ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue.2 April 2024

FRAUD DETECTION IN BANKING DATA BY MACHINE LEARNING TECHNIQUES

Budimudi Yadidhya[1], Choodi Yashwanth Reddy[2],
Mohammad Abdul Zameer Pasha[3],
Dr. A. Sudhakar[4]
[1,2,3] UG Student, Department of ECE, CMR Institute of Technology, Hyderabad
[4] Associate Professor, Department of ECE, CMR Institute of Technology, Hyderabad

ABSTRACT

As technology advanced and e-commerce services expanded, credit cards became one of the most popular payment methods, resulting in an increase in the volume of banking transactions. Furthermore, the significant increase in fraud requires high banking transaction costs. As a result, detecting fraudulent activities has become a fascinating topic. In this study, we consider the use of class weight-tuning hyper parameters to control the weight of fraudulent and legitimate transactions. We use Bayesian optimization in particular to optimize the hyper parameters while preserving practical issues such as unbalanced data. We propose weight-tuning as a per-process for unbalanced data, as well as Cat Boost and Boost to improve the performance of the Limelight method by accounting for the voting mechanism. Finally, in order to improve performance even further, we use deep learning to fine-tune the hyper parameters, particularly our proposed weight-tuning one. We perform some experiments on real-world data to test the proposed methods. To better cover unbalanced datasets, we use recall-precision metrics in addition to the standard ROCOCO. Cat Boost, Limelight, and Boost are evaluated separately using a 5-fold cross-validation method. Furthermore, the majority voting ensemble learning method is used to assess the performance of the combined algorithms. Limelight and Boost achieve the best level criteria of ROCOCO D 0.95, precision 0.79, recall 0.80, F1 score 0.79, and MCC 0.79, according to the results. By using deep learning and the Bayesian optimization method to tune the hyper parameters, we also meet the ROCOCO D 0.94, precision D 0.80, recall D 0.82, F1 score D 0.81, and MCC D 0.81. This is a significant improvement over the cutting-edge methods we compared it to.

INTRODUCTION

In recent years, there has been a significant increase in the volume of financial transactions due to the expansion of financial institutions and the popularity of web-baside-commerce. Fraudulent transactions have become a growing problem in online banking, and fraud detection has always been challenging [1], [2]. Along with credit card development, the pattern of credit card fraud has always been updated. Fraudsters do their best to make it look legitimate, and credit card fraud has always been updated. Fraudsters do their best to make it look The associate editor coordinating the review of this manuscript and approving it for publication was Than Bu legitimate. They try to learn how fraud detection systems work and continue to stimulate these systems, making fraud detection more complicated. Therefore, researchers are constably trying to find new ways or improve the performance of the existing methods. People who commit fraud usually use security, control, and monitoring weaknesses in commercial applications to achieve their goals. However, technology can be a tool to combat fraud [4]. To prevent further possible fraud, it is important to detect the fraud right away after its occurrence Fraud can be defined as wrongful or criminal deception intended to result in financial or personal gain. Credit card fraud is related to the illegal use of credit card information for purchases in a physical or digital manner. In digital transactions, fraud can happen over the line or the web, since the cardholders usually provide the card number, expiration date, and card verification number by telephone or website [6]. There are two mechanisms, fraud prevention and fraud detection, that can be exploited to avoid fraud-related losses. Fraud prevention is a proactive method that stops fraud from happening in the first place. On the other hand, fraud detection is needed when a fraudster attempts a fraudulent transaction. Fraud detection in banking is considered a binary classification problem in which data classified as legitimate or fraudulent [8]. Because banking data is large in volume and with datasets containing a large amount of transaction data, manually reviewing and finding patterns for fraudlent transactions either impossible or takes a long time.

LITERATURE REVIEW

R. Alistair, A. Goodhearted, A. R. Both a, and E. Jaycees, ‘‘Analyzing credit card fraud detection based on machine learning models,’’ in Prof. *IEEE Int. Io T, Electron. Mechanics Con. (IEMTRONICS)*, Jun. 2022, pp. 1–8

This comprehensive review paper investigates the current status of credit card fraud detection by using both traditional and advanced machine learning techniques. The text presents a range methods, each with its own advantages and disadvantages. These methods include Decision Trees (DT), Logistic Regression (LR), K-Nearest Neighbour (KNN), Neural Networks (NN), Naive Bayes (NB), Genetic Algorithms (GA), Hidden Markov Models (HMM), Support Vector Machines (SVM), Fuzzy Logic-based Systems (FLBS), Hybrid Approaches, and Privacy-preserving Techniques. DT sacrifice generalization capacity in exchange for interpret ability, making them prone to over fitting. On the other hand, R's performance is hindered by its susceptibility to outliers. Although NN excel at detecting intricate patterns, they may be relatively demanding in terms of computational resources.

A. C. Baseness, D. Douala, A. Stochastic, and B. Otters ten, ‘‘Feature and is often used

] **A. C. Baseness, D. Douala, A. Stochastic, and B. Otters ten, ‘‘Feature engineering strategies for credit card fraud detection,’’ *Expert Cyst. Apply.*, vol. 51, pp. 134–142, Jun. 2016.**

The system never implements Majority Voting model which leads less effective.The use of credit cards and the advent of online purchasing have significantly facilitated the lives of both consumers and retailers [1]. Regrettably, the advent of the digital revolution has seen a significant surge in instances of credit card theft. Credit card fraud is a significant challenge for financial institutions and individuals worldwide, including unauthorized transactions, identity theft, and account hijacking [2]. Credit card fraud is a pressing issue for effective remedies, given the financial ramifications and the erosion of trust in digital payment mechanisms. The detection of fraud has become inadequate with the rise of intricate fraudulent schemes, rendering rule-based systems and human evaluations insufficient [3]. Manual assessments are characterized by their time-consuming nature, high costs, and susceptibility to human error. Conversely, rule-based systems sometimes lack the necessary adaptability to effectively address emerging fraud tendencies.

H. Wang, P. Thu, X. Zoe, and S. Sin, ‘‘An ensemble learning framework for credit card fraud detection based on training set partitioning and clustering,’’ in Prof. *IEEE Smart World, Ubiquitous Intel. Com put., Adv. Trusted Com-put., Salable Com put. Com mun., Cloud Big Data Compute., Internet People Smart City Innovate. (SmartWorld/SCALCOM/UIC/ATC/CBD Com/IOP/SCI)*, Oct. 2018, pp. 94–98

Timely detection of fraudulent credit card transactions is a business critical and challenging problem in Financial Industry. Specifically, is, the ratio of fraud to normal transactions is very small. In this work, we present an ensemble machine learning approach as a possible solution to this problem. Our observation is that *Random Forest* is more accurate in detecting normal instances, and *Neural Network* is for detecting fraud instances. We present an ensemble method – based on a combination of random forest and neural network which keeps the best of both worlds, and is able to predict with high accuracy and confidence the label of a new sample. We experimentally validate our observations on real world datasets Over last decade, due to rise of e-commerce, the use of credit cards has increased dramatically. This has also increased the risk of fraudulent transactions. Further, credit card transactions are considered as an easy fraud target because of its low risk and high reward nature. Incidence of credit card fraud is limited to about 0.1% of all card transactions, but it may result into huge financial losses as transactions can be of quite large amount

] J. Durian, Y.-W. Li, K. Yang, and Y. Ria, ‘‘E commerce fraud detection through fraud islands and multi-layer machine learning model,’’ in Prof. *Future Inf. Com-mun. Cong.*, in Advances in Information and Communication. San Francisco, CA, USA: Sp ringer, 2020, pp.556–570

Main challenge for e-commerce transaction fraud prevention is that fraud patterns are rather dynamic and diverse. This paper introduces two innovative methods, fraud islands (link analysis) and multi-layer machine learning model, which can effectively tackle the challenge of detecting diverse fraud patterns. Fraud Islands are formed using link analysis to investigate the relationships between different fraudulent entities and to uncover the hidden complex fraud patterns through the formed network. Multi-layer model is used to deal with the largely diverse nature of fraud patterns. Currently, the fraud labels are determined through different channels which are banks' declination decision, manual review agents' rejection decisions, banks' fraud alert and customers' chargeback requests. It can be reasonably assumed that different fraud patterns could be caught though different fraud risk prevention forces (i.e. bank, manual review team and fraud machine learning model). The experiments showed that by integrating few different machine learning models which were trained using different types of fraud labels, the accuracy of fraud decisions can be significantly improved..

EXISTINGSYSTEM

Hallie&Akbar study a new model called the AIS-based fraud detection model (AFDM). They use the Immune System Inspired Algorithm (AIRS) to improve fraud detection accuracy. The presented results of their paper show that their proposed AFDM improves accuracy by up to 25%, reduces costs by up to 85%, and reduces system response time by up to 40% compared to basic algorithms [11].

Baseness At AL. developed a transaction aggregation strategy and created a new set of features based on the periodic behaviouranalysis of the transaction time by using the con Moses distribution. In addition, they propose a new cost-based criterion for evaluating credit card fraud detection's models and then, using a real credit card dateset, examine how different feature sets affect results. More precisely, they extend the transaction aggregation strategy to create new offers based on an analysis of the periodic behaviour of transactions [12]. Rwandan e a. study the application of machine learning algorithms to detect fraud in credit cards. They _est use Nave Bayes, stochastic forest and decision trees, neural networks, linear regression (LR), and logistic regression, as well as support vector machine standard models, to evaluate the available datasets. Further, they propose a hybrid method by applying Ada Boost and majority voting. In addition, they add noise to the data samples for robustness evaluation.

Disadvantages

The system never use a sequential model, which is a linear stack of layers to construct an artificial neural network model. Our model has a dense class, which is a very common layer **]**

Proposed System

The system proposes an efficient approach for detecting credit card fraud that has been evaluated on publicly available datasets and has used optimized algorithms SVM and logistic regression individually, as well as majority voting combined methods, as well as deep learning and hyper parameter settings. An ideal fraud detection system should detect more fraudulent cases, and the precision of detecting fraudulent cases should be high, i.e., all results should be correctly detected, which will lead to the trust of customers in the bank, and on the other hand, the bank will not suffer losses due to incorrect detection. propose a group learning frameworkbased on partitioning and clustering of the training set. Theirproposed framework has two goals: 1) to ensure the integrityof the sample features, and 2) to solve the high imbalanceof the dataset. The main feature of their proposed frameworkis that every base estimator can be trained in parallel, whichimproves the effectiveness of their framework.

Advantages

_ We adopt Bayesian optimization for fraud detection and propose to use the weight-tuning hyper parameter to solve the unbalanced data issue as a per-process step. We also suggest using CatBoost and Boost alongside Limelight to improve performance. We use the Boost algorithm due to the high speed of training in big data as well as the regularization term, which overcomes over fitting by measuring the complexity of the tree, and it does not require much time to set the hyper parameters. We also use the Cat boost algorithm because there is no need to adjust hyper parameters for over fitting control, and it also obtains good results without changing hyper parameters compared to other machine learning algorithms.

MODULES

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such asTrain & Test Bank Datasets, View Trained and Tested Datasets Accuracy in Bar Chart, View Trained and Tested Datasets Accuracy Results, View Prediction Of Bank Fraud Detection, View Prediction Of Bank Fraud Detection Ratio, Download Predicted Data Sets, View Bank Fraud Detection Ratio Results View All Remote Users

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user’s details such as, user name, email, address and admin authorizes the users.

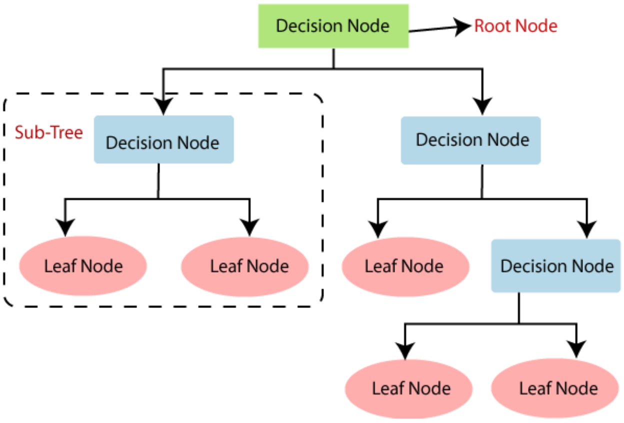
Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations likeREGISTER AND LOGIN,PREDICT BANK FRAUD DETECTION TYPE, VIEW YOUR PROFILE.

ALGORITHMS

DECISION TREE CLASSIFICATION ALGORITHM

- Decision Tree is a **Supervised learning technique** that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where **internal nodes represent the features of a dataset, branches represent the decision rules** and **each leaf node represents the outcome.**
- In a Decision tree, there are two nodes, which are the **Decision Node** and **Leaf Node.** Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.



K-NEAREST NEIGHBOR(KNN) ALGORITHM FOR MACHINE LEARNING

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category

that is most similar to the available categories.

- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K-NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
- It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

KNN Classifier
Input value
Predicted Outpu

Logistic regression Classifiers

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Naïve Bayes

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .

Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).

SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels fornewly acquired instances. Unlike generative machine learning approaches, which require computations ofconditional probability distributions, a discriminant classification function takes a data point *x* and assignsit to one of the different classes that are a part of the classification task. Less powerful than generativeapproaches, which are mostly used when prediction involves outlier detection, discriminant approachesrequire fewer computational resources and less training data, especially for a multidimensional featurespace and when only posterior probabilities are needed. From a geometric perspective, learning a classifieris equivalent to finding the equation for a multidimensional surface that best separates the different classesin the feature space.

CONCLUSION

In this paper, we studied the credit card fraud detectionproblem in real unbalanced datasets.We proposed a machinelearningapproach to improve the performance of frauddetection.We used a publicly available ``credit card" dataset with 28featuresand0.17 percent of the fraud data. We proposedtwo methods. In the proposed LightGBM, we usedclass weight tuning to choose the proper hyperparameters.We used the common evaluation metrics, including accuracy,precision, recall, F1-score, and AUC. Our experimentalresults showed that the proposed LightGBM methodimproved the fraud detection cases by 50% and the F1-scoreby 20% compared with the recently presented method in [17].We improve the performance of the algorithm with the helpof the majority voting algorithm. We also improved the criteriaby using the deep learning method. The assurance ofthe results of MCC for unbalanced data proved that, comparedto other criteria of evaluation, it's stronger. In thispaper, by combining the LightGBM and XGBoost methods,we obtained 0.79 and 0.81 for the deep learning method.

REFERENCES

[1] J. Nanduri, Y.-W. Liu, K. Yang, and Y. Jia, ``Ecommerce fraud detection through fraud islands and multi-layer machine learning model,'' in *Proc. Future Inf. Commun. Conf.*, in Advances in Information and Communication. San Francisco, CA, USA: Springer, 2020, pp. 556_570.

[2] I. Matloob, S. A. Khan, R. Rukaiya, M. A. K. Khattak, and A. Munir, ``A sequence mining-based novel architecture for detecting fraudulent transactions in healthcare systems,'' *IEEE Access*, vol. 10, pp. 48447_48463, 2022.

[3] H. Feng, ``Ensemble learning in credit card fraud detection using boosting methods,'' in *Proc. 2nd Int. Conf. Comput. Data Sci. (CDS)*, Jan. 2021, pp. 7_11.

[4] M. S. Delgosha, N. Hajiheydari, and S. M. Fahimi, ``Elucidation of big data analytics in banking: A four-stage delphi study,'' *J. Enterprise Inf. Manage.*, vol. 34, no. 6, pp. 1577_1596, Nov. 2021.

[5] M. Puh and L. Brki¢, ``Detecting credit card fraud using selected machine learning algorithms,'' in *Proc. 42nd Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2019, pp. 1250_1255.

[6] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, ``Credit card fraud detection using AdaBoost and majority voting,'' *IEEE Access*, vol. 6, pp. 14277_14284, 2018.

[7] N. Kumaraswamy, M. K. Markey, T. Ekin, J. C. Barner, and K. Rascati, ``Healthcare fraud data mining methods: A look back and look ahead,'' *Perspectives Health Inf. Manag.*, vol. 19, no. 1, p. 1, 2022.

[8] E. F. Malik, K. W. Khaw, B. Belaton, W. P. Wong, and X. Chew, ``Credit card fraud detection using a new hybrid machine learning architecture,'' *Mathematics*, vol. 10, no. 9, p. 1480, Apr. 2022.

[9] K. Gupta, K. Singh, G. V. Singh, M. Hassan, G. Himani, and U. Sharma, ``Machine learning based credit card fraud detection_A review,'' in *Proc. Int. Conf. Appl. Artif. Intell. Comput. (ICAAIC)*, 2022, pp. 362_368.

[10] R. Almutairi, A. Godavarthi, A. R.Kotha, and E. Ceesay, ``Analyzing credit card fraud detection based on machine learning models,'' in *Proc. IEEE Int. IoT, Electron. Mechatronics Conf. (IEMTRONICS)*, Jun. 2022, pp. 1_8.

[12] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Design of a Dual Doping Less Double Gate Tfet and Its Material Optimization Analysis on a 6t Sram Cells."

[13] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Low power process, voltage, and temperature (PVT) variations aware improved tunnel FET on 6T SRAM cells." Sustainable Computing: Informatics and Systems 21 (2019): 143-153.

[14] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Survey on improvement of PVT aware variations in tunnel FET on SRAM cells." In 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), pp. 703-705. IEEE, 2017

[15] Karne, R. K. ., & Sreeja, T. K. . (2023). PMLC- Predictions of Mobility and Transmission in a Lane-Based Cluster VANET Validated on Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5s), 477–483. https://doi.org/10.17762/ijritcc.v11i5s.7109

[16] Radha Krishna Karne and Dr. T. K. Sreeja (2022), A Novel Approach for Dynamic Stable Clustering in VANET Using Deep Learning (LSTM) Model. IJEER 10(4), 1092-1098. DOI: 10.37391/IJEER.100454.

[17] Reddy, Kallem Niranjan, and Pappu Venkata Yasoda Jayasree. "Low Power Strain and Dimension Aware SRAM Cell Design Using a New Tunnel FET and Domino Independent Logic." International Journal of Intelligent Engineering & Systems 11, no. 4 (2018).

[18] X. Kewei, B. Peng, Y. Jiang, and T. Lu, ``A hybrid deep learning model for online fraud detection,'' in *Proc. IEEE Int. Conf. Consum. Electron. Comput. Eng. (ICCECE)*, Jan. 2021, pp. 431_434.

[19] T. Vairam, S. Sarathambekai, S. Bhavadharani, A. K. Dharshini, N. N. Sri, and T. Sen, ``Evaluation of Naïve Bayes and voting classi_er algorithm for credit card fraud detection,'' in *Proc. 8th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2022, pp. 602_608.

[20] P. Verma and P. Tyagi, ``Analysis of supervised machine learning algorithms in the context of fraud detection,'' *ECS Trans.*, vol. 107, no. 1,p. 7189, 2022.

[21] J. Zou, J. Zhang, and P. Jiang, ``Credit card fraud detection using autoencoder neural network,'' 2019, *arXiv:1908.11553*.

[22] D. Almhaithawi, A. Jafar, and M. Aljnidi,``Example-dependent costsensitive credit cards fraud detection using SMOTE and Bayes minimumrisk,'' *Social Netw. Appl. Sci.*, vol. 2, no. 9, pp. 1_12, Sep. 2020.

[23] J. Cui, C. Yan, and C.Wang, ``Learning transaction cohesiveness for onlinepayment fraud detection,'' in *Proc. 2nd Int. Conf. Comput. Data Sci.*, Jan. 2021, pp. 1_5.

[24] M. Rakhshaninejad, M. Fathian, B. Amiri, and N. Yazdanjue,``An ensemble-based credit card fraud detection algorithm using anef_cient voting strategy,'' *Comput. J.*, vol. 65, no. 8, pp. 1998_2015, Aug. 2022.

[25] A. H. Victoria and G. Maragatham, ``Automatic tuning of hyperparameters using Bayesian optimization,'' *Evolving Syst.*, vol. 12, no. 1, pp. 217_223, Mar. 2021.