ICS CLASSIFICATION CYPER SECURITY

Ahmed Al-Muaybid
Abdulrahman Al-Rifae

Wonder Team

Dunkin Tech

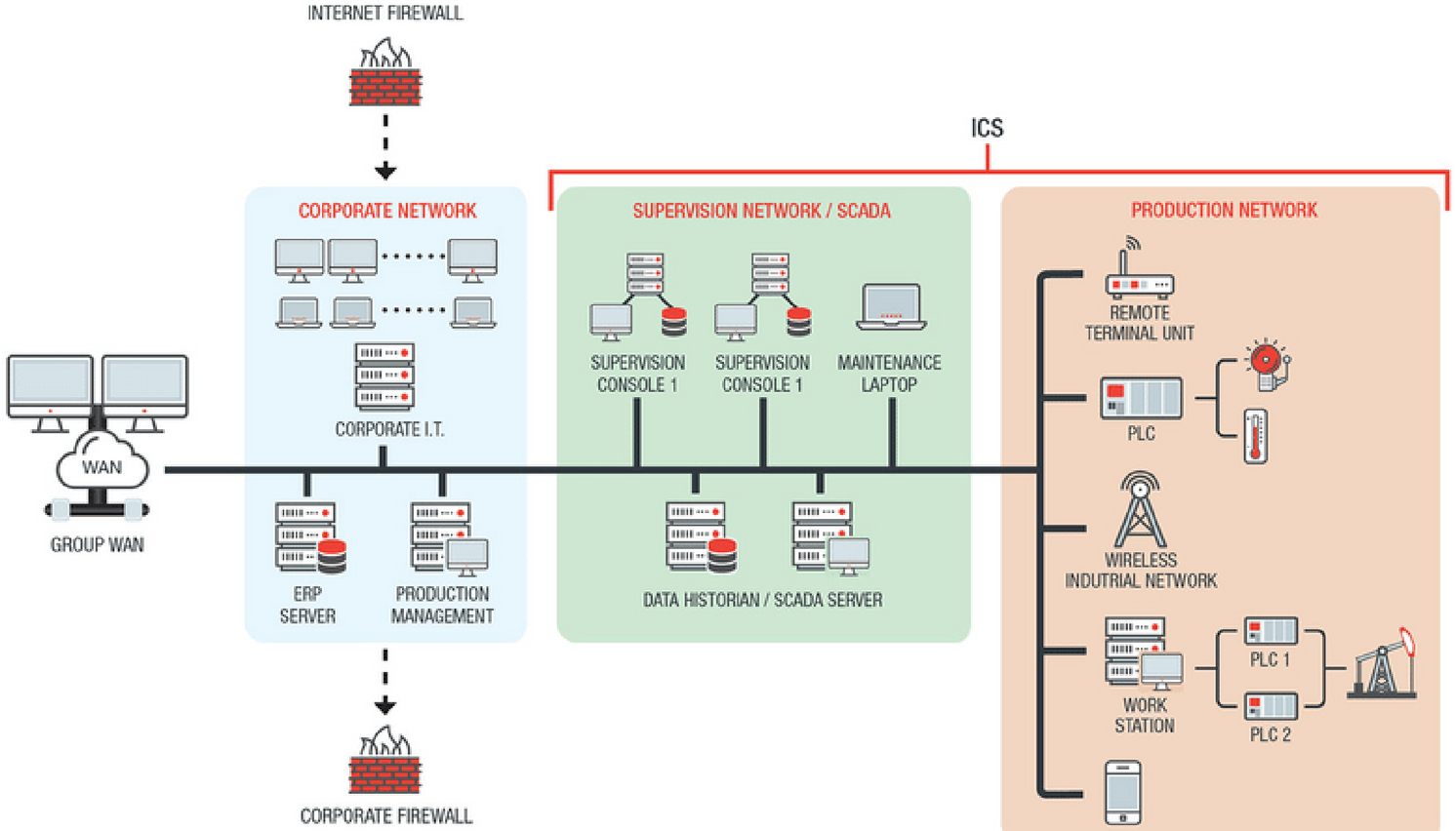
Agenda

- Introduction
- Methadology
- DATA
- Model Selection
- Conclusion

Introduction

ICS

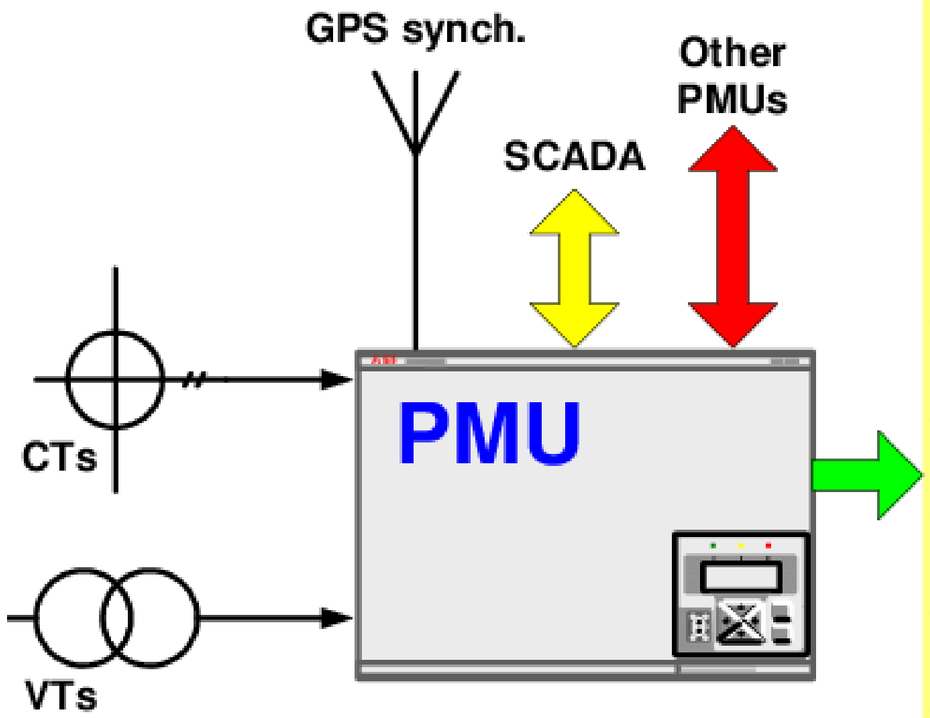
-Industrial Control Systems (ICS) are commonly used in various domains, where it consists of many devices to automate the industrial process.



Introduction

SCADA

-Control systems for ICS is Supervisory Control and Data Acquisition (SCADA) systems



OUTPUT, examples:

- **Voltage and Current Phasors**
- **Real and Reactive Power**
- **Phasor differences**
- **Stability Indices**
- **Power Transfer Margins**
- **Trip signals**
- **Control signals**

Data Science Methodology

Identifying the problem and the approach to fix the problem

Data requirements and collection methods

Understand the data

Generate models and evaluate them

Deploy the model and get feedback

DATA

Thanks to Tommy Marries

Understanding and Preparing the Data

is this data going to answer problem that I am having?

+10,000 rows

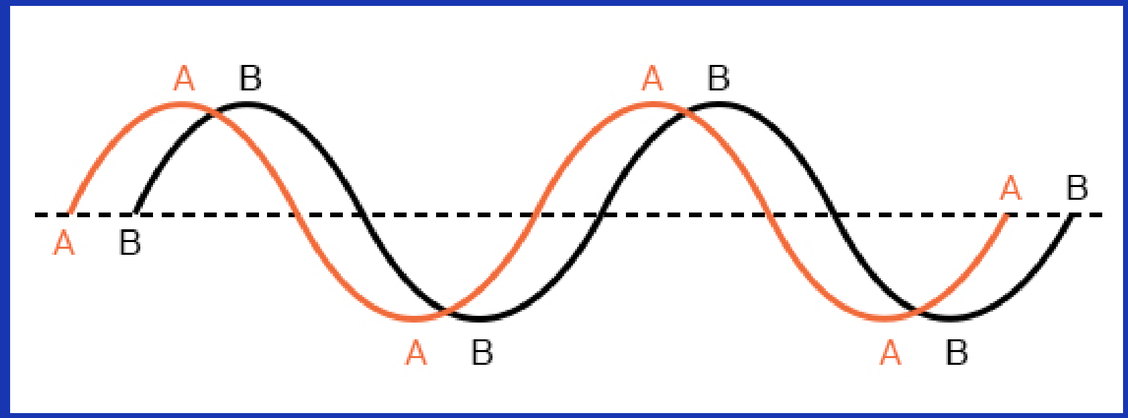
129 Columns\Featurs

-4 PMU

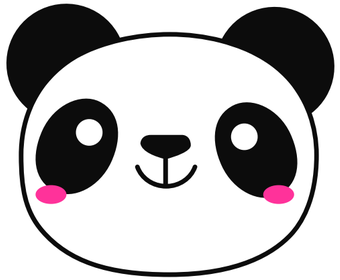
- Each PMU give 29 featurs

-Other 12 featurs

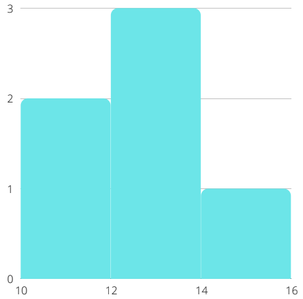
- Target is called [Marker]



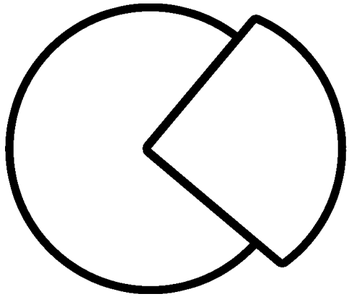
Tools



Pandas, Numpy

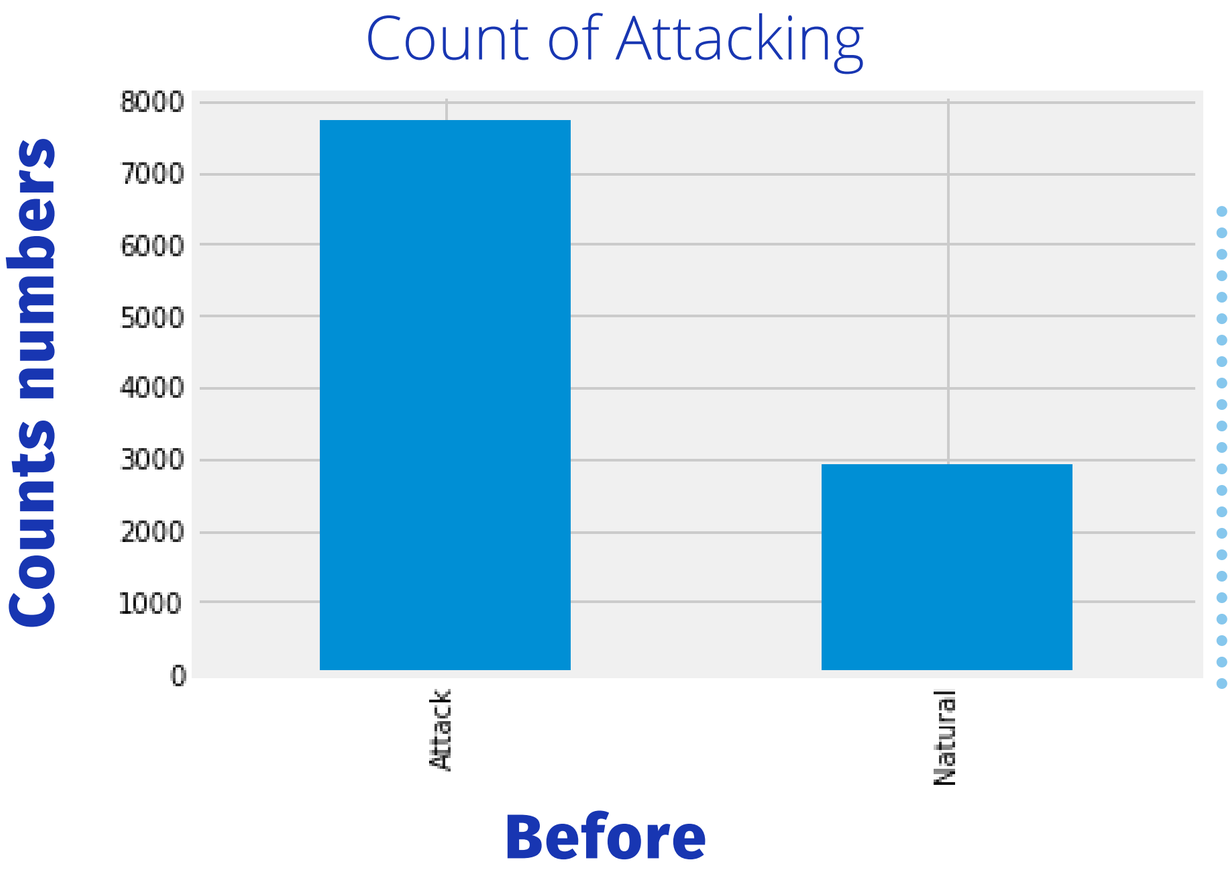


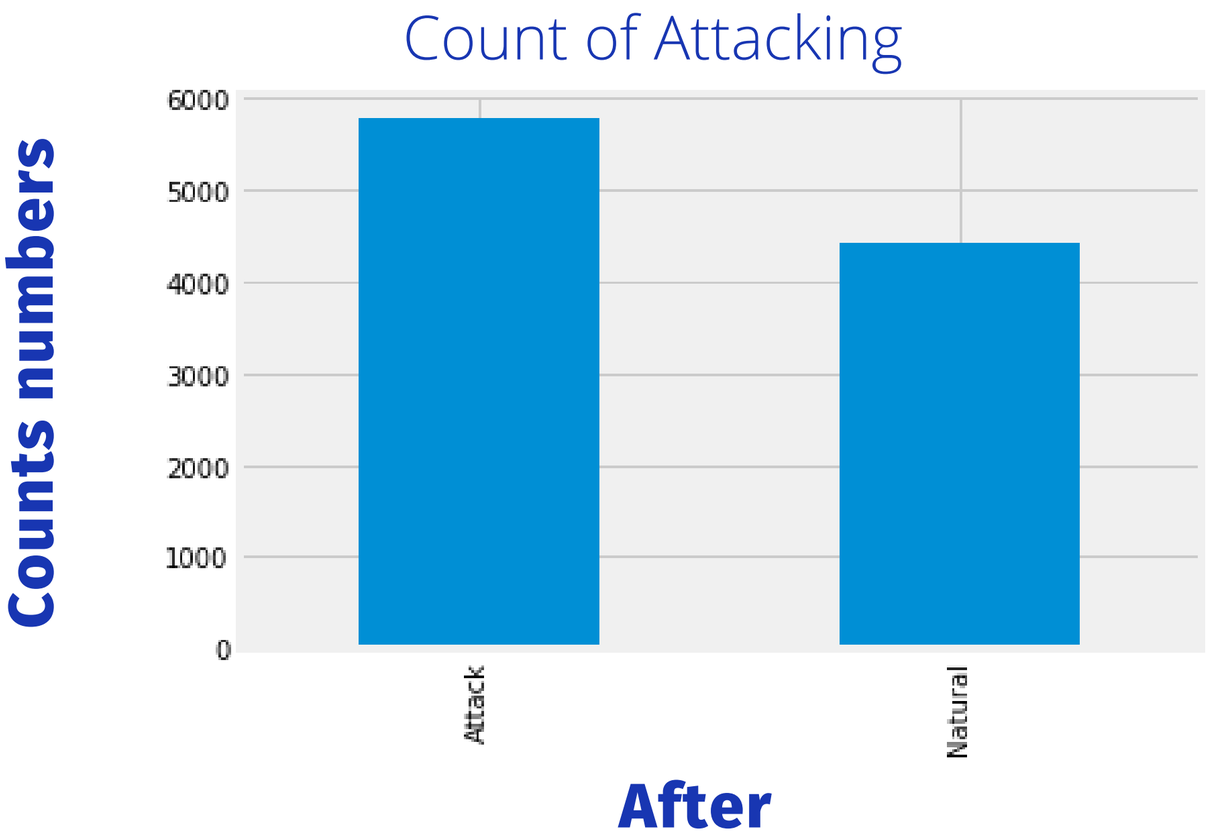
Seaborn,Matplotlib



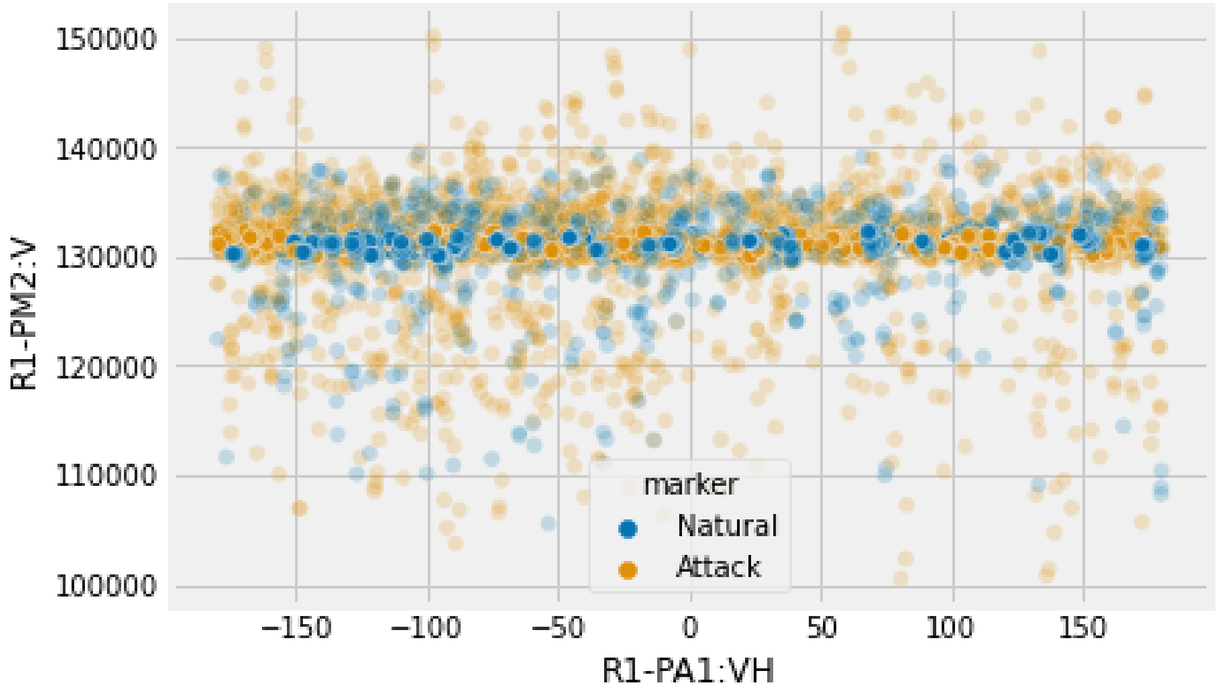
Sklearn

Imbalance


Count of Attacking
Counts numbers
8000
7000
6000
5000
4000
3000
2000
1000
0
Attack
Natural
Before


Count of Attacking
Counts numbers
6000
5000
4000
3000
2000
1000
0
Attack
Natural
After

PairPlot



PairPlot (Selected Featuers)

Model Experiments

	Exp.Name	F1	Precision	Recall	Accuracy	AUC	Time
0	KNN-5	0.842677	0.728947	0.998455	0.728126	0.499228	0.670456
1	KNN-5_Scaled	0.918300	0.907901	0.928939	0.879459	0.837563	0.698523
2	KNN-5_Scaled_cv10	0.899252	0.886721	0.912252	0.852113	0.916127	10.902511
3	KNN_best_acc	0.924939	0.926252	0.923770	0.891549	0.865511	3726.142044
4	KNN_best_F1	0.924939	0.926252	0.923770	0.891549	0.865511	3726.142044
5	KNN_best_Precision	0.900838	0.945364	0.860513	0.863028	0.895221	3726.142044
6	KNN_best_Recall	0.849228	0.773344	0.941771	0.757981	0.775913	3726.142044
7	KNN_best_AUC	0.899252	0.886721	0.912252	0.852113	0.916127	3726.142044
8	Logisitic_C1000	0.843852	0.732651	0.994851	0.731506	0.508521	0.420177
9	Logisitic_C1000_PCA	0.843825	0.731219	0.997425	0.730755	0.504954	0.026924
10	DecisionTree_depth4	0.855163	0.755529	0.985067	0.756665	0.563269	0.265449
11	DecisionTree_depth4_PCA	0.741130	0.756844	0.726056	0.630116	0.548881	0.045846
12	DecisionTree_depth10_PCA	0.804614	0.754394	0.861998	0.694705	0.553052	0.101723
13	DecisionTree_depth4	0.889642	0.839269	0.946447	0.828765	0.729118	0.585146
14	RandomForest_n100	0.965535	0.943925	0.988157	0.948554	0.915021	5.003665
15	RandomForest_n100_PCA	0.832735	0.737490	0.956231	0.719865	0.519724	2.186644
16	KNN_best_acc_pca	0.944728	0.944044	0.945499	0.919953	0.899289	996.218218
17	KNN_best_F1_pca	0.944728	0.944044	0.945499	0.919953	0.899289	784.917531
18	KNN_best_Recall_pca	0.836442	0.735135	0.970154	0.725469	0.672736	784.917531
19	KNN_best_AUC_pca	0.926956	0.917459	0.936740	0.893192	0.923348	784.917531
20	Logit_best_acc	0.840225	0.730611	0.988808	0.727817	0.687563	1533.411719
21	Logit_best__F1	0.840225	0.730611	0.988808	0.727817	0.687563	1533.411719
22	Logit_best_Precision	0.837545	0.734963	0.973560	0.726761	0.687014	1533.411719
23	Logit_best_Recall	0.838008	0.724595	0.993512	0.722066	0.689512	1533.411719
24	Logit_best_AUC	0.838245	0.726966	0.989781	0.723592	0.692307	1533.411719
25	Logisitic_PCA	0.843825	0.731219	0.997425	0.730755	0.504954	0.443292
26	Logisitic_C10	0.843558	0.734172	0.991246	0.731881	0.512267	0.409369
27	Stacking_with_best_para	0.965778	0.944390	0.988157	0.948930	0.915715	0.409369
28	Stacking_with_best_para_balanced	0.972053	0.959378	0.985067	0.958693	0.936361	10.027806
29	Stacking_with_best_para_balanced_pca	0.834660	0.731216	0.972194	0.719114	0.504821	4.316928
30	DecisionTree_best_acc	0.929802	0.932476	0.927170	0.898709	0.875687	334.737644
31	DecisionTree_best__F1	0.929802	0.932476	0.927170	0.898709	0.875687	334.737644
32	DecisionTree_best_Precision	0.929802	0.932476	0.927170	0.898709	0.875687	334.737644
33	DecisionTree_best_Recall	0.839791	0.732647	0.983779	0.728404	0.574142	334.737644
34	DecisionTree_best_AUC	0.929802	0.932476	0.927170	0.898709	0.875687	334.737644
35	rf_best_acc	0.965563	0.947268	0.984590	0.949178	0.986423	996.218218
36	rf_best__F1	0.965563	0.947268	0.984590	0.949178	0.986423	996.218218
37	rf_best_Precision	0.888754	0.952846	0.832766	0.849178	0.904680	996.218218
38	rf_best_Recall	0.964163	0.943084	0.986212	0.946948	0.985850	996.218218
39	rf_best_AUC	0.965563	0.947268	0.984590	0.949178	0.986423	996.218218

Model Selection

Model Name	F1	Precision	Recall	Accuracy	AUC	Time (s)
Knn	0.842	0.728	0.998	0.728	0.499	0.670
Knn Tunning	0.918	0.907	0.928	0.879	0.837	3726
Logistic	0.837	0.734	0.973	0.726	0.687	0.026
Logistic_PCA Tunning	0.843	0.731	0.997	0.730	0.504	1533

Model Selection

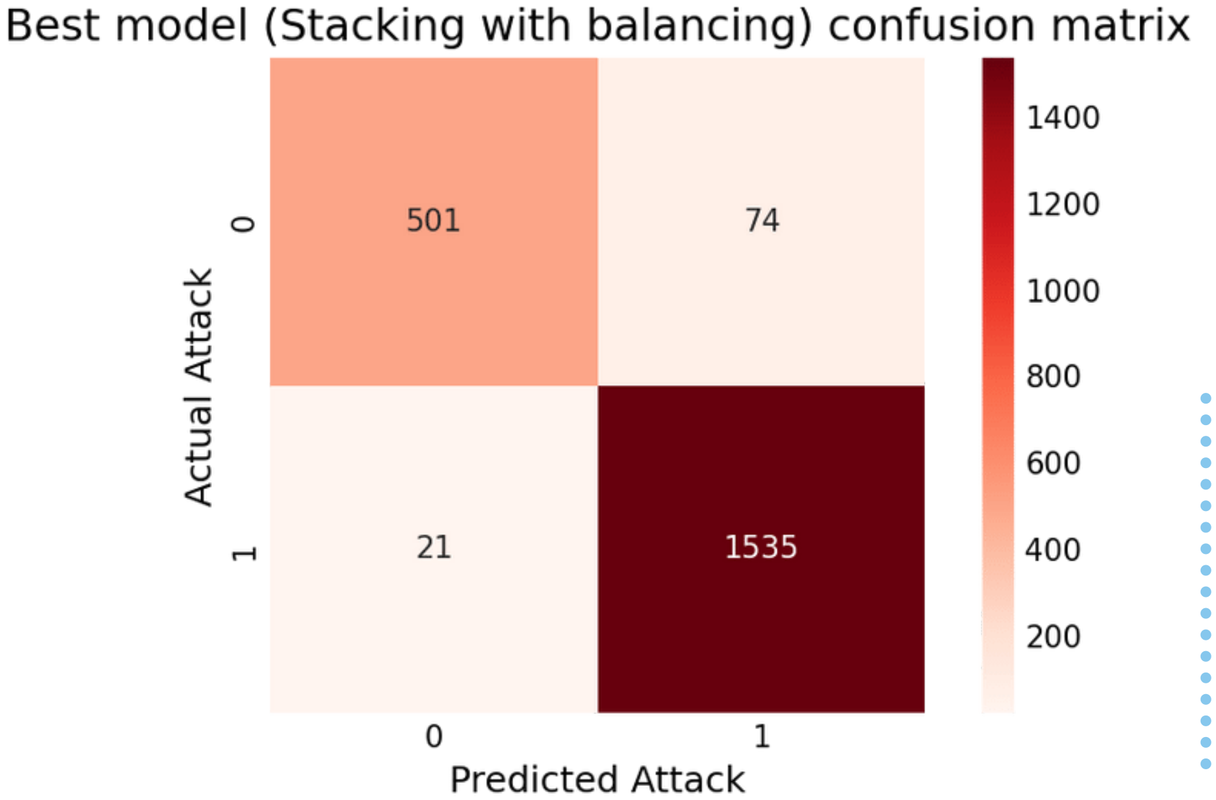
Model Name	F1	Precision	Recall	Accuracy	AUC	Time (s)
Decision Tree	0.929	0.932	0.927	0.898	0.875	334.737
Random Forest	0.965	0.943	0.988	0.948	0.915	5.003
Random Forest PCA	0.832	0.737	0.956	0.719	0.519	2.186
Stacking	0.972	0.959	0.985	0.958	0.936	10.

Golden Model

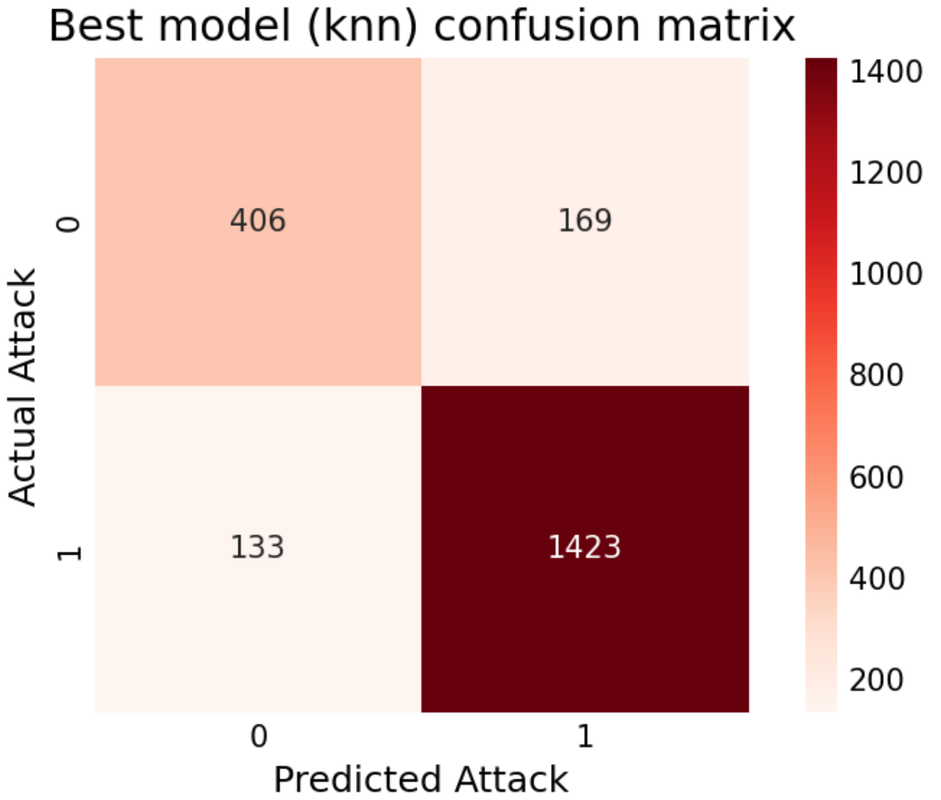
Stacking

Model Name	F1	Precision	Recall	Accuracy	AUC	Time
Stacking	0.972	0.959	0.985	0.958	0.936	10.027

Confusion Matrix



Stacking



Knn

Deliverable

01

Classify the scenario based on SCADA measrments

02

Stacking Model is the Best Model for this problem

03

GridSearch Tunning take a long time to run

04

KNN is very fast to excute

Future Plan

01

What type of Attack
- Data Injection
-Tripping Injecction

02

What type of Normal fault
- Normal operation
- Line maintenance

Conclusion

Finding the best model

GridSearchCV for Tuning

Random forest is a very powerful model

Thanks for Listening
Any Questions?