Development of Algorithms for Searching, Analyzing and Detecting Fraudulent Activities in the Financial Sphere

Marina Pavlovna Khrestina¹, Dmitry Ivanovich Dorofeev², Polina Andreevna Kachurina³, Timur Rinatovich Usubaliev⁴, Aleksey Sergeevich Dobrotvorskiy⁵,

Abstract:

According to Digital Evolution Index 2017, Russia is included to the category of so-called “Break Out” countries. The major problem to be encountered at transfer to the digital economy is adaptation of new technologies – such as Big Data, Blockchain, Internet of Things, Cryptocurrency, machine learning. No less important field is development of friendly informative environment facilitating international cooperation, cyber safety problems resolving, etc.

This example provides the data of the report prototype of a system to detect suspicious transactions. This system shall read and analyze the transaction database and, in accordance with search algorithms, it detects suspicious transactions within the entire data base.

The algorithm consists of several stages: development of a graph, selection of suspicious and trusted transactions, calculation of signs and machine learning. The methods of social connections analysis, parallel processing of graphs and mathematical apparatus of neural networks are used as the basis of this research.

Keywords: Digital economy, antifraud, AML-systems, digital transformation, machine learning, Big Data, financial control, parallel processing, neural networks.

JEL Classification: F50, F52, F59.

¹ Moscow Technological University (MIREA), e-mail: khrestinam@mail.ru
² Quality Software Solutions LLC, e-mail: dmitry.dorofeevivanovich@mail.ru
³ ITMO University, e-mail: vlpolina@yandex.ru
⁴ Quality Software Solutions LLC, e-mail: usubaliev.timur@mail.ru
⁵ Moscow Technological University (MIREA), e-mail: aleksey.dobrotvorskiy@gmail.com
1. Introduction

Modern realities require application of significantly higher technological solutions than it used to be earlier - especially in the sphere of finance where the highest level of security shall always be provided. In terms of technologies, everything is already provided to develop various systems to perform safe and trusted financial operations. Nowadays the world is being obsessed with the idea of digital transformation, and some shifts towards it are inevitable. ‘Digital organizations’ with no branches take the leading positions. Thus, the urgency of development of this type of business is extremely high. In terms of technologies, we have covered all the way to reach totally different level of establishment. Systems of artificial intellect, neuronets, machine learning and machine recognition systems enable design and implementation of various high-technological systems to resolve the challenges impossible to implement earlier.

This study defines “fraud” in the following way:

- attempts to launder proceeds obtained illegally;
- first party fraud;
- fraudulent debit card charges.

This article considers the first type of fraud – money laundering and various Anti-Money Laundering (AML) solutions to be applied in the Russian Federation within the existing legislation. The requirements of mandatory implementation of procedures of counteractions to legalization (laundering) of money obtained illegally and terrorism backing are enforced at the level of federal legislation and normative documents of regulator: The Federal Law N 115-FZ, the Federal Law N 134-FZ (Digital Evolution Index 2017; Hosmer and Lemeshow 2000), The Criminal Code of the Russian Federation, the normative documents of the Central Bank of Russia. The campaign against criminal money laundering and terrorism backing is global.

Violation of requirements of counteractions to legalization (laundering) of money obtained illegally and terrorism backing is the basic factor for punitive sanctions, significant reputational losses and finally termination of activities of financial organizations. To prevent such risks, effective implementation of counteractions to legalization (laundering) of money obtained illegally and terrorism backing is required.

The existing solutions, including those that employ machine learning, are not capable to process operations online. The intellectual analysis of the existing solutions is executed filtration of transactions in accordance with the rules. For instance, it is affected by the preset limitations. All the filtered transactions are considered ‘legitimate’ and do not undergo the procedure of analysis, which may be used in illegal purposes. The existing solutions consider only the indicators directly associated with an operation or its participants and do not take indirect indicators
into account. As a result, false activations and errors occur. Thus, the top of the agenda challenges in the field of automation of counteractions to legalization (laundering) of money obtained illegally and terrorism backing are as follows:

- Incapability to automate online those factors and signs that are unique for specific types of financial activity or financial agencies;
- Absence of a centralized tool to analyze transactions;
- Essential requirement to use a combination of software and manual analysis;
- High numbers of false activations and errors;
- Significant costs of maintenance of analysts’ team to provide manual analysis of transactions.

2. Digital Economy

In recent years the topic of digital economy has been widely discussed at different levels – in scientific, business and governmental circles. The modern world has been developing rapidly. The boost of informative and communication technologies, innovations, and scientific and technical sector, modernizations of many industries – all this promotes to form a new vision of economy and the society’s positive attitude to it.

One of definitions for “digital economy” is as follows – it is a sort of commercial activity applied to production and sales of electronic goods and services. In recent years the concept of digital economy went beyond the scope of buying and selling electronic products on the internet. Today this idea also includes application of virtual processes within financial activities of large companies and corporations, which requires analysis of a significant amount of financial transactions, automated detection and blockage of illegal and suspicious operations based on decisions provided by machine analysis. The opportunities of digital economy are clear – it has a great potential, but there are some suspicions and questions regarding it. The basic are as follows – what are the risks associated with the transfer to the new technological procedures? What challenges can be predicted?

According to Digital Evolution Index 2017, Russia is included to the category of so-called “Break Out” countries: “Though still at relatively lower absolute levels of digital advancement, these countries demonstrate the fastest momentum, are poised for growth and are attractive to investors” (Digital Evolution Index, 2017).

The major problem to be encountered at transfer to the digital economy is adaptation of new technologies – such as Big Data, Blockchain, Internet of Things, Cryptocurrency, machine learning. No less important field is development of friendly informative environment facilitating international cooperation, cyber safety problems resolving, etc.
This example provides the data of the report prototype of a system to detect suspicious transactions. This system shall read and analyze the transaction database and, in accordance with search algorithms, it detects suspicious transactions within the entire database. The algorithm consists of several stages: development of a graph, selection of suspicious and trusted transactions, calculation of signs and machine learning. The logistic regression is a method of machine learning.

1) **Stage 1**: Development of a graph whose peaks consist of persons and the edges, of transactions. 
2) **Stage 2**: Selection of multitude of edges for the machine learning stage: all suspicious transactions and the same number of regular transactions (randomly chosen). 
3) **Stage 3**: Selection of communities around each sender in order to calculate the signs. A community consists of an original sender and all the neighbors around (search depth 1) in accordance with initiated transactions. 
4) **Stage 4**: Calculation of signs for each edge (transaction) (9 signs in total):
   - Basic for a given edge:
     - Degree of sender and recipient
     - Transaction amount
   - In a sender’s community:
     - minimum, maximum and average degree of peaks of a given community
     - number of peaks
     - number of transactions and amount of all transactions.
5) **Stage 5**: Machine learning.
   - The method applied is logistic regression (Hosmer and Lemeshow 2000) (it is applied to predict a probability of occurrence of an event by the values of a multitude of signs)
   - Signs selection method is chi-square test.
   - Marking out of selection - random, 70% for learning stage (with cross-validation), 30% - to test the results.

The logistic regression method does not provide forecast for the value of numeric variable based on selection of original values. Instead of it, the value of function is the probability of belonging of a given original value to a specific class.

1) We may consider the function q(x), where x is a point of data of teaching selection. The value of function is the probability of belonging of a given original value to a specific class.
2) The logistic regression learning mechanism tends to maximize an average value q(x).

The chi-square test allows to estimate the significance of differences between actual (detected as a result of investigation) number of outcomes or qualitative properties of selection for the each of categories and theoretical number to be expected in the
studies groups in case of validity of zero hypothesis. To put it simple, the method helps estimate the statistical significance of differences of two and more relative indicators (frequencies, shares). Precision of classification of this method (the share of correctly classified items) is 89.53%. The selection of signs has showed that all the 9 signs are a matter of importance (the quality of model drops in case some signs are taken out). The other quality metrics are as follows (the closer to the value of 1.0 the each of them is, the better the result is):

- Sensitivity (also known as "recall") = 0.868.
- NPV (negative predictive value) = 0.859 (= number of real proper transactions/predicted proper transactions).
- F1 score = 0.899.

Thus, we can see that application of machine learning algorithms provide a high percentage of correct solutions, which allows to widely apply the similar solutions in the sphere of counteraction to illegal operations in the financial market.

3. Analytical review of existing solutions

During the analysis the following expected functional capacities of the developed system of prompt analysis of risks and assessment of states of financial processes’ participants shall be considered (note 1):

1) processing and analysis of transactions in offline mode;
2) organization’s financial processes monitoring automation;
3) extended analysis of clients’ data; assessment of reliability of legal bodies through the database for legal bodies data;
4) high rate of decision making;
5) an option of automated adjustment of algorithms of data analysis for various variants and principles in order to detect sustainable regularities and direct and indirect links between objects;
6) examination of formal requirements of the Central bank of Russia and Rosinfo monitoring.

It is expected that the system developed will use the following source data for the analysis:

1) transaction details;
2) detailed information of transaction’s counteragents;
3) transactions history (transactions’ graph);
4) training extract (a set of transactions with tags – suspicious/appropriate).

It is required, with the use of all the data provided, to develop a system capable to detect suspicious transactions (Ray 2015; Pesce, 2014; Stroeva et al., 2016; Suryanto and Thalassinos, 2017; Thalassinos, 2008; Thalassinos et al., 2015).
3.1 SAS AML

SAS AML (sas.com) is being applied in such banks as Gazprombank, Commonwealth Bank of Australia. The basic advantages of the solution are as follows:

1) Complete automation of organization’s financial processes monitoring, including detection of assessment of suspicious transactions, conduction of investigations;
2) Implementation of Know Your Customer (KYC) principle;
3) High rate of decision making for suspicious transactions;
4) Funds traffic in transactions is provided as a graph;
5) Report constructor.

The disadvantages are as follows:

1) Processing and analysis of transactions are possible only in deferred mode;
2) Impossibility of system’s self-training on actual data – only by manual adjustment of rules;
3) Absence of analysis for extended data level on legal bodies – transactions’ participants;
4) Extremely high cost of licensing, implementation and ownership. The solution can only be affordable by the major banks.

3.2 Oracle Mantas AML

The product is an analogue of SAS AML. Oracle Mantas (oracle.com) is used by several lending agencies, such as Deutsche Bank AG, ABN AMRO Bank N.V., Credit Suisse Group, Barclays Bank, Citibank N.A., Merrill Lynch, Charles Schwab, Goldman Sachs, BBVA, M&T Bank. Among the Russian agencies the major user of Oracle Mantas is Sberbank.

The advantages of solution are as follows:

1) Complete automation of agency’s financial processes monitoring, including detection of assessment of suspicious transactions, conduction of investigations;
2) Implementation of KYC principle;
3) Scenario analysis;
4) High rate of decision making for suspicious transactions;
5) Report constructor.

The disadvantages are as follows:

1) Processing and analysis of transactions are possible only in deferred mode;
2) Impossibility of system’s self-training on actual data – only by manual adjustment of rules;

3) Absence of analysis for extended data level on legal bodies – transactions’ participants;

4) Extremely high cost of licensing, implementation and ownership. The solution can only be affordable by the major banks.

3.3 Nice Actimize AML

Nice Actimize AML (niceactimize.com) has been implemented in the major global financial agencies, including the controlling units:

- The Central Bank of the Russian Federation (the Federal service for financial markets in the Russian Federation);
- AFM (The Netherlands Authority for the Financial Markets);
- CFTC (The U.S. Commodity Futures Trading Commission);
- FINRA (The US Financial Industry Regulatory Authority);
- FINTRAC (The Financial Transactions and Reports Analysis Centre of Canada);
- IMPA (The Israel Money Laundering and Terror Financing Prohibition Authority).

The advantages are as follows:

1) Capacities to process and analyze transactions both in the deferred mode and online;

2) Complete automation of agency’s financial processes monitoring, including detection of assessment of suspicious transactions, conduction of investigations;

3) Self-training capacity of the system on actual data;

4) Implementation of KYC principle;

5) Scenario analysis;

6) High rate of decision making for suspicious transactions;

7) Report constructor.

The disadvantages of Nice Actimize AML include:

1) Absence of analysis for extended data level on legal bodies – transactions’ participants;

2) Relatively high cost of licensing, implementation and ownership.

3.4 FLEXTERA Counteraction of legalization of revenues
AML-module of the line of banking products FLEXTERA by the Russian company ‘Diasoft’(diasoft.com). Equipped with a basic functional that can partially automate AML-processes, in combination with affordable cost.

The advantages are as follows:

1) Low cost of the licensing (~870 thousand rubles), implementation (~2 mln rubles), ownership (technical support ~500 thousand rubles).

The disadvantages are as follows:

1) Processing and analysis of transactions are possible only in deferred mode;
2) Partial automation of agency’s financial processes monitoring (transactions’ control for compliance with the manually preset rules only);
3) Low rate of decision making for suspicious transactions; as the investigations are conducted manually;
4) Impossibility of system’s self-training on actual data – only by manual adjustment of rules, with involvement of developers;
5) Absence of extended analysis for data level on legal bodies

3.5 CFT ‘Control of Income Legalization (Laundering) Risks’

The product made by the Center of Financial Technologies(cft.ru) is similar, by its properties, to its analogue by ‘Diasoft’. This is why it demonstrated similar advantages (cost) and disadvantages (limited automation for AML-processes, low efficiency of detection of suspicious transactions) (Markets and Markets, n. d.).

3.6 Counteragents’ control services (Spark/X-Compliance, Kontur-Fokus)

The advantages of existing services, such as, for instance, Spark, X-Compliance(xco-interfax.ru) (Counteraction of the Legitimization {Laundering} of Proceeds of Crime and the Financing of Terrorism spin-off Spark), Kontur-Fokus (focus.contur.ru), are as follows:

1) High recognizability, awareness and habituality for users;
2) Wide range of data provided, up to analysis of publications in mass-media, social networks and official websites of legal bodies.

The disadvantages are as follows:

1) Absence of analysis for AML challenges. Data interpretation is being performed manually.
2) The criteria of reliability assessment of legal bodies are not transparent, not accessible to be corrected by users and, consequently, are not always suitable to consider in case of investigation.
4. The examples of typical schemes of laundering of Proceeds of Crime

- Placement — cash transfer to mobile financial tools, territorial relocation of funds away from the location of their origin.
- Layering — criminal revenues separation away from their sources of origin through a sophisticated chain of financial operations in order to mask verifiable tracks of such revenues.
- Integration — the last stage of legitimization process directly aimed to make the assets of criminal origin to appear as legitimate.

4.1 Rotary scheme

The scheme is based on VAT (value added tax) refund principle. The example is as follows:

1) company A, located in country 1, acquires some goods from a supplier in country 2, with zero VAT rate;
2) upon acquisition, the goods are being supplied to another trader (company B, located in the same country) at the same price with VAT. However, the company A does not pay out VAT to the favor of state and turns into a ‘Disappearing trader’;
3) company B then sells the goods to another country (often – to the same company that has been an original supplier) and claims to refund the VAT that has been paid at acquisition of the goods from the company A.

4.2 The laundering scheme based on invoices for fictitious jobs

Company A’s CEO (chief executive officer) signs several contracts with subcontractors for building and assembly jobs that the Company A is performing on their own. Then the jobs are transferred to a subcontractor and further along the chain (through short-lived companies). The each of them will receive a commission fee for money laundering and confirm the execution of jobs under the contracts that included fictitious reports. In the end of chain an owner of Company D is trying to recall the cash, and it is implied that this money will be received by the management of Company A in ‘under the table’ form (possibly through a trusted agent connected here with Company C) (Jedrzejek et al., 2009; Albekov et al., 2017; Stroeva et al., 2016; Thalassinos et al., 2012; Vovchenko et al., 2017; Bogdanova et al., 2016). The laundering scheme based on invoices for fictitious jobs as described in the paper is shown Figure 1. In this scenario the funds are being withdrawn from companies to individuals.
Figure 1. The laundering scheme based on invoices for fictitious jobs as described in the paper.

4.3 Simple scheme (corruption)

Figure 2. Simple scheme (corruption) as described in the paper
Subject A opens a banking account 1. Subject B opens a banking account 2. Then an international investor, through a number of transactions, transfers money to the account 1. Subject A withdraws a part of money from the account 1 (in cash), and another part he transfers from the account 1 to account 2. Subject A places a significant amount of cash on the subject B’s (account 2). Then Subject B transfers a part of funds from the account 2 to his foreign account (Irwin et al., 2012). See Figure 2.

4.4 The scheme of placement of funds into financial system by outsiders

The suspects, allegedly acting in some third persons’ favor (who wanted to stay unrevealed) have opened accounts in various foreign banks. There is a special person – a representative – who is engaged in placement of cash. There are also some other persons who will also place great amounts of cash into the banking accounts considered. Eventually the suspects will transfer the large sums of money from the accounts considered to the various foreign accounts.

The suspicious signs are as follows:

1) absence of rational business need to perform such operations;
2) multiple and repeated transfers between the accounts;
3) participation of suspicious persons;
4) ‘layering’ schemes application.

4.5 The scheme of placement of funds into financial system through fake companies

The scheme of placement of funds into banking accounts of various companies is employed. The funds shall be placed regularly and in form of large sums of cash. This money is being transferred immediately to the accounts of foreign banks owned by foreign companies (those who benefit by the given scheme).

The indicators of suspicious activities for the given scheme are as follows:

1) questionable level of real business activity;
2) multiple and repeated transfers between the accounts;
3) funds’ source cannot be verified;
4) employment of fake companies (or short-lived companies);
5) ‘layering’ schemes application.
5. Examples of money laundering associated with tax violations

5.1 The scheme of tax avoidance illustrated by the example of a jewelry company

A large jewelry company with their own jewelry production sells the goods through their own retail chain. See Figure 3. The suspicious indicators and the signs of this scheme are as follows:

1) simultaneous payments for goods supply in favor of a fake company B, executed by jewelry companies and individual entrepreneurs;
2) sales of gold bars of the jewelry company by a bank, through a chain of two intermediates, with obvious signs of fictitiousness;
3) there are payments in favor of other fake companies, executed by intermediate companies:

*Figure 3. The scheme of tax avoidance as described in the paper*
Schemes of money laundering, involving placement of cash, high number of transactions, tax evasion are also possible.

6. Conclusions

In the conditions of modern realities, both the legislation requirements and sanctions for late detection of untrusted operations are becoming stricter. Along with significant reputational losses, financial agencies might be subject to pay high fines and even suspension of activities, or even be deprived of license. Large companies perform up to 20 million transactions per day where detection of suspicious operations is associated with manual work (Dow 2015; Suryanto, 2016). This might cause the risk of errors, false detections and lead to excessive efforts. So far this is the reason why AntiFraud-systems development and implementation is of a high demand. Thus, the modern tendencies require development of high-tech systems with capacities and abilities complying with the requirements of digital economics.

Any mathematical model within the systems of machine learning will work in accordance with the principle of search of patterns. These models will help teach the neuron networks to make much more credible decisions, at the same time, the goal is not to develop systems that would resemble a human outwardly but those that would act as a human, or design self-organizing databases that would adjust themselves to a specific goal without the need of administration.

Many banks confirm that they find it difficult to track the constantly changing rules, to interpret and implement global regulatory modifications at their local operational level, gather and update information from multiple sources and systems. Traditional solutions based on the rules require significant engagement of human factors, particularly at the stage of investigation, which is a costly and basically ineffective procedure, error-prone.

The decisions that support machine learning and artificial intellect are able not only to automate essential share of operations but also provide high reliability of information due to enhanced capabilities of analysis of structured and unstructured data (Ray and Katkov, 2016; Bogdanova et al., 2016).

Making four less of a chore (Gabriel, 2016):

1) Research and analysis of clients’ activity
2) Communications monitoring - emails analysis by key words
3) Multi-language monitoring - Compared with existing tools, Artificial Intelligence (AI) solutions offer a better option to translate or transliterates client names, along with code words and other data in languages and scripts other than those used in languages with Latin roots.
4) Watch list monitoring - industry lists tend to be long and include names in non-Latin languages – which makes ongoing analysis a challenge for
existing tools.

For example, in the 3rd quarter of 2015 “VTB 24” opened 29 thousand settlement accounts, and every fifth client of the bank was refused to open an account. According to some data, in two months Alfa-bank has closed a few thousand accounts of small companies with signs of suspiciousness.

Acknowledgements:
The research is being conducted with the financial support of the Ministry of Education and Science of the Russian Federation (Contract №14.574.21.0142) Unique ID for Applied Scientific Research (project) RFMEFI57417X0142. The data presented, the statements made, and the views expressed are solely the responsibility of the authors.

References:


