

Analysis of Insurance Anti-fraud Game Based on Random Forest and XGBoost Model

Zheyu Li

Guangdong university of finance, Guangzhou, China

Abstract: With the rapid development of China's health insurance industry, cases of health insurance fraud are also on the rise year by year, which has had a negative impact on the development of the health insurance industry and social stability. This article analyzes the insurance anti fraud game based on the random forest and XGBoost model. The purpose of game analysis is to use game theory to analyze the occurrence process of health insurance fraud behavior, and then identify the main factors that affect moral hazard and insurance agency expenditure costs. And an insurance anti fraud game model was constructed in the article. Game theory can be used to analyze the process of health insurance fraud and identify the main factors that affect moral hazard. The most common type of health insurance fraud in China is doctor-patient collusion fraud. We will only discuss this type of fraud and not analyze other types. And through research, it has been shown that the XUBoost model has significant advantages in accuracy and recall, while the random forest has a higher accuracy. Obviously, the XUBoost model has good performance.

Keywords: Random Forest, XGBoost Model, Insurance Anti Fraud, Game Analysis.

1. Introduction

Today, the problem of insurance fraud has been concerned by international insurance, because insurance fraud has caused many negative effects. Insurance fraud refers to the illegal and criminal acts that the insured, the insured or the beneficiary take various deceptive means to defraud the insurer for the purpose of illegally possessing insurance indemnity [1]. The interests of insurers will be affected by insurance fraud, and social property and social stability will also be threatened by insurance fraud. What is more serious is that the losses caused by insurance anti-fraud will be difficult to estimate, and insurance and other related industries will be threatened and affected by insurance fraud [2]. A large number of insurance anti-fraud have done great harm to insurance companies and society. For insurance companies, paying compensation that should not be spent reduces the company's income and solvency, and also forces insurance companies to raise insurance rates, which is not conducive to participating in the competition in the insurance market, but also increases the burden on consumers [3]. The competition in the insurance industry is very fierce, and there is little exchange of information about fraudulent insurance and compensation within the industry. This enables repeat insurance fraudsters to defraud several insurance companies. However, in order to maintain their reputation, some insurance companies deal with insurance fraud in a low-key way, and even condone the behavior of fraudsters. Because of the complexity of insurance anti-fraud, the research work is difficult, and because of the variety of insurance anti-fraud, it is not convenient for researchers to make detailed statistics, so both domestic and foreign insurance companies feel "headache" about insurance fraud [4]. Some insured people ignore the terms of the contract, violate the insured's obligation to tell the truth as stipulated in the insurance contract, and keep silent about their past medical history. After underwriting, the insurance company sought medical treatment for related diseases, and concealed the past medical history during the medical treatment, resulting in the

insurance company being unable to obtain evidence and having to pay compensation. Therefore, preventing insurance fraud is not only the guarantee of the insurance industry, but also an important means of self-control for insurance companies. In this regard, this paper analyzes the insurance anti-fraud game based on random forest and XGBoost model. The purpose of game analysis is to analyze the occurrence process of health insurance fraud by using game theory, and then find out the main factors that affect moral hazard and the expenditure cost of insurance institutions [5]. This paper defines the additional expenses of insurance companies caused by health insurance fraud as the cost of fraudulent claims. Random forest and XGBoost algorithm are used to evaluate the importance of features, and data-based heuristic statistical analysis is used to rank the features in a weighted comprehensive way. XGBoost algorithm is used to establish a model, which is compared according to the model evaluation index, thus providing reference for credit evaluation researchers and related enterprises and institutions to choose models [6].

2. Causes of Insurance Anti-fraud

2.1. The insurance company's own management is insufficient

The phenomenon of "valuing quantity over quality" generally exists in China's insurance companies, with extensive management and assessment of the treatment and promotion of business personnel taking business volume as a single measure. Due to the lack of standardization in management, the problem of insurance fraud is more serious in China, and the proportion is estimated to be between 28% and 39% [7]. According to the statistics of the United States in 2022, the amount of health insurance fraud is \$28 billion, and it is rising every year. In the field of health insurance in China, the most common is the collusion fraud between the insured and the medical service provider. Insured or beneficiary often colludes with hospitals, clinics and doctors to defraud insurance money. Doctors may have a good

personal relationship with the insured or beneficiary, or doctors may take kickbacks into account their own interests, or hospitals may prescribe more drugs to attract more patients to see a doctor, even if they are not sick, they can prescribe hospitalization expenses and medical expenses. There are also some insured people who let others pretend to see a doctor to defraud the insurance money. It is more difficult for an insurance company to control if the insurance company's claim adjusters, policyholders or beneficiaries collude with doctors to deceive the insurance company[8]. Limited by space, this paper only discusses the game analysis of their collusion, and at the same time, only discusses the fraudulent behavior of agents in the fact of illness, without considering the fraud of different types of diseases. Some underwriting and compensation personnel are of low quality, the company's management rules are not strict, insurance companies absorb a large number of unqualified policies, and they fail to conduct detailed investigations when making claims, which makes some insurance fraud cases successfully compensated[9]. There are even cases where insiders of insurance companies participate in insurance fraud.

2.2. Incomplete social medical system

China's medical system is currently in a period of reform, and the relevant systems are not yet perfect. The sharing of medical information has not yet been established in most regions. Health insurance fraud is an illegal and criminal activity in which policyholders or related personnel deceive insurers into compensating for insurance benefits by fabricating or exaggerating insurance accidents and concealing the truth[10]. In addition, some insured individuals may excessively exaggerate symptoms or are unwilling to clearly explain existing symptoms. Even if the policyholder deliberately conceals their past medical history or existing symptoms with superb techniques, it is inevitable to reveal some traces of fraud, which can sometimes be detected by vigilant claims personnel. Due to the serious impact of insurance fraud on the reputation of insurance companies, the survival and development of insurance companies will be seriously threatened by insurance fraud, which will discourage buyers who originally had the desire to purchase insurance. Insurance companies will also reduce their business scale due to the increase in their own costs, which will have a very serious impact on the entire insurance market. At present, there is no effective supervision mechanism to regulate insurance fraud involving collusion between doctors and patients, which poses certain difficulties for insurance companies in risk control and investigation and evidence collection.

2.3. Incomplete laws and regulations

The Insurance Law and Criminal Law stipulate the crime of insurance fraud, but there is no relevant provision for attempted insurance fraud. Some policyholders, without undergoing any treatment, make doctors panic and issue documents for surgical and hospitalization expenses, and then claim compensation from insurance companies based on these documents. In the United States, poor people and new immigrants who have medical insurance but are struggling in life sign cooperation agreements with "Snake Head" in order to obtain insurance benefits obtained together with "Snake Head" and clinics, and are then sent to "designated hospitals" in batches. In addition to fabricating insurance subject matter insurance fraud claims, it is common in health insurance to

intentionally delay the observation period for illness or death, report the case after the observation period, place the accident time within the responsibility period, and deceive the insurance company for insurance benefits. Some insured individuals collude with medical personnel or entrust counterfeit vendors to provide false medical records, receipts, and other false evidence to claim compensation for the purpose of payment. In practice, the public security organs are unable to pursue criminal responsibility for attempted insurance fraud cases, and there are no relevant punishment measures. In insurance fraud, insurance companies are in a disadvantaged position, and some judicial rulings blindly lean towards the insured, leading to judicial injustice and objectively condoning insurance fraud.

3. Analysis of Insurance Anti-fraud Game Based on Random Forest and XGBoost Model

3.1. Construction of Insurance Anti-fraud Game Model

By establishing a game model of health insurance fraud, we can analyze the occurrence process of health insurance fraud by using game theory, and then find out the main factors affecting moral hazard. The most common type of health insurance fraud in China is doctor-patient collusion fraud. Here, only this type of fraud is discussed, and other types are not analyzed. This paper attempts to analyze the insurance anti-fraud game based on random forest and XGBoost model. XGBoost is called Extreme Gradient Boosting, which can be regarded as the optimization of GBDT. GBDT adopts the idea of gradient descent when generating each tree, taking all single decision trees as the basis and taking the goal of minimizing the loss function as one step further. Different from GBDT model, XGBoost model can automatically use CP U for multi-thread parallel calculation, and expand the loss function by Taylor formula in the second order, adding regularization term after the loss function to restrain the decline of the loss function and the overall complexity of the model. The overall objective function of XUBoost is:

$$L(\varphi) = \sum_i l(y_i) + \sum_k \Omega \quad (1)$$

Where $L(\varphi)$ is the loss function, which is used to ask the difference between the predicted value y_i and the real value; $\Omega(\varphi)$ is the regularization term of the model, which is used to reduce the complexity and over-fitting problem of the model, and the model goal is to minimize the objective function.

In this model, the claimant is a concrete person with a tendency to maximize benefits. Claims adjuster has two optional strategies: violation or non-violation. What stimulates its violation is the huge gains when the violation is not discovered. There are four results in this game: first, the policyholder cheats and the insurer investigates, and the policyholder's income is a-b; Second, if the insured cheats and the insurer does not investigate, the income of the insured is C-A; Third, the insured does not cheat, the insurer investigates, and the insured has no income; Fourth, the

insured does not cheat, the insurer does not investigate, and both the insured and the insurer have no income. The expanded form of this game is shown in Figure 1.

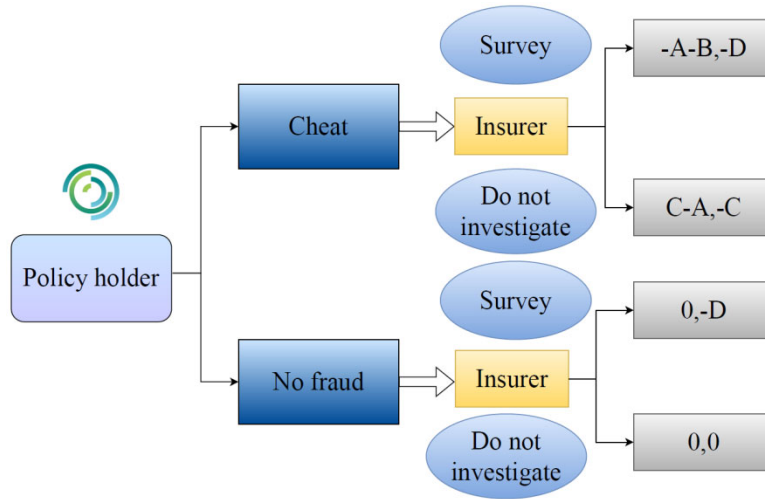


Figure 1. The extended form of the game

The problem of insurance fraud caused by information asymmetry in health insurance belongs to a static game problem with incomplete information. So, if the policyholder Zhang Yanzha, the insurer will investigate. According to the game tree, the benefits of the insurer's investigation are greater than the benefits of the insurer's investigation, because $C > D$, so $-C < -D$. If the policyholder is not fraudulent, the insurer will not investigate. Because when the policyholder is not fraudulent, according to the game tree, the benefits of the insurer's investigation are greater than the benefits of the insurer's investigation, i.e. $0 > -D$.

3.2. Analysis of Game Results

The higher the review cost, lower the fines, and lower the compensation for insurance companies, the greater the probability of insurance fraud occurring. According to general classification, insurance fraud can be divided into

three categories, namely direct fraud committed by claimants, agency fraud committed by claims agents, and internal and external collusion fraud committed by company claims personnel. The claims agent and the company's claims personnel are actually the company's agents when handling claims. The construction of the models in this article is implemented using Python's Sklearn program package. In addition to building the XUBoost algorithm model, a random forest model and a UBDT algorithm model are also established. When using the same dataset, these three models are compared and used. Table 1 shows the accuracy, precision, and recall of three algorithm models on the same dataset. From the table, it can be seen that the XUBoost model has significant advantages in accuracy and recall, while the random forest has a higher accuracy. Obviously, the XUBoost model has good performance.

Table 1. Comparison of Evaluation Index Values of Three Algorithm Models

Model	Accuracy	Precision	Recall
Random Forest	0.722	0.87	0.73
XGBoost	0.753	0.84	0.75
GBDT	0.708	0.81	0.78

Establish a comprehensive claims settlement procedure and a complete set of supervision and constraint mechanisms to continuously reduce supervision and inspection costs. In practice, the higher the amount of compensation the insured applies for, the greater the probability that the insurance company will review it, the greater the likelihood of fraudulent behavior being discovered, and the probability of fraudulent behavior occurring will decrease accordingly. Therefore, the probability of fraudulent behavior occurring in major illness insurance is relatively low, while in some common diseases such as colds and fever, doctor-patient collusion is more common. Under the random forest and XGBoost model, strict standards should be established for various stages within the company, including case filing, exploration, loss assessment, and compensation. It is

necessary to cooperate with relevant departments and sister units and take various measures to reduce the chances of claimant fraud.

4. Conclusions

This paper analyzes the insurance anti-fraud game based on random forest and XGBoost model, and constructs an insurance anti-fraud game model in this paper. From the model, it can be concluded that in order to reduce the probability of policyholders, insurance companies should strengthen the specialization and standardization of underwriting and claims review; Improve the internal supervision mechanism of insurance companies; Strictly deal with insurance fraud. Only in this way can we really reduce the proportion of insurance fraud. Moreover, we should

strengthen the training of relevant personnel of insurance companies, enhance their professional skills and professionalism, and let salesmen have a sense of responsibility. The results show that the models in this paper are all built by python's sklearn package. In addition, in addition to the XUBoost model, the random forest model and UBDT model are also built, and the three models are compared under the same data set. Among the three models, XUBoost model has obvious advantages in accuracy and recall, while the random forest has higher accuracy. Obviously, XUBoost model has better performance. The admission system should be implemented for the underwriting and compensation personnel of health insurance, and their determination and attitude to prevent fraud should be publicized to strengthen the anti-fraud efforts. In addition, speed up the reform of personnel system and establish an effective talent incentive mechanism. At present, the fraudulent means of health insurance are diversified, and insurance companies urgently need to strengthen the construction of internal control system, improve the comprehensive quality of business personnel, identify fraud and establish a complete anti-fraud system, so as to promote the development of health insurance.

References

- [1] Yu L. Thinking about Regulating Basic Medical Insurance Payment Fraud[J]. China Health Insurance, 2022, 26(7):10-13.
- [2] Guang-Hao H, Economics S O, University P. Applied Research of Digital Fraud and Financial Technology Anti-fraud[J]. Taxation and Economy, 2019, 20(7):12-16.
- [3] Pathmanathan P R, Aseh K. Identifying Predictors of Perceived Claims of Insurance Fraudulance[J]. Archives of Business Research, 2021, 9(6):68-76.
- [4] Li P, Shen B, Dong W. An Anti-fraud System for Car Insurance Claim Based on Visual Evidence[J]. The Singapore Economic Review, 2019, 16(7):10-17.
- [5] Verykios V S, Stavropoulos E C, Zorkadis V, et al. Sensitive data hiding in financial anti-fraud process[J]. International Journal of Electronic Governance, 2022, 20(11):10-14.
- [6] Wu T C M, Yen C T, Huang C C, et al. Social Insurance under Fraud and Redistributive Taxation[J]. The Singapore Economic Review, 2019, 16(9):10-15.
- [7] María skarsdóttir, Ahmed W, Antonio K, et al. Social Network Analytics for Supervised Fraud Detection in Insurance[J]. Risk Analysis, 2021, 14(4):3-7.
- [8] Haiche L. Research on Anti-Fraud Legal Measures of Commercial Medical Insurance[J]. Medicine and Society, 2022, 23(7):10-15.
- [9] Wang S, Zhuang J, Zheng J, et al. Application of Bayesian hyperparameter optimized random forest and XGBoost model for landslide susceptibility mapping[J]. Frontiers in Earth Science, 2021, 9(8):10-16.
- [10] Wang Xiaohui, Sean, Li Junqing, et al. Research on the improved method of XGBoost based on genetic algorithm and random forest [J]. Computer Science, 2020, 47 (s02): 6-9.