International Journal of Applied Science and Mathematics
Volume 10, Issue 4, ISSN (Online): 2394-2894

Insurance Fraud Identification Based on XGboost

Jie Huang
School of Economics, Zhejiang University of Finance and Economics, Hangzhou, China.
Corresponding author email id: eaasy0620@163.com

Date of publication (dd/mm/yyyy): 24/07/2023

***Abstract* – In China, motor vehicle insurance plays a very important role in property insurance, and its operating status is closely related to the operating stability of property insurance companies. The growing fraud problem has many adverse effects on other honest insureds and insurance companies. The fraud identification model can identify high-risk insured persons, so that the insurer can identify suspicious cases more quickly and accurately through this model in practical work. However, due to some special characteristics of the auto insurance claim data set, such as unbalanced data distribution and various types, it has not yet been able to build a fraud identification model that adapts to a wide range. In view of the above problems, this paper cleans the data according to the characteristics of imbalance and high feature dimension of auto insurance data, and selects the features with strong prediction ability by filtering feature selection. Then, on the basis of comparing various models, this paper chooses the best-performing XGboost model to construct the insurance fraud identification model and evaluate the model. The analysis found that the indicators of XGboost are high, indicating that the performance of the model is good and can be used to accurately identify insurance fraud. This provides a reference for the insurance company platform to provide users with insurance products.**

***Keywords* – Insurance Fraud Identification, Filtered Feature Selection, Model Comparison, XGboost.**

I. INTRODUCTION

In recent years, the development of China's insurance industry is booming, in 2020 under the influence of the epidemic insurance industry is still growing steadily, the original premium reached 4.53 trillion yuan, an increase of 6.12%, a total of 1.39 trillion yuan of premium payouts, an increase of 7.86%, the overall industry solvency is sufficient to achieve the general economic situation to help support the role of the weak links of the social economy. Meanwhile, with China becoming the largest motor vehicle consumer market and the implementation of the commercial vehicle fee reform, the auto insurance market has huge potential, and the hidden auto insurance fraud problem has become more and more serious, as disclosed by the China Insurance Institute, auto insurance fraud brings more than $20 billion in losses to the industry every year, with a leakage ratio of up to 20%. On the one hand, auto insurance fraud increases the pressure of insurance companies to pay insurance claims, causing loss of operating profit for insurance companies and affecting the normal operation order of the insurance market; on the other hand, auto insurance fraud violates the provisions of insurance contracts, which is contrary to the social values of fairness, honesty and justice. In fact, in recent years, China's motor insurance fraud cases have shown the development trend of concealment, professionalism and group, and new ways of auto insurance fraud are emerging and more specialized. At the same time, because the subject of insurance is mobile and information asymmetry, its fraud risk is difficult to prevent, making car insurance fraud frequently. Therefore, motor vehicle insurance fraud not only harms the interests of both the insured and the insurer, but also hinders the healthy development of the auto insurance market and disrupts the public order, so the task of strengthening the anti-fraud efforts of motor vehicle insurance and finding new identification methods cannot be delayed.

Academic as well as practical attempts on insurance fraud identification methods have been carried out, such

as: common expert systems, generalized linear model Probit, etc., but all the above methods have disadvantages such as long computing time and low identification correct rate. Although some scholars have also applied ELM, SVM, Logistic regression and other techniques, they are limited to data sets with few variables with good results. Other scholars have also conducted research on other advanced algorithms, but most of them choose a single model and do not take into account the impact of data imbalance problem on the model. With the advent of big data era, the data will present multi-scale, high-frequency, high-dimensional and other characteristics, so all the above techniques have their limitations, which shows the urgent need for the introduction of new technologies in the insurance industry, for which this paper will try to explore the insurance fraud identification problem based on the integrated learning approach.

II. RELATED WORK

Back in 2010, Ye based on the empirical data of motor insurance claims collected from car insurance research in Jiang, Zhejiang and Shanghai, and obtained the factor characteristics and related explanations about motor insurance fraud identification in China through the empirical analysis of binary choice model under logistic distribution; meanwhile, Ye pointed out the shortcomings of the current insurance fraud identification system in China and proposed relevant countermeasures for the construction of insurance integrity we also point out the shortcomings of the current insurance fraud identification system in China, and propose relevant countermeasu- -res to build insurance integrity [1].

In 2011, Ye tested the effectiveness of BP neural network in China's insurance fraud identification based on the sample data of auto insurance claims from property insurance companies in China; meanwhile, in order to try the effective integration of statistical regression and neural network, the indicators obtained from Logit discrete model were used as refined explanatory variables to train the neural network, and the two prediction results were compared and analyzed to construct China's insurance claims indicators perfection and the countermeasu- -res for the improvement of the neural network fraud identification technology [2].

In 2017, Yu achieved effective identification of gangs with very small probability of occurrence but highly suspicious by introducing the concept of generalized gangs and using matrix-based similarity calculation, rank ordering and transformation algorithms, which have more accurate and efficient practical applications compared to traditional methods [3].

In the same year, Li studied the auto insurance fraud model based on extreme learning machine from the theory of extreme learning machine, introduced a generalized linear model, proposed a generalized linear model-extreme learning machine (GLM-ELM) auto insurance fraud identification model, conducted empirical analysis and drew conclusions. The results show that: Compared with the traditional model, the GLM-ELM based auto insurance fraud identification model can better identify the fraud information in the claim data [4].

In 2018, Liu used BP neural network to achieve active identification of health insurance fraud, and used logistic regression analysis to improve the neural network model to reduce the interference of weak factors on neural network identification. In addition, to cope with the problem of scarcity of fraud data, the model of training the neural network to simulate the function curve by taking only normal data is used. The empirical evidence shows that the method has a good ability to identify health insurance fraud [5]. Liu explores the mechanism of insurance fraud by extracting the fraud correlation factors of typical insurance litigation cases,

establishing regression models, clarifying the institutional causative factors of insurance fraud, and assessing the risk of insurance fraud in different fields based on empirical studies [6].

In 2019, Li empirically tested the insurance fraud problem based on Bagging, Random Subspace, and Random Patches algorithms and analyzed the applicability issues of each algorithm and with base learners, and the effect of the number of base learners on the performance of the algorithms. The analysis found that the Random Patches algorithm performed best and Bagging had the shortest runtime for the insurance fraud identification problem [7]. Lu proposed a TLSTM-based health insurance fraud identification framework to address the challenges of unbalanced samples and uneven temporal distribution, using the user's historical medical visit behavior sequence as the input to the TLSTM model to predict the reasons for patient readmission and the algorithm compares the difference between the model output and the user's current medical behavior to determine the possibility of fraud. Experiments show that the algorithm is significantly better than existing algorithms in terms of fraud recognition accuracy [8].

In 2021, Yang designed and implemented a GAPSO-BP neural network model based on a combination of particle swarm algorithm and genetic algorithm for optimization. The model optimizes the initial connection weights and thresholds to solve the problem of easy convergence to local extremes and improve the performance of the neural network. The empirical results showed that the overall prediction accuracy of the optimized model increased to 94.5%, especially for fraudulent claims cases [9]. Finally, Chen summarized the findings and made some suggestions for the application of machine learning in auto insurance fraud identification in China, and made a research outlook [10].

In the same year, Lu introduced the specific research progress of machine learning models applied to auto insurance fraud detection from foreign and domestic perspectives and conducted a macro comparison; selected representative machine learning models for modeling based on high-quality auto insurance data provided by a domestic auto insurance company in the past five years and conducted a comprehensive test and analysis; discussed the future of auto insurance fraud detection technology [11]. Xie used COSO theory to help analyze the problems of X insurance company in dealing with fraud risk and the reasons for the problems. In addition, the company's risk response strategy was analyzed, and a suitable risk strategy was selected to guide the optimization of the response measures [12].

In 2022, Yan proposed an improved particle swarm algorithm (PSO) to optimize T-S fuzzy neural network (TSFNN) for auto insurance fraud detection model, using the improved particle swarm algorithm to iteratively find the optimal network coefficients and affiliation function parameters of TSFNN, and introducing nonlinear time-varying inertia weights and natural selection mechanism to construct an improved particle swarm algorithm. The simulation results show that the improved PSO-TSFNN auto insurance fraud detection model is easy to implement and has higher fraud recognition rate, prediction accuracy, and good robustness compared with the traditional TSFNN, PSO-TSFNN, and LDWPSO-TSFNN models [13].

III. Empirical Analysis

A. *Data Source*

The data in this paper uses the user claim data of the insurance company, which contains 38 columns of data features, such as policy_id, age, customer_months, policy_bind_date, policy_state, policy_csl, policy_deductab-

-le, policy_csl annual_premium, umbrella_limit, etc. The original data are shown in Table 1.

Table 1. The sources of data.

Policy_id	Age	Customer_Months	Policy_State	...	Auto_Model	Fraud
122576	37	189	C	...	Maxima	0
937713	44	234	B	...	Civic	0
680237	33	23	B	...	Wrangler	0
513080	42	210	A	...	Legacy	1
192875	29	81	A	...	F150	0

B. *Data Cleaning*

In daily life, the data obtained is often incomplete, there are some missing values, and some features have large differences between them, and are not stable enough to appear many outliers, so it is particularly important to pre-process the data. Therefore, it is necessary to deal with this problem before modeling to ensure that the quality of the data can meet the task of data mining. In this paper, the work of data cleaning includes filling in missing values, exploring outliers and transforming data features.

The first is the treatment of missing values, there are many missing values in this dataset, in order to avoid the impact of missing values on the prediction performance of the model, the missing values need to be filled. The approach taken in this paper is to eliminate this feature if there are more missing values (more than 30% missing); if there are fewer missing values, the Lagrangian interpolation method is used to interpolate the missing data.

Next, outliers are handled. Through data exploration, it is found that some features contain some outliers, however, in the field of financial risk control, the outliers may represent some special cases and can be considered as a risk, so the outliers are not handled in this paper.

Finally, the features are transformed. Firstly, the obvious irrelevant variables and the obvious redundant variables are eliminated based on business experience. The date of insurance binding (policy_bind_date) and the date of accident (incident_date) are used to create the derivative variable insured_days by subtracting these two dates, and the original variable is eliminated. For the time variable date of purchase, the date of purchase is converted into the length of time (in years) until the car is purchased at the time of the accident. Subsequently, the categorical variables were coded using the category mapping method for ordered categories and the Label Encoder method of the sklearn data preprocessing module for unordered categories. After all the data cleaning was completed, there were 36 features in the dataset.

C. *Factor Selection*

Feature selection is the process of finding and selecting the most useful features in a dataset, which is a key step in the machine learning process. Not all of these features are useful for subsequent modeling, and according to their usefulness for modeling or not, the features can be classified into two categories: useful and useless, which can be called “relevant features” and “irrelevant features”, respectively. Then the process of selecting relevant features from these features is called feature selection. Feature selection before modeling can not only reduce the dimensionality of the data, reduce the processing time of useless features, i.e., reduce the complexity

of the computation time, but also improve the credibility of the model, and can more fully understand the information implied in the features.

In this paper, we mainly use filtered feature selection for factor selection. The filtering method first selects features for the dataset according to some rules (scoring each feature according to divergence or relevance, setting a threshold or the number of thresholds to be selected, and thus selecting features that satisfy the conditions), and then train the learner. This is equivalent to "filtering" the initial features with the feature selection process and then training the model with the filtered features.

(1) *Removal of Covariance Features*

Covariance features are features that are highly correlated with each other. In the field of machine learning, high variance and low model interpretability lead to reduced generalization ability on the test set. In this paper, we find covariance features based on a specified correlation coefficient value (set a threshold of 0.9). For each pair of correlated features, it identifies one of them to be removed.

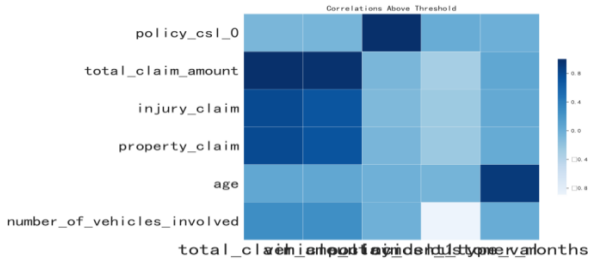

Fig. 1. Characteristic heat map for exceeding the threshold.

The covariance can be well visualized using heat maps. Figure 1 shows all features with at least one correlation above a threshold (0.9), for a total of five covariance features, removed.

(2) *Remove Zero Importance Feature*

For supervised machine learning problems, where the labels of the training models exist and are non-deterministic, the features with zero importance are found according to the gradient boosting machine (GBM) learning model. In this paper, we mainly use a tree-based machine learning model to find the feature importance. It is also possible to use feature importance in feature selection by removing zero-importance features. In a tree-based model, zero-importance features are not used to segment any nodes, so they can be removed without affecting model performance. The gradient boosting machine from the LightGBM library in python is used to obtain the feature importance. In addition, the model is trained using early stopping (early stopping) to prevent overfitting on the training data.

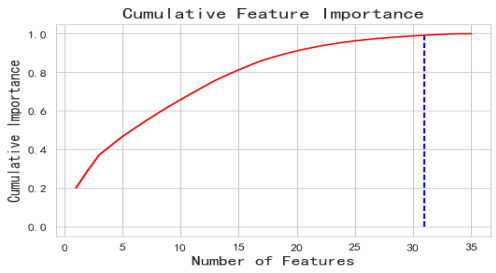

Fig. 2. Cumulative Importance of Number of Features Chart.

The model identifies 1 feature with zero importance and removes it. Figure 2 shows the cumulative importance of the corresponding number of features, and the blue vertical line marks the threshold of 95% cumulative importance.

(3) *Remove Low Importance Features*

The next approach is based on zero importance identification, using the feature importance from the model for further selection. The features with the lowest importance are found by the previous model, and these features do not contribute to the specified total importance. A threshold of 0.95 is set to find the least important features, which achieve 95% importance even without these features.

Table 1. Importance table of all features.

Feature	Importance	Normalized_Importance	Cumulative_Importance
insured_hobbies_val	101.6	0.200078771	0.200078771
incident_severity_val	45.2	0.089011422	0.289090193
insured_days	41.9	0.0825128	0.371602993
...	...	...	...
property_damage_val	1.1	0.002166207	1
policy_csl_1	0	0	1

Table 2 shows the importance ranking of all features, and based on the previous cumulative importance graph and this information, the gradient boosting machine considers many features irrelevant to learning. Twelve factors with low importance are identified, and the remaining 23 factors can reach 95% importance.

After data cleaning and feature selection, a total of 19 most valuable features were retained from the original features, which will be used to model the various prediction models to be built later, and to facilitate the subsequent comparison of the prediction ability of various models.

D. *Model Construction*

The insurance fraud identification problem studied in this paper is a dichotomous problem, i.e., both fraud and non-fraud results. Therefore, five commonly used evaluation metrics for classification algorithms, Accuracy, Precision, Recall, fl-score and AUC (Areaunder Curve), are used to evaluate the classification effectiveness of the model. All data are also disrupted and divided into training and test sets in the ratio of 7:3. The training set is used to train the model and the test set is used for model evaluation.

Firstly, weak classifiers like decision trees and logistic regression are used for model building, and it is found that the classification effect is poor. Therefore, integrated algorithms are considered, and the decision tree based bagging and random forest in integrated learning bagging algorithm, and AdaBoost, XGBoost and Gradient Boosting Classifier in boosting algorithm are built respectively, and three models with default parameters are built, and these models are compared according to the evaluation index. Models according to the evaluation indexes, as shown in Table 3, to comprehensively analyze and select the appropriate model as the prediction model.

Table. 3. Effect comparison of model.

Type	Classifier	Accuracy	Precision	Recall rate	F1-score	AUC
Weak Classifier	SVM(linear)	0.5353	0.5227	0.6013	0.5593	0.5593
	SVM(SGD)	0.5417	0.5240	0.7124	0.6039	0.6039
	Native Bayesian	0.5385	0.5257	0.6013	0.5610	0.5610
	Logistic Regression	0.5449	0.5325	0.5882	0.5590	0.5590
	Decision Tree	0.4904	0.4904	1.0000	0.6581	0.6581
	KNN	0.6410	0.5887	0.8889	0.7083	0.7083
Bagging	Bagging	0.8077	0.8298	0.7647	0.7959	0.7959
	Random Forest	0.7949	0.7697	0.8301	0.7987	0.7987
Boosting	Ada Boost	0.8397	0.8411	0.8301	0.8355	0.8355
	Xg boost	0.8558	0.8506	0.8562	0.8534	0.8534
	Gradient Boosting Classifier	0.7981	0.7922	0.7974	0.7948	0.7948

Since this paper studies insurance fraud identification, the purpose is to be able to predict fraudulent users, so the metric of accuracy rate, although it represents the overall correct rate and has some reference value, is not an appropriate metric for what is studied in this paper, so this metric is not considered for now. For the rest of the indicators, the recall rate and accuracy rate are the relationship between each other, the recall rate can be understood as the proportion of actual fraudulent customers can be predicted by the model, and the accuracy rate refers to the proportion of customers predicted by the model as fraudulent, often we pay more attention to whether the fraudulent customers can be identified, if the model identifies fraudulent users accurately, it can avoid the majority of fraudulent cases and the loss and harm to the insurance company. If the model is accurate in identifying fraudulent users, the loss and harm caused to the insurance company can be avoided in most of the fraudulent cases, while for the case that the insurance company refuses to issue insurance due to the prediction of fraud but the customer is not actually fraudulent, the insurance company will also have certain losses, such as a decrease in business volume leading to a decrease in revenue, but compared with the two, the former loss is often greater. The AUC value, on the other hand, tends not to be affected when the positive and negative sample ratios change, while the other metrics will be more affected, so the AUC value should also be focused on.

From the data in Table 3, it can be seen that among the weak classifiers, KNN works better, with the highest return rate and AUC values, indicating that among the weak classifiers, KNN has the best prediction effect. In Bagging, the metrics are similar with Decision Tree-based Bagging and Random Forest. In Boosting, XGboost has better results than AdaBoost and Gradient Boosting Classifier. Collectively, the best performance in both recall and AUC values among all models is XGboost. Therefore, XGboost is finally used for the construction of insurance fraud identification model in this paper.

XGboost is a gradient boosting algorithm, residual decision tree, whose basic idea is to gradually add one tree after another to the model, making the overall effect (objective function decreases) improve with each addition of a CRAT decision tree. Using multiple decision trees (multiple single weak classifiers) to form a combined classifier, each time a new decision tree is constructed, a second-order Taylor expansion is used to optimize the

loss function based on the established decision tree residuals to obtain the weights of the new decision tree. It is fast in training because of automatic parallel computation using multiple threads of CPU and also accuracy improvement in algorithmic precision.

The accuracy of the model obtained according to the default parameters is 85.58%. Based on this, the parameters were adjusted using sklearn's grid search. The specific process is as follows.

Step 1: Determine the number of estimators for learning rate and tree_based parameter tuning.

max_depth is generally selected between 3 and 10, starting at 5. min_child_weight picks a relatively small value, starting at 1. Initial gamma = 0. subsample, colsample_bytree is generally taken as 0.5-0.9, with a starting value of 0.8. scale_pos_weight = 0. The optimal number of n_estimators is 300, which is obtained by incrementing the number of n_estimators from 100 to 900.

Step 2: Max_depth and min_child_weight parameter tuning.

Start by tuning these two parameters, as they have a big impact on the final result. First coarse tuning the parameters from a large range, and then small fine tuning. After outputting the results, we get the optimal max_depth of 5 and min_child_weight of 1.

Step 3: Tuning of gamma parameters.

Search gamma from [0,0.1,0.2,0.3,0.4] and get the optimal gamma as 0.1.

Step 4: Adjust the subsample and colsample_bytree parameters.

The ideal values for the subsample and colsample_bytree parameters are 0.8 and 0.8. We then take values around this value in steps of 0.05. The ideal values for the subsample and colsample_bytree parameters are still 0.8 and 0.9. The ideal values for the _bytree parameter are still 0.8 and 0.8. Then they are taken as final values.

Step 5: Regularization parameter tuning.

The next step is to reduce the overfitting by regularization, here using an adjustment of the reg_alpha parameter. The ideal value here is 0.1.

Step 6: Reduce the efficiency of learning.

The learning rate is reduced by a factor of 10 to 0.001, trees' number is increased to 5000, and score is increased.

The final tuning results are shown in Table 4:

Table. 4. Table of tuning results.

Parameter Abbreviation	Parameter Meaning	Initial Value Parameter	Adjustment Results
max_depth	Maximum depth	None	5
min_child_weight	Decision tree	1	1
gamma	Number of	0	0.1
subsample	Minimum sample weight of leaf nodes	1	0.8
colsample_bytree	Random sample	1	0.8

reg_alpha	Spanning Tree Column Sampling	0	0.1
learning_rate	L1 of weights	0.1	0.001
n_estimators	Regularization term	60	5000

Finally, this paper uses accuracy, precision, recall and F1-score to evaluate the XGboost model. The evaluation results are shown in Table 5 :

Table 5. Table of tuning results.

Classification Report	Precision	Recall	F1-Score
Not Default	0.88	0.89	0.88
Default	0.88	0.88	0.88
accuracy			0.88
macro avg	0.88	0.88	0.88
weighted avg	0.88	0.88	0.88

The accuracy, precision, recall, F1-score and AUC values reach 88%, indicating that the model has a good performance and can be used to accurately identify insurance fraud situations. Thus, it can be assumed that the integrated learning algorithm has improved the predictive capability of the model. In addition integrated learning has a strong advantage in the stability of the model in comparison, from which it can be inferred that the XGboost model has a huge advantage in identifying insurance fraud.

IV. CONCLUSION

Insurance fraud has caused a negative impact on China's insurance market. In this paper, we take motor vehicle insurance fraud data as an example, and research and apply insurance fraud identification through statistical model and machine learning model training and prediction, so as to help auto insurance anti-fraud system and combat auto insurance fraud. The summary of this paper is as follows.

(1) In the exploratory analysis and feature engineering of the data, this paper uses a variety of filtered feature selection for feature screening, and a total of 26 features are retained after screening, and the quality of the screened features is high.

(2) When comparing the modeling effects of the weak classifier and the integrated model, it can be found that the prediction ability of the weak classifier is poor and unstable when the amount of data is large, while the integrated model performs better, and models are found to be in the optimal state after combining all indicators, thus determining that the integrated learning model is more advantageous in identifying insurance fraud.

(3) Since in the actual insurance business, fraudsters generally account for a relatively small number, i.e., there is a positive and negative sample imbalance in the dataset, when accuracy, precision, and F1 scores are affected, while recall and AUC values are almost unaffected, the XGboost model is chosen as the final insurance fraud identification model for identifying whether insurance users will commit fraud.

In summary, the XGboost-based insurance fraud identification model constructed in this paper can be used as

a reference for insurance company platform when offering insurance products to users.

With the continuous development of machine learning algorithms and data applications, there may be differentiation in research and application fields in the future, which may lead to the emergence of oligarchs with absolute advantages in number. From the evolution of data islands to data oligarchs, insurance companies with a long history, rich data resources, and algorithm advantages may occupy more advantages in a favorable position. How to prevent resources from being controlled and not infringed on the basis of using data and algorithms to generate commercial value will become a topic of future research.

At the same time, in the direction of data collection, property insurance companies can work with other industries to build a data ecosystem, such as AIC detection system, maintenance intelligent detection, real-time traffic data and other fields, making the control of relevant risk information more accurate, so that property insurance companies can reduce information asymmetry in the closest part to consumers, which is also an important direction for future research.

REFERENCES

[1] Ye M.H. Factor analysis of motor vehicle insurance fraud identification in China-based on motor vehicle insurance claim data in Jiang, Zhejiang and Shanghai [J]. East China Economic Management, 2010, 24(02): 84-86.
[2] Ye M.H. Research on insurance fraud identification based on BP neural network-a case study of motor vehicle insurance claims in China [J]. Insurance Research, 2011, No. 275(03): 79-86. DOI: 10.13497.
[3] Yu Wei, Feng Genfu, Zhang Wenjun. Research on motor vehicle insurance fraud detection system and gang identification [J]. Insurance Research, 2017, No. 346(02): 63-73. DOI: 10.13497.
[4] Li Yaki,Yan Chun, Sun Haitang. Construction and research of auto insurance fraud identification model based on extreme learning machine [J]. Technology and Economy, 2017, 30(03): 96-100. DOI: 10.14059.
[5] Liu, Chong, Zhu, Xiyong. Medical insurance fraud identification based on BP neural network [J]. Computer System Applications, 2018, 27(06): 34-39. DOI: 10.15888.
[6] Liu Yi, Dong Jie. Key factors of insurance fraud risk and legal regulation [J]. Financial Theory Exploration, 2018, No. 180(04): 60-70. DOI: 10.16620.
[7] Li Xiufang, Huang Zhiguo, Chen Xiaowei.Research on the application of Bagging integration method in insurance fraud identification [J]. Insurance Research, 2019(04): 66-84. DOI: 10.13497.
[8] Cao Luhui, Qin Fenglin,Yan Zhongmin. TLSTM-based health insurance fraud detection [J]. Computer Engineering and Applications, 2020, 56(21): 237-241.
[9] Yang, Runze. Research on auto insurance fraud identification based on GA-PSO-BP neural network [D]. Hunan University, 2021. DOI: 10.27135/d.cnki.ghudu.2021.000764.
[10] Chen Xiyu. Exploring the problem of auto insurance fraud identification based on machine learning algorithm [D]. Southwest University of Finance and Economics, 2021. DOI: 10.27412.
[11] Bingjie Lu, Weizhuo Li, Chongning Na, et al. Research progress of machine learning models in auto insurance fraud detection[J]. Computer Engineering and Applications, 2022, 58(05):34-49.
[12] Xie Linling. Research on the countermeasures of fraud risk of X insurance company [D]. East China Normal University, 2022. DOI: 10.27149/d.cnki.ghdsu.2022.000150.
[13] Yan Chun,Chi Xiao Ying,Liu Xin Hong. An auto insurance fraud detection model based on improved PSO-TSFNN [J]. Computer Simulation, 2022, 39(07): 168-173.

AUTHOR'S PROFILE

Jie Huang, School of Economics, Zhejiang University of Finance and Economics, Hangzhou, China.