Credit Card Fraud Detection using Data Mining Techniques: Critical Review Study

Adel Habib Abdulaziz
Information Science Dept.– College of Arts and Humanities
King Abdulaziz University - Jeddah - KSA

Abstract. Nowadays due to the revolution of Information Technology in all aspects of our life, Electronic Payments is becoming a raising trend. With the global direction and tremendous increase of online transactions for example, the use of credit cards has significantly increased. According to the increase of online payments and transactions, fraud cases also became a night mare and risk to anyone of us, if no precautionary measures and tools are taken. Such risks and losses are huge and enormous financial losses. People need to feel safe to shop and buy products and services online and to be confident that their money (credit cards) are safe and not used by others. As a result, banks and financial institutions enforced some tools and techniques to enhance and improve their fraud detections methods and capabilities. Data Mining is one of the most important tools used for credit card fraud detection. The goal of this study is to explore some techniques to be used by banks to reduce credit card fraud cases. By applying these tools and methods, consumers will be more safe and secure to use their credit cards while they are running their shopping online.

Keywords: credit card, data mining, Fraud detection, data mining techniques, electronic payments, online transactions.

1 INTRODUCTION

1.1 Background

Due to the fast developments of electronic commerce technologies, use of credit cards has significantly increased. As a result, credit cards are the most common option of payments and because of this, credit card frauds are increasing in recent years (Patil et al., 2018). Fraud detection includes checking the spending behavior of users in order to determination, discovery or avoidance of undesirable behavior (Pushpalatha and Joseph, 2017).

Nowadays, Data Mining became a popular methodology to battle credit card frauds. Data mining is a well-defined process which takes data as input and generates output in the form of patterns. In other words, the task of data mining is to analyze huge amount of data and to extract some usable information that can interact with for future purposes. If we have the correct model for the data, it can be used for forecasting figure events by classifying the data (Amanze and Onukwugha, 2018). Data mining states to a group of machine learning techniques capable to analyze and extract patterns of data. It is also known as knowledge discovery because it can expose previous unknown information which was hidden in the data.

One of the major challenges to be overcome is to look for ways of using data mining techniques and other statistical tools to identify the reasons in advance and to take immediate actions and efforts. This is doable if the past history of customers is analyzed. The huge amount of billing information, call detail data and network data can be generated and used by data mining techniques. These data will be vague and will require tools to analyze and study them in different perspectives (Desai and Deshmukh, 2013).

1.2 The Research Importance

With the tremendous growth of businesses depending on E transactions, E payments and the global direction towards digitalization, dealing with credit cards is becoming an integral part of any payment methods. Therefore, the need to ensure the safety and privacy of credit cards transactions are vital and gaining more and more importance. This is why this topic is important and requires continuous research and studies. Also, as credit card fraud dilemma is a major challenge to banks and financial institutions, the search for recommendations and solutions is becoming a must.

1.3 Research Problem

There are many options and techniques should be used to minimize credit card fraud cases. This study will discuss and analyze some techniques and will answer the following question: *How can banks and financial institutions minimize credit card fraud cases based on the adoption of Datamining techniques?*

2 RESEARCH METHDOLOGY

This study is a qualitative type and is done by conducting literature reviews on studies and research papers on Data Mining and related topics. By reading and analyzing them, recommendations and finding of this study and the answer to the study question were reached.

2 LITERATURE REVIEW

Data Mining is a raising technology nowadays but what is it? Data Mining is structured procedure where it takes data as input and generates output in the form of patterns (Amanze and Onukwugha, 2018). This means that the function of data mining is to analyze huge volume of data and to extract some useful information which can interact with future needs.
Also, Data Mining is defined as complicated data search capability where is uses statistical algorithms to identify patterns and correlations in data (Desai and Deshmukh, 2013).. Basically, it discovers unseen relationships in data. Data Mining techniques are the outcomes of lengthy studies, research and development processes. The beginning of data mining is with the first storage of data on PCs continuous with developments in data access.
In the first stage, data collection, individuals find gathered data used to make basic calculations like totals and averages. The second step, organization wide policies and procedures for data collection and reporting were recognized. Finally, online tools provided live feedback and information with cooperating business units, which is Data Mining.
The main parts of data mining technology have been evolving for years for research such as AI, ML and statistics. Below we will explore the following studies:

Table 1. Studies covered in this study

#	Study	Year	Primary Goal	Main Results
1	Review on Credit Card Fraud Detection using Data Mining Classification Techniques and Machine Learning	2020	Present an indication of different Data Mining and Machine Learning Techniques for detecting credit card fraud.	• Importance of the topic and some remarks on traditional practices. • Advanced algorithms are required to

				better solve the frauds.
2	Machine Learning Techniques for Credit Card Fraud Detection	2019	Provide 14 different techniques of how data mining obtain high fraud coverage.	Comparison of the advantages and disadvantages of every technique.
3	Survey Paper for Credit Card Fraud Detection Using Data Mining Techniques	2019	Survey results for (4) techniques	Understanding the mechanism of how detections tools are working.
4	An Overview of Credit Card Fraud Detection Using Data Mining Techniques.	2019	• Detect different electronic business frauds. • Investigate the methods used of detections.	There are many techniques for detecting fraud cases.
5	Fraud Detection of Credit Card using Data Mining Techniques	2019	Present several techniques and some fraud cases	Understanding detection techniques is vital to identify fraud cases.
6	An Empirical Study of AML Approach for Credit Card Fraud Detection – Financial Transactions	2019	Help new researchers to identify limitations of current fraud detection techniques and provide some directional research insights	• Pros and cons of some techniques are discussed. • Further development is needed on the current techniques
7	Data Mining Application in Credit Card Fraud Detection System	2018	To understand how credit card fraud are being committed.	The accuracy of fraud detection needs to be improved.
8	A Survey on Different Data Mining and Machine Learning Methods for Credit Card Fraud Detection	2018	Present review of different data mining and machine learning techniques.	Common data mining methods are not enough for some case. Advanced algorithms are needed.
9	Credit Card Fraud Detection Based on the Transaction by Using Data Mining Techniques	2017	Enhance current fraud processes by improving the prediction of fraudulent accounts.	It is important to understand how fraud mechanism is working to find solutions.
10	Study on Credit Card Fraud Detection Using Data Mining Techniques	2017	Compare some data mining techniques.	There are different techniques and tools to detect fraud.

11	Data Mining Techniques for Fraud Detection	2013	To provide general review of data mining and different techniques used to detect fraud.	Presented features of fraud types and need for fraud detection systems.

3 DISCUSSIONS

All research's, papers and surveys included in this study agreed on the importance and value of Fraud Detecting Tools and Techniques. Below we cover the discussions of the used studies:

a- Study Review on Credit Card Fraud Detection using Data Mining Classification Techniques and Machine Learning which was in (2020) stated that fraud detection system faces the following challenges:

Table 2. Fraud Detection System Challenges (Goyal and Manjhvar 2020)

Imbalanced Data	Different Misclassification importance	Overlapping data
Lack of adaptability	Fraud detection cost	Lack of standard metrics

The study also identified the following 2 types of frauds:

1. Offline Fraud: for example, in a call center on a physical stolen card.
2. Online Fraud: online fraud is by a cardholder with shopping online or web pages.

Also, the study mentioned the techniques for credit card fraud detection:

Table 3. Techniques for Credit Card Fraud Detection (Goyal and Manjhvar 2020)

Genetic Algorithm	Decision Tree	Artificial Neural Network
Convolution Neural Network	Outlier Detection	Clustering Techniques
Logistic Regression	Deep learning	Rule Based method
Hidden Markov Model		

This study could be a basic level study for beginners in this area. It is more into an introduction with general information on some tools and techniques without detailed statistics – analysis.

b- Machine Learning Techniques for Credit Card Fraud Detection. This study was in 2019. Compared to study #1 above, this study spoke about (14) different techniques with advantages and disadvantages for each one of them. One of the main challenges faced in this study was the absence of open access for databases, especially visa. Below are tables for datasets used by researchers:

Table 4. Datasets Used in this study (Abd Elhamid *et al* 2019)

Technique	**Source of Data**	**Data Set Size**
Genetic Algorithm	Synthetically generated data	320,000,000 different transactions of 1050 credit cards
	Synthetically generated data	around 1,000,000 transactions
Artificial Immune System	Big Brazilian Bank	41,647 transactions with 3.14% fake fraudulent transactions

	Financial institute in Ireland	More than 4 million credit card transactions, 5417 were fake fraudulent transactions
	Large Australian Bank	640,361 different transactions , 21,746 for credit cards.
Data Mining Techniques	First Union bank & Chase Bank	Each bank provided 500,000 transactions
	Synthetically generated data	10,000 financial transactions
Multiple criteria linear programming	Major US bank	More than 6,000 credit card data.

Availability of accurate and open data is important for the success of such studies, which requires data to perform required analysis and tests.

c- Survey Paper for Credit Card Fraud Detecting Using Data Mining Techniques. This survey was conducted back in 2019. The interesting point about this survey that it to come with a higher level of accuracy to detect frauds, we need to implement combination of different algorithms. Depending on the fraud scenarios, there is no one methodology – solution and this is why we have many techniques and tools for detecting credit card fraud cases. This study mentioned the following tools:

Table 5. Tools used in this Study (Deepa and Akila 2019)

Decision Tree	k-means clustering	Random forest
k-nearest neighbors	Anomaly detection	

This survey didn't provide much details (statistics, analysis, diagrams). On the contrary, it is similar to study #1 in terms of poor statistics and being more into general information and definitions.

d- An Overview of Credit Card Fraud Detection Using Data Mining Techniques. This simple and basic study dated back in 2019 is a good starting material to read for beginners in this field. It uses simple and basic terms and language providing an overview of the topic. One more positive point is it mentioned pros and cons for each of the (5) types of techniques stated.

e- Fraud Detection of Credit Card using Data Mining Techniques. This study also criticized the absence of open and free databases related to this topic. It also generally mentioned how fraudster users behave. One advantage to this study is the use of flow charts in explaining and clarifying the understanding of concepts. Here is how this study presented the Data Mining Process:

f- An Empirical Study of AML Approach for Credit Card Fraud Detection – Financial Transactions. One of the strong elements of this study, is the statistics and analysis mentioned in it. For example:

- Acceding to fraud facts report 2017, the UK payment credit card losses has been increased by 9% in 2016 from Euro 567 million in 2015 to Euro 618 million.
- In India, State Bank of India has blocked 0.6 million debit cards due to a cyber-attack in 2016. It was one of the largest security gaps in history of India.

- According to the Nilson Report, the card fraud losses reached to $21.84 billion in 2015, $24.71 billion in 2016 and $ 27.69 billion in 2017.

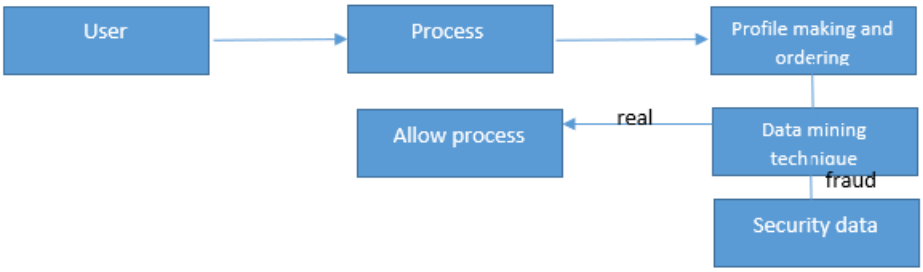

Fig. 1 Data Mining Process (Sharma, Verma and Gupta 2019)

All these figures are solid and facts and reflects the actual dilemma size faced in this area. The authors presented the research process of this paper in the following simple chart:

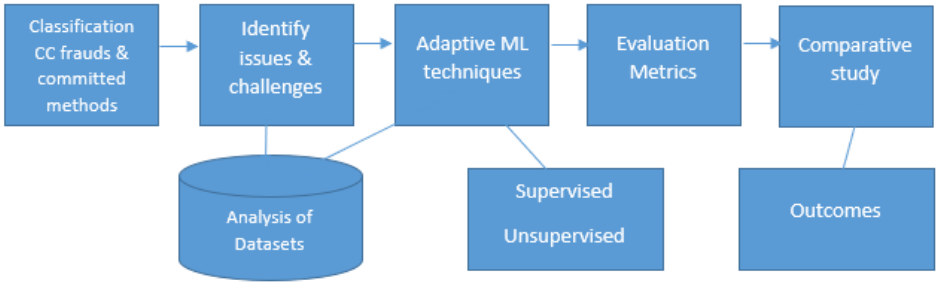

Fig. 2 Research Process of this paper (Singh and Jain 2019)

Furthermore, this study mentioned some algorithms used in some of the detection techniques, for example:

- Bayesian network:

$$P(\text{Fraudulent} / \text{Evidence}) = \frac{(P(\text{Evidence} / \text{Fraudulent}) * P(\text{Fraudulent}))}{(P(\text{Evidence})}$$

The study also presented a comparison table of various detection techniques where advantages and disadvantages of each technique were stated. Among the researches and surveys used in this study, this is the strongest study as it contains deep analysis and figures. It depended on three measures – aspects: sensitivity, specificity and accuracy of the used techniques. Nevertheless, this study is not easy to comprehend by beginners due to the depth and complexity of the data.

g- Data Mining Application in Credit Card Fraud Detection System. This study dated in 2018 provided background about different types of fraudsters that commit fraud the techniques used by them. The study mainly focused on the features and characteristics of both Credit Card Hacking and Fraudsters. These aspects were not covered in this depth by other papers mentioned earlier. One drawback of this study is it mentioned only (2) Fraud Detection tools.

h- A Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection. This study dated in 2018 and mentioned (Sharma et al., 2019) techniques and methodologies. What is different from other studies, it mentioned some challenges related to this topic which are:

1. Typical classification problems.
2. Fraud type and detection methods.
3. Privacy considerations.
4. Computational performance.
5. Evolving problem.
6. Disproportionate misclassification costs.
7. Generic frameworks.

i- Credit Card Fraud Detection Based on the Transaction by Using Data Mining Techniques. One of the pros of this study dated back in 2017, is the use of algorithms in its content. The authors provided algorithms and an ontology algorithm where classes, properties and individuals of Ontology were defined. As an example, in this study, identifying the credit card fraud is based on customer behavioral variables. The parameters involved in the data set are:

- **C_Freq:** frequency of credit card used
- **C_Loc:** Location at which credit card are in the hands of fraudulent.
- **C_OD:** Rate of Over Draft time.
- **C_BB:** balance available at the bank of credit card
- **C_Ds:** average daily spending amount.

Another advantage of this paper is the table of advantages and disadvantages of the techniques covered it. Yet, this paper is more suited to researchers with previous understanding of the topic due to the high level of context.

j- Study on Credit Card Fraud Detection Using Data Mining Techniques. The authors have compared some techniques and tools used in detecting credit card fraud cases. Although the study is dated 2017, the tools covered here are similar to newer papers. It gave a general brief about the tools covered without any comparisons, figures or algorithms.

k- Study on Credit Card Fraud Detection Using Data Mining Techniques. As this is the oldest study among the ones used, the paper provided basic concepts of data mining and the tools used in the fraud detection. The two authors provided basic and simple source of information on the topic without any technical – numerical analysis.

4 CONCLUDED COMMENTS AND RECOMMENDATIONS

At the end of this study and after analyzing above researches and surveys, we can summarize the main conclusions in the following points:

- The world in vastly moving towards digitalization and part of that is the online payment transactions which is increasing every year.
- The use of credit cards is becoming a standard living practice. Therefore, solutions to safe use of them are mandatory.

- Data Mining Techniques is a critical and essential tool for detecting credit card fraud cases.

The study recommends the following main points:

- More studies and reviews on credit card fraud cases are needed for the Middle East area.
- More access and flexibility is needed to open and free databases are encouraged for better researches and studies.
- Such studies and researches are limited in Arabic content. More efforts are required to enrich the Arabic content on this topic.
- More studies and research are required to further minimize the risk of credit card fraud cases by improving current tools or developing new techniques.

References

Abd El-Hamid, H., Khalifa, W., Roushdy, M., and Salem, A. (2019), Machine Learning Techniques for Credit Card Fraud Detection, *Future Computing and Informatics Journal:* Vol. 4 : Issue. 2. Available at: https://digitalcommons.aaru.edu.jo/fcij/vol4/iss2/5

Amanze, B and Onukwugha, C (2018), Data Mining Application in Credit Card Fraud Detection Systems. *IJTRD*, Vol 5, Issue 4.

Deepa, M. , Akila, D (2019), Survey Paper for Credit Card Fraud Detection Using Data Mining Techniques, *IJIRASE*, Vol 3, Issue 6

Desai, A., Deshmukh, R (2013), Data Mining Techniques for Fraud Detection, *IJCSIT*, Vol 4, Issue 1.

Goyal, Rahul and Manjhvar, Amit (2020), Review on Credit Card Fraud Detection using Data Mining Classification Techniques & Machine Leaning Algorithms, *IJRAR*, Vol7, Issue 1

Kavipriya, T., Geetha, N (2017) Study on Credit Card Fraud Detection Using Data Mining Techniques, *IJCRCST*, Vol 3, Issue 3.

Mantri, A., Sen, C. and Kumar, S (2019) An Overview of Credit Card Fraud Detection Using Data Mining Techniques, *IJSART*, Vol 5, Issue 4.

Patil, Vipul and Lilhore, Umesh (2018) A Survey on Different Data Mining & Machine Leaning Methods for Credit Card Fraud Detection, *IJSR, CSEIT*, Vol 3, Issue 5.

Pushpalatha,B. and Joseph, C. (2017) Credit Card Fraud Detection Based on the Transaction by Using Data Mining Techniques, *IJIRCCE*, Vol 5, I 2,1785-1794.

Sharma, A. , Verma, A. and Gupta, D (2019) Fraud Detection of Credit Card Using Data Mining Techniques, *IJITEE*, Vol 8, Issue 12.

Singh, Ajeet, Jain, Anurag, An Empirical Study of AML Approach for Credit Card Fraud Detection–Financial Transactions. *International Journal of Computers Communications & control*, [S.l.], v. 14, n. 6, p. 670-690, Feb. 2020. ISSN 1841-9844. Available at: <http://univagora.ro/jour/index.php/ijccc/article/view/3498>. Date accessed: 06 Mar. 2021.