

RoboGarden Bootcamp Capstone Project

Credit Card Fraud Detection
July 2019

RoboGarden Bootcamp

Credit Card Fraud Project

- **Description:** The dataset consists 284,807 credit card transactions made by European cardholders in September 2013
- **Link:** <https://data.world/raghuv543/credit-card-fraud-data>
- **File:** creditcard.csv
- **Features:**
 - Time (integer): transaction time (seconds since first transaction)
 - V1 – V28 (decimal): results of PCA transformation
 - Amount (decimal): transaction amount
 - Class (T/F): fraudulent/genuine transaction (target)
- **Notes:**
 - Feature V1-V28 are the principal components obtained with PCA
 - The original features and background information about the data are not provided due to confidentiality issues
- **License:** Public domain (CC0)

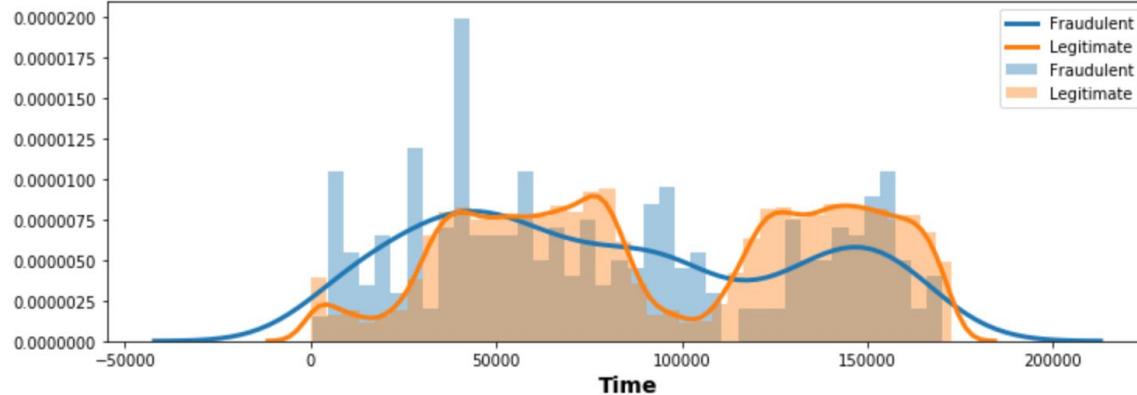
RoboGarden Bootcamp

Credit Card Fraud Project – By the Numbers

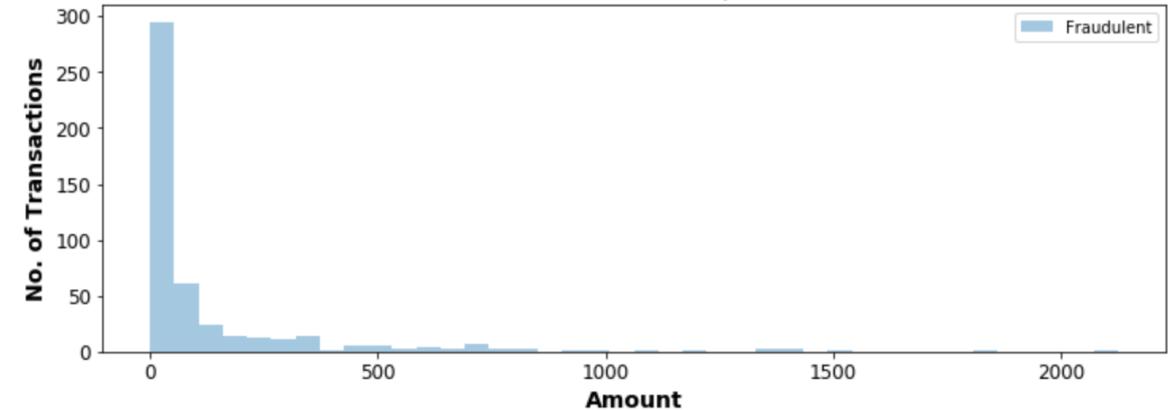
- *Samples:* **284,807 transactions over 2 days (average 1.6 per second)**
- *Fraud vs. Normal Transactions:* **492 fraud Extremely Unbalanced**
- *Amounts:* Normal total: **25,043,410** vs. **58,591** fraud (without duplicates)
- *0.00 Amount Transactions:* **1798 zero amount normal transactions (0.6%)**
24 zero amount fraud transactions (5.5%)
 - Considered dropping these if 0.00 fraud was a transaction that was caught
 - This doesn't explain the normal transactions.
 - Could be a card verification for reservation etc.
 - Not enough information to remove these transactions
- *Duplicates:* **1062 (0.37 %) non-fraud duplicates and 19 (4.0 %) fraud duplicates**

RoboGarden Bootcamp Visualization

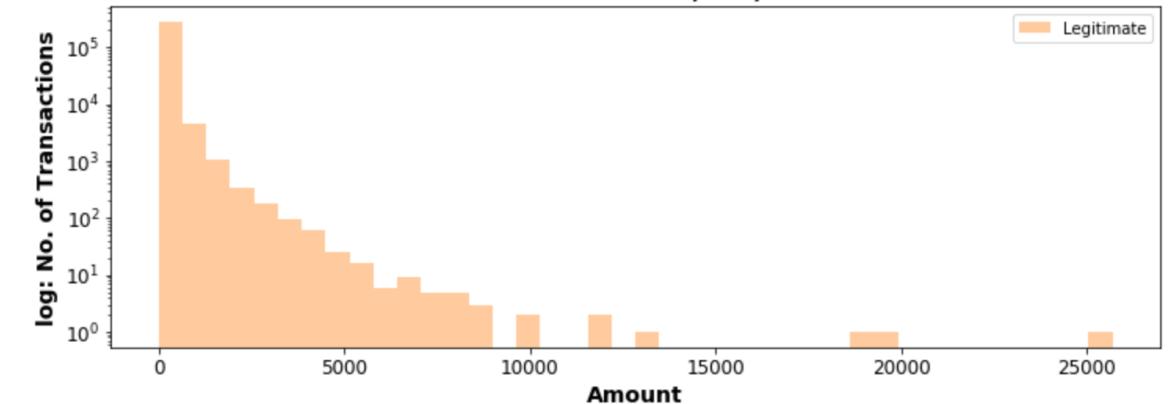
Transaction Time Histogram



Fraudulent Transactions Histogram
Total Amount 58,591



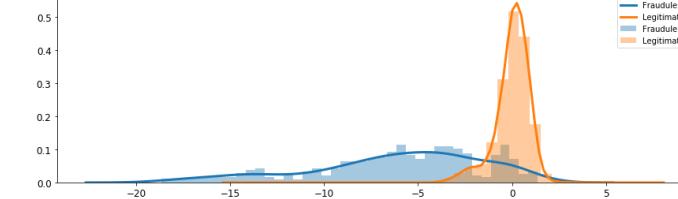
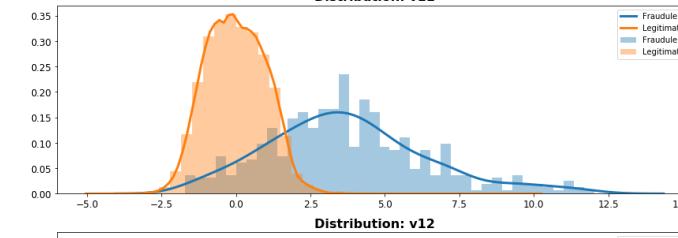
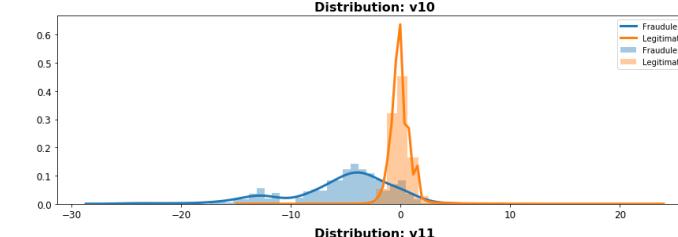
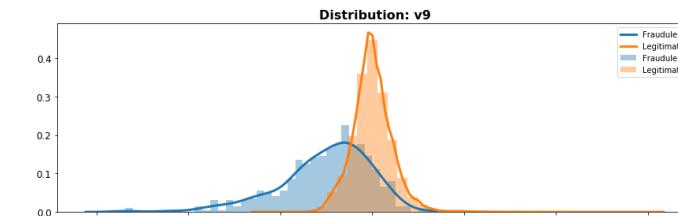
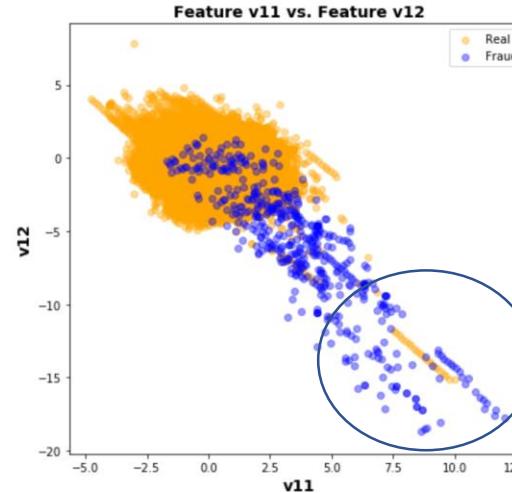
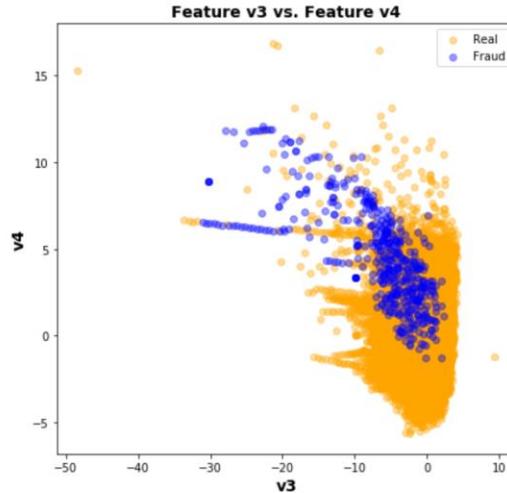
Legitimate Transactions Histogram
Total Amount 25,043,410



- Frauds are more frequent in the late hours*
- Lots of small value transactions
- Frauds are more weighted to small amounts although a few go up to ~2,000
- Normal transactions have higher spending amounts

* Assume time starts at midnight

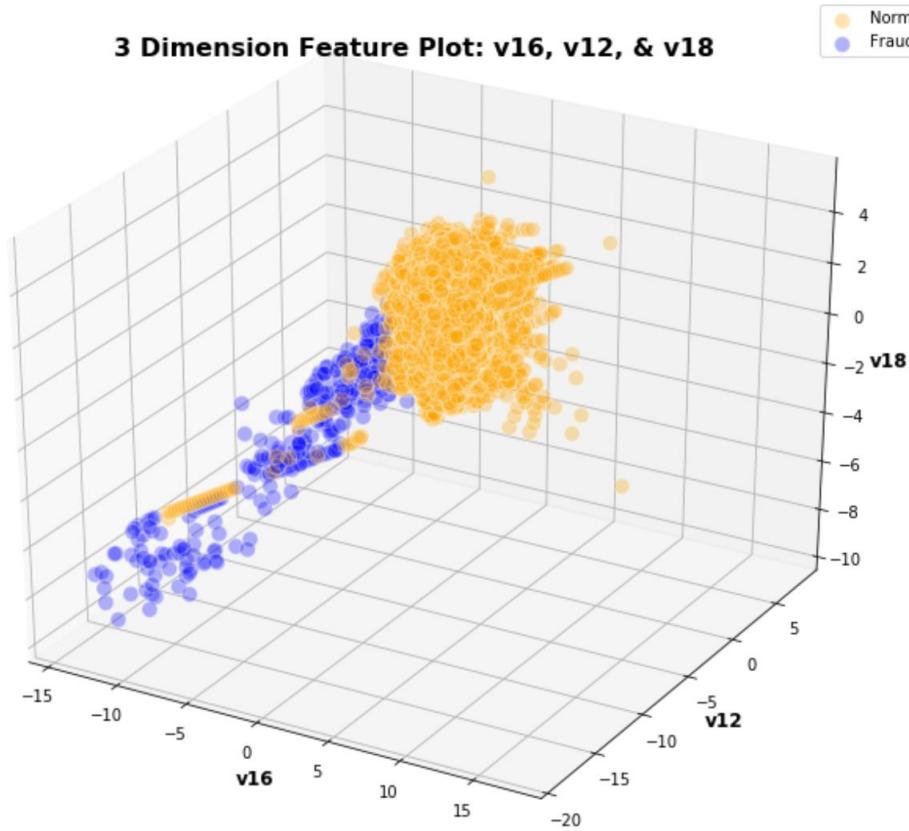
RoboGarden Bootcamp Visualization



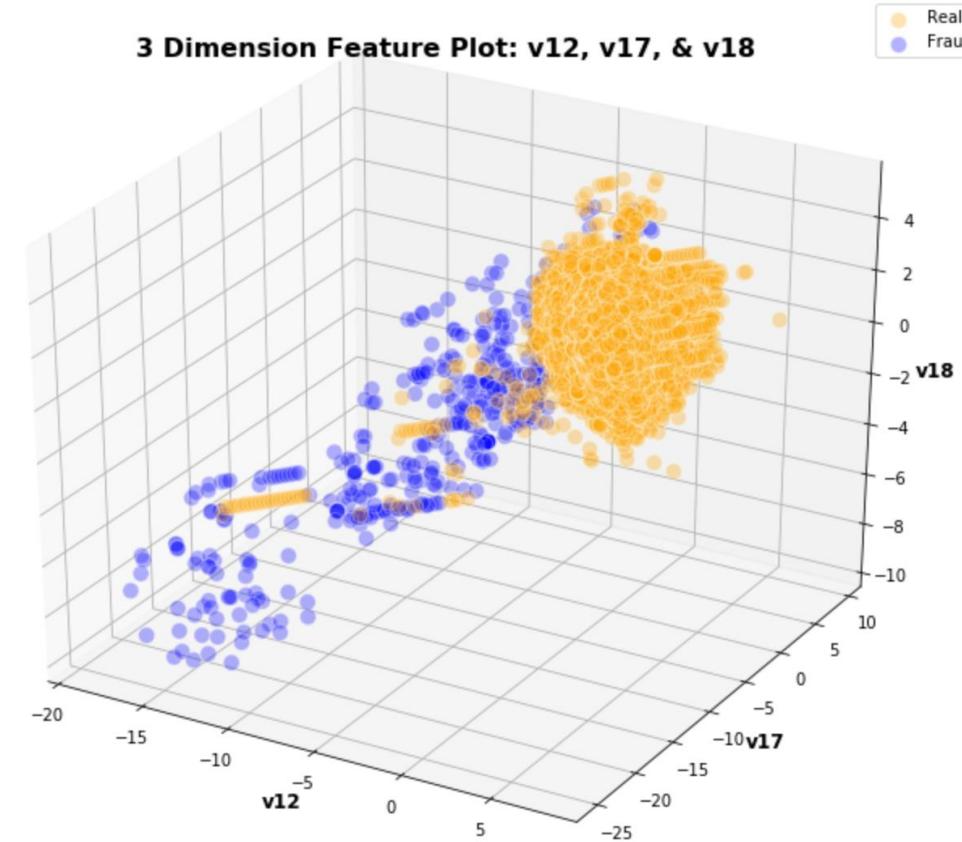
- Frauds vs Normal histograms show overlap but distinct differences. Some distributions are aligned.
- The frauds that align with the normal distributions may be difficult to detect
- 2D Scatter plots of some features show the same overlap & separation.

RoboGarden Bootcamp Visualization

3 Dimension Feature Plot: v16, v12, & v18



3 Dimension Feature Plot: v12, v17, & v18



RoboGarden Bootcamp

Model Application

	All Features	Some Features	
All Samples	<ul style="list-style-type: none">• Logistic Regressor• SGD Classifier• Random Forest• AdaBoost• QuadraticDiscriminantAnalysis*	<ul style="list-style-type: none">• Random Forest (10)• MLP (10)• Autoencoders (10, 4)	All Samples
Some Samples	<ul style="list-style-type: none">• Random Forest• MLP		

← Under Sampling

* Eliminated after first pass

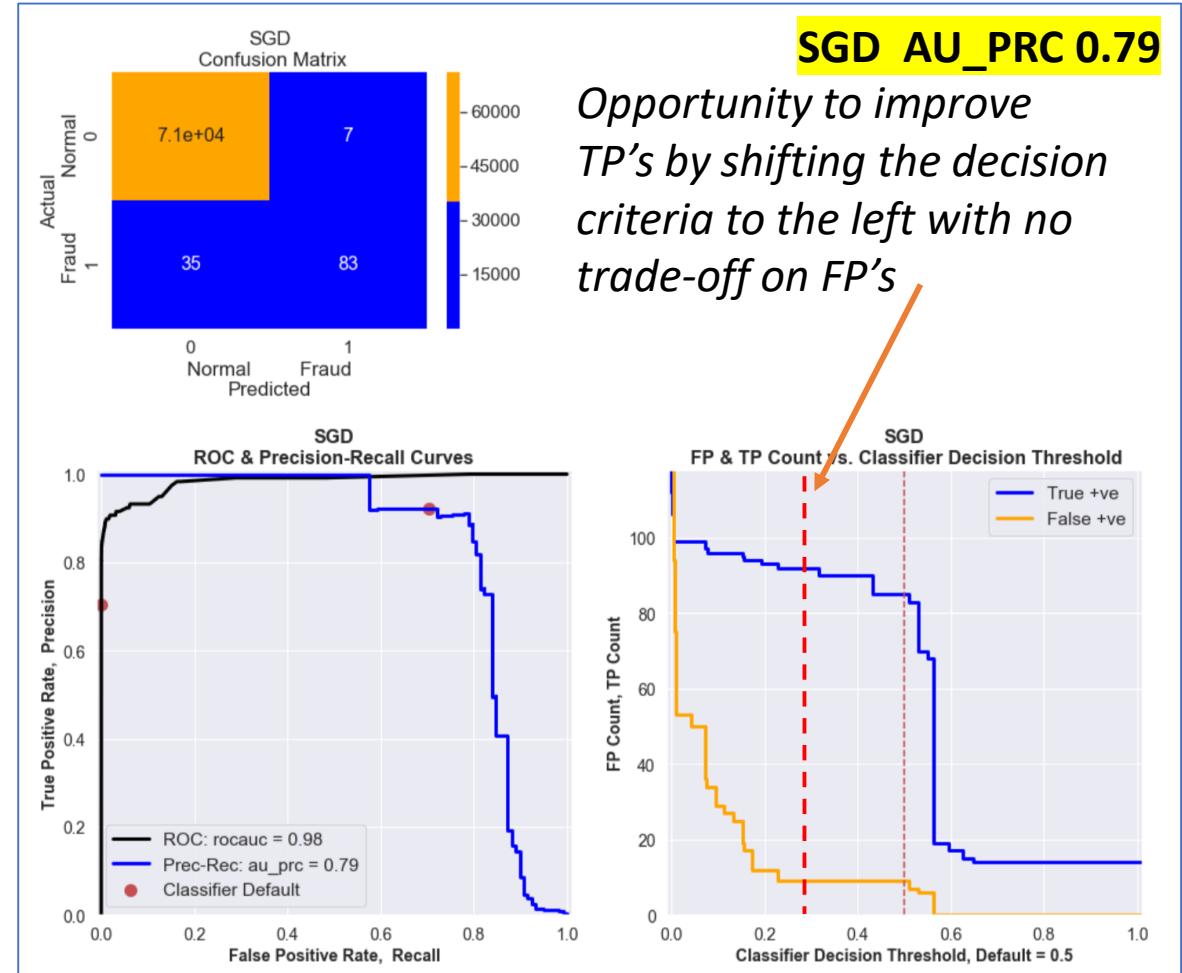
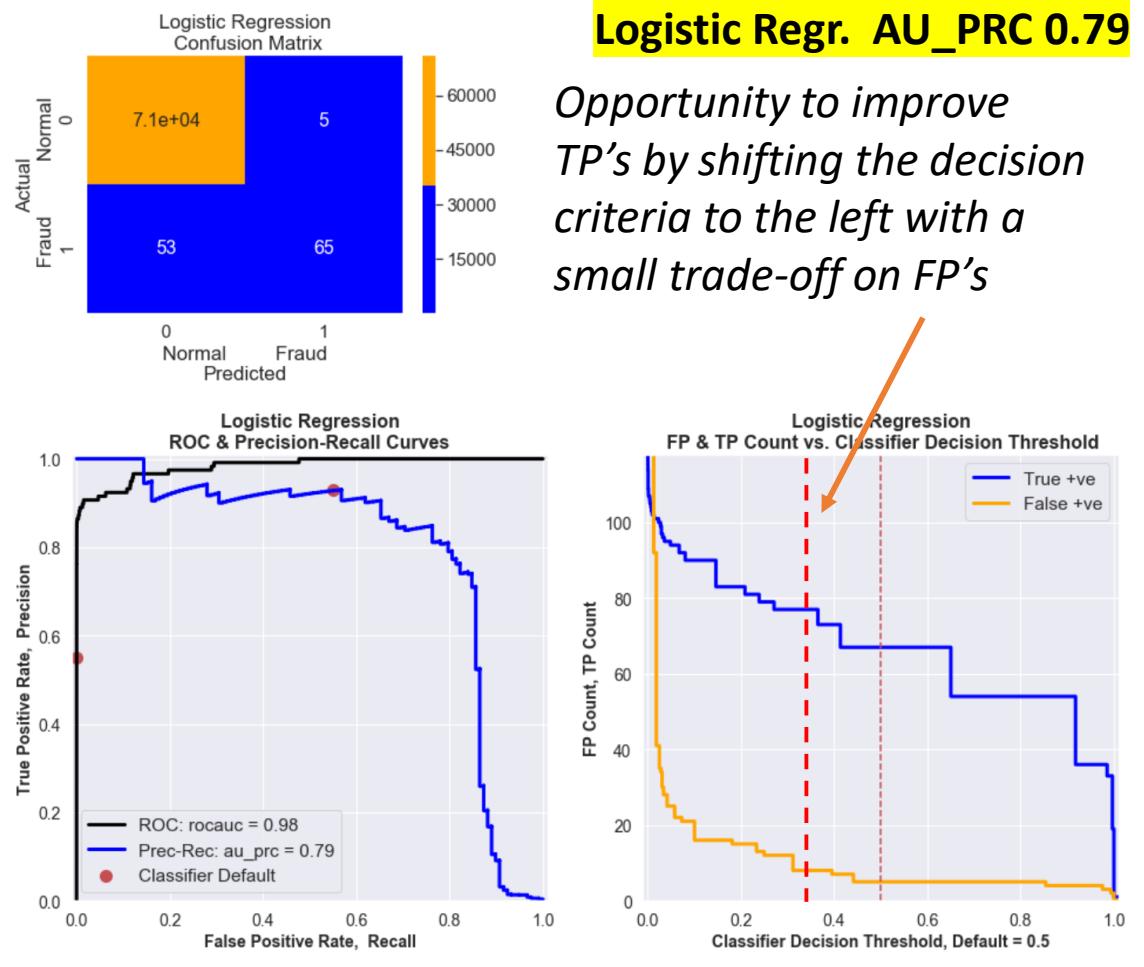
RoboGarden Bootcamp

Model Results

Model	Application	# Features	AU-PRC	TP	FP	FN
Logistic Regressor	All Samples	29	0.79	65	5	53
SVC	All Samples	29	0.79	83	7	35
Random Forest	All Samples	29	0.86	92	4	26
SVC	All Samples	29	0.80	77	3	41
AdaBoost	All Samples	29	0.80	87	8	31
MLP	All Samples	29	0.86	95	6	23
Random Forest	All Samples	10	0.84	90	4	28
MLP	All Samples	10	0.83	89	9	29
Random Forest	Under Sampling	29	0.76	116	29,000	2
MLP	Under Sampling	29	0.77	118	22000	0
Autoencoder	All Samples	10	0.57	66	45	52
Autoencoder	All Samples	4	0.75	84	13	34

RoboGarden Bootcamp

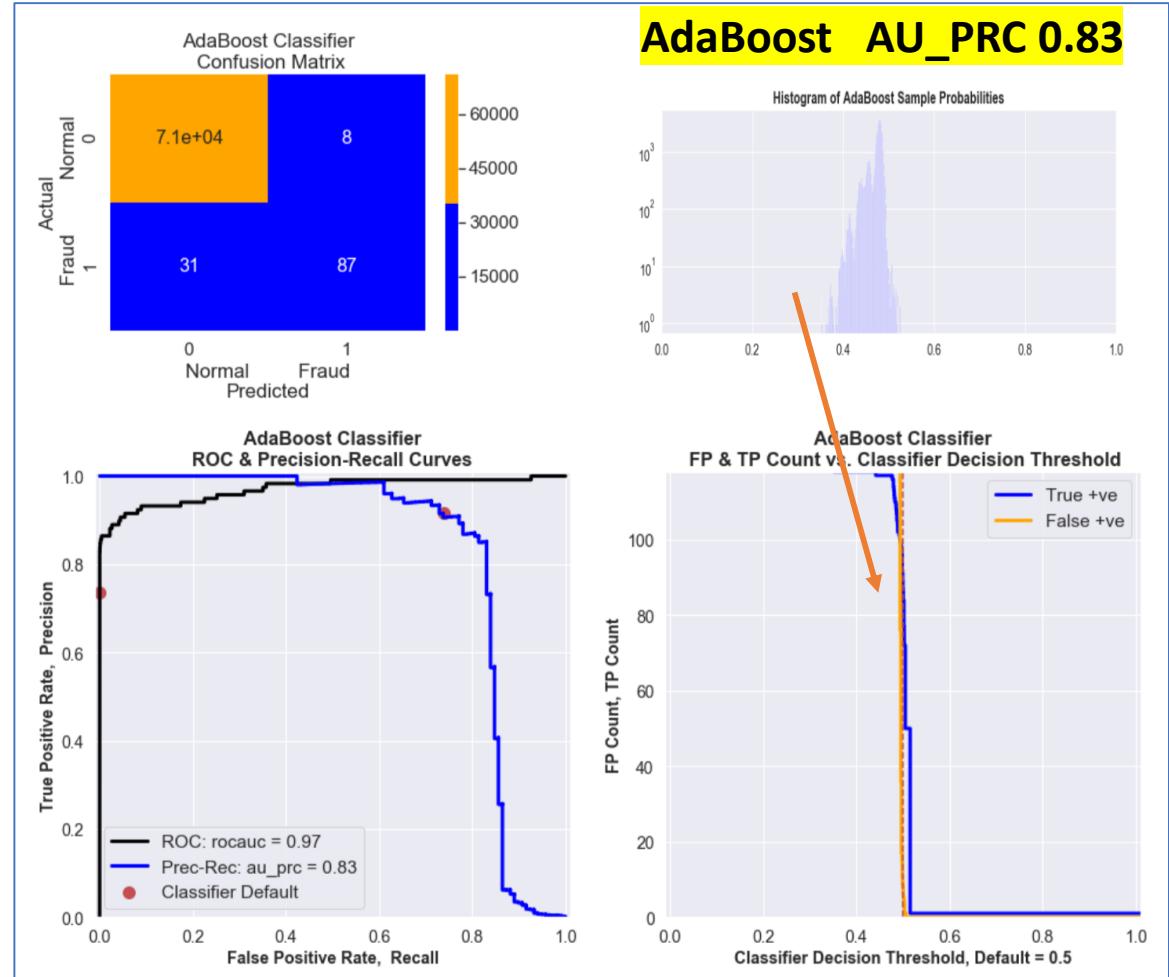
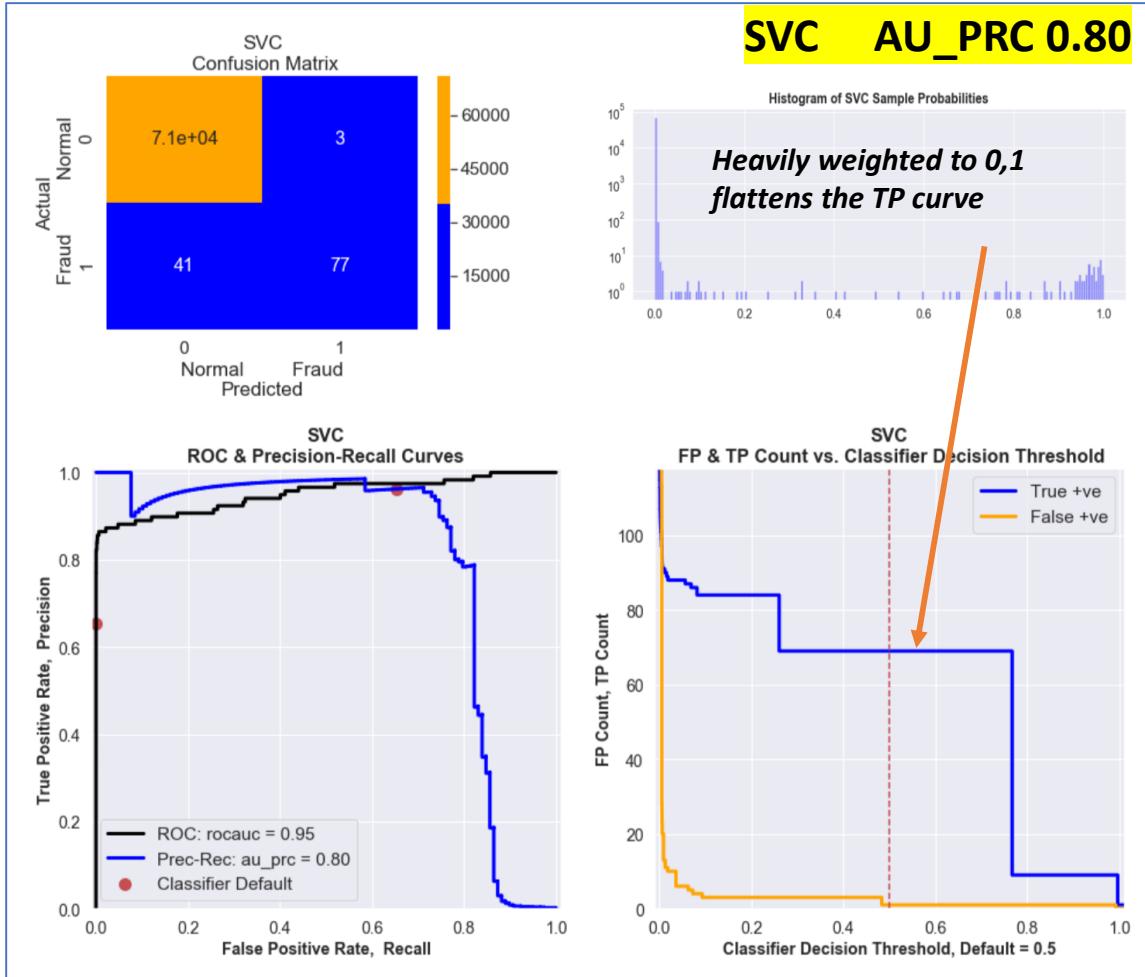
Logistic Regression & SGD Trained on 29 Features



* 29 Features are Amount + Feature v1 to v28

RoboGarden Bootcamp

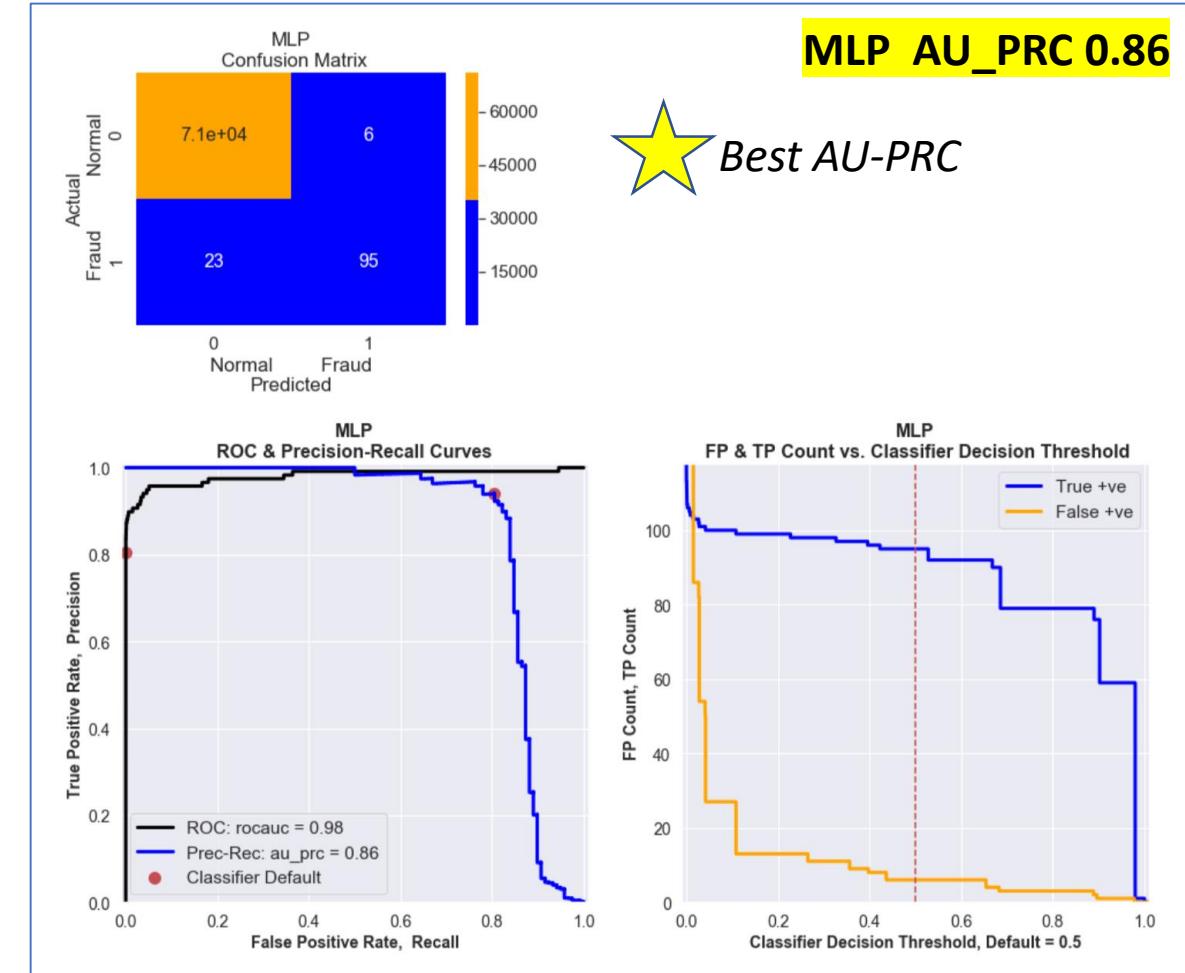
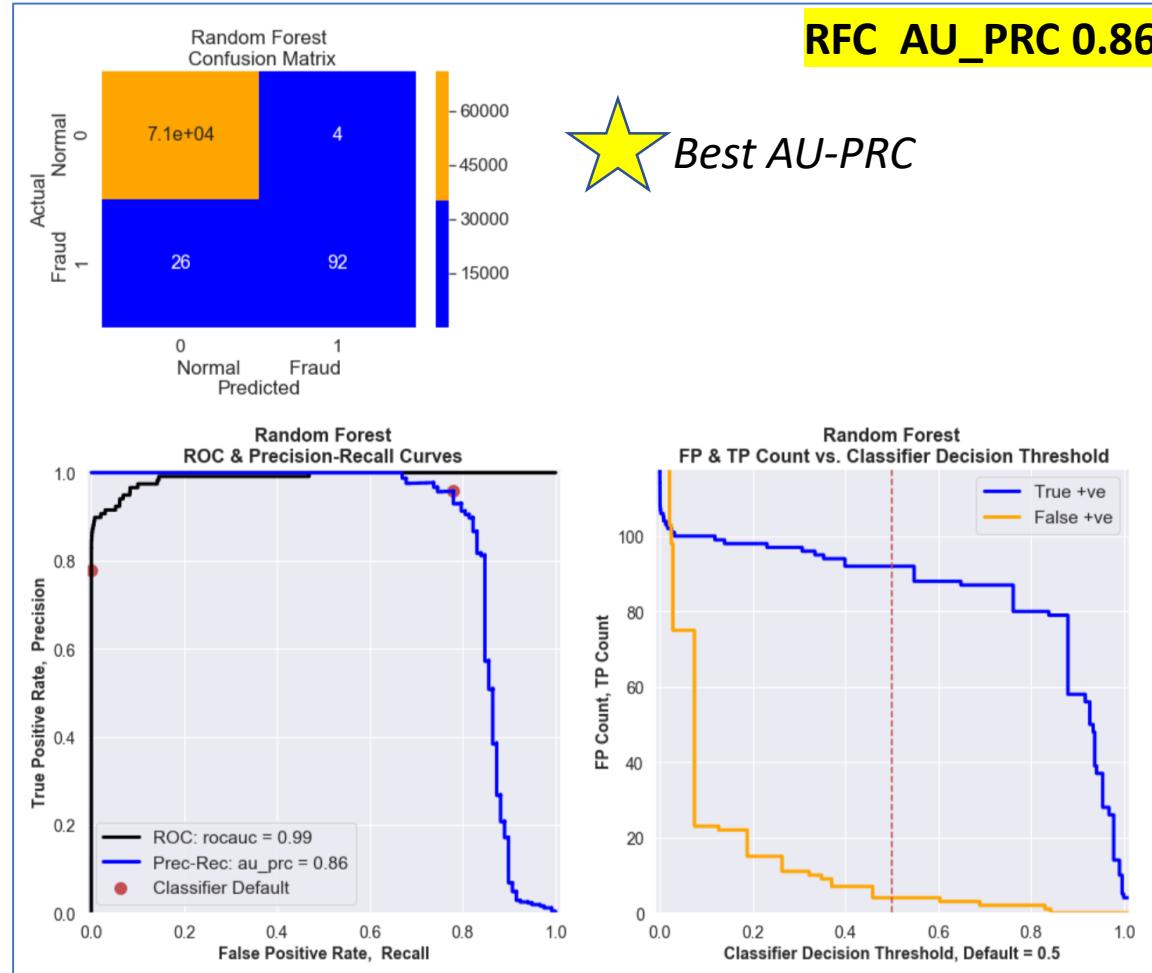
SVC & AdaBoost Trained on 29 Features



* 29 Features are Amount + Feature v1 to v28

RoboGarden Bootcamp

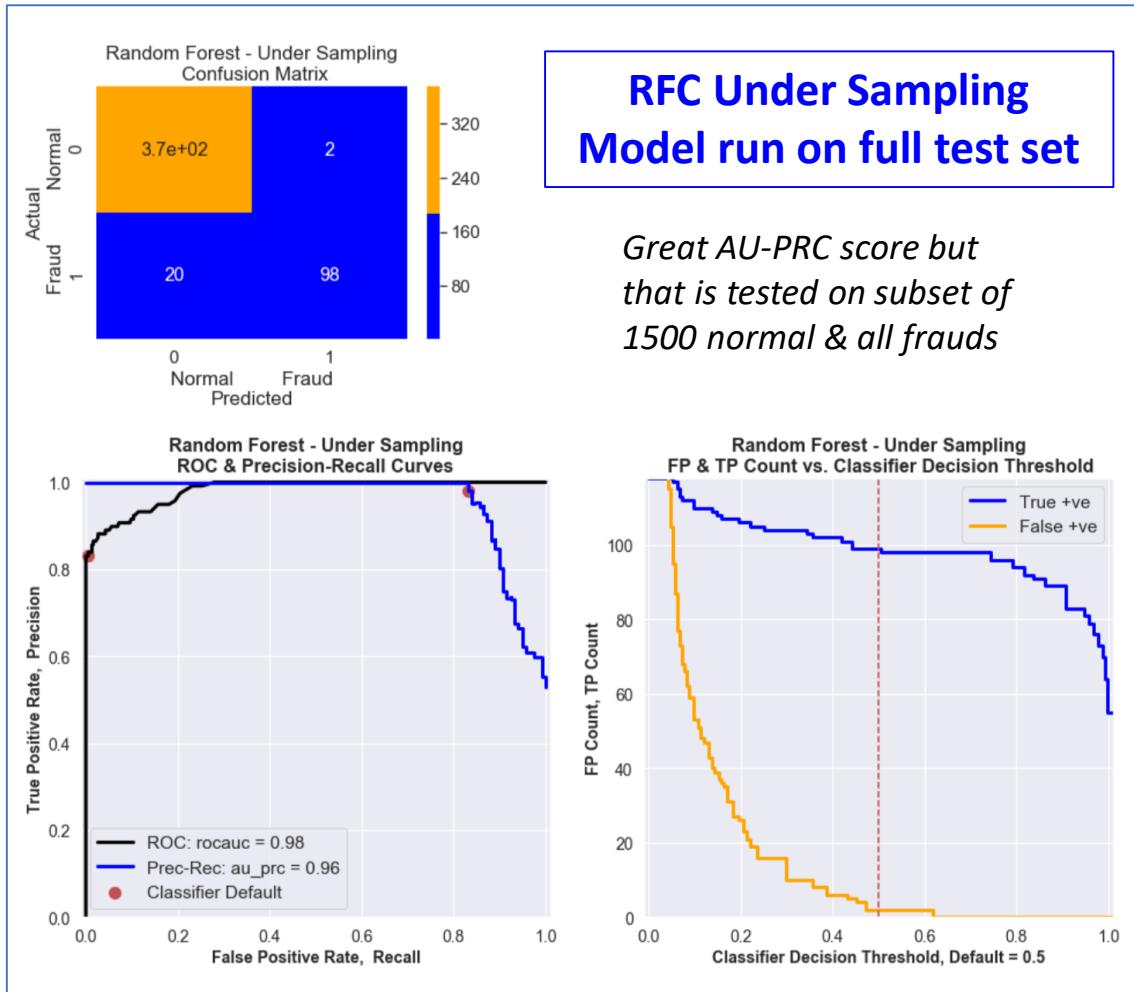
Random Forest & MLP Trained on 29 Features



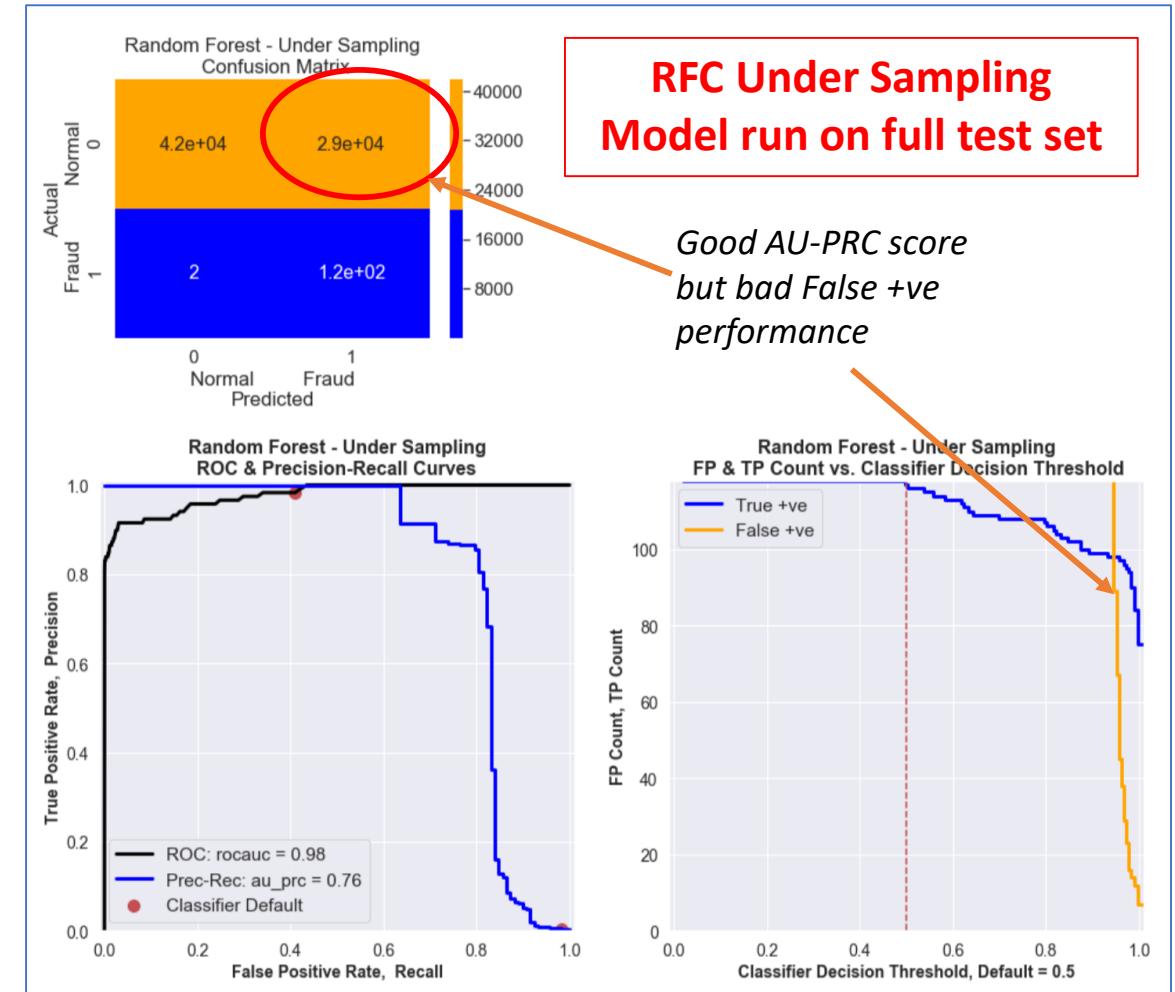
* 29 Features are Amount + Feature v1 to v28

RoboGarden Bootcamp

Random Forest on Under Sample Dataset after Optimization



Model trained on U-S train set & tested on U-S test set →

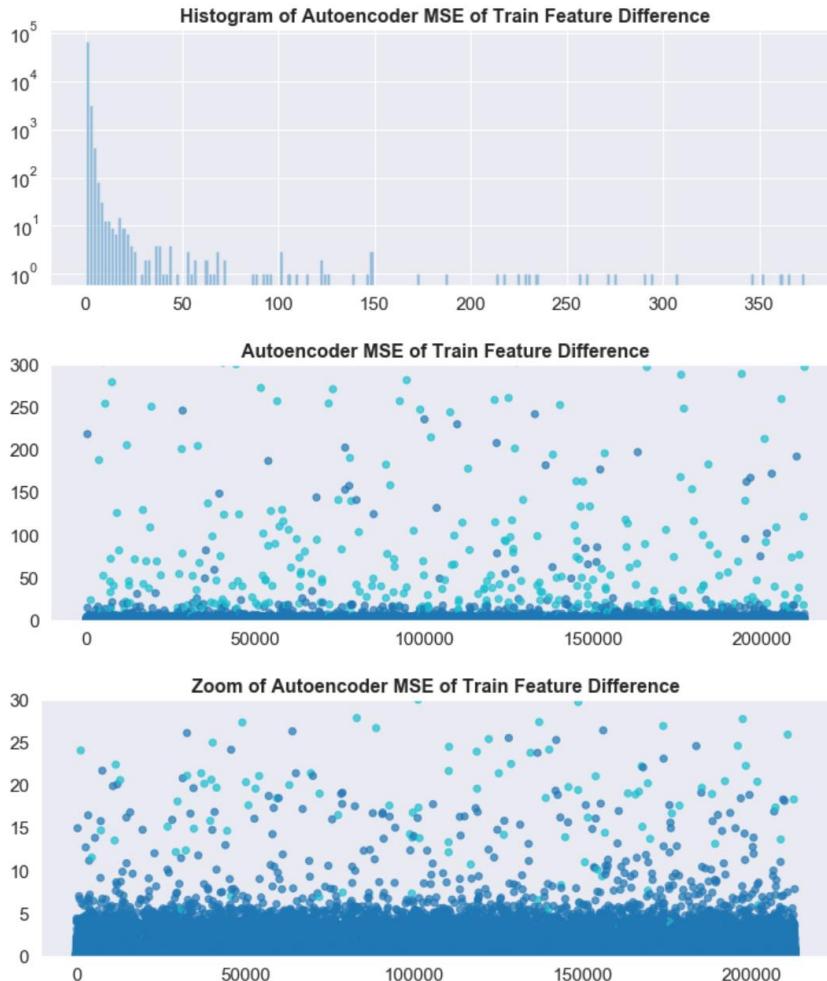


Then Model tested on original test set (this is target test).

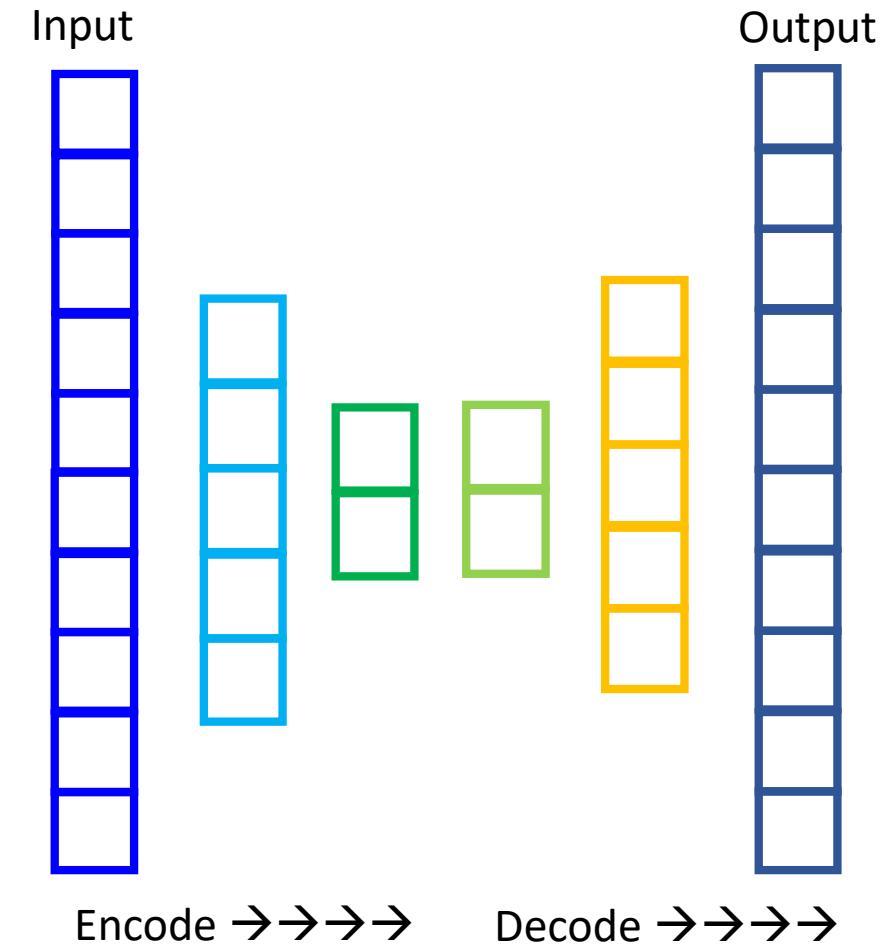
RoboGarden Bootcamp

Autoencoders

MSE stats for train set: min 0.0003, max 373.4034, mean 0.7072, std 5.2781

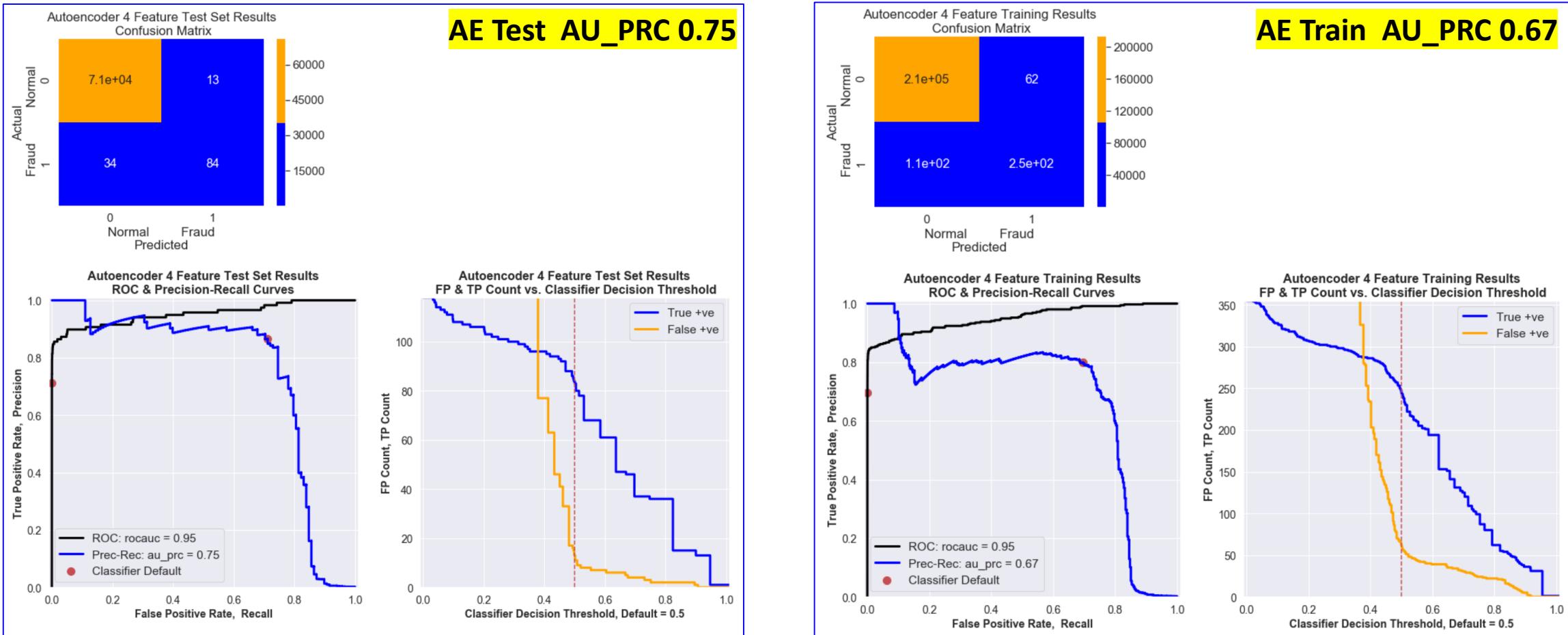


Compare the error Out vs In



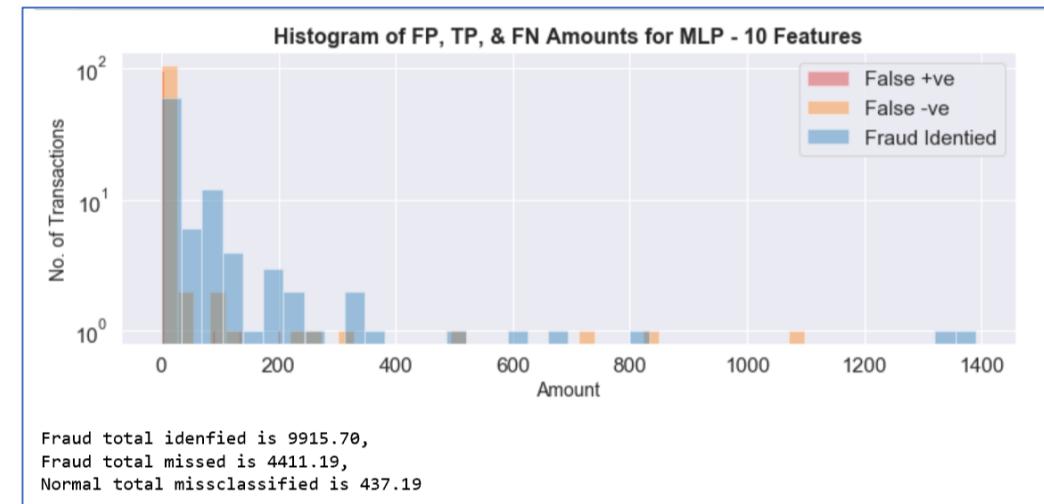
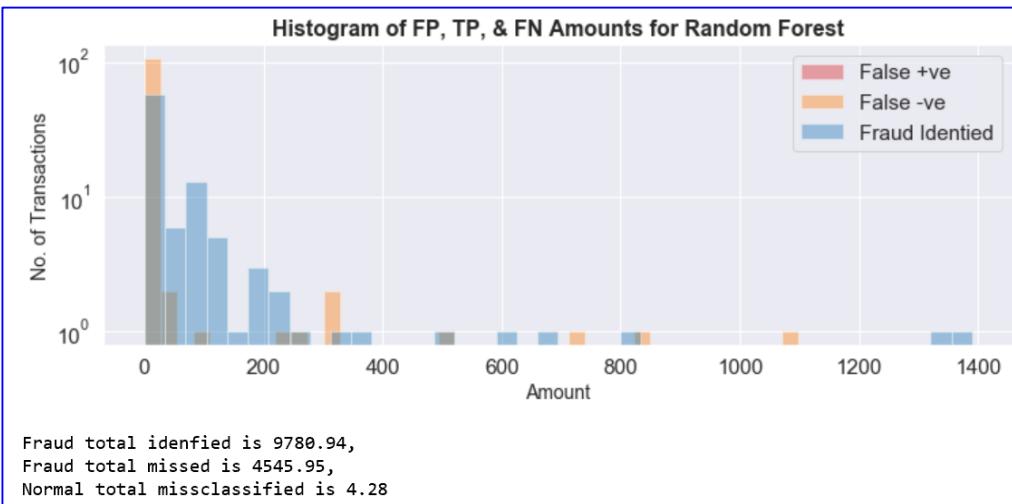
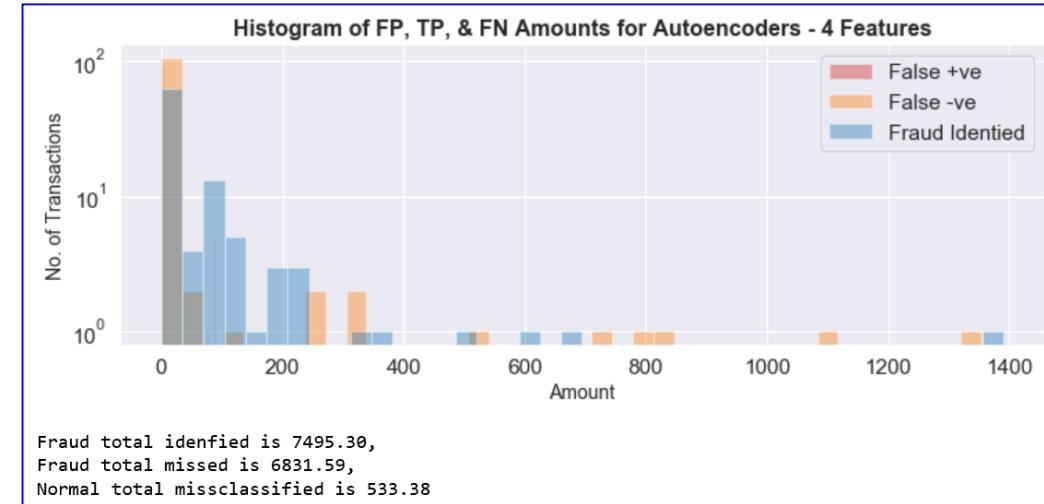
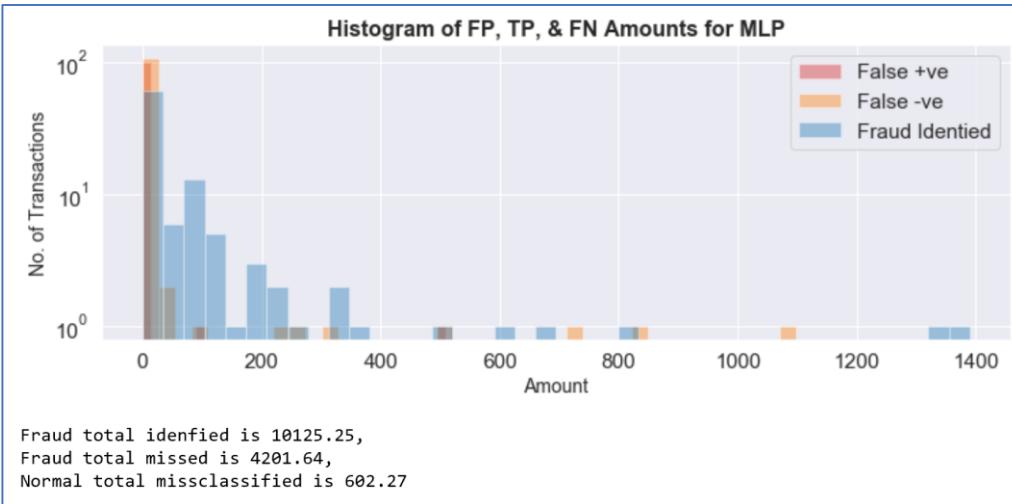
RoboGarden Bootcamp

Autoencoders



RoboGarden Bootcamp

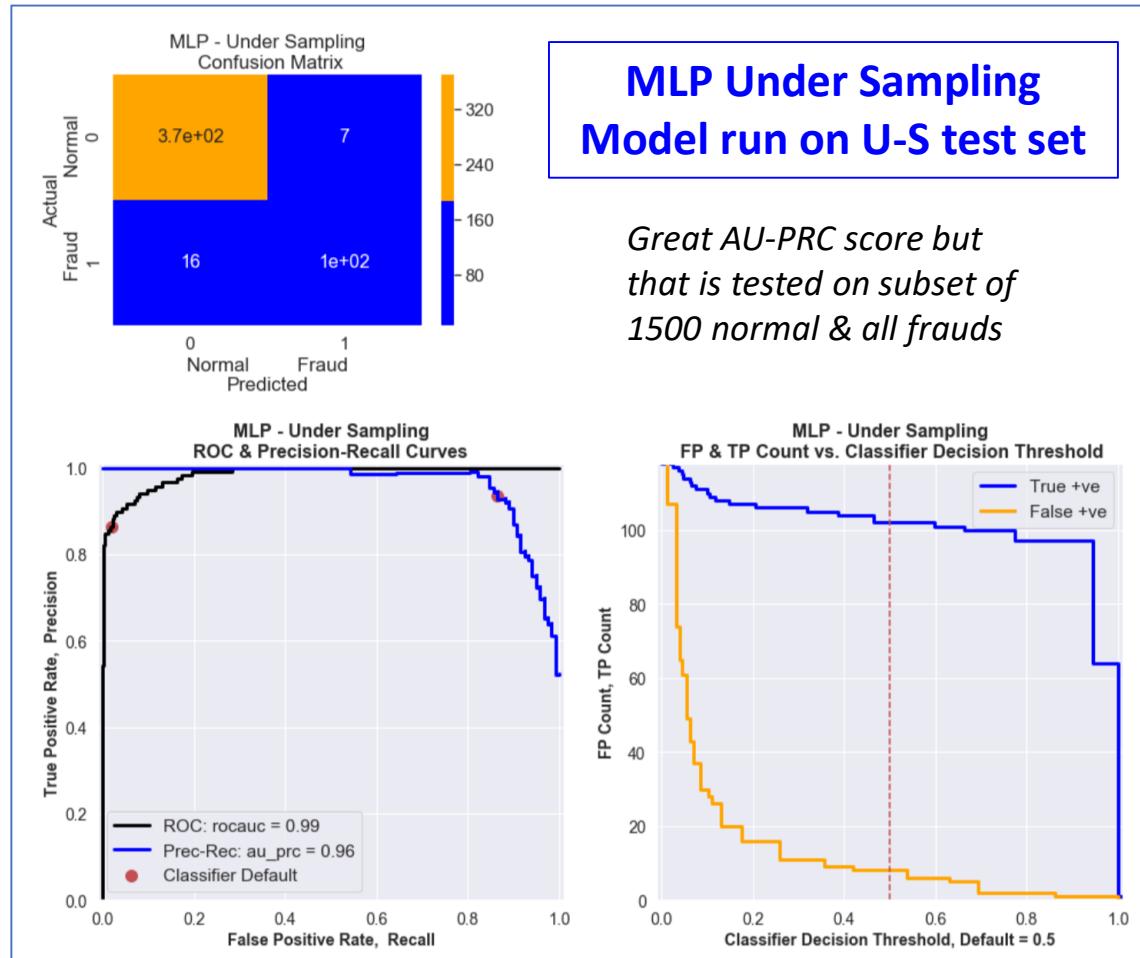
Fraud Value Identified



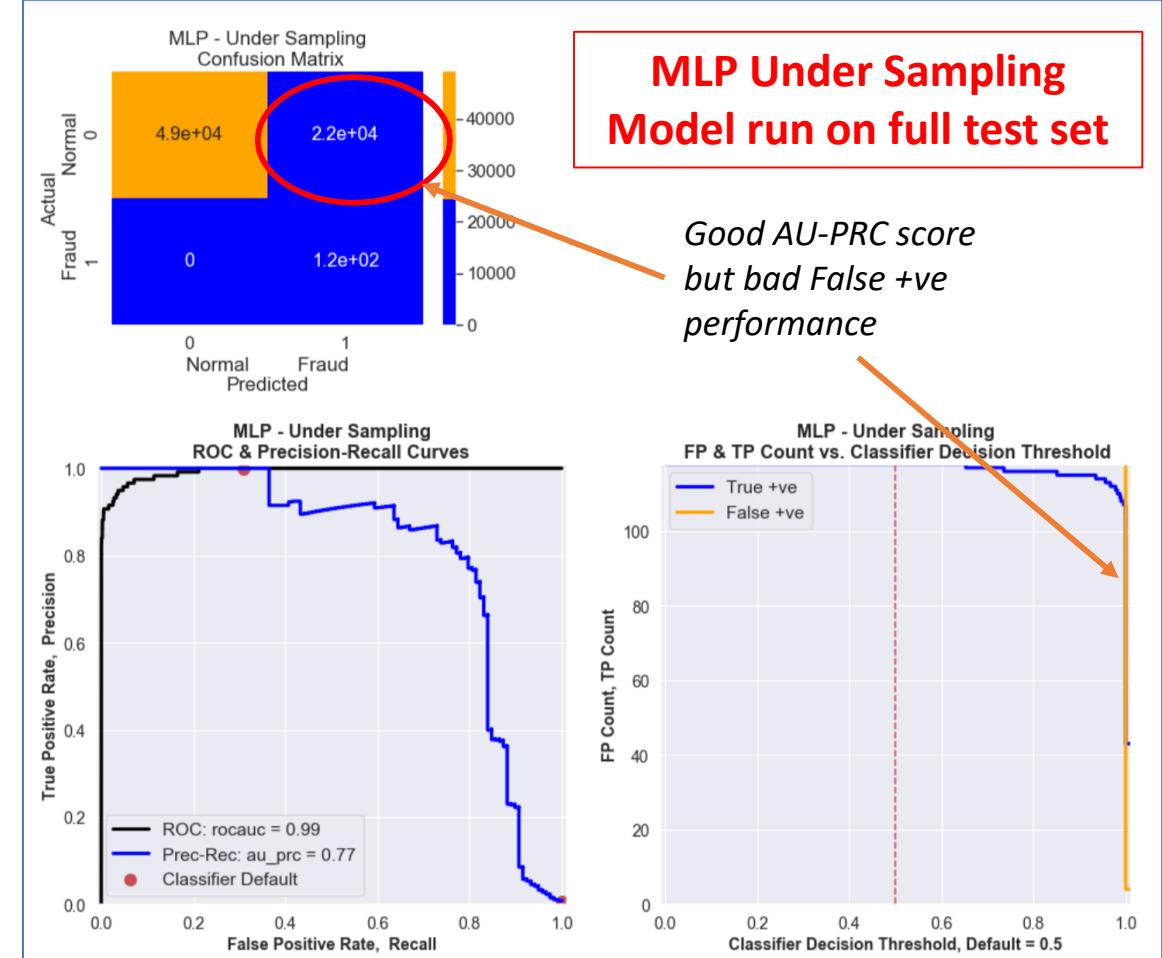
Back-up – Additional Charts

RoboGarden Bootcamp

MLP on Under Sample Dataset after Optimization



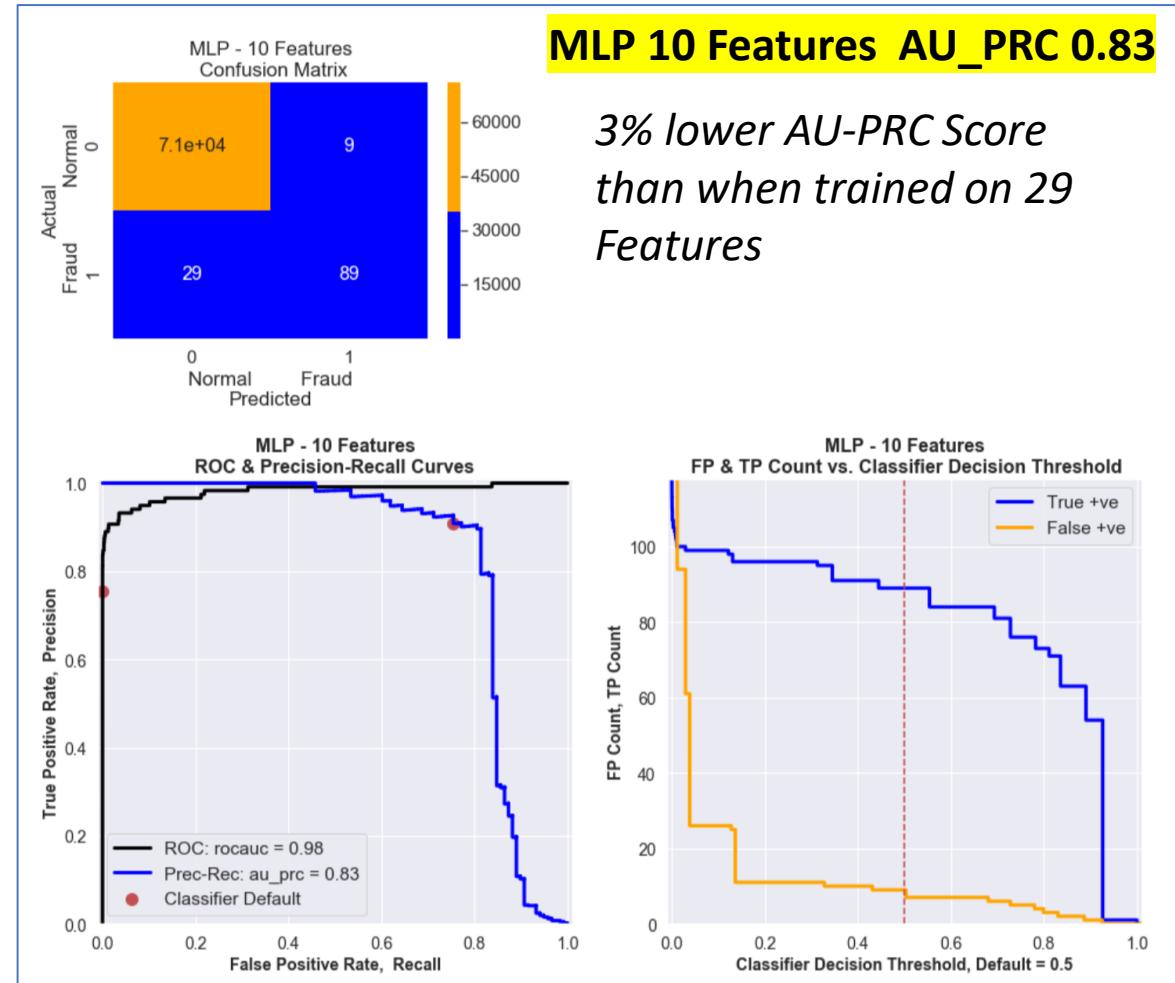
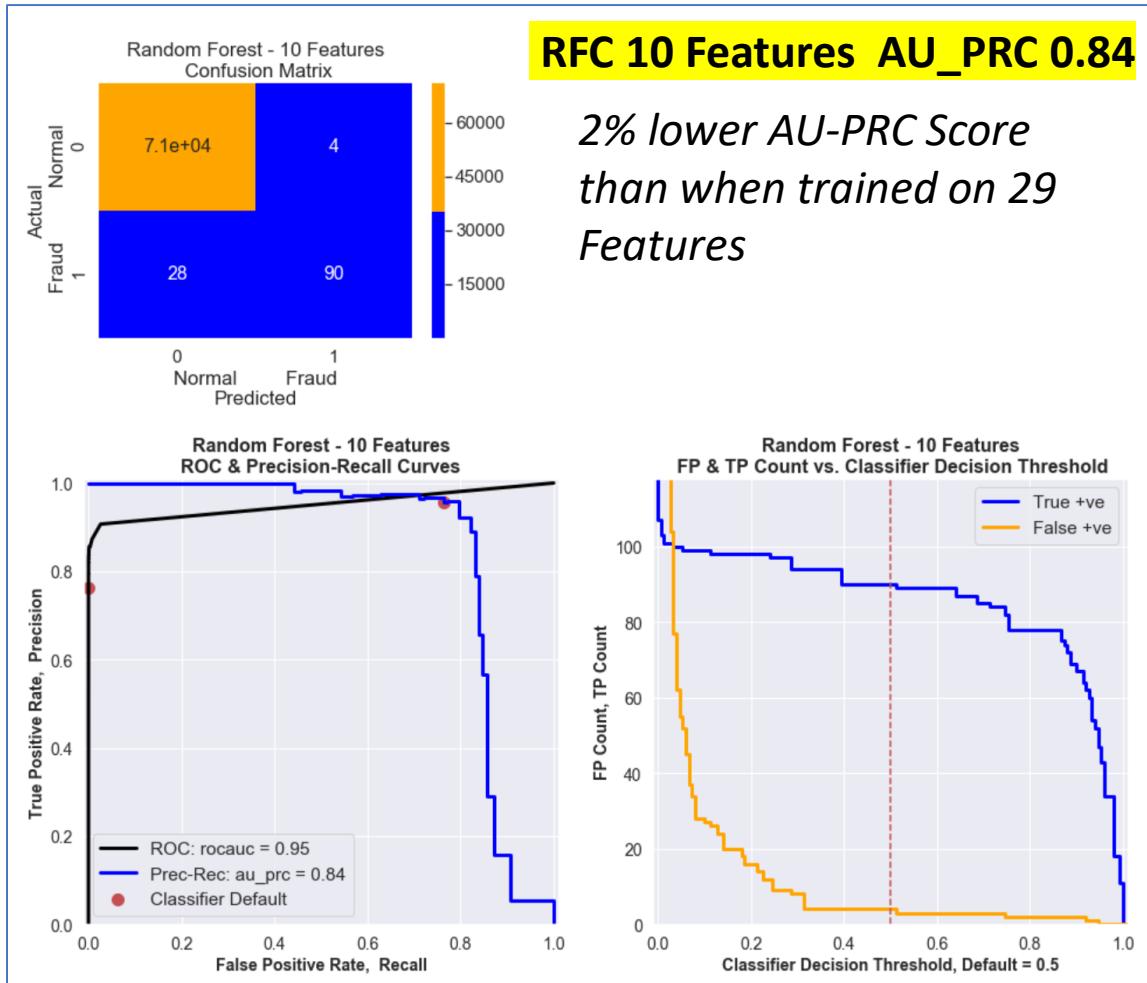
Model trained on U-S train set & tested on U-S test set →



Then Model tested on original test set (this is target test).

RoboGarden Bootcamp

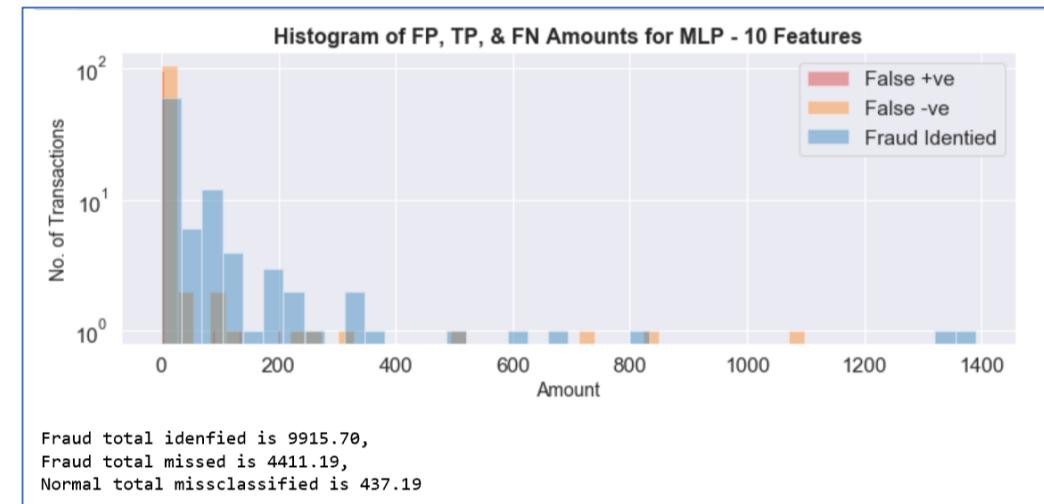
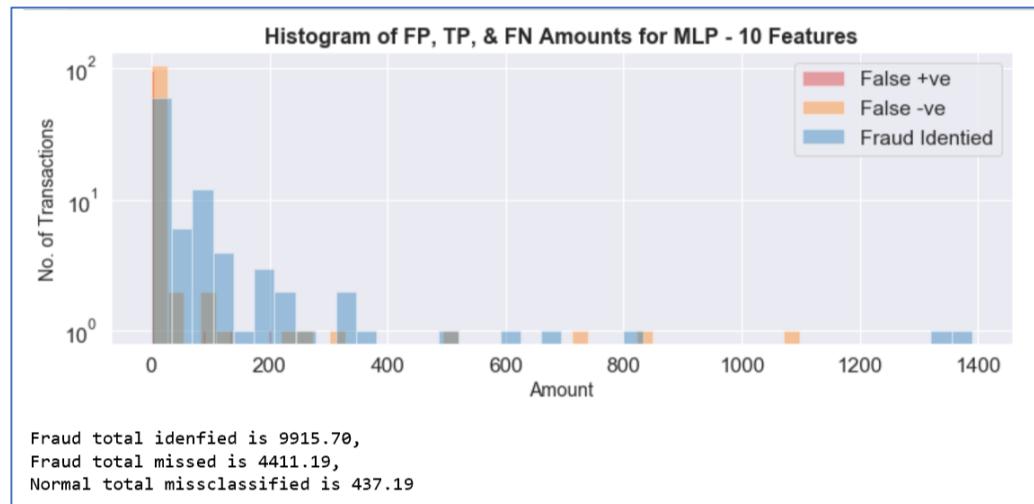
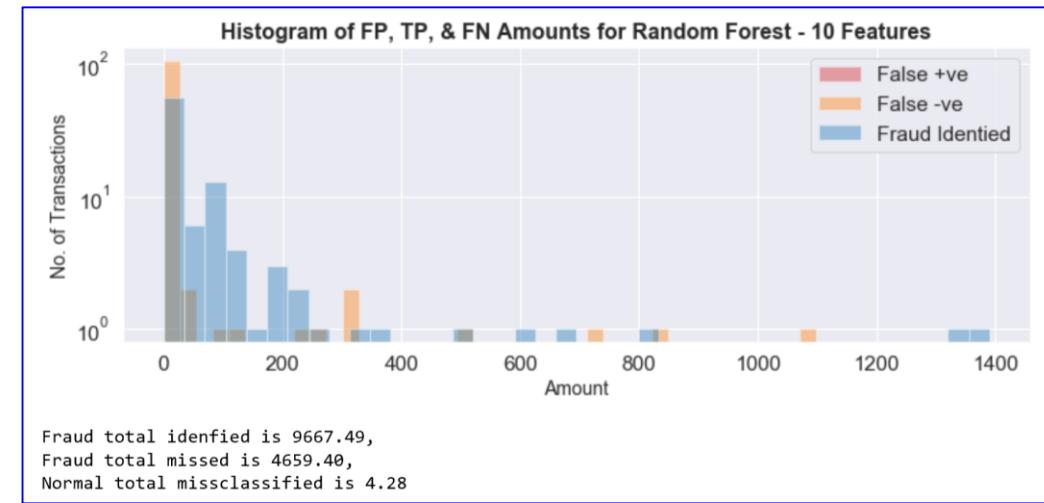
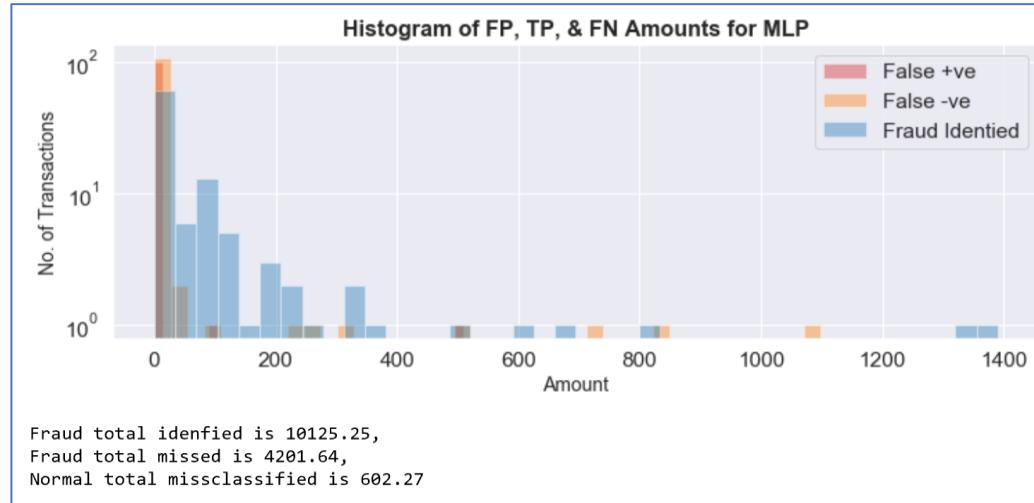
Random Forest & MLP Trained on 10 Features



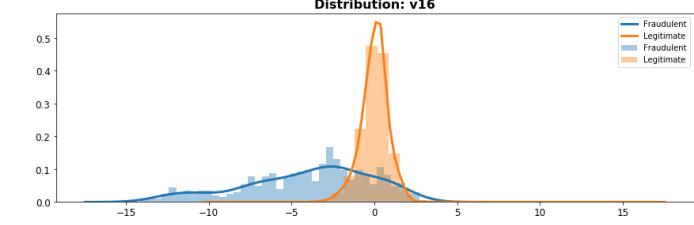
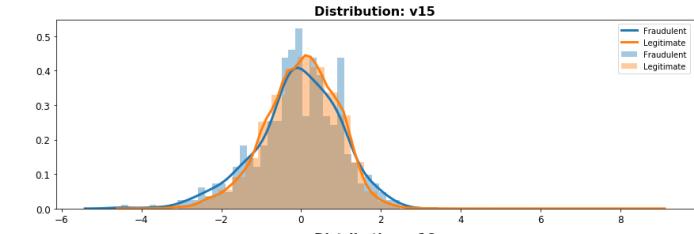
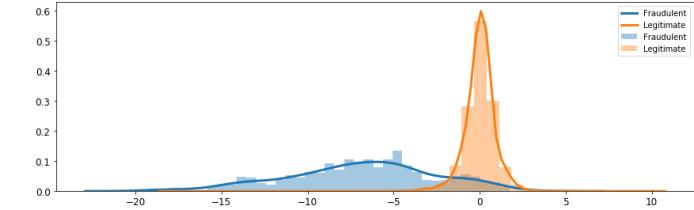
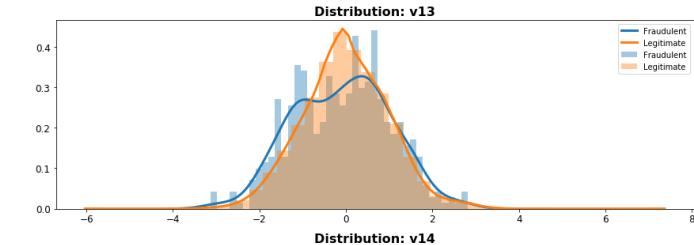
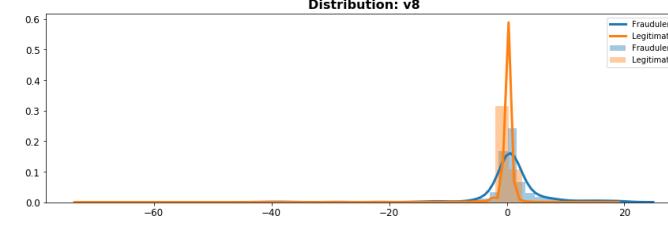
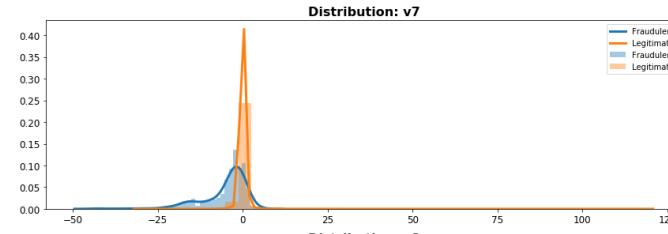
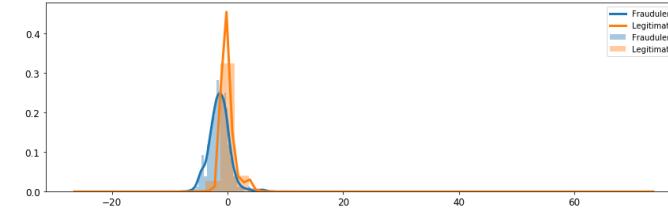
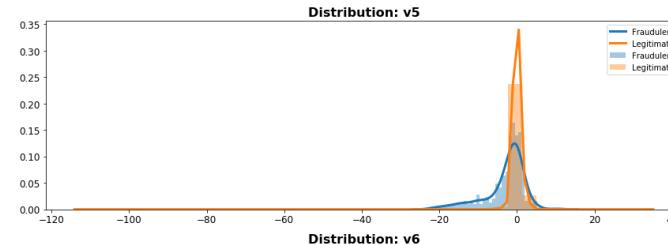
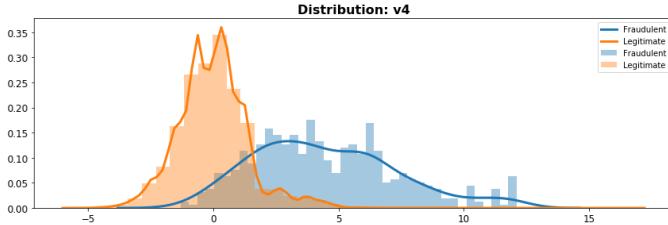
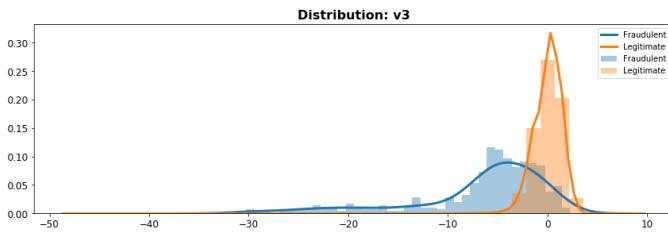
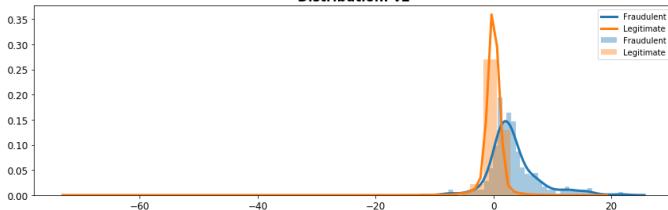
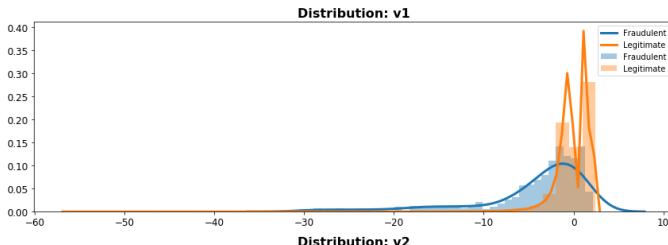
* 10 Features are Features v3, v4, v7, v9, v10, v11, v12, v14, v16, & v17

RoboGarden Bootcamp

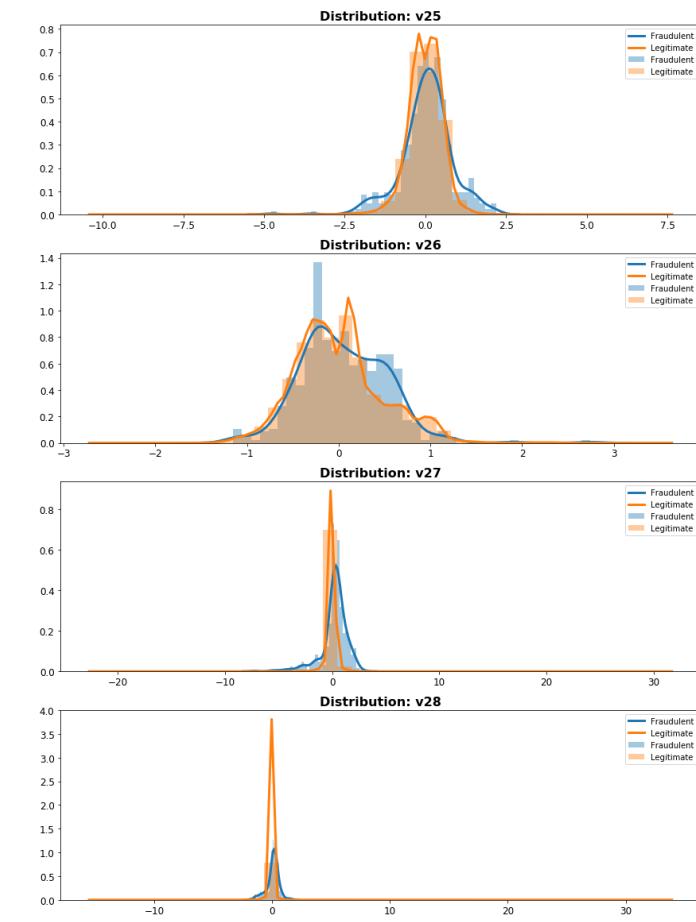
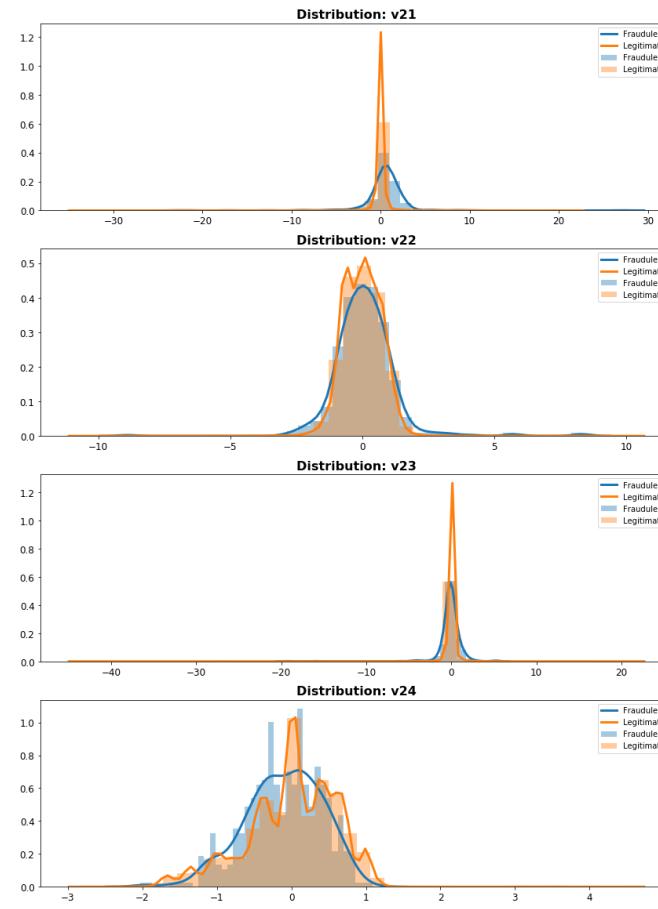
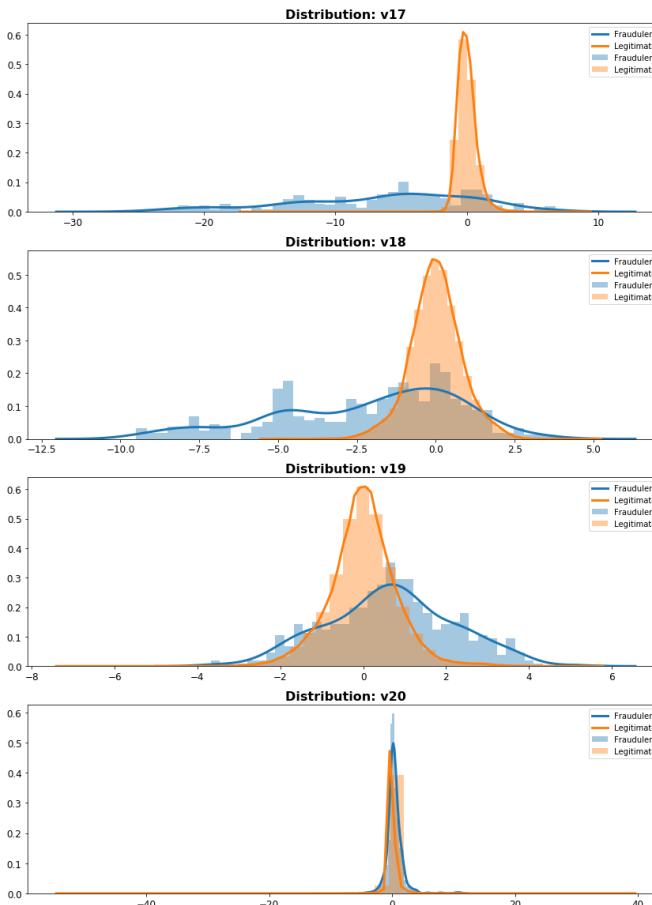
Fraud Value Identified



RoboGarden Bootcamp Visualization



RoboGarden Bootcamp Visualization



RoboGarden Bootcamp

Data Cleaning – Examples of Duplicates

	time	v1	v2	v3	v4	v5	v6	v7	v8	v9	...	v21	v22	v23	
13560	24050	1.216761	0.698963	-0.137686	2.527629	0.618533	-0.314776	0.486486	-0.303207	0.330209	...	-0.200891	-0.448693	-0.226691	
13561	24050	-0.841458	0.918286	1.504540	-0.521650	1.046457	-0.714847	0.855367	-0.089329	0.533366	...	0.029330	0.180174	-0.290965	
13562	24050	0.783460	-0.766538	1.331255	1.812482	-0.929991	1.318111	-0.979965	0.472786	2.619051	...	0.027365	0.410787	-0.294227	
13563	24050	0.783460	-0.766538	1.331255	1.812482	-0.929991	1.318111	-0.979965	0.472786	2.619051	...	0.027365	0.410787	-0.294227	
13564	24050	0.783460	-0.766538	1.331255	1.812482	-0.929991	1.318111	-0.979965	0.472786	2.619051	...	0.027365	0.410787	-0.294227	
13565	24050	0.783460	-0.766538	1.331255	1.812482	-0.929991	1.318111	-0.979965	0.472786	2.619051	...	0.027365	0.410787	-0.294227	
13566	24052	0.019196	1.060485	2.078401	1.703168	-0.332444	-0.210411	0.065809	-0.106047	0.782455	...	-0.095117	0.090776	0.000921	
13567	24052	1.124959	-0.283011	0.454480	0.910154	-0.616918	-0.344330	-0.297557	-0.024893	2.232117	...	-0.472669	-1.231524	-0.007479	
13568	24053	-0.254508	1.082387	-0.593772	-0.109728	3.051428	3.289752	0.430333	0.595560	0.318167	...	-0.055309	-0.054035	-0.247829	
13569	24053	-0.802007	0.283655	1.559359	-0.032721	0.952904	0.032269	0.046522	0.159000	0.352412	...	0.084018	0.397200	-0.028074	
	v6	v7	v8	v9	...	v21	v22	v23	v24	v25	v26	v27	v28	amount	class
	-0.928678	0.344191	0.262082	-1.143424	...	0.179758	0.290998	-0.346720	0.435099	0.564471	-0.200885	-0.038164	-0.057133	5.00	False
	5.760059	-18.750889	-37.353443	-0.391540	...	27.202839	-8.887017	5.303607	-0.639435	0.263203	-0.108877	1.269566	0.939407	1.00	True
	5.760059	-18.750889	-37.353443	-0.391540	...	27.202839	-8.887017	5.303607	-0.639435	0.263203	-0.108877	1.269566	0.939407	1.00	True
	5.760059	-18.750889	-37.353443	-0.391540	...	27.202839	-8.887017	5.303607	-0.639435	0.263203	-0.108877	1.269566	0.939407	1.00	True
	5.760059	-18.750889	-37.353443	-0.391540	...	27.202839	-8.887017	5.303607	-0.639435	0.263203	-0.108877	1.269566	0.939407	1.00	True
	5.760059	-18.750889	-37.353443	-0.391540	...	27.202839	-8.887017	5.303607	-0.639435	0.263203	-0.108877	1.269566	0.939407	1.00	True
	5.760059	-18.750889	-37.353443	-0.391540	...	27.202839	-8.887017	5.303607	-0.639435	0.263203	-0.108877	1.269566	0.939407	1.00	True
	3.183007	-0.499724	0.803819	-0.074543	...	-0.340246	-1.150620	0.091744	0.946327	0.368241	0.110198	-0.022594	0.024337	5.99	False
	0.960209	0.338479	0.151942	-0.182027	...	0.191189	0.865940	-0.272396	-1.333896	-0.154556	-0.161314	0.215079	-0.099068	22.72	False
	3.098005	-0.013022	0.649896	-0.465807	...	-0.043931	-0.258826	-0.189341	1.009328	0.998140	-0.266138	-0.009990	0.006402	14.90	False
	-0.822682	0.633459	-0.335497	-1.189777	...	-0.002347	0.007407	-0.119860	0.403321	0.758846	0.065612	-0.034713	0.012673	7.53	False