PROPOSAL

DETECTION OF CREDIT CARD FRAUD USING MACHINE LEARNING

# Introduction

## Background

Credit card payment system is gradually gaining the upper hand in the current society as people seek to go for everything credit card to ease their operations (Swartz, 2020). This increase has promoted extraordinary credit card fraud, which has led to massive losses of cash and erosion of the customers’ confidence. Normative systems are usually used for identification of conventional frauds which do not incorporate dynamic change in terms of their operations as are the complicit frauds (Carter, 2023). There are new and more advanced frauds that appear and it is necessary to have intelligent adaptable systems to detect frauds in real time. Due to the increasing volume of transaction data, machine learning remains a relevant tool that will enable the organization to analyse large volumes of transactions to identify several suspicious activities. Applied with both the supervised and unsupervised learning methods, ML can lower the occurrences of false positives while at the same time enhancing the rate of the detection of frauds at a short time. This study contribution can be considered major in that it seeks to apply and enhance the application of some of the machine learning techniques in the detection of fraud in credit card usage, with an aim of coming up with a safer financial system.

## Problem Statement

Credit card fraud is one of the most significant risks to the financial organizations and customers globally which continues to claim billions of finance every year (Beju & Făt, 2023). However, the application of these conventional anti-fraud tools, most of the methodologies used has the disadvantage that they work with rule-based frameworks alongside with historical trends which makes difficult to combat the advanced and fresh, not five minutes, fraudulent schemes. In addition, the credit card transaction data sets possess an imbalance nature where the instance having interests of fraud are significantly minuscule portion of all the instances and it becomes a challenging task in the detection. The existing models either do not identify frauds in real time or else pinpoint too many aberrations due to which the clean users are offended and do not trust the system any more. This study would establish that there is a need for the more intelligent, accurate and adaptive fraud detection. To solve this problem, there are machine learning algorithms that are capable of analysing patterns of the transactional data and identifying fraudulent activities with higher accuracy and effectiveness.

## Aim

To develop and evaluate a machine learning-based system for accurately detecting credit card fraud using real-world transactional data.

## Objectives

1. To analyze and preprocess credit card transaction datasets, addressing challenges such as data imbalance and noise.
2. To implement and compare the performance of various machine learning algorithms (e.g., Logistic Regression, Random Forest, XGBoost, and Neural Networks) for fraud detection.
3. To optimize the selected model(s) to improve detection accuracy and reduce false positives and false negatives.

## Research Questions

1. How effective are different machine learning algorithms in detecting fraudulent credit card transactions?
2. What preprocessing techniques can improve the performance of fraud detection models, particularly in handling imbalanced datasets?
3. Which model provides the best trade-off between fraud detection accuracy and computational efficiency?

# Literature Review

Due to the increase in the number of transactions and conformity with this kind of fraud, credit card fraud detection is considered an open field of study (Al-Hashedi & Magalingam, 2021). In the previous methods used in the detection of frauds, the ways used in detecting the frauds are preprogrammed and cannot be changed at any one point in time or even as the fraudsters adopt new ways of perpetration of the crime. As for the main characteristics and findings, ML algorithms are revealed in the recent studies as superior in detection accuracy, as well as in adaptation. For example, the study by Sahin (2020), have compared various ways how to include ensembles, for example, Random Forest and Gradient Boosting in the case of imbalanced datasets. Neural networks as well as deep learning models are also able to extract more complicated structures and nonlinear dependencies of the transaction data (Ahmed et al., 2023). Regardless this the speed is not very high but at the same time it does not report a high number of false alarms. There are many works that deal with issues such as data imbalance, feature selection and on-line processing. Such a problem is solved using methods like SMOTE (Synthetic Minority Over Sampling Technique) and cost sensitive learning. The authors of ML studies recommend the use of more models at once or integration of domain specific knowledge as it is considered that this would help to increase the result.

# Methodology

The study will employ a quantitative research methodology using machine learning technique. In the first instance of the study, it will look for a raw dataset of actual world obtained from kaggle credit card fraud dataset that consist of anonymized transaction record classified as either genuine or a fraud. The data will be pre-processed to deal with missing values, feature scaling, it also the problem of having less samples in the minority class will have to be tackled so that the models can learn from them, this will be in form of oversampling by the use of SMOTE. The main features that shall be used for purpose of detecting fraud transaction will be determined through feature selection. This study will feature multple models that will be tested that include Logistic regression, Decision trees, and Random forest etc. The model performance measures for this study will be accuracy, precision, recall, F- TEST, F1 score, and area under the curved specific to this model will be paid to the recall score to minimize on cases where frauds go unnoticed. The model generalization and optimization will then be done by applying cross validation and hyper tuning. The aim is to develop an efficient model for fraud detection on credit card transaction.

# References

Swartz, L. (2020). New money. Yale University Press. https://books.google.com.pk/books?hl=en&lr=&id=TXXgDwAAQBAJ&oi=fnd&pg=PP1&dq=Swartz,+L.+(2020).+New+money.+Yale+University+Press.&ots=PVfcaxfiQQ&sig=OOQJzxnClFuWXQTfmf6dOTpO5Kk&redir\_esc=y#v=onepage&q=Swartz%2C%20L.%20(2020).%20New%20money.%20Yale%20University%20Press.&f=false

Carter, E. (2023). Confirm not Command: Examining fraudsters’ use of language to compel victim compliance in their own exploitation. The British Journal of Criminology, 63(6), 1405-1422. https://doi.org/10.1093/bjc/azac098

Beju, D. G., & Făt, C. M. (2023). Frauds in banking system: Frauds with cards and their associated services. In Economic and financial crime, sustainability and good governance (pp. 31-52). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-34082-6\_2

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. Computer Science Review, 40, 100402. https://doi.org/10.1016/j.cosrev.2021.100402

Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences, 2(7), 1308. https://doi.org/10.1007/s42452-020-3060-1

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., ... & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review, 56(11), 13521-13617. https://doi.org/10.1007/s10462-023-10466-8