

Customer Ethics In Health Insurance Claims: A Study Using Genetic Algorithm

K S Hari Sankar
CB.BU.P2MBA15041
Amrita School of Business,
Amrita Vishwa Vidyapeetham University
Coimbatore, India

Shyam A V,
Assistant Professor,
Amrita School of Business,
Amrita Vishwa Vidyapeetham University
Coimbatore, India

Abstract— Health Insurance industry is one of the fastest growing industries in India. One major issue faced by them is insurance fraud. “Fraud in health insurance is done by intentional deception or misrepresentation for gaining some undue benefit in the form of health expenditures”. Fraudulent claims add to the expense for the insurance providers. This paper aims at using analytics and statistical techniques for analyzing the past data and predicting the possibility of claims being a fraudulent one. Claims data of one of the regional offices of a prominent health insurance provider was used for this study. The variables to be measured and studied were identified based on the inputs from the literature review and the expert interview. Expert Interview was done with an insurance expert who has experience in detecting insurance fraud. Once the variables were identified the data for the same was analyzed to predict the probability of the claim being a fraudulent one and flagged for further investigation. Based on the results the experts can have a closer look at only those cases that are flagged of having higher probability of being a fraudulent one thereby reducing the cost involved in detestation of insurance fraud.

Keywords— Health Insurance; Health Insurance Fraud; Analytics; Decision Tools;

I. INTRODUCTION

Health Care in India is undergoing huge transition. Increase in the average income, awareness about its importance and privatization of health care are the major factors that lead to this change. Since India is a huge nation with large population and also the fact that a large percentage of the population in India lives rural areas or urban slums and are also below the poverty line make health care a gargantuan problem. This has motivated the people and the government to explore financing options in health care and cover for the increasing cost of health care and the changing pattern of diseases. Health Insurance which remained underutilized till recently is now being projected as a tool that can be used to manage the financial needs of people who need health care. Health Insurance is comparatively more complex with respect to other segments of insurance business because of the very nature of health insurance. Due to the complex nature of health insurance, it is vulnerable to fraudulent claims.

Health insurance fraud in particular is a costly form of unethical customer behavior and has been the prime focus of many of the recent studies. Currently detection of fraudulent claims is done manually, involving a medical practitioner with rich experience in health care. Details regarding the claim are manually checked by him/her and then based the findings the final decision is made to approve the claim or go for further investigation before approving the claim. In this paper we use the experience of an expert in detection of insurance fraud and try to find out the pattern in the data, which could be used to help flag fraudulent claims, thus reducing the burden of manual intervention, as the expert will have lesser number of claims that are to be manually verified. Next part of the paper would discuss about the literature review that was done and about the expert interview that was conducted. After that the discussion would be about the hypotheses formulation for this study. In the fourth section the discussion would be on the data that was collected and the methodology used for research. The last section discusses about the results and interpretation of the results followed by the conclusion

II. LITERATURE REVIEW

There has been a significant growth in the amount of research done in developing algorithms and using analytics to predict and identify consumer fraud and unethical behavior. Fraud essentially involves doing set of activities which could range from dishonesty to deception with the intention of creating loss to others or which could lead to a loss to others or with the intention of making a personal gain [1]. Insurance fraud is one of the main unethical consumer behavior that is being researched upon. According to [2], Insurance claim fraud is the second largest white collar crime and it burdens the government, insurance providers and the customers billions of dollars each year. It claims that the perception of the policy holders that inflating an insurance claim is not unethical is the major issues faced by the insurance providers in the United States of America. The results show that the higher deductible amounts has induced a feeling among the customers that it is fair to inflate the claim amount. There is not sense of

unethicality in inflating the claim amount from the point of view of the customer. The paper also discusses about the implications of this behavior in the insurance industry. To predict the probability of insurance fraud, many researchers have used analytical techniques by analyzing large datasets looking for patterns or using data mining techniques to predict the probability of a claim being a fraudulent one.

In [3], the claim attributes that could help in predicting the probability of a claim being fraudulent were identified and used. The age of the person making the claim, the time between the occurrence and the claim being actually made and the a few other factors were identified as indicators of potential fraud. However not much research is done on the methods used by the expert in India to identify the fraudulent claims. The parameters used for the predication would also depend on the country as fraudulence methods are affected by the local culture and local legislations, as mentioned in [4]. For example, people from a country with individualistic culture may have a lower tendency to misuse the system when compared to an individual from a collectivist background. There is every possibility that an uninsured person would disguise himself or herself as an insured person and get the benefits of insurance unlawfully in a collectivist culture, as the family ties are stronger and this type of fraud would require consent of the actual insurance holder. If it is within the family the chances are higher in collectivist culture. It also states that a large majority of fraudulent activity in the insurance industry follow a limited number of patterns. These patterns are known to insurance experts who gain it through experience. This research was done based on the insurance industry in Turkey and data mining technique was used to identify patterns using the past data and then flagging the entries that needs further investigation, in this case it was a manual intervention. Research that has focused specifically on customer dishonesty and unethical behavior has documented the importance of customer attitudes toward insurance dishonesty [5]. For example, in [6], the research was to create Consumer Ethics Survey, which includes insurance fraud, to understand consumer perceptions of ethically questionable consumer behaviors. The more favorable individuals' ethical perceptions of fraud are, the more likely individuals are to engage in such behavior [7].

In [8], study is done to identify the type of hospitals and patients involved in fraudulent activities with respect to health insurance claims and frauds related to it. The study was done based on the data of elderly people who had one or more of the six medical conditions that are prone to fraudulence. The data was collected for a period of four years for the study. Researchers obtained longitudinal claims data linked with social security death records, hospital characteristics, and state/year-level antifraud enforcement efforts. The study showed that with increased enforcement, certain types of hospitals and patients returned with lower bill amounts without any adverse effects or change in quality of the treatment. There is a common assumption made in economics that if there is monetary benefit associated, people always misrepresent their personal information, and this was studied in [9]. A

representative sample was selected and the extent of lying cost was measured and it was found that in spite of having monetary benefits for lying people did not lie as much as it was expected. Further experiments were also done to make sure that mode of communication did not play an influencing factor in it. Earlier studies, like in [10], [11], [12] and [13] have proved that certain observable and measurable characteristics play an important role in influencing the amount of lies that is being told. In [14], preference for honesty is studied and based on the class of economic environments a positive result is obtained. In [15], the study was on the social, economic and the moral dimensions of insurance fraud and in this paper it was mentioned as to how complex insurance fraud is to detect and that about 10% of the claims are fraudulent. However, it has also been mentioned that the number could be much higher because most of the cases are not reported and an amicable solution is reached privately. The study also discusses on how to reduce insurance claim fraud and one major input was on using moral cues during claims can reduce the chances of fraud. Having the internal honesty monitors at the time of decision making would help a lot in reducing the chances of customers taking advantage of chances when they can cheat. There has also been studies like in [16] which state that the researcher interaction can also influence the response of the responder both knowingly and unknowingly, for example, the introduction by the researcher can have a huge impact on how the respondent replies, especially in case of ethics related studies. In [17], the study is on ethically questionable consumer behavior and how this is influenced by the individual perception about ethicality, social support, benefits and the consequences of being unethical. Ethics does play an important role in insurance fraud, it is the choice that the customer takes, between right and wrong as discussed in [18].

III. EXPERT INTERVIEW

An interview was conducted with an expert in this field who is involved in identification of insurance fraud in India. From the literature review it was clear that the techniques would depend on the country as the culture and the attitude of customers towards fraud would change based on country. Enforcement of law also would change based on the country that is being studied upon and thus the pattern and the methods used for fraud would differ. An insurance expert would be the right person to identify the most apt methods that are currently used to detect insurance fraud and these methods could be very useful in developing the model to identify the pattern of frauds that could be used in flagging the claim. Based on the input from the expert around ten variables were identified and divided into two groups as measurable and non-measurable. Measurable are those variables that could be identified from the claims data and the non-measurable are those variables which are subjective in nature and cannot be identified from the claims data. So based on the input from the expert, and

availability of data most ideal and apt variables from the measurable variables were selected and used for the study.

IV. HYPOTHESES DEVELOPMENT

Hospitals having different billing for insurance patients and change in physician comprised the non-measurable variables. Hospitals visited, claimed amount, type of medical condition, type of payment, maximum claimable amount, cost on investigations, age, gender, location and amount spent room rent as the measurable variables. Among these age, gender, total insured amount, total bill amount and type of payment were used for this study. The focus was to identify patterns based on these data.

A. Age

Study was done to check if the age of the patient had a significant impact on the claim being a fraudulent one. The hypothesis (alternate) developed for the same is as follows.

Ha1 – Age of the patient does have an impact on the claim being a fraudulent claim

B. Gender

Study was done to check if the gender of the patient had a significant impact on the claim being a fraudulent one. The hypothesis (alternate) developed for the same is as follows.

Ha2 – Gender of the patient does have an impact on the claim being a fraudulent claim

C. Type of Payment

Study was done to check if the type of the payment had a significant impact on the claim being a fraudulent one. If we could identify that claims from a particular type of payment were fraudulent, that would help a lot in flagging the claim. The hypothesis (alternate) developed for the same is as follows.

Ha3 – Type of the patient does have an impact on the claim being a fraudulent claim

D. Total Insured Amount

Study was done to check if the total insured amount had impact on the claim being a fraudulent one. If we could identify that claims from a particular range of insured amount were fraudulent, that would help in flagging the claim. The hypothesis (alternate) developed for the same is as follows.

Ha3 – Insured amount does have an impact on the claim being a fraudulent claim

E. Total Bill Amount

Study was done to check if the total bill amount had impact on the claim being a fraudulent one. If we could identify that claims from a particular range of total bill amount were fraudulent, that would help in flagging the claim. The hypothesis (alternate) developed for the same is as follows.

Ha3 – Total Bill amount does have an impact on the claim being a fraudulent claim

V. DATA AND METHODOLOGY

The data was collected based on the claims requests received by one of the regional offices of a prominent health insurance provider. The data contains details of the claim being accepted or rejected and the details about the insurance. Based on this data prediction is made to flag the data as one with higher chances of being fraudulent one and then it is checked with the actual data to measure the success rate. The data collected has more than three thousand five hundred observations of which about ten percentage are claims that are rejected. Supervised learning using artificial neural network were employed using the tool– NeuralTools 7.5 by Palisade Corporation

Neural Network Analysis

Neural Tools are widely used in helping managers take business decisions based on the past data by predicting the likelihood of an event taking place. The system develops the most optimal model and checks for the fitness of the model by splitting the data into train data and test data. Then the predictions are made on test data and the error percentage is calculated. Once the model is finalized it can be used to predict the results in future. NeuralTools from Palisade Corporation was used and the model was developed based on City, Age, Gender, Total Insured Amount, Type of Claim and Total Bill Amount.

VI. RESULTS AND DISCUSSION

PNN and MLFN 2 to 6 nodes were done on the data. City, Gender of the applicant and type of claim were the categorical predictor variables and Age of the applicant, Total insured amount, total bill amount where the continuous independent variables used for the analysis. If the amount was paid to the applicant or not was the categorical dependent variable. The results from the analysis are as follows.

TABLE. I SUMMARY OF ANALYSIS

Summary	
Net Information	
Name	Net Trained on Data Set #1

Configurations Included in Search	PNN, MLFN 2 to 6 nodes
Best Configuration	MLFN Category Predictor (4 nodes)
Independent Category Variables	3 (City, Gender, Type Of Claim)
Independent Numeric Variables	3 (Age, Sum Insured Amt, Bill Amt)
Dependent Variable	Category Var. (Insurance Paid)
Training	
Number of Cases	2940
Training Time	02:00:00
Number of Trials	2035577
Reason Stopped	Auto-Stopped
% Bad Predictions	6.3265%
Testing	
Number of Cases	722
% Bad Predictions	5.9557%
Data Set	
Name	Data Set #1
Number of Rows	3689
Manual Case Tags	NO

The following table shows the error percentage from various analysis done.

TABLE. II BESTNET SEARCH RESULTS

	% Incorrect
PNN	6.51%
MLFN 2 Nodes	5.96%
MLFN 3 Nodes	5.96%
MLFN 4 Nodes	5.96%
MLFN 5 Nodes	6.65%
MLFN 6 Nodes	6.93%

VII.CONCLUSION

Insurance fraud is one of the most prevalent and pressing issues in the health insurance sector. The amount of time and money spent on detecting health insurance fraud is very high and with the higher availability of data organizations can use data to automate their fraud detection mechanisms. Decision tools can be used for the same and from the analysis and the results of the data we can conclude that, for the given set of data the best fitting model is MLFN 4 nodes model. The independent variables are city of claim by the applicant, gender of the applicant, type of claim, i.e. it is reimbursement or cashless, age of the applicant, Total sum insured by the applicant and the total bill amount for the ailment. Based on the past data for these variables, neural networks is built to develop a model and the MLFN 4 nodes model has the least error percentage. This

model can be used in future to predict the possibility of a fraud based on the variables. This can help in reducing the manual work involved in detecting the fraud. The model would flag the potential fraudulent claims and manual intervention would be required only for the flagged ones, thereby reducing the cost and time involved. Since machine learning approach is used the precision of the predictions will improve with more data flowing in and the model training itself.

VIII.LIMITATION AND DIRECTION FOR FURTHER RESEARCH

The present study has not considered major factors like the type of hospital and the type of ailment due to the constraints in data that was available. If these factors are also considered, the analysis and the predictions would be more accurate.

REFERENCES

- [1] Yusuf Ibrahim Arowosaiye. (2012). Fighting Economic and Financial Crimes For National Development: The Malaysian Visioan 2020. Current Law Journal (CLJ), 1(LNS(A)), xxix.
- [2] Miyazaki, A. D. (2009). Perceived ethicality of insurance claim fraud: Do higher deductibles lead to lower ethical standards?. Journal of Business Ethics, 87(4), 589-598.
- [3] Viaene, S., Derrig, R. A., Baesens, B., & Dedene, G. (2002). A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection. Journal of Risk and Insurance, 69(3), 373-421.
- [4] Kirlidog, M., & Asuk, C. (2012). A fraud detection approach with data mining in health insurance. Procedia-Social and Behavioral Sciences, 62, 989-994.
- [5] Brinkmann, J., & Lentz, P. (2006). Understanding insurance customer dishonesty: Outline of a moral-sociological approach. Journal of Business Ethics, 66, 177-195.
- [6] Vitell, S. J., & Muncy, J. (1992). Consumer ethics: An empirical investigation of factors influencing ethical judgments of the final consumer. Journal of Business Ethics, 11, 585-597.
- [7] Vitell, S. J. (2003). Consumer ethics research: Review, synthesis and suggestions for the future. Journal of Business Ethics, 43, 33-47.
- [8] Becker, D., Kessler, D., & McClellan, M. (2005). Detecting medicare abuse. Journal of Health Economics, 24(1), 189-210.
- [9] Abeler, J., Becker, A., & Falk, A. (2014). Representative evidence on lying costs. Journal of Public Economics, 113, 96-104.
- [10] Dreber, A., & Johannesson, M. (2008). Gender differences in deception. Economics Letters, 99(1), 197-199.
- [11] Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. Psychological science, 18(9), 803-809.
- [12] Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. European Economic Review, 56(8), 1645-1655
- [13] Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. Journal of marketing research, 45(6), 633-644.
- [14] Kartik, N., Tercieux, O., & Holden, R. (2014). Simple mechanisms and preferences for honesty. Games and Economic Behavior, 83, 284-290.
- [15] Tennyson, S. (2008). Moral, social, and economic dimensions of insurance claims fraud. Social Research, 1181-1204.
- [16] Miyazaki, A. D., & Taylor, K. A. (2008). Researcher interaction biases and business ethics research: Respondent reactions to researcher characteristics. Journal of Business Ethics, 81(4), 779-795.

[17] Fukukawa, K. (2002). Developing A Framework for Ethically Questionable Behavior in Consumption. *Journal of Business Ethics*, 41(1), 99-119.

[18] Brinkmann, J. (2005). Understanding insurance customer dishonesty: Outline of a situational approach. *Journal of Business Ethics*, 61(2), 183-197.