Computers, Materials & Continua CMC, vol.61, no.1, pp.185-195, 2019

Credit Card Fraud Detection Based on Machine Learning

Yong Fang[1], Yunyun Zhang[2] and Cheng Huang[1, *]

Abstract: In recent years, the rapid development of e-commerce exposes great vulnerabilities in online transactions for fraudsters to exploit. Credit card transactions take a salient role in nowadays' online transactions for its obvious advantages including discounts and earning credit card points. So credit card fraudulence has become a target of concern. In order to deal with the situation, credit card fraud detection based on machine learning is been studied recently. Yet, it is difficult to detect fraudulent transactions due to data imbalance (normal and fraudulent transactions), for which Smote algorithm is proposed in order to resolve data imbalance. The assessment of Light Gradient Boosting Machine model which proposed in the paper depends much on datasets collected from clients' daily transactions. Besides, to prove the new model's superiority in detecting credit card fraudulence, Light Gradient Boosting Machine model is compared with Random Forest and Gradient Boosting Machine algorithm in the experiment. The results indicate that Light Gradient Boosting Machine model has a good performance. The experiment in credit card fraud detection based on Light Gradient Boosting Machine model achieved a total recall rate of 99% in real dataset and fast feedback, which proves the new model's efficiency in detecting credit card fraudulence.

Keywords: Credit card fraud detection, imbalanced data, LightGBM model, smote algorithm.

1 Introduction

E-commerce has flourished in the world for last decades. As a credit voucher and a payment tool, the credit card has the superiority of convenience. Cardholders do not need to pay cash when purchasing goods. People are getting used to purchasing goods and services by using credit cards. Credit card fraud has ensued. Credit card fraud is a form of financial fraud [Sahin, Bulkan and Duman (2013); Adewumi and Akinyelu (2017)], referring to the fraudulent activities conducted by violating the regulations of credit card management and using a credit card for the purpose of illegal possession. Usually, credit card fraud is a behavior in which a fraudster uses the cardholder's credit card to defraud money and property [Bhatla, Prabhu and Dua (2003)]. With the increase of e-commerce transaction volume, credit card fraud is becoming more and more rampant [Kumari and Mishra (2019); Dhankhad, Mohammed and Far (2018); Patil, Nemade and Soni (2018)]. The lag of merchant risk management is becoming one of the main reasons for the expansion of the

[1] College of Cybersecurity, Sichuan University, Chengdu, 610065, China.

[2] College of Electronics and Information Engineering, Sichuan University, Chengdu, 610065, China.

[*] Corresponding Author: Cheng Huang. Email: opcodesec@gmail.com.

credit card fraud phenomenon. In order to reduce the losses caused by merchant risks and improve the overall management level of credit card merchants, merchant risk management has become an important part of the private financial business.

To commit credit card fraud, fraudsters are racking their brains to get sensitive information such as credit card number [Zareapoor and Yang (2017)], email address, phone number [Óskarsdóttir, Bravo, Sarraute et al. (2019)] and so on. Credit card holders should pay great attention to protecting their personal information when making online payments. Credit card holders develop good habits of not transferring bank cards and passwords in chat software or mobile phone text messages. In the past, there were some ways to deal with credit card fraud. The neural network method was actually used in a credit card fraud system at a local bank and improved the level of detection [Ghosh and Reilly (1994)]. The combination of Bayesian belief networks and artificial neural networks can significantly improve the experimental results [Maes, Tuyls, Vanschoenwinkel et al. (2002)]. Hidden Markov model [Govind and Hazari (2014)] works well in helping banks improve their security gateways. A fraud model is built by digging the relationship between the cardholder's consumption and personal habits [Quah and Sriganesh (2008)]. With the improvement of anti-fraud means, fraudsters are constantly changing their fraudulent means. For this reason, the method of fraud detection must be improved accordingly.

LightGBM model is used to detect credit card fraud in this paper. There is a common basic assumption that the training samples of different categories are equal in number. If the difference is very large, it will cause trouble in the learning process. The real-life credit card transaction dataset is not in accord with this common basic assumption. To tackle the imbalance of the data, two solutions are proposed in the data analysis. They are the next sampling and oversampling. The next sampling is relatively simple to implement. That's to say, the amount of normal data is reduced to the same as fraudulent data. We want the experiment results to be better, then modified oversampling called Smote algorithm is used. Confidentiality is another challenge when analyzing credit card transaction dataset. Some features could not be obtained. The credit card fraud has long been appraised as a serious problem in the financial industry, in the sense that the bank credit risk is magnified and the financial order is seriously disturbed. In this paper, we construct a credit card fraud detection model based on Light Gradient Boosting Machine algorithm, to determine new transactions as fraudulent or legitimate transactions. The dataset that is used in this paper is publicly available. Our major contributions are as follows:

- In a large number of credit card transactions, LightGBM model improves training efficiency, which achieves a recall rate of 99% in real dataset.
- Some classification values which are not the numeric type are identified by one-hot encoding, thus strengthening the training effect of the model.
- To tackle the imbalance of the data, Smote algorithm is used. This can balance fraudulent data with normal data effectively.

2 Related work

In order to detect fraud in credit card transactions, lots of methods have been proposed.

AC Bahnsen et al. [Bahnsen, Aouada, Stojanovic et al. (2016)] proposed a new set of features which analyzed the consumer spending behavior. The features are useful in credit card fraud detection. However, this model may take too long to classify a new transaction by calculating the features.

N Mahmoudi et al. [Mahmoudi and Duman (2015)] proposed a method named Linear Fisher Discriminant Analysis which belonged to a supervised learning method. The method used the category label of the data to reduce the dimension issues. As the main contribution of this study, the weighted average is used in calculating the within-class variance. J Jurgovsky et al. [Jurgovsky, Granitzer, Ziegler et al. (2018)] chose Long Short-Term Memory (LSTM) networks to accomplish feature processing and model calculation as quickly as possible. The dataset is split into two parts: e-commerce transactions (ECOM) and face-to-face transactions (F2F). In order to make the experimental results more objective, LSTM and Random Forest are compared. P Kulkarni et al. [Kulkarni and Ade (2016)] proposed an improved algorithmic system to solve the imbalance of the data. Only in this way can it be adapted to the real credit card fraud environment. C Liu et al. [Liu, Chan, Alam Kazmi et al. (2015)] introduced Random Forest (RF) for predicting credit card fraud. The detection efficiency is improved significantly by using Random Forest (RF). RF can handle high-dimensional data without making feature selection. F Carcillo et al. [Carcillo, Dal Pozzolo, Le Borgne et al. (2018)] also chose RF to train their model because of the advantages of RF.

G. Rushin et al. [Rushin, Stancil, Sun et al. (2017)] compared the three supervised classification models: logistic regression, Gradient Boosted Machine (GBM), and deep learning in detecting fraud. And two methods were used in feature processing. However, this process takes a long time because feature selection is not applied. M Óskarsdóttir et al. [Óskarsdóttir, Bravo, Sarraute et al. (2019)] used phone data to build the network. This approach can be a very intuitive presentation of the relationship, thus helping them identify fraud. VV Vlasselaera et al. [Vlasselaera, Bravo, Caelen et al. (2015)] proposed a probabilistic model to express the correlation of variables by graphs. The relevant variables are mainly merchants and credit card holders. The time spent on the transaction is also taken into account. However, the result is highly imbalanced.

Although there are many proposed methods for detecting credit card fraud, there still have some shortages, we use Smote algorithm to process the unbalanced data and choose LightGBM algorithm to build our model, which improves the training efficiency.

3 Proposed approach

The goal of credit card fraud detection based on machine learning is to judge whether a credit card transaction is legal or fraudulent accurately and quickly. In this section, we analyze how to preprocess the input data and select Light Gradient Boosting Machine algorithm to establish LightGBM model.

3.1 Overview

Fig. 1 shows the proposed method of classifying fraud from normal. Firstly, the database is collected from kaggle. com website. We preprocess the input data, which includes feature category, one-hot encoding, and Smote algorithm. In order to test the model later, the data

is divided into two pieces by us. We put training data into Light Gradient Boosting Machine algorithm to train a LightGBM model. The test data is used to evaluate the test errors of the LightGBM model.

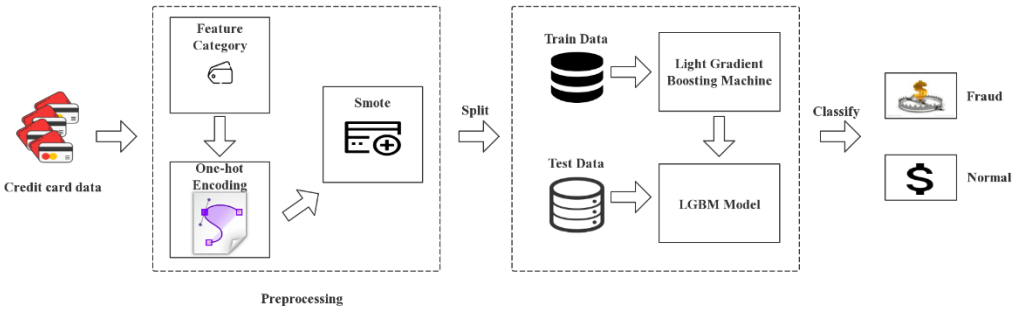


Figure 1: Flow graph of the proposed approach

3.2 Feature processing

3.2.1 Feature groups

To better detect fraud in the credit card dataset, some features are considered. They are as follows:

Age of customer: age can reflect the user's cognitive ability to the network. Customers over the age of 50 seldom have credit cards unless they have large fixed assets or a large number of bank statements. We choose the 0-7 number to present the distribution of age.

The zip code: the zip code reflects the city where users trade. Through the zip code, we can learn which city is more likely to have a credit card fraud, thus strengthening credit card anti-fraud. We selected the most frequent zipcode values by a number of transactions and amounts.

Amount: the amount of credit card transactions is very important to us. In some instances, if a person spends a fixed amount of money with the credit card every month, but suddenly the amount of money has multiplied, it is clear that a suspicious order is available.

Category of consumption: through this feature, we can understand the consumption habits of credit card customers. There are sixteen merchant categories in the data. It is important to identify if the payments made in a restaurant or in other places. For example, nowadays many people purchase cars by using credit cards for early consumption.

Job category of customer: there are some stable professions, such as civil servants, teachers and doctors. In general, these professionals are less likely to engage in credit card fraud.

The telephone number of customer: a credit card is usually bound to a phone number. If a phone number receives credit card spending text messages more than ten times within 24 hours, it is clear that this is likely a credit card fraud transaction.

3.2.2 One-hot encoding

According to the data we obtained, the values of characteristics are not always continuous, but moreover categorical. For example, the type of customer transactions is not continuous. And, the input format for classifying a model needs to be sequential. One-hot encoding is

introduced to be preprocessed before constructing our model. One-hot encoding, also known as One-bit valid encoding, uses N states encoded by an N-bit status register. Each state has its own register bit, and only one bit is valid at any time. Its values are only 0 and 1. Different types are stored in vertical space. To a certain extent, one-hot encoding also plays a role in expanding features. So we use this method to handle certain features.

3.2.3 Smote

In the data set of a credit card, it can be found that the entire data set is extremely unbalanced. There are two common ways to reconcile the imbalance in data analysis. They are undersampling and oversampling. Undersampling is relatively easy to implement. Undersampling removes some samples of the majority class so that the majority class samples are as many as that of the minority. Oversampling adds the minority class by simply copying the sample. The full name of Smote is synthetic minority oversampling technique. It is an improved scheme based on oversampling [Chawla, Bowyer, Hall et al. (2002)]. In the paper, we decide to use the Smote algorithm. The basic idea of Smote algorithm is analyzing the minority class samples and synthesizing new samples into datasets, based on the minority class samples. The algorithm flow is in four steps.

- Step 1. Based on Euclidean distance, the distance from each sample x_i in the minority class to all samples in the minority class samples is calculated. And k minority classes' nearest neighbors are obtained.
- Step 2. According to the proportion of the unbalanced sample, a sampling ratio N is determined. For each sample x_i in the fraud class, we select several samples from their k nearest neighbors randomly, assuming that the nearest neighbor was $\tilde{x}_i$.
- Step 3. For each selected neighbor $\tilde{x}_i$, a new sample is constructed by means of the following formula.

$$x_{new} = x_i + \text{rand}(0,1) \times (\tilde{x}_i - x_i) \quad (1)$$

- Step 4. Repeat Step 3 for N times, and we can synthesize N new samples.

3.3 Light gradient boosting machine algorithm

Random Forest (RF) is one of the most commonly used algorithms for classification and other tasks due to its ideal experimental results in most cases. However, the main limitation of the random forest is that the use of a large number of trees makes the speed slow. So the random forest can't predict in real time. Light Gradient Boosting Machine (LightGBM) is based on the traditional Gradient Boosted Decision Tree (GBDT) algorithm, which can accelerate the training speed of GBDT model without damaging its accuracy. In order to achieve this effect, LightGBM uses Histogram algorithm and leaf-wise growth strategy. The idea of Histogram algorithm is to turn the continuous floating point eigenvalues into discrete values (k) and construct a histogram with width k. The training data is then traversed and the cumulative statistic of each discrete value in the histogram is counted. When making a feature selection, we only need to traverse the discrete values of the histogram to find the optimal segmentation point. LightGBM eliminates the level-wise decision tree growth strategy used by most GBDT today, using a leaf-wise strategy with depth limitations. In fact, level-wise is an inefficient algorithm because it treats the leaves

of the same layer indiscriminately, bringing a lot of unnecessary overhead. Leaf-wise strategy finds the leaf with the highest split gain from all the current leaves each time, then splits, and so on. Therefore, compared with level-wise, the leaf-wise can reduce more errors and obtain better accuracy in the case of the same splitting times. Taking the over-fitting caused by leaf-wise into consideration, LightGBM adds a factor-the maximum depth limit into the model. Consequently, over-fitting can be prevented; meanwhile, a high efficiency can be guaranteed as well. LightGBM optimizes the supports for category features, which could be entered directly without additional 0/1 expansion.

4 Experiments and evaluations

In order to resolve data imbalance and have a fast feedback, a novel framework based on Light Gradient Boosting Machine algorithm is proposed. In this section, the process of the proposed approach is detailed. A five-fold cross-validation is used. In simple terms, five-fold cross-validation reduces variance by averaging the results of five different training groups. The results of the cross-validation of the proposed model are then collected, including the average score of AUC and the training time. Also, the results are compared with Random Forest (RF) and Gradient Boosting Machine algorithm (GBM).

4.1 Dataset

The dataset collected from kaggle.com website includes over 410,000 records of credit card transactions. It contains transactions that occurred between November 2012 and April 2013. And transactions were restricted to taking place in Madrid and Barcelona. We select 178,393 records of credit card transactions as the test data. The rest of the records are the training data. The last feature named fraud is our training basis. What we need to do now is to detect anomalies in the data. Observing the distribution of the data is an important factor. The frequency of fraud only accounts for 1.2% of the transaction frequency. This shows that the credit card datasets are quite unbalanced. A fraud label of 1 means a fraudulent transaction. A fraud label of 0 means a legitimate transaction. The basic information provided by the queries is mainly statistical information about payments such as number, average, minimum and maximum values. As is shown in Fig. 2, a fraudulent transaction usually occurs when small amounts are traded. The smaller the amount is, the higher the number of fraudulent transactions is. Meanwhile, fraud often occurs between the ages of 26 and 35. By using one-hot encoding, the number of features is expanded from 6 to 21. In the process of data preprocessing with the Smote algorithm, parameter values are as follows: k=5, N=100. The previous research showed that the Recall rate(R) of fraud detection was the best when the ratio of legitimate and fraudulent transactions was 1:1 [Xuan, Liu, Li et al. (2018)]. After sampling the unbalanced data with Smote algorithm, the length of oversampled data is 822,522, the length of the legitimate transaction is 411,261, and the length of the fraudulent transaction is 411,261. The ratio of legitimate and fraudulent transactions has reached 1:1.

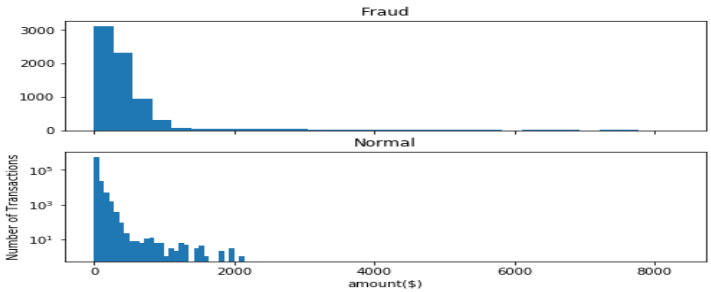


Figure 2: The relation between fraud and amount

4.2 Experimental steps

The dataset is running on a virtual machine which is based on the Ubuntu operating system. The version number of Ubuntu is 5.4.0-6ubuntu1~16.04.9. The code is written to follow the rules of Python 2.

GBM, Neural network [Adewumi and Akinyelu (2017)], Support vector machines [Bhatla, Prabhu and Dua (2003); Sadiq, Faris, Ala'M et al. (2019)], Random Forest, discriminant analysis and social network analysis are used in credit card fraud detection. A majority of these techniques are part of machine learning methods. Their proposed algorithms are compared with practical binary classification. For the binary classification problem, according to the combination of its real category and the learning tool prediction category, the samples can be divided into four cases: true positive, false positive, true negative, false negative. Tab.1 shows the classification confusion matrix.

Table 1: Classification confusion matrix

	Actual positive	Actual negative
Predicted positive	True positive (TP)	False positive (FP)
Predicted negative	False negative (FN)	True negative (TN)

From Tab. 1, we can get several metrics:

- True Positive (TP): A normal transaction is considered normal.
- False Positive (FP): A fraudulent transaction is considered normal.
- True Negative (TN): A fraudulent transaction is considered fraudulent.
- False Negative (FN): A normal transaction is considered fraudulent.

In particular：

Precision rate (P):

$$P = TP/(TP + FP) \quad (2)$$

Recall rate (R):

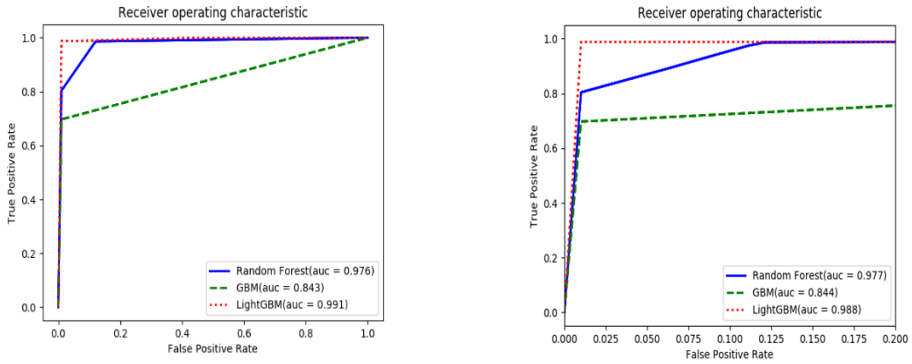
$$R = TP/(TP + FN) \quad (3)$$

We are trying to find out the credit card frauds as many as possible, so the measure we take in this paper is the Recall rate(R).

ROC (receiver operating characteristic curve) can reflect the same sensitivity at each point on the curve [Kannan and Vasanthi (2019)]. They are results concluded from several different criteria but responses to the same signal stimulus. ROC is a graph in which the false positive rate is the horizontal axis and the true positive rate is the vertical axis. And the results obtained by the subject under different stimuli conditions are so different because of the use of different judgment criteria. In this section, we use three models to train data and plot the ROC curve of each model to the same coordinate so that we can intuitively identify the advantages and disadvantages.

4.3 Evaluation and result

In the experiment, the effectiveness of LightGBM model is compared with RF and GBM. Random Forest (RF) and Gradient Boosted Machine (GBM) have a good performance for the binary classification problem [Kannan and Vasanthi (2019)]. Also, the previous research showed that RF and GBM performed well in credit card fraud detection [Li, Yan, Liu et al. (2016); Xuan, Liu, Li et al. (2018); Shaik and Srinivasan (2019)]. Based on the cross-validation scores of the five rounds, the ROC curve is plotted. The AUC average score and training time of the three models are compared. First, we compared the AUC score of the three models in Fig. 3.



(a) (b)

Figure 3: The Roc curve of models

From Fig. 3 in which (b) is the larger version of (a), it is known that the AUC score of RF is 98%, the AUC score of GBM is 84%, and the AUC score of LightGBM is 99%. The model that has the highest AUC score is LightGBM.

The results of the experiment indicate that although LightGBM only improves a little compared with RF and GBM, it still makes a big difference. There is one notable thing that the number of credit card transactions is large. So the fraud detection rate is increased by 1%, and a good deal of new illicit transactions per year can be identified.

To better demonstrate the generality of the LightGBM model, other real dataset is also trained. Most of the datasets used in the research on the direction of credit card fraud are publicly available online. Due to sensitive information, the original feature tags of the data are hidden by Principal Component Analysis (PCA) method, replaced with simple tags of V1 to V28 and three named features (Amount, Time and Class) [Awoyemi, Adetunmbi and

Oluwadare (2017)]. The meaning of the 28 labels (V1 to V28) is unclear. It is difficult to choose the suitable features to build the model. In contrast, our data set is more practical. It is helpful to identify fraud in a real environment. Data 1 is the dataset used in this paper. Data 2 contains 284,807 transactions made by credit cards in September 2013 by european cardholders, including 492 fraudulent transactions. It contains onsly numerical input variables which are the result of a PCA transformation. Tab. 2 shows the results of two datasets under LightGBM model. From Tab. 2, it is clear that the AUC scores of Data 1 and Data 2 are 0.991 and 0.982, respectively. Obviously, the performance of LightGBM model is very good.

Table 2: Comparison of auc scores of two datasets under LightGBM model

Datasets	AUC
Data 1	0.991
Data 2	0.982

Next, we compare the time that the three models take, as shown in Tab. 3. The unit of time is second.

Table 3: Comparison of training time

Model	Time of training a model
LightGBM	46.62
GBM	198.59
Random Forest	127.29

From Tab. 3, it is known that the time of LightGBM is 46.62 seconds, the time of GBM is 198.59 seconds, and the time of RF is 127.29 seconds. The model that has the shortest time is LightGBM. Through the experiment, it is shown that our proposed method based on machine learning is very effective in detecting credit card fraud.

5 Conclusion and future work

Credit card fraud has become a challenge for the banking industry. The ability to identify fraudulent transactions online needs to be improved accordingly. In this paper, a credit card fraud detection model based on Light Gradient Boosting Machine algorithm is constructed. Firstly, the feature category is used for the original data. Secondly, we use one-hot encoding to handle some features so that the data is numeric and continuous. It plays a key role in the process of preprocessing. Finally, the Smote algorithm is used to tackle the imbalance of the data. The result of the experiment shows the proposed model performs well in the average score of AUC and the training time of the model.

However, this paper is only used to identify the single fraudulent user. Then, how to judge the fraud ring based on the relation maps is a research direction.

Acknowledgement: This work is supported by Sichuan University Postdoc Research Foundation under Grant 19XJ0002.

References

Adewumi, A. O.; Akinyelu, A. A. (2017): A survey of machine-learning and nature-inspired based credit card fraud detection techniques. *International Journal of System Assurance Engineering and Management*, vol. 8, no. 2, pp. 937-953.

Awoyemi, J. O.; Adetunmbi, A. O.; Oluwadare, S. A. (2017): Credit card fraud detection using machine learning techniques: a comparative analysis. *Proceedings of the IEEE International Conference on Computing, Networking and Informatics*.

Bahnsen, A. C.; Aouada, D.; Stojanovic, A.; Ottersten, B. (2016): Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications*, vol. 51, pp. 134-142.

Bhatla, T. P.; Prabhu, V.; Dua, A. (2003): Understanding credit card frauds. *Cards Business Review*, vol. 1, no. 6, pp. 1-15.

Carcillo, F.; Dal Pozzolo, A.; Le Borgne, Y. A.; Caelen, O.; Mazzer, Y. et al. (2018): Scarff: a scalable framework for streaming credit card fraud detection with spark. *Information Fusion*, vol. 41, pp. 182-194.

Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P. (2002): Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research,* vol. 16, no. 1, pp. 321-357.

Dhankhad, S.; Mohammed, E.; Far, B. (2018): Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. *IEEE International Conference on Information Reuse and Integration*.

Ghosh, S.; Reilly, D. L. (1994): Credit card fraud detection with a neural-network. *Proceedings of the 27th Hawaii International Conference on System Sciences*, vol. 3, pp. 621-630.

Govind, V.; Hazari, S. R. (2014): Credit card fraud detection using hidden markov model. *Indian Streams Research Journal,* vol. 4, no. 4, pp. 37-48.

Jurgovsky, J.; Granitzer, M.; Ziegler, K.; Calabretto, S.; Portier, P. E. et al. (2018): Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, vol. 100, pp. 234-245.

Kannan, R.; Vasanthi, V. (2019): Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease. *Soft Computing and Medical Bioinformatics*, pp. 63-72.

Kulkarni, P.; Ade, R. (2016): Logistic regression learning model for handling concept drift with unbalanced data in credit card fraud detection system. *Proceedings of the Second International Conference on Computer and Communication Technologies.*

Kumari, P.; Mishra, S. P. (2019): Analysis of credit card fraud detection using fusion classifiers. *Computational Intelligence in Data Mining*, pp. 111-122.

Li, Y.; Yan, C.; Liu, W.; Li, M. (2016): Research and application of random forest model in mining automobile insurance fraud. *International Conference on Natural Computation.*

Liu, C.; Chan, Y.; Alam Kazmi, S. H.; Fu, H. (2015): Financial fraud detection model: based on random forest. *International Journal of Economics and Finance*, vol. 7, no. 7.

Maes, S.; Tuyls, K.; Vanschoenwinkel, B.; Manderick, B. (2002): Credit card fraud detection using Bayesian and neural networks. *Proceedings of the 1st International Naiso Congress on Neuro Fuzzy Technologiess*, pp. 261-270.

Mahmoudi, N.; Duman, E. (2015): Detecting credit card fraud by modified fisher discriminant analysis. *Expert Systems with Applications*, vol. 42, no. 5, pp. 2510-2516.

Óskarsdóttir, M.; Bravo, C.; Sarraute, C.; Vanthienen, J.; Baesens, B. (2019): The value of big data for credit scoring: enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing*, vol. 74, pp. 26-39.

Patil, S.; Nemade, V.; Soni, P. K. (2018): Predictive modelling for credit card fraud detection using data analytics. *Procedia Computer Science*, vol. 132, pp. 385-395.

Quah, J. T.; Sriganesh, M. (2008): Real-time credit card fraud detection using computational intelligence. *Expert Systems with Applications*, vol. 35, no. 4, pp. 1721-1732.

Rushin, G.; Stancil, C.; Sun, M.; Adams, S.; Beling, P. (2017): Horse race analysis in credit card fraud-deep learning, logistic regression, and Gradient Boosted Tree. *Systems and Information Engineering Design Symposium*, pp. 117-121.

Sadiq, A. S.; Faris, H.; Ala'M, A. Z.; Mirjalili, S.; Ghafoor, K. Z. (2019): Fraud detection model based on multi-verse features extraction approach for smart city applications. *Smart Cities Cybersecurity and Privacy*, pp. 241-251.

Sahin, Y.; Bulkan, S.; Duman, E. (2013): A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications*, vol. 40, no. 15, pp. 5916-5923.

Shaik, A. B.; Srinivasan, S. (2019): A brief survey on random forest ensembles in classification model. *International Conference on Innovative Computing and Communications*, pp. 253-260.

Van Vlasselaer, V.; Bravo, C.; Caelen, O.; Eliassi-Rad, T.; Akoglu, L. et al. (2015): Apate: a novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, vol. 75, pp. 38-48.

Xuan, S.; Liu, G.; Li, Z. (2018). Refined weighted random forest and its application to credit card fraud detection. *International Conference on Computational Social Networks*, pp. 343-355.

Xuan, S.; Liu, G.; Li, Z.; Zheng, L.; Wang, S. et al. (2018): Random forest for credit card fraud detection. *15th International Conference on Networking, Sensing and Control*, pp. 1-6.

Zareapoor, M.; Yang, J. (2017): A novel strategy for mining highly imbalanced data in credit card transactions. *Intelligent Automation & Soft Computing*, pp. 1-7.