USING QUANTITATIVE METHODS TO IDENTIFY SECURITY AND UNUSUAL BUSINESS OPERATIONS

Antonín Korauš¹, Miroslav Gombár², Pavel Kelemen³, Stanislav Backa⁴

¹ Academy of the Police Force in Bratislava, Sklabinská 1, 835 17 Bratislava 35, Slovak republic
²,³,⁴ University of Prešov in Prešov, Faculty of Management, Konštantínova 16, 080 01 Prešov, Slovak Republic

E-mails:¹ antonin.koraus@minv.sk, ² miroslav.gombar@unipo.sk, ³ kelemen.pavel@gmail.com, ⁴ stanislav.backa@gmail.com

Received 16 October 2018; accepted 29 January 2019; published 30 March 2019

Abstract. Financial institutions are the first vertical level in the fight against money laundering and to improve security. Therefore, it is essential that tools are available to enable effective detection and analysis of suspicious transactions, or unusual business operations. These, in line with the legislative requirements, report to responsible entities - FIUs representing the second vertical plane in the fight against money laundering. However, special software tools are available for obligated persons, especially for financial institutions that carry out tens of millions of financial transactions a day. These can trigger the alert to most unusual operations. The software automatically creates customer profiles, including expected behavior and executed transactions. Using advanced statistical analyses, it identifies unusual business operations, i.e. financial transactions significantly different from those carried out in the past. It is very useful to apply software support in form of electronic detection of indicators of legalization of crime proceeds. However, the output of such support software requires a more detailed and demanding investigation of the nature of operation and is based on the use of special algorithms based on mathematical and statistical methods. The software builds on the results of scientific research.

Keywords: security; unusual business operations; crime; money laundering; corruption; legal acts

Reference to this paper should be made as follows: Korauš, A.; Gombár, M.; Kelemen, P.; Backa, S. 2019. Using quantitative methods to identify security and unusual business operations, Entrepreneurship and Sustainability Issues 6(3): 1101-1112. http://doi.org/10.9770/jesi.2019.6.3(3)

JEL Classifications: E26, E42, G21

1. Introduction

In mid-eighties, computational models based on biological neural networks begin to emerge in scientific contributions aimed at detecting unusual business operations associated with money laundering or terrorism. Mid-nineties are marked by the onset of the theory of fuzzy sets, fuzzy logic, and fuzzy decision making, as well as
spatial statistics and, in particular, data mining. At the end of the last century, genetic algorithms were becoming popular.

Software support specifically designed for financial institutions to identify cases of legalization of proceeds of crime is based on the use of mathematical and statistical methods, especially those dealing with identification of remote values, probability distribution of data - Benford law, correlation and regression, time series, spatial statistics, core component, cluster analysis, decision trees, naive Bayes classifier, graph theory, game theory, fuzzy logic, fuzzy sets, and neural networks.

Data mining can be characterized as data stripping, discovery of knowledge in a database. It is a specific process of gathering new useful information. The process of modeling results in a description of patterns and relationships appearing in the data. It covers existing techniques of data analysis as well as knowledge and information acquisition (Mura et al., 2018; Zulova et al. 2018). The discovery of knowledge in the database is of interdisciplinary character. Data mining uses logistic regression and decision trees. Unlike the classical use of mining data, where the criterion of model quality lies in its accuracy, in case of fraud detection and money laundering, the criterion of quality lies in profitability, or in other words, in the probability of detecting fraud or money laundering (Jančíková, Veselovská 2018; Jančíková, Pásztorová 2018; Grančay et al., 2015; Mura et al. 2017).

2. Theoretical background

Gao, Ye (2007) dealt with using the process of data mining to identify money laundering. The FAIS (Financial Crimes Enforcement Network AI Systems), described by Senator (1995), allows mining data facilities to identify activities potentially related to the legalization of criminal proceeds.


Article of Ngai et al. (2011) reviews and presents a detailed classification system for the use of data mining techniques to detect financial fraud. This document presents the first systematic, identifiable, and comprehensive review of professional literature on data mining techniques used to detect financial frauds. Forty-nine journal articles published from 1997 to 2008 were analyzed. The authors identified four categories of financial fraud (bank fraud, insurance fraud, securities and commodity frauds as well as other financial frauds). They identified six mining data classes (classification, regression, clustering, prediction, outlying values and visualization). The findings of the survey clearly show that data mining techniques have been used extensively to detect insurance fraud. Nevertheless, corporate frauds and credit card frauds have also attracted great attention of the authors in recent years. On the other hand, there is a clear lack of research on mortgage fraud and fraud in money laundering and securities and commodities. The main data mining techniques for detecting financial fraud are logistic models, neural networks, Bayesian networks and decision trees, all of which provide primary solutions to fraudulent data detection and classification problems.
Financial frauds affect millions of people every year and financial institutions must use methods for protecting themselves and their clients. Using statistical methods to address these issues faces many challenges. Financial frauds are rare occurrences leading to extreme imbalances in the data. The volume and complexity of financial data require algorithms that need to be effective. The article of Sudjianto et al. (2010) focuses on two important types of financial crime, namely fraud and money laundering. It describes some of the traditional statistical methods that have been used, as well as the newer neural networks and data mining. The aim of this article is to provide an overview of a wide range of methods accompanied by selected illustrative examples.

In recent years, the number of money laundering offenses has grown through foreign exchange and payment transactions. Cybernetic security issues, which are often perceived as synonymous with the safety of critical infrastructure (e.g., Dobrovič et al., 2017, Veselovska et al., 2017, Korauš, Kelemen 2018, Šišulák 2017, Limba, Šidlaukas 2018), need to be emphasized.

The Sohn study (2005) proposes four scoring models providing early warnings of money laundering in foreign exchange transactions for incoming and outgoing transfers of funds. Models of logistic regression, decision trees, and neural network are used as well as that combining all three previous models. It appears that the accumulated number of transactions is the most important indicator.

Hawkins (1994) defines an outlier as an observation that deviates so much from other observations as to arouse a suspicion that it was generated by a different mechanism. Frauds of a different nature, including the legalization of proceeds of crime, can be revealed as anomalies in the data. For this reason, outliers may become subject of interest in uncovering unusual operations.

Multiple statistical methods were used to identify outliers (Barnet & Lewis, 1994). We can divide outlier detection methods into two large groups. Methods based on probability distribution and distance-based methods. In the case of money laundering issues, distance-based methods (for example, Euclidean, Manhattan distance) are appropriate. These methods are also used for the detection of outlier in multidimensional cases (Knor, Ng 1997; Kosinski 1999; Knorr, Ng, Tucakov 2000). An unusual distance is that from the nearest neighbor. The literature lists three types of outlier identified by distance. Outliers are objects, whose average distances to the kth closest neighbors are the largest. This definition was used by Angiulli and Pizzuti (2002). Ramaswamy at al. (2000) have labeled n outlying objects whose distances to the kth closest neighbors are greatest. Knorr, Ng (1999) and Knorr at al. (2000) have labeled outlying objects for which there are fewer than p objects at a distance greater than d. Many outlier detection algorithms that work on small files cannot be used for databases containing tens or hundreds of millions of data. Hung, Cheung (1999), Hand and Blunt (2001) reported an algorithm for the detection of outliers in large databases. The Sherlock system for auditors uses a combination of detection of outliers with classification techniques (Bay et al. 2006). According to Zhu (2006), the detection of outliers is a key element for intelligent financial supervision systems to identify frauds and money laundering by discovering unusual behavior of customers. Detection procedures generally fall into two categories, namely by comparing each transaction against its account history, and then, in order to find out whether its behavior is unusual, by comparing it against the reference set. The paper presents an approach to reduce false positivity.

The Benford’s law describes the probability distribution of the first and other significant numbers. It demonstrates that real data sets remarkably contradict the even probability distribution. To come to this conclusion, Benford analyzed 20,229 files from a wide variety of areas. From the statistical point of view, it is the conformity testing (for example, using the Chi-square goodness of fit test) of the empirical distribution of the first and other significant numbers of the studied accounting data with the theoretical division according to Benford’s law. The Benford’s Law is a simple and effective tool for detecting fraud in accounting (Durttschi at al. 2004). Nigrini (1999) used Benford's law for fraud detection in accounting. He also pointed out the difficulty in generating data that comply with Benford's law. To verify the validity of the Benford's Law, Nigrini and Mittermaier (1997) and
York (2000) have pointed it out as a technique suitable for audits. Not only is the Bernford's law applicable in detection of fraud in accounting, it can also be used for identifying money laundering from financial statements.

3. Research objective

Individual phenomena are always found in certain mutually dependent and determining relationships. The examination of their dependence requires an important step of choosing the appropriate statistical features that characterize given phenomena. The correlation number solves two basic tasks, namely the correlation task, which assesses the tightness of dependence and determines the characteristics describing to what extent the independent variables explain the variability of the dependent variable, and the regression task, which determines the shape of a regression function and estimates its parameters. Regression analysis provides a set of estimates that can predict the effect of certain variables on the variables examined. An important role is played by testing the statistical significance of variables. Nevertheless, caution should be exercised since mechanical testing of statistical significance may conceal actual significance by the size of the studied variable. Statistically significant relationships can be considered negligible if the effect size is too small. Statistical insignificance can occur with many effects that show a high degree of uncertainty. This could cause ignorance of potentially significant money laundering effects. Statistical significance itself is neither a necessary nor sufficient condition for real significance of the observed variable. Logistic regression models are among the generalized linear models. They are used to identify factors associated with false published financial statements of Greek companies. The accuracy of the correct classification exceeds 84 percent. Bell and Carcello (2000) on a sample of 77 fraudulent statements and 305 statements without fraud designed a logistic regression model to estimate the probability of fraud in the financial statement. On the basis of a set of 130 companies with detected fraud in 1989-2004 and 83 companies without fraud, Jay et al. (2006) have constructed logistic models. Annual averages, shares, percentage changes, and dummy variables entered the models. The likelihood of correct classification by model for the year in which the fraud was committed is 60.0 percent, while for the year before fraud, it is 55.8 percent, and for a year after fraud, it is 61.2 percent. Jay and others (2006) used a regression analysis to elucidate the causes of crime. Regression analysis plays an important role in the macroeconomic approach to measuring the extent of money laundering, especially in quantification based on currency demand, material input-based approach, usually electricity consumption and econometric approach when considering an unobservable variable between certain observable causes and consequences (Walker 1999; Schneider 2007; Straková et al. 2017). Reliable risk assessment method RM/RA CRAMM applicable for a crime risk assessment was described by Mullerova (2016), Mamojka and Mullerova (2017) for multi-purpose use and by Palkova et al.(2018a) for National Risk management system and for disasters (Palkova 2018b) and its legal questions by Mullerova and Mamojka (2017). At the micro level, it is mainly an estimate of the amount of proceeds from the sale of drugs, stolen goods, etc. The range of these activities is usually estimated in a few years and the values between them are approximated.

Cluster analysis is a method that enables to find internal structure in large-scale data sets in the form of so-called clusters. In case of large files, it is a key question to choose a suitable distance and right algorithm. Cluster analysis can also serve as a means for detecting outliers. Data remote from existing clusters can be considered as outliers. For clustering of huge and massive databases, special clustering algorithms need to be used (Williams, Huang 1997, Yamanishi et al. 2000, Zhang et al. 1996). Jiang et al. (2001) presented a two-stage clustering algorithm for outlier detection. The clustering approach suggested by Yang et al. (2014) is tested on bulk data provided by the bank. The result shows that this method can automatically detect suspicious cases of financial
transactions. Rohit and Patel (2015) claim that clustering techniques are the best techniques for detecting non-cyclical operations.

Graph theory represents the discipline of applied mathematics. It is part of network theory. Organizational networks are important from the point of view of the analyzed issue. Special types of networks are often included in artificial intelligence and expert systems applications. Presenting the data in form of charts is useful in several areas. Relationship analysis uses a variety of theoretical techniques of graph theory.

Xu and Chen (2005) dealt with the analysis of visualized various types of criminal organizations. For fraud detection, but also for money laundering, it is important to find anomalies in Ditas which are in graph forms. Based on the entropy method, Shetty and Adibi (2005) identified the most important people in the Enron scandal based on email addresses. The most important are the vertices (in case of Enron, the people), whose omission from the chart results in the greatest entropy change. Noble and Cook (2003) transformed the structural anomalies in the graph to detect anomalous subgraphs. Lin and Chalupsky (2003) defined different metrics to quantify matching edges between the vertices. It is necessary to note that the most important peaks do not necessarily have to be anomalies. Social Network Analysis (Scott, Carrington 2011) is also useful for relationship analysis. It draws attention to the consistency of notaries who enter the notarial records of companies in which the same persons appear. Similarly, the fact that the same doctor confirms multiple injuries in the same person may signal a potential fraud. Network analysis is also useful for identifying a network of people involved in money laundering.

Decision making is the process of selecting one of several variants. The situations in which one of the more variants is to be chosen are the decision situations. A rational participant in this process is one who selects the best option in some ways. An indifferent participant is indifferent to the outcome of the decision. It is a random mechanism that selects variants according to the probability distribution. The situations when the outcome of the decision depends on the participants' decisions are referred to as being conflicting. It is assumed that the outcome of decision-making from the point of view of a rational participant can be evaluated by one (scalar evaluation) or multiple criteria (vector ratings). Game theory deals with decision-making situations with more rational participants with scalar rating. Conflict decision situations with one rational participant and vector valuation are dealt with by the theory of multicriteria optimization. Conflict decision situations with one or more rational participants and one indiscriminate participant with scalar rankings deal with the theory of decision being indefinite. Jones (2004) uses a game theory machine to model the behaviour of managers of companies with fraud. They studied the impact of four factors on the interaction between auditors and managers. The results indicate the relationship between testing, fraud detection and fraud prevention in the company. An audit model involving internal control is analyzed as a non-cooperative game. Matsumara and Tucker (1992) also used the theory of games to detect fraud.

Fuzzy sets and fuzzy logic are a generalization of classical double-valued logic and set theory. Instead of binary assignments, the interval is used for true values $\langle 0, 1 \rangle$. Knowledge and information are linguistically shaped. However, computers need accurate information. Fuzzy sets are useful for translating inaccurate verbal information into numerical information. The linguistic variable is a variable whose values are words or phrases of a natural or artificial language. Fuzzy sets are successfully used to express the content of a linguistic variable. The design of the function of affiliation of the fuzzy set is a fundamental problem. This is possible based on expert knowledge and also by using neural networks. Mastroeleo et al. (2001) used a fuzzy expert system for detecting insurance frauds. Estévez et al. (2006) used a combination of fuzzy decisions and a neural network to prevent fraud in telecommunications. The fuzzy set and fuzzy logic apparatus are applied mainly in conjunction with neural networks.

The neural network is a massively parallel processor that has the ability to memorize knowledge gained experimentally and to further exploit this knowledge. The characteristic feature of the neural network lies in its
structure formed by individual neurons linked by synaptic connections. Neural networks can be used successfully in addressing prediction and process management problems, as well as in classifying the objects. Neural networks are able to solve problems that are difficult to solve using classical algorithmic techniques. The neural network does not follow a predetermined algorithm. It is able to learn from examples and use the learned information.

The most common type of neural network is represented by multilayer perceptron (MLP). Another type of neural network used to detect fraud are the self-organizing maps (SOM). Their specificity is given by the fact that under certain conditions it allows a display that retains the typology and displays the characteristic features of the trained set of data. Neural networks are used in all situations where it is necessary to understand complex relationships between variables. Neural networks are non-linear in their design and do not need to explicitly specify the shape of the dependency function. They can also find interactive effects. The acquired knowledge is implicitly stored in the network setup vector. Neural networks based on fuzzy rules allow for improved performance of solutions. Neural networks are usually linked with an expert system into hybrid systems. Another option lies in extracting knowledge from a neural network into an expert system. Genetic algorithms are suitable for setting neural network parameters, difficult optimization problems, and machine learning with classification systems.

Neural networks based on decision trees are used in BAYES, FOIL, RIPPER and other fraud detection systems. The ASPeCT project of the European Commission, Vodaphone and other European telecommunication companies for detecting cell phone fraud also uses neural networks. Neural network technology is also used in the FALCON software suited to detect credit card fraud. Bolton and Hand (2002) gave an overview of neural network applications for detecting legalization of proceeds of crime, credit card fraud, and telecom fraud. The BRUTUS system allows for detecting fraud with mobile phones using neural networks. Ezawa and Norton (1996) also used neural networks to detect fraud based on telecommunication accounts. Stefano and Gisella (2001) developed a fuzzy expert system for the detection of insurance frauds. In order to identify a suspicious transaction and gain information it is possible to use intelligent multi-agent technology (MAT).

Literature brings many debates about the risk and several definitions. There are two interpretations - subjective probability and operationalism (both based on the same source of David Hume's empiricism). In a document on risk definition, Ngai et al. (2011) claims that two ingredients are needed for the risk - the first is the uncertainty of possible experimental results, and the second is that the results are material in nature. Financial institutions and other statutory bodies face two categories of risk: the first is the regulatory risk associated with the violation of rules on combating the legalization of criminal proceeds, and the second business risk (Parkitna et al., 2016) is that a liable entity must face the fact that unknowingly or otherwise, it can be providing services that can be associated with facilitating money laundering or terrorist financing. In addition, it is important to note that regulatory and business risks may overlap.

The risk-based approach is a systematic, continuous process of identifying and measuring the potential risk of money laundering and terrorist financing. At the same time, a strategy to mitigate these risks is emerging and being adapted, especially in the areas mostly at risk. An overview of methods used to detect money laundering based on a risk-based approach is listed by Hong et al. (2015).

4. Patent protection of procedures

Many excellent scientific results from the field of identifying unusual operations and money laundering are rather patented than published in scientific journals. Their publication would make it impossible to have them patented since their disclosure would be classified as a defect in novelty.
In some countries, computer algorithms that serve as bases for enhancing the functionality of computer software may be protected by a patent. In other countries, however, they are explicitly excluded from patentability and considered non-patentable. Software inventions may be patentable provided that the software provides a technical contribution to the state of technology. In most countries, the machine or source code of computer programs may be copyrighted. Although in some countries it is possible, or even necessary, this protection is not dependent on registration. Compared to patent protection, copyright protection is more limited because it only relates to the expression of ideas, not to thoughts themselves. Many companies protect the unit code of computer programs by copyright, while the source code is subject to business secrets.

USA, Canadian and other patent law allows the patenting of algorithms. The following examples are an example of patenting algorithms used to detect unusual business operations and in AML. An example of several patents follows:

Networked system for generating suggestions for exchanging foreign currency for credit in restricted account, has suggestion module for generating suggestion information for transaction locations and transmitting information to user device
Patent Number: US2016140555-A1
Patent Assignee: EBAY INC
Inventor(s): SCIPIONI G.

Computerized method for identifying accounts stored in database of computerized account management system, involves analyzing and assigning all accounts to per-jurisdiction buckets or default bucket, and no account is left unassigned
Patent Number: US2016104166-A1
Patent Assignee: MORGAN STANLEY
Inventor(s): COLE M; CHAN P; ENG K; et al.

RMB crown word number management system, has financial organization center for transmitting data to total center through bank intranet, and self-service equipment arranged on line type teller machine to process paper currency
Patent Number: CN104657818-A
Patent Assignee: SICHUAN JUNYI DIGITAL TECHNOLOGY DEV CO
Inventor(s): GOU J; ZENG L.

System for detecting unusual activity such as money laundering in cash vault transactions, determines dynamic threshold indicating whether change in proportion of large denomination currency transacted by customer is unusual
Patent Number: US2015142629-A1
Patent Assignee: BANK OF AMERICA CORP
Inventor(s): SUPLEE C; HUGHES C B; ZHOU J; et al.

Apparatus for generating graphical user interface for investigating e.g. illegal financial transaction of individual in financial institution, has module for generating interface that displays link to provide correlation between parties
Patent Number: US2015142627-A1
Patent Assignee: BANK OF AMERICA CORP
Inventor(s): LEE A.

Computer-implemented method for removing personally identifiable data, using fraud detection system, involves generating reverse hashing map and which explains how to restore identifiable information of set of hashed data
Conclusions

On the basis of literature assessment, it can be claimed that the current anti-money laundering research and work policies are primarily focused on the financial sector. Studies also highlight the need to have an international cooperation network, capacity building, enhancement of supervisory processes, and so on. However, in-depth studies of the main weaknesses are also required from a legislative and implementation point of view. It is also necessary to develop typologies and potential security measures using sophisticated mathematical and statistical methods.

References


1111


Short biographical note about the contributors at the end of the article (name, surname, academic title and scientific degree, duties, research interests):


ORCID ID: https://orcid.org/0000-0003-2384-9106

Ass. Prof. Ing. Miroslav GOMBÁR, PhD. is an associate professor in the Department of Management, Faculty of Management at the University of Prešov in Prešov since 2016. Since 2016, he works as head of the Department of Management, and he teaches school subjects: statistics, management, operations management, and logistics.

ORCID ID: https://orcid.org/0000-0002-8383-7820

Mgr. Pavel KELEMEN, Ph. D. Candidate at the Faculty of Management at the University of Prešov in Prešov, Slovak Republic

ORCID ID: https://orcid.org/0000-0001-7563-3142

JUDr. Stanislav BACKA, Ph. D. Candidate at the Faculty of Management at the University of Prešov in Prešov, Slovak Republic

ORCID ID: https://orcid.org/0000-0002-0411-4158

Copyright © 2019 by author(s) and Vsl Entrepreneurship and Sustainability Center
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/