

Mental Health in Tech

Urbane Technologien

Ilia Ozhmegov, 07.01.2021

Content

1. Why Mental Health in Tech dataset
2. Business Value
3. Dataset Introduction
4. Data Issues
5. Customer Questions
6. Imputation
7. Feature Selection
8. Modeling



Why Mental Health in Tech dataset

1. Many my friends and acquaintances in this area started suffer from different mental issues in the recently passed 2020 year from anxiety to depression.
2. According to OSMI data, 51% of tech professionals have been diagnosed with a mental health condition. By comparison, 19.1% of U.S. adults experience mental illness, according to the National Alliance on Mental Illness.
3. Mental Health affects your performance at work whereas your performance can affect BV the company. As a result it can make the future of the company less bright.

Stockwell, the AI-vending machine startup formerly known as Bodega, shutting down July 1

Ingrid Lunden @ingridlunden 9:27 AM GMT+2 • June 16, 2020

Psyomics secures £1.5M funding for Censeo mental health platform

EIS fund manager Parkwalk led the funding round alongside existing investors Jonathon Milner, Cambridge Enterprise and Martlet.

By Sophie Porter | October 27, 2020 | 02:51 am

SHARE Share 35 [f](#) [in](#) [t](#)

 Latest

Your Sneak Preview of the Future

DE NL Login Subscribe

START-UPS

EIT InnoEnergy: 'Start-ups that go bankrupt due to corona amounts to capital destruction'

EIT InnoEnergy aims to drag start-ups that it invests in as well as it can through the corona crisis. Read how they go about that in this series.

24 October 2020

LUCETTE MASCINI



LATEST

EIT InnoEnergy: 'Start-ups that go bankrupt due to corona amounts to capital destruction'

The c
desig

[Menu](#)

[Search](#)

'Say
coron

A hy
ships
Antw

Rese
the e

NEW

Your

ScaleFactor Raised \$100 Million In A Year Then Blamed Covid-19 For Its Demise. Employees Say It Had Much Bigger Problems.

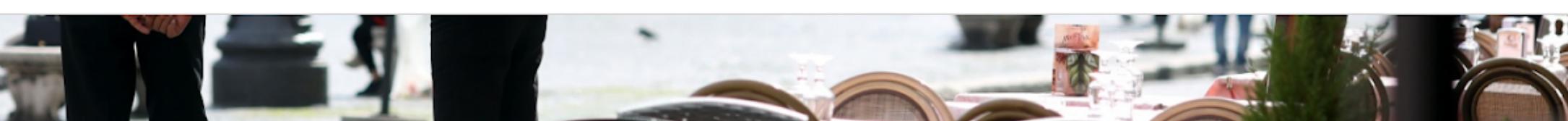


Bloomberg

Sign In

Subscribe

What you need to know: [Tracking the U.S. Small Business Impact From Coronavirus](#)



Photographer: Alessia Pierdomenico/Bloomberg

Economics

Half of Europe's Smaller Businesses Risk Bankruptcy Within Year

By Lucy Meakin

22 октября 2020 г., 05:01 GMT+2

Dataset Introduction

- Data Source: OSMI Mental Health in Tech Survey (2014)
- 1259 obs. of 27 variables
- Features:
 - 25 categorical
 - 2 numerical

```
## 'data.frame': 1259 obs. of 
##   $ Timestamp : F
##   $ Age      : n
##   $ Gender   : F
##   $ Country  : F
##   $ state    : F
##   $ self_employed : F
##   $ family_history : F
##   $ treatment : F
##   $ work_interfere : F
##   $ no_employees : F
##   $ remote_work : F
##   $ tech_company : F
##   $ benefits  : F
##   $ care_options : F
##   $ wellness_program : F
##   $ seek_help  : F
##   $ anonymity  : F
##   $ leave     : F
##   $ mental_health_consequence: F
##   $ phys_health_consequence : F
##   $ coworkers  : F
##   $ supervisor : F
##   $ mental_health_interview : F
##   $ phys_health_interview  : F
##   $ mental_vs_physical   : F
##   $ obs_consequence : F
##   $ comments   : F
```

Dataset Introduction

```
## 'data.frame': 1259 obs. of 27 variables:  
## $ Timestamp : Factor w/ 1246 levels "2014-08-27 11:29:31",...: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Age : num 37 44 32 31 31 33 35 39 42 23 ...  
## $ Gender : Factor w/ 49 levels "A little about you",...: 16 24 30 30 30 30 16 24 16 30 ...  
## $ Country : Factor w/ 48 levels "Australia","Austria",...: 46 46 8 45 46 46 46 46 8 46 8 ...  
## $ state : Factor w/ 45 levels "AL","AZ","CA",...: 11 12 NA NA 38 37 19 NA 11 NA ...  
## $ self_employed : Factor w/ 2 levels "No","Yes": NA NA NA NA NA NA NA NA NA ...  
## $ family_history : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 2 2 1 2 1 ...  
## $ treatment : Factor w/ 2 levels "No","Yes": 2 1 1 2 1 1 2 1 2 1 ...  
## $ work_interfere : Factor w/ 4 levels "Never","Often",...: 2 3 3 2 1 4 4 1 4 1 ...  
## $ no_employees : Factor w/ 6 levels "1-5","100-500",...: 5 6 5 3 2 5 1 1 2 3 ...  
## $ remote_work : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 2 2 1 1 ...  
## $ tech_company : Factor w/ 2 levels "No","Yes": 2 1 2 2 2 2 2 2 2 2 ...  
## $ benefits : Factor w/ 3 levels "Don't know","No",...: 3 1 2 2 3 3 2 2 3 1 ...  
## $ care_options : Factor w/ 3 levels "No","Not sure",...: 2 1 1 3 1 2 1 3 3 1 ...  
## $ wellness_program : Factor w/ 3 levels "Don't know","No",...: 2 1 2 2 1 2 2 2 2 1 ...  
## $ seek_help : Factor w/ 3 levels "Don't know","No",...: 3 1 2 2 1 1 2 2 2 1 ...  
## $ anonymity : Factor w/ 3 levels "Don't know","No",...: 3 1 1 2 1 1 2 3 2 1 ...  
## $ leave : Factor w/ 5 levels "Don't know","Somewhat difficult",...: 3 1 2 2 1 1 2 1 4 1 ...  
## $ mental_health_consequence: Factor w/ 3 levels "Maybe","No","Yes": 2 1 2 3 2 2 1 2 1 2 ...  
## $ phys_health_consequence : Factor w/ 3 levels "Maybe","No","Yes": 2 2 2 3 2 2 1 2 2 2 ...  
## $ coworkers : Factor w/ 3 levels "No","Some of them",...: 2 1 3 2 2 3 2 1 3 3 ...  
## $ supervisor : Factor w/ 3 levels "No","Some of them",...: 3 1 3 1 3 3 1 1 3 3 ...  
## $ mental_health_interview : Factor w/ 3 levels "Maybe","No","Yes": 2 2 3 1 3 2 2 2 2 1 ...  
## $ phys_health_interview : Factor w/ 3 levels "Maybe","No","Yes": 1 2 3 1 3 1 2 2 1 1 ...  
## $ mental_vs_physical : Factor w/ 3 levels "Don't know","No",...: 3 1 2 2 1 1 1 2 2 3 ...  
## $ obs_consequence : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 1 1 1 1 1 ...  
## $ comments : Factor w/ 160 levels " ","-","(yes but the situation was unusual and involved a
```

Data Issues

- Almost all features is categorical
- Gender feature was filled badly
- Age feature was filled badly

gender	n
<fct>	<int>
1 "Male"	615
2 "male"	206
3 "Female"	121
4 "M"	116
5 "female"	62
6 "F"	38
7 "m"	34
8 "f"	15
9 "Make"	4
10 "Male "	3
11 "Woman"	3
12 "Cis Male"	2
13 "Female "	2
14 "Female (trans)"	2
15 "Man"	2
16 "A little about you"	1
17 "Agender"	1
18 "All"	1
19 "Androgynous"	1
20 "Cis Female"	1

Customer Questions

- What is the most important factors to define the necessity of mental treatment? (it would be great to ask only relevant questions)
- Is it possible to build any ML model based on those feature to evaluate the risk of mental health problems? (all features categorical)
- What accuracy would be of a model if it is possible? (no signs that it is feasible to build a model with a decent accuracy)

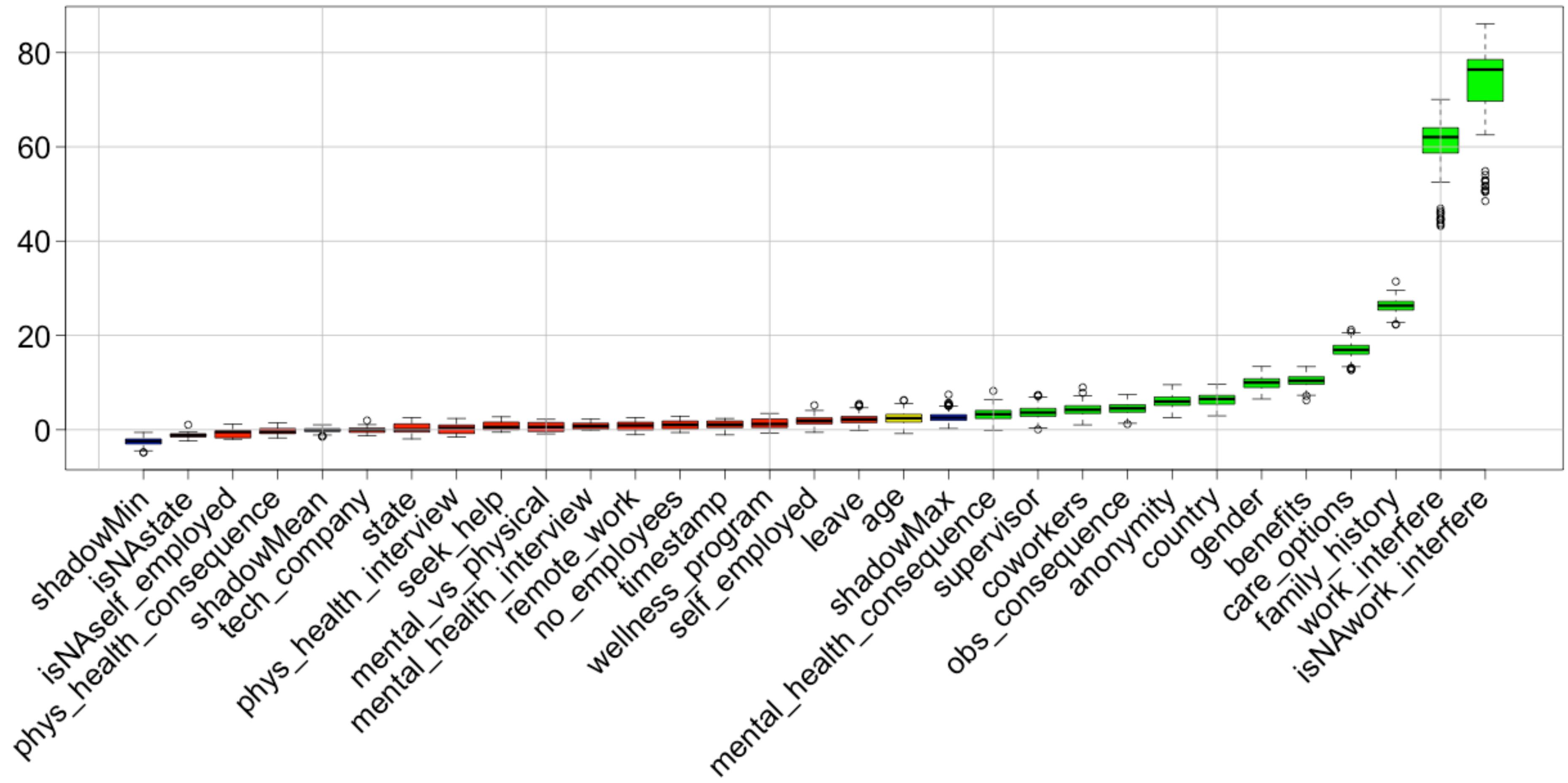
Imputation

- There were only 4 variable that had missing values
 - comments - 1090 (removed)
 - state - 513 (for not USA countries were labelled as “None”)
 - work_interfere - 262
 - self_employed - 18
- Before imputing were created a dummy variables to mark that value was imputed
- Used K-nn approach to impute ($k = 3$)

Feature Selection

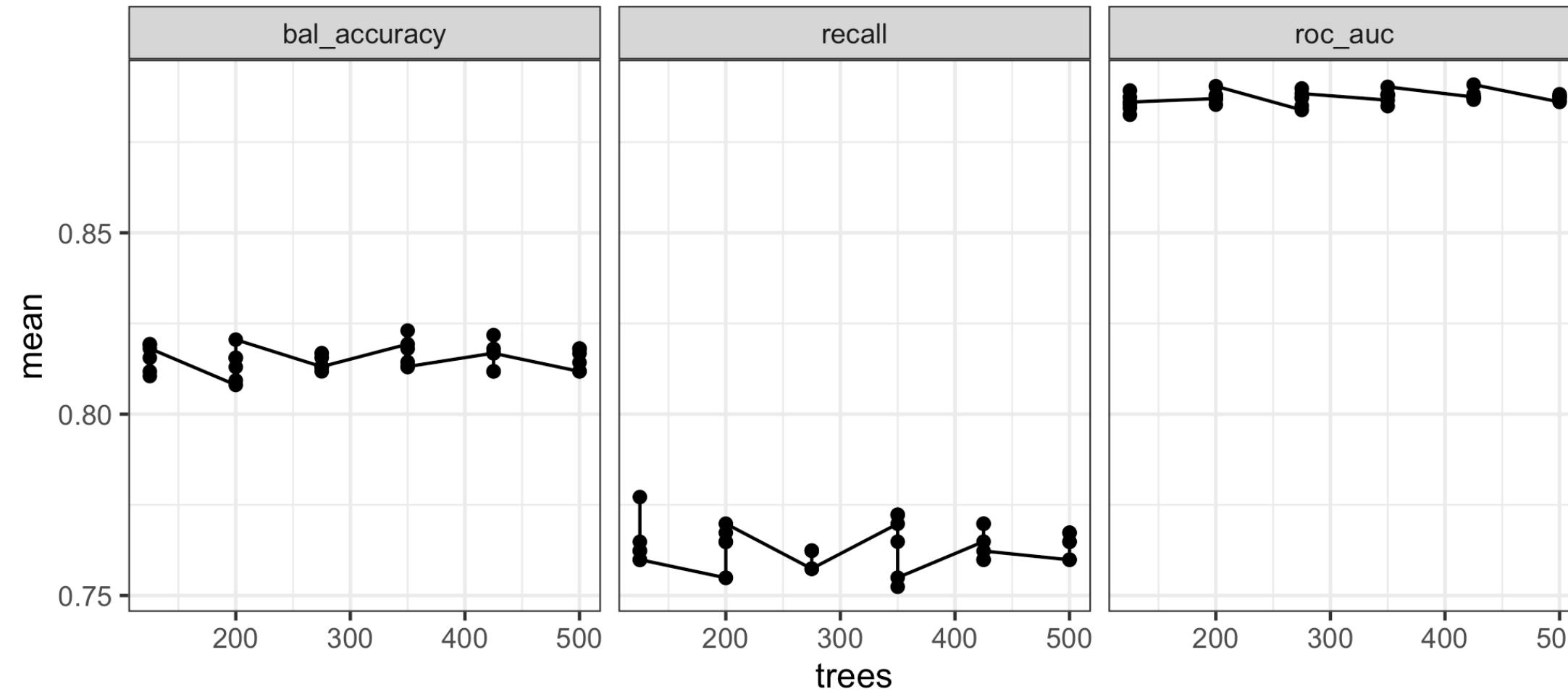
- Boruta
- Correlation matrix

Variable Importance

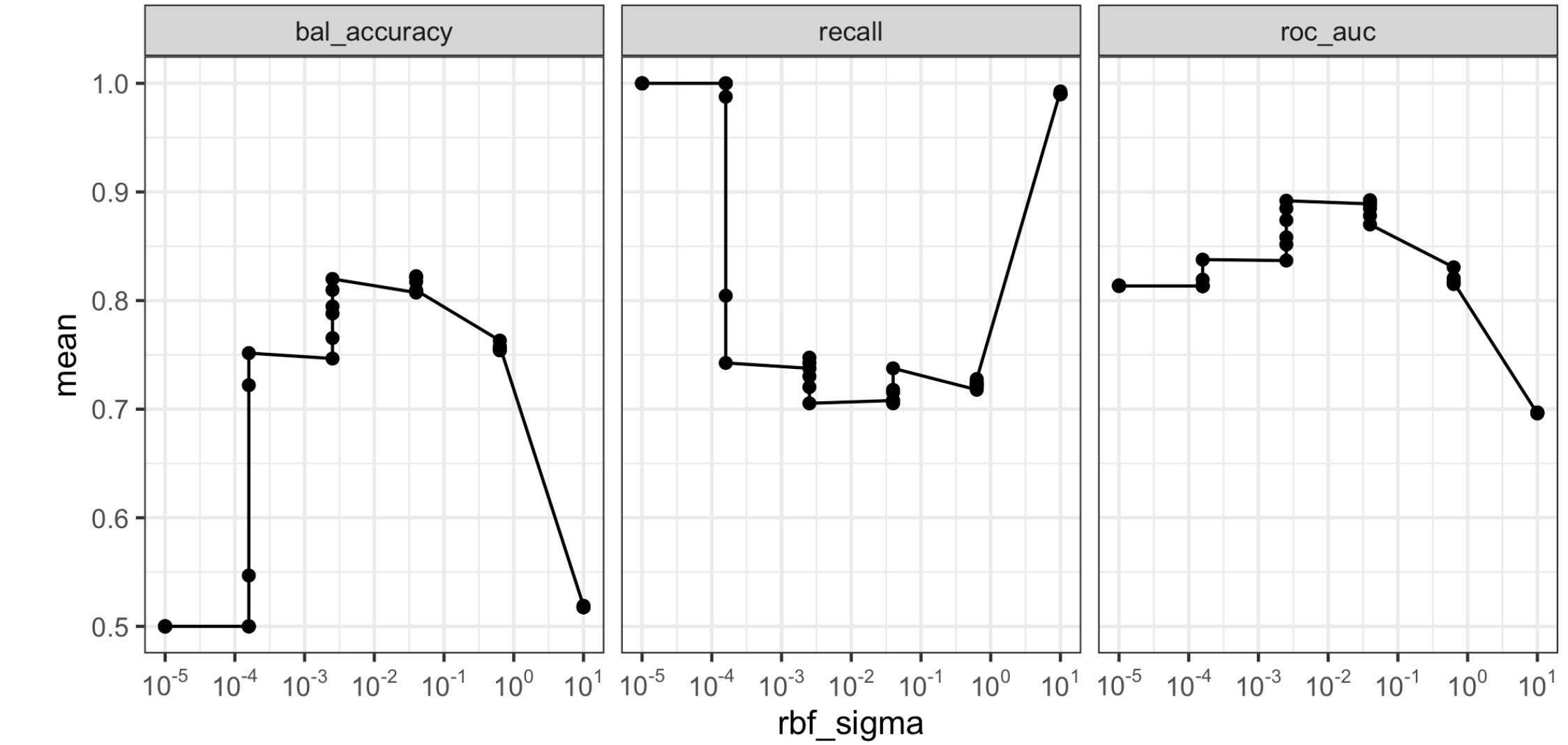
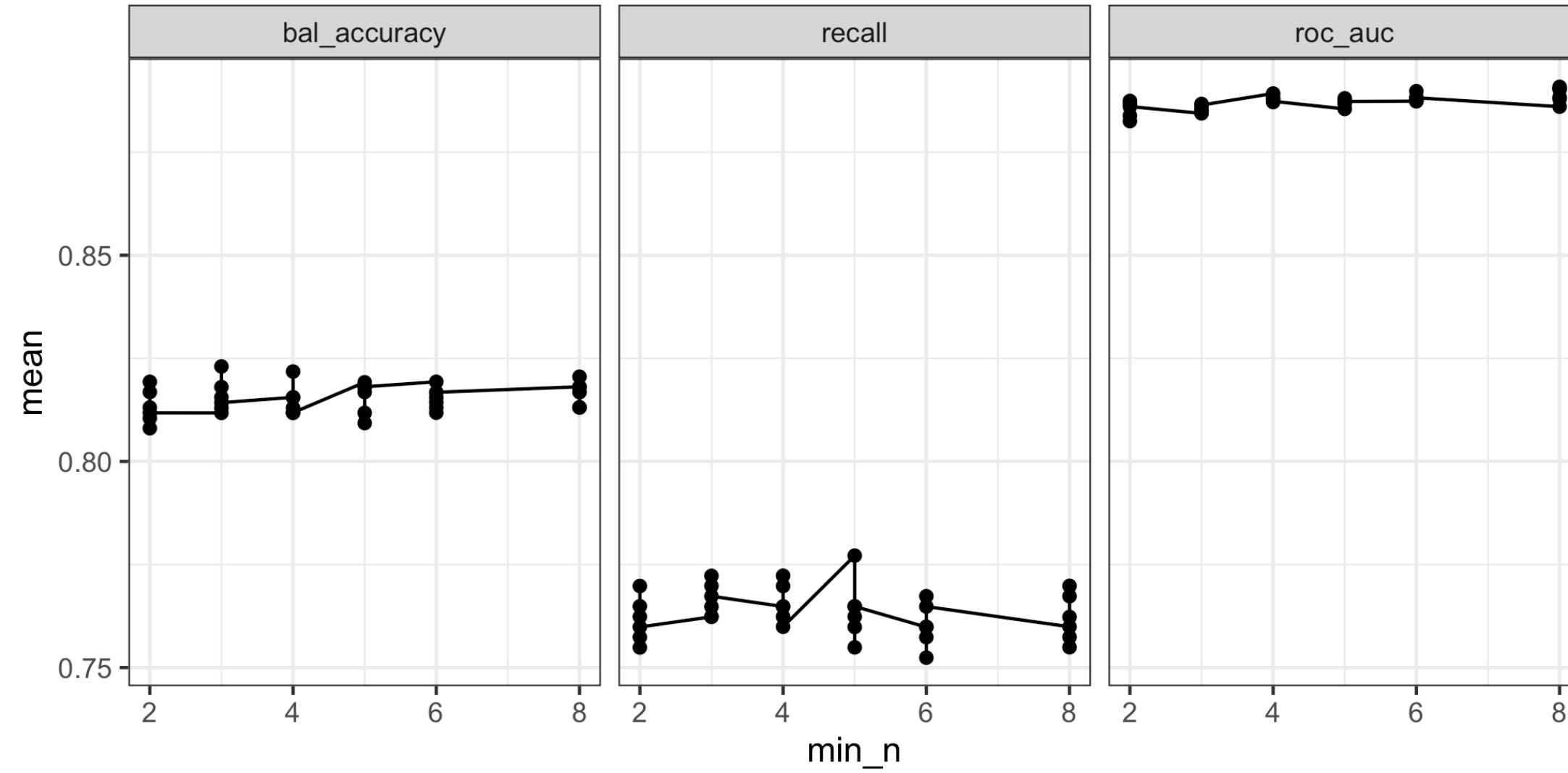
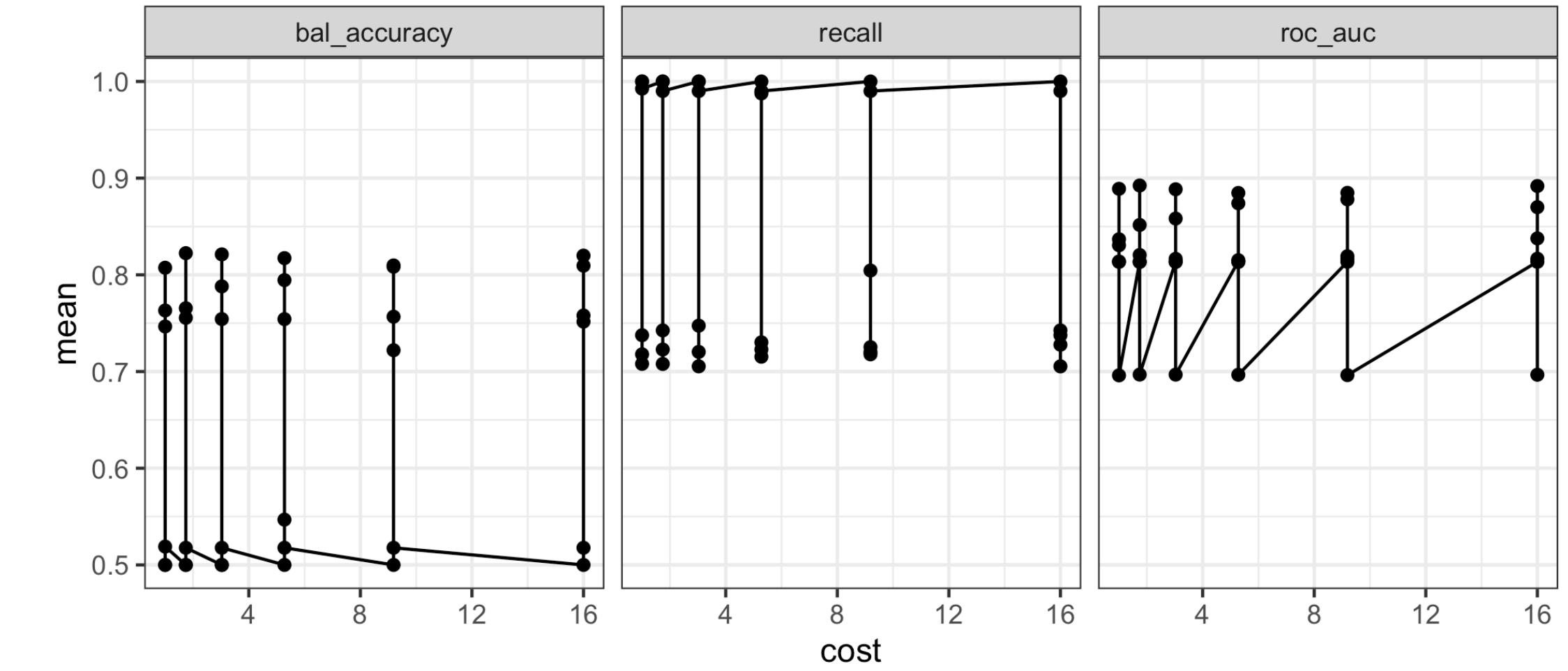


Modeling (HPO)

Random Forest penalty plot

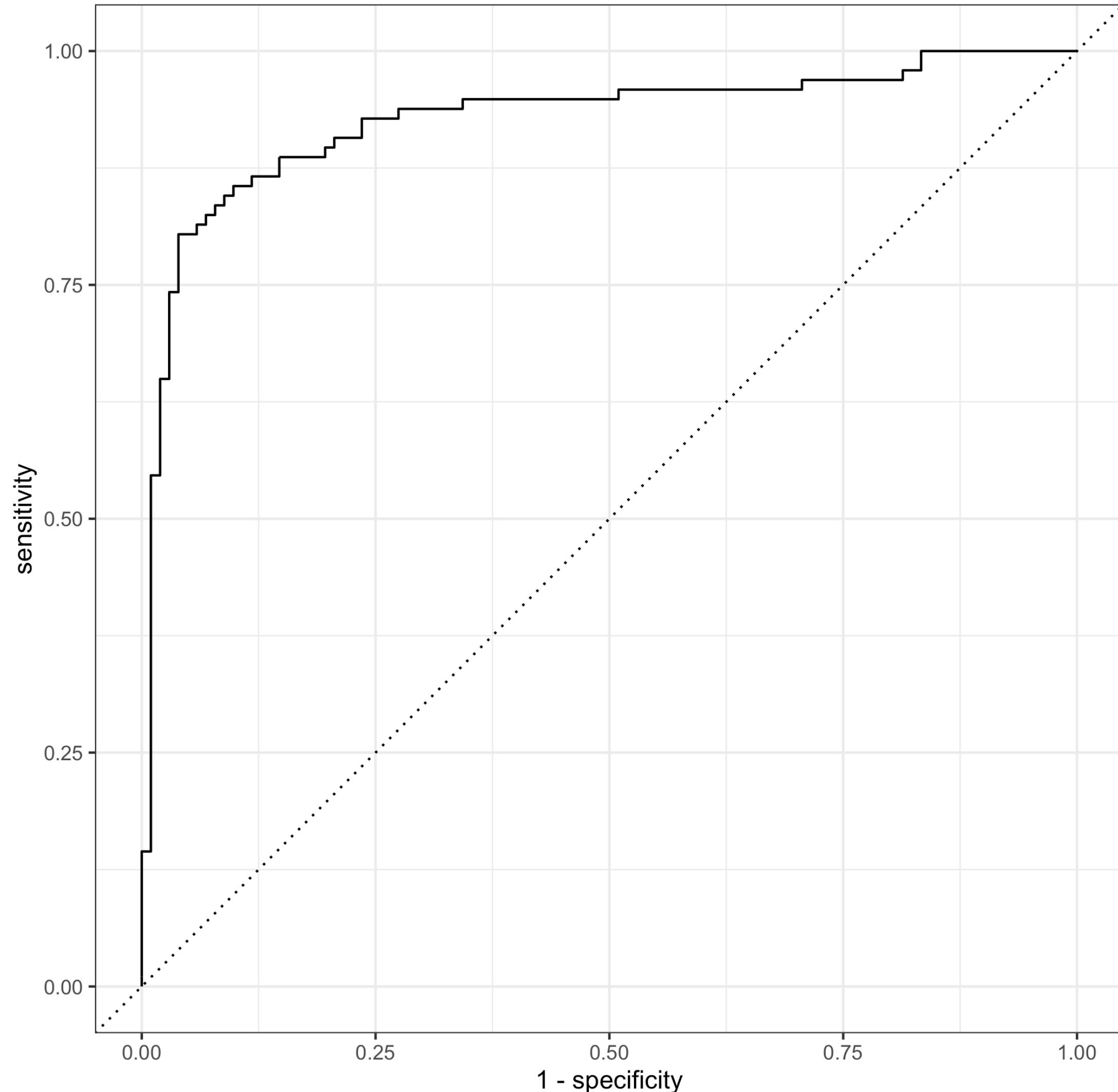


SVM penalty plot

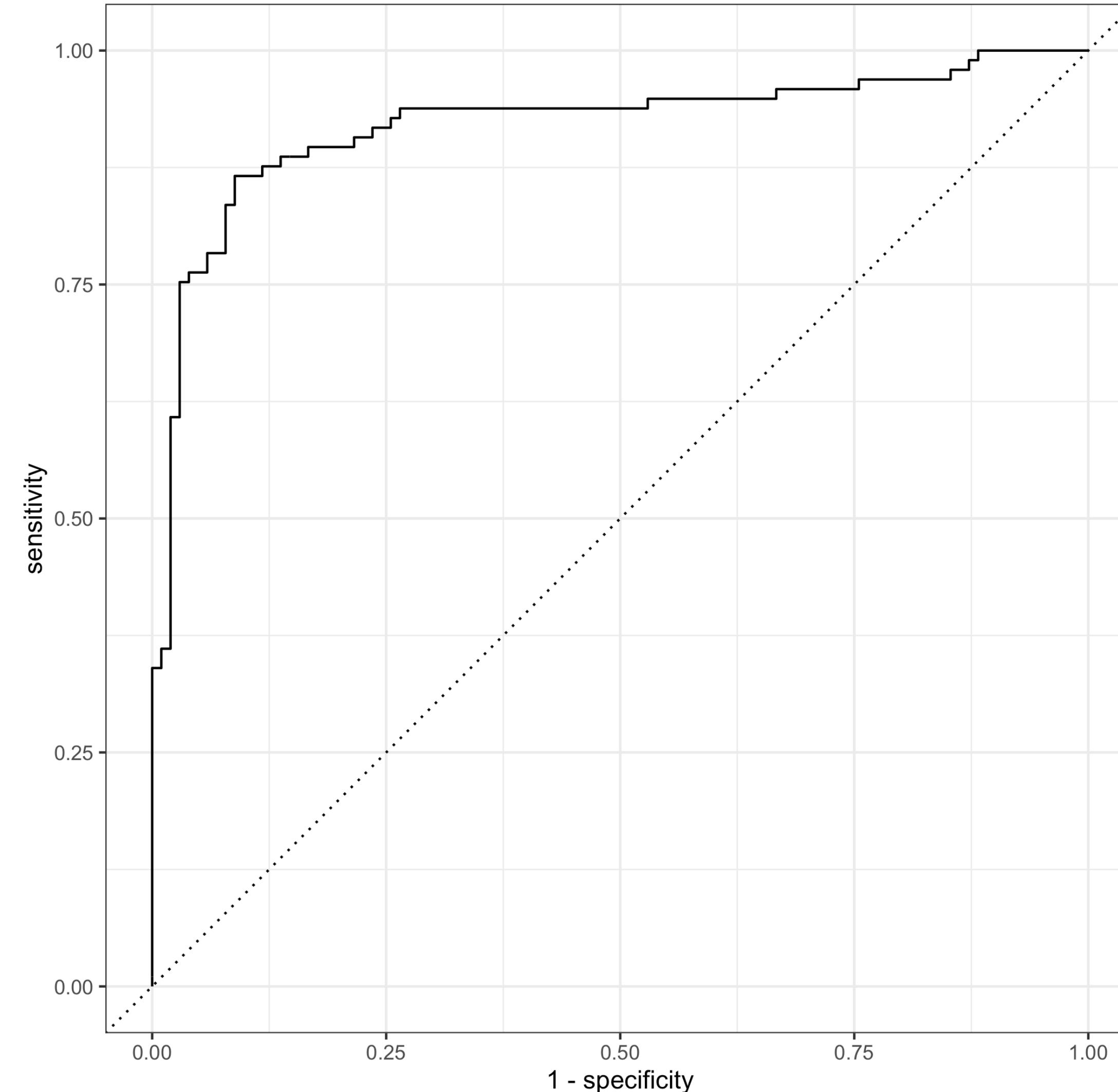


Modeling (ROC)

Random Forest: AUC 0.927

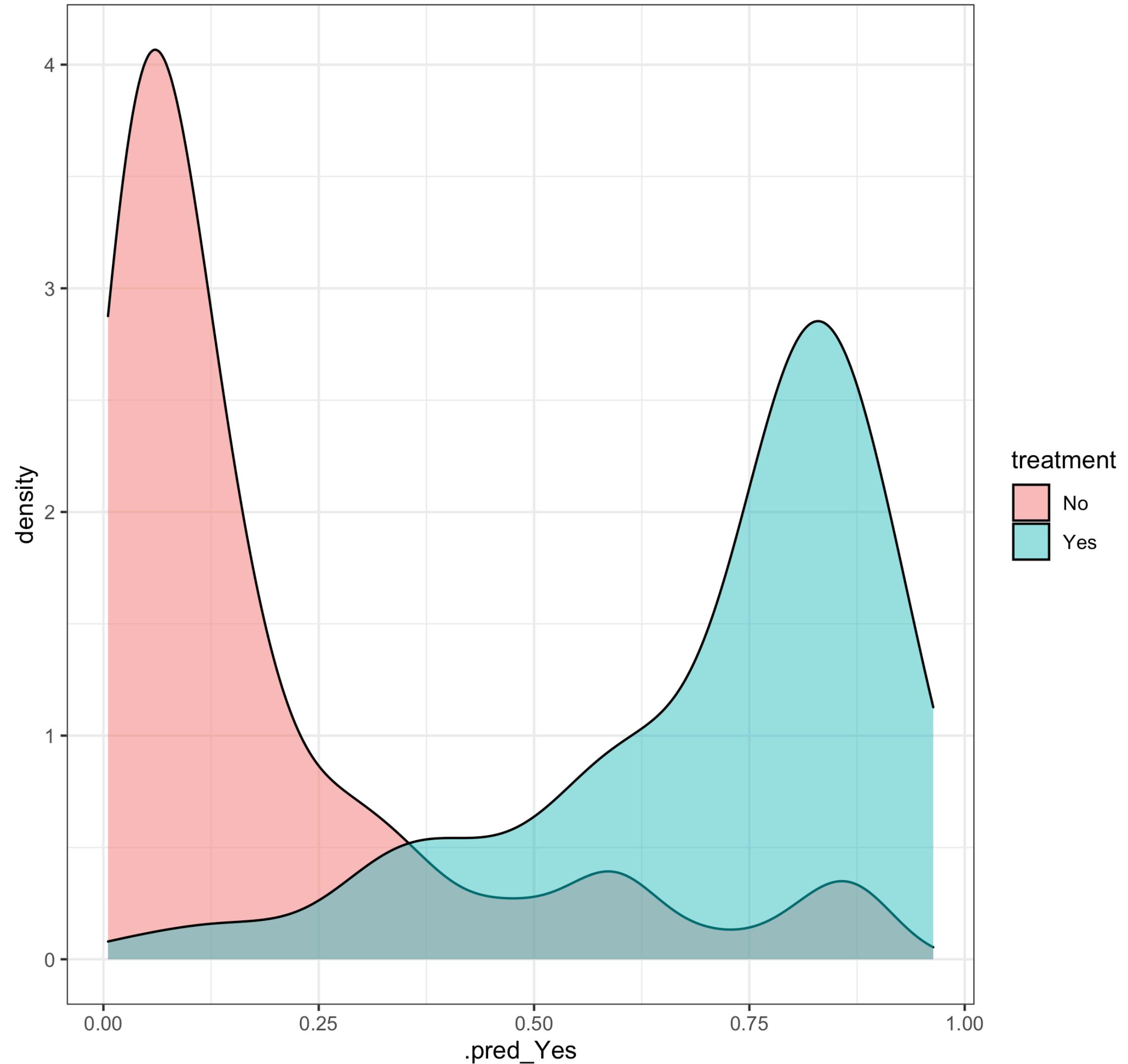


SVM: AUC 0.921

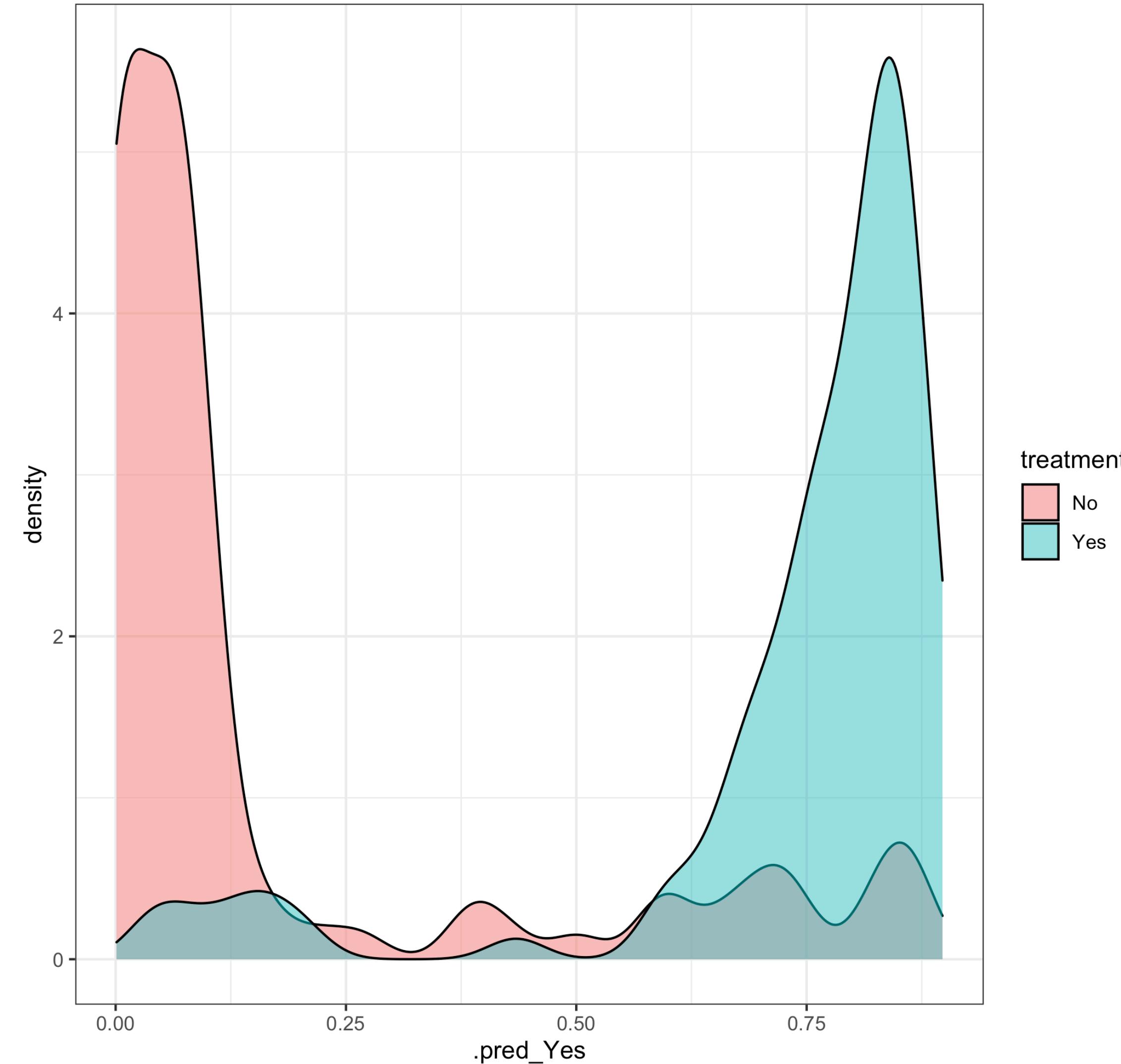


Modeling (Distribution)

Random Forest, Validation Distribution



SVM, Validation Distribution



Modeling: Validation

	AUC ROC	Recall	Accuracy	Balanced Accuracy	Kappa
Random Forest	0.927	0.853	0.869	0.870	0.739
SVM	0.921	0.922	0.864	0.863	0.728

Modeling: Testing

	Recall	Accuracy	Balanced Accuracy	Kappa
Random Forest	0.880	0.809	0.804	0.613