 doi:10.20944/preprints202106.0172.v1

Article

Blockchain-Enabled Transaction Scanning Method for Money Laundering Detection

Ammar Oad [1], Abdul Razaque [2,*], Askar Tolemyssov [2], Munif Alotaibi [3], Bandar Alotaibi [4,5], Zhao Chenglin [1]

[1] Faculty of Information Engineering, Shaoyang University, Shaoyang 422000, China; ammar.oad@hnsyu.edu.cn (A.O.); chenglinzhao@126.com (Z.C.)
[2] Department of Computer Engineering & Information Security, IITU, Almaty 050000, Kazakhstan; a.razaque@edu.iitu.kz; a.tolemysov@gmail.com
[3] Department of Computer Science, Sharqa University, Sharqa, Saudi Arabia; munif@su.edu.sa
[4] Sensor Networks and Cellular Systems Research Center, University of Tabuk, Tabuk 71491 , Saudi Arabia; b-alotaibi@ut.edu.sa
[5] Department of Information Technology, University of Tabuk, Tabuk 71491, Saudi Arabia
* Correspondence: a.razaque@edu.iitu.kz

Abstract: Currently, life cannot be imagined without the use of bank cards for purchases or money transfers; however, their use provides new opportunities for money launders and terrorist organizations. This paper proposes a Blockchain-enabled transaction scanning (BTS) method for the detection of anomalous actions. The BTS method specifies the rules for outlier detection and rapid movements of funds, which restrict anomalous actions in transactions. The specified rules determine the specific patterns of malicious activities in the transactions. Furthermore, the rules of the BTS method scan the transaction history and provide a list of entities that receive money suspiciously. Finally, the Blockchain-enabled process is used to restrict the money laundering. To validate the performance of the proposed BTS, the Spring Boot application is built based on the Java programming language. Based on experimental results, the proposed BTS method automates the process of investigating transactions and restricts the money laundering incident.

Keywords: transactions; money transfers; money laundering; originator; beneficiary

Currently, monetary transactions can be committed easily by using bank transactions. Shopping, money transfers, and ordering services are all services that are available to clients anywhere in the world [1,2]. However, these transactions open many opportunities for third parties to commit illegal activities with money without punishment [3]. One of these activities is called money laundering. Money laundering is a process whose purpose is to hide illegal sources of profit [4,5] . The need for money laundering arises in three cases. The first case is if the origin of income is criminal, e.g., illegal drug trafficking, racketeering, or corruption. Criminals receiving such income are forced to launder the money to be able to spend it freely [6,7]. The second case is if an entrepreneur or firm hides a portion of their legal income from increased taxation by underestimating revenue or overcharging, using unaccounted cash, etc. [8,9]. The third case is if the recipients of the money do not want to show their real source for security, ethics, or political reasons [10].

Anti-money laundering (AML) includes a number of measures intended to counter the legalization of proceeds of crime and curb financial flows intended for terrorist activities [11,12]. Most people have participated in these activities. They use many rules to identify suspicious activities in transaction histories [13,14]. During the process of reading and analyzing a significant amount of data, the human factor has an effect. Therefore, the contribution of this paper is the automation of the process by creating a Java-based application, where the rules of recognition of suspicious activity are implemented with a BTS method. AML teams suffer from false alerts that result in significant additional costs [15]. As the human resources of investigators are spent checking many false warnings, real money laundering actions continue to happen.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

There are three fundamental problems that cause this to occur. First, they use detailed rules for each scenario, resulting in many warnings that are not actually suspicious. Second, they only check a fraction of the available data, which limits the number of signals they can use to detect money laundering. Third, they have strict data format requirements that require a painful process of data integration, which often leads to poor data quality [16,17].

The application solves these problems as follows: First, it uses only the data that the user provides. Second, it checks the user's data in full, meticulously checking each value. Third, it allows the user to initially assign keys that will be exposed during the suspicious activity search process.

The contributions of this paper are summarized as follows:

- The novel rules have been formulated for the anomalous transaction detection that supports against the money laundering
- Blockchain transaction scanning method is employed that involves the rich features of data mining and Blockchain for deciding the confirmed malicious and legitimate transaction.

For defining money laundering applications, rules such as "outlier detection" [18] or "rapid mvmt funds" are used to check transactions by comparing actions with a template. If the sequence of actions matches the template, a case can be created for the beneficiary. While other applications use difficult methods such as machine learning and hardcoding the rules into the architecture, our application will use simple methods to easily add other rules and scale the system in the future.

The remainder of the paper is organized as follows: Section 1 identifies the problem. Section 2 briefly analyzes the state-of-the-art related work. Section 3 describes the system model. Section 4 defines the proposed BTS. Section 6 provides experimental results. Section 7 presents a conclusion of the entire paper.

1. Problem Identification

Money laundering is considered a criminal process that allows illegally earned money to flow into the basic cash flow of society. Given that financial products and services around the world fall under anti-money laundering regulations, the international community believes that money laundering is a threat to the world economy. As a result of these activities, "dirty" money becomes "clean." The point of this activity is that the origin of the illegally obtained money becomes impossible to determine, and criminals can spend it with impunity.

From a social point of view, the largest problem of money laundering is that it finances and creates favorable conditions for organized crime. Often, dirty money initially appears because of drug trafficking, tax evasion, the sale of illicit goods or trafficking, and support for terrorist acts. According to the calculations of The United Nations Office on Drugs and Crime, approximately two to five percent of global GDP (between $800 million and $2 billion) has been integrated into the global banking system through money laundering. With that in mind, it is a global problem.

This problem is so urgent that to counter it in Europe, EU law requires companies to hire financial services to conduct checks on their clients' AML to prevent this practice. Anti-money laundering measures include verifying the identity of each client by a financial service or agency and monitoring his or her operations. As part of the fight against money laundering, the financial institution may also request additional information from the client if it discovers any suspicious activity. The financial institution can ask the client that is depositing a large sum of money into his or her account to provide documents confirming the origin of the funds.

AML organizations try to define money laundering by using rules. Rules are a template of a sequence of actions that may be defined as a money laundering process. If during an investigation a specific beneficiary receives many alerts based on the rule matching, a case can be created for this person. The only difficulty of this entire process is that it requires a significant amount of human resources and time. Customer verification,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

data analysis, and mathematical calculation of revenue/waste are all the responsibility of specialists. To make the overall process easier, the most optimistic option is to automate some of these processes.

2. Related Works

The main characteristics of existing approaches are discussed in this section. A prototype application AML2ink provides a visualization of relations between accounts in transactions to further identify suspicious activity [19]. An SQL query is responsible for data processing, while GraphViz renders and produces the visualization. By using this output, the process of investigation is simplified. However, it is impossible to determine the maximum amount of data GraphViz can render because it produces only one image file, where all entities are presented as nodes and relations as links; light data with 100 rows can require significant time to be rendered.

Kolhatkar [20] fully introduced the process of a multichannel data-driven, real-time AML system, providing detailed schemas. In this approach, some methods and algorithms for defining money laundering are described. However, the paper provides little information about automatization of the entire process and no information concerning the technical aspects. Raza and Haider [21] proposed the SARDBN tool, which identifies abnormalities in the sequence of transactions. As a basic algorithm, a dynamic Bayesian network is used, which generates output by filtering transactions with outlier detection rules. Although the data of the results are diverse, they provide no useful information.

Weber [22] proposed deep learning, which works with a massive amount of graph data. Then, he used graph learning to display available information to the users. Although the application provides visual data, it is impossible to manipulate the database structure. Luo [23] introduced a framework with a data mining system for detecting suspicious transactions. In this paper, specific rules such as attribute filtering and a correlation matrix between trade accounts are provided. Although the data of the results are diverse, they are useful.

Colladon and Elisa [24] proposed a social network analysis to determine money laundering. Based on network metrics, this paper presents predictive models showing the risk profiles of clients. However, this model cannot provide complete information because not everyone uses social networks. After examining these approaches, we decided that many of the proposed approaches are essential. All these approaches are aimed at improving data collection, identifying money laundering, and visualizing the outcome. Our approach will not be inferior to them; additionally, it will combine all of these advantages.

3. System Model

This section addresses the main values in transaction data for further use in rules, which are originator, beneficiary, transaction committed date, and amount of money. These four modules can be defined as follows:

- Originator
- Beneficiary
- Transaction committed date
- Amount of money

An example of bank transaction data with the columns that include these modules is shown in Figure 1.

As shown in Figure 2, the transaction committed date and amount of money are marked in blue and green, respectively. The originator and beneficiary, as mentioned before, consist of multiple lines.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

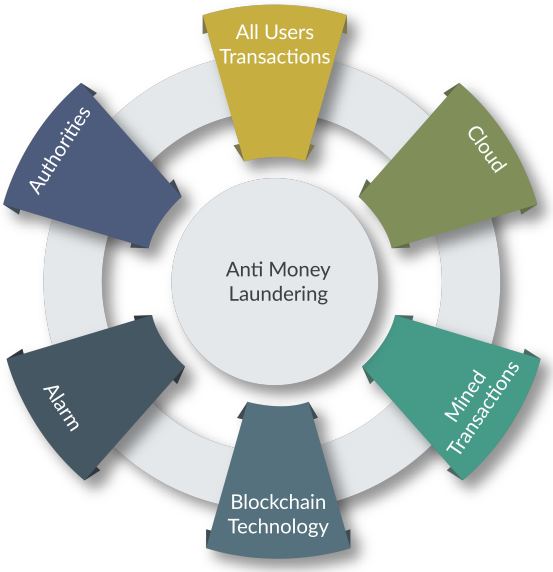
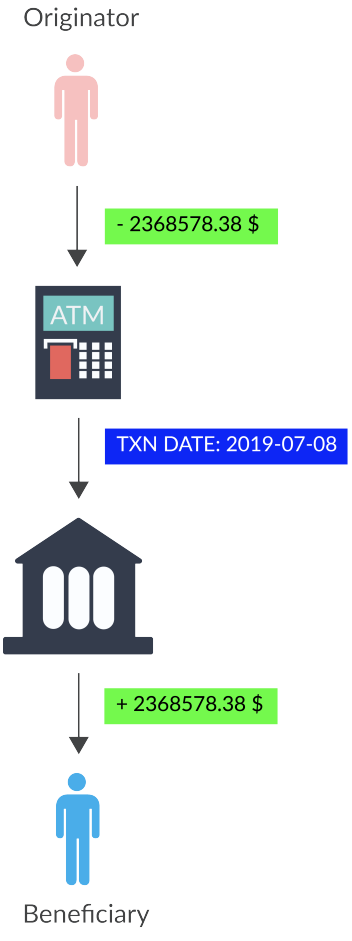


Figure 1. Simple Anti money laundering process with the Data mining and Blockchain-enabled features.



Transaction
id_o: 1
id_b: 2
txn_key: 310202140002882
orig_id_type: NULL
orig_line1: "UNIDO BANCO S.A NEW YORK BRANCH"
orig_line4: NULL
orig_line3: "NEW YORK, NY 10153"
orig_line2: "707 FIFTH AVENUE"
orig_id: UNDOUS33XXX
orig_bank_id_type: "ACCOUNT"
orig_bank_id: "8038821045"
orig_bank_line1: "FEDERAL RESERVE BANK"
orig_bank_line2: "33 LIBERTY STREET"
orig_bank_line3: "NEW YOURK, N.Y. 10045"
orig_bank_line4: NULL
TXN DATE: "2019-07-08"
amount: 2368578.38
int_bank_id: NULL
int_bank_line1: NULL
int_bank_line2: NULL
int_bank_line3: NULL
int_bank_line4: NULL
bene_line1: "UNIDO BANCO S.A-GRAND CAYMAN"
bene_line2: "UNDOKYKCIGC"
bene_line3: NULL
bene_bank_id: NULL
bene_bank_id_type: NULL
bene_bank_line1: NULL
bene_bank_line2: NULL
bene_bank_line3: NULL
bene_bank_line4: NULL
bene_id_type: NULL
bene_id: "599955785"
obi: NULL
bbi: NULL
dr_cr: "/BNF/YANKEE CD SETTLEMENT OF PYMNT"
dr_cr: "D"
id: 1

Figure 2. Structure of a transaction table.

3.1. Originator

The account that is the source of the transaction flow. This can be described differently in transaction history depending on the bank. For the record of the originator, the bank can

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

use the address of the ATM from which the transaction was performed. Often, for better integration into the bank system, the address of the ATM can be divided into multiple columns. For example, the name of the country, the city, and the address can be inserted into four separate columns.

3.2. Beneficiary

The account that is the destination of the transaction flow. Its structure is the same as the originator's. It also has columns by which it is possible to identify the location where the person received the money. It is not possible to say if the beneficiary can be the main suspect in money laundering just because he receives all the money. The beneficiary can also be the originator, which means that he or she acts in the long sequence of the money laundering process.

3.3. Transaction Committed Date

The date the transaction flow occurred. The outlier detection rule identifies transaction anomalies by checking how much money the originator spends in one month, by week. The rapid mvmt funds rule uses a two-week range to check whether the originator sends or the beneficiary receives a specific percent of the money. Therefore, it is important to identify when the transaction was committed to perform an investigation.

3.4. Amount of Money

The amount of money that is used in the transaction flow. This information alone is useless. Even if the amount is very low or very high, it does not define any suspicious behavior because income is an individual element. Only by using the combination of the above elements is it possible to perform an investigation and identify anomalies in account transactions. In the money laundering process, generally, a small amount of money is used to avoid additional attention.

4. Problem Formulation Rules for Transaction Anomaly Detection

The transaction-scanning algorithm uses two rules to define anomalies in transaction data tables. The following rules are described below:

- Outlier detection
- Rapid mvmt funds

4.1. Outlier Detection

An outlier is a data point that is significantly different from the others. Conversely, inliers are data that are within a stable distribution. It is not easy to define outliers because they are highly diverse and unpredictable; however, inliers are often stable, which can help define outliers.

Outlier detection in our case works with a monthly income scenario. Most people receive a wage at the beginning of the month. Then, every week until the end of the month, the person spends a fixed amount of money. This information can be considered an inlier and can be written as follows:

$$T \leftarrow \left\{x'_{i'}\right\}_{i'}^{n'} = 1 \tag{1}$$

where T is the data set of transactions with fixed expenses in one month; x is the data set of transactions; i is the sequence number of the transactions; and n is the number of transactions within a one-month period.

Therefore, the main data that can be useful to the rule is the expense of the last week, which can be recorded as follows:

$$T_l \leftarrow \left\{x_i\right\}_{i}^{n} = 1 \tag{2}$$

where T_l is the data set of transactions with anomaly data in last week of the month.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

Data can be defined as anomalous only if the expense of the last week is extremely high compared to the expenses of the other weeks of the month. This type of activity can be considered suspicious because most people try to save money in the last week of the month to sustain them until the next wage payment. This outlier can be defined by using the density ratio given by

$$W(\gamma) = \frac{p(T)}{p(T_l)} \tag{3}$$

where $p'(T)$ is the inlier density and $p(T_l)$ is the test sample density. The density ratio is close to 1 when x is an inlier and it is close to 0 when γ is an outlier.

Theorem 1. *The highest expenses occur at the beginning of the month.*

Proof. As a rule, at the beginning of the month, people receive their wages. Before spending this money, first, they identify their necessary costs. These required costs can be divided into the following three categories: □

4.1.1. Investment or Savings

A certain percentage is immediately separated from the earned and set aside for a predetermined overall goal. Depending on the level of wealth, this percentage may vary. In most cases, people save 10% of their wages [25].

4.2. Mandatory Payments

After the savings funds are removed, payments that cannot be avoided will follow. First, people repay money borrowed from friends or payments on bank loans. Then, the funds required to pay for housing and communal services are calculated. Finally, the necessary costs for public transport, payment for kindergarten, medicines, gasoline for a car, etc. are deducted.

4.3. Variable Costs

This includes all other family expenses, such as food, shoes and clothing, household expenses, spouses' personal expenses, entertainment, holidays, birthdays, vacations and unexpected expenses.

The formula of the weekly expense can be written as:

$$E_{wn} = E_{inv} + E_{mand} + E_{costs} \tag{4}$$

where n identifies the week number $(1 \leq n \leq 4)$; E_{inv}, E_{mand}, and E_{costs} are expenses for investment, mandatory, and variable costs, respectively. As mentioned before, only in the first week, after receiving wages, do people primarily pay investment and mandatory payments because they are necessary. Thus, for the next three weeks they are not considered, as follows:

$$E_{w_{\{2,3,4\}}} = E_{costs} \tag{5}$$

Corollary 1. *As a result, people spend more money in the beginning of the month because of expenses for needs and budget allocation. The outlier detection process is described in Figure 3.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

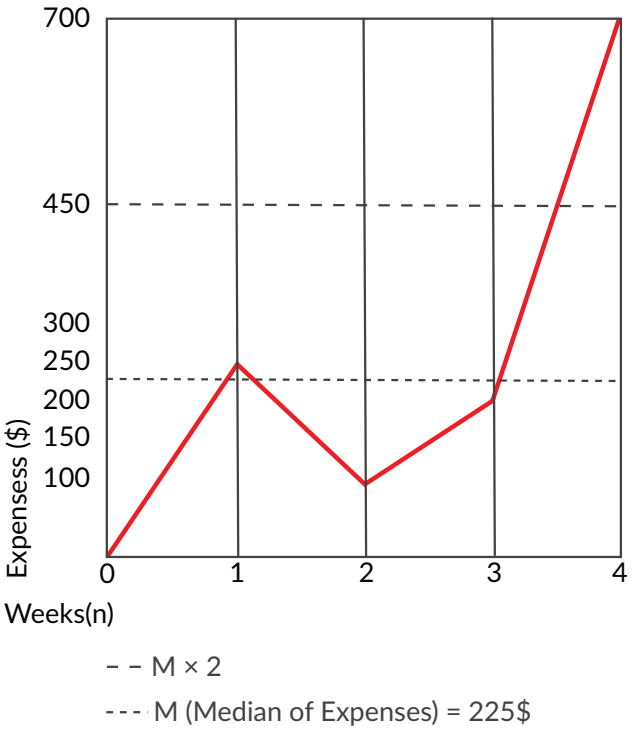


Figure 3. Outlier Detection.

The outlier detection process is explained in Algorithm 1.

Algorithm 1 Outlier Detection

Input: $\{T[], T_l, T_s\}$ in
Output: $\{S\}$ out
1: **if then**
2: **Initialization:** $\{T[]$: *entity's list of weekly transaction expenses;* T_l*: entity's last week transaction expense;* T_s*: size of* T*;* S*: entity's suspicious state;* M*: median of transaction expenses* $\}$
3: **else if** T_s is odd **then**
4: **Set** $MT[\frac{T_s}{2}]$
5: **else**
6: Set $M\left(\frac{T\left[\frac{T_s}{2}\right]+T\left[\frac{T_s}{2}+1\right]}{2}\right)$
7: **end if**
8: **if** $M \times 2 \leq T_l$ **then**
9: **Set** S True
10: **else**
11: **Set** S False
12: **end if**

Algorithm 1 explains the detecting anomaly transaction process by using outlier detection. At the beginning of the algorithm, the input and output are shown, respectively. In step 1, the initialization process of given variables is explained. Steps 2 through 6 check the size of transactions for odd/even and according to their set median transaction expense. Steps 7 through 11 check the median at the expense of the last week. If the expense exceeds the median, then the entity's state is set to suspicious; otherwise, it is not.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

4.4. Rapid Mvmt Funds

Most people spend their money buying something or lending to someone. The rapid mvmt funds rule, for the most part, is related to the second scenario. This rule is based on transferring money from one account to another. To launder money, scammers divide it and send it to multiple accounts. These accounts can also repeat this action. As a result, many accounts will have a portion of the money; then, a reverse process will begin in which accounts will collect the money by sending it to one account. In the end, the last beneficiary will receive the laundered money.

The rapid mvmt funds process uses two weeks of income and outflow of a specific entity. The main rule of this method is that if income is between 80 and 120 percent of outflow during two weeks' transactions, then the entity can be suspected to be an actor of money laundering. This can be written as follows:

$$S = \sum_{l=0}^{n} T_o \tag{6}$$

$$S = \sum_{l=0}^{n} T_b \tag{7}$$

where n is the number of transactions over a two-week period and T_o and T_b are transactions where the entity acted as an originator and beneficiary, respectively.

Total inequality can be written as follows:

$$S \times 0.8 \leq R \leq S \times 1.2 \tag{8}$$

Theorem 2. *Defining money laundering by inspecting remittance transactions is difficult.*

Proof. T he money transfer industry is currently growing rapidly. In 2018, more than $689 billion in money transfer transactions were made. Hence, we can conclude that this is a good platform for money laundering [26]. □

Key risks associated with remittances are the following:

- *Digital services:* Internet money transfer services are not only more difficult for authorities to control, but also allow criminals to bypass identity verification processes.
- *Prepaid cards:* Every bank consumer can use prepaid cards to send and withdraw money through ATMs.
- *Third party engaging:* Money launderers can hire third parties on their behalf to perform transactions. Such third parties are called money mules.
- *Ownership:* Given the spread of money transfer services, to avoid the rules, money launderers can obtain ownership of a money transfer company.
- *Structuring:* Different accounts may be used by money launderers to participate in multiple money transfer transactions.

The probability P_{md} definition of money laundering can be calculated as:

$$P_{md} = \frac{N_t}{N_a} \times 100\% \tag{9}$$

where N_t is the number of transactions and N_a is the number of accounts. The number of accounts can be converted as:

$$N_a = \sum_{j=0}^{m} \{\rho_1, \rho_2, ..., \rho_{n-1}, \rho_n\}_j \tag{10}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

where m is the number of unique entities; ρ are the columns in the transaction table that are defined as originator or beneficiary; and n is the number of columns of ρ. Thus, the number of transactions can be converted as:

$$N_t = \sum_{i=0}^{r} d_n \tag{11}$$

where r is the number of records in the transaction table and d is the column "id" in the table that must be unique for every row. Thus, the entire probability of detecting the money laundering is written as:

$$P_{md} = \frac{\sum_{i=0}^{r} d_n}{\sum_{j=0}^{m} \{\rho_1, \rho_2, ..., \rho_{n-1}, \rho_n\}_j} \times 100\% \tag{12}$$

As observed, N_a is inversely proportional to probability, which means that a large number of accounts cause a small probability of defining money laundering.

Corollary 2. *Because of the large size of the industry and common uses of money transfers in people's lives, it is difficult to define money laundering. The rapid mvmt funds actions are depicted in Figure 4.*

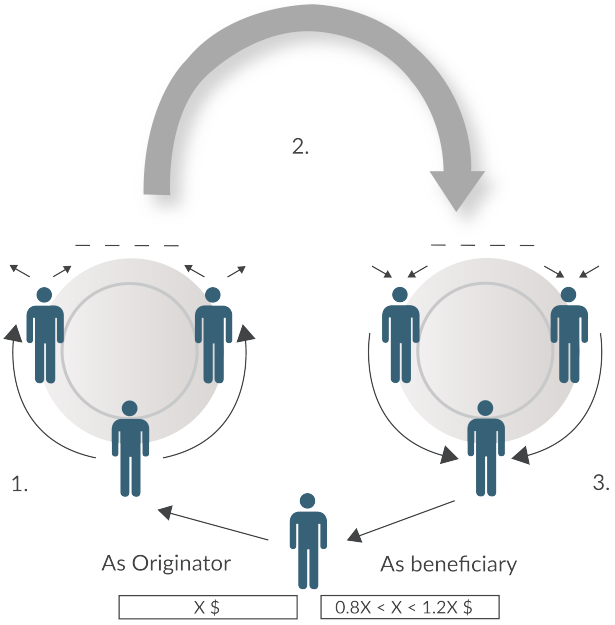


Figure 4. Rapid Mvmt Funds example.

In Algorithm 2, the rapid mvmt funds algorithm for detecting suspicious transaction activity is explained. The input and output are shown at the beginning of the algorithm. Step 1 explains the initialization process of the given variables. Steps 2 through 4 check if the amount of sent money ranges between approximately 80 and 120 percent of the amount of received money; if true, then the entity's state is set to suspicious. Steps 5 and 6 check if the amount of received money ranges between approximately 80 and 120 percent of the amount of sent money; if true, then the entity's state is set to suspicious. Steps 7 and 8 set the entity's suspicious state to false because no anomaly is detected.

Hypothesis 1. *A transaction scanning processes require little processing time because of filtered columns.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

Proof. The abstract useful information that can be used in TS methods can be written as follows:

$$I_e = T_{O_e} + T_{B_e} \tag{13}$$

where T_{O_e} is the entity's transactions where he or she acted as originator and T_{B_e} is the entity's transactions where he or she acted as beneficiary. □

Algorithm 2 Rapid Movements of Funds

Input: $\{T_o, T_b\}$ in
Output: $\{S\}$ out
1: **Initialization:** $\{T_o$: *money amount of certain two-week transactions as an originator;* T_b: *money amount of certain two-week transactions as a beneficiary; S: entity's suspicious state* $\}$
2: **if** $T_o \geq T_b \times 0.8$ & $T_o \geq T_b \times 1.2$ **then**
3: **Set** S True
4: **end if**
5: **if** $T_b \geq T_o \times 0.8$ & $T_b \geq T_o \times 1.2$ **then**
6: **Set** S True
7: **else if** $T_o \leq T_o \times 0.8$ & $T_o \geq T_b \times 1.2$ & & $T_b \leq T_o \times 0.8$ & $T_b \times \geq T_o \times 1.2$ **then**
8: **Set** S False
9: **end if**

In the bank database, there will be a table where all transaction histories are stored. The structure of this table contains many columns because a bank system usually uses an NoSQL structured database, which means that data have no relations; thus, it is not necessary to divide data into a table. The rows of a table T_r can be recorded as:

$$T_r = \sum_{i=0}^{R} (C_r)i \tag{14}$$

where R is the total number of rows, and C_r is the column of the row.

The transaction scanning method does not require many columns C_r. It primarily uses the following process for columns:

$$C_r = \{D, A, O, B\} \tag{15}$$

where D is the transaction committed date; A is the amount of money; O is the account of the originator; and B is the account of the beneficiary.

Therefore, the rows of the table can be rewritten as follows:

$$T_r = \sum_{i=0}^{R} \{D_i, A_i, O_i, B_i\} \tag{16}$$

In the result, removing data about originators and beneficiaries in transactions where entities acted as originators and beneficiaries, respectively, thus above equation can be described as:

$$I_e = \sum_{i=0}^{n} \{D_i, A_i, B_i\} + \sum_{i=0}^{n} \{D_i, A_i, O_i\} \tag{17}$$

Corollary 3. *By reducing the usage of rows of table as*

$$T_r = \sum_{i=0}^{R} (C_r)i \rightarrow T_r = \sum_{i=0}^{R} \{D_i, A_i, O_i, B_i\}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

Thus it is possible to increase the performance of transaction scanning.

5. Proposed BTS Method for Money Laundering Detection

Money laundering postures a severe threat to financial bodies that leads to national impairment. Thus, detecting doubtful transactions concerning money laundering is of paramount significance. Detecting money laundering, the BTS method is employed depicted in Figure 5, consisting of the following steps.

- Content Pre-processing
- Content Mining and Blockchain-enabled Features

5.1. Content Pre-Processing

The leading role of content pre-processing is to excerpt datasets from different locations. Subsequently, it merges different datasets into an integrated database, then extraction, transformation, and loading (ETL) processes are applied as employed in [27]. This process experiences due to content quality issues because the financial institutions possess a different set of quality issues at the content level. Most of the problems are associated with the customer information in our case. These problems include:

- Null or dummy values: It happens in most of the data fields of the databases except the identity of the user, the user type (individual, joint, or company), and fund name.
- Misspelling, usually phonetic and typos errors. Additionally, banking datasets are mostly organized in a distributed fashion to maintain security and flexibility. The heterogeneity of the contents can pose the threat to the content quality particularly when an integrating process is required. Therefore, the basic content quality issues can be addressed using pre-processing.

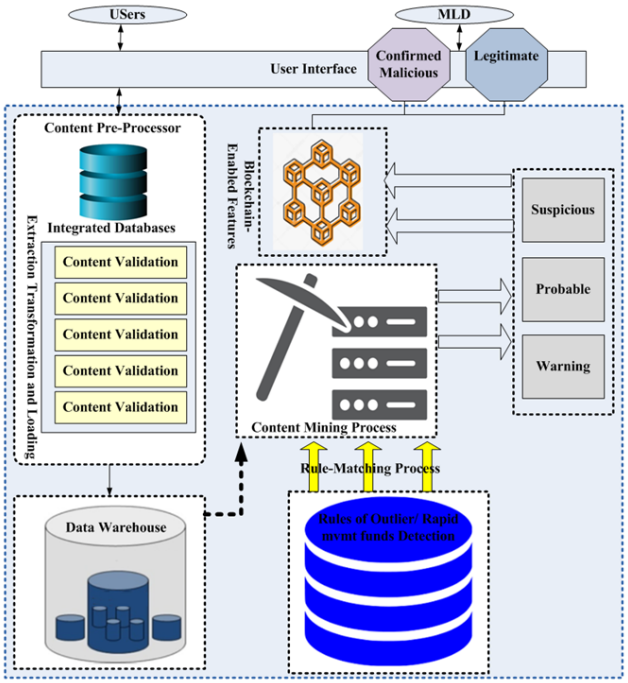


Figure 5. Blockchain-enabled transaction scanning system.

Numerous datasets possess the different variations between the features ranging from minimum to maximum, for example 0.001 and 10,000. If such variations occur, then there is need of the scaling down to make the attribute acceptable and appropriate. This process

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

supports various classifiers to be compatible for content processing. The scaling process of determining the new feature F_n is given by:

$$F_n = \frac{F_o - V_{Min}}{V_{Max} - V_{Min}} \tag{18}$$

where F_o: The original features of the contents, V_Max: Maximum value of the features, and V_Min: Minimum value of the features.

It is highly important to determine all of the features from all datasets. Thus, scaling process of identifying the new features from entire contents are given by:

$$F_n = \frac{F_o - V_{Min}}{V_{Max} - V_{Min}} \{(\beta V_{Max}) - (\beta V_{Min}) + \beta V_{Min}\} \tag{19}$$

where βV_{Max}: New maximum features, and βV_{Min}: New minimum features.

Once, a new feature values are obtained, then there is need of standardization of the feature value is given by:

$$F_n(s) = \frac{(F_o - M_v)}{\sigma}$$

Substitute M_v and σ, we get as:

$$F_n(s) = \frac{F_o - (\frac{1}{n}\sum_{i=1}^{n} V_{ci})}{\sqrt{\frac{\sum\{(a_i)-(\mu)\}^2}{D_{si}}}} \tag{20}$$

where V_{ci}: Data content value, n: Number of values, M_v: Mean value, σ: content standard value, D_{si}: Dataset size; a_i: Each value from the data content, and μ: data content mean value.

5.2. Content Mining and Blockchain-Enabled Features

This process collects and stores the information obtained from anti-money laundering (AML) experts and uses case studies and past money laundering cases. Based on the collected and stored information, the rules for anomaly detection are generated to identify the malicious activities of the outlier and rapid mvmt funds. This process is responsible to control the entire data mining process by employing the rules to obtain better performance. It matches the data obtained from the warehouse with the associated rules. The rules provide three types of classifications (warning, probable and suspicious) for each transaction to money laundering depicted in Figure 5. When the customer makes the transactions, then it goes to the "Rule-matching process" which is part of the content mining process. The Rule-matching process consists of several rules, which are matched against each transaction. Each transaction is initially marked as a "warning transaction" and sent for further investigation. Based on the investigation, if the transaction matches more than 60% of the rules, then the transaction is considered as a "probable transaction" if the transaction matches less than 60% of the given rules, then the transaction is decided as a safe transaction. The transactions which match more than 60% of the given rules are sent for final investigation. If the transactions match with the rules ≥ 95, then it is considered as the "suspicious transaction". Finally, the report of the suspicious transaction is forwarded to the Blockchain-enabled feature server.

The Blockchain-enabled features are stored on the Blockchain-enabled server that is responsible to block the suspicious transaction and releases the legitimate transactions. When the Blockchain-enabled server receives the message from content mining process component as the suspicious, then it declares it as the "confirmed malicious" On the other hand, if the transaction is received as a "non-suspicious", then the Blockchain-enabled server declares it as the "legitimate transaction". Finally, the legitimate transaction is allowed for further process. The Rule-matching process is given in Algorithm 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

Algorithm 3 Rule matching process using content mining

Input: $\{t\}$ in
Output: $\{S_t; P_t; Su_t\}$ out
1: **Initialization:** $\{S_t$: *Safe transaction; t: Transaction;* P_t: *Probable transaction;* Su_t: *Suspicious transaction;* W_t: *Warning transaction;* g': *Rule-matching process* $\}$
2: **Set** g'
3: **Set** $W_t \cong t$
4: **if** $g' \geq 60$ **then**
5: $t = P_t$
6: **else**
7: $t = S_t$
8: **end if**
9: **if** $g' \geq 95$ **then**
10: $t = Su_t$
11: **else**
12: $t = S_t$
13: **end if**

The probability model Pr_m is generated that reports the confirmed malicious state. Let us assume if the transaction t_i is reported as the "confirmed malicious" to the authorities that holds the value 1 otherwise 0. The variables associated with the transaction i are denoted as Ψ_i that can be written as:

$$Pr_m(t_i = 1|\Psi_i)\&f(\Psi_i) \in \{0, 1\}$$

Minimizing the money laundering M_m process is derived by

$$Pr_m\Big(t_i, f(\Psi_i)\Big) = t_i \log\Big(f(\Psi_i)\Big) + (1 - t_i) \log\Big(1 - f(\Psi_i)\Big) \tag{21}$$

The content-mining process component takes the features from the iterative local search and random methods explained in [28] that help to develop the final predicative model FPr_m for the confirmed malicious transaction.

$$FPr_m(t_i) = \frac{1}{5} \geq \sum_{k=1}^{5} \partial\forall(t_i)k \tag{22}$$

where $\partial\forall(t_i)$: the matching-rules which do not match with the suspicious transaction.

From the above equation, we deduce that ≥ 95 matching-rules match with suspicious transaction and remaining rules do not match with transaction. Based on the result of the predictive model, the transaction is either considered as the confirmed malicious or legitimate transaction decided by the Blockchain-enable server.

6. Experimental Results

To validate the quality of the BTS method, the complete model is written using the Spring Boot application based on the Java programming language. To store and retrieve data, the PostgreSQL relational database version 9.6.14 was used. Application Netdata is used for the monitoring system. The laptop configuration on which the application is executed is described in Table 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

Table 1. Machine characteristics.

OS	Ubuntu 20.04.1 LTS
Processor	Intel® Core™ i7-8550U CPU @ 1.80 GHz × 8
Processor architecture	x64
Graphic card	GeForce MX150/PCIe/SSE2
Hard drive	256 GB SSD
RAM	15.5 GB

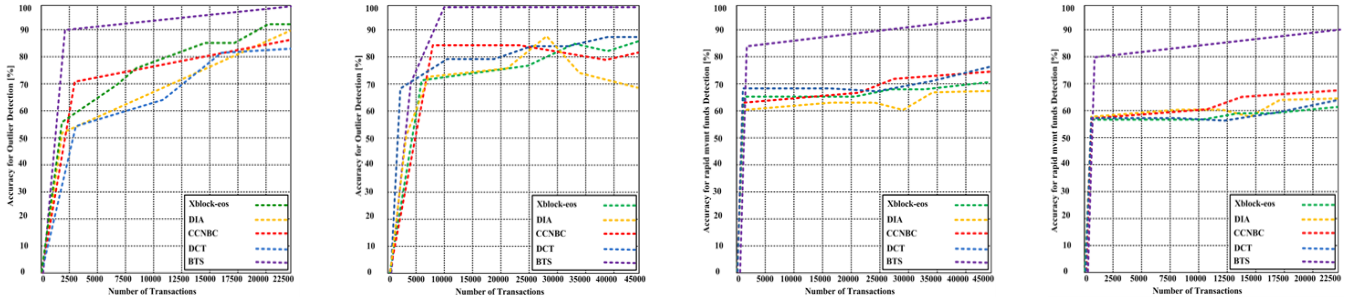
Mock transaction backup data provided by HSBC is used as input data. It contains 21,602 rows and every row has 37 columns. The columns that are defined as originator and beneficiary are "orig_line1", "orig_line2", "orig_line3", and "orig_line4", and "bene_line1", "bene_line2", "bene_line3", and "bene_line4", respectively. The column "amount" contains the amount of money in the transaction and "TXN DATE" describes the date when the transaction was committed. The proposed BTS is compared with the state-of-the-art methods: Extracting and exploring Blockchain data from eosio (Xblock-eos) [29], Convolutional Neural Network Biometric Cryptosystem (CNNBC) [30], detecting cryptocurrency transactions (DCT) [31] and Detection of illicit accounts (DIA) [32] to ensure the validity of the propose algorithm. Based on the testing, interesting results are determined for the following:

- Outlier detection
- Rapid mvmt funds

6.1. Outlier Detection

For the outlier detection method, transactions with role originators and committed dates between March 1, 2021 and April 30, 2021 are used. All these data are retrieved by using the SQL script. Further calculations are performed in the Java language and the results are stored in the database. The outlier detection scanning process takes 138,606 ms, and as a result, 581 rows are generated. The method requires 10% of processor usage and 2 GB of RAM for processing. By using the provided results of the transaction money-laundering state, the accuracy of outlier detection is determined. As shown in Figure 6a, the method provides high accuracy because it uses Algorithms 1–3, which can define only specific actions in transactions; the chances that innocent entities could unintentionally trigger this rule are minimal. Additionally, it should be noted that the accuracy of the method is more stable than contending methods: (Xblock-eos, CNNBC, DCT and DIA) depicted in Figure 6a. While our accuracy plateaus at approximately 98.7%, in the contending methods, the accuracy begins at 60 percent and grows slowly. The contending methods: Xblock-eos, CNNBC, DCT and DIA produce outlier accuracy 83.3–92.2%. When the number of transactions increase that greatly affact the accuracy of contending methods depicted in Figure 6b, while the proposed BTS remain stable and produces 99.4% outlier accuracy. Thus, it is proved even if the number of transactions increase that do not affact the accuracy of the proposed BTS. On the other hand, the contending methods are not appropriate to deal with the number of the increased transactions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1



(**a**) (**b**) (**c**) (**d**)

Figure 6. (**a**) The outlier detection accuracy of the proposed BTS and contending methods with the maximum 22,500 transactions. (**b**) The outlier detection accuracy of the proposed BTS and contending methods with the maximum 45,000 transactions. (**c**) Total accuracy. (**d**) Comparison of total accuracy.

6.2. Rapid Mvmt Funds

For rapid mvmt funds, the same date span as the outlier detection method is used, but SQL scripts are used to retrieve transactions with originator and beneficiary roles. The rapid mvmt funds scanning process took 139,701 ms. It is required slightly longer than the outlier detection process because in this method, entities with originator and beneficiary roles are checked. As a result, 143 rows are generated. This process uses 2 GB of RAM and 20% of processor usage. The accuracy of rapid mvmt funds is shown in Figure 6c,d. Based on the result, it is observed that the proposed BTS and contending methods produce a lower accuracy than outlier detection. This occurs because of frequent money transfers. People often give and receive money, especially when lending and borrowing. This creates anomalies that are followed by the rapid mvmt funds rule, which triggers the algorithm. Therefore, multiple false alarms are generated. In summary, it can be observed that the application generated sufficient data. The comparison of the data with the output of the bank transaction history is depicted in Figure 6b,c.

According to the Figure 6c, the contending methods started with 60–69% accuracy; with increase in the transactions, the accuracy increased, reaching maximum 68.4–77.8%. All these processes are automated and executed within a monolith application, which provides better performance than DCT method, where a real-time AML system is described that has no automation and a poor technical description. Although Xblock-eos, and CNNBC ignore the process of creating the information, it provides visual data that are better than our method's provided raw text. However, it seems that Weber hardcoded the process of retrieving transaction tables, which means that these methods are flexible, and for other ATMs, transaction table application editing is needed. DIA attempts to use e a social network to predict the risks of money laundering; however, in reality, not every ATM stores information about originators or beneficiaries, which makes this approach useless, while our proposed BTS uses transaction-specific columns to define originators, beneficiaries, and money laundering risks.

In Figure 6d, a comparison of our proposed method is shown with the contending methods. Data for comparison are generated using results in every method scaled with our result CCNBC provides a good example of data generation, using SQL for processing; however, it shows the worst result because GraphViz takes a large number of computer resources for data visualization. Our proposed BTS provides better performance than other methods, and at the end of the process, the accuracy of the proposed BTS exceeds approximately 19.8 to 29.1% higher than contending methods.

Based on the results, the performance of BTS method and machine performance consumption are demonstrated. The overall results show that during the BTS process, the application constantly uses 2 GiB of RAM, but the CPU is loaded differently according to the specific method. Outlier detection uses only transactions of an entity where he or she acts as an originator, whereas the rapid mvmt method needs both originator and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 June 2021 doi:10.20944/preprints202106.0172.v1

beneficiary transactions to look for an anomaly. The processing requirements of the methods are as follows: outlier detection uses 10%, while rapid mvmt sometimes consumes an additional 5 to 10% more than outlier detection.Outlier detection takes 138,606 ms to process 21,602 transactions and generates 581 rows, which means that approximately every 37th transaction is suspected of involving money laundering. Additionally, the rapid mvmt method is processed five times more quickly than outlier detection because of the easy implementation of the algorithm.

Table 2. Summary of experimental results.

	Outlier Detection	**Rapid Mvmt**
Execution time	138,606 ms	139,701 ms
Number of rows generated	581	143
CPU usage	1.40 GHz	1.50 GHz
RAM usage	2 GB	2 GB

Based on the above, it is possible to say that the application can process a large amount of data and provide results in a minimum amount of time. Moreover, the provided results are more useful than the solutions proposed in [9–11]. The application records only the record generated date, the rule of a detected anomaly, and the entity that is suspected. However, this paper does not include ordinary user visualization, which is provided in [7–10] and can be the first priority. However, this can be easily fixed by including the frontend side of the application.

7. Conclusions

This paper introduces the Blockchain-enabled transaction scanning method that uses outlier detection and rapid mvmt funds rules to detect anomalies in bank transaction history. To validate the accuracy of the approach, the methods of the algorithm are executed by using the mock HSBC transaction history. Outlier detection works with one month's income and checks if the income of the last week is suspicious. Rapid mvmt funds works with two weeks' income and outflow and checks if income is between 80 and 120 percent of the outflow. Based on the simulation results, it is discovered that the outlier detection method works more accurately than rapid mvmt funds, and the algorithm in this paper does not require a super machine to execute ML to define the method. Outlier detection uses an algorithm to define anomaly actions that are not easy to unintentionally commit, while rapid mvmt selects transaction actions that may be innocent because of simple money transfers between accounts. Therefore, the accuracy of the second rule is low. However, a combination of these two rules provides good results that investigators could use for further cases. The proposed BTS is compared with other methods: Xblock-eos, CNNBC, DCT and DIA. Based on the testing process, we observed that proposed BTS produces the higher result than contending methods from the accuracy perspective.

Author Contributions: A.O and A.R., conceptualization, writing, idea proposal, methodology, and results; A.T, data curation, software development, and preparation; M.A. and B.A., conceptualization, draft preparation, and visualization; Z.C review, and editing. All authors have read and agreed to this version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Haller, D.R.; Nguyen, T.; Rowney, K.T.; Berger, D.A.; Kramer, G.A. System, Method and Article of Manufacture for Managing Transactions in a High Availability System. U.S. Patent 6,363,363, 26 March 2002.
2. Dodgson, M.; Gann, D.; Wladawsky-Berger, I.; Sultan, N.; George, G. Managing digital money. 2015.
3. Adrian, T.; Mancini-Griffoli, T. The rise of digital money. *Annu. Rev. Financ. Econ.* **2019**, *13*.
4. Kruisbergen, E.W.; Leukfeldt, E.R.; Kleemans, E.R.; Roks, R.A. Money talks money laundering choices of organized crime offenders in a digital age. *J. Crime Justice* **2019**, *42*, 569–581.
5. Chong, A.; Lopez-De-Silanes, F. Money laundering and its regulation. *Econ. Politics* **2015**, *27*, 78–123.
6. Weber, J.; Kruisbergen, E.W. Criminal markets: The dark web, money laundering and counterstrategies-an overview of the 10th research conference on organized crime. *Trends Organ. Crime* **2019**, *22*, 346–356.
7. Levi, M. Money for crime and money from crime: Financing crime and laundering crime proceeds. *Eur. J. Crim. Policy Res.* **2015**, *21*, 275–297.
8. Chaikin, D. Money laundering and tax evasion–the assisting of the banking sector. In *The Handbook of Business and Corruption*; Emerald Publishing Limited: 2017.
9. Alldridge, P. Tax avoidance, tax evasion, money laundering and the problem of 'offshore'. In *Greed, Corruption, and the Modern State*; Edward Elgar Publishing: 2015.
10. Kurnia, D.A. Study on money laundering practices from the criminal action results of political parties. *Translitera: J. Kaji. Komun. Dan Studi Media***2018**, *6*, 24–36.
11. Zhakenova, A. Features of the system of combating money laundering and financing of terrorism. *Bull. St. Petersburg State Univ. Econ.* **2018**, *3*.
12. Sullivan, K. *Anti-Money Laundering in a Nutshell: Awareness and Compliance for Financial Personnel and Business Managers*; Apress: 2015.
13. Maximillian, F.; Teichmann, J. Twelve methods of money laundering. *J. Money Laund. Control.* **2017**, *20*, 130–137.
14. Salehi, A.; Ghazanfari, M.; Fathian, M. Data mining techniques for anti money laundering. *Int. J. Appl. Eng. Res.* **2017**, *12*, 10084–10094.
15. Julian, W. Using ai to reduce false alerts and improve compliance. *Datavisor Inc.*
16. Pol, R.F. Anti-money laundering: The world's least effective policy experiment? together, we can fix it. *Policy Des. Pract.* **2020**, *3*, 73–94.
17. Alldridge, P. *What Went Wrong with Money Laundering Law?* Springer: Berlin/Heidelberg, Germany,2016.
18. Bergþórsdóttir, K. Local Explanation Methods for Isolation Forest: Explainable Outlier Detection in Anti-Money Laundering. Master's Thesis, Delft University of Technology, Delft, The Netherlands, 28 August 2020.
19. Singh, K.; Best, P. Anti-money laundering: Using data visualization to identify suspicious activity. *Int. J. Account. Inf. Syst.* **2019**, *34*, 100418.
20. Kolhatkar, J.S.; Fatnani, S.S.; Yao, Y.; Matsumoto, K. Multi-Channel Data Driven, Real-Time Anti-Money Laundering System for Electronic Payment Cards. U.S. Patent 8,751,399, 10 June 10 2014.
21. Raza, S.; Haider, S. Suspicious activity reporting using dynamic bayesian networks. *Procedia Comput. Sci.* **2011**, *3*, 987–991.
22. Weber, M.; Chen, J.; Suzumura, T.; Pareja, A.; Ma, T.; Kanezashi, H.; Kaler, T.; Leiserson, C.E. .; Schardl, T.B. Scalable graph learning for anti-money laundering: A first look. *arXiv* **2018**, arXiv:1812.00076.
23. Luo, X. Suspicious transaction detection for anti-money laundering. *Int. J. Secur. Its Appl.* **2014**, *8*, 157–166.
24. Colladon, A.F.; Remondi, E. Using social network analysis to prevent money laundering. *Expert Syst. Appl.* **2017**, *67*, 49–58.
25. Warren, E.; Tyagi, A.W. *All Your Worth: The Ultimate Lifetime Money Plan*; Simon and Schuster: 2005.
26. Force, F.A.T. Money laundering through money remittance and currency exchange providers. 2010.
27. VTheodorou; Jovanovic, P.; Abellò, A.; Nakuçi, E. Data generator for evaluating etl process quality. *Inf. Syst.* **2017**, *63*, 80–100.
28. Jullum, M.; Løland, A.; Huseby, R.B.; Ånonsen, G.; Lorentzen, J. Detecting money laundering transactions with machine learning. *J. Money Laund. Control.* **2020**.
29. Zheng, W.; Zheng, Z.; Dai, H.-N.; Chen, X.; Zheng, P. Xblock-eos: Extracting and exploring blockchain data from eosio. *Inf. Process. Manag.* **2021**, *58*, 102477.
30. Albakri, A.; Mokbel, C. Convolutional neural network biometric cryptosystem for the protection of the blockchain's private key. *Procedia Comput. Sci.* **2019**, *160*, 235–240.
31. Baek, H.; Oh, J.; Kim, C.Y.; Lee, K. A model for detecting cryptocurrency transactions with discernible purpose. In Proceedings of the 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), Zagreb, Croatia, 2–5 July 2019; IEEE: 2019; pp. 713–717.
32. Farrugia, S.; Ellul, J.; Azzopardi, G. Detection of illicit accounts over the ethereum blockchain. *Expert Syst. Appl.* **2020**, *150*, 113318.