Literature Review: Credit Card Fraud detection using Machine Learning and Deep Learning models

Subject: CM3070
Student: 190128812

Table of contents

Introduction

Frauds perpetrated in digital payments, and in particular using credit cards, are a constant threat to users and cause damages for billions of dollars every year.
It is therefore important to have a secure system that allows users to use their credit cards in absolute safety; for this reason it is necessary to have an efficient system for detecting fraud in real time.
I will analyze some research that uses different methods of Machine Learning and Deep Learning for the implementation of models that can detect fraud attempts in real time.

End-to-end neural network architecture for fraud scoring in card payments[1]

In this research, a dataset containing 900 million transactions of the BBVA bank, over a period of one and a half years, is used. A set of Artificial Neural Networks (ANN) was used to detect fraud; the first ANN is used in combination with two filters: the first one, to reduce the ratio between genuine transactions and fraud, from an initial value of 5000:1 to about 100:1, and the second one to classify the transaction as genuine or fraud.

The evaluation of the quality of this fraud detection method takes into account the costs for the company that manages the payments, and therefore also considers the transactions classified as false positives; the evaluation of the quality of this method of fraud detection takes into account the costs for the company that manages the payments, and therefore also considers the transactions classified as false positives; this is because this type of transaction also requires work by an operator who must analyze the payment operation and change its status. Furthermore, unfairly blocking user transactions because they are detected as fraud could have a cost in the trust of the payment company. However, this last data cannot be easily classified and is not the subject of the research.

The results obtained with the classification method created in this research are not clearly reported, but the authors simply mention a result similar to that of another research: Ramanathan, Paypal - fraud detection with h2o deep learning, 2015. Furthermore, even in this case there are no evaluations relating to the time needed to train the model nor to the time required for the classification of each transaction

Generative Adversarial Neural Networks based Oversampling Technique for Imbalanced Credit Card Dataset[2]

In this research, a Generative Adversarial Network (GAN) is used to overcome the problem that arises with highly unbalanced datasets. Three different machine learning algorithms were tested for transaction classification, to evaluate the best performing one in fraud detection. The use of a GAN is innovative in this context: to manage highly unbalanced datasets, different techniques are usually used, such as the Synthetic Minority Oversampling Technique (SMOTE), the Random Oversampling Technique (ROS) or the Adaprive Synthetic Sampling Approach (ADSYN).

To balance a dataset, one tries to reduce the number of records of the majority class until reaching the number of records of the minority class (undersampling) or vice versa (oversampling). The dataset of this research, containing a number of fraud records of about 0.2%, does not lend itself to undersampling techniques because it would drastically reduce the data available for training. It is therefore preferable to use oversampling techniques, through which new records are artificially created; the difficulty in this case is to create records as similar as possible to fraud records, to avoid the creation of false negatives. The creation of a GAN network in this case aims to create an oversampling model that outperforms the other most commonly used methods.

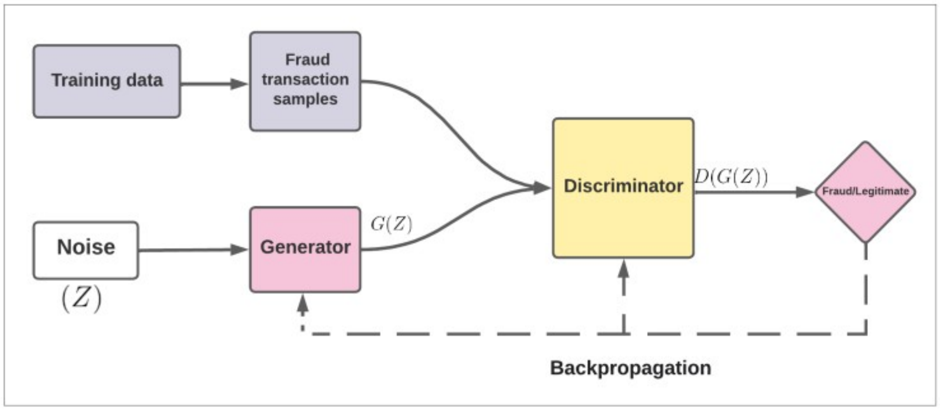


Figure 1: GAN architecture

Three different ML algorithms were used to evaluate the model: LightGBM (LBM), XGBoost (XGB), Cat-Boost (CB).
Below are the results of tests carried out with the different models created:

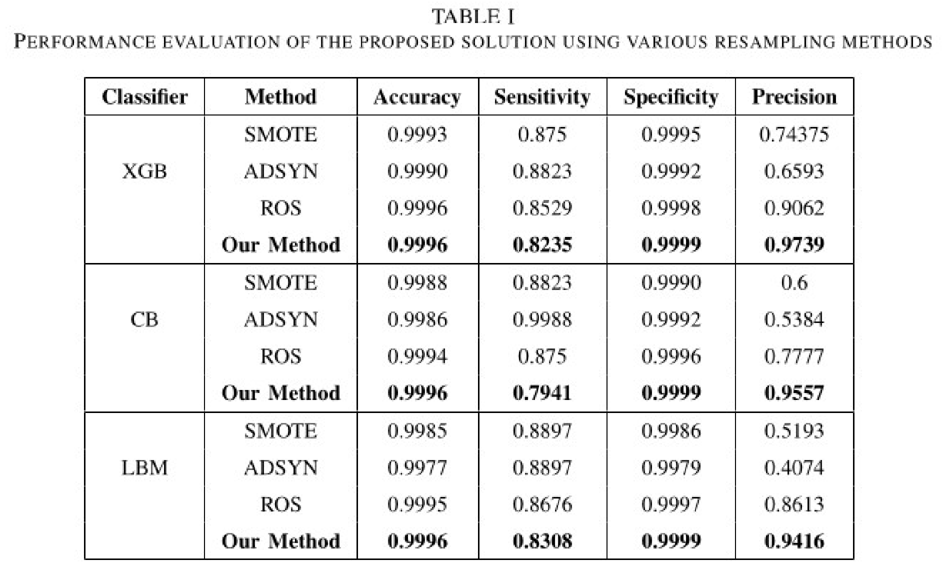


TABLE I
PERFORMANCE EVALUATION OF THE PROPOSED SOLUTION USING VARIOUS RESAMPLING METHODS

Classifier	Method	Accuracy	Sensitivity	Specificity	Precision
XGB	SMOTE	0.9993	0.875	0.9995	0.74375
	ADSYN	0.9990	0.8823	0.9992	0.6593
	ROS	0.9996	0.8529	0.9998	0.9062
	Our Method	**0.9996**	**0.8235**	**0.9999**	**0.9739**
CB	SMOTE	0.9988	0.8823	0.9990	0.6
	ADSYN	0.9986	0.9988	0.9992	0.5384
	ROS	0.9994	0.875	0.9996	0.7777
	Our Method	**0.9996**	**0.7941**	**0.9999**	**0.9557**
LBM	SMOTE	0.9985	0.8897	0.9986	0.5193
	ADSYN	0.9977	0.8897	0.9979	0.4074
	ROS	0.9995	0.8676	0.9997	0.8613
	Our Method	**0.9996**	**0.8308**	**0.9999**	**0.9416**

Figure 2: Results

The results show that the XGB algorithm applied to resampling techniques with GAN is the best in fraud detection.
It would have been interesting to also analyse the data on false positives and false negatives generated, as this is where they can determine the effectiveness of an algorithm in a real-world scenario.

Also in this research, there are no performance measures related to tranining time and real time transaction analysis.

Generative Adversarial Network for Oversampling Data in Credit Card Fraud Detection[3]

This research also uses a GAN neural network for the generation of new minority class data, to overcome the problem of the highly unbalanced dataset.

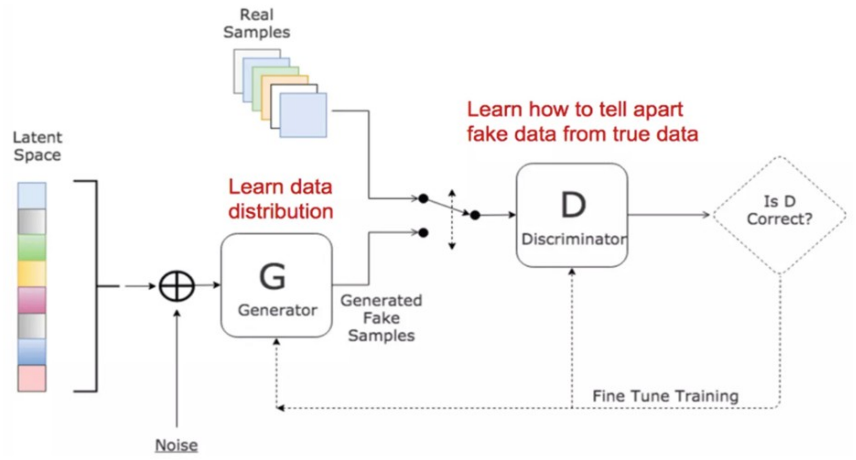


Figure 3: GAN architecture

This research also uses a GAN neural network for the generation of new minority class data, to overcome the problem of the highly unbalanced dataset.

Compared to the research illustrated above, a WGAN (Wassertein GAN) is implemented here, which makes use of the Wasserstein distance to measure the distance between two probability distributions. The aim is to demonstrate the superiority of a WGAN in generating synthetic data more similar to the originals, and thus decrease the generation of false positives, in oversampling the dataset.

The proposed model is as follows:

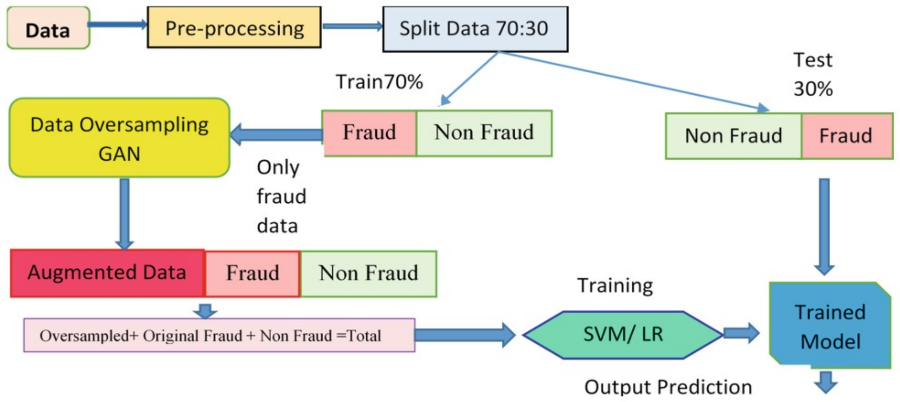


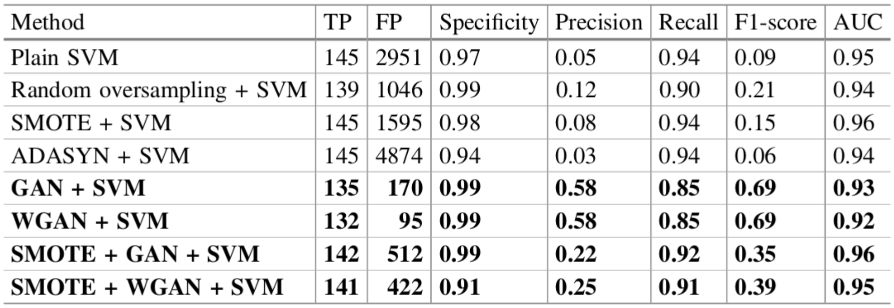
Figure 4: DNN model

The dataset is divided into two parts, training and test, containing 70% and 30% of the transactions. Then an oversampling of only the fraud data is done, using a WGAN neural network.
The architecture of the generator consists of an input layer, three hidden layers of 500 neurons each, and an output layer. The input layer receives data with random values, while the output layer generates records similar to real ones.

The Discriminator architecture consists of an input layer, two hidden layers of 500 and 300 neurons respectively, and an output layer. The output of the Discriminator is used to modify the weights of the Generator, in order to improve the veracity of the records created.

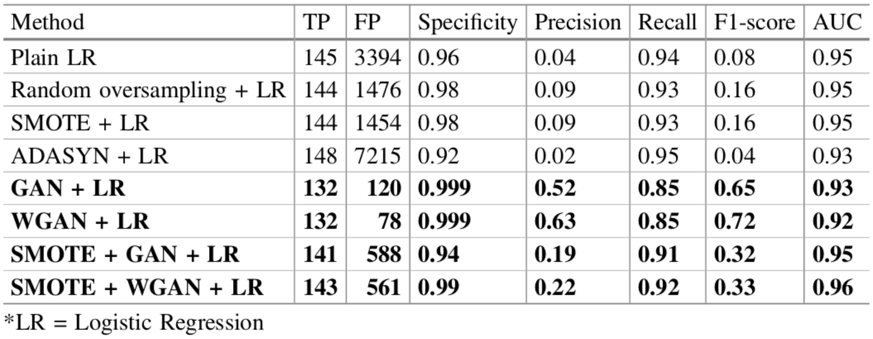
Two different classifiers using Logistic Regression and SVM were used to classify the data; no particular importance was attached to the choice of classifiers as this was not the aim of the research.

For the evaluation of the model, 15000 records were generated for each oversampling method in order to balance the dataset.
The evaluation criteria and the comparison of the results can be seen in the table below, in which the False Positive values are also analysed.



Method	TP	FP	Specificity	Precision	Recall	F1-score	AUC
Plain SVM	145	2951	0.97	0.05	0.94	0.09	0.95
Random oversampling + SVM	139	1046	0.99	0.12	0.90	0.21	0.94
SMOTE + SVM	145	1595	0.98	0.08	0.94	0.15	0.96
ADASYN + SVM	145	4874	0.94	0.03	0.94	0.06	0.94
GAN + SVM	**135**	**170**	**0.99**	**0.58**	**0.85**	**0.69**	**0.93**
WGAN + SVM	**132**	**95**	**0.99**	**0.58**	**0.85**	**0.69**	**0.92**
SMOTE + GAN + SVM	**142**	**512**	**0.99**	**0.22**	**0.92**	**0.35**	**0.96**
SMOTE + WGAN + SVM	**141**	**422**	**0.91**	**0.25**	**0.91**	**0.39**	**0.95**

Figure 5: SVM results



Method	TP	FP	Specificity	Precision	Recall	F1-score	AUC
Plain LR	145	3394	0.96	0.04	0.94	0.08	0.95
Random oversampling + LR	144	1476	0.98	0.09	0.93	0.16	0.95
SMOTE + LR	144	1454	0.98	0.09	0.93	0.16	0.95
ADASYN + LR	148	7215	0.92	0.02	0.95	0.04	0.93
GAN + LR	**132**	**120**	**0.999**	**0.52**	**0.85**	**0.65**	**0.93**
WGAN + LR	**132**	**78**	**0.999**	**0.63**	**0.85**	**0.72**	**0.92**
SMOTE + GAN + LR	**141**	**588**	**0.94**	**0.19**	**0.91**	**0.32**	**0.95**
SMOTE + WGAN + LR	**143**	**561**	**0.99**	**0.22**	**0.92**	**0.33**	**0.96**

*LR = Logistic Regression

Figure 6: LR results

The results confirm the superiority of the WGAN method in the generation of synthetic data, in comparison with the other techniques, both in the model using Logistic Regression and in the model using SVM.

A measurement of time performance in the creation of synthetic data is also briefly reported, albeit inaccurately. The estimate reports a time between 10 and 15 seconds for the generation of 15,000 records.

It would also be interesting to compare the time performance in the real time analysis of transactions.

Credit Card Fraud Detection Using Autoencoder Neural Network[4]

This research aims to implement an unsupervised deep learning model for the detection of credit card fraud.
For the realisation of the model, we want to create a Denoising Auto Encoder (DAE) that is able to minimise the ‘noise’ created by the balancing of the dataset.

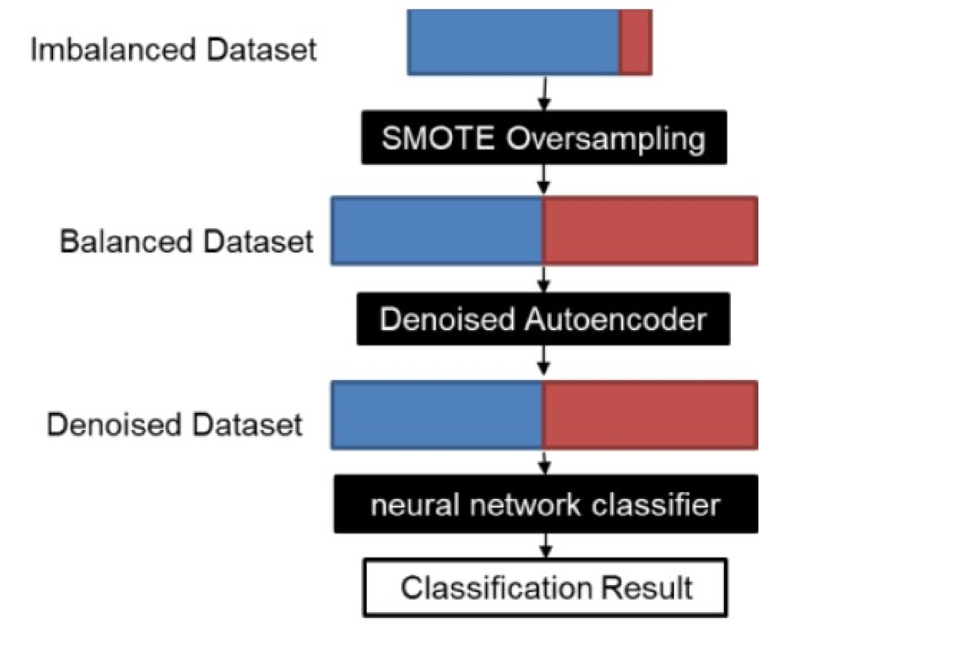


Figure 7: Model architecture

The SMOTE technique was used to balance the dataset and an AED was applied to the newly balanced dataset to clean the synthetic data generated.
A DAE is in fact able to eliminate the noise generated by the oversampling process, so that the newly generated transactions are as close as possible to the original transactions, reducing the generation of transactions classified as False Positives.

The SMOTE technique was used to balance the dataset and a DAE with 7 layers was applied to the newly balanced dataset to clean the generated synthetic data.
A DAE is in fact able to eliminate the noise generated by the oversampling process, so that the newly generated transactions are as close as possible to the original transactions, reducing the generation of transactions classified as False Positives.

For the classification of transactions, another DNN was created with 6 fully-connected layers, using SoftMax with cross-entropy as loss function.

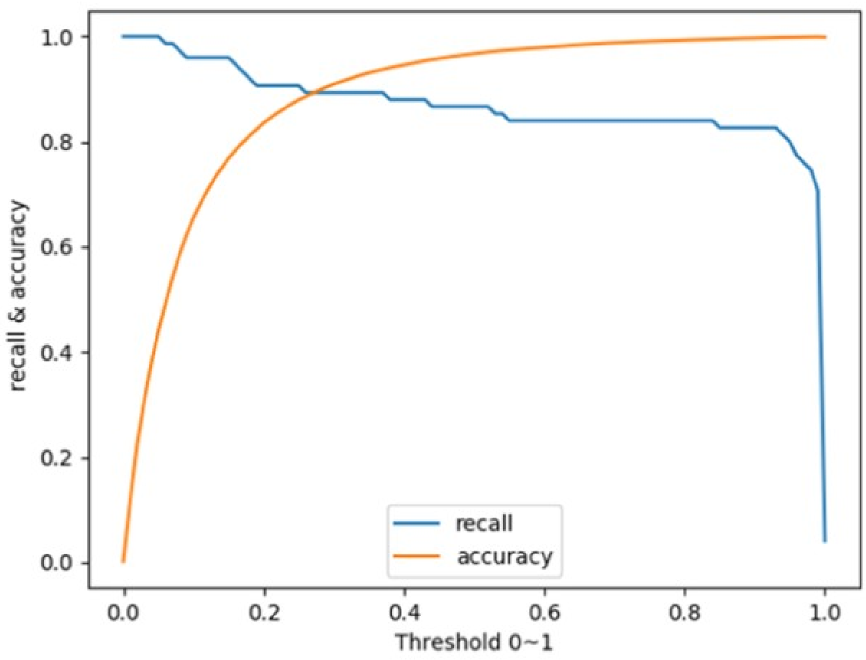


Figure 8: Recall and Accuracy results

The results show good recall (around 90% with a threshold of 0.2) and accuracy (> 97% with a threshold of 0.6).

The research reports a training time of about 10 minutes with a mid-range PC. No measurements are reported in the detection of 'live' data.

This research uses a new type of AutoEncoder for cleaning synthetic data; I will include it in my project, combining it with other techniques used in other research.

Conclusions

I consider all the research analysed to be reliable, as it can be found on sites such as IEEE or otherwise downloaded via the University of London's online library.
Each research introduces one or more innovative techniques that I will use, combining them, in my project.
The difficulty will be to replicate the results obtained, as I will have to write the code from scratch, since the research does not include it but only mentions the type of architecture and/or techniques used.

The missing part in the researches analysed is a measurement of time performance. Some research briefly mentions the time needed to train the model, but no analysis of the time needed to analyse new transactions that have to be examined in real time was ever made.
What I will do in my project is to add an analysis of the time performance of the various models created, comparing their results with those of accuracy in fraud detection.

References

[1]. Jon Ander Gómez, Juan Arévalo, Roberto Paredes, and Jordi Nin. 2018. End-to-end neural network architecture for fraud scoring in card payments. *Pattern Recognit. Lett.* 105, (April 2018), 175–181. https://doi.org/10.1016/j.patrec.2017.08.024

[2]...........Said El Kafhali and Mohammed Tayebi. 2022. Generative Adversarial Neural Networks based Oversampling Technique for Imbalanced Credit Card Dataset. In *2022 6th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI)*, December 01, 2022. IEEE, Colombo, Sri Lanka, 1–5. https://doi.org/10.1109/SLAAI-ICAI56923.2022.10002630

[3].........Akhilesh Kumar Gangwar and Vadlamani Ravi. 2019. WiP: Generative Adversarial Network for Oversampling Data in Credit Card Fraud Detection. In *Information Systems Security*, Deepak Garg, N. V. Narendra Kumar and Rudrapatna K. Shyamasundar (eds.). Springer International Publishing, Cham, 123–134. https://doi.org/10.1007/978-3-030-36945-3_7

[4] Cheng-Yuan Liou, Wei-Chen Cheng, Jiun-Wei Liou, and Daw-Ran Liou. 2014. Autoencoder for words. *Neurocomputing* 139, (September 2014), 84–96. https://doi.org/10.1016/j.neucom.2013.09.055