Combating Fraudulent Transactions: How Data Science Is Revolutionizing the World of Finance

Ajit Kolekar
Bellevue University
akolekar@my365.bellevue.edu

Arun Thangaraj
Bellevue University
athangaraj@my365.bellevue.edu

Derrick Lee
Bellevue University
derlee@my365.bellevue.edu

Saurabh Shrestha
Bellevue University
saushrestha1@my365.bellevue.edu

Abstract
Data Science is a new wave in the Technological field and it is used everywhere. It is an interdisciplinary field that comprises Data Analysis models, Machine Learning algorithms, and statistics for scientific methods. The goal of this paper is to elaborate more on Data Science use in the financial sector, focusing on detecting and/or predicting fraud in the banking industry, and also learning what necessary planning methods are used to create a system to tackle banking fraud. For our project, we will research on key technologies involved in classifying transactions as either legitimate or fraudulent, based on transaction details, such as amount, merchant, location, time, and many more; identity fraud, where someone impersonates you and uses your personal information to steal money; phishing, where internet banking clients receive deceptive emails asking them to give account login, password, and personal details on a website which looks like the bank's legitimate website.

Author Keywords
Fraud Detection; Financial Fraud; Machine Learning; Classification Models; Naïve Bayes

Introduction

The official dictionary definition of fraud is "deceiving others for personal gain". Fraud in the financial sector dates back to the year 300 B.C., where borrowed money with interest was not paid. In today's era, fraud in the financial sector could encompass a wide variety including credit card fraud, insurance fraud, mortgage fraud and so on. [6] Financial Fraud can be committed through many methods, including mail, wire, phone, and the internet. The difficulty of checking identity and legitimacy online, and the ease with which hackers can divert browsers to dishonest sites and steal credit card details, the international dimensions of the web and ease with which users can hide their location, all contribute to making internet fraud the fastest growing area of fraud. Many fraud cases involve complicated financial transactions conducted by 'white collar criminals', business professionals with specialized knowledge and criminal intent. With the advent of Data Science and use of modern machine learning algorithms, we are trying to classify financial transactions as either legitimate or fraudulent, and detect other financial fraud activities involving identity theft and spoofing. All these fraudulent activities may make people too afraid to shop online. Credit card fraud affects not only the victims, but also the credit card companies and merchants. Some individuals do not realize they have become a victim of fraud, thus making the true extent of fraud difficult to determine.

Application to Data Science

Financial fraud has been one of the easy targets for fraudsters. The credit card fraud has been increasing consistently over the last few years as shown in figure 1. The credit card numbers and debit card numbers can easily be stolen using different avenues. The fraudsters would use various methods, from creating plastic cards

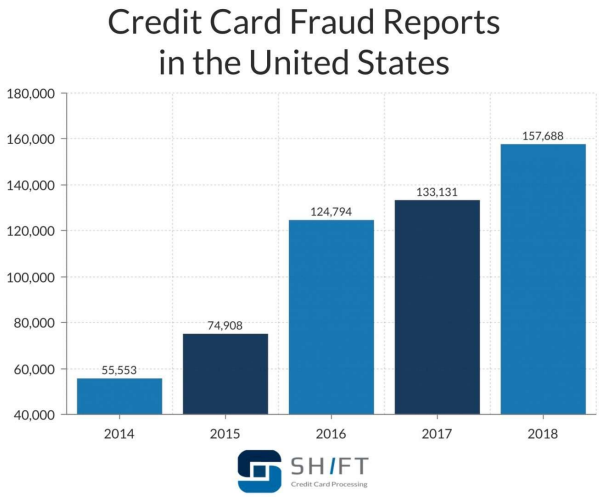

Figure 1. Credit Card Fraud Reports in the United States [21]

and using stolen card numbers to use at stores and to make online purchases. The financial institutions have taken significant steps to prevent financial fraud. One such step that Visa and MasterCard took in 2015 was to build chip card technology so that the merchants are required to ask for PIN to validate the transactions.

The fraudsters are up-to-date on their technological capabilities and are using new methods, such as conducting identity theft, sending phishing emails, and creating deceptive websites to commit financial fraud. Conventional rule-based programming systems cannot be changed quickly enough to detect the new ways of financial fraud, so they would not provide appropriate protection to the financial institutions. Machine Learning can be used in such cases to detect and

prevent financial fraud by utilizing past data on fraudulent transactions, historical patterns and come up with new techniques.

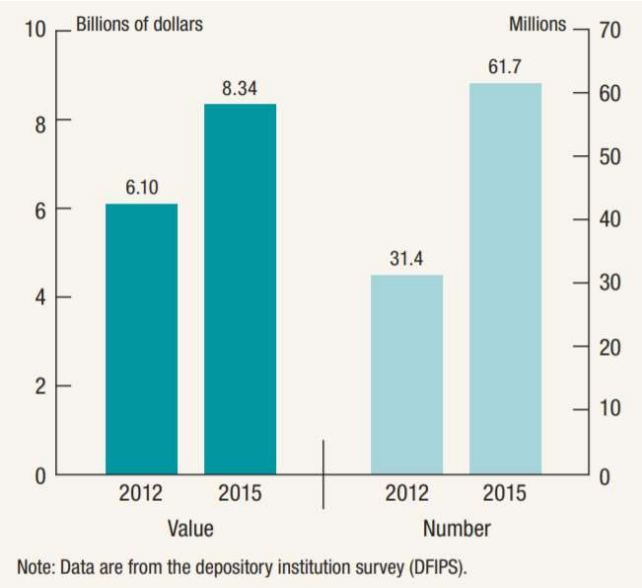

Figure 2. Total Payments Fraud, by value and number, 2012 and 2015 [24]

As shown in figure 2, the online credit card fraud was $6.10 billion in 2012 and $8.34 billion in 2015. It is estimated that online credit card fraud will soar to $32 Billion in 2020. With the huge amount at stake, financial institutions have been investing a significant amount in Data Science to identify and prevent financial fraud. A typical fraud detection system is shown in figure 3.

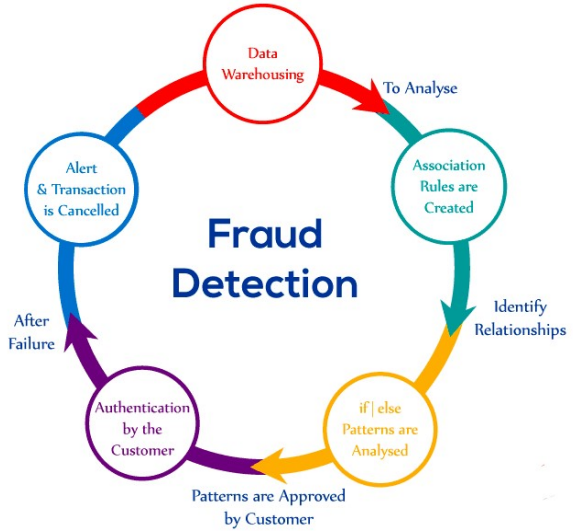

Figure 3. Typical Fraud Detection System [20]

Machine Learning can be used to detect the financial fraud using Classification Models. Classification models generally use supervised Machine Learning and they attempt to draw some conclusion and predict values of one or more output attributes based on one or more input attributes. The well-defined classification models can properly classify transactions as either legitimate or fraud, based on the transaction details.

What does Naïve Bayes classification model do?

Some examples of classification models are Logical Regression, Decision Trees, Random Forest, and Naïve Bayes [22].

This paper presents the implementation of Naïve Bayes algorithm on credit card dataset to identify the fraudulent transactions in the dataset.[3] Bayesian network classifiers are very popular in the area of machine learning and it comes under the category of supervised classification models. Naïve Bayes classifier is also a well-known Bayesian Network that is based on Bayes theorem of conditional probability and hence, is a classifier based on probability which considers Naïve i.e., strong independence assumption. It was formerly used in the text retrieval community as a baseline technique for categorizing text because there was a problem of deciding in which category the document belongs to, with word frequencies as the feature.

The Naïve Bayes machine learning classifier tries to predict a class which is known as outcome class based on probabilities, and also conditional probabilities of how many times it occurred from the training data. This kind of learning is very efficient, fast and high in accuracy for real-world scenarios, and is known as supervised learning. Also, this is highly efficient because it estimates the parameters by using very small training data which is used for classification and is based upon word independence. Though Naïve Bayes is quite simple to implement and understand and uses strong assumptions, it gives pretty accurate results and also it has been proven over and over the time that Naïve Bayes works effectively in various areas related to machine learning. In Naïve Bayes, the basic concept to computing the probabilities of various categories given a text is performed by using joint probabilities of categories and words.

Application of Naïve Bayes in Financial Fraud Detection

Naïve Bayes has been a popular classification system among researchers that have researched on the use of classification systems to detect financial fraud. Maes et al. [23] performed experiments and compared Bayesian Belief Networks and Artificial Neural Networks for credit card fraud detection. Their experimentation proved that the Bayesian Belief Network had better fraud detection capability than the Artificial Neural Networks.

[3][22] The Naïve Bayes machine learning classifier tries to predict a class which is known as outcome class based on probabilities, and also conditional probabilities of its occurrence from the training data. The initial step for Naïve Bayes classification algorithm is the Bayes theorem for conditional probability, where 'x' is given data point and 'C' is a class: $P(C/x) = P(x/C)/P(x)$. The further steps are done by making the assumption for a data point $x = \{x1 \text{ to } xj\}$, and the occurrence probability of each of its attribute within given class is independent. So the probability of x can be computed as follows: $P(C/x) = P(C).\Pi\ P(xi/C)$.

In their research, Er. Monika et al. [2] constructed N-Dimensional hyperplane and divided the data into two categories, normal transactions and fraudulent transactions. They generated two parallel hyperplanes on each side of the hyperplane, so that the data can be separated easily. The distance between the two hyperplanes was increased by separating the hyperplanes and regression model with squared distance was used for analysis. Naïve Bayes classification then helped to perform regression analysis and numerical calculations. They proposed an algorithm

to utilize hyperplanes using Naïve Bayes classification model as shown in figure 4.

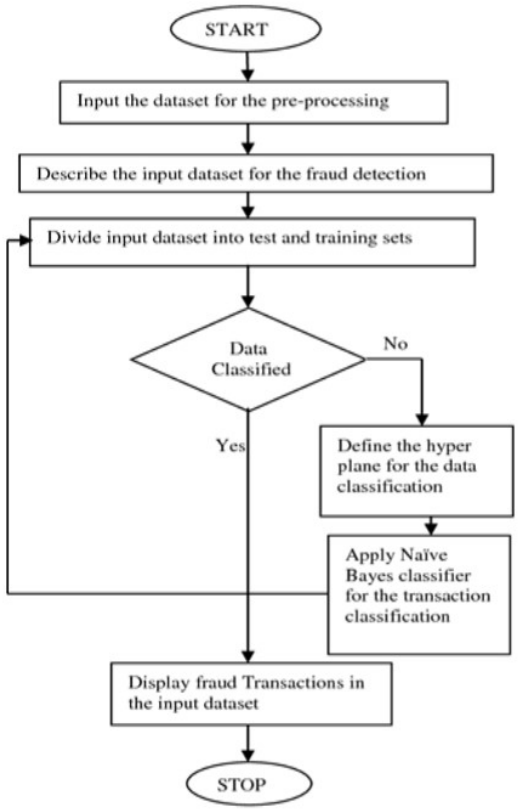

Figure 4. Proposed Flowchart [2]

Conclusion:

In conclusion, the significance of data science in catching credit card fraud is unquestionable. As we have mentioned, based on the information we have collected, pinpointing specific anomalies in the cluster of data gives us the decisive solution. It is important to note that there are various methodologies which could be used and the models are chosen based on the accuracy and precision. Detecting anomalies from regular transactions is like policing. By using Machine learning algorithms, models can be created where we can classify legitimate vs illegitimate transactions. These models would be run on a credit card dataset which aids us in distinguishing our regular customers with scammers and fraudsters. They are the ones who find loopholes in the banking system to take advantage of it for financial reasons. We conclude that Naïve Bayes Classification System has proven to have better fraud detection capability.

Acknowledgements:

We would like to thank Professor Shankar Parajulee for letting us create a team project with understanding of all the readings (articles/journals/books) provided to us which in turn is beneficial in understanding of the data Science process. We would also like to thank John D. Kelleher and Brendan Tierney who wrote the book Data Science. With the readings from the book, we have been able to understand data science, its methodologies, objectives, model names, and so on. Finally, we would like to express our gratitude to Bellevue University for starting the Data Science subject online because of which all of our teammates were able to join the course without attending the college in person (especially helpful during pandemic).

References:

1. Joshi, Bhaavika, et al. "Credit Card Fraud Detection with Tools Provided by Payment Gateways." *Chargebee's SaaS Dispatch*, 11 June 2020, www.chargebee.com/blog/credit-card-fraud-detection-tools/.

2. Monika, Er, and Er Amarpreet Kaur. "Fraud Prediction for Credit Card Using Classification Method." *International Journal of Engineering & Technology*, 2018, www.sciencepubco.com/index.php/ijet/article/view/12577/5978.

3. Kiran, Sai, et al. "Credit Card Fraud Detection Using Naïve Bayes Model Based and KNN Classifier: Semantic Scholar." Semanticscholar.org, Jan. 2018, www.semanticscholar.org/paper/Credit-card-fraud-detection-using-Na%C3%AFve-Bayes-model-Kiran-Guru/60a0221ba986ad612fff7ef10005ca684640b351.
4. Husejinovic, Admel. "Credit Card Fraud Detection Using Naïve Bayesian and C4.5 Decision Tree Classifiers." SSRN, 11 Feb. 2020, papers.ssrn.com/sol3/papers.cfm?abstract_id=3521283
5. Seeja, K. R., and Masoumeh Zareapoor. "FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining." *The Scientific World Journal*, Hindawi, 11 Sept. 2014, www.hindawi.com/journals/tswj/2014/252797/.

6. Joyner, Ellen. "Enterprise wide Fraud Management." *http://support.sas.com/*, SAS Institute, 2011, support.sas.com/resources/papers/proceedings11/029-2011.pdf.

7. "Best Practices For Detecting Banking Fraud." *https://Guardiananalytics.com/*, Guardian Analytics, 2019, guardiananalytics.com/wp-content/uploads/2018/06/Best_Practices_for_Detecting_Fraud_white_paper.pdf.

8. Journals, GATR, et al. "The Comparison of Two Data Mining Method to Detect Financial Fraud in Indonesia." *Academia.edu*, 2017, www.academia.edu/34136185/The_Comparison_of_Two_Data_Mining_Method_to_Detect_Financial_Fraud_in_Indonesia.

9. Singh, Ajeet, and Anurag Jain. "An Empirical Study of AML Approach for Credit Card Fraud Detection–Financial Transactions." *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL*, 2020, univagora.ro/jour/index.php/ijccc/article/view/3498.

10. Alison Lui & George William Lamb (2018) Artificial intelligence and augmented intelligence collaboration: regaining trust and confidence in the financial sector, Information & Communications Technology Law, 27:3, 267-283, DOI: 10.1080/13600834.2018.1488659

11. Repousis, S., Lois, P. and Veli, V. (2019), "An investigation of the fraud risk and fraud scheme methods in Greek commercial banks", Journal of Money Laundering Control, Vol. 22 No. 1, pp. 53-61. https://doi.org/10.1108/JMLC-11-2017-0065

12. Jadhav, Anirudha. "Fighting Fraud in Financial Services." *https://www.tcs.com/*, Tata Consultancy Services, 2018, www.tcs.com/content/dam/tcs/pdf/Industries/Banking%20and%20Financial%20Services/fighting-fraud-in-financial-services.pdf.

13. Hasham, Salim, et al. "Combating Payments Fraud and Enhancing Customer Experience." *McKinsey & Company*, 2018,

www.mckinsey.com/industries/financial-services/our-insights/combating-payments-fraud-and-enhancing-customer-experience.

14. Lamont, Judith. "Financial Services: Real-Time Fraud Countermeasures." *KMWorld*, 3 July 2010, www.kmworld.com/Articles/Editorial/Features/Financial-services-Real-time-fraud-countermeasures-68146.aspx.

15. Blaschka, Todd, and Gaurav Deshpande. "How the World's Largest Banks Use Advanced Graph Analytics to Fight Fraud." *RTInsights*, 22 Jan. 2020, www.rtinsights.com/how-the-worlds-largest-banks-use-advanced-graph-analytics-to-fight-fraud/.

16. Vergara, David. "Solving the Multi-Billion Dollar Fraud Prevention Problem." *OneSpan*, OneSpan, 2019, www.onespan.com/blog/solving-multi-billion-dollar-fraud-prevention-problem.

17. Zareapoor, Masoumeh, and Pourya Shamsolmoali. "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier." *Procedia Computer Science*, Elsevier, 22 May 2015, www.sciencedirect.com/science/article/pii/S1877050915007103.

18. Pierre, Rafael. "Detecting Financial Fraud Using Machine Learning: Winning the War Against Imbalanced Data." *Medium*, Towards Data Science, 4 Jan. 2019, towardsdatascience.com/detecting-financial-fraud-using-machine-learning-three-ways-of-winning-the-war-against-imbalanced-a03f8815cce9.

19. Institute, Software Engineering. "Insider Fraud in Financial Services." *Resources.sei.cmu.ed*, Carnegie Mellon University, 2012, resources.sei.cmu.edu/asset_files/Brochure/2012_015_001_28207.pdf

20. Group, SPD. "Credit Card Fraud Detection Solutions - How to Implement Them in Your Business." *Technology Partner for Innovative Companies*, 18 June 2020, spd.group/machine-learning/credit-card-fraud-detection/.

21. Company, Shift Processing. "Credit Card Fraud Statistics." *Shift Credit Card Processing*, 2 July 2020, shiftprocessing.com/credit-card-fraud-statistics/.

22. Kelleher, John D., and Brendan Tierney. Data Science. The MIT Press, 2018.

23. S. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick, "Credit card fraud detection using Bayesian and neural networks," in Proceedings of the 1st International NAISO Congress on Neuro Fuzzy Technologies, pp. 261–270, 1993.

24. 24. Board of Governors of The Federal Reserve System. (n.d.). *Changes in U.S. Payments Fraud from 2012 to 2016: Evidence from the Federal Reserve Payments Study* (October 2018 ed.). Retrieved July 25, 2020, from https://www.federalreserve.gov/publications/files/changes-in-us-payments-fraud-from-2012-to-2016-20181016.pdf