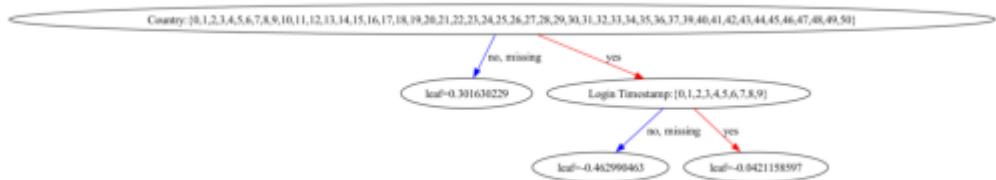


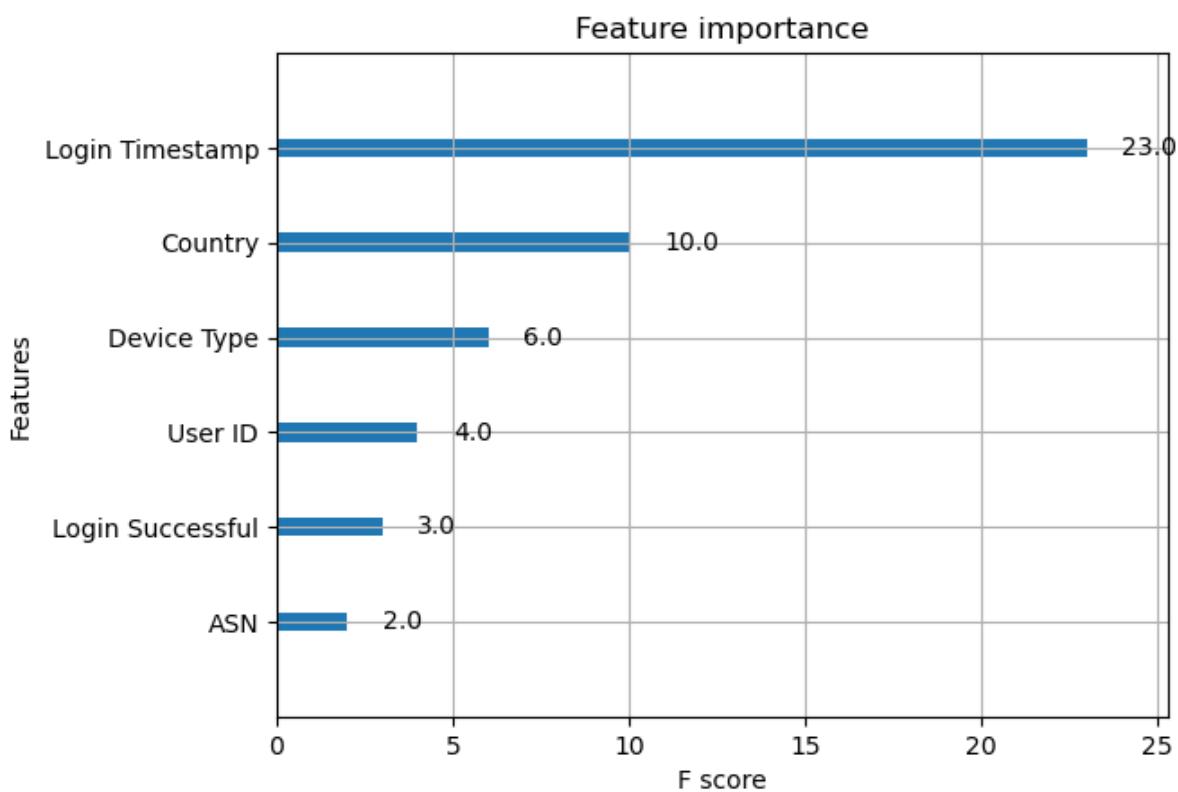
ATO 1000 rows

I. 10/1000 → 1%

1. XGB Experimental

```
--- -----  
0 ASN           1000 non-null  float64  
1 Country        1000 non-null  category  
2 Device Type   1000 non-null  category  
3 IP Address    1000 non-null  category  
4 Is Attack IP  1000 non-null  bool  
5 Login Successful 1000 non-null  bool  
6 Login Timestamp 1000 non-null  category  
7 User ID        1000 non-null  float64  
8 Blacklisted    1000 non-null  bool  
9 Browser Type   1000 non-null  category  
10 Is Account Takeover 1000 non-null  bool  
dtypes: bool(4), category(5), float64(2)  
memory usage: 90.7 KB  
Feature importances:  
[0.13135657 0.49352363 0.07387897 0.          0.          0.05633408  
 0.20868494 0.03622184 0.          0.          ]
```





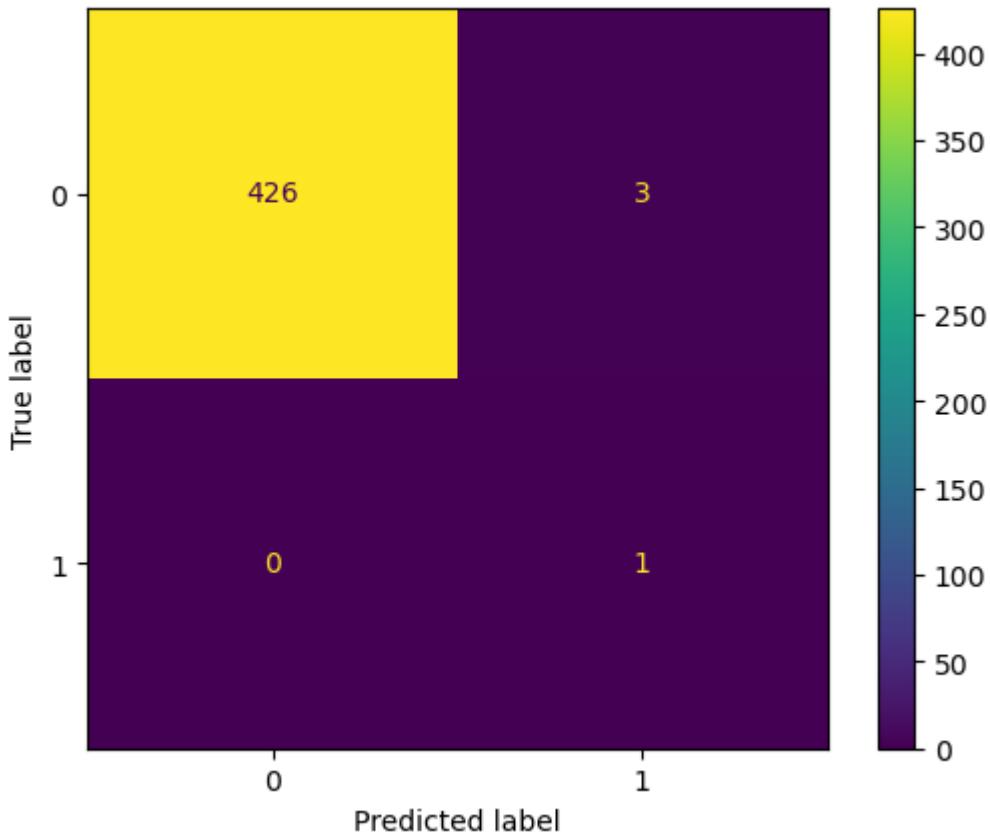
2. Label encoding (cat.codes) test_size increase 0.1 → 0.43 or "UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 due to no predicted samples. Use `zero_division` parameter to control this behavior."

2A. XGBoost

memory usage: 26.5 KB

Accuracy: 99.30%				
	precision	recall	f1-score	support
False	1.00	0.99	1.00	429
True	0.25	1.00	0.40	1
accuracy			0.99	430
macro avg	0.62	1.00	0.70	430
weighted avg	1.00	0.99	1.00	430

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 800
The number of records in the test dataset is 200
The training dataset has 794 records for the majority class and 6 records for the minority class.

	precision	recall	f1-score	support
False	0.00	0.00	0.00	196
True	0.02	1.00	0.04	4
accuracy			0.02	200
macro avg	0.01	0.50	0.02	200
weighted avg	0.00	0.02	0.00	200

The customized score threshold for 2% of outliers is 0.00

	precision	recall	f1-score	support
False	0.98	1.00	0.99	196
True	0.00	0.00	0.00	4
accuracy			0.98	200
macro avg	0.49	0.50	0.49	200
weighted avg	0.96	0.98	0.97	200

```
[[ 0 196]
 [ 0  4]]
```

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

```
The number of records in the training dataset is 700
The number of records in the test dataset is 300
The training dataset has 695 records for the majority class and 5 records for the minority class.

precision    recall   f1-score   support

      False       0.00      0.00      0.00      295
       True       0.02      1.00      0.03        5

accuracy                           0.02      300
macro avg       0.01      0.50      0.02      300
weighted avg    0.00      0.02      0.00      300

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

      False       0.98      1.00      0.99      295
       True       0.00      0.00      0.00        5

accuracy                           0.98      300
macro avg       0.49      0.50      0.50      300
weighted avg    0.97      0.98      0.98      300

[[ 0 295]
 [ 0  5]]
```

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

```
The number of records in the training dataset is 500
The number of records in the test dataset is 500
The training dataset has 498 records for the majority class and 2 records for the minority class.

precision    recall   f1-score   support

      False       0.00      0.00      0.00      492
       True       0.02      1.00      0.03        8

accuracy                           0.02      500
macro avg       0.01      0.50      0.02      500
weighted avg    0.00      0.02      0.00      500

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

      False       0.98      1.00      0.99      492
       True       0.00      0.00      0.00        8

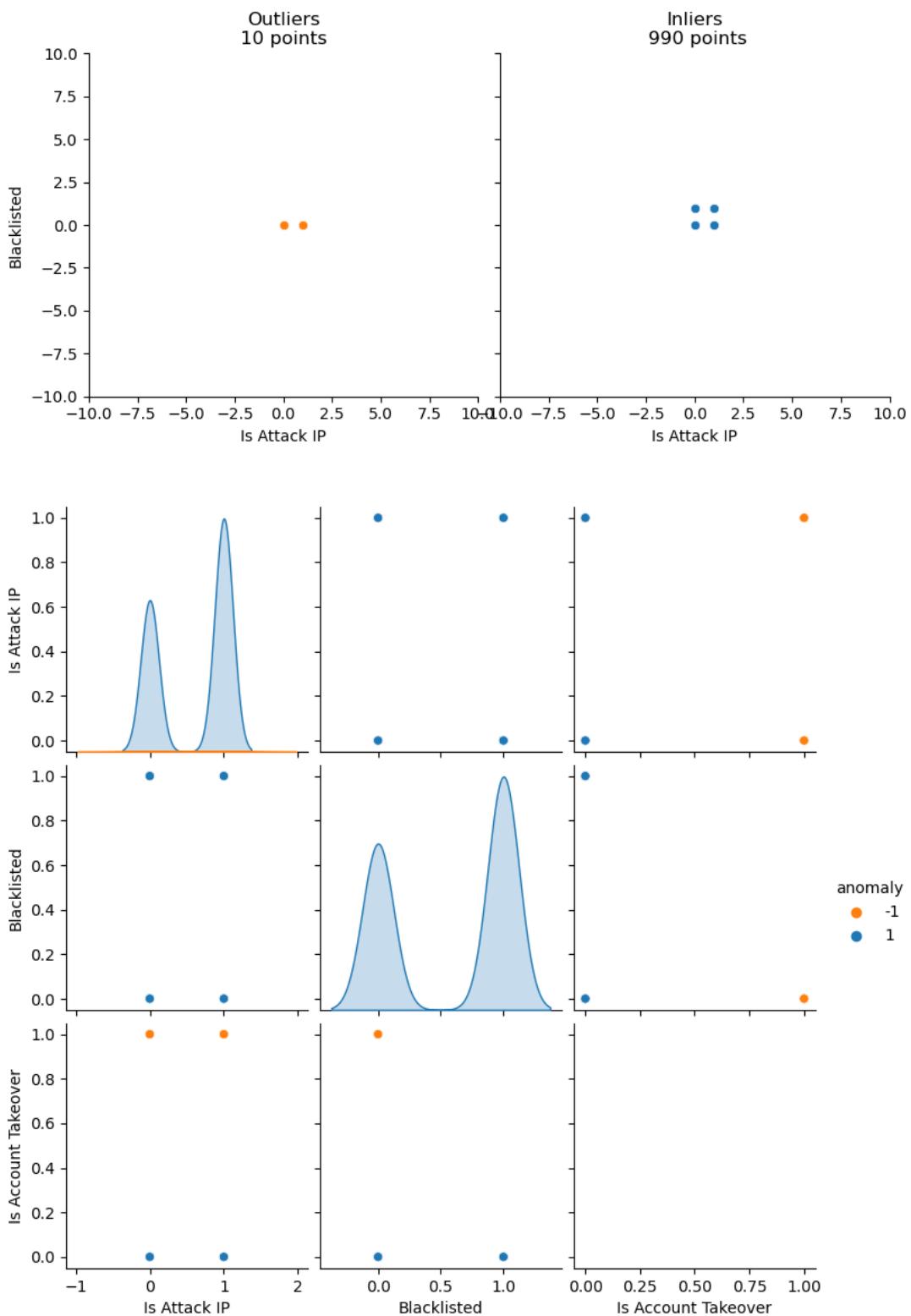
accuracy                           0.98      500
macro avg       0.49      0.50      0.50      500
weighted avg    0.97      0.98      0.98      500

[[ 0 492]
 [ 0  8]]
```

2C. Isolation Forest

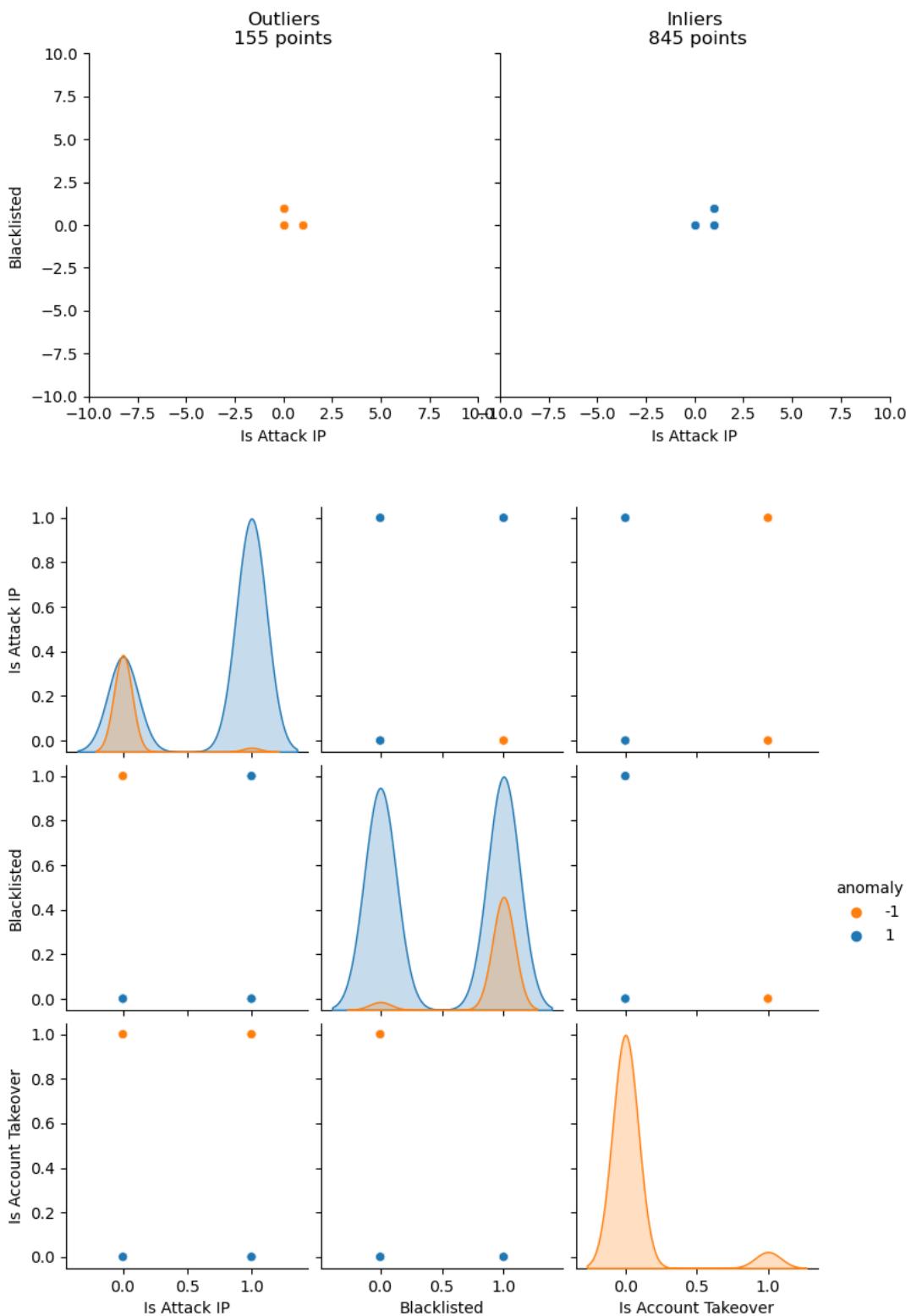
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1

Outlier Method: Isolation Forest



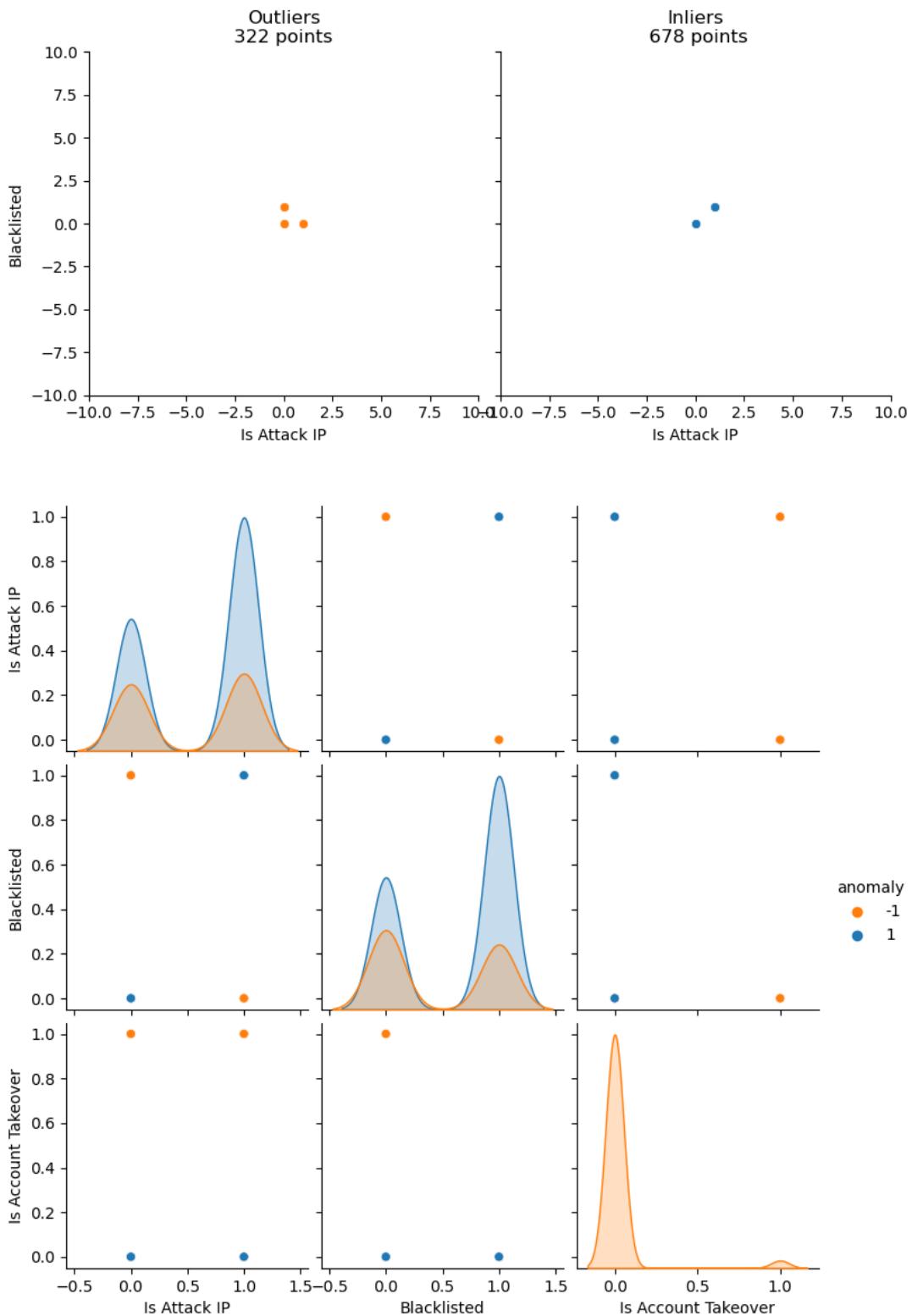
- increasing contamination value to 0.3

Outlier Method: Isolation Forest



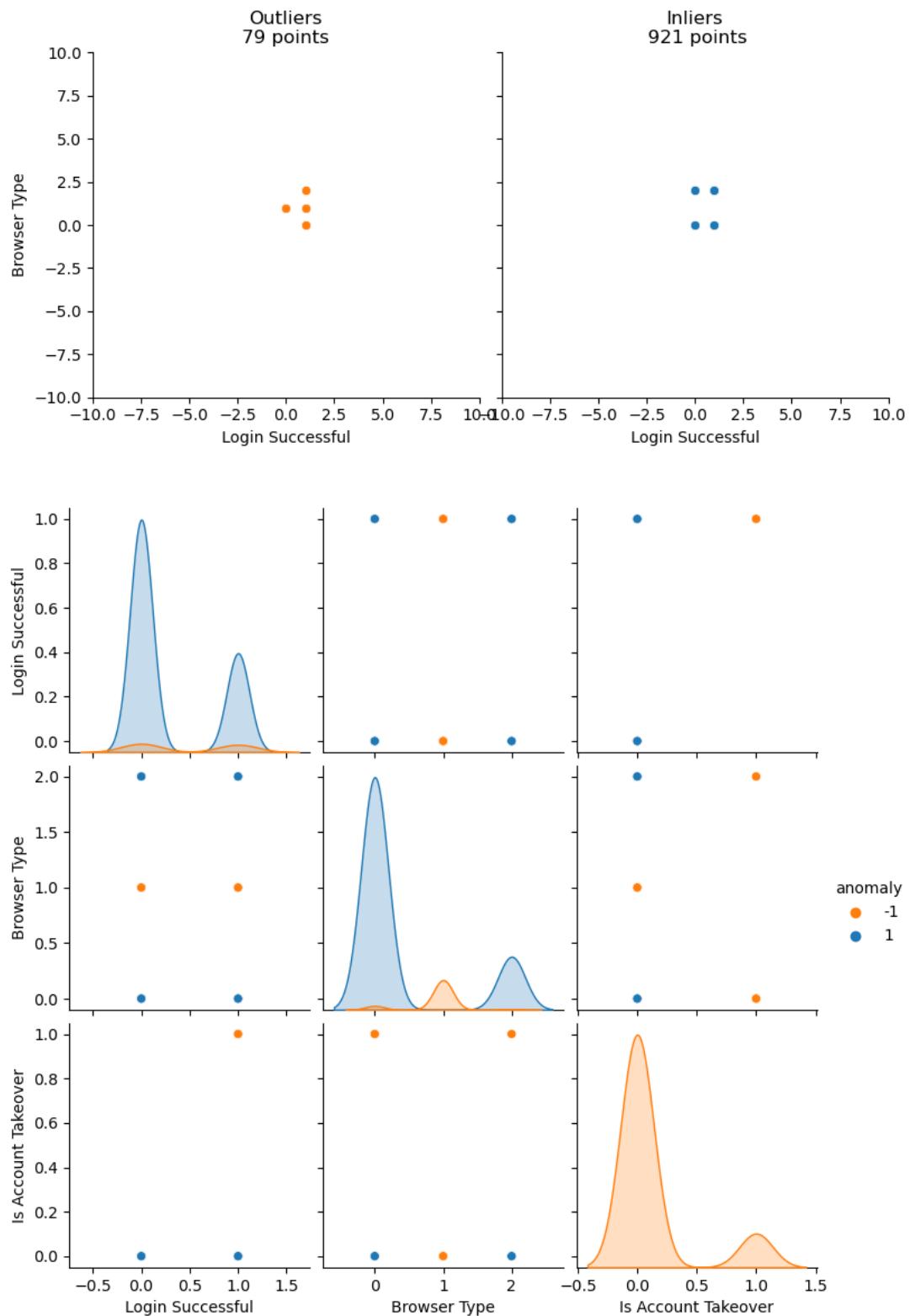
- contamination value == 0.5

Outlier Method: Isolation Forest



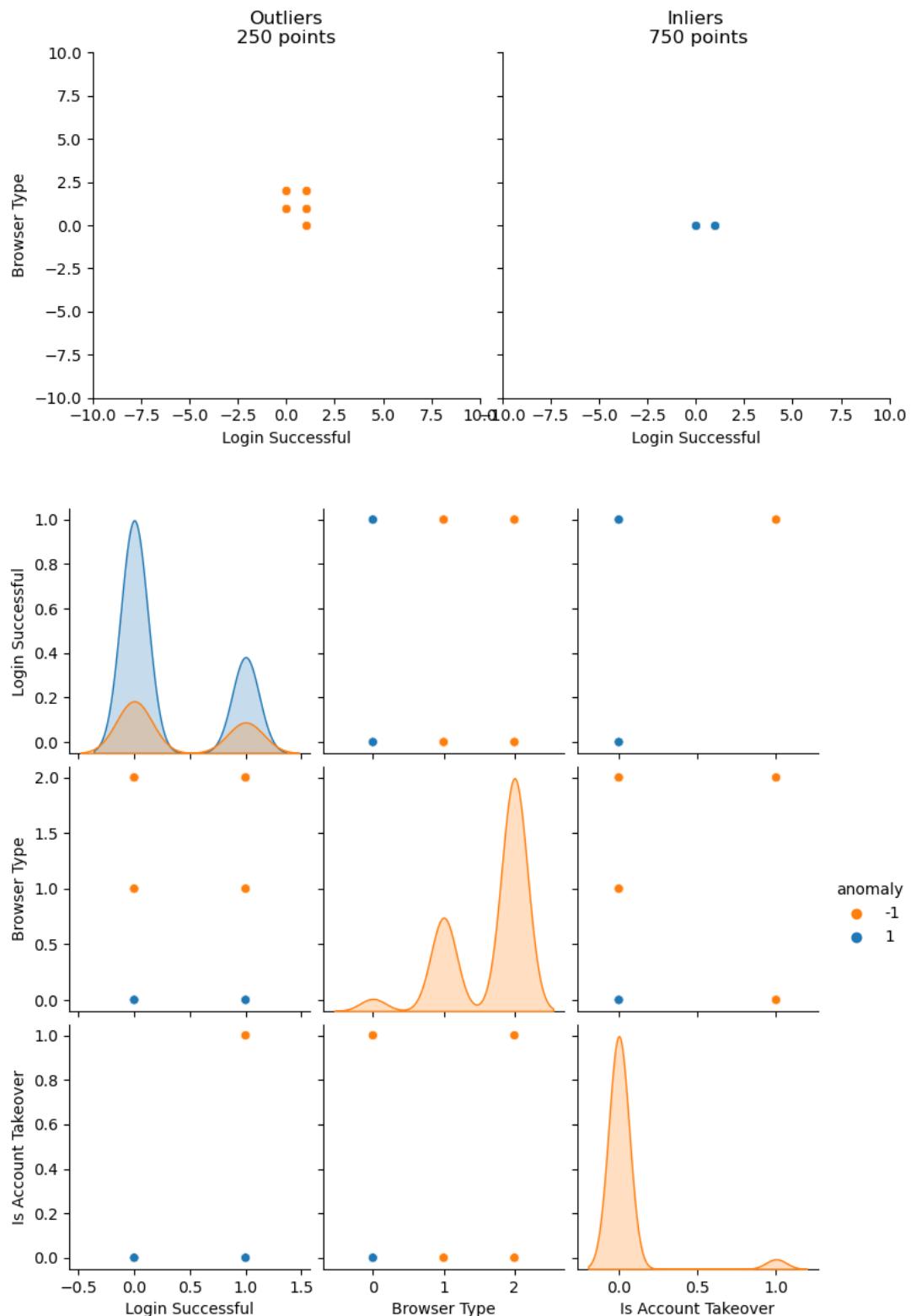
- `anomaly_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']`
 - contamination value == 0.1

Outlier Method: Isolation Forest



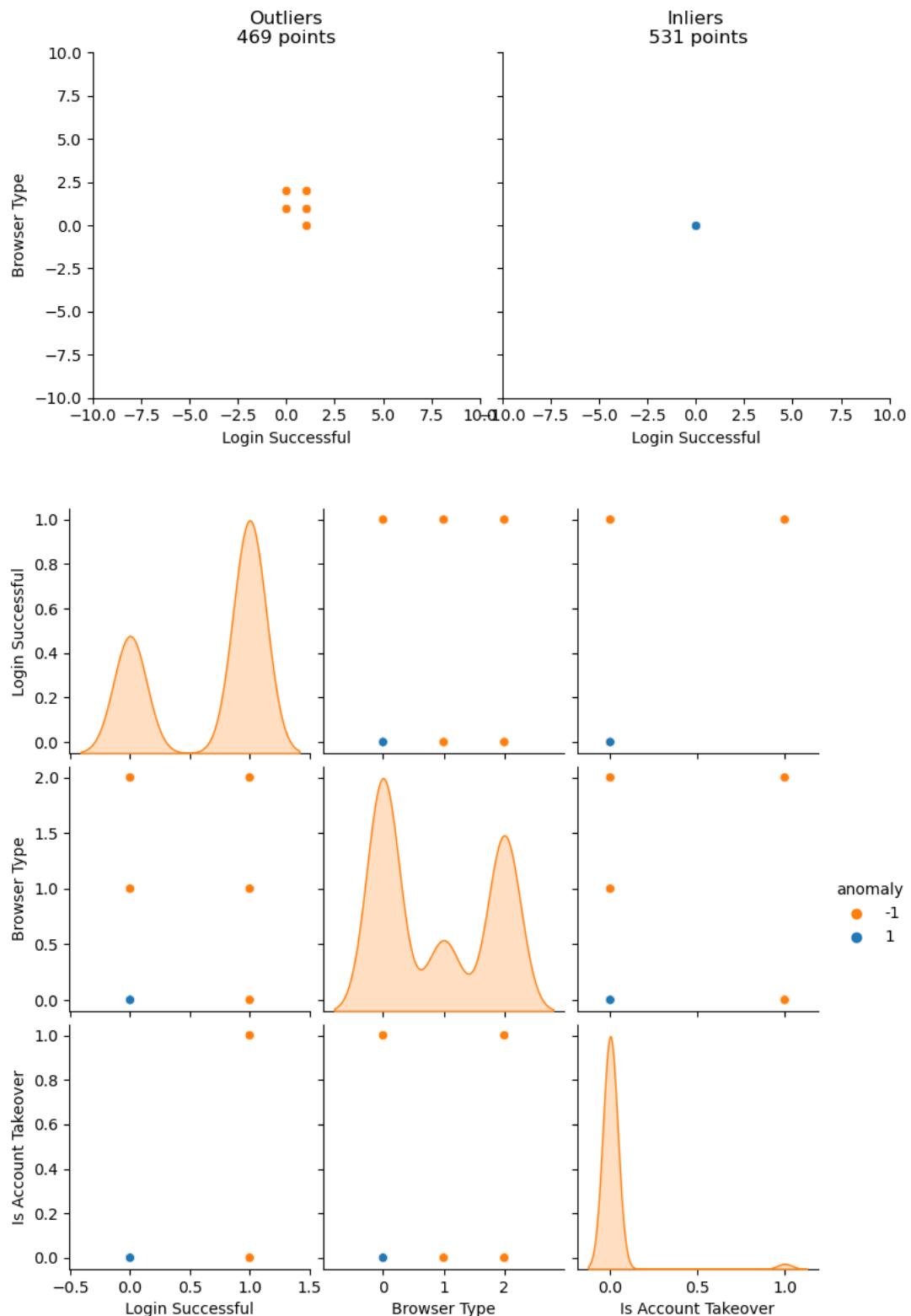
- contamination value == 0.3

Outlier Method: Isolation Forest



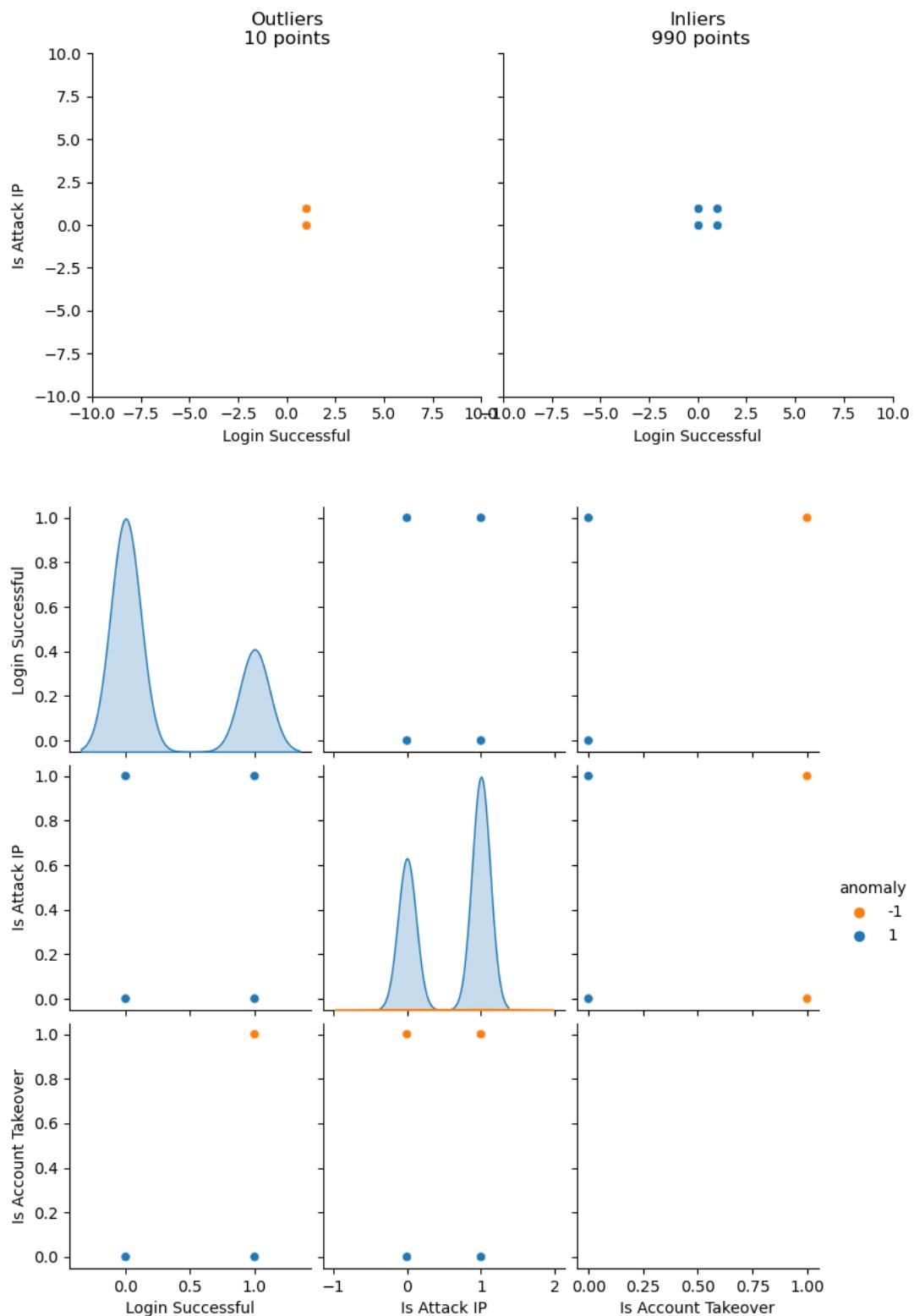
- contamination value == 0.5

Outlier Method: Isolation Forest



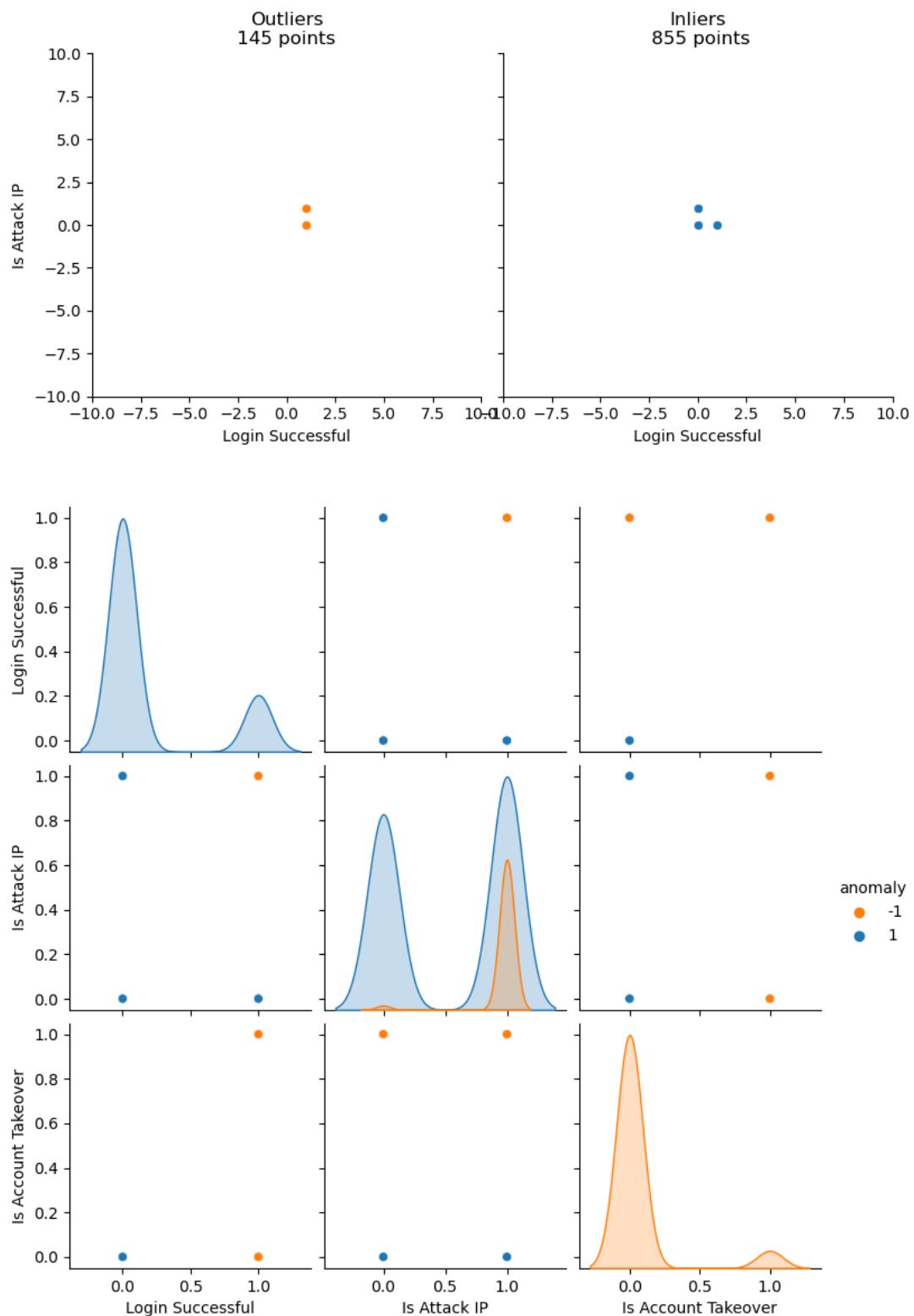
- `anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']`
 - contamination value == 0.1

Outlier Method: Isolation Forest



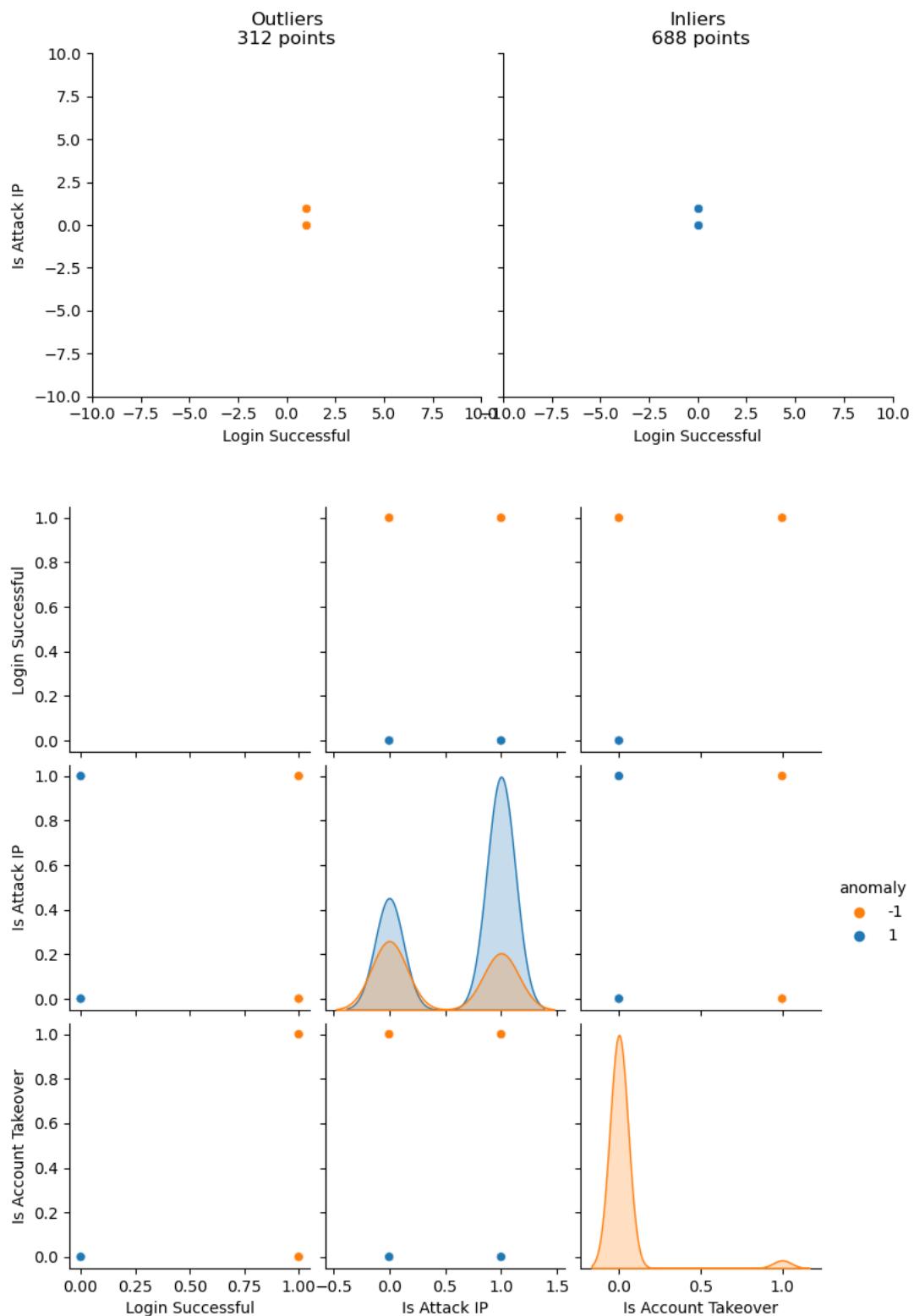
- contamination value == 0.3

Outlier Method: Isolation Forest

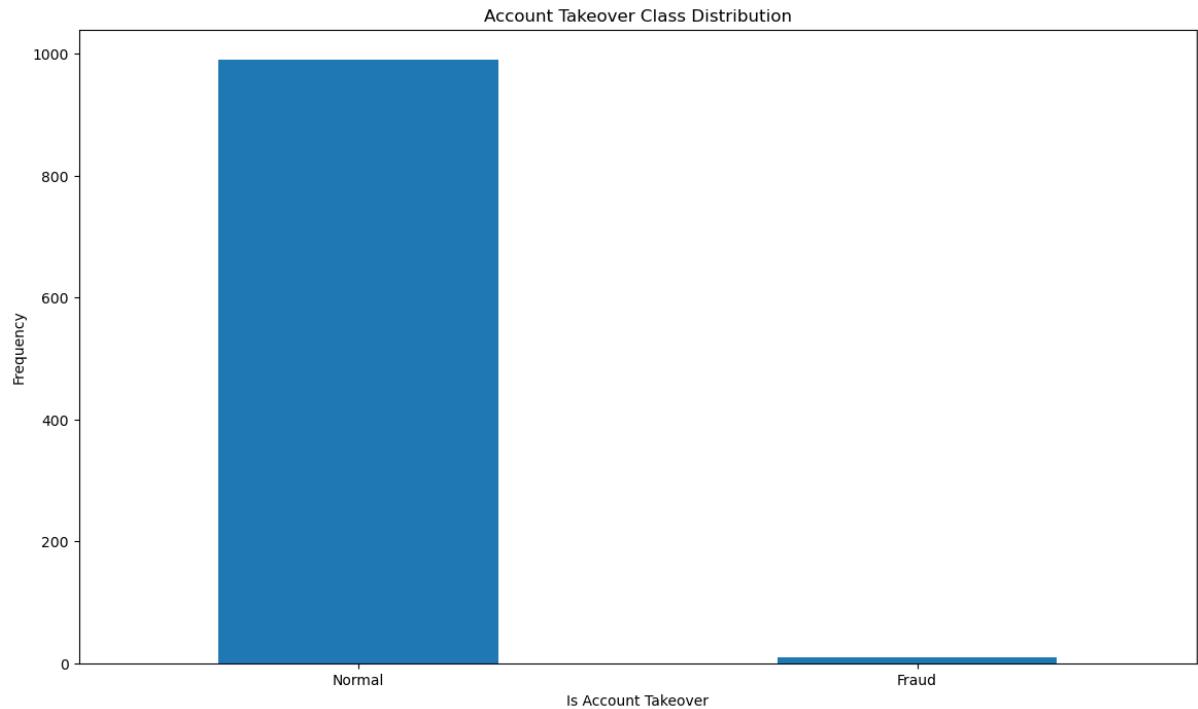


- contamination value == 0.5

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 :

0.979

Classification Report :

	precision	recall	f1-score	support
False	0.99	0.99	0.99	990
True	0.00	0.00	0.00	10
accuracy			0.98	1000
macro avg	0.49	0.49	0.49	1000
weighted avg	0.98	0.98	0.98	1000

Local Outlier Factor: 19

Accuracy Score :

0.981

Classification Report :

	precision	recall	f1-score	support
False	0.99	0.99	0.99	990
True	0.09	0.10	0.10	10
accuracy			0.98	1000
macro avg	0.54	0.54	0.54	1000
weighted avg	0.98	0.98	0.98	1000

Support Vector Machine: 743

Accuracy Score :

0.257

Classification Report :

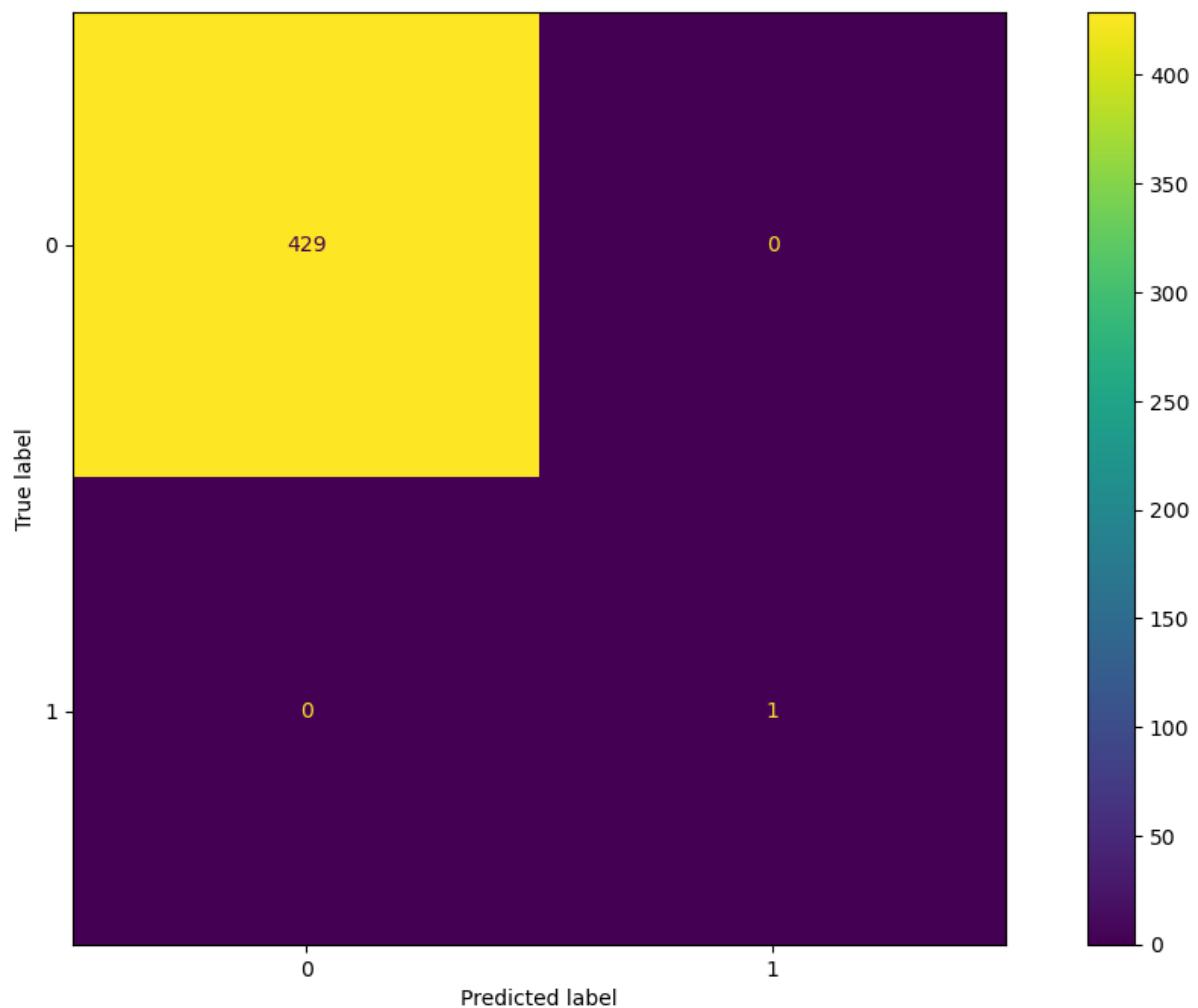
	precision	recall	f1-score	support
False	0.99	0.25	0.40	990
True	0.01	0.70	0.02	10
accuracy			0.26	1000
macro avg	0.50	0.48	0.21	1000
weighted avg	0.98	0.26	0.40	1000

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: 1.7 MB

Accuracy: 100.00%				
	precision	recall	f1-score	support
False	1.00	1.00	1.00	429
True	1.00	1.00	1.00	1
accuracy			1.00	430
macro avg	1.00	1.00	1.00	430
weighted avg	1.00	1.00	1.00	430



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 800
The number of records in the test dataset is 200
The training dataset has 794 records for the majority class and 6 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.08     0.15      196
      True         0.02     1.00     0.04       4
      accuracy          0.10      200
      macro avg       0.51     0.54     0.10      200
      weighted avg    0.98     0.10     0.15      200
      The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.98     1.00     0.99      196
      True         0.00     0.00     0.00       4
      accuracy          0.98      200
      macro avg       0.49     0.50     0.49      200
      weighted avg    0.96     0.98     0.97      200
[[ 16 180]
 [ 0  4]]

```

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

```

The number of records in the training dataset is 700
The number of records in the test dataset is 300
The training dataset has 695 records for the majority class and 5 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.23     0.37      295
      True         0.02     1.00     0.04       5
      accuracy          0.24      300
      macro avg       0.51     0.62     0.21      300
      weighted avg    0.98     0.24     0.37      300
      The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.98     1.00     0.99      295
      True         0.00     0.00     0.00       5
      accuracy          0.98      300
      macro avg       0.49     0.50     0.50      300
      weighted avg    0.97     0.98     0.98      300
[[ 68 227]
 [ 0  5]]

```

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

```

The number of records in the training dataset is 500
The number of records in the test dataset is 500
The training dataset has 498 records for the majority class and 2 records for the minority class.
      precision    recall   f1-score   support
False        1.00     0.05     0.09     492
True        0.02     1.00     0.03      8

accuracy          0.06     500
macro avg       0.51     0.52     0.06     500
weighted avg    0.98     0.06     0.09     500

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.98     1.00     0.99     492
True        0.00     0.00     0.00      8

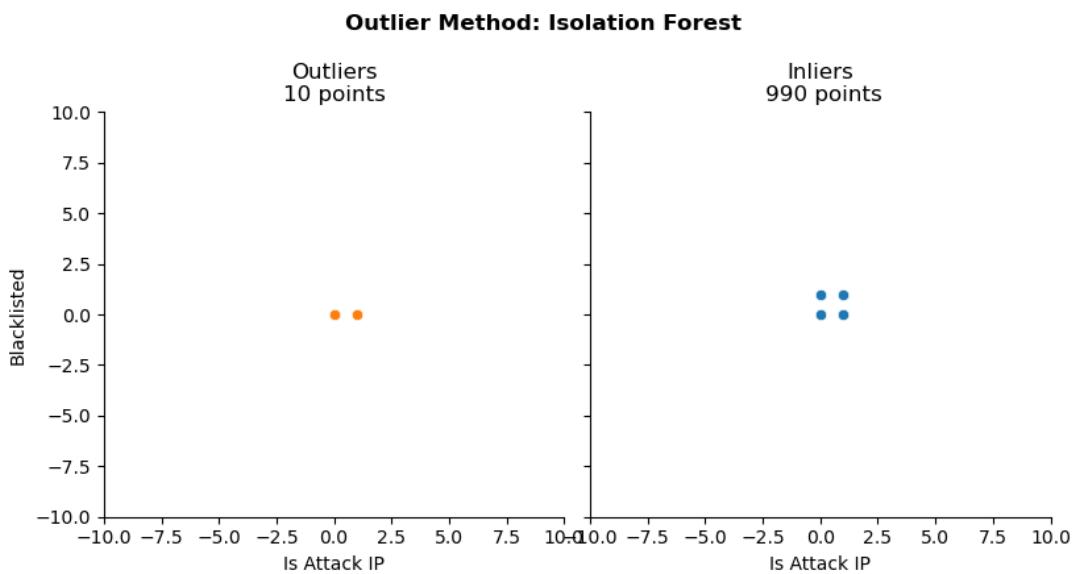
accuracy          0.98     500
macro avg       0.49     0.50     0.50     500
weighted avg    0.97     0.98     0.98     500

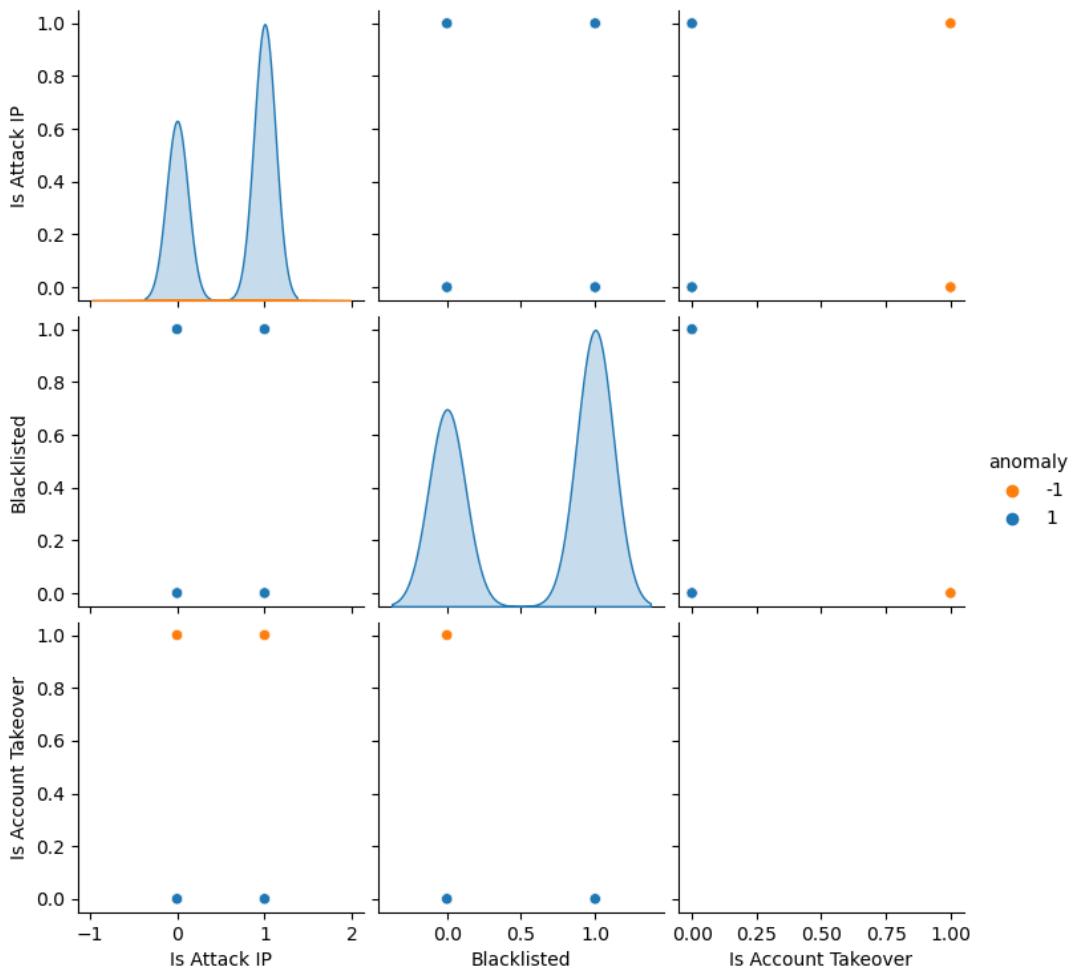
[[ 23 469]
 [ 0  8]]

```

3C. Isolation Forest

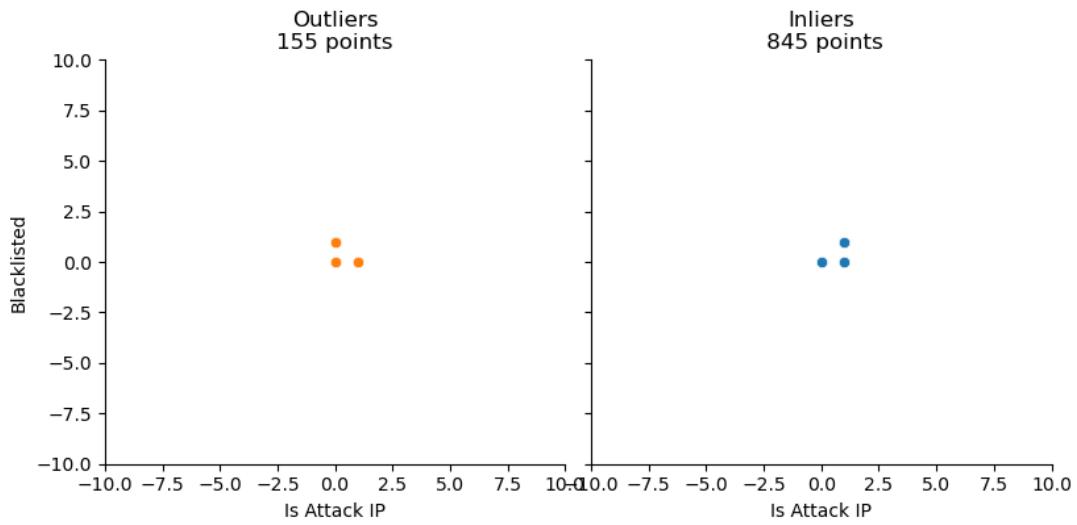
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1

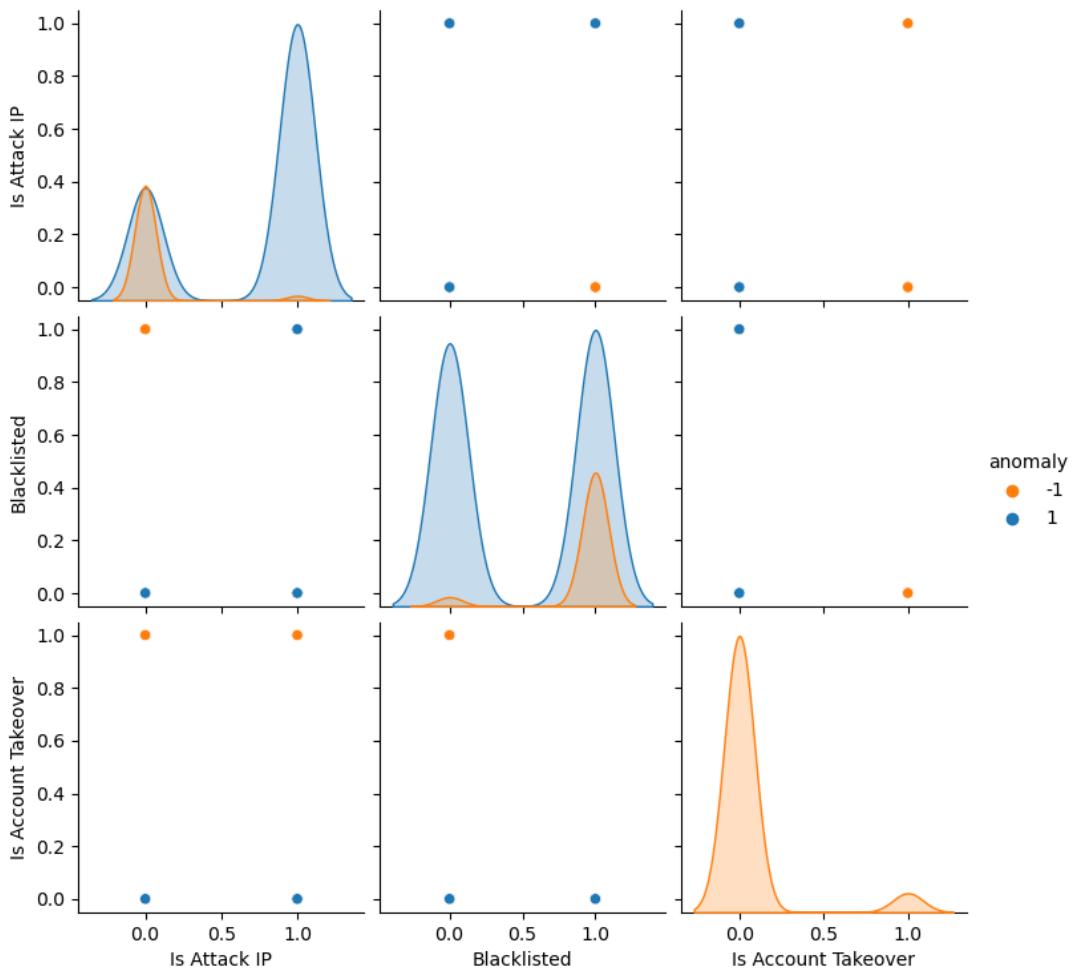




- contamination value == 0.3

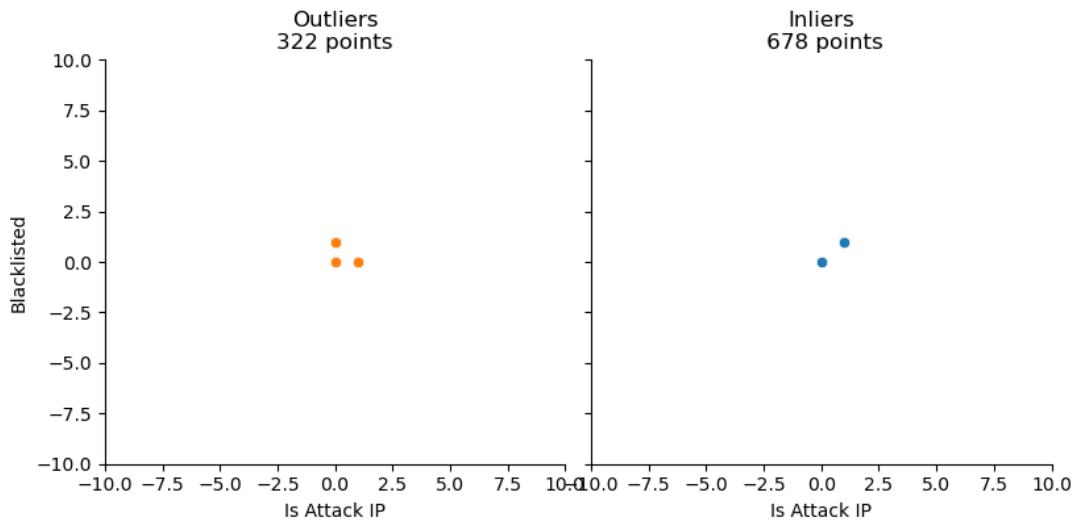
Outlier Method: Isolation Forest

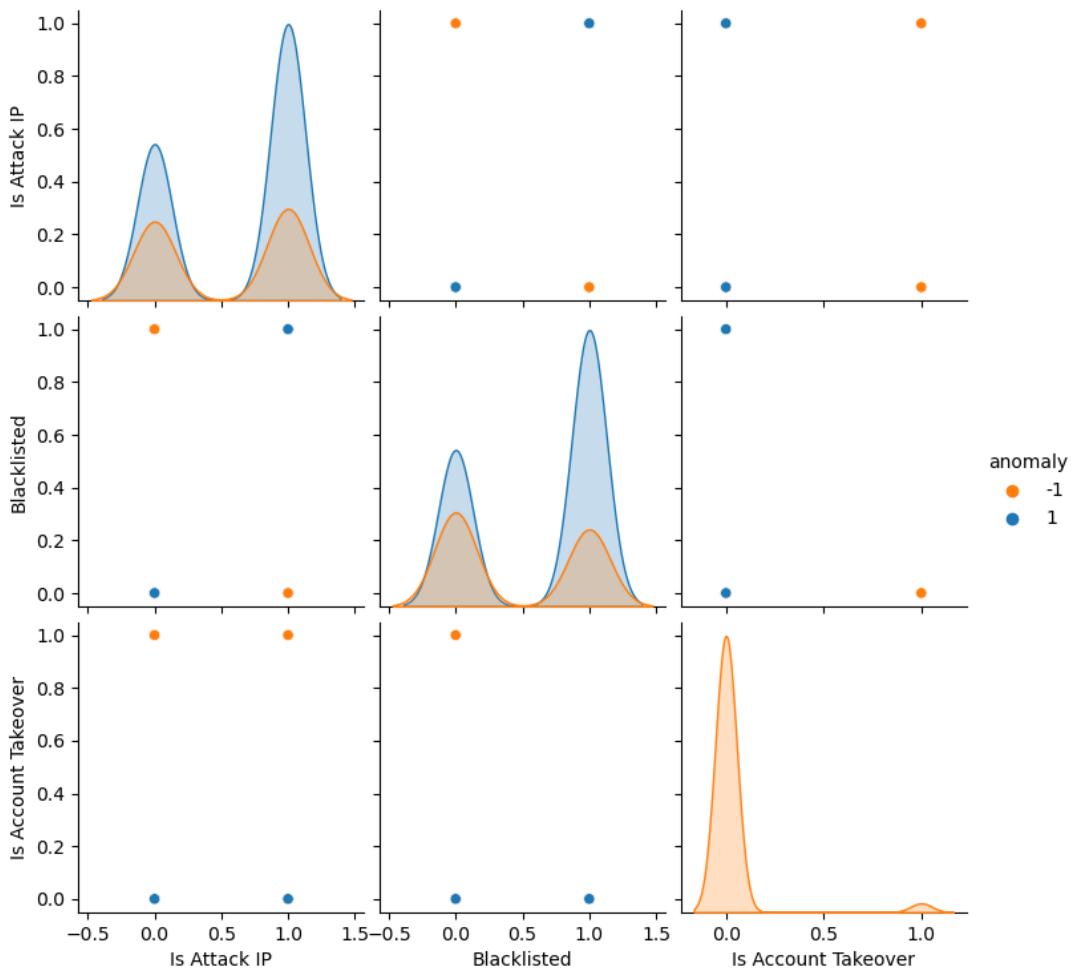




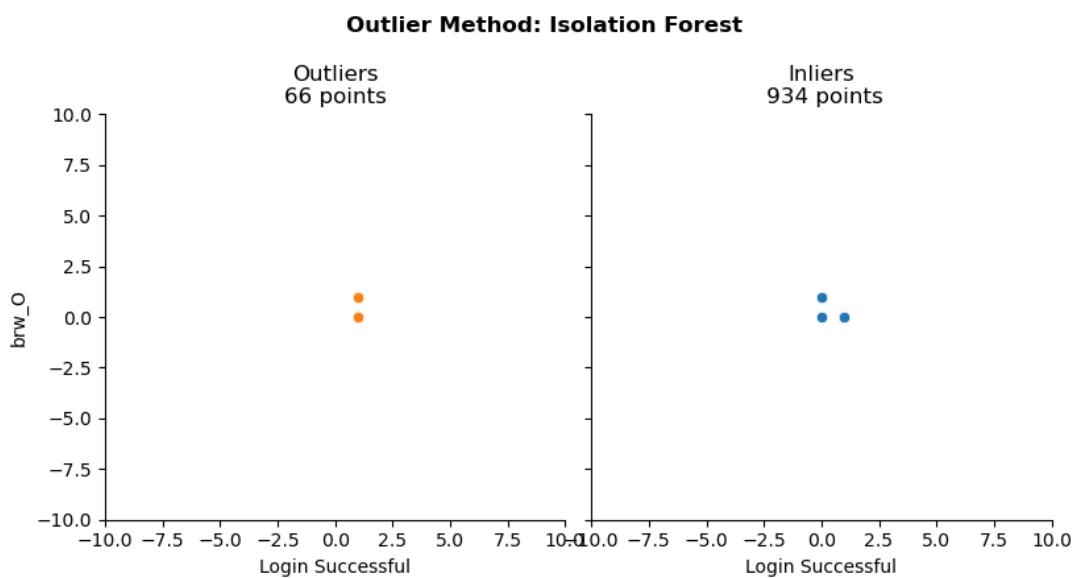
- contamination value == 0.5

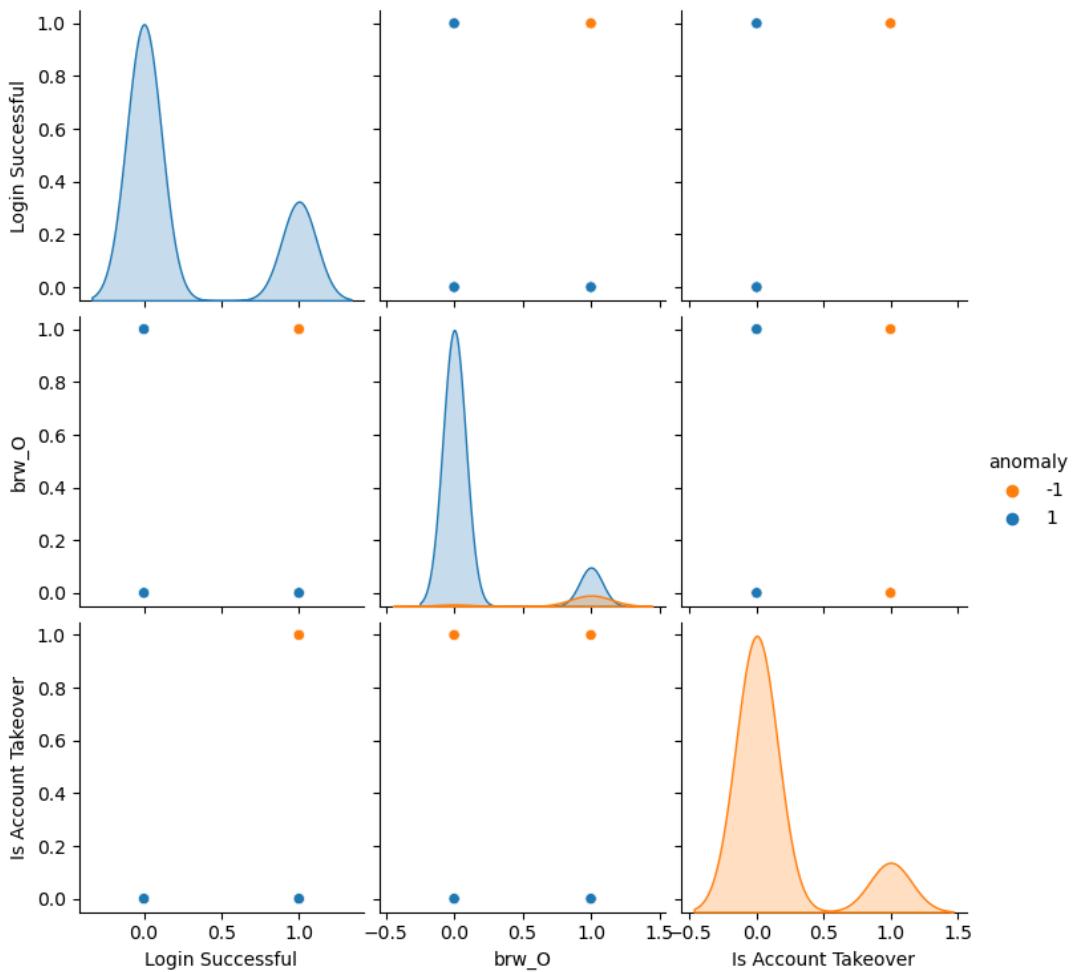
Outlier Method: Isolation Forest



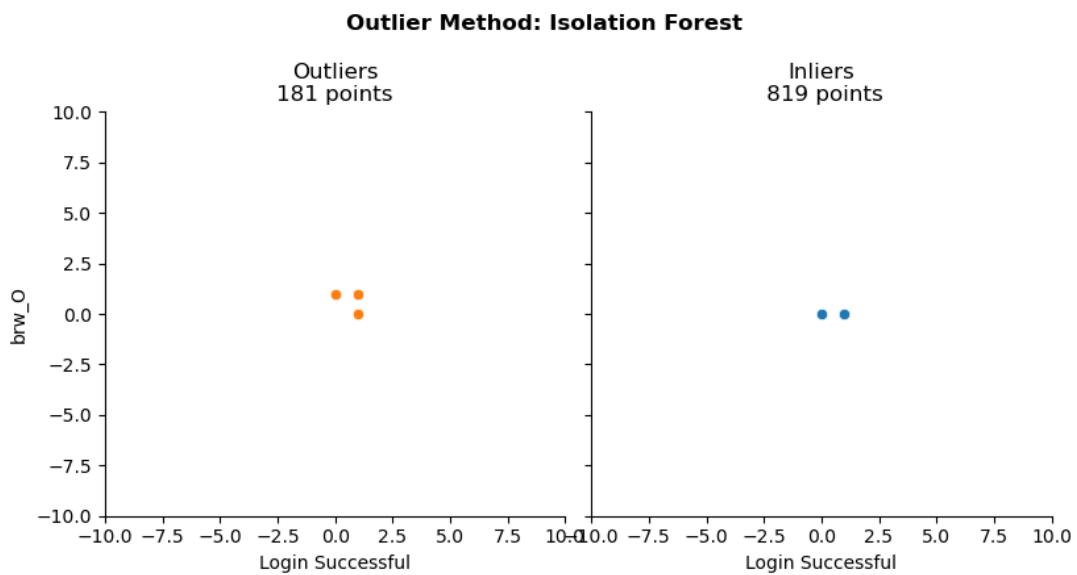


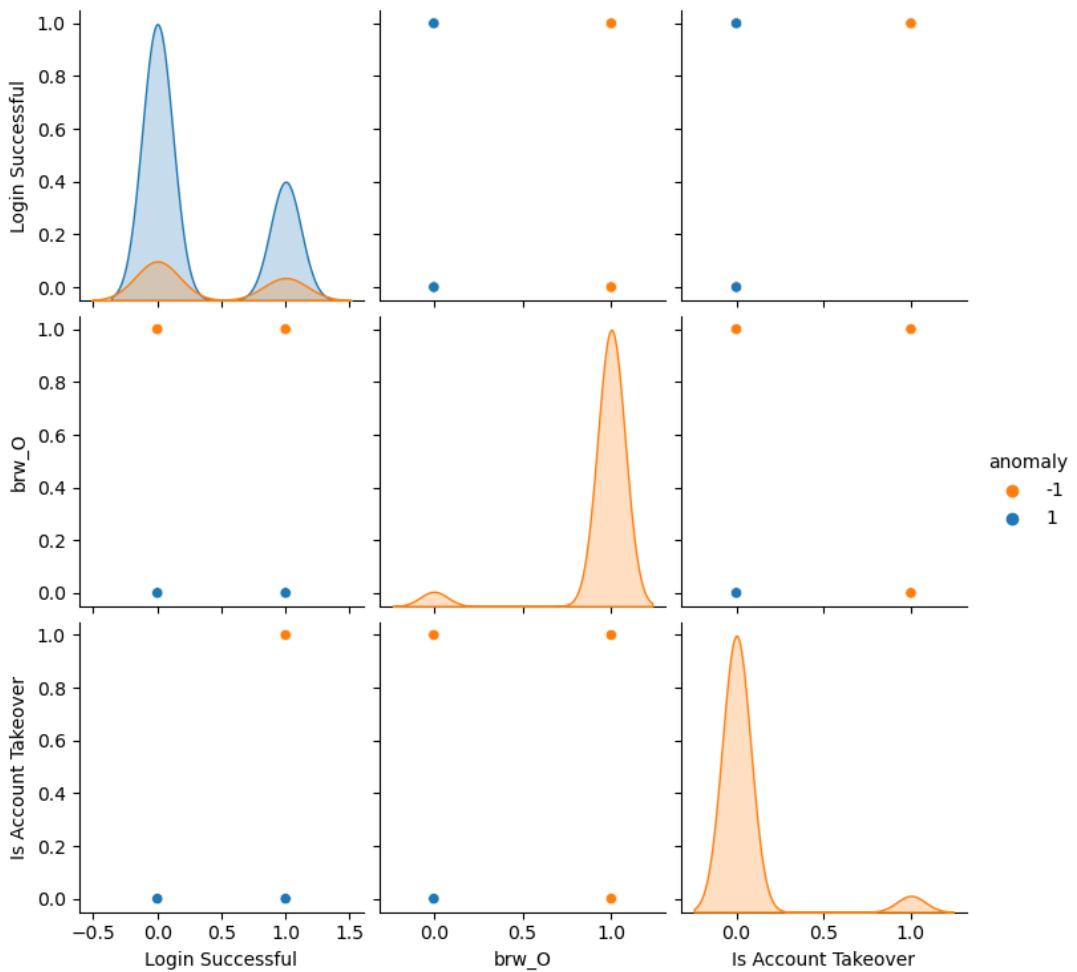
- anomaly_inputs = ['Login Successful', 'Browser Type (brw_O)', 'Is Account Takeover']
 - contamination value == 0.1





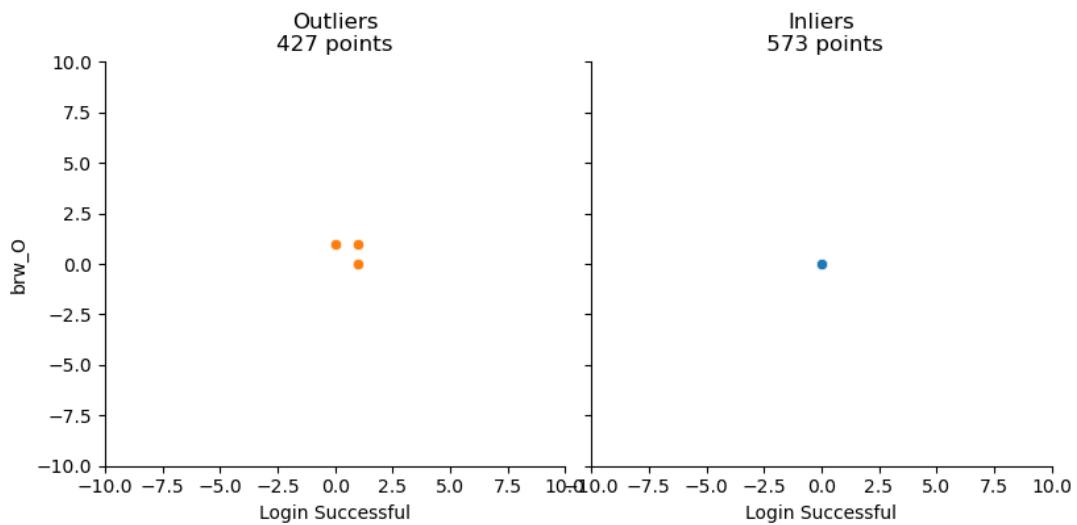
- contamination value == 0.3

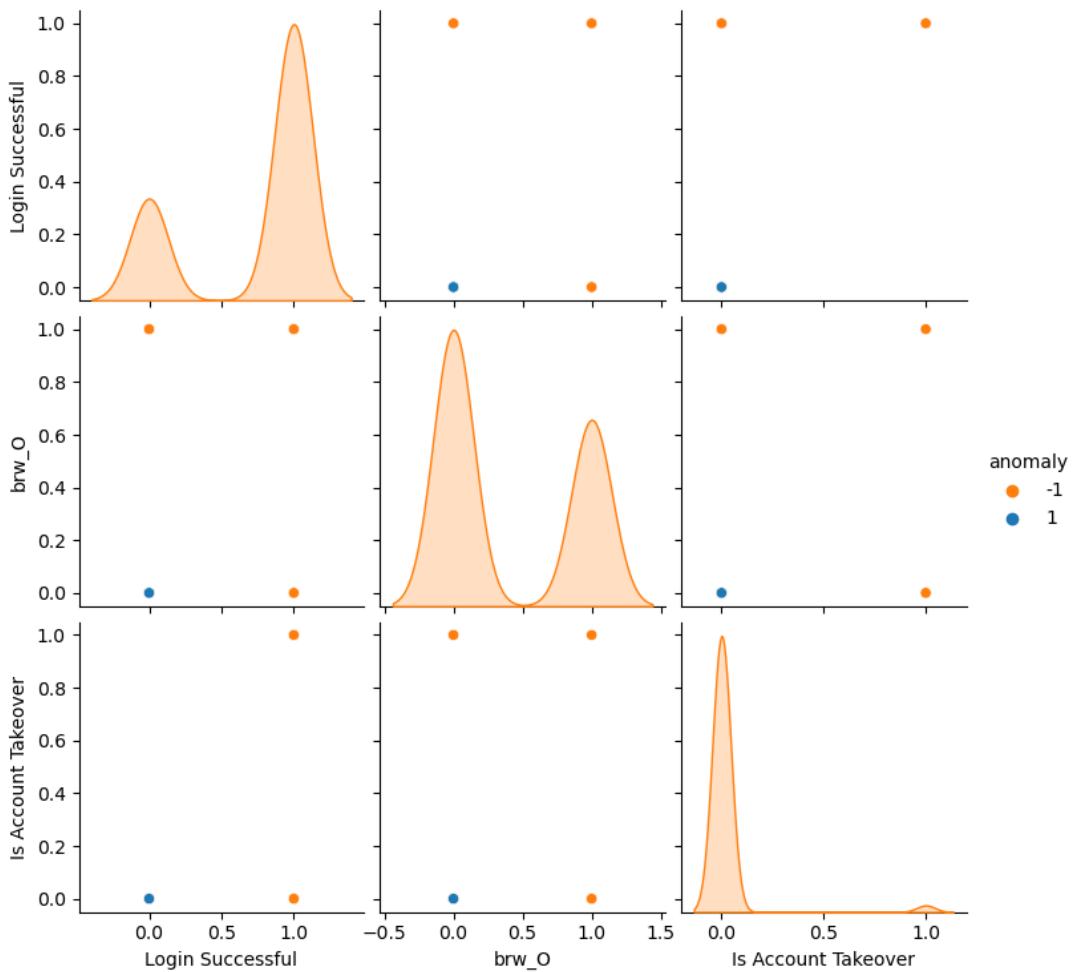




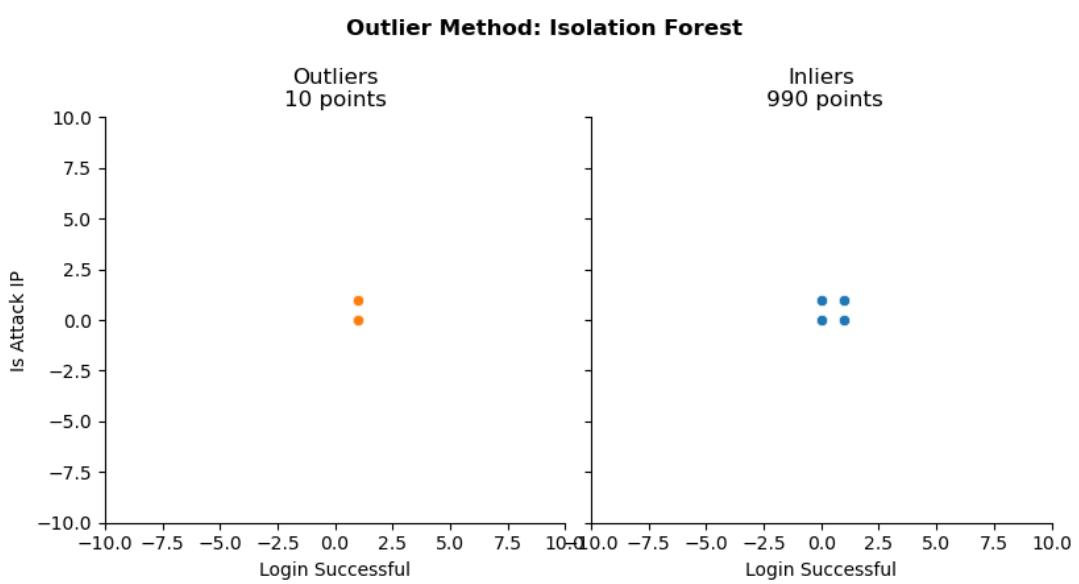
- contamination value == 0.5

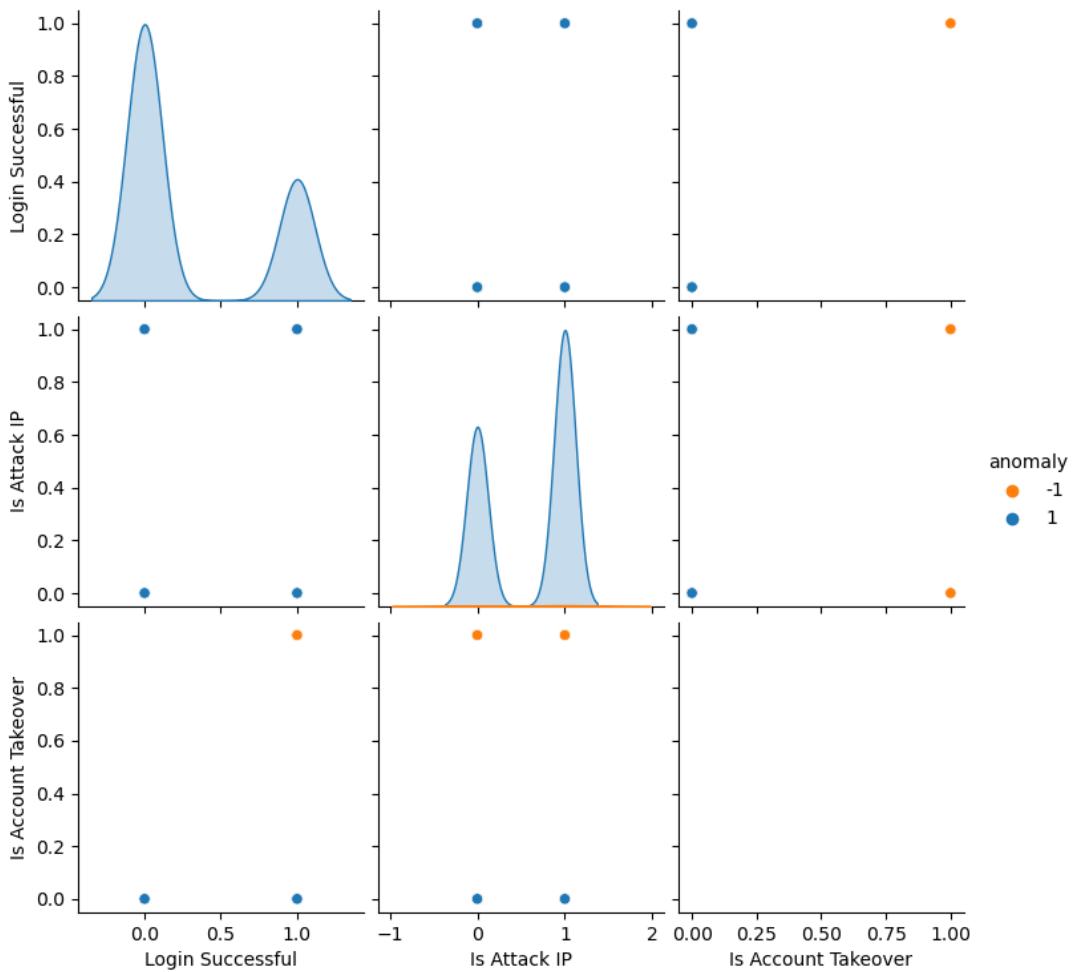
Outlier Method: Isolation Forest



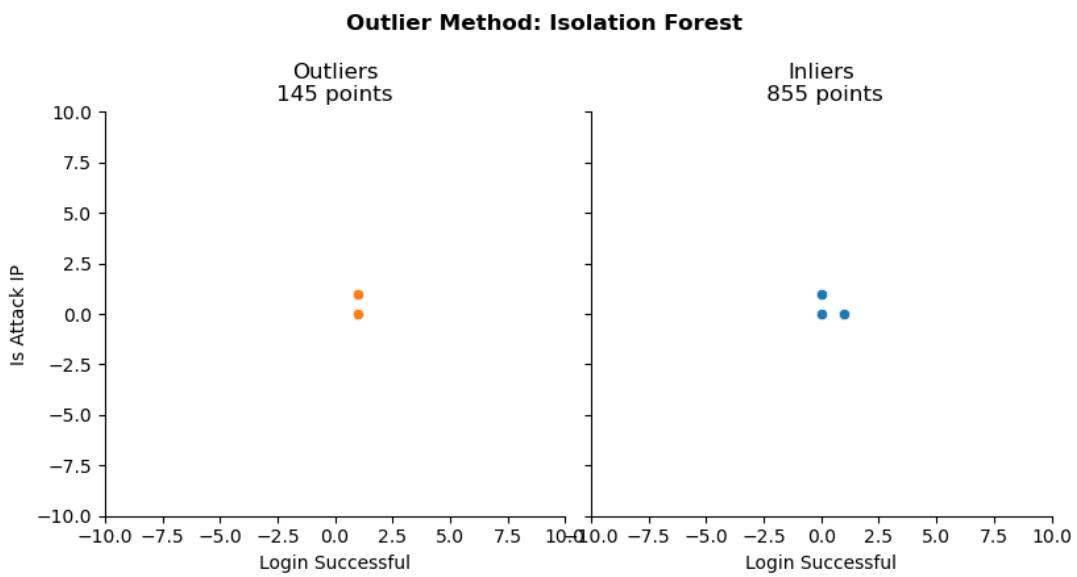


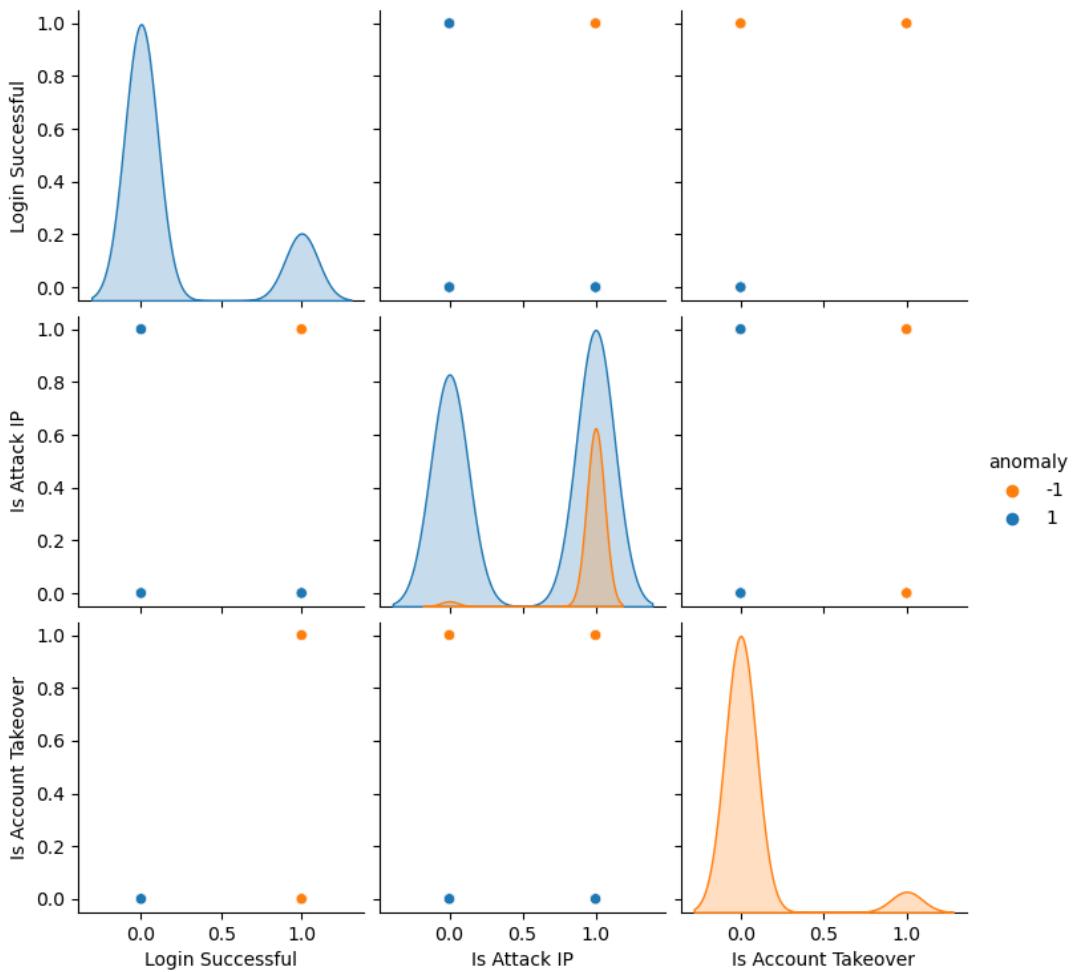
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1



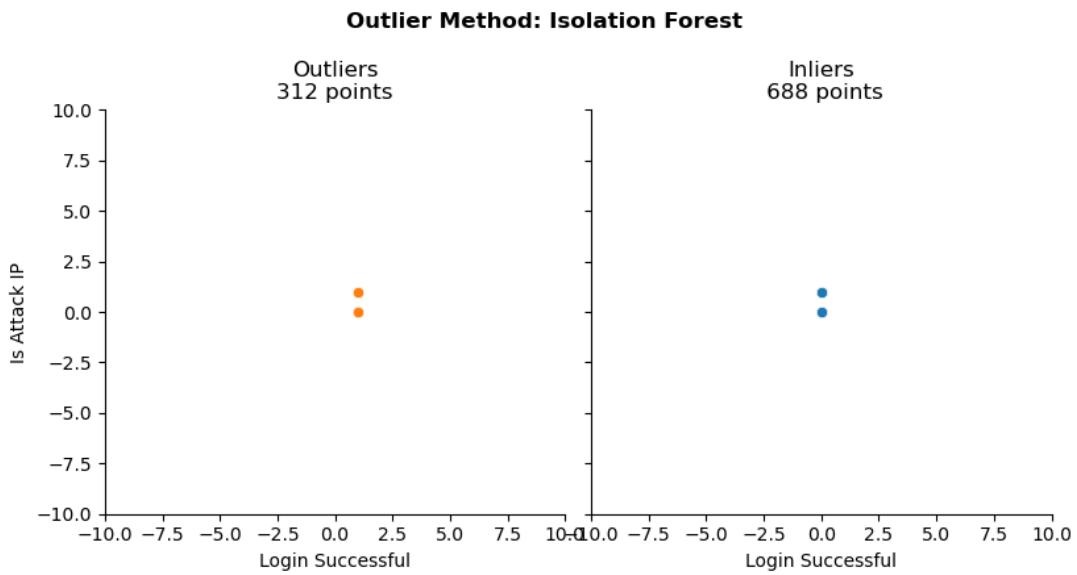


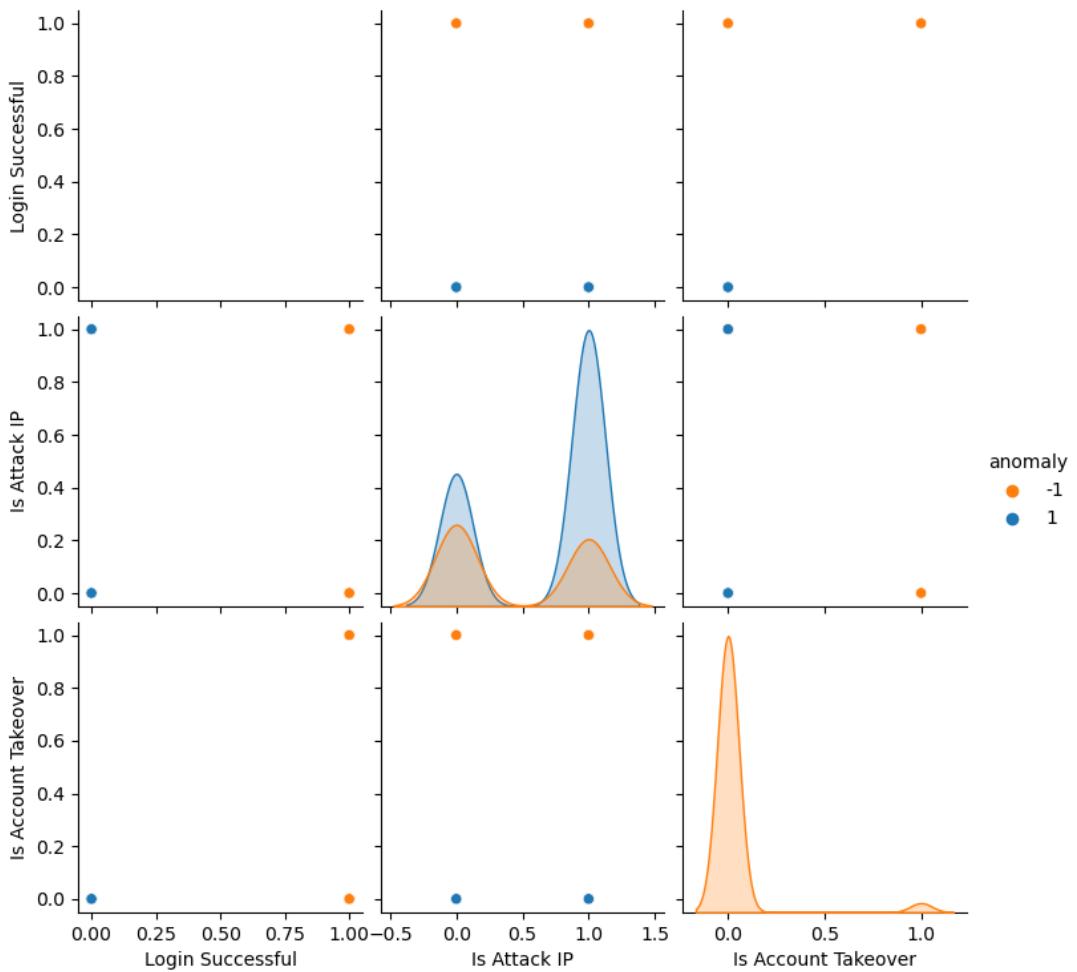
- contamination value == 0.3



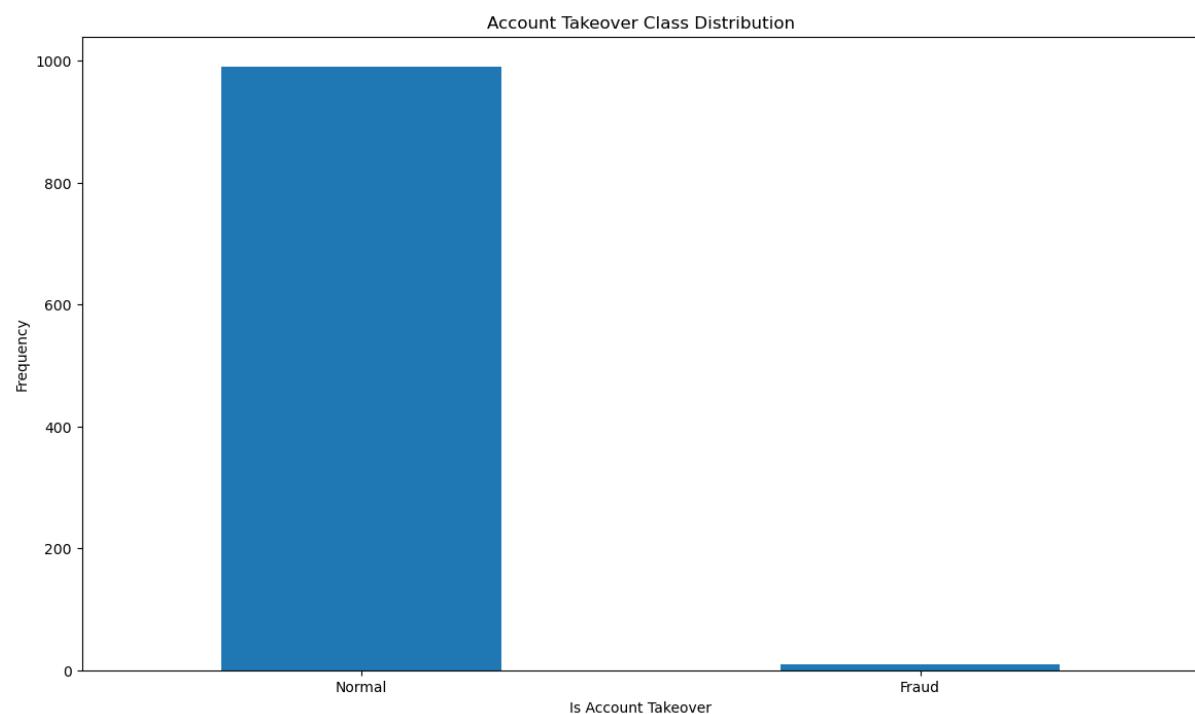


- contamination value == 0.5





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



Isolation Forest: 21

Accuracy Score :

0.979

Classification Report :

	precision	recall	f1-score	support
False	0.99	0.99	0.99	990
True	0.00	0.00	0.00	10
accuracy			0.98	1000
macro avg	0.49	0.49	0.49	1000
weighted avg	0.98	0.98	0.98	1000

Local Outlier Factor: 19

Accuracy Score :

0.981

Classification Report :

	precision	recall	f1-score	support
False	0.99	0.99	0.99	990
True	0.09	0.10	0.10	10
accuracy			0.98	1000
macro avg	0.54	0.54	0.54	1000
weighted avg	0.98	0.98	0.98	1000

Support Vector Machine: 741

Accuracy Score :

0.259

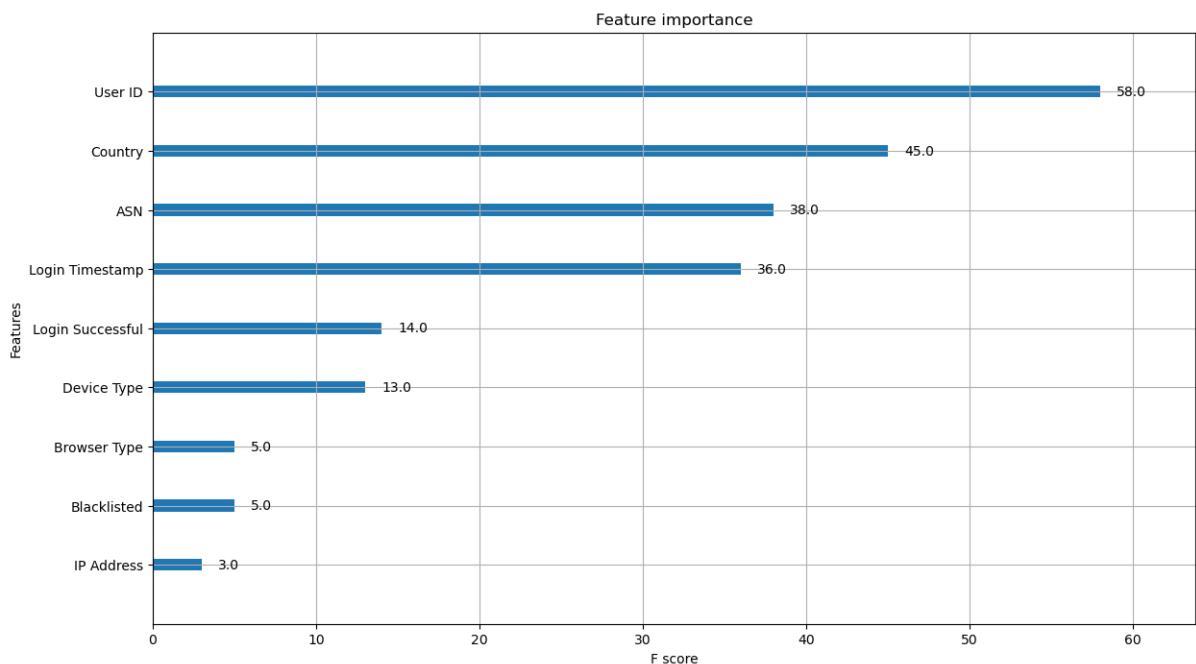
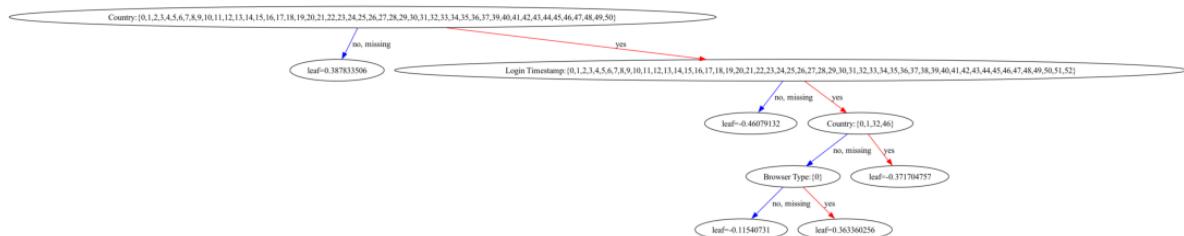
Classification Report :

	precision	recall	f1-score	support
False	0.99	0.25	0.40	990
True	0.01	0.70	0.02	10
accuracy			0.26	1000
macro avg	0.50	0.48	0.21	1000
weighted avg	0.98	0.26	0.40	1000

II. 30/1000 → 3%

1. XGB Experimental

```
-- 0 ASN 1000 non-null float64
-- 1 Country 1000 non-null category
-- 2 Device Type 1000 non-null category
-- 3 IP Address 1000 non-null category
-- 4 Is Attack IP 1000 non-null bool
-- 5 Login Successful 1000 non-null bool
-- 6 Login Timestamp 1000 non-null category
-- 7 User ID 1000 non-null float64
-- 8 Blacklisted 1000 non-null bool
-- 9 Browser Type 1000 non-null category
-- 10 Is Account Takeover 1000 non-null bool
dtypes: bool(4), category(5), float64(2)
memory usage: 90.7 KB
Feature importances:
[0.02946164 0.39781916 0.02500814 0.06304331 0.          0.19330929
 0.11321067 0.05228835 0.01595914 0.10990029]
```

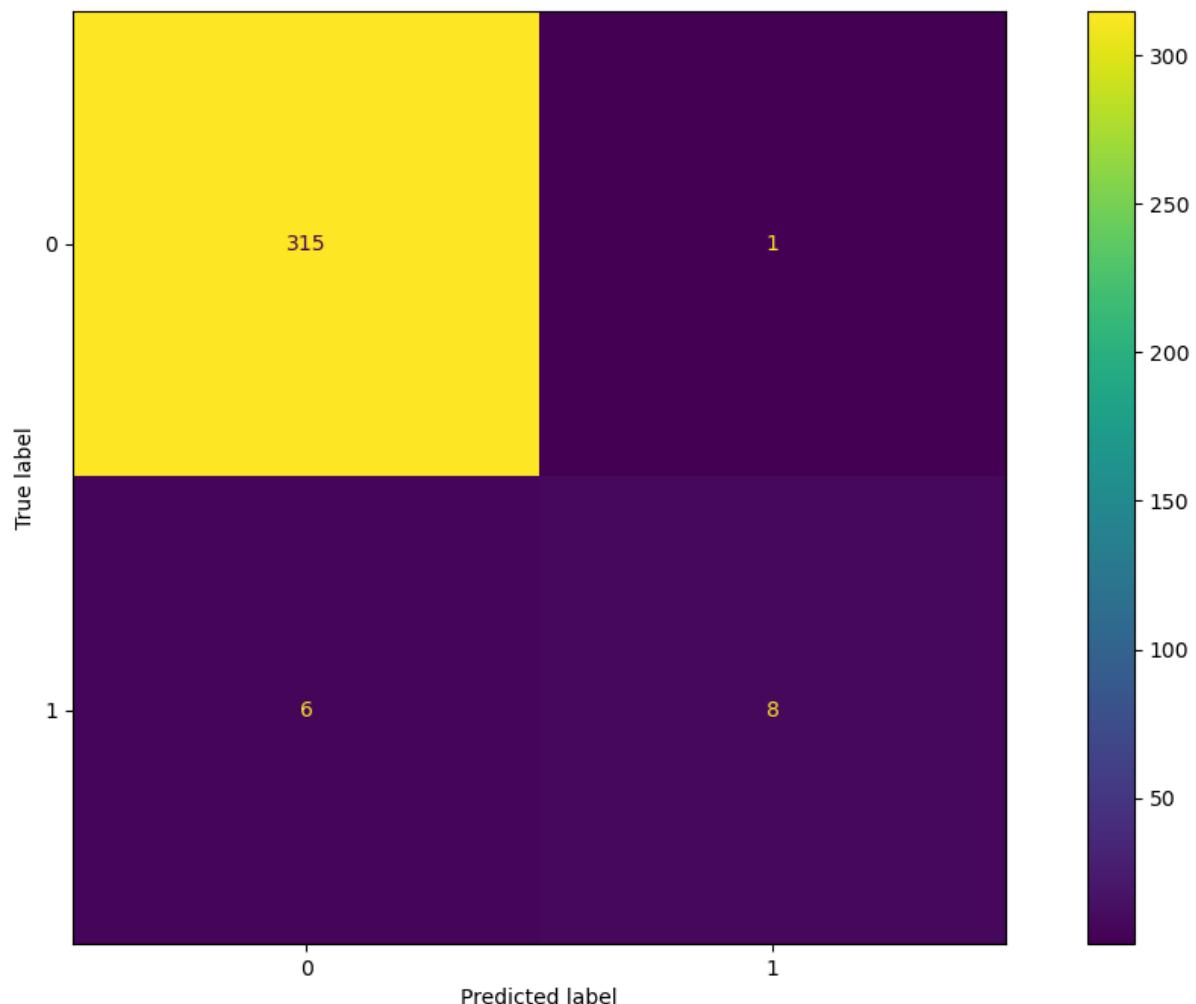


2. Label encoding (cat.codes)

2A. XGBoost

memory usage: 26.5 KB

Accuracy: 97.88%				
	precision	recall	f1-score	support
False	0.98	1.00	0.99	316
True	0.89	0.57	0.70	14
accuracy			0.98	330
macro avg	0.94	0.78	0.84	330
weighted avg	0.98	0.98	0.98	330



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 800
The number of records in the test dataset is 200
The training dataset has 776 records for the majority class and 24 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      194
      True         0.03     1.00     0.06       6
      accuracy          0.03     200
      macro avg       0.01     0.50     0.03     200
      weighted avg    0.00     0.03     0.00     200
      The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.97     1.00     0.98      194
      True         0.00     0.00     0.00       6
      accuracy          0.97     200
      macro avg       0.48     0.50     0.49     200
      weighted avg    0.94     0.97     0.96     200
[[ 0 194]
 [ 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 700
The number of records in the test dataset is 300
The training dataset has 679 records for the majority class and 21 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      291
      True         0.03     1.00     0.06       9
      accuracy          0.03     300
      macro avg       0.01     0.50     0.03     300
      weighted avg    0.00     0.03     0.00     300
      The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.97     1.00     0.98      291
      True         0.00     0.00     0.00       9
      accuracy          0.97     300
      macro avg       0.48     0.50     0.49     300
      weighted avg    0.94     0.97     0.96     300
[[ 0 291]
 [ 0  9]]

```

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

```

The number of records in the training dataset is 500
The number of records in the test dataset is 500
The training dataset has 484 records for the majority class and 16 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00      486
  True        0.03     1.00     0.05      14

accuracy                           0.03      500
macro avg       0.01     0.50     0.03      500
weighted avg    0.00     0.03     0.00      500

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.97     1.00     0.99      486
  True        0.00     0.00     0.00      14

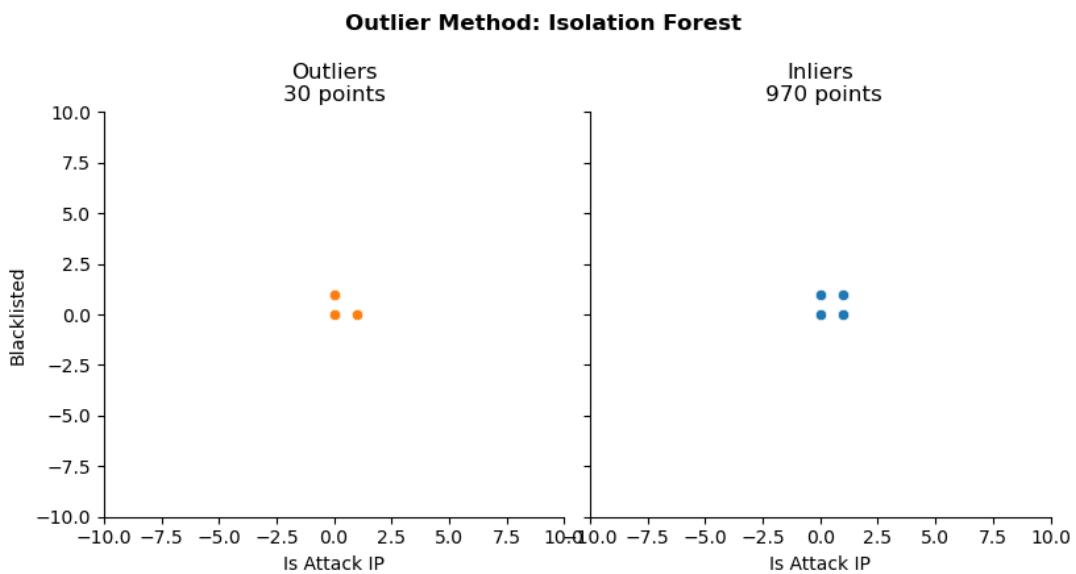
accuracy                           0.97      500
macro avg       0.49     0.50     0.49      500
weighted avg    0.94     0.97     0.96      500

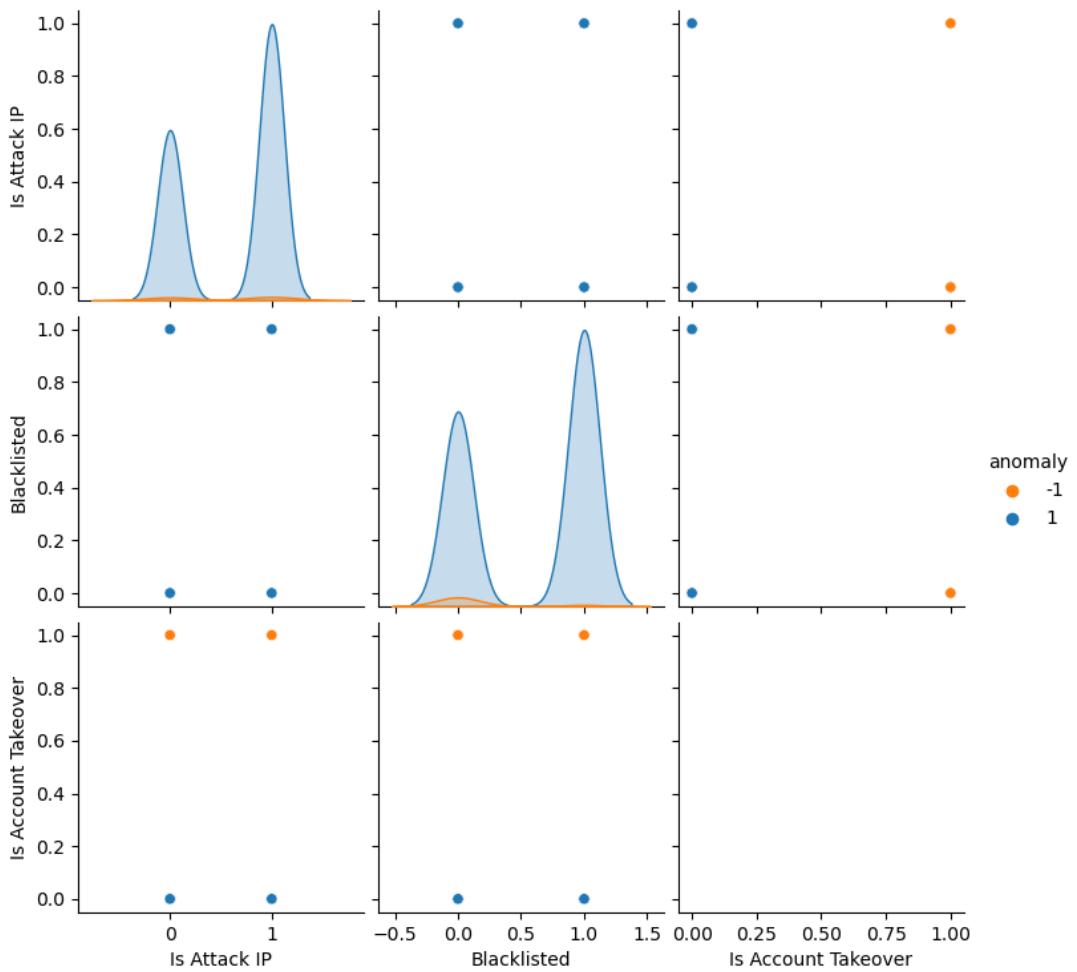
[[ 0 486]
 [ 0 14]]

```

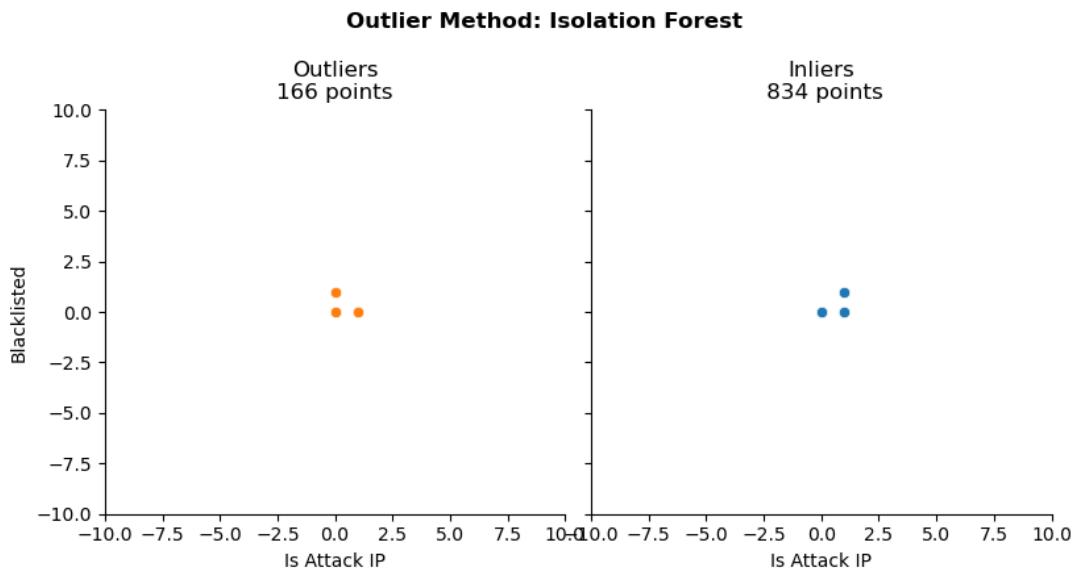
2C. Isolation Forest

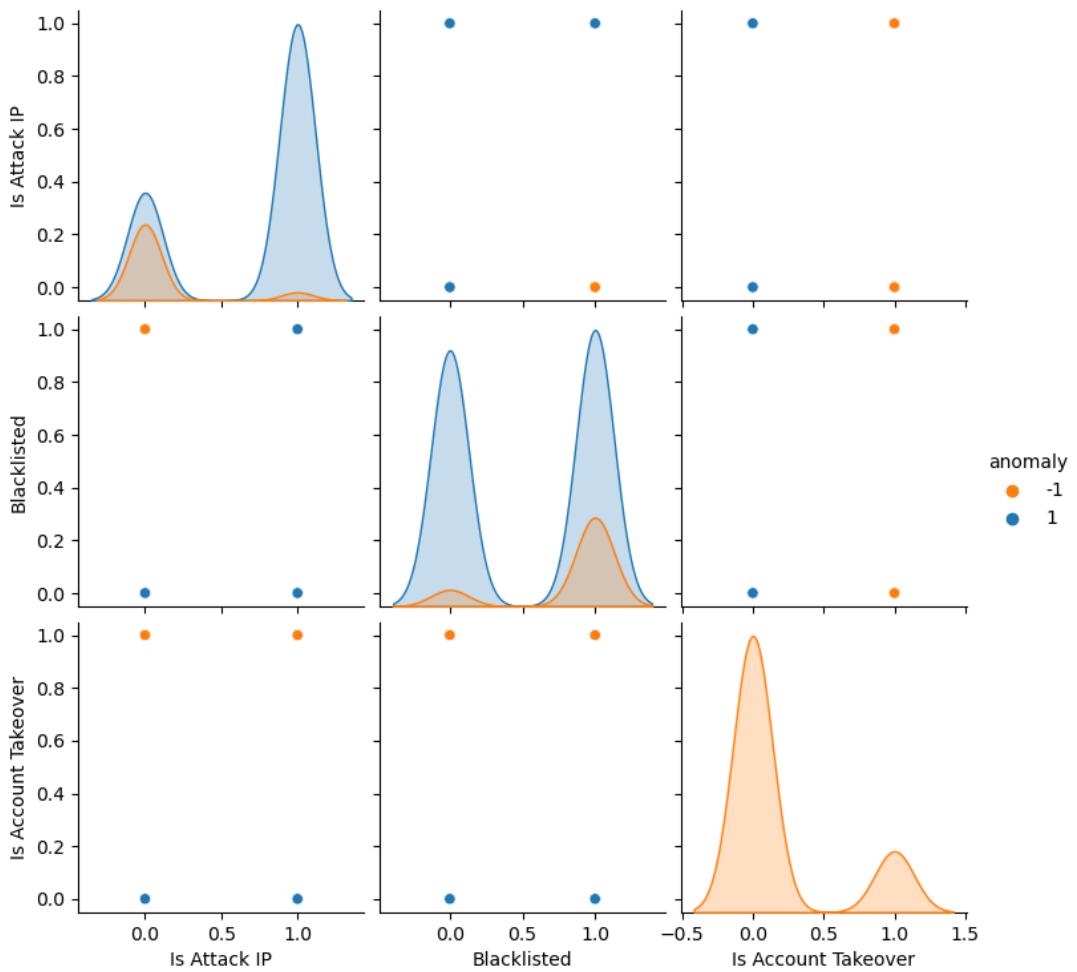
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



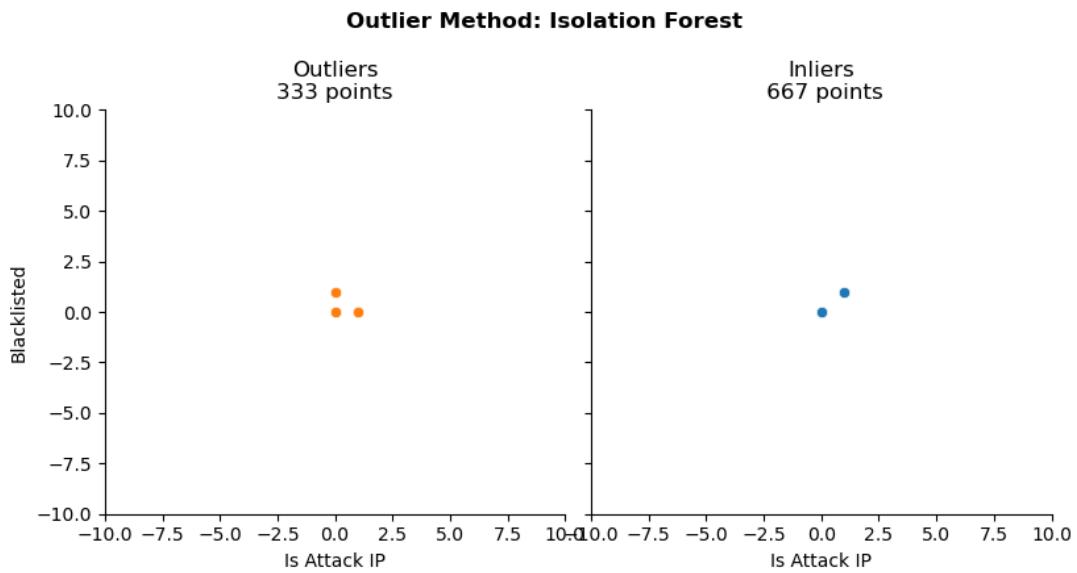


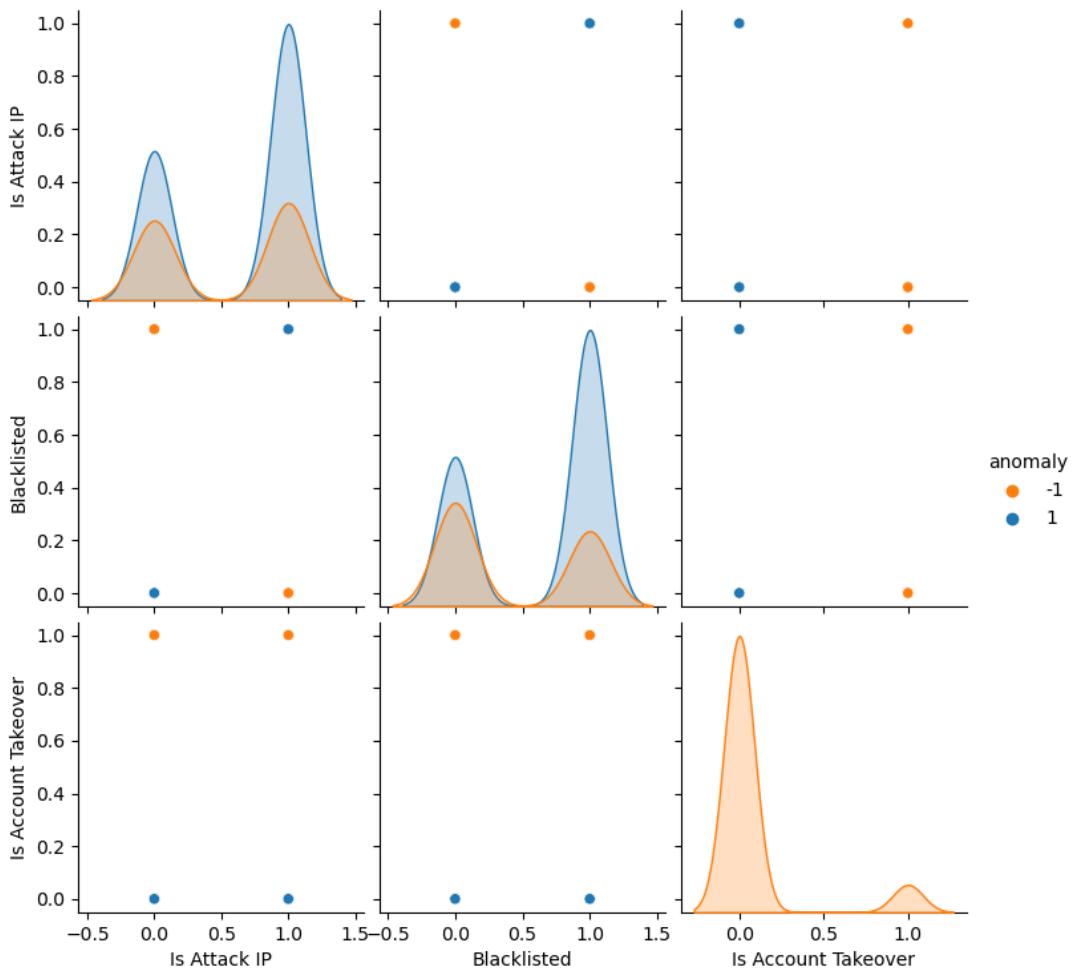
- increasing contamination value to 0.3



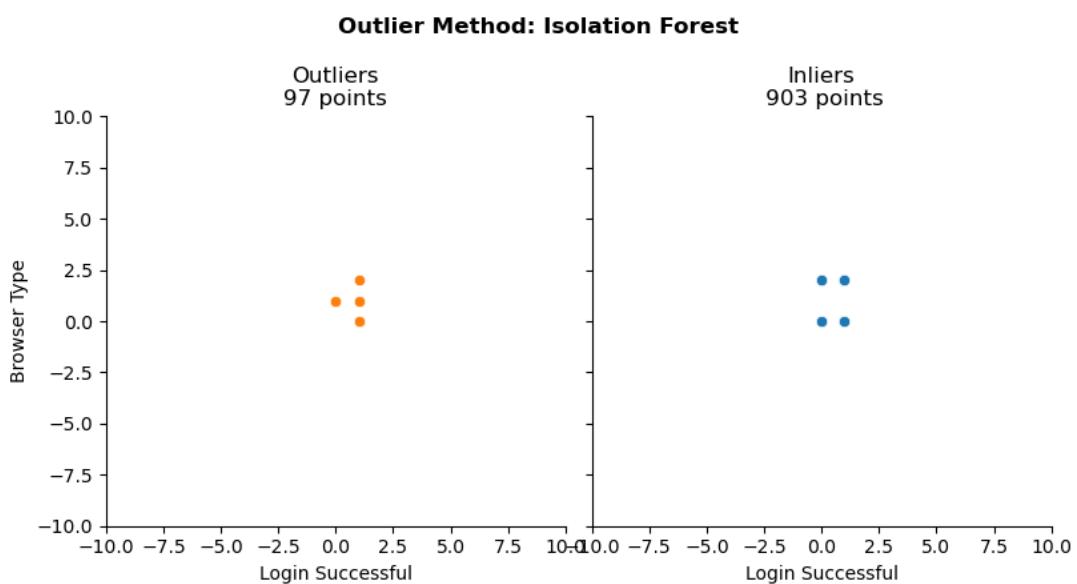


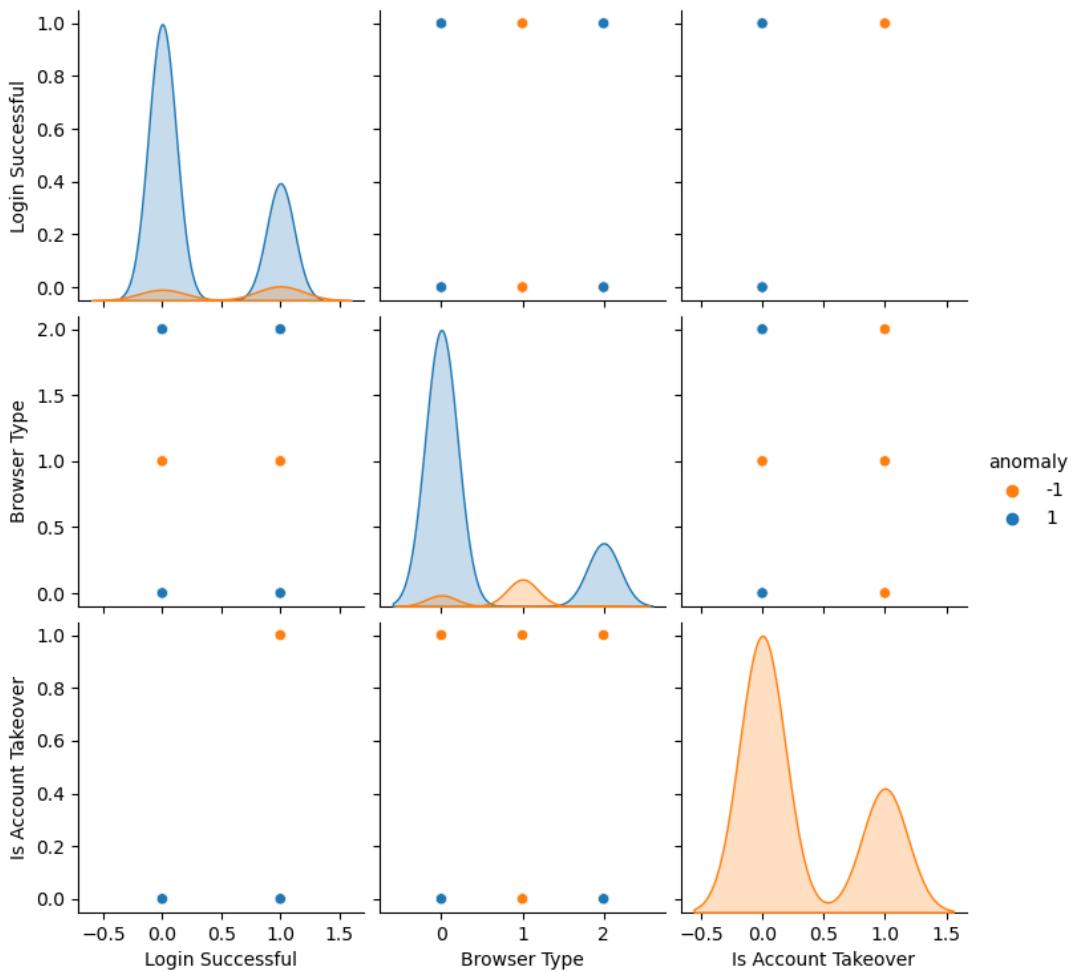
- contamination value == 0.5



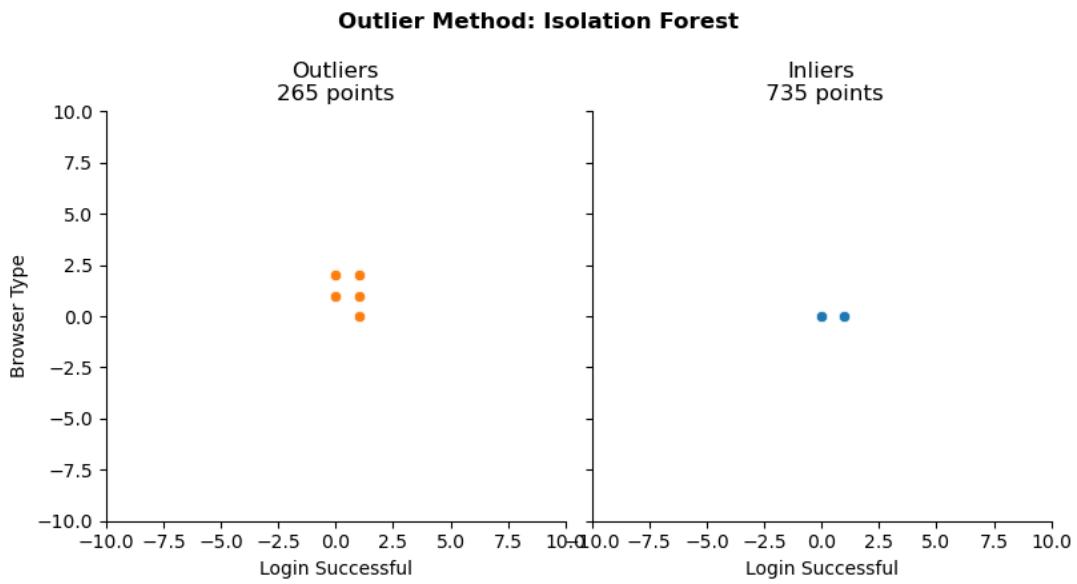


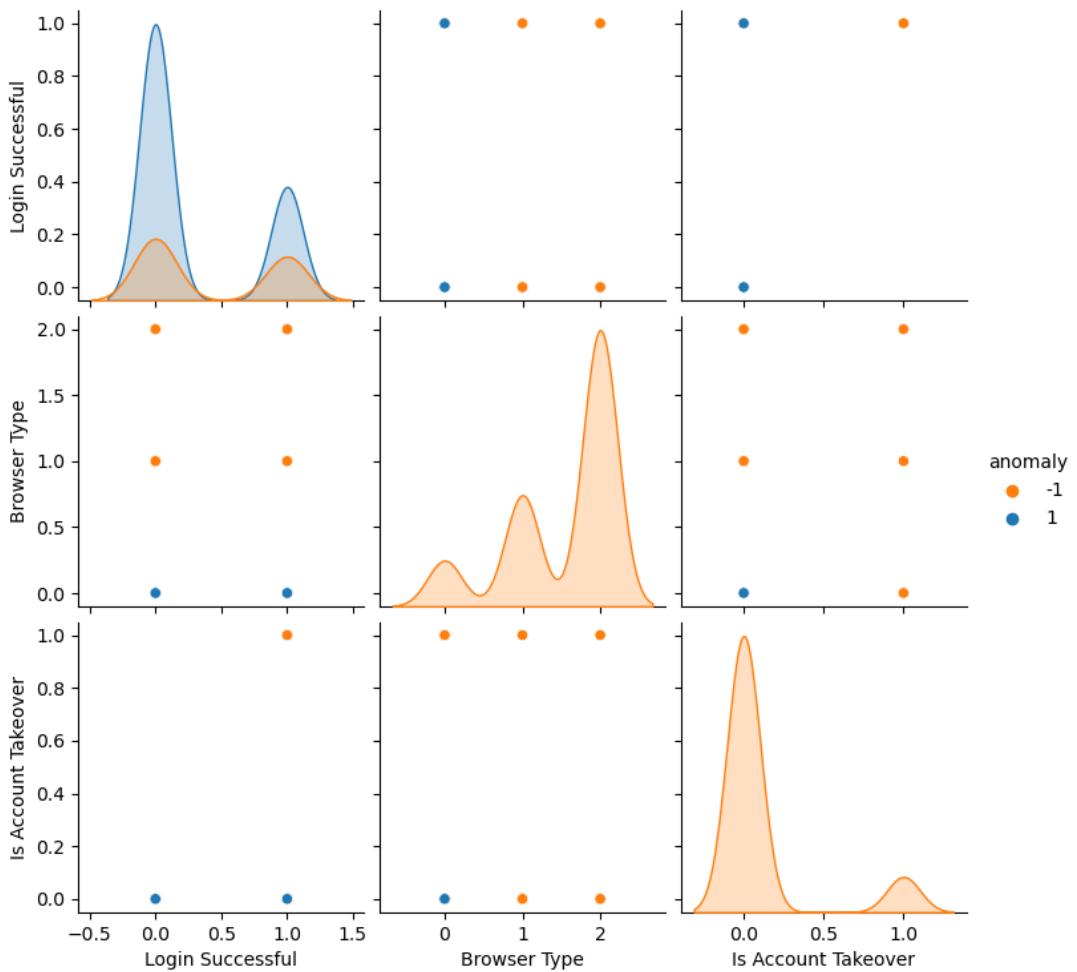
- `anomaly_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']`
 - contamination value == 0.1



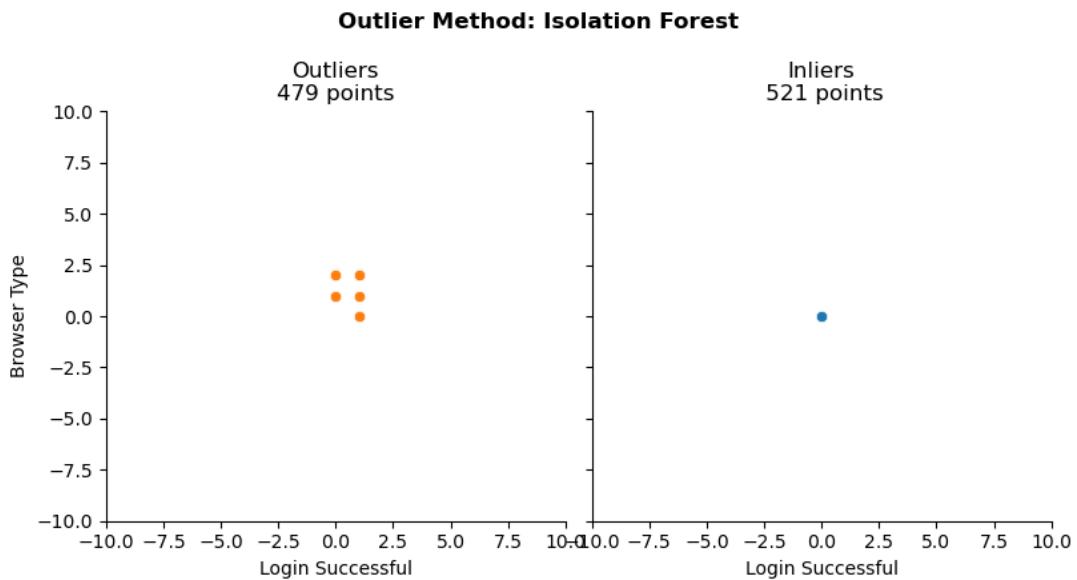


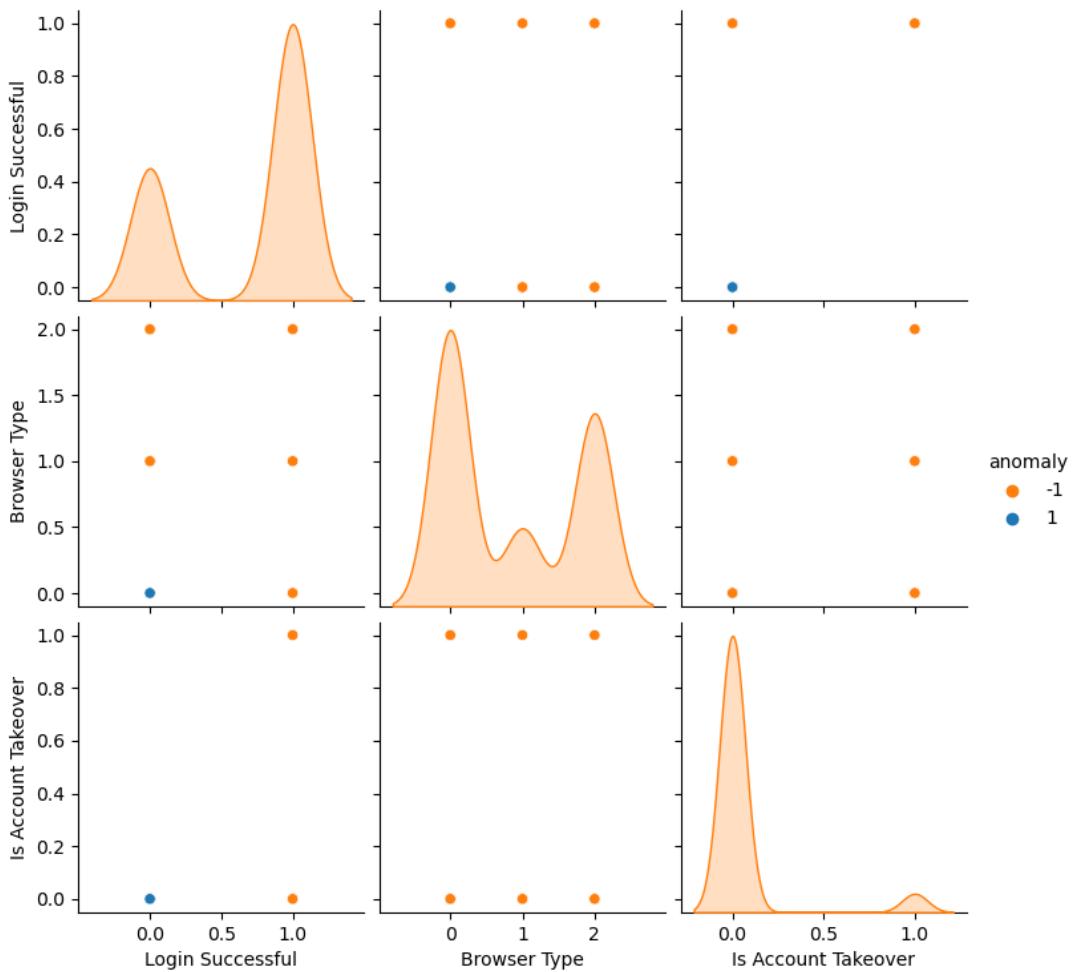
- contamination value == 0.3



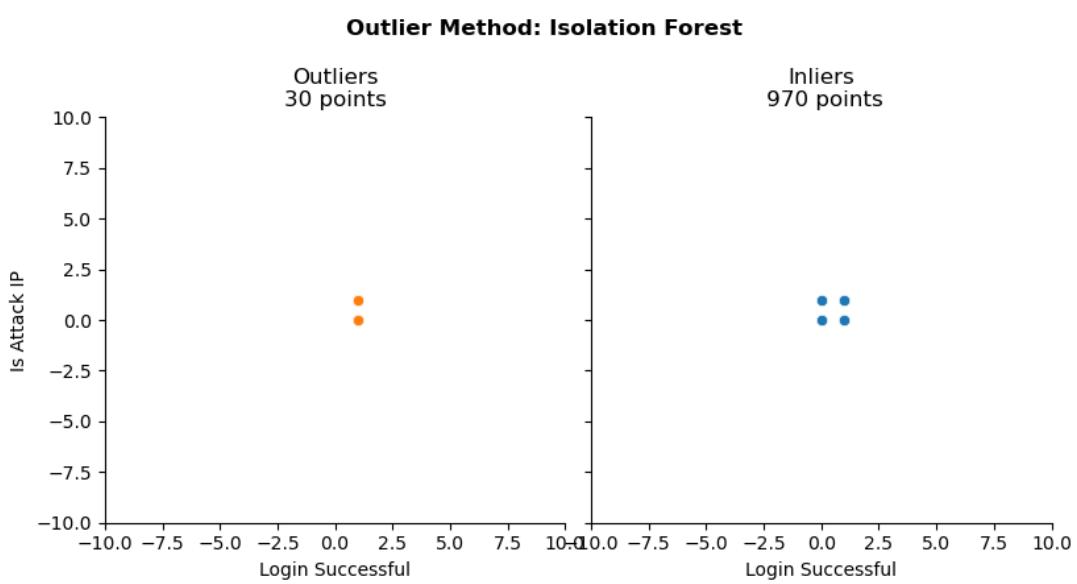


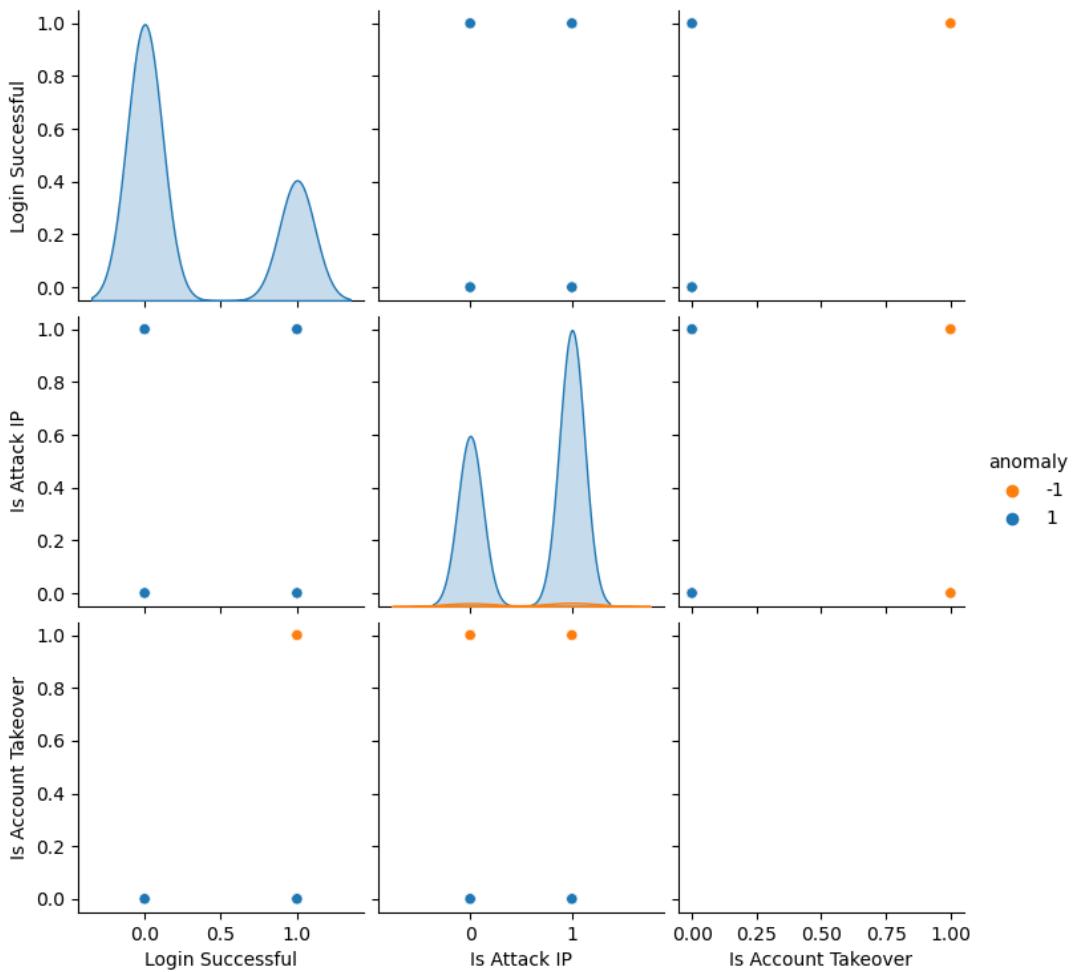
- contamination value == 0.5



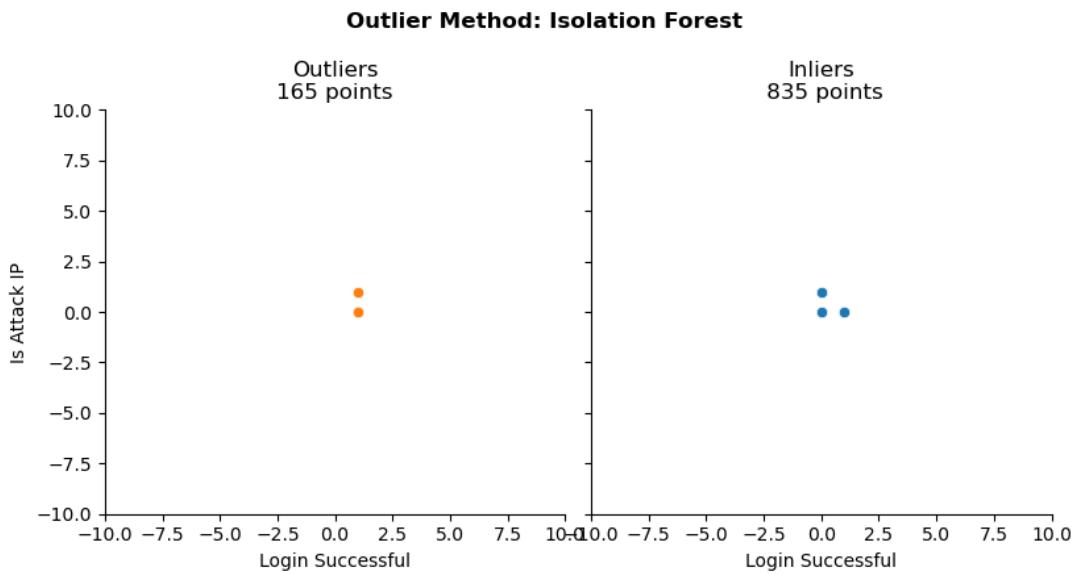


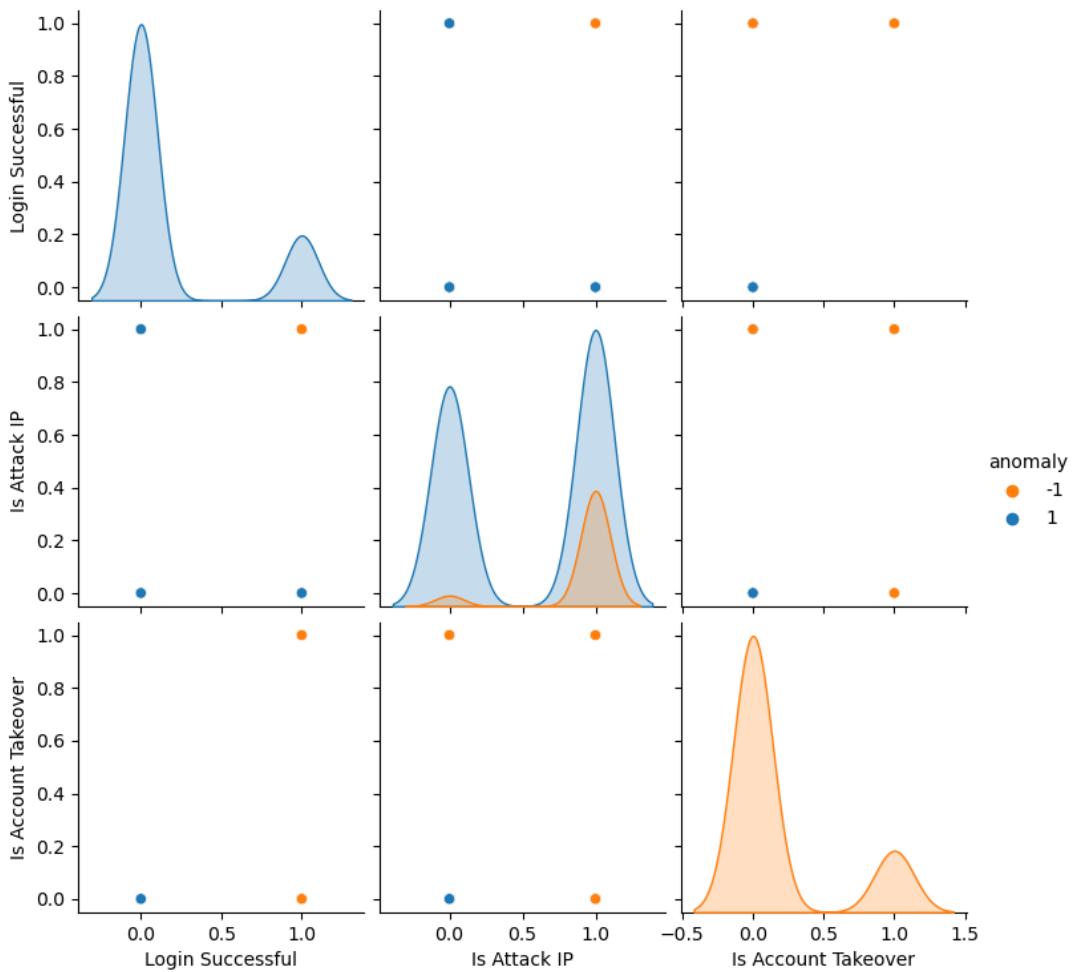
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1



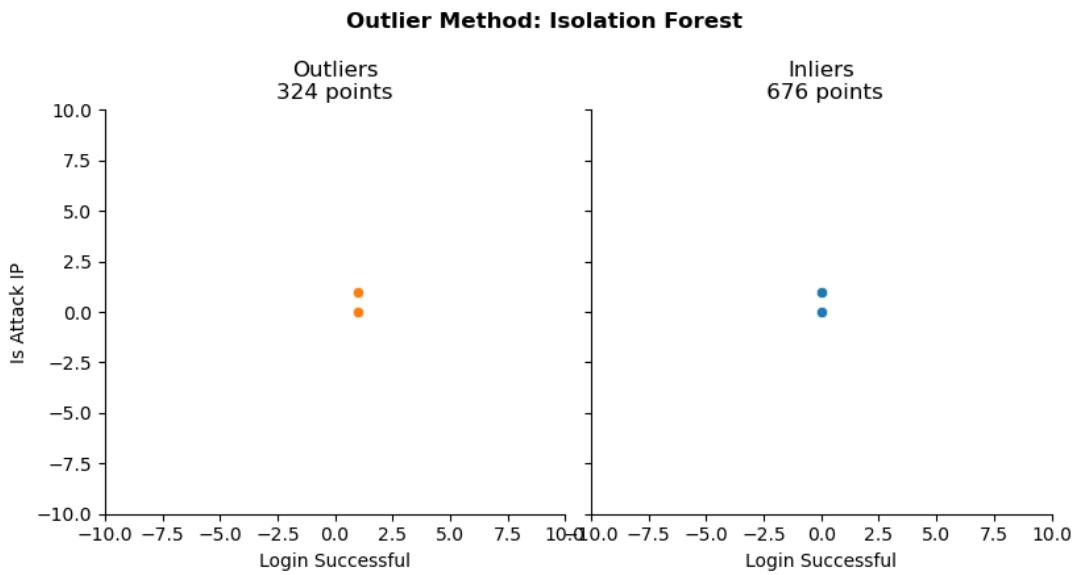


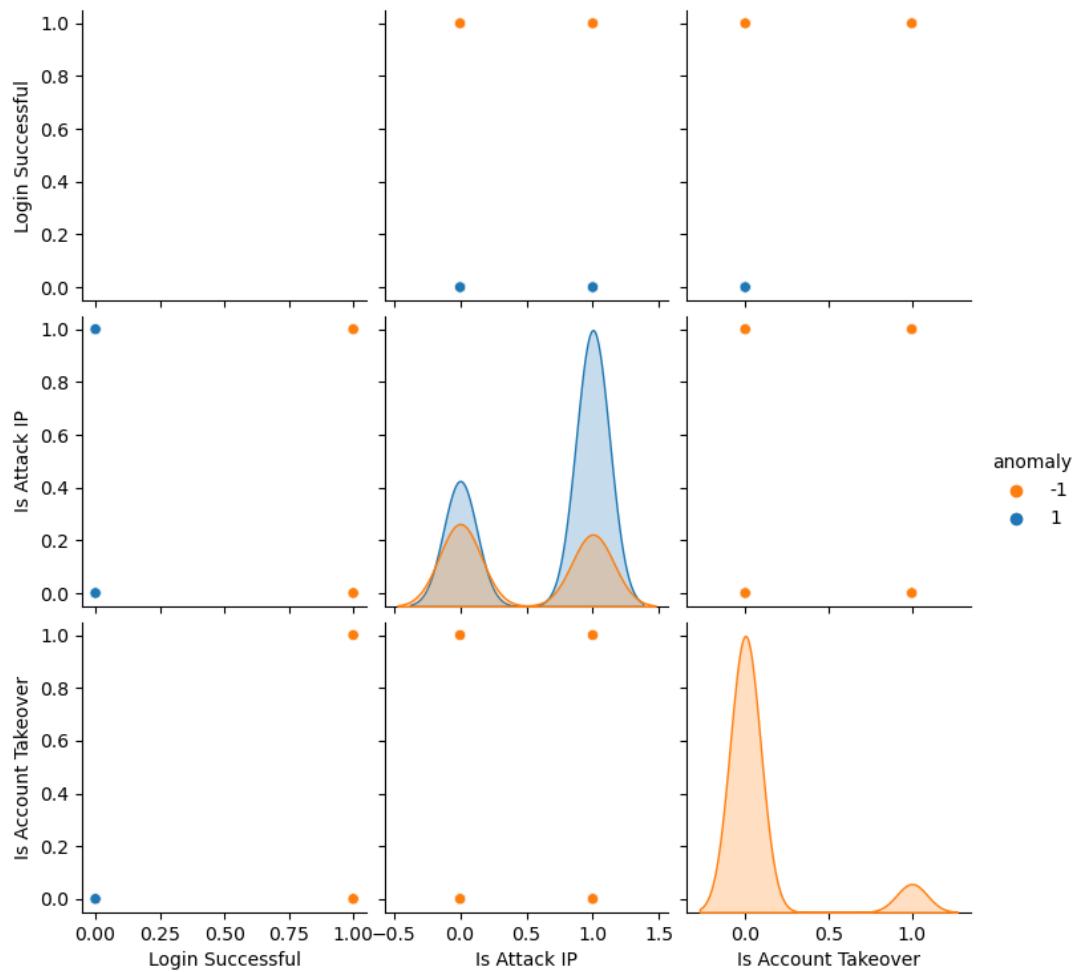
- contamination value == 0.3



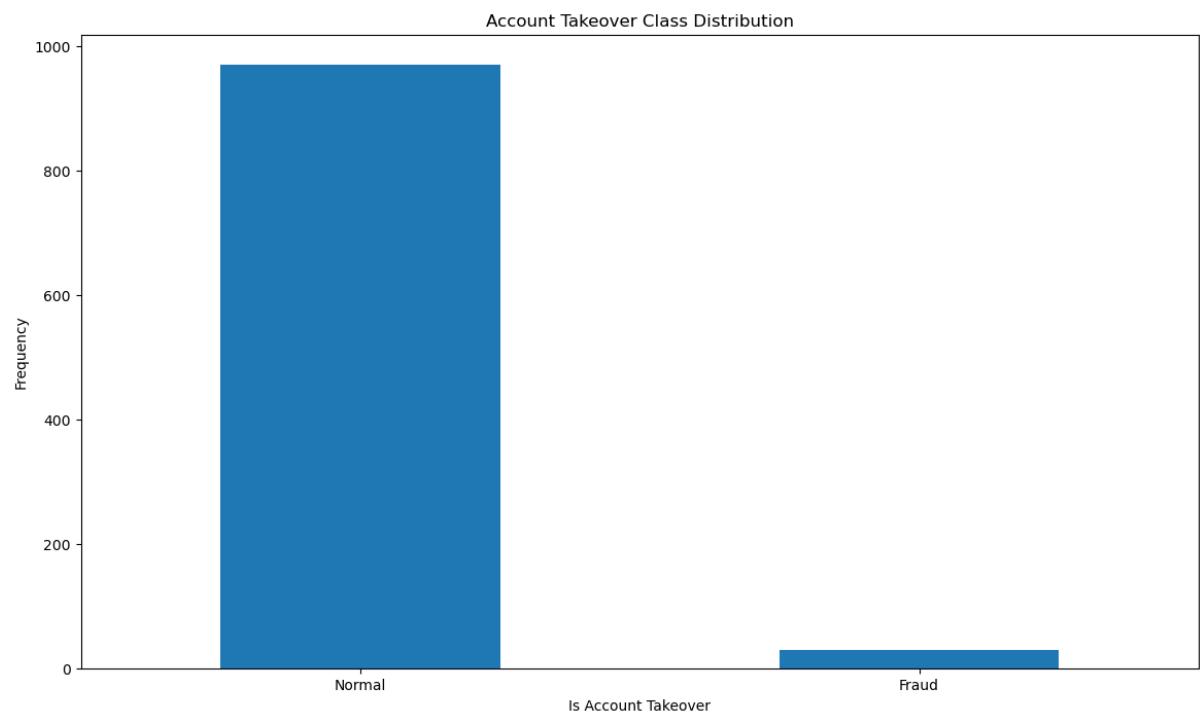


- contamination value == 0.5





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



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

Isolation Forest: 59

Accuracy Score :

0.941

Classification Report :

	precision	recall	f1-score	support
False	0.97	0.97	0.97	970
True	0.03	0.03	0.03	30
accuracy			0.94	1000
macro avg	0.50	0.50	0.50	1000
weighted avg	0.94	0.94	0.94	1000

Local Outlier Factor: 59

Accuracy Score :

0.941

Classification Report :

	precision	recall	f1-score	support
False	0.97	0.97	0.97	970
True	0.03	0.03	0.03	30
accuracy			0.94	1000
macro avg	0.50	0.50	0.50	1000
weighted avg	0.94	0.94	0.94	1000

Support Vector Machine: 824

Accuracy Score :

0.176

Classification Report :

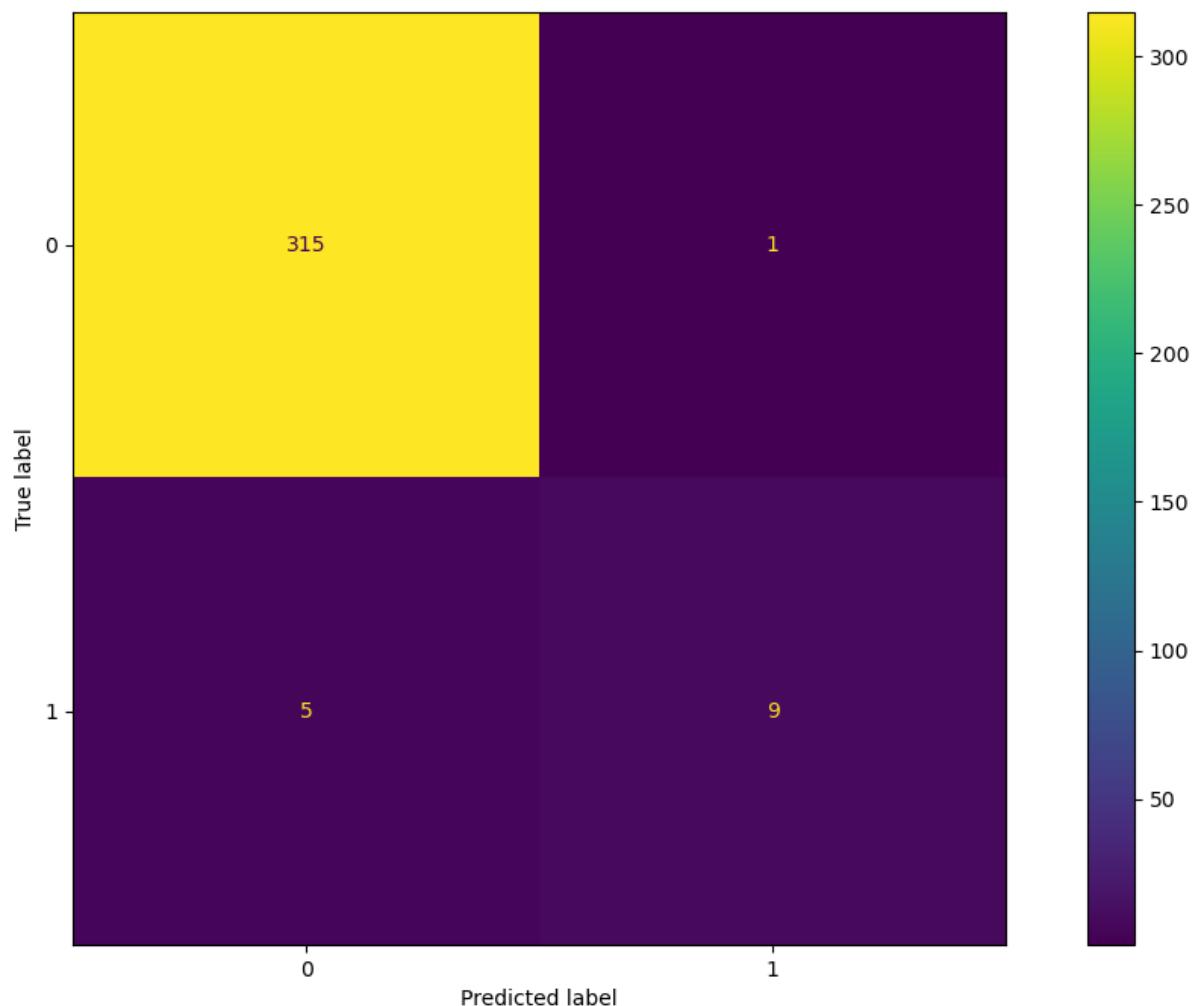
	precision	recall	f1-score	support
False	0.93	0.16	0.28	970
True	0.02	0.60	0.04	30
accuracy			0.18	1000
macro avg	0.48	0.38	0.16	1000
weighted avg	0.90	0.18	0.27	1000

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: 1.7 MB

Accuracy: 98.18%				
	precision	recall	f1-score	support
False	0.98	1.00	0.99	316
True	0.90	0.64	0.75	14
accuracy			0.98	330
macro avg	0.94	0.82	0.87	330
weighted avg	0.98	0.98	0.98	330



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 800
The number of records in the test dataset is 200
The training dataset has 776 records for the majority class and 24 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.31     0.48     194
      True         0.04     1.00     0.08       6

      accuracy          0.34     200
      macro avg       0.52     0.66     0.28     200
      weighted avg    0.97     0.34     0.47     200

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.97     1.00     0.98     194
      True         0.00     0.00     0.00       6

      accuracy          0.97     200
      macro avg       0.48     0.50     0.49     200
      weighted avg    0.94     0.97     0.96     200

[[ 61 133]
 [ 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 700
The number of records in the test dataset is 300
The training dataset has 679 records for the majority class and 21 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.01     0.03     291
      True         0.03     1.00     0.06       9

      accuracy          0.04     300
      macro avg       0.52     0.51     0.04     300
      weighted avg    0.97     0.04     0.03     300

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.97     1.00     0.98     291
      True         0.00     0.00     0.00       9

      accuracy          0.97     300
      macro avg       0.48     0.50     0.49     300
      weighted avg    0.94     0.97     0.96     300

[[ 4 287]
 [ 0  9]]

```

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

```

The number of records in the training dataset is 500
The number of records in the test dataset is 500
The training dataset has 484 records for the majority class and 16 records for the minority class.
      precision    recall   f1-score   support
False        1.00     0.04     0.07     486
True        0.03     1.00     0.06     14

accuracy          0.06     500
macro avg       0.51     0.52     0.06     500
weighted avg    0.97     0.06     0.07     500

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.97     1.00     0.99     486
True        0.00     0.00     0.00     14

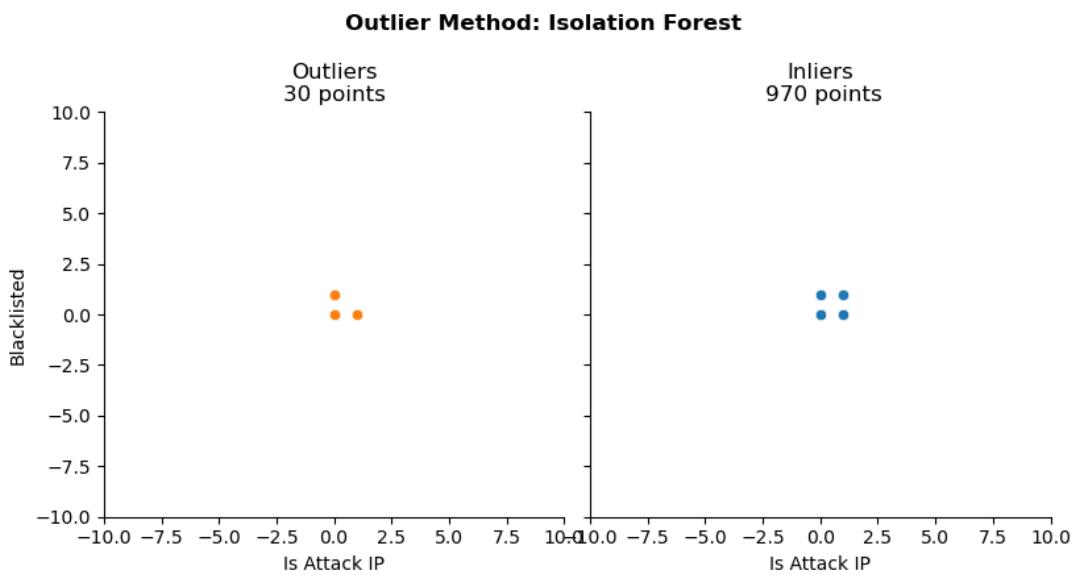
accuracy          0.97     500
macro avg       0.49     0.50     0.49     500
weighted avg    0.94     0.97     0.96     500

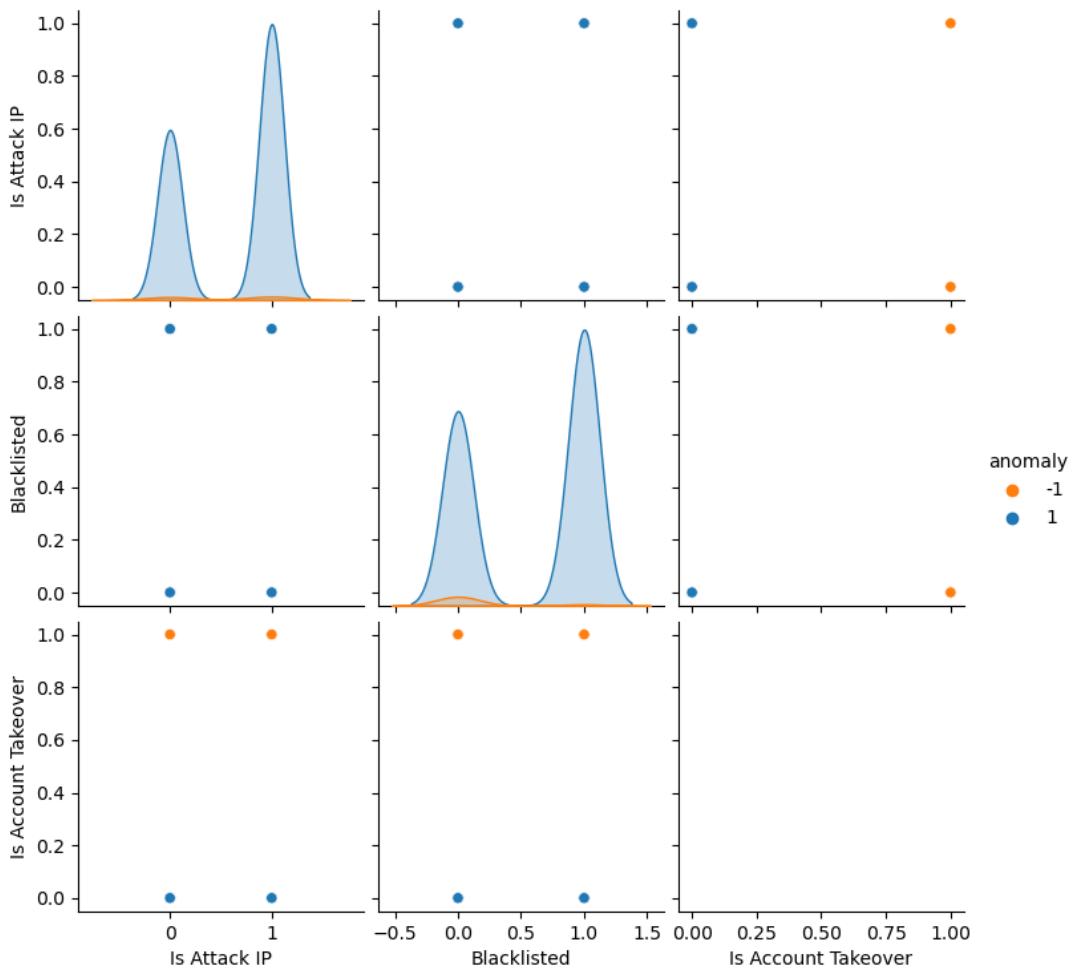
[[ 18 468]
 [ 0 14]]

```

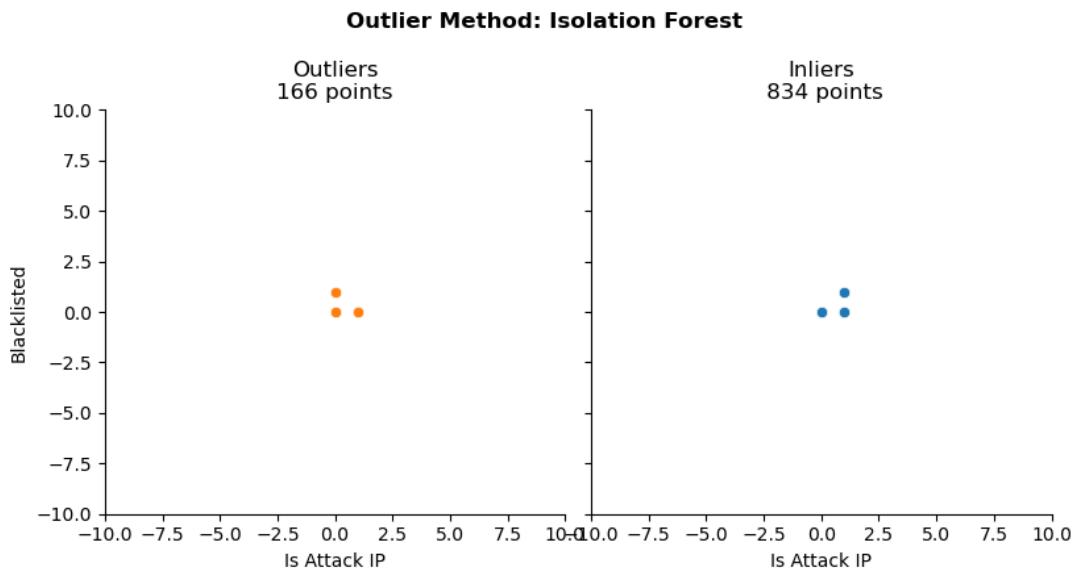
3C. Isolation Forest

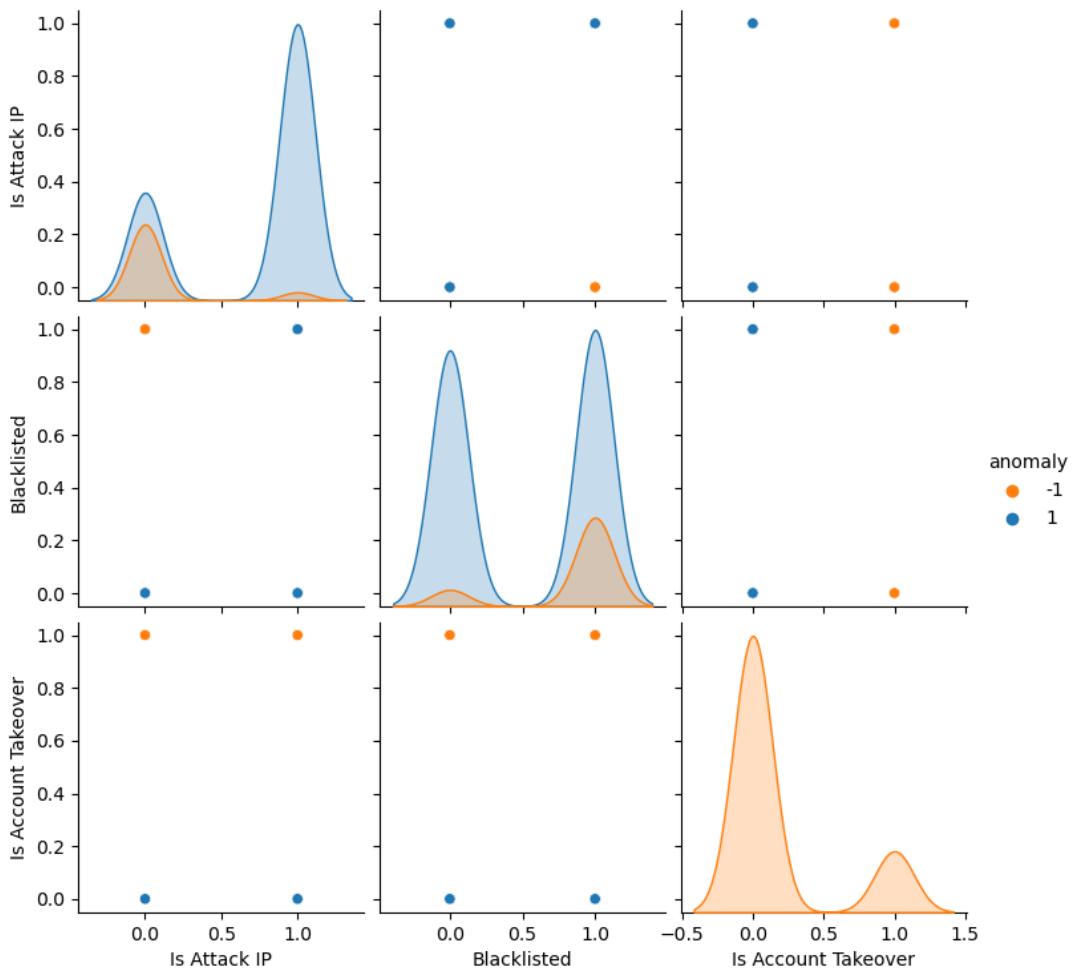
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



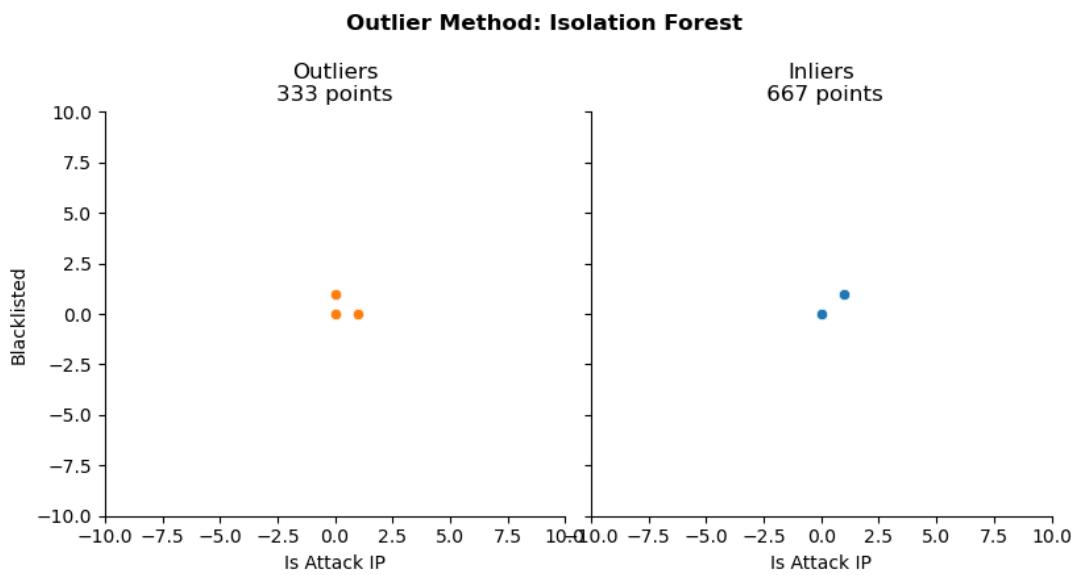


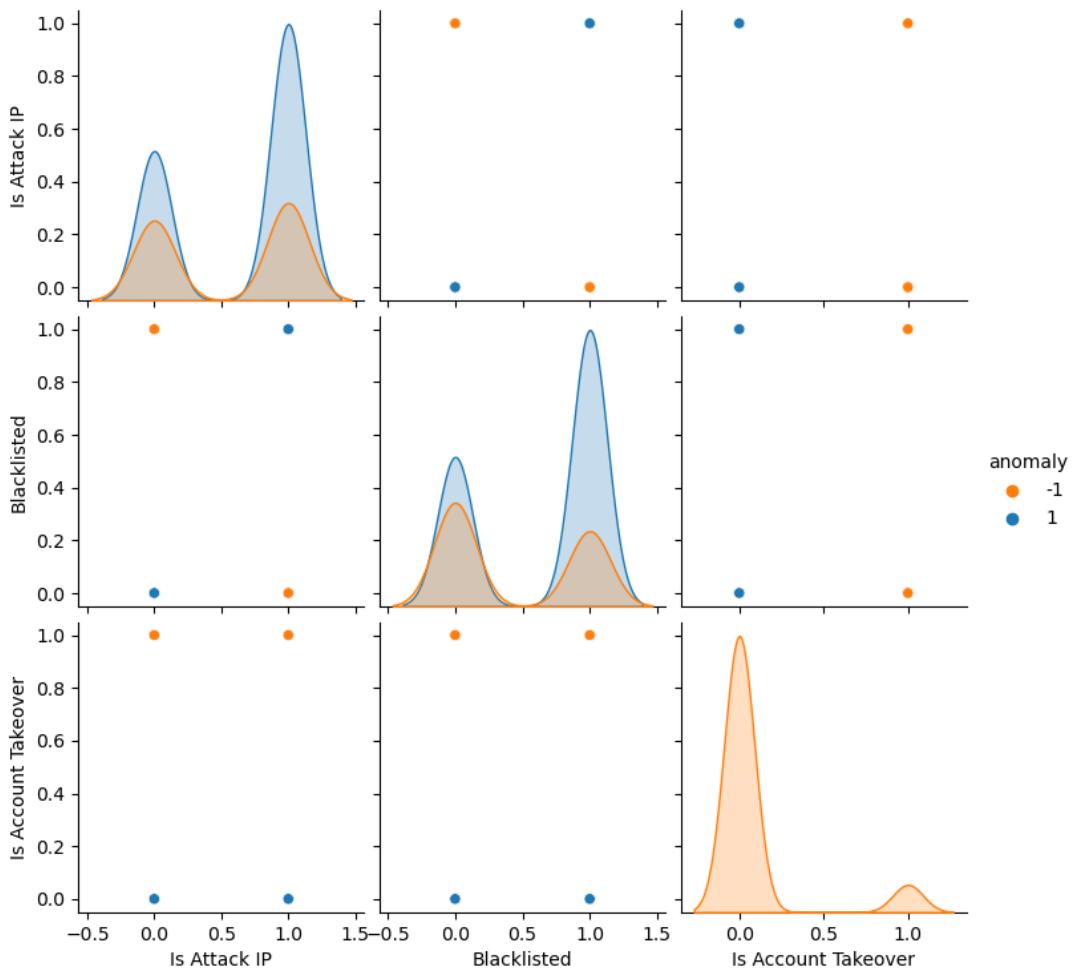
- contamination value == 0.3



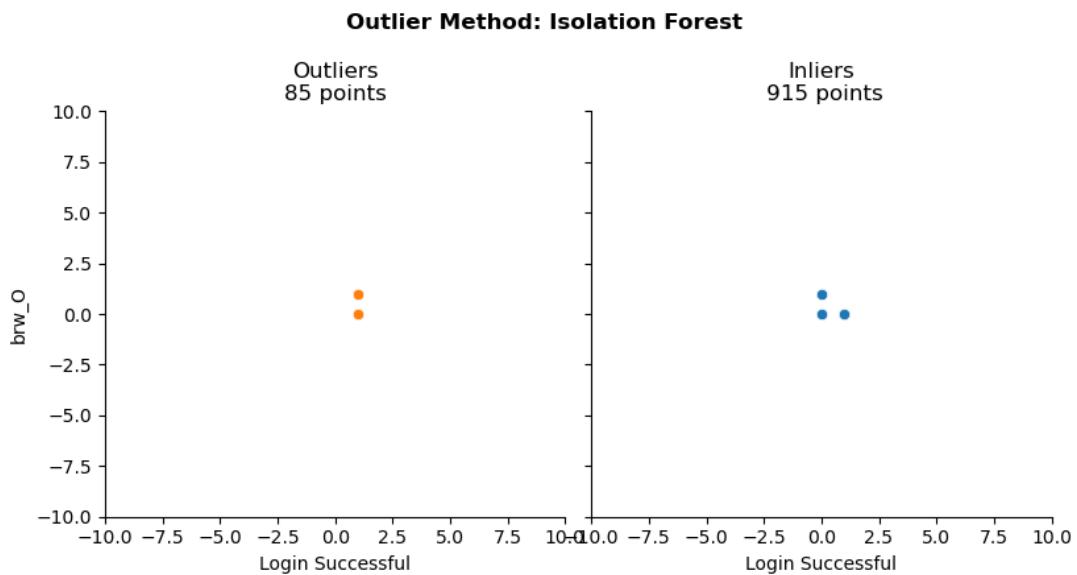


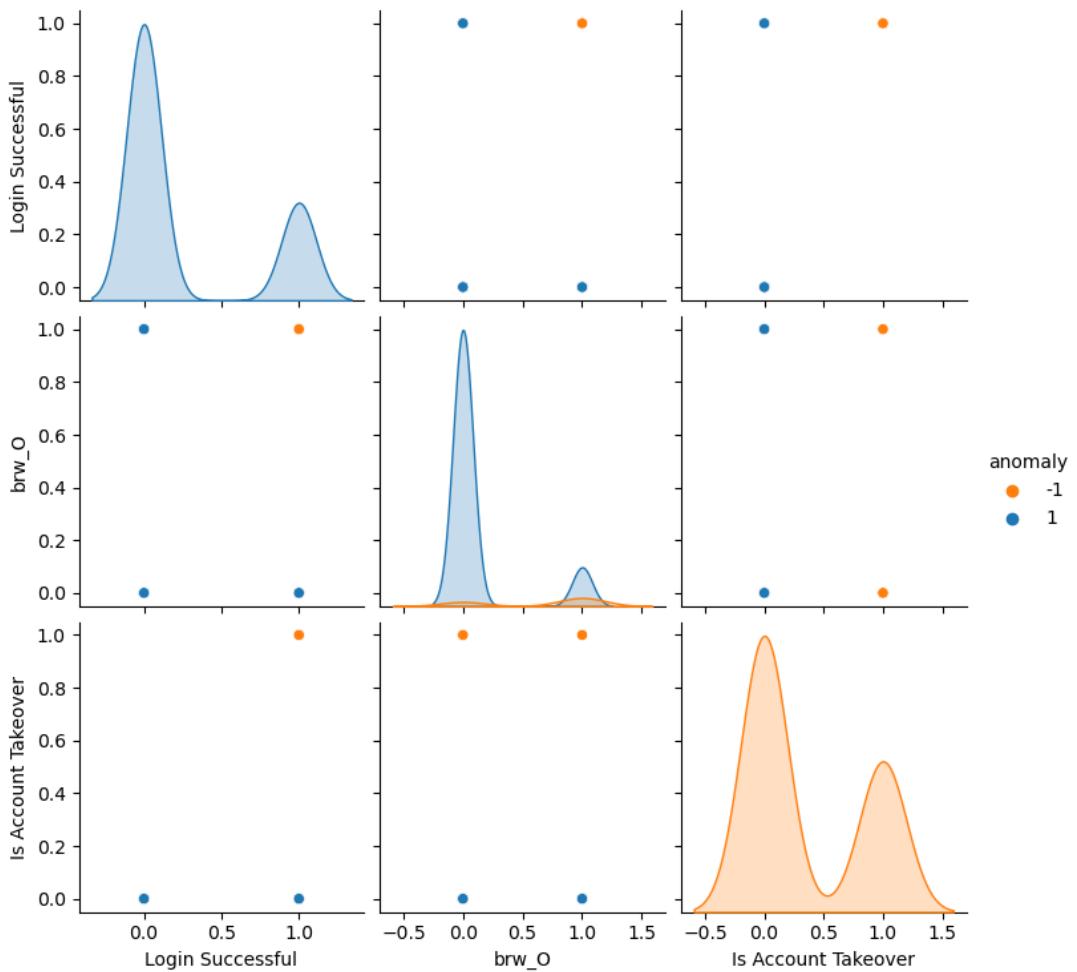
- contamination value == 0.5



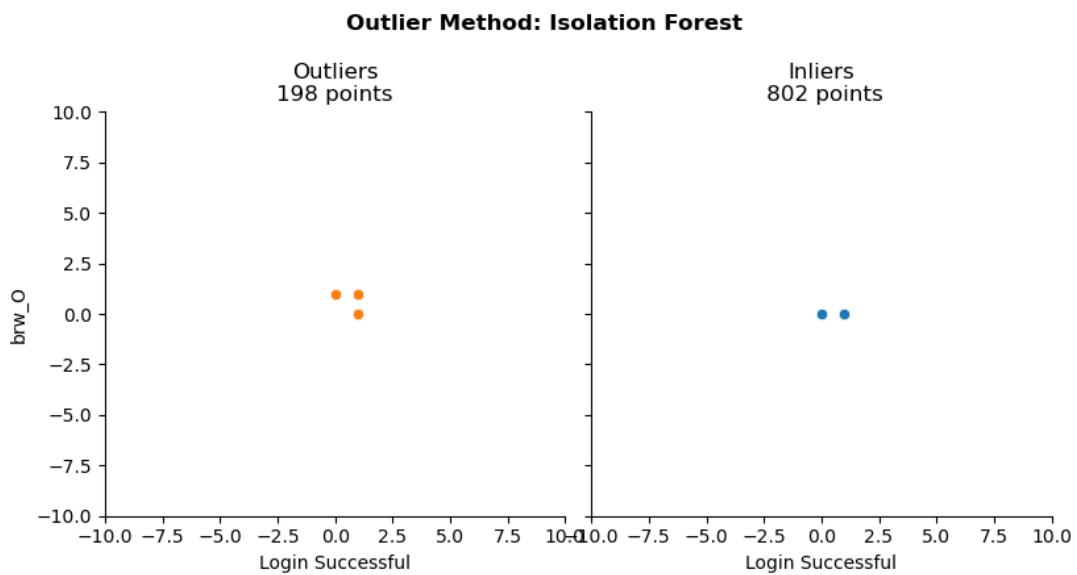


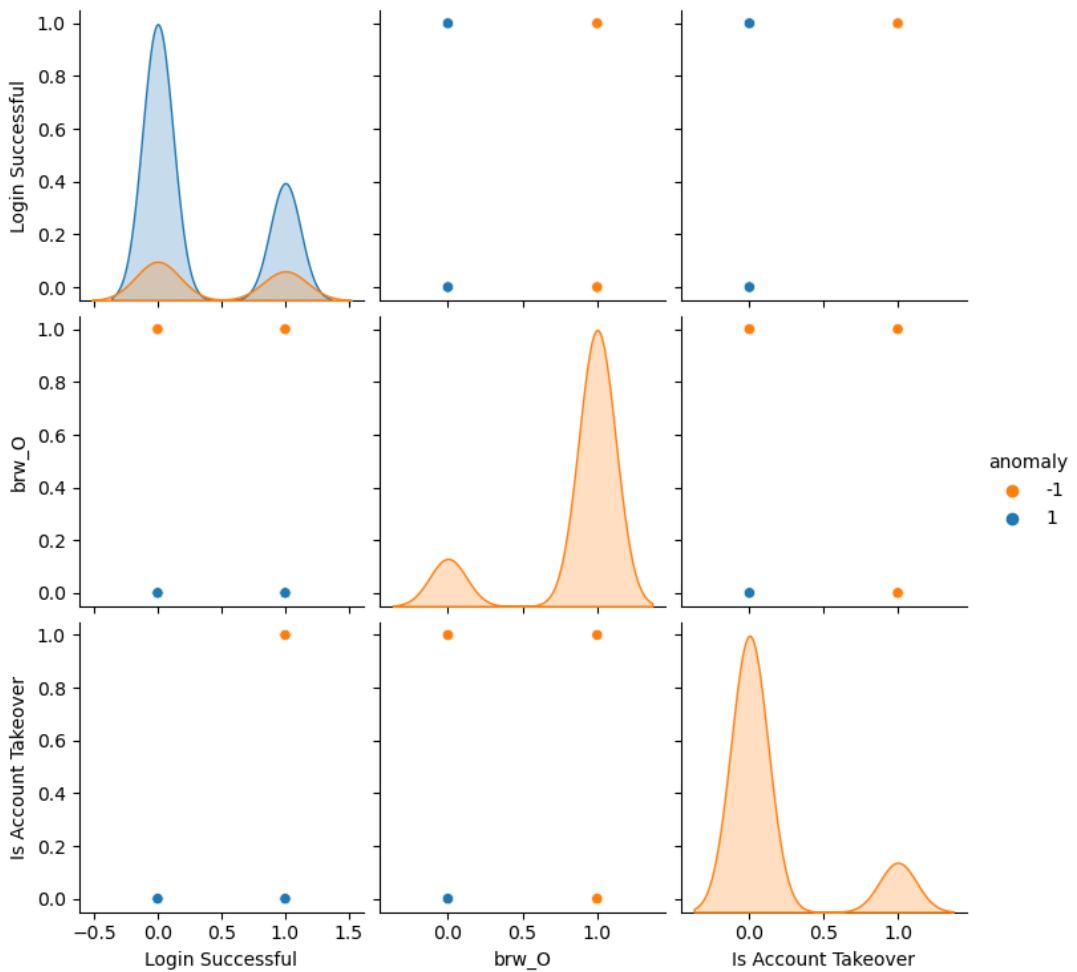
- anomaly_inputs = ['Login Successful', 'Browser Type (brw_O)', 'Is Account Takeover']
 - contamination value == 0.1



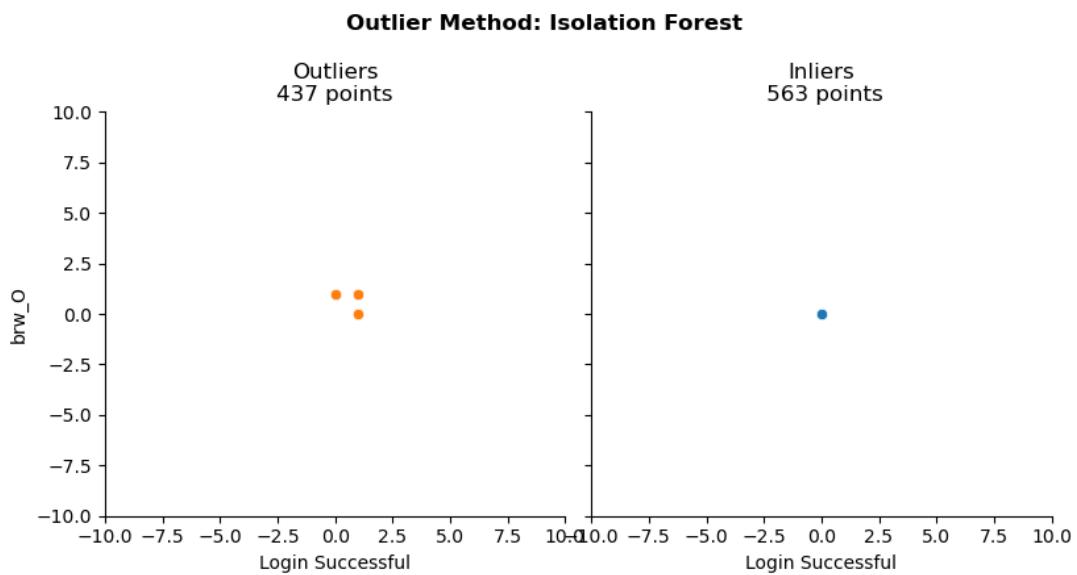


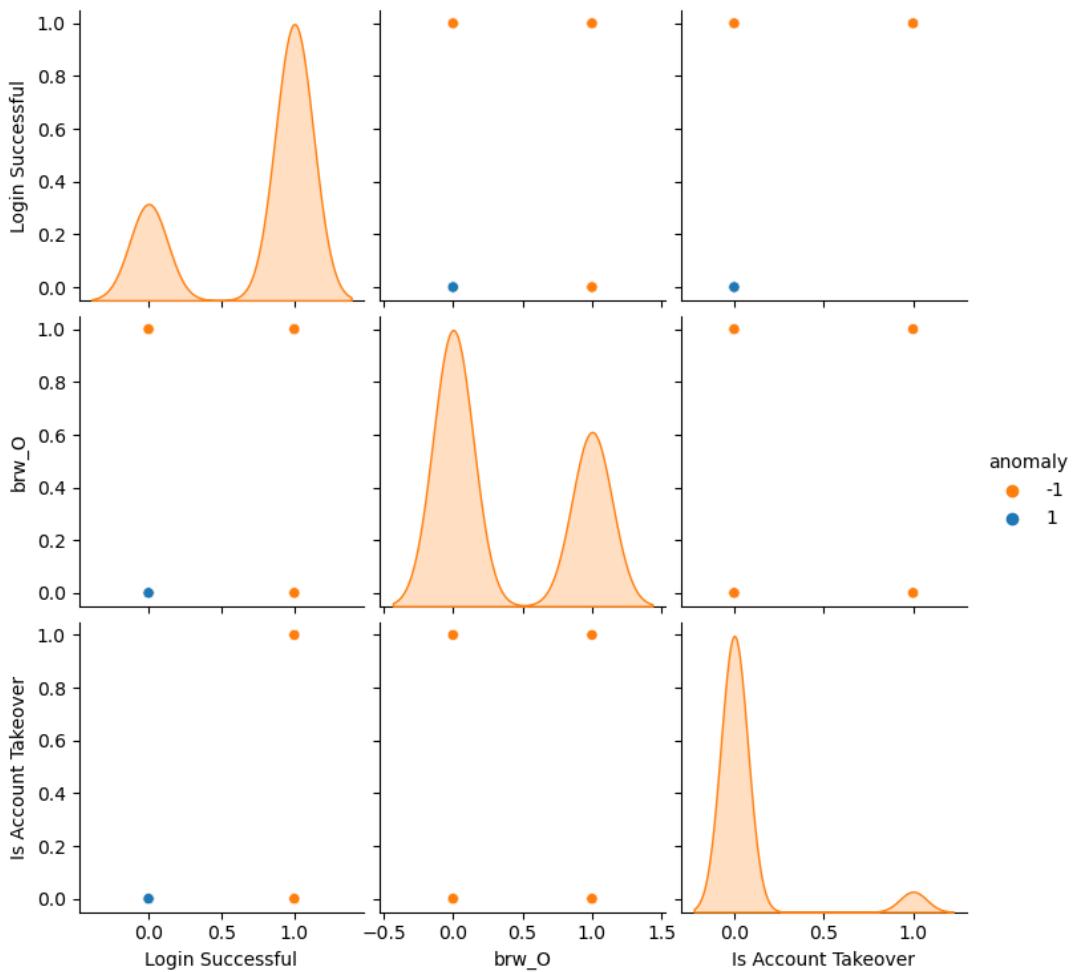
- contamination value == 0.3



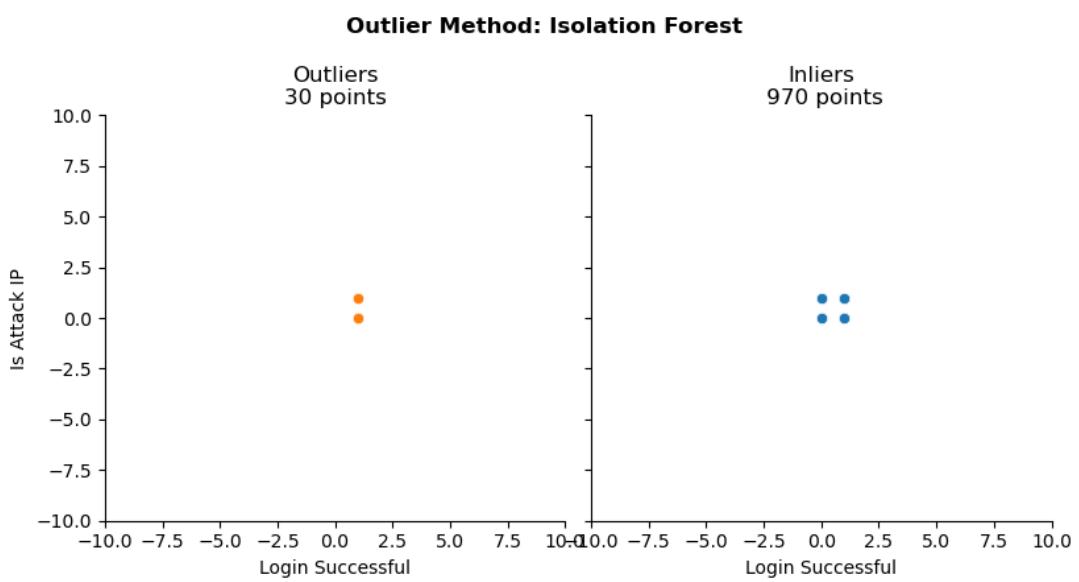


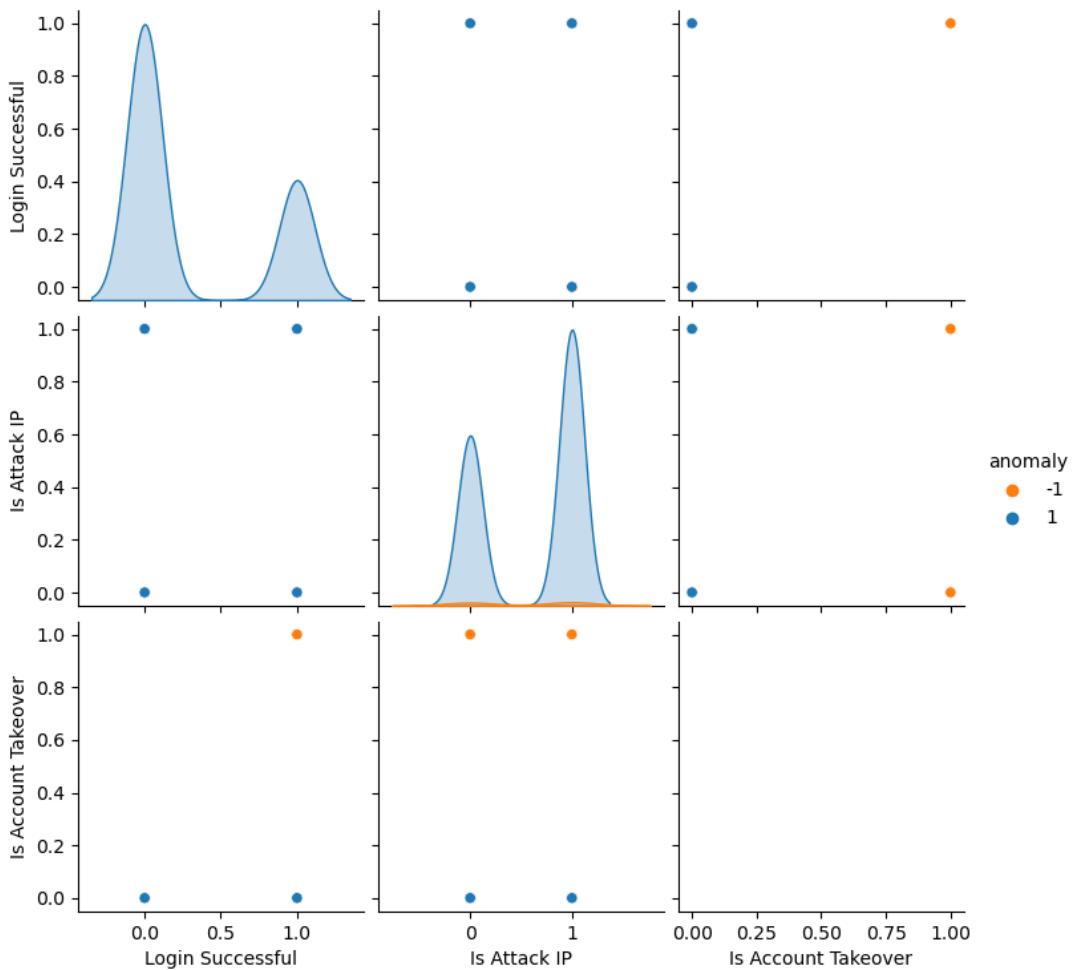
- contamination value == 0.5



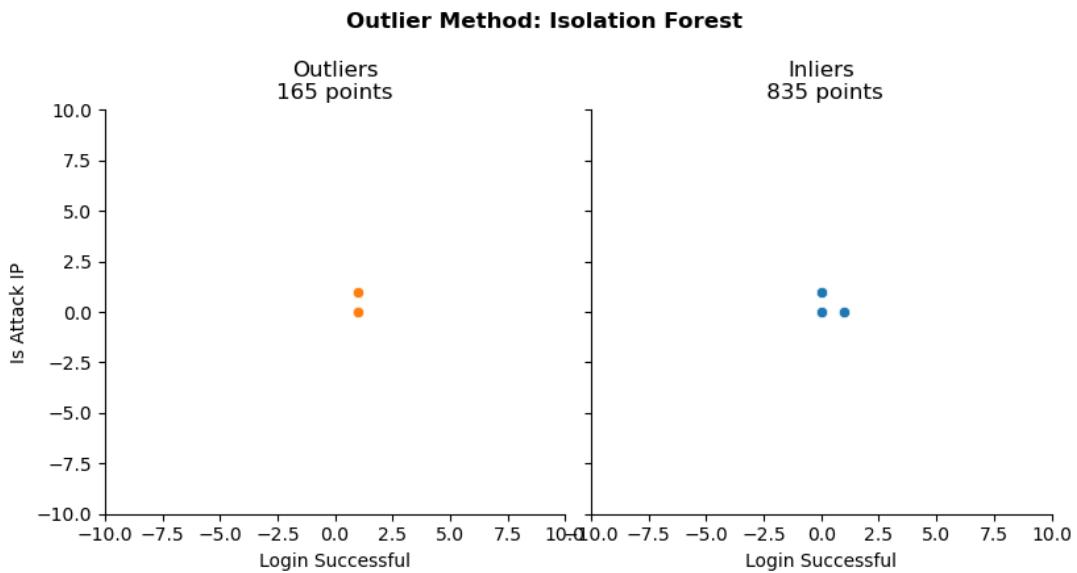


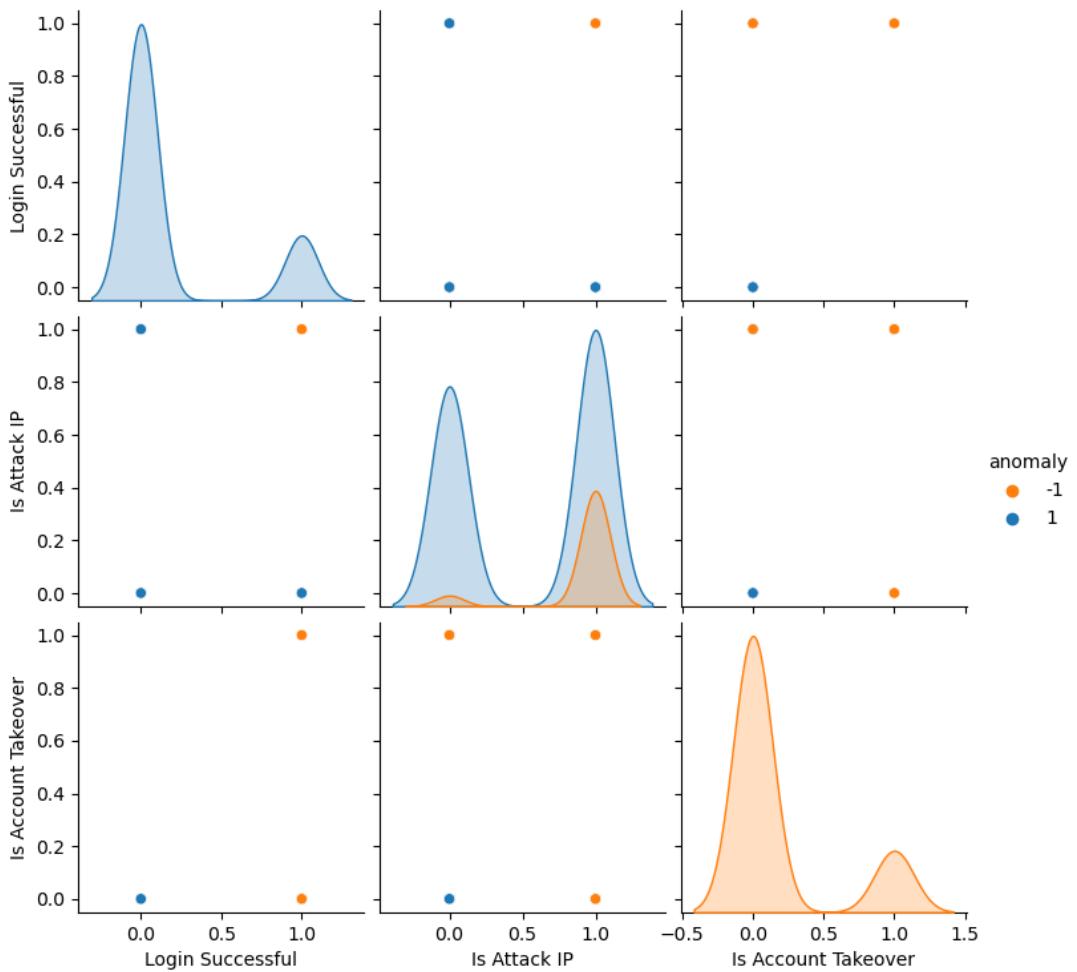
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1





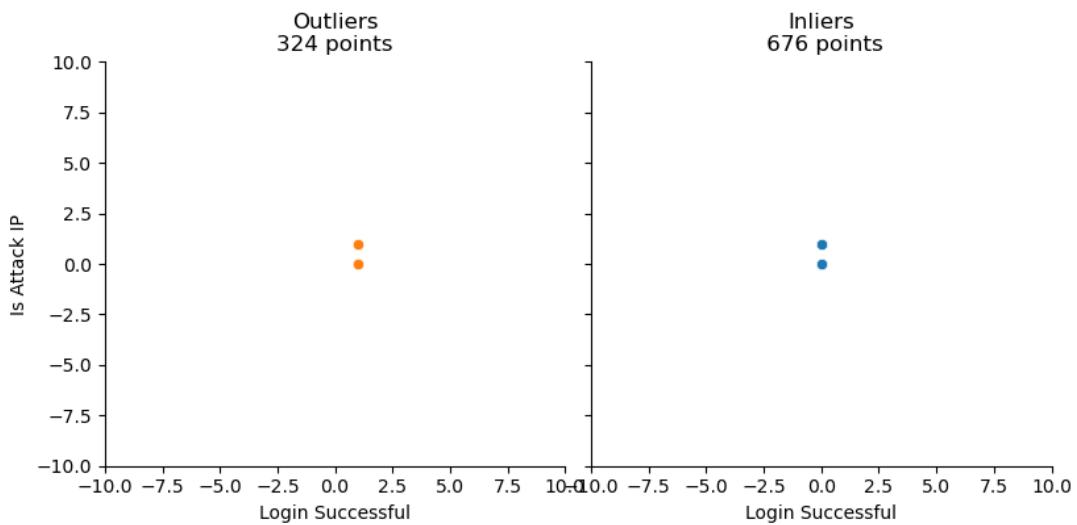
- contamination value == 0.3

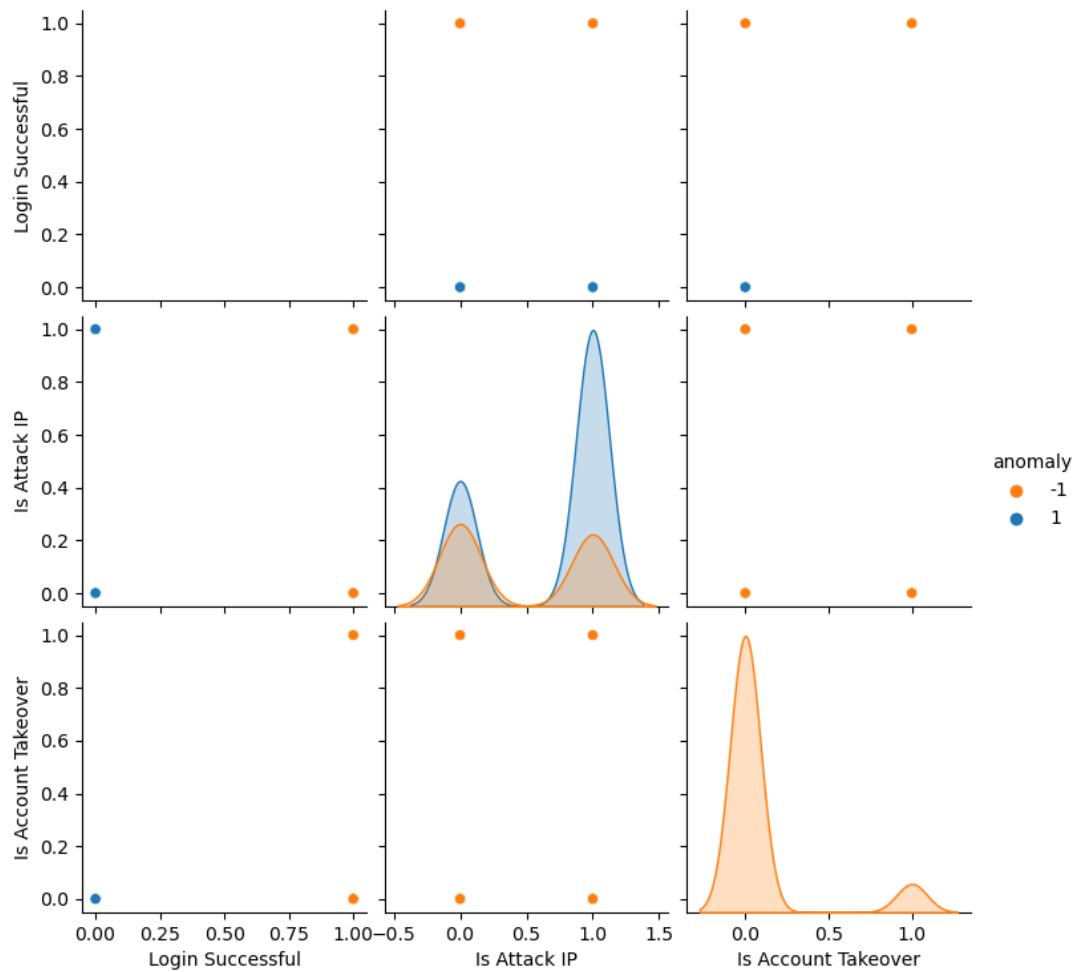




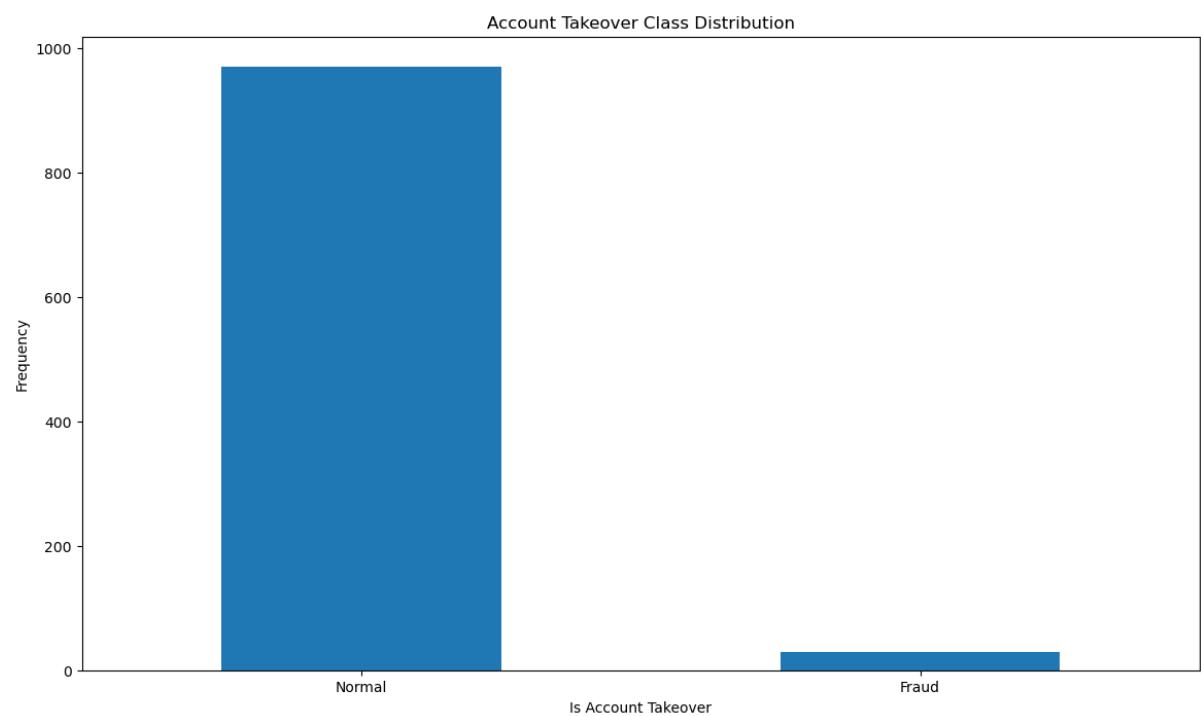
- contamination value == 0.5

Outlier Method: Isolation Forest





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



Isolation Forest: 61

Accuracy Score :

0.939

Classification Report :

	precision	recall	f1-score	support
False	0.97	0.97	0.97	970
True	0.00	0.00	0.00	30
accuracy			0.94	1000
macro avg	0.48	0.48	0.48	1000
weighted avg	0.94	0.94	0.94	1000

Local Outlier Factor: 55

Accuracy Score :

0.945

Classification Report :

	precision	recall	f1-score	support
False	0.97	0.97	0.97	970
True	0.10	0.10	0.10	30
accuracy			0.94	1000
macro avg	0.53	0.54	0.53	1000
weighted avg	0.95	0.94	0.95	1000

Support Vector Machine: 824

Accuracy Score :

0.176

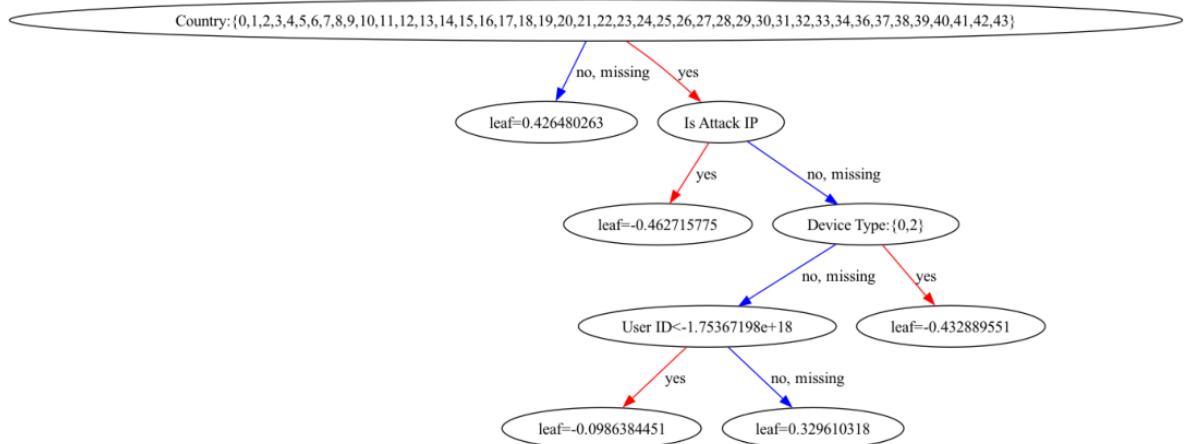
Classification Report :

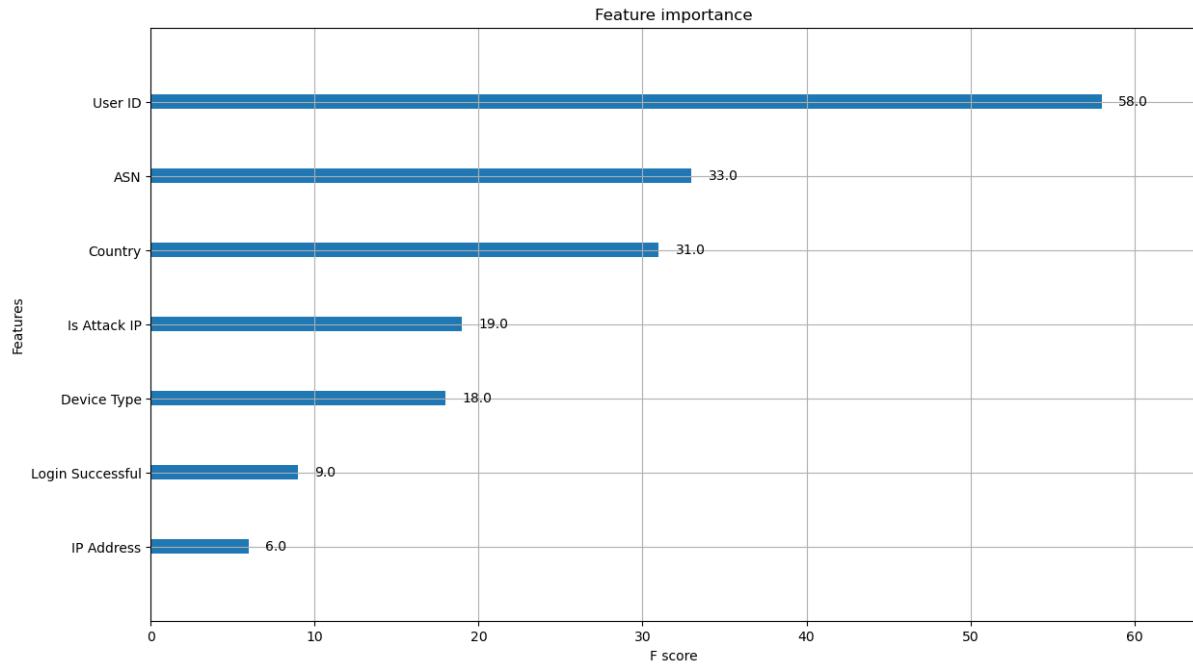
	precision	recall	f1-score	support
False	0.93	0.16	0.28	970
True	0.02	0.60	0.04	30
accuracy			0.18	1000
macro avg	0.48	0.38	0.16	1000
weighted avg	0.90	0.18	0.27	1000

III. 77/1000 → 7.7%

1. XGB Experimental

```
-----  
0  Login Timestamp      1000 non-null  category  
1  User ID              1000 non-null  int64  
2  IP Address            1000 non-null  category  
3  Country               1000 non-null  category  
4  ASN                  1000 non-null  int64  
5  Device Type           1000 non-null  category  
6  Login Successful       1000 non-null  bool  
7  Is Attack IP          1000 non-null  bool  
8  Blacklisted            1000 non-null  bool  
9  Browser Type           1000 non-null  category  
10 Is Account Takeover    1000 non-null  bool  
dtypes: bool(4), category(5), int64(2)  
memory usage: 107.6 KB  
Feature importances:  
[0.          0.01772542 0.03462604 0.70550615 0.00749496 0.09745706  
 0.07308619 0.0641042  0.          0.          ]
```





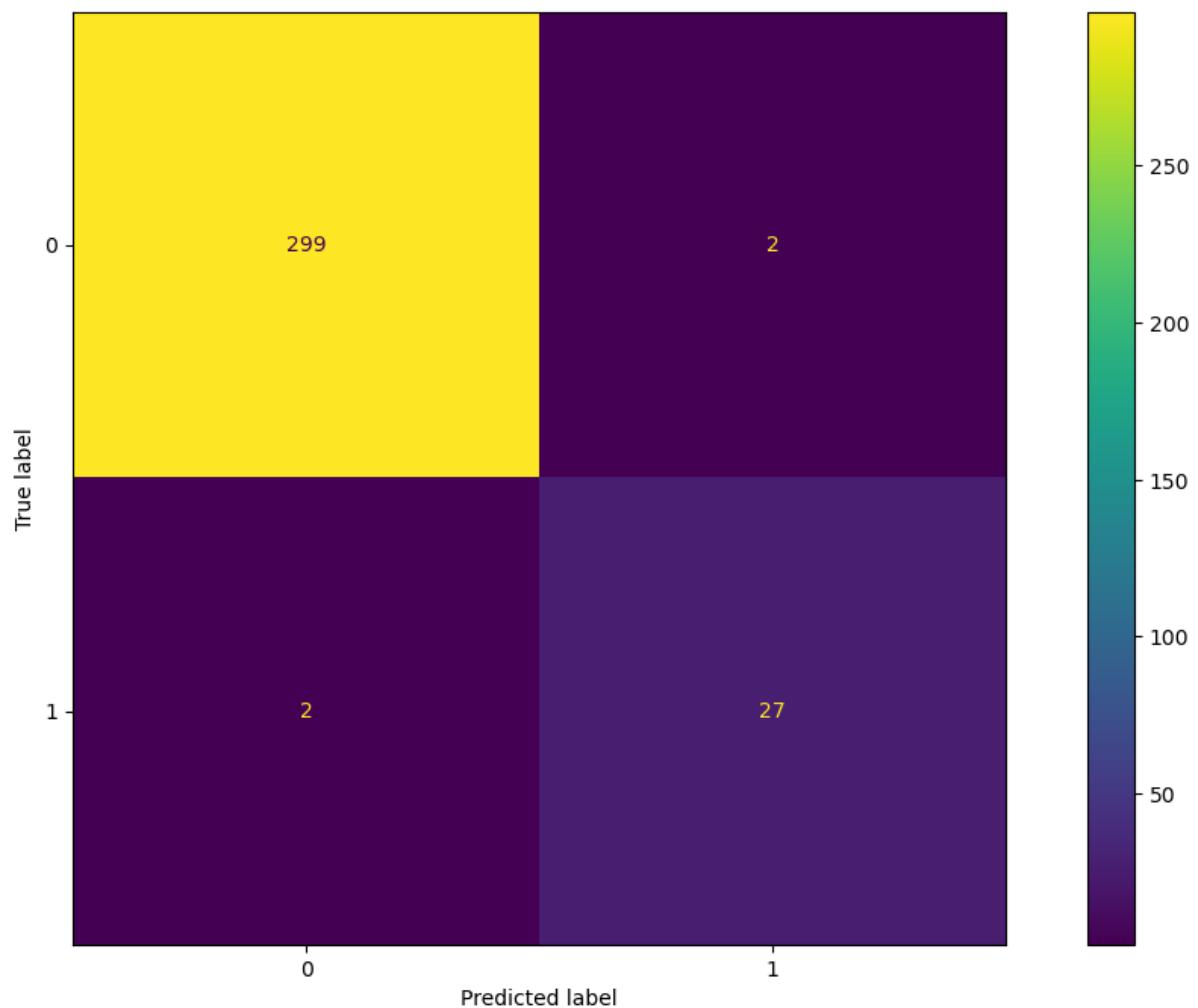
2. Label encoding (cat.codes)

2A. XGBoost

memory usage: 26.5 KB

Accuracy: 98.79%

	precision	recall	f1-score	support
False	0.99	0.99	0.99	301
True	0.93	0.93	0.93	29
accuracy			0.99	330
macro avg	0.96	0.96	0.96	330
weighted avg	0.99	0.99	0.99	330



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 800
The number of records in the test dataset is 200
The training dataset has 742 records for the majority class and 58 records for the minority class.

	precision	recall	f1-score	support
False	0.00	0.00	0.00	181
True	0.10	1.00	0.17	19
accuracy			0.10	200
macro avg	0.05	0.50	0.09	200
weighted avg	0.01	0.10	0.02	200

The customized score threshold for 2% of outliers is 0.00

	precision	recall	f1-score	support
False	0.91	1.00	0.95	181
True	0.00	0.00	0.00	19
accuracy			0.91	200
macro avg	0.45	0.50	0.48	200
weighted avg	0.82	0.91	0.86	200

```
[[ 0 181]
 [ 0 19]]
```

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

The number of records in the training dataset is 700
The number of records in the test dataset is 300
The training dataset has 655 records for the majority class and 45 records for the minority class.

	precision	recall	f1-score	support
False	0.00	0.00	0.00	268
True	0.11	1.00	0.19	32
accuracy			0.11	300
macro avg	0.05	0.50	0.10	300
weighted avg	0.01	0.11	0.02	300

The customized score threshold for 2% of outliers is 0.00

	precision	recall	f1-score	support
False	0.89	1.00	0.94	268
True	0.00	0.00	0.00	32
accuracy			0.89	300
macro avg	0.45	0.50	0.47	300
weighted avg	0.80	0.89	0.84	300

```
[[ 0 268]
 [ 0 32]]
```

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

```

The number of records in the training dataset is 500
The number of records in the test dataset is 500
The training dataset has 472 records for the majority class and 28 records for the minority class.
      precision    recall   f1-score   support
False        0.00     0.00     0.00    451
True        0.10     1.00     0.18     49

accuracy                           0.10    500
macro avg       0.05     0.50     0.09    500
weighted avg    0.01     0.10     0.02    500

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.90     1.00     0.95    451
True        0.00     0.00     0.00     49

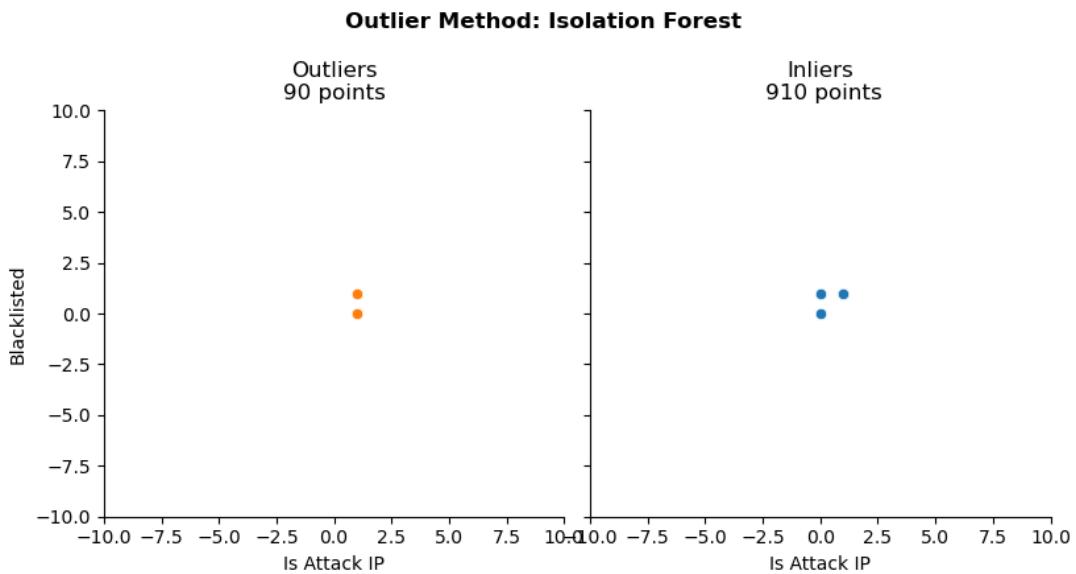
accuracy                           0.90    500
macro avg       0.45     0.50     0.47    500
weighted avg    0.81     0.90     0.86    500

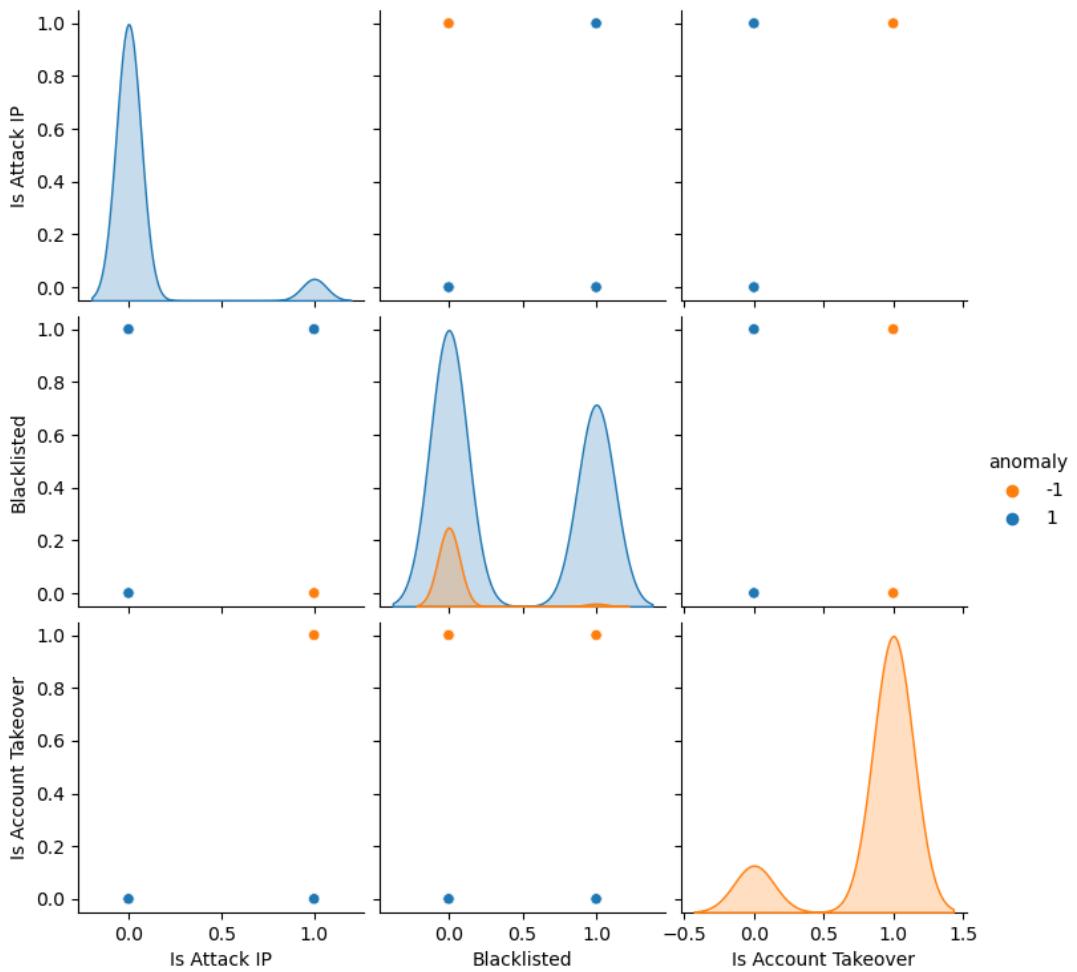
[[ 0 451]
 [ 0 49]]

```

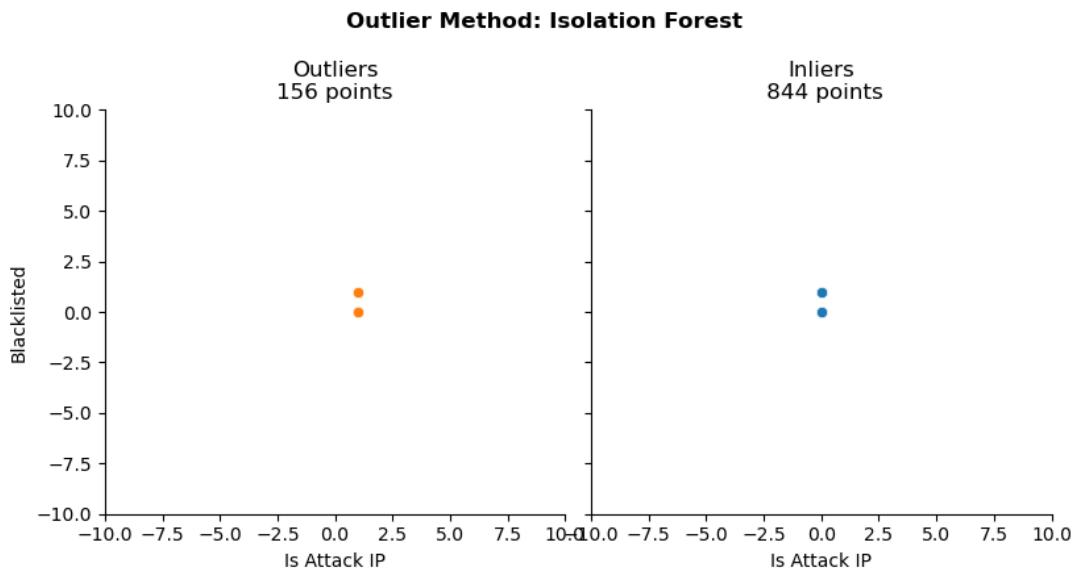
2C. Isolation Forest

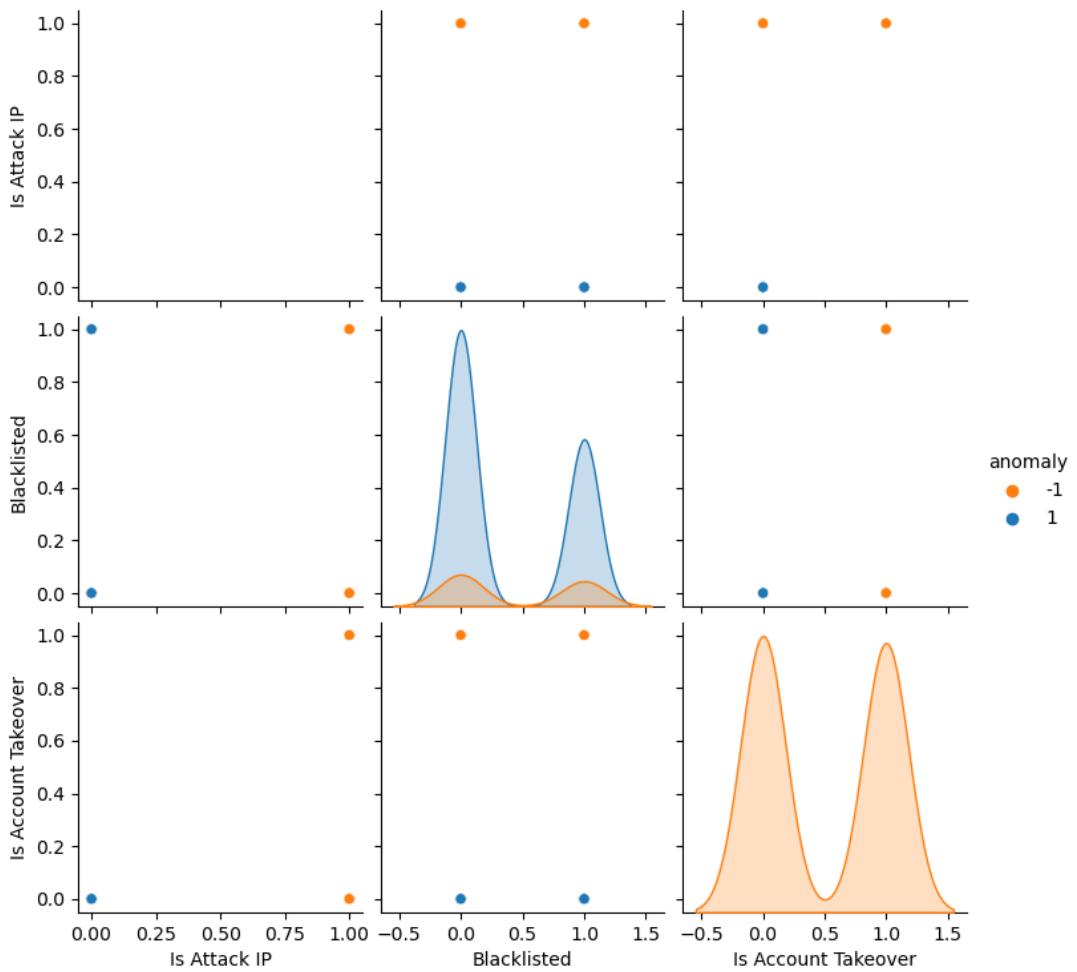
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



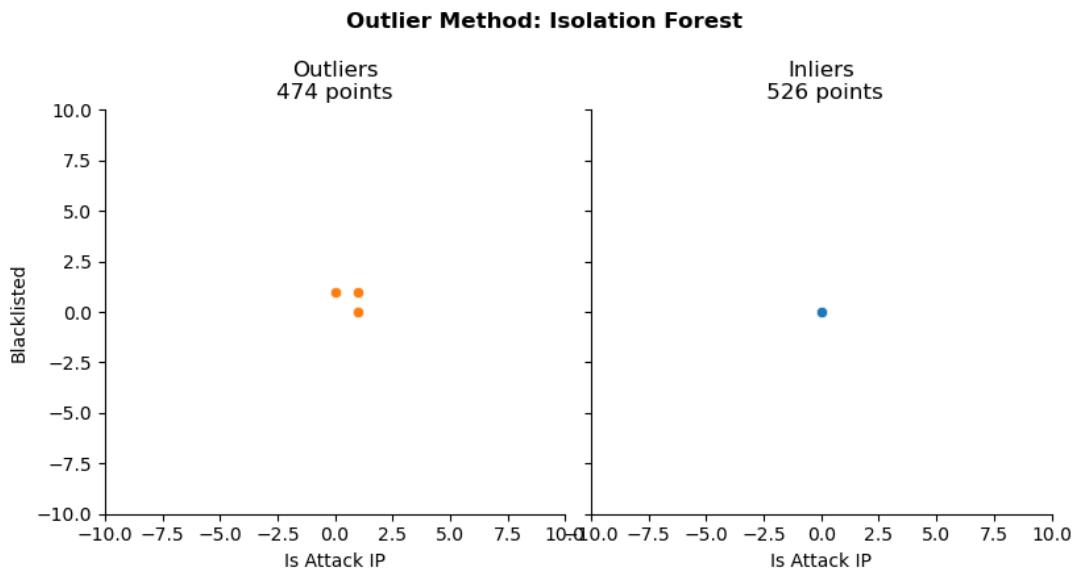


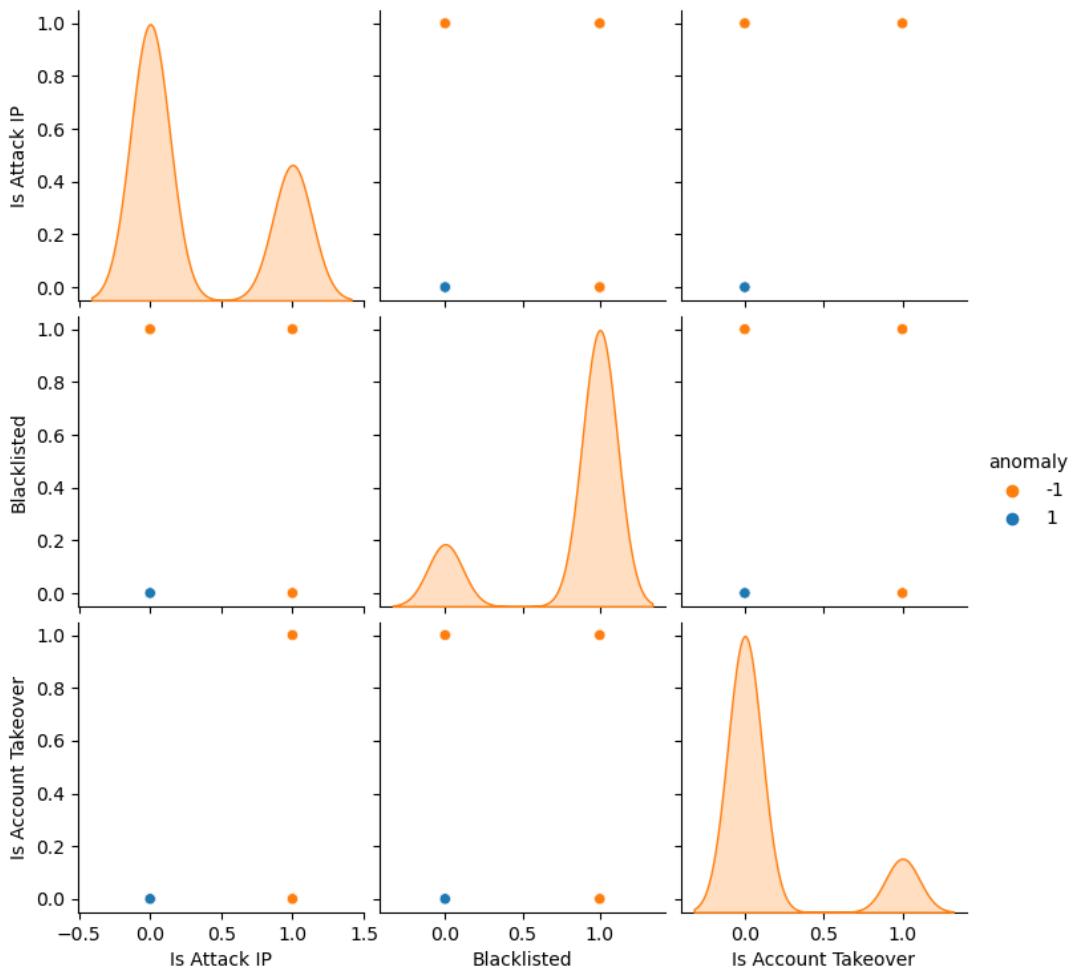
- increasing contamination value to 0.3



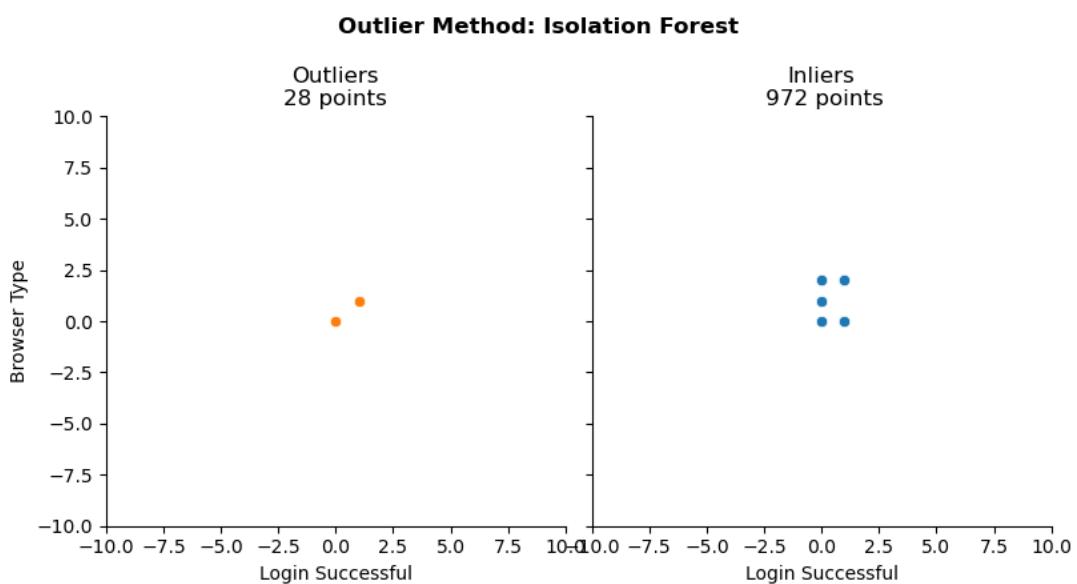


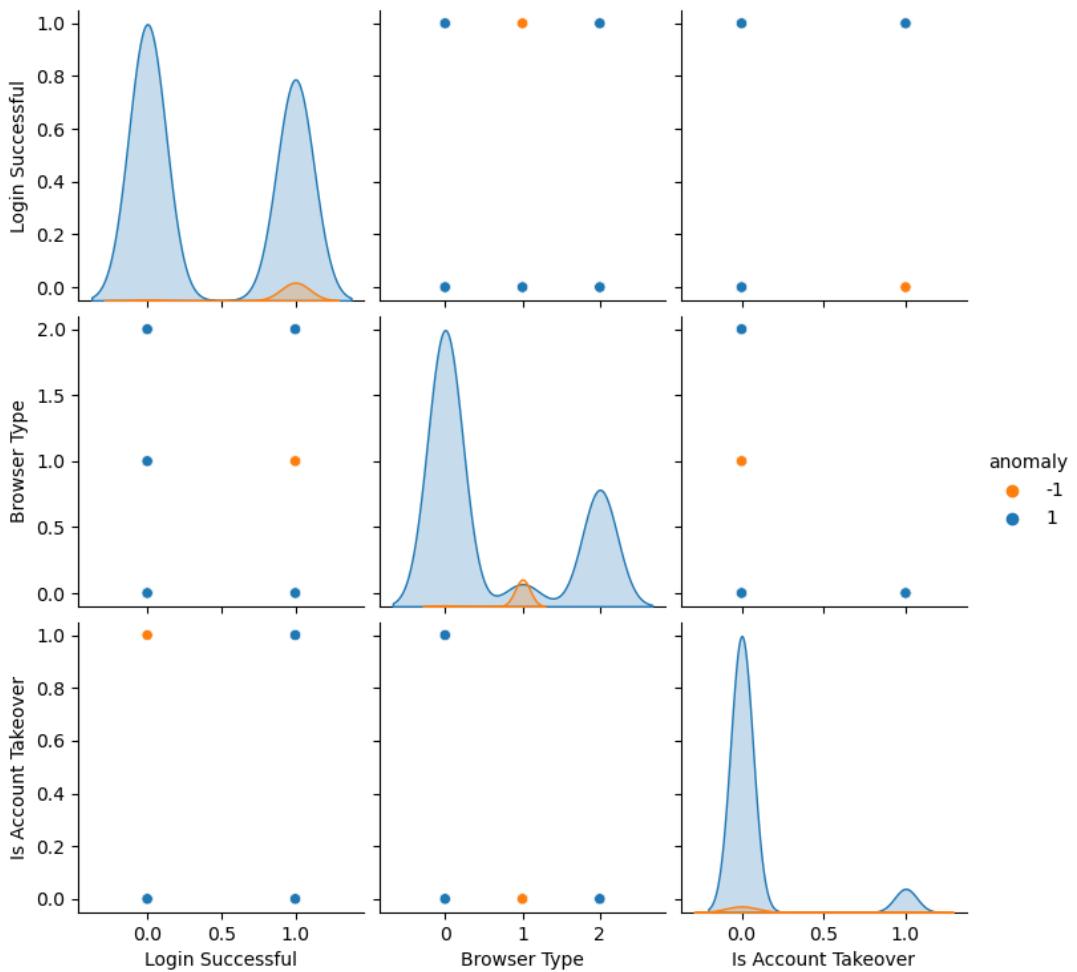
- contamination value == 0.5



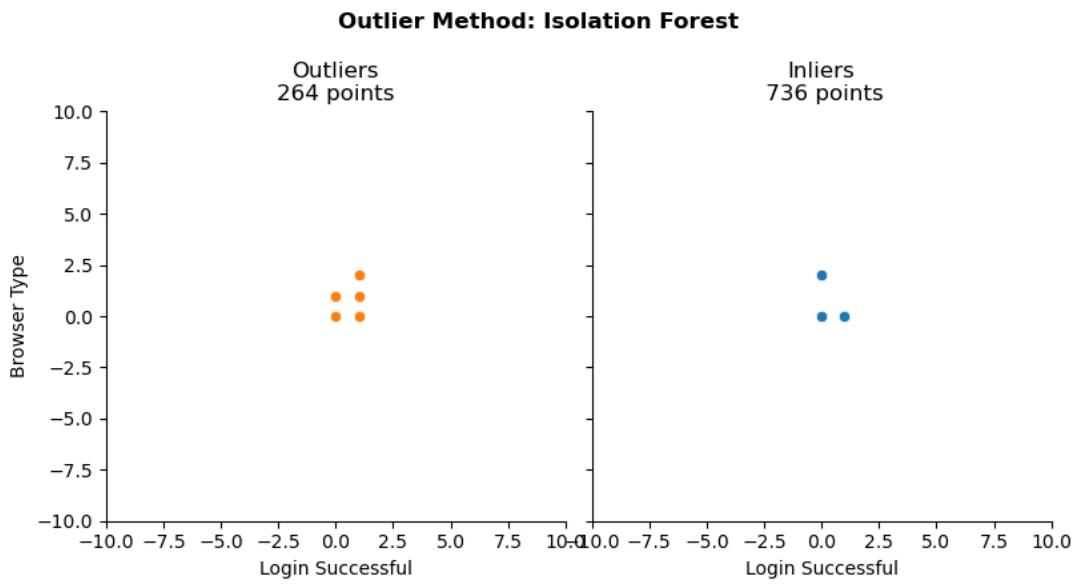


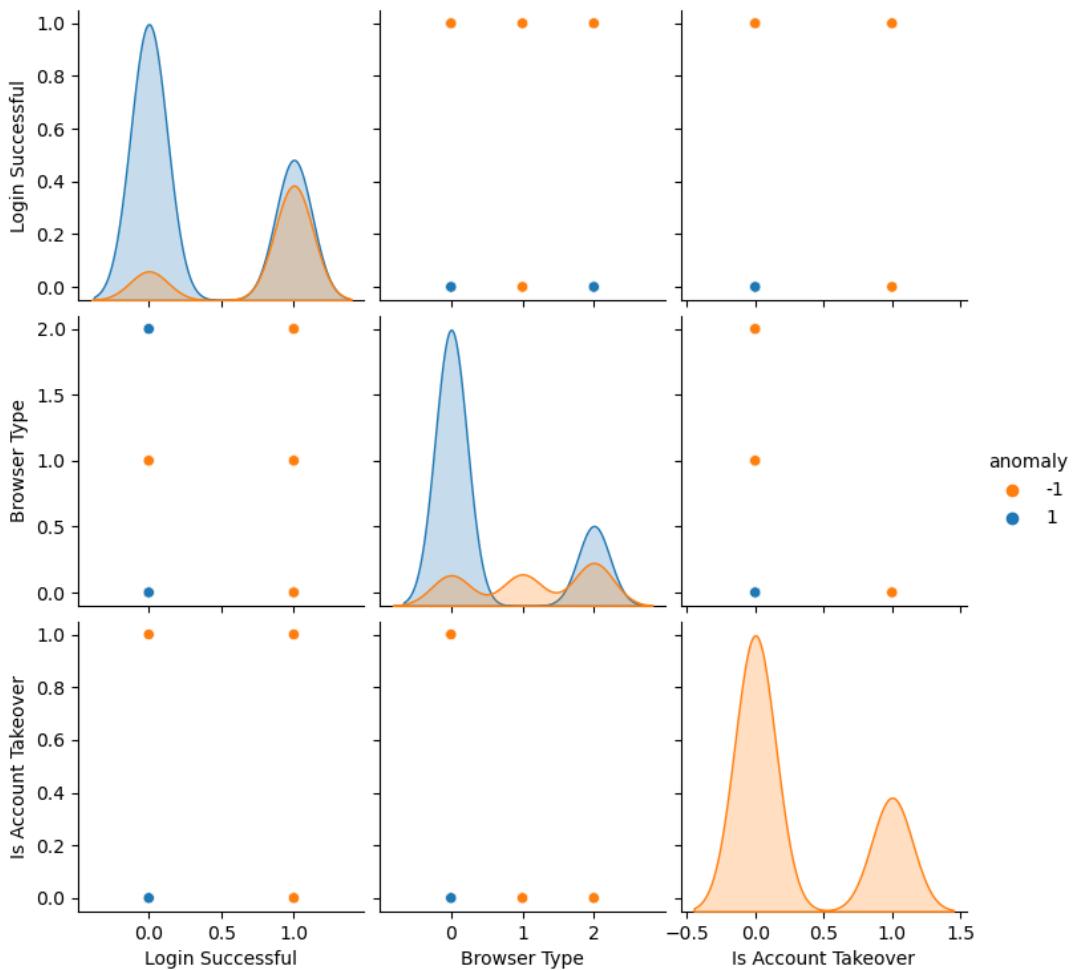
- anomaly_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']
 - contamination value == 0.1



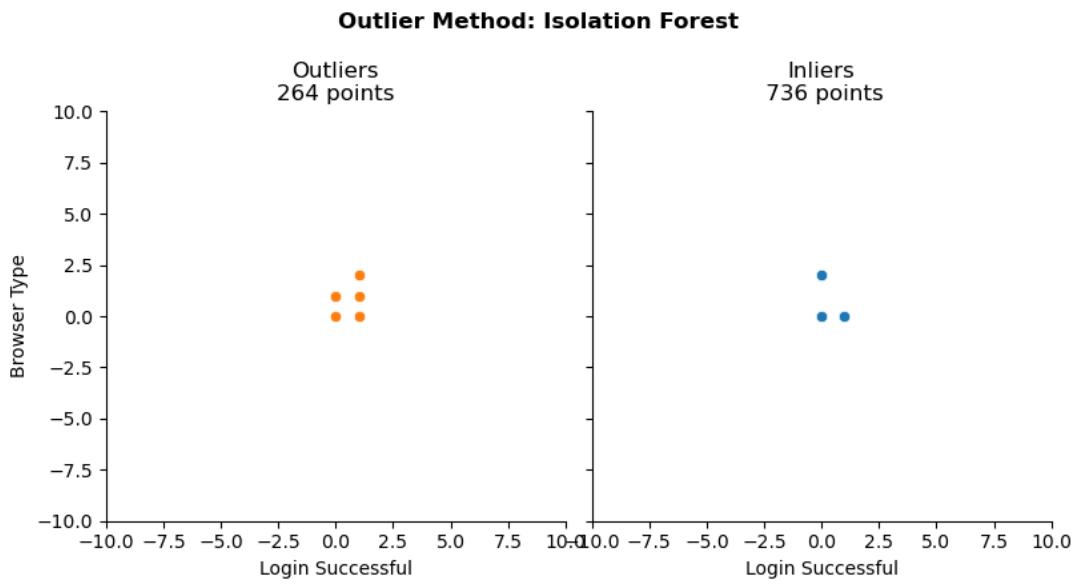


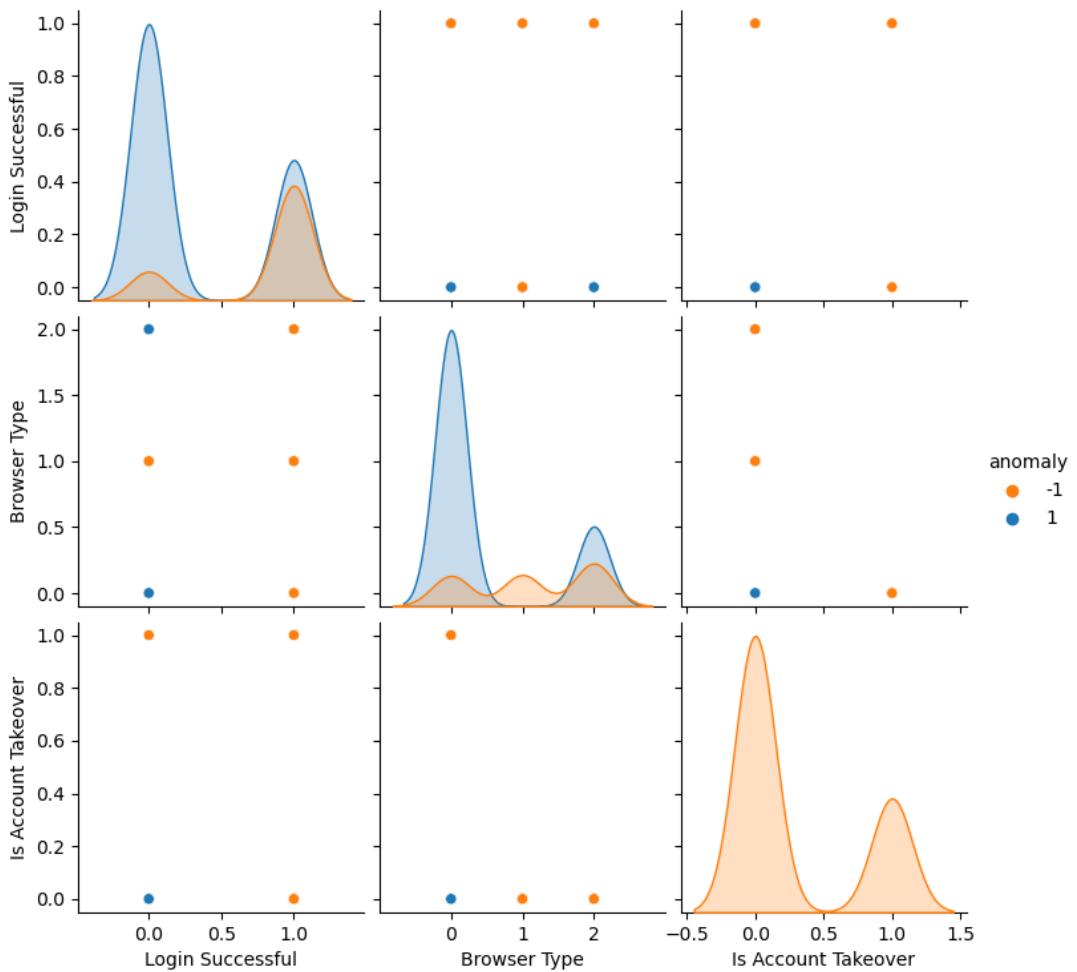
- contamination value == 0.3



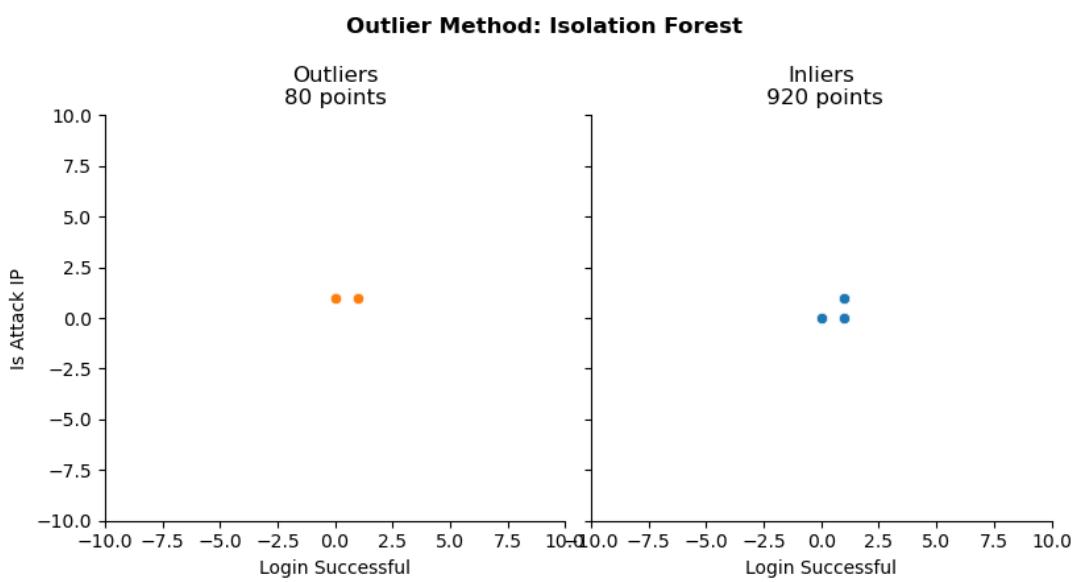


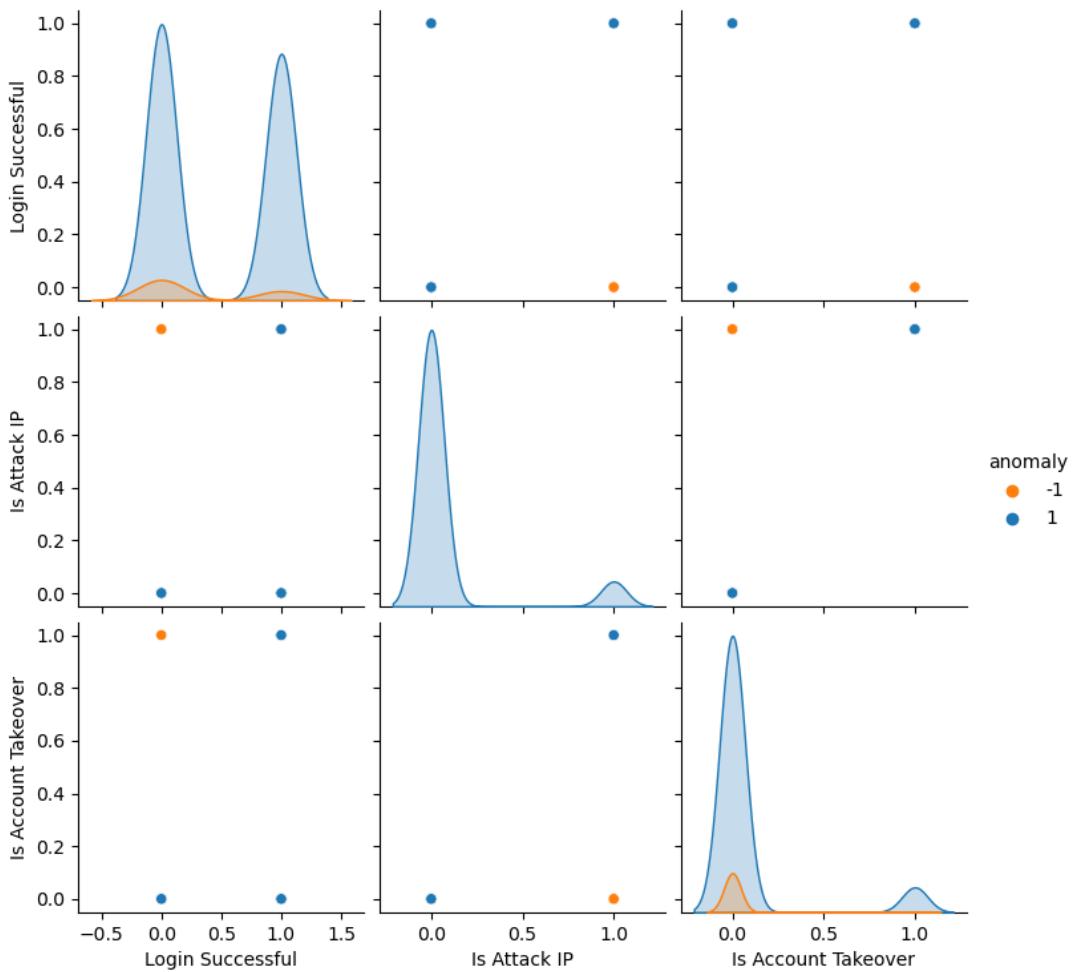
- contamination value == 0.5





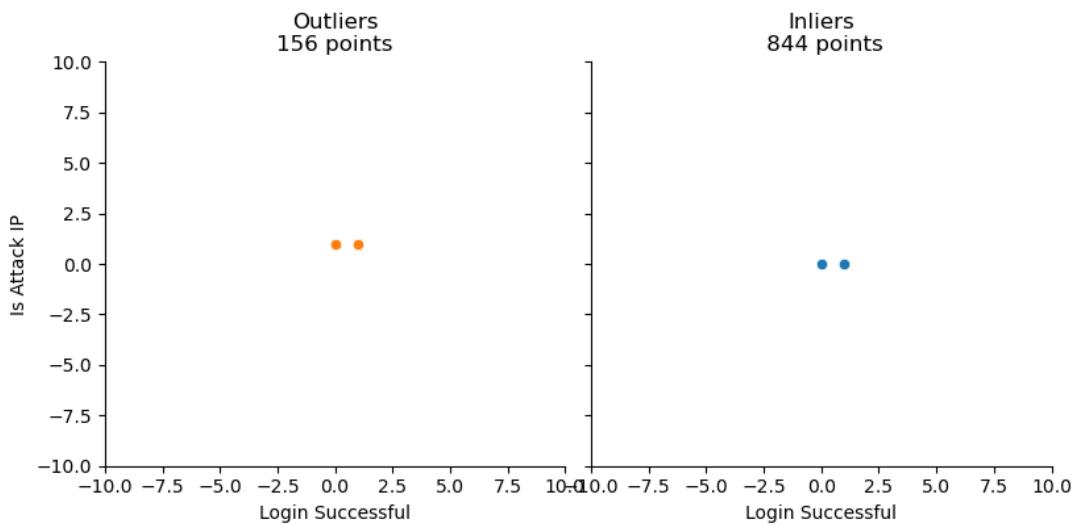
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1

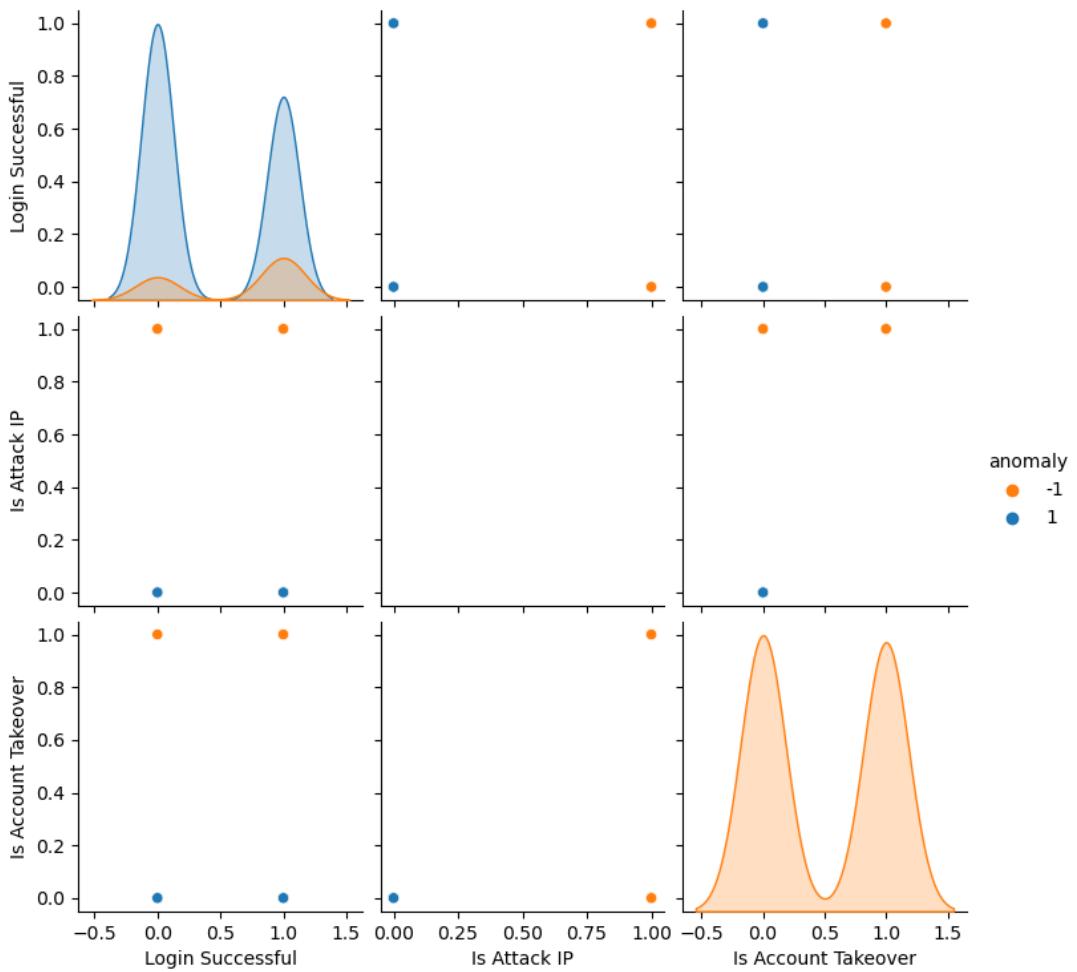




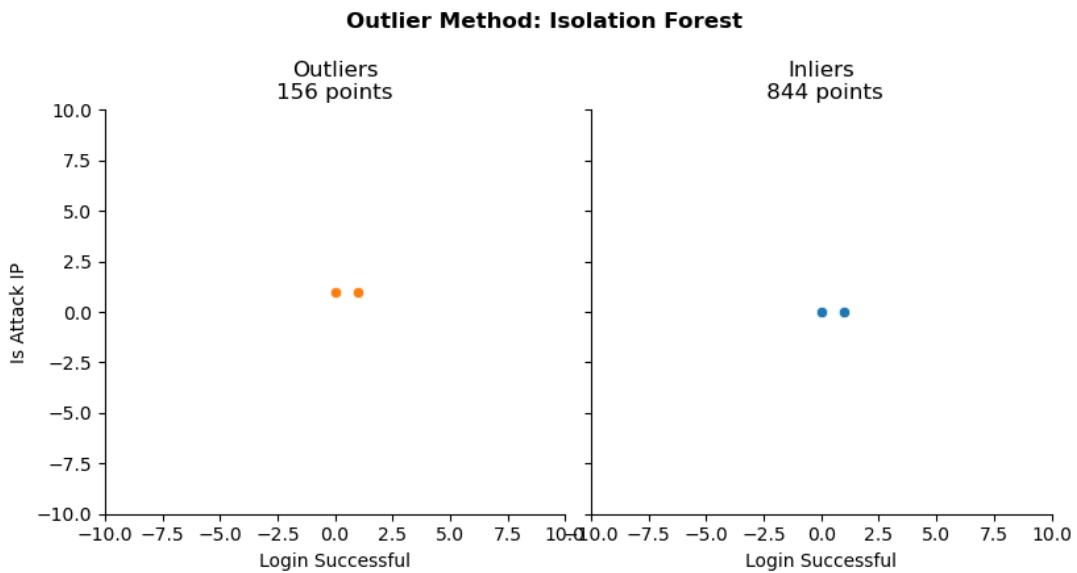
- contamination value == 0.3

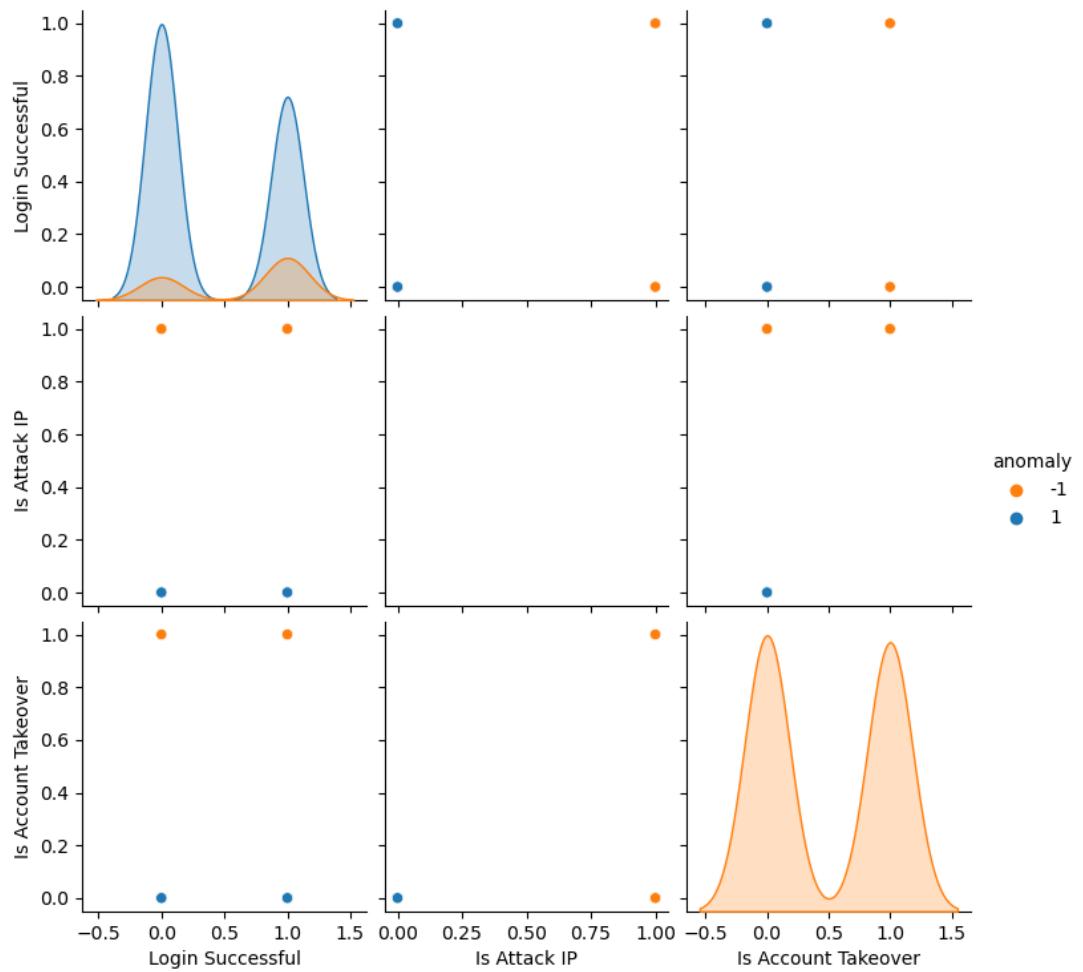
Outlier Method: Isolation Forest



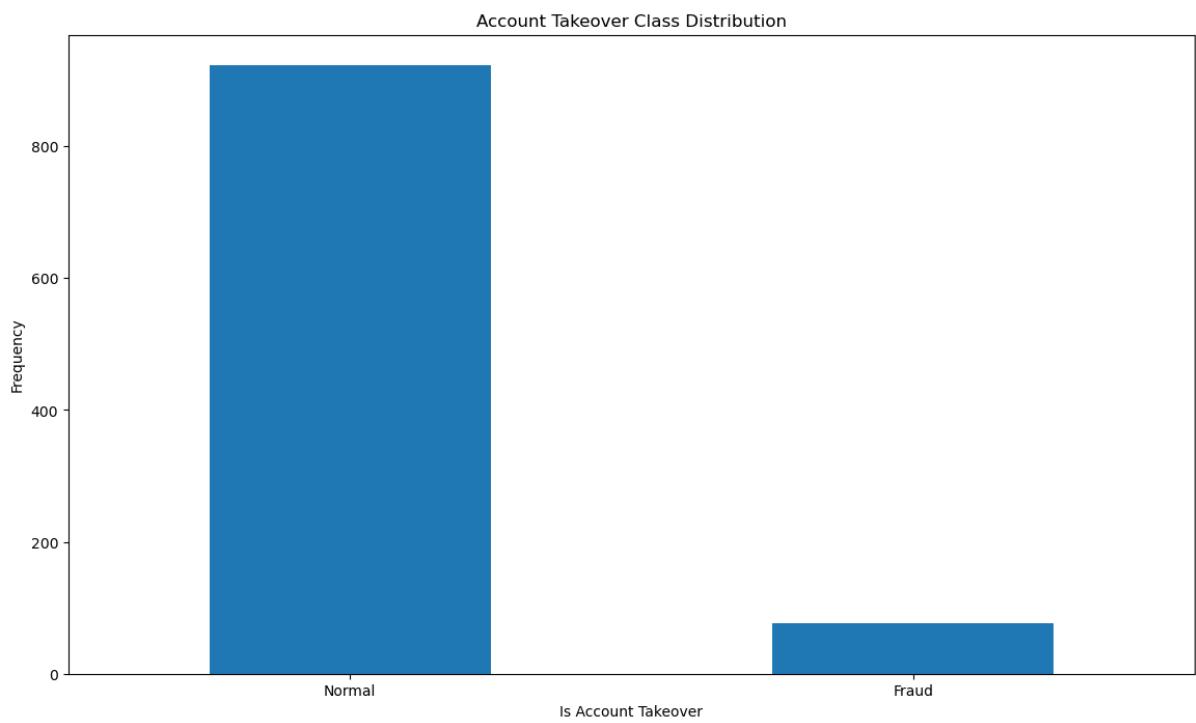


- contamination value == 0.5





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



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

Isolation Forest: 159

Accuracy Score :

0.841

Classification Report :

	precision	recall	f1-score	support
False	0.92	0.90	0.91	923
True	0.09	0.12	0.10	77
accuracy			0.84	1000
macro avg	0.51	0.51	0.51	1000
weighted avg	0.86	0.84	0.85	1000

Local Outlier Factor: 169

Accuracy Score :

0.831

Classification Report :

	precision	recall	f1-score	support
False	0.92	0.90	0.91	923
True	0.04	0.05	0.05	77
accuracy			0.83	1000
macro avg	0.48	0.47	0.48	1000
weighted avg	0.85	0.83	0.84	1000

Support Vector Machine: 603

Accuracy Score :

0.397

Classification Report :

	precision	recall	f1-score	support
False	0.88	0.40	0.55	923
True	0.04	0.32	0.08	77
accuracy			0.40	1000
macro avg	0.46	0.36	0.31	1000
weighted avg	0.81	0.40	0.52	1000

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

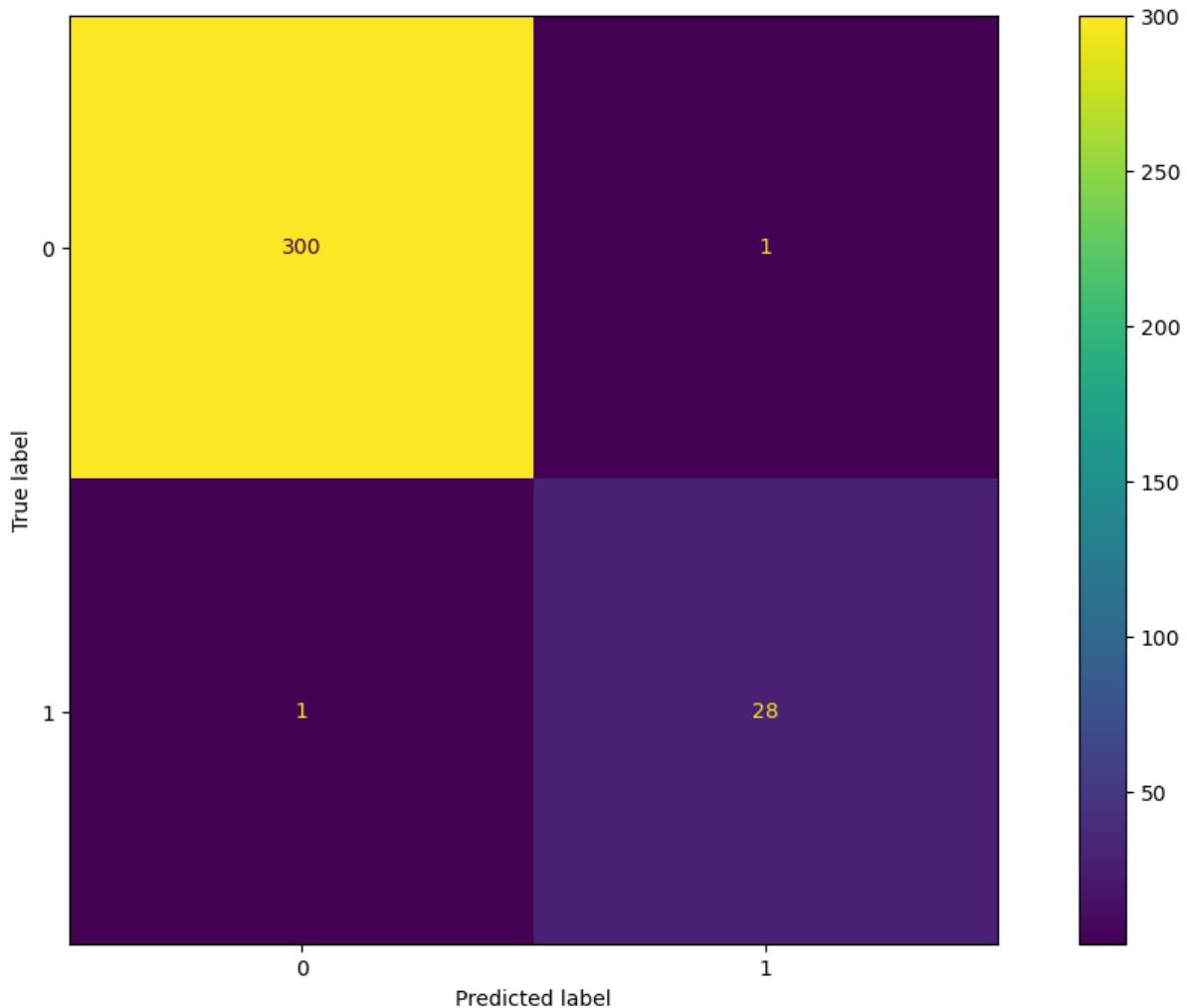
Type', 'IP Address', 'Login Timestamp', 'Browser Type'], prefix=["cntr", "dvc", "ip", "Ignt", "brw"])

3A. XGBoost

memory usage: 1.9 MB

Accuracy: 99.39%

	precision	recall	f1-score	support
False	1.00	1.00	1.00	301
True	0.97	0.97	0.97	29
accuracy			0.99	330
macro avg	0.98	0.98	0.98	330
weighted avg	0.99	0.99	0.99	330



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 800
The number of records in the test dataset is 200
The training dataset has 742 records for the majority class and 58 records for the minority class.
```

	precision	recall	f1-score	support
False	0.00	0.00	0.00	181
True	0.10	1.00	0.17	19
accuracy			0.10	200
macro avg	0.05	0.50	0.09	200
weighted avg	0.01	0.10	0.02	200

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

	precision	recall	f1-score	support
False	0.91	1.00	0.95	181
True	0.00	0.00	0.00	19
accuracy			0.91	200
macro avg	0.45	0.50	0.48	200
weighted avg	0.82	0.91	0.86	200

```
[[ 0 181]
 [ 0 19]]
```

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

```
The number of records in the training dataset is 700
The number of records in the test dataset is 300
The training dataset has 655 records for the majority class and 45 records for the minority class.
```

	precision	recall	f1-score	support
False	0.00	0.00	0.00	268
True	0.11	1.00	0.19	32
accuracy			0.11	300
macro avg	0.05	0.50	0.10	300
weighted avg	0.01	0.11	0.02	300

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

	precision	recall	f1-score	support
False	0.89	1.00	0.94	268
True	0.00	0.00	0.00	32
accuracy			0.89	300
macro avg	0.45	0.50	0.47	300
weighted avg	0.80	0.89	0.84	300

```
[[ 0 268]
 [ 0 32]]
```

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

```

The number of records in the training dataset is 500
The number of records in the test dataset is 500
The training dataset has 472 records for the majority class and 28 records for the minority class.
      precision    recall   f1-score   support
False        1.00     0.00     0.00    451
True        0.10     1.00     0.18     49

accuracy          0.10    500
macro avg       0.55     0.50     0.09    500
weighted avg    0.91     0.10     0.02    500

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.90     1.00     0.95    451
True        0.00     0.00     0.00     49

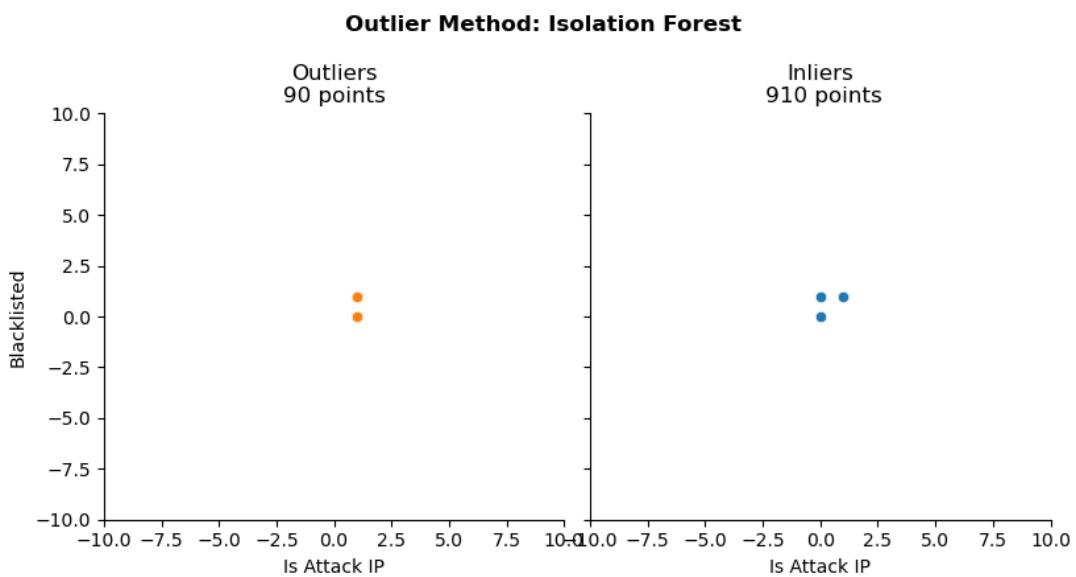
accuracy          0.90    500
macro avg       0.45     0.50     0.47    500
weighted avg    0.81     0.90     0.86    500

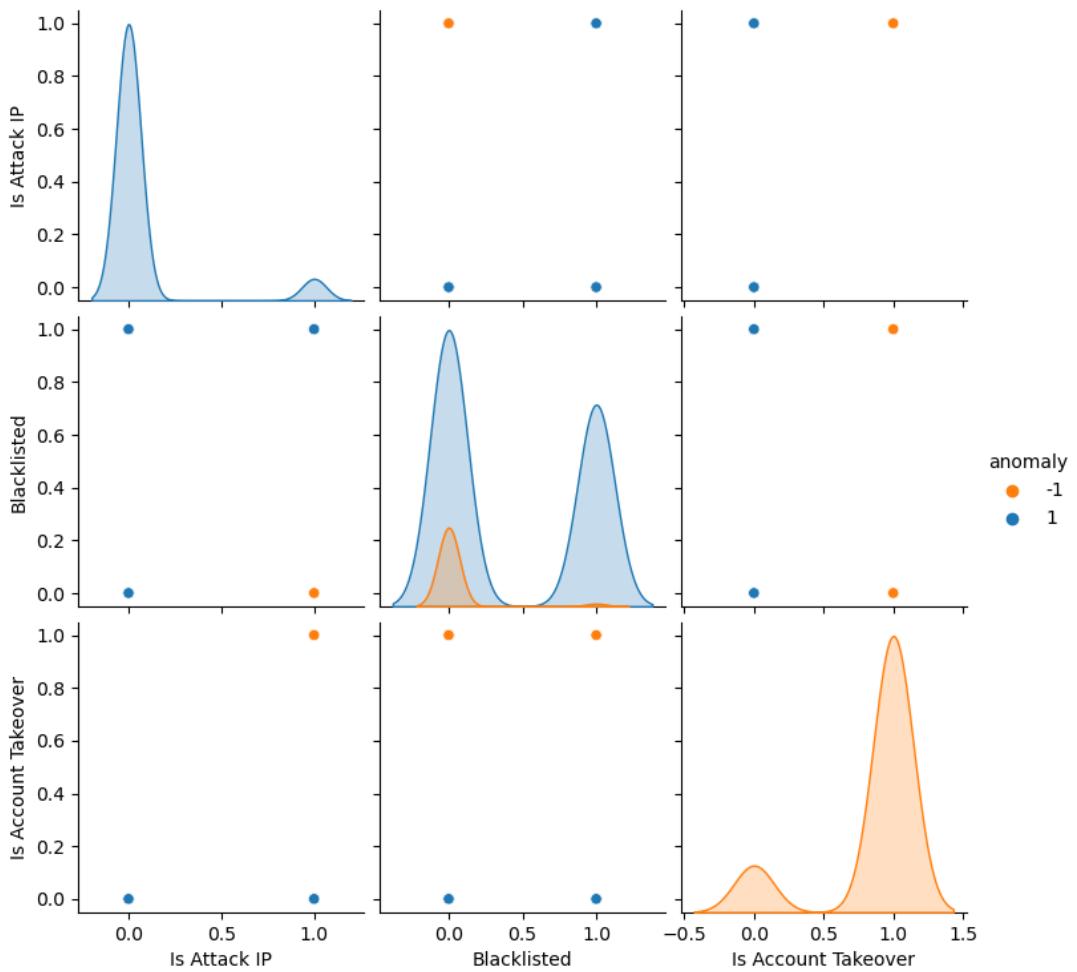
[[ 1 450]
 [ 0 49]]

```

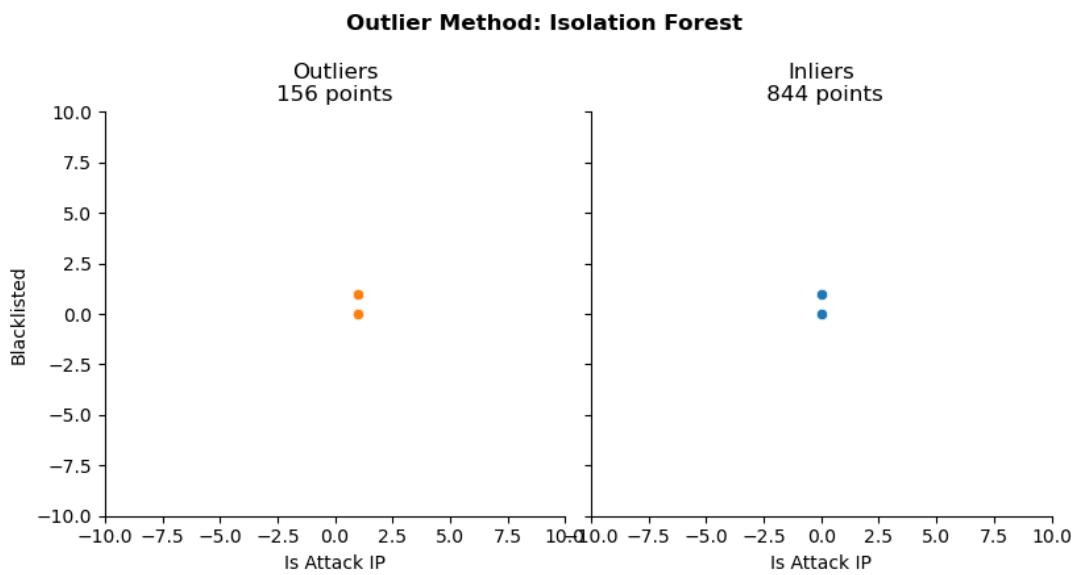
3C. Isolation Forest

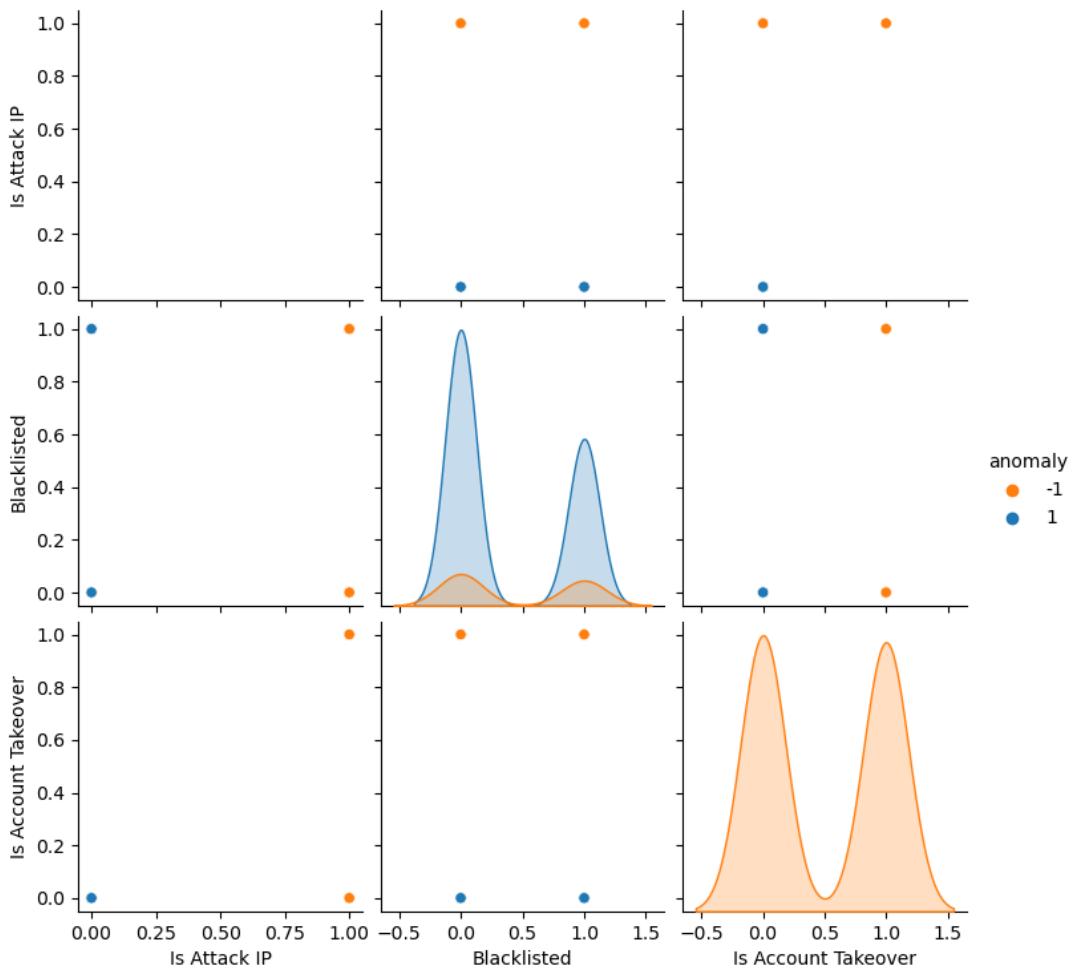
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



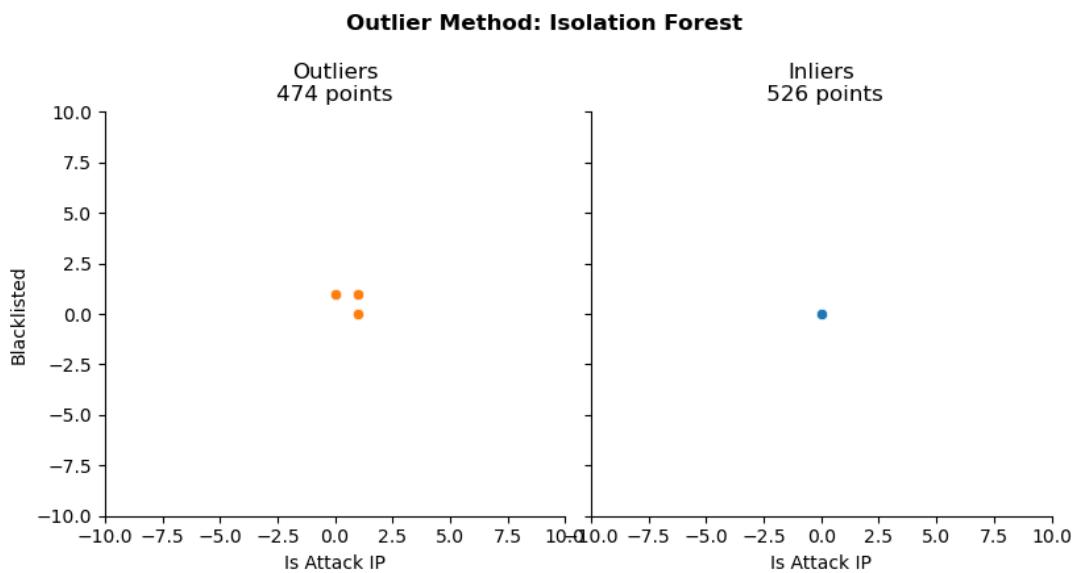


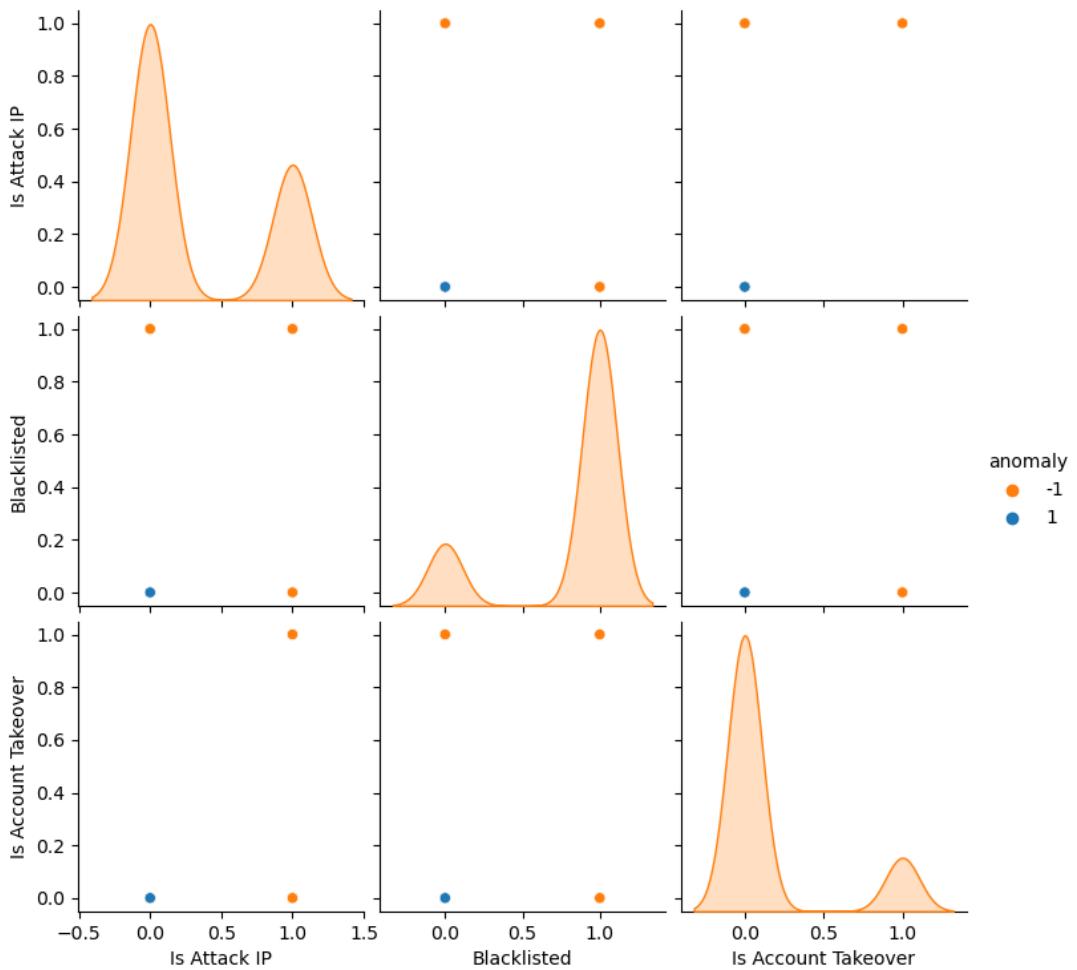
- contamination value == 0.3



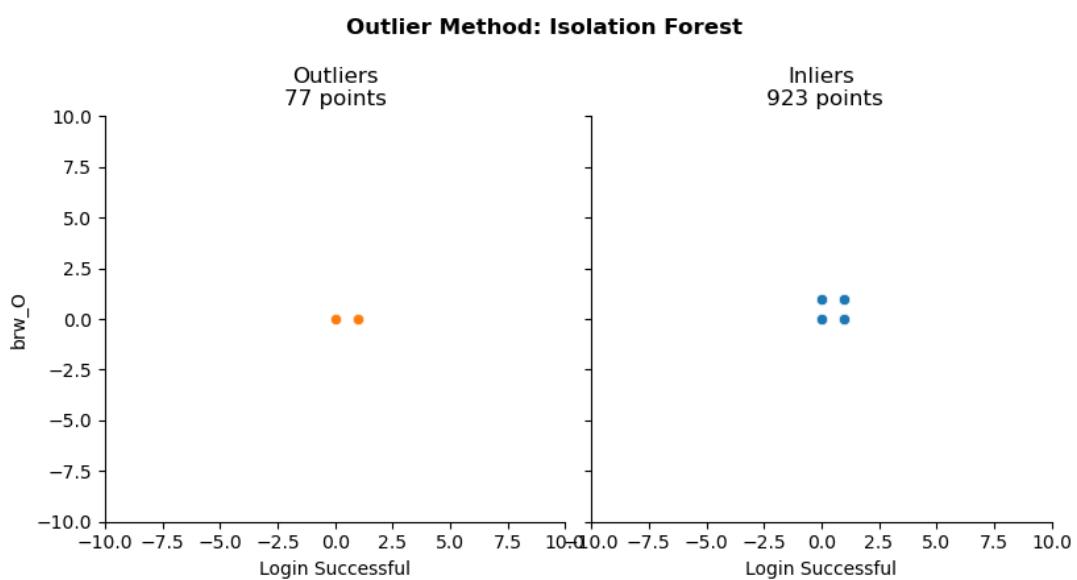


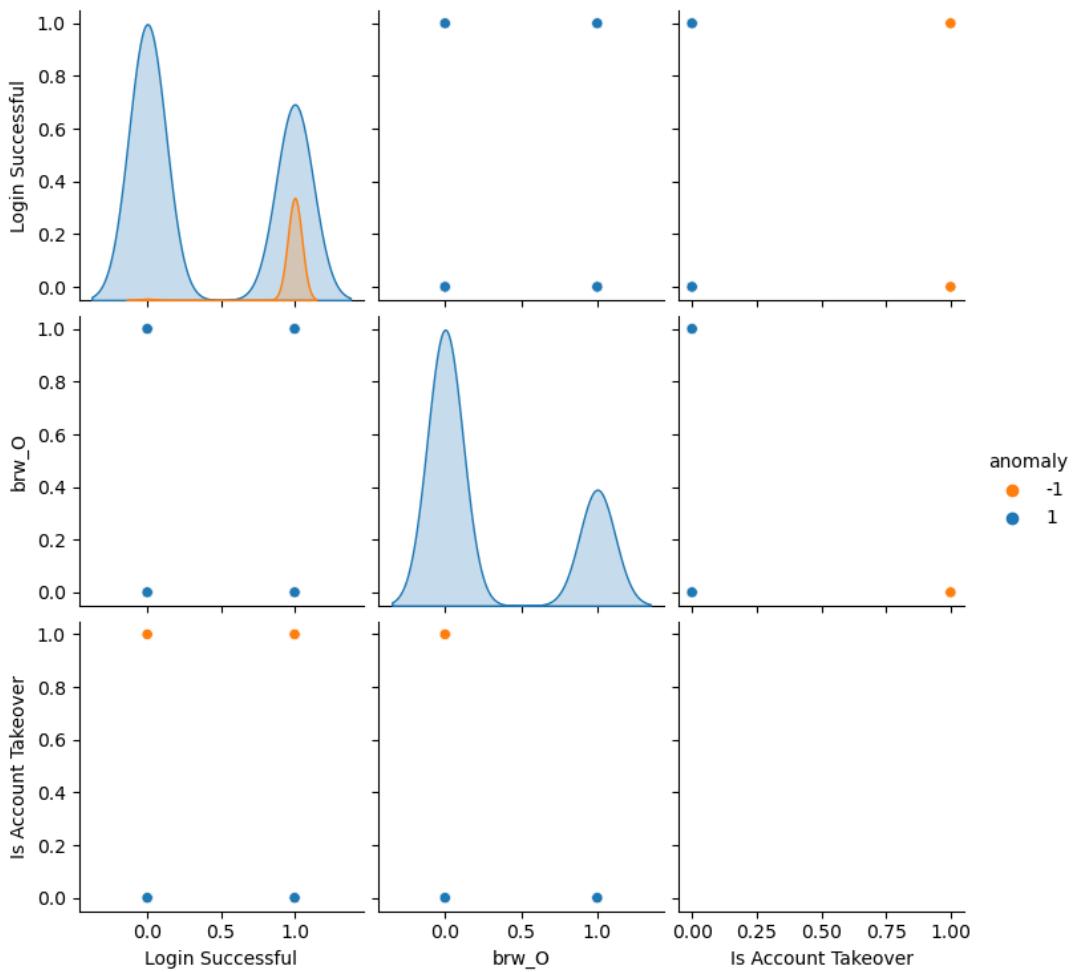
- contamination value == 0.5





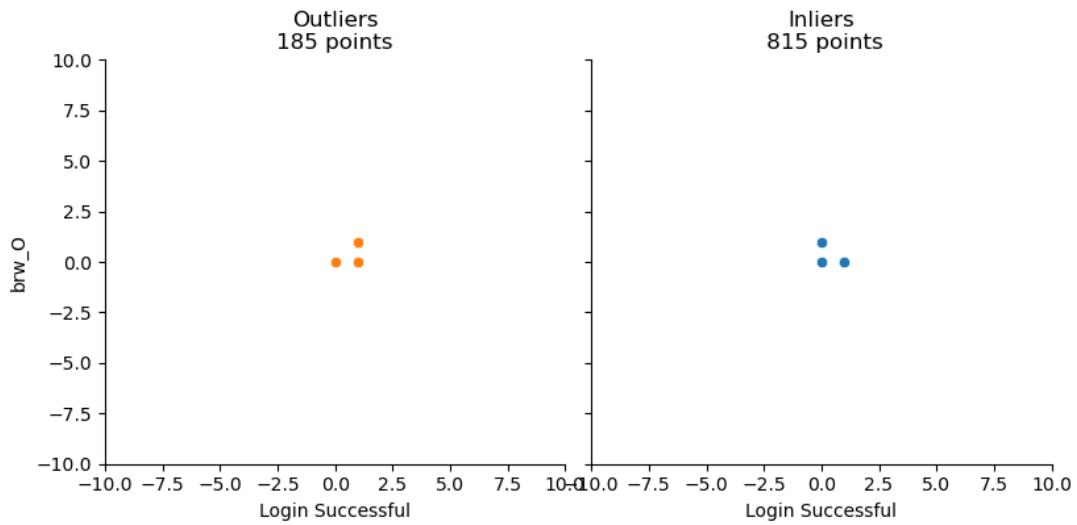
- anomaly_inputs = ['Login Successful', 'Browser Type (brw_O)', 'Is Account Takeover']
 - contamination value == 0.1

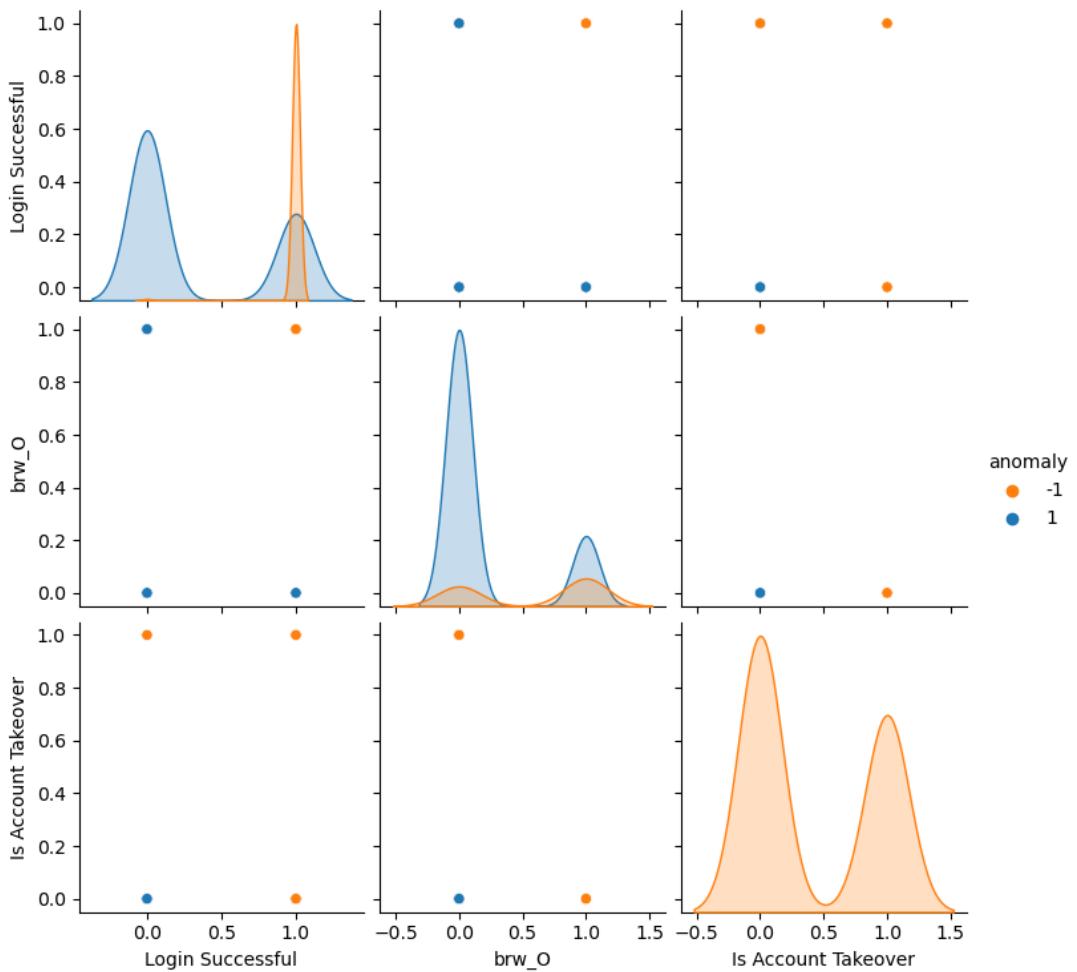




- contamination value == 0.3

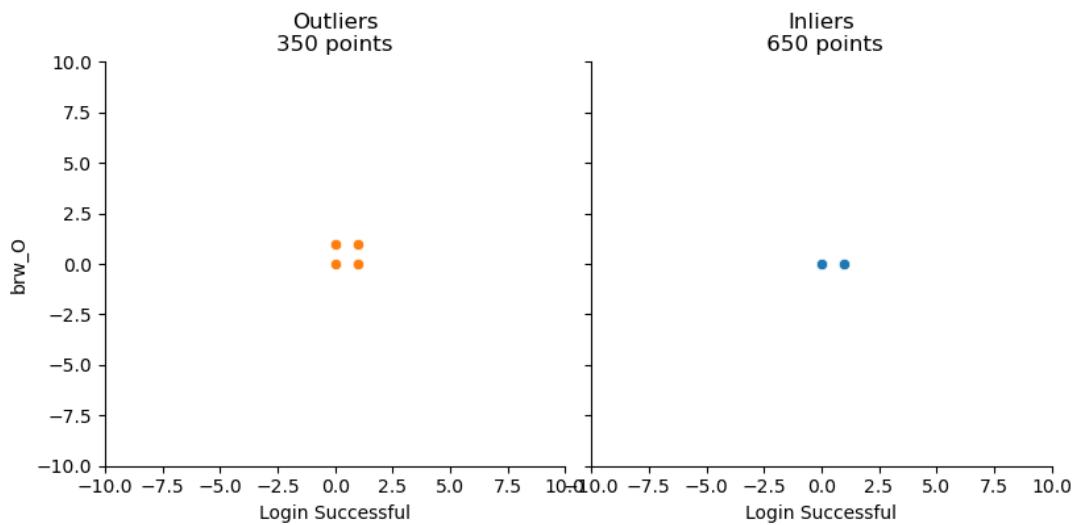
Outlier Method: Isolation Forest

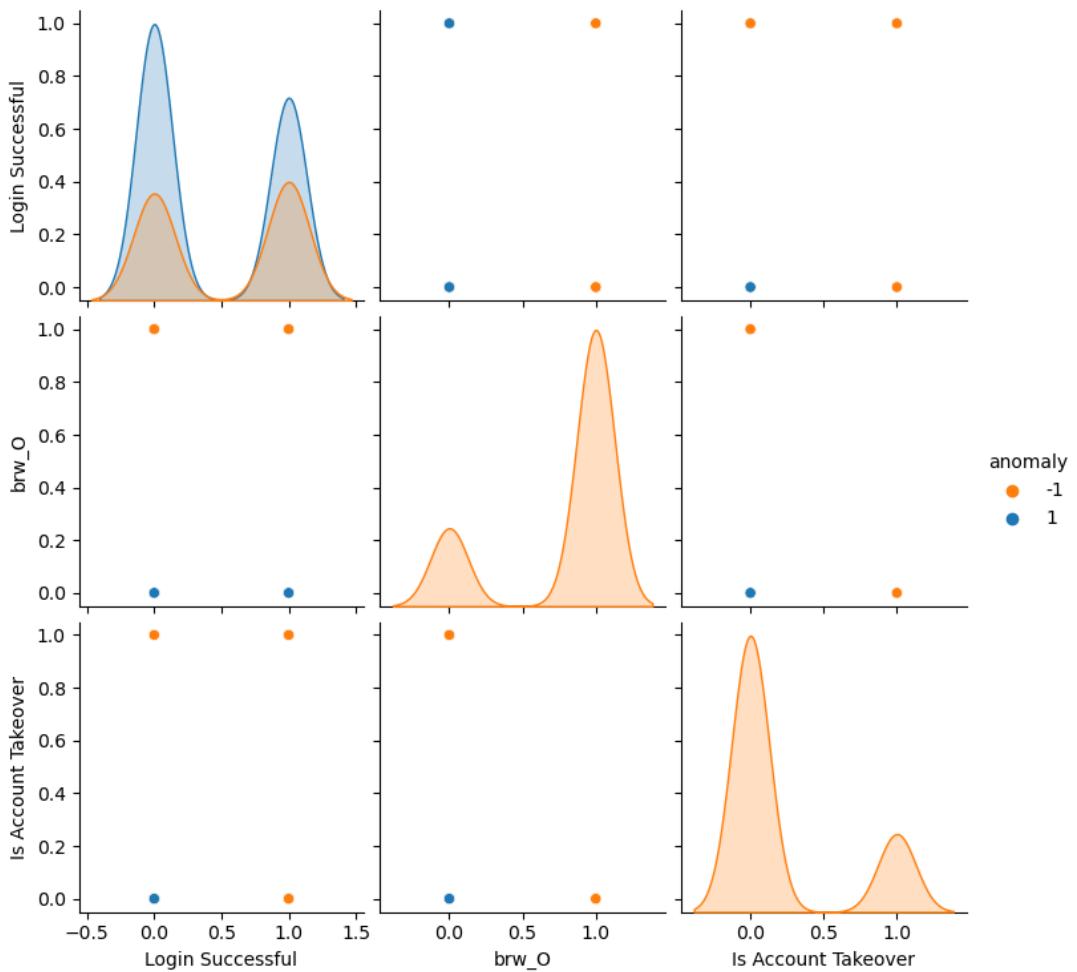




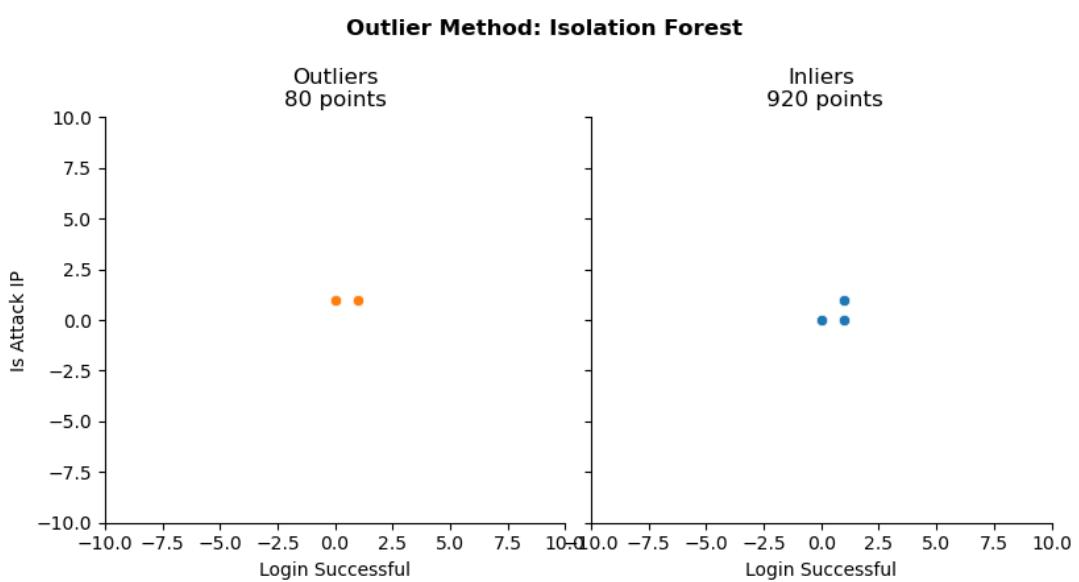
- contamination value == 0.5

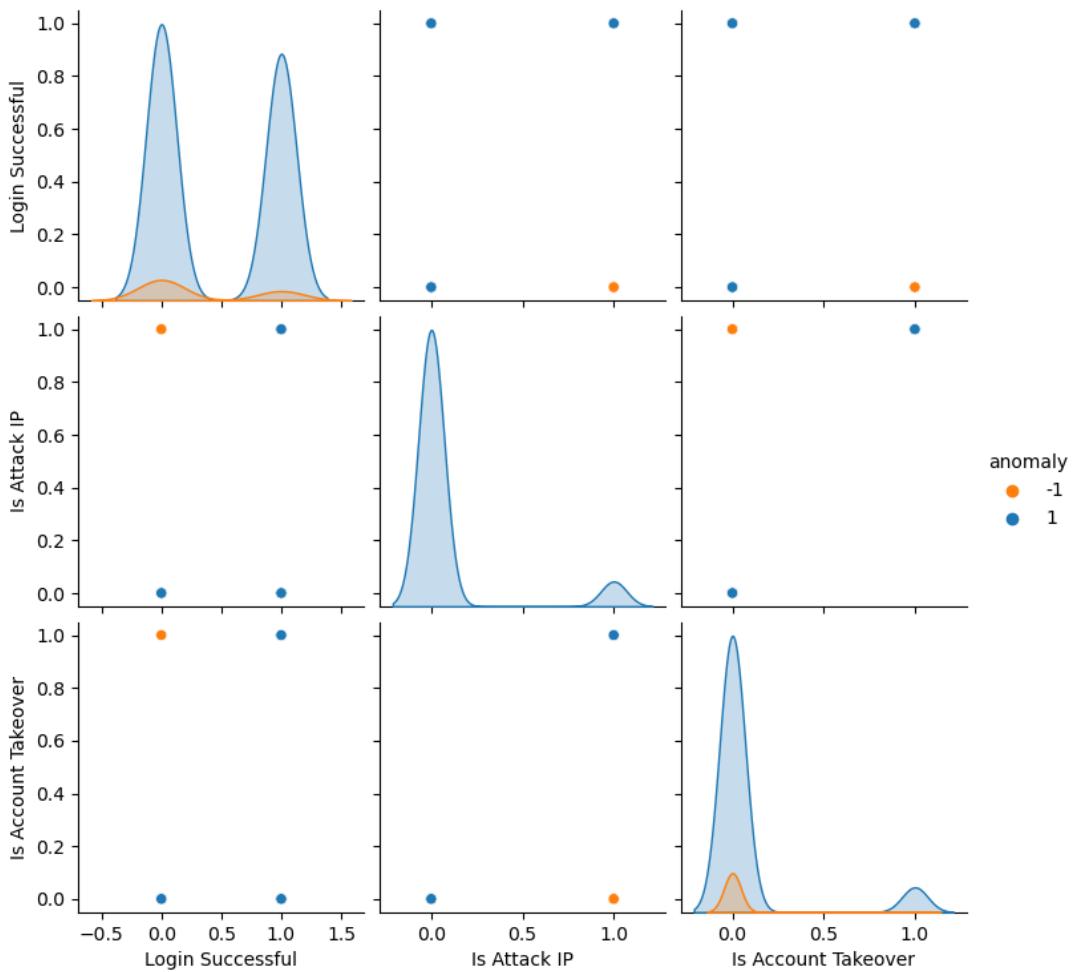
Outlier Method: Isolation Forest





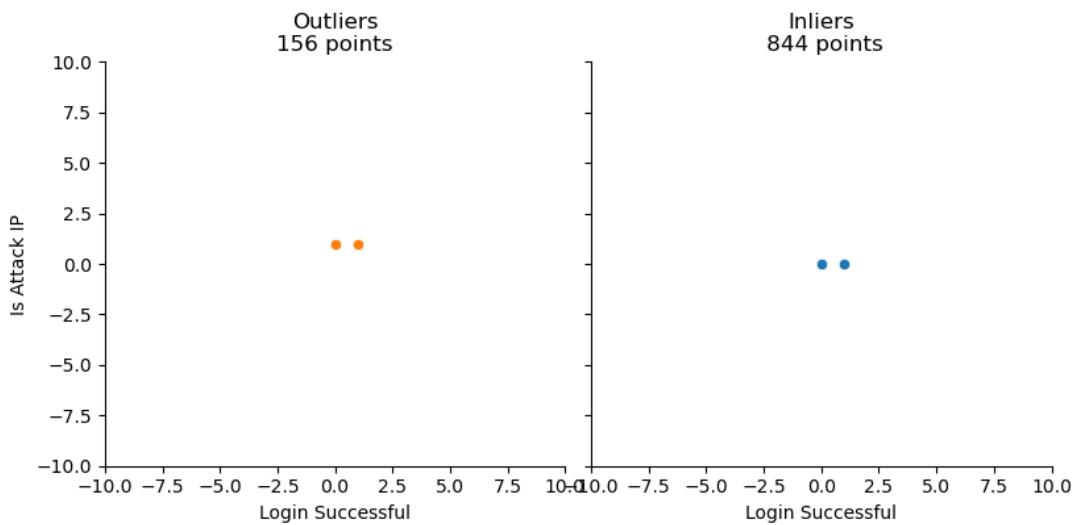
- `anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']`
 - contamination value == 0.1

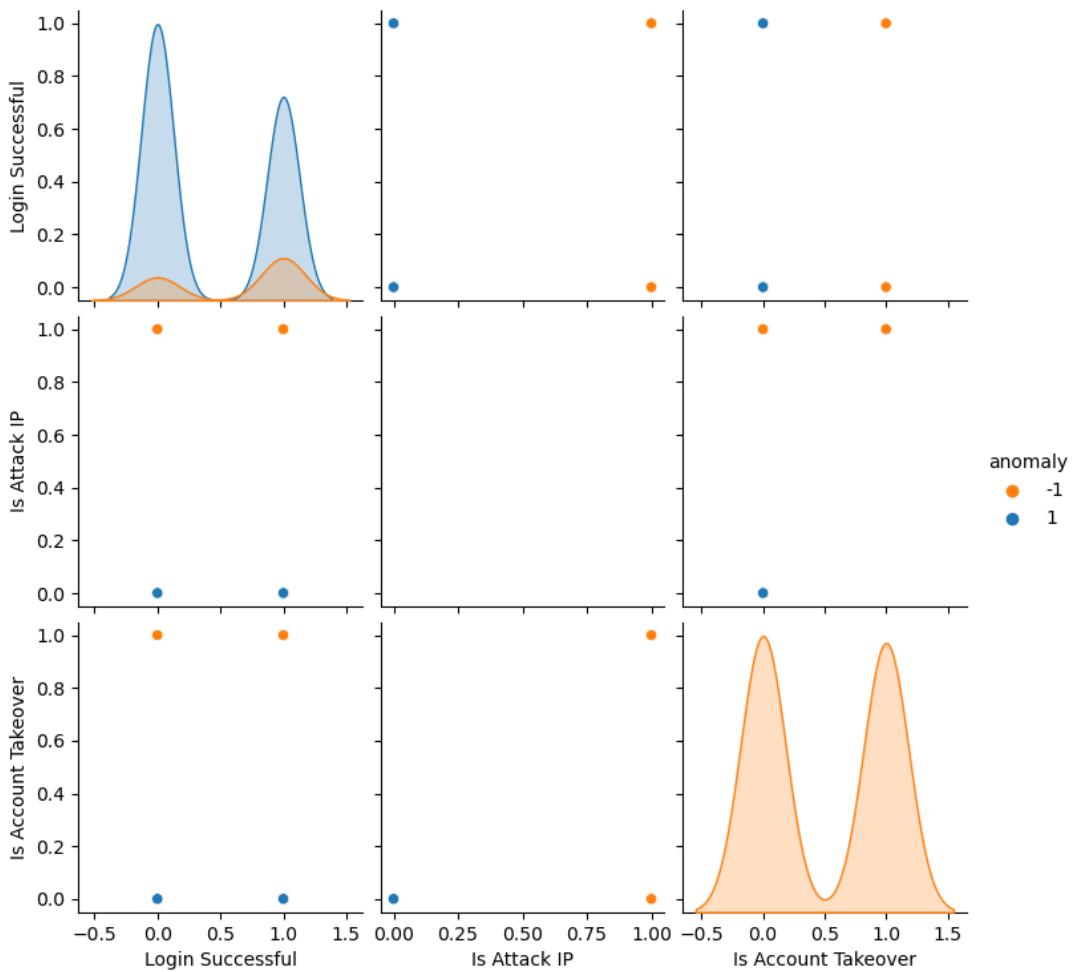




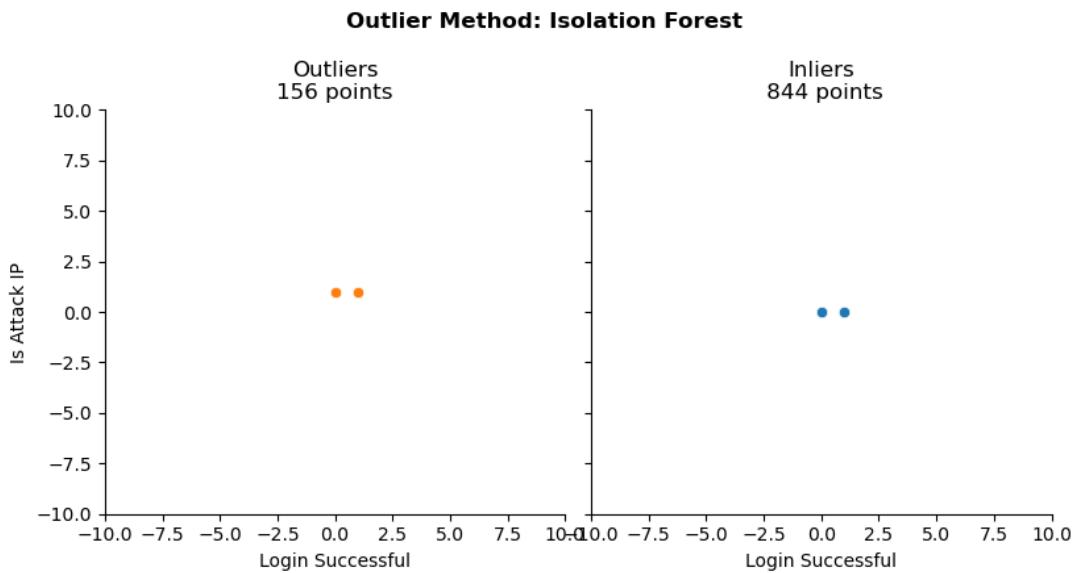
- contamination value == 0.3

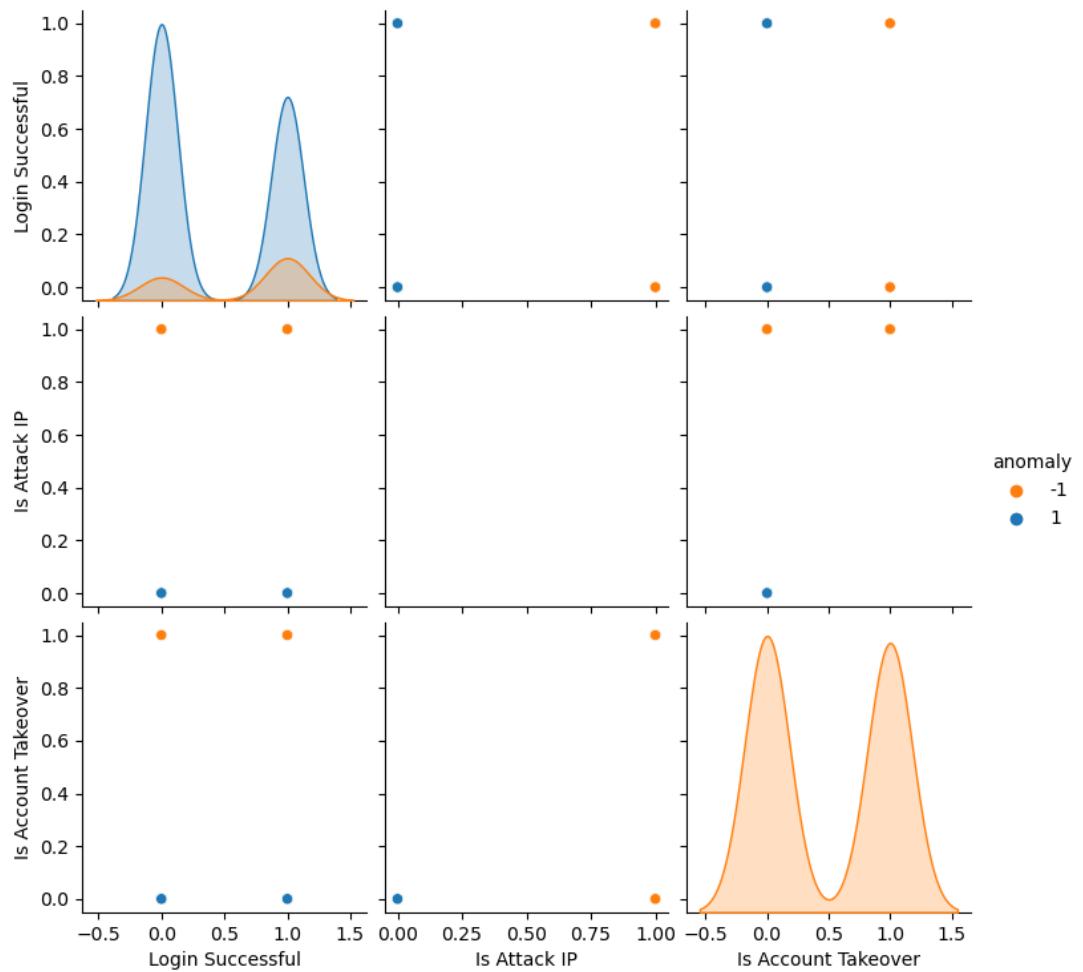
Outlier Method: Isolation Forest



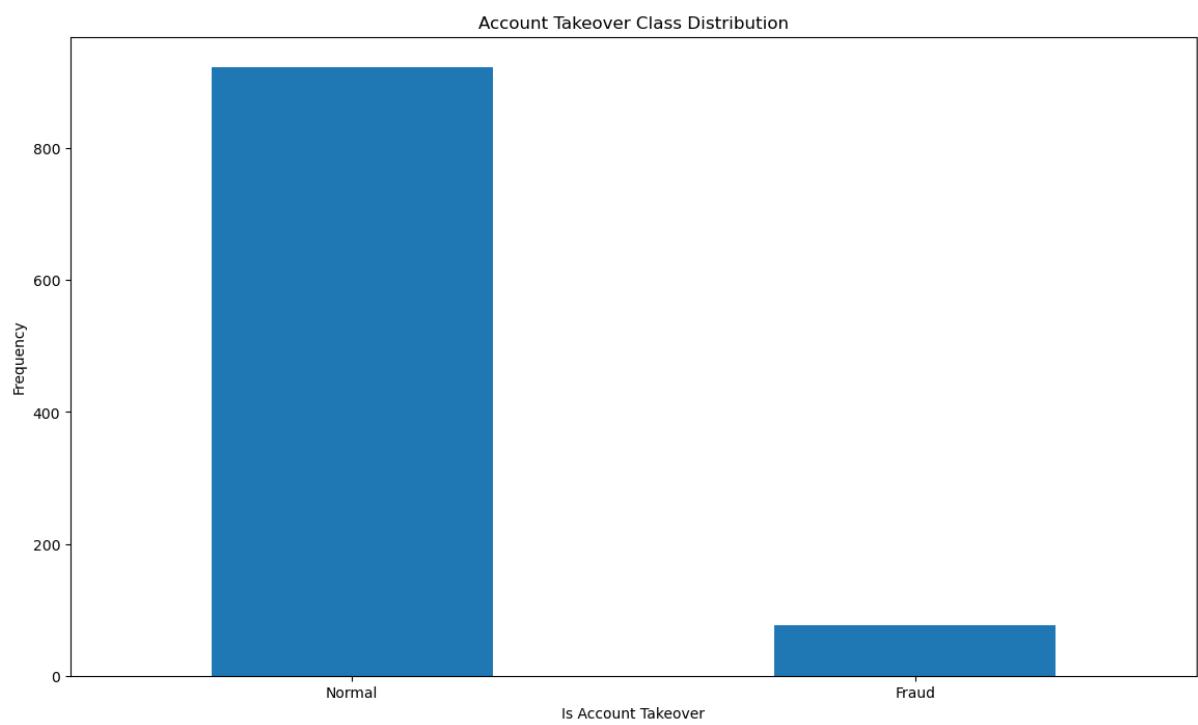


- contamination value == 0.5





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



Isolation Forest: 159

Accuracy Score :

0.841

Classification Report :

	precision	recall	f1-score	support
False	0.92	0.90	0.91	923
True	0.09	0.12	0.10	77
accuracy			0.84	1000
macro avg	0.51	0.51	0.51	1000
weighted avg	0.86	0.84	0.85	1000

Local Outlier Factor: 147

Accuracy Score :

0.853

Classification Report :

	precision	recall	f1-score	support
False	0.93	0.91	0.92	923
True	0.15	0.19	0.17	77
accuracy			0.85	1000
macro avg	0.54	0.55	0.54	1000
weighted avg	0.87	0.85	0.86	1000

Support Vector Machine: 603

Accuracy Score :

0.397

Classification Report :

	precision	recall	f1-score	support
False	0.88	0.40	0.55	923
True	0.04	0.32	0.08	77
accuracy			0.40	1000
macro avg	0.46	0.36	0.31	1000
weighted avg	0.81	0.40	0.52	1000

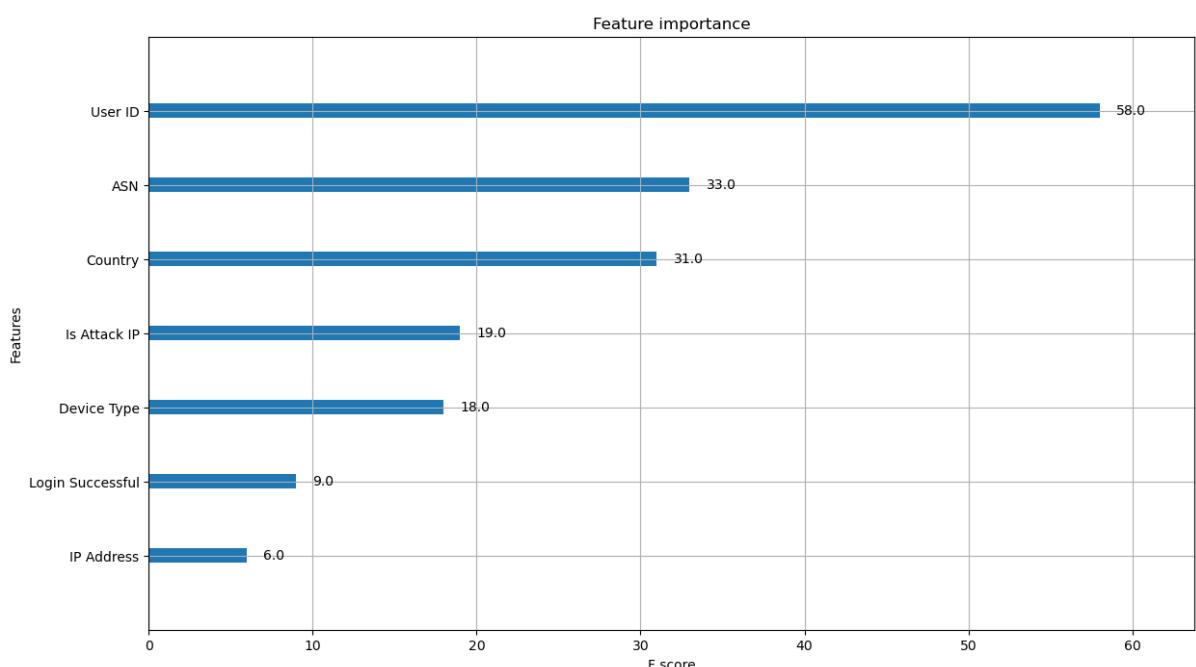
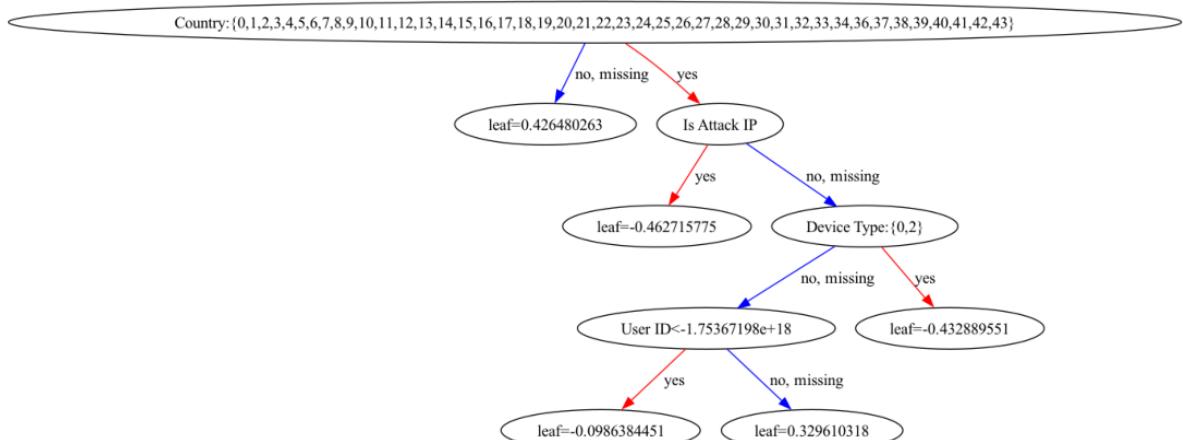
IV. 141/1000 → 14.1%

1. XGB Experimental

```

-----
0 Login Timestamp      1000 non-null  category
1 User ID             1000 non-null  int64
2 IP Address          1000 non-null  category
3 Country              1000 non-null  category
4 ASN                 1000 non-null  int64
5 Device Type          1000 non-null  category
6 Login Successful     1000 non-null  bool
7 Is Attack IP         1000 non-null  bool
8 Blacklisted           1000 non-null  bool
9 Browser Type          1000 non-null  category
10 Is Account Takeover 1000 non-null bool
dtypes: bool(4), category(5), int64(2)
memory usage: 107.6 KB
Feature importances:
[0.          0.01772542 0.03462604 0.70550615 0.00749496 0.09745706
 0.07308619 0.0641042  0.          0.          ]

```



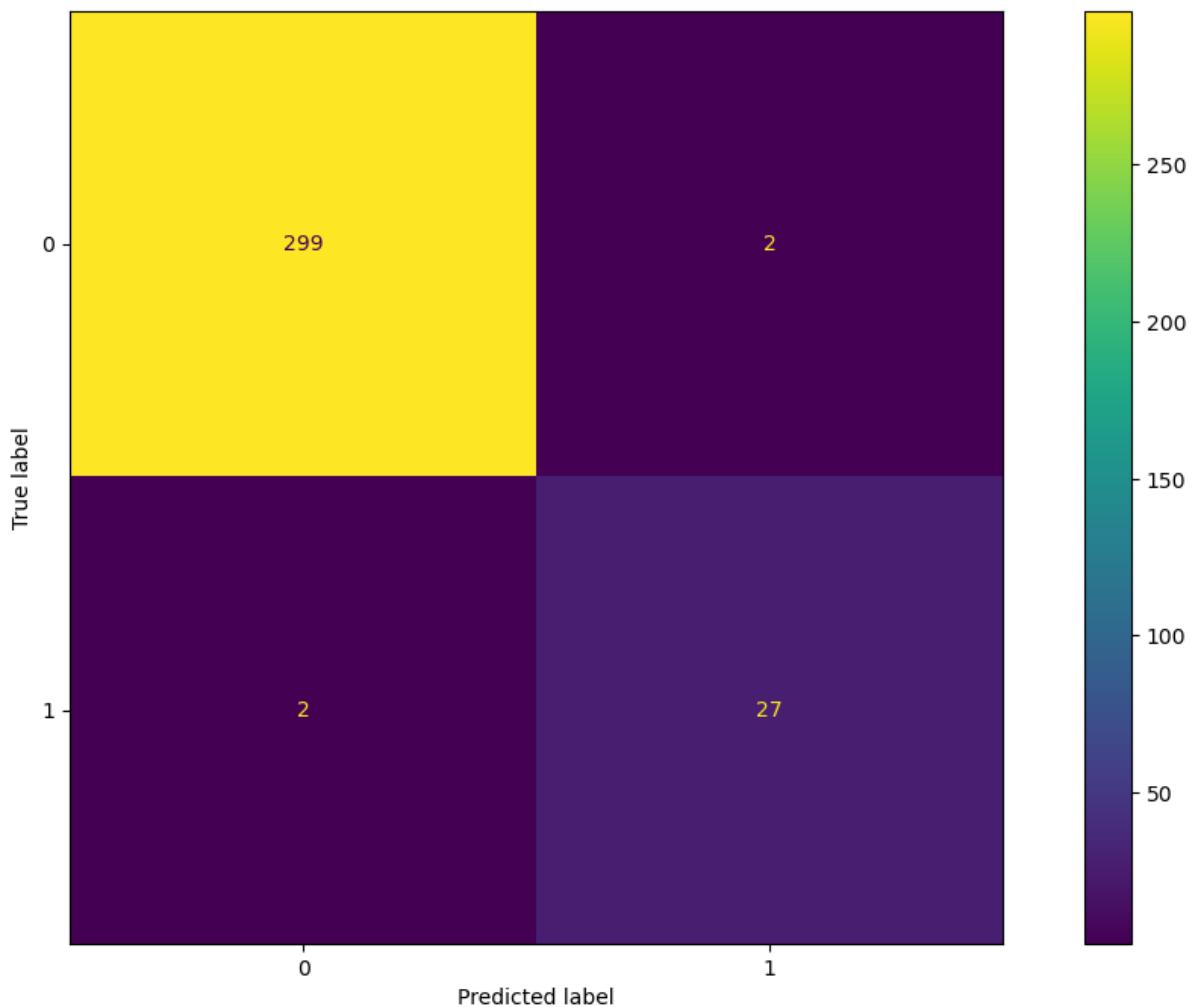
2. Label encoding (cat.codes)

2A. XGBoost

memory usage: 25.5 KB

Accuracy: 98.79%

	precision	recall	f1-score	support
False	0.99	0.99	0.99	301
True	0.93	0.93	0.93	29
accuracy			0.99	330
macro avg	0.96	0.96	0.96	330
weighted avg	0.99	0.99	0.99	330



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 800
The number of records in the test dataset is 200
The training dataset has 742 records for the majority class and 58 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00      181
  True        0.10     1.00     0.17      19

accuracy                           0.10      200
macro avg       0.05     0.50     0.09      200
weighted avg    0.01     0.10     0.02      200

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.91     1.00     0.95      181
  True        0.00     0.00     0.00      19

accuracy                           0.91      200
macro avg       0.45     0.50     0.48      200
weighted avg    0.82     0.91     0.86      200

[[ 0 181]
 [ 0 19]]
```

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

```
The number of records in the training dataset is 700
The number of records in the test dataset is 300
The training dataset has 655 records for the majority class and 45 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00      268
  True        0.11     1.00     0.19      32

accuracy                           0.11      300
macro avg       0.05     0.50     0.10      300
weighted avg    0.01     0.11     0.02      300

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.89     1.00     0.94      268
  True        0.00     0.00     0.00      32

accuracy                           0.89      300
macro avg       0.45     0.50     0.47      300
weighted avg    0.80     0.89     0.84      300

[[ 0 268]
 [ 0 32]]
```

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

```

The number of records in the training dataset is 500
The number of records in the test dataset is 500
The training dataset has 472 records for the majority class and 28 records for the minority class.
      precision    recall   f1-score   support
False        0.00     0.00     0.00    451
True        0.10     1.00     0.18     49

accuracy                           0.10      500
macro avg       0.05     0.50     0.09      500
weighted avg    0.01     0.10     0.02      500

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.90     1.00     0.95    451
True        0.00     0.00     0.00     49

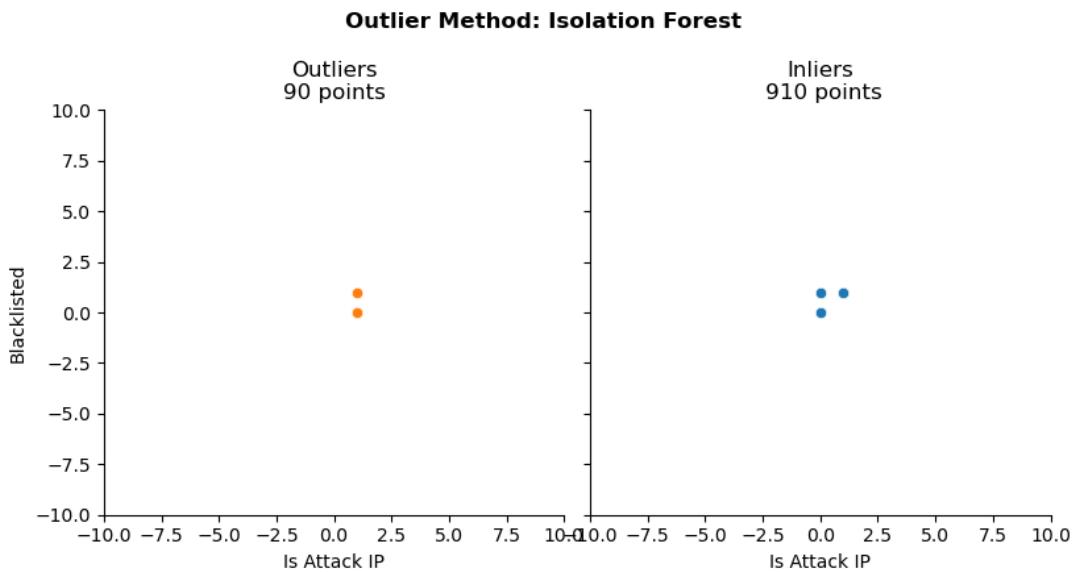
accuracy                           0.90      500
macro avg       0.45     0.50     0.47      500
weighted avg    0.81     0.90     0.86      500

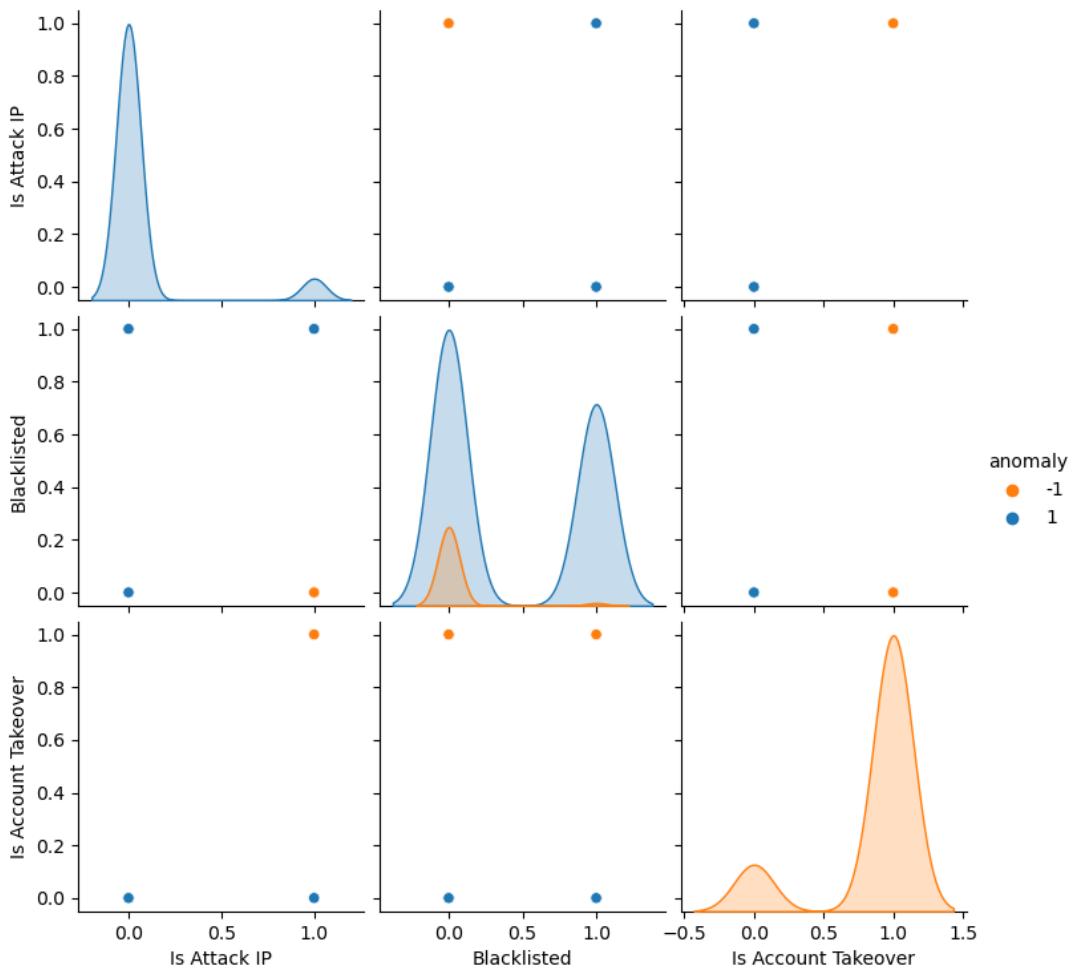
[[ 0 451]
 [ 0 49]]

```

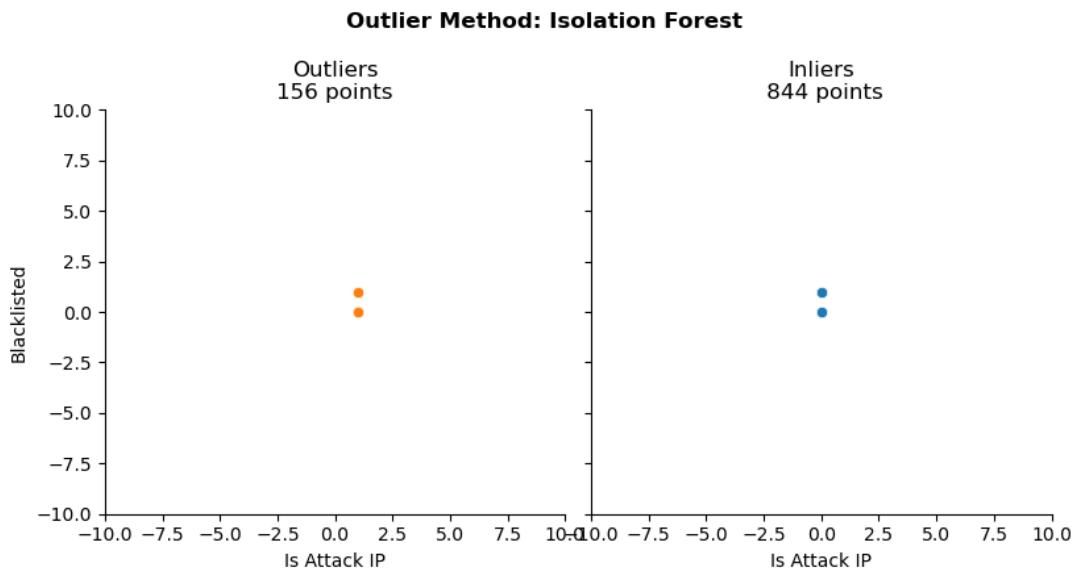
2C. Isolation Forest

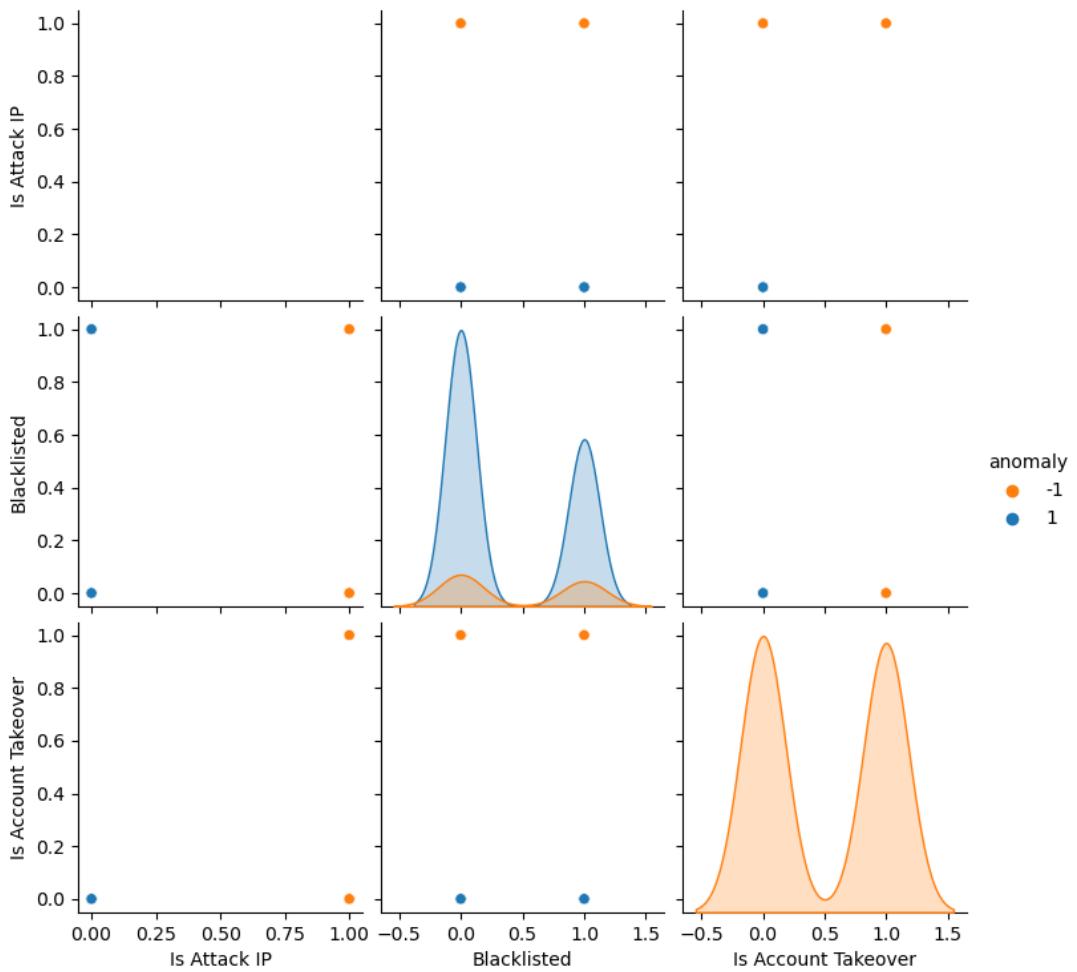
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



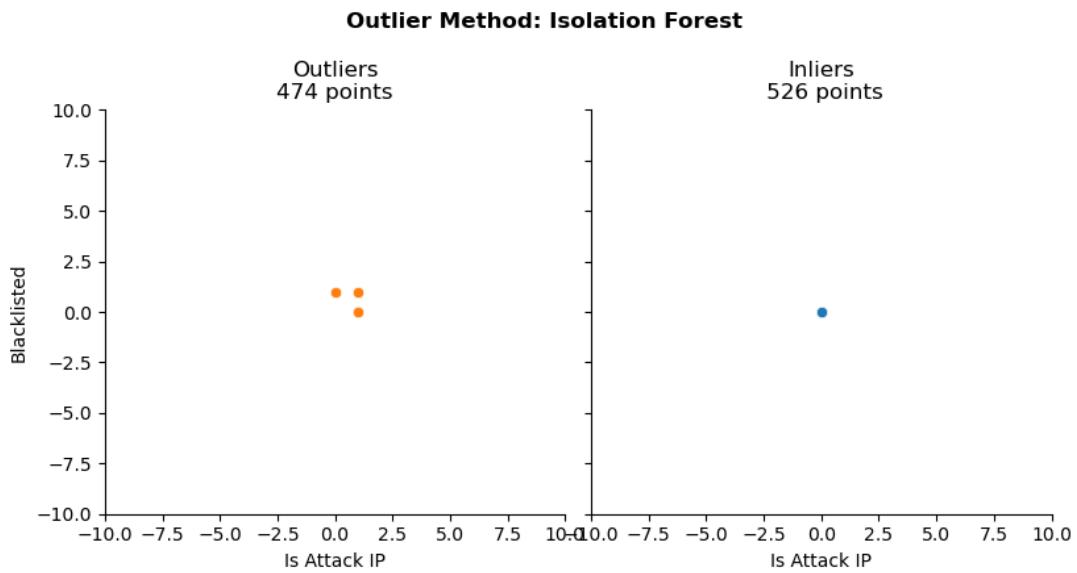


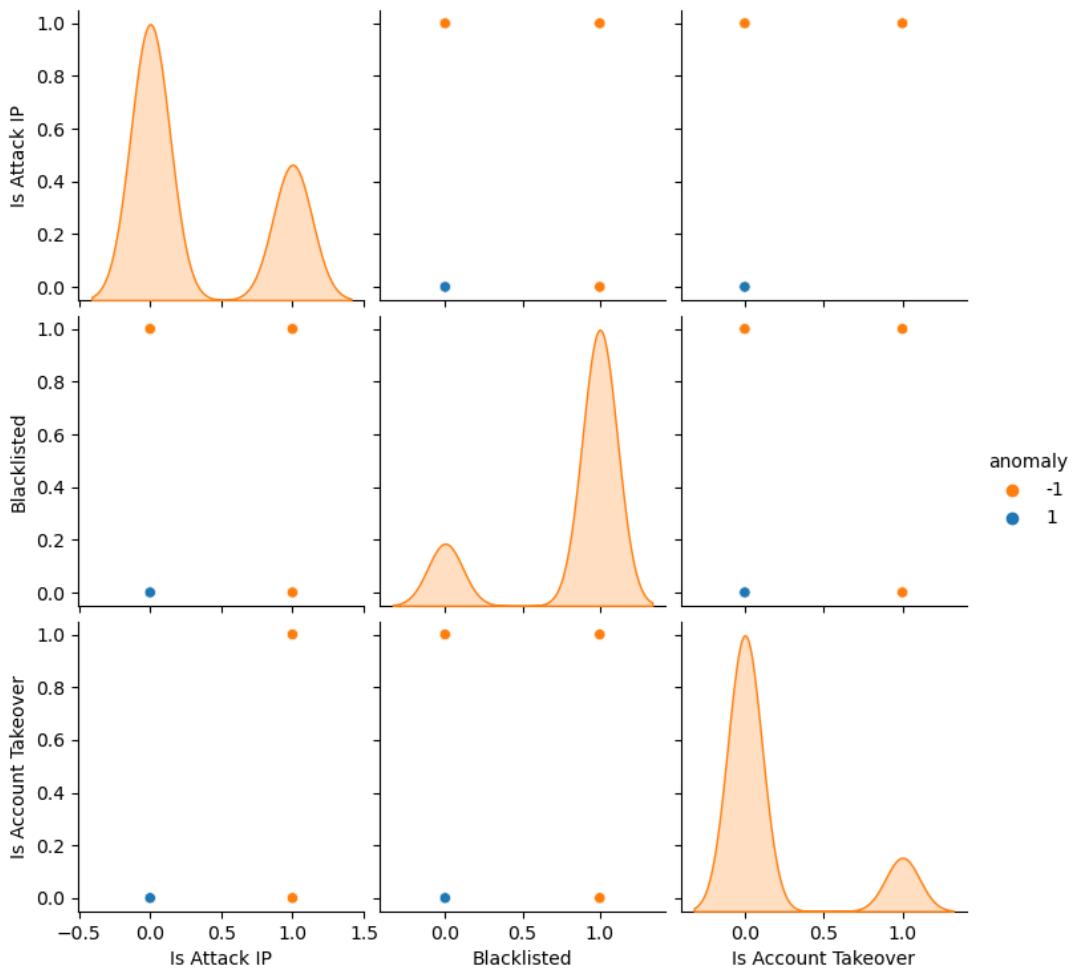
- increasing contamination value to 0.3



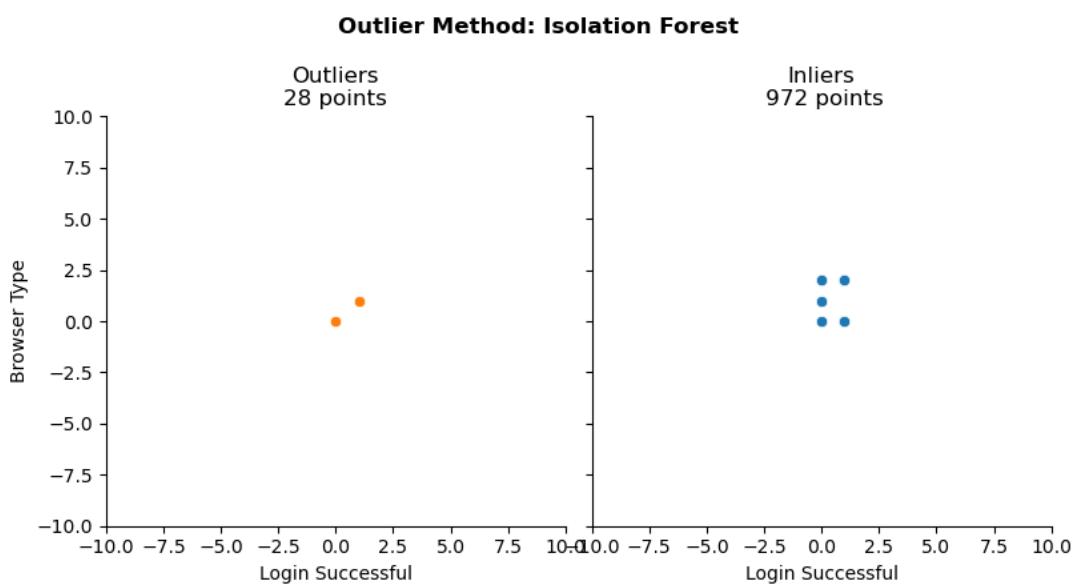


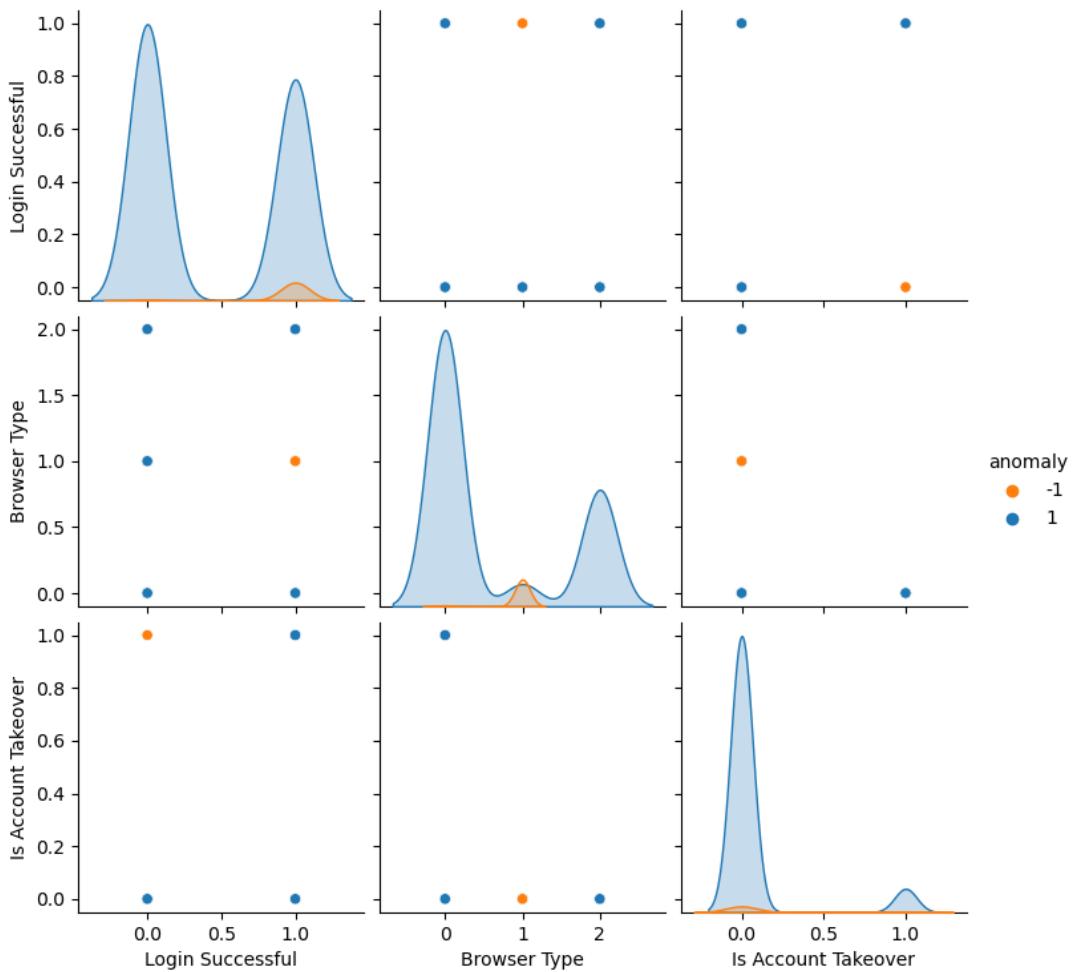
- contamination value == 0.5



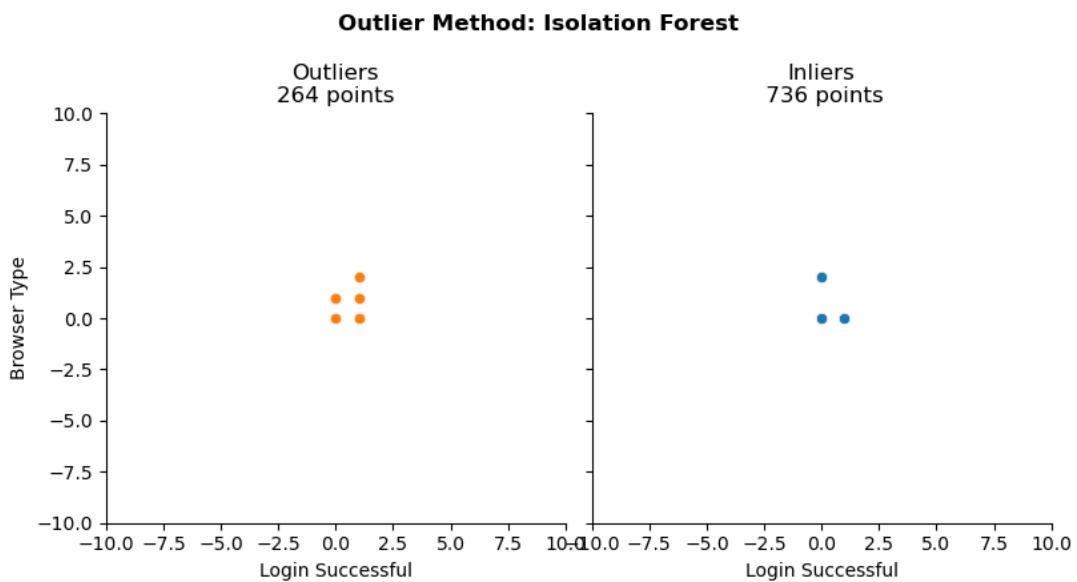


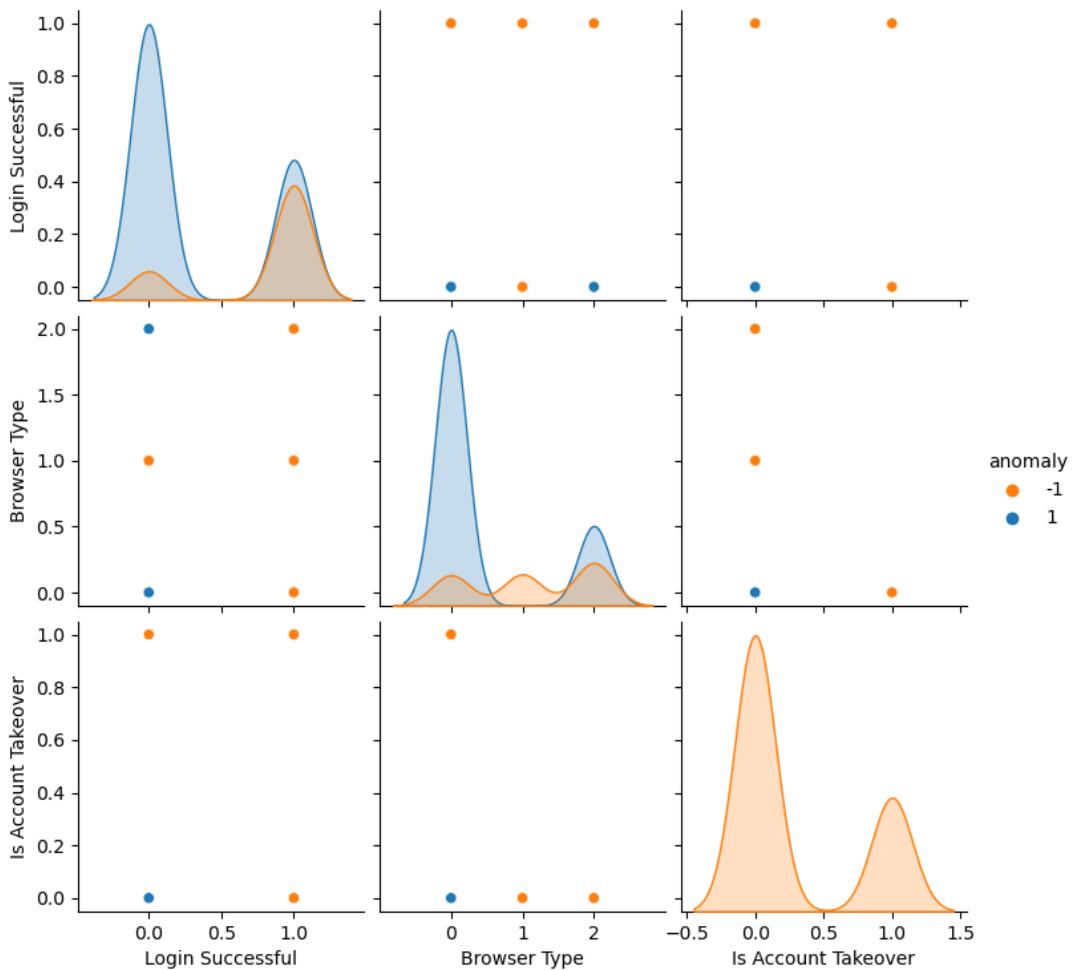
- anomaly_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']
 - contamination value == 0.1



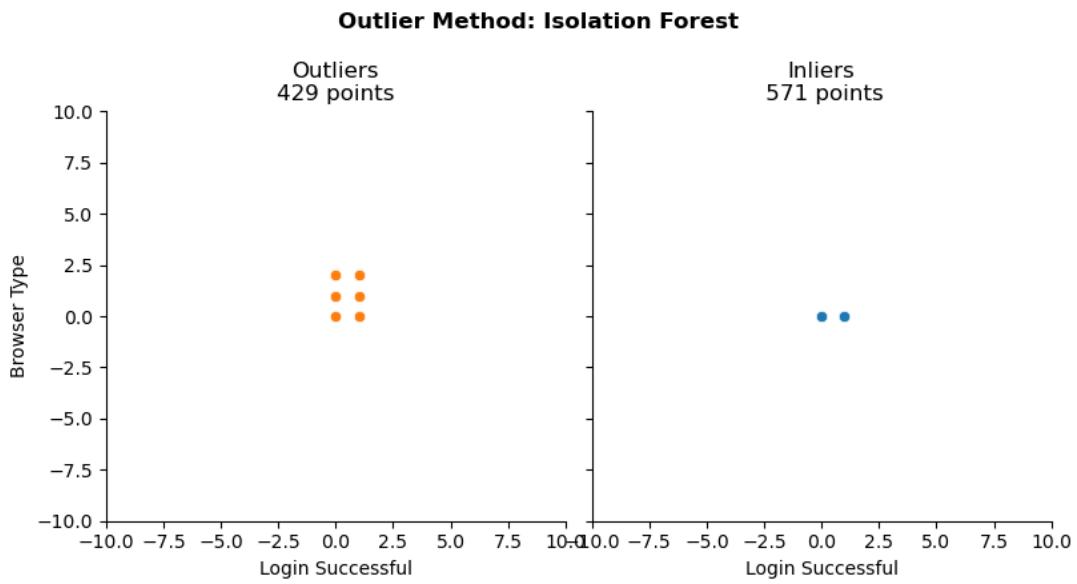


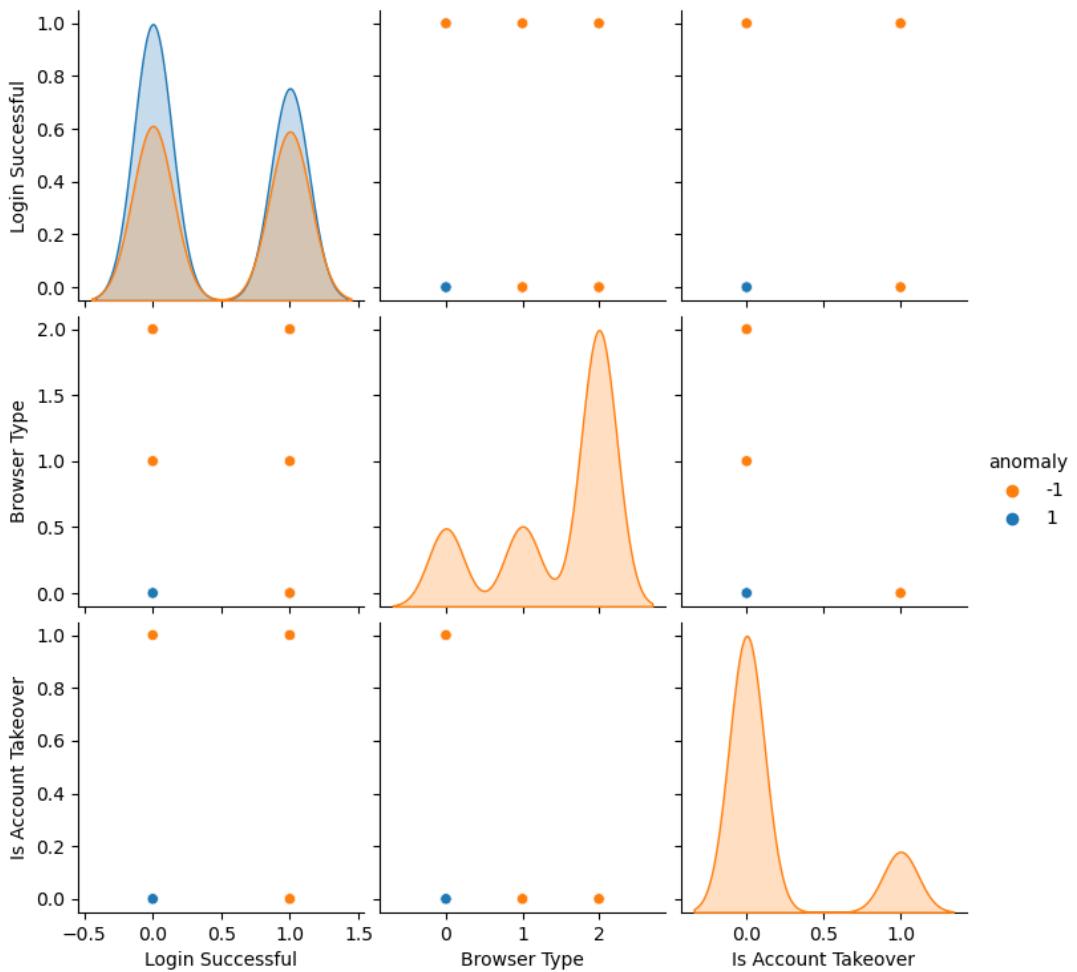
- contamination value == 0.3



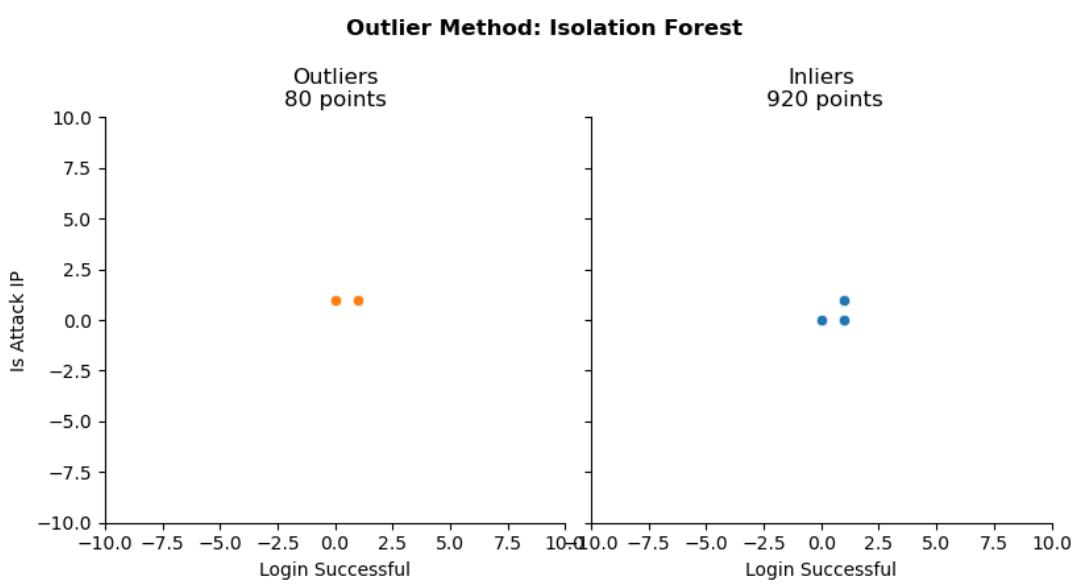


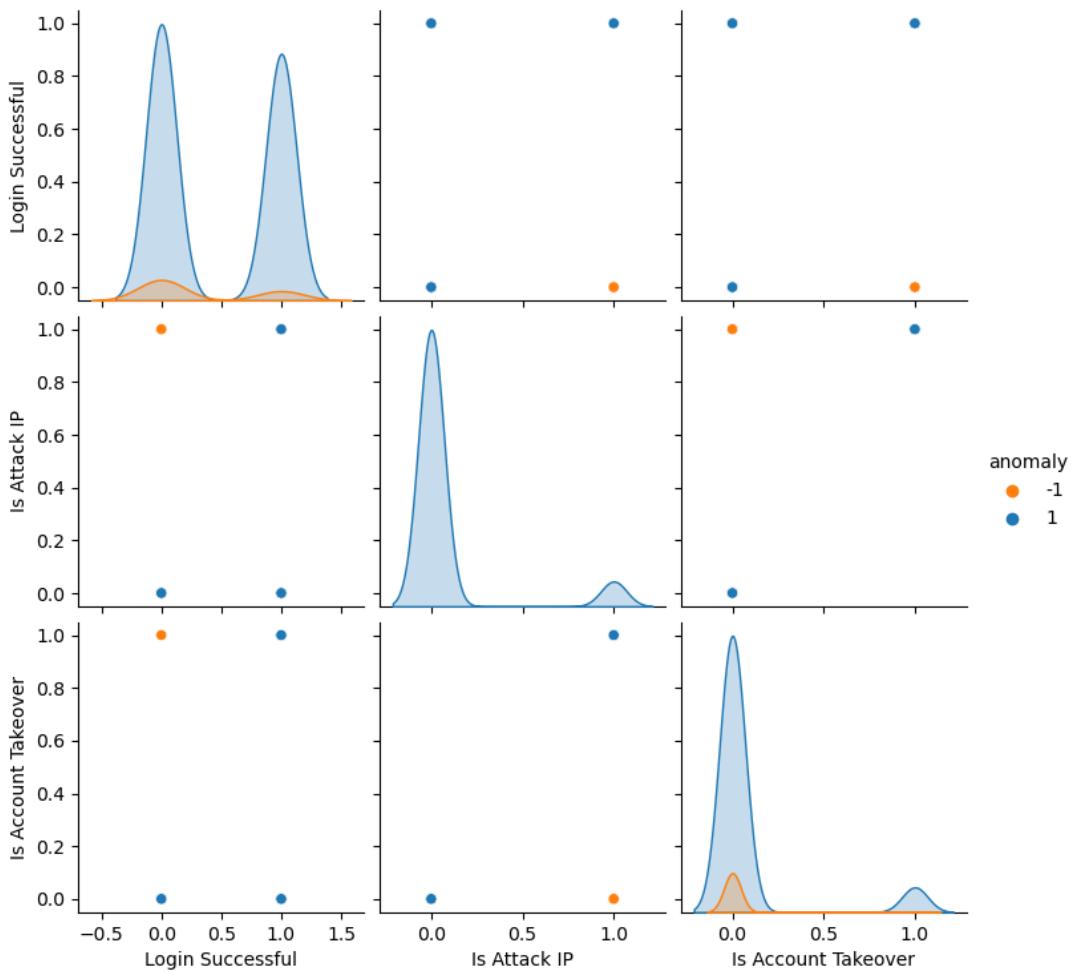
- contamination value == 0.5





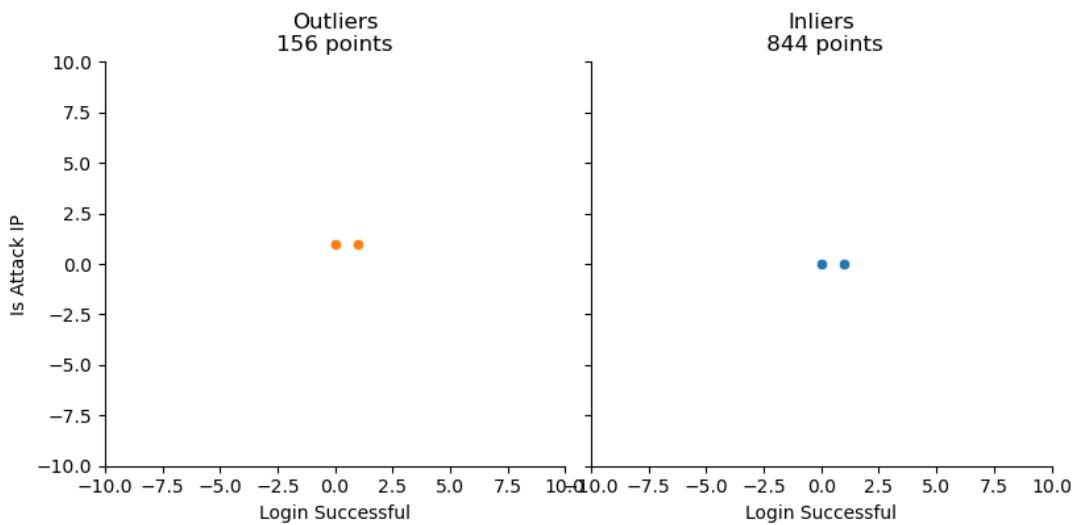
- `anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']`
 - contamination value == 0.1

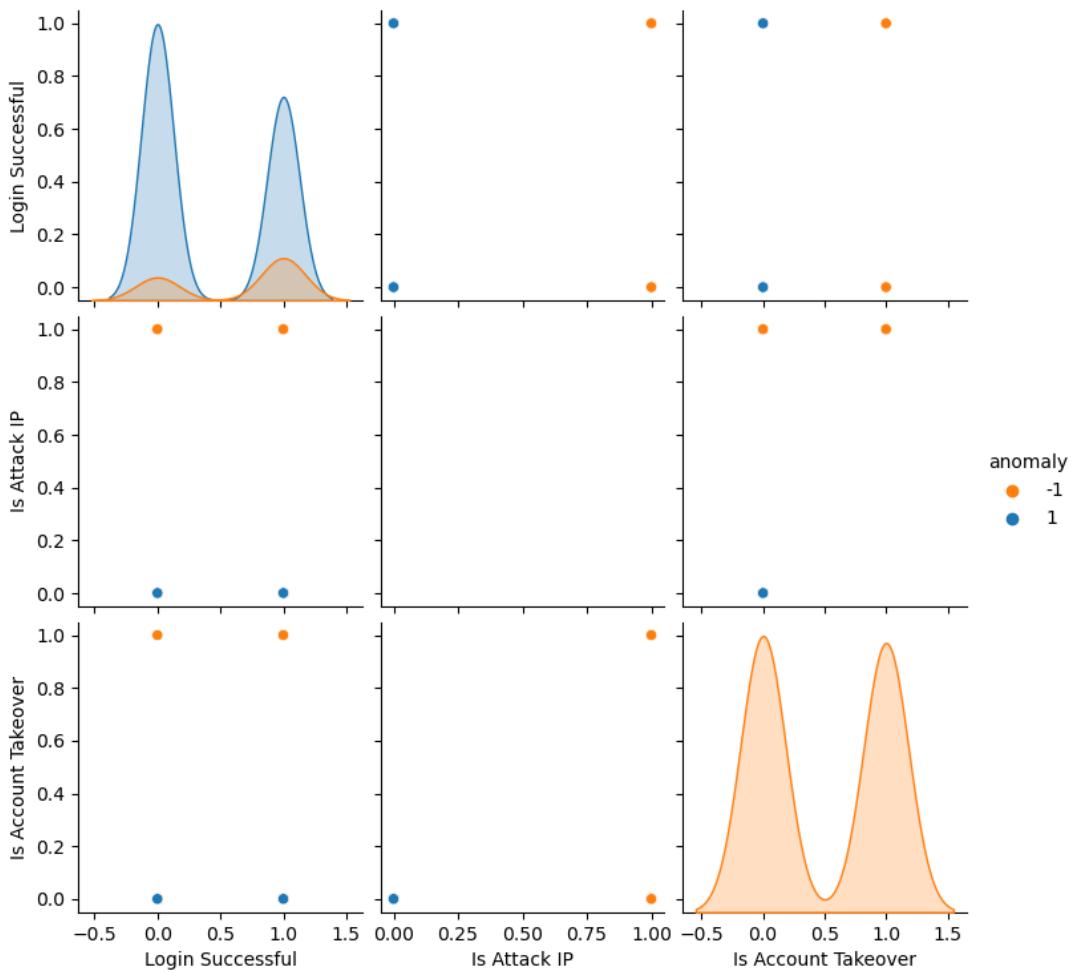




- contamination value == 0.3

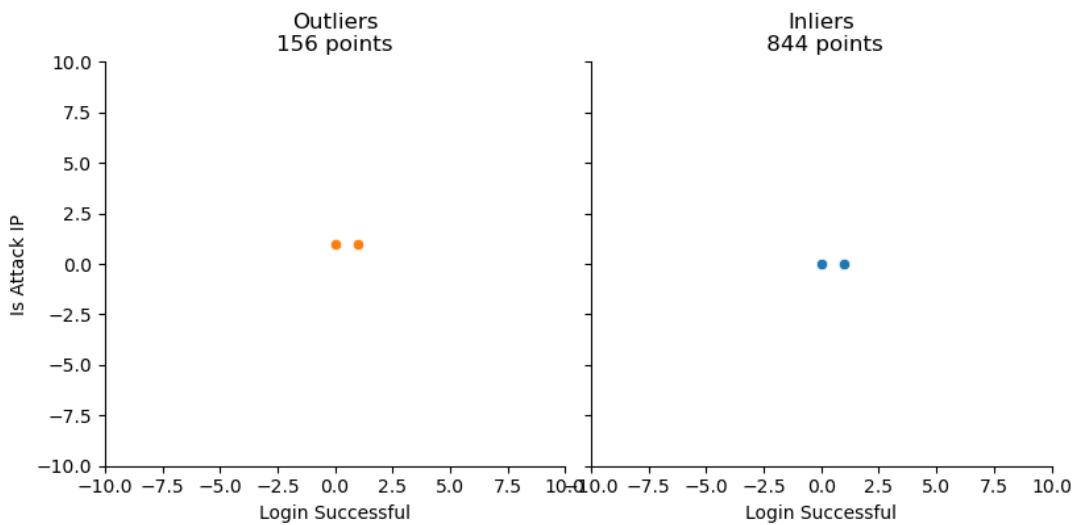
Outlier Method: Isolation Forest

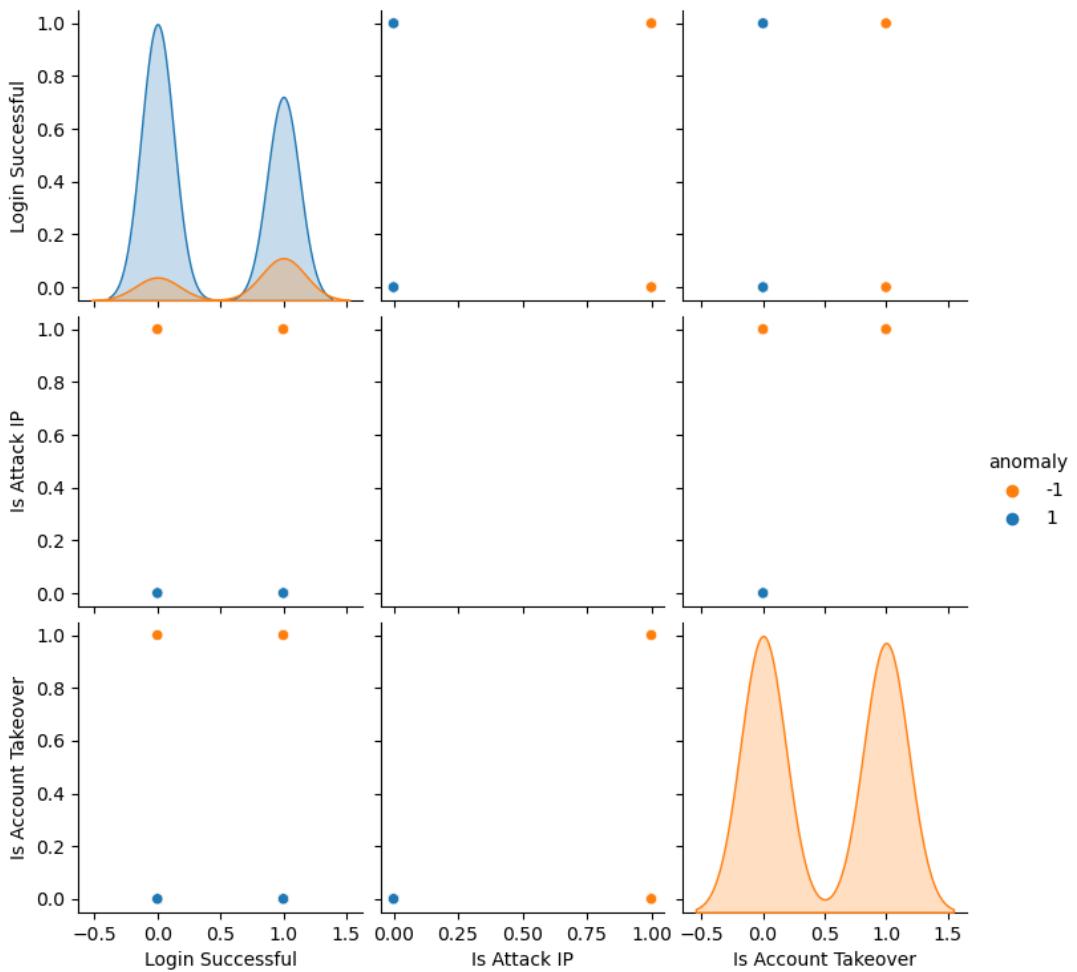




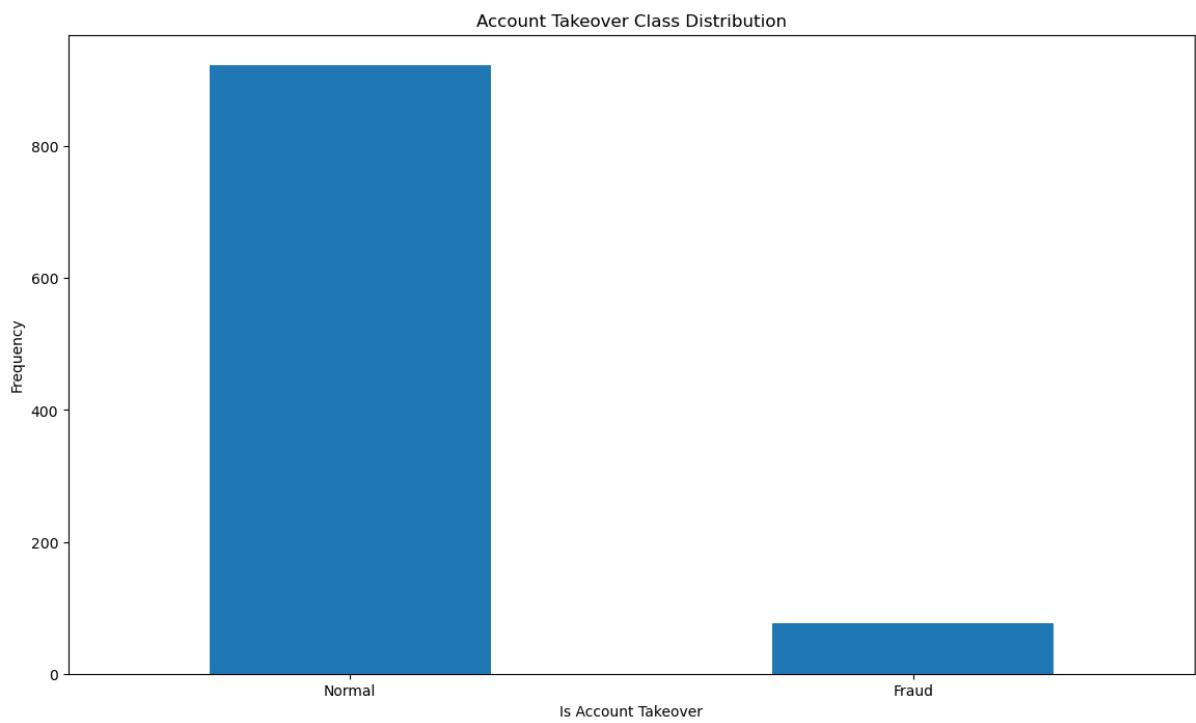
- contamination value == 0.5

Outlier Method: Isolation Forest





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



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

Isolation Forest: 159

Accuracy Score :

0.841

Classification Report :

	precision	recall	f1-score	support
False	0.92	0.90	0.91	923
True	0.09	0.12	0.10	77
accuracy			0.84	1000
macro avg	0.51	0.51	0.51	1000
weighted avg	0.86	0.84	0.85	1000

Local Outlier Factor: 169

Accuracy Score :

0.831

Classification Report :

	precision	recall	f1-score	support
False	0.92	0.90	0.91	923
True	0.04	0.05	0.05	77
accuracy			0.83	1000
macro avg	0.48	0.47	0.48	1000
weighted avg	0.85	0.83	0.84	1000

Support Vector Machine: 603

Accuracy Score :

0.397

Classification Report :

	precision	recall	f1-score	support
False	0.88	0.40	0.55	923
True	0.04	0.32	0.08	77
accuracy			0.40	1000
macro avg	0.46	0.36	0.31	1000
weighted avg	0.81	0.40	0.52	1000

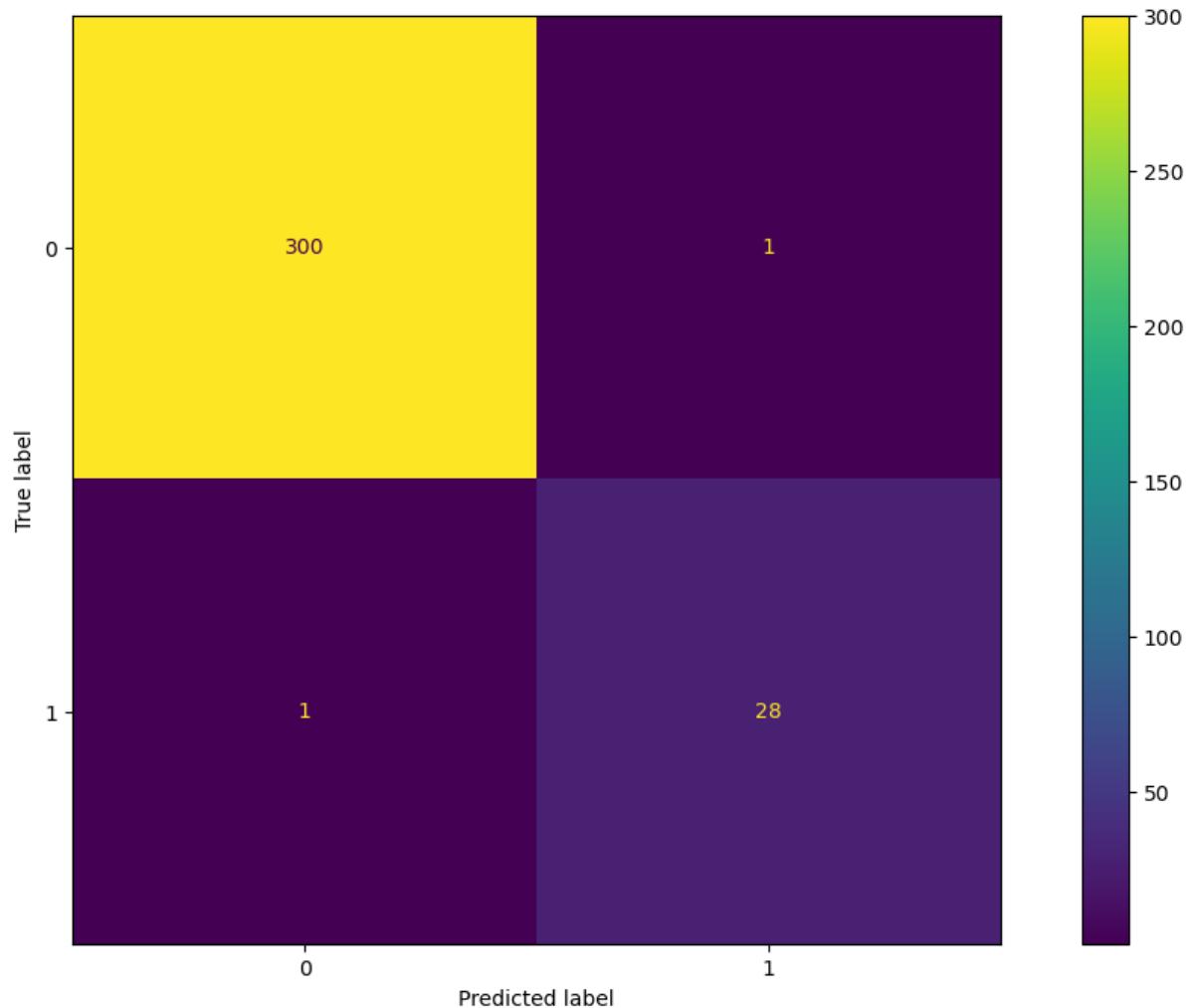
3. Dummy Variable Encoding → df3 =
`pd.get_dummies(df3, columns=['Country', 'Device Type', 'IP Address', 'Login Timestamp', 'Browser Type'], prefix=["cntr", "dvc", "ip", "lgn", "brw"])`

3A. XGBoost

memory usage: 1.9 MB

Accuracy: 99.39%

	precision	recall	f1-score	support
False	1.00	1.00	1.00	301
True	0.97	0.97	0.97	29
accuracy			0.99	330
macro avg	0.98	0.98	0.98	330
weighted avg	0.99	0.99	0.99	330



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 800
The number of records in the test dataset is 200
The training dataset has 742 records for the majority class and 58 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      181
      True         0.10     1.00     0.17      19
      accuracy          0.10      200
      macro avg       0.05     0.50     0.09      200
      weighted avg    0.01     0.10     0.02      200

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.91     1.00     0.95      181
      True         0.00     0.00     0.00      19
      accuracy          0.91      200
      macro avg       0.45     0.50     0.48      200
      weighted avg    0.82     0.91     0.86      200

[[ 0 181]
 [ 0 19]]

```

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

```

The number of records in the training dataset is 700
The number of records in the test dataset is 300
The training dataset has 655 records for the majority class and 45 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      268
      True         0.11     1.00     0.19      32
      accuracy          0.11      300
      macro avg       0.05     0.50     0.10      300
      weighted avg    0.01     0.11     0.02      300

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.89     1.00     0.94      268
      True         0.00     0.00     0.00      32
      accuracy          0.89      300
      macro avg       0.45     0.50     0.47      300
      weighted avg    0.80     0.89     0.84      300

[[ 0 268]
 [ 0 32]]

```

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

```

The number of records in the training dataset is 500
The number of records in the test dataset is 500
The training dataset has 472 records for the majority class and 28 records for the minority class.
      precision    recall   f1-score   support
False        1.00     0.00     0.00      451
True        0.10     1.00     0.18      49

accuracy                           0.10      500
macro avg       0.55     0.50     0.09      500
weighted avg    0.91     0.10     0.02      500

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.90     1.00     0.95      451
True        0.00     0.00     0.00      49

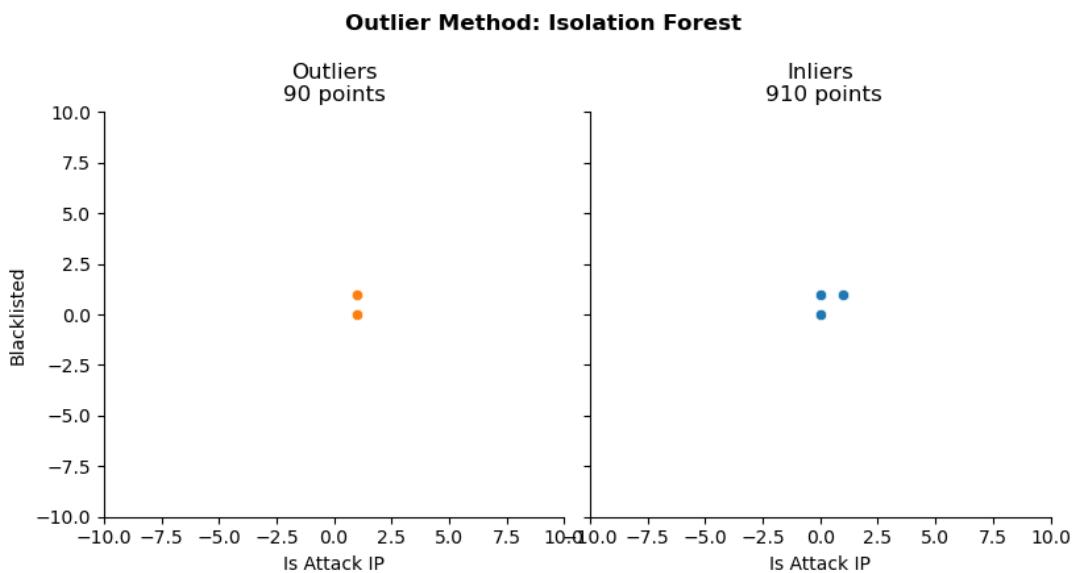
accuracy                           0.90      500
macro avg       0.45     0.50     0.47      500
weighted avg    0.81     0.90     0.86      500

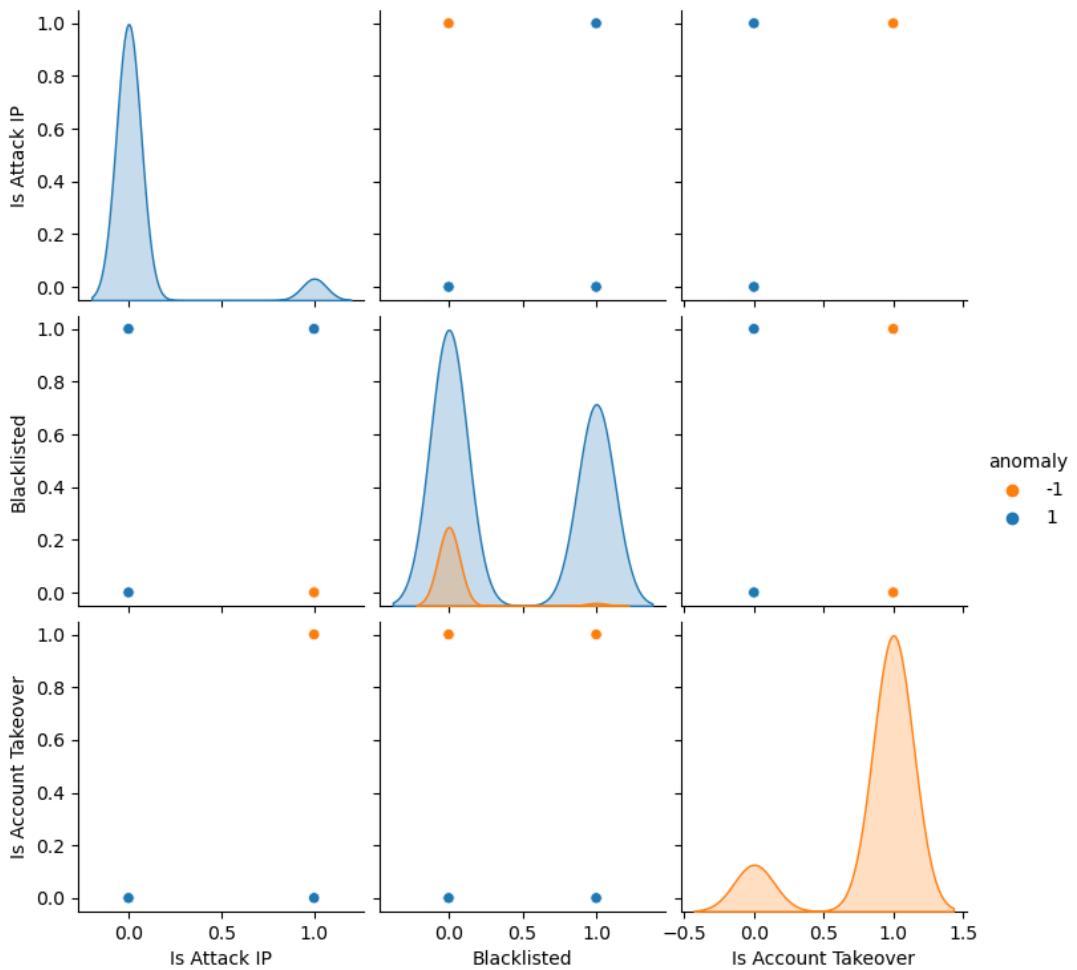
[[ 1 450]
 [ 0 49]]

```

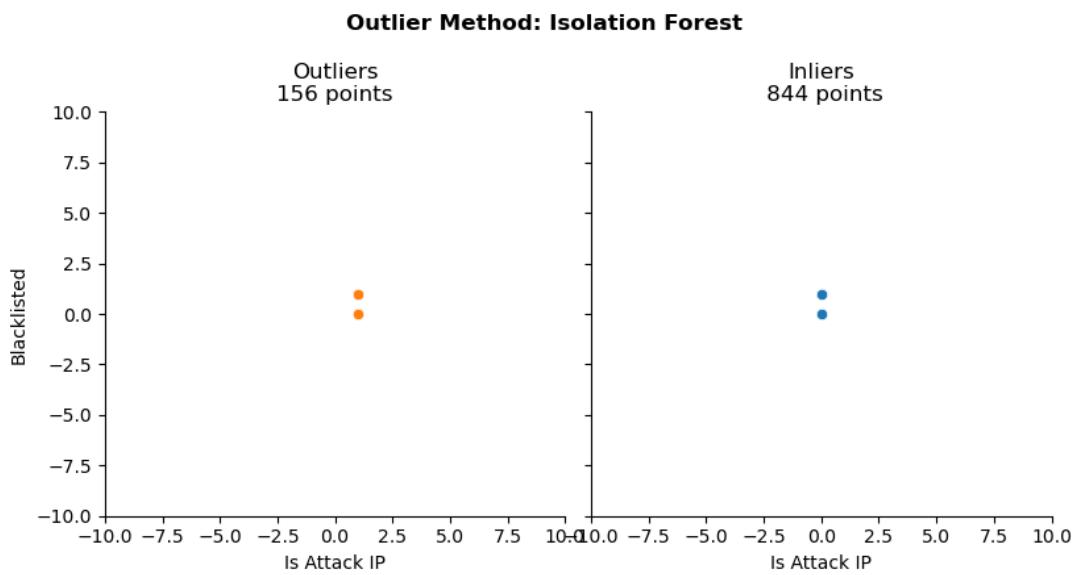
3C. Isolation Forest

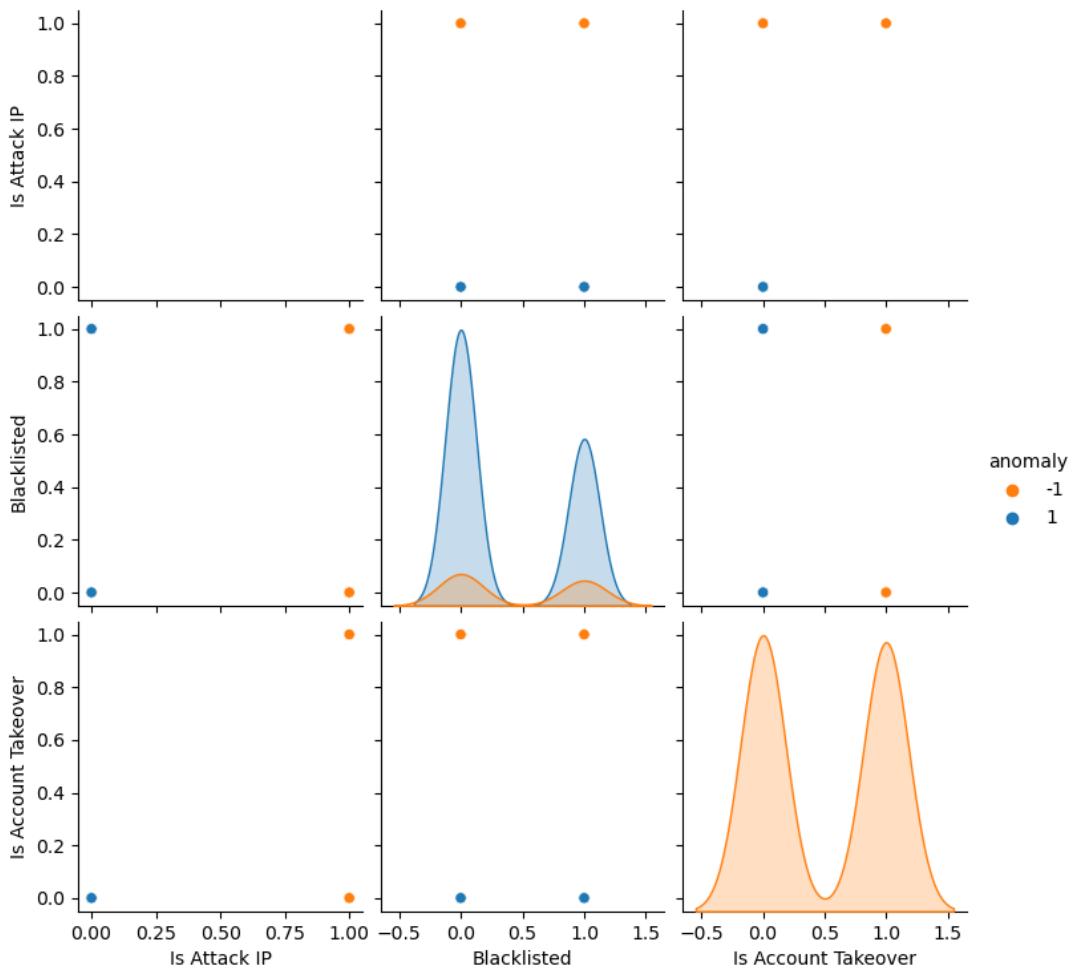
- anomaly_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
 - contamination value == 0.1



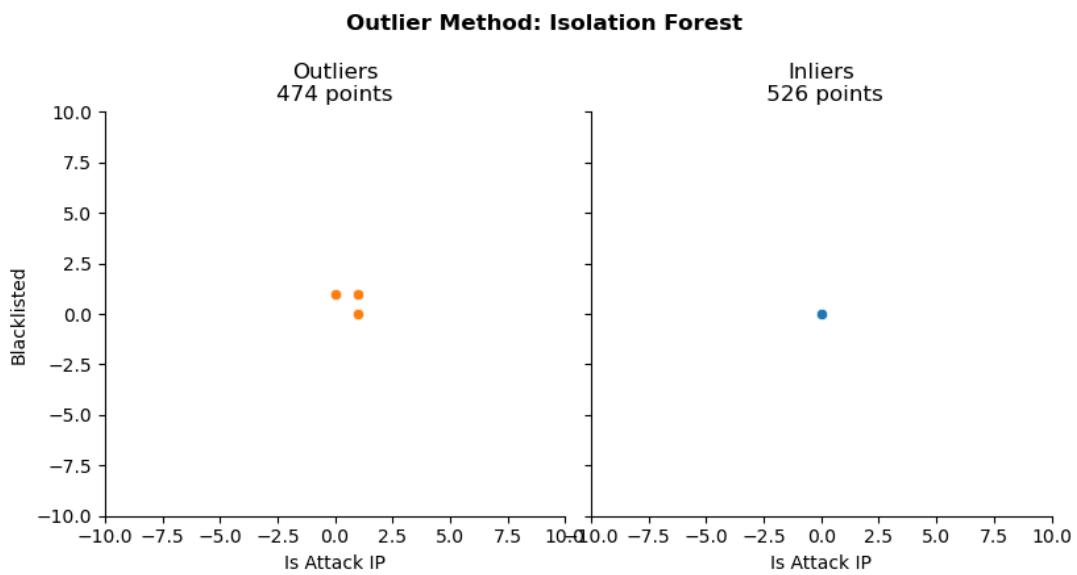


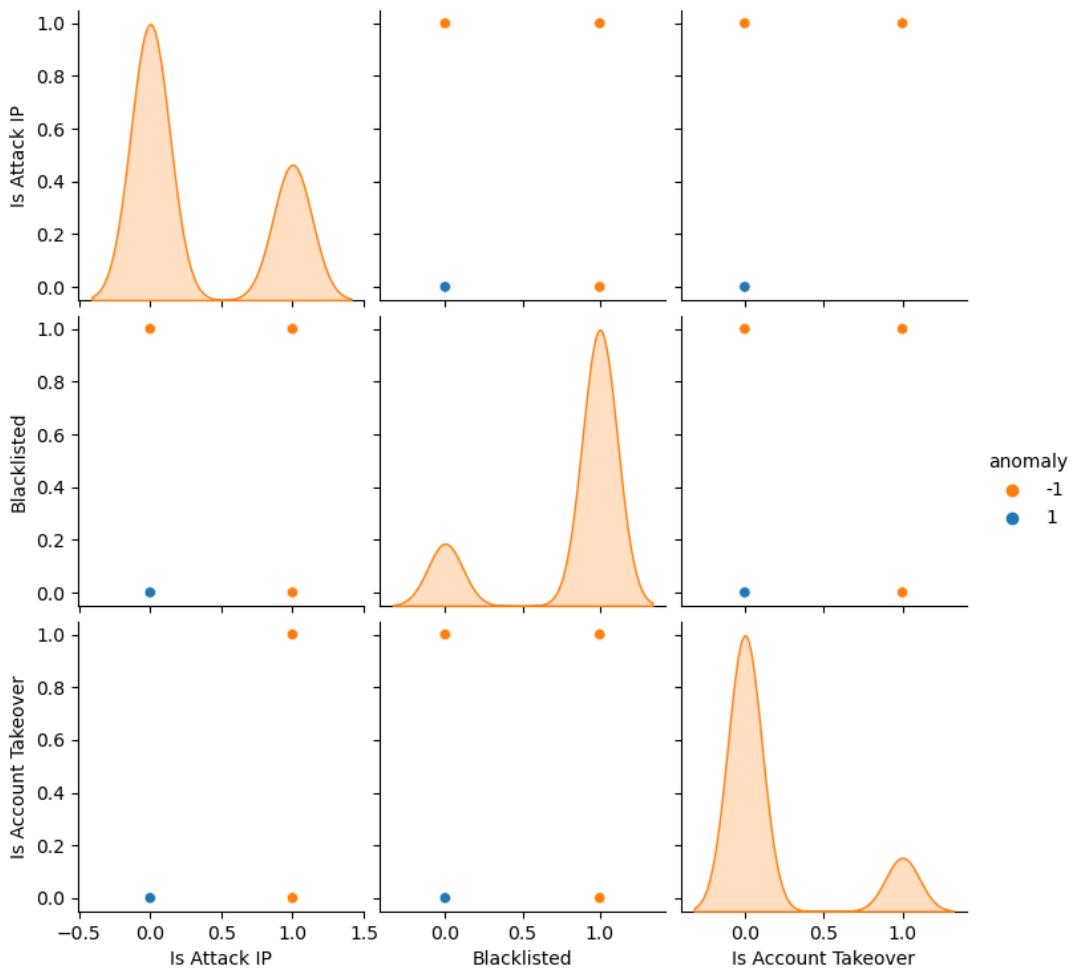
- contamination value == 0.3



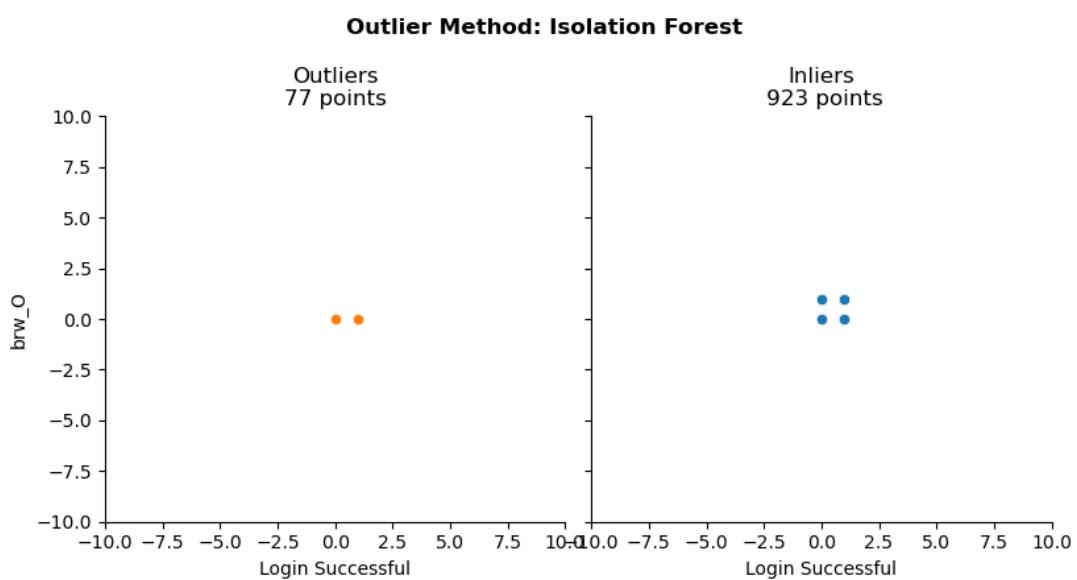


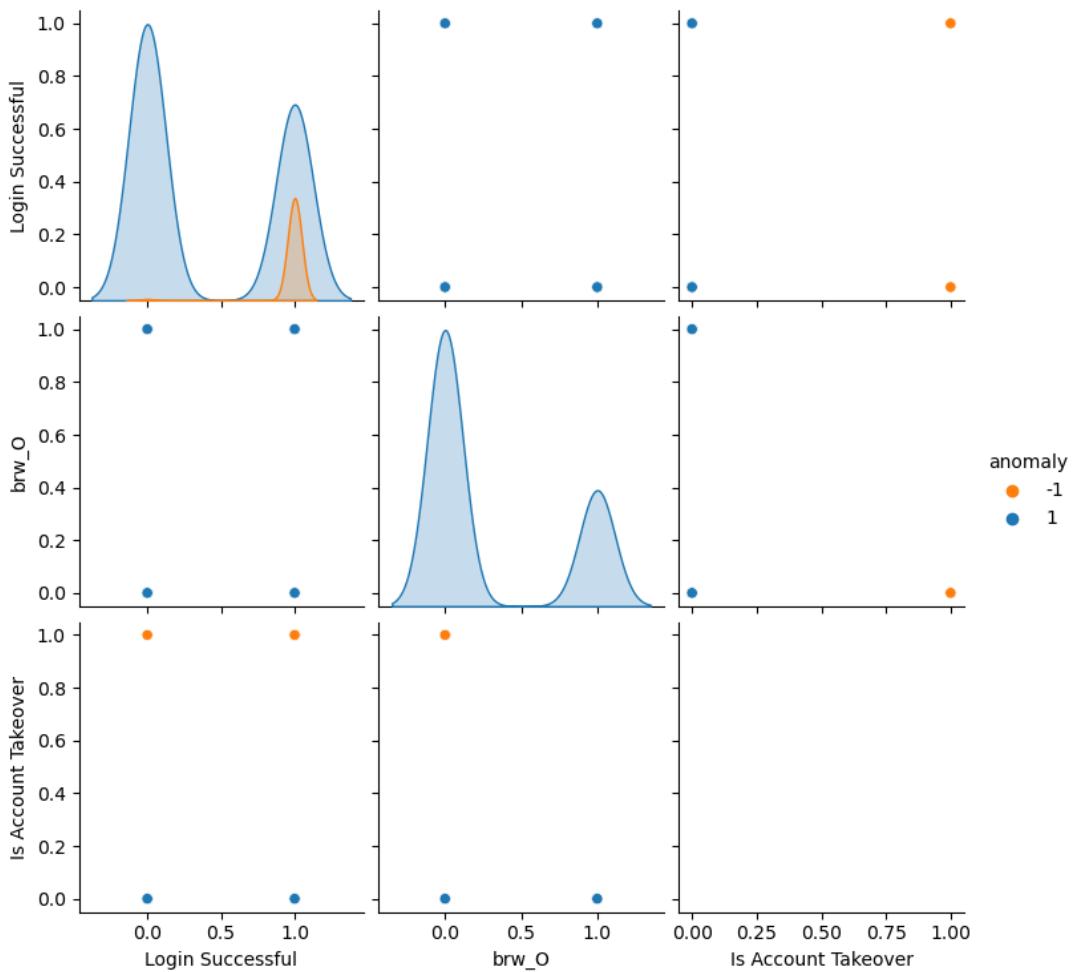
- contamination value == 0.5





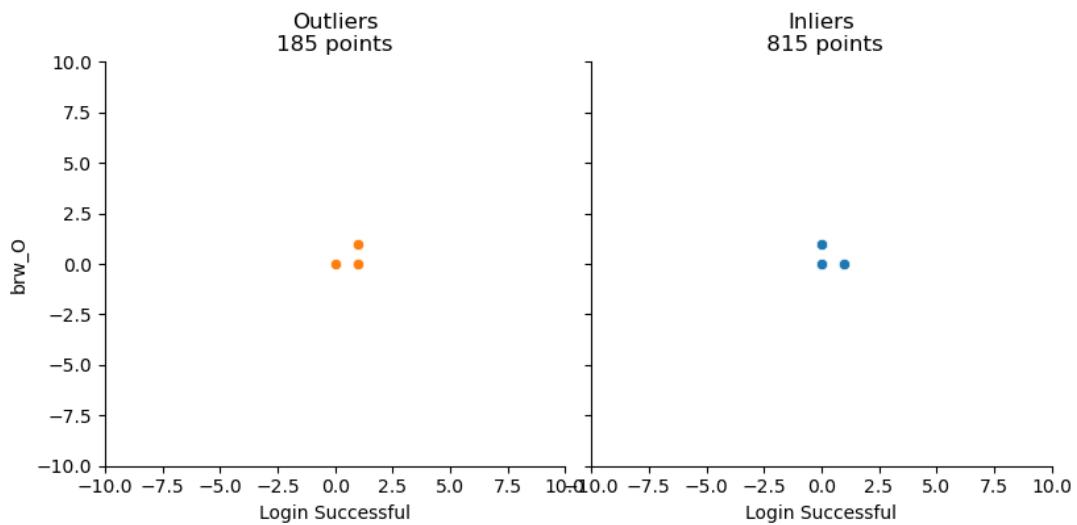
- anomaly_inputs = ['Login Successful', 'Browser Type (brw_O)', 'Is Account Takeover']
 - contamination value == 0.1

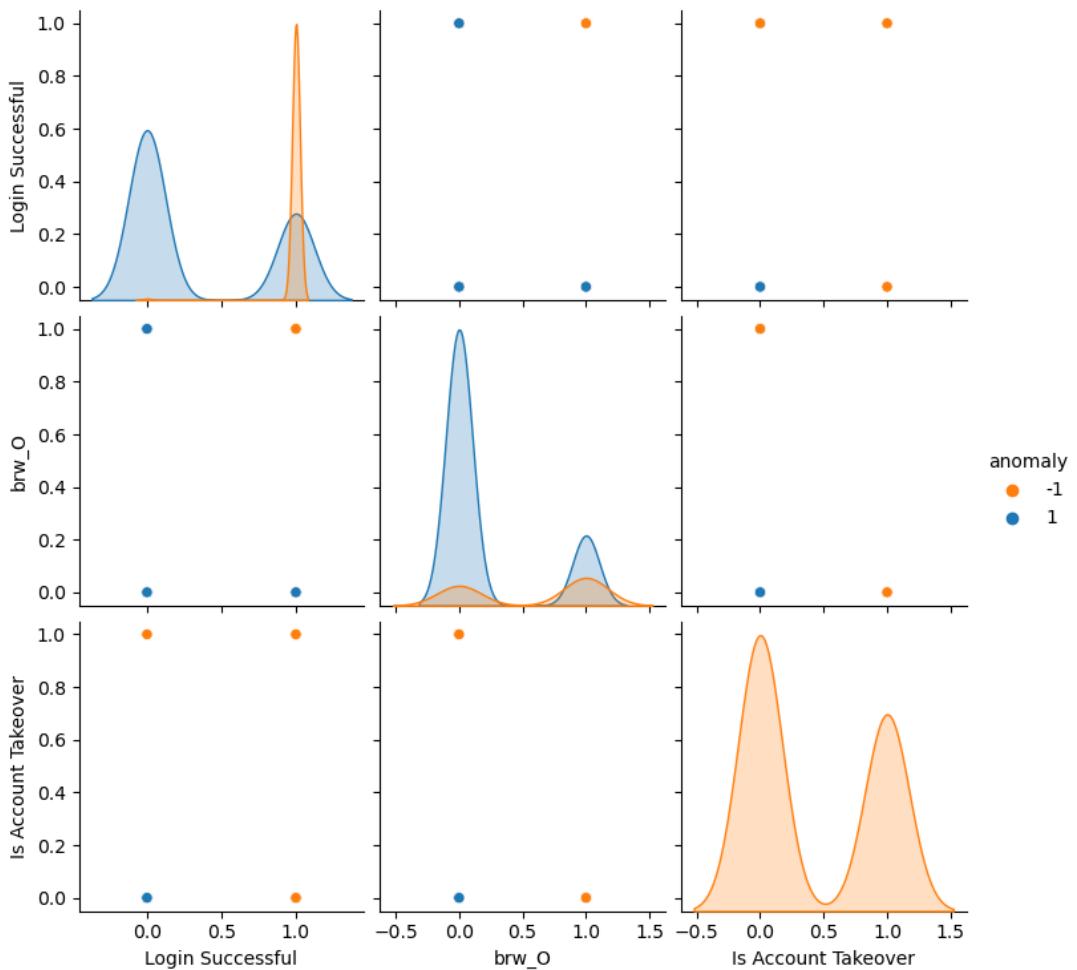




- contamination value == 0.3

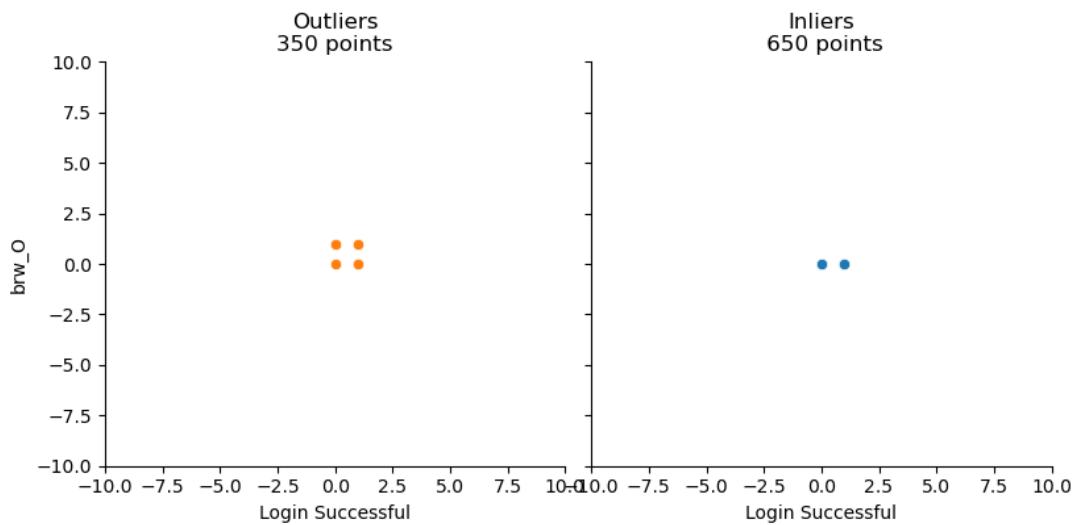
Outlier Method: Isolation Forest

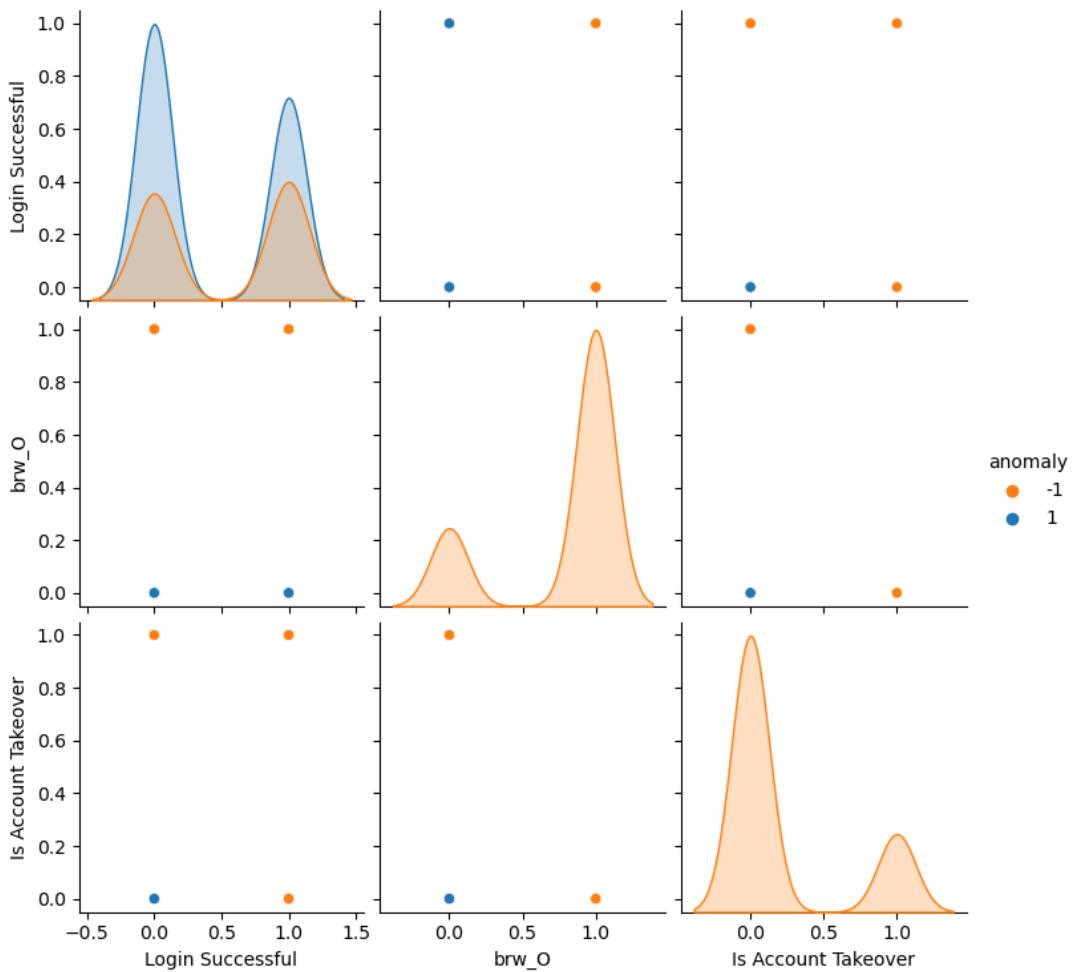




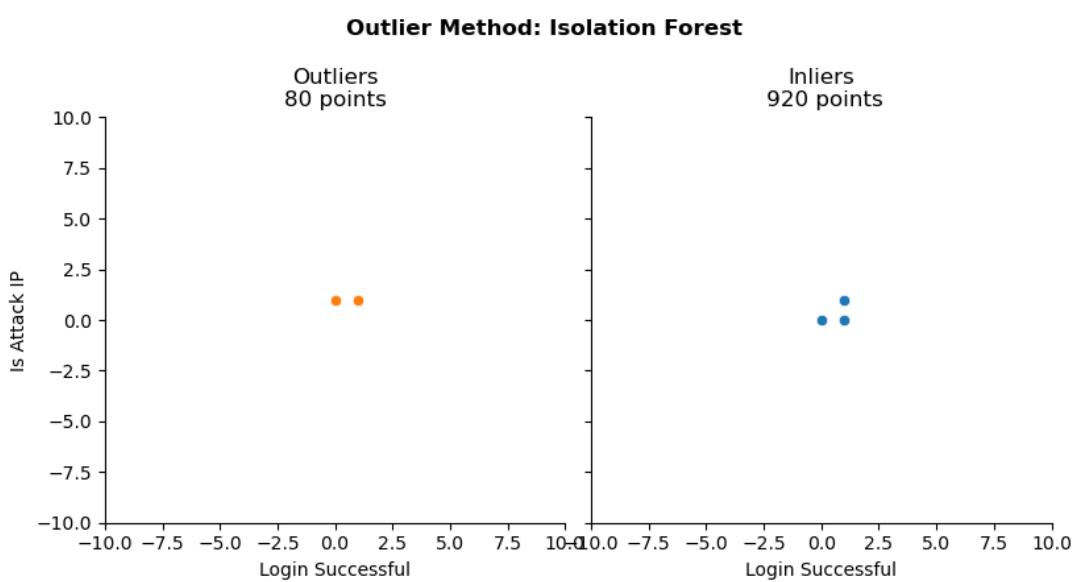
- contamination value == 0.5

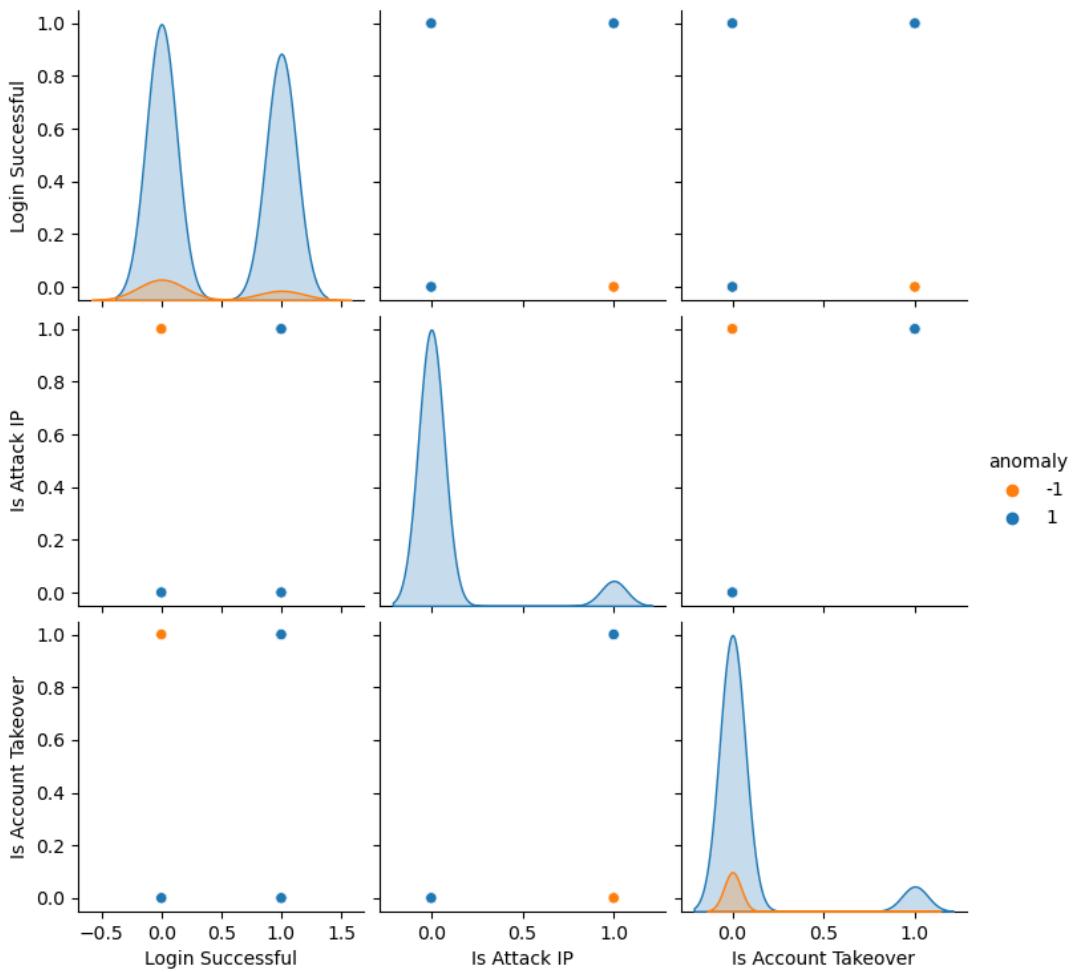
Outlier Method: Isolation Forest





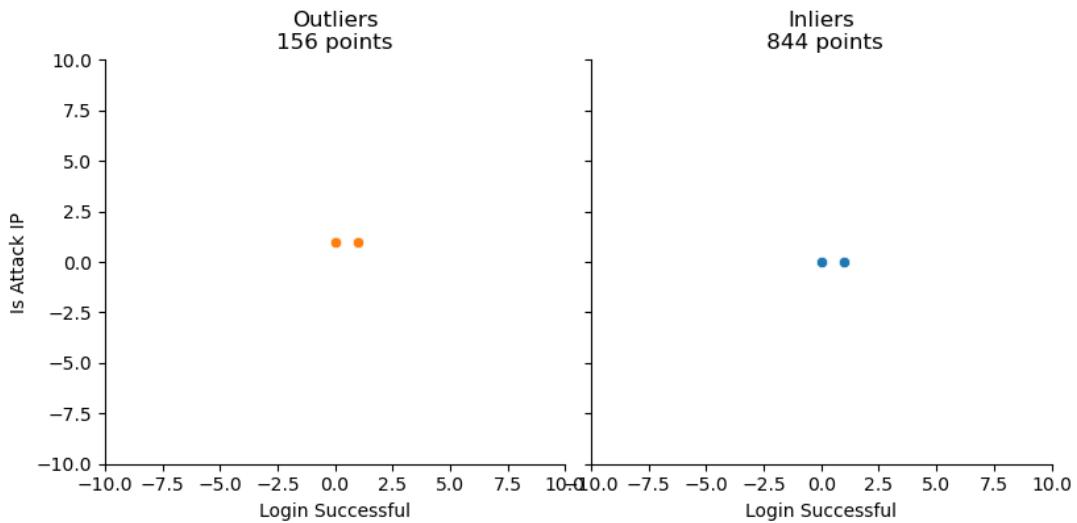
- anomaly_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
 - contamination value == 0.1

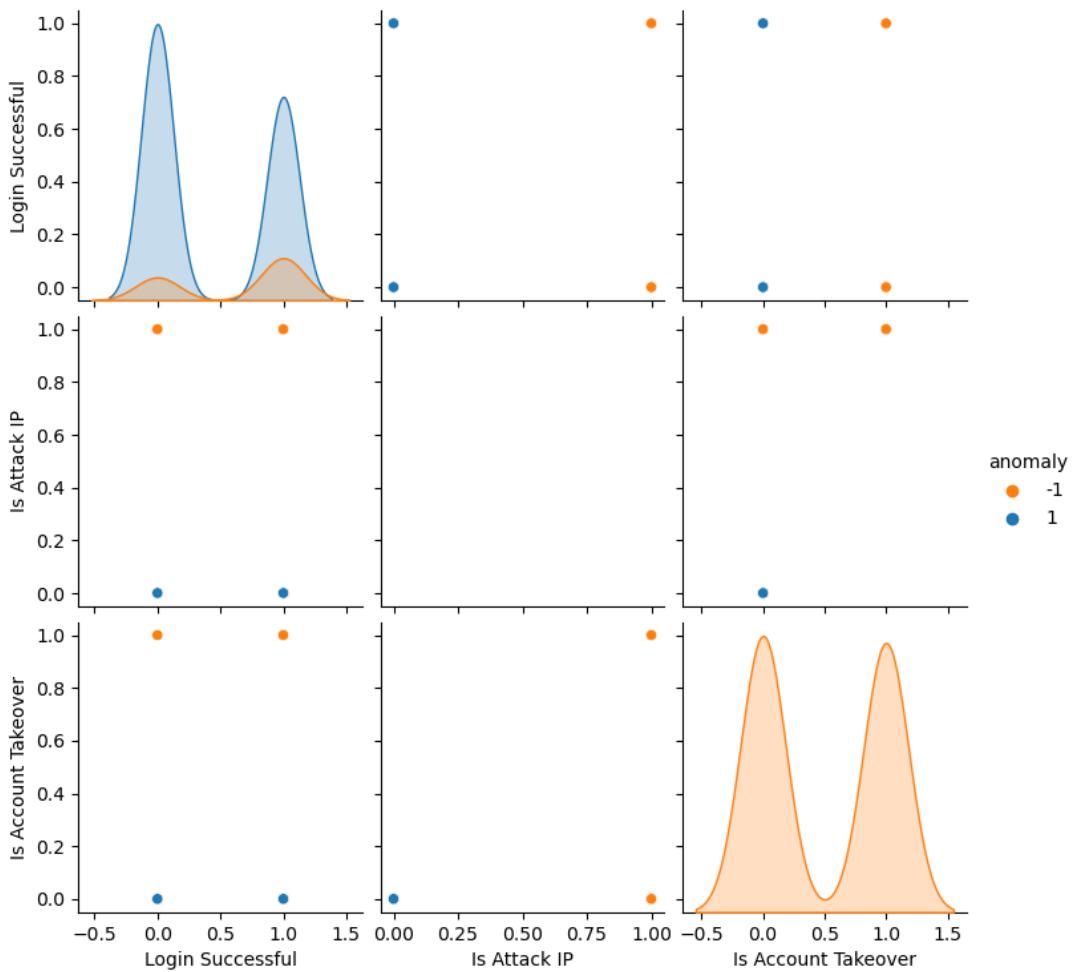




- contamination value == 0.3

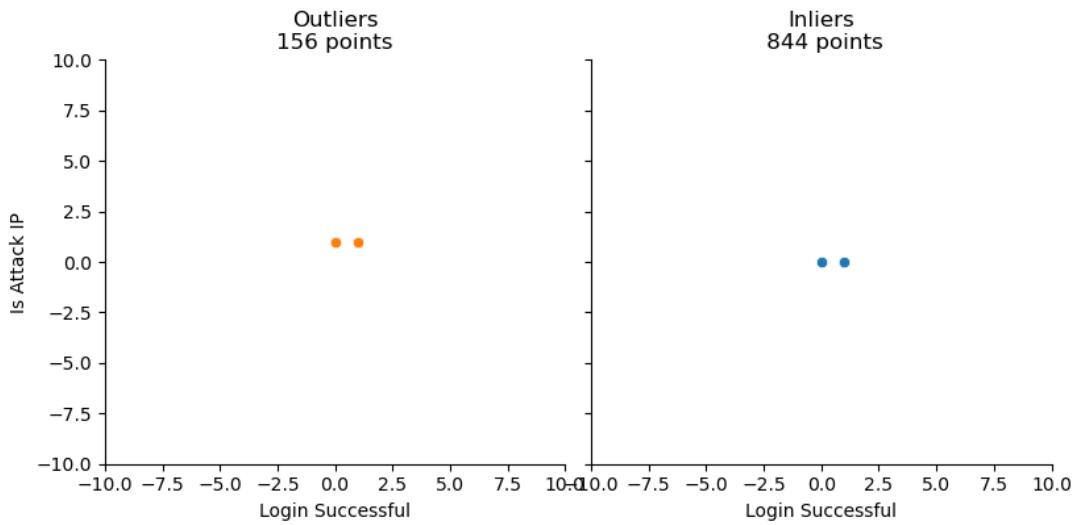
Outlier Method: Isolation Forest

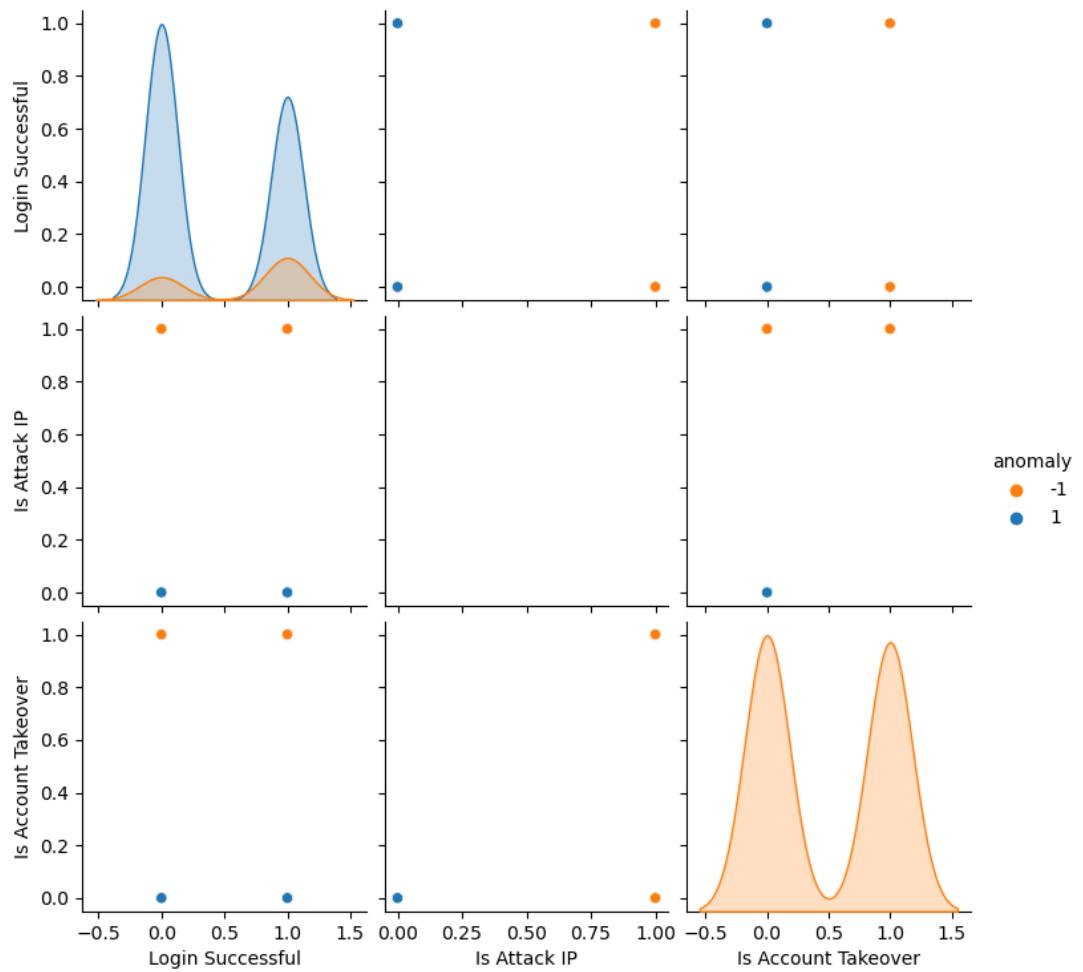




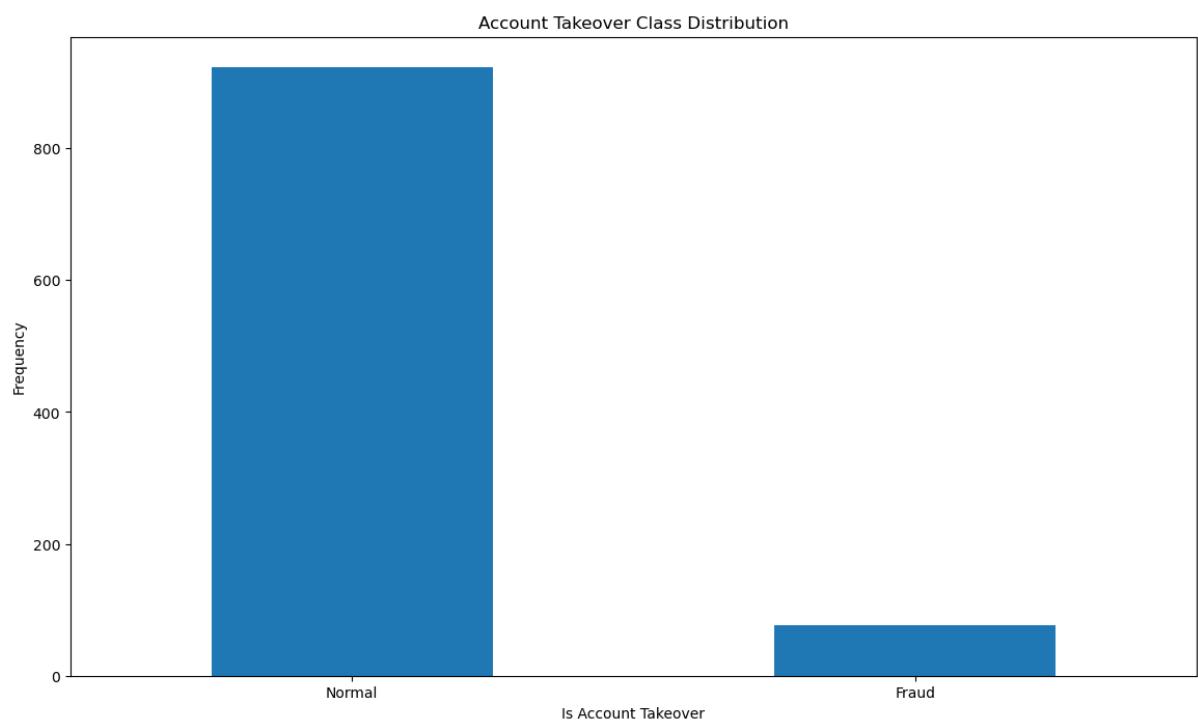
- contamination value == 0.5

Outlier Method: Isolation Forest





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



Isolation Forest: 147

Accuracy Score :

0.853

Classification Report :

	precision	recall	f1-score	support
False	0.93	0.91	0.92	923
True	0.15	0.19	0.17	77
accuracy			0.85	1000
macro avg	0.54	0.55	0.54	1000
weighted avg	0.87	0.85	0.86	1000

Local Outlier Factor: 147

Accuracy Score :

0.853

Classification Report :

	precision	recall	f1-score	support
False	0.93	0.91	0.92	923
True	0.15	0.19	0.17	77
accuracy			0.85	1000
macro avg	0.54	0.55	0.54	1000
weighted avg	0.87	0.85	0.86	1000

Support Vector Machine: 603

Accuracy Score :

0.397

Classification Report :

	precision	recall	f1-score	support
False	0.88	0.40	0.55	923
True	0.04	0.32	0.08	77
accuracy			0.40	1000
macro avg	0.46	0.36	0.31	1000
weighted avg	0.81	0.40	0.52	1000