

**A machine learning approach to fraudulent payment  
detection of the payment aggregator business model in Sri  
Lanka**

S. N. Rambukkanage

198768G

Master of Science in Information Technology  
Department of Information Technology  
Faculty of Information Technology  
University of Moratuwa

June 2022

**A machine learning approach to fraudulent payment detection of the  
payment aggregator business model in Sri Lanka**

By

S. N. Rambukkanage

198768G

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfilment of the requirements of the Degree of Master of Science in Information Technology.

June 2022

# **A machine learning approach to fraudulent payment detection of the payment aggregator business model in Sri Lanka**

June 2022



*A research on detecting payments which are suspicious for fraud under card-not-present financial transactions of a special business model for payment acquiring services operating as a regulated payment aggregator in the financial technology industry in Sri Lanka*

## **Research Report**

### **IN 6900 Research Project**

By S. N. Rambukkanage

198768G

Supervisor – Dr. S. C. Premaratne

Master of Science in Information Technology  
Department of Information Technology  
Faculty of Information Technology  
University of Moratuwa

## Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of the Student:

Signature of the Student:

S. N. Rambukkanage

***UOM Verified Signature***

Date: 30/07 /2022

Supervised by:

Signature of the Supervisor:

Dr. S. C. Premaratne

***UOM Verified Signature***

Date: 30/07 /2022

## **Acknowledgement**

First of all, I intend to extend my appreciation to my supervisor, Dr. Saminda Premaratne, Senior Lecturer, Faculty of Information Technology, University of Moratuwa for his supervision and advice given to improve my research project.

Next, I would like to thank the payment service provider who was willing to extend their support for my research study.

Finally, I wish to thank and acknowledge my family and friends for the support given for me to complete my research while balancing other work.

## **Abstract**

When a payment aggregator is accepting payments on behalf of the merchants as a financial technology service provider, it is possible for a portion of those payments to be fraudulent payments. In order to detect fraudulent payments and avoid related losses, it is important to use an algorithm-based model which adapts to the changing circumstances instead of having fixed rules. This research uses machine learning concepts in order to detect a given card-not-present online transaction being a suspicious-for-fraud payment in the context of a Central Bank approved payment aggregator business model in Sri Lanka. Further to that, this research also investigates the conditions which are highly influential in deciding whether a given payment is suspicious for fraud under the payment aggregator model. Under machine learning, classification approach is used, as the dataset used is categorized as fraud and not fraud. The attributes related to the payment data are different given the context and feature engineering was required to obtain a meaningful outcome. It was discovered that the conditions which influence a payment to become suspicious for fraud are the name registered with the payment method being different to the name of the actual payee who made that particular transaction and the originating country of the transaction being different to the name of the country entered by the payee who made that particular transaction. Fourteen different supervised learning algorithms were tested on a payment dataset and were evaluated based on the accuracy of the predicted class label. As part of the outcome of this study, decision tree algorithm was identified as the most effective algorithm with the highest prediction accuracy and a model was built and saved using PyCaret for future use.

*Keywords: payment aggregator, fraudulent payments, machine learning, financial technology, card-not-present transactions*

## TABLE OF CONTENTS

ABSTRACT .....	vi
LIST OF FIGURES .....	x
LIST OF TABLES.....	xi
ABBREVIATIONS.....	xi
CHAPTER 1	
INTRODUCTION.....	1
1.1 Prolegomena .....	1
1.2 Background of the study .....	1
1.3 Problem statement.....	3
1.4 Aims and Objectives .....	3
1.5 Proposed Solution .....	4
1.6 Structure of the thesis.....	4
CHAPTER 2	
LITERATURE REVIEW.....	5
2.1 Introduction .....	5
2.2 Theoretical Background.....	5
2.3 Payment Aggregator business model .....	5
2.4 Fraudulent Payments .....	6
2.5 Approaches in fraudulent payment detection .....	6
2.6 Empirical Literature and findings.....	8
2.7 Machine Learning applications in fraudulent payment detection.....	9
2.8 Research problem identification .....	13
2.9 Summary .....	13
CHAPTER 3	
TECHNOLOGY ADAPTED.....	14
3.1 Introduction .....	14
3.2 Machine Learning Approach .....	14
3.3 Supervised Learning .....	16
3.4 Classification Algorithms .....	17
3.5 Environment setup with Google Colab and PyCaret .....	19
3.6 Summary .....	19
CHAPTER 4	
RESEARCH METHODOLOGY.....	20

4.1 Introduction.....	20
4.2 Hypothesis.....	20
4.3 Input.....	20
4.4 Output.....	20
4.5 Process.....	21
4.6 Data collection.....	21
4.7 Data pre-processing .....	21
4.8 Attribute selection .....	21
4.9 Machine Learning Approach.....	23
4.10 Features.....	23
4.11 Users.....	23
4.12 Evaluation .....	24
4.13 Summary .....	24
CHAPTER 5	
RESEARCH DESIGN AND ANALYSIS.....	25
5.1 Introduction.....	25
5.2 Research Design.....	25
5.3 Detailed Research Design.....	26
5.3.1 Sub Research Question One.....	26
5.3.2 Sub Research Question Two.....	26
5.4 Summary.....	26
CHAPTER 6	
IMPLEMENTATION.....	27
6.1 Introduction.....	27
6.2 Solution for Research Question Once.....	27
6.3 Solution for Research Question Two.....	27
6.4 Data Collection and Pre-processing.....	27
6.5 Exploratory Data Analysis.....	30
6.6 Feature Engineering and Feature Selection.....	31
6.7 Algorithm Selection and Model Comparison.....	33
6.8 Selecting the best model.....	35
6.9 Model Creation.....	35
6.10 Tuning hyperparameters.....	36
6.11 Prediction accuracy on test data.....	36
6.12 Model finalization.....	36
6.13 Prediction accuracy on unseen data.....	37
6.14 Saving the model.....	37
6.15 Loading the saved model for future predictions.....	38
6.16 Summary.....	38



CHAPTER 7

EVALUATION..... 39

7.1 Introduction.....39

7.2 Evaluating the best model..... 39

7.3 Performance Metrics of the best model..... 39

7.4 Summary..... 43

CHAPTER 8

CONCLUSION AND FUTURE WORK..... 44

8.1 Introduction.....44

8.2 Overview of the research..... 44

8.3 Limitations..... 45

8.4 Future work and further development.....45

8.5 Summary..... 45

REFERENCES.....46

APPENDICES

Appendix A..... 50

Appendix B..... 56

Appendix C..... 57

## LIST OF FIGURES

Fig. 2.1 Mediated model with payment aggregator.....	6
Fig. 2.2 End-to-end non-bank payment aggregator transactional flow.....	8
Fig. 2.3 Process of the research.....	9
Fig. 2.4 Data mapping.....	9
Fig. 2.5 The parties and the actions involved in a 3DS 2.0 transaction process.....	10
Fig. 2.6 Game theoretic fraud detector.....	11
Fig. 2.7 Process of detecting credit card frauds.....	11
Fig. 2.8 Model Selection and Performance Metrics.....	12
Fig. 3.1 Machine learning framework for payment fraud detection.....	16
Fig. 6.1 Correlation Matrix.....	31
Fig. 6.2 Feature Importance Plot.....	32
Fig. 6.3 Best Model – ROC curve.....	35
Fig. 7.1 ROC/Area under the Curve - decision tree classifier.....	40
Fig. 7.2 Confusion Matrix - decision tree classifier.....	40
Fig. 7.3 Discrimination Threshold - decision tree classifier.....	41
Fig. 7.4 Precision Recall Curve - decision tree classifier.....	41
Fig. 7.5 Class Prediction Error - decision tree classifier.....	42
Fig. 7.6 Classification Report - decision tree classifier.....	42
Fig. 7.7 Learning Curve - decision tree classifier.....	42
Fig. 7.8 Calibration Curve - decision tree classifier.....	43
Fig. 10.1 GPU – Google Colab.....	55
Fig. 10.2 Pipeline Plot.....	57
Fig. 10.3 KS Statistic Plot.....	57
Fig. 10.4 Manifold Learning.....	58

## LIST OF TABLES

Table. 6.1 Null Value Check.....	28
Table. 6.2 PyCaret - Setting up the environment.....	29
Table. 6.3 Dataset at a glance.....	30
Table. 6.4 Attribute Information.....	30
Table. 6.5 Count, mean and standard deviation.....	31
Table. 6.6 Algorithms Selected.....	33
Table. 6.7 Comparing models.....	34
Table. 6.8 Test Data Prediction.....	36
Table. 6.9 Unseen Data Prediction.....	37
Table. 7.1 Comparing Model Accuracy.....	39
Table. 7.2 Model Statistics - decision tree classifier.....	40
Table. 10.1 View the dataset.....	56
Table. 10.2 Prediction result on unseen data.....	56
Table. 10.3 Hyperparameters.....	57

## ABBREVIATIONS

AVS: Address Verification Systems

Dtype: Data type

PA: Payment Aggregators

SVM: Support Vector Machine

Std: Standard Deviation

MCC: Matthew's Correlation Coefficient