Online Transaction Fraud Detection Using Machine Learning

Ashish Ranjan
(Scholar)
School of Computer Science & Engineering
Galgotias University, Greater N
ashish.21scse1010688@galgotiasuniversity.edu.in

Raushav Royal
(Scholar)
School of Computer Science & Engineering
Galgotias University, Greater N
raushav.21scse1011102@galgotiasuniversity.edu.in

Department of Computer Science and Engineering, Galgotias University

ABSTRACT
The rising volume of electronic payments and internet transactions has transformed online transaction fraud into a key concern for both financial institutions and individual account holders.This research paper presents an exhaustive literature review of machine learning methods for online transaction fraud detection to improve detection system precision with efficacy. Conventional rule-based methodologies determined identity risks are inadequate to protect against advanced fraud schemes in the modern era.

1. INTRODUCTION
Over the past decade the Internet has seen uncontrolled expansion. The vast growth of the Internet encouraged the ubiquitous use of numerous services like e-commerce and tap and pay systems and internet bill payment systems. Online transaction fraud has grown due to increased fraudster activity on internet payment systems. Online transactions are safeguarded by combined systems that include data encryption and tokenization procedures. These safeguard systems are effective for the majority of transactions but their scope is not of total protection from online transaction fraud.

Machine Learning (ML) refers to the AI subset by which computers learn capabilities to enhance prediction after training with data that has been processed before without being explicitly instructed. Machine Learning methods are used to detect fraudulent transactions in online transactions throughout this research. Online transaction fraud refers to the term used to explain how unauthorized users initiate payment without the use of debit or credit cards to enable fraudulent usage.

This expansion, however, will naturally lead to an increase in fraudulent activities along with it. The Nilson Report for 2023 showed that the amount of global card frauds was at $38 billion which makes it evident that we still need systems that can detect such frauds appropriately.

Because of the sophistication of the fraud strategies, traditional rule based fraud detection methods are not sufficient anymore. In contrast to that, machine learning models with the ability to recognize complicated patterns and change over time proved to be better solution. In this paper, various types of machine learning algorithms are applied to detect fraudulent online transactions, examined for their comparison, and the challenges are discussed such as imbalanced datasets and adversarial attacks.Internet transaction fraud activity was the most common type mentioned in 1579 data breaches that affected 179 million data points total in the course of the year, according to Federal Trade Commission (FTC) reports. Proper use of secure internet transaction fraud detection systems to safeguard financial assets of users is necessary. The most important challenge to apply ML techniques for online application arises due to this backdrop.

2. PREVIOUS WORK
Firstly, they were mostly of traditional static rules based systems with the manual development done in the context of domain experts to mark suspicious activity and were adapted for online transaction fraud detection.

There is attention to approaches on machine learning such as logistics regression and decision trees in the period from 2005 to 2012. Chan et al. (1999) and Bhattcharyya et al. (2011). For this purpose, ensemble methods such as random forest, adoption or Xgboost, data resampling methods (smote (synthetic minority overtamping technology) or cost sensitive learning models,) in the literature were introduced to deal with imbalances in data records in order to provide awareness, behavioral features, and to prove that imbalances in data records are critical.

Trick the model and get the model, without sharing the sensitive user data to increase the data protection level of your training model.

Feature selection:

In this study, we employed statistical methods and domain knowledge to identify the most relevant features from transactional data. Key attributes such as transaction amount, time of transaction, location, device ID, user behavior patterns, and historical fraud labels were analyzed for their predictive power. Not only does it improve the accuracy of the model, but it also helps simplify computation so that real time fraud detection can be achieved. A combination of expert knowledge and algorithmic techniques are usually used in feature selection in fraud detection. Redundancy among features is dealt with in a way so that the model remain efficient and doesn't get over fited using certain techniques like correlation analysis.

Research methodology:

The methodology is broken down into the following five core as its methodology: dataset acquisition, preprocessing, feature selection, model training and performance evaluation.

Training multiple supervised machine learning algorithms was a part of the model development phase. Logistic Regression, Decision Tree, Random Forest and Gradient Boosting classifiers have been implemented since they are widely popular to utilize in a classification tasks.

Various classification metrics such as precision, recall, and F1 score were used for the performance evaluation and the area under the Receiver Operating Characteristic curve (AUC-ROC). Recall and F1-score were particular emphasized as the cost of a missed fraudulent case is usually higher than that of a false alarm. Models of the model and the analysis are implemented in Python by the use of the Scikit-learn, Pandas, NumPy, and Imbalanced-learn.

Dataset:

The research conducts analysis using transaction data for two days in September 2013 that European cardholders performed. The total number of 284807 transactions includes 0.172% fraudulent transactions.The complete dataset consists only of numerical value attributes. A fraudulent transaction receives the value 1 in the last column while all other transactions represent 0 in the same column. The dataset features V1 through V28 were given labels for security confidentiality reasons. Our proposed framework incorporates SMOTE as its primary technique for resolving class imbalance during the Data-Preprocessing stage in . With SMOTE the method selects nearby feature space points to draw straight lines between adjacent data points.

Logistic regression

The supervised machine learning method known as Logistic Regression (LR) is a Logit classifier that is utilized to perform binary classification tasks according to research articles [6]. The Logistic

Regression system is a particular type of linear regression system which operates on the basis of an application of linear functions to the logit function.

The logit function accepts the linear function as input.

The value range of q is 0 to 1 during prediction to determine the class probability.The class becomes more predictable with the value of q approaching value 1.

Decision trees and random forest

The supervised literacy algorithm Decision Tree is a classifier and retrogression problem result. A DT has three top knot types that include the root knot and decision bumps and splint bumps. A decision tree algorithm begins its operation from the root knot. The decision knot is the knot that provides a selection process needed for tree splitting. A splint knot specifies the ultimate decision affair of the system. The RF approach makes prognostications grounded on its operation of an ensemble of DTs. The decision in RF is the result of collaborative voting of its constituent rudiments. A fine description of the RF is as follows

The RF model includes the following description when a tree number k is employed

An RF utilizes independent identically distributed trees that bounce on the input vector X. Prediction is made grounded on the voting outgrowth of the most suggested markers.

Naive Bayes

Supervised machine Literacy employs Naive Bayes(NB) as a system on the base of Bayes' theorem. In NB the experimenter specifies class information and also states that all trait dyads are independent of one another. For this study the experimenters have employed Gaussian NB(GNB) as their classifier. The GNB is working on the supposition that attributes are distributed on the base of a Gaussian distribution as stated in Equation.

The values of probability and are set up using maximum probability computation.

Related work

The authors of ref. developed an online sale fraud discovery system using ML styles, which were LR, DT, SVM and RF. These classifiers were tested on sale fraud data attained from European cardholders in 2013. The data have a veritably uneven rate ofnon-fraud to fraud deals and hence it's a veritably imbalanced dataset. The experimenter compared each ML system on the base of its bracket delicacy measure.The authors stated that freshpre-processing ways can enhance classifier results indeed though the reported results were successful.

Varmedja et al. experimenters came up with an ML- grounded system for relating online sale fraud from Kaggle data. The maturity of the dataset consists of European online sale possessors' deals that were made over a period of two days. In this problem, the smoothing system in the dataset was applied by the experimenter using is Synthetic nonage Oversampling fashion(SMOTE) system. To test the proposed system three styles of ML were used RF, NB and multilayer perceptron(MLP). According to experimental findings the RF system was optimal by rightly relating frauds at 99.96. NB and MLP delicacy yielded 99.23 and 99.93 independently. The authors identify the need for carrying out further exploration that would design a point selection system to break

limitations in other ML styles delicacy situations.

The authors made a comparison among five ML styles similar as DT with KNN and LR and RF and NB. The authors used a specific largely imbalanced European cardholder data set for comparing the performance issues of their ML styles. Precision value attained by each classifier came the top measure for measuring performance in experimental study. Results of experimental study showed perfection values for DT, KNN, LR and RF were 85.11, 91.11, 87.5 and 89.77, besides NB attained 6.52.

Awoyemi et al. study examined colorful ML ways to dissect their performance on European cardholders online sale fraud dataset. Authors of this study used a mongrel slice approach to handle imbalances set up within their dataset. The study used NB along with KNN and LR as part of the ML ways. The study performed the trials using a Python- erected ML frame. Each ML approach had its performance carried out through primary performance measure delicacy. delicacy situations attained from trials proved that NB had achieved 97.92, while LR achieved 54.86 and KNN achieved 97.69. point selection was n't included in trials by authors although NB and KNN ways achieved successful issues.

The authors used the European online transactionholder fraud dataset in their exploration. SMOTE slice was employed by the authors to resolve the severe class imbalance in the dataset. DT, LR and IF were the authors' chosen primary ML styles for exploration. delicacy measures were employed as one of the most important performance- measuring factors.

Dataset

This study examined European cardholder online transactions on two September days in 2013. There are 284807 transactions in this database and fraud transactions are merely 0.172% of the total collection. The database contains thirty features like V1 to V28 features and Time and Amount features. All elements of the dataset contain numeric values only. The class column in the database schema employs one as a value to represent fraud transactions and zero for all other transactions.

Data security and integrity requirements necessitate the unnamed features V1 to V28. The database collection reflects one salient finding which consists of low detection accuracy scores due to its unbalanced nature.The SMOTE method consists of discovering pairs of close samples to draw lines of connection between them and creating additional instances from the minority class along those lines.

Feature selection

There's an important step during machine literacy deployment called point selection(FS). When training and testing are executed with the same data, it frequently has lesser than one point that produces inimical performance goods on the overall process performance of models. Experimenter performance necessitates choosing among colorful FS approaches for working particular problems. This paragraph describes colorful scripts when choosing the applicable features using FS approaches produces bettered ML model performance.

Kasongo enhanced intrusion discovery system model performance using his operation of GA- grounded point selection ways in ML models. Mienye et al. studied PSO and demonstrate how it can be applied to SSAE and softmax unit in the FF NN used in heart complaint vaticination. Optimal parameters

adaptation in SSAE point literacy were enhanced through using PSO. At 97.3, the PSO- SSAE approach was suitable to achieve delicacy using the Framingham heart complaint dataset. Hemavathi et al. created a successful FS system that's incorporated into enhanced top element analysis(EPCA) for use in the terrain. A FS system with mongrel features of FS and GA was employed by Pouramirarsalani et al. fore-banking fraud discovery purposes. Experimental testing proved that fractional selection styles used on fiscal fraud datasets have positive impacts on final functional model performance

Genetic algorithm feature selection

The Genetic Algorithm (GA) is used as an Evolutionary inspired Algorithm (EA) to optimize various tasks of optimization with the capacity to reduce computational costs.

The Population-based EAs have a set of potential solutions to which they assign the name population.

A single member of the population is the single individual under fitness. Every individual possesses a gene encoding and a fitness value assigned to it.

The evolutionary process of humans occurs through mutations which take their cue from biological gene evolution.

The main alternative to RF algorithms within the field is tree-based ML techniques Extra-Trees and Extreme Gradient Boosting. A potential solution that is a feature vector enters the fitness method to discover its fitness state. The RF method testing under the GA platform uses a definite attribute vector during its testing phase to discover the fitness measure.

Six working steps constitute this algorithm. The first 20% of the total Online transaction Fraud data is split into training data amounting to 70% and testing data amounting to 30%. The instantiation of the RF classifier is carried out as Step 2 commences. The training of the RF instance is achieved through Step 3 using the training data. The processed model is tested for evaluation using the testing dataset during Step 4. The process predictions are stored in. The evaluation process is carried out through. It is used the most crucial performance metric of the evaluation process. Least accuracy measure is provided by the optimal model of the equation. In the same pseudo code as Algorithm 2, the computation procedure of a candidate feature vector is the same. The first step which was Error free loading of Online transaction Fraud dataset. In the second phase, Computational phase, all the variables in the computation of candidate feature vector will be initialized. Depending on how you see it, the procedure is handled as a list of variables that contain feature labels in the Online transaction Fraud dataset from A to y as the target outcome variable and the B vector for storage of optimal features.

Fraud detection framework

There's a graphical illustration of the structure of the methodology. The homogenize Inputs block starts with original calculation that does min-maximum scaling to homogenize training data through Equation. The input values need a fixed range when they're regularized so the scaling operation is done. The GA point Selection block processes data handed to it by the homogenize Inputs block, the affair of

which executes the GA algorithm using these inputs. The GA point Selection block allows the GA to produce seeker trait vectors whose models are trained in the Training block containing Training data and Train the models blocks. When testing over the test data, the same vector tests the run models. To complete one model, each asked issues must test with only one unique testing case.

Performance metrics

The research model is a part of ML process, and this paper is a binary classification. In the test on the test set, the primary performance measure is from accuracy (AC). In the end, the evaluation is performed between this methodology by and all the models by recalling (RC) with precision (PR) and F1-Score (F-Measure). The quality of classification for each model is represented by the AUC. AUC is a measure to show how good is a classifier for any given classification task. Therefore, the AUC measure is between 0 and 1, as a perfect classifier will have results close to 1.

A correct identification of intrusive activities produces true positive results (TP).

Routine traffic patterns/traces are effective in classification since routine activities which we categorize as True Negatives (TN).

Experiments

Experimental configuration

Experimental procedures were performed using Google Colab. Experimental hardware is Intel(R) Xeon(R) CPU operation at 2.30GHz speed with two cores.

Results and discussions

There were two factors of the examinations. was employed in the bracket of the first step. We had named RF, DT, ANN, NB and LR as trained and tested algorithms for all point vectors of F. Tables 2 to 6 include the tables of data donation. As seen from the Table 2, the test delicacy was over to 99.94 using the RF and ANN. Indeed though other models had inversely respectable test delicacy situations, RF offered the stylish performance in terms of perfection. Results presented confirm that the RF bracket algorithm was 99.93 accurate and that the model handed the stylish results grounded on those results. exercising the stylish fraud discovery delicacy of 99.94 was attained by the RF algorithm. The evaluation of revealed that exercising Decision Tree the perfection values were between 81.17 and the delicacy values were 99.1. The results presented by these had the stylish issues. In similar case, the RF model had 99.98 fraud discovery delicacy and 95.34 perfection. The stylish discovery results were attained in the exploration work of. We relate to numbers 2, 3, 4, 5, 6 and notice that the performance of NB system is before when the Recall, Precision, and F1- Score.

For every vector in F, we calculated its AUC value. The results are in the form of images given in Figs. 4, 5, 6, 7, 8. The RF eliminates the stylish bracket quality between the NB and LR with AUC values of 0.96, 0.97 and 0.97. The whole RF combined with NB achieved the stylish AUC values of 0.95 and 0.96 when applied on 5. The given exploration GA- RF system was tested and vindicated by an illustration that the GA- RF was superior to the model by a 2.28 better delicacy. In comparison to the DT model given in(14), the fraud discovery model by GA- DT in

this exploration has an delicacy of 4.42 advanced. 2.41 enhancement over the stylish result of the SVM handed in(13) are offered by the GA- LR model created in this exploration. In(16), discovery delicacy of about 1.75 advanced over KNN model delicacy is reported for this exploration using the GA- NB model as given. In this exploration, the delicacy position of the GA- DT exceeded the measures taken by a difference of 17.23. The bracket stylish redounded with 5 still the stylish model was RF. The delicacy of the proposed model on online sale fraud discovery was excellent with an delicacy of 99.98.

Experiments on synthetic dataset

We also performed another experiment with publicly available synthetic data and really did so over User, Card, Year, Month, Day, Time, Amount, Use Chip, Merchant Name, Merchant City, Merchant State, Zip, MCC, Errors, and Is Fraud as our feature sets, but Is Fraud is the target variable. Finally, reference is made to the dataset of choices in the form of a total of 24357143 legitimate online transaction transactions and 29757 fraudulent ones [36]. In the experiment, the methods under-use RB, DT, ANN, NB, and LR. The execution of data starts with the framework. The GA module used the features boxed as 0. This feature was taken either from training or testing session of the ML models. The results of which the experimental results obtained when the convergence occurred are as shown in Table 9. Then the GA-RF and GA-LR accuracies were 99.95 and 99.96 % respectively. To keep a lower level of AUC value, 0.63, GA-LR still persisted.

In the experiments researched, all ML models with evaluation are displayed in ROC curve. In the analysis, they gave an AUC of 1 which was paired with the RF and the DT. The detection models, upon achieving the full accuracy, have perfect capacity to detect fraud.

Conclusion

This study examined several boost technologies and models to identify online transaction fraud.This highlights the time and accuracy issues of fraud detection to minimize financial losses and improve the safety of businesses. By comparing the performance of such algorithms,such as vector machines,neuron network support,
We presented the promise of data control identification against fraudulent behaviour problems.Although there is no such model to create 100% accuracy, the method integrated into actual monitoring provied better identification rates.