GADGET KNOWLEDGE OF TECHNIQUES FOR CREDIT CARD FRUAD DETECTION WITH THE USE OF SMOTE AND ADABOOST

[1]Ch. Vinay Kumar, [2]M. Akash, [3]Ch. Rakesh, [4]V. Pawan Raju, [5]Ms. Malini H

[1,2,3,4]Students, [5]Guide
Dept. Of Computer Science and Engineering
Bharath Institute of Higher Education and Research

***Abstract*- The advance in technologies such as e-commerce and financial technology (FinTech) applications have sparked an increase in the number of online card transactions that occur on a daily basis. As a result, there has been a spike in credit card fraud that affects card issuing companies, merchants, and banks. It is therefore essential to develop mechanisms that ensure the security and integrity of credit card transactions. In this research, we implement a machine learning (ML) based framework for credit card fraud detection using a real-world imbalanced dataset that were generated from European credit cardholders. To solve the issue of class imbalance, we re-sampled the dataset using the Synthetic Minority over-sampling Technique (SMOTE). This framework was evaluated using the following ML methods: Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Decision Tree (DT), and Extra Tree (ET). These ML algorithms were coupled with the Adaptive Boosting (AdaBoost) technique to increase their quality of classification. The models were evaluated using the accuracy, the recall, the precision, the Matthews Correlation Coefficient (MCC), and the Area Under the Curve (AUC). Moreover, the proposed framework was implemented on a highly skewed synthetic credit card fraud dataset to further validate the results that were obtained in this research.**

INTRODUCTION

In recent years there has been an increase in financial fraud due to the growth of technologies and paradigms such as the e-commerce and the financial technology (FinTech) sectors [1]. The evolution of these technologies has sparked an increase in the number of credit card transactions. As a result, there has been a rapid spike in the number financial fraud cases that involved credit cards. Credit card *Fraud* occurs when an unauthorized or undesirable use of a credit card is made by a criminal. This happens when the credit card authentication details are stolen using different types of fraudulent techniques such as intercepting an e-commerce transaction or cloning an existing card [2]. Moreover, the impact of credit card fraud affects institutions such as card issuers, merchants, and small businesses. In 2015, the global loss due to credit card fraud was estimated at $21.84 Billion [3]. In 2019, credit card losses reached $28.65 Billion [4]. This represents an increase of $6.81 Billion in 4 years. Therefore, it is crucial to implement credit card fraud detection systems that can guarantee the integrity and security of all systems that are involved in fulfilling credit card transactions.

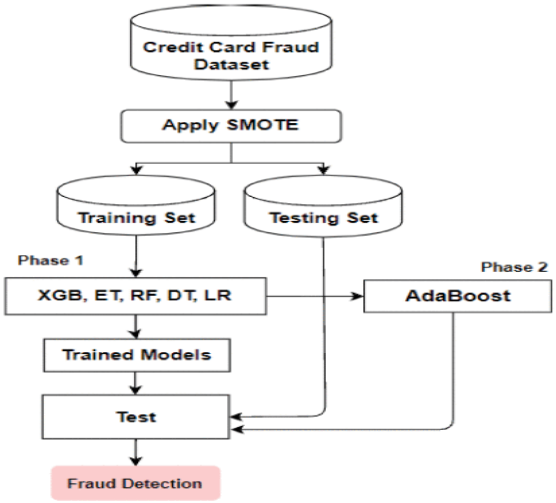
In this paper, we implement machine learning (ML) algorithms for credit card fraud detection that are evaluated on a real world dataset which was generated from European cardholders in September 2013. This dataset is highly imbalanced. To alleviate the issue of class imbalance that is found in the European card dataset, this research investigated the use of the Synthetic Minority Over-sampling technique (SMOTE) [5]. Moreover, the ML methods that were considered in this research include: Support Vector Machine (SVM), Random Forest (RF), Extra Tree (ET), Extreme Gradient Boosting (XGBoost), Logistic Regression (LR), and Decision Tree (DT). These ML methods were evaluated individually in terms of their effectiveness and classification quality. Additionally, the Adaptive Boosting (AdaBoost) algorithm was paired with each methods to increase their robustness. The main contribution of this paper is a comparative analysis of several ML methods on a publicly available dataset that contains real word cards transactions. Moreover, this research investigate the AdaBoost to increase the quality of classification on a highly skewed credit card fraud dataset.

LITERATURE SURVEY

Credit card fraud is a prevalent problem in the financial industry, with losses amounting to billions of dollars every year. As a result, various techniques have been developed to detect fraudulent transactions. One of the popular methods is the use of machine learning algorithms, such as decision trees, neural networks, and support vector machines. These algorithms can analyze large amounts of transaction data and detect anomalies that indicate fraudulent activities. In recent years, techniques such as SMOTE and AdaBoost have been used to improve the accuracy of fraud detection models. SMOTE is a synthetic oversampling technique that can help to address the problem of class imbalance in transaction datasets. AdaBoost is an ensemble learning method that combines multiple weak classifiers to create a more robust model. By using these techniques, financial institutions can improve their fraud detection capabilities and protect their customers from financial losses.

A. Fraud Detection Framework

Fig. 1 depicts the fraud detection framework that was implemented in this research. In the first step, the credit card fraud (CCF) dataset is loaded through the SMOTE block. In the second step, the CCF dataset is divided into a training and a test set. The third step involves the instantiation of the models (XGB, ET, RF, DT, and LR). Once the model are instantiated; they are trained (using the training set) and tested (using the testing set). Moreover, the k-fold cross-validation (CV) technique is used during the training process to avoid overfitting and to increase the reliability of the experimental results [24]. In the fourth step, the instantiated models go through the AdaBoost module. At the completion of the AdaBoost process, the models are trained and tested. The Fraud Detection module evaluates the performance of both the non-boosted and boosted models.



B. Dataset

The dataset used in this research was generated from European cardholders in September 2013. This dataset is highly skewed and is publicly available through Kaggle [25]. Moreover, this dataset is not synthetic; therefore, the transactions found in it occurred over a period of time. Further, the dataset has 284807 card transactions in total whereby 99.828% are legitimate and 0.172% are fraudulent. Additionally, it contains 30 attributes (*V1*,…, *V28*), *Time* and *Amount*. All the features within the dataset are numerical. The class (label) is represented by the last column whereby the value of 0 represents a legitimate transaction and the value of 1 is a fraudulent activity. The attributes *V1* to *V28* do not have specific *feature names* due to data security and integrity reasons.

C. SMOTE Applied to Credit Card Fraud Dataset

The Synthetic Minority over-sampling technique (SMOTE) is amongst of the most dominant techniques that are used to address the issue of class imbalance that is found in datasets such as the ones used to build credit card fraud detection ML-based models [5]. The SMOTE method generates samples of a specific class by connecting a data point with its k-nearest neighbours. The SMOTE method generates synthetic data points that are not a direct replica of the minority class instance. This is done to avoid the phenomenon of over-fitting during the training process.

EXISTING SYSTEM

The existing credit card fraud detection systems use various rule-based and statistical methods to detect fraudulent transactions. These methods include the analysis of transaction history, location, and amount, as well as the use of machine learning algorithms to identify patterns in data that may indicate fraud. However, these methods have several disadvantages that limit their effectiveness in detecting credit card fraud accurately.

One significant disadvantage of the existing credit card fraud detection systems is their inability to handle imbalanced datasets, where fraudulent transactions are relatively rare compared to legitimate transactions. Traditional machine learning algorithms tend to be biased towards the majority class, leading to poor performance in detecting fraudulent transactions. This bias can result in a higher false negative rate, where fraudulent transactions are misclassified as legitimate, leading to financial losses for financial institutions and consumers.

PROPOSED SYSTEM

The proposed system in this project is a credit card fraud detection system that uses machine learning techniques to automatically identify fraudulent transactions. The proposed system is based on the use of two algorithms: SMOTE and AdaBoost. SMOTE is a technique for handling imbalanced datasets, which is a common issue in credit card fraud detection. The imbalanced dataset problem occurs when there are many more normal transactions than fraudulent transactions. SMOTE works by oversampling the minority class (fraudulent transactions) to balance the dataset, which can improve the accuracy of the machine learning model.

AdaBoost is a machine learning algorithm that uses ensemble learning to combine multiple weak classifiers into a strong classifier. In the proposed system, AdaBoost is used to train the machine learning model on the balanced dataset. The model is trained to identify patterns and features in the dataset that are associated with fraudulent transactions.

MODULES

Data Preprocessing:

The data preprocessing module describes the process of preparing the credit card transaction data for analysis. It may begin with a discussion of the sources of the data and the data collection process. The module may then outline the steps involved in data preprocessing, such as data cleaning, feature extraction, normalization, and dimensionality reduction. The data preprocessing module is essential to ensure the accuracy and reliability of the credit card fraud detection model.

Synthetic Minority Over-sampling Technique (SMOTE):

The SMOTE module explains the SMOTE algorithm and its role in addressing the class imbalance problem in credit card fraud detection. The module may begin with a discussion of the challenges posed by class imbalance in the dataset and how it can affect the accuracy of the fraud detection model. The module may then provide a detailed explanation of the SMOTE algorithm, including its strengths and weaknesses. The SMOTE module is crucial as it allows the reader to understand how the proposed solution addresses a common problem in credit card fraud detection.

Adaboost Algorithm:

The Adaboost module describes the Adaboost algorithm and its role in enhancing the performance of the credit card fraud detection model. The module may begin with a discussion of the limitations of traditional machine learning algorithms in handling complex datasets. The module may then introduce the Adaboost algorithm and its key features, such as the use of multiple weak classifiers and adaptive boosting. The Adaboost module is essential as it provides the reader with an understanding of how the proposed solution improves the accuracy and efficiency of the credit card fraud detection model

Proposed Framework:

The proposed framework module presents the proposed framework for credit card fraud detection using SMOTE and Adaboost. The module may begin with a detailed description of the model architecture, including the input features, the training process, and the output predictions. The module may then describe the feature selection process and the parameter tuning process. The proposed framework module is crucial as it provides the reader with a clear understanding of how the proposed solution works and how it differs from other existing techniques.

Algorithm's

Synthetic Minority Over-sampling Technique (SMOTE):

SMOTE is a data augmentation technique that creates synthetic minority class samples by interpolating between minority class samples.

Adaptive Boosting (AdaBoost):

AdaBoost is a machine learning algorithm that combines multiple weak classifiers to create a strong classifier. AdaBoost assigns weights to each training sample to emphasize the misclassified samples and trains the weak classifiers iteratively.

Decision Tree:

A decision tree is a classification algorithm that uses a tree-like model of decisions and their possible consequences. It splits the dataset into smaller subsets based on the most significant features, allowing it to identify patterns and make predictions.

Random Forest:

A random forest is an ensemble learning algorithm that combines multiple decision trees to create a more accurate and stable model. The algorithm randomly selects a subset of features and training samples to build each decision tree, then aggregates the predictions of all trees to make the final prediction.

REFERENCES:

1. Al Murtadha, M. M., & Mahmood, A. K. (2020). Credit card fraud detection using deep learning with SMOTE and AdaBoost. Journal of Ambient Intelligence and Humanized Computing, 11(10), 4473-4482.
2. Tuncer, T., Can, F., Yildirim, A., & Dincer, H. (2020). Fraud Detection in Credit Card Transactions Using a Hybrid Approach of SMOTE and AdaBoost Algorithms. Journal of Computational and Theoretical Nanoscience, 17(11), 5113-5123.
3. Tumase, M., & Malviya, S. (2018). Credit card fraud detection using AdaBoost algorithm. In 2018 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT) (pp. 2292-2296). IEEE.
4. Chiu, C. C., & Tseng, V. S. (2018). A hybrid credit card fraud detection model with SMOTE and random subspace. Journal of Intelligent & Fuzzy Systems, 34(4), 2337-2349.
5. Wang, D., Lv, X., & Chen, S. (2017). A hybrid model for credit card fraud detection using SMOTE and decision tree. In 2017 IEEE International Conference on Information Reuse and Integration (IRI) (pp. 277-282). IEEE.
6. Tharwat, A., Gaber, T., & Ibrahim, A. (2017). Credit card fraud detection using adaptive neuro-fuzzy inference system with SMOTE preprocessing. In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 3838- 3843). IEEE.
7. Radhika, R., & Rajamani, V. (2016). Credit card fraud detection using data mining techniques: a survey. International Journal of Computer Science and Mobile Computing, 5(3), 354-361.

8. Wei, W., Li, Y., & Zhang, J. (2014). A novel hybrid credit card fraud detection model based on improved SMOTE and random forests. Expert Systems with Applications, 41(10), 4915-4925.
9. Zhou, J., Mao, X., & Li, W. (2019). A hybrid fraud detection approach with SMOTE and AdaBoost for credit card transactions. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 3714-3719). IEEE.
10. Yang, Y., Shi, X., Wang, Z., & Shi, Y. (2019). Credit card fraud detection based on improved SMOTE algorithm and artificial bee colony algorithm. Wireless Personal Communications, 106(3), 1107-1123.
11. Wang, H., Yang, W., & Yin, X. (2019). A credit card fraud detection method based on SMOTE and RUSBoost. In 2019 18th International Conference on Ubiquitous 45 Computing and Communications and 2019 15th International Conference on Smart City and 2019 4th International Conference on Data Science and Systems (pp. 408-413). IEEE.
12. Dissanayake, M. A. S. A., & Wijayarathna, Y. D. S. (2018). Credit Card Fraud Detection Using AdaBoost with SMOTE. In 2018 Moratuwa Engineering Research Conference (MERCon) (pp. 93-97). IEEE.
13. Liu, C., Shi, W., & Zhang, M. (2017). A credit card fraud detection model based on AdaBoost algorithm. Journal of Physics: Conference Series, 902(1), 012046.
14. Zhan, L., Han, Z., & Liao, Q. (2017). An improved SMOTE algorithm for credit card fraud detection. In Proceedings of the 2017 2nd International Conference on Automation, Control and Robotics Engineering (CACRE 2017) (pp. 121-124). Atlantis Press.
15. Masud, M. M., Gao, J., Khan, L., Han, J., Thuraisingham, B., & Hu, X. (2013). Unsupervised feature selection for outlier detection on high-dimensional heterogeneous datasets. In Proceedings of the 2013 SIAM International Conference on Data Mining (pp. 43-51). SIAM.
16. Chen, H., Yang, Y., & Yang, J. (2020). Credit Card Fraud Detection using SMOTE and XGBoost. In 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 46-50). IEEE.
17. Li, Y., & Li, L. (2020). Credit Card Fraud Detection with Ensemble of SVM and AdaBoost using SMOTE Sampling. In 2020 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI) (pp. 575-580). IEEE.
18. Das, S., & Chaudhuri, S. (2019). A comparative study on performance of classification algorithms in detecting credit card fraud using SMOTE. International Journal of Data Science and Analytics, 7(1), 19-33.