# Prediction of sales and probability of adding a new fund

Capstone Project by Dmitry Amanov. Aug 6, 2020

# Agenda

- Background
- Objectives
- Methodology
- Analysis Results
- Recommendations

nuveen
A TIAA Company

Columbia University in the City of New York

Emeritus Institute of Management

nuveen
A TIAA Company

## Background

Nuveen is a mutual fund company headquartered in Chicago, with major offices in New York City, Charlotte, San Francisco, London and secondary offices in Frankfurt, Los Angeles, Shanghai, Singapore, Rio de Janeiro, Vienna, Stockholm, Minneapolis, Montreal, Washington DC, Tokyo, Luxembourg, Madrid, Milan, Paris, and Miami.

Nuveen is tasked with marketing and selling mutual funds through investment professionals such as brokers, financial planners, and financial advisors.

# Objectives

- Assist sales and marketing by improving their targeting.
- Predict sales for 2019 using the data for 2018.
- Estimate the probability of adding a new fund in 2019.

# Summary

- Analysis performed using Nuveen transactions data (2018 & 2019).
- Initial features selection is performed based on EDA and statistical hypothesis tests.
- Final features selection is performed using feature importance evaluation technics on pre-trained models.
- Models are trained on 70% of the initial dataset using cross validation.
- The final evaluation of models is performed on the rest 30% of the initial dataset.

nuveen
A TIAA Company

## Methodology

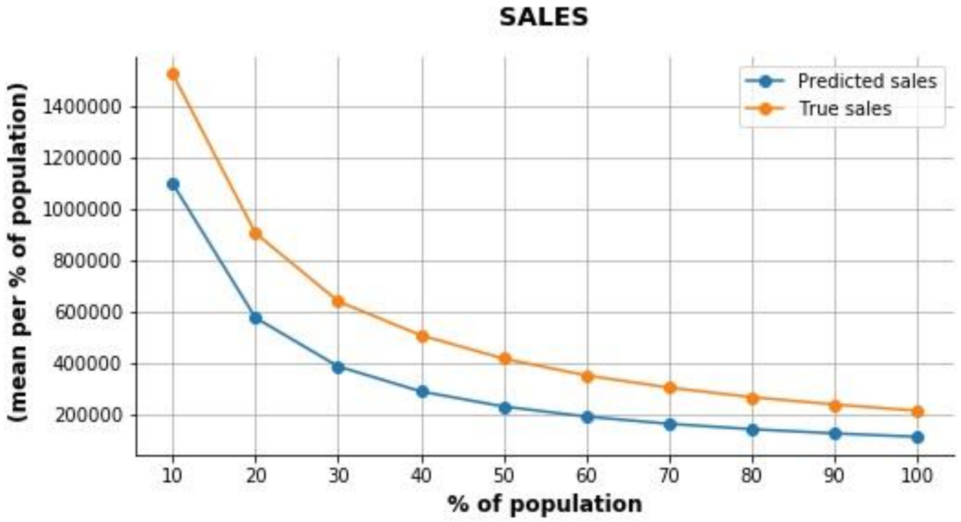
# Selected features for regression model (sales prediction)

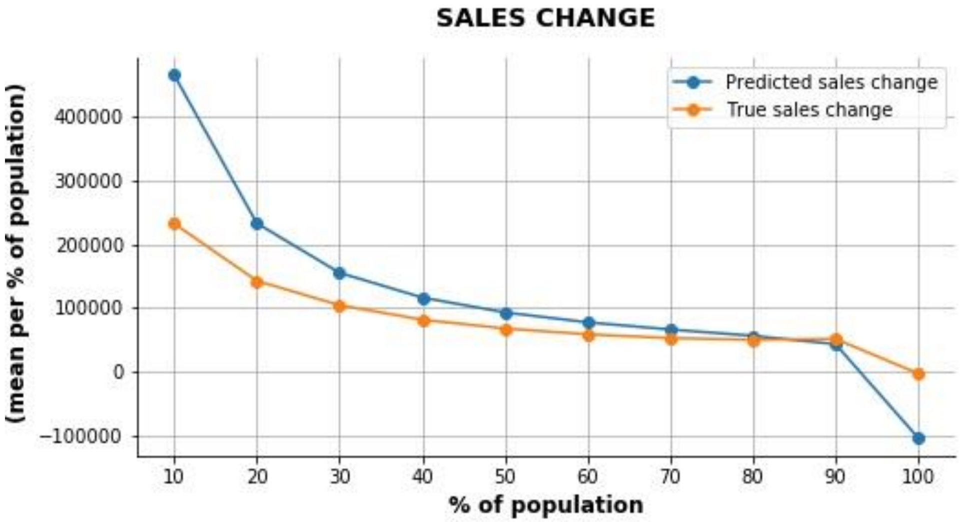
- Total sales in current month
- Total redemption in current month
- Number of sales in last 12 months that more than $1
- Number of redemptions in last 12 months that more than $1
- Number of sales in last 12 months that more than $10K
- Number of redemptions in last 12 months that more than $10K
- AUM (asset class EQUITY)
- AUM (asset class FIXED INCOME TAXABLE
- AUM (asset class TARGET)
- AUM (product type CEF)
- AUM (product type ETF)
- AUM (product type SMA)

# Selected features for classification model (probability of adding new fund)

- Total sales in current month
- Total redemption in current month
- Number of sales in last 12 months that more than $1
- New funds added in the last 12 months excluding current month
- AUM (asset class FIXED INCOME TAXABLE
- AUM (asset class MULTIPLE)
- AUM (asset class REAL ESTATE)
- AUM (asset class TARGET)
- AUM (product type MF)
- AUM (product type UIT)

# Regression Model Performance



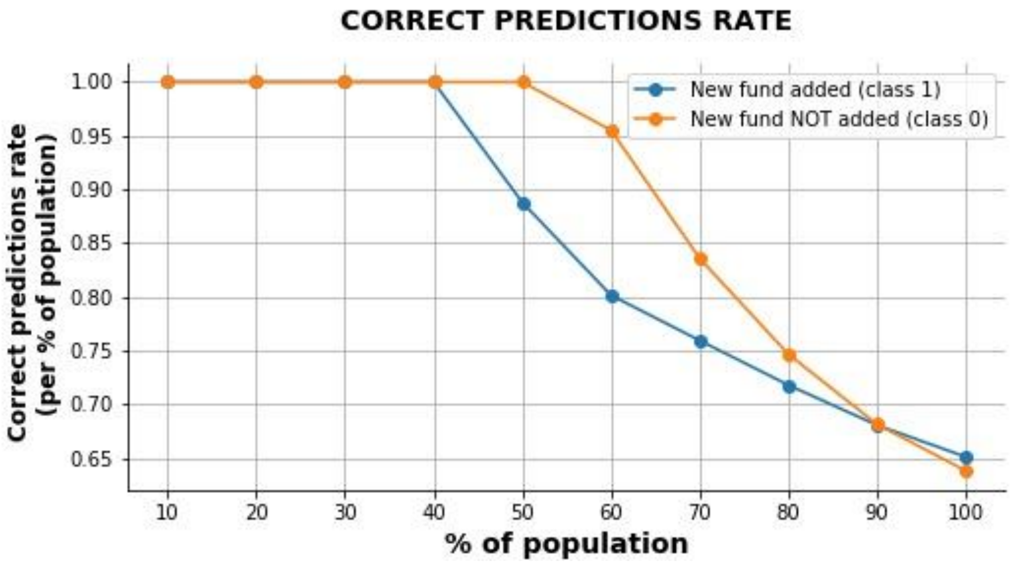



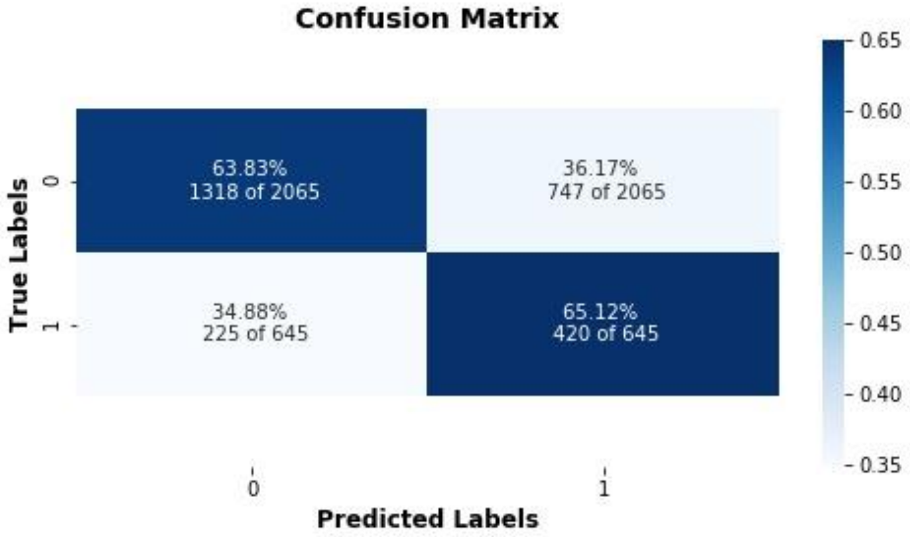

nuveen
A TIAA Company

## Analysis Results

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

# Classification Model Performance




Analysis Results

nuveen
A TIAA Company

Columbia University
in the City of New York

Global Ivy Education
EMERITUS
INSTITUTE OF MANAGEMENT

# Lift Charts (regression model)

Total population: 2726
Mean sales: $217828.41

| Decile | Number of advisors | Sales $ (avg. per decile) | Lift over average | Cumulative number of advisors | Cumulative sales $ | Cumulative lift |
|---|---|---|---|---|---|---|
| 1 | 272 | 1949312 | 795% | 272 | 1949312 | 795% |
| 2 | 272 | 174977 | -20% | 544 | 1062144 | 388% |
| 3 | 272 | 46704 | -79% | 816 | 723664 | 232% |
| 4 | 272 | 10830 | -95% | 1088 | 545456 | 150% |
| 5 | 272 | 1267 | -99% | 1360 | 436618 | 100% |
| 6 | 272 | 0 | -100% | 1632 | 363848 | 67% |
| 7 | 272 | 0 | -100% | 1904 | 311870 | 43% |
| 8 | 272 | 0 | -100% | 2176 | 272886 | 25% |
| 9 | 272 | 0 | -100% | 2448 | 242565 | 11% |
| 10 | 278 | 0 | -100% | 2726 | 217828 | 0% |



## Analysis Results

# Lift Charts (classification model)

Total population: 2710
Mean Probability of adding new fund: 45%

| Decile | Number of advisors | Probability of adding new fund (avg. per decile) | Lift over average | Cumulative number of advisors | Cumulative probability of adding new fund | Cumulative lift |
|---|---|---|---|---|---|---|
| 1 | 271 | 76% | 66% | 271 | 76% | 66% |
| 2 | 271 | 64% | 41% | 542 | 70% | 54% |
| 3 | 271 | 57% | 26% | 813 | 66% | 44% |
| 4 | 271 | 53% | 15% | 1084 | 63% | 37% |
| 5 | 271 | 49% | 7% | 1355 | 60% | 31% |
| 6 | 271 | 41% | -10% | 1626 | 57% | 24% |
| 7 | 271 | 33% | -29% | 1897 | 53% | 17% |
| 8 | 271 | 30% | -34% | 2168 | 50% | 10% |
| 9 | 271 | 29% | -37% | 2439 | 48% | 5% |
| 10 | 271 | 25% | -44% | 2710 | 46% | 0% |

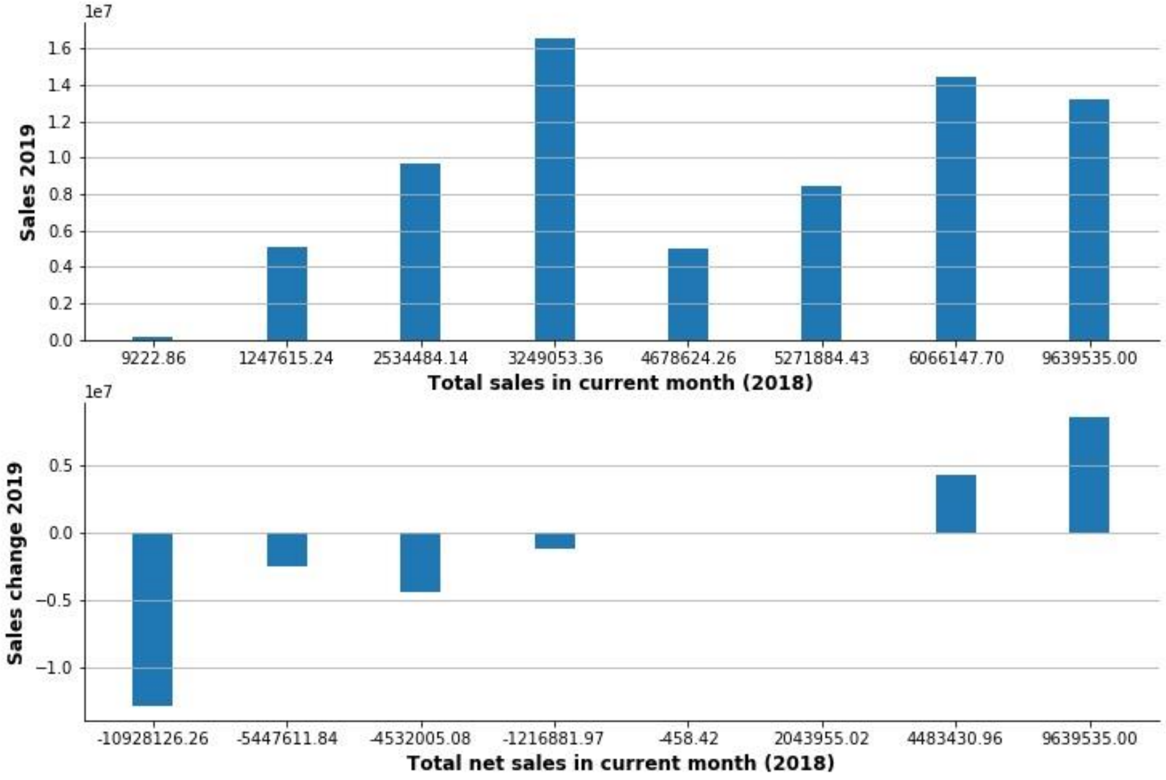
nuveen
A TIAA Company

## Analysis Results

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

# Relational charts (regression model)




- Advisors that have higher sales in current month will likely have higher sales next year.
- Advisors that have higher net sales in current month will likely increase sales next year and vice-versa.

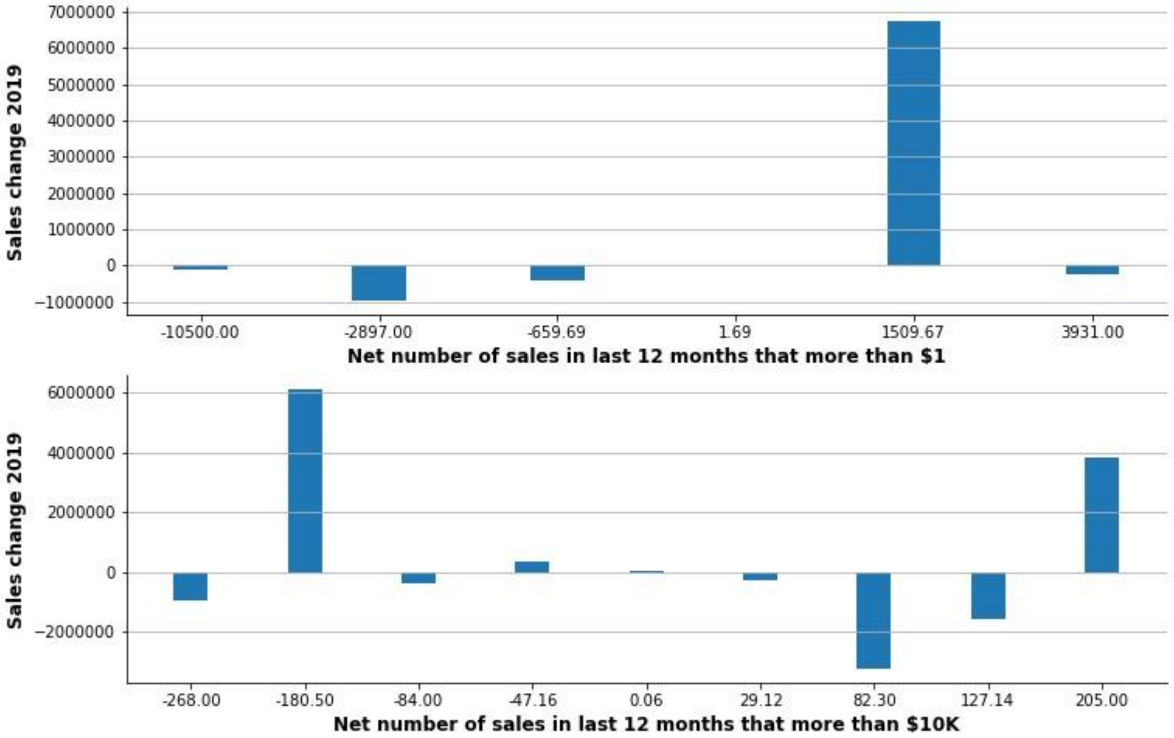
nuveen
A TIAA Company

## Analysis Results

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

# Relational charts (regression model)




- Advisors that have positive net number of sales that are more than $1 in the last 12 months will likely increase their sales next year. However, advisors with positive net number of sales that more than $10K in the last 12 months will likely decrease their sales next year.

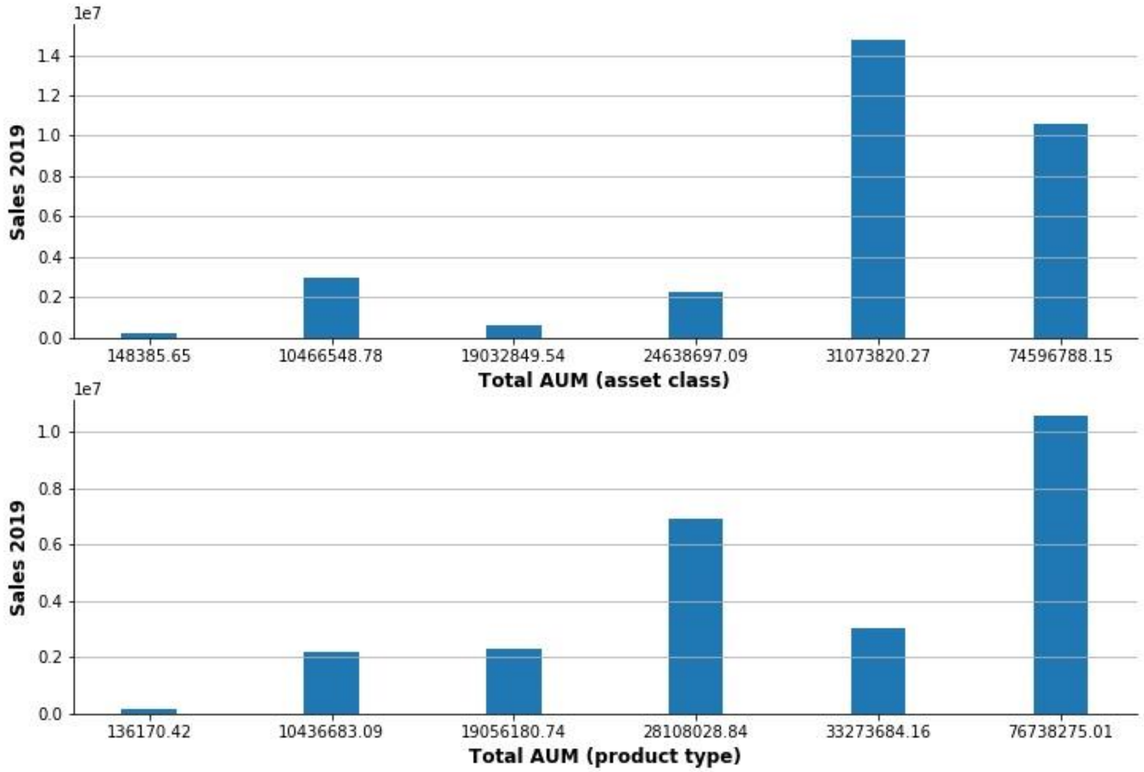
nuveen
A TIAA Company

## Analysis Results

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

GLOBAL IVY EDUCATION EMERITUS INSTITUTE OF MANAGEMENT

# Relational charts (regression model)



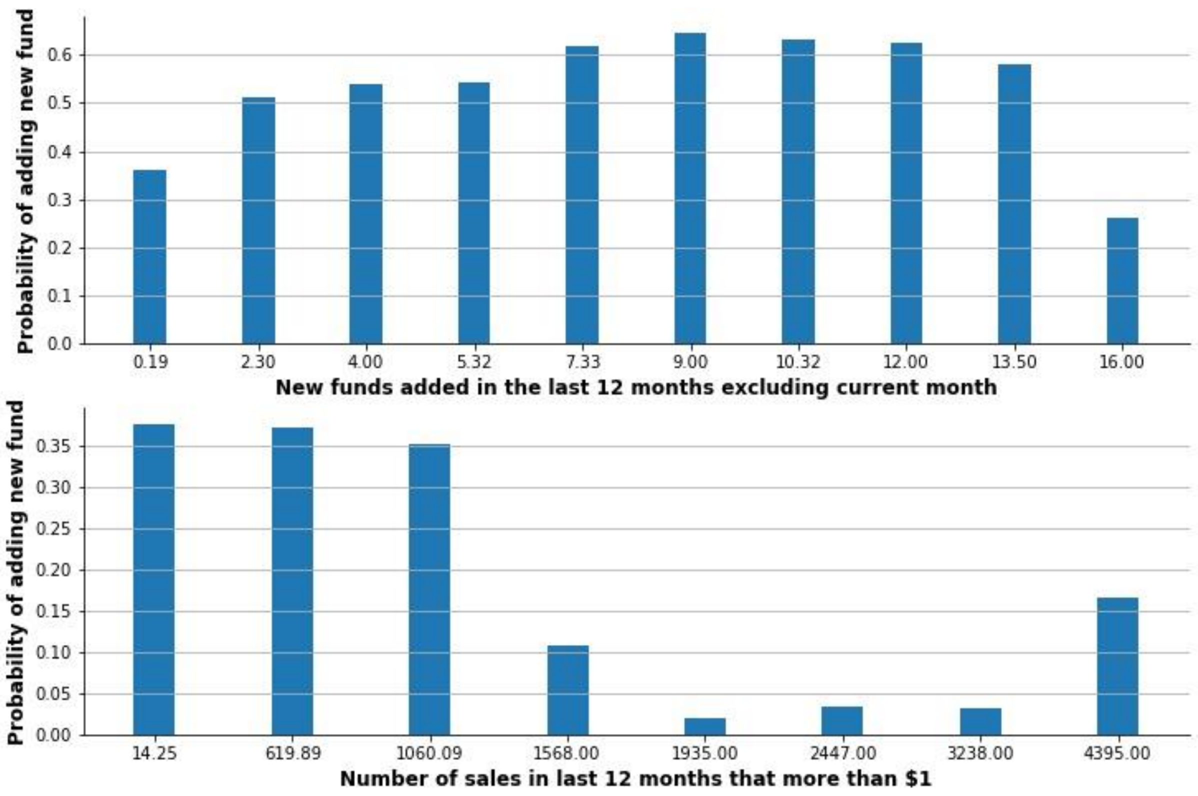

- Advisers with higher AUM will likely have higher sales next year.

nuveen
A TIAA Company

## Analysis Results

Columbia University in the City of New York
Emeritus Institute of Management

# Relational charts (classification model)




- Advisors who added more funds in the last 12 months are likely to add a new fund in next year. However, those advisors who have highest number of funds added in the last 12 months are less probable to add a new fund in next year.
- Advisors that have smaller number of sales that more than $1 in the last 12 months are more likely to add a new fund in next year.

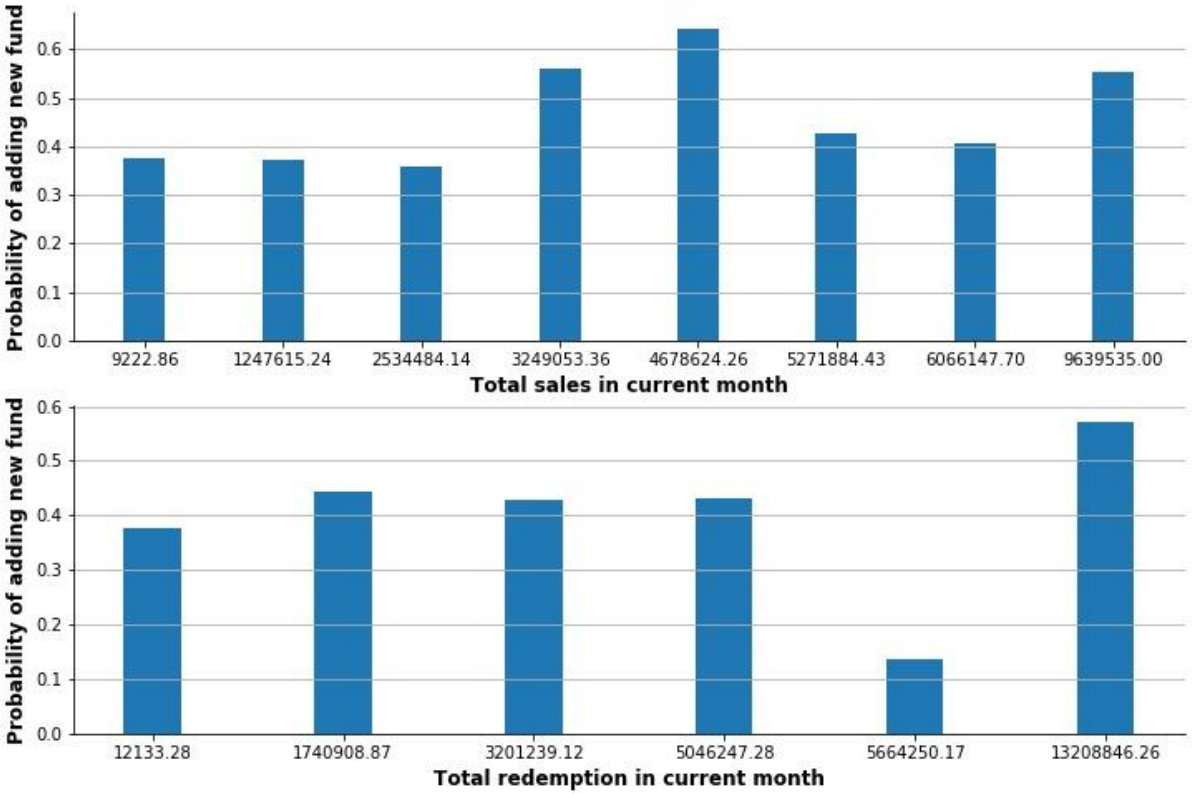
**Analysis Results**

nuveen
A TIAA Company

Columbia University
in the City of New York

Emeritus
Institute of Management

# Relational charts (classification model)




- Advisors with higher total sales in current month have slightly higher probability of adding new fund in next 12 months.

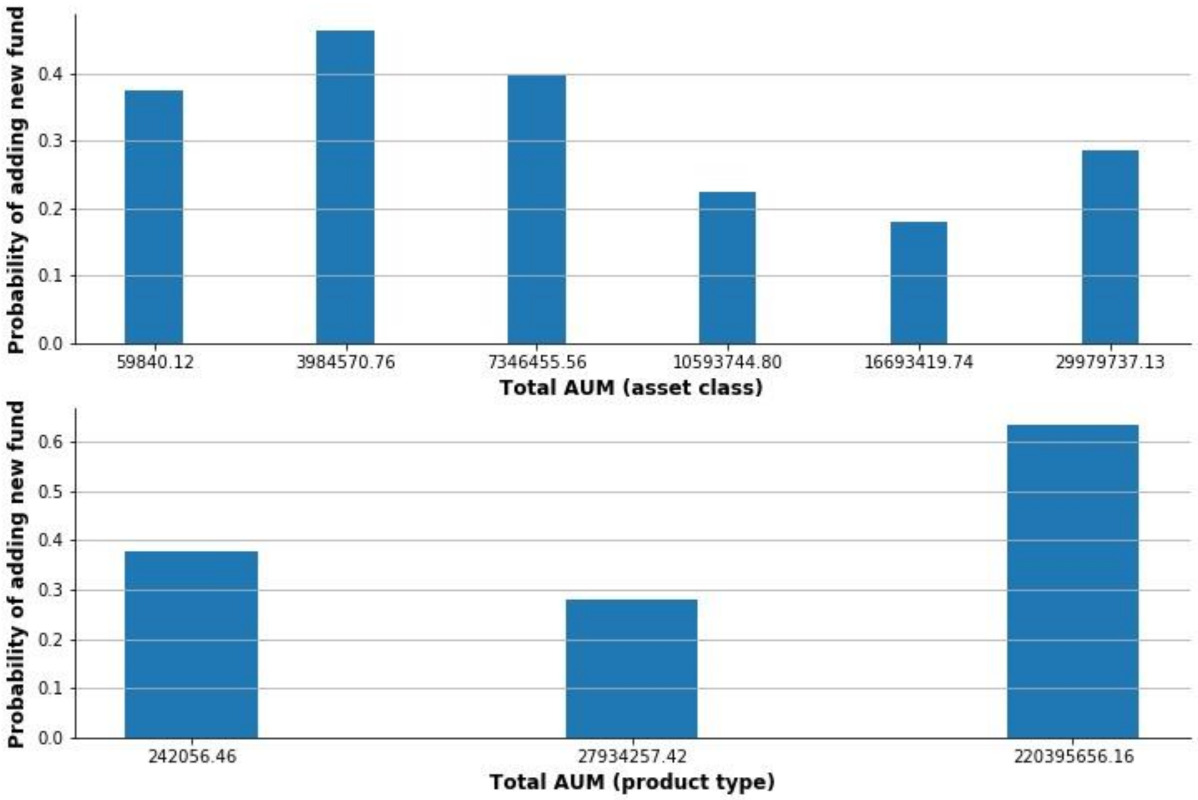
nuveen
A TIAA Company

## Analysis Results

Columbia University in the City of New York

Global Ivy Education Emeritus Institute of Management

# Relational charts (classification model)




- Advisors with higher total AUM in asset class MULTIPLE, FIXED INCOME TAXABLE, TARGET and REAL ESTATE have less probability of adding new fund in next 12 months. However, advisors with higher total AUM in product class UIT and MF have higher probability of adding new fund in next 12 months.



## Analysis Results

# Inference

- Advisors that have higher sales in current month will likely have higher sales next year.
- Advisors that have higher net sales in current month will likely increase sales next year and vice-versa.
- Advisors with higher total sales in current month have slightly higher probability of adding new fund in next 12 months.
- Advisors that have positive net number of sales that are more than $1 in the last 12 months will likely increase their sales next year. But advisors with positive net number of sales that more than $10K in the last 12 months will likely decrease their sales next year.
- Advisors that have smaller number of sales that more than $1 in the last 12 months are more likely to add a new fund in next year.
- Advisors who added more funds in the last 12 months are likely to add a new fund in next year. However, those advisors who have highest number of funds added in the last 12 months are less probable to add a new fund in next year.
- Advisers with higher AUM will likely have higher sales next year.
- Advisors with higher total AUM in asset class MULTIPLE, FIXED INCOME TAXABLE, TARGET and REAL ESTATE have less probability of adding new fund in next 12 months. However, advisors with higher total AUM in product class UIT and MF have higher probability of adding new fund in next 12 months.

# Recommendations

Monitor advisor performance monthly. Retain advisors that continuously increase their net sales. Define a development strategy that will target advisors who show decrease (or no increase) in their net sales.

Retain advisors that have high net number of sales that are more than $1 in the last 12 months. However, advisors with the highest net number of sales should be targeted for development of further sales increase and for adding a new funds.

Retain advisors that show higher number of funds added in the last 12 months but pay a special attention to advisors who added the highest number of funds as they likely to add less in next year.

Retain advisors with higher AUM but target those who have higher AUM in asset class MULTIPLE, FIXED INCOME TAXABLE, TARGET and REAL ESTATE to encourage them to continue add more funds.

# Appendix

Technical Report

# Contents

# 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%
```

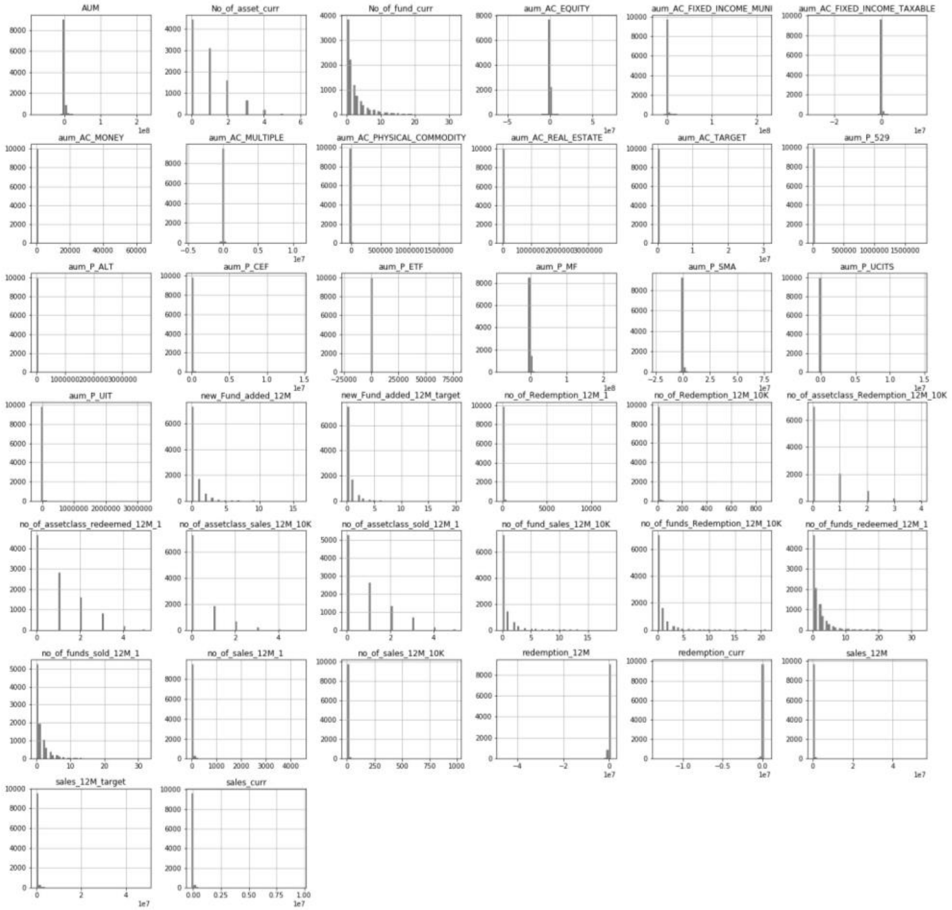
```
data.isnull().sum()

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.

EDA

# Features distributions




- The data is highly biased toward zero. But also, there are a noticeable number of samples with abnormaly big 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

EMERITUS INSTITUTE OF MANAGEMENT

nuveen
A TIAA Company

## Feature Engineering

- Drop all samples that have negative sales and positive 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 entire dataset (including target variable for regression model).
- Apply one-hot encoder for target variable for classification model.

# 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

The models were found to perform better on the subset_1.

# 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*

nuveen
A TIAA Company

## Feature Selection

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

EMERITUS
INSTITUTE OF MANAGEMENT

# Model Selection

- ## Regression model
  - **Model and hyperparameters**

```
gbr = GradientBoostingRegressor(loss='huber',
                                criterion='friedman_mse',
                                learning_rate=0.1,
                                n_estimators=100,
                                random_state=random_state)
```

  - **Cross-validation strategy**
    - Train-test split: 70% train / 30% test
    - Repeated Kfold cross-validation: 5 folds & 10 repeats
    - Metrics: MAE, R2, explained variance

- ## Classification model
  - **Model and hyperparameters**

```
gbr = GradientBoostingClassifier(loss='deviance',
                                 criterion='friedman_mse',
                                 learning_rate=0.1,
                                 n_estimators=100,
                                 random_state=random_state)
```

  - **Cross-validation strategy**
    - Train-test split: 70% train / 30% test
    - Repeated stratified Kfold cross-validation: 5 folds & 10 repeats
    - Metrics: accuracy, ROC-AUC

# Training

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

# Regression model

nuveen
A TIAA Company

## Training & Evaluation

### Features importance

| Feature | drop |
|---|---|
| sales_curr | 0.056292 |
| redemption_curr | 0.005405 |
| no_of_Redemption_12M_1 | 0.004925 |
| no_of_sales_12M_10K | 0.003073 |
| pos_aum_AC_FIXED_INCOME_TAXABLE | 0.001714 |
| neg_aum_P_SMA | 0.001693 |
| neg_aum_AC_FIXED_INCOME_TAXABLE | 0.001027 |
| pos_aum_AC_EQUITY | 0.000890 |
| no_of_Redemption_12M_10K | 0.000825 |
| no_of_sales_12M_1 | 0.000576 |
| pos_aum_AC_TARGET | 0.000483 |
| pos_aum_P_CEF | 0.000356 |
| pos_aum_P_ETF | 0.000237 |
| neg_aum_AC_MULTIPLE | -0.000014 |
| pos_aum_AC_PHYSICAL_COMMODITY | -0.000134 |
| new_Fund_added_12M | -0.000632 |
| pos_aum_AC_MULTIPLE | -0.000653 |
| pos_aum_P_SMA | -0.001013 |
| neg_aum_P_MF | -0.001065 |
| pos_aum_AC_REAL_ESTATE | -0.001136 |
| neg_aum_AC_FIXED_INCOME_MUNI | -0.001399 |

Important features

### CV results (all features)

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

| | mae | explained_variance | r2 |
|---|---|---|---|
| mean_train | 2.207898 | 0.642501 | 0.641211 |
| std_train | 0.024959 | 0.005219 | 0.005276 |
| mean_test | 2.362490 | 0.598390 | 0.596744 |
| std_test | 0.063948 | 0.022856 | 0.023132 |

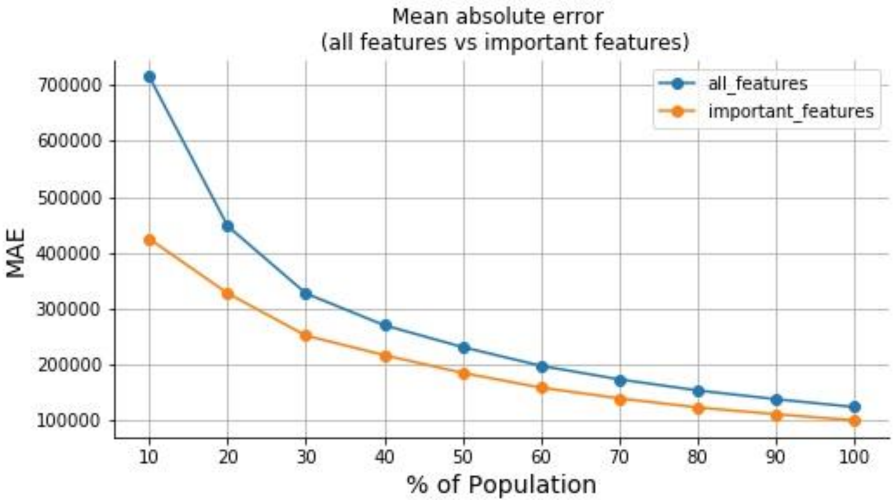
```
Validation on the test set
==============================================
mean_absolute_error:  2.393461641149341
explained_variance_score:  0.5825111780287197
r2_score:  0.5824634994425928
```

### CV results (important features)

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

| | mae | explained_variance | r2 |
|---|---|---|---|
| mean_train | 2.253700 | 0.634185 | 0.632711 |
| std_train | 0.026154 | 0.005408 | 0.005467 |
| mean_test | 2.389155 | 0.595501 | 0.593682 |
| std_test | 0.064844 | 0.023310 | 0.023634 |

```
Validation on the test set
==============================================
mean_absolute_error:  2.419445734427349
explained_variance_score:  0.5819876478478977
r2_score:  0.5819081454357539
```




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

# Classification model

nuveen
A TIAA Company

## Training & Evaluation

### Features importance

| | |
|---|---|
| new_Fund_added_12M | 0.007850 |
| neg_aum_AC_MULTIPLE | 0.002617 |
| neg_aum_P_UIT | 0.002568 |
| pos_aum_AC_FIXED_INCOME_TAXABLE | 0.002374 |
| pos_aum_AC_TARGET | 0.001599 |
| no_of_sales_12M_1 | 0.001599 |
| pos_aum_AC_MULTIPLE | 0.001066 |
| neg_aum_AC_REAL_ESTATE | 0.001017 |
| pos_aum_P_UIT | 0.001017 |
| redemption_curr | 0.000775 |
| sales_curr | 0.000775 |
| neg_aum_P_MF | 0.000049 |
| neg_aum_AC_FIXED_INCOME_TAXABLE | -0.000193 |
| pos_aum_AC_EQUITY | -0.000533 |
| pos_aum_AC_PHYSICAL_COMMODITY | -0.000824 |
| pos_aum_AC_REAL_ESTATE | -0.000824 |
| neg_aum_AC_PHYSICAL_COMMODITY | -0.001259 |
| pos_aum_P_ETF | -0.001308 |
| no_of_sales_12M_10K | -0.002277 |
| neg_aum_P_ETF | -0.003150 |
| pos_aum_P_529 | -0.003245 |
| no_of_Redemption_12M_1 | -0.003876 |
| neg_aum_AC_FIXED_INCOME_MUNI | -0.004651 |
| no_of_Redemption_12M_10K | -0.005233 |
| neg_aum_P_529 | -0.005573 |
| pos_aum_P_SMA | -0.007801 |
| pos_aum_P_CEF | -0.008236 |

Important features

### CV results (all features)

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

| | accuracy | roc_auc |
|---|---|---|
| mean_train | 0.688677 | 0.789007 |
| std_train | 0.008173 | 0.003681 |
| mean_test | 0.638096 | 0.697467 |
| std_test | 0.014014 | 0.013697 |

Validation on the test set
============================
accuracy: 0.6442804428044281
roc_auc: 0.6423822662687464

### CV results (important features)

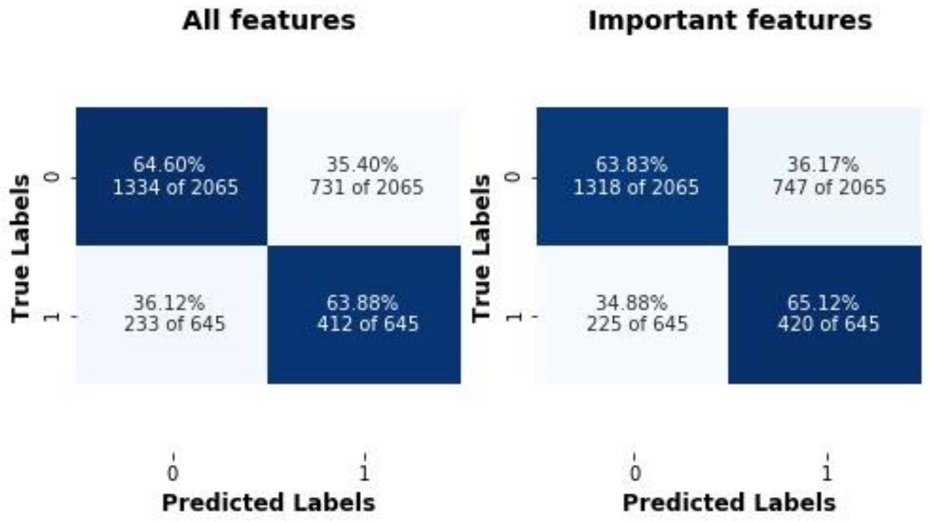
CV Results
============================

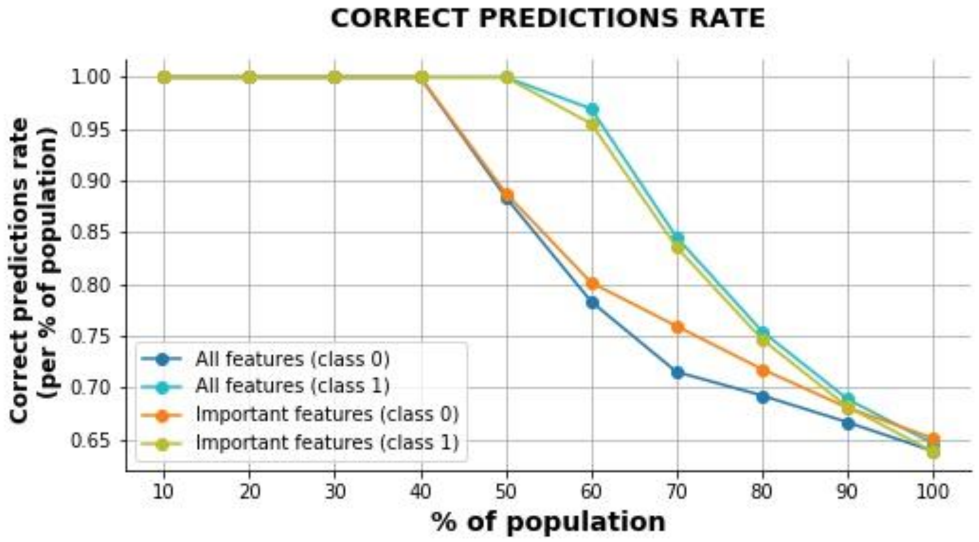
| | accuracy | roc_auc |
|---|---|---|
| mean_train | 0.681368 | 0.765319 |
| std_train | 0.008018 | 0.004119 |
| mean_test | 0.639520 | 0.683099 |
| std_test | 0.014126 | 0.016493 |

Validation on the test set
============================
accuracy: 0.6413284132841328
roc_auc: 0.6447097246466581

**NOTE:** Data imbalance was compensated by using sample weights.
The weights of class 0 samples were set to be equal to class 1 prior and vice-versa.

# Classification model (continued)




## Training & Evaluation

nuveen
A TIAA Company

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

Compressed (zipped) Folder

nuveen
A TIAA Company

Attachments