American Journal of Data Mining and Knowledge Discovery
2019; 4(2): 70-74
http://www.sciencepublishinggroup.com/j/ajdmkd
doi: 10.11648/j.ajdmkd.20190402.13
ISSN: 2578-7810 (Print); ISSN: 2578-7837 (Online)

Fraud Detection and Fraudulent Risks Management in the Insurance Sector Using Selected Data Mining Tools

Leonard Mushunje

Department of Applied Mathematics and Statistics, Midlands State University, Gweru, Zimbabwe

Email address:
leonsmushunje@gmail.com

To cite this article:
Leonard Mushunje. Fraud Detection and Fraudulent Risks Management in the Insurance Sector Using Selected Data Mining Tools. *American Journal of Data Mining and Knowledge Discovery*. Special Issue: *Wider Thoughts on the Application of Data Mining Tools and Predictive Modelling in Finance*. Vol. 4, No. 2, 2019, pp. 70-74. doi: 10.11648/j.ajdmkd.20190402.13

Received: October 2, 2019; **Accepted**: November 6, 2019; **Published**: November 19, 2019

Abstract: Knowledge discovery, shortly known as Data mining plays a crucial role within the insurance sector. Serious troublesome cases such as fraudulent cases can be well managed in the insurance sector through data mining application. In this paper, we aim to put on surface the two forms of fraud that is softy and hard fraud, to give out the causes of such fraudulent acts and to state out different suggested data mining techniques that can be applied to the insurance data to detect fraud. Also, we aim to highlight other benefits that can be enjoyed from using data mining in the insurance sector. We conjectured and found that, application of data mining helps to quickly detect fraud, reduce operation cost and to improve profit margins and increased competitive advantages. We put forward that techniques such as association, clustering, classification and regression are good when detecting fraud from the insurance claims data and should be acquired and applied. We then recommended that, underwriters and insurance officials should contribute much in preventing fraudulent cases in the insurance sector. This is so because, prevention is better than cure. Above all, we concluded that, application of data mining techniques through sequential pattern mining can help much to predict any future and potential fraudulent cases. This is helpful on planning and to keep the insurers alert before the fraudulent risk occurs.

Keywords: Fraud, Data Mining, Insurance Sector, Knowledge Discovery

1. Introduction

Cases where insurance services are obtained and enjoyed through false pretences are now common in the insurance sector. Such cases are shortly known as Fraud. Fraudulent actions create a special type of risk called fraudulent risk. As such proper investigation and management of such risks should be well practised in any business sector like insurance. Examples of such fraud cases include exaggeration of losses or causing an accident with the sole intent of the payout. Insurance companies have been trying to set up different methods to detect the root causes to such cases. Some of these include use of external auditors, tip lines application and detection by dedicated departments and accident. These methods however proved to be inefficient as evidenced by persisting and prevailing fraudulent cases during insurance transactions. These fraudulent acts are believed to be coming from a number of avenues. As such, this paper aims to put forward different root causes of fraud in the insurance sector and to provide better methods and techniques that can be used to analyse insurance data to detect fraudulent cases. We aim to suggest and emphasis the application of selected data mining tools as an effective way of managing risks. We conjectured that, application of such tools can result in better management of fraudulent risks by insurers. Application of data mining tools is not new. However, their practical appreciations seem not to be widely known within the insurance sector. We found some existing literature on data mining tools and fraud detection by scholars such as Sumiran [1] who discussed and gave an overview of data mining including the concepts behind what it is and the variations on how it is accomplished. They applied the data mining techniques in the engineering industry. They bring on a predictive tool known as Sequential pattern mining. Additionally, Agrawal [2] highlighted that the most known way of applying these data mining techniques is in the field of business. However, due to time progression, such applications are now in more than business sections of

economies. For example the latest applications of data mining can be found in Torabi et al [3] where they developed a new prediction model for energy production with wind turbines. On the same note, Mostafa [4] took a review of some techniques and applications of data mining concepts. First, they described the fundamentals of databases and its standards, and advantages of DBMS. Afterwards, they explained the data-mining concept, data mining techniques or methods, how the processes of data mining are accomplished, then the most categories that can be used by data mining in the informational world. Data mining is a new and upcoming technique based on knowledge discovers and so it is also called Knowledge discovery in databases (KDD). It can be seen in Freitas [5] who said that data mining application can extract meaningful new patterns from existing databases and valuable and satisfactory decisions can be well made. In the insurance sector, data mining has been applied in both health and general insurance. The applied techniques such as neural networks, K-means algorithms were used to detect fraud and to minimize costs. This paper aims to give an overview of causes of fraud in the insurance sector, to propose and to state out the number of data mining techniques that can be used to detect fraud in the insurance sector. In addition to that, Guo [6] concentrated on property insurance and elaborated how data mining techniques are used within the sector. Also, Goonetilleke and Caldera [7] summarised the analysis of customer analysis when mining a life insurance data. Moreover, Bhowmik [8] detected the auto insurance fraud using the various fraud anomaly detection techniques. He further focused and emphasis on identifying the behaviour of customer and analyzed the profitable customers for the insurance company. The results showed the preciseness of data mining tools when dealing with big data. In addition, Srinivasta and Balaji [9] summarized the classification techniques used for the prediction of data over life insurance customer data. He analyzed the various classifier algorithms (Naïve Bayes, Bayesian Network classifier etc) and all of them proved to efficient and wonderful when analysing insurance data. Again, LookmanSithic and Balasubramanian [10] analyzed the various review for the fraud detection by listing related research papers on fraud detection in the insurance sector. Also, JayanthiRanjan [11] summarized the applications of customer relationship management in insurance company. She did the case study with insurance data predicting fraudulent claims & medical coverage and predicting the customer's pattern which customer will buy policies. All this with other massive related works were in support of the application of data mining tools and techniques. Our paper provided a theoretical discussion of causes of fraud, data mining and applications and the consequences of using such tools in the insurance sector. An emphasis at the end was given on the wider and serious application of such tools and techniques.Therefore the rest of the paper proceeds as causes of fraud in the insurance sector, a review of selected data mining techniques, other importance of data mining to the insurers, discussions and recommendations and finally the conclusions.

2. Causes of Fraud in the Insurance Sector

There are two forms of Insurance fraud which are "hard" or "soft." Hard fraud is when someone deliberately fabricates claims or fakes an accident whiles soft insurance fraud, also known as opportunistic fraud, occurs when someone goes for legitimate claims, for example, in the case of business owners, list fewer employees or misrepresents the work they do to pay lower premiums for workers compensation. People who commit insurance fraud range from organized criminals, who steal large sums through fraudulent business activities and insurance claim mills, to professionals and technicians, who inflate the cost of services or charge for services not rendered, to ordinary people who want to cover their deductible or view filing a claim as an opportunity to make a little money. Some lines of insurance are more vulnerable to fraud than others. Healthcare, workers compensation and auto insurance are believed to be the sectors most affected. As such below is an overview of suggested causes of fraud in the insurance sector.

2.1. Information Asymmetry

The flow of information from one point to another and from person to person is of vital importance in the business sector. Usually people act according to what they know and what they hear. Information within the insurance sector should be permeable and accessible to every active part involved. This means that, officials who are responsible for the insurance games between insurers and policyholders should disclose and clearly pass all the necessary information to the players into the market. The players should know what each one is doing and how he/she is doing that. This gives no room to find one trying to trick the other part (fraud). This means that, the information defining the games should be well defined disclosed and well communicated to the players so as to avoid such fraudulent cases. Where such crucial information is unavailable, surely fraud will be practised.

2.2. The rationality Principle

The idea is that, every player in the insurance sector aims to maximize his/her benefits to be derived from every action they take. For example insurers aim to maximize their payoff from the harvested premiums whereas policyholders aim to maximise their benefits from their insured sums and pre-paid premiums. Sometimes, the risks that policyholders do insure against, may not happen and as such, they may find it unfair if their contributions are just eaten up by insurers that is in some contracts. For example, the Medicaid policy. One pays the entire regular and up to date premiums and sometimes he/she may not get ill so he will find it unfair and maybe decide to place a fake sick or to give someone the rights to enjoy the benefits where in actual fact, there is no claim, thus fraud. Therefore, the suggestion of this paper is that, after identifying that fraud may arise as a result of rationality, insurers and policyholders should aim to operate at a point

which is stable, equilibrium and fair to both and as well under equally favourable conditions. This helps to encounter the challenge of fraud from this angle.

2.3. Underwriters' Inefficiency

In some cases, underwriters do not take their work efficiently. Sometimes they do not record whatever information about the policyholders seriously and carefully. For example, when the policyholder is joining the policy or when the claim hits the insurer. The underwriters may decide due to their laziness/unintentionally to skip some information when it is of importance, for example, skipping gender of the policyholders. Such skipped information may be used as a way of committing insurance sin (fraud) by policyholders so as to enjoy extra benefits than what they deserve. Additionally, underwriters may lack precision when recording any important information about the policyholders. For example, they may mistype the names, age or any other information which may act as a cool driver of fraud in future.

3. Data Mining Techniques

Data mining is based on a number of techniques. The techniques as well are based on different tools and algorithms. In this section we gave out the selected techniques and algorithms or each that can be used in the insurance sector to detect fraud. Each technique has unique applications based on the aims and prevailing conditions and structure of the data. The detailed explanation and application for each technique are given below.

3.1. Classification

It is a data analysis task whose interests are pivoted on describing and distinguishing classes within our data, Hajiadeh et al [12]. It is worried with identifying the family class to which either a data point (s) belongs. Based on this tool, a thorough data check is done to establish its feasibility for the study. A classifier is actually required when identifying classes within our data. A classifier is used to either classify our data as risky or safe. The risky class will be representing the potential fraudsters while the safe class is for the clean and potentially successful applicants. So, from such classes insurers are can safely detect the potential future risks and pre-plan for these. Classification is performed using such algorithms as random forest, Bayesian classifiers, neural networks, support vector machines (SVM) and K-Nearest Neighbour (KNN). So generally, the classification technique helps to group the claims data as either risky or safe class. The risky will be comprising of the potential fraud and true otherwise. This makes it easy and convenient for the insurers when treating their claims data before allocating any compensation and benefits.

3.2. Outlier Detection

It is a technique used to identify unusual patterns that do not conform to expected behaviour. It is also called anomaly detection. Also it is almost similar to noise removal and novelty detection as it deals with new pattern discovery and identification. This method aims to identify and fish out individuals who maybe misbehaving or failing to follow the expected and recommended pattern. It is an efficient method of monitoring policyholders' data by insurers. We have different anomaly boundaries that we can safely use to demark and classify our data. These include point anomalies (when a single data point is too far from the rest), contextual (when data is failing to match and it is common with time series data) and collective anomalies (collective data analysis). This technique to identify such anomalies from claims data and make our further classification analysis. Other techniques that can be used are factor analysis, time series Boris and Evingenii [13] and neural networks. These techniques are all efficient and they can as much help to identify interesting patterns within our data. As such, insurers can make use of the technique and algorithms to identify extreme cases from the applicants, claims and policy data.

3.3. Clustering

According to Berry et al [14], it is the process of grouping a set of data objects such that objects with similarities fall into a single cluster and as well those with dissimilarities in one single cluster. This merely implies that the claims with similar features fall into one clusters. In credit data analysis, this technique helps to identify and cluster all the potential defaulters together. Also, this helps to curb all potential fraud that can promise to happen. Clustering is well done using either supervised or unsupervised learning based algorithms, but the supervised one is of more relevancies since we will be dealing with real life data. Unsupervised is based on human physical and manual observed clustering. Different algorithms such as K-means clustering and K-mediod algorithms can be well employed to come up with meaningful data analytic results. Other classification tools include the hierarchical clustering and multiphase clustering. A recommendation on the application of such tools at different times to the same data is crucial and should be well considered. This helps users (insurers) to identify some hidden patterns from their data (claims data). Potential fraud can be well detected, thus rendering safety to them.

3.4. Regression

This technique is based on the associative rules of data mining. It is a technique which aims to fish out ant-relationships and associations among data objects in a data set, DeVault [15]. Basically we have either simple linear or multiple linear regressions. Other forms like logistic and poly regressions are also important. Investigating relationships among any variables of interests is of use when analysing data. Using claims data, insurers can safely identify the associations between or among the applicants' information such as age, gender, income among others. Regression analysis in data mining is well performed using random forest algorithm based on the association rules. These tools

can be also used in connection with the association rules.

4. Other Importance of Data Mining

Data mining is a broad subject whose application is now appearing on a broad base. Within and outside business sector, a quite number of benefits are realized as a result of data mining application. Such benefits can also be attained and enjoyed in the insurance sector. We therefore, gave out a detailing of the benefits other than fraud detection that can be obtained from using data mining. Some of the main benefits are:

Reduced costs-data mining tools helps to quickly identify potential fraudsters from the insurance data. This means that other costs of fraud analysis like auditing costs, management and administrative costs will be minimized.

Increased profit margins- by minimized costs margins and increased efficiency ratios, profit margins for insurers will be more likely to rise. This ensures their long-term survival in the insurance field.

Improved decision making-new knowledge discovery and new pattern discoveries plays great roles in improving decision quality and value of insures in the insurance industry. The decisions that will be made will be appropriate and as such the insurers will have more chances of operating at the right side of profit making. Additionally, reduced risk margins, improved forecasting and competitive advantage over their competitors are some of the benefits derived from data mining. Forecasting is well done using the sequential pattern mining techniques. Such predictive modelling tools provide insurers with better, quick and precise ways of predicting their future outcomes.

5. Discussions and Recommendations

Risk management by insurers is surely the key reason behind their existence. As such, we discussed a new method of dealing with insurance data to mitigate some toping risks such as fraudulent risks. We discussed the use of data mining techniques in the insurance sector. The techniques and tools are applicable in all insurance business that is both life and non-life insurance including health insurance. We identified some of the causes of fraud cases in the insurance sector. Theseinclude imperfect and information asymmetry, the power of utility theory, inefficiency of insurance laws and inefficiency of underwriters themselves. Such factors do contribute much to fraud in the insurance sector and as such they should be prevented and avoided. It should also noted that application of data mining tools have other extended sheets of benefits than fraud detection. These include reduces cost, improved competitive advantages among others. This means that, knowledge discovery and pattern search are important in the insurance sector. Our recommendation is that, every insurance company must strive to adopt the use of such techniques in order to ensure a better and concrete stance against its competitors and to finely manage its fraudulent risks. If possible a data science and analytics section/department must be opened which focuses on developing and application of such beautiful machine based methods.

6. Conclusions

Our study, managed to put forward different data mining tools that can be used by insurance firms to detect fraud and to manage the associated fraudulent risks. We identified and put forward the sources of fraud in the insurance sector. A primer discussion of different data mining techniques that can be used was done. Additionally, a couple of benefits that can be extracted from the tools was put forward and discussed. This paper therefore, recommended insurers to aim to make use of the techniques whenever analysing their claims data before allocating any benefits and compensation to their policyholders.

References

[1] Sumiran, K (2018). An Overview of Data Mining Techniques and Their Application in Industrial Engineering. Asian Journal of Applied Science and Technology (AJAST). Open Access Quarterly International Journal) Volume 2, Issue 2, Pages 947-953.

[2] Agrawal, R, Imielinski, T & Swami, A (1993). Mining Association Rules between Sets of Items in Large Databases. Proceedings Of The 1993 Acm Sigmod International Conference On Management Of Data, Washington Dc (Usa).

[3] Torabi, A., KiaianMousavy, S., Dashti, V., Saeedi, M., &Yousefi, N. (2018). A New Prediction Model Based on Cascade NN for Wind Power Prediction. Computational Economics.

[4] MOSTAFA, A (2016). Review of Data Mining Concept and its Techniques. DOI: 10.13140/RG.2.1.3455.2729. SIGMOD '93 Proceedings of the 1993 ACM SIGMOD international conference on Management of data, Pages 207-216.

[5] Freitas, A (2013). Data Mining and Knowledge Discovery with Evolutionary Algorithms. Natural Computing Series.

[6] Lijia Guo. "Applying Data mining Techniques in Property/Casualty Insurance", Casualty Actuarial Society Forum Casualty Actuarial Society - Arlington, Virginia Winter 2003.

[7] T. L. OshiniGoonetilleke and H. A. Caldera Mining Life Insurance Data for Customer Attrition Analysis, Journal of Industrial and Intelligent Information Vol. 1, No. 1, March 2013.

[8] RekhaBhowmik, "Detecting Auto Insurance Fraud by Data Mining Techniques", Journal of Emerging Trends in Computing and Information Sciences, Volume 2 No. 4, April 2011.

[9] S. Balaji and Dr. K. Srinivasta, "Naïve Bayes Classification Approach for Mining Life Insurance Databases for Effective Prediction of Customer Preferences over Life Insurance Products", International journal of Computer Applications, vol. 51, No. 3, pp. 22-26. August 2012.

[10] H. Lookman Sithic, T. Balasubramanian, "Survey of Insurance Fraud Detection Using Data Mining Techniques", International Journal of Innovative Technology and Exploring Engineering, Vol-2, Issue-3, February 2013.

[11] Jayanthi Ranjan "Data mining in pharmacy sector: benefits", IJHCQA, Vol. 22 No. 1, 2009, pp. 82-92.

[12] Hajiadeh, E, Ardakani, H. D and Shahrabi, J, (2010). Application of data mining techniques in stock markets: A survey. Journal of Economics and International Finance, vol (2).

[13] Boris, K and Evgenii, V, (2005). Data Mining for Financial Applications. Data Mining and Knowledge Discovery Handbookpp 1203-1224.

[14] Berry, Michael J. A. danLinoff, Gordon S. 2004, Data Mining Techniques, 1st ed., John Wiley & Sons Inc., Indianapolis, Indiana.

[15] Gigi De Vault, (2019). A Basic Statistics Approach to Analyzing Quantitative Data. available at https://www.thebalancesmb.com/what-is-simple-linear-regression-2296697.