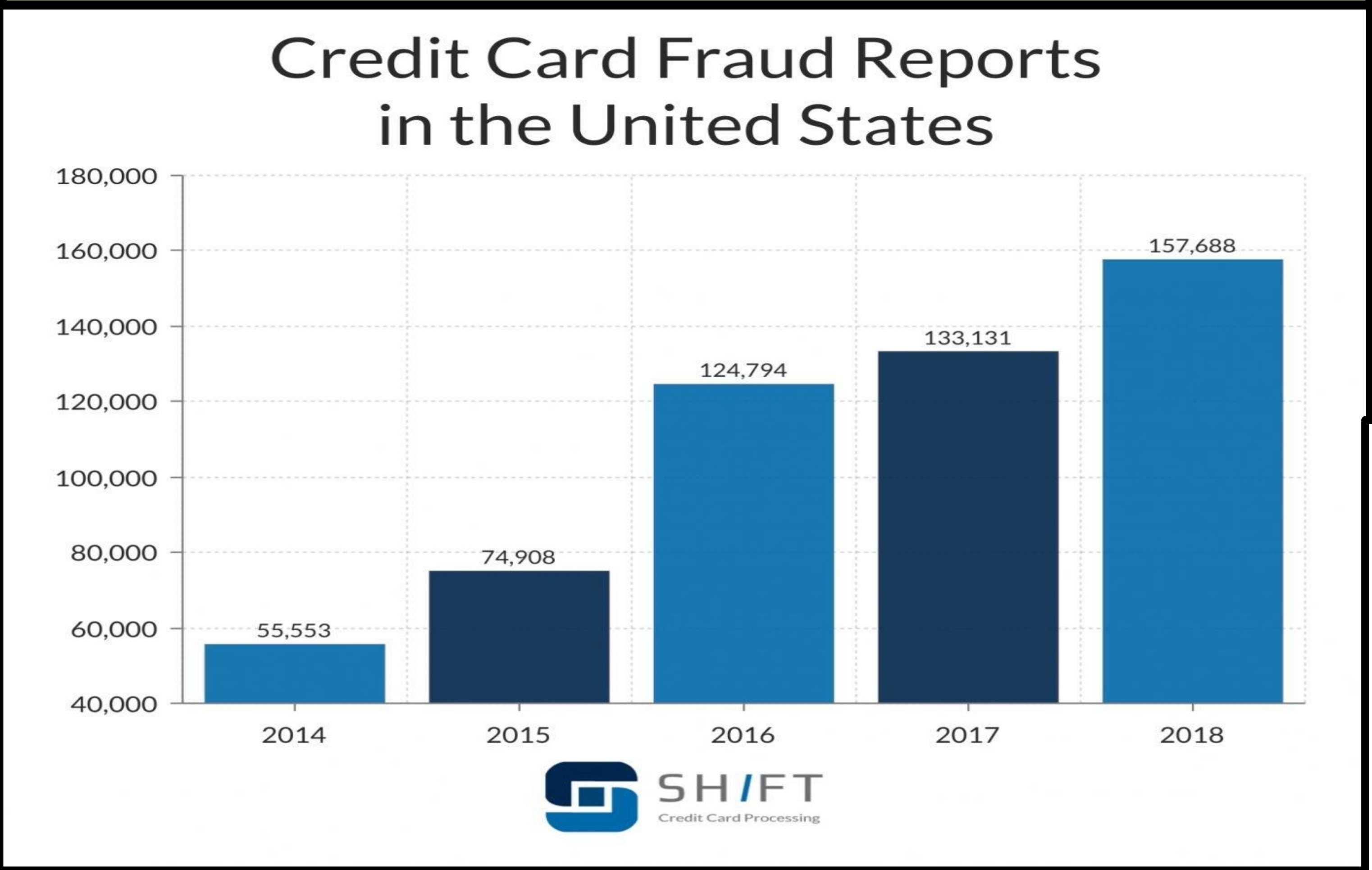
Combating Fraudulent Transactions: How Data Science Is Revolutionizing the World of Finance

Ajit Kolekar, Arun Thangaraj, Derrick Lee, Saurabh Shrestha
Bellevue University MS in Data Science Program

Introduction

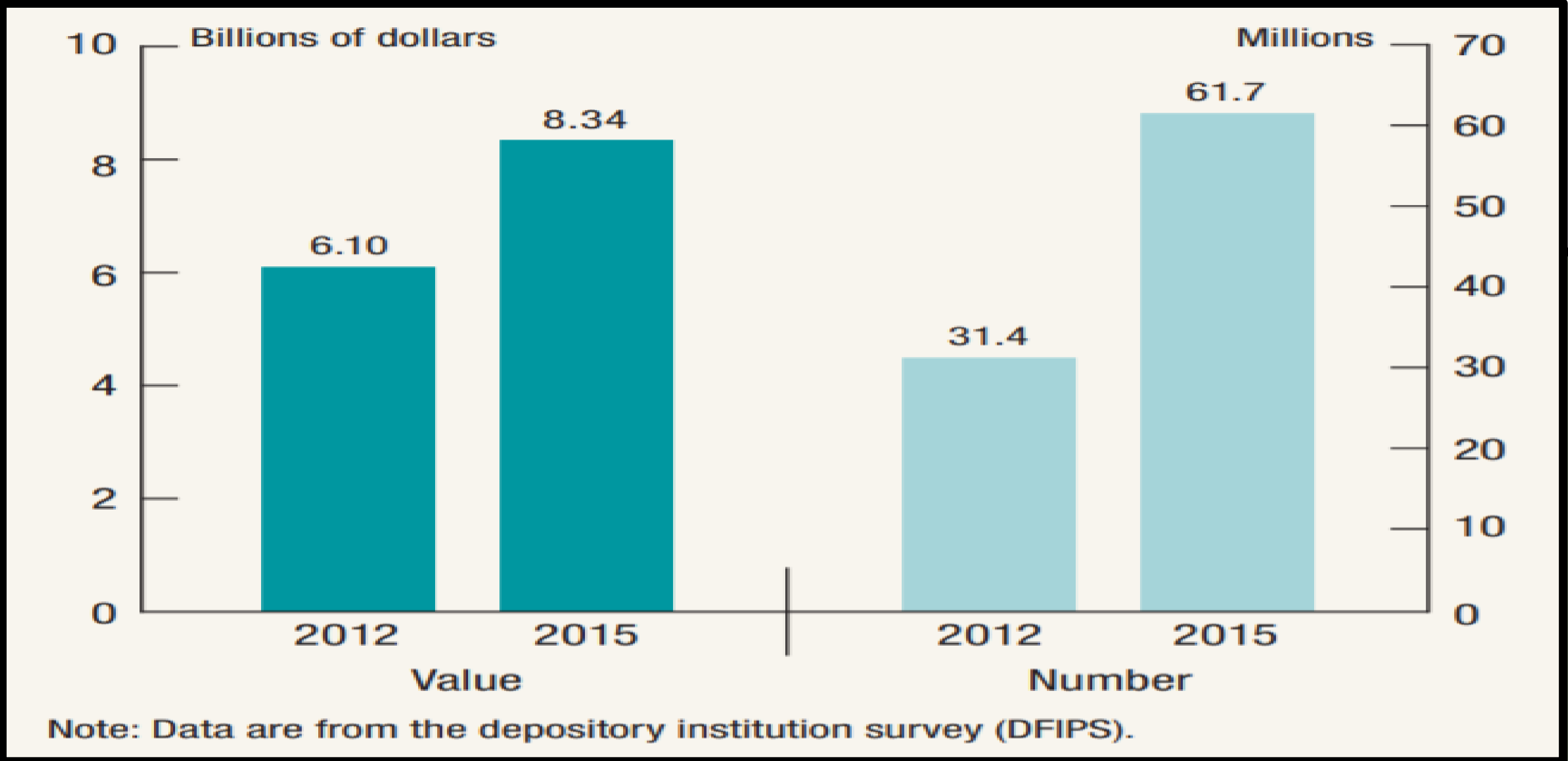
In today's era, fraud in the financial sector could encompass a wide variety including credit card fraud, insurance fraud, mortgage fraud and so on. Financial Fraud can be committed through many methods, including mail, wire, phone, and the internet. The difficulty of checking identity and legitimacy 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. 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. 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. 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 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. 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.

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. 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.



Note: Data are from the depository institution survey (DFIPS).

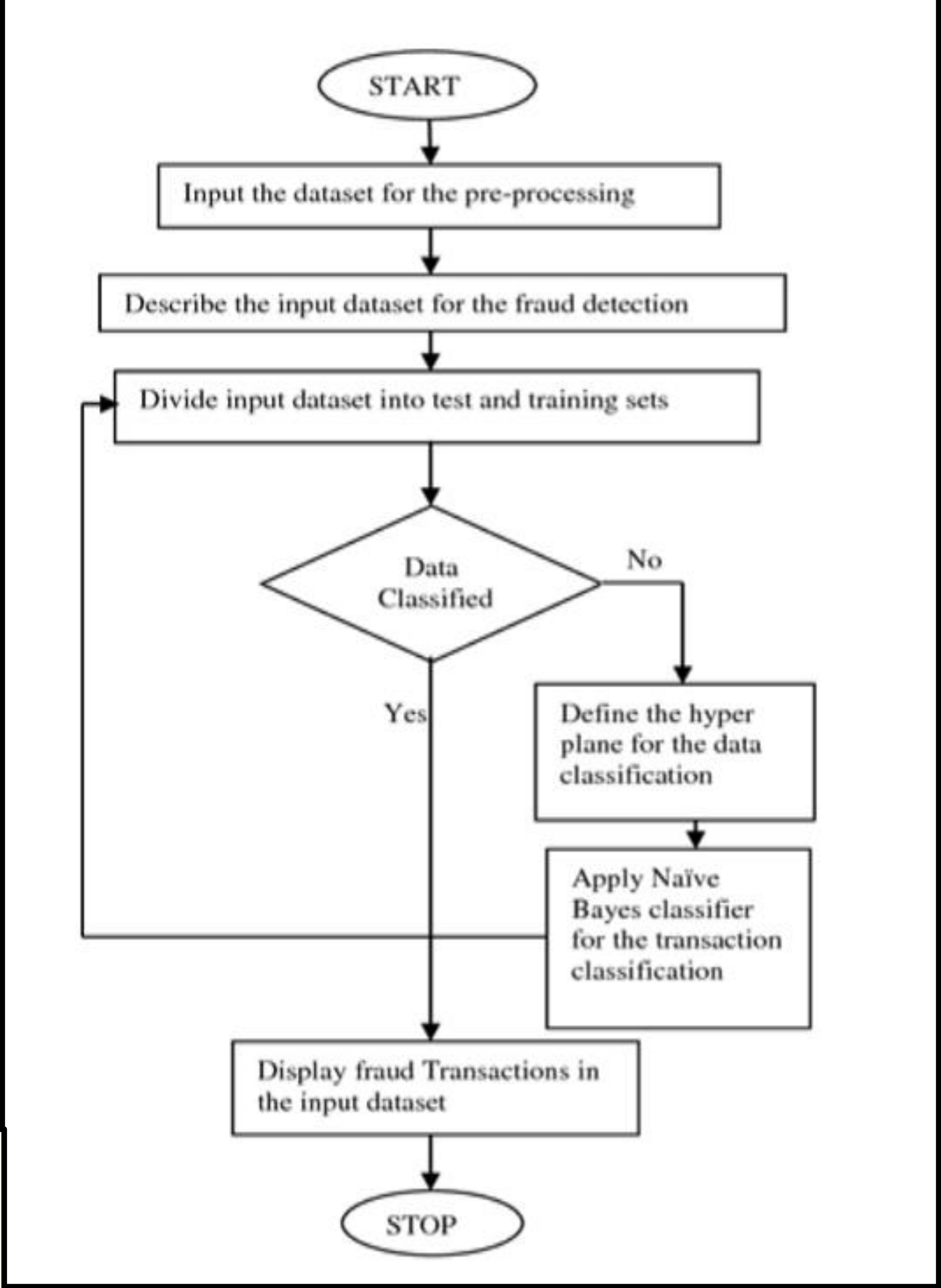
Deliverables

- ❖Analysis of how Data Science is used to combat financial fraud.
- ❖Determining the best classification model financial fraud.
- ❖Understanding what Naïve Bayes classification model does.
- ❖Understanding of how Naïve Bayes is applied in financial fraud detection.

What does Naïve Bayes classification model do?

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.

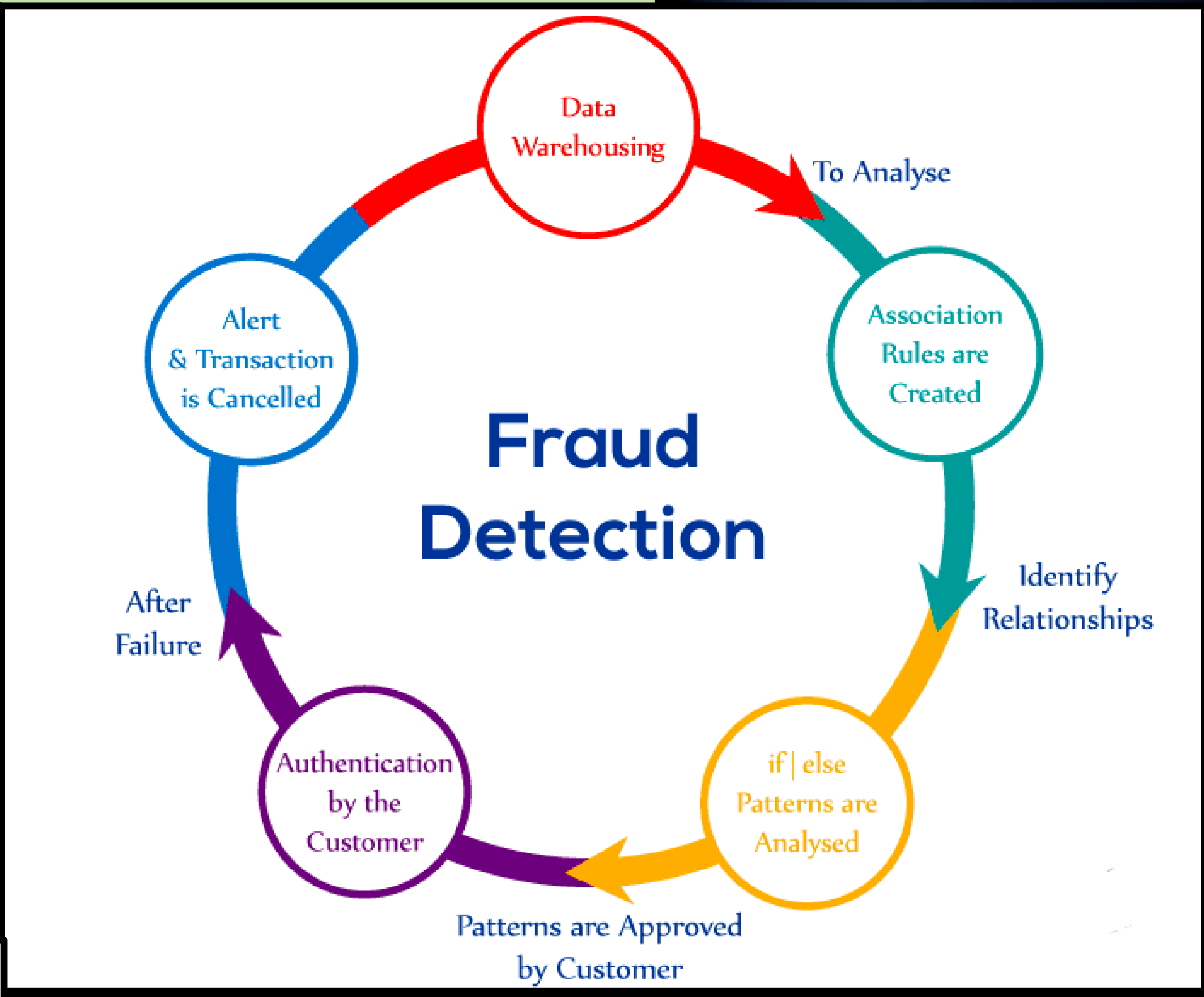
The Naïve Bayes machine learning classifier tries to predict a class which is known as outcome class based on probabilities, and also conditional probabilities of how many times it occurred from the training data. This kind of learning is very efficient, fast and high in accuracy for real-world scenarios, 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.


START
Input the dataset for the pre-processing
Describe the input dataset for the fraud detection
Divide input dataset into test and training sets
Data Classified
No
Yes
Define the hyper plane for the data classification
Apply Naïve Bayes classifier for the transaction classification
Display fraud Transactions in the input dataset
STOP

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. 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.

In their research, Er. Monika et al. 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.


Fraud Detection
Data Warehousing
To Analyse
Association Rules are Created
Identify Relationships
if | else Patterns are Analysed
Patterns are Approved by Customer
Authentication by the Customer
After Failure
Alert & Transaction is Cancelled

References

Pierre, R. (2019, January 04). Detecting Financial Fraud Using Machine Learning: Winning the War Against Imbalanced Data. Retrieved July 07, 2020, from https://towardsdatascience.com/detecting-financial-fraud-using-machine-learning-three-ways-of-winning-the-war-against-imbalanced-a03f8815cce9

Sajjad Daliri. (2020). Using Harmony Search Algorithm in Neural Networks to Improve Fraud Detection in Banking System. Computational Intelligence and Neuroscience, 2020. https://doi.org/10.1155/2020/6503459

Elena-Adriana MINASTIREANU, & Gabriela MESNITA. (2019). An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection. Informatică Economică, 23(1), 5–16. https://doi.org/10.12948/issn14531305/23.1.2019.01

Albashrawi, M. (2016). Detecting Financial Fraud Using Data Mining Techniques: A Decade Review from 2004 to 2015. Journal of Data Science, 14(3), 553–569.

Demush, R. (2020, June 09). Solving Financial Fraud Detection with Machine Learning Methods. Retrieved July 07, 2020, from https://perfectial.com/blog/fraud-detection-machine-learning/

Luis Jose S. Santos, & Shirlee R. Ocampo. (2018). Bayesian Method with Clustering Algorithm for Credit Card Transaction Fraud Detection. Revista Română de Statistică, 66(1), 103–120

Bart Baesens, Veronique Van Vlasselaer, & Wouter Verbeke. (2015). Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques : A Guide to Data Science for Fraud Detection. Wiley.

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).