Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 8, Issue 02, February -2021

Review of smart credit card fraud detection approach using DM techniques

Payal Boda[1], Prof. Dixita Kagathara[2]

[1]M.E. Scholar, Computer Engineering, Darshan Institute of Engineering & Technology, Rajkot
[2] Assistant Professor, Computer Engineering, Darshan Institute of Engineering & Technology, Rajkot

Abstract — *Due to the surge of intrigued in online retailing, the utilize of credit cards has been quickly extended in later a long time. Taking the card details to perform online exchanges, which is called extortion, has moreover seen more habitually. Preventive arrangements and moment extortion location methods are broadly considered due to basic monetary misfortunes in numerous industries. In this work, Naïve Bayes, D-TREE and Multiple Additive Regression Tree Classifier (MART) show for the detection of credit card fakes on the spilling transactions is explored with the utilize of diverse qualities of card transactions. I am applying Naïve Bayes, D-TREE and Multiple Additive Regression Tree Classifier (MART) algorithm to detect the CC fraud then compare the result with all algorithms for getting higher CC fraud accuracy.*

Keywords*- Data mining, Credit card fraud, Fraud detection, Naïve Bayes, D-TREE, Multiple Additive Regression Tree Classifier*

I. INTRODUCTION

Data mining aims to extract useful information from huge amount of data. In the process of data mining, large data sets are first sorted, then patterns are identified and relationships are established to perform data analysis and solve problems. The credit card payment system is one of the simplest payment methods and the most common type of financial transaction. However, it is observed that a good number of fraudulent credit card transactions are occurring. Credit card fraud means the unauthorized use of a credit card account. Thus, the fraud occurs when the third party starts unauthorized using of the credit card without the consent of the card owner. These cards can be used for making purchases in both online and offline modes. Online credit purchases need customers to endorse payments by showing at the point of sale their personal identity numbers (PINs), while offline transactions need customers to sign purchase receipts. The method of Credit Card Fraud Detection (CCFD) mainly involves separating fraudulent financial details from genuine data. The models help to classify the pattern of fraud in the databases by applying machine learning algorithms. Various problems are associated with credit card fraud detection and hinder the direction of fraud detection, such as non-availability of real dataset, size of dataset, determining the appropriate evaluation parameters and complex actions of the fraudsters. Generally, credit card fraud is classified as three types: Traditional card related frauds, Merchant related frauds and Internet frauds. Traditional card related frauds include application fraud, lost and stolen cards, account takeover and fake cards. Merchant related frauds include merchant collusion and triangulation. Internet frauds include site cloning, card generators and false merchant sites. In all these categories of fraud, the fraudsters obtain the information of the card without the knowledge of the cardholders and then use them for various fraudulent activities to steal the money from the account.

II. LITERATURE SURVEY

According to [1], the use of various card transaction attributes is searched with the Gradient Boosting Tree (GBT) model of the credit card fraud detection in real time on streaming Card-Not-Present transactions (CNP). To form a feature vector to be used as a training example, numerical, hand-crafted numerical, categorical, and textual attributes are merged. In this research, the main focus is two points such as character-level word embedding and sliding window-based automated training dataset generation technique. Character-level word embedding is necessary to map the name to a vector of real numbers and the name of the merchant can be used as a unique feature to detect fraudulent behavior. The sliding window-based automated training dataset generation technique is retrained by itself over time to prevent concept drift adaptively. Three features used in these experiment metrics like that encoded features, aggregated and encoded Feature, embedding, and aggregated and encoded feature. Experiments are evaluated by the following metrics; the False-Positive Rate (FPR), recall, precision, Area Under Curve (AUC). Sliding the training set increase the fraud detection performance in terms of AUC by 0.028%, and in fixed 0.3 FPR, Recall is improved by 0.029%.

According to [2], the behavior of frauds and legitimate transactions change constantly. Also, the issue with the credit card data is that it is highly skewed which leads to an inefficient prediction of fraudulent transactions. The three various proportions of datasets were used in this study and the random under-sampling technique was used for skewed dataset. The three machine learning algorithms used in this work are Logistic Regression, Naïve Bayes, and K-Nearest Neighbour. The performance of these algorithms is recorded and analyze that how accurately they differentiate and classify the fraud and non-fraud transactions of the credit card dataset with random under sampling method (RUS) and to check out if the performance is improved or not. The analysis is done in python and the performance of the algorithms is calculated based on precision, sensitivity, specificity, accuracy, F-measurement, and area under curve. On the basis these measurements Logistic Regression (LR) showed the optimal performance for all the data proportions as compared to Naïve Bayes (NB) and K-Nearest Neighbour (KNN).

According to [3], in a relatively small-time frame, which can range from micro to milliseconds, the mechanism of acceptance or denial of a transaction occurs and a large number of related forms of transactions occur at the same time. Therefore, to be able to distinguish between a legitimate and a fraud transaction, an appropriate Fraud Detection Mechanism must be put into work. In this paper, they used an imbalanced dataset to check the suitability of various supervised machine learning models to forecast the probability of occurrence of a fraudulent transaction. They used sensitivity, precision, and time as the deciding parameters to come to a clear conclusion. Accuracy has not been used as a parameter as it is not susceptible to imbalanced data and does not have a definitive answer. They used kNN, Naive Bayes, Decision Tree, Logistic Regression and Random Forest models for predicting the chances of occurrence of a fraudulent credit card transaction out of a given number of transactions and the analysis shows that the sensitivity of the kNN model is greater than that of Decision tree, but as time taken by kNN for testing the data is very large. To ensure that minimal time is required for prediction in the case of fraud identification, so the preferred model is the Decision Tree.

According to [4], classifier ensembles are used successfully in either data mining or data stream mining to increase the performance of single classifiers. This paper therefore proposes an Online Boosting (OLBoost) approach, which first uses the Extremely Fast Decision Tree (EFDT) as a base (weak) learner, in order to assemble them into a single strong online learner, to achieve great success in prediction with virtually no increasing memory and time costs.

According to [5], sometimes the learning models used by them are too weak to fit the large scale of data. This paper extends the fraud detection method and uses lightgbm to propose a detection algorithm. Used by many data scientists to achieve state-of-the-art results to solve many machine learning problems, it is a scalable end-to-end tree boosting method. It also implemented other classical machine learning models in this task like SVM, logistic Regression, and Xgboost. They used it to tune some key parameters like learning rate, number of estimators, a sample rate of rows, sample rate of columns, max depth of each tree, and boosting types. Experiments showed that the lightgbm model outperformed the other Logistic Regression, SVM, and Xgboost models on both Auc-Roc score and accuracy.

According to [6], this paper proposes an intelligent approach using an optimized light gradient boosting machine (OLightGBM) to detect fraud in credit card transactions. A Bayesian-based hyperparameter optimization algorithm is intelligently implemented in the proposed approach to change the parameters of a light gradient boosting machine (LightGBM). Experiments were carried out using two real-world public credit card transaction data sets consisting of fraudulent transactions and legitimate ones to demonstrate the efficacy of our proposed OLightGBM for detecting fraud in credit card transactions. The output of the proposed method was evaluated by comparison with other research findings and state-of-the-art machine learning algorithms, including random forest, logistic regression, the radial support vector machine, the linear support vector machine, kNN, decision tree, and naive bayes, based on a comparison with other approaches using the two data sets. The experimental results show that the method proposed outperformed the other machine learning algorithms and obtained the highest accuracy performance (98.40%), Area under receiver operating characteristic curve (AUC) (92.88%), Precision (97.34%), and F1-score (56.95%).

According to [7], the information of the credit card fraud detection are highly skewed or unbalanced. Due to this class imbalance problem, several generic machine learning algorithms such as Random Forest (RF), Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Logistic Regression (LR) have been applied to a balanced dataset with different sampling techniques such as Oversampling, Undersampling, Both sampling, ROSE and SMOTE. In reality, these algorithms have been effective in correctly predicting non-fraudulent transactions rather than fraudulent ones. The minority class of fraudulent exchanges is ignored as noise-based exchanges. The work effectively handles this misclassification by altering the raw data provided to the classification algorithms to correctly forecast the minority class, not the majority class. According to the experiments, SMOTE sampling is better conducted on the basis of its system of synthetic sampling rather than the nearest values. Among the five methods used, logistic regression performs well with 97.04% accuracy and 99.99% precision. SMOTE sampling along with logistic regression is also recommended for the detection of credit card fraud.

According to [8], five standard machine learning classification models from heterogeneous family such as K-Nearest Neighbor (K-NN), Extreme Learning Machine (ELM), Random Forest (RF), Multilayer Perceptron (MLP), and Bagging classifier are investigated and compared with each other. An ensemble of machine learning algorithms with majority voting yields a better hybridized model that can correctly classify the fraudulent and non-fraudulent transactions. An ensemble of the machine learning algorithm is one of the novel approach for the credit card fraud detection technique. Performance parameters are measured for accuracy, sensitivity, specificity, precision, F1-Score, and Matthews Correlation Coefficient. The prediction accuracy of the proposed model is observed to be 83.83%, which is significantly improved as compared to other single classification models. The error of detection of fraud for the proposed model is decreased and the rate of prediction of fraud is increased.

III. PROPOSED APPROACH

PROPOSED FLOWCHART:

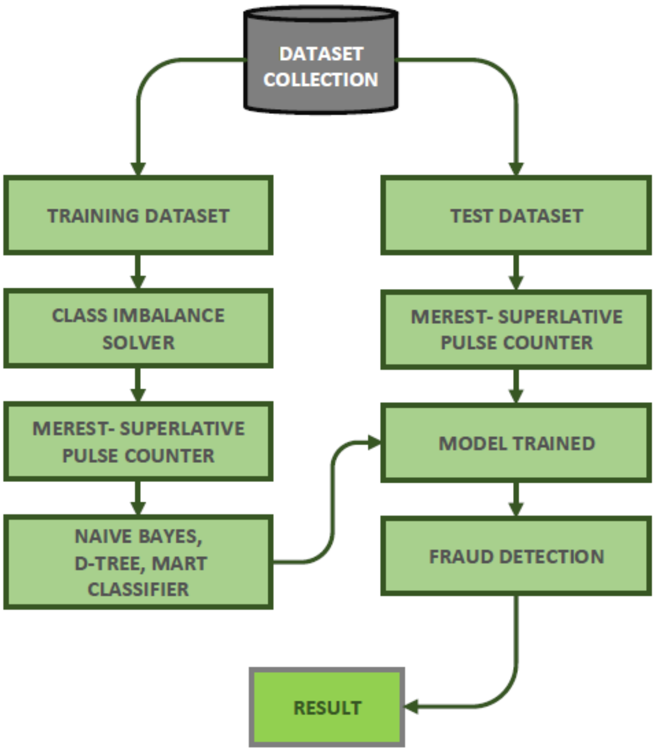
DATASET COLLECTION
TRAINING DATASET
TEST DATASET
CLASS IMBALANCE SOLVER
MEREST- SUPERLATIVE PULSE COUNTER
MEREST- SUPERLATIVE PULSE COUNTER
MODEL TRAINED
NAIVE BAYES, D-TREE, MART CLASSIFIER
FRAUD DETECTION
RESULT

Figure 1. Proposed Flowchart Method

PROPOSED ALGORITHM:
BEGIN
Step 1: Take input from Dataset
Step 2: Data-preprocessing from Dataset
Step 3: Divide Training and Testing data from Dataset
Step 4: Utilize class imbalance solver technique on Dataset
Step 5: Apply Merest – Superlative pulse counter on Dataset
Step 6: Train Model using Naïve Bayes, D-TREE and Multiple Additive Regression Tree (MART) Classifier algorithm
Step 7: Model Trained
Step 8: Fraud Detection
Step 9: Result
End

IV. FUTURE WORK

In future, I am going to use assist by applying Multiple Additive Regression Tree (MART) classifier algorithms and machine learning calculations over the preparing information and comparing their accuracies so that I can conclude the result.

V. CONCLUSION

In this research, I have prepared proposed architecture and studied various classification methods such as Naïve Bayes, D-TREE, and Multiple Additive Regression Tree (MART) Classifier for implementation to get better results and accuracy among all algorithms.

VI. REFERENCES

[1] Ali Ye，silkanat(B), Barı，s Bayram, Bilge K¨oro˘glu, and Se，cil Arslan, “An Adaptive Approach on Credit Card Fraud Detection Using Transaction Aggregation and Word Embeddings” © IFIP International Federation for Information Processing 2020.

[2] Fayaz Itoo, Meenakshi, Satwinder Singh, “Comparison and analysis of logistic regression, Naı¨ve Bayes and KNN machine learning algorithms for credit card fraud detection” © Bharati Vidyapeeth’s Institute of Computer Applications and Management 2020.

[3] Samidha Khatri, Aishwarya Arora, Arun Prakash Agrawal, “Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison” © 2020 IEEE.

[4] Aye Aye Khine, Hint Wint Khin, “Credit Card Fraud Detection Using Online Boosting with Extremely Fast Decision Tree” © 2020 IEEE.

[5] Dingling Ge, Shunyu Chang, “Credit Card Fraud Detection Using Lightgbm Model” © 2020 IEEE.

[6] Altyeb Altaher Taha, Sharaf Jameel Malebary, “An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine ” © 2020 IEEE.

[7] J. V. V. Sriram Sasank, G. Ram sahith, K.Abhinav, Meena Belwal, “Credit Card Fraud Detection Using Various Classification and Sampling Techniques: A Comparative Study” Proceedings of the Fourth International Conference on Communication and Electronics Systems (ICCES 2019).

[8] Debachudamani Prusti, Santanu Kumar Rath, “Fraudulent Transaction Detection in Credit Card by Applying Ensemble Machine Learning techniques”, 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT 2019).