Credit Card Fraud Detection

...

Data Science Career Track
Egor Petrov
August 22, 2020

Overview

This project describes the various data wrangling, data analysis, fraud detection methods applied to the “Credit Card Fraud Detection” dataset.

Understanding the problem

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

Project objective:

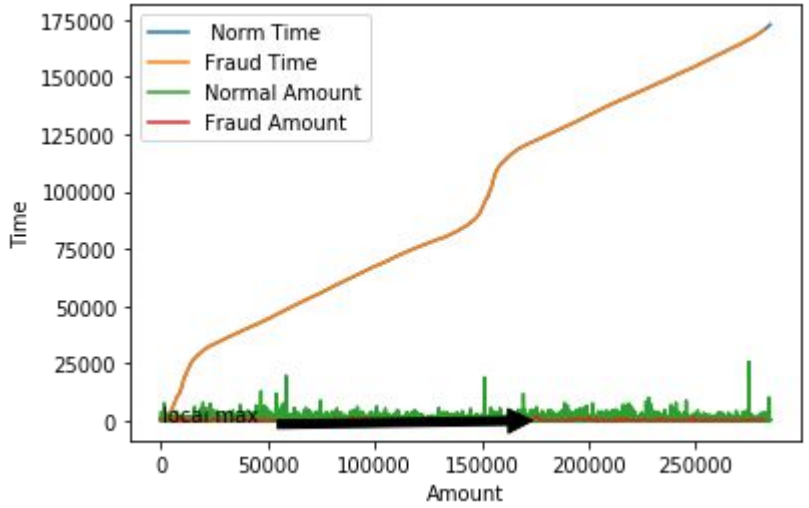
My aim here is to detect majority of the fraudulent transactions while minimizing the incorrect fraud classifications.

Target audience

- Banks
- Money Transfer Systems(PayPal)
- Credit Card Companies
- The best high-tech technologies to deal with fraudulent transactions. It requires a touch of a data scientist’s job and training a classification model based on given data.

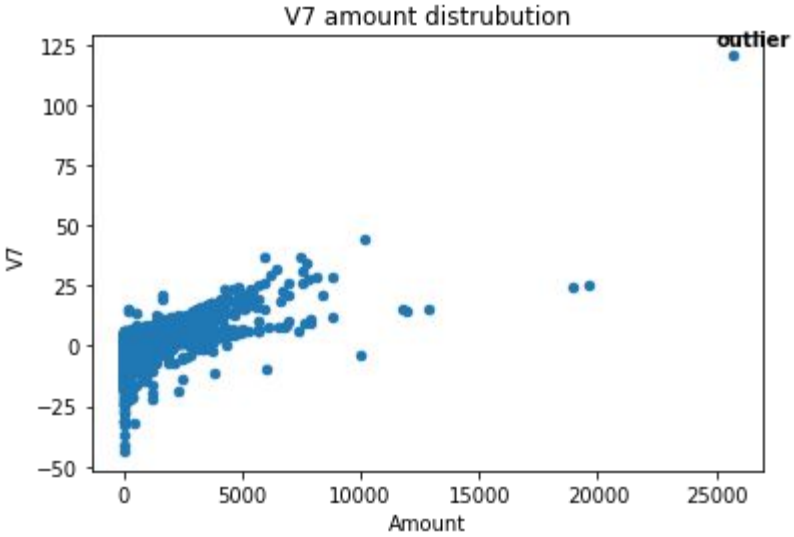
Exploratory Data Analysis

Finding Trends



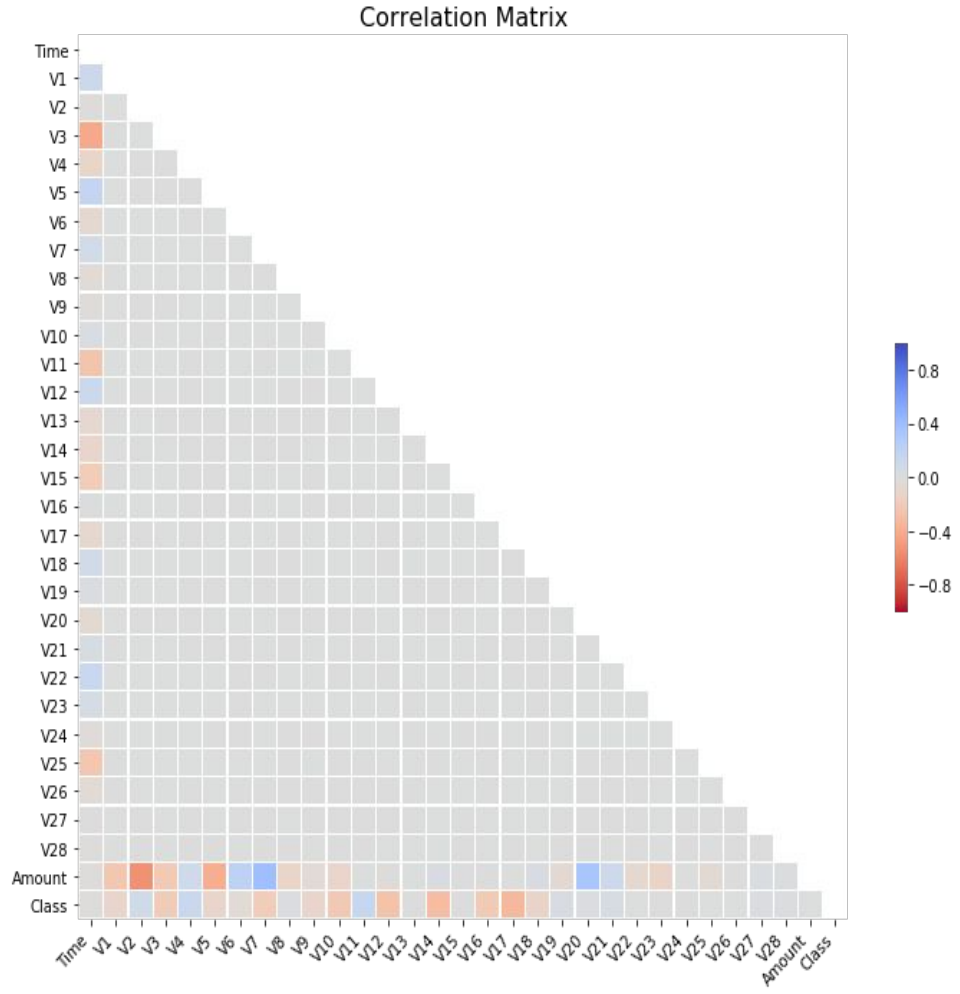
Outliers and Inliers

Finding outliers in my dataset



Statistical Analysis

Finding Correlations



Hypothesis Testing

Null hypothesis is comparing two groups for statistical significant difference .

Ttest

Ttest_indResult(statistic=18.761176823315797, pvalue=1.2885797612095705e-59)

P value is less than 0.5 so it indicates to reject Null hypothesis in favor of Alternative.

Anomaly Detection with Machine Learning

Finding the best model

Model Algorithm	Confusion Matrix
IsolationForest:	[401, 283914] [175, 317]
Support Vector Machine	[284305, 10] [81,411]
KNeighborsClassifier	[284315,0] [98,394]
Autoencoder	[284038,277] [352, 140]

Conclusion

Conclusion 1

- Chosen Algorithm is Support Vector Machine
  Confusion Matrix Metrics
  Detected 411 Fraud Transactions out of 492

Conclusion 2

- Undersampling and Oversampling as improving detection factors

Final Thoughts

Banks and Credit Card Companies can integrate a high-tech ML model and expertise of a Data Scientist to save time and money to do a job and invest in bottom line increase.

End Use Suggestions:

As an integrated application.
For Analysis and educational use.
Hiring a data scientist