# Use of Machine Learning to Detect Credit Card Fraud

Sachin Agwan and Nitin Patil



June 19, 2022

# Use of machine learning to detect credit card fraud

**Sachin Kumar Agwan[1], Nitin Patil[2]**

[1]Department of Technology,
*Savitribai Phule Pune University,*
*sachinagwan1994@gmaill.com*

[2]Department of Computer Science
*Savitribai Phule Pune University,*
*nitindocs@gmail.com*

***Abstract*: *For business owners, payment card issuers and transaction service providers, payment card fraud is a major problem, leading to significant and growing financial losses every year. As we all know, detecting fraudulent systems in payment card or plastic money dealings and transactions is a very complicated task. As the amount of data produced by payment card or plastic money dealings and transactions continues to increase, it is not possible for human analysts to identify fraud patterns in dealings and transaction records, which are usually characterized by high sample numbers, multi-dimensional and online updates. In the past decade, the development of methods for detecting payment card fraud has increasingly focused on machine learning (ML) methods, which automate the process of detecting large-scale fraud programs.***

***Integrating the LD method into the credit card fraud detection system greatly improves their ability to detect fraud more effectively and help payment intermediaries detect illegal transactions. In 2016, the number of fraud cases started to decline-a counter-trend related to the increasing adoption of machine learning solutions. Now, the introduction of machine learning-based fraud detection systems not only helps save money, but also becomes an obligation for institutions and companies to gain customer trust.***

***In this emerging DL field of DL card fraud detection, a well-known problem that persists is that most published studies on this topic lack reproducibility. On the one hand, due to confidentiality reasons, no data on payment card transactions cannot be released. On the other hand, the authors did not go to great lengths to provide their code and make the results reproducible.***

***Take the first step of repeatability to compare and analyse payment card fraud detection methods. Due to the large amount of research in this field, it is impossible to comprehensively analyse and implement all existing methods. Some of the most important technologies for us***

***Some of the proposed techniques, such as model set or concept drift in the case of class imbalance, are widely regarded as an important part of designing a credit card fraud detection system. Among other things, this includes the design aspects of the simulation process, such as the selection of performance indicators and verification strategies, as well as promising preprocessing and learning strategies, such as inclusion of functions and active and portable investigation.***

***Keywords*: Machine Learning, Convolution, Max Pooling, Transfer Learning.**

**(Article history: Received: ##/##/#### and accepted ##/##/####)**

## I. INTRODUCTION

Today, businesses and government agencies are facing more and more fraudulent activities and require automated systems for fraud detection. To detect the fraud specimen in transaction, automated systems are much essential as it not practically possible for human analyst to detect the all possible fraud records, which are usually characterized by high sample numbers, multi-dimensional and online updates. Many training algorithms are known to be ineffective when used on unbalanced data sets, and methods to improve their performance (for example, sampling) is being introduced. We cannot rely on imbalance as one and only component to determine the trouble of the classification work. We can consider another component i.e. the degree of intersecting between categories of interest due to restricted information regarding the nature of the operation provided by transaction records.

The recognition problem is usually solved in two non-identical ways: In a fixed study environment, the recognition replica is learned from afresh on regular basis (such as once a year or once a month) because the entire data set can be stored and processed at the same time. Configuration, as long as there is new data, the recognition model will be updated. However, this strategy is best for solving non-stationary problems. Due to the evolution of the consumption behavior of ordinary cardholders or fraudsters, little attention has been paid to unbalanced environments in the literature. Another problem with credit card verification is insufficient data due to privacy issues, leaving the community almost no opportunity to share real records and evaluate existing methods.

This article aims to fill this gap by comprehensively comparing a large number of algorithms and modelling techniques of two real data sets, focusing on some unsolved problems, such as B: What algorithm should you choose for machine learning? Is it sufficient to study the replica once a year? Bimonthly do you have to make changes to the model every day? Finding out the number of transactions required to form the model? Do we require to analyze the data in its real unbalanced phase? If not, what could be the most effective way to balance them? What could be the best performance indicator for measuring results? In this article, we will focus on these topics and evaluate their importance in real data from a professional perspective. Questions and provide a formal form of identifying the problem. Then, we discuss existing alternatives to classic classification indicators that should be considered when solving the identification problem. Then, we compare some online learning methods to determine what important information is retained or forgotten in the ever-changing transient environment. Finally, we show the impact of equalization techniques on the final performance. All results are obtained through experiments. Two sets of actual credit card transaction data used.

## II. RESEARCH METHOD

Every time a transaction on credit card takes place or plastic money transaction occurs, the transaction produced data consists of many unique attributes (such as CC (credit/debit card) ID, date of transaction, amount recipient, transaction money) stored in service provider's database. However, the information of a one transaction usually not enough to identify fraud, so it needs to be analysed at the entire credit card phase. At the entire card phase, all transactions can be grouped with the same card & we can check the behavior of individual cards. For each card, attributes such as total expenditure per day, number of transactions per week, or average transaction amount are retrieved and analysed to detect card fraud. credit card.

### *A. Detection*

As far as classification is concerned, it is not necessarily a good description of performing detection tasks (for example, detecting fraud). In the detection problem, the most important thing is whether the algorithm can put several useful elements (such as fraud) on top of the rest. Therefore, the focus is not on accurately predicting each class, but on returning to the correct minority class range. In detection equipment such as our industrial partner equipment, every time the detection system generates a fraud alert, it must be checked Investigators before taking action (such as contacting customers or restricting card numbers). Given the limited number of researchers, only a limited number of warnings can be tested. With this setting, it is important that the first X warnings are as accurate as possible. In the following, we refer to precision as PrecisionRank the highest level of observation

### *B. Supervised Vs Unsupervised Detection*

We can find supervised and unsupervised methods which uses one type of transaction (such as real or fraud). The traceable method requires that the classification of earlier transaction is obtainable and dependable rather is usually restricted to detecting pre-existing fraud models. But, unsupervised method aim to identify those transactions that deviate the most from the standard by simulating the distribution of most transactions. This article focuses on controlled methods. Various supervised fraud detection methods are used in the literature, such as: B. Algorithms such as Tree-based algorithm (C4.5 and CART), Neural Networks(NN) and methods such as (BAYES, RIPPER) which are rule-based. However, as we all know, an unresolved problem is how to manage the size of imbalanced categories, because there are usually more legitimate transactions than fraudulent transactions.

### *C. Unbalanced Problem*

Researching unbalanced data sets is challenging for the reason that most training structure are not made to control the enormous difference among total quantity belonged per category cases. Conventional method of sampling techniques is used to balance the unbalanced data sets. Algorithms and methods can be differentiated that needs to work on the data plane. At the initial stage, equalization techniques are being utilize at the data level to equilibrate the data set or rectify the errors. Before applying any algorithm, switch in the middle of the two categories. At the design level, the categorization design is for handling the minority classes detetction. Here, we will concentrate on data-level sampling and assembly methods. Invalid sampling method. When removing or adding observations from the class, consider certain information that is easy to implement and understand. Under sampling involves reducing the dimensions of the bulk categories by eliminating random monitorings until the data set is balanced. CNN oversamples minority categories by generating synthetic minority samples that are very close to the observed samples. The idea is to create new examples of the minority by incorporate in the middle of specimen of the similar category. Which leads to the formation of bunches on all sides every few monitoring points. Weighing procedures with classifiers can study the distribution of majority and minority ethnic categories. This method explores all aspects of the primitive majority in an uncontrolled way. For this, different training sets are created, which are balanced by down sampling, then the model is checked for every data set, and then each and every forecast (for example, packing) are combined together.

### *D. Discussion for static methods, which algorithm is recommended?*

The static method is one of the most widely used methods by those skilled in the art due to its simplicity and speed. However, there are still questions about which training algorithm to use along with the acuity of the accuracy of the outcome to the practice measure. Analysis of two dissimilar algorithms with standard parameters provided by R software: Support Vector Machine (SVM), Random Forest and Neural Network (NNET). In these two sets of data, as expected, Random Forests is significantly better than its competitors in terms of accuracy. It can be improved by increasing exercise. In view of the advantages of random forest over other algorithms, we will limit ourselves to this learning algorithm in the future.

*1) Impingement on precise results by balancing techniques:* In our experiments, we have only regarded subsampling as a balancing technique. In this segment, understand and evaluate the effect of utilizing different procedures such as CNN. In particular, the best strategies in the two datasets include CNN for PrecisionRank.

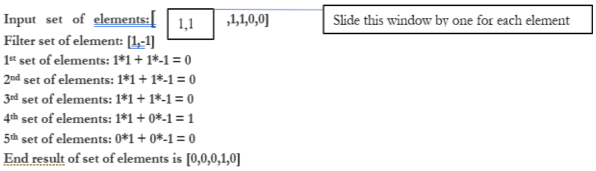
*2) By considering all the scenario and results which is the most efficient method:* Among the considerable alternatives (with reference to incremental learning strategies, equalization techniques and learning classifiers) essential combined evaluation of numerous fusions to show up at a suggested method. Determine the strategy for the upgrade method. This confirms studying online in terms of balancing data, the balancing strategy adopted can play an important role. Unsurprisingly, static methods have low scores and cannot adapt to changing distributions.

## III. RESULTS AND DISCUSSION

Classic subsampling. We only tested three balanced methods, but other unbalanced methods can be easily added to the structure. Our structure breaks the classic incremental learning structure by preserving the information in the early fragments. However, the recommended strategy uses a small number of previous cassette tape transactions. The experimental part shows that when learning online, it is very important to keep the previous examples when keeping the data consistent with the class.

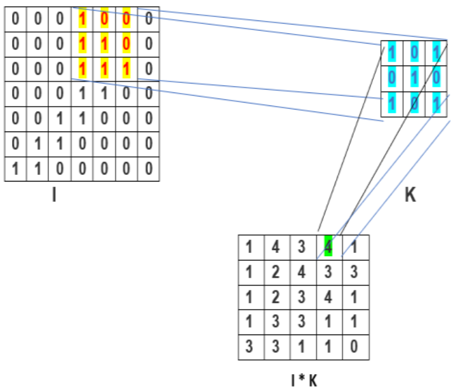
### A. *Principles of CNN (Convolution neural networks)*

- ***Convolution***: Convolution traverses the window in the image, then calculates the input and filters out pixel values from the dot product. This allows the folding to highlight related purpose.



1D convolution operation with filter (above)

Check out this article. We encapsulate the window element in a small window, multiply it with the filter element at this point, and then save the result. We will repeat each operation to get 5 output gates of [0,0,0,1,0]. From this output, we can see that the characteristics have changed in sequence 4 (1 becomes 0). The filter can well recognize the input values 2D turns and the same thing happened.



2D convolution operation with filter

Through this calculation, you can identify specific features in the input image and create feature maps (reduced features) that highlight important features. These confusing properties always change according to the filter value affected by gradient descent to minimize prediction loss.

- ***Max Pooling***:

**Max Pooling slices of 2x2 filtered area**

CNN uses the max pool to replace the output with the maximum number to reduce data size and processing time. This will help you identify the most influential features and reduce the risk of overfitting. The maximum pool requires two hyperparameters: stride and size. The stride will identify the skip of value pools and the size identifies how big he values of group in every skips.

- ***Transfer Learning***: As use cases become more and more complex, the complexity of the model should also increase. With some CNN layers. However, in the deep learning stage, you may want to classify more complex objects in the dataset and use more data. Transfer learning allows you to quickly classify using existing models instead of teaching them yourself. Transfer learning is a technique that reuses existing models in existing models. You can use an existing model that has been carefully designed. There are some restrictions: First, you must change the last level to match the number of possible classes. Second, you need to freeze the parameters and set the immutable variables of the training model. dramatic change.

## IV. CONCLUSION

This paper formalizes the problem of fraud detection and proposes accuracy and total cost of detection as the correct indicators to measure detection efficiency. As far as we know, several attempts have been made in the literature to combine incremental learning research and unbalanced data. In this article, we propose a fraud detection framework that can use any technology developed for unbalanced data sets. In particular, we used method (CNN) and showed that both methods can improve performance. Classic subsampling. We only tested three balanced methods, but other unbalanced methods can be easily added to the structure. Our structure breaks the classic incremental learning structure by preserving the information in the early fragments. However, the recommended strategy uses a small number of previous cassette tape transactions. The experimental part shows that when learning online, it is very important to keep the previous examples when keeping the data consistent with the class. The proposed second method and the third method have significant differences in the number of minority groups in the current part, but the impact is limited due to the low level of fraud. Live in time. Latte caused more deception, thereby reducing the distortion of training. In the collection, it is created by combining the inspected models into simple blocks. Our experiments show that by combining the information in the current segment and the previous segment, we can improve the performance of each model in the collection. It has been shown that many models can improve performance through subsampling. However, this is not better than a single model using CNN before training. Our skeleton solves the issue of unstable data flow by introducing a new design in each block. And using the axial window in the clips in the training set shows better results than upgrades. Model with lower frequency. Further work focuses on improving the combination of imbalanced methods in online learning.


ACKNOWLEDGMENT *(Heading 5)*

NONE


REFERENCES


[1] G. Ditzler, R. Polikar, N. Chawla, Incremental learning algorithm for non-stationary environments and class imbalance, in: Pattern

Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010, pp.2997-3000.

[2] H. Wang, W. Fan, P. S. Yu, J. Han, Mining concept-drifting data streams using ensemble classi_ers, in: Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2003, pp. 226-235.I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] F. Provost, Machine learning from imbalanced data sets 101, in: Proceeding of the AAAI2000 Workshop on Imbalanced Data Set, 2000.

[4] S. Chen, H. He, Sera: selectively recursive approach toward nonstationary imbalanced stream data mining, in: Neural Networks,

## AUTHOR PROFILE

**Sachin Agwan**
Savitribai Phule Pune University, M.Tech. (Computer & Information Technology) [Appearing] and B.E.(Computer).
Working on this project under Prof. Nitin Patil guidance. Prof. Nitin Patil is working as Assistant Professor in Savitribai Phule University in Department of Computer Science.