A State-of-the-Art Review of Machine Learning Techniques for Fraud Detection Research

SINAYOBYE Janvier Omar
Makerere University
P.O.Box 7062
Kampala-Uganda
sijaom2@gmail.com

KIWANUKA Fred
Makerere University
P.O.Box 7062
Kampala-Uganda
kiwanoah@gmail.com

KAAWAASE KYANDA Swaib
Makerere University
P.O.Box 7062
Kampala-Uganda
kswaibk@gmail.com

ABSTRACT

The area of fraud detection[1] has been traditionally correlated with data mining and text mining. Even before the "big data" phenomena started in 2008, text mining and data mining were used as instruments of fraud detection. However, the limited technological capabilities of the pre-big data technologies made it very difficult for researchers to run fraud detection algorithms on large amounts of data. This paper reviews the existing research done in fraud detection across different areas with the aim of investigating the machine learning techniques used and find out their strengths and weaknesses. It used the systematic quantitative literature review methodology to review the research studies in the field of fraud detection research within the last decade using machine learning techniques. Various combinations of keywords were used to identify the pertinent articles and were retrieved from ACM Digital Library, IEEE Xplorer Digital Library, Science Direct, Springer Link, etc. This search used a sample of 80 relevant articles (peer-reviewed journals articles and conference papers). The most used machine learning techniques were identified, and their strengths and weaknesses. Finally, the conclusion, limitations and future work have been shown.

KEYWORDS

Data mining, systematic literature review, Machine learning, Fraud detection, Areas of fraud.

ACM Reference format:

Omar J. Sinayobye, Fred Kiwanuka, Swaib Kaawaase Kyanda. 2018. A State-of-the-Art Review of Machine Learning Techniques for Fraud Detection Research. In SEIA '18: SEIA '18: Symposium on Software Engineering in Africa, May 27–28, 2018, Gothenburg, Sweden. ACM , 8 pages. *https://doi.org/10.1145/3195528.3195534*

1 INTRODUCTION

Nowadays, most organizations, private companies and government agencies have adopted Electronic commerce to increase their productivity or efficiency in trading products or services; in areas such as Credit card, Telecommunication, Health-care Insurance, Automobile Insurance, Online Auction, etc. [1,2]. Electronic commerce systems are used by both legitimate users and fraudsters; hence they become more vulnerable to large scale and systematic fraud.

The term fraud means to obtain services, goods, and money by the unethical way, and grow a problem in the world today. Fraud is a crime where the purpose is to appropriate money by illegal means. The deals of fraud with cases involving criminal purposes mostly are difficult to identify [3-7]. The main reason behind the commitment of fraud is to achieve gain on false ground by an illegal means. This has a dramatic impact on the economy, law and even the human moral values [8]. Almost all technological system that involves money and services can be compromised by fraudulent acts; for example the credit card, telecommunication, health care insurance, automobile insurance and online auction system [9]. Therefore, frauds in these systems are considered as cyber-crime, causing huge amount of financial losses.

1Produces the permission block, and copyright information
The full version of the author's guide is available as acmart.pdf document
It is a data type.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org
SEIA '18, May 27–28, 2018, Gothenburg, Sweden
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5719-7/18/05...$15.00
https://doi.org/10.1145/3195528.3195534

Fraud detection, being part of the overall fraud control, automates and helps reduce the manual parts of a screening/checking process. This area has become one of the most established industry/government data mining applications. It is impossible to be absolutely certain about the legitimacy of and intention behind an application or transaction. Given the reality, the best cost effective option is to tease out possible evidences of fraud from the available data using mathematical algorithms. Evolved from numerous research communities, especially those from developed countries, the analytical engine within these solutions and software are driven by artificial immune systems, artificial intelligence, auditing, database, distributed and parallel computing, econometrics, expert systems, fuzzy logic, genetic algorithms, machine learning, neural networks, pattern recognition, statistics, visualization and others[10].

There are many surveys in this field that cover similar topics of fraud detection using data mining techniques as the main approach [11-16]. There are also plenty of specialized fraud detection solutions and software which protect businesses such as credit card, e-commerce, insurance, retail, telecommunications industries. There are often two main criticisms of data mining-based fraud detection research: the deficiency of public available real data in this domain to perform experiments on I.e Obtaining appropriate access to financial data to perform research in this area is extremely difficult due to privacy and competitive reasons, and the lack of published well-researched methods and techniques.

To counter both of them, In this paper, we surveyed fraud detection related literatures across six areas that frauds usually occur which are Credit card, Telecommunication, Healthcare Insurance, Automobile Insurance, Online Auction and Smart meter data frauds. The contribution of our work in this paper were; to investigate the techniques used and find out their strengths and weaknesses, to provide an up-to-date and comprehensive analysis about the topic and to provide scholars and practitioners with an excellent source of machine learning techniques used in fraud detection research for their fast access and use.

The rest of this paper is outlined as follows. Section 2 contains the related review and survey papers in the fraud detection system. Section 3 addresses the methodology used to generate the SLR, Section 4 interpreting the results and discussion the weaknesses and strengths of the techniques identified, Finally, Section 5 concludes the paper.

2 LITERATURE REVIEW

Fraud detection system is important in several significant and sensitive sectors or areas. Therefore, fraud detection has been the topic of various surveys and review articles; that may be based on topics such as fraud areas, fraud detection approaches and techniques. [1,2,17] surveyed fraud detection done on different areas based on data mining and statistical techniques. [18] reviewed fraud detection utilizing artificial intelligence techniques. In their study, [18] also covered on the challenges that can be faced by Fraud Detection System. [19,20] and [21] surveyed and analyzed fraud detection statistical methods for health care fraud detection.

From another aspect, [22] presented the different types of credit card frauds, such as bankruptcy fraud, counterfeit fraud, theft fraud, application fraud and behavioral fraud, and discussed on the appropriate techniques to fight them; such as a pair wise matching, decision trees, clustering techniques, neural networks, and genetic algorithms. In the same area, [23] analyzed different kind of methods that are used to detect credit card fraud. [24] presented the VoIP fraud problem and surveyed the fraud detection systems proposed in various areas, and their usability in the VoIP context. [25,26] provided a comprehensive survey and review for different data mining techniques used to detect financial fraud. [27] presented an extensive survey for fraud types in medical and motor insurance systems and many types of data mining techniques are used to detect fraud in these insurance sectors. [28] introduce a classification framework for financial fraud detection which expands the typical data mining process and adds into consideration the specific characteristics of financial data for fraud detection.

The survey covers research on data mining detection which includes work on regression, neural networks and statistical tests. [29] surveys the field of accounting fraud detection. The survey focuses more specifically on data mining-based techniques for financial statements. It points to the factor that hampers a proper comparison of these methods is that there is a notorious difference between the data sets, methods and the evaluation techniques.

In previous years, manual fraud audit techniques such as discovery sampling have been used to detect fraud, such as in [30]. These complicated and time consuming techniques transact with various areas of knowledge like economics, finance, law and business practices. Therefore, to raise the effectiveness of detection, computerized and automated Fraud Detection Systems was invented. However, FDS capabilities were limited because the detection fundamentally depends on predefined rules that are stated by experts [19]. More complex FDSs integrating a wide range of data mining methods are required and are being developed for effective fraud detection [31-34]. Data mining involves statistical, mathematical, artificial intelligence and machine learning techniques to

extract and identify useful information and subsequent knowledge from large databases. These systems have three main advantages: fraud pattern are obtained automatically from data; specification of “fraud likelihood ” for each case, consequently that efforts in investigating suspicious cases can be prioritized; and revelation of new fraud types that were not defined before [19].

Data mining methods consist of six main categories which are Classification, Clustering, Regression, Outlier detection, Visualization and Prediction [35]. Each of these methods is supported by specific techniques. For example, neural network technique and support vector machine technique are used for data mining classification method. K-means technique is used for data mining clustering method. Furthermore, data mining has incorporated many techniques from other domains such as statistics, machine learning, pattern recognition, database and data warehouse systems, information retrieval, visualization, algorithms, high-performance computing, and many application domains [36]. Recently, fraud detection integrates anomaly based detection approach and misuse based detection approach by using data mining techniques [37].

As we can see, there are numerous articles that surveyed Fraud Detection System techniques, although almost all existing surveys do not highlight the most machine learning techniques used in FDS across different application domain. Thus, this survey attempts to make a structured and comprehensive overview of the research on fraud detection. This is done by covering the fraud types, fraud detection techniques, as well as the strengths and weaknesses of techniques used in six identified areas: credit card, telecommunications, healthcare insurance, automobile insurance, online auction and smart meter data fraud.

3 METHODOLOGY

We used Systematic Literature Review (SLR) [38] method for conducting this review. A systematic literature review is an evidence-based approach to thoroughly search studies relevant to some pre-defined research questions and critically select, appraise, and synthesize findings for answering the research questions at hand. We followed the SLR guidelines reported in [38]. As part of the review process, we developed a protocol that provides a plan for the review in terms of the method to be followed, including the research questions and the data to be extracted.

3.1 Research questions

This study is aimed at reviewing and summarizing the research done in fraud detection across different fraud areas, investigating the most machine learning techniques used, and finding out their strengths and weaknesses.

Research questions are:

RQ1: *What are the application domains used in fraud detection research?*

RQ2: *Which machine learning techniques are most used in fraud detection research?*

RQ3: W*hat are the strengths and weaknesses of machine learning techniques applied in fraud detection research?*

3.2 Search strategy

According to the guidelines provided in [39] and [40], we define a search strategy to retrieve as many relevant peer-reviewed papers as possible. Our search strategy is described as follows.

3.2.1 Search method

To cover all the relevant publications, we used the following five electronic libraries: IEEEXplore, ScienceDirect, ACM Digital Library, SpringerLink, and Wiley Online Library. These libraries were searched using the search terms introduced in Section 3.2.2

3.2.2 Search terms

We designed the search string according to the guidelines provided in [38]. Based on the research question and pilot studies, we found the following basic:

1. *Data mining **OR** Fraud detection **OR** Areas of fraud.*

2. *Machine Learning Techniques **OR** Fraud detection **OR** Areas of fraud.*

To construct the search string, all these search terms were combined using Boolean “AND” as follows: **1 AND 2.**

The search terms in the string were matched only with the title, abstract, and keywords of the papers in the digital databases.

2.2.3 Data sources

As previously mentioned, we searched five digital databases, which are shown in Table 1.

Table 1. Database sources.

Source	URL
IEEE Xplore	http://ieeexplore.ieee.org/
ScienceDirect	http://www.sciencedirect.com/
ACM	http://portal.acm.org/
SpringerLink	http://link.springer.com/
Wiley	http://onlinelibrary.wiley.com/

The next decision was to find the suitable field (i.e. title, abstract and full-text) to apply the search string on. In our experience, searching in the 'title' alone does not always provide us with all relevant publications. Thus, 'abstract' or 'full-text' of publications should potentially be included. On the other hand, since the search on the full-text of studies results in many irrelevant publications, we chose to apply the search query additionally on the 'abstract' of the studies. This means a study is selected as a candidate study if its title or abstract contains the keywords defined in the search string. In addition, we limited our search to the publications that are written in English and are published after 2007 and We did not use Google scholar due to its low precision of the results and the tendency of returning a large number of irrelevant papers [36].

3.3 Study selection

Some of the studies might contain the keywords used in the search string but might still be irrelevant for our research questions. Therefore, a study selection has to be performed to include only studies that contain useful information for answering the research question. We conducted our review in early December 2017. As a consequence, our review included studies that were published and/or indexed before that date. As shown in Table 2, we first applied the search query on each data source separately and we merged the results obtained from the different data sources, we got 2320 papers. Subsequently. To remove irrelevant studies, we scanned the articles by title and thereby reduced the number of studies to 720. Then, removing duplicate studies we got 342, we read the abstract of each publication carefully and further decreased the number of studies to 171. Finally, we added a list of additional papers recommended by experts and then scanned the full-text of the publications. We checked the full-text of studies to see if they fit with our predefined selection criteria. The result comprised 80 publications that represented our final set of primary studies.

Table 2.Steps followed to scope the search results.

Source	Automatic search	Title scanning	Duplication Removal	Abstract scanning	Full text scanning
IEEE Xplore	672	232	129	65	27
ScienceDirect	599	196	93	43	11
ACM	487	124	65	30	19
SpringerLink	250	90	31	15	9
Wiley	312	78	24	18	14
Results merged	**2320**	**720**	**342**	**171**	**80**

3.4 Data extraction and synthesis

This section describes the process of data extraction from the selected papers and the analysis of the extracted data to answer the research questions of this SLR.

3.4.1 Data extraction

We extracted the required data from the selected papers using a pre-designed data extraction form that included the data items envisaged necessary to answer the research questions of this SLR. The data extraction form is shown in Table 3. The extracted data were recorded in an MS Excel spreadsheet for analysis.

Table 3. Data extraction form.

#	Data Item	Description	Research question
D1	Author(s)	The author(s) of the paper	Demographics Data
D2	Year	Year of publication	
D3	Title	The title of the paper	
D4	Publication type	From journals or conferences	
D5	Publication venue	Where paper is published	
D6	Data sources	Database source	
D7	Fraud detection	Application domain	RQ1
D8	Techniques	ML techniques used	RQ2
D9	discussion	Strengths and weaknesses	RQ3

3.4.2 Data synthesis

We extracted four types of data – demographic data, Fraud detection application domain, ML techniques most used and the most used ML techniques strengths and weaknesses. We analyzed the demographic data using descriptive statistics. The results of our analysis of the demographic data, the data items D7, D8 and D9 have been presented in Section 4.

4 RESULTS INTERPRETATION AND DISCUSSION

4.1 Demographics data.

The reviewed papers were published between 2007 and 2017. Our SLR covers the paper published before 30th December 2017 when the search process for selecting the potentially relevant papers was completed. It used only the peer-reviewed journals and conferences papers. We have classified the reviewed papers based on five databases sources and We have categorized the included papers based on the six application domains(Fraud Areas).

4.2 RQ1 Application domains used in fraud detection research.

This section reports the results based on analysis of the data about Fraud detection research done across different fraud areas. The analysis is meant to answer RQ1, We answer this question from statistical analysis of the number of percentage of papers in each fraud area.

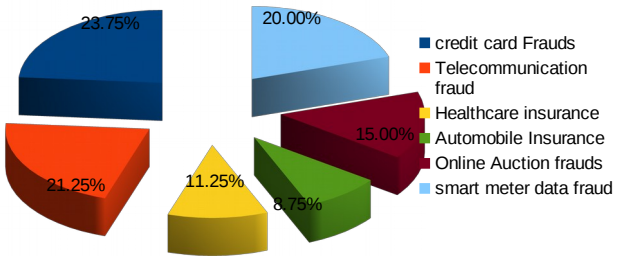

Figure 1: Number of papers reviewed in each fraud area for a decade

As we can observe from Fig.1, credit card Frauds covered 23.75% of the reviewed papers. This fraud is considered as common type of credit fraud that takes many forms [40]. Existing literatures classify them into several categories based into; Offline Credit Card Fraud which Happens when the plastic card is stolen by fraudsters, using it in stores as the actual owner and Online Credit Card Fraud which is A popular and very dangerous fraud, credit cards' information are stolen by fraudsters to be used later in online transactions by internet or phone.

Telecommunication fraud covered 21.25% of the reviewed papers as from Fig. 1, its problem in mobile telecommunications has grown dramatically over the past ten years [41]. It is a complex and dynamic problem for telecommunication operators because these frauds threaten both the prepaid and post-paid services. telecommunication fraud is categorized into : Subscription Fraud, Superimposed fraud, Premium Rate Fraud, Roaming Fraud, Prepaid Fraud, SIM Surfing, SIM Cloning Fraud, SIM BOX/Gateway Bypass, Fraud Private Branch Exchange, Hacking (PBX) Fraud, Voucher Fraud.

Healthcare insurance systems have become the main concern of modern life. The most common known healthcare fraud types are Phantom claims, Duplicate claims, Bill Padding, Upcoding, Unbundling, Excessive or Unnecessary Services Kickbacks, Claims in short time, Unpaid instalments, Incorrect dates, Medications without examination and Excessive numbers of small bills. It covered 11.25% of the reviewed papers.

Automobile insurance fraud become the main issue for companies and consumers. The most common and most addressed issues in the literature on automobile insurance fraud detection system are; Ditching, Past Posting, Vehicle Repair, Vehicle Smuggling, Phantom Vehicles, Staged Accidents, Vehicle Identification Number (VIN) Switch and Rental Car Fraud. It covered 8.75% of the reviewed papers.

Online auction frauds have been classified into six categories by the internet Fraud Complaint Centre (IFCC) which are: non-delivery of goods, misrepresentation of the items, triangulation, fee staking, selling of black-market goods, multiple bidding and shill bidding. The common types of online auction fraud from the victim's viewpoint as presented in [42] are Competitive shilling, Bid shielding, Non-delivery of goods, Multiple bidding, False bids, Bid shading and Credit card phantom. It covered 15% of the reviewed papers.

Smart meter data fraud is a type of fraud dealing with data from smart meters. It covered 20% of the reviewed papers.

4.3 RQ2: Machine learning techniques used in fraud detection research.

This section reports the results of analysis of data machine learning used in fraud detection research in general for this particular review. This part of the analysis is meant to answer RQ2, "Which machine learning techniques are most used in fraud detection research?. It shows the machine learning techniques used only for this SLR, identify the most used for each fraud area and the most used in general for all application domains considered in this review paper.

Table 4.1: Distribution of Machine learning Techniques used in one decade for this review.

ML Techniques	Credit card Fraud	Telecommunication Frauds	Health-care insurance frauds	Automobile insurance Frauds	Online auction frauds	Smart meter data fraud	**Total**
Logistic regression					1		**1**
Hidden makov model	1				1		**2**
ANN	5	2	1	2	3	4	**17**
Adaptive Neuro-Fuzzy		1					**1**
Bayesian NN		1		1			**2**
Decision tree	4	1	3		2	4	**14**
Rule based		2					**2**
Regression Analysis					1		**1**
SVM	3	1	2	1	2	3	**12**
Artificial Immuno system		3					**3**
Naive bayes	1	2		1	1	3	**8**
Fuzzy NN		4					**4**
Association rule analysis	1			1			**2**
Random forest	2		1		1	2	**6**
k-NN	2		2		1		**5**
Total	**19**	**17**	**9**	**7**	**12**	**16**	**80**

As we can observe from Table 4.1, the most frequently used machine learning methods in fraud detection systems; For credit card frauds are Artificial Neural Networks, Decision tree and support vector machine techniques. Fuzzy Neural Network and Artificial Immune system techniques for Telecommunication frauds. Decision tree for Healthcare insurance frauds, Artificial Neural Network for Automobile insurance frauds and online auction frauds. Finally, decision tree in [43-46] is one such widely used machine learning technique that has been effective for classification, its usefulness results from the ability to compensate for missing values and having a highly flexible hypothesis space. A neural network in [47-50] is one of the most common machine learning techniques used in fraud detection due to its noise-robustness and fast response qualities. Support Vector Machines (SVMs) to classify fraudulent activity in [44][51-53].

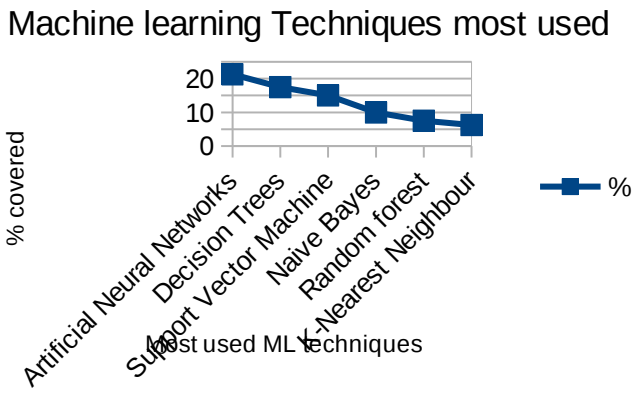

Figure 2: Machine learning Techniques most used in this SLR

As we can observe from Fig. 2, this research demonstrates that Artificial neural network, Decision tree, SVM, Naives Bayes, Random forest and K-Nearest Neighbour techniques with 21.25%, 17.5%, 15%, 10%, 7.5 and 6.25% respectively, are the most machine learning techniques used and it means they can be used as a computational intelligent tools for detecting different types of frauds. They cover 78% of the papers in this review-paper.

4.4 RQ3. The strengths and weaknesses of ML techniques most used

No single machine learning technique can uniformly outperform other technique over all datasets. They are in common having the same strengths and/or weaknesses, the next section talk about each machine learning techniques strengths and weaknesses from evidence of existing empirical and theoretical studies.

Artificial Neural Network's strengths are for better accuracy in general, speed of classification, tolerance to highly interdependent attributes and attempts for incremental learning while it's weaknesses are the speed of learning with respect to number of attributes and the number of instances, tolerance to; missing values, irrelevant attributes, and noise, Dealing with danger of overfitting and Model parameter handling.

Decision Trees's strengths are for speed of classification, dealing with discrete/binary/ continuous attributes, and explanation ability /transparency of knowledge /classifications while its weaknesses are Tolerance to; redundant attributes,

highly interdependent attributes and noise, attempts for incremental learning.

Support Vector Machine's strengths are for Accuracy in general, Speed of classification, tolerance to irrelevant attributes and redundant attributes. But its weaknesses are Speed of learning with respect to number of attributes and the number of instances, Model parameter handling, Explanation ability/ transparency of knowledge/ classifications.

Naive Bayes's strengths are mainly for the speed of learning with respect to number of attributes and the number of instances, speed of classification, Tolerance to missing values, attempts for incremental learning and model parameter handling. But its weaknesses are Accuracy in general, tolerance to redundant attributes and highly interdependent attributes.

Random forest's strengths are for speed of classification, explanation ability/transparency of knowledge/classifications and model parameter handling. but it weaknesses are attempts for incremental learning and tolerance to noise,

K-Nearest Neighbour's strengths are for speed of learning with respect to number of attributes and the number of instances and attempts for incremental learning. But weak at speed of classification, tolerance to; missing values, highly interdependent attributes, noise.

When faced with the decision "Which machine learning technique will be most accurate on a studied problem?", the simplest approach is to estimate the accuracy of the candidate technique on the problem and select the one that appears to be most accurate.

5 CONCLUSION

This review paper suffers from some limitations: First, a decade review may not be sufficient to address this growing problem as it started when the business started. second, the Eighty articles explored may not reveal the entire story of Machine learning techniques usage across the fraud areas investigated, several online databases need to be included in the sample for more powerful presentation and analysis. It is, however, crucial to have a wide-ranging review on fraud detection research to increase the understanding and to expand the knowledge in this field among researchers and professionals.

We have explored the state-of-the-art fraud detection systems in six areas of fraud. Furthermore, the fraud detection techniques have been categorized and reviewed. However, it is noticed that most fraud detection systems in all areas use supervised approach. In addition, the most commonly used machine learning techniques are Artificial Neural Networks (ANN), Decision tree, Support Vector Machines (SVM), Naive Bayes, Random forest and K-NN algorithms. These techniques can be used alone or combined with an ensemble or meta-learning techniques to build strong detection classifiers. Their weaknesses and strengths have been Identified.

Prior research in the area of fraud detection has particularly focused on using private data sets to develop statistical methods, and taking advantage of power of data mining in finding hidden patterns or anomalies in financial data. However it is hard to test and compare some of these methods due to the lack of generally available data. Future work in the area of fraud detection research should focus ; - On shared data sets (available to the general public and researchers) that contains samples of different malicious behavior which will be a great help to compare and test the performance of new methods with the existing ones. - On organizations with useful information regarding fraud ares and data mining techniques may be able to select the suitable technique once considering its particular usage context and frequency. - The concept of combining classifiers is proposed as a new direction for the improvement of the performance of individual classifiers. Besides other benefits, researchers can take advantage of knowing the most used methods and in which context so that they can develop a research project to either investigating such method in a different context or suggesting a new innovative method in a similar context.

REFERENCES

[1] Tareq Allan, Justin Zhan, and South Dako. 2010. Towards Fraud Detection Methodologies e. (2010).

[2] Mirjana Pejic-Bach. 2010. Invited Paper: Profiling Intelligent Systems Applications in Fraud Detection and Prevention: Survey of Research Articles. In 2010 International Conference on Intelligent Systems, Modelling and Simulation. IEEE, 80–85. DOI:http://dx.doi.org/10.1109/ISMS.2010.26

[3] J. P. Linda Delamaire, Hussein Abdou , "Credit card fraud and detection techniques : a review," Banks Bank Syst., vol. 4, no. 2, 2009.

[4] C. Classifier, J. Kim, K. Choi, G. Kim, and Y. Suh, "Expert Systems with Applications Classification cost : An empirical comparison among traditional classifier," Expert Syst. Appl., vol. 39, no. 4, pp. 4013–4019, 2012.

[5] K. Ramakalyani and D. Umadevi, "Fraud Detection of Credit Card Payment System by Genetic Algorithm," International Journal of Scientific & Engineering Research, vol. 3, no. 7, pp. 1–6, 2012.

[6] T. Tian, J. Zhu, F. Xia, X. Zhuang, and T. Zhang, "Crowd Fraud Detection in Internet Advertising," International World Wide Web Conference Committee, 2015.

[7] F. Louzada and A. Ara, "Expert Systems with Applications Bagging k-dependence probabilistic networks : An alternative powerful fraud detection tool," Expert Syst. Appl., vol. 39, no. 14, pp. 11583–11592, 2012.

[8] Panos Alexopoulos, Kostas Kafentzis, Xanthi Benetou, and Tassos Tagaris. 2007. TOWARDS A GENERIC FRAUD ONTOLOGY IN E- GOVERNMENT DETECTION IN THE E-.(2007).

[9] Miguel Pironet San-bento Almeida. 2009. Classification for Fraud Detection with Social Network Analysis . Dissertation for the obtaining of a Masters Degree in Engenharia Informática e de Computadores Júri. (2009).

[10] Aisha Abdallah, Mohd Aizaini Maarof and Anazida Zainal, Fraud Detection System: A survey, Journal of Network and ComputerApplications, http://dx.doi.org/10.1016/j.jnca.2016.04.007

[11] S. Benson Edwin Raj and A. Annie Portia. 2011, "Analysis on credit card fraud detection methods". In: nternational Conference on Computer, Communication and Electrical Technology (ICCCET). IEEE, pp. 152–156.

[12] E Kirkos, C Spathis, and Y Manolopoulos. 2007. "Data Mining techniques for the detection of fraudulent financial statements". In: Expert Systems with Applications 32.4, pp. 995–1003.

[13] E. E. Ngai et al. 2011, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature". In: Decision Support Systems 50.3, pp. 559–569.

[14] A. Sorin. 2012, "Survey of Clustering based Financial Fraud Detection Research". In: Informatica Economica 16.1, pp. 110–123.

[15] A. Sudjianto et al., 2010, "Statistical Methods for Fighting Financial Crimes". In: Technometrics 52.1, pp. 5–19.

[16] J. West and M. Bhattacharya. 2015, "Intelligent financial fraud detection: a comprehensive review". In: Computers & Security 57, pp. 47–66.

[17] Clifton Phua, Vincent Lee, Kate Smith, and Ross Gayler. 2005. A Comprehensive Survey of Data Mining-based Fraud Detection Research. (2005).

[18] Mohammad Behdad, Luigi Barone, Mohammed Bennamoun, and Tim French. 2012. Nature-Inspired Techniques in the Context of Fraud Detection. IEEE Trans. Syst.Man, Cybern. Part C (Applications Rev. 42, 6 (November 2012), 1273–1290. DOI:http://dx.doi.org/10.1109/TSMCC.2012.2215851

[19 Ll. Bermúdez, J.M. Pérez, M. Ayuso, E. Gómez, and F.J. Vázquez. 2008. A Bayesian dichotomous model with asymmetric link for fraud in insurance. Insur. Math. Econ. 42, 2 (April 2008), 779–786. DOI:http://dx.doi.org/10.1016/j.insmatheco.2007.08.002

[20] Peter Travaille, Dallas Thornton, and Roland M. Müller. 2011. Electronic Fraud Detection in the U . S . Medicaid Healthcare Program : Lessons Learned from other Industries. (2011), 1–10.

[21] Han Tao, Liu Zhixin, and Song Xiaodong. 2012. Insurance fraud identification research based on fuzzy support vector machine with dual membership. 2012 Int. Conf.Inf. Manag. Innov. Manag. Ind. Eng. (October 2012), 457–460. DOI:http://dx.doi.org/10.1109/ICIII.2012.6340016

[22] Linda Delamaire, Hussein Abdou, and John Pointon. 2009. Credit card fraud and detection techniques : a review. 4, 2 (2009).

[23] S. Benson Edwin Raj, A. Annie Portia, and Assistant Sg. 2011. Analysis on Credit Card Fraud Detection Methods. (2011), 152–156.

[24] Yacine Rebahi, Mohamed Nassar, Thomas Magedanz, and Olivier Festor. 2011. A survey on fraud and service misuse in voice over IP (VoIP) networks. Inf. Secur. Tech.Rep. 16, 1 (February 2011), 12–19. DOI:http://dx.doi.org/10.1016/j.istr.2010.10.012

[25] E.W.T. Ngai, Yong Hu, Y.H. Wong, Yijun Chen, and Xin Sun. 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decis. Support Syst. 50, 3 (February 2011), 559–569. DOI:http://dx.doi.org/10.1016/j.dss.2010.08.006

[26] Shiguo Wang. 2010. A Comprehensive Survey of Data Mining-Based Accounting-Fraud Detection Research. Intell. Comput. Technol. Autom. (ICICTA), 2010 Int. Conf. 1(May 2010), 50–53. DOI:http://dx.doi.org/10.1109/ICICTA.2010.831

[27] H. Lookman Sithic and T. Balasubramanian. 2013. Survey of Insurance Fraud Detection Using Data Mining Techniques. , 3 (2013), 62–65.Sparrow. 2000. License To Steal: How Fraud bleeds America's health care system,

[28] S. Wang. 2010, "A Comprehensive Survey of Data Mining-Based Accounting-Fraud Detection Research". In: 2010 International Conference on Intelligent Computation Technology and Automation. Vol. 1. IEEE, pp. 50–53.

[29] D. Yue, X. Wu, and Y. Wang. 2007, "A Review of Data Mining-Based Financial Fraud Detection Research". In: 2007 Wireless Communications, Networking and Mobile Computing. Ieee,, pp. 5514–5517.

[30] Sharon Tennyson and Pau Salsas-forn. 2008. IN AUTOMOBILE CLAIMS AUDITING INSURANCE : FRAUD AND DETERRENCE DETECTION OBJECTIVES. 69, 3 (2008), 289–308.

[31] John Akhilomen. 2013. Data mining application for cyber credit-card fraud detection system. In Lecture Notes in Engineering and Computer Science. 1537–1542.

[32] T.A.O. Guo and Gui-yang Li. 2008. NEURAL DATA MINING FOR CREDIT CARD FRAUD DETECTION. , July (2008), 12–15.

[33] Francisca Nonyelum Ogwueleka. 2011. DATA MINING APPLICATION IN CREDIT CARD FRAUD DETECTION SYSTEM. 6, 3 (2011), 311–322.

[34] Anita B. Desai and Ravindra Deshmukh. 2013. Data mining techniques for Fraud Detection. 4, 1 (2013), 1–4.

[35] P. Saravanan, V. Subramaniyaswamy, N. Sivaramakrishnan, M. Arun Prakash And, and T. Arunkumar. 2014. Data Mining Approach For Subscription-Fraud Detection in Telecommunication Sector. 7, 11 (2014), 515–522.

[36] N.M.M. Noor, S.H.a Hamid, R. Mohemad, M.a Jalil, and M.S. Hitam. 2015. A Review on a Classification Framework for Supporting Decision Making in Crime Prevention. J. Artif. Intell. (2015). DOI:http://dx.doi.org/10.3923/jai.2015.17.34

[37] Han Tao, Liu Zhixin, and Song Xiaodong. 2012. Insurance fraud identification research based on fuzzy support vector machine with dual membership. 2012 Int. Conf.Inf. Manag. Innov. Manag. Ind. Eng. (October 2012), 457–460. DOI:http://dx.doi.org/10.1109/ICIII.2012.6340016

[38] M. Sasirekha, I. Sumaiya Thaseen, and J. Saira Banu. 2012. An Integrated Intrusion Detection System for Credit Card Fraud Detection. (2012), 55–60.

[39] Kitchenham, B. and S. Charters, Guidelines for performing systematic literature reviews in software engineering, in Technical report, Ver. 2.3 EBSE Technical Report. EBSE. 2007, sn.

[40] Zhang, H., M.A. Babar, and P. Tell, Identifying relevant studies in software engineering. Information and Software Technology, 2011. 53(6): p. 625-637.

[41] Giannis Potamitis. 2013. Design and Implementation of a Fraud Detection Expert System using Ontology- Based Techniques A dissertation submitted to the University of Manchester.

[42] Byungtae Lee, Hyungjun Cho, Myungsin Chae, and Seonyoung Shim. 2010. Empirical analysis of online auction fraud: Credit card phantom transactions☆. Expert Syst. Appl.

37, 4 (April 2010), 2991–2999. DOI:http://dx.doi.org/10.1016/j.eswa.2009.09.034

[43] I. Monedero, F. Biscarri, C. Leon, J. I. Guerrero, J. Biscarri, and R. Millan,2012. “Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees,” Int. J. Electr. Power Energy Syst.,.

[44] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, 2016. “Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid,” IEEE Trans. Ind. Informatics.

[45] Félix Biscarri, Iñigo Monedero, Carlos León, Juan I. Guerrero, 2013. "FRAUD Detection In Electric Power Distribution Networks Using An ANN-Based Knowledge-Discovery Process"International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 4, No. 6.

[46] B. Coma-Puig, J. Carmona, R. Gavalda, S. Alcoverro, and V. Martin, 2016. “Fraud detection in energy consumption: A supervised approach,” Proc. - 3rd IEEE Int. Conf. Data Sci. Adv. Anal. DSAA 2016, pp. 120–129.

[47] J. E. Cabral, J. O. P. Pinto, E. M. Martins, and A. M. A. C. Pinto, 2008. “Fraud Detection in High Voltage Electricity Consumers Using Data Mining,” pp. 3–7.

[48] B. C. Costa, B. L. a Alberto, A. M. Portela, M. W, and E. O.Eler, 2013. “Fraud Detection in Electric Power Distribution Networks using an Ann-Based Knowledge-Discovery Process,” Int. J. Artif. Intell. Appl., vol. 4, no. 6, pp. 17–23.

[49] I. Monedero, F. Biscarri, C. Leon, J. Biscarri, and R. Millan, 2006, “MIDAS: Detection of Non-Technical Losses in Electrical Consumption Using Neural Networks and Statistical Techniques,” in Proceedings of the Computational Science and Its Applications Conference – ICCSA, Vol. 3984, pp. 725-734, M.

[50] V. Ford, A. Siraj, W. Eberle, 2014. "Smart grid energy fraud detection using artificial neural networks", IEEE Symposium on Computational Intelligence Applications in Smart Grid, 2014, pp.1,6.

[51] H. Pok, K. S. Yap, and I. Abidin, 2007. “Abnormalities and fraud electric meter detection using hybrid support vector machine and modified genetic algorithm,” Proc. 19th …, no. 291, pp. 21–24.

[52] J. Nagi, A. Mohammad, S. Yap, S. Tiong, and S. Ahmed, 2008. “Non-technical loss analysis for detection of electricity theft using support vector machines,” IEEE Int. Conf. Power Energy, no. PECon 08, pp. 907–912.

[53] S. Shekara, S. Reddy, L. Wang, and V. Devabhaktuni, 2011. “Support Vector Machine Based Data Classification for Detection of Electricity Theft,” pp. 1–8.