EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Legal system-oriented telecom fraud detection, identification and prevention

Z. L. Liu*

School of Law, Hebei Finance University, Baoding, Hebei 071000, China

Abstract

INTRODUCTION: With the development of technology, telecom fraud is appearing more and more frequently and causing more and more harm.
OBJECTIVES: This paper focused on the detection, identification, and prevention of telecom fraud.
METHODS: Firstly, the telecom fraud crime was analyzed, the existing legal system was explained, and some suggestions on the protection of telecom fraud were proposed at the legal level. Then, the characteristics of telecom fraud users were analyzed to point out the differences between fraud users and normal users in terms of call, message, and traffic behavior. Finally, the Boosting algorithm was used to detect and identify telecom fraud.
RESULTS: The experiments found that the boosting algorithm had advantages in the detection and recognition of telecom fraud compared with the algorithms such as support vector machine and random forest algorithms. Among several boosting algorithms, the CatBoost algorithm performed the best, with an accuracy of 0.9465 and an F1 value of 0.9047.
CONCLUSION: The results demonstrate the reliability of the CatBoost algorithm in detecting and recognizing telecom fraud, and it can be applied in practice.

Keywords: legal system, telecom fraud, boosting algorithm, CatBoost, accuracy

Received on 12 May 2023, accepted on 15 September 2023, published on 18 September 2023

doi: 10.4108/eetsis.3335

1. Introduction

Under the influence of the rapid development of network technology, mobile technology, etc., the probability of personal private data being leaked is also increasing, so lawbreakers can easily access information and commit telecom fraud [1]. Telecom fraud refers to the act in which the perpetrator sets up a scam to obtain property through the Internet, telephone, and messages [2,3]. Compared with traditional fraud, telecom fraud has a higher targeting. Criminal methods are complex and constantly changing, leading to a higher possibility of victims. There are also many shortcomings in the existing criminal legal system for regulating telecom fraud, and there is still a possibility that telecom fraud will further intensify. The high incidence of telecom fraud causes both personal property and mental losses [5] and also seriously affects the harmony and stability of society [6]; therefore, the detection, identification and prevention of electronic fraud is particularly important. With the development of artificial intelligence, machine learning and other technologies, more and more methods have been applied in the detection and identification of telecom fraud [7]. Zamini et al. [8] designed a fraud detection method combining autoencoder and K-means clustering and found that the method had 98.9% accuracy for 284,807 transactions in European banks. Yao et al. [9] used trajectory clustering and FP-growth algorithm for mining on telecom fraud to find individuals with multiple phones and monitored them by analyzing the trajectory data. The method provided technical support to prevent fraudulent activities. Hou et al. [10] combined voice recognition techniques with Gaussian mixture

*Corresponding author. Email: izl617278@163.com

models to determine the identity of the speaker to identify telecom frauds. They found through experiments that the accuracy of voiceprint recognition reached 72%. Kashir et al. [11] used call detail records of normal and fraudulent users as input and used neural networks and support vector machines (SVM) for classification and found that Bayesian regularization had the best performance. Wu et al. [12] proposed a method that combines machine learning and area algorithms to achieve telecom fraud detection, significantly enhancing the accuracy of detection results through the integration of common algorithms. Chang et al. [13] designed a centrality-oriented deep random walk method to identify key fraudulent roles in telecom fraud networks and found that this approach outperformed other baseline methods. Nawawi et al. [14] conducted an investigation on actual fraud cases in a telecommunications company in Malaysia and found that weak compliance with internal controls provided opportunities for fraud to occur. Tseng et al. [15] developed a graph mining-based framework for detecting fraudulent calls, conducted comprehensive experiments on the dataset, and proved the effectiveness of this method. Amuji et al. [16] developed a linear classifier to distinguish between genuine users and fraudulent users, achieving a posterior probability of 0.7368 on a sample of 80 users. They found that the most important parameters were the number of calls per hour and call duration. In order to further improve the efficiency of detecting and identifying telecom fraud, as well as provide a more practical and effective method for detection and identification, this article first conducted a simple analysis of telecom fraud based on legal systems and proposed some protective suggestions from a legal perspective. Then, by analyzing the characteristics of telecom fraud users, it aims to understand the correlation between different characteristics and telecom fraud. Then, the detection and identification of telecom fraud using several Boosting algorithms were studied. The performance of AdaBoost, XGBoost, and CatBoost algorithms was compared and analyzed on real datasets, demonstrating the superiority of the CatBoost algorithm. This paper provides theoretical support for a better and faster discovery of telecom fraud users.

2. Telecom fraud detection and identification by boosting algorithm

2.1. Telecom fraud crimes

Telecom fraud refers to a new type of special fraud using modern telecommunications, network and other technologies, and is not an independent crime in the current law, but as a special form of fraud crime. Compared with traditional fraud, telecom fraud has the following characteristics.

(1) The social harm is greater. The target of telecom fraud is not specific and large in scope. Under the influence of technological development, the range of victims of telecom fraud is increasing, the economic loss is getting bigger and bigger. Even many victims make extreme choices, which seriously affects the security and stability of the society.

(2) Fraudulent techniques are more professional. There are many highly technical and intelligent means of telecom fraud, using hacking techniques, fake websites, fake documents, etc. to gain the trust of victims and transfer stolen money through remote operations, electronic payments, etc., making it much more difficult to solve the case.

(3) The infringement of legal interests is more serious. Impersonating public prosecutors and law enforcement officers in telecom fraud seriously affects the credibility of public officials, and the process of fraud also severely threatens information security and financial order.

2.2. Provisions of the existing legal system

There is no direct provision for the crime of telecom fraud in the Criminal Law of the People's Republic of China. In judicial practice, the crime of telecom fraud is generally convicted and sentenced as the crime of ordinary fraud; however, due to the characteristics of telecom fraud, there are shortcomings in the application of charges and penalties.

The Law of the People's Republic of China on the Protection of Personal Information regulates the activities of handling personal information and increases the protection of personal information, thereby combating telecom frauds that use personal information illegally.

The Anti-Telecommunication Network Fraud Law (Draft) is submitted for consideration in 2021 [17], which proves that the country is legislatively focusing on telecom fraud crimes.

The current deficiencies in the legal system bring higher requirements for the detection and identification of telecom fraud. Timely and effective telecom fraud detection and identification can avoid suffering greater losses. Therefore, this paper investigated the monitoring and identification of telecom fraud through boosting algorithms.

2.3. Analysis of the characteristics of telecom fraud users

According to the statistics of telecom operators, the following characteristics of telecom fraud users exist.

(1) The frequency of calls is high. Fraudsters lure users by casting a wide net and making a large number of calls, so they call with high frequency and are predominantly the calling party.

(2) The use of traffic is little. Fraudsters mainly make phone calls and use far less traffic than normal telecom users, and they may not even have traffic services open at all.

(3) The length of the call is long. Fraudsters will use a variety of tactics to trick their victims, so it takes a long time.

However, in addition to fraudulent users, the behavior of people who are engaged in some occupations, such as deliveryman, may also have characteristics similar to those of fraudulent users, so a more in-depth analysis of their characteristics is needed to further classify fraudulent users from normal users.

At present, telecom fraud is dominated by cell phone users, and most of them are individual users; therefore, the source of the experimental dataset used in this paper was the individual cell phone user data from a telecom company, all of which were desensitized, and the dataset contained the numbers that have been marked as being belonging to fraudsters as well as the numbers of normal users. Then, 18,567 fraudulent users and 20,124 normal users were randomly selected for the experiment. The data from January 2022 to August 2022 were selected, including the basic data, call data, messages, and Internet access data of users. Based on the characteristics of fraudulent users, the indicators shown in Table 1 were extracted from the dataset for analysis.

Table 1. Fraudulent user characteristics indicators

Indicators	Description
Duration of cell phone access	Duration from the time when the card was activated to the current time when the data are extracted
Number of cards activated	Number of cards activated by users with the same identify document
Average number of calls per day	Average number of calls per day from users
Average calling duration per day	Average time of calling per day made by users
Average number of messages sent per day	Average number of messages sent per day by users
Average daily traffic used	Average daily traffic used by users
Average monthly consumption amount	Average monthly spending of users
Is the user a fraudster?	1: Yes; 0: No

Based on the above indicators, the experimental data were analyzed to understand the connection between these indicators and whether the user was fraudulent or not. First, the analysis of the duration of cell phone access and whether the user is a fraudster or not is shown in Figure 1.

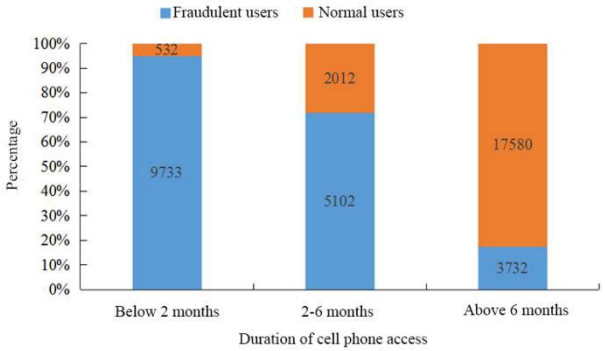

Figure 1. Duration of cell phone access and fraud analysis

From Figure 1, it was found that the percentage of fraudulent users was significantly higher than that of normal users among users who have been accessed to the network for less than two months, 9,733 fraudulent users have been accessed to the network for less than two months, and among users who have been accessed to the network for more than six months, the percentage of normal users was significantly higher than that of fraudulent users, indicating that there was a clear difference between fraudulent users and normal users in terms of duration of cell phone access.

The analysis of the number of cards activated and whether the user is a fraudster or not is shown in Figure 2.

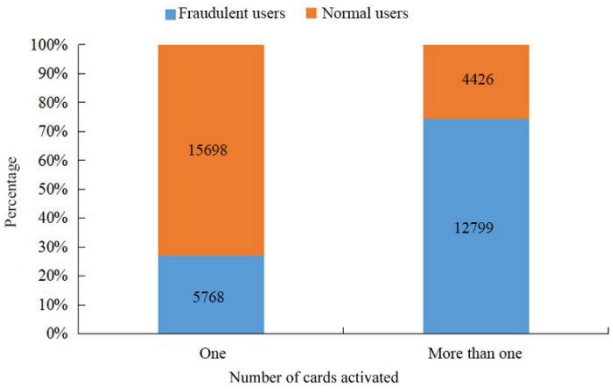

Figure 2. Number of cards activated and fraud analysis

From Figure 2, it was found that users who activated one phone card were usually normal users, with the number reaching 15,698, while among the users who activated multiple phone cards, the percentage of fraudulent users was high, over 70%, indicating that the more the number of cards activated, the higher the possibility of fraudulent users.

The analysis of users' call, message, and traffic behavior and whether the users are fraudsters or not is shown in Table 2.

Table 2. Analysis of user call, SMS and traffic behavior and fraud

Calling behavior		Fraudulent users	Normal users
Average number of calls per day	0-5 times	0 (0%)	11,023 (100%)
	6-10 times	2,880 (29.80%)	6,785 (70.20%)
	More than 10 times	15,687 (87.14%)	2,316 (12.86%)
Average calling duration per day	Under 120 s	1,017 (18.73%)	4,414 (81.27%)
	121-1,500 s	2,865 (18.46%)	12,658 (81.54%)
	More than 1,500 s	14,685 (82.79%)	3,052 (17.21%)
Average number of messages sent per day	0	17,189 (48.81%)	18,025 (51.19%)
	1-5	514 (20.88%)	1,948 (79.12%)
	More than 5	864 (85.12%)	151 (14.88%)
Average daily traffic used	0 M	14,672 (79.02%)	6,785 (31.62%)
	Under 100 M	3,785 (20.39%)	6,987 (64.86%)
	100 M or more	110 (1.70%)	6,352 (98.30%)

According to Table 2, the percentage of fraudulent users reached 87.14% among users with more than ten calls every day, while the percentage of normal users was only 12.86%, indicating that the higher the average number of calls per day, the higher the possibility of fraudulent users. Among users with an average daily call duration of 1,500 s or more, the percentage of fraudulent users reached 82.79%, indicating that there was a connection between the average daily call duration and fraudulent behavior. Among users who sent more than five messages per day, the percentage of fraudulent users was significantly higher than that of normal users, while among users who used more than 100 M of traffic every day, the percentage of fraudulent users was extremely small, only 1.70%. Overall, users' call, message, and traffic behavior were all related to fraudulent behavior.

The analysis of the average monthly consumption amount and whether the user is a fraudster is shown in Figure 3.

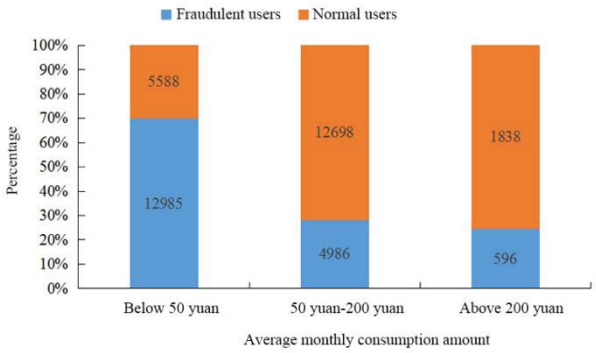

Figure 3. Average monthly consumption amount and fraud analysis

From Figure 3, it was found that among users with average monthly consumption below 50 yuan, the percentage of fraudulent users was relatively high, accounting for about 70% of the total, while among users with average monthly consumption above 200 yuan, the number of normal users was obviously higher than the number of fraudulent users, and the fraudulent users accounted for about 20%, indicating that there was a difference between fraudulent users and normal users in terms of average monthly consumption amount.

2.4. Boosting algorithm

The boosting algorithms are a class of integrated learning algorithms [18], which combine multiple models with low accuracy to improve the accuracy of classification. The detection and identification of telecom fraud can be considered as a problem of classifying fraudulent users from normal users; therefore, this paper analyzed the performance of several boosting algorithms.

The Adaboost algorithm trains different classifiers with the same training set [19]. it is assumed that an input sample is $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, $y_i \in (-1,1)$. The sample weight is initialized: $\begin{cases} w_{1,i} = \frac{1}{2m}, y_i = 1 \\ w_{1,i} = \frac{1}{2v}, y_i = -1 \end{cases}$, where m and v are the number of positive and negative samples. Then, the weak classifier is trained to minimize error rate ε_t:

$$h(x, f, p, \theta) = \begin{cases} 1, pf(x) < \theta \\ -1, else \end{cases}, \quad (1)$$

$$\varepsilon_t = \sum_{i=1}^{k} w_n^{(t)} I(h_t(x_n) \neq y_n), \quad (2)$$

where $I(h_t(x_n) \neq y_n)$ is the indicator function, which takes the value of 1 when $y_i \neq f(x_i)$ and 0 in the other cases.

The weighted value for every weak classifier is calculated: $\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$. The weight of the training set is updated according to $w_{t+1,i} = \left[w_{t,i}exp\left(-\alpha_t y_i G_t(x_i)\right)\right]/z_t$ to obtain the final classifier $F(X) = sign(\sum_{i=1}^{T}\alpha_t h_t(x))$.

The XGBoost algorithm [20] uses multiple weak classifiers to fit the residuals and then accumulates the results to obtain the final predicted value. It is expressed by the following equation:

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i), \tag{3}$$

where $f_k(x_i)$ is the weak classifier, k is the number of iterations, and $\hat{y}_i^{(t)}$ denotes the predicted value. The objective function of the XGBoost algorithm consists of two components, the loss function and the regular term. It is expressed by the following equation:

$$obj = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda} + \gamma T, \tag{4}$$

where G_j is the sum of the first-order partial derivatives of all samples in leaf node j: $G_j = \sum_{i\in I_j} g_i$, H_j is the sum of the second-order partial derivatives, $H_j = \sum_{i\in I_j} h_i$, λ and γ are penalty factors, and T is the number of leaf nodes in the tree.

Similar to the XGBoost algorithm, the CatBoost algorithm is an improvement of the gradient boosted decision tree (GBDT) [21]. For dataset $\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$, where $x_i = (x_{i1},x_{i2},\cdots,x_{im})$, the equation of the CatBoost algorithm for the category feature processing is:

$$x_{ik} = \frac{\sum_{j=1}^{n}[x_{jk}=x_{ik}]\cdot Y_j}{\sum_{j=1}^{n}[x_{jk}=x_{ik}]}. \tag{5}$$

Then, to avoid overfitting, the data set is randomly permute. It is assumed that $\sigma = (\sigma_1,\sigma_2,\cdots,\sigma_n)$ is a randomly permuted sequence. The sample value of the category feature is:

$$x_{\sigma_p,k} = \frac{\sum_{j=1}^{P-1}\left[x_{\sigma_j k}=x_{\sigma_p k}\right]\cdot Y_{\sigma_j}+\alpha\cdot P}{\sum_{j=1}^{P-1}\left[x_{\sigma_j k}=x_{\sigma_p k}\right]+\alpha}, \tag{6}$$

where P is the a priori value and α is the weight of the a priori value.

During the computation, the CatBoost algorithm generates a symmetric tree of trees, and before each round of generating the tree, a randomly permuted sequence is reselected to process the category features, thus mining richer information; therefore, it has good operational efficiency and classification accuracy.

3. Results and analysis

3.1. Algorithm parameter setting

The experiment was conducted in the MATLAB environment. The algorithm codes were implemented using Python. Several boosting algorithms have many parameters that need to be set. In order to obtain better results in the detection and recognition of telecom fraud, this paper used the particle swarm optimization (PSO) algorithm [22] to optimize the parameters of the boosting algorithm. The population size of the PSO algorithm was set as 30, the maximum number of iterations was 30, the inertia weight was set as 0.8, and the acceleration factor was set as 1.49. The parameter search ranges of several boosting algorithms and the final parameters obtained after PSO are shown in Tables 3-5.

Table 3. The parameter settings of the AdaBoost algorithm

	Parameter search range	Optimal parameter
learning_rate	(0.01,0.3)	0.05
max_leaf_nodes	(5,30)	8
n_etimators	(50,250)	105
max-features	(0.1,0.999)	0.42
max-depth	(5,15)	13

Table 4. The parameter settings of the XGBoost algorithm

	Parameter search range	Optimal parameter
learning_rate	(0.01,0.3)	0.16
max_child_weight	(0,20)	18
n_etimators	(10,200)	168
subsample	(0.4,1)	0.65
max-depth	(5,15)	14
colsample_bytree	(0.01,1)	0.75
gamma	(0.001,10)	5.35
reg_alpha	(0,5)	2.33
reg_lambada	(0,5)	4.83

Table 5. The parameter settings of the CatBoost algorithm

	Parameter search range	Optimal parameter
learning_rate	(0.11,1)	0.41
iterations	(10,1000)	817
depth	(5,15)	7

3.2. Algorithm evaluation indicators

The effectiveness of the algorithm on telecom fraud detection and identification was evaluated based on the confusion matrix (Table 6).

Table 6. Confusion matrix

		True value	
		Positive category	Negative category
Predicted value	Positive category	TP	FP
	Negative category	TN	FN

In Table 6, TP represents the number of fraudulent users correctly predicted as fraudulent; FP represents the number of normal users incorrectly predicted as fraudulent; TN represents the number of fraudulent users incorrectly predicted as normal; FN represents the number of normal users correctly predicted as normal.

(1) Accuracy: the proportion of correctly classified samples to the total number of samples, $Accuracy = (TP + TN)/(TP + FN + FP + TN)$

(2) Precision: the proportion of true positive samples among the samples predicted as positive, $Precision = TP/(TP + FP)$

(3) Recall rate: the proportion of samples predicted as positive among all true positive samples, $Recall = TP/(TP + FN)$

(4) F1 value: harmonic mean of precision and recall rate, $F1 = (2 \times Precision \times Recall)/(Precision + Recall)$

(5) Area under the curve (AUC) value: The area under the receiver operator characteristic (ROC) curve plotted by taking false positive rate ($\mathrm{FPR} = \mathrm{FP}/(\mathrm{FP} + \mathrm{TN})$) as the horizontal coordinate and true positive rate ($\mathrm{TPR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$) as the vertical coordinate, which can effectively measure the classification performance of algorithms in situations where there is an imbalance between positive and negative samples.

3.3. Analysis of results

The indicators in Table 1 were used as inputs to the algorithm for the detection and identification of telecom fraud. To demonstrate the effectiveness of the boosting algorithm, it was compared with several other algorithms:

(1) Support vector machine (SVM) [23],

(2) Random Forest (RF) [24],

(3) Classification and Regression Tree (CART) [25],

(4) the call detail records (CDR)-based classifier proposed in literature [26].

These algorithms were cross-validated by 5-fold, and the results were averaged. First, the algorithms were compared in terms of accuracy, as shown in Figure 4.

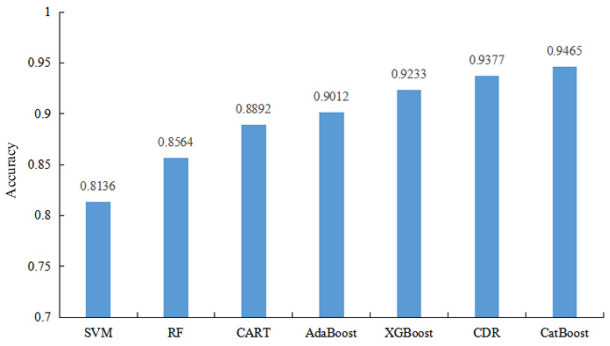

Figure 4. Comparison of the accuracy between different algorithms for telecom fraud detection and identification

From Figure 4, it was found that the SVM algorithm had the lowest accuracy in telecom fraud detection and identification, only 0.8136, RF was 0.8564, CART had an accuracy of 0.8892, slightly higher than the SVM algorithm but still below 0.9, while the boosting algorithms had accuracies above 0.9 in fraud detection and identification. The accuracy of the AdaBoost algorithm was 0.9012, the XGBoost algorithm was 0.9233, and the CatBoost algorithm had the highest accuracy, 0.9465, which was 0.0453 higher than the AdaBoost algorithm and 0.0232 higher than the XGBoost algorithm. The accuracy of the CDR-based classifier was 0.9377, which was 0.0088 higher than that of the CatBoost algorithm. This suggested the effectiveness of the CatBoost algorithm in the detection and identification of telecom fraud.

The comparison of the algorithms in terms of precision and recall rate is shown in Figure 7.

Table 7. Comparison of the precision and recall rate between different algorithms

	Precision	Recall rate
SVM	0.7542	0.7963
RF	0.7729	0.8124
CART	0.8133	0.8256
AdaBoost	0.8326	0.8567
XGBoost	0.8578	0.8749
CDR	0.8721	0.9042
CatBoost	0.8869	0.9233

From Table 7, it was found that both SVM and RF algorithms performed poorly in terms of precision and recall rate, and the CART algorithm had an accuracy of

0.8133 and a recall rate of 0.8256, which were slightly higher than SVM and RF algorithms. The accuracy and recall rates of Boosting algorithms were significantly better than those of SVM and the other algorithms, where the CatBoost algorithm had the highest values in both indicators, 0.8869 (0.0543 higher than the AdaBoost algorithm, 0.0291 higher than the XGBoost algorithm, and 0.0148 higher than the CDR-based classifier) and 0.9233 (0.0666 higher than the AdaBoost algorithm, 0.0484 higher than the XGBoost algorithm, and 0.0191 higher than the CDR-based classifier). These results proved the precision of the CatBoost algorithm in telecom fraud detection and recognition.

The comparison results of the F1 value are shown in Figure 5.

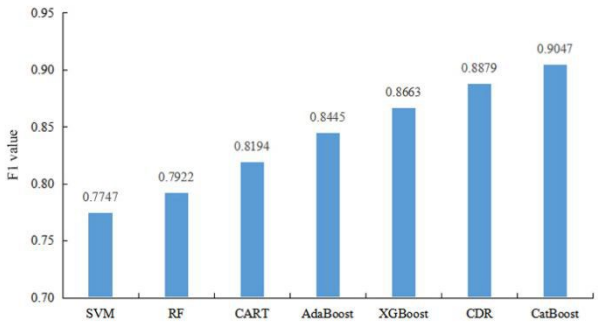

Figure 5. Comparison results of the F1 value between different algorithms for telecom fraud detection and identification

From Figure 5, it was found that the F1 values of SVM and RF algorithms were below 0.8; the CART algorithm had an F1 value of 0.8194, the AdaBoost algorithm was 0.8445, the XGBoost algorithm was 0.8663, and the CDR-based classifier was 0.8879, all below 0.9. Only the CatBoost algorithm had an F1 value of 0.9047, which was 0.0602 higher than the AdaBoost algorithm, 0.0384 higher than the XGBoost algorithm, and 0.0168 higher than the CDR-based classifier. Overall, the CatBoost algorithm performed the best in the detection and identification of telecom frauds, and it could accurately identify fraudulent users according to the call, message, and traffic characteristics of different users. Therefore, this algorithm has good application values in reality.

The comparison results of the AUC value are shown in Table 8.

Table 8. Comparison of the AUC value between different algorithms for the detection and identification of telecom fraud

	AUC value
SVM	0.8536
RF	0.8775
CART	0.9123
AdaBoost	0.9327
XGBoost	0.9521
CDR	0.9715
CatBoost	0.9826

From Table 8, it can be observed that SVM and RF algorithms had relatively low AUC values, both below 0.9. CART and AdaBoost algorithms also had AUC values that did not exceed 0.95, indicating poor performance in detecting telecom fraud. The XGBoost algorithm had an AUC value of 0.9521. The CDR-based classifier had a value of 0.9715. The CatBoost algorithm had an AUC value of 0.9826, showing a significant improvement of 0.0305 compared to the XGBoost algorithm and an increase of 0.0111 compared to the CDR-based classifier. These results further confirmed the superior performance of the CatBoost algorithm.

According to the CatBoost algorithm, the top five features ranked in terms of their importance are shown in Table 9.

Table 9. Feature importance

Feature	Importance
Average number of calls per day	27.18
Number of cards activated	25.33
Average daily traffic used	17.62
Average number of messages sent per day	8.98
Average calling duration per day	5.33

According to Table 9, it can be observed that the average number of calls per day had the greatest impact on evaluating telecom fraud. The more frequent the daily calls were, the higher the likelihood of being a fraudulent user. Additionally, the number of cards activated also held significant importance, indicating that there was a greater possibility of fraud when there were more new card activations. This finding aligned with the previous analysis. In conclusion, the CatBoost algorithm not only effectively distinguished between normal and fraudulent users but also identified important features, showing good performance.

4. Discussion

Compared with traditional fraud, telecom fraud differs in the way and object of fraud. Telecom fraud adopts a non-contact approach, and its fraud process relies on telecom technology. In terms of fraud objects, it is aimed at unspecified people. With the development of technology, the forms of telecom fraud are constantly renovated, and the impact on society is also increasing. The current legal system has many shortcomings, for example, the determination of the number of crimes and accomplices are still unclear; therefore, in order to further meet the current judicial needs, strengthen the fight against telecom fraud, and achieve legal protection, this paper put forward some suggestions from the following perspectives.

(1) The crime of telecom network fraud should be established separately. The current legislation cannot meet the needs of telecom fraud crimes and cannot really achieve the severe punishment of telecom fraud crimes. Including telecom fraud crimes in the criminal law will have a positive effect on both crime fighting and prevention.

(2) The existing criminal law should be perfected. At present, the crime of fraud mainly considers the amount of the crime and the circumstances of the crime; however, in telecom fraud, these two elements are difficult to measure, and there are also shortcomings in the subject of the crime and the sentence. For example, the sentence for ordinary fraud is less than three years in prison, which is not able to bring deterrence to criminals, so the sentence needs to be adjusted. In terms of the subject of the crime, telecom fraud usually involves people under 16 years old. Therefore, the existing criminal law should be improved.

(3) Financial institution prevention and control should be strengthened. The transfer link in telecom fraud generally relies on financial institutions, so the control of financial accounts needs to be strengthened. There is a need to increase the management of financial accounts, improve the transfer system, and implement the real name system for activated bank card accounts.

5. Conclusion

This paper analyzed the detection identification and prevention of telecom fraud for the legal system and compared the performance of several boosting algorithms for detection and identification based on the characteristics of fraudulent users and normal users. The experiments found that the performance of the boosting algorithm was significantly better than SVM and the other methods, and among the boosting algorithms, the accuracy and F1 value of the CatBoost algorithm were higher, which proved the reliability of the method for telecom fraud detection and recognition. At present, there are still many shortcomings in the prevention of telecom fraud at the legal level, and the telecom fraud detection and identification method designed in this paper provides a reference for further strengthening the prevention of telecom fraud.

References

[1] Ali M A, Azad M A, Parreno-Centeno M, Hao F, van Moorsel A. Consumer-facing technology fraud: Economics, attack methods and potential solutions. Future Gener. Comp. Sy., 2019; 100:408-427.

[2] Gong H. The Dilemma of Telecommunication Fraud Crime——An Analysis of China's Governance Model as a Sample. SHS Web Conf., 2022; 148:1-5.

[3] Shut O A. Fraud in Social Networks and Ways to Implement. Herald Omsk Univ. Ser. Law, 2020; 17(4):97-106.

[4] Starostenko OA. Nature and methods of committing fraud using information-telecommunication technologies. Bull. Udmurt Univ. Ser. Econ. Law, 2020; 30(4):576-582.

[5] Mawgoud AA, Ali I. Statistical Insights and Fraud Techniques for Telecommunications Sector in Egypt. 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), 2020; 143-150.

[6] Chen G, Ding L, Chen G, Qin P. Reliable Security Strategy for Message-Oriented Middleware. Int. J. Digit. Crime Fo., 2018; 10(1):12-23.

[7] Zhong R, Zhang Z, Lin R, Zou H. Encoding Broad Learning System : An Effective Shallow Model For Anti-fraud. 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020; 5496-5504.

[8] Zamini M, Montazer G. Credit Card Fraud Detection using autoencoder based clustering. 2018 9th International Symposium on Telecommunications (IST), 2018; 486-491.

[9] Yao R, Wang F, Chen S, Zhao S. Assisting Telecommunication Fraud Prediction: Detect Individuals Carrying Multiple Phones Based on Trajectory Data Mining. 2020 Information Communication Technologies Conference (ICTC), 2020; 158-165.

[10] Hou D, Han H, Novak E. TAES: Two-factor Authentication with End-to-End Security against VoIP Phishing. 2020 IEEE/ACM Symposium on Edge Computing (SEC), 2020; 340-345.

[11] Kashir M, Bashir S. Machine Learning Techniques for SIM Box Fraud Detection. 2019 International Conference on Communication Technologies (ComTech), 2019; 4-8.

[12] Wu B, Li M, Zhou C. Application of adaboost algorithm and immune algorithm in telecommunication fraud detection. 2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018), 2018; 159-163.

[13] Chang YC, Lai KT, Chou SCT, Chiang WC, Lin YC. Who is the boss? Identifying key roles in telecom fraud network via centrality-guided deep random walk. Data Technol. Appl., 2021; 55(1):1-18.

[14] Nawawi A, Salin ASAP. Employee fraud and misconduct: empirical evidence from a telecommunication company", Inf. Comput. Secur., 2018; 26(1):129-144.

[15] Tseng VS, Ying JC, Huang CW, Kao Y, Chen K. FrauDetector: A Graph-Mining-based Framework for Fraudulent Phone Call Detection. KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015; 2157-2166.

[16] Amuji HO, Chukwuemeka E, Ogbuagu EM. Optimal Classifier for Fraud Detection in Telecommunication Industry. Open J. Optim., 2019; 08(1):15-31.

[17] Shan Y. The Transformation of Digital Society to Front-end Preventive Crime Governance—A Case Study of A Draft Law on Anti-Telecom and Online Fraud. J. Shanghai Norm. Univ. (Philos. Soc. Sci.), 2022; 51(3):58-66.

[18] Osman A H, Aljahdali H. An Effective of Ensemble Boosting Learning Method for Breast Cancer Virtual Screening Using Neural Network Model. IEEE Access, 2020; 8: 39165-39174.
[19] Rajesh K, Dhuli R. Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier. Biomed. Signal Proces., 2018; 41(mar.):242-254.
[20] Le NQK, Do D T, Nguyen TTD, Le QA. A sequence-based prediction of Kruppel-like factors proteins using XGBoost and optimized features. Gene, 2021; 787(4):145643.
[21] Coronado-Blázquez J. Classification of Fermi-LAT unidentified gamma-ray sources using catboost gradient boosting decision trees. Mon. Not. R. Astron. Soc., 2022; 515(2): 1807-1814.
[22] Yu H, Zheng M, Zhang W, Nie W, Bian T. Optimal design of helical flute of irregular tooth end milling cutter based on particle swarm optimization algorithm. P. I. Mech. Eng. C-J. Mec., 2022; 236(7):3323-3339.
[23] Marston Z, Cira T M, Knight J F, Mulla D, Alves TM, Erin W Hodgson, Arthur V Ribeiro, Ian V MacRae, Robert L Koch. Linear Support Vector Machine Classification of Plant Stress From Soybean Aphid (Hemiptera: Aphididae) Using Hyperspectral Reflectance. J. Econ. Entomol., 2022; 115(5):1557-1563.
[24] Noi P T, Kappas M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors, 2018; 18(1):1-20.
[25] Ghiasi M M, Zendehboudi S, Mohsenipour A A. Decision Tree-Based Diagnosis of Coronary Artery Disease: CART Model. Comput. Meth. Prog. Bio., 2020; 192(6):105400.
[26] Xing J, Yu M, Wang S, Zhang Y, Ding Y. Automated Fraudulent Phone Call Recognition through Deep Learning. Wirel. Commun. Mob. Com., 2020; 2020(2):1-9.