International Journal of Research Studies in Computer Science and Engineering (IJRSCSE)
Volume 3, Issue 4, 2016, PP 1-11
ISSN 2349-4840 (Print) & ISSN 2349-4859 (Online)
DOI: http://dx.doi.org/10.20431/2349-4859.0304001
www.arcjournals.org

Fraud Analytics Using Data Mining

Sreeparna Mukherjee
MCA, MACS Department
NIT-Karnataka
Surathkal, India

Triparna Mukherjee
Department of Computer Science
St. Xavier's College (Autonomous)
Kolkata, India

Dr. Asoke Nath
Department of Computer Science
St. Xavier's College (Autonomous)
Kolkata, India

Abstract: *This paper deals with the data mining techniques used to combat fraud in different industries. The main challenges in fraud analytics have been addressed and the type of fraudsters involved have been studied in the light of different situations. The two main types of data mining methods i.e. supervised and unsupervised learning methods used in fraud analytics have been intensively reviewed with the help of existing literature. This is aimed to help in recognizing which data mining techniques help the most in fraud analysis in respective domains.*

Keywords: *Data mining, fraud pattern detection, clustering, supervised learning, unsupervised learning, stratified sampling, K-means algorithm etc.*

1. INTRODUCTION

Fraud includes a person or a community who/which deliberately act furtively to deny others something of worth, for their own advantage. Extortion is as old as humankind itself and is capable of boundless assortment of various structures. Nonetheless, as of late, the advancement of new innovations has additionally given further routes in which offenders may submit misrepresentation. Notwithstanding that, business reengineering, revamping or cutting back may debilitate or dispense with control, while new information frameworks may introduce extra chances to confer extortion. Recent research uncovers that data fraud influences a huge number of individuals a year, costing the victims innumerable hours and money in character recuperation and repair. What causes this example of online robbery and extortion? It's a mix of elements: an absence of purchaser information in regards to ensuring that you can protect your character online; developing solace with, and trust in, social stage suppliers; the requirement for social stages to produce income; and an absence of models or policing of these norms. Despite the fact that this issue is not yet in the standard awareness, it likely will be within the near future. Fraud is an ever-growing billion dollar industry. It has been suggested by the PwC global economic crime survey carried out in the year 2009 that almost 30 percent of companies worldwide have informed of being victimized by fraud. [1]

2. MAIN CHALLENGES IN FRAUD ANALYTICS

Today's intense monetary atmosphere is driving a surge in first party fraud for some associations, whilst identity fraud, the complexity of fraudsters and cybercrime are additionally on the expansion. The key is to work intimately with customers and learn to deal with fraud risk management and build up a multi-layered procedure that is comparable with danger and quality all through the business.

- The first concerns the way that fraud is uncommon. Autonomous of the definite setting or application, just a minority of the included populace of cases regularly concerns misrepresentation, of which moreover just a predetermined number will be known as recognized extortion. This truly convolutes the estimation of investigative models.

- Fraud detection frameworks enhance and learn by illustration. In this manner the systems and traps fraudsters embrace develop in time alongside, or better in front of misrepresentation location components. Fraudsters attempt to mix into the earth and not act not the same as others all together not to get saw and to stay secured by non-fraudsters. This viably makes fraud subtly hid, since fraudsters do succeed by being secluded from everything by well considering and arranging how to exactly confer extortion. By receiving and creating progressed logical misrepresentation location and anticipation instruments, associations do figure out how to decrease misfortunes because of extortion since fraudsters, as different lawbreakers, tend to search for the easy way and will search for other, simpler open doors.
- A key test concerns the dynamic way of fraud. Fraudsters attempt to always out beat location and aversion frameworks by growing new procedures and strategies. Accordingly versatile analytical models, identification and preventive frameworks are required, so as to identify and resolve misrepresentation at the earliest opportunity. Identifying misrepresentation as ahead of schedule as could reasonably be expected is crucial. The key thought here is to confirm whether the extortion show still performs as per the organization standards. Changing extortion strategies can create a concept drift suggesting that the relationship between the objective fraud marker and the changes in data are on an on-going basis.
- Fraud is regularly a precisely composed offence, implying that fraudsters frequently don't work autonomously, have partners, and may impel impersonators. In addition, a few extortion sorts, for example, government evasion, and money laundering and carousel frauds include complex structures that are set up with a specific end goal to submit extortion in a composed way. This makes extortion not to be a disconnected occasion and makes it very important to consider the context as well. This makes the analytical models very complicated. [2][3]

Techniques used for fraud detection can be divided into two primary classes: statistical techniques and artificial intelligence. [4] This paper discusses the role of data science in fraud analytics.

2.1. The Role of Data Anlytics in Fraud Analysis

Big data analytics tools and technologies are used in combatting threats. These techniques combine text mining, machine learning and ontology modelling to help in secured threat prediction, detection and prevention at an early stage. Intelligence led investigation processes are much at ease with the help of these techniques and through improved collaborative systems threats can easily be detected. Organizations are hence opting to move away from the conventional firewall and endpoint vendor techniques to adopting big data and cloud solutions to maintain security in the organization. Data analytic techniques can thus be concluded to have a very important part to play in proactive prediction, identification and discerning of fraud. These strategies can permit The association can be allowed to eliminate, scrutinize, decipher and change business information to recognize potential occasions of extortion and fraud depending on reports generated by these systems. It would thus be possible to realize efficacious fraud monitoring projects with the help of these hands-on strategies. [5]

In a nutshell using data analytics in fraud prevention would have the following benefits

- Improved efficiency – Automated method for detecting and monitoring potentially fraudulent behavior.
- Repeatable tests – Repeatable fraud tests that can be run on your data at any time.
- Wider coverage – Full coverage of testing population rather than ‘spot checks’ on transactions – better chance of finding exceptional items.
- Early warning system – Analytics solutions can help you to quickly identify potentially fraudulent behavior before the fraud becomes materialized.

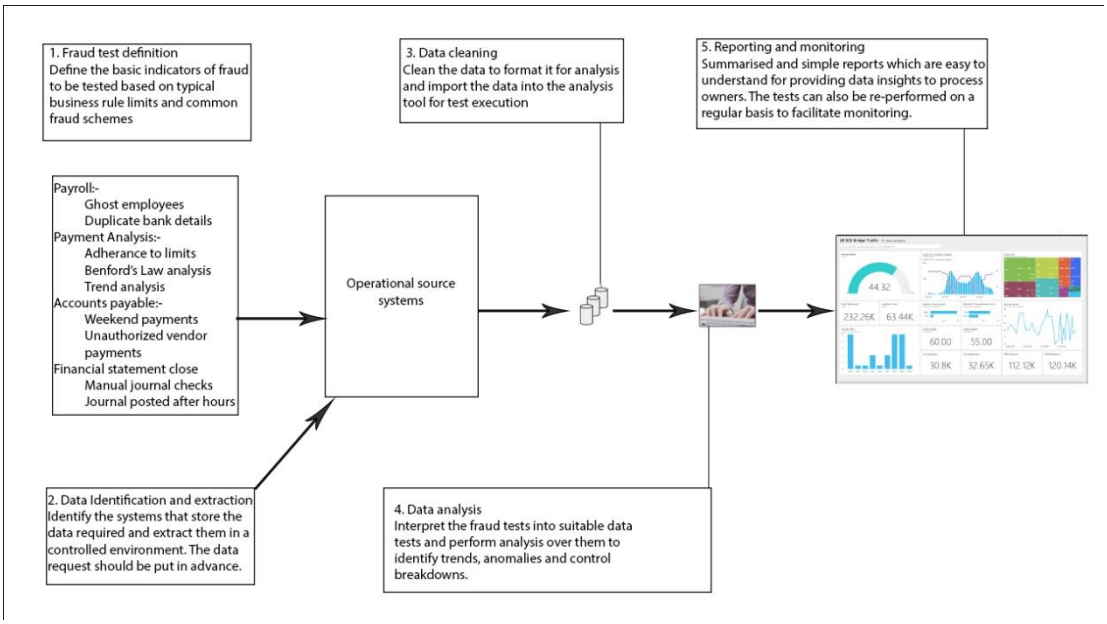

Figure1. *The role of data analytics in aiding fraud prevention [5]*

2.2. Types of Fraudsters

There are different types of fraudsters. The fraudster who is typically enthused by profit has collaboration with the subjective business in question. Generally, every business is constantly vulnerable to interior misrepresentation or defilement from its administration (abnormal state) and workers who cannot be held responsible for upholding administration principles. Notwithstanding inside and outside reviews for misrepresentation control, information mining can likewise be used as a scientific instrument. It is conceivable that the fraudster is an outside gathering. Likewise, the fraudster can either submit misrepresentation as an imminent/existing client (purchaser) or a imminent/prevailing (contractor).

The external fraudster is usually categories to be of three rudimentary sketches: the normal guilty party, felonious offender, and composed criminal. Normal guilty parties show arbitrary and/or periodic exploitative conduct when there is a clear prospect of intrusion, unexpected allurement, or when experiencing budgetary adversity. Conversely, the more dangerous of these external fraudsters are lawbreakers and sorted out guilty parties (proficient/vocation fraudsters) since they exceedingly camouflage their actual personalities and/or advance their business as usual after some time to rough authoritative documents and to counter recognition frameworks. In this way, it is essential to represent the vital communication, alternately moves and counter moves, between a misrepresentation recognition framework's calculations and the expert fraudsters' business as usual. It is also plausible that interior and protection misrepresentation will probably be submitted by normal wrongdoers; credit and information transfers misrepresentation is more defenseless against expert fraudsters. For some organizations where they have collaborations with up to a huge number of outer gatherings, it is cost-restrictive to physically check most of the outer gatherings' personalities and exercises. So the most hazardous ones decided through data mining would result in reports such as suspicion marks, guidelines, and visual anomalies that are be explored. [6]

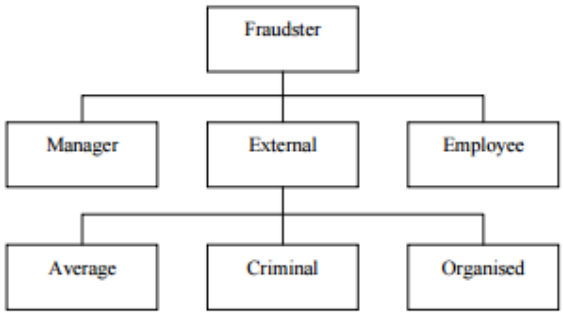

Figure2. *Firm-level and community-level perspectives of fraudsters- a hierarchy charts. [6]*

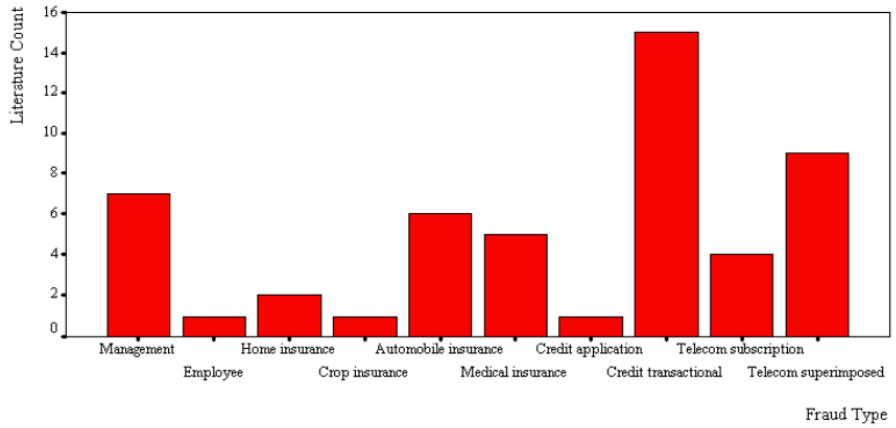

Figure3. *Bar chart of fraud types from 51 unique and published fraud detection papers. [6]*

2.3. Data Mining Techniques Used For Fraud Analytics

There are different ways of approaching fraud detection one of which is to consider it as a predictive modelling problem i.e by correctly anticipating an event that is hopefully not too frequent. the typical useful predictive modeling workflow can be redirected if historical data are available where fraud or opportunities for preventing loss have been identified and verified. This can help in increasing the chances to capture those occasions. In practice, for example, analytical units are supported by several insurance companies, to appraise opportunities for redeeming money based on previously submitted claims. The main objective of this is to find out a screening mechanism so that detailed investigation claims are applied selectively where the cumulative probability for recovery in the form of data data, money etc. is more. Thus along with efficient predictive models subsequent manual resources must also be employed to analyze a claim properly in order to reduce loss.

Fraud Detection as a Predictive Modeling Problem

If one can accurately anticipate a rare event, the issue of fraud detection can be treated as a predictive modeling problem. The success of predictive modeling will increase if we can take into account the previously found frauds in our calculation. Fraud detection mechanism is widely used in investigating claims to reduce loss. Moreover, most companies evaluate opportunities for saving money. The idea is to selectively apply claims where there is a higher probability of recovery. This method is widely used by insurance companies to reduce the loss incurred.

Predicting Rare Events

The above-described fraud detection methodology is quite similar to standard predictive modeling problem. The main perspective is to identify the best predictors as well as create a valid model which provides the greatest lift so as that one gets the most accurate associations between the predicted observations and the actual loss. This can be used to make decisions for accepting credit applications, insurance claims, credit card purchases, etc. With less the 30% cases, frauds are quite occasional. So even though the above-mentioned technique can create redundant samples from the group, it is useful in model building. The data mining models are more successful in finding patterns and relationships that are needed to detect frauds. Creating a good data model is quite essential to yield desired results. The stratified sampling strategies have to be appropriately applied to generate fraudulent vs. non-fraudulent observations with an equal probability which in turn is dependent on the base rate of fraudulent events.

Fraud Detection as Anomaly Detection, Intrusion Detection

If a good training data set is not available, fraud detection is manifested as an interference or irregularity discovery problem. Moreover, in these cases, the datasets are ambiguously assembled where the distinction between fraudulent and non-fraudulent observations are not properly defined. The main objective is to create an elaborate data set to cover the large extent of rules and procedure covered. So, if we consider the example of insurance use case, we would take into consideration all the claims inclusive of all the recoveries as a result of further investigation. However, data which were used in further investigation, can help in the formation of more elaborate dataset. But ordinarily, the data that is to be investigated do not have any powerful indicators of fraud initially. So, the dataset is

simply large and complex with few useful indicators for predictive modeling or supervised learning. Hence, we apply unsupervised learning to find out the outliers. In case of health insurance, the data set is often diverse in order to cover large set of various health issues where with each claim there is a related expectation.

Anomaly Detection

To indicate a process problem, the main objective of anomaly detection is to identify "outliers" in multivariate space. One uses the method of Least Square Fit to identify these. Moreover, this can be used for fraud detection also in other data streams. If we refer to the previously described health care example, our main objective is to identify all claims including fraudulent claims involving reduced payments.

The following four modules of task are used in fraud analytics

1) Classification - During the classification process the data being considered are organized into pre-labelled groups with the use of different types of data mining algorithms. Classification is the process of combination of data in predefined classes. This is sometimes called supervised classification as it uses various class labels to sort the objects in the data group. This usually includes using a known class label as obtained by previous algorithms. These sets are called training sets and a structure is created for this. Different classification techniques are used for different kinds of fraud patterns. There are two ways to examine the performance of classifiers: i) confusion matrix, and ii) to use a ROC graph. Given a class, Cj, and a tuple, ti, that tuple may or may not be assigned to that class while its actual membership may or may not be in that class. With two classes, there are four possible outcomes with the classification as: i) true positives (hits), ii) false positives (false alarms), iii) true negatives (correct rejections), and iv) false negatives. False positive occurs if the actual outcome is legal but incorrectly predicted as fraud. False negative occurs when the actual outcome is fraud but incorrectly predicted as legal

2) Clustering – Clustering is similar to classification but this does not use predefined training classes. It is simply meant to cluster similar objects together. Thus it is a type of unsupervised classification. This follows principle of similarity maximization among intra class objects and similarity minimization techniques among inter class objects.

3) Regression – Regression or genetic programming as it is usually called attempts to obtain a function which models the data of the minimum error.

4) Association rule – Association rules are used to find relationship among data objects by observing the frequency sets occurring together in transactional database. Threshold values known as support and confidence are used to find how frequent an item set is in a particular transaction. Support identifies the frequent item sets and confidence is the conditional probability that an item appears in a transaction when another item appears. [7]

2.4. Applying Data Mining Methods in Different Scenarios for Fraud Analytics

2.4.1. In Cellular Networks

Parameters such as call details, billing data etc are used to analyse and classify on the basis of attributes such as Gender, Account Type, voucher types, billing, calling history etc. This can be used in clustering of clients according to usage patterns. This procedure is usually aided by k means algorithm, kohenon neural networks or hierarchical agglomerative clustering with the help of data mining techniques like Decision tree, Association rule, Neural Networks for training sets and test sets. [8] The results are based on performance measures like sensitivity and precision. The steps used in analyzing and detecting communication fraud include understanding the required data set and instituting the associations among data sets to know behavior of mobile users. K-means algorithm works faster than other clustering methods. Hence, it is more popular. Redundant data variables are removed by task extraction techniques. Now the relationship among variables are analyzed to find out the structure of variables and to detect anomalies helping in fraud analytics. [9]

2.4.2. In Swap Card Fraud Patterns

In swap card fraud detection classification is used to distinguish between real and fraudulent connections. Swap card frauds are usually categorized into three classes- conventional card frauds, commercial frauds and Internet frauds. Swap Card Fraud detection is a conventional case of fraud detection with the help of standard data mining techniques.

2.4.3. In Fraud Claim Data Pattern Detection

For finding fraud claim data patterns the following steps must be followed.

1) It is required to analyze the data sets which are deliberately hidden by the claimer and how the data can be managed. It is also mandatory to measure and analyze the data variables which are provided as inputs.
2) The previous step is usually enough to determine how much amount of data variables is fraudulent and to find out the factors which can help to create prediction rules for swap card fraud analytics. [9]

2.4.4. In Company Fraud Pattern Detection

For company related data fraud detection can be done with the help of models supported by Rule-based classification techniques, Decision Tree visualization and Bayesian Naïve Visualization. [9][10] It is recommended however to collect more data for validation and then apply subject expertise to improve the analysis.

2.4.5. In Retailing

Fraud Detection in retail industries are meant to help in identifying risks of fraud or improper behavior resulting in losses. It is possible for an application to assign a fraud risk score to any event with the help of a mix of business rules, anomaly detection and predictive analysis.

Key features used in fraud detection in retail industries include the following.

- Business rules at all levels (store, cashier, retail)
- Risk matrix
- Clustering and anomaly detection
- Anomaly indices and risk scores
- Risk drivers and impact analysis
- New rules from anomaly detection algorithms
- Ability to manage events / transaction ("fraud" label assigned by the user)
- Predictive models
- Social Network Analysis at multiple levels
- Ability to manage real-time monitoring and alerting logic [11]

2.5. Supervised Vs Unsupervised Learning Methods Used in Fraud Analytics

In case of unsupervised learning both fraudulent and non-fraudulent records are considered. The records are classified as 'fraudulent' and 'non-fraudulent data'. The foundation of the model is reliant on identifying the true classes of training data. It is therefore necessary to obtain such information before going on to create the model. However it is obvious that this method is only capable of previously recorded frauds. Labeled records cannot be used by unsupervised methods. These methods try to trace the behavioral anomaly of accounts, suppliers, customers, retailers etc. with the help of suspicion scores, rules or visual anomalies depending on the context. [12][13]It must be noted that irrespective of whether supervised or unsupervised methods are used the output from such statistical models are capable of only indicating fraud likelihood. It is not possible for any statistical analysis to assure with complete certainty that the object in question is a fraudulent one. For a more detailed overview [13] and [14] maybe referred. In the following section first the supervised methods used in the literature has been briefly discussed, then the unsupervised.

2.5.1. Supervised Methods Used in Fraud Analytics

One of the most exhaustively researched methods of supervised data mining for fraud detection are neural networks. Neural networks have been extensively used as an important tool in fraud analysis for mobile phone networks and in financial statement fraud. [15], [16], [17]. Fuzzy neural networks can also be quite effective for fraudulent financial reporting. [18] Some supervised methods of data

mining use fuzzy rules for situations such as in credit card fraud detection while some use traditional association rules. [19]Traditional association rules have mostly been used for telecommunication companies where it is easier to trace the behavioral models of clients. [20] The indispensability, ever flourishing growth and huge money have turned the health care sector into a very popular fraud target.

Systems that help in processing electronic claims have been deployed to perform automatic audits and reviews of claimed data for quite some time. These systems help to detect areas that need or want special attention. These areas may be categorized to have incomplete data inputs, copy claims or medically uncovered services. However these methods are dependent on pre-defined rules that are usually not too complex and are specified by domain experts without considering the dynamic nature of the problem. Different types of statistical methods that have been applied to health care fraud detection include neural networks [21, 22, 23, 24, 25], decision trees [26, 27] association rules [28], Bayesian networks [29], and genetic algorithms [30, 31]

It must also be noted here that while applying supervised methods in the context of fraud detection in the health care sector there is a tendency to combine several supervised methods in order to improve classification performance. For example, It was propsed by Ormerod et al.[29] that, in order to detect fraud by a Bayesian network (BN),whose weights were refined a rule generator called Suspicion Building Tool (SBT) had to be used. According to He et al. [30] k-nearest neighbor algorithm (where distance metric was optimized by a genetic algorithm) could be used to identify two types of fraud: inappropriate practice of service providers and "doctor-shoppers". In 2005 Vianee et al.[32] presented a tool for coping with insurance fraud. Major and Riedinger in 2002 [33] devised a method of detecting fraud in medical insurance. A hybrid knowledge/statistical system was proposed by them where expert knowledge is integrated with statistical power. Riedinger has originally proposed a program that learns to find out indicators of fraudulent behavior from a big database of customer transactions. In [34] ways have been proposed to aid against cellular clone fraud. A selection is made by taking help of the generated fraud rules to apply in monitors. MADAMID (Mining Audit Data for Automated Models for Intrusion Detection), JAMs (Java Agents for Meta-learning) have been used to help in detecting intrusion via supervised systems by relating cooperative information developed by discrete local managers. [35], [36], [37] Cahill et al [38] proposed a very pioneering idea in 2000. It was based on the principle of mapping data of fraudulent cells to identify telecommunication fraud and came to be known as of fraud signature. In this method to designate a transaction as fraudulent its probability under the account signature is compared to its probability under a fraud signature. Rule-learning and decision tree analysis has also been widely used. [39][40] Link analysis transmits known fraudsters to other individuals with the help of record linkage and social network methods. The fact that fraudsters do not work in isolation from other fraudsters is quite apparent inspired Cortes et al. [41] to come up with some innovative measures.

2.5.2. Classification of Structured Data in Fraud Detection

In this subsection examples of different experimental studies and actual systems used for previous fraud detection has been discussed. It can be expected that further studies on fraud detection will be aided by the review in order to decide if they can work with real data or need to create synthetic data.

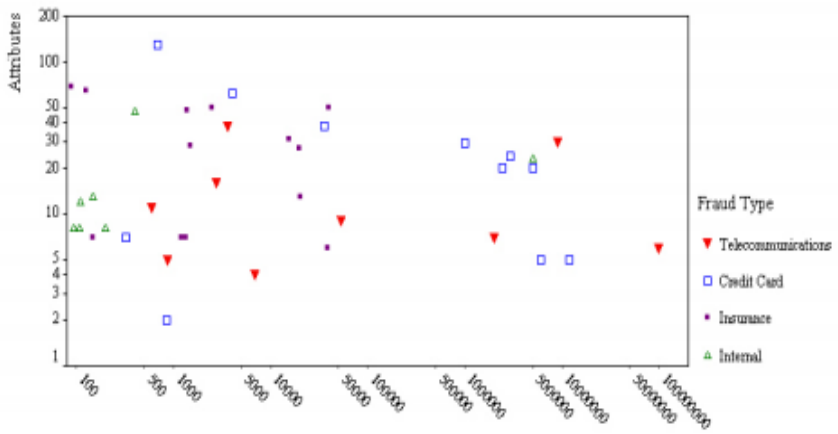

Figure4. *Scatter plot of the data size within common fraud types as collected from 40 unique and published fraud detection papers [6]*

According to the above figure-4 the vertical axis represents the original attributes and the horizontal axis consists of the pre-sampled examples from internal, insurance, credit card, and telecommunications fraud detection literature. Usually, these attributes are binary, numerical (interval or ratio scales), categorical (nominal or ordinal scales), or a mixture of the three. 16 data sets have less than 10 attributes, 18 data sets have between 10 to 49 attributes, 5 data sets have between 50 to 99 attributes, and only 1 data set used more than 100 attributes [42] The lowest of these data sets are seen to be the management data sets (all have less than 500 examples), except for employee/retail data with more than 5 million transactions [43]. The largest data set is the insurance data sets containing thousands of instances and the biggest contain 40000 examples [44]. Most credit transactional data are seen to consist of more than 1 million transactions and the leading data set encompass more than 12 million transactions per year [45]. Telecommunications data includes transactions produced by millions of accounts and hence is huge. The biggest one that was reported consisted of at least 100 million telecommunications accounts [41].

2.5.3. Unsupervised Methods Used in Fraud Analytics

Supervised learning for fraud detection is more popular than unsupervised learning. An "unusual claim" can be characterized by many attributes. But there are two fundamental ways of looking at a problem. One is by identifying outliers in the multivariate space, i.e., unusual combinations of data fields that are unlike typical claims, or by identifying "in-liers", that is, claims that are "too typical", and hence suspect of having been "made up". The difference between usual and unusual claims can be understood by unsupervised learning. Nevertheless unsupervised learning it is not explored as meticulously as supervised learning. Peer Group Analysis have been used by scientists over the ages to monitor behavior. Discrete objects that start to behave in a method unlike from objects to which they had previously been similar are identified readily by analyzing tools. Break Point Analysis have been used to develop methods of behavioral fraud discovery by these scientists [46][47]. A break point is an observation where irregular comportment for a particular account is detected. Both the tools are applied on spending behavior in credit card accounts. [46][47] In the year 1999, two scientists named Murad and Pinkas [48] concentrated on behavioral changes for the purpose of fraud detection and presented three-level-profiling. As the Break Point Analysis from Bolton and Hand, the three-level-profiling method works at the account level and any substantial deviance from an account's standard behavior is designated as a potential fraud. In order to do this, 'normal' profiles are created (on three levels), based on data without fraudulent records. Nonetheless, these methods should rightly be called semi-supervised instead of unsupervised. In the year 2001 behavior profiling has been used by two scientists named Burge and Shawe-Taylor for fraud detection. [49] In the year 1997 Cox et al, in a brief paper has used domain specific interfaces that associate human pattern recognition skills with automated data algorithms. [50] These include some of the important works concerning unsupervised learning in fraud detection but with respect to supervised learning the field is not too well researched.

Nevertheles research on complex, nonlinear supervised algorithms such as neural networks and support vector machines have been emphasized. In the coming years, less complicated and quicker calculations, for example, naive Bayes (Viaene et al, 2002)[32] and logistic regression (Lim et al, 2000) [51] will deliver parallel, if not better results. In the event that the data stream must be handled quickly in an event driven framework or markers are not promptly accessible, then semi supervised and unsupervised methodologies will pose to be the primary data mining alternatives.

3. CONCLUSION

This survey paper has categorized and compared from almost all published technical and review articles using data mining algorithms for fraud detection. This research paper offers methods and techniques to deal with fraud in several business contexts by citing common scenarios. Data mining methods are aimed to help deal with these problems with higher cost savings. Several innovative data visualization and data mining technologies are being deployed by security firms to help in identifying data patterns. These are intended to flush out cyber spies, terrorists and hackers. Many suspicious criminal activities and fraudulent transactions can be stopped by identifying suspicious behavior patterns with the help of historical fraud patterns. Conventional dominions of fraudulent practices have been conferred as probable explanations along with the respective algorithms that can help in such situations. Thus this paper can be regarded as a review guide for related domains. Specifically, future fraud detection research can be aided by unsupervised approaches from counterterrorism work,

actual monitoring systems and text mining from law enforcement, and semi supervised and game-theoretic approaches from intrusion and spam detection communities.

REFERENCES

[1] PricewaterhouseCoopers LLP (2009). "2009 Global Economic Crime Survey". Retrieved June 29, 2016

[2] Fraud Detection and Prevention Biggest Challenge Facing Businesses in EMEA http://www.acquisition-intl.com/2015-fraud-detection-and-prevention-biggest-challenge-facing-businesses-in-emea

[3] http://www.odbms.org/blog/2015/09/on-fraud-analytics-and-fraud-detection-interview-with-bart-baesens/

[4] Nigrini, Mark (June 2011). "Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations". Hoboken, NJ: John Wiley & Sons Inc. ISBN 978-0-470-89046-2.

[5] The role of data analytics in fraud prevention http://www.ey.com/Publication/vwLUAssets/EY_-_Forensic_Data_Analytics/$FILE/EY-Data-Analytics-The-role-of-data-analytics-in-fraud-prevention.pdf

[6] A Comprehensive Survey of Data Mining-based Fraud Detection Research CLIFTON PHUA, VINCENT LEE, KATE SMITH & ROSS GAYLER https://arxiv.org/ ftp/arxiv/papers/ 1009/1009.6119.pdf

[7] Usama Fayyad, Gregory Piatetsky-Shapiro and Padhraic Smyth" From Data Mining to Knowledge Discovery in Databases" Article.

[8] Data Mining - Clustering Lecturer: JERZY STEFANOWSKI Institute of Computing Sciences Poznan University of Technology Poznan, Poland 2008/2009 http://www.cs.put.poznan.pl/ jstefanowski/sed/DM-7clusteringnew.pdf

[9] http://www.ijcsit.com/docs/Volume%204/vol4Issue3/ijcsit2013040303.pdf

[10] Data Mining Applications for Fraud Detection in Securities Market KooshaGolmohammadi, Osmar R. Zaiane.

[11] https://www.accenture.com/us-en/service-fraud-detection-retail

[12] Bolton, R. and D. Hand. Unsupervised profiling methods for fraud detection.

[13] Bolton, R. and D. Hand (2002). Statistical fraud detection: A review. Statistical Science 17 (3), 235–255.

[14] Phua, C., V. Lee, K. Smith, and R. Gayler (2005). A comprehensive survey of datamining-based fraud detection research.

[15] Barson, P., S. Field, N. Davey, G. McAskie, and R. Frank (1996). The detection of fraud in mobile phone networks. Neural Network World 6 (4), 477–484.

[16] Fanning, K. and K. Cogger (1998). Neural network detection of management fraud using published financial data. International Journal of Intelligent Systems in Accounting, Finance & Management 7, 21–41.

[17] Green, B. and J. Choi (1997, Spring). Assessing the risk of management fraud through neural network technology. Auditing 16 (1).

[18] Lin, J., M. Hwang, and J. Becker (2003). A fuzzy neural network for assissing the risk of fraudulent financial reporting. Managerial Auditing Journal 18 (8), 657–665

[19] Brause, R., T. Langsdorf, and M. Hepp (1999). Neural data mining for credit card fraud detection.

[20] Est´evez, P., C. Held, and C. Perez (2006). Subscription fraud prevention in telecommunications using fuzzy rules and neural networks. Expert Systems with Applications 31, 337–344.

[21] Cooper C (2003) Turning information into action. Computer Associates: The Software That Manages eBusiness, Report, available at http://www.ca.com

[22] Hall C (1996) Intelligent data mining at IBM: new products and applications. IntellSoftwStrateg 7(5):1–11

[23] He H, Wang J, Graco W, Hawkins S (1997) Application of neural networks to detection of medical fraud. Expert SystAppl 13:329–336

[24] Ortega PA, Figueroa CJ, Ruz GA (2006) A medical claim fraud/abuse detection system based on data mining: a case study in Chile. In Proceedings of International Conference on Data Mining, Las Vegas, Nevada, USA

[25] Shapiro AF (2002) The merging of neural networks, fuzzy logic, and genetic algorithms. Insurance: Mathematics and Economics 31:115–131

[26] Bonchi F, Giannotti F, Mainetto G, Pedreschi D (1999) A classification-based methodology for planning auditing strategies in fraud detection. In Proceedings of SIGKDD99, 175–184

[27] Williams G, Huang Z (1997) Mining the knowledge mine: The Hot Spots methodology for mining large real world databases. Lect Notes ComputSci 1342:340–348

[28] Viveros MS, Nearhos JP, Rothman MJ (1996) Applying data mining techniques to a health insurance information system. In Proceedings of the 22nd VLDB Conference, Mumbai, India, 286–294

[29] Ormerod T, Morley N, Ball L, Langley C, Spenser C (2003) Using ethnography to design a Mass Detection Tool (MDT) for the early discovery of insurance fraud. In Proceedings of the ACM CHI Conference

[30] He H, Hawkins S, Graco W, Yao X (2000) Application of Genetic Algorithms and k-Nearest Neighbour method in real world medical fraud detection problem. Journal of Advanced Computational Intelligence and Intelligent Informatics 4(2):130–137

[31] Williams G (1999) Evolutionary Hot Spots data mining: an architecture for exploring for interesting discoveries. Lect Notes ComputSci 1574:184–193

[32] Viaene, S., G. Dedene, and R. Derrig (2005). Auto claim fraud detection using bayesian learning neural networks. Expert Systems with Applications 29, 653–666

[33] Fawcett, T. and F. Provost (1997). Adaptive fraud detection. Data Mining and Knowledge Discovery 1 (3), 291–316.

[34] Fawcett, T. and F. Provost (1999). Activity monitoring: Noticing interesting changes in behavior. In Chaudhuri and Madigan (Eds.), Proceedings on the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, pp. 53–62

[35] Pathak, J., N. Vidyarthi, and S. Summers (2003). A fuzzy-based algorithm for auditors to detect element of fraud in settled insurance claims. Odette School of Business Administration Working Paper No. 03-9.

[36] Bordoni, S., R. Emilia, and G. Facchinetti (2001). Insurance fraud evaluation – a fuzzy expert system. In FUZZ-IEEE, pp. 1491

[37] Deshmukh, A. and L. Talluru (1998). A rule based fuzzy reasoning system for assessing the risk of management fraud. Journal of Intelligent Systems in Accounting, Finance & Management 7 (4), 223–241.

[38] Cahill, M., D. Lambert, J. Pinheiro, and D. Sun (2000). Detecting fraud in the real world.

[39] Rosset, S., U. Murad, E. Neumann, Y. Idan, and G. Pinkas (1999). Discovery of fraud rules for telecommunications: Challenges and solutions. In KDD '99: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, pp. 409–413. ACM Press.

[40] Shao, H., H. Zhao, and G. Chang (2002). Applying data mining to detect fraud behaviour in customs declaration. In Proceedings of the First International Conference on Machine Learning and Cybernetics.

[41] Cortes, C., D. Pregibon, and C. Volinsky (2002). Communities of interest. Intelligent Data Analysis 6, 211–219.

[42] Multiple Algorithms for Fraud Detection Richard Wheeler Stuart AtkeinArti?cial Intelligence Applications Institute, The University of Edinburgh https://www.researchgate.net/ publication/222300540_Multiple_Algorithms_for_Fraud_Detection

[43] Kim, J., Ong, A. &Overill, R. (2003). Design of an Artificial Immune System as a Novel Anomaly Detector for Combating Financial Fraud in Retail Sector. Congress on Evolutionary Computation.

[44] Williams, G. (1999). Evolutionary Hot Spots Data Mining: An Architecture for Exploring for Interesting Discoveries. Proc. Of PAKDD99

[45] Neural fraud detection in credit card operations Article in IEEE Transactions on Neural Networks 8(4):827-34 · February 1997 José R. Dorronsoro et al https://www.researchgate.net/publication/5595886_Neural_fraud_detection_in_credit_card_operations

[46] Bolton, R. and D. Hand. Unsupervised profiling methods for fraud detection.

[47] Bolton, R. and D. Hand (2002). Statistical fraud detection: A review. Statistical Science 17 (3), 235–255.

[48] Murad, U. and G. Pinkas (1999). Unsupervised profiling for identifying superimposed fraud. Lecure Notes in Computer Science 1704, 251–262

[49] Burge, P. and J. Shawe-Taylor (2001). An unsupervised neural network approach to profiling the behavior of mobile phone users to use in fraud detection. Journal of Parallel and Distributed Computing 61, 915–925.

[50] Cox, K., S. Eick, and G. Wills (1997). Viual data mining: Recognizing telephone calling fraud. Data Mining and Knowledge Discovery 1, 225–231.

[51] Lim, T., Loh, W. & Shih, Y. (2000). A Comparison of Prediction Accuracy, Complexity, and Training Time for Thirty-Three Old and New Classification Algorithms. Machine Learning 40: 203-228.

AUTHORS' BIOGRAPHY

Sreeparna Mukherjee is a student of National Institute of Technology, Karnataka pursuing Master's in Computer Application, Department of Mathematical and Computational Sciences. Her area of interest are Data Science and Big Data Analytics and Web Service Architecture.

Triparna Mukherjee is a student of M.Sc,. Computer Science, Department of Computer Science, St. Xavier's College (Autonomous), Kolkata. Currently she is doing research work in areas of cognitive science and data science in relation to innovative business models. Her area of interest includes cognitive science in the business industry and digital innovation.

Dr. Asoke Nath is Associate Professor in department of Computer Science, St. Xavier's College (Autonomous), Kolkata. Apart from his teaching assignment he is involved with various research in cryptography and network security, Visual Cryptography, Steganography, Mathematical modelling of Social Networks, Green Computing, Big data analytics, MOOCs, Quantum computing, e-learning. He has already published more than 191 papers in reputed Journals and conference proceedings. Dr. Nath is the life member of MIR Labs(USA), CSI Kolkata Chapter.