

April 2024

Two-Stage Direct Response Model

GWSB Business Analytics Practicum Group 8

# Client Introduction



The Nature Conservancy, a global environmental non-profit, which is advancing conservation in all 50 states and U.S. territories and in 70 countries around the world.

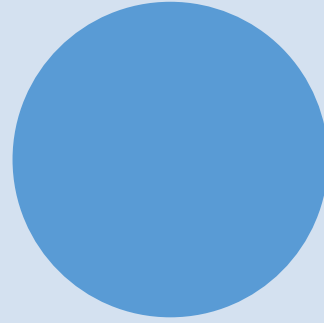
Founded in the U.S. through grassroots action in 1951, The Nature Conservancy (TNC) has grown to become one of the most effective and wide-reaching environmental organizations in the world. Thanks to more than a million members and the dedicated efforts of our diverse staff and over 400 scientists, we impact conservation in 79 countries and territories: 37 by direct conservation impact and 42 through partners.

- Mission: To conserve the lands and waters on which all life depends.
- Vision: A World where the diversity of life thrives, and people act to conserve nature for its own sake and its ability to fulfill our needs and enrich our lives.

# Executive Summary and Problem Understanding

- ❑ TNC relies on fundraising efforts to support its mission, so we need to develop a two-stage direct response model to maximize revenue efficiently.
  - ❖ Highlight the importance of maximizing net revenue for The Nature Conservancy and ensure financial sustainability.
  - ❖ The Nature Conservancy faces the challenge of optimizing its direct-mail fundraising appeals program to maximize net revenue.
  - ❖ Traditional approaches lack precision, leading to inefficiencies. Appeal to all donors risks financial losses, while targeting only the most responsive donors neglects high-value, less responsive ones.
  - ❖ Balancing responsiveness and value are critical for success.
- Our Project Objectives

## Project Objectives



Build a two-stage direct response model to predict (1) probability of donor response and (2) expected gift amount.



Maximize net revenue of direct-mail fundraising appeal campaign by identifying donors with highest expected gift potential.

# Methodology: Data Analyzed

## Raw text files

File name	Description	Data source
Target1stPrtyfile	Internally collected data reflecting past donor history, including key target variables "TGTresp" and "TGTgiftamt".	Internal
Demofile	Includes demographic information such as age, gender, geographic location; donation history and responses to marketing campaigns.	Third-Party
Transfile	Performance metrics from previous marketing campaigns, including response rates and revenue generated.	Third-Party

## Data overview

Metric	Value
Data shape	3276648 rows * 62 columns
Target variable	'TGTresp' and 'TGTgiftamt'
Number of unique donors	992714 donors
Max/min/mean of target gift amount	\$9636/\$0/\$0.88
Number of columns with missing values	9 columns
Number of categorical columns	21 columns

# Data Preprocessing

- **Merge three files**
  - Used 'masterprimaryid' as the key, where each 'masterprimaryid' corresponds to a donor.
- **Field type manipulation**
  - Converted currency values from string format to float data type to facilitate numerical calculations.
- **Handling missing values**
  - Dropped the 'append\_enviroconquintile' column due to a high proportion of missing data.
  - Removed two rows in the 'append\_HomeValue' column where data was absent.
  - Filled missing values with the mean for variables:
    - 'append\_age';
    - 'append\_mt\_OnlineInsuranceBuyer' for online insurance purchasing likelihood;
    - 'append\_mt\_LowDollarDonor' and 'append\_mt\_HighDollarDonor' for donation propensity;
    - 'append\_mt\_CultureArtsEvents' for cultural event attendance likelihood.
- **Deal with duplicate values**
  - Removed columns 'birth\_year' and 'append\_age\_indicator' as these fields contain overlapping age information.
- **Created dummy variables**
  - 'Monthly\_Donor', 'YE\_Behavior', 'Gift\_Behavior' reflect donation timing and patterns, identifying year-end donors, regular monthly contributors, and responses to varied fundraising campaigns.
  - 'First\_gift\_channel', 'MRG\_channel', 'HPG\_channel', 'Prior\_Channel\_Behavior' captured the initial, most recent, highest, and preferred donation channels.
  - 'LifeCycle', 'LifeCycleDetail', 'Donor\_status' for classifying the duration of continuous donations or the time since lapse.
- **Final data**
  - Average target donation amount \$0.884 VS average target donation amount among respondents \$34.123.
  - Overall Response Rate 2.592%.

# Feature Selection

## Phase 1

column name	Description
append_animal	Past year participants supporting animal well-being non-profits.
append_env	Number of transactions with environmental non-profits within lifetime.
append_wildlifewarriors	Past year participants in non-profits for animal health and welfare.
append_fundraisingpartners	Cumulative number of participants donating to non-profit merchandise categories in the past 12 months.
append_competitors	Transactions with competitor non-profits.
append_env_enthusiasts	Environment enthusiasts' participants in a lifetime.
append_direct_trans	Past year number of direct transactions.
YE_Behavior_A: Both YE and Non YE	Donor year-end behavior: both YE and non-YE
Gift_Behavior_A: All	Kinds of donor behavior: all types
Gift_Behavior_D: Others Only	Kinds of donor behavior: others only
LifeCycleDetail_Consecutive 5+ Years	Donors with 5+ consecutive years of giving.
LifeCycle_Consecutive Givers	Donors with consecutive giving years.
GivingYears	Number of donation years.
cumul_amount	cumulative donation amount
append_trans_life	Overall transactions lifetime.
Monthly_Donor_Y	Donors who give monthly.
Monthly_Donor_N	Non-monthly donors.
gift_count	Count of donation transactions.
append_parks_nature	Cumulative number of transactions with parks and nature non-profits.

Target Variable 1 = Response; 0 = No Response (84931 VS 3191715)

## Phase 2

Column Name	Description
HPG_amount	Highest previous donation amount.
MRG_amount	Most recent donation amount
avg_gift_amt	Average donation amount.
first_gift_amount	Amount of the first donation.
append_direct_trans	Cumulative number of direct transaction in the past 12 months.
append_fundraisingpartners	Cumulative number of participants donating to non-profit merchandise categories in the past 12 months.
append_HouseHoldEducation	Median school years completed by adults over 18 in the household.
append_WealthResources	Epsilon's newest estimate of value of savings and investments.
append_age	Advantage Individual Age reports known ages, employs algorithms for inferred exact ages.
append_web	Cumulative number of transactions in the past 12 month that made through online.
append_competitors	Transactions with competitor non-profits.
append_parks_nature	Cumulative number of transactions with parks and nature non-profits.

Target Variable: Gift amount in dollars (use value of 250 for 129 rows where TGTgiftamt > 250)

Feature Selection done by a combination of scikitlearn SelectKBest function and distribution analysis.

# Phase 1 Model Process



Feature selection for phase 1 is done via:

`SelectKBest(score_func=chi2, k=20)`

Hyper-parameter tuning is done through a grid search function.

Best model is chosen by F1 score.

The following model is trained by 70% of the data:

Random Forest

Classifier:

`class_weight={0: 1, 1: 20}`

`max_depth=16`

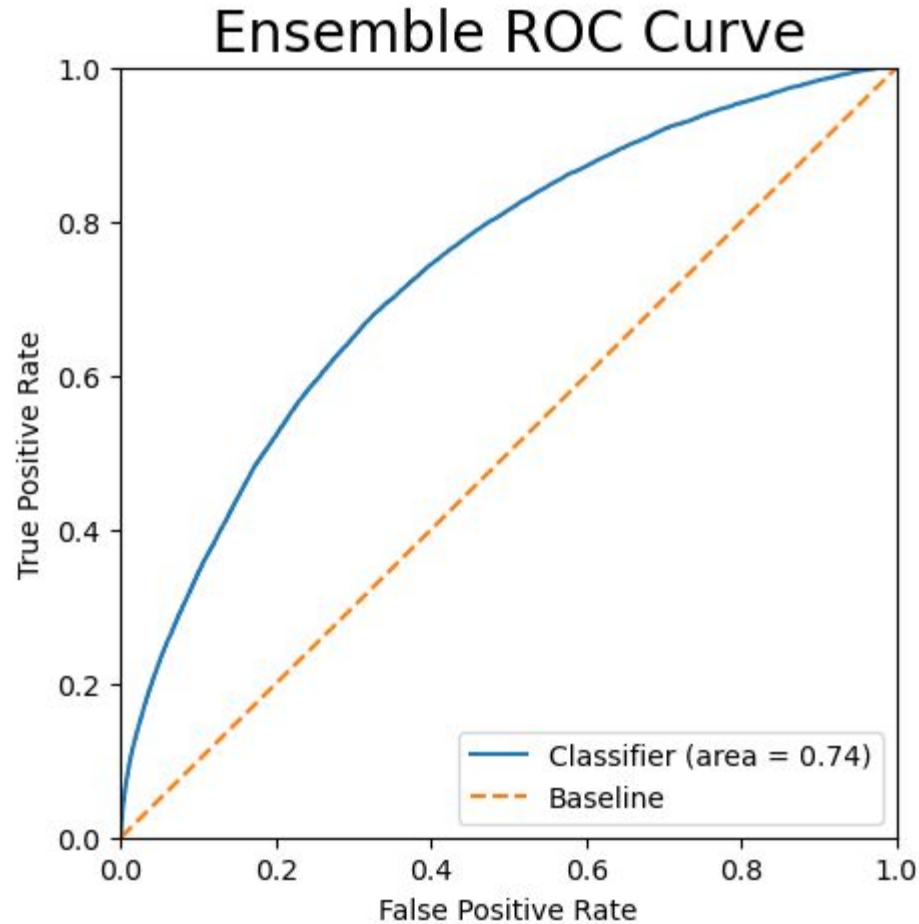
`min_samples_leaf=19`

`n_estimators=110`

30% of the data is used to test model predictions and results are recorded.



# Model 1 Performance



## Top 5 Predictor Variables

- Gift Behavior A: All
- append\_fundraisingpartners
- cumul\_amount
- append\_trans\_life
- LifeCycle\_Consecutive Givers

# Phase 2 Model Process

## Addressing Outliers

Boxplots are drawn to view the presence of outliers in the data.

A target variable threshold of 250 is set and data is changed accordingly.

## Feature Selection

Feature selection for phase 1 is done via:

SelectKBest(score\_func=chi2, k=10),

for each selected model.

## Model Training

2 models are trained with 70% of our data:

Random Forest Regressor

Decision Tree Regressor

## Remediation, Re-train

We identify and plot residuals.

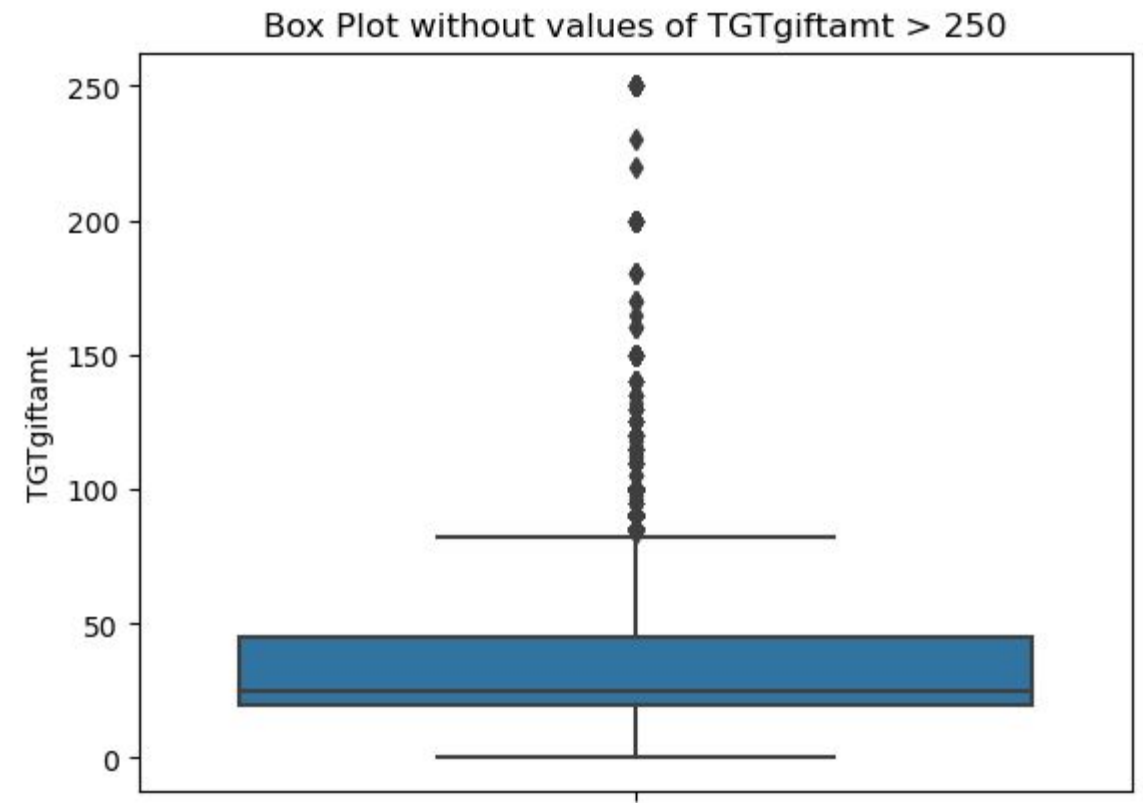
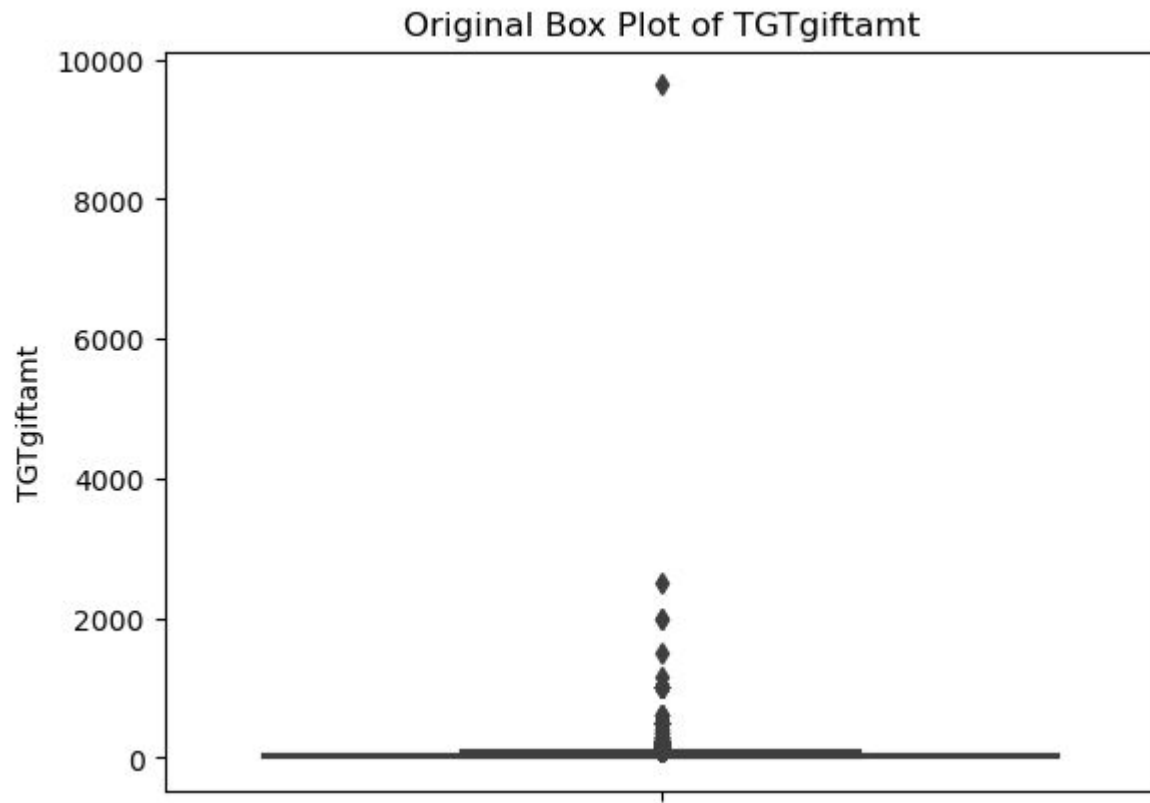
Removed residual outliers from the training data, and train the models an additional time.

## Generate Predictions

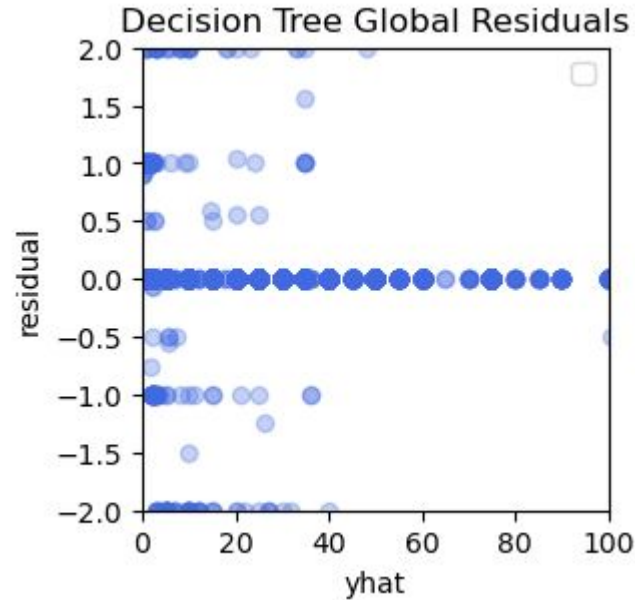
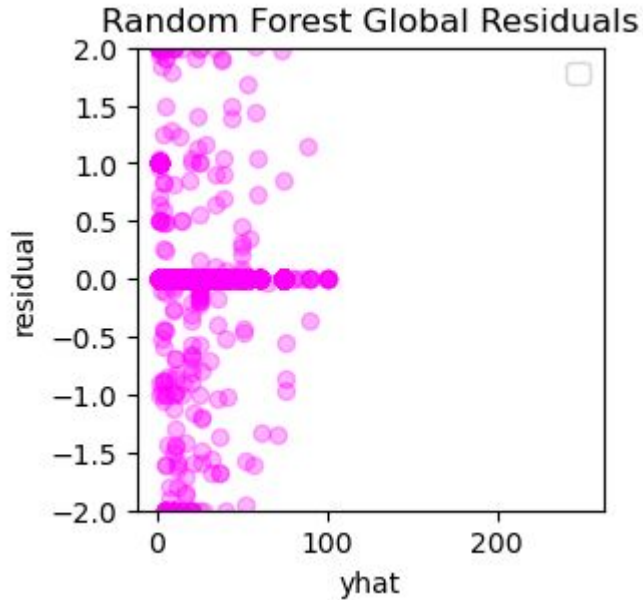
30% of the data is used to test both models predictions.

The models are ensemble by prediction result averages.

## Model 2 Outlier Visualization



# Model 2 Residual Analysis



## Effect of Remediation on MAE\*:

### Initial model MAE results:

- Random Forest: 32.72
- Decision Tree: 29.29

### Final MAE results:

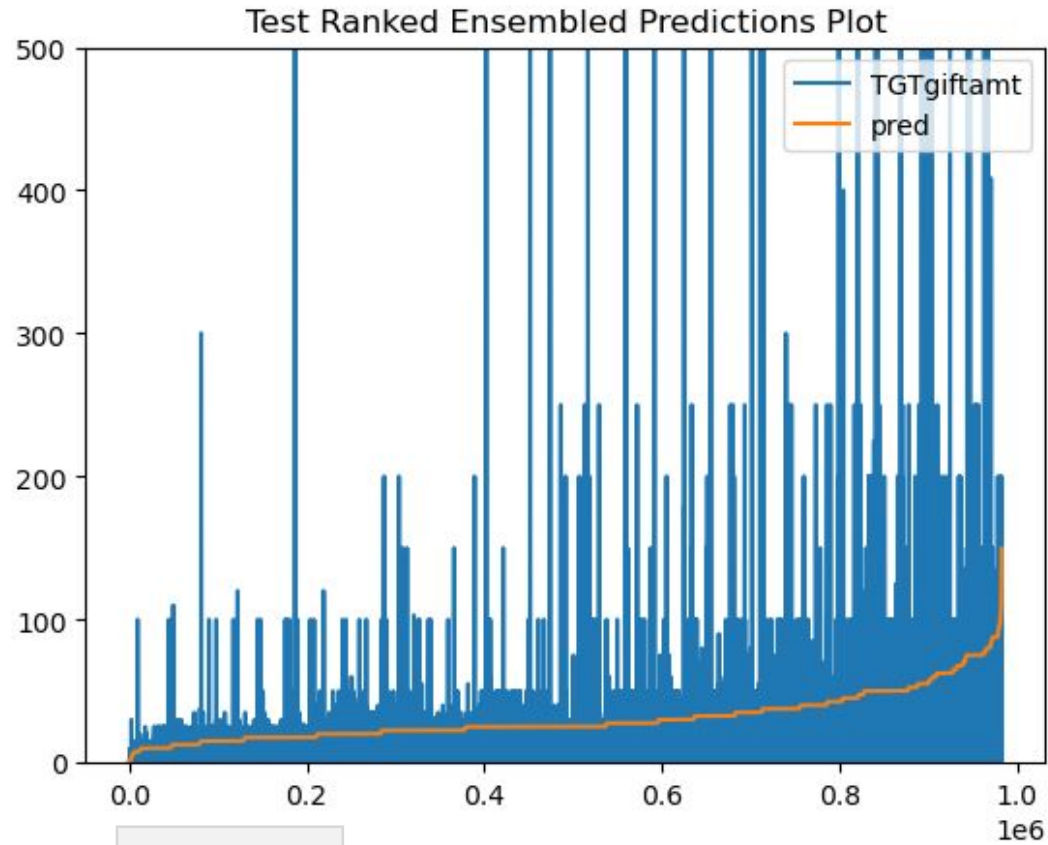
- Random Forest: 33.02
- Decision Tree: 30.97

### Steps performed:

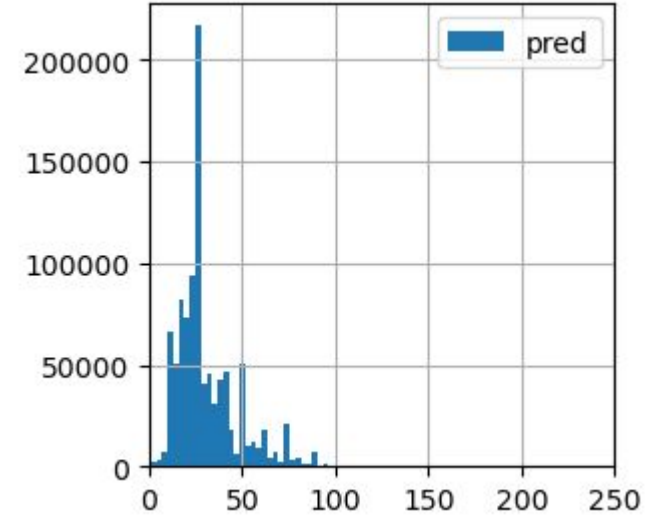
- Quantile feature transformation
- Data remediation by way of removing rows with some of the largest residual outliers

\* MAEs above calculated with test data used for final result predictions

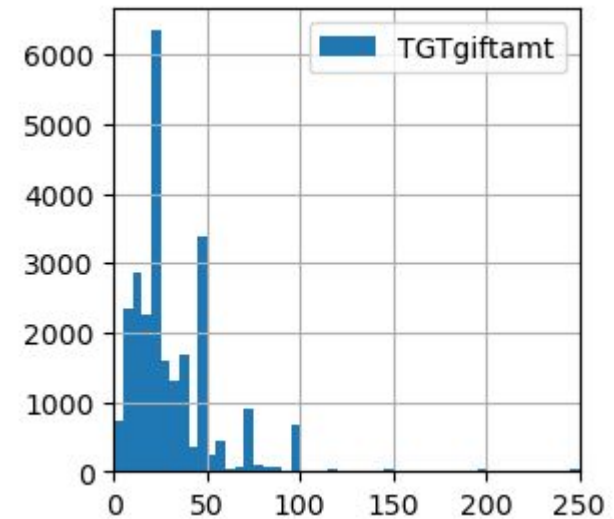
# Model 1 Performance



MAE: 30.12



Distribution of gift amount predictions from final validation data



Distribution of TGTgiftamt from positive responder data

# Expected Gift Calculation

## Formula

*classifier['predict\_proba'] \* regressor['predict']*

Final output decile analysis:

Decile	Responses	% of Total
1	5987	23.3%
2	4338	16.9%
3	3464	13.5%
4	2895	11.3%
5	2264	8.8%
6	1852	7.2%
7	1547	6%
8	1373	5.4%
9	1009	3.9%
10	933	3.6%
Total	25662	100%

# Next Steps

## Potential Next Steps

- Future enhancements to the models should focus on handling outlier donors. This model's output is limited to predicting expected gift amounts within our min and max (\$0.00-\$94.80).
  - In our original training data, analysis found that outliers were gift amounts over \$72.00
- Explore integration with other fundraising channels, such as online campaigns or events.
  - Conduct further analysis on lower deciles to determine if they are more responsive to other donation channels
- Continuously monitor and update the model to adapt to evolving donor behaviors and campaign dynamics.





## Risk Considerations

- Over-reliance on predictive models may overlook the human element and unique donor motivations.
- Ethical considerations regarding data privacy and transparency in model deployment.
- Potential for unintended consequences, such as alienating donors or reducing engagement if targeting strategies are too aggressive.