Predicting Financial Well-Being

Matt Calvert and Jeff DuBois

Project Overview

What is Financial Well-Being?

In 2016, the Consumer Financial Protection Bureau (CFPB) led a rigorous research effort to develop the following consumer-driven definition of financial well-being:

> ***"A state of being wherein a person can fully meet current and ongoing financial obligations, can feel secure in their financial future, and is able to make choices that allow them to enjoy life."***

The CFPB then created and tested a set of questions – a "scale" – to measure financial well-being.

Four Elements of Financial Well-Being?

	Present	Future
Security	Control over your day-to-day, month-to-month finances	Capacity to absorb a financial shock
Freedom of choice	Financial freedom to make choices to enjoy life	On track to meet your financial goals

The CFPB created the questionnaire and a scoring method as a tool that can help you take stock of your financial well-being and reassess over time.

Financial Well-Being Scale

The CFPB collected responses to the financial well-being scale in a nationwide survey of U.S. adults. The following 10 questions were ultimately selected for the final version of the scale, based on item performance, overall scale reliability, and item content (ensuring that the four substantive areas of interest were all represented)

Part 1: How well does this statement describe you or your situation?

1. I could handle a major unexpected expense
2. I am securing my financial future
3. Because of my money situation, I feel like I will never have the things I want in life
4. I can enjoy life because of the way I'm managing my money
5. I am just getting by financially
6. I am concerned that the money I have or will save won't last

Part 2: How often does this statement apply to you?

1. Giving a gift for a wedding, birthday or other occasion would put a strain on my finances for the month
2. I have money left over at the end of the month
3. I am behind with my finances
4. My finances control my life

Additional Measures in the Survey

The survey also collected a host of other measures including the following:

- Individual characteristics
- Household and family characteristics
- Income and employment characteristics
- Savings and safety nets
- Financial experiences
- Financial behaviors, skills, and attitudes (which included the following tools)
    - CFPB Financial Skill Scale (10 items)
    - Knoll and Houts Financial Knowledge Scale (10 items)
    - Lusardi and Mitchell Financial Knowledge Scale (3 items)

Project Objectives

- Understand what is financial well-being and the factors that influence it
- Create a machine learning model to classify individuals by their financial well-being scores and predict their scores based on these factors
- Discover insights on the relationship of these factors and areas of further study

Question: If a model works, can this be used to identify individuals in need of financial education or resources?

Process

Data Exploration and Preprocessing

Initial dataset: csv file, 216 columns x 6,394 rows

Load and Review Data

1. Read data from PostgreSQL database table and load into a pandas DataFrame
2. Create lists to store needed columns in groups/measures for future use
3. Review scoring scales and summary statistics

Preprocessing

1. Drop unneeded columns and rows/columns with negative values
2. Encode columns with non ordinal values using pandas.get_dummies

Final dataset for analysis: 114 columns x 5,931 rows

Model Building and Data Analysis

1. Load preprocessed data
2. Define target (FWB score) and features
3. Define functions for repetitive tasks
4. Create and evaluate linear regression model to predict FWB score
5. Use pandas.qcut for quantile binning of FWB scores
6. Create and evaluate various decision tree and random forest models by adjusting quantile binning and hyperparameter tuning

Early Lessons Learned

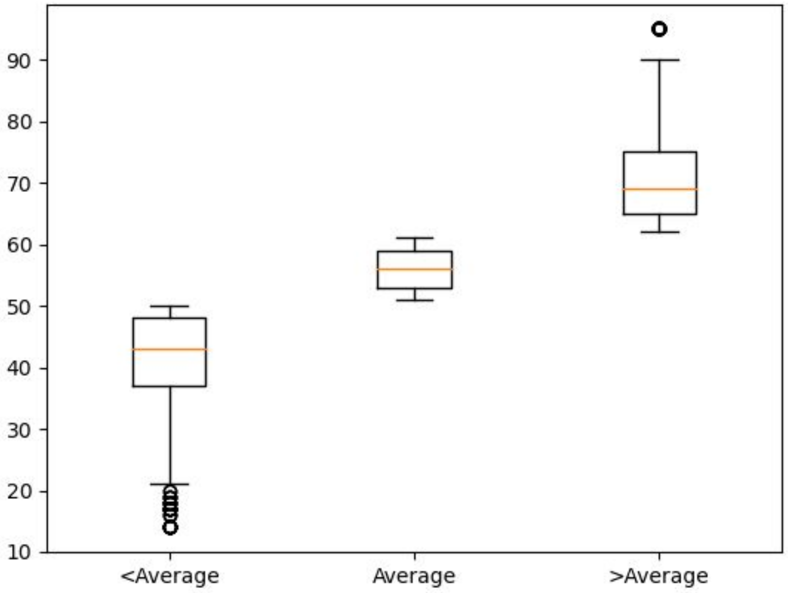
- Too much emphasis on achieving a specific accuracy score instead of building a model that addresses our initial objectives
- Not considering data leakage when selecting features

Analysis

Financial Well-Being Classes

The dataset was broken in 3 classes based on the 25th, 50th and 75th quartiles of the Financial Well-Being Score.

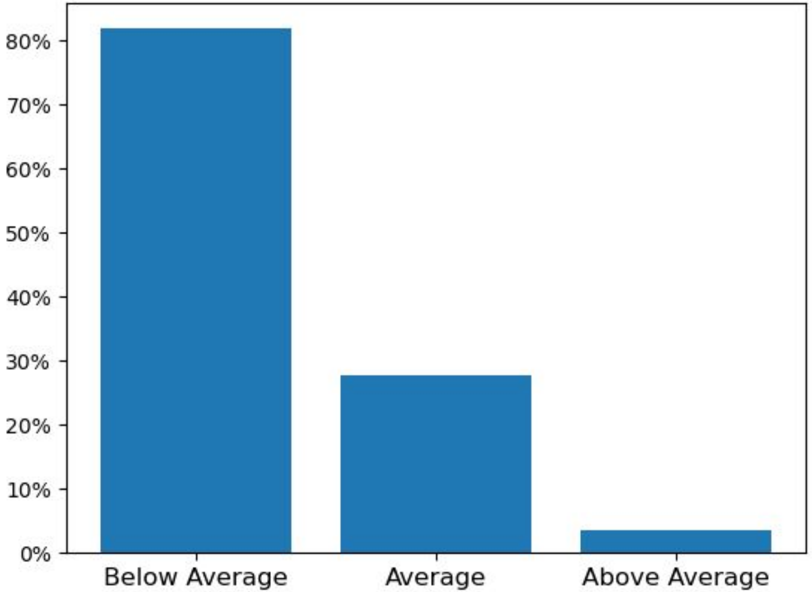
	Below Average	Average	Above Average
Samples	1483	1491	1474
Score	< 48	48 - 65	> 65
Mean	41.22	56.06	71.04
Min	14	51	62
Max	50	61	95



Difficulty Making Ends Meet by Class

In a typical month, how difficult is it for you to cover your expenses and pay all your bills?

1. Not at all difficult
2. Somewhat difficult
3. Very difficult



Datasets and Model Accuracy

1. All variables except the 3 additional financial scales
2. Removed variables variables related to FWB questions
3. Removed household income, SNAP and federal poverty level variables

2 Classes			
	Dataset 1	Dataset 2	Dataset 3
Decision Tree	0.84	0.71	NA
Random Forest	0.86	0.78	NA

3 Classes			
	Dataset 1	Dataset 2	Dataset 3
Decision Tree	0.65	0.57	0.55
Random Forest	0.69	0.62	0.60

Note: Some variables and records were removed in which the respondent refused to answer the question, gave an invalid response, or did not know the answer.

Model performance with 2 Classes (Good, Bad)

Decision Tree

Accuracy Score

```
# DT Model #1: Default settings
model_dt2c.append(DecisionTreeClassifier(max_depth=None, max_features=None, min_samples_split=2, min_samples_leaf=1))
```

0.65

```
# DT Model #1: Simple tree
model_dt2c.append(DecisionTreeClassifier(max_depth=4, max_features=None, min_samples_split=5, min_samples_leaf=3))
```

0.69

```
# DT Model #3: Best results after testing
model_dt2c.append(DecisionTreeClassifier(max_depth=5, max_features=None, min_samples_split=10, min_samples_leaf=5))
```

0.71

```
# DT Model #4: One of the best estimators from RandomizedSearchCV
model_dt2c.append(DecisionTreeClassifier(max_depth=7, max_features=None, min_samples_split=4, min_samples_leaf=5))
```

0.70

Random Forest

Accuracy Score

```
# RF Model #1
model_rf2c.append(RandomForestClassifier(n_estimators=200, random_state=78, bootstrap=True,
                                         max_depth=5, min_samples_split=2, min_samples_leaf=1))
```

0.77

```
# RF Model #2
model_rf2c.append(RandomForestClassifier(n_estimators=500, random_state=78, bootstrap=False,
                                         max_depth=10, min_samples_split=5, min_samples_leaf=3))
```

0.78

```
# RF Model #3: One of the best estimators from RandomizedSearchCV
model_rf2c.append(RandomForestClassifier(n_estimators=250, random_state=78, bootstrap=False,
                                         max_depth=20, min_samples_split=5, min_samples_leaf=4))
```

0.78

Model performance with 3 Classes (> Avg, Avg, <Avg)

Decision Tree

Accuracy Score

```
# DT Model #1: Default settings
model_dt3c.append(DecisionTreeClassifier(max_depth=None, max_features=None, min_samples_split=2, min_samples_leaf=1))

# DT Model #1: Simple tree
model_dt3c.append(DecisionTreeClassifier(max_depth=4, max_features=None, min_samples_split=5, min_samples_leaf=3))

# DT Model #3: Best results after testing
model_dt3c.append(DecisionTreeClassifier(max_depth=5, max_features=None, min_samples_split=10, min_samples_leaf=5))

# DT Model #4: Best estimator from RandomizedSearchCV
model_dt3c.append(DecisionTreeClassifier(max_depth=7, max_features=None, min_samples_split=4, min_samples_leaf=8))
```

0.50

0.55

0.57

0.55

Random Forest

Accuracy Score

```
# RF Model #1
model_rf3c.append(RandomForestClassifier(n_estimators=200, random_state=78, bootstrap=True,
                                         max_depth=5, min_samples_split=2, min_samples_leaf=1))

# RF Model #2
model_rf3c.append(RandomForestClassifier(n_estimators=600, random_state=78, bootstrap=False,
                                         max_depth=10, min_samples_split=5, min_samples_leaf=3))

# RF Model #3: One of the best estimators from RandomizedSearchCV
model_rf3c.append(RandomForestClassifier(n_estimators=250, random_state=78, bootstrap=False,
                                         max_depth=10, min_samples_split=7, min_samples_leaf=2))
```

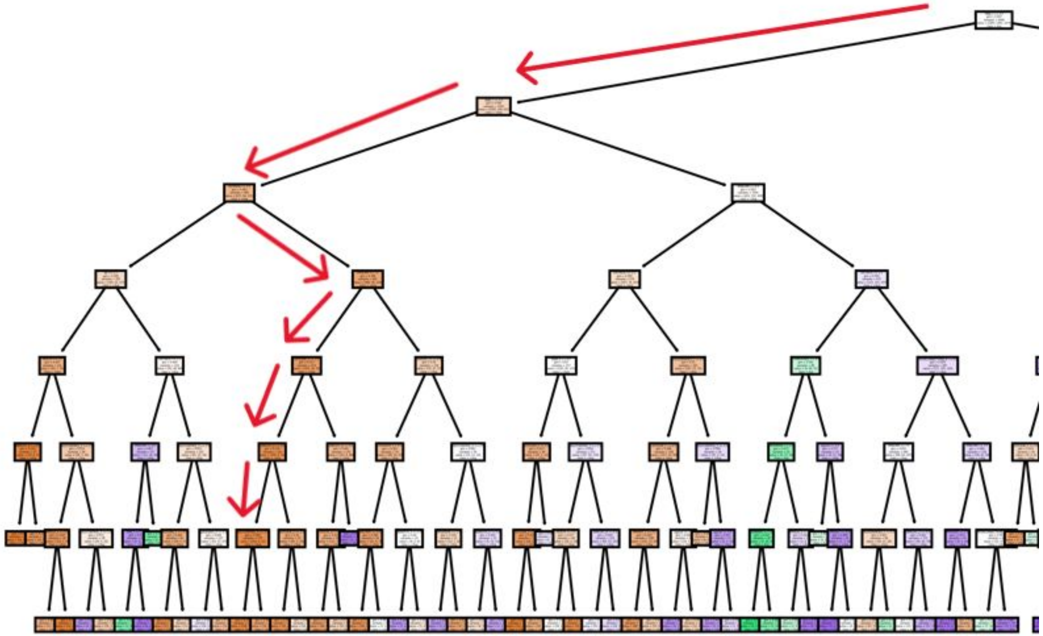
0.59

0.61

0.62

Sample Decision Tree

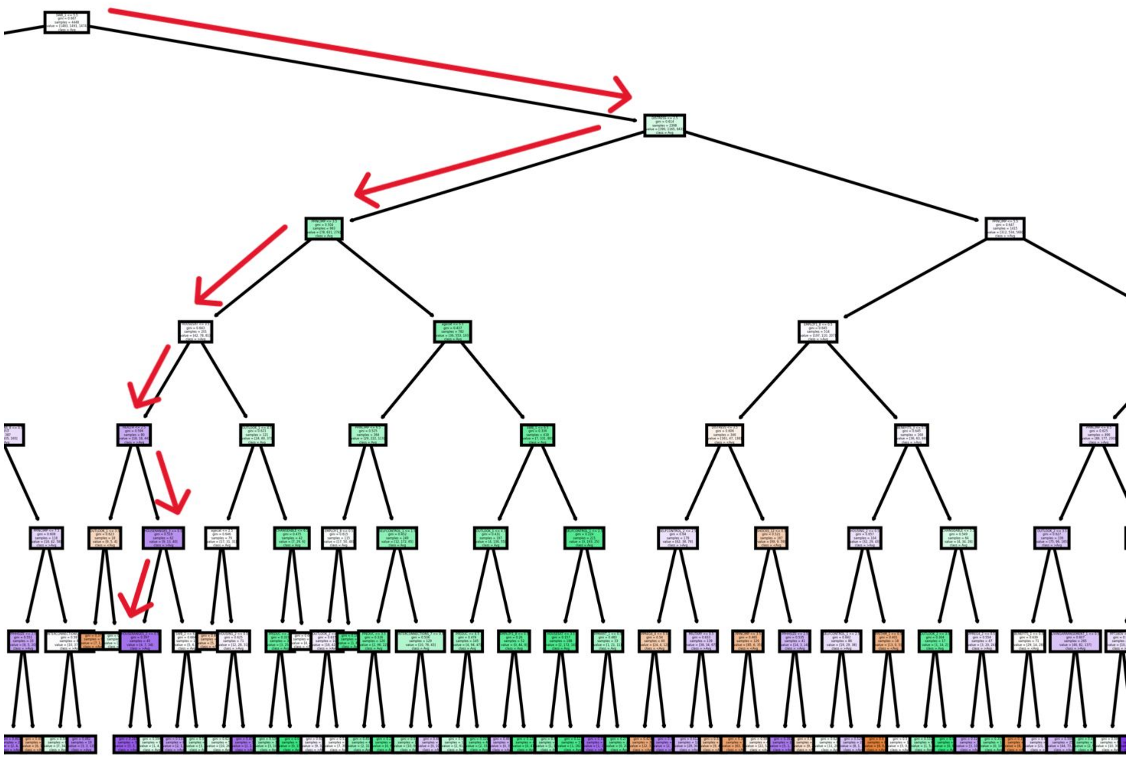
with Household Income



I am satisfied with my life.	SWB_1 <= 5.5 gini = 0.667 samples = 4448 value = [1483, 1491, 1474] class = Avg	1 Strongly Disagree → 7 Strongly Agree
I am satisfied with my life.	SWB_1 <= 4.5 gini = 0.596 samples = 2050 value = [1093, 326, 631] class = <Avg	1 Strongly Disagree → 7 Strongly Agree
I have a lot of stress in my life.	DISTRESS <= 3.5 gini = 0.483 samples = 996 value = [671, 94, 231] class = <Avg	1 strongly disagree → 5 strongly agree
Household Income	PPINCIMP <= 7.5 gini = 0.389 samples = 641 value = [486, 39, 116] class = <Avg	1 Less than $20K → 9 $150K or more
There are good work opportunities for me, if I choose to take them.	OUTLOOK_2 <= 3.5 gini = 0.317 samples = 513 value = [416, 18, 79] class = <Avg	1 Strongly Disagree → 5 Strongly Agree
Retired	EMPLOY1_8 <= 0.5 gini = 0.216 samples = 334 value = [294, 9, 31] class = <Avg	0 No 1 Yes
I am able to work diligently toward long-term goals.	SELFCONTROL_3 <= 3.5 gini = 0.165 samples = 279 value = [254, 4, 21] class = <Avg	[how it describes you] 1 Not at all → 4 Completely well

Sample Decision Tree

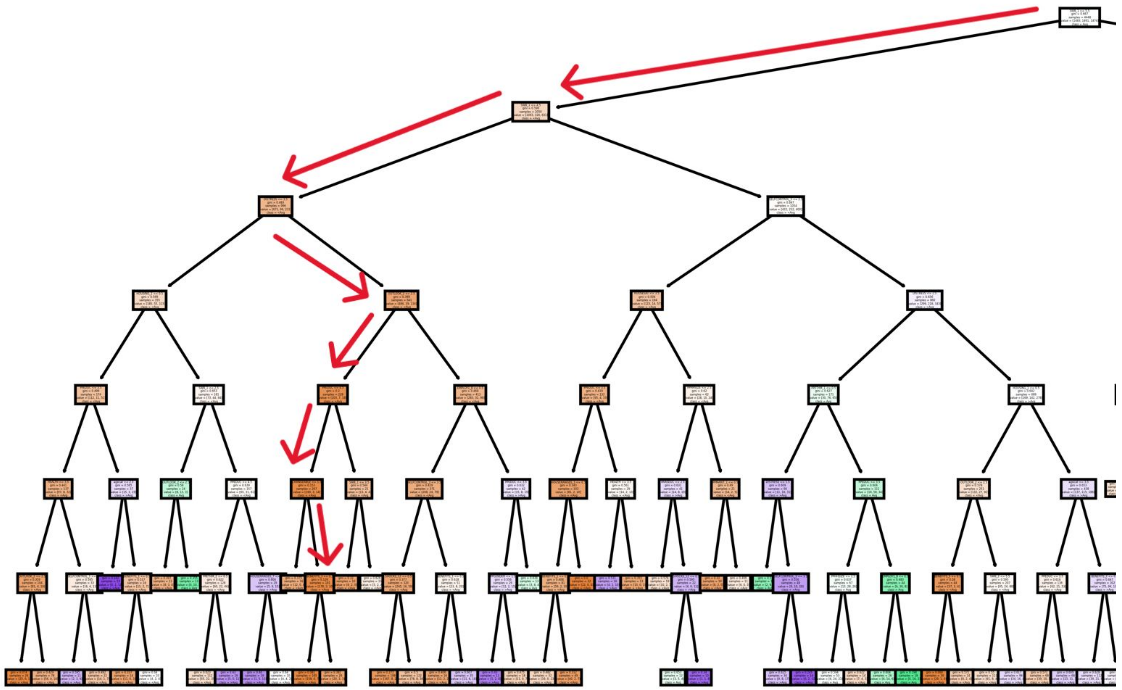
with Household Income



I am satisfied with my life.	SWB_1 <= 5.5 gini = 0.667 samples = 4448 value = [1483, 1491, 1474] class = Avg	1 Strongly Disagree → 7 Strongly Agree
I have a lot of stress in my life.	DISTRESS <= 2.5 gini = 0.614 samples = 2398 value = [390, 1165, 843] class = Avg	1 strongly disagree → 5 strongly agree
Household Income	PPINCIMP <= 3.5 gini = 0.504 samples = 983 value = [78, 631, 274] class = Avg	1 Less than $20K → 9 $150K or more
How satisfied are you with the place you live currently?	HOUSESAT <= 3.5 gini = 0.643 samples = 201 value = [42, 78, 81] class = >Avg	1 Not at all satisfied → 5 very satisfied
In general, would you say your health is . . .	HEALTH <= 2.5 gini = 0.596 samples = 80 value = [18, 18, 44] class = >Avg	1 Poor → 5 Excellent
About how much do you pay for your home each month? $500-$749 (dummy column)	HOUSERANGES_3 <= 0.5 gini = 0.519 samples = 62 value = [9, 13, 40] class = >Avg	0 No 1 Yes
About how much do you pay for your home each month? $300-$499 (dummy column)	HOUSERANGES_2 <= 0.5 gini = 0.397 samples = 45 value = [4, 7, 34] class = >Avg	0 No 1 Yes

Sample Decision Tree

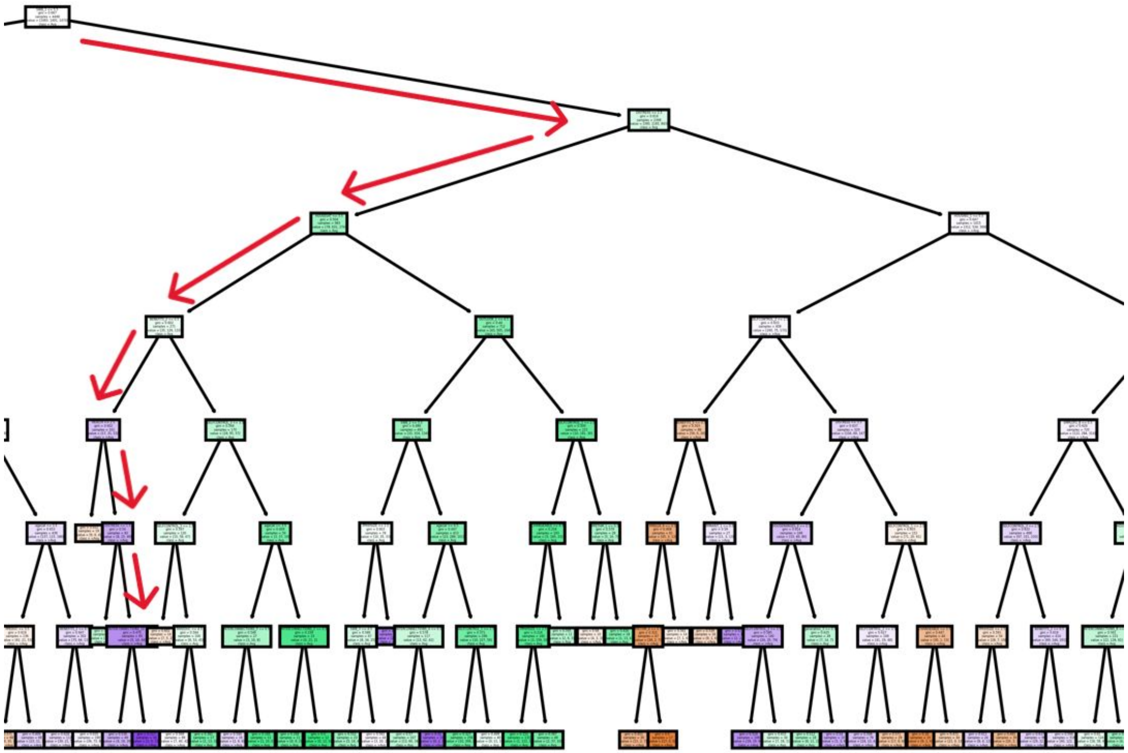
excluding Household Income



I am satisfied with my life.	SWB_1 <= 5.5 gini = 0.667 samples = 4448 value = [1483, 1491, 1474] class = Avg	1 Strongly Disagree → 7 Strongly Agree
I am satisfied with my life.	SWB_1 <= 4.5 gini = 0.596 samples = 2050 value = [1093, 326, 631] class = <Avg	1 Strongly Disagree → 7 Strongly Agree
I have a lot of stress in my life.	DISTRESS <= 3.5 gini = 0.483 samples = 996 value = [671, 94, 231] class = <Avg	1 strongly disagree → 5 strongly agree
There are good work opportunities for me, if I choose to take them.	OUTLOOK_2 <= 2.5 gini = 0.389 samples = 641 value = [486, 39, 116] class = <Avg	1 Strongly Disagree → 5 Strongly Agree
Education (Highest Degree Received)	PPEDUC <= 4.5 gini = 0.2 samples = 228 value = [203, 7, 18] class = <Avg	1 Less than high school → 5 Grad/Professional Degree
Belief that ability to manage money is NOT changeable	CHANGEABLE <= 1.5 gini = 0.153 samples = 207 value = [190, 3, 14] class = <Avg	1 Strongly disagree → 7 Strongly agree
Defined-Benefit Pension [access through current or former employer]	BENEFITS_3 <= 0.5 gini = 0.129 samples = 191 value = [178, 3, 10] class = <Avg	0 No 1 Yes

Sample Decision Tree

excluding Household Income



I am satisfied with my life.	SWB_1 <= 5.5 gini = 0.667 samples = 4448 value = [1483, 1491, 1474] class = Avg	1 Strongly Disagree → 7 Strongly Agree
I have a lot of stress in my life.	DISTRESS <= 2.5 gini = 0.614 samples = 2398 value = [390, 1165, 843] class = Avg	1 strongly disagree → 5 strongly agree
How satisfied are you with the place you live currently?	HOUSESAT <= 3.5 gini = 0.504 samples = 983 value = [78, 631, 274] class = Avg	1 Not at all satisfied → 5 very satisfied
401(k) [access through current or former employer]	BENEFITS_2 <= 0.5 gini = 0.602 samples = 271 value = [35, 126, 110] class = Avg	0 No 1 Yes
In general, would you say your health is . . .	HEALTH <= 2.5 gini = 0.602 samples = 101 value = [17, 31, 53] class = >Avg	1 Poor → 5 Excellent
I have a lot of stress in my life.	DISTRESS <= 1.5 gini = 0.54 samples = 82 value = [8, 25, 49] class = >Avg	1 strongly disagree → 5 strongly agree
Do you seek advice on matters involving money from any of the following types of people or organizations? (Did not select any)	INTERCONNECTIONS_10 <= 0.5 gini = 0.475 samples = 65 value = [5, 16, 44] class = >Avg	0 No 1 Yes

Conclusions

Conclusions

Understanding Financial Well-Being and the Scoring Tool

- Financial well-being is very complex and goes well beyond financial measures such as income, expenditures, savings, and investments.
- Individuals with a financial well-being score below 48 are experiencing a significant financial insecurity.
- Life satisfaction and stress levels appear to be more influential to financial well-being than income.
- Housing satisfaction, physical health, and life outlook, also appear to highly influence financial well-being.
- A non-financial survey could sufficiently predict financial well-being and identify individuals in need of financial education or resources?

Conclusions

Machine Learning Models

- Valuable insights can be gained in models with multiple classes even with lower accuracy.
- Decision Trees and Random Forests require larger datasets to ensure sufficient samples in deeper nodes/leaves and discover insights across multiple variables.
- Machine learning could be used by researchers to further refine and test to the financial well-being scale.

Conclusions

Original Question

If the model works, can this be used to identify individuals in need of financial education or resources?

Finding

Yes. In addition, a non-financial survey could be developed to predict financial well-being and identify individuals in need of financial education or resources.