IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of
Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going

[1] K SUPARNA, [2] S. HALIMUNNISA

[1](Assistant Professor), MCA, DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH

[2]MCA, scholar, DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH

ABSTRACT

Recently, Internet finance is increasingly popular. However, bad debt has become a serious threat to Internet financial companies. The fraud detection models commonly used in conventional financial companies is logistic regression. Although it is interpretable, the accuracy of the logistic regression still remains to be improved. This paper takes a large public loan dataset, *e.g.* Lending club, for example, to explore the potential of applying deep neural network for fraud detection. We first _all the missing values by a random forest. Then, an XGBoost algorithm is employed to select the most discriminate features. After that, we propose to use a synthetic minority oversampling technique to deal with the sample imbalance. With the pre-processed data, we design a deep neural network for Internet loan fraud detection. Extensive experiments have been conducted to demonstrate the outperformance of the deep neural network compared with the commonly-used models. Such a simple yet effective model may brighten the application of deep learning in anti-fraud for Internet loans, which would benefit the financial engineers in small and medium Internet financial companies.

1.INTRODUCTION

Internet fraud methods are increasing dramatically in recent years, together with the rapid development of Internet financial models and the Internet business used to be handled by traditional financial institutions. In this regard, Internet lending companies face an unprecedented risk of online fraud. Luckily, the rapid development of computer technology, the accumulating data, and the emerging data analysis techniques bring new opportunities to financial risk management and analysis on the big data in the financial industry.

Researchers have developed various anti-fraud measures and fraud prevention systems over the years. Leonard [1] proposed a rule-based expert system for fraud detection. The rules of this model were manually constructed by the fraud experts from the bank. Sanchez *et al.* [2] proposed to use association rules to detect fraud and help risk analysts extract more fraud rules. Edge and Sampaio [3] proposed a set of a financial fraud modelling language (FFML) for better describing and combining fraud rule sets to assist fraud analysis. However, the rule-based models require sufficient and accurate expertise knowledge and cannot be updated timely to new frauds.

To this end, machine learning models have been introduced for fraud detection. Ghosh and Reilly [4] uses neural networks to detect credit card fraud. Kokkinaki [5] proposed decision trees and Boolean logic functions to characterize normal transaction patterns to detect fraudulent transactions. Peng *et al.* [6] compared nine machine learning models for fraud detection. The results demonstrate linear logistic and IEEE Access andTransaction on Deep Learning, Volume:9,Issue Date:12.January.2021 Bayesian networks are more effective. Lei and Ghorbani [7] proposed a new clustering algorithm namely improved competitive learning network (ICLN) and supervised an improved competitive learning network (SICLN). Sahin *et al.* [8] designed a decision tree based on cost sensitivity. Halvaiee and Akbari [9] proposed to use an AIRS improved algorithm for fraud detection.

However, these traditional machine learning methods heavily rely on manual subjective rules and easily lead to model risk. These methods also tend to over due to the imbalance training dataset with serious pollution by noises. Thus, ensemble learning methods have also been introduced to integrate different models for complicated fraud detection. Louzada and Ara [10] proposed a bagging ensemble model that integrates k-dependence probabilistic networks. The results show that the proposed ensemble model has stronger modelling capabilities.

Carminati *et al.* [11] proposed a combination of semi-supervised and unsupervised fraud and anomaly detection methods, mainly using a histogram-based outlier score (HBOS) algorithm to model the user's past behaviour.

Recently, deep learning techniques have attracted a lot of academic and industrial attention that provides a new insight for financial data analysis. Fu *et al.* [12] used con volitional neural networks to effectively reduce feature redundancy. Tu *et al.* [13] design a deep feature representation technique for fraud detection. To incorporate with prior knowledge with the deep network, Greiner and Wang [14] pointed out the borrower is likely to conceal information that is not beneficial to him or even fictitious favourable information before obtaining the loan. After obtaining the loan, the borrower is likely to default unilaterally. Pope and Sydnor [15] also found it difficult to judge the risk of the personal information provided by the borrower unilaterally because the authenticity of this information cannot be verified. Freedman and Jin [16] uncovered that the borrower may commit fraudulent behaviour by reporting false information, which exacerbates the information asymmetry between the two parties.

Herzen stein *et al.* [17] also found that the borrowers' repayment ability and credit rating are the factors that have the greatest impact on personal credit risk. They concluded that economic strength is the determinant of judging the availability of borrowing. At the same time, Herzen stein *et al.* [18] depicted the borrowers' spending power can also directly affect the success rate of borrowing. These methods reveal the characteristics of the borrowers would be helpful for fraud detection.

Motivated by such an idea, we propose a deep learning technique to mine the fraud in a public lending dataset with 200,000 records. We analyse the customer credit rating, which can help us to identify customers' actual situations. Intuitively, the lower a customer has a credit rating, such as the rating, the greater the likelihood of being a fraudulent user. Internet finance small loan companies set different thresholds on their customer credit rating data to build anti-fraud rules based on the true information of their customers.

This paper aims to provide small financial credit companies a simple yet effective model to improve their risk control and the level of anti-fraud. Such companies often have a poor-risk control capacity with limited capacity for data engineering, modelling, and optimization.

The main contribution of this paper is summarized as follows:

- First, we analyse the real-world Internet financial data for the missing data and sample imbalance. We propose to all the missing with a random forest and deal with the sample imbalance with a synthetic minority oversampling technique.
- We train a deep neural network by the pre-processed data. We make comprehensible experiments for the setting of the network architecture and hyper parameters.
- Extensive experiments have been conducted to demonstrate the outperformance comparing with the commonly-used loan fraud detection models.

The rest of this paper is organized as follows. The second part is the methodology, and the third part is an empirical study from real-world data. The fourth part is the conclusion.

2. EXISTING SYSTEM

Ghosh and Reilly [4] uses neural networks to detect credit card fraud. Kokkinaki [5] proposed decision trees and Boolean logic functions to characterize normal transaction patterns to detect fraudulent transactions. Peng *et al.* [6] compared nine machine learning models for fraud detection. The results demonstrate linear logistic andBayesian networks are more effective. Lei and Ghorbani [7] proposed a new clustering algorithm namely improved competitive learning network (ICLN) and supervised an improved competitive learning network (SICLN). Sahin *et al.* [8] designed a decision tree based on cost sensitivity.

Halvaiee and Akbari [9] proposed to use an AIRS improved algorithm for fraud detection. However, these traditional machine learning methods heavily rely on manual subjective rules and easily lead to model risk. These methods also tend to over fit due to the imbalance training dataset with serious pollution by noises. Thus, ensemble learning methods have also been introduced to integrate different models for complicated fraud detection.

Louzada and Ara [10] proposed a bagging ensemble model that integrates k-dependence probabilistic networks. The results show that the proposed ensemble model has stronger modelling capabilities. Carminati et *al.* [11] proposed a combination of semi-supervised and unsupervised fraud and anomaly detection methods, mainly using a histogram-based outlier score (HBOS) algorithm to model the user's past behaviour.

DISADVANTAGES:

1)The system doesn’t analyse for large number of data sets due to lack of ml classifies.

2)The system couldn’t implement to detect the following:

(i) Level of Loan activity.

(ii) Level of Loan Prediction.

(iii) Loan Profile information.

3. PROPOSED SYSTEM

This paper aims to provide small financial creditcompanies a simple yet effective model to improve their risk control and the level of anti-fraud. Such companies often have a poor-risk control capacity with limited capacity for data engineering, modelling, and optimization. The main contribution of this paper is summarized as follows.

First, we analyse the real-world Internet financial data for the missing data and sample imbalance. We propose to fill the missing with a random forest and deal with the sample imbalance with a synthetic minority oversampling technique.

We train a deep neural network by the pre-processed data. We make comprehensible experiments for the setting of the network architecture and hyper parameters.

Extensive experiments have been conducted to demonstrate the outperformance comparing with the commonly-used loan fraud detection models.

DISADVANTAGES

Identity theft:

Criminals steal user's personal financial information in order to conduct fraudulent financial transaction activities or withdraw money from your account.

Investment fraud:

Selling investment or securities with false, misleading, or fraudulent information.

Mortgage and loan fraud:

The borrower uses false information to open a mortgage or loan, or the lender uses a high-pressure sales strategy to sell the mortgage or loan or predatory loan to users.

Large-scale marketing fraud:

Criminals usually use a lot of mail, telephone, or spam to steal users' personal financial information or request donations and fees from fraudulent organizations, usually involving fake checks, charities, sweepstakes, lotteries, and exclusive clubs or honour society invites.

4. OUTPUTSCREENS

Home screen

Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going
Login

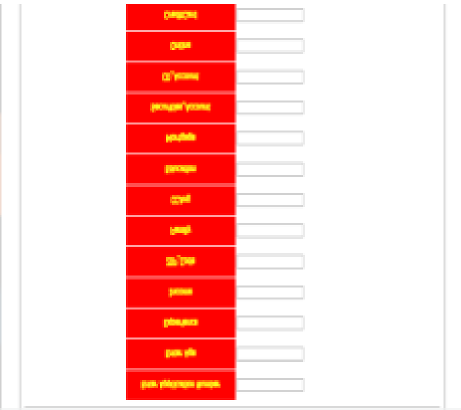
USER REGISTRATION:

Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going
REGISTER NOW!

USER LOGIN:

Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going
Login

PREDITION OF LOAN APPROVED STATUS:



USER PROFILE:

Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going

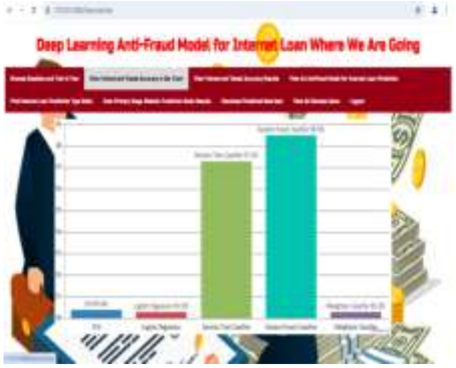
SERVICE PROVIDER LOGIN:

Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going
Login

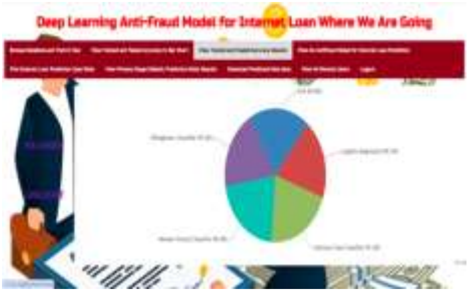
BROWSE DATASET AND TRAIN&TEST DATA

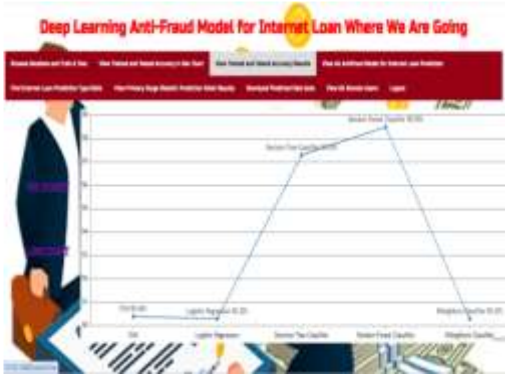
Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going

VIEW TRAINED AND TESTED ACCURACY IN BAR CHAT:


Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going

VIEW TRAINED AND TESTED ACCURACY RESULTS:


Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going


Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going

INTERNET LOAN PREDICTION DETAILS:

INTERNET LOAN PREDICTION TYPE RATIO:

Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going

PREDICTION RATIO RESULTS:

Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going

Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going

ALL REMOTE USERS:

Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going

5. CONCLUSION

In this paper, we take the real customer information of the public loan data set of the lending club company as a sample. Then, we build a deep learning-based Internet fraud detection model. We introduce the main parameters of the model and optimizes to find the optimal parameter combination of the model. Finally, the most popular logistic regression in the financial industry as well as other comparisons are used as a baseline to evaluate the performance of the proposed model. The results reveal the deep neural network achieves better Performance, which is promising to be used in the financial industry for Internet fraud detection. In the future, we plan to cooperate with mature Internet financial technology companies and banks in China for blacklists and white lists. The deep neural network combined with such blacklists and white lists and the expertise anti-fraud rules is promising to increase fraud detection capability.

6. REFERENCE

[1] K. J. Leonard, ``The development of a rule based expert system model for fraud alert in consumer credit," *Eur. J. Oper. Res.*, vol. 80, no. 2, pp. 350_356, Jan. 1995.

[2] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, ``Association rules applied to credit card fraud detection,'' *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3630_3640, Mar. 2009.

[3] M. E. Edge and P. R. F. Sampaio, ``The design of FFML: A rule-based policy modelling language for proactive fraud management in _nancial data streams,'' *Expert Syst. Appl.*, vol. 39, no. 11, pp. 9966_9985, Sep. 2012.

[4] S. Ghosh and D. L. Reilly, *Credit Card Fraud Detection With a Neural-Network*. Wailea, HI, USA: IEEE, 1994.

[5] A. I. Kokkinaki, ``On atypical database transactions: Identi_cation of probable frauds using machine learning for user prowling,'' in *Proc. IEEE Knowl. Data Eng. Exchange Workshop*, 1997, pp. 229_238.

[6] Y. Peng, G.Wang, G.Kou, andY. Shi, ``An empirical study of classi_cation algorithm evaluation for _nancial risk prediction,'' *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2906_2915, Mar. 2011.

[7] J. Z. Lei and A. A. Ghorbani, ``Improved competitive learning neural networks for network intrusion and fraud detection,'' *Neurocomputing*, vol. 75, no. 1, pp. 135_145, Jan. 2012.

[8] 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, Nov. 2013.

[9] N. Soltani Halvaiee and M. K. Akbari, ``A novel model for credit card fraud detection using arti_cial immune systems,'' *Appl. Soft Comput.*, vol. 24, pp. 40_49, Nov. 2014.

[10] F. Louzada and A. Ara, ``Bagging k-dependence probabilistic networks: An alternative powerful fraud detection tool,'' *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11583_11592, Oct. 2012.