

A Machine Learning Approach to Fundraising Success in Higher Education

by

Liang Ye

B.Sc., Xi'An University of Posts and Telecommunications, 2014

M.Sc., University of Victoria, 2017

A Thesis Submitted in Partial Fulfillment of the  
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

© Liang Ye, 2017

University of Victoria

All rights reserved. This Thesis may not be reproduced in whole or in part, by photocopying or other means, without the permission of the author.

A Machine Learning Approach to Fundraising Success in Higher Education

by

Liang Ye

B.Sc., Xi'An University of Posts and Telecommunications, 2014

M.Sc., University of Victoria, 2017

Supervisory Committee

---

Dr. Kui Wu, Supervisor  
(Department of Computer Science)

---

Dr. Alex Thomo, Departmental Member  
(Department of Computer Science)

## Supervisory Committee

---

Dr. Kui Wu, Supervisor  
(Department of Computer Science)

---

Dr. Alex Thomo, Departmental Member  
(Department of Computer Science)

### ABSTRACT

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. This thesis presents machine learning solutions as targeting strategies for the both programs based on readily available alumni data in almost any institution. The targeting strategy for new donor acquisition is modeled as a donor identification problem. The Gaussian naïve bayes, random forest, and support vector machine algorithms are used and evaluated. The test results show that having been trained with enough samples, all three algorithms can distinguish donors from rejectors well, and big donors are identified more often than others. While there is a trade off between the cost of soliciting candidates and the success of donor acquisition, the results show that in a practical scenario where the models are properly used as the targeting strategy, more than 85% of new donors and more than 90% of new big donors can be acquired when only 40% of the candidates are solicited. The targeting strategy for donor promotion is modeled as a promising donor (i.e., those who will upgrade their pledge) prediction problem in machine learning. The Gaussian naïve bayes, random forest, and support vector machine algorithms are tested. The test results show that all the three algorithms can distinguish promising donors from non-promising donors (i.e., those who will not upgrade their pledge). When the age information is known, the best model produces an overall accuracy of 97% in the test set. The results show that in a practical scenario where the models are properly used as the targeting strategy, more than 85% of promising donors can be acquired when only 26% candidates are solicited.

Keywords: machine learning, fundraising, support vector machine, random forest, naïve bayes, predictive analysis.

# Contents

<b>Supervisory Committee</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Table of Contents</b>	<b>v</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>ix</b>
<b>Acknowledgements</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.1.1 Fundraising in Higher Education . . . . .	1
1.1.2 Related Work . . . . .	2
1.1.3 Machine Learning Techniques . . . . .	3
1.2 Contributions of the Thesis . . . . .	3
<b>2 Data Exploration and Analysis</b>	<b>6</b>
2.1 Donor and Fund Composition . . . . .	6
2.2 Personal Attributes . . . . .	10
2.2.1 Age . . . . .	10
2.2.2 Wealth . . . . .	12
2.2.3 Gender . . . . .	13
2.2.4 Marital Status . . . . .	14
2.2.5 Other Features . . . . .	14
2.3 Affiliation Attributes . . . . .	16
2.3.1 Number of Degrees Achieved . . . . .	16
2.3.2 Number of Events Registered . . . . .	16

2.3.3	Family Relation . . . . .	17
2.3.4	Other Affiliation Features . . . . .	18
<b>3</b>	<b>Prospective Donor Prediction</b>	<b>20</b>
3.1	Problem Description . . . . .	20
3.2	Model Description . . . . .	21
3.2.1	Naïve Bayes Classifier . . . . .	21
3.2.2	Random Forest Classifier . . . . .	22
3.2.3	Support Vector Machine . . . . .	23
3.3	Data Preparation . . . . .	25
3.3.1	Feature Selection . . . . .	25
3.3.2	Categorical Variables Transformation . . . . .	26
3.3.3	Data Normalization . . . . .	26
3.3.4	Correlated Continuous Numeric Variables . . . . .	27
3.4	Model Validation . . . . .	28
3.4.1	Evaluation Metrics . . . . .	28
3.4.2	Performance For Targeting New Donors . . . . .	30
3.4.3	Performance For Targeting New Big Donors . . . . .	31
3.4.4	ROC Curves and Learning Curves . . . . .	33
3.5	Performance in a Practical Scenario . . . . .	37
3.6	Summary . . . . .	40
<b>4</b>	<b>Promising Donor Prediction</b>	<b>41</b>
4.1	Problem Description . . . . .	41
4.2	Analysis and Modeling . . . . .	41
4.3	Data Preparation . . . . .	44
4.3.1	Data Selection . . . . .	44
4.3.2	Feature Selection . . . . .	45
4.3.3	Categorical Features Transformation and Data Normalization . . . . .	45
4.3.4	Correlated Numerical Feature Transformation . . . . .	45
4.4	Model Validation . . . . .	45
4.4.1	Evaluation Metrics . . . . .	45
4.4.2	Performance For Targeting Promising Donors . . . . .	45
4.4.3	ROC Curves and Learning Curves . . . . .	48
4.5	Performance in a Practical Scenario . . . . .	52

4.6	Summary . . . . .	54
<b>5</b>	<b>Conclusions and Future Work</b>	<b>56</b>
5.1	Conclusions . . . . .	56
5.2	Limitations and Further Works . . . . .	57
5.2.1	Limitations . . . . .	57
5.2.2	Future Work . . . . .	57
	<b>Bibliography</b>	<b>58</b>

# List of Tables

Table 2.1 Donor Composition by Donor Category . . . . .	7
Table 2.2 Fund Contribution by Donor Category . . . . .	7
Table 3.1 Transformed Categorical Variables . . . . .	26
Table 3.2 The performance of the three algorithms for distinguishing donors from rejectors when all donors are used for evaluation, with Maximum Precision, Maximum Accuracy and Maximum F-score, respectively . . . . .	31
Table 3.3 The performance of the three algorithms for distinguishing donors from rejectors when only big donors are used for evaluation, with Maximum Precision, Maximum Accuracy and Maximum F-score, respectively . . . . .	33
Table 3.4 Test result for all 3 algorithms with default predicting thresholds	38
Table 4.1 Performance of algorithms with thresholds for Maximum Precision, Maximum Accuracy and Maximum F-score, respectively, when age information is included . . . . .	47
Table 4.2 Performance of algorithms with thresholds for Maximum Precision, Maximum Accuracy and Maximum F-score, respectively, when age information is not included . . . . .	48
Table 4.3 Prediction of promising donors from current small donors . . . . .	52

# List of Figures

Figure 2.1 Donor composition by category in 3 different times. The vertical axis represents the percentage in the total donor population. . .	7
Figure 2.2 Donation composition by category in 3 different times. The chart shows the contribution proportion to the total fund raised during each 5 years by each category. . . . .	8
Figure 2.3 The number of donors and amount raised vs. the threshold on minimum total donation from 2011 to 2015. The threshold on the minimum personal total donation increases exponentially. The left vertical axis represents the number of donors above threshold. The right vertical axis represents the total fund raised by the donors below the threshold. . . . .	9
Figure 2.4 Donation behaviour in different ages. The horizontal axis represents the ages when their first donations were made. The vertical axis represents the ages when their largest donations were made. The area of each dot represents the total amount in that case, and the color indicates the average largest amount: red means big, green means small. . . . .	10
Figure 2.5 The distribution of donors' ages when they gave their first and largest gifts. The horizontal axis represents the age when they make donation, and the vertical axis represents the number of donors. . . . .	11
Figure 2.6 Average first and largest donation amount by different ages. The trend line is drawn with 6th order polynomial fitting. . . . .	11
Figure 2.7 Estimated average household annual income versus average largest donation amount . . . . .	12
Figure 2.8 Estimated average contribution to charities versus average largest donation amount . . . . .	13
Figure 2.9 Average total and largest donation amount by gender . . . . .	13

Figure 2.10	Average total and largest gift amount by marital status. . . . .	14
Figure 2.11	Average total and largest donation amount by the number of degrees received from UVic . . . . .	16
Figure 2.12	Average donation amount and largest donation amount by the number of events registered . . . . .	17
Figure 2.13	Donation amounts by the number of alumni in family . . . . .	18
Figure 2.14	Donation amounts when spouse is or is not an alumnus . . . . .	18
Figure 3.1	Illustration of how a Gaussian Naive Bayes classifier works(by Lee, Yune-Sang[26] 2013) . . . . .	21
Figure 3.2	Illustration of random forest classifier . . . . .	23
Figure 3.3	Illustration of support vector machine . . . . .	24
Figure 3.4	ROC for naïve Bayes classifier when all donors are involved for evaluation . . . . .	34
Figure 3.5	ROC for naïve Bayes classifier when only big donors are involved for evaluation . . . . .	34
Figure 3.6	ROC for random forest classifier when all donors are involved for evaluation . . . . .	34
Figure 3.7	ROC for random forest classifier when only big donors are in- volved for evaluation . . . . .	34
Figure 3.8	ROC for support vector classifier when all donors are involved for evaluation . . . . .	35
Figure 3.9	ROC for support vector machine when only big donors are in- volved for evaluation . . . . .	35
Figure 3.10	Learning curves of AUROC for all three classifiers when all donors are used for evaluation . . . . .	36
Figure 3.11	Learning curves of AUROC for all three classifiers when only big donors are used for evaluation . . . . .	36
Figure 3.12	Learning curves of accuracy for all three classifiers when all donors are used for evaluation . . . . .	36
Figure 3.13	Learning curves of accuracy for all three classifiers when only big donors are used for evaluation . . . . .	36
Figure 3.14	Learning curves of recall rate for all three classifiers when all donors are used for evaluation . . . . .	36

Figure 3.15	Learning curves of recall rate for all three classifiers when only big donors are used for evaluation . . . . .	36
Figure 3.16	Prediction recall curve for all donors. The horizontal axis represents the percentage of candidates labeled as potential donors. The vertical axis represents the percentage of donors acquired. . . . .	38
Figure 3.17	Prediction recall curve for big donors. The horizontal axis represents the percentage of candidates labeled as potential donors. The vertical axis represents the percentage of big donors acquired. . . . .	39
Figure 4.1	Last small donation from upgraded big donors. The horizontal axis represents the amount of the previous donation before the donor's first big donation. The vertical axis represents the proportion of upgraded big donors. . . . .	43
Figure 4.2	Year gap between the first big gift and the previous gift. The horizontal axis represents the year gap between the first big donation and the previous donation. The vertical axis represents the proportion of upgraded big donors. . . . .	43
Figure 4.3	ROC for naïve bayes classifier when age information is not included	49
Figure 4.4	ROC for naïve bayes classifier when age information is included	49
Figure 4.5	ROC for random forest classifier when age information is not included . . . . .	49
Figure 4.6	ROC for random forest classifier when age information is included	49
Figure 4.7	ROC for support vector classifier when age information is not included . . . . .	50
Figure 4.8	ROC for support vector classifier when age information is included	50
Figure 4.9	Learning curve measured in AUROC for all three algorithms when age information is included . . . . .	51
Figure 4.10	Learning curve measured in AUROC for all three algorithms when age information is not included . . . . .	51
Figure 4.11	Learning curve measured in accuracy for all three algorithms when age information is included . . . . .	51
Figure 4.12	Learning curve measured in accuracy for all three algorithms when age information is not included . . . . .	51
Figure 4.13	Learning curve measured in recall for all three algorithms when age information is included . . . . .	51

Figure 4.14 Learning curve measured in recall for all three algorithms when age information is not included . . . . .	51
Figure 4.15 Prediction recall curve for the model with age information. The horizontal axis represents the percentage of candidates labeled as a promising donor. The vertical axis represents the percentage of upgraded big donors acquired . . . . .	53
Figure 4.16 Prediction recall curve for the model without age information. The horizontal axis represents the percentage of candidates labeled as a promising donor. The vertical axis represents the percentage of upgraded big donors acquired. . . . .	54

## ACKNOWLEDGEMENTS

I would like to thank:

**my supervisor, Dr. Kui Wu**, for supervising me, encouraging me to choose research topic according to my own interest and providing me with not only knowledge and direction but also experience and support.

**my parents**, for their endless love and support, their love and encouragements are my eternal power source.

**UVic Data Analytics Officer, Atsuko Umeki**, who led me into the fundraising field and helped me applying the machine learning knowledge to fundraising.

**Gregory Churchill** and

**Stephanie Rowe** for their full support of data, equipment, and friendship. This research would not be accomplished without their support and efforts.

**Xinyi Hao** for cooking and accompanying me when I was experiencing a hard period.

**Dr. Alex Thomo and Dr. Hong-Chuan Yang** for serving in my thesis oral defense committee.

# Chapter 1

## Introduction

### 1.1 Background

#### 1.1.1 Fundraising in Higher Education

Fundraising has been playing a prominent role in the development history of higher education in North America. According to Miller [16]. Fundraising for higher education started in the twelfth century and can be traced directly to the opening of the medieval universities.

“As these institutions opened for the first time and matured, college founders were forced to take measures to secure the money and resources necessary for the college’s operation, such as living arrangements for students, book acquisitions, and faculty incentives. In order to accomplish this early fund raising, the college founders and “president” [i.e., rector, principal, master, etc.] solicited businessmen, merchants, and other college supporters for cash and in-kind contributions. The concept of the chief college faculty member being responsible for fund raising was transported to the Colonial Colleges in New England, and was common at institutions such as the College at Cambridge [sic] (later Harvard) where head faculty members solicited, in person, gifts of brick, mortar, food, books, and cash and other valuables.”

Today, fundraising provides support for more areas of higher education than ever before, and is playing a more and more important role in the development of modern higher education institutions. Using Harvard University as an example, its 2015-2016

financial report discloses that the non-federal sponsors and gifts account for 14% of the total revenue in the last year [32], which is even more than the sponsored support from the federal government. Canadian universities, in the recent years, have also achieved considerable success in fundraising. In 2016, the non-government grants, contracts and donation for University of British Columbia accounted for 6.9% of their total revenue [19]. The University of Victoria received more than 15 million from donation and non-government grants in 2016 [20]. These donations not only help the university to build new buildings, research facilities, and gyms, but also help financially-stretched students to support their education.

### 1.1.2 Related Work

The research related to boosting fundraising for higher education falls into four major categories. The first category is qualitative analysis based on case studies, as Wastyn did in 2009 [34] with 14 years of expertise in fundraising. Such research highly depends on the researchers' academic vision and personal experience. As a result, its value may be limited to a specific circumstance similar to that of the researchers.

The second category is statistical analysis based on historical data. Wesley and Christopher (1992 [36] used logit analysis in 1992 to predict the individuals who would give higher (e.g., \$100,000) or lower (\$1,000) donations based on the data from the alumni database as well as the geo-demographic information. Their result showed that 92% of the dollars could be collected with 36.5% prospects selected in the annual fund model. Later with their upgraded model (1994) [12], a slightly better performance was achieved for major gift prediction. McDearmon and Kathryne [15] conducted a research based on online survey and multiple regression analysis. They showed that an alumnus' positive experiences at the university are good predictors of their donation behaviour. Sun et al. (2007) [30] conducted a multivariate casual model using the data from a two-year alumni survey. They produced an accuracy of 81% for predicting donors. However, their model was built upon the survey responses. These statistical models have shown the potential of predicting donors based on quantitative analysis. However, although some of the models did perform well in their cases, they are custom-built statistical models, and may be difficult to apply into other cases.

The third category is alumni clustering and segmentation. Louis and Conway (2008) [10] studied 33,000 university alumni records with cluster analysis. Their research illustrated the existence of Pareto effects in alumni donors that 2.6% of

the donors contribute to 88% fund raised. Pablo and Elizabeth (2013) [3] segmented alumni based on their personal information and affiliation factors. The study suggests that alumni whose spouses or children are also alumni are more likely to belong to the segments with higher contributions. Also their segmentation results show that personal features and affiliation factors can be very good predictors for donation behaviour.

The fourth category is based on supervised machine learning models. Although few research in this type was done in the past, Heiat (2011) [5] used artificial neural network and decision tree models to predict donors and non-donors with 5 features, i.e., Years since graduation, Bachelor of Secondary Education (BSED), Degree, Major, and Gender. However, the prediction result is not good because of the improper handling of imbalanced data as well as the insufficient features involved.

### 1.1.3 Machine Learning Techniques

As one of the most important sub-fields of artificial intelligence, machine learning is a computer science subject that “gives computers the ability to learn without being explicitly programmed” (Arthur Samuel [29], 1959). Instead of executing statically programmed instructions, machine learning algorithms are able to make their own judgements based on the knowledge acquired from a data set, which is called “the training set”. Machine learning is *roughly* divided into two different classes, supervised learning and unsupervised learning. Supervised learning does the task of inferring a function from labeled training data [17]. When training data is not available or hard to obtain, unsupervised learning can be used to infer a function that describes the hidden structure from unlabeled data. Nowadays, supervised learning is widely used in pattern recognition and data mining, such as detection [9], prediction [13], and classification [21].

While there are many newly emerging machine learning methods, such as reinforcement learning, deep learning, active learning, and transfer learning, this thesis only focuses on supervised learning.

## 1.2 Contributions of the Thesis

Despite readily available alumni data warehoused at nearly every institution, the potential of alumni data has not been fully explored by fundraisers to enhance fundrais-

ing outcome. In this thesis, using machine learning methods, we 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.

In particular, we investigate and test three popular machine learning algorithms, naïve bayes classifier, support vector machine, and random forest classifier, in the context of higher education fundraising. The following questions are studied, and our answers are summarized as below.

- Question 1: Are donors predictable with their personal features and affiliation factors using machine learning algorithms?

Answer 1: Machine learning algorithms can indeed predict donors and rejectors based on their personal features and affiliation factors. Besides, compared with small donors, big donors are more recognizable.

- Question 2: What are the differences among naïve bayes classifier, support vector machine and random forest classifier when they are used to predict prospective donors?

Answer 2: Among the three machine learning algorithms, the random forest classifier performs the best if enough training samples are provided, in terms of accuracy, recall and precision. The test results show that in a practical scenario where the targeting models is properly used, more than 85% of the potential donors and more than 90% of big donors can be acquired when only 40% of the candidates are contacted.

- Question 3: Are promising donors predictable with their personal features and affiliation factors using machine learning algorithms?

Answer 3: Machine learning algorithms can predict promising donors with their personal features and affiliation factors. The results show that in a practical scenario where the targeting models are properly used, more than 85% of promising donors can be acquired with only 26% candidates contacted.

- Question 4: What are the differences among naïve bayes classifier, support vector machine and random forest classifier when they are used to predict promising donors?

Answer 4: Among the three machine learning algorithms, random forest classifier and support vector machine work better than naïve bayes classifier in distinguishing promising donors from non-promising donors. Generally, support vector machine has a better recall rate, while the random forest classifier has a slightly higher accuracy. In addition, random forest classifier and support vector machine produce nearly the same Area under Receiver Operating Characteristic (AUROC).

The rest of the thesis is organized as follows. In Chapter 2, the data set used in this research is explored and briefly analyzed. This analysis discloses the features in the given data set and how the features are related to the donation behaviour. In Chapter 3, the problem of predicting prospective donors is modeled as a donor identification task in machine learning. Three popular machine learning algorithms are tested for distinguishing donors from rejectors. We also performance a study to illustrate how well the machine learning algorithms can predict prospective donors in a practical scenario. In Chapter 4, the prediction of promising donors is modeled as a promising donor identification task in machine learning. The same three machine learning algorithms are tested for distinguishing bigger donors from smaller ones. We also performed a study to better understand the performance of the algorithms in the prediction of promising donors in a real world scenario. Finally, Chapter 5 outlines the main conclusions and discusses the limitations of the models and possible future work.

## Chapter 2

# Data Exploration and Analysis

### 2.1 Donor and Fund Composition

Knowing the composition of the donation can help fundraisers better understand their goal and importance of their work. Understanding the change of the composition over the years can also help fundraisers make better targeting strategies in the future. Since the analysis in the thesis is based on the data from University of Victoria, the use of the statistical results should be limited to the universities with similar background, because universities with different fundraising history and development stage might have different composition.

In this thesis, individual donors are classified into 6 categories according to the amount of their largest donations, i.e., Small Donors (less than \$100), Common Donors (between \$100 and \$1,000), Serious Donors (between \$1,000 and \$10,000), Major Donors (between \$10,000 and \$100,000), Large Donors (between \$100,000 and \$1,000,000), and Leadership Donors (over \$1,000,000).

Checking the donation data from the past 15 years, we noticed that the population of Small Donors and Common Donors are always the vast majority (88+% of the total), while they contribute to a small portion of the fund raising performance. From 2001 to 2005, small donors and common donors accounted for 90.29% of the total donors, while they only contributed to 2.79% of the total fund raised. From 2006 to 2010, small donors and common donors accounted for 89.03% of the population, while they only contributed to 2.38% of the total fund raised. In the latest 5 years, they have accounted for 88.27% of the population, and their donation was 12.25% of the total fund raised.

Table 2.1: Donor Composition by Donor Category

Donor Composition						
Category	Small Donors	Common Donors	Serious Donors	Major Donors	Large Donors	Leadership Donors
2001 - 2005	50.52%	39.77%	7.19%	1.95%	0.47%	0.09%
2006 - 2010	43.91%	45.12%	7.83%	2.34%	0.64%	0.18%
2011 - 2015	43.46%	44.81%	8.77%	2.24%	0.60%	0.13%

Table 2.2: Fund Contribution by Donor Category

Fund Contribution						
Category	Small Donors	Common Donors	Serious Donors	Major Donors	Large Donors	Leadership Donors
2001 - 2005	0.53%	2.26%	4.53%	8.93%	22.84%	60.92%
2006 - 2010	0.36%	2.02%	3.72%	9.43%	24.02%	60.44%
2011 - 2015	0.95%	11.30%	8.32%	15.02%	31.75%	32.66%

In contrast, although Large Donors and Leadership Donors were less than 1% in population, they contributed to more than 60% of the total fund raised. From 2001 to 2011, they contributed even more than 80%. Tables 2.1 and 2.2 show the statistics.

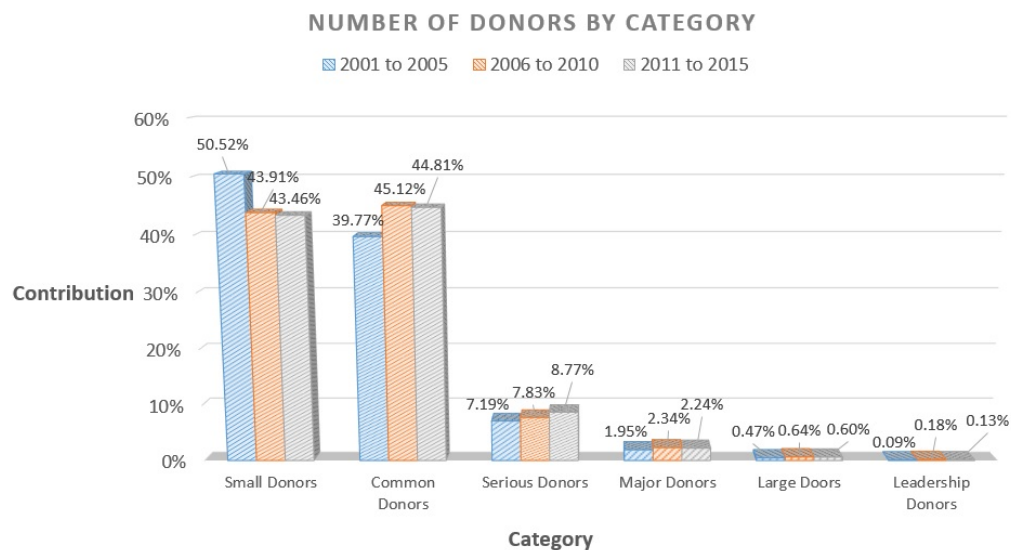


Figure 2.1: Donor composition by category in 3 different times. The vertical axis represents the percentage in the total donor population.

Figure 2.1 shows the different donor composition by category in different times, the proportion of small donors proportion has decreased by 7% over the years, while

serious donors proportion has increased by 1.6%. The common donors proportion has increased by 5%.

Fifteen years ago, leadership donors played a dominating role in fundraising success, as most fund was raised from leadership donors. Things have changed in these years. Compared with the composition 15 years ago, nowadays large donors, major donors and serious donors are becoming the major contributors in fund raising success. In the last 5 years (from 2011 to 2015), although the leadership donors were still contributing the most to the total fund raised, the leadership's composition decreased by half to 32.66%, while it was 60.92% during 2001 to 2005. And the large donors, major donors and serious donors together were contributing to 55.09% of the total fund in the latest five years, while it was 36.3% during 2001 to 2005, as shown in Figure 2.2.

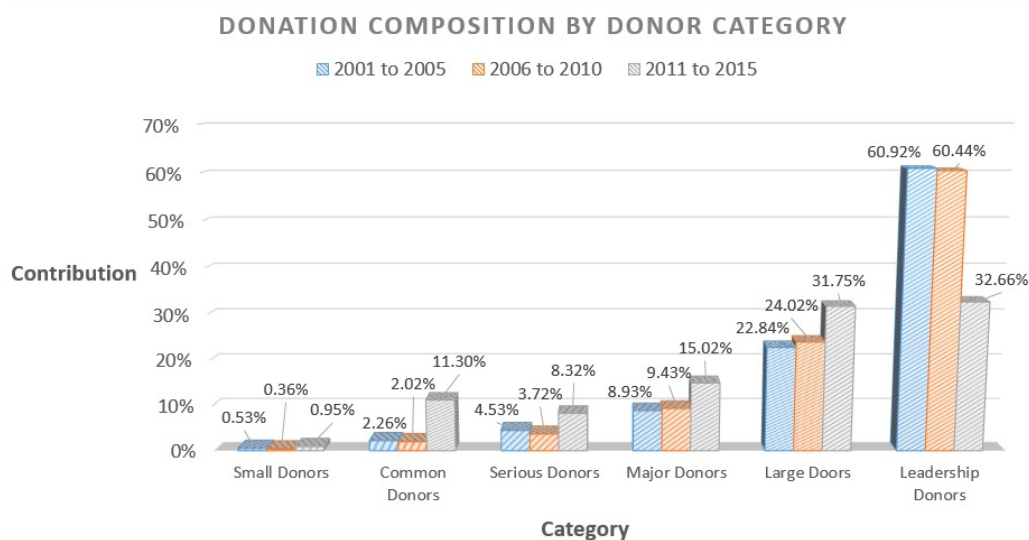


Figure 2.2: Donation composition by category in 3 different times. The chart shows the contribution proportion to the total fund raised during each 5 years by each category.

Figure 2.3 shows the donors' contribution to fundraising, as well as the change of number of donors by setting different thresholds on the minimum personal total donation. Along the x-axis, we increase the threshold on the minimum personal total donation exponentially. It can be seen that with the increase of this threshold, the number of donors declines quickly, and the total dollar amount contributed by donors with minimum personal total donation below this threshold increases (but at a smaller rate). The chart also implies that more than 90% of the fund raised comes from less

than 10% donors, and donors with more than \$1,000 total donation contribute more than 95% of the total fund raised.

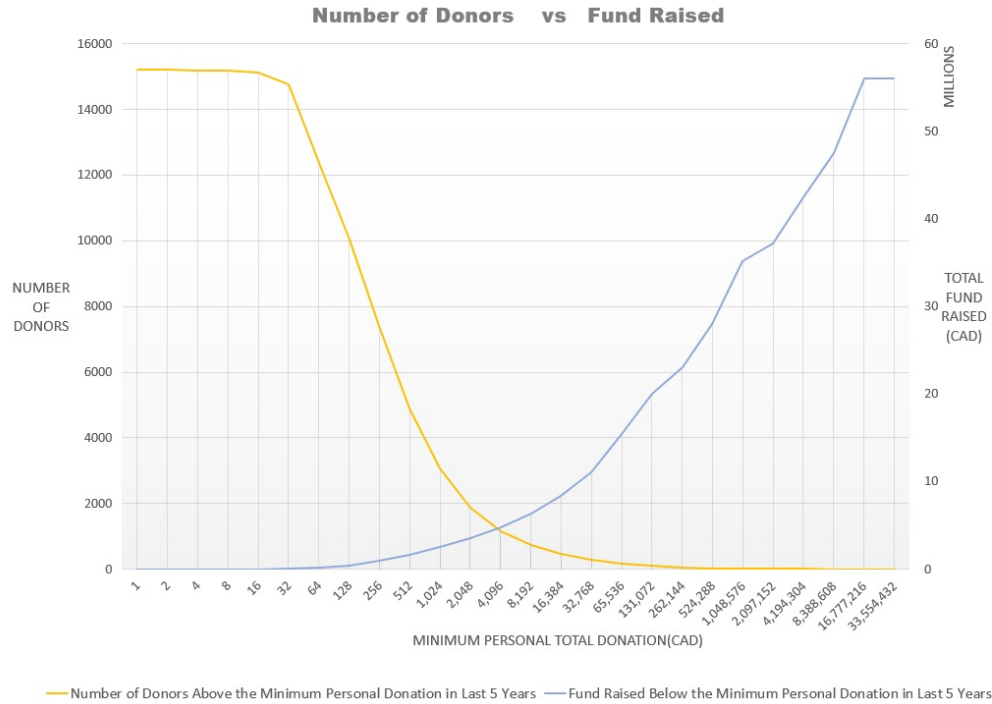


Figure 2.3: The number of donors and amount raised vs. the threshold on minimum total donation from 2011 to 2015. The threshold on the minimum personal total donation increases exponentially. The left vertical axis represents the number of donors above threshold. The right vertical axis represents the total fund raised by the donors below the threshold.

## 2.2 Personal Attributes

### 2.2.1 Age

Many experienced fundraisers believe that age can be one of the most important personal features that make a huge difference to their donation behaviour. 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 behaviour.

To illustrate impact of the age on donation behaviour, 20,000 donors are randomly selected to draw the following 3 charts. Figure 2.4 shows the largest amount and total amount from different donating ages.

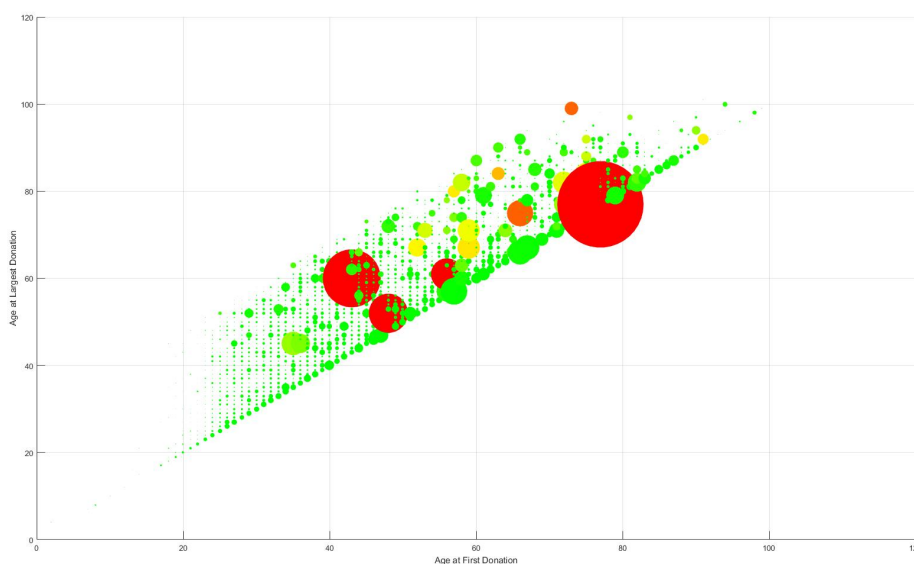


Figure 2.4: Donation behaviour in different ages. The horizontal axis represents the ages when their first donations were made. The vertical axis represents the ages when their largest donations were made. The area of each dot represents the total amount in that case, and the color indicates the average largest amount: red means big, green means small.

Figure 2.5 shows the distribution of donors' ages when they gave their first donation, as well as the distribution of donors' ages when they gave their largest gift. Figure 2.6 shows the different average amount for the first gift and the largest gift in different donating ages.

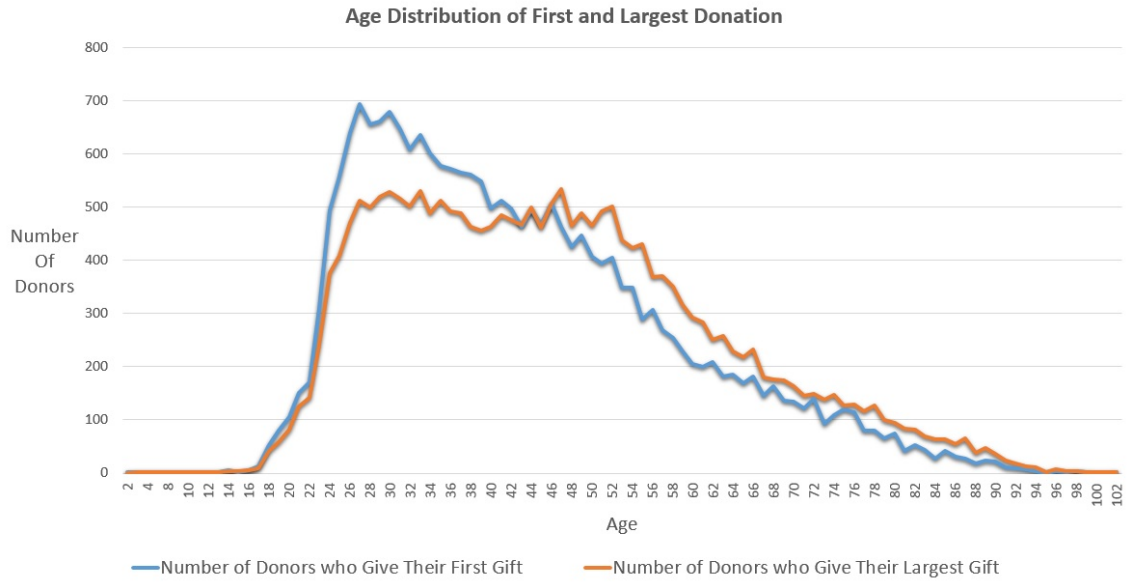


Figure 2.5: The distribution of donors' ages when they gave their first and largest gifts. The horizontal axis represents the age when they make donation, and the vertical axis represents the number of donors.

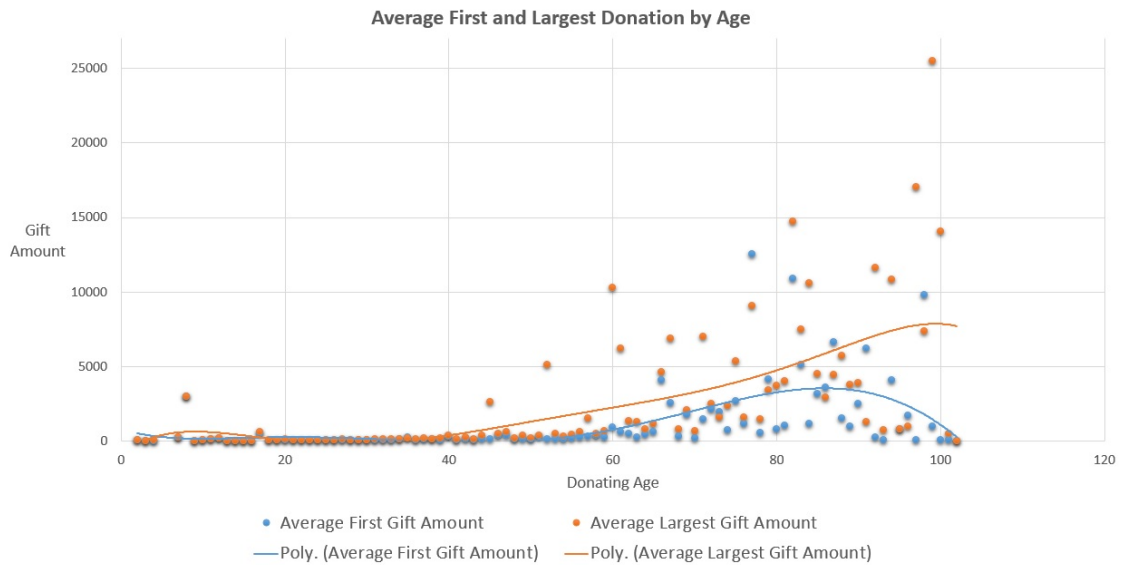


Figure 2.6: Average first and largest donation amount by different ages. The trend line is drawn with 6th order polynomial fitting.

From the above figures, we can conclude that:

- Most of the large total donations are given by the donors who gives their largest gift between 60 and 90 years old.

- Many donors start their donation when they are 20s or 30s.
- Many donors give their largest donation between 25 to 55 years old, while the large donations (in dollar amount) are donated mostly after the age of 60.
- The average first donation and average largest gift are both much higher when the donor's age is over 60.

### 2.2.2 Wealth

Wealth information is the most important information since it is believed to be highly correlated to the donation behavior [31]. Nevertheless, the annual household income for a family is private information, which may not be always available in the dataset. To estimate the wealth condition of the randomly selected donors, their annual household income is inferred, based on the location of their home (e.g., the postal codes), as well as their estimated contribution to charities. Figure 2.7 and Figure 2.8 show the estimated results regarding the donors' wealth condition in our dataset.

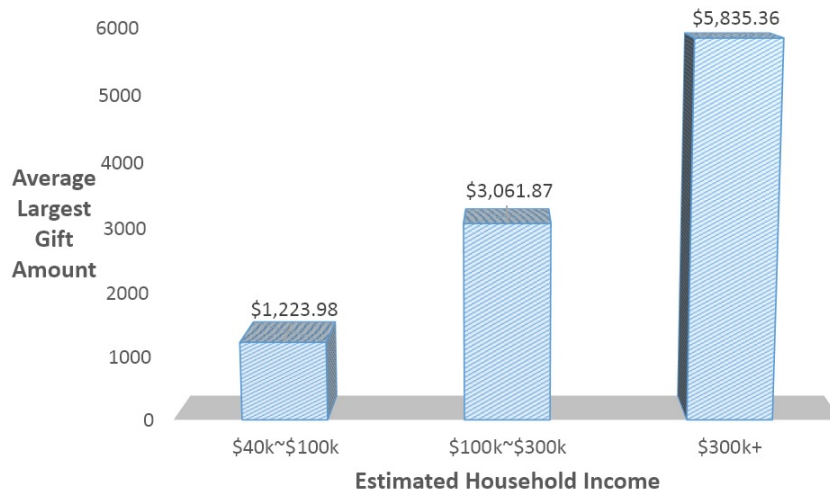


Figure 2.7: Estimated average household annual income versus average largest donation amount

From the estimation, we observe that although donors who have higher household income tend to give a smaller proportion of their total income [31], the wealthier donors donate larger amount on average as expected.

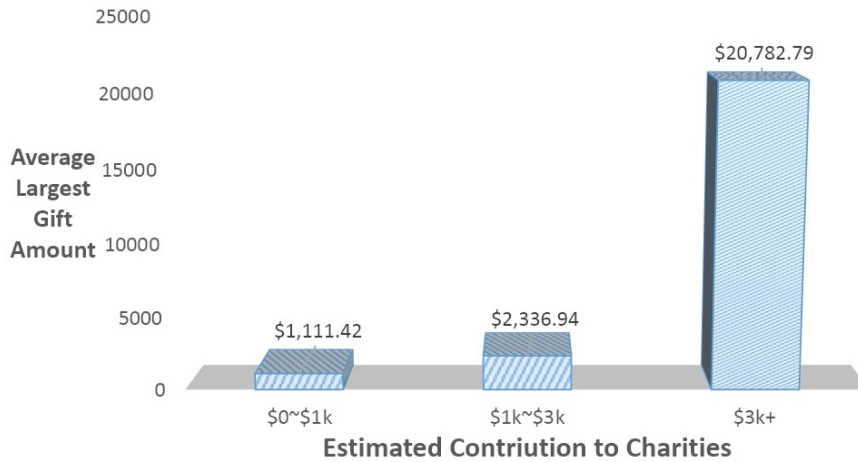


Figure 2.8: Estimated average contribution to charities versus average largest donation amount

### 2.2.3 Gender

Gender may be another important factor impacting the donating behaviour. Gender information is easy to obtain, and in our dataset, among the 35122 randomly selected donors, only 98 donors are missing gender records. According to the statistics, the male donors averagely donate 3 times more in terms of total gift amount, and 5 times more in terms of largest gift amount than female donors, as shown in Figure 2.9.

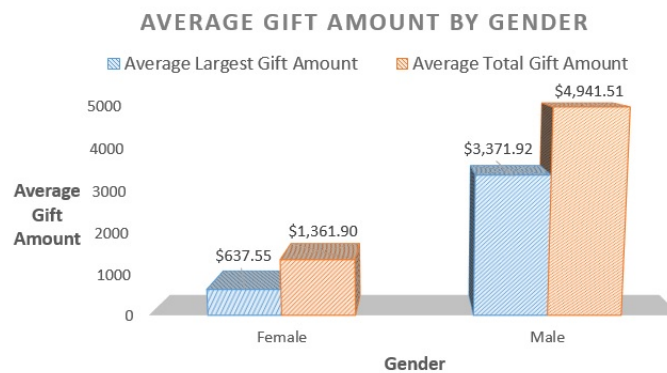


Figure 2.9: Average total and largest donation amount by gender

## 2.2.4 Marital Status

Marital status, mostly available in an alumni database, might indicate the person’s living condition, which may make a difference on their donating behaviour. From our dataset, there are 9688 married donors, 5168 single donors, 1148 widow donors, 162 divorced donors, and 29 donors declaring married but separated. The statistical result shows that the donors involved in a relationship (including the widowed) donate significantly more than single or divorced donors. The widowed donors donate the most in terms of largest donation and total donation, while the donors with a positive marital status (married) contribute the most to the whole fundraising program because of their population and high average individual contribution. Although the “separated” donors also donate high amount, their total contribution is not significant due to their scarce in number.

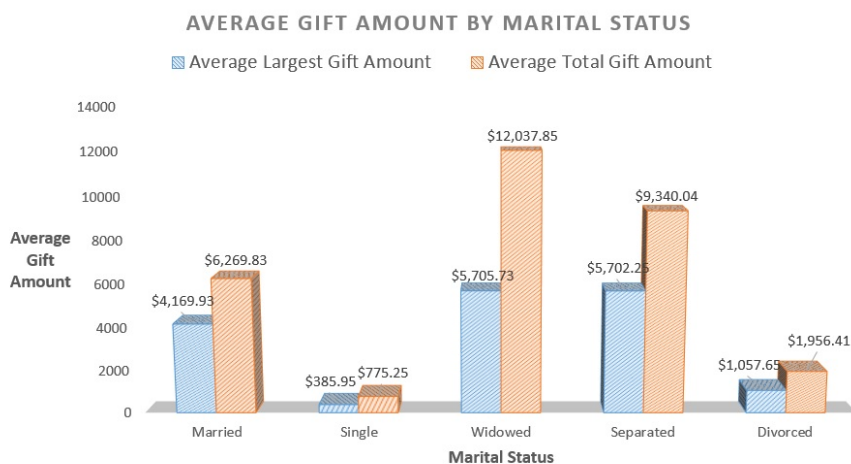


Figure 2.10: Average total and largest gift amount by marital status.

## 2.2.5 Other Features

Besides the features discussed above, our data set also includes the following features. While these features might be helpful and will be used in the following chapters, they will not be discussed in detail here. These features include:

- Job title category
- MEM membership count

- Guest speaker counts
- UVic volunteer counts
- Whether he/she lives in BC
- Whether he/she lives in big cities
- Is mail address active
- Other estimated wealth indexes
- Other estimated charity contribution indexes

## 2.3 Affiliation Attributes

### 2.3.1 Number of Degrees Achieved

Because the number of degrees achieved may indicate the number of years that the donor have 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. In our dataset, among the randomly selected 35149 donors, 15320 donors do not have a degree from UVic, 15847 donors have one UVic degree, 3437 donors have 2 UVic degrees, and 518 donors have 3 or more UVic degrees.

It is interesting to see that the alumni do not perform well as expected. In terms of the average total donation amount, the donors with one or more UVic degrees (alumni) do not give more than donors who did not have a UVic degree. Besides, the donors with more UVic degrees tend to donate smaller gifts at a time instead of one time big donation. The result is shown in Figure 2.11.

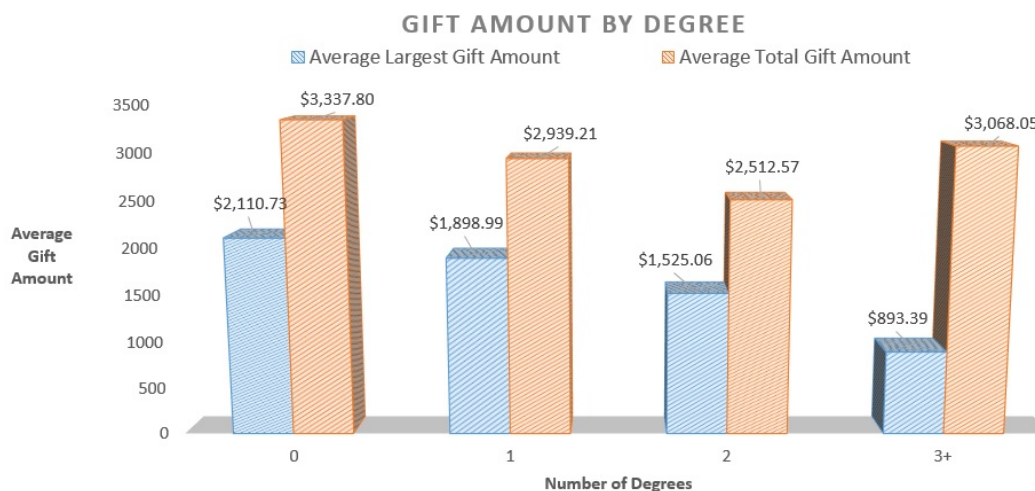


Figure 2.11: Average total and largest donation amount by the number of degrees received from UVic

### 2.3.2 Number of Events Registered

The number of events registered is another important affiliation information, because it indicates whether the person is willing to keep in touch and maintain the connection with the university.

According to the statistics and Figure 2.12, both average largest donation and average total donation increase as the number of events registered increases. This phenomenon indicates that hosting events may be a good way to boost fundraising success.

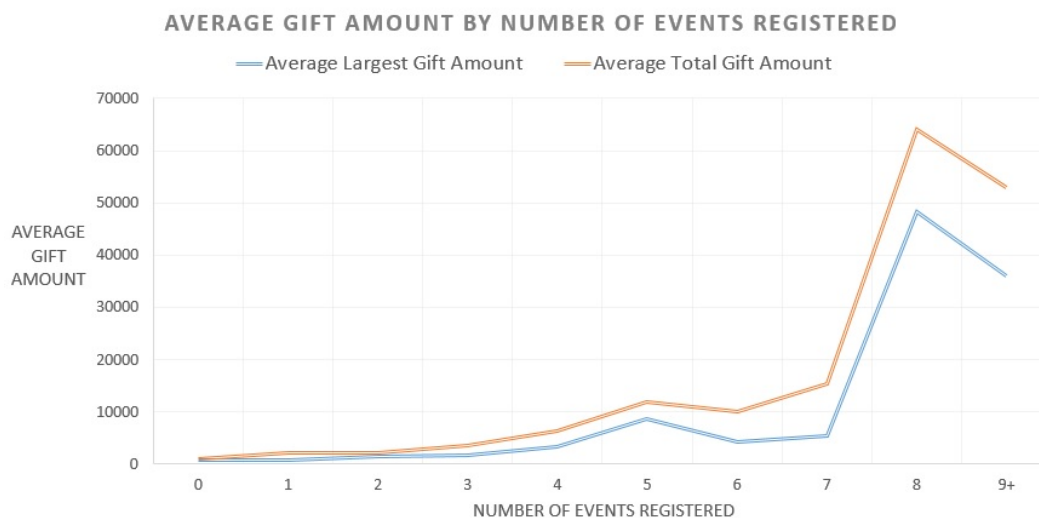


Figure 2.12: Average donation amount and largest donation amount by the number of events registered

### 2.3.3 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. Among the 35122 randomly selected donors that have family member information, 3287 donors have 1 alumnus in their family, 229 donors have 2 alumni in their family, and 138 donors have 3 or more alumni in their family. Through Figures 2.13 and 2.14, we observe that the family relation has a great impact on people's donation behaviour in terms of total and largest donation amounts. Compared with donors who do not have UVic alumnus in their family, donors who have 1 alumnus in the family donate 10 times more on average, donors who have 2 alumni family members donate 20 times more on average, and donors who have 3 or more alumni in their family donate 40 times more on average. Besides, donors whose spouses are alumni donate 6 times more than others.

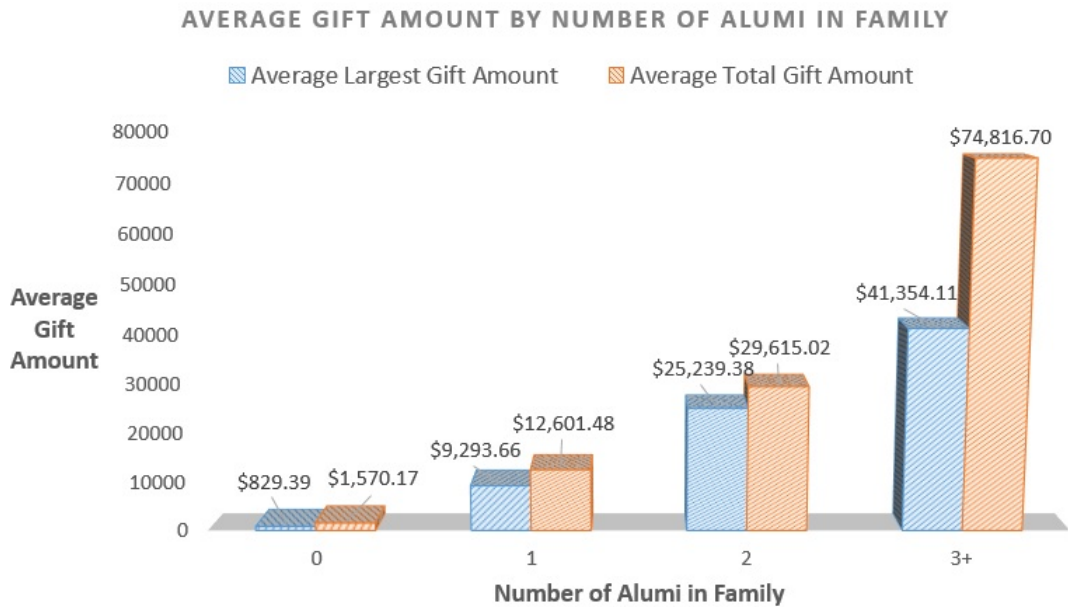


Figure 2.13: Donation amounts by the number of alumni in family

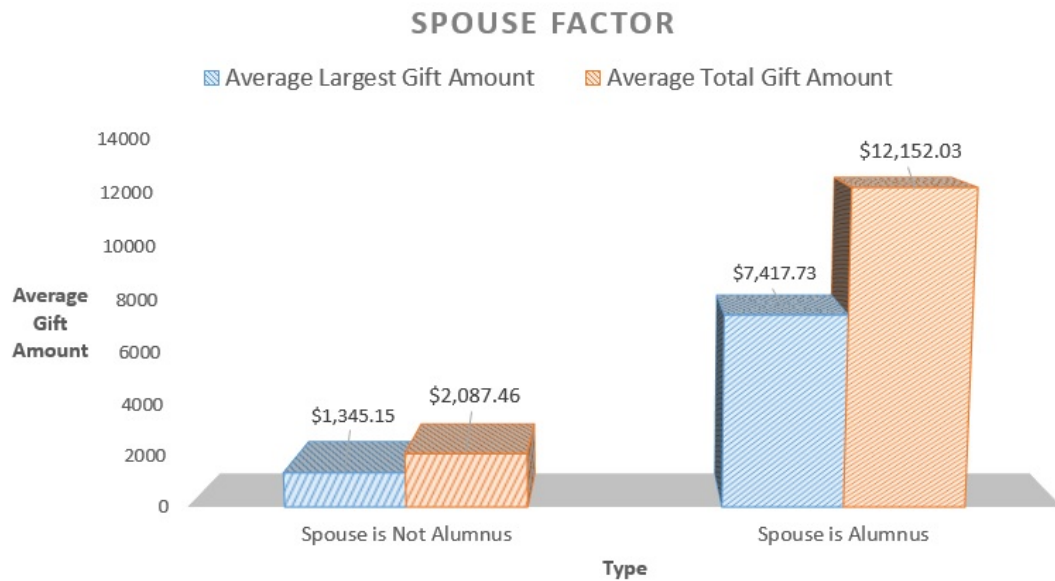


Figure 2.14: Donation amounts when spouse is or is not an alumnus

### 2.3.4 Other Affiliation Features

Besides the affiliation features discussed above, our data set also includes the following features. While these features might be helpful and will be used in the following

chapters, they will not be discussed in detail here. These features include:

- Education record by each department
- Relationship with board member
- Relationship with organizations
- Indirect relations
- Relationship with Vikes team
- Records for specific events

## Chapter 3

# Prospective Donor Prediction

### 3.1 Problem Description

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 ten-fold cross validation. The results are the average values from the cross validation.

## 3.2 Model Description

### 3.2.1 Naïve Bayes Classifier

Naïve Bayes Classifier is a classic and well known classifier based on Bayes Theory. The Naïve Bayes classifier uses the following formula[22]:

$$P(y|x_1, x_2, \dots, x_m) = \frac{P(y) \prod_{i=1}^m P(x_i|y)}{P(x_1, x_2, \dots, x_m)} \quad (3.1)$$

where  $m$  is the number of features,  $y$  is the label of classes. The probability of a class for a given instance will be estimated by the product of the prior probability  $y$  and the probability of each feature value  $x_i$ , given  $y$ . From Equation (3.1), the classifier is based on the assumption that features are independent from each other given the class. The accuracy of Naïve Bayes Classifier is sensitive to the predictive power of each feature.

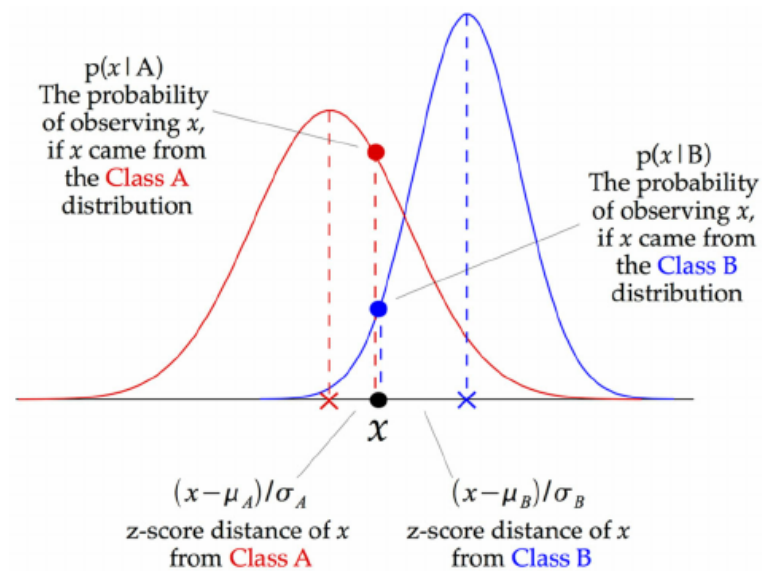


Figure 3.1: Illustration of how a Gaussian Naive Bayes classifier works (by Lee, Yune-Sang[26] 2013)

As one of the most popular machine learning classifier in the world, the naïve Bayes classifier is popular for its simplicity and fast speed in training modeling and making

prediction. As a probabilistic classifier, one of its biggest advantages is the Reject Option [18], which means that we can tell how uncertain we are about the prediction, and there is always an option of not classifying a subject and pass it to human experts for manual decision, when the uncertainty does not meet our requirements. This feature gives us flexibility for tasks at hand and is important for the task of fundraising.

Besides, since the data imbalance problem is very common in fundraising (the donors are much less than the rejectors), the chosen algorithm should be able to handle the data imbalance problem. As prior probabilities are used in naïve Bayes classifier to calculate the probabilistic prediction, the classifier takes the original proportion of each class into account. As a result, the data imbalance problem will be naturally handled as long as the composition of the training data represents the real-world distribution.

It is worthy noting that naïve Bayes classifiers are not able to learn interactions between features, as naïve Bayes classifiers are based on the assumption of independent features. This is the main reason that we need to evaluate different classifiers.

### 3.2.2 Random Forest Classifier

Random Forests Classifier (RFC) [7] is a way of blending the results from many classification trees (decision tree models). A forest is a group of decision trees and Random Forests are simply a collection of different decision tree models [28]. Every tree in the RFC is built from samples drawn with replacement from the original training set, and when splitting a node during the construction of each tree, the choice of every split feature will be the best among a random subset of the features instead of the best among all features. An simple illustration of RFC is in Figure 3.2.

Unlike the naïve Bayes classifier, the decision tree models are good at handling feature interactions (e.g., someone loves banana and yogurt but the person hates the mixture of banana and yogurt). In addition, while the outliers in the training set may create a biased model in the naïve Bayes classifier, they will not cause a big problem in the decision tree models. Unlike the support vector machine which will be introduced later, random forest classifier is resilient to whether the data is linearly separable or not.

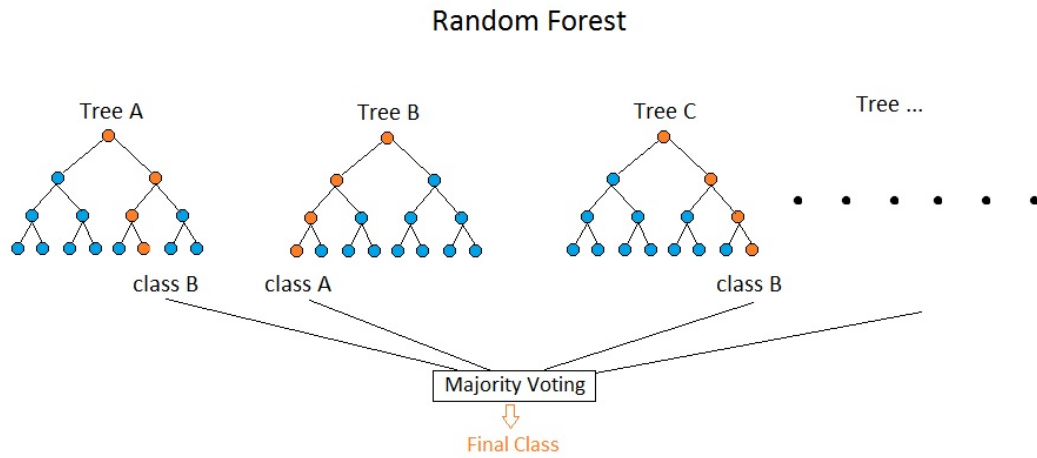


Figure 3.2: Illustration of random forest classifier

As an derived and improved model from decision tree classifier, the random forest classifier overcomes the overfitting problem in the traditional decision tree classifiers [7]. There are two parameters in standard random forest classifier, number of estimator and the maximum number features. The number of estimator (NE) sets the number of decision trees, which will be built when the RFC model is being trained; the maximum number features (MF) limits the number of features used for building each decision tree. To reach an optimal performance, exhaustive grid search for parameters will be performed for RFC with the step size of 1 for each parameter. The search range for each parameter is from 1 to 100.

### 3.2.3 Support Vector Machine

A Support Vector Machine (SVM) [1] is a supervised machine learning classifier formally defined by a separating hyperplane. Given labeled training data, SVM outputs an optimal hyperplane which categorizes new examples. The operation of the SVM algorithm is based on finding the hyperplane that gives the largest minimum distance to the training examples. Therefore, the optimal separating hyperplane maximizes the margin of the training data. An example is illustrated in Figure 3.3. Since the problems in the real world are often not linearly separable, the parameters for handling non-separable data are introduced [6] as follows:

$$\min \left( \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \right) \quad (3.2)$$

subject to:

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad (3.3)$$

where:

$$\xi_i \geq 0, i = 1, 2, \dots, N \quad (3.4)$$

where  $C$  is the trade off parameter between error and margin,  $w$  is the coefficients for the support vectors,  $b$  is a constant, and  $\xi$  is the penalty factor dealing with non-separable training samples. The index  $i$  labels the  $N$  training samples;  $y \in \pm 1$  represents the class labels;  $x_i$  represents the independent variables.

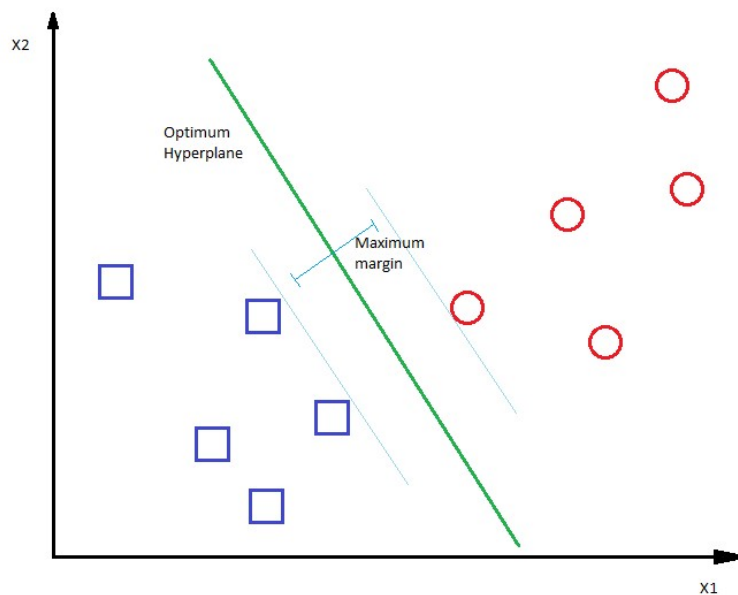


Figure 3.3: Illustration of support vector machine

SVM usually has a high accuracy, and it has nice theoretical guarantees to avoid overfitting. In the SVM with linear kernel,  $C$  is the penalty parameter which makes huge difference on the learning performance. In this thesis, the exhaustive search of parameter for SVM will be performed in the range  $[0.1, 20]$  with a variable step size from 0.01 to 0.1 to find an optimal parameter which gives the largest area under the receiver operating characteristic curve (AUROC)[4] of cross validation.

## 3.3 Data Preparation

### 3.3.1 Feature Selection

To enhance classifier performance and reduce the computational complexity, we need to exclude redundant and irrelevant features. For instance, Naïve Bayes Classifier suffers from correlated attributes [27]; Weston [35] demonstrated that the performance of standard SVM suffers from the presence of irrelevant features. In the context of the prospective donor identification, the feature selection requires extra care to handle two special problems.

First, we must exclude the features that only the donors possess. Although these features have great power to distinguish donors from non-donors in the training set, the trained model will be biased by these features and become useless when predicting new donors. For example, the number of donor appreciation events that a person has registered is larger than zero only when the person is a donor already. In other words, these features record the direct consequence of becoming a donor. If the model is trained with these feature, these features will dominate due to its high identification power, and the trained model will highly rely on these features. Nevertheless, when the trained model is used to predict new donors from candidates who never donated before, the model is unlikely to select any candidates because they are not donors. We call this type of features *consequence feature*, meaning their existence is the consequence of being a donor.

Second, the features whose values are missing for most records need to be excluded. This type of features are called *sparse features*. The trained model will not work reliably if sparse features are included, because the model lacks enough information on the sparse features to make a right decision. Intuitively, a model built using attributes that only few people have will make the model useless for most people who are missing that information.

In the context of higher education fundraising, the features can be divided into two big categories. The first category is personal features, such as age, gender and income information. The second category is affiliation features, such as alumnus relation, number of volunteered university events or the membership of some university club. Since it is unknown beforehand whether or not a feature will impact the donor's behavior, we evaluate and test all these features after excluding consequence features and sparse features.

### 3.3.2 Categorical Variables Transformation

Categorical features need to be transformed before being used. A common transformation of categorical variables creates several binary indicator variables for a feature but one for each category. Because some special variables, such as Job Title, have far too many categories to create binary indicator variables, the original categories of these variables need to be clustered and redefined as several big categories. For example, software engineer, hardware engineer, and firmware engineer could be merged and redefined as engineer.

In our data set, the transformed categorical data is listed in Table 3.1.

Table 3.1: Transformed Categorical Variables

Categorical Feature Transformation		
Feature Name	Original Variable	Transformed Variables
Marital Status	$ms \in \{1, 2, 3, 4, 5\}$	$ms_i \in \{0, 1\}; \quad i \in \{1, 2, 3, 4, 5\}$
Job Title Category	$tc \in \{1, 2, 3, 4, 5, 6, 7, 8\}$	$tc_i \in \{0, 1\}; \quad i \in \{1, 2, 3, 4, 5, 6, 7, 8\}$

### 3.3.3 Data Normalization

Normalization is another important data preprocessing procedure in machine learning. It places the values of numerical attribute with different range on a same scale to prevent the attributes with a larger range from having a larger weight inadvertently. In this thesis, as there are binary attributes, the normalization needs to bring the numerical attributes to the same scale (0,1) as the binary attributes. For some integer features including but not limited to Total Events Registered, the number of Family Members, the actual difference between 0 and 1 can be more significant than others. To better interpret the difference, an equation is introduced to normalize those features:

$$f(x) = 1 - e^{-cx} \quad c \in (0.2, 1], \quad (3.5)$$

where  $c$  is an constant. For each attribute using this equation, a different  $c$  will be determined according to the characteristics of the attribute. For the integer attributes which have largest value greater than 10,  $c$  will be set to 0.2, otherwise,  $c$  will be set to 1. As a result, higher weights are given to the differences between smaller

values. The purpose of using this equation is to emphasize the affiliation difference between smaller values for some features, e.g., for the Events Registered attribute, the affiliation difference between 0 and 1 should be much greater than the difference between 10 and 11.

### 3.3.4 Correlated Continuous Numeric Variables

The performance of Naïve Bayes Classifier suffers from correlated features. If two or more attributes are highly correlated, they will have higher weight in the final decision as to which class an instance belongs to. Unfortunately, some of the numeric variables in our data set are highly correlated with each other, such as Average Household Discretionary, Average Household Disposable and Average Household Income.

Principal component analysis (PCA) is a mathematical process that reduces the dimensionality of the data while retaining most of the variation in the data set [8]. PCA uses an orthogonal transformation to convert a set of observations of correlated variables into a set of values of linearly uncorrelated variables. And these new synthetic variables are called principal components.

There are 11 estimated financial features in the experiment data set:

- AHDC: Average Household Discretionary
- AHDP: Average Household Disposable
- AHI: Average Household Income
- MHI: Median Household Income
- AWA: Average WealthScapes Assets
- AWD: Average WealthScapes Debt
- AWLA: Average WealthScapes Liquid Assets
- ARSV: Average Real Estate Value
- CCH: Contributions to Charity per Household
- CRH: Contributions to Religious Charity per Household
- CNH: Contributions to Non-Religious Charity per Household

The normalized features show that the above features have strong correlations. After applying PCA to these variables, the first 3 principal components are selected and normalized with 96.69% of the original variation. And their coefficients of the linear combination of the original variables are calculated by using Matlab's PCA tool[14]:

- Principal component 1:  $WC1 = 0.26 \times AHDC + 0.29 \times AHDP + 0.24 \times AHI + 0.31 \times MHI + 0.46 \times AWA + 0.25 \times ASLA + 0.43 \times ARSV + 0.28 \times CCH + 0.28 \times CRH + 0.28 \times CNH$
- Principal component 2:  $WC2 = -0.28 \times AHDC - 0.20 \times AHDP - 0.19 \times AHI - 0.01 \times MHI + 0.71 \times AWA - 0.14 \times ASLA - 0.28 \times ARSV - 0.28 \times CCH - 0.28 \times CRH - 0.28 \times CNH$
- Principal component 3:  $WC3 = -0.35 \times AHDC - 0.34 \times AHDP - 0.28 \times AHI + 0.38 \times MHI - 0.40 \times AWA + 0.39 \times ASLA + 0.49 \times ARSV + 0.02 \times CCH + 0 \times CRH - 0.03 \times CNH$

## 3.4 Model Validation

### 3.4.1 Evaluation Metrics

To measure the performance of the algorithms, the ROC (Receiver Operating Characteristics) curve [4], which is plotted with the predicted probability for both classes, is introduced to evaluate the performance of machine learning algorithms [24, 25]. We also evaluate the algorithms' performance by using the area under the curve of ROC (AUROC), which was proposed in [2] to compare popular machine learning algorithms.

In addition to ROC curves and AUROC, F-score[23], Accuracy, Precision, and Recall Rate [33] are also used to measure algorithms' performance in different views. They are defined as:

$$Precision = \frac{TP}{TP + FP} \quad (3.6)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.7)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.8)$$

$$F_{Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3.9)$$

Where TP stands for true positive rate, FP stands for false positive rate; TN stands for true negative rate; FN stands for false negative rate.

During the validation process, the maximum precision, maximum accuracy and maximum f-score achieved by each algorithm are calculated and compared. It is worth noting that there is no algorithm that is absolutely better than the other, so the performance should be evaluated with different metrics to facilitate decision making. Under different circumstances, fundraisers may want to choose different prediction thresholds to get best accuracy, or best precision or best f-score (f-score is a balanced measure for both precision and recall rate). For instance, when the models are used in a small institution with very limited fundraiser resources or fundraising budget, the predicting precision will be the most important measure, and a higher precision model with less recall rate may be their best choice. On the other hand, a large institutions with more resources and does not want to lose any of the prospects, the model with maximum f-score could be a better choice as high f-score gives both great recall rate and decent accuracy.

Therefore, it is important to know the potential of algorithm in different measurements (accuracy, precision and f-score). In our experiment, the optimal predicting thresholds for maximizing each measurement are estimated for each algorithm, as well as their corresponding performance. The predicting thresholds are probabilities in probabilistic classifiers (naïve Bayes gaussian and random forest classifier). For support vector machine (a non-probabilistic binary classifier), a probabilistic prediction output is calculated using scikit-learn's built-in function [22], which is an implementation of Platt scaling method [11].

In this experiment, scikit-learn [22] python package is used for all three machine learning models. To get the optimal performance of each tested algorithm, the optimal parameters for each algorithm are found using exhaustive grid search before the validation process. For SVM, the linear kernel SVM binary classifier from scikit-learn package is used with an optimal penalty factor  $C = 1.94$ . For random forest, the optimal pair of parameters are: Maximum Features (MF)= 40 and Number of Estimators (NE) = 45. In order to avoid occasional factors, all three algorithms are validated using ten-fold cross validation.

In order to validate the performance of machine learning algorithms with respect to distinguishing potential donors from rejectors, 13,248 persons who is either a known donor or a rejector are randomly selected from the past 5 years. Among them there are 4,671 donors in total, and 1175 of them donated more than \$1,000.

Furthermore, we have found that the data set is imbalanced in that some classes appear more frequently than the others. As a result, the “Balanced” option is used in both SVM and RFC in the scikit-learn package. The balanced mode uses the class label  $y$  to automatically adjust the weights inversely proportional to class frequencies in the input data [22], i.e.,

$$w_i = \frac{\sum_{i=1}^n S_i}{nS_i}, i = 1, 2, \dots, n \quad (3.10)$$

where  $w_i$  is the weight for class  $i$ ,  $n$  is the total number of classes and  $S_i$  is the number of samples in class  $i$ .

### 3.4.2 Performance For Targeting New Donors

We define the donors who contributed less than \$1,000 as *small donors* and otherwise *big donors* in this chapter. When small donors are also considered as the targets, the misclassification of all kinds of donors including very small donors (e.g. donated \$1 in total) matters. The validation result shows that in this case, random forest classifier (RFC) and support vector machine (SVM) perform better than naïve Bayes gaussian (NBG) in terms of f-score, recall rate and accuracy.

The random forest classifier in this experiment has the highest f-score of 0.59 with a precision of 57.45% and a recall of 61.25%. When the random forest classifier is set to maximize the precision, it produces a high precision of 99.49% with a recall of 16.83%. However, their performances are not very satisfying because one third of samples are the donors in the validation data set which gives a 30% of baseline precision for random guessing.

Table 3.2: The performance of the three algorithms for distinguishing donors from rejectors when all donors are used for evaluation, with Maximum Precision, Maximum Accuracy and Maximum F-score, respectively

Performance of Algorithms with Different Thresholds								
Measure	TP	FP	TN	FN	Precision	Recall	Accuracy	F-Score
NBG Max Precision	1159	337	8240	3512	77.47%	24.81%	70.95%	0.38
RFC Max Precision	786	4	8573	3885	99.49%	16.83%	70.64%	0.29
SVM Max Precision	1	0	8577	4670	100%	0.02%	64.75%	0
NBG Max Accuracy	2184	1130	7447	2487	65.90%	46.76%	72.70%	0.55
RFC Max Accuracy	1838	331	8246	2833	84.74%	39.35%	76.12%	0.54
SVM Max Accuracy	1846	360	8217	2825	83.68%	39.52%	75.96%	0.54
NBG Max F-Score	3132	3260	5317	1539	49.00%	67.05%	63.78%	0.57
RFC Max F-Score	2861	2119	6458	1810	57.45%	61.25%	70.34%	0.59
SVM Max F-Score	3136	2562	6015	1535	55.04%	67.14%	69.07%	0.60
NBG Default	1455	555	8022	3216	72.39%	31.15%	71.54%	0.44
RFC Default	2044	572	8005	2627	78.13%	43.76%	75.85%	0.56
SVM Default	2191	864	7713	2480	71.72%	46.91%	74.76%	0.57

### 3.4.3 Performance For Targeting New Big Donors

According to the statistics in Chapter 2, the donors who donate more than \$1,000 in total contribute 95% more in the fundraising program. As such targeting these donors is most important for the success of fundraising. In order to know the performance of

the trained models for identifying new big donors, the accuracy, precision and f-score of targeting new small donors will not be involved for calculating the performance in this section.

The results show that the prediction performance for all three algorithms looks significantly better when the small donors are excluded for evaluation. The support vector machine in this case has the highest f-score of 0.76 with a precision of 75.92% and a recall rate of 75.15%. The random forest classifier produces the second highest f-score of 0.74 with a precision of 84.86% and a recall rate of 65.36%.

When the random forest classifier is set to maximize the precision, it produces a precision of 98.96% with a recall rate of 32.43%. As the baseline precision of random guessing for targeting new big donors is 12.8%, the performance of targeting new big donors of the machine learning algorithms is delightful.

Table 3.3: The performance of the three algorithms for distinguishing donors from rejectors when only big donors are used for evaluation, with Maximum Precision, Maximum Accuracy and Maximum F-score, respectively

Performance of Algorithms with Different Thresholds								
Measure	TP	FP	TN	FN	Precision	Recall	Accuracy	F-Score
NBG Max Precision	667	337	8240	508	66.43%	56.77%	91.34%	0.61
RFC Max Precision	381	4	8573	794	98.96%	32.43%	91.82%	0.49
SVM Max Precision	1	0	8577	1174	100%	0.09%	87.96%	0.00
NBG Max Accuracy	667	337	8240	508	66.43%	56.77%	91.34%	0.61
RFC Max Accuracy	731	96	8481	444	88.39%	62.21%	94.46%	0.73
SVM Max Accuracy	762	127	8450	413	85.71%	64.85%	94.46%	0.74
NBG Max F-Score	764	504	8073	411	60.25%	65.02%	90.62%	0.63
RFC Max F-Score	768	137	8440	407	84.86%	65.36%	94.42%	0.74
SVM Max F-Score	883	280	8297	292	75.92%	75.15%	94.13%	0.76
NBG Default	781	555	8022	394	58.46%	66.47%	90.27%	0.62
RFC Default	918	572	8005	257	61.61%	78.13%	91.50%	0.69
SVM Default	971	864	7713	204	52.92%	82.64%	89.05%	0.65

### 3.4.4 ROC Curves and Learning Curves

The ROC curves show that the predicting efficiency for big donors is much better in terms of AUROC, recall and accuracy than other donors. Both RFC and SVM can produce a 0.7 true positive rate with less than 0.1 false positive rate. The naïve bayes

classifier performs a bit worse, but still has a decent prediction ability.

Although the prediction performance looks worse when small donors are included for evaluation, the prediction accuracy is very good when fewer donors are predicted. When all donors are used for evaluation, the ROC curves show that the prediction accuracy decreases drastically for all three algorithms when the TP rate goes beyond 0.3, which implies that 30% of positive samples in the validation set can be predicted at a much lower cost (in terms of FP rate) while the rest donors are costly to predict (in terms of FP rate).

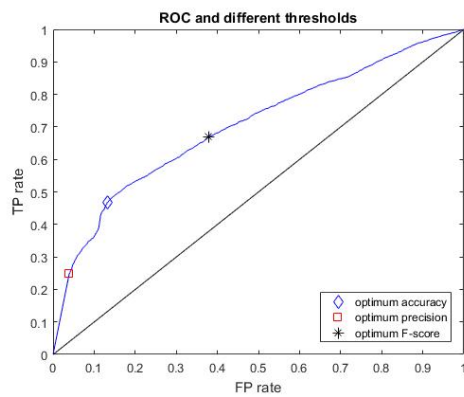


Figure 3.4: ROC for naïve Bayes classifier when all donors are involved for evaluation

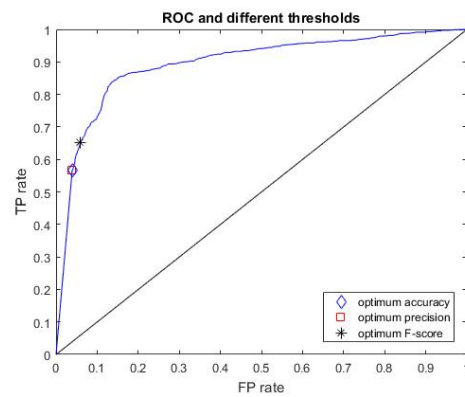


Figure 3.5: ROC for naïve Bayes classifier when only big donors are involved for evaluation

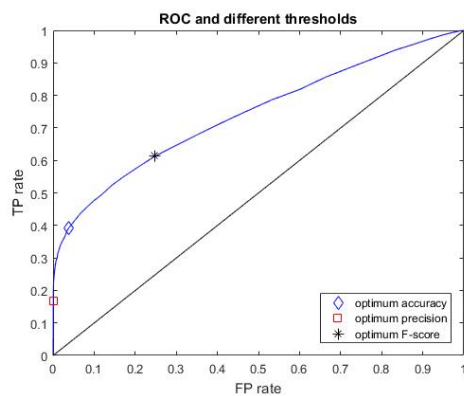


Figure 3.6: ROC for random forest classifier when all donors are involved for evaluation

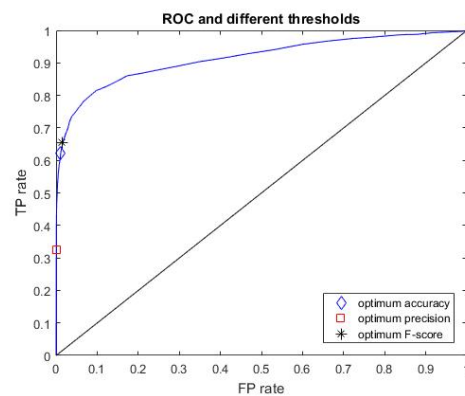


Figure 3.7: ROC for random forest classifier when only big donors are involved for evaluation

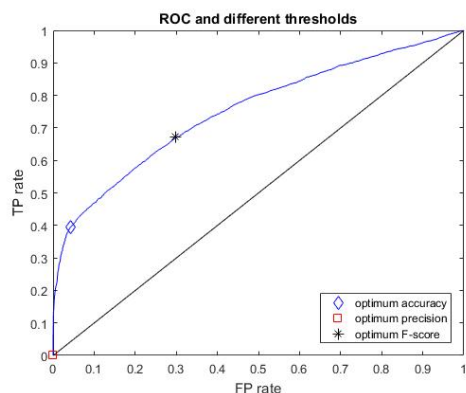


Figure 3.8: ROC for support vector classifier when all donors are involved for evaluation

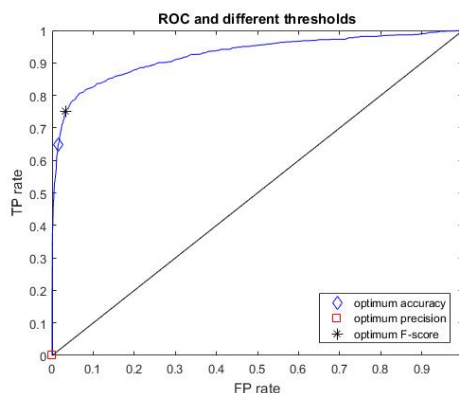


Figure 3.9: ROC for support vector machine when only big donors are involved for evaluation

The learning curves show that all three algorithms learn very fast, but the prediction power does not improve much when more than 3,000 training samples are used. With only 3,000 training samples used, all three algorithms almost reached the same prediction capability with 10,000 training samples.

When all donors (including small donors) are involved for evaluating the prediction, learning curve scores are significantly lower than those when only big donors are involved for evaluation.

The learning curves also indicate that the SVM and RFC generally produce a significantly higher recall rate than naïve bayes classifier. SVM has the highest recall rate and the highest AUROC. Although SVM and RFC outperform naïve bayes classifier in recall rate when enough training samples are used, the naïve bayes classifier produces a better accuracy than SVM and RFC when predicting big donors. All three algorithms perform similarly in terms of AUROC and require at least 1000 training samples to produce stable predicting result.

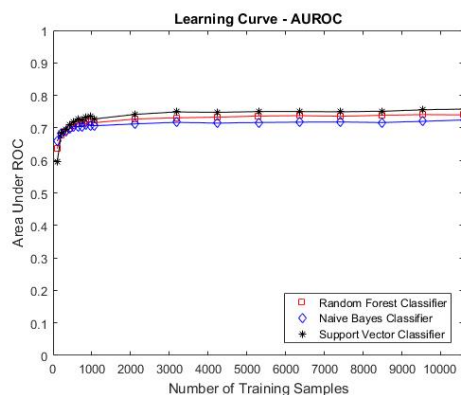


Figure 3.10: Learning curves of AUROC for all three classifiers when all donors are used for evaluation

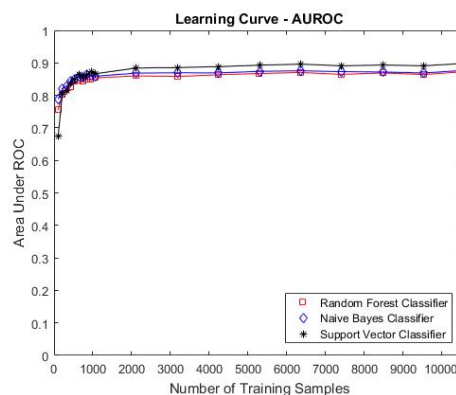


Figure 3.11: Learning curves of AUROC for all three classifiers when only big donors are used for evaluation

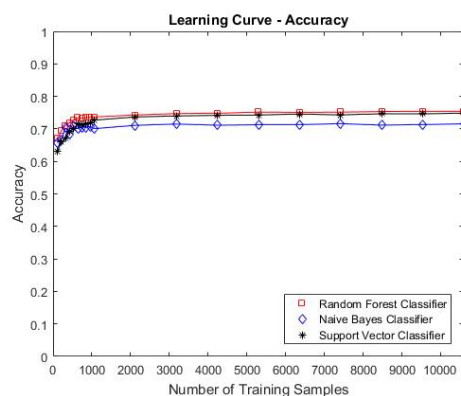


Figure 3.12: Learning curves of accuracy for all three classifiers when all donors are used for evaluation

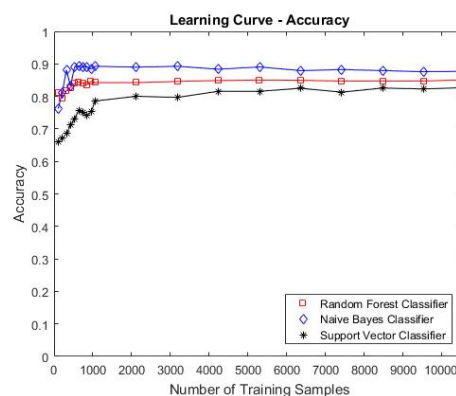


Figure 3.13: Learning curves of accuracy for all three classifiers when only big donors are used for evaluation

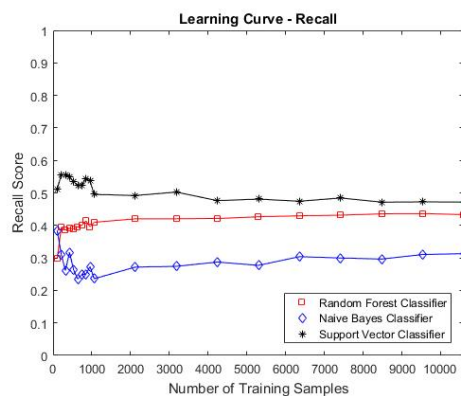


Figure 3.14: Learning curves of recall rate for all three classifiers when all donors are used for evaluation

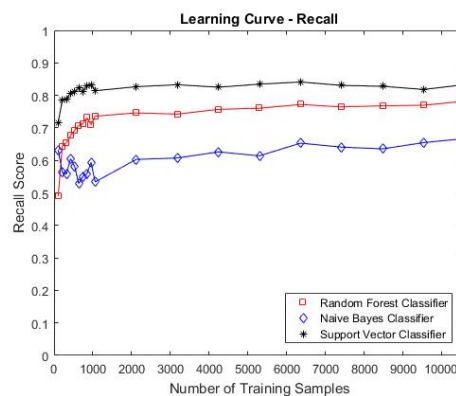


Figure 3.15: Learning curves of recall rate for all three classifiers when only big donors are used for evaluation

## 3.5 Performance in a Practical Scenario

The validation process shows the ability of machine learning models for distinguishing donors from rejectors by learning their personal features and affiliation factors. Nevertheless, their performance needs further evaluation in a practical scenario for the following reasons:

1. Most of the samples for the donor class in the cross validation data set have donated for many years. Their affiliation features (e.g., events registered) could change when they made more donations, and being a donor for a long time will make themselves more identifiable. However, in the practical scenario, all the candidates are those never donated before, which makes the identification task much harder.
2. The proportion of small donors, big donors and rejectors in the practical scenario is unknown. As discussed above, big donors are identified more often. If the proportion of small donors is too big in the practical scenario, the performance could be worse than the validation test.

Therefore, in order to know the performance of the machine learning models in a practical scenario, a test scenario similar to the real application environment is necessary and important.

In this experiment, the new test set consists of all 2015 new donors, who gave their first donation in the year of 2015, and 2015 rejectors, who never donated before and have been reported to had rejected the campaign over phone during the 2015 calling program. The training set is the whole data set used in the validation section, excluding the people in the new test set.

Table 3.4: Test result for all 3 algorithms with default predicting thresholds

Test Result				
	NBG_Def	RFC_Def	SVM_Def	Everyone
Candidates Picked Up	138	363	340	2495
Donors Included	56	124	125	458
Precision	40.58%	34.16%	36.76%	18.36%
Recall	12.23%	27.07%	27.29%	100%
Big Donors Included	11	19	19	22
Big Donors Precision	7.97%	5.23%	5.59%	0.88%
Big Donors Recall	50.00%	86.36%	86.36%	100%

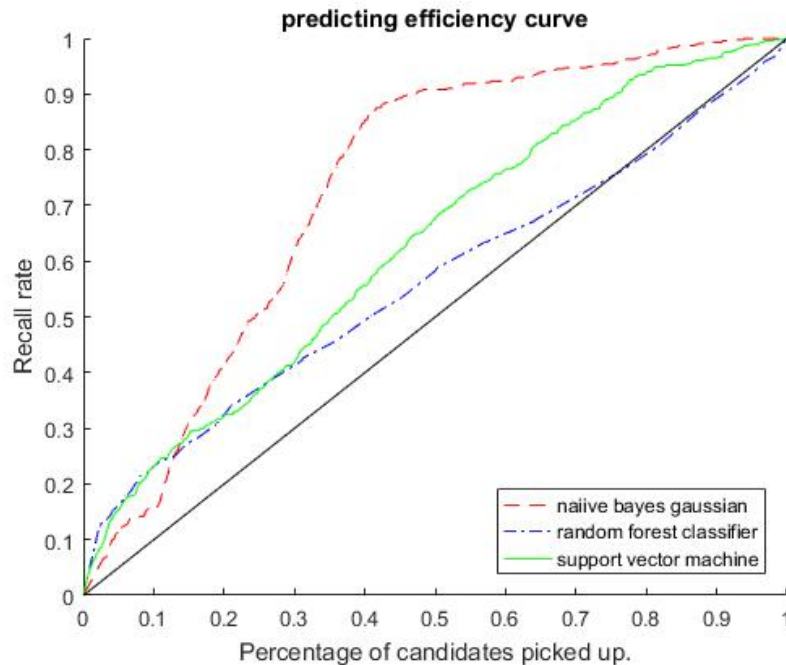


Figure 3.16: Prediction recall curve for all donors. The horizontal axis represents the percentage of candidates labeled as potential donors. The vertical axis represents the percentage of donors acquired.

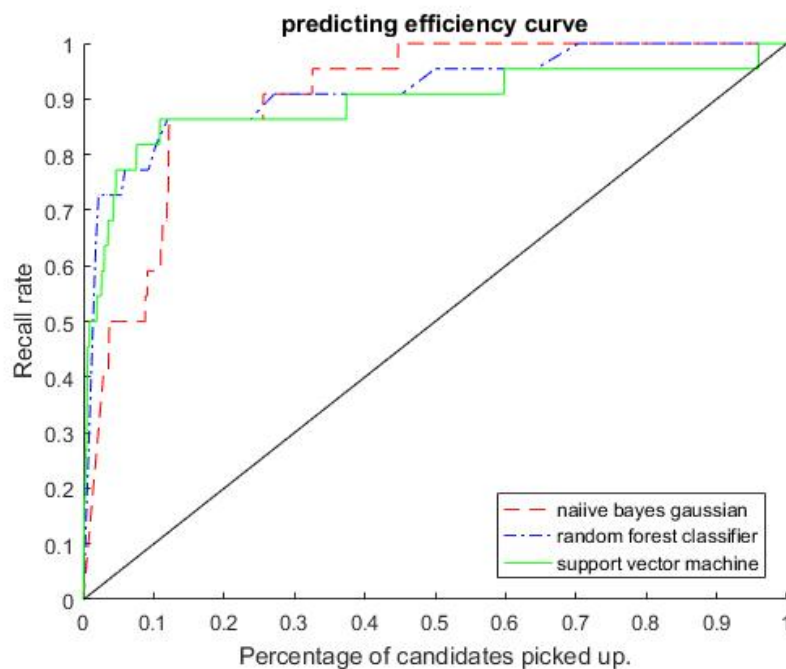


Figure 3.17: Prediction recall curve for big donors. The horizontal axis represents the percentage of candidates labeled as potential donors. The vertical axis represents the percentage of big donors acquired.

The test results are shown in Table 3.4. Noting that the default predicting threshold equals 0.5 in probabilistic classifiers. The probabilistic prediction output for SVM is calculated using scikit-learn’s built-in function [22], which is an implementation of Platt scaling method [11]. The prediction recall curve for all donors and that for big donors are shown in Figure 3.16 and 3.17, respectively.

From the results, among the 240 out of 2,495 candidates who are predicted to be donors, 86.36% of the total big donors are covered with a big-donor precision of 5.59%, which is impressive because the big-donor precision for random selecting is only 0.88%. The results also show that the candidates labeled as donors will have 6 times more chance to become a big donor than normal candidates. Besides, the candidates labeled as donors by any of the models have almost twice chance to become donors.

The prediction recall curves show that the SVM and RFC predict more efficiently than NBG when the predicting thresholds are set for labeling less than 20% candidates as donors. However, when the predicting thresholds are set for labeling more candidates, the NBG produces a higher recall rate than SVM and RFC for both general donors and big donors.

## 3.6 Summary

The problem of targeting prospective donors can be modeled as a donor identification problem with machine learning. In the model validation section, three different machine learning algorithms were tested (SVM, RFC and Naïve Bayes classifier), and their performance were compared. For a robust test, the validation process were conducted using ten-fold cross validation method. In the validation section, the SVM performed the best in terms of f-score, accuracy and recall rate. When the small donors were not used for evaluation, all three algorithms showed better performances, and SVM could produce the highest overall predicting accuracy of 94.46% with a recall of 64.85%. When small donors were included for evaluation, however, the highest prediction accuracy was 76.12% with a recall of 39.35% produced by RFC.

Another test for the real world scenario of 2015 fiscal year was built and presented. The results show that most of the new big donors in 2015 could be predicted, and all candidates predicted as donors have a much higher chance to become a donor than others. Besides, the big donors are actually much more predictable than small donors. This result has significant practical meaning, because according to the statistics in the previous chapter, the big donors make 95% more contribution to the total fund raised. In this test, among the 13.6% out of 2495 candidates predicted by SVM, 86.36% of the big donors were covered. When only 5.5% candidates were labeled as donors, half of the big donors were still covered, implying that it is possible to acquire most of the big donors when only a very small portion of candidates are solicited. If the machine learning algorithms are set for labeling 40% of the candidates as potential donors, more than 90% of big donors will be captured by any of the algorithms, and more than 85% of general donors will be captured by the naïve bayes classifier.

According to the learning curves, all three algorithms have a similar learning trace. The predicting ability of all three algorithms do not benefit a lot from involving more than 3,000 training samples. They build up their predicting ability quickly as one to two thousand training samples are involved. And 3,000 training samples will get all three algorithms close to their performance with 10,000 training samples. When more than 3,000 training samples are involved, the predicting efficiency of all three algorithms stays almost steady.

## Chapter 4

# Promising Donor Prediction

### 4.1 Problem Description

According to the analysis from Chapter 2, the big donors (Major Donors, Large Donors and Leadership Donors) contribute to more than 90% of the total donation. Among the big donors, half of them were not big donors when they gave their first gifts. As such, donor promotion plays an important role in big donor acquisition. In other words, it is critical to identify “*promising donors*”, a term referring to donors who will upgrade their gift size or current pledge to become a big donor.

Nevertheless, the number of donors for any institution may be tens of thousands of individuals. In a typical donor promotion program, the limited fundraising resources for individual solicitations reinforce the need to reduce the number of candidates to no more than a few hundreds. Making efforts blindly will build up the fundraising cost quickly, because most of the current small donors will finally end up with a small donation due to their personal or affiliation factors. Thus, we should more focus on the candidates who have the real potential to become big donors. Developing a proper targeting model for distinguishing promising donors from non-promising donors becomes an important task, which can boost fundraisers’ efficiency and reduce fundraising cost.

### 4.2 Analysis and Modeling

Among the 35,122 randomly-selected donors, 543 of them had a largest gift greater than or equal to \$10,000, and they are defined as big donors (Major Donors, Large

Donors and Leadership Donors) in this thesis. 256 of the big donors are upgraded big donors, i.e., they donated less than \$10,000 when they gave their first gift. In fact, for most of the upgraded big donors, the last donation before their first big gift above \$10,000 are still not big, as shown in Figure 4.1.

To predict promising donors who will upgrade their donation in the future to become a big donor, training set for both positive and negative samples are required. In this problem the positive training samples are easy to acquire, as those big donors (including upgraded big donors over the years) can be the samples and they are easily found in the data set. The negative training samples, however, is hard to locate, because all small donors still have the chance to upgrade their donations, and some current small donors are potential big donors who should not be considered as the example of non-promising donors.

However, the above problem can be alleviated based on the following observation: From our data set, for most big donors, the time interval between their first big gift and the previous gift is mostly within 2 years, and 83% of the promising donors gave their first big gift within 5 years from their previous donation. Only less than 5% of the promising donors had a longer than 10 years' no gift period before they became a big donor. The above phenomenon is shown in Figure 4.2. The above phenomenon implies that if we treat the donors who have stopped donating for 10 or more years as the samples of non-promising donors, the purity of the negative training set will be at least 95%. Therefore, we regard those donors as negative samples.

We thus have two types of samples:

- Big donors: Donors whose largest gift is greater than or equal to \$10,000.
- Non-promising donors: Donors whose largest gift is less than \$10,000 and have not donated in the past ten years

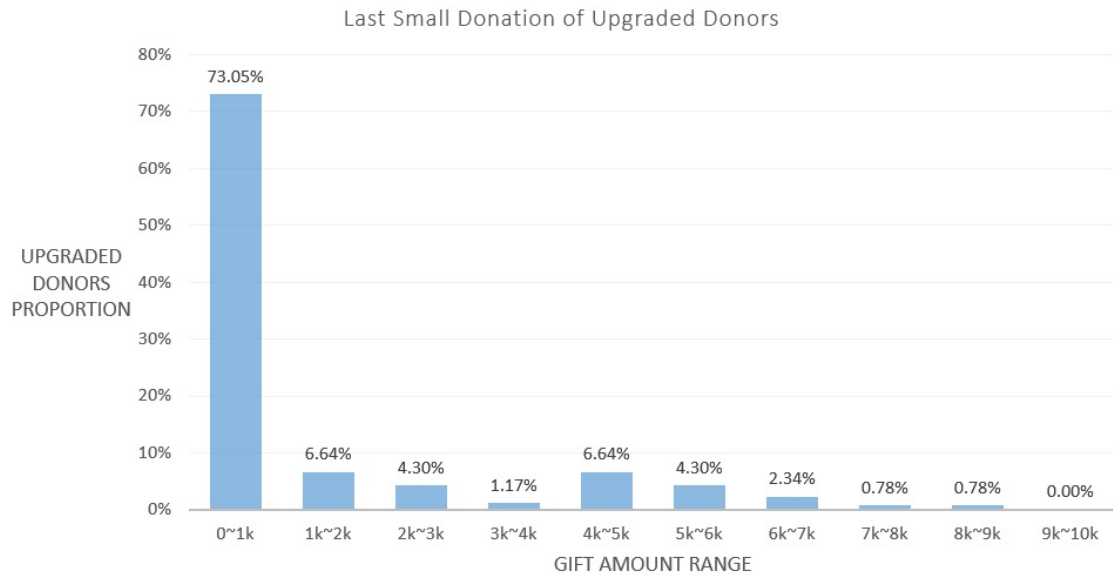


Figure 4.1: Last small donation from upgraded big donors. The horizontal axis represents the amount of the previous donation before the donor's first big donation. The vertical axis represents the proportion of upgraded big donors.

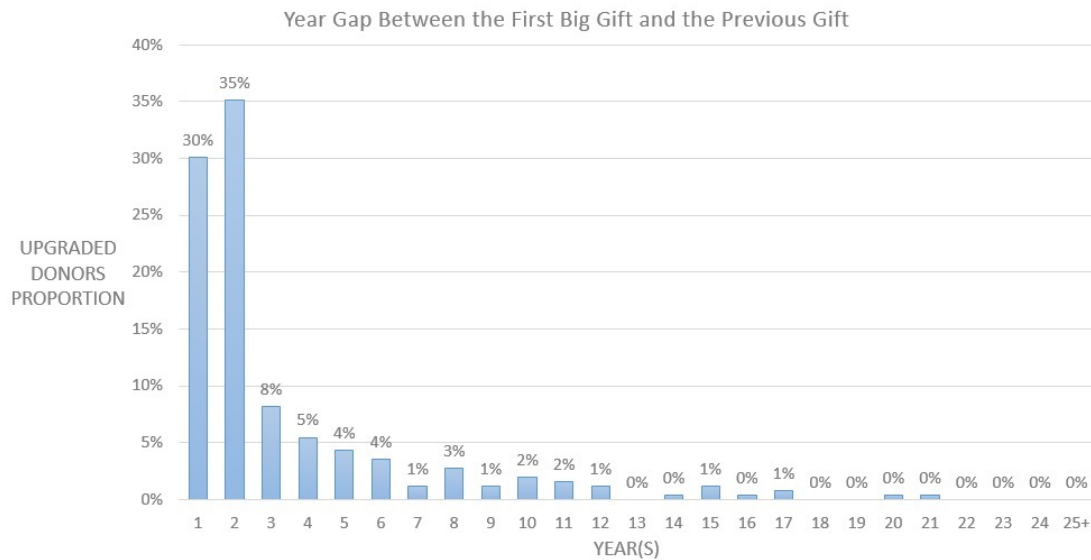


Figure 4.2: Year gap between the first big gift and the previous gift. The horizontal axis represents the year gap between the first big donation and the previous donation. The vertical axis represents the proportion of upgraded big donors.

In the following, we use and evaluate the Naïve Bayes(Gaussian) Classifier, the Linear Support Vector Machine (SVM) and the Random Forest Classifier in the pre-

diction of promising donors.

## 4.3 Data Preparation

### 4.3.1 Data Selection

To test the algorithms' ability of distinguishing promising donors from non-promising donors, data samples for both classes are required. Because there are not many big donors in the database (big donors are only a very small portion in donors) and also because the models need enough samples for both classes to be trained properly, all the big donors through the years whose personal information is in the database are selected as samples for the big donors class. Nevertheless, as the big donor information was collected from different times when they donated, their current age do not necessarily reflect their donation behavior at that time. For this reason, the ages of the big donors used in the training set will be the ages when they made their first big donation.

Compared with the big donor samples, there are much more samples for the non-promising donors. Therefore, only the non-promising donors in the recent five years are selected as the samples (e.g., a non-promising donor in 2015 is a donor who stopped donating by 2005). As a result, there are 503 big donors and 4290 non-promising donors in total for our study.

Furthermore, it is noticed that the age information for many donors in the data set, including some big donors, are missing. Excluding the samples and candidates without age information from the data set is not a great solution, because donors whose real ages are not recorded can have other great affiliation factors and should be predicted as promising donors as well. In addition, in a real world scenario, there will be many big donors whose real ages are not recorded. In order to predict candidates without age information, two models are built and tested with different number of attributes. To be specific, in one model, age is used as a feature and the model is trained with candidates that all have age information; in the other model, age is not used as a feature and the model is trained with all candidates. The candidates with age information will be predicted using the first model, and the candidates without age information will be predicted using the second model.

### **4.3.2 Feature Selection**

As discussed in the previous chapter, the feature selection is an important part in machine learning, because redundant and irrelevant features will have negative impact on the prediction performance. In this chapter, the same rules and methods of Chapter 3 will be applied. Furthermore, since the candidates in the promising donor prediction problem are all donors, the age of first donation and the years of donation will be included as important features about their donation history.

### **4.3.3 Categorical Features Transformation and Data Normalization**

The categorical features will be transformed, and numerical features will be normalized using the same technique described in Chapter 3.

### **4.3.4 Correlated Numerical Feature Transformation**

The correlated numerical attributes are the same as those in Chapter 3. The Principal Components Analysis (PCA) is used to transform the highly-correlated financial and charity information into three orthogonal principal components.

## **4.4 Model Validation**

### **4.4.1 Evaluation Metrics**

Referring to the same discussion in Chapter 3, the same evaluation metrics will be used here for evaluating the model by different views. These metrics are: AUROC, Accuracy, Precision, F-score and Recall Rate.

### **4.4.2 Performance For Targeting Promising Donors**

In this section, we use Scikit-learn [22] python package for validating both models with Naïve Bayes (Gaussian) Classifier (NBG), Linear Support Vector Machine (SVM) and Random Forest Classifier (RFC). For a robust test, each algorithm will be validated using a ten-fold cross validation.

To optimize the performance for each algorithm, optimal parameters for the algorithms are found before conducting the validation test by doing the exhaustive

grid search within an assigned domain as discussed in Chapter 3. As a result, the linear kernel SVM binary classifier from the scikit-learn package is used with an optimal penalty factor  $C = 0.1$ . For random forest, the optimal pair of parameters are: Maximum Features (MF) = 25 and Number of Estimators (NE) = 37.

Referring to the same discussion in Chapter 3, we know that the data set is unbalanced. The 'Balanced' option [22] is used for both SVM and RFC in the scikit-learn package, which uses the values of class label  $y$  to automatically adjust weights inversely proportional to class frequencies in the input data:  $w_i = \frac{\sum_{i=1}^n S_i}{nS_i}$ ,  $i = 1, 2, \dots, n$ , where  $w_i$  is the weight for class  $i$ ,  $n$  is the number of classes and  $s_i$  is the number of samples in class  $i$ .

In order to know the best performance in different measurements (accuracy, precision and f-score), the optimal prediction thresholds for all different measurements are calculated for each algorithm. The prediction thresholds are the probabilities in the probabilistic classifiers (naïve bayes gaussian and random forest classifier). For support vector machine (a non-probabilistic binary classifier), a probabilistic prediction output is calculated using scikit-learn's built-in function [22], which is an implementation of Platt scaling method [11].

The experiment shows that all three machine learning algorithms can distinguish promising donors from non-promising donors very efficiently with or without age information. The highest f-score is produced by the RFC, and the f-score reflects both recall rate and accuracy.

In the model trained with age information, the maximum accuracies achieved by the naïve bayes gaussian (NBG), random forest classifier (RFC) and support vector machine (SVM), are 91.28%, 96.85% and 95.53%, respectively. The maximum f-score produced by NBG RFC and SVM were 0.57, 0.83 and 0.76, respectively. The RFC and SVM performed better than NBG in terms of f-score and accuracy when all training samples are provided. In the case that the age information was included, the RFC with the threshold maximizing the f-score produced a precision of 80.58% with a recall rate of 86.16%, and the SVM with the f-score threshold produced a precision of 75.42% with a recall rate of 77.51%. When the RFC was set to maximize the precision, it produced a precision of 97.44% in the test set with a recall rate of 52.60%.

In the model trained without age information, the performance results w.r.t. f-score, accuracy and recall rate of the three algorithms were not as good as the corresponding ones in the model with age information. The maximum accuracies achieved

Table 4.1: Performance of algorithms with thresholds for Maximum Precision, Maximum Accuracy and Maximum F-score, respectively, when age information is included

Performance of Algorithms for the Model with Age Information								
Algorithm	TP	FP	TN	FN	Precision	Recall	Accuracy	F-Score
NBG Max Precision	179	159	2636	110	52.96%	61.94%	91.28%	0.57
RFC Max Precision	152	4	2791	137	97.44%	52.60%	95.43%	0.68
SVM Max Precision	1	0	2795	288	100%	0.35%	90.66%	0.01
NBG Max Accuracy	179	159	2636	110	52.96%	61.94%	91.28%	0.57
RFC Max Accuracy	224	32	2763	65	87.50%	77.51%	96.85%	0.82
SVM Max Accuracy	223	72	2723	66	75.59%	77.16%	95.53%	0.76
NBG Max F-Score	192	191	2604	97	50.13%	66.44%	90.66%	0.57
RFC Max F-Score	249	60	2735	40	80.58%	86.16%	96.76%	0.83
SVM Max F-Score	224	73	2722	65	75.42%	77.51%	95.53%	0.76
NBG Default	229	360	2435	60	38.88%	79.24%	86.38%	0.52
RFC Default	210	21	2774	79	90.91%	72.66%	96.76%	0.81
SVM Default	257	191	2604	32	57.37%	88.93%	92.77%	0.70

by the NBG, RFC and SVM were 89.53%, 95.47% and 94.49%, respectively. The maximum f-scores for NBG, RFC and SVM were 0.53, 0.76 and 0.74, respectively. The RFC and SVM still performed better than NBG in terms of f-score and accuracy. The RFC using the f-score threshold produced a precision of 78.98% with a recall rate of 73.96%. When RFC used the default prediction threshold, it produced a precision of 89.17% with a recall rate of 63.82%. When SVM was set to maximize the precision, it produced a precision of 98.82% with a recall rate of 33.40%.

Overall, the prediction performance of all algorithms improves much over that with random selection, whose precision is 10.2% for the test set with age information and 9.2% for the test set without age information.

Table 4.2: Performance of algorithms with thresholds for Maximum Precision, Maximum Accuracy and Maximum F-score, respectively, when age information is not included

Performance of Algorithms for the Model without Age Information								
Measure	TP	FP	TN	FN	Precision	Recall	Accuracy	F-Score
NBG Max Precision	258	257	4033	245	50.10%	51.29%	89.53%	0.51
RFC Max Precision	168	2	4288	335	98.82%	33.40%	92.97%	0.50
SVM Max Precision	1	0	4290	502	100%	0.20%	89.53%	0
NBG Max Accuracy	258	257	4033	245	50.10%	51.29%	89.53%	0.51
RFC Max Accuracy	318	32	4258	185	90.85%	63.22%	95.47%	0.75
SVM Max Accuracy	377	138	4152	126	73.20%	74.95%	94.49%	0.74
NBG Max F-Score	303	347	3943	200	46.62%	60.24%	88.59%	0.53
RFC Max F-Score	372	99	4191	131	78.98%	73.96%	95.20%	0.76
SVM Max F-Score	385	150	4140	118	71.96%	76.54%	94.41%	0.74
NBG Default	353	524	3766	150	40.25%	70.18%	85.94%	0.51
RFC Default	321	39	4251	182	89.17%	63.82%	95.39%	0.74
SVM Default	417	345	3945	86	54.72%	82.90%	91.01%	0.66

#### 4.4.3 ROC Curves and Learning Curves

The ROC curves, as shown in Figures 4.3, 4.4, 4.5, 4.6, 4.7, and 4.8, indicate that the prediction efficiency in the model with age information is better than that with the model without age information for every algorithm. When age information was

not included, both RFC and SVM produced a true positive rate of 80% with a false positive rate less than 10%. When the age information was included, they both produced a true positive rate above 90% at the cost of a false positive rate less than 10%. Among the three algorithms, the naïve bayes classifier performed slightly worse than RFC and SVM, but still had a good prediction ability and produced a true positive rate above 60% with a false positive rate of 10% without age information. When the age information was included, the naïve bayes classifier produced a true positive rate above 70% with a false positive rate less than 10%.

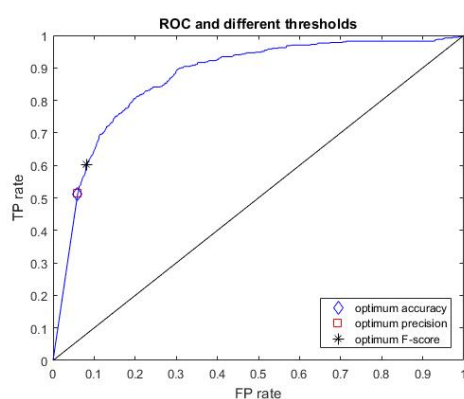


Figure 4.3: ROC for naïve bayes classifier when age information is not included

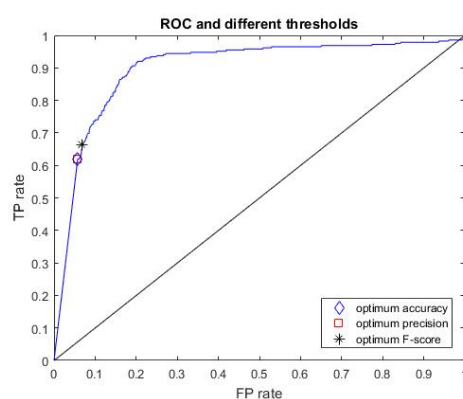


Figure 4.4: ROC for naïve bayes classifier when age information is included

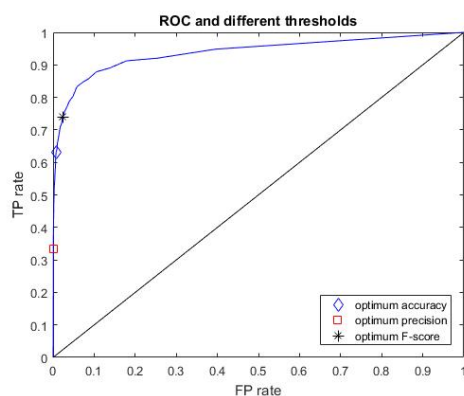


Figure 4.5: ROC for random forest classifier when age information is not included

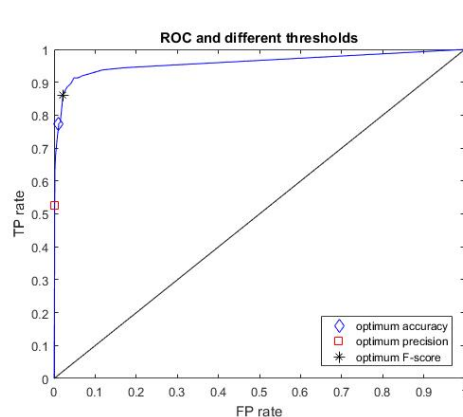


Figure 4.6: ROC for random forest classifier when age information is included

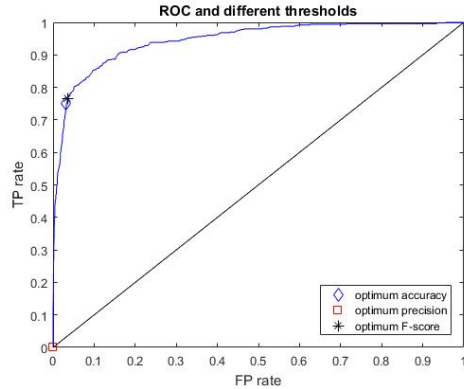


Figure 4.7: ROC for support vector classifier when age information is not included

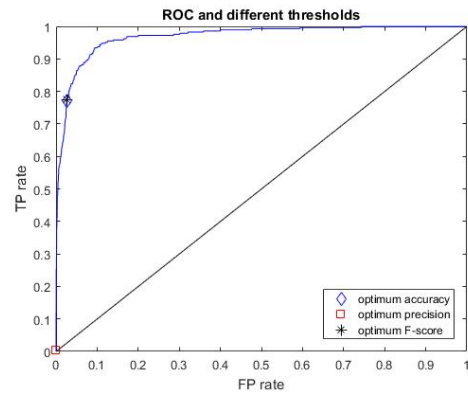


Figure 4.8: ROC for support vector classifier when age information is included

The learning curves show that SVM and RFC outperform naïve bayes classifier in terms of the area under ROC (AUROC) and the accuracy, when all training samples are involved. Besides, the RFC generally produces a higher accuracy while NBG and SVM generally produce a higher recall. In addition, RFC and SVM need less than 1,000 training samples to get a decent and stable prediction ability, while the naïve bayes classifier needs more training samples. Furthermore, although less training samples were involved in the training with age information, all of the three algorithms made better prediction than in the case where age information was excluded.

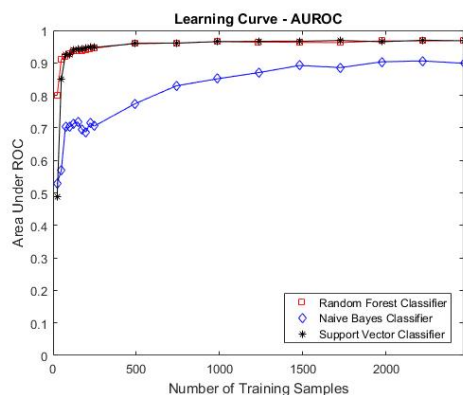


Figure 4.9: Learning curve measured in AUROC for all three algorithms when age information is included

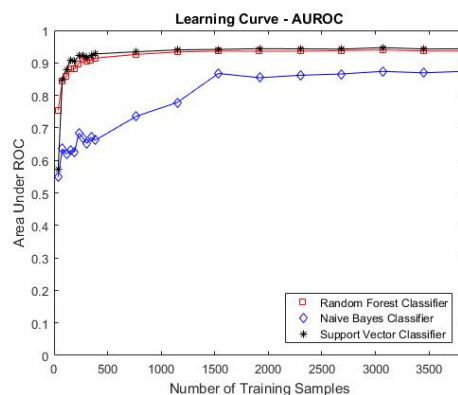


Figure 4.10: Learning curve measured in AUROC for all three algorithms when age information is not included

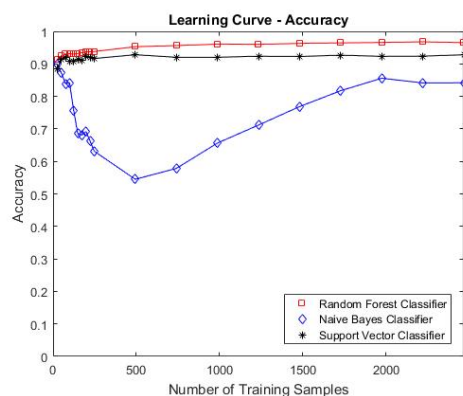


Figure 4.11: Learning curve measured in accuracy for all three algorithms when age information is included

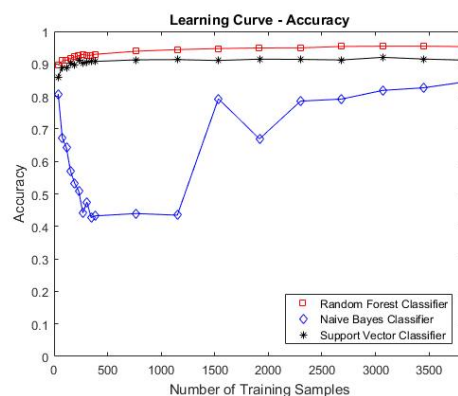


Figure 4.12: Learning curve measured in accuracy for all three algorithms when age information is not included

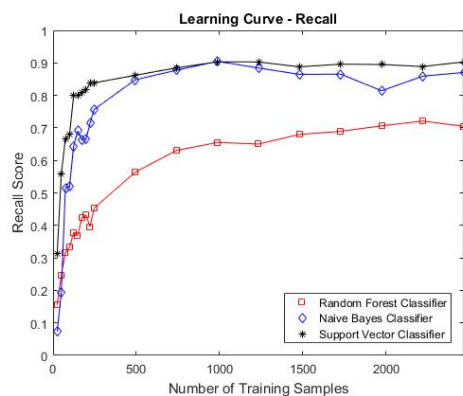


Figure 4.13: Learning curve measured in recall for all three algorithms when age information is included

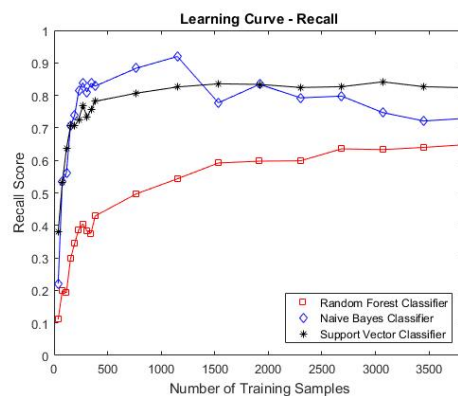


Figure 4.14: Learning curve measured in recall for all three algorithms when age information is not included

## 4.5 Performance in a Practical Scenario

Although the validation results in the previous section showed the potential of machine learning algorithms for distinguishing promising donors from non-promising donors, the performance of the machine learning models need further test in the practical scenario, with the similar reasons listed in Chapter 3.5.

In this section, we perform a test with a case simulating the real application environment in 2011. All the big donors and non-promising donors by 2011 will be the training samples. And all the donors who are neither big donors nor considered as non-promising donors by 2011 will be the candidates. The two models are built with and without age information as in the previous section. In this experiment, the candidates with age information will be predicted by model trained with age information, and the candidates without age information will be predicted by model trained without age information. Therefore, the performances of both models in the practical scenario will be tested and presented. All machine learning algorithms will be used with default prediction threshold which equals 0.5 in probabilistic classifiers.

Table 4.3: Prediction of promising donors from current small donors

Predicting 10889 Current Donors Who Have Age Information			
Algorithms	NBG_Default	RFC_Default	SVM_Default
Candidates Selected	2691	896	2072
Current Recall	86.44%	54.24%	83.05%
Expected Recall	79.24%	72.60%	88.93%
Predicting 4689 Current Donors Who Do Not Have Age Information			
Algorithms	NBG_Default	RFC_Default	SVM_Default
Candidates Selected	1381	403	1109
Current Recall	82.35%	52.94%	76.47%
Expected Recall	70.18%	63.82%	82.90%

The machine learning models are trained with personal features and affiliation factors, and the predicted promising donors are picked up according to their good personal and affiliation factors. Even if someone does have very good personal features and a strong affiliation with the university and finally becomes a big donor, it may take very long time to see this happen because they might just keep donating small gifts until they finally accomplish their life goals or just pay off the debt.

Because the prediction accuracy in this study cannot be correctly calculated until all the candidates are marked as either big donor or non-promising donor, the only

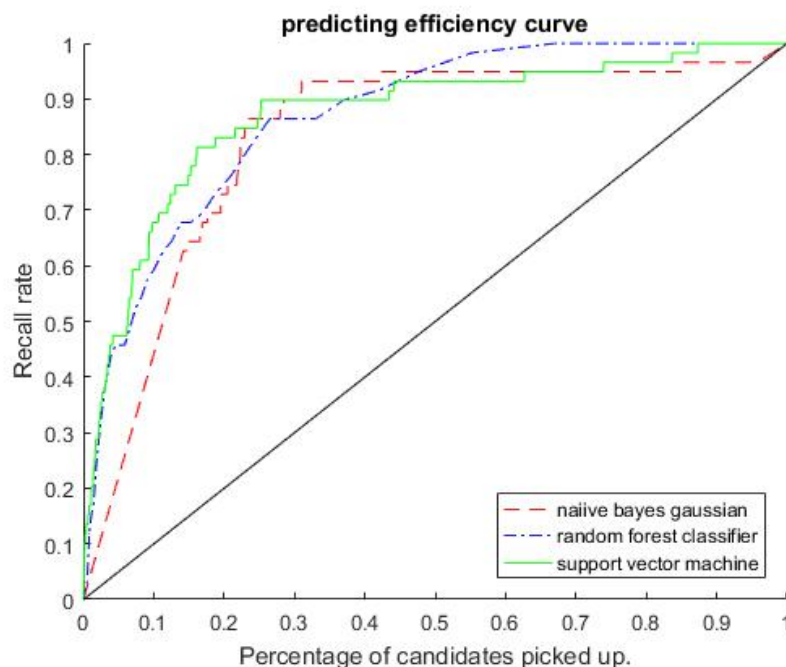


Figure 4.15: Prediction recall curve for the model with age information. The horizontal axis represents the percentage of candidates labeled as a promising donor. The vertical axis represents the percentage of upgraded big donors acquired

measurement will be the current recall rate, which shows how many upgraded big donors so far have been the predicted candidates. And the expected recalls shown in Table 4.3 are the recall rates calculated in the “model validation” section.

When the age information was included, RFC predicted 896 promising donors out of 10889 candidates. Only 8% candidates were labeled as promising donors, and 54.24% of the upgraded big donors by now (5 years from 2011) have been covered in the promising donors group. 86.44% of the promising donors are covered in the NBG selected group which only has 24.71% of the total candidates.

When the age information was not included, RFC in this test predicted 403 promising donors out of 4689 candidates. Within the predicted 8.6% of the total, 52.94% of the upgraded big donors by now have been covered. And 82.35% of the promising donors are covered in the NBG selected group which only has 29.45% of the total candidates.

According to the prediction recall curves, it is noticed that it is possible to acquire more than 90% of upgraded big donors when less than 40% of the candidates are labeled as promising donors (or solicited in a practical scenario). When the age

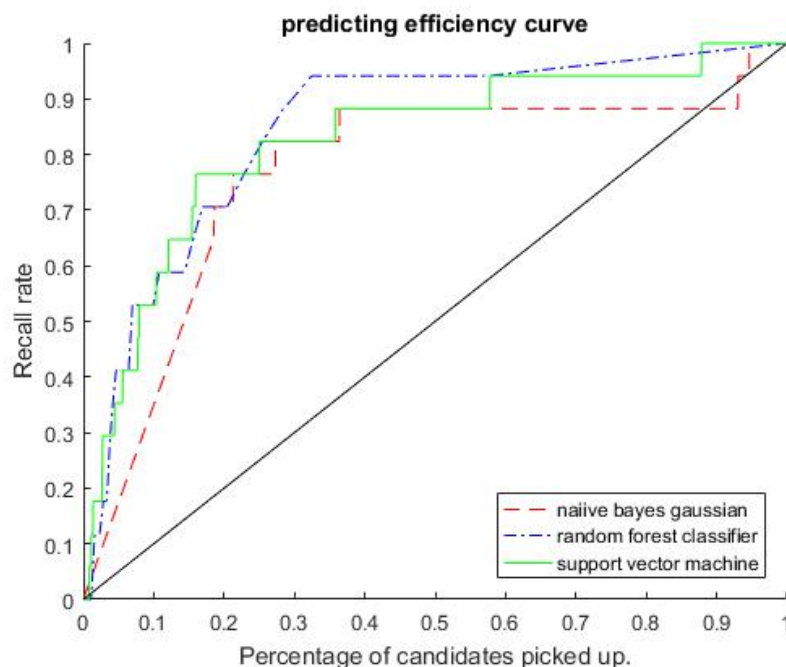


Figure 4.16: Prediction recall curve for the model without age information. The horizontal axis represents the percentage of candidates labeled as a promising donor. The vertical axis represents the percentage of upgraded big donors acquired.

information is not given, SVM and RFC predict better than NBG. However, when the age information is given, the RFC and SVM predict better than NBG only when the algorithms are set to label less than 20% of the total candidates as promising donors.

## 4.6 Summary

In this chapter, we studied the prediction of promising donors. In the model validation section, three different machine learning algorithms (SVM, RFC and Naïve bayes classifier) are validated with ten-fold cross validation. The results show that all three algorithms have great ability of predicting promising donors in the given data set.

Among the algorithms tested, RFC works the best for predicting promising donors. Generally, SVM has a better recall rate while the RFC has a higher accuracy. The algorithms trained with age information (including current age and first donation age) perform better than those trained without age information. When the age information was included, the RFC with the predicting threshold that maximizes the f-score

produced a precision of 80.58% with a recall rate of 86.16%. When the predicting threshold of RFC is set to maximize the precision, the RFC produced a precision of 97.44% in the test set with a recall rate of 52.60%. When the age information was not included, the RFC with the predicting threshold that maximizes the f-score produced a precision of 78.98% with a recall rate of 73.96%. When the predicting threshold of RFC was set to maximize the precision, the RFC produced a precision of 98.82% in the test set with a recall rate of 33.40%.

According to the learning curves, RFC and SVM need at least 1000 training samples to get a decent and stable predicting ability while the naïve bayes classifier needs more training samples. Furthermore, When the training set has age information, all three algorithms predicted better in this case. This phenomenon shows that the age information is important for the performance of the predictors.

The test results in Section 4.5 show that more than 90% of potential big donors can be predicted by well trained machine learning algorithms and acquired with less than 40% of the candidates being solicited. In the test, when the age information was given, with 24.71% of candidates labeled as promising donors, 86.44% of the potential big donors were predicted by NBG using default prediction threshold. For the candidates without age information, 82.35% of the potential big donors were predicted by NBG when 29.45% of the candidates were labeled as promising donors. These results demonstrate that it is practical to acquire most of the potential big donors when only a small portion of candidates are solicited. And also, machine learning algorithms do have the ability to predict potential big donors based on their personal and affiliation features which are normally available in alumni database.

# Chapter 5

## Conclusions and Future Work

### 5.1 Conclusions

In this thesis, we tackled two major fundraising challenges by modeling the problems with supervised machine learning. Machine learning solutions are presented and their effectiveness are investigated.

The prospective donor prediction problem is modeled as a donor identification task. Naïve bayes (Gaussian) classifier (NBG), support vector machine (SVM) and random forest classifier (RFC) are trained and tested with current donors and recent rejectors. Test result shows that SVM produces the best predictive model for distinguishing donors from rejectors. It is found that the big donors are identified significantly more often. In the experiment where small donors are excluded from the test set, SVM identified 75.15% of the donors with a precision of 75.92%. RFC identified 65.36% of the donors with a precision of 84.86%. In the experiment where small donors are not excluded from the test set, however, SVM identified 67.14% donors with a precision of 55.04%. In a practical scenario where the model is properly used as a targeting strategy, 86.36% of the big donors can be acquired when only 14% of the candidates are solicited.

For the problem of donor promotion, NBG, SVM and RFC are also used to build predictive models for distinguishing promising big donors from not-promising donors. RFC produces the best predictive models in the test set. RFC generally predicts more accurately while SVM produces a higher recall. Because there are some big donors having no age information in the data set, the algorithms are also trained and tested without age information. However, it is found that age information is a

good predictor which can be used to increase the prediction efficiency. In the test, when the age information, including real age and age at the first gift, was given, RFC produced an prediction accuracy of 96.76% with a recall rate of 86.16%. When the age information was not given, the RFC produced an accuracy of 95.20% with a recall rate of 73.96%, and SVM produced an accuracy of 94.41% with a recall rate of 76.54%. Moreover, the test from a real world scenario indicates that in a practical scenario where the machine learning models are properly used as a targeting strategy, more than 85% of the upgrading donors can be acquired when only 26% of the candidates are solicited, and more than 90% of the upgrading donors can be acquired when 40% of the candidates are solicited.

## 5.2 Limitations and Further Works

### 5.2.1 Limitations

- Because the samples in the data set are alumni, the families or relatives of alumni who have their detailed information in the alumni database, the application of the proposed models are limited to predicting alumni and alumni families or relatives. Other types of donors have not been considered in this study.

### 5.2.2 Future Work

- In this paper, naïve bayes (Gaussian) classifier, support vector machine, and random forest classifier are used and tested. Nevertheless, there are other machine learning algorithms. Their performance when applied to solve our problems is unclear. While it is impossible to evaluate all possible classifiers, it is worth performing more experiments to evaluate other classifiers in the future.
- Parameters play an essential role in algorithms' performance. In this paper, the optimal parameters are estimated by using a brutal grid-based search within an assigned domain. In the future, better methods could be researched to optimize the parameters quickly and effectively.

# Bibliography

- [1] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.
- [2] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.
- [3] Pablo L Durango-Cohen, Elizabeth J Durango-Cohen, and Ramón L Torres. A bernoulli–gaussian mixture model of donation likelihood and monetary value: An application to alumni segmentation in a university setting. *Computers & Industrial Engineering*, 66(4):1085–1095, 2013.
- [4] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Academic press, 2013.
- [5] Nafisseh Heiat. Data mining applications in fund raising. *World Journal of Social Sciences*, 1(4):14–22, 2011.
- [6] P. Hill, T. & Lewicki. *STATISTICS: Methods and Applications*. StatSoft, Tulsa, OK., 2007.
- [7] Tin Kam Ho. Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, pages 278–282. IEEE, 1995.
- [8] I.T. Jolliffe. *Principal component analysis*. Springer, 2002.
- [9] Miroslav Kubat, Robert C. Holte, and Stan Matwin. Machine learning for the detection of oil spills in satellite radar images. *Machine Learning*, 30(2):195–215, 1998.

- [10] Louis A Le Blanc and Conway T Rucks. Data mining of university philanthropic giving: Cluster-discriminant analysis and pareto effects. *International Journal of Educational Advancement*, 9(2):64–82, 2009.
- [11] Hsuan-Tien Lin, Chih-Jen Lin, and Ruby C Weng. A note on platt’s probabilistic outputs for support vector machines. *Machine learning*, 68(3):267–276, 2007.
- [12] Wesley E. Lindahl and Christopher Winship. A logit model with interactions for predicting major gift donors. *Research in Higher Education*, 35(6):729–743, 1994.
- [13] Robert Lowe, Robert C Glen, and John BO Mitchell. Predicting phospholipidosis using machine learning. *Molecular pharmaceutics*, 7(5):1708–1714, 2010.
- [14] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.
- [15] J Travis McDearmon and Kathryn Shirley. Characteristics and institutional factors related to young alumni donors and non-donors. *International Journal of Educational Advancement*, 9(2):83–95, 2009.
- [16] Michael T Miller. The college president’s role in fund raising. 1991.
- [17] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2012.
- [18] Kevin P Murphy. Naive bayes classifiers. *University of British Columbia*, 2006.
- [19] University of British Columbia. Consolidated Financial Statements. <http://vpfinance-dev.sites.olt.ubc.ca/files/2015/04/UBC-FIA-Mar-31-2016-Final-website.pdf>, April 2016.
- [20] University of Victoria. Consolidated financial statements of university of victoria. <https://www.uvic.ca/vpfo/accounting/assets/docs/financial/uvicfinancialstatements/financial-statements-2015-16.pdf>, April 2016.
- [21] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics, 2002.

- [22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [23] David Martin Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. 2011.
- [24] Foster J Provost, Tom Fawcett, and Ron Kohavi. The case against accuracy estimation for comparing induction algorithms. In *ICML*, volume 98, pages 445–453, 1998.
- [25] J Ross Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014.
- [26] Rajeev DS Raizada and Yune-Sang Lee. Smoothness without smoothing: why gaussian naive bayes is not naive for multi-subject searchlight studies. *PloS one*, 8(7):e69566, 2013.
- [27] CHOTIRAT "ANN" RATANAMAHATANA and DIMITRIOS GUNOPULOS. Feature selection for the naive bayesian classifier using decision trees. *Applied Artificial Intelligence*, 17(5/6):475, 2003.
- [28] Lior Rokach and Oded Maimon. *Data mining with decision trees: theory and applications*. World scientific, 2014.
- [29] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 44(1.2):206–226, Jan 2000.
- [30] Xiaogeng Sun, Sharon C Hoffman, and Marilyn L Grady. A multivariate causal model of alumni giving: Implications for alumni fundraisers. *International Journal of Educational Advancement*, 7(4):307–332, 2007.
- [31] Martin Turcotte. Spotlight on canadians: Results from the general social survey-charitable giving by individuals. 2015.
- [32] Harvard University. Financial Report-Fiscal year 2016. [http://finance.harvard.edu/files/fad/files/harvard\\_ar\\_11\\_12016\\_final.pdf](http://finance.harvard.edu/files/fad/files/harvard_ar_11_12016_final.pdf), 2016.
- [33] C. J Van Rijsbergen. *INFORMATION RETRIEVAL*. University of Glasgow, Butterworth, 1979.

- [34] M Linda Wastyn. Why alumni don't give: A qualitative study of what motivates non-donors to higher education. *International Journal of Educational Advancement*, 9(2):96–108, 2009.
- [35] S. Mukherjee Weston, J. Feature selection for svms. *Advances in Neural Information Processing Systems*, 2000.
- [36] Christopher Winship and Wesley E. Lindahl. Predictive models for annual fundraising and major gift fundraising. *Nonprofit Management and Leadership*, 1992.