INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 05, May 2023, pp : 1293-1296

e-ISSN : 2583-1062

Impact Factor : 5.725

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

Mr M. Devandren[1]

[1]Assistant Professor, Dept of Computer Science and Engineering, Hindusthan College of Engineering and Technology, Coimbatore, Tamilnadu, India

DOI: https://www.doi.org/10.58257/IJPREMS31179

ABSTRACT

Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds.

Increase in fraud rates, researchers started using different machine learning methods to detect and analyse frauds in online transactions.

The main aim of the paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyse the past transaction details of the customers and extract the behavioural patterns. Where cardholders are clustered into different groups based on their transaction amount.

Then using sliding window strategy [1], to aggregate the transaction made by the cardholders from different groups so that the behavioural pattern of the groups can be extracted respectively.

Later different classifiers [3],[5],[6],[8] are trained over the groups separately. And then the classifier with better rating score can be chosen to be one of the best methods to predict frauds. Thus, followed by a feedback mechanism to solve the problem of concept drift [1]. In this paper,

we worked with European credit card fraud dataset.

1. INTRODUCTION

Credit cards are now the most preferred way for customers to transact either offline or online. There are a number of reasons, as illustrated below, due to which consumers are slowly shifting from debit card transactions to credit cards, especially in developing countries like India.

Lucrative cashback and reward point options are present for each credit card transaction. These are generally not offered by financial institutions for debit cards.

- Tie up credit cards with online and offline merchants, especially during festive seasons like Diwali, Eid, and Christmas, to offer further discounts on transactions. Several online merchants run their own promotional campaigns, which are tied up with credit cards—for example, Amazon Prime day, which happens at least once a year.

Immediate needs can be fulfilled (for example, medical emergencies, lifetime events, etc.) quickly instead of having sufficient account balance for the same. Most credit cards offer 0% EMI options, so it makes it all the more worth pursuing this goal.

Having a good credit history helps to build a good CIBIL score which then, in turn, helps customers to avail themselves better and competitive interest rates on longer-term needs like home loans or car loans.

Credit cards are tailored to suit individual customer needs. For example, customers who want to use a credit card for daily usage are usually offered a card with no annual fees or joining fees (marketed as lifetime free credit cards). On the other hand, we have premium cards with annual fees or joining fees for affluent people who offer golf membership, airport lounge access, seamless transactions at international and domestic merchants with lower transaction transfer fees, 5x to 10x reward points, etc.

However, with all these advantages, we also have the additional advantage of the ease of usage without having to carry currency around, and we can get a record of all our digital transactions through credit card statements far more easily compared with cash transactions or bank statements. One downside that has been witnessed over the past few years of this increasing digital phenomenon is the rise of fraud on the credit card. Fraud can be of several types, as we will try to understand a bit later on in this blog. You can also take up a free online credit card defaulter prediction course and enhance your knowledge about the same.

Before going into details of credit card fraud detection, let us try to understand the size of the overall credit card industry, especially in the large western economies like in some of the European countries, the US, and the UK. Below are some of the numbers around the credit card industry in general at a worldwide level.

There are 1.06 billion credit cards in use in America and 2.8 billion credit cards worldwide.

A US citizen, on average, has four active credit cards.

In the European Union (EU), the number of cards carried per person ranges from 0.8 to 3.9

In the UK, there were 32.3 million people with credit cards or charge cards in 2016, which roughly translates to 6 in every ten adults. The numbers have only grown since then from 2016 to now.

There were 368.92 billion card transactions worldwide in 2018. However, the average value per card payment is decreasing in most of the major economies, as a credit card is used more and more as a preferred financial product compared to other means. The average value per card payment drop indicates that customers are using a credit card more and more for daily use compared to one-off events like big purchases.

2. RELATED STUDIES

Authors implemented a credit card fraud detection system using several ML algorithms including logistic regression (LR), decision tree (DT), support vector machine (SVM) and random forest (RF). These classifiers were evaluated using a credit card fraud detection dataset generated from European cardholders in 2013. In this dataset, the ratio between non-fraudulent and fraudulent transactions is highly skewed; therefore, this is a highly imbalanced dataset. The researcher used the classification accuracy to assess the performance of each ML approach. The experimental outcomes showed that the LR, DT, SVM and RF obtained the following accuracy scores: 97.70%, 95.50%, 97.50% and 98.60%, respectively. Although these outcomes are good, the authors suggested that the implementation of advanced pre-processing techniques could have a positive impact on the performance of the classifiers.

Varmedja et al. [14] proposed a credit card fraud detection method using ML The authors used a credit card fraud dataset sourced from Kaggle [19]. This dataset contains transactions made within 2 days by European credit card holders. To deal with the class imbalance problem present in the dataset, the researcher implemented the Synthetic Minority Oversampling Technique (SMOTE) oversampling technique. The following ML methods were implemented to assess the efficacy of the proposed method: RF, NB, and multilayer perceptron (MLP). The experimental results demonstrated that the RF algorithm performed optimally with a fraud detection accuracy of 99.96%. The NB and the MLP methods obtained accuracy scores of 99.23% and 99.93%, respectively. The authors concede that more research should be conducted to implement a feature selection method that could improve on the accuracy of other ML methods.

Khatri et al. [15] conducted a performance analysis of ML techniques for credit card fraud detection. In this research, the authors considered the following ML approaches: DT, k-Nearest Neighbor (KNN), LR, RF and NB. To assess the performance of each ML method, the authors used a highly imbalanced dataset that was generated from European cardholders. One of the main performance metric that was used in the experiments is the precision which was obtained by each classifier. The experimental outcomes showed that the DT, KNN, LR, and RF obtained precisions of 85.11%, 91.11%, 87.5%, 89.77%, 6.52%, respectively.

Awoyemi et al. [16] presented a comparison analysis of different ML methods on the European cardholders credit card fraud dataset. In this research, the authors used an hybrid sampling technique to deal with the imbalanced nature of the dataset. The following ML were considered: NB, KNN, and LR. The experiments were carried out using a Python based ML framework. The accuracy was the main performance metric that was utilized to assess the effectiveness of each ML approach. The experimental results demonstrated that the NB, LR,and KNN achieved the following accuracies, respectively: 97.92%, 54.86%, and 97.69%. Although the NB and KNN performed relatively well, the authors did not explore the possibility to implement a feature selection method.

In ref. [4] the authors utilized several ML learning based methods to solve the issue of credit card fraud. In this work, the researchers used the European credit cardholder fraud dataset. To deal with the highly imbalanced nature of this dataset, the authors employed the SMOTE sampling technique. The following ML methods were considered: DT, LR, and Isolation Forest (IF). The accuracy was one of the main performance metrics that was considered. The results showed that the DT, LR, and IF obtained the accuracy scores of 97.08%, 97.18%, and 58.83%, respectively.

3. METHODOLOGY

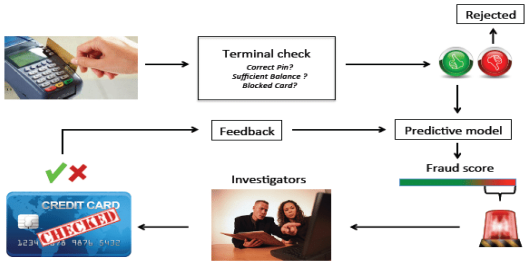
Dataset In this research, we use a dataset that includes credit card transactions that were made by European cardholders for 2 days in September 2013. This dataset contains 284807 transactions in total in which 0.172% of the transactions are fraudulent. The dataset has the following 30 features (V1,.., V28), Time and Amount. All the attributes within the dataset are numerical. The last column represents the class (type of transaction) whereby the value of 1 denotes a fraudulent transaction and the value of 0 otherwise. The features V1 to V28 are not named for data security and integrity reasons [19]. This dataset has been used in ref. [4, 13, 14, 16] and one of the key issues that we discovered is the low detection accuracy score that was obtained by those models because of the highly imbalanced nature of the dataset. In order to solve the issue of class imbalance, we applied the Synthetic Minority Oversampling Technique (SMOTE) method in the Data-Preprocessing phase of the proposed framework in Fig. 5 [18]. The SMOTE method works by picking samples that are close to each other within the feature space, drawing a

line between the data points in the feature space and creating a new instance of the minority class at a point along the line.

Feature selection

Feature selection (FS) is a crucial step when implementing machine learning methods. This is partly because the dataset used during the training and testing processes may have a large feature space that may negatively impact the overall performance of the models. The choice of which FS method to use depends on the kind of problem a researcher is trying to solve. The following paragraph provides an overview of instances where using a FS method improved on the performance of ML models.

Kasongo [20] implemented a GA-based FS in order to increase the performance of ML based models applied to the domain of intrusion detection systems. The results demonstrated that the application of GA improved the performance of the RF classifier with an Area Under the Curve (AUC) of 0.98. Mienye [21] et al. implemented a particle swarm optimization (PSO) technique to increase the performance of stacked sparse autoencoder network (SSAE) coupled with the softmax unit for heart disease prediction. The PSO technique was used to improve the feature learning capability of SSAE by optimally tuning its parameters. The results demonstrated that the PSO-SSAE achieved an accuracy of 97.3% on the Framingham heart disease dataset. Hemavathi et al. [22] implemented an effective FS method in an integrated environment using enhanced principal component analysis (EPCA). The results demonstrated that using the EPCA yields optimal results in supervised and unsupervised environments. Pouramirarsalani et al. [23] implemented a FS method using hybrid FS and GA for fraud detection in an e-banking environment. The experimental results demonstrated that using a FS method on a financial fraud datasets has a positive impact on the overall performance of the models that were used. In ref. [24], the authors implemented the GA-based FS method in conjunction with NB, SVM and RF algorithms for credit card fraud detection. The experimental output demonstrated that the RF yielded a better performance in comparison to the NB and SVM.



Genetic algorithm feature selection

The Genetic Algorithm (GA) is a type of Evolutionary inspired Algorithm (EA) that is often used to solve a number of optimization tasks with a reduced computational overhead. EAs generally possess the following attributes [25, 26]:

Population EAs approaches maintain a sample of possible solutions called population.

Fitness A solution within the population is called an individual. Each individual is characterized by a gene representation and a fitness measure.Variation The individual evolves through mutations that are inspired from the biological gene evolution.

In this study, the RF approach is used as the fitness method inside the GA. Further, the RF method is employed because it resolves the problem of over-fitting that is generally encountered when using regular Decision Trees (DTs). Moreover, RF performs well with both continuous and categorical attributes and RF are known to perform optimally on datasets that have a class imbalance problem. Additionally, the RF is a rule-based approach; therefore, the normalising of data is not required [27]. The alternative to the RF include tree-based ML algorithms such as Extra-Trees and Extreme Gradient Boosting [28, 29]. The fitness method is defined a function that receives a candidate solution (a feature vector) and determines whether it is fit or not. The measure of fitness is determined by the accuracy that is yielded by a particular attribute vector in the testing process of the RF method within the GA. Algorithm 1 provides more details about the implementation of RF in the GA. Algorithm 1 denotes the pseudo code implementation of the fitness function that was used in the GA. This algorithm consists of 6 main steps. In step 1, the data (20% of the full Credit Card Fraud dataset) is divided into a training ([Math Processing Error] and [Math Processing Error]) and testing ([Math Processing Error] and [Math Processing Error]) subsets. In Step 2, an instance of the RF classifier is instantiated. In Step 3, the RF instance is trained using the training set. In Step 4, the resulting model is then evaluated using the testing data [Math Processing Error]. In Step 5, the predictions are stored in [Math Processing Error]. In the last step, the evaluation process is conducted using [Math Processing Error]. During the evaluation procedure, the accuracy is used as the main performance metric. The most optimal model is one that yields the highest accuracy score.

4. CONCLUSION

Credit card fraud is most common problem resulting in loss of lot money for people and loss for some banks and credit card company. This project want to help the peoples from their wealth loss and also for the banked company and trying to develop the model which more eciently separate the fraud and fraud less transaction by using the time and amount feature in data set given in the Kegel. rst we build the model using some machine learning algorithms such as logistic regression, decision tree, support vector machine, this all are supervised machine learning algorithm in machine learning.

In feature solving this problem statement using another part of articial intelligence that is time series analysis, in our present project we used both and time and amount feature mainly for predicting the weather the transaction is fraud or Nonfraud transaction, in time series analysis we can reduce the number of parameters that is feature required for the model and we can achieve this model by using average method ,moving average or window method, naive method and sessional naive methods but all this method havesome advantages.

5. REFRENCES

[1] Changjun Jiang, et al.

[2] "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism."IEEE Internet of Things Journal, 5 (2018), pp. 3637-3647 View article CrossRefView in ScopusGoogle Scholar.

[3] Pumsirirat, A. and Yan, L. (2018). Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine. International Journal of Advanced Computer Science and Applications, 9(1).

[4] Mohammed, Emad, and Behrouz Far. "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study." IEEE Annals of the History of Computing, IEEE,1uly2018, ieee computer society. org/10.1109/IRI.2018.00025

[5] Kuldeep Randhawa, et al.

[6] "Credit Card Fraud Detection Using AdaBoost and Majority Voting" IEEE Access, 6 (2018), pp. 14277-14284 doi:10.1109/access. 2018.2806420 View article Cross Ref View in Scopus Google Scholar

[7] Roy, Abhimanyu, et al. "Deep Learning Detecting Fraud in Credit Card Transactions." 2018 Systems and Information Engineering Design Symposium (SIEDS), 2018, doi:10.1109/sieds.2018.8374722.

[8] Xuan, Shiyang, et al. "Random Forest for Credit Card Fraud Detection." 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018, doi:10.1109/icnsc.2018.8361343.

[9] Google Scholar Awoyemi, John O., et al. "Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis." 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017, doi:10.1109/iccni.2017.8123782.

[10] Google Scholar Melo-Acosta, German E., et al. "Fraud Detection in Big Data Using Supervised and Semi-Supervised Learning Techniques." 2017 IEEE Colombian Conference on Communications and Computing (COLCOM),2017doi:10.1109/ colcomcon.2017.80882https://www.ftc.gov/news-eventeleases/

[11] https://www.kaggle.com/mlg-ulb/creditcardf

[12] https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset.