# Cyber data analysis – fraud detection结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Ds1 | Ds2 | Ds3 | Ds4 |
| Fraud rate | 15% | 10% | 5% | 2% |
| After smote | 1:1 (340) | 1:1  (630) | 1:1  (1330) | 1:1  (3430) |

## BEFORE SMOTE

DS1:

random forest accuracy:

0.95794254330376

TP: 28

FP: 1949

FN: 42

TN: 45321

[[45321 1949]

[ 42 28]]

precision: [0.00147866 0.01416287 1. ]

recall: [1. 0.4 0. ]

DS2:

random forest accuracy:

0.9624841571609632

TP: 28

FP: 1734

FN: 42

TN: 45536

[[45536 1734]

[ 42 28]]

precision: [0.00147866 0.01589103 1. ]

recall: [1. 0.4 0. ]

DS3:

random forest accuracy:

0.9930502746092099

TP: 3

FP: 262

FN: 67

TN: 47008

[[47008 262]

[ 67 3]]

precision: [0.00147866 0.01132075 1. ]

recall: [1. 0.04285714 0. ]

DS4:

random forest accuracy:

0.9985213350232361

TP: 0

FP: 0

FN: 70

TN: 47270

[[47270 0]

[ 70 0]]

precision: [0.00147866 1. ]

recall: [1. 0.]

## AFTER SMOTE

DS1:

Counter({0: 340, 1: 70})

Resampled dataset shape Counter({0.0: 340, 1.0: 340})

random forest accuracy:

0.8376003379805661

TP: 61

FP: 7679

FN: 9

TN: 39591

[[39591 7679]

[ 9 61]]

precision: [0.00147866 0.00788114 1. ]

recall: [1. 0.87142857 0. ]

DS2:

Counter({0: 630, 1: 70})

Resampled dataset shape Counter({0.0: 630, 1.0: 630})

random forest accuracy:

0.8953738910012674

TP: 59

FP: 4942

FN: 11

TN: 42328

[[42328 4942]

[ 11 59]]

precision: [0.00147866 0.01179764 1. ]

recall: [1. 0.84285714 0. ]

DS3:

Counter({0: 1330, 1: 70})

Resampled dataset shape Counter({0.0: 1330, 1.0: 1330})

random forest accuracy:

0.9043937473595268

TP: 58

FP: 4514

FN: 12

TN: 42756

[[42756 4514]

[ 12 58]]

precision: [0.00147866 0.01268591 1. ]

recall: [1. 0.82857143 0. ]

DS4:

Counter({0: 3430, 1: 70})

Resampled dataset shape Counter({0.0: 3430, 1.0: 3430})

random forest accuracy:

0.9385086607520068

TP: 57

FP: 2898

FN: 13

TN: 44372

[[44372 2898]

[ 13 57]]

precision: [0.00147866 0.01928934 1. ]

recall: [1. 0.81428571 0. ]

#---------------------10 fold cross validation

('TRAIN:', array([ 0, 1, 2, ..., 236693, 236694, 236695]), 'TEST:', array([ 8, 10, 14, ..., 236637, 236656, 236689]))

Counter({0: 212714, 1: 312})

random forest accuracy:

0.834980988593

[[19735 3902]

[ 4 29]]

recall: [ 1. 0.87878788 0. ]

('TRAIN:', array([ 1, 2, 3, ..., 236692, 236694, 236695]), 'TEST:', array([ 0, 5, 9, ..., 236657, 236691, 236693]))

Counter({0: 212710, 1: 316})

Resampled dataset shape Counter({0.0: 212710, 1.0: 212710})

random forest accuracy:

0.838318546684

[[19817 3824]

[ 3 26]]

recall: [ 1. 0.89655172 0. ]

('TRAIN:', array([ 0, 1, 2, ..., 236693, 236694, 236695]), 'TEST:', array([ 12, 23, 31, ..., 236616, 236619, 236686]))

Counter({0: 212720, 1: 306})

Resampled dataset shape Counter({0.0: 212720, 1.0: 212720})

random forest accuracy:

0.836924376848

[[19779 3852]

[ 8 31]]

recall: [ 1. 0.79487179 0. ]

('TRAIN:', array([ 0, 1, 2, ..., 236693, 236694, 236695]), 'TEST:', array([ 18, 25, 33, ..., 236668, 236669, 236673]))

Counter({0: 212718, 1: 308})

Resampled dataset shape Counter({0.0: 212718, 1.0: 212718})

random forest accuracy:

0.83810730883

[[19806 3827]

[ 5 32]]

recall: [ 1. 0.86486486 0. ]

('TRAIN:', array([ 0, 1, 2, ..., 236693, 236694, 236695]), 'TEST:', array([ 11, 20, 38, ..., 236687, 236688, 236690]))

Counter({0: 212715, 1: 311})

Resampled dataset shape Counter({0.0: 212715, 1.0: 212715})

random forest accuracy:

0.83206590621

[[19666 3970]

[ 5 29]]

recall: [ 1. 0.85294118 0. ]

('TRAIN:', array([ 0, 1, 2, ..., 236693, 236694, 236695]), 'TEST:', array([ 21, 28, 45, ..., 236681, 236683, 236692]))

Counter({0: 212716, 1: 310})

Resampled dataset shape Counter({0.0: 212716, 1.0: 212716})

random forest accuracy:

0.835065483735

[[19737 3898]

[ 6 29]]

recall: [ 1. 0.82857143 0. ]

('TRAIN:', array([ 0, 1, 4, ..., 236693, 236694, 236695]), 'TEST:', array([ 2, 3, 6, ..., 236603, 236635, 236639]))

Counter({0: 212716, 1: 311})

Resampled dataset shape Counter({0.0: 212716, 1.0: 212716})

random forest accuracy:

0.842071908403

[[19906 3729]

[ 9 25]]

recall: [ 1. 0.73529412 0. ]

('TRAIN:', array([ 0, 2, 3, ..., 236692, 236693, 236695]), 'TEST:', array([ 1, 17, 41, ..., 236679, 236685, 236694]))

Counter({0: 212708, 1: 319})

Resampled dataset shape Counter({0.0: 212708, 1.0: 212708})

random forest accuracy:

0.834213528244

[[19720 3923]

[ 1 25]]

recall: [ 1. 0.96153846 0. ]

('TRAIN:', array([ 0, 1, 2, ..., 236692, 236693, 236694]), 'TEST:', array([ 13, 34, 39, ..., 236682, 236684, 236695]))

Counter({0: 212709, 1: 318})

Resampled dataset shape Counter({0.0: 212709, 1.0: 212709})

random forest accuracy:

0.835650006337

[[19758 3884]

[ 6 21]]

recall: [ 1. 0.77777778 0. ]

('TRAIN:', array([ 0, 1, 2, ..., 236693, 236694, 236695]), 'TEST:', array([ 4, 29, 30, ..., 236671, 236674, 236677]))

Counter({0: 212733, 1: 294})

Resampled dataset shape Counter({0.0: 212733, 1.0: 212733})

random forest accuracy:

0.832523554016

[[19664 3954]

[ 10 41]]

recall: [ 1. 0.80392157 0. ]