

Literature Review: Credit card fraudulent detection using data analysis

Darab Qasimi
Daqasimi18@earlham.edu
Earlham College
Richmond, Indiana, USA

ABSTRACT

Credit card fraudulent detection using data analysis is a capstone project that'll be written in form of a program to detect activities of credit cards. The project will rely on online resources such as research papers, thesis papers, blogs, and other sources of analysis which will be used for the program and analysis side of my project.

ACM Reference Format:

Darab Qasimi. 2022. Literature Review: Credit card fraudulent detection using data analysis. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Millions of credit cards have been stolen or hacked worldwide in the past few years. Most hacked or stolen credit cards have been used for fraudulent or illegal activities. Given that credit card fraud is a pressing issue in most societies, the Data Science Capstone Project will be used as an opportunity to study the likelihood of a credit card being used for fraudulent activities. Data modeling and data analysis will be used to find the best fit line between the number of transactions and the number of fraud activities in the dataset. I'll also look at what the trend of credit cards being used for fraudulent activities look like in the future. I'll also investigate whether other variables correlate with the ratio of stolen credit cards?

2 COMMON FRAUDULENT TRENDS

To identify the trend of credit card fraud activities we must understand the factors that affect fraud. One of the resources that will be used in the research paper is a thesis by Ayorinde [2] that looks at the common trends in credit card fraud in the banking, retail, financial services, and healthcare industries. Credit cards that are stolen or misplaced, synthetic fraud, data breaches, mail interception, skimming, and merchant collusion are common examples of ways credit cards are hacked or used for fraudulent activities. The thesis has used machine learning models to classify fraudulent transactions on a dataset taken from Kaggle which is a simulated credit card transaction that contains legit fraudulent transactions. At the end of the thesis, the writer reports the accuracy. For the

analysis part of my capstone project I'll use some of the concepts and methods that are used in the research paper such as Random Forest Decision Tree algorithm. Random Forest Decision Tree will be used as the main algorithm for detecting fraudulent activities of credit card transactions.

3 DECISION TREE

The decision tree will be used for analysing the credit card fraud dataset. In an online blog on DataFlair called Detect Credit Card Fraud with Machine Learning in R [1], many algorithms are used in R programming language to detect fraudulent credit card activity. The blog has a complete set of instructions and implementations of credit card fraud detection algorithms. The program has used algorithms and methods such as Decision Trees, Logistic Regression, Artificial Neural Networks, and Gradient Boosting Classifier for data analysis. The blog has used the Card Transactions dataset that's available on google drive (there is a link to the dataset in the cited website), which contains a mix of fraudulent and non-fraudulent transactions in order to have an unbiased result. Sometimes some datasets can be unstructured, and for that purpose, the blog has used scale in data to standardize the dataset. Standardizing data helps in specifying a specific range of data or values and getting rid of outliers in the dataset. I might need to use the mentioned blog as a source to learn some programming skills for the implementation part of the capstone project along with using it to learn how fraud is detected in credit cards.

4 DATA NORMALIZATION

Data normalization is part of analyzing and producing a result that's close to the reality. In Sharma's [6] thesis the paper has applied data normalization as part of credit card fraud detection. For the experiment, the thesis paper used a dataset of transactions made by European cardholders during July 2015 through various credit cards. The dataset has been collected during a research collaboration with Université Libre de Bruxelles. Before data analysis the data in the dataset is scaled so that the overall dataset is ready for a standardized modeling and analysis. Part of the analysis for the thesis is done using the Random Forest Decision Algorithm for classification and regression. Random Forest Decision Algorithm is a tree-like model of decision with many possible outcomes. The results of many outcomes from the decision trees are then aggregated into one result which in my research is detecting if a credit card activity is fraudulent or not. Random Forest Decision Algorithm is also used to correct the habit of over-fitting training sets, and a subset of the training set is randomly sampled to train each single decision tree, and finally a decision tree is built. The research paper is a good example of how I want to conduct my research

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2022 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM... \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

through categorizing data and feeding different branches of the Random Forest Decision Algorithms with different parts of dataset and finally producing a predictive detection of credit card fraud.

5 MACHINE LEARNING

In Sulaiman et al's research [4], machine learning is an effective way of determining which transactions might be fraudulent and which are not. Sulaiman et al have used the Random Forest algorithm for constructing a a decision tree for training and machine learning purposes. In the research paper, the Random Forest decision tree is described as a slow algorithm in real time fraud analysis. The artificial Neural Network (ANN) Method is seen as a convenient algorithm for being an unsupervised method for predicting fraud. ANN and RFDT find patterns in detecting fraudulent activities in credit cards, but ANN seems to be an effective solution. Although RFDT is seen as a weak algorithm for performing fraudulent activities of credit cards I believe RFDT is an effective way for producing a result that is aggregated from many other decision trees. For the purpose of my research paper RFDT will be an appropriate way to analyze my intended dataset. Since Sulaiman et al's research paper has analyzed credit card fraudulent activities with two different algorithms each has different time complexity and space complexity. Before starting the implementation of the project I might analyze with ANN and RFDT to determine what time and space complexity I should be expecting for my program.

6 SUPERVISED AND UNSUPERVISED

There are two approaches of detecting credit cards frauds; supervised and unsupervised. For the purpose of capstone project I'm interested in supervised credit card fraud detection. Mekterović [5] has looked at detecting fraudulent activities of credit cards when cards are not present during transactions. The research has taken two approaches to predict fraudulent activities of credit cards; the first approach is supervised data analysis, and the second approach is unsupervised data analysis. The novel dataset has used 197,471 transactions from three months. The transactions are divided chronologically to achieve a realistic scenario when an actual fraudulent credit card activity happens. The dataset has also been modeled visually, where it's easier to see the data flow. On the programming side of the capstone project I'll be building a supervised program and machine learning concept for detecting credit card fraud. For the given purpose, during implementation of my project I might be referring to the cited data and use it for help and guidance specifically for supervised method.

7 TRADITIONAL AND NEW FRAUD DETECTION METHODS

To detect credit card fraud an essential part of the research paper will be understanding the use of machine learning. Credit Card Fraud Detection With Machine Learning [3] is an online blog which elaborates on the concept of credit card fraud and the types of fraud. The blog analyzes why the traditional credit card fraud detection methods such as detecting by location don't work and why machine learning is a way to reduce credit card fraudulence. The blog has stated that greater accuracy, less manual work, fewer declines of normal transactions, and the ability to adapt and evolve machine

learning elements have helped reduce credit card fraud. The blog contains facts and data about credit card fraudulence and how we can use that information to predict it. The blog is a good tutorial which will be used as a source to learn about machine learning and why the traditional way of detecting credit card fraudulence didn't work, which can help take a better decision in what method I want to use in my capstone project for detecting fraud in credit cards.

8 CONCLUSION

Detecting credit card activities is an area where deep analysis is required. For the purpose of familiarizing with the methods of analyzing fraud activities and analyzing datasets the above mentioned online resources will be used as reference and tutorial. The aim is to study the supervised approach of detecting or analyzing credit card frauds and implement it in program that can help people avoid credit card fraudulence.

REFERENCES

- [1] 2022. Data Science Project – Detect Credit Card Fraud with Machine Learning in R.
- [2] Kayode Ayorinde. 2021. *A Methodology for Detecting Credit Card Fraud*. Ph.D. Dissertation. Minnesota State University, Mankato.
- [3] Kirsten Barkved. 2022. Credit Card Fraud Detection With Machine Learning. (2022).
- [4] Rejwan Bin Sulaiman, Vitaly Schetinin, and Paul Sant. 2022. Review of Machine Learning Approach on Credit Card Fraud Detection. *Human-Centric Intelligent Systems* (2022), 1–14.
- [5] Igor Mekterović, Mladen Karan, Damir Pintar, and Ljiljana Brkić. 2021. Credit card fraud detection in card-not-present transactions: Where to invest? *Applied Sciences* 11, 15 (2021), 6766.
- [6] Nishant Sharma. 2020. Credit Card Fraud Detection Predictive Modeling. (2020).