An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection

Elena-Adriana MINASTIREANU[1], Gabriela MESNITA[2]
[1]"Alexandru Ioan Cuza" University of Iași, Doctoral School of Economics and Business Administration, Iași, 700057, Romania,
[2]"Alexandru Ioan Cuza" University of Iasi, Faculty of Economics and Business Administration, Business Information Systems Department, Iași, Romania
adrianastan3@gmail.com, gabriela.mesnita@feaa.uaic.ro

Today illegal activities regarding online financial transactions have become increasingly complex and borderless, resulting in huge financial losses for both sides, customers and organizations. Many techniques have been proposed to fraud prevention and detection in the online environment. However, all of these techniques besides having the same goal of identifying and combating fraudulent online transactions, they come with their own characteristics, advantages and disadvantages. In this context, this paper reviews the existing research done in fraud detection with the aim of identifying algorithms used and analyze each of these algorithms based on certain criteria. To analyze the research studies in the field of fraud detection, the systematic quantitative literature review methodology was applied. Based on the most called machine-learning algorithms in scientific articles and their characteristics, a hierarchical typology is made. Therefore, our paper highlights, in a new way, the most suitable techniques for detecting fraud by combining three selection criteria: accuracy, coverage and costs.

Keywords: *Bank fraud, Detection algorithms, Machine-Learning algorithms, Online transactions*

1 Introduction

During the last decades, the dependency on e-commerce and online payments has increasingly grown. As the area of information technology is developing every day to be better over the time, illegal attempts in online transactions have been increased worldwide and because of that most organizations and people are suffering substantial financial losses [1]. In the literature [2], fraud is defined as "the abuse of a profit organization system without necessarily leading to direct legal consequences".

Online bank fraud is continuously evolving and is difficult to analyze and detect because of the fraudulent behavior which is dynamic, spread across different customer profiles and dispersed in very large and dynamic datasets. Complex decision-making systems based on algorithms and analytical technologies have been developed. These can learn from previous experiences and create patterns that can detect proactively potentially fraudulent transactions.

Going through a number of important research studies within the last few years, this paper aims to provide a review of up-to-date techniques for fraud detection based on the most outstanding criteria:

- The algorithm should achieve high accuracy while processing large volumes of transaction data => high accuracy
- The algorithm should help to obtain high fraud coverage combined with low false positive rate => high coverage
- The algorithm should be useful for both the organizations and individual users in terms of cost and time efficiency => cost

The structure of the paper is divided as follows. The first part offers background over the machine-learning algorithms used in fraud detection highlighting the chosen criteria. The second part presents the methodology of the research and the classification of the various techniques used in fraud detection based on the defined criteria. Finally, the paper presents the research results and conclusions.

DOI: 10.12948/issn14531305/23.1.2019.01

2 Background

Online banking fraud has become a serious issue in financial crime management for all bank institutions. It is becoming ever more challenging and leads to massive losses, due to the emergence and evolution of complex and innovative online banking fraud, such as phishing scams, malware infection and ghost websites. The detection of online banking fraud needs to be instant because it is very difficult to recover the loss if fraud is undiscovered during the detection period. Most customers usually rarely check their online banking history regularly and are therefore not able to discover and report fraud transactions immediately after an occurrence of fraud. This makes the possibility of loss recovery very low. In this context, online banking detection systems are expected to have high accuracy, high detection rate, and low false positive rate for generating a small, manageable number of alerts in complex online banking business. These characteristics greatly challenge existing fraud detection techniques for protecting credit card transactions, e-commerce, insurance, retail, telecommunication, computer intrusion, etc. These existing methods demonstrate poor performance in efficiency and/or accuracy when directly applied to online banking fraud detection [3]. For instance, credit card fraud detection often focuses on discovering particular behavior patterns of a specific customer or group, but fraud-related online banking transactions are very dynamic and appear very similar to genuine customer behavior. Some intrusion detection methods perform well in a dynamic computing environment, but they require a large amount of training data with complete attack logs as evidence. However, there is no obvious evidence to show whether an online banking transaction is fraudulent.

As stated in the work of Wei et al. (2013) [1], the essence of online fraud reflects the abuse of interaction between resources in three worlds:

- the fraudster’s intelligence abuses in the social world,
- the abuse of web technology and Internet banking resources in the cyber world
- the abuse of trading tools and resources in the physical world.

In the same work we find that most online fraud detection have the following characteristics and challenges:

- The data set is large and highly imbalanced – for example, in a very large data set of more than 300 000 transactions in one day there were present only 5 cases of fraud which results in the task of detecting very rare fraud dispersed among a massive number of genuine transactions.
- Fraud detection needs to be real time – taking into account the fact that the interval between a customer making a payment and the payment being transferred to its destination account is usually very short; to prevent instant money loss, a fraud detection alert should be generated as quickly as possible. This requires a high level of efficiency in detecting fraud in large and imbalanced data.
- The fraud behavior is dynamic – with the everyday advances in information technology, fraudsters continually advance their techniques to defeat online banking defenses.
- The customer behavior patterns are diverse – in this context, fraudsters tend to simulate genuine customer behavior. Also, they change their behavior frequently to compete with advances in fraud detection. All of these make it difficult to characterize fraud and even more difficult to distinguish it from genuine behavior.
- The online banking system is fixed - customer accesses the same banking system which can lead to good references for characterizing common genuine behavior sequences, and for identifying suspicions in fraudulent online banking.

The above characteristics make the detection of fraud very challenging, which is the reason why there have been developed many machine-learning techniques to fix this problem [5].

Seeja and Masoumeh (2014) [6] proposed a credit card fraud detection model for highly and anonymous dataset. Frequent item set mining was used to handle the class imbalance problem thereby finding legal and illegal transaction patterns for each customer. A matching algorithm was then used to determine the pattern of an incoming transaction whether it was genuine or fraud. The evaluation of this model confirmed that it is possible to detect fraudulent transaction and also improve imbalance classification.

Duman and Ozcelik (2011) proposed a novel combination of the genetic algorithm and the scatter search algorithm to detect credit card fraud in a large Turkish bank. From this novel combination, the authors were able to improve the bank's existing fraud detection strategy by obtaining a high coverage of 200% [41].

Krenker et al. (2009) [7] proposed a model for real time fraud detection based on bidirectional neural networks. In their study, they used a large data set of cell phone transactions provided by a credit card company. The results confirmed that the proposed model outperforms the rule-based algorithms in terms of false positive rate.

In the same context of false positive rate, in 2011 Bhusari V. et al. [11] used Hidden Markov Model in order to detect credit card fraud during transactions. Their experiment confirmed that HMM model helps to obtain a high fraud reporting combined with a low false positive. HMM model represents a great value solution for addressing detection of fraud transaction through credit card [11]. Also, Delio Panaro et al. (2015) [12] proposed a two layer statistical classifier for sensitive, highly skewed and massive data sets to detect fraud. The algorithm has been inspired by the necessity of analyzing a data set of about fifteen million real world online banking transactions, spanning from 2011 to 2013 with the aim of detecting frauds from legitimate operations. Results confirmed that the algorithm is particularly effective in detecting anomalies, achieving high true positive rates and reasonably low false positive rates. Therefore, several other studies [71-73] have been made to develop classifiers in this sense of high coverage, which include techniques based on Naïve Bayes, boosting, neural networks, and ensemble learning.

In a study made by Mishra et al (2014) [8] the analysis of credit card fraud detection has been done through three classification models on two datasets. The approaches were compared according to their accuracy and elapsed time. The comparison of its performance was done with two approaches like decision tree for fraud detection and multilayer perceptron network.

Azeem Ush Shan et al. (2014) [9] proposed an algorithm named Simulated Annealing algorithm that was used to train the neural networks for the detection of credit card frauds in a real-time scenario. The proposed technique was useful for individual users and also for the organizations in terms of cost and time efficiency.

In this context of cost efficiency, in 2013 Sahin et al. [10] proposed a new cost-effective tree decision approach to minimize the total cost of categorization which addresses the problem of detecting fraud.

Analyzing the so far published literature it is pragmatic that most of the articles focus on detection of fraud in the context of high accuracy while processing large volumes of transaction data, cost and time efficiency, high fraud coverage combined with low false positive rate etc. Which represents the reason why the focus of our research is mainly on these three criteria.

3 Research methodology

The research aims to analyze and classify machine-learning techniques suitable to detect bank fraud in the online environment taking into account the following criteria: high accuracy, high coverage and low costs.

Therefore, a meta-analysis was performed on a various number of specialized articles (peer-reviewed journals articles and conference papers) from the period of 2010 till present that fit into the defined criteria (high accuracy, high coverage, low costs) and follow these descriptors found in the abstract of the articles and also in their title: detecting bank fraud, online bank fraud, bank fraud and machine-learning, bank fraud detection, and detecting bank fraud via machine-learning. The descriptors were chosen based on the scope of this article that well describe the most called machine-learning techniques used in detecting fraud in the online transactions. Following the descriptors, the articles were retrieved from ACM Digital Library, IEEE Xplorer Digital Library, Science Direct, Springer Link, etc. The analysis based on types of fraud reveals that 38 articles include credit frauds, 3 financial frauds and 9 e-commerce frauds as we present in table 1.

Table 1. Types of fraud mentioned in the specialized scientific articles

No. crt.	Type of fraud	Number of articles	References
1.	Credit card fraud	38	[4], [6], [8], [13-14], [16-19], [20-22], [24], [27-30], [32-33], [36-46], [48-49], [52-56], [70]
2.	Financial fraud	3	[15], [26], [31]
3.	E-commerce fraud	9	[8], [11], [12], [34], [35], [47], [51], [53], [57]

We find 19 machine-learning techniques mentioned in these articles analysed with different numbers per each article, from 1 to maximum 5. Also, there are 4 techniques (Artificial Neural Network, Decision Tree, Genetic algorithm, and Support Vector Machine) which appear in more 10 articles. On the other hand, there are 4 techniques (Expert system, Gradient Descendent, K-means, and Scatter Search) mentioned in only one article. These 4 techniques were kept in the methodology because they presented high results for accuracy [29, 34], but they have been removed from the results section because of the low frequency. We summarize these findings in table 2, where we have ordered the machine-learning techniques based on the two main categories in which these algorithms can be organized, supervised and unsupervised learning techniques.

Table 2. The machine-learning techniques mentioned in the specialized scientific articles

No. crt.	Machine learning technique	Number of articles	References
1.	Artificial Immune System (AIS)	4	[42-43], [59-60]
2.	Artificial Neural Network (ANN)	20	[8], [9], [13], [15], [18], [29-31], [33], [38], [43], [46-48], [50-51], [53-54], [57-58]
3.	Bayesian network	8	[13], [16-18], [26], [32], [51], [57]
4.	Support Vector Machine (SVM)	17	[4], [6], [12-15], [17], [20], [34], [37], [39], [43], [45], [52], [56], [63], [70]
5.	Decision Tree (DT)	19	[4], [8-9], [13], [15], [17-19], [21-22], [27], [31-32], [43-46], [55-56]
6.	Logistic regression	8	[4], [14-15], [34], [39], [55], [57], [70]
7.	Naïve Bayes	7	[6], [27], [34], [40], [45], [56], [70]
8.	Random forest	8	[4], [6], [14], [34], [39, 55, 57, 70
9.	Fuzzy logic based system	4	[13], [43], [46], [49]
10.	K-nearest neighbor (KNN)	10	[6], [17], [40], [43], [45], [49], [50], [64- 65], [70]

DOI: 10.12948/issn14531305/23.1.2019.01

11.	Hidden Markov Model (HMM)	8	[11], [13], [25], [37], [51], [53], [61-62]
12.	Self-organizing map (SOM)	4	[6], [23-24], [30]
13.	Genetic algorithm (GA)	13	[9], [13], [39], [41], [43], [46-50], [54], [68-69]
14.	K-means	1	[34]
15.	DBSCAN	2	[16], [28]
16.	Expert system	1	[50]
17.	Gradient Descendent	1	[39]
18.	Scatter search	1	[41]

From all these scientific articles, 36 of them used data sets in order to sustain the research, as we can see in table 3. We mainly concentrated on those articles that used a huge volume of data and presented good results for the main criteria: high coverage, high accuracy and low costs. In terms of coverage, SVM, DT, Naïve Bayes and KNN obtained high and medium rate on more than 1 million of transactions or data records and also on those that presented from thousands to several hundred thousand of transactions or data records [4, 6, 12, 17, 19, 20, 21, 31, 35, 39, 40, 44, and 70]. In terms of accuracy, SVM, DT, Naïve Bayes and KNN presented high and medium rate on more than 1 million of transactions or data records and also on those that presented from thousands to several hundred thousand of transactions or data records [4, 17, 20, 21, 31, 35, 40, and 44]. In terms of costs, these were high for all the algorithms applied on huge volume of data (more than 1 million transactions or data records).

Table 3. Data sets used in the specialized scientific articles

No. crt.	Number of transactions or records	Number of articles	References
1.	More than 1 million transactions or data records	9	[12], [14], [17], [19], [22], [28], [31], [42], [55]
2.	From thousands to several hundred thousand of transactions or data records	19	[4], [6], [15], [20-21], [24], [26-27], [29], [30], [34-35], [38-41], [44-45], [70]
3.	Hundreds of transactions or data records	2	[37], [43]
4.	Only mentioned the use of data sets	8	[9], [11], [16], [18], [36], [52], [56-57]

Machine-learning techniques intensively use math statistics, as well as knowledge and results from fields such as artificial intelligence, mathematics, psychology, neurobiology, information technology. Thus, depending on the type of learning, machine-learning algorithms can be organized into two main categories:

- **supervised learning** algorithms are mainly used for accurate classification and prediction, being a method of classification with labeled data (Artificial Immune System, Artificial Neural Network, Bayesian network, Support Vector Machine, Decision Tree, Logistic regression, Naïve Bayes, Random forest, Fuzzy logic based system, and K-nearest neighbor)
- **unsupervised learning** algorithms cluster unlabeled data with similar attributes, usually performs lower accuracy than supervised learning algorithms (Hidden Markov Model, Self-organizing map, Genetic algorithm, K-means, DBSCAN, Expert system, Gradient Descendent, and Scatter search).

DOI: 10.12948/issn14531305/23.1.2019.01

To sum up, the analysis of a number of various relevant articles retrieved mainly from Data Science was meant to identify a set of machine-learning techniques that present similar properties and meet the defined criteria. For this, we grouped the machine-learning algorithms by supervised and unsupervised techniques and by the volume of data on which these algorithms were applied. The conclusion to all this process was that the supervised learning algorithms are the most called techniques and offer high accuracy, high coverage with the disadvantage of high costs than the unsupervised ones in detecting online bank fraud.

4 Research results

In this paper we presented a comparative study of 14 most called algorithms in scientific articles regarding online fraudulent transactions (artificial immune system, artificial neural network, Bayesian network, DBSCAN, decision tree, Fuzzy logic based system, genetic algorithm, Hidden Markov Model, k-nearest Neighbour, logistic regression, Naïve Bayes, random forest, Self-organizing map, support vector machine) based on their usage frequency and on certain criteria:

- The algorithm should achieve high accuracy while processing large volumes of transaction data => high accuracy
- The algorithm should help to obtain high fraud coverage combined with low false positive rate => high coverage
- The algorithm should be beneficial for both the organizations and individual users in terms of cost and time efficiency => cost

The classification of the algorithms that we can see in table 4 was made based on the positive and negative instances that a classifier predicts correctly, metrics that we find in the following formulas:

$$TPR = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \ (1)$$

$$TNR = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}} \ (2)$$

$$FPR = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \ (3)$$

$$FNR = \frac{\mathrm{FN}}{\mathrm{TP} + \mathrm{FN}} \ (4)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \ (5)$$

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \ (6)$$

The positive and negative instances that a classifier predicts correctly are called true positives TP and true negatives TN. The incorrectly classified instances are called false positives FP and false negative FN. Based on that True Positive Rate, True Negative Rate, False Positive Rate, False Negative Rate, Precision, and Accuracy concepts will occur which helps in classifying the techniques. The correctitude in the use of this metrics represented also a point for which we have chosen the articles for review and also the three criteria, as it is important to have algorithms that can deal with these issues in an efficient manner. As we didn’t have access the datasets to apply the metrics, for future work we will use these formulas on our dataset and verify if the obtained set of machine-learning techniques present the same accuracy as stated in the reviewed literature.

According to the analysis of the reviewed articles and the above metrics, in the following part we present in a comparative way the results of the most applied fraud detection techniques in the online environment based on their usage frequency and the defined criteria – accuracy, coverage, costs, where 1 – means low, 2 – means medium, 3 – means high.

Table 4. Classification of algorithms based on the defined criteria and frequency

Machine Learning algorithm	**Type of technique: supervised (st) or unsupervised (ut)**	**Frequency**	**Accuracy**	**Coverage**	**Costs**
Artificial Neural Network (ANN)	st	40%	2	2	**3**
Decision Tree (DT)	st	38%	2	2	**3**
Support Vector Machine (SVM)	st	34%	**3**	**3**	**3**
Genetic algorithm (GA)	ut	26%	2	2	1
K-nearest Neighbor (KNN)	st	20%	2	2	**3**
Bayesian Network	st	16%	**3**	2	**3**
Hidden Markov Model (HMM)	ut	16%	1	1	**3**
Logistic regression	st	16%	**3**	2	2
Random forest	st	16%	**3**	2	2
Naïve Bayes	st	14%	2	2	**3**
Self-organizing map (SOM)	ut	8%	2	1	**3**
Fuzzy Logic Based system (FL)	st	8%	**3**	2	**3**
Artificial Immune System (AIS)	st	8%	2	**3**	1
DBSCAN	ut	4%	**3**	2	**3**

The above table illustrates that the supervised learning algorithms are used more frequently than the unsupervised ones, findings that are also sustained by the most used public platform for data science competitions [74]. As the results show, the fraud detection systems based on SVM, Bayesian Network, Fuzzy logic based and DBSCAN have very high accuracy with 100% true positive but with the disadvantage of high costs when processing large datasets. In another view, Genetic algorithm and AIS present medium accuracy with low costs when processing large datasets. For comparing the other techniques such as: ANN, DT, KNN and Naïve Bayes we see that they present medium accuracy and medium coverage with the disadvantage of high costs. At the same time, SOM presents medium accuracy with high costs but low coverage and HMM present high costs but low accuracy and low coverage.

Based on literature review [32, 39, and 70] the supervised learning algorithms appear to be the most called techniques in the detection of online fraudulent transactions with the disadvantage of high costs. Also, the literature [17, 29] states that the most popular fraudulent transactions in the online environment are those made with credit card, the result confirmed also by our findings in table 1.

According to this, the classification made on the selected criteria aims to develop more efficient and trustable fraud detection systems that should also take into consideration factors like risk ranges, bank and customer behavior, geographic areas and so on.

To conclude, the results show high accuracy and high coverage for supervised techniques with the disadvantage of high costs. However, it is important to consider that these results are good because the datasets used were highly unbalanced with lots of negative examples.

One of our research limits was the lack of access to the data sets used in the reviewed articles to determine the characteristics of the techniques. Finally, we consider our paper highlights, in a new way, the most suitable techniques for detecting fraud by combining three selection criteria: accuracy, coverage and costs.

5 Conclusions and future direction

Relatively to our study it can be stated that the problem of credit card fraud in the online environment has gained the most attention in the literature, although there are a number of significant problems that have not been addressed closely by the researchers, like online intellectual property theft, pagejacking, fake money orders, wire-transfer fraud. Our classification criteria were chosen based on the most common difficulties encountered by credit card fraud detection techniques. The classification of the algorithms showed that the best results in terms of accuracy and coverage were achieved by the supervised learning techniques: support vector machine, artificial neural network and decision tree. These three algorithms also were the most called in the reviewed articles which demonstrates the fact that they present the best results.

The classification was made with the intention to design an efficient and trustable fraud detection systems that should also take into account other variables like risk ranges, bank and customer behavior, geographic areas and so on. Thus, as far as the research direction is concerned, it is desirable to investigate possible improvements that can be made to algorithms in order to extend their applicability to the other types of online fraudulent transactions with high accuracy, high coverage and low costs. The research will focus mainly on the hybridization of the most used machine-learning techniques to help in improving the efficiency of the fraud detection process in other important areas and validate this on our own dataset.

References

[1] W. Wei, J. Li, L. Cao, Y. Ou and J. Chen, “Effective detection of sophisticated online banking fraud on extremely imbalanced data”, World Wide Web, 2013

[2] C. Phua, V. Lee, K. Smith, R. Gayler, “A comprehensive survey of data mining based fraud detection research”, Artificial Intelligence Review, pp 1 – 14, 2005

[3] K.N. Karlsen and T. Killingberg, “Profile based intrusion detection for Internet banking systems”, Norwegian University of Science and Technology, 2008

[4] K. Navanshu and S. Y. Sait, “Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models”, International Journal of Pure and Applied Mathematics, Volume 118, No. 20, pp. 825-838, 2018

[5] S. J. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach”, 3rd edition. Prentice Hall, 2010

[6] K. R. Seeja and Z. Masoumeh, “Fraud miner. A novel credit card fraud detection model based on frequent item set mining”, The Scientific World Journal, Article ID 252797, 2014

[7] A. Krenker, M. Volk, U. Sedlar, J. Bester and A. kosh, “Bidirectional artificial neural networks for mobile phone fraud detection”, Journal of Artificial Neural Networks, Volume 31, No. 1, pp. 92 - 98, 2009

[8] M. Mukesh Kumar and R. Dash, “A comparative study of Chebyshev functional link artificial neural network, multi-layer perceptron and decision tree for credit card fraud detection”, Information Technology (ICIT), International Conference on IEEE, 2014

[9] K. A. U. Shan, N. Akhtar and M. N. Qureshi, “Real-time credit-card fraud detection using artificial neural network tuner by simulated annealing algorithm”, Proceedings of International Conference on Recent Trends in Information, Telecommunication and Computing, ITC, 2014

[10] Y. Sahin, B. Serol and D. Ekrem, “A cost-effective decision tree approach for fraud detection”, Expert systems with applications, Elsevier, Volume 40, No. 15, 2013

[11] V. Bhusari and S. Patil, “Study of hidden Markov Model in credit card fraudulent detection”, International Journal of Computer Applications, Volume 20, No. 5, 2011

[12] D. Panaro, E. Riccomagno and F. Malfanti, “A Fraud Detection Algorithm For Online Banking”, 2015

[13] S. B. E. Raj and A. A. Portia, “Analysis on Credit Card Fraud Detection Methods”, IEEE – International Conference on Computer, Communication and Electrical Technology, pp. 152 – 156, 2011

[14] S. Bhattacharyya, S. Jha, K. Tharakunnel and J.C. Westland, “Data Minining for Credit Card Fraud: A comparative study”, Decision Support Systems, Volume 50, pp 602 – 613, 2011

[15] J. Perols, “Financial Statement Fraud Detection: An analysis of statistical and Machine Learning Algorithms”, Auditing: A Journal of Practice & Theory, Volume 30, pp 19 – 50, 2011

[16] J. N. Dharwa and A. R. Patel, “A Data Mining with Hybrid Approach Based Transaction Risk Score Generation Model (TRSGM) for Fraud Detection of Online Financial Transaction”, International Journal of Computer Applications, Volume 16, pp 18 – 25, 2011

[17] J. K.-F. Pun, “Improving Credit Card Fraud Detection using a Meta-Learning strategy”, a thesis submitted in conformity with the requirements for the degree of Master of Applied Science Graduate Department of Chemical Engineering and Applied Chemistry University of Toronto, 2011

[18] K. K. Sherly, “A comparative assessment of supervised data mining techniques for fraud prevention”, TIST Int. J. Sci. Tech. Res, Volume 1, pp. 1 – 6, 2012

[19] Y. Sahin, S. Bulkan and E. Duman, “A cost-sensitive Decision Tree Approach for Fraud Detection”, Expert Systemts with Applications, Volume 40, pp 5916 – 5923, 2013

[20] Q. Lu and C. Ju, “Research on Credit Card Fraud Detection Model Based on Class Weighted Support Vector Machine”, Journal of Convergence Information Technology, pp. 62, 2011

[21] D. D. Patil, V. M. Wadhai and J. A. Gokhale, “Evaluation of Decision Tree Pruning Algorithms for Complexity and Classification Accuracy”, International Journal of Computer Applications, Volume 11, No. 2, pp. 23 – 30, 2010

[22] Y. Shahin and E. Duman, “An overview of business domains where fraud can take place and a survey of various fraud detection techniques”, Proceedings of the 1st International Symposium on Computing in Science and Engineering, 2010

[23] Y.-Y. Nguwi and S.-Y. Cho, “An unsupervised self-organizing learning with support vector ranking for imbalanced datasets”, Expert Systems with Applications, Volume 37, pp 8303 – 8312, 2012

[24] D. Olszewski, “Fraud detection using Self-Organizing Map Visualizing the User Profiles”, Knowledge-Based Systems, Volume 70, pp 324-334, 2014

[25] S. Ding, H. Jia, J. Chen and F. Jin, “Granular neural networks”, Springer Artificial Intelligence review, 2012

[26] S. H. Li, D. C. Yen, W. H. Lu and C. Wang, “Identifying the Signs of Fraudulent Accounts using data mining techniques”, Computers in Human Behavior, Volume 28, pp 1002 – 1013, 2012

[27] N. Mahmoudi and E. Duman, “Detecting Credit Card Fraud by Modified Fisher Discriminant Analysis”, Expert Systems with Applications, Volume 42, pp. 2510 – 2516, 2015

[28] O. Ayano and S. O. Akinola, “A multi-algorithm data mining classification approach for bank fraudulent transactions”, African Journal of Mathematics and Computer Science Research, Volume 10, No 1, pp 5-13, 2017

[29] C. Mishra, D. Lal Gupta and R. Singh, “Credit Card Fraud Identification Using Artificial Neural Networks”, International Journal of Computer Systems, Volume 04, Issue 07, 2017

[30] F. N. Ogwueleka, “Data mining applica-

DOI: 10.12948/issn14531305/23.1.2019.01

tion in credit card fraud detection system", Journal of Engineering Science and Technology, Volume 6, No 3, pp. 311 – 322, 2011

[31] S. Soltaniziba, M. A. Balafar, "The Study of Fraud Detection in Financial and Credit Institutions with Real Data", Computer Science and Engineering, Volume 5, No 3, pp 30-36, 2015

[32] J. R. Gaikwad, A. B. Deshmane, H. V. Somavanshi, S. V. Patil and R. A. Badgujar, "Credit Card Fraud Detection using Decision Tree Induction Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 4, No. 6, 2014

[33] R. Patidar and L. Sharma, "Credit Card Fraud Detection Using Neural Network", International Journal of Soft Computing and Engineering, Volume 1, Issue NCAI2011, 2011

[34] D. Choi and K. Lee, "Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System", IT CoNvergence PRActice (INPRA), Volume 5, No. 4, pp. 12-24, 2017

[35] M. Carminati, R. Caron, F. Maggi, I. Epifani and S. Zanero, "Banksealer: A decision support system for online banking fraud analysis and investigation", Computers and Security, 53:175–186, 2015

[36] K. RamaKalyani and D. UmaDevi, "Fraud detection of credit card payment system by genetic algorithm", International Journal of Scientific & Engineering Research, 3(7):1–6, 2012

[37] T. Kavipriya and N. Geetha, "An identification and detection of fraudulence in credit card fraud transaction system using data mining techniques", International Research Journal of Engineering and Technology, Volume 05, No 01, 2018

[38] S. Ghosh and D. L. Reilly, "Credit Card Fraud Detection with a Neural-Network", Proceeding IEEE First International Conference on Neural Networks, 2014

[39] A. Sinha and S. Mokha, "Classification and Fraud Detection in Finance Industry", International Journal of Computer Applications, Volume 176, No 3, 2017

[40] S. Kiran, N. Kumar, J. Guru, D. Katariya, R. Kumar and M. Sharma, "Credit card fraud detection using Naïve Bayes model based and KNN classifier", International Journal of Advance Research, Ideas and Innovations in Technoloy, Volume 4, Issue 3, 2018

[41] E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search", Expert Systems with Applications, 2011

[42] A. Brabazon, J. Cahill, P. Keenan and D. Walsh, "Identifying Online Credit Card Fraud using Artificial Immune Systems", IEEE Congress on Evolutionary Computation, 2010

[43] R. Jain, B. Gour and S. Dubey, "A Hybrid Approach for Credit Card Fraud Detection using Rough Set and Decision Tree Technique", International Journal of Computer Applications, Volume 139, No.10, 2016

[44] A. C. Bahnsen, S. Villegas, D. Aouada and B. Ottersten, "Fraud Detection by Stacking Cost-Sensitive Decision Trees", Data Science for Cyber-Security, 2017

[45] M. Zareapoor and P. Shamsolmoalia, "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier", Procedia Computer Science, Volume 48, 2015

[46] K. Chaudhary, J. Yadav and B. Mallick, "A review of Fraud Detection Techniques: Credit Card", International Journal of Computer Applications, Volume 45, No.1, 2012

[47] A. Pouramirarsalani, M. Khalilian and A Nikravanshalmani, "Fraud detection in E-banking by using the hybrid feature selection and evolutionary algorithms", International Journal of Computer Science and Network Security, Volume 17, No. 8, 2017

[48] N.Malini and Dr.M.Pushpa, "Analysis on Credit Card Fraud Detection Techniques by Data Mining and Big Data Approach", International Journal Of Research In Computer Applications And Robotics, Volume 5 Issue 5, pp 38-45, 2017

[49] C. Sudha and T. N. Raj, "Credit Card Fraud Detection In Internet Using K-Nearest Neighbor Algorithm", International Journal of Computer Science, Volume 5, Issue 11, 2017

[50] R. F. Nejad, "The Fraud Detection in the Bank Payments and its Methods", Asian Journal of Information Technology, 14 (6), pp 239 – 245, 2015

[51] P. J. Rana and J. Baria, "A Survey on Fraud Detection Techniques in Ecommerce", International Journal of Computer Applications, Volume 113, No. 14, 2015

[52] S. Patel and S. Gond, "Supervised Machine (SVM) Learning for Credit Card Fraud Detection", International Journal of Engineering Trends and Technology (IJETT), Volume 8, No. 3, 2014

[53] R. Rajamani and M. Rathika, "Credit Card Fraud Detection using Hidden Morkov Model and Neural Networks", Proceedings of the UGC Sponsored National Conference on Advanced Networking and Applications, 2015

[54] Z. Monirzadeh, M. Habibzadeh and N. Farajian, "Detection of Violations in Credit Cards of Banks and Financial Institutions based on Artificial Neural Network and Metaheuristic Optimization Algorithm", International Journal of Advanced Computer Science and Applications, Volume 9, No. 1, 2018

[55] A. C. Bahnsen, D. Aouada, A. Stojanovic and B. Ottersten, "Feature engineering strategies for credit card fraud detection", Expert Systems With Applications 51, pp. 134–142, 2015

[56] M. Fahmi, A. Hamdy and K. Nagati, "Data Mining Techniques for Credit Card Fraud Detection: Empirical Study", Sustainable Vital Technologies in Engineering & Informatics, 2016

[57] E. Caldeira, G. Brandao and A. C. M. Pereira, "Fraud Analysis and Prevention in e-Commerce Transactions", 9th Latin American Web Congress, 2014

[58] C.L. Cocianu and H. Grigoryan, "An Artificial Neural Network for Data Forecasting Purposes", Informatica Economică, Volume 19, No. 2/2015, pp.34-45

[59] L. N. De Castro Silva and F.J. V. Zuben, "An evolutionary immune network for data clustering", Proceedings of the IEE SBRN (Brazilian Symposium on Artificial Neural Networks), pp 84-89, 2000

[60] L. De Castro and J. Timmis, "Artificial immunce systems: a new computational approach", London, UK: Springer – Verlag., 2002

[61] A. Srivastava, A. Kundu, S. Sural and A. K. Majumdar, "Credit Card Fraud Detection using Hidden Markov Model", IEEE Transactions on dependable and secure computing, Volume 5, pp 37-48, 2008

[62] A. Singhand D. Narayan, "A survey on Hidden Markov Model for Credit Card Fraud Detection", International Journal of Engineering and Advanced Technology, Volume 1, No. 3, pp 49 – 52, 2012

[63] L. State, C. Cocianu, C. Uscatu and M. Mircea, "Extensions of the SVM Method to the Non-Linearly Separable Data", Informatica Economică vol. 17, no. 2/2013, 173-182

[64] M. J. Islam, Q. M. Jonathan Wu, M. Ahmadi and M. A. Sid-Ahmed, "Investigating the Performance of Naïve-Bayes Classifiers and K-Nearest Neighbor Classifiers", IEEE, International Conference on Convergence Information Technology, pp 1541 – 1546, 2007

[65] A. Rohilla, "Comparative Analysis of Various Classification Algorithms in the Case of Fraud Detection", International Journal of Engineering Research & Technology, Volume 6, No. 09, 2017

[66] G. K. Venayagamoorthy, "Teaching Neural Networks Concepts and Their Learning Techniques", Proceedings of the American Society for Engineering Education Midwest Section Conference, 2004

[67] R. Patidar and L. Sharma, "Credit Card Fraud Detection Using Neural Network", International Journal of Soft Computing and Engineering, Volume 1, pp 32-38, 2011

[68] M. Kim and I. Han, "The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms", Elsevier, Expert Systems with Applications, pp 637-646, 2003

[69] C.-H. Wu, G.-H. Tzeng, Y.-J. Goo and W.-C. Fang, "A real-valued genetic algorithm to optimize the parameters of support vector machine for prediction bankruptcy", Expert Systems with Applications, Volume 32, No. 2, pp 397-408, 2007

[70] R. Banerjee, G. Bourla, S. Chen, M. Kashyap, S. Purohit and J. Battipaglia, "Comparative Analysis of Machine Learning Algorithms through Credit Card Fraud Detection", New Jersey's Governor's School of Engineering and Technology, 2018

[71] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem", Knowledge and Data Engineering, IEEE Transactions on 18, pp 63-77, 2006

[72] H.-N. Qu, G.-Z. Li and W.-S. Xu, "An asymmetric classifier based on partial least squares", Pattern Recognition, volume 43, pp. 3448-3457, 2010

[73] W.-T. Yih, J. Goodman and G. Hulten, "Learning at low false positive rates", Proceedings of the third conference on email and anti-spam, pp. 1-8

[74] Kaggle. (2003). The Home of Data Science & Machine Learning, Internet: https://www.kaggle.com/agpickersgill/credit-card-fraud-detection/data

[75] J. E. T. Akinsola, "Supervised Machine Learning Algorithms: Classification and Comparison", International Journal of Computer Trends and Technology (IJCTT), volume 48, No. 3, pp. 128 -138, 2017.

Elena-Adriana MINASTIREANU has graduated the Faculty of Automatic and Computer Science in 2010. She holds a Master diploma in Project Management from 2017. Currently she is acting as Project Manager / Scrum Master at a multinational IT company. She is PhD student at Doctoral School of Economics and Business Administration. Her research focuses on the analysis of fraud detection in the banking system.

Gabriela MESNITA is PhD professor of Business Information Systems at "Alexandru Ioan Cuza" University of Iasi, Faculty of Economics and Business Administration. She received her PhD in Business Information Systems in 1999. Currently professor Mesnita is PhD adviser. She published as author or coauthor over 20 books and university courses or chapters. Also she presented and/or published over 80 articles in Romania, Poland, China, Turkey, Italy etc. Her research interests include information systems (analysis and design, outsourcing); e-business; project management; IoT.

DOI: 10.12948/issn14531305/23.1.2019.01