# DEEP LEARNING AND FRAUD DETECTION IN CREDIT CARD TRANSACTIONS

Dinesh Ghanta
Student, PGDM in Business Analytics
RACE, REVA University
Bangalore, India
Dineshg.ba01@reva.edu.in

Kavitha Mailengi*
Student, PGDM in Business Analytics
RACE, REVA University
Bangalore, India
Kavitham.ba01@reva.edu.in

Ratnakar Pandey
India Head, Analytics and Data Science
Kabbage
ratnakarpandey20@gmail.com

## ABSTRACT

As per a recent article from CNBC News, around 15.4 Million American Consumers were victims of identity theft or some other fraudulent activities in 2016, resulting in almost $16 Billion losses. Needless to say that real-time fraud detection has become an imperative for the global banking and financial services industry to survive in a very competitive and fast-changing environment.

Currently, most of the players in the BFSI industry use Machine learning techniques such as logistic regression, decision tree, Elastic Net, Gradient Boosting Tree to identifying potential fraud cases. However, with ever-evolving modus operandi of fraudsters, we need to deploy self-learning/ deep learning algorithms such as Autoencoder, LSTM and other Deep Learning Neural Network to stay a step ahead of fraudsters.

In this paper, we have built an algorithm using a combination of two different deep learning networks- Multilayer Perceptron and Autoencoders to classify the labelled fraud cases of a European Credit Card company to achieve the highest levels of precision and recall. By adjusting hyper-parameters through grid search and using Autoencoder output of data, we are able to get good results for MLP technique.

**Keywords:** *Fraud Detection, Keras, Autoencoders, Multilayer Perceptrons, Artificial Neural Network (ANN)*

## 1. INTRODUCTION

As per a recent article from CNBC News, around 15.4 Million American consumers were victims of identity theft or some other fraudulent activities, resulting in almost $16 Billion losses in 2016 alone. Needless to say that real-time fraud detection has become an imperative for the global banking and financial services industry to survive in a very competitive and fast-changing environment. Making fast and accurate decision on whether a transaction is fraudulent can make or break the relationship with a customer. (FICO report, 2013)

Currently, most of the players in the BFSI industry use Machine learning techniques such as logistic regression, decision tree, Elastic Net, Gradient Boosting Tree to identifying potential fraud cases. However, with ever-evolving modus operandi of fraudsters, we need to deploy self-learning/ deep learning algorithms such as Autoencoder, Multilayer Perceptron (MLP) and other Deep Learning Neural Networks to stay a step ahead of fraudsters.

In this paper, we have built an algorithm using a Deep learning network- MLP and optimized hyper-parameters using Grid search to classify the labeled fraud cases of a European Credit Card company to achieve the highest levels of precision and recall.

## 2. LITERATURE SURVEY

We performed a literature survey on the extant research in the areas of fraud detection and management, credit card fraud detection, deep learning techniques of Multilayer Perceptron, Hyper parameters and Autoencoders.

### 2.1 Fraud and fraud detection

Fraud is a general word for illegal use of a system to obtain some benefit, usually resulting in a loss to another person or institution. Financial fraud is obtaining financial benefit illegally from a transaction. Billions of US dollars have been lost to fraud (Hurwitz et al., 2013). The proliferation of internet usage has made it easier to communicate and connect from a distance. It has also made it easier for fraudsters to target financial institutions from a distance. This increases the threats to security systems. Hence, fraud prevention and detection are important issues to all financial institutions. (Isa, 2016)

Bolton et al., refers Fraud prevention as all the measures to prevent frauds from happening, while Fraud detection refers to mechanisms to detect Frauds as quickly as possible when prevention fails (2002).

### 2.2 Credit card fraud detection

The advent of credit cards has helped people tremendously and has become a target for financial fraud too. Some of the credit card fraud techniques involve stealing or copying the card information or vendors charging more money than agreed without the customer being aware of it. All these results in financial losses to the customer as well as the banks. It is in the interest of both parties that fraudulent transactions are detected very quickly. Fraud detection is the process of identifying those transactions that are fraudulent i.e. classifying the transactions into genuine and fraudulent transactions (Maes et al., 2002).

### 2.3 Challenges in credit card fraud detection

Some of the challenges in credit card fraud detection are:

- Nonavailability of real-time data, since banks and financial institutions, are not ready to reveal sensitive customer information.
- Unbalanced dataset as the number of fraudulent transactions is very small compared to the number of genuine transactions. Generally, in the real case, 98% of the transactions are legal while only 2% are fraud (Piotr et al., 2008).
- Size of the data set: Analyzing enormous amounts of data requires techniques which are competent and scalable.
- Determining the appropriate evaluation metrics: Accuracy is not a suitable metric for credit card fraud detection as the data is highly imbalanced (Bolton & Hand, 2001). The error cost of misclassifying fraudulent transactions is higher than error cost of misclassifying genuine ones. Hence, it is important not only to study precision (correctly classified), also sensibility (correct classified fraudulent cases) of each case (Zareapoor & Shamsolmoali, 2015).

### 2.4 Techniques for credit card fraud detection

All techniques for credit card fraud detection should consider the requirements of preventive and detective systems to ensure success in real-world scenarios. It is necessary for preventive systems to be as precise as possible. Detection systems, on the other hand, need to adapt to the constant evolution of threats. Therefore, in addition to being predictive, Fraud detection systems need to be adaptive. A related concern is the time required to detect fraudulent transactions (Isa, 2016).

There are several approaches to credit card fraud detection used by researchers and practitioners.

#### 2.4.1 Rule-based approach

These methods list all the fraud characteristics based on past experience and use these as rules to detect future fraudulent transactions (Isa, 2016). While these supervised learning methods can detect fraud, they have issues in scalability due to their manual nature. An example is RIPPER algorithm (Cohen, 1995).

### 2.4.2 Statistical methods

The transactional data is assumed to follow a statistical distribution, and thus, transactional data points that fall out of the normal distribution are considered suspicious. Such methods include Linear Discriminant Analysis and Logistic Regression (Bolton & David, 2002). A challenge with the statistical method is determining the most suitable distribution to fit a data set, and as dimension increases, it becomes more difficult to estimate the distribution (Aggarwal & Yu, 2001; Tan et al., 2005).

### 2.4.3 Machine Learning

While there are several machine learning techniques that can be used to detect fraud, the quest is for faster and better models to detect fraud. Hence, the focus of this paper is, to use deep learning technique of multi-layer perceptron and optimizing hyper-parameters through grid search.

### 2.4.4 Deep Learning - Multilayer Perceptrons

Machinelearningmastery.com defines a Perceptron as "a single neuron model that was a precursor to larger neural networks. Multi-layer perceptron or forward feed neural networks help investigate how simple models of biological brains can be used to solve difficult computational tasks like the predictive modeling tasks in machine learning. The goal is to develop robust algorithms and data structures that we can use to model difficult problems.

The power of neural networks comes from their ability to learn the representation in training data and how to best relate it to the output variable that you want to predict. In this sense, neural networks learn a mapping. Mathematically, they are capable of learning any mapping function and have been proven to be a universal approximation algorithm."

**2.4.4.1 Model Parameter**: A model parameter is a configuration variable that is internal to the model and whose value can be estimated from data. They are required by the model when making predictions, they are estimated or learned from data, are often not set manually by the practitioner and are often saved as part of the learned model. Parameters are key to machine learning algorithms. They are the part of the model that is learned from historical training data. An example includes the weights in an artificial neural network.

**2.4.4.2 Model hyper-parameter** is a configuration that is external to the model and whose value cannot be estimated from data. They are often used in processes to help estimate model parameters, are often specified by the practitioner, can often be set using heuristics, are often tuned for a given predictive modeling problem. When a machine learning algorithm is tuned for a specific problem, such as when you are using a grid search or a random search, then you are tuning the hyper parameters of the model or order to discover the

parameters of the model that result in the most skillful predictions. Some examples of model hyperparameters include the learning rate for training a neural network.

**2.4.4.3 Hyper-parameter optimization using grid search**: Hyper-parameter optimization should be regarded as a formal outer loop in the learning process (Bergstra, 2011). Hyperparameters can be continuous or categorical, but what they have in common is that there is no single efficient learning algorithm for them. Many researchers rely on searching them on a grid. The performance of a model on test data trained with specific hyper-parameters depends on the data set where the machine learning model should be learned, and therefore hyper-parameter optimization usually starts from the scratch for each new data set (Schilling et al., 2015).

**2.4.4.4 Autoencoder:** While supervised learning has been a powerful use of neural networks, one of the limitations has been that we still need to manually specify the input features given to the algorithm. This manual intensive approach may not be scalable to all new problems like credit card fraud detection. Autoencoder learning algorithm is one approach to automatically learn features from unlabeled data. An Autoencoder neural network is an unsupervised learning algorithm that applies backpropagation, setting the target values to be equal to the inputs (Andrew, 2011). This transformation can be used as input for another model. This is especially useful in situations like fraud detection where we can input large amounts data, with several variables. These variables will be reduced without manual intervention to provide transformed data quickly required to detect fraud.

A lot of work has been done in the past on credit card fraud detection using various techniques. Currently, there is a lot of interest on using deep learning techniques including Multilayer Perceptron, optimizing hyperparameters and using autoencoders to better the solution.

## 3. DATA DESCRIPTION

The data used for this work contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that have occurred in two days, where 492 transactions were fraudulent out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

The data sets contain only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, the original features and more background information about the data is not available. Features V1, V2,… V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the Transaction Amount. Feature

'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise (Andrea et al., 2015).

## 4. METHODOLOGY

The objective of this paper is to build different models to predict the number of fraudulent credit card transactions and to highlight that the model using Multilayer perceptron (MLP) yields the best results. As a first step, data was explored to check the classification of the fraudulent and normal transactions. Distribution of all the variables was checked to get a better understanding of the data. Since the variables had different scales, they were scaled down between 0 and 1. The resulting output was plotted to confirm the distribution.

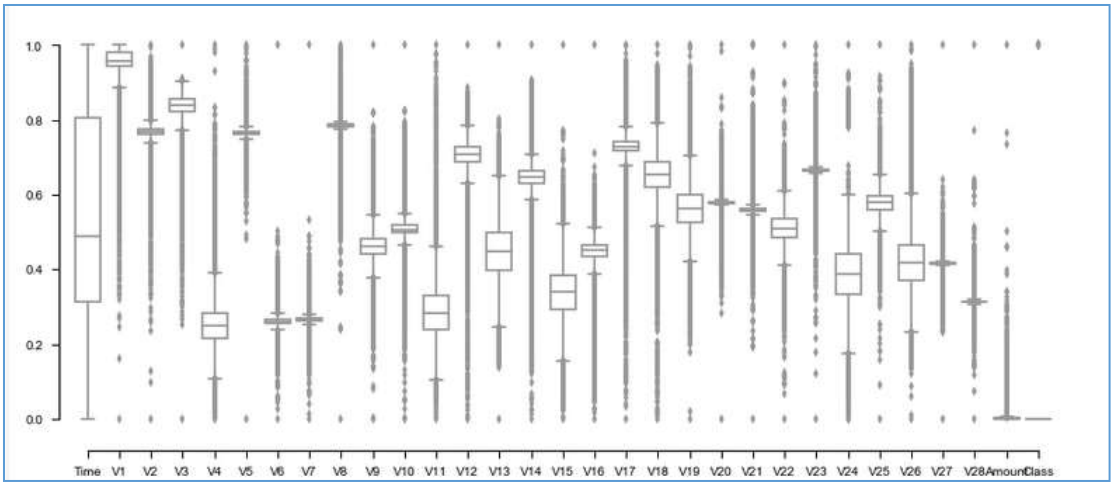
Data were split into train and test data. The model was trained using a sample dataset from the training dataset and validated on the test dataset. Performance metrics for evaluation were determined and models were evaluated based on these metrics. MLP model emerged as the best model based on the performance metrics evaluation. The link to the dataset and Python script used for the study is uploaded to GitHub and the link is here. The procedure followed is discussed below.

### 4.1 Data understanding and pre-possessing

As seen in Table 1, data is imbalanced, there are 2,84,315 normal transactions and 492 fraudulent transactions. All the variables have been scaled between 0 and 1 using MinMax Scalar. As seen in Graph 1, there are outliers in many variables. Since the purpose was to predict fraudulent transactions i.e. anomaly to normal transactions, these outliers are not removed before building models.

| Classification | No. of transactions |
|---|---|
| Fraud | 492 |
| Non Fraud | 284315 |

**Table 1** No. of normal and fraudulent transactions




**Graph 1:** Box plot of the variables

### 4.2 Train / Test data sets and sampling strategy

Data were split into training dataset (80%) and testing dataset (20%), keeping the same proportion of normal and fraudulent transactions as per the population. This is required to ensure that our models are trained and tested on data that represent the population. Within the training dataset, due to the imbalanced nature of data, customized cluster sampling method is used. In this method, all fraudulent transactions from the train data were taken from the sample. Additionally, all normal transactions were clustered using K-Means and sample picked from each of the clusters.

### 4.3 Model fitting

The following non-deep learning and deep learning techniques were used for preparing models using data sampled from train dataset - Random forest, Adaboost, Gradient Descent and MLP. Data was input to Autoencoder and the resulting data output was used in MLP. Also, all hyperparameters were optimized through grid search.

| Hyperparameter | Name |
|---|---|
| Optimizer | Adam |
| Activation function | Tanh |
| Learning rate | 0.001 |
| Dropout | 0 |
| Weight/Kernel Initilizer | Go to normal |
| Epochs | 100 |
| Batch size | 20 |

**Table 2:** Optimized hyper-parameters through grid search

### 4.4 Model Evaluation:

Normally, accuracy results are considered for model evaluation. Accuracy results of all the models are very high (results are summarized in Table 3). In this paper, we have not considered accuracy as a performance metric as it is considered to be a biased metric (Zareapoor & Shamsolmoali, 2015). Other metrics to consider are precision and recall. We can balance what is more important for to the business: precision or recall. It is possible to get both up: one may choose to optimize a measure that combines precision and recall into a single value, such as the F-measure, in this case (Alan, 2013).

Since the number of fraudulent transactions is very small and the cost of wrongly classifying a normal transaction as fraudulent is high, precision, recall, F1 score and Lift% have been considered for evaluation (Tryolabs.com).

$$Precision\ \% \div Base\ \% = Lift\ \%$$

For example, for the MLP model, Lift % is calculated as $(\frac{0.5690608}{0.001738} = 3.27422\%)$

This metric shows how much percentage the model will determine the number of frauds more than the base percentage.

| Algorithm | Accuracy | Precision | Recall | F1 Score | Base metric of frauds from population | Lift % |
|---|---|---|---|---|---|---|
| MLP | 0.9986 | 0.5691 | 0.99963 | 0.7253 | 0.0017 | 3.2742 |
| Random Forest | 0.9978 | 0.4353 | 0.9998 | 0.6065 | 0.0017 | 2.5043 |
| Adaboost | 0.9959 | 0.2886 | 0.9998 | 0.4479 | 0.0017 | 1.6605 |
| Gradient Decent | 0.9935 | 0.20223 | 0.9999 | 0.3371 | 0.0017 | 1.1638 |

**Table 3** Comparison of Accuracy, Precision, Recall, F1 score and Lift

## 5. FINDINGS

Based on the metrics in Table 3, MLP model has the highest F1 score of 0.725 and Lift of 3.27%. This means that MLP model can determine 3.25% more fraudulent transactions than the base percentage. For determining credit card fraud, this is a good result as even a small percentage of lift will result in huge savings for financial institutions.

## 6. CONCLUSION

In this paper, we have developed and implemented Multi-Layer Perceptron technique for credit card fraud detection. By adjusting hyper-parameters through grid search and using Autoencoder output of data, we are able to get good results for MLP technique. We have been able to showcase that it outperforms the models built using traditional machine learning techniques such as Random Forest, Ada Boosting, and Gradient Descent.

We evaluated the proposed algorithm on a test data set and demonstrated that the conclusions can be used as a guideline for working on a classification problem where variable misclassification cost is high. We have shown that well-known performance metrics such as accuracy may not be suitable for such problems, instead lift percentage would be a better metric.

By implementing such deep learning algorithms in fraud detection systems, financial institutions can save money on financial losses due to fraud. The performance difference between deep learning model (MLP) and traditional models in our test data corresponds to 1% and higher. An improvement in performance by even 1% may result in millions of dollar savings. Due to high savings, financial institutions will benefit from such improvements. This results in higher customer retention and loyalty too.

**REFERENCES**


Aggarwal, C. C., & Yu, P. S. (2001, May). Outlier detection for high dimensional data. In ACM Sigmod Record (Vol. 30, No. 2, pp. 37-46). ACM.

Andrea Dal Pozzolo, Olivier Caelen, Reid A. Johnson and Gianluca Bontempi. Calibrating Probability with Undersampling for Unbalanced Classification. In Symposium on Computational Intelligence and Data Mining (CIDM), IEEE, 2015

Bergstra, J. S., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. In Advances in Neural Information Processing Systems (pp. 2546-2554).

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical science, 235-249.

Cohen, W. W. (1995, July). Fast effective rule induction. In Proceedings of the twelfth international conference on machine learning (pp. 115-123).

Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2013). Big data for dummies. John Wiley & Sons.

Isa, I. M. (2016). An adaptive predictive financial fraud detection approach using deep learning methods on a Big Data platform (Doctoral dissertation).

J. Piotr., A.M. Niall, J.D. Hand, C. Whitrow, J. David (2008). Off the peg and bespoke classifiers for fraud detection. Computational Statistics and Data Analysis, 52, pp:4521-4532.

Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002, January). Credit card fraud detection using Bayesian and neural networks. In Proceedings of the 1st international naiso congress on neuro-fuzzy technologies (pp. 261-270)

Ng, A. (2011). Sparse autoencoder. CS294A Lecture notes, 72(2011), 1-19.

Tan, P. N., Steinbach, M., & Kumar, V. (2005, May). Introduction to data mining: Pearson addison wesley. ISBN 0-321-32136-7.

R.J. Bolton, D.J. Hand (2001). Unsupervised profiling methods for fraud detection. In Conference on credit scoring and credit control, Edinburgh.

Richard J. Bolton and David J. Hand. Statistical Fraud Detection: A Review. Journal of Statistical Science, 17:235–255, 2002.

Schilling, N., Wistuba, M., Drumond, L., & Schmidt-Thieme, L. (2015, September). Hyperparameter optimization with factorized multilayer perceptrons. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 87-103). Springer, Cham.

Zareapoor, M., & Shamsolmoali, P. (2015). Application of credit card fraud detection: Based on bagging ensemble classifier. Procedia Computer Science, 48, 679-685

**WEBLIOGRAPHY**

Is It Fraud? Or New Behavior? Innovative streaming analytics anticipate customer behavior and interpret what change means, July 2013.

Available at http://www.fico.com/en/latest-thinking/white-papers/is-it-fraud-or-new-behavior (Last accessed on October 28th 2017)

https://machinelearningmastery.com/neural-networks-crash-course/
https://machinelearningmastery.com/difference-between-a-parameter-and-a-hyperparameter

(Last Accessed on October 20th 2017)

https://tryolabs.com/blog/2013/03/25/why-accuracy-alone-bad-measure-classification-tasks-and-what-we-can-do-about-it/. Last Accessed on November 2nd 2017.

http://www.damienfrancois.be/blog/files/modelperfcheatsheet, Last Accessed on November 2nd 2017.