CREDIT CARD FRAUD DETECTION USING PREDICATIVE MODELLING

K.SUPARNA[1], TANGELLA VIJAY PRAKASH [2]

[1] Assistant Professor MCA, DEPT, Dantuluri Narayana Raju College , Bhimavaram, Andhrapradesh
Email id:- suparnakalidindi@gmail.com
[2]PG Student of MCA, Dantuluri Narayana Raju College , Bhimavaram, Andhrapradesh
Email id :- prakashtangella5@gmail.com

ABSTRACT

Billions of dollars of loss are caused every year by fraudulent credit card transactions. The design of efficient fraud detection algorithms is key for reducing these losses, and more and more algorithms rely on advanced machine learning techniques to assist fraud investigators. The design of fraud detection algorithms is however particularly challenging due to the non-stationary distribution of the data, the highly unbalanced classes distributions and the availability of few transactions labeled by fraud investigators. At the same time public data are scarcely available for confidentiality issues, leaving unanswered many questions about what is the best strategy. In this thesis we aim to provide some answers by focusing on crucial issues such as: i) why and how under sampling is useful in the presence of class imbalance (i.e. frauds are a small percentage of the transactions), ii) how to deal with unbalanced and evolving data streams (non-stationarity due to fraud evolution and change of spending behavior), iii) how to assess performances in a way which is relevant for detection and iv) how to use feedbacks provided by investigators on the fraud alerts generated. Finally, we design and assess a prototype of a Fraud Detection System able to meet real-world working conditions and that is able to integrate investigators' feedback to generate accurate alerts

1. INTRODUCTION

Credit cards are used for purchasing goods and services with the help of virtual card and physical card where as virtual card for online transaction and physical card for offline transaction. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. If the cardholder does not realize the loss of card, it canlead to a substantial financial loss to the credit card company. In online payment mode, attackers need only little information for doing fraudulent transaction (secure code, card number, expiration date etc.). In this purchase method, mainly transactions will be done through Internet or telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Fraud is an offensive activity, carried out by an unauthorized person by cheating innocent. Credit card fraud involves stealing the essential credentials from the cardholder and using it unauthorized manner by the fraudsters either by using phone calls or SMS. This fraud in credit card may also happen using some software applications that are under the control of fraudsters. The credit card fraud detection takes place as: the user or the customer enters the necessary credentials in order to make any transaction using credit card and the transaction should get approved only upon being checked for ay fraud activity. For this to happen, we first pass the transaction details to the verification module where, it is classified under fraud and non-fraud categories. Any transaction that is put under fraud category is rejected. Otherwise, the transaction gets approved.

Most of the time, the genuine way to detect this kind of fraud is to analyse the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns.Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds. Since humans tend to exhibit specific behavioristic profiles, every

cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is a potential threat to the system.

2. LITERATURE SURVEY AND RELATED WORK

1. Fraudulent Detection in Credit Card System Using SVM & Decision Tree

With growing advancement in the electronic commerce field, fraud is spreading all over the world, causing major financial losses. In the current scenario, Major cause of financial losses is credit card fraud; it not only affects tradesperson but also individual clients. Decision tree, Genetic algorithm, Meta learning strategy, neural network, HMM are the presented methods used to detect credit card frauds. In contemplating system for fraudulent detection, artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solvethe problem. Thus by the implementation of this hybrid approach, financial losses can be reduced to greater extent.

2.Machine Learning Based Approach to Financial Fraud Detection Process in MobilePayment System

Mobile payment fraud is the unauthorized use of mobile transaction through identity theft or credit card stealing to fraudulently obtain money. Mobile payment fraud is a fast-growing issue through the emergence of smartphone and online transition services. In the real world, a highlyaccurate process in mobile payment fraud detection is needed since financial fraud causes financial loss. Therefore, our approach proposed the overall process of detecting mobile payment fraud based on machine learning, supervised and unsupervised method to detect fraud and process large amounts of financial data. Moreover, our approach performed sampling process and feature selection process for fast processing with large volumes of transaction dataand to achieve high accuracy in mobile payment detection. F-measure and ROC curve are usedto validate our proposed model.

3. The Use of Predictive Analytics Technology to Detect Credit Card Fraud in Canada

This research paper focuses on the creation of a scorecard from relevant evaluation criteria, features, and capabilities of predictive analytics vendor solutions currently being used to detect credit card fraud. The scorecard provides a side-byside comparison of five credit card predictive analytics vendor solutions adopted in Canada. From the ensuing research findings, a list of credit card fraud PAT vendor solution challenges, risks, and limitations were outlined.

4. BLAST-SSAHA Hybridization for Credit Card Fraud Detection.

This paper propose to use two-stage sequence alignment in which a profile Analyser (PA) firstdetermines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyser are next passed on to a deviation analyser (DA) for possible alignment with past fraudulent behaviour. The final decision about the nature of a transaction is taken on the basis of the observations by these two analysers. In order to achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithmsBLAST and SSAH

3. EXISTING SYSTEM

Since the credit card fraud detection system is a highly researched field, there are many different algorithms and techniques for performing the credit card fraud detection system. One of the earliest system is CCFD system using markov model. credit card fraud detection(CCFD) is also proposed by using neural networks. The existing credit card fraud detection system using neural network follows the whale swarm optimization algorithm to obtain an inceptive value. It the uses BP network to rectify the values which are found error. All of these techniques has some serious disadvantages such as decreasing accuracy levels, lack of efficiency, sometimes classifying the normal transactions as fraud transactions and vise versa. These disadvantages are overcomed in the upcoming proposed systems.

This was on k-means Algorithm implementation, Only the two features with the most variance were used to train the model. The model was set to have 2 clusters, 0 being non-fraud and 1 being fraud. We also experimented with different values for the hyper parameters, but they all produced similar results. Changing the dimensionality of the data (reducing it to more dimensions than 2) also made little difference on the final values

Disadvantages of Existing system

In existing System, a research about a case study involving credit card fraud www.jespublication.com detection, where data normalization is applied before Cluster Analysis and with results obtained from the use of Cluster Analysis and Artificial Neural Networks on fraud detection has shown that by clustering attributes neuronal inputs can be minimized. This research was based on unsupervised learning. Significance of this paper was to find new methods for fraud detection and to increase the accuracy of results.

The data set for this paper is based on real life transactional data by a large European company and personal details in data is kept confidential. Accuracy of an algorithm is around 50%. Significance of this paper was to find an algorithm and to reduce the cost measure. The result obtained was by 23% and the algorithm they find was Bayes minimum risk.The Clustering doesn't produce the less accuracy when compared to Regression methods in scenarios like credit card fraud detection. Comparatively with other algorithms k-means produce less accurate scores in prediction in this kind of scenarios

4. PROPOSED SYSTEM

The Credit card fraud detection system is initiated for detecting the fraud transactions from number of transactions made by the card holders. The transactions done by credit card holders are derived in the form of datasets.

Our goal is to implement machine learning model in order to classify, to the highest possible degree of accuracy, credit card fraud from a dataset gathered from Kaggle. After initial data exploration, we knew we would implement a logistic regression model for best accuracy reports.

Logistic regression, as it was a good candidate for binary classification. Python sklearn library was used to implement the project, We used Kaggle datasets for Credit card fraud detection, using pandas to data frame for class ==0 forno fraud and class==1 for fraud, matplotlib for plotting the fraud and non fraud data, train_test_split for data extraction (Split arrays or matrices into

random train and test subsets) and used Logistic Regression machine learning algorithm for fraud detection and print predicting score according to the algorithm. Finally Confusion matrix was plotted on true and predicted.

Advantages

1. The results obtained by the Logistic Regression Algorithm is best compared to any other Algorithms.
2. The Accuracy obtained was almost equal to cent percent which proves using of Logistic algorithm gives best results.
3. The plots that were plotted according to the proper data that is processed during the implementation
4. The 'amount' feature is the transaction amount. Feature 'class' is the target class for the binary classification and it takes value 1 for positive case (fraud) and 0 for negative case (not fraud).
5. Random forest ranks the importance of variables in a regression or classification problem in a natural way can be done by Random Forest.
6. and produce the efficient output as compare to the existing Kaggle algorithms

5. IMPLEMENTATION

DATASET

This paper utilizes the dataset provided by revolution analytics for the detection of the fraudulent credit card transaction from Kaggle. Dataset has 51149 legal transactions and 3312 fraudulent transactions. The dataset is divided as 60%, 20% and, 20% in the Train, Valid and Test set, respectively

DATA PREPROCESSING

For efficient implementation of the classification algorithm, data preprocessing is performed before feature selection. Under-sampling is performed to make the dataset balanced to avoid the biasing of the classification algorithm towards the majority class. Feature Selection is implemented on a balanced dataset.

FEATURESELECTION

Feature selection methods are used to remove unnecessary, irrelevant, and redundant attributes from a dataset that do not contribute to the accuracy of a predictive model or which might reduce the accuracy of the model. In this paper seven feature selection techniques namely Select-K-best, Feature Importance, Extra tress classifier, Person's correlation, Mutual Information, Step forward selection and Recursive feature elimination are used.

FEATUREIMPORTANCE

Feature importance is a class of techniques for assigning scores to input features to a predictive model that indicates the relative

importance of each feature at the time of making a prediction. It reduces the number of input features. In this paper, feature importance is implemented using an extra tree classifier from the decision tree. Extra Trees is similar to Random Forest, it builds multiple trees and splits nodes using random subsets of features, but unlike Random Forest, Extra Tree samples without replacement and nodes are split on random

6. CONCLUSION AND SCOPE

CONCLUSION:

This machine learning fraud detection tutorial showed how to tackle the problem of credit card fraud detection using machine learning. It is fairly easy to come up with a simple model, implement it in Python and get great results for the Credit Card Fraud Detection task on Kaggle. Credit card fraud detection system using whale optimization algorithm and SMOTE (Synthetic minority optimization technique) aims in indentifying the fraud transactions occurring during the transactions made by the card holder. The system also aims to improve the convergence speed and solves the data imbalance.

In this research, we have proposed a method to detect the fraud in credit card transactions that is based on deep learning. We first compare it with machine learning algorithms such as k- Nearest Neighbor, Support vector machine etc. Finally we have used the neural network, even though tough to train the model which would fit fine to model for detecting a fraud in credit card Transactions. In our model, by using an artificial neural network (ANN) which gives accuracy approximately equal to 100% is best suited for credit card fraud detection. It gives ac-curacy more than that of the unsupervised learning algorithms.

FUTURE SCOPE:

From the above analysis, it is clear that many machine learning algorithms are used to detect the fraud but we can observe that the results are not satisfactory ,so we would like to implement deep learning algorithms to detect credit card fraud accurately.

7.REFERENCE

1. A. Taha, S.J. Malebary, An intelligent approach to credit card fraud detection using an optimizedlight gradient boosting machine, IEEE Access 8 (2020) 25579–25587, doi:10.1109/ACCESS.2020.2971354.
2. S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid, H. Zeineddine, An experimental study with imbalanced classification approaches for credit card fraud detection, IEEE Access 7 (2019) 93010–93022, doi:10.1109/ACCESS.2019.2927266.
3. C. Jiang, J. Song, G. Liu, L. Zheng, W. Luan, Credit card fraud detection: a novel approach using aggregation strategy and feedback mechanism, IEEE Internet Things J. 5 (5) (Oct. 2018) 3637–3647,

doi:10.1109/JIOT.2018.2816007.

4. https://www.ijeat.org/wp-content/uploads/papers/v8i5/D6468048419.pdf
5. https://www.ijert.org/credit-card-fraud-detection-using-machine-learning
6. https://www.sciencedirect.com/science/article/pii/S2666285X21000066