Simulated examples for the survival ensemble methods

Select data type

ELSA_Diabetes_Study2

Time point for event prediction:

10

Random seed for calibration and validation

42

K_Outer loop CV (for validation)

3

K_ inner CV folds (model tuning)

3

Simulated data: random seed (generation):

4242

Sample size:

150

Observation time

5

Expected event prevalence by study end

0.5

Expected drop out rate

0.3

Custom data: path to data file

~/Desktop/Study_KCL/PhD P

Predictors to use in the model

"baseline_age_", "genderdur

Time variable name

time

Event indicator variable name

event

Sample statistics CoxPH SRF Ens1: CoxPH->SRF Ens2: CoxPH in clusters

Ens3: extended CoxPH Summary Conclusions

Internally cross-validated results:

Show 25 **entries** **Search:**

	AUCROC	BS	BS_scaled	C_score	Calib_slope	Calib_alpha	T
test	0.778	0.0769	0.1017	0.7468	1.0648	0.046	10
train	0.7909	0.0754	0.1195	0.7567	1.1225	0.045	10

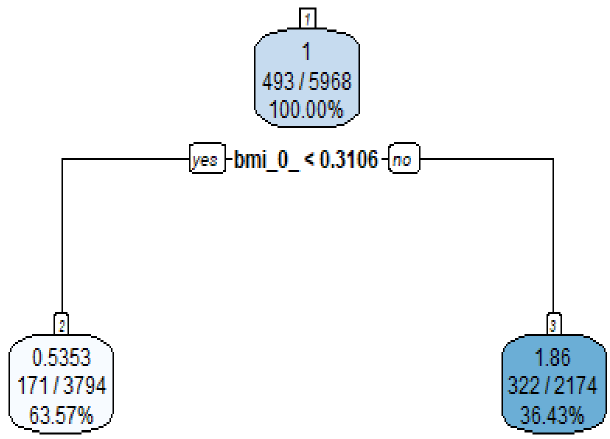
Showing 1 to 2 of 2 entries Previous 1 Next

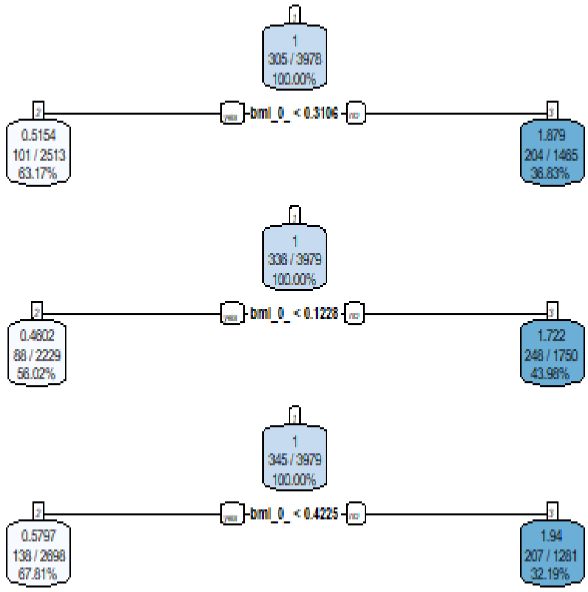
Internally cross-validated Test results for each CV fold:

Show 25 **entries** **Search:**

	AUCROC	BS	BS_scaled	C_score	Calib_slope	Calib_alpha	T
test.1	0.7693	0.086	0.1027	0.7404	1.069	0.0743	10
test.2	0.7625	0.0756	0.0767	0.7236	0.9232	0.0365	10
test.3	0.8023	0.0691	0.1258	0.7765	1.2022	0.0273	10

Showing 1 to 3 of 3 entries Previous 1 Next





Show 10 entries Search:

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
sz20_	0.0314	1.0319	0.0454	0.6925	0.4886
pc1_	0.0212	1.0214	0.0453	0.4673	0.6403
pc2_	-0.0651	0.9369	0.0452	-1.4404	0.1498
pc3_	0.0072	1.0072	0.0472	0.1528	0.8786
pc4_	-0.0105	0.9896	0.0444	-0.2357	0.8137
age_	0.1754	1.1917	0.0563	3.1128	0.0019
sex	-0.3062	0.7363	0.1029	-2.9766	0.0029
bmi_0_	0.3085	1.3614	0.0604	5.109	0
hyp_0	0.4904	1.6329	0.0941	5.2111	0
cvd_0	-0.0331	0.9674	0.1327	-0.2499	0.8027

Showing 1 to 10 of 23 entries Previous 1 2 3 Next

```
$test
   T    AUCROC         BS  BS_scaled   C_score
1 10 0.7693149 0.08599898 0.10270605 0.7403549
2 10 0.7624956 0.07564321 0.07669279 0.7235991
3 10 0.8022541 0.06909455 0.12579893 0.7764802
  Calib_slope Calib_alpha test cv_n
1   1.0689787  0.07425385    1    1
2   0.9231891  0.03649639    1    2
3   1.2021587  0.02730576    1    3

$train
   T    AUCROC         BS BS_scaled   C_score
1 10 0.7982134 0.07084518 0.1236720 0.7581750
2 10 0.7982208 0.07649145 0.1272103 0.7687648
3 10 0.7763923 0.07894850 0.1075090 0.7432108
  Calib_slope Calib_alpha test cv_n
1    1.144057  0.04176389    0    1
2    1.110478  0.04582521    0    2
3    1.113096  0.04752583    0    3

$testaverage
          T      AUCROC          BS   BS_scaled
10.00000000  0.77802152  0.07691225  0.10173259
    C_score Calib_slope Calib_alpha        test
 0.74681141  1.06477551  0.04601867  1.00000000

$trainaverage
          T      AUCROC          BS   BS_scaled
10.00000000  0.79094217  0.07542838  0.11946377
    C_score Calib_slope Calib_alpha        test
 0.75671687  1.12254366  0.04503831  0.00000000

$model_list
$model_list[[1]]
$model_list[[1]]$treemodel
n= 3978

node), split, n, deviance, yval
      * denotes terminal node

1) root 3978 2090.3410 1.0000000
  2) bmi_0_< 0.310594 2513  856.6503 0.5153783 *
  3) bmi_0_>=0.310594 1465 1109.4060 1.8792320 *

$model_list[[1]]$modcoxmodel
Call:
coxph(formula = as.formula(paste("Surv(df_train$time, df_train$event) ~",
    paste(predict.factors, collapse = "+"))), data = df_train,
    x = TRUE)

                          coef exp(coef)  se(coef)
sz20_                -0.031330  0.969156  0.055656
pc1_                  0.016134  1.016265  0.057637
pc2_                 -0.093042  0.911155  0.057887
pc3_                 -0.024301  0.975992  0.060319
```

```
pc4_                 -0.038358  0.962369  0.058286
age_                  0.144791  1.155798  0.071801
sex                  -0.353514  0.702216  0.131318
bmi_0_                0.252468  1.287199  0.082062
hyp_0                 0.538283  1.713064  0.120599
cvd_0                -0.025864  0.974468  0.168937
B_dep_0               0.340963  1.406301  0.151434
trig_0                0.096417  1.101218  0.035706
baseline_hdl         -0.297892  0.742381  0.197442
stroke_0              0.315922  1.371523  0.303216
B_smokstatus_0        0.409015  1.505334  0.151475
exercise_light       -0.003738  0.996269  0.244836
exercise_vig         -0.330648  0.718458  0.145688
EduLevel_low          0.332327  1.394208  0.220431
EduLevel_med          0.160220  1.173769  0.209952
wealth_med            0.060801  1.062687  0.159759
wealth_low            0.234254  1.263965  0.161578
t2dm_                 0.321573  1.379296  0.059779
cluster_tree1.879232  0.690132  1.993978  0.176159
                          z        p
sz20_                -0.563  0.57349
pc1_                  0.280  0.77954
pc2_                 -1.607  0.10799
pc3_                 -0.403  0.68704
pc4_                 -0.658  0.51048
age_                  2.017  0.04374
sex                  -2.692  0.00710
bmi_0_                3.077  0.00209
hyp_0                 4.463 8.07e-06
cvd_0                -0.153  0.87832
B_dep_0               2.252  0.02435
trig_0                2.700  0.00693
baseline_hdl         -1.509  0.13136
stroke_0              1.042  0.29746
B_smokstatus_0        2.700  0.00693
exercise_light       -0.015  0.98782
exercise_vig         -2.270  0.02323
EduLevel_low          1.508  0.13165
EduLevel_med          0.763  0.44539
wealth_med            0.381  0.70352
wealth_low            1.450  0.14712
t2dm_                 5.379 7.47e-08
cluster_tree1.879232  3.918 8.94e-05

Likelihood ratio test=273.3  on 23 df, p=< 2.2e-16
n= 3978, number of events= 305

$model_list[[1]]$clusters
[1] 1.879232 0.515378


$model_list[[2]]
$model_list[[2]]$treemodel
n= 3979

node), split, n, deviance, yval
```

```
      * denotes terminal node

1) root 3979 2236.1680 1.0000000
  2) bmi_0_< 0.1228089 2229  743.9276 0.4601545 *
  3) bmi_0_>=0.1228089 1750 1358.8730 1.7221500 *

$model_list[[2]]$modcoxmodel
Call:
coxph(formula = as.formula(paste("Surv(df_train$time, df_train$event) ~",
    paste(predict.factors, collapse = "+"))), data = df_train,
    x = TRUE)

                         coef exp(coef)  se(coef)      z
sz20_                0.056481  1.058107  0.056765  0.995
pc1_                 0.021115  1.021339  0.054847  0.385
pc2_                -0.013904  0.986193  0.055185 -0.252
pc3_                 0.036449  1.037122  0.059384  0.614
pc4_                 0.034536  1.035140  0.052262  0.661
age_                 0.243151  1.275261  0.067700  3.592
sex                 -0.265403  0.766897  0.124144 -2.138
bmi_0_               0.322843  1.381048  0.068618  4.705
hyp_0                0.557509  1.746316  0.114264  4.879
cvd_0               -0.044285  0.956681  0.159025 -0.278
B_dep_0              0.331283  1.392754  0.148077  2.237
trig_0               0.104463  1.110114  0.031824  3.282
baseline_hdl        -0.566958  0.567249  0.192600 -2.944
stroke_0             0.519228  1.680729  0.251577  2.064
B_smokstatus_0       0.375297  1.455424  0.148538  2.527
exercise_light       0.003652  1.003659  0.234295  0.016
exercise_vig        -0.215276  0.806319  0.133434 -1.613
EduLevel_low         0.396542  1.486675  0.222224  1.784
EduLevel_med         0.331270  1.392735  0.210705  1.572
wealth_med           0.001424  1.001425  0.149084  0.010
wealth_low           0.167137  1.181916  0.152430  1.096
t2dm_                0.307085  1.359456  0.056194  5.465
cluster_tree1.72215  0.613150  1.846237  0.161722  3.791
                           p
sz20_               0.319731
pc1_                0.700258
pc2_                0.801084
pc3_                0.539358
pc4_                0.508717
age_                0.000329
sex                 0.032528
bmi_0_              2.54e-06
hyp_0               1.07e-06
cvd_0               0.780644
B_dep_0             0.025271
trig_0              0.001029
baseline_hdl        0.003243
stroke_0            0.039028
B_smokstatus_0      0.011517
exercise_light      0.987564
exercise_vig        0.106668
EduLevel_low        0.074354
EduLevel_med        0.115905
```

```
wealth_med          0.992379
wealth_low          0.272869
t2dm_               4.64e-08
cluster_tree1.72215 0.000150

Likelihood ratio test=325  on 23 df, p=< 2.2e-16
n= 3979, number of events= 336

$model_list[[2]]$clusters
[1] 1.722150 0.460155


$model_list[[3]]
$model_list[[3]]$treemodel
n= 3979

node), split, n, deviance, yval
      * denotes terminal node

1) root 3979 2275.804 1.0000000
  2) bmi_0_< 0.422484 2698 1074.604 0.5797049 *
  3) bmi_0_>=0.422484 1281 1076.318 1.9398860 *

$model_list[[3]]$modcoxmodel
Call:
coxph(formula = as.formula(paste("Surv(df_train$time, df_train$event) ~",
    paste(predict.factors, collapse = "+"))), data = df_train,
    x = TRUE)

                         coef exp(coef) se(coef)      z
sz20_                 0.06340   1.06545  0.05533  1.146
pc1_                  0.02630   1.02665  0.05463  0.481
pc2_                 -0.08886   0.91497  0.05411 -1.642
pc3_                  0.01407   1.01417  0.05469  0.257
pc4_                 -0.04171   0.95914  0.05374 -0.776
age_                  0.14376   1.15461  0.06817  2.109
sex                  -0.31699   0.72834  0.12434 -2.549
bmi_0_                0.31586   1.37143  0.07165  4.408
hyp_0                 0.40267   1.49581  0.11239  3.583
cvd_0                -0.02466   0.97565  0.16111 -0.153
B_dep_0               0.24557   1.27835  0.14924  1.645
trig_0                0.10870   1.11483  0.04025  2.700
baseline_hdl         -0.42387   0.65451  0.18952 -2.237
stroke_0              0.45491   1.57604  0.28328  1.606
B_smokstatus_0        0.39766   1.48834  0.14121  2.816
exercise_light       -0.14408   0.86582  0.24121 -0.597
exercise_vig         -0.20064   0.81821  0.12899 -1.556
EduLevel_low          0.38005   1.46236  0.21311  1.783
EduLevel_med          0.31994   1.37704  0.20092  1.592
wealth_med           -0.05387   0.94755  0.14335 -0.376
wealth_low            0.09223   1.09662  0.14809  0.623
t2dm_                 0.29383   1.34156  0.05609  5.239
cluster_tree1.939886  0.56730   1.76350  0.15965  3.554
                            p
sz20_                 0.25188
pc1_                  0.63017
```

```
pc2_                     0.10052
pc3_                     0.79700
pc4_                     0.43764
age_                     0.03496
sex                      0.01079
bmi_0_                  1.04e-05
hyp_0                    0.00034
cvd_0                    0.87837
B_dep_0                  0.09988
trig_0                   0.00693
baseline_hdl             0.02532
stroke_0                 0.10830
B_smokstatus_0           0.00486
exercise_light           0.55029
exercise_vig             0.11983
EduLevel_low             0.07453
EduLevel_med             0.11131
wealth_med               0.70706
wealth_low               0.53339
t2dm_                   1.62e-07
cluster_tree1.939886  0.00038

Likelihood ratio test=270.2  on 23 df, p=< 2.2e-16
n= 3979, number of events= 345

$model_list[[3]]$clusters
[1] 1.939886 0.579705



$time
Time difference of 35.42474 secs
```