I. Introduction

While non-profits are mission-driven rather than profit-driven, they still have opportunities to generate revenue, expand impact, and improve sustainability. Direct mail marketing is one of the most common fundraising strategies where donation requests are sent to potential donors. However, mailing to every individual is not cost efficient and profitable, not all the mails sent would be able to collect donations and the expected profit from each mailing is negative. Taking advantage of data analytics, I develop both predictive classification and regression models that can effectively target the right individuals to maximize fundraising outcomes, and improve cost-effectiveness.

Business Opportunities

- **Increasing donation size:** Use historical donation patterns to personalize suggested donation amounts
- **Use donation predictions to improve financial planning:** Identify potential funding gaps and adjust strategies in real time.
- **Identify non-donors:** identifies unresponsive donors to limit and avoid sending mails to them, saving on postage and material costs.

Predicting Donors

- What are the key characteristics of individuals who are likely to donate in response to a mailing campaign?
- Which demographic, financial, or behavioral features are the most significant predictors?
- How accurate is the model to predict the potential donors?

Predicting Donation Amounts

Which are the common features of high value donations?

Among individuals who donate, what factors influence the donation amount?

Which predictive model provides the best estimate of donation amounts while minimizing errors?

By answering these questions, I aim to develop a data-driven strategy that enhances fundraising efficiency, strengthens donor engagement, and ensures long-term financial sustainability for the nonprofit.

II. Data Understanding/Exploratory Data Analysis (EDA)

Data dictionary:

Label	**Data type**	**Meaning**
id number	ID	Unique identification number for each person
region	nominal	Five geographic regions including ter1, ter2, ter3, ter4, ter5
ownd	binary	(1 = homeowner, 0 = not a homeowner)
kids	interval	Number of children
inc	ordinal	Household income (7 categories)
sex	binary	Gender (0 = Male, 1 = Female)
wlth	ordinal	Wealth Rating (Wealth rating uses median family income and population statistics from each area to index relative wealth within each state. The segments are denoted 0-9, with 9 being the highest wealth group and 0 being the lowest.)
hv	interval	Average Home Value in potential donor's neighborhood in $ thousands
incmed	interval	Median Family Income in potential donor's neighborhood in $ thousands
incavg	interval	Average Family Income in potential donor's neighborhood in $ thousands

low	interval	Percent categorized as “low income” in potential donor's neighborhood
npro	interval	Lifetime number of promotions received to date
gifdol	interval	Dollar amount of lifetime gifts to date
gifl	interval	Dollar amount of largest gift to date
gifr	interval	Dollar amount of most recent gift
mdon	interval	Number of months since last donation
lag	interval	Number of months between first and second gift
gifa	interval	Average dollar amount of gifts to date
donr	binary	Classification Response Variable (1 = Donor, 0 = Non-donor)
damt	interval	Prediction Response Variable (Donation Amount in $).

*Note that the donr and damt variables are set to missing for the score data.

Target Variables

The dataset contains **two target variables for different types of model:**

- **Classification**: Predicting **donr** (whether someone donates).
- **Regression**: Predicting **damt** (amount donated in $).

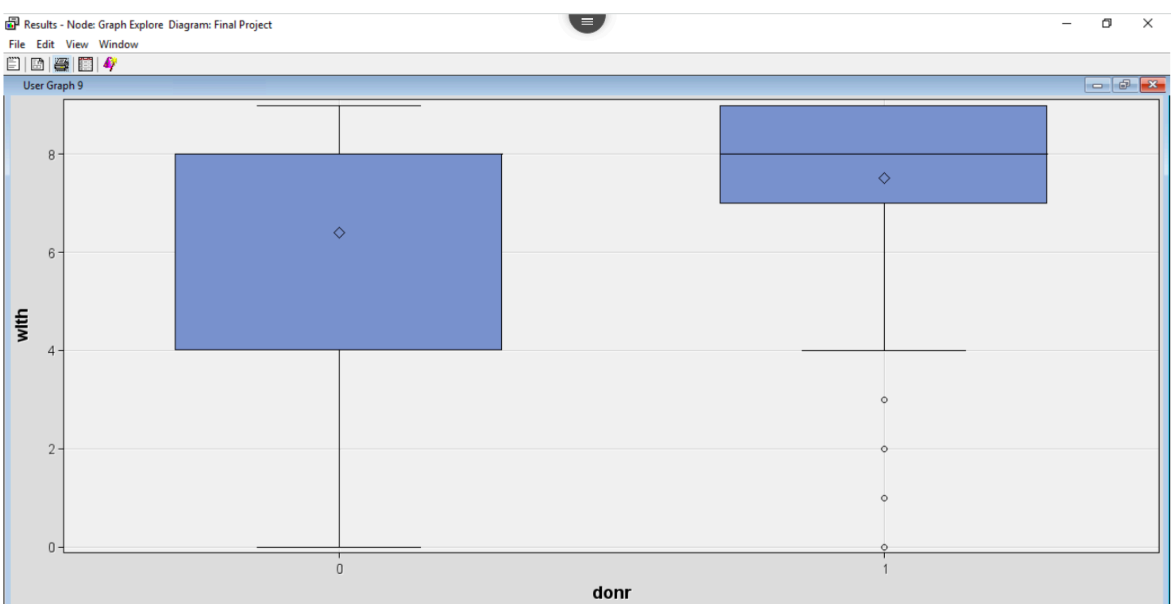
Missing Values: There is no missing data.

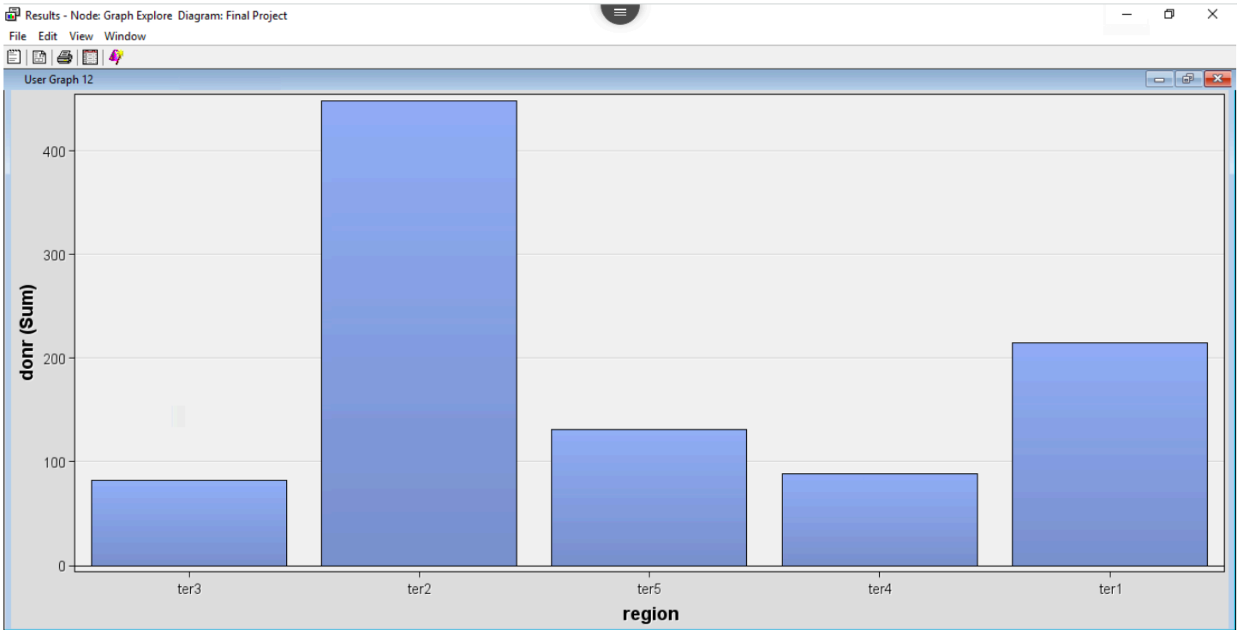
Class Variable Summary Statistics

The dataset includes several categorical variables. For Sex, there are 3,648 female donors and 2,354 male donors. For Homeownership, 5,309 people own a home, and 693 don’t. The Region variable has five categories: ter1 (1,209 people), ter2 (2,083), ter3 (728), ter4 (795), and ter5 (1,187). Donor status is almost evenly split—2,994 are donors and 3,008 aren’t. As for Income, there are seven categories: category 1 has 333 people, category 2 has 706, category 3 has 604, category 4 is the biggest with 2,754, then category 5 has 888, category 6 has 390, and category 7 has 327. Lastly, Wealth is broken into ten categories, ranging from 152 people in category 0 to 1,626 in category 9. Category 8 has the most with 2,314.

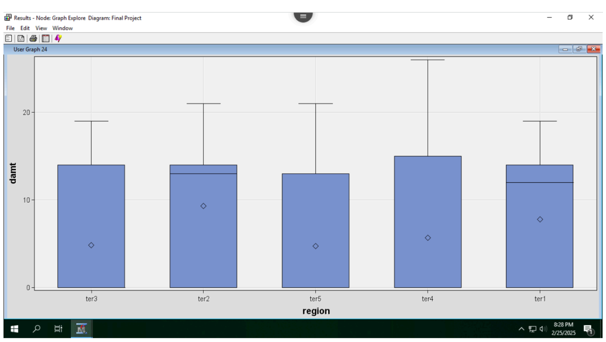
Examine Potential Predictors

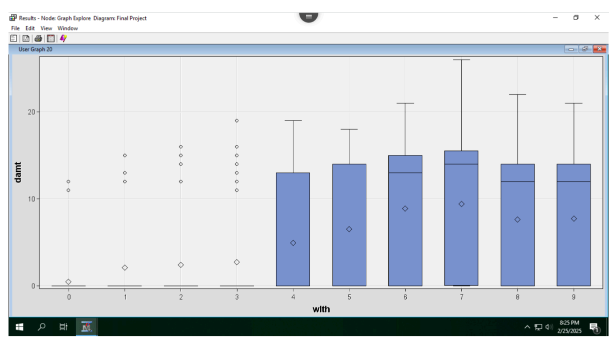
Based on the initial observations, several factors may influence whether someone is a donor or not. The number of kids could play a role in donation behavior, as financial responsibilities might impact a person's ability to give. Additionally, average income could be a significant factor, as higher earnings may increase donation potential. Wealth level also appears to have a strong influence, with donors predominantly coming from higher wealth groups. Region may be another contributing factor, as most donors seem to come from ter2. Furthermore, there is a noticeable imbalance in donor distribution by demographics—there are significantly more female donors compared to male donors, and a large portion of donors are homeowners. However, some of these groups, such as females, homeowners, and higher-wealth individuals, have a larger sample size, which may influence the observed patterns. This imbalance could impact the prediction model, potentially skewing results toward the dominant groups.

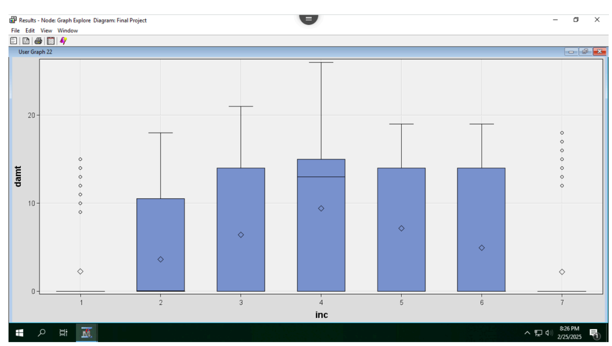




For classification, based on the variable correlation in the clustering diagram, kids appear to have a significant negative correlation with damt. Additionally, when conducting box plots for other categorical variables on damt, the box plots illustrate that wlth and region could also be strong predictors of donation amount.







Based on the initial analysis, several variables appear to be potentially irrelevant for predicting donor classification and donation amount. The percentage of individuals categorized as low may not provide significant additional predictive power beyond individual wealth and income variables. Similarly, npro may not have a direct relationship with donation behavior. Donation-related variables such as gifdol, gifl, gifr, and gifa may not contribute much to predicting whether someone is a donor but could be more relevant for estimating donation amounts among existing donors. Additionally, time-based variables like mdon and lag may have limited influence on donor classification, though they could offer insights into donor retention. While these variables initially seem less relevant based on exploratory analysis, further statistical testing, such as feature importance in modeling, is necessary to confirm whether they should be excluded or retained.

III. Data Preparation

I used Pandas to conduct data wrangling and preprocessing. I checked for duplicates and missing values—none were found. I also used the IQR method to detect outliers and confirmed that there were no outliers in the dataset. To ensure sufficient data for training and evaluation, I used the Data Partition node to split the dataset: 70% for training and 30% for validation to evaluate model generalization and reduce overfitting risk. Since the dataset contains values with different units (e.g., thousands, single digits, decimals), I used the Transform Variables node to standardize all interval variables on the same scale, except for the damt variable. Input variables were kept consistent across models, and all interval variables were range standardized to support proper feature scaling.

IV. Modeling

Classification Models

For the KNN classification models, I set ‘donr’ as a target variable and 'damt' as a reject variable.

1. K-Nearest Neighbor (KNN) Model

   After partitioning the data input, transform variables are necessary to standardize all the variables to the same scale. Using 13 MBR nodes to calculate the distance between points, as AS automatically includes that point as one of the neighbors, increase k value by 1 in the MBR nodes, then compare all the MBR nodes on misclassification rate to find the best k value.

2. Decision Tree Model

   Two models are built for comparison: one with 6 branches and a depth of 20, and another with 2 branches and a depth of 6. During training, interactive node splitting is enabled to allow manual adjustments. The model prioritizes more important variables for splitting—specifically, 'inc' (income) and 'ownd' (home ownership). After all nodes are split, a control point node is added to streamline the process flow, enabling multiple paths to execute simultaneously and simplifying model structure. The final step involves comparing both models on their misclassification rate to evaluate performance.

3. Neural Network Models

   To evaluate performance across different architectures, I developed three neural networks with varying complexities. All models were trained using the backpropagation technique. The specific configurations included: Model 1 with 3 hidden units and a learning rate of 0.1, Model 2 with 5 hidden units and a learning rate of 0.01, and Model 3 with 16 hidden units, direct connections, and a learning rate of 0.1. Each model was allowed a maximum run time of 30 minutes for training.

In addition to manually configured models, I explored the use of SAS AutoNeural to automate architecture and iteration tuning. Two Artificial Neural models were tested: AutoNeural 1 with 15 iterations and 3 hidden layers, and AutoNeural 2 with 20 iterations and 2 hidden layers. The final step is comparing all models on their misclassification rate to evaluate performance.

Regression Models

For the regression models, I set ‘damt’ as a target variable and 'donr' as a reject variable.

1. KNN

The architecture of the KNN Regression model is identical to that of the classification model. Variable transformations node is applied to standardize all variables to the same scale. The optimal *k* value is then determined by comparing the average squared error (ASE) across all MBR nodes.

2. Decision Tree Model

The architecture of the decision tree regression model mirrors that of the classification model. All settings remain unchanged. Two models are utilized for comparison: a 2-branch, 6-depth tree and a 6-branch, 20-depth tree. The default assessment measure, “Decision,” is applied to both. Once the control point node is run, it is linked to the Compare Models node to generate the final evaluation results based on ASE.

3. Neural Network

All modeling settings were retained from the previous neural network setup, including the backpropagation training method, a maximum training time of 30 minutes per model, and the default activation functions provided by SAS. Since the task involved predicting a continuous outcome, the model selection criterion was adjusted from misclassification rate to average error, which better reflects prediction accuracy in regression modeling. Three regression-based neural network models were manually configured: Model 1 included 3 hidden units with a learning rate of 0.1, Model 2 used 5 hidden units with a learning rate of 0.01, and Model 3 had 16 hidden units with direct connections and a learning rate of 0.1. In addition to these manually specified networks, two AutoNeural models were tested under regression settings. AutoNeural 1 ran for 15 iterations with 3 hidden layers, while AutoNeural 2 completed 20 iterations with 2 hidden layers.

V. Evaluation

The average donation is $14.50 and it costs $2.00 to produce and send.

Classification

To evaluate classification models with donr as our target variable, I set Selection Statistics to Misclassification Rate for the Model Comparison node.

1. KNN

The best k value is k = 10 according to the output shown below, with a misclassification rate of 0.24928. This indicates that about 25% of the classification made by the model would be incorrect, this could significantly impact profitability depending on how errors are distributed (see Figure 1).

Confusion matrix:

	FN	**TN**	**FP**	**TP**	**Misclassification Rate**
K = 1	246	659	242	652	0.30636
K = 2	234	616	285	664	0.26806

K = 3	331	711	190	567	0.27240
K = 4	238	623	278	660	0.26373
K = 5	289	690	211	609	0.27168
K = 6	229	632	269	669	0.26590
K = 7	274	677	224	624	0.25650
K = 8	226	627	274	672	0.25867
K = 9	258	674	227	640	0.25795
K = 10	**233**	**633**	**268**	**665**	**0.24928**
K = 11	260	664	237	638	0.25578
K = 12	218	639	262	680	0.25361
K =13	246	659	242	652	0.25723

Profits:

	Total Predicted Positive (TPP)	Expense in $ (TPP x $2.00)	Revenue in $ (TP x $14.50)	Total Profit in $ (Rev - Expense)
K = 1	894	1788	9454	7666
K = 2	949	1898	9628	7730
K = 3	757	1150	8221.5	7071.5
K = 4	938	1876	9570	7694
K = 5	820	1640	8830.5	7190.5
K = 6	938	1876	9700.5	7824.5
K = 7	848	1696	9048	7352
K = 8	946	1892	9744	7852
K = 9	867	1734	9280	7546
K = 10	933	1866	9642.5	7776.5
K = 11	875	1750	9251	7501
K = 12	**942**	**1884**	**9860**	**7976**

K = 13	894	1788	9454	7666

Among the compared models, K = 12 model demonstrates the greatest profitability.

2. Decision Tree

The results indicate that the 6 branch 20 depth model performs better and is therefore selected as the preferred model for this project (see Figure 2).

Confusion matrix:

	FN	**TN**	**FP**	**TP**	**Misclassification Rate**
6 branch 20 depth	**130**	**798**	**105**	**769**	**0.13041**
2 branch 6 depth	101	741	162	798	0.14595

Profits:

	Total Predicted Positive (TPP)	Expense in $ (TPP x $2.00)	Revenue in $ (TP x $14.50)	Total Profit in $ (Rev - Expense)
6 branch 20 depth	874	1748	11150.5	9402.5
2 branch 6 depth	**960**	**1920**	**11571**	**9651**

The 2-branch, 6-depth model yields higher profits compared to the alternative model.

3. ANN

Based on the Model Comparison node's result, the auto neural network with 20 iterations has the lowest misclassification rate on validation data of 0.09989 (see Figure 3).

Confusion matrix:

	FN	**TN**	**FP**	**TP**	**Misclassification Rate**
Neural Network 3	80	770	133	819	0.11820
Neural Network 16	84	779	124	815	0.11543
AutoNeural 20	**83**	**806**	**97**	**816**	**0.09989**
Neural Network 5	100	792	111	799	0.11709
AutoNeural 15	80	799	104	819	0.10211

Based on the validation data AutoNeural 15 has the highest profit compared to the rest of the neural network models.

	Total Predicted Positive (TPP)	Expense in $ (TPP x $2.00)	Revenue in $ (TP x $14.50)	Total Profit in $ (Rev - Expense)
Neural Network 3	952	1,904	11,875.5	9,971.5
Neural Network 16	939	1,878	11,817.5	9,939.5
AutoNeural 20	913	1,826	11,832.0	10,006.0
Neural Network 5	910	1,820	11,585.5	9,762.5
AutoNeural 15	**923**	**1,846**	**11,875.5**	**10,029.5**

Regression

To evaluate regression models with damt as our target variable, I set Selection Statistics to ASE.

1. KNN

Among the K values, K = 13 shows the best fitting model according to the ASE criterion with ASE for validation dataset equals 32.01032 (see Figure 4).

2. Decision Trees

The final result indicates that 6 branch 20 depth is the selected model with ASE equals 23.12256 (see Figure 5).

3. ANN

Based on the result, the auto neural network with 15 iterations also has the lowest average squared error on validation data of 17.8185 (see Figure 6).

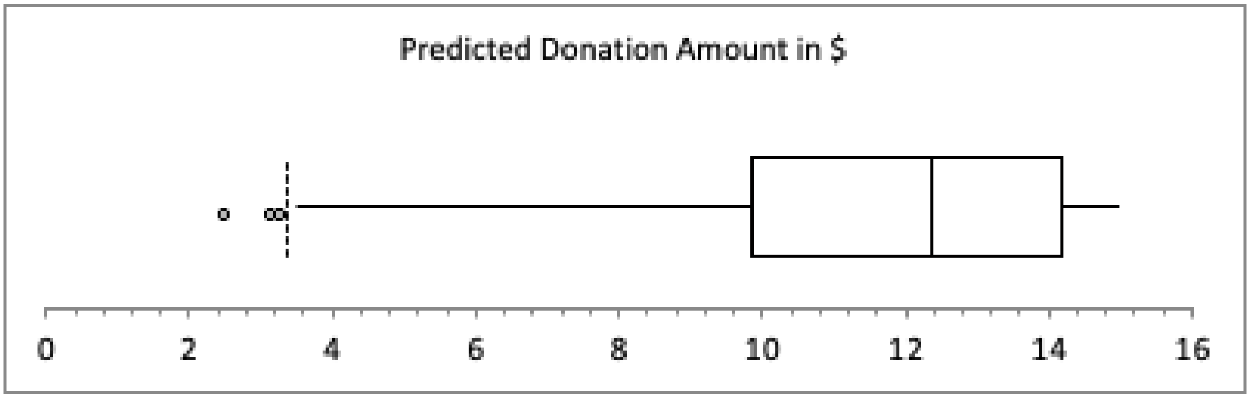
VI. Deployment

To improve the effectiveness and profitability of fundraising efforts, I implemented a data-driven strategy that targets only individuals most likely to donate and contribute more than the cost of mailing. Specifically, I applied the best-performing models—Artificial Neural Network (with 20 iterations) for classification and Artificial Neural Regression (with 15 iterations) for donation amount prediction—to a score dataset. These models were chosen based on their performance, with the classification model achieving a misclassification rate of 9.99% and the regression model recording the lowest average squared error of 17.8185.

To improve the cost-effectiveness of the nonprofit's direct mail campaigns, I deployed predictive classification and regression models on a score dataset of 2,007 prior donors. According to the organization's historical records, only 10% of individuals typically respond to mailings, with an average donation of $14.50. Given a mailing cost of $2.00 per person, the expected profit from mailing everyone is negative: (10% × $14.50) – $2.00 = –$0.55 per person, or a total loss of $1,104 when mailing to all 2,007 recipients.

To address this inefficiency, I implemented a model-based strategy that filtered individuals using two criteria: (1) they are predicted to be donors (donr = 1), and (2) their expected donation (damt) is greater than the $2.00 mailing cost. This resulted in 408 selected individuals from the score dataset. The sum of their predicted donation amounts totaled $4,757. However, to account for the model's 9.99% misclassification rate, I estimate that approximately 41 individuals are false positives—predicted to donate but unlikely to do so.

To conservatively estimate revenue, I accounted for the wasted mailing cost to these 41 individuals (41 × $2 = $82) and subtracted their predicted contributions using the median predicted donation value ($12.36), which was chosen due to outliers and non-normal distribution (see Figure 7). This resulted in an estimated misclassification cost of $506.76, bringing the total adjustment to $588.76. After deducting this and the mailing cost of $816 (408 × $2), the net profit from the targeted campaign is approximately $3,434.24.



This approach highlights the value of predictive analytics in nonprofit fundraising. By narrowing the target audience to those most likely to give—and to give more than the cost of contact—the organization significantly increases profitability while reducing waste. The table below summarizes the financial comparison between traditional and model-driven strategies:

Result from Score Data:

Criteria	Number of Individuals	Expense in $ ($2 each)	Expected Revenue in $	Net Profit (Loss) in $
Mass Mailing (all 2007)	2007	4,014	~2,910 ($10\% \times 14.50 \times 2007$)	–1,104
Final Mailing (donr=1 & damt > $2)	408	816	~4,757 (sum of predicted donation amount)	3,434.24 ($4,757 - 12.36 \times 41 - 816$)

Two .csv files have been included with this report to support implementation of the campaign, one contains the predicted donor status (donr), and the other one has predicted donation amount (damt).

VII. Conclusion

With a misclassification rate of 0.09989, meaning about 90% of the observations were classified correctly, the ANN model not only outperforms the other two in the classification model, but its regression version with 15 iteration also delivers the most accurate gift‑amount forecasts with the lowest average squared error of 17.8185. After completing the initial analysis in both Python and Excel, the result shows that the top 30 donation amounts record the consistent high donor profiles, most of them are affluent female homeowners with two to three children (wealth scores of 8–9), who tend to support and contribute to charitable causes.

Deploying best-performing models, Artificial Neural Network for donor classification and Artificial Neural Regression for gift-amount prediction, which target individuals who are likely to donate and contribute more than the cost of mailing. However, the classification model misclassified 9.99% of the observations and the regression model showed the average squared error of 17.8185, indicating there is room to improve both classification accuracy and prediction precision. Having filtered individuals by two criteria, the likelihood to donate and amount prediction, I can manage to address this inefficiency by calculating the expected revenue and net profit loss. However, the model accounts for a 9.99% misclassification rate, about 41 individuals predicted to donate but unlikely to do so. Subtracting the total mailing cost of $816 (408 x 2) and the total adjustment of $588.76, it is possible to turn from loss to net profit which is approximately $3,434.24.

A key limitation is the feature set. As certain mailings produce a net loss, the model predicts negative donation amounts. By incorporating more precise features such as actual outreach costs, and detailed demographic data, it can significantly improve prediction accuracy and target the most promising donors. In addition, resetting the activation function for the validation set will ensure all predictions to be zero or above.

Appendix

1. The misclassification rate for validation dataset of different KNN classification models.

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Number of Estimated Weights	Train: Sum of Frequencies
Y	MBR12	MBR12	MBR (12)	donr	donr	0.266815	12	2400
	MBR9	MBR9	MBR (9)	donr	donr	0.269594	12	2400
	MBR13	MBR13	MBR (13)	donr	donr	0.271262	12	2400
	MBR11	MBR11	MBR (11)	donr	donr	0.276265	12	2400
	MBR7	MBR7	MBR (7)	donr	donr	0.27682	12	2400
	MBR6	MBR6	MBR (6)	donr	donr	0.27682	12	2400
	MBR8	MBR8	MBR (8)	donr	donr	0.277932	12	2400
	MBR5	MBR5	MBR (5)	donr	donr	0.277932	12	2400
	MBR10	MBR10	MBR (10)	donr	donr	0.278488	12	2400
	MBR4	MBR4	MBR (4)	donr	donr	0.286826	12	2400
	MBR2	MBR2	MBR (2)	donr	donr	0.288494	12	2400
	MBR3	MBR3	MBR (3)	donr	donr	0.289605	12	2400
	MBR	MBR	MBR	donr	donr	0.316287	12	2400

2. The misclassification rate for validation dataset of different Decision Tree classification models.

Fit Statistics
Model Selection based on Valid: Cumulative Percent Captured Response (_VCAPC_) at Decile 10

Selected Model	Model Node	Model Description	Valid: Cumulative Percent Captured Response	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error	Valid: Misclassification Rate
Y	Tree2	6 branch 20 depth	19.4105	0.10029	0.13024	0.10637	0.13041
	Tree3	2 branch 6 depth	18.4330	0.12062	0.14738	0.11835	0.14595
	Tree	Decision Tree	18.4330	0.17863	0.25071	0.16508	0.22586

3. The misclassification rate for validation dataset of different ANN classification models.

Enterprise Miner - Homeworks - Quynh Nguyen

Results - Node: Model Comparison Diagram: Final Project

File Edit View Window

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Total Degrees of Freedom	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Train: Number of Estimated Weights	Train: Akaike's Information Criterion	Train: Schwarz's Bayesian Criterion	Train: Average Squared Error	Train: Maximum Absolute Error	Train: Divisor for ASE	Train: Sum of Frequencies
Y	AutoNeural	AutoNeural	AutoNeural ...	donr	donr	0.099889	4200	4129	71	71	2210.793	2661.135	0.076313	0.993173	8400	4200
	AutoNeural2	AutoNeural2	AutoNeural ...	donr	donr	0.102109	4200	3989	211	211	2561.121	3899.46	0.077794	0.992849	8400	4200
	Neural2	Neural2	Neural Net...	donr	donr	0.115427	4200	3606	594	594	3702.4	7470.047	0.092803	0.98608	8400	4200
	Neural3	Neural3	Neural Net...	donr	donr	0.117092	4200	4024	176	176	3025.616	4141.956	0.098171	0.99327	8400	4200
	Neural	Neural	Neural Net...	donr	donr	0.118202	4200	4094	106	106	2816.258	3488.599	0.094853	0.962622	8400	4200

4. The average squared error for validation dataset of different KNN regression models.

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error
Y	MBR26	MBR26	13	damt	damt	32.01032
	MBR24	MBR24	11	damt	damt	32.13991
	MBR25	MBR25	12	damt	damt	32.15102
	MBR23	MBR23	10	damt	damt	32.38242
	MBR22	MBR22	9	damt	damt	32.59565
	MBR21	MBR21	8	damt	damt	32.77746
	MBR20	MBR20	7	damt	damt	32.95612
	MBR19	MBR19	6	damt	damt	33.46334
	MBR18	MBR18	5	damt	damt	34.04399
	MBR17	MBR17	4	damt	damt	34.87223
	MBR16	MBR16	3	damt	damt	36.6545
	MBR15	MBR15	2	damt	damt	39.44833
	MBR14	MBR14	1	damt	damt	44.15491

5. The average squared error for validation dataset of different Decision Tree regression models.

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error	Train: Sum of Frequencies	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Train: Average Squared Error
Y	Tree5	Tree5	6 branch 20...	damt	damt	23.12256	4201	17.17308	86501.46	20.59068
	Tree6	Tree6	2 branch 6 ...	damt	damt	23.43696	4201	18.73494	92685.94	22.06283

6. The average squared error for validation dataset of different ANN regression models.

Results - Node: Model Comparison Diagram: Final Project

File Edit View Window

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target	Target Label	Selection Criterion: Valid: Average Squared Error	Train: Total Degrees of Freedom	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Train: Number of Estimated Weights	Train: Akaike's Information Criterion	Train: Schwarz's Bayesian Criterion	Train: Average Squared Error	Train: Maximum Absolute Error	Train: Divisor for ASE	Train: Sum of Frequencies
Y	AutoNeural4	AutoNeural4	AutoNeural ...	damt	damt	17.81852	4201	4129	72	72	11963.22	12419.92	16.667	17.70325	4201	4201
	AutoNeural3	AutoNeural3	AutoNeural ...	damt	damt	18.36915	4201	3814	387	387	12867.17	15321.94	17.79007	15.89577	4201	4201
	Neural5	Neural5	Neural Net...	damt	damt	20.87624	4201	3607	594	594	13996.27	17764.05	21.09131	15.61162	4201	4201
	Neural4	Neural4	Neural Net...	damt	damt	21.60162	4201	4095	106	106	13277.64	13950.01	22.42389	18.48727	4201	4201
	Neural6	Neural6	Neural Net...	damt	damt	23.27013	4201	4025	176	176	13536.94	14653.33	23.06981	16.84705	4201	4201

7. The descriptive statistics of the predicted donation amount for the score dataset.

Descriptive statistics

	#1
count	408
mean	11.659313
sample standard deviation	2.937415
sample variance	8.628407
minimum	2.48000535
maximum	14.9498628
range	12.4698574
empirical rule	
mean - 1s	8.721898
mean + 1s	14.596728
percent in interval (68.26%)	65.7%
mean - 2s	5.784483
mean + 2s	17.534143
percent in interval (95.44%)	95.1%
mean - 3s	2.847068
mean + 3s	20.471558
percent in interval (99.73%)	99.8%
skewness	-0.912221
kurtosis	0.003742
coefficient of variation (CV)	25.19%
1st quartile	9.858433
median	12.359085
3rd quartile	14.200889
interquartile range	4.342456
mode	#N/A
low extremes	0
low outliers	3
high outliers	0
high extremes	0