Predicting Changes in Recovery Goal for Alcohol Use Disorder

PIPER DEAN (CARLETON COLLEGE), KELSEY MULLIGAN (WHEATON COLLEGE), JULIAN SOTO (UNIVERSITY OF PUERTO RICO, MAYAGUEZ)

FACULTY MENTOR: DR. GRANT BROWN

Outline

Introduction

- Research purpose
- Data

Modeling

Results/Summary

Future Explorations

Project Purpose

- This project uses machine learning models on data collected from survey responses of those recovering from Alcohol Use Disorder (AUD) to predict changes in participants' recovery goals and to better understand the process of recovery.

- This is important for identifying primary influencers of drinking outcomes.

- Our goal was to create a model that could make accurate predictions.

Data

2020: baseline data collected from 1498 individuals

2021: follow-up 1 included 255 individuals

2022: follow-up 2 included 235 individuals

- ❖ We utilized the baseline and "best" follow-up data (most recent follow-up available), leaving us with 276 individuals

Data

- Question 115: goals for recovery
  - Abstinent
  - Less than Abstinent
  - Controlled

- We want to predict...
  - Less Strict
  - Constant
  - Stricter

Variables

There were over 600 variables in the survey

We focused on 27, including...

- Social support
- Gender identity
- Anxiety levels
- Economic status
- Recovery capital
- Environmental stressors
- Baseline recovery goal

Models - Overview

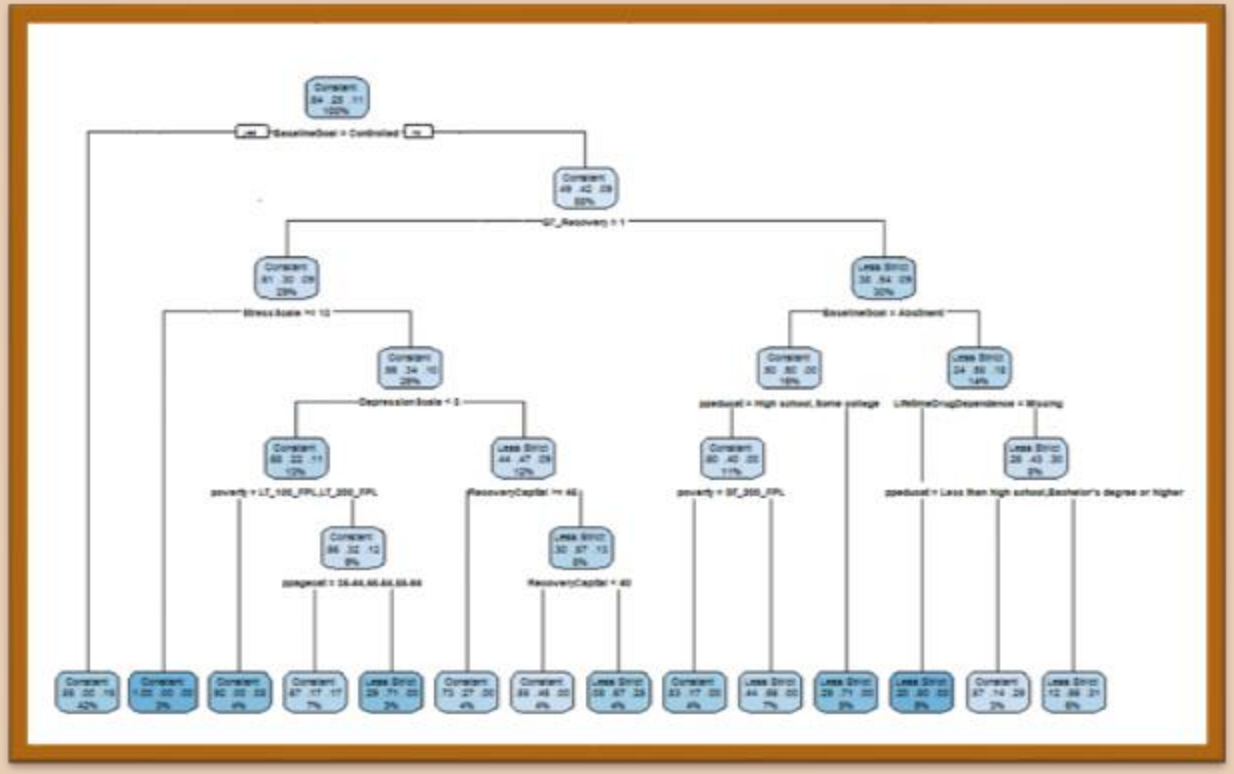
- Machine Learning Approaches
  - Single Decision tree
  - Random forest
  - Neural Network
- Traditional Approach
  - Bayesian multinomial regression

Validation Method

- Packages: caret
- Cross Validation: resampling method utilized to train and test a model on different iterations
- Confusion Matrix: separates our predictions and actual values into two dimensions

Prediction	Reference: Constant	Less Strict	Stricter
Constant	128	43	22
Less Strict	46	20	9
Stricter	3	5	0

NIR = 64%
Accuracy = 53%



Single Decision Tree

- Package: rpart
- Contains nodes and branches
- Each split classifies our data based on the inputs
- Process is repeated until final leaf node, which contain our predictions

Random Forest

- Packages: ranger, RandomForest
- Utilizes many single decision trees and chooses the output given by most trees
- Examined the variable importance plots
- Model is less predictive than the no information rate

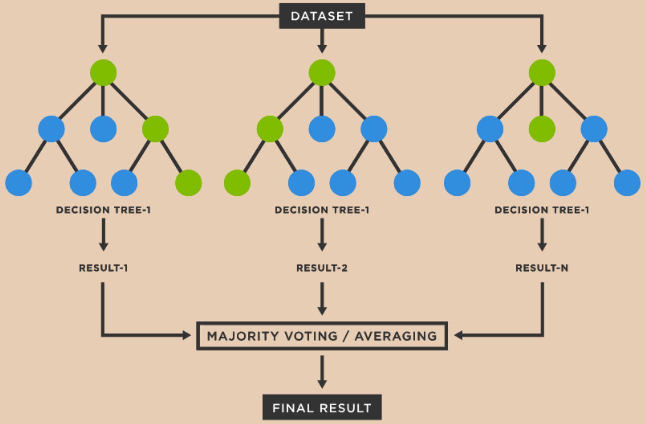
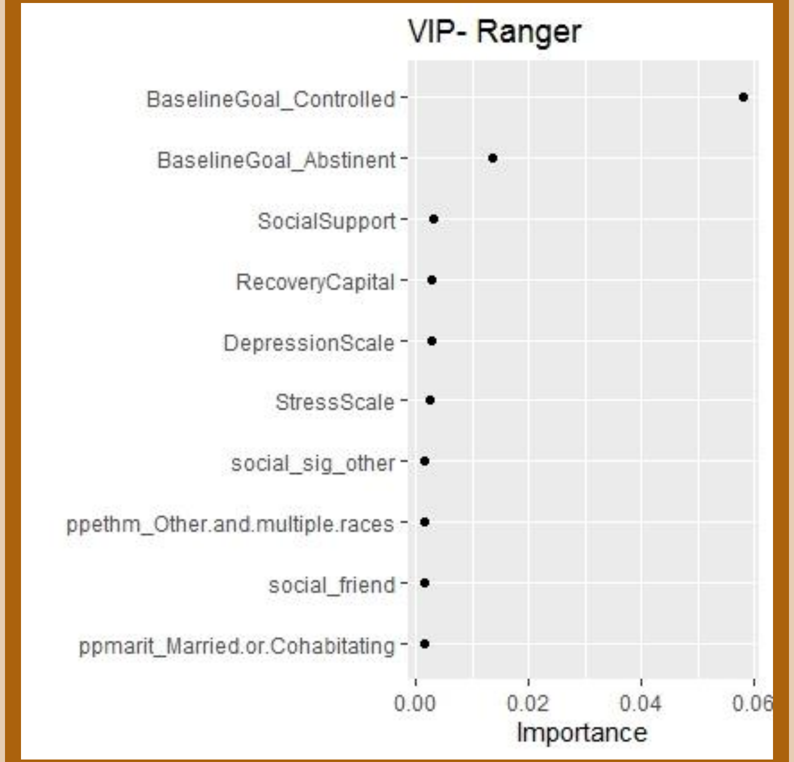
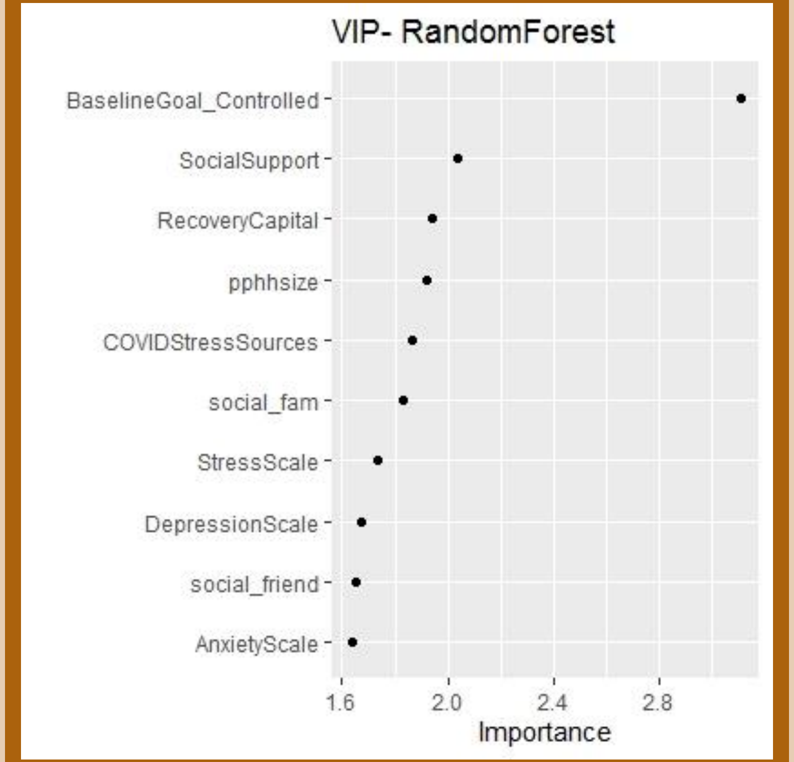
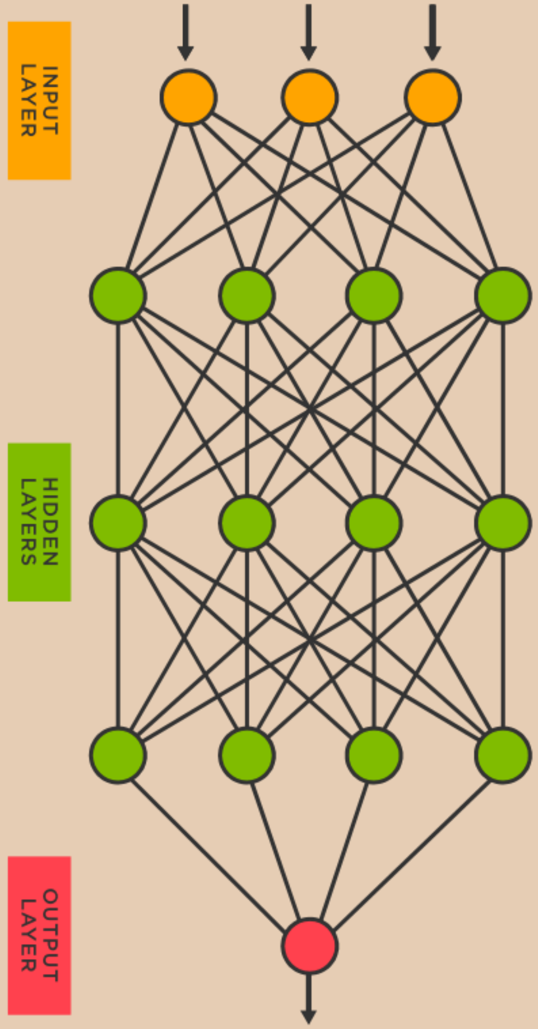


Image from: https://www.tibco.com/reference-center/what-is-a-random-forest





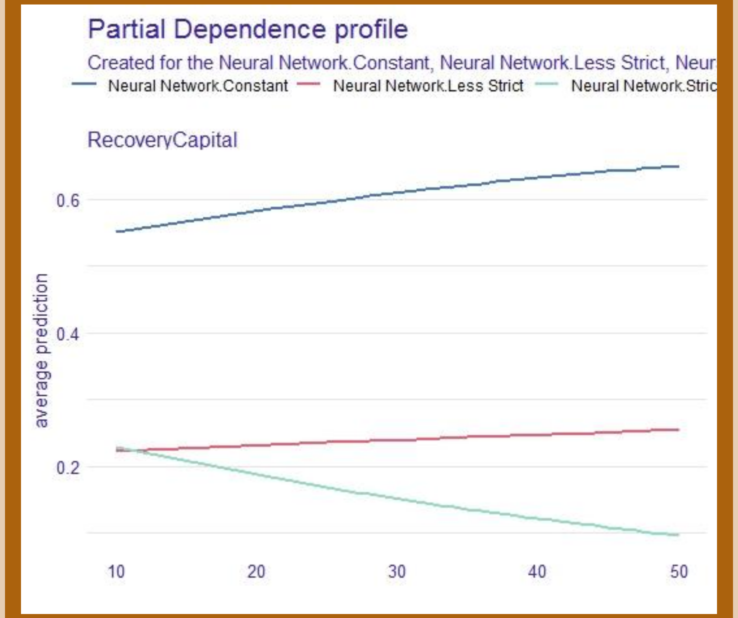
Variable Importance Plots

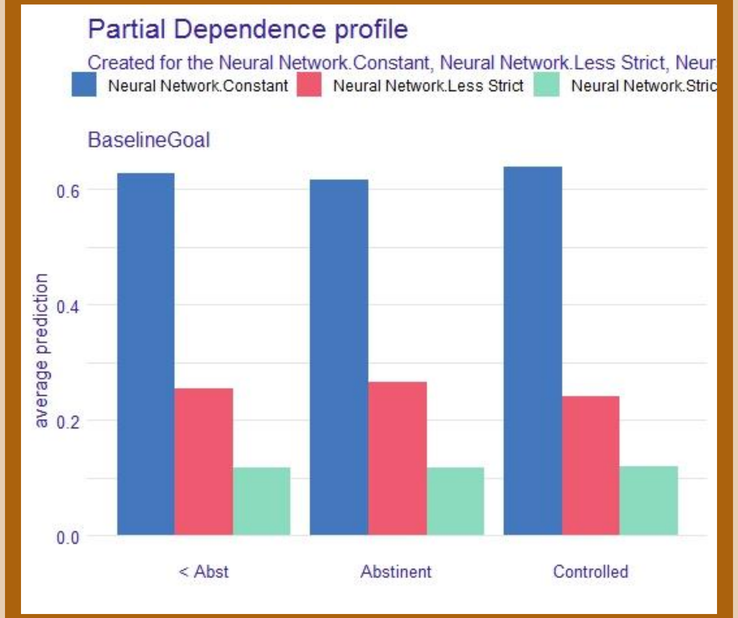


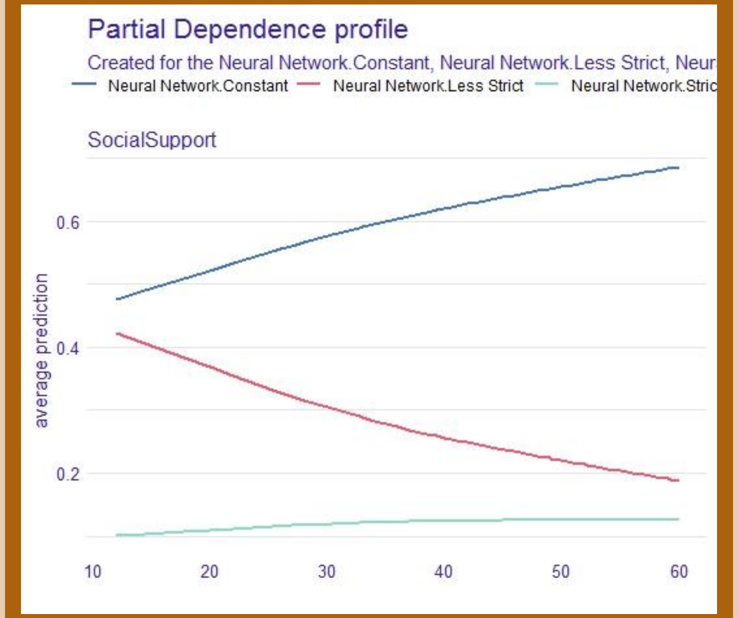
Neural Network

- Packages: Brulee
- Fit a multilayer perceptron network
- Weight assigned to each input in order to make predictions
- Model is an ineffective predictor
- Examined the partial dependence of variables in this model

Image from: https://www.tibco.com/reference-center/what-is-a-neural-network







Partial Dependence Plots – Neural Network

Results (Machine Learning)

- Our ML models lacked predictive power
- Models could not accurately determine changes in recovery goals
- Illuminated important variables
- Turn to Bayesian approach

Bayesian Multinomial Regression

$$\ln \frac{\Pr(Y_i = 1)}{\Pr(Y_i = K)} = \boldsymbol{\beta}_1 \cdot \mathbf{X}_i$$

$$\ln \frac{\Pr(Y_i = 2)}{\Pr(Y_i = K)} = \boldsymbol{\beta}_2 \cdot \mathbf{X}_i$$

$$\cdots\cdots$$

$$\ln \frac{\Pr(Y_i = K-1)}{\Pr(Y_i = K)} = \boldsymbol{\beta}_{K-1} \cdot \mathbf{X}_i$$

- Used a multinomial logistic regression model in Stan
- Predicted outcomes as odd ratios compared to the baseline of constant recovery group

Results (Traditional Approach)

Variable	Odds Ratio	Probability
Recovery Definition	1.524	0.976
Drug Use	7.077	0.992
Poverty	5.893	0.981
Education Status- HS Degree	20.325	0.980
Minor Children	0.005	1.000
Relapse	0.002	1.000

- Identified key associations:
  - Positive associations between less strict and constant: **recovery definition**
  - Positive associations between stricter and constant: **drug use, poverty, education status**
  - Negative associations between less strict and constant: **baseline goal: controlled, minor children**
  - Negative associations between stricter and constant: **baseline goal: abstinent, previous relapse**

Summary

Machine Learning models: ineffective predictors of change in recovery status

Traditional model: insight into the association between certain variables and the outcome of interest

Future Explorations

- Testing with other models to look for a stronger predictive power
- Focusing on a different set of variables and their influence
- Predicting the occurrence of relapse
  - Prioritizes changes in one direction of the strictness scale

Acknowledgements

- National Heart Lung and Blood Institute (NHLBI), grant # HL161716-01
- Iowa Biostatistics Department
- Grant Brown

References

Gilbert, P. A., Pro, G., Zemore, S. E., Mulia, N., & Brown, G. (2019). Gender differences in use of alcohol treatment services and reasons for Nonuse in a national sample. *Alcoholism: Clinical and Experimental Research*, *43*(4), 722–731. https://doi.ord/10.1111/acer.13965

Kuhn et al., (2020). Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. https://www.tidymodels.org

Fit neural networks — brulee_mlp. (n.d.). Brulee.tidymodels.org. Retrieved July 19, 2023, from https://brulee.tidymodels.org/reference/brulee_mlp.html

Molnar, C. (n.d.). 5.1 Partial Dependence Plot (PDP) | Interpretable Machine Learning. In *christophm.github.io*. https://christophm.github.io/interpretable-ml-book/pdp.html

Questions?