Detection of credit card fraud: State of art

Article *in* International Journal of Computer Network and Information Security · November 2018

CITATIONS
18

READS
2,520

3 authors:

Imane Sadgali

14 PUBLICATIONS **86** CITATIONS

SEE PROFILE

Nawal Sael

Université Hassan II de Casablanca

62 PUBLICATIONS **166** CITATIONS

SEE PROFILE

Faouzia Benabbou

University Hassan II of Casablanca

95 PUBLICATIONS **222** CITATIONS

SEE PROFILE

Detection of credit card fraud: State of art

Imane Sadgali, Nawal Sael and Faouzia Benabbou

Laboratory of Modelling and Information Technology
Faculty of Sciences Ben M'SIK, University Hassan II, Casablanca, Morocco

Summary
Credit card fraud is costing the payment card industry, literally billions of dollars annually. Financial institutions try to improve continually their fraud detection systems, but fraudsters are in same time inventing new techniques to hack systems. That said; the prevention and detection of credit card fraud become an emergency. Data mining techniques are providing great help in financial fraud detection, since dealing with the large and complex among of financial data are big challenges for financial institutions. In recent years, several studies have used machine learning and data mining techniques to face this problem. In this paper, we propose a state of the art on various techniques of credit card fraud detection. The purpose of this study is to give a review of implemented techniques for credit card fraud detection, analyse their incomes and limitless, and synthesise the finding in order to identify the techniques and methods that give the best results so far.

Key words:
Fraud Detection, Financial Fraud, Machine-Learning, Credit Card

1. Introduction

Credit card fraud is an expensive problem for many financial institutions, that why they use a variety of fraud prevention models to address this problem. However, fraudsters are adaptive, and over time, they conceive several ways of intruding such protectives models. Despite the best effort of financial institutions, law enforcement and government, financial fraud continues to grow. Fraudsters today can be a very inventive, intelligent and fast fraternity. Today, credit card fraud in Morocco is focusing on foreign cards and online payments [1] says the interbank electronic banking center of Morocco (CMI). Several arrests of Moroccans and foreigners have been made since the beginning of this year. Stolen cards have been used to book hotel stays or purchase smartphones. The prevention and detection of credit card fraud become an emergency.
Machine learning techniques are providing great help in financial fraud detection, since dealing with the large and complex among of financial data are big challenges for financial institutions. The data mining techniques have the potential to solve the contradiction between effect and efficiency of fraud detection [2]. Data mining plays an important role in financial fraud detection, as it is often applied to extract and discover the hidden patterns in very large collection of data [3].

Manuscript received November 5, 2018
Manuscript revised November 20, 2018

Several studies have addressed this problem and have proposed solutions using machine learning and datamining, and it is the purpose of this paper. A survey of the related works in the existing literature was carried out, with a comparative analysis. Our objective is to point out their strength and weaknesses and to identify the open issues of credit card fraud analysis. The aim of this study is to find compromising techniques to face this problem.
The rest of this paper is organized as follows. Section 2 contains a definition and types of credit card fraud, Section 3 presents machine learning techniques implemented for credit card fraud detection. Section 4 describes review of related works in the form of comparatives tables. In Section 5, we present a statistical review of results. Finally, we conclude and propose a future work in Section 6.

2. Credit card fraud

Fraud definition, according to the Association of Certified Fraud Examiners (ACFE) "ACFE Association of Fraud Examiners Certificates", fraud includes any intentional or deliberate act of depriving another of property or money by cunning, deception or other unfair acts [4]. There are several types of financial fraud; insurance fraud, financial statement fraud, credit card fraud…
In our study, we focused on credit card fraud; which is a fraudulent transaction made by an unauthorized person on his own account while the cardholder and its provider are totally unaware of the existence of this transaction at the time of its realization. [23]. Credit card transactions data are mainly characterized by an unusual phenomenon. Both legitimate transactions and fraudulent ones tend to share the same profile. Fraudsters learn new ways to mimic the spending behaviour of legitimate card (or cardholder). Thus, the profiles of normal and fraudulent behaviours are constantly dynamic [25].
In the literature, we found that credit card fraud is essentially of two types: application and behavioural fraud [24]. Application fraud; is where fraudsters obtains new cards from issuing companies using false information or other people's information [16]. Behavioural fraud can be of four types [11]:

- Mail theft fraud: occurs when fraudsters intercept credit cards in mail before they reach cardholders or pilfer personal information from bank and credit

card statements

- Stolen/lost card fraud: happens when fraudsters get hold of credit cards through theft of purse/wallet or gain access to lost cards.
- Counterfeit card fraud: Cardholders information is obtained through a variety of ways, such as employees stealing information through unauthorized 'swipers', 'phishing' scams, or through intrusion into company computer networks.
- 'Card holder not present' fraud: happens when credit cards details are used remotely to conduct fraudulent transactions through mail, phone, or the Internet.

3. Machine learning techniques

Machine learning refers to analytic techniques that "learn" patterns in datasets without being guided by a human analyst. It helps data scientists efficiently to determine which transactions are most likely to be fraudulent, while significantly reducing false positives. The techniques are extremely effective in fraud prevention and detection, as they allow the automated discovery of patterns across large volumes of streaming transactions [5].

Various Fraud Detection techniques have been proposed in the past years, each one, has advantages and shortcomings. In this section, we present works, grouped by techniques category.

Association Rules (AR): the Fuzzy Logic (FL) is used for representing the cognitive uncertainties, measuring the intensity of the truth-values for unquantifiable measures or probabilistic measures within the range of zero and one. In 2009, Sanchez [13] propose a novel methodology based on using Fuzzy Association Rules (FAR) to detect credit card fraud. The applied methodology overcomes the difficulties of minimum support and confidence, optimizes the execution times, reduces the excessive generation of rules, and helps make the results more intuitive, thereby facilitating the work of fraud analysts. Askari[23], also, proposed fraud detection algorithm based on Fuzzy-ID3 in 2017. The applied methodology overcomes the difficulties of minimum support and confidence, optimizes the execution times, reduces the excessive generation of rules, and helps make the results more intuitive, thereby facilitating the work of fraud analysts.

Support Vector Machines (SVM) use a linear model to implement nonlinear class boundaries by mapping input vectors nonlinearly into a high-dimensional feature space. In the new space, an optimal separating hyperplane is constructed. In 2011, Bhattacharyya evaluated two advanced data mining approaches, support vector machines and random forests, together with the well-known logistic regression, the results showed that, while sensitivity and accuracy decreased with lower proportions of fraud in the training data, precision showed an opposite trend [16].

In 2013, Hejazi investigated two-class and one-class support vector machines (SVM) for detection of fraudulent credit card transactions and shown the superiority of one-class SVM (OCSVM) for the anomaly detection problem. Phuong present a Real Time Data-Driven approaches for Credit Card Fraud Detection in 2018, using OCSVM with the optimal kernel parameter selection and T2 control chart and shown that the proposed approach achieved a high-level of detection accuracy and a low false alarm rate.

Artificial Immune System (AIS): Artificial neural network (ANN) was first created with the purpose to imitate the behaviour of the human brain. Multilayer Perception Algorithm (MPL) is an artificial neural network and is a non-parametric estimator that can be used for classifying and detecting intrusions. Parenclitic Network (PN) is a network reconstruction technique, that allows highlighting the differences between one instance and a set of standard, the addition of parenclitic features to the raw data set enhance the obtained results, with the error dropping from a 19,2 to a 12,23%. In 2017, Mubalaik proposed an ANN-MPL multilayer perception [20], based on implementation of well-known machines learning techniques, it helps to anticipate and quickly detect fraud. Zanin [14] proposed hybrid data mining/complex network, classification algorithm, able to detect illegal instances in a real card transaction data set

Artificial Immune Recognition System (AIRS) both self/non-self-cells and detector cells are represented as feature vectors. In order to reduce redundancy. In 2012, Halvaiee developed a novel model for credit card fraud detection using AIS and introduced a new model called AIS-based Fraud Detection Model (AFDM), which increase the accuracy up to 25%, reduce the cost up to 85%, and decrease system response time up to 40% compared to the base algorithm [19].

Genetic Algorithm (GA) is an inspired from natural evolution, randomly generated rules are considered as an initial population. Scatter Search (SS) is an evolutionary algorithm, which shares some common characteristics with the GA. Duman [8] suggested a novel combination of the two well-known meta-heuristic approaches, namely GA, SS. This method tried to improve performance by just playing with the values of the parameters and the statistics related with the popular and unpopular regions for a credit card holder were found to be most important.

Neural Network (NN): Self-Organization Map (SOM) is a neural network technique but used as unsupervised learning. It allows users to visualize data from high to low dimension. In 2008, Quah and Sriganesh [6] proposed, for real-time fraud detection, a new and innovative approach based on self-organization map. It helps in identifying new hidden patterns in input data, the filtering of transactions and reduces the overall cost as well as processing time.

Neuro-fuzzy inference system (NFIS), this technique allows utilizing the advantage of both neural networks and fuzzy inference system. In 2017, S. Shaji developed a model a hybrid of neural networks along with fuzzy inference to improve the efficiency of fraud detection process. The training time required for the proposed model increases as the number of samples increases. Despite, time taken for testing after the initial training is very less [11].

Dempster Shafer Theory (DST) or evidence theory is a general framework for reasoning with uncertainty. The role of DST is to combine evidences from the rules R1and R2 and compute an overall belief value for each transaction. Panigrahi [7] investigated a fusion approach using Dempster–Shafer theory and Bayesian learning, the positive point in this approach that the architecture has been kept flexible so that new rules using any other effective technique can also be included at a later stage.

Cost-sensitive decision tree (CSDT) is an induction algorithm developed to identify fraudulent credit card transactions. In the well-known decision tree algorithms, the splitting criteria are either insensitive to costs and class distributions or the cost is fixed to a constant ratio. In 2013, Sahin [18] developed a methodology for fraud detection using decision tree and showed that it outperforms the models built using the traditional data mining methods such as decision trees, ANN and SVM.

Aggregation: the authors said that Transaction aggregation is advantageous in many researches but not in all circumstances. Whitrow [15] have developed a framework for transaction aggregation, using a variety of classification methods and a realistic cost-based performance measure for credit card fraud detection, in his approach he found that aggregation period has a major impact upon the performance of classifiers for fraud detection

Logistic Regression (LR): Density based spatial clustering of applications with noise (DBSCAN) is a density based clustering algorithm which can be used to filter out outliers and discover clusters of arbitrary shapes. Hidden Markov Model (HMM) differs from the normal statistical Markov model by having invisible states, but each state randomly generates one of the visible states. LR is a type of generalized linear model. In 2016, Dai [10] developed a hybrid framework with Big Data technologies to improve the accuracy, handle data storage, model training, data sharing and online detection.

Synthetic Minority Over-sampling Technique (SMOTE) [24] is used to balance the class ratio by generating synthetic instances of fraudulent transactions [29]. Gaussian Mixture Models (GMM) is used to segment the distribution space of continuous attributes as a means to find possible adversarial strategies. In 2017, MF Zeager developed a game theoretical adversarial learning approach in order to model the fraudster's best strategy, using a logistic regression classifier as the fraud detection mechanism, this approach give the credit card company the ability to pre-emptively react to the changing transaction strategies

K-nearest Neighbour (KNN) works extremely well in credit card fraud detection systems using supervised learning techniques. Outlier detection (OD) [22] is used to detect unusual behaviour of a system using a different mechanism. Malini implemented KNN algorithm and outlier detection methods to optimize the best solution for the fraud detection problem; he showed that KNN can suit for detecting fraud with the limitation of memory.

Active learning (AL), which can be considered as a specific instance of semi-supervised learning [30] [29], is to select unlabelled training samples which, once labelled, can improve the accuracy. F. Carcillo, investigated the combination of semi supervised and active learning techniques in the context of streaming fraud detection [25]. This approach had increase the fraud detection accuracy by up to five percent.

Invariant Diversity (ID) is based on the concept that the more generic an attribute is, the more diversity will be seen on transactions associated with that attribute. In 2017, R. Laurens, proposed ID as a unique way of combining the strength of both generic and specific correlation, tested on real online merchant transactions and it managed to find several instances of previously undetected fraudulent transactions using a common rule-based approach [12].

Cardwatch (CW) [26] is based on a neural network, consists of five main modules: Global Constants, Graphical User Interface, Database Interface, and Learning Algorithm Interface. Aleskerov proposed this technique for credit card fraud detection and shown a good rate for detection. The strong point of this implementation was the ability to be easily extensible and able to work directly on a large variety of commercial databases.

Deep Learning (DL) presents a promising solution to the problem of credit card fraud detection by enabling institutions to make optimal use of their historic customer data as well as real-time transaction details that are recorded at the time of the transaction [28]. In 2017, Rushin proposed a comparative study between DL, LR and Gradient Boosted Tree [27] and found that deep learning has the largest value for the majority of the feature sets. In 2018, Roy evaluated different DL topologies and showed that, the Long Short-term Memory (LSTM) and Gated Recurrent Units (GRUs) model significantly outperformed the baseline ANN, which indicates that order of transactions for an account contains useful information in differentiating between fraud and non-fraudulent transactions.

4. Related works review

In this section, we will analyse the contribution of each technique and its effectiveness, in order to find a promising combination for future work. We first listed works for credit

card fraud detection, extract the most important criteria, description of the dataset and validation of implemented solution. Then, we study the implemented methodology, his contribution and limitless. Finally, we summarize our finding in form of tables, to have a best visualization of our results.

4.1 Criteria

In our comparative tables, we regroup and synthesize most used criteria in anterior credit card fraud detection works in order to have most complete comparison:

- Fraud category; means that the proposed solution is dedicate for which category:
  - Online (O): fraud of e-commerce transaction.
  - Merchant (M): fraud when transaction is initiated from a terminal in point of service of a merchant.
  - All (A): all type of fraud even from ATM.
- Real time: parameter show if the technique bellow is able to run in real time.
- Accuracy: a validation parameter of precision (TP+TN)/(TP+FP+TN+FN)
- True positive: the rate of fraudulent case, which is a fraud.
- Dataset size.
- Dataset type: particular, standard or generic.
- Validation: if the approach was implemented for real data (I), under validation (U), or proposed solution (P).

TP: true positive / TN: true negative / FP: False positive / FN: False negative

4.2 Analysis and discussion

In this paragraph, we present three tables, for each category of fraud, we give a summary of techniques and observations belong this study.

Table 1: Contributions and limitations of techniques applied to online credit card frauds

Technique	Real time	Precision		Data set		Validation	Reference
		Accuracy (%)	TP (%)	Size	Type		
SOM +NN + RI	R	NA	NA	over 200 million customers	P	U	[6]
DST + NB	R	NA	98	NA	G	U	[7]
GA + SS	NR	NA	NA	100,000 fraudulent	P	V	[8]
SOM	NR	100	NA	10,000 accounts of selected credit card (1.01.2005 - 1.03.2005)	P	U	[9]
DBSCAN+ HMM + LR	R	NA	NA	5 dataset, 10,000 ,... 1,000,000 within one year.	G	V	[10]
NFIS	R	NA	NA	contains 20 attributes and 1000 instances	S	U	[11]
ID	R	NA	NA	9,547 transactions	P	V	[12]

The first table presents techniques applied to online credit card fraud, which is considered as the most critical and spreader fraud in credit card fraud. We can observe that, not all of them are able to work in real time; even they are used for online detection.

Invariant Diversity technique is tested on real online merchant transactions, and managed to find several instances of previously undetected fraudulent transactions using a common rule-based approach.

Results based on accuracy, reveal that, SOM clustering helps in identifying new hidden patterns in input data, and the filtering of transactions for further review reduces the overall cost as well as processing time, outperforms other techniques and works well in real time.

In addition, we observe that, NFIS's training time required for the proposed system increases as the number of samples increases but time taken for testing after the initial training is very less.

Table 2: Contributions and limitations of techniques applied to merchant credit card frauds

Technique	Real time	Precision		Data set		Validation	Reference
		Accuracy (%)	TP (%)	Size	Type		
AR + FAR	R	NA	NA	12,107 transactions	P	U	[13]
PN+ ANN + MPL	NR	NA	NA	Transactions 01-2011 →12- 2012.	P	V	[14]

The table 2, is dedicated to techniques implemented for merchant credit card frauds, for transactions of credit card, which done from a service point of a specific merchant. We

found that, FAR methodology overcomes the difficulties of minimum support and confidence, optimizes the execution times, reduces the excessive generation of rules, and helps make the results more intuitive, thereby facilitating the work of fraud analysts. We can see that, ANN-MPL helps to anticipate and quickly detect fraud.

The table 3 presents techniques used to detect all type of credit card fraud even from ATM. The results of SVM technique showed that, while sensitivity and accuracy decreased with lower proportions of fraud in the training data, precision showed an opposite trend.

For Fuzzy-ID3, the applied methodology overcomes the difficulties of minimum support and confidence, optimizes the execution times, reduces the excessive generation of rules, and helps make the results more intuitive, thereby facilitating the work of fraud analysts.

Halvaiee introduced a new model called AIS-based Fraud Detection Model (AFDM), which increase the accuracy up to 25%, reduce the cost up to 85%, and decrease system response time up to 40% compared to the base algorithm.

Table 3: Contributions and limitations of techniques applied to merchant credit card frauds

Technique	Real time	Precision		Data set		Validation	Reference
		Accuracy (%)	TP (%)	Size	Type		
RF + AG	NR	NA	NA	175 million transactions 1.1million transactions	P	V	[15]
SVM + AG	NR	NA	NA				
LR + AG	NR	NA	NA				
KNN + AG	NR	NA	NA				
SVM	NR	95,3	NA	2420 fraudulent transactions	G	U	[16]
RF	NR	90,8	NA				
LR	NR	94,2	NA				
AIS	NR	80	88*	640 361 total transactions	P	U	[17]
SVM	NR	NA	90.0	978 fraudulent records 22 million normal transactions	P	V	[18]
CART	NR	NA	83,1				
CSDT	NR	NA	92.1				
AIRS + CC	NR	NA	83*	3.74% fraudulent transactions	P	V	[19]
DT	NR	91,03	NA	NSL-KDD dataset [34]	S	U	[20]
NB	NR	99,02	NA				
ANN + MPL	NR	99,47	NA				
SVM	NR	98,8	NA	NSL-KDD	S	V	[21]
NB	NR	74,9	NA				
KNN + OD	NR	NA	NA	NA	P	P	[22]
FL+ ID3	NR	89	NA	NA	P	U	[23]
SMOTE + GMM + LR	R	84	NA	86 million transactions in one year	P	V	[24]
AL + HRQ	R	NA	NA	12 million transactions	P	U	[25]
CW	R	NA	85	323 transactions	G	U	[26]
DL	R	87.5	NA	80 million account transactions	P	U	[27]
LR	R	82.4	NA				
GBT	R	86.4	NA				
LTSM	R	91.2	NA	80 million account transactions	P	U	[28]
RNN		90.4	NA				
GRU		91.6	NA				
ANN		88.2	NA				

From last two records in table 3, we can observe that deep learning (DL) has the largest value comparing to LR and GBT. In addition, for DL topologies it showed that, the Long Short-term Memory (LSTM) and Gated Recurrent Units (GRUs) model significantly outperformed the baseline ANN.

OD and FL work fast and well on online large datasets, while NN requires a high computing power for learning and functioning, which makes it makes it unfit to operate in real time. Complex networks can be used as a way to improve data mining models; they may be integrated as complementary tools, in order to improve the clustering rates obtained by classical data mining algorithms. Huge

challenge in several works was the imbalanced dataset, the use NSL-KDD dataset, show that it is a best candidate data set to simulate and test the performance of IDS.
As can be seen, the detection of credit card fraud uses several ML techniques, especially those of artificial intelligences and combines them with optimization techniques such as aggregation. However, the majority of these techniques provide results based on a particular dataset, which is itself characterized by unbalanced data.

5. Statistical review

In this section, we present our finding in form of graphs, to give a statistical review of the most important parameters as used parameters by techniques, distribution of techniques by years and used techniques by category.
The graph in figure 1 presents a repartition of techniques, that is able, or not, to work in real time.

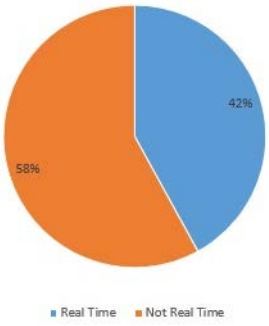

Fig. 1 Techniques ability in real time.

As we can see, not all techniques can work fast, in real time, to detect credit card fraud, and to give a decision during the transaction processing time.
In graph below, we present parameters considered by techniques for credit card fraud detection.

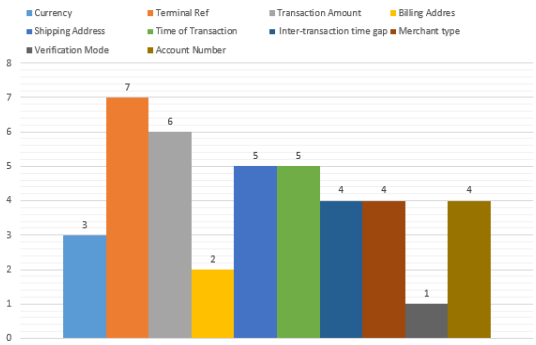

Fig. 2 Parameters used by techniques.

We can observe that, the most used parameters were Terminal reference, Transaction amount, Shipping address and inter-transaction time gap.
These parameters are qualified as the most significant, and can give us a transaction certificate, the others like Transaction time, Merchant type and Account number; can be examined to be more specific.
In figure 3, we give the distribution of reviewed works, in this paper, over last ten years.

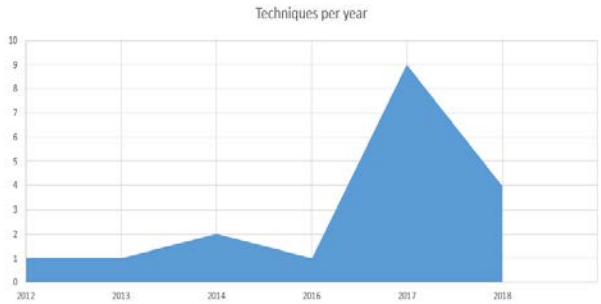

Fig. 3 Distribution of works by years.

It should be noted that work on credit card frauds has become more and more common in recent years, given the current need, which is in continue grow.
This last graph present a summary of used techniques by category.

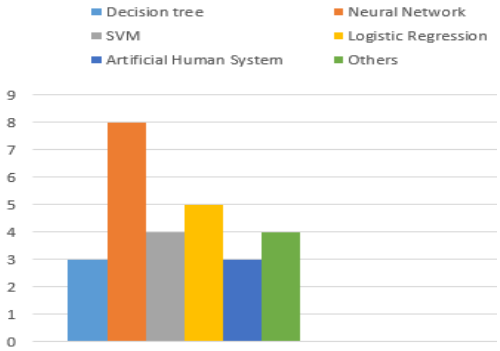

Fig. 4 techniques Used by category.

We can deduce that neural network category was the most used by anterior works as SOM, DL, ANN… following by both Logistic regression and SVM.

4. Conclusion and future work

In this study, we investigated the ML techniques for credit card fraud detection. We analyse the latest articles in this field, made tables and graph, synthetize our finding. The aim of this work, to have a global view of techniques in

anterior researchs, to evaluate contributions and shortcoming of these techniques, and to define the real need for credit card fraud detection. We can conclude that, there is a need for effective technologies to detect fraud in order to maintain the viability of the payment system. The main challenge identified by most of them is:

- Detect frauds in a huge dataset where the legal transactions are more and the fraudulent transactions are bare minimum or close to negligible.
- Ability to minimize false alarms. The goal of a reliable detection system is to learn the behaviours of users dynamically to minimize its own loss.
- How to improve detection accuracy.

From our study, we show that the most used and adequate technique to detect credit card fraud, is Neural Network category, and the use NSL-KDD dataset, show that it is a best candidate data set to simulate and test the performance of IDS.

In our future work we attempt to propose a hybrid model that is both able to handle imbalanced dataset and the real-time problem (to have a response during the financial transaction runtime) with an improved accuracy. We will use results of this study, as NN category in our model and testing with NSL-KDD dataset to achieve our objective.

References

[1] Moroccan Monetary Activity Report as of March 31, 2018, CMI Website.

[2] Wang, S. A Comprehensive Survey of Data Mining-Based Accounting-Fraud Detection Research. International Conference on Intelligent Computation Technology and Automation, vol. 1, pp.50-53, 2010.

[3] Ngai, E.W.T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature, Decision Support System (2010), doi:10.1016/j.dss.2010.08.006.

[4] (Date last accessed 15-July-2014). Online Available: http://www.acfe.com/uploadedfiles/acfewebsite/content/documents/rttn-2010.pdf.

[5] https://www.fico.com/blogs/analytics-optimization/5-keys-to-using-ai-and-machine-learning-in-fraud-detection/

[6] Quah.J and Sriganesh.M, 2008, Real-time credit card fraud detection using computational intelligence, Elsevier, Expert Systems with Applications, Volume 35, Issue 4, 1721-1732.

[7] Panigrahi.S, Kundu.A, Sural.S and Majumdar.A.K, 2009, Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning, Elsevier, Information Fusion, Volume 10, Issue 4, 354-363.

[8] Duman.E and Ozcelik.H, 2011, Detecting credit card fraud by genetic algorithm and scatter search, Elsevier, Expert Systems with Applications, Volume 38, Issue 10, 13057-13063.

[9] Olszewski.D, 2014, Fraud detection using self-organizing map visualizing the user profiles, Elsevier, Knowledge-Based Systems, Volume 70, 324-333.

[10] Dai.Y, Yan.J, Tang.X, Zhao.H and Guo.M, 2016, Online Credit Card Fraud Detection: A Hybrid Framework with Big Data Technologies, IEEE, Trustcom/BigDataSE/ISPA, 1644-1652.

[11] J.Shaji, Panchal.D, 2017, Improved Fraud Detection in e-Commerce Transactions, IEEE, Communication Systems, Computing and IT Applications (CSCITA), 121-126.

[12] Laurens.R, Jusak.J and Zou.C, 2017, Invariant Diversity as a Proactive Fraud Detection Mechanism for Online Merchants, IEEE Global Communications Conference.

[13] Sanchez.D, Vila. M.A, Cerda.L and Serrano. J.M, 2009, Association rules applied to credit card fraud detection, Elsevier, Expert Systems with Applications, Volume 36, Issue 2, Part 2, 3630-3640.

[14] Zanin.M, Romance.M, Moral.S and Criado.R, 2017, Credit card fraud detection through parenclitic network analysis, arXiv; 1706.01953v1, 1-8.

[15] Whitrow.C, Hand. D. J, Juszczak.P, Weston.D and Adams. N. M, 2009, Transaction aggregation as a strategy for credit card fraud detection, Springer, Data Mining and Knowledge Discovery, Volume 18, Issue 1, 30-55.

[16] Bhattacharyya.S, Jha.S, Tharakunnel.K and Westland.J.C, 2011, Data mining for credit card fraud: A comparative study, Elsevier, Decision Support Systems, Volume 50, Issue 3, p602-613.

[17] Wong.N, Ray.P, Stephens.G and Lewis.L, 2012, Artificial immune systems for the detection of credit card fraud: an architecture, prototype and preliminary results, Information Systems Journal, Volume22, Issue1, 53-76.

[18] Sahin.Y, Bulkan.S and Duman.E, 2013, A cost-sensitive decision tree approach for fraud detection, Elsevier, Expert Systems with Applications, Volume 40, Issue 15, 5916-5924.

[19] Halvaiee.N.S and Akbari.M.K, 2014, A novel model for credit card fraud detection usingArtificial Immune Systems, Elsevier, Applied Soft Computing, Volume 24, 40-49.

[20] Mubalik.A (Mubarek) and Adali.E, 2017, Multilayer Perception Neural network technique for fraud detection, IEEE, Computer Science and Engineering (UBMK), International Conference, 383-387.

[21] Dhanabal.L and Dr.Shantharajah.S.P, 2015, A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 6, 446-452.

[22] Malini.N and Pushpa.M, 2017, Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection, IEEE, Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Third International Conference.

[23] Askari.S and Hussain.A, 2017, Credit Card Fraud Detection Using Fuzzy ID3, IEEE, Computing, Communication and Automation (ICCCA), 446-452.

[24] Zeager. M.F, Sridhar.A, Fogal.N, Adams.S, Brown. D.E, and Beling. P.A, 2017, Adversarial Learning in Credit Card Fraud Detection, IEEE, Systems and Information Engineering Design Symposium (SIEDS), 112-116.

[25] F.Carcillo, Le Borgne.Y, Caelen.O and Bontempi.G, 2017, An Assessment of Streaming Active Learning Strategies for Real-Life Credit Card Fraud Detection, Springer, International Journal of Data Science and Analytics, 1–16.

[26] Aleskerov. E, fieisleben.B and Rao.B, Cardwatch A Neural Network Based Database Mining Systern for Credit Card Fraud Detection, IEEE, Computational Intelligence for Financial Engineering (CIFEr), 220-226, 1997.
[27] G. Rushin, C. Stancil, M. Sun, S. Adams, P. Beling, Horse Race Analysis in Credit Card Fraud—Deep Learning, Logistic Regression, and Gradient Boosted Tree, IEEE, Systems and Information Engineering Design Symposium (SIEDS), 117- 121, 2017.
[28] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, P. Beling, Deep Learning Detecting Fraud in Credit Card Transactions; IEEE, Systems and Information Engineering Design Symposium (SIEDS), 129-134, 2018.
[29] Chapelle.O, Scholkopf.B, and Zien.A, 2009, Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews], IEEE Transactions on Neural Networks, vol. 20, no. 3, 542–542.
[30] Settles.B, 2010 , Active learning literature survey, University of Wisconsin, Madison, vol. 52, no. 55-66, 11.

Imane SADGALI received the engineer degree in software engineering from INPT, Morocco, in 2009, and worked for eight years for a payment system company. Currently, she is preparing her PhD in computer Science in faculty of Science Ben M'sik. Her research interests card fraud detection and prevention using machine learning.
Email: sadgali.imane@gmail.com

Nawal SAEL is a professor of Computer Science and member of Computer Science and Information Processing laboratory at faculty of science Ben M'sik (Casablanca, Morocco). She received her Ph.D. in Computer Science from the Faculty of Sciences, University Hassan II Casablanca, Morocco, 2013 and her engineer degree in software engineering from ENSIAS, Morocco, in 2002. Her research interest include data mining, educational data mining, machine learning and Internet of things.
Email: saelnawal@hotmail.com

Faouzia Benabbou is a professor of Computer Science and member of Computer Science and Information Processing laboratory. She is Head of the team "Cloud Computing, Network and Systems Engineering (CCNSE)". She received his Ph.D. in Computer Science from the Faculty of Sciences, University Mohamed V, Morocco, 1997. His research areas include cloud Computing, data mining, machine learning, and Natural Language Processing. She has published several scientific articles and book chapters in these areas.
Email: Faouzia.benabbou@univh2c.ma