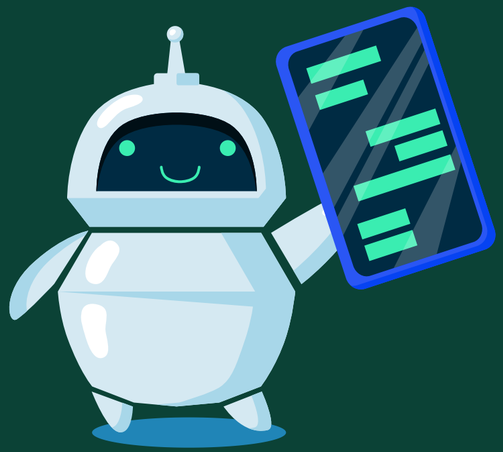
Investor Risk Tolerance Prediction and Robo-Advisory Dashboard

Presented by: Anand Ramchandani



Contents

WHAT IS A ROBO-ADVISOR?

Problems Observed with Robo-Advisors

01

Long and tedious risk assessment questionnaires that require a lot of time and manual input

02

Studies show* these questionnaires are prone to error due:

- Behavioural biases
- People are poor judges of their own risk perceptions
- Questionnaires sometimes reinforce these biases

03

May not reflect investors risk tolerance during times of financial turmoil

*Studies references cited at end of presentation

Background on Risk Profiling

Risk Tolerance	Ability to take Risk
• Subjective aspects • Willingness to Take Risk • Investors Personality • Personal goals and priorities	• Objective Factors • Time Horizon • Need for Income • Family situation, etc.

Using Machine Learning...

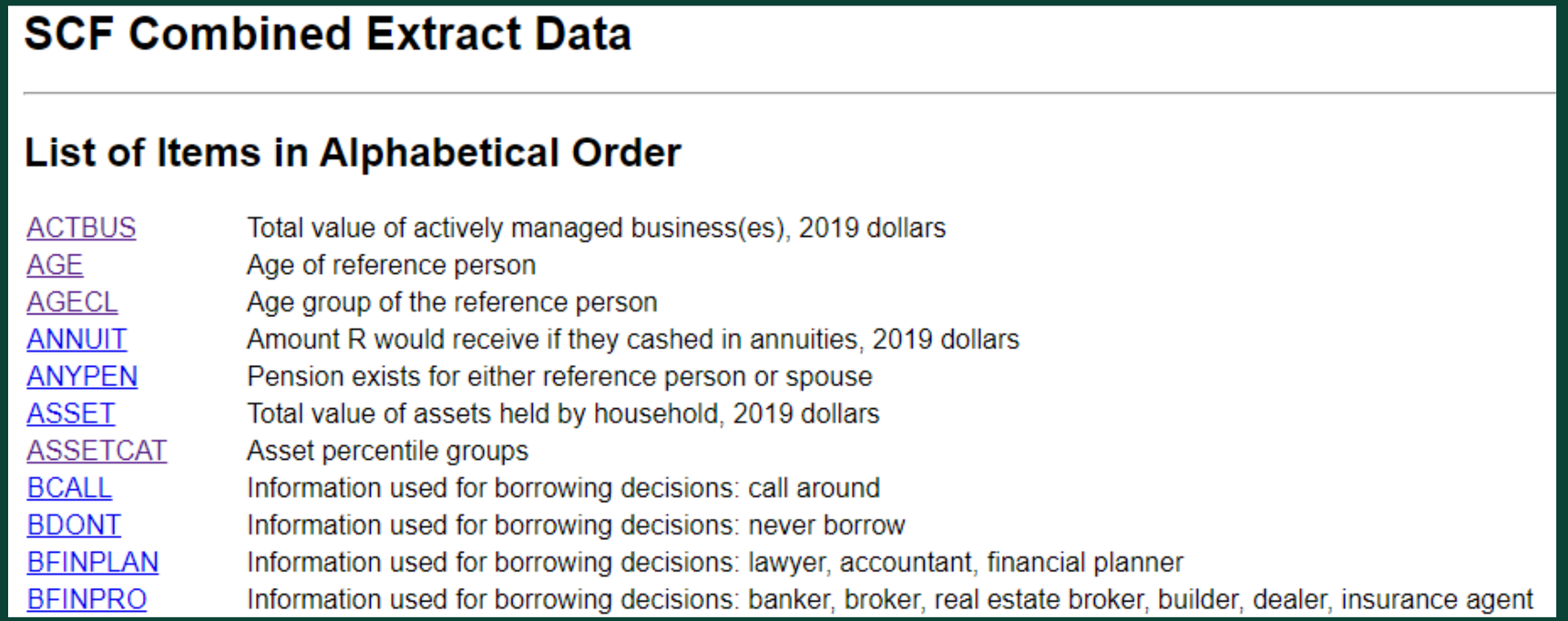
- Can machine learning provide a better assessment of an investors risk profile than a risk tolerance questionnaire?
- Can the machine learning algorithm better reflect how a client would deal in different financial market scenarios?
- Can we automate the investment and portfolio management process with minimal need for human intervention?

Problem Statement

A Robo-Advisory firm hires a data scientist to create a machine learning model that is able to predict an investors risk tolerance from key feature attributes. Along with the investors stated willingness to take risk, a Robo Advisory prototype dashboard will also be created.

Dataset Used

- Dataset comes from the US Federal Reserve - Survey of Consumer Finances
- Statistical survey taken every 3 years of US Households
- Latest dataset taken in 2019
- Dataset used in this project from 2007 to 2019
- 2019 Dataset - 351 features and 28,885 Households



SCF Combined Extract Data

List of Items in Alphabetical Order

ACTBUS	Total value of actively managed business(es), 2019 dollars
AGE	Age of reference person
AGECL	Age group of the reference person
ANNUIT	Amount R would receive if they cashed in annuities, 2019 dollars
ANYPEN	Pension exists for either reference person or spouse
ASSET	Total value of assets held by household, 2019 dollars
ASSETCAT	Asset percentile groups
BCALL	Information used for borrowing decisions: call around
BDONT	Information used for borrowing decisions: never borrow
BFINPLAN	Information used for borrowing decisions: lawyer, accountant, financial planner
BFINPRO	Information used for borrowing decisions: banker, broker, real estate broker, builder, dealer, insurance agent

How do we calculate Risk Tolerance?

$$Risk\ Tolerance = \frac{Risky\ Assets}{Risky\ Assets + Risk\ Free\ Assets} = \frac{Risky\ Assets}{Total\ Assets}$$

Risky Assets

- Stocks, bonds, commodities, real estate, etc.

Risky Assets of Families	
NMMF	Total value of directly held pooled investment funds held by household, 2019 dollars
STOCKS	Total value of directly held stocks held by household, 2019 dollars
BOND	Total value of directly held bonds held by household, 2019 dollars

Risk Free Assets

- Checking and saving balances, certificates of deposit, other cash balances, etc.

Risk-Free Assets of Households	
CDS	Total value of certificates of deposit held by household, 2019 dollars
LIQ	Total value of all types of transactions accounts, 2019 dollars
SAVBND	Total value of savings bonds held by household, 2019 dollars
CASHLI	Total cash value of whole life insurance held by household, 2019 dollars

Attributes that could be correlated with Risk Tolerance

Demographic Attributes	Financial Attributes	Behavioural Attributes
Age	Income	Willingness to Take Risk
Education	Spending	
Occupation	Net Worth	
Marital Status / No. of Children		

Data Cleaning & EDA

- Dataset from Federal Reserve website is generally clean
- No null values
- No non-numerical data
- Years Analysed: 2019, 2016, 2013, 2010, 2009 and 2007

1

Create columns in each year for:

- Risky Assets
- Risk-Free Assets
- Risk Tolerance

2

Remove Households with Null Values in resultant Risk Tolerance*

3

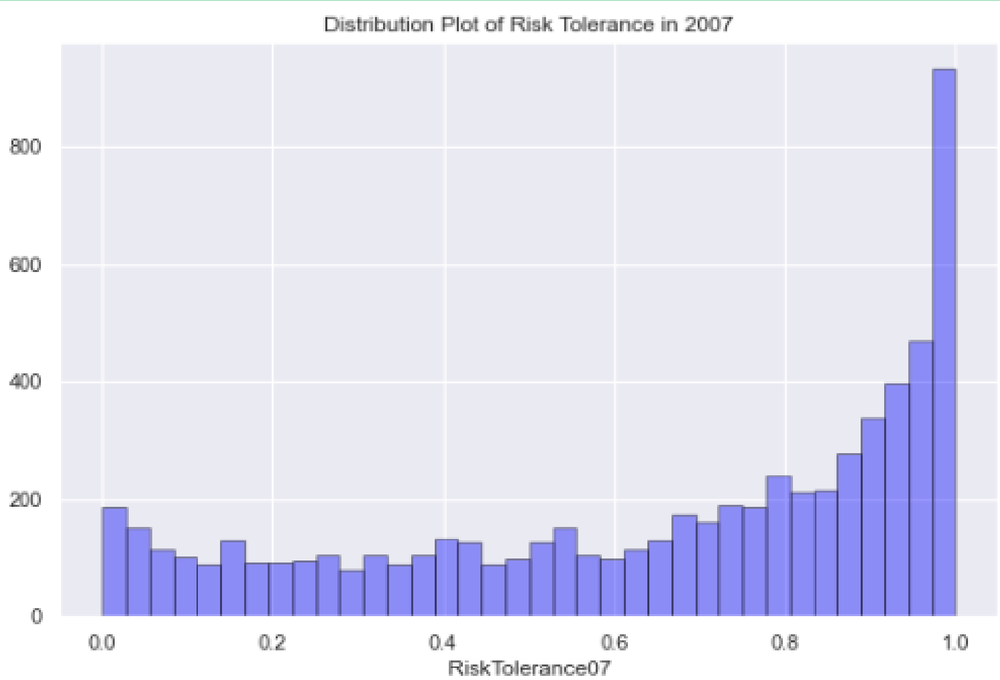
Analyse Risk Tolerance Distribution Plot

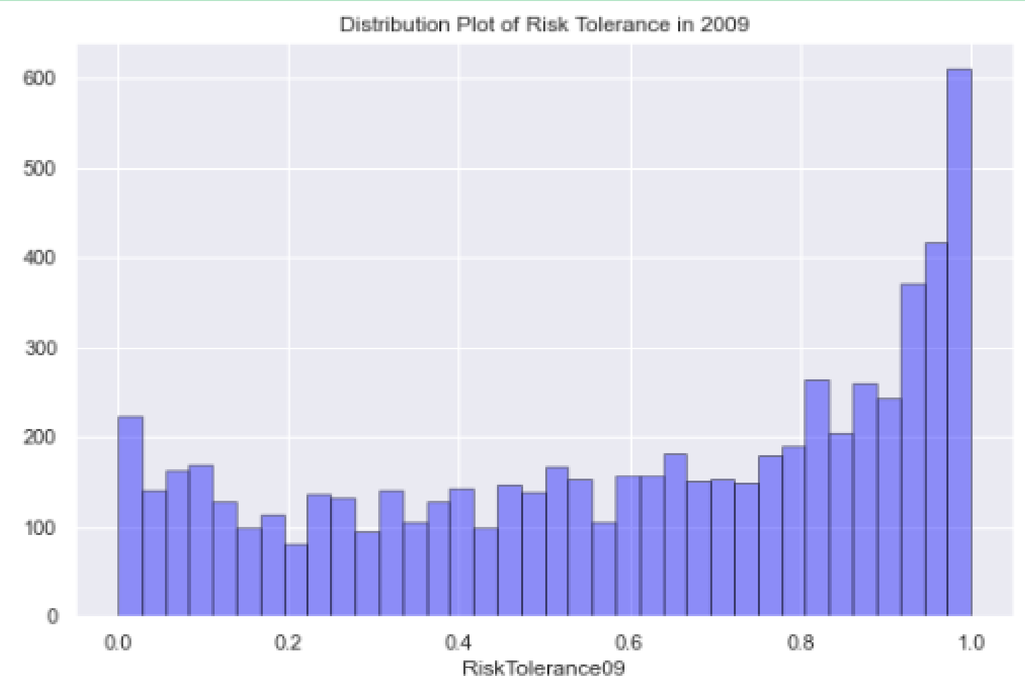
4

Merge all Datasets into 1 Big Dataset!

*Some households have no Risky or Risk Free Assets, resulting in 0 in denominator and produces a null value for Risk Tolerance

EDA – Risk Tolerance Before and After the 2008 Financial Crisis





2007

2009

Feature Selection & Engineering

Top 10 Features Correlated with Risk Tolerance (Precursory Analysis)

1. Alternate net worth percentile groups
2. Net worth percentile groups
3. Asset percentile groups
4. Age of reference person
5. Age group of the reference person
6. Alternate Normal income percentile groups
7. Number different companies in which hold stock
8. Alternate income percentile groups
9. Normal income percentile groups
10. Have a brokerage account

01 PROBLEMS

1. High degree of Multicollinearity
2. Features not easily available to average investor
3. Simple features required to reduce manual input time of data

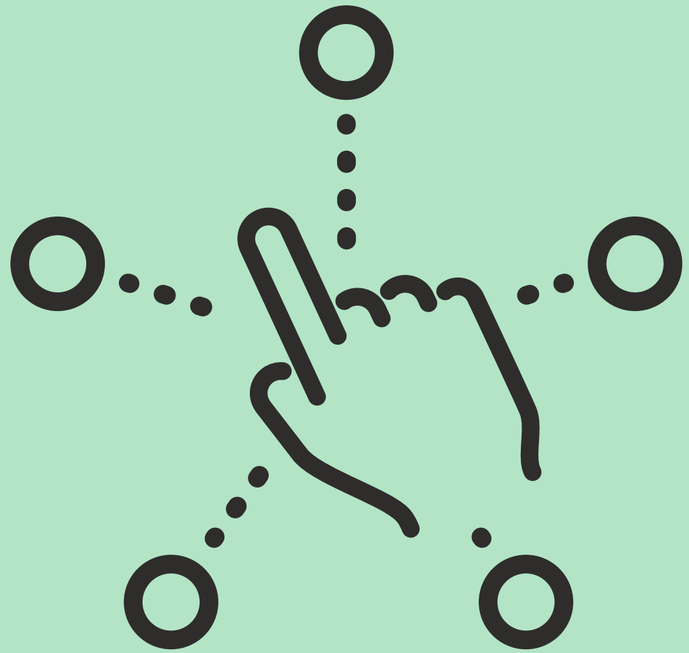
02 SOLUTION

Select features that are:

1. Highly correlated with Risk Tolerance
2. Easily available to average investor
3. Choose one feature between those that have high multicollinearity

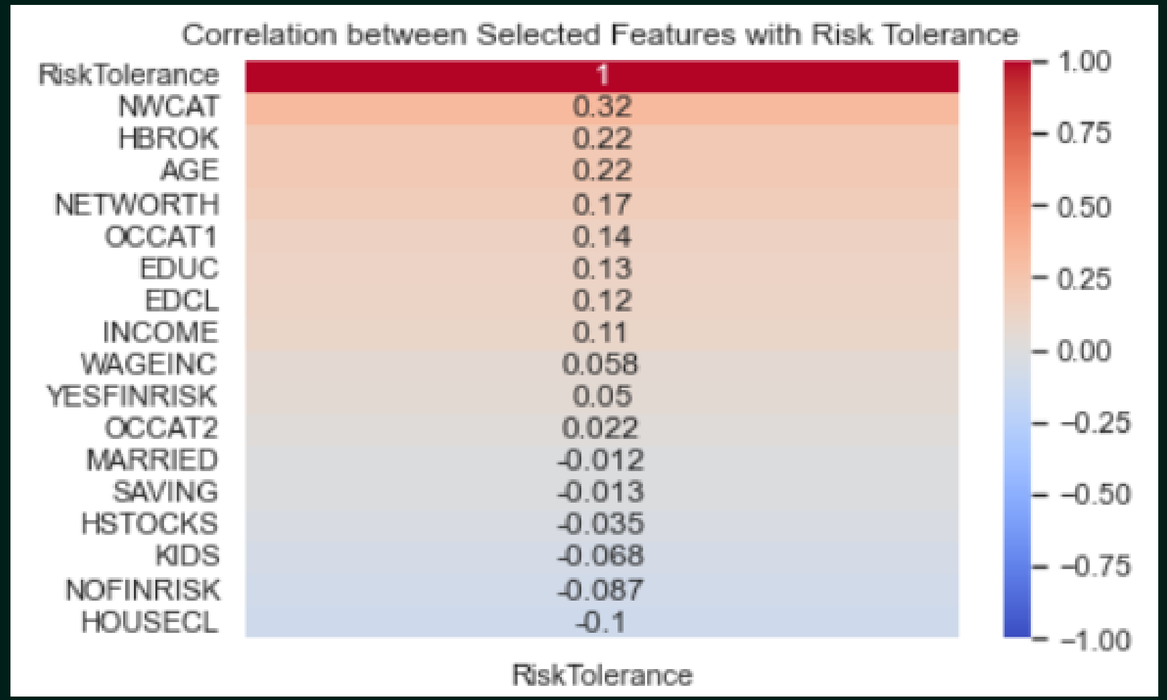
FEATURES SELECTED FOR FURTHER EXPLORATION AND CANDIDACY FOR SELECTION

1. Age
2. Education/Education Category
3. Marital Status
4. Number of Children
5. Occupation Category/Classification
6. Networth/Networth Percentile Groups
7. Income/Wage and Salary Income
8. Willingness to take risk
9. Savings
10. Whether or not the individual has stocks
11. Whether or not the individual has a brokerage account
12. Home-ownership category of household

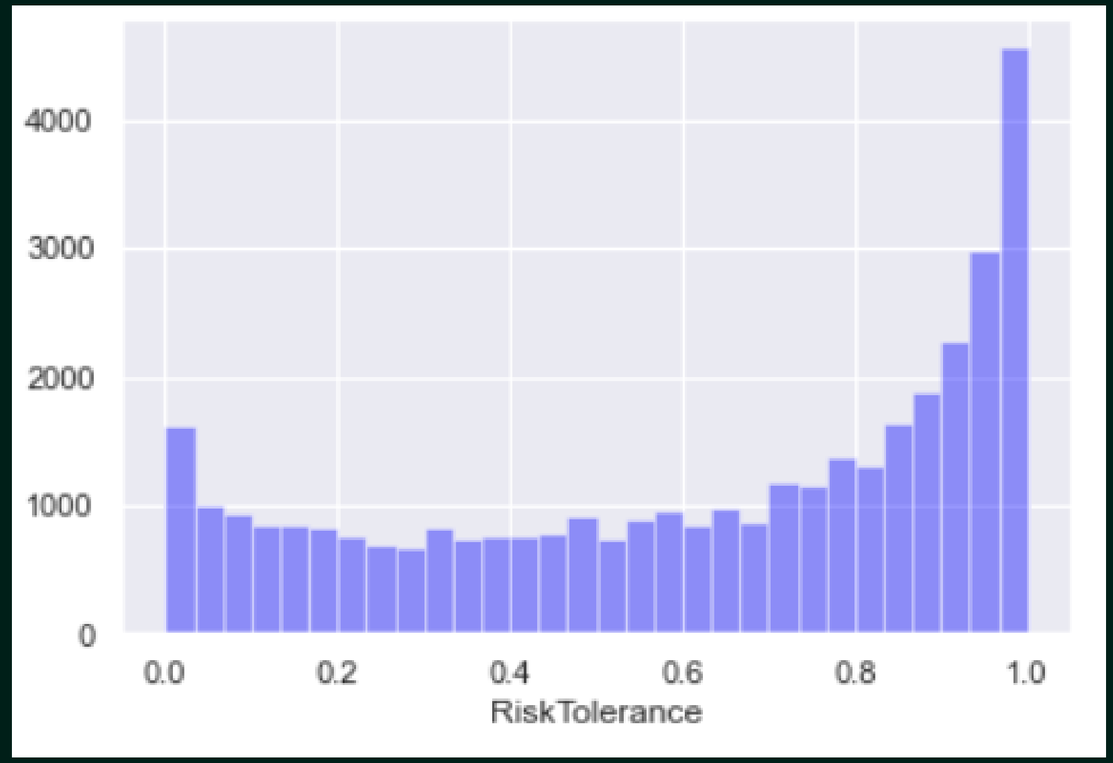


Merged Dataset

Correlation of Features to Risk Tolerance

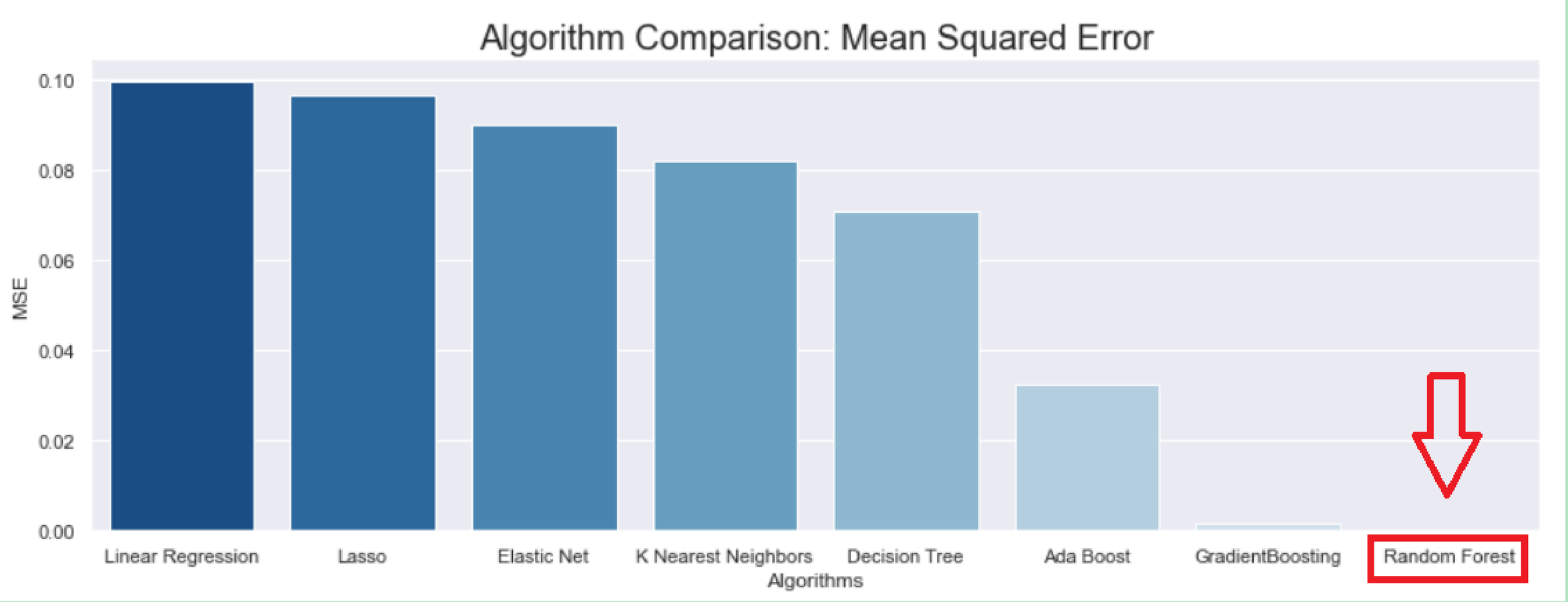


Risk Tolerance Distribution



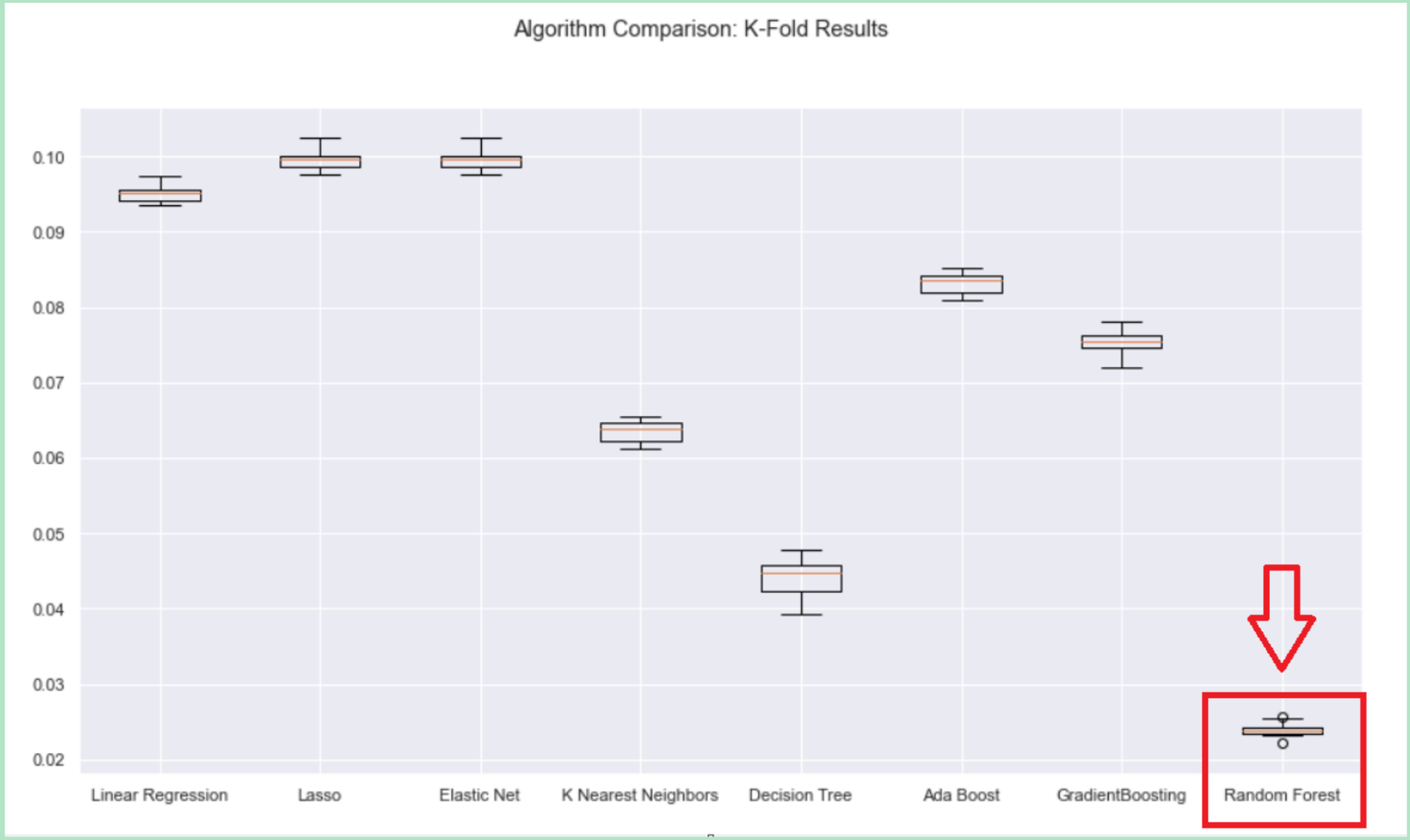
Model Evaluation

- We use Mean Squared Error (MSE) as the evaluation metric and select 10 number of folds for cross validation



Comparing Regression Models using Cross Validation Results

- Using Negative Mean Squared Error as scoring parameter, which returns the negated value of the metric [Lower the better]*



*https://scikit-learn.org/stable/modules/model_evaluation.html

MODEL SELECTION

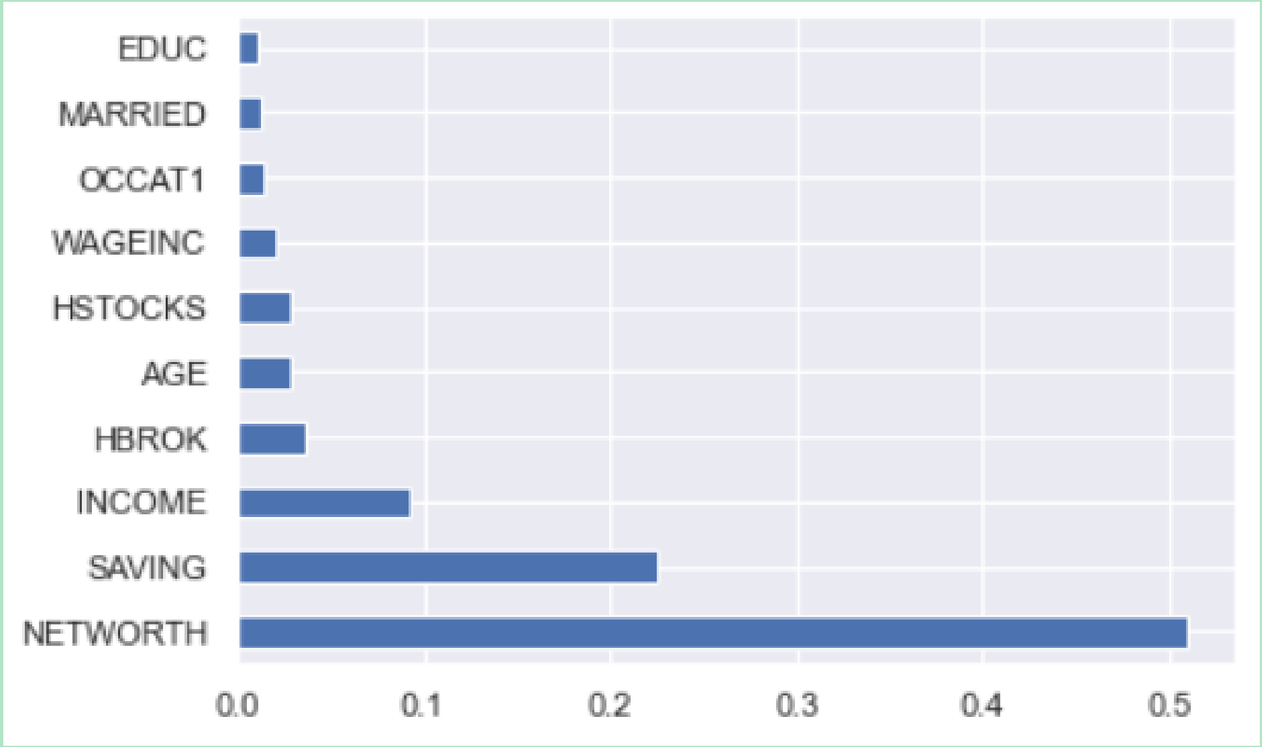
- Non-linear models perform better
- Implies non-linear relationship between Risk Tolerance and variables

Random Forest Model Chosen!

Random Forest Model
(Post-Hyperparameter Tuning)

MSE Train	MSE Test	MSE Diff (%)
0.0712	0.0734	-3.13 %

Model Feature Importance

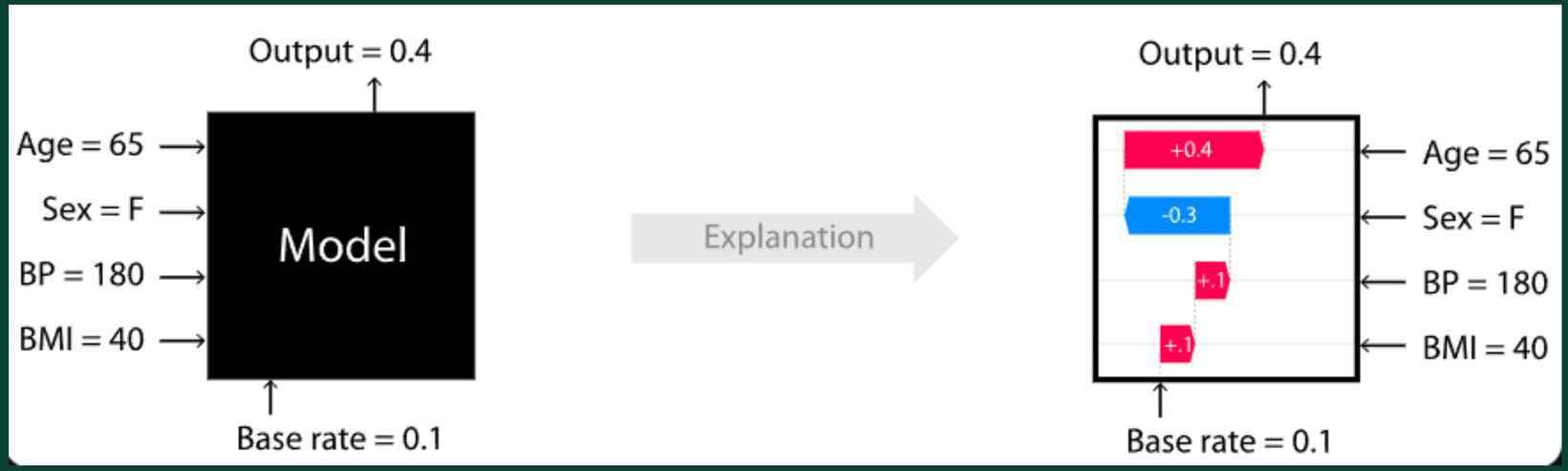


TOP 6 MOST IMPORTANT FEATURES BASED ON THE RANDOM FOREST MODEL

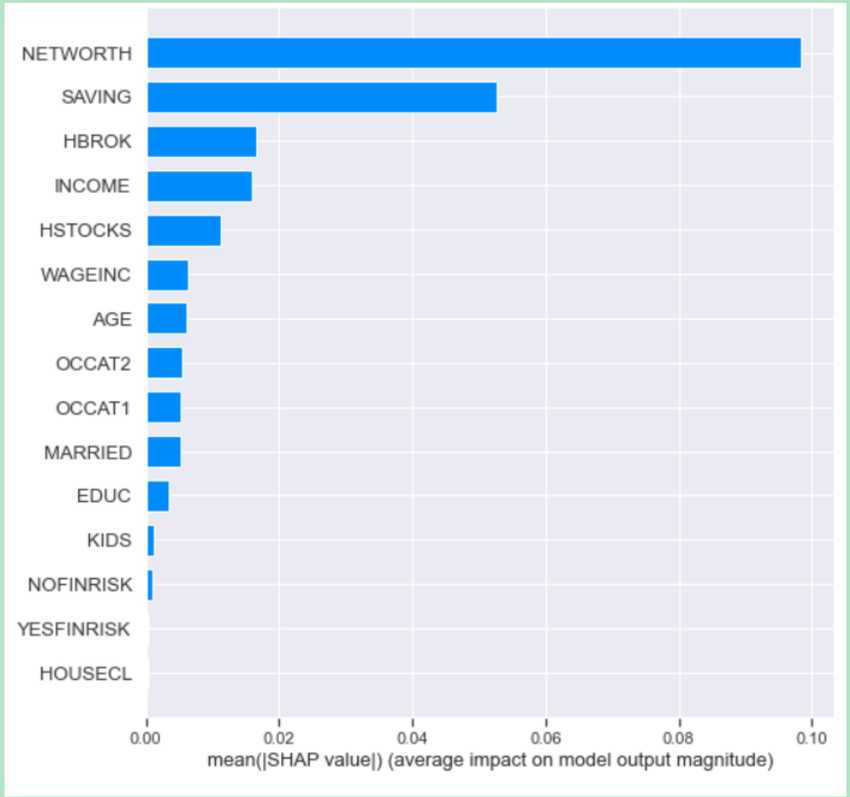
1. NETWORTH
2. SAVINGS
3. INCOME
4. WHETHER OR NOT ONE HAS A BROKERAGE ACCOUNT
5. AGE
6. WHETHER OR NOT ONE HAS STOCKS

Shapley Values for Model Interpretation

SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model.



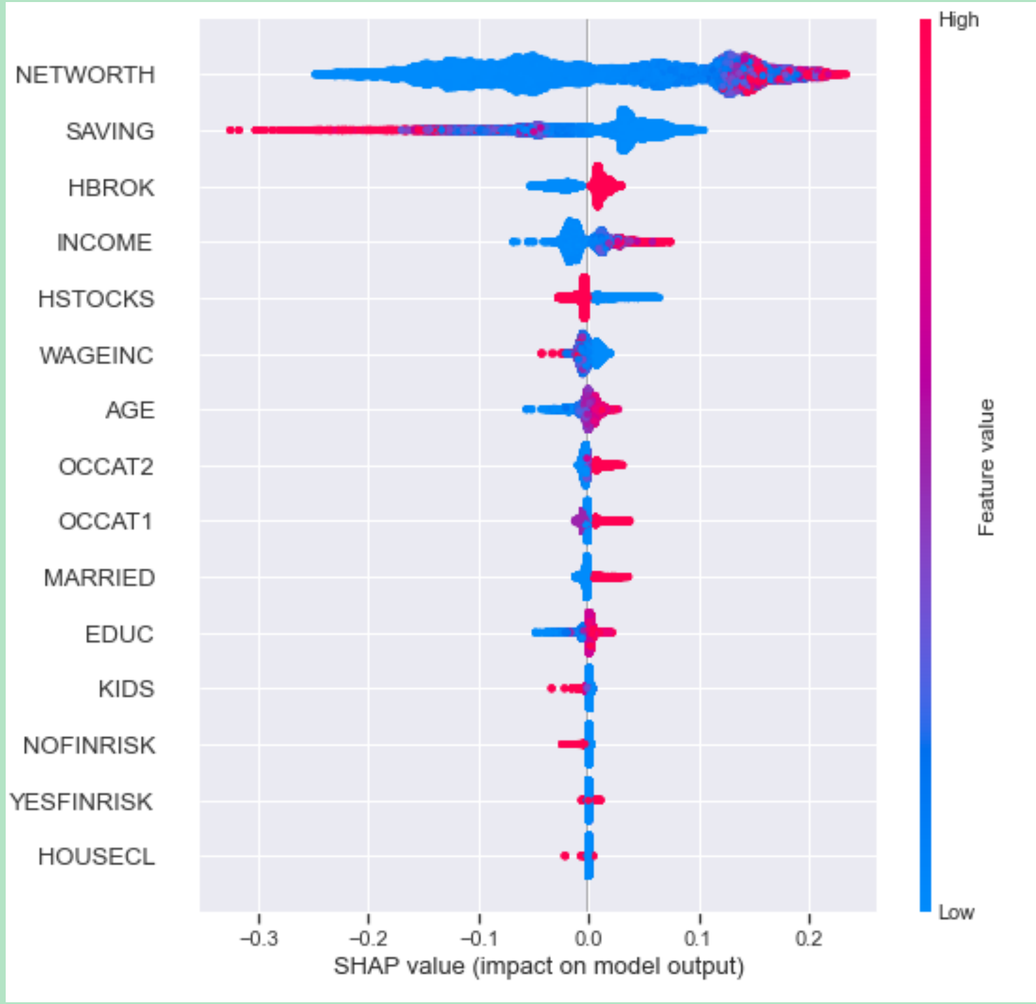
Shapley Feature Importance



TOP 6 MOST IMPORTANT FEATURES BASED ON THE RANDOM FOREST MODEL W/ SHAPLEY

1. NETWORTH
2. SAVINGS
3. WHETHER OR NOT ONE HAS A BROKERAGE ACCOUNT
4. INCOME
5. WHETHER OR NOT ONE HAS STOCKS
6. WAGES/SALARY

Summary effect of all Shapley Features



Limitations of this Model

1. Model is purely based on US Households
2. Based on Households and not individual investors
3. Figures not adjusted for inflation throughout the years
4. Model at risk of Data Drift
5. Willingness to take risk in dataset is Yes or No rather than a range
6. There are a lot of households with no savings
7. Regular ML Ops required to keep model relevant
8. Easy use by investor hinders predictive ability of model
9. Susceptible to new financial instruments, innovation and technology, like cryptocurrency, NFTs or even new laws

NETWORTH

RANGE:
-1,933,339.62-
1,688,567,600.0

SAVINGS

RANGE:
0.0 -
60,946,249.78

OCCUPATION

1 MANAGERIAL/PROFESSIONAL
2 TECHNICAL/SALES/SERVICES
3 OTHER
(INCL. WORKERS, OPERATORS, LABORERS, FARMERS, FORESTERS, FISHERS)
4 NOT WORKING/RETIRED

HAVE BROKERAGE ACCOUNT?

0 NO
1 YES

INCOME

RANGE:
0.0-42,5777,653.23

WILLING TO TAKE RISK?

0 NO
1 YES

ROBO-ADVISOR PROTOTYPE DASHBOARD!

Conclusions

- Net Worth and Savings have by far the largest impact on Risk Tolerance
- The higher ones Savings , the lower the Risk Tolerance
- This model could be used to not only decipher one's Risk Tolerance but also to determine how each feature inform this prediction
- This could possibly used to target customers directly to portfolios in line with their Risk Tolerance

Recommendations

- Longer time period, more features and more households/investor data
- Run more GridSearch to Hyperparameter Tune for each model
- Gather data from other countries besides the US
- Gather data from investors rather than just households
- Adjust figures for inflation
- Use range for willingness to take risk
- Possible design improvements to make model more versatile and susceptible to market conditions and new financial instruments

Studies showing issues with Risk Questionnaires

- https://www.cfainstitute.org/-/media/documents/article/rf-brief/rfbr-v2-n1-1-pdf.ashx
- https://www.kitces.com/blog/risk-tolerance-questionnaire-and-risk-profiling-problems-for-financial-advisors-planplus-study/
- https://assets.ey.com/content/dam/ey-sites/ey-com/en_us/topics/wealth-and-asset-management/ey-behavioral-finance-whitepaper-v1.pdf?download

Thank You!

ANY QUESTIONS?