|  |  |
| --- | --- |
|  | |
| Scientific Research + Literature –Assessment 3  TU060 : Literature Review | |
|  | |
| Ciaran Finnegan – Part Time – First Year 2021/2022  MSc in Computer Science (Data Science)  Student No : D21124026  1/5/2022 |  |
|  |  |

Word Count: 2210

Title

**An analysis to compare Machine Learning techniques used in the application of Credit Card fraud detection from 2012-2017 against more recent optimised approaches.**

Abstract

This review describes Machine Learning techniques that were applied from 2012 to 2017 to build models to detect credit card fraud and compares them against emerging approaches documented circa 2020. The review objective is to determine if the more modern ML strategies were delivering significantly better performance, despite possible limitations due to their inherent complexities.

American and European datasets formed the basis of the modelling experiments, and ‘fraud’ is defined as the unauthorised use of card services by a third party.

A significant characteristic of the majority of these datasets is that instances of fraud make up a very small proportion of the total record set. Hence, data resampling considerations are a major factor throughout the review.

The two initial papers comprise the pre-2019 research and employed what the authors themselves described as a range of ‘traditional’ ML Classification approaches. The third paper explained a later (2019) ensemble approach to resampling and anomaly detection. The fourth paper (2020) looked at recent algorithm optimisations to avoid resampling of imbalanced data and circumvent possible data corruption. The last three papers (2020-2021) detailed experiments into more contemporary credit card fraud detection approaches using Neural Networks. Further research papers are included in this analysis to add context to the assumptions made by the various authors.

Comparing the findings across the review is challenging as authors use datasets of different sizes and a variety of model comparison criteria. However, F1 Scores, from the best performing models, show a steady increase from 0.729 with enhanced resampling and feature reduction in 2017, up to 0.849 when Neural Network approaches are employed (2020), to a score of 0.941 with an Optimised XGBoost (OXGBoost) algorithm (2020).

All metrics in this review need to be considered in the context of their research experiments, but an upward trend in fraud detection rates over time is evident.

The more recent Neural Network approaches performed well and have crucial benefits in terms of computation efficiency.

However, Neural Network predictions can be difficult to audit, although the 2021 FDIC CNN research offers solutions in this area. The OXGBoost approach has the benefit of both greater transparency and the avoidance of data corruption through resampling. Both techniques are therefore emerging areas of interest for further research.

**Key Words**: Credit Card Fraud Detection, Data Resampling, Feature Engineering, Deep Learning, Neural Network Auditing

Table of Contents

1 Introduction 5

2 Concept A: Addressing the Imbalance and Resampling Challenge 7

3 Concept B: Neural Networks – Handling Data Volume and Training Time for Fraud Detection 8

4 Concept C: Neural Networks and Auditing Fraud Detection Results 9

5 Conclusion: Better Ways to Capture CC Fraud? 10

6 References 11

# Introduction

Financial analysts such as Whatman (2019) have collated industry statistics to show that credit card fraud remains an evolving multi-billion Euro challenge**[1]**, which threatens global financial institutions with both loss of revenue and reputation.

Sinanc, Demirezen, Sağıroğlu (2021) provide a modern, but very pertinent, definition of the problem by saying that “*Credit card fraud activities occur when fraudsters exploit credit cards for personal interests without the knowledge of the cardholder and the card provider*.”**[2]**

This report begins by looking at earlier research by Bhattacharyya et al (2011) that emphasized the value of specialist algorithms for fraud detection techniques**[3]**, and which is followed by proposals from Lima and Pereira (2017) for combining more sophisticated resampling in conjunction with selective feature engineering**[4]**. Later research from Nguyen, Tahir, Abdelrazek, and Babar (2020)**[5]** posits that Deep Learning techniques may offer more promising solutions, and ones that will actually be performant enough for the ‘big data’ and real-time requirements of the 2020s. A similar conclusion is reached in 2020/21 by Anowar and Sadaoui and Sinanc, Demirezen, and Sağıroğlu, although with more specific Incremental Neural Network**[10]** and FDIC techniques**[2]** respectively. Priscilla and Prabha (2020)**[6]** also propose that resampling itself could be distorting credit card fraud data and that optimising the XGBoost algorithm can sidestep this issue for more contemporary solutions. The criterion for comparison focuses on how the different approaches handle typical fraud dataset imbalances (Section 2), the need for computation efficiency in both speed and data volumes (Section 3), and the transparency of the process used to label a transaction as possibly ‘fraudulent’ (Section 4). Obviously, the general accuracy of successive fraud classification approaches is an essential supplementary characteristic.

The historical datasets used in this literature review are sourced from credit card operators providing services in the European and American marketplaces. Therefore, any fraud patterns that are specific to other markets, such as those in Africa and Asia, will not be considered. Neither does any research in this review focus on debit card or prepaid card transactions.

Although the historical fraud datasets in this review vary in size between tens of thousands to more than 10 million rows, the incidents of actual fraud usually constitute less than 1% of any given record set. This imbalance is one of the defining characteristics of the fraud detection challenge. Researchers such as Sohony, Pratap, and Nambair (2018) have documented that such a degree of imbalance is a major complication against which learning algorithms must adapt**[7]**.

Due to data confidentiality concerns, there are still relatively few historical credit card fraud datasets upon which to conduct ML experiments for fraud detection. This is a limitation noted in research conducted by Dal Pozzolo et al (2014)**[8]**. Credit Card Fraud Detection is a very active field of research, but the lack of variety in source data requires experimentation with an increasingly sophisticated array of algorithms and parameters.

# Concept A: Addressing the Imbalance and Resampling Challenge

The heavily skewed distribution of data towards ‘non-fraud’ records in many credit card fraud datasets can pose problems for learning algorithms with a possible bias towards the majority class. The 2019 research from Ceronmani Sharmila et almakes a somewhat informal description of applications needing to capture “*fraudulent transactions that take place every once in a while*” **[9]**. In addition, Lima and Pereira (2017) reference the fact that e-commerce systems tend to generate a significant numbers of features for each transaction, and that this combines with the fraud imbalance problem to further complicate the creation of detection models**[4]**.

*Undersampling* is a technique to remove rows from the majority (non-fraud) class to address potential bias in the ML learning process. The earliest research in this review, from Bhattacharyya, Jha, Tharakunnel, and Westland in 2011, determined by experiment that random undersampling contributed to their best detection results, with an F1 Score of *0.787* with RF. However, Lima and Pereira (2017) call out that such a methodology may result in important information being lost. A more elaborate sequence of resampling techniques was proposed, including *oversampling* by artificially creating fraudulent transaction records with the SMOTE method. In these experiments, using Bayes Network and Logistic Regression algorithms, the best model F1 Score result was *0.729*, but with a larger dataset and less input data removed than in the 2011 experiments by Bhattacharyya et al.

Despite the positive outcomes in the Lima and Pereira research, later papers in this review declare that oversampling techniques may also introduce distortions in the data and cause potential overfitting. Sharmila (2019) and Priscilla and Prabha (2020) both recommend approaches that avoid resampling of credit card fraud data and instead employ enhanced Outlier Detection (F1 Score: *0.849*) and Tree Boosting (F1 Score: *0.940*) techniques respectively.

Future credit card fraud research that examines techniques to avoid data resampling will be a key future topic of interest.

# Concept B: Neural Networks – Handling Data Volume and Training Time for Fraud Detection

There is a contemporary expectation that credit card fraud detection systems can handle very high volumes of data at high speed. This is very much the keynote sentiment of separate research papers in 2020 by Nguyen et al**[5]**, and Anowar and Sadaoui**[10]**.

The 2011-2017 ML experiments in this review included datasets of 350K**[3]** and 1M**[4]**, whereas the later work by Nguyen et al**[5]** processed results on a ‘Tall Data Set’ of 10M records. Crucially, modelling failed entirely (timed out) in that 2020 research when using the more ‘traditional’ SVM and RF algorithms, commonly employed in the earlier papers. The 2020 sets of experiments**[5][10]** were able to exploit GPU computing support and refinements in the TensorFlow/Scikit-Learn libraries to dramatically reducing NN model learning times.

Computational efficiency on high volumes is only pertinent if the Deep Learning techniques are proving their effectiveness at detecting fraud. The best F1 score in the 2020 research using LSTM NN techniques was 0.*8485,* showing a small but marked improvements on the earlier 2011-2017 research.

Despite the obvious potential of Neural Network techniques, both of the 2020 papers highlight two area of potential inconsistency;

The first is that despite Anowar and Sadaoui(2020) drawing attention, in both the *Abstract* and *Conclusion* of their research paper, to the advantage of their Multi-Layer Perceptrons methods with ‘big data’, their experiments only use established datasets with 284K rows. It would be ideal to see their work repeated on datasets in the 10M+ range.

The second is that in Section 2 of this review, a warning is raised on the danger of overfitting data when resampling techniques are used**[9]**, and the results in the experiments from Nguyen et al acknowledge this tendency in their own result. Their recommendations for future research suggest a re-examination of the ensemble techniques of Sohony, Pratap, and Nambair (2018), using combined multiple NN ML classifiers**[7]**, to avoid resampling data.

# Concept C: Neural Networks and Auditing Fraud Detection Results

Deep Learning approaches promise a solution for credit card fraud detection that can address the high data volume processing demands of the 2020s. However, the inherent complexity of the models in this review, such as LSTM Recurrent Neural Networks**[5]** or Multi-Layer Perceptrons**[6]**, can be a major stumbling block. Fraud analysts in 2021 were recommending that Financial Institutions seek out fraud detection toolkits that utilise ML techniques but also “*provide evidence about why a transaction was declined or accepted*”.**[11]**

The RF algorithms in the 2020 Nguyen et al experiments may not perform as well as the LSTM models, but the decision process can be more easily exposed, whereas the NN models are effectively a ‘black box’ activity.

The 2021 ML experiments by Sinanc, Demirezen, and Sağıroğlu address this challenge by using an approach called *Fraud Detection with Image Conversion* (FDIC)**[2]**. Credit Card transactions are converted into images and classified through an extended CNN architecture. The experiments detected fraud with a strong F1 score of *0.8549*, but also provided a clear heat map visualisation of the features most dominant in predicting fraud.

Another very significant characteristic of the FDIC approach is that data is not resampled. Instead, data of any scale is converted into what is described in the research as ‘*smart data*’, which does not bias the classification process.

As mentioned in the Introduction, there is a relative scarcity of credit card fraud datasets and Sinanc et al use the same ULB data as Anowar and Sadaoui in their research during the previous year. Thus, a similar criticism can be raised that by not using a dataset with millions or rows the final analysis may be lacking.

Research in the general area of explainable credit card fraud NN solutions is still relatively threadbare but larger data volume experiments on FDIC-like solutions would appear to be a compelling next step. Of particular interest would be hardware/processing time metrics, and how the ‘smart data’ concept performs with larger datasets.

# Conclusion: Better Ways to Capture CC Fraud?

Table

Description automatically generated

*Figure 1 - key characteristics of the major papers in this review*

Many of the research papers in this review conducted multiple ML experiments with a variety of algorithms and resampling techniques. The table in *Figure 1* highlights the algorithm selections that provided the best results for the authors, and which supported their research outcomes.

**F1** is a better score for fraud detection problems, as opposed to simple accuracy, because of the uneven class distribution seen in almost all credit card datasets. This score takes the numbers of false positives and false negatives into a weighted average. Our later 2019-2021 models, using more sophisticated algorithms, are clearly performing better than the earlier ‘traditional’ approaches.

Only Nguyen et al worked with very large data volumes. A later challenge is to re-build this results table when OXGBoost and the two remaining NN approaches (MLP and FDIC) have been tested with significantly more rows of data.

Two sets of authors**[6][9]** sounded warnings about the potential for data distortion when resampling techniques are used in ML experiments. This process is hard to avoid with heavily imbalanced credit card datasets and explains why four out of the seven entries in *Figure 1* were dependent on data rebalancing. This review would suggest that there are potentially a number of fascinating avenues for future research in algorithms or ML process refinements to avoid resampling of credit card event data.

Only one paper (LTSM NN) had metrics on model training time, hence no figures are record in *Figure 1*, but further grounding and comparison of the computation efficiency of NN models for fraud detection is another rich vein of further research to explore.

# References

[1] Whatman, P., 2019. *Credit Card Statistics 2020: 65+ Facts For Europe, UK, And US*. [online] Blog.spendesk.com. Available at: <https://blog.spendesk.com/en/credit-card-statistics-2020>

[2] Sinanc, D., Demirezen, U., & Sağıroğlu, Ş. (2021). Explainable Credit Card Fraud Detection with Image Conversion. *ADCAIJ: Advances in Distributed Computing And Artificial Intelligence Journal*, *10*(1), 63-76. doi: 10.14201/adcaij20211016376

[3] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, *50*(3), 602-613. doi: 10.1016/j.dss.2010.08.008

[4] Lima, R., & Pereira, A. (2017). Feature Selection Approaches to Fraud Detection in e-Payment Systems. *Lecture Notes in Business Information Processing*, 111-126. doi: 10.1007/978-3-319-53676-7\_9

[5] Nguyen, T., Tahir, H., Abdelrazek, M., & Babar, A. (2020). Deep Learning Methods for Credit Card Fraud Detection. Retrieved 25 March 2022, from https://doi.org/10.48550/arXiv.2012.03754

[6] Priscilla, C., & Prabha, D. (2020). Influence of Optimizing XGBoost to handle Class Imbalance in Credit Card Fraud Detection. *2020 Third International Conference On Smart Systems And Inventive Technology (ICSSIT)*, pp. 1309-1315. doi: 10.1109/icssit48917.2020.9214206

[7] Sohony, I., Pratap, R., & Nambiar, U. (2018). Ensemble learning for credit card fraud detection | Proceedings of the ACM India Joint International Conference on Data Science and Management of Data. Retrieved 21 April 2022, from https://doi.org/10.1145/3152494.3156815

[8] Dal Pozzolo, A., Caelen, O., Le Borgne, Y., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems With Applications*, *41*(10), 4915-4928. doi: 10.1016/j.eswa.2014.02.026

[9] Ceronmani Sharmila, V., R., K., R., S., D., S., & R., H. (2019). Credit Card Fraud Detection Using Anomaly Techniques. *2019 1St International Conference On Innovations In Information And Communication Technology (ICIICT)*, *1*(1), 1-4. doi: 10.1109/iciict1.2019.8741421

[10] Anowar, F., & Sadaoui, S. (2020). Incremental Neural-Network Learning for Big Fraud Data. *2020 IEEE International Conference On Systems, Man, And Cybernetics (SMC)*. doi: 10.1109/smc42975.2020.9283136

[11] Miteva, R. (2021). 4 Essential Things to Look for in a Fraud Detection Solution. Retrieved 23 April 2022, from https://www.onespan.com/blog/4-essential-things-look-fraud-detection-solution