ABSTRACT
Today's most immediate threat to address is terrorism. Terror organizations use illegal methods to raise their fund, such as scamming banks, fraud, donation, ransom and oil. This illicit money needs be laundered to be used within legal economy through financial institutions (FI). This paper is a complementary to our previous research. And it’s proposes an unsupervised machine learning technique for detecting Money Laundering hidden patterns, groups and transactions in a timely manner to counter terrorism finance. Two different algorithms were implemented and performance was measured, compared and summarized. The preliminary experimental results show the effectiveness of the proposed technique. Domain experts confirm that the proposed method has produced efficient accurate results by identifying and detecting similarities, hidden patterns, grouping across all transactions and all the suspicious accounts involved.

KEYWORDS
Terrorism Financing, Money Laundering, Anti-Money Laundering, Machine Learning

1. INTRODUCTION
Considering the rapid growing threat of terrorism, Counter terrorism financing is becoming the most pressing matter for the globe. Terrorism financing even poor flows of money, can have a major impact by playing a part in serious terrorist crimes. Terrorism financing has become a major political issue, especially after the September 11th terrorist attacks in the United States. In the aftermath of these attacks, Financial Action Task Force (FATF) developed a series of Recommendations (FATF, 2006). Later, such recommendations approved officially as an international standard for counter money laundering, terrorism financing and weapons proliferation, issued in 1990, modified in 1996, 2001, 2003 and 2012 to guarantee that these recommendations are up to date and applicable. Terrorism Financing is often closely linked with money laundering, because criminals and terrorism organizations need to use illegally obtained money without drawing attention of law enforcement officials and deposit this money into financial institutions (FI). This can happen only if the money appears to be obtained through legal sources (to be laundered). So that money laundering acts as a mechanism for aiding terrorism financing. “Simply, Money laundering is the process that disguises illegal profits and integrate it with legal financial systems without exposing criminals who wish to benefit from these profits” (FATF, 2009). The amount of money laundered globally in one year is approximately 2-5 percent of global Gross Domestic Product (GDP) (UNODC, 2011).

Criminals use financial institutions (FI) to exploit weakness in the global financial system to blur the trail of illicit funds. FI are delegated by law to actively monitor report and detect suspicious activities to prohibit money laundering, finance of terrorism and corruption. Using all procedures, laws, policies, regulations, pieces of legislation, programs known as “counter financing of terrorism (CFT)” or “Anti money laundering (AML)”.

Money laundering has a three dynamic stage process that requires Placement, Layering and Integration. The placement stage where criminal or lauderer moves black money from its illegal sources into financial institution both locally and abroad. The layering stage main purpose is to make it difficult to detect or trace
all chaining transactions of black money by making different financial transfers like wire transfers between different accounts (one to many or many to many) various bank to bank transfers and changing the money's currency. The Integration stage for criminals and launderers is the easiest stage as money had entered legitimate financial institutions as if money has come from a legal transaction and it’s impossible to detect a launderer with no documentation from the previous stages (Salehi et al., 2017). Detecting suspicious activities is becoming more and more difficult and tricky since the number of suspicious activities is increasing extraordinarily and the rules based systems alone is not enough to monitor those suspicious activities.

More recently, researchers have applied various artificial intelligence and machine learning techniques such as neural networks, genetic algorithms, fuzzy logic, clustering, sequence matching, Bayesian network, support vector machines (SVM), and outlier detection, in order to better detect and prevent suspicious transactions, activities and hidden patterns.

This paper aims at proposing an unsupervised machine learning technique for detecting Money Laundering hidden patterns, groups and transactions in a timely manner to counter terrorism finance. Paper is organized as follows: a literature review that deals with related work of different machine learning approaches and techniques used for money laundering detection, the methodology that describes the proposed model of money laundering detection using unsupervised machine learning approach in detail, and the evaluation and analysis that contains implementation detailed steps, environment and experimental results. The paper ends with a conclusion and future directions.

2. LITERATURE REVIEW

Detecting suspicious activities related to terrorism financing and money laundering is substantial and inevitable since it affects economy, industries, financial institutions and the entire nation. Because of large amount of banking data and enormous number of transactions involved, banks create a convenient environment for money launderers to hide the origins of money laundering and terrorism financing, so that money laundering techniques became more sophisticated and hard to trace. Hence, the solution for detecting money laundering must be balanced between accuracy and the process time. Finding a suitable technique for money laundering detection in financial institutions / banks is the most significant step for the solution, applying a suitable machine learning approach with a perspective of dataset. Money laundering detection approaches and techniques are varied and this makes comparisons difficult. “However, analyzing, reviewing and comparing various methods of money laundering detection is essential to detect money laundering crimes, patterns, unusual behavior and money laundering groups‖ (Labib et al, 2020). In this part, different machine learning approaches, methods and techniques will be introduced. They specifically do not belong to a particular approach of machine learning technique and are a combination of several methods used in pervious works.

Late 1990s statistical methods and sequence matching were implemented for money laundering detection (Phua et al, 2010), and then rule based methods were used. These methods depend on rules and reputation lists that can be implemented without AI expertise. However, it has to be permanently updated and it detects the most naive launderers only. Later on, financial institutions began to adopt statistical and machine learning models in their automated AML/CFT programs. Those models are highly challenging, due to the rising number of customer transactions and automated interaction with customers and this made anti money laundering compliance become more difficult. Some recent researches referenced that the large dataset size problem can be handled by applying supervised learning techniques which requires a labeled dataset to learn from, and then model classify new data into different label categories. Therefore, supervised learning models can only detect suspicious activities, transactions and patterns which are similar to training data. Nonetheless, many sophisticated modern day money launderers are still able to get around these Supervised Machine Learning models and fabricate new money laundering techniques, literally every day. (Wang et al, 2007) Validated the use of decision tree to create the detection rules of money laundering risk by customer’s profiles of Chinese bank. Based on the rules generated, 12% of 160 thousand current customer profiles are considered high AML risk and need to be further monitored with regard to their future transactions. The study results specified the effectiveness of decision tree in generating anti money laundering rules from company’s customer profiles and determining money laundering risks of bank customers. (Lv, L et al, 2008) Provided a new method to detect suspicious transactions using neural networks composed of three layers.
This method can adapt to the changing means of money laundering and used one million transaction records. The RBF neural network model is compared against SVM and outlier detection. The RBF neural network result was promising with a high detection rate and low false positives. (Keyan et al, 2011) Proposed the cross validation method to optimize selection of parameters of the support vector machine (SVM) model so as to improve the identification of suspicious financial transactions. Using 1.2 million transactions records, the results show that training SVM by employing the selection parameters by cross validation method has more effect than employing the randomly selected parameters on enhancing performance and improving detection of suspicious transactions. (Michalak et al, 2011) Presented a graph mining method to detect suspicious transactions and identify sub graphs within a network that closely matches known typologies. Their method is parameterized using fuzzy numbers which represent parameters of transactions and of the transaction sub graphs to be detected with a synthetically generated data set by simulating a mini economy. The study provided greater flexibility than simple studies using fuzzy matching where the sub graphs might deviate from given typology. (Khan et al, 2013) Used a new Bayesian network (BN) based technique to analyze transactions in a financial institution for suspicious behavior detection. Testing was on more than 8.2 million real financial transaction and more than hundred thousand accounts. Research indicate that new transactions will be considered suspicious activities if the activity has considerable variation to the pattern, but this might result a high false positive rate. (Savage et al, 2016) Described money laundering detection system that analyses group behavior, using network analysis and supervised machine learning. Using (AUSTRAUC) unpublished data, System indicates that suspicious pattern and activity is detected with a suitable level of accuracy, high precision and a low false positives. Furthermore, other studies suggest applying semi supervised techniques which can learn from both labeled and unlabeled datasets; new data will be counted invalid if their behavior doesn’t match with the training set. (Le Khac et al, 2010) Implemented a knowledge based approach that combines data mining and neural networks in bank to detect suspicious transactions or activities of money laundering using K-means clustering, neural network and genetic algorithm for 10 million real transactions records from CE-Bank. The results show that money laundering cases can successfully be detected within the investment activities, and, thus, improves learning process and the performance of current CE bank’s solution in terms of running. (Liu et al, 2011) Presented a core decision tree algorithm to identify money laundering activities, apply a combination of decision trees with BRICH and K-means clustering algorithms in order to detect money laundering using unpublished data source. The study has identified suspicious transactions more effectively but it is too dependent on the strong performance of IBM mainframe and did not gain popularity due to the high cost. (Drezewski et al, 2015) Examined an approach applying the social network analysis for detecting money laundering. The system constructs social networks from the bank statements and National Court Register. The assignment of the roles to persons from the network shows that there are true leaders and vulnerabilities inside the network. The dataset is taken from the X banks disk files and web pages. The study showed that the SNA is a reliable and effective technique for money laundering detection. It also made it clear that finding algorithm is always faster than the interconnection analysis.(Shaikh et al, 2018) Proposed a model for identifying association and relationships within accounts and transactions, using Social Networks Analysis (SNA). For groups and gang mafia, those who play a vibrant role in money laundering activities can be identified by using (AUSTRAUC) unpublished data. The study has successfully identified specific relations among illegal transactions and suspicious customers, such as businesses, parents, spouses, friends, and siblings.

Moreover, other recent researchers pointed out to apply unsupervised machine learning techniques that try to separate data points into different groups without having labeled datasets, these groups known as clusters, which hold data point’s unique characteristics / patterns and are defined by similar member instances. Since financing systems, technologies and operations may vary from time to time, there is a need for new and up to date anti-money laundering systems. Unsupervised machine learning approaches, techniques and models that are more effective and able to detect new money laundering patterns earlier and identify all of the accounts associated with these money laundering groups without knowing what suspicious activity looks like. (Wang et al, 2009) Proposed a new algorithm for money laundering detection using improved minimum spanning tree clustering. The experiment on real 65,001 financial records from X bank shows that the new algorithm has not been affected by noise data, detection of suspicious money laundering transactions; it is complete and effective and has decreased dependency on the parameters. (Cao et al, 2012) Presented an implemented system for money laundering detection in Vietnam’s banking industry using CLOPE algorithm for 8,020 transactions records from Bank X. The results of the study have proved that CLOPE is a proper algorithm to detect money laundering based on the ability of analysts to provide a
criterion for validating clusters after clustering. (Chen et al, 2014) Examined the use of expectation maximization for detecting suspicious financial transaction with 30 million transactions from local X bank in Malaysia. The results show that Expectation Maximization (EM) defeated traditional clustering method k-means for AML in detecting true suspicious transactions with low false positive rate, giving the advantage to EM to be employed in this field. (Paula et al, 2017) Presented results from implementing unsupervised deep learning model for detecting fraud suspects in Brazilian exports, using Auto-Encoder to detect anomalous transactions with regards to the data pattern. The datasets used in this work are 819,990 records from foreign trade of the Secretariat of Federal Revenue of Brazil. The experimental results reduced dimensionalities about twenty times faster, using Deep Learning Auto-Encoder in contrast with PCA proved to be successful detecting suspicious accounts of exporters. The model was able to detect anomalies in at least 20 exporters. (Camino et al, 2017) Presented the use of unsupervised learning methods without the use of ground truth to analyze financial transactions from different sources. The dataset used is from a publicly available data source (Ripple transaction data) and X private company. User ranking results have led to suspicious findings, detected expected anomalies and cases that weren’t known before, confirmed that anomaly detection of user behavior is a must in both traditional and modern financing systems.

3. METHODOLOGY

This section describes the proposed model of money laundering detection using unsupervised machine learning approach in detail. The Model starts by reviewing the input dataset which contains banking transactions dataset wherein the researcher considers only cash debit and credit transactions, the pre-processing of the dataset, and the feature extraction, and then searches for anomaly transactions. Anomaly transactions are financial transactions that have common attributes such as similarity in deposit and withdrawal amounts, structuring and time series (transactions have been conducted in the same time window).

Next, the model then applies an unsupervised machine learning algorithm method to find anomaly transactions and hidden patterns as frequent accounts should not be the only criteria for ending out the suspicious transaction as the case may be when the transaction does not occur frequently but even then they are illegal. With the chaining of accounts, suspicious account which identities the sure relation between these identified model will return then a list containing potential groups of money laundering accounts, Finally, the extracted list of (accounts, groups, and transactions) is rearranged, sorted and visualized to be validated by domain expertise as the output of the framework.
4. EVALUATION AND ANALYSIS

The implementation environment is Windows 10 Pro with 6 GB RAM Intel(R) Core(TM) i7- @ 3777 GHz. Implementation is carried out based on NumPy and mostly on its arrays. The Pandas package builds on the NumPy array object and contains a large number of mathematical functions that can operate on them. And Matplotlib library is a widely used package for Python that allows plotting. The sklearn ensemble includes an implementation of Isolation forest and One Class SVM that was used for this project.

4.1 Dataset and Feature Engineering

One of the difficulties with money laundering domain is the lack of datasets. There is a significant challenge to get realistic data, as real datasets are generally private and hard to get since transaction data and customer profiles are privacy sensitive. In experimentation, the researcher used 118250 transactions with three fraud patterns (Fan-in, Fan-out and Cycle) from a tailored synthetic dataset for an AML problem from (AMLSim) (Weber et al, 2018). As privacy of transaction data and customer profiles is not impacted, conducting and injecting different suspcious activity are controlled and defined by domain expertise. Transactions behave like normal bank account transfer money to other accounts and a small fraction of accounts make suspiscious transactions modeled on real world patterns.

Feature engineering was mainly done by domain expertise, not analytically, only four relevant features numerical attributes remained, the input to the algorithm is a set of financial transactions; each transaction has the following four attributes:

- **Source Node Id**: The unique ID of the account sending the funds.
- **Target Node Id**: The unique ID of the account receiving the funds.
- **Amount**: The amount of funds being transferred.
- **Time**: The date and time of when the transfer takes place.

The dataset is comprised of three files containing three different money laundering patterns (Fan-in-Cycle-Fan-out) and normal transactions in CSV (comma separated values) format.

4.2 Experimental Results

The experimental results from implementing two unsupervised outlier detection Isolation Forest and One Class Support Victor machine for the synthetic AML data are generated from (AMLSim). The results and performance of the two algorithms can be seen in the following figures. Figure (2) shows the anomaly score of both algorithms where histogram x axis represents the anomaly score of the dataset and y axis represents the frequency of the transactions. Figure (3) determines the number of fraud and normal transactions in the entire dataset using both algorithms. Figure (4) represents consumed time for training of the two algorithms. Figure (4) represents consumed time for prediction of the two algorithms. Then Domain expertise results and evaluation, detailed results in Table (1). Hence, both algorithms were implemented on Synthetic AML data generated from (AMLSim).

![Isolation Forest Anomaly Score](image1.png)

![One Class SVM Anomaly Score](image2.png)

Figure 2. Shows the anomaly score of both algorithms
Figure 2 shows the transactions class distribution of the two algorithms which is to help know the number of fraud and normal transactions. Isolation Forest predicted 106425 transactions as normal transactions and 11825 transactions as suspicious (Fraud) transactions. On the other hand, One Class SVM prediction result was 107028 transactions as normal transaction and 11222 as suspicious (Fraud) transaction.

Isolation Forest performed much better than One Class SVM, as Isolation Forest detected 11825 anomalies versus One Class SVM detecting 11222 anomalies. Hence, the overall Isolation Forest method performed much better in determining the fraud cases.

Figure 3. N.O. fraud and normal transactions in the dataset

As discussed above, the One Class SVM algorithm performed badly in the consumed time for training / prediction and memory usage. As a comparison, Figure 4 and Figure 5 are the results of training time parameter where Isolation Forest consumed 7.0073 seconds. On the other hand, One Class SVM consumed 4836.9632. As for the prediction time parameter, Isolation Forest consumed 5.1798 seconds while One Class SVM consumed 911.1394.
Figure 5. Prediction time for both algorithms

About One Class SVM, it runs fine for somewhere in the 10s of thousands, but 100s of thousands take very long. However, financial institutions have tons of data and do not want to wait for days, and maybe even more for nothing to come of it, and miss the chance of preventing fraud transactions before they take place. Nonetheless, Isolation Forest showed a promising performance, low time complexity and low memory usage, and, thus, it is clear that the Isolation Forest was much better than the One Class SVM algorithm at time complexity and memory usage, as predicted by the theory.

Evaluation was done by domain experts as a sample test of 5000 records and the results of suspicious transactions were analyzed by anti-money laundering analysts and officers. It is important to highlight the investigation results of Isolation Forest implementation and the following aspects are about the analysed cases as a sample test: All cases classified by the system are considered high risk, whereas 50 cases were clearly identified as suspects with a 414 suspicious transaction and 4 Money Laundering groups.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of matched transactions</td>
<td>414</td>
</tr>
<tr>
<td>Number of Money Laundering nodes identified</td>
<td>50</td>
</tr>
<tr>
<td>Number of Money Laundering groups identified</td>
<td>4</td>
</tr>
</tbody>
</table>

5. SYNTHESIS AND DISCUSSION

The method of using Isolation Forests for the anomaly detection in money laundering / terrorism financing detection prevention field is still relatively new. It is no secret that detecting suspicious transactions, patterns and groups has become more challenging as money launderers have become more sophisticated. Advanced outlier detection methods, such as Isolation Forests, are imperative for financial institutions detecting and preventing money laundering / terrorism financing because isolation forest (Liu et al, 2008) is an ensemble regressor and uses concept of isolation to “explain / separate away anomalies, no profiling of normal instances and no point based distance calculation. Instead, isolation forests builds an ensemble of random trees for a given dataset and anomalies are points with the shortest average path length”, mainly different from all existing techniques, such as One Class SVM which is different from (Support Vector Machine) a boundary is created around normal data points to separate them from the anomalous ones. It can find nonlinear boundaries with the help of kernel functions, but its accuracy is degraded if too many anomalies are present in the training dataset. However, Isolation Forests exploit subsampling to achieve a small memory requirement, low linear time complexity and to deal with swamping and masking, building it’s model using small subsamples of fixed size regardless of dataset size and detects anomalies based on isolation without using density measure. This improves detection rates and reacts in a timely manner to new money
laundering / terrorism financing patterns. On the other hand, One Class SVM classifier training time increases exponentially with the size of training data and also prediction time. However, no one has figured out the problem yet. The researcher considers time parameter an issue as it eliminates the chance of detecting and preventing the suspicious transaction before it is completed.

Interpreting the results from an unsupervised learning scheme is a very difficult task; interpreting and evaluating the results from detected suspicious/money laundering transactions have been difficult in this research as well. Isolation Forest showed a great, promising result and should be a better algorithm for the current problem than One Class SVM. As discussed above, both algorithms were good at handling outliers, but Isolation Forest results are easier to interpret. Furthermore, for the consumed training/prediction time and memory usage, there is a huge difference from One Class SVM, achieving low linear time complexity and a small memory requirement, dealing with the effects of swamping and masking effectively, higher transparency and interpretability of the results.

6. CONCLUSION

This paper motivated the challenge of detecting money laundering which is a crucial step and mechanism for countering terrorism financing, encouraged technical research community to help financial institutions and law enforcement harness artificial intelligence technologies to counter terrorism financing. Normal financial activities / transactions behavior is chaotic, but money laundering / terrorism financing activities make patterns, whether criminals recognize it or not. From a theoretical point of view, It is confirmed that unsupervised machine learning approach is a more suitable machine learning approach, than supervised and semi supervised approaches for the current problem due to its ability to detect similarities, hidden patterns, structures or grouping across all transactions without prior training, it detects suspicious activities without knowing what suspicious activity behavior or pattern looks like and this is a huge differentiator from and rule based models, supervised and semi supervised which requires knowledge of suspicious patterns / activities to detect similar ones.

Researcher reviewed current AML methods and provided implementation to two unsupervised approaches (Isolation Forest – One Class SVM) for transaction monitoring using a tailored synthetic dataset for an AML problem (AMLSim).

Isolation Forest was much better than One Class SVM algorithm at time complexity and memory usage. Therefore, Isolation Forest algorithm seems more useful, effective and can detect suspicious transactions in a timely manner before transaction is completed or money gets to criminals’ hands in the context of anti-money laundering / counter terrorism financing.

Domain experts confirmed that Isolation Forest has produced higher efficient results and accurate findings, and improved the efficiency of the existing anti money laundering techniques by identifying and detecting similarities, hidden patterns, structures or grouping across all transactions and all chaining suspicious accounts involved.

7. FUTURE DIRECTION

In near future, to better assess the effectiveness and evaluate performance of the proposed model, researcher aims to install the proposed model or conduct experiments on real financial data and real life scenarios from financial institutions which would help in fine tuning the proposed approach. The researcher is planning to focus on the detection of unexpected changes in user’s behavior and will extend the analysis methods in order to automatically discover the money laundering patterns in real time.
REFERENCES


