|  |  |
| --- | --- |
|  | |
| 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  1/4/2022 |  |
|  |  |

Title

**An analysis to determine if in the past decade (2012-2022) that the application of Machine Learning techniques for Credit Card Fraud detection have improved in accuracy and performance.**

Abstract

*Objectives*

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

*Methods*

ML detection Models were built based on American and European datasets, where fraud is considered to be the unauthorised use of card services by a third party.

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

The first two papers employed what the authors themselves described as ‘traditional’ ML Classification approaches. The third paper looks at an ensemble approach to resampling and anomaly detection. The fourth paper moves into the more contemporary approach of using Neural Networks. The last paper looks at recent algorithm optimisations to avoid resampling of imbalanced data and circumvent possible data corruption. Further research papers are included in the analysis to add context to the assumptions made by the various authors.

*Results*

Comparing the findings across the review is challenging as authors use datasets of varying sizes and a selection of model comparison criteria. However, F1 Scores, from the best performing models, show a steady increase from 0.767 with enhanced resampling and feature reduction, up to 0.849 when Neural Network approaches are employed, and a score of 0.941 with an Optimised XGBoost (OXGBoost) algorithm.

*Conclusion*

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

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

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

**Key Words**:

Table of Contents

1 Introduction 5

2 Concept A: Imbalance and Resampling Challenge 7

3 Concept B: Too Many Features 8

4 Concept C: Neural Networks and Auditing 9

5 Conclusion: Better ways to capture CC Fraud? 10

6 References 11

# Introduction

* *A concise* ***definition of a topic*** *under consideration (this may be a descriptive or argumentative thesis, or proposal), as well as the* ***scope*** *of the related literature being investigated. (Example: If the topic under consideration is ‘women’s wartime diaries’, the scope of the review may be limited to published or unpublished works, works in English, works from a particular location, time period, or conflict, etc.)*
* *The introduction should also note intentional* ***exclusions****. (Example: “This review will not explore the diaries of adolescent girls.”)*
* *Another purpose of the introduction is to state the* ***general findings*** *of the review (what do most of the sources conclude), and comment on the* ***availability*** *of sources in the subject area.*

Credit Card fraud remains an evolving multi-billion Euro challenge each year, in terms of lost revenue and reputation, for global Financial Institutions and their customers.

‘*Fraud’* in the domain of this literature review is defined as a historical credit card event, reported to the card provider, in which a third party has conducted a transaction (online or in person) without the permission of the card holder.

As of 2021, because of data confidentiality concerns, there are still relatively few historical credit card fraud datasets upon which to conduct Machine Learning experiments for fraud detection. In addition, fraudsters are continuing to adapt their techniques and need for real-time (or near real-time) protection continue to grow.

A further complication is that, although the record sizes in the datasets in this review vary from tens of thousands to more than 10 million, a high proportion of historical fraud datasets are also highly imbalanced, often with less than 1% of records reflecting incidents of actual fraud.

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 others 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.

Therefore, the drive in this area of research is to produce continuously more sophisticated Machine Learning models that will respond to all of these challenges and offer the best fraud detection performance.

This article looks at five major studies, conducted from 2012-2020, which recommend a series of feature engineering and algorithm selections refinements to better detect credit card fraud, and tackle the inherent imbalance challenge. The objective is to show how ongoing research in the past decade is responding to need to generate more accurate and performant fraud detection applications in this domain.

# Concept A: Imbalance and Resampling Challenge

* *There are many ways to organize the evaluation of the sources.* ***Chronological and thematic approaches*** *are each useful examples.*
* *Each work should be critically summarized and evaluated for its* ***premise****,* ***methodology****, and* ***conclusion****. It is as important to address inconsistencies, omissions, and errors, as it is to identify accuracy, depth, and relevance.*
* *Use logical connections and* ***transitions*** *to connect sources.*

# Concept B: Too Many Features

# Concept C: Neural Networks and Auditing

# Conclusion: Better ways to capture CC Fraud?

* *The conclusion* ***summarizes the key findings*** *of the review in general terms. Notable commonalities between works, whether favourable or not, may be included here.*
* *•This section is the reviewer’s opportunity to* ***justify a research proposal****. Therefore, the idea should be clearly re-stated and supported according to the findings of the review.*

The computation speed demonstrated in the Neural Network experiments has implications in terms of real-time detection applications.

This review presents compelling arguments that Neural Networks, more often used for image classification, can deliver solutions that are both more accurate and, crucially, much faster at detection credit card fraud.

Neural Network solutions can be criticised for lack of auditability in terms of classifications, and that remains a problem in this domain, but the potential is obvious.

# References

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

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

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

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

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