ADAPTIVE FRAUD DETECTION THROUGH MACHINE LEARNING

[1]Muthu Dayalan
Researcher & Software Professional
Chennai & Tamil Nadu

Abstract- ***Machine learning relies on the technique of identification of the influential cause-and-effect relationship, for making accurate future prediction. It involves the application of artificial intelligence to enable the computer learn without subjection to artificial programming. Machine learning is an effective technique for improving risk management through enhanced detection of the frauds and compliance of control violations. The machine-learning model applies accumulated historic data to pinpoint the fraudulent transactions. The strengths, which makes machine learning an effective fraud detector, include facilitating real time decision-making, improved accuracy and a rapid response to change. The assembles used for detecting fraud include regression analysis, artificial neutral networks, and decision trees.***

Keywords-machine learning, fraud detection, model, prediction, pattern recognition

I. INTRODUCTION

Machine learning relies on a simple concept. Through identification of the most influential cause-and-effect relationship in the past, the machine could 'learn' to make future accurate predictions. It implies that the application of the artificial intelligence is to enable the computers to learn without being subjected to an explicit programming. This learning and prediction rely on a complex statistical technique and high-octane computing power. The machine learning is initiated by high-pow computers, which are instructed by the human intelligence through huge historic data points for identification of the historic data points. The whole information is fed into a wide range of algorithms. In the recent past, the big data detection techniques have been applied in improving the risk management through enhanced detection of the frauds, and compliance of control violations. The advancement in machine learning is vital and applicable in the boosting the alerts accuracy and develop actions which could be relied and acted upon. Due to its accuracy in patter recognition in data, it is also equally accurate in recognition of the anomalies in these patterns. Therefore, it is considered as a perfect approach in preventing fraud. Under this background, this paper conducts a comprehensive analysis of the adaptive fraud detection techniques by the use of machine learning.

II. MACHINE LEARNING IN FRAUD DETECTION

The machine learning has been considered as an effective measure of the fraud detection.[1] There is a great data, which is transferred during the online transactions, which results in the binary results. These may be either genuine or fraudulent. The online business identifies the fraudulent activities because they receive chargebacks on them. However, the chargebacks are initiated after the transactions, hence being reactive rather than proactive.

Machine learning works based on large and historic dataset, accumulated from a wide range of clients and industries. Even the companies that undertake little transactions can take control of the full range of their dataset, which would allow them to access accurate data on every transaction. The aggregated data offers a highly accurate training data set, which would be allowed by the business to choose the right model for optimization of the levels of call and precision. Among all the transactions conducted, the model can pinpoint the fraudulent or recall, the proportion of these is precision [2]

There are features, which are constructed within the dataset. These are the data points such as the customer's account value, age, credit card origin among others. There could be many features, which contribute, in different extents, the probability of a fraud. Each feature has a peculiar contribution to the fraud score, which is determined and generated by the artificial intelligence machine. The intelligence machine is driven by the training set. Therefore, these features embedded in the machine learning based system make it possible for identification of the significant fraud contributors.

A. Strengths making machine learning a powerful fraud detector

Facilitating real-time decision-making- machine learning facilitates creation of ad hoc rules that determine the orders that could be rejected or accepted, which require time consumption, and which require manual interactions.[3] As a result, the machine learning is vital in evaluating a large amount of transactions in real time.

Improve accuracy-In the recent past, the level of crime has become more sophisticated, and increasingly adept in disguising fraud. In this case, the machine learning is considered more efficient as compared to human being, in the process of detection of non-intuitive patterns in the identification of fraudulent transactions. Through improved accuracy, machine learning could help avoid false positives, which implies the good orders, which are erroneously indicated as fraudulent.

[1] "Mobileum Announces Anti-Fraud Analytics that Discovers Telecom Frauds before they Occur." *PR Newswire*, Mar 03, 2015, *ABI/INFORM Collection*.

[2] "Equifax Offers Tips to Detect and Deter Fraud during Customer Acquisition."*PR Newswire*, Mar 29, 2016

[3] LaComb, Christina, John Interrante, and Kareem S. Aggour. "Monitoring Key Company Events through Deliberative Learning." *Information Systems and eBusiness Management*, vol. 5, no. 4, 2007, pp. 295, 0044-7.

Rapid response to change-Fraudsters is active in changing their fraud tactics, in the form of a constant cat-and-mouse game [4]. However, the machine algorithms could be applied in a continuous analysis and processing of data, as well as updating its models to detect the latest trends by the fraudsters.

Lower costs-technology advancement is a significant aspect of machine learning. This technological advancement has contributed significantly to reduction of the cost involved in the machine learning solutions and computer systems for running them.[5] In addition to improving accuracy, the machine learning reduces the cost of false positives, time and manual review expenses.

B. Machine Learning ensemble for detecting fraud

There are wide techniques used in machine learning, which range from basic to sophisticated. However, there are few of these techniques, which are effectively applicable and useful in the fraud management. These are discussed below

Regression analysis

Regression analysis is a long-standing statistical technique, which assesses the cause-and-effect relationship strength in a structured data set.[6] Due to the variable numbers and data set size, the regression analysis is considered sophisticated in the application of fraud detection. Regression analysis is valuable in the assessment of the prediction power of the individual variables during the process of fraud detection[7].

Artificial neural networks

The artificial neural networks imitate the working of the brain through establishing interconnected networks with a large layer of neurons. Each layer of a neuron is made up of combined inputs and functions. In fraud detection, it relies on the transformation layer, which is applied to convert the raw data into a meaningful information, which could be applied in the neural networks and processes. However, it is critical to note that though the neural networks are effectively applied in finding a suitable interaction between variables, they are not in a position to explain why a given score is produced.

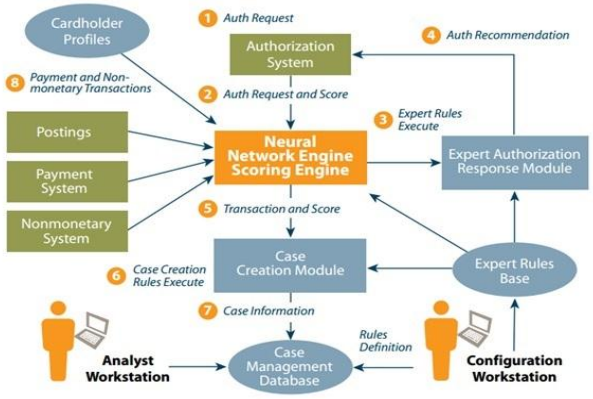

Figure 1: Artificial neural networks

Decision trees

The decision trees consist of the decision points, which predict varying data elements, where data is groped into minor groups.[8] The end of a tree path represents the outcome and accuracy level predicted. In the case of fraud detection, the decision tree enables the use of unstructured data with reduced transformations. More importantly, a readily available insight is provided by the logic constructed behind the score. Some decision trees techniques, which could numerously be applied together, include boosted trees, random forest, and stochastic gradient boosted trees.

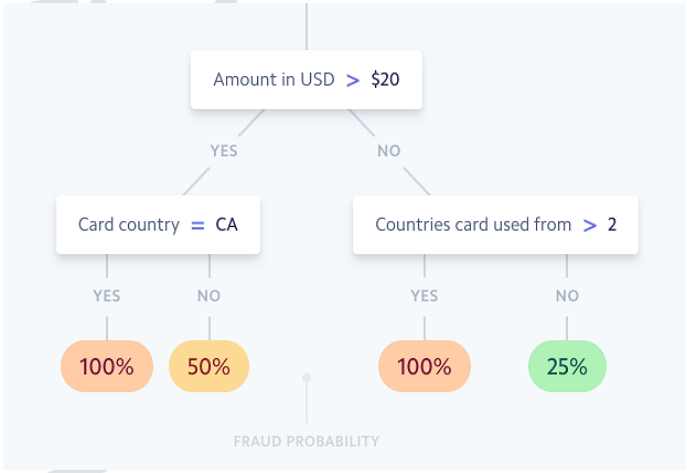

Figure 2: Decision trees

Random Forest

The random forest technique applies a wide range of multiple decision trees techniques to enhance the performance and regression analysis or classification of the machine learning applications. The random forest makes it easy to smoothen the error which might be existing in the single tree.[9] More importantly, a random forest is applied in the fraud detection process to increase the model performance and accuracy, as well as maintain the ability for interpreting the results and provision of scores which could be

[4] Smith, Kate. "Intelligent Design." *Best's Review*, no. 8, 2014, pp. 56-57.

[5] Zhang, Mei. "Evaluation of Machine Learning Tools for Distinguishing Fraud from Error." *Journal of Business & Economics Research (Online)*, vol. 11, no. 9, 2013, pp. 393.

[6] Li, Jing, et al. "A Survey on Statistical Methods for Health Care Fraud Detection." *Health Care Management Science*, vol. 11, no. 3, 2008, pp. 275-87,

[7] Ryoo, Jungwoo. "Machine Learning and Big Data Know it Wasn't You Who just Swiped Your Credit Card." *Scientific Computing*, 2015.

[8] Violino, Bob. "Machine Learning Proves its Worth to Business." *InfoWorld.Com*, 2017.

[9] Ai, Jing. *Supervised and Unsupervised PRIDIT for Active Insurance Fraud Detection*, The University of Texas at Austin, Ann Arbor, 2008.

explained to the users[10]. One aspect of the adaptive fraud detection is that they are very fast and could effectively be used to deal with missing and unbalanced data. However, the random forest has weaknesses in that when they are applied to carry out the regression analysis; they could not predict extensively beyond the training data range. Further, they may also outfit the data set which is quite noisy. This makes the critical consideration of algorithm as the effectiveness in which it works with the concerned data.

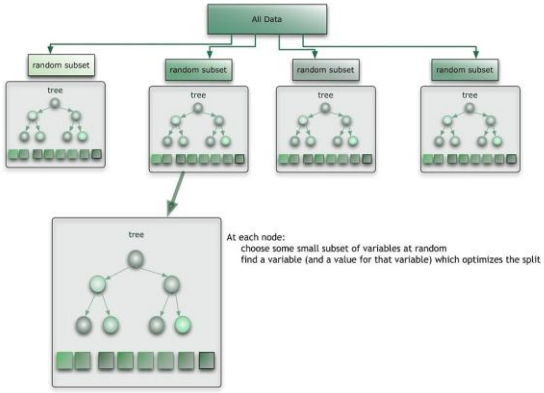

Figure 3: Random Forest

C. Restrictions on using Machine Learning for Detecting Fraud

Fundamentally, this learning in question does not characterize as a universal remedy for revealing fraudulent activities. Instead, it serves as an extremely advantageous know-how that enables individuals to discover configurations of different inconsistencies in routine business operations. Undoubtedly, this method on the agenda tends to be more advanced in comparison to the other approaches that the former establishments utilized to uncover irregularities. Ideally, some of these other approaches under discussion include conducting manual assessments and the use of systems based on rules. Nonetheless, just like the other methods, the machine learning for fraud detection also possesses several limitations that hinder its usability. They include:

Absence of inspectability

In most technical laboratories, the managers preserve the server-side machine-learning prototype for their users. For this reason, these labs in question are usually obligated to elucidate the motives for a purchaser or supplier being signposted as a swindler and hence prohibited from operating the scheme in hand. Additionally, this kind of accountability is also important since it aids their clients in sanctioning shams and thereby indoctrinate the structure. This kind of learning under discussion tends to solely be as noble as its inventors, who were mostly the anthropoid statistics technologists. In other words, this means that not even the most advanced technology could substitute both the proficiency and conviction it requires to successfully screen and sorts out documents, in addition to assessing the connotation of the danger mark[11]. As such, while technical laboratories have effectively eradicated this potential setback in hand via rule-centered procedures, the method's deficiency of inspectability could pose as a shortcoming of other particular machine techniques founded on learning.

Inauspicious inception

These machine-learning models tend to consume a considerable amount of information for them to be precise. Ideally, this drawback mainly affects smaller institutions, which possess lesser capacities of files. Ideally, this is because, for outsized corporations, this file size on the agenda is not usually a problem [12] However, for the former, they tend to lack sufficient points of records required to ascertain the authentic base and outcome associations. Consequently, the absence of the correct facts and figures makes the machines in question to pick up the incorrect interpretations thereby making either faulty or inappropriate counterfeit valuations.[13] For this reason, it is frequently nobler to employ a basic array of guidelines originally and permit the prototypes under discussion to limber up with additional information. Most of the times, this methodology is put into operation with less significant sets of data.

Sightlessness to networks in data

In essence, these machine-learning archetypes tend to act on deeds, conducts, and ventures. At the outset, in cases where the sets of information tend to be small, the models in hand are usually indiscriminate to systems in records. Furthermore, these prototypes could also deliberately ignore an apparently palpable association.[14] This kind of connection on the agenda could include a mutual identification in the middle of two financial records. In turn, to respond to this limitation, technical workrooms tend to augment their archetypes using Chart complexes. Ideally, the use of these particular approaches assists in that they aid in the location of numerous counterfeit performers for each one prohibited by recording [15] In addition to this, databases from the networks under discussion also sanction technicians to obstruct both dubious and phony records prior to their engagement in duplicitous activities.

REFERENCES

[1] "Equifax Offers Tips to Detect and Deter Fraud during Customer Acquisition."PR Newswire, Mar 29, 2016.

[2] "Kaspersky Fraud Prevention Cloud Enables Machine Learning and Big Data Analysis for Enhanced Multi-Channel Protection." Al Bawaba, May 24, 2017.

[3] "Mobileum Announces Anti-Fraud Analytics that Discovers Telecom Frauds before they Occur." PR Newswire, Mar 03, 2015, ABI/INFORM Collection.

[10] Perols, Johan. "Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms." *Auditing*, vol. 30, no. 2, 2011, pp. 19-50,

[11] Krambia-Kapardis, Maria, Chris Christodoulou, and Michalis Agathocleous. "Neural Networks: The Panacea in Fraud Detection?" *Managerial Auditing Journal*, vol. 25, no. 7, 2010, pp. 659-678.

[12] Ehramikar, Soheila. *The Enhancement of Credit Card Fraud Detection Systems using Machine Learning Methodology*, University of Toronto (Canada), Ann Arbor, 2000.

[13] Sudjianto, Agus, et al. "Statistical Methods for Fighting Financial Crimes."*Technometrics*, vol. 52, no. 1, 2010, pp. 5-19.

[14] "Kaspersky Fraud Prevention Cloud Enables Machine Learning and Big Data Analysis for Enhanced Multi-Channel Protection." *Al Bawaba*, May 24, 2017.

[15] Viaene, Stijn, et al. "A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection." *Journal of Risk and Insurance*, vol. 69, no. 3, 2002, pp. 373-421.

[4] Ai, Jing. Supervised and Unsupervised PRIDIT for Active Insurance Fraud Detection, The University of Texas at Austin, Ann Arbor, 2008.

[5] Ai, Jing, et al. "A ROBUST UNSUPERVISED METHOD FOR FRAUD RATE ESTIMATION." Journal of Risk and Insurance, vol. 80, no. 1, 2013, pp. 121-143.

[6] Ehramikar, Soheila. The Enhancement of Credit Card Fraud Detection Systems using Machine Learning Methodology, University of Toronto (Canada), Ann Arbor, 2000.

[7] Fawcett, Tom, and Foster Provost. "Adaptive Fraud Detection." Data Mining and Knowledge Discovery, vol. 1, no. 3, 1997, pp. 291-316.

[8] Hodge, Victoria, and Jim Austin. "A Survey of Outlier Detection Methodologies."The Artificial Intelligence Review, vol. 22, no. 2, 2004, pp. 85-126.

[9] Jans, Mieke, Nadine Lybaert, and Koen Vanhoof. "A Framework for Internal Fraud Risk Reduction at IT Integrating Business Processes: The IFR^sup 2^ Framework." International Journal of Digital Accounting Research, vol. 9, 2009, pp. 1-29.

[10] Jha, Sanjeev. Credit Card Fraud Detection with Discrete Choice Models and Misclassified Transactions, University of Illinois at Chicago, Ann Arbor, 2009.

[11] Krambia-Kapardis, Maria, Chris Christodoulou, and Michalis Agathocleous. "Neural Networks: The Panacea in Fraud Detection?" Managerial Auditing Journal, vol. 25, no. 7, 2010, pp. 659-678.

[12] LaComb, Christina, John Interrante, and Kareem S. Aggour. "Monitoring Key Company Events through Deliberative Learning." Information Systems and eBusiness Management, vol. 5, no. 4, 2007, pp. 295, 0044-7.

[13] Li, Jing, et al. "A Survey on Statistical Methods for Health Care Fraud Detection." Health Care Management Science, vol. 11, no. 3, 2008, pp. 275-87,

[14] Perols, Johan. "Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms." Auditing, vol. 30, no. 2, 2011, pp. 19-50,

[15] Ryoo, Jungwoo. "Machine Learning and Big Data Know it Wasn't You Who just Swiped Your Credit Card." Scientific Computing, 2015.

[16] Smith, Kate. "Intelligent Design." Best's Review, no. 8, 2014, pp. 56-57.

[17] Sudjianto, Agus, et al. "Statistical Methods for Fighting Financial Crimes."Technometrics, vol. 52, no. 1, 2010, pp. 5-19.

[18] Viaene, Stijn, et al. "A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection." Journal of Risk and Insurance, vol. 69, no. 3, 2002, pp. 373-421.

[19] Violino, Bob. "Machine Learning Proves its Worth to Business." InfoWorld.Com, 2017.

[20] Zhang, Mei. "Evaluation of Machine Learning Tools for Distinguishing Fraud from Error." Journal of Business & Economics Research (Online), vol. 11, no. 9, 2013, pp. 393.