

# Credit Card Fraud Detection using ML

Name:	Abhay Kumar
Registration No./Roll No.:	19354
Institute/University Name:	IISER Bhopal
Program/Stream:	DSE 302
Problem Release date:	January 19, 2022
Date of Submission:	April 24, 2022

## 1 Introduction

The topic that I chose to work on is Credit Card Fraud Detection using Machine Learning. I chose this because Credit Cards are important part of our life. Thus, predicting and avoiding frauds in Credit Cards is crucial.

The dataset (.csv file) which I will work on contains transactions made by credit card users in September 2013 by European cardholders. It presents transactions that occurred in two days, where there were 492 frauds out of 284,807 transactions (i.e., only 0.172 percent transactions are fraud).

## 2 Methods

Firstly, we used different ways to analyze and interpret the data, such as Time-Density plot, Features-Density plot, etc.

Then, we used 6 predictive classifying models, namely KNN Classifier, RandomForrestClassifier, XGBoost, AdaBoostClassifier, LightGBM and CatBoostClassifier. We used hypertuning in LightGBM model.

Lastly, since the ratio of fraudulent transactions to normal transactions is highly skewed and imbalanced, we also used Area Under Precision-Recall Curve (AUPRC) in addition to the usual metrics like Accuracy, Precision, Recall, etc.

I did AdaBoostClassifier, LightGBM, CatBoostClassifier, and did the final prediction.

Table 1: Comparing the performance of different predictive models using evaluation metrics

Models	Accuracy	ROC AUC	Precision	Recall
KNN	0.9977	0.9232	0.9942	0.9974
RF	0.9995	0.9422	0.9951	0.9964
AdaBC	0.9992	0.9951	0.9975	0.9953
CBC	0.9975	0.9231	0.9950	0.9949
XGB	0.9969	0.9244	0.9949	0.9958
LightGBM	0.9963	0.9782	0.9941	0.9961

### 3 Experimental Analysis

From the given table, we can see that Random Forest Classifier has the highest accuracy of all.

Thus, we used RF model to predict our results.

The resulting table we got is attached with the name "class labels.txt".

### 4 Discussions

Since the data set was small, getting high accuracy wasn't a problem. For large data sets too, this works fine, although with a minor mishaps here and there.

I think this problem is pretty important for all of us, since majority of us use credit cards in our daily life. So the scope of any solutions on this problem are high enough.

### 5 References

1. Dal Pozzolo, Andrea Caelen, Olivier Le Borgne, Yann-Aël Waterschoot, Serge Bontempi, Gianluca. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications. 41. 4915–4928. 10.1016/j.eswa.2014.02.026.
2. Lebichot, Bertrand Le Borgne, Yann-Aël He, Liyun Oblé, Frédéric Bontempi, Gianluca. (2019). Deep-Learning Domain Adaptation Techniques for Credit Cards Fraud Detection. 10.1007/978-3-030-16841-4<sub>8</sub>.
3. Le Borgne, Yann-Aël Bontempi, Gianluca. (2021). Reproducible Machine Learning for Credit Card Fraud Detection - Practical Handbook.