Review on Credit Card Fraud Detection using Data Mining Classification Techniques & Machine Learning Algorithms

Rahul Goyal
Dept. of CSE & IT
MITS
Gwalior, India

Amit Kumar Manjhvar
Dept. of CSE & IT
MITS
Gwalior, India

Abstract— Data mining (DM) involves a core algorithm that enables data deeper than basic insights and knowledge. In fact, data mining is more part of knowledge discovery process. Credit card (CC) providers provide multiple cards to their customers. All credit card users must be genuine and sincere. Giving a card to any kind of mistake can lead to a financial crisis. Due to the rapid growth in cashless transactions, it is unlikely, Fake transactions can also be increased. A fraudulent transaction can be identified by studying credit cards of various behaviors as a previous transaction history dataset. If there is any deviation from the available cost pattern, it is a bogus transaction. DM & machine learning techniques (MLT) are widely applied in credit card fraud detection (CCFD). In this survey paper we show an indication of various widely available DM & MLT for detecting credit card fraud.

Keywords— Data Mining (DM), Credit card fraud , Detection Classification, Techniques & Challenges.

I. Introduction

Data mining is process of finding statistically reliable, anonymous & actionable information. In addition, DM problem needs to be well definite, cannot be explained with query & reporting tools, & can be directed in DM process model. This data must be available, relevant, adequate & clean [1].

The bank is financial institution that receives investments from community. Being vulnerable to any type of fraud becomes a major disqualification for the bank. 'K Chan & J Stolpho et al' note that numerous forms of fraud & financial fraud are ones most affected by bank. Owing to fast-growing online banking activity, we came to know that 44% of US people used these online transactions. 'John T The MistyLook Theme' stated that It is estimated to have loss $ 8.2 billion in 2006 with $ 3 billion in US alone. 'Philip K Keener' says that DM is newly developing machinery that can detect CCF very quickly. Defined by 'Chan & Wei Fan et al' in their opinion, data mining can help us find relationships between hidden patterns & data sets. Fraud or criminal fraud as a result of financial or personnel benefits. Therefore, CCF is use of illegal or complete cards or unusual transaction behavior. As shown in Figure 1, many frauds were found to disturb banks, traders & consumers. Some of these are listed below:

a) Recently distributed mail card.
b) Copy card data through cloned websites.
c) Phishing that hacks credit card numbers & passwords via email.
d) Triangulation In this type of fraud, fraudsters create websites & advertisements that appear to be very cheap. Unknown operators attract those sites & make online transactions. They submit card data to purchase those items. Fraudsters use data on this card to perform the actual transaction.

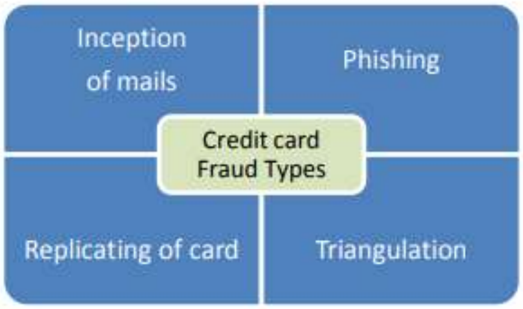

Fig.1 General Types of Credit Card Fraud

It is worth noting that CCF affects merchants the most. Card issuing bank must bear administrative and infrastructure costs. Studies say average time amid fraudulent transaction dates & chargebacks can be up to 72 days, giving fraudster enough time to deal serious harm [2].

Online credit cards or offline transactions for physical cards are used for daily life credit cards for good & services. In physical transactions, a credit card is inserted into a payment machine at the merchant's store to purchase the goods. This mode may not be able to track forged transactions because the attacker already theft a credit card. By online payment mode, attackers have very little data to counterfeit transactions (safe codes, card numbers, end dates, so on [3].

II. Credit Card Fraud

Unauthorized procedure of CC or information deprived of owner's data is called CCF. The dissimilar CCF trick applications & behaviors are related to two groups of frauds. When app fraud occurs, fraudsters apply for a new card from the bank or provide it to companies that use false or other information. A user can file multiple applications with a single usual of describes (named duplicate fraud), or a different user with similar describes (named identity fraud).

Instead, there are practically 4 main types of behavioral fraud: stolen / lost cards, mail theft, fake cards, & 'current card holder does not exist' fraud. When a stolen / lost card fraud occurs, fraudsters steal a credit card or get lost card. Mail theft fraud when a fraudster receives personal information from a bank in the mail before a credit card or original card holder. Fake & Card Holders Fraud & credit card describes are not presented. In past, remote communications can be done using card details via mail, phone or internet. Second, fake cards are created on card data.

A. Classification

Established on the statistics reported by 2012, the countries with the highest risk of credit card fraud are depicted in Figure 2

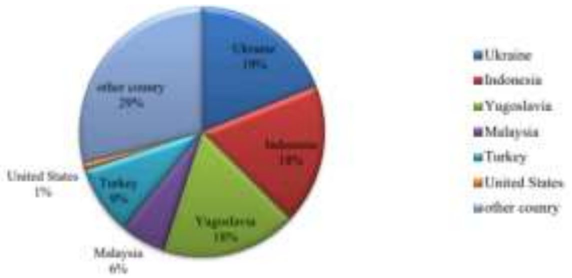

Fig.2 Statistical Classification of Credit Card Fraud Occurrences

Ukraine has the highest rate of fraud at 19%, surveyed in Indonesia at 18.3%.

Afterward these 2, Yugoslavia is most at-risk country at 17.8%. The next highest fraud rate is Malaysia (5.9%), Turkey (9%) & lastly the US. In other states, they are 1% below the rate associated with CCF.

B. Difficulties of Credit Card Fraud Detection:

Fraud detection systems face many difficulties & challenges. An effective fraud detection method should address capabilities & adjust these difficulties.

1) **Imbalanced data:** CCFDs information is of unbalanced nature. This means that entirely CC transactions are fraudulent. Fraudulent transactions are difficult & impossible to detect.
2) **Different misclassification importance:** By fraud detection process, dissimilar diversification errors have dissimilar significance. Typical transaction of abortion is not fraud as fraud. As if you make first mistake, classification will be investigated further.
3) **Overlapping data:** Numerous transactions can be measured fraudulent, but in reality they are common (false positive) &, in addition, fraudulent transaction may seem valid (false negative). Therefore, the key to procurement low rate of false positives & false proposals is fraud detection systems.
4) **Lack of adaptability:** The classification algorithm is the most common problematic of finding new types of normal or deceptive patterns. Monitoring & outdated fraud detection systems are ineffective for detecting new, common & fraudulent practices, respectively.
5) **Fraud detection cost:** System must proceeds in account fraudulent behavior of detected cost & cost of preventing it. E.g., stopping fraudulent transactions of a few dollars and getting no income.
6) **Lack of standard metrics:** There are no standard evaluation criteria to evaluate & compare the results of fraud detection systems.

There are several advantages of using a credit card e.g.**:**

1. Ease of purchase

CCs make life easier. A payment made over the Internet, by telephone, & by an ATM allows customers to borrow credit at a time, place & amount without paying for an efficient payment method.

2. Keep customer credit history

Having good credit history is often key to finding loyal customers. This history is valuable not simply to CCs, but also for other financial services, e.g. loans, rental application or certain jobs. Lender & issuer of credit mortgage companies, CC companies, retail stores & utility companies can evaluation credit scores, timely & responsible customers' history of how well they operate on their loans.

3. Protection of Purchases

Credit cards can provide other protection to customers if they are lost, damaged or stolen. Buyer's CC statement & corporation can ensure that original receipt has been lost or taken. Additionally, specific CC companies offer large purchases for insurance [4].

C. Types of Frauds

This letter covers credit cards fraud, telecommunication fraud, computer penetration, bankruptcy fraud, theft / fake fraud, application fraud & conduct fraud. CCF: CCF is classified in 2 categories:

(1) Offline Fraud: At a call center or other location on a physical card stolen using offline fraud.

(2) Online Fraud: Online fraud is by a cardholder with shopping, Internet, phone, web or absence.

- ➢ Telecom fraud: Use of telecom services for other types of fraud. Its victims are consumers, businesses & communications service providers.
- ➢ Computer Intrusion: Intrusion is distinct a warranty or invasion without entering work; this means "unauthorized attempts to access data, & manipulate data. Infiltrators can be since any environment, outsider (or hacker), & person who recognizes layout of system.
- ➢ Bankruptcy Fraud: This column attentions on bankruptcy fraud. Bankruptcy fraud resources not by CC. One of most complex scams is bankruptcy fraud.
- ➢ Theft Fraud/ Counterfeit Fraud: In this section, we attention on each other's related theft & Counterfeit fraud. Theft fraud states card that is not yours. Once holder gives some feedback & approaches bank, bank will proceed action to investigate thief as soon as likely. Similarly, credit fraud is used remotely when fraud is committed, Wherever CC details are required only.
- ➢ Applications Fraud: Once a person relates to credit card, he or she is given false data, which is called application fraud. Toward detect application fraud, two dissimilar scenarios need to be considered. While apps with the same information from similar user, it is termed duplicate, & when applications derived as of different people by same information, it is called identity fraud. Phua et al. describes application fraud as "demonstration of identity crime, occurs when application forms contain possible, & synthetic (identity fraud), or real but also stolen identity information (identity theft)" [5].

III. Credit Card Fraud Detection Techniques

Following approaches [6] are widely used for CC fraud detection-

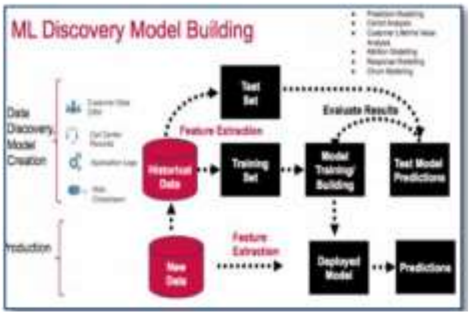

Fig. 3 ML Process in Credit Card Fraud Detection

Genetic algos- Algos are often recommended as fraud prediction methods. An algorithm developed by Bentley is based on genetic software design to create the classification of CC transactions in questionable & non-doubtful classes. Essentially, this method follows scoring procedure. In their study, database consisted of 62 regions with over 4,000 transactions. As similar point of view, training & testing models were utilized. Dissimilar types of rule were verified by different fields. Best rule is to have best prediction. Their technique has proven outcomes of real home insurance data, & is an effective way to combat credit card fraud.

Decision Tree- Decision perspective is a graphical demonstration of probable solution to an option based on positive circumstances. The decision view starts from root node, divided into separate spaces, which are linked to added nodes. Decision tree termination up node is named leaf node. At every node, decision view signifies an experiment, related by branch, representing its outcomes, & leaf node is class of labels. Through this strategic method to differentiation & decision-making, decision perspectives are usually simplified in a complex problem.

Artificial Neural Network (ANN)- ANN is most influential classifiers with different characteristics among hidden patterns. ANN functions similarly to human brain. The first layer is input layer & last layer is output layer. It may have either any number of hidden layers. If neural networks have more hidden layer of stability, it is intensive learning. Each layer has dissimilar neurons & every neuron is associated with heavier edges. Every neuron of output has its private unit of action. This function is named activation function. E.g., various beginning functions are used: linear function, step function, threshold function, sigmoid function, & so on. There is commonly applied function is public sigmoid function.

Convolution Neural Network (CNN)- CNN is measure of intensive education. The feature map represents the hidden layer within the mapping. Each feature map represents a feature. The feature map in the compressing neurons of the process is called convolution. The feature of the sub-sample reduces the map parameters. The fully connected layer is the same neural network.

Outlier Detection- Outlier are basic method of substandard attention that can be applied to detect fraud. An observation that deviates so much from other explanations that it is suspicious another observation is known externally. This model uses unsupervised learning approach. In general, outcome of unread study is new description or demonstration of detected information, followed by better future decisions. Unfeasible approaches do not require prior information of fraudulent & non-fraudulent transactions, but rather sense variations by unfeasible learning behavior & uncommon transactions.

Clustering techniques- Two clustering methods to behavior fraud reported in Bolton & Hand (2002). Peer group study is system that identifies account that act otherwise as of others in a moment. These are some of the accounts that are called suspicious. & then there are cases of fraud. Peer cluster study behind assumption is that if an account is still operating differently for specified period of time, then this account needs for reported. Other method, Breakpoint Analysis, usages another theory that suggests that the card should be investigated if the change in card procedure is on separate beginning.

Logistic Regression- There are more & more statistical models that discriminate data mining functions such as study, regression analysis, & multiple logistic logic. Logistic regression (LR) is a set of predictive variables that are valuable to predicting presence or deficiency of attribute or outcome. This is parallel to linear regression model, but it is suite for model with reliant on variable dichotomies.

Deep learning - Deep Learning is a sophisticated technology that has recently attracted the attention of IT circles. Deep Learning Theory is an ANN with many hidden layers. In contrast, deep learning forward neural networks have only one hidden layer.

Rule based method- Association rules have been created by perceive fraud-based transactions & common transactions. In fraud detection, the rules created can be applied to categorize fraud & legal relations. There are rules for created behavior. This technique is related to the decision perspective.

Hidden Markov Model (HMM)- The HHM is modeling of hybrid, embedded stochastic procedure. This generalized process of complexity exceeds the Markov model. If the learner with a high potential probability does not approve the hidden Markov model bank transaction, this is measured dangerous & fake transaction. Baum Welch algo is applied for model learning, & K-Means algo to data classification. Model categorizes transactions in high, average, & low levels.

IV. Challenges In Credit Card Fraud Detection

1) **Data deficiency-** Basically, CCFD scientifically addresses biggest problem of real-time data exploration, due to the confidentiality of the problem [7]. However, investigators are not discouraged because they can frequently perform scientific work by an industrial partner. Additionally, some people suggest using synthetic data that mimics the transactions of datasets.
2) **Behavioral variation-** Fraudulent behavior to avoid detecting allergies over time.

V. Literature Survey

Anuruddha Thennakoon et al. [2019] Real-world transactions in four major fraud cases. Every scam is solved by ML model, & best way is through valuation. This evaluation gives comprehensive guide to selecting the optimal algo for the types of scams & weights we consider to be most appropriate mitigation measures. Another key part that we statement in our project is real-time CCFD. To do that, we usage predictive analytics to determine whether particular transaction to machine learning models & API module is real or fraudulent. We are also evaluating new approach that addresses distorted distribution of data. Information applied in our experiments are as of confidential disclosure agreement [8].

Chunzhi Wang et al.[2018] BP Neural Network, a fast tracking system that optimizes the BP neural network, is based on solving slow convergence rate problems, which can result in local optimal, network outages, & poor system stability. Using the Whale Group Optimization algo to enhance weight of BP network, we first procedure WOA algo to obtain primary value, & Next BP network algo to precise fault values. The optimal values are obtained [9].

Sahil Dhankhad et al. [2018] In numerous supervised ML algos, detect CC counterfeit transactions & execute real-world datasets. In addition, we use these algos to implement super classifier by embedded learning approaches [10].

Krishna Modi et al. [2017] previous transaction data from customers analyzing cost behavior. If there is any deviation from the available cost pattern, it is a bogus transaction. Banks & credit card companies use a variety of data mining methods to detect fraud, e.g. decision perspectives, rule-based mining, NN, fuzzy clustering approaches, hidden Markov models or hybrid approaches. Both of that approaches are used to explore common usage patterns based on the past actions of the customers. This paper provides comparison of numerous methods for detecting fraud [11].

Zahra Kazemi et al. [2017] To remove the best features from credit card transactions, suggest an in-depth auto encoder, and then add a soft max network to the class label. Such data using super-complete auto encoder can be used to map to a large amount of space, & the sparse model can be useful for classifying targets [12].

Kosemani Temitayo Hafiz et al. [2016] focus on building scorecards as of relevant assessment principles, aspects & abilities of prognostic analytics vendor solutions presently utilized to CCFD. Record gives simultaneous comparison of five vendor CC prediction analytics vendor solutions in Canada. Confirming study results, list of CCFs has outlined PAT vendor's tests, threats & restrictions [13].

V.Mareeswari et al. [2016] Owing to limitations of current system, this paper suggested new algo with present algo. Limitations of current scalability issues, highly unbalanced classes, & time constraints. Fraud detection for community & spike detection using CC application hybrid support vector machine (HSVM). HSVM is commonly utilized technique to pattern recognition & classification [14].

Fahimeh Ghobadi et al. [2016] Progress CCFD Model Based on ANN & Meta Cost Process to Improve Risk & Loss. ANN strategy was used for credit card fraud prevention and detection. Due to unbalanced nature of information (fraud & non-fraud cases), fraud can be difficult to detect. Added Meta Cost Process to deal with issue of unstable information. Cost Sensitive NN (CSNN) is based on abuse detection method. Based on the comparison of the Artificial Immune System (AIS), this model found cost savings & growth rates. This study on data was derived from a large Brazilian credit card issuer who provided real transaction data [15].

VI. Conclusion

In the current paper, credit card investigations have been conducted on various methods of detecting fraud. First, it stated the importance of the topic & mentioned the current shortcomings in traditional practices. Counterfeit transactions have different levels of risk, & they must find ways to quickly & accurately detect high-risk transactions. Typical data mining methods are not sufficient to identify these transactions. Advanced algorithms should be used to find the best answer.

References

[1] Clifton Phua, Vincent Lee, Kate Smith & Ross Gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research", https://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf 2014.

[2] Suman, Mitali Bansal, "Survey Paper on Credit Card Fraud Detection", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), ISSN: 2278 – 1323, Pp 827-832, Volume 3 Issue 3, March 2014.

[3] Bilonikar Priya, Deokar Malvika, Puranik Shweta, Sonwane Nivedita4, Prof.B.G.Dhake "Survey on Credit Card Fraud Detection Using Hidden Markov Model", International Journal of Advanced Research in Computer & Communication Engineering, ISSN (Online) : 2278-1021, ISSN (Print) : 2319-5940, Vol. 3, Issue 5, Pp 6445-6448, May 2014.

[4] Samaneh Sorournejad, Zahra Zojaji, Reza Ebrahimi Atani, Amir Hassan Monadjemi, "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective", sorournejad@yahoo.com, 1611.06439, Pp 1-26.

[5] Suman, Nutan, "Review Paper on Credit Card Fraud Detection", International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 7–July 2013, ISSN: 2231-2803 , Pp 2206-2215.

[6] Vipul Patil, Dr. Umesh Kumar Lilhore "Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection", International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2018 IJSRCSEIT, | Volume 3 | Issue 5 ISSN : 2456-3307, [(3)5 : 320-325.

[7] Maja puh, Ljiljana Brkic, " Detecting credit card fraud using selected machine learning algorithms", maja.puh@fe.hr, ljiljana.brkic@fe.hr, MIPRO 2019, May 20-24, 2019, Opatija Croatia, Pp 1250-1255.

[8] Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, Nuwan Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning", *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence).* 2019, IEEE, pp. 488-493.

[9] Chunzhi Wang Yichao Wang Zhiwei Ye Lingyu Yan Wencheng Cai Shang Pan, "Credit card fraud detection based on whale algorithm optimiz HG BP neural network", The 13th International Conference on Computer Science & Education (ICCSE 2018), 978-1-5386-5495-8/18/$31.00 ©2018 IEEE, Pp 614-617.

[10] Sahil Dhankhad, Emad A. Mohammed, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study", 2018 IEEE International Conference on Information Reuse and Integration for Data Science, 978-1-5386-2659-7/18/$31.00 ©2018 IEEE, DOI 10.1109/IRI.2018.00025, Pp 122-125.

[11] Krishna Modi, Reshma Dayma, "Review On Fraud Detection Methods in Credit Card Transactions", 2017 International Conference on Intelligent Computing and Control(I2C2017),krishnamodi1994@gmail.com,ceradayma@gmail.com .

[12] Zahra Kazemi, Houman Zarrabi, "Using deep networks for fraud detection in the credit card transactions", IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI) I Dec. 22QG, 2017, 978-1-5386-2640-5/17/$31.00 ©2017 IEEE, Pp 0630-0633

[13] Kosemani Temitayo Hafiz, Dr. Shaun Aghili, Dr. Pavol Zavarsky, "The Use of Predictive Analytics Technology to Detect Credit Card Fraud in Canada", tkoseman@student.concordia.ab.ca,{shaun.aghili,pavol.zavarsky}@concordia.ab.ca, 2016, Pp 1-6.

[14] V.Mareeswari, Dr G. Gunasekaran, "Prevention of Credit Card Fraud Detection based on HSVM", International Conference On Information Communication And Embedded System(ICICES 2016), 978-1-5090-2552-7.

[15] Fahimeh Ghobadi, Mohsen Rohani, "Cost Sensitive Modeling of Credit Card Fraud Using Neural Network Strategy",Ghobadi.Fahimeh@Gmail.com,m.Rohani@niopdc.ir, 2016.