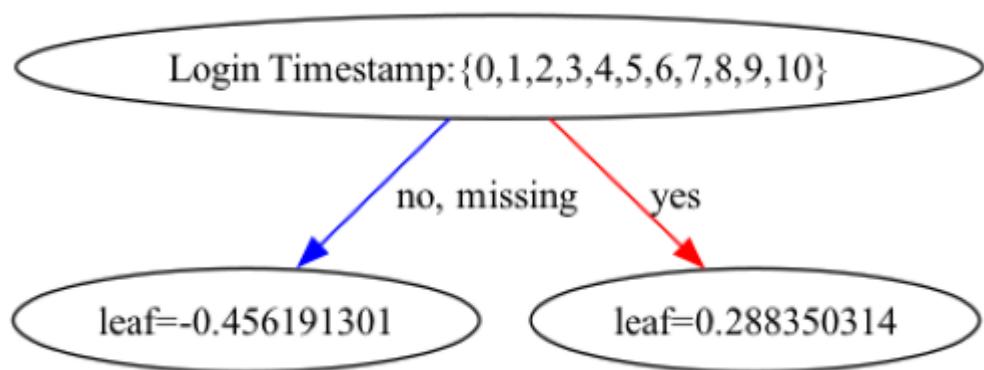
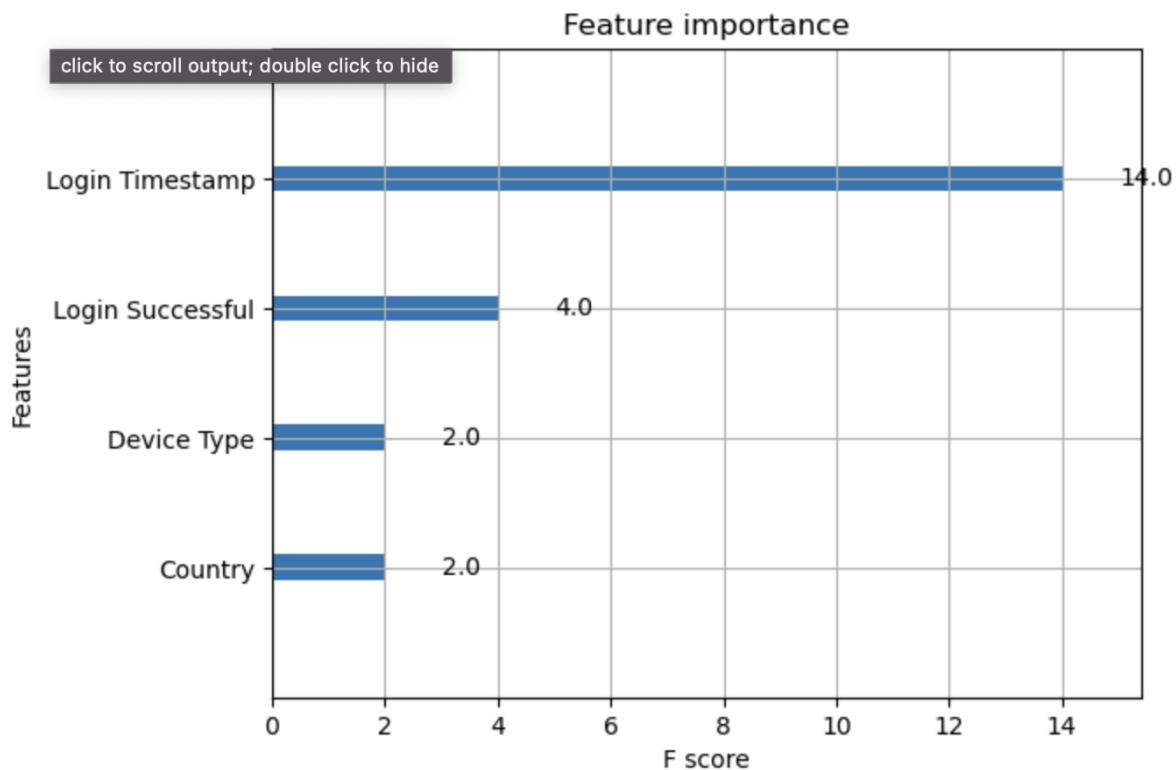


# ATO 200 rows

I. 10/200 → 5%

## 1. XGB Experimental

```
memory usage: 22.0 KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   ASN               200 non-null    float64
 1   Country           200 non-null    category
 2   Device Type       200 non-null    category
 3   IP Address        200 non-null    category
 4   Is Attack IP     200 non-null    bool   
 5   Login Successful  200 non-null    bool   
 6   Login Timestamp   200 non-null    category
 7   User ID           200 non-null    float64
 8   Blacklisted       200 non-null    bool   
 9   Browser Type      200 non-null    category
dtypes: bool(3), category(5), float64(2)
memory usage: 21.8 KB
Feature importances:
[0.          0.18094781 0.07451553 0.          0.          0.         0.09844579
 0.64609087 0.          0.          0.          ]
```



## 2. Label encoding (cat.codes)

### 2A. XGBoost

memory usage: 5.4 KB

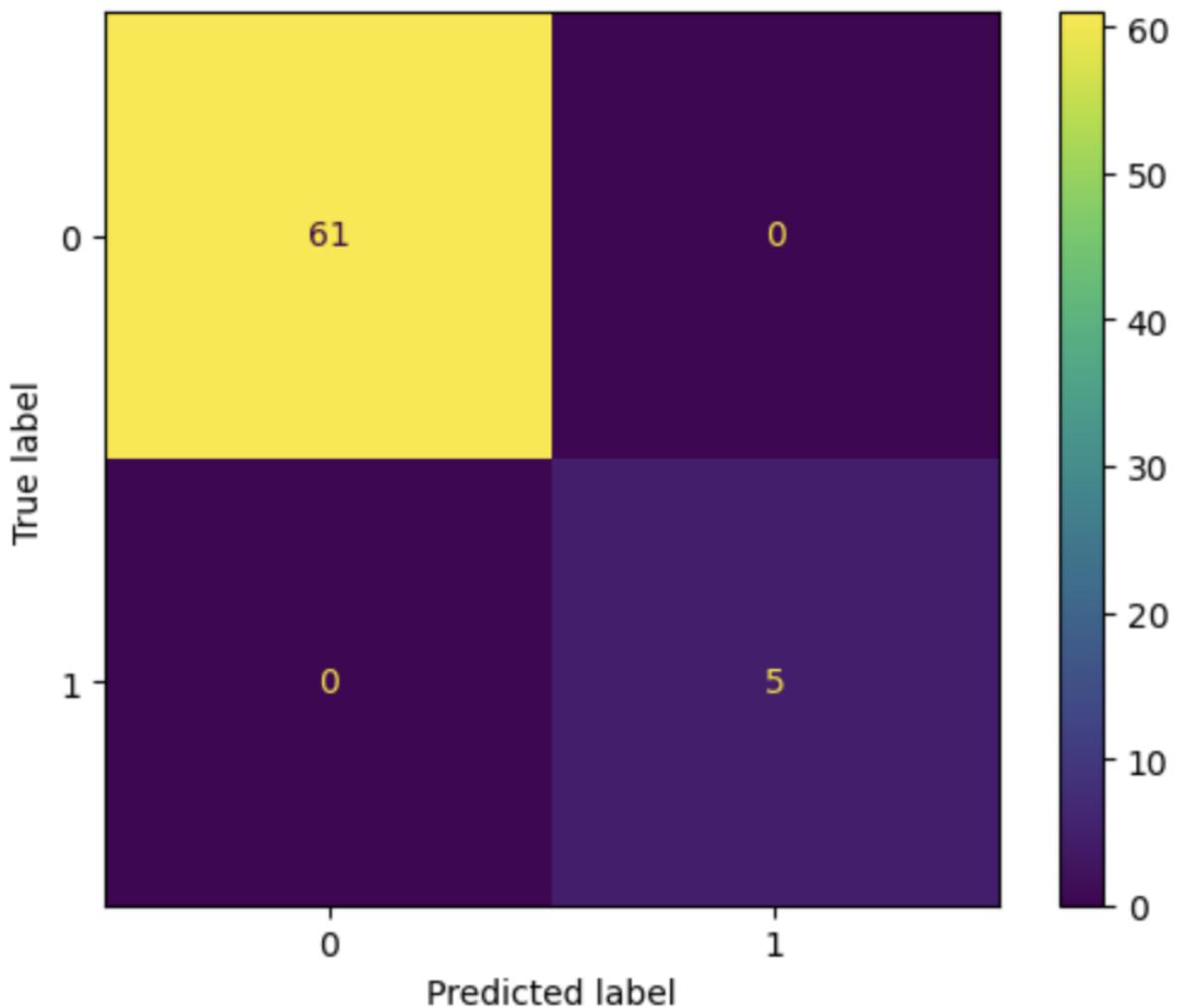
```

Accuracy: 100.00%
      precision    recall   f1-score   support

      False       1.00     1.00     1.00      61
      True        1.00     1.00     1.00       5

  accuracy          1.00     1.00     1.00      66
  macro avg       1.00     1.00     1.00      66
weighted avg      1.00     1.00     1.00      66

```



## 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 160
The number of records in the test dataset is 40
The training dataset has 153 records for the majority class and 7 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      37
      True         0.07     1.00     0.14       3
      accuracy          0.07      40
      macro avg       0.04     0.50     0.07      40
      weighted avg    0.01     0.07     0.01      40
The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.93     1.00     0.96      37
      True         0.00     0.00     0.00       3
      accuracy          0.93      40
      macro avg       0.46     0.50     0.48      40
      weighted avg    0.86     0.93     0.89      40
[[ 0 37]
 [ 0  3]]

```

- columns=['Is Account Takeover', 'Login Timestamp'] test\_size=0.3

```

The number of records in the training dataset is 140
The number of records in the test dataset is 60
The training dataset has 133 records for the majority class and 7 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      57
      True         0.05     1.00     0.10       3
      accuracy          0.05      60
      macro avg       0.03     0.50     0.05      60
      weighted avg    0.00     0.05     0.00      60
The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.95     1.00     0.97      57
      True         0.00     0.00     0.00       3
      accuracy          0.95      60
      macro avg       0.47     0.50     0.49      60
      weighted avg    0.90     0.95     0.93      60
[[ 0 57]
 [ 0  3]]

```

- columns=['Is Account Takeover', 'Login Timestamp'] test\_size=0.5

```

The number of records in the training dataset is 100
The number of records in the test dataset is 100
The training dataset has 95 records for the majority class and 5 records for the minority class.
      precision    recall   f1-score   support
False        0.00     0.00     0.00      95
True        0.05     1.00     0.10       5

accuracy          0.05      100
macro avg       0.03     0.50     0.05      100
weighted avg    0.00     0.05     0.00      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.95     1.00     0.97      95
True        0.00     0.00     0.00       5

accuracy          0.95      100
macro avg       0.47     0.50     0.49      100
weighted avg    0.90     0.95     0.93      100

[[ 0 95]
 [ 0  5]]

```

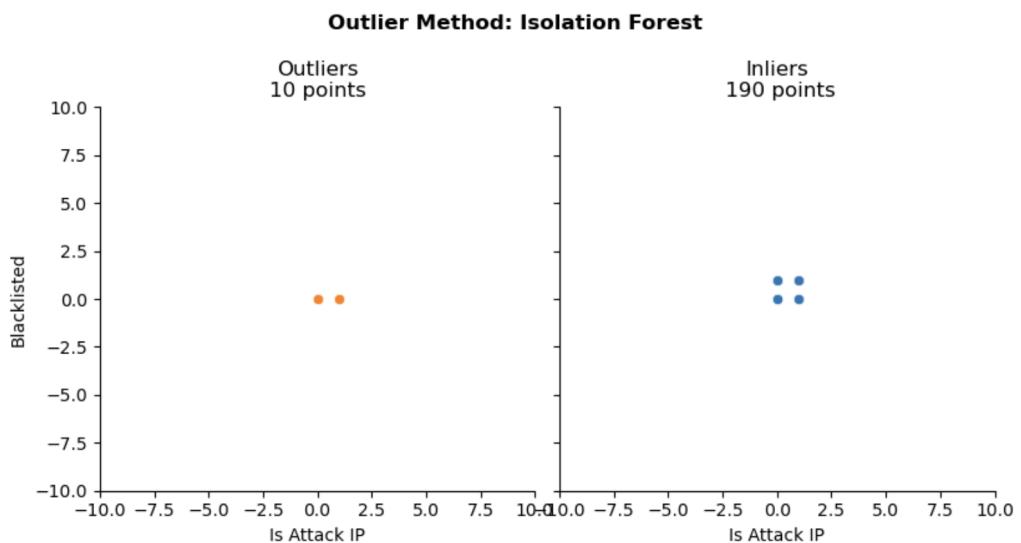
## 2C. Isolation Forest

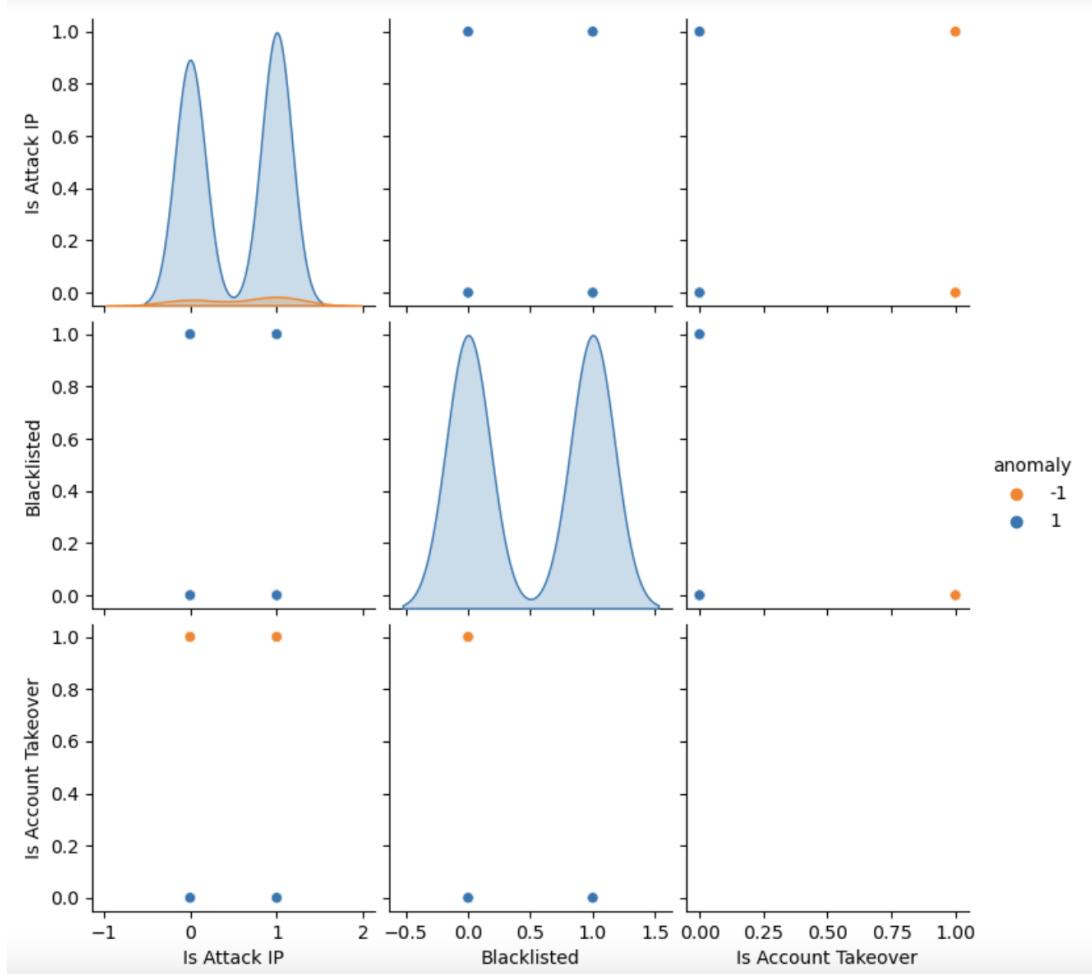
- anomaly\_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 10
Number of non anomalous values 190
Total Number of Values: 200

```



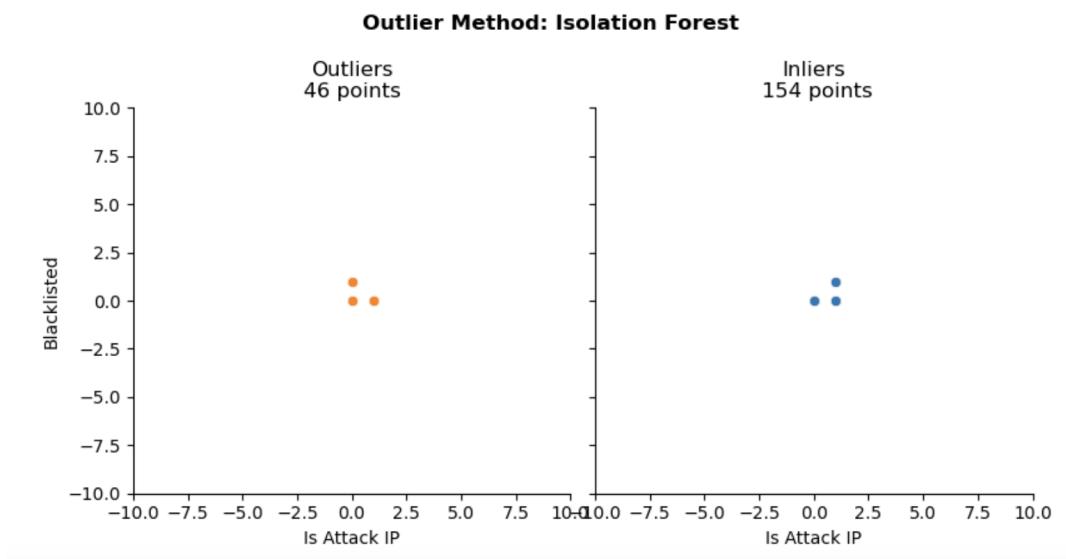


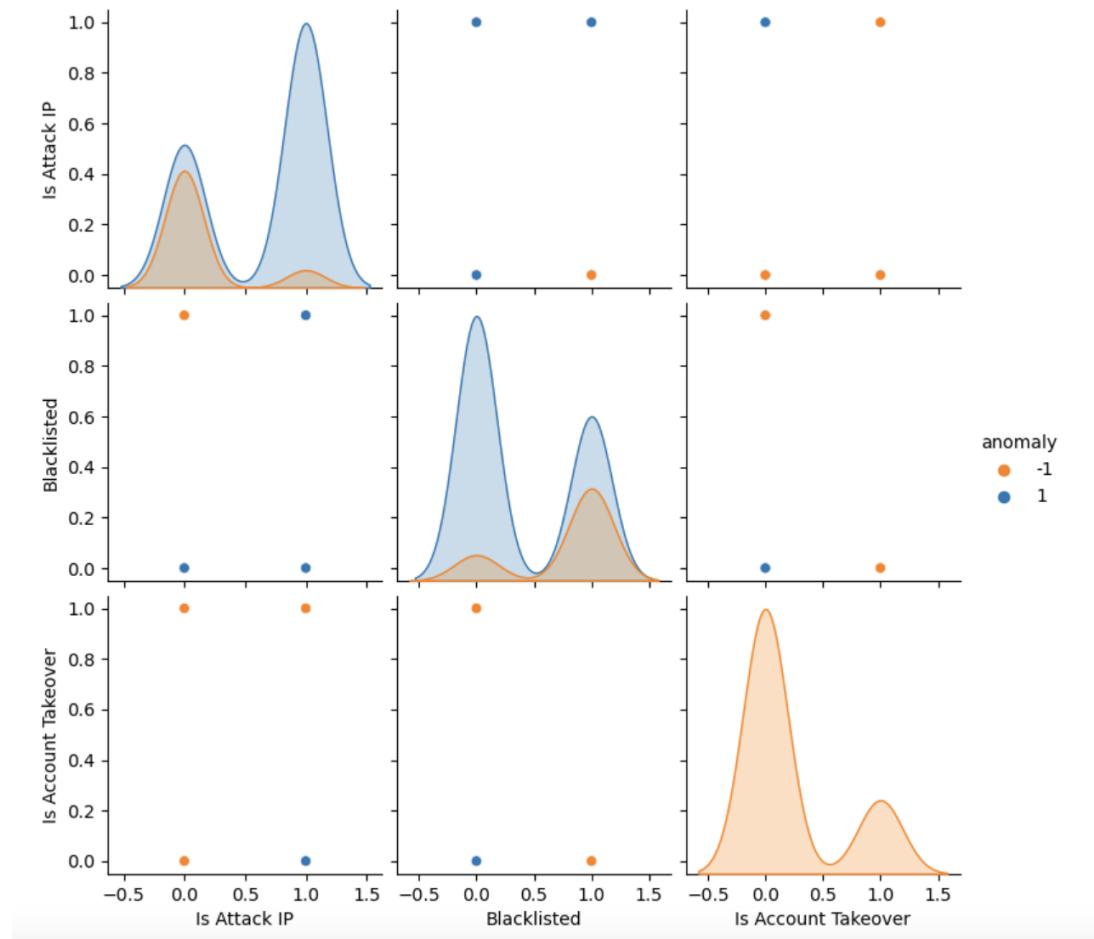
- increasing contamination value to 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 46
Number of non anomalous values 154
Total Number of Values: 200
:<seaborn.axisgrid.PairGrid at 0x165d16790>

```



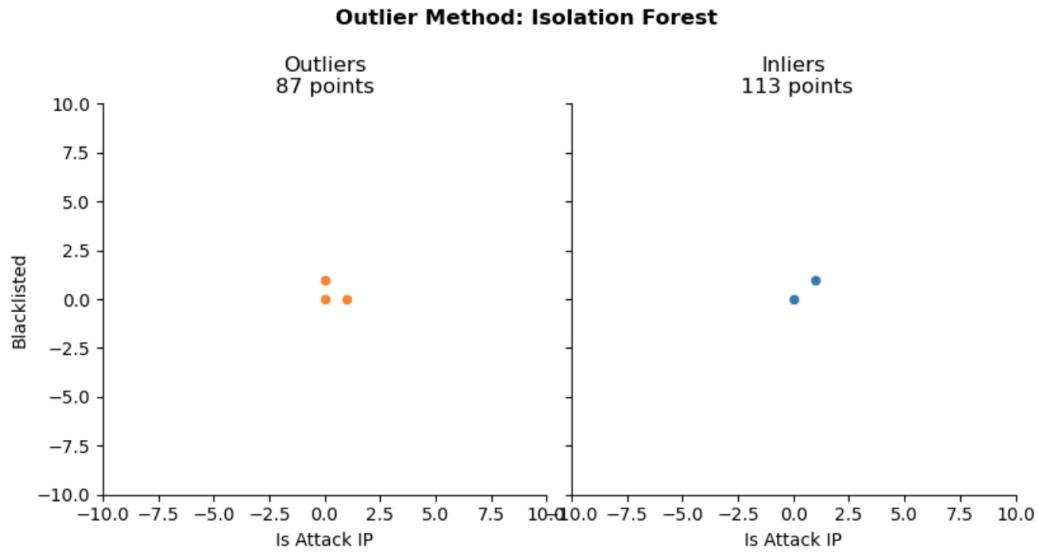


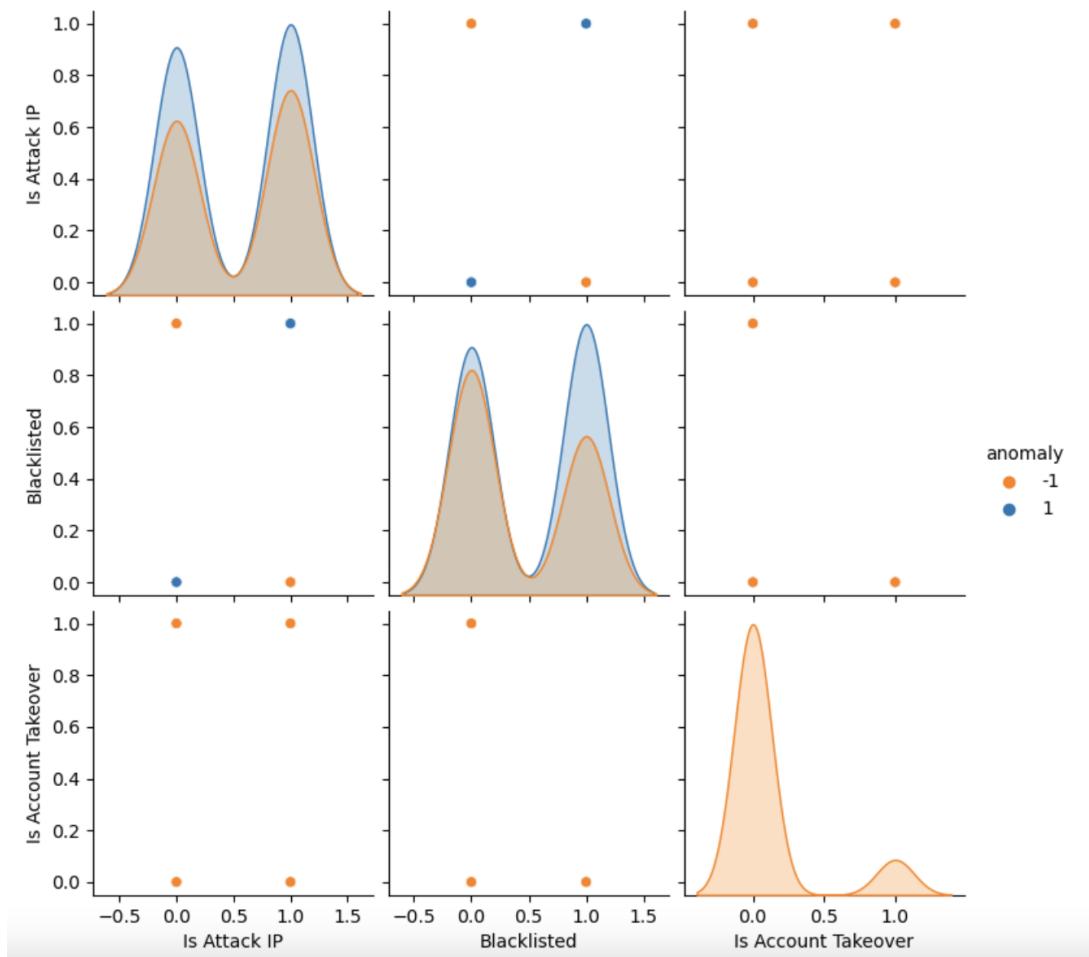
- contamination value == 0.5

```

Outlier Method: Isolation Forest
Number of anomalous values 87
Number of non anomalous values 113
Total Number of Values: 200
: <seaborn.axisgrid.PairGrid at 0x166f5dd50>

```





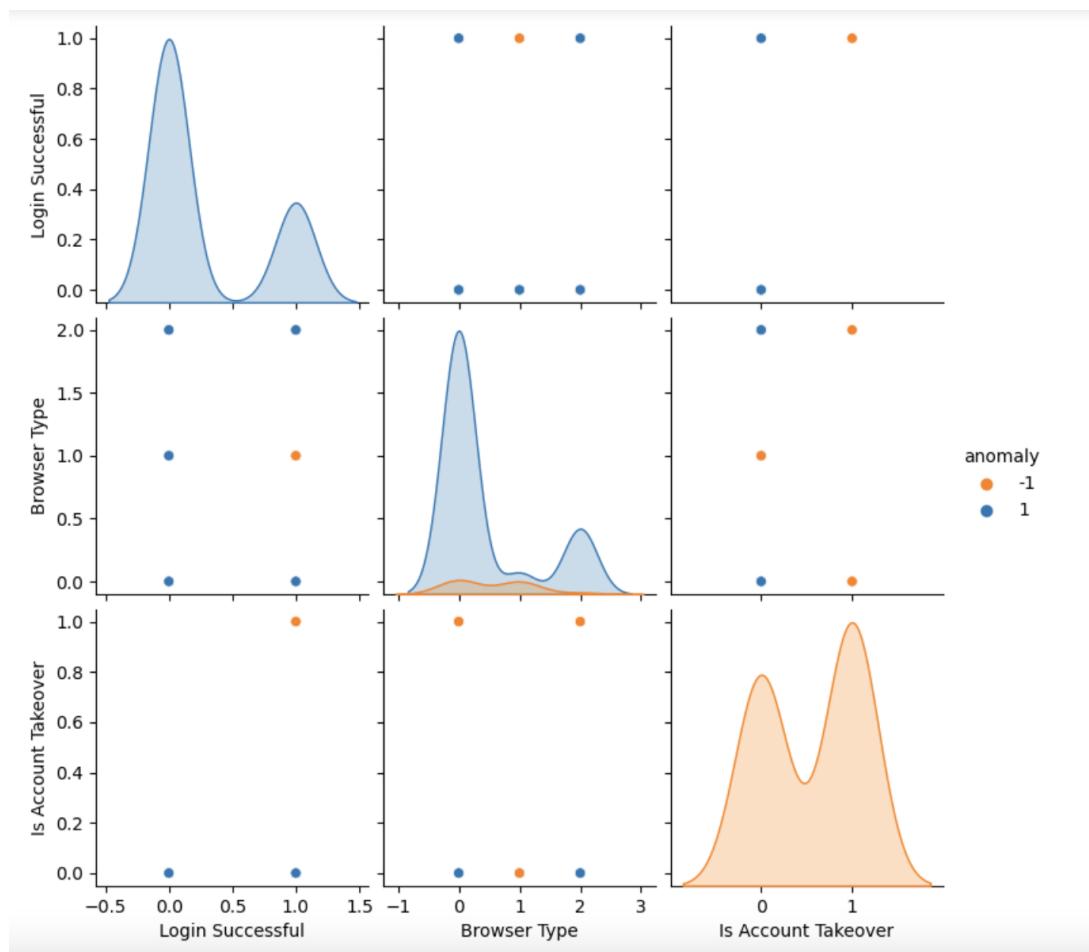
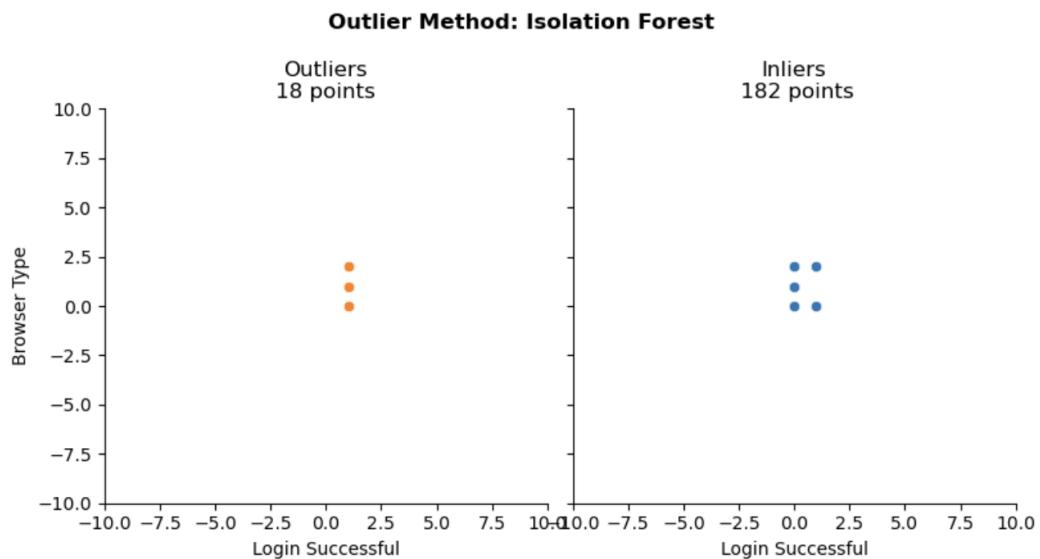
- anomaly\_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 18
Number of non anomalous values 182
Total Number of Values: 200

<seaborn.axisgrid.PairGrid at 0x2b5c4d950>

```



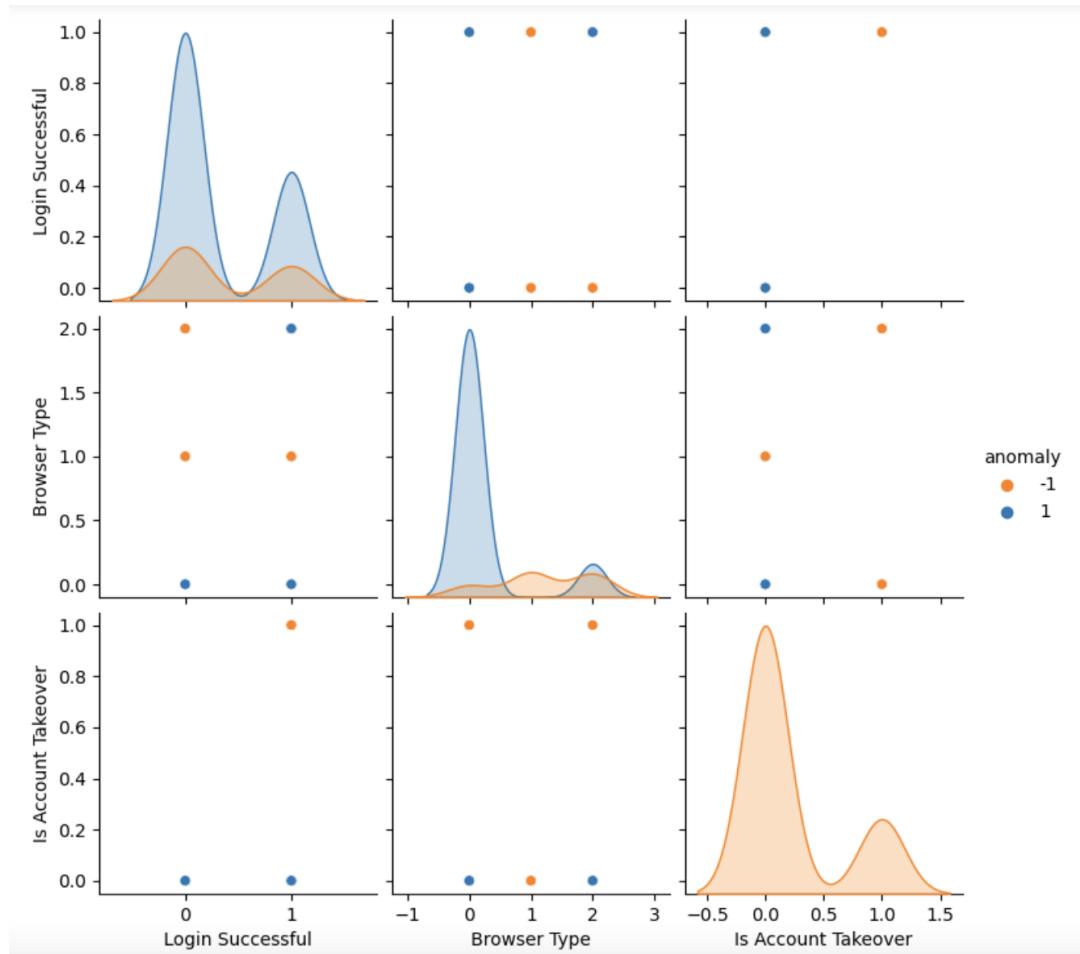
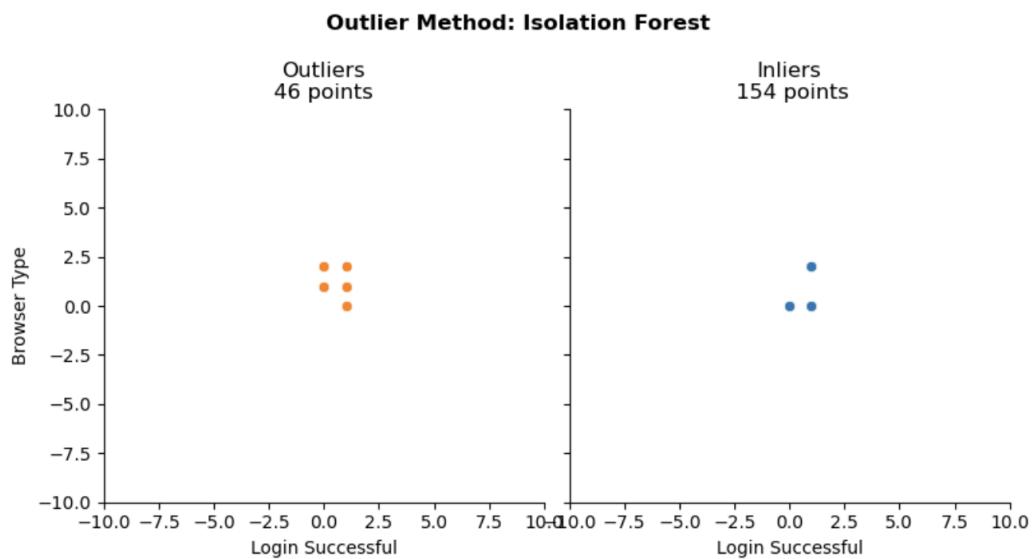
- contamination value == 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 46
Number of non anomalous values 154
Total Number of Values: 200

```

<seaborn.axisgrid.PairGrid at 0x2b6902890>



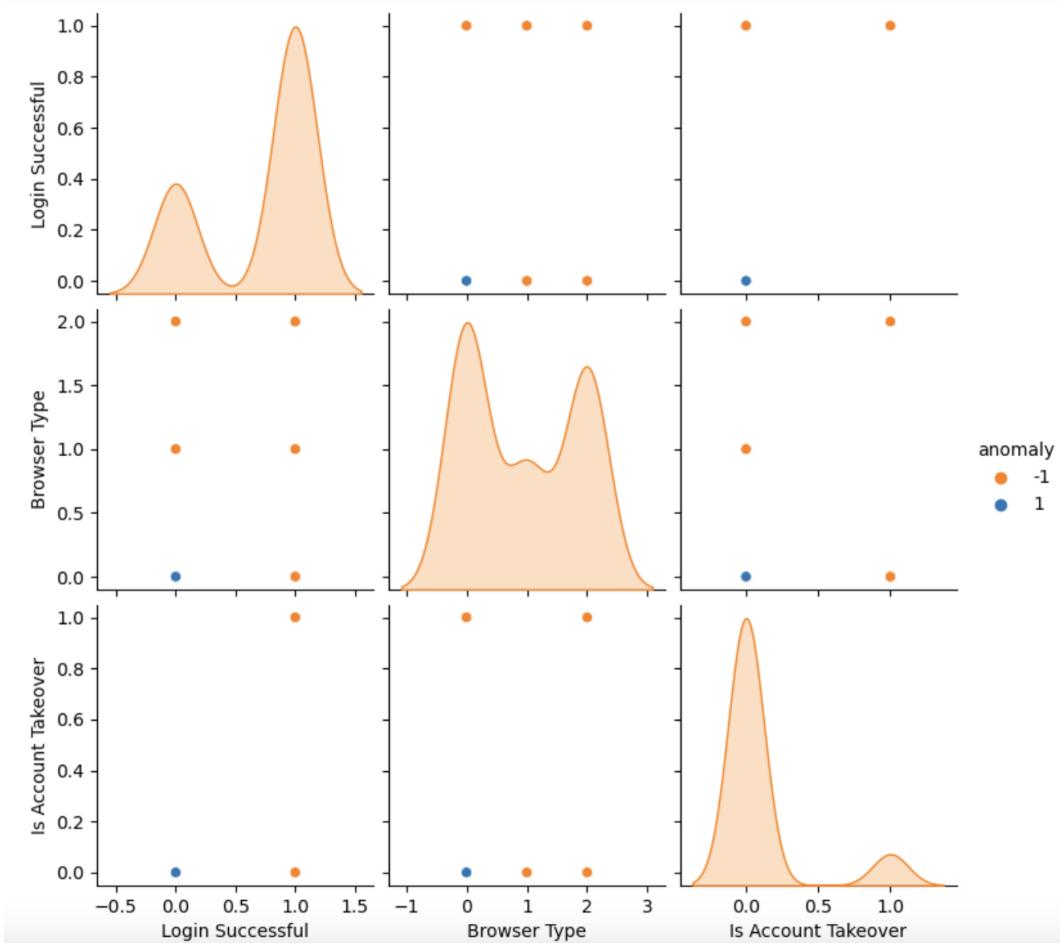
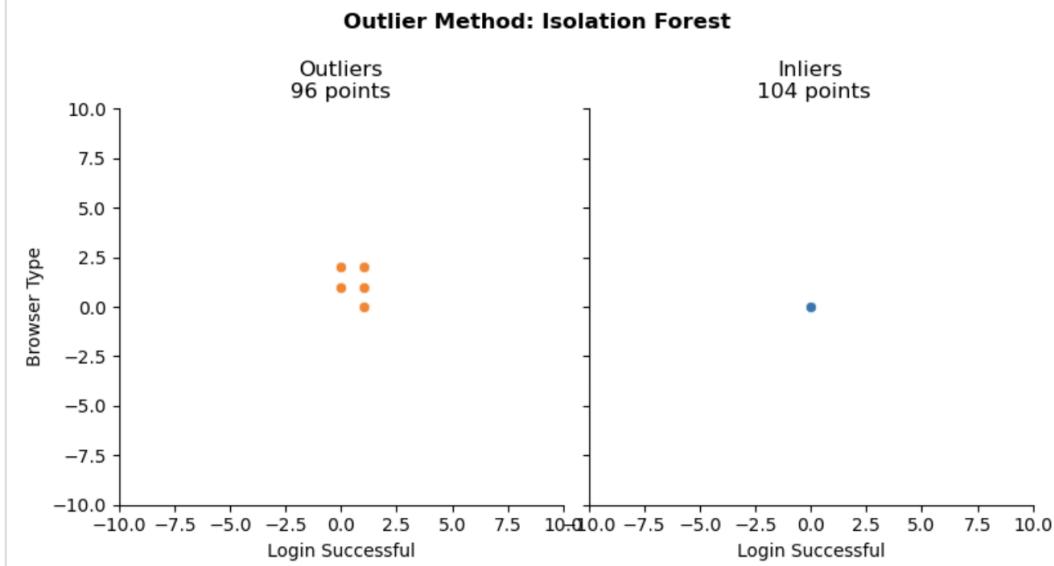
- contamination value == 0.5

```

Outlier Method: Isolation Forest
Number of anomalous values 96
Number of non anomalous values 104
Total Number of Values: 200

<seaborn.axisgrid.PairGrid at 0x2b7383650>

```

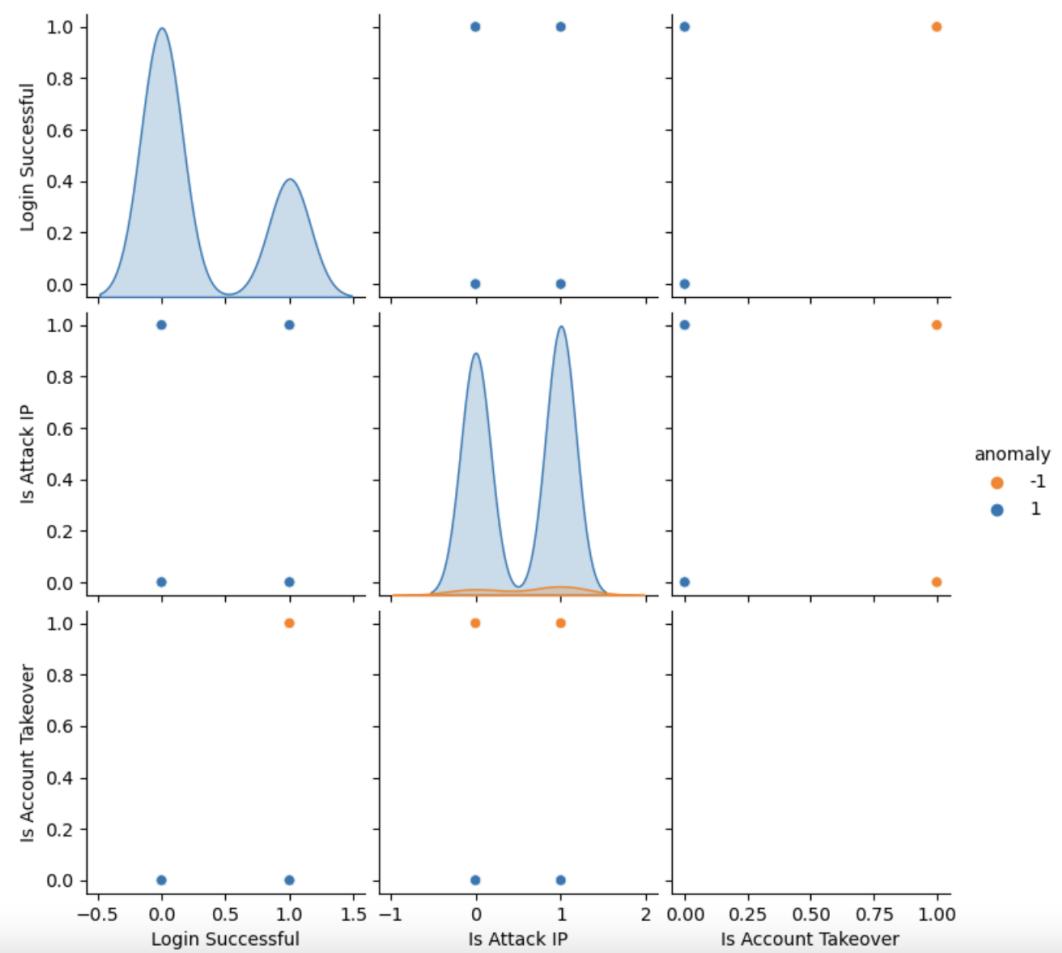
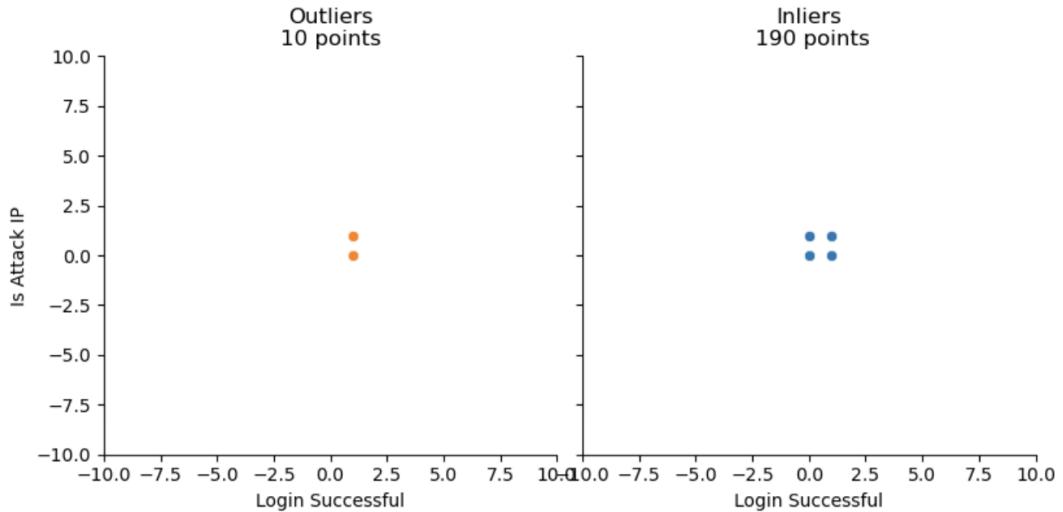


- anomaly\_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']

- contamination value == 0.1

```
Outlier Method: Isolation Forest
Number of anomalous values 10
Number of non anomalous values 190
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x292da5fd0>
```

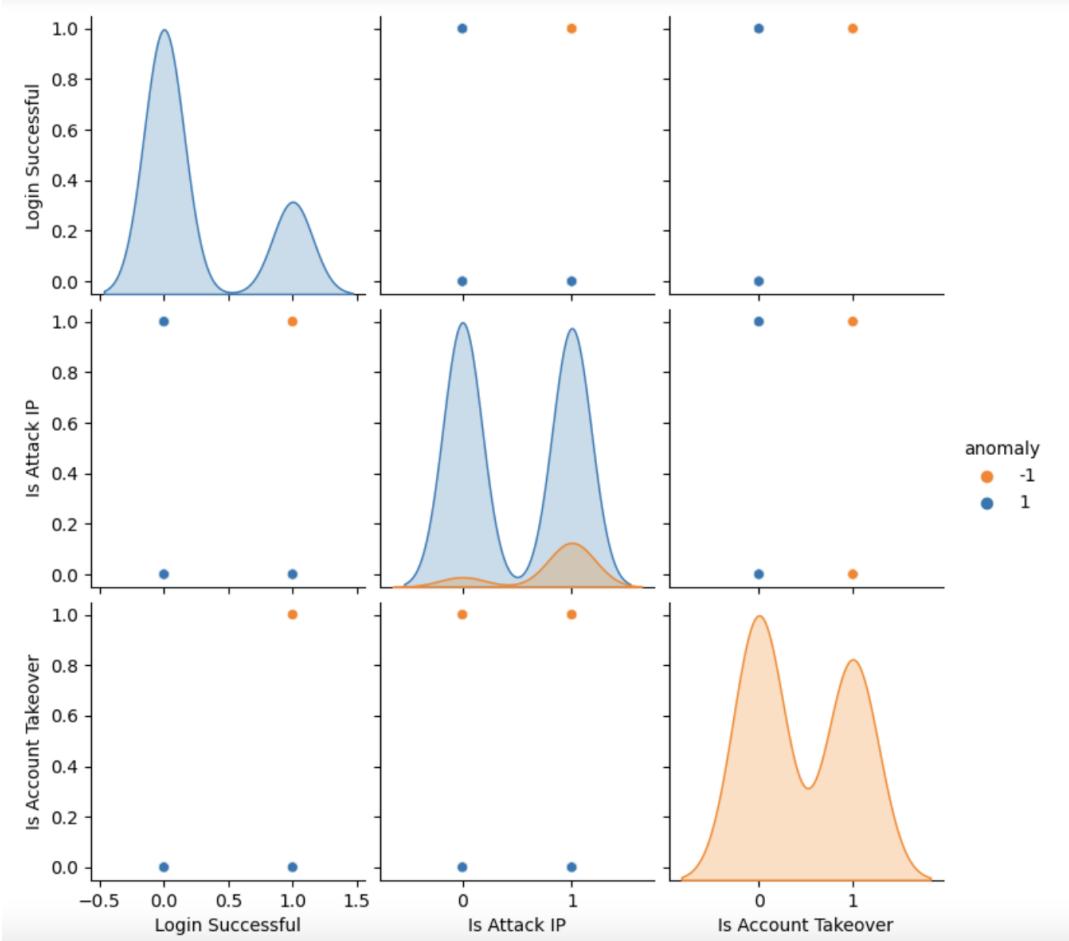
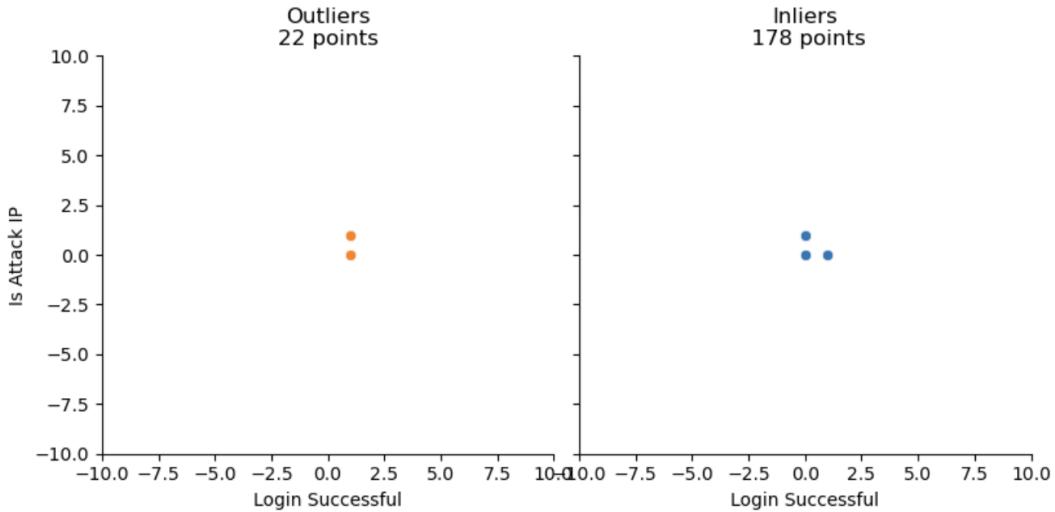
**Outlier Method: Isolation Forest**



- contamination value == 0.3

```
Outlier Method: Isolation Forest
Number of anomalous values 22
Number of non anomalous values 178
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x292ed1090>
```

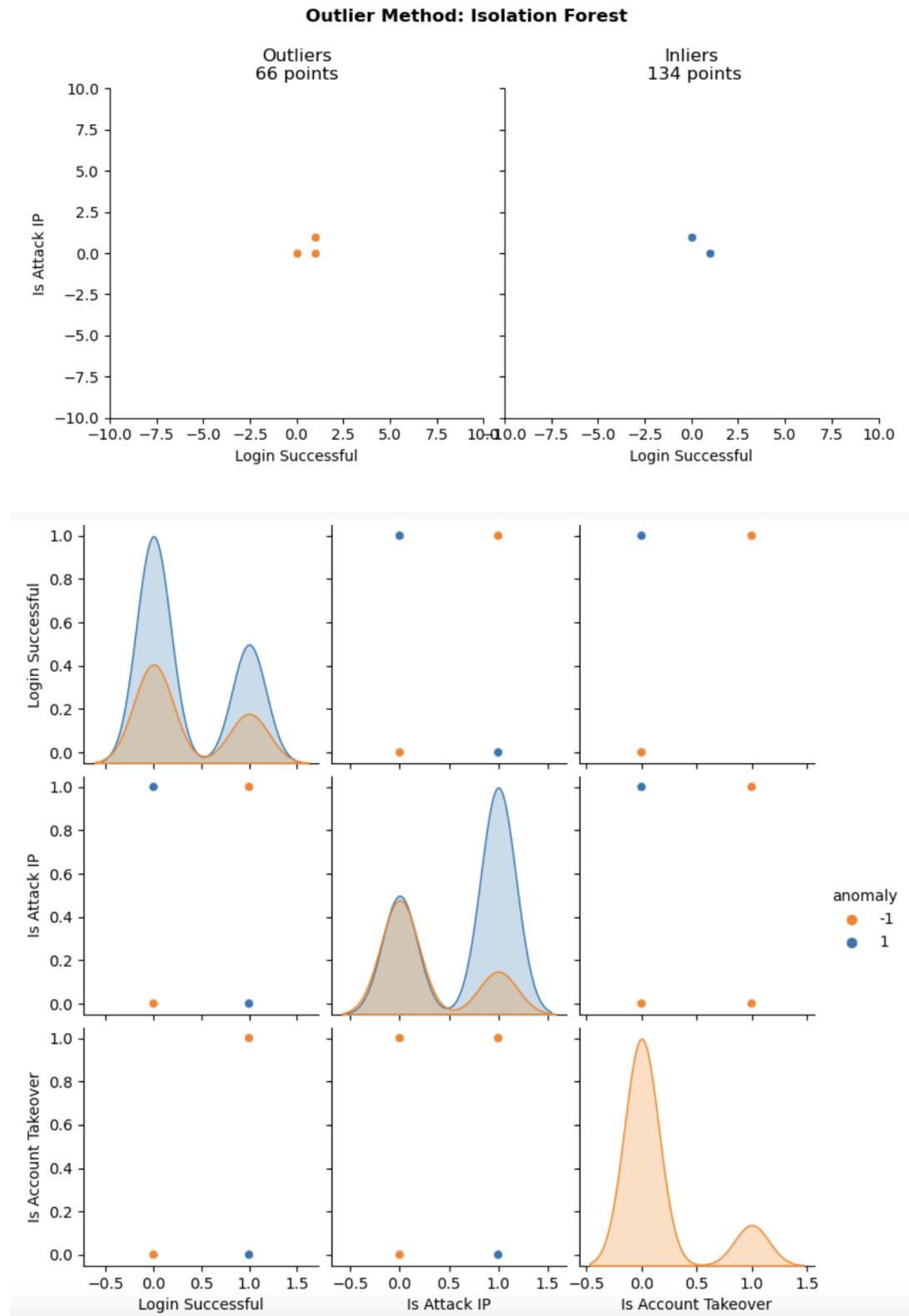
**Outlier Method: Isolation Forest**



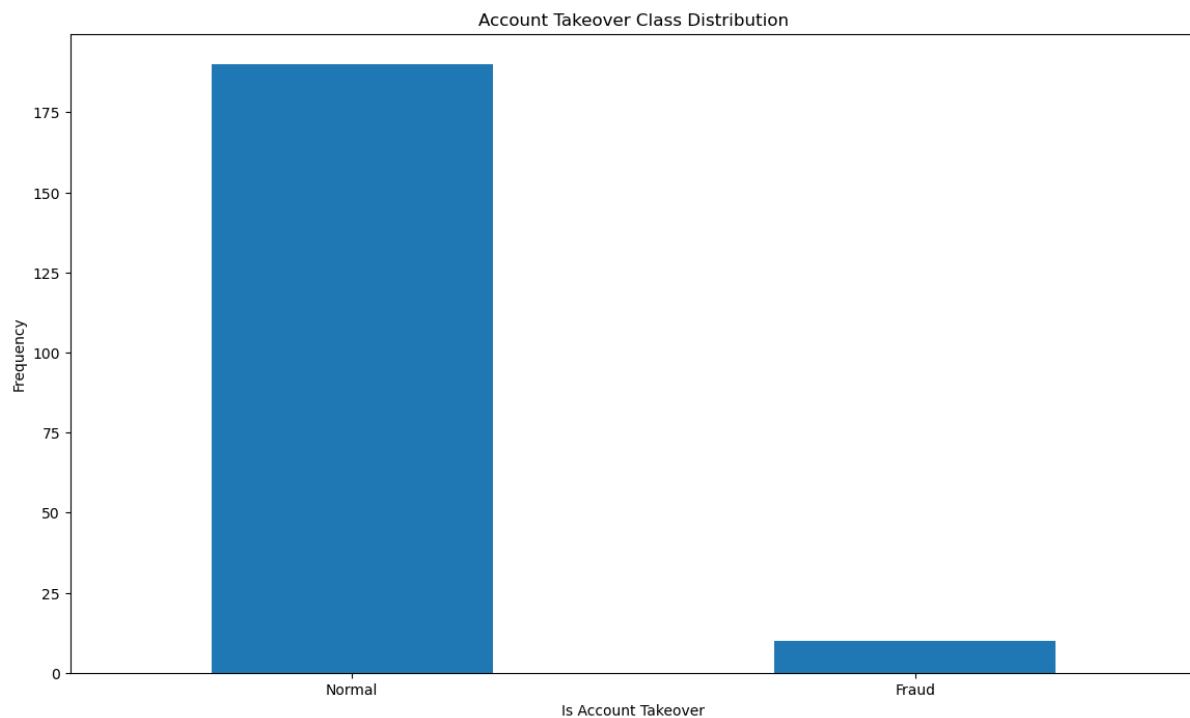
- contamination value == 0.5

```
Outlier Method: Isolation Forest
Number of anomalous values 66
Number of non anomalous values 134
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x2941271d0>
```



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



"{}: {}".format(clf\_name,n\_errors) → np. Isolation Forest: 21

Isolation Forest: 21

Accuracy Score :

0.895

Classification Report :

	precision	recall	f1-score	support
False	0.95	0.94	0.94	190
True	0.00	0.00	0.00	10
accuracy			0.90	200
macro avg	0.47	0.47	0.47	200
weighted avg	0.90	0.90	0.90	200

Local Outlier Factor: 17

Accuracy Score :

0.915

Classification Report :

	precision	recall	f1-score	support
False	0.96	0.95	0.96	190
True	0.18	0.20	0.19	10
accuracy			0.92	200
macro avg	0.57	0.58	0.57	200
weighted avg	0.92	0.92	0.92	200

Support Vector Machine: 152

Accuracy Score :

0.24

Classification Report :

	precision	recall	f1-score	support
False	0.90	0.23	0.36	190
True	0.03	0.50	0.06	10
accuracy			0.24	200
macro avg	0.46	0.36	0.21	200
weighted avg	0.85	0.24	0.35	200

### 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: 81.4 KB

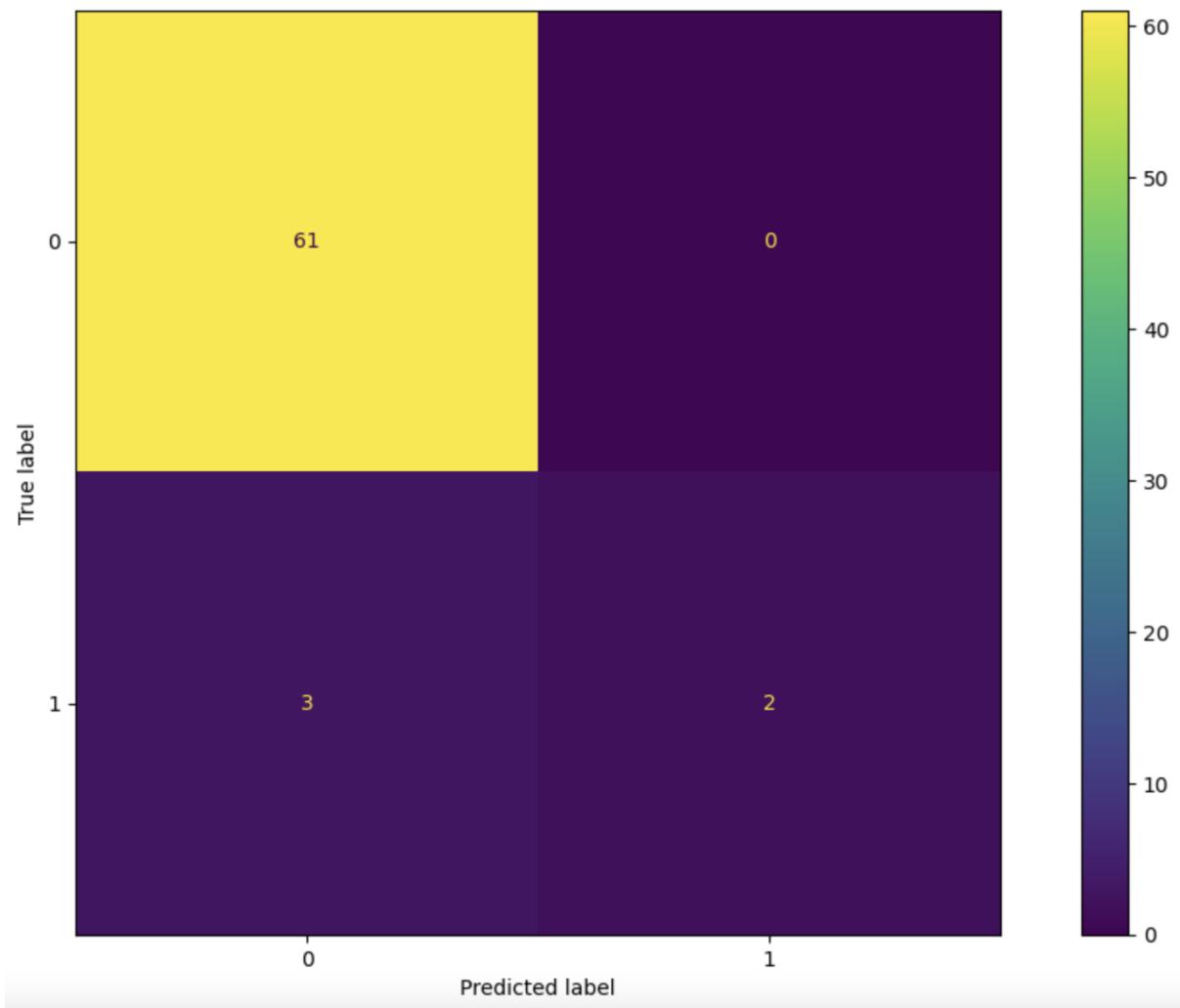
    predictor=None, random_state=None, ...)

Accuracy: 95.45%
          precision    recall   f1-score   support

      False        0.95     1.00      0.98      61
      True        1.00     0.40      0.57       5

accuracy                           0.95      66
macro avg        0.98     0.70      0.77      66
weighted avg     0.96     0.95      0.95      66

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 160
The number of records in the test dataset is 40
The training dataset has 153 records for the majority class and 7 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.16     0.28      37
      True         0.09     1.00     0.16       3
      accuracy          0.23      40
      macro avg       0.54     0.58     0.22      40
      weighted avg    0.93     0.23     0.27      40
      The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.93     1.00     0.96      37
      True         0.00     0.00     0.00       3
      accuracy          0.93      40
      macro avg       0.46     0.50     0.48      40
      weighted avg    0.86     0.93     0.89      40
[[ 6 31]
 [ 0  3]]

```

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

```

The number of records in the training dataset is 140
The number of records in the test dataset is 60
The training dataset has 133 records for the majority class and 7 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.02     0.03      57
      True         0.05     1.00     0.10       3
      accuracy          0.07      60
      macro avg       0.53     0.51     0.07      60
      weighted avg    0.95     0.07     0.04      60
      The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.95     1.00     0.97      57
      True         0.00     0.00     0.00       3
      accuracy          0.95      60
      macro avg       0.47     0.50     0.49      60
      weighted avg    0.90     0.95     0.93      60
[[ 1 56]
 [ 0  3]]

```

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

```

The number of records in the training dataset is 100
The number of records in the test dataset is 100
The training dataset has 95 records for the majority class and 5 records for the minority class.
      precision    recall   f1-score   support
False          0.00     0.00     0.00      95
True          0.05     1.00     0.10       5

accuracy                           0.05      100
macro avg                      0.03     0.50     0.05      100
weighted avg                     0.00     0.05     0.00      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False          0.95     1.00     0.97      95
True           0.00     0.00     0.00       5

accuracy                           0.95      100
macro avg                      0.47     0.50     0.49      100
weighted avg                     0.90     0.95     0.93      100

[[ 0 95]
 [ 0  5]]

```

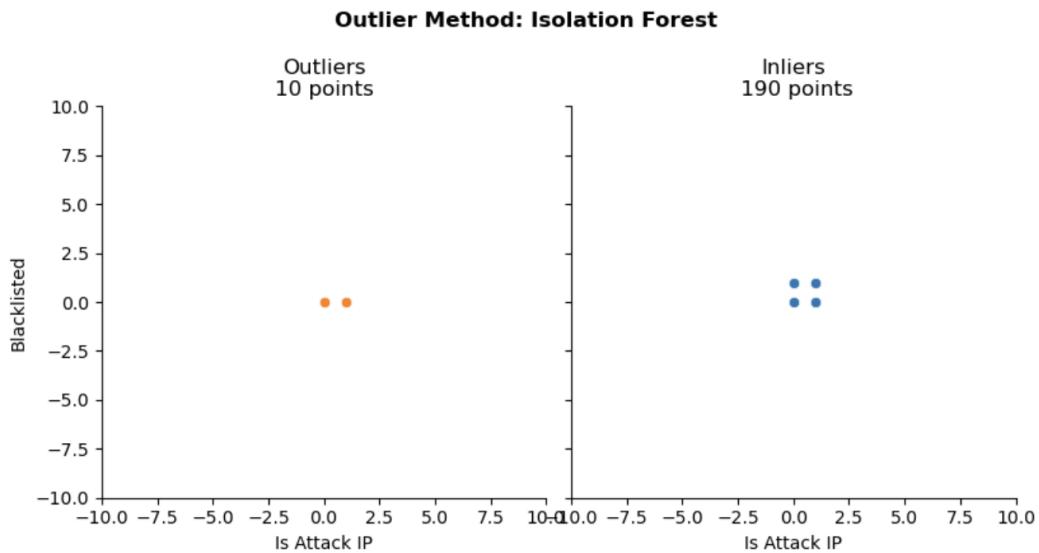
### 3C. Isolation Forest

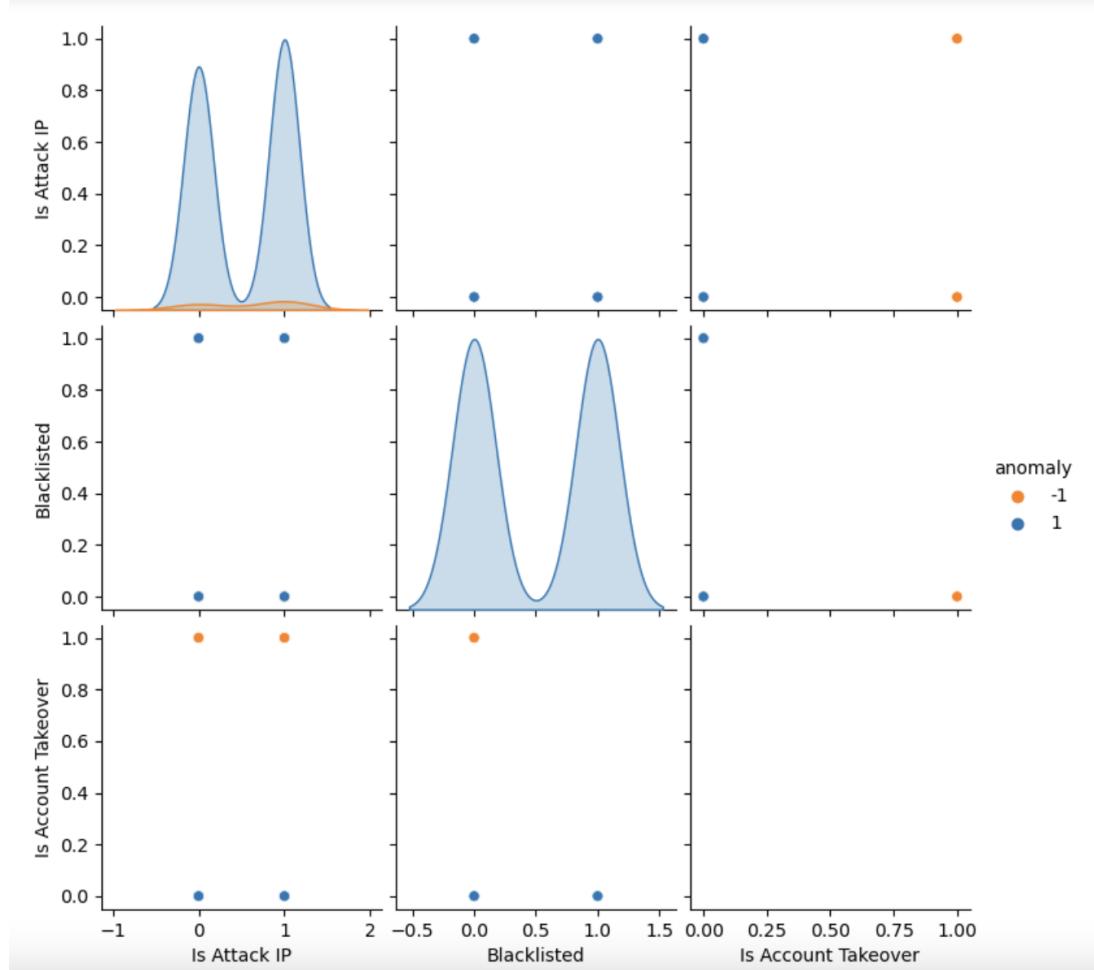
- anomaly\_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 10
Number of non anomalous values 190
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x2929d0190>

```

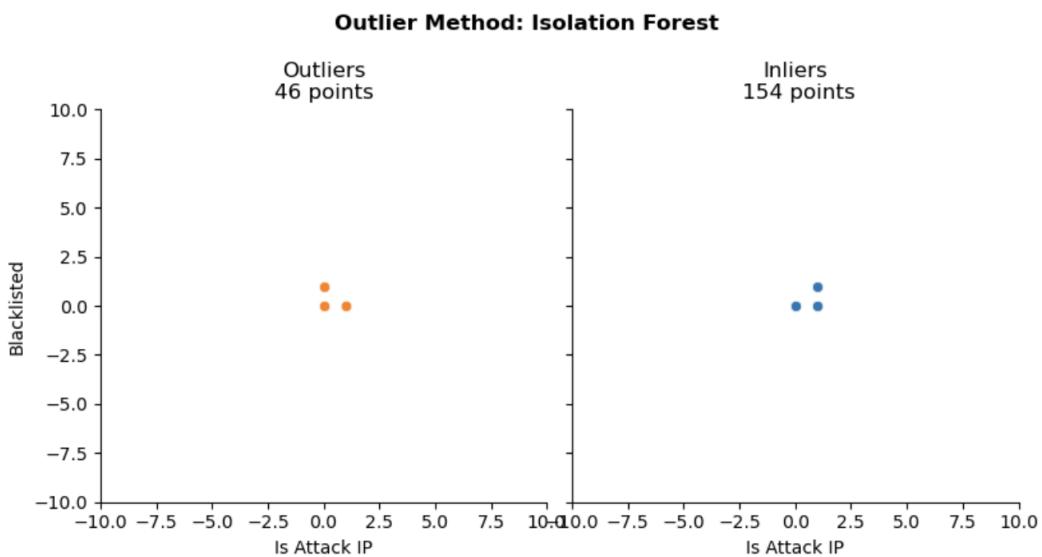


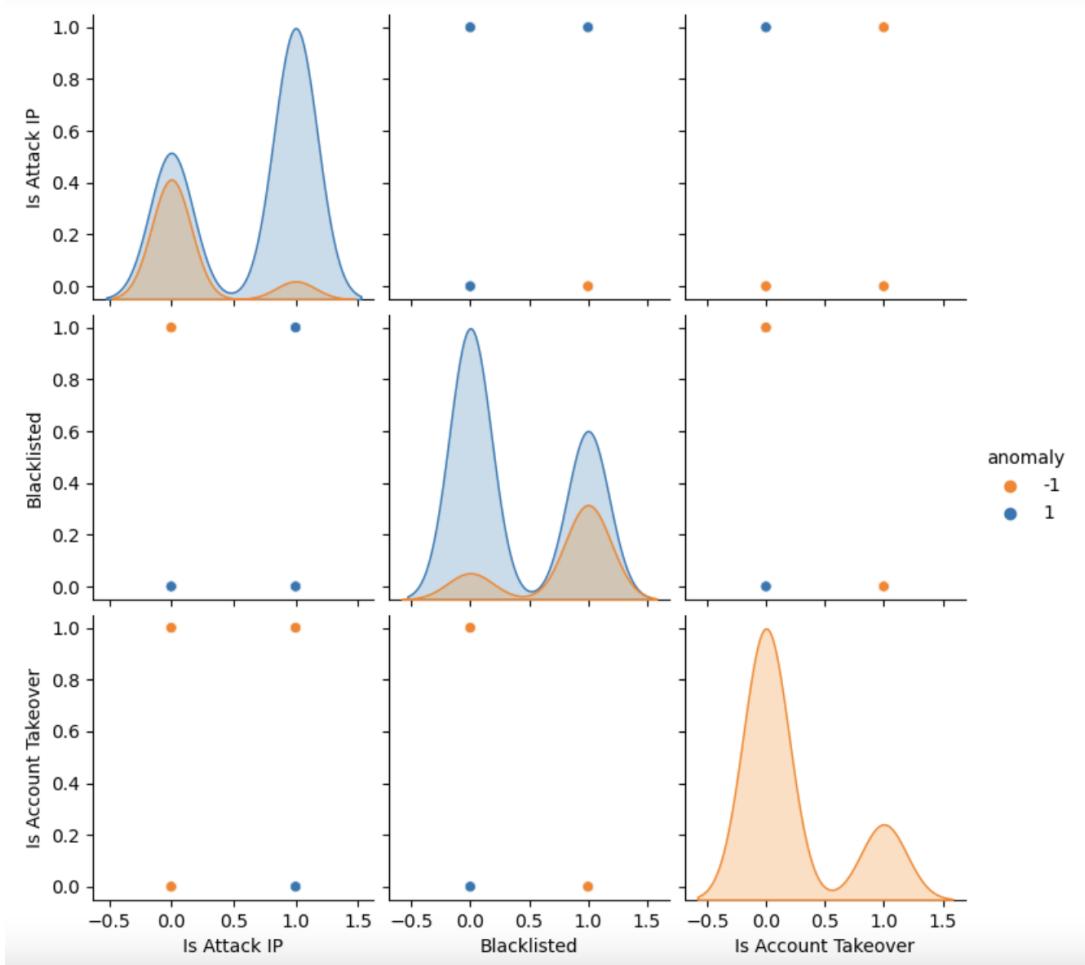


- contamination value == 0.3

```
Outlier Method: Isolation Forest
Number of anomalous values 46
Number of non anomalous values 154
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x2957e3310>
```

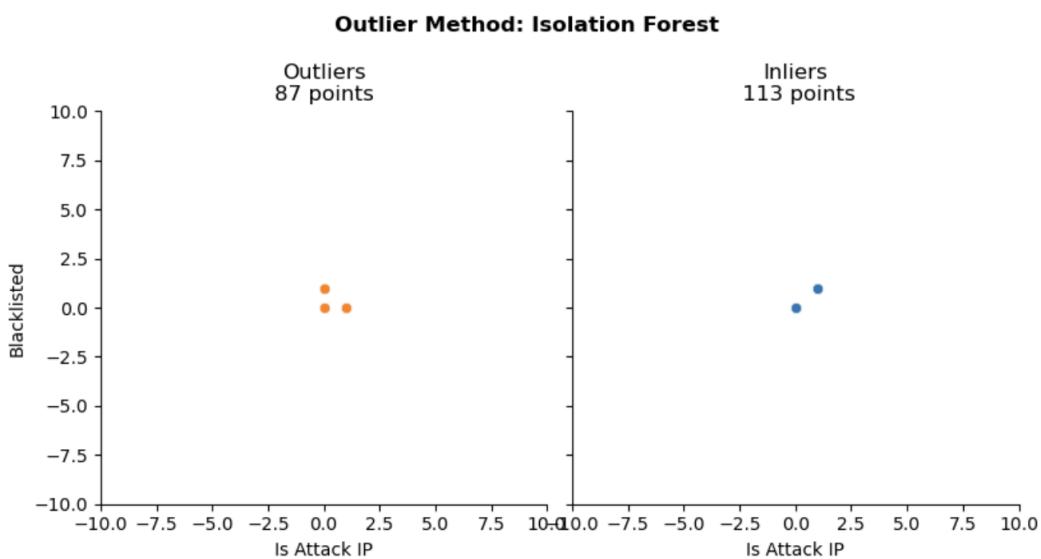


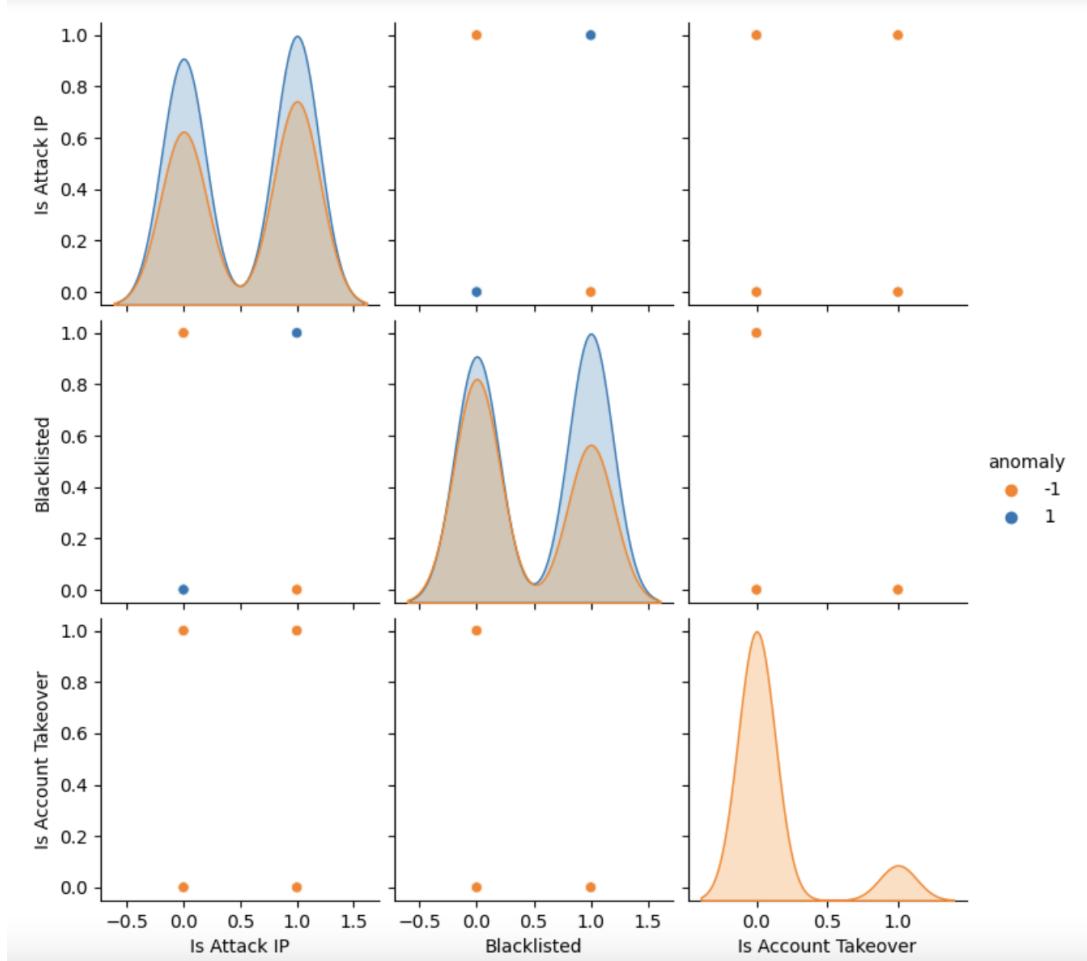


- contamination value == 0.5

```
Outlier Method: Isolation Forest
Number of anomalous values 87
Number of non anomalous values 113
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x2929aeed0>
```



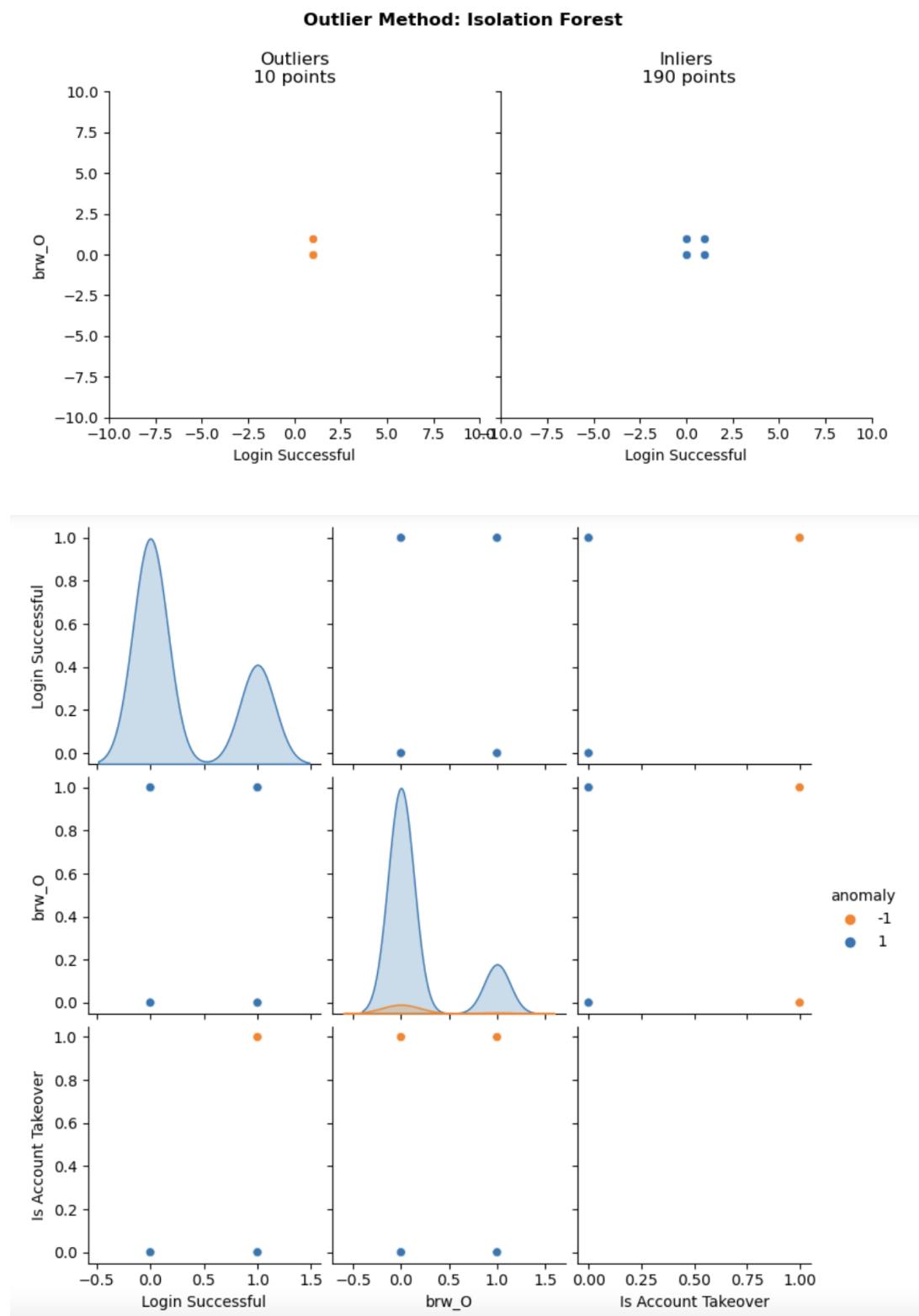


- anomaly\_inputs = ['Login Successful', 'Browser Type (brw\_O)', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 10
Number of non anomalous values 190
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x2b9da3dd0>

```



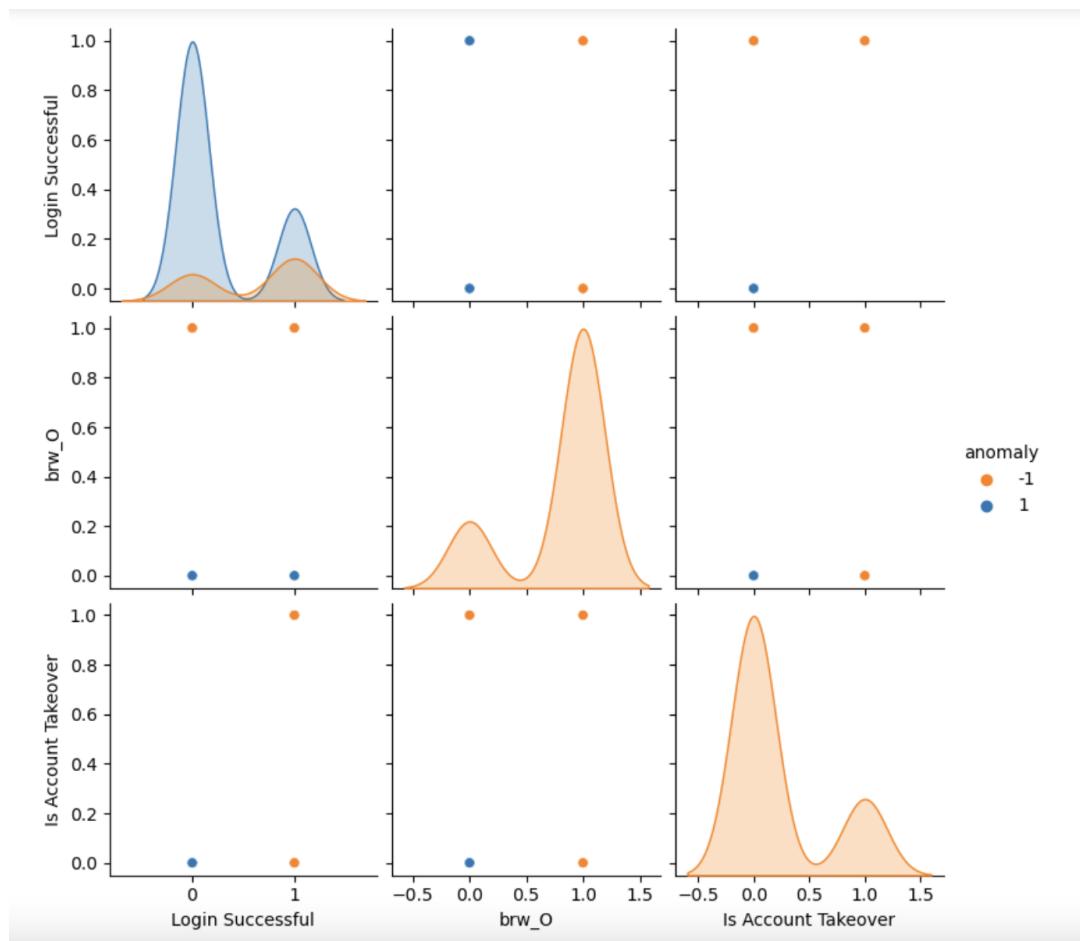
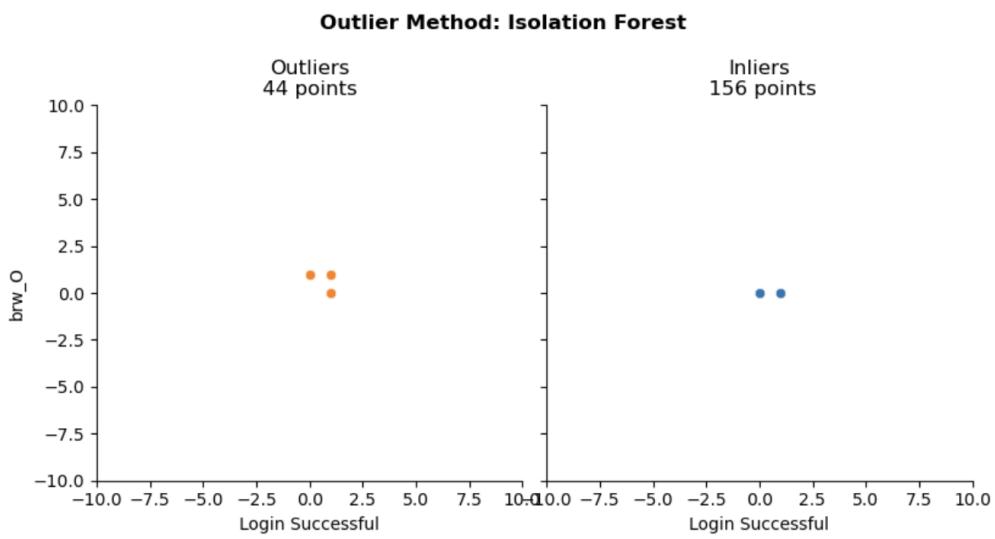
- contamination value == 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 44
Number of non anomalous values 156
Total Number of Values: 200

<seaborn.axisgrid.PairGrid at 0x2bbb1ba90>

```

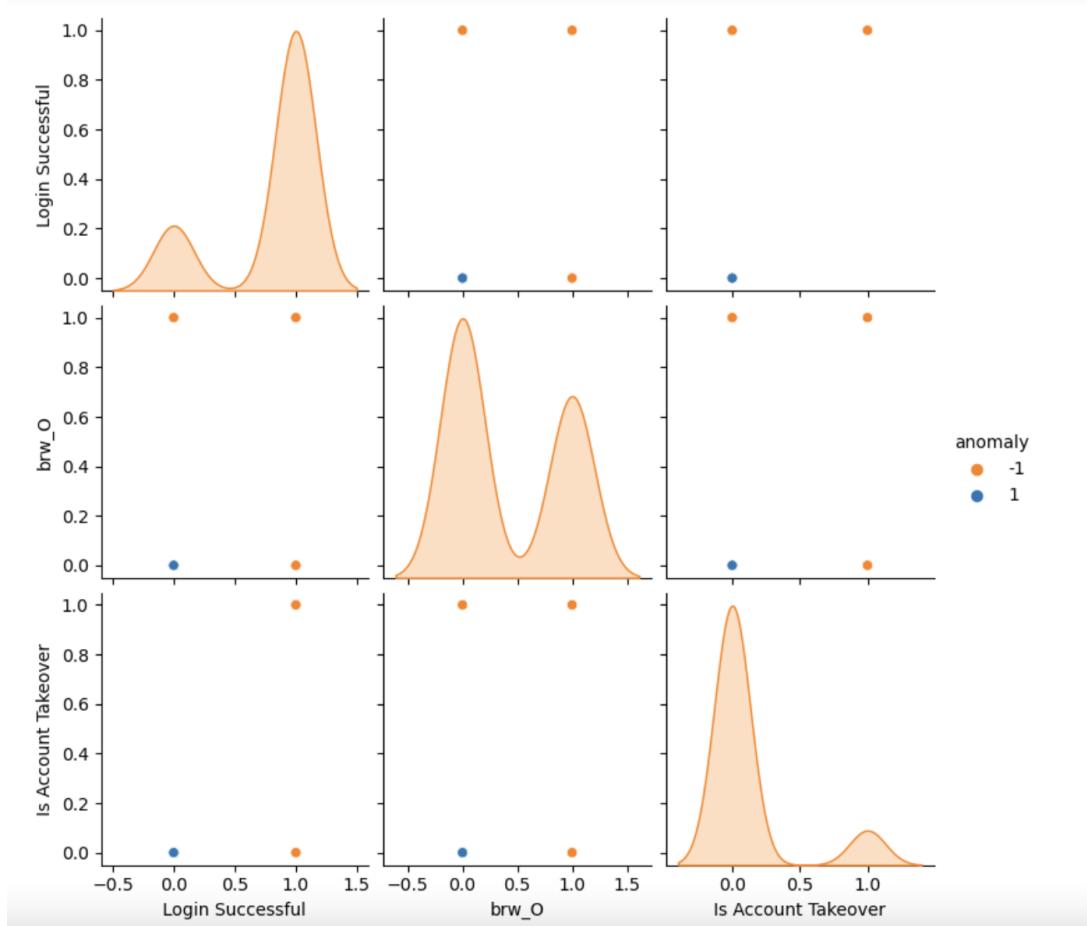
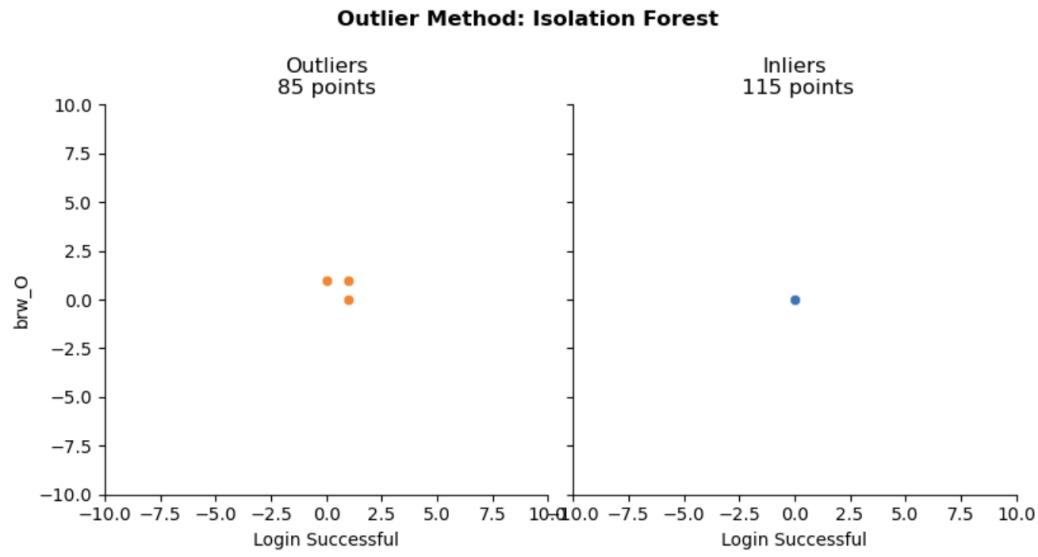


- contamination value == 0.5

```

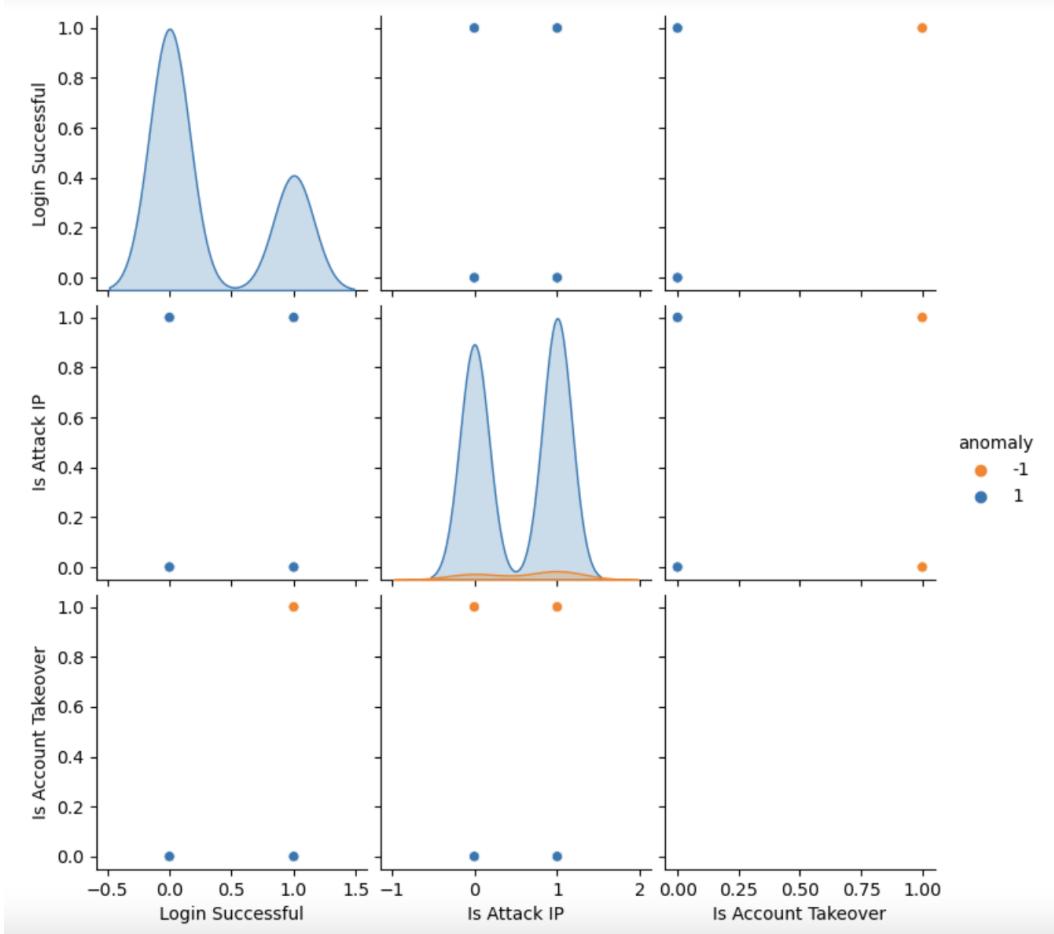
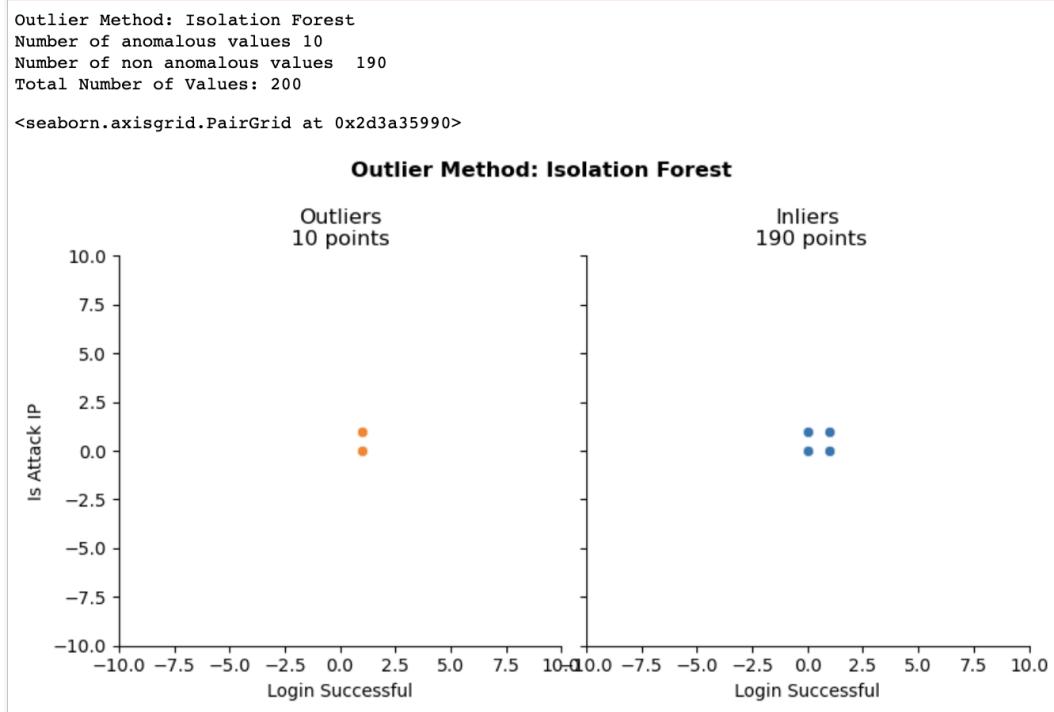
Outlier Method: Isolation Forest
Number of anomalous values 85
Number of non anomalous values 115
Total Number of Values: 200
: <seaborn.axisgrid.PairGrid at 0x2bc18a450>

```



- anomaly\_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']

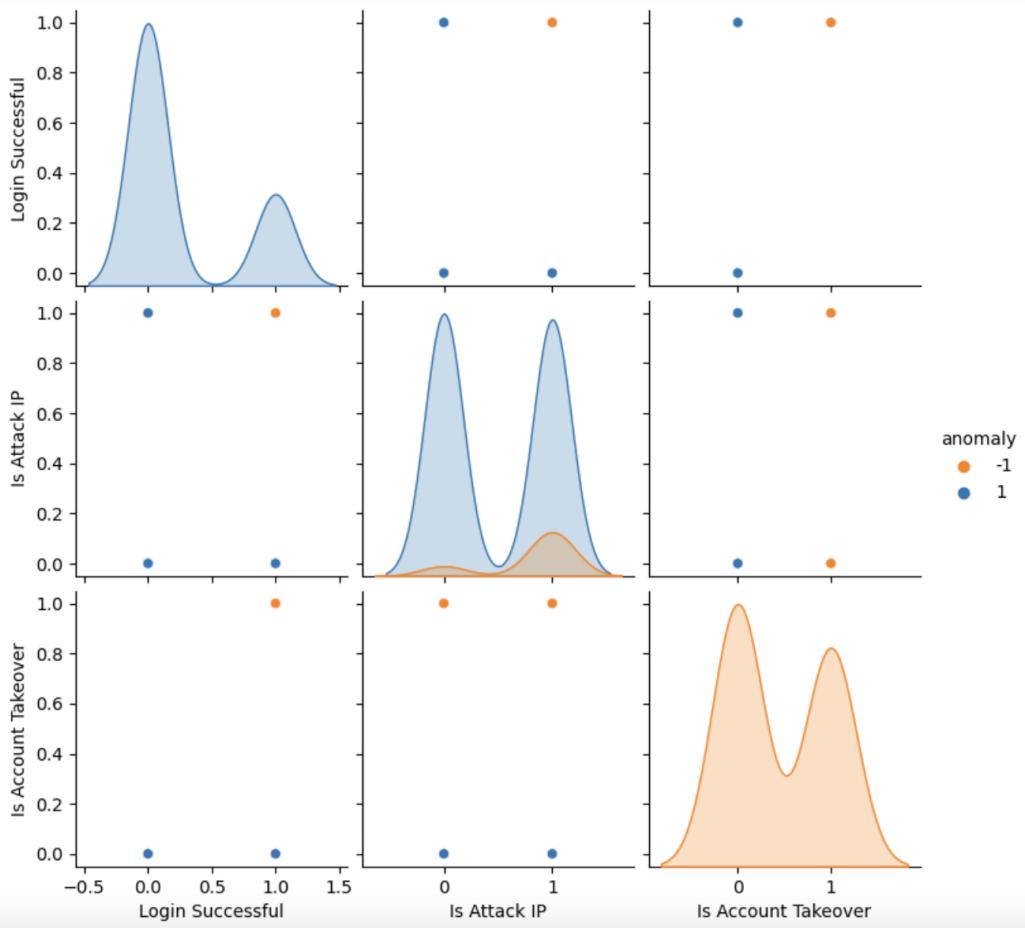
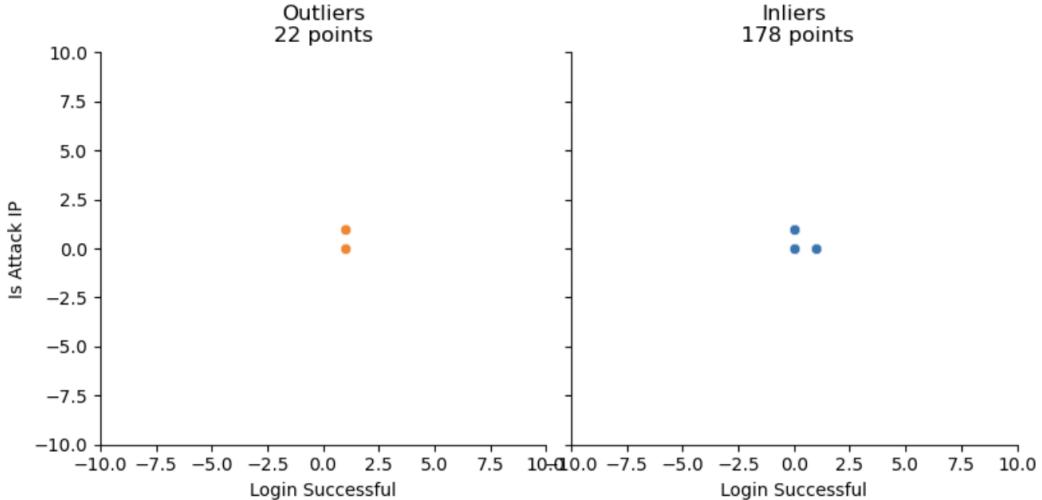
- contamination value == 0.1



- contamination value == 0.3

```
Outlier Method: Isolation Forest
Number of anomalous values 22
Number of non anomalous values 178
Total Number of Values: 200
: <seaborn.axisgrid.PairGrid at 0x2d431d1d0>
```

**Outlier Method: Isolation Forest**



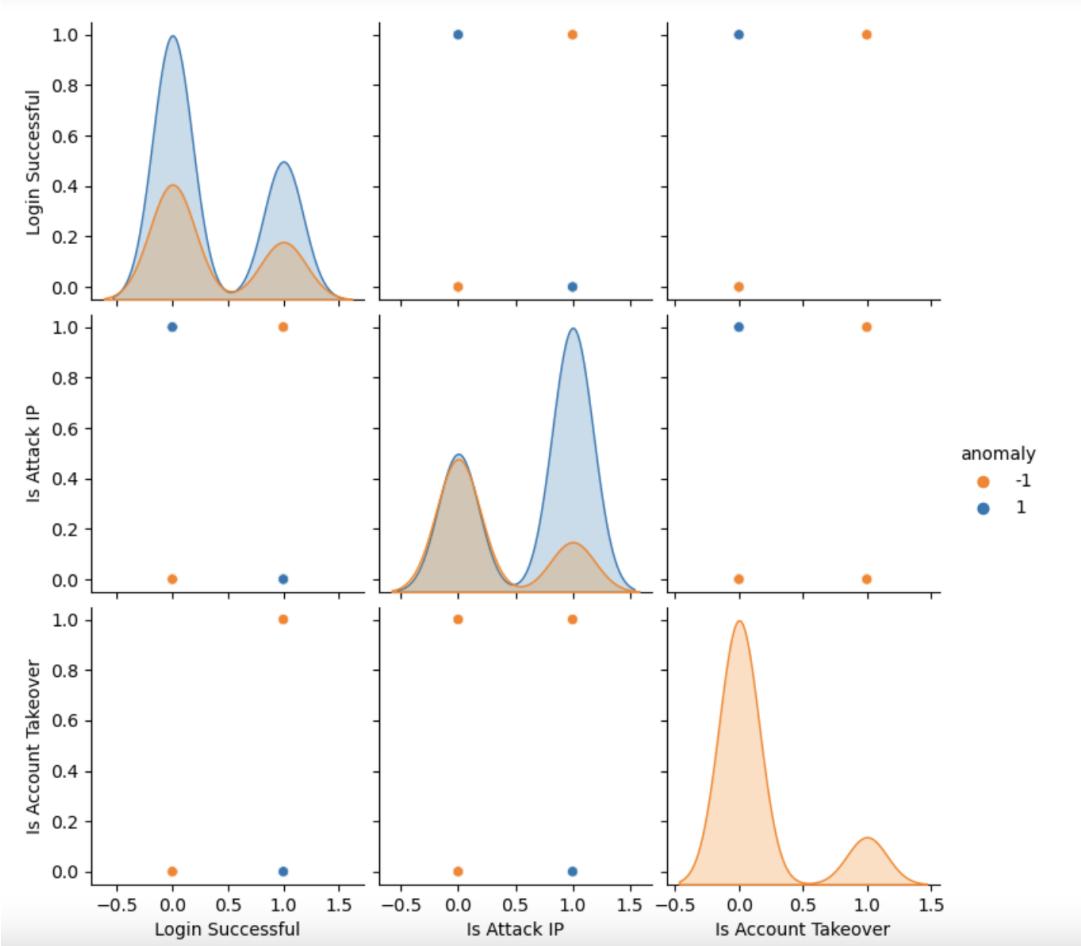
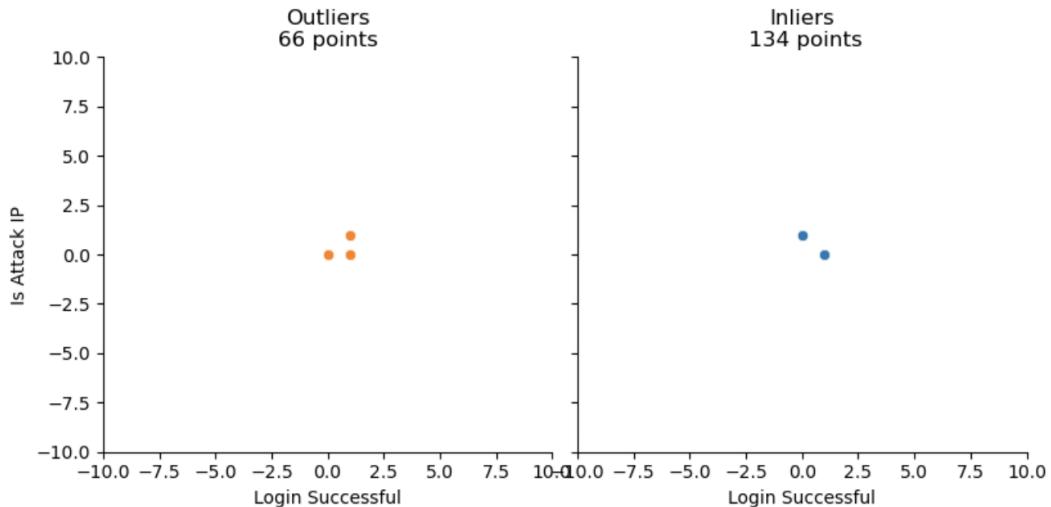
- contamination value == 0.5

```

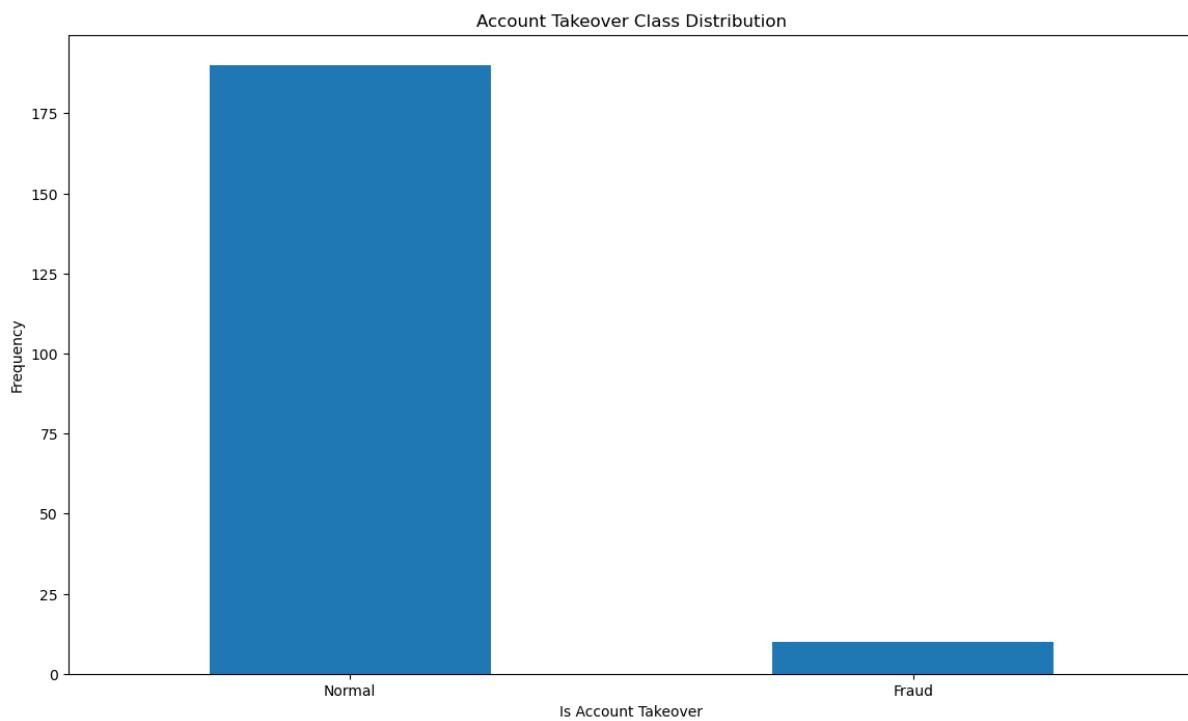
Outlier Method: Isolation Forest
Number of anomalous values 66
Number of non anomalous values 134
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x2d4aca7d0>

```

**Outlier Method: Isolation Forest**



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



**Isolation Forest:** 21

**Accuracy Score :**

0.895

**Classification Report :**

	precision	recall	f1-score	support
<b>False</b>	0.95	0.94	0.94	190
<b>True</b>	0.00	0.00	0.00	10
<b>accuracy</b>			0.90	200
<b>macro avg</b>	0.47	0.47	0.47	200
<b>weighted avg</b>	0.90	0.90	0.90	200

---

```
Local Outlier Factor: 17
```

```
Accuracy Score :
```

```
0.915
```

```
Classification Report :
```

	precision	recall	f1-score	support
False	0.96	0.95	0.96	190
True	0.18	0.20	0.19	10
accuracy			0.92	200
macro avg	0.57	0.58	0.57	200
weighted avg	0.92	0.92	0.92	200

```
Support Vector Machine: 152
```

```
Accuracy Score :
```

```
0.24
```

```
Classification Report :
```

	precision	recall	f1-score	support
False	0.90	0.23	0.36	190
True	0.03	0.50	0.06	10
accuracy			0.24	200
macro avg	0.46	0.36	0.21	200
weighted avg	0.85	0.24	0.35	200

## II. 30/200 → 15%

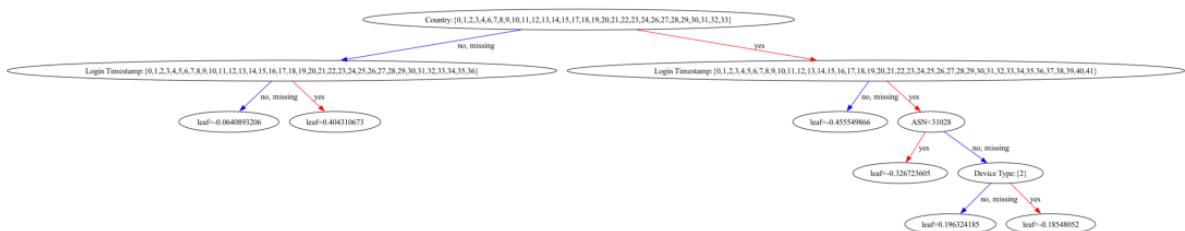
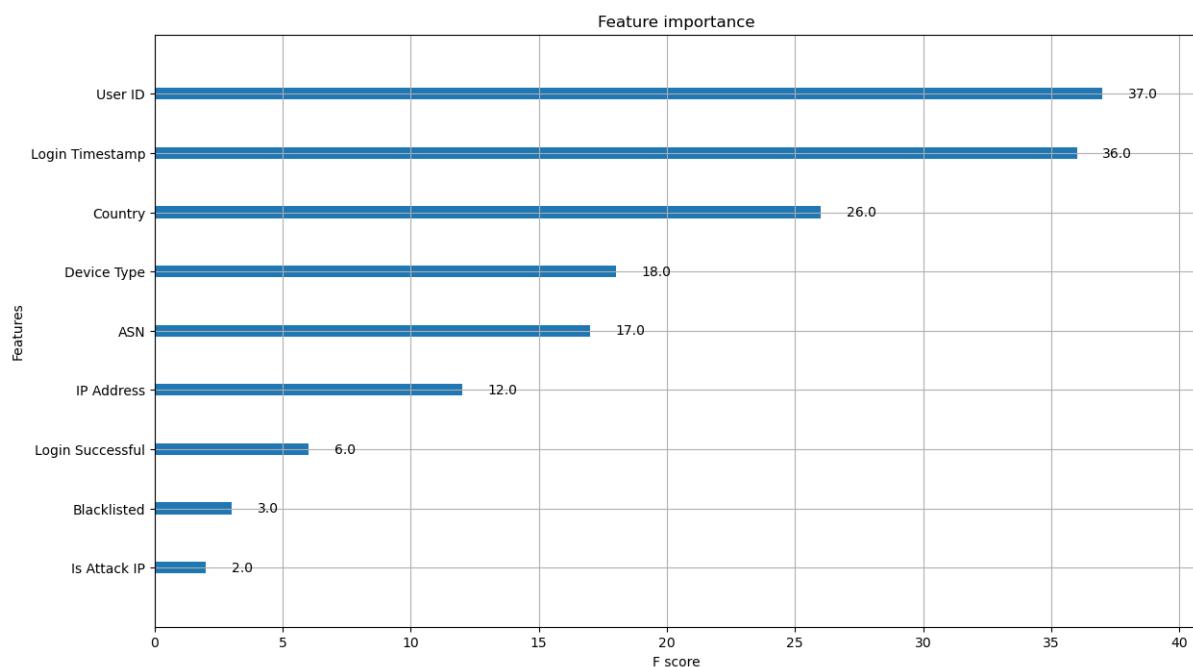
### 1. XGB Experimental

```
Features importances:
```

```
memory usage: 22.1 KB

array([0.04656677, 0.4416981 , 0.02399971, 0.01092903, 0.08391444,
       0.11133045, 0.21102063, 0.01387566, 0.05666523, 0.        ],
      dtype=float32)
```

0	ASN	200	non-null	float64
1	Country	200	non-null	category
2	Device Type	200	non-null	category
3	IP Address	200	non-null	category
4	Is Attack IP	200	non-null	bool
5	Login Successful	200	non-null	bool
6	Login Timestamp	200	non-null	category
7	User ID	200	non-null	float64
8	Blacklisted	200	non-null	bool
9	Browser Type	200	non-null	category
10	Is Account Takeover	200	non-null	bool

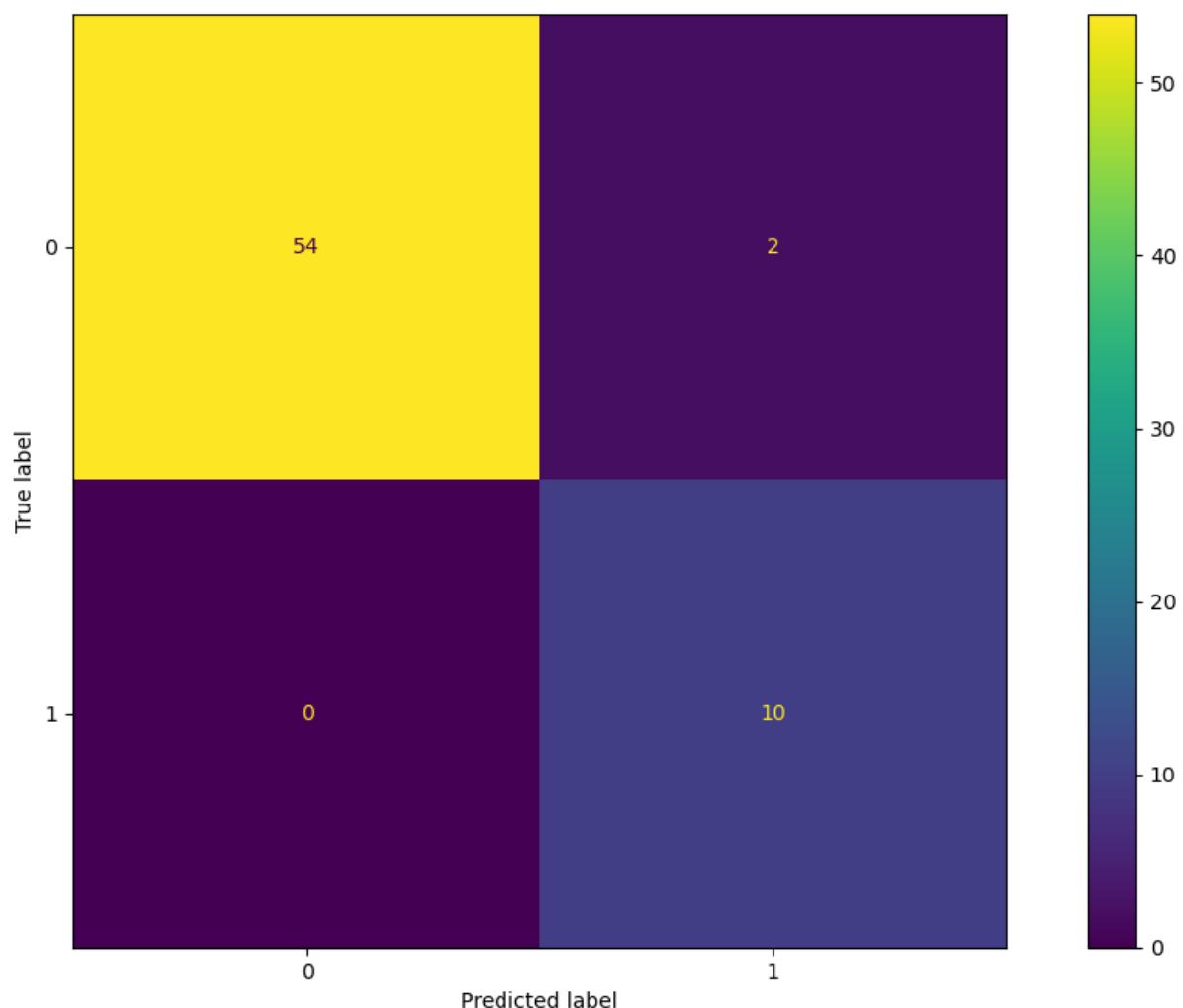


## 2. Label encoding (cat.codes)

### 2A. XGBoost

memory usage: 5.4 KB

```
Precision: None, RandomState=None, ...,
Accuracy: 96.97%
      precision    recall   f1-score   support
      False        1.00     0.96     0.98      56
      True        0.83     1.00     0.91      10
      accuracy           0.97
      macro avg       0.92     0.98     0.95      66
      weighted avg    0.97     0.97     0.97      66
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 160
The number of records in the test dataset is 40
The training dataset has 131 records for the majority class and 29 records for the minority class.

precision    recall   f1-score   support
False        0.00     0.00     0.00      39
True         0.03     1.00     0.05      1

accuracy          0.03      40
macro avg       0.01     0.50     0.02      40
weighted avg    0.00     0.03     0.00      40

The customized score threshold for 2% of outliers is 0.00
precision    recall   f1-score   support
False        0.97     1.00     0.99      39
True         0.00     0.00     0.00      1

accuracy          0.97      40
macro avg       0.49     0.50     0.49      40
weighted avg    0.95     0.97     0.96      40

[[ 0 39]
 [ 0  1]]
```

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

```
The number of records in the training dataset is 140
The number of records in the test dataset is 60
The training dataset has 116 records for the majority class and 24 records for the minority class.

precision    recall   f1-score   support
False        0.00     0.00     0.00      54
True         0.10     1.00     0.18       6

accuracy          0.10      60
macro avg       0.05     0.50     0.09      60
weighted avg    0.01     0.10     0.02      60

The customized score threshold for 2% of outliers is 0.00
precision    recall   f1-score   support
False        0.90     1.00     0.95      54
True         0.00     0.00     0.00       6

accuracy          0.90      60
macro avg       0.45     0.50     0.47      60
weighted avg    0.81     0.90     0.85      60

[[ 0 54]
 [ 0  6]]
```

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

```

The number of records in the training dataset is 100
The number of records in the test dataset is 100
The training dataset has 82 records for the majority class and 18 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00      88
  True        0.12     1.00     0.21      12

accuracy                           0.12      100
macro avg       0.06     0.50     0.11      100
weighted avg    0.01     0.12     0.03      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.88     1.00     0.94      88
  True        0.00     0.00     0.00      12

accuracy                           0.88      100
macro avg       0.44     0.50     0.47      100
weighted avg    0.77     0.88     0.82      100

[[ 0 88]
 [ 0 12]]

```

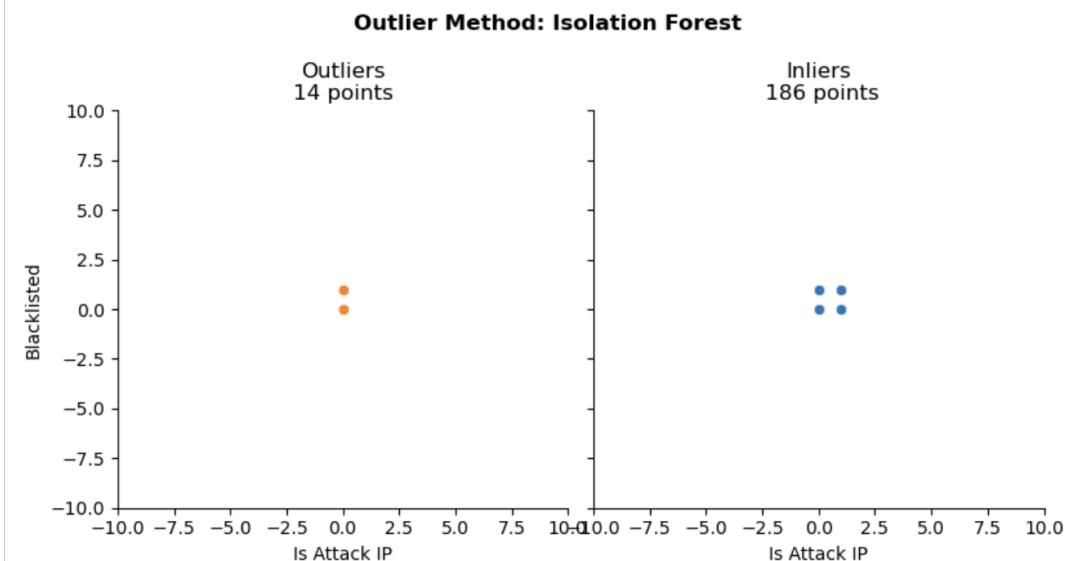
## 2C. Isolation Forest

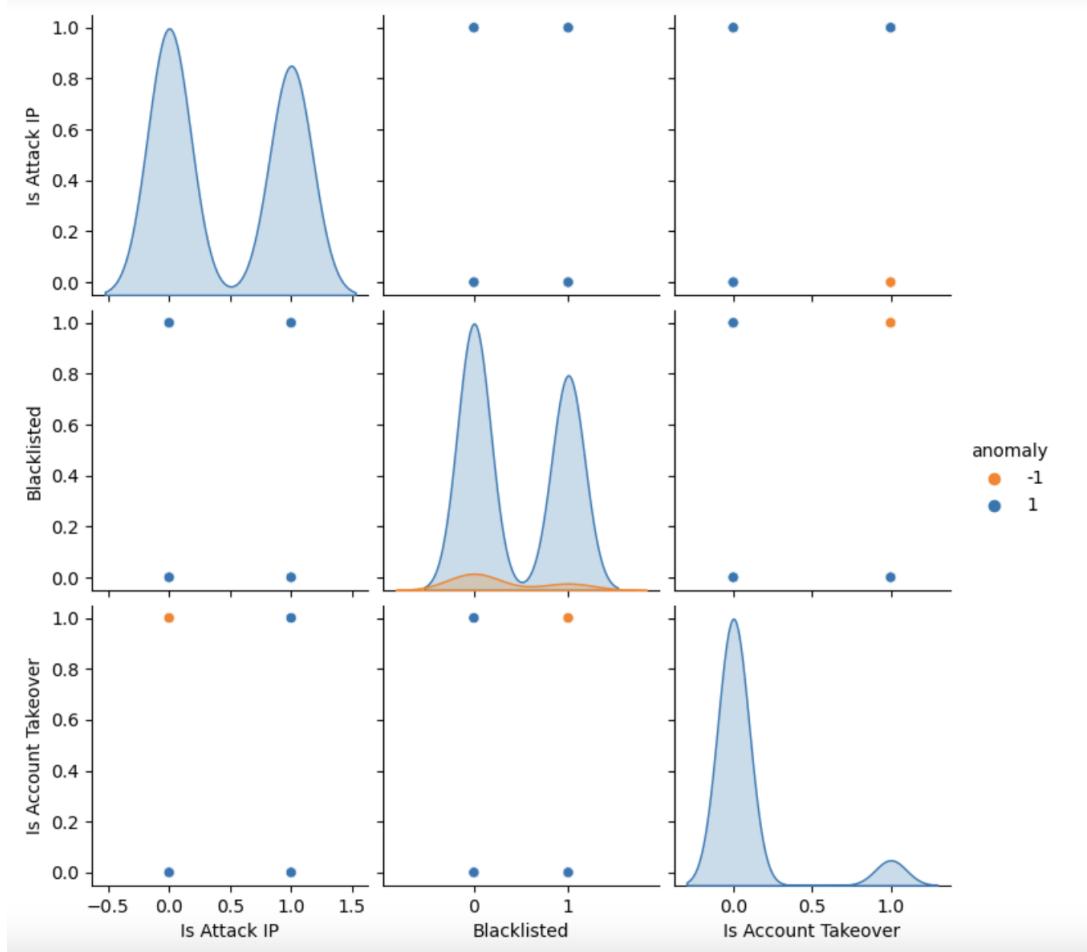
- anomaly\_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 58
Number of non anomalous values 142
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x2d933f050>

```



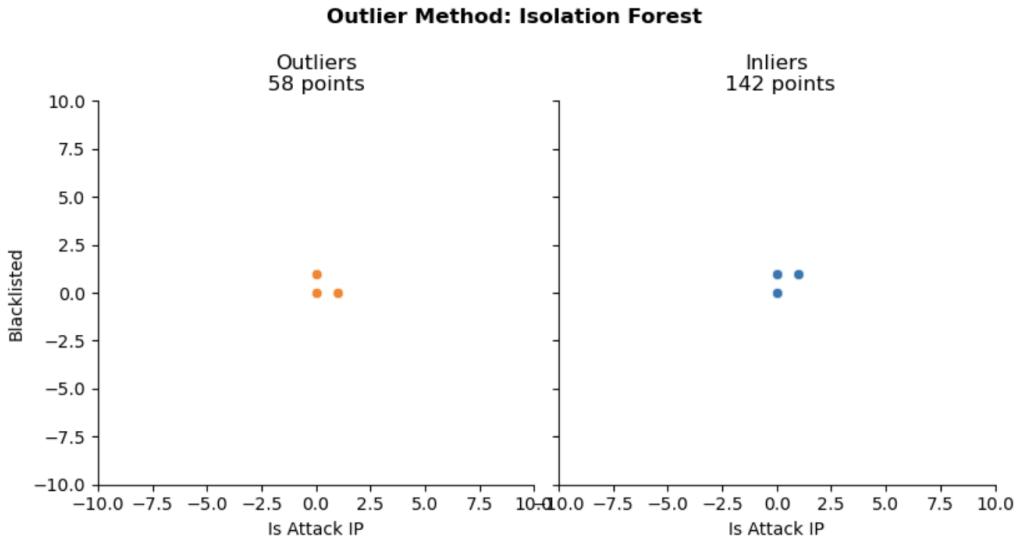


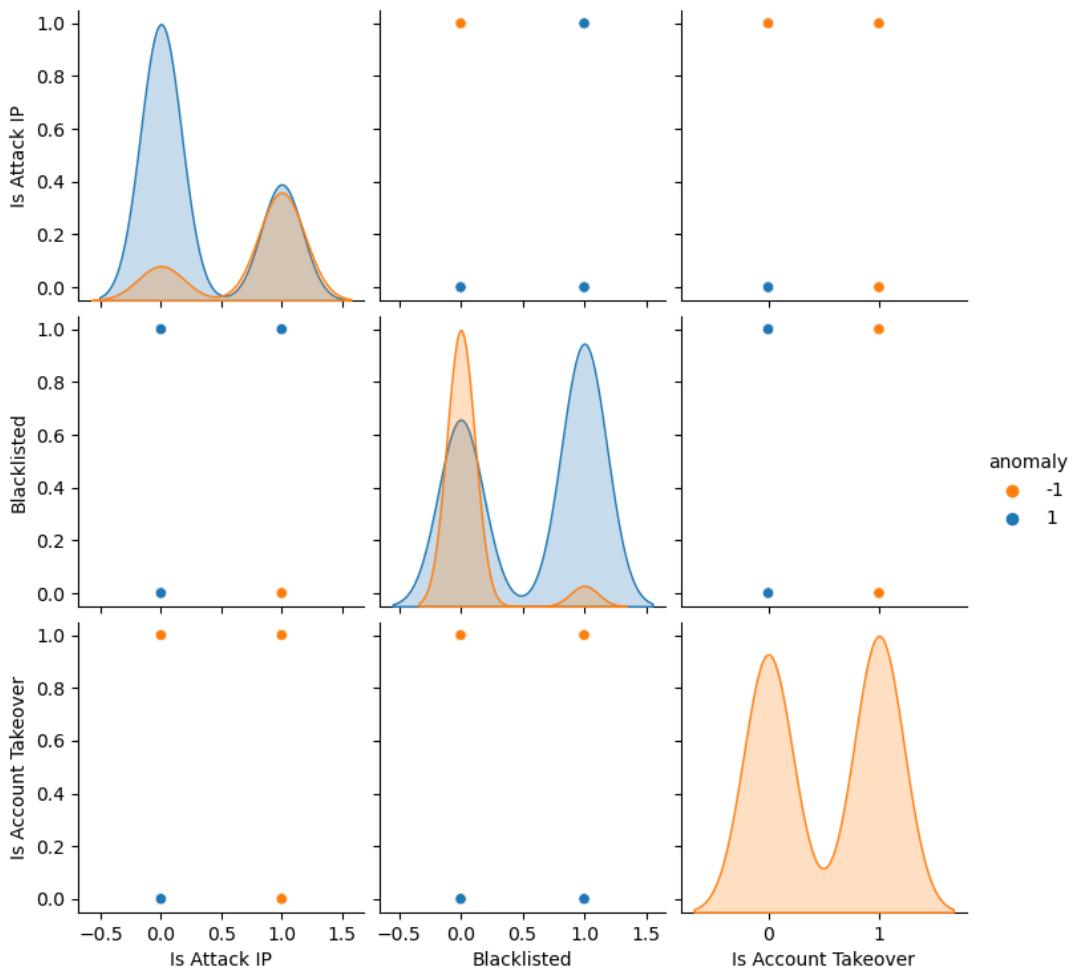
- increasing contamination value to 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 58
Number of non anomalous values 142
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x177b2c410>

```



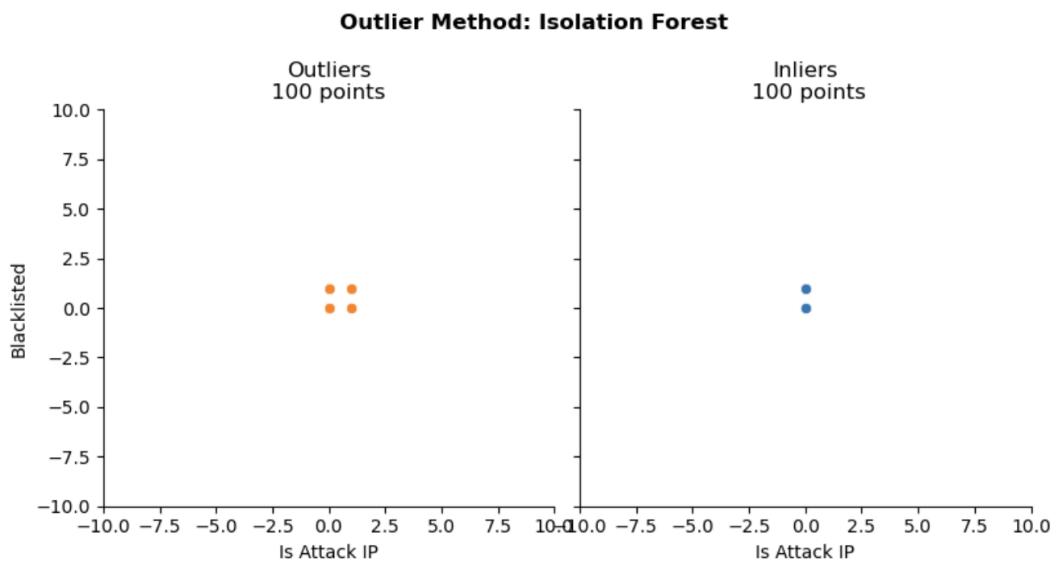


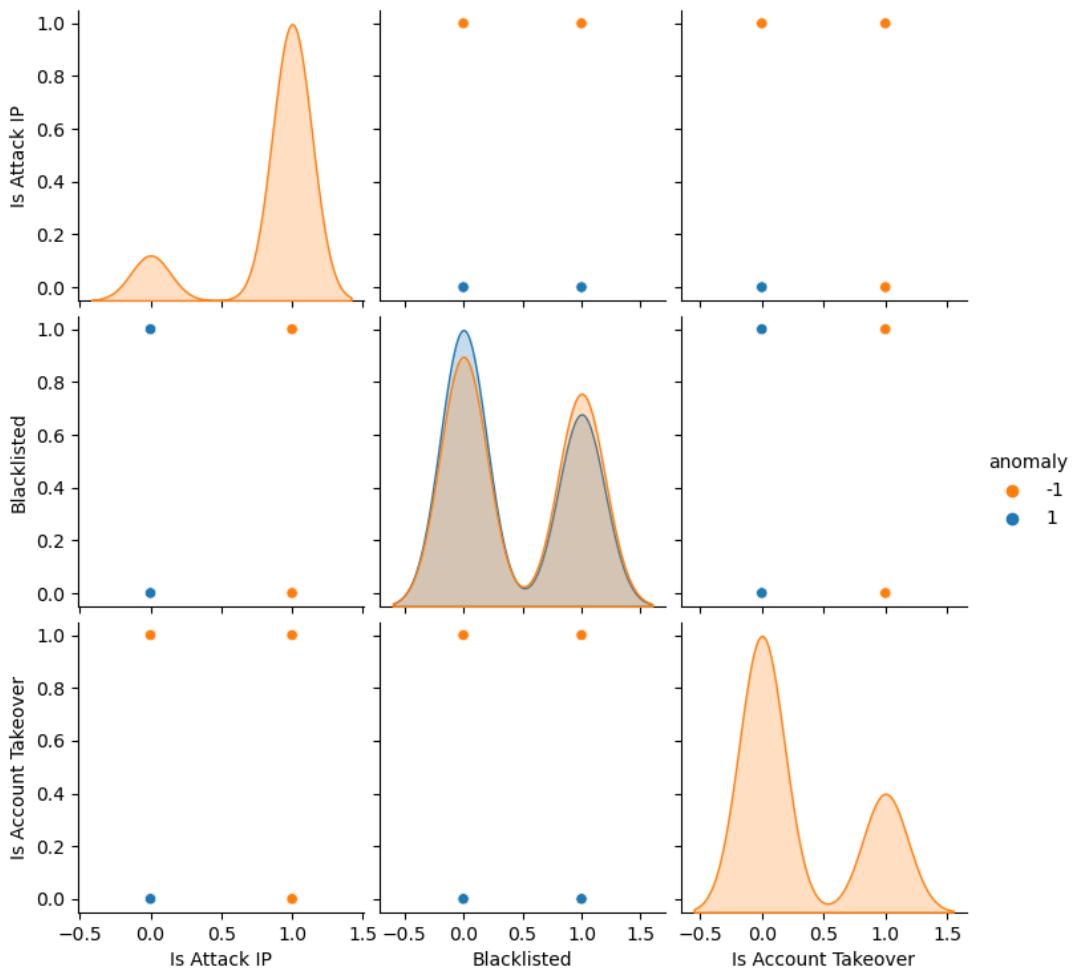
- contamination value == 0.5

```

Outlier Method: Isolation Forest
Number of anomalous values 100
Number of non anomalous values 100
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x2da203790>

```





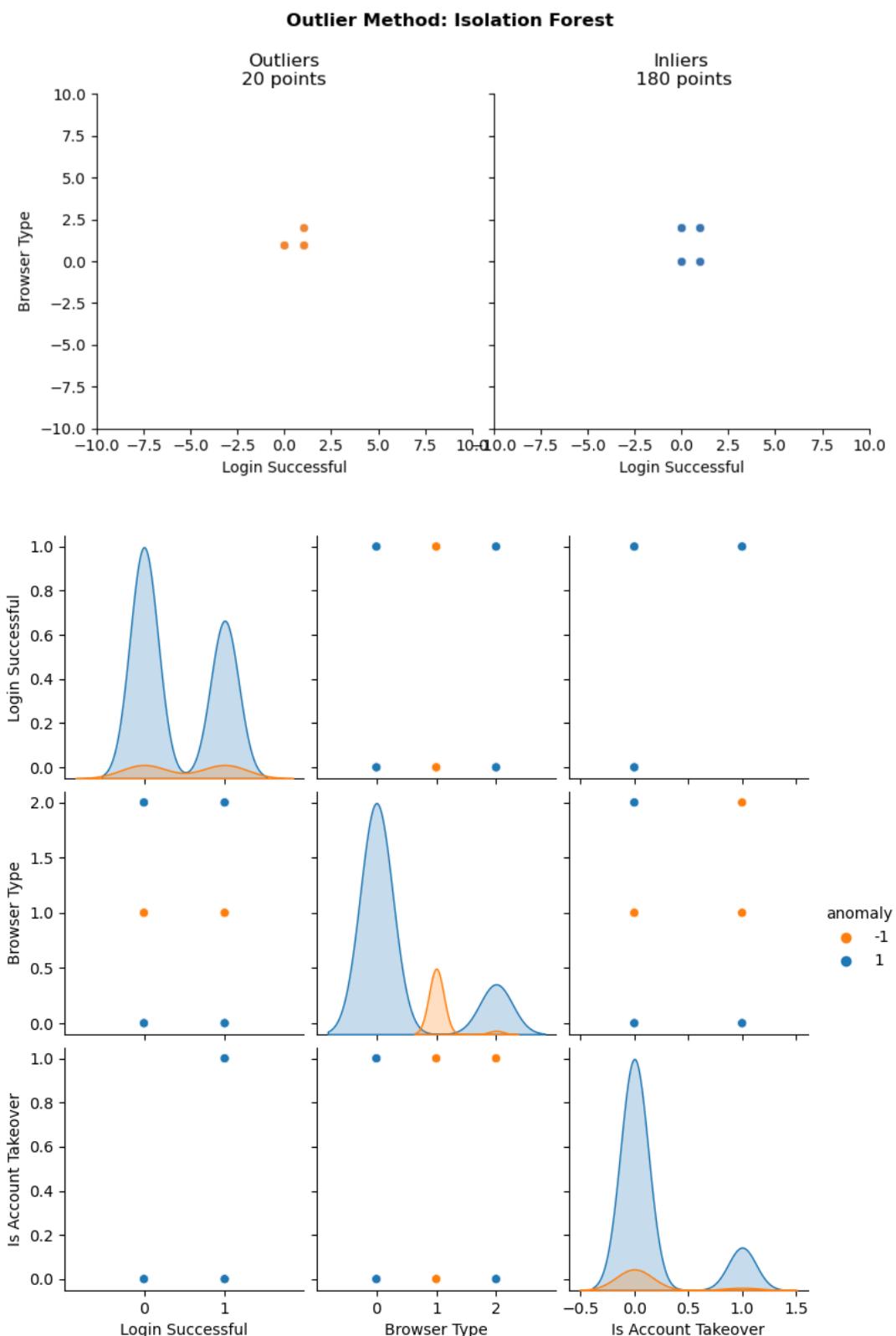
- anomaly\_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 20
Number of non anomalous values 180
Total Number of Values: 200

```

<seaborn.axisgrid.PairGrid at 0x2e6b8b9d0>

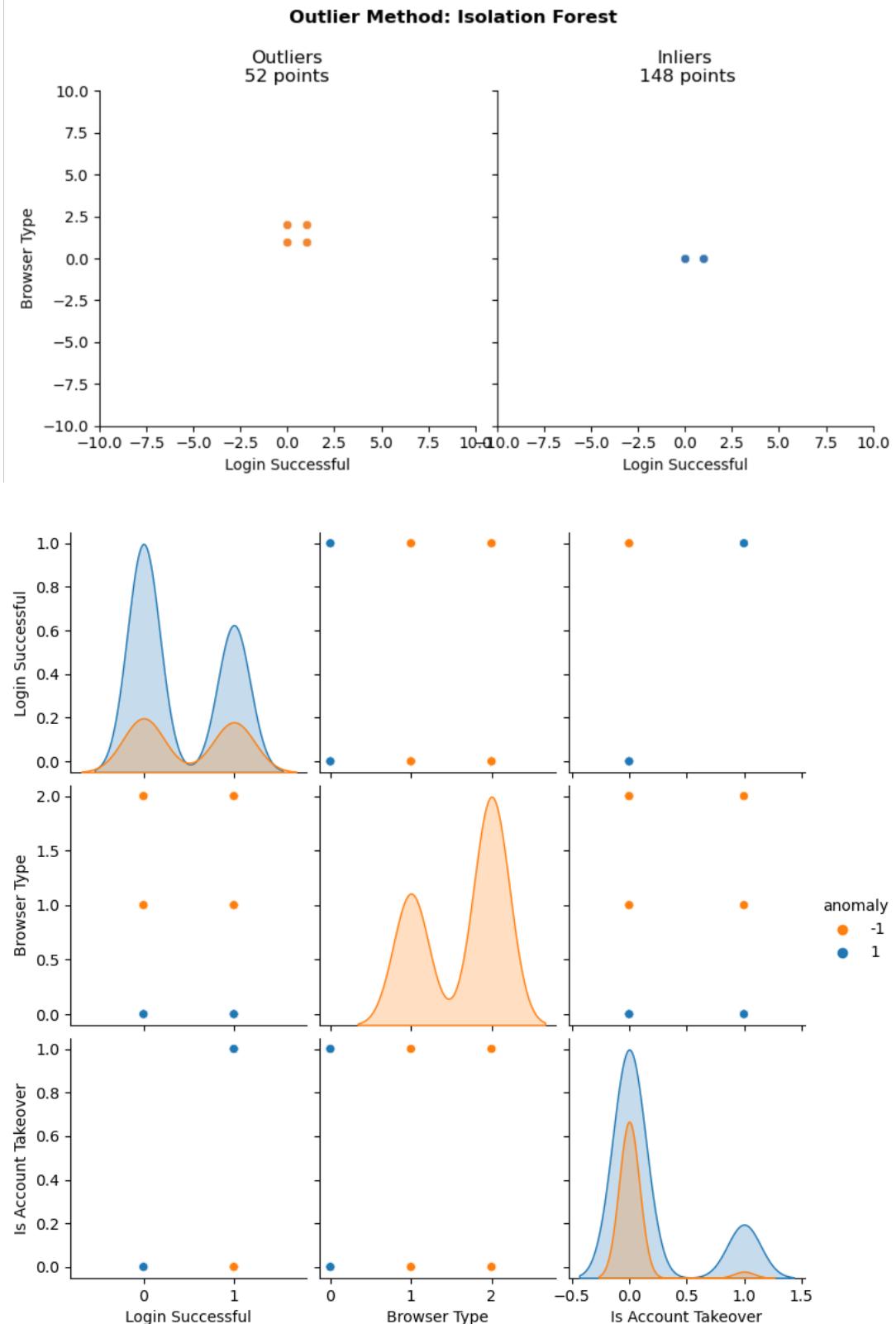


- contamination value == 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 52
Number of non anomalous values 148
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x2e7a4a790>

```

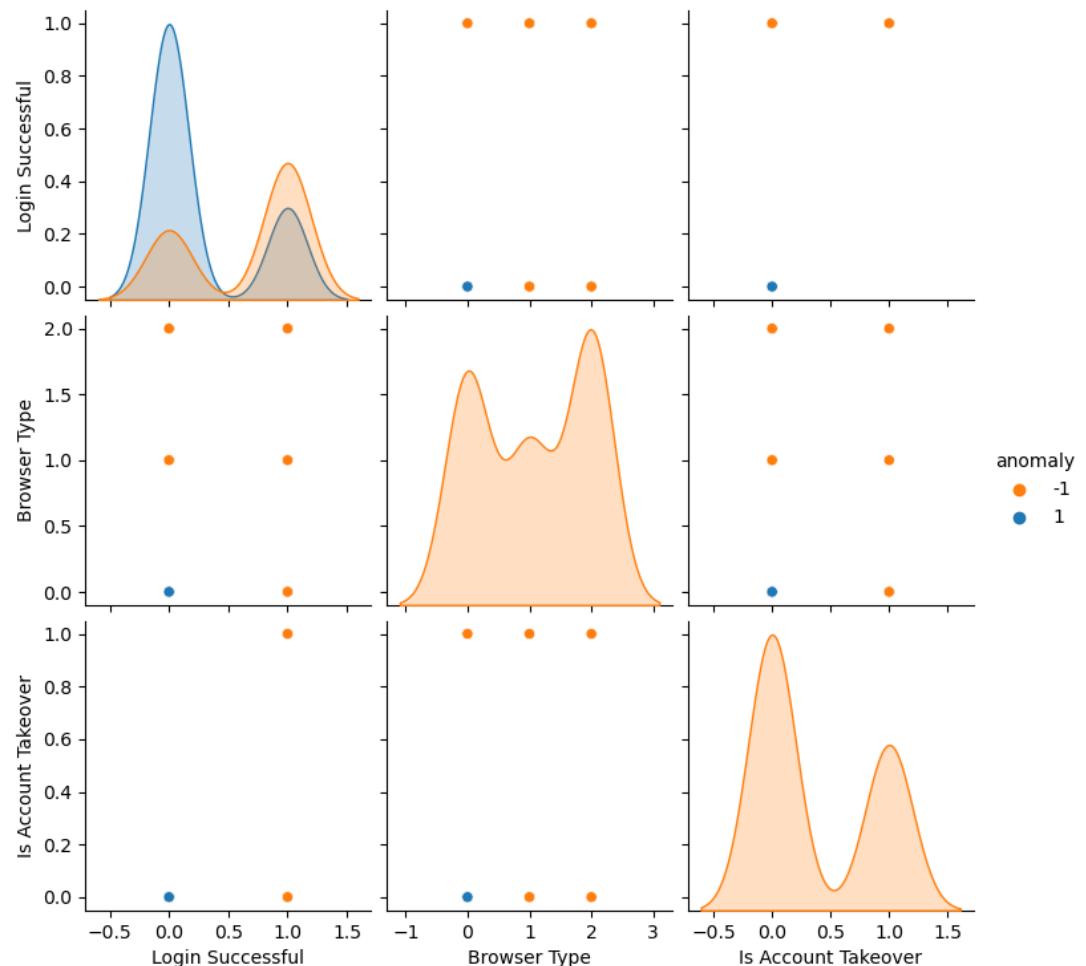
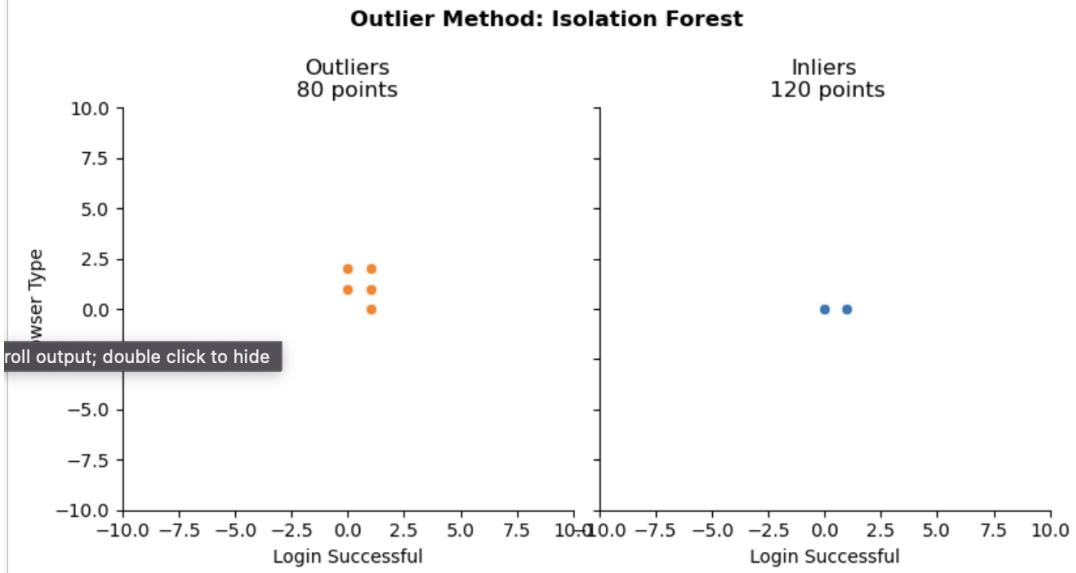


- contamination value == 0.5

```

Outlier Method: Isolation Forest
Number of anomalous values 80
Number of non anomalous values 120
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x16586dc90>

```

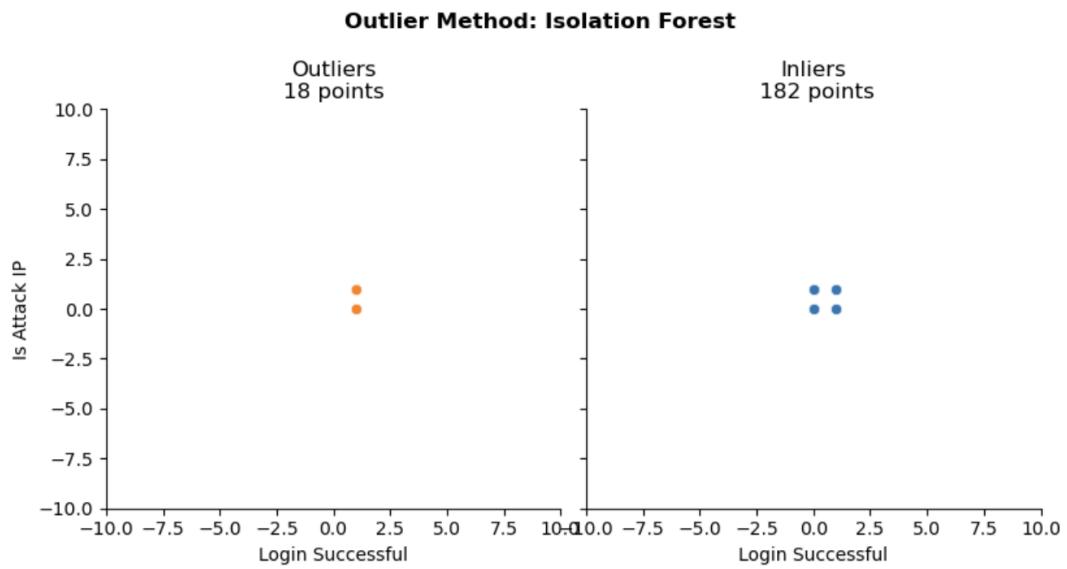


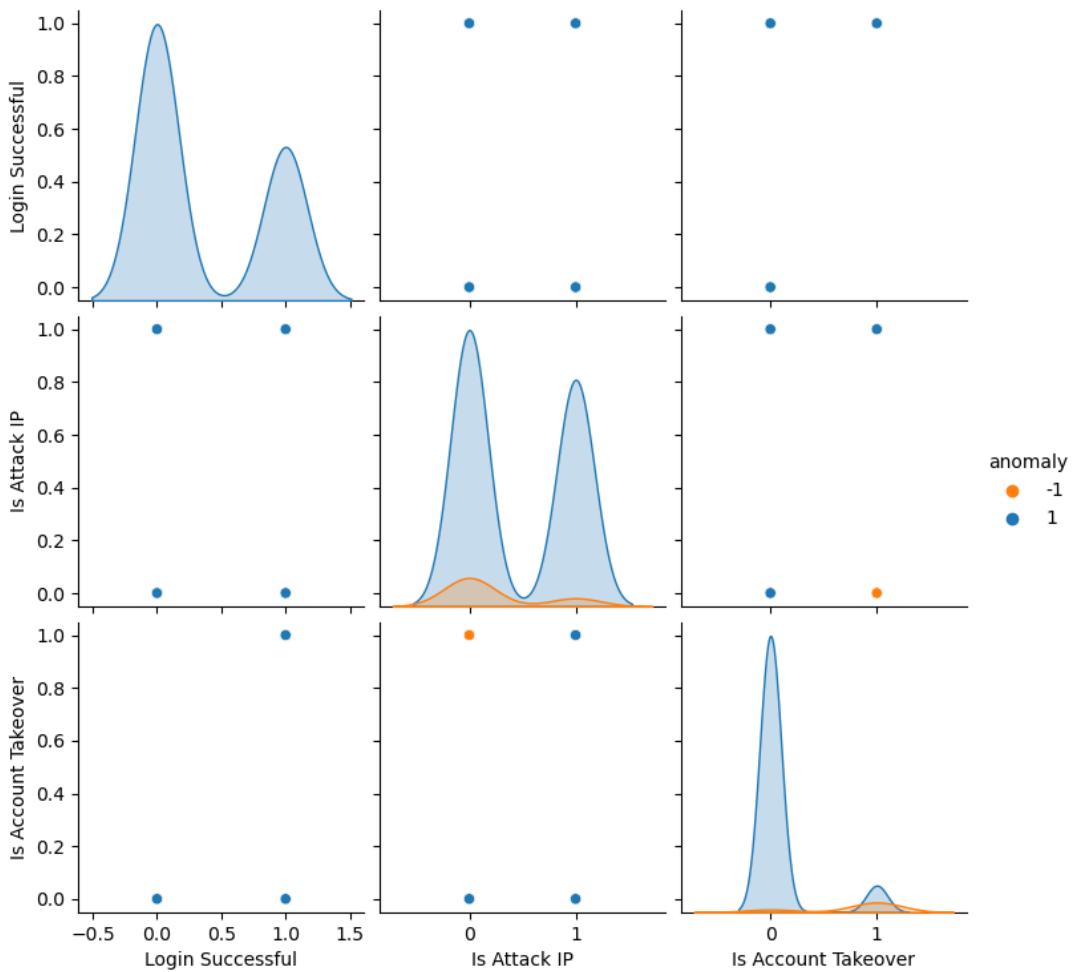
- anomaly\_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 18
Number of non anomalous values 182
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x2f03ec550>

```





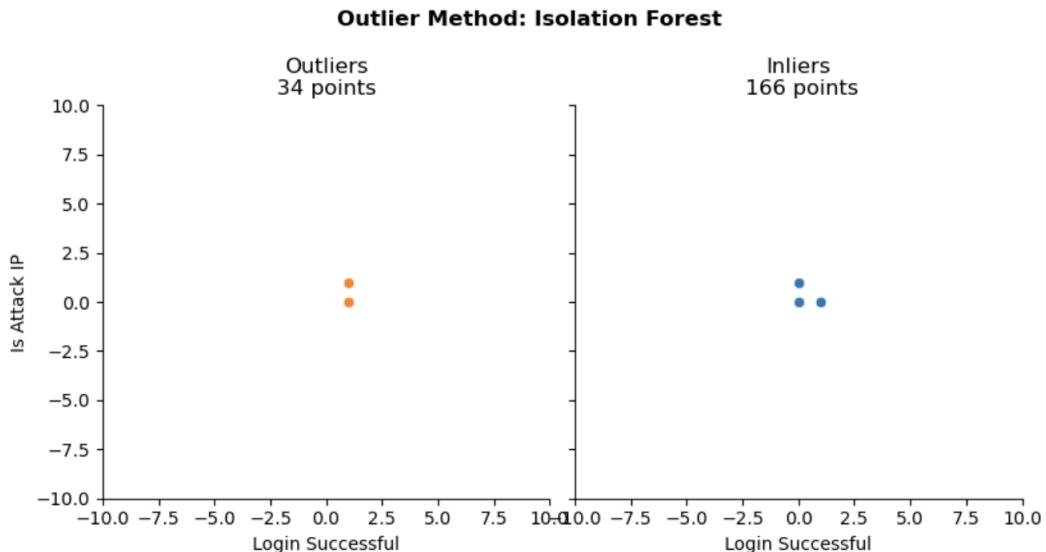
- contamination value == 0.3

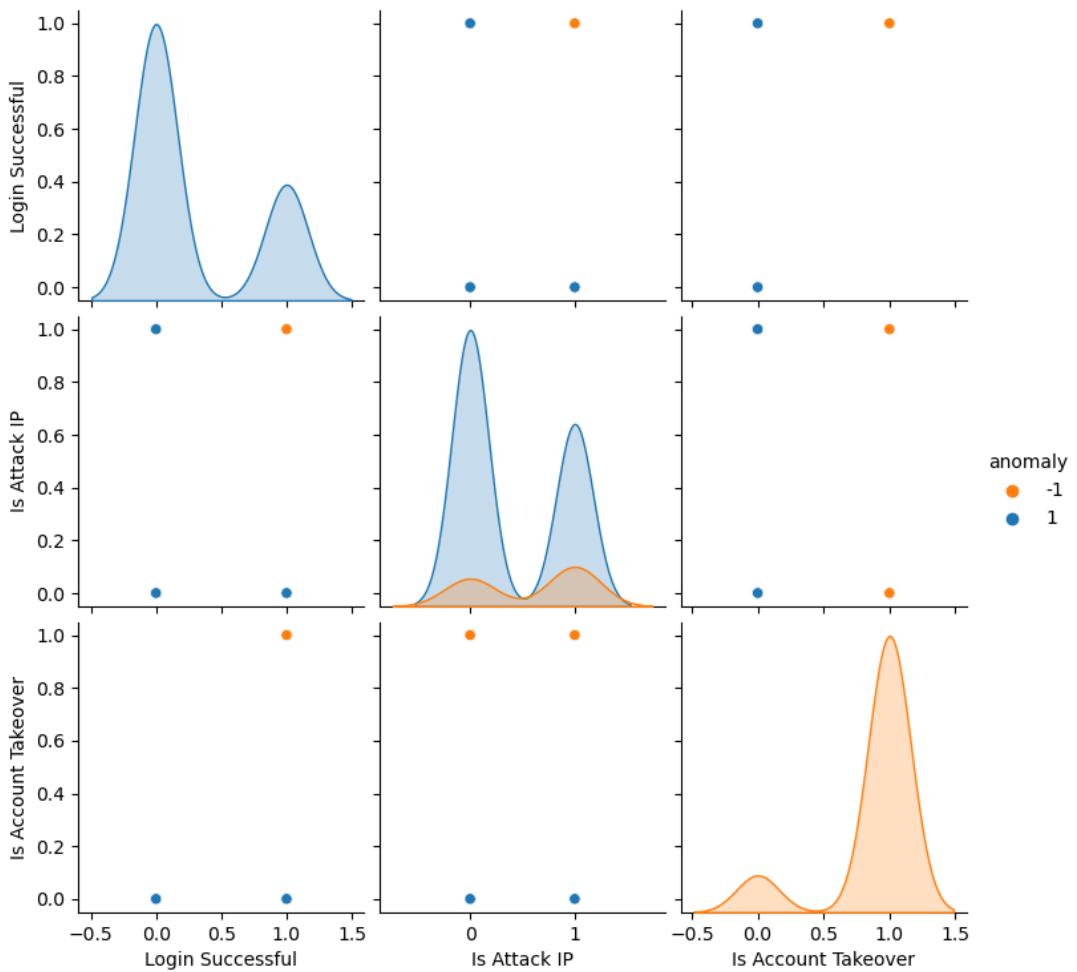
```

Outlier Method: Isolation Forest
Number of anomalous values 34
Number of non anomalous values 166
Total Number of Values: 200

<seaborn.axisgrid.PairGrid at 0x2da2de690>

```





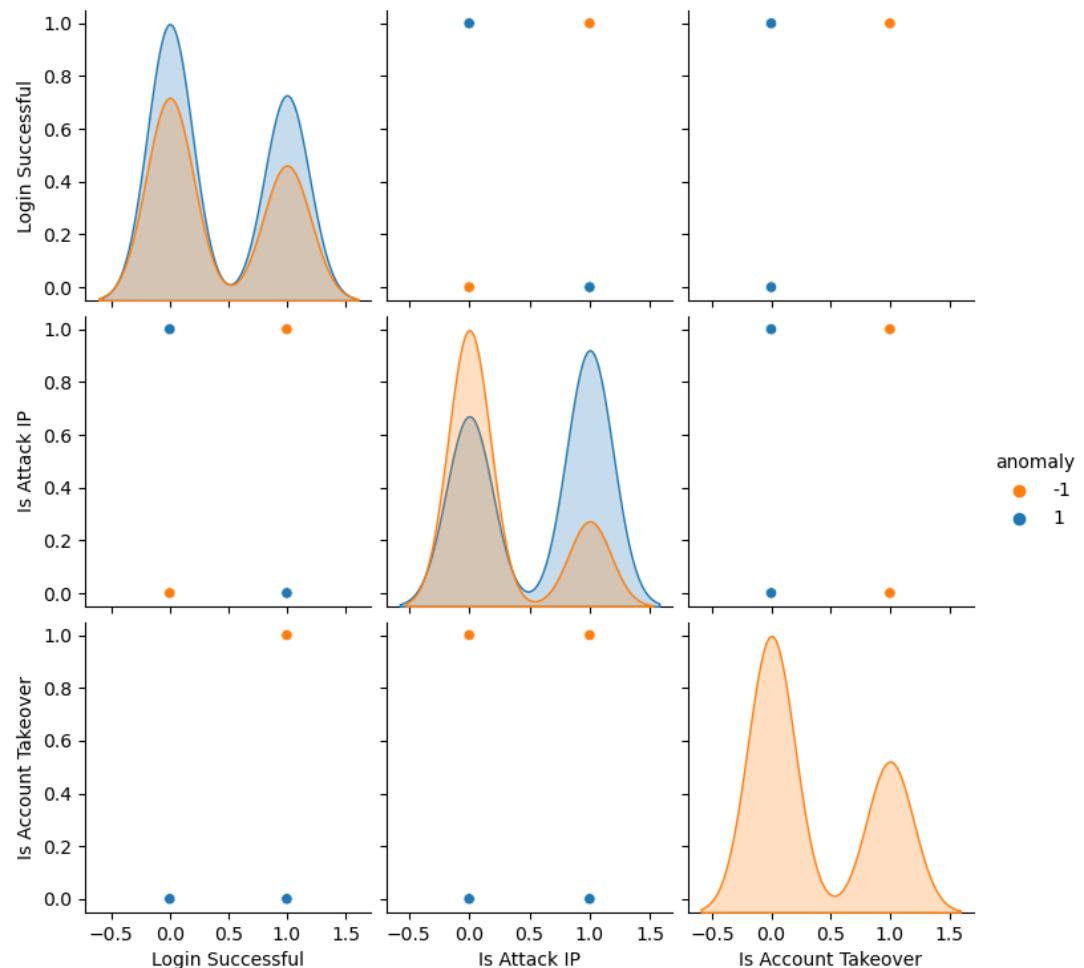
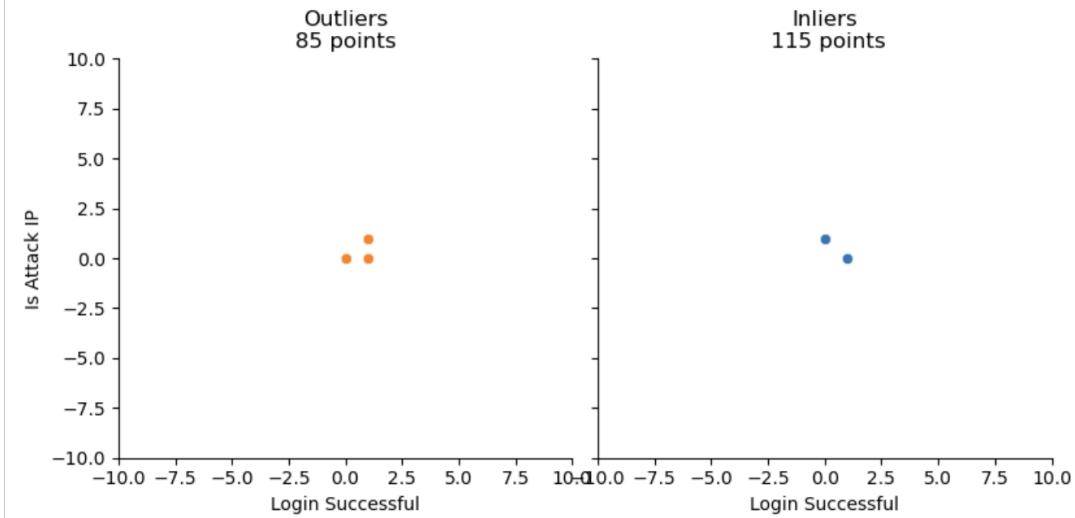
- contamination value == 0.5

```

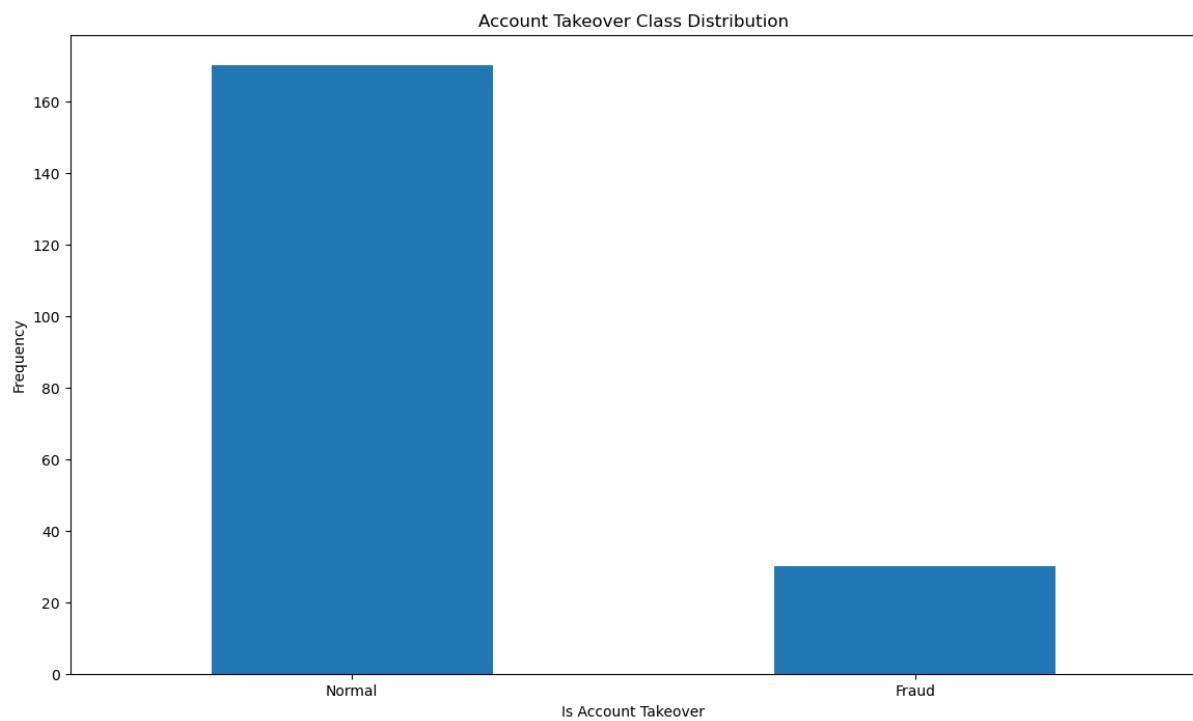
Outlier Method: Isolation Forest
Number of anomalous values 85
Number of non anomalous values 115
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x3099f60d0>

```

### Outlier Method: Isolation Forest



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



"{}: {}".format(clf\_name,n\_errors) → np. Isolation Forest: 21

-----  
Isolation Forest: 52

Accuracy Score :

0.74

Classification Report :

	precision	recall	f1-score	support
False	0.86	0.83	0.84	170
True	0.19	0.23	0.21	30
accuracy			0.74	200
macro avg	0.53	0.53	0.53	200
weighted avg	0.76	0.74	0.75	200

Local Outlier Factor: 52

Accuracy Score :

0.74

Classification Report :

	precision	recall	f1-score	support
False	0.86	0.83	0.84	170
True	0.19	0.23	0.21	30
accuracy			0.74	200
macro avg	0.53	0.53	0.53	200
weighted avg	0.76	0.74	0.75	200

Support Vector Machine: 148

Accuracy Score :

0.26

Classification Report :

	precision	recall	f1-score	support
False	0.73	0.21	0.32	170
True	0.11	0.57	0.19	30
accuracy			0.26	200
macro avg	0.42	0.39	0.25	200
weighted avg	0.64	0.26	0.30	200

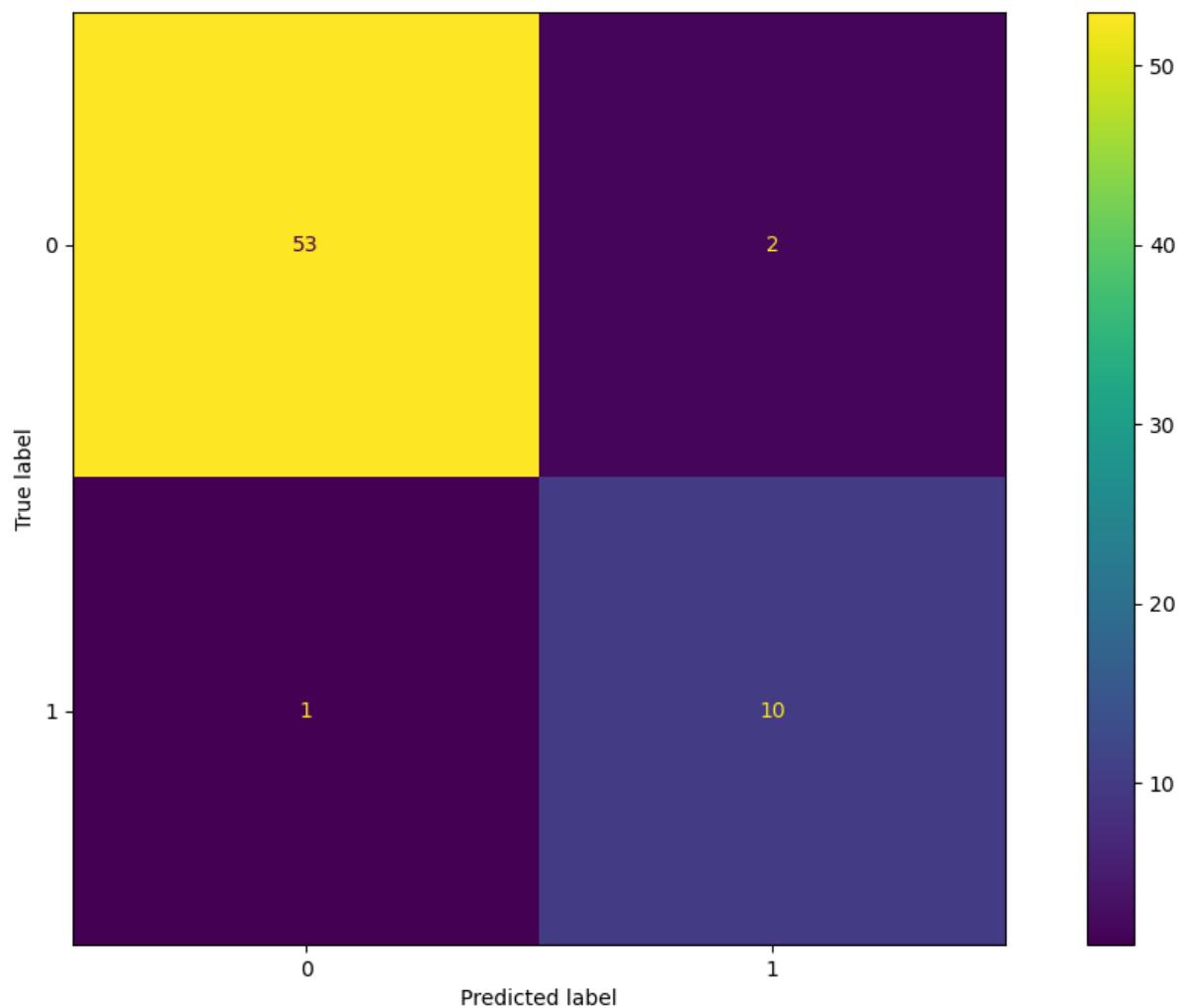
**3. Dummy Variable Encoding → df3 =**  
`pd.get_dummies(df3, columns=['Country', 'Device Type', 'IP Address', 'Login Timestamp', 'Browser Type'], prefix=["cntr", "dvc", "ip", "lgn", "brw"])`

### 3A. XGBoost

```
memory usage: 84.7 KB
```

Accuracy: 95.45%				
	precision	recall	f1-score	support
False	0.98	0.96	0.97	55
True	0.83	0.91	0.87	11
accuracy			0.95	66
macro avg	0.91	0.94	0.92	66
weighted avg	0.96	0.95	0.96	66

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 160
The number of records in the test dataset is 40
The training dataset has 135 records for the majority class and 25 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      35
      True         0.12     1.00     0.22       5

      accuracy          0.12      40
      macro avg       0.06     0.50     0.11      40
      weighted avg    0.02     0.12     0.03      40

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.88     1.00     0.93      35
      True         0.00     0.00     0.00       5

      accuracy          0.88      40
      macro avg       0.44     0.50     0.47      40
      weighted avg    0.77     0.88     0.82      40

[[ 0 35]
 [ 0  5]]

```

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

```

The number of records in the training dataset is 140
The number of records in the test dataset is 60
The training dataset has 118 records for the majority class and 22 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.31     0.47      52
      True         0.18     1.00     0.31       8

      accuracy          0.40      60
      macro avg       0.59     0.65     0.39      60
      weighted avg    0.89     0.40     0.45      60

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.87     1.00     0.93      52
      True         0.00     0.00     0.00       8

      accuracy          0.87      60
      macro avg       0.43     0.50     0.46      60
      weighted avg    0.75     0.87     0.80      60

[[16 36]
 [ 0  8]]

```

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

```

The number of records in the training dataset is 100
The number of records in the test dataset is 100
The training dataset has 81 records for the majority class and 19 records for the minority class.
      precision    recall   f1-score   support
False        0.00     0.00     0.00      89
True        0.11     1.00     0.20      11

accuracy                           0.11      100
macro avg                          0.06      100
weighted avg                       0.01      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.89     1.00     0.94      89
True         0.00     0.00     0.00      11

accuracy                           0.89      100
macro avg                          0.45      100
weighted avg                       0.79      100

[[ 0 89]
 [ 0 11]]

```

### 3C. Isolation Forest

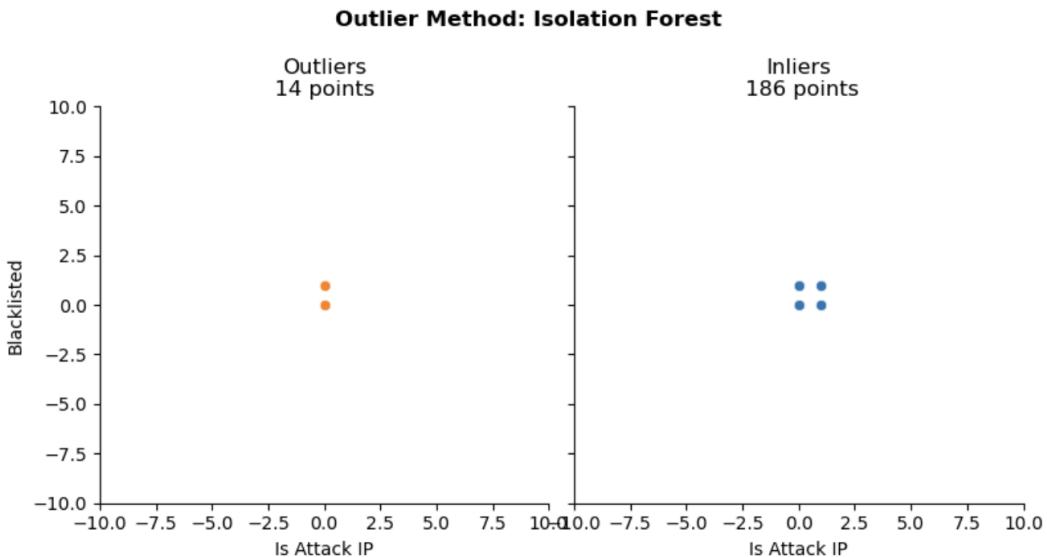
- anomaly\_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
  - contamination value == 0.1

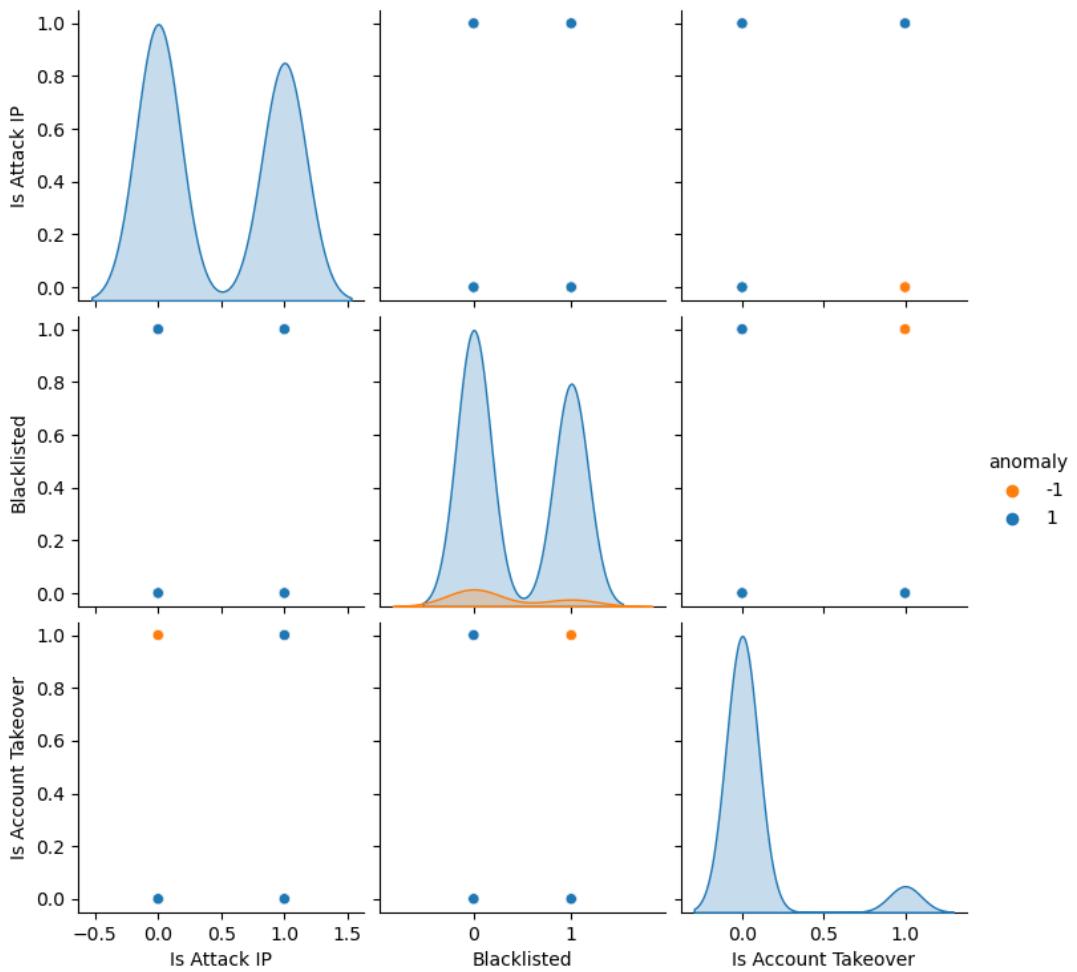
---

```

Outlier Method: Isolation Forest
Number of anomalous values 14
Number of non anomalous values 186
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x30bc41b50>

```



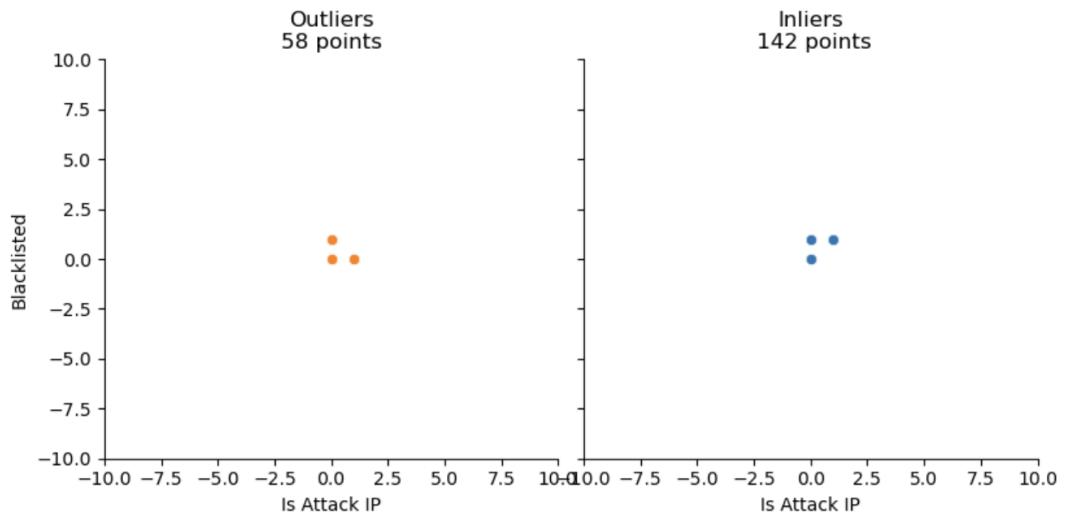


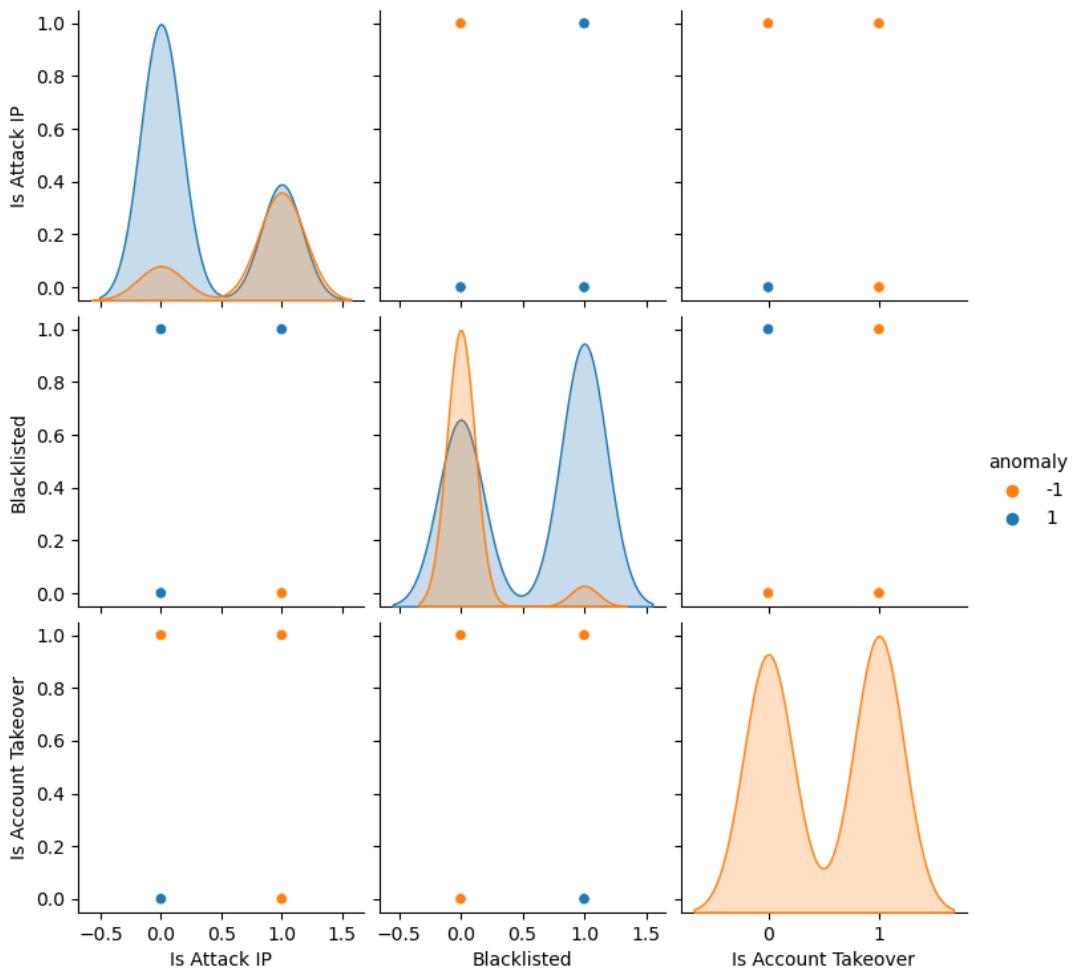
- contamination value == 0.3

```
Outlier Method: Isolation Forest
Number of anomalous values 58
Number of non anomalous values 142
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x2b6282690>
```

### Outlier Method: Isolation Forest

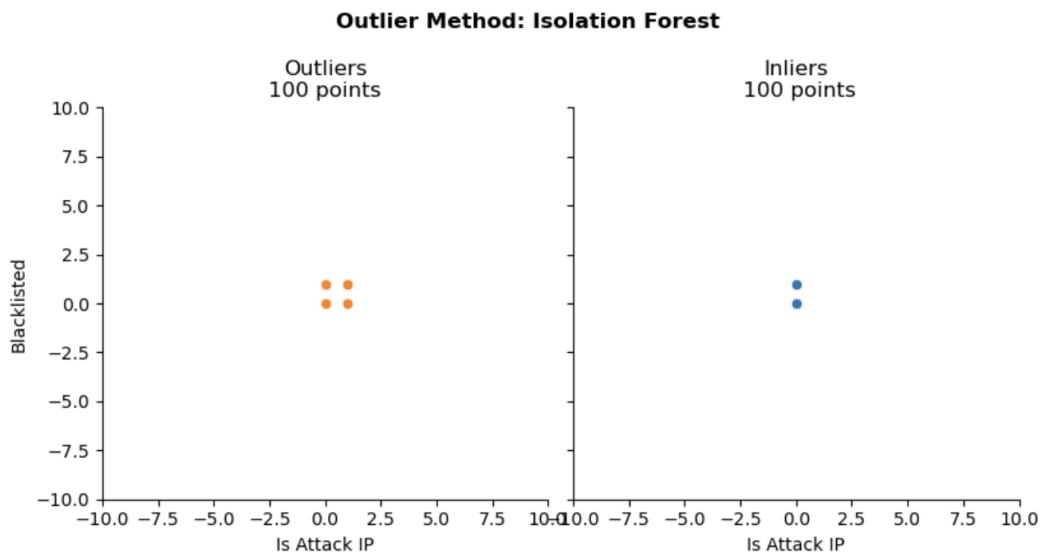


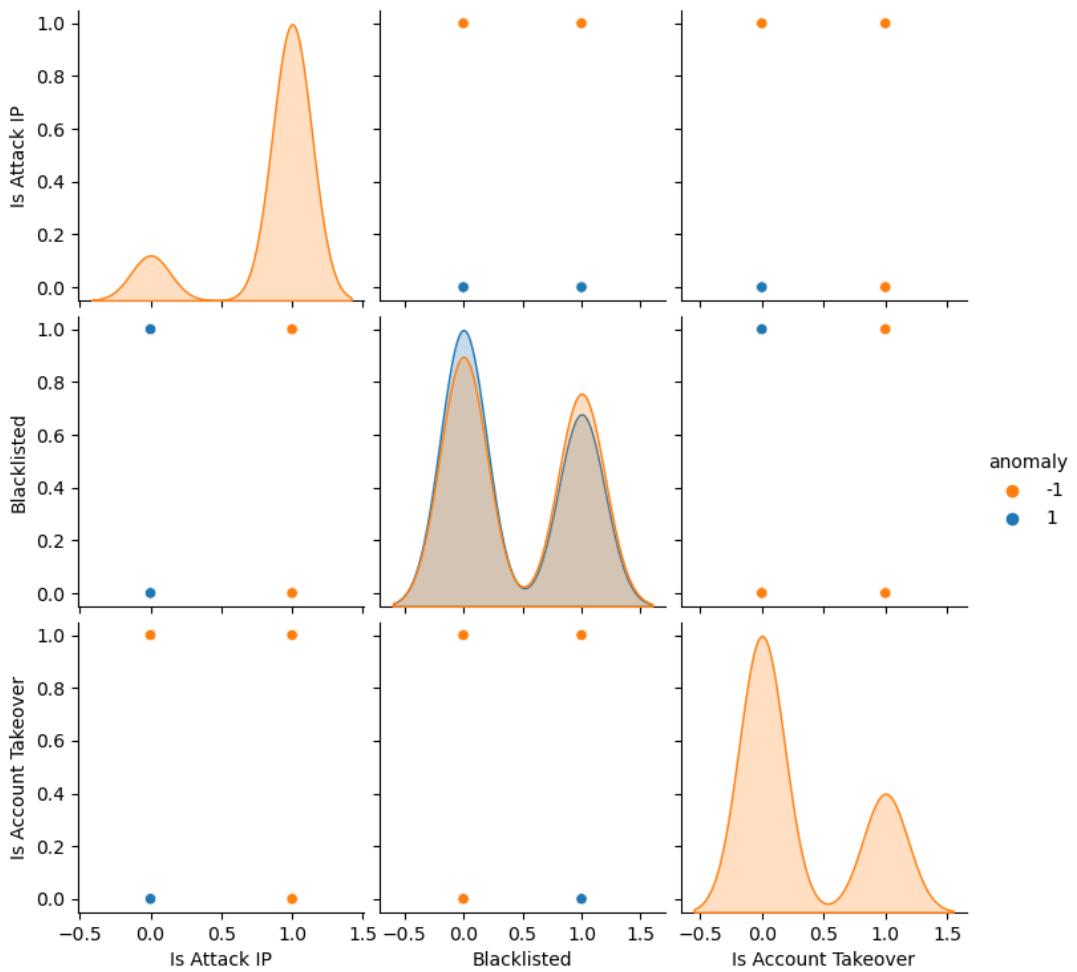


- contamination value == 0.5

```
Outlier Method: Isolation Forest
Number of anomalous values 100
Number of non anomalous values 100
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x30bb57010>
```



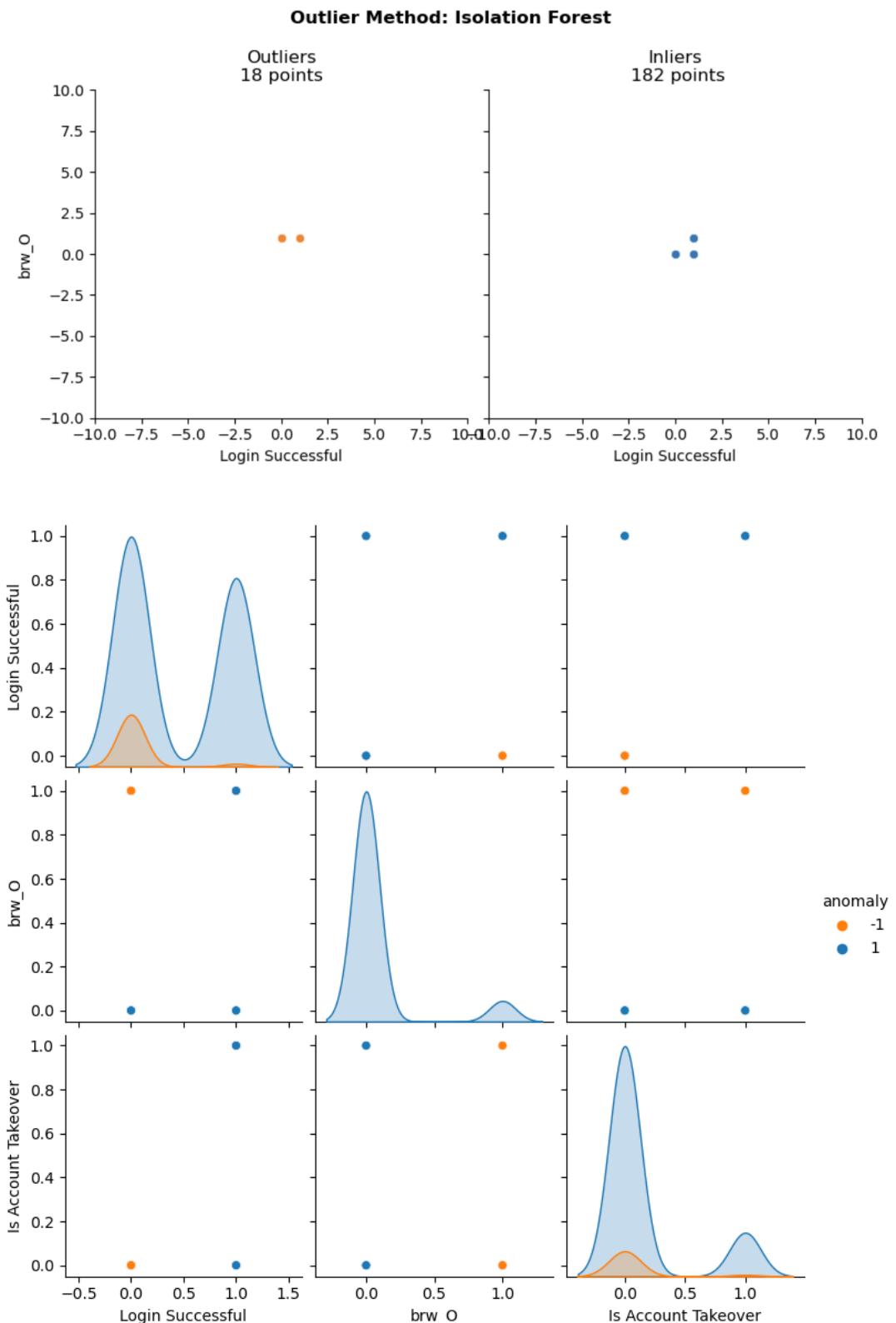


- `anomaly_inputs = ['Login Successful', 'Browser Type (brw_O)', 'Is Account Takeover']`
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 18
Number of non anomalous values 182
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x2d93640d0>

```



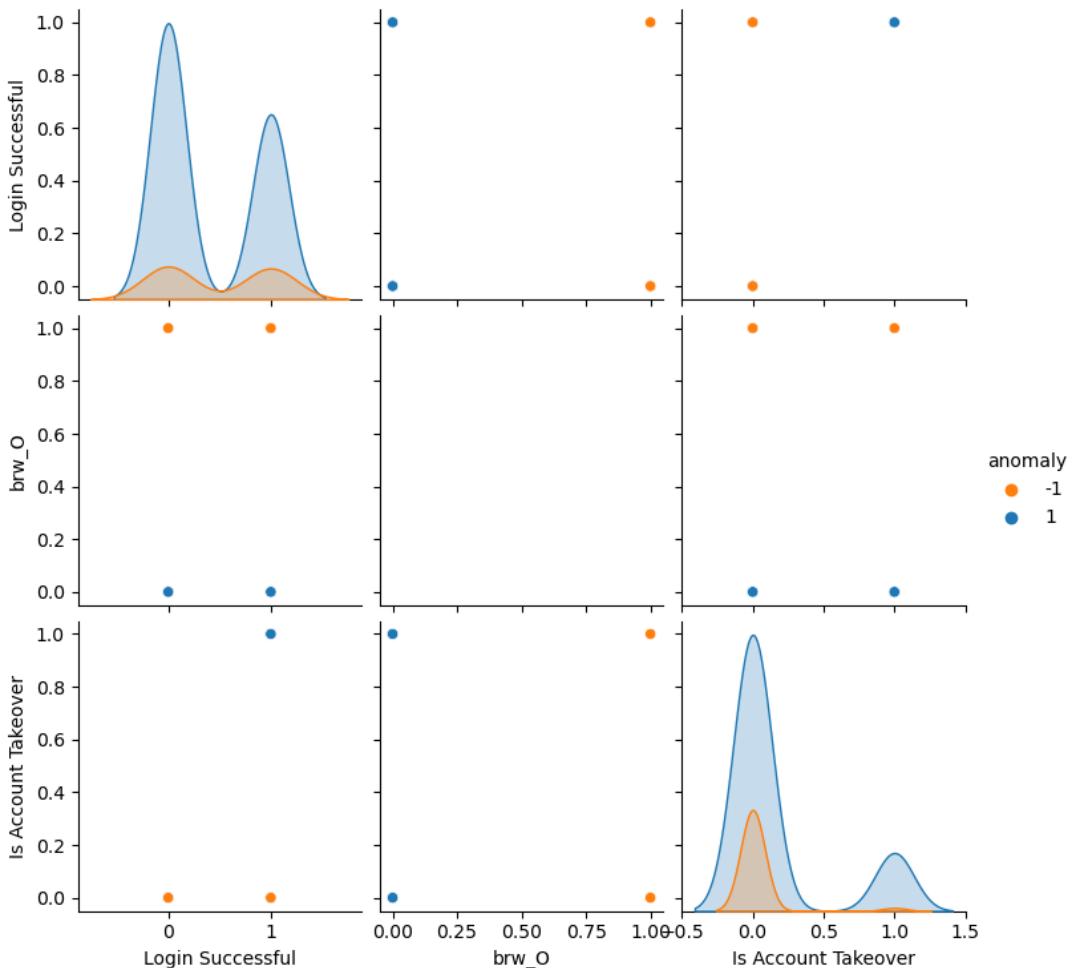
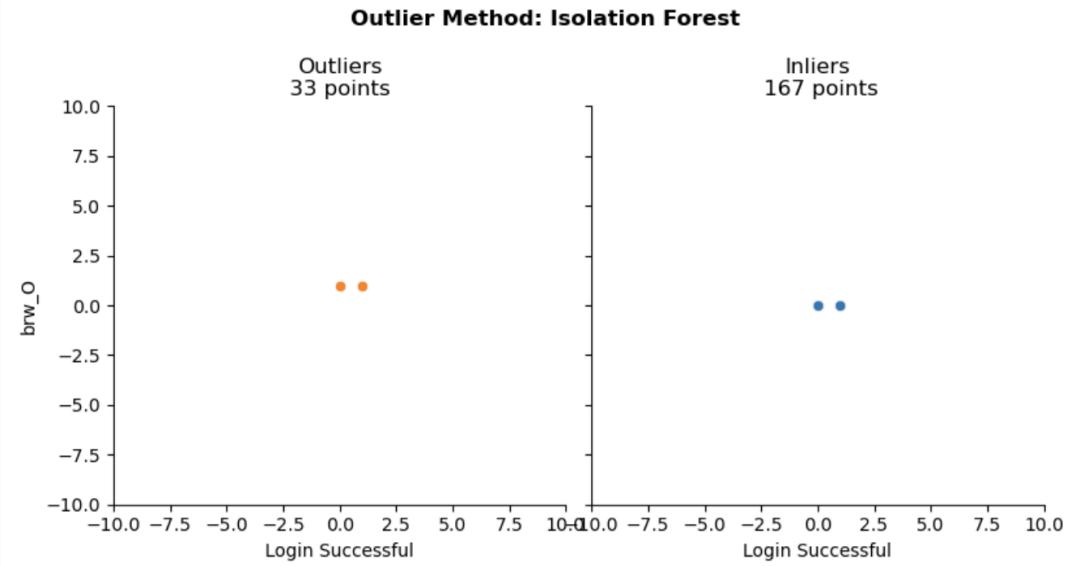
- contamination value == 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 33
Number of non anomalous values 167
Total Number of Values: 200

<seaborn.axisgrid.PairGrid at 0x30add4750>

```



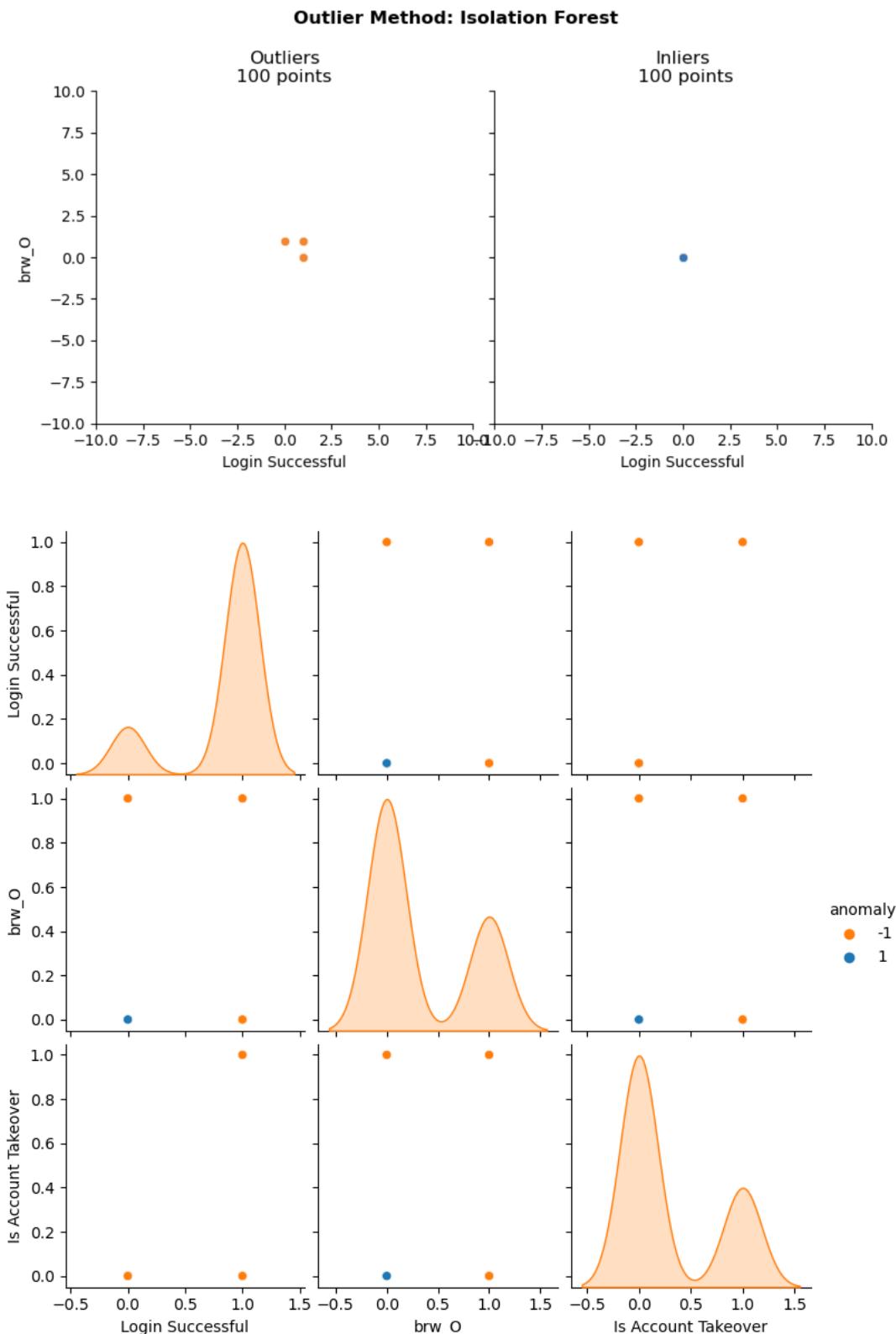
- contamination value == 0.5

```

Outlier Method: Isolation Forest
Number of anomalous values 100
Number of non anomalous values 100
Total Number of Values: 200

<seaborn.axisgrid.PairGrid at 0x30cce5150>

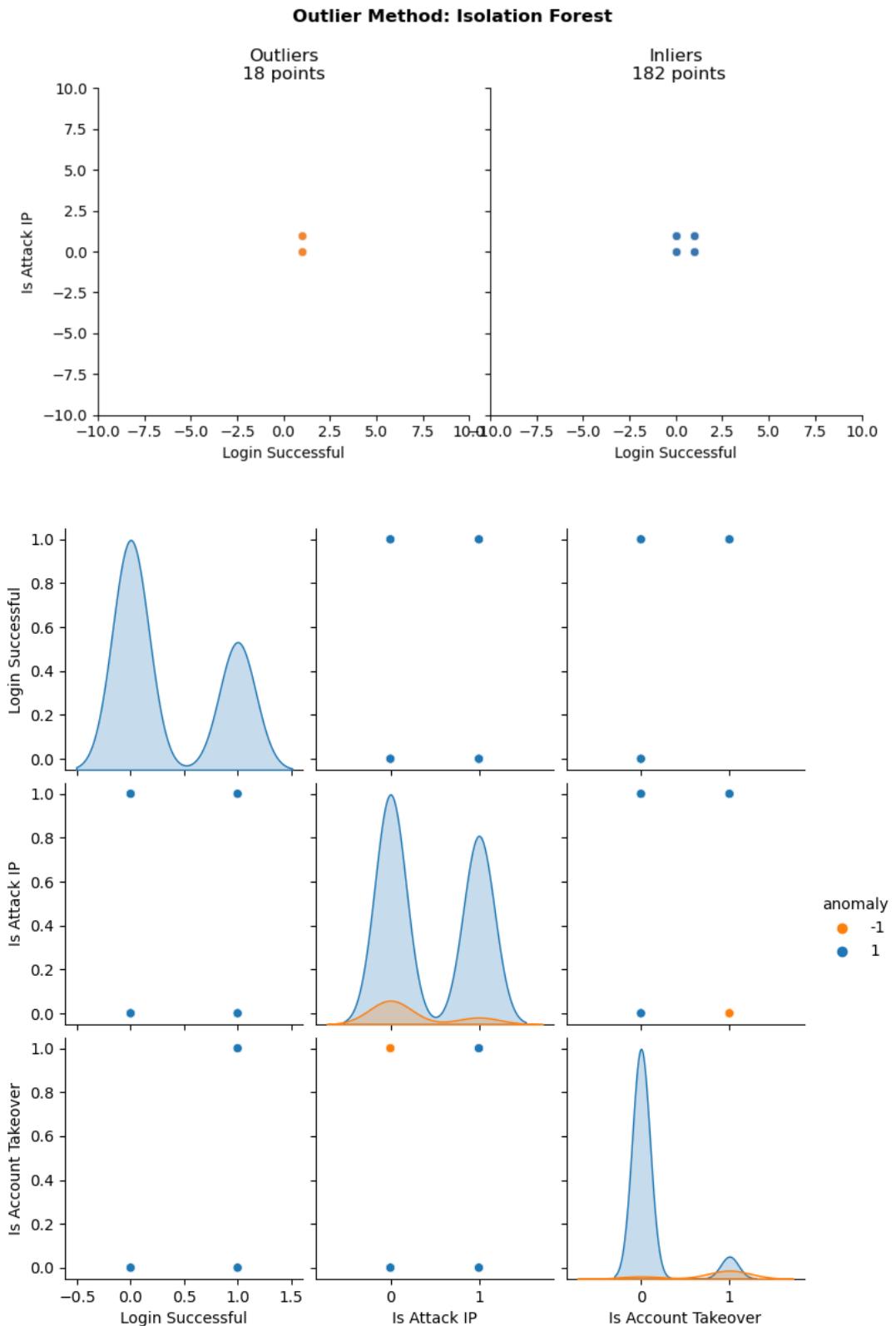
```



- anomaly\_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']

- contamination value == 0.1

```
Outlier Method: Isolation Forest
Number of anomalous values 18
Number of non anomalous values 182
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x30d3db5d0>
```

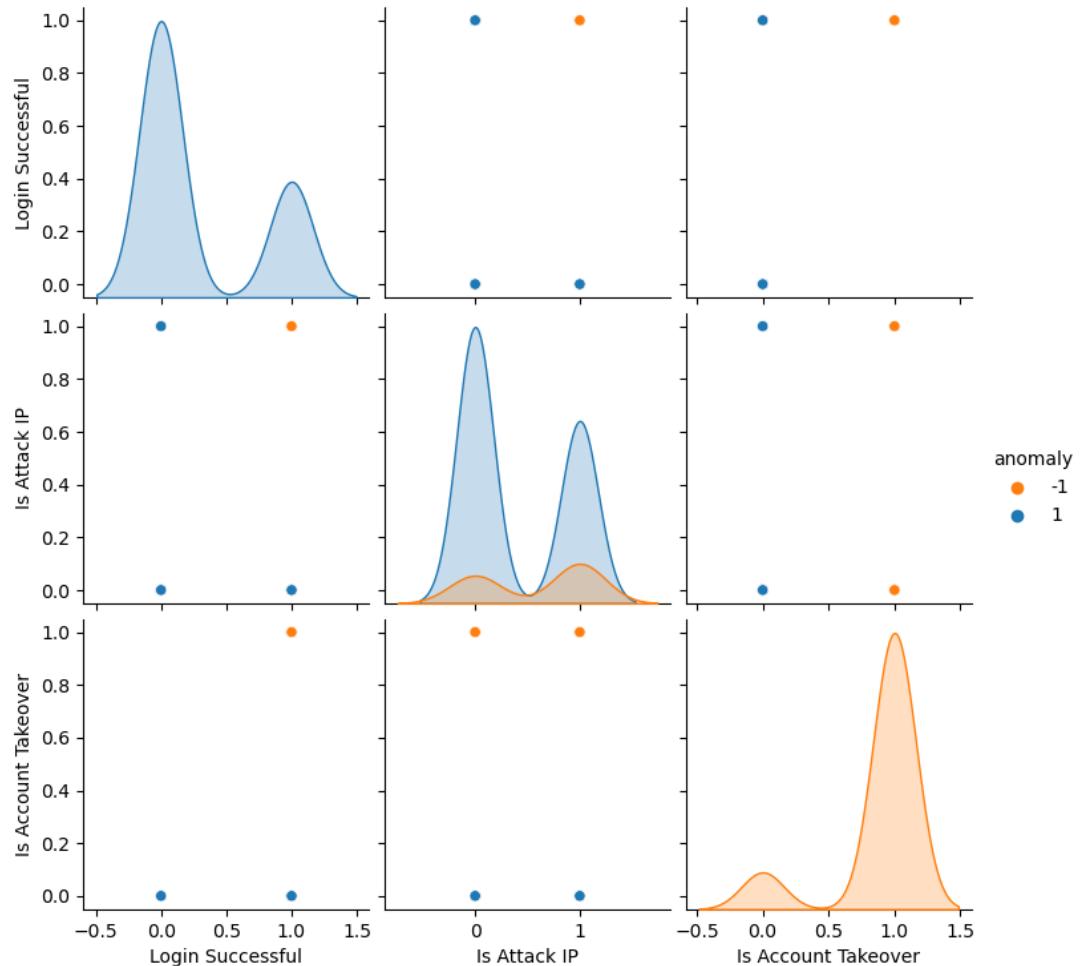
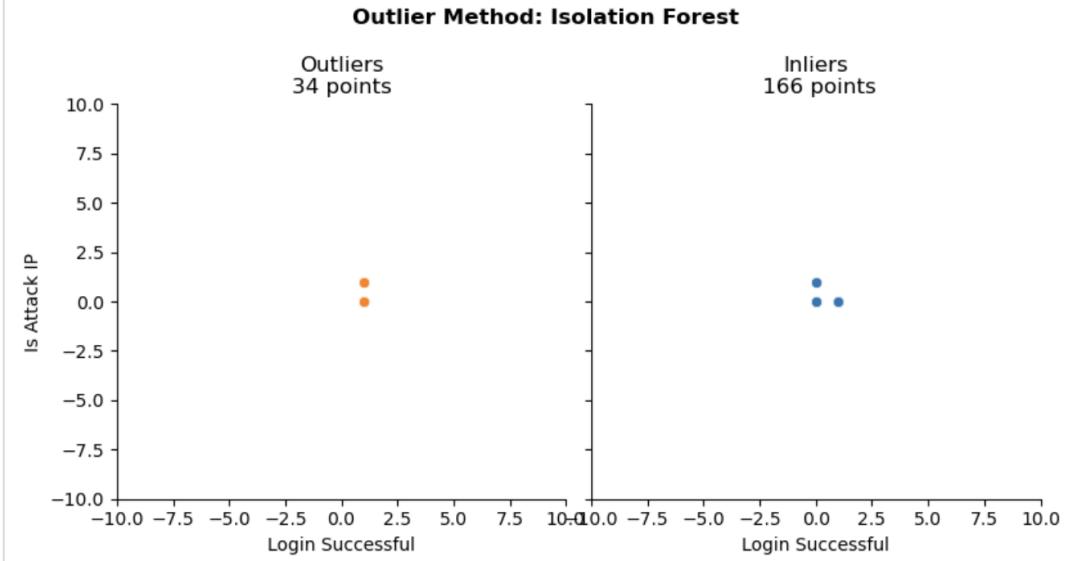


- contamination value == 0.3

```

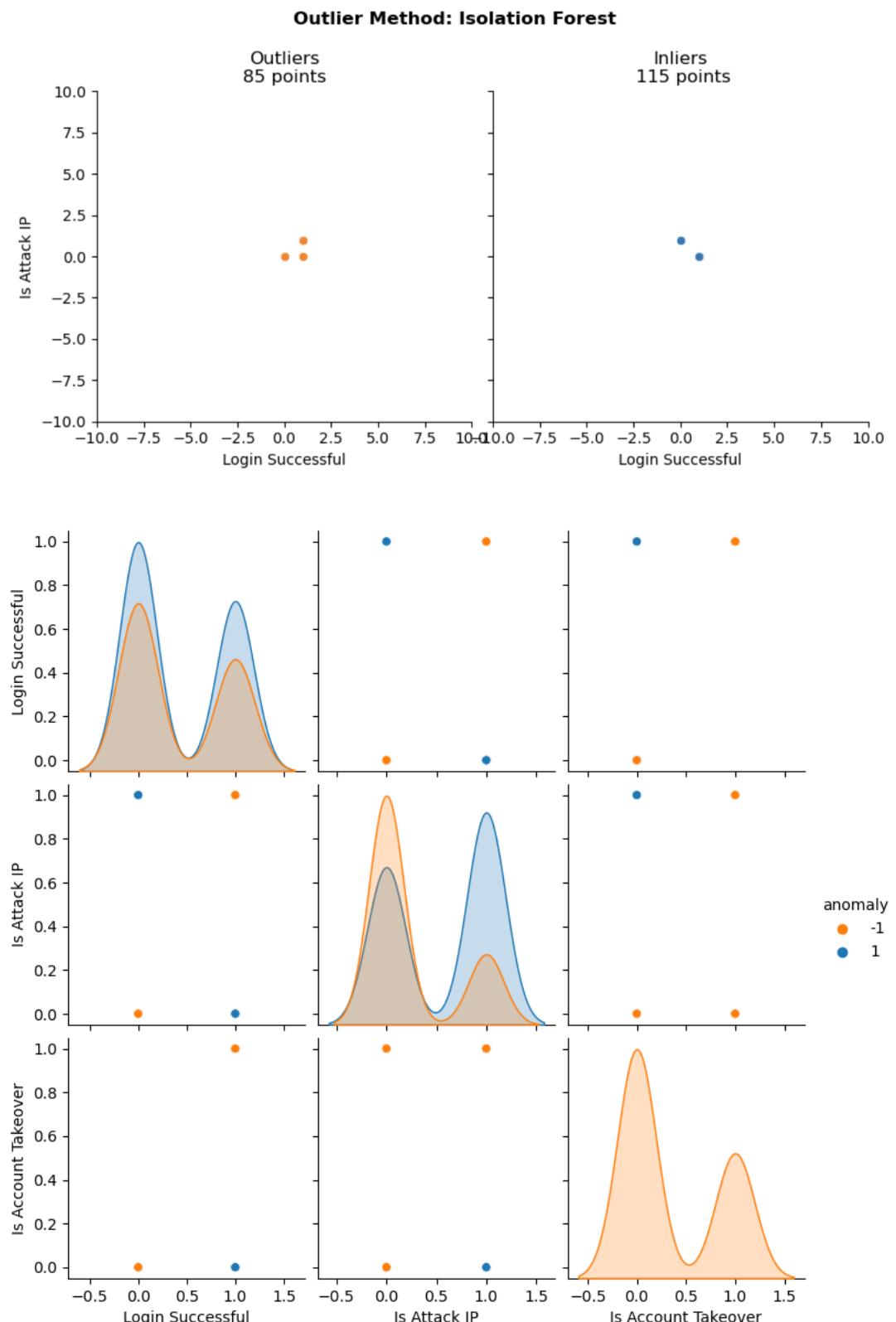
Outlier Method: Isolation Forest
Number of anomalous values 34
Number of non anomalous values 166
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x30d4104d0>

```

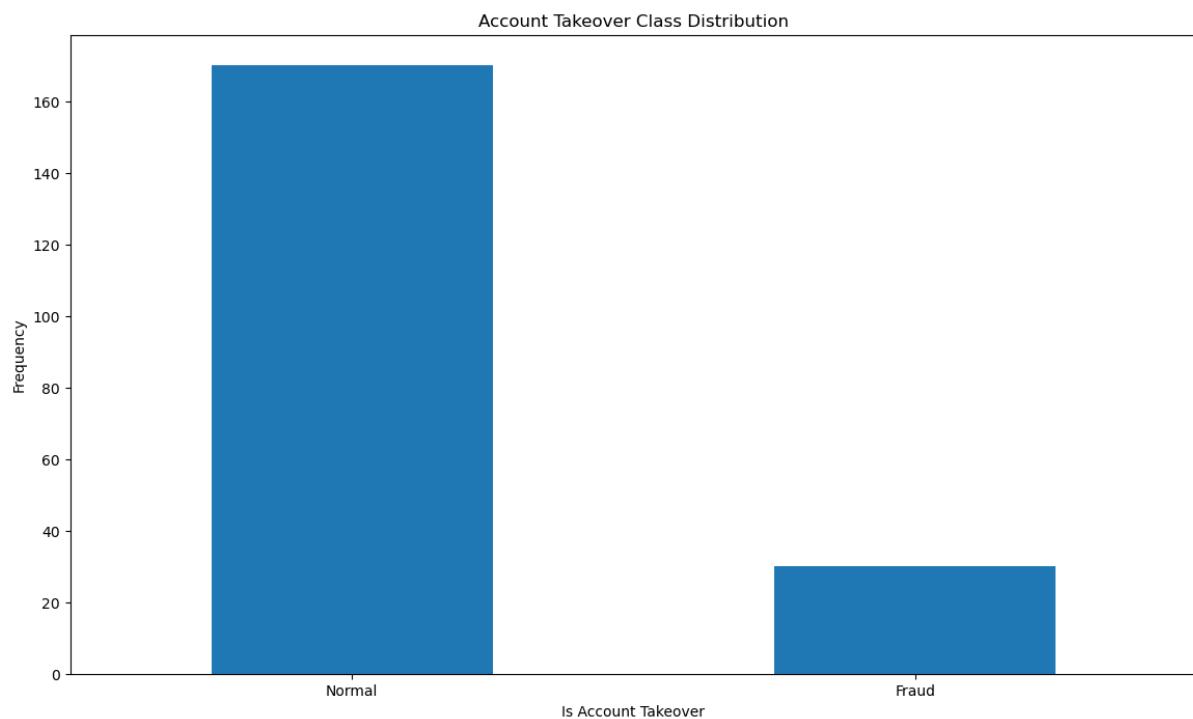


- contamination value == 0.5

```
Outlier Method: Isolation Forest
Number of anomalous values 85
Number of non anomalous values 115
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x175752d10>
```



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



"{}: {}".format(clf\_name,n\_errors) → np. Isolation Forest: 21

**Isolation Forest:** 58

**Accuracy Score :**

0.71

**Classification Report :**

	precision	recall	f1-score	support
False	0.84	0.81	0.83	170
True	0.11	0.13	0.12	30
accuracy			0.71	200
macro avg	0.48	0.47	0.47	200
weighted avg	0.73	0.71	0.72	200

**Local Outlier Factor:** 45

**Accuracy Score :**

0.775

**Classification Report :**

	precision	recall	f1-score	support
False	0.88	0.85	0.87	170
True	0.29	0.33	0.31	30
accuracy			0.78	200
macro avg	0.58	0.59	0.59	200
weighted avg	0.79	0.78	0.78	200

**Support Vector Machine:** 148

**Accuracy Score :**

0.26

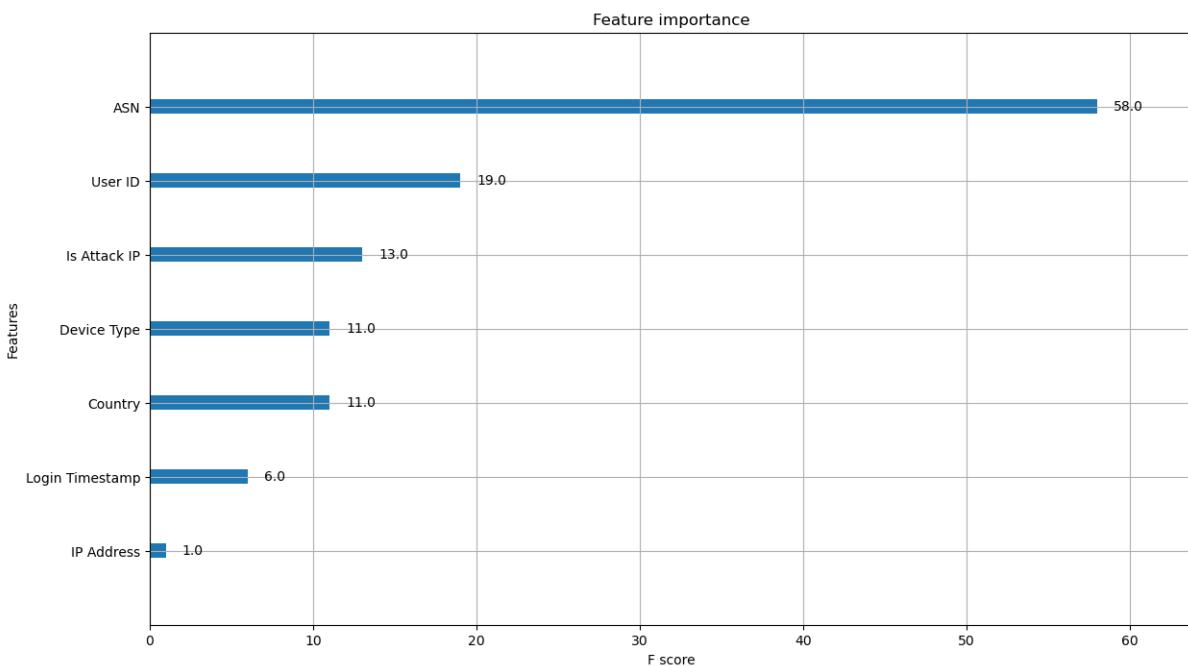
**Classification Report :**

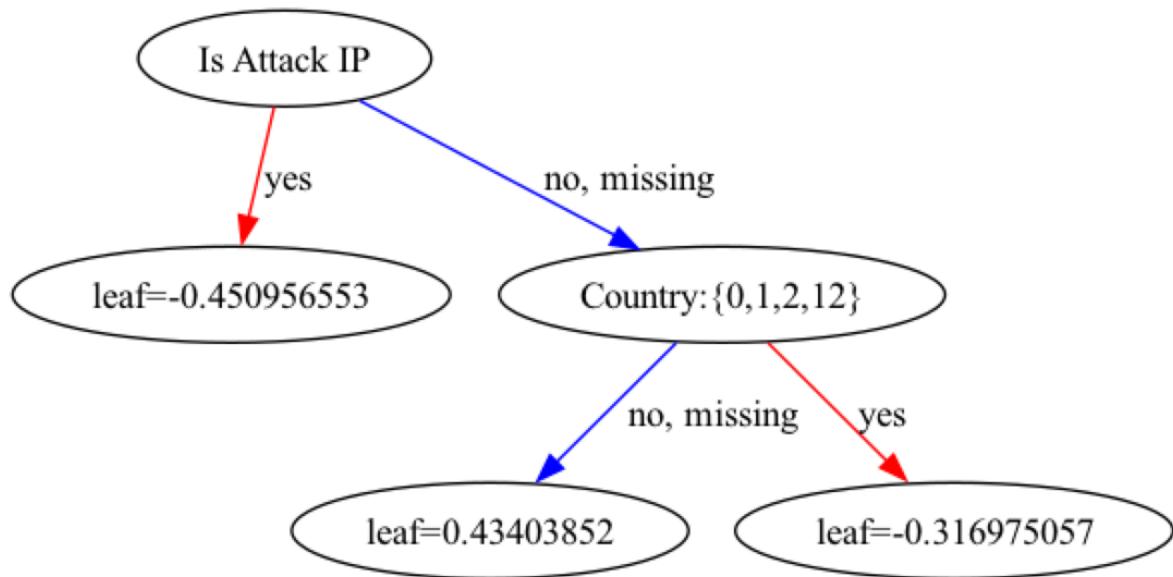
	precision	recall	f1-score	support
False	0.73	0.21	0.32	170
True	0.11	0.57	0.19	30
accuracy			0.26	200
macro avg	0.42	0.39	0.25	200
weighted avg	0.64	0.26	0.30	200

# III. 77/200 → 38.5%

## 1. XGB Experimental

```
--- ----- ---  
0 Login Timestamp    200 non-null   category  
1 User ID           200 non-null   int64  
2 IP Address         200 non-null   category  
3 Country            200 non-null   category  
4 ASN                200 non-null   int64  
5 Device Type        200 non-null   category  
6 Login Successful   200 non-null   bool  
7 Is Attack IP       200 non-null   bool  
8 Blacklisted         200 non-null   bool  
9 Browser Type       200 non-null   category  
10 Is Account Takeover 200 non-null bool  
dtypes: bool(4), category(5), int64(2)  
memory usage: 21.4 KB  
Feature importances:  
[0.118387  0.0012349  0.00642419 0.2238188  0.00738945 0.00833171  
 0.          0.634414   0.          0.          ]
```



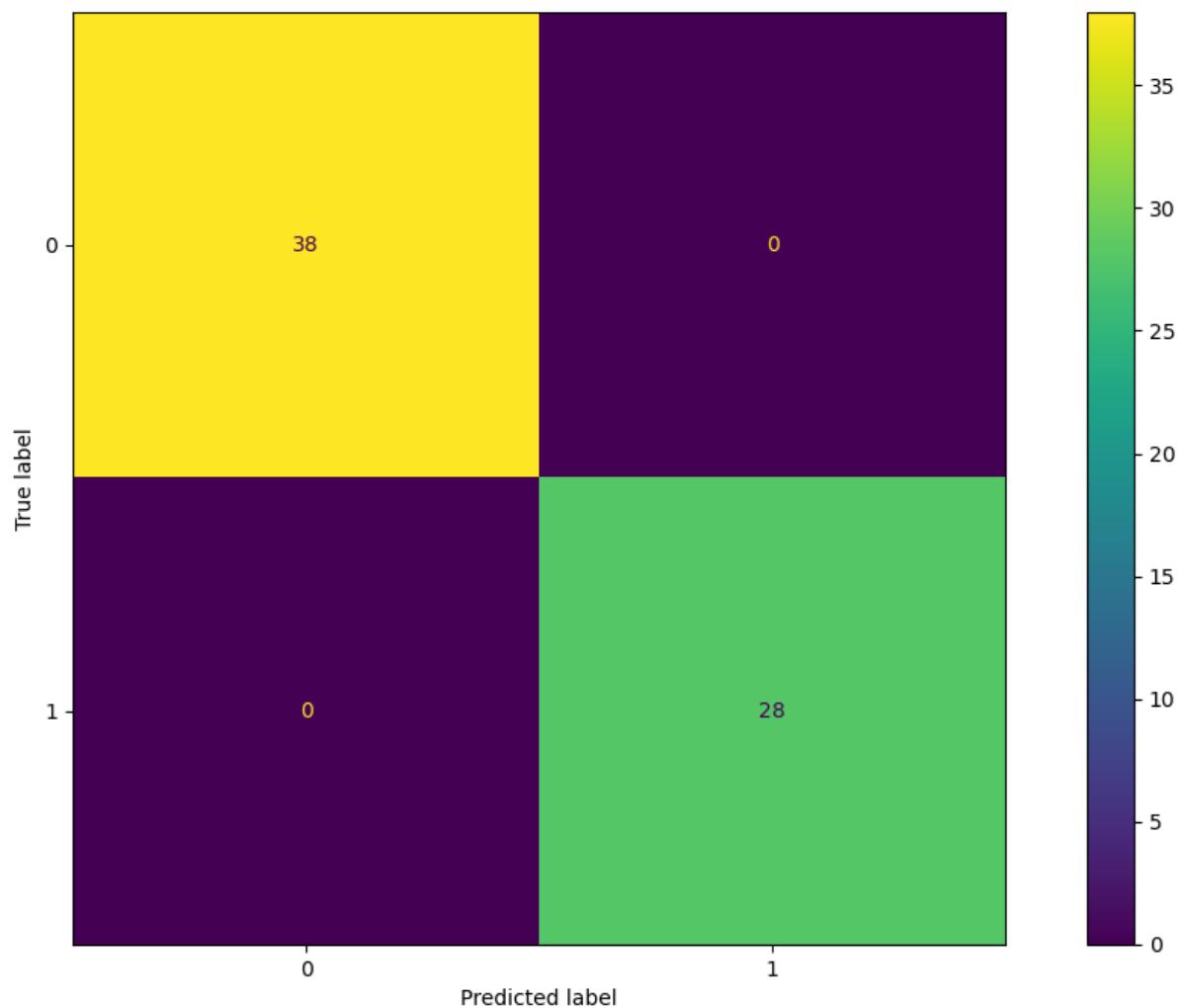


## 2. Label encoding (cat.codes)

### 2A. XGBoost

`memory usage: 5.4 KB`

Accuracy: 100.00%				
	precision	recall	f1-score	support
False	1.00	1.00	1.00	38
True	1.00	1.00	1.00	28
accuracy			1.00	66
macro avg	1.00	1.00	1.00	66
weighted avg	1.00	1.00	1.00	66
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 160
The number of records in the test dataset is 40
The training dataset has 98 records for the majority class and 62 records for the minority class.

precision    recall   f1-score   support
False        0.00     0.00     0.00      25
True         0.38     1.00     0.55      15

accuracy          0.38      40
macro avg       0.19     0.50     0.27      40
weighted avg    0.14     0.38     0.20      40

The customized score threshold for 2% of outliers is 0.00
precision    recall   f1-score   support
False        0.62     1.00     0.77      25
True         0.00     0.00     0.00      15

accuracy          0.62      40
macro avg       0.31     0.50     0.38      40
weighted avg    0.39     0.62     0.48      40

[[ 0 25]
 [ 0 15]]

```

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

```

The number of records in the training dataset is 140
The number of records in the test dataset is 60
The training dataset has 86 records for the majority class and 54 records for the minority class.

precision    recall   f1-score   support
False        0.00     0.00     0.00      37
True         0.38     1.00     0.55      23

accuracy          0.38      60
macro avg       0.19     0.50     0.28      60
weighted avg    0.15     0.38     0.21      60

The customized score threshold for 2% of outliers is 0.00
precision    recall   f1-score   support
False        0.62     1.00     0.76      37
True         0.00     0.00     0.00      23

accuracy          0.62      60
macro avg       0.31     0.50     0.38      60
weighted avg    0.38     0.62     0.47      60

[[ 0 37]
 [ 0 23]]

```

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

```

The number of records in the training dataset is 100
The number of records in the test dataset is 100
The training dataset has 62 records for the majority class and 38 records for the minority class.

      precision    recall   f1-score   support

  False        0.00     0.00     0.00      61
  True        0.39     1.00     0.56      39

    accuracy          0.39      100
   macro avg       0.20     0.50     0.28      100
weighted avg     0.15     0.39     0.22      100

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

  False        0.61     1.00     0.76      61
  True        0.00     0.00     0.00      39

    accuracy          0.61      100
   macro avg       0.30     0.50     0.38      100
weighted avg     0.37     0.61     0.46      100

[[ 0 61]
 [ 0 39]]

```

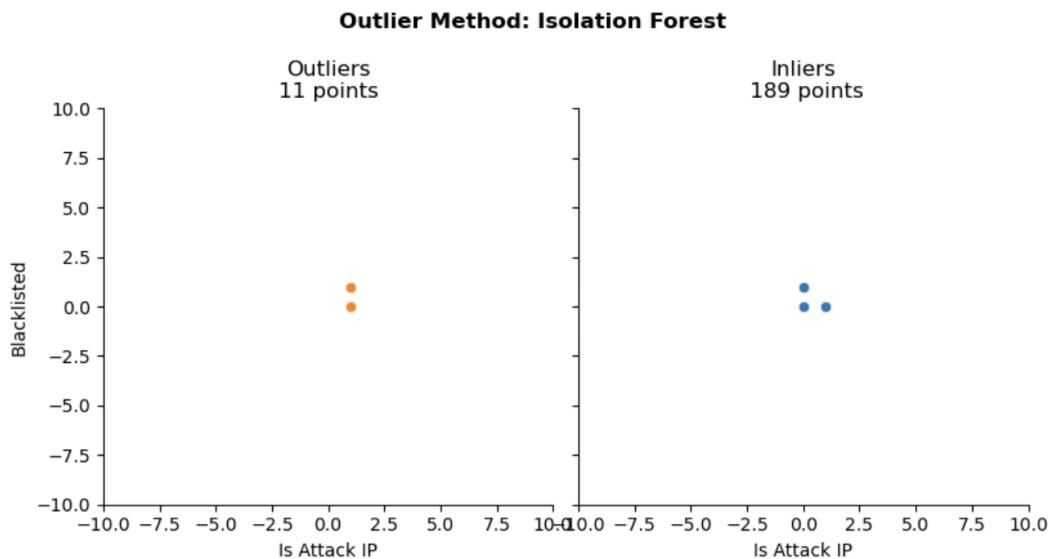
## 2C. Isolation Forest

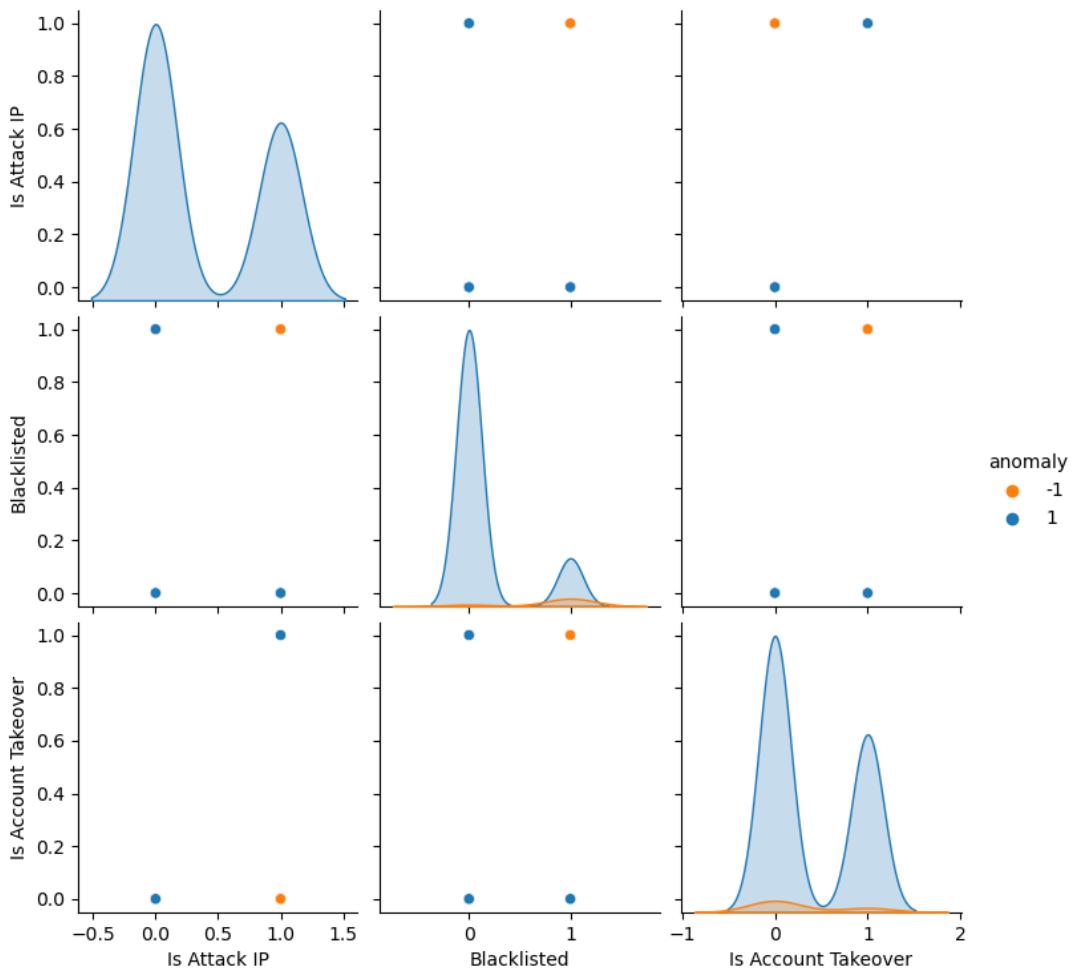
- anomaly\_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 11
Number of non anomalous values 189
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x453e9f790>

```



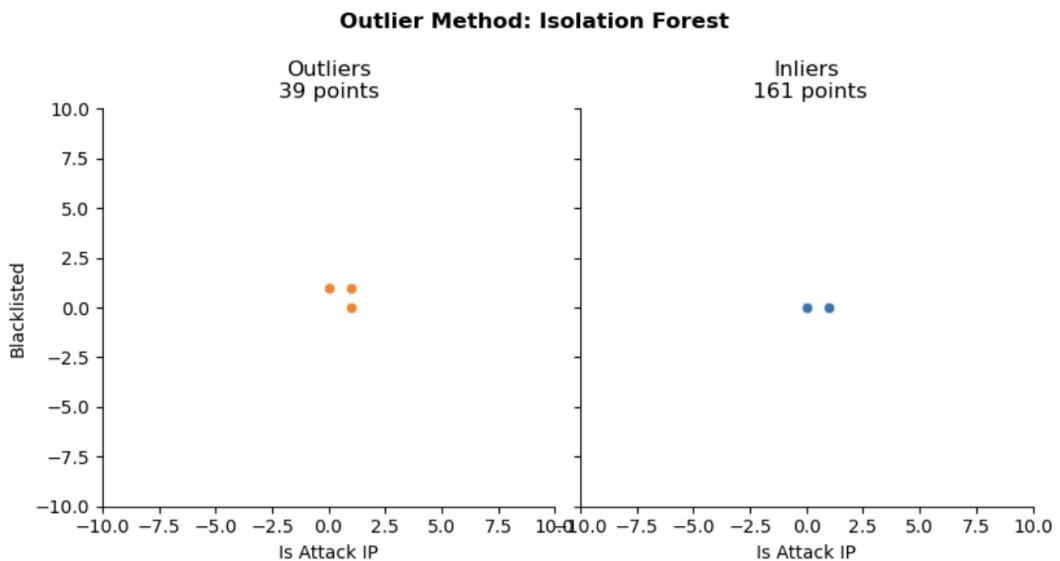


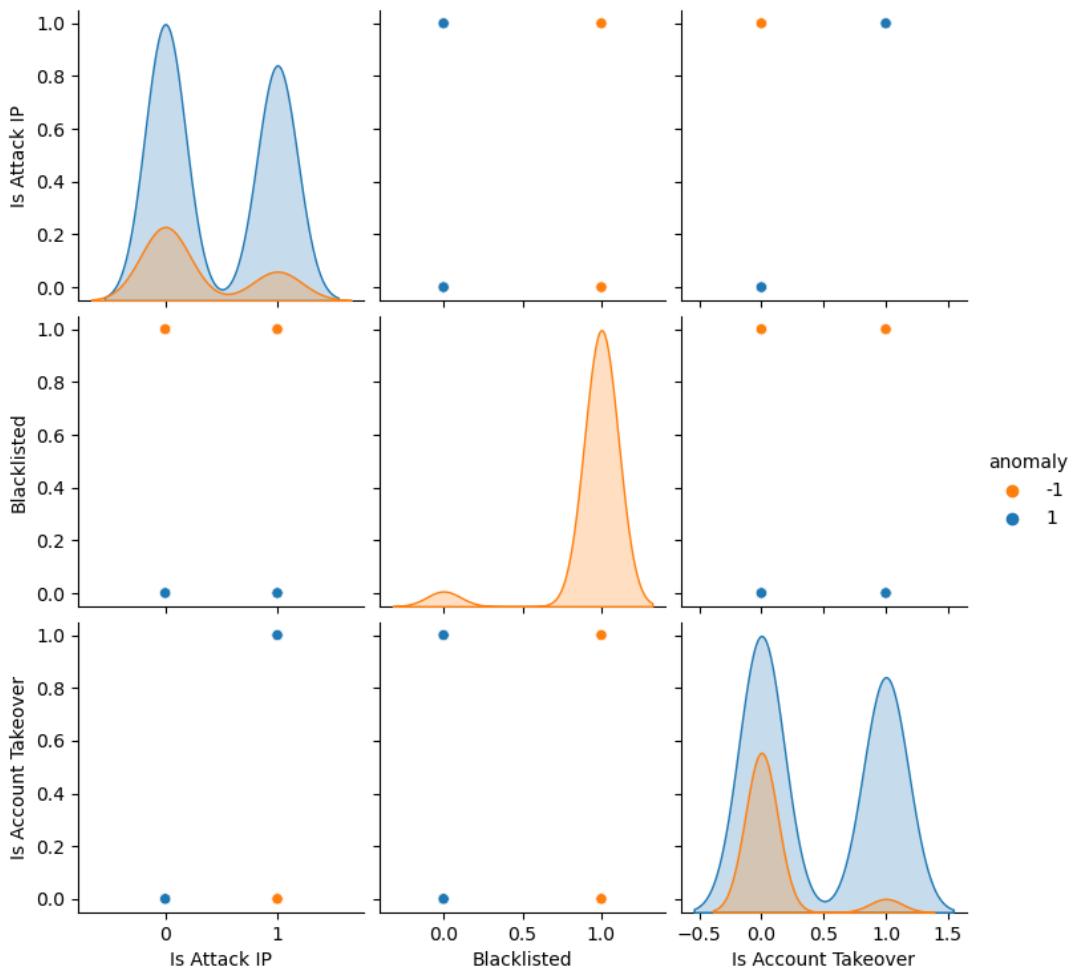
- increasing contamination value to 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 39
Number of non anomalous values 161
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x4544c39d0>

```

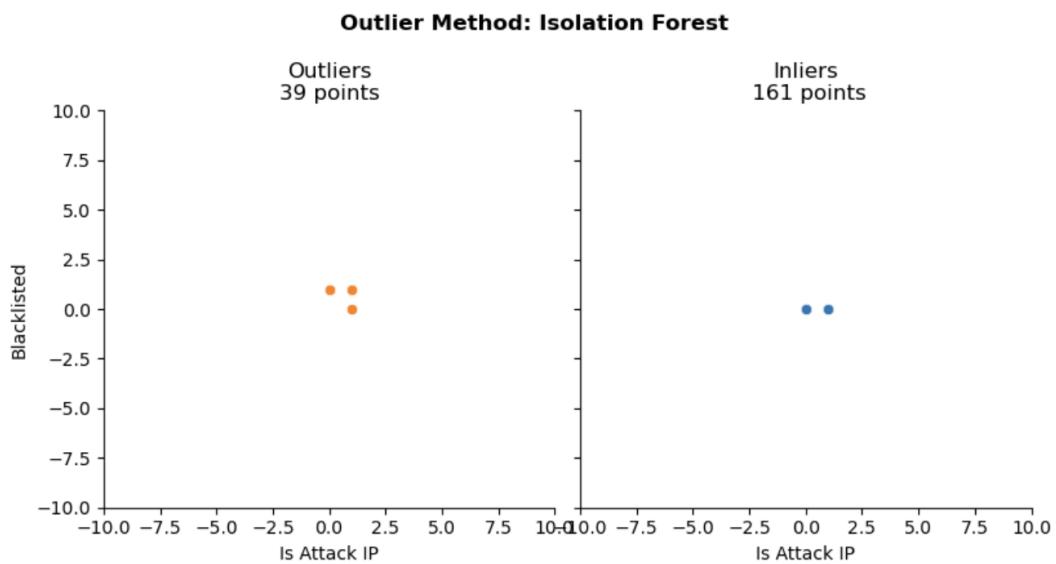


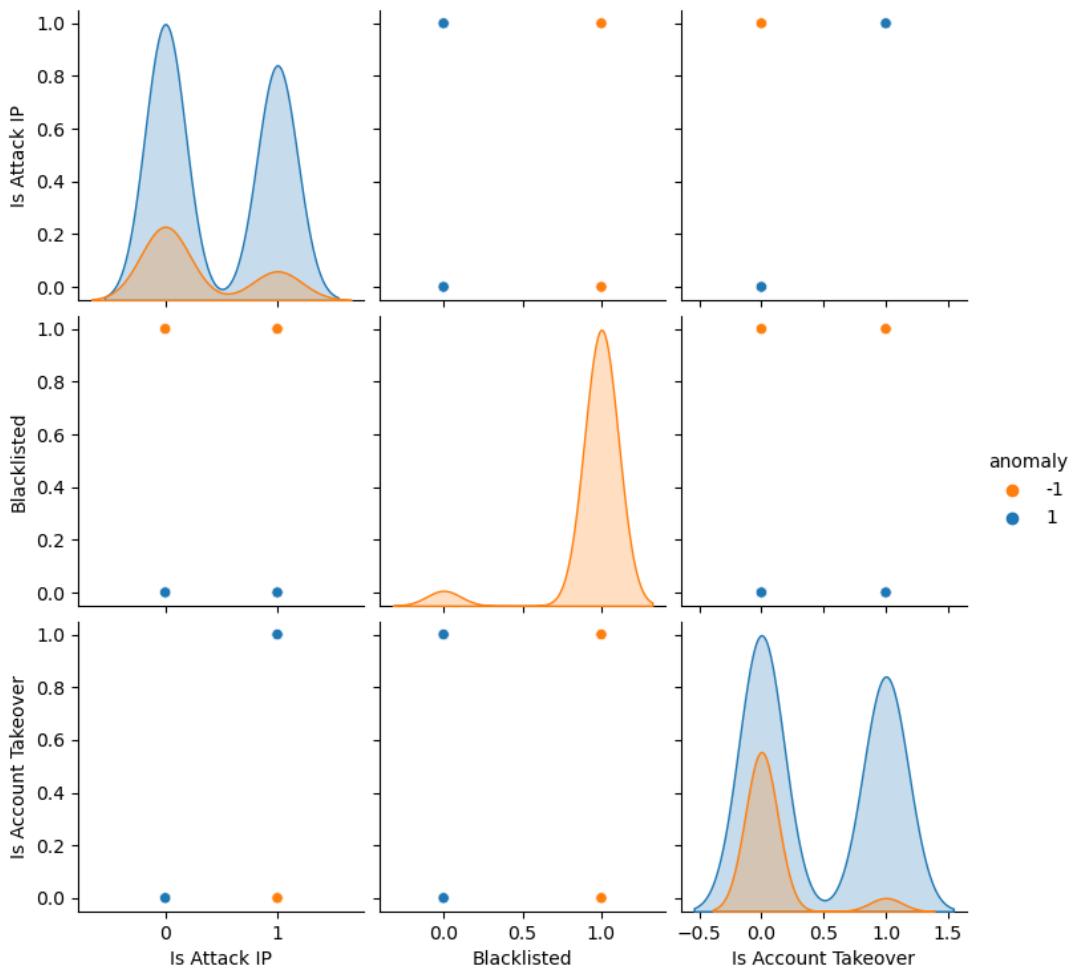


- contamination value == 0.5

```
Outlier Method: Isolation Forest
Number of anomalous values 39
Number of non anomalous values 161
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x453156950>
```



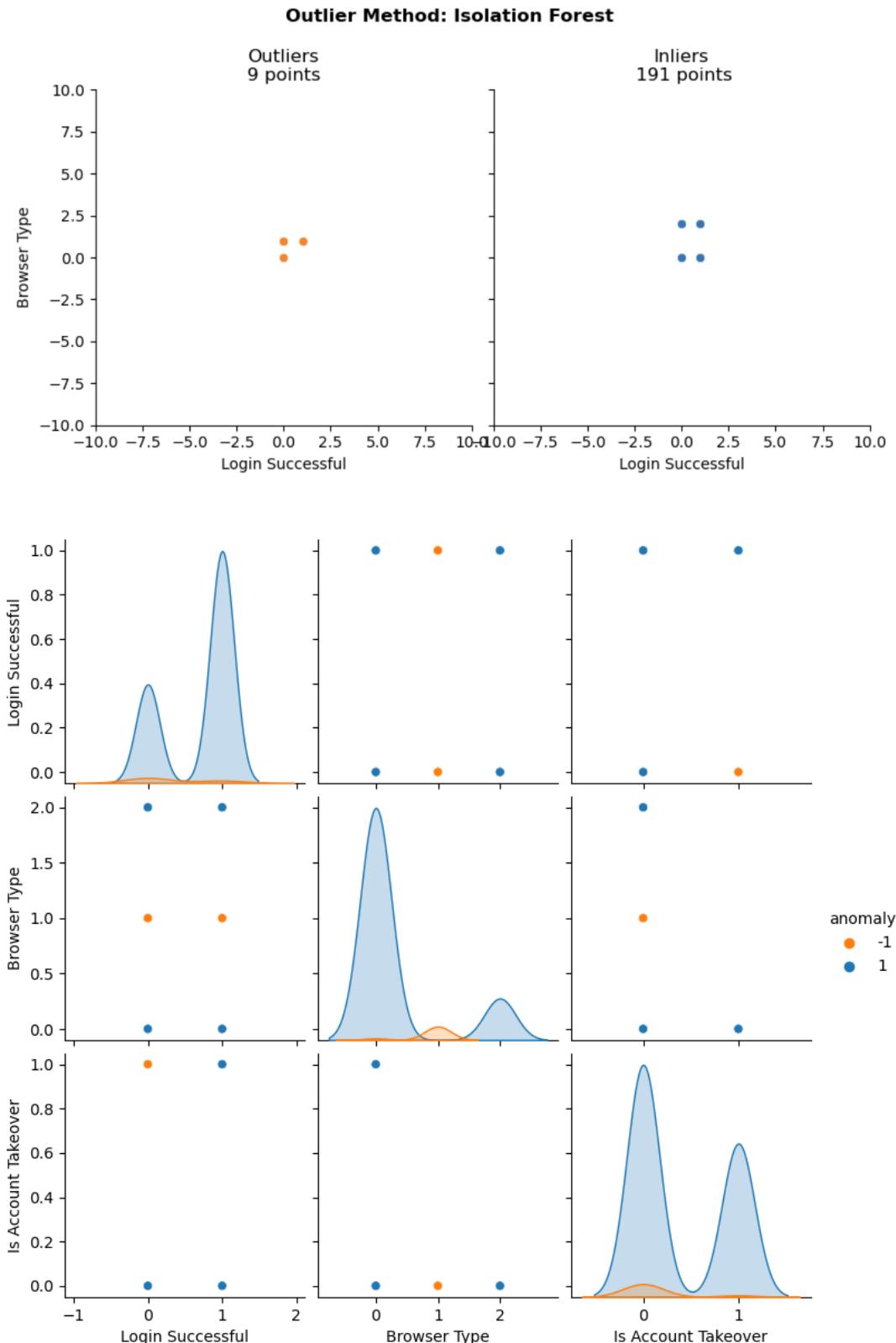


- `anomaly_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']`
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 9
Number of non anomalous values 191
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x4503aa910>

```



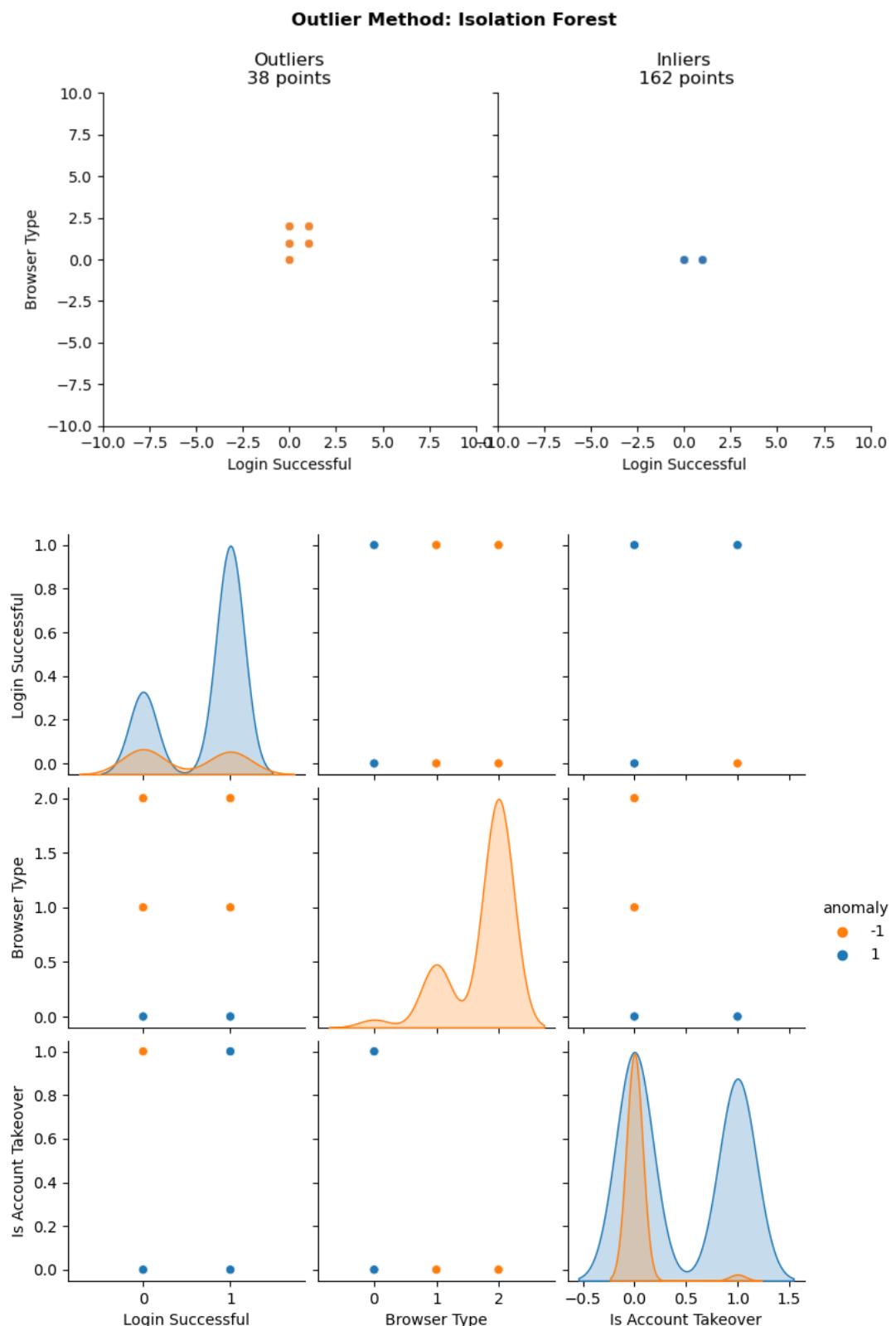
- contamination value == 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 38
Number of non anomalous values 162
Total Number of Values: 200

```

<seaborn.axisgrid.PairGrid at 0x45170fb50>



- contamination value == 0.5

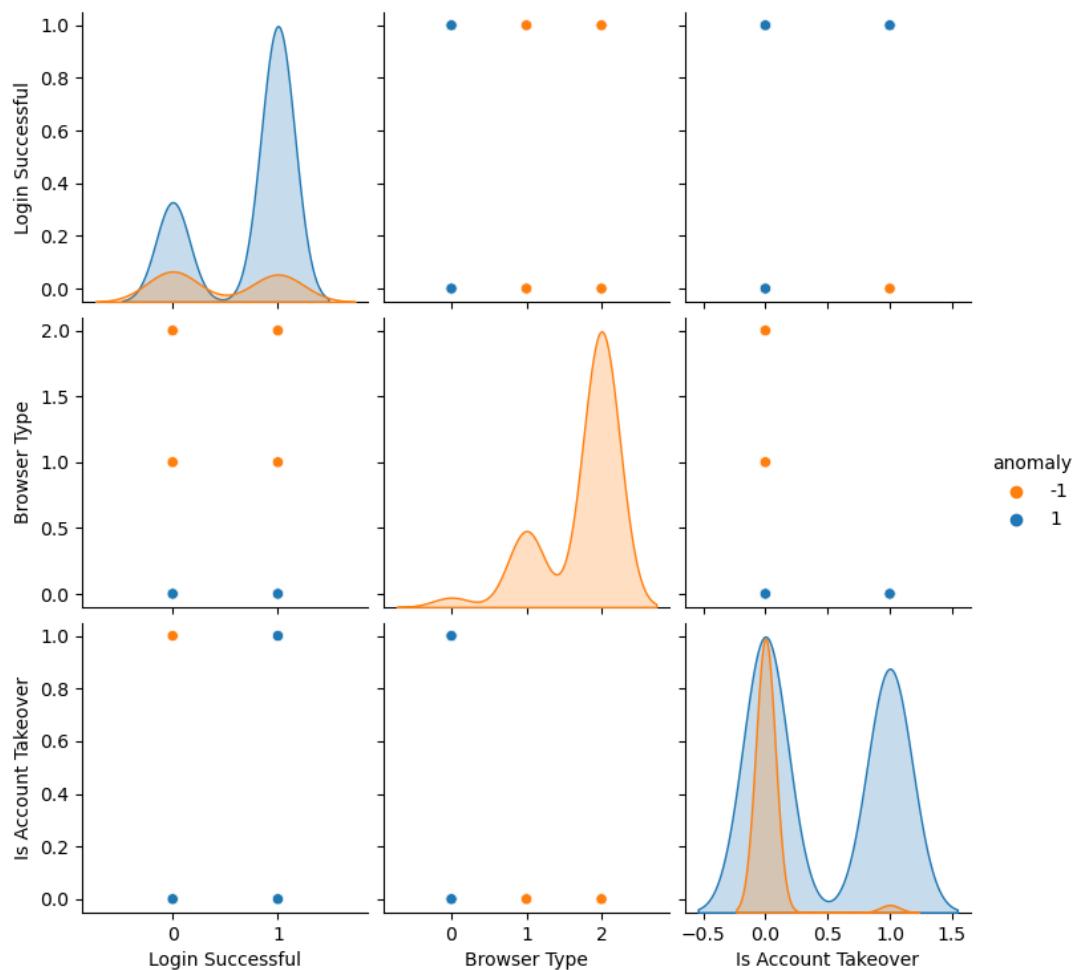
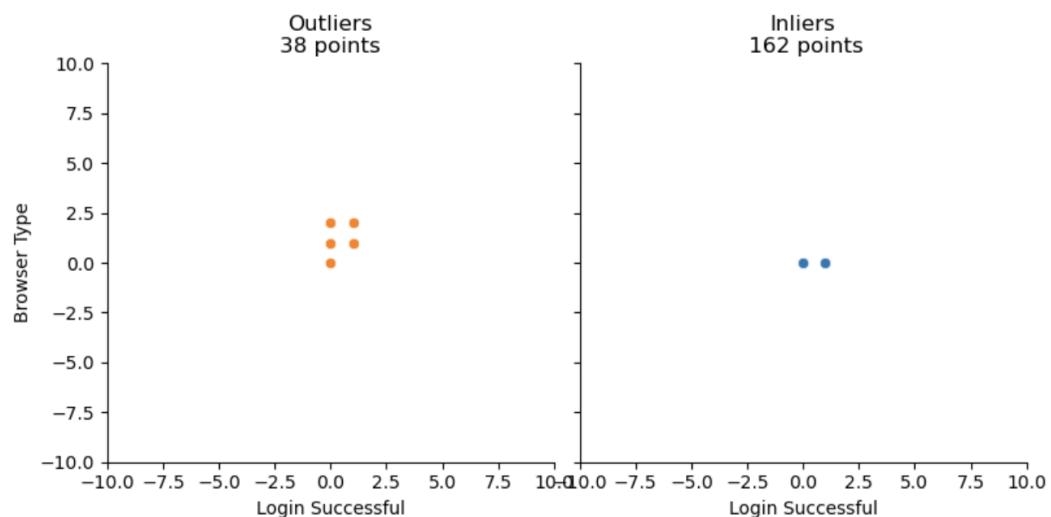
```

Outlier Method: Isolation Forest
Number of anomalous values 38
Number of non anomalous values 162
Total Number of Values: 200

```

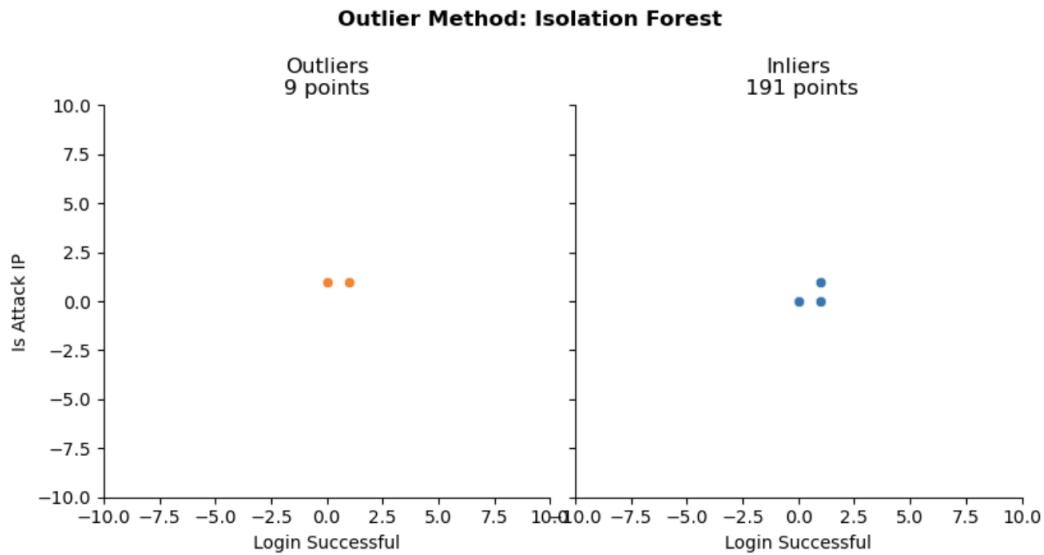
<seaborn.axisgrid.PairGrid at 0x44f691690>

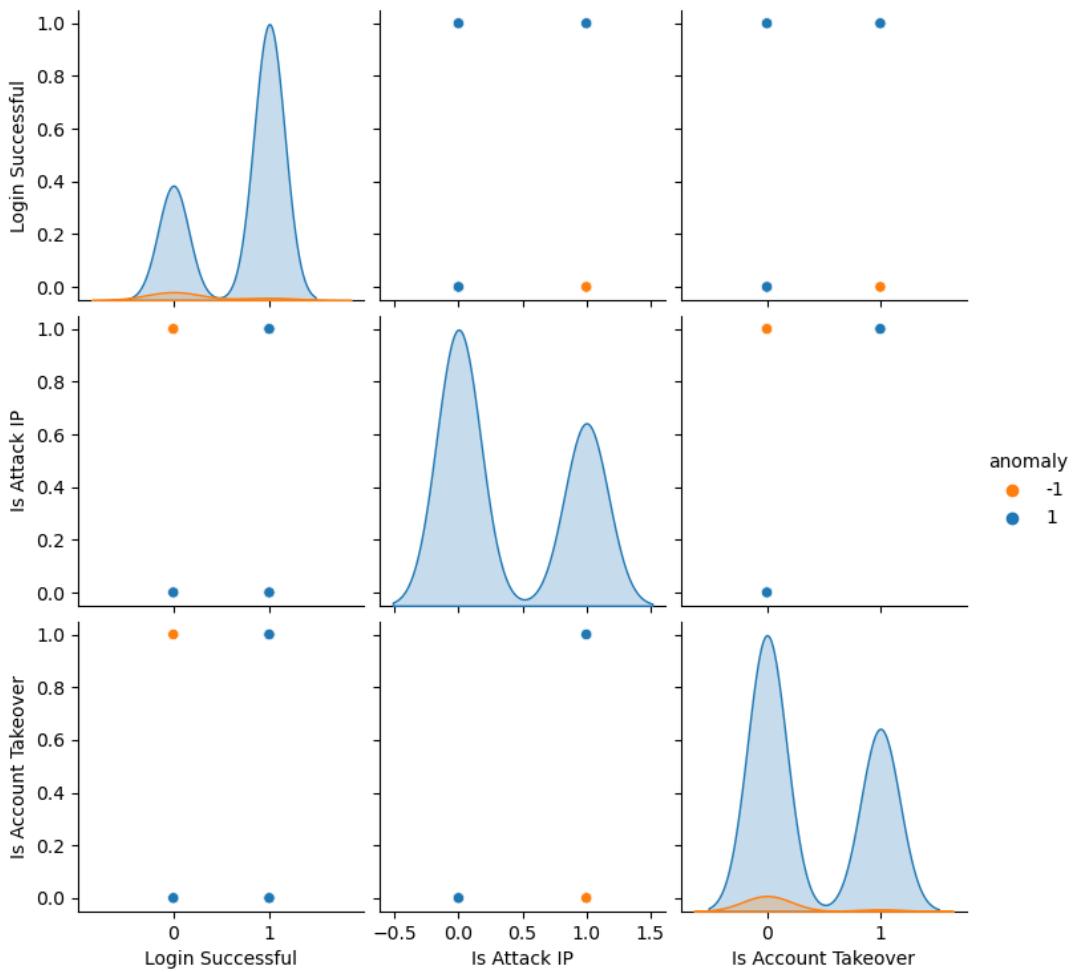
### Outlier Method: Isolation Forest



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

```
Outlier Method: Isolation Forest
Number of anomalous values 9
Number of non anomalous values 191
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x44fc17b50>
```



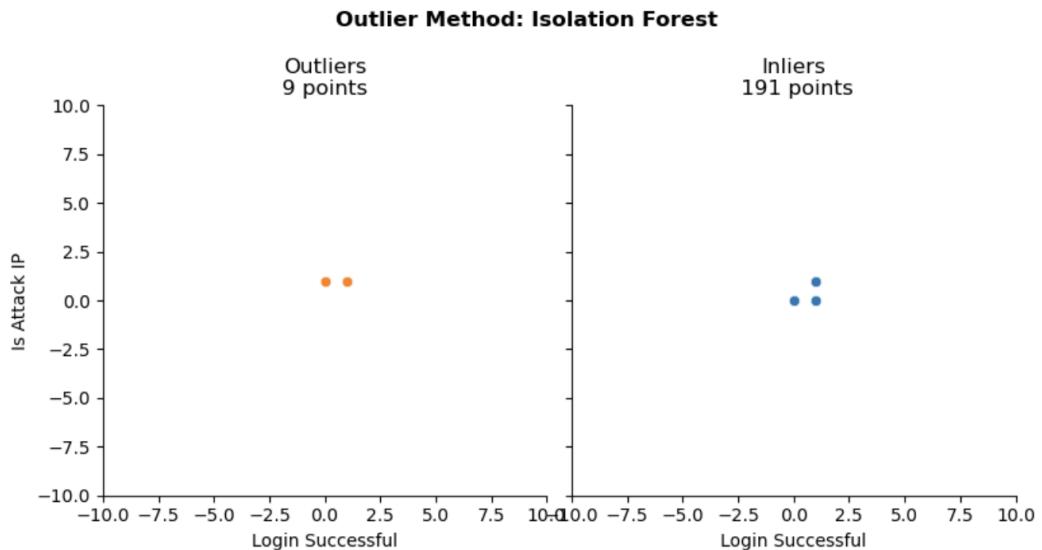


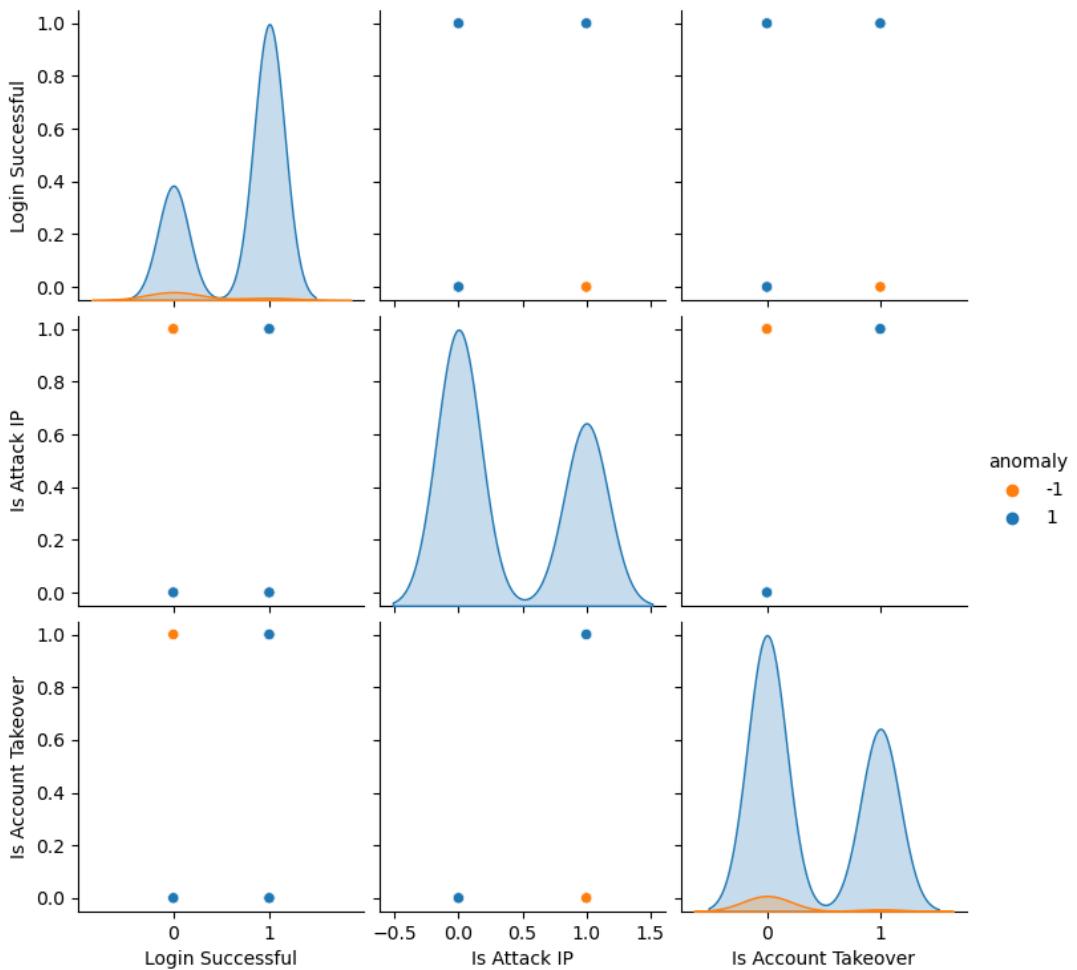
- contamination value == 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 9
Number of non anomalous values 191
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x4246360d0>

```

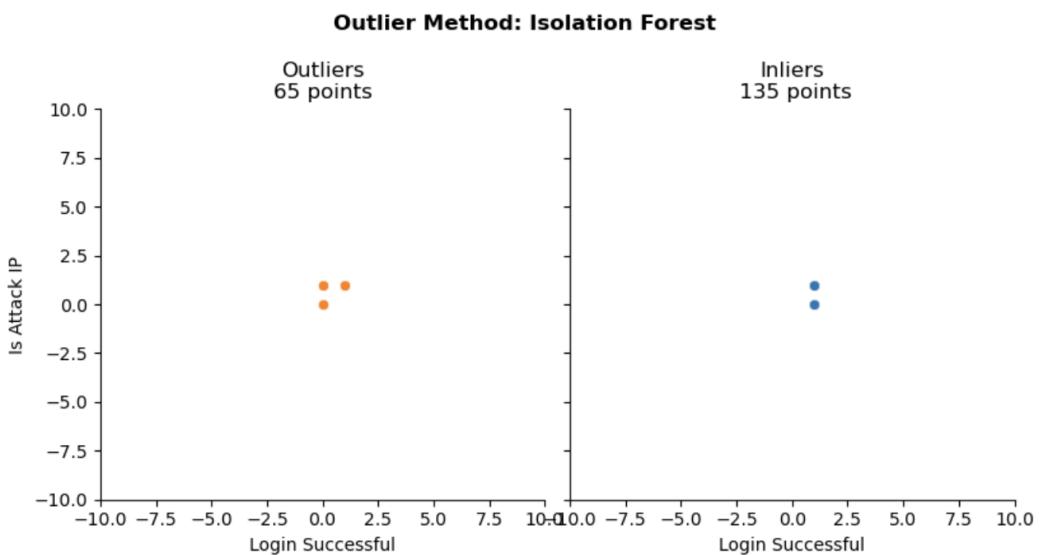


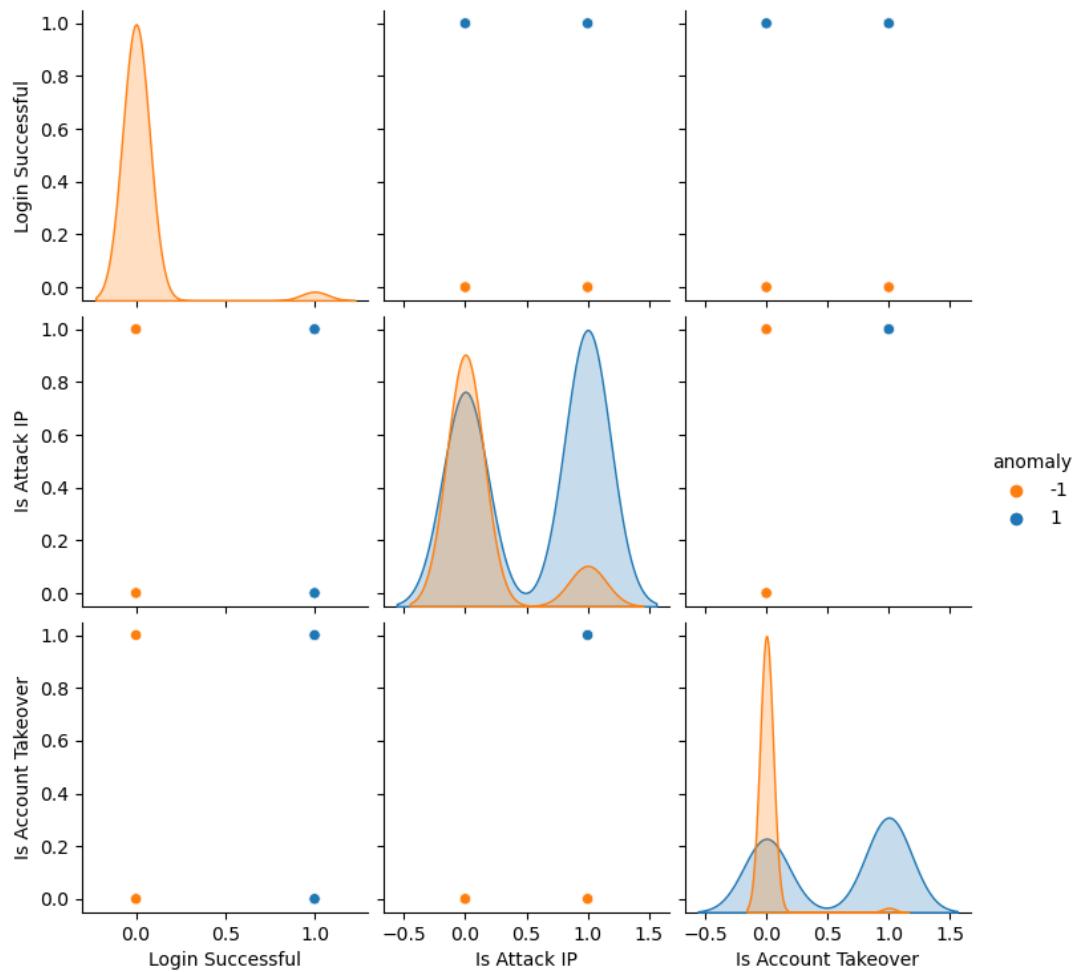


- contamination value == 0.5

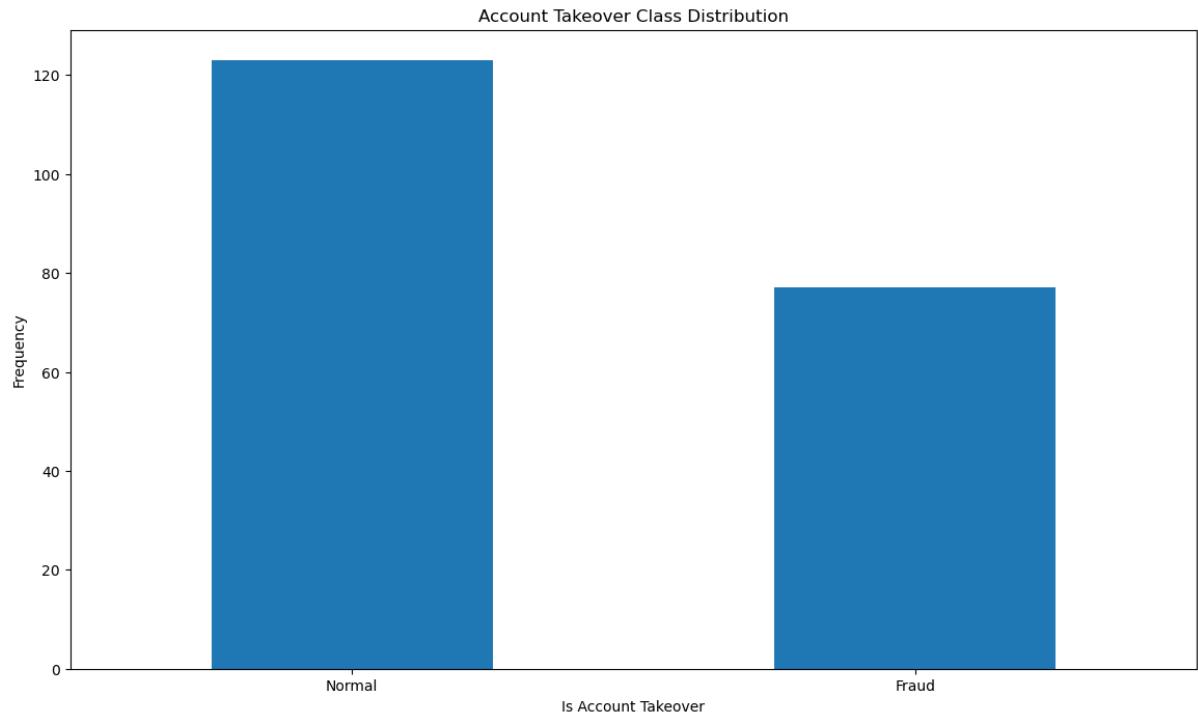
```
Outlier Method: Isolation Forest
Number of anomalous values 65
Number of non anomalous values 135
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x30ce4b1d0>
```





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



"{}: {}".format(clf\_name,n\_errors) → np. Isolation Forest: 21

Isolation Forest: 91

Accuracy Score :

0.545

Classification Report :

	precision	recall	f1-score	support
False	0.59	0.86	0.70	123
True	0.15	0.04	0.06	77
accuracy			0.55	200
macro avg	0.37	0.45	0.38	200
weighted avg	0.42	0.55	0.45	200

Local Outlier Factor: 81

Accuracy Score :

0.595

Classification Report :

	precision	recall	f1-score	support
False	0.62	0.90	0.73	123
True	0.40	0.10	0.16	77
accuracy			0.59	200
macro avg	0.51	0.50	0.45	200
weighted avg	0.53	0.59	0.51	200

Support Vector Machine: 123

Accuracy Score :

0.385

Classification Report :

	precision	recall	f1-score	support
False	0.50	0.04	0.08	123
True	0.38	0.94	0.54	77
accuracy			0.38	200
macro avg	0.44	0.49	0.31	200
weighted avg	0.45	0.39	0.25	200

**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"])

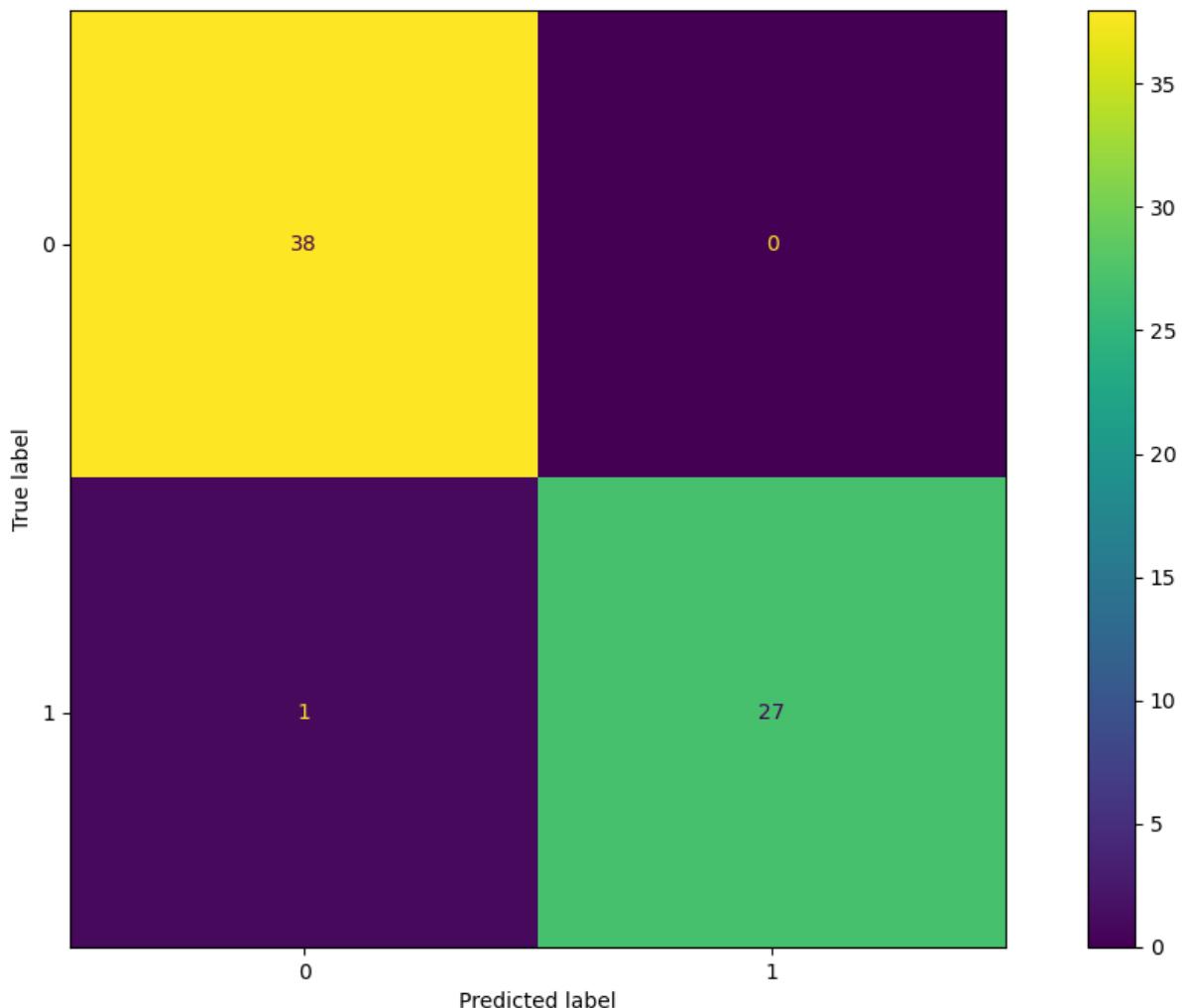
### 3A. XGBoost

memory usage: 80.6 KB

Accuracy: 98.48%

	precision	recall	f1-score	support
False	0.97	1.00	0.99	38
True	1.00	0.96	0.98	28
accuracy			0.98	66
macro avg	0.99	0.98	0.98	66
weighted avg	0.99	0.98	0.98	66

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 160
The number of records in the test dataset is 40
The training dataset has 98 records for the majority class and 62 records for the minority class.
      precision    recall   f1-score   support
  False        0.75     0.12     0.21      25
  True        0.39     0.93     0.55      15

accuracy                           0.42      40
macro avg       0.57     0.53     0.38      40
weighted avg    0.61     0.42     0.34      40

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.62     1.00     0.77      25
  True        0.00     0.00     0.00      15

accuracy                           0.62      40
macro avg       0.31     0.50     0.38      40
weighted avg    0.39     0.62     0.48      40

[[ 3 22]
 [ 1 14]]
```

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

```
The number of records in the training dataset is 140
The number of records in the test dataset is 60
The training dataset has 86 records for the majority class and 54 records for the minority class.
      precision    recall   f1-score   support
  False        1.00     0.03     0.05      37
  True        0.39     1.00     0.56      23

accuracy                           0.40      60
macro avg       0.69     0.51     0.31      60
weighted avg    0.77     0.40     0.25      60

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.62     1.00     0.76      37
  True        0.00     0.00     0.00      23

accuracy                           0.62      60
macro avg       0.31     0.50     0.38      60
weighted avg    0.38     0.62     0.47      60

[[ 1 36]
 [ 0 23]]
```

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

```

The number of records in the training dataset is 100
The number of records in the test dataset is 100
The training dataset has 62 records for the majority class and 38 records for the minority class.
      precision    recall   f1-score   support
  False        0.00     0.00     0.00      61
  True        0.39     1.00     0.56      39

      accuracy          0.39      100
   macro avg       0.20     0.50     0.28      100
weighted avg     0.15     0.39     0.22      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
  False        0.61     1.00     0.76      61
  True        0.00     0.00     0.00      39

      accuracy          0.61      100
   macro avg       0.30     0.50     0.38      100
weighted avg     0.37     0.61     0.46      100

[[ 0 61]
 [ 0 39]]

```

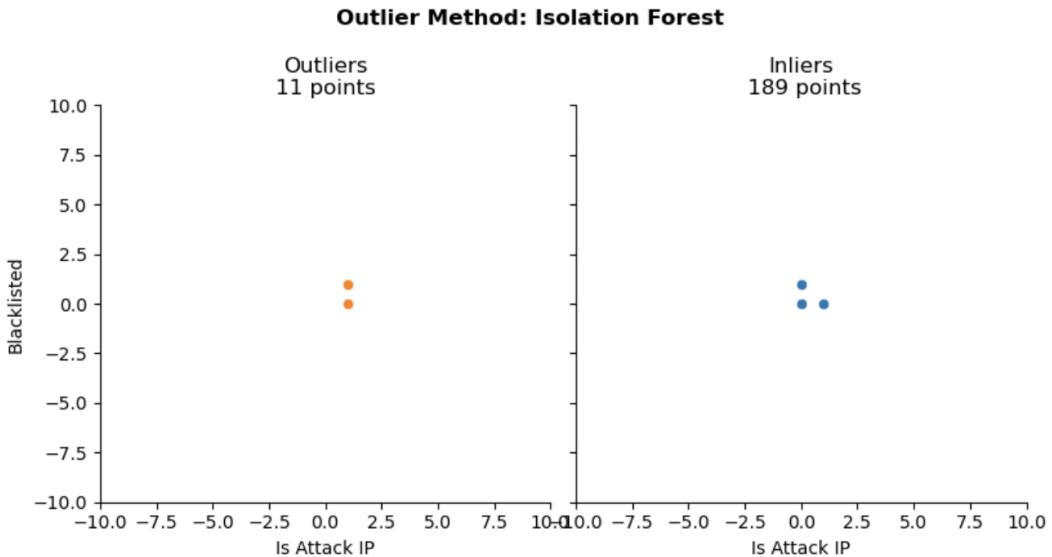
### 3C. Isolation Forest

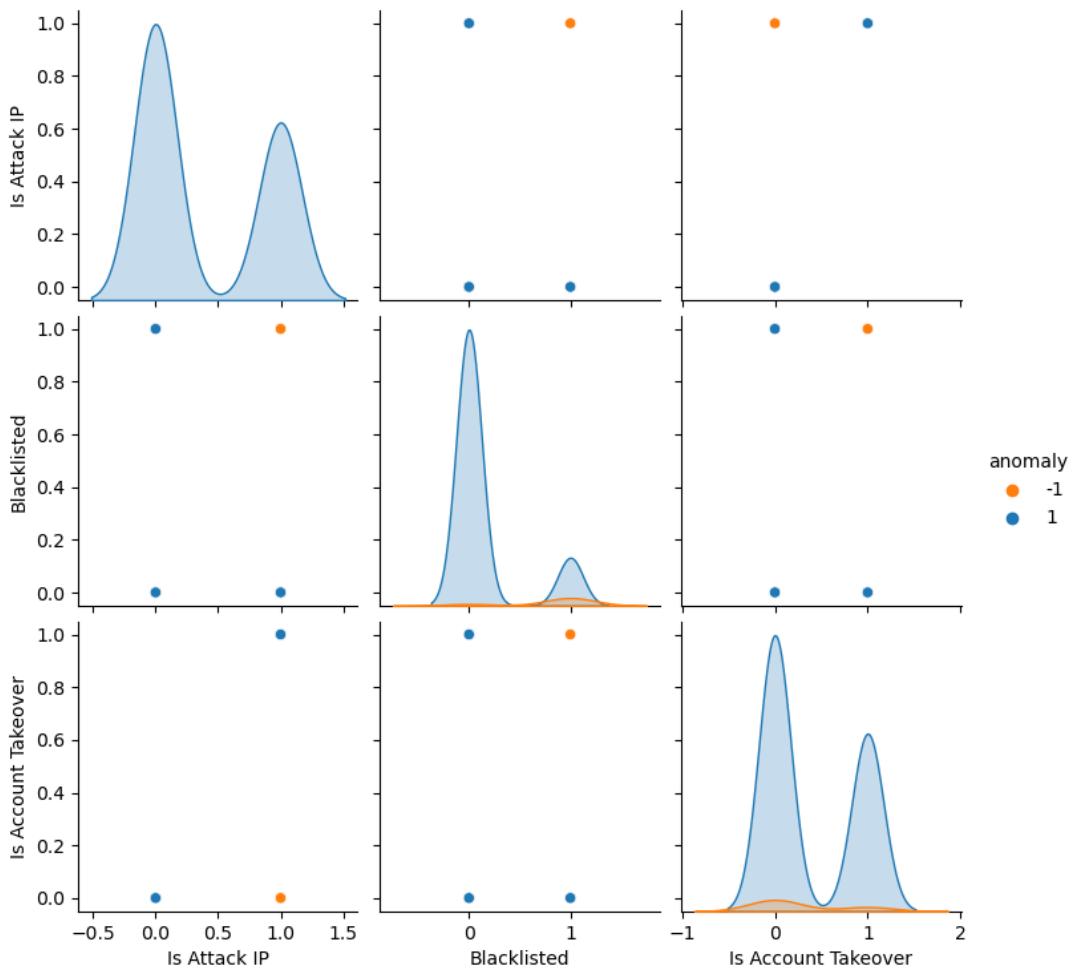
- anomaly\_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 11
Number of non anomalous values 189
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x44e910550>

```



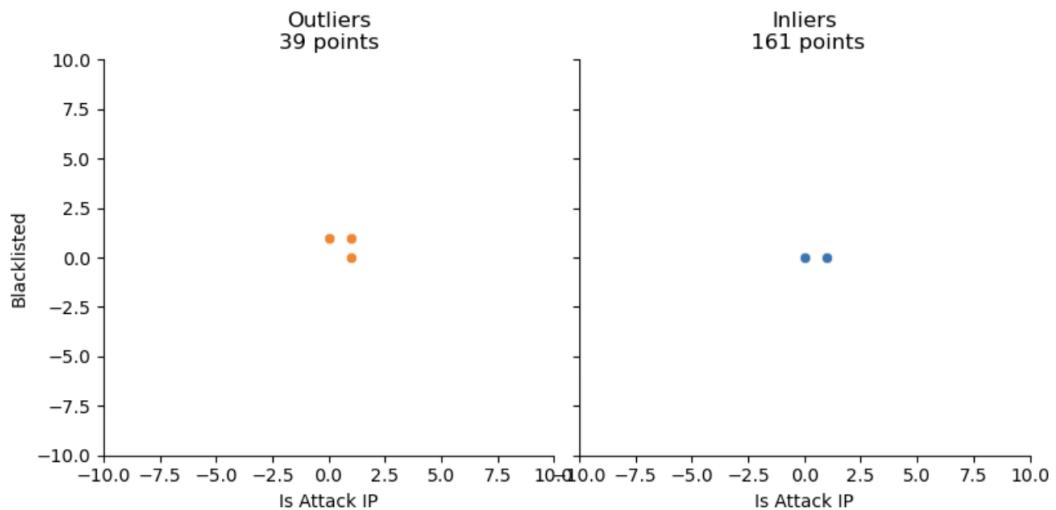


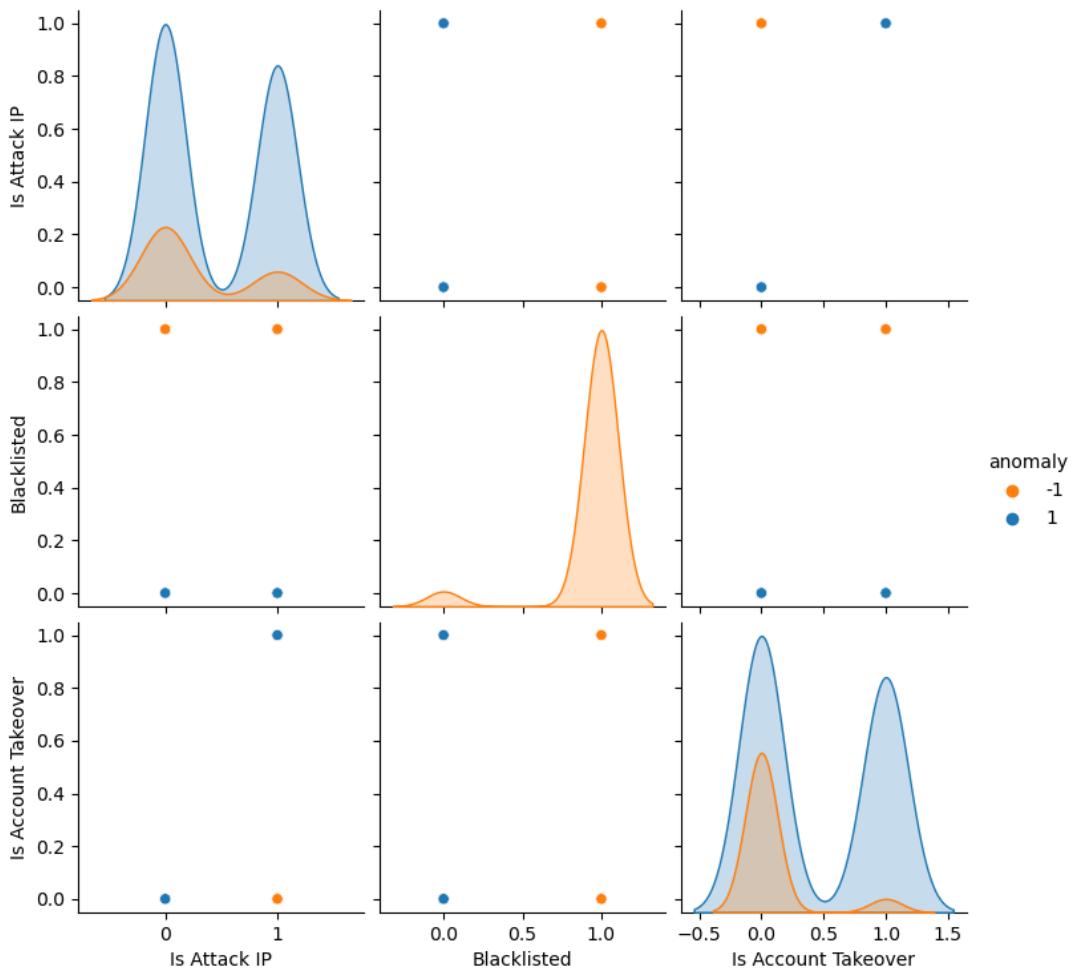
- contamination value == 0.3

```
Outlier Method: Isolation Forest
Number of anomalous values 39
Number of non anomalous values 161
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x44dc30610>
```

### Outlier Method: Isolation Forest



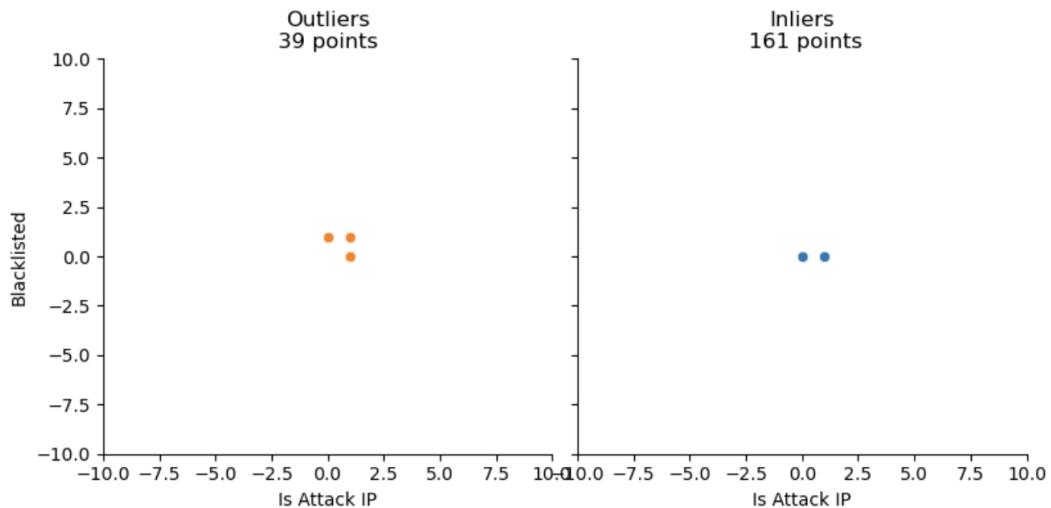


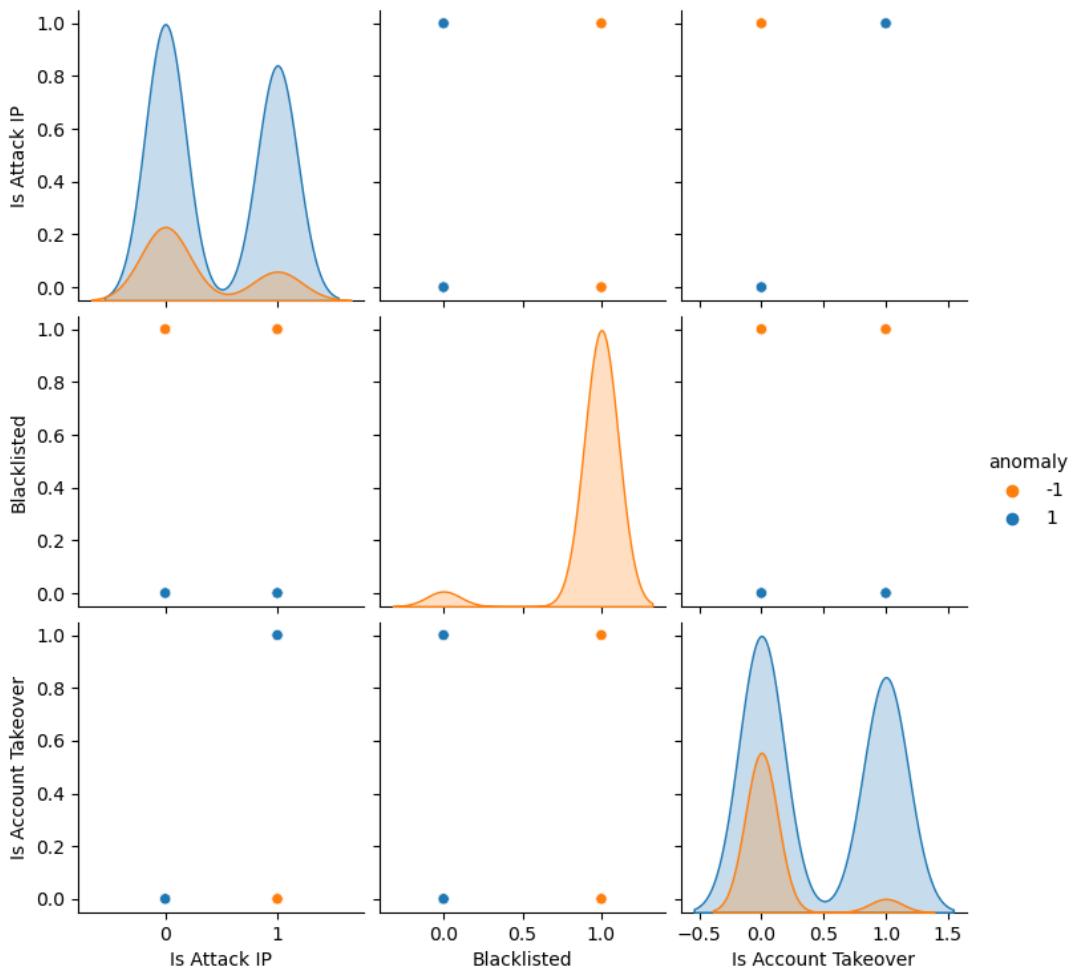
- contamination value == 0.5

```
Outlier Method: Isolation Forest
Number of anomalous values 39
Number of non anomalous values 161
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x44cf22c10>
```

#### Outlier Method: Isolation Forest





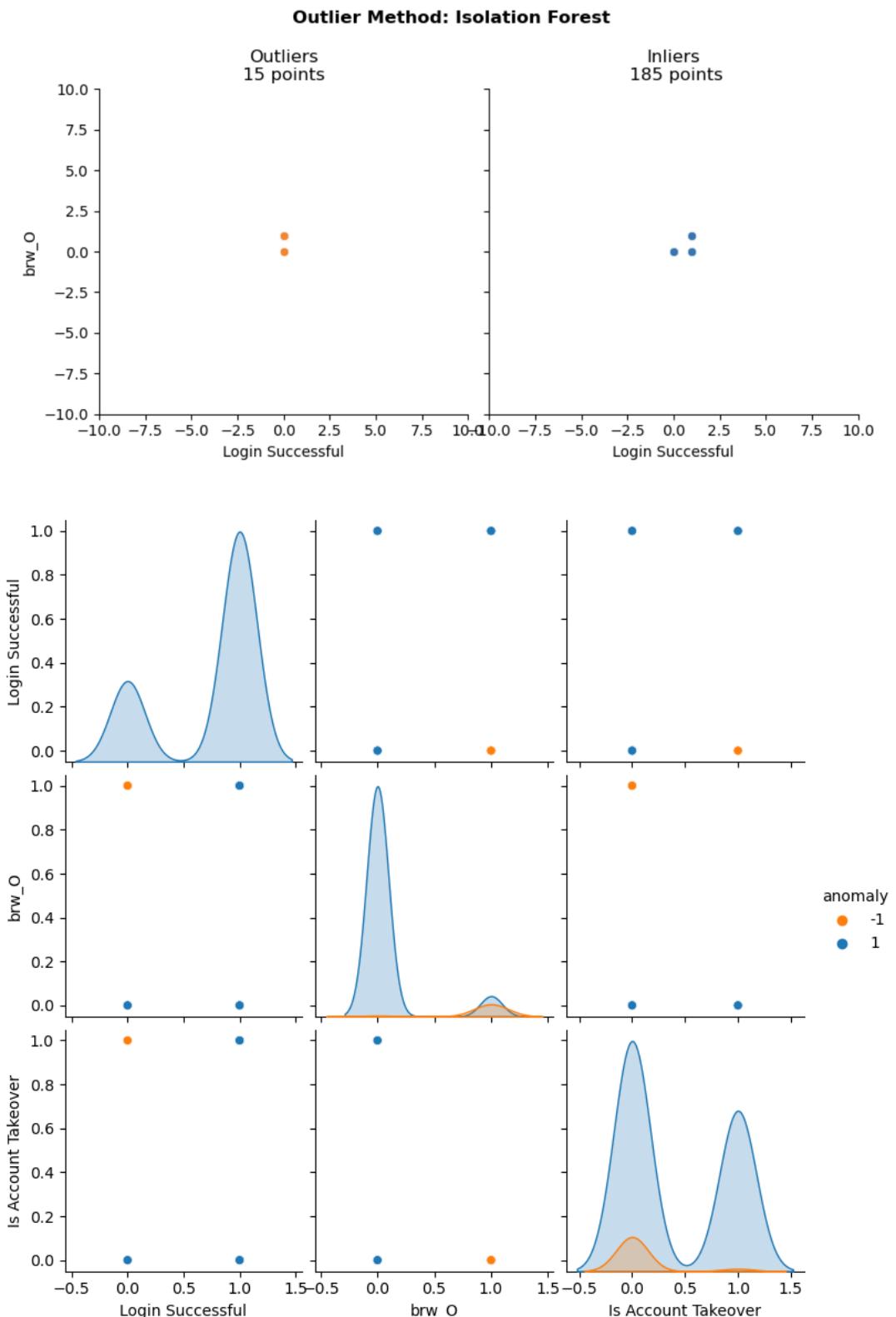
- anomaly\_inputs = ['Login Successful', 'Browser Type (brw\_O)', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 15
Number of non anomalous values 185
Total Number of Values: 200

```

<seaborn.axisgrid.PairGrid at 0x44c1d4050>



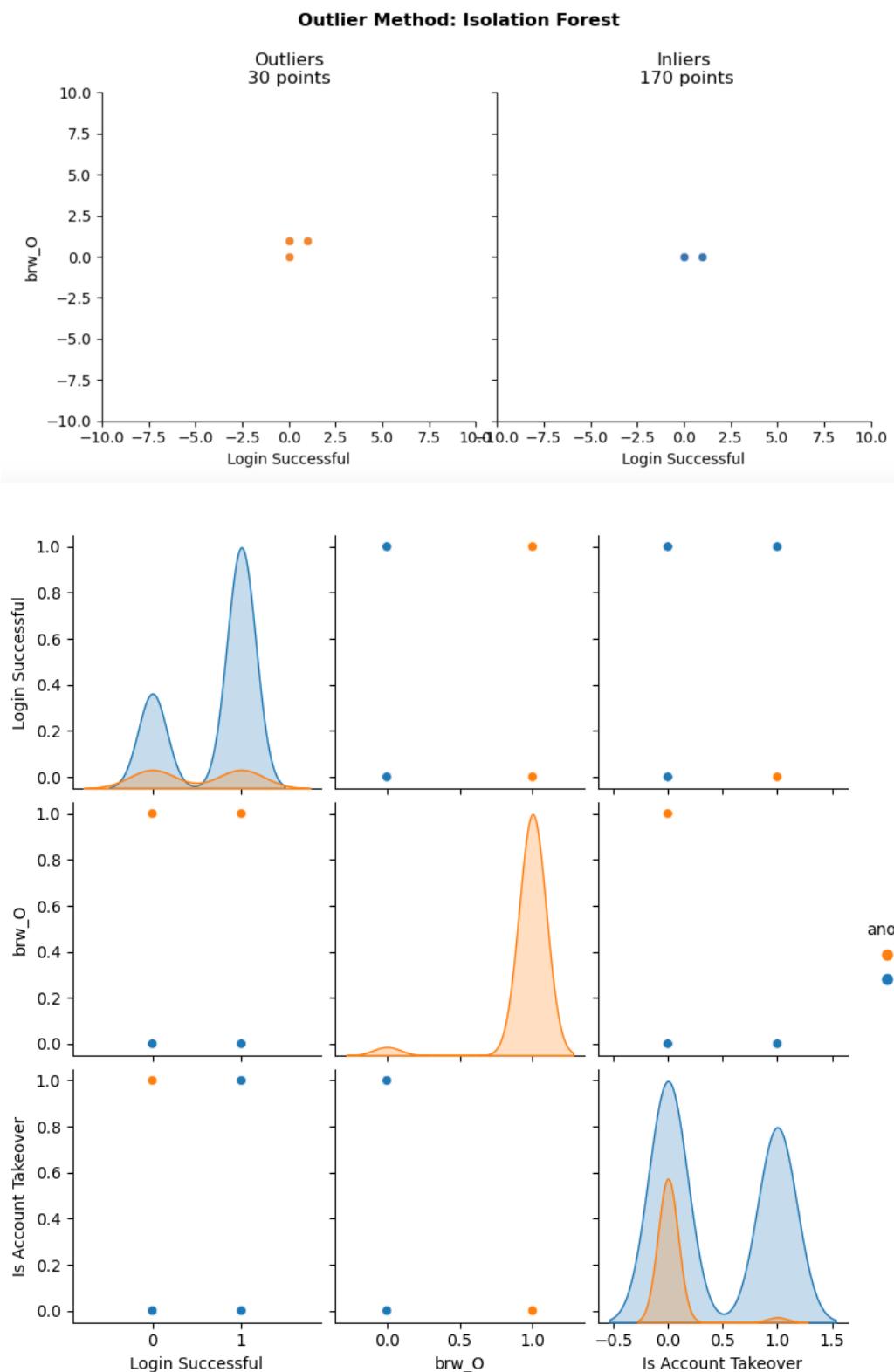
- contamination value == 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 30
Number of non anomalous values 170
Total Number of Values: 200

```

<seaborn.axisgrid.PairGrid at 0x44c155d10>



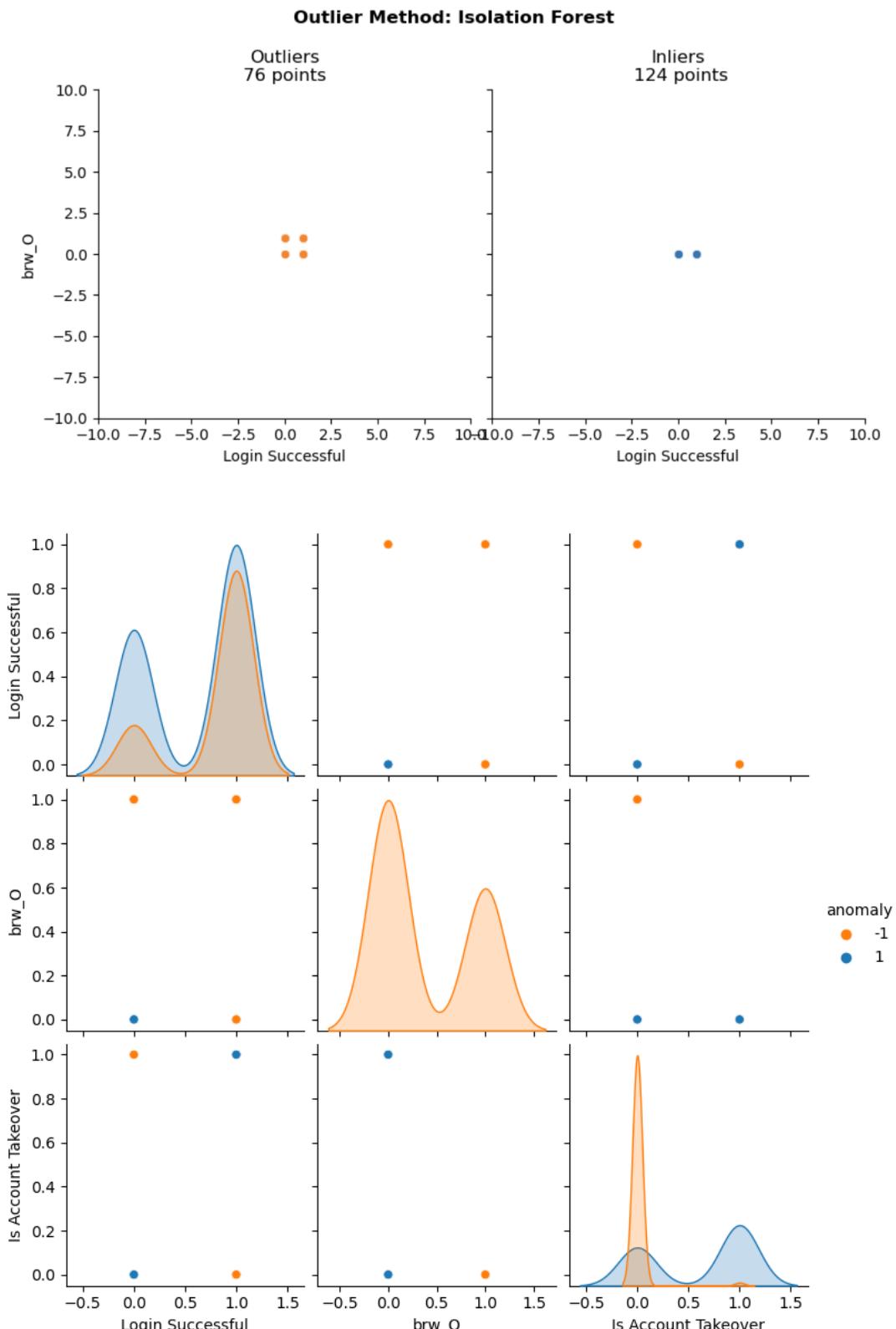
- contamination value == 0.5

```

Outlier Method: Isolation Forest
Number of anomalous values 76
Number of non anomalous values 124
Total Number of Values: 200

<seaborn.axisgrid.PairGrid at 0x44ad10ad0>

```



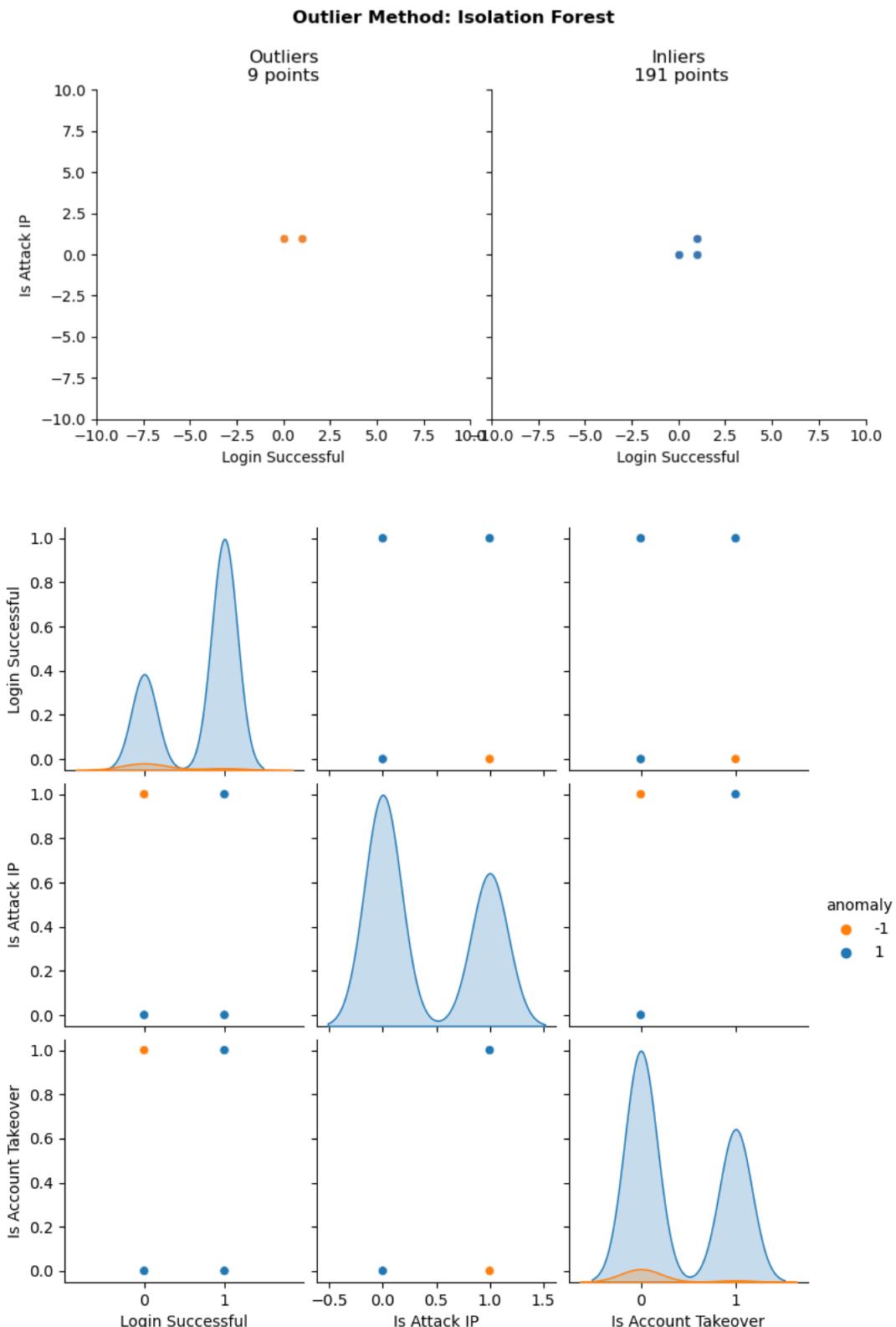
- anomaly\_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']

- contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 9
Number of non anomalous values 191
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x449fd8a90>

```

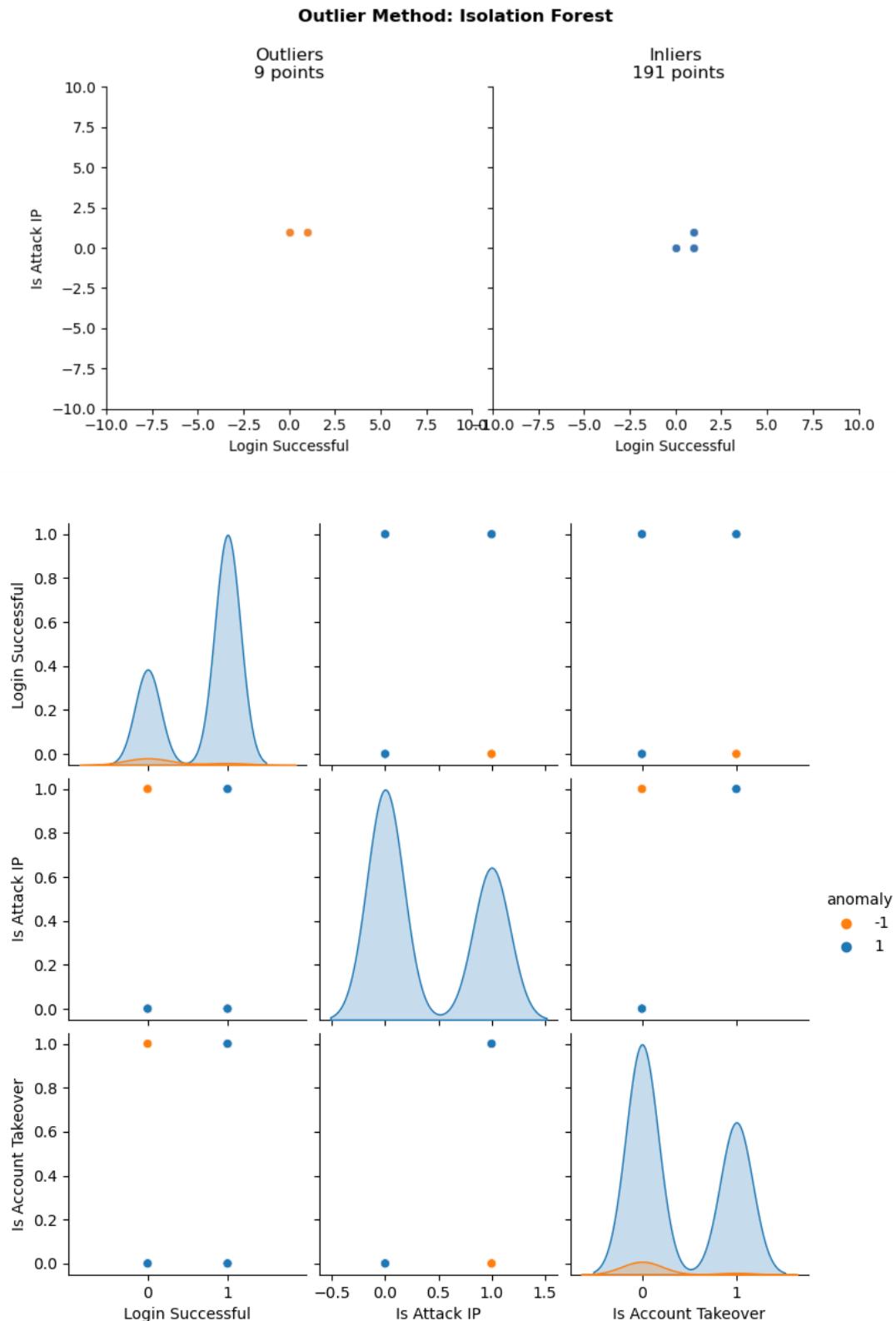


- contamination value == 0.3

```

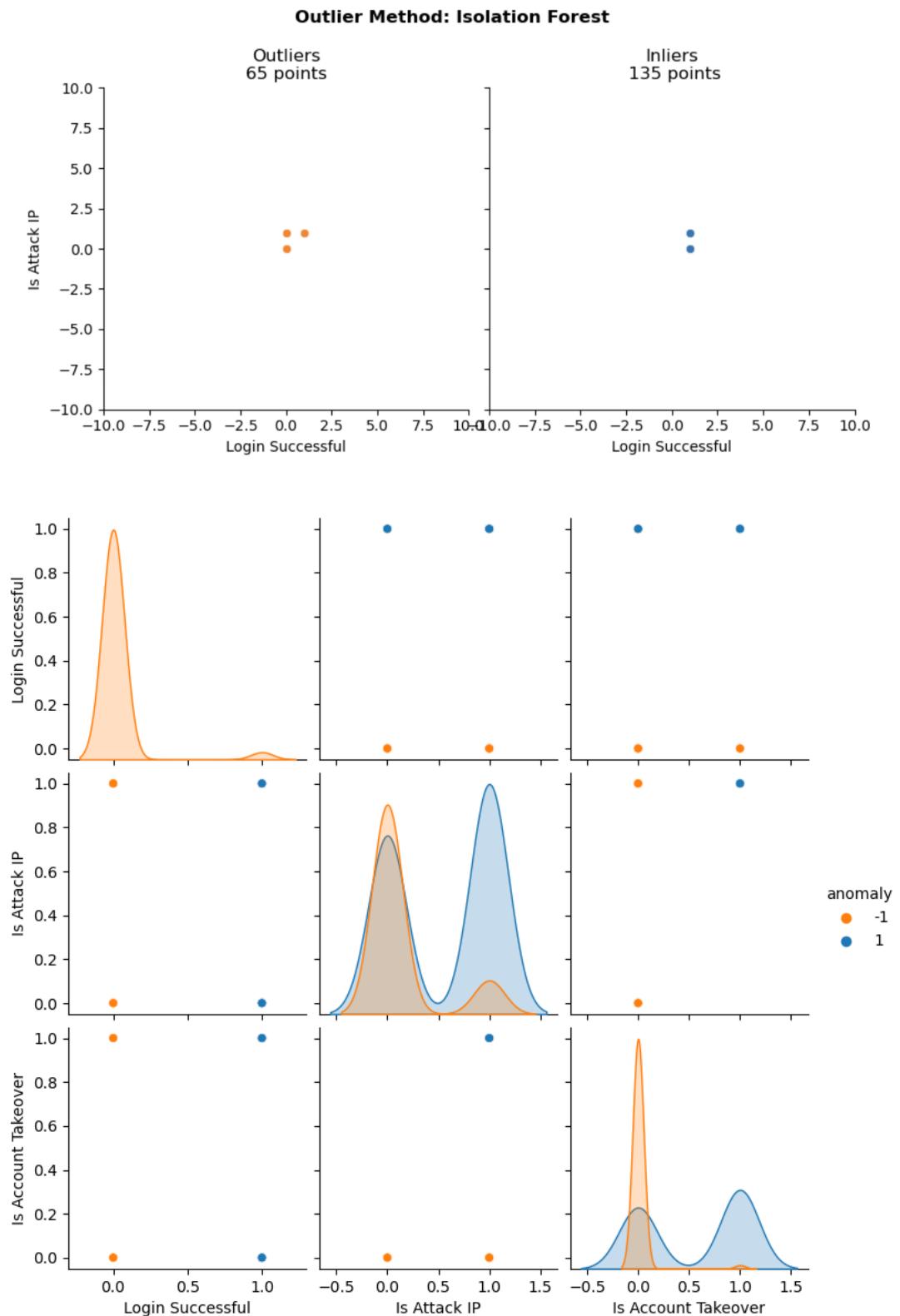
Outlier Method: Isolation Forest
Number of anomalous values 9
Number of non anomalous values 191
Total Number of Values: 200
: <seaborn.axisgrid.PairGrid at 0x44982b790>

```

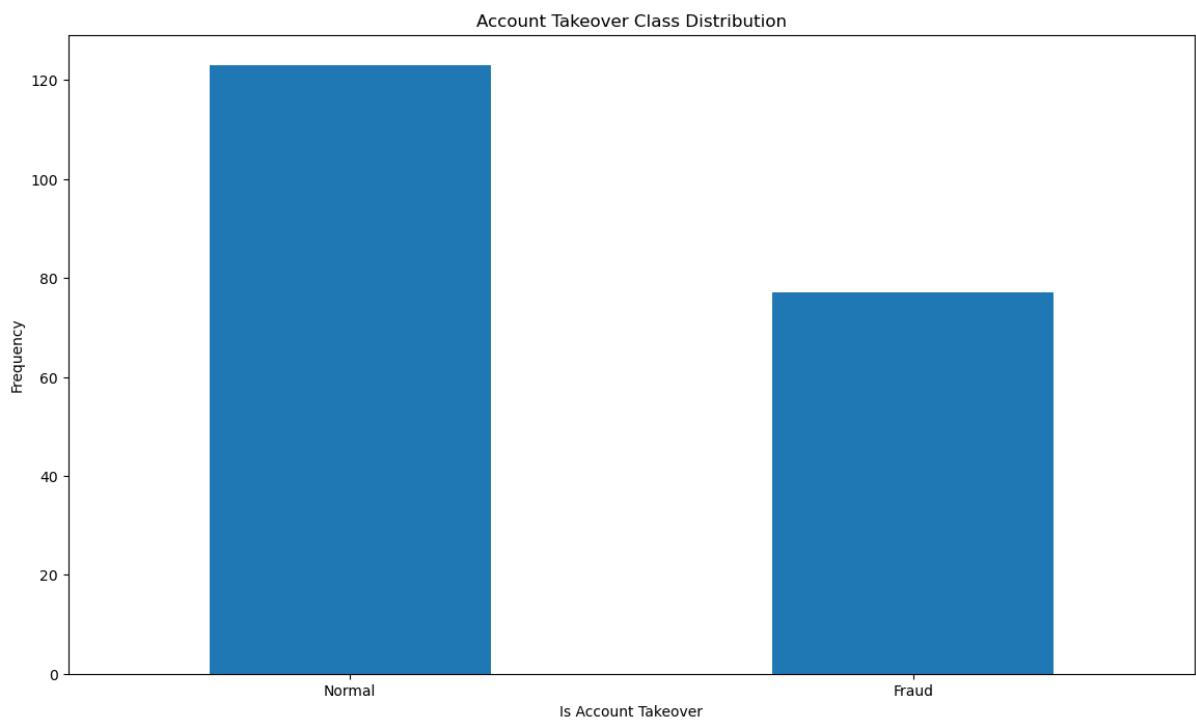


- contamination value == 0.5

```
Outlier Method: Isolation Forest
Number of anomalous values 65
Number of non anomalous values 135
Total Number of Values: 200
: <seaborn.axisgrid.PairGrid at 0x448b735d0>
```



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



"{}: {}".format(clf\_name,n\_errors) → np. Isolation Forest: 89

**Isolation Forest:** 89

Accuracy Score :

0.555

Classification Report :

	precision	recall	f1-score	support
False	0.59	0.87	0.71	123
True	0.20	0.05	0.08	77
accuracy			0.56	200
macro avg	0.40	0.46	0.39	200
weighted avg	0.44	0.56	0.47	200

**Local Outlier Factor:** 81

Accuracy Score :

0.595

Classification Report :

	precision	recall	f1-score	support
False	0.62	0.90	0.73	123
True	0.40	0.10	0.16	77
accuracy			0.59	200
macro avg	0.51	0.50	0.45	200
weighted avg	0.53	0.59	0.51	200

**Support Vector Machine:** 123

Accuracy Score :

0.385

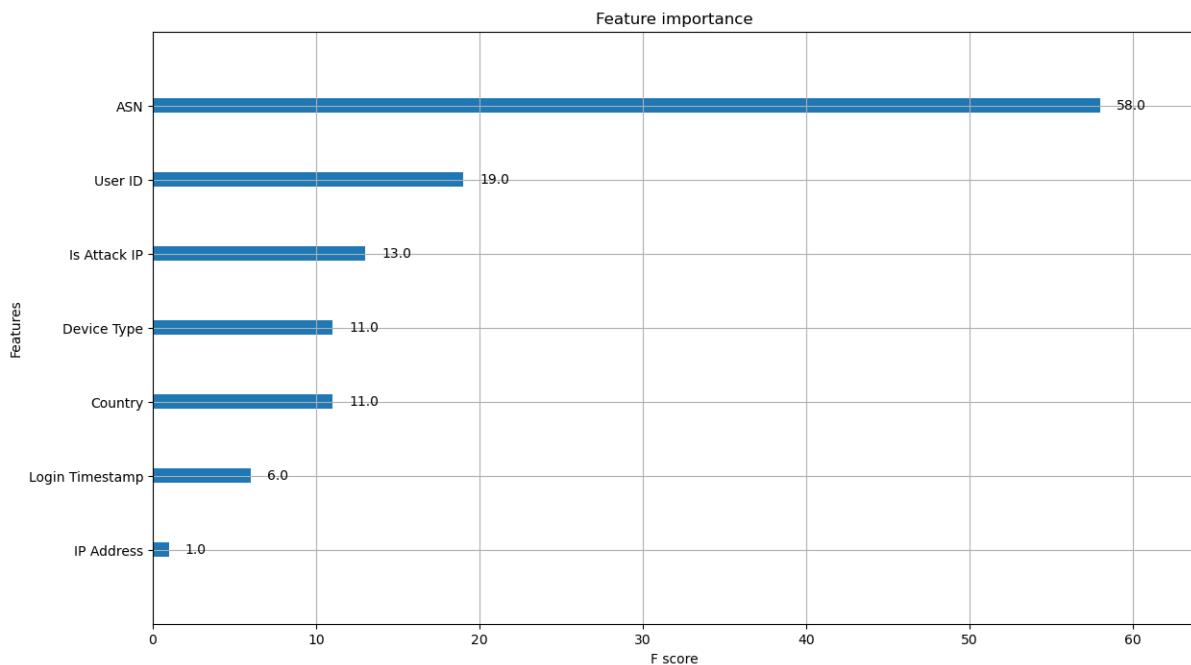
Classification Report :

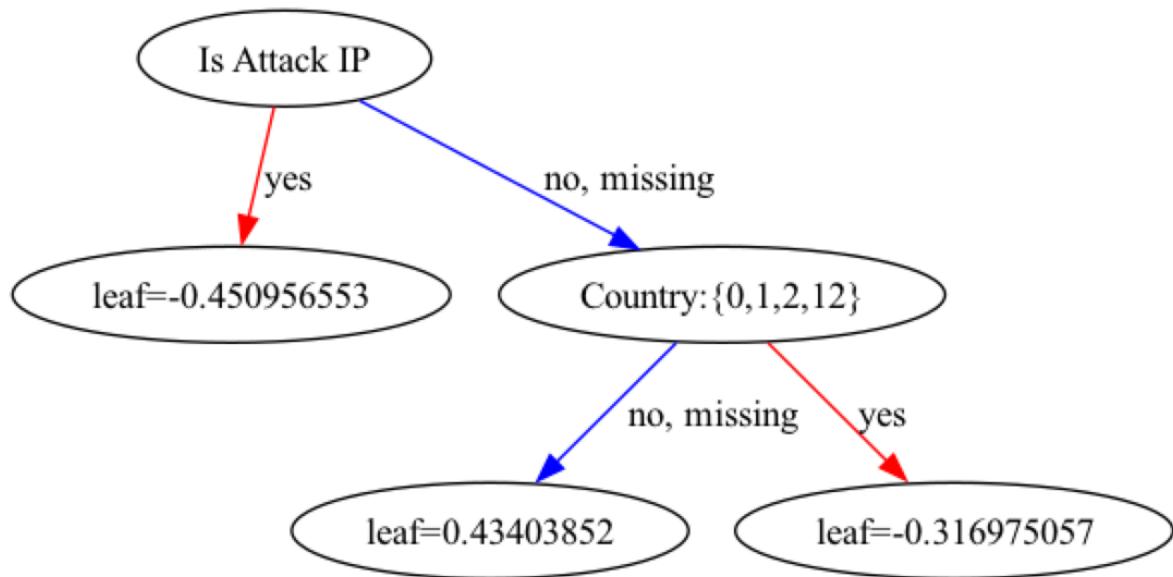
	precision	recall	f1-score	support
False	0.50	0.04	0.08	123
True	0.38	0.94	0.54	77
accuracy			0.38	200
macro avg	0.44	0.49	0.31	200
weighted avg	0.45	0.39	0.25	200

**IV. 141/200 → 70.5%**

# 1. XGB Experimental

```
Data columns (total 11 columns):
 #   Column           Non-Null Count Dtype  
 ---  -- 
 0   Login Timestamp    200 non-null   category
 1   User ID          200 non-null   int64    
 2   IP Address        200 non-null   category
 3   Country           200 non-null   category
 4   ASN               200 non-null   int64    
 5   Device Type       200 non-null   category
 6   Login Successful  200 non-null   bool    
 7   Is Attack IP     200 non-null   bool    
 8   Blacklisted       200 non-null   bool    
 9   Browser Type      200 non-null   category
 10  Is Account Takeover 200 non-null   bool    
dtypes: bool(4), category(5), int64(2)
memory usage: 21.4 KB
Feature importances:
[0.118387  0.0012349  0.00642419 0.2238188  0.00738945 0.00833171
 0.         0.634414   0.         0.         ]
```





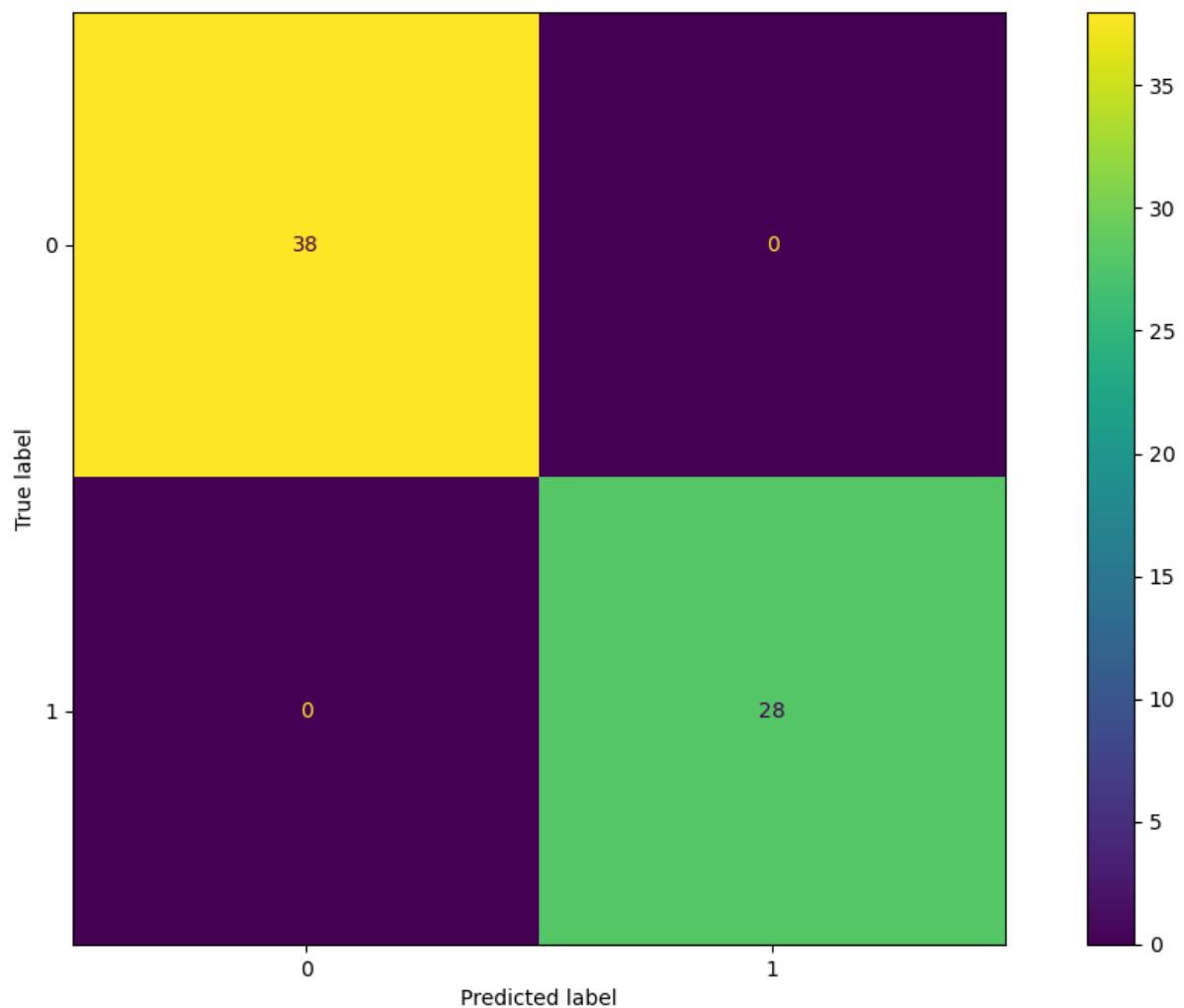
## 2. Label encoding (cat.codes)

### 2A. XGBoost

memory usage: 5.4 KB

	precision	recall	f1-score	support
False	1.00	1.00	1.00	38
True	1.00	1.00	1.00	28
accuracy			1.00	66
macro avg	1.00	1.00	1.00	66
weighted avg	1.00	1.00	1.00	66

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 160
The number of records in the test dataset is 40
The training dataset has 98 records for the majority class and 62 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      25
      True         0.38     1.00     0.55      15

      accuracy          0.38      40
      macro avg       0.19     0.50     0.27      40
      weighted avg    0.14     0.38     0.20      40

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.62     1.00     0.77      25
      True         0.00     0.00     0.00      15

      accuracy          0.62      40
      macro avg       0.31     0.50     0.38      40
      weighted avg    0.39     0.62     0.48      40

[[ 0 25]
 [ 0 15]]

```

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

```

The number of records in the training dataset is 140
The number of records in the test dataset is 60
The training dataset has 86 records for the majority class and 54 records for the minority class.
      precision    recall   f1-score   support
      False        0.00     0.00     0.00      37
      True         0.38     1.00     0.55      23

      accuracy          0.38      60
      macro avg       0.19     0.50     0.28      60
      weighted avg    0.15     0.38     0.21      60

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.62     1.00     0.76      37
      True         0.00     0.00     0.00      23

      accuracy          0.62      60
      macro avg       0.31     0.50     0.38      60
      weighted avg    0.38     0.62     0.47      60

[[ 0 37]
 [ 0 23]]

```

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

```

The number of records in the training dataset is 100
The number of records in the test dataset is 100
The training dataset has 62 records for the majority class and 38 records for the minority class.
      precision    recall   f1-score   support
False        0.00     0.00     0.00      61
True        0.39     1.00     0.56      39

accuracy          0.39      100
macro avg       0.20     0.50     0.28      100
weighted avg    0.15     0.39     0.22      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.61     1.00     0.76      61
True         0.00     0.00     0.00      39

accuracy          0.61      100
macro avg       0.30     0.50     0.38      100
weighted avg    0.37     0.61     0.46      100

[[ 0 61]
 [ 0 39]]

```

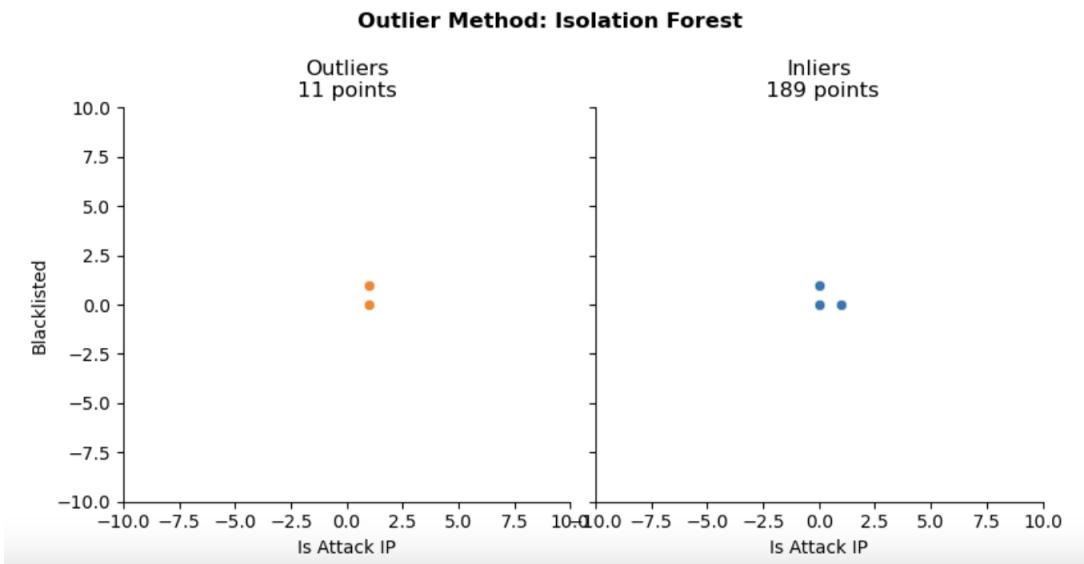
## 2C. Isolation Forest

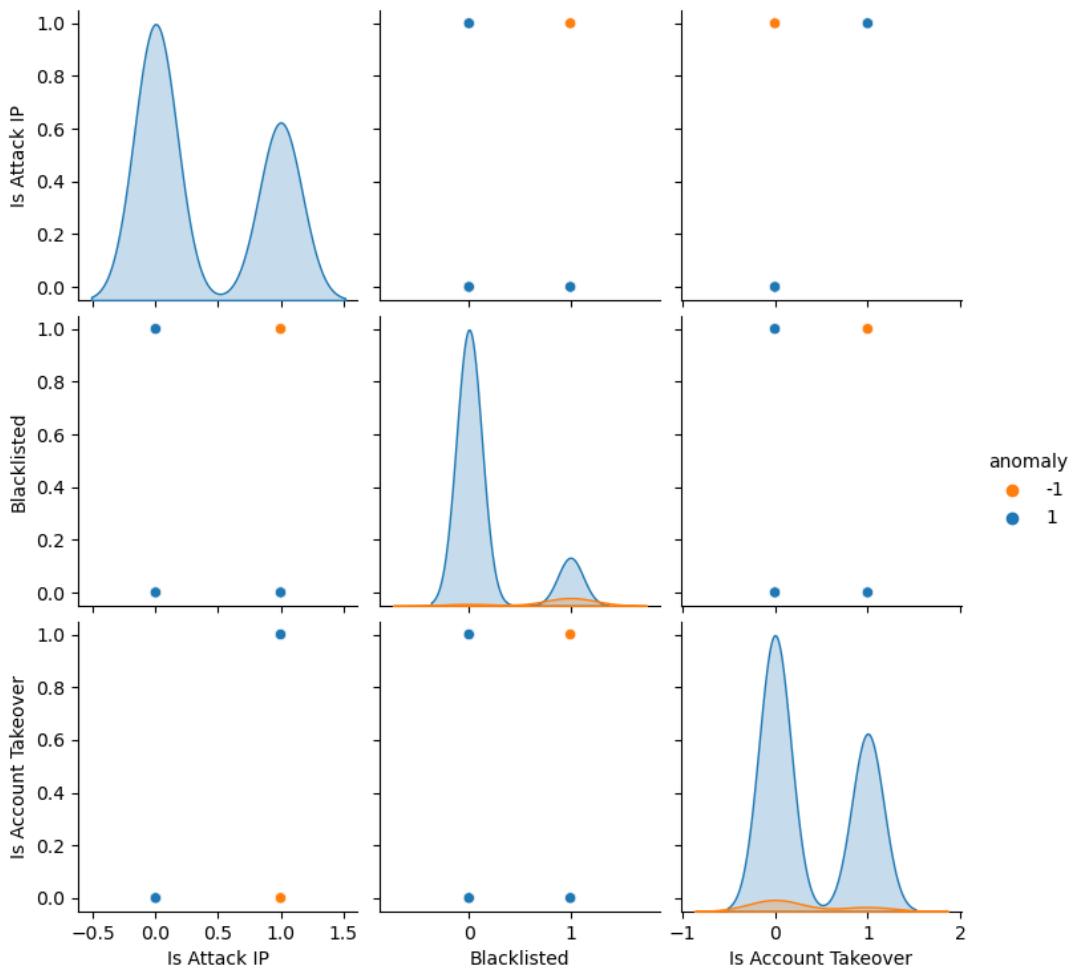
- anomaly\_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 11
Number of non anomalous values 189
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x449f56090>

```



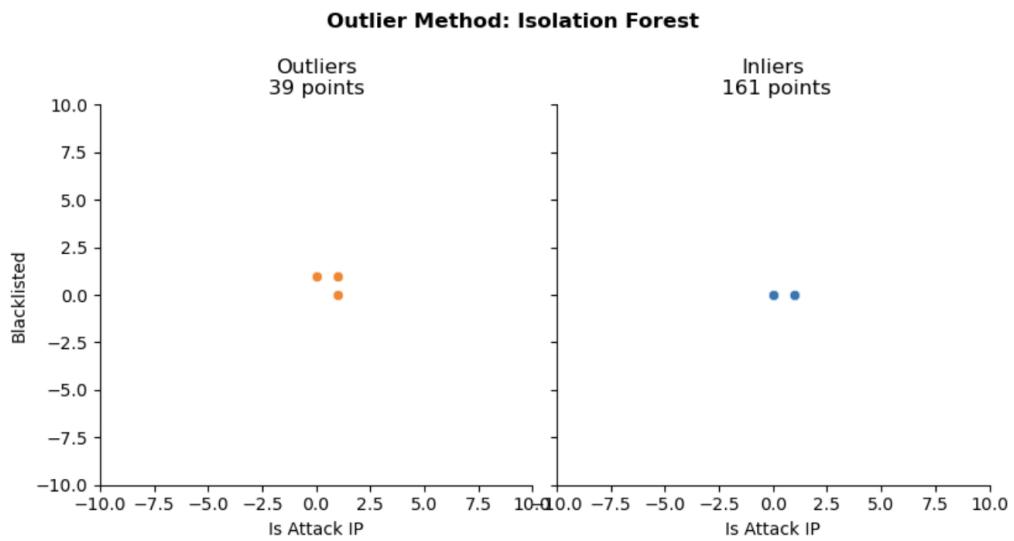


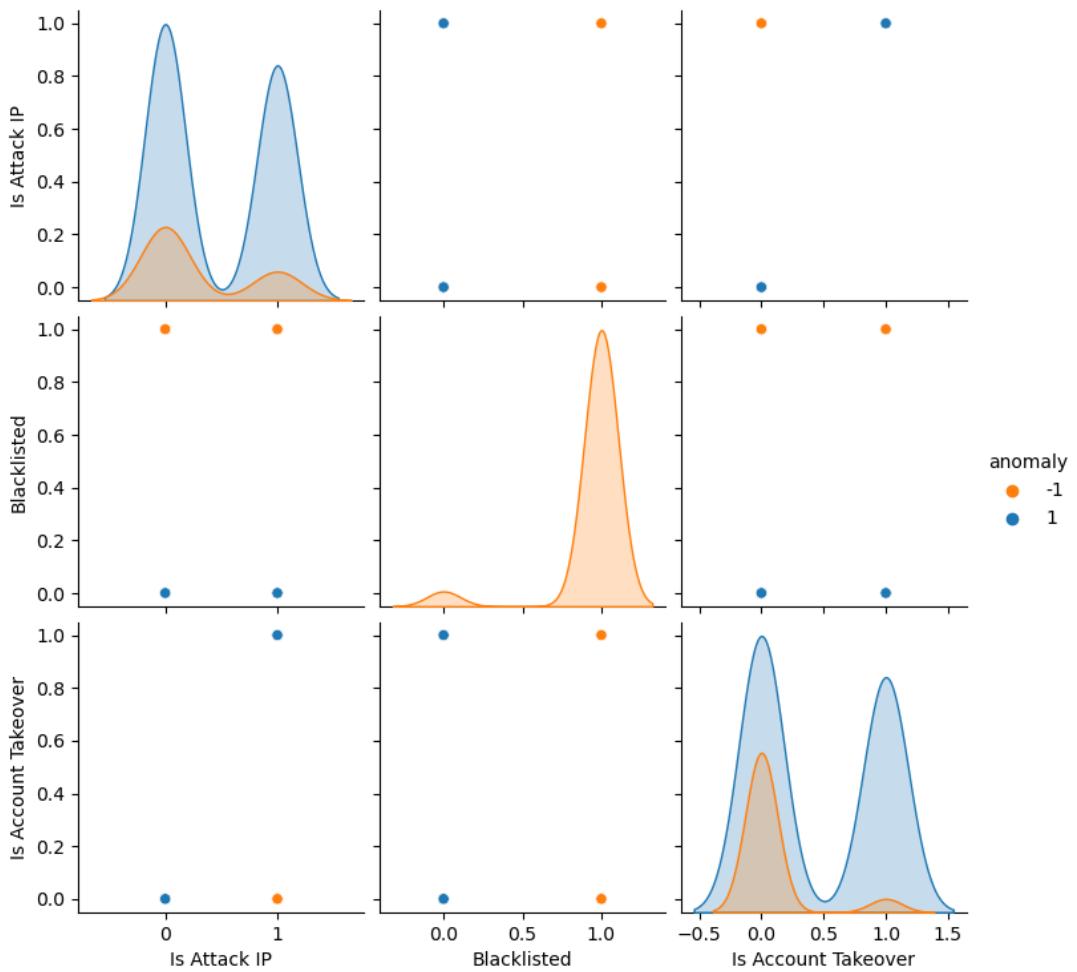
- increasing contamination value to 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 39
Number of non anomalous values 161
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x449876a90>

```



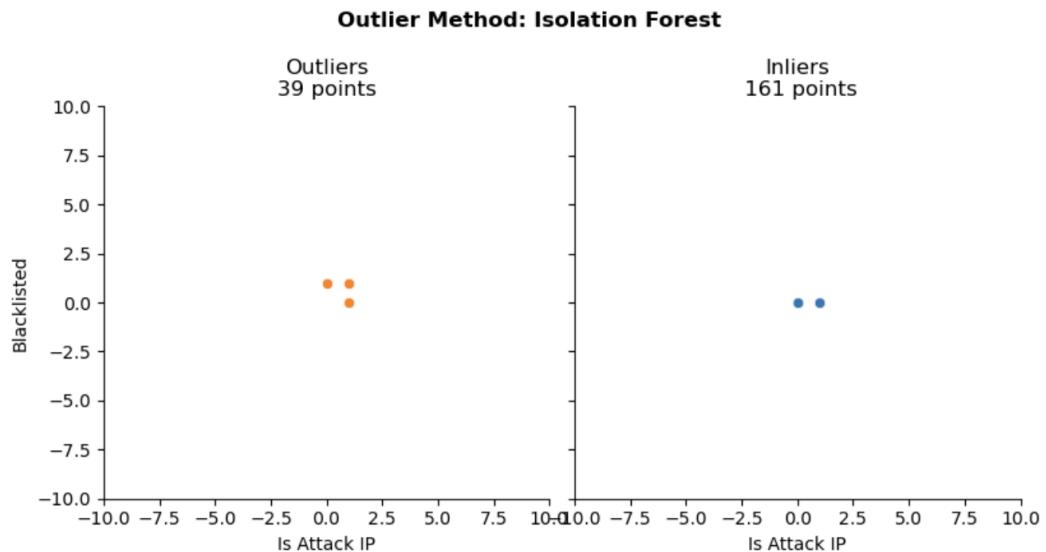


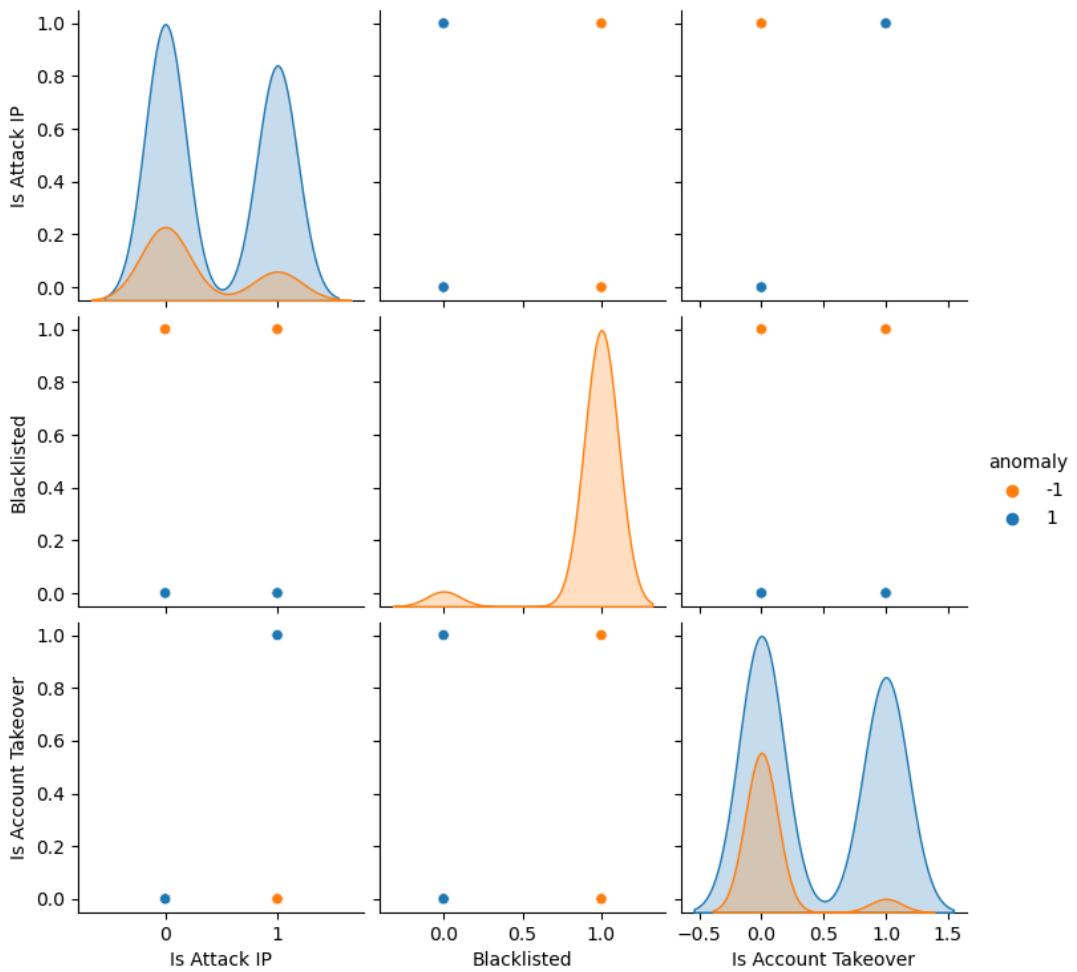
- contamination value == 0.5

```

Outlier Method: Isolation Forest
Number of anomalous values 39
Number of non anomalous values 161
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x30fc40cd0>

```



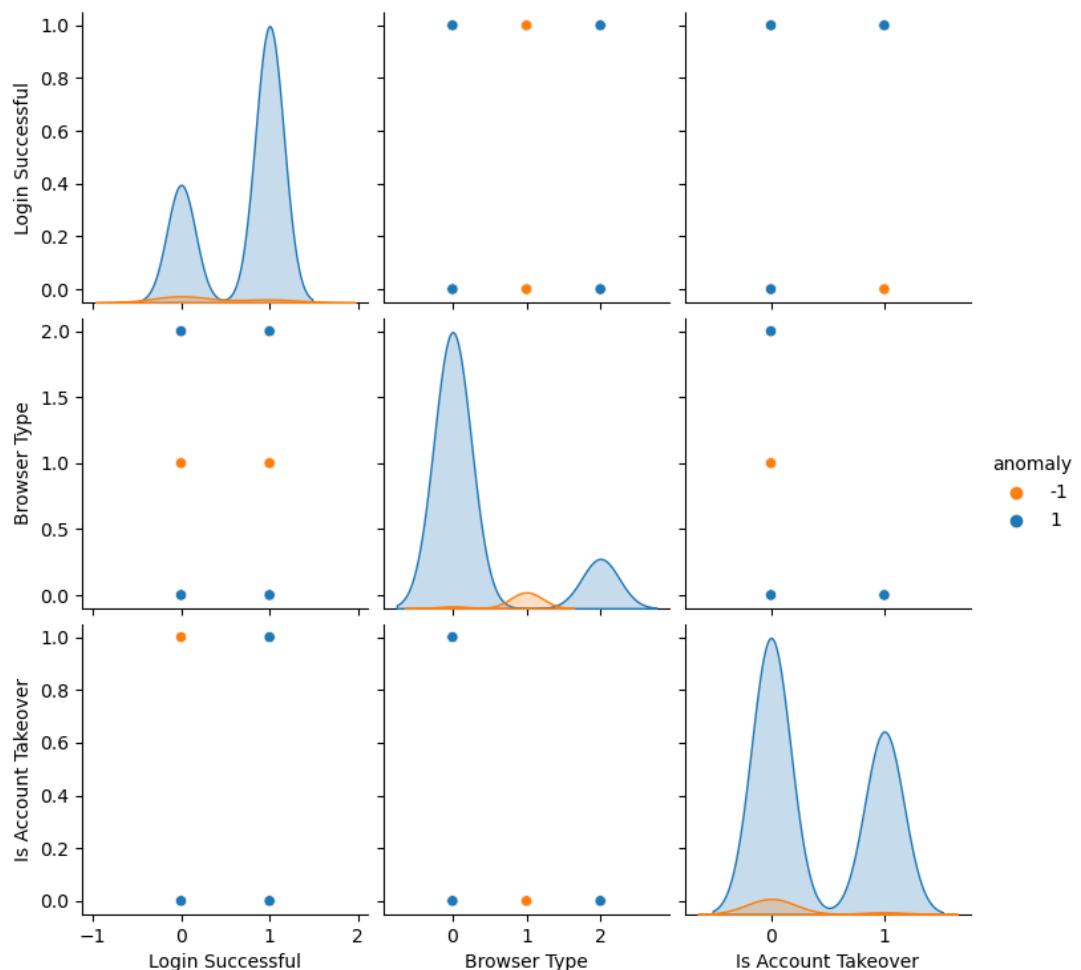
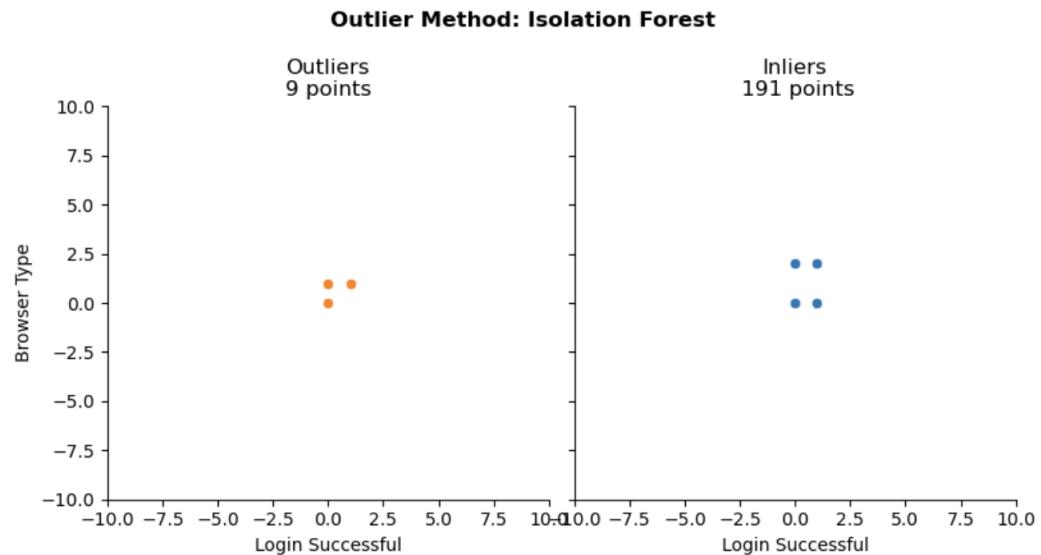


- anomaly\_inputs = ['Login Successful', 'Browser Type', 'Is Account Takeover']
  - contamination value == 0.1

```

Outlier Method: Isolation Forest
Number of anomalous values 9
Number of non anomalous values 191
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x453ecf6d0>

```

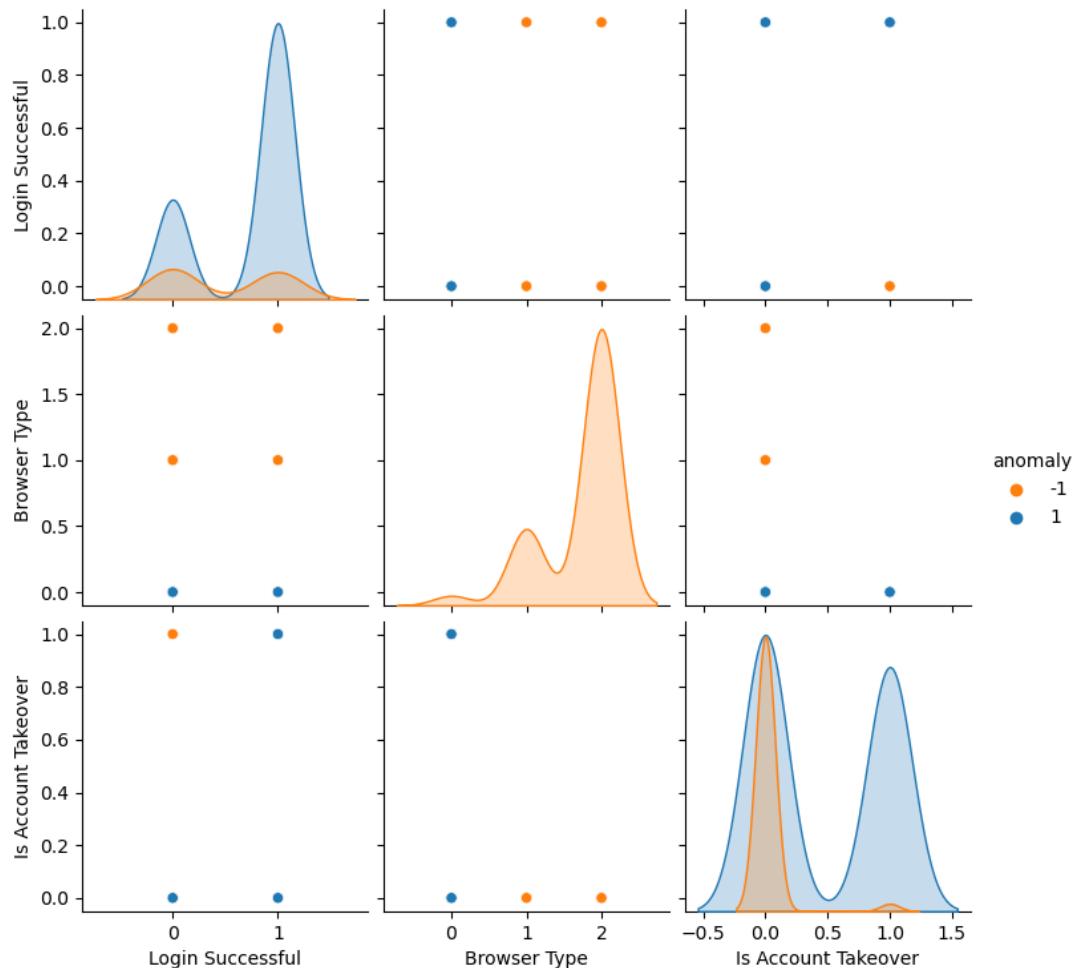
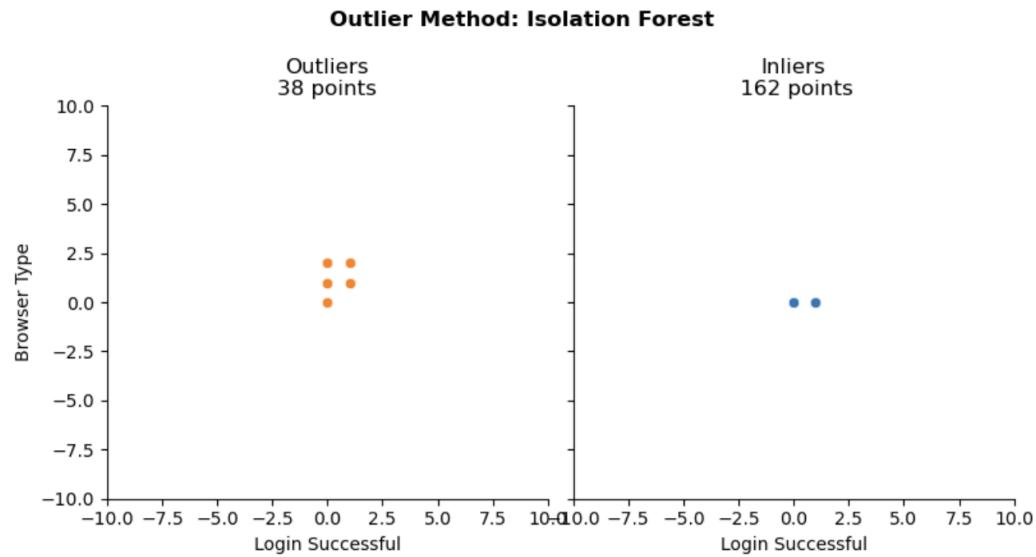


- contamination value == 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 38
Number of non anomalous values 162
Total Number of Values: 200
: <seaborn.axisgrid.PairGrid at 0x3b1f11710>

```



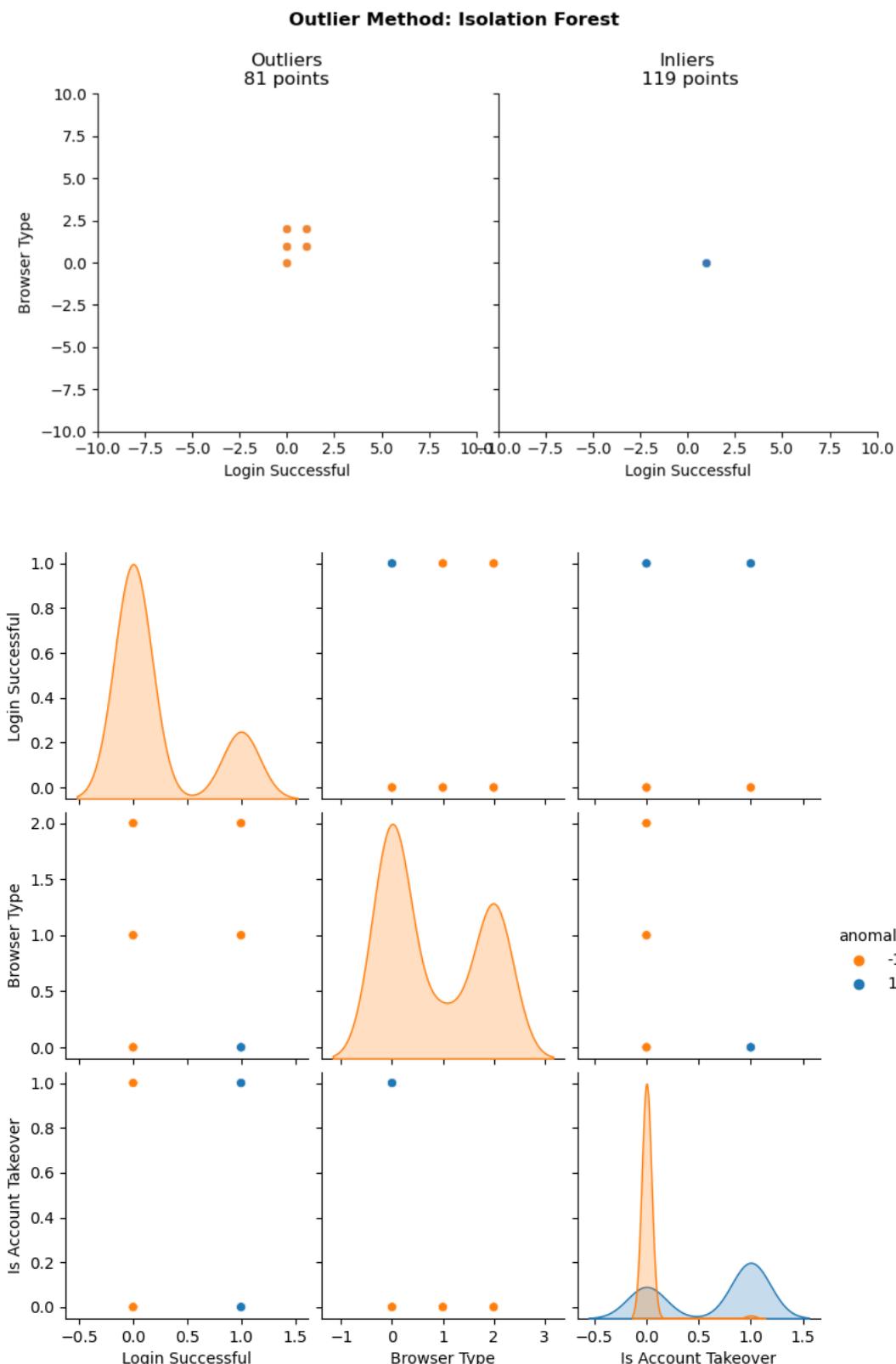
- contamination value == 0.5

```

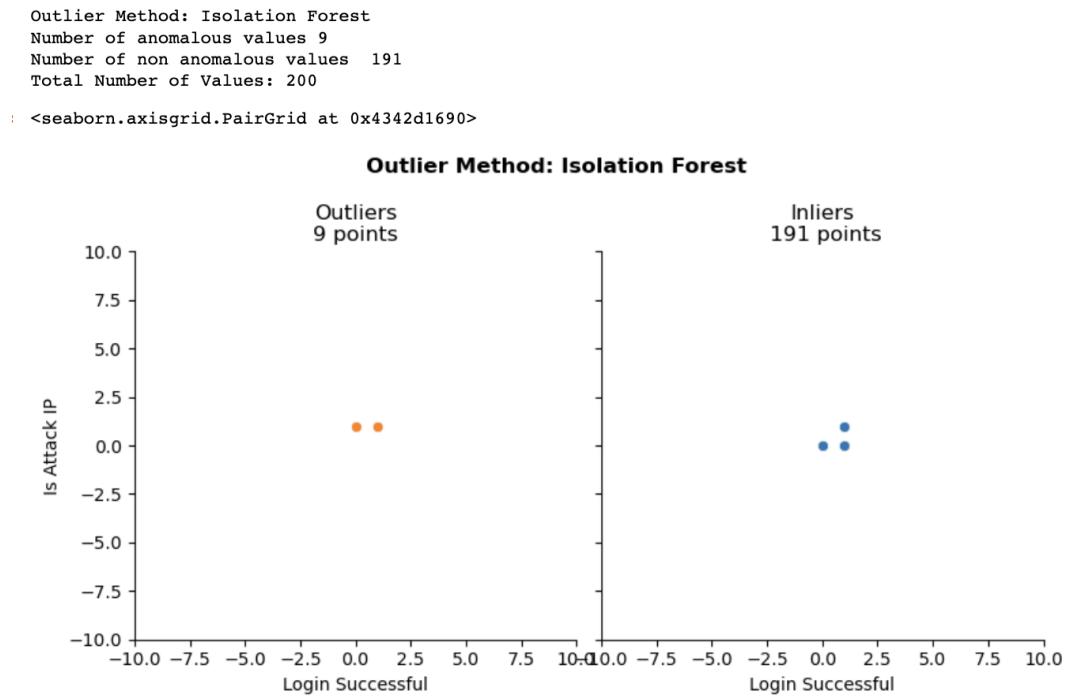
Outlier Method: Isolation Forest
Number of anomalous values 81
Number of non anomalous values 119
Total Number of Values: 200

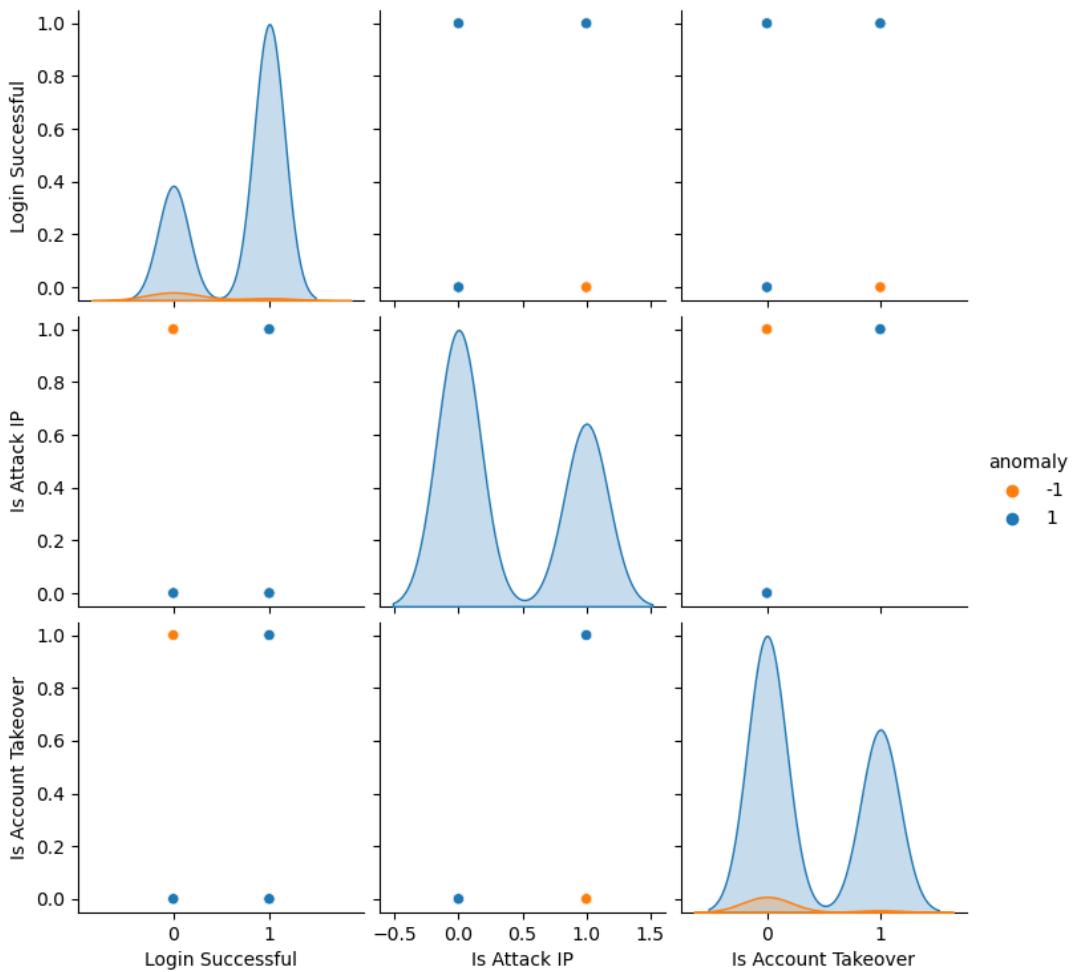
```

<seaborn.axisgrid.PairGrid at 0x2e6673790>



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



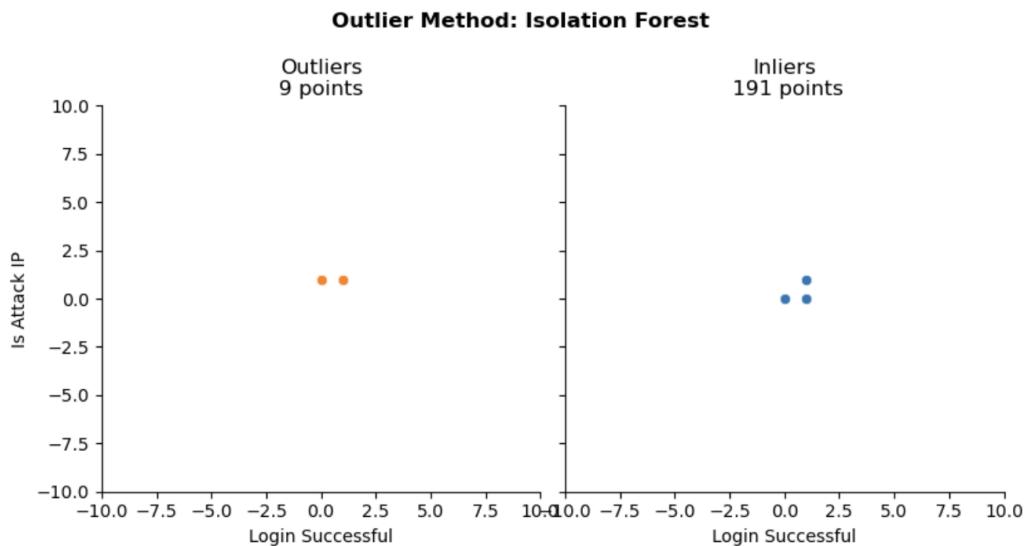


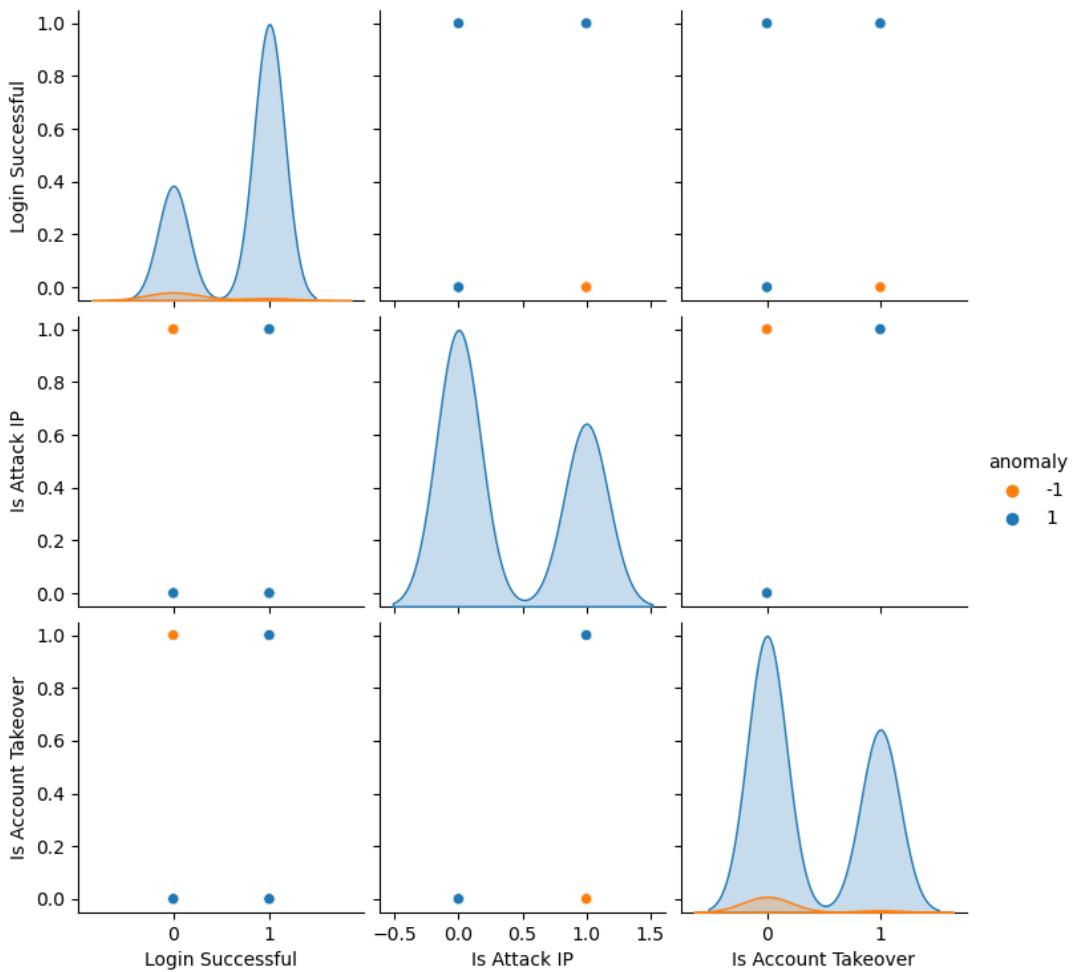
- contamination value == 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 9
Number of non anomalous values 191
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x432bae410>

```



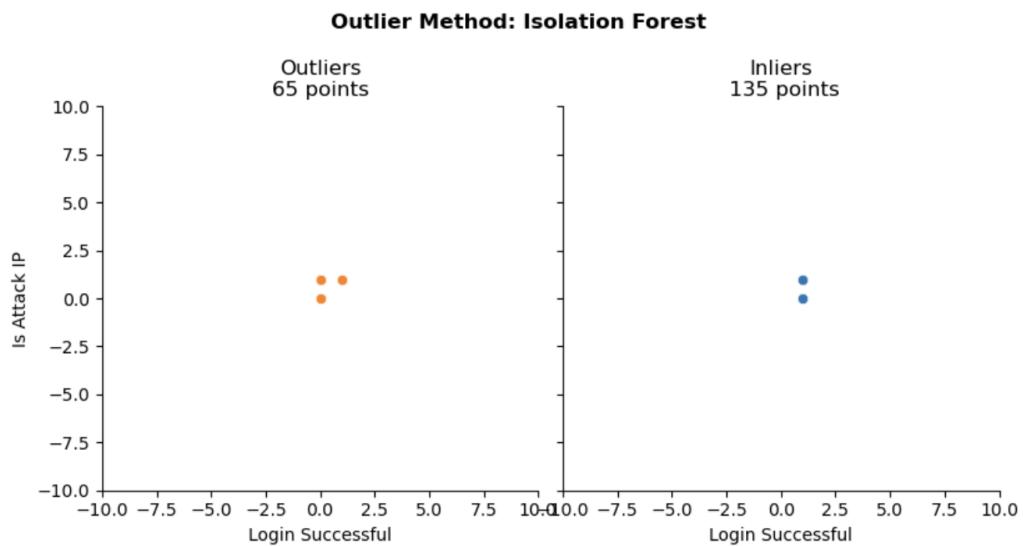


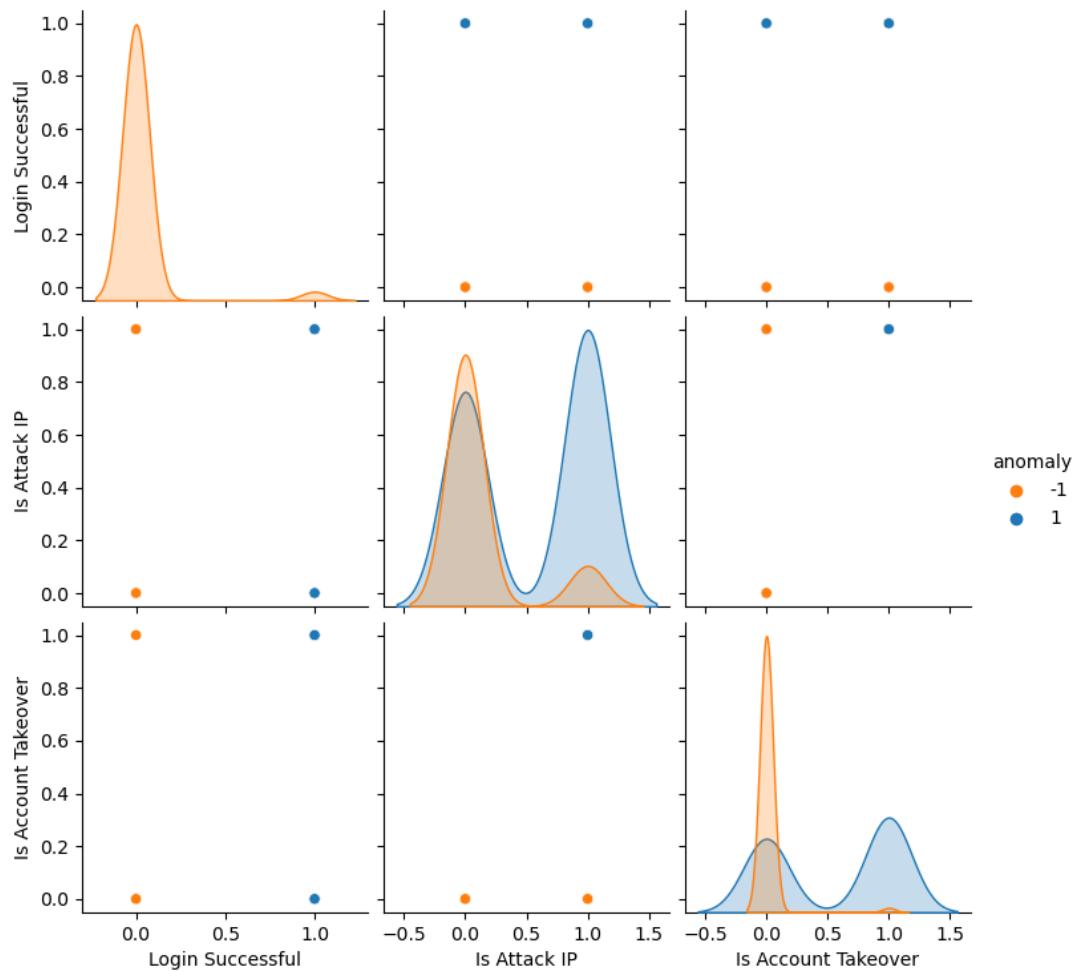
- contamination value == 0.5

```

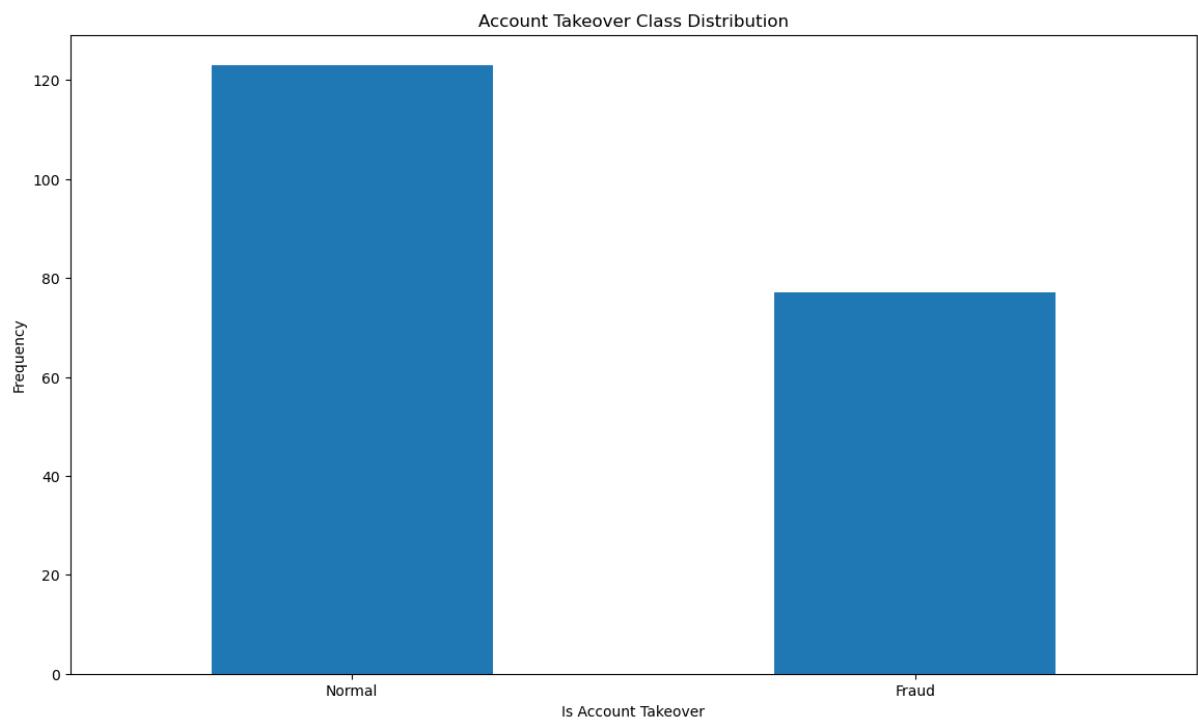
Outlier Method: Isolation Forest
Number of anomalous values 65
Number of non anomalous values 135
Total Number of Values: 200
: <seaborn.axisgrid.PairGrid at 0x4318d8a90>

```





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



"{}: {}".format(clf\_name,n\_errors) → np. Isolation Forest: 91

Isolation Forest: 91

Accuracy Score :

0.545

Classification Report :

	precision	recall	f1-score	support
False	0.59	0.86	0.70	123
True	0.15	0.04	0.06	77
accuracy			0.55	200
macro avg	0.37	0.45	0.38	200
weighted avg	0.42	0.55	0.45	200

Local Outlier Factor: 81

Accuracy Score :

0.595

Classification Report :

	precision	recall	f1-score	support
False	0.62	0.90	0.73	123
True	0.40	0.10	0.16	77
accuracy			0.59	200
macro avg	0.51	0.50	0.45	200
weighted avg	0.53	0.59	0.51	200

Support Vector Machine: 123

Accuracy Score :

0.385

Classification Report :

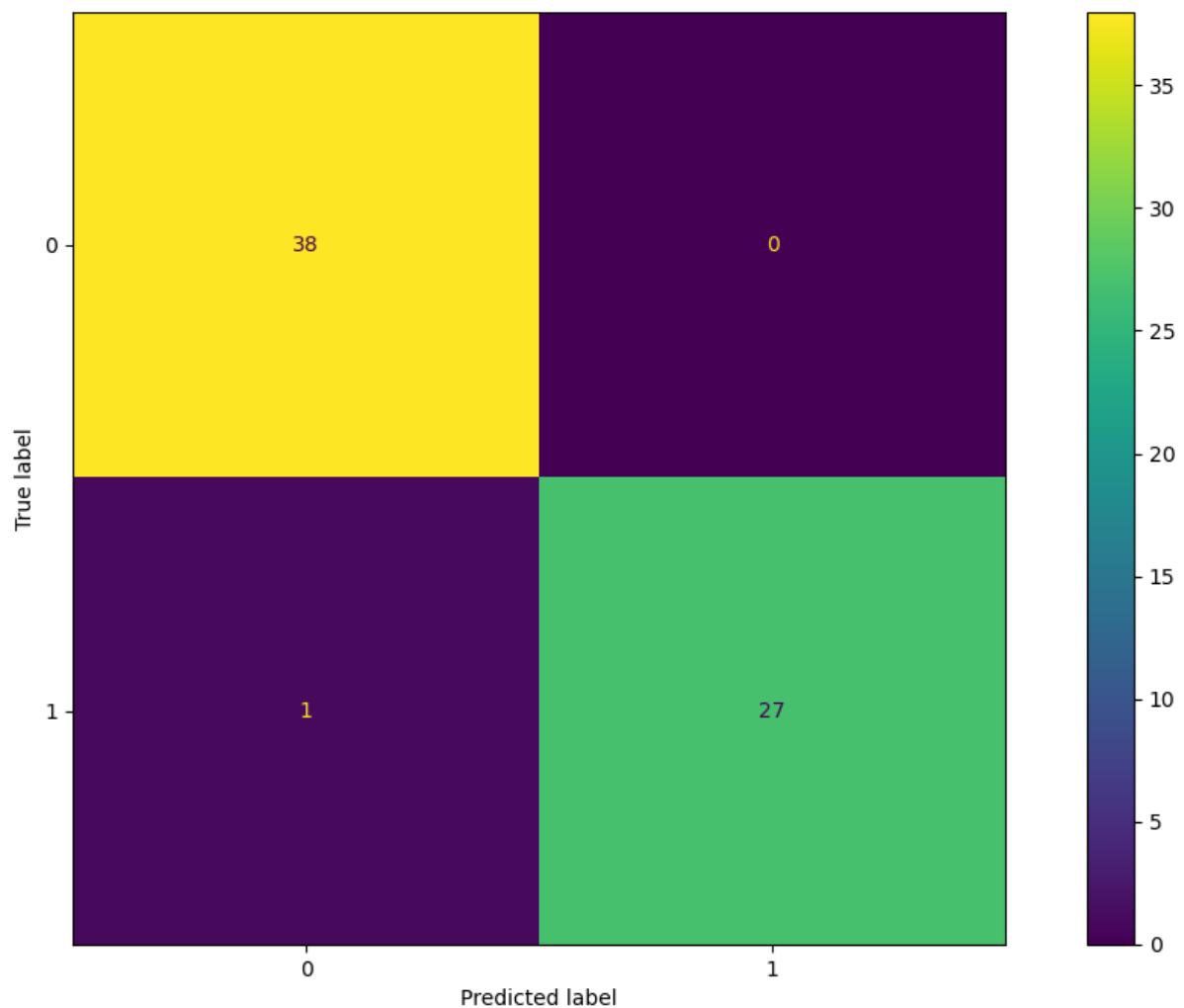
	precision	recall	f1-score	support
False	0.50	0.04	0.08	123
True	0.38	0.94	0.54	77
accuracy			0.38	200
macro avg	0.44	0.49	0.31	200
weighted avg	0.45	0.39	0.25	200

**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: 80.6 KB
```

```
precision    recall   f1-score   support
Accuracy: 98.48%
          precision    recall   f1-score   support
          False        0.97      1.00      0.99      38
          True         1.00      0.96      0.98      28
          accuracy           0.98
          macro avg       0.99      0.98      0.98      66
          weighted avg     0.99      0.98      0.98      66
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 160
The number of records in the test dataset is 40
The training dataset has 98 records for the majority class and 62 records for the minority class.
      precision    recall   f1-score   support
      False        0.75     0.12     0.21      25
      True         0.39     0.93     0.55      15

      accuracy          0.42      40
      macro avg       0.57     0.53     0.38      40
      weighted avg    0.61     0.42     0.34      40

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.62     1.00     0.77      25
      True         0.00     0.00     0.00      15

      accuracy          0.62      40
      macro avg       0.31     0.50     0.38      40
      weighted avg    0.39     0.62     0.48      40

[[ 3 22]
 [ 1 14]]

```

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

```

The number of records in the training dataset is 140
The number of records in the test dataset is 60
The training dataset has 86 records for the majority class and 54 records for the minority class.
      precision    recall   f1-score   support
      False        1.00     0.03     0.05      37
      True         0.39     1.00     0.56      23

      accuracy          0.40      60
      macro avg       0.69     0.51     0.31      60
      weighted avg    0.77     0.40     0.25      60

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
      False        0.62     1.00     0.76      37
      True         0.00     0.00     0.00      23

      accuracy          0.62      60
      macro avg       0.31     0.50     0.38      60
      weighted avg    0.38     0.62     0.47      60

[[ 1 36]
 [ 0 23]]

```

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

```

The number of records in the training dataset is 100
The number of records in the test dataset is 100
The training dataset has 62 records for the majority class and 38 records for the minority class.
      precision    recall   f1-score   support
False        0.00     0.00     0.00      61
True        0.39     1.00     0.56      39

accuracy                           0.39      100
macro avg                         0.20      100
weighted avg                      0.15      100

The customized score threshold for 2% of outliers is 0.00
      precision    recall   f1-score   support
False        0.61     1.00     0.76      61
True        0.00     0.00     0.00      39

accuracy                           0.61      100
macro avg                         0.30      100
weighted avg                      0.37      100

[[ 0 61]
 [ 0 39]]

```

### 3C. Isolation Forest

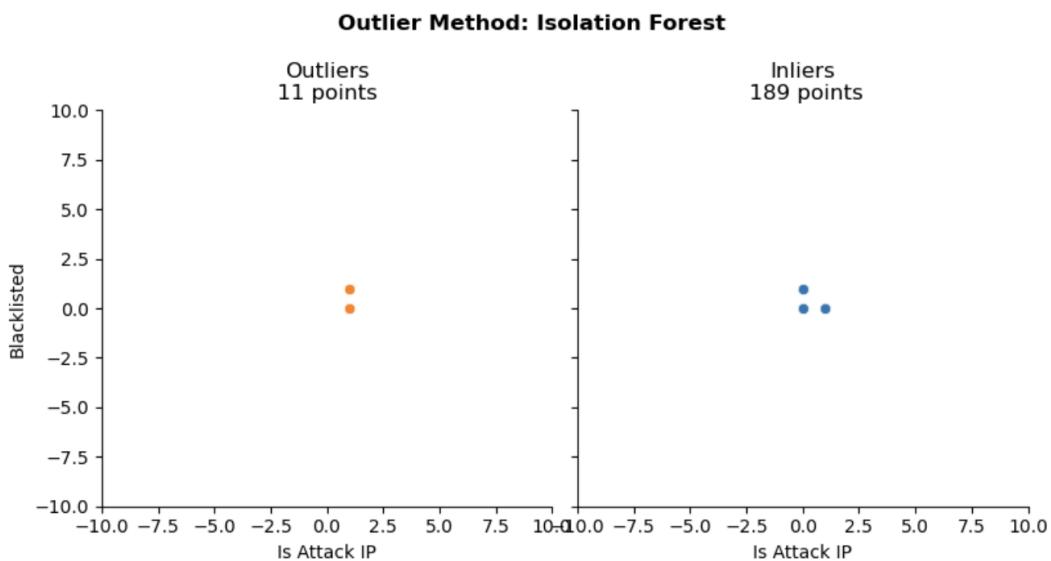
- anomaly\_inputs = ['Is Attack IP', 'Blacklisted', 'Is Account Takeover']
  - contamination value == 0.1

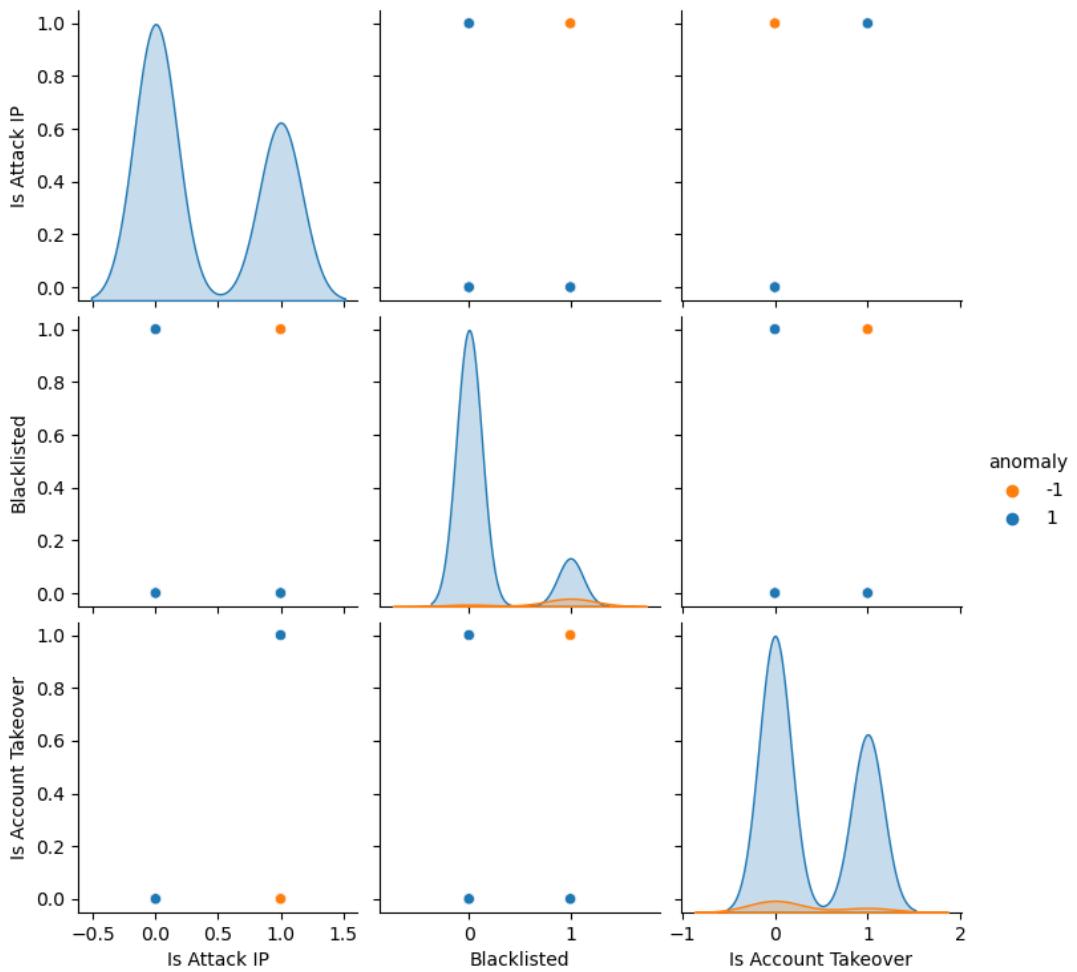
```

Outlier Method: Isolation Forest
Number of anomalous values 11
Number of non anomalous values 189
Total Number of Values: 200

<seaborn.axisgrid.PairGrid at 0x427a1cad0>

```

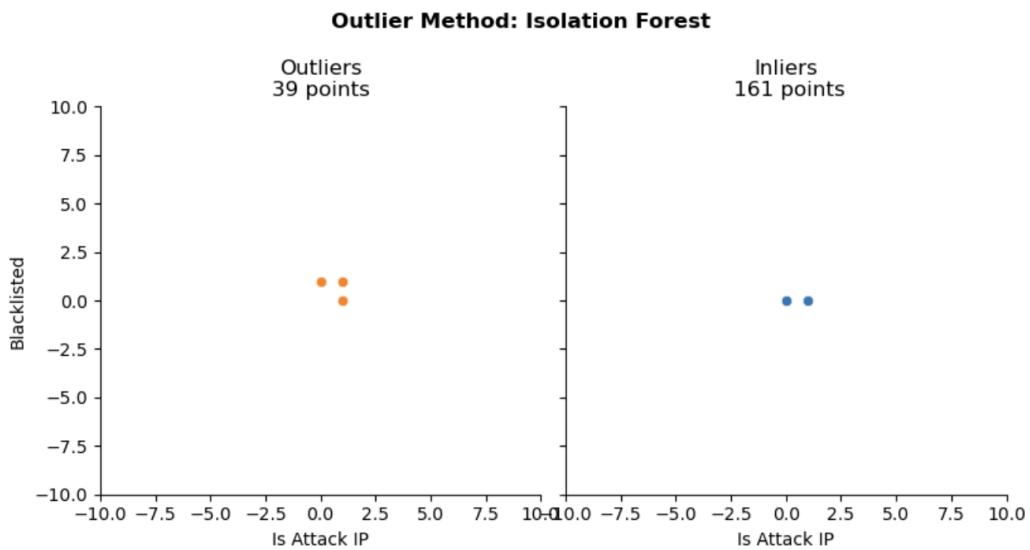


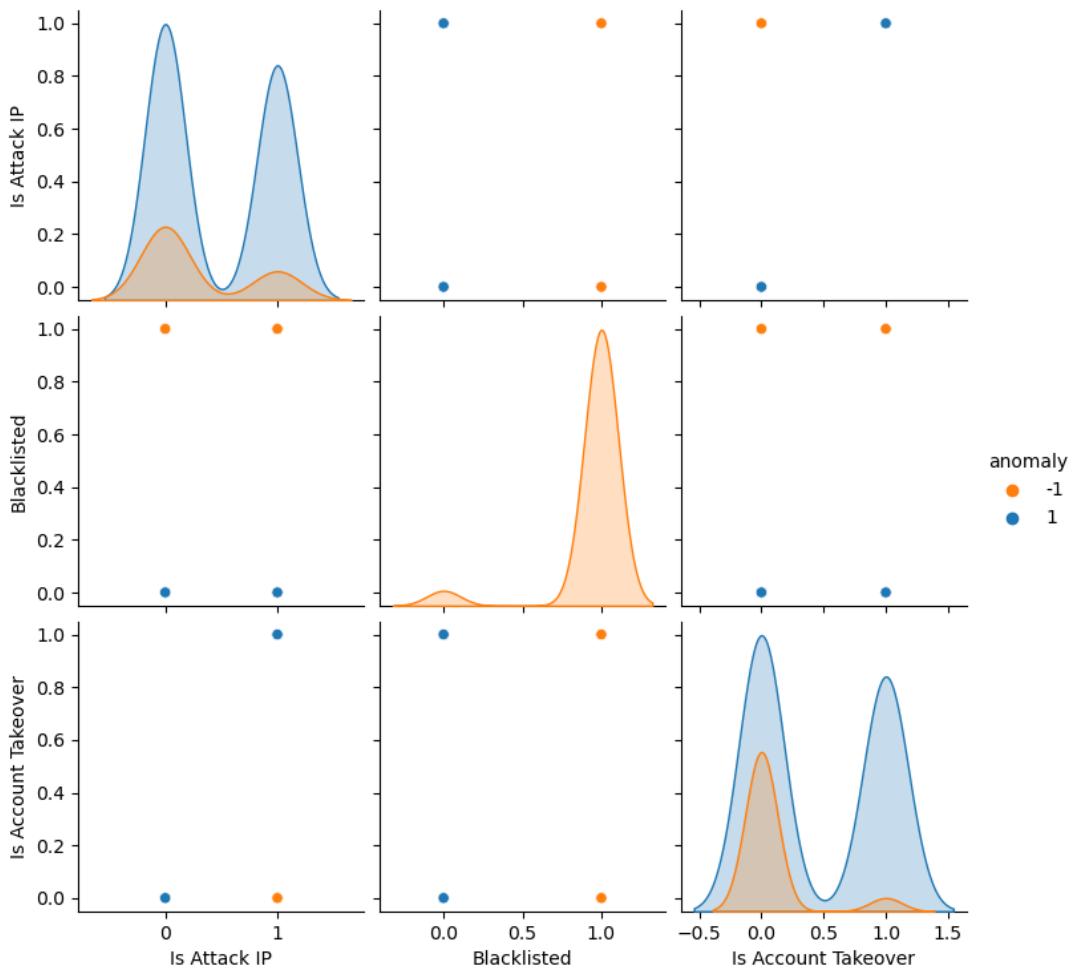


- contamination value == 0.3

```
Outlier Method: Isolation Forest
Number of anomalous values 39
Number of non anomalous values 161
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x42669bb90>
```

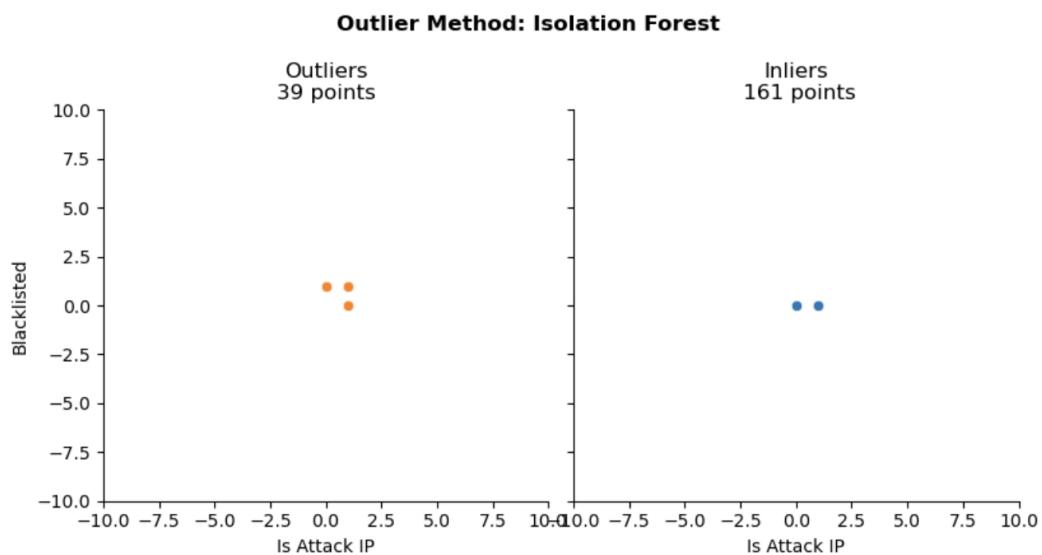


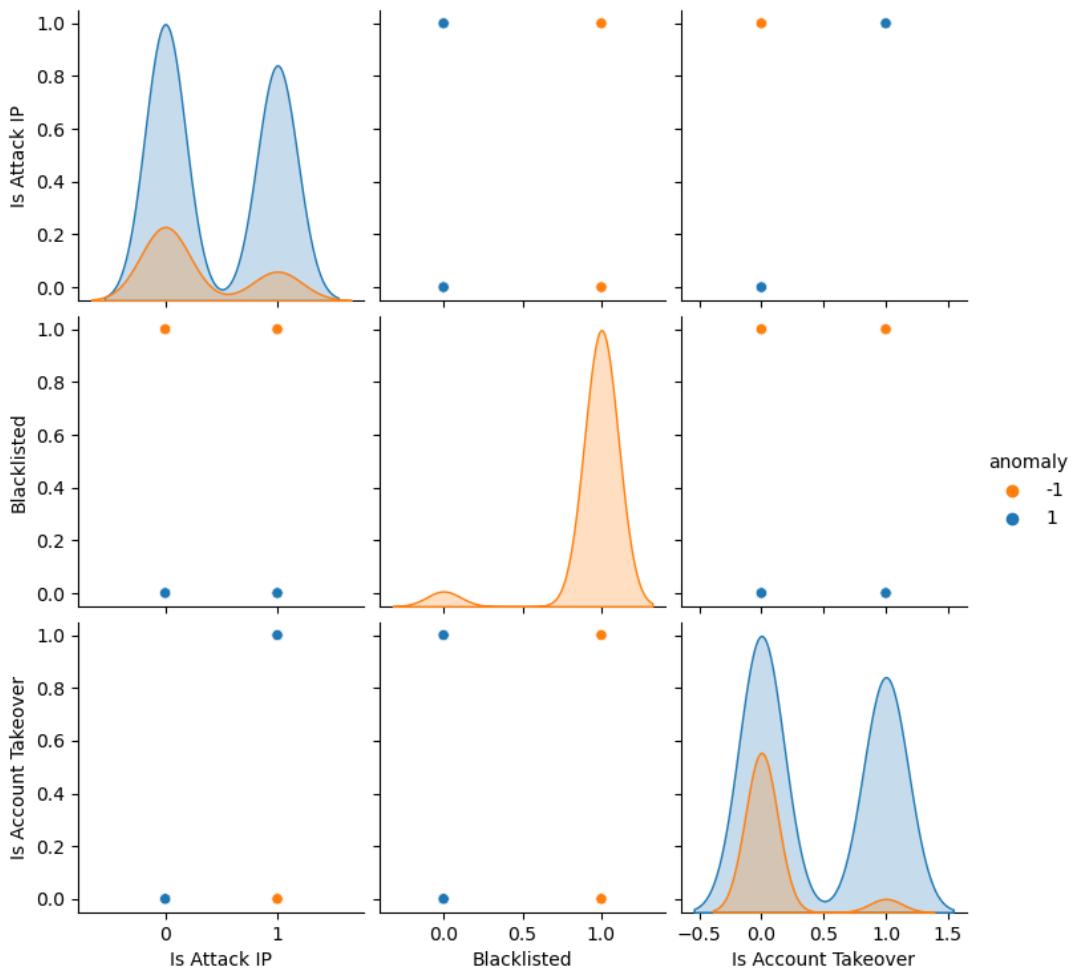


- contamination value == 0.5

```
Outlier Method: Isolation Forest
Number of anomalous values 39
Number of non anomalous values 161
Total Number of Values: 200
```

```
<seaborn.axisgrid.PairGrid at 0x4271f2f50>
```

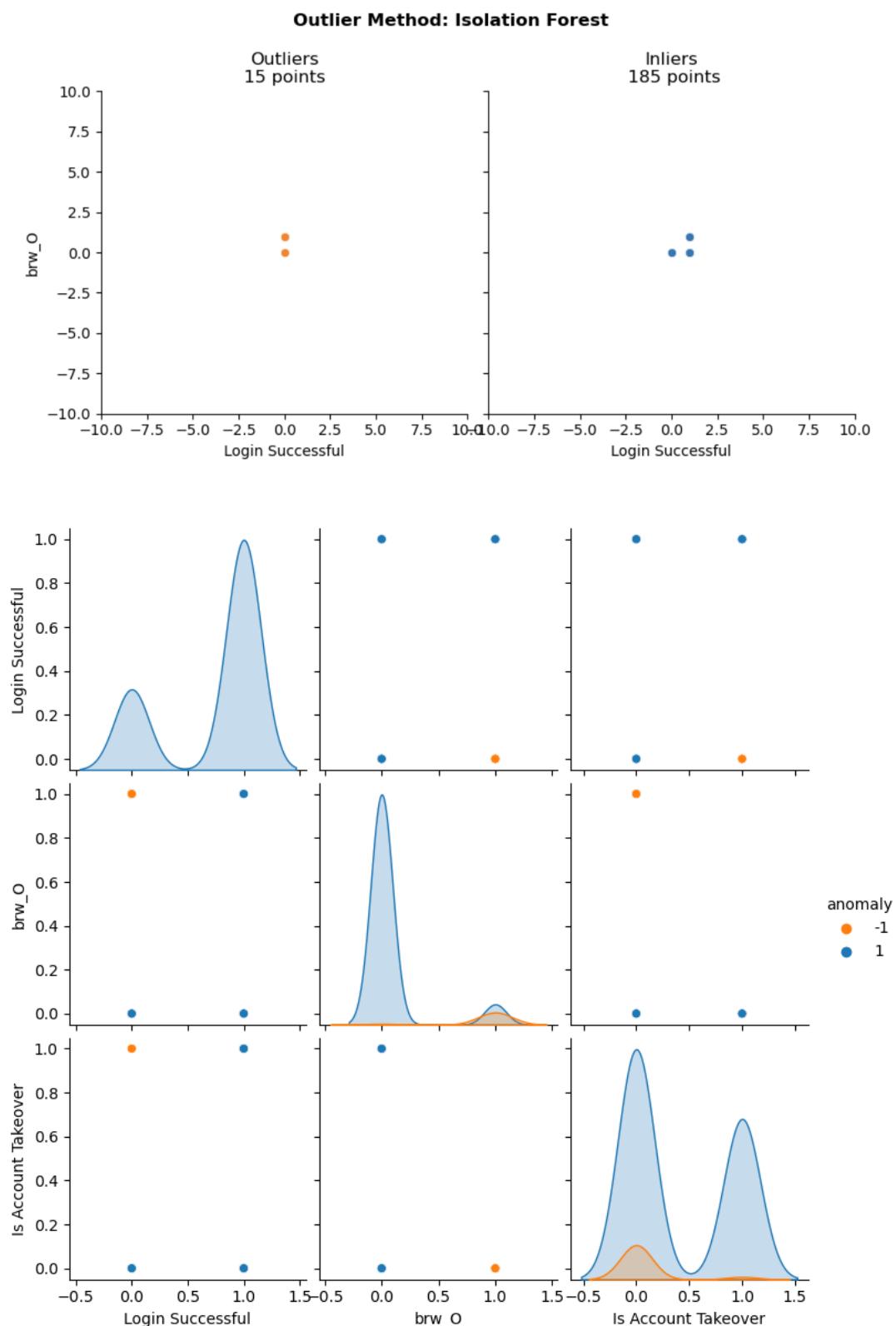




- anomaly\_inputs = ['Login Successful', 'Browser Type (brw\_O)', 'Is Account Takeover']
  - contamination value == 0.1

Outlier Method: Isolation Forest  
 Number of anomalous values 15  
 Number of non anomalous values 185  
 Total Number of Values: 200

<seaborn.axisgrid.PairGrid at 0x423ec74d0>



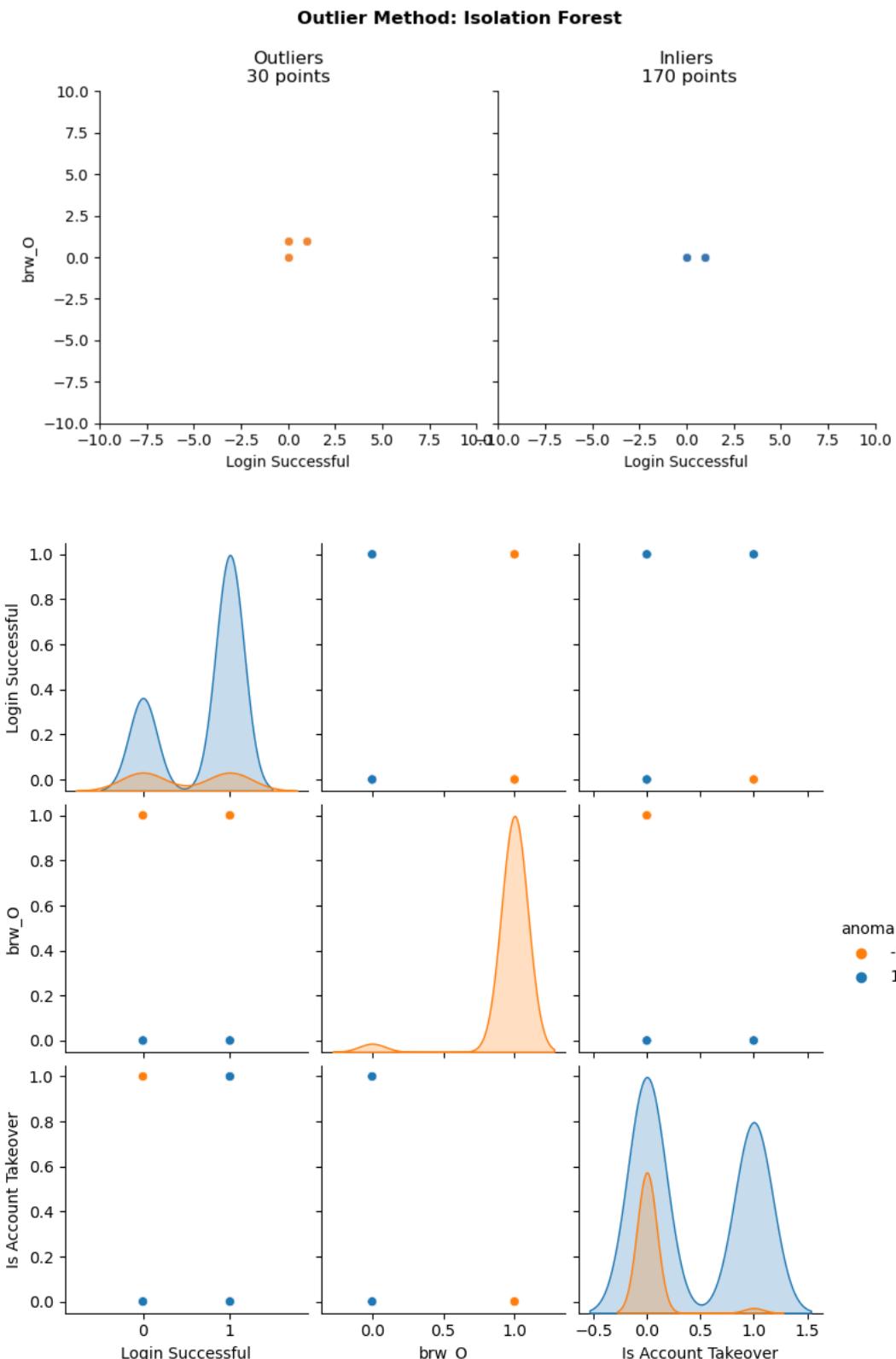
- contamination value == 0.3

```

Outlier Method: Isolation Forest
Number of anomalous values 30
Number of non anomalous values 170
Total Number of Values: 200

```

```
<seaborn.axisgrid.PairGrid at 0x4235f9bd0>
```

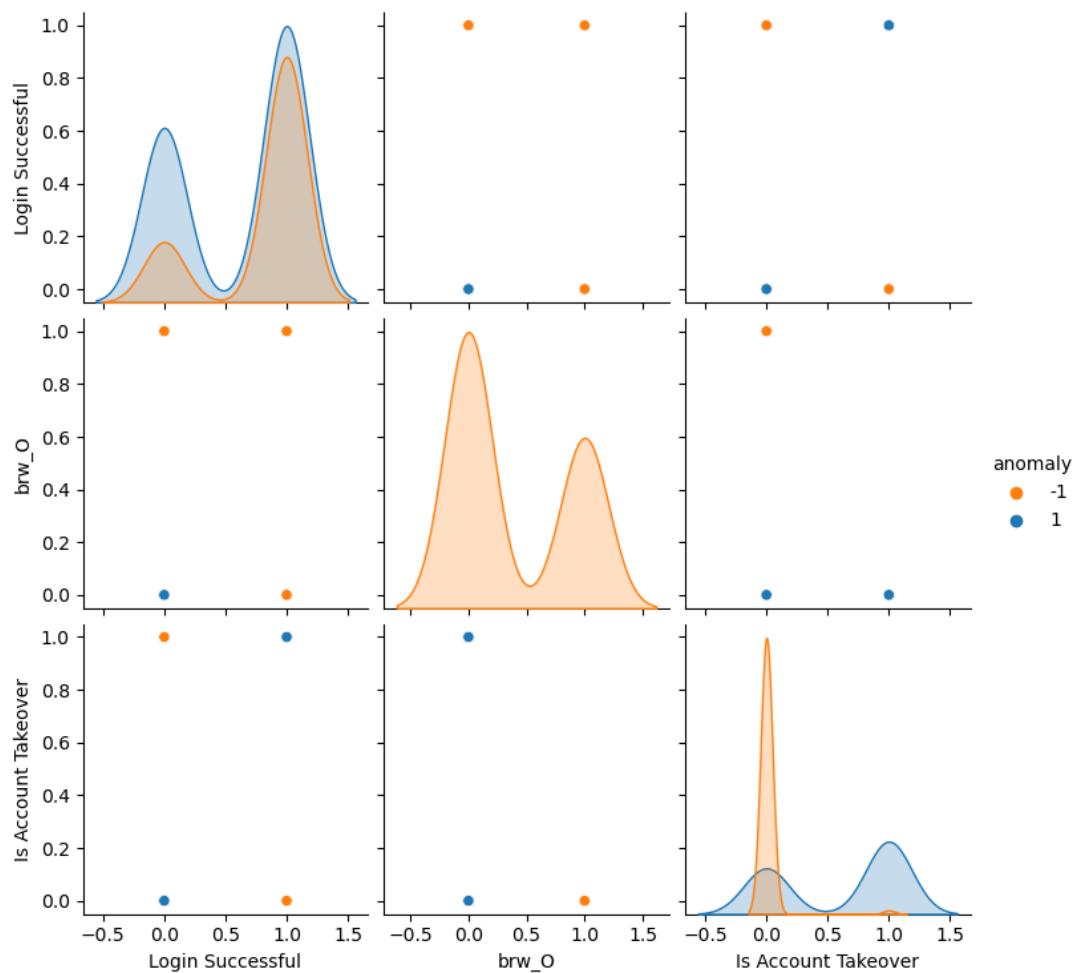
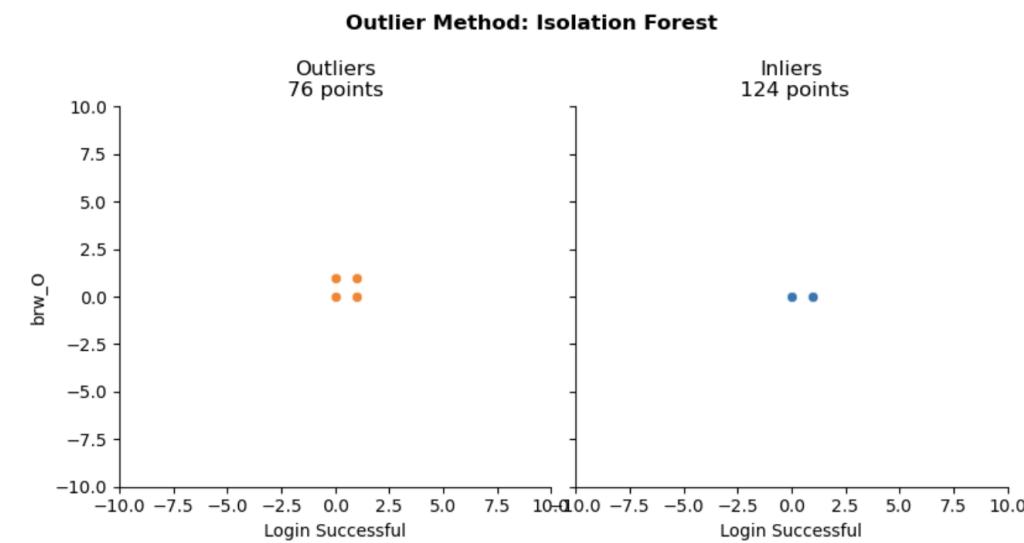


- contamination value == 0.5

```

Outlier Method: Isolation Forest
Number of anomalous values 76
Number of non anomalous values 124
Total Number of Values: 200
<seaborn.axisgrid.PairGrid at 0x421ff5210>

```



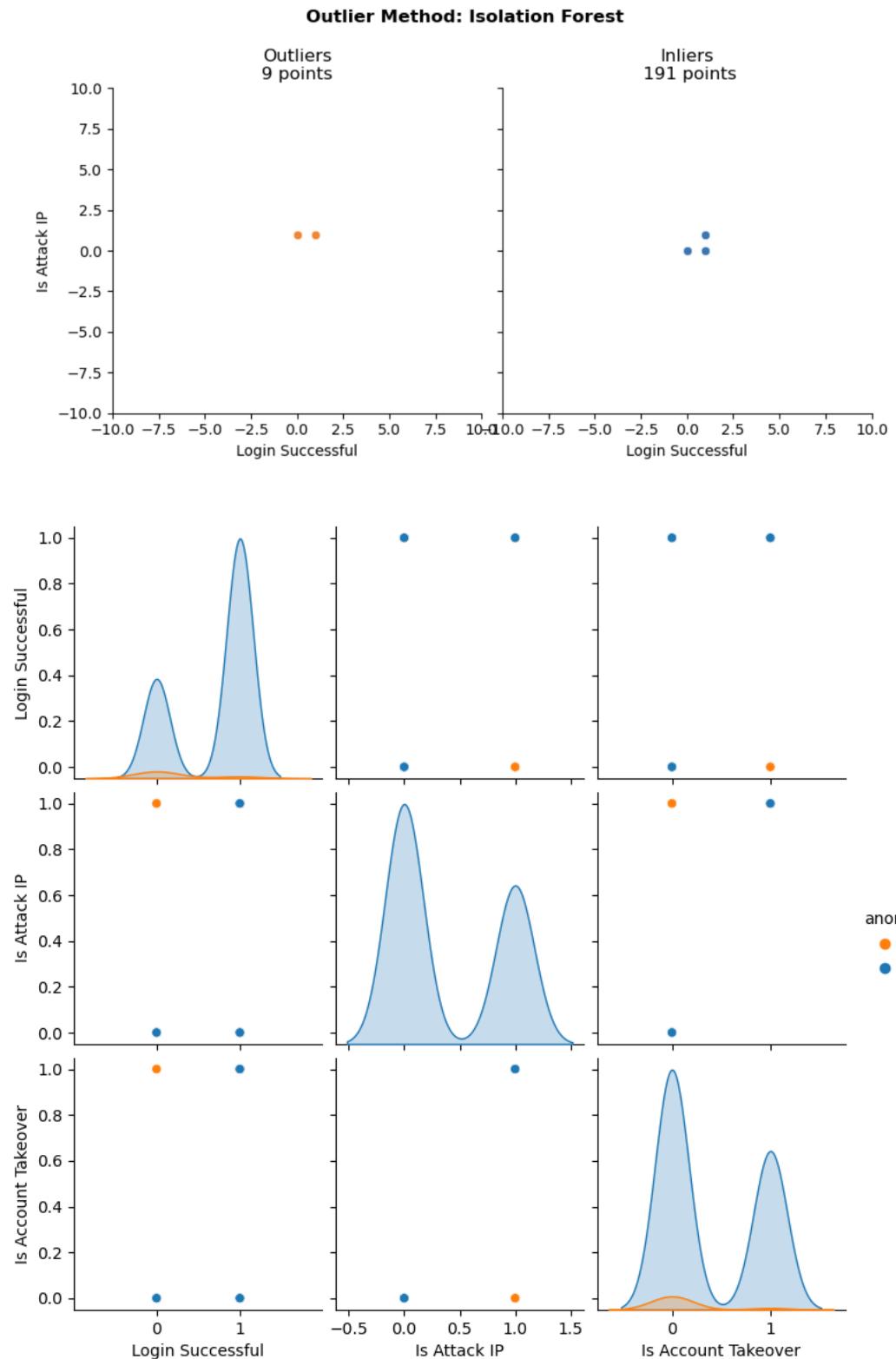
- anomaly\_inputs = ['Login Successful', 'Is Attack IP', 'Is Account Takeover']

- contamination value == 0.1

```

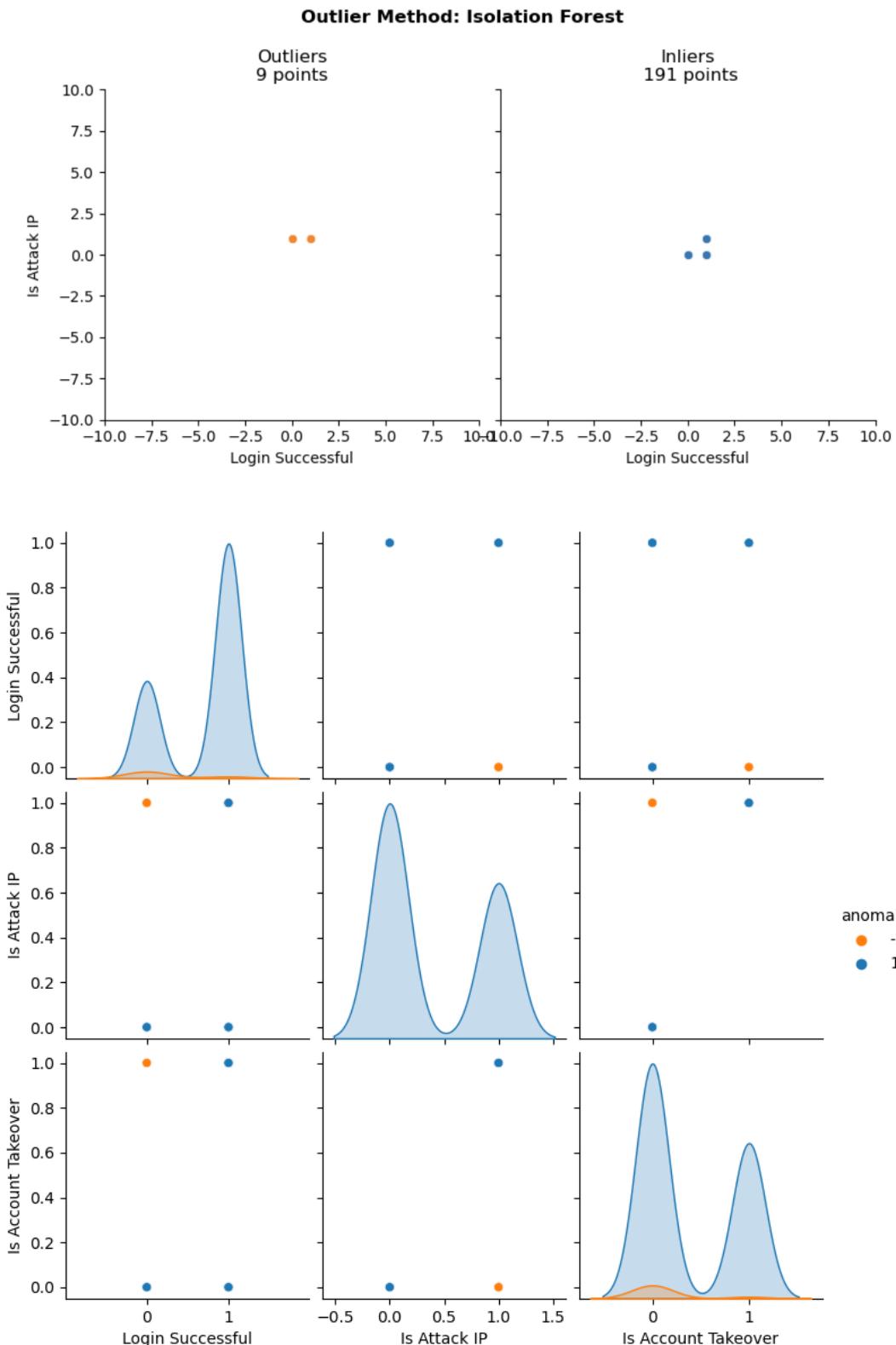
Outlier Method: Isolation Forest
Number of anomalous values 9
Number of non anomalous values 191
Total Number of Values: 200
: <seaborn.axisgrid.PairGrid at 0x420c0b5d0>

```

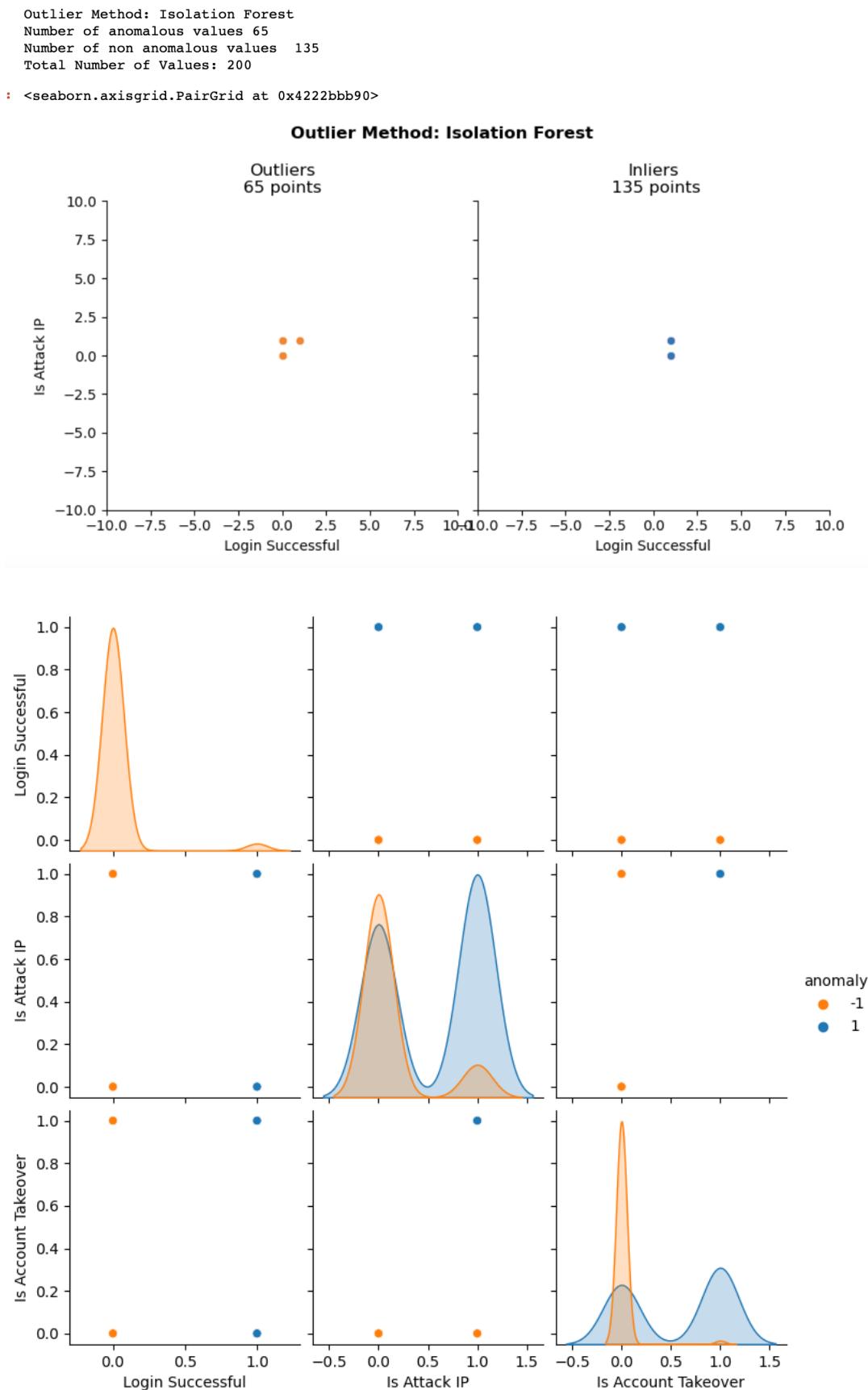


- contamination value == 0.3

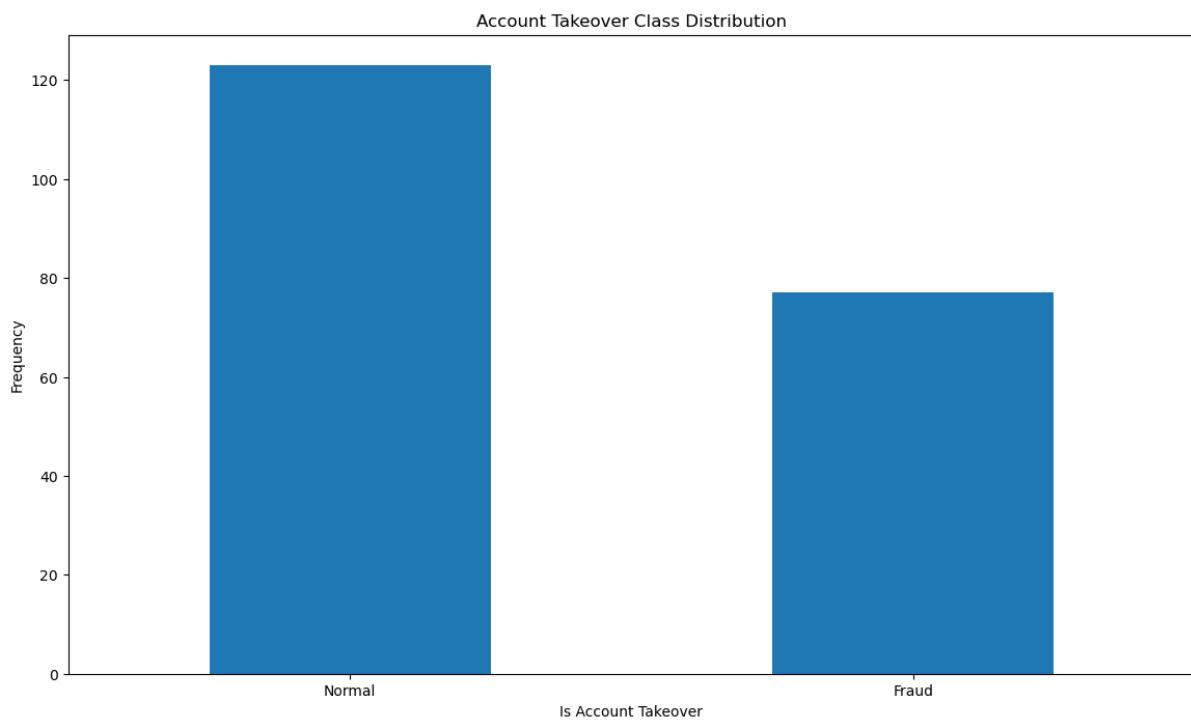
```
Outlier Method: Isolation Forest
Number of anomalous values 9
Number of non anomalous values 191
Total Number of Values: 200
: <seaborn.axisgrid.PairGrid at 0x2c782b5d0>
```



- contamination value == 0.5



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



"{}: {}".format(clf\_name,n\_errors) → np. Isolation Forest: 87

**Isolation Forest:** 87

**Accuracy Score :**

0.565

**Classification Report :**

	precision	recall	f1-score	support
False	0.60	0.88	0.71	123
True	0.25	0.06	0.10	77
accuracy			0.56	200
macro avg	0.42	0.47	0.41	200
weighted avg	0.47	0.56	0.48	200

**Local Outlier Factor:** 81

**Accuracy Score :**

0.595

**Classification Report :**

	precision	recall	f1-score	support
False	0.62	0.90	0.73	123
True	0.40	0.10	0.16	77
accuracy			0.59	200
macro avg	0.51	0.50	0.45	200
weighted avg	0.53	0.59	0.51	200

**Support Vector Machine:** 123

**Accuracy Score :**

0.385

**Classification Report :**

	precision	recall	f1-score	support
False	0.50	0.04	0.08	123
True	0.38	0.94	0.54	77
accuracy			0.38	200
macro avg	0.44	0.49	0.31	200
weighted avg	0.45	0.39	0.25	200