Classifying Nonprofit Donor Strategy from Cultivating Volunteers

Sanjay Regi Philip
Applied Data Science
Master's Program
Shiley Marcos School of Engineering/ University of San Diego
sregiphilip@sandiego.edu

Maha Jayapal
Applied Data Science
Master's Program
Shiley Marcos School of Engineering/ University of San Diego
mjayapal@sandiego.edu

Abstract
This article examines the use of machine learning applications to assist nonprofit organizations in identifying high-level donors for fundraising efforts. As fundraising is vital to supporting nonprofit organizations in serving their communities, efforts to help organizations be more efficient in their fundraising efforts can have positive impacts on many peoples' lives. Leveraging data from the U.S. census for a proof of concept, a classification application was successfully developed that will help organizations identify individuals who are able to provide large financial gifts .

1 Introduction

The United States is home to many nonprofit organizations that address problematic issues facing society. Many of these organizations raise awareness and support to solve issues such as homelessness, hunger, social justice, and more, by mobilizing large numbers of passionate volunteers to take action to help others. Despite the noble purposes for which these organizations operate, as organizations scale their operations, they need to have high performing fundraising strategies to engage new donors who can sustain their organization's operations with donations.

1.1 Background

For many nonprofit organizations, volunteers can be one of their most valuable assets as it comes to physical help, but also for monetary support. Georgiev (2023) explained that over 1 billion people around the world are volunteers, and in the United States alone, 64 million people spend 50 hours volunteering every year. With over two-thirds of volunteers now finding opportunities online, in tandem with the widespread adoption of digital volunteer databases, many organizations now possess key contact information for individuals who have demonstrated a passion for an organization's cause (Georgiev, 2023). Bonded by a passion for a specific social cause, these individuals are often champions for an organization's purpose and are willing to provide monetary support in the role of a donor, just as much as they are willing to donate their time as a volunteer.

Though these individuals can be a great source of support, volunteers come from extensively diverse backgrounds and often have a large variety of motivating factors driving them. Therefore, organizations need to have specific outreach strategies for different types of potential donors, as is often observed by how nonprofits distinguish strategies for midlevel and high-level donors, where midlevel donors are defined as middle class individuals who donate a relatively small amount on a recurring ongoing basis, and whereas high-level donors are upper-middle to upper class individuals who often make large individual contributions.

1.2 Motivation

As organizations compete for the attention and support of invaluable high-level donors, nonprofits have a need to identify these donors who need to be cultivated with a distinct strategy where they are engaged on a personal level with frequent direct contact, as compared to middle class donors who will likely respond to mass communication even if less personal. For many organizations today, an inordinate amount of time is spent on identifying whether individuals should be cultivated using a midlevel or high-level donor strategy, which takes away from time and effort that could be used towards advancing their actual cause.

By cultivating high-level donors, organizations will be able to increase their fundraising results to ultimately leverage more resources for the important causes for which they exist to serve.

1.3 Objective

The objective of this project will be to provide a method for nonprofit organizations to identify potential high-level donors for fundraising efforts. The hope is to be able to identify these individuals from nonprofits' already existing databases of volunteers, using personal information and data that individuals have already elected to provide. As high-level donors can provide substantial contributions to a nonprofit's cause, being able to identify and cultivate these donors can provide a tremendous benefit to organizations and the people they serve.

1.4 Hypothesis

With the motivation and objective established, our hypothesis for this research project is that machine learning models can be used to distinguish individuals who are likely to be high-level donors from individuals who are likely to be midlevel donors, by leveraging classification techniques with data of prospective donors.

1.5 Planned Methodology

To assist nonprofits in identifying high-level donors within their volunteer and supporter databases, the research project will first be training a classification machine learning model using basic demographic, personal, and occupational data from an anonymized U.S. census report. Feature selection will be conducted to only use features that a standard nonprofit would be able to collect on their volunteer database to ensure that the resulting models can be used by individual nonprofits in the future. Feature selection will also focus on transforming the wide variety of categorical factors to be used for classification.

After feature selection was performed, exploratory data analysis will be conducted to look for class imbalances, to better understand what potential gaps in this initial research application are given the training dataset. Exploratory data analysis will also be performed to understand relationships between the different personal and demographic factors and how they may relate to an individual's income or wealth. Lastly, insights from the exploratory data analysis initiative will be recorded to direct future research and understand what steps will need to be taken for this model to be used widely.

We will be creating multiple machine learning classification models such as logistic regression, naive bayes, decision trees, and more. These models will be trained given a training data set portioned from the census dataset and will be compared using accuracy, precision, recall, and F1 score after given a test dataset. The model that shows the best results, as measured by the aforementioned metrics of precision, recall, and F1 score, will be selected as our model for organizations to use with their actual data.

1.6 Real World Impact

As many nonprofits transform the lives of people in need, efficiency in fundraising efforts will provide opportunities for more people to be helped. Researchers find that the 25 largest charities in the world spend (5%–25%) of their funding on fundraising expenses further validating that efficiencies in fundraising efforts can lead to more allocation of resources for the actual causes that these organizations exist to tackle (Cagala et al., 2021). A successful execution for this project will allow users at nonprofits to take information from their volunteer database and receive an informed suggestion on what type of strategy to leverage when engaging with that individual for fundraising. Further, this execution will allow nonprofits to better compete for the attention of high-level donors, many of which can provide tremendous benefits when relationships are cultivated personably and allow organizations to better allocate their time for fundraising ultimately to help more people.

2 Literature Review

The benefits of data science applications in nonprofit operations have been studied and tested by many researchers in the last several years and provide merit that further research in this field can provide tremendous benefits for nonprofit organizations.

2.1 *Exploring the applications & challenges of data analytics in nonprofit organizations.*

In one such study, Nault et al. (2020) described how opportunities for the application of machine learning to drive benefits for nonprofits are endless. The use cases described in the study include the functions of marketing, operations, and strategy, and emphasize how organizations today find the clearest opportunities for machine learning in fundraising. Furthermore, Nault et al. detailed applications currently in production being used for segmenting donor databases, applying clustering to develop targeting strategies, and other donor and donation identification techniques which have already brought great efficiencies to organizations. Though the researchers detailed an increase in the application of machine learning models in nonprofit organizations, it also details that there is much untapped potential and possibilities for organizations to leverage their own techniques for fundraising (Nault et al., 2020).

2.2 *A comparative study of machine learning models for fundraising success*

Within the realm of fundraising, many studies and applications have been developed for different uses to help fundraising efforts. Some of these studies are focused on reviewing applications and their efficiency in live production use cases, while other studies focus on research conducted for the purpose of identifying the viability of machine learning models within the nonprofit setting.

One such example of a viability study is detailed by, Nabar (2020) who explained how a collection of Machine Learning techniques including Gradient Boosting, XGB, KNN, SVM, Naïve Bayes, Logistic Regression, and Decision Trees, were used to predict future donation amounts and donor upgradation success rates (Nabar, 2020). With the models detailed in this study, the author explains how they were able to make predictions with a high accuracy rate, which validates the author's hypothesis that by using a donor database, an actionable model could be developed for the aid of fundraising (Nabar, 2020). Although this study is focused on donation amounts and donor upgradation, it validates that accurate machine learning models can be developed for fundraising predictions in a wide scope.

2.3 *Using machine learning to help nonprofits with fundraising activities.*

Another viability study centered around donation prediction is detailed by George (2021). In this study, George used historical donation data in multiple machine learning models in an attempt to predict future donation amounts. The author focused primarily on how location information can be used in tandem with historical donation records to predict a future donation amount pledged by an individual donor. Insights from this exercise shows that an individual's personal information can be very valuable as a marker for future donations and suggests for future research to be conducted using a wider example of personal and demographic data.

2.4 *Predicting fundraising performance in medical crowdfunding campaigns using machine learning.*

Studies based on real world production models also share important insights as to possible areas for future research. In one example of study detailing past experience, Peng et al. (2021) explained how they used web scraping crawling techniques to monitor people's engagement with the COVID-19 pandemic news on social media, with the goal of predicting how likely someone would be to give to an organization that combated the pandemic (Peng et al., 2021). This study was unique in that it did not use data that a nonprofit organization already had, but instead they were able to scrape for new data from external sources to help with fundraising. This viability of this approach may be unique, as the nature of the COVID-19 pandemic was unique in the amount of attention the issue collected, but it shows the value of possessing external data sources and shows how concern in a social cause can drive fundraising interest for organizations by training predictive models.

2.5 *Optimal targeting in fundraising: A causal machine-learning approach.*

Lastly, in another study focused on data science applications for fundraising, Cagala et al. (2021) detailed the development of a machine learning model for fundraising but also explored the relationships uncovered by the model. The model was used to optimize spending for fundraising campaigns and finds that different individuals have differing reasons as to why they give to a charitable cause. The researchers found that when engaging with an individual for giving, that the fundraising organizations would need to appeal to the specific motivation for that individual, and that sending an individual an appeal with messaging that does not appeal to their motivation for the cause, will most often lead to no response by the recipient. This study shows that by developing machine learning models for nonprofit fundraising purposes, we can uncover relationships and learnings that extend beyond the use of the specific model, but also for actionable insights that can drive important decisions and impacts in the future (Cagala et al., 2021).

3 Data Exploration and Data Cleaning

For many nonprofit organizations, volunteer and donor data are highly sensitive and safeguarded. For that reason, it proves to be difficult to find real examples of nonprofit volunteer or donor lists for public research. However, the UCI Machine Learning Repository makes available a U.S. census dataset which contains much of the same information that is collected by nonprofits in their volunteer databases and can be used for the study of this topic (Kohavi & Becker, 1996).

The census income dataset contains a wide variety of information describing the personal characteristics and income of over 40,000 individuals. As we are focused on building a solution that will be used by nonprofit's using their volunteer databases, we cut down the feature set to only include features that a nonprofit could realistically collect and store. We chose to be highly selective in which features those would be, to features that a nonprofit likely already stores or could easily ascertain from their contacts lists.

3.1 Dropped Variables

Examples of features that were dropped from the census dataset included features that were likely to have high correlation with other features. These include "education-num" which was similar to "education," and "relationship" which was similar to "marital-status." Other features were ones that a nonprofit organization would not have a justification to record or store such as capital-gain and capital-loss, which describe how much money an individual made from investments. Lastly, we removed a feature which described the race of an individual. We felt that collecting race information to be used in our model would lead to biases that we could not properly account and ethically should not be a factor in determining donor classification.

Furthermore, the binary "income" feature in this dataset was chosen as a proxy for whether an individual would likely be a high-level or midlevel donor. If an individual had an income of less than $50k as denoted in the income field, that would be considered a midlevel donor, and if an individual had an income greater than $50k as denoted by the income field, that would be considered a high-level donor.

3.2 Bias, Ethics, and Privacy

To ensure individuals have a right to privacy and to avoid any issues as it comes to bias, almost all personal identification information was removed from the census dataset before being used for analysis. No identifiers such as names or addresses were included in this analysis, and as previously mentioned, race information was also excluded from the analysis in order to avoid any inadvertent biases into the model. With the remaining personal information that will be included for research, it is highly improbable that an individual can be personally identified.

3.3 Data Gaps

The one major gap in these data, which could likely help make this solution extremely accurate, was the unavailability of an individual's previous donation amounts. As intuition would suggest that an individual's future donation amounts are likely to be highly correlated with their past donations, the inclusion of previous donations as a feature would have likely been extremely insightful but alas was not available. Nevertheless, by not including previous donation amounts, it will help us uncover what are the general characteristics that make an individual more likely to be a high-level donor and can prove to be advantageous to identify new donors who have never monetarily contributed previously.

3.4 Initial Data Cleaning

Table 1

Null Values

```
age               0
workclass         0
education         0
marital-status    0
sex               0
native-country    0
income            0
dtype: int64
```

Table 2

Rows with '?' Value

```
workclass         2799
education            0
marital-status       0
sex                  0
native-country     857
income               0
dtype: int64
```

The dataset was first read into a Python 3 environment, and then examined to impute any missing values in the dataset. As shown in **Table 1**, using Python's built in functions to check for null values misleadingly shows that there are no null values in the dataset, but further exploration, as shown in **Table 2**, uncovered those missing values in this dataset were actually denoted by a '?' character in place of actual missing values.

As the missing values were all for categorical features that did not have a clear logic to how they should be filled, and as the number of records with a missing value was relatively small, those records were removed for this research.

3.5 Exploratory Data Analysis

To learn more about this dataset, each of the features included were examined individually for univariate analysis, and with combinations of different features for multivariate analysis.

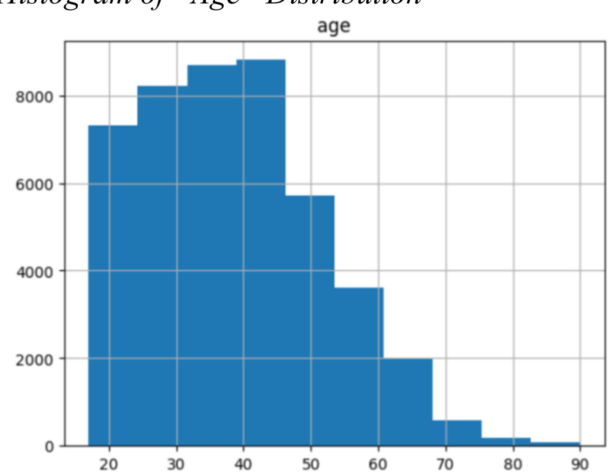
Table 3

Description of "Age"

```
count    45232.000000
mean        38.543819
std         13.219436
min         17.000000
25%         28.000000
50%         37.000000
75%         47.000000
max         90.000000
Name: age, dtype: float64
```

Figure 1

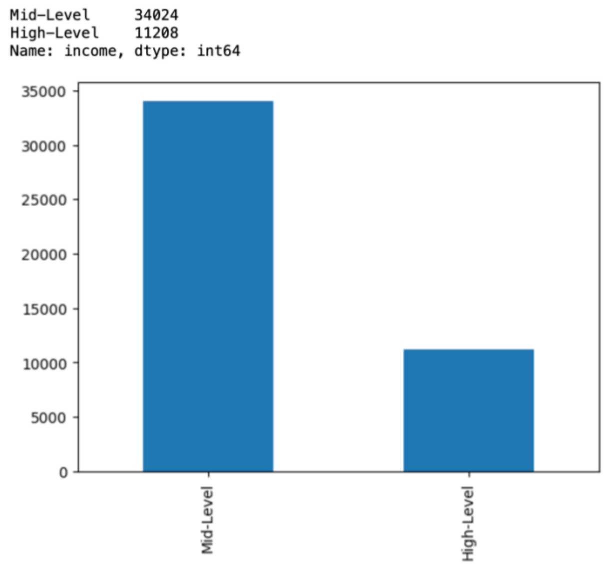
Histogram of "Age" Distribution



Within the dataset used for this research, age is the only numerical feature. It was found that the mean age of individuals in this dataset was 39 years old, with an age range of 17–90 years old. Also discovered was that 50% of individuals were between the ages of 17–37 years old, with the other 50% between the ages of 37–90 years old. Furthermore, **Figure 1** demonstrates that there is a dramatic right skew in the age of individuals.

Figure 2

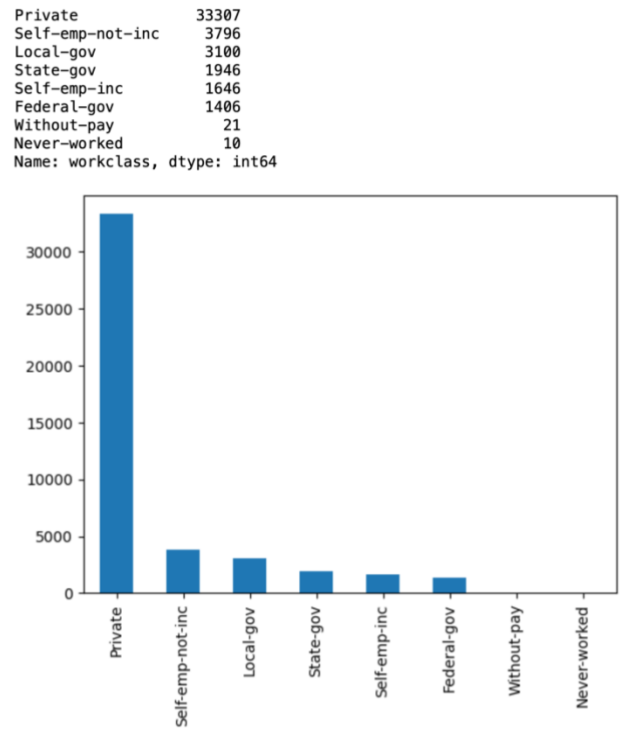
Histogram of "Income" Distribution



Following, the distribution of the "income" feature was observed. As illustrated in **Figure 2**, we found that this feature had a high class imbalance with 75% of records belonging to the midlevel class, and only 25% for the high-level class. As this feature will be the target variable and focus of this research, this class imbalance was noted to be considered during modeling to ensure that the models perform with high precision and recall.

Figure 3

Histogram of "Workclass" Distribution

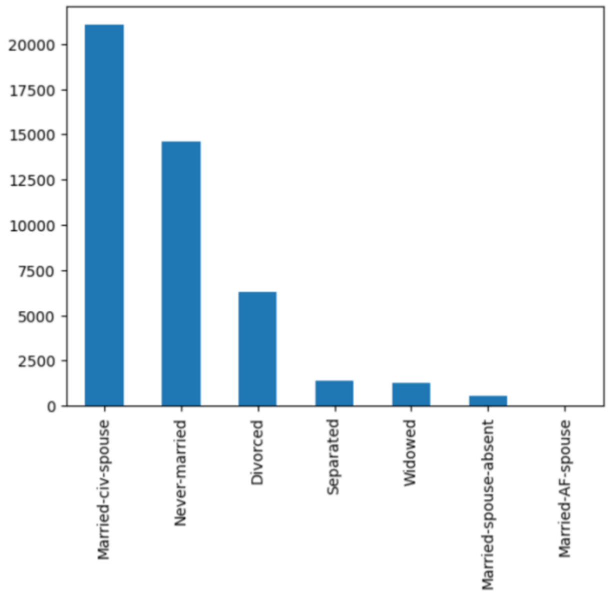


Next, the distribution of the "workclass" feature as pictured in **Figure 3** uncovered another class imbalance with about 74% of individuals working in the private sector. Also noted was that many of the workclass categories were different types of government positions, or categories of individuals who do not work or make any money, which again are useful insights to consider for data preprocessing.

Figure 4

Histogram of "Marital Status" Distribution

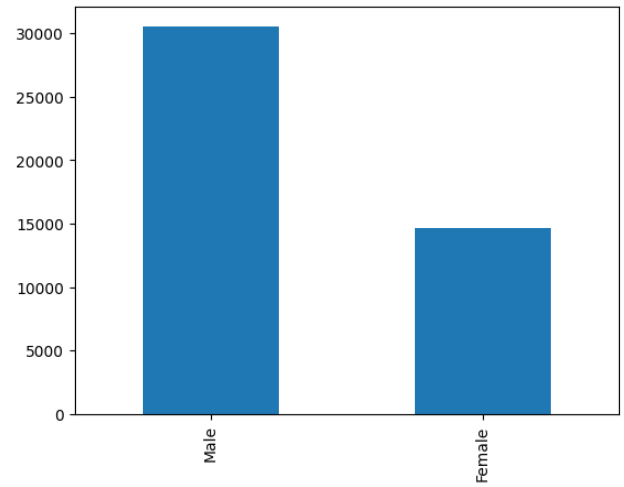
```
Married-civ-spouse        21056
Never-married             14605
Divorced                   6298
Separated                  1411
Widowed                    1277
Married-spouse-absent       553
Married-AF-spouse            32
Name: marital-status, dtype: int64
```



Further, when looking into "marital status" similar to the "working class" feature, **Figure 4** shows that "marital status" appears to have many categories that can be combined. As currently categorized, married with a civilian spouse seems to be the largest class.

Figure 5

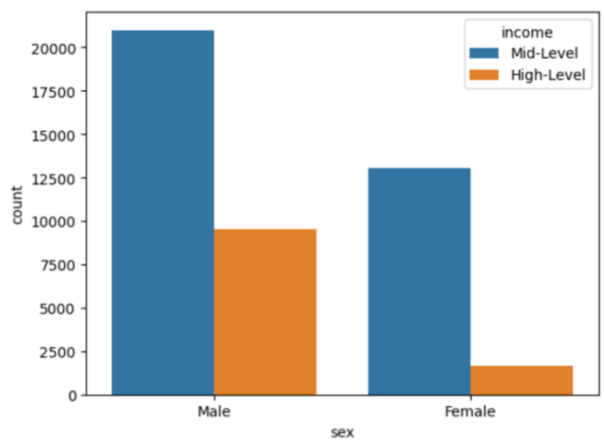
Histogram of "Sex" Distribution



Consequently, when looking into the distribution of individuals as denoted by "sex," **Figure 5** shows 68% of individuals in this dataset are male, which shows another class imbalance as it comes to the sex of individuals in this dataset.

Figure 6

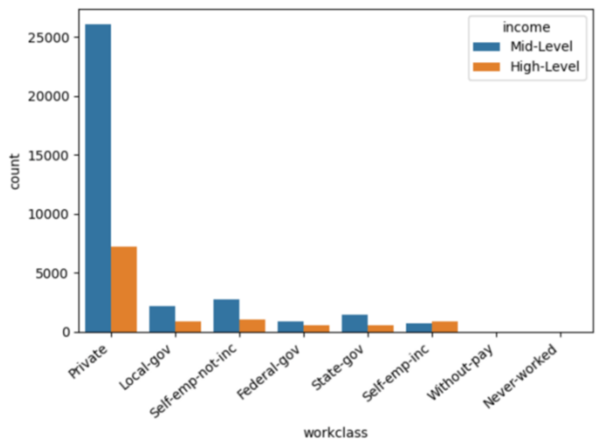
Histograms displaying "Income" by "Sex"



When looking into relationships of different features together for the purposes of bivariate analysis, **Figure 6** shows there appears to be a larger difference in the ratio of potential high-level to midlevel donors for females than there are for males.

Figure 7

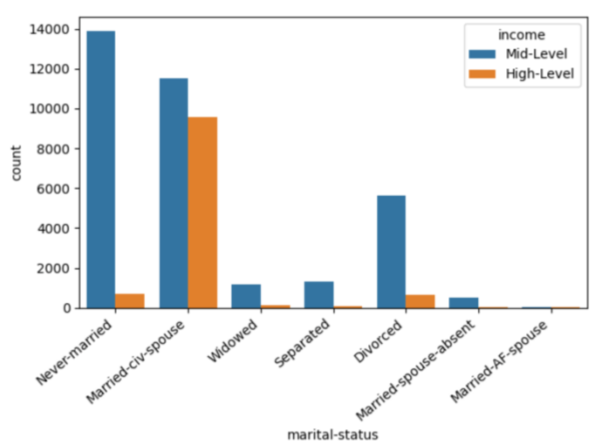
Histogram displaying "Income" by "Workclass"



When looking into the relationship of "income" with "workclass," **Figure 7** interestingly shows the self-employed work class is the only "workclass" where there are more potential high-level donors than there are midlevel.

Figure 8

Histogram displaying "Income" by "Marital Status"



Lastly, when observing the relationships between "income" and "marital status," **Figure 8** shows the married with a civilian spouse marital status class has almost as many potential high-level as midlevel donors. The other classes have dramatically higher disparities.

3.6 Data Wrangling and Cleaning

Table 4

Counts of "Marital Status" Values

Married-civ-spouse	21056
Never-married	14605
Divorced	6298
Separated	1411
Widowed	1277
Married-spouse-absent	553
Married-AF-spouse	32
Name: marital-status, dtype: int64	

Table 5

Table Describing Counts of "Marital Status" After Data Processing

Married	21641
Never-married	14605
Divorced	6298
Separated	1411
Widowed	1277
Name: marital-status, dtype: int64	

Table 6

Table Describing Counts of "Marital Status" After Further Consolidation of Values

Single	23591
Married	21641
Name: marital-status, dtype: int64	

As mentioned in the Exploratory Data Analysis section, it was observed that there were many marital status values that could be consolidated. To complete the consolidation, the values that denoted an individual as "married" were combined into a single value, as shown in **Table 5**. From there, values were further consolidated by combining categories that showed someone currently was not married into a new value labeled "single." This leaves the marital status feature to only have two remaining categories, as shown in **Table 6**.

Table 7

Counts of "Workclass" Values

Private	33307
Self-emp-not-inc	3796
Local-gov	3100
State-gov	1946
Self-emp-inc	1646
Federal-gov	1406
Without-pay	21
Never-worked	10
Name: workclass, dtype: int64	

Table 8

Counts of "Workclass" Values After Data Processing

Private	33307
Government	6452
Self-emp	5442
No-income	31
Name: workclass, dtype: int64	

Following the processing of the "marital status" values, the "workclass" values were consolidated. For this feature, we chose to group together all of the values that described someone as having some type of government job and combined the values of individuals having no income. This again simplified the values dramatically as reflected in the differences between **Table 7** showing the values before processing, and **Table 8,** which shows the values after processing.

Table 9

Counts of "Native Country" After Data Processing

United-States	41302
Non-US	3930
Name: native-country, dtype: int64	

Although there was a large class imbalance in "native country" with over 91% of individuals being from the United States, there was a long list of countries where a small number of individuals were from. To consolidate this list, we combined all the countries outside of the United States into a single value described as "Non-U.S." as shown in **Table 9.**

Table 10

Counts of "Education"

HS-grad	14785
Some-college	9901
Bachelors	7570
Masters	2514
Assoc-voc	1959
11th	1622
Assoc-acdm	1507
10th	1225
7th-8th	824
Prof-school	785
9th	676
12th	577
Doctorate	544
5th-6th	449
1st-4th	222
Preschool	72
Name: education, dtype: int64	

Table 11

Counts of "Education" After Data Processing

HS	24686
Bachelors	7570
DNF HS	5667
Assoc	3466
Masters	2514
Prof-school	785
Doctorate	544
Name: education, dtype: int64	

As displayed in **Table 10**, Education originally had very many different levels of schooling listed, which was very noisy for the purposes of analysis. To rectify this, we put together any education values into a select few categories. Most of the effort was in grouping together those that did not finish high school into a category labeled DNF HS, creating a new value for individuals with an associate degree, and combining those with small college degree with those who only finished High School. The remaining list of values was dramatically shortened as shown in **Table 11**.

3.7 Train/Test Split

For the last step of the data preparation, the dataset was split into a train and test datasets. For the split, 20% of the records were held out to be used for testing and the other 80% for training. The data was partitioned stratification based on income, which is the desired target variable for modeling, and which has a 75/25 class imbalance split.

4 Modeling

To build our desired solution for donor selection, we trained and tuned several classification models that would be able to consider characteristics of a given individual and predict whether they are likely to be a high-level or midlevel donor. The models were trained using differing classification techniques that include logistic regression, decision trees, random forest, naïve bayes, and K nearest neighbor. Each of these models were trained using the partitioned training data which consists of 80% of the original dataset. The models were then evaluated with the remaining 20% hold out test data. Furthermore, each of these models had different hyperparameters that were evaluated and tuned with different user selected values, ultimately to better fit the data.

4.1 Baseline Model

To evaluate if the created models had significantly valuable results, a baseline model was created for evaluation purposes using the DummyClassifier function. This baseline model made predictions without evaluating the input data and would be a valuable baseline to later compare against to show the trained models performed more effectively than a random guess.

4.2 Model Building and Tuning

During the model building process, the grid search method was leveraged to find the best values for each hyperparameter for each example of the selected classification techniques. Each model's accuracy with training data was recorded for each permutation of a model and its selected hyperparameters. The model that had the highest accuracy for each of the five techniques, was then saved and recorded to later be compared against the best performing model of the other classification techniques.

Table 12

Hyperparameter Tuning for Logistic Regression

```
Best: 0.810695 using {'penalty': 'l2', 'solver': 'liblinear'}
0.810667 (0.004762) with: {'penalty': 'l2', 'solver': 'newton-cg'}
0.810557 (0.004630) with: {'penalty': 'l2', 'solver': 'lbfgs'}
0.810695 (0.004661) with: {'penalty': 'l2', 'solver': 'liblinear'}
```

The first classification technique to be tested was logistic regression. For this classification technique, the hyperparameters that were chosen to be evaluated included ***solver***, which is the specific algorithm used by the model for fitting, and ***penalty,*** which is a given penalty type to be included for calculations. The different solvers used for the comparison included ***liblinear***, ***newton-cg***, and ***lbfgs*** but the only penalty that was used was ***l2***, because the other possible penalty types were not compatible with using different types of ***solver***. After the models were trained using the grid search, as shown in **Table 12**, the best performing model for logistic regression used an ***l2*** penalty with the ***liblinear*** solver.

Table 13

Hyperparameter Tuning for Decision Tree

```
Best: 0.816416 using {'criterion': 'gini', 'max_depth': 10, 'min_samples_split': 10}
0.801879 (0.006852) with: {'criterion': 'gini', 'max_depth': 5, 'min_samples_split':
0.801879 (0.006852) with: {'criterion': 'gini', 'max_depth': 5, 'min_samples_split':
0.801879 (0.006852) with: {'criterion': 'gini', 'max_depth': 5, 'min_samples_split':
0.816029 (0.005709) with: {'criterion': 'gini', 'max_depth': 10, 'min_samples_split'
0.816084 (0.005596) with: {'criterion': 'gini', 'max_depth': 10, 'min_samples_split'
0.816416 (0.005678) with: {'criterion': 'gini', 'max_depth': 10, 'min_samples_split'
0.800691 (0.004496) with: {'criterion': 'gini', 'max_depth': 20, 'min_samples_split'
0.801327 (0.004252) with: {'criterion': 'gini', 'max_depth': 20, 'min_samples_split'
0.803427 (0.004112) with: {'criterion': 'gini', 'max_depth': 20, 'min_samples_split'
0.801548 (0.006060) with: {'criterion': 'entropy', 'max_depth': 5, 'min_samples_spli
0.801548 (0.006060) with: {'criterion': 'entropy', 'max_depth': 5, 'min_samples_spli
0.801548 (0.006060) with: {'criterion': 'entropy', 'max_depth': 5, 'min_samples_spli
0.815144 (0.005825) with: {'criterion': 'entropy', 'max_depth': 10, 'min_samples_spl
0.815144 (0.005729) with: {'criterion': 'entropy', 'max_depth': 10, 'min_samples_spl
0.815476 (0.006117) with: {'criterion': 'entropy', 'max_depth': 10, 'min_samples_spl
0.800415 (0.004118) with: {'criterion': 'entropy', 'max_depth': 20, 'min_samples_spl
0.800636 (0.004240) with: {'criterion': 'entropy', 'max_depth': 20, 'min_samples_spl
0.803068 (0.003579) with: {'criterion': 'entropy', 'max_depth': 20, 'min_samples_spl
```

The decision tree classification technique was then tuned for the best selection of hyperparameters. The decision tree technique had three different hyperparameters that were used for tuning which included ***criterion***, which is the function used to measure the quality of a tree split, ***max depth***, which is a maximum depth constraint for a decision tree, and ***min split*** which is the minimum number of splits a decision tree should make. After tuning, **Table 13** shows the best performing decision tree includes the gini ***criterion*** with a ***max depth*** of 10 and with 10 ***minimum splits***.

Table 14

Hyperparameter Tuning for Random Forest

```
Best: 0.805444 using {'criterion': 'entropy', 'max_features': None, 'n_estimators': 100}
0.804477 (0.004542) with: {'criterion': 'gini', 'max_features': 'sqrt', 'n_estimators': 100}
0.803980 (0.004732) with: {'criterion': 'gini', 'max_features': 'sqrt', 'n_estimators': 200}
0.804809 (0.004681) with: {'criterion': 'gini', 'max_features': 'sqrt', 'n_estimators': 300}
0.804477 (0.004542) with: {'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 100}
0.803980 (0.004732) with: {'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 200}
0.804809 (0.004681) with: {'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 300}
0.805251 (0.005233) with: {'criterion': 'gini', 'max_features': None, 'n_estimators': 100}
0.804974 (0.005427) with: {'criterion': 'gini', 'max_features': None, 'n_estimators': 200}
0.805361 (0.005311) with: {'criterion': 'gini', 'max_features': None, 'n_estimators': 300}
0.804394 (0.004867) with: {'criterion': 'entropy', 'max_features': 'sqrt', 'n_estimators': 100}
0.804118 (0.004944) with: {'criterion': 'entropy', 'max_features': 'sqrt', 'n_estimators': 200}
0.804698 (0.004755) with: {'criterion': 'entropy', 'max_features': 'sqrt', 'n_estimators': 300}
0.804394 (0.004867) with: {'criterion': 'entropy', 'max_features': 'log2', 'n_estimators': 100}
0.804118 (0.004944) with: {'criterion': 'entropy', 'max_features': 'log2', 'n_estimators': 200}
0.804698 (0.004755) with: {'criterion': 'entropy', 'max_features': 'log2', 'n_estimators': 300}
0.805444 (0.005023) with: {'criterion': 'entropy', 'max_features': None, 'n_estimators': 100}
0.805057 (0.005274) with: {'criterion': 'entropy', 'max_features': None, 'n_estimators': 200}
0.805417 (0.005041) with: {'criterion': 'entropy', 'max_features': None, 'n_estimators': 300}
```

The next model technique observed for tuning was Random Forest. For this technique, we evaluated models based on ***criterion***, which like decision trees, was the criterion used to decide a tree split, the number of ***n estimators***, which is the number of trees in the forest, and lastly ***max features***, which details the number of features that should be considered when looking for the best split. The results of tuning showed that the model that performed best used entropy as the ***criterion***, with no ***max features***, and 100 ***n-estimators***.

Table 15

Hyperparameter Tuning for Naïve Bayes

```
Best: 0.781677 using {'var_smoothing': 1e-08}
0.781677 (0.016651) with: {'var_smoothing': 1e-08}
0.773525 (0.032578) with: {'var_smoothing': 1e-09}
0.764046 (0.051395) with: {'var_smoothing': 1e-10}
```

The next model technique to be evaluated was for models using naïve bayes. The only hyperparameter that was used for tuning was ***var smoothing*** which details the portion of the largest variance for all features. The default value for this parameter is ***1e-9,*** but tuning comparisons were performed with ***1e-8, 1e-9,*** and ***1e-10.*** Ultimately, the best performing naïve bayes model was with a ***var smoothing*** of ***1e-8.***

Table 16

Hyperparameter Tuning for K Nearest Neighbor

```
Best: 0.800470 using {'algorithm': 'brute', 'n_neighbors': 7, 'weights': 'uniform'}
0.791350 (0.005238) with: {'algorithm': 'auto', 'n_neighbors': 5, 'weights': 'uniform'}
0.787205 (0.005470) with: {'algorithm': 'auto', 'n_neighbors': 5, 'weights': 'distance'}
0.795330 (0.004791) with: {'algorithm': 'auto', 'n_neighbors': 7, 'weights': 'uniform'}
0.791239 (0.005332) with: {'algorithm': 'auto', 'n_neighbors': 7, 'weights': 'distance'}
0.792151 (0.005363) with: {'algorithm': 'ball_tree', 'n_neighbors': 5, 'weights': 'uniform'}
0.787426 (0.005518) with: {'algorithm': 'ball_tree', 'n_neighbors': 5, 'weights': 'distance'}
0.796711 (0.004952) with: {'algorithm': 'ball_tree', 'n_neighbors': 7, 'weights': 'uniform'}
0.791129 (0.005387) with: {'algorithm': 'ball_tree', 'n_neighbors': 7, 'weights': 'distance'}
0.791350 (0.005238) with: {'algorithm': 'kd_tree', 'n_neighbors': 5, 'weights': 'uniform'}
0.787205 (0.005470) with: {'algorithm': 'kd_tree', 'n_neighbors': 5, 'weights': 'distance'}
0.795330 (0.004791) with: {'algorithm': 'kd_tree', 'n_neighbors': 7, 'weights': 'uniform'}
0.791239 (0.005332) with: {'algorithm': 'kd_tree', 'n_neighbors': 7, 'weights': 'distance'}
0.795302 (0.005657) with: {'algorithm': 'brute', 'n_neighbors': 5, 'weights': 'uniform'}
0.790493 (0.006228) with: {'algorithm': 'brute', 'n_neighbors': 5, 'weights': 'distance'}
0.800470 (0.005866) with: {'algorithm': 'brute', 'n_neighbors': 7, 'weights': 'uniform'}
0.795661 (0.006068) with: {'algorithm': 'brute', 'n_neighbors': 7, 'weights': 'distance'}
```

The last model technique to undergo model tuning was K nearest neighbor. For this technique, three hyperparameters were compared which included ***n_neighbors***, which is the number of neighbors used by the technique to measure similarity, ***weights*** which is the given weight function used for prediction, and **algorithm** which is the algorithm used to compute the nearest neighbors. Ultimately, the model that performed best was one that used the brute ***algorithm***, with 7 ***n neighbors***, and uniform ***weights***.

4.3 Model Outcomes and Deployment Strategy

After the model building and tuning with the five different classification techniques, the best performing models of each technique were then evaluated and compared against each other using a selection performance metrics that include recall, F1, accuracy, precision, and kappa, in that order of significance. The trained models would not be used in parallel or interact with one another, but instead the best performing model given the evaluation was chosen to be the model of choice to be deployed and to be our ultimate deliverable.

5 Results and Findings

A selection of five standard classification performance metrics, which include recall, F1, accuracy, precision, and kappa were used to evaluate the performance of each of the trained classification models. The models chosen for evaluation were the best performing model of each of the five classification techniques that were leveraged and tuned, and each model was evaluated on predictions calculated by each model for the hold out test data. As the donor classification problem must account for the large class imbalance between high-level and midlevel donors, recall is the most important measure of performance, as misclassifying a high-level donor will lead to a large loss in fundraising opportunity. From there, naturally it is important to score high in F1, as F1 is the harmonic mean of recall with precision and another important marker of model accuracy. Furthermore, the metrics of precision, accuracy, and kappa were also reviewed to see if there were any other big differences between performance of the models.

5.1 Evaluation Results

Table 17

Hyperparameter Tuning for K Nearest Neighbor

Model Performance

	Accuracy	Precision	Recall	f1	Kappa
Baseline	0.62	0.24	0.24	0.24	−0.01
Logit	0.80	0.67	0.42	0.52	0.40
Decision Tree	0.81	0.69	0.45	0.54	0.43
Random Forest	0.81	0.64	0.48	0.55	0.43
Naive Bayes	0.79	0.59	0.50	0.54	0.41
K Nearest Neighbor	0.80	0.62	0.48	0.54	0.42

After calculating the performance metrics of each of the five models, we chose the naïve bayes model to be our model of choice given that it had the highest score for recall and the second highest score for F1. All five of the models performed similarly well in terms of accuracy, precision, and kappa, and so recall became the deciding factor to select the model of choice as it is valuable to be as accurate as possible in identifying the minority class of high-level donors.

The recall of value 0.5 for the naïve bayes model explains that the model can correctly identify 50% of the high-level donors within the dataset. Although 50% shows that there are areas for opportunity, it is important to keep in mind that only 25% of individuals in this dataset are high-level donors due to the large class imbalance in the data.

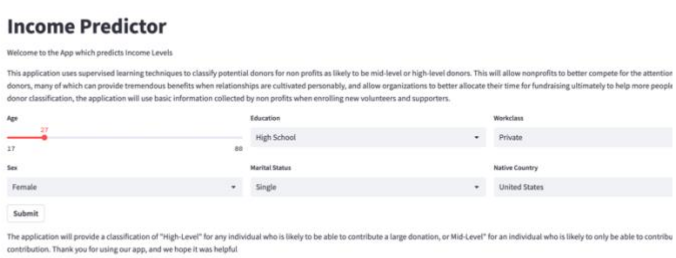
Furthermore, the performance of each model as compared to the baseline model shows that the trained models perform significantly more effectively than a random guess and so the input data is significant in being able to correctly distinguish and predict high-level and midlevel donors.

Therefore, by using this classification model, our original hypothesis is validated, that nonprofit organizations will be able to identify untapped high-level donors for lucrative fundraising opportunities by reaching out to the individuals identified as high-level donors, by this model.

5.2 Model Deployment

Figure 9

Screenshot of Deployed Application on steamlit.io



Once the naïve bayes model was identified to be the best classification technique for this given problem, we deployed our trained model using streamlight.io. The deployed model can take input from a user and then predict whether that individual is likely to be a high-level or midlevel donor. Like the feature set used for model building, the deployed model will make its decisions based on the inputs of age, marital status, sex, workclass, native country, and education.

6 Results and Discussion

The results of our application prove our original hypothesis that it is indeed viable for nonprofits to identify high-level donors using classification techniques and the basic personal information that they collect from their volunteers. The results of our model and application prove that it is possible to successfully distinguish potential high-level and midlevel donors for the majority of potential donors, which will directly help nonprofits in strategizing how to best solicit donations from those individuals.

6.1 Discussion

These results are consistent with other research performed in this area such as detailed by George (2021) and Nabar (2020), which indicated machine learning techniques are proficient in being able to helpfully identify individuals to aid in fundraising opportunities. We hope that the results of this research project, in conjunction with the success of other projects within the realm of nonprofit fundraising will establish trust for nonprofits to partner with data scientists to accelerate their fundraising efforts.

Though our model was successful in accurately identifying most of the individuals we tested, there is area of opportunity for the model to be more accurate in predicting individuals from the high-level donor specifically, as almost half of those donors were miscategorized. As it relates to the results of this study, the result is because of the class imbalance where midlevel donors are far more represented in the dataset than high-level donors, but can hopefully be rectified if the model can access data of more individuals for model training.

6.2 Conclusion

Given the results of this research project, we have found there is a tremendous availability of high-level donor fundraising opportunities for nonprofit organizations using classification applications. We also found that by using our solution, we were able to better identify who are individuals that are likely to donate large contributions if correctly cultivated which would provide lucrative opportunities for nonprofits looking to increase their fundraising. Furthermore, we also found there are tremendous insights available from leveraging even the simplest personal information that is already available to many organizations as is.

6.3 Recommendation and Next Steps

To further develop classification applications for donor classification, and to improve their ability to correctly classify donors, we recommend further research be done in partnership with individual nonprofit organizations, so research can leverage historic donor records for further model training. By demonstrating the efficacy of this application as is, we believe we will be able to gain the trust of organizations to have access to their data for training, and in turn be used in a live setting to guide their donor cultivation strategy.

Acknowledgments

We would like to thank Professor Mikael Graindorge (University of San Diego, Masters Applied Data Science) for supervising this research project and providing guidance for this capstone. Furthermore, we would like to acknowledge the University of San Diego Masters Applied Data Science program for encouraging this research focus, and for Dr. Ebrahim Tarshizi for encouraging practical research projects including this capstone project.

References

Cagala, T., Glogowsky, U., Rincke, J., & Strittmatter, A. (2021, September 30). *Optimal targeting in fundraising: A causal machine-learning approach.* CESifo Working Paper No. 9037. CESifo Network. https://arxiv.org/pdf/2103.10251.pdf

George, M. (2021, April 28). Using machine learning to help nonprofits with fundraising activities. *AWS Public Sector Blog*. https://aws.amazon.com/blogs/publicsector/using-machine-learning-nonprofits-fundraising-activities/

Georgiev, D. (2023, January 26). *22 eye-opening volunteering statistics to know in 2023*. Techjury. https://techjury.net/blog/volunteering-statistics/#gref

Kohavi, R., & Becker, B (1996). *UCI machine learning repository*. University of California, School of Information and Computer Science. Retrieved March 5, 2023 from https://archive.ics.uci.edu/ml/datasets/census+income

Nabar, R. (2020, August). *A comparative study of machine learning models for fundraising success*. Dublin Business School.

https://esource.dbs.ie/bitstream/handle/10788/4229/msc_nabar_r_2020.pdf?sequence=1&isAllowed=y

Nault, K., Ruhi, U., & Livvarcin, O. (2020). Exploring the applications & challenges of data analytics in nonprofit organizations. https://aisel.aisnet.org/amcis2020/social_inclusion/social_inclusion/8

Osula, B., & Ng, E. (2014). Toward a collaborative, transformative model of nonprofit leadership: Some conceptual building blocks. *Administrative Sciences*, *4*(2), 87–104. https://doi.org/10.3390/admsci4020087

Peng, N., Zhou, X., Niu, B., & Feng, Y. (2021, January 11). Predicting fundraising performance in medical crowdfunding campaigns using machine learning. *Electronics*, *10*(2), 143. https://doi.org/10.3390/electronics10020143