ISSN: 2754-6659

Journal of Artificial Intelligence & Cloud Computing

SCIENTIFIC Research and Community

Review Article

Open Access

Credit Card Fraud Detection Using Data Science

Chandra Mouli Yalamanchili

USA

ABSTRACT

We have witnessed an enormous evolution in credit card processing over last few years, issuing chip-based credit cards, starting mobile device-based wallets like Apple Pay are some of the significant changes done to secure credit card transactions.

Despite financial institutions (banks) working hard to eliminate fraud in credit card transactions, credit card fraud has been continuously rising over the last few years. Fraudsters are getting smarter and using latest technologies to steal cardholder's information, either through hacking or through social engineering.

Increasing fraud in the industry makes fraud prediction very critical to be able to identify and stop fraud in real time, and data science plays a significant role in analyzing and being able to predict fraud based on transactional and cardholder information. The scope of this project is to research and identify different types of predictive analysis algorithms available that can be applied to determine and stop fraudulent transactions.

***Corresponding author**

Chandra Mouli Yalamanchili, USA.

Received: March 08, 2023; **Accepted:** March 15, 2023; **Published:** March 22, 2023

Keywords: Credit Cards, Debit Cards, Fraud, Prediction, e-Commerce, Neural Network

Introduction

Credit card processing is one of the fast-growing industries due to rapid advances in technology and with more and more customers switching to use credit cards instead of cash for purchases. Innovations like mobile wallets provided by Apple, Google, and other major technology firms have played an enormous role in increased usage of credit cards in recent years.

On a very high level, credit card transactions can be of two types, card present, and the card not present transactions. Card present transactions are the transactions from retail stores or gas stations where cardholder is present during the transaction, and that makes fraud a little bit difficult as the fraudster must either steal the physical card or copy the card details, to create a duplicate card. Fraud in card present transactions has reduced in recent years due to the introduction of chip cards (challenging to copy and reproduce) and increased usage of mobile wallets which have the same security as chip cards. That leaves us with the card not present transactions, where we are seeing an increased number of fraudulent transactions in recent years. These are usually e-commerce or online portal-based transactions. In this case, fraudsters needed very less information about the physical card and cardholder to perform the transactions.

Fraud transactions can be of different types, below are some examples of fraudulent transaction types

- **Merchant Fraud:** Merchant POS device is compromised and used to run fraudulent transactions.
- **Application Fraud:** Fraudster applying for a new credit card on behalf of the cardholder.
- **Counterfeit Card Fraud:** Usually committed through skimming. Information from the card is stolen and used to create a fake magnetic stripe card with stolen data.
- **Lost/Stolen Fraud:** Transactions are performed using the cards that are either stolen from the cardholder or lost by the cardholder.
- **Not Received as Issued (NRI):** Fraudsters intercepts the mail and steal the credit cards issued to the cardholder.

Any fraudulent transaction will add liability to different parties in the transaction flow like the merchant, merchant processor, networks like Visa/MasterCard, issuing processor, issuing bank and even cardholder depending on who was the weak link for that transaction.

Although financial institutions are working hard to eliminate fraud in credit card transactions, it has been continuously rising as fraudsters are using the latest technologies to steal cardholder's information, either through hacking or through social engineering.

We can detect these fraudulent transactions by analyzing parameters from different segments of information like transactional information, historical information, etc. Also, considering the liability burden on banks, it is critical to be able to identify these fraudulent transactions in real time.

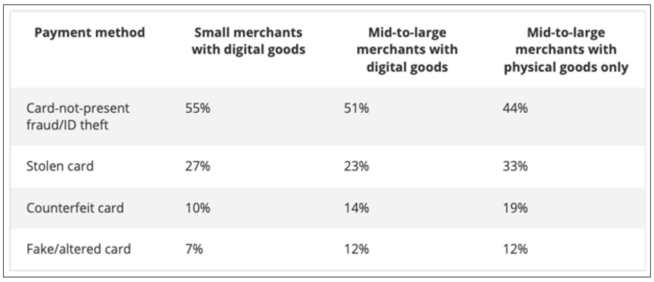
Capabilities of data science will add a great value to predict fraudulent transactions in real time and help financial institutions in preventing fraud. There are several techniques within data science to predict fraud. The goal of this project is to review and understand several fraud prediction models that are already documented.

Credit Card Fraud Statistics

There have been massive data breaches in the past few years, and these data breaches will make cardholder information available to fraudsters, subsequently increasing fraud. Below are some of the well-known data breaches happened in recent years

- Yahoo data breach in 2016 has impacted 3 billion user accounts [1].
- Equifax data breach in 2017 has affected close to 150 million users [2].
- Identity theft has been close to 400,000, during 2021 [3].

Below statistics from figure 1 shows the card-not-present being primary area of fraud in different market segment, the trend of the fraudulent transaction from the card present transaction has shifted to CNP after 2015 with introduction of EMV secure payments [2].



Payment method	Small merchants with digital goods	Mid-to-large merchants with digital goods	Mid-to-large merchants with physical goods only
Card-not-present fraud/ID theft	55%	51%	44%
Stolen card	27%	23%	33%
Counterfeit card	10%	14%	19%
Fake/altered card	7%	12%	12%

Figure 1: This Image Depicts the Distribution of US Retail Merchants Fraud Losses by Payment Method from Year 2020 [2].

Conventional Fraud Detection Applications

Currently, we have so many fraud prediction applications some of them are rules-based, and some of them are score based. These solutions use transactional and historical information to come up with fraud prediction. Below are a few drawbacks I have noticed with these types of applications

- **Rules Based:** Complex to build and manage the rules, we have seen clients setting up bad rules resulting in false positives impacting cardholders [1].
- **Score Based:** These applications use cardholder's shopping pattern, distance from the location of the previous transaction, etc. to come up fraud score indicating how risky the transaction is. Primary issue I have noticed with this solution is that a new model would take a minimum of 3 months to be ready for production.
- **Common:** Both applications are highly dependent on human support who has very good domain knowledge.
- **Common:** Some of the fraud prediction applications predict the fraud after the transaction got processed, even though it would stop subsequent fraudulent transactions, cardholder is already impacted for that first fraudulent transaction.

How Can Data Science Help Prevent Fraud?

We need a robust fraud detection system that can accommodate all the complexities involved with credit card transactions like high volume processing, volatility, variety of transactions, and criticality, and be able to consider the vast number of attributes available in transactional or historical data and predict fraudulent transactions with high precision in real time.

The current solution of static rules-based fraud prediction tools won't stand a chance before rapidly evolving credit card industry as well as increasing fraud in the industry.

There are several machine learning algorithms that can be used to implement fraud prediction, I have mentioned few of them here based on the research done through several articles.

Hidden Markov Model (HMM): HMM is a statistical Markov model with finite set of hidden states. Only the outcome is visible to an external observer and so this can be a good solution for credit card fraud detection as it does not require fraud signatures and still it is capable to detect frauds just by bearing in mind a cardholder's spending habit [4]

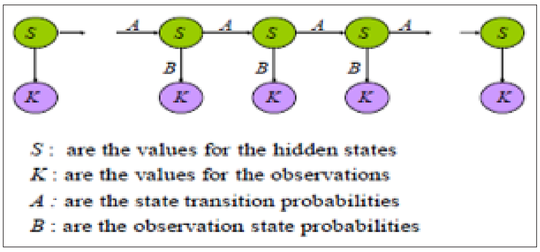

S
A
B
K
S : are the values for the hidden states
K : are the values for the observations
A : are the state transition probabilities
B : are the observation state probabilities

Figure 2: Image Depicting HMM Finite Set of States. Picture Created by Shailesh S Dhok for 'Credit Card Fraud Detection Using Hidden Markov Model' Research Paper [4].

Convolutional Neural Network (CNN): CNN is a class of deep neural network that is most used for detecting credit card fraud [5].

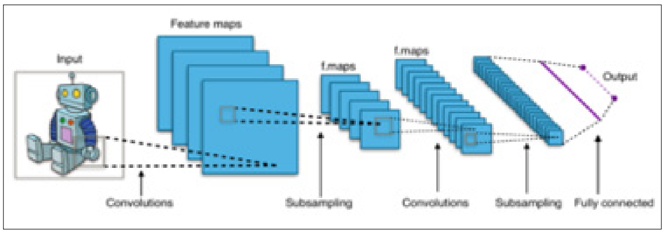


Figure 3: Image Depicting Typical CNN Architecture. Picture Created by Wikipedia for 'Convolutional Neural Network' Page on Wikipedia [5].

CNN models have multiple hidden layers, and they are on lower extreme with regards to complexity as they have advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. They have mechanism to avoid the model over-fitting [5].

Machine learning with neural network models is the best solution for fraud prediction applications as they can continue to learn new trends in the transactions and predict fraudulent transactions more precisely [6].

Apache Spark is interesting as it is analytics engine built to process large scale transactions in real time. Below is a potential data flow structure showing Fraud detection implementation using Apache Spark and Event Streaming [7,8].

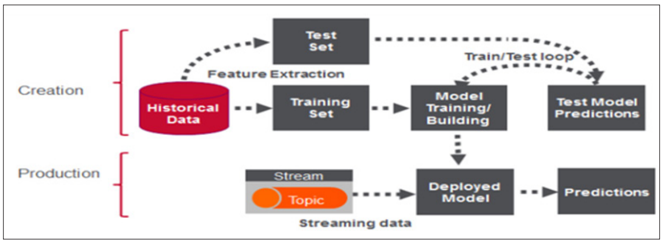


Figure 4: Image Depicting Two Phase Approach for Real Time Fraud Detection Application. Picture Created by Carol McDonald for Real-Time-Credit-Card-Fraud-Detection-Apache-Spark-and-Event-Streaming Article on mapr.com [7].

Conclusion

Credit card fraud is a growing concern in today's world with both credit card usage and credit card fraud on the rise. As shown in the stats, the number of data breaches, identity theft cases, and credit card fraudulent transactions are rising at an alarming level.

Data science plays a significant role in improving the fraud prediction tools we are currently using, by analyzing credit card transactions and being able to predict fraud based on transactional data, cardholder data, and historical data.

Machine learning algorithms running on neural networks should replace the current static rules-based fraud detection products to improve the fraud prediction precision and to stay up to speed with market trends in detecting new strategies of fraudsters and stopping them.

References

1. Edward J McAndrew (2018) The Hacked & the Hacker-for-Hire: Lessons from the Yahoo Data Breaches (So Far). Available at: https://natlawreview.com/article/hacked-hacker-hire-lessons-yahoo-data-breaches-so-far.
2. Jason Steele (2021) Credit card fraud and ID theft statistics. Available at: https://www.creditcards.com/statistics/credit-card-security-id-theft-fraud-statistics-1276/.
3. Taylor Schulte (2022) Identity Theft and Credit Card Fraud Statistics 2022. Available at: https://www.definefinancial.com/blog/identity-theft-credit-card-fraud-statistics/.
4. Dhok SS, Bamnote GR (2012) Credit Card Fraud Detection Using Hidden Markov Model. International Journal of Advanced Research in Computer Science 3: 816.
5. Wikipedia (2023) Convolutional neural network. Available at: https://en.wikipedia.org/wiki/Convolutional_neural_network.
6. Ghosh S, Reilly DL (1994) Credit card fraud detection with a neural-network. In System Sciences. Proceedings of the Twenty-Seventh Hawaii International Conference 3: 621-630.
7. Carol McDonald (2020) Real Time Credit Card Fraud Detection with Apache Spark and Event Streaming. Available at: https://developer.hpe.com/blog/real-time-credit-card-fraud-detection-with-apache-spark-and-event-stream/.
8. Brian Fung (2018) Equifax's massive 2017 data breach keeps getting worse. Available at: https://www.washingtonpost.com/news/the-switch/wp/2018/03/01/equifax-keeps-finding-millions-more-people-who-were-affected-by-its-massive-data-breach/?noredirect=on&utm_term=.0b854d8c3526.