

COMBATING FRAUD BY LEVERAGING MACHINE LEARNING AND HUMAN BEHAVIOR

Casey Astiz

Adviser: Professor Michael Linderman

A Thesis

Presented to the Faculty of the Computer Science Department
of Middlebury College

in Partial Fulfillment of the Requirements for the Degree of
Bachelor of Arts

December 2018

ABSTRACT

I am writing about fraud and how to detect it! Wooh!

ACKNOWLEDGEMENTS

Thank you Professor Linderman and Professor Scharstein!!

TABLE OF CONTENTS

1	Introduction	1
2	Modern Day Fraud	3
2.1	Credit Card Fraud	3
2.1.1	Online Payments Fraud	3
2.2	Current Research	3
2.3	The Human Element	3
3	Fraud in Context	4
3.1	Interviews from Professionals	4
3.2	Research Versus Reality	4
4	Implementing a Fraud Detection System	5
4.1	Model	5
4.2	Experiment	5
4.2.1	Hypothesis	5
4.2.2	Data	5
4.2.3	Methods	5
4.2.4	Results	5
4.2.5	Discussion	5
5	Further Applications	6
5.1	Anomaly Detection	6
5.1.1	Cell Phone Fraud	6
5.1.2	Biometric Fraud	6
5.1.3	Bioinformatics	6
6	Conclusions	7
	Bibliography	8

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In today's world, there are more and more ways to pay for goods and services. Cash is a way of the past; mobile, peer to peer, and online payments are the way of the future. While these convenient payment methods are very exciting, the additional methods of payment lead to more gaps in security, in particular fraud. Fraudsters have taken advantage of every system in some form or another, and it is the financial institution's job to block or detect these fraudsters or else deal with the payments. For reference, credit card fraud through online payments in Australia hit \$476 million in 2017, rising \$58 million from the year before [2].

The main purpose of this thesis is to investigate current methods of fraud detection and implement one or more of those methods in a financial context. I will discuss the most prevalent methods for fraud detection are currently, as well as past fraud detection methods. Fraud detection is a difficult field of study because it falls in the subset of anomaly detection and has very skewed classes as more transactions are not fraud than are fraud, and the classification of fraud is always changing. In order to build an effective fraud detector, it must be adaptive to new patterns and categories of fraud that are developing as fraudsters find new methods to deceive the system.

The main focus of this thesis will be a review of different methods used in the industry in the past and today for fraud detection, as well as my own implementation using some dataset of transactions to be determined. This background section will include information from published research papers as well as information I learned from interviewing people at different companies in this field to provide some context for the literature review. Comparing business interactions with fraud to published research in the field will allow me to somewhat analyze the performance of different machine learning methods in real life versus a research context.

Chapter 2 discusses past and current research into the field of fraud detection. Chapter 3 gives context for the problem at hand, synthesizing the information I learned from interviewing professionals in the fraud world. In Chapter 4, I describe the methodology and experiments I ran.

This thesis will outline the problem at hand, describe current solutions to the problem, explain an implementation of a solution, provide context with some interviews of experts in the field, and conclude.

CHAPTER 2

MODERN DAY FRAUD

The text for Section 1 goes here.

2.1 Credit Card Fraud

2.1.1 Online Payments Fraud

2.2 Current Research

2.3 The Human Element

In this section I will discuss some game theory and behavioral elements of fraud and ways of eliminating frauds and fraudsters.

CHAPTER 3

FRAUD IN CONTEXT

3.1 Interviews from Professionals

This section will be about what professionals are doing in the fraud field.

3.2 Research Versus Reality

This section I will compare whatever I find out from professionals to the most modern research on this topic. I assume research will be slightly behind what these professionals are dealing with and how they are solving these problems.

CHAPTER 4

IMPLEMENTING A FRAUD DETECTION SYSTEM

4.1 Model

This section is where I will talk about the overall system I am planning on implementing.

4.2 Experiment

4.2.1 Hypothesis

4.2.2 Data

4.2.3 Methods

LOTS of Machine Learning incorporated into bigger fraud detection and rule based learning systems! Overall system will be discussed in Model section.

4.2.4 Results

This section will be full of tables and figures. There will probably also be an appendix with the entire set of experiments run. The most interesting or a subset of the experimental runs will be in a table in this section.

4.2.5 Discussion

This section will compare my hypothesis to the results, as well as my results to the model I'm basing it off of. I also plan on adding any notes I learn along the way of this implementation.

CHAPTER 5
FURTHER APPLICATIONS

5.1 Anomaly Detection

5.1.1 Cell Phone Fraud

5.1.2 Biometric Fraud

5.1.3 Bioinformatics

CHAPTER 6
CONCLUSIONS

BIBLIOGRAPHY

- [1] Mac Wang. How to protect your business from online payments fraud - CSO — The Resource for Data Security Executives, 2018.