Credit Card Fraud Detection

Several method of anomaly detection from Sk learn packages. We are going to be using a local outlier factor to calculate anomaly scores along with an isolation force algorithm. Here we have downloaded dataset from kaggle, just go to the kaggale dot com and create an account, find there credit card csv file just download the file and upload it in your project folder.

In the dataset we have 284808 transaction of credit card and we have to predict then which ones are the fraudulent one.

I have made the whole project using jupyter notebook. Initially we have to import the packages “sys”, “numpy”, pandas, “matplotlib”. For correlation matrix we have to import “seaborn” along with “sci py” packages also don’t forget to import “Sk learn” package. The we have print out the version of those packages.

My packages versions are:

Python: 3.7.3 (default, Mar 27 2019, 17:13:21) [MSC v.1915 64 bit (AMD64)]

Numpy: 1.16.2

Pandas: 0.24.2

Matplotlib: 3.0.3

Seaborn: 0.9.0

Scipy: 1.2.1

Sklearn: 0.20.3

After then we have loaded the data set from the csv file using pandas. After that exploring the dataset. We have found 31 different columns in the dataset.

We have plotted a histogram of each parameter. We have found very few 1 value or fraudulent transaction in comparison to our valid transaction. Then we have determined the number of fraud cases and number of valid cases with their percentage but dividing fraud cases with the valid cases. The results will be like this

0.0017234102419808666

Fraud Cases: 49

Valid Cases: 28432

In local outlier factor we have 97 total errors but don’t be frustrated now look at the accuracy it is almost 1 most sharply ninety nine point six five nine so it was very accurate. At the same time if you look down here at precision recall and f1 you will see that we are not quite as good as we think. Precision accounts false positive and recall accounts false negative and f1 score is the combination of those. However our isolation was little bit better. We are correctly identifying thirty percent actual fraudulent cases.

Sheikh Saif Al Islam

Computer Scienece and Engineering