Credit Card Fraud Detection Using Supervised Learning Approach

Rashmi S. More, Chetan J. Awati, Dr. Suresh K. Shirgave, Dr. Rashmi J. Deshmukh, Sonam S. Patil

Abstract: Fraud is a set of illegal activities that are used to take money or property using false pretenses. Transaction fraud using credit card is one of the growing issue in the world of finance. A huge financial loss has significantly affected individuals using credit cards and furthermore vendors and banks. One of the most successful techniques to identify such fraud is Machine learning. This paper proposes a fraud detection algorithm using Random Forest which can help in solving this real world problem. The accuracy of detecting fraud in credit card transaction is increased using this proposed system. The proposed system also uses learning to rank approach to rank the alert that effectively reduces the number of alert generated by FDS thereby providing investigator a small reliable fraud alerts.

Index Terms: Random Forest, class imbalance, credit card fraud, Learning to Rank, concept drift

1 INTRODUCTION

In recent years credit card transaction fraud is a major issue in different sectors like banking, insurance, finance, etc . Fraud is a set of illegal activities to obtain goods and funds. The reason of such illegal transaction might be to get items without giving money. Identifying such illegal activities or fraud is a troublesome and may risk the business and business organizations. In the traditional-approach [1], the algorithm, were written based on strict rules. If a new fraud is detected, then new changes in algorithm is done by fraud analyst either by writing new algorithm or by changing the already existing algorithm. All this changes are done by fraud analyst. In this approach, as number of customers and data increases, human efforts also increase. Also in the real world FDS [1], investigator are not able to check all transactions alert. In this approach, the Fraud Detection System monitors all the approved transactions and alerts the most doubtful one. Verification of all these suspicious alerts are done by investigator and feedback is provided to FDS which indicate that if the transaction was authorized or fraudulent. So this traditional rule-based approach is time-consuming and costly. Only few alerts each day are verified while the rest of the transactions alert remains unchecked until customer identifies them and report them as a fraud. Also, fraudstersuse different techniques in finding a loophole in FDS so that they can do their illegal activities in transaction. The spending behavior of cardholder also changes over time. This change in behavior of cardholder and fraudster during credit card transaction is called as concept drift [1] [7]. So most of the time it is difficult to identify fraud in the credit card transaction.

- *Rashmi More is M. Tech. Student, Department of Technology, Shivaji University Kolhapur, India, E-mail: rashmimore107@gmail.com*
- *Chetan J. Awati is Assistant Professor, Department of Technology, Shivaji University Kolhapur, India, E-mail: cja_tech@unishivaji.ac.in*
- *Dr. Suresh K. Shirgave is Associate Professor, DKTE Society's Textileand Engineering Institute, Ichalkaranji, India, E-mail: skshirgave@gmail.com*
- *Dr. Rashmi J. Deshmukh is Assistant Professor, Department of Technology, Shivaji University Kolhapur, India, E-mail: rvm_tech@unishivaji.ac.in*
- *Sonam S. Patil is Assistant Professor, Department of Information Technology, D Y Patil College of Engineering, Akurdi, Pune India, E-mail: skh9624@gmail.com*

Machine Learning is considered as one of the most successful technique used for creating a fraud detection algorithm for fraud identification. In rule-based approach, algorithm cannot recognize the hidden patterns as they are strictly rule based. We use Machine Learning as it makes machine to learn by itself using classification and regression approach for recognizing fraud in credit card transaction. Due to its fast computing power it has became one of the efficient ways of detecting fraud. The machine learning algorithms are divided into two types, supervised [14][18] and unsupervised [16] learning algorithm. Many supervised and unsupervised machine learning techniques have been presented for fraud detection in credit card transaction which includes logistic regression [3], decision tree [4][15], neural networks [14][19][21][22], Naive Bayes [6], K-Nearest Neighbors [6] ,Support Vector Machines [5] and Random Forest [1][2]. This paper proposes a FDS using Random Forest which can identify transaction fraud in credit card. Random Forest is the advance version of Decision Tree and has better efficiency and accuracy than any other Machine Learning algorithm. The system also uses learning to rank approach to rank the alert generated by the model so that alert with highest rank will only be notified thereby reducing the number of alert detected by rule-based approach FDS.

2. PROPOSED SYSTEM

Increase in online transactions using payment methods like credit card has also increased the fraudulent activities. Every year, a large amount of financial losses are caused by these illegal credit card transactions. No system is 100% secure and there is always a loophole in them. Therefore there is need to solve the issues of detecting fraud in transactions done by credit cards. To overcome this problem the proposed system for fraud identification in credit card transactions is designed using Random Forest algorithm. This algorithm uses combination of Decision Tree to solve the problem. Each tree is trained using dataset and based on this training each tree gives probability of transaction been fraud or legal. After that model predicts the result.

2.1 Objectives

The proposed system will achieve following main objectives:

- To train the model using feedbacks and delayed samples and sum up their likelihood to identify alert.
- To address class imbalance and concept drift issues

by implementing machine learning techniques.
- To increase alert precision develop a learning to rank approach.
- To apply performance measure those are considered in real-world FDS.

The block diagram of proposed system is shown below.

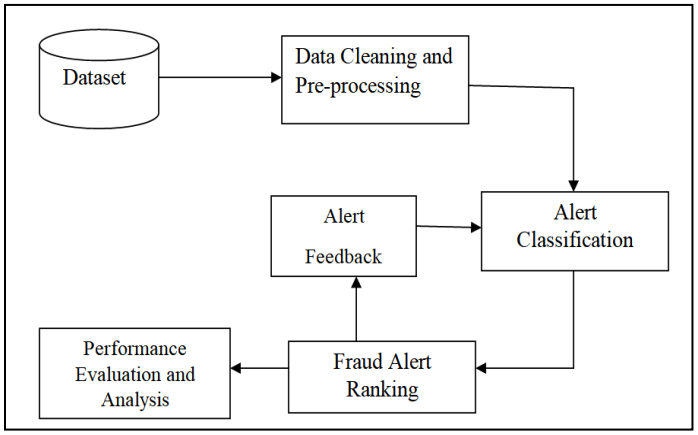

Figure 1 *Proposed System Block Diagram*

2.2 Modules

As shown in the block diagram, following are the modules of system:

a. Data Cleaning and Preprocessing
b. Alerts Classification
c. Alert Ranking
d. Performance Analysis

2.2.1 Data Cleaning and Preprocessing

The model accuracy depends on amount of data on which it is trained. The more amounts of data better will be the performance of model. In this first module the selected data is cleaned and preprocessed as follow:

a. Cleaning: Fixing of missing data or removal of duplicate data from dataset is called as cleaning. The dataset may contain record which may be duplicate, incomplete or may have null values. Such records need to remove by cleaning.
b. Sampling: As number of frauds in dataset is less than overall transaction, class distribution is unbalanced in credit card transaction. Hence sampling method is used to solve this issue.

2.2.2 Alert Classification

Here machine learning model is used that trains the model based on features associated with transactions like location from where transaction is made, zip code, IP address, time and identity of customer. All this dataset is fed as input to the classifier and classifier splits them into multiple decision trees. The sub-trees check this input for an authorized transaction and give probabilities of transaction to be fraud or legal. Combining the results of all sub-trees, the model will alert the fraudulent transaction.

2.2.3 Alert Ranking

This module ranks each alert using learning to rank algorithm. The algorithm ranks each alert identified by the model using likelihood. If it is found that alert has greater rank then a security question is generated. If the individual answers the security question correctly then the transaction is allowed otherwise it is blocked. The IP address and location of fraudster is then tracked by the system. This security questions will be created every time whenever the transaction is identified to be suspicious and rank of alert is highest. This makes the FDS user friendly and helps to launch complaint against fraudster. Also the number of alert generated by the system is reduced as compared to rule-based approach system.

2.3 Algorithm of Proposed Strategy

User inputs v1,v2,…vn
Dataset D

Step1: Initiate User Input
Step2: for each transaction x do
deploy pattern P
compute probability P_D
generate alerts A
Step3: Implement RankNet Model to generate rank
return rank
Step 4: if rank>=n then
Display Security Question SQ
verify SQ
Step 5: if SQ verified then
allow transaction
else
block transaction and track fraudster location
[end if]

3. Result

In this study a dataset used contained 100000 transactions made by cardholder. 0.262 % of all transactions belong to fraud activities. This dataset is highly imbalanced. A data preprocessing is carried out on this imbalanced dataset and 80% of dataset was used for training the model whereas 20% of dataset was used for testing.

3.1 Performance Evaluation

The performance evaluation of proposed system is done based on F1 score, precision, recall (sensitivity) and accuracy. Figure 2 shows proposed system output where we see that the accuracy of system is 0.9793. This shows that proposed strategy had showed better accuracy for large number of training data.

```
Confusion matrix :
 [[19451    34]
 [  379   136]]
Outcome values :
 tp= 136 fn= 379 fp= 34 tn= 19451
Classification report :
              precision    recall  f1-score   support

           1       0.80      0.26      0.40       515
           0       0.98      1.00      0.99     19485

    accuracy                           0.98     20000
   macro avg       0.89      0.63      0.69     20000
weighted avg       0.98      0.98      0.97     20000

rf acc is  0.97935
```

Figure 2 *Performance Evaluation of Proposed System*

There are total 20,000 transactions in our test dataset out of which 19,830 transactions belongs to class 0 and 170 transactions belongs to class 1. Now from confusion matrix shown in Fig.2 we see that value of True Negative (tn) is 19,451 which means that out of 19,830 transactions which belongs to class 0, 19,451 are predicted as class 0. Furthermore, from confusion matrix we see that value of True Positive (tp) is 136 which means that out of 170 transactions that belongs to class 1, 136 are predicted as class 1. 34 transactions from class 1 are detected falsely. To conclude we can say that despite being an imbalanced dataset our model is working well.

```
127.0.0.1 - - [16/Sep/2020 11:56:44] "POST /formdata HTTP/1.1" 200 -
maxtransactionfromuseratthislocation 2
3   20.7   70.9833   123 Diu
Rank is :  2
Fraud detected
127.0.0.1 - - [16/Sep/2020 11:57:29] "POST /checkingquestion HTTP/1.1" 200 -
{'country_code': 'IN', 'country_name': 'India', 'city': 'Diu', 'postal':
'362520', 'latitude': 20.7, 'longitude': 70.9833, 'IPv4': '152.57.235.160',
'state': 'Daman and Diu'}
```

Figure 3 *Result after applying Ranking approach*

3.2 Comparative Performance

The performance evaluation of proposed system and other two classifier i.e Decision Tree and Naive Bayes is shown in figure 3. We can clearly see that our proposed system using Random Forest technique performed much better than Decision Tree and Naïve Bayes Technique.

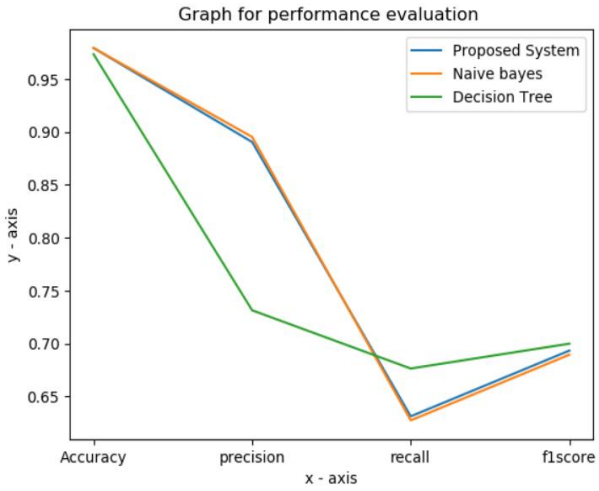

Figure 4 *Comparative Performance Evaluation of Proposed System*

4. Conclusion

This paper has proposed a system to classify alerts in fraud detection system using supervised learning technique Random Forest to classify alert as fraudulent or non-fraudulent. Further we have also used learning to rank approach to rank the fraudulent alert generated by classifier based on priority. The performances of all this techniques are examined based on precision,F1 score, recall (sensitivity) and accuracy. A comparative study is also done where proposed system is compared with Decision Tree and Naive Bayes Technique. It showed that the proposed system showed better accuracy for a larger dataset.

Future work concerns the classification of alerts by applying semi-supervised learning methods in FDS.

REFERENCES

[1] Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi and GianlucaBotempi, "Credit card Fraud Detection : A realistic Modeling and a Novel Learning Strategy", IEEE Trans. on Neural Network and Learning system,vol.29,No.8, August 2018.

[2] ShiyangXuan,GuanjunLiu,ZhenchuanLi,LutaoZheng,Shuo Wang, Jiang,"Random Forest for credit card fraud detection",Int.conf.onNetworking,Sensing and control,2018.

[3] Y. Sahin , and Duman,E.,(2011) "Detecting credit card fraud by ANN and logistic regression." In Innovations in Intelligent Systems and Applications(INISTA),2011 international Symposium on (pp.315-319).IEEE

[4] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," Expert Syst. Appl., vol. 40, no. 15,pp. 5916–5923, 2013

[5] Sahin Y. and Duman E. (2011),"Detecting Credit Crad Fraud by Decision Trees and Support Vector Machines", International Multi-Conference Of Engineers and Computer Scientists(IMECS 2011),Mar 16-18,Hong Kong,Vol.1,pp.1-6

[6] Sai Kiran, Jypti Guru, Rishabh Kumar, Naveen Kumar, Deepak Katariya,"Credit card fraud detection using Naïve Bayes model based and KNN classifier", Int. Journal of Adv. Research , Ideas and Innovations in Technology,vol.4,2018.

[7] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection and concept-drift adaptation with delayed supervised information," in Proc. Int. Joint Conf. Neural Netw., 2015,pp. 1–8.

[8] A. C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive decision trees," Expert Syst. Appl., vol. 42, no. 19,pp. 6609–6619, 2015

[9] A. Dal Pozzolo, O. Caelen, and G. Bontempi, "When is undersampling effective in unbalanced classification tasks?" in Machine Learning and Knowledge Discovery in Databases. Cambridge, U.K.: Springer, 2015

[10] N. Mahmoudi and E. Duman, "Detecting credit card fraud by modified fisher discriminant analysis," Expert Syst. Appl., vol. 42, no. 5,pp. 2510–2516, 2015

[11] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Detecting credit card fraud using periodic features," in Proc. 14th Int. Conf. Mach. Learn. Appl., Dec. 2015, pp. 208–213.

[12] AnuruddhaThennakoon, Chee Bhagyani, SasithaPremadasa, ShalithaMihiranga,"Realtime Credit Card Fraud Detection Using Machine Learning ,"Int. Conf. on Cloud Computing, Data Science & Engineering,2019.

[13] S. Wang, L.L. Minku, and X. Yao, "Resampling-based ensemble methods for online class imbalance learning." Trans, Knowl., Data Eng., vol 27, no. 5, pp. 1356-1368, May 2015.

[14] Jan may Kumar Behera, SuvasiniPanigrahi, "Credit Card Fraud Detection: A Hybrid Approach using Fuzzy Clustering and Neural Network", 2015 IEEE Second International Conference on Advances in Computing and Communication Engineering.

[15] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, BankSealer: “A Decision Support System for Online Banking Fraud Analysis and Investigation”, Berlin, Germany: Springer, 2014, pp. 380–394

[16] R. J. Bolton and D. J. Hand, “Unsupervised profiling methods for fraud detection,” in Credit Scoring Credit Control VII. London, U.K.: Imperial College London, 2001, pp. 235–255

[17] R. Elwell and R. Polikar, “Incremental learning of concept drift in nonstationary environments,” Trans. Neural Netw., vol. 22, no. 10,pp. 1517–1531, 2011.

[18]S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, “Data
mining for credit card fraud: A comparative study,” Decision Support
Syst., vol. 50, no. 3, pp. 602–613, 2011.

[19] Tao Guo ,Gui-Yang Li ,”Neural data mining for credit card fraud detection”,Int.Conf. on Machine Learning and Cybernetics, Sept 2008

[20]J. Gao, B. Ding, W. Fan, J. Han, P.S. Yu, “Classifying data streams with skewed class distributions and concept drifts”, IEEE internet comput., vol.12, no. 6, pp. 37-49, Nov 2008

[21]E. Aleskerov, B. Freisleben, and B. Rao, “CARDWATCH: A neural network based database mining system for credit card fraud detection,” in Proc. IEEE/IAFE Computat. Intell. Financial Eng., Mar. 1997, pp. 220–226.

[22] J.R. Dorronsoro, F. Ginel, C. Sgnchez and C.S. Cruz, “Neural fraud detection in credit card operations”, IEEE transaction neural network vol. 8, no. 4, pp. 827-834, Jul.1997.

[23] D. J. Weston, D. J. Hand, N. M. Adams, C. Whitrow, and P. Juszczak,“Plastic card fraud detection using peer group analysis,” Adv. Data Anal.Classification, vol. 2, no. 1, pp. 45–62, 2008.