nuveen
A TIAA Company

Prediction of sales and estimation of probability of adding a new funds in the next 12 months

Prepared by Dmitry Amanov Aug 6 2020

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Agenda

- Background
- Objectives
- Methodology
- Analysis Results
- Recommendations

Columbia University
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Nuveen

- One of the world's largest asset managers, serving institutions, financial intermediaries and individual investors in more than 30 countries.
- Recognized as a leader in income generation, alternative investments and responsible investing.
- Honors a 100-plus year legacy of service and innovation based on enduring principles established by John Nuveen and Andrew Carnegie.

Background

Columbia University
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Objectives

- Help sales and marketing to improve their targeting of the most profitable advisors.
- Develop models for sales prediction and for estimation of probability of adding a new funds in the next 12 months.

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK
EMERITUS INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Methodology

The key points of analysis methodology

- Analysis performed using Nuveen transactions data (2018 & 2019) and firm information data.
- Predictors selected based on data analysis and statistical hypothesis.
- Training performed on the 50% of the initial dataset using cross validation.
- The final evaluation is performed on the rest 50% of the initial dataset.

nuveen
A TIAA Company

Data Contents

Nuveen transactions data contents

- Sales and redemptions in the past 12 months
- Sales and redemptions in the current month
- AUM of different asset class and product type

nuveen
A TIAA Company

Regression Model

Top Predictors

Top predictors for the sales regression model

- Sales in the current month
- Number of sales over $10K in the last 12 months
- Number of redemptions over $10K in the last 12 months
- Net sales in the last 12 months

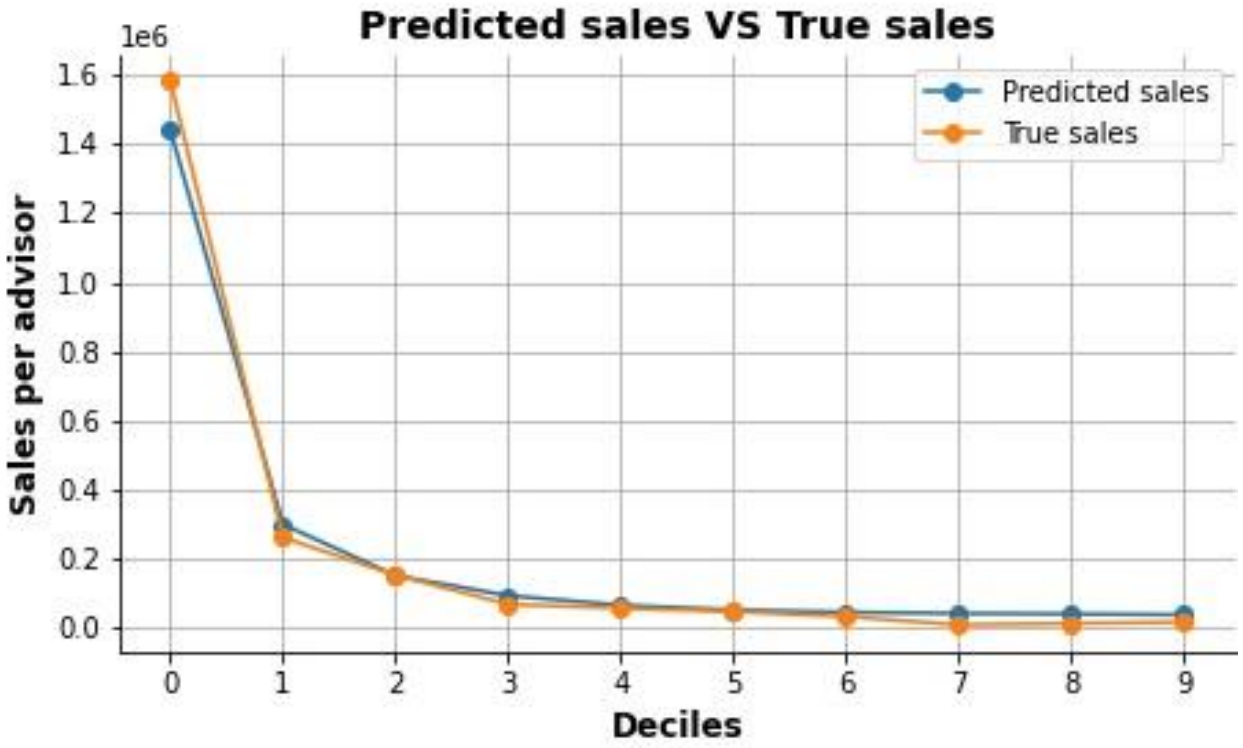
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK
EMERITUS INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Regression Model

Model Performance

Model is demonstrated a good match between predicted sales and true sales



Columbia University in the City of New York

Global Ivy Education EMERITUS Institute of Management

nuveen
A TIAA Company

Regression Model

Lift Chart

Total number of advisors: 4684
Average sales: $226193

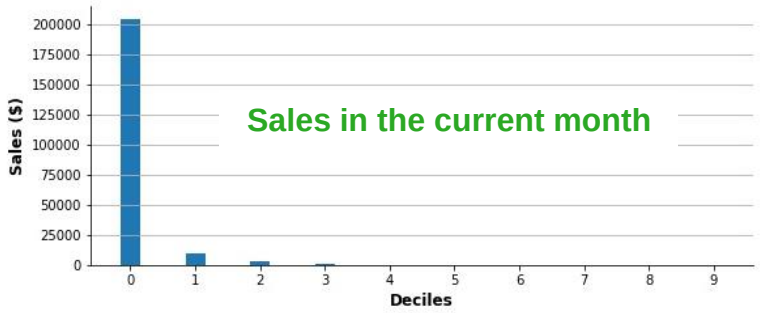
Decile	Number of advisors	Sales $ (per advisor)	Lift over average	Cumulative number of advisors	Cumulative sales $	Cumulative lift	Cumulative gain
0	468	1,440,016	537%	468	1,440,016	537%	64%
1	468	299,236	32%	936	1,739,252	569%	77%
2	468	150,731	-33%	1404	1,889,983	536%	83%
3	468	92,840	-59%	1872	1,982,823	477%	88%
4	468	65,344	-71%	2340	2,048,167	405%	90%
5	468	51,418	-77%	2808	2,099,584	328%	93%
6	468	44,778	-80%	3276	2,144,363	248%	95%
7	468	41,468	-82%	3744	2,185,831	166%	97%
8	468	40,138	-82%	4212	2,225,969	84%	98%
9	472	37,582	-83%	4684	2,263,550	1%	100%

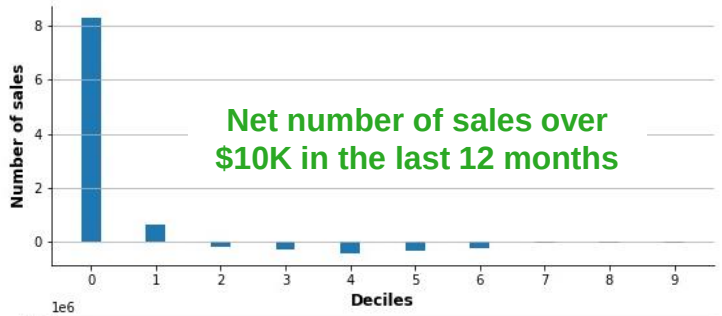
- Top 20% of advisors make 77% of the total sales.
- The first 10% has 5 times higher sales than the next decile.

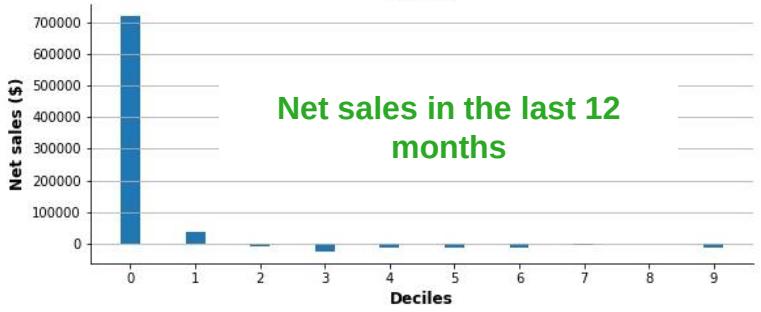
nuveen
A TIAA Company

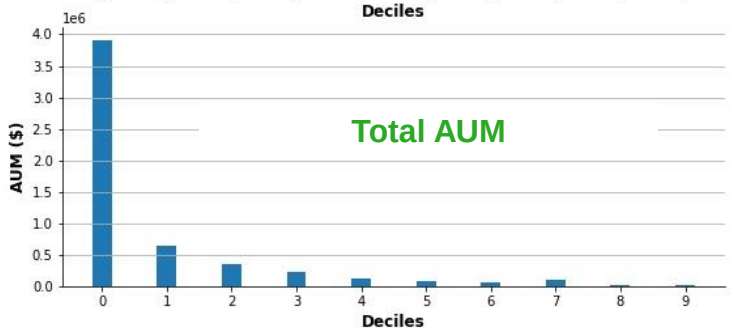
Regression Model

Relational Charts









Key attributes of advisors with the highest sales in the next 12 months

- The highest sales in the current month
- The highest net sales in the last 12 months
- The highest AUM

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK
EMERITUS INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Classification Model

Top Predictors

Top predictors for adding new funds

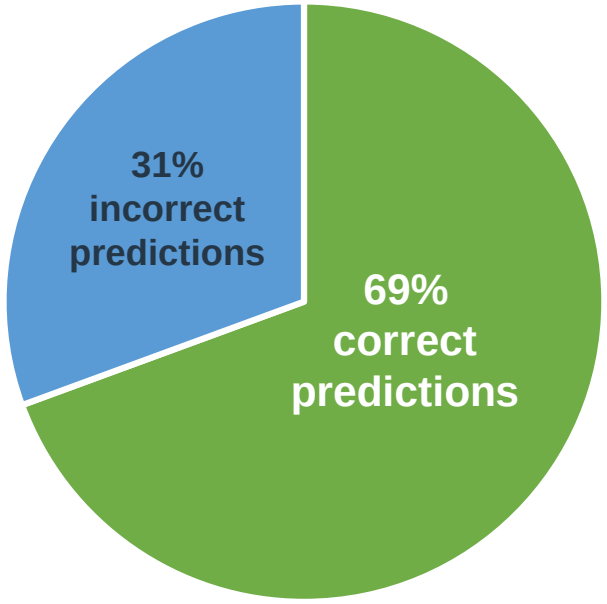
- Net sales in last 12 months
- Number of new funds added in the last 12 months
- Number of sales over $1 in the last 12 months
- Net number of fund sales over $1 in the last 12 months

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Predictions of adding new funds in the next 12 months (test dataset)



- 69% correct predictions for advisors that actually added new funds.

Classification Model

Model Performance

nuveen
A TIAA Company

Classification Model

Lift Chart

Total number of advisors: 4663
Average probability of adding new fund: 45%

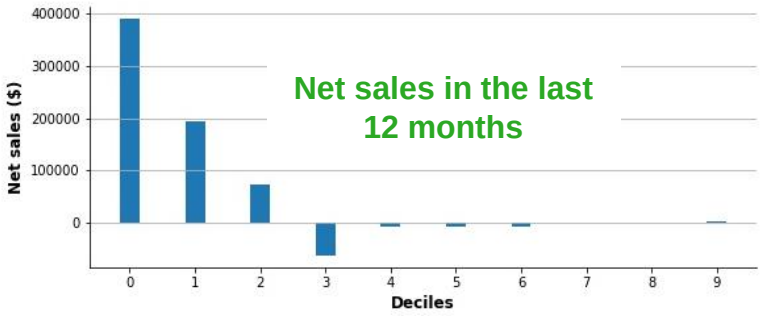
Decile	Number of advisors	Probability of adding new fund (avg. per advisor)	Lift over average	Cumulative number of advisors	Cumulative probability of adding new fund	Cumulative lift
0	466	78%	71%	466	78%	71%
1	466	66%	44%	932	72%	58%
2	466	59%	30%	1398	68%	49%
3	466	54%	19%	1864	64%	41%
4	466	50%	9%	2330	61%	35%
5	466	45%	-2%	2796	58%	29%
6	466	35%	-23%	3262	55%	21%
7	466	24%	-47%	3728	51%	13%
8	466	23%	-50%	4194	48%	6%
9	469	22%	-52%	4663	45%	0%

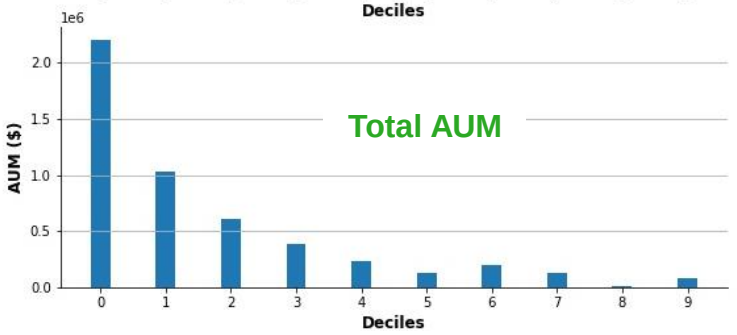
- Top 40% of advisors are well above the average probability of adding new funds.
- The top 10% has almost twice higher probability of adding funds than the rest 30%.

nuveen
A TIAA Company

Classification Model

Relational Charts





Key attributes of advisors with the highest probability of adding new funds

- The highest net sales in the last 12 months
- The highest numbers of new funds added in the last 12 months
- The highest AUM

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK
EMERITUS INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Findings

- The majority of advisors that have the highest sales and the highest probability of adding new funds are **independent dealers** and **national broker-dealers**. And the major subchannels are **NACS** and **IBD**.
- The most contributing firms in the top decile are **Merrill Lynch** and **Morgan Stanley Wealth Management**. They make 18% and 14% out of 64% of the total sales correspondingly. And each of them holds 16% among advisors with the highest probability of adding new fund in the next 12 months.

nuveen
A TIAA Company

Advisors with the highest sales in the next 12 months

- Target 20% of advisors with the highest sales in the current month, the highest net sales in the last 12 months and with the highest AUM.
- These advisors have their sales above average, they make 77% of total sales and they will have the highest sales in the next 12 months.

Advisors with the highest probability of adding new funds in the next 12 months

- Target 40% of advisors with the highest net sales in the last 12 months, the highest numbers of new funds added in the last 12 months and with the highest AUM.
- The top 10% of these advisors has almost twice higher probability of adding funds than the rest 30%.

Recommendations

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK
EMERITUS INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Appendix

Technical Report

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Contents

- Explanatory Data Analysis (EDA)
- Feature Engineering
- Feature Selection
- Model Selection
- Training & Evaluation

Columbia University
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

Overview and dealing with NaNs

Provided transactions data consist of 10005 samples and 38 columns.

All columns are numeric and represent either ordinal (number of something) or continuous (sales in dollars) data.

There is a big number of missing values (around 36%).

```
print('Percent of missing values: {0:.0%}'.format(data.isnull().sum().mean() / len(data)))
```

```
Percent of missing values: 36%
```

```
data.isnull().sum()
```

no_of_sales_12M_1	5242
no_of_Redemption_12M_1	4644
no_of_sales_12M_10K	7293
no_of_Redemption_12M_10K	7029
no_of_funds_sold_12M_1	5242
no_of_funds_redeemed_12M_1	4644
no_of_fund_sales_12M_10K	7293
no_of_funds_Redemption_12M_10K	7029
no_of_assetclass_sold_12M_1	5242
no_of_assetclass_redeemed_12M_1	4644
no_of_assetclass_sales_12M_10K	7293
no_of_assetclass_Redemption_12M_10K	7029
No_of_fund_curr	3822
No_of_asset_curr	4426
AUM	585
sales_curr	7574
sales_12M	5237
redemption_curr	7429
redemption_12M	4621
new_Fund_added_12M	7310
aum_AC_EQUITY	585
aum_AC_FIXED_INCOME_MUNI	585
aum_AC_FIXED_INCOME_TAXABLE	585
aum_AC_MONEY	585
aum_AC_MULTIPLE	585
aum_AC_PHYSICAL_COMMODITY	585
aum_AC_REAL_ESTATE	585
aum_AC_TARGET	585
aum_P_529	585
aum_P_ALT	585
aum_P_CEF	585
aum_P_ETF	585
aum_P_MF	585
aum_P_SMA	585
aum_P_UCITS	585
aum_P_UIT	585
sales_12M_target	4931
new_Fund_added_12M_target	7484

dtype: int64

All missing values are set to 0.

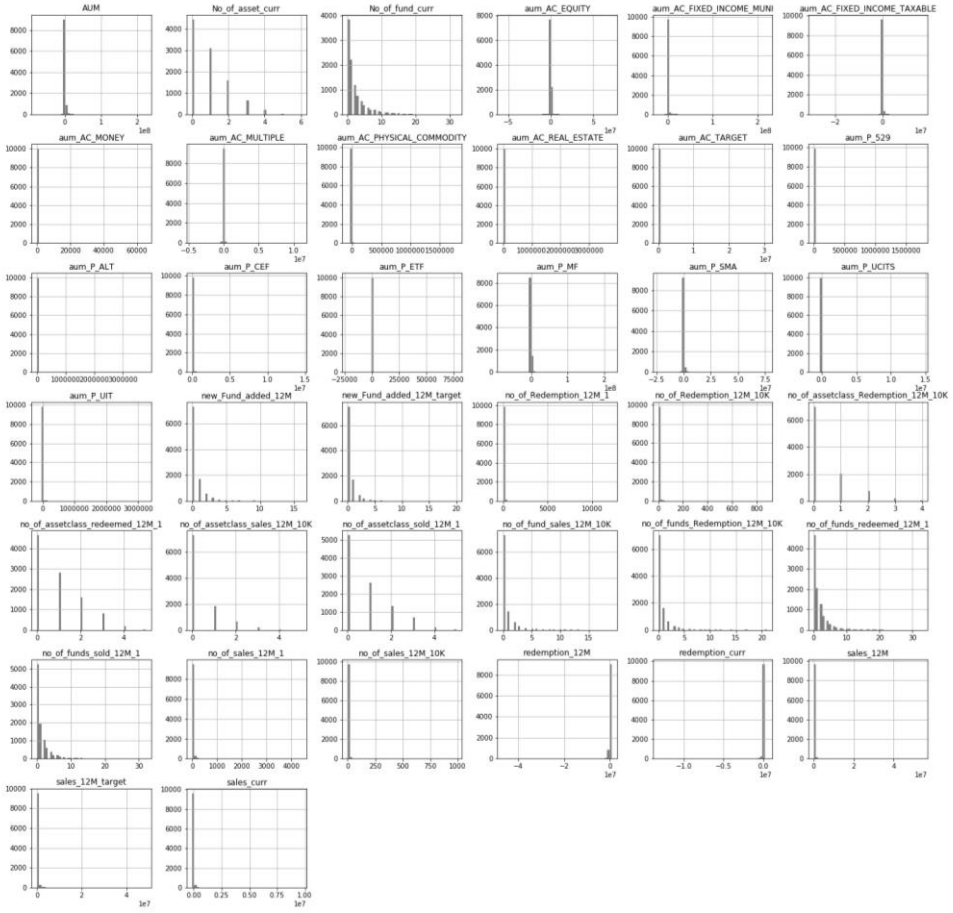
nuveen
A TIAA Company

EDA

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

EMERITUS INSTITUTE OF MANAGEMENT

Features distributions



- The data is highly biased toward zero. But also, there are a noticeable number of samples with very large values.
- The data is highly unbalanced in terms of classification target

```
Class priors:
Class 0 (NO new funds added): 75%
Class 1 (new funds added): 25%
```

nuveen
A TIAA Company

EDA

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

GLOBAL IVY EDUCATION
EMERITUS
INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Feature Engineering

- Drop all samples that have negative sales and positive redemption.
- Use absolute value for redemption.
- Create “net” columns (sales – redemption) then split them into “positive net” (values >= 0) and “negative net” (values < 0). Both positive_net and negative_net columns have positive values.
- Replace AUM columns with positive/negative pair.
- Apply log(x + 1) transform to the all features.
- Apply one-hot encoder for target variable for classification model.

nuveen
A TIAA Company

Feature Selection

Subsets

Models were tested on two subsets of features:

Subset_1 = x_noof + x_aum + ['sales_curr', 'sales_12M', 'redemption_curr', 'redemption_12M', 'new_Fund_added_12M']

Subset_2 = x_net + x_aum + ‘new_Fund_added_12M’

Where:

x_noof: original sales / redemptions columns

x_net: all positive / negative net columns

x_aum: all positive / negative AUM columns

Statistical tests

Initial features selection was performed using statistical tests.

Feature – target correlation tests

- Regression
  - Pearson's Correlation Coefficient
  - Spearman's Rank Correlation
  - Kendall's Rank Correlation
- Classification
  - Point biserial correlation
  - Kruskal-Wallis H-test

Features multicollinearity test

- Variance inflation factor (vif)

- *Only features with p_value <= 0.05 for all tests were selected.*
- *Only features with vif < 5 were selected*

NOTE: Statistical tests are performed on both subsets. Then selected features from both subsets are merged into a single subset.

nuveen
A TIAA Company

Feature Selection

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

EMERITUS INSTITUTE OF MANAGEMENT

Regression Model

- **Model Type:**
  - Stacking Regressor
- **Base estimators:**
  - Gradient Boosting Regressor
  - Extra Trees Regressor
  - Decision Tree Regressor
- **Final Estimator:**
  - LinearRegression
- **Cross-validation strategy**
  - Train-test split: 50% train / 50% test
  - Repeated Kfold cross-validation: 5 folds & 10 repeats
  - Metrics: MAE, R2, explained variance

nuveen
A TIAA Company

Model Selection

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

EMERITUS INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Classification Model

- **Model Type:**
  - Voting Classifier
- **Base estimators:**
  - Gradient Boosting Classifier
  - Logistic Regression
- **Cross-validation strategy**
  - Train-test split: 50% train / 50% test
  - Repeated stratified Kfold cross-validation: 5 folds & 10 repeats
  - Metrics: accuracy, ROC-AUC

Model Selection

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

EMERITUS INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

Training

- Train model on the pre-selected features
- Run permutation feature importance and keep only features with positive score
- Retrain model on selected features
- Compare results before and after feature importance study

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

Regression model

nuveen
A TIAA Company

Training & Evaluation

Features importance

Feature	importance
sales_curr	0.1508
no_of_sales_12M_10K	0.0695
no_of_Redemption_12M_10K	0.0352
pos_net_sales_12M	0.0313
pos_aum_AC_FIXED_INCOME_MUNI	0.0308
no_of_Redemption_12M_1	0.0267
redemption_curr	0.0216
pos_aum_AC_FIXED_INCOME_TAXABLE	0.0193
no_of_sales_12M_1	0.0088
pos_aum_AC_TARGET	0.0028
neg_net_sales_curr	0.0023
pos_aum_AC_PHYSICAL_COMMODITY	0.0019
pos_net_no_of_sales_12M_1	0.0017
pos_net_no_of_assetclass_sales_12M_10K	0.0014
neg_net_no_of_sales_12M_1	0.0013
neg_net_no_of_assetclass_sales_12M_10K	0.0007
pos_net_no_of_fund_sales_12M_10K	0.0007
neg_aum_P_SMA	0.0007
pos_aum_AC_MULTIPLE	0.0004
neg_aum_P_MF	0.0004
neg_aum_AC_FIXED_INCOME_MUNI	0.0003
neg_aum_AC_MULTIPLE	0.0003
pos_net_no_of_funds_sold_12M_1	0.0001
pos_aum_P_ETF	0.0001
pos_aum_AC_REAL_ESTATE	0.0000
pos_net_no_of_sales_12M_10K	-0.0001
pos_aum_AC_EQUITY	-0.0002
neg_net_no_of_fund_sales_12M_10K	-0.0003
neg_net_no_of_funds_sold_12M_1	-0.0004
neg_aum_AC_FIXED_INCOME_TAXABLE	-0.0007
neg_net_no_of_sales_12M_10K	-0.0008
pos_net_no_of_assetclass_sold_12M_1	-0.0018
pos_aum_P_CEF	-0.0040
new_Fund_added_12M	-0.0086
pos_net_sales_curr	-0.0127
pos_aum_P_SMA	-0.0227

Important features

CV results (all features)

CV Results
==========

	mae	explained_variance	r2
mean_train	106375.992151	0.826594	0.826522
std_train	16709.981042	0.047588	0.047577
mean_test	218992.712576	0.309704	0.308429
std_test	22800.677533	0.103610	0.103443

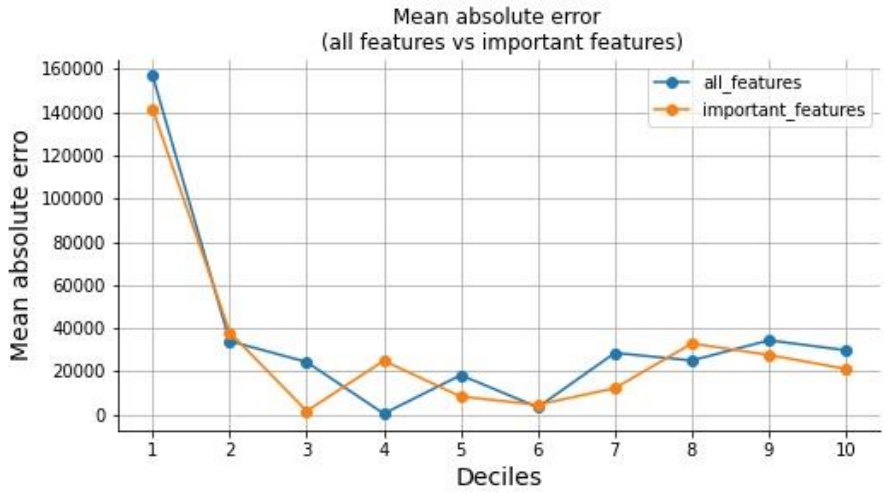
Validation on the test set
==========
mean_absolute_error: 214850.17769694887
explained_variance_score: 0.398818764726113
r2_score: 0.3988079933413986

CV results (important features)

CV Results
==========

	mae	explained_variance	r2
mean_train	109513.614908	0.811716	0.811585
std_train	15783.495423	0.052522	0.052502
mean_test	218974.136626	0.314009	0.312927
std_test	22434.782710	0.120451	0.120295

Validation on the test set
==========
mean_absolute_error: 212601.70701347786
explained_variance_score: 0.44591683692611594
r2_score: 0.4459123432422083



Model that trained on the important features subset demonstrates considerably higher score.

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK
EMERITUS INSTITUTE OF MANAGEMENT

Classification model

nuveen
A TIAA Company

Training & Evaluation

Features importance

Feature	importance
pos_net_sales_12M	0.0306
neg_net_sales_12M	0.0262
new_Fund_added_12M	0.0237
no_of_sales_12M_1	0.0190
neg_net_no_of_funds_sold_12M_1	0.0159
pos_aum_AC_FIXED_INCOME_TAXABLE	0.0086
neg_net_sales_curr	0.0074
neg_net_no_of_fund_sales_12M_10K	0.0048
no_of_Redemption_12M_1	0.0026
pos_aum_P_CEF	0.0022
neg_aum_AC_MULTIPLE	0.0022
pos_net_no_of_fund_sales_12M_10K	0.0020
neg_aum_AC_FIXED_INCOME_TAXABLE	0.0017
neg_aum_AC_FIXED_INCOME_MUNI	0.0017
no_of_sales_12M_10K	0.0016
pos_aum_P_SMA	0.0013
neg_net_no_of_sales_12M_10K	0.0012
pos_net_no_of_assetclass_sold_12M_1	0.0008
pos_net_no_of_funds_sold_12M_1	0.0006
no_of_Redemption_12M_10K	0.0004
pos_aum_AC_TARGET	0.0004
pos_aum_AC_REAL_ESTATE	0.0003
redemption_curr	0.0003
pos_aum_P_UIT	0.0001
neg_net_no_of_assetclass_sold_12M_1	0.0001
pos_aum_AC_MULTIPLE	0.0001
sales_curr	0.0001
pos_aum_P_ETF	0.0000
pos_aum_AC_EQUITY	0.0000
neg_aum_P_529	0.0000
neg_aum_AC_REAL_ESTATE	0.0000
neg_net_no_of_assetclass_sales_12M_10K	0.0000
neg_aum_P_ETF	-0.0001
pos_aum_AC_PHYSICAL_COMMODITY	-0.0002
neg_aum_AC_PHYSICAL_COMMODITY	-0.0003
pos_aum_P_529	-0.0004
pos_net_no_of_sales_12M_1	-0.0005
pos_net_no_of_sales_12M_10K	-0.0006
pos_aum_AC_FIXED_INCOME_MUNI	-0.0006
neg_aum_P_UIT	-0.0012
pos_net_no_of_assetclass_sales_12M_10K	-0.0018
neg_aum_P_MF	-0.0020
pos_net_sales_curr	-0.0044

Important features

CV results (all features)

CV Results

	accuracy	roc_auc
mean_train	0.699614	0.791622
std_train	0.005747	0.003644
mean_test	0.660938	0.719954
std_test	0.015517	0.016215

Validation on the test set

accuracy: 0.6450782757881193
roc_auc: 0.6553828524744366

CV results (important features)

CV Results

	accuracy	roc_auc
mean_train	0.694396	0.786160
std_train	0.005803	0.003627
mean_test	0.660746	0.721738
std_test	0.014115	0.015540

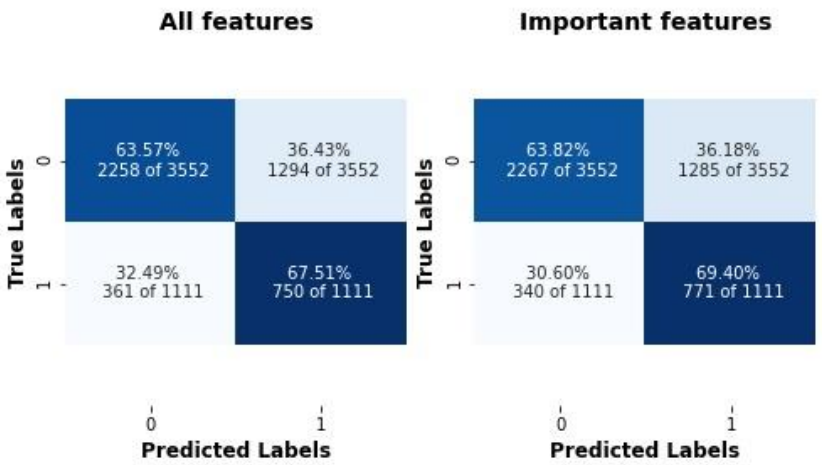
Validation on the test set

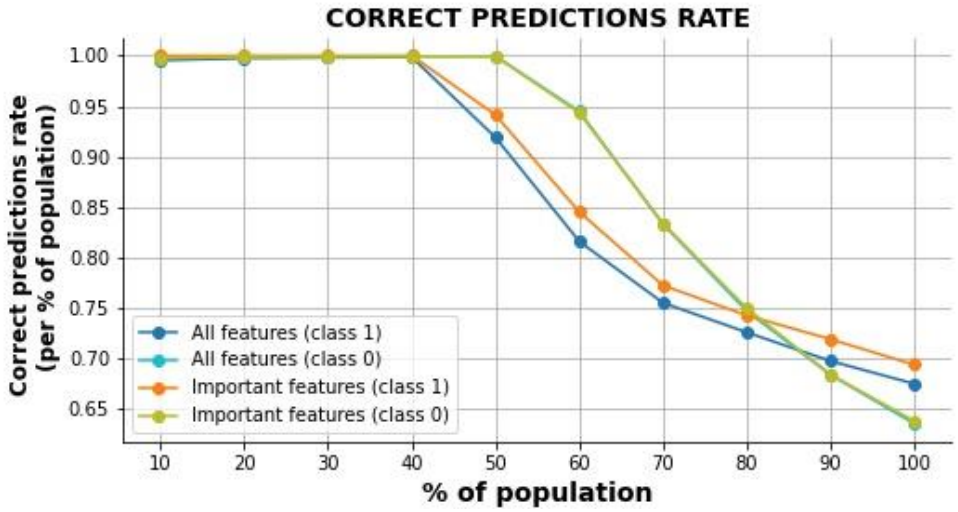
accuracy: 0.6515119022088784
roc_auc: 0.666100689460838

Model is fit on weighted samples. Sample weights equal to inverted class priors.

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK
EMERITUS INSTITUTE OF MANAGEMENT

Classification model (continued)





nuveen
A TIAA Company

Training & Evaluation

Columbia University
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

Attachments

- EDA-FINAL.ipynb - EDA
- Sales_reg_x_noof_x_aum_GBR_FINAL.ipynb – regression model
- Sales_reg_lift.xlsx – lift chart for regression model
- NewFund_cls_x_noof_x_aum_GBC_FINAL.ipynb – classification model
- NewFund_cls_lift.xlsx – lift chart for classification model

nuveen
A TIAA Company

Attachments

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT