

Data Mining for Fraud Detection in Pawning Services offered by Banks

D.T.D Gamage
179461A

Faculty of Information Technology
University of Moratuwa

2020

Data Mining for Fraud Detection in Pawning Services offered by Banks

D. T. D. Gamage

179461A

Supervisor: Mr. Saminda Premarathna

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

June 2020

Declaration

We declare that is our 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 Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.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 Student

Signature of Student

D.T.D Gamage

.....

Date: 2020-06-09

Supervised by

Name of Supervisor

Signature of Supervisor

Mr. S.C. Premarathne

.....

Date:

Acknowledgements

First and foremost, I would like to express my sincere gratitude towards my supervisor, Mr. Saminda Premarathne, Senior Lecturer, Faculty of Information Technology, University of Moratuwa for his guidance, supervision, advices and sparing valuable time thorough the research project.

Furthermore, big thank should goes to Prof. Mohamad Ferdihous who taught Research Methodology and Literature Review and thesis writing subjects which were the basis for this research work.

I would also like to thank my family specially my father for the support provided me through the completion of this project and all the batch mates of the M.Sc. in IT degree program and my office mates who gave their valuable feedbacks to improve the results of the research. I must acknowledge Mr. M.C Rajapaksha, my boss, here for the support given me to collect data for this research and my future husband for the support and encouragement given throughout this project.

ABSTRACT

Pawning or gold-pledged loans has been popular instrument/service among the various products provided by banks and other financial institutions in Sri Lanka in recent decades. Major reason for pawning becoming that much popular is gold items is one of the most reliable sources of credit for lower and middle cast households in Sri Lanka. With the increasing demand for pawning services offered by financial institutions exposure for frauds become incremental accordingly. With the increasing market value of gold people who are seeking for chances to do a fraud are activated and find various methods to achieve their goals. This Paper attempts to explore how data mining techniques can be used in detecting the fraudulent Transactions with related to pawning services offered by banking sector in Sri Lanka. This descriptive analysis on discovering frauds in pawning services will analyses the main parameters associated with pawning and how they can be effectively use for detecting fraudulent transactions. Finally, at the end this research will emphasizes how data mining techniques can be used for detecting fraudulent transactions by providing a solution for pawning fraud detection based on classification model.

Keywords— Pawning fraud, Banking fraud, fraud detection, Data Mining

Table of Contents

Declaration	ii
ABSTRACT	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
1 CHAPTER ONE	1
INTRODUCTION	1
1.1 Prolegomena	1
1.2 Background of the Study	2
1.3 Problem Statement	3
1.4 Aims and Objectives of the research.....	3
1.5 Proposed Solution	4
1.6 Structure of the thesis	5
2 CHAPTER TWO	6
REVIEW OF LITERATURE	6
2.1 Introduction	6
2.2 Pawning procedures in banking sector.....	6
2.3 Pawning Frauds	7
2.4 Traditional pawning fraud detection procedures in banking sector.....	7
2.5 Issues with existing pawning fraud detection procedures in banking sector	8
2.6 Data mining for fraud detection History	9
2.7 Data Mining Applications in Use for Fraud Detection	13
2.8 Research problem identification	14
2.9 Summary	14
3 CHAPTER THREE	15
TECHNOLOGY ADAPTED IN PAWNING FRAUD DETECTION.....	15
3.1 Introduction	15
3.2 What is the data mining?.....	15
3.3 Reasons for choosing classification for pawning fraud detection	17
3.4 Why Classification?.....	17
3.4.1 Supervised Learning algorithms	18
3.5 Classification Algorithms.....	19

3.6	Microsoft SQL SERVER	23
3.7	Rapid Minor Studio	23
3.8	Summary.....	23
4	CHAPTER FOUR.....	24
	METHODOLOGY FOR DETECTING AND PREDICTING FRAUDS IN PAWNING SERVICES	24
4.1	Introduction	24
4.2	Hypothesis	24
4.3	Input	24
4.4	Output.....	25
4.5	Process.....	25
4.5.1	Data collection and selection.....	25
4.5.2	Data Pre-processing	26
4.5.3	Attribute Selection	26
4.6	Data mining	27
4.7	Users	27
4.8	Features.....	27
4.9	Summary.....	28
5	CHAPTER FIVE.....	29
	RESEARCH DESIGN AND ANALYSIS	29
5.1	Introduction	29
5.2	Research Design	29
5.3	Detailed Research Design	30
5.3.1	Sub Research Question One	30
5.3.2	Sub Research Question Two	30
5.3.3	Sub Research Question 3.....	30
5.4	Summary	30
6	CHAPTER SIX.....	31
6.1	Introduction	31
6.2	Solution for Research Question One.....	31
6.3	Solutions for Research Question Two	31
6.3.1	Data Preprocessing.....	31

6.3.2	Attribute Selection	32
6.3.3	Correlation Matrix	33
6.3.4	Feature Weights	34
6.4	Solution for Sub Research Question 3	35
6.4.1	Model Creation	35
6.5	Summary	36
7	CHAPTER SEVEN	37
7.1	Introduction	37
7.1.1	Evaluation for Classification	37
7.1.2	Confusion Matrix	37
7.2	Evaluation Results	39
7.3	Summary	44
8	CHAPTER EIGHT	45
	CONCLUSION AND FUTURE WORK	45
8.1	Introduction	45
8.2	Overview of the research	45
8.3	Limitations	46
8.4	Future work of the project	46
8.5	Summary	46
	REFERENCES	47
9	Appendix A	49
	Data Preprocessing	49
10	Appendix B	51
	Attribute Selection	51
11	Appendix C	53
	Model Creation	53
12	Appendix D	55
	Evaluation Results	55
	Results for Correlation Matrix Attributes	55
	Results for Information gain attributes	61

LIST OF TABLES

Table 6-1 Attributes selected by correlation matrix.....	33
Table 6-2 Attributes selected by weight by Information Gain	34
Table 7-1 Classification Evaluation Measurements.....	38
Table 7-2 Comparison of classification techniques dataset1 cross validation	39
Table 7-3 Comparison of classification techniques_dataset1 split validation.....	39
Table 7-4 Comparison of classification techniques dataset2 cross validation	40
Table 7-5 Comparison of classification techniques dataset2 split validation.....	40

LIST OF FIGURES

Figure 2.1 Fraud detection technique.....	11
Figure 3.1 KDD Process in datamining.....	16
Figure 3.2 supervised learning techniques.....	18
Figure 3.3 Structure of Artificial Neural networks.....	22
Figure 6.1 Selecting and removing outlier values.....	49
Figure 6.2 Converting class variable type to nominal.....	49
Figure 6.3 Generate new categorized attributes from the existing attributes 1.....	50
Figure 6.4 Generate new categorized attributes from the existing attributes 2.....	50
Figure 6.5 Attribute selection using correlation matrix.....	51
Figure 6.6 Attribute selection using correlation matrix results.....	51
Figure 6.7 Attribute selection using weight by information gain.....	52
Figure 6.8 Attribute selection using weight by information gain results	52
Figure 6.9 Model creation cross validation.....	53
Figure 6.10 Model creation cross validation parameters.....	53
Figure 6.11 Model creation split validation.....	54
Figure 6.12 Model creation split validation parameters.....	54
Figure 7.1 Confusion matrix.....	38
Figure 7.2 Cross validation Bagging	55
Figure 7.3 Cross validation Decision tree	55
Figure 7.4 Cross validation KNN.....	56
Figure 7.5 Cross validation Naive Bayes.....	56
Figure 7.6 Cross validation Random Forest.....	57
Figure 7.6 Cross validation Random Forest.....	57
Figure 7.8 Split validation Decision tree	58
Figure 7.9 Split validation KNN.....	58
Figure 7.10 Split validation Naive Bayes	59

Figure 7.11 Split validation Random Forest.....	59
Figure 7.12 Cross validation Bagging.....	60
Figure 7.13 Cross validation Decision tree.....	60
Figure 7.14 Cross validation KNN.....	61
Figure 7.15 Cross validation Naive Bayes.....	61
Figure 7.16 Cross validation Random Forest.....	62
Figure 7.17 Split validation Bagging.....	62
Figure 7.18 Split validation Decision Tree.....	63
Figure 7.19 Split validation KNN.....	63
Figure 7.20 Split validation Naive Bayes.....	64
Figure 7.21 Split validation Random Forest.....	64
Figure 7.22 Algorithm Performance comparison for Dataset 1 –Accuracy.....	41
Figure 7.23 Algorithm Performance comparison for Dataset 2 –Accuracy.....	42
Figure 7.24 Algorithm Performance comparison for Dataset 1 –Precision	42
Figure 7.25 Algorithm Performance comparison for Dataset 2 –Precision	43
Figure 7.26 Algorithm Performance comparison for Dataset 1 –Recall	43
Figure 7.27 Algorithm Performance comparison for Dataset 2 –Recall.....	44