Credit Card Fraud Detection: A Systematic Literature Review

Hossain, Md. Mobarak[1], Sams, A.B.M. Toufiqul Alam[2], Sufian Mustaq [3], and Course Supervisor: DR. M. M. MAHBUBUL SYEED[1]

[1]18-38054-2, G2, C, Department of Computer Science, 18-38054-2@student.aiub.edu
[2]19-39297-1, G2, C, Department of Computer Science, t101alam52@gmail.com
[3]19-39292-1, G2, C, Department of Computer Science, sufianmustaq3599@gmail.com

August 25, 2021

Abstract

It has been reported that one tenth of world's population is currently doing their shopping online [1]. As credit card is becoming the most popular mode of payment, fraudulent cases which are associated with it are also increasing and this has become a serious issue. This issue should be solved efficiently and safely for keeping people's money safe and also keeping their trust in the digital system.

This paper reports on a systematic literature review aimed at the credit card fraud detection algorithms, how they work, where are they good at, where are they lacking and what data are used for get the algorithms working. After reviewing the papers, we can compare the algorithm and get the results and based on the results we can have the appropriate algorithm for use.

1 INTRODUCTION

Banking is a system offered by banks which deals with cash, credit and other financial transactions in a secure way. With the help of modern technology, people have successfully implemented the online version of banking which his called E-banking. E-banking has made our life easier with cashless transactions. The most impact of credit card can be seen in the sector of E-Commerce because of rapid growth of using credit cards for purchasing or selling goods online which need the use of cashless payment system. Credit card is a modern mode of payment which is a plastic card [2] specifically issued to number of users which allows the cardholder to purchase goods online as well as offline. In UK, 2016, 1 out of 7 people reported that they do not prefer to carry cash and 75% of them responded to choose credit or debit card as a payment method [3]. As credit card has become one of the most popular method of payment for both online and as well as offline, fraudulent cases which are associated with it are also increasing [4]. Credit cards fraud are increasing because it gives a lot of money within a very short time without taking too many risks [5]. Keeping money secure is the most important part of a bank where modern banking system provides technological securities. As modern banking system is based on technology, some people called as fraudster do financial abuse which is the main cause of losing significant financial resources. This harm decreases the trust of a customer in using modern banking system. "Between July 2005 and mid-January 2007, a breach of systems at TJX Companies exposed data from more than 45.6 million credit cards [6]." "In 2012, about 40 million sets of payment card information were compromised by a hack of Adobe Systems [7]." These are the known and famous attack which is known for exposing huge amount of information and loss but there are many personal attacks which are unknown to all. These activities are related with human and computer and being computer science students, we have responsibility to prevent these fraudulent activities to keep the trust of people in modern system. We will review some papers relevant to our topic and will try to compare the data mining techniques used in those papers and we will do these things by a systematic literature review.

2 RESEARCH METHODOLOGY

Our main research methodology is a systematic literature review as mentioned earlier. At first, we will select the papers related to our topic with the keywords and will find the relevant papers and then select some set of papers form the selected papers which are suitable for our purpose. We will try to see what algorithms were used there, how they are used and what data were used and how data were collected. After reviewing all these things, we will try to compare the algorithms and will try to find the best working algorithm for credit card fraud detection.

2.1 Research Objective

The main objective is to reviewing the Data Mining techniques used for credit card fraud detection. We will try to know which algorithms were used, how were they were used and how data was collected and how data were applied to these algorithms. After knowing all these things, we can analysis the data and compare the algorithms to find the best algorithms for credit card fraud detection. After conducting the research, we can provide the best working algorithms for use and prevent credit card fraud detection.

2.2 Research Questions

The research questions we have defined fall within the context of Credit Card Fraud detection and Data mining techniques. In total we have formulated 7 questions, as presented in Table 1. These questions proposed to know about all the things about the data mining techniques they used and at the end to compare them to find the best algorithms. This covers aspects like the focus of the study, methodological details, data sources and validation mechanism.

Category	**Research Questions**	**Main Motivation**
Target	a.What is their main target to solve the problem, did they used algorithms or another approaches?	a. To know about working process of the research papers.
Approach	a. What algorithms were used for credit card fraud detection in the studies? b.How data were collected for testing the algorithms? c. How data were applied to the algorithms?	a. To know about the available algorithms in use to find the frauds. b. Data collection might be an issue because we need real life data. So, we need to find out how data were collected. c. To know about how data were used for the approaches.
Outcome	a. Do the algorithms give the expected results? b. Which algorithms give the best results? c. How the research approaches and results of the articles typically validated?	a. To know the algorithms performance. b. To find the best working algorithms. c. To identify the approaches employed to evaluate the research approaches and study results.

Table 1: Research Questions

2.3 Article Selection

We have selected all the articles from the digital libraries, Google Scholar search and by checking references of the searched articles.

2.3.1 Keywords and Search String

The keywords and search string that were used for searching for the papers and articles were “credit card”, “detection”, “credit card fraud”,” credit card fraud detection”, “review”, “systematic literature”, “systematic literature review”.

2.3.2 Digital Libraries to Search

We searched in ACM digital library, IEEE digital library, Google Scholar for the articles. We have searched with advanced search option with the keywords and timeline for search was ¿1990.

2.3.3 Keyword search and Manual Selection

We have searched with the keywords “credit card”, “detection”, “credit card fraud”,” credit card fraud detection”, “review”, “systematic literature”, “systematic literature review” and after finding the resulting articles and papers we had manually searched for the relevant papers form the searched papers by reading the titles, abstract and introduction part also we had read the references part for finding any papers that we had missed. We had found 30 articles by keyword search and then by reference checking we have found additional 12 articles. In total we had 42 articles (17 journal articles and 25 conference articles).

2.3.4 Final set of Articles

The article selection process finally ended up with 11 articles (4 journals and 7 conference paper). The articles are already added in the references section.

3 DISCUSSIONS

In this part, we will discuss about the research questions and our findings about the answer of the questions.

RQ1. What is their main target to solve the problem, did they used algorithms or another approaches?

The articles [10][11] they just surveyed the credit card fraud incidents; they did not give any solution based on the problems. The rest of the articles gave some Data Mining algorithms as solution.

RQ2. What algorithms were used for credit card fraud detection in the studies?

In the article [1] Hidden Markov Model was used as a solution, in [3] Capsule Network was used as a solution, in [2, 4] some algorithms together were reviewed, [5] Bayesian and Neural Networks was used, [8] K-Nearest Neighbor algorithm was used, [9] Dempster Shafer theory and Bayesian learning, [10] Fuzzy Darwinian Detection, [11] BLAST-SSAHA Hybridization algorithm was used.

RQ3. How data were collected for testing the algorithms?

Data relevant to credit card fraud detection is very sensitive because all the data are connected to banking information that is why it is very hard to collect real world's data. So, data set were made by the scientist in most of cases.

RQ4. How data were applied to the algorithms?

The main approach of using data in the Data Mining techniques is to train the algorithms on the available dataset to teach them how to react when this type of incident occurs.

RQ5. Do the algorithms give the expected results?

There were some parameters based on which we evaluated the algorithms. Seems like none of them were 100% accepted for use. We used 5 parameters for evaluation true positive%, false positive%, processing speed, cost, accuracy.

RQ6. Which algorithms give the best results?

We have created a table with the 5 performance parameters for comparing the algorithms. From the above table we can say that

Parameter	***Dempster–Shafer theory and Bayesian learning***	***Hybridization of BLAST and SSAHA***	***HMM***	***Bayesian and Neural Network***	***Fuzzy Darwinian detection***
True Positive%	98	86	70	74	100
False Positive%	10	10	20	10	5.79
Processing Speed	Medium	Very High	High	High	Low
Cost	Expensive	Inexpensive	Quite Expensive	Expensive	Highly Expensive
Accuracy	High	High	Medium	Medium	Very High

Table 2: Comparison of various fraud detection algorithms

among these algorithms Fuzzy Darwinian detection method is more capable than others in all parameters except speed and cost. As it is highly expensive and the speed is comparatively low, if we could manage to reduce the cost and increase the processing speed then Fuzzy Darwinian will be the most effective and efficient algorithm we could ever have.

RQ7. How the research approaches and results of the articles typically validated?

The approaches and results were validated using the data sets and also some algorithms were used in real life.

4 FUTURE RESEARCH DIRECTIONS

In future we are planning to merge the available algorithms of credit card fraud detection and will try to provide the perfect algorithm to use. For this we will need to research on the previous algorithms and learn all available information about them. Bu doing these tasks in future, we can provide more secure technique for credit card fraud detection.

5 VALIDITY THREAT

Carrying out a survey is mostly a manual task. Thus, most threats to validity relate to the possibility of re-searcher bias. To minimize this, we adopted guidelines on conducting SLR. In particular, we documented and reviewed all steps we made in advance, including selection criteria and attribute definitions.

6 CONCLUSIONS

As modern banking system is becoming a must use for everyone in the world, as credit card system is becoming more famous its fraud rate is also increasing. There are many data mining techniques in use to preventing the fraud from their misconduct but many of them are still failing to their job. So, we need to find the best technique for preventing this kind of misconduct. In future it will be very much needed for a 100% working algorithm for preventing these frauds. This research work has a bigger potential for the banking system for giving the people a safe digital banking system. By conducting this research banks can update their system according the research result and the system will be able to identify if the transaction is by being a legit user or not. After that people can have a safe and reliable credit card usage.

References

[1] Abhinav, Srivastava, and et al, "Credit card fraud detection using hidden markov model," *IEEE Transactions on dependable and secure computing*, vol. 5, pp. 37–48, 2008.

[2] K. Chaudhary and et al, "A review of fraud detection techniques: Credit card," vol. 45-No.1, May 2012.

[3] S. Wang and et al, "Credit card fraud detection using capsule network," *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3679–3684, 2018.

[4] A. A. Portia and S. B. E. Raj, "Analysis on credit card fraud detection methods," *International Conference on Computer, Communication and Electrical Technology*, pp. 152–156, 2011.

[5] B. Manderick, B. Vanschoenwinkel, K. Tuyls, and S. Maes, "Credit card fraud detection using bayesian and neural networks," pp. 261–270, 1993.

[6] "Tjx hacker gets 20 years in prison." https://www.wired.com/2010/03/tjx-sentencing/, March 2018.

[7] "Adobe hacked: customer data, source code compromised." https://www.smh.com.au/technology/adobe-hacked-customer-data-source-code-compromised-20131004-hv1wl.html, October 2015.

[8] M. Fartash and M. Khodabakhshi, "Fraud detection in banking using knn (k-nearest neighbor) algorithm," *5th International Conference on Research in Science and Technology*, pp. 26–34, 2016.

[9] A. Kundu, A. K. Majumdar, S. Sural, and S. Panigrahi, "Credit card fraud detection: A fusion approach using dempster shafer theory and bayesian learning," *Special Issue on Information Fusion in Computer Security*, vol. 10,Issue no 4, pp. 354–363, October 2009.

[10] G.-H. Jung, J. Kim, J.-U. Choi, and P. J. Bentley, "Fuzzy darwinian detection of credit card fraud," *In the 14th Annual Fall Symposium of the Korean Information Processing Society*, 14th October 2000.

[11] A. Kundu, A. K. Majumdar, S. Sural, and S. Panigrahi, "Blast-ssaha hybridization for credit card fraud detection," *IEEE Transactions On Dependable and Secure Computing*, vol. 45, Issue no. 4, pp. 309–315, October-December 2009.

A Contribution Record

Detail of each group members contributions.

A.1 Paper Assessment

Paper reviewing information.

Student id & name	**Paper no from references**	**Paper Title**
18-38054-2 Hossain, Md. Mobarak	[1]	Credit card fraud detection using hidden Markov model
19-39292-1 Sufian Mustaq	[2]	A review of Fraud Detection Techniques: Credit Card
18-38054-2 Hossain, Md. Mobarak	[3]	Credit Card Fraud Detection Using Capsule Network.
18-38054-2 Hossain, Md. Mobarak	[4]	Analysis on Credit Card Fraud Detection Methods.
19-39292-1 Sufian Mustaq	[5]	Credit card fraud detection using Bayesian and neural networks
19-39297-1 Sams ,A.B.M. Toufiqul Alam	[8]	Fraud Detection in Banking Using KNN (K-Nearest Neighbor) Algorithm
19-39292-1 Sufian Mustaq	[9]	Credit card fraud detection: A fusion approach using Dempster Shafer theory and Bayesian learning,
18-38054-2 Hossain, Md. Mobarak	[10]	Fuzzy Darwinian Detection of Credit Card Fraud.
19-39297-1 Sams, A.B.M. Toufiqul Alam	[11]	BLAST-SSAHA Hybridization for Credit Card Fraud Detection
19-39297-1 Sams, A.B.M. Toufiqul Alam	[6]	TJX Hacker Gets 20 Years in Prison
19-39297-1 Sams, A.B.M. Toufiqul Alam	[7]	Adobe hacked: customer data, source code compromised

Table 3: Paper collected and read by the group member

A.2 Paper writing contribution

Paper writing information.

Student id & name	**Section No**	**Section Title**
19-39297-1 Sams, A.B.M. Toufiqul Alam	1	Introduction
18-38054-2 Hossain, Md. Mobarak	2.1	Research Objective
18-38054-2 Hossain, Md. Mobarak	2.2	Research Questions
19-39292-1 Sufian Mustaq	2.3	Article Selection
18-38054-2 Hossain, Md. Mobarak	3	Discussions
19-39297-1 Sams, A.B.M. Toufiqul Alam	4	Future Research Directions
19-39297-1 Sams, A.B.M. Toufiqul Alam	5	Validity Threat
19-39292-1 Sufian Mustaq	6	Conclusions.

Table 4: Section(s) Written in the paper by the group member