Recursive feature selection

Outer resampling method: Cross-Validated (10 fold, repeated 5 times)

Resampling performance over subset size:

Variables RMSE Rsquared RMSESD RsquaredSD Selected

1 0.4591 0.1743 0.009807 0.02654

5 0.3929 0.3848 0.009898 0.02921

10 0.3930 0.3845 0.010476 0.03087

15 0.3915 0.3886 0.010953 0.03241

20 0.3898 0.3935 0.011173 0.03338

22 0.3897 0.3938 0.011090 0.03315 \*

The top 5 variables (out of 22):

A\_follower\_count, B\_listed\_count, A\_listed\_count, B\_network\_feature\_1, B\_follower\_count

**TOP Predictors**

[1] "A\_follower\_count" "B\_listed\_count" "A\_listed\_count"

[4] "B\_network\_feature\_1" "B\_follower\_count" "A\_network\_feature\_1"

[7] "B\_mentions\_received" "A\_mentions\_received" "A\_retweets\_received"

[10] "A\_network\_feature\_3" "B\_retweets\_received" "B\_network\_feature\_3"

[13] "A\_network\_feature\_2" "B\_mentions\_sent" "A\_mentions\_sent"

[16] "B\_network\_feature\_2" "A\_posts" "A\_following\_count"

[19] "B\_posts" "B\_following\_count" "B\_retweets\_sent"

[22] "A\_retweets\_sent"

**Multi-Layer Perceptron**

4401 samples

16 predictor

2 classes: '0', '1'

Pre-processing: scaled (16), centered (16)

Resampling: Cross-Validated (5 fold, repeated 5 times)

Summary of sample sizes: 3521, 3521, 3521, 3521, 3520, 3521, ...

Resampling results across tuning parameters:

size Accuracy Kappa

16 0.7229308 0.4455287

24 0.7172015 0.4342571

32 0.7155688 0.4304072

Accuracy was used to select the optimal model using the largest value.

The final value used for the model was size = 16.

Accuracy = 0.47

Multi-Layer Perceptron

4401 samples

22 predictor

2 classes: '0', '1'

Pre-processing: scaled (22), centered (22)

Resampling: Cross-Validated (5 fold, repeated 5 times)

Summary of sample sizes: 3521, 3521, 3521, 3521, 3520, 3521, ...

Resampling results across tuning parameters:

size Accuracy Kappa

22 0.7311071 0.4619286

33 0.7296050 0.4589879

44 0.7308785 0.4614511

Accuracy was used to select the optimal model using the largest value.

The final value used for the model was size = 22.

Accuracy = 0.53

GBM

Tuning parameter 'shrinkage' was held constant at a value of 0.1

Tuning parameter 'n.minobsinnode' was held constant at a value of 20

Accuracy was used to select the optimal model using the largest value.

The final values used for the model were n.trees = 100, interaction.depth =

5, shrinkage = 0.1 and n.minobsinnode = 20.

**Stochastic Gradient Boosting**

4401 samples

22 predictor

2 classes: '0', '1'

Pre-processing: scaled (22), centered (22)

Resampling: Cross-Validated (5 fold, repeated 5 times)

Summary of sample sizes: 3520, 3520, 3521, 3521, 3522, 3521, ...

Resampling results across tuning parameters:

interaction.depth n.trees Accuracy Kappa

1 1 0.6850679 0.3693730

1 2 0.6857984 0.3653965

1 3 0.7036091 0.4067256

1 4 0.7009764 0.3967721

1 5 0.7099319 0.4192199

1 6 0.7086092 0.4145249

1 7 0.7190619 0.4375537

1 8 0.7175604 0.4332741

1 9 0.7202437 0.4401120

1 10 0.7197445 0.4379192

1 11 0.7210151 0.4416917

1 12 0.7246053 0.4475940

1 13 0.7284677 0.4567595

1 14 0.7279673 0.4548292

1 15 0.7325586 0.4649856

1 16 0.7352381 0.4697043

1 17 0.7409631 0.4816130

1 18 0.7416893 0.4831296

1 19 0.7417373 0.4828860

1 20 0.7432386 0.4858116

1 21 0.7465573 0.4925287

1 22 0.7469197 0.4938971

1 23 0.7466008 0.4929056

1 24 0.7497369 0.4989817

1 25 0.7530078 0.5055270

1 26 0.7580984 0.5159890

1 27 0.7580095 0.5158374

1 28 0.7600086 0.5198787

1 29 0.7569622 0.5138538

1 30 0.7602357 0.5204530

2 1 0.6947935 0.3866734

2 2 0.7071560 0.4112079

2 3 0.7161541 0.4317498

2 4 0.7230602 0.4442253

2 5 0.7299223 0.4588551

2 6 0.7357392 0.4701117

2 7 0.7386933 0.4769987

2 8 0.7429198 0.4853037

2 9 0.7440539 0.4874374

2 10 0.7498271 0.4990259

2 11 0.7510996 0.5015206

2 12 0.7513256 0.5017905

2 13 0.7538247 0.5069613

2 14 0.7567780 0.5130434

2 15 0.7602321 0.5200429

2 16 0.7622793 0.5241820

2 17 0.7618686 0.5233440

2 18 0.7635493 0.5266520

2 19 0.7657776 0.5312682

2 20 0.7665950 0.5329333

2 21 0.7676398 0.5350291

2 22 0.7681394 0.5360038

2 23 0.7683669 0.5364948

2 24 0.7680944 0.5358427

2 25 0.7686866 0.5370277

2 26 0.7703214 0.5403302

2 27 0.7707764 0.5412720

2 28 0.7716392 0.5430078

2 29 0.7715942 0.5429083

2 30 0.7725934 0.5449127

3 1 0.6955226 0.3855078

3 2 0.7209675 0.4386526

3 3 0.7356482 0.4695812

3 4 0.7449175 0.4878837

3 5 0.7492812 0.4969745

3 6 0.7522363 0.5033353

3 7 0.7538717 0.5065183

3 8 0.7585985 0.5164615

3 9 0.7590974 0.5176961

3 10 0.7628699 0.5254664

3 11 0.7659595 0.5314943

3 12 0.7644595 0.5282614

3 13 0.7647322 0.5288951

3 14 0.7659597 0.5313854

3 15 0.7678215 0.5351474

3 16 0.7703201 0.5402119

3 17 0.7713190 0.5422716

3 18 0.7723202 0.5442955

3 19 0.7716377 0.5429688

3 20 0.7742739 0.5482294

3 21 0.7731375 0.5459746

3 22 0.7735922 0.5468735

3 23 0.7735924 0.5468940

3 24 0.7731382 0.5459877

3 25 0.7745922 0.5488905

3 26 0.7748196 0.5493556

3 27 0.7751379 0.5500070

3 28 0.7756830 0.5511015

3 29 0.7774552 0.5546298

3 30 0.7768199 0.5533831

Tuning parameter 'shrinkage' was held constant at a value of 0.1

Tuning parameter 'n.minobsinnode' was held constant at a value of 20

Accuracy was used to select the optimal model using the largest value.

The final values used for the model were n.trees = 29, interaction.depth =

3, shrinkage = 0.1 and n.minobsinnode = 20.

**Stochastic Gradient Boosting**

4401 samples

22 predictor

2 classes: '0', '1'

Pre-processing: scaled (22), centered (22)

Resampling: Cross-Validated (5 fold, repeated 5 times)

Summary of sample sizes: 3520, 3521, 3522, 3521, 3520, 3521, ...

Resampling results across tuning parameters:

interaction.depth n.trees Accuracy Kappa

3 10 0.7611012 0.5213696

3 20 0.7713698 0.5423862

3 30 0.7770490 0.5538016

3 40 0.7794580 0.5586443

3 50 0.7810933 0.5619744

3 60 0.7811387 0.5620445

3 70 0.7815932 0.5629378

3 80 0.7814115 0.5625820

3 90 0.7819110 0.5635948

3 100 0.7810931 0.5619348

3 110 0.7818658 0.5634736

3 120 0.7824107 0.5645716

3 130 0.7825920 0.5649231

3 140 0.7810020 0.5617252

3 150 0.7814112 0.5625593

3 160 0.7813652 0.5624526

3 170 0.7813195 0.5623609

3 180 0.7812288 0.5621963

3 190 0.7808653 0.5614847

3 200 0.7809562 0.5616762

3 210 0.7816830 0.5631231

3 220 0.7820017 0.5637542

3 230 0.7819563 0.5636566

3 240 0.7819565 0.5636614

3 250 0.7820925 0.5639285

3 260 0.7802753 0.5602840

3 270 0.7795025 0.5587413

3 280 0.7806842 0.5611123

3 290 0.7805028 0.5607440

3 300 0.7802762 0.5602790

4 10 0.7697789 0.5390732

4 20 0.7747328 0.5491864

4 30 0.7783685 0.5564968

4 40 0.7792765 0.5583435

4 50 0.7803669 0.5605508

4 60 0.7810030 0.5618148

4 70 0.7819570 0.5637188

4 80 0.7818204 0.5634533

4 90 0.7813665 0.5625206

4 100 0.7825472 0.5648736

4 110 0.7821386 0.5640374

4 120 0.7831385 0.5660487

4 130 0.7824113 0.5645846

4 140 0.7825473 0.5648638

4 150 0.7820924 0.5639323

4 160 0.7812754 0.5622930

4 170 0.7816387 0.5630145

4 180 0.7828661 0.5654609

4 190 0.7832745 0.5662822

4 200 0.7827298 0.5652003

4 210 0.7828194 0.5653855

4 220 0.7830470 0.5658297

4 230 0.7830466 0.5658315

4 240 0.7836366 0.5670143

4 250 0.7830010 0.5657468

4 260 0.7824113 0.5645643

4 270 0.7832746 0.5663012

4 280 0.7824110 0.5645343

4 290 0.7819566 0.5636428

4 300 0.7807753 0.5612899

5 10 0.7689153 0.5374359

5 20 0.7750961 0.5499047

5 30 0.7778229 0.5554006

5 40 0.7805484 0.5608757

5 50 0.7813666 0.5625061

5 60 0.7830925 0.5659589

5 70 0.7831386 0.5660659

5 80 0.7826387 0.5650398

5 90 0.7831383 0.5660495

5 100 0.7828660 0.5655269

5 110 0.7830480 0.5658741

5 120 0.7825470 0.5648674

5 130 0.7821837 0.5641451

5 140 0.7818194 0.5634104

5 150 0.7818650 0.5634963

5 160 0.7820020 0.5637862

5 170 0.7820914 0.5639311

5 180 0.7810915 0.5619336

5 190 0.7806375 0.5610438

5 200 0.7795913 0.5589394

5 210 0.7793653 0.5584755

5 220 0.7796379 0.5590293

5 230 0.7790014 0.5577838

5 240 0.7782745 0.5563325

5 250 0.7779562 0.5556853

5 260 0.7775466 0.5548539

5 270 0.7776374 0.5550397

5 280 0.7777737 0.5553121

5 290 0.7788188 0.5574153

5 300 0.7773642 0.5544840

6 10 0.7727777 0.5451404

6 20 0.7755053 0.5507248

6 30 0.7782312 0.5562153

6 40 0.7797315 0.5592338

6 50 0.7813219 0.5624249

6 60 0.7815951 0.5629828

6 70 0.7824128 0.5646225

6 80 0.7818677 0.5634868

6 90 0.7825932 0.5649687

6 100 0.7824113 0.5646046

6 110 0.7833653 0.5665057

6 120 0.7827754 0.5653355

6 130 0.7831844 0.5661431

6 140 0.7826843 0.5651323

6 150 0.7828667 0.5655082

6 160 0.7815477 0.5628686

6 170 0.7822753 0.5643292

6 180 0.7818196 0.5634111

6 190 0.7802759 0.5603118

6 200 0.7801857 0.5601293

6 210 0.7797311 0.5592099

6 220 0.7803661 0.5604711

6 230 0.7793664 0.5584928

6 240 0.7791837 0.5581350

6 250 0.7794109 0.5585890

6 260 0.7784109 0.5566064

6 270 0.7760488 0.5518912

6 280 0.7765033 0.5527850

6 290 0.7759582 0.5516696

6 300 0.7765936 0.5529578

Tuning parameter 'shrinkage' was held constant at a value of 0.1

Tuning parameter 'n.minobsinnode' was held constant at a value of 20

Accuracy was used to select the optimal model using the largest value.

The final values used for the model were n.trees = 240, interaction.depth =

4, shrinkage = 0.1 and n.minobsinnode = 20.

**Results**

Accuracy was used to select the optimal model using the largest value.

The final values used for the model were n.trees = 110, interaction.depth =

5, shrinkage = 0.1 and n.minobsinnode = 20.

Random forests

Random Forest

5500 samples

22 predictor

2 classes: '0', '1'

Pre-processing: scaled (22), centered (22)

Resampling: Cross-Validated (5 fold, repeated 5 times)

Summary of sample sizes: 4399, 4400, 4401, 4401, 4399, 4401, ...

Resampling results across tuning parameters:

|  |  |  |  |
| --- | --- | --- | --- |
| Training set size | Test set size | Feature vectors | Classification |
| 5500 | 5952 | 22 | binary |

mtry Accuracy Kappa

1 0.7729462 0.5456173

2 0.7724010 0.5445959

3 0.7726548 0.5451103

4 0.7726548 0.5451241

5 0.7715650 0.5429408

6 0.7706922 0.5411948

Accuracy was used to select the optimal model using the largest value.

The final value used for the model was mtry = 1.