helped

Let $A_j^{[i]}$ be the amount we invest in university j of the class i, and $T_j^{[i]}$ be the tuition of low-income students. Then we can compute the number of students who we help as:

$$n_j^{[i]} = \lfloor \frac{A_j^{[i]}}{T_j^{[i]}} \rfloor \tag{2}$$

2.4.2 Adjusted Poverty

Next we computed the adjusted poverty rate. We make a crucial assumption here about the correlation between grant-money and retention rates. We assume that every student who receive the grant will stay in the program and university (this assumption can be adjusted by a future study on the probabilities between grant and retention rates).

Let $\phi_j^{[i]}$ be the total number of students in university j for class i.
The adjusted rate of existing poor students based on our assumptions is:

$$OP_j^{[i]*} = \frac{OP_j^{[i]} \phi_j^{[i]} - n_j^{[i]}}{\phi_j^{[i]}} \tag{3}$$

We can then provide a rank to this university, in the similar fashion we computed the rank for the benchmark universities (recall eq. (1)):

$$R_j^{[i]*} = \frac{IP_j^{[i]} - OP_j^{[i]*}}{IP_j^{[i]}}$$

We use this to compute the increase to the social well-being of students going to this university.

2.4.3 Return on Investment (ROI)

The return on investment for our model will be the change in the level of alleviation of poverty provided by our investment.
We do this as follows:

$$R_j^{[i]}(1 + r_j^{[i]}) = R_j^{[i]*} \tag{4}$$

where $r_j^{[i]}$ is the return on investment for a particular school.
We also note that (4) can be rewritten using (1) and (3):

$$r_j^{[i]} = \frac{OP_j^{[i]} - OP_j^{[i]*}}{IP_j^{[i]} - OP_j^{[i]}} \tag{5}$$

This will later be used for simplification purpose of the maximization problem.

We then average these rates, to get the approximated return:

$$\bar{r} = \frac{\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} r_j^{[i]}}{\Phi} \tag{6}$$

Where Φ is the total number universities.

2.4.4 Maximization of return

Now that we have a measure of return on investment, we wish to maximize this measure based on the physical restriction that are given to us.
Since we only have a limit of \$100M, and would not want to overshoot our investment by giving any particular university more than what the top-tier schools are given. We put the following restriction forth.

Where $g_j^{[i]} = \max(\frac{G_o^{[i]}}{\phi_o^{[i]}} - \frac{G_j^{[i]}}{\phi_j^{[i]}}\ ,\ 0)$, is the difference between per student investment with the benchmark school. We formulate the problem as follows:

$$\begin{aligned} &\underset{A}{\text{maximize}} && \bar{r} \\ &\text{subject to} && \sum\sum A_j^{[i]} = 100,000,000 \\ & && \forall i,j \quad 0 \leq A_j^{[i]} \leq g_j^{[i]}\phi_j^{[i]} \end{aligned} \tag{7}$$

2.4.5 Simplifying the complexity

In order to simplify the complexity of the previous problem we outlined before. We reformulated the problem as follows:
First, we rewrote (5) using (2) as,

$$\begin{aligned} r_j^{[i]} &= \frac{1}{\phi_j^{[i]}(IP_j^{[i]} - OP_j^{[i]})} n_j^{[i]} \\ r_j^{[i]} &= \Delta_j^{[i]} n_j^{[i]} \end{aligned} \tag{8}$$

We note that $\Delta_j^{[i]}$ from (4) is the information given by our data. It must be positive as

$$\Delta_j^{[i]} \implies IP_j^{[i]} < OP_j^{[i]}$$

Which means that the considered university is getting more in funding than the benchmark one. If this happens, we omit that university from our dataset.

Whereas $n_j^{[i]}$ is the number of students we invest in, which is a variable that will be controlled by us.

Hence the problem as outlined in (6) return can be written as:

$$\bar{r} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\frac{1}{\Phi}r_j^{[i]} = \frac{1}{\Phi}\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\Delta_j^{[i]}n_j^{[i]} \tag{9}$$

So our new maximization problem can be reformulated as follows:

$$\begin{aligned} \underset{\forall n_j^{[i]}}{\text{maximize}} \quad & \frac{1}{\Phi}\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\Delta_j^{[i]}n_j^{[i]} \\ \text{subject to} \quad & \sum\sum n_j^{[i]}T_j^{[i]} = 100,000,000 \\ & \forall i,j \quad 0 \leq n_j^{[i]} \leq \frac{g_j^{[i]}\phi_j^{[i]}}{T_j^{[i]}} \end{aligned}$$

2.5 Limiting Schools

The model outlined in the previous section, is given the ability to reject schools. This can be seen in the condition:

$$\forall i,j \ \ 0 \leq n_j^{[i]} \leq \frac{g_j^{[i]}\phi_j^{[i]}}{T_j^{[i]}}$$

However, we wanted to eliminate this ability, in order to provide the capacity to maximize over any number of schools. So we reworte te condition as

$$\forall i,j \ \ 1 \leq n_j^{[i]}T_j^{[i]} \leq g_j^{[i]}\phi_j^{[i]} \tag{10}$$

As a side note, this would mean that $g_j^{[i]}\phi_j^{[i]} \geq 1$. Meaning, that some schools may be omitted before the optimization process starts.

3 Model Application

3.1 Partitioning and Benchmark

As discussed in the previous section, the first goal in the analysis is to find a list of benchmark schools. However, before doing so, we partitioned the schools by the type of degrees that they predominantly provide. This was due to the fact that otherwise, many pieces of data become unusable, and the kNN algorithm becomes biased towards schools that have less-years-to-graduation.

INSTNM	GG_NO_re
York County School of Technology-Adult & Continui...	0.655913943
Greater Johnstown Career and Technology Center	0.550817740
Western Technical College	0.534803138
Lenape Technical School Practical Nursing Program	0.514687109
Eastern Suffolk BOCES-Practical Nursing Program	0.498404042
Lancaster County Career and Technology Center	0.486810085
Greenville Technical College	0.485671590
Western Technical College	0.468055285
Washburn Institute of Technology	0.427829299
Ashtabula County Technical and Career Campus	0.411111103
Nicolet Area Technical College	0.409556778
GateWay Community College	0.408813708

(a) 2 Year Program

INSTNM	GG_NO_re
Century College	0.6141397
Salt Lake Community College	0.5373160
Clackamas Community College	0.5035577
Three Rivers Community College	0.4773438
Tarrant County College District	0.4648729
Nashua Community College	0.4373494
Quincy College	0.4307892
Southeast Technical Institute	0.4258058
George C Wallace State Community College-Hanceville	0.4115761
NHTI-Concord's Community College	0.4006864
Eastern Oklahoma State College	0.3863382
Parkland College	0.3631325

(b) Bachelor Program

INSTNM	GG_NO_re
Indiana University-Purdue University-Fort Wayne	0.5790924257
Sam Houston State University	0.4605967075
Georgia Institute of Technology-Main Campus	0.4339397553
Texas Woman's University	0.4215454649
California State University-Los Angeles	0.3948813640
New Jersey Institute of Technology	0.3696700168
University of Colorado Denver	0.3696529120
University of California-Berkeley	0.3529841955
University of California-Irvine	0.3358830137
University of Mississippi	0.3288047228
University of Arkansas-Fort Smith	0.3237021395
Clayton State University	0.3120537286

6 Year Program

Once we acquired this data, we chose the top 10 schools from each school type as our benchmark schools.

3.2 Variable Selection

3.2.1 Filtering by Grant

Since we did not want to overlap with grants given by other institutions, we filtered the next list using the data given by the Gates[4] and Lumina [5] Foundations.

3.2.2 Scorecard Variables

First we created a dictionary for all the variables that were given to us by Scorecard. We had to categorize these variables as follows:

- **Irrelevant [0]**, values such as location of website, or miscellaneous ids.
- **Informative [1]**, flags for women-only, historically black colleges and others.
- **Selected [2]**, the values we chose to focus on, such as SAT, ACT, retention rates etc.
- **Percentile [3]** values that refer to the composition of the student body (such as racial, sexual or income composition).

We did make use of [1], and [3], for cleaning and producing other variables (such as the PES rank), however we did not run the kNN algorithm on them.

3.2.3 IPEDS Variables

Since many of the variables provided by scorecard, overlapped with the ones from IPEDS, we only withdrew the following variables:

- **Number of students**, which we used for computation of kNN and for the maximization algorithm.
- **Average Tuition of Low Income Students**, which we used mainly for the maximization.
- **Total Grants Received**, we summed the various grants that each university received to get this number. We then used it for both kNN and maximization.

3.3 kNN Selection

After this point, we used the kNN algorithm as the method of selection for the nearest best schools. As we wish to avoid investing in the top tier schools, and would rather invest in schools with the potential to become the top-tier.

Prior to the application of the kNN algorithm we normalized the data by transforming each variable to a scale in $[0, 1]$, in order to avoid bias for larger scaled-values.

This identification provided us with the following schools (please note that this is not the entire dataset, the full dataset will be provided separately). As can be seen in fig.1.

INSTNM	GG_NO_re
Indiana University-Purdue University-Fort Wayne	0.5790924257
Sam Houston State University	0.4605967075
Georgia Institute of Technology-Main Campus	0.4339397553
Texas Woman's University	0.4215454649
California State University-Los Angeles	0.3948813640
New Jersey Institute of Technology	0.3696700168
University of Colorado Denver	0.3696529120
University of California-Berkeley	0.3529841955
University of California-Irvine	0.3358830137
University of Mississippi	0.3288047228
University of Arkansas-Fort Smith	0.3237021395
Clayton State University	0.3120537286

Figure 1: Sample of kNN results

As this point, we used the kNN algorithm as the method of selection for the nearest best schools. As we wish to avoid investing in the top tier schools, and would rather invest in schools with the potential to become the top-tier.
Prior to the application of the kNN algorithm we had to normalize the data by transforming each variable to a scale in $[0, 1]$, as otherwise variables measure in larger values would be favored to smaller variables.

This identification provided us with the following schools (please note that this is not the entire dataset, the full dataset will be provided separately) 1.

4 Discussion

4.1 Sensitivity Analysis

In order to test model stability, we impose a radius condition to the kNN selection algorithm. The model testing outlined in Section 3 used an unrestricted subset of the data. We compute the the mean and standard deviation for each class i. Then each value exceeding k from k = 1,..3 standard deviations away was omitted from the data passed to the investment algorithm. The goal is to partition the dataset such that the subset data will have 68%, 95%, and 99.7% of the data.

With all other variables the same the ROI for each standard deviation(STD) is:

1. STD 1 ROI: 0.0353
2. STD 2 ROI: 0.0499
3. STD 3 ROI: 0.0474

All ROI are relatively the same suggesting that the model does well with any size of data input.

4.2 Strengths

4.3 Weaknesses

5 Conclusions

6 Letter to Mr. Alpha Chiang

Dear Mr. Alpha Chiang,

References

[1] Trombitas, Kate. *Financial Stress: An Everyday Reality for College Students*. Inceptia, July 2012. PDF.

[2] Carnevale, Anthony, Ban Cheah, and Martin Van Der Werf. *Ranking Your College*. Georgetown University Center on Education and the Workforce, Dec. 2015. PDF.

[3] Conkey,Christopher 2006. "Big donations to charity often include spending advice *The Wall Street Journal Asia* 05:

[4] "Bill & Melinda Gates Foundation." *Bill & Melinda Gates Foundation*. Web. 29 Jan. 2016.

[5] "Grants Database." *Grants Database*. Web. 29 Jan. 2016. <https://www.luminafoundation.org/grants-database/strategy/student-financial-supports>.

[6] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. "Prototype Methods and Nearest-Neighbors". *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. New York: Springer, 2009. Print.

[7] "IPEDS Data Center." *IPEDS Data Center.* Web. 30 Jan. 2016. <https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx>.

[8] Hanushek, Eric and Margaret Raymond. 2005. "Does School Accountability Lead to Improved School Performance? *Journal of Policy Analysis and Management* 24(2): 297-329.

[9] Winston, Wayne L. "Linear Programming." *Operations Research: Applications and Algorithms.* 4th ed. Belmont, Calif.: Duxbury, 2003. Print.

[10] "Using Federal Data To Measure And Improve The Performance Of U.S. Institutions of Higher Education" Sept 2015. <https://collegescorecard.ed.gov/assets/UsingFederalDataToMeasureAnd-ImprovePerformance.pdf>.