Critical Analysis of Machine Learning Based Approaches for Fraud Detection in Financial Transactions

Thushara Amarasinghe
Department of Computing
Informatics Institute of Technology,
University of Westminster
Colombo, Sri Lanka
a.amarasinghe@my.westminster.ac.uk

Achala Aponso
Department of Computing
Informatics Institute of Technology,
University of Westminster
Colombo, Sri Lanka
achala.a@iit.ac.lk

Naomi Krishnarajah
Department of Computing
Informatics Institute of Technology,
University of Westminster
Colombo, Sri Lanka
naomi.kr@iit.ac.lk

ABSTRACT

Fraud has become a trillion-dollar industry today. Some finance companies have separate domain expert teams and data scientists who are working on identifying fraudulent activities. Data Scientists often use complex statistical models to identify frauds. However, there are many disadvantages to this approach. Fraud detection is not real-time and therefore, in many cases fraudulent activities are identified only after the actual fraud has happened. These methodologies are prone to human errors. In addition, it requires expensive, highly skilled domain expert teams and data scientists. Nevertheless, the accuracy of manual fraud detection methodologies are low and due to that, it is very difficult to handle large volumes of data. More often, it requires time-consuming investigations into the other transactions related to the fraudulent activity in order to identify fraudulent activity patterns. Finance companies are not getting adequate return of interest (ROI) despite the resources and money spent on these traditional methodologies. Most of the traditional fraud detection methodologies focused on discrete data points. (User accounts, IP addresses devices, etc…) However, these methodologies are no longer sufficient for today's needs. As fraudsters and hackers are using more advanced and cutting edge techniques to mask their fraudulent activities even from the sharpest eyes. These methodologies can only detect known types of attacks, therefore an analytical approach is required to address these drawbacks of the traditional methodologies.

The aim of this paper is to review selected machine learning and outlier detection techniques that can be integrated into a fraud detection system for financial transactions. Various machine-learning algorithms such as Bayesian Networks, Recurrent Neural Networks, Support Vector Machines, Fuzzy Logic, Hidden Markov Model, K-Means Clustering, K-Nearest Neighbor and their existing implementations on fraud detection domain will be discussed to find a better approach for a fraud detection system.

ICMLT 2018, May 19--21, 2018, JINAN, China
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-6432-4/18/05…$15.00
DOI: https://doi.org/10.1145/3231884.3231894

CCS Concepts

• Computing methodologies~Machine learning approaches • Computing methodologies~Supervised learning • Computing methodologies~Unsupervised learning • Computing methodologies~Anomaly detection

Keywords

Fraud Detection; Machine Learning; Supervised Anomaly Detection & Unsupervised Anomaly Detection

1. INTRODUCTION

Fraud can be defined as a "Wrongful or criminal deception intended to result in Financial or Personal gain" according to the Oxford dictionary [1]. In recent years, various kinds of frauds have captured the world's attention. The World Fact Book [2] reports gross world product (GWP) as $75.27 trillion. Also "ACFE Report [3] estimates organizations worldwide loss 5% of Revenues to fraud every year". Which means worldwide loss due to frauds is $3.76 trillion. That is more than 46 times of Sri Lanka's Gross Domestic Product [4]. (GDP of 2016 is $81.32). None of the industries are safe from frauds. According to the Kroll report [5] "75% of companies reported they have fallen victim to a fraud incident within the past year". Today, many industries are struggling to prevent frauds. Even though the Fraud affects different industries in different ways, the most affected industry is the banking and finance sector. According to Maxwell Locke & Ritter [6] "The banking and financial services industry is the most victimized industry with nearly 17 percent of all fraud cases reported".

For an example, A Telegraph report [7] states that 3 UK citizens were arrested for running over 200 million credit card fraud ring. Another incident is that Anup Patel [8] managed to steal 2 million US dollars through fake credit cards, with a help of a special device called as skimmers to clone credit card details. According to Telegraph [9] Kaspersky Lab uncovered that 680 Euros were stolen from more than 100 financial institutions worldwide by a Russian hacker group. This paper describes different supervised and unsupervised machine learning techniques, which can be used to identify fraudulent transactions.

The rest of the paper is organized as follows: Section 2 describes how the machine learning and anomaly detection techniques can be used to identify fraudulent transactions, which includes a discussion of different supervised and unsupervised machine learning techniques. Section 3 introduces a few other fraud detection techniques, which can be used to identify transaction anomalies. Section 4 provides the conclusion. Section 5 describes future works and Section 6 presents the acknowledgement.

2. MACHINE LEARNING

More robust and accurate solutions are required to identify today’s highly skilled fraudsters and their activities. This is where anomaly detection and machine learning techniques comes to the picture. Anomaly detection techniques can be used to identify data points in the data that deviates from the rest of the data. There are three types of anomalies aka outliers. Which are namely, Point outlier, Contextual outlier and Collective outlier [10].When individual data points deviate from the rest of the data, it is considered as a Point outlier (ex: Abnormal transaction amounts). When an instance of data is abnormal in a specific context but not otherwise, it is considered as a Contextual outlier aka conditional outlier (ex: Transactions occurs at middle of the night). When a set of related data points is anomalous with respect to the whole data set, but not to individual values, it is considered as a Collective outlier. Point outliers can occur in any data set but the collective outliers occur only on the data sets which has data instances related to each other. According to Tang et al [11] contextual outliers deviates from the normal behavior of the data set when considered with a set of specific attributes but not otherwise. Any point or collective outlier can be converted to a contextual outlier if the context is considered for analysis. When the training data is available, supervised learning techniques can be employed to identify transaction anomalies, whereas unsupervised anomaly detection techniques can be used when the training data is unavailable.

2.1 Supervised Anomaly Detection/ Classification

Transaction frauds can be identified by detecting anomalies in transaction streams. As per Omar et al [12], one way to identify anomalies in transactions is to use classification algorithms such as Random Forests, Support Vector Machines (SVM), Decision Trees or Bayesian Network. But According to Goernitz et al [13] there are few drawbacks. It is required to have labeled data, anomalous and normal classes’ needs to be balanced (at least 1:5) and data points should not depend on earlier data points. But as shown Srinath [14] this is often not the case for time series data as they are auto-correlated. Most of the time real-world data are auto-correlated and their classes are not balanced. Also it is hard to find labelled data. Bayesian networks, recurrent neural networks, support vector machine and fuzzy logic can be discussed under supervised machine-learning algorithms.

2.1.1 Bayesian Network

Heckerman [15] defined that a Bayesian network as a “Graphical model that encodes probabilistic relationships among variables of interest “. Bayesian belief networks (BBN) can be used for fraud detection as it has few advantages. It has the ability to encode interdependencies among all variables and predicting events with limited data. According to Maes et al [16], it has the ability to incorporate prior knowledge and data. They have used reasoning under uncertainty technique of Bayesian belief networks for credit card fraud detection. Moreover, results of real world financial data set provided by Serge Waterschoot at Europay International (EPI) indicated that learning time is much shorter for BBN (20 min) compared to ANN (approximately 1hour). According to Mukhanov [17] the BBN detects 8% more fraudulent transactions. He also described a BBN based approach for credit card fraud detection. Initially Naive Bayesian Classifier was used to calculate the probabilities for fraudulent transactions and later moves to Bayesian Belief Networks. Conditional independence of attributes has a great influence on Bayesian networks. For an example, transaction currency has an impact on transaction amount. As a solution to these situations, Rissanen’s Minimal Description Length principle was employed to find the structure of dependence network between used attributes of transactions. The results show that Bayesian Networks is more accurate than the Naive Bayesian Classifier.

2.1.2 Recurrent Neural Networks(RNN)

Recurrent neural networks allow previous states to be persisted. Feedforward neural networks report high false positive rates as they are considering only individual transaction values. Benard and Christian [18] proposed RNN for frauds detection. They treated sequence of transaction records as a collection to be used with time series data. They used real-world credit card dataset to test the recurrent neural network against a FFNN and a SVM. Results indicated that RNN could successfully identify frauds in time series data. They have considered using real time recurrent learning (RTRL) by forwarding activity gradient information instead of back propagation. In addition, the LSTM was used to address the vanishing gradient problem.

2.1.3 Support Vector Machine (SVM)

Abdelhamid et al [19] proposed a hybridization mechanism of binary and single class SVM methods for detecting credit card frauds, money laundering and mortgage frauds. They were able to achieve 80% of precision for single class SVM method. Sahin and Duman [20] compares the performance of SVM and Decision Trees in terms of performance for credit card fraud detection with a real data set. Data pre-processing was done to address the class unbalancing. Stratified sampling has been used to undersample the normal records. Kernels of polynomial, sigmoid, linear and radial functions were used in SVMs. It was noted that decision trees outperform SVMs. Figure 1. shows how the SVM categorize abnormal and normal data points with the help of support vectors.

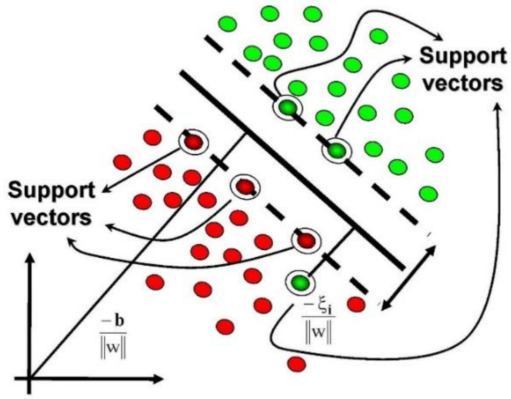

Figure 1. Support Vector Machine

2.1.4 Fuzzy Logic

Alam and Lenard [21] proposed fuzzy logic to cluster information into different fraud risk categories. This method is called FANNY due to the fuzziness of risk categorization. Objects were partitioned into different categories without hard decisions, which could result in fixed set of categories. Euclidian distances were used to cluster data and identify similar categories. In addition, they described few critical issues related to fuzzy logic on fraud detection domain. Fuzzy logic method, fuzzy rules, membership functions, fuzzification and defuzzification were discussed [22] by Razooqi et al. Their results using an ANN showed that ANN is 33% more accurate than fuzzy logic. However, they were able to achieve a mean square error of 0.476 for the fuzzy logic model.

Bentley [23] proposed a credit card fraud detection mechanism which utilizes genetics to evolve fuzzy logic rules to better classify transactions into anomalies and normal classes. At first data were clustered and three membership functions were applied to the clustered data. (LOW, MEDIUM, HIGH). Fuzzy rule evaluation was done with a genetic algorithm and the results were assessed to identify frauds.

2.2 Unsupervised Anomaly Detection

Kadous et al [13] states that when training data is available classification can be done using Recurrent Neural Networks or time series classification techniques such as Hidden Markov Models, Dynamic Time Warping, Dynamic Bayes Nets. Constructive induction of temporal feature, extracting prototype examples and applying relational learning techniques. But as Brownlee [4] shows, when there is no training data available it is possible to conduct unsupervised learning or semi-supervised learning. The major disadvantage to this approach is, there is no way to figure out its capabilities, as there is no data to test it against. Outlier detection techniques can be used when training data is unavailable.

2.2.1 Point Outliers

Point anomalies can be identified with the help or percentiles for numeric data such as transaction values. According to Goldstein and Dengel [10] Histograms can be used to detect anomalies in categorical data Ex: high quantities of abnormal transaction occurs in multiple cities. The solution is to find rate occurrences of data ranges or values and flag them as frauds if they happen again. Ex: If someone purchase goods from the same city. Whenever a transaction happens from another city, it can be flagged as a fraudulent transaction.

To determine a transaction spike is an anomaly or not, Z-Score calculation can be used according to Baesens et al. [3] If the z-score is greater than the threshold (default 3) it is considered as a fraudulent transaction. This is useful to reduce the noise in data. Figure 2. shows how transaction anomalies are spread in a data set.

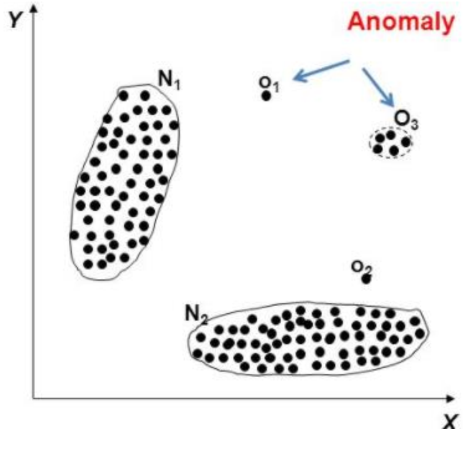

Figure 2. Point Anomalies

2.2.2 Collective Outliers

According to Filzmoser and Hron [8] There are two types of Collective Outliers, which are Univariate Collective Outliers and Multivariate Collective Outliers.

2.2.2.1 Univariate Collective Outliers

Univariate Collective Outliers anomalies happen, when the regular transaction pattern get disturbed. The anomaly is not itself a fraud but when compared with the rest of the transactions it deviates from the rest. Thus, it can be considered as a fraudulent activity. There are few methods to handle these use cases. Predictive analytics can be employed to solve this use case. With predictive analytics the system can predict the next value based on historical data and then apply centile on the error (centile value – true value). This model detects points with low probability as Univariate Collective Outliers and it can be built using Recurrent Neural Networks, regression, or time series models.

Hidden Markov Model - (Univariate Collective Outliers)

Chetcuti and Dingli, [5] used Hidden Markov Model with two different clustering algorithms (SimpleKMeans and Expectation Maximization algorithm) to identify credit card frauds. Building a Markov model is three-step process – State Classification, Probability Calculation, and Metric Comparison. Markov model can be trained with historical data. In addition, it is possible to let Markov model updates from real time data as well. Figure 3. Illustrates the probabilistic parameters of a HMM.

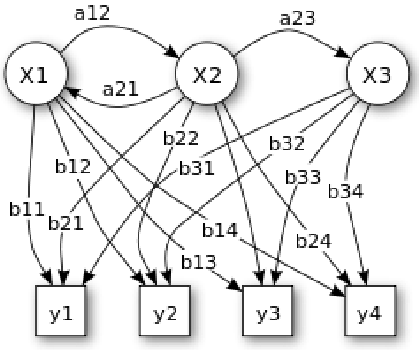

Figure 3. Hidden Markov Model

Sungkono and Sarno [18] proposed coupled HMM for credit card fraud detecting. Fraud detection was done with two steps. First sequence of strategies were obtained based on the activity sequences in fraud detection event logs of the existing system. Then the map model was determined by using the coupled HMM with sequences and activities obtained from event logs.

2.2.2.2 Multivariate Collective Outliers

As mentioned by Srinath [12] Multivariate Collective Outliers can be divided in to two categories as ordered Outliers and unordered Outliers. In some cases there can be unordered data with multiple readings such as transactions records (unordered multivariate dataset). For an example, low transaction values and high transaction quantities might be an anomaly even though both transaction value and quantities are in normal range when considered individually.

Clustering (K-Means) - Multivariate Collective Outliers (Unordered)

Transaction anomalies are very rare in real world data. In order to identify anomalies one approach would be to group the data into similar clusters [6] as stated by James McCaffrey. According to Srinath [12] centroids and density should be calculated for each cluster. Whenever a new data point arrives, distance to the known large cluster should be calculated. If the distance is large, then it can be identified as an anomaly. If there is no significant distance to the normal data cluster, then it can be identified as a genuine transaction. Then a new data point is assigned to the respective cluster and the average is calculated for all data points in that cluster. The centroid is moved to that average location. It is possible to manually label normal and anomalous clusters this will improve the model. The value of K determines the number of clusters needs to be created. At the start, “K” is initialized with a random number. K-means usually runs many times with different cluster sizes and calculate the clusters’ distortion. Lowest distortion value is the best clustering. According to Kevin Davenport [7] K-means attempts to minimize the cluster distortion defined by the sum of squared differences between each data point and its corresponding centroids. Figure 4 illustrates how the normal and abnormal transactions have been spread into 2 clusters.

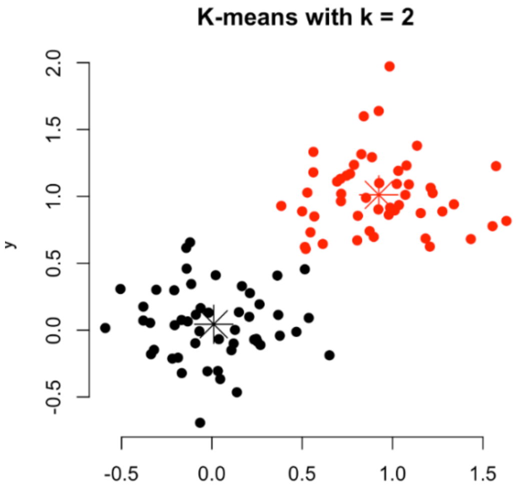

Figure 4. K -means cluster (normal and anomaly data points)

Meiping [16] proposed a credit card fraud detection solution by combining K-means clustering with Bayesian classification. Bayesian model has been setup on the K average computation method. Meiping identified that this model conciseness and functions of the proposed model exceeds the traditional Bayesian model. Weerathunga [2] discusses about anomaly detection with K-means clustering. Cluster boundaries were identified using Percentile distance values instead of maximum distance point of each cluster from their cluster centroids. It helps to prevent anomalous data being added into normal clusters. Weerathunga emphasizes the importance of giving high priority to true positives, as anomalous data is very rare. (ex.1:10,000). And further recommends few accuracy measures in Table 1.

Table 1. Accuracy measures

Accuracy Measure	Description	Formula
Recall (sensitivity)	Gives the true positive rate	TP / (TP + FN)
Precision	Gives the probability of predicting a true positive from all positive predictions	TP / (TP + FP)
PR curve	Plots precision vs. recall	
F1 score	Gives the harmonic mean of precision and recall (sensitivity)	2TP / (2TP + FP + FN)

To determine how good, the model is PR curve and F1 score can be used. Detecting all the fraudulent transactions are necessary while reducing the false positives. However, there is a tradeoff between precision and recall. If reducing the false positive is important, precision needs to be improved. If detecting all the anomalies is important, recall (sensitivity) needs to be improved. However, both precision and recall cannot be improved at the same time.

K -Nearest Neighbor

This model is based on the assumption that the normal data points are closer to each other and anomalies are closer to known anomalies. In this method, anomalies are detected with the distance to k-anomalies or with the relative density to other anomalies nearby. Hence, it is named as nearest neighbor. For numerical data, space can be broken into hypercube and often Euclidean distances can be used according to Angelo and Giaccari [1]. With categorical data space can be broken into bins using histograms as stated by Goldstein and Dengel [11]. According to Leung [14] often simple matching coefficient is used. If the data is multivariate then distance is calculated for every attribute and then combined. The result of the KNN algorithm depends on three factors, the distance metric used to decide the NN, the distance rule that has been used in the classification from KNN and the number of neighbors (K).

Multivariate Collective Outliers (Ordered)

According to Srinath [12] this class is the most general and consider ordering as well as value combinations. High value transactions happen to same merchant multiple times, may be normal in combination. However, it could be anomalous as combinations happen in wrong order. As an example, customer making high value transactions to the same merchant may be normal but it may be an anomaly when this happens within shorter period. Hidden Markov model and clustering methods can be combined. As stated by Fernando [9], first data should be clustered and used to build a probability matrix. Clusters will group common value and can be used to derive probabilities.

3. OTHER TECHNIQUES

Lu et al [15] showed that Adaptive Benford’s Law can be used even with incomplete datasets in order to identify fraudulent transactions. They combined this law with reinforced learning and created a new fraud detection approach. Fraudulent transactions were detected by calculating the deviations from expected Benford’s Law distributions. Any anomaly indicates a high probability of a fraudulent transaction.

Following Table 2 compares different types of fraud detection techniques with their advantages and disadvantages.

Table 2. Anomaly Detection Techniques Comparison

Technique	Pros	Cons
Traditional Methods	Since there is human interaction for fraud detection there is a high probability of identifying unknown fraud patterns	Accuracy is low. Difficult handle large volumes of data. Requires time consuming investigations into data Requires skilled professionals cost is high
Static Rules	Easy to implement Simple Easy to understand	According to Fernando [9] Can only be used with known frauds. Requires domain

		experts. Rules grow overtime and become hard to manage. It is difficult to implement complex rules due to the limitation of underlying languages.
Decision Trees	Robust, Simple to understand and interpret. Requires little data preparation. Able to handle numerical and categorical data. Can handle nonlinear data. Uses a white box model. Possible to validate a model using statistical tests. Perform well with large data sets in a short time.	Decision trees tends to create complex trees that do not generalize the data well. There are concepts that are hard to learn because decision trees do not express them easily.
Support Vector Machine	Finds the optimal separation hyper plane. Can deal with very high dimensional data. Some kernels have infinite Vapnik-Chervonenkis dimension, so it can learn very elaborate concepts. Usually work very well. Can be used for real-time fraud detection	Require both positive and negative examples. Need to select a good kernel function. Require lots of memory and CPU time.
K-means	Low complexity	K must be specified. Sensitive to noise sensitive to initial assignment of centroids.
Hidden Markov Model	Can be used for real-time fraud detection. The HMM based models reduce the False Positive (FP) transactions predict as fraud though they are really genuine customers.	It is not possible to detect frauds in initial few transactions.
Bayesian Networks	According to Mukhanov [17] one advantage of the method is the automatic decision and training processes. Clarity of the process. Simplicity of calculation.	Computationally expensive.
Fuzzy Logic	Works well with uncertainty and nonlinearity; Can be trained with minimal amount of data	Rules are not robust. Gives the same importance to all factors.

4. CONCLUSION

The objective of this paper was to identify existing fraud detection algorithms, techniques and critically evaluating each one of them. Advantages and disadvantages of existing research on fraud detection domain were thoroughly discussed in this paper. It was identified that each algorithm had its own advantages and disadvantages. Some of them were good on one aspect but had few shortcomings at the same time.

For a real-world fraud detection system, human interaction is necessary to determine whether a flagged transaction is actually fraudulent or not. Human user should be able to look into the past transaction patterns or contact the customer in order to determine whether the identified transaction is actually fraudulent or not. Therefore, it was identified that flagging a few genuine transactions as fraudulent is acceptable rather than ignoring an actual fraudulent transaction. Thus, improving recall is important rather than having a good precision. Most of the time supervised learning algorithms outperform unsupervised learning algorithms unless test data does not have unknown fraud types. It was identified that most of the classification algorithms does not perform well for fraud detection domain because the normal and anomalous classes are not balanced. Therefore, special oversampling and underdamping techniques should be used to overcome this issue. Accuracy is not a suitable matric to be used with imbalanced data sets. Hence, more suitable metrics measures like precision, recall and F1-Score should be used for evaluations.

5. FUTURE WORKS

Neural networks are one of the emerging deep learning technique, which can be used for detecting transaction anomalies in financial data. As a future goal, implementing an artificial neural network (ANN) can be done. The ANN can be combined with a genetic algorithm to obtain a higher accuracy. And the results obtained from the ANN can be compared with other algorithms.

6. ACKNOWLEDGMENTS

It would have not been possible without the support and guidance given by the people who stood behind me. I want to extend my gratitude to all of them. I would like to express my gratitude to my supervisor Achala Chathuranga Aponso for the constant support and guidance given to me. Special thanks to Nilantha Herath, Chatushka Dilhan, Isuru Jayasooriya and IIT.

7. REFERENCES

[1] Oxford Dictionary, “fraud | Definition of fraud in English by Oxford Dictionaries.” [Online]. Available: https://en.oxforddictionaries.com/definition/fraud. [Accessed: 12-Nov-2017].

[2] Central Intelligence Agency, “The World Factbook,” 2016. [Online]. Available: https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html. [Accessed: 10-Oct-2017].

[3] Association of Certified Fraud Examiners, “ACFE Report Estimates Organizations Worldwide Lose 5 Percent of Revenues to Fraud,” 2012.

[4] Trading Economics, “Sri Lanka GDP Growth Rate,” 2017. [Online]. Available: tradingeconomics.com/sri-lanka/gdp-growth. [Accessed: 10-Oct-2017].

[5] Kroll, “Global Fraud Report: Vulnerabilities on the Rise,” 2015.

[6] Maxwell Locke & Ritter, “Which industries are hardest hit by fraud?,” MLR, 2016. [Online]. Available: http://www.mlrpc.com/articles/which-industries-are-hardest-hit-by-fraud/. [Accessed: 10-Oct-2017].

[7] Telegraph, “Three UK men arrested over $200m credit card fraud,” 05-Jun-2013.

[8] BBC, “Men jailed over fake credit cards,” 29-Oct-2008.

[9] M. Evans, “Hackers steal £650 million in world’s biggest bank raid - Telegraph,” 15-Feb-2015. [Online]. Available: http://www.telegraph.co.uk/news/uknews/crime/11414191/Hackers-steal-650-million-in-worlds-biggest-bank-raid.html. [Accessed: 10-Dec-2017].

[10] S. Mukherjee, F# for Machine Learning Essentials. Packt Publishing Ltd, 2016.

[11] G. Tang, J. Pei, J. Bailey, and G. Dong, Mining Multidimensional Contextual Outliers from Categorical Relational Data ∗. .

[12] S. Omar, A. Ngadi, and H. H. Jebur, Machine Learning Techniques for Anomaly Detection: An Overview. 2013.

[13] N. Goernitz, M. M. Kloft, K. Rieck, and U. Brefeld, “Toward Supervised Anomaly Detection,” ArXiv14016424 Cs, Jan. 2014.

[14] P. Srinath, “Introduction to Anomaly Detection: Concepts and Techniques | My views of the World and Systems,” 2016. [Online]. Available: https://iwringer.wordpress.com/2015/11/17/anomaly-detection-concepts-and-techniques/. [Accessed: 10-Oct-2017].

[15] D. Heckerman, “A Tutorial on Learning With Bayesian Networks,” Mar. 1995.

[16] S. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick, “Credit Card Fraud Detection. Applying Bayesian and Neural networks,” Oct. 2017.

[17] L. Mukhanov, “Using Bayesian Belief Networks for credit card fraud detection,” 2008, pp. 221–225.

[18] B. Wiese and C. Omlin, “Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks,” vol. 247, 1970, pp. 231–268.

[19] D. Abdelhamid, S. Khaoula, and O. Atika, Automatic Bank Fraud Detection Using Support Vector Machines. .

[20] Y. Sahin and E. Duman, “Detecting Credit Card Fraud by Decision Trees and Support,” in Vector Machines”, International Multiconference of Engineers and computer scientists, 2011.

[21] P. Alam and M. Lenard, “Application of Fuzzy Logic to Fraud Detection,” Jan. 2009.

[22] T. Razooqi, P. Khurana, K. Raahemifar, and A. Abhari, “Credit Card Fraud Detection Using Fuzzy Logic and Neural Network,” in Proceedings of the 19th Communications & Networking Symposium, San Diego, CA, USA, 2016, p. 7:1–7:5.

[23] P. Bentley, “Fuzzy Darwinian Detection of Credit Card Fraud - Semantic Scholar,” 2004. [Online]. Available: /paper/Fuzzy-Darwinian-Detection-of-Credit-Card-Fraud-Bentley/364c45aeacd872370d0dab30456afa11f0ccc23c. [Accessed: 19-Jan-2018].

[24] M. W. Kadous, M. W. Kadous, and S. C. Sammut, “Temporal Classification: Extending the Classification Paradigm to Multivariate Time Series,” The University of New South Wales, 2002.

[25] J. Brownlee, “Supervised and Unsupervised Machine Learning Algorithms,” Machine Learning Mastery, 16-Mar-2016. .

[26] M. Goldstein and A. Dengel, Histogram-based Outlier Score (HBOS): A fast Unsupervised Anomaly Detection Algorithm. .

[27] B. Baesens, V. V. Vlasselaer, and W. Verbeke, Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection, 1st ed. Wiley Publishing, 2015.

[28] P. Filzmoser and K. Hron, “Outlier detection for compositional data using robust methods,” Math. Geosci., pp. 233–248, 2008.

[29] T. Chetcuti and A. Dingli, “Using Hidden Markov Models in Credit Card Transaction Fraud Detection,” 2018.

[30] K. R. Sungkono and R. Sarno, “Patterns of fraud detection using coupled Hidden Markov Model,” in 2017 3rd International Conference on Science in Information Technology (ICSITech), 2017, pp. 235–240.

[31] J. McCaffrey, “Data Clustering - Detecting Abnormal Data Using k-Means Clustering,” Feb-2013. [Online]. Available: https://msdn.microsoft.com/en-us/magazine/jj891054.aspx?f=255&MSPPError=-2147217396. [Accessed: 20-Feb-2018].

[32] K. Davenport, “The Cost Function of K-Means,” Kevin Davenport, 14-Feb-2014. .

[33] X. Meiping, “Application of Bayesian Rules Based on Improved K-Means Cassification on Credit Card,” in 2009 International Conference on Web Information Systems and Mining, 2009, pp. 13–16.

[34] A. Weerathunga, “[Article] Anomaly Detection Using K-Means Clustering,” 06-Jan-2016. [Online]. Available: https://wso2.com/library/articles/2016/01/article-anomaly-detection-using-k-means-clustering/. [Accessed: 01-Nov-2017].

[35] L. D. Angelo and L. Giaccari, “An efficient algorithm for the nearest neighbourhood search for point clouds,” 2011.

[36] M. Goldstein and A. Dengel, “Histogram-based Outlier Score (HBOS): A fast Unsupervised Anomaly Detection Algorithm - Semantic Scholar,” 2012. [Online]. Available: /paper/Histogram-based-Outlier-Score-HBOS-A-fast-Unsuperv-Goldstein-Dengel/5cf881d1db19834f123fcfc79ad32097aeafe17f. [Accessed: 12-Nov-2017].

[37]M. Leung, “k-Nearest Neighbor Algorithm for Classification,” 13-Nov-2007.

[38] S. Fernando, “Fraud Detection and Prevention: A Data Analytics Approach,” 2015. [Online]. Available: http://wso2.com/whitepapers/fraud-detection-and-prevention-a-data-analytics-approach/. [Accessed: 10-Oct-2017].

[39] F. Lu, J. Boritz, and D. Covvey, “Adaptive Fraud Detection Using Benford’s Law,” 2006, vol. 4013, pp. 347–358.