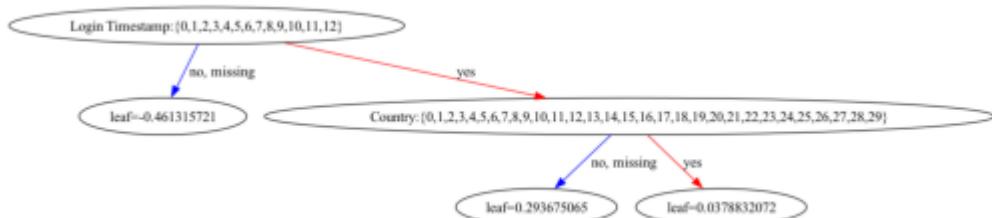


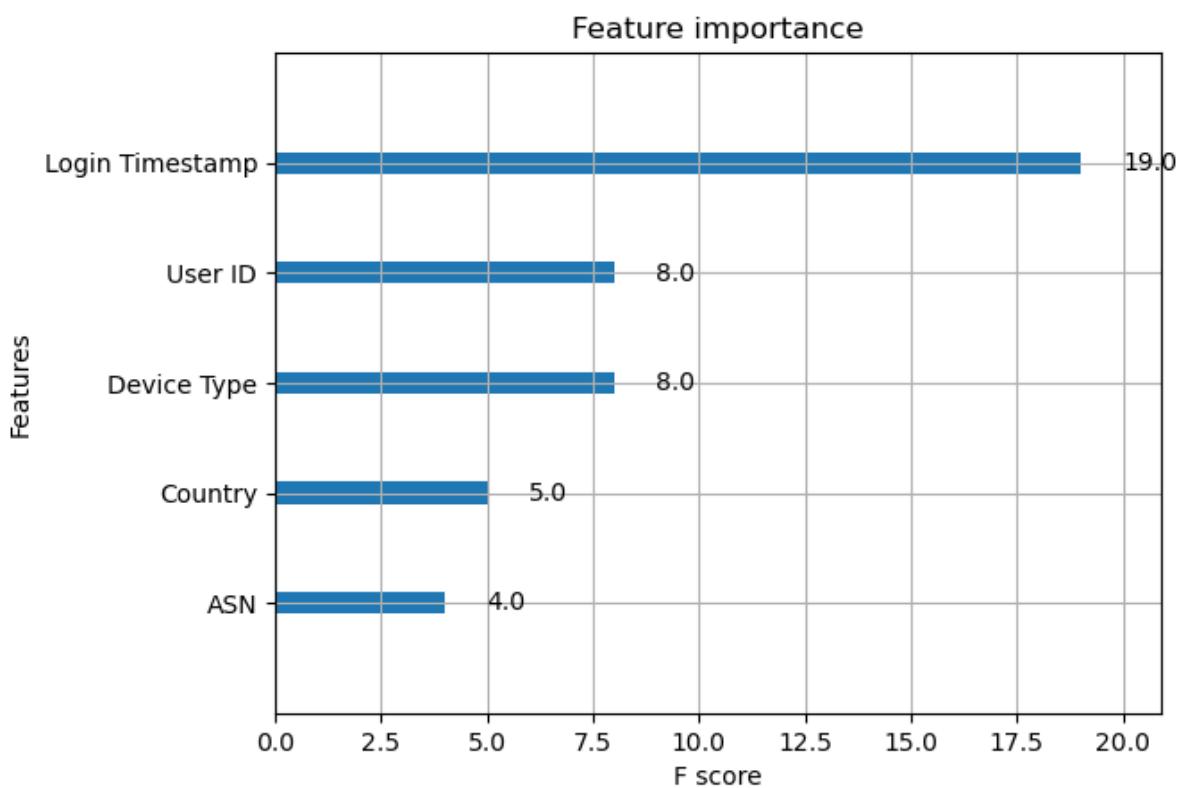
ATO 500 rows

I. 10/500 → 2%

1. XGB Experimental

```
---      ---      ---  
0 ASN          500 non-null    float64  
1 Country       500 non-null    category  
2 Device Type   500 non-null    category  
3 IP Address    500 non-null    category  
4 Is Attack IP  500 non-null    bool  
5 Login Successful 500 non-null    bool  
6 Login Timestamp 500 non-null    category  
7 User ID        500 non-null    float64  
8 Blacklisted    500 non-null    bool  
9 Browser Type   500 non-null    category  
dtypes: bool(3), category(5), float64(2)  
memory usage: 45.4 KB  
Feature importances:  
[0.02398985 0.27313262 0.11431003 0.          0.  
 0.56790614 0.02066133 0.          0.          ]
```





2. Label encoding (cat.codes)

2A. XGBoost

```
memory usage: 39.7+ KB
```

```

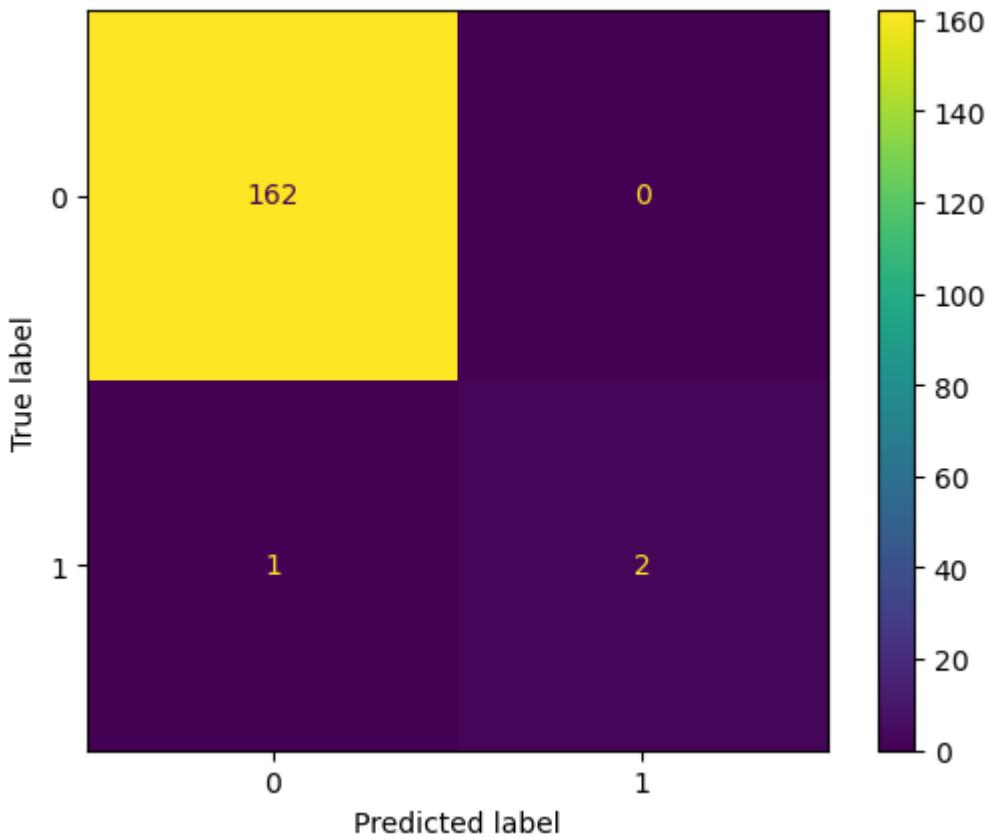
predictor=none, random_state=None, ...)

Accuracy: 99.39%
      precision    recall  f1-score   support

        False       0.99     1.00     1.00      162
         True       1.00     0.67     0.80       3

    accuracy          0.99
   macro avg       1.00     0.83     0.90      165
weighted avg       0.99     0.99     0.99      165

```



2B. ONE CLASS SVM

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.2 (same for any other col)

```

The number of records in the training dataset is 400
The number of records in the test dataset is 100
The training dataset has 392 records for the majority class and 8 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00      98
  True        0.02     1.00     0.04      2
  accuracy         -         -         -      100
  macro avg       0.01     0.50     0.02      100
  weighted avg    0.00     0.02     0.00      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.98     1.00     0.99      98
  True        0.00     0.00     0.00      2
  accuracy         -         -         -      100
  macro avg       0.49     0.50     0.49      100
  weighted avg    0.96     0.98     0.97      100

[[ 0 98]
 [ 0  2]]

```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.3 (same for any other col)

```
The number of records in the training dataset is 350
The number of records in the test dataset is 150
The training dataset has 342 records for the majority class and 8 records for the minority class.

precision    recall   f1-score   support
False        0.00     0.00     0.00      148
True         0.01     1.00     0.03       2

accuracy          0.01
macro avg       0.01     0.50     0.01      150
weighted avg    0.00     0.01     0.00      150

The customized score threshold for 2% of outliers is 0.00
precision    recall   f1-score   support
False        0.99     1.00     0.99      148
True         0.00     0.00     0.00       2

accuracy          0.99
macro avg       0.49     0.50     0.50      150
weighted avg    0.97     0.99     0.98      150

[[ 0 148]
 [ 0  2]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.5 (same for any other col)

```
The number of records in the training dataset is 250
The number of records in the test dataset is 250
The training dataset has 243 records for the majority class and 7 records for the minority class.

precision    recall   f1-score   support
False        0.00     0.00     0.00      247
True         0.01     1.00     0.02       3

accuracy          0.01
macro avg       0.01     0.50     0.01      250
weighted avg    0.00     0.01     0.00      250

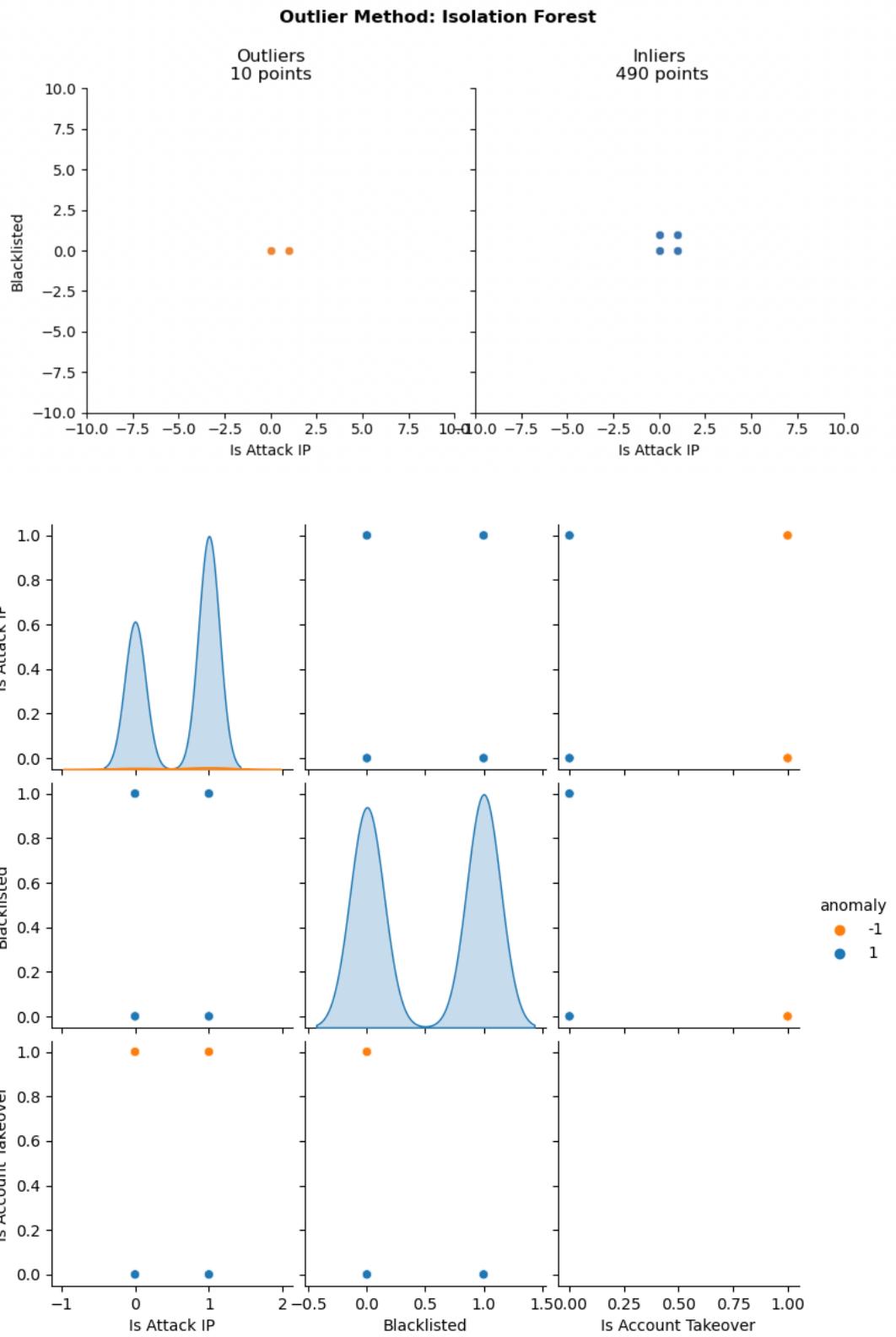
The customized score threshold for 2% of outliers is 0.00
precision    recall   f1-score   support
False        0.99     1.00     0.99      247
True         0.00     0.00     0.00       3

accuracy          0.99
macro avg       0.49     0.50     0.50      250
weighted avg    0.98     0.99     0.98      250

[[ 0 247]
 [ 0  3]]
```

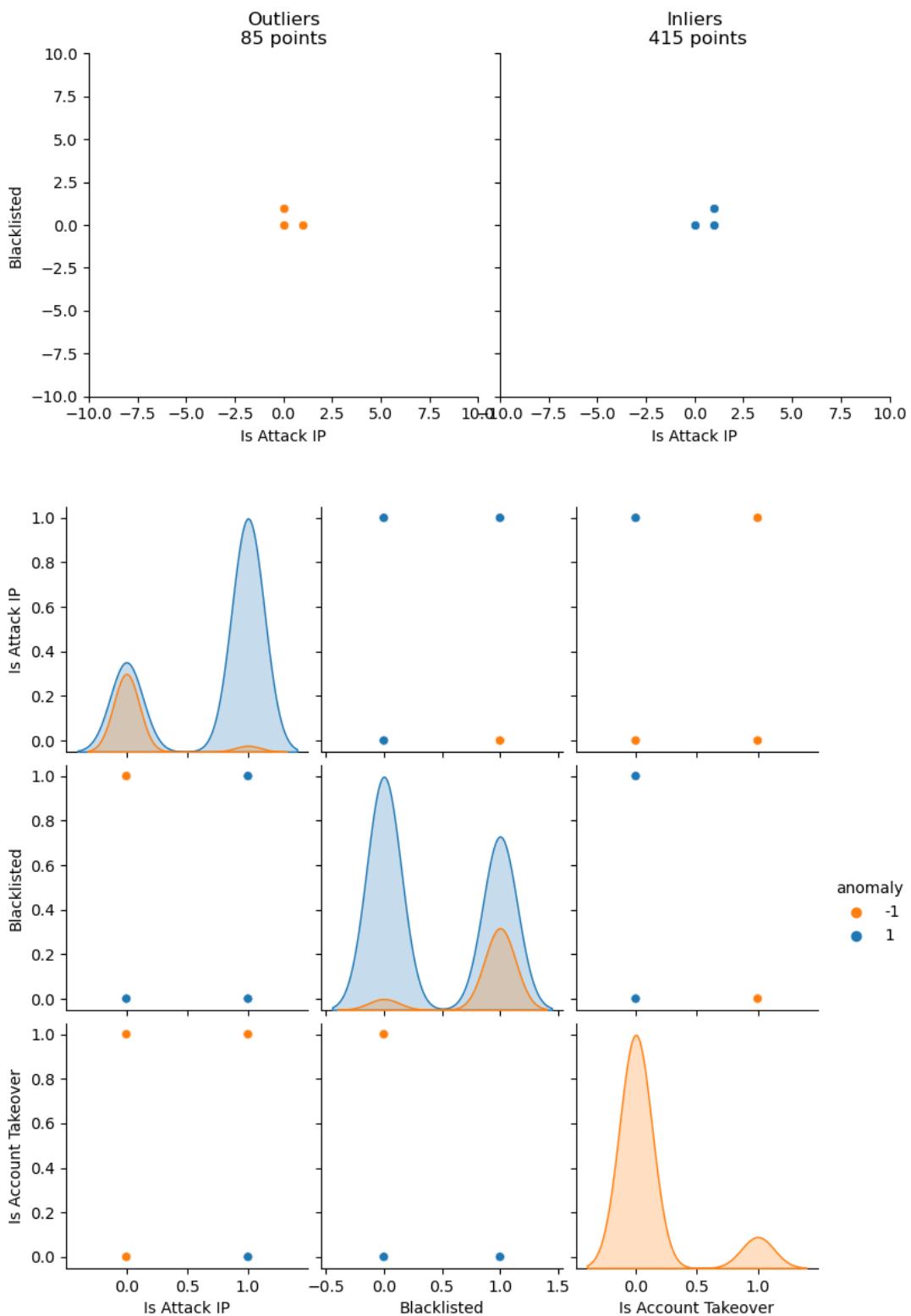
2C. Isolation Forest

- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



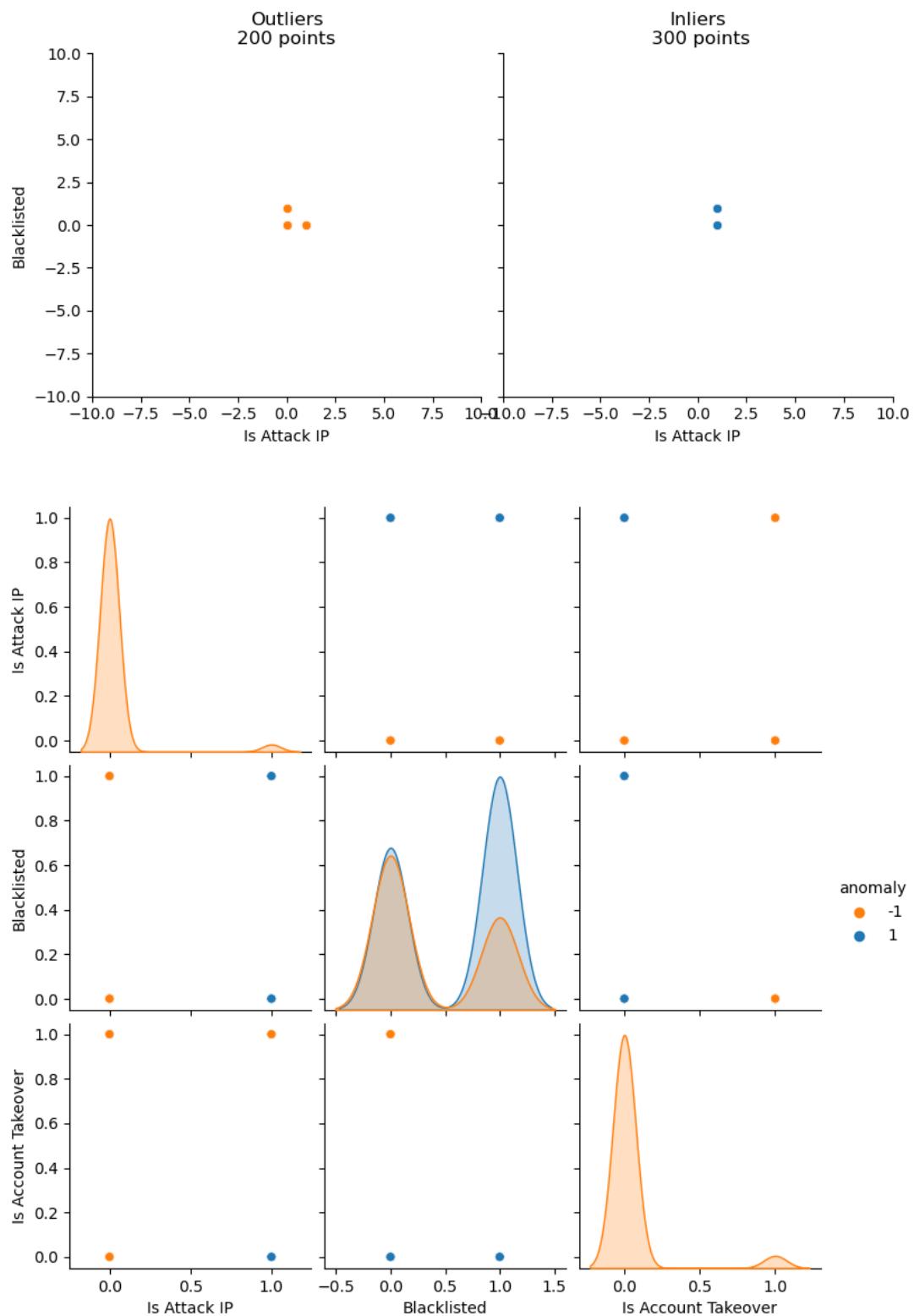
- increasing contamination value to 0.3

Outlier Method: Isolation Forest



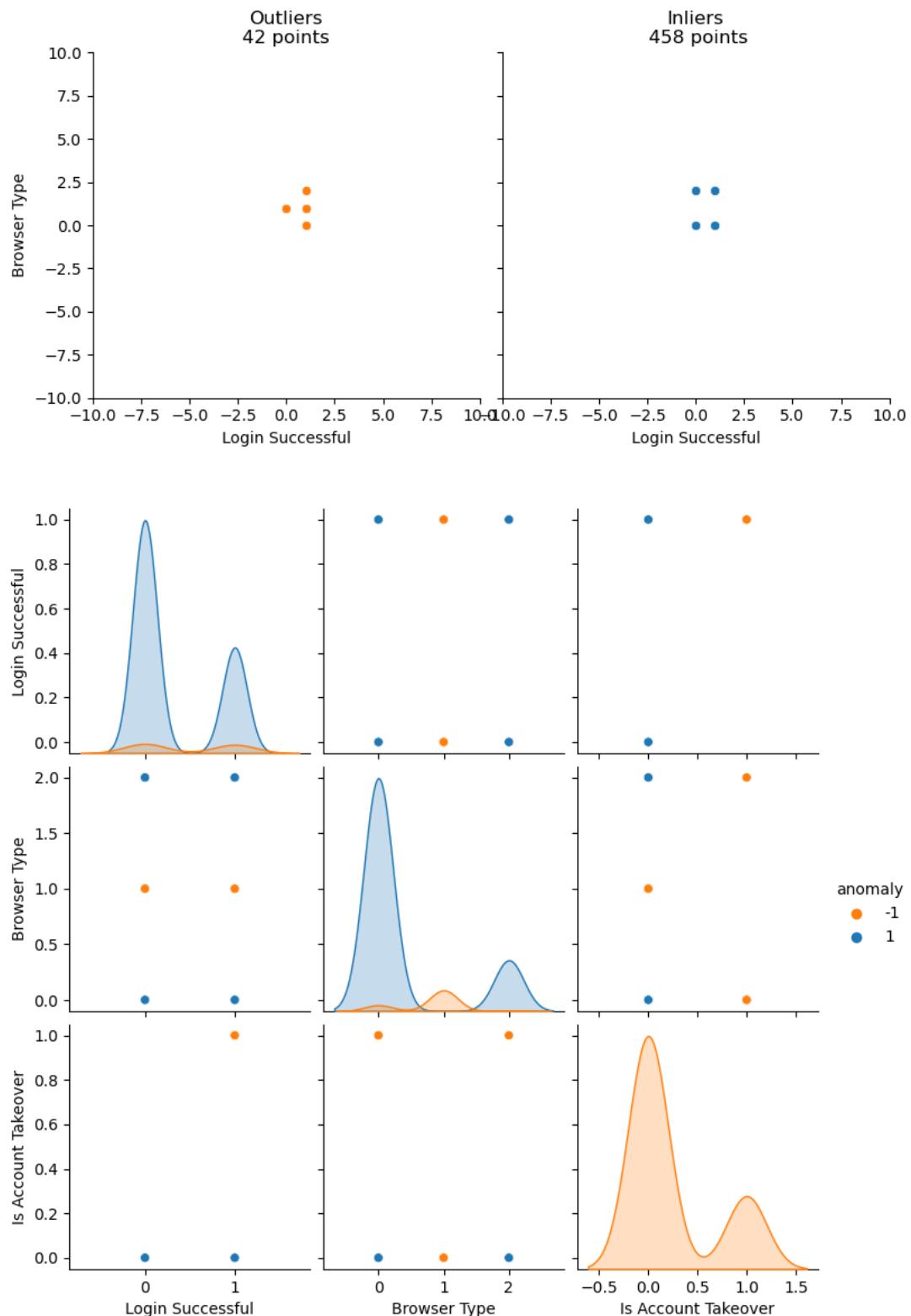
- contamination value == 0.5

Outlier Method: Isolation Forest



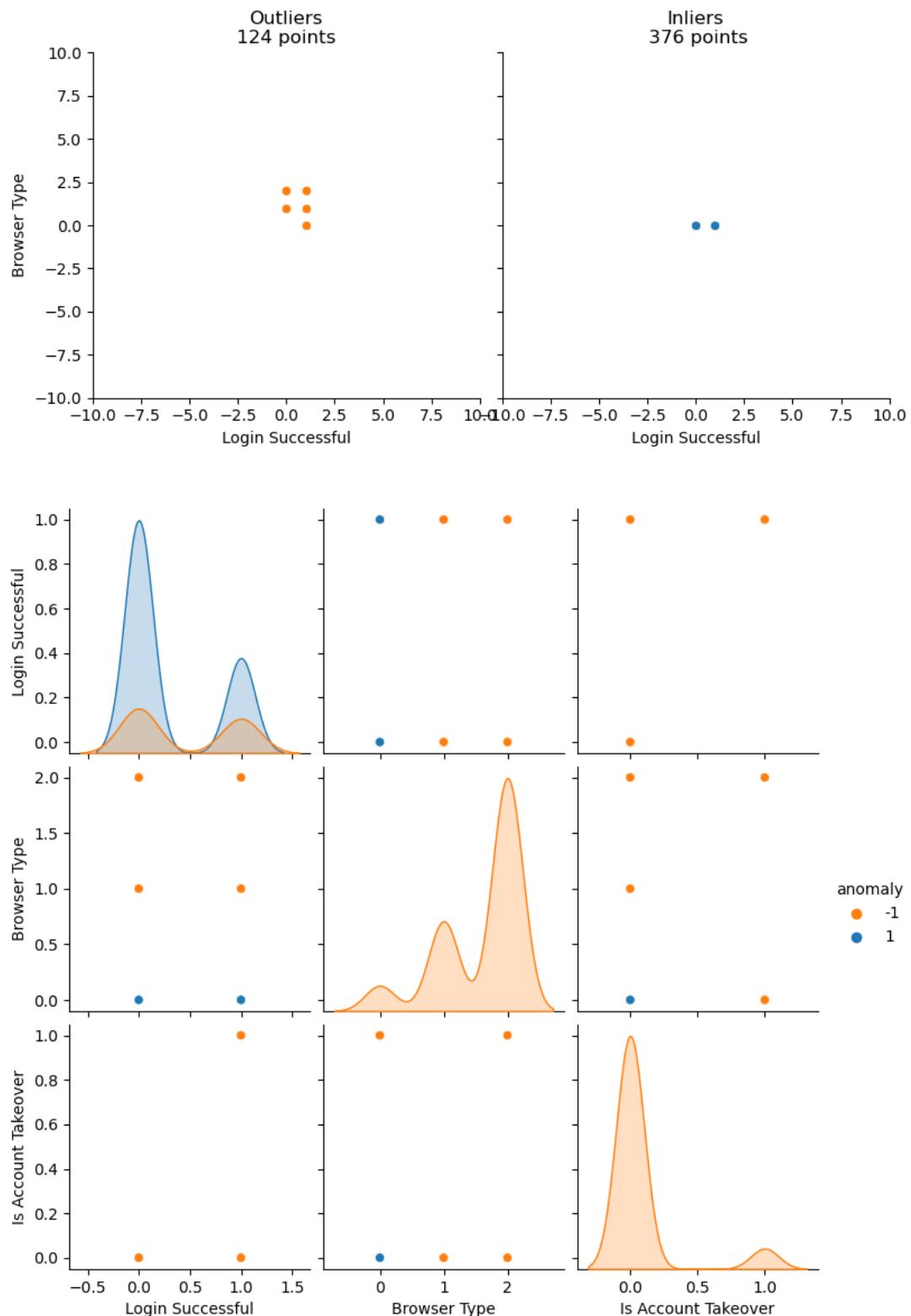
- `anomaly_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']`
 - contamination value == 0.1

Outlier Method: Isolation Forest



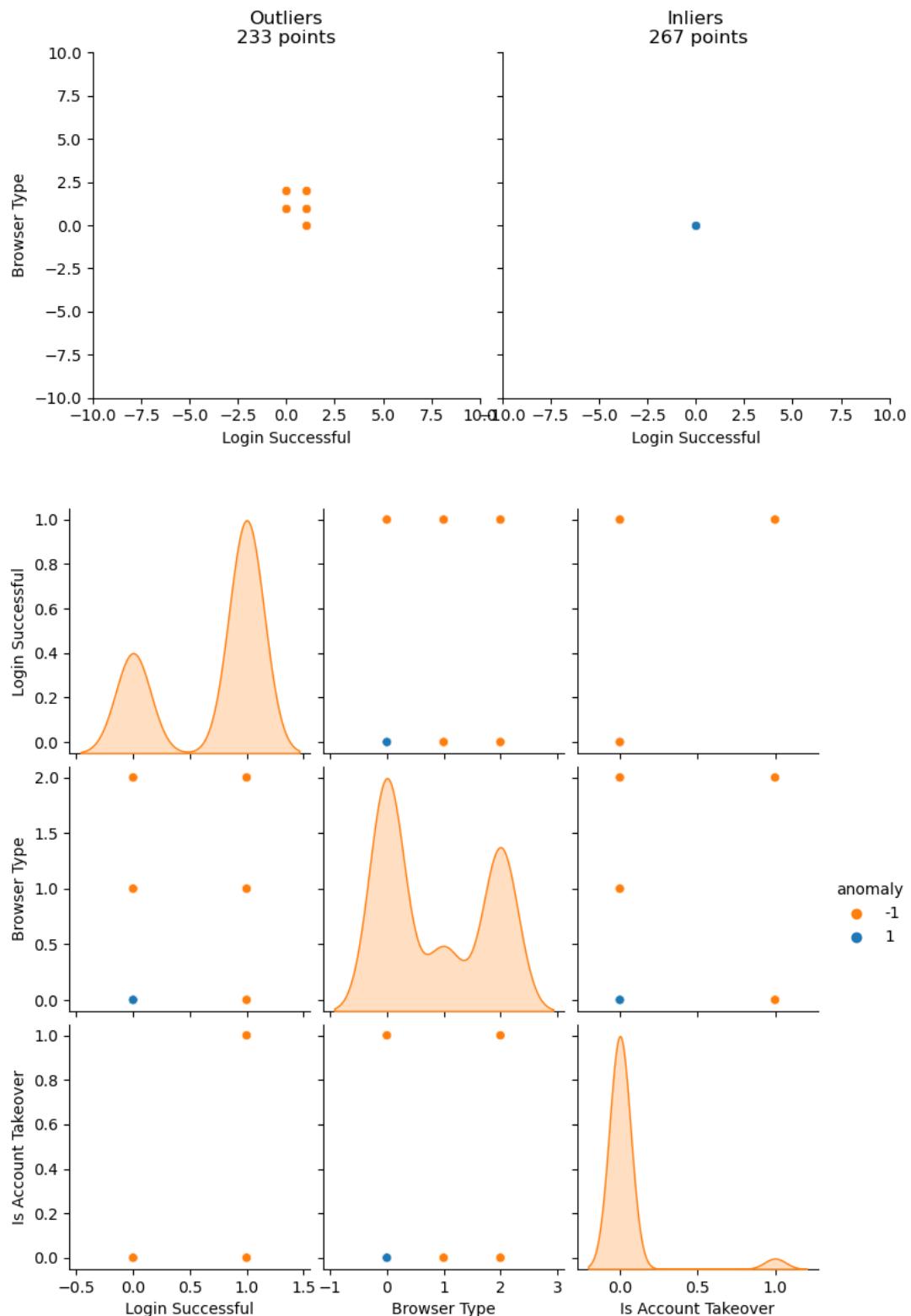
- contamination value == 0.3

Outlier Method: Isolation Forest



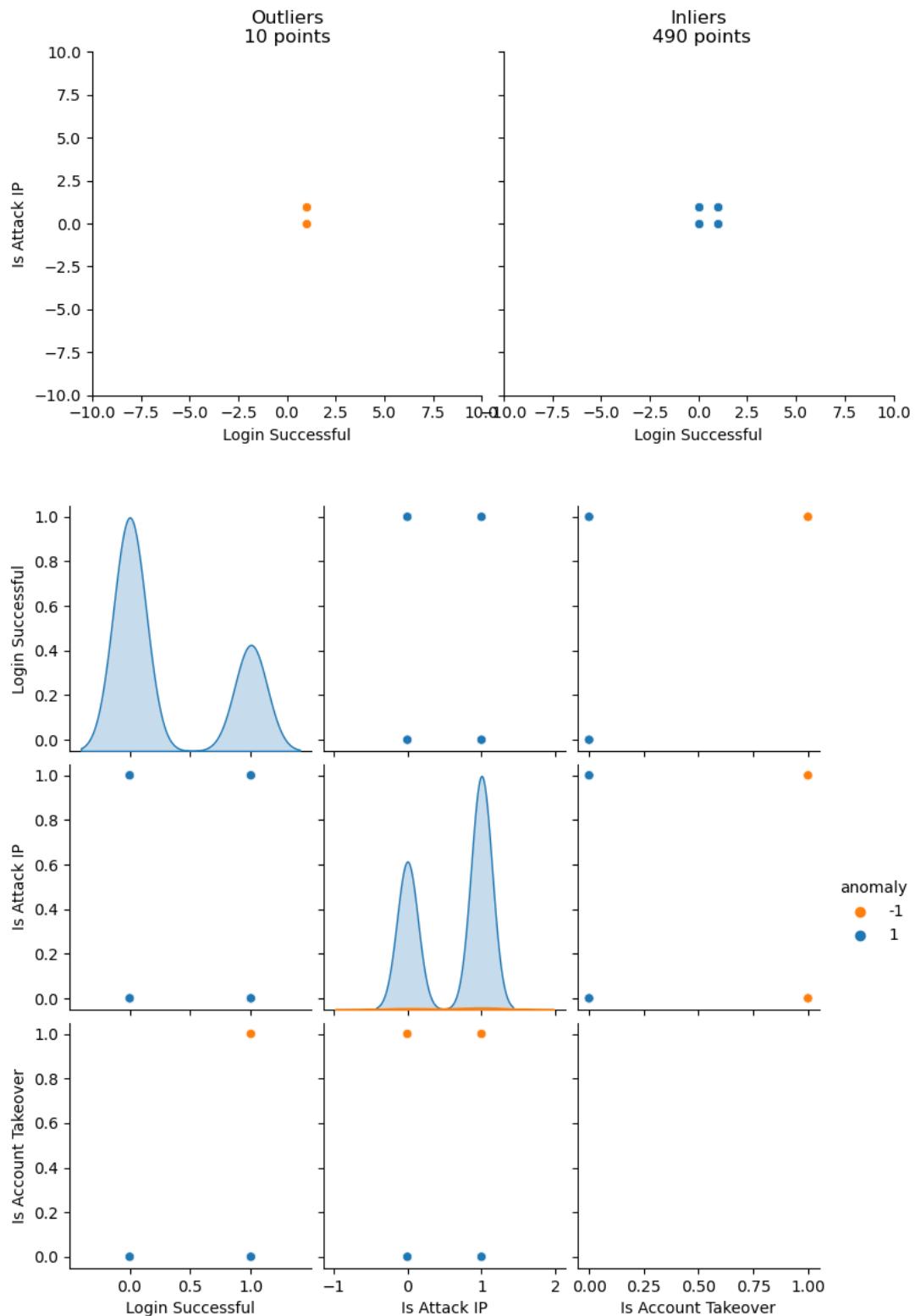
- contamination value == 0.5

Outlier Method: Isolation Forest



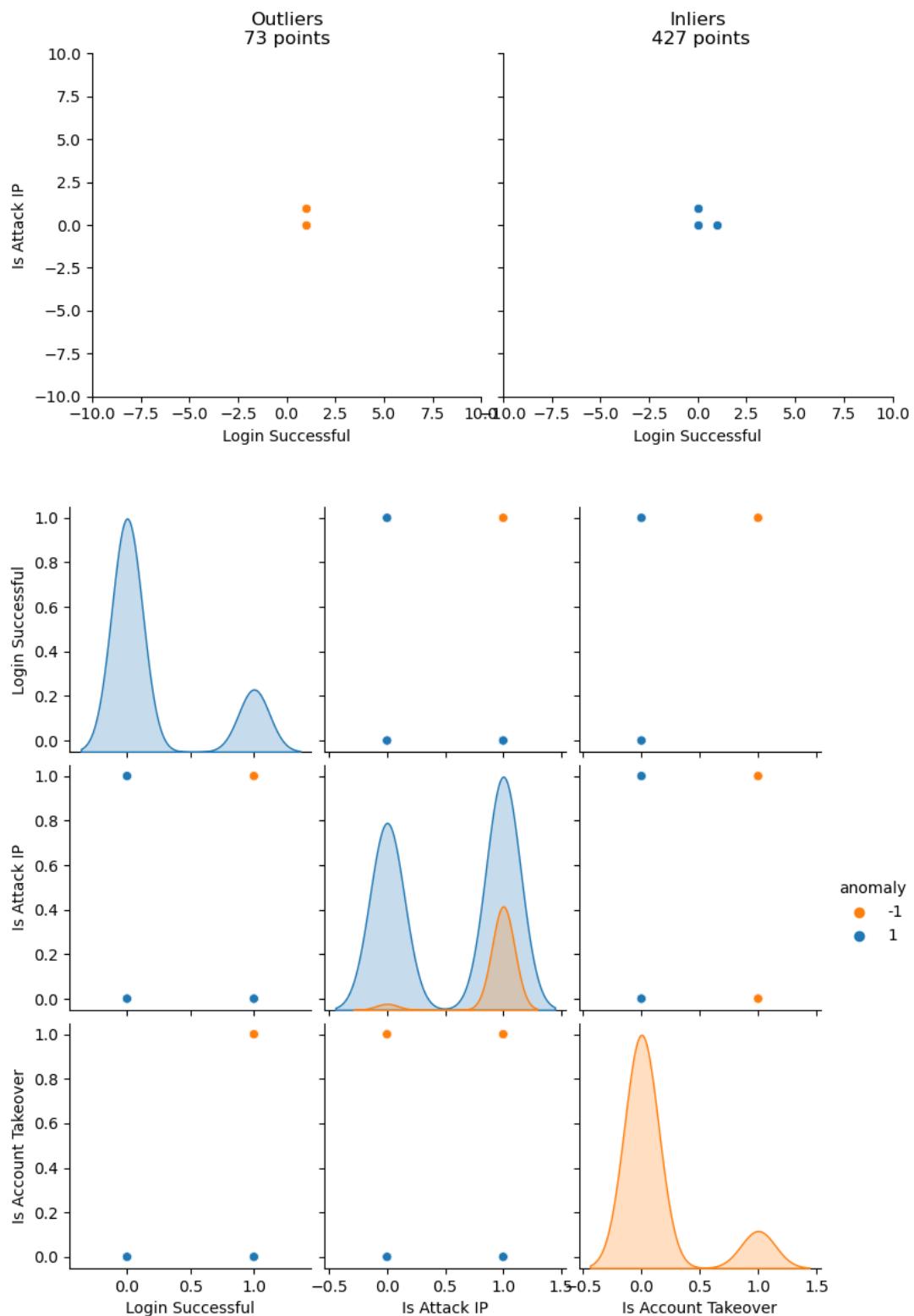
- `anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']`
 - contamination value == 0.1

Outlier Method: Isolation Forest



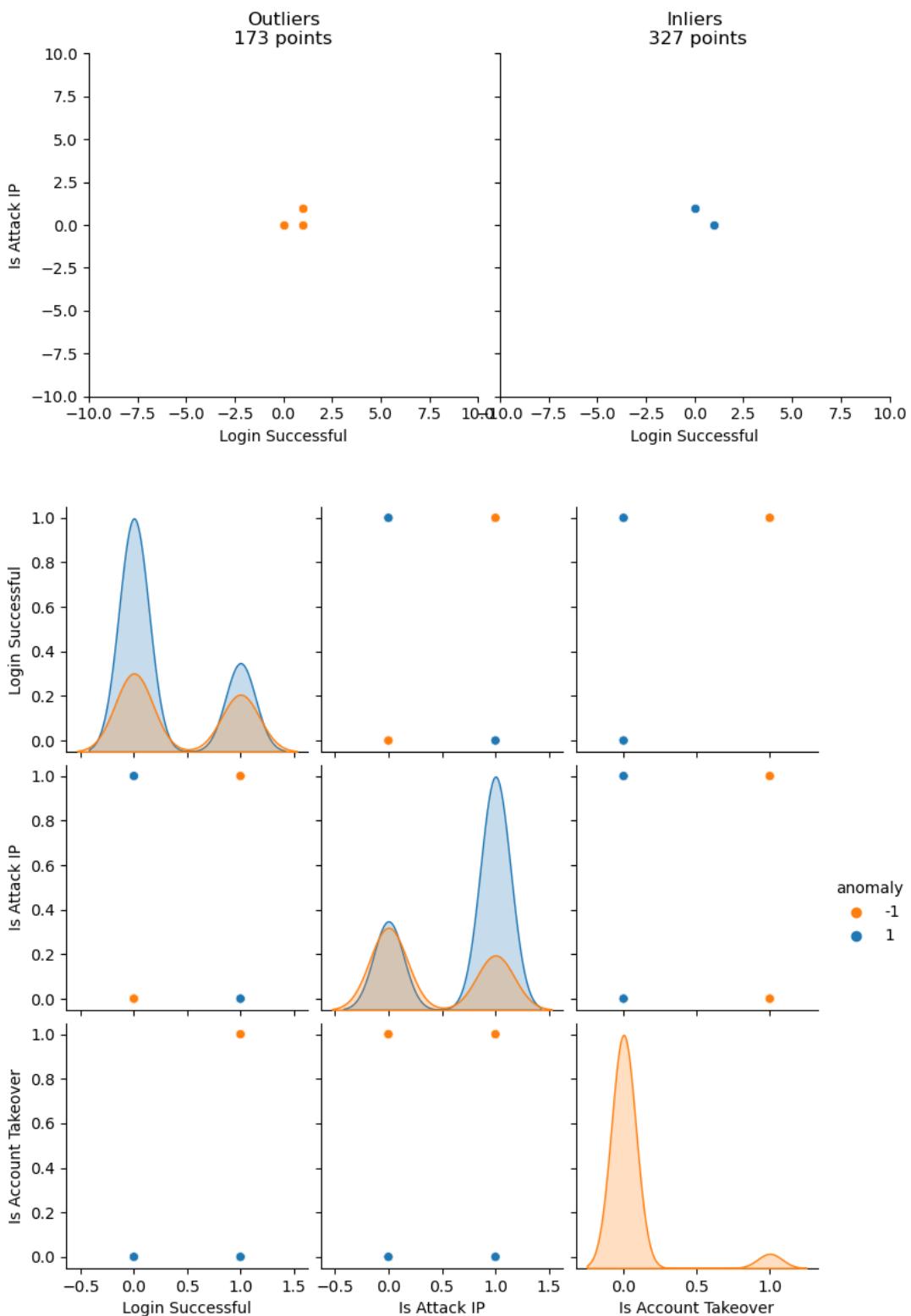
- contamination value == 0.3

Outlier Method: Isolation Forest

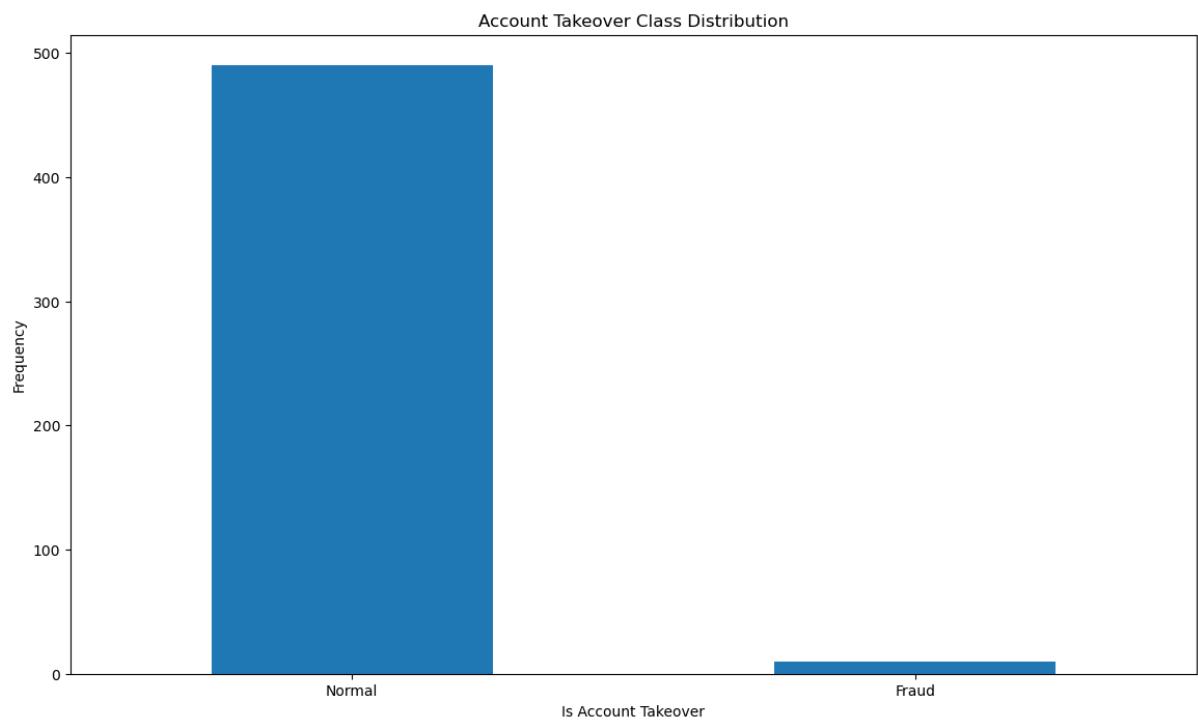


- contamination value == 0.5

Outlier Method: Isolation Forest



2D. iForest, Local Outlier Factor, Support Vector Machine



"{}: {}".format(clf_name,n_errors) → np. Isolation Forest: 21

Isolation Forest: 21

Accuracy Score :

0.958

Classification Report :

	precision	recall	f1-score	support
False	0.98	0.98	0.98	490
True	0.00	0.00	0.00	10
accuracy			0.96	500
macro avg	0.49	0.49	0.49	500
weighted avg	0.96	0.96	0.96	500

Local Outlier Factor: 19

Accuracy Score :

0.962

Classification Report :

	precision	recall	f1-score	support
False	0.98	0.98	0.98	490
True	0.09	0.10	0.10	10
accuracy			0.96	500
macro avg	0.54	0.54	0.54	500
weighted avg	0.96	0.96	0.96	500

Support Vector Machine: 322

Accuracy Score :

0.356

Classification Report :

	precision	recall	f1-score	support
False	0.95	0.36	0.52	490
True	0.00	0.10	0.01	10
accuracy			0.36	500
macro avg	0.48	0.23	0.26	500
weighted avg	0.93	0.36	0.51	500

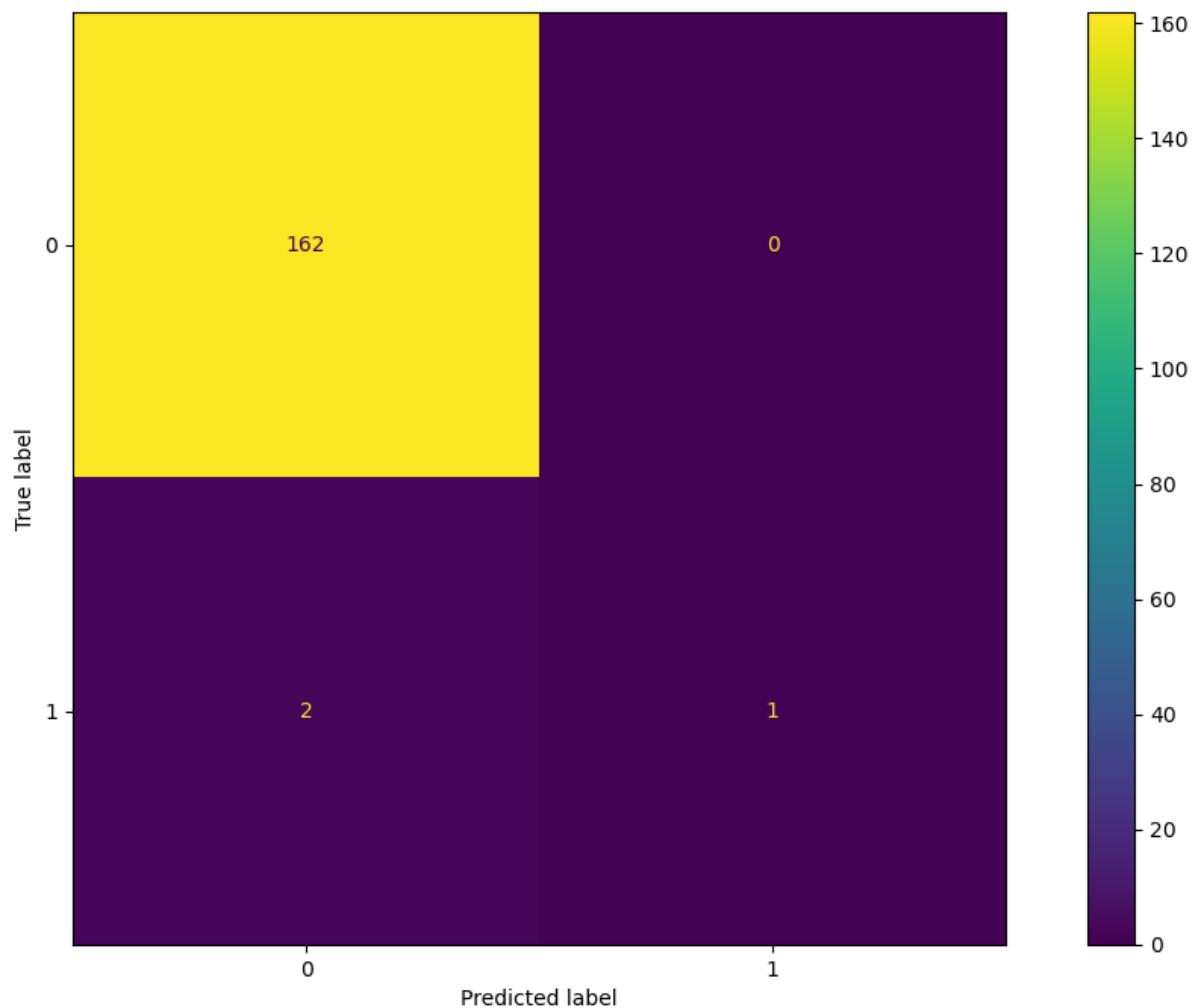
3. Dummy Variable Encoding → df3 =
`pd.get_dummies(df3, columns=['Country', 'Device Type', 'IP Address', 'Login Timestamp', 'Browser Type'], prefix=["cntr", "dvc", "ip", "lgnt", "brw"])`

3A. XGBoost

memory usage: 444.5 KB

Accuracy: 98.79%				
	precision	recall	f1-score	support
False	0.99	1.00	0.99	162
True	1.00	0.33	0.50	3
accuracy			0.99	165
macro avg	0.99	0.67	0.75	165
weighted avg	0.99	0.99	0.98	165

TN: 0, FP: 2, FN: 1, TP: 1



3B. One Class SVM

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.2 (same for any other col)

```

The number of records in the training dataset is 400
The number of records in the test dataset is 100
The training dataset has 392 records for the majority class and 8 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.02     0.04      98
      True         0.02      1.00     0.04       2

      accuracy          0.04      100
      macro avg       0.51      0.51     0.04      100
      weighted avg    0.98      0.04     0.04      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.98     1.00     0.99      98
      True         0.00      0.00     0.00       2

      accuracy          0.98      100
      macro avg       0.49      0.50     0.49      100
      weighted avg    0.96      0.98     0.97      100

[[ 2 96]
 [ 0  2]]

```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.3 (same for any other col)

```

The number of records in the training dataset is 350
The number of records in the test dataset is 150
The training dataset has 342 records for the majority class and 8 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.21     0.35     148
      True         0.02      1.00     0.03       2

      accuracy          0.22      150
      macro avg       0.51      0.60     0.19      150
      weighted avg    0.99      0.22     0.34      150

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.99     1.00     0.99     148
      True         0.00      0.00     0.00       2

      accuracy          0.99      150
      macro avg       0.49      0.50     0.50      150
      weighted avg    0.97      0.99     0.98      150

[[ 31 117]
 [ 0  2]]

```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.5 (same for any other col)

```

The number of records in the training dataset is 250
The number of records in the test dataset is 250
The training dataset has 243 records for the majority class and 7 records for the minority class.
      precision    recall   f1-score   support
False        0.00     0.00     0.00    247
True        0.01     1.00     0.02      3

accuracy                           0.01    250
macro avg                          0.01    250
weighted avg                       0.00    250

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.99     1.00     0.99    247
True        0.00     0.00     0.00      3

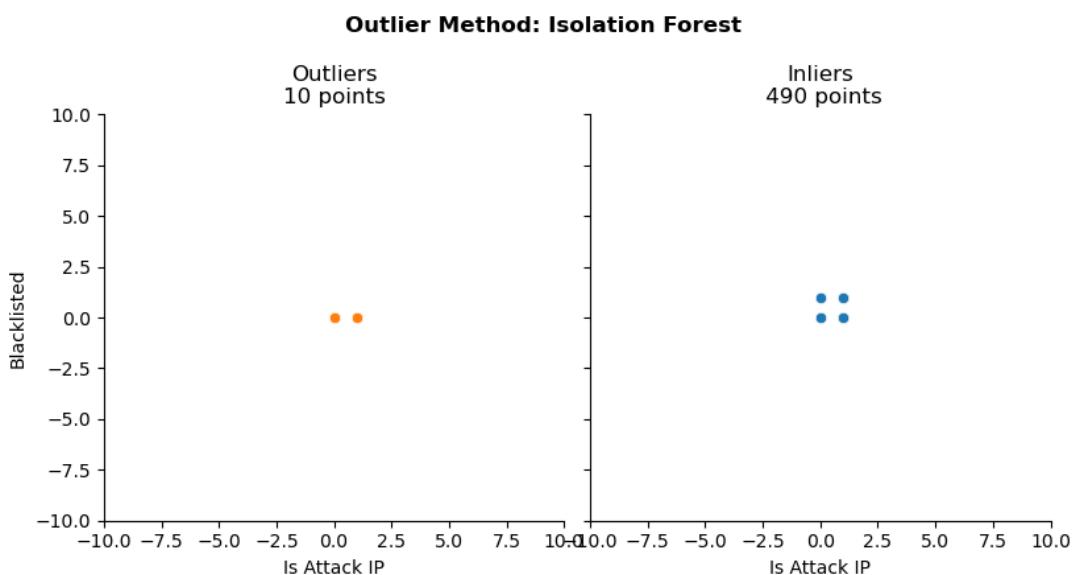
accuracy                           0.99    250
macro avg                          0.49    250
weighted avg                       0.98    250

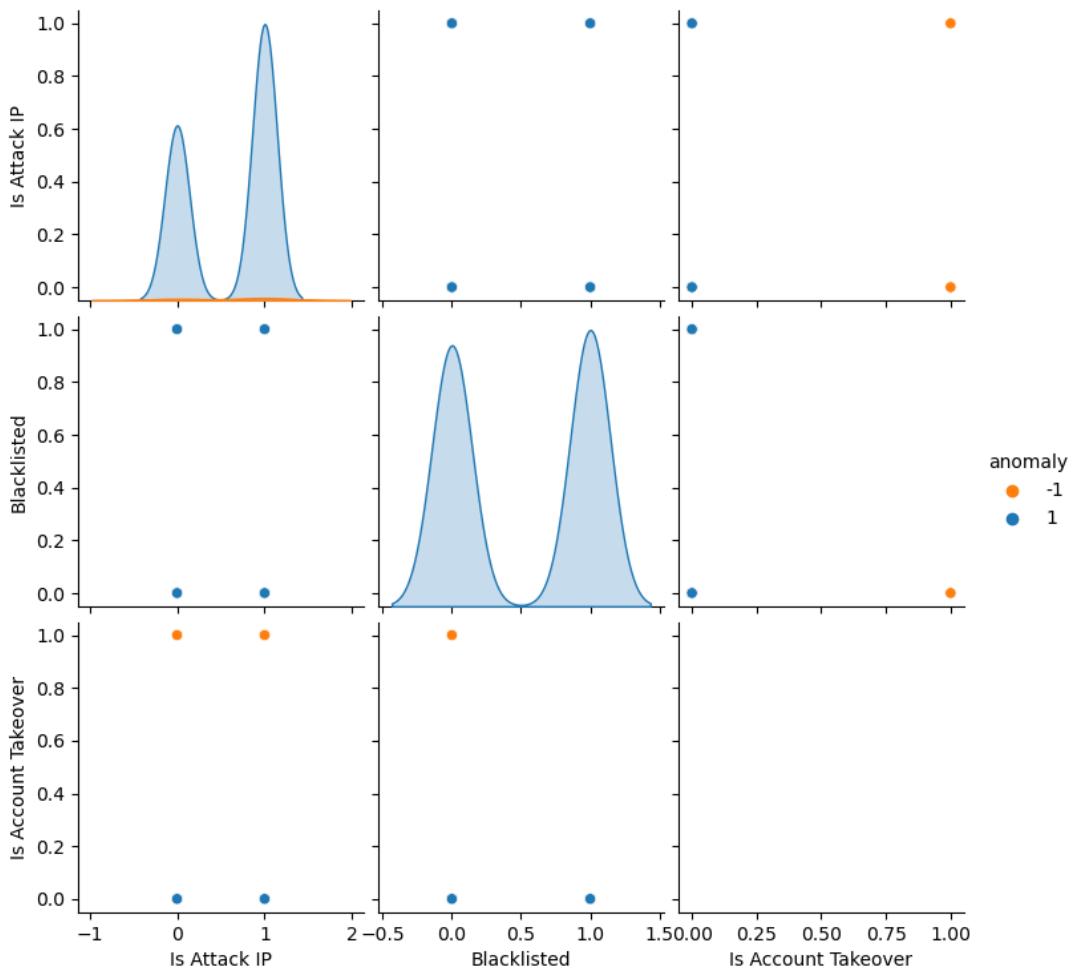
[[ 0 247]
 [ 0  3]]

```

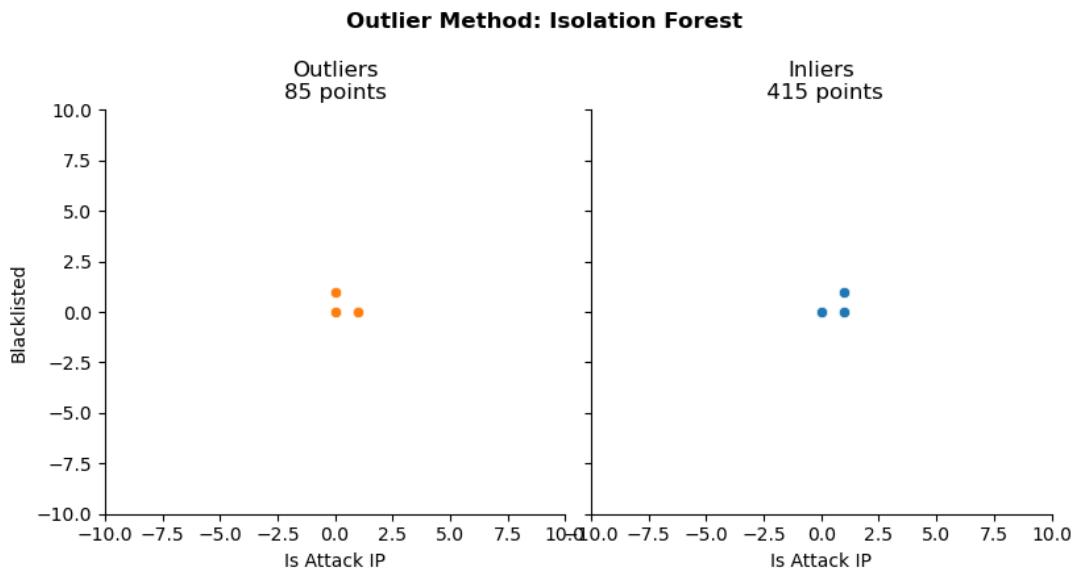
3C. Isolation Forest

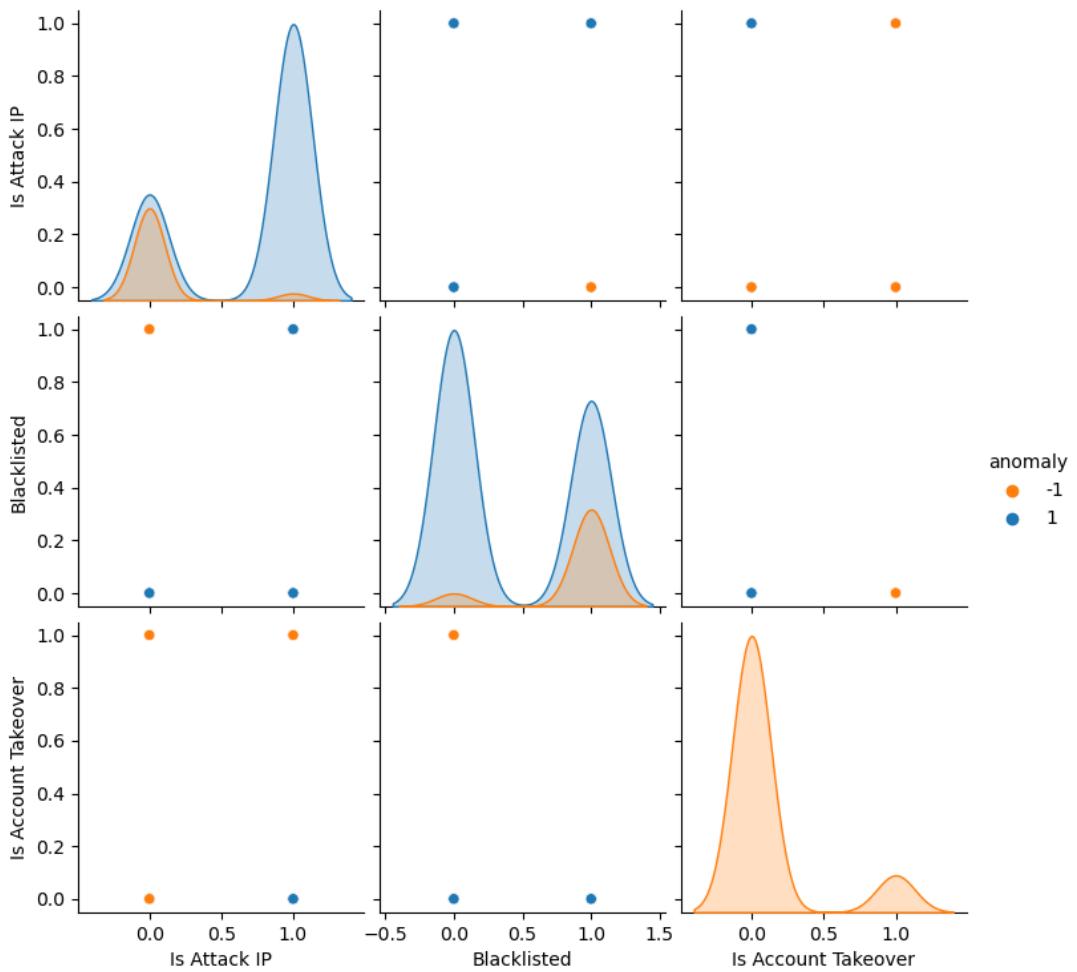
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



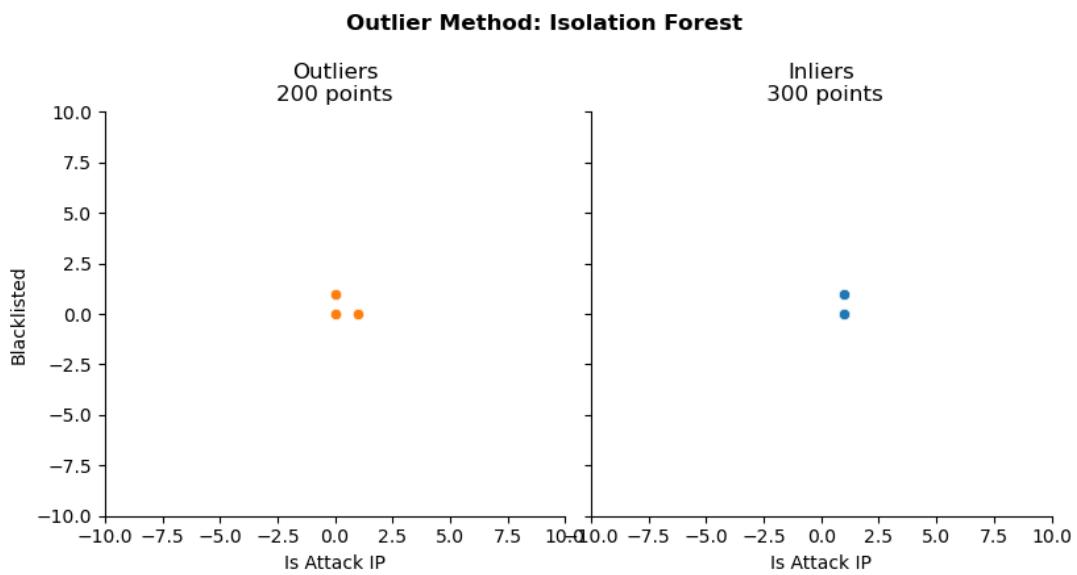


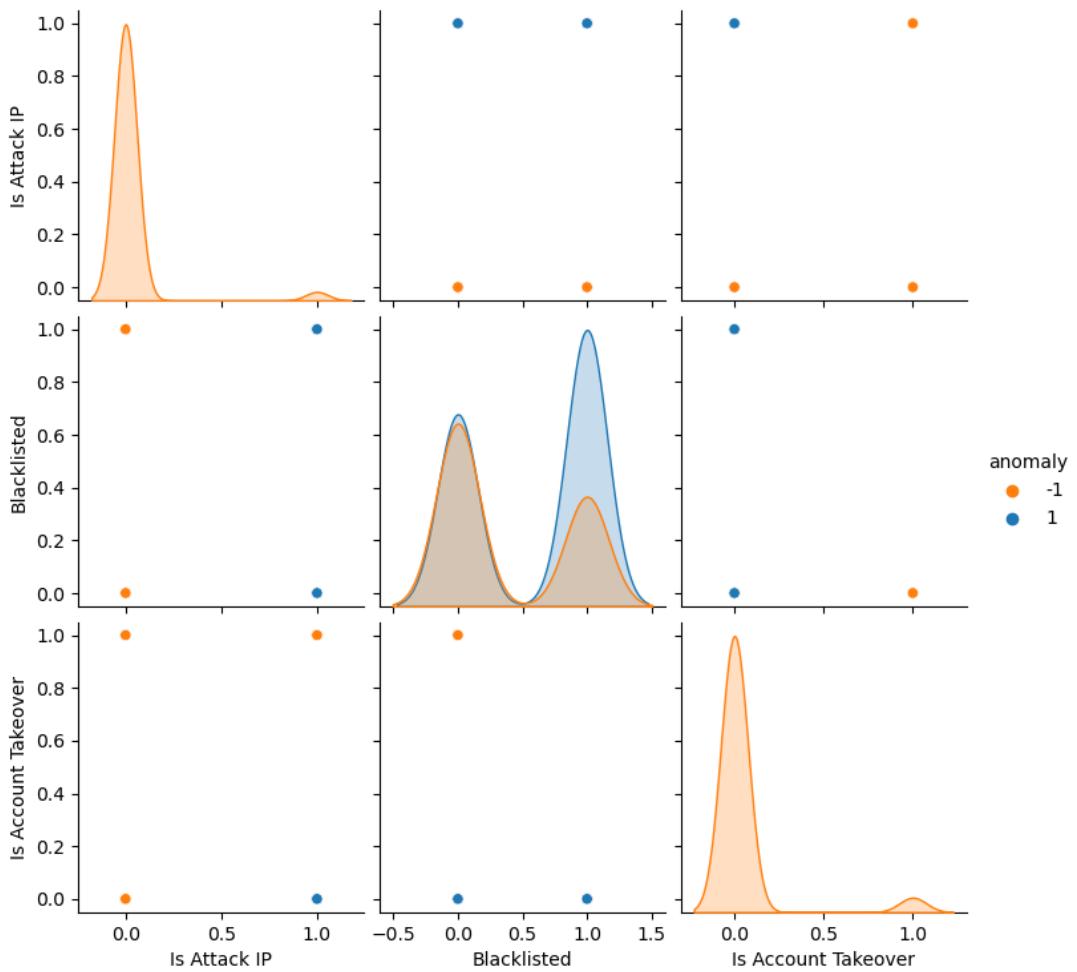
- contamination value == 0.3



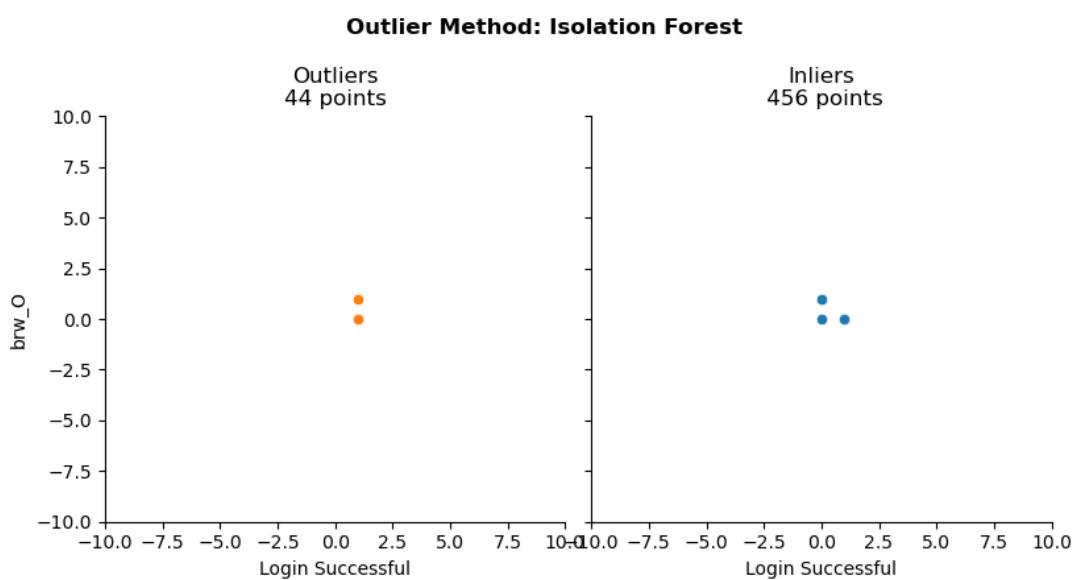


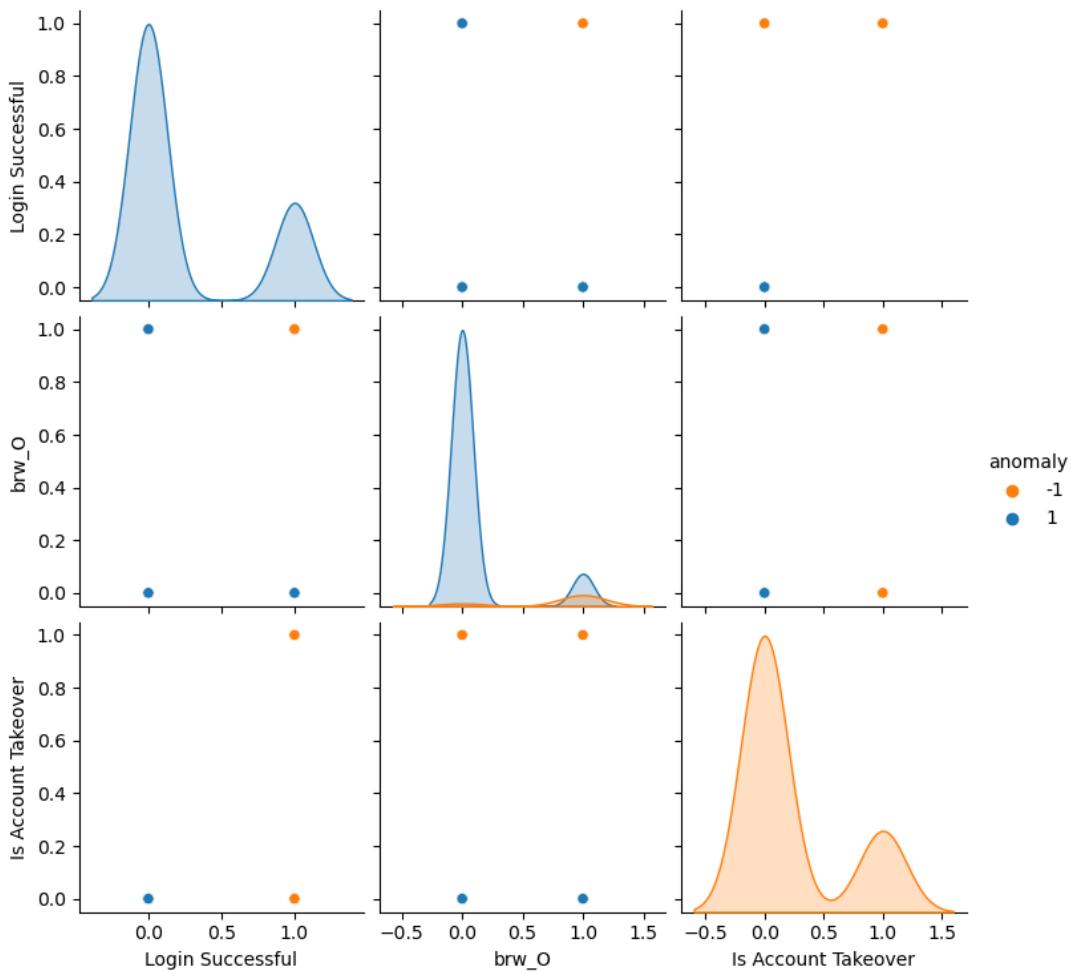
- contamination value == 0.5



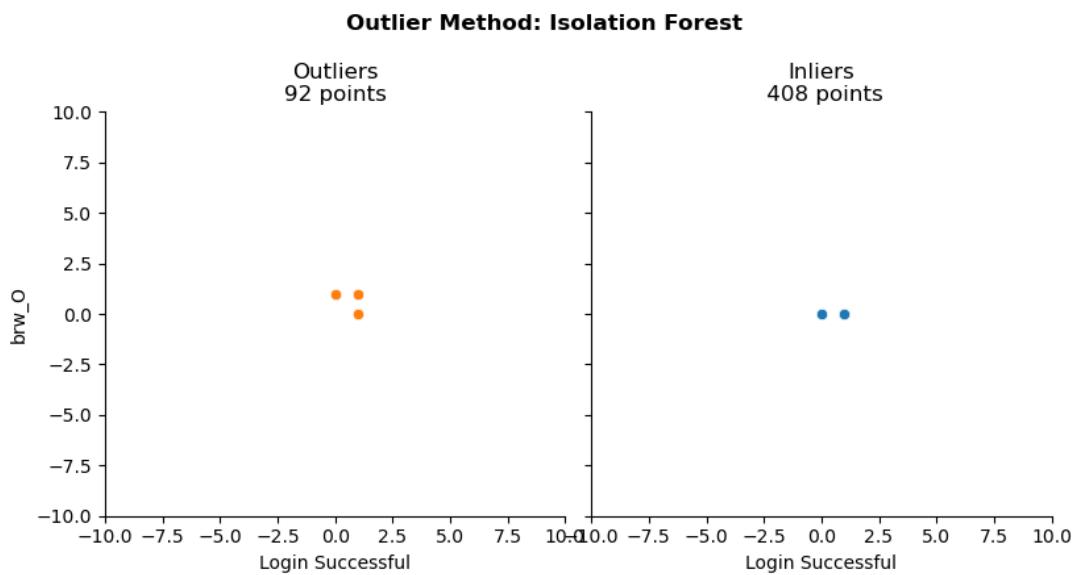


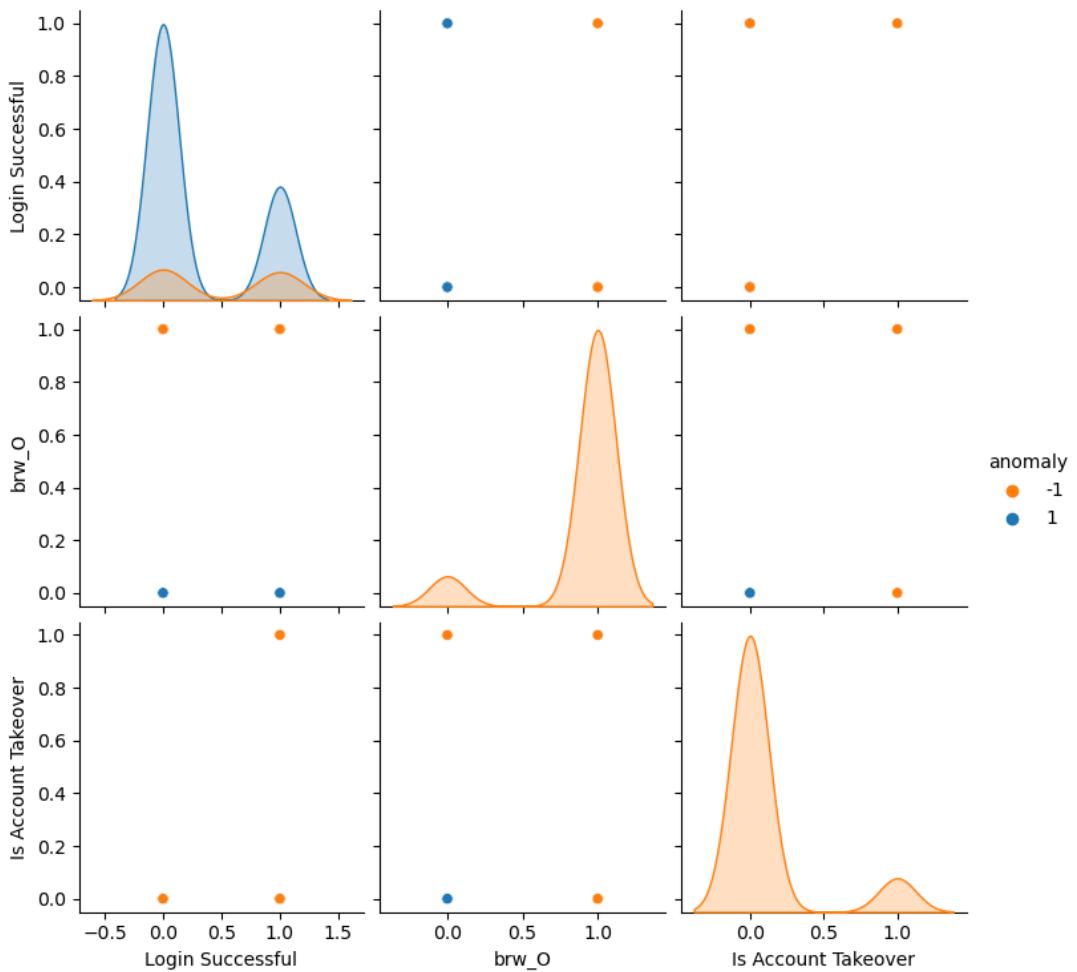
- anomaly_inputs = ['Login Successful', 'Browser Type (brw_O)', 'Is Account Takeover']
 - contamination value == 0.1



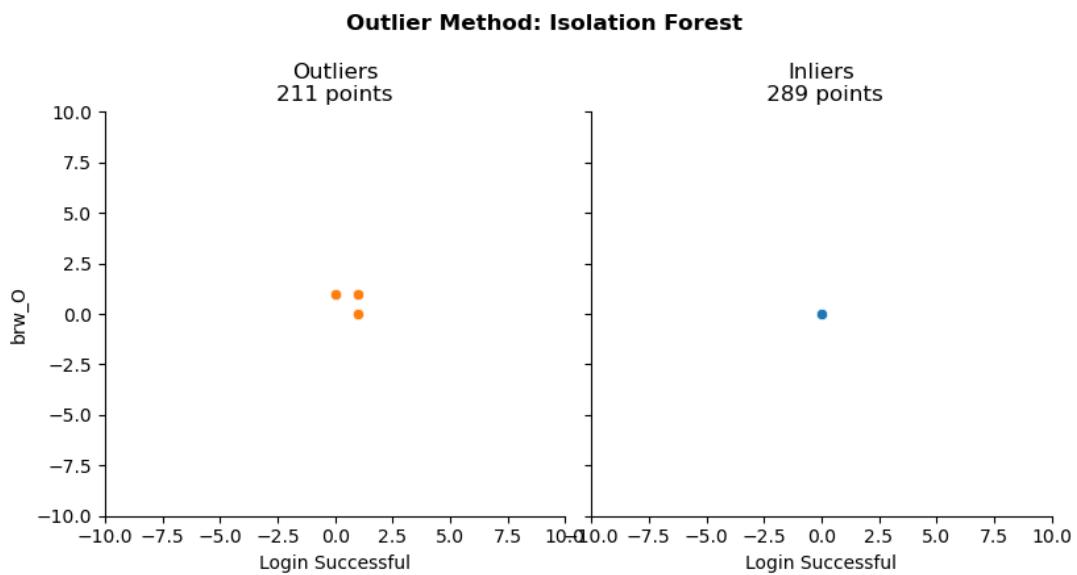


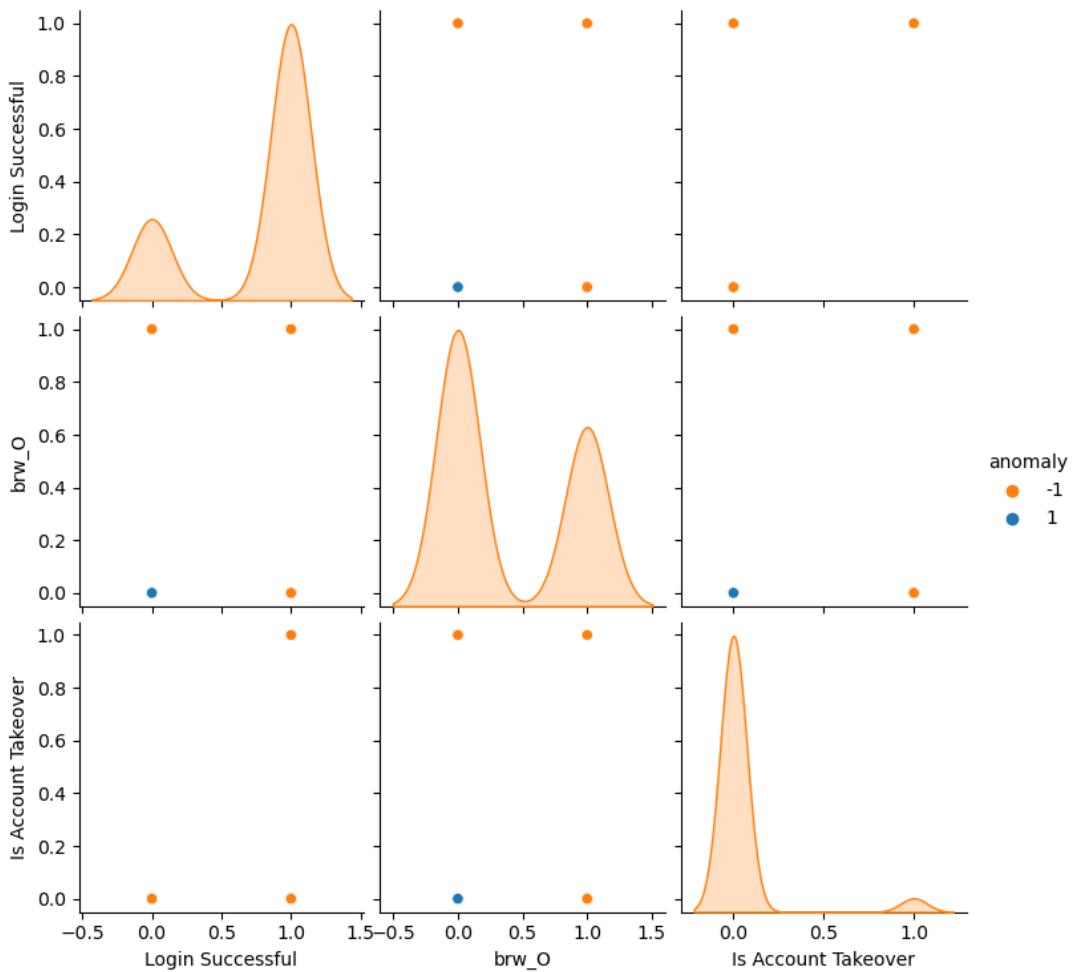
- contamination value == 0.3



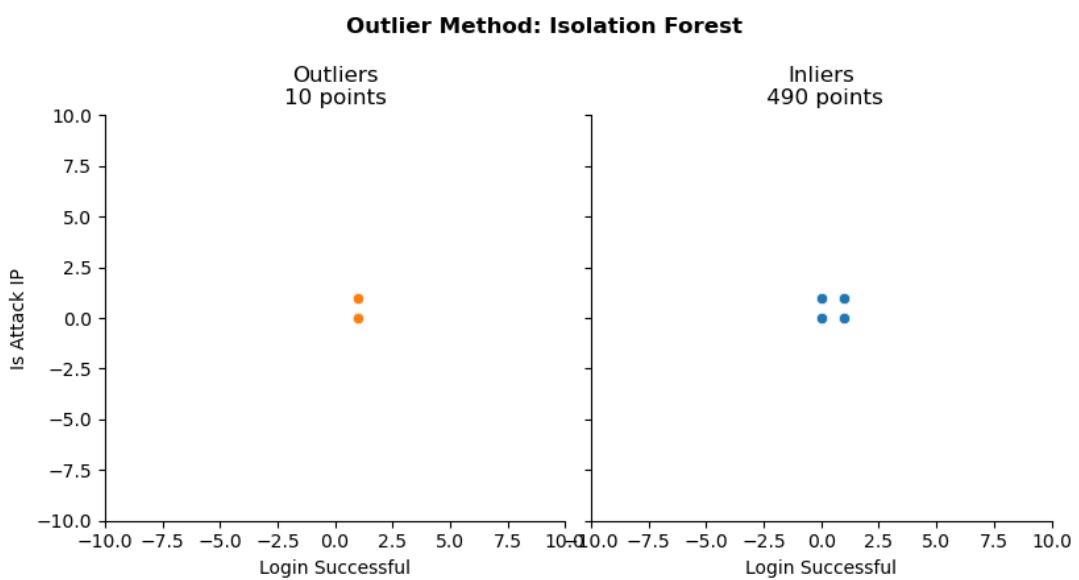


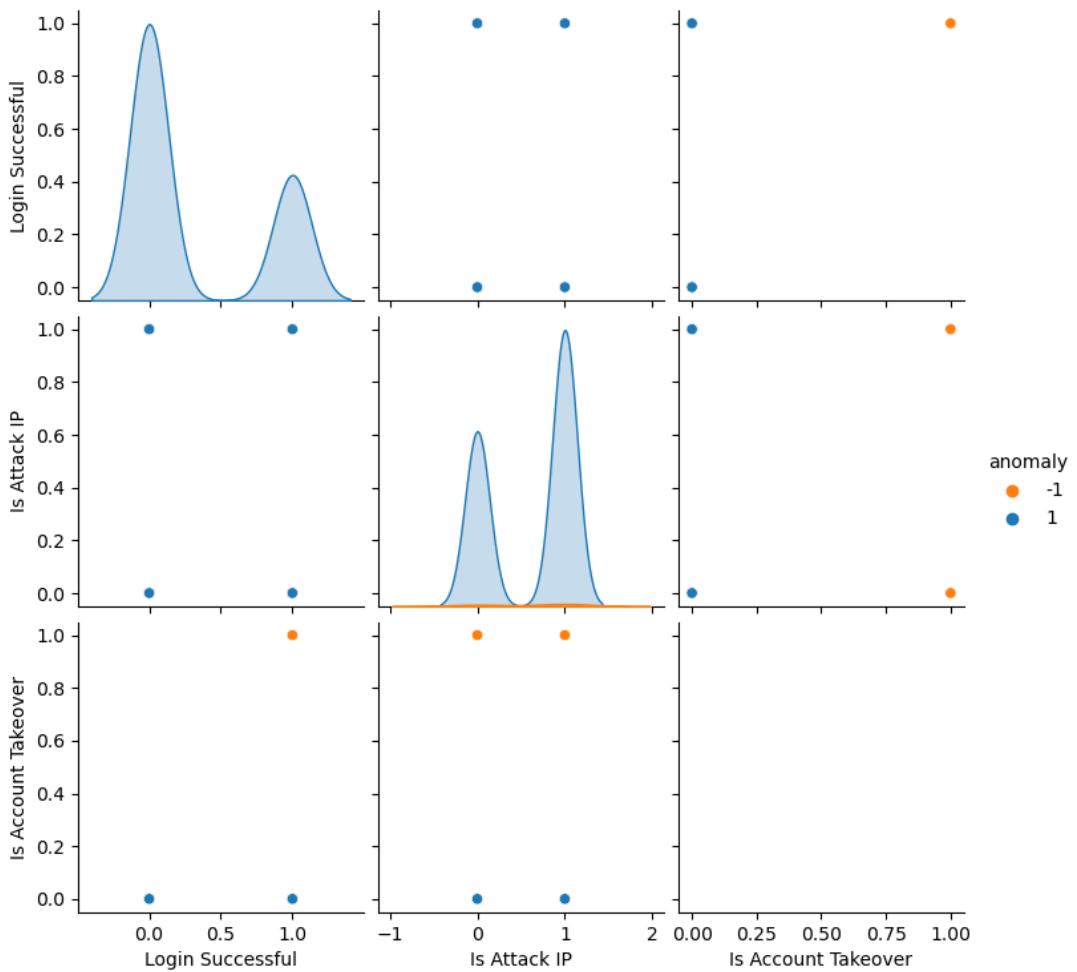
- contamination value == 0.5



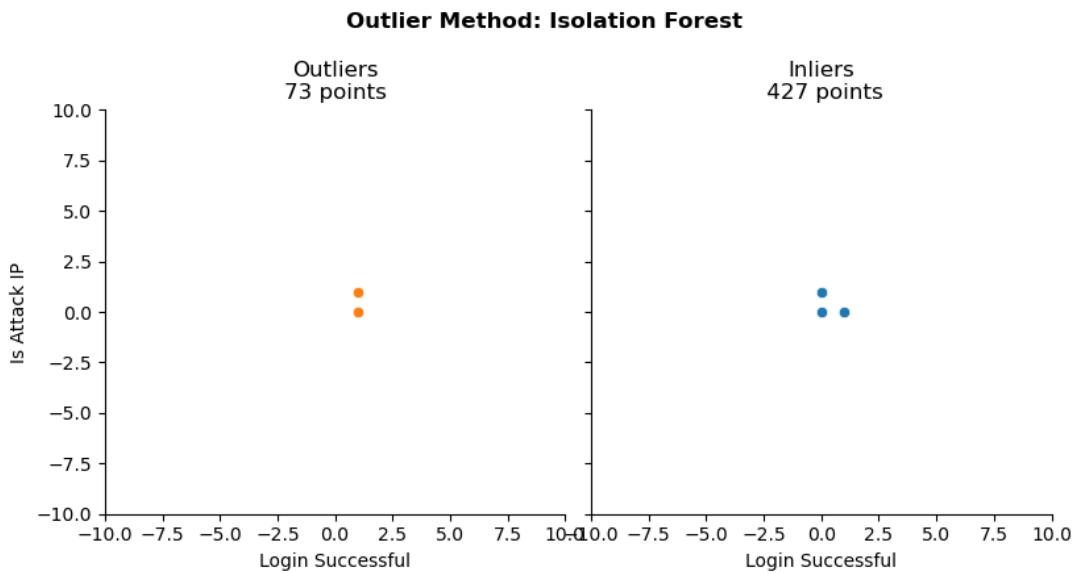


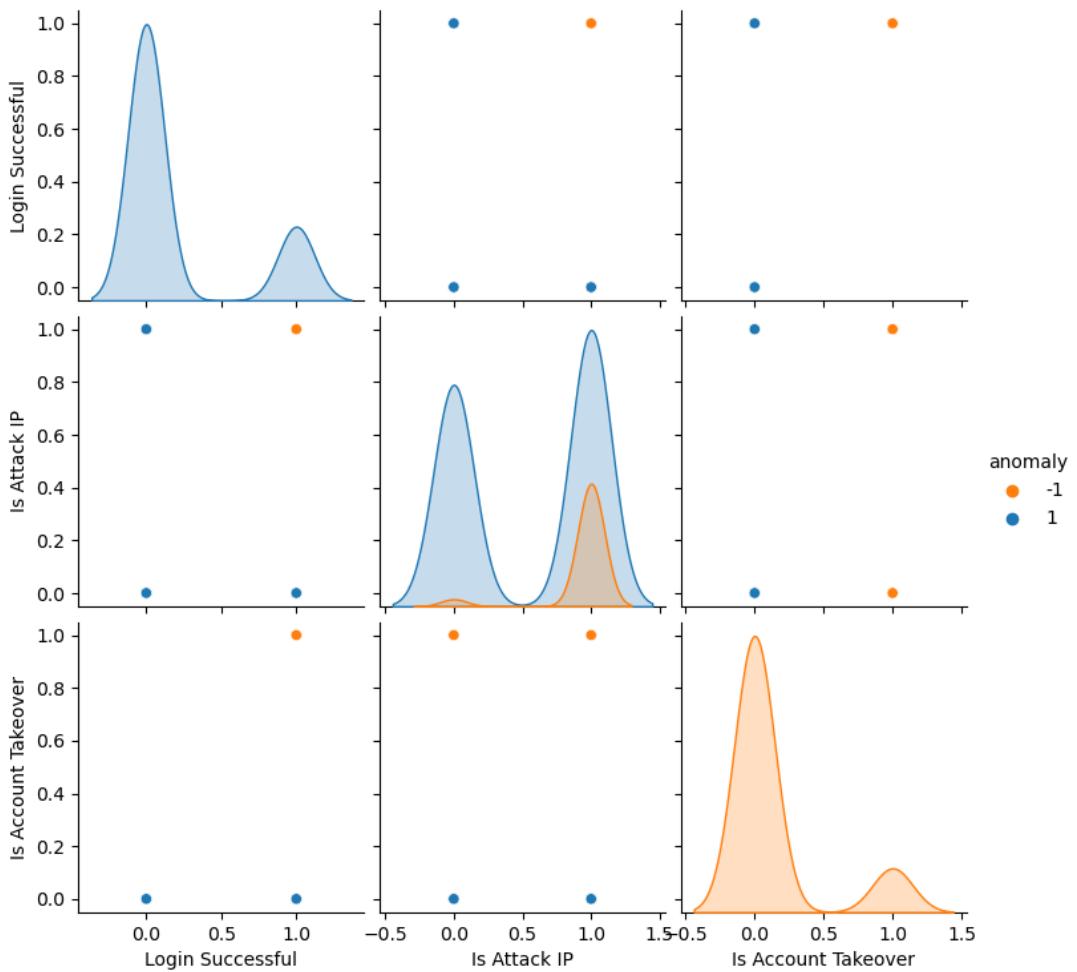
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1



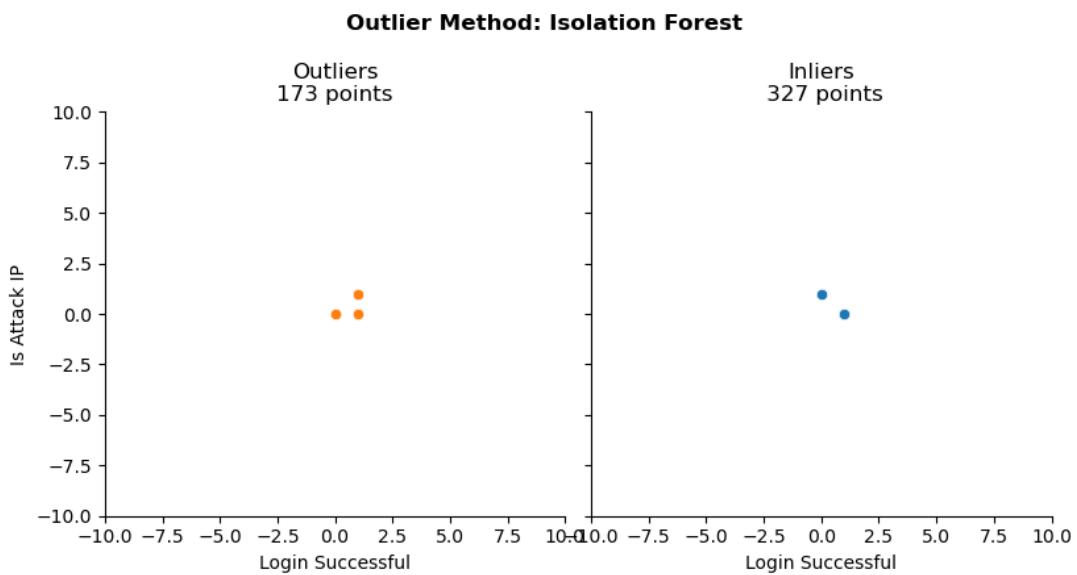


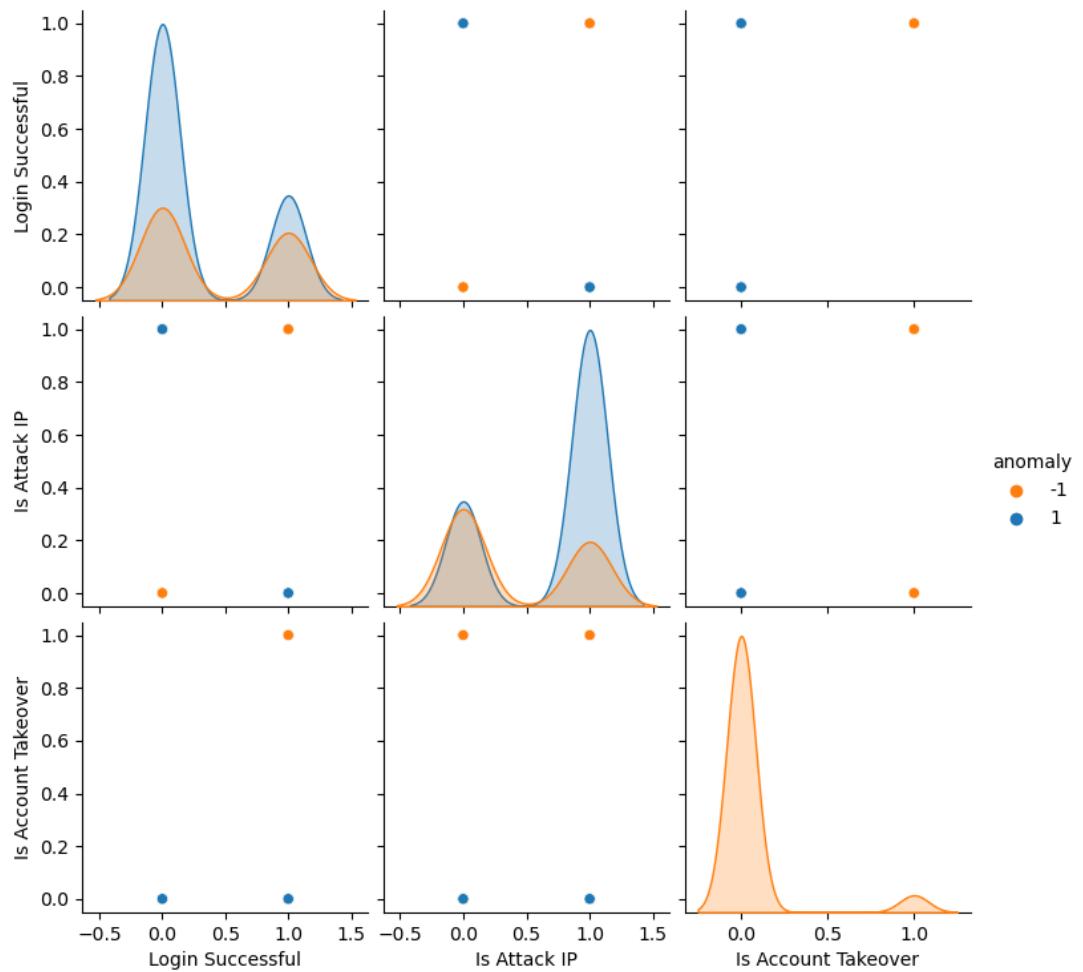
- contamination value == 0.3



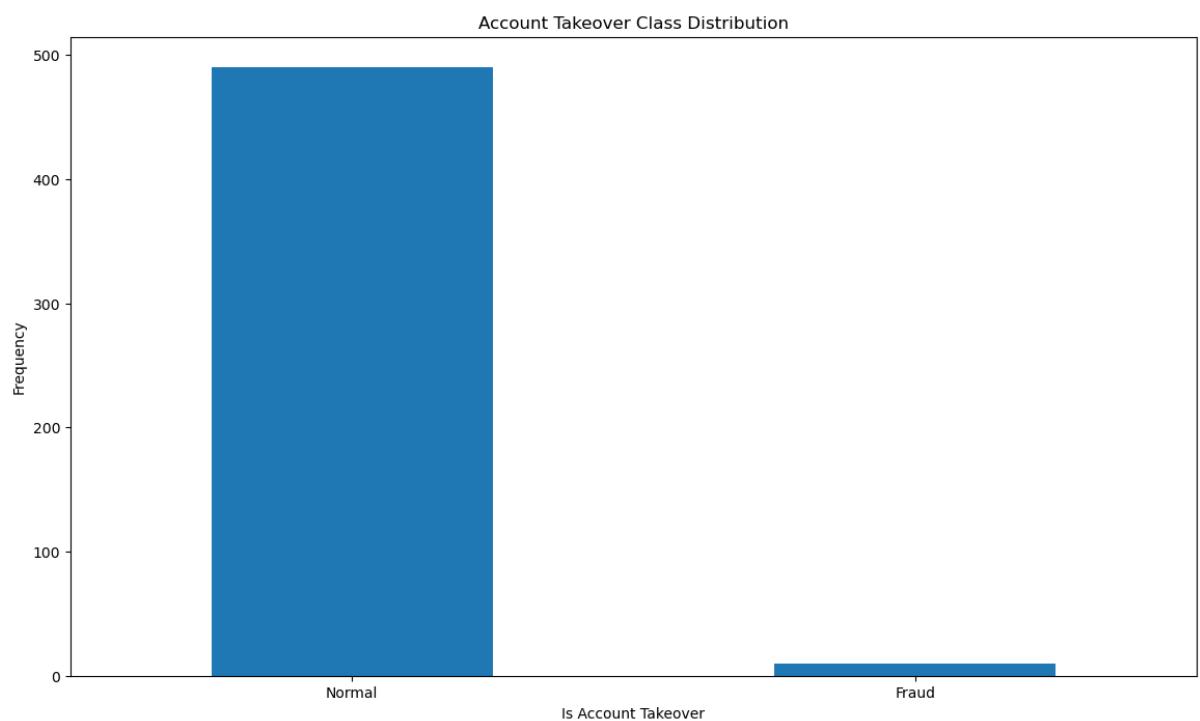


- contamination value == 0.5





3D. iForest, Local Outlier Factor, Support Vector Machine



Isolation Forest: 21

Accuracy Score :

0.958

Classification Report :

	precision	recall	f1-score	support
False	0.98	0.98	0.98	490
True	0.00	0.00	0.00	10
accuracy			0.96	500
macro avg	0.49	0.49	0.49	500
weighted avg	0.96	0.96	0.96	500

Local Outlier Factor: 21

Accuracy Score :

0.958

Classification Report :

	precision	recall	f1-score	support
False	0.98	0.98	0.98	490
True	0.00	0.00	0.00	10
accuracy			0.96	500
macro avg	0.49	0.49	0.49	500
weighted avg	0.96	0.96	0.96	500

Support Vector Machine: 322

Accuracy Score :

0.356

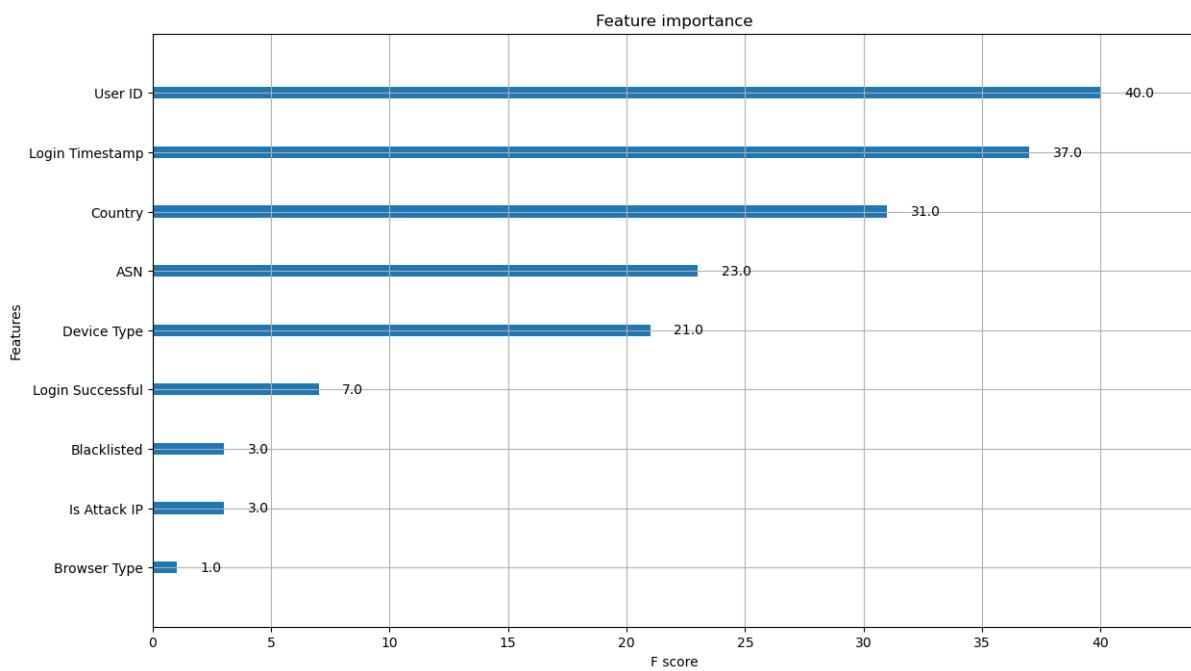
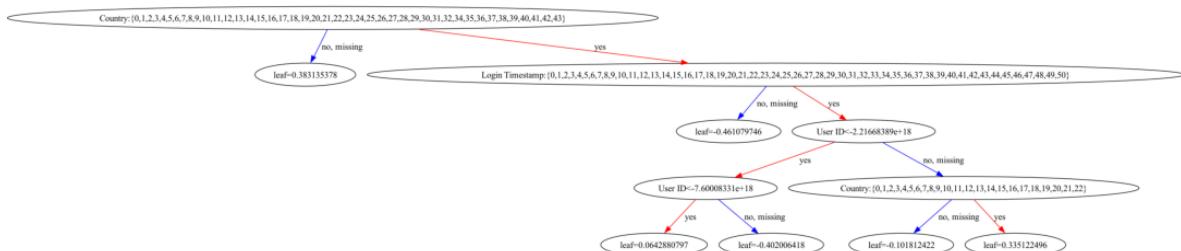
Classification Report :

	precision	recall	f1-score	support
False	0.95	0.36	0.52	490
True	0.00	0.10	0.01	10
accuracy			0.36	500
macro avg	0.48	0.23	0.26	500
weighted avg	0.93	0.36	0.51	500

II. 30/500 → 6%

1. XGB Experimental

```
-----
0 ASN                         500 non-null      float64
1 Country                      500 non-null      category
2 Device Type                  500 non-null      category
3 IP Address                   500 non-null      category
4 Is Attack IP                 500 non-null      bool
5 Login Successful              500 non-null      bool
6 Login Timestamp               500 non-null      category
7 User ID                      500 non-null      float64
8 Blacklisted                   500 non-null      bool
9 Browser Type                 500 non-null      category
10 Is Account Takeover         500 non-null      bool
dtypes: bool(4), category(5), float64(2)
memory usage: 45.9 KB
Feature importances:
[0.0278866  0.20840558 0.03784616 0.           0.0766107  0.1206836
 0.30358383 0.03665654 0.01446437 0.17386264]
```



2. Label encoding (cat.codes)

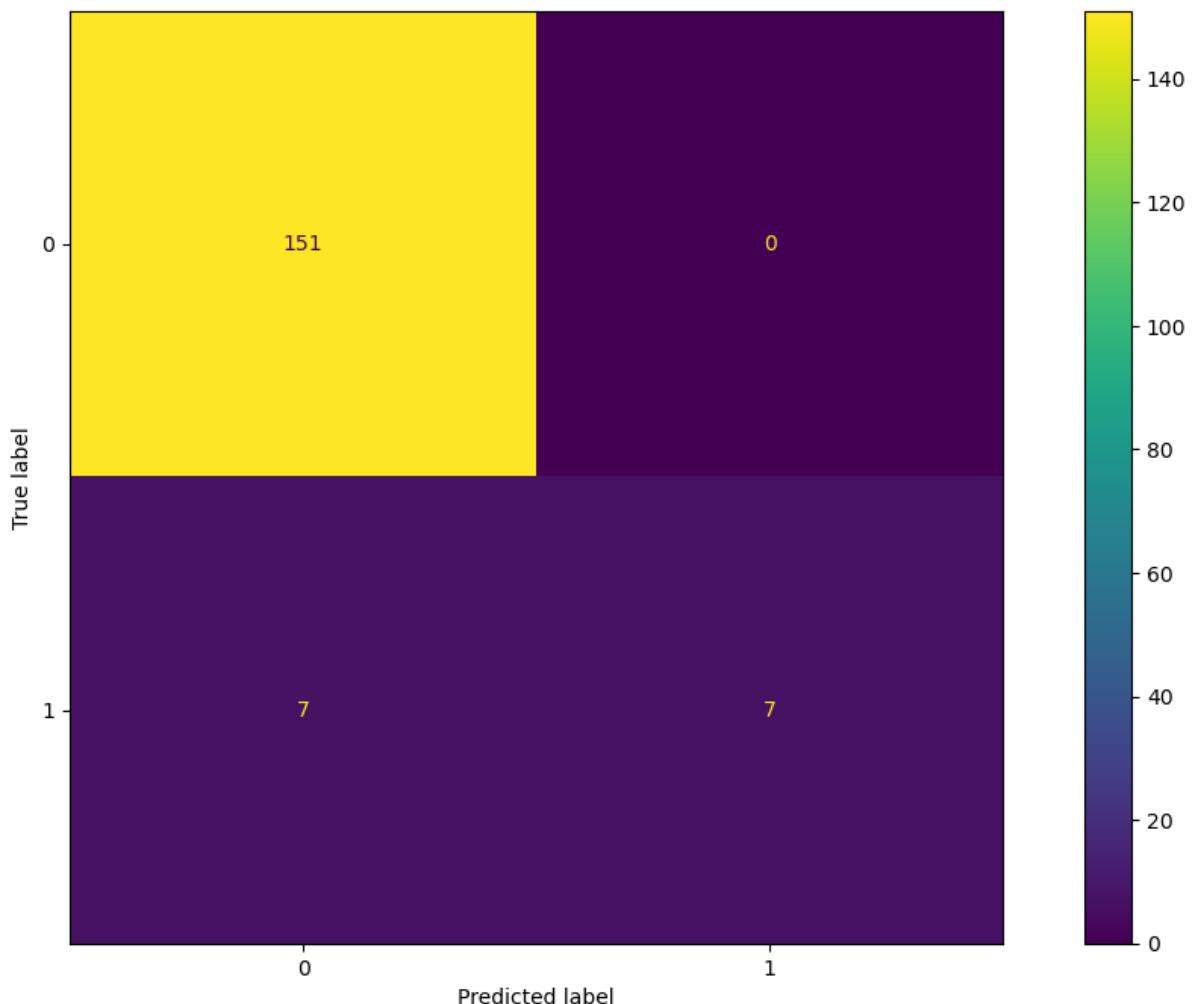
2A. XGBoost

memory usage: 13.3 KB

Accuracy: 95.76%

	precision	recall	f1-score	support
False	0.96	1.00	0.98	151
True	1.00	0.50	0.67	14
accuracy			0.96	165
macro avg	0.98	0.75	0.82	165
weighted avg	0.96	0.96	0.95	165

TN: 0, FP: 2, FN: 1, TP: 1



2B. ONE CLASS SVM

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.2 (same for any other col)

```
The number of records in the training dataset is 400
The number of records in the test dataset is 100
The training dataset has 376 records for the majority class and 24 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00      94
  True        0.06     1.00     0.11       6

  accuracy          0.06      100
  macro avg       0.03     0.50     0.06      100
  weighted avg    0.00     0.06     0.01      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.94     1.00     0.97      94
  True        0.00     0.00     0.00       6

  accuracy          0.94      100
  macro avg       0.47     0.50     0.48      100
  weighted avg    0.88     0.94     0.91      100

[[ 0 94]
 [ 0  6]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.3 (same for any other col)

```
The number of records in the training dataset is 350
The number of records in the test dataset is 150
The training dataset has 330 records for the majority class and 20 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00     140
  True        0.07     1.00     0.12      10

  accuracy          0.07      150
  macro avg       0.03     0.50     0.06     150
  weighted avg    0.00     0.07     0.01     150

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.93     1.00     0.97     140
  True        0.00     0.00     0.00      10

  accuracy          0.93      150
  macro avg       0.47     0.50     0.48     150
  weighted avg    0.87     0.93     0.90     150

[[ 0 140]
 [ 0  10]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.5 (same for any other col)

```

The number of records in the training dataset is 250
The number of records in the test dataset is 250
The training dataset has 235 records for the majority class and 15 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00     235
  True        0.06     1.00     0.11      15

  accuracy          0.06     250
 macro avg       0.03     0.50     0.06     250
weighted avg     0.00     0.06     0.01     250

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.94     1.00     0.97     235
  True        0.00     0.00     0.00      15

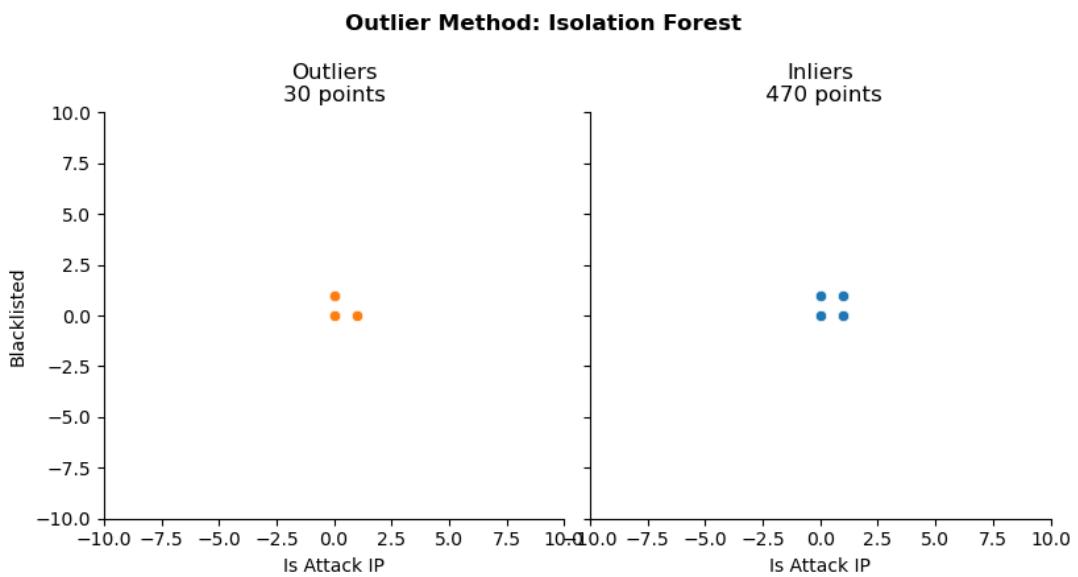
  accuracy          0.94     250
 macro avg       0.47     0.50     0.48     250
weighted avg     0.88     0.94     0.91     250

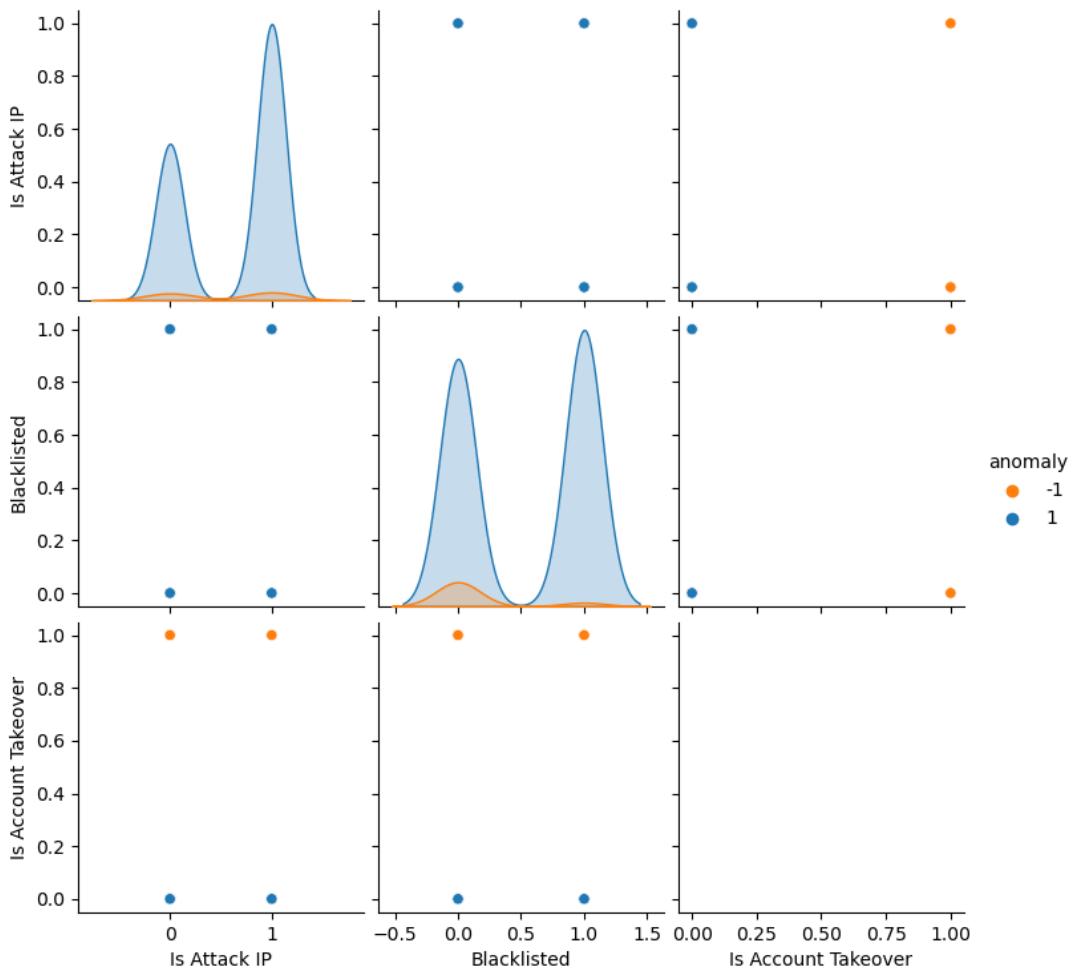
[[ 0 235]
 [ 0 15]]

```

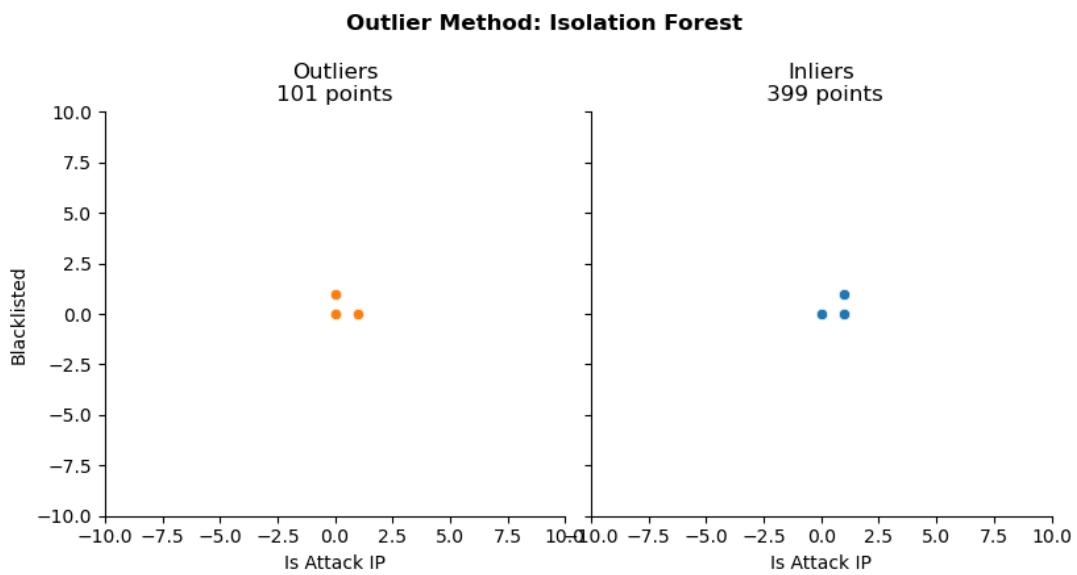
2C. Isolation Forest

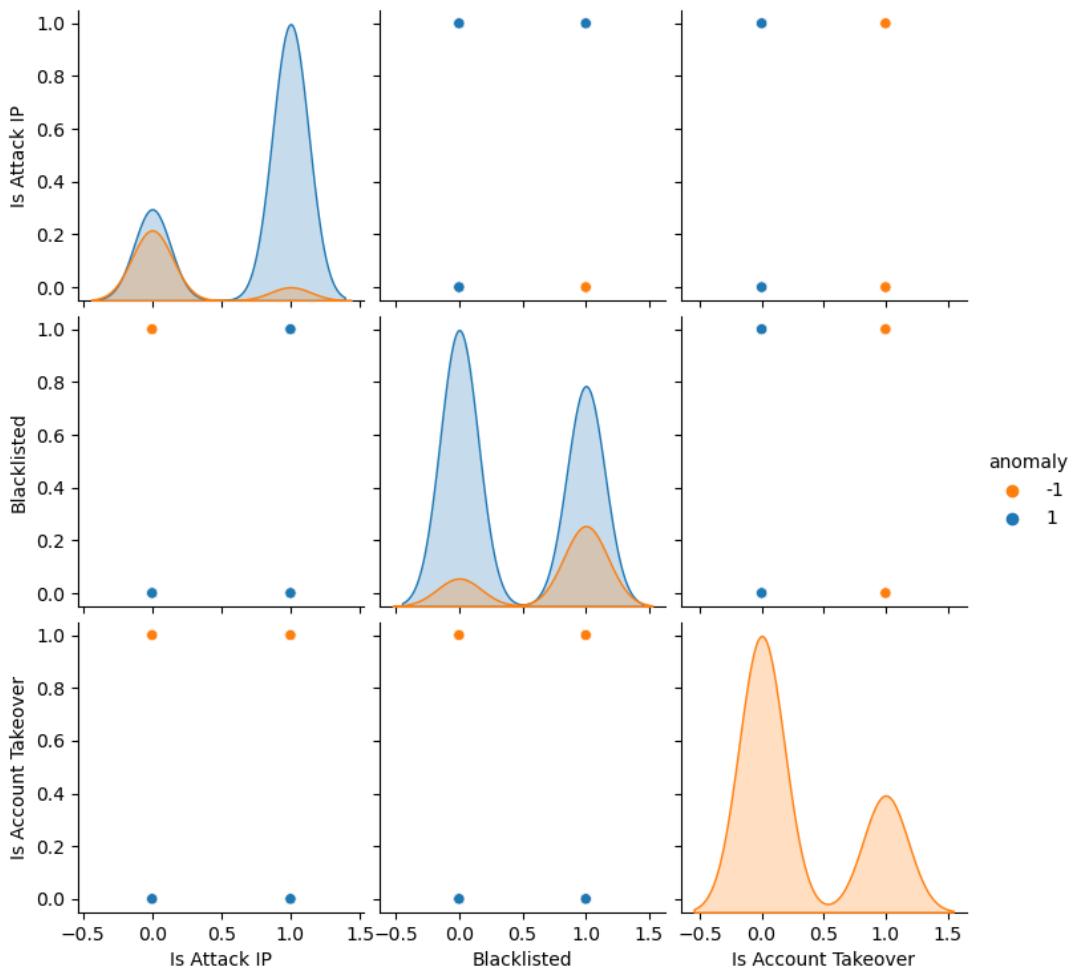
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



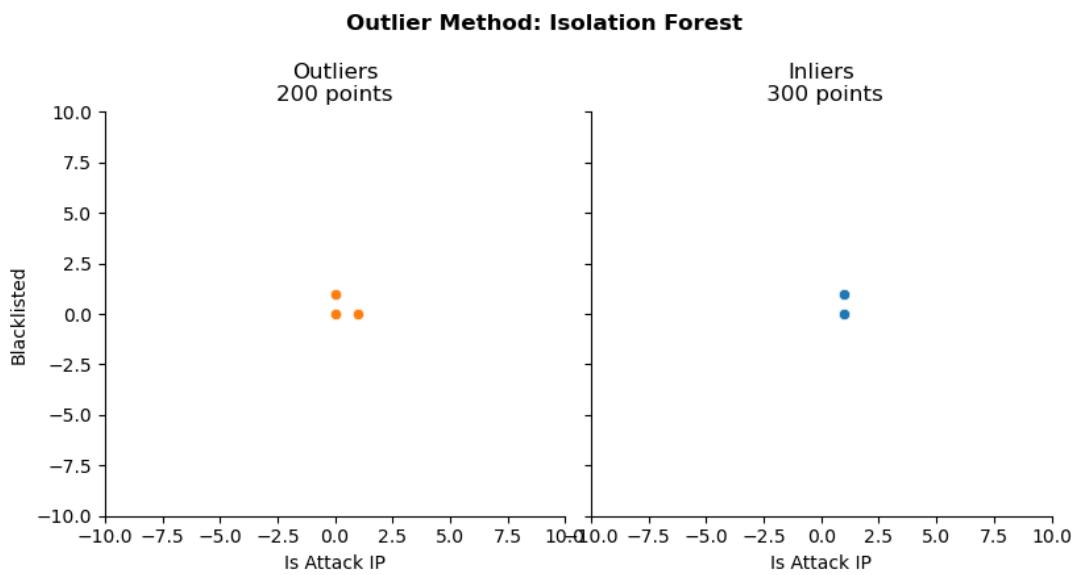


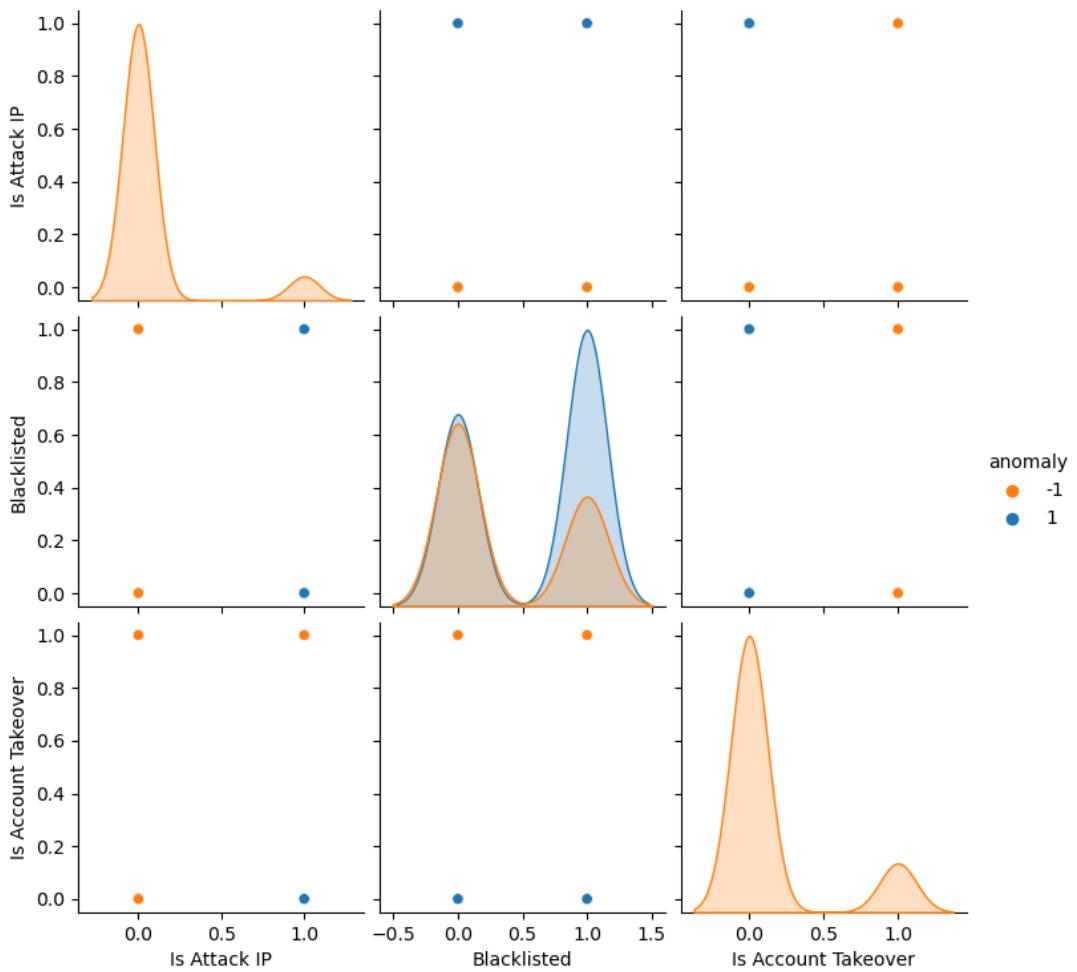
- increasing contamination value to 0.3



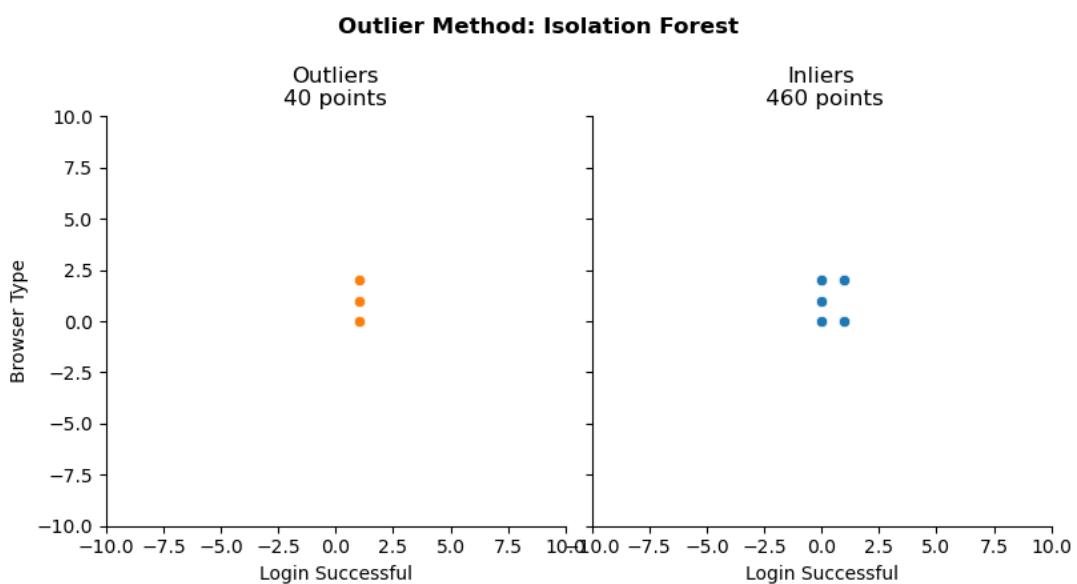


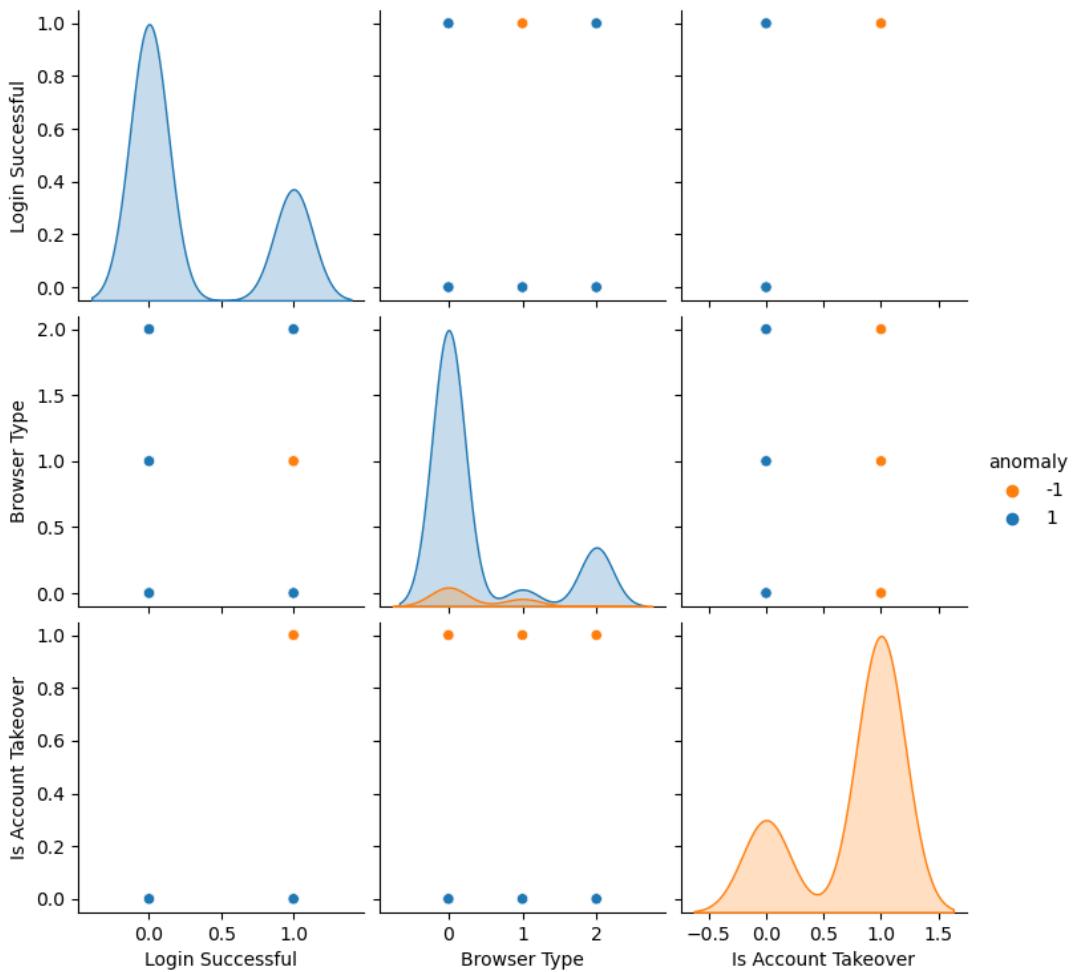
- contamination value == 0.5



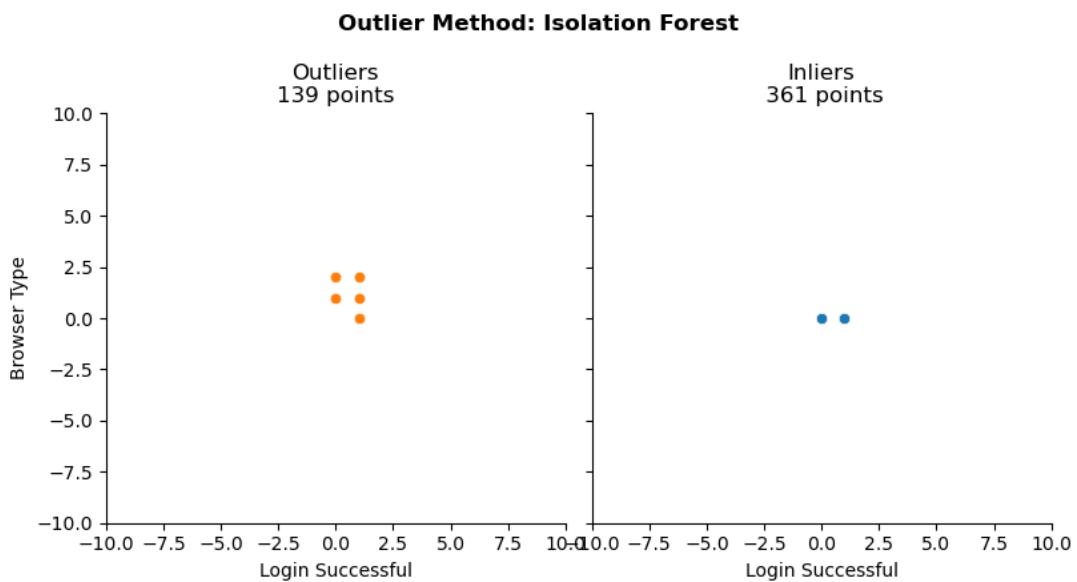


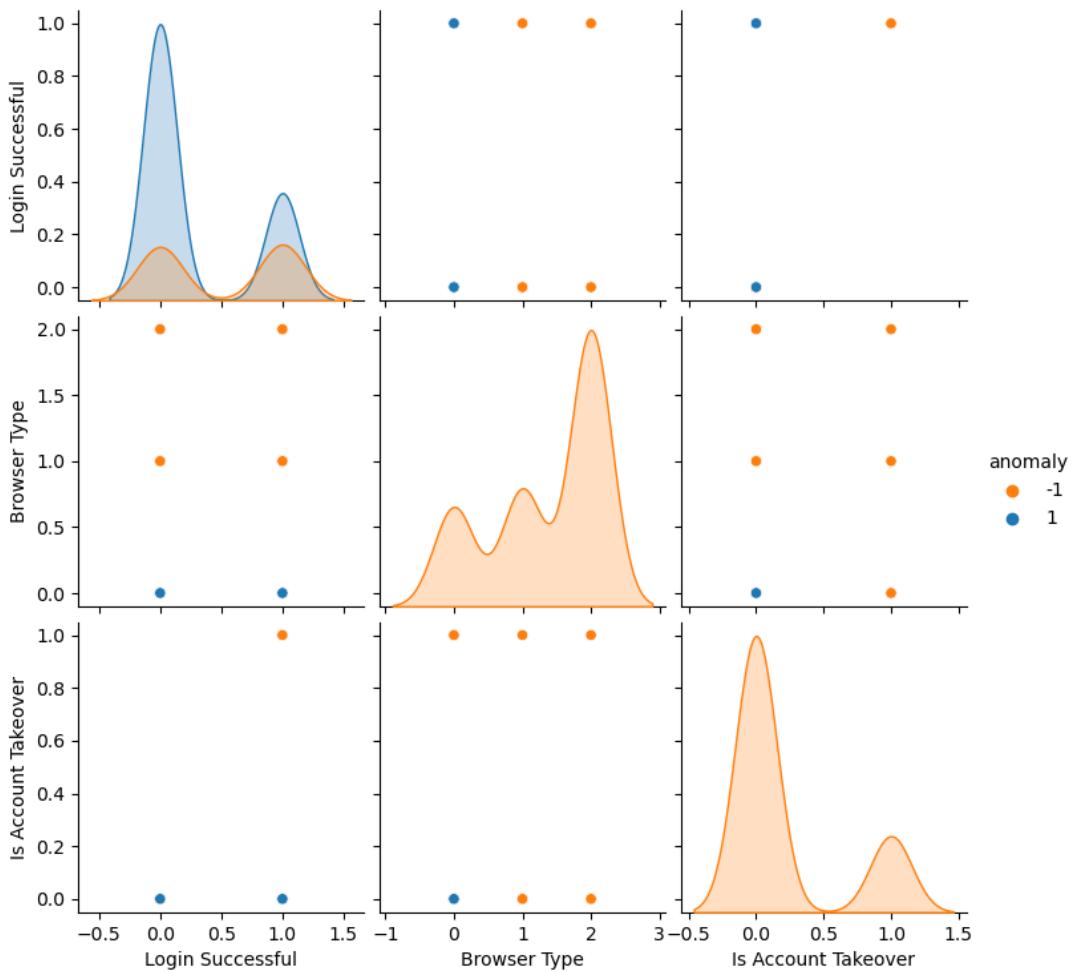
- anomaly_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']
 - contamination value == 0.1



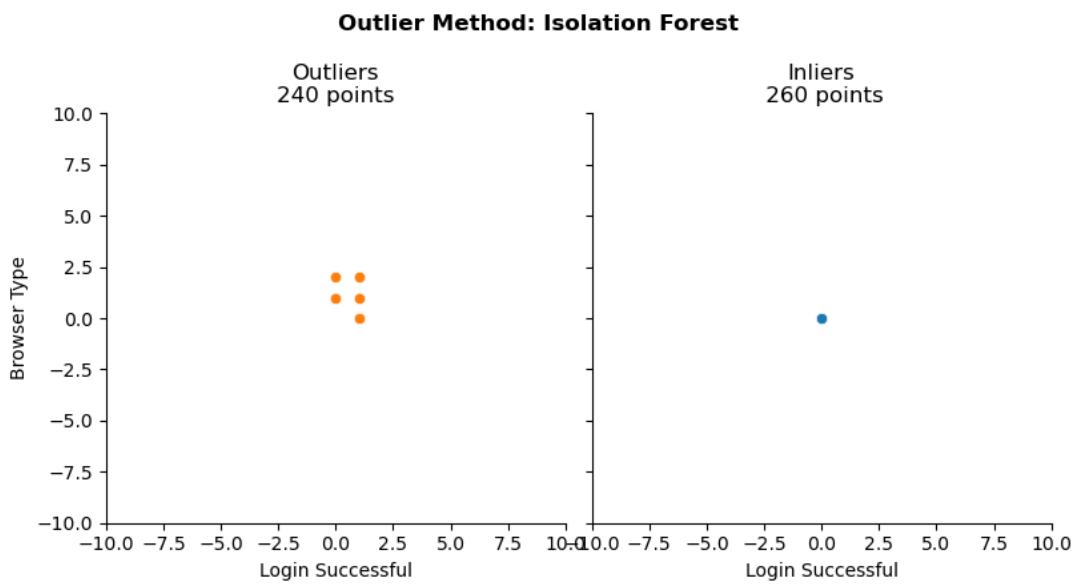


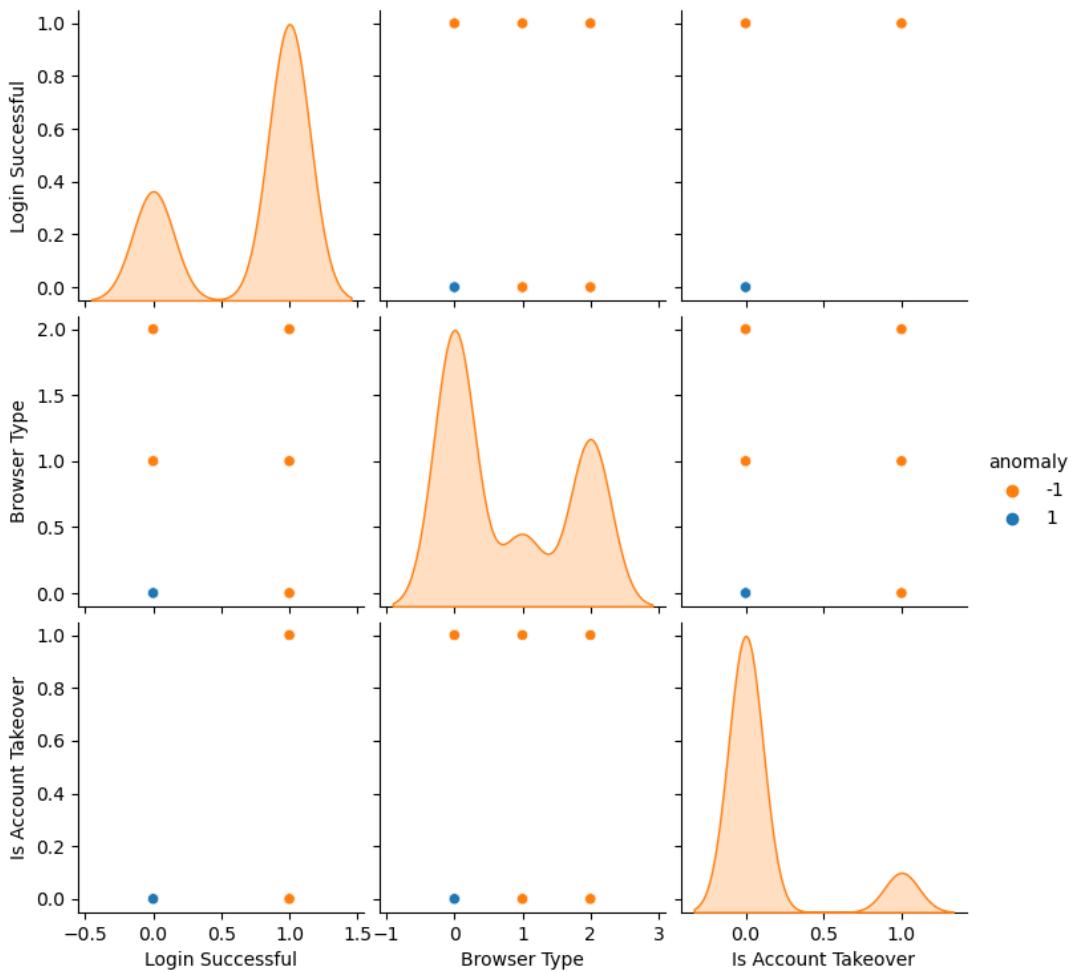
- contamination value == 0.3



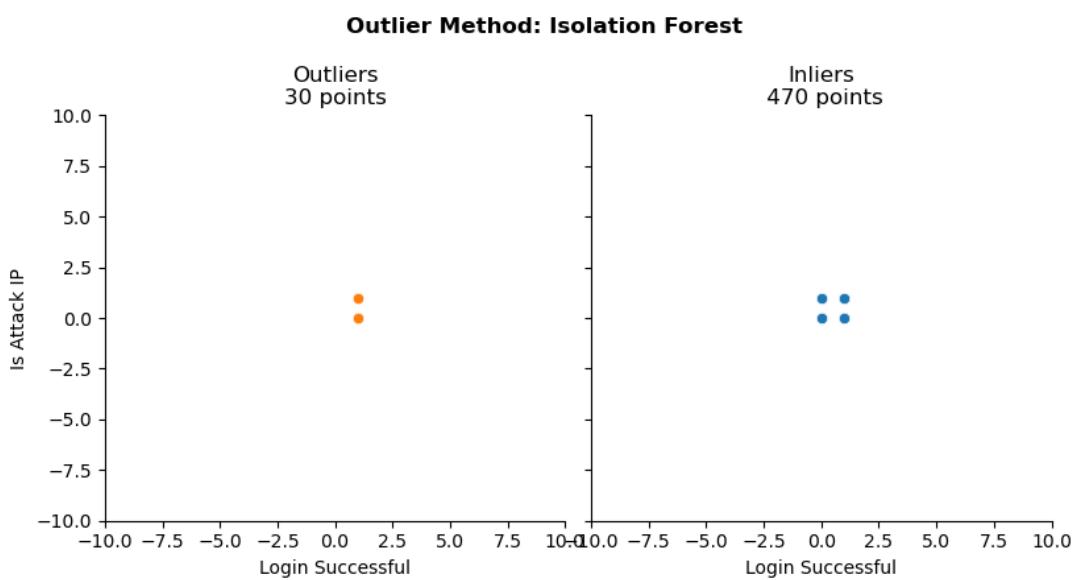


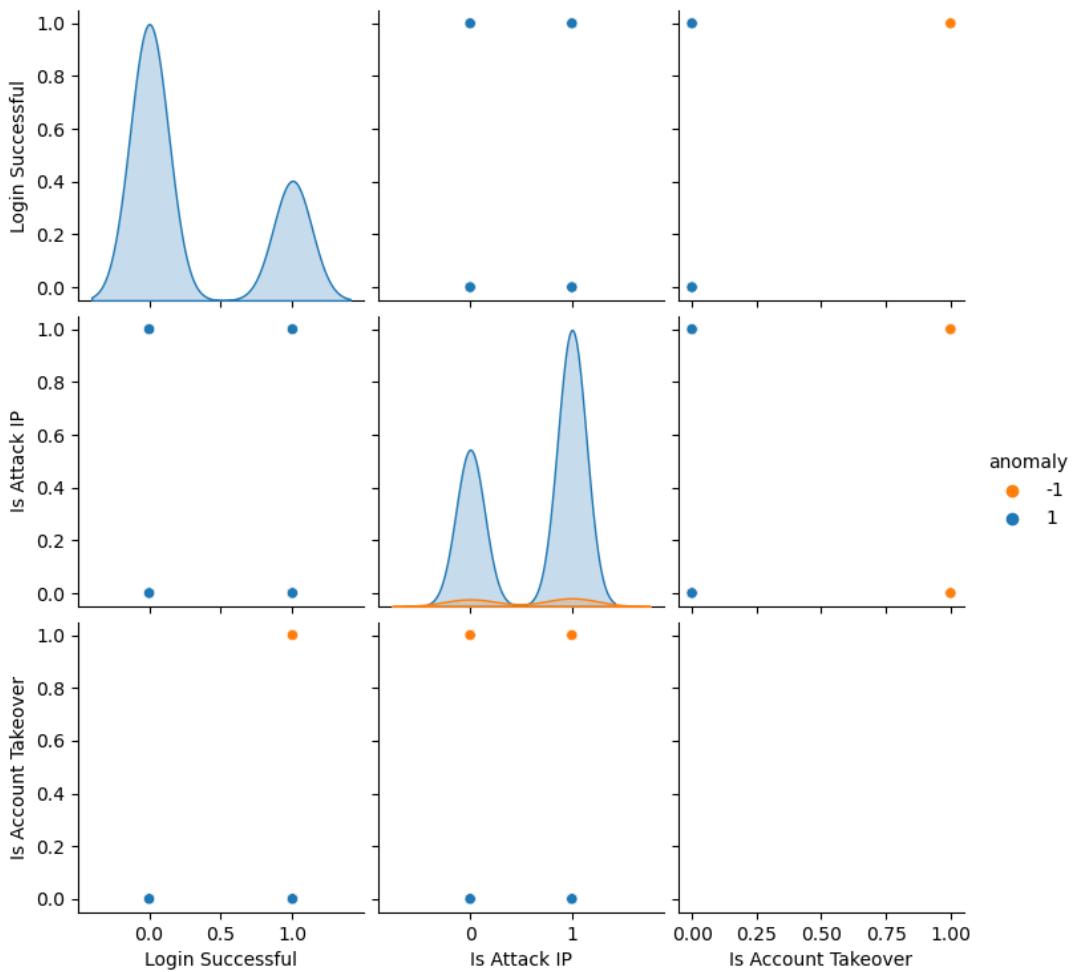
- contamination value == 0.5



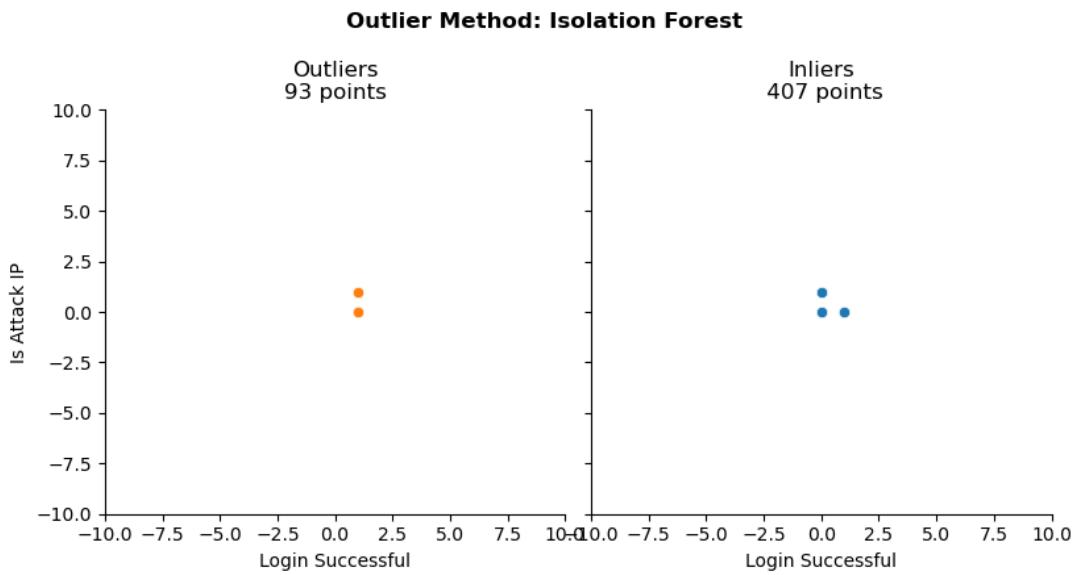


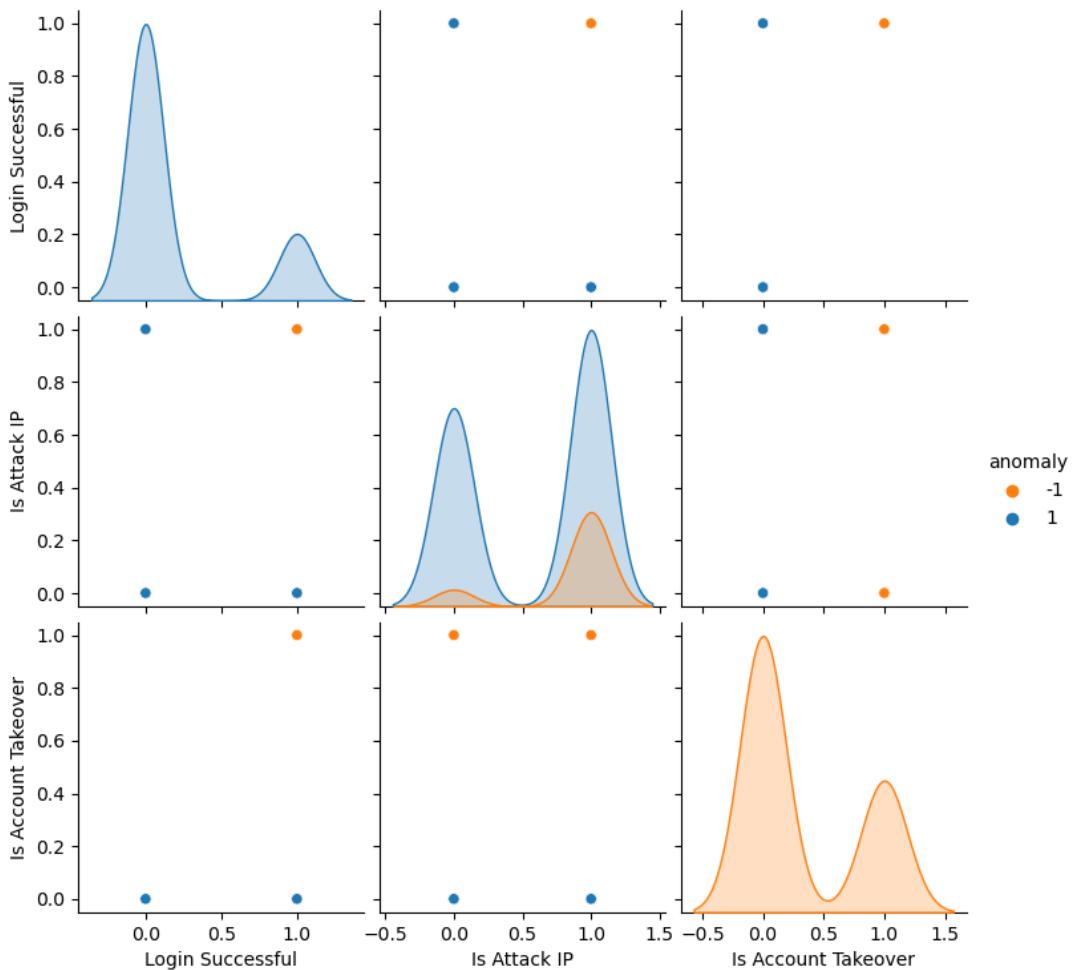
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1



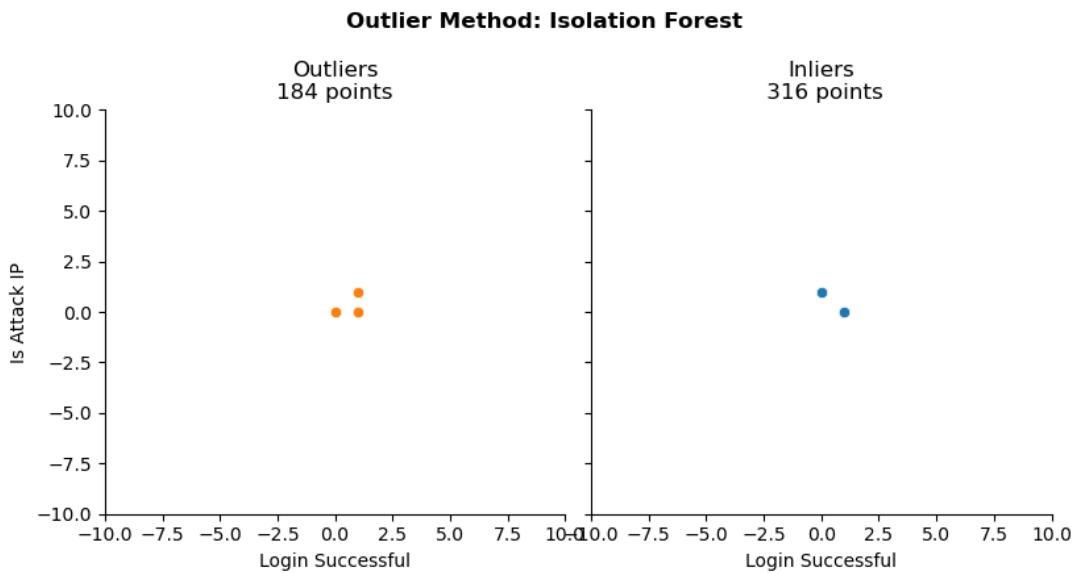


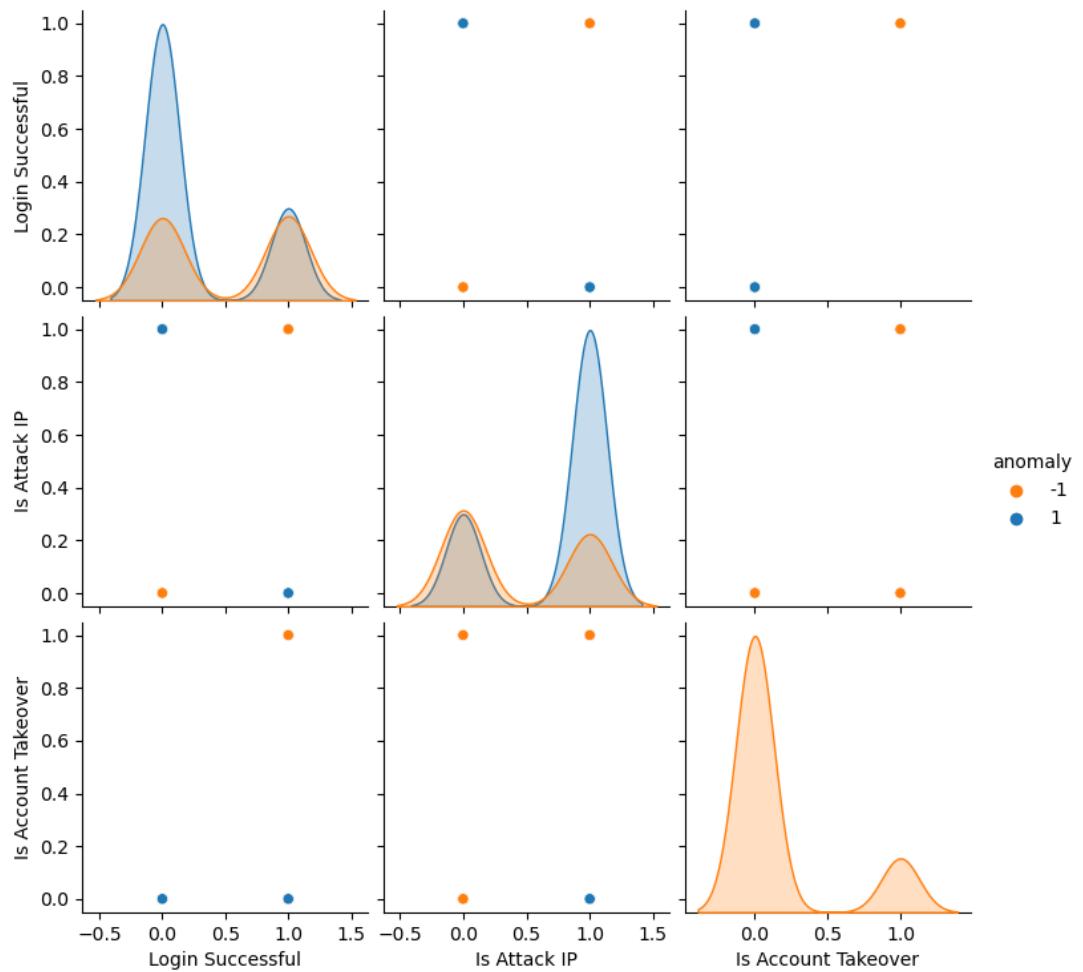
- contamination value == 0.3



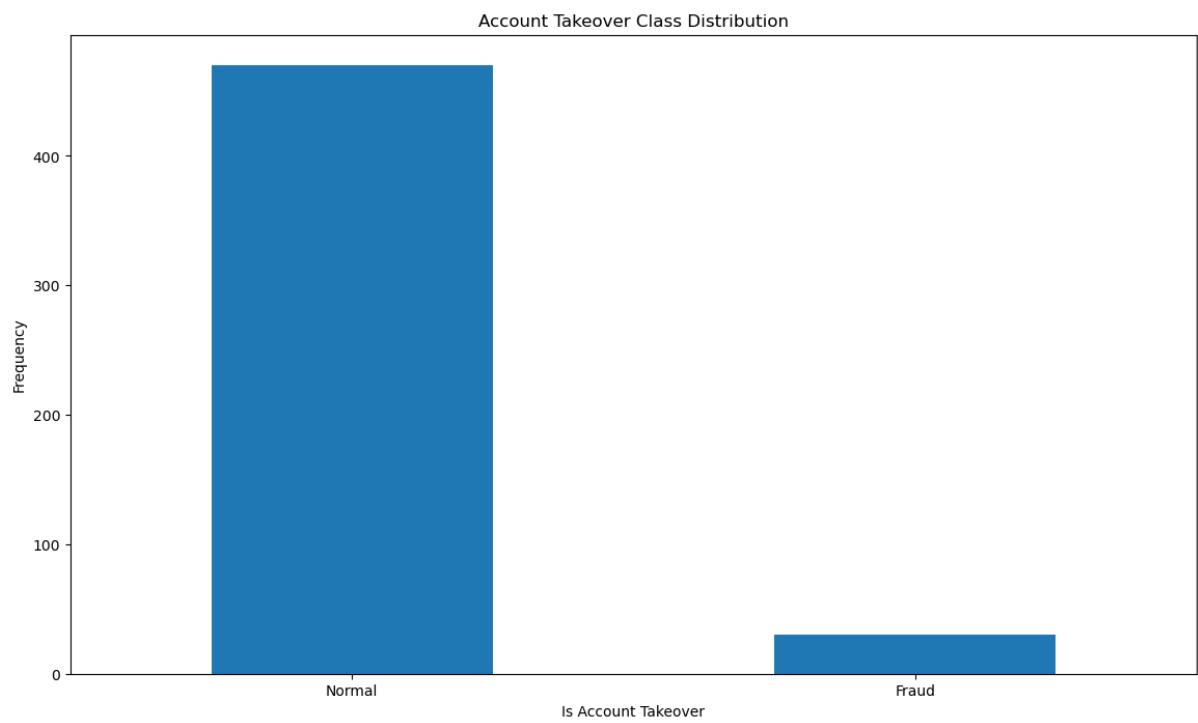


- contamination value == 0.5





2D. iForest, Local Outlier Factor, Support Vector Machine



"{}: {}".format(clf_name,n_errors) → np. Isolation Forest: 21

Isolation Forest: 54

Accuracy Score :

0.892

Classification Report :

	precision	recall	f1-score	support
False	0.94	0.94	0.94	470
True	0.12	0.13	0.13	30
accuracy			0.89	500
macro avg	0.53	0.54	0.54	500
weighted avg	0.90	0.89	0.89	500

Local Outlier Factor: 58

Accuracy Score :

0.884

Classification Report :

	precision	recall	f1-score	support
False	0.94	0.94	0.94	470
True	0.06	0.07	0.06	30
accuracy			0.88	500
macro avg	0.50	0.50	0.50	500
weighted avg	0.89	0.88	0.89	500

Support Vector Machine: 340

Accuracy Score :

0.32

Classification Report :

	precision	recall	f1-score	support
False	0.86	0.33	0.48	470
True	0.02	0.17	0.03	30
accuracy			0.32	500
macro avg	0.44	0.25	0.25	500
weighted avg	0.81	0.32	0.45	500

3. Dummy Variable Encoding → df3 =
`pd.get_dummies(df3, columns=['Country', 'Device Type', 'IP Address', 'Login Timestamp', 'Browser Type'], prefix=["cntr", "dvc", "ip", "lgnt", "brw"])`

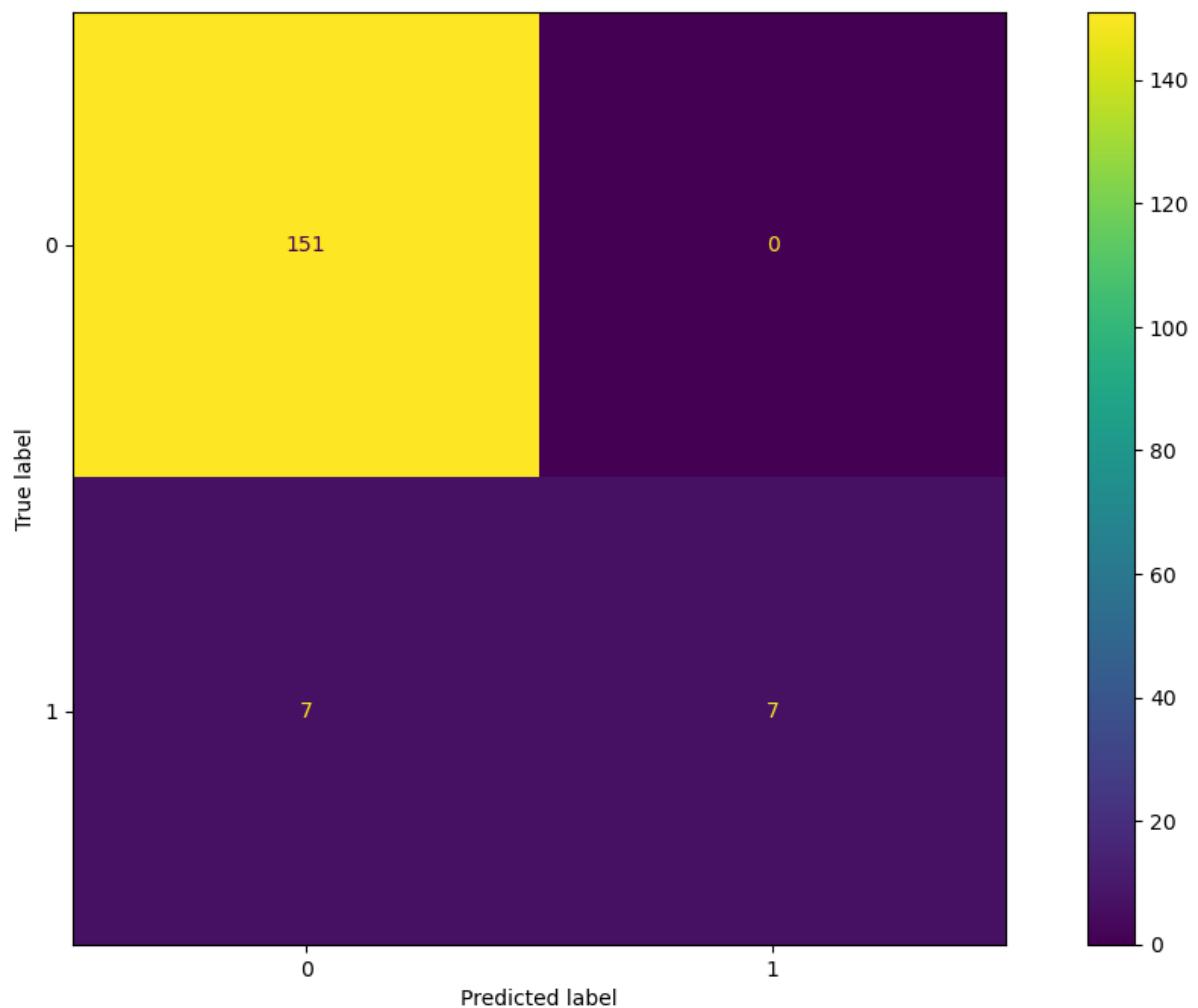
3A. XGBoost

memory usage: 445.0 KB

Accuracy: 95.76%

	precision	recall	f1-score	support
False	0.96	1.00	0.98	151
True	1.00	0.50	0.67	14
accuracy			0.96	165
macro avg	0.98	0.75	0.82	165
weighted avg	0.96	0.96	0.95	165

TN: 0, FP: 2, FN: 1, TP: 1



3B. One Class SVM

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.2 (same for any other col)

```

The number of records in the training dataset is 400
The number of records in the test dataset is 100
The training dataset has 376 records for the majority class and 24 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.07     0.14      94
      True         0.06     1.00     0.12       6
      accuracy          0.13      100
      macro avg       0.53     0.54     0.13      100
      weighted avg    0.94     0.13     0.14      100
      The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.94     1.00     0.97      94
      True         0.00     0.00     0.00       6
      accuracy          0.94      100
      macro avg       0.47     0.50     0.48      100
      weighted avg    0.88     0.94     0.91      100
[[ 7 87]
 [ 0 6]]

```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.3 (same for any other col)

```

The number of records in the training dataset is 350
The number of records in the test dataset is 150
The training dataset has 330 records for the majority class and 20 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.24     0.38     140
      True         0.09     1.00     0.16      10
      accuracy          0.29      150
      macro avg       0.54     0.62     0.27      150
      weighted avg    0.94     0.29     0.37      150
      The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.93     1.00     0.97     140
      True         0.00     0.00     0.00      10
      accuracy          0.93      150
      macro avg       0.47     0.50     0.48      150
      weighted avg    0.87     0.93     0.90      150
[[ 33 107]
 [ 0 10]]

```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.5 (same for any other col)

```

The number of records in the training dataset is 250
The number of records in the test dataset is 250
The training dataset has 235 records for the majority class and 15 records for the minority class.
      precision    recall   f1-score   support
False        1.00     0.03     0.06     235
True        0.06     1.00     0.12      15

accuracy                           0.09      250
macro avg                         0.53      250
weighted avg                      0.94      250

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.94     1.00     0.97     235
True        0.00     0.00     0.00      15

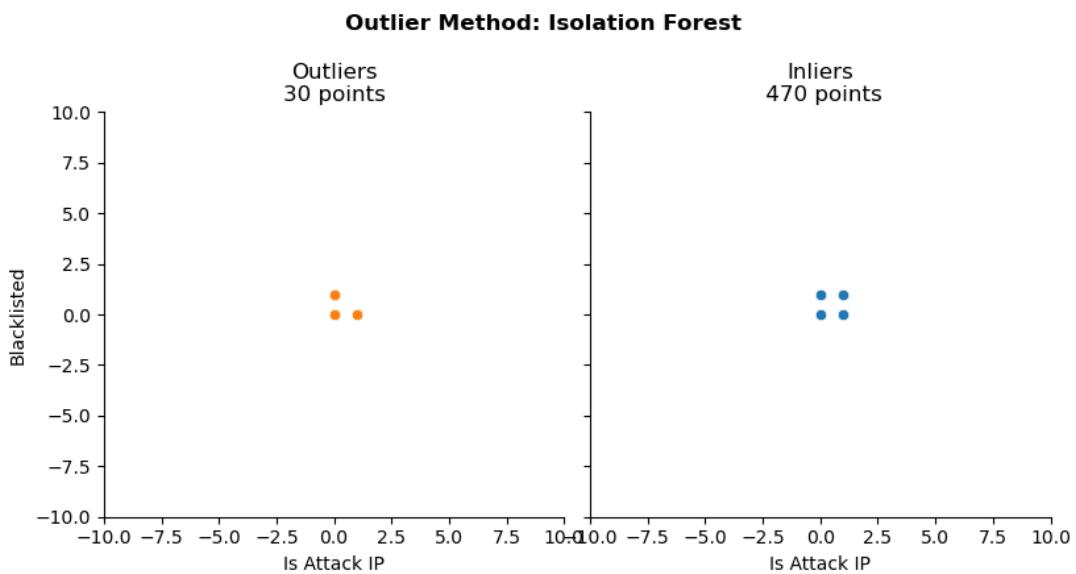
accuracy                           0.94      250
macro avg                         0.47      250
weighted avg                      0.88      250

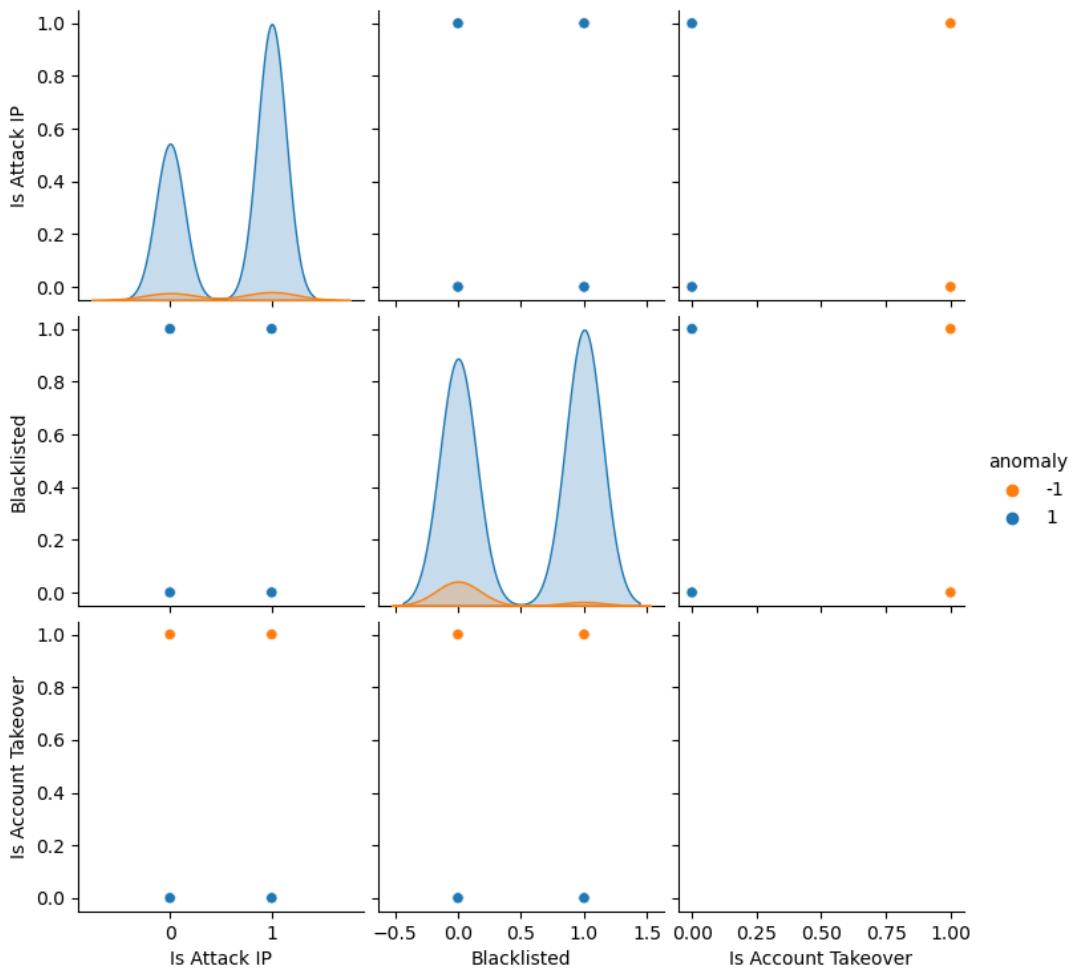
[[ 7 228]
 [ 0 15]]

```

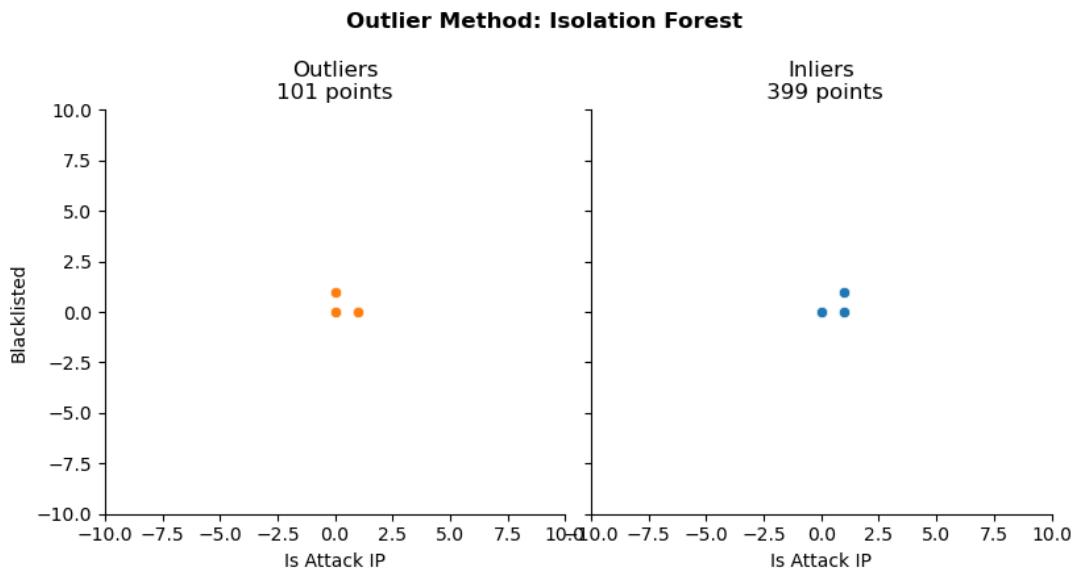
3C. Isolation Forest

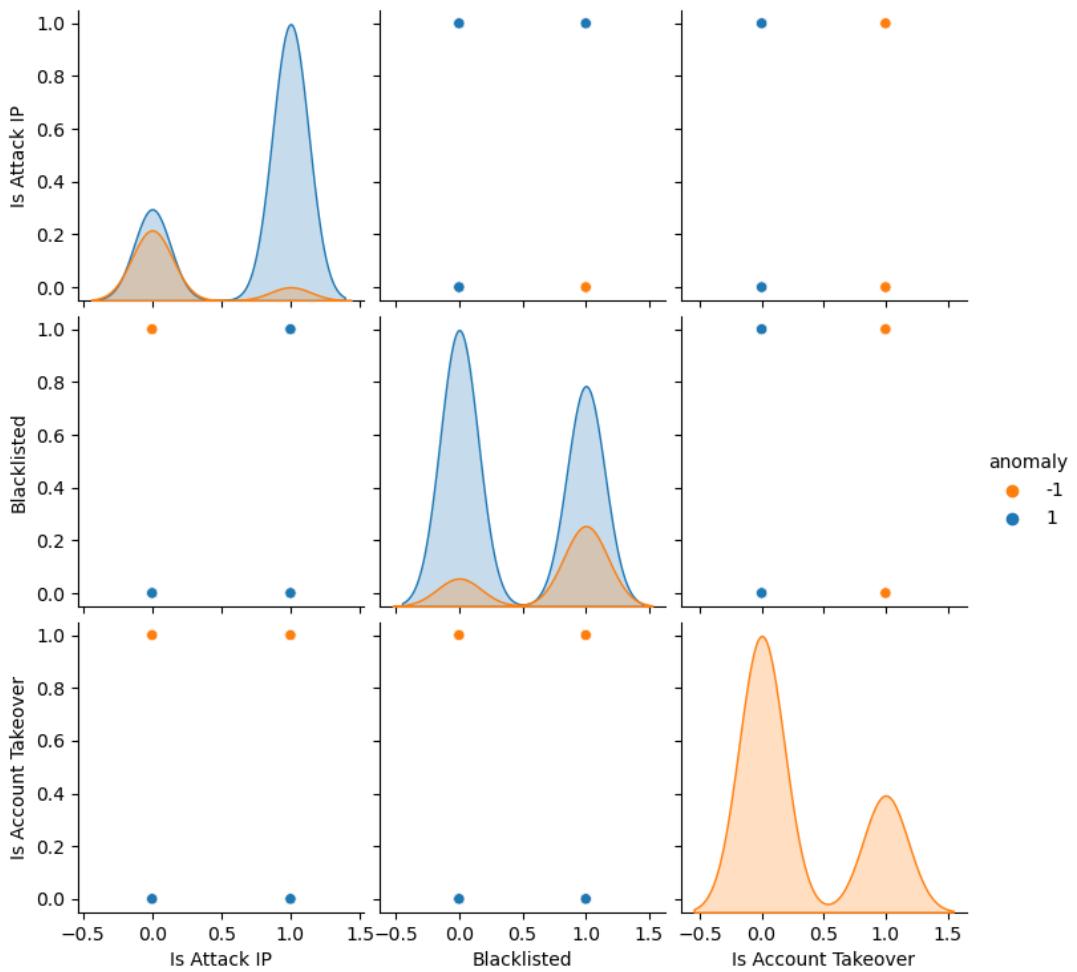
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



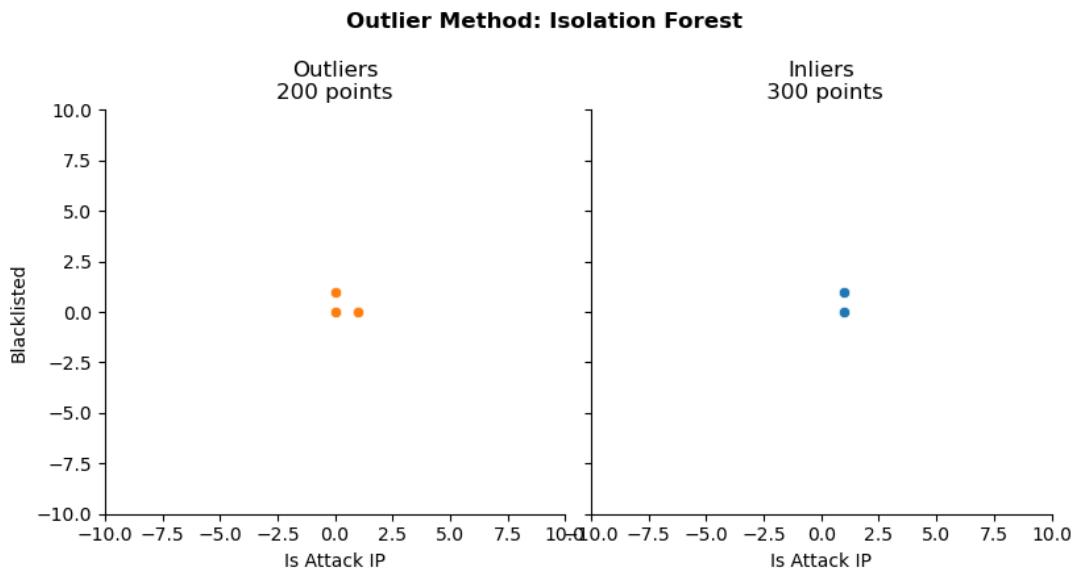


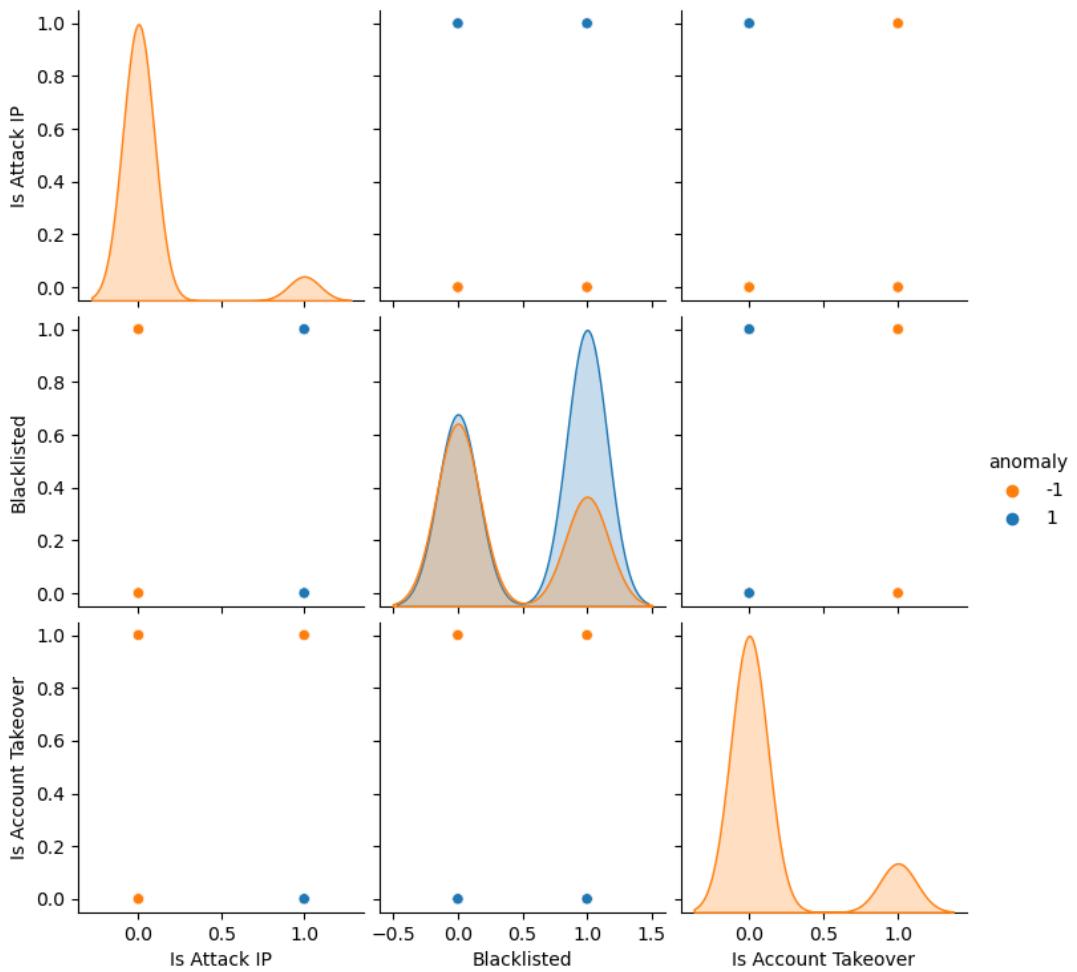
- contamination value == 0.3



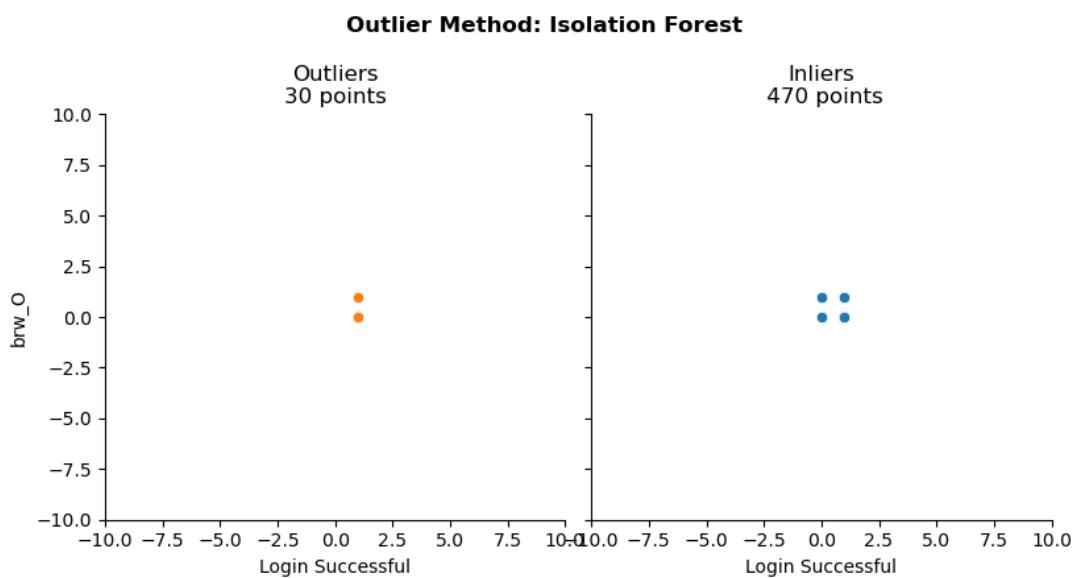


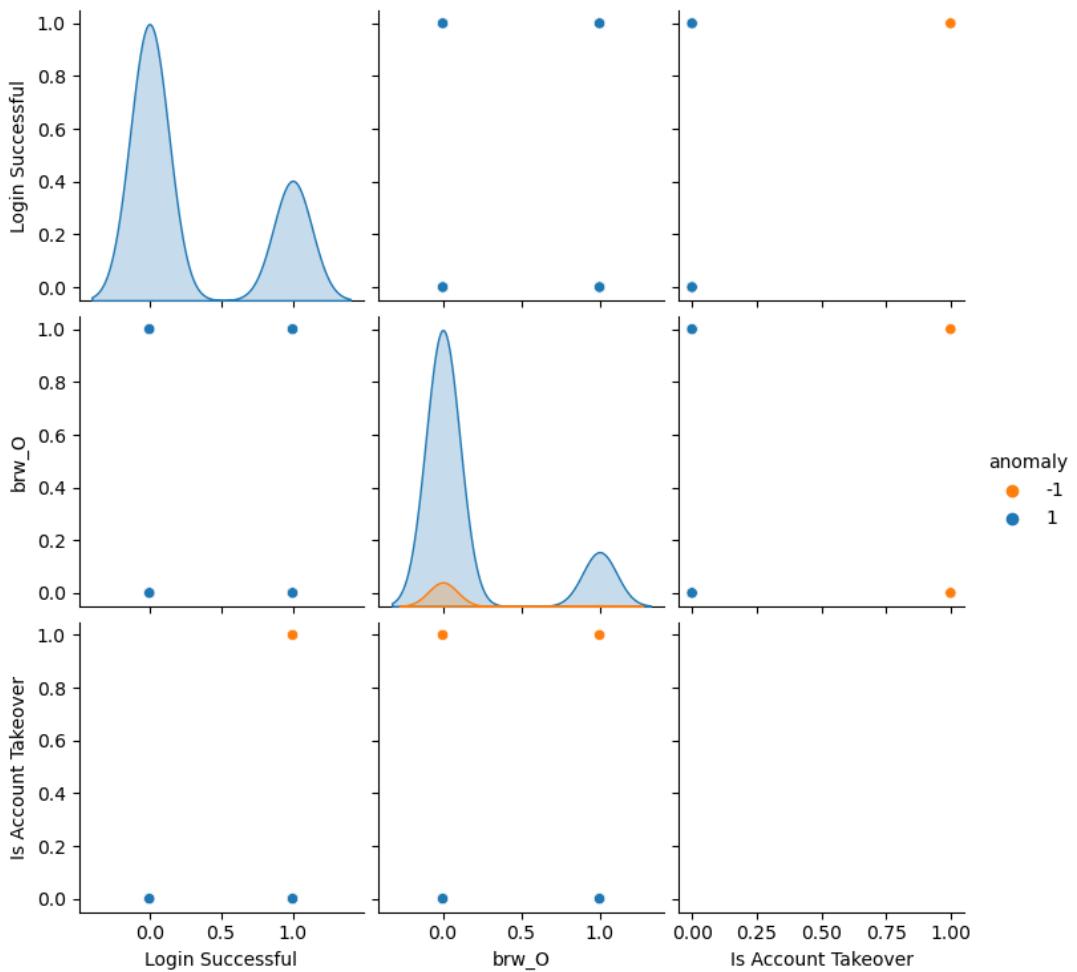
- contamination value == 0.5



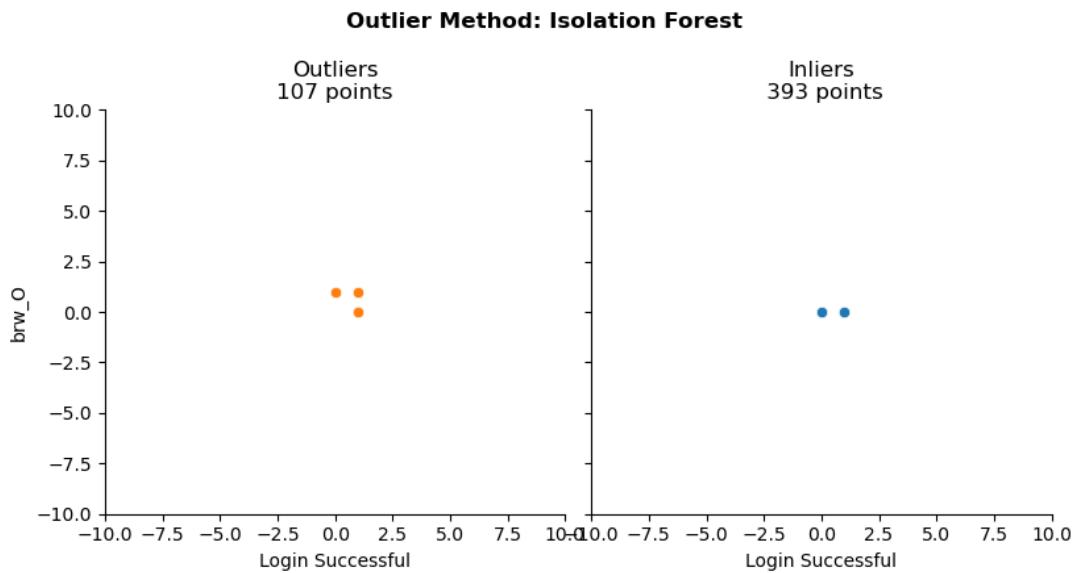


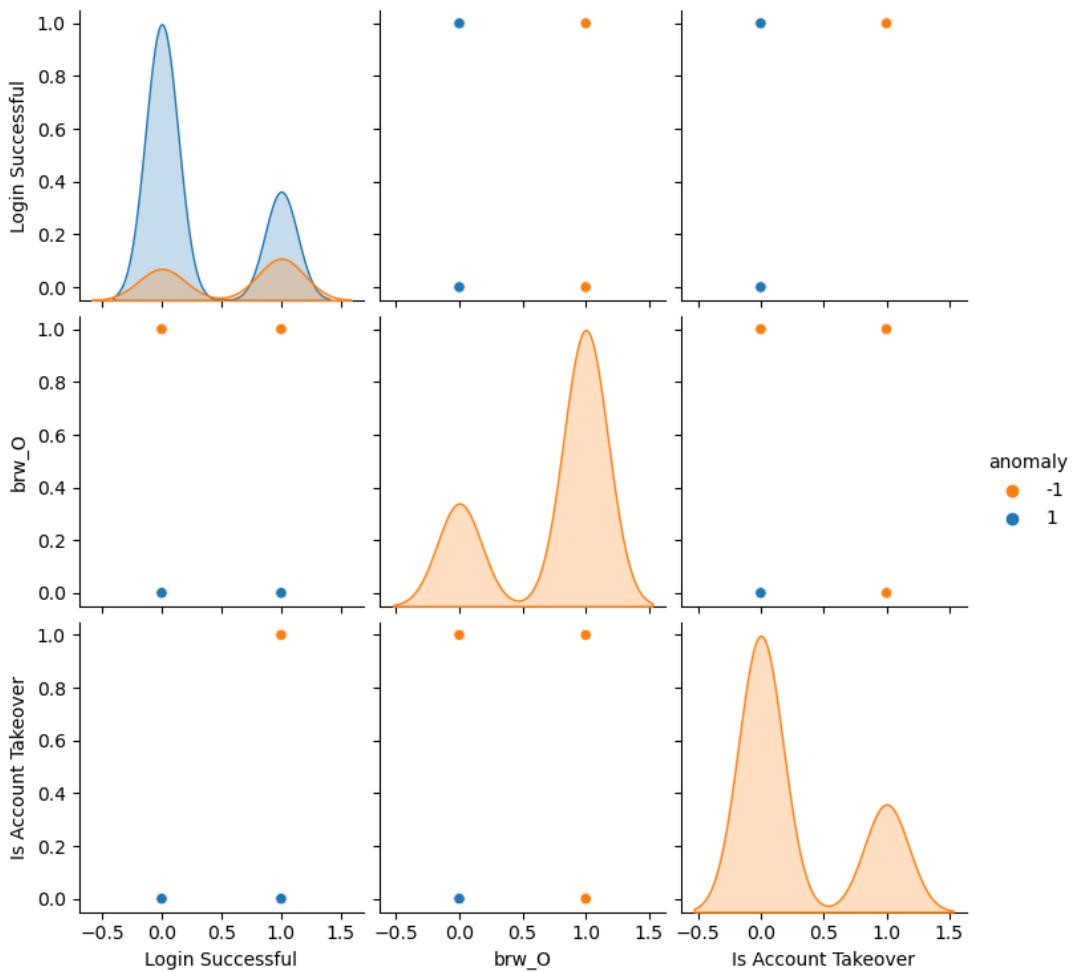
- anomaly_inputs = ['Login Successful', 'Browser Type (brw_O)', 'Is Account Takeover']
 - contamination value == 0.1





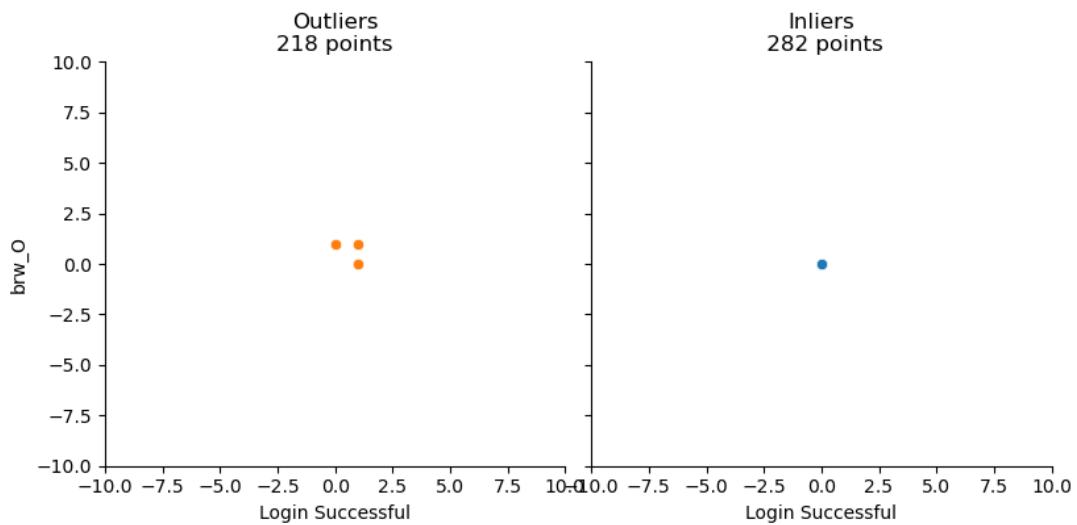
- contamination value == 0.3

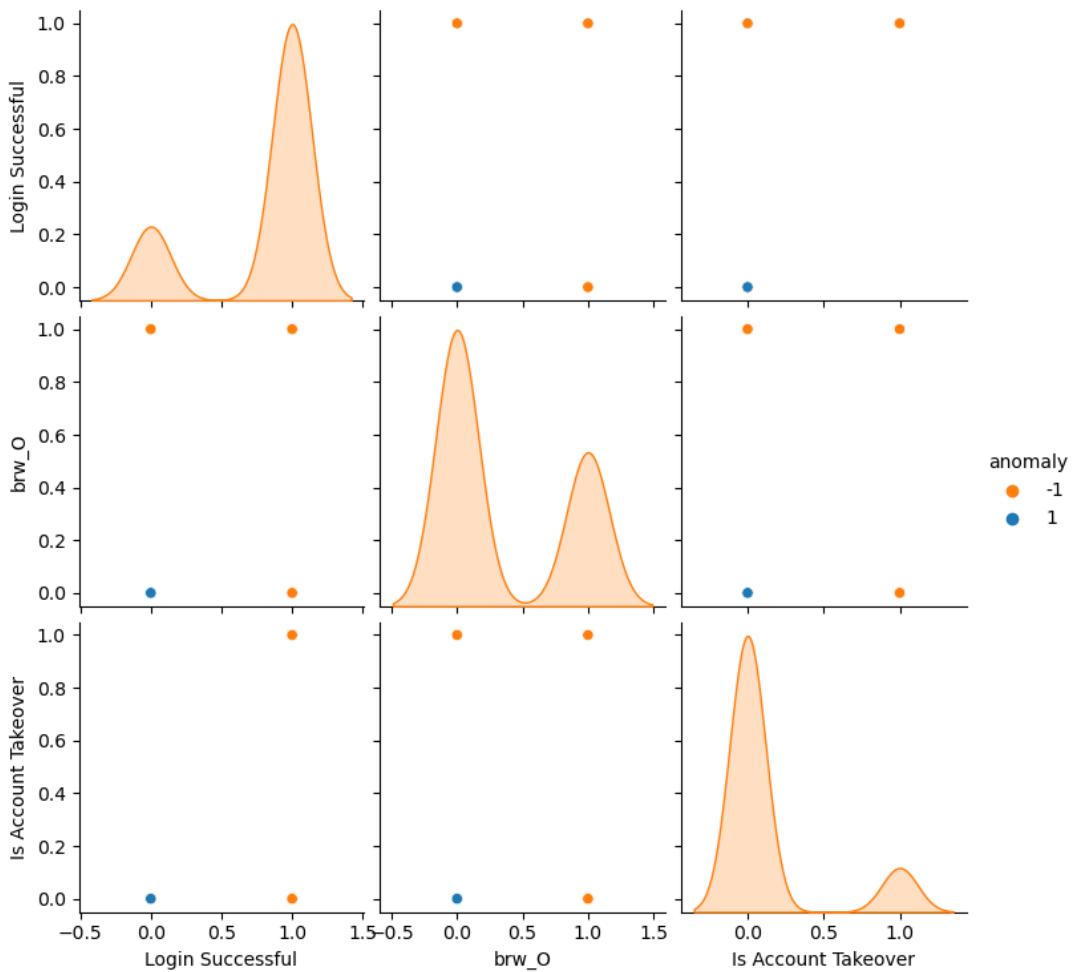




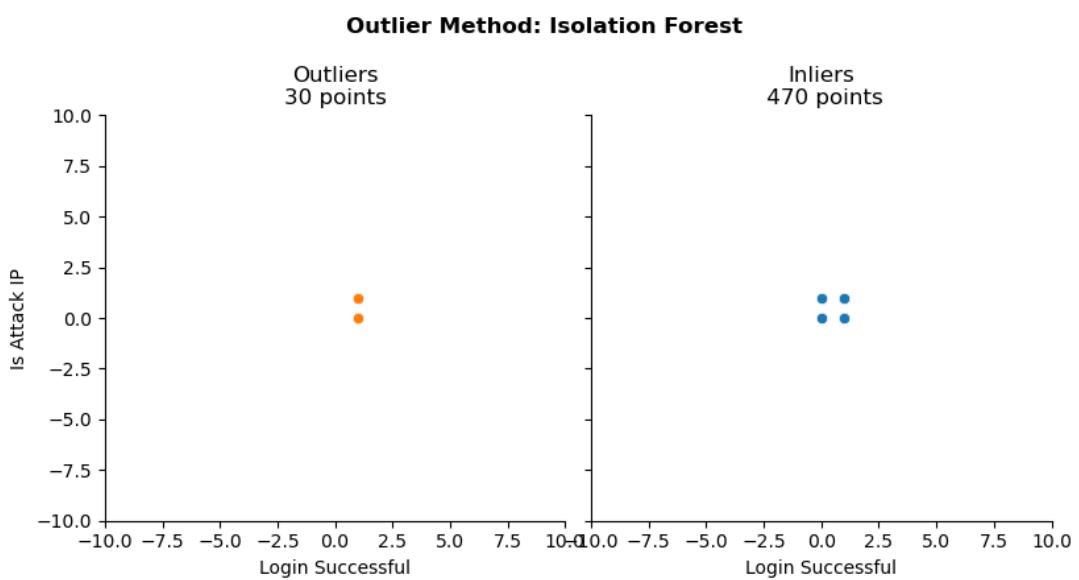
- contamination value == 0.5

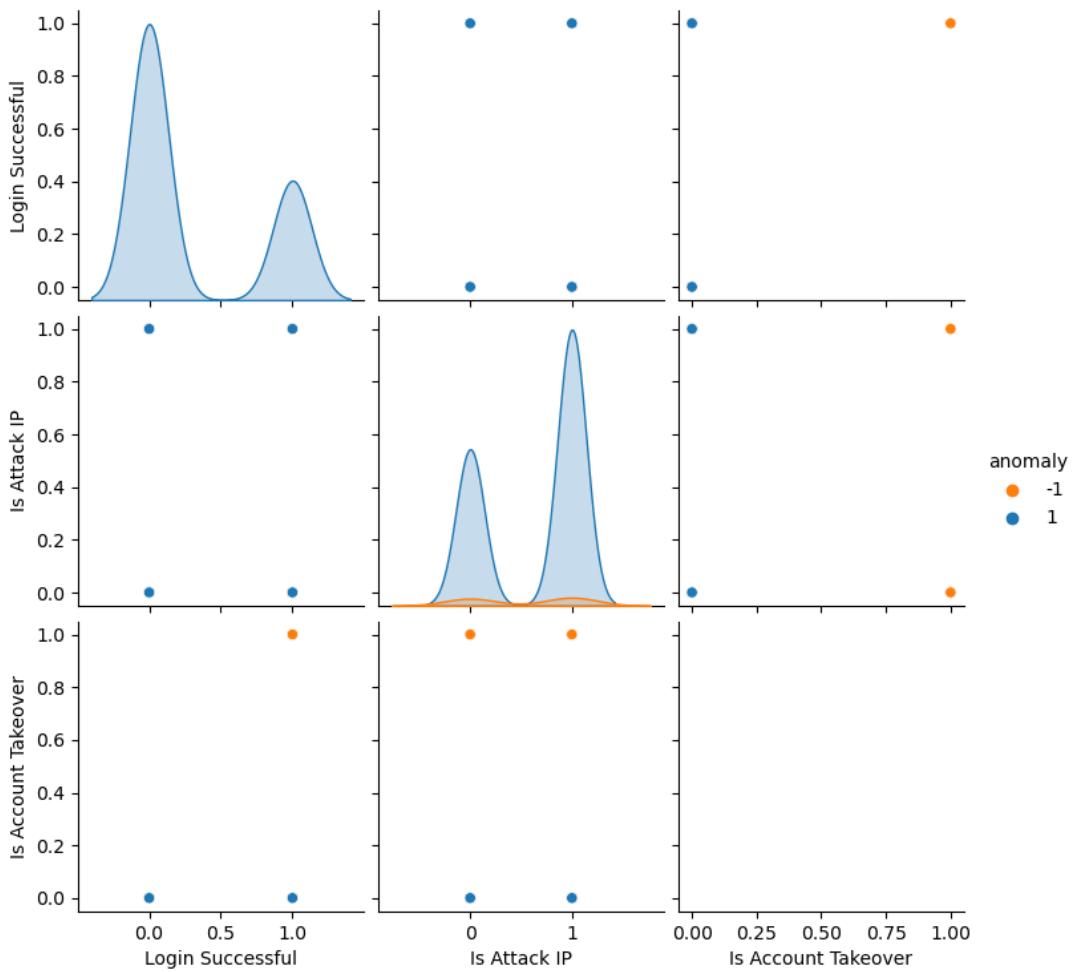
Outlier Method: Isolation Forest



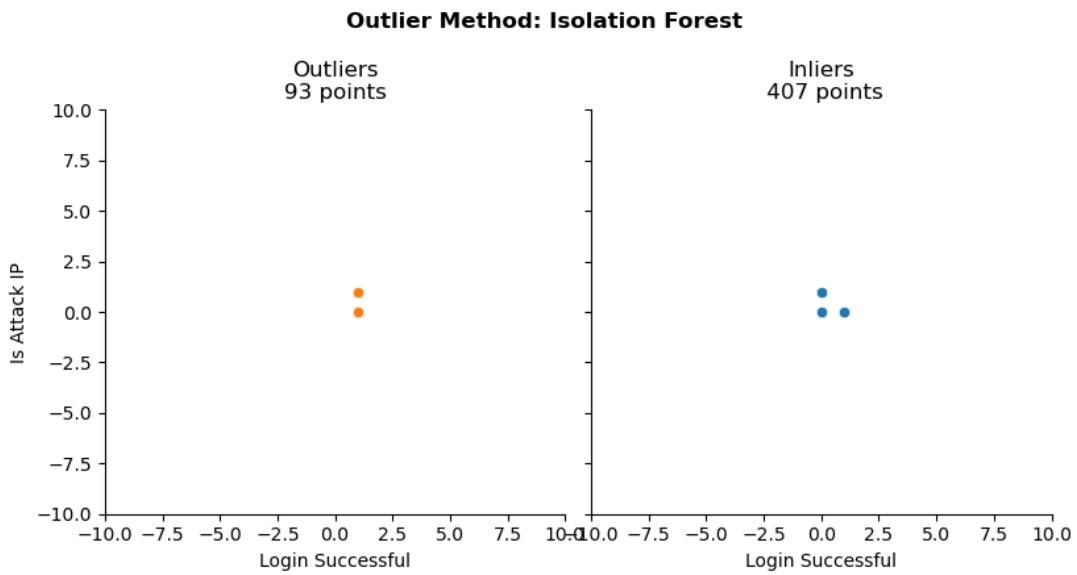


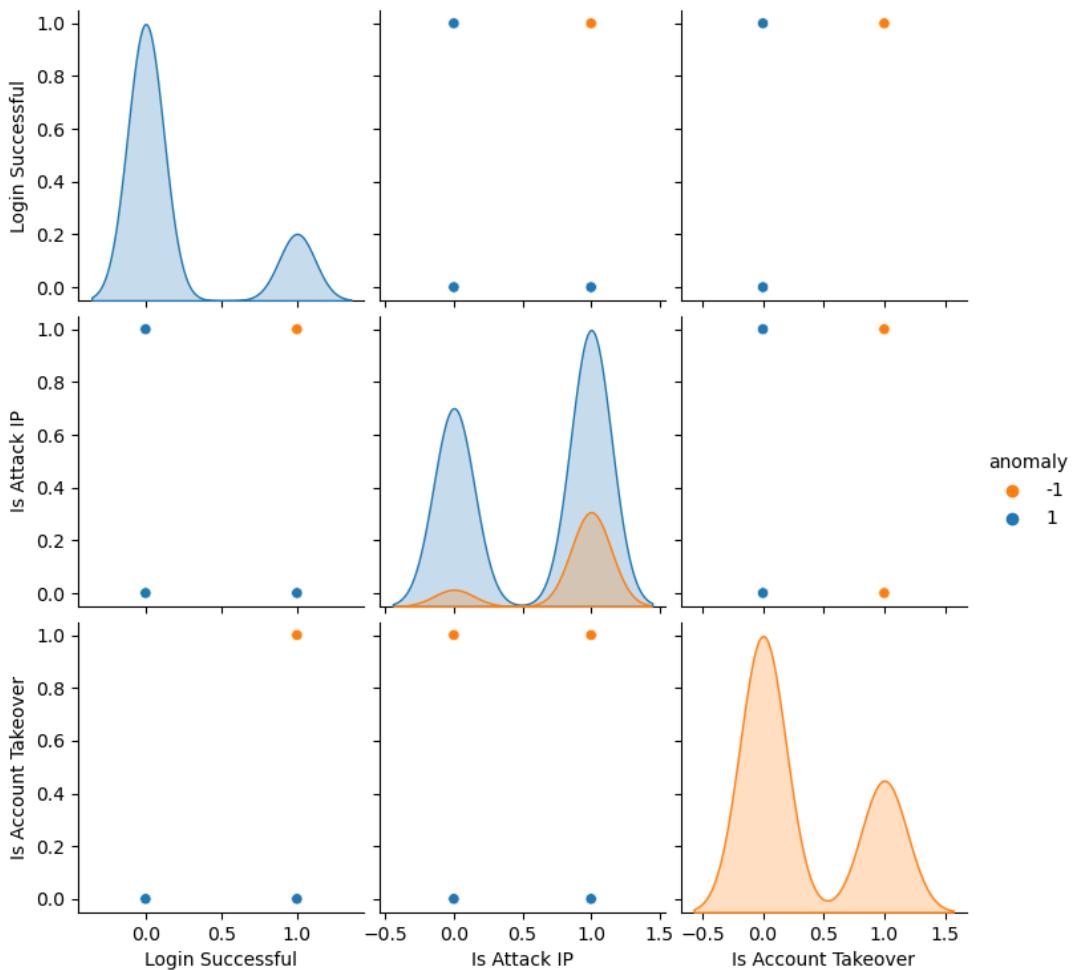
- `anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']`
 - contamination value == 0.1



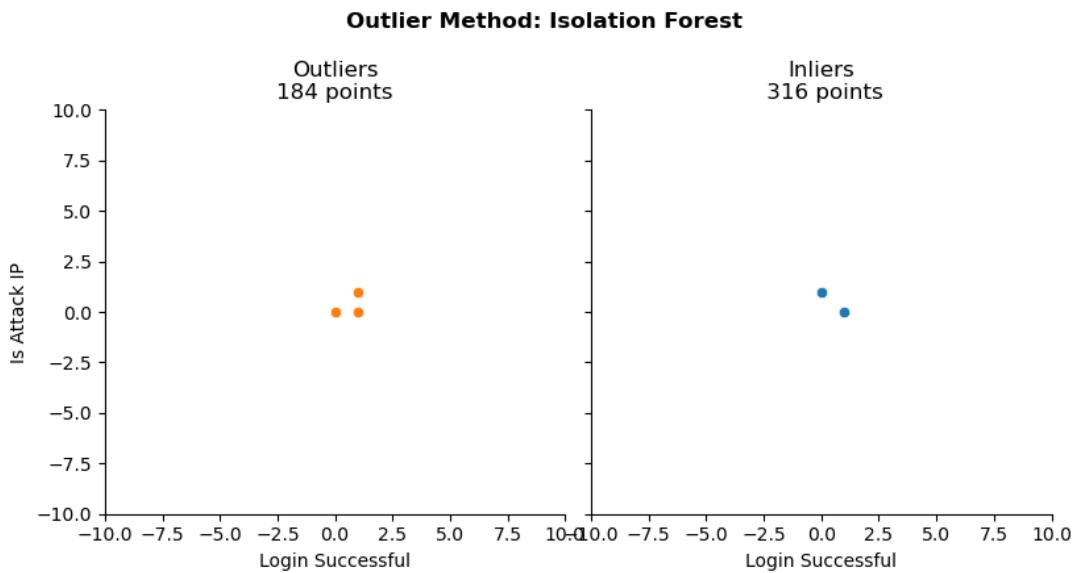


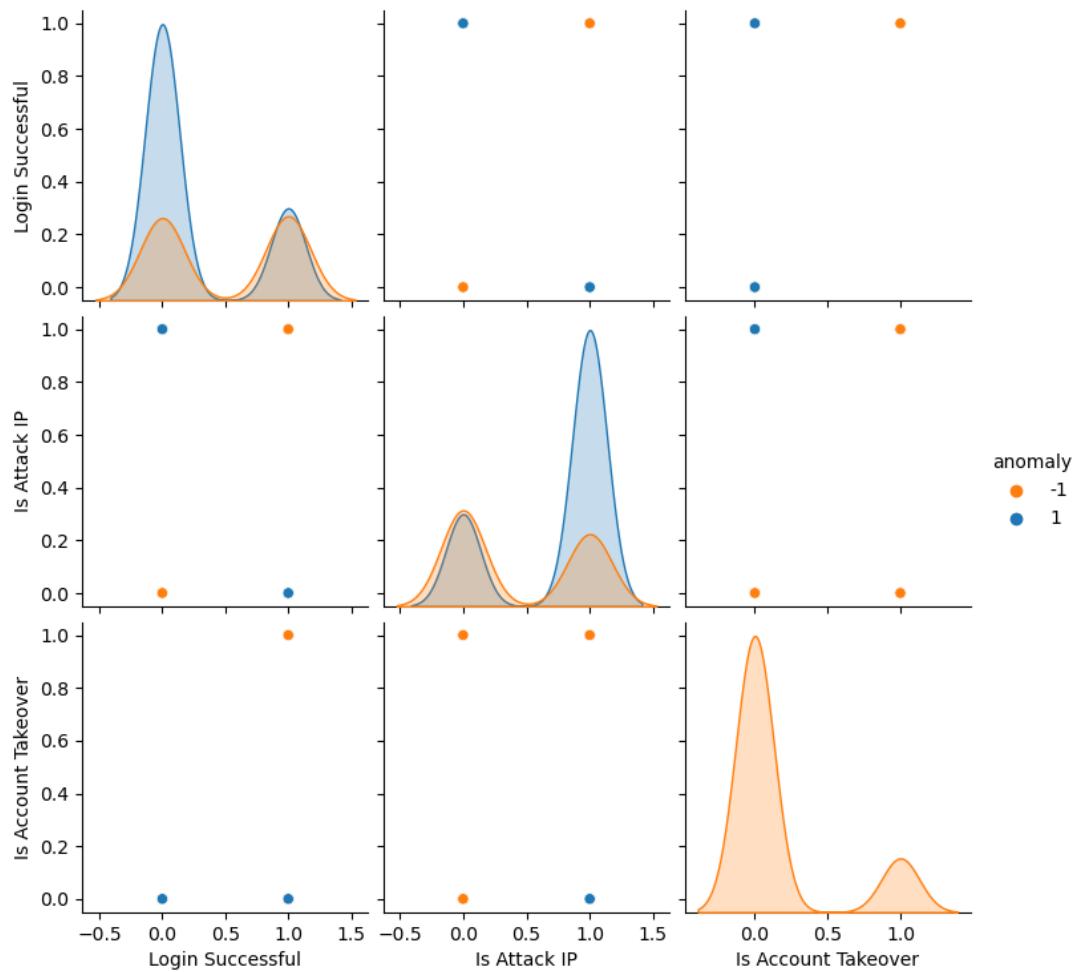
- contamination value == 0.3





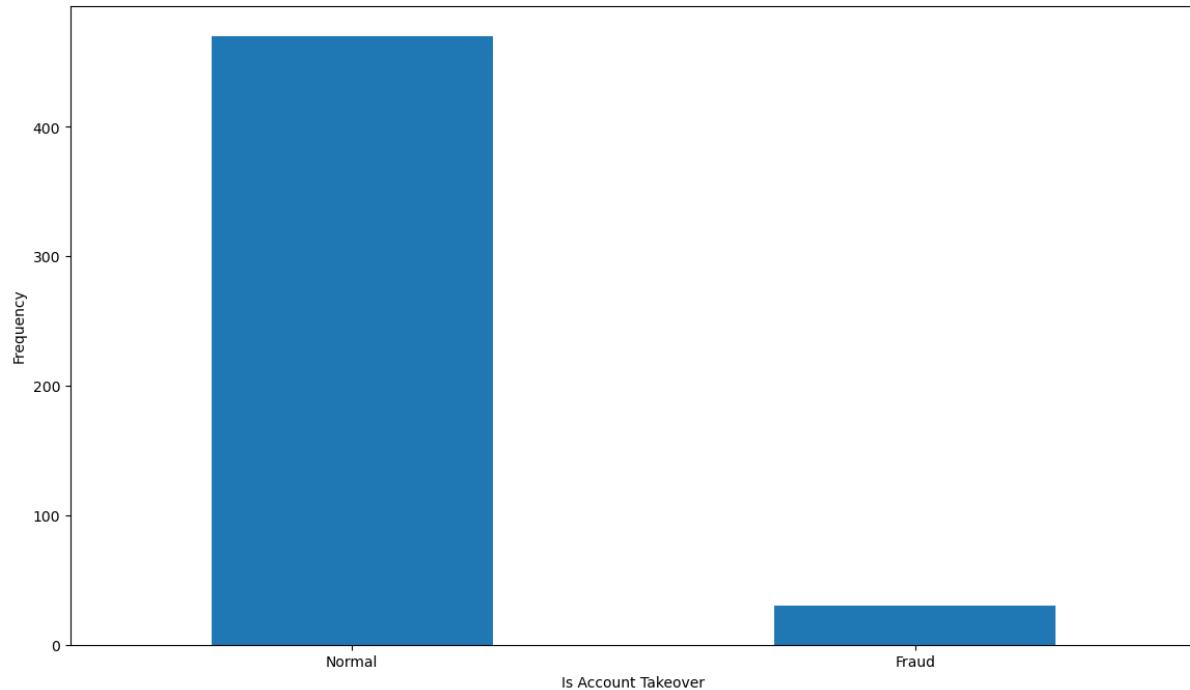
- contamination value == 0.5





3D. iForest, Local Outlier Factor, Support Vector Machine

Account Takeover Class Distribution



Isolation Forest: 54

Accuracy Score :

0.892

Classification Report :

	precision	recall	f1-score	support
False	0.94	0.94	0.94	470
True	0.12	0.13	0.13	30
accuracy			0.89	500
macro avg	0.53	0.54	0.54	500
weighted avg	0.90	0.89	0.89	500

Local Outlier Factor: 54

Accuracy Score :

0.892

Classification Report :

	precision	recall	f1-score	support
False	0.94	0.94	0.94	470
True	0.12	0.13	0.13	30
accuracy			0.89	500
macro avg	0.53	0.54	0.54	500
weighted avg	0.90	0.89	0.89	500

Support Vector Machine: 340

Accuracy Score :

0.32

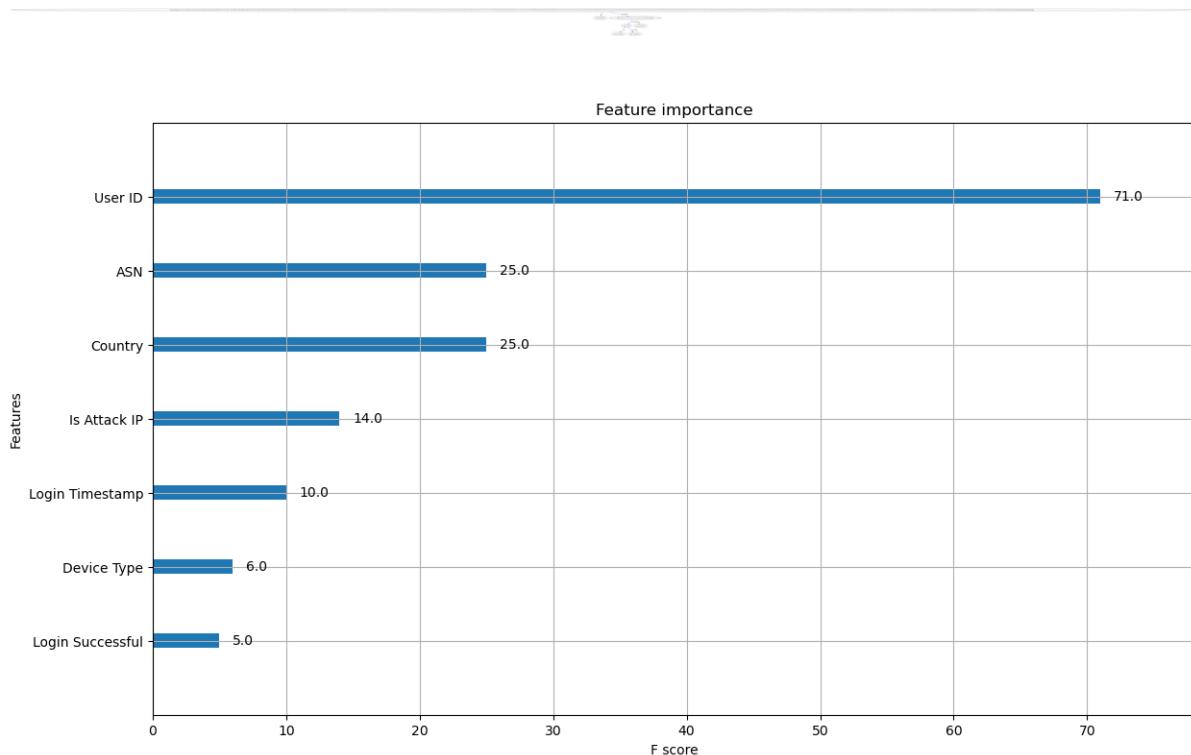
Classification Report :

	precision	recall	f1-score	support
False	0.86	0.33	0.48	470
True	0.02	0.17	0.03	30
accuracy			0.32	500
macro avg	0.44	0.25	0.25	500
weighted avg	0.81	0.32	0.45	500

III. 77/500 → 15.4%

1. XGB Experimental

```
-----  
0 Login Timestamp      500 non-null   category  
1 User ID              500 non-null   int64  
2 IP Address            500 non-null   category  
3 Country               500 non-null   category  
4 ASN                  500 non-null   int64  
5 Device Type           500 non-null   category  
6 Login Successful       500 non-null   bool  
7 Is Attack IP          500 non-null   bool  
8 Blacklisted            500 non-null   bool  
9 Browser Type           500 non-null   category  
10 Is Account Takeover  500 non-null  bool  
dtypes: bool(4), category(5), int64(2)  
memory usage: 54.2 KB  
Feature importances:  
[0.5381381  0.00136504 0.          0.22627188 0.04183649 0.00924834  
 0.04554712 0.13759302 0.          0.          ]
```



2. Label encoding (cat.codes)

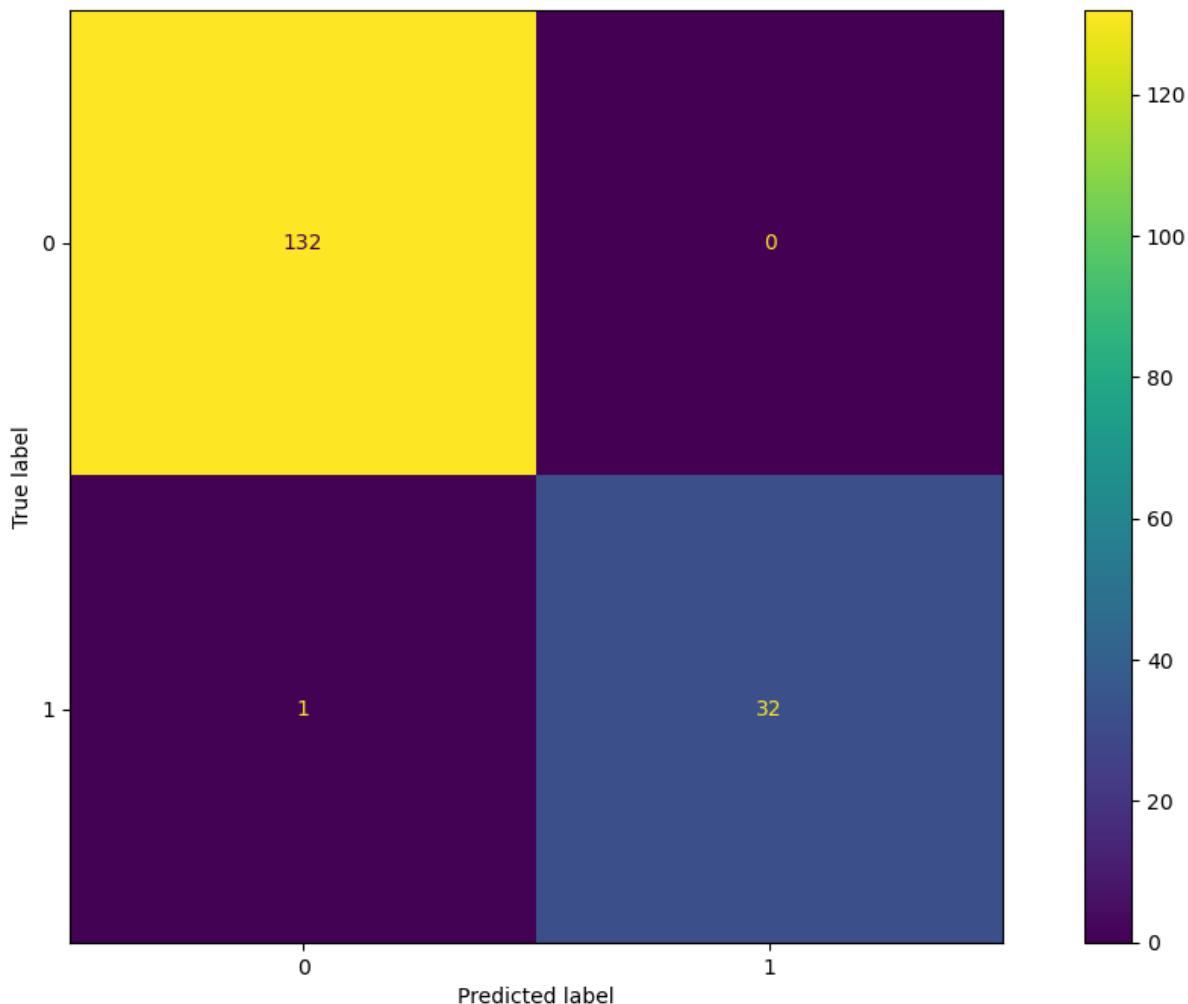
2A. XGBoost

memory usage: 13.3 KB

Accuracy: 99.39%

	precision	recall	f1-score	support
False	0.99	1.00	1.00	132
True	1.00	0.97	0.98	33
accuracy			0.99	165
macro avg	1.00	0.98	0.99	165
weighted avg	0.99	0.99	0.99	165

TN: 0, FP: 2, FN: 1, TP: 1



2B. ONE CLASS SVM

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.2 (same for any other col)

```
The number of records in the training dataset is 400
The number of records in the test dataset is 100
The training dataset has 339 records for the majority class and 61 records for the minority class.

precision    recall   f1-score   support
False        0.00     0.00     0.00      84
True         0.16     1.00     0.28      16

accuracy          0.16      100
macro avg       0.08     0.50     0.14      100
weighted avg    0.03     0.16     0.04      100

The customized score threshold for 2% of outliers is 0.00
precision    recall   f1-score   support
False        0.84     1.00     0.91      84
True         0.00     0.00     0.00      16

accuracy          0.84      100
macro avg       0.42     0.50     0.46      100
weighted avg    0.71     0.84     0.77      100

[[ 0 84]
 [ 0 16]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.3 (same for any other col)

```
The number of records in the training dataset is 350
The number of records in the test dataset is 150
The training dataset has 291 records for the majority class and 59 records for the minority class.

precision    recall   f1-score   support
False        0.00     0.00     0.00     132
True         0.12     1.00     0.21      18

accuracy          0.12      150
macro avg       0.06     0.50     0.11      150
weighted avg    0.01     0.12     0.03      150

The customized score threshold for 2% of outliers is 0.00
precision    recall   f1-score   support
False        0.88     1.00     0.94     132
True         0.00     0.00     0.00      18

accuracy          0.88      150
macro avg       0.44     0.50     0.47      150
weighted avg    0.77     0.88     0.82      150

[[ 0 132]
 [ 0 18]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.5 (same for any other col)

```

The number of records in the training dataset is 250
The number of records in the test dataset is 250
The training dataset has 206 records for the majority class and 44 records for the minority class.

      precision    recall   f1-score   support

  False       0.00     0.00     0.00      217
  True       0.13     1.00     0.23      33

   accuracy          0.13      250
macro avg       0.07     0.50     0.12      250
weighted avg    0.02     0.13     0.03      250

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support

  False       0.87     1.00     0.93      217
  True       0.00     0.00     0.00      33

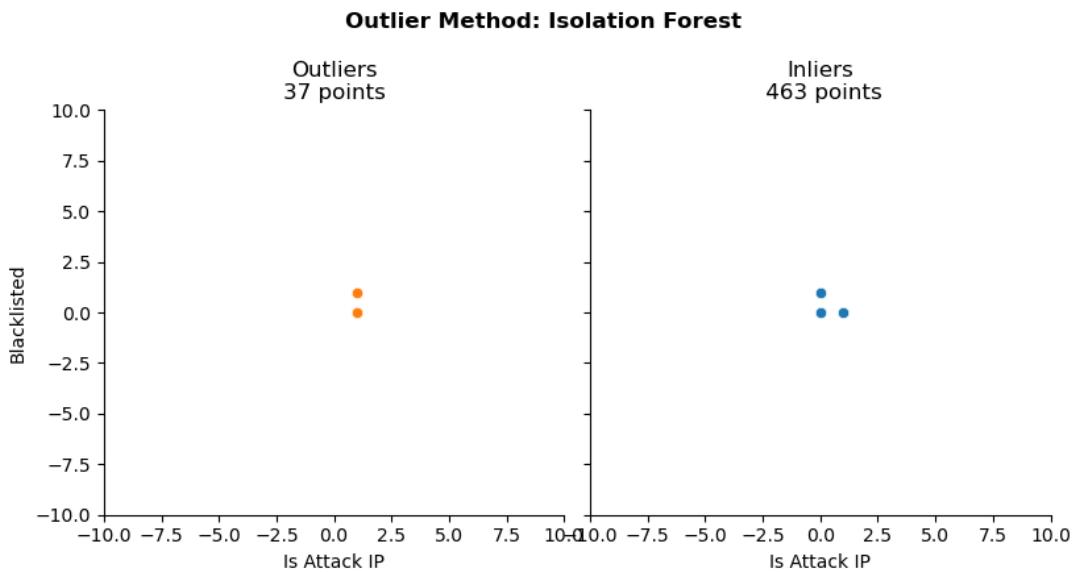
   accuracy          0.87      250
macro avg       0.43     0.50     0.46      250
weighted avg    0.75     0.87     0.81      250

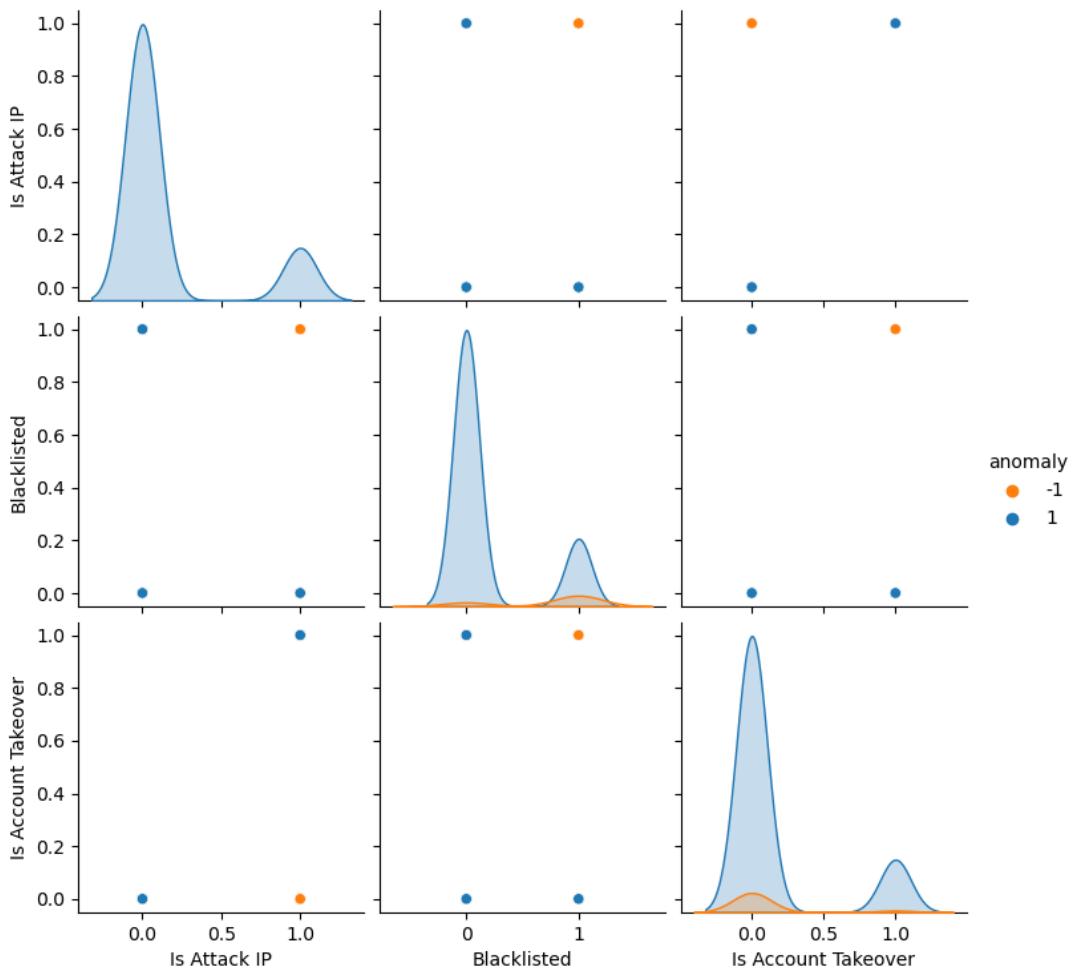
[[ 0 217]
 [ 0 33]]

```

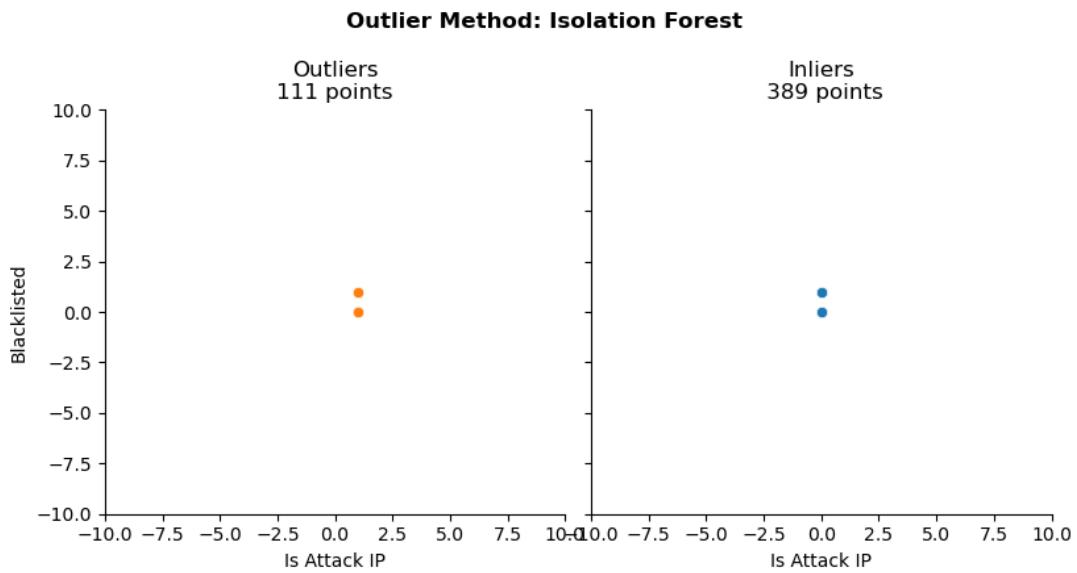
2C. Isolation Forest

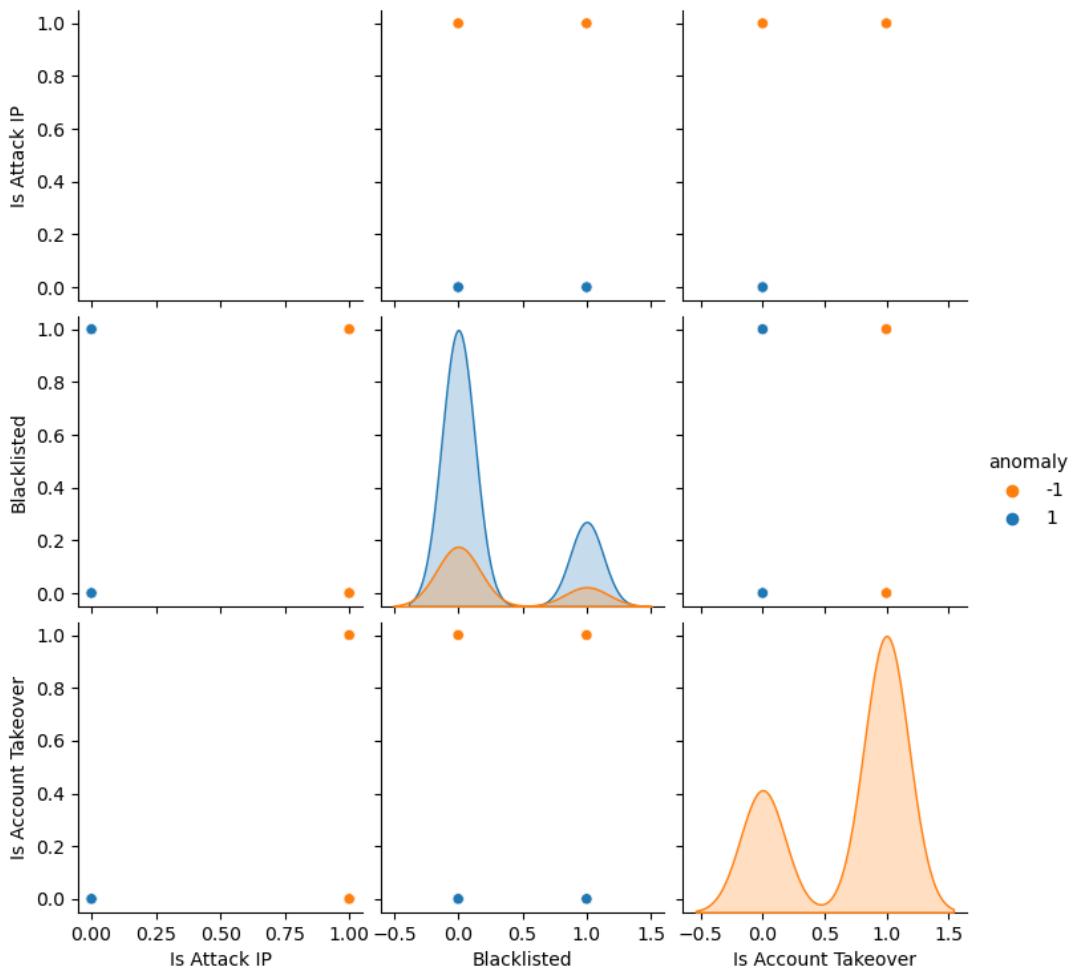
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



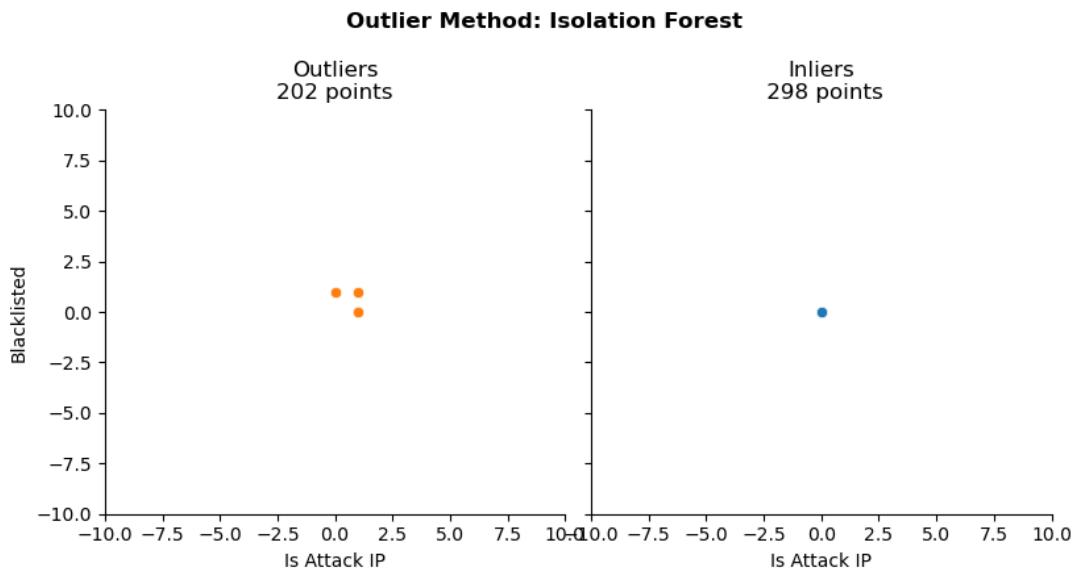


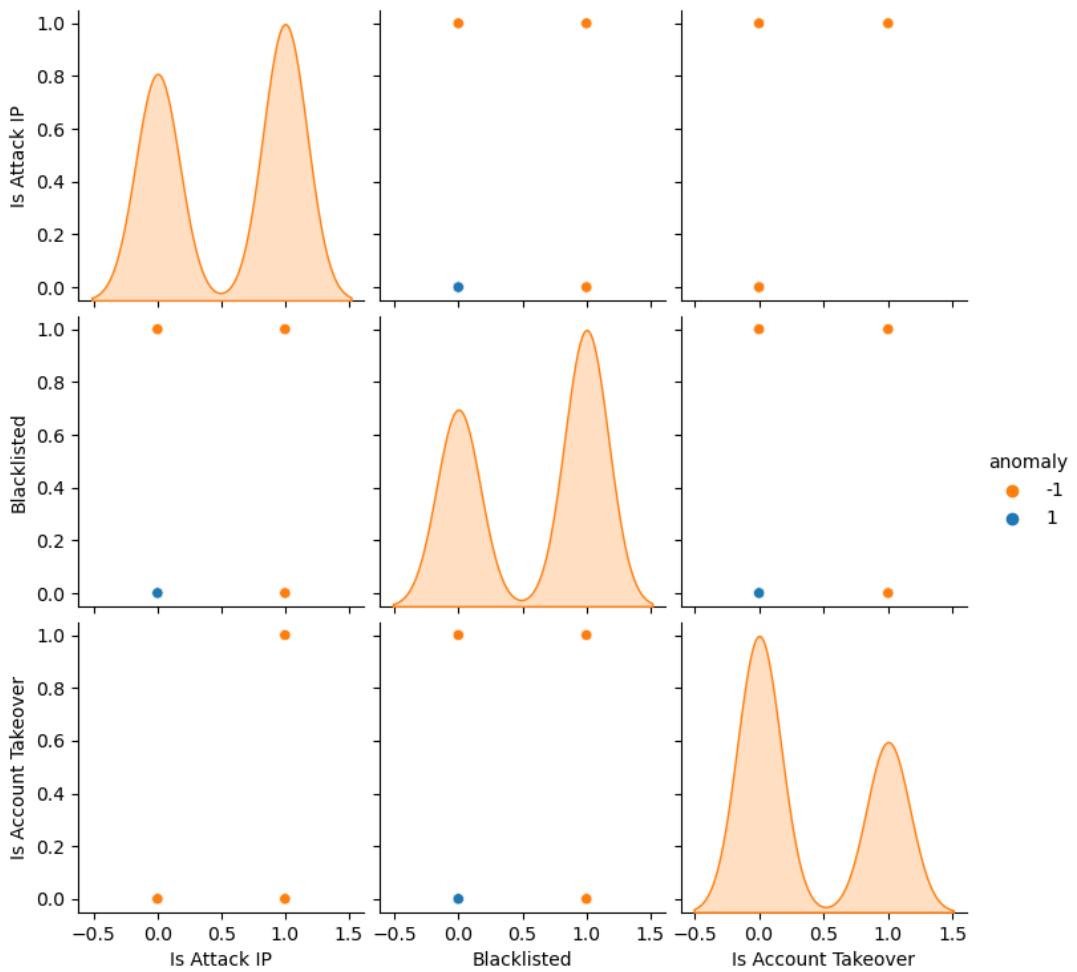
- increasing contamination value to 0.3



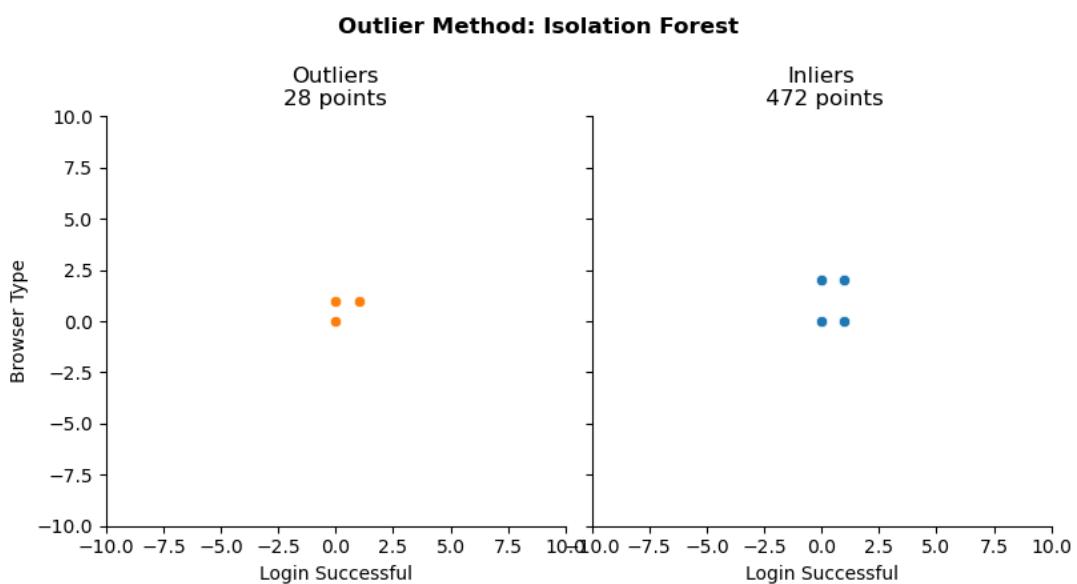


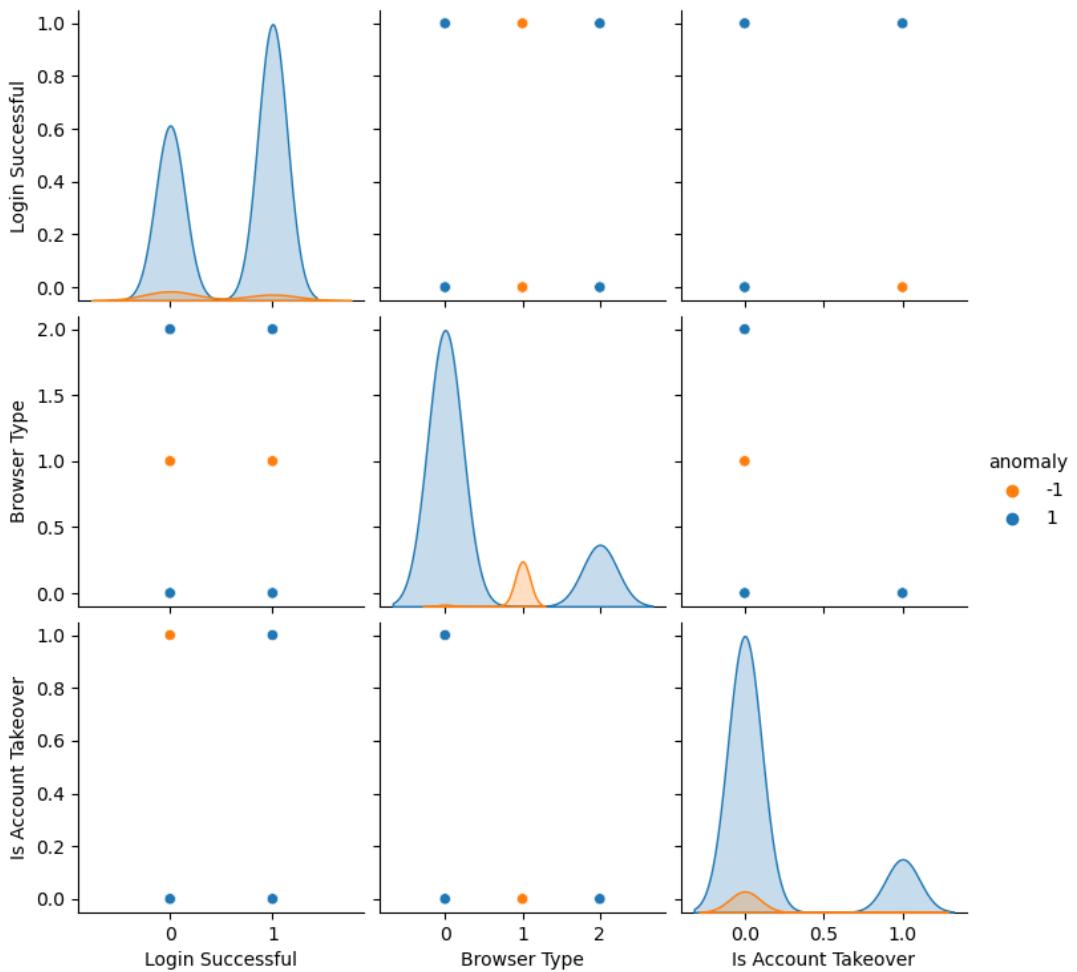
- contamination value == 0.5



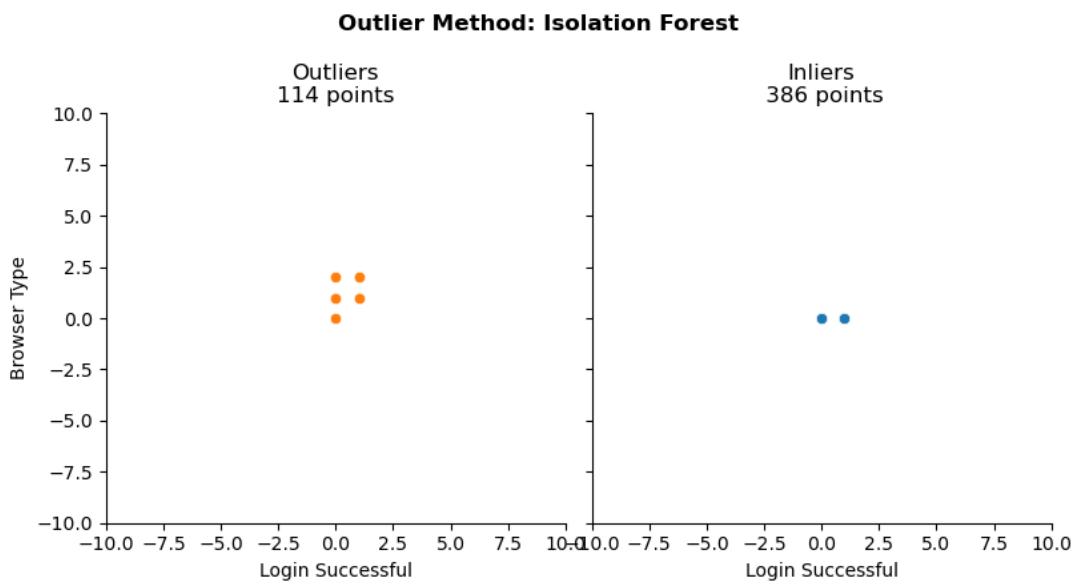


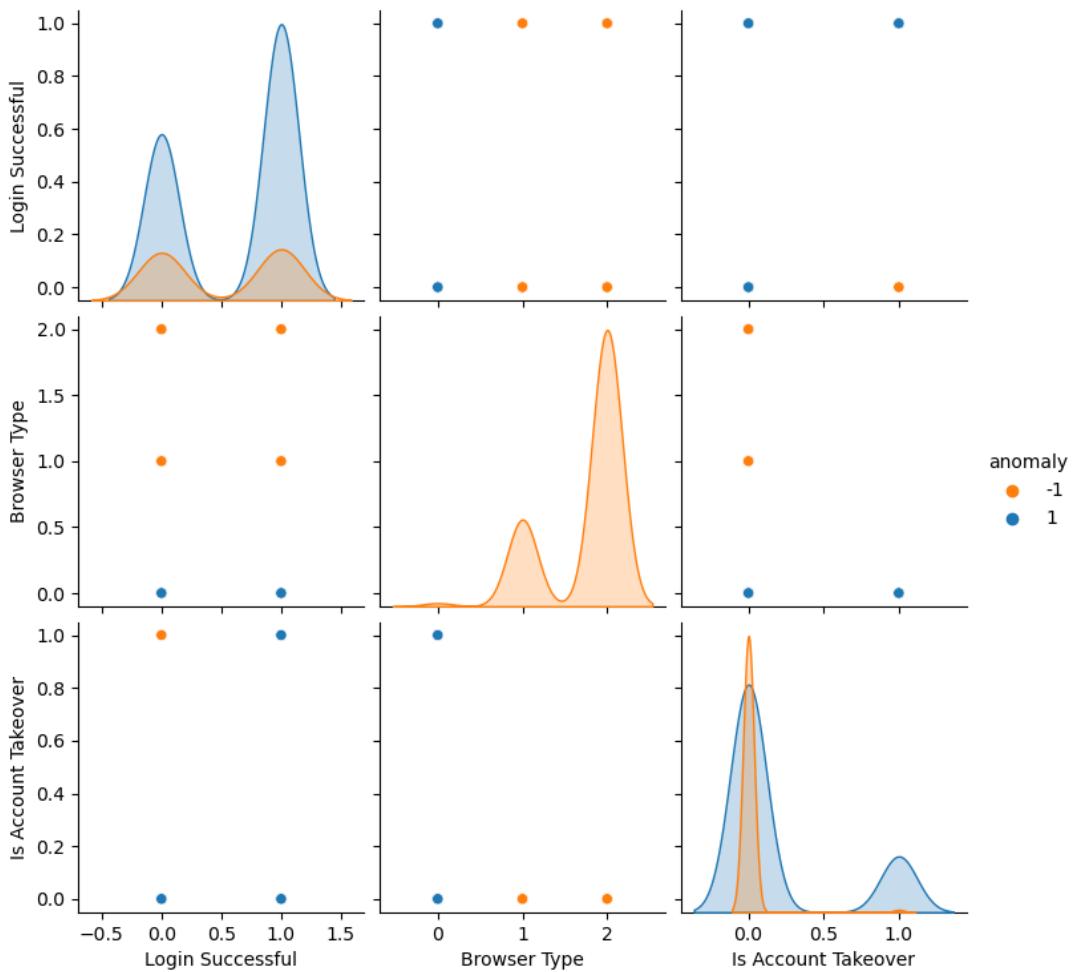
- anomaly_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']
 - contamination value == 0.1



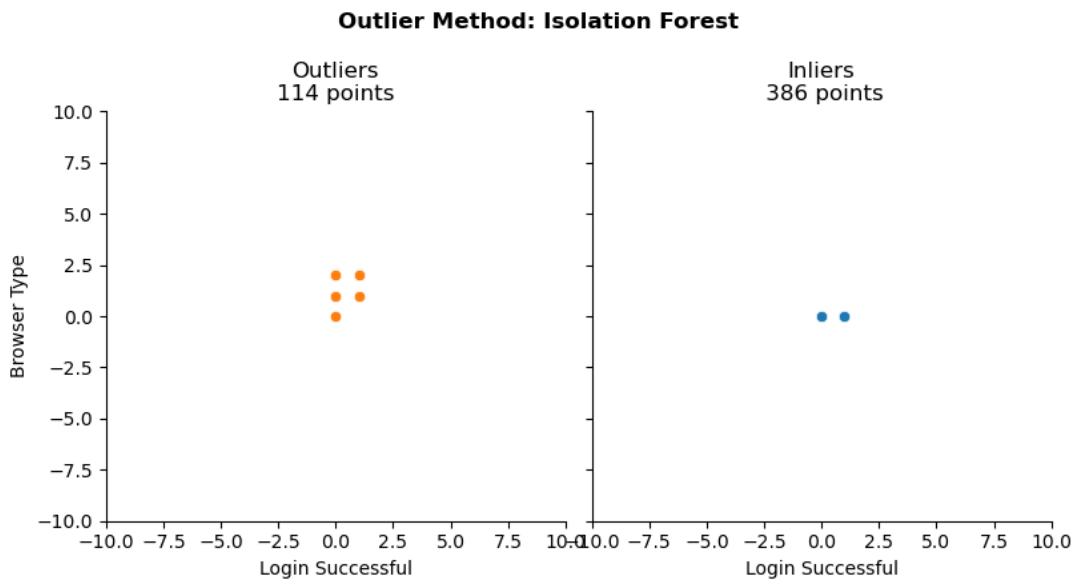


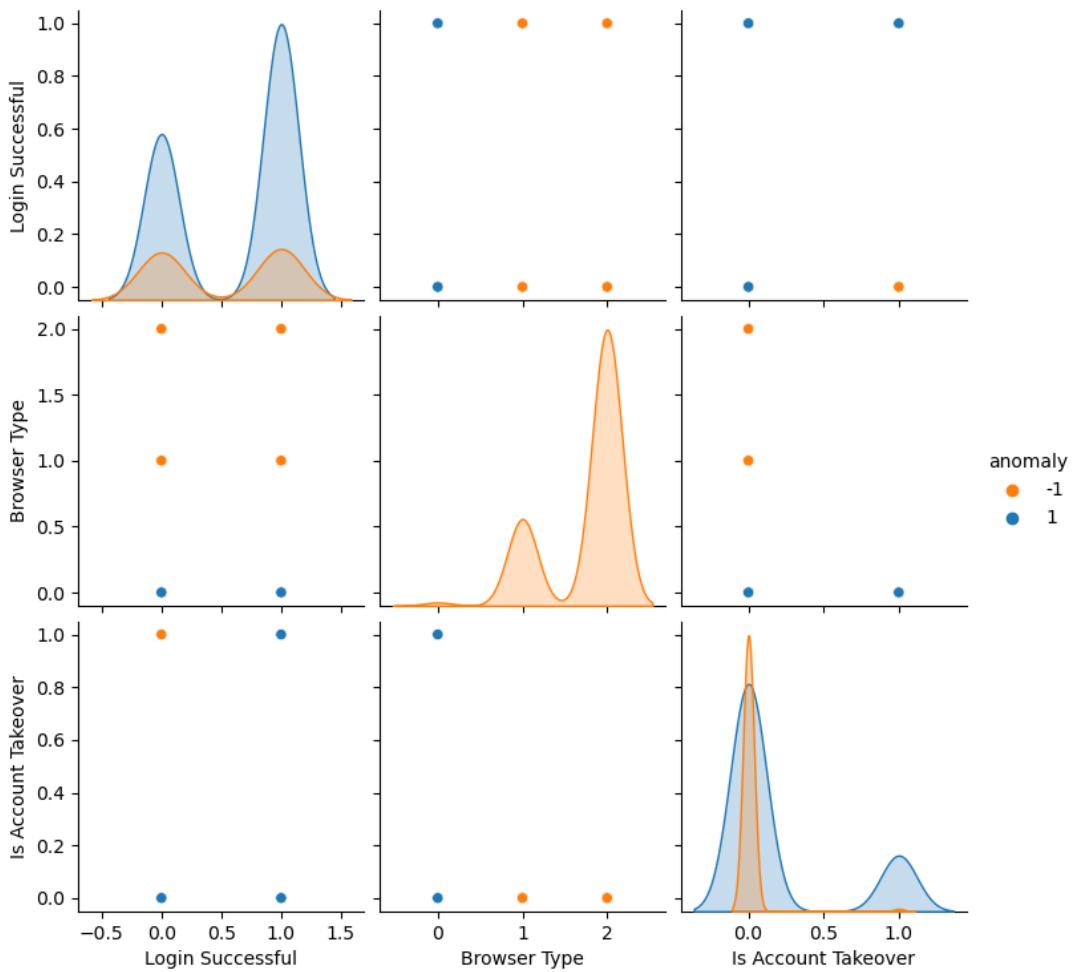
- contamination value == 0.3



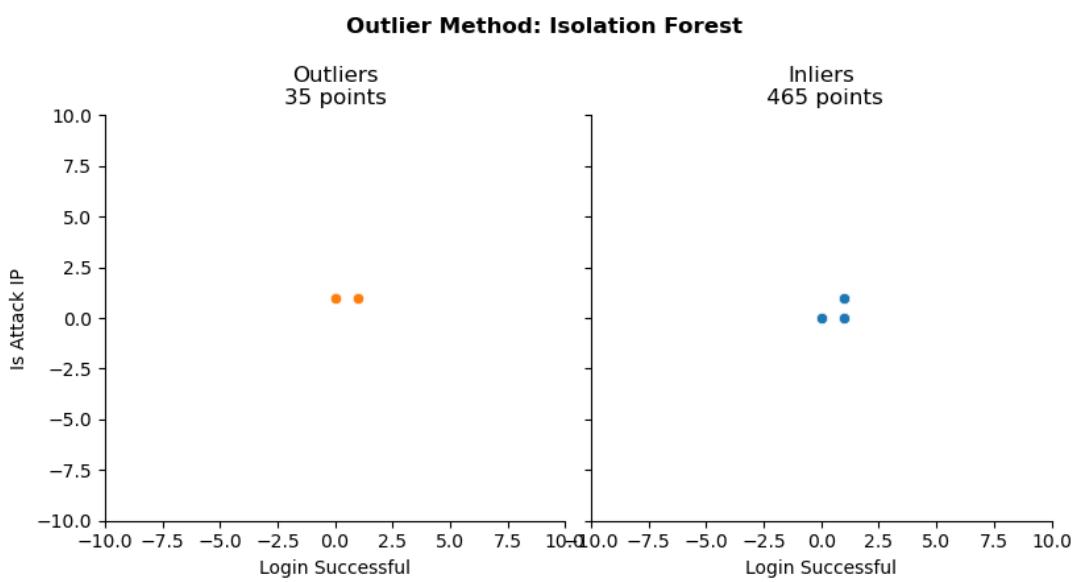


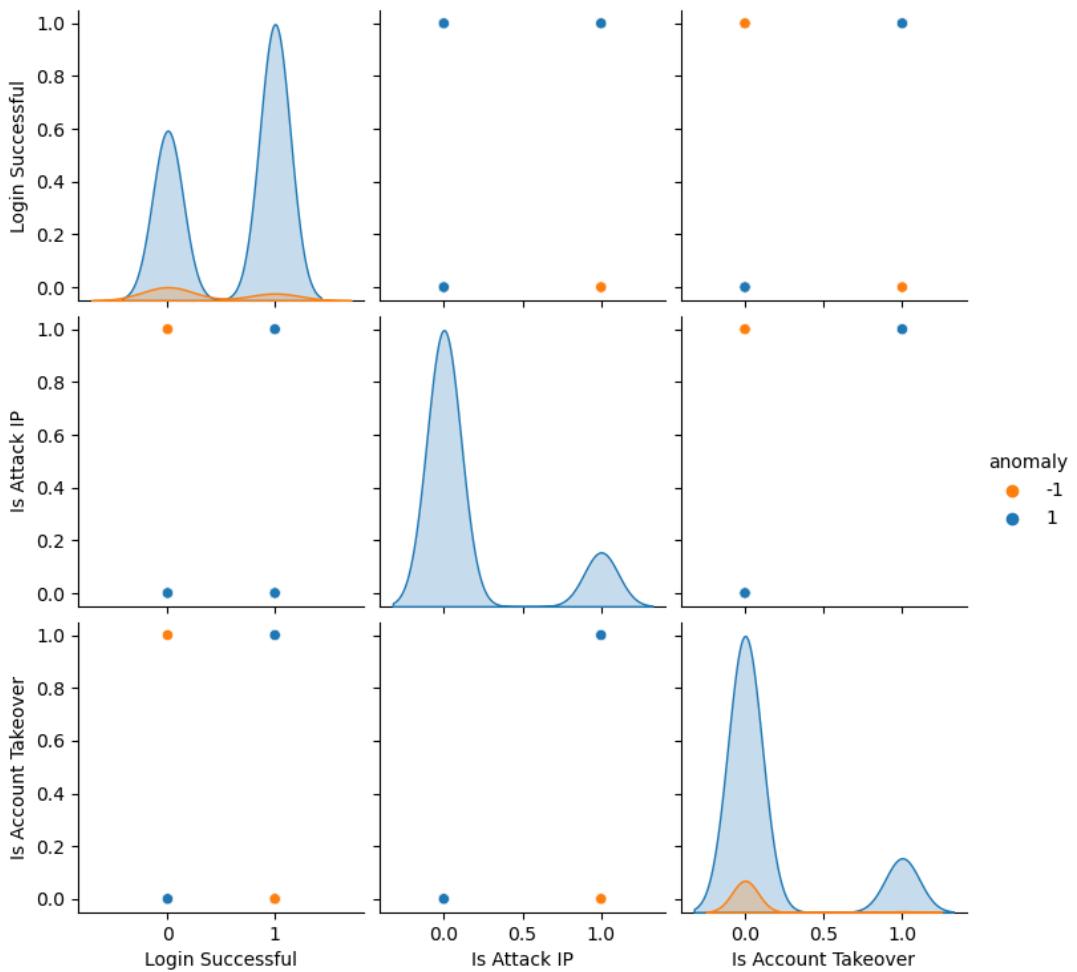
- contamination value == 0.5





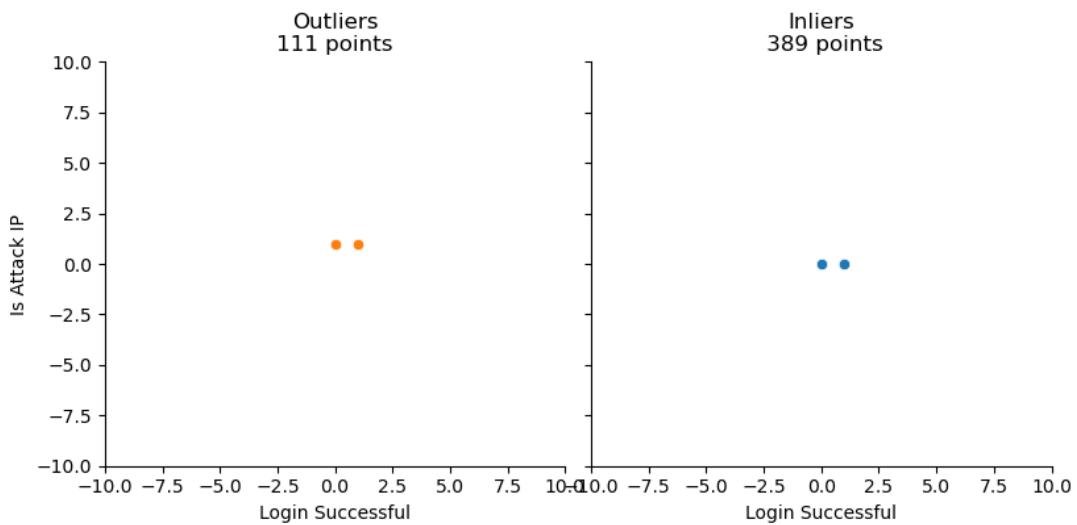
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1

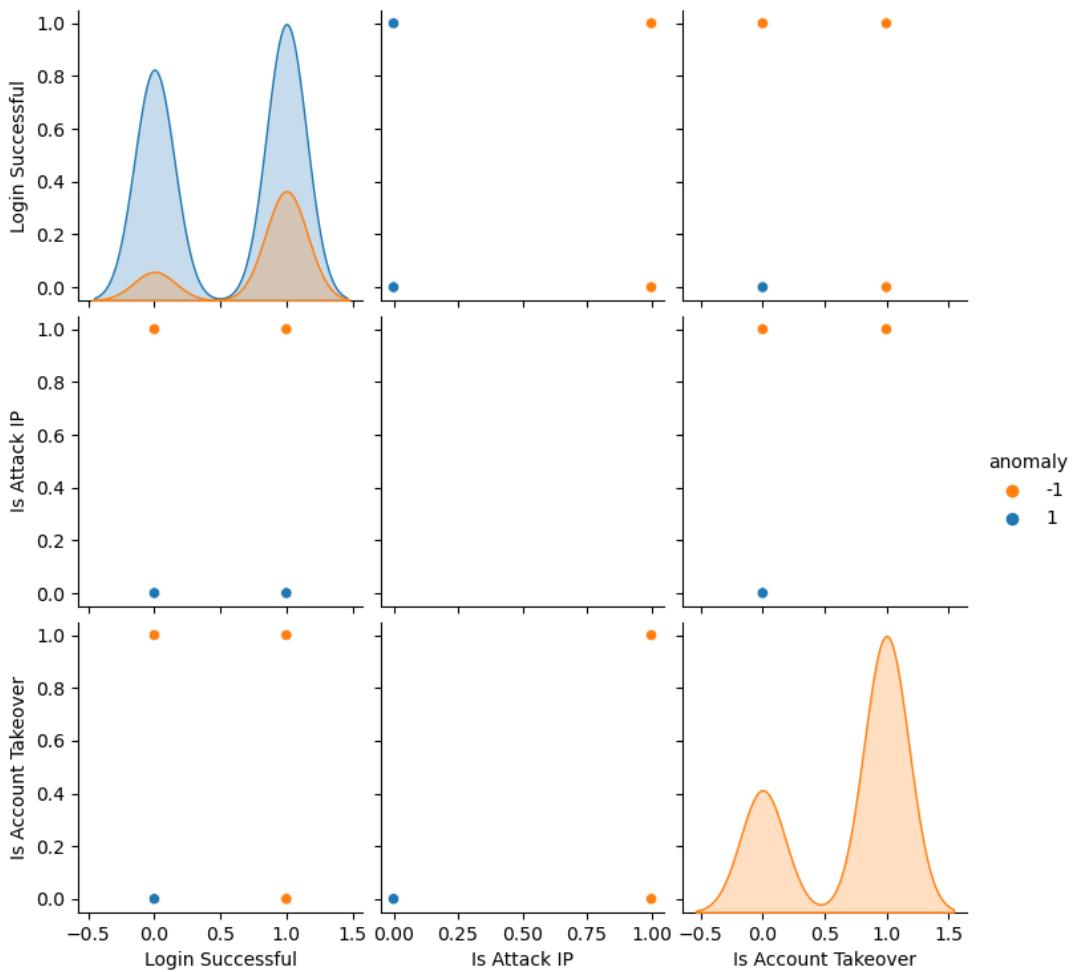




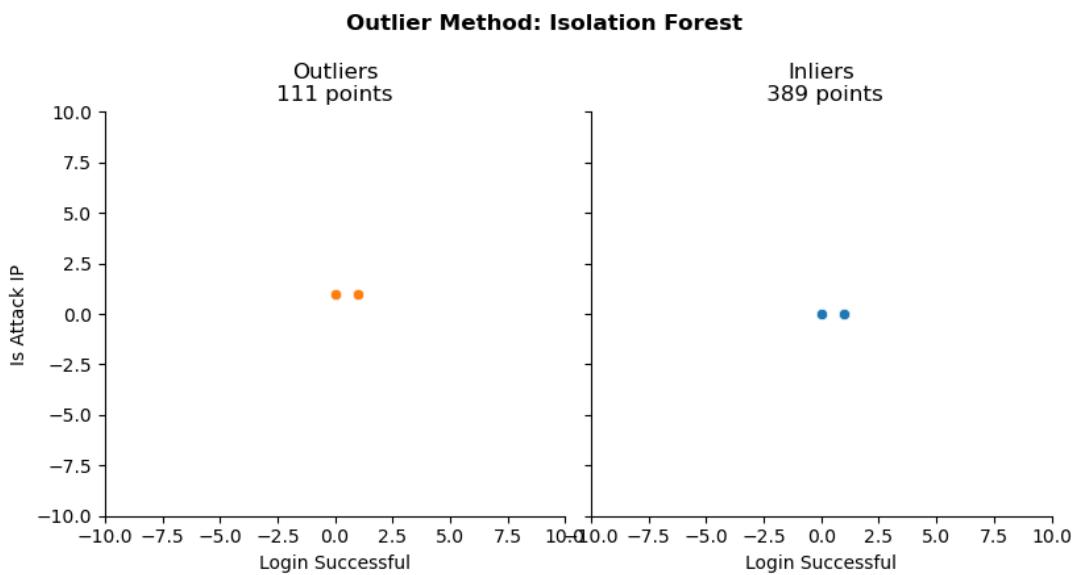
- contamination value == 0.3

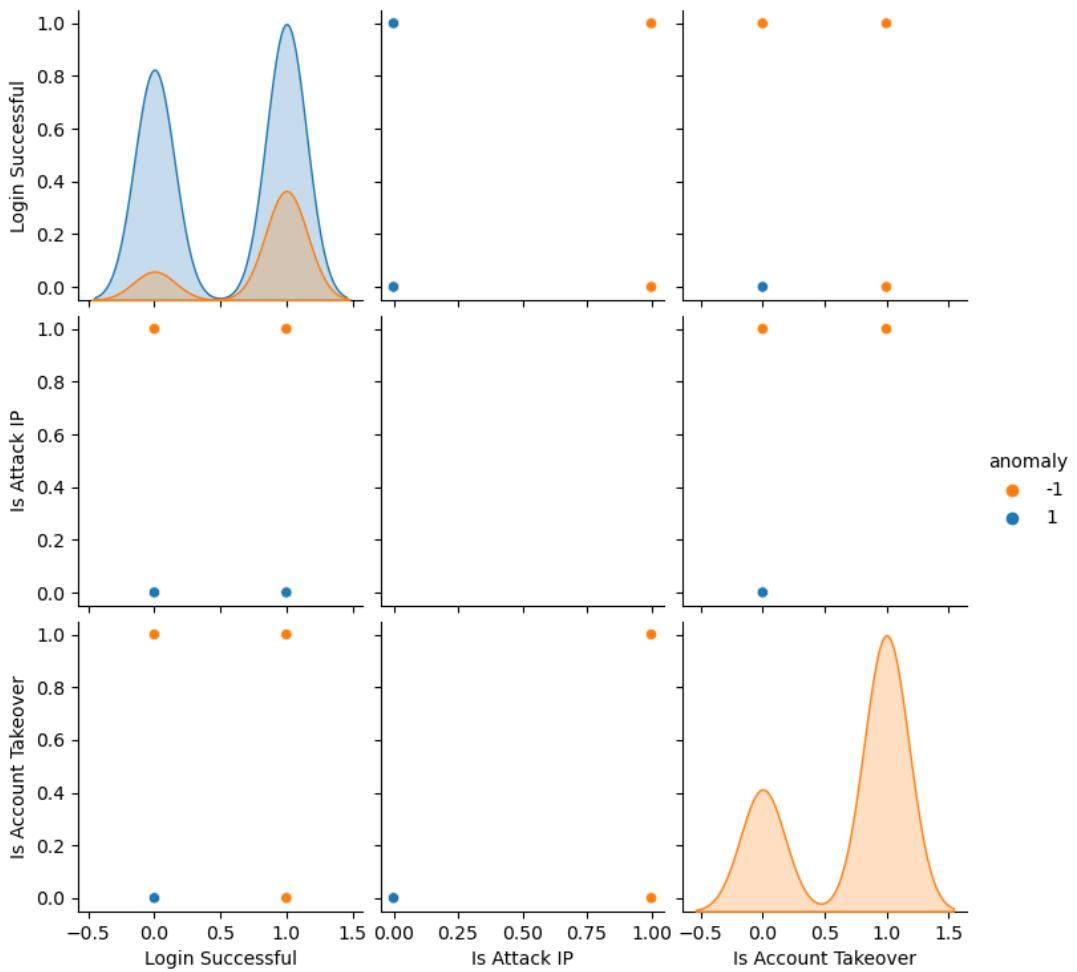
Outlier Method: Isolation Forest



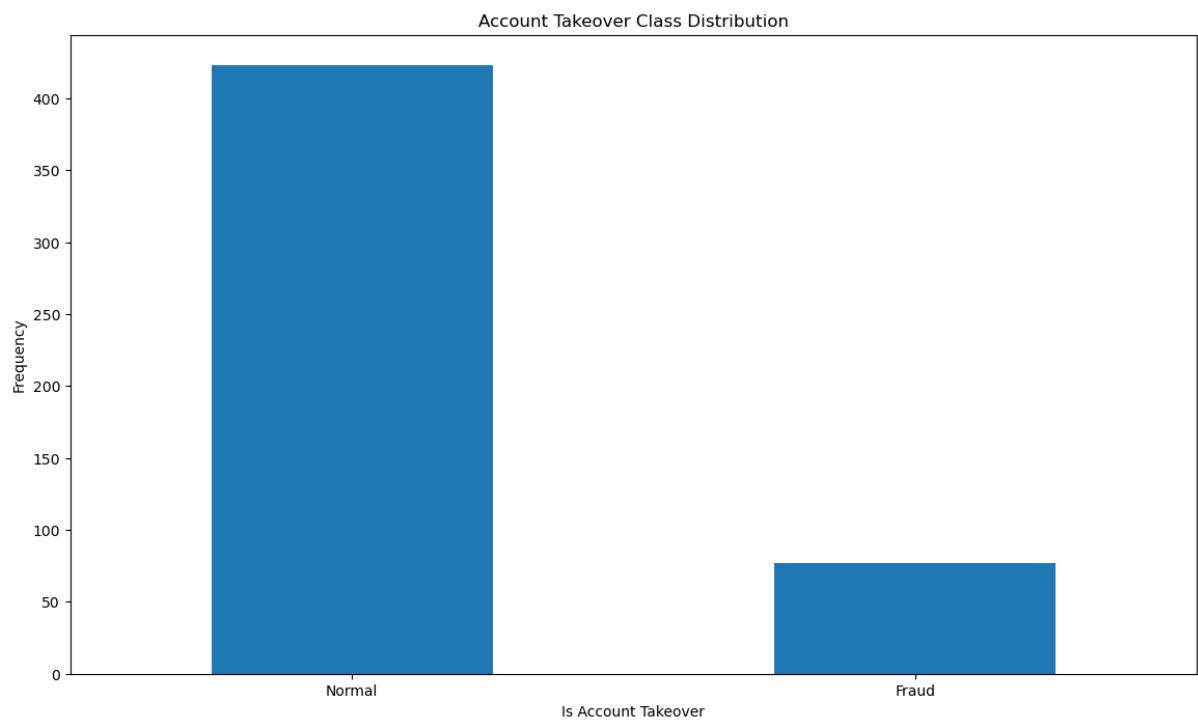


- contamination value == 0.5





2D. iForest, Local Outlier Factor, Support Vector Machine



"{}: {}".format(clf_name,n_errors) → np. Isolation Forest: 21

Isolation Forest: 113

Accuracy Score :

0.774

Classification Report :

	precision	recall	f1-score	support
False	0.84	0.90	0.87	423
True	0.14	0.09	0.11	77
accuracy			0.77	500
macro avg	0.49	0.49	0.49	500
weighted avg	0.74	0.77	0.75	500

Local Outlier Factor: 117

Accuracy Score :

0.766

Classification Report :

	precision	recall	f1-score	support
False	0.84	0.89	0.87	423
True	0.10	0.06	0.08	77
accuracy			0.77	500
macro avg	0.47	0.48	0.47	500
weighted avg	0.73	0.77	0.74	500

Support Vector Machine: 343

Accuracy Score :

0.314

Classification Report :

	precision	recall	f1-score	support
False	0.79	0.26	0.39	423
True	0.13	0.61	0.22	77
accuracy			0.31	500
macro avg	0.46	0.44	0.30	500
weighted avg	0.68	0.31	0.36	500

3. Dummy Variable Encoding → df3 =
pd.get_dummies(df3, columns=['Country', 'Device

Type', 'IP Address', 'Login Timestamp', 'Browser Type'], prefix=["cntr", "dvc", "ip", "Ignt", "brw"])

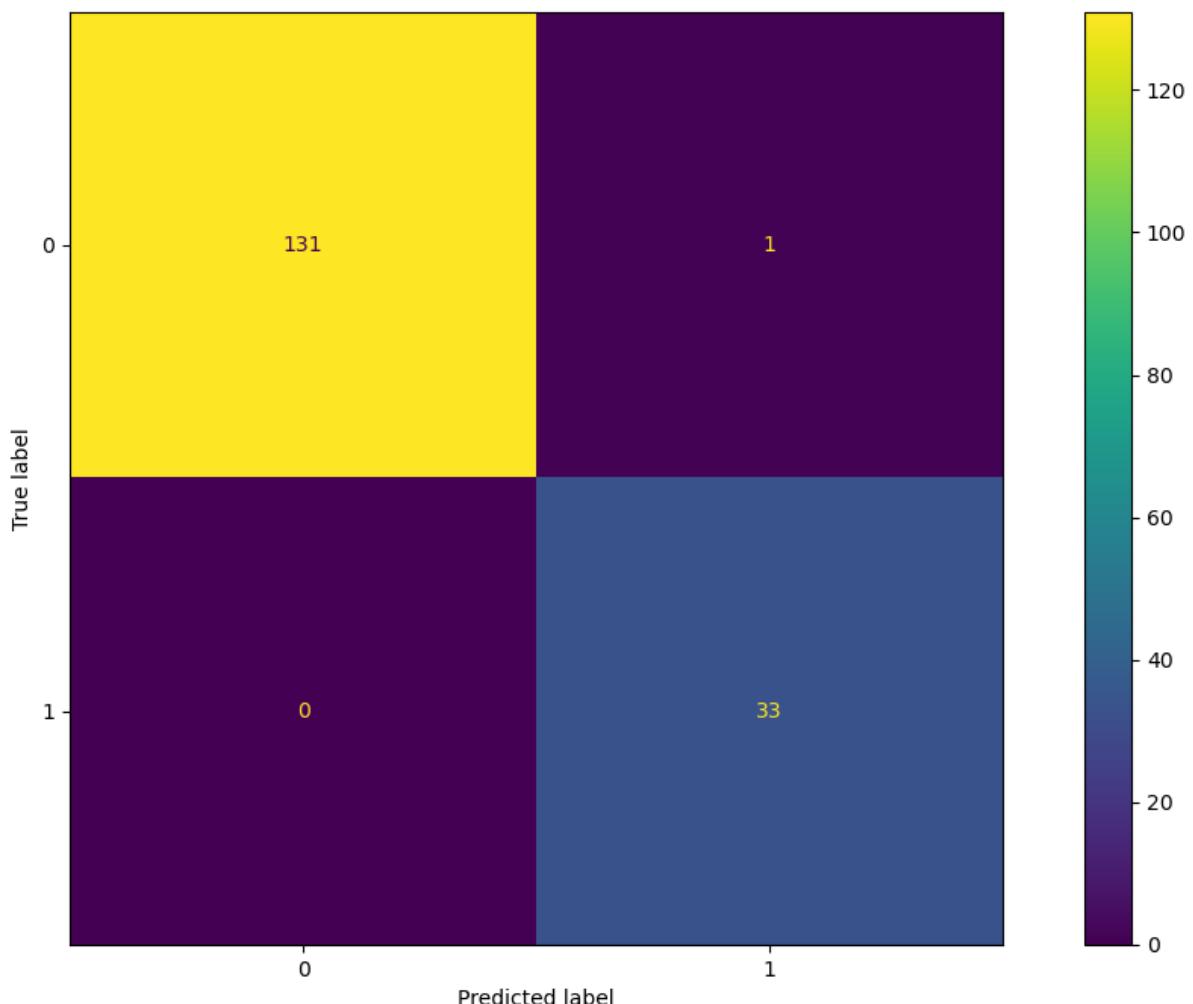
3A. XGBoost

memory usage: 466.9 KB

Accuracy: 99.39%

	precision	recall	f1-score	support
False	1.00	0.99	1.00	132
True	0.97	1.00	0.99	33
accuracy			0.99	165
macro avg	0.99	1.00	0.99	165
weighted avg	0.99	0.99	0.99	165

TN: 0, FP: 2, FN: 1, TP: 1



3B. One Class SVM

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.2 (same for any other col)

```
The number of records in the training dataset is 400
The number of records in the test dataset is 100
The training dataset has 339 records for the majority class and 61 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      84
      True        0.16     1.00     0.28      16

      accuracy          0.16      100
      macro avg       0.08     0.50     0.14      100
  weighted avg     0.03     0.16     0.04      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.84     1.00     0.91      84
      True        0.00     0.00     0.00      16

      accuracy          0.84      100
      macro avg       0.42     0.50     0.46      100
  weighted avg     0.71     0.84     0.77      100

[[ 0 84]
 [ 0 16]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.3 (same for any other col)

```
The number of records in the training dataset is 350
The number of records in the test dataset is 150
The training dataset has 291 records for the majority class and 59 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00     132
      True        0.12     1.00     0.21      18

      accuracy          0.12      150
      macro avg       0.06     0.50     0.11      150
  weighted avg     0.01     0.12     0.03      150

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.88     1.00     0.94     132
      True        0.00     0.00     0.00      18

      accuracy          0.88      150
      macro avg       0.44     0.50     0.47      150
  weighted avg     0.77     0.88     0.82      150

[[ 0 132]
 [ 0 18]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.5 (same for any other col)

```

The number of records in the training dataset is 250
The number of records in the test dataset is 250
The training dataset has 206 records for the majority class and 44 records for the minority class.
      precision    recall   f1-score   support
False        0.91     0.05     0.09     217
True         0.13     0.97     0.24      33

accuracy                           0.17      250
macro avg       0.52     0.51     0.16      250
weighted avg    0.81     0.17     0.11      250

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.87     1.00     0.93     217
True         0.00     0.00     0.00      33

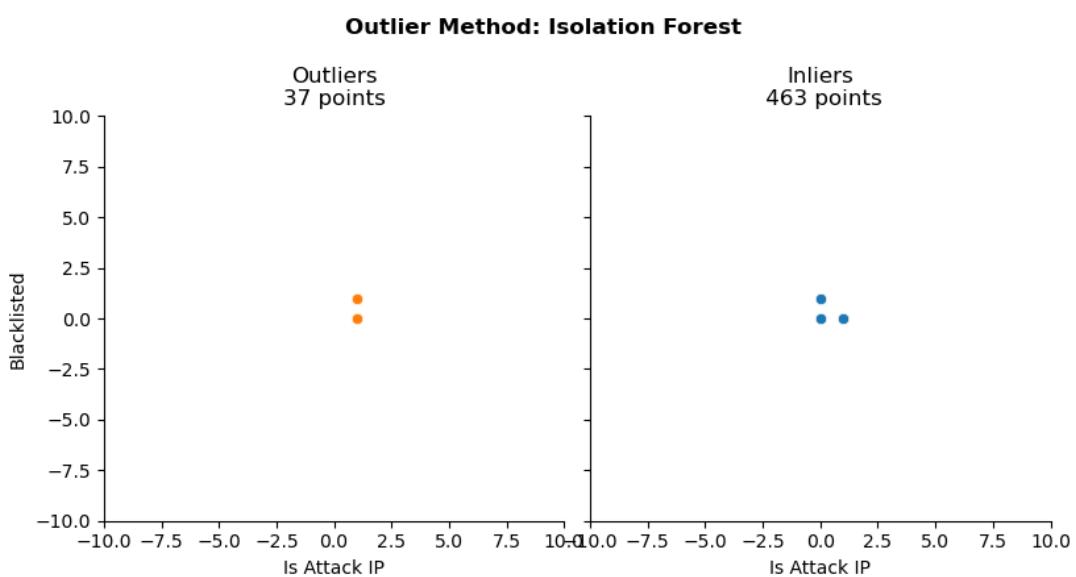
accuracy                           0.87      250
macro avg       0.43     0.50     0.46      250
weighted avg    0.75     0.87     0.81      250

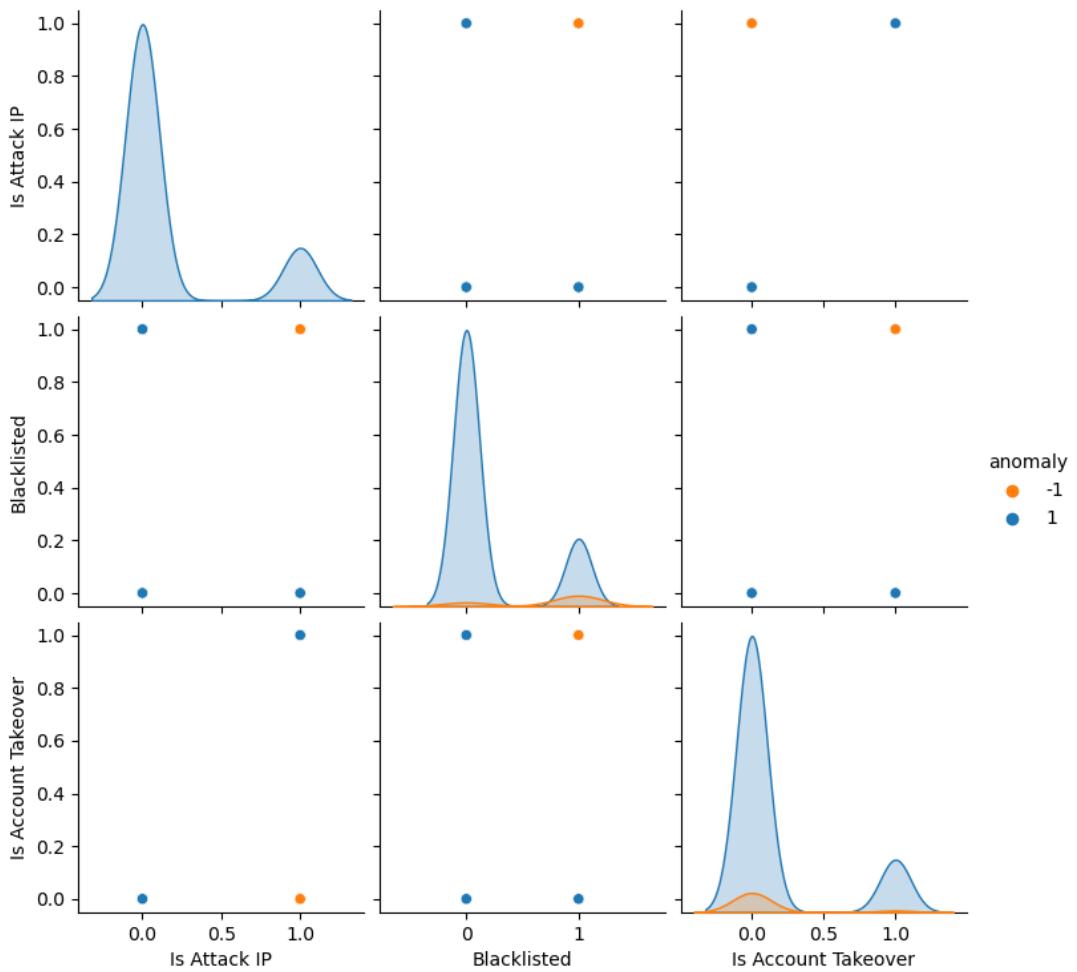
[[ 10 207]
 [ 1 32]]

```

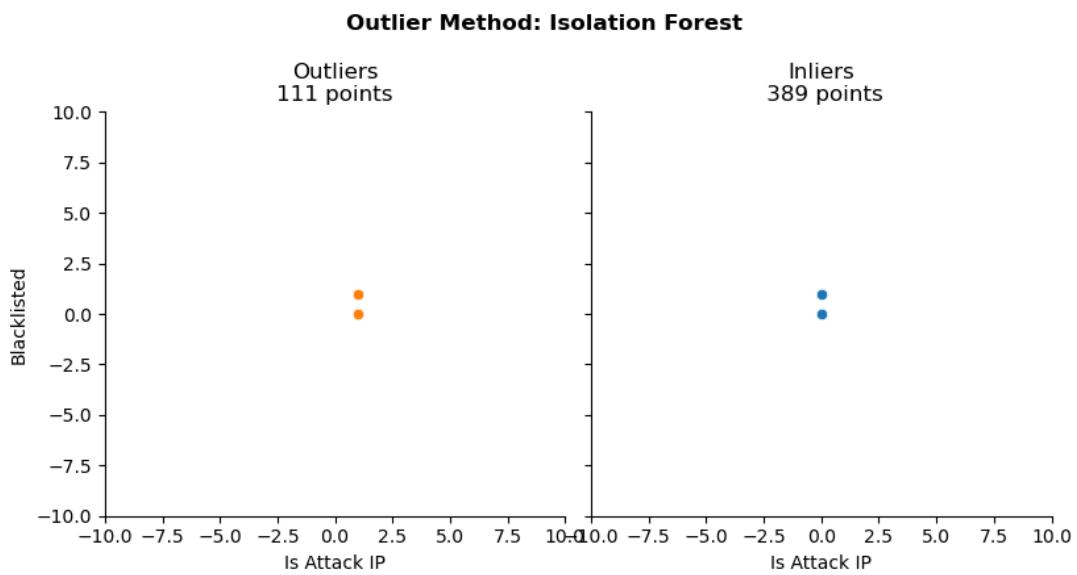
3C. Isolation Forest

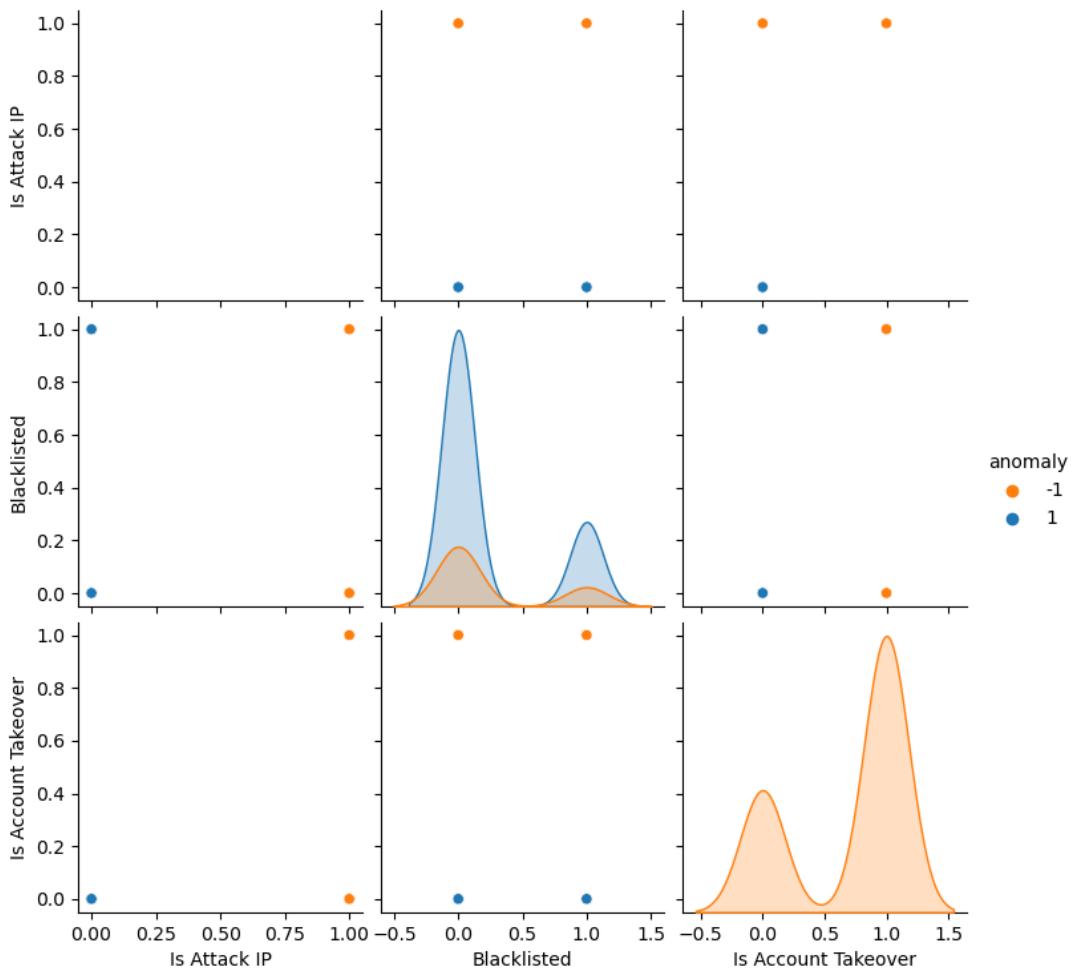
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



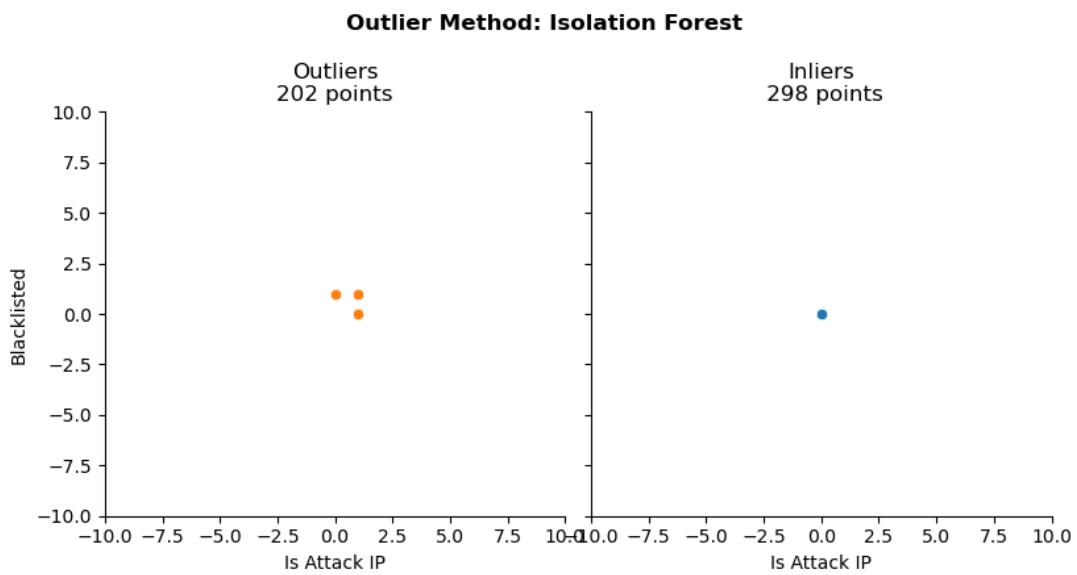


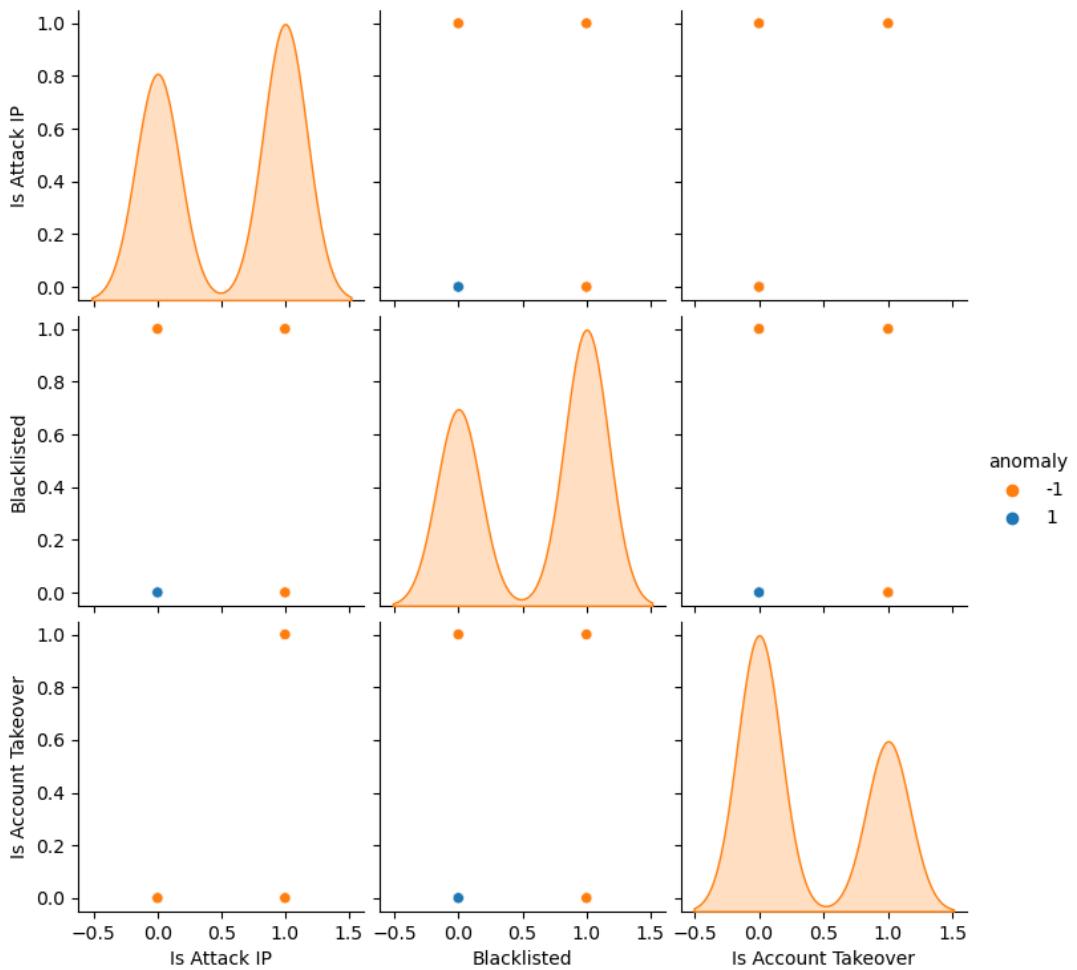
- contamination value == 0.3



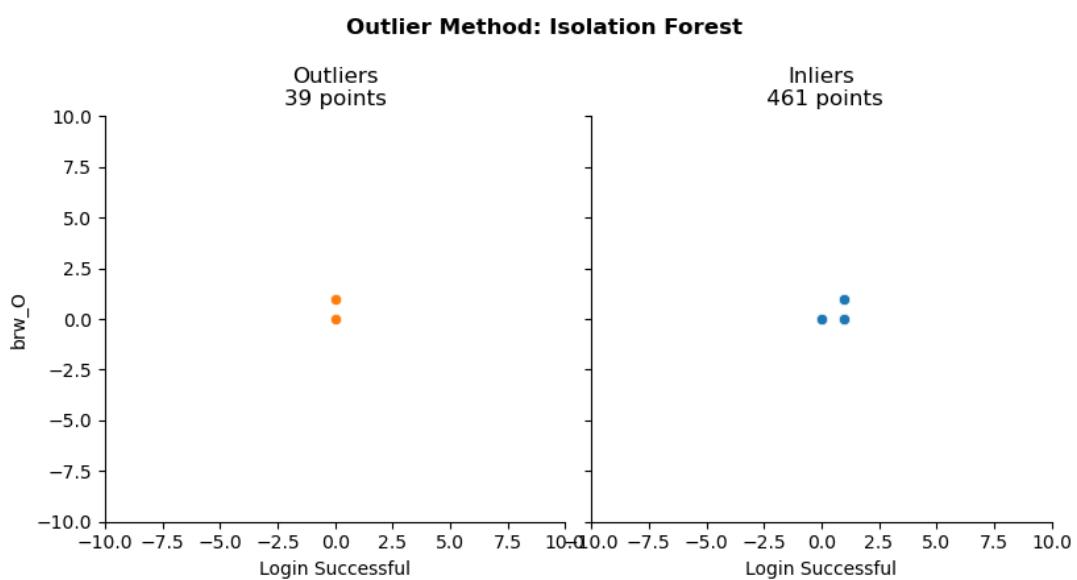


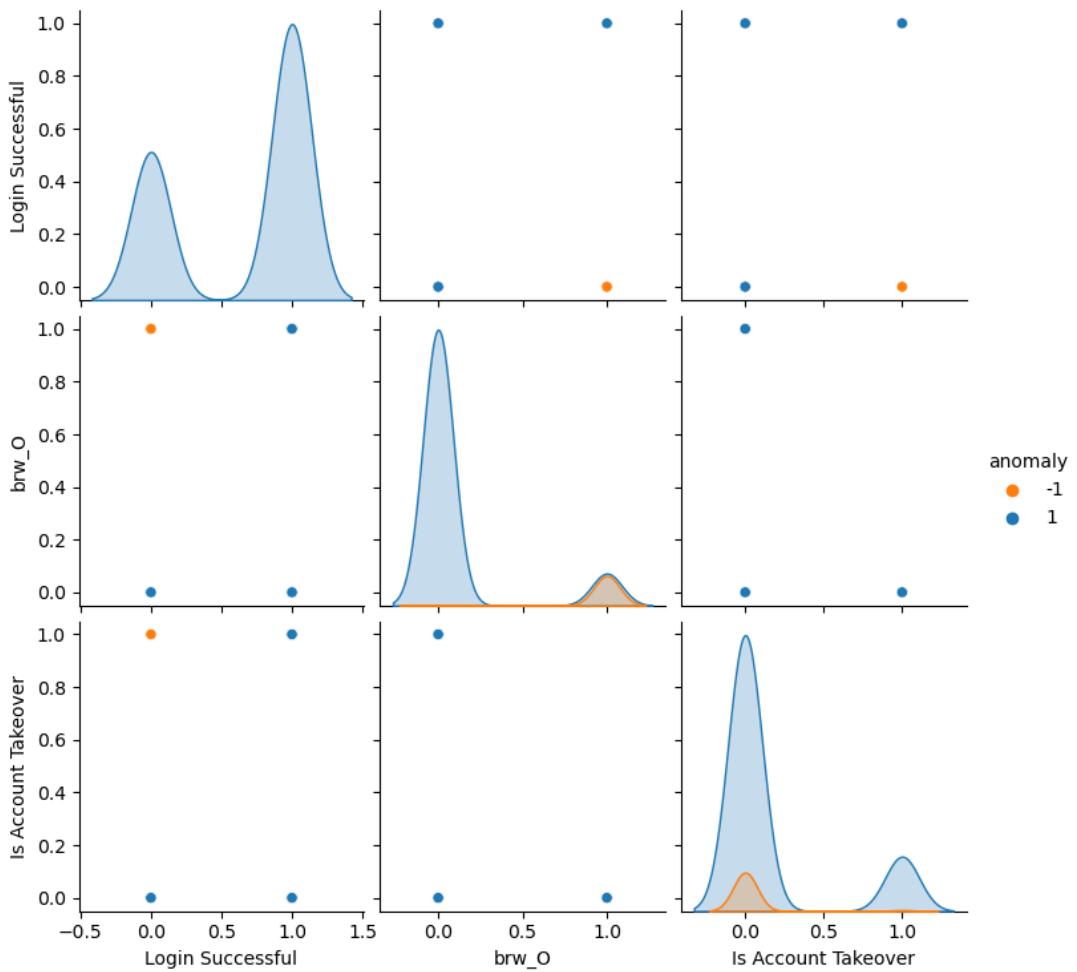
- contamination value == 0.5



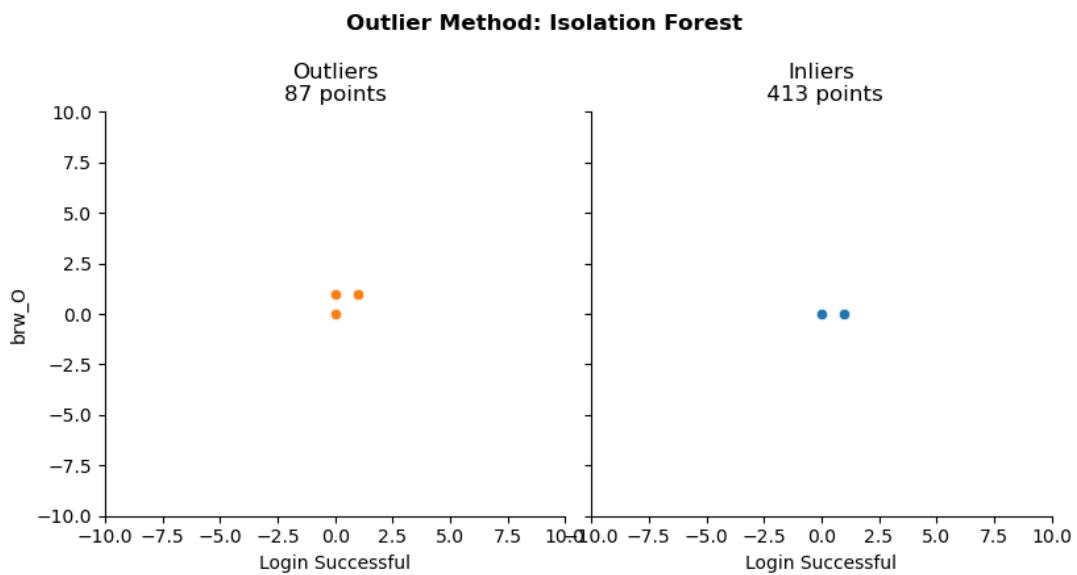


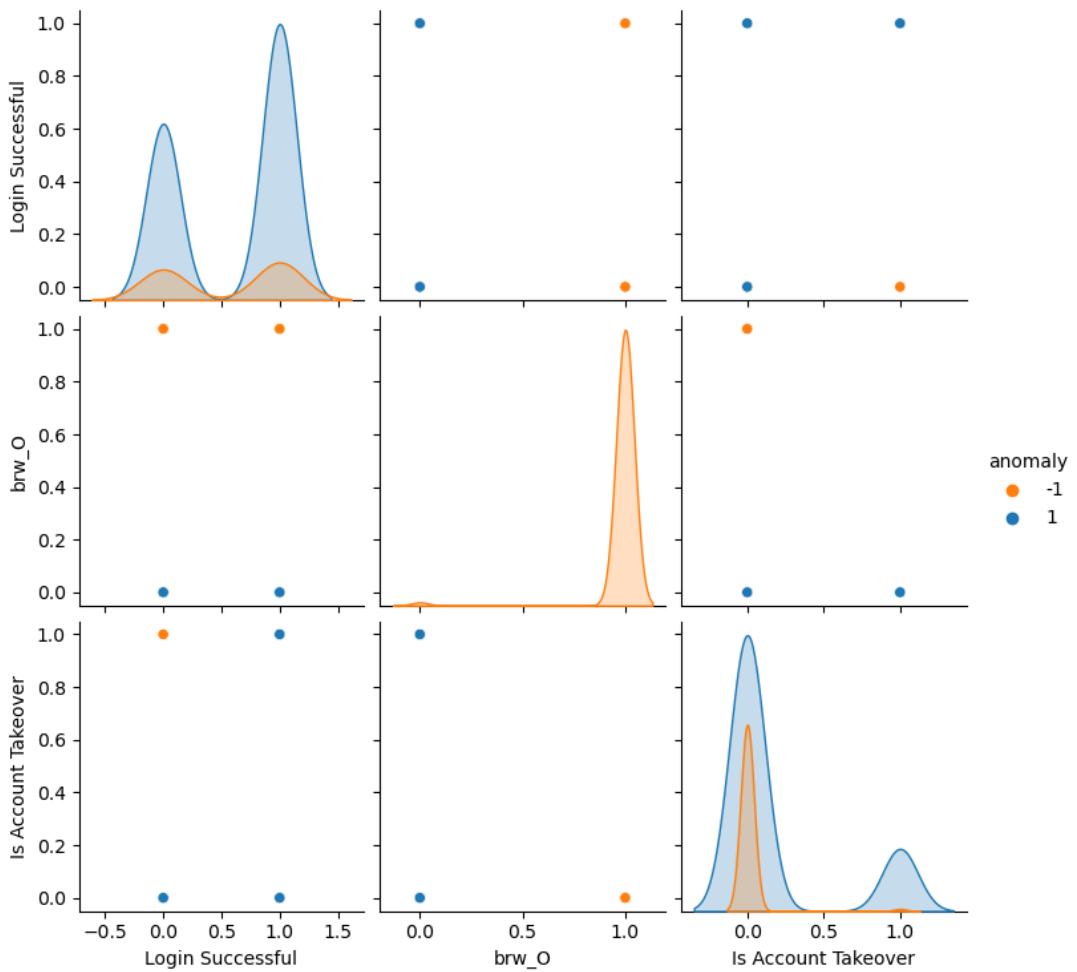
- anomaly_inputs = ['Login Successful', 'Browser Type (brw_O)', 'Is Account Takeover']
 - contamination value == 0.1





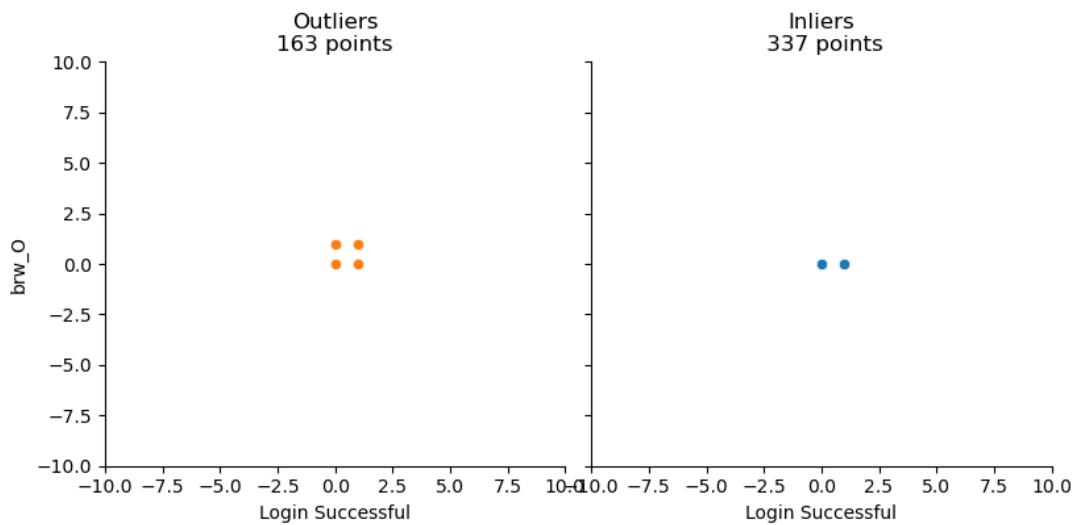
- contamination value == 0.3

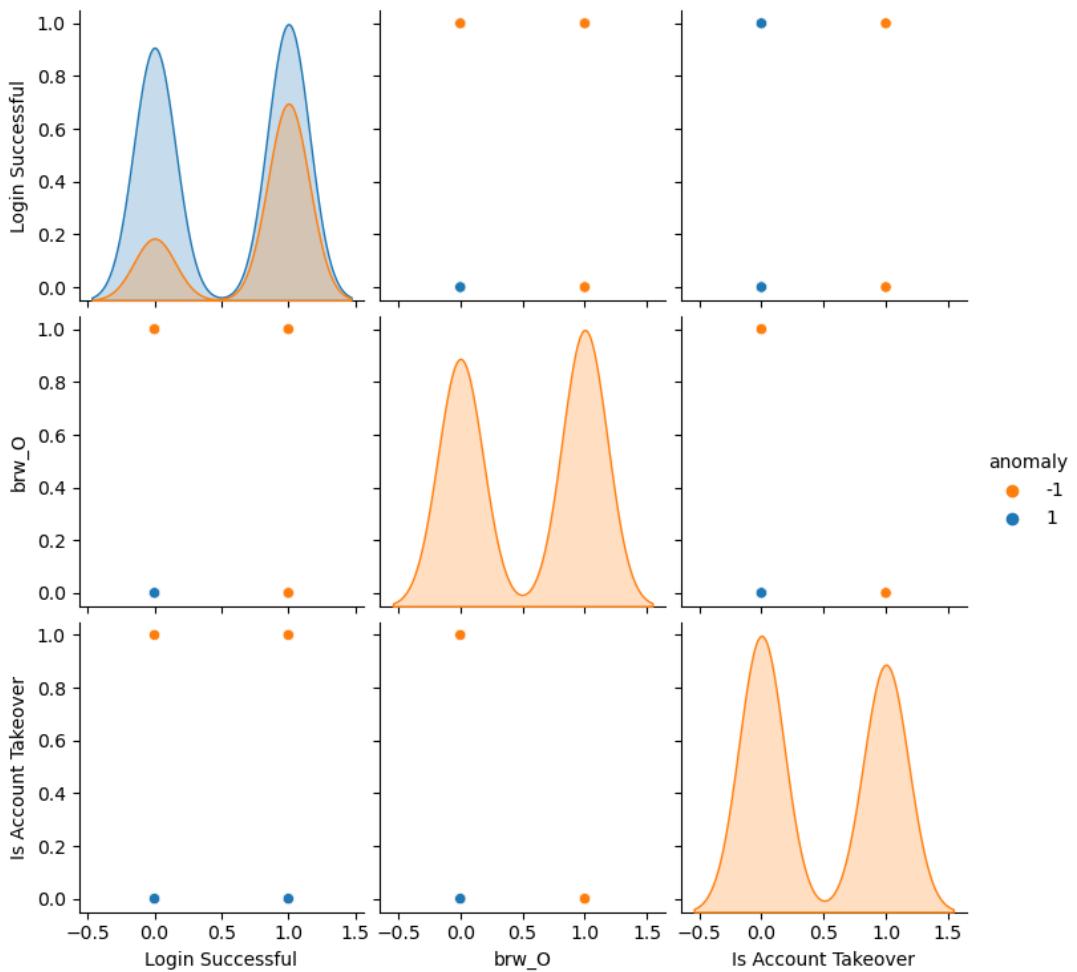




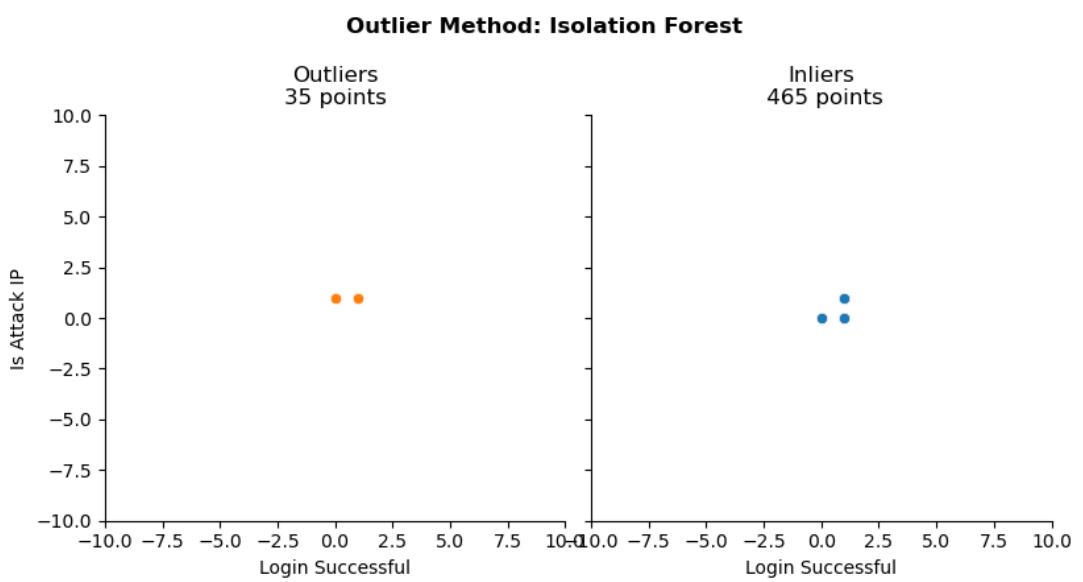
- contamination value == 0.5

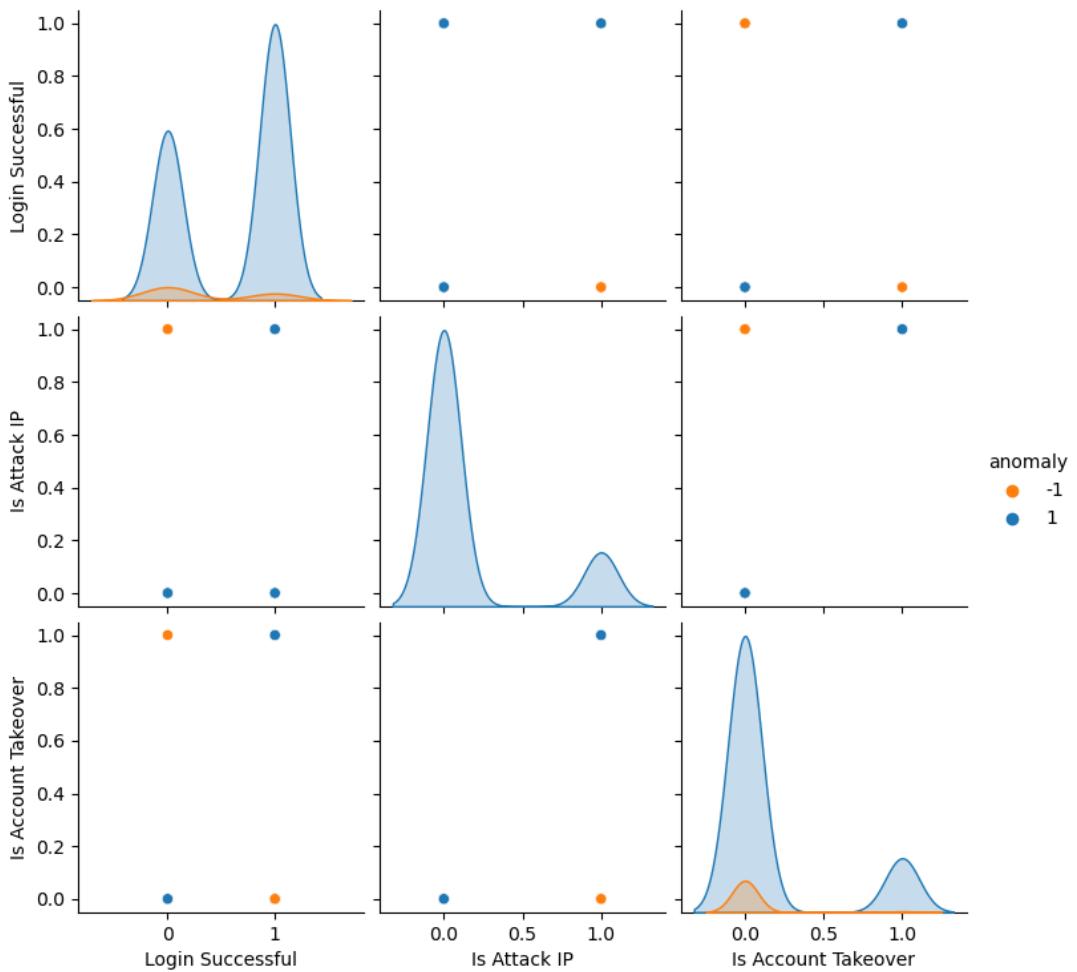
Outlier Method: Isolation Forest



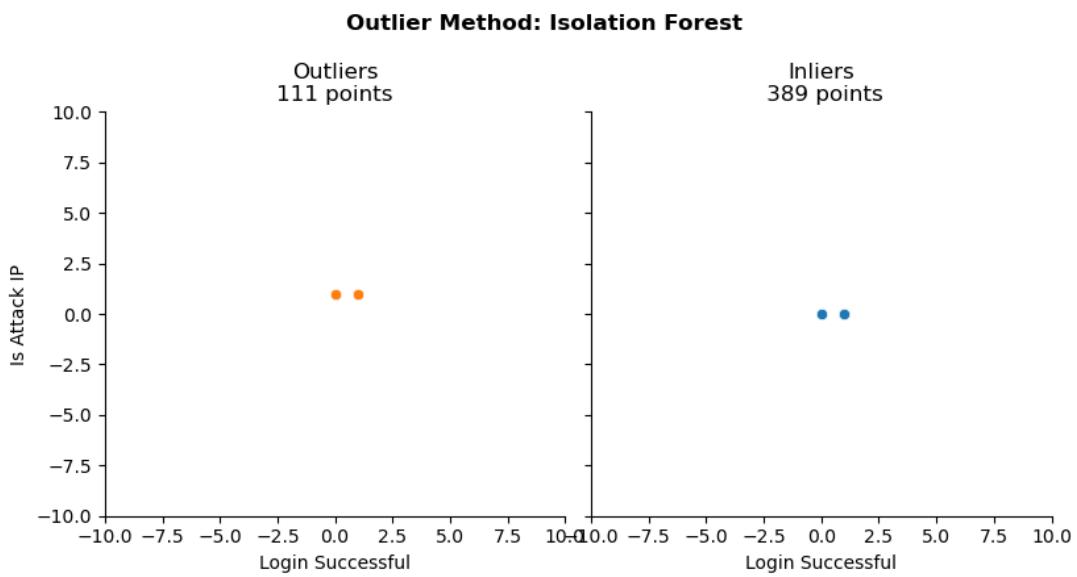


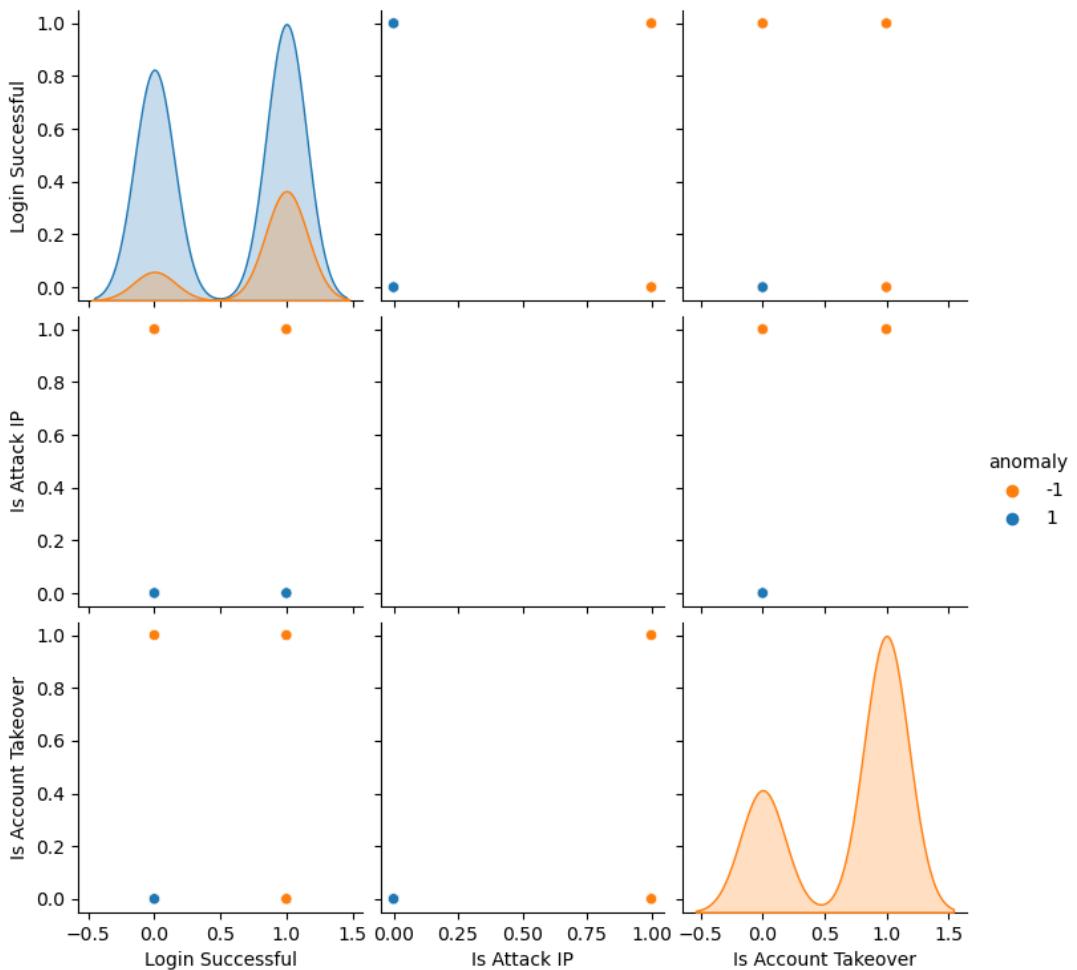
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1



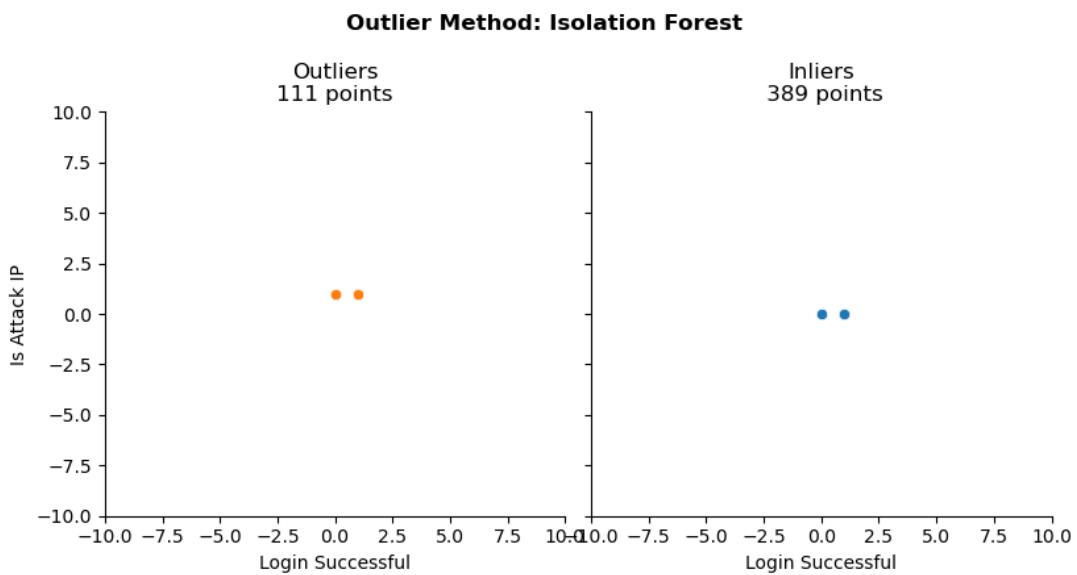


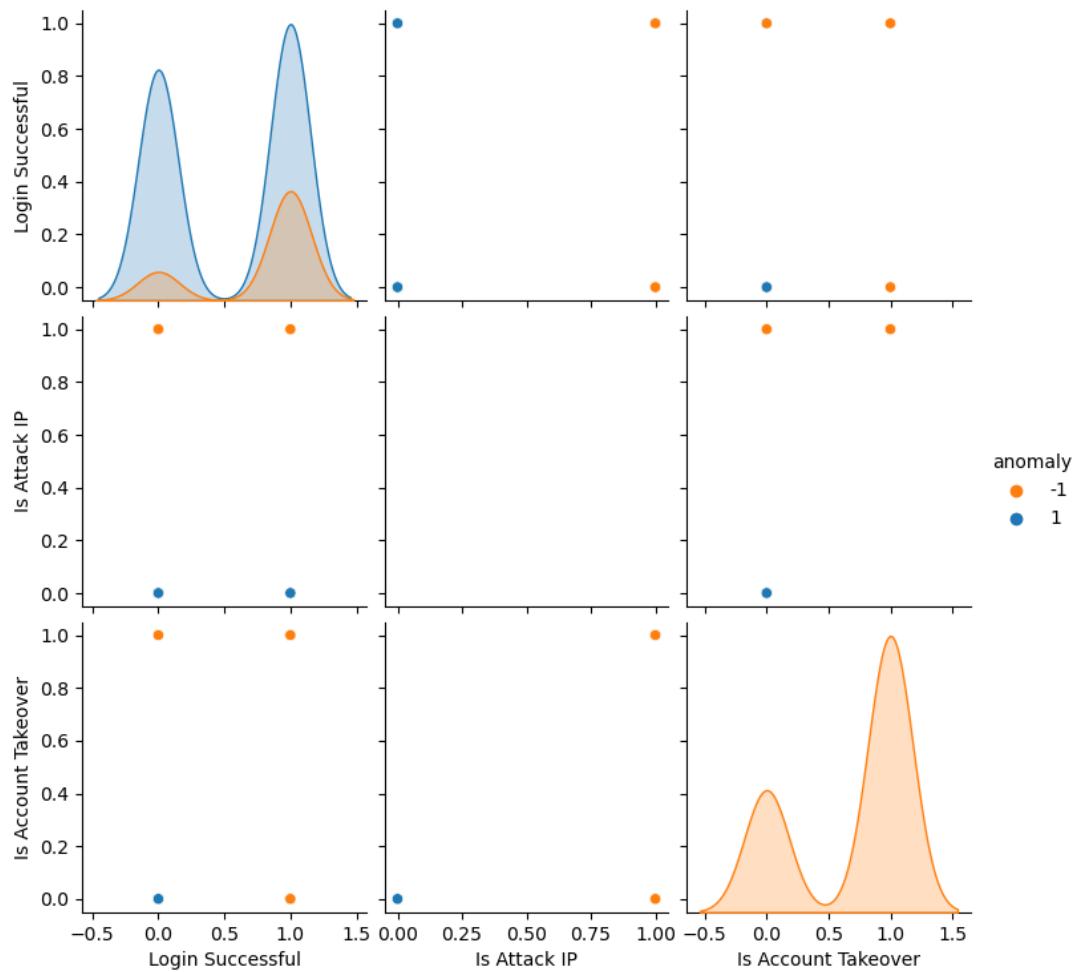
- contamination value == 0.3



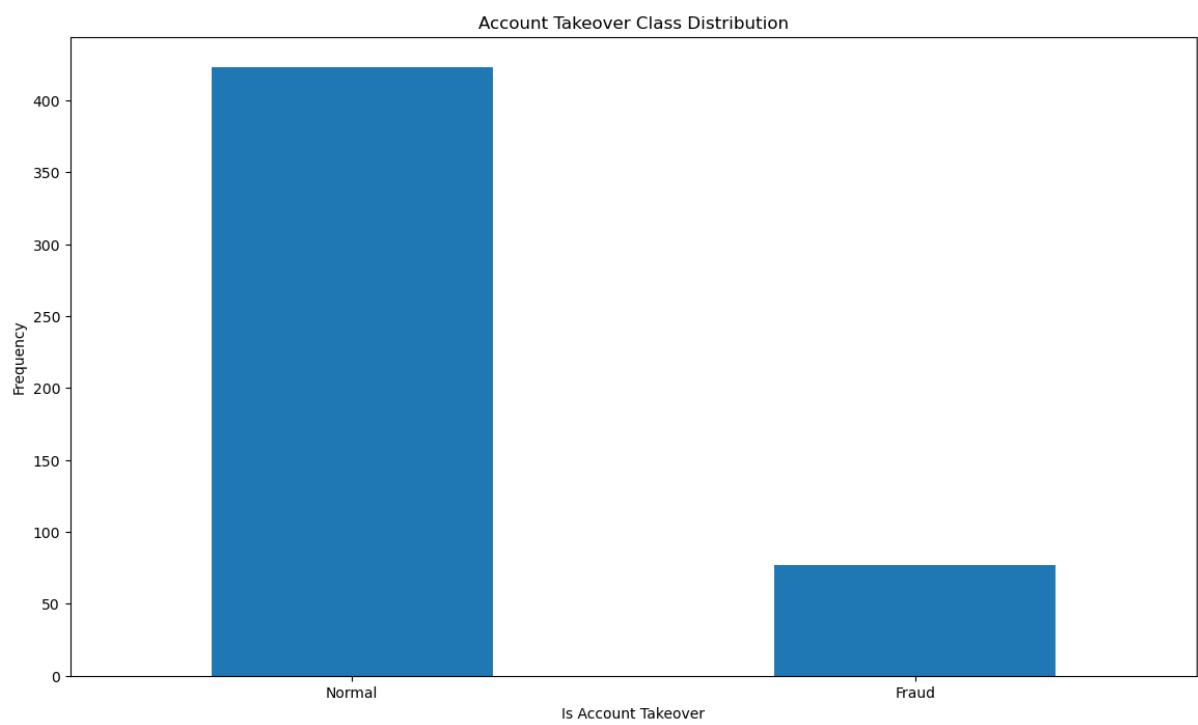


- contamination value == 0.5





3D. iForest, Local Outlier Factor, Support Vector Machine



Isolation Forest: 117

Accuracy Score :

0.766

Classification Report :

	precision	recall	f1-score	support
False	0.84	0.89	0.87	423
True	0.10	0.06	0.08	77
accuracy			0.77	500
macro avg	0.47	0.48	0.47	500
weighted avg	0.73	0.77	0.74	500

Local Outlier Factor: 111

Accuracy Score :

0.778

Classification Report :

	precision	recall	f1-score	support
False	0.85	0.90	0.87	423
True	0.16	0.10	0.13	77
accuracy			0.78	500
macro avg	0.50	0.50	0.50	500
weighted avg	0.74	0.78	0.76	500

Support Vector Machine: 343

Accuracy Score :

0.314

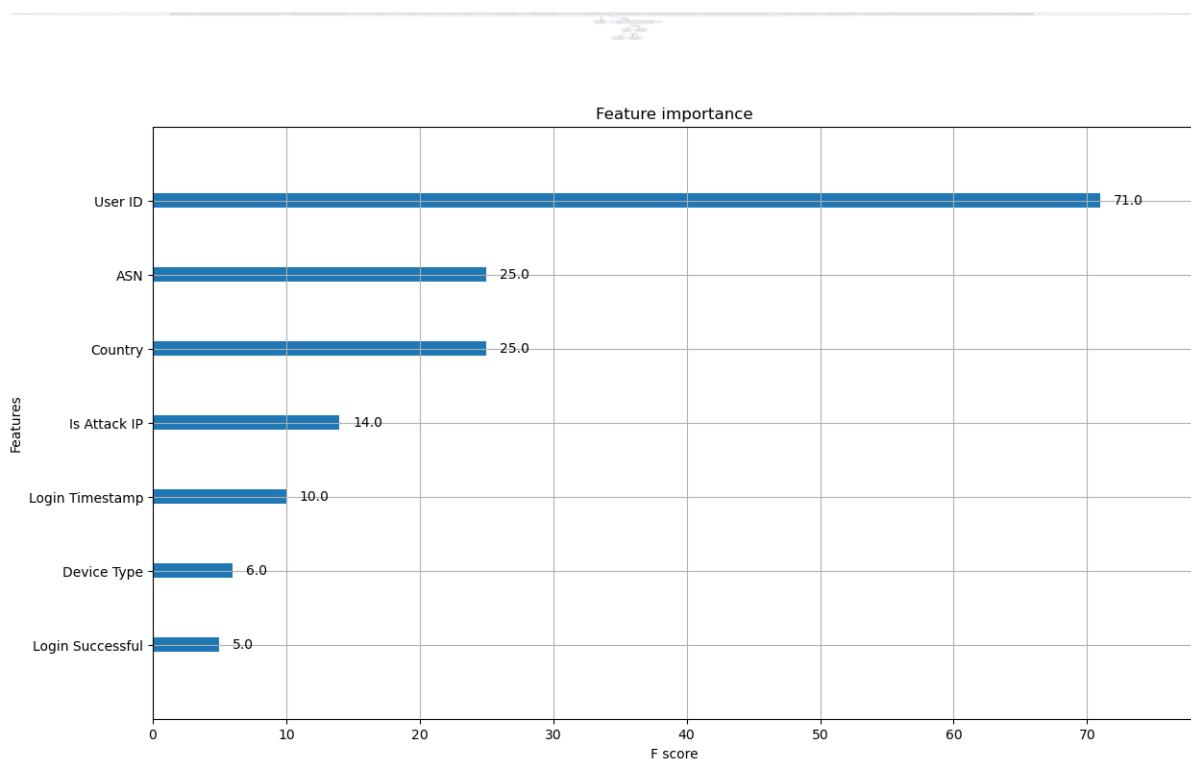
Classification Report :

	precision	recall	f1-score	support
False	0.79	0.26	0.39	423
True	0.13	0.61	0.22	77
accuracy			0.31	500
macro avg	0.46	0.44	0.30	500
weighted avg	0.68	0.31	0.36	500

IV. 141/500 → 28.2%

1. XGB Experimental

```
----  
0 Login Timestamp      500 non-null   category  
1 User ID              500 non-null   int64  
2 IP Address            500 non-null   category  
3 Country               500 non-null   category  
4 ASN                  500 non-null   int64  
5 Device Type           500 non-null   category  
6 Login Successful       500 non-null   bool  
7 Is Attack IP          500 non-null   bool  
8 Blacklisted            500 non-null   bool  
9 Browser Type           500 non-null   category  
10 Is Account Takeover  500 non-null  bool  
dtypes: bool(4), category(5), int64(2)  
memory usage: 54.2 KB  
Feature importances:  
[0.5381381  0.00136504 0.          0.22627188 0.04183649 0.00924834  
 0.04554712 0.13759302 0.          0.          ]
```



2. Label encoding (cat.codes)

2A. XGBoost

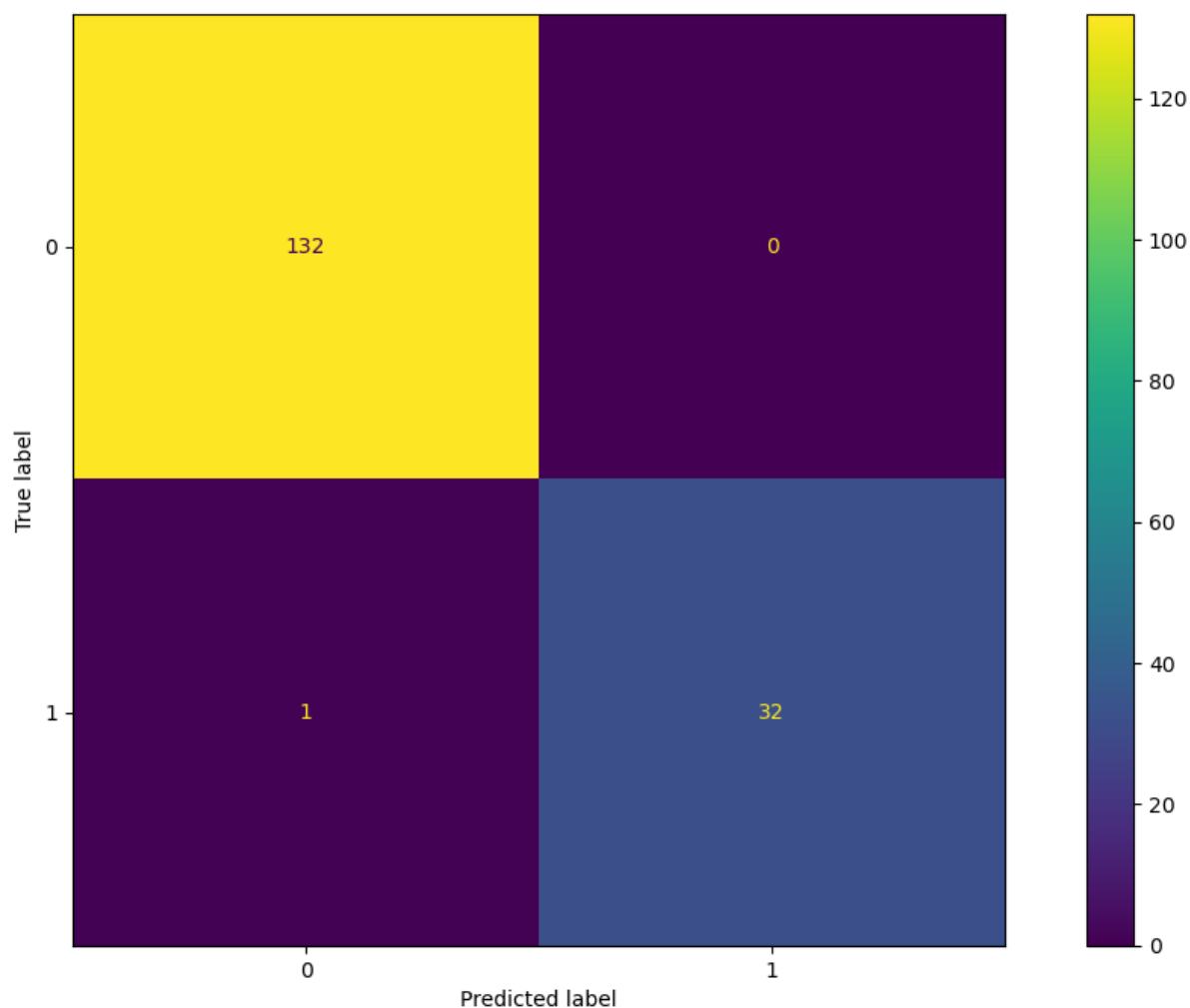
memory usage: 12.8 KB

```
Accuracy: 99.39%
      precision    recall  f1-score   support

      False        0.99     1.00     1.00      132
       True        1.00     0.97     0.98      33

  accuracy                           0.99      165
 macro avg        1.00     0.98     0.99      165
weighted avg     0.99     0.99     0.99      165

TN: 0, FP: 2, FN: 1, TP: 1
```



2B. ONE CLASS SVM

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.2 (same for any other col)

```
The number of records in the training dataset is 400
The number of records in the test dataset is 100
The training dataset has 339 records for the majority class and 61 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      84
      True        0.16     1.00     0.28      16

      accuracy          0.16      100
macro avg       0.08     0.50     0.14      100
weighted avg    0.03     0.16     0.04      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.84     1.00     0.91      84
      True        0.00     0.00     0.00      16

      accuracy          0.84      100
macro avg       0.42     0.50     0.46      100
weighted avg    0.71     0.84     0.77      100

[[ 0 84]
 [ 0 16]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.3 (same for any other col)

```
The number of records in the training dataset is 350
The number of records in the test dataset is 150
The training dataset has 291 records for the majority class and 59 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00     132
      True        0.12     1.00     0.21      18

      accuracy          0.12      150
macro avg       0.06     0.50     0.11      150
weighted avg    0.01     0.12     0.03      150

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.88     1.00     0.94     132
      True        0.00     0.00     0.00      18

      accuracy          0.88      150
macro avg       0.44     0.50     0.47      150
weighted avg    0.77     0.88     0.82      150

[[ 0 132]
 [ 0 18]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.5 (same for any other col)

```

The number of records in the training dataset is 250
The number of records in the test dataset is 250
The training dataset has 206 records for the majority class and 44 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      217
      True         0.13     1.00     0.23      33

      accuracy          0.13      250
macro avg       0.07     0.50     0.12      250
weighted avg    0.02     0.13     0.03      250

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.87     1.00     0.93      217
      True         0.00     0.00     0.00      33

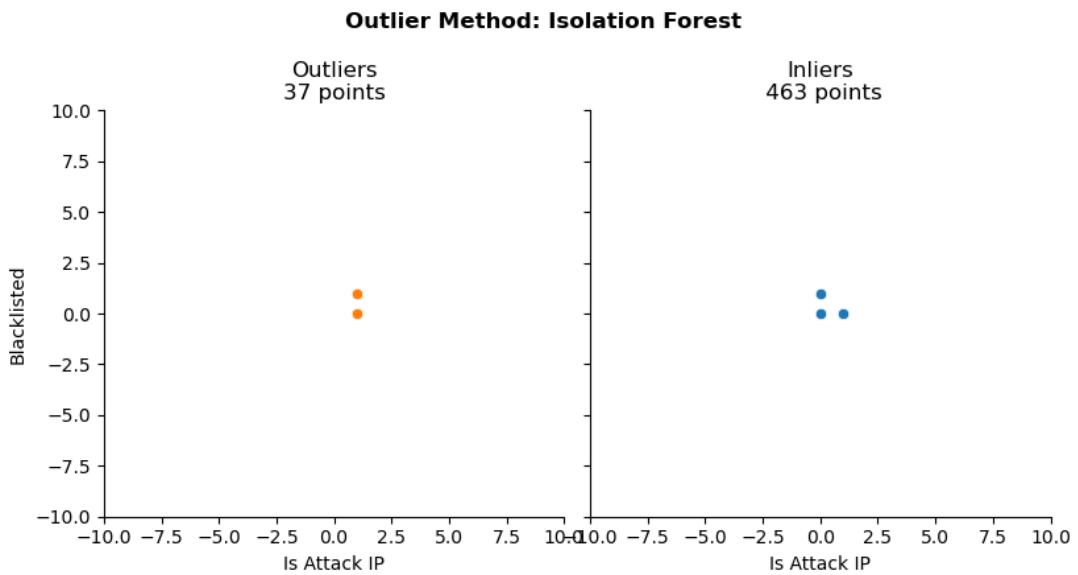
      accuracy          0.87      250
macro avg       0.43     0.50     0.46      250
weighted avg    0.75     0.87     0.81      250

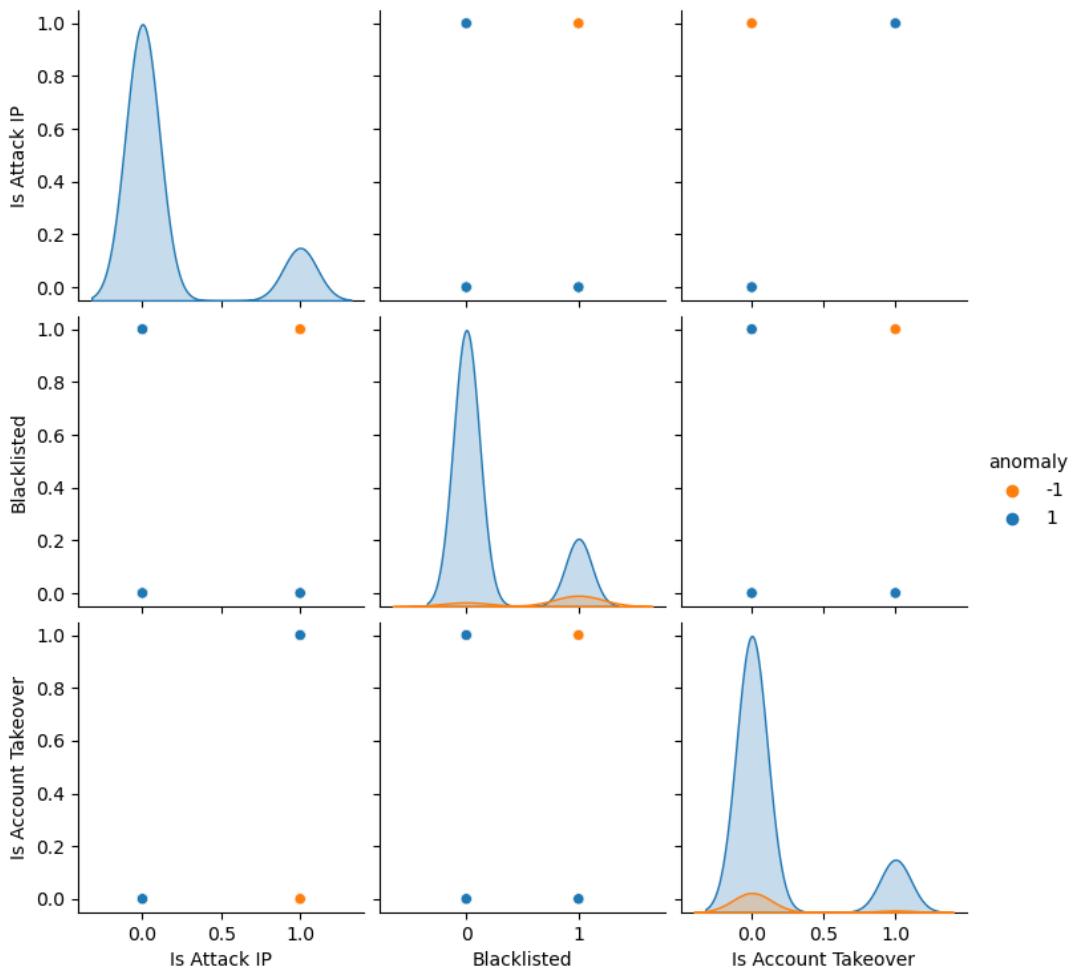
[[ 0 217]
 [ 0 33]]

```

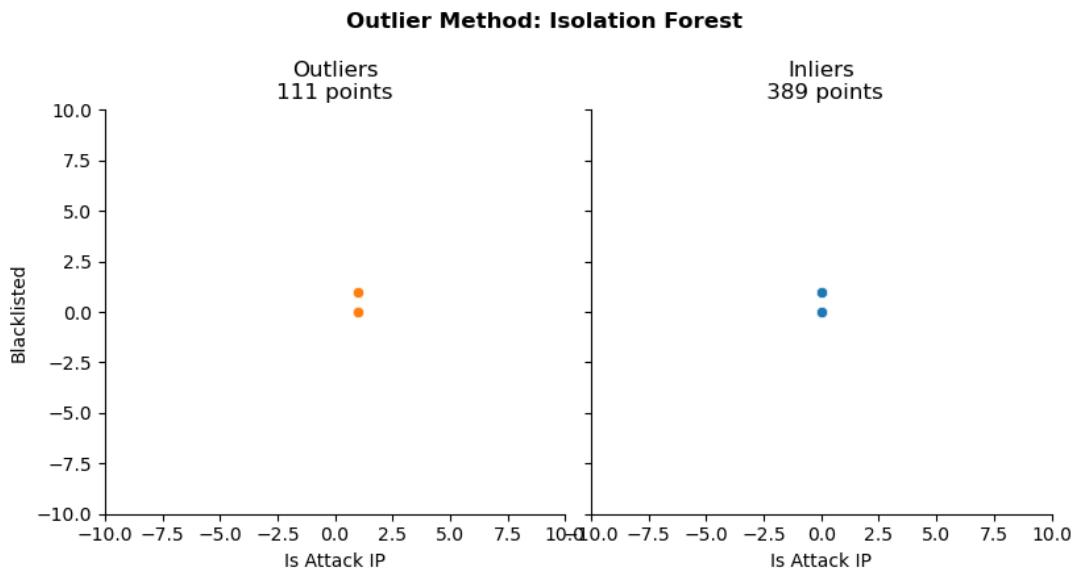
2C. Isolation Forest

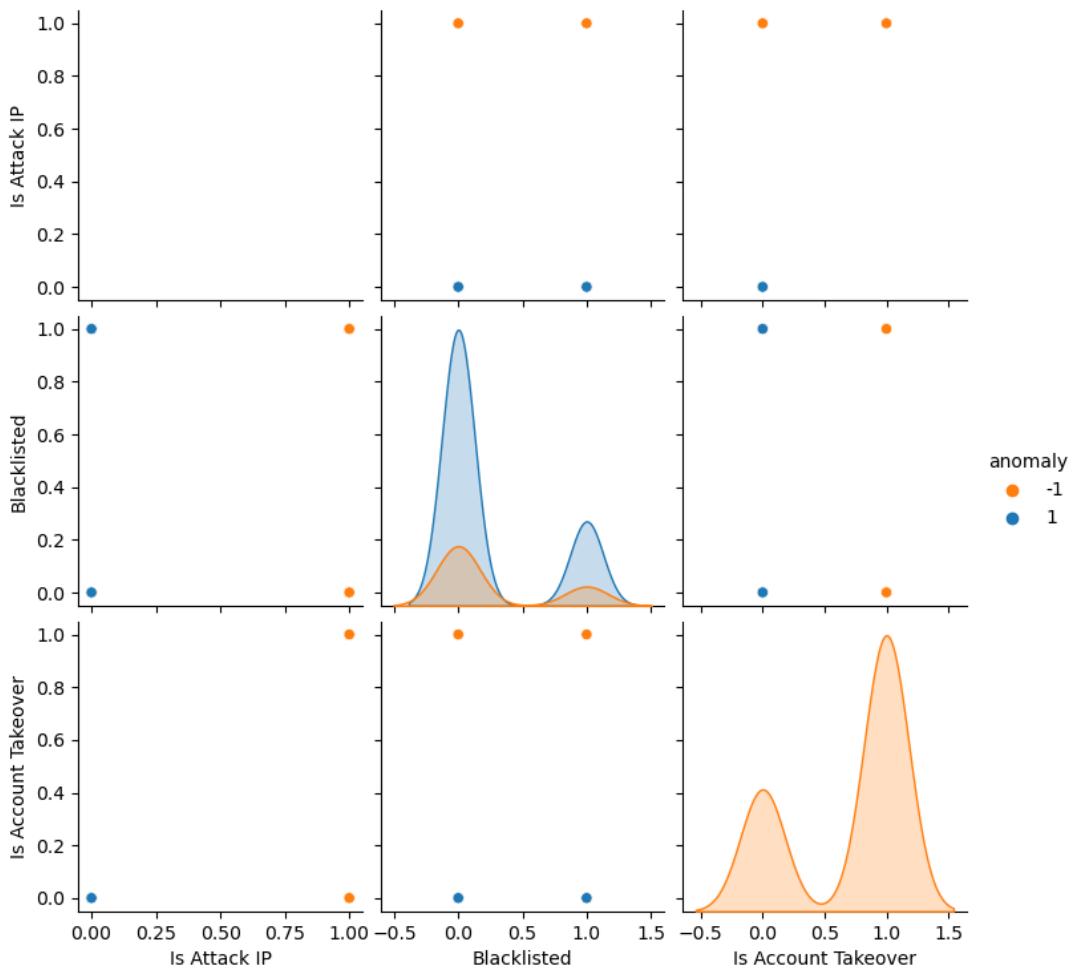
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



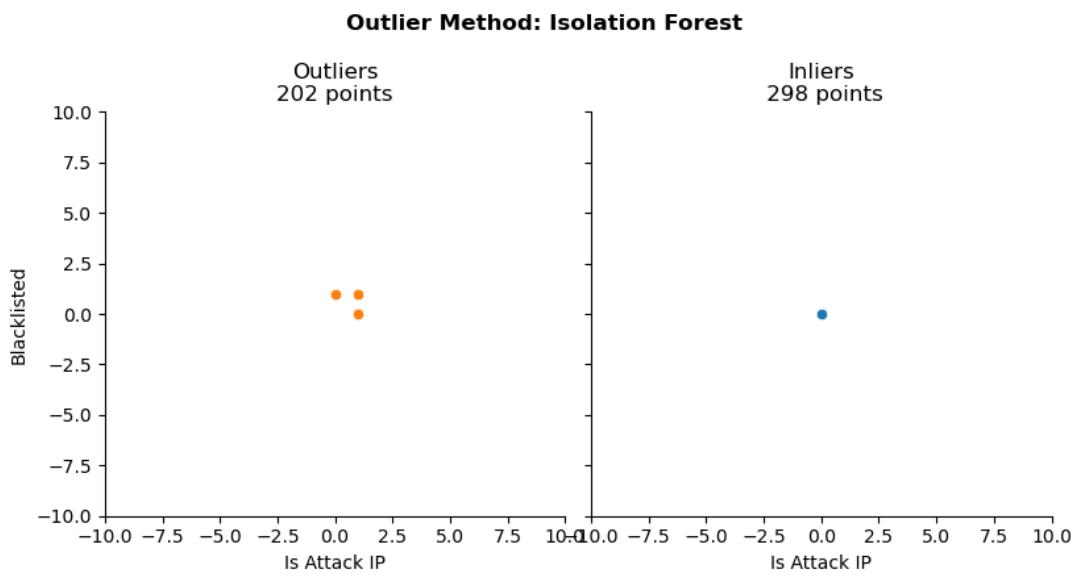


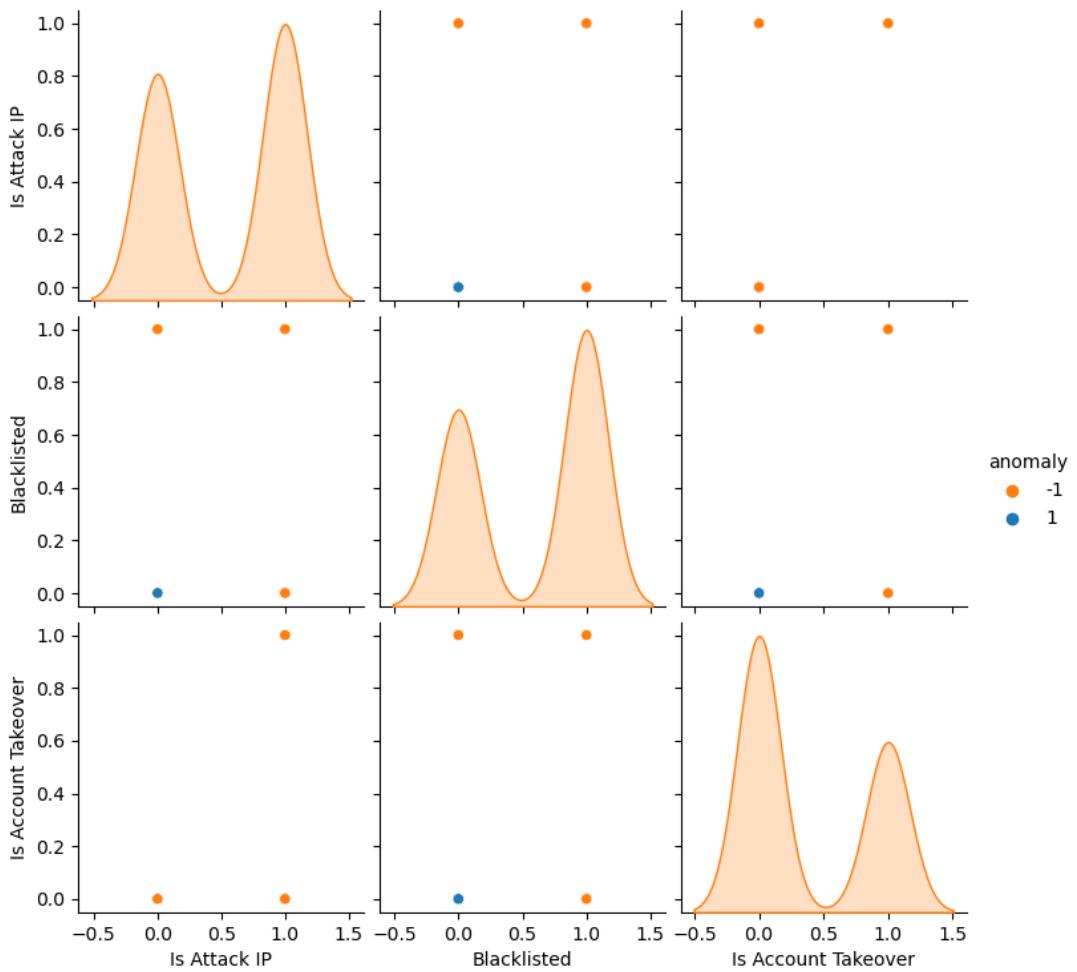
- increasing contamination value to 0.3



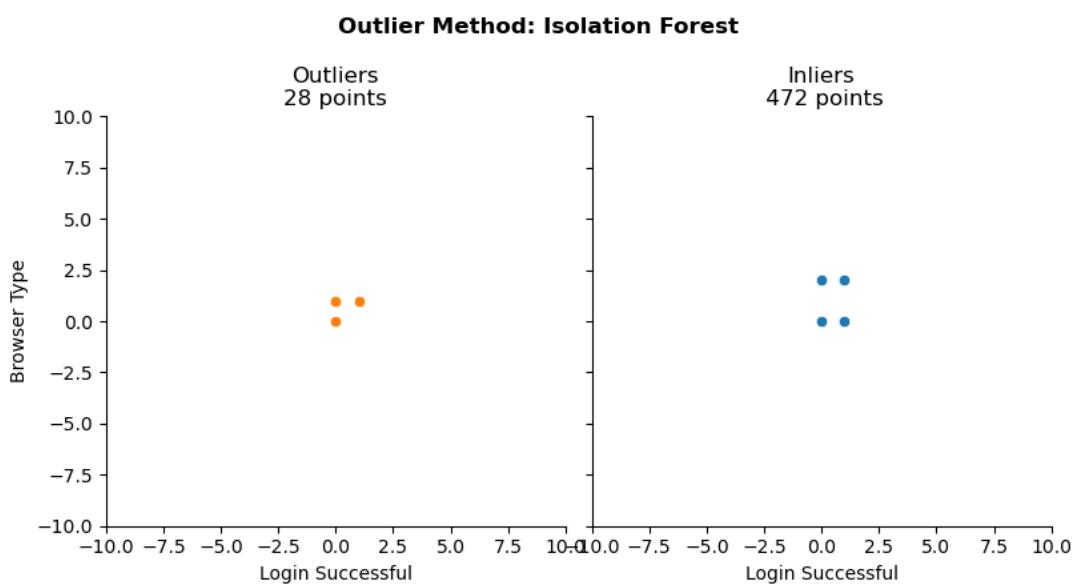


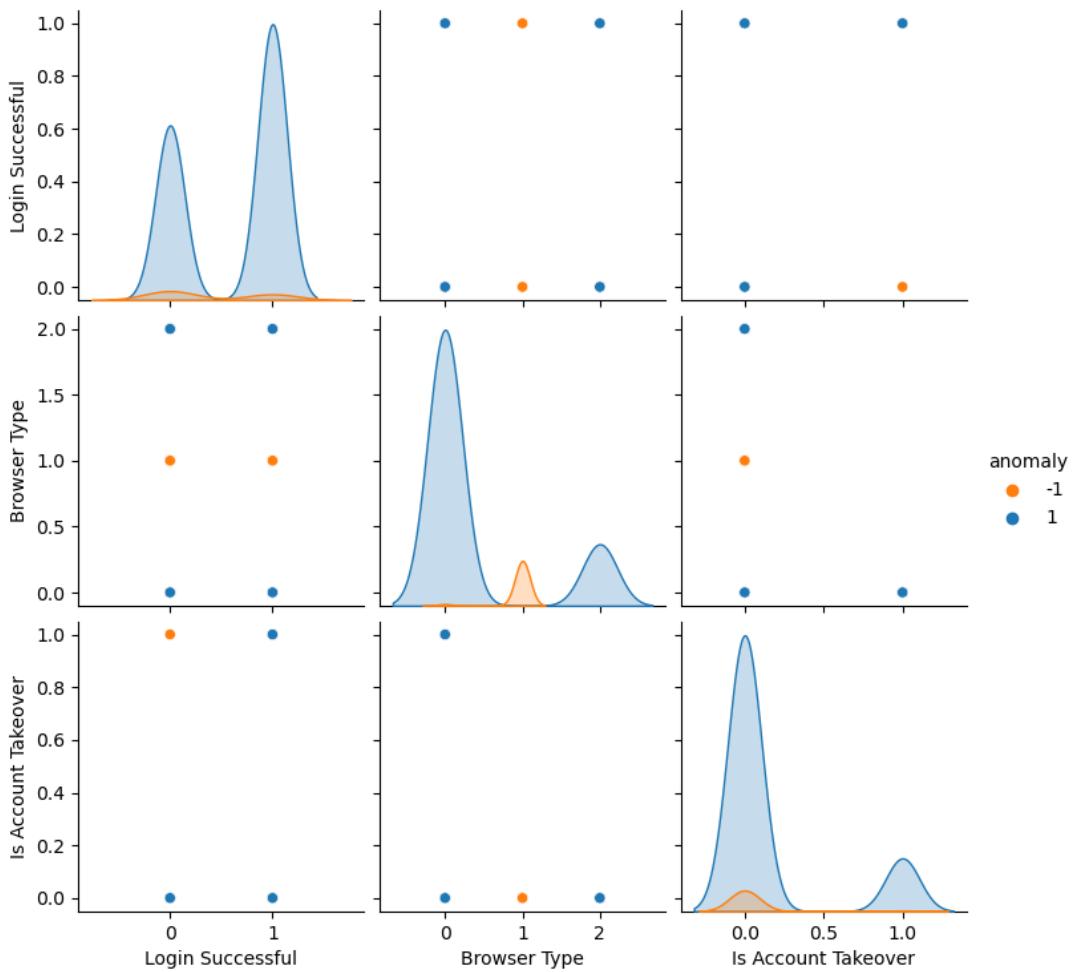
- contamination value == 0.5



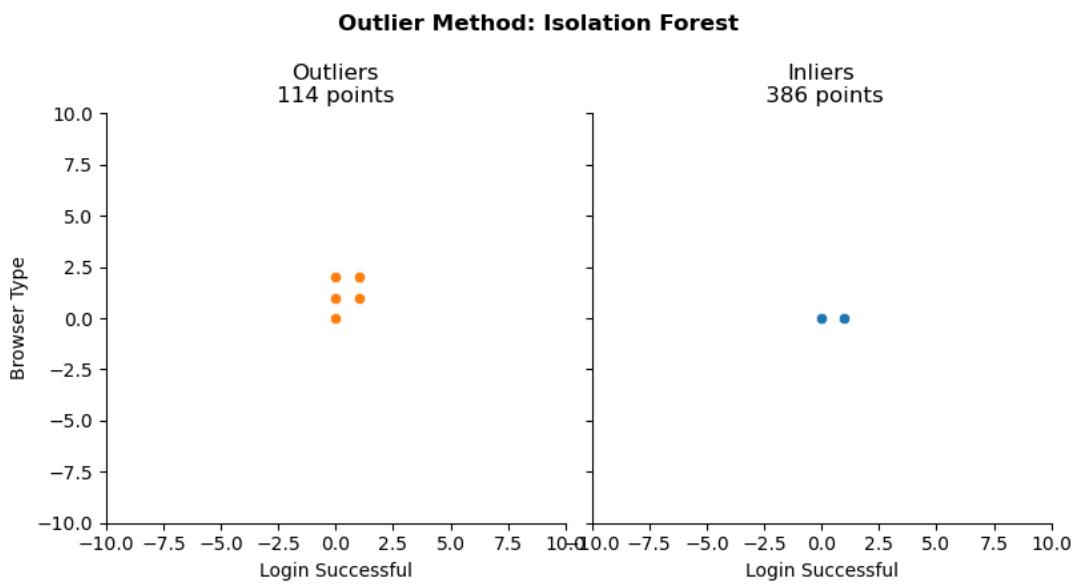


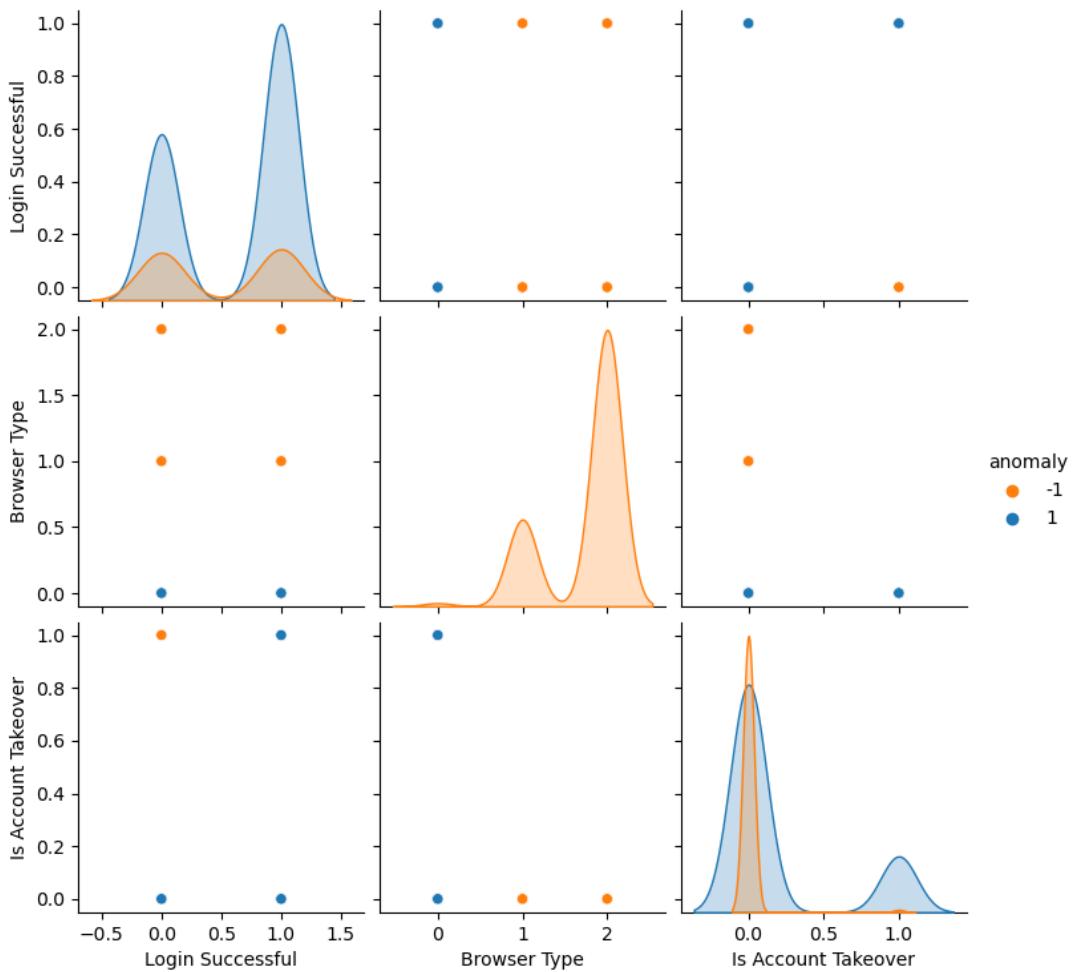
- anomaly_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']
 - contamination value == 0.1



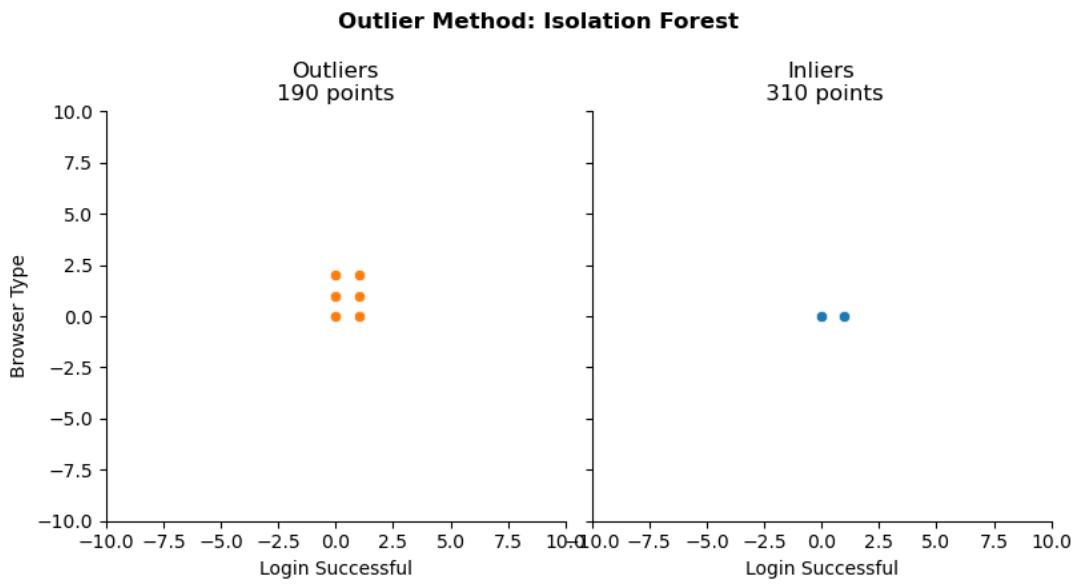


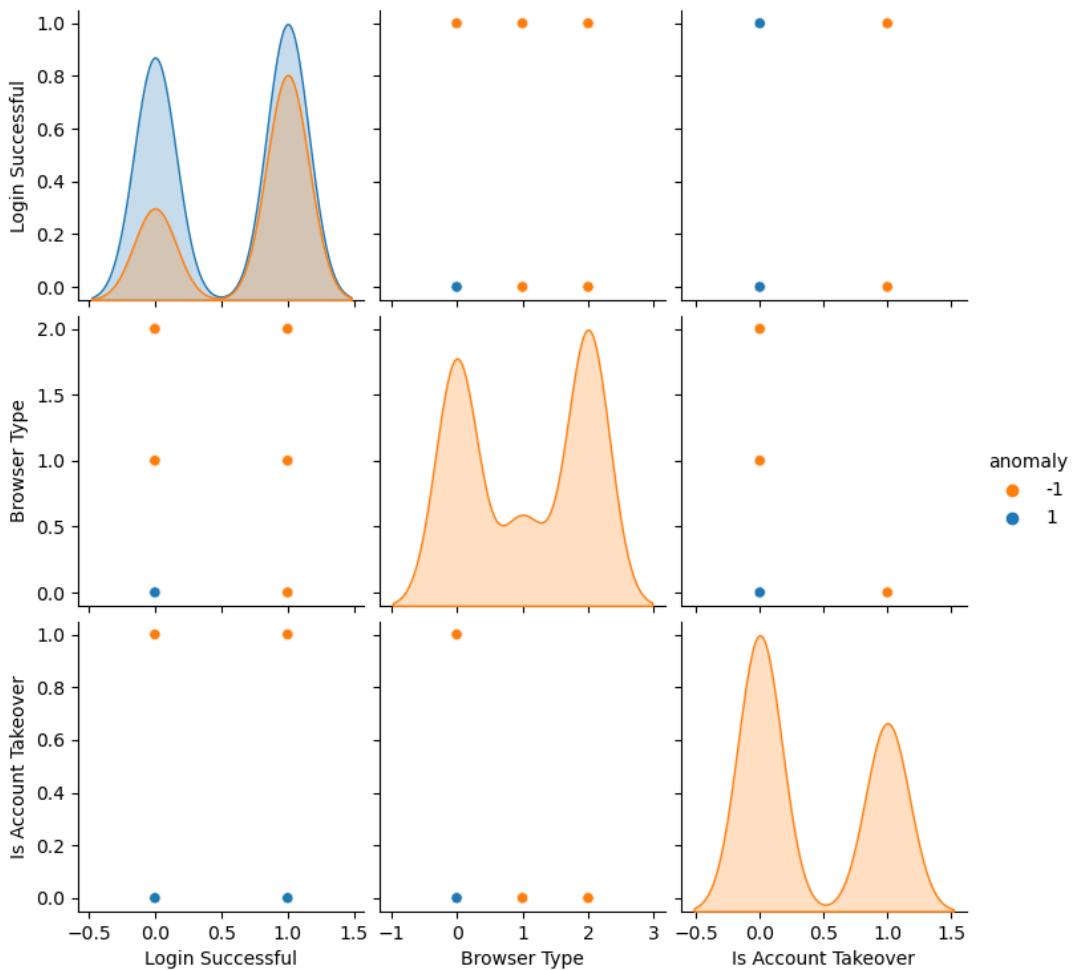
- contamination value == 0.3



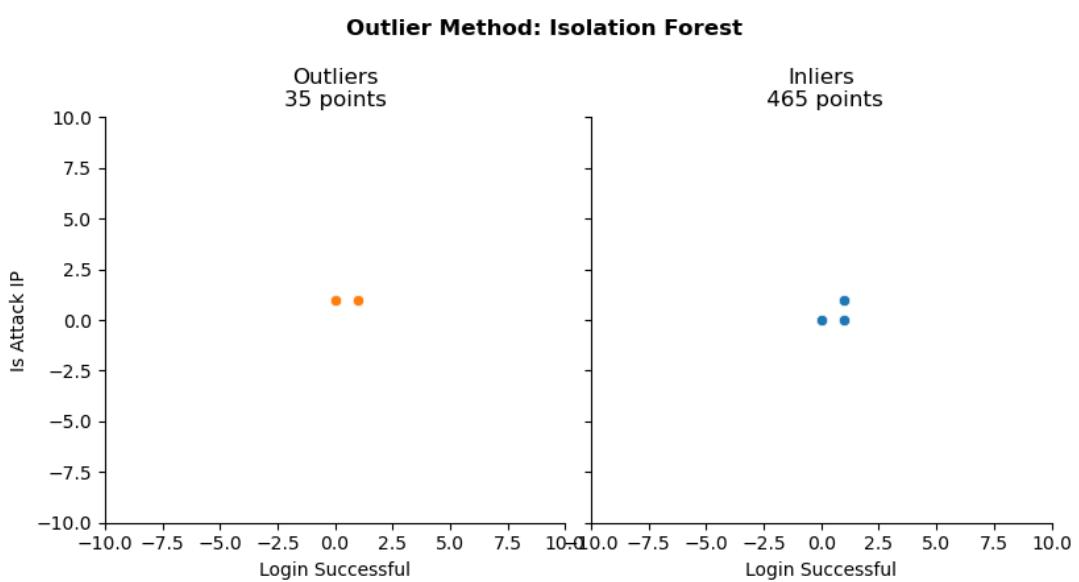


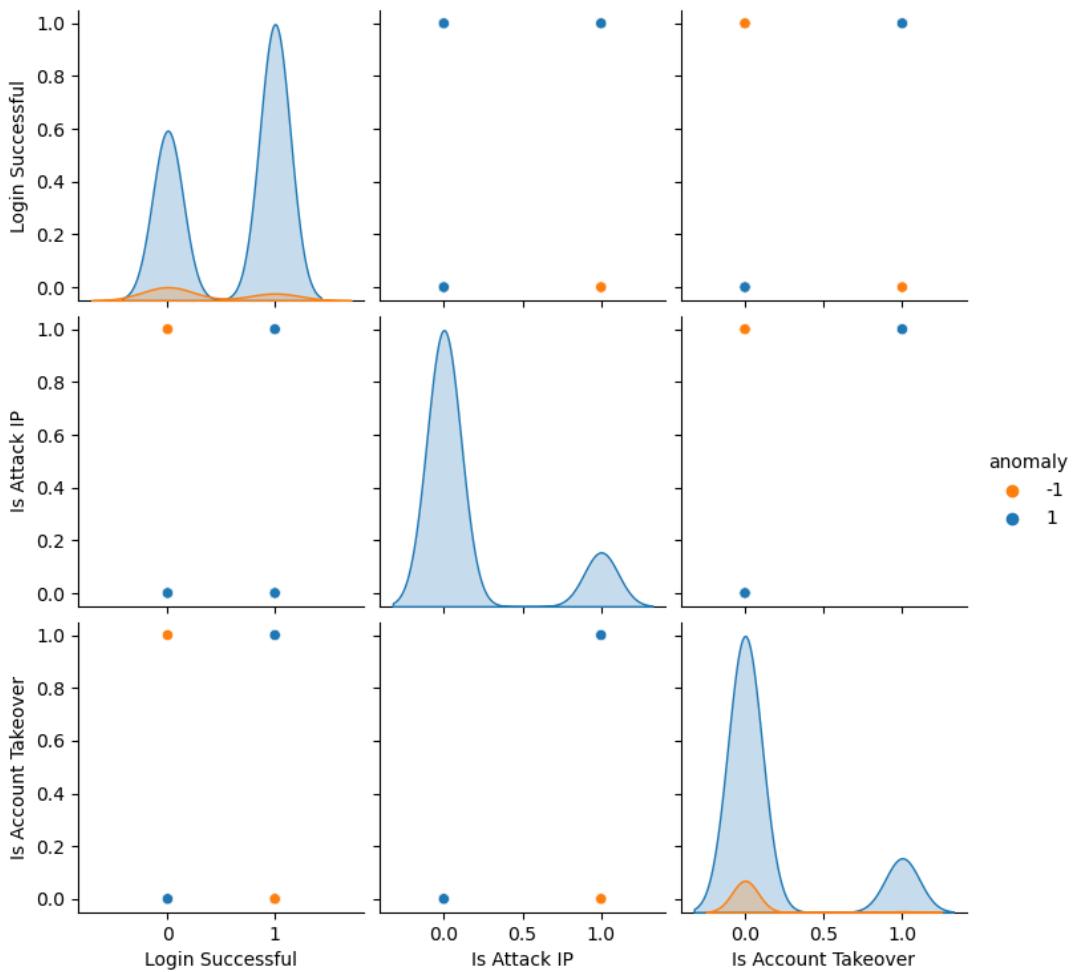
- contamination value == 0.5





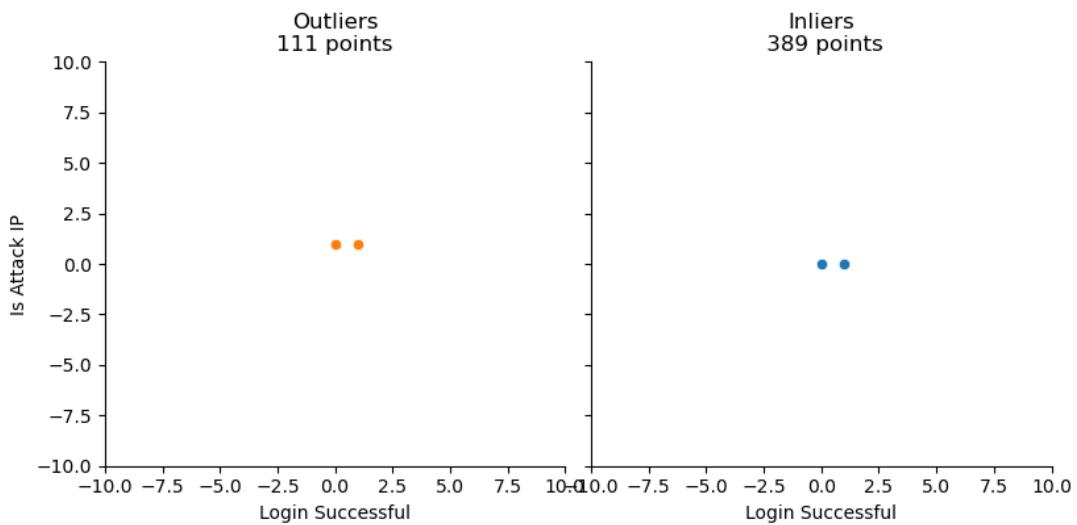
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1

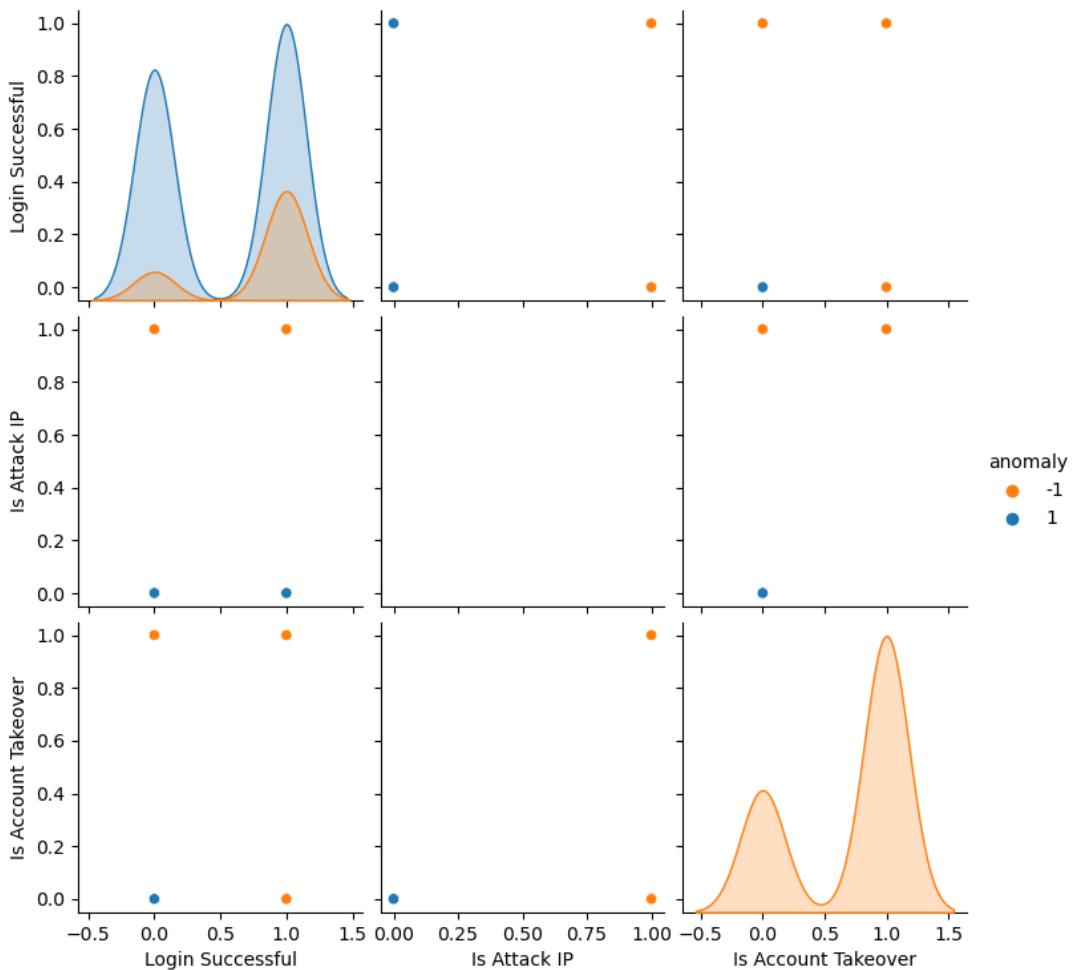




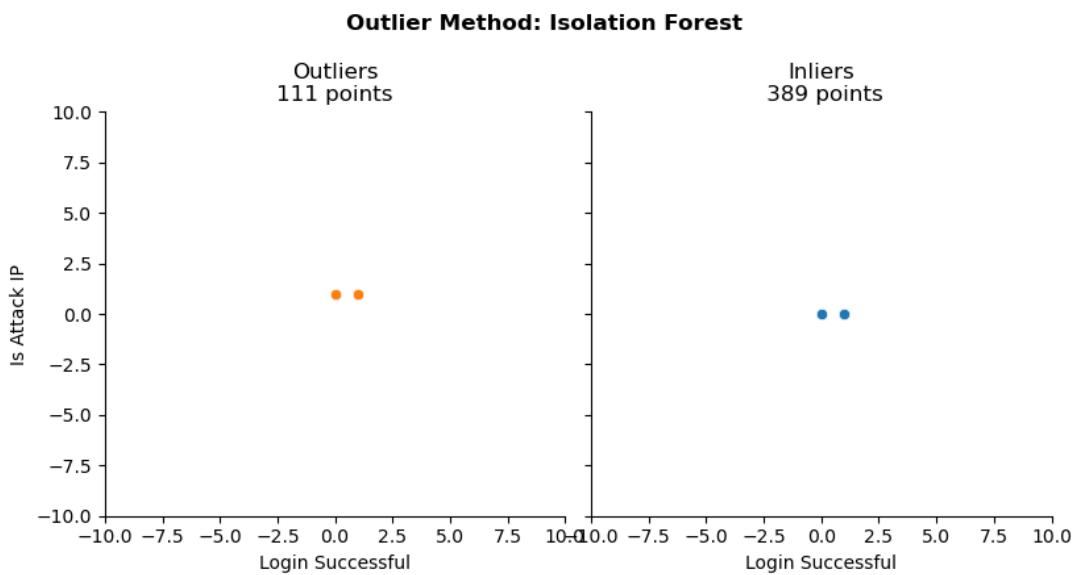
- contamination value == 0.3

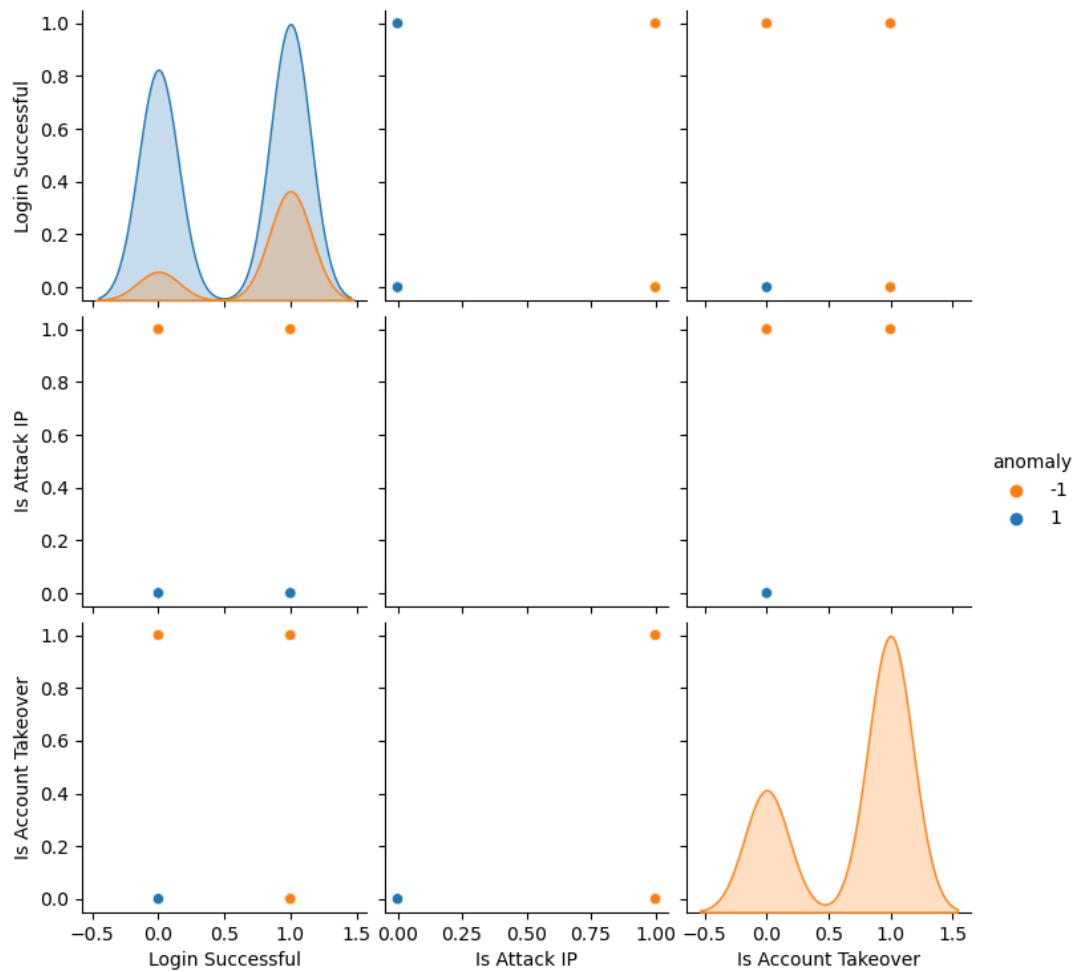
Outlier Method: Isolation Forest



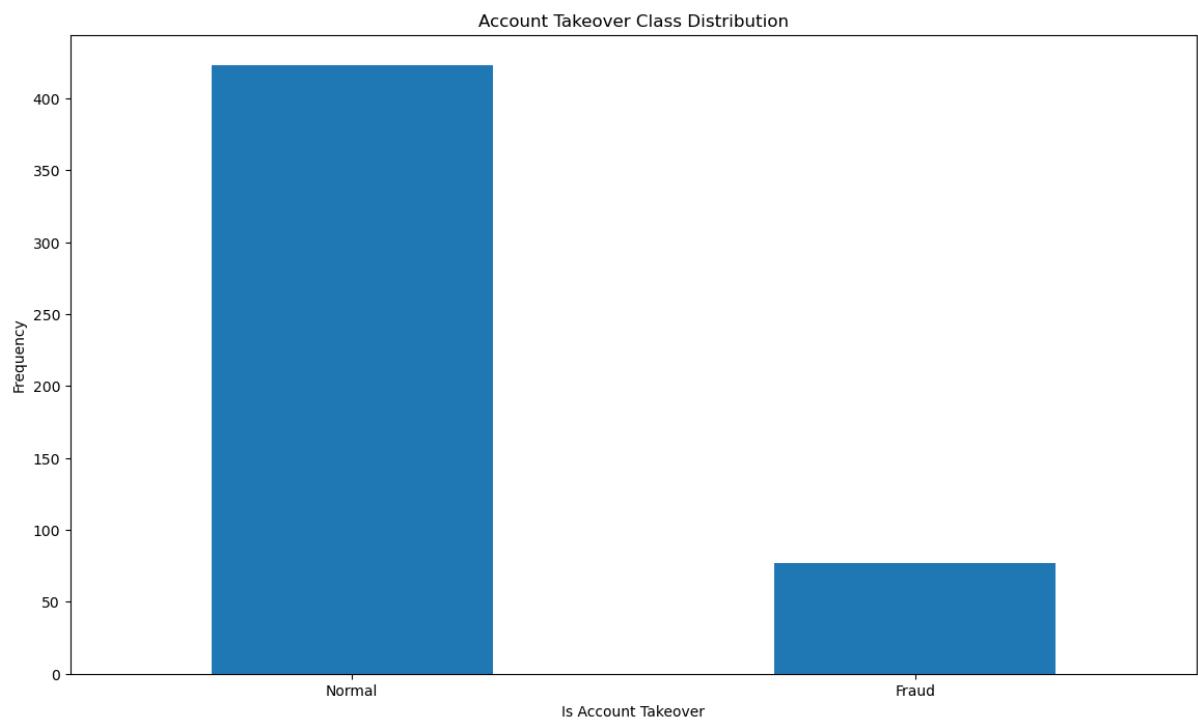


- contamination value == 0.5





2D. iForest, Local Outlier Factor, Support Vector Machine



"{}: {}".format(clf_name,n_errors) → np. Isolation Forest: 113

Isolation Forest: 113

Accuracy Score :

0.774

Classification Report :

	precision	recall	f1-score	support
False	0.84	0.90	0.87	423
True	0.14	0.09	0.11	77
accuracy			0.77	500
macro avg	0.49	0.49	0.49	500
weighted avg	0.74	0.77	0.75	500

Local Outlier Factor: 117

Accuracy Score :

0.766

Classification Report :

	precision	recall	f1-score	support
False	0.84	0.89	0.87	423
True	0.10	0.06	0.08	77
accuracy			0.77	500
macro avg	0.47	0.48	0.47	500
weighted avg	0.73	0.77	0.74	500

Support Vector Machine: 343

Accuracy Score :

0.314

Classification Report :

	precision	recall	f1-score	support
False	0.79	0.26	0.39	423
True	0.13	0.61	0.22	77
accuracy			0.31	500
macro avg	0.46	0.44	0.30	500
weighted avg	0.68	0.31	0.36	500

3. Dummy Variable Encoding → df3 = pd.get_dummies(df3, columns=['Country', 'Device

Type', 'IP Address', 'Login Timestamp', 'Browser Type'], prefix=["cntr", "dvc", "ip", "Ignt", "brw"])

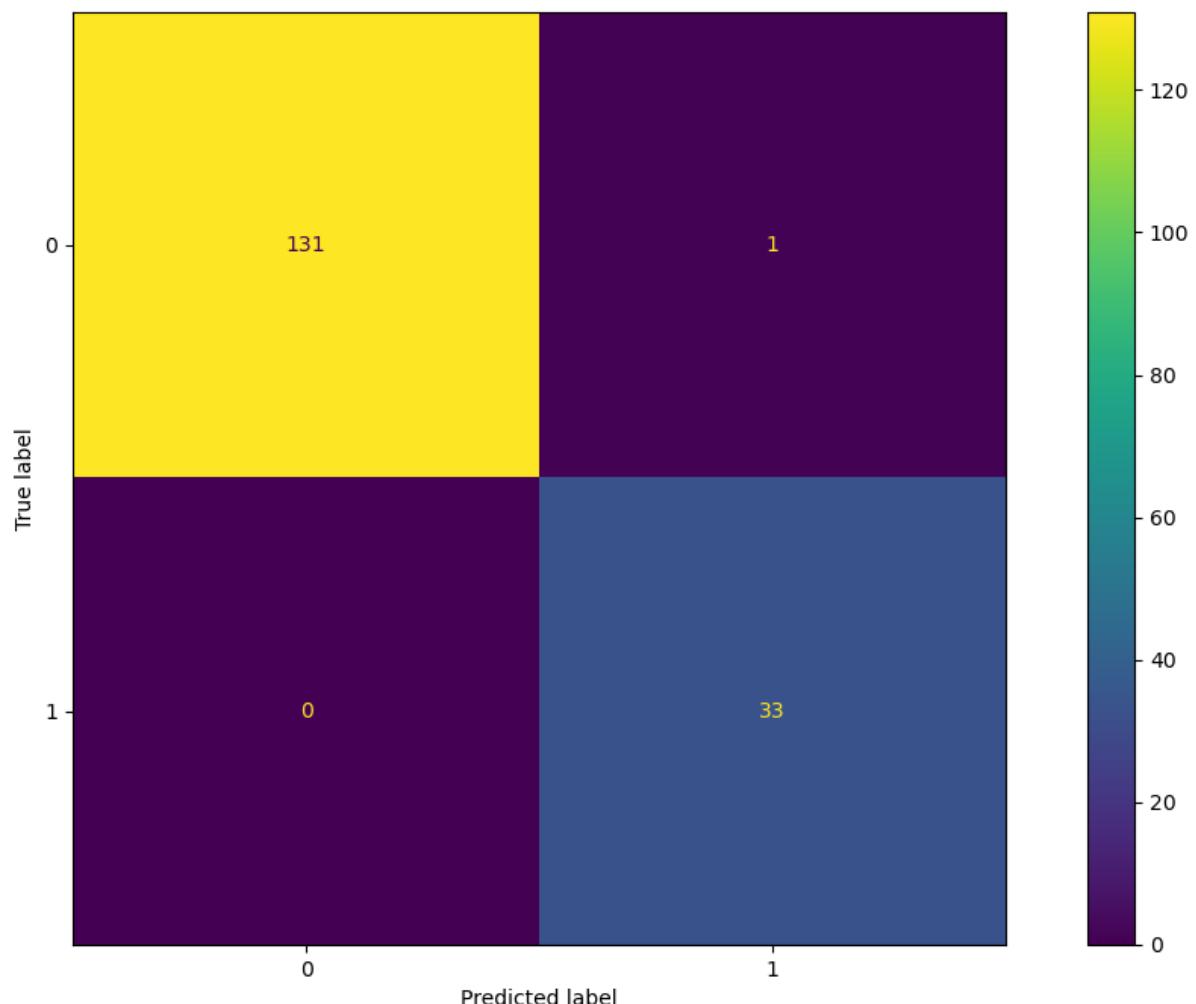
3A. XGBoost

memory usage: 466.4 KB

Accuracy: 99.39%

	precision	recall	f1-score	support
False	1.00	0.99	1.00	132
True	0.97	1.00	0.99	33
accuracy			0.99	165
macro avg	0.99	1.00	0.99	165
weighted avg	0.99	0.99	0.99	165

TN: 0, FP: 2, FN: 1, TP: 1



3B. One Class SVM

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.2 (same for any other col)

```
The number of records in the training dataset is 400
The number of records in the test dataset is 100
The training dataset has 339 records for the majority class and 61 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00      84
  True        0.16     1.00     0.28      16

  accuracy          0.16      100
  macro avg       0.08     0.50     0.14      100
  weighted avg    0.03     0.16     0.04      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.84     1.00     0.91      84
  True        0.00     0.00     0.00      16

  accuracy          0.84      100
  macro avg       0.42     0.50     0.46      100
  weighted avg    0.71     0.84     0.77      100

[[ 0 84]
 [ 0 16]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.3 (same for any other col)

```
The number of records in the training dataset is 350
The number of records in the test dataset is 150
The training dataset has 291 records for the majority class and 59 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00     132
  True        0.12     1.00     0.21      18

  accuracy          0.12      150
  macro avg       0.06     0.50     0.11      150
  weighted avg    0.01     0.12     0.03      150

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.88     1.00     0.94     132
  True        0.00     0.00     0.00      18

  accuracy          0.88      150
  macro avg       0.44     0.50     0.47      150
  weighted avg    0.77     0.88     0.82      150

[[ 0 132]
 [ 0 18]]
```

- columns=['Is Account Takeover', 'Blacklisted'] test_size=0.5 (same for any other col)

```

The number of records in the training dataset is 250
The number of records in the test dataset is 250
The training dataset has 206 records for the majority class and 44 records for the minority class.
      precision    recall   f1-score   support
      False        0.91     0.05     0.09      217
      True         0.13     0.97     0.24      33
      accuracy          0.17      250
      macro avg       0.52     0.51     0.16      250
      weighted avg    0.81     0.17     0.11      250

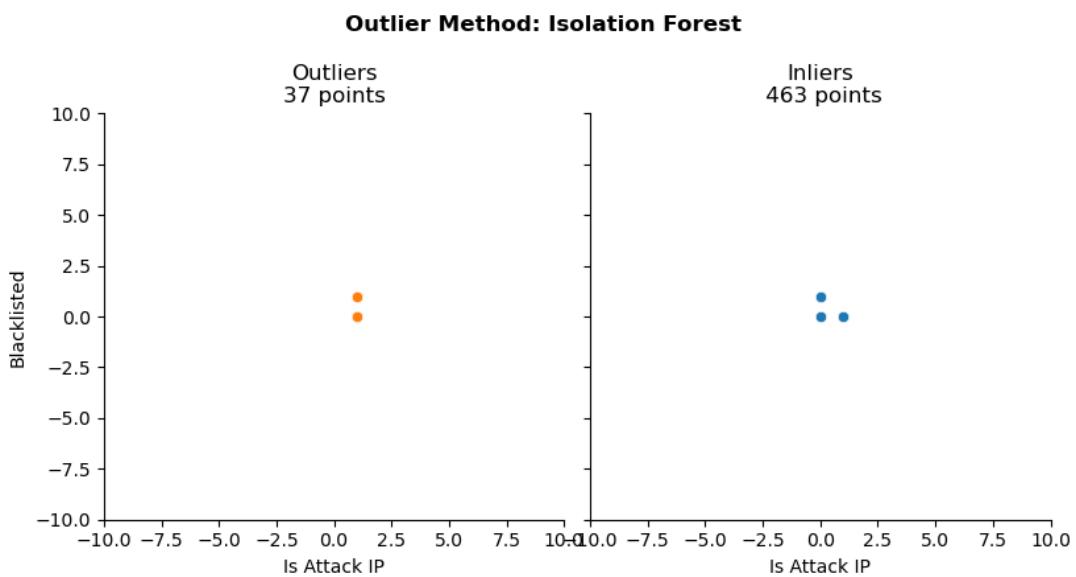
The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.87     1.00     0.93      217
      True         0.00     0.00     0.00      33
      accuracy          0.87      250
      macro avg       0.43     0.50     0.46      250
      weighted avg    0.75     0.87     0.81      250

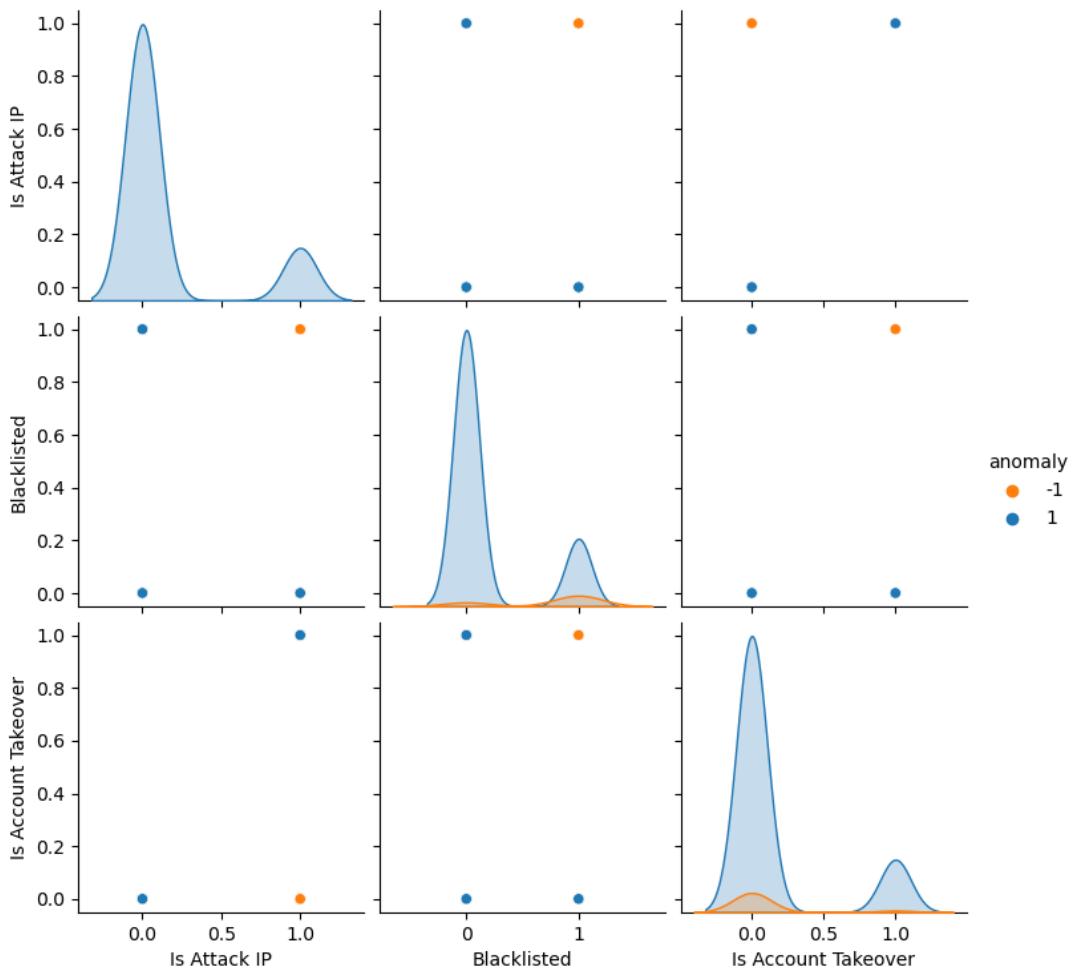
[[ 10 207]
 [ 1 32]]

```

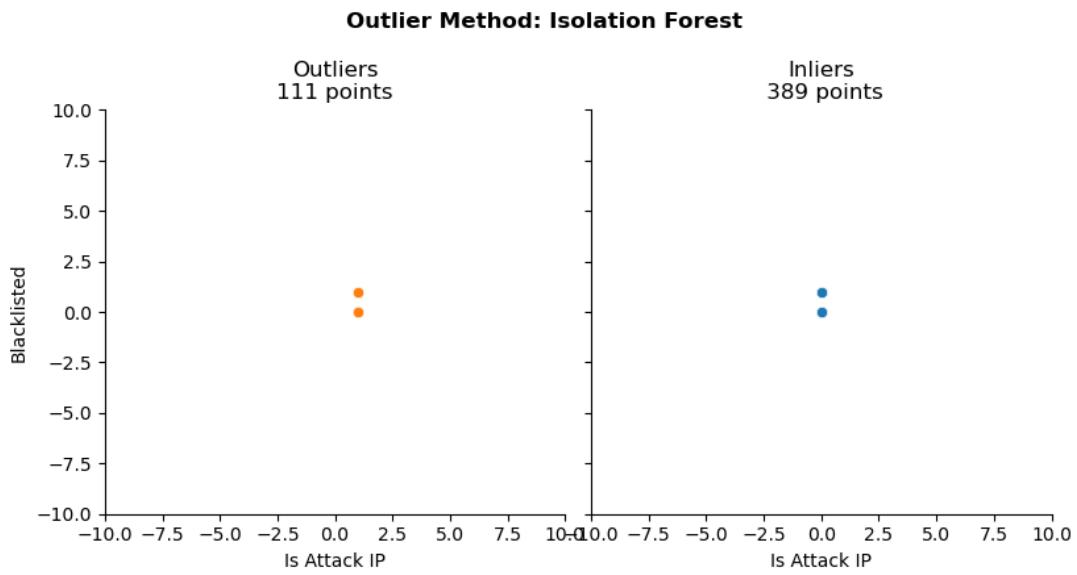
3C. Isolation Forest

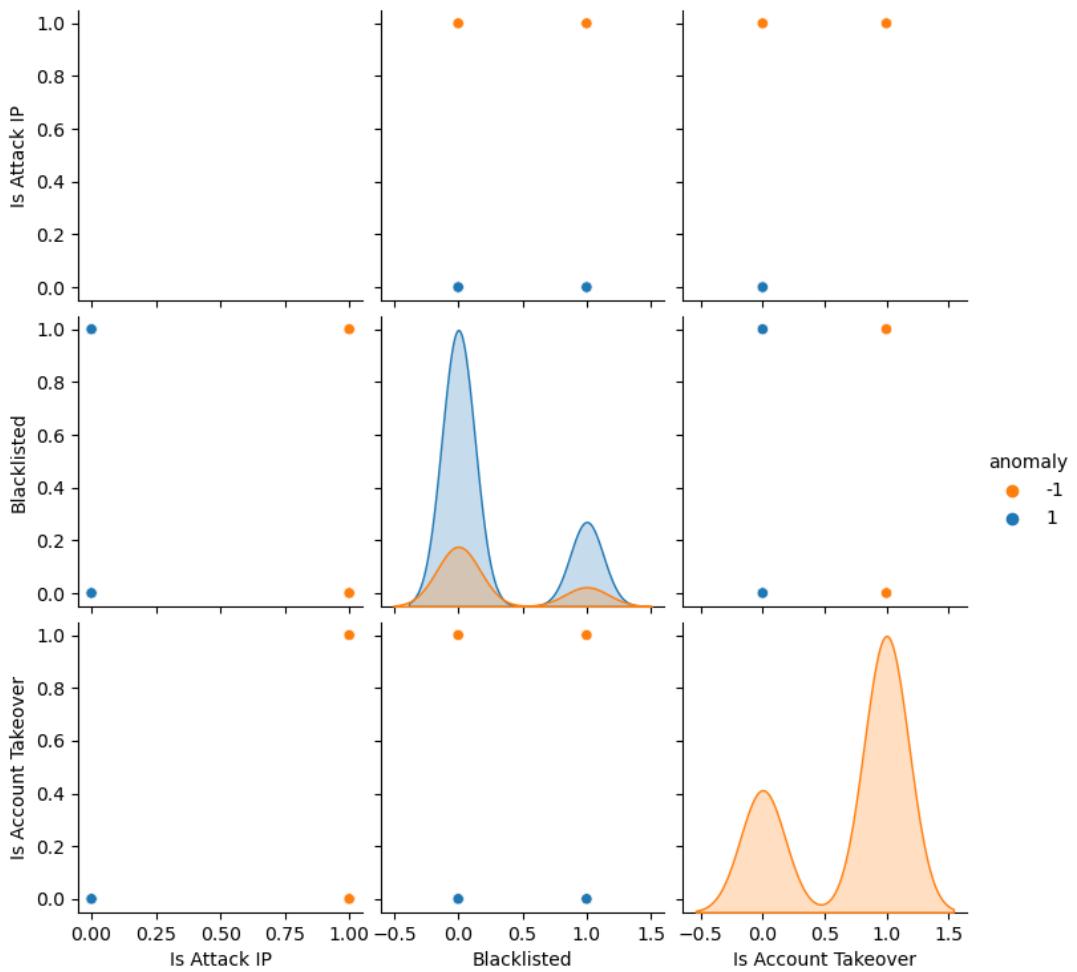
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1





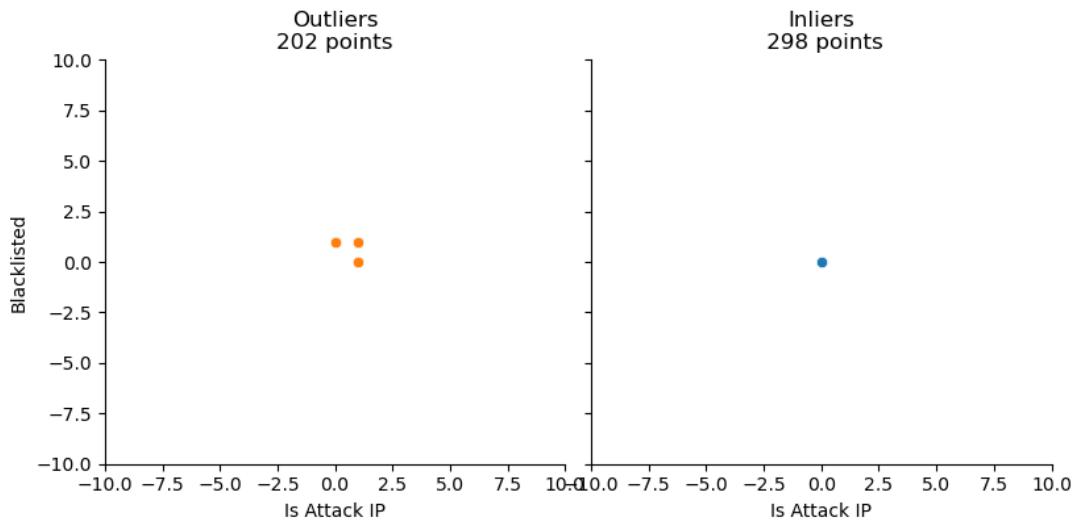
- contamination value == 0.3

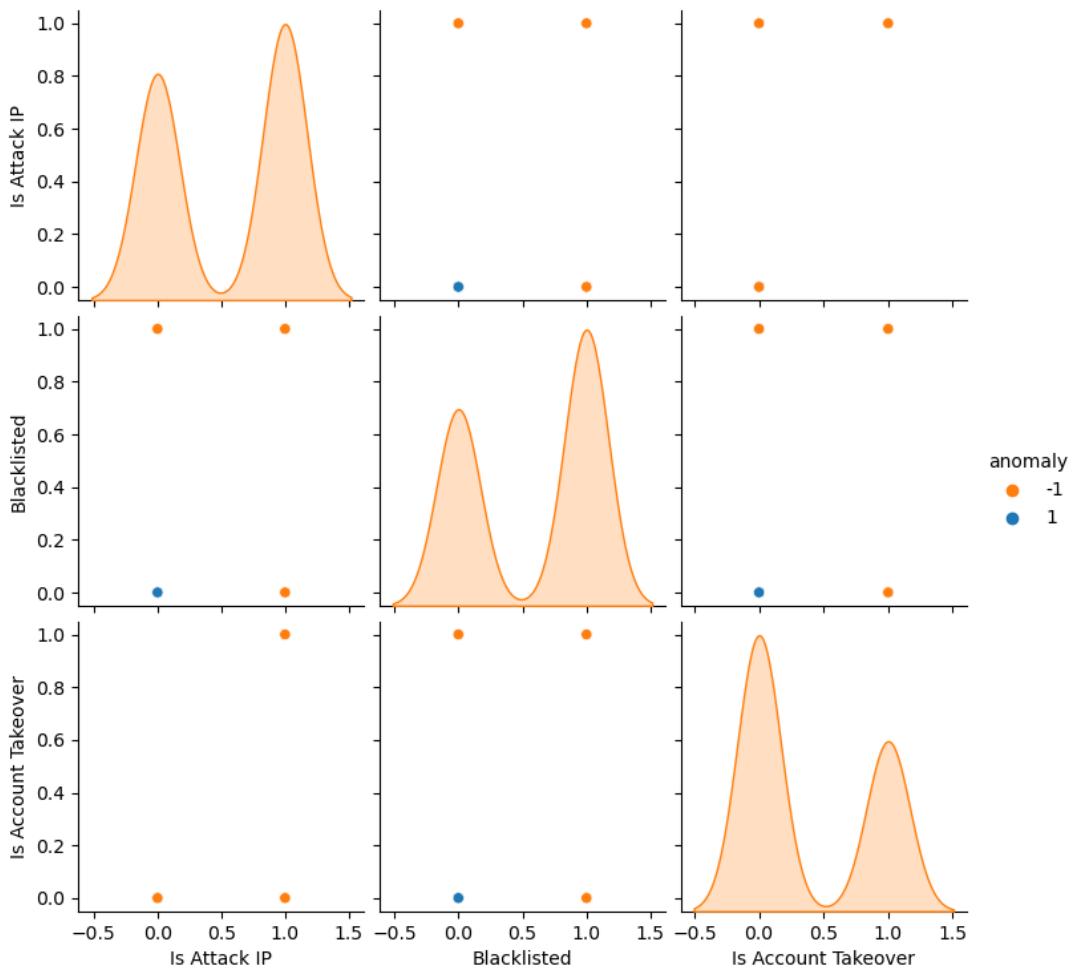




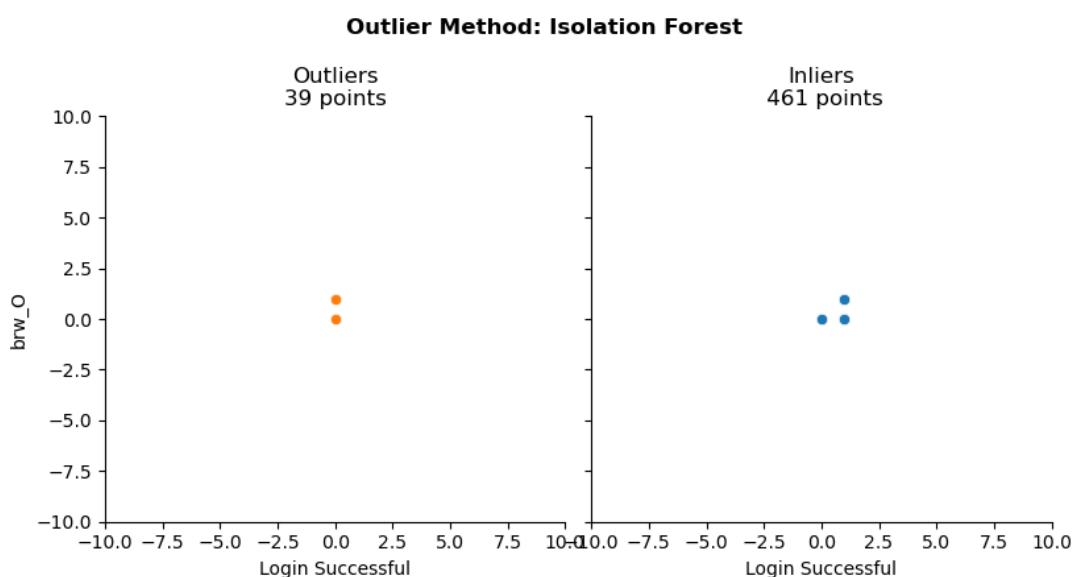
- contamination value == 0.5

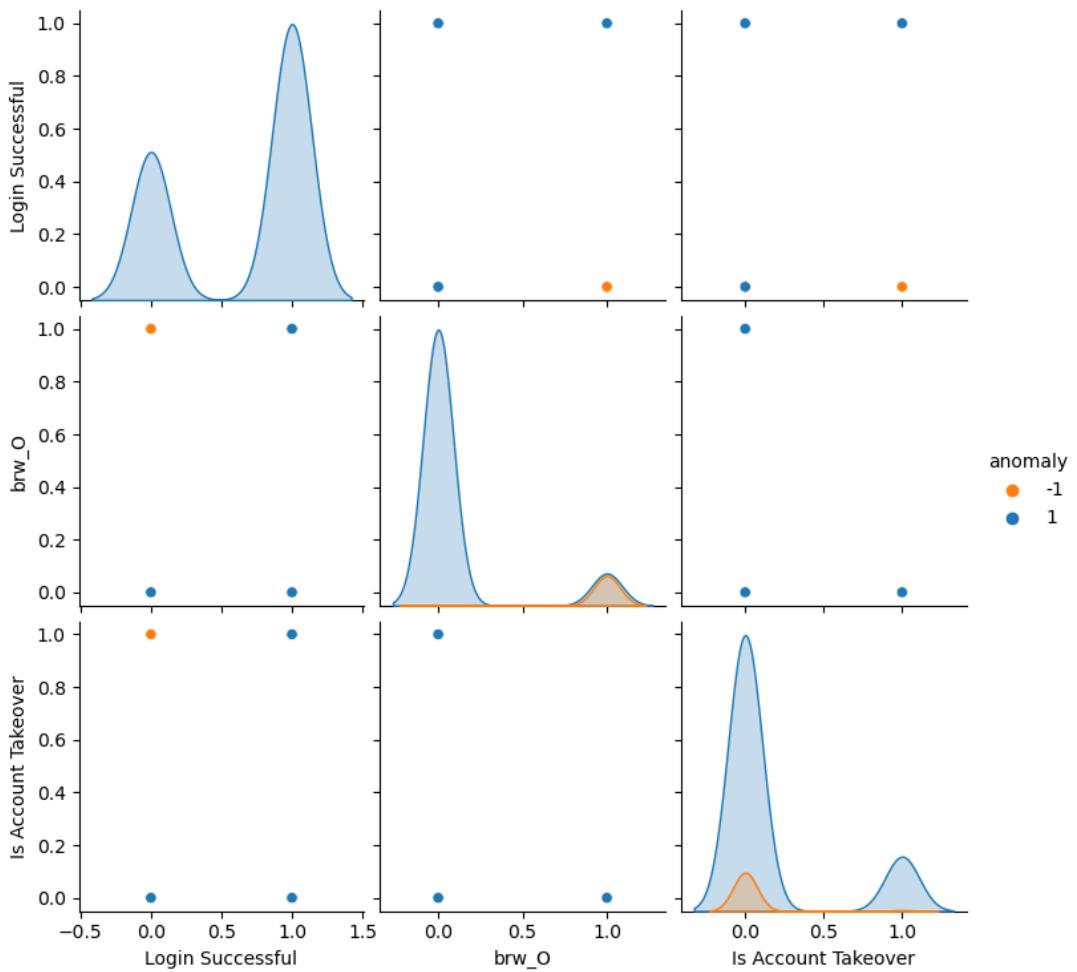
Outlier Method: Isolation Forest



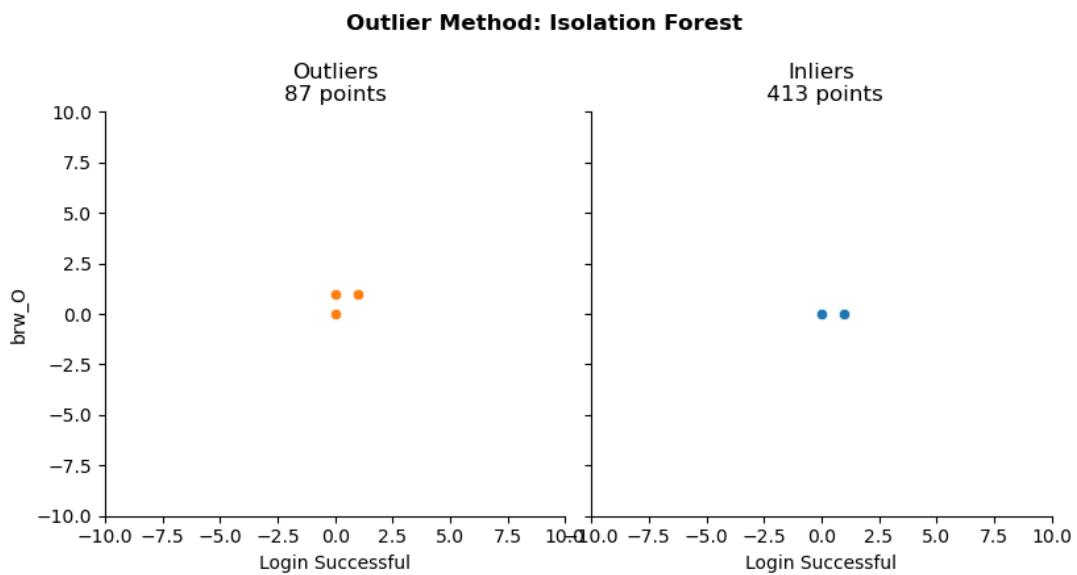


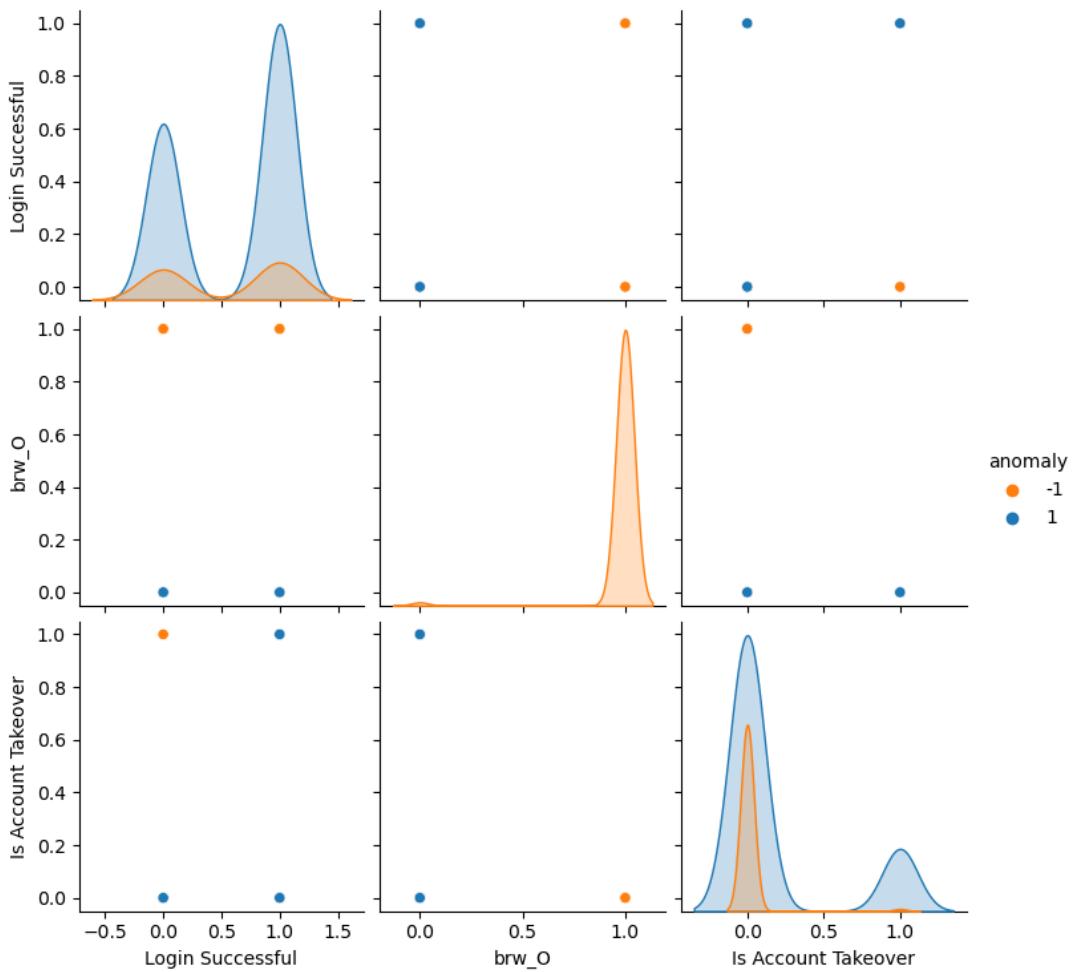
- anomaly_inputs = ['Login Successful', 'Browser Type (brw_O)', 'Is Account Takeover']
 - contamination value == 0.1





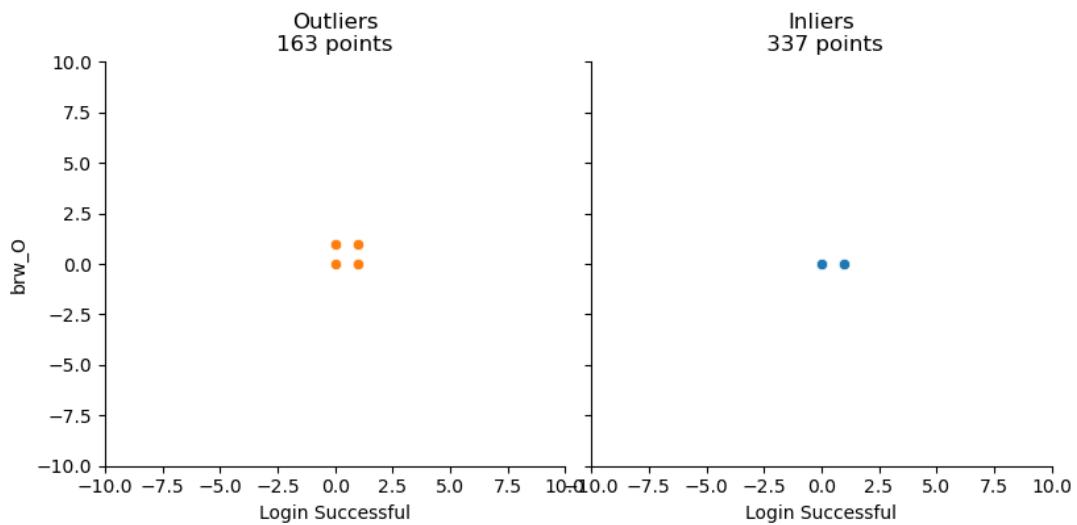
- contamination value == 0.3

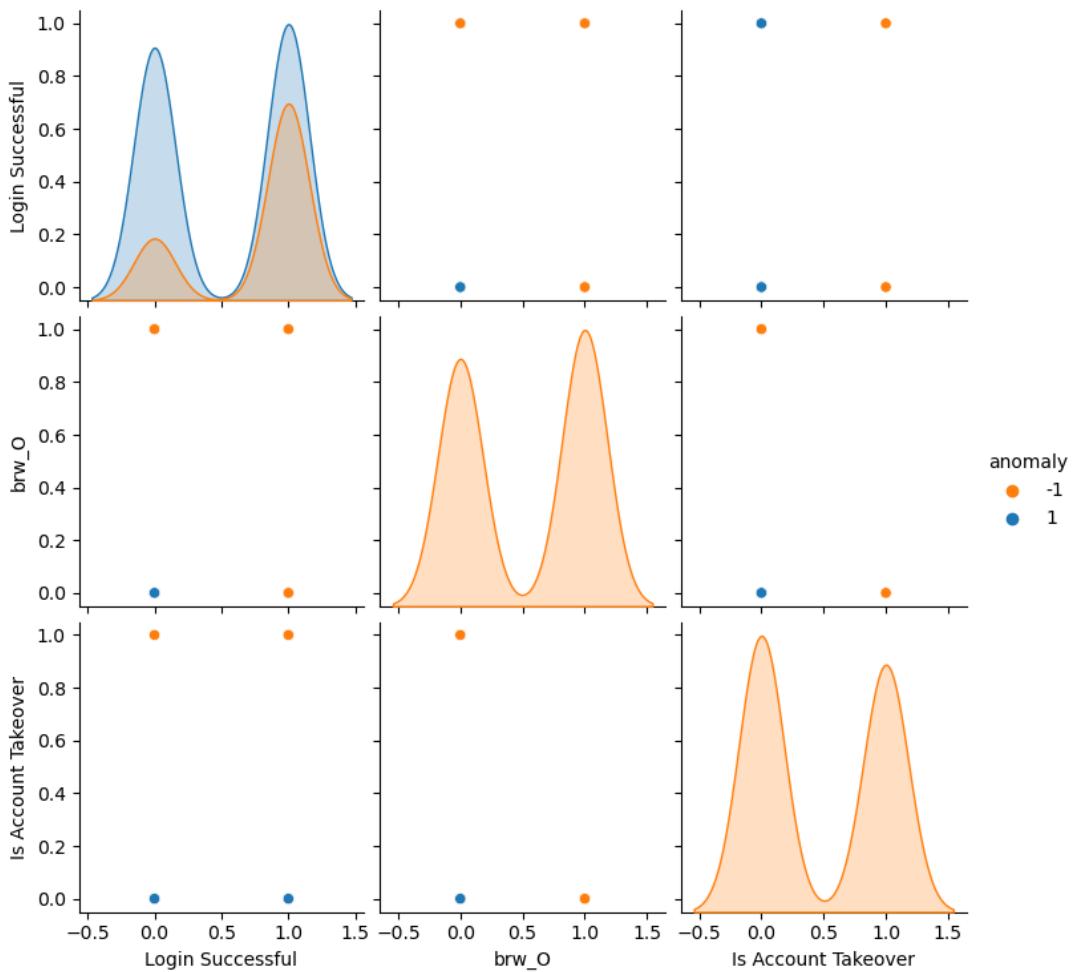




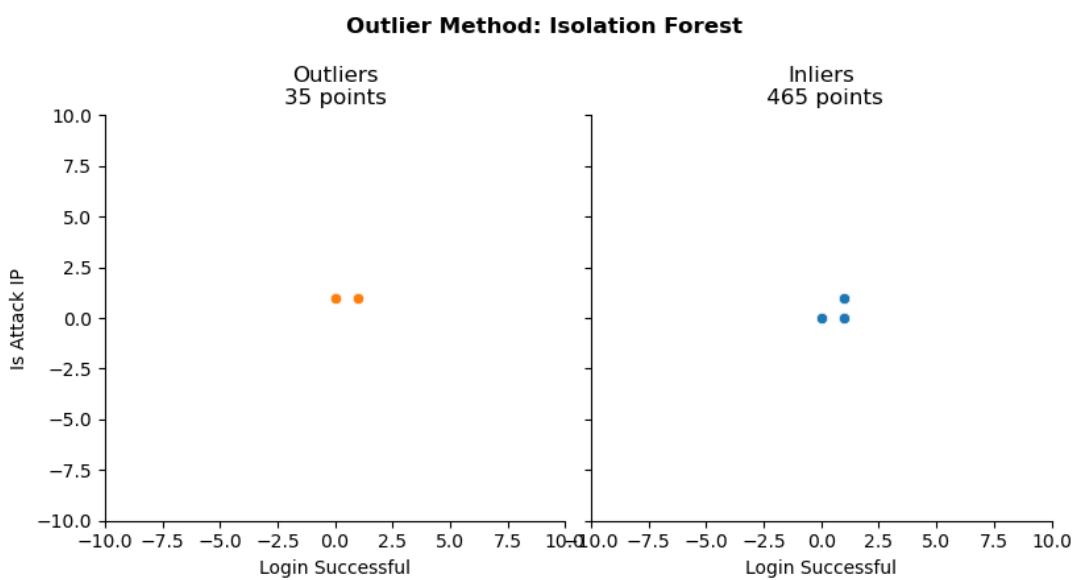
- contamination value == 0.5

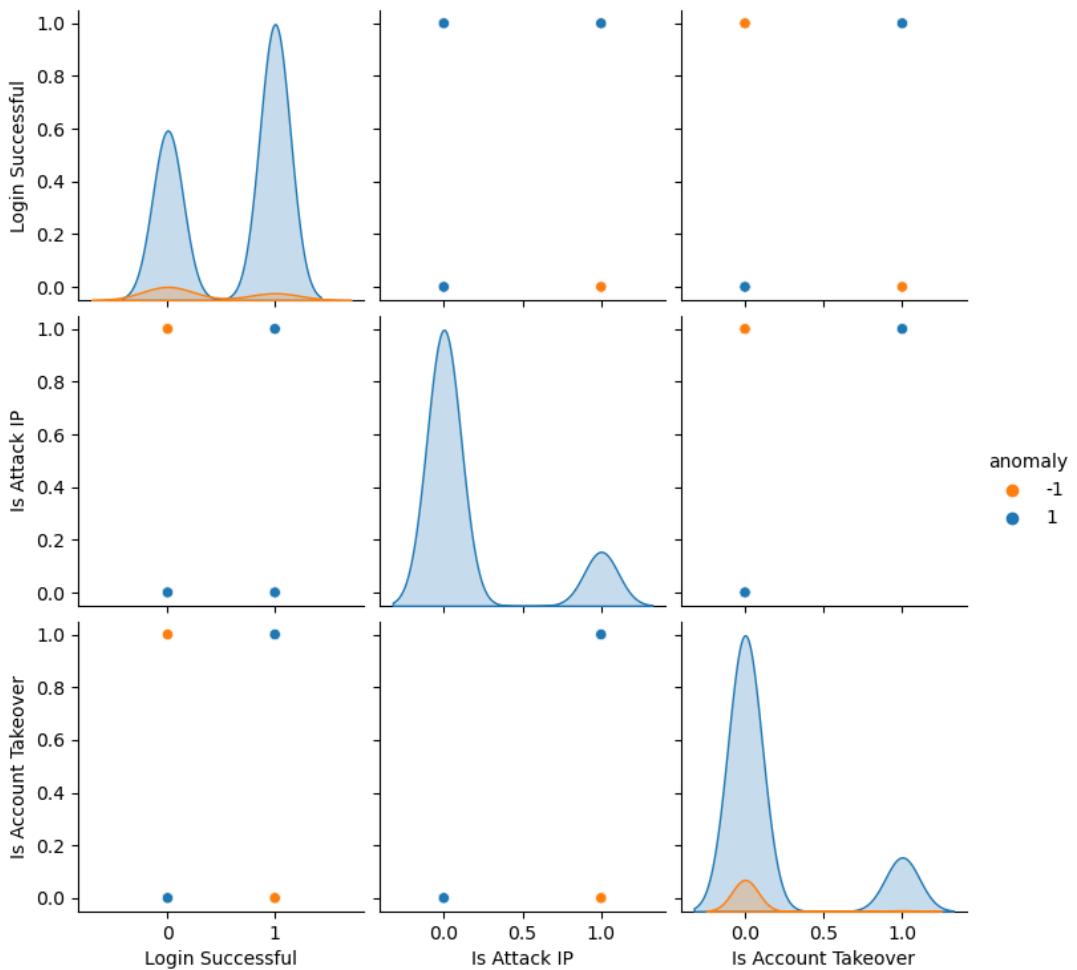
Outlier Method: Isolation Forest



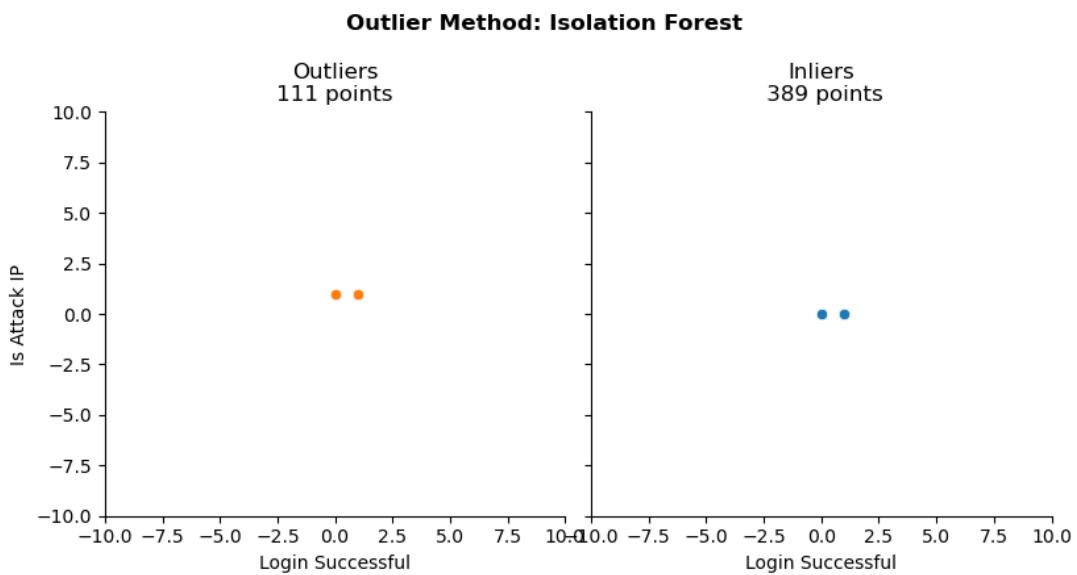


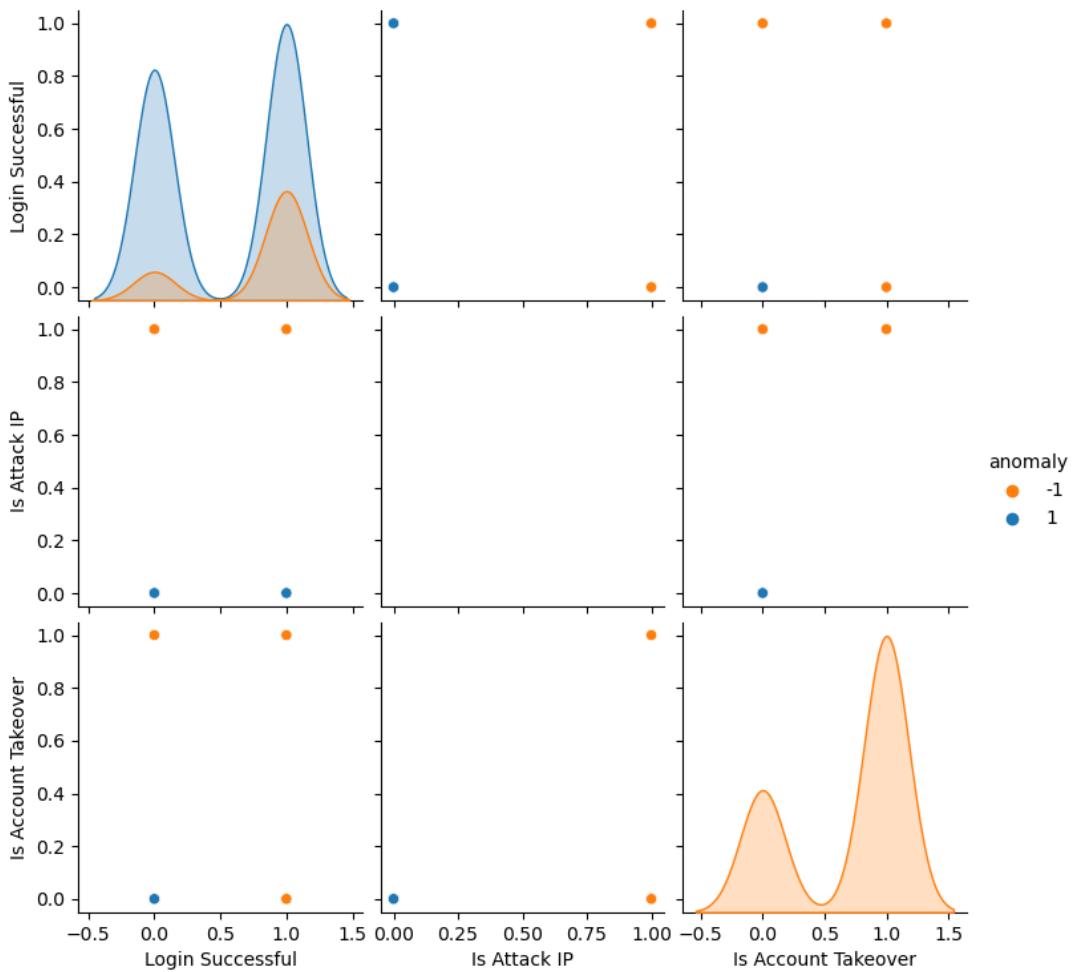
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1





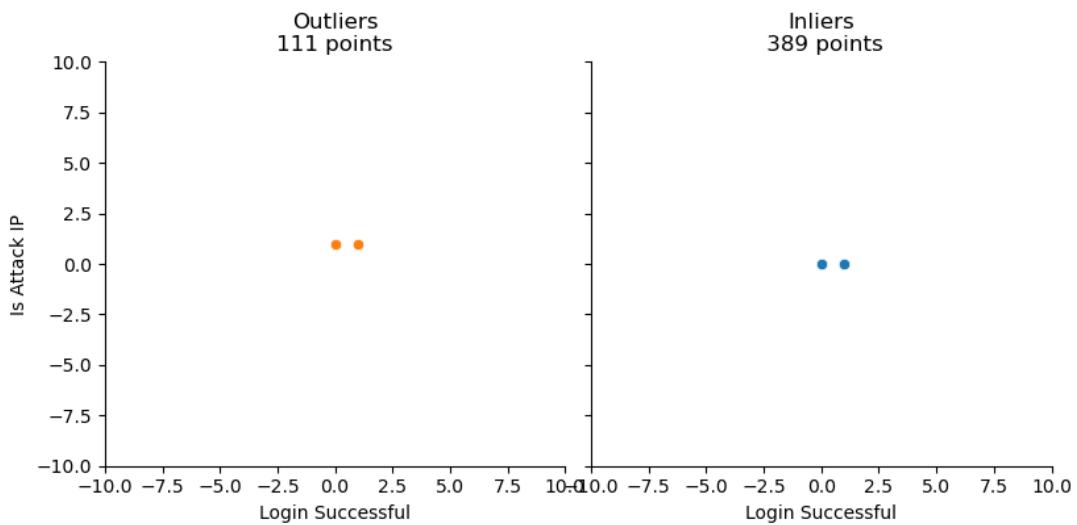
- contamination value == 0.3

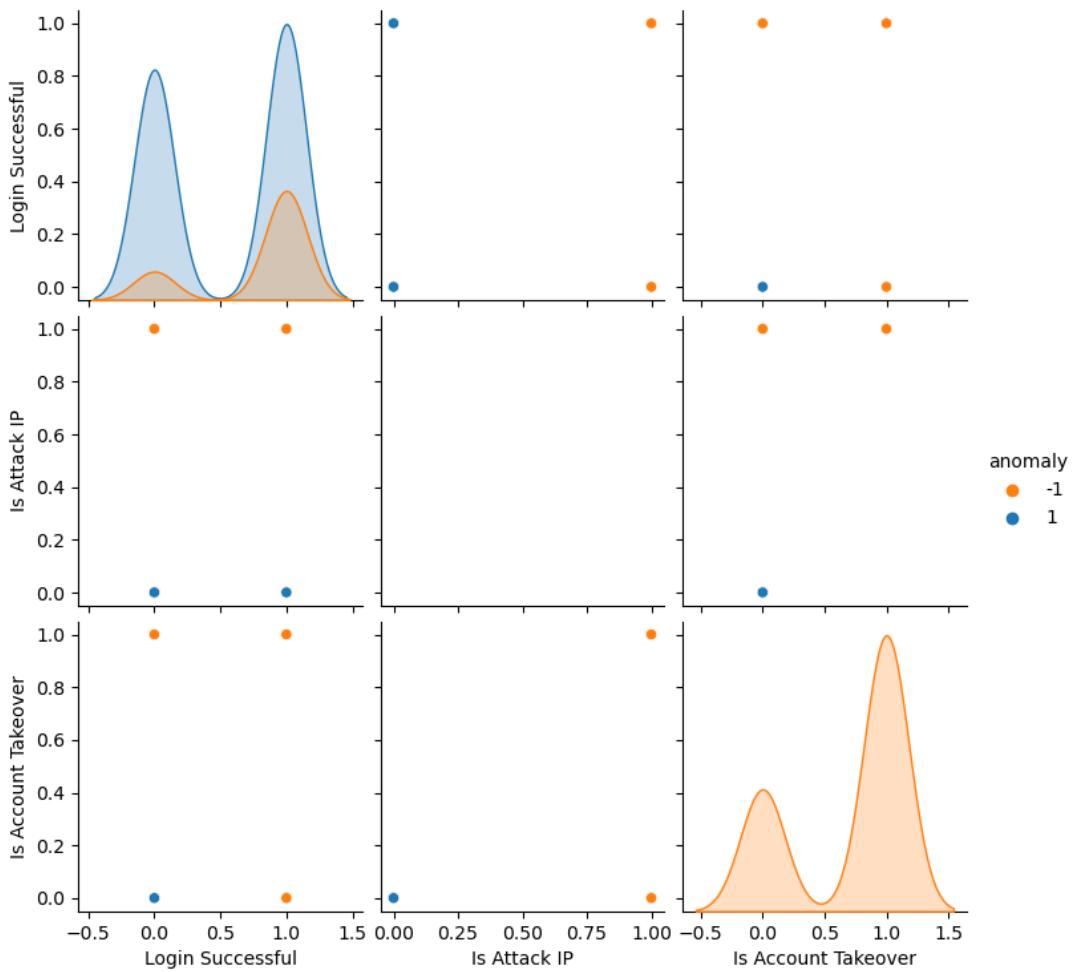




- contamination value == 0.5

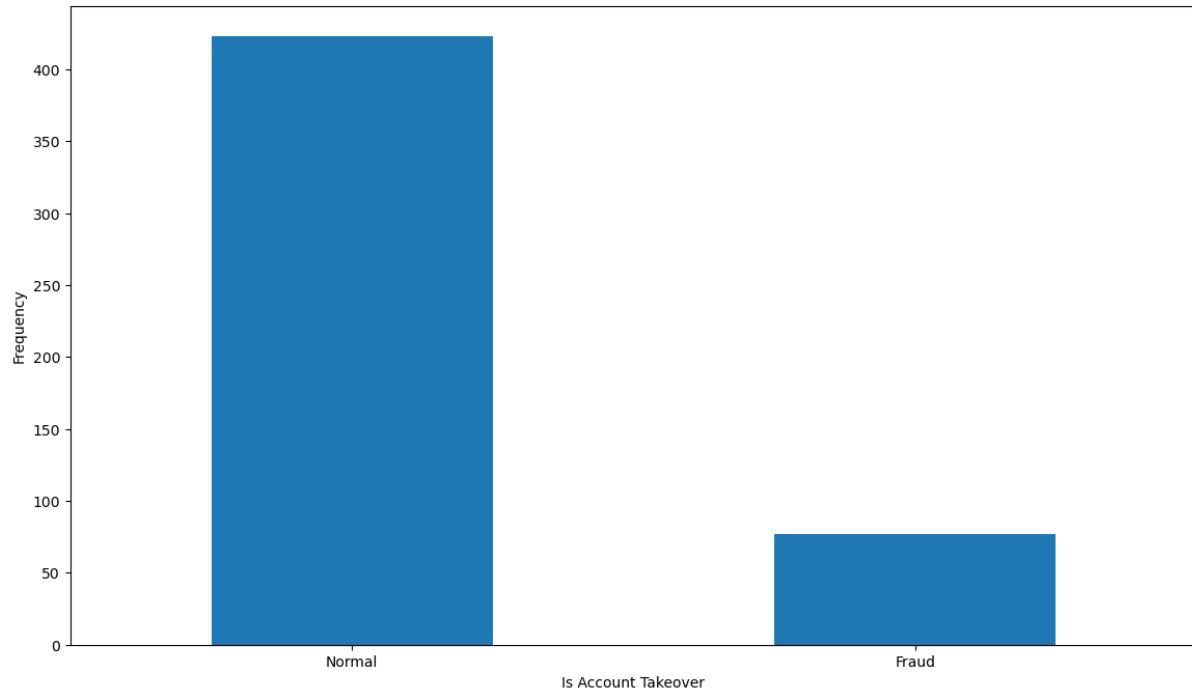
Outlier Method: Isolation Forest





3D. iForest, Local Outlier Factor, Support Vector Machine

Account Takeover Class Distribution



Isolation Forest: 107

Accuracy Score :

0.786

Classification Report :

	precision	recall	f1-score	support
False	0.85	0.91	0.88	423
True	0.20	0.13	0.16	77
accuracy			0.79	500
macro avg	0.53	0.52	0.52	500
weighted avg	0.75	0.79	0.77	500

Local Outlier Factor: 111

Accuracy Score :

0.778

Classification Report :

	precision	recall	f1-score	support
False	0.85	0.90	0.87	423
True	0.16	0.10	0.13	77
accuracy			0.78	500
macro avg	0.50	0.50	0.50	500
weighted avg	0.74	0.78	0.76	500

Support Vector Machine: 343

Accuracy Score :

0.314

Classification Report :

	precision	recall	f1-score	support
False	0.79	0.26	0.39	423
True	0.13	0.61	0.22	77
accuracy			0.31	500
macro avg	0.46	0.44	0.30	500
weighted avg	0.68	0.31	0.36	500