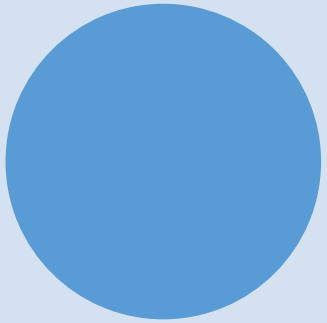
April 2024

Two-Stage Direct Response Model Update

GWSB Business Analytics Practicum Group 8

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.

Data Preprocessing Steps

Handling non-numeric x variables

- Created dummy variables for categorical variables and dropped columns converted to dummy variables and others we didn't intend to use
- Converted currency strings to float data type

Handling missing values

- Applied mean imputation to fill missing values for columns with missing values
  - append_mt_OnlineInsuranceBuyer
  - append_mt_CultureArtsEvents
  - append_mt_HighDollarDonor
  - append_mt_LowDollarDonor
  - append_age
- Dropped append_enviroconquintile as it was no longer needed
- Deleted rows with missing values in the append_HomeValue column

Phase 1 Model Data

Population: Previous Donors

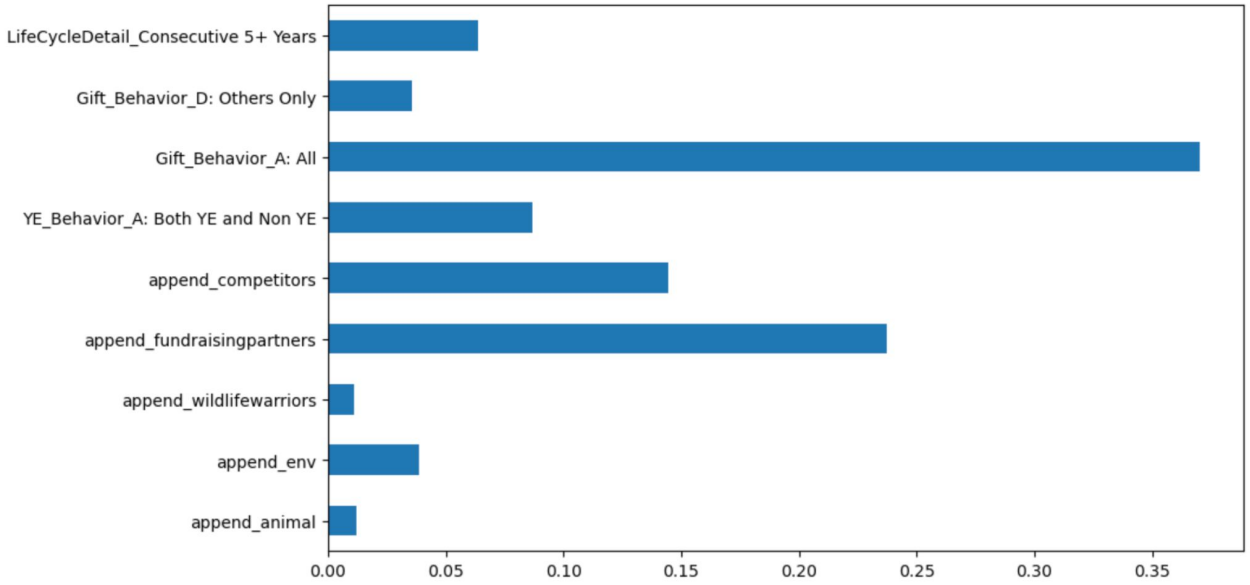
Target Variable[Mean=0.88]**:** 1 = Response; 0 = No Response (84931 VS 3191715)

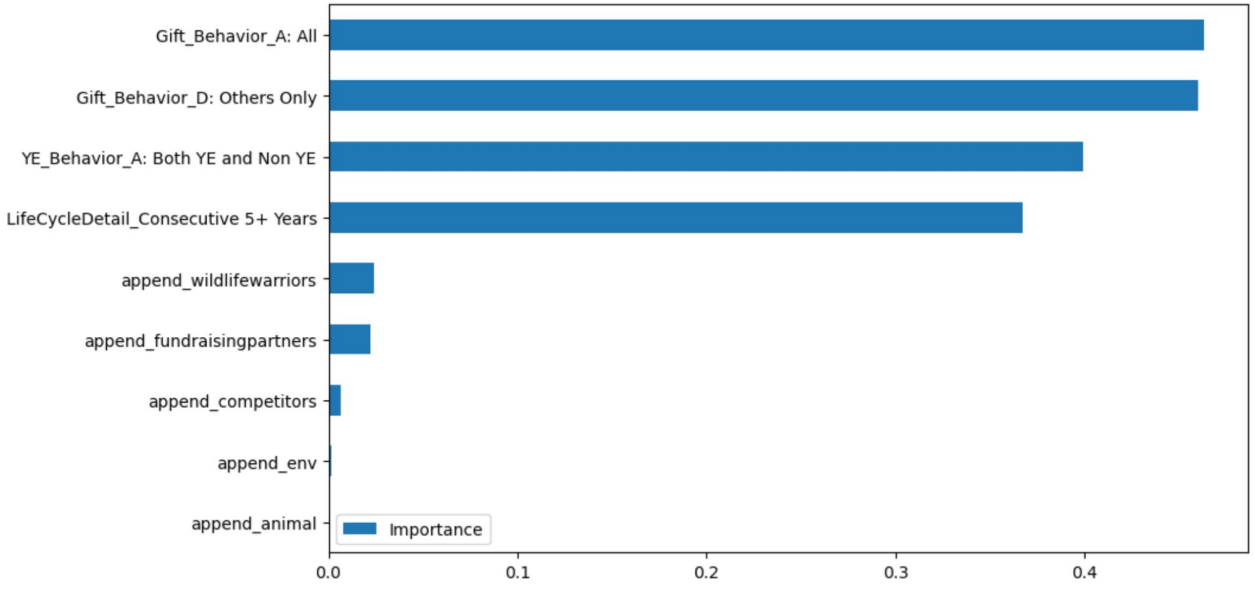
Feature Selection:

- 'append_animal',
- 'append_env',
- 'append_wildlifewarriors',
- 'append_fundraisingpartners',
- 'append_competitors',
- 'append_env_enthusiasts',
- 'append_direct_trans',
- 'YE_Behavior_A: Both YE and Non YE',
- 'Gift_Behavior_A: All',
- 'Gift_Behavior_D: Others Only',
- 'LifeCycleDetail_Consecutive 5+ Years',
- 'LifeCycle_Consecutive Givers',
- 'GivingYears',
- 'cumul_amount',
- 'append_trans_life',
- 'Monthly_Donor_Y',
- 'Monthly_Donor_N',
- 'gift_count',
- 'append_parks_nature'

Proportion of Features

Classification Tree Model VS. Logistic Regression Modelopulation:





Phase 1 Model Overview

Champion Model is an ensemble of two models

- DecisionTreeClassifier(class_weight={0: 0.3, 1: 1.2}, max_depth=9, min_samples_leaf=5, min_samples_split=3, random_state=12345)
- LogisticRegression(solver='newton-cholesky', C=.3, max_iter=10, random_state=12345, class_weight={0: 0.3, 1: 1.2})

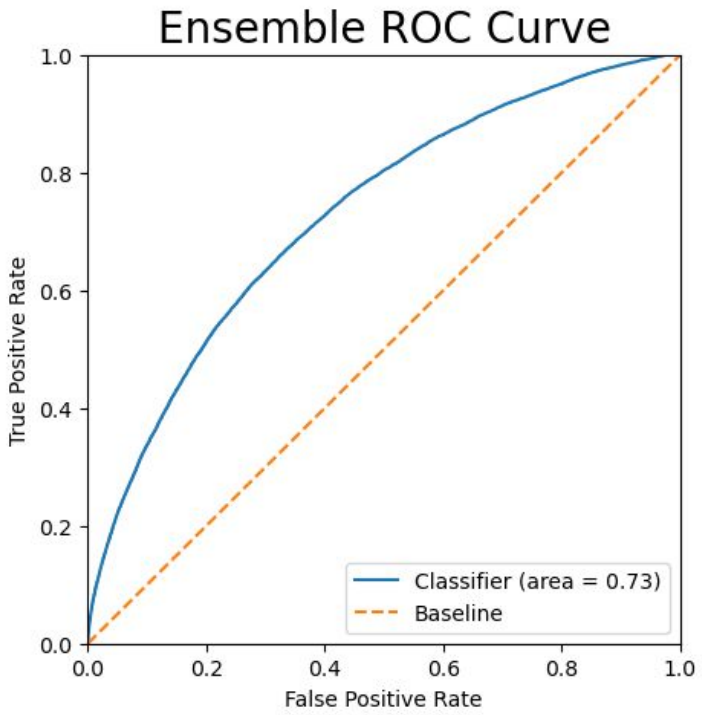
70% of the dataset was used for training the model, 30% for validation

The model was chosen based on the AUC (Area under Curve) score for the validation data

Top 5 Predictor Variables

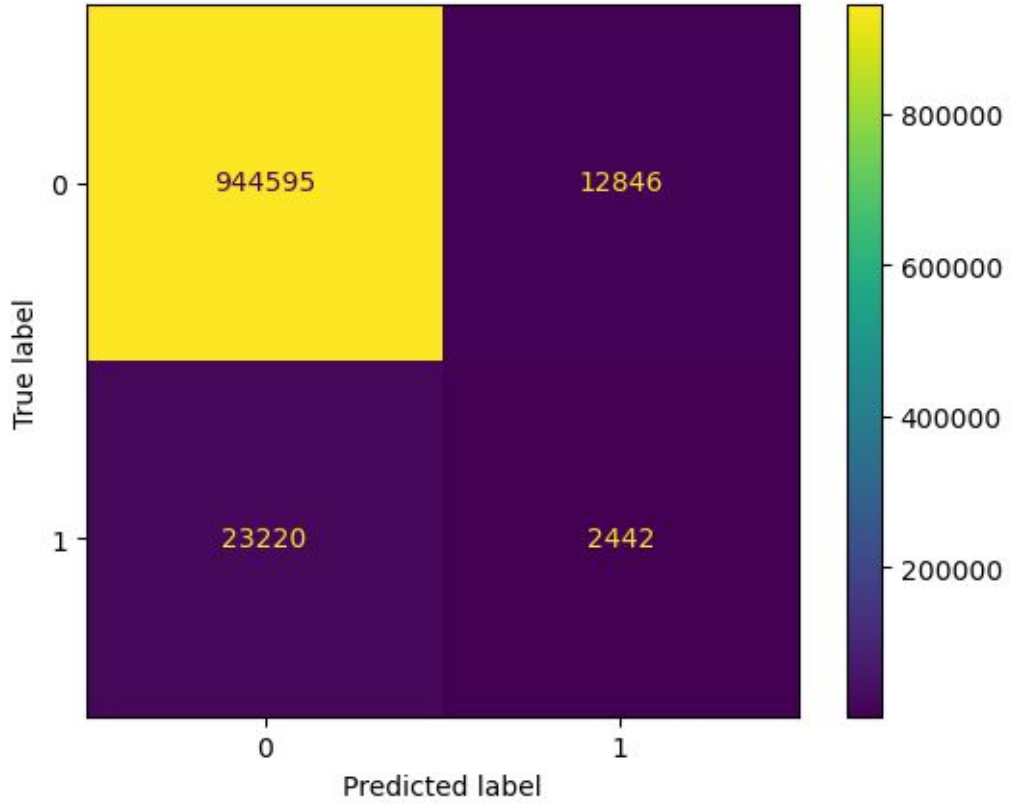
- Monthly_Donor_Y
- Monthly_Donor_N
- YE_Behavior_A: Both YE and Non YE
- Gift_Behavior_A: All
- Gift_Behavior_D: Others Only

Model 1 Performance



AUC = 0.73
KS = 0.97

Visualizing the confusion matrix with a lowered threshold (threshold=.3)



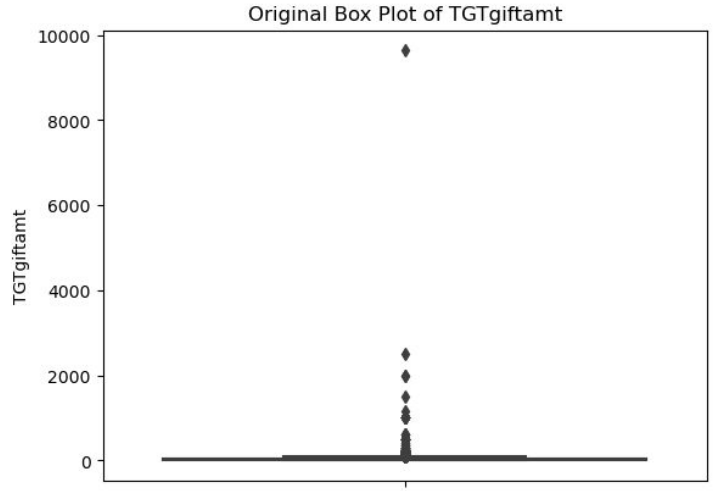
Phase 2 Model Data

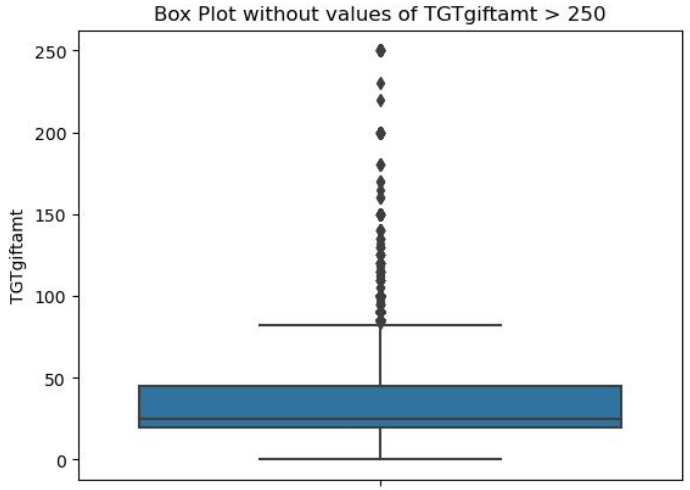
Population: Previous Donors (need to confirm what year(s) the data is from) with TGTresp = 1

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

Feature Selection:

- 'HPG_amount',
- 'MRG_amount',
- 'avg_gift_amt',
- 'first_gift_amount',
- 'append_direct_trans',
- 'append_fundraisingpartners',
- 'append_HouseHoldEducation'
- 'append_WealthResources',
- 'append_age',
- 'append_web',
- 'append_competitors',
- 'append_parks_nature'





Model Overview

Champion Model is an ensemble of two models

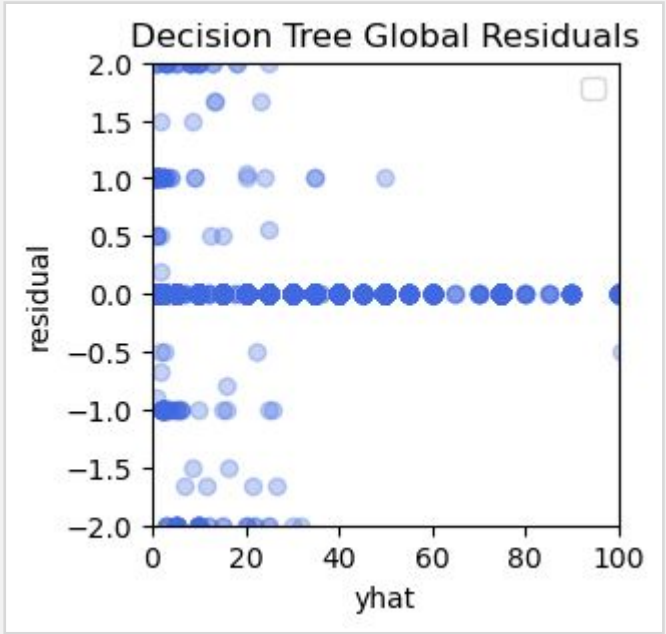
- DecisionTreeRegressor(random_state=12345)
- RandomForestRegressor(random_state=12345)

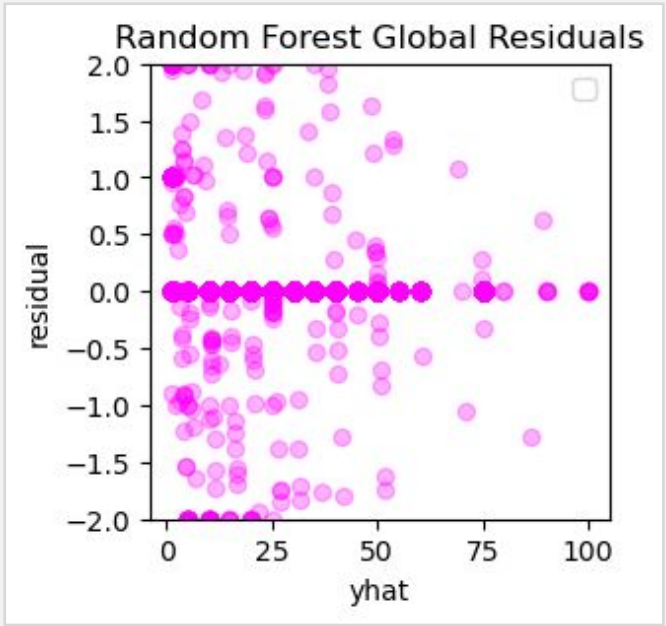
70% of the dataset was used for training the model, 30% for validation

The model was chosen based on the MAE (Mean Absolute Error)
for the validation data

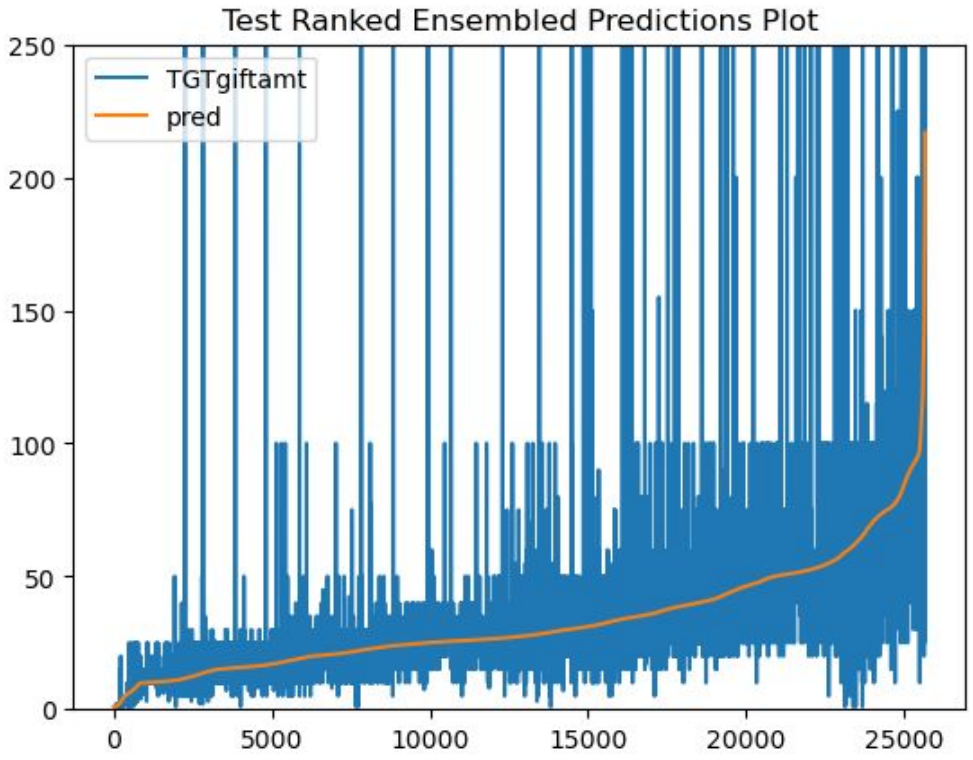
Top 5 Predictor Variables

- MRG_amount
- append_age
- avg_gift_amt
- append_direct_trans
- append_fundraisingpartners

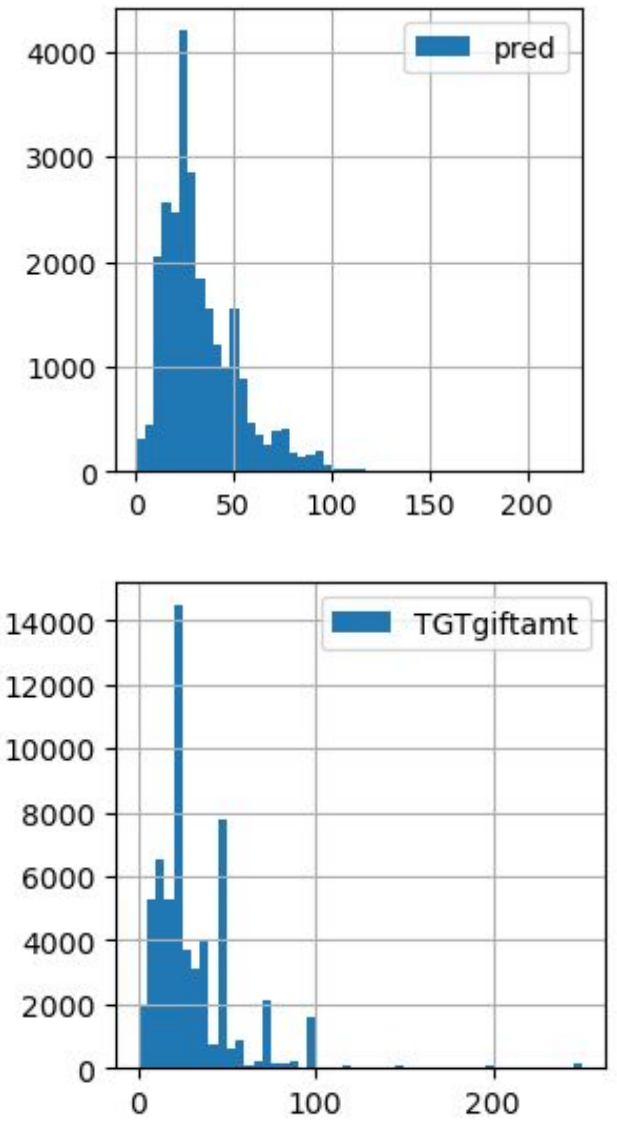




Phase 2 Model Performance



MAE: 5.947



Comparing TGTgiftamt distribution to ensemble model predictions distribution

Next Steps

Expected Gift Calculation

Formula

*classifier['predict_proba'] * regressor['predict']*

Final output dataframe:

	masterprimaryid	TGTresp	TGTgiftamt	expected_value
0	000004E1-D14E-42AB-A384-40A2773F507C	0	0.0	0.000000
648800	A7FFE446-3CB2-4EDE-BADD-3AD20CD8C82C	0	0.0	0.000000
648801	A7FFF811-9B13-40DE-A32B-4A4CEBB8ADA3	0	0.0	0.000000
648802	A7FFFBC5-7654-4629-9A61-BE246AA183B1	0	0.0	0.000000
648803	A800098C-CCC8-4055-8E26-68724177189C	0	0.0	0.000000
...	...	...	...	...
22771	05D7A30B-B6F5-434A-B779-F6B6B2EDCC3C	1	80.0	67.027449
149975	26E5D364-054F-47DB-8BCC-A9A56DBCA6E6	1	75.0	67.759939
149977	26E5D364-054F-47DB-8BCC-A9A56DBCA6E6	0	75.0	67.914802
149976	26E5D364-054F-47DB-8BCC-A9A56DBCA6E6	0	75.0	67.914802
185739	30024BB9-65A5-4854-9BD0-77FD06D6F87C	1	75.0	72.187903