Fraud Credit Card Transaction Detection

Final Report

Sushil Chaudhary
The Beacom College of Computer and Cyber Sciences
Dakota State University
Madison, South Dakota, USA
sushil.chaudhary@trojans.dsu.edu

***Abstract*—There are various types of payment method and one of the most popular methods is using credit card. With increase in use of credit card, fraud transactions are alarmingly increasing. Due to large number of transaction in each day, it is very difficult to identify the fraud transaction with traditional approach. It is carried out so precisely that it is difficult to detect by human eyes. So, Machine learning plays a vital role to identify the fraudulent transactions and timely taking legal action against the fraudulent actor. Fraud detection is very challenging due to ever changing transaction behavior or the requirement of precise and fast fraud detection algorithm. There is high demand of such a robust system by all kinds of financial sectors such as banks and credit card service providers. There are various types of machine learning approach for Credit-card fraud detection, however I am using XGBOOST algorithm for my class project. It was not specified which datasets to use, so I used "Credit Card Fraud Detection" datasets on Kaggle.com website.**

***Index Terms*—transaction, XGBOOST, fraudulent, algorithm, Credit-card, Kaggle.com**

I. Introduction

Twenty-first century is an era of digitization due to availability of internet, advancement of cellular networks and enhancement of other technologies. Growth of technology and growth of attacks to the technology goes side by side. One way it is convenient to use credit card another way it is insecure to use. We are in a situation that we cannot say any system is 100% secure. On provider side vast number of transaction to process it is almost impossible to detect fraudulent transaction with human eyes. It is easy to do fraud using credit-card and without the knowledge of owner, significant amount can be withdraw in a short period of time. Fraudsters will always try to make every fraudulent transaction so legitimate that it is almost impossible to detect it. "According to PwC's global economic crime and fraud survey [3], 47% of companies have experienced fraud in the past 24 months". "The total fraud losses reported by respondents reach a number of eye-watering 42 billion dollars [1]".Thus, our goal is to show how machine learning can be applicable to identify fraud using history of transaction.

In section II, it will be discussed about datasets and in section III, it will about Machine learning model used.

II. Dataset

It is very difficult to get data about credit card transaction because of privacy and confidentiality issue. The datasets that I used, is form Kaggle.com called "Credit Card Fraud Detection" [4]. It is a datasets of transactions made by European credit cardholders in September 2013 [4]. There are total of 284,807 transaction and out of which there are 492 frauds [4]. As you can observe from the numbers the data sets is highly uneven. There are variables from V1 to V28 are already in decoded form due to user identity protection and because of sensitive features. We can observe Time and Amount columns. The values of Class column represents the types of transaction, value 1 indicates fraud and value 0 indicates normal transactions.

III. Machine Learning models

As I have mention my goal is to use XGBoost classifier for this project. However, I will also use Logistics regression and Random forest classifier for model comparison.

IV. Data Processing

First of all data is loaded into Jupiter-Notebook as a dataframe and then data processing is begin. Since data value in Time and Amount column are large, data standardization is done. I plotted the feature density with respect to the class type as you can see in "Fig. 1". As you can see in the figure each class has different density for same feature and for different features both of them are different. This indicate that some features are important for Class = 0 and some are important for Class = 1. After visualizing the feature density plot, training data is breakdown into X-features and Y-label. These X-features and Y-label are again splitted for training and testing into the ratio of 70% to 30%. After that model is trained for 1st time and for the second time data will be down sampled to size of fraud class to detect more fraud.

V. Model comparison

The accuracy score is not good metric to measure the performance of the models. Because we need to focus on identifying the fraud transactions than the normal ones. Accuracy measure overall correct prediction score. So my goal is to select the model based on the lowest number of frauds predicted

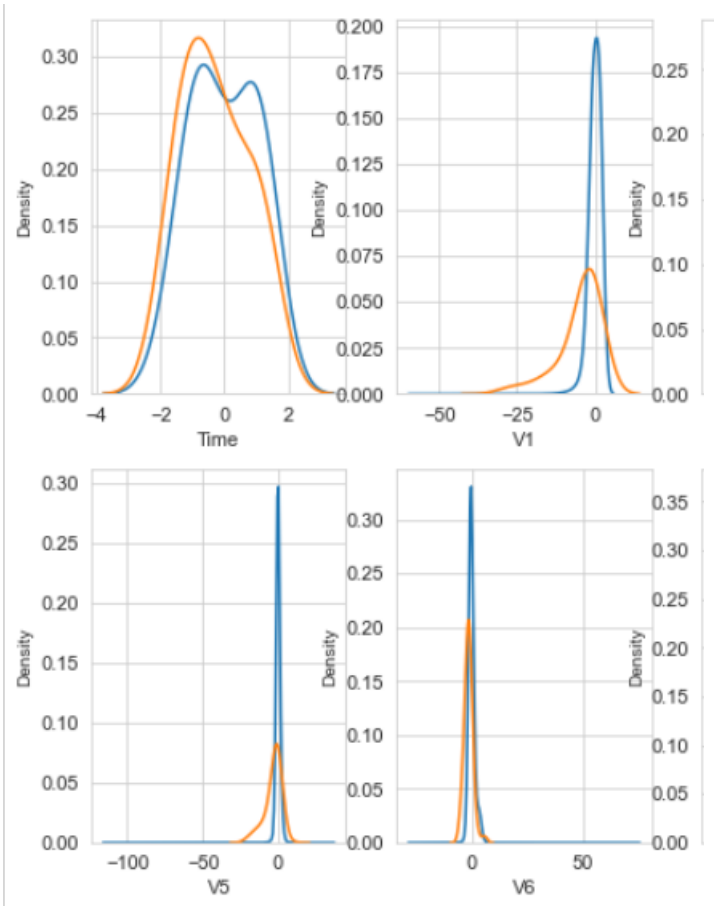

Fig. 1. Feature density

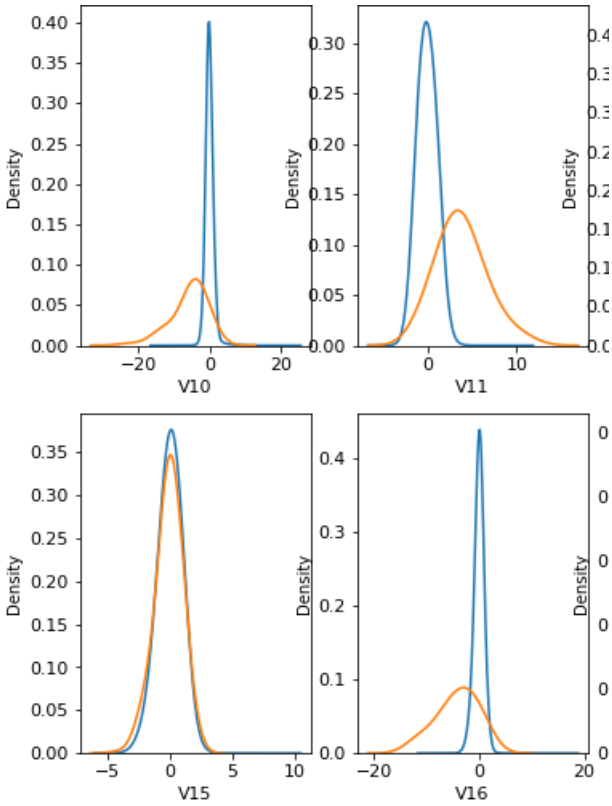

Fig. 2. Feature density

as normal. The detail evaluation is shown in "Table. I". "Table. II" represents the percentage of true fraud predicted. As you can see XGBoost works well in both cases i.e., before down sampling and after down sampling.

TABLE I
MODEL EVALUATION

Fraud as Normal %	**Classifier**		
	XGBoost	***RandomForest***	***Logistic regression***
Without down-sampling	16%	17%	37%
With down-sampling	9.3%	9.3%	7.3%

TABLE II
MODEL EVALUATION

True Fraud %	**Classifier**		
	XGBoost	***RandomForest***	***Logistic regression***
Without down-sampling	84%	83%	63%
With down-sampling	91%	91%	93%

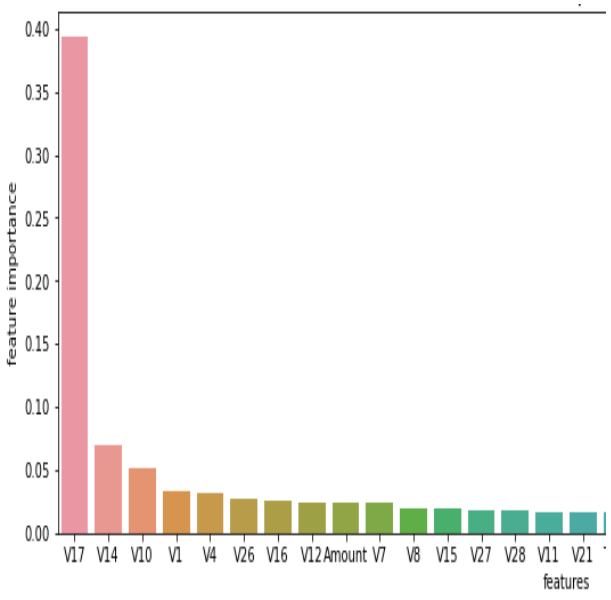

Fig. 3. Feature importance

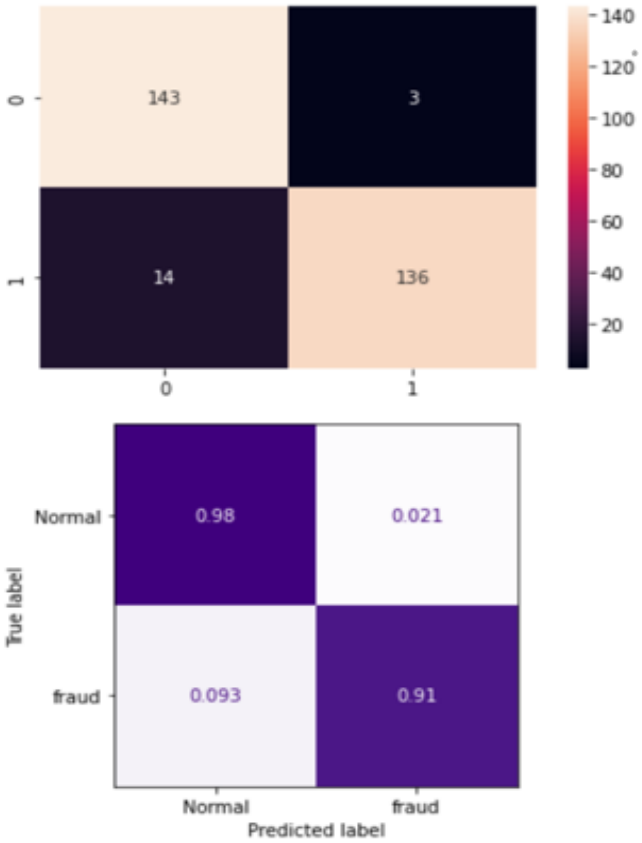

Fig. 4. Confusion Matrix of XGBoost Classifier

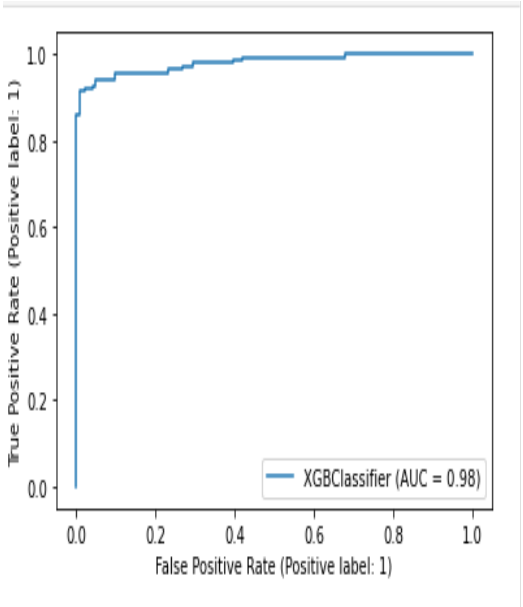

Fig. 5. ROC of XGBoost Classifier

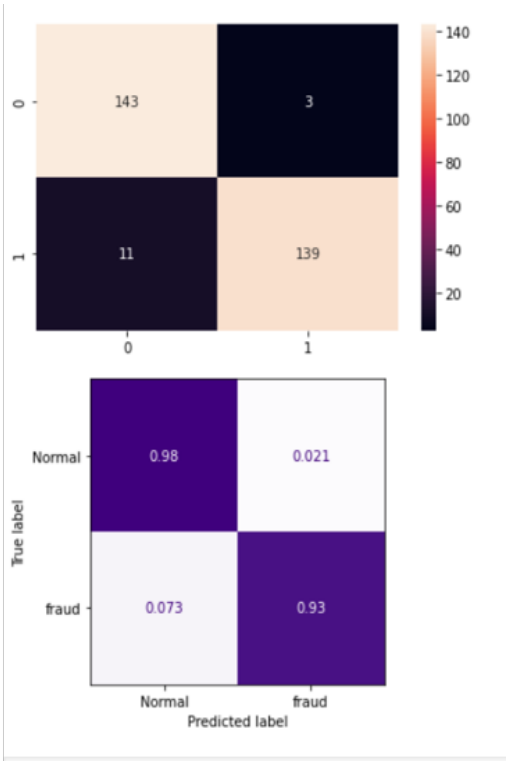

Fig. 6. Confusion Matrix of Logistic Regression

VI. Conclusion

From above experiment it can be concluded that Machine learning can be used for Credit card fraud detection based on the history of user's transactions. Three Different model are compared and among them XGBoost classifier performed well for both after and before down sampling. Logistic Regression perform well in down sampling case but worst before down sampling. Thus XGBoost could be the one algorithm that can be implemented in real world for Credit card fraud detection.

References

[1] Y. Chen and X. Han, "CatBoost for Fraud Detection in Financial Transactions," 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 2021, pp. 176-179, doi: 10.1109/ICCECE51280.2021.9342475.

[2] N. Malini and M. Pushpa, "Analysis on credit card fraud identification techniques based on KNN and outlier detection," 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017, pp. 255-258, doi: 10.1109/AEEICB.2017.7972424.

[3] Global economic crime and fraud survey 2020. (2020). Retrieved 2 October 2020, from https://www.pwc.com/gx/en/forensics/gecs2020/pdf global-economic-crime-and-fraud-survey-2020

[4] https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud