6501 Progress Report Group 4

Improving Credit Card Fraud Detection using Generative Adversarial Networks

Group 4 Team Member: Hao Ning, Jun Ying

Working Schedule

Time	Milestone
09/21/2020	Exploratory Data Analysis (EDA): Jun Base Model: Hao
09/28/2020	Original data + GAN
10/05/2020	Network & Framework Development WGAN: Hao BEGAN: Jun
10/12/2020	WGAN & BEGAN Evaluation & Analysis
10/19/2020	Network & Framework Development BAGAN: Hao SNGAN: Jun
10/26/2020	Preliminary Presentation
11/02/2020	BAGAN, SNGAN Evaluation & Analysis
11/09/2020 & 11/16/2020	Summary of Results
11/23/2020	Manuscript
11/30/2020 & 12/07/2020	Mock Presentation & Presentation and Journal Submission

6501 Progress Report Group 4

09/21/2020

EDA

```
Index(['Time', 'V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10',
       'V11', 'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20',
       'V21', 'V22', 'V23', 'V24', 'V25', 'V26', 'V27', 'V28', 'Amount',
       'Class'],
```

About the dataset, there are 30 features and 1 class (normal:0, fraud:1)

	Time	V1	V2	V3	...	V27	V28	Amount	Class
0	0.0	-1.359807	-0.072781	2.536347	...	0.133558	-0.021053	149.62	0
1	0.0	1.191857	0.266151	0.166480	...	-0.008983	0.014724	2.69	0
2	1.0	-1.358354	-1.340163	1.773209	...	-0.055353	-0.059752	378.66	0
3	1.0	-0.966272	-0.185226	1.792993	...	0.062723	0.061458	123.50	0
4	2.0	-1.158233	0.877737	1.548718	...	0.219422	0.215153	69.99	0

There is no null value in the dataset.

Total null values in the dataset
0

As we know, the dataset is extremely imbalanced(0.173%).

The amounts of normal transactions (class 0) & fraud transactions (class 1)
0 284315
1 492

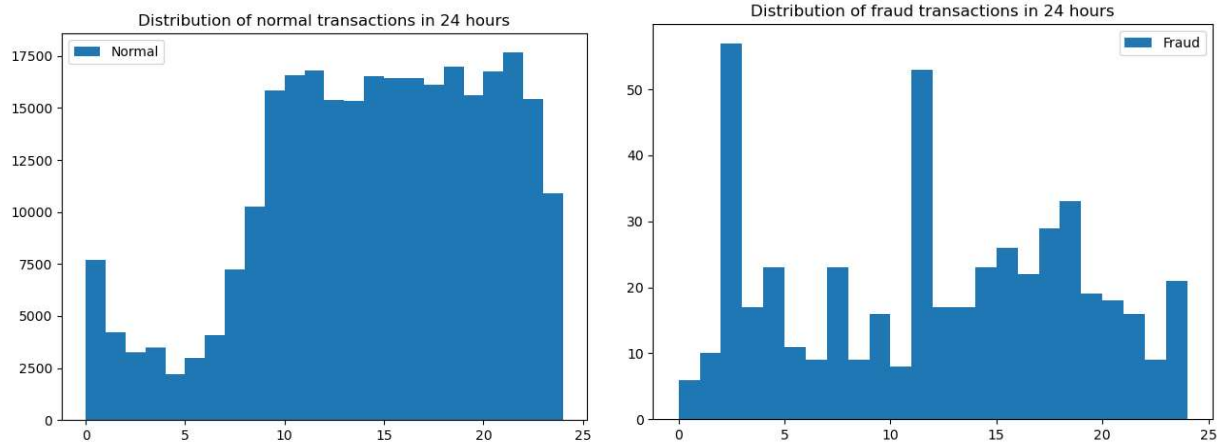
We have observed that there are some transactions which are 0.

	Time	V1	...	Amount	Class
count	284807.000000	2.848070e+05	...	284807.000000	284807.000000
mean	14.537951	3.919560e-15	...	88.349619	0.001727
std	5.847061	1.958696e+00	...	250.120109	0.041527
min	0.000000	-5.640751e+01	...	0.000000	0.000000
25%	10.598194	-9.203734e-01	...	5.600000	0.000000
50%	15.010833	1.810880e-02	...	22.000000	0.000000
75%	19.329722	1.315642e+00	...	77.165000	0.000000
max	23.999444	2.454930e+00	...	25691.160000	1.000000

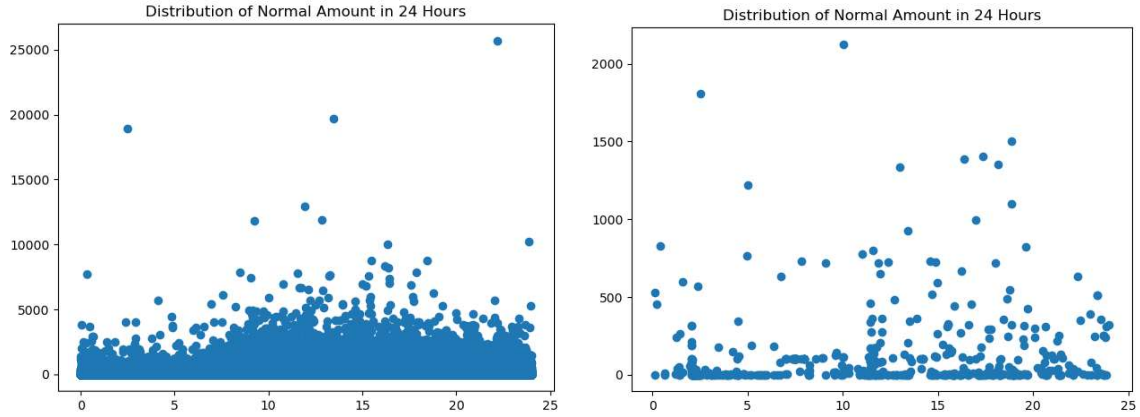
The total number of 0 amount: 1825 (1.479% fraud)

```
The null amounts of normal transactions (class 0) & fraud transactions (class 1)
0   1798
1     27
Name: Class, dtype: int64
```

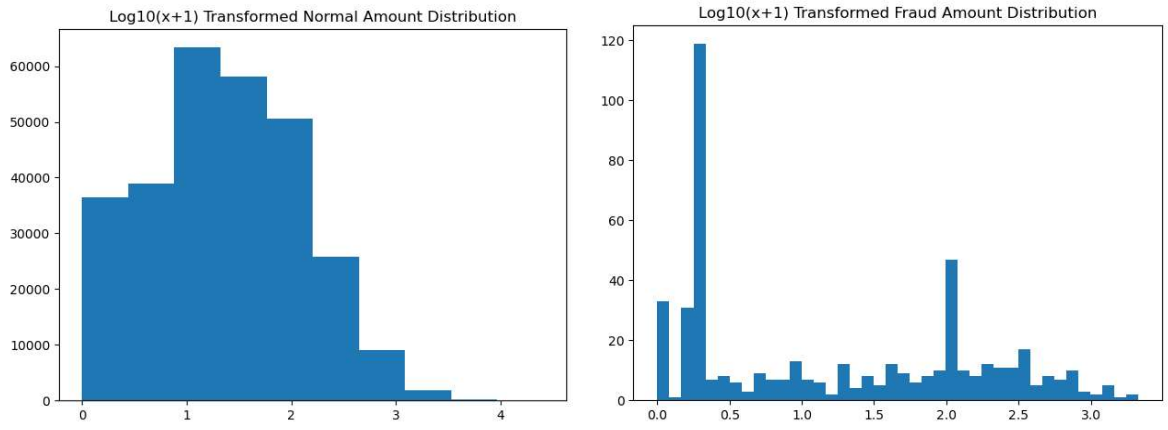
6501 Progress Report Group 4



From the histogram, we can observe that normal transactions generally occur from 9 am to 0 am. However, the fraud transactions occur particularly frequently at 2 am and 12 pm.

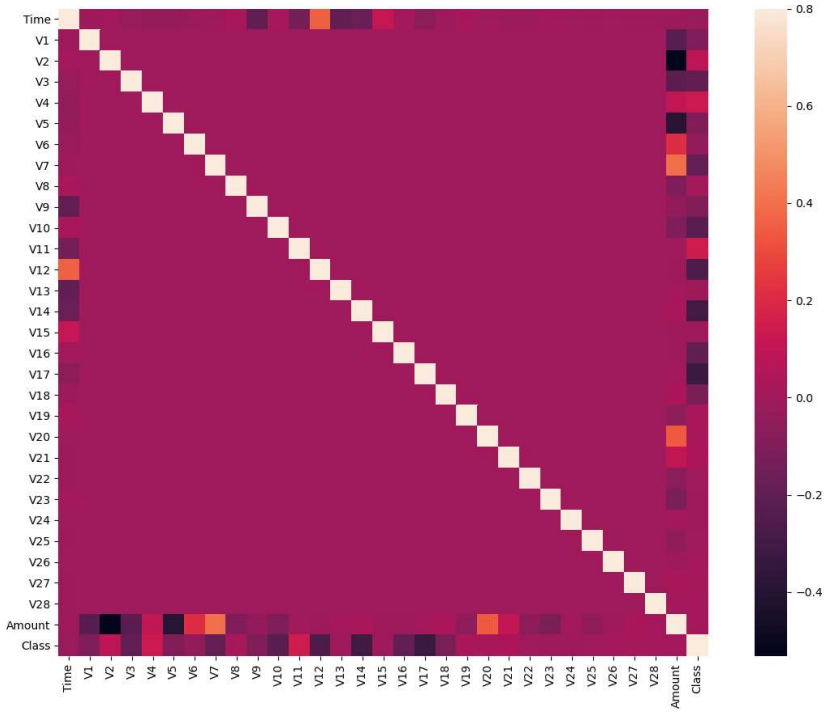


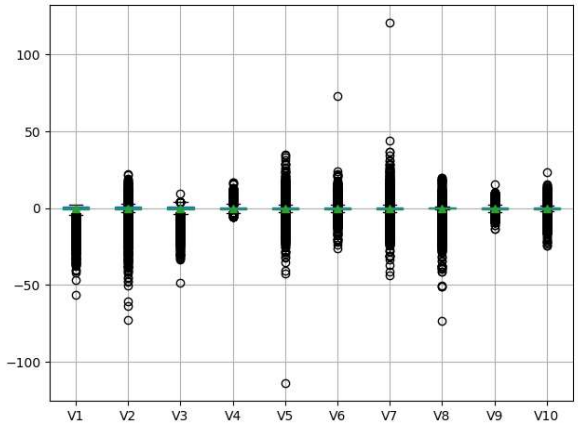
We can find from this scatter plot that the amount of super large transactions is very small. In comparison, the largest amount of nurmal transactions is over €25,000. However, the largest amount of fraud transactions is only €2,000.

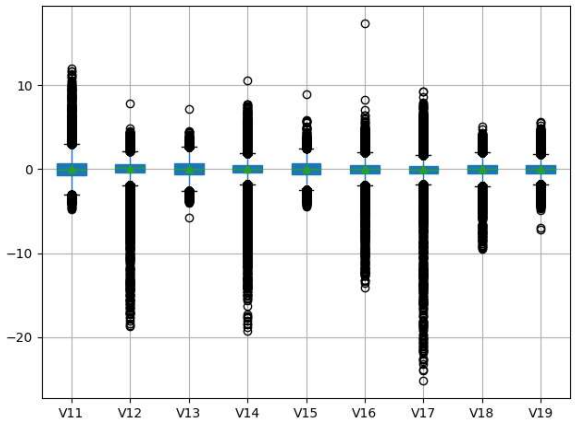


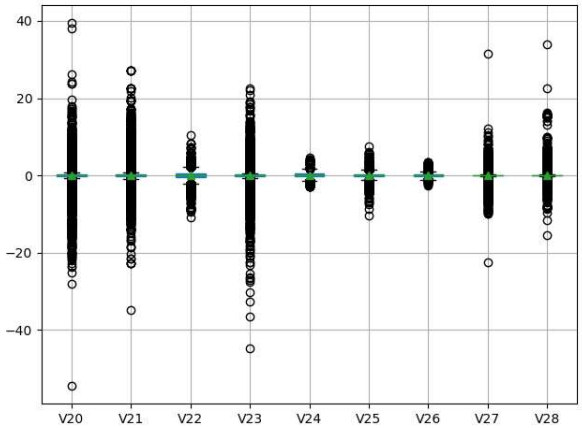
Normal amount was from ten to hundred. Fraud Amount distributed in less than €1.

6501 Progress Report Group 4









6501 Progress Report Group 4

Base Model:

1. Train Test Split & Stratified: 80% 227845 (**394 fraud**), 20% 56962 (**98 fraud**)
2. Random Under Sampling (RUS) and Random Over Sampling (ROS)
3. GridsearchCV for XGBoostClassifier
4. Predict with best_params
5. Test result comparison

RUS 1 394 0 394	ROS 1 227451 0 227451
Accuracy: 0.9691548751799445 Precision: 0.049429657794676805 Recall: 0.9285714285714286 F1 score: 0.09386281588447654 ROC AUC score: 0.9488981228394566	Accuracy: 0.9995084442259752 Precision: 0.8571428571428571 Recall: 0.8571428571428571 F1 score: 0.8571428571428571 ROC AUC score: 0.9284483278398585