3

The use of Big Data and Artificial Intelligence to prevent and detect fraud

(https://doi.org/10.47907/livro2021_4c4)

José Ricardo Marcondes Ramos[1]

Abstract:
The development and increased adoption of different IT trends is fostering the evolution and application of Big Data and Artificial Intelligence in many contexts of society and different business sectors. The financial services are the leading field in investment in advanced technology for continuous monitoring, transaction analysis, anomaly detection, data review, pattern recognition and other emerging techniques. In this article, we are going to analyse the role played by Big Data and artificial intelligence for fraud detection, analysing algorithms applied to prevent and detect payment frauds and the use of these techniques in the context of corporate fraud and investigations against financial statement frauds.

Keywords: big data; artificial intelligence; data mining; fraud detection; fraud prevention; forensic accounting.

[1] PhD student at the University of Coimbra. Univ. Coimbra, University of Coimbra Institute for Legal Research, Fac. Law - Collaborator.

I. INTRODUCTION

Ranging from the digital wallet to all sorts of smart gadgets oriented by sensors that perceive their surroundings[2] – such as autonomous vehicles, smart watches or fridges and everything that stands in between –, the development of new technologies is guiding the digital transformation of society, changing many aspects of life and introducing new ways of social interaction. Within the financial sector, for example, just as the development of ATMs (Automated Teller Machines) and online banking portals were revolutionary in the 1960s-1970s and 1990s, respectively, the creation of new technologies for financial transactions and the transfer of funds, such as crypto currencies and smart mobile payment systems (like Apple Pay, Samsung Pay, Google Pay and Amazon Pay, to name a few), are gradually replacing the need and the use of cash and facilitating the direct transfer of money and the purchase of goods and services[3].

As the rising use of electronic gadgets, due to this digital revolution, is increasing the production of digital information[4] (for instance, more than 98% of all information stored is currently electronic

---

[2] The ability to perceive the surrounding environment is usually associated with cloud-based systems that guide these gadgets with the use of the Internet of the Things (IoT), technology which "refers to the concept of connecting things like objects, people and animals, to the Internet using sensors that allow them to send and receive data in real-time". Bawack, R.E., Fosso Wamba, S. and Carillo, K.D.A. A framework for understanding artificial intelligence research: insights from practice. *Journal of Enterprise Information Management*, Vol. 34 No. 2, 2021, p. 653.

[3] Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies. *Forensic Science International: Digital Investigation*, Volume 33, 2020, p. 01-02.

[4] As outlined by Bernard Marr "[w]e have created more data in the past two years than in the entire previous history of mankind. By 2020, it is predicted that about 1.7 megabytes of new data will be created every second, for every human being on the planet. This data is coming not just from the tens of millions of messages and emails we send each other every second via email, WhatsApp, Facebook, Twitter, etc. but also from the one trillion digital photos we take each year and the increasing amounts of video data we generate (every single minute we currently upload about 300 hours of new video to YouTube and we share almost three million videos on Facebook). On top of that, we have data from all the sensors we are now surrounded by". Marr, Bernard. Big Data in Practice. How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. Chichester: Wiley, 2016, p. 02.

whereas this figure was near 25% in 2000[5]), the improvement in the ability to collect and store data as well as to analyse different types of digital data is fostering transformations "from the way banks and shops operate to the way we treat cancer and protect our world from terrorism"[6]. From the two advertising giants Google and Facebook, which developed a business model of targeted advertising using Big Data and artificial intelligence techniques applied to massive databases of personal data gathered from its platforms; to the Royal Bank of Scotland[7], which applied data analytics techniques to redesign its customer relationship, creating a more personal service (the so-called "personology" philosophy) based on enormous amounts of information about its clients, there are plenty of examples of the usage of new informational technologies to boost efficiency for both business and government organizations[8].

Alongside the expansion of electronic databases, the development and increased adoption of different IT trends like the Internet of the Things (IoT), business intelligence and analytics (BI&A), Big Data, cloud computing and Machine Learning (ML) is fostering the evolution and application of yet another disruptive technology, namely, Artificial Intelligence (AI)[9] [10]. Based on its unique abilities to perceive

---

[5] Rezaee, Z. and Wang, J. Relevance of big data to forensic accounting practice and education. *Managerial Auditing Journal*, Vol. 34 No. 3, 2019, p. 270.

[6] Marr, Bernard. Big Data in Practice. How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. *Cit.*, p. 01.

[7] As Bernard Marr describes, "RBS use data on their customers, including their account transactional history and personal information, to determine what products or services would be most useful" Marr, Bernard. Big Data in Practice. How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. *Cit.*, p. 84.

[8] As noted by Martin Fleming, VP and Chief Economist at IBM, "AI technology has the potential to increase the productivity of workers as well as productivity in all walks of life". Wilson, C. (2019), "IBM Tech trends to watch in 2020 . . . and beyond", IBM.

[9] Bawack, R.E., Fosso Wamba, S. and Carillo, K.D.A. A framework for understanding artificial intelligence research: insights from practice. *Cit.,* p. 645 and 655-657.

Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies.. *Cit.,* p. 01-02.

[10] There are several different perspectives to understand artificial intelligence, such as a *field of study* in which AI is perceived "as the branch of knowledge that investigates the possibility of giving human intelligence capabilities to nonhuman entities", a *concept* that understands it as "an abstract concept that corresponds to any manifestation of human intelligence by machines or technology", an *ability* by which

the environment, learn from experience, understand intention and context and take appropriate action with autonomous decisions[11] – carried out using complex algorithms that recognize patterns, understand written and spoken words, identify images and make predictions and recommendations – AI is one of the main technological and strategic trends in the digital transformation of business and society today, being the main responsible for business automation processes[12].

Although its consolidated adoption is so far limited to a handful of industries such as financial services, healthcare, marketing and fraud detection, there are an increasing number of studies and experiments designed to test the application of artificial intelligence in fields like education, telecommunication, transportation, automotive, energy and so on[13]. In this article, we will analyse the use of artificial intelligence in the combat of financial fraud, with a focus on algorithms used to prevent the occurrence of payment frauds and money laundering as well as to detect financial statement fraud in the context of forensic accounting. In order to do so, firstly, we are going to analyse the role played by digital tools, Big Data and artificial intelligence in the context of forensic investigations. Thus, our focus will be moved for fraud detection algorithms applied to prevent and detect payment frauds, notably, credit card frauds and, finally, the use of these techniques in the context of corporate fraud and investigations.

II. FROM FORENSIC SCIENCE AND DIGITAL FORENSICS TO ARTIFICIAL INTELLIGENCE AND BIG DATA APPLIED TO FRAUD DETECTION AND PREVENTION

Just as new technologies are facilitating and transforming several aspects of life interaction, the emergence of new technological trends

---

"AI is a skill given to a technology artifact for it to behave like an intelligent human being" and even as a *system*, perspective that sees AI as a set of technologies that "can perceive, learn, reason, assist in decision-making and solve problems in ways like humans". Bawack, R.E., Fosso Wamba, S. and Carillo, K.D.A. A framework for understanding artificial intelligence research: insights from practice. *Cit.,* p. 651-652.

[11] Despite the different points of view to understand artificial intelligence, the four capabilities of perception, comprehension, learning and acting are seen by practitioners as the main features that characterize Artificial Intelligence. Idem, p. 651.

[12] Idem, p. 652.

[13] Idem, p. 658.

is also being used to leverage both old and new criminal behaviours[14]. Either by tampering or meddling with technical and technological mechanisms (such as hardware, software, network protocols and cryptography) or by smoothing social engineering schemes in order to exploit personal weaknesses and enable fraud[15], the development of new operational techniques based on technological advances and emerging gadgets is aiding the occurrence of different forms of fraud, money laundering and other underground criminal activity[16].

If, however, technology may help crime be committed, it is also evolving to improve old investigative methods as well as to develop new ones[17]: first, the growing essentiality of digital gadgets (be it computers, smartphones, tablets or others) has led to the development of a whole new field of digital forensics[18] specialized in the extraction

---

[14] As Richard Bolton and David Hand argue "in recent years, the development of new technologies (which have made it easier for us to communicate and helped increase our spending power) has also provided yet further ways in which criminals may commit fraud. Traditional forms of fraudulent behavior such as money laundering have become easier to perpetrate and have been joined by new kinds of fraud such as mobile telecommunications fraud and computer intrusion". Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review, *Statistical Science*, Vol. 17, No. 3 (Aug., 2002), Institute of Mathematical Statistics, p. 235.

[15] Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. PRECEPT: a framework for ethical digital forensics investigations. *Journal of Intellectual Capital*, 2020, Vol. 21 No. 2, pp. 259. Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies. *Cit.,* p. 06-07. van Beek, H.M.A.; van den Bos, J.; Boztas, A.; van Eijk, E. J.; Schramp, R.; Ugen, M. Digital forensics as a service: Stepping up the game. *Forensic Science International: Digital Investigation*, Volume 35, 2020, p. 01.

[16] As Sunger Gee points out, "[n]ew technology allows for new, more convenient payment methods for consumers and also provides new opportunities for money laundering ... [In addition,] criminals can mix the old with the new to move money to further reduce the risks of detection". Gee, Sunder. Fraud and fraud detection: a data analytics approach. New Jersey: Ed. Wiley, 2015, p. 257.

[17] Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology. *Accounting Research Journal*, 28(1), p. 06.

[18] As a branch of forensic science, digital forensics is responsible for the process of identification, collection, processing and interpretation of digital data from any given device and, as such, can be understood as "the process of applying scientific methods to analyze stored information and to determine the events of a particular incident, thus making evidence usable in court". Oliveira Júnior, Edson; Zorzo, Avelino F.; Neu, Charles Varlei (2020). Towards a conceptual model for promoting digital forensics experiments. *Forensic Science International: Digital Investigation*, 35(), 301014, p. 01. It is important to recognize, though, that digital investigation's

and analysis of data produced, stored and processed within these devices[19] – be it software, hardware or a combination of both[20]. Second, the improvement of data extraction and analysis capacity, enabled by new technological advances such as data mining and data analysis techniques, is helping to upgrade crime-related forensic methods in diverse areas such as neurocriminology, handwriting analysis and forensic accounting[21].

While the increasing adoption of new information and communication technology by society is fostering new investigative sources and techniques for traditional crimes[22], the enormous amount of digital data produced by new digital devices is fuelling the development of digital forensics as a new and independent field[23] specialized in the

---

appliance is not restricted to judicial controversies, also being commonly used in the corporate ecosystem as a preventive and investigative tool related to behavioral and disciplinary concerns. However, even recognizing that the digital forensics has several applications within the legal frameworks – namely, public sector security and operation as well as corporate investigations – it is essential to keep in mind that the "main purpose of digital evidence is to support or rebut a thesis or argument on which court decision is based on". V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. *2020 43rd International Convention on Information, Communication and Electronic Technology* (MIPRO), 2020, p. 2096.

[19] van Baar, R.B.; van Beek, H.M.A.; van Eijk, E.J. (2014). Digital Forensics as a Service: A game changer. *Digital Investigation*, 2014, 11, p. 54.

[20] V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. *Cit.,* p. 2095. Wu, Tina; Breitinger, Frank; O'Shaughnessy, Stephen. Digital forensic tools: Recent advances and enhancing the status quo. *Forensic Science International: Digital Investigation*, 2020, 34, p. 04. Netherlands Register of Court Experts NRGD, 2016. Standards 008.0 Digital Forensics. Technical Report Netherlands Register of Court Experts, p. 06.

[21] Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology. *Cit.*, p. 07-08.

[22] Illustrative examples are the use of Google searches and other online activities to prove premeditation and, in a more concrete stance, the extraction of information from the Apple Watch app to unveil the disappearance and assassination of Saudi dissident Jamal Khashoggi in Turkey. Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. PRECEPT: a framework for ethical digital forensics investigations. *Cit.*, p. 260. Lee, Jae-Ung; Soh, Woo-Young. Comparative analysis on integrated digital forensic tools for digital forensic investigation. *IOP Conference Series: Materials Science and Engineering*, 2020, 834, 012034, p. 01.

[23] As Ferguson *et al* explain "the field of digital forensics, though relatively young, has earned the right to call itself a discipline, and that law enforcement and

understanding of how this data is produced and how it can be collected and analysed[24]. Usually, crimes and felonies involving different sorts of technologies are very technical in nature[25], which implies different types of analysis of hardware, software systems, malware, network protocols, APIs and cryptography[26]. Even though there are different criteria for classifying digital forensics tools, the diversity of data sources and the need for expertise on the underlying technology is the base for its taxonomy, which separates the digital forensics in different sub-fields such as computer forensics, software forensics, multimedia forensics, device forensics, network forensics, malware forensics and memory forensics[27].

Regarding the old investigative methodology, on the other hand, ever since the introduction of the fingerprinting method (the first

---

educational institutions are developing training to ensure that effective investigations can indeed be carried out in the digital world to support law enforcement". Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. PRECEPT: a framework for ethical digital forensics investigations. *Cit.,* p. 260.

[24] van Beek, H.M.A.; van den Bos, J.; Boztas, A.; van Eijk, E. J.; Schramp, R.; Ugen, M. Digital forensics as a service: Stepping up the game. *Cit.,* p. 01. van Baar, R.B.; van Beek, H.M.A.; van Eijk, E.J. (2014). Digital Forensics as a Service: A game changer. *Digital Investigation*, 2014, 11, p. 54.

[25] As Bruce Nikkel explains, there are several kinds of different criminal activities exploiting different technological bases. Considering solely financial frauds, there are felonies raging from phishing, attacks against ATMs and payment card terminals, online banking trojans, rogue mobile banking apps, extortion and ransom attacks, online social engineering attacks, online money laundering and others. Once the crime could be committed throughout different technological means, the investigative process may vary. A good example is the practice of phishing, tricking people into giving personal or financial information through "spoofed" messages, which could be committed by SMS (smishing), voice (vishing), twitter (twishing). Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies. *Cit.,* p. 03-05.

[26] Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies. *Cit.*, p. 06.

[27] This taxonomy is proposed by Tina Wu *et al* (2020) as an updated version of the distinction made by the Netherlands Register of Court Experts (NRGD, 2016. Standards 008.0 Digital Forensics. Technical Report Netherlands Register of Court Experts, p. 08). As the authors describe, their version has two central differences: firstly, "due to the lack of available database forensic tools", the data base sub-field is placed under the software category; and, secondly, the taxonomy was extended to include the categories of malware and memory forensics. Wu, Tina; Breitinger, Frank; O'Shaughnessy, Stephen (2020). Digital forensic tools: Recent advances and enhancing the status quo. *Cit.,* p. 04.

significant investigation technique), forensic science[28] developed a series of scientific processes in fields ranging from biology, chemistry and physics to anthropology and accounting to assist investigators identify and enquire information and objects related to a crime scene (a mute witness of the crime) and collect relevant evidence such as stains, hair or DNA samples, soil and so on[29]. Although several of the forensic investigation techniques and tools are not particularly new – for instance, the creation of the polygraph machine, which set the standard for lie-detection methodology, dates back to the 1880s[30] – it was not until the end of the 20th century that the use of computers to perform investigative tasks contributed to the development of the digital forensics field[31]. Despite being used to perform, with enhanced capacity, traditional forensic tasks like fingerprinting, hair or DNA analysis and even autopsies, the development of the digital forensics field was highly influenced both by the ubiquitous adoption of new digital devices and the rising incidence of cybercrime[32].

As a scientific field, the development of new applied research[33] and new technologies helped forensics science to improve its methods

---

[28] The field of forensic science emerged and was developed due to the difficulty of unveiling the circumstances in which a crime may have occurred and to overcome the excessive reliance on confessions or witness testimony to identify the offender. With this goal, the field of forensic science developed a series of techniques and methods to aid the investigative process by acquiring, analysing and interpreting evidence through a coordinated process in order to base scientifically investigative conclusions. V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. *Cit.*, p. 2094.

[29] V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. *Cit.*, p. 2094.

[30] Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology. *Cit.,* p. 07.

[31] Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. PRECEPT: a framework for ethical digital forensics investigations. *Cit.*, p. 259. V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. *Cit.*, p. 2094-2095.

[32] Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. PRECEPT: a framework for ethical digital forensics investigations. *Cit.*, p. 259.

[33] As Tina Wu *et al* describe, "[c]ompared to other domains, the digital forensics community has a very applied focus, meaning that we are not solving problems in theory but practically. Consequently, research endeavors frequently come with prototype implementations". Wu, Tina; Breitinger, Frank; O'Shaughnessy, Stephen. Digital forensic tools: Recent advances and enhancing the status quo. *Cit.,* p. 04.

and create advances in its techniques for interviewing and interrogation, handwriting analysis, data analysis and others[34]. With respect to the interviewing and interrogation processes, for example, lie-detection techniques previously based on alterations in breathing, blood pressure, pulse rate and sweat, measured by the polygraph, can now be performed by neurocriminology instruments, which identify whether someone is lying or telling the truth based on neural mapping and the areas of the brain displayed as active when the person is confronted with evidences of the crime[35]. Similarly, handwriting analysis, which used to be done personally by experts, can now be performed by algorithms that examine features such as pen pressure and letter dimensions[36].

Finally, the development of new technological advances is also revolutionizing audit and forensic accounting practices and, as a consequence, not only innovative techniques, such as word mapping software that identify bribery-related terms, are being used to fight corruption[37], but the emergence of new data processing capacities and statistical tools for fraud detection is also improving the fight against financial crimes, corporate fraud and money laundering[38] – topics that will be analysed next. As described by Michael Young[39],

> Forensic computers can be deployed to look for fraud. Based on years of accumulated experience, savvy forensic accountants at the big accounting and consulting firms have developed computerized searching tools that, once plugged into a company's general ledger system, will at high speed start combing through thousands of entries and kicking out those that for any number of reasons look unusual or suspicious.

---

[34] Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology.*Cit.,* p. 07.

[35] Idem, p. 07-08. The author also adds that "[u]sing similar technology, a recent study of New Mexico inmates using brain scans correctly predicted which prisoners were more likely to commit another crime once released".

[36] Idem, p. 07.

[37] Idem, p. 08.

[38] As Richard Bolton *et al* describes, "[p]rocessing these data sets in a search for fraudulent transactions or calls requires more than mere novelty of statistical model, and also needs fast and efficient algorithms: data mining techniques are relevant". Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 236.

[39] Young, Michael R. *Financial Fraud Prevention and Detection. Governance and Effective Practices*. New Jersey: Wiley, 2014, p. 169.

III. THE APPLICATION OF BIG DATA AND ARTIFICIAL INTELLIGENCE TO DETECT AND PREVENT FINANCIAL FRAUD

The expansion of digital data and the increasing importance of electronic databases for economic development and strategic management are reshaping many aspects of society as well as several different business sectors – so much so, that the market of Big Data has grown dramatically from U$16,1 billion in 2014 to more than U$50 billion by the end of 2016[40]. Despite its rising adoption in a growing number of industries, though, when it comes to the fight against fraud, the financial services sector is the leading field in investment in advanced technology for continuous monitoring, transaction analysis, anomaly detection, data review, pattern recognition and other emerging techniques[41].

This scenario is justified by several reasons such as the better cost and operational efficiency that data-driven or statistically based fraud-detection methodologies present, or its greater precision and improved detection power when compared to the classic human-driven approach[42]. However, aside from the fact that data analysis allows the entire database of financial transactions or journal entries to be tested instead of a selected sample[43], perhaps the main reason may be the essentiality of Big Data and analytics techniques when dealing with huge amounts of dynamic and constantly evolving data[44], after all "by processing massive volumes of information, fraud patterns may be uncovered that are not sufficiently apparent to the human eye"[45].

---

[40] Rezaee, Z. and Wang, J. Relevance of big data to forensic accounting practice and education. *Cit.,* p. 270.

[41] Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Managerial Auditing Journal*, Vol. 34 No. 5, 2019, p. 602-622.

[42] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. A Guide to Data Science for Fraud Detection. New Jersey: Wiley, 2015, p. 17-18.

[43] Gee, Sunder. Fraud and fraud detection: a data analytics approach. *Cit.*, p. 67.

[44] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 236.

[45] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. A Guide to Data Science for Fraud Detection. *Cit.*, p. 17.

As a rule, the use of these techniques is separated into two different stages: *data mining*, when the computer uses artificial intelligence, neural network techniques and advanced statistical tools (e.g. cluster analysis) to perform searches on large amounts of data with the specific goal of finding trends, patterns and relationships within the dataset, but without testing any pre-established hypothesis; and, *data analysis*, when the evaluation of each component of the data has the purpose of testing a hypothesis to be confirmed or dismissed, reaching a conclusion based on the inference from the findings[46]. The data analytics process may be subsequently separated into three other steps of *exploratory data analysis* (EDA) in which very little in known about the data's relationships where "hypotheses are formed and new patterns of features of the data are discovered"; *confirmatory data analysis* (CDA), when "testing takes place and the hypotheses are proven correct or false"; and *qualitative data analysis* (QDA), stage "used to draw conclusions from non quantitative or non-numerical data such as images or text"[47].

The hypotheses tested by the data analysis techniques, in turn, are developed using two types of methods, namely, *supervised* ones, in which selected samples of behaviours labelled as both fraudulent and non-fraudulent are applied to train algorithms that assign a suspicion score to evaluated cases; and *unsupervised methods*[48], in which the algorithm is trained with a baseline of what represents the normal behaviour and focus on detecting anomalies outlining observations that depart from this norm[49]. Once supervised methods learn with historical observation from which they extract patterns of fraudulent and non-fraudulent behaviour, it is mostly applied when there is a wide and complete dataset about each type of fraud[50]. This, however, requires not only that the training set be composed of both classes of cases, but also that the labelling of each is reliable and balanced,

---

[46] Gee, Sunder. Fraud and fraud detection: a data analytics approach. *Cit.*, p. 10-11.

[47] Idem.

[48] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 19-20.

[49] Bhattacharyya, Siddhartha; Jha, Sanjeev; Tharakunnel, Kurian; Westland, J Christopher. Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50, (3), 2011, p. 602.

[50] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review.*Cit.*, p. 237.

avoiding a biased algorithm caused either by over- or under-sampling and, thus, reducing the occurrence of false negatives and positives[51].

Furthermore, besides the lack of testable and open access data[52] (considering not all victims of fraud publicly disclosure this type of information) and the limited exchange of information regarding fraud detection methods (notably because "it does not make sense to describe fraud detection techniques in great detail in the public domain, as this gives criminals the information that they require to evade detection"[53]), the use of supervised methods faces yet another problem: the dynamic character of fraud, which makes it unable to detect new types of fraudulent behaviour or frauds that uses unknown mechanisms or methods[54]. As fraud detection algorithms evolve, so do fraudsters, adapting their approaches and strategies with inventive and refined new methods to make the fraudulent behaviour less apparent and detectable by upgraded algorithms[55]. Also, because there are still new criminals trying old methods (sometimes unaware of the consolidated detection techniques), the latest upgrades on the algorithms should be applied jointly with earlier tools[56].

More than showing that the growing availability of fraud data is an important driver for the improvement of fraud prevention and detection techniques, this dynamic character of fraud and the need to reduce the pernicious consequences of new frauds evidence the importance of regularly updating the algorithms throughout the fraud cycle and its four steps of fraud detection, investigation, confirmation and

---

[51] Beasens, Bart. Analytics in a Big Data World. The Essential Guide to Data Science and its Applications. New Jersey: Wiley, 2014, p. 165-166.

[52] Shiguo Wang goes further and argues that "[t]here are two kinds of critical suggestions concerning applying data mining technology on detecting fraud. One is lack of testable, open accessible data. The other is lack of mature methods and technologies". Wang, Shiguo. A comprehensive survey of data mining-based accounting-fraud detection research. *2010 International Conference on Intelligent Computation Technology and Automation*, ICICTA 2010, 1, 50.

[53] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 236.

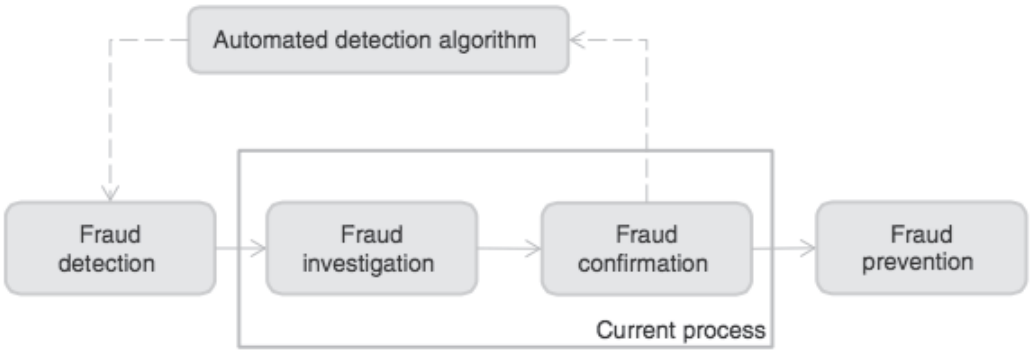
[54] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 237.

[55] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 18.

[56] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 236.

prevention (as shown in Figure 1)[57]. Although the required frequency for retraining the algorithm is influenced by several factors such as the volatility of the unknown fraud behaviour, the rate at which new cases are confirmed, the detection power of the existing model and the cost-benefit relation associated with the upgrading process, the main element for this feedback loop is the correct and reliable labelling of cases as fraudulent in a careful *ex post* analysis[58]. Besides, these limitations on the supervised methods also "illustrates the complementarity of supervised and unsupervised methods and motivates the use of both types of methods as complementary tools in developing a powerful fraud-detection and prevention system"[59].

Figure 1. The fraud cycle



Source: Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. **Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques.** *Cit.,* p. 23.

Alongside the supervised and unsupervised methods, there is a third complementary tool used to enhance the abilities for fraud detection, known as social network analysis. Based on researches showing that fraudsters seldom act in an isolated fashion and are usually highly interconnected with other fraudulent individuals and companies[60], this method creates a so-called spider construction (Figure 2) in order to analyse network-related information and identify potentially suspicious activities[61].

---

[57] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 11-12.

[58] Idem, p. 23.

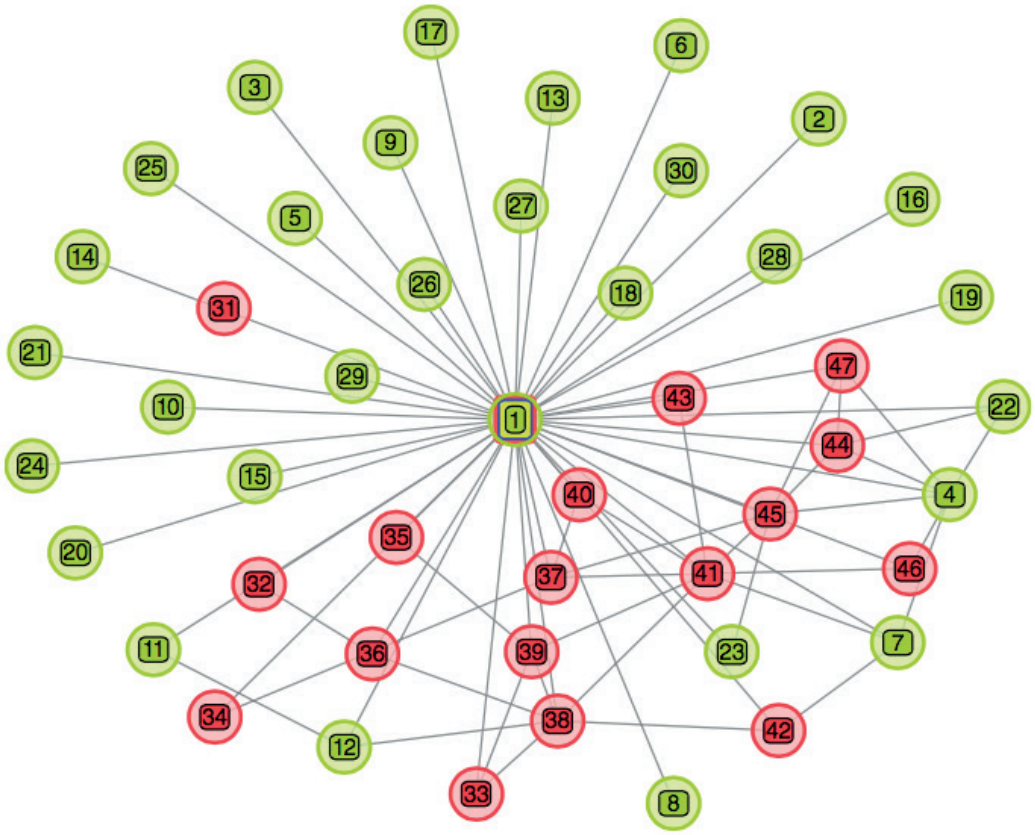
[59] Idem, p. 21.

[60] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 237-238.

[61] Rezaee, Z. and Wang, J. Relevance of big data to forensic accounting practice and education. *Cit.,* p. 271. Beasens, Bart. Analytics in a Big Data World. The Essential Guide to Data Science and its Applications. *Cit.,* p. 166-167.

For instance, within the analysis of banking accounts, by examining variables such as the *fraudulent degree*, obtained by analysing the number of immediate contacts a node has and its number of direct fraudulent connections; the *triangles* it belongs to (that is, structure of three nodes connected to each other) and the potential of being a fraudulent node by integrating with a fraudulent triangle (after all "nodes that are involved in many suspicious triangles have a higher probability to commit fraud themselves"[62]); and the *cliques*, or extensions of triangles that may indicate undirected connections with other fraudsters, are important techniques used to identify money laundering and complex structure for carousel fraud[63].

Figure 2. Spider constructions of social network analysis



Green nodes represent legitimate individuals, while red ones represent fraud.

Source: Beasens, Bart. Analytics in a Big Data World. The Essential Guide to Data Science and its Applications. *Cit.,* p. 168.

---

[62] Idem, p. 168.

[63] Idem, p. 167-168. Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 04.

Even recognising that the identification and analysis of the agents involved in a specific group of financial transactions play a key role in the fight against frauds and money laundering, it is important to mention that the connection with other parties and potential fraudsters is not the only type of linked analysis performed by fraud detection algorithms. As such, there are also several levels of analysis that could be made (sometimes simultaneously) to reconstruct patterns of transactions and distinguish legitimate sets from illegitimate ones, including the participants engaged ("an obvious and simplistic illustration is the fact that a transaction with a known criminal may rouse suspicion"[64]), the individual transaction or its association with other sets of transactions ("a single deposit of just under $10,000 is not suspicious, but multiple such deposits are; a large sum being deposited is not suspicious, but a large sum being deposited and instantly withdrawn is"[65]) and even the geographical location of either the origination or destination of the funds, which orients rules such as "flag transactions from countries X and Y", based on international lists of countries considered to be at high risk of money laundering or that have some form of connection with terrorism[66].

It is important to mention, though, that these three fraud detection techniques are not mutually exclusive and, once each one focuses on a different aspect of fraud, are usually complementary. As a matter of fact, because each of these methods has different capacities and limitations, the development of an effective mechanism to prevent and detect different types of financial fraud is commonly based on the combination of the three, reinforcing each other's strength and compensating for its vulnerabilities[67]. Notably, however, the selection

---

[64] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 241.

[65] Idem.

[66] The evaluation of jurisdictions and their compliance to international standards related to the combat of money laundering and terrorist financing (AML/CFT) is made by the Financial Action Task Force (FAFT), which releases, three times a year, two lists containing the countries, first, with weak AML/CFT regimes and, second, under increased monitoring, that is, countries working with the FAFT to address their strategic deficiencies. Both lists can be accessed at https://www.fatf-gafi.org/publications/high-risk-and-other-monitored-jurisdictions/?hf=10&b=0&s=desc(fatf_releasedate).

[67] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 22.

of the ideal fraud prevention and detection system and the development of the most adequate algorithm depends on the type of crime to be avoided. In general, the solutions currently existing in the market aim at two different types of misconduct associated, on the one hand, with illegal financial flows in crimes such as credit card fraud, financial identity scams, money laundering and terrorist financing and, on the other hand, with different forms of financial fraud, which range from internal corporate crimes, such as money embezzlement and reckless management, to financial statement fraud in the form of market manipulation and investment fraud. Once each form of fraud assumes a different *modus operandi*, the recommended algorithm to detect and prevent it vary based on expertise and technique required, a topic that we shall address next.

IV. THE USE OF ARTIFICIAL INTELLIGENCE TO FIGHT PAYMENT FRAUDS, MONEY LAUNDERING AND TERRORIST FINANCING

Among other serious problems emerging in the financial sector, two main issues with major harmful consequences for the economy are located within the capital flows system: firstly, the occurrence of payment frauds, in the form of credit card fraud, account takeover and other scams; and the incidence of illegal financial flows, namely money laundering and terrorist financing. Regarding payment frauds, for example, in its latest report the European Central Bank points out that worldwide €1.80 billion was lost in 2018 alone, with a total of €0.94 billion defrauded from credit cards issued in the euro area[68]. As a fraudulent behaviour, credit card fraud integrates a larger form of financial crime known as *financial identity scams* in which a fraudster reaches for personal information of a victim in order to perform fraudulent money transfers or payments[69].

---

[68] European Central Bank. Sixth report on card fraud. August 2020, p. 02.

[69] As Arjan Reurink explains, there are several terms used to describe financial identity scams such as "identity crime", "identity theft", "identity fraud", "credit card fraud" and "payment fraud". Still according to the author, among this fraudulent behaviour it is possible to distinguish three main types of fraud, namely "financial identity theft, which entails the use of personal identifying information to establish credit lines in the name of the victim; criminal identity theft, which involves a criminal giving

With a similar *modus operandi* of, first, obtaining the client's identifying information – either using *technical subterfuge schemes* by installing malware and malicious software in digital devices (technique known as *pharming*[70]) or *social engineering schemes* for deceiving the victims into giving their information away (such as *phishing*[71]) or allowing someone to obtain it (using techniques like "skimming" or "shoulder surfing"[72]) – and, second, using it to realize financial gain, credit card fraud and account takeover alike can be classified essentially in two types: application and behavioural fraud[73]. While application

---

another person's identifying information to law enforcement; and identity cloning, whereby imposters, illegal immigrants, or wanted felons use the victim's personal information to establish a new life". Reurink, Arjan. Financial fraud: A literature review. MPIfG Discussion Paper, No. 16/5, Max Planck Institute for the Study of Societies, Cologne, 2016, p. 47.

70 Arjan Reurink explains that technical subterfuge schemes "are more technical in nature and rely much less on persuasion to entice victims into the scheme, which enables a much wider victim base". Regarding the pharming technique, for example, "fraudsters send out e-mails which, when opened, plant malware – malicious software – in the victims' personal computers. The malware then directs traffic from those PCs that is destined for a legitimate website, say, a bank, to the pharmer's bogus website, which looks just like the real one. Without the victim's knowledge or consent, all the information the victim thinks is being sent to the bank's website is sent directly to the pharmer. Another possible mode of operation for pharmers is to alter a website's internet protocol (IP) address in the domain name server (DNS). In so doing, pharmers redirect all users who type in the URL (the web address) of, say, a bank to the illegitimate website controlled by the pharmer". Idem, p. 48.

71 "In a typical phishing attack, a scam artist pretending to be an agent from a bank or credit card company sends out e-mails to customers in which the operator prompts them to click on a hyperlink that brings them to a website, controlled by the phisher, where they will be asked to further process their account details. To appear credible and to trick the recipient's into participating in the scheme, the scam artist's e-mails contain company logos and use scare tactics – such as threats of account closure – and urgency cues that short-circuit victims' elaboration on clues that could reveal the deceptive nature of the invitation". Idem, p. 48.

72 According to Richard Bolton and David Hand "skimming" is a technique "where employees illegally copy the magnetic strip on a credit card by swiping it through a small handheld card reader", while "shoulder surfers" are fraudsters "who enter card details into a mobile phone while standing behind a purchaser in a queue". Alongside there fraudulent techniques, the authors also mention yet another in which "people posing as credit card company employees taking details of credit card transactions from companies over the phone". Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 238-239.

73 Bhattacharyya, Siddhartha; Jha, Sanjeev; Tharakunnel, Kurian; Westland, J Christopher. Data mining for credit card fraud: A comparative study. *Cit.,* p. 603.

fraud occurs when a fraudster uses false information or other people's information to obtain a credit card, behaviour fraud is characterized by the fraudulent use of other people's credit card or another payment means without its knowledge and approval. This last type of fraud, is separated into four types: mail theft, in which credit cards are intercepted before reaching the cardholder; stolen or lost card; counterfeit card and "card holder not present" fraud[74]. Similarly, frauds involving bank accounts range from the *account takeovers*, in which a fraudster takes control over an existing account and extracts its balance (sometimes even creating additional accounts and using the victims' credit lines); to *fictitious identity fraud*, where pieces of real information are combined to fabricate a fake identity in order to defraud the banking institution by establishing credit lines and stealing the money[75].

In the literature, there is widespread agreement that these types of payment fraud have several harmful consequences that affect directly and indirectly three groups of victims consisting of the *consumers* and *businesses* that had their financial identity stolen and credit cards/bank accounts misused; *merchants* and *credit providers* tricked into giving credit, money and goods to scammers; and, finally, *banks*, *credit card companies* and *e-retailers* whose brands were associated with these felonies and may need to review and reform their cyber security programmes and policies[76]. Due to its pernicious consequences, thus, it is in all stakeholder's interests to avoid the occurrence of payment fraud or, at least, to detect it as soon as possible in order to reduce direct losses and preserve the confidence in the whole payment system[77].

Because a typical credit card transaction has abundant data availability by recording hundreds of characteristics that describe each transaction in detail, credit card companies are not only among the early adopters of Big Data approaches, but also of the use of artificial intelligence algorithms, one of the main weapons at their disposal[78].

---

74 Idem, p. 603. Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 238-239.

75 Reurink, Arjan. Financial fraud: A literature review. *Cit.*, p. 48-49.

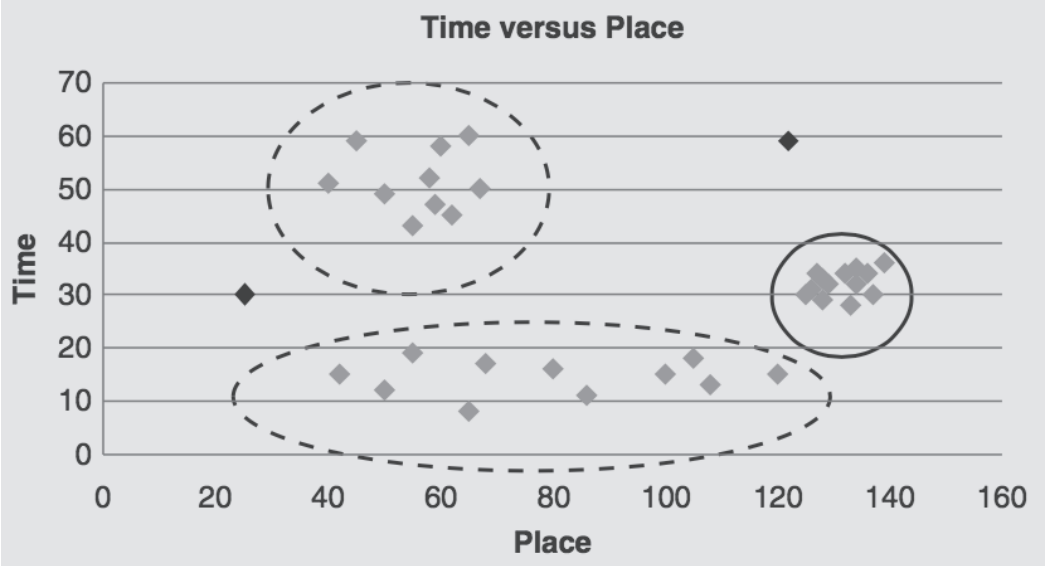
76 Idem, p. 50.

77 Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 238.

78 Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 24.

By applying different data mining techniques like neural networks, random forests or logistic regression, the predictive models trained to detect credit card fraud analyses a broad variety of information attributed either to the transaction itself or the parties engaged in it. This includes both numerical attributes, like its amount, and categorical attributes, as the example of the merchant code and name, the date of the transaction and its type or its geographical location, among others[79]. Using these pieces of information, the algorithms use descriptive analytical methods such as *outlier detection techniques* in order to identify abnormal or anomalous behaviours that might indicate suspicious activities[80]. As Figures 3.1 and 3.2 show, by pinpointing transactions that deviate from the clusters of regular and frequently occurring pattern, it is possible to detect outliers that do not comply with the overall behaviour and, hence, flag a transaction for further human investigation[81].

Figure 3.1. Outlier detection at the data item level

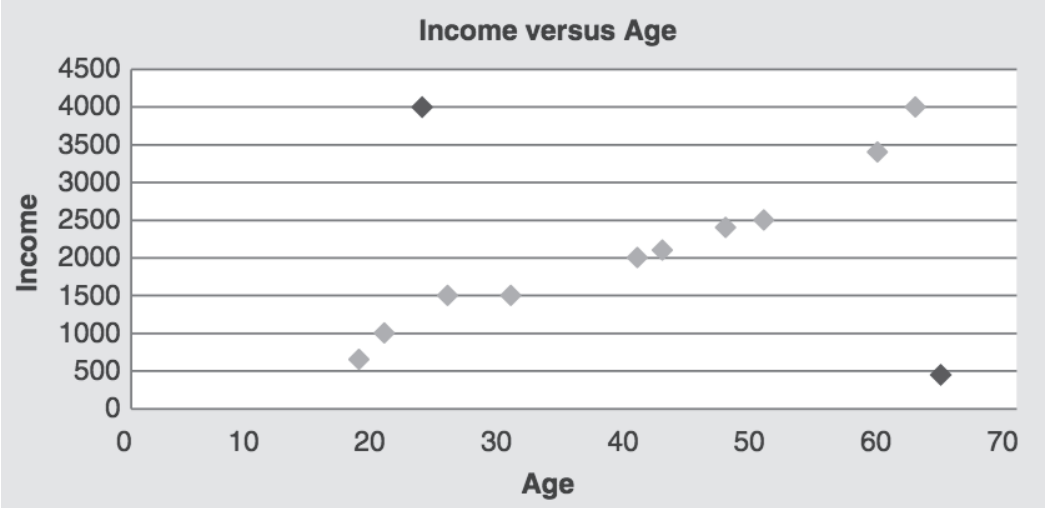


[79] Bhattacharyya, Siddhartha; Jha, Sanjeev; Tharakunnel, Kurian; Westland, J Christopher. Data mining for credit card fraud: A comparative study. *Cit.,* p. 603-604.

[80] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 24.

[81] Idem, p. 24-25. Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Cit.,* p. 617.

Figure 3.2. Outlier detection at the data set level



Source: Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques*. *Cit.,* p. 25.

The standards of normality used as benchmark to analyse transactions, in turn, are obtained by the application of a wide range of statistical tools, machine learning methods and data mining techniques which examine and identify both individual patterns of previous usage as well as general patterns of use and consumption according to the type of establishment, personal profile, age, location, etc.[82] (Figures 4.1 and 4.2). In addition, these predictive analytics tools also use algorithms trained to identify transaction patterns known to be intrinsically suspicious as the example of small purchases followed by big ones[83], a large number of online transactions made in a short period of time, the immediate use of a new card in a wide range of different locations as quickly as possible[84],

---

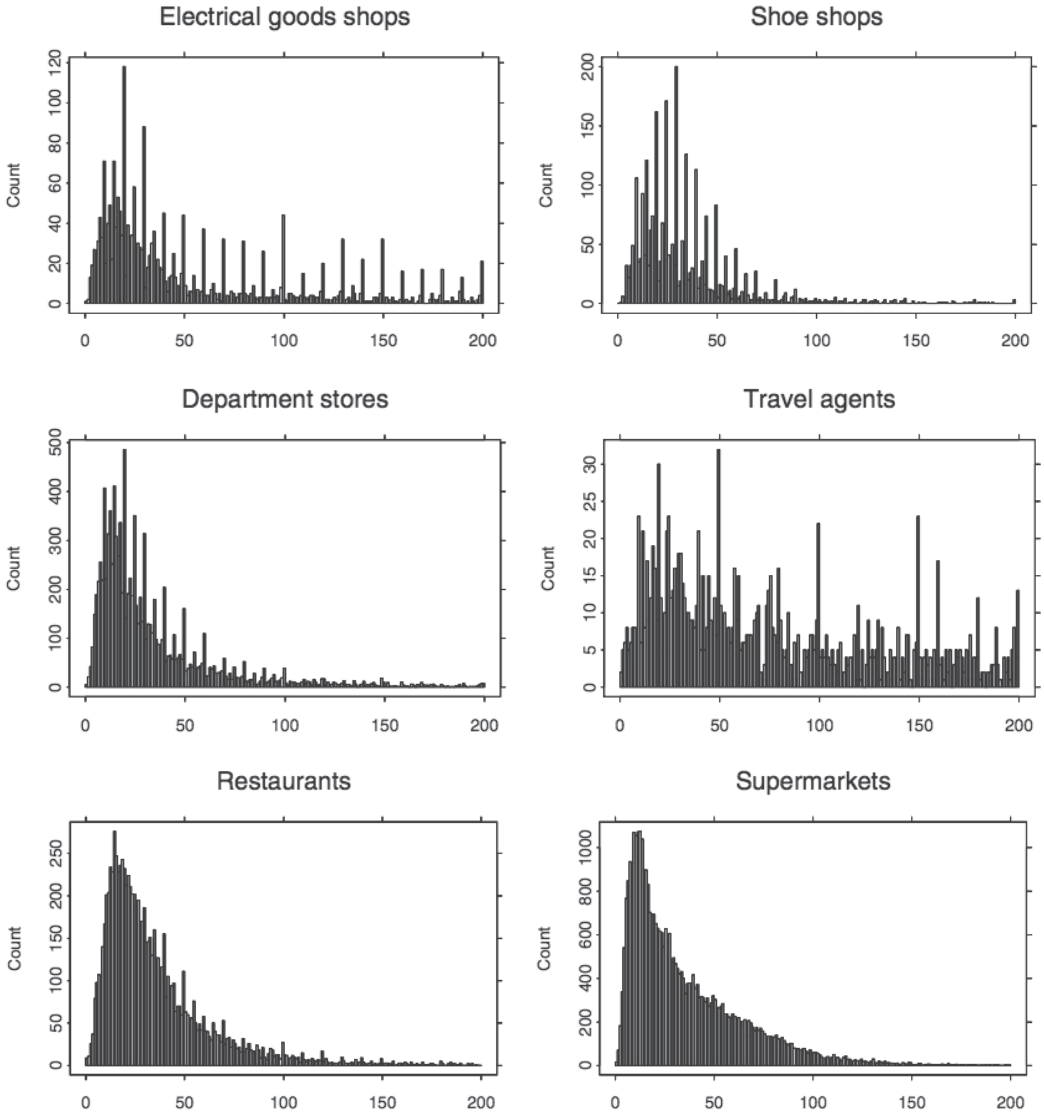
[82] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 239.

[83] Regarding this pattern, it is worth mentioning that "credit card fraudsters often try out a stolen credit card for a low amount to see whether it works, before making a big purchase, resulting in a recent and low monetary value transaction followed by a recent and high monetary value transaction" Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 39.

[84] As Bhattacharyya *et al* explains, "past research suggests that fraudsters try to maximize spending within short periods before frauds get detected and cards are withdrawn". Bhattacharyya, Siddhartha; Jha, Sanjeev; Tharakunnel, Kurian; Westland, J Christopher. Data mining for credit card fraud: A comparative study. *Cit.,* p. 603.

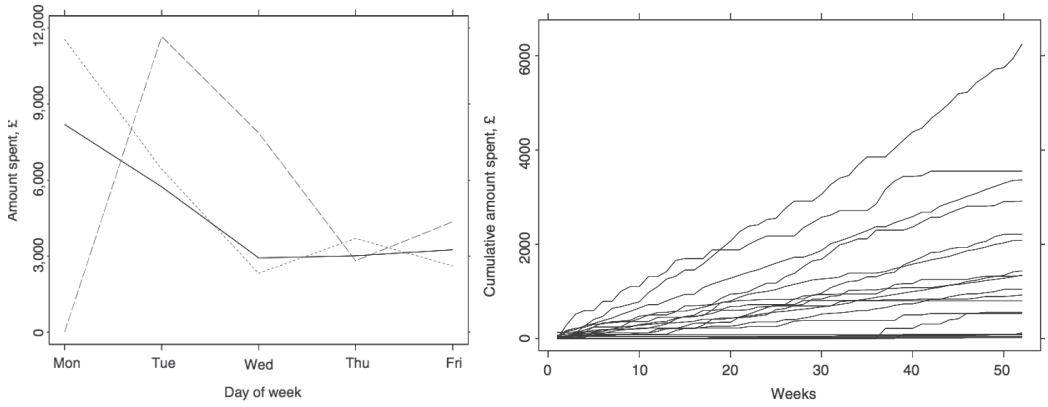
the sudden purchase of numerous goods of high value and that can be easily resold on the black market (namely jewellery or electronic devices), and so on[85].

Figure 4.1. Transaction size distributions for selected trade sectors



[85] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 239. Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 25-26.

Figure 4.2. Transaction patterns for supermarkets



Source: Hand, D. J. and Blunt, G. Prospecting for gems in credit card data. *IMA Journal of Management Mathematics*, 2001, n. 12, pgs. 181, 182 and 185.

Among the types of data used to feed fraud detection algorithms, an important type of aggregated transactional information are the so-called RMF variables, that identify the *recency* (R), that measures the time lapse since the last transaction; the *monetary* aspect (M), notably the minimum, maximum, median and average of historical transactions, as well as the value of the most recent one; and the *frequency* (F) responsible for quantifying the number of transactions made each day, week, month, year and so forth[86]. Besides being useful in detecting credit card misuse alongside other types of fraud[87], the RMF variables are also a powerful technique for identifying and combatting money laundering and terrorist financing, by uncovering patterns typically used by money launderers[88].

As a matter of fact, because a single transaction is very unlikely to appear to be a money laundering event, the application of data mining techniques within the RMF variables help to unveil the three steps associated with money laundering of *placement* (introduction of illegal capital into the banking system), *layering* (undertaking multiple transactions in the legitimate financial system) and *integration* (merging

---

[86] Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 39.

[87] As Bart Baesens *at al* points out, the RMF variables "may be operationalized for insurance claim fraud detection by constructing variables such as time since previous claim, number of claims submitted in the previous twelve months, and total monetary amount of claims since subscription of insurance contract". Idem, p. 39.

[88] Idem, p. 39.

the illegal funds with capital from legitimate activities) [89]. One of the ways it does so, for example, is by identifying several transactions made strategically close together but still under the legal threshold by which banks are obligated to report a transaction to authorities – strategy known as smurfing or structuring[90].

Another technique that plays a key role in the fight against money laundering and terrorist financing is *link analysis* through which algorithms can flag transactions as suspicious by confronting them with a database of recorded fraudsters and money launderers, as well as of countries with known connections with terrorist organizations[91]. Finally, when a legitimate company is used to launder money or aid terrorist financing, practices of income statement laundering done by overstating income and expenses can be detect by the analysis of its balance sheet throughout accounting techniques such as Benford's Law test, first two digits test or last two digits tests[92], among others, a theme that shall be addressed in the next section.

V. DATA MINING TECHNIQUES AND FORENSIC ACCOUNTING

The last form of financial fraud against which Big Data and artificial intelligence techniques are applied to prevent and detect are the so-called *financial statement fraud*, a form of misbehaviour by which market participants make false or incomplete statements about the real nature or financial health of a company[93]. Following three main objectives of either covering up the misappropriation or misapplication of funds, misleading investors or regulators about the profitability and the future prospects of the firm or, finally, to facilitate and hide other criminal activities (such as money laundering or tax evasion), this kind

---

[89] Gee, Sunder. Fraud and fraud detection: a data analytics approach. *Cit.*, p. 254-256.

[90] Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review. *Cit.*, p. 239.

[91] Idem, p. 237-239.

[92] Gee, Sunder. Fraud and fraud detection: a data analytics approach. *Cit.*, p. 257-259. Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology. *Cit.*, p. 08.

[93] Reurink, Arjan. Financial fraud: A literature review. *Cit.*, p. 08.

of fraud is committed within the balance sheet of a company by using different forms of accounting manipulation and deception[94]. The association of several harmful consequences caused by these forms of white collar crime in the financial sector and in the real economy with the difficulty of unveiling their occurrence until it is too late, led to the development of the specialized field of forensic accounting[95]. As a merger of both forensic science and accounting, based on the combination of skills and techniques emerging from law, accounting and audit, this field is responsible for the process of assessment, interpretation, summary and presentation of complex financial issues with the main purpose of preventing and detecting fraud[96].

With the increasing volume and complexity of corporate information and the huge amount of structured and unstructured data creating "large samples that will usually be too extensive to review given the auditor or forensic accountant's time constraints"[97], the use of Big Data and artificial intelligence techniques to perform and improve audit and forensic practices is on the rise, in a search for greater operational efficiency[98]. Within the forensic accounting framework, the

---

[94] Young, David. Financial Statement Fraud: Motivation, Methods, and Detection. Baker, H.K., Purda-Heeler, L. and Saadi, S. (Ed.) Corporate Fraud Exposed, Emerald Publishing Limited, Bingley, 2020, p. 325. Regarding the techniques used to commit this type of fraud, Arjan Reurink clarifies that a "review of the literature ... shows that this myriad of techniques can be broken down into five broad categories. The first two of these, *revenue-based schemes* and *expense-based schemes*, aim at artificially boosting a firm's current profitability as reported on the in- come statement. The third and fourth categories, *asset-based schemes* and *liability-based schemes*, involve the fraudulent strengthening of the balance sheet through misrepresentations of asset values and risk exposures, in order to increase a company's financial health and perceived future earnings power. The final category, *other financial statement schemes*, represents a residual one". Reurink, Arjan. Financial fraud: A literature review. *Cit.*, p. 09.

[95] Kumari Tiwari, Reshma; Debnath, Jasojit. Forensic accounting: a blend of knowledge. *Journal of Financial Regulation and Compliance*, 25(1), 2017, p. 73.

[96] Akinbowale, Oluwatoyin Esther; Klingelhöfer, Heinz Eckart; Zerihun, Mulatu Fikadu. An innovative approach in combating economic crime using forensic accounting techniques, *Cit,* p. 1263-1266.

[97] Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Cit.,* p. 620.

[98] Rezaee, Z. and Wang, J. Relevance of big data to forensic accounting practice and education. *Cit.,* p. 270-271. Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. *Cit.,* p. 18.

use of advanced digital tools is increasingly playing a critical role in the steps of data acquisition and management, data analysis and deep investigation and, finally, presentation of findings[99]. During the first phase of preparation of the investigation, for example, not only is it possible to extract files and recover deleted ones, but it is also possible to perform, with enhanced precision and quicker results, a *content examination*, identifying the type of each data file in the system and comparing it with known documents, as well as a *transaction examination*, reviewing the time of its occurrence and the sequence of creation of its data in the system[100].

When it comes to the analysis of large data sets, on the other hand, data mining techniques are a powerful tool for continuous monitoring and periodic analysis, improving the process of transaction testing, proactive fraud detection, detection of abnormalities, unstructured data reviews and pattern recognition[101]. In this context, these data mining techniques can either be *predictive*, used to reduce the risks of fraud in selected business processes, or *descriptive*, employed to detect anomalies in large data sets. While predictive techniques use large historical data sets to predict outcomes of a targeted variable and avoid fraud and manipulation, descriptive ones identify clusters and underlying associations within the data in a search for deviating behaviours that may need further investigation[102].

Even recognizing, following Mark Nigrini, that "[t]he literature lacks a case-based guide for external auditors and forensic accountants to determine which analytics tests might be effective in a proactive fraud detection exercise"[103], it is possible to argue that the use of data mining techniques to analyse the balance sheet of a company is

---

[99] Akinbowale, Oluwatoyin Esther; Klingelhöfer, Heinz Eckart; Zerihun, Mulatu Fikadu. An innovative approach in combating economic crime using forensic accounting techniques. *Cit.,* p. 1259-1260.

[100] Kumari Tiwari, Reshma; Debnath, Jasojit. Forensic accounting: a blend of knowledge. *Cit.,* p. 79.

[101] Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Cit.,* p. 603.

[102] Kumari Tiwari, Reshma; Debnath, Jasojit. Forensic accounting: a blend of knowledge. *Cit.,* p. 78.

[103] Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Cit.,* p. 603.

associated with a wide range of fraud detection tests[104] designed to identify "seven categories of fraudulent number patterns: round numbers, rising numbers, threshold numbers, non-Benford numbers, repeated numbers, outlier numbers, and rounded numbers"[105]. Because the volume of information is usually large, the use of Big Data to perform predefined audit testes combined with a data extraction tool aids the forensic accounting process, allowing the entire data set to be tested and evaluated[106].

Even though the analysis can only provide a list of anomalies and not a set of confirmed fraud cases, the greater analytical capacities provided by these algorithms reveals patterns of interest, reducing notable transactions eligible for a human review to a manageable number of entries which can subsequently be analysed using fraud-audit procedures[107]. Once a fraudulent event is confirmed, the forensic accountant can review its previous hypothesis, adjust its tests and, with a revised plan, perform additional analytical investigations and procedures, in a circular process that may continue several times.

Finally, despite its clear importance, however, this is not the only use of Big Data in the context of forensic accounting practice, as Zabihollah Rezaee and Jim Wang explain[108]:

> First, when forensic accountants investigate fraud, corruption or bribery cases, they take industry-specific norms or regulations into consideration and use keyword phrases to identify potential fraud.

---

[104] For instance, it is possible to mention the Benford's Law Test, Number Duplication Test, Z-Score Test, Relative Size Factor Test, Same-Same-Same Test, Same--Same-Different Test, Trend Analysis, GEL-1 and GEL-2 Tests, Relative Size Factor Test, Even Dollar Amounts, Payments without Purchase Orders Test, Length of Time between Invoice and Payment Dates Test, Payroll Master and Commission Tests. For an overview of each test, see Gee, Sunder. Fraud and fraud detection: a data analytics approach. New Jersey: Ed. Wiley, 2015.

[105] Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Cit.,* p. 603-604. As the author argues, these patterns of fraudulent numbers are associated with three white collar crimes, namely, asset misappropriation, corruption and financial statement fraud.

[106] Kumari Tiwari, Reshma; Debnath, Jasojit. Forensic accounting: a blend of knowledge. *Cit.,* p. 78.

[107] Gee, Sunder. Fraud and fraud detection: a data analytics approach. *Cit.*, p. 67-68.

[108] Rezaee, Z. and Wang, J. Relevance of big data to forensic accounting practice and education. *Cit.,* p. 270-271.

Second, by using historical activities or transaction data, forensic accountants can use predictive modeling and other advanced analytics to detect suspicious and anomalous transactions, high-risk events, or potential fraudulent behavior or activities. Third, by mining across multiple databases (such as customer or third-party databases), forensic accountants can use entity resolution algorithms to identify hidden relationships, addresses and aliases and investigate conflicts of interest, fake identities or sanctioned individuals and entities. Fourth, forensic accountants use social network analytics to detect hidden relationships, bogus vendors or fake bank accounts when they analyze both structured and unstructured data in the format of visuals and links from social media. Fifth, a large amount of unstructured text data is available from the free text field of journal entries, payment description, expense details, e-mails, social media, documents, presentations and hard drives of individual employees or organizations. Forensic accountants use text mining or text analytics with heuristic rules and statistical techniques to discover the sentiments and conceptual meanings of large amounts of text data, which help to identify potential fraud or non- compliance in the organization. Finally, besides traditional simple spreadsheets or static charts and graphs, forensic accountants use data visualization techniques and interactive dashboards to present evidence in an easy to understand manner.

VI. CONCLUSIONS

The development of new technologies and the digital transformation of society are changing many aspects of life and introducing new ways of social interaction. Alongside the expansion of electronic databases, the development and increased adoption of different IT trends is fostering the evolution and application of Big Data and Artificial Intelligence. The expansion of digital data and the increasing importance of electronic databases for economic development and strategic management are reshaping many aspects of society as well as several different business sectors. The financial services are the leading field in investment in advanced technology for continuous monitoring, transaction analysis, anomaly detection, data review, pattern recognition and other emerging techniques. Simultaneously, the improvement of the data extraction and analysis capacity, enabled by new technological

advances such as data mining and data analysis, is revolutionizing audit and forensic accounting practices and improving the fight against financial crimes, corporate fraud and money laundering.

Data analytics applied to fraud detection uses supervised and unsupervised methods. Supervised methods learn from historical observations from which they extract patterns of fraudulent and non-fraudulent behaviour based on selected samples applied to train algorithms that latter assign a suspicion score to evaluated cases. Unsupervised methods, on the other hand, are trained with a baseline of what represents normal behaviour and focus on detecting anomalies outlining observations that departure from this norm. Along with these methods, social network analysis is also applied by creating a so-called spider construction to analyse network-related information and identify potentially suspicious activities. These three fraud detection techniques are not mutually exclusive and, once each one focuses on a different aspect of fraud, are usually complementary, reinforcing each other's strength and compensating for their vulnerabilities.

The solutions in the market aim at two different types of misconduct associated, firstly, with illegal financial flows in crimes such as credit card fraud, financial identity fraud, money laundering and terrorist financing and, secondly, with different forms of financial fraud, which range from internal corporate crimes, such as money embezzlement and reckless management, to financial statement fraud in the form of market manipulation and investment scams.

Credit card fraud detection algorithms use artificial intelligence and data mining techniques to analyse transactions data as well as its numerical and categorical attributes. It also uses descriptive analytics methods such as *outlier detection techniques* to identify abnormal or anomalous behaviours that might indicate suspicious activities, based on standards of normality obtained by the application of statistical tools, machine learning methods and data mining techniques which examine and identify both individual patterns of previous usage and general patterns of consumption. The so-called RMF variables, that identify the recency (R), monetary aspect (M) and frequency (F) are an important type of aggregated transactional information that are useful in detecting credit card misuse and identifying and combatting money laundering and terrorist financing.

Big Data and artificial intelligence techniques are also applied in order to prevent and detect financial statement fraud by providing

greater operational efficiency for forensic accounting processes, aiding the steps of data acquisition and management, data analysis and deep investigation and, at last, presentation of findings. By being either predictive or descriptive, data mining techniques are a powerful tool for continuous monitoring and periodic analysis, improving the process of transaction testing, proactive fraud detection, detection of abnormalities, unstructured data reviews and pattern recognition. These techniques are used to analyse the balance sheet of a company and perform predefined audit tests designed to identify seven categories of fraudulent number patterns. With its enhanced analytical capacities, data mining techniques reveal patterns of interest, reducing notable transactions eligible for a human review to a manageable number of entries, which can subsequently be analysed using fraud-audit procedures.

REFERENCES

Akinbowale, Oluwatoyin Esther; Klingelhöfer, Heinz Eckart; Zerihun, Mulatu Fikadu. An innovative approach in combating economic crime using forensic accounting techniques, *Journal of Financial Crime*, Vol. 27 n. 4, 2020, p. 1253-1271.

Beasens, Bart. Analytics in a Big Data World. The Essential Guide to Data Science and its Applications. New Jersey: Wiley, 2014.

Baesens, Bart; Van Vlasselaer, Véronique; Verbeke, Wouter. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. A Guide to Data Science for Fraud Detection. New Jersey: Wiley, 2015.

Bawack, R.E., Fosso Wamba, S. and Carillo, K.D.A. A framework for understanding artificial intelligence research: insights from practice. *Journal of Enterprise Information Management*, Vol. 34 No. 2, 2021, p. 645-678.

Bhattacharyya, Siddhartha; Jha, Sanjeev; Tharakunnel, Kurian; Westland, J Christopher. Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50, (3), 2011, p. 602-613.

Bolton, R. J., Hand, D. J., Provost, F., Breiman, L., Statistical Fraud Detection – A Review, Statistical Science, Vol. 17, No. 3 (Aug., 2002), Institute of Mathematical Statistics, pp. 235-249.

European Central Bank. Sixth report on card fraud. August 2020.

Ferguson, R.I., Renaud, K., Wilford, S. and Irons, A. "PRECEPT: a framework for ethical digital forensics investigations", *Journal of Intellectual Capital*, 2020, Vol. 21 No. 2, pp. 257-290.

Gee, Sunder. Fraud and fraud detection: a data analytics approach. New Jersey: Ed. Wiley, 2015.

Hand, D. J. and Blunt, G. Prospecting for gems in credit card data. *IMA Journal of Management Mathematics*, 2001, n. 12, 173– 200.

Kumari Tiwari, Reshma; Debnath, Jasojit. Forensic accounting: a blend of knowledge. *Journal of Financial Regulation and Compliance*, 25(1), 2017, p. 73–85.

Lee, Jae-Ung; Soh, Woo-Young. Comparative analysis on integrated digital forensic tools for digital forensic investigation. *IOP Conference Series: Materials Science and Engineering*, 2020, 834(), 012034.

Louwers, Timothy J. (2015). The past, present, and future (?) of crime-related forensic accounting methodology. Accounting Research Journal, 28(1), p. 4–9.

Marr, Bernard. Big Data in Practice. How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. Chichester: Wiley, 2016.

Netherlands Register of Court Experts NRGD, 2016. Standards 008.0 Digital Forensics. Technical Report Netherlands Register of Court Experts. Available at https://www.nrgd.nl/binaries/Standards%20Digital%20Forensics_tcm39-82994.pdf

Nigrini, Mark J. The patterns of the numbers used in occupational fraud schemes. *Managerial Auditing Journal*, Vol. 34 No. 5, 2019, p. 602-622.

Nikkel, Bruce. Fintech forensics: Criminal investigation and digital evidence in financial technologies. *Forensic Science International: Digital Investigation*, Volume 33, 2020, 200908.

Oliveira Júnior, Edson; Zorzo, Avelino F.; Neu, Charles Varlei. Towards a conceptual model for promoting digital forensics experiments. *Forensic Science International: Digital Investigation*, 35, 2020, 301014.

Reurink, Arjan. Financial fraud: A literature review. MPIfG Discussion Paper, No. 16/5, Max Planck Institute for the Study of Societies, Cologne, 2016.

Rezaee, Zabihollah; Wang, Jim. Relevance of big data to forensic accounting practice and education. *Managerial Auditing Journal*, Vol. 34 No. 3, 2019, p. 268-288.

V. Rajič, M. Milenković and G. Vojković. Digital forensics appliance in corporate ecosystem considering limitations in the EU legal framework. 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, p. 2094-2100.

van Baar, R.B.; van Beek, H.M.A.; van Eijk, E.J. (2014). Digital Forensics as a Service: A game changer. Digital Investigation, 11(), S54–S62.

van Beek, H.M.A.; van den Bos, J.; Boztas, A.; van Eijk, E. J.; Schramp, R.; Ugen, M. Digital forensics as a service: Stepping up the game. *Forensic Science International: Digital Investigation*, Volume 35, 2020, 301021, ISSN 2666-2817.

Young, David. Financial Statement Fraud: Motivation, Methods, and Detection. Baker, H.K., Purda-Heeler, L. and Saadi, S. (Ed.) Corporate Fraud Exposed, Emerald Publishing Limited, Bingley, 2020, p. 321-339.

Young, Michael R. Financial Fraud Prevention and Detection. Governance and Effective Practices. New Jersey: Wiley, 2014.

Wang, Shiguo. A comprehensive survey of data mining-based accounting-fraud detection research. *2010 International Conference on Intelligent Computation Technology and Automation*, ICICTA 2010, 1, 50–53

Wilson, C. (2019), "IBM Tech trends to watch in 2020 . . . and beyond", IBM, available at: https://www.forbes.com/sites/ibm/2019/12/09/ibm-tech-trends-to-watch-in-2020–and-beyond/#5511ae004c1c (accessed on 22nd September 2021).

Wu, Tina; Breitinger, Frank; O'Shaughnessy, Stephen (2020). Digital forensic tools: Recent advances and enhancing the status quo. *Forensic Science International: Digital Investigation*, 34(), 300999.