Credit Card Fraud Detection - Machine Learning methods

Dejan Varmedja, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, Andras Anderla
Faculty of Technical Sciences
University of Novi Sad
Novi Sad, Serbia
varmedja@uns.ac.rs, mkaranovic@uns.ac.rs, sladojevic@uns.ac.rs, arsenovic@uns.ac.rs, andras@uns.ac.rs

Abstract— **Credit card fraud refers to the physical loss of credit card or loss of sensitive credit card information. Many machine-learning algorithms can be used for detection. This research shows several algorithms that can be used for classifying transactions as fraud or genuine one. Credit Card Fraud Detection dataset was used in the research. Because the dataset was highly imbalanced, SMOTE technique was used for oversampling. Further, feature selection was performed and dataset was split into two parts, training data and test data. The algorithms used in the experiment were Logistic Regression, Random Forest, Naive Bayes and Multilayer Perceptron. Results show that each algorithm can be used for credit card fraud detection with high accuracy. Proposed model can be used for detection of other irregularities.**

Keywords: Fraud; Logistic Regression; Multilayer Perceptron; Naive Bayes; Random Forest;

I. INTRODUCTION

There are growing number of new companies all around the world [1]. All of that companies are trying to provide best service quality for their customers. In order to succeed in that, companies are processing a lot of data on a daily basis. These data comes from vast number of resources and are in different formats. Moreover, this data contains some of the key parts of the company's future business. Because of that, companies need to store that data, to process it and what is really important, to keep it safe. Without securing data, a lot of it can be used by other companies or even worse, it can be stolen. In most cases, financial information is stolen, which can harm whole company or individual.

There are several types of frauds [2]. Check Fraud occurs when person forges a check or pays for something with check knowing that there is not enough money. Internet sales is fraud where fraudster sale fake items or counterfeit items, or taking payment without delivering the item. There are a couple more, such as charities fraud, identity theft, credit card fraud, debt elimination, Insurance fraud and others. Due to increasing popularity of cashless transactions, one of the most common frauds are credit card frauds. Credit card fraud refers to the situation where fraudster uses credit card for their needs while owner of that credit card is not aware of that. Fraudulent transactions conducted using credit cards acquired worldwide amounted to €1.8 billion in 2016 [3]. Although there is a tremendous volume increase in credit card transactions, the amount of frauds is proportionally the same or have decreased due to sophisticated fraud detection systems. However, fraudsters are constantly coming up with new ways to steal information [4].

There are two types of credit card frauds. One is theft of physical card, and other one is stealing sensitive information from the card, such as card number, cvv code, type of card and other. By stealing credit card information, a fraudster can broach a large amount of money or make a large amount of purchase before cardholder finds out. Because of that, companies use various machine learning methods to recognize which transactions are fraudulent and which are not.

The purpose of this paper is to analyze various machine-learning algorithms, such as Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB) and Multilayer Perceptron (MLP) in order to determine which algorithm is most suitable for credit card fraud detection.

The rest of the paper is structured as follows: in Section II researches that deal with specified problem are presented, Section III gives a brief description of the dataset which is used in the experiment, after which the results are presented in Section IV. Finally, concluding remarks are discussed in Section V, followed by a list of literature.

II. RELATED WORK

Fraudulent activities are causing major loss, which motivated researchers to find a solution that would detect and prevent frauds. Several methods have already been proposed and tested. Some of them are briefly reviewed below.

Classical algorithms such as Gradient Boosting (GB), Support Vector Machines (SVM), Decision Tree (DT), LR and RF proven useful. In paper [5] GB, LR, RD, SVM and a combination of certain classifiers was used, which led to high recall of over 91% on a European dataset. High precision and recall were achieved only after balancing the dataset by undersampling the data. In paper [6], European dataset was also used, and comparison was made between the models based on LR, DT and RF. Among the three models, RF proved to be the best, with accuracy of 95.5%, followed by DT with 94.3% and LR with accuracy of 90%.

The experiment was sponsored by the PanonIT company

978-1-5386-7073-6/19/$31.00 ©2019 IEEE

According to [7] and [8], k-Nearest neighbors (KNN) and outlier detection techniques can also be efficient in fraud detection. They are proven useful in minimizing false alarm rates and increasing fraud detection rate. KNN algorithm also performed well in experiment for paper [9], where the authors tested and compared it with other classical algorithms.

Unlike so far mentioned papers, in paper [10] a comparison was made between some classical algorithms and deep learning techniques. All of the tested techniques achieved accuracy of approximately 80%. Authors of paper [11], set side by side following algorithms: RF, GB, LR, SVM, DT, KNN, NB, XGBoost (XGB), MLP and stacking classifier (a combination of multiple machine learning classifiers), while using European dataset. As a result of thorough data preprocessing, all of the algorithms accomplished high accuracy of over 90%. Stacking classifier was most successful with accuracy of 95% and recall value of 95%. In paper [12], a neural network was tested on the European dataset. Experiment included back propagation neural network that was optimized with Whale algorithm. Neural network consisted of 2 input layers, 20 hidden and 2 output layers. Due to optimization algorithm, they achieved exceptional results on 500 test samples: 96.40% accuracy and 97.83% recall. Authors of paper [13] and [14] used neural networks, in order to demonstrate improvement in results when ensemble techniques are used. In paper [15] three datasets were used for comparison between Auto-encoder and Restricted Boltzmann Machine algorithms, which led to the conclusion that algorithms like MLP can be suitable for credit card fraud detection.

Numerous papers are focused on detecting fraudulent transactions using deep neural networks. However, these models are computationally expensive and perform better on larger datasets [16]. This approach may lead to great results, as we saw in some papers, but what if same results, or even better, can be achieved with less amount of resources? Our main goal is to show that different machine learning algorithms can give decent results with appropriate preprocessing. Authors of most of the mentioned paper used undersampling technique, and that was a motivation for using a different approach – oversampling technique.

Considering given facts, authors of this paper decided to compare the suitability of LR, RF, NB and MLP for credit card fraud detection. In order to achieve that, an experiment was conducted.

III. MATERIALS AND METHODS

A. Dataset

In this research the Credit Card Fraud Detection dataset was used, which can be downloaded from Kaggle [17]. This dataset contains transactions, occurred in two days, made in September 2013 by European cardholders.

The dataset contains 31 numerical features. Since some of the input variables contains financial information, the PCA transformation of these input variables were performed in order to keep these data anonymous. Three of the given features weren't transformed. Feature "Time" shows the time between first transaction and the every other transaction in the dataset. Feature "Amount" is the amount of the transactions made by credit card. Feature "Class" represents the label, and takes only 2 values: value 1 in case of fraud transaction and 0 otherwise.

Dataset contains 284,807 transactions where 492 transactions were frauds and the rest were genuine. Considering the numbers, we can see that this dataset is highly imbalanced, where only 0.173% of transactions are labeled as frauds.

Since distribution ratio of classes plays an important role in model accuracy and precision, preprocessing of the data is crucial.

B. Preprocessing

Feature selection is a fundamental technique, which selects the variables that are most relevant in the given dataset. Carefully choosing appropriate features and removing the less important one can reduce overfitting, improve accuracy and reduce training time. Visualization techniques can be helpful in that process. Feature selector tool [18] by Will Koehrsen was used in this experiment for that purpose. By using this tool it has been determined which features are the most important. Furthermore, features that do not contribute to the cumulative importance of 95% were removed. After the feature selection technique, 27 features were selected for additional experiment.

Machine learning algorithms have trouble learning when classification categories are not approximately equally distributed. Considering given data is highly imbalanced, it is necessary to perform some kind of balancing, so that model can be efficiently trained. Frequently used methods for adjusting the class distribution include undersampling the majority class, oversampling the minority class, or combination of those two. Synthetic Minority Oversampling Technique (SMOTE) is a popular oversampling method that has proven useful when used on imbalanced dataset [19], [20]. SMOTE was proposed method to improve random oversampling (Fig. 1).

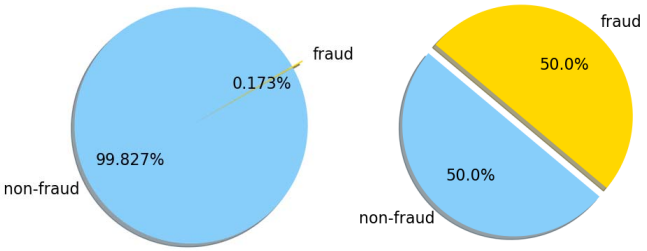

Figure 1. Class distribution before and after sampling

Many machine-learning algorithms expect the scale of the input. Taking into account that values of time and amount are highly varying, scaling is done in order to bring all features to the same level of magnitudes.

The experiment system environment is Windows 10 operating system, and the software operating environment is Spyder, scientific python development environment, which is part of the Anaconda platform. Used libraries include: numpy, pandas, matplotlib, sklearn and imblearn.

Previously mentioned algorithms used in the experiment are described in the following section.

C. Experiment

Logistic regression is one of the most popular classification algorithm in machine learning. The logistic regression model describes relationship between predictors that can be continuous, binary, and categorical. Dependent variable can be binary. Based on some predictors we predict whether something will happen or not. We estimate the probability of belonging to each category for a given set of predictors.

Naive Bayes is one of the supervised learning algorithms in which there are not dependencies between attributes. It's based on Bayes theorem. Depending of the type of distribution there are following algorithms: Gaussian distribution, Multinomial distribution, Bernoulli distribution. In this research, Bernoulli distribution is used for detecting fraud transactions.

Random forest is an algorithm that can be used in both classification and regression problems. It consists of many decision trees. This algorithm gives better results when there is higher number of trees in the forest and preventing model to overfitting. Each decision tree in forest gives some results. These results are merged together in order to get more accurate and stable prediction.

Multilayer perceptron is feedforward artificial neural network that consists of minimum 3 layers of nodes: input layer, hidden layer and output layer. Each node use activation function. Activation function calculates weighted sum of its inputs and adds bias. This allows us to decide which neuron should be removed and not considered in outside connections.

ANN used in the experiment consisted of 4 hidden layers with 50, 30, 30 and 50 units in each hidden layer, respectively, with relu activation function. It has been shown that deeper networks acquire better results than those with smaller number of layers [20]. Following this experience, we started with a smaller number of layers gradually increasing them in order to get acceptable results. Therefore, the best hyper-parameters were chosen based on exhaustive research. Further enhancing of the network caused greater computational time and obtained results didn't differ much from the chosen architecture. Weight optimization was accomplished with Adam, stochastic gradient-based optimizer.

Train and test set were split in 80:20 ratio and the model was updated through multiple epochs, based on tolerance for the optimization (TOL). When the loss or score is not improving by at least TOL for specified consecutive iterations, convergence is considered to be reached and training stops.

IV. Results and Discussion

To determine which algorithm is most suitable for the problem of detecting fraud transactions, different criteria for algorithm comparison have been used. Most used metrics for determining the results of machine learning algorithms are accuracy, recall and precision. All of the mentioned metrics can be calculated from a Confusion matrix.

Evaluation of a model's performance was made in accordance to these metrics. Models were tested both on original and over-sampled data and the results have shown that sampling is very important.

Since the test set consists of 20% of the whole dataset, total sum of samples is 56962. Of the total of 98 fraud transactions, LR model (Table 1) achieved:

- precision: 58.82%,
- recall: 91.84%,
- accuracy: 97.46%.

TABLE 1:CONFUSION MATRIX FOR LR

		Predicted	
		0	1
Actual	0	55424	1440
	1	8	90

NB model obtained following results (Table 2):

- precision: 16.17%,
- recall: 82.65%,
- accuracy: 99.23%

TABLE 2:CONFUSION MATRIX FOR NB

		Predicted	
		0	1
Actual	0	56444	420
	1	17	81

RF model obtained following results (Table 3):

- precision: 96.38%,
- recall: 81.63%,
- accuracy: 99.96%.

TABLE 3:CONFUSION MATRIX FOR RF

		Predicted	
		0	1
Actual	0	56861	3
	1	18	80

MLP model obtained following results (Table 4):

- precision: 79.21%,
- recall: 81.63%,
- accuracy: 99.93%

TABLE 4:CONFUSION MATRIX FOR MLP

		Predicted	
		0	1
Actual	0	56843	21
	1	18	80

By analyzing obtained results, it is obvious that accuracy is extremely high, although that doesn't mean that results are perfect. Accuracy must be taken "with a grain of salt" – desirably it should be interpreted in combination with some other metrics. According to given results, it is shown that classic algorithm, like RF can give similar results as a simple neural network.

Comparison of the obtained results with results achieved in researches on the same dataset, with classical algorithms [5] and [8], show that oversampling the data can improve fraud detection rate. Like in papers [10] and [11], it is proven that classical algorithms can be as successful as deep learning algorithms. Although papers [12] and [15] represent deep learning algorithms as optimal for this type of problems, it should be decided according to the situation which of these should be used. For example, deep networks work better with more data and can be adapted to different domains more easily than classical algorithms. On the other hand, if there is not much data, it is probably better to work with classical algorithms. These algorithms are also easier to interpret and cheaper, both in financial and computational sense [21].

V. CONCLUSION

Credit card frauds represent a very serious business problem. These frauds can lead to huge losses, both business and personal. Because of that, companies invest more and more money in developing new ideas and ways that will help to detect and prevent frauds.

The main goal of this paper was to compare certain machine learning algorithms for detection of fraudulent transactions. Hence, comparison was made and it was established that Random Forest algorithm gives the best results i.e. best classifies whether transactions are fraud or not. This was established using different metrics, such as recall, accuracy and precision. For this kind of problem, it is important to have recall with high value. Feature selection and balancing of the dataset have shown to be extremely important in achieving significant results.

Further research should focus on different machine learning algorithms such as genetic algorithms, and different types of stacked classifiers, alongside with extensive feature selection to get better results.

ACKNOWLEDGMENT

This work has been funded by the SENSors and Intelligence in BuLt Environment (SENSIBLE) project with Grant agreement ID: 734331. Authors would also like to thank the PanonIT company for their support.

REFERENCES

[1] Global Facts (2019). Topic: Startups worldwide. [online] Available at: https://www.statista.com/topics/4733/startups-worldwide/ [Accessed 10 Jan. 2019].

[2] Legal Dictionary (2019). Fraud - Definition, Meaning, Types, Examples of fraudulent activity. [online] Available at: https://legaldictionary.net/fraud/ [Accessed 15 Jan. 2019].

[3] European Central Bank (2018). Fifth report on card fraud, September 2018. [online]. Available at: https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport201809.en.html#toc1 [Accessed 21 Jan. 2019].

[4] En.wikipedia.org. (2019). Credit card fraud. [online] Available at: https://en.wikipedia.org/wiki/Credit_card_fraud [Accessed 24 Jan. 2019].

[5] A. Mishra, C. Ghorpade, "Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques" 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) pp. 1-5. IEEE.

[6] S. V. S. S. Lakshmi, S. D. Kavilla "Machine Learning For Credit Card Fraud Detection System", unpublished

[7] N. Malini, Dr. M. Pushpa, "Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection", Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017 Third International Conference on pp. 255-258. IEEE.

[8] Mrs. C. Navamani, M. Phil, S. Krishnan, "Credit Card Nearest Neighbor Based Outlier Detection Techniques"

[9] J. O. Awoyemi, A. O. Adentumbi, S. A. Oluwadare, "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis", Computing Networking and Informatics (ICCNI), 2017 International Conference on pp. 1-9. IEEE.

[10] Z. Kazemi, H. Zarrabi, "Using deep networks for fraud detection in the credit card transactions", Knowledge-Based Engineering and Innovation (KBEI), 2017 IEEE 4th International Conference on pp. 630-633. IEEE.

[11] S. Dhankhad, B. Far, E. A. Mohammed, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study", 2018 IEEE International Conference on Information Reuse and Integration (IRI) pp. 122-125. IEEE.

[12] C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai, S. Pan, "Credit card fraud detection based on whale algorithm optimized BP neural network", 2018 13th International Conference on Computer Science & Education (ICCSE) pp. 1-4. IEEE.

[13] N. Kalaiselvi, S. Rajalakshmi, J. Padmavathi, "Credit card fraud detection using learning to rank approach", 2018 Internat2018

International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC) ional conference on computation of power, energy, Information and Communication (ICCPEIC) pp. 191-196. IEEE

[14] F. Ghobadi, M. Rohani, "Cost Sensitive Modeling of Credit Card Fraud using Neural Network strategy", 2016 Signal Processing and Intelligent Systems (ICSPIS), International Conference of pp. 1-5. IEEE.

[15] A. Pumsirirat, L. Yan, "Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine", 2018 International journal of advanced computer science and applications, 9(1), pp. 18-25

[16] Learning – Towards Data Science. [online] Available at: https://towardsdatascience.com/deep-learning-vs-classical-machine-learning-9a42c6d48aa [Accessed 19 Jan. 2019].

[17] Kaggle.com. (2019). Credit Card Fraud Detection. [online] Available at: https://www.kaggle.com/mlg-ulb/creditcardfraud [Accessed 10 Jan. 2019].

[18] Github (2019). Feature selector. [online] Available at: https://github.com/WillKoehrsen/feature-selector [Accessed 18 Jan. 2019].

[19] Garćıa, Salvador and Nitesh V. Chawla. "SMOTE for Learning from Imbalanced Data : Progress and Challenges, Marking the 15-year Anniversary." (2018), Journal of Artificial Intelligence Research, 61, pp. 863-905.

[20] J. Wang, M. Xu, H. Wang and J. Zhang, "Classification of Imbalanced Data by Using the SMOTE Algorithm and Locally Linear Embedding", Signal Processing, 2006 8th International Conference on (Vol. 3). IEEE. 2006 8th international Conference on Signal Processing, Beijing, 2006

[21] Deeplearningbook.org. (2019). Deep Learning. [online] Available at: https://www.deeplearningbook.org/ [Accessed 11 Jan. 2019].