Μηχανική Μάθηση

Intrusion Detection System with the power of ML

Student:Κωνσταντακόπουλος Δημήτρης
Teacher: Δρ. Γιαννακόπουλος Θεόδωρος

Table Of Contents

Details about the dataset

-TCP/IP simulated attacks in a military grade LAN
-Samples are the connections
-Label Class: Normal or Anomalous
-Every Connection concludes 100 bytes of data
-41 features from every connection

Extra details from the dataset

Samples: 25192
-Class Normal: 13449
-Class Anomalous:11743
-Imbalance difference 1706

Types of Data Features:
3 Categorical
15 Float quantitive
23 Int quantitive

Null Values:
None

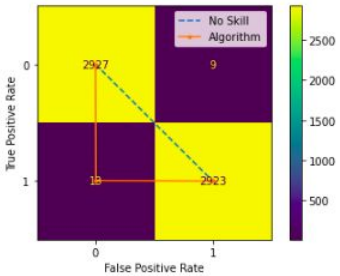
Duplicates:
None

Data Preprocessing

-Categorical to Discrete
-Scaling
-Outliers and features with no corrolation deletion
-Max corrolation features deletion
-Changes in Csv
-PCA when Needed
-Shuffling
-Correction of Imbalance

Classifiers Part 1

Desicion Tree with: Gini index, Max Depth=50, random_state=5, Splitter=best

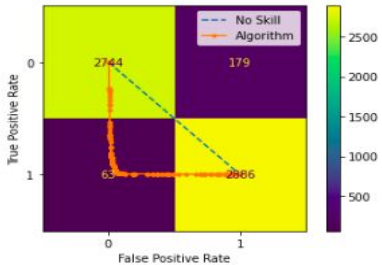



```
------------------------------
DECISIONTREE
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      2936
           1       1.00      1.00      1.00      2936

    accuracy                           1.00      5872
   macro avg       1.00      1.00      1.00      5872
weighted avg       1.00      1.00      1.00      5872
```

Quadratic Discriminant Analysis with: store_covariance=True,reg_param=0.5 WITH PCA



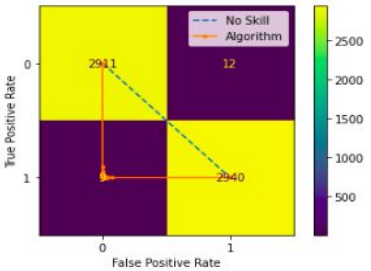

```
Quadratic
              precision    recall  f1-score   support

           0       0.98      0.94      0.96      2923
           1       0.94      0.98      0.96      2949

    accuracy                           0.96      5872
   macro avg       0.96      0.96      0.96      5872
weighted avg       0.96      0.96      0.96      5872
```

Classifiers Part 2

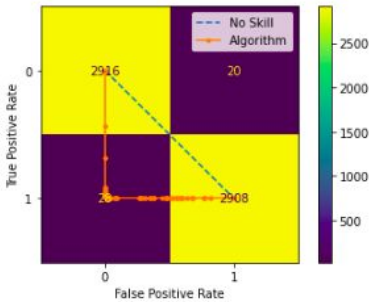
Random forest with bootstrap=False,class_weight=None,criterion='entropy',max_depth=30,n_estimators=50,warm_start=True AND PCA



RANDOM FOREST

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2923
1	1.00	1.00	1.00	2949
accuracy			1.00	5872
macro avg	1.00	1.00	1.00	5872
weighted avg	1.00	1.00	1.00	5872

MLP with activation='relu', solver='adam',learning_rate='adaptive', max_iter=200,nesterovs_momentum=True

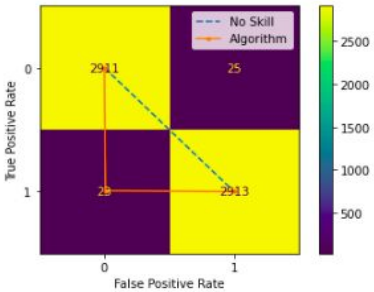


Multilayer-Perceptron

	precision	recall	f1-score	support
0	0.99	0.99	0.99	2936
1	0.99	0.99	0.99	2936
accuracy			0.99	5872
macro avg	0.99	0.99	0.99	5872
weighted avg	0.99	0.99	0.99	5872

Classifiers Part 3 AND More..

Kneighboors with n_neighbors=1,weights='uniform',algorithm='auto',n_jobs=1,leaf_size=1



Kneighbors

	precision	recall	f1-score	support
0	0.99	0.99	0.99	2936
1	0.99	0.99	0.99	2936
accuracy			0.99	5872
macro avg	0.99	0.99	0.99	5872
weighted avg	0.99	0.99	0.99	5872

More things to know...

-Cross Validation
-PCA
-Hypertuning
-ROC curves
-Learning Curves
-Pre-Trained Models
-Confusion Matrices

Conclusions

-Which Classifier is the best?
-Could be done more?
-Possible future exploration
-Are anomaly-based ML IDS applicable in real world?

THANK YOU FOR YOUR TIME

QUESTIONS?