

CREDIT CARD FRAUD DETECTION

Thesis submitted in partial fulfillment of the
requirements for

Postgraduate diploma in data science

By

Jothi Prakash Anandan

(Reg no: 17125760069)

Under the guidance of

Ramakrishna Ganesan

Faculty

Manipal Prolearn

Bangalore, Karnataka



CREDIT CARD FRAUD DETECTION

Thesis submitted in partial fulfillment of the
requirements for

Postgraduate diploma in data science

By

Jothi Prakash Anandan

(Reg no: 17125760069)

Examiner 1

Examiner 2

Signature:

Signature:

Name:

Name:

CERTIFICATE

This is to certify that the project titled

CREDIT CARD FRAUD DETECTION

Is a bonafide record of the work done by

Jothi Prakash Anandan

Reg no: 17125760069

In partial fulfillment of the requirements for the award of **Postgraduate Diploma in data science** under Manipal University, Manipal and the same has not been submitted elsewhere for any kind of certification/recognition.

Ramakrishna Ganesan

Faculty

Manipal Prolearn

Manipal University

ACKNOWLEDGMENT

I would like to express my sincere gratitude to the Director of Manipal global academy of data science, **Dr. Ramesh Babu** for having given me the opportunity to work on the project titled “**Credit card fraud detection**”.

I would like to thank my project guide, **Mr. Ramakrishna Ganesan** for guiding me throughout this project term, providing valuable feedback and advice needed for successful completion of this project.

This project helped me research and learn important aspects of how credit defaults happen and also helped me sharpen my data analysis skills by working on this real business case.

2. TABLE OF CONTENTS

1. Acknowledgment
2. Table of contents
3. Abstract
4. Introduction
 - a. Motivation
 - b. Project scope
 - c. Project focus
5. Project description
 - a. Business and domain understanding
 - b. Datasets understanding
 - c. Data limitations
 - d. Business objectives
6. Tools and technologies
 - a. RStudio
7. Exploratory data analysis
 - a. Data exploration
 - b. Data cleaning and transformation
8. Modeling
 - a. Selection of model/technique
 - b. Challenges faced
 - c. Evaluation
 - d. Validation metrics
 - e. Variable importance and model interpretation
9. Conclusion
 - a. Summary of project outcome
 - b. Future works
10. References

3. ABSTRACT

In this project, machine learning, network analysis and UI building tools and techniques were used to analyze credit defaulting from credit card transactions in September 2013 by European cardholders. The dataset has been collected and analyzed during a research collaboration of Worldline and the Machine Learning Group (<http://mlg.ulb.ac.be>) of ULB (Université Libre de Bruxelles) on big data mining and fraud detection.

This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, they cannot provide the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Further, various machine learning and visualization techniques were leveraged on the data to find defaults.

4. INTRODUCTION

a. Motivation

The health of the credit card industry is best measured not by the number of people with cards, but rather the number who pay their bills. Bad payment habits begin by nicking you with more fees and lower credit scores, and, in advanced cases, can lead to the loss of a vehicle or home, garnishment, and bankruptcy.

Credit card defaults, after a lengthy decline, are starting to tick up slightly.

Delinquencies in bank cards rose in the third quarter of 2016, but remain near historical lows, according to the American Bankers Association's Consumer Credit Delinquency Bulletin.

Bank card delinquencies increased 26 basis points to 2.74 percent of all accounts in the third quarter, but remain well below their 15-year average of 3.68 percent. The ABA report defines a delinquency as a late payment that is 30 days or more overdue.⁴

The trend looks similar when examining accounts that have been overdue for three months. TransUnion's Industry Insights Report found that the credit card delinquency rate reached 1.79 percent in Q4 2016, an increase of 12.6 percent from 1.59 percent in Q4 2015. The credit card delinquency rate remains more than a full point below its peak in Q4 2009 (2.97 percent).

Credit card fraud detection

b. Project scope