A Cascaded Artificial Neural Network and K-Medoids Method for Money Laundering Detection in Financial Organizations

J. K. Alhassan, *Member, IAENG*, and N. E. Yusuf

Abstract— **This work is on money laundering discovery in monetary organizations with a cascaded artificial neural network and k-medoids approach. The overbearing of averting cash laundering wrongdoings is not because the launderers takes unjustified benefit of the economy for illegal financial achievement but also since cash laundering is a misconduct that is typically associated to criminalities like fraud, Illegal trafficking in drugs, abduction, corruption, arms trafficking, terrorism, among others. This research used a cascade model of K-medoids and Artificial neural network (ANN) to design a robust detection model for money laundering. K-medoids method is used to cluster the dataset into two clusters, its output is filtered, and then passed into ANN for training and testing purposes and classified into doubtful and un-doubtful transactions. Other algorithms like Support Vector Machine (SVM) and ANN were used separately on the dataset for detection purposes. Their performances were matched with the Cascade K-medoid-ANN. Results shows that ANN has an accuracy of 80.3%. SVM has an accuracy of 70.6% and Cascade K-medoid-ANN has an accuracy of 74.5%. Other metrics used includes; Sensitivity, Specificity, and Precision. Cascade K-medoid-ANN outperformed other algorithms used by correctly identifying doubtful transactions as doubtful and un-doubtful transactions to be un-doubtful. It is recommended that all monetary organizations work together to offer a warehouse of cash laundering dataset for tackling menace and other bank frauds**.

***Index Terms*— Artificial Neural Network, Financial organization, K-Medoids, Money laundering**

I. Introduction

ARTIFICIAL Artificial neural network (ANN) encompasses of elements termed processed neurons. Artificial neuron tries to replicate the structure of natural neuron and conduct. A neuron contains of single yield (synapse through axon) and single input or dendrites. The neuron's activation defines its feature. A neuron is a component that procedures information that is vital for the operation of ANN. It is a parallel spread processor containing of easy handling components with a disposition which stock and make available for empirical knowledge usage. It looks like the brain: the network obtains information from its setting vial a technique of learning and inter-neuron connection powers, denoted as synaptic weights, they are used to stock the knowledge gotten. ANN have been useful in diverse areas such as monetary market predicting, credit card fraud discovery and risk assessment.

The *k*-medoids is an algorithm which is used for clustering which partitions (dividing the dataset to clusters) and try to decrease the space amid points considered to be in a group and a point allocated the centre of that group. *K-medoids* selects data plugs as centres (exemplars or medoids) random distances can be used. These k medoids denotes diverse essential facets of the scope N data set that are separated to exhaustive clusters and mutually exclusive around k medoids, wherever a medoid is the entity of the cluster which the amount of spaces to all other cluster is negligible. This method is robust in contradiction of outliers, noise or barely disseminated information as a result of medoids [1].

All over the world, there are huge volumes of monetary dealings vial diverse means each minute. These comprise of fraudulent transactions by criminals that have initiated damage to monetary organizations or clienteles, or cause reputational injury, that is tough to overhaul [2]. Money laundering is one of such fraudulent transaction, which is a way of hiding the source of moneys illicitly gotten and accumulated after a while [3]. By cash laundering, criminals attempt to opaque the real source of cash acquired from unlawful action.

Cash laundering is a world-wide problematic issue with profuse damaging effects on society. The influence remains extra overwhelming in the emerging financial prudence with feeble monetary controlling scheme and starting where the moneys are relocated to advanced countries to obtain glamorous extravagance substances. In specific, laundering of cash offers a way of obtaining proceeds for the offenders, providing free money which may be for financing additional criminal actions. Furthermore, cash laundering can severely undermine sureness in monetary structures and financial organizations, and cause harm to local economies.

It is perceived that cash laundering takes countless consequence on the Nigerian budget [4]. Notwithstanding the policies and laws passed, cash laundering with additional monetary and fiscal crimes flourish in the Nigeria [5]. This research is on a cascaded approach of ANN and K-medoid for noticing cash laundry in Nigeria by means of unidentified transaction data from manifold banking sources. The remains part of this work is organized as

Manuscript received December 28, 2019; revised February 3, 2020.

J. K. Alhassan is with the Federal University of Technology, Department of Computer Science, Niger State, Minna, Nigeria, Phone: +2348035961620; (e-mail: jkalhassan@futminna.edu.ng)

N. E. Yusuf was with the Federal University of Technology, Department of Computer Science, Niger State, Minna, Nigeria, (e-mail: nimah93@gmail.com).

follows: related studies, research methodology, results and discussion and references.

II. RELATED STUDIES

[4] analyzed the effects of money laundering on the economy of Nigeria; socio-economically, monetarily, administratively, in engineering of local goods and in the oil and gas area. They resolved that, it has a vast influence on the budget of Nigeria. It is vital to annotate that in spite of the nation's regulation and rules, cash laundering and additional monetary and fiscal crimes are thriving in the country due to immoral means by civic officers. Cash laundering has been perceived to take excessive consequence on the Nigeria budget [6]. Administrations are worried about hazards of cash laundering, it supports financing of terrorism and reduces administration tax incomes in numerous means [7].

Monetary criminalities in Nigeria generally affected the banking sector according to [6]. Banks likewise worsen the situation since they refuse to cooperate and offer the right institutions with data on dealings transitory vial their institution on cash laundering. [8] discovered the concerns of with machine learning approaches to recognize cash laundering in a data set collected of synthetic monetary dealings and aimed at noticing anomalies inside an information set of mobile cash monetary transactions by categorizing a collection of dealings as doubtful or un-doubtful.

[4] stated that, data excavating methods might be engaged in determining deal forms that lead to cash laundering. By allowing for customer danger valuation data, deal peril quantity data and behaviour outlines in noticing cash laundering forms. Centred on the resemblances in data, dealings are grouped [3]. The impression of merging classifiers was advised as a pure method of enhancing the competence of distinct classifiers [9].

Funds gotten unlawfully in cash laundering are combined into monetary organizations by lawfully gotten revenue and later allows differentiating lawfully gotten income and unlawfully gotten funds problematic. In examples wherever anti-cash laundering rules and guidelines are not active and rule dodging consequences are not measured, cash laundering incidence is unavoidable [10].

The cases wherever crowding approaches used to group previously highlighted, perhaps via classifier of deceitful items. In this situation, the grouping objective is to recognize a classification of the scam items previously known to present countermeasures for each group of exposed scam [11].

[12] experimentally recognised that ANN not merely gives optimistic foretelling exactness, but can have a superior discovery capability and abridged rate of incorrect classification. As grouping is a supervised knowledge, it need the understanding of classes earlier, absence of preceding understanding of laundered accounts may well be a limitation.

[9] indicated that, SVM's powers are recognised for correctness, quickness of taxonomy, patience to immaterial and redundant qualities, whereas nearly of its errors are knowledge rate due to quantity of qualities and cases, prototype factor treatment, capability to enlighten, exactness in taxonomies. To sustain the exactness of finding and drop statistics of false positive, anti-money laundry (AML) scheme must be clever to grip finding of uncommon forms from covered dealings and estimate unseen cases with no importance on the performance correctness and mechanize analysis of dealings. A unique AML architectural system involves a set of training data (categorized) comprising information of earlier acknowledged doubtful business deal and normal dealings [12].

[13] noted that, data excavating structures can be betrothed in finding business deal forms associated to cash laundering by bearing in mind customer danger valuation data, business deal danger dimension data and behaviour forms in observing cash laundering forms. They detailed roughly keys for defining cash laundering built on the resemblances in the data, dealings and clustered.

[14] derived a procedure for analysis of monetary dealings by means of unsupervised knowledge tactics, they applied irregularity uncovering algorithm to get doubtful ranking for user financial records.

[15] suggested outlier of auto-regression algorithm for cash laundering finding by means of Mean, Inter- quartile Range (IQ), auto-regression and Zero-mean procedures on mined data for revealing of cash laundering deeds.

[16] suggested an algorithm of network-based to sieve dealings, further used clustering way to recognise doubtful cash laundering forms inside network. Forms are re-arranged, organized and reverted by way of outcome by the structure.

[8] examined the inferences of smearing machine learning strategies in AML starting with set of artificial monetary dealings meant at realizing irregularities inside a dataset of mobile cash financial dealings by means of client profiling to cluster dealings as doubtful or un-doubtful.

[17] assessed four (4) classification performance of kinds of ANN for AML; Probabilistic neural network, Multi-layer neural network (MLP), Linear neural network (LNN) and Radial basis function (RBF). LNN ensured the finest grouping performance degree and highest performance of training, choosing and testing the data with abridged mistake degree.

[14] suggested the mix of Binary class SVM and sole class SVM for cash laundering recognition and additional bank scams, that stops the circumstance of untrue generality in usage of the twofold learning by sieving the single class learning with positives. They suggested that situations wherever there is no preceding database on dishonest dealings should use binary class and normal dealings. In incidences wherever such record does not exist in the bank and all dealings are supposed thorough, it is suitable to only learn in solitary class. They achieved 80% of exactness for solitary class SVM system.

[18] suggested the usage of cross authentication to enhance SVM factors once the total working of the model is finest, to evade above knowledge and below knowledge. He established that SVM model taught by cross authentication was extra in effect than the model gotten with arbitrarily elected classification result parameters. [15] offered a group of OCSVM, GMM and IF in finding of doubtful financial

ISBN: 978-988-14049-2-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

records, luckily their model revealed doubtful circumstances, the existing rule-based scheme can recognize and additional circumstances of doubtful financial records.

[19] established network based scheme for realizing doubtful clusters of dealings. The scheme prototypes the association among groups that uses a graph with traits. Subsequently, groups are mined from the business deal graph and computed. After that supervised learning was used on the groups to get classifiers that are trained. The performance of the system revealed a countless level of correctness and exactness.

[2] Applied a scheme by picking main research variables for cash laundering in venture banks, at that juncture suggesting a neural network-based examination mode and clustering to recognize dubious cash laundering cases. Heuristics like suspect screening have been applied to improve running time.

III. Research Methodology

A. Data Collection

Data acquired from a virtual source (github) were used for this study were. Current literature used parameters in observing cash laundering financial records; Total received, debit/credit business deal incidence, amount withdrawn, danger value, sender/receiver distinct account history and individuals' salary information. For data groundwork, guaranteeing the excellence of data, pre-processing is achieved to eliminate immaterial features, noise, redundant features, and examples that are characterized erroneously. MATLAB R2017b was used first to pre-processed the dataset by take-out non-suspicious and suspicious dealings by means of the after procedures (William-McKee, cash laundering test).

i). The second transaction sender Id matches first transaction receiver Id

ii). Between 90% and 100% of first transaction amount is second transaction amount

iii. On the same day the two transactions occurred.

The summation of 500 doubtful dealings were sifted as of the dataset and balance the portion of mutually classes, 500 instances of usual dealings were additional to form 1000 testing and training dataset. The goal tag for doubtful dealings is fixed to 1 whereas 0 for non-doubtful (usual) dealings. The timestamp trait was changed to number data kind afterward that standardization was completed to avert prejudice of the classifier about traits with great values.

B. Design of the Model

MATLAB was used to train the dataset with a fixed of doubtful dealings, afterward learning; used to categorize group of dealings to classes of non- doubtful and doubtful dealings. In probing factors, several training algorithms were explored counting scaled conjugate, Levenberg-Marquardt and Bayesian regularization training algorithm, Bayesian regularization was nominated as it performed well than others. So was the space metric for K-medoid, Mahalanobis presented appropriate. Out of the scrutinized data, greatest of the doubtful dealings happened at the close of month, few at mid of the month. It was decided that timing could be a gage for lawbreakers, they might aim ending of month to distribute cash (payment of salary) or few might aim initial days or closely extra time of the month.

C. K-Medoid and ANN Cascade

The used data was separated to (2) two, the training group and testing group. The training set was handed for clustering to K-medoid algorithm; It was used to group the dataset to clusters of two that was dogged via the quantity of classes planned for the dataset outcome, meanwhile clustering of non-doubtful dealings and doubtful dealings were used. K-medoid used mahalanobis space which was a measure of no unit computed by means of the standard deviation and mean of the data, and financial statement for correlation in the data. Afterward, the data was sieved and ANN was used for training model. Subsequently test dataset was passed for testing to ANN and lastly its performance stood assessed.

Adjusting K-medoid model equation by Mahalanobis distance equation provides equation (1).

$$Z = \sum_{i-1}^{k} \sum |\Delta| \qquad (1)$$

Where;

Δ = Mahalanobis distance $d(x,y)$ amid n-dimensional points x and y, with reverence to a certain n-by-n covariance matrix S, is

$$\mathrm{d}(\mathrm{x},\mathrm{y}) = \sqrt{(\mathrm{x}-\mathrm{y})^{\mathrm{T}}\mathrm{S}^{-1}\ (\mathrm{x}-\mathrm{y})} \qquad (2)$$

$$\mathrm{Z} = \sum_{\mathrm{i}-1}^{\mathrm{k}} \sum |\mathrm{d}(\mathrm{x},\mathrm{y}) = \sqrt{(\mathrm{x}-\mathrm{y})^{\mathrm{T}}\mathrm{S}^{-1}\ (\mathrm{x}-\mathrm{y})}| \qquad (3)$$

A distinctive model neural network is shown in (4).

$$\mathrm{A} = \mathrm{f}(\mathrm{w}*\mathrm{p}+\mathrm{b}) \qquad (4)$$

Apiece component of the network with vector **p** input is linked to neuron each input via the weight matrix **W**.

$$\mathrm{a} = \mathrm{Z}(\mathrm{wp} + \mathrm{b}) \qquad (5)$$

Equation (5) gives the cascade of ANN and K-medoid resultant by adjusting (4). Two layers ANN was used, therefore

$$\mathrm{a}_2 = \mathrm{Z}(\mathrm{LW}_{2,1}\left(\mathrm{IW}_{1,1}\mathrm{P} + \mathrm{b}^1\right) + \mathrm{b}^2) \qquad (6)$$

Where;

a = signifies the model output,

Z = signifies sieved data inputted to network that are clustered

ISBN: 978-988-14049-2-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

W = signifies weights
P = signifies input vector

Indicated in Figure 1 is the flowchart of the cascade K-medoid and ANN model.

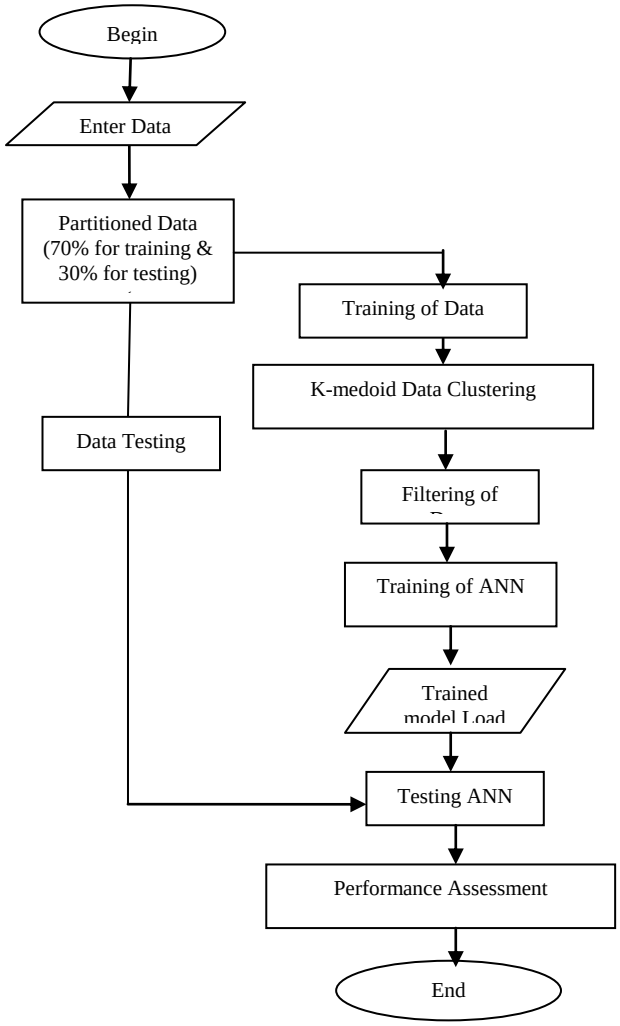

Fig. 1: Flowchart of the Cascade Model

IV. Results and Discussion

A. Cascade of K-Medoid and ANN

Data divided to (2) two; training dataset and testing dataset in ratio 70:30. Obtainable of 1100 cases, 770 inputted as training data into K-medoids for assembling, subsequently data was sieved to realize accurate grouping built on forecasters set. Aggregate of 425 trials were acquired. The result of K-medoid was inputted to ANN for training; subsequently the outstanding 30% fraction was used for testing the model that produced an accuracy of 74.5% of properly organizing the data in the confusion matrix. Shown in Figure 2 is the confusion matrix of the cascade model of K-medoid and ANN. After testing, 81 instances were correctly classified to be non-suspicious, 84 instances were incorrectly categorized to non-suspicious. While 165 instances were correctly categorized as doubtful and no instance was wrongly classified to be suspicious

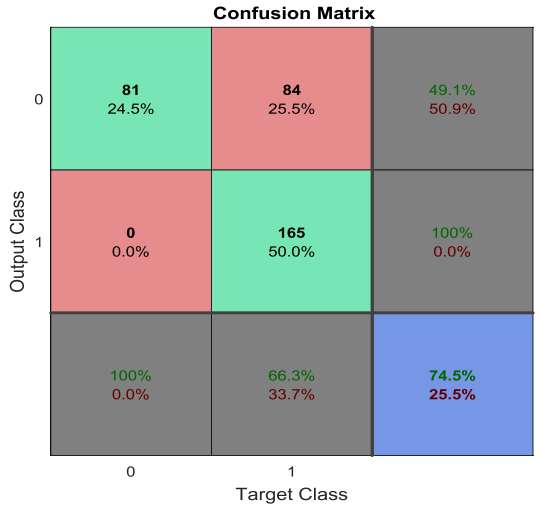

Fig. 2: K-medoid-ANN

Figure 3 is a performance plot portraying the training block, testing performance of the model and validation block. Its best authentication performance was at seventeen (17) iterations.

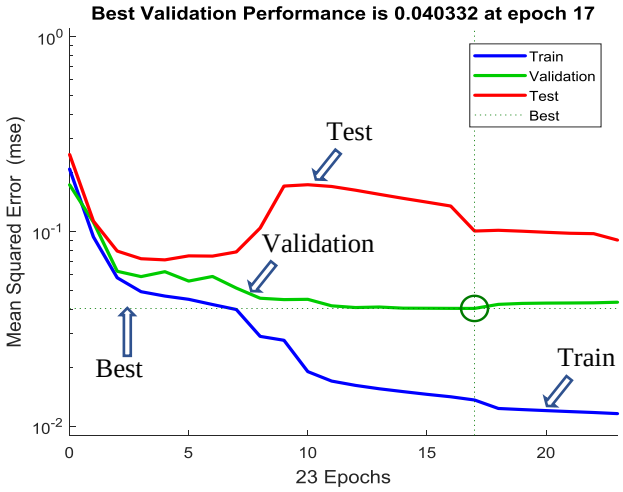

Fig. 3: Performance of the model

The training state performance is shown in figure 4, of K-medoid-ANN cascade at twenty-three (23) epochs with a gradient slope value of 0.057187 and twenty-three (23) epochs against Validation check plot is six (6).

ISBN: 978-988-14049-2-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

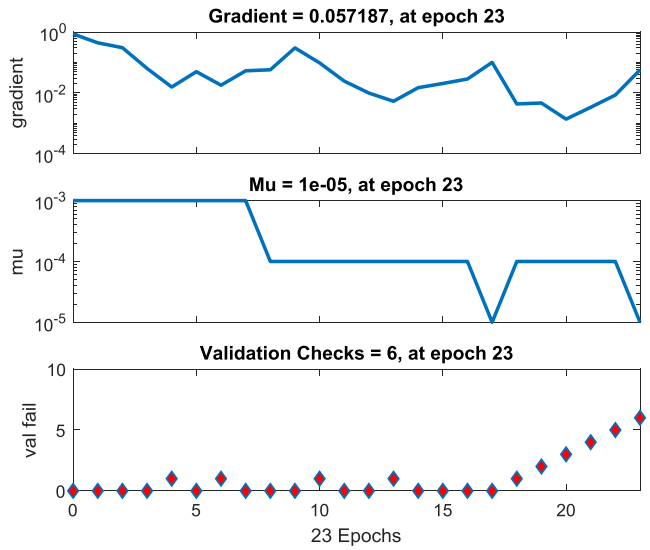

Fig. 4: Performance of K-medoid-ANN at training state

The performances of each of the algorithms used in the course of this study are outlined in Table 1.

TABLE I
PERFORMANCE OF ALGORITHMS USED

SNo	Algorithm	Accuracy	Precision	SENSITIVITY/RECALL	Specificity
1	ANN	80.30%	69.70%	88.50%	75.00%
2	SVM	70.60%	49.70%	85.40%	64.50%
3	K-Medoid-ANN	74.50%	49.10%	100.00%	66.30%

V. CONCLUSION

The number of neurons and layers on the model was single-minded, also the number of groups and space measure for K-medoid. The trained model with algorithms of Bayesian regularization (BR) presented well outcome contrasting scaled conjugate and Levenberg-Marquardt (LM) algorithm. The Cascade model of K-medoid-ANN was attained by using K-medoid for clustering the transactions into two clusters, then filtering the dataset based on some well-defined guidelines, and lastly passing the filtered dataset into ANN for classification into doubtful and non- doubtful transactions. The model performance was assessed using metrics like; Specificity, Accuracy, Precision and Sensitivity. A contrast of SVM and ANN outcomes disclosed that ANN outdid SVM exceptionally. On contrast with the new method Cascade of ANN and K-medoid, outcome of accuracy is 74.5%. It is recommended that all monetary organizations work together to offer a storehouse of cash laundering dataset for confronting cash laundering and other bank scams.

REFERENCES

[1] S. Modak, T. Chattopadhyay, A. K. Chattopadhyay, (2019). Unsupervised classification of eclipsing binary light curves through k -medoids clustering. *Journal of Applied Statistics*, *0*(0), 1–17. Available at https://doi.org/10.1080/02664763.2019.1635574

[2] N. Le Khac, S. Markos, M. T. Kechadi, (2010). A data mining-based solution for detecting suspicious money laundering cases in an investment bank.

[3] K. V. Manjunath, (2015). *Data Mining Techniques for Anti Money Laundering*. *2*(8), 819–823.

[4] A. O. Oluwadayisi, M. O. Mimiko, (2016). *Effects of Money Laundering on the Economy of Nigeria*. (June), 158–169.

[5] S. Agarwal, S. Upadhyay, (2014). *A Fast Fraud Detection Approach using Clustering Based Method*. *1*(10), 33–37

[6] E. Christopher, E. A. Ibanichuka, (2016). Money Laundering and Forensic Accounting Skills in Nigerian Banks. *Research Journal of Finance and Accounting*, *7*(15), 149–155.

[7] C. Suresh, K. T. Reddy, N. Sweta, (2016). *A Hybrid Approach for Detecting Suspicious Accounts in Money Laundering Using Data Mining Techniques*. (May), 37–43. https://doi.org/10.5815/ijitcs.2016.05.04

[8] E. A. Lopez-rojas, S. Axelsson, (2012). *Money Laundering Detection using Synthetic Data*. (071), 14–15.

[9] S. J. Omar, K. Fred, K. K. Swaib, (2018). *A State-of-the-Art Review of Machine Learning Techniques for Fraud Detection Research*. 11–19.

[10] B. Bidabad, (2016). *Money Laundering Detection System (MLD) (A Complementary System of Rastin Banking)*. (Mld)

[11] S. Soltaniziba, M. A. Balafar, (2015). *The Study of Fraud Detection in Financial and Credit Institutions with Real Data*. *5*(2), 30–36. https://doi.org/10.5923/j.computer.20150502.02

[12] Z. Chen, L. Dinh, V. K. Ee, N. Teoh, A. Nazir, E. Kandasamy, S. Lam, (2018). *Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection : a review*

[13] D. Abdelhamid, S. Khaoula, O. Atika, (2014). *Automatic Bank Fraud Detection Using Support Vector Machines*. 10–17

[14] R. Camino, V. Petko, (2017). *Finding Suspicious Activities in Financial Transactions and Distributed Ledgers*. https://doi.org/10.1109/ICDMW.2017.109

[15] S. Kannan, K. Somasundaram, (2017). *Article information : Users who downloaded this article also downloaded : About Emerald www.emeraldinsight.com*. https://doi.org/10.1108/JMLC-07-2016-0031

[16] R. Soltani, U. T. Nguyen, Y. Yang, M. Faghani, A. Yagoub, A. An, (2016). *A New Algorithm for Money Laundering Detection Based on Structural Similarity*.

[17] A. Khalaf, N. El Khamesy, (2016). *Data Mining Techniques for Anti-Money Laundering*. *146*(12), 28–33.

[18] L. Keyan, (2011). *An Improved Support-Vector Network Model for Anti-Money Laundering*. https://doi.org/10.1109/ICMeCG.2011.50

[19] D. Savage, Q. Wang, P. Chou, X. J. Zhang, (2016). *Detection of money laundering groups using supervised learning in networks*. (August).

ISBN: 978-988-14049-2-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)