RoboGarden Bootcamp Capstone Project

Credit Card Fraud Detection

July 2019

RoboGarden Bootcamp
Credit Card Fraud Project

Description:

- 284,807 credit card transactions made by European cardholders in September 2013

Features:

- Time: seconds since first transaction
- V1 – V28: Anonymous data – Confidentiality
- Amount: Transaction value (Unspecified currency)
- Class (T/F): fraudulent / genuine

License: Public domain (CC0)

Available: *Data World :* https://data.world/raghu543/credit-card-fraud-data (also available on Kaggle)

File: creditcard.csv

Reference use of dataset: Dal Pozzolo, Olivier Caelen, Reid A. Johnson and Gianluca Bontempi. Calibrating Probability with Undersampling for Unbalanced Classification. In Symposium on Computational Intelligence and Data Mining (CIDM), IEEE, 2015

RoboGarden Bootcamp
Credit Card Fraud Project Dataset

- *normal amount total* ***25,043,410***
- *fraud amount total* ***58,591*** ***(0.25%)***
- *# transactions over 2 days* ***284,807***
- *# fraud transactions* ***492*** ***(0.17%)***
- *# non-fraud duplicates* ***1062*** ***(0.4 %)***
- *# fraud duplicates* ***19*** ***(4.0 %)***
- *# zero amount normal transactions* ***1798*** ***(0.6%)****
- *# zero amount fraud transactions* ***27*** ***(5.5%)****

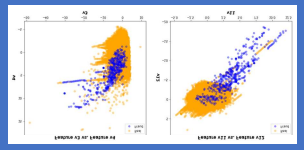
* Retained zero amount transactions. Insufficient information to remove them.

RoboGarden Bootcamp
Work Process

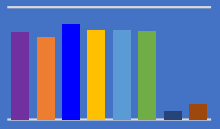
Clean Data

Remove Duplicates

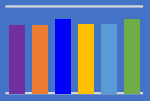
Visualize Data



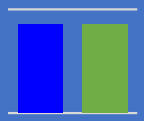
Screen 8 Classifiers



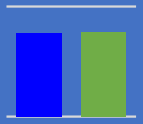
Optimize 6 Classifiers
All Features



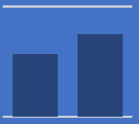
Optimize RF & MLP on
10 Features
Feature Importance



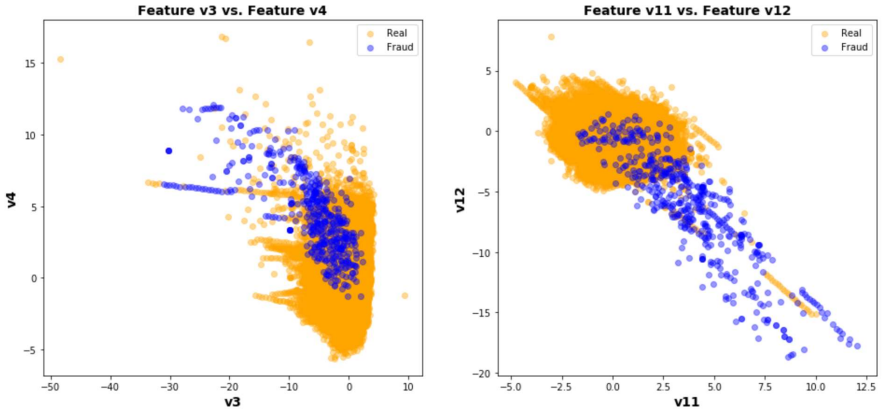
Optimize RF & MLP on
Undersampled
Dataset All Features



Run Autoencoders on
10 Features
4 Features



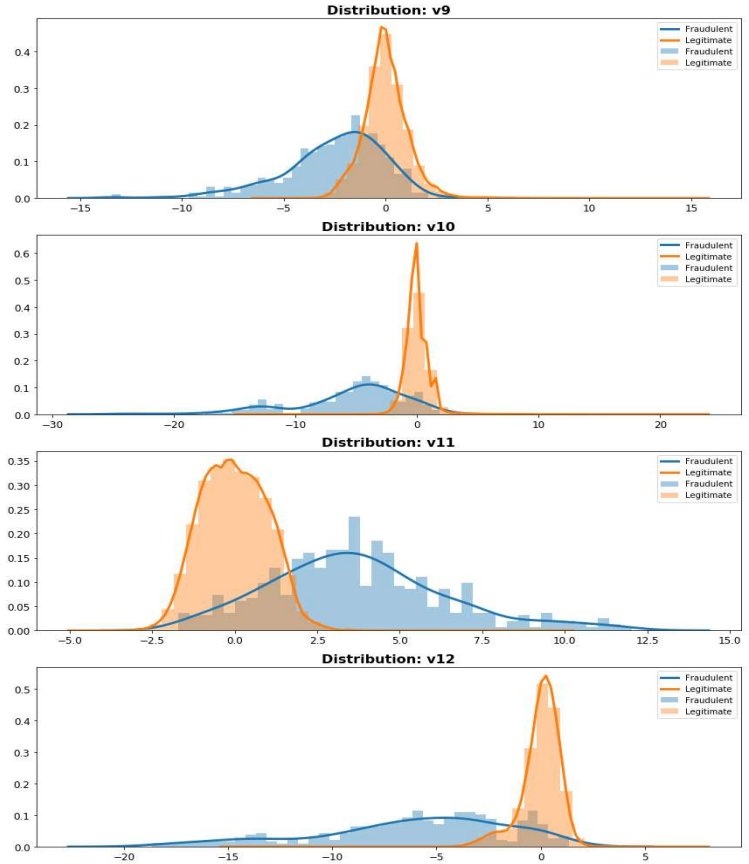
RoboGarden Bootcamp
Visualization Examples



- **2D Scatter plots show some overlap & separation.**

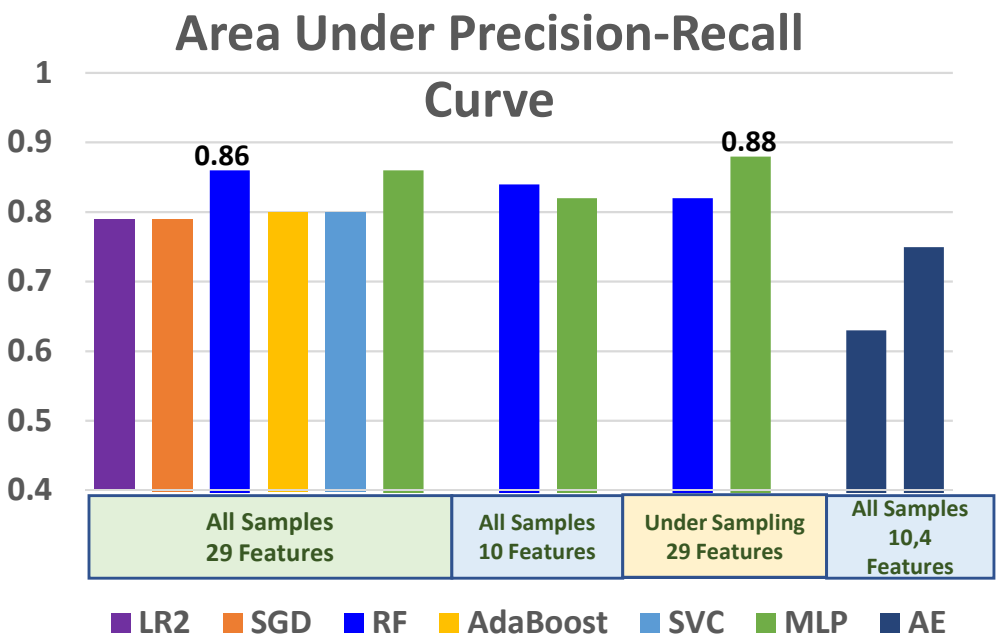
Showing 4 of 28 Features: Fraud vs. Normal Histograms

- **Several have distinctive range differences.**
- **Some distributions are aligned, (not shown).**



RoboGarden Bootcamp
Modelling Results

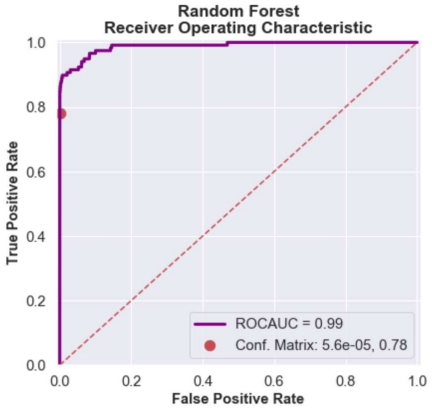
Model	Application	Features	Scores AU-ROC	AU-PRC*	F-1	TP	FP	FN
LR	All Samples	29	0.98	0.79	0.69	65	5	53
SGD	All Samples	29	0.98	0.79	0.79	83	7	35
RF ★	All Samples	29	0.96	0.86	0.86	92	4	26
SVC	All Samples	29	0.95	0.80	0.78	77	3	41
AdaBoost	All Samples	29	0.97	0.80	0.80	84	8	34
MLP	All Samples	29	0.99	0.86	0.86	92	5	26
RF	All Samples	10	0.96	0.84	0.85	90	4	28
MLP	All Samples	10	0.98	0.82	0.83	88	6	30
RF **	Undersampling	29	0.99	0.82	0.87	97	8	21
MLP ** ★	Undersampling	29	0.98	0.88	0.86	96	10	22
AE	All Samples	10	0.97	0.57	0.58	66	45	52
AE	All Samples	4	0.96	0.75	0.78	83	11	35

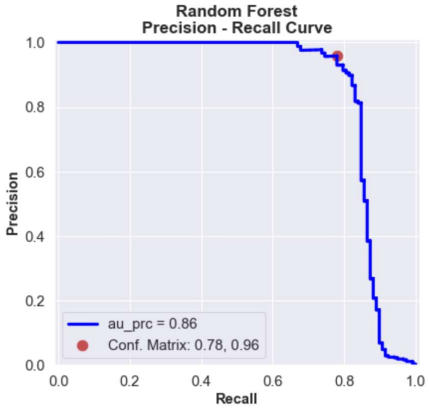


★ *Model results shown on following slides*

** AU-PRC: Area under the Precision-Recall Curve is the recommended measure of accuracy stated by the dataset provider due to the imbalance in the dataset.*
***Under Sampling applies calibration to the sample probabilities.*

RoboGarden Bootcamp
Random Forest Results



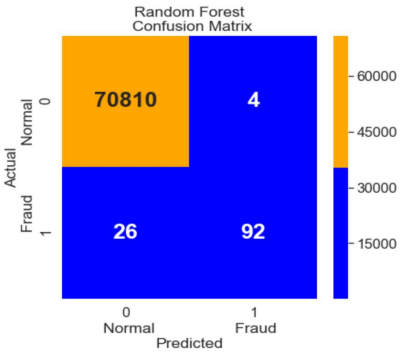


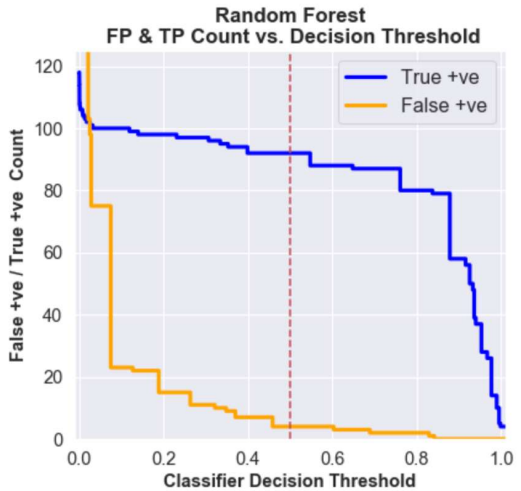
ROC and Precision-Recall Curves:

- Area under Precision-Recall of 0.86.
- Area under ROC of 0.99.
- Confusion Matrix corresponds to markers on ROC & PRC and decision threshold line on the FP & TP Count plot.

Random Forest trained on the full dataset *:

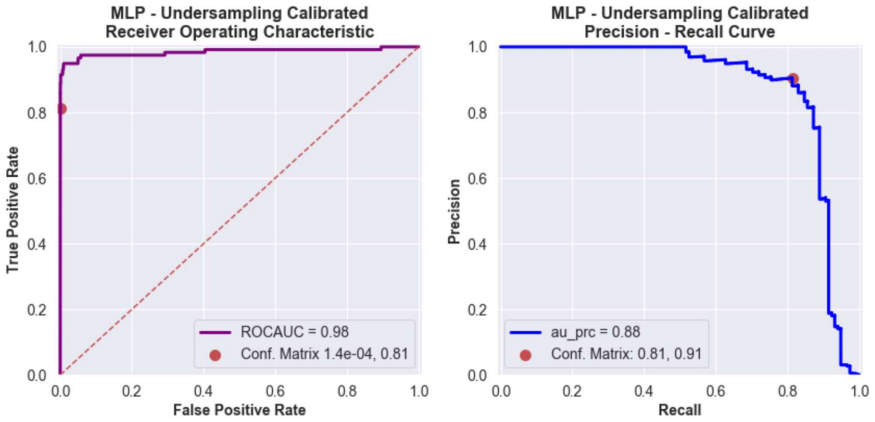
- Found 78% of frauds at 0.5 decision threshold.
- Has a low false positive rate.
- 67% of frauds have a classification probability over 0.8.





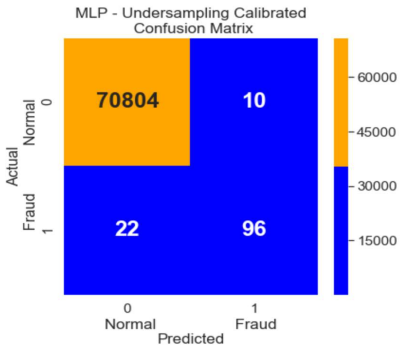
**Trained on 29 Features: Amount + Features v1 to v28 (time was dropped)*

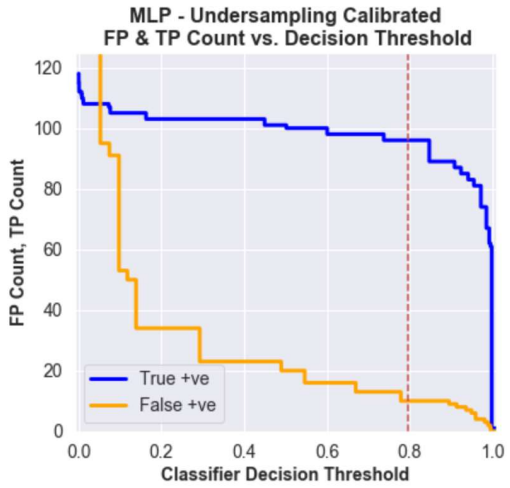
RoboGarden Bootcamp
MLP – Undersampling Results



ROC and Precision-Recall Curves:

- Area under Precision-Recall of 0.88.
- Area under ROC of 0.98.
- Confusion Matrix corresponds to markers on ROC & PRC and decision threshold line on the FP & TP Count plot.





MLP trained on fewer samples *:

- Tested on the full 70,000+ test sample set.
- Found 81% of frauds using a 0.8 decision threshold after undersampling calibration.
- Has a low overall false positive rate, but higher than the Random Forest model.

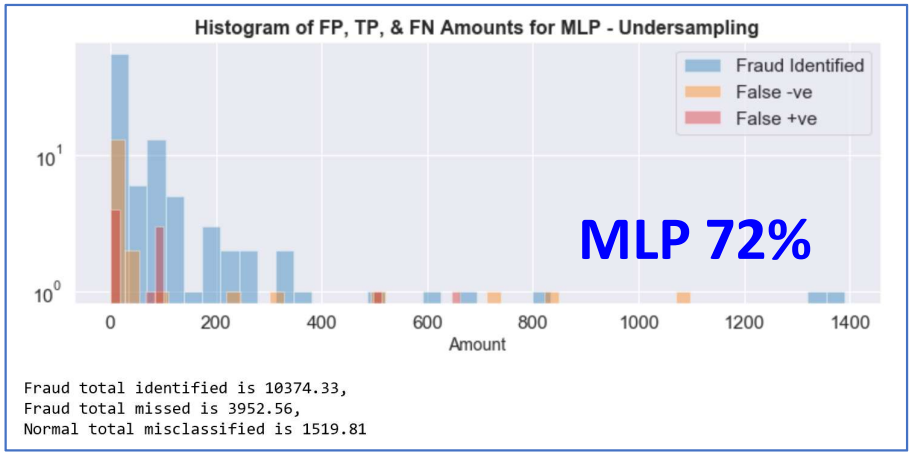
**Trained on reduced set containing 10% of non-fraud transactions and 29 Features (Amount + Features v1 to v28)*

RoboGarden Bootcamp

Fraud Value Identified in Test Set (25% of Dataset)

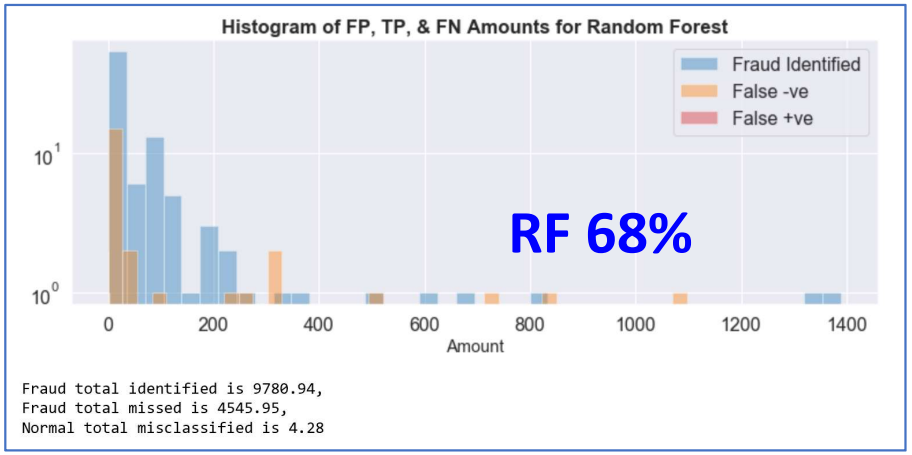
96 Frauds & 72% of value*

10 FP's – 15% of the Fraud value



92 Frauds & 68% of value*

4 FP's – .03% of the Fraud value



- Random Forest identified a slightly lower fraud amount, but misclassified a substantially lower amount than the MLP Classifier.

** Amounts will vary with new data.*

RoboGarden Bootcamp
Conclusions / Future Work

Conclusions:

- Despite the extreme unbalanced nature of the dataset, Random Forest classified 78% of the fraudulent transactions with few false positives (4% of frauds identified).
- The undersampling technique improved the area under the Precision-Recall Curve score and identified 81% of the fraudulent transactions. This MLP model had a higher false positive rate and misclassified a higher value amount of legitimate transactions.
- Performance degraded when features were dropped except for the Autoencoder model which improved with fewer more distinct features.

Future Work:

- Include more model parameters in a broader optimization search.
- Use time feature by setting it to time of day vs. time from first transaction.
- Investigate a hybrid classifier by combining multiple classifiers.