# 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. ]