

Credit Card Fraud Detection



Presented by
Artem Ramus

Background and Introduction

The goal of this project is to predict if a transaction is fraudulent from 28 anonymous features and an amount.

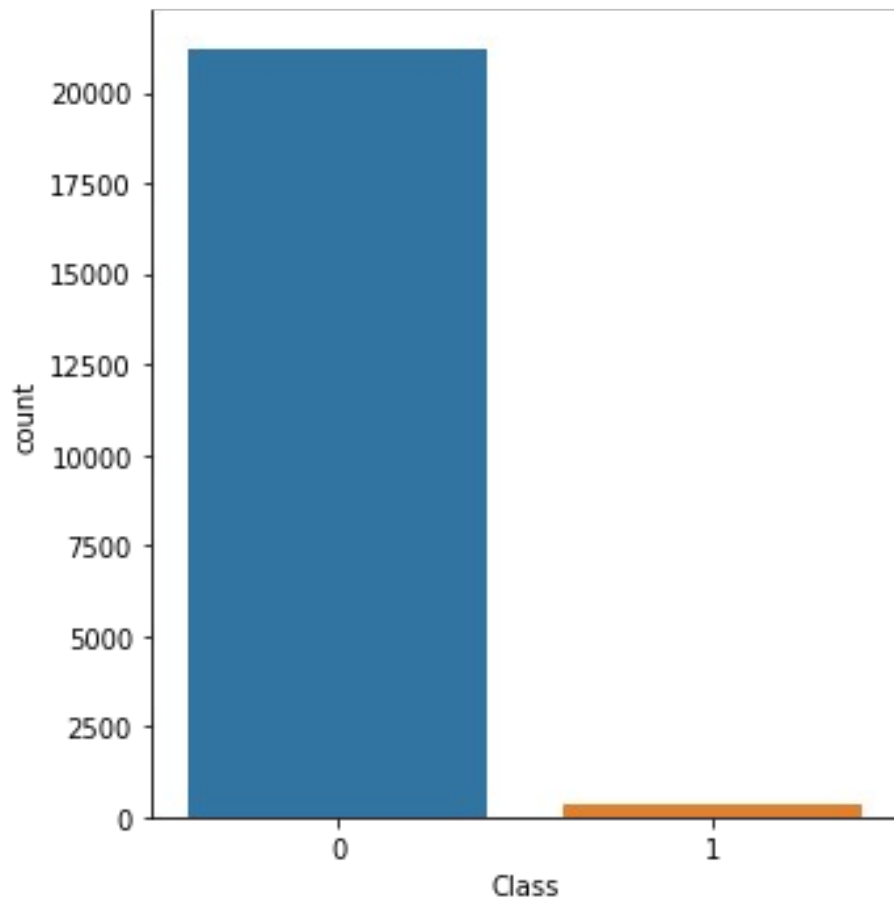
This data set is about credit card fraud detection. It was published by Rashida Sucky. The feature names aren't available. It has a binary column that indicates if a transaction is fraudulent or not. This data set can be used for classification models.

Link to the data set: https://github.com/rashida048/Datasets/blob/master/fraud_data.csv

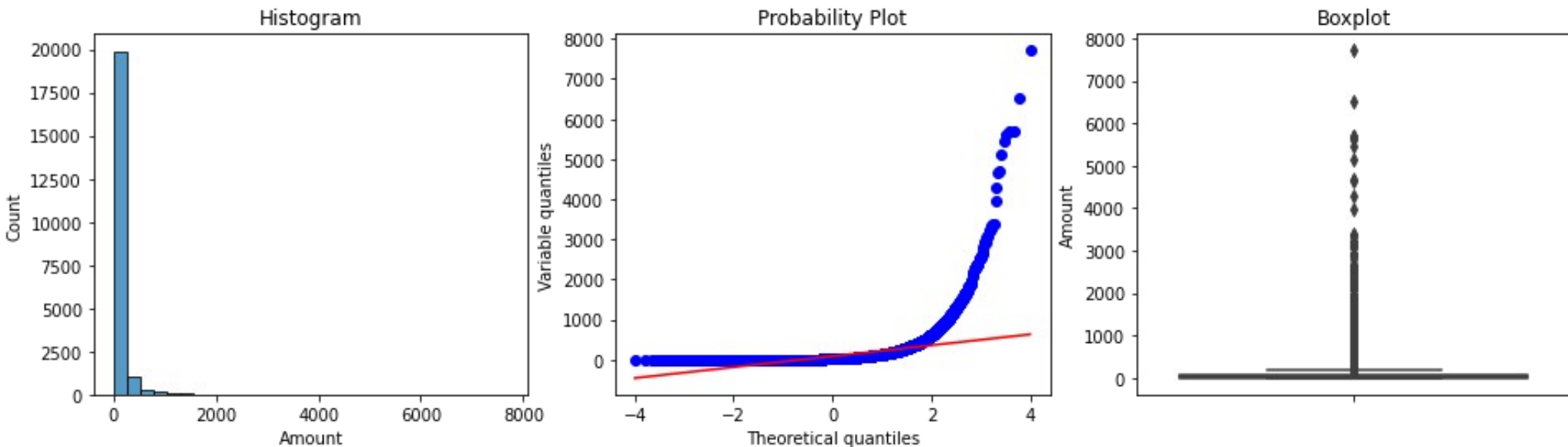
Exploratory Data Analysis - target

The target is imbalanced with under-represented class 1, fraudulent.

class	0	1
count	21204	346



Exploratory Data Analysis



Amount is right skewed with high kurtosis and possible outliers at high values. Most of the variable are skewed or/and have high kurtosis. Possible outliers do not seem acute.

Exploratory Data Analysis

All the 28 variables are numerical. Each variable's mean differs significantly for classes 0 and 1. This means that populations of classes 0 and 1 are different.

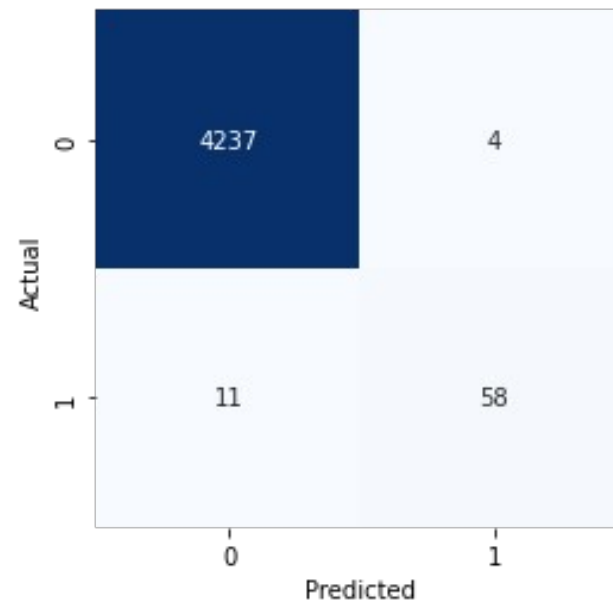
Variable	Mean trustworthy	Mean fraudulent
V1	0.0315	-4.2053
V2	-0.0094	3.3247
V3	0.0258	-6.5457
...
...
...
V26	0.0016	0.041
V27	-0.0017	0.1993
V28	0.0012	0.1177
Amount	86.7	116.4

Feature Engineering and Selection

- Using all the features for modeling
- Original features' distribution
- Modeling with Random Forest and XGBoost

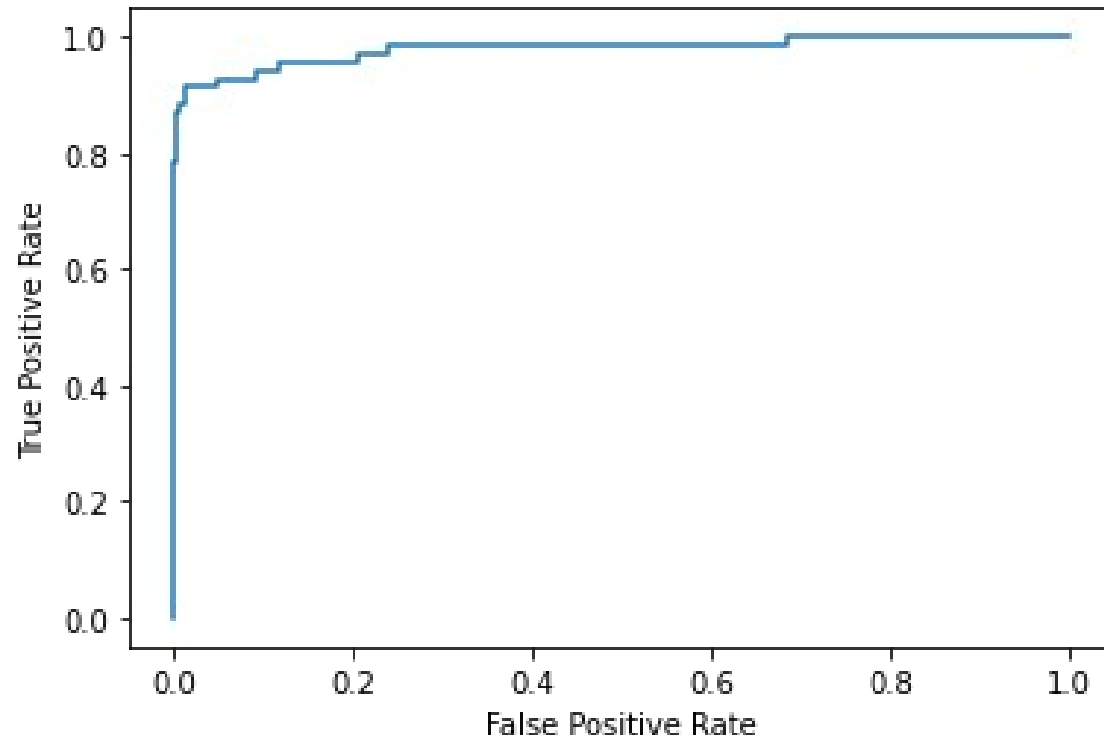
Performance

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4241
1	0.94	0.84	0.89	69
accuracy			1.00	4310
macro avg	0.97	0.92	0.94	4310
weighted avg	1.00	1.00	1.00	4310



Probability cutoff 0.5 results in precision 0.94 and recall 0.84.

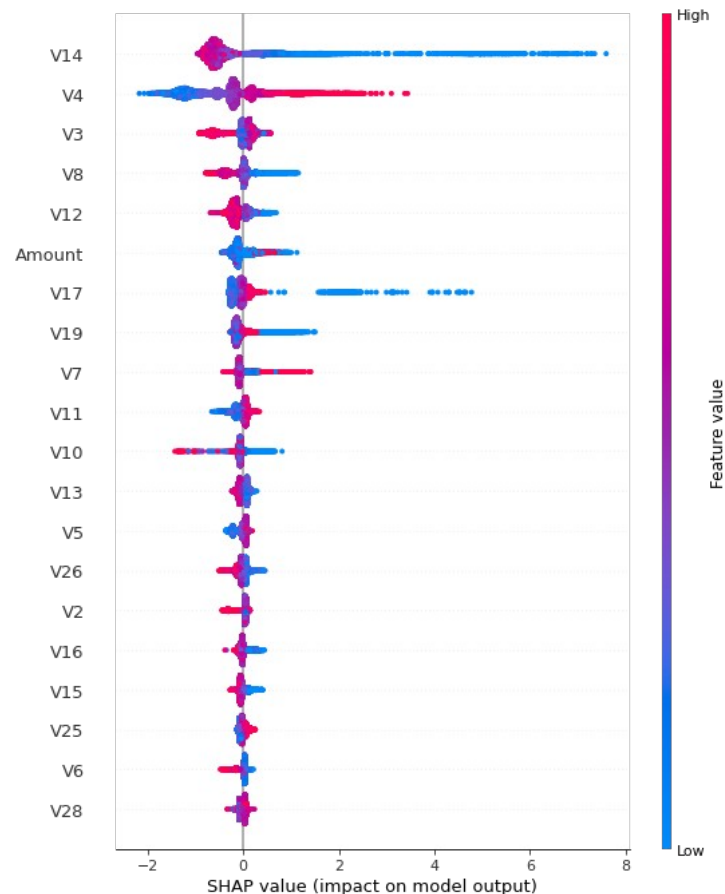
Performance



ROC AUC score: 0.9794

Feature Importance

Most influencing variables by SHAP are V14, V4, V3, V8, V12, Amount, V17 and V19.



Summary and Conclusions

With all variables included and originally distributed, XGBoost classifier demonstrates ROC AUC score 0.979.

Probability cutoff 0.5 results in precision 0.94 and recall 0.84.

The end

Thank you for your attention!