A Multifaceted Approach to Bitcoin Fraud Detection: Global and Local Outliers

Conference Paper · December 2016

DOI: 10.1109/ICMLA.2016.0039

Patrick Monamo
Eskom

Bhekisipho Twala
Tshwane University of Technology

A Multifaceted Approach to Bitcoin Fraud Detection: Global and Local Outliers

Patrick M. Monamo
Council for Scientific
and
Industrial Research
Pretoria, South Africa
Email: PMonamo@csir.co.za

Vukosi Marivate
Council for Scientific
and
Industrial Research
Pretoria, South Africa
Email: VMarivate@csir.co.za

Bhekisipho Twala
University of Johannesburg
Institute of Intelligent Systems
Johannesburg, South Africa
Email: btwala@uj.ac.za

***Abstract*—In the Bitcoin network, lack of class labels tend to cause obscurities in anomalous financial behaviour interpretation. To understand fraud in the latest development of the financial sector, a multifaceted approach is proposed. In this paper, Bitcoin fraud is described from both global and local perspectives using trimmed k-means and kd-trees. The two spheres are investigated further through random forests, maximum likelihood-based and boosted binary regression models. Although both angles show good performance, global outlier perspective outperforms the local viewpoint with exception of random forest that exhibits nearby perfect results from both dimensions. This signifies that features extracted for this study describe the network fairly.**

***Keywords*—kd-trees, anomaly, outlier, data mining, random forest, regression**

I. Introduction

In the data mining arena, the key objective is often attributable to extraction of insightful patterns from large datasets. Related to data mining in security applications, is anomaly detection whereby of interest are rare events that deviate significantly from majority of the instances as to arouse suspicion they were generated by a mechanisms not aligned to protocols that generated the bulk of the objects [18]. Such instances are referred to as outliers or anomalies and the application area is commonly known as outlier or anomaly detection. The concept of anomaly detection can be categorized into two broad approaches, that is either $global$ or $local$. By global, reference in terms of being an anomaly is made with respect to the majority of instances under study while local outlierness only considers the neighbourhood. In machine learning nomenclature, anomaly detection is encapsulated through the main groups depending on a particular angle under review in [7]. The groups include $classification$, $nearest$-$neighbourhood$, $clustering$ and $statistical$ based. In $classification$ problems, machine learning algorithms seek to distinguish between objects belonging to either normal or anomalous classes while in $clustering$ based techniques the learning algorithm will assign non-meaningful labels to objects based on similarities. With regard to $nearest$-$neighbourhood$, normal instances are found in denser neighbourhood as opposed to anomalies that lie in more sparse regions. The first three approaches stated above are non-parametric in nature. In some situations, it is assumed or known in advance that normality follows one of the known distributions and, any deviations are categorized as outliers. Such situations falls under $statistical$ based algorithms.

This paper takes a closer look at the open financial markets involving the Bitcoin network. The transaction data since inception of the cryptographic currency is publicly available on a ledger known as $blockchain$, but no instance is labelled as either an anomaly or normal. It is believed that majority of Bitcoin users adhere to network protocols with only a few deviations and as such deemed to be outliers or fraudsters in line with the definition provided. This paper will take a two-pronged approached towards anomaly detection over the Bitcoin network. From a methodological perspective, kd-trees and trimmed k-means under the broad groups of nearest-neighbourhood and clustering based approaches are adopted to serve as proxies in terms of what is deemed as fraud or non-fraud. A predetermined proportion of, maximum top 1% instances, will be categorized as fraud. The resultant kd-trees and trimmed k-means labelled dataset will thus serve as proxy for characterization of normal versus anomalous activity that ultimately allows supervised assessments.

II. Detect Fraud Without Labels

Unlike in the advent anonymous transactions of the Bitcoin network, traditional financial systems or institutions know the real world identities of its users (customers). As a result thereof, criminals are often aware of compliance requirement of this institutions and hence will break large transactions into amounts lower than thresholds set by statutory agencies and distribute across multiple accounts (or third parties) to avoid alarms by suspicious activity report (SAR) as per anti-money laundering protocols [28]. This practice is referred to as smurfing in the financial fraternity. While machine learning techniques have been hailed for novelty detection in recent years, [31] proposed a model-based fraud detection technique for event-driven process security analysis (PSA@R) as a basis for Fraud Chain Detection (FCD) on mobile money transfers which was compared with some of the known machine learning methods. According to [23] the technique is inspired by the desire to comply with processes involving security requirements as well as evaluating events associated with well-defined workflows and security properties. Based on synthetic data, the performance of FCD was found to be better in terms of both recall and precision when compared to known supervised machine learning algorithms.

To detect local outliers, [14] [3] [20] [29] used peer group analysis to model and monitor behavioural patterns of objects in a time series setup. According to the authors, the capabilities of the technique extend to detection of instances that start behaving in an odd manner to those within the group they were similar to at the previous time. Noting that traditional peer group analysis considered equal importance of members within a peer group, [20] improved the method by incorporating weight per group member into behavioural summary which they updated continuously based on new observed data. Larger weights in this regard are assigned to peer group members that are much closer to their targets. [3] furthered peer group analysis by developing Break Point Analysis for behavioral fraud detection. In contrast to peer group analysis that profiles groups, the techniques use patterns established within individual user account activity over a fixed length of time. The incoming new transactions are compared with a proportion of old ones within user profile and as such significant deviations through statistical tests are categorized as anomalies.

Related to peer group analysis in terms of objectives is the combination of k-means clustering and Breunig's [5] local outlier factor (LOF) with its variations used by [10] [13] [30]. Clustering and outlier detection tend to be treated differently by the community of data scientists [10] [8], but yet used to complement one another. On the one hand k-means algorithm groups instances together based on distance from the centroids while on the other hand LOF compares the density/distance of an instance to the density of its nearest neighbours. While k-means, LOF and peer group analysis can detect outliers without priori information but, they also have associated weaknesses associated with them. Of concern with peer group analysis according to [3] is caution that need to be exercised when deciding on the parameter *n-peer*, the peer group size which explains how local the model is and also control the sensitivity of the technique. With k-means the problem is attributable to initializing the value of k, number of clusters. The densities of the neighbourhood plays a pivotal role in LOF, but similar to peer group analysis the challenge can be on deciding on the number of nearest-neighbours since considering larger neighbourhood might lead to outliers looking normal.

III. Holistic Approach to Fraud Detection

The latest research in anomaly detection approaches considers the problem from multiple angles. A study by [12] takes cognizance a three-way look at anomaly detection problem as experienced by domain researchers. This include facts such as 1. datasets are unbalanced nature, 2. static approaches being common in usage and 3. the essence of the real-world operate in dynamical settings. Motivated by the notion of concept drifting from [26], [16], the study highlights the importance of previous patterns of anomalous behaviour together with emerging trends due to continuous incoming data streams. The approach proposed by [12] maintain this concept and hence assist in terms of being conscious of old forms of anomalies that may possibly strike in future. Because new information is being fed into detectors on continuous basis the *forget* approach takes that into consideration by concentrating on a predetermined number of latest data chunks and, thus taking control of computational burden. The three-way approach helps learn in the presents class imbalances by incorporating strategies like SMOTE as captured in [9]. While the general approach has been the static one which is faster but cannot adapt to changing distribution circumstances. It is both the $update$ and $forget$ approaches that are capable of adapting from changing circumstantial distribution changes. The disadvantage associated with the former is the need for several number of chunks containing the minority class while the former will need to propagate instances. Strategies employed in this regard addresses challenges prone to the anomaly detection space that emanate from the class imbalance problem, data stationarity versus streaming.

Related to the *update* and *forget* strategies that originated from the notion concept-drifting is an algorithm that mimics functioning of the human immune system (Artificial Immune Recognition System) explained by [6], [27] respectively and later adopted in [17] to detect online fraud. The AIS_Based Fraud Detection Model (AFDM) use scoring functions and cost of alarms in decision making. This anomaly detection paradigm when linked with the Numenta Anomaly Benchmark (NAB) [21] can be used to evaluate streaming data on real-world datasets.

This paper will take a dual approach in terms of defining fraud and non-fraud activity using global and local reference frameworks. Global outliers will be those instances further away from the nearest centroids established through trimmed k-means clustering while local outliers will be based on distances from the neighbourhood using kd-trees. An equivalent proportion of the top 1% from both approaches will be labelled out as fraud and thereafter assess the two spheres using classification-based methodological orientation outline in the following section.

IV. Methodology

In this section we provide a brief outline of the dataset used, followed by a description of all features that were extracted from the dataset. The section is concluded by describing the machine learning algorithm proposed for the study. For the anomaly detection techniques, it is assumed that the majority of the transactions on the network are legitimate with at maximum of only 1% being fraudulent.

A. Data Description

This study will use the Bitcoin dataset housed by the Laboratory for Computational Biology at the University of Illinois. All transactions from the genesis block to blockchain 230686 dated 7 April 2013. The blockchain under study is contains 6 336 769 users which in our case we refer to as nodes. In between the users are 37 450 461 edges (transactions) that link interactions amongst users.

B. Feature Extraction

Based on the measured variables provided by the Bitcoin network, the attempt is to build more meaningful features that will assist the learning algorithm in terms achieving the desired objectives. A total of 14 features were derived from the transaction data of the Bitcoin network. The following 14 features were derived from the dataset:

- Currency features: total amount sent, total amount received, average amount sent, average amount received, standard deviation received, standard deviation sent
- Network/graph features: in degree, out degree, clustering coefficient, number of triangles,
- Average neighbourhood (source target) whereby with reference to each query node: source refers to origin on incoming transaction and target is the destination. The four features identified: in-in, in-out, out-out, out-in. Source-target in this regard represent node in-degree and out-degree as reflected by in and out.

C. *Pre−Processing*

Given that the dataset lists all transactions that took place during the period under study, it is noted that some nodes were involved only in either sending or receiving Bitcoins . This led to the existence of missing values in the final dataset and hence imputation was in this regard exercised. All the missing values were imputed with zeroes based on the premise of equivalence to sending or receiving 0 BTC.

To have appropriate metrics between instances in our multivariate environment, we opted to transform our data by centering around mean zero and unit variance.

D. *Proposed Algorithms*

The objects contained by the Bitcoin network dataset as described in Part A are unlabelled, hence this study will start off with unsupervised learning techniques to group instances into sets based on outlierness. We broadly contextualize outliers into two main spheres, i.e. global and local outliers. Global outliers in this study are those instances farther away from the centroids in terms of Euclidean distance that will be attained through the use trimmed k-means founded by [11]. In contrast to global view, local outliers are defined as the top 1% instances with large distance to nearest predetermined number of neighbours when all instances are ranked according to ascending distance by using [1] and Bently [2].

The top 1% based on kd and trimmed k-means will be labelled as anomalies while the remainder of instances will be assigned $normal$ label. In this regard the two algorithms serve as proxies for normality and abnormality paving a way for assessment through supervised learning techniques. The methods are discussed in the following subsections.

1) Labeling global outliers: The clustering algorithm is adopted to assess node normality with reference to centroids based on Euclidean distances to nearby centres. Furthermore, a proportion of the top 1% of instances that are farthest from centroids are presumed to be fraud and labelled as such. Trimmed k-means in this regard provide a means to segment the instance space into two groups with regard to an instance being normal or an anomaly. Such a binary space will be categorized into global anomaly and global normality subspaces.

2) Labeling local outliers: The kd-trees was founded by Bently [2] in the information retrieval domain. The technique is a data structure type for associative searches that forms part of the broader binary space partitioning schemes. Generally the method regard the entire data space as the root node which is recursively partitioned into smaller hyperspaces by predefined rule according to [19].

3) Contextual Binary Classification: Based on the two broad categories by kd-trees and trimmed k-means clustering, three binary classification algorithms are proposed to further explain the detected outliers. The algorithms include are maximum-likelihood based logistic regression,boosted logistic regression and random forest whose general background can be found on [24], [15], [4] respectively. These models are used in this study to further deepen understanding of the relationship between fraud (outcome of interest) and predictors. Given the data is unlabelled, of importance is to establish whether fraudulent activities can be explained by the extracted feature intrinsic to the network graph and transactions [25].

V. Experimental results

In this section, a holistic view on the labeling of the instance space is briefly explained as well as how anomalous behaviour relates with the 30 known fraudulent activity on the Bitcoin network with reference to both global and local anomalies. The two algorithms differs significantly in terms of computational resource requirements; hence kd-trees was implemented on the full dataset while trimmed k-means was only limited to the first 1 million instances when arranged in descending node/user number. This decision is motivated by the need for optimal resource usage coupled with the fact that 28 of the known outliers falls within the first million.

1) Known fraudulent network users: The key challenge in this study relates to lack of ground truth which plays a pivotal role in machine learning applications. Out of nearly 37.5 million transactions under study, only 76 have been labelled as fraud which when aggregated to user level result with only 30 fraudsters of over 6 million nodes/users. This situation prompted for detection of fraudulent activities through unsupervised algorithms as per the results below.

A. *Unsupervised Detection Level*

1) Detect and assign labels with trimmed k-means and kd-trees: Clustering and binary space partitioning algorithms are adopted to partition instances into two groups under the assumption of the existence of 1% fraudsters on the Bitcoin peer-2-peer network. The variability between clusters show that the resultant number of clusters is 8 as shown on Figure 1. Subsequent to trimmed k-means, a predetermined bucket of k=7 (number of neighbours) similar to [22] is used with kd-trees under similar quantitative assumption of normality.It must be noted that the number of outliers based on kd-trees exceeds those derived from k-means by a margin of 193. This disparity is a direct result of the established fact that the cut-of value based on k-means is a single observation while with kd-trees multiple observations existed and hence all where considered as outliers.

Both global and local outlier search methods successfully detected some of the known crimes[1] as shown on Table I. On the one hand, from a global outlier perspective only 5 users have been detected while on the other hand local

[1] https://bitcointalk.org/index.php?topic=576337

approach filtered out one more (i.e. 6 in total) successfully when considering the first 1 million instances. On the full dataset, kd-trees detected 7 of the known thefts. This study takes cognizant of the fact that though kd-trees detected all those filtered with trimmed k-means. It must be noted that of the total outliers presumed by the study, the two algorithms remained in agreement in 22% of the time.

In the midst of detection success in relation to known incidents as well as total overlaps, it will remain difficult to assess superiority of the techniques regarding performance based on few known criminal incidents. In the following subsection, classification-based approach is used to assess how well the extracted features describes the Bitcoin peer-2-peer networks from the two outlined perspectives. It must be noted that interpretation of this approach must be limited to the two spheres established in this paper.

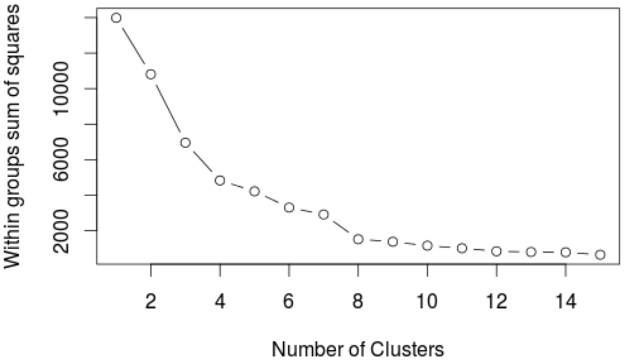


Fig. 1: The elbow chart showing optimal clustering attained at k=8

TABLE I: Known crime associated with both global and local detection algorithms

ID	Name of Entity	Method of Detection	
		k-means	kd-trees
11	Mt. Gox	X	X
224 319	Linode Hacks	X	X
377 918	2012 50BTC Theft	X	X
442 450	Linode Hacks	X	X
586 816	Linode Hacks	X	X
727 888	Allinvain Theft		X
4 725 674	Stone Man Loss		X

B. Supervised Detection Level

The binary classification findings based on logistic regression from both general linear models (GLM) and General Additive Models (GAM) as well as random forests using a 10-fold cross validation with 70% of instances during training and remainder reserved for testing. The results associated with each model are encapsulated in Table II to IV. Furthermore, a snapshot on feature selection is included to compare modeling results based on attribute importance.

1) GLM Logistic Regression: This algorithm is adopted to serve as a baseline to understanding the ground necessary towards complex binary models. Given the scenario is two-class, the Binomial link function is employed. The model shows excellent results from a global perspective (Table IV) with regard to recall and precision levels of approximately 98% and 75% respectively. Although the local model is capable of detecting the positive class (98%), precision in this instance dropped to lowest levels of 33% given the class prevalence. This vindicates that the local model suffers more than the global one with regard to imbalances associated with established class labels.

There is a wide gap in terms of classification agreements between global and local viewpoints as measure by the Kappa statistic. While the global classification exhibit excellent agreement with a Kappa value of 85%, the local sphere was found to be moderate at 48%. Although this findings are considered good, they vindicate that the percentage agreement that would occur by chance lean more towards the local classification scheme.

TABLE II: Consolidated confusion matrix of the three models from a global contextual viewpoint

Global Reference						
	Fraud			Normal		
Prediction	GLM	GAM	RF	GLM	GAM	RF
Fraud	2258	179	2977	742	2821	23
Normal	30	0	23	296970	297000	296977

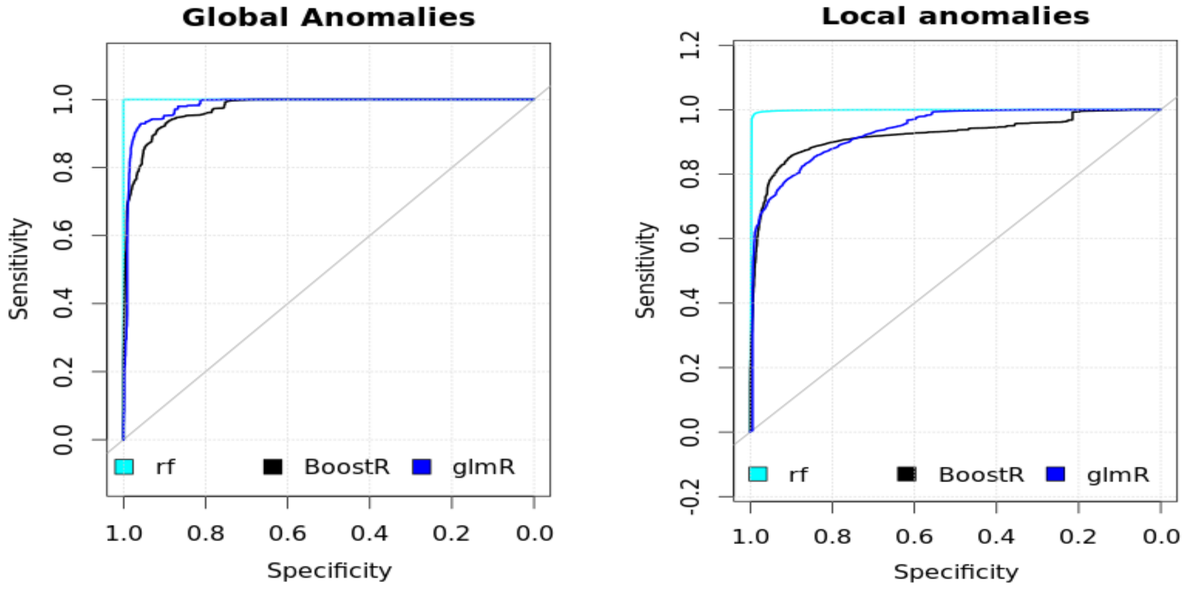
TABLE III: Consolidated confusion matrix of the three models from a local contextual viewpoint

Local Reference						
	Fraud			Normal		
Prediction	GLM	GAM	RF	GLM	GAM	RF
Fraud	1012	107	2381	2045	2950	242
Normal	155	5	676	296787	296937	296700

TABLE IV: Model comparison based on three performance measures in alignment with context

	Global			Local		
Model	**Recall**	**Kappa**	**Precision**	**Recall**	**Kappa**	**Precision**
GLM	0.98	0.85	0.75	0.86	0.48	0.33
GAM	1.00	0.11	0.06	0.95	0.07	0.04
RF	0.99	0.99	0.99	0.78	0.84	0.91

2) Random Forest: The final random forest model was attained to create optimal hyper-parameters. The algorithm was ran using different number of features at each split with optimality realized at 8 features. Although the results shows that optimality is realized with 8 features from both a local and global angle, loss value appears to be fairly small. This implies that similar results attained with 8 features can be attained when using only 2 features. Although results on Table IV shows that random forest performs well from both contextual viewpoints, there appeared to be a slight drop when crossing to the local perspective. In terms of the out of bag error rate, the globalized random forest perform better with 0.02% in contrast to the localized model with 0.3%. The results are duped more reliable with a Kappa-value of 99% and 83% from global and local dimension respectively.



(a)

(b)

Fig. 2: Performance comparison of MLE-based logistic regression, boosted regression and random forest using (a). global and (b). local outliers .

3) Additive/Boosted Logistic Regression: This model is adopted to compensate for potential bias associated with generalized linear model and because of its feature selection capabilities that takes place during model-building process. Optimal model performance was attained at 250 boosting iterations. The performance of this model with regard to the class of interest (anomaly or fraud) is poor as outlined by the hit rate on Table II and III that is far below competing models from both sides. Other performance measures on Table IV bear evidence that boosted regression suffers more from class imbalances when compare with GLM regression and random forest. This is proved true by high recall values accompanied by poor Kappa statistics showing that the detection of positives is likely by chance.

4) Importance of Predictors: To under the importance of features for training, the score for each variable was computed using the entire training dataset. The score ranges from 0 to 100, but for ease of readability scoring containers of the form (10,20],....,(90,100) were replaced with value labels 1 to 10 across all models. The result from feature importance are categorized as per Table V below.

The logistic regression built during training phase vindicated that not all predictors add value to the model. It is also important that variable importance differs according to which direction one is looking at it from, either global or local viewpoint.

TABLE V: Feature importance by model-type categorized according to outlier context

Feature	Global			Local		
	GLM	Boost	RF	GLM	Boost	RF
AverageIn	10	3	8	6	1	1
AverageOut	4	2	3	10	2	2
TotalIn	1	2	1	7	9	10
TotalOut	1	2	1	4	9	3
DeviationIn	1	6	1	2	3	3
DeviationOut	3	2	2	4	1	1
Triangles	7	7	3	9	10	6
ClusterCoeff	3	8	3	4	1	3
NumberOutDeg	1	1	1	3	7	1
NumberInDeg	1	4	1	1	9	2
AvgNeighborDeg(In_In)	5	10	10	6	3	2
AvgNeighborDeg(In_Out)	3	10	9	6	3	2
AvgNeighborDeg(Out_Out)	3	10	2	1	4	4
AvgNeighborDeg(Out_In)	3	10	2	1	4	4

From a global sphere the variable found to be statistically significant considered both an inward and outward movements of average value of Bitcoins involved as well as the associated standard deviations. The total value of Bitcoins played no pivotal role in this space. Of all the graphical attributes used to model the logistic regression, it is the number triads/triangles and the node degree (number in and out) found to be instrumental in terms of describing anomalous behaviour of the Bitcoin network. Of all graphical parameters adopted in the global outlier viewpoint, none of the average-neighbourhood attributes were found to add value to the predictive model.

When taking a peep from local outlier perspective, it was found that of all financial attributes, only average value sent, total received, standard deviation sent and standard deviation received played a critical role in terms predicting outliers. In contrast to the global viewpoint with reference to graphical parameters, local outlierness can only be predicted by node out-degree, triads, clustering coefficient couple with one of the four average-neighbourhood attributes (out-degree by source and target).

When considering the two ensembles, feature selection was

TABLE VI: Computed AUC for random forest, GLM and GAM regression models by context

Model	Global	Local
GLM Logistic	0.98	0.93
Boosted Logistic	0.97	0.92
Random Forest	0.99	0.99

found to vary by both the model type as well as contextual viewpoint. From a global classification perspective, boosted logistic shows higher scores leaning towards graphical parameters which is in contrast to random forest which is fairly balanced between financial and graphical features. Furthermore, the algorithms tends to be in agreements from a local perspective as vindicated the extend of score similarities.

5) Validation of Predictive Abilities: From receiver operating characteristics (ROC) curve shown Figure 2, the area under the curve was calculated for each model as tabulated in Table VI. The performance of all the three models exceeds 90% with random forests leading (99.99%), followed by GLM logistic and boosted regression models.

VI. Conclusion

This paper used global (trimmed k-means) and local (kd-trees) algorithms for fraud detection over the Bitcoin network. Furthermore, based on findings of this algorithms instances were assigned class labels as either global or local normal/anomaly. The labeling paved a way for classification-based learning to further deepen understanding of the links between the response and predictor variables. Although local detection techniques performed better than global in terms of known fraudulent incidents, the latter surpassed the former across all models under study with regard to performance metrics adopted in this paper. Of the three models, random forest was found to be the best performing classifier with 8 features irrespective of class imbalances. The findings by the adopted supervised learning algorithms revealed the most important features used in this particular study that were useful in finding perceived anomalous activities on the network.

References

[1] Deepali Aggarwal, Pankaj Singhal, Sharanjit Kaur, and Vasudha Bhatnagar. Pdod: Tree based algorithm for outlier detection. In *Proceedings of Int'l Conference on Computer Vision and Information Technology*, 2007.

[2] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

[3] Richard J Bolton, David J Hand, et al. Unsupervised profiling methods for fraud detection. *Credit Scoring and Credit Control VII*, pages 235–255, 2001.

[4] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[5] MM Breunig, HP Kriegel, RT Ng, and J Sander. Lof: identifying outliers in large dataset. In *Proc. of ACM SIGMOD*, pages 93–104, 2000.

[6] Jerome H Carter. The immune system as a model for pattern recognition and classification. *Journal of the American Medical Informatics Association*, 7(1):28–41, 2000.

[7] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

[8] Khyati Chaudhary, Jyoti Yadav, and Bhawna Mallick. A review of fraud detection techniques: Credit card. *International Journal of Computer Applications (0975–8887) Volume*, 2012.

[9] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[10] Sanjay Chawla and Aristides Gionis. k-means-: A unified approach to clustering and outlier detection. In *SDM*, pages 189–197. SIAM, 2013.

[11] JA Cuesta-Albertos, Alfonso Gordaliza, Carlos Matrán, et al. Trimmed k-means: An attempt to robustify quantizers. *The Annals of Statistics*, 25(2):553–576, 1997.

[12] Andrea Dal Pozzolo, Olivier Caelen, Yann-Ael Le Borgne, Serge Waterschoot, and Gianluca Bontempi. Learned lessons in credit card fraud detection from a practitioner perspective. *Expert systems with applications*, 41(10):4915–4928, 2014.

[13] Lian Duan, Lida Xu, Ying Liu, and Jun Lee. Cluster-based outlier detection. *Annals of Operations Research*, 168(1):151–168, 2009.

[14] Zakia Ferdousi and Akira Maeda. Unsupervised outlier detection in time series data. In *Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on*, pages x121–x121. IEEE, 2006.

[15] Jerome Friedman, Trevor Hastie, Robert Tibshirani, et al. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2):337–407, 2000.

[16] Jing Gao, Wei Fan, Jiawei Han, and S Yu Philip. A general framework for mining concept-drifting data streams with skewed distributions. In *SDM*, pages 3–14. SIAM, 2007.

[17] Neda Soltani Halvaiee and Mohammad Kazem Akbari. A novel model for credit card fraud detection using artificial immune systems. *Applied Soft Computing*, 24:40–49, 2014.

[18] Douglas M Hawkins. *Identification of outliers*, volume 11. Springer, 1980.

[19] Seung Kim, Nam Wook Cho, Young Joo Lee, Suk-Ho Kang, Taewan Kim, Hyeseon Hwang, and Dongseop Mun. Application of density-based outlier detection to database activity monitoring. *Information Systems Frontiers*, 15(1):55–65, 2013.

[20] Yoonseong Kim and So Young Sohn. Stock fraud detection using peer group analysis. *Expert Systems with Applications*, 39(10):8986–8992, 2012.

[21] Alexander Lavin and Subutai Ahmad. Evaluating real-time anomaly detection algorithms–the numenta anomaly benchmark. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 38–44. IEEE, 2015.

[22] Phillip Thai Pham and Steven Lee. Anomaly detection in bitcoin network using unsupervised learning methods.

[23] Roland Rieke, Maria Zhdanova, Jurgen Repp, Romain Giot, and Chrystel Gaber. Fraud detection in mobile payments utilizing process behavior analysis. In *Availability, Reliability and Security (ARES), 2013 Eighth International Conference on*, pages 662–669. IEEE, 2013.

[24] Saroje K Sarkar, Habshah Midi, Md Rana, et al. Detection of outliers and influential observations in binary logistic regression: An empirical study. *Journal of Applied Sciences*, 11(1):26–35, 2011.

[25] Véronique Van Vlasselaer, Cristián Bravo, Olivier Caelen, Tina Eliassi-Rad, Leman Akoglu, Monique Snoeck, and Bart Baesens. Apate: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, 75:38–48, 2015.

[26] Haixun Wang, Wei Fan, Philip S Yu, and Jiawei Han. Mining concept-drifting data streams using ensemble classifiers. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 226–235. ACM, 2003.

[27] Andrew Watkins, Jon Timmis, and Lois Boggess. Artificial immune recognition system (airs): An immune-inspired supervised learning algorithm. *Genetic Programming and Evolvable Machines*, 5(3):291–317, 2004.

[28] Sarah N Welling. Smurfs, money laundering, and the federal criminal law: The crime of structuring transactions. *Fla. L. Rev.*, 41:287, 1989.

[29] David J Weston, David J Hand, Niall M Adams, Christopher Whitrow, and Piotr Juszczak. Plastic card fraud detection using peer group analysis. *Advances in Data Analysis and Classification*, 2(1):45–62, 2008.

[30] Gao Zengan. Application of cluster-based local outlier factor algorithm in anti-money laundering. In *2009 International Conference on Management and Service Science*, 2009.

[31] Maria Zhdanova, Jurgen Repp, Roland Rieke, Chrystel Gaber, and Baptiste Hemery. No smurfs: Revealing fraud chains in mobile money transfers. In *Availability, Reliability and Security (ARES), 2014 Ninth International Conference on*, pages 11–20. IEEE, 2014.