International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 2 June - Dec 2019 pp. 115-121 www.csjournals.com

A Review on Insurance Fraud Prediction Approach in Data Mining

[1]Bhavna Batra, [2]Sheetal Kundra
[1]ME(CSE) Student, [2]Associate Professor, AIT
Chandigarh University, Chandigarh, INDIA

Abstract: An approach which can be used for the prediction of future potentials on the basis of present information is known as prediction analysis. This study is relied on the fraudulent discovery in the insurance business. A number of approaches have been projected up to now for the fraudulent discovery in insurance sector. These approaches mainly rely on machine learning algorithms. The method of logistic regression is provided in association with clustering for the fraudulent discovery. The data sample is utilized in the form of input in the machine learning arrangement which is clustered using the k-mean clustering. The clustered data is applied as input to logistic regression for the forecasting. The secondary methods which are projected for the discovery of merely four wheelers insurance deceptions are relied on data analysis. These kinds of schemes attain high precision for the fraudulent forecasting. In the earlier study, the different methods of machine learning are compared as well for finding finest algorithm for the fraudulent forecasting. In this investigative study, the amalgam machine learning algorithm is projected for the insurance fraudulent discovery. It is anticipated that projected method will give good performance in terms of accurateness for the insurance fraudulent discovery.

Introduction

Data mining is known as the process through which the data is stored and retrieved efficiently within certain application depending upon its requirement. The prediction analysis is the approach which can predict future possibilities based on the current information. In this research work, the hybrid machine learning algorithm will be proposed for the automobile insurance fraud detection**.**

1.1 Introduction to Data Mining

The huge quantity of information is generated inside dissimilar applications in daily routine. The proper management of this data or information is a complex task. Generally, large size databases or folders are utilized for the storage of this huge quantity of information. This information is retrieved by the customers according to their obligation. The large sized storage areas and folders are formed for the storage of this big sized data. The suitable extraction of significant information from such enormous database is the major area of concern which rises inside these schemes. According to the prerequisite of consumer, the significant data is to be withdrawal from the folder which is the main reason behind the development of different mechanisms. Different types of neural networks, machine learning algorithms, sample detection and numerous other approaches have been introduced in the recent years in this domain [1]. The procedure which is used for the efficient storage and extraction of information inside some applications according to the need is specified as data mining process. Information detection is the other name given to the data mining. The information amassed inside enormous database is retrieved with the help of this procedure. The extracted or retrieved information can be utilized as significant data for some other purpose further. The data mining process is an imperative branch of KDD although both data mining and KDD are very much alike. The unprocessed information is retrieved through database. Then KDD procedure transformed this data into valuable information which is utilized in different applications.

1.2 Introduction to Insurance Frauds

The occurrence of fraud directly affects any kind of insurance cover. A brief description of various kinds of deceptions in this scenario is given below:

- The data inside the claims of insurance or the solutions of queries rising from an insurance application structure is not ingenuous or absolute.
- A assert is proposed on the basis of mislead or false situations involving the overstatement of a authentic allege
- With a purpose of profit increase inside an insurance agreement, ambiguous or making false contract with an insuring agent.

The insurance plan holder or third parties assert can entrust an insurance deception beside an insurance plan. The asserts for apparition travelers, untrue damage in highway catastrophe, cosmopolitan allege or prearranged offense spheres are only mere of these kind of circumstances in which deception is present in huge manner.

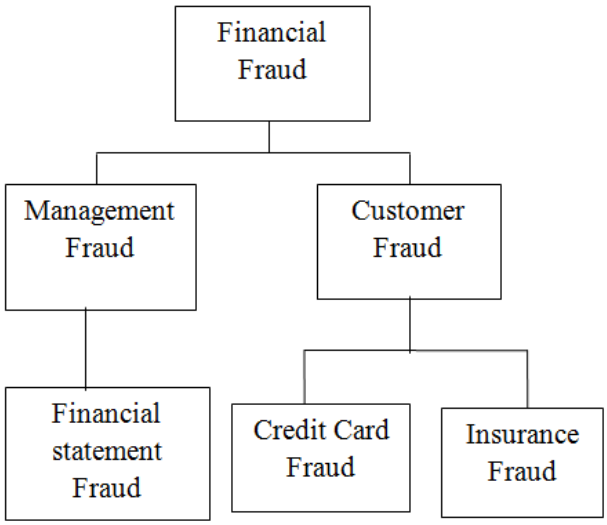

Fig 1.2.2: Categorization of monetary frauds [5]

1.3 Insurance Fraud Detection Techniques

The significant dissimilarities amid consumer activities scrutiny and fraudulent scrutiny methods are to be highlighted importantly. A low optimistic pace can be utilized for the detection of recognized fraudulent behaviors of the deception scrutiny technique. The name and form of fraudulent actions that are given inside the prophesy data sample are retrieved through the scheme [10]. The kinds of frauds being faced by the scheme can be found out easily. If the experiment information does not comprise any fraudulent signatures, no alarm rings. Thus, there can be tremendous diminishing of the false positive rate. The new deceptions cannot be identified as the partial and precise fraudulent proceedings are utilized for the learning of fraudulent scrutiny scheme. Therefore, the false negatives rate turn out to be very elevated on the basis of the information that how clever the impostors are. The consumer performance study can be

utilized for addressing the issue of new deceptions identification. The future actions are evaluated beside the structured form of the legal consumer activities inside these techniques as a substitute of penetrating for exacting fraudulent samples. A probable fraudulent will be affirmed when some action is dissimilar from the scheme. Enormous rates of fake alarm are experienced by the consumer performance scrutiny mechanisms even in a case when they identify the pioneering deceptions generally. The fraud activities will go through the baseline manner and will be implicated as usual for additional scrutiny in case if a fraudulent is identified during the preparation segment.

a. Artificial Neural Network

A suite of interconnected nodules that are intended for emulating the functioning of person intellect is identified as artificial neural network (ANN). A prejudiced link is allocated to every nodule present in the neighboring layer of every nodule. The key acknowledged from the linked nodules is collected and the masses are utilized in association with an easy purpose for computing the yield standards by the solitary nodules [11]. The neural systems comprise dissimilar dimensions and structural designs. The neural system structural design is created by the particular consumer according to the difficulty of constraints which consists the amount of concealed layers, sum of nodules present in the exacting concealed layer and also according to their relatedness. Various supervised, unsupervised and amalgam methods are utilized for the configuration of artificial neural network.

b. Artificial Immune System

The natural immune scheme is a multifaceted scheme which involves specific tissues, units, chemical particles and organs for the generation of a complicated system. An interrelationship is present amid these rudiments and they perform in a coordinated way after the identification of any kind of issue inside them for ensuring that it is eradicated. When immune structure identifies any kind of element then it is recognized as Antigen [12]. The identifiers of immune structure are the antibodies which comprise the capacity of identification and elimination of the damaging and dangerous antigens. The Artificial Immune System (AIS) is a recently evolved sub-domain which is developed on the basis of organic metaphor of the immune structure. The demarcation can be made amid the self and non-self-cells or amid the injurious and non-injurious cells by the immune structure. The investigators nowadays have been inventing such kind of schemes in all domains as the diversity in samples is acknowledged and the issues are recognized and detached exactly by them. The ideas of immunology have been utilized by investigators for introducing a suite of algorithms.

c. Genetic Algorithm

The natural development was the inspiration behind the evolution of genetic algorithms. The chromosomes are identified as a populace of applicant resolutions which occur in the structure of dual threads. The chromosomes are considered as best possible resolution and provided through genetic algorithm. The idea that the probability of continued existence and imitation is higher in the suite of robust associates of definite populace is applicable in this scenario. The ability of a resolution for resolving a problem shows its robustness level as well and defines the power of that resolution. The novel cohort is chosen for the robustness from the earlier populace and the freshly developed offspring. The

genetic algorithms in data mining perform the characteristic assortment procedure mostly. During the amalgamation of genetic algorithm and various other algorithms, the stricture tuning and optimization are executed. The credit card fraudulent discovery procedure comprises genetic algorithm as it is obtainable with different encoding languages and quite trendy and robust as well. Though, the costs of genetic algorithm are high by means of time and memory utilization. A number of applications also use genetic programming in the structure of categorization technology.

d. Hidden Markov Model

The hidden Markov model is a twofold entrenched speculative procedure which is used for the generation of extremely complex speculative procedures [13]. Inside the fundamental scheme, a Markov procedure that comprises unnoticed state is implicated to be accessible. The merely unidentified constraints are the exact conversion of the state inside the easier Markov mock-ups. Though, the state reliant yields are observable but the situations are concealed in case of Hidden Markov Model. The hidden Markov model (HMM) model is trained in the credit card fraudulent discovery for the modeling of usual activities which is encrypted in the clients' profiles. When the recent key transaction is not acknowledged by the scheme with sufficient elevated prospect, this scheme will categorize it as fraudulent.

Literature Review

Li, et.al (2016) projected the random forest fraudulent mining approach which was based on vehicle insurance fraudulent mining by choosing the genuine information of a vehicle insurance corporation. This study demonstrated that the patterns inside the authentic vehicle insurance claim information were not reasonable [14]. Scrutiny was executed on the fault of this approach and experiential study was carried out for the verification. The tested outcomes depicted that the proposed approach showed good outcomes for enormous and disturbed dataset for vehicle insurance fraudulent mining. The proposed approach also showed better precision and sturdiness.

Bauder, et.al (2016) presented a new machine learning approach for the recognition of irregular activities of doctors in the medical insurance alleges. Any kind of misuse, fraudulent or lack of information could be recognized during the billing processes. The projected approach utilized the genuine data obtained from U.S. Medicare system [15]. The multinomial Naïve Bayes algorithm was implemented in this study and numerous recital constraints were computed for performing assessments. This model attained an F-score more than 0.9 for the prediction of different types of doctors. The doctors were classified into individual domains with the help of procedures.

John et.al (2016) proposed a research relevant to the bank fraudulent discovery procedure which involved data mining methods for scrutinizing the user's information [16]. This approach helped in the recognition of any type of samples which were the main reason of frauds. A superior level of verification could be added to the banking procedures after the identification of samples. An ideal form of verification was applied in biometrics according to the attained outcomes. It was identified that it was not necessary for two clients having alike attributes to act in analogous way in the threat appraisal procedure.

Roy, et.al (2017) stated that insurance fraudulent is a situation when an individual or object made fake insurance alleges for obtaining reimbursement or profit to which he/she is not permitted [17]. This study focused on the detection of the automobile fraudulent with the help of machine learning method. The performance of proposed and existing approaches was compared by computing the confusion matrix. The tested results demonstrated that the proposed approach gave better performance in comparison with various other algorithms. The proposed approach showed better accuracy, precision, and recall value. The future work will involve more algorithms.

Supraja, et.al (2017) proposed improvement in fraud detection by using Fuzzy Logic by framing fuzzy rules. This technique will be implemented on more number of datasets and variables [18]. The Fuzzy Logic Technique is used to predict and present fraud. It was illustrated how S curves could be interpreted and avoided ambiguity based on the "degree of goodness" using Fuzzy Logic membership functions. Fuzzy Logic technique was used for high dimensional and large datasets with accuracy and also reliable. Time consumption was low and interpreted results easily. In real time Fuzzy Logic comprised many applications in identifying frauds.

Verma, et.al (2017) proposed a research in which the deceptions of health insurance information were acknowledged and estimated. The involvement of this insurance alleged fraudulent discovery investigational work untangled the fraudulent identification normal samples with the help of rule based sample mining [19]. The outcomes of projected approach were scrutinized. Statistical Decision rules and k-means clustering were implemented on Period based alleged irregularity outliers discovery. The tested outcomes demonstrated that projected scheme was efficient in the identification of deception insurance allege with the help of rule based mining.

kareem, et.al (2017) projected a new scheme for the detection of fraudulent in health insurance alleges. Some definite qualities on the claim credentials were scrutinized for recognizing the association or connection amid them for the identification of deceptions [20]. The necessities of health insurance business were not fulfilled. Therefore, the recognition of deceptions in health insurance enlarged the level of investigators attention in data mining. Therefore, the thriving fortitude of connected qualities could deal with the inconsistency of information in the fraudulent alleges which could provide assistance in reducing the possibilities of fraudulent in the health insurance.

Kenyon, et.al (2017) presented an investigation of big data and data science applications utilized for the prediction of fraudulent insurance alleges. Large data, data science and predictive analytics were applied along with a use-case in an interim insurance business [21]. A privacy protection approach was projected for predicting the insurance alleges fraud. The proposed approach utilized enormous data samples for the generation of policies beside with the interest of cross-broker and cross-insurer exploitations. The tested outcomes depicted that the regulations generated in this technique were more precious than the existing approaches.

Anbarasi, et.al (2017) proposed a research relevant to the fraudulent discovery in the health insurance information. The proposed approach integrated the proactive and retrospective analysis [22]. The disjointed character of anomalous actions was managed by incorporating the reactive and proactive schemes for generating an enhanced method. The doubtful activity of health concerned trace was

recognized with the help of outlier relied predictors for health insurance fraudulent discovery. In future, the character of health care fraudulent existence can be identified by realizing future enhanced online fraudulent discovery technologies.

Wang, et.al (2017) projected a new scheme for the identification of automobile insurance fraudulent based on deep learning [23]. The proposed approach utilized LDA relied text analytics. Latent Dirichlet Allocation technique was used for the extraction of the text characteristics incasing in the text images of accidents. Further deep neural networks were trained on the presented information. The tested outcomes demonstrated that for the detection of automobile insurance fraudulent, the combination of Latent Dirichlet Allocation and deep neural network approach was extremely important and competent.

Lamberti, et.al (2018) proposed a analysis in which smart contract sensors information was used to understand a scheme for on-demand insurance. A mobile application and an electronic tool were fitted on the vehicles of customers for generating a prototype. For scheduling the automatic modifications and collecting pictures of vehicles, the interaction with smart contract was carried out for modify the policy coverage physically [24]. The environmental circumstances were observed and the alterations were launched with the help of electronic equipment. The projected resolution was capable for lowering the policy modification costs.

Subudhi, et.al (2018) projected a novel fraudulent detection approach in vehicle insurance area. A method identified as adaptive oversampling technique was utilized for the implementation of the projected methodology [25]. Within the alternative class spaces, an adaptive oversampling method was utilized for the information inequity accessible in the formerly maintained dataset. Many tests on the basis of automobile insurance dataset were executed in order to evaluate the projected method. Different tests performed on practical dataset clearly demonstrated the competence of the projected scheme.

Conclusion and Future Scope

The procedure which is used for the efficient storage and extraction of information inside some applications according to the need is specified as data mining process. The forecast scrutiny or prediction analysis is a methodology which is implemented for the prediction of future instance from the present data. This approach is executed in the two steps which are feature extraction or characteristic retrieval and classification. Different classification methods are implemented up to now for the forecast scrutiny such as support vector machine, decision tree and voting techniques. This investigative study presents a new approach for the forecast scrutiny. The projected method aims to enlarge accurateness and decrease implementation time for the insurance plan fraudulent recognition.

References

[1] Chenfei Sun ; Qingzhong Li ; Hui Li ; Yuliang Shi ; Shidong Zhang ; Wei Guo, “Patient Cluster Divergence Based Healthcare Insurance Fraudster Detection” , IEEE Access Year: 2019 , Volume: 7 Page s: 14162 – 14170
[2] S. Viaene, R.A. Derrig ; G. Dedene, “case study of applying boosting naive Bayes to claim fraud diagnosis”, IEEE Transactions on Knowledge and Data Engineering, Year: 2004 , Volume: 16 , Issue: 5 Page s: 612 – 620
[3] M. Sternberg; R.G. Reynolds, “Using cultural algorithms to support re-engineering of rule-based expert systems in dynamic performance environments: a case study in fraud detection”, IEEE Transactions on Evolutionary Computation Year: 1997 , Volume: 1 , Issue: 4 Page s: 225 - 243

[4] J.R. Dorronsoro; F. Ginel ; C. Sgnchez ; C.S. Cruz, "Neural fraud detection in credit card operations", IEEE Transactions on Neural Networks Year: 1997 , Volume: 8 , Issue: 4 Page s: 827 – 834
[5] Aastha Bhardwaj, Rajan Gupta, "Financial Frauds: Data Mining based Detection – A Comprehensive Survey", 2016, International Journal of Computer Applications (0975 – 8887)
[6] Yongchang Gao ; Chenfei Sun ; Ruican Li ; Qingzhong Li ; Lizhen Cui ; Bin Gong, "An Efficient Fraud Identification Method Combining Manifold Learning and Outliers Detection in Mobile Healthcare Services", IEEE Access Year: 2018 , Volume: 6 Page s: 60059 – 60068
[7] Yufeng Kou, "Survey of fraud detection techniques", 2004, Networking, Sensing and Control,IEEE International Conference
[8] Shunzhi Zhu, Yan Wang, Yun Wu, "Health Care Fraud Detection Using Nonnegative Matrix Factorization", The 6th International Conference on Computer Science & Education (ICCSE 2011) August 3-5, 2011. SuperStar Virgo, Singapore
[9] Zhongyuan Zhang, Tao Li, Chris Ding, Xiangsun Zhang, "Binary Matrix Factorization with Applications", Proceeding ICDM '07 Proceedings of the 2007 Seventh IEEE International Conference on Data Mining Pages 391-400.
[10] Mohammad Sajjad Ghaemi. Class Lecture, Topic: "Clustering and Nonnegative Matrix Factorization". DAMAS LAB, Computer Science and Software Engineering Department, Laval University. Apr.12, 2013.
[11] Haesun Park. Class Lecture, Topic: "Nonnegative Matrix Factorization for Clustering". School of Computational Science and Engineering Georgia Institute of Technology Atlanta, GA, USA, July 2012.
[12] Fashoto Stephen G., Owolabi Olumide, Sadiku J., Gbadeyan Jacob A, "Application of Data Mining Technique for Fraud Detection in Health Insurance Scheme Using Knee-Point K-Means Algorithm", Australian Journal of Basic and Applied Sciences, 7(8): 140-144, 2013 ISSN 1991- 8178.
[13] Leonard Wafula Wakoli. "APPLICATION OF THE K-MEANS CLUSTERING ALGORITHM IN MEDICAL CLAIMS FRAUD/ABUSE DETECTION." MSc Thesis, Jomo Kenyatta University Of Agriculture And Technology, 2012.
[14] Yaqi Li, Chun Yan, Wei Liu, Maozhen Li, "Research and Application of Random Forest Model in Mining Automobile Insurance Fraud", 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)
[15] Richard A. Bauder, Taghi M. Khoshgoftaar, Aaron Richter, Matthew Herland, "Predicting Medical Provider Specialties to Detect Anomalous Insurance Claims", 2016 IEEE 28th International Conference on Tools with Artificial Intelligence
[16] S. N. John, Okokpujie Kennedy O. C. Anele, F. Olajide, Chinyere Grace Kennedy, "Real time Fraud Detection in the Banking Sector Using Data Mining Techniques/Algorithm", 2016, IEEE
[17] Riya Roy, Thomas George K, "DETECTING INSURANCE CLAIMS FRAUD USING MACHINE LEARNING TECHNIQUES", 2017 International Conference on circuits Power and Computing Technologies [ICCPCT]
[18] K. Supraja, S.J. Saritha, "Robust Fuzzy Rule based Technique to Detect Frauds in Vehicle Insurance", International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017)
[19] Aayushi Verma, Anu Taneja, Anuja Arora, "Fraud Detection and Frequent Pattern Matching in Insurance claims using Data Mining Techniques", Proceedings of 2017 Tenth International Conference on Contemporary Computing (IC3)
[20] Saba kareem, Dr. Rohiza Binti Ahmad, Dr. Aliza Binit Sarlan, "Framework for the Identification of Fraudulent Health Insurance Claims using Association Rule Mining", 2017 IEEE Conference on Big Data and Analytics (ICBDA)
[21] David Kenyon, J.H.P Eloff, "Big Data Science for Predicting Insurance Claims Fraud", 2017, IEEE
[22] Dr.M.S. Anbarasi, S. Dhivya, "FRAUD DETECTION USING OUTLIER PREDICTOR IN HEALTH INSURANCE DATA", INTERNATIONAL CONFERENCE ON INFORMATION, COMMUNICATION & EMBEDDED SYSTEMS (ICICES 2017)
[23] Yibo Wang, Wei Xu, "Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud", 2017, DECSUP 12895
[24] Fabrizio Lamberti, Valentina Gatteschi, Claudio Demartini, Matteo Pelissier, Alfonso Gómez, and Victor Santamaria, "Blockchains Can Work for Car Insurance: Using Smart Contracts and Sensors to Provide On-Demand Coverage", 2018, IEEE Consumer Electronics Magazine, Volume: 7, Issue: 4
[25] Sharmila Subudhi, Suvasini Panigrahi, "Effect of Class Imbalanceness in Detecting Automobile Insurance Fraud", 2018, 2nd International Conference on Data Science and Business.