**Prospective Donors Prediction Model**

we selected the following 14 variables for model building：

1. Affiliation Percent
2. Age（less than 65）
3. Total Score without Giving
4. Grand Total Affiliation Score
5. Communication (binary: 0 or 1)
6. **EGC Score**

(Rating Code and Rating:   
For Rating Code:  
It has Z, Y, P… F, E, D, which correspond to   
Rating:   
Note Rated, Insufficient Data, <$2,500, …, $50M - $74.99M.  
As Z means Not Rated, we rewrite ‘#N/A’ to be Z and Not Rated. And to be easy to calculate for model, we set up a **new column called EGC Score** (estimated Giving Capacity): 0,1,2,3,4,5,6,7,8,9,10,11,12,13.   
Because Z and Y (Not Rated, Insufficient Data) has same meaning for prediction, we give them ‘0’ in EGC Score.  
Then, 1 corresponds to <$2,500, 2 corresponds to $2,500 - $9,999, and so on, until 13 corresponds to $50M - $74.99M.)

1. Whether in NTEE (binary: 0 or 1)
2. **Total Campus Score**

(UI Degree Campus: create a new column named Total Campus Score:   
= sum (UI Degree Campus),   
U: 1,   
S or C: -1,   
0: 0.   
U means Urbana campus so that who always study in Urbana campus of business may more likely donate money to Gies.)

1. Experiential (binary: 0 or 1)
2. Philanthropic (binary: 0 or 1)
3. Trust Association (binary: 0 or 1)
4. Volunteer (binary: 0 or 1)
5. Affiliation Rank
6. Small Business Owner (binary: 0 or 1)

**Prediction Logic:**

The goal of this prediction task was to identify whether alumni who did **not donate in the previous year** **would donate in the following year**—a binary classification problem (0/1). This approach aims to help identify and target potential new donors.

We used data from three time periods: **2020–2021, 2021–2022, and 2022–2023**. Considering concerns about the **i.i.d. (independent and identically distributed)** assumption, we did **not** combine the datasets across years. Instead, we built **three separate machine learning models**, one for each period.

As the number of people who did not donate in the previous year but donated in the following year is small, and the people who did not donate in the following year are too many, we need to balance it.

**20-21**,

1: 393

0: 500 (randomize)

Test set: 20%

**21-22**:

1: 414

0: 500 (randomize)

Test set: 20%

**22-23**:

1: 451

0: 500 (randomize)

We experimented with **logistic regression**, a **simple artificial neural network (ANN)** (limited by computing power), and **random forest**. Among these, the **random forest model** consistently performed the best and was selected for final use.

For example, there are the parameters of the ANN we tried:

| **Layer** | **Description** |
| --- | --- |
| **Input Layer** | 64 neurons, ReLU activation; input dimension equals the number of features |
| **Dropout** | 30% dropout to reduce overfitting |
| **Hidden Layer** | 32 neurons, ReLU activation |
| **Dropout** | Another 30% dropout |
| **Output Layer** | 2 neurons, softmax activation for binary classification (class 0 and class 1) using one-hot encoding |

Table 1: Layer Settings (ANN)

| **Parameter** | **Value** | **Description** |
| --- | --- | --- |
| Optimizer | adam | Adaptive optimizer, widely used and effective |
| Loss Function | categorical\_crossentropy | Suitable for one-hot encoded binary classification |
| Metric | accuracy | To monitor model performance |
| Epochs | 100 | Maximum training iterations |
| Batch Size | 32 | Number of samples per training batch |
| Validation Split | 0.2 | 20% of the training set used as validation data |
| EarlyStopping | Patience = 10 | Stops training if validation loss doesn't improve for 10 consecutive epochs |
| Data Split | train\_test\_split | Stratified to maintain class balance |

Table 2: Layer Parameters (ANN)

Using the same variables, it shows that **RF is better in this case**. Each of the three random forest models achieved approximately **77% accuracy** on the test set.

The Tree depths from the 3 RF models are from 14~30, and the distribution of each is like normalization distribution. Most of the depths are near 18~20. After we tried depth = 10, 15 or the modes of the depth from each model itself to tune the models, they always show obvious bad performance in the test sets. Therefore, we just remain the natural structure from original training. Sometimes, it shows better but most important problem is that the recall became too low when we just set up a certain tree depth for every tree.

Based on the accuracy metrics below, the model consistently performs better at predicting class ‘0’ than class ‘1’, as indicated by the higher **precision** and **F1-score** for class ‘0’. This suggests the model is more effective at identifying alumni who are unlikely to donate in the following year.

A screenshot of a graph

AI-generated content may be incorrect.

Figure 1: Accuracy of 20-21 RF Model

A screenshot of a computer

AI-generated content may be incorrect.

Figure 2: Accuracy of 21-22 RF Model

A screenshot of a graph

AI-generated content may be incorrect.

Figure 3: Accuracy of 22-23 RF Model

For the Importance Plots of RF model, from the 3 years conditions, the top 6 most important feature among the 14 features are

**Affiliation Percent**;

**Age**;

**Total Score without Giving**;

**Grand Total Affil Score**;

**Communication**;

**Experiential**.

A graph with blue bars

AI-generated content may be incorrect.

Figure 4: Feature Importance for 20–21 RF Model

A graph with blue bars

AI-generated content may be incorrect.

Figure 5: Feature Importance for 21–22 RF Model

A graph with blue bars

AI-generated content may be incorrect.

Figure 6: Feature Importance for 22–23 RF Model

After standardization, the total importance of the 14 variables sums to 1. As shown in the table below, **Affiliation Percent** contributes the most to the prediction with an importance of approximately 0.22, followed by **Age** (0.17) and **Total Score without Giving** (0.16), and so on in descending order.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Feature | FY21 Importance | FY22 Importance | FY23 Importance | Average Importance |
| 1 | Affiliation Percent | 0.21 | 0.21 | 0.23 | 0.22 |
| 2 | Age | 0.2 | 0.2 | 0.12 | 0.17 |
| 3 | Total Score without Giving | 0.18 | 0.15 | 0.16 | 0.16 |
| 4 | Grand Total Affil Score | 0.14 | 0.13 | 0.13 | 0.13 |
| 5 | Communication | 0.06 | 0.09 | 0.06 | 0.07 |
| 6 | Experiential | 0.05 | 0.05 | 0.07 | 0.06 |

Table 3: Feature Importance of RF Model

However, although the three models performed well on the training sets, we further evaluated their practical usability by testing them on a new dataset derived from part of the 2023–2024 alumni records. The 2024 data is completely new and was not involved in any model training. The models were trained and validated on the 2020–2021, 2021–2022, and 2022–2023 datasets.

The overall prediction performance on the new data is summarized as follows:

A screenshot of a test

AI-generated content may be incorrect.

Figure 7: Outcome of 2024 Data

At first glance, the overall accuracy of the results above may appear average. However, a closer look reveals that the **lower accuracy is primarily driven by misclassifications of class 0** — alumni who did not donate. When focusing specifically on class 1 — those who did donate — the models demonstrate strong predictive performance. This distinction is important, as the primary goal of this task is not to identify the majority who are unlikely to donate, but rather to accurately **identify the smaller group of alumni with a high likelihood of making donation**.

Next, we examine the performance metrics specifically for class 1 predictions. This evaluation uses the 2024 data without any further model tuning or retraining. The fact that the models achieved reasonably good prediction results **on the first attempt** highlights their potential practicality and usability in real-world applications.

A screenshot of a computer program

AI-generated content may be incorrect.

Figure 8: Performance Metrics of Class 1

You may notice that the recall is relatively low, which leads to a lower F1-score. However, our goal is **not to capture every alumnus with a low probability** of donating. Instead, we aim to ensure that when the model predicts someone is likely to donate, **they truly have a high chance of donating**—especially after we follow up with a phone call.

Narrowing down the list of potential donors allows us to be more targeted and efficient, as we don't have the time or resources to engage in deep conversations with every alumnus. Therefore, to improve precision, we adjusted the model threshold to 60%, meaning only those with a predicted donation probability above 60% will be marked as likely donors.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 9: Performance Metrics with 60% Threshold

As we can see, the precision is relatively high, which means that **if the model predicts an alumnus is likely to donate, there is approximately an 80% chance** that they will actually donate.

Finally, we applied the trained models and developed a simple interactive website using a 60% threshold. Users only need to **upload an Excel file containing 14 feature columns along with the alumni ID**. The app will then predict the donation probability and **append two new columns: the average predicted probability and a binary prediction** indicating whether the likelihood is ≥60%. ([https://alumni-donation-app.onrender.com](https://alumni-donation-app.onrender.com/), if nobody use the website for 15 minutes, it will be inactive, and when you open it which needs 1 minute to restart)

A screenshot of a computer program

AI-generated content may be incorrect.

Figure 10: Interactive Website