

Credit Card Fraud Detection using Machine Learning: A Systematic Literature Review

Harish Paruchuri

Senior AI Engineer, Department of Information Technology, Anthem, Inc., USA

*Corresponding Contact:

Email: harishparuchuri9999@gmail.com

ABSTRACT

Companies want to give more and more facilities to their customers. One of these facilities is the online mode of buying goods. The customers now can buy the required goods online but this is also an opportunity for criminals to do frauds. The criminals can theft the information of any cardholder and use it for online purchases until the cardholder contacts the bank to block the card. This paper shows the different algorithms of machine learning that are used for detecting this kind of transaction.

Keywords: Credit card, Machine learning, online transactions, Fraud detection

9/15/2017

Source of Support: Nil, No Conflict of Interest: Declared

This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Attribution-NonCommercial (CC BY-NC) license lets others remix, tweak, and build upon work non-commercially, and although the new works must also acknowledge & be non-commercial.



INTRODUCTION

The internet is the great invention of the modern times. The users of the internet are increasing day by day. The business organizations or companies also started their business through this online medium (Khan, Akhtar, & Qureshi, 2014). These business companies are providing the facility of online buying to their customers. Customers can buy the required products through the website or ecommerce stores of these companies. Most customers use credit card for buying things online. In this way, some of the customers can be the thief who has stolen the card of a person to make the online transactions. This is considered as the credit card fraud that must be detected. This fraud can also be in the form of any purchase by using the credit card in an unauthorized way. The cases of this kind of fraud are increasing. It is necessary to solve this challenging issue. Artificial intelligence is saving the time of humans in different fields. Especially machine learning, which is the branch of artificial intelligence is very helpful in performing the complex and difficult tasks. Many researchers used this sub field of artificial intelligence as a solution to various problems. It is necessary to handle the credit card fraud problem through the machine learning because this cannot be done by a human being in a proper way.

Online customers are increasing day by day. The customers now want to purchase the goods by sitting at their homes because of different reasons. For example, purchasing goods online saves the time of the customers. The increasing number of online customers makes credit card fraud, a more challenging and important problem. Electronic payment has several issues but the major issue is the credit card fraud (Vadlamudi, 2015). The

ecommerce stores also provide the facility of cash on delivery but the customers mostly purchase the goods through the credit card. Customer prefers to use credit card while purchasing the goods because of different benefits like cash back, discount (Seeja & Zareapoor, 2014). According to the researchers (Zareapoor & Shamsolmoali, 2015) credit card fraud (CCF) is the reason for the loss of billions dollars to the financial companies and online customers. The fraudulent transactions are performed in a way that looks like the original transactions and these transactions cannot be detected by the traditional pattern matching techniques (Excell & Security, 2012). This is a problem that cannot be compromised because with the passage of time the number of frauds are increasing. One percent of the online transactions are the fraud transactions (Juszczak et al., 2008). It is necessary to have an efficient mechanism that can be used to deal with this problem.

Every day, new and new researches are performed by the researchers in the different fields. Many researchers of finance field considered this problem as a challenging and important problem. The use of machine learning is proposed by the researchers to deal with this problem. Researchers trained the algorithms of machine learning with different kinds of data set. These algorithms are proved helpful in detecting this kind of fraud. Researchers (Ujwala et al., 2012) used the different machine learning (ML) algorithms for the CCF detection. Researchers said the different ecommerce websites are offering the online payment modes and it is also increasing the number of online frauds. They applied the various ML algorithms and random forest detected frauds with more accuracy. Researchers (Bhattacharyya, Jha, Tharakunnel, & Westland, 2011), (Ngai, Hu, Wong, Chen, & Sun, 2011), (Zareapoor, Seeja, & Alam, 2012) discussed the various kinds of frauds and fraud detection techniques. Researchers (Aleskerov, Freisleben, & Rao, 1997) created a system that is based on neural network. This system is helpful for detecting the CCF.

In our study, we will discuss various type of machine learning algorithms that are used by different researchers for the purpose of CCF detection. Our main aim is providing a complete literature review about this problem. Different attributes of the customers that can be used for training the algorithm for this purpose, will also be identified in this literature review.

The next section of this literature review is Methodology, section three shows the Research questions of this literature review, section four shows the Search process, section five is the Result and discussion section, section six shows the Conclusion.

METHODOLOGY

The systematic literature review is the commonly used methodology for the systematic literature review that is followed for writing this systematic literature review. According to this methodology we searched and discussed the relevant papers in this literature review.

Inclusion and Exclusion

For including papers in this literature review, we followed a criteria. This criteria ensures that the added papers are completely available as well as written in the English language.

Quality Assessment

Before adding the papers to this literature review, their quality is assessed. The quality assessment process is performed on the basis of the researcher's work that is presented in that paper.

RESEARCH QUESTION (RQ)

A systematic literature review has to answer one or more research questions. These are the RQ's that will be answered in result and discussion section.

RQ 1: What are the different kinds of credit card frauds and how machine learning algorithms can be trained for prevention?

RQ 2: What are the different machine learning algorithms that are used for credit card fraud detection?

SEARCH PROCESS

The process of search that we followed for finding the most relevant papers is based on different steps. In the first step, we checked the title of all the papers that we found against the search string. The papers with relevant titles were considered in the next phase that was reading the abstract of these papers. In this second step, the paper's whose abstract was relevant to CCF detection were collected. Complete content of papers was studied in the last phase to get only the papers that are completely relevant to the title of this systematic literature review.

RESULT AND DISCUSSION

Many companies are offering their products in online mode to the customers. The customers are allowed to make payments in different ways such as using the credit card or cash on delivery. Customer prefers the credit card method because of several reasons such as the cash back etc. The use of credit cards is increasing continuously (Vadlamudi, 2016). The transactions that are made through the credit card can be a fraud and this a major problem in the financial research (Donepudi, 2015). Because of the popularity of credit cards, this is considered as the most easy and popular way of online purchases (Raj & Portia, 2011). This increasing use of credit cards also increasing the fraud cases that are associated with it. A huge amount of losses are made every year because of the fraudulent transactions (Dal Pozzolo, Caelen, Le Borgne, Waterschoot, & Bontempi, 2014).

Researchers (Ghosh & Reilly, 1994) created a CCF detection system by using neural network. They trained a neural network with larger data set so that it can detect the fraud. This system was created for a bank to detect the various kinds of CCFs. Researchers (Ogwueleka & Technology, 2011), (Zaslavsky, Strizhak, & Security, 2006) used neural network for detecting CCF on the basis of the behavior of the customers. Researchers (Duman & Ozcelik, 2011) tried to enhance the CCF detection solution that was using by a bank. They applied the scatter search and genetic algorithm to improve accuracy of the existing solution. The improved solution that the researchers provided is able to detect the credit card frauds with 200% more accuracy.

In this digital world, it is very important for the finance sector to detect the frauds that are connected with the credit card. Researchers (Maes, Tuyls, Vanschoenwinkel, & Manderick, 2002) said the machine learning can be helpful to resolve this issue. They applied the two different machine learning algorithms for this purpose. These are the Bayesian network and neural network. Through the results they proved that the Bayesian network can detect the fraud in a faster way than the neural network. SVM based approach is used by researchers (Lu & Ju, 2011) for detecting the CCFs.

The understanding about the credit card frauds and its prevention techniques is necessary to stop this fraud (Masuda, 1993), (Delamaire, Abdou, Pointon, & systems, 2009), (Bhatla, Prabhu, & Dua, 2003). We will first identify the types of credit card frauds and after that a detailed discussion will be made on ML algorithms that are used for preventing this fraud.

RQ 1: “What are the different kinds of credit card frauds and how machine learning algorithms can be trained for prevention?”

Due to the increasing use of the internet, many companies are offering their services or products in online mode. Credit is now the popular way of making the payments (Srivastava, Kundu, Sural, Majumdar, & computing, 2008). This can be in the form of online payment or any regular payment. The number of transactions that are made through the credit cards are increasing continuously (Chan, Fan, Prodromidis, Stolfo, & Applications, 1999). The credit card frauds are also increasing in the number because of the increasing use of credit cards for transactions (Shen, Tong, & Deng, 2007). This is now easy for the fraudsters to make various types of frauds related to the credit card (Quah & Sriganesh, 2008).



Figure 1: Credit Card Fraud Types

Bankruptcy Fraud

This is the kind of CCF that cannot be detected easily. To detect this, the use of latest techniques and methods is necessary. In this kind of fraud, the fraudster tries to make transactions when he knows he cannot pay through that card. As a result the bank faces the loss. A model is proposed by the researchers (Foster & Stine, 2004) to detect this kind of fraud.

Theft Fraud

In this kind of fraud, a person uses the credit card of another person in an illegal way. In simple words, we can say that the card of a person is stolen by a thief to make different transactions. The thief makes the transactions as much as possible until the card owner does not contact the bank to block the card. This kind of fraud is usually made while buying the goods from the internet. The card does not need physically on these websites to buy things. The thief only enters the details given on the card. This kind of fraud is easy and increasing in the number because the companies have to offer their products online and more and more companies are now offering the products online. This kind of fraud can be detected through the over limit report. The banks record the transactions that are made through an account in a day and if it exceeds the limit, the bank contacts the customer. In case of no response from the customer, the banks block the card.

Application Fraud

This is a kind of fraud that is made while applying for a credit card with the wrong details. The applicant provides false information while submitting the application for a credit card. This kind of fraud can also be made by entering the details of a stolen ID card while filling the application form (Phua, Gayler, Lee, Smith, & Control, 2005). A system can be developed to detect this kind of fraud. The system should be able to tell the bank if two or more applications contain the same ID card details or other information that must not be repeated.

Behavioral Fraud

The fraudulent makes this type of fraud without accessing the card of a person, only the details that are given on the card are known to a person and he makes the online transactions. This includes any transaction that only needs the card details such as the e-commerce transactions (Bolton & Hand, 2002). A data set is necessary to train any machine learning algorithm for prediction or detection purposes. The data set is based on various attributes. These attributes are according to the entity. For the detection of CCF, the data set can be based on the following attributes to train the machine learning algorithm.

Table 1: Attributes for Credit Card Fraud Detection

Transaction ID	Card Owner ID
Amount	Time
Fraud Label	Transaction Type

RQ 2: What are the different machine learning algorithms that are used for credit card fraud detection?

ML is using in different fields for different purposes. Each company or institute uses the machine learning according to the requirement. Mostly these algorithms are used in the companies where prediction or detection is required. Different kinds of frauds are made in the banking sector that are related to the credit card. The prevention of this kind of frauds is of vital importance for the prediction techniques (Brause, Langsdorf, & Hepp, 1999). The credit card frauds are the reason behind losing the billions of Euros each year (Bahnsen, Aouada, Stojanovic, & Ottersten, 2016). It is necessary for financial institutes to keep their fraud detecting system updated with new and new techniques. Companies are moving towards the online mode to give the latest facilities to their customers and this is also a reason behind the increasing number of credit card frauds (Patidar, Sharma, & Engineering, 2011). To handle or detect this fraud, the researchers used different algorithms of machine learning. The use of these algorithms is described below.

Neural Network

This algorithm is used by various researchers for detecting the CCF. Researchers (Murli, Jami, Jog, Nath, & Management, 2015) said the credit frauds are increasing because of the increasing interest in online shopping. They implemented a mechanism based on neural network to detect this kind of fraud. Researchers (Lakshmi Narayana et al., 2012) also used this algorithm for the detection of increasing CCF activities.

Decision Tree

Researchers (Save, Tiwarekar, Jain, & Mahyavanshi, 2017) used the decision tree for the detection of credit card frauds. Researchers said, increasing use of internet and online business is also increasing so it is necessary to have a mechanism to deal with this

problem. Researchers (Dhanapal & Gayathiri, 2012) also used this algorithm for CCF detection. CCF affects both i.e. merchants and the individual clients. This makes it an interesting topic in the machine learning (Gaikwad et al., 2014).

SVM

Researchers (Dheepa & Dhanapal, 2012) proposed a new approach to detect the CCF. This approach is based on SVM and detects the fraud on the basis of the customer's behavior. Researchers (Movva et al., 2012) also used a support vector based approach to save the bank from different kind of frauds on the basis of customers' behavior.

Logistic Regression

Researchers (Kulkarni & Ade, 2016) used this algorithm to detect the increasing frauds related to the credit. Researchers said this is an important and rising issue in the financial sector. Researchers (Movva et al., 2012) compared performance of different ML algorithms for CCF detection. This algorithm is considered better than the other algorithms on the basis of results.

Genetic Algorithm

Researchers (RamaKalyani, UmaDevi, & Research, 2012) used an optimization technique for the detection of increasing frauds in online transactions. This technique is the genetic algorithm. To improve the performance of an existing fraud detection system, researchers (Özçelik, Duman, Işık, & Çevik, 2010) used this algorithm. Researchers (Patel, Singh, & Engineering, 2013) also used this algorithm for the prevention of CCFs.

Random Forest

Researcher (Donepudi, 2014a) used this algorithm for detecting the increasing CCFs in online and offline payments. Criminal's theft the credit information by using different techniques such as phishing or Trojan (Donepudi, 2014).

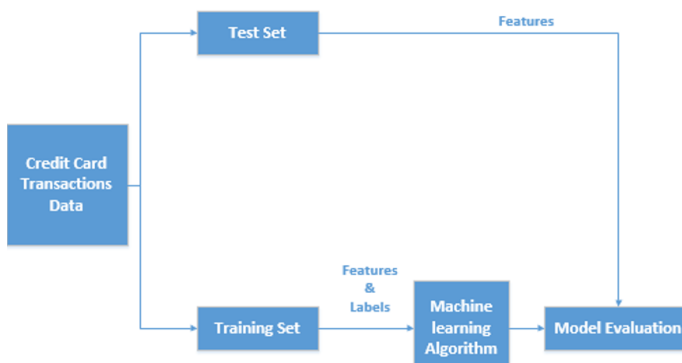


Figure 2: Algorithm Training

CONCLUSION

The research shows the CCF is the major issue of financial sector that is increasing with the passage of time. More and more companies are moving towards the online mode that allows the customers to make online transactions. This is an opportunity for criminals to theft the information or cards of other persons to make online transactions. The most popular techniques that are used to theft credit card information are phishing and Trojan. So a fraud detection system is needed to detect such activities.

REFERENCES

- Aleskerov, E., Freisleben, B., & Rao, B. (1997). Cardwatch: A neural network based database mining system for credit card fraud detection. Paper presented at the Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFER).
- Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. J. E. S. w. A. (2016). Feature engineering strategies for credit card fraud detection. 51, 134-142.
- Bhatla, T. P., Prabhu, V., & Dua, A. J. C. b. r. (2003). Understanding credit card frauds. 1(6), 1-15.
- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. J. D. s. s. (2011). Data mining for credit card fraud: A comparative study. 50(3), 602-613.
- Bolton, R. J., & Hand, D. J. J. S. s. (2002). Statistical fraud detection: A review. 17(3), 235-255.
- Brause, R., Langsdorf, T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. Paper presented at the Proceedings 11th International Conference on Tools with Artificial Intelligence.
- Chan, P. K., Fan, W., Prodromidis, A. L., Stolfo, S. J. J. I. S., & Applications, T. (1999). Distributed data mining in credit card fraud detection. 14(6), 67-74.
- Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. J. E. s. w. a. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. 41(10), 4915-4928.
- Delamaire, L., Abdou, H., Pointon, J. J. B., & systems, B. (2009). Credit card fraud and detection techniques: a review. 4(2), 57-68.
- Dhanapal, R., & Gayathiri, P. J. I. J. o. C. S. I. (2012). Credit card fraud detection using decision tree for tracing Email and IP. 9(5), 406.
- Dheepa, V., & Dhanapal, R. J. I. J. o. S. c. (2012). Behavior based credit card fraud detection using support vector machines. 2(07), 2012.
- Donepudi, P. K. (2014). Technology Growth in Shipping Industry: An Overview. *American Journal of Trade and Policy*, 1(3), 137-142. <https://doi.org/10.18034/ajtp.v1i3.503>
- Donepudi, P. K. (2014a). Voice Search Technology: An Overview. *Engineering International*, 2(2), 91-102. <https://doi.org/10.18034/ei.v2i2.502>
- Donepudi, P. K. (2015). Crossing Point of Artificial Intelligence in Cybersecurity. *American Journal of Trade and Policy*, 2(3), 121-128. <https://doi.org/10.18034/ajtp.v2i3.493>
- Duman, E., & Ozelik, M. H. J. E. S. w. A. (2011). Detecting credit card fraud by genetic algorithm and scatter search. 38(10), 13057-13063.
- Excell, D. J. C. F., & Security. (2012). Bayesian inference—the future of online fraud protection. 2012(2), 8-11.
- Foster, D. P., & Stine, R. A. J. J. o. t. A. S. A. (2004). Variable selection in data mining: Building a predictive model for bankruptcy. 99(466), 303-313.
- Gaikwad, J. R., Deshmane, A. B., Somavanshi, H. V., Patil, S. V., Badgujar, R. A. J. I. J. o. I. T., & Engineering, E. (2014). Credit Card Fraud Detection using Decision Tree Induction Algorithm. 4(6).
- Ghosh, S., & Reilly, D. L. (1994). Credit card fraud detection with a neural-network. Paper presented at the System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on.
- Juszczak, P., Adams, N. M., Hand, D. J., Whitrow, C., Weston, D. J. J. C. S., & Analysis, D. (2008). Off-the-peg and bespoke classifiers for fraud detection. 52(9), 4521-4532.
- Khan, A. U. S., Akhtar, N., & Qureshi, M. N. (2014). Real-time credit-card fraud detection using artificial neural network tuned by simulated annealing algorithm. Paper presented at the Proceedings of International Conference on Recent Trends in Information, Telecommunication and Computing, ITC.
- Kulkarni, P., & Ade, R. (2016). Logistic regression learning model for handling concept drift with unbalanced data in credit card fraud detection system. Paper presented at the Proceedings of the Second International Conference on Computer and Communication Technologies.
- Lakshmi Narayana S., Suneetha Devi J., Bhargav Reddy I., Harish Paruchuri. (2012). Optimizing Voice Recognition using Various Techniques. *CiT International Journal of Digital Signal Processing*, 4(4), 135-141
- Lu, Q., & Ju, C. J. J. o. C. I. T. (2011). Research on credit card fraud detection model based on class weighted support vector machine. 6(1).

- Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002). Credit card fraud detection using Bayesian and neural networks. Paper presented at the Proceedings of the 1st international nairo congress on neuro fuzzy technologies.
- Masuda, B. J. C. p. s. (1993). Credit card fraud prevention: A successful retail strategy. 1, 121-134.
- Movva, L., Kurra, C., Koteswara Rao, G., Battula, R. B., Sridhar, M., & Harish, P. (2012). Underwater Acoustic Sensor Networks: A Survey on MAC and Routing Protocols. *International Journal of Computer Technology and Applications*, 3(3).
- Murli, D., Jami, S., Jog, D., Nath, S. J. I. J. o. S. R. i. T., & Management. (2015). Credit card fraud detection using neural networks. 2(2), 84-88.
- Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. J. D. s. s. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. 50(3), 559-569.
- Ogwueleka, F. N. J. J. o. E. S., & Technology. (2011). Data mining application in credit card fraud detection system. 6(3), 311-322.
- Özçelik, M. H., Duman, E., Işık, M., & Çevik, T. (2010). Improving a credit card fraud detection system using genetic algorithm. Paper presented at the 2010 International Conference on Networking and Information Technology.
- Patel, R. D., Singh, D. K. J. I. J. o. S. C., & Engineering. (2013). Credit card fraud detection & prevention of fraud using genetic algorithm. 2(6), 292-294.
- Patidar, R., Sharma, L. J. I. J. o. S. C., & Engineering. (2011). Credit card fraud detection using neural network. 1(32-38).
- Phua, C., Gayler, R., Lee, V., Smith, K. J. P. o. C. S., & Control, C. (2005). On the approximate communal fraud scoring of credit applications. 1-10.
- Quah, J. T., & Sriganesh, M. J. E. s. w. a. (2008). Real-time credit card fraud detection using computational intelligence. 35(4), 1721-1732.
- Raj, S. B. E., & Portia, A. A. (2011). Analysis on credit card fraud detection methods. Paper presented at the 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET).
- RamaKalyani, K., UmaDevi, D. J. I. J. o. S., & Research, E. (2012). Fraud detection of credit card payment system by genetic algorithm. 3(7), 1-6.
- Save, P., Tiwarekar, P., Jain, K. N., & Mahyavanshi, N. J. I. J. o. C. A. (2017). A novel idea for credit card fraud detection using decision tree. 161(13).
- Seeja, K., & Zareapoor, M. J. T. S. W. J. (2014). Fraudminer: A novel credit card fraud detection model based on frequent itemset mining. 2014.
- Shen, A., Tong, R., & Deng, Y. (2007). Application of classification models on credit card fraud detection. Paper presented at the 2007 International conference on service systems and service management.
- Srivastava, A., Kundu, A., Sural, S., Majumdar, A. J. I. T. o. d., & computing, s. (2008). Credit card fraud detection using hidden Markov model. 5(1), 37-48.
- Ujwala, D., Ram Kiran, D. S., Jyothi, B., Fathima, S. S., Paruchuri, H., Koushik, Y. M. S. R. (2012). A Parametric Study on Impedance Matching of A CPW Fed T-shaped UWB Antenna. *International Journal of Soft Computing and Engineering*, 2(2), 433-436.
- Vadlamudi, S. (2015). Enabling Trustworthiness in Artificial Intelligence - A Detailed Discussion. *Engineering International*, 3(2), 105-114. <https://doi.org/10.18034/ei.v3i2.519>
- Vadlamudi, S. (2016). What Impact does Internet of Things have on Project Management in Project based Firms?. *Asian Business Review*, 6(3), 179-186. <https://doi.org/10.18034/abr.v6i3.520>
- Zareapoor, M., & Shamsolmoali, P. J. P. c. s. (2015). Application of credit card fraud detection: Based on bagging ensemble classifier. 48(2015), 679-685.
- Zareapoor, M., Seeja, K., & Alam, M. A. J. I. j. o. c. a. (2012). Analysis on credit card fraud detection techniques: based on certain design criteria. 52(3).
- Zaslavsky, V., Strizhak, A. J. I., & Security. (2006). Credit card fraud detection using self-organizing maps. 18, 48.