International Journal For Technological Research In Engineering
Volume 7, Issue 10, June-2020 *ISSN (Online): 2347 - 4718*

REVIEW OF CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

Apoorva[1], Ashwini S H[2], Bindushree S[3], Sinchana N[4], Prof. K N Prashanth Kumar[5]
[1,2,3,4]BE (Student), [5]Assistant Professor
Dept of CSE BIT, Bangalore, INDIA,

ABSTRACT: Credit card fraud is a wide-ranging term for theft and fraud committed using or involving a payment card, such as a credit card or debit card, as a fraudulent source of funds in a transaction. If this kind of fraud is not detected quickly, it could cause more problems to the owner of the card as the fraudster can miss use the identity of the person. Therefore, an automated system for detection of fraud can be implemented. This paper proposes one such automated system for credit card fraud detection using a Machine Learning approach, which makes use of the SVM-PSO algorithm.
KEYWORDS: Credit card fraud, SVM, SVM-PSO, Feature Extraction, Dimensionality Reduction

I. INTRODUCTION

Credit Card Fraud can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used. Due to the rise and acceleration of E-Commerce, there has been a tremendous use of credit cards for online shopping which led to High amount of frauds related to credit cards. In the era of digitalization, the need to identify credit card frauds is necessary. Fraud detection involves monitoring and analysing the behaviour of various users in order to estimate, detect or avoid undesirable behaviour. In order to identify credit card fraud detection effectively, we need to understand the various technologies, algorithms and types involved in detecting credit card frauds.

Fraud has been increasing drastically with the progression of state-of-art technology and worldwide communication. Fraud can be avoided in two main ways: prevention and detection. Prevention avoids any attacks from fraudsters by acting as a layer of protection. Detection happens once the prevention has already failed. Therefore, detection helps in identifying and alerting as soon as a fraudulent transaction is being triggered. However, there has been an extreme increase in fraudulent transactions that affect the economy dramatically. Credit card fraud can be classified into several categories. The two types of frauds that can be mainly identified in a set of transactions are Card-not-present (CNP) frauds and Card-present (CP) frauds. Those two types can be described further by bankruptcy fraud, theft/counterfeit fraud, application fraud, and behavioural fraud.

Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent. The investigators provide feedback to the automated system which is used to train and update the algorithm to eventually improve the fraud-detection performance over time.

II. LITERATURE REVIEW

Addresses credit card fraud detection through a semi supervised approach, in which clusters of account profiles are created and used for modelling classifiers. Accounts are first profiled and then clustered. Profiling is done on the basis of the behavioural trends of the customer and they are clustered into similar groups. Four different types of views based on which profiling was done were: Spend, which calculates the mean and variance of the data associated with an account's transaction. Spread, in which diversity of spending was measured. Safety, under which safety preferences of an account are measured. Sketch, which stores the most preferred account type on which transactions are recorded and mean and variance of money in account before each transaction are computed. The performance was measured for Random Forest and XG Boost.

Proposes the novel approach for fraud detection which consists of four stages. Cardholder's

historical transaction data is used to divide all cardholders into different groups. Window-sliding approach is used for aggregating the transactions in each group. Based on historical transactions and aggregated transactions, behavioural patterns are extracted. Then the classifiers are trained to detect fraud instances. Feedback-mechanism is used to solve concept drift. Experimental results show that AggRF only performs well on FFDR, and (RawLR) only performs well on CDDR.

Two kinds of random forests are used to train the behaviour features of normal and abnormal transactions. In Random-Tree based Random forest, the collection of bootstrapped samples selected from the standard training set randomly with replacement is used as the training set for each tree. The subset of attributes is selected randomly at each internal node and the centres of different classes of data at the current node is calculated. The base classifier used in

CART-based Random Forest is CART (Classification and Regression Trees). The training set is obtained by the bootstrapped samples selected randomly from the standard set. The best attribute from the subset of attributes at each node is chosen according to Gini impurity which measures the uncertainty of the dataset. The performance of these two

 Copyright 2020.All rights reserved.

models, which differ in their base classifiers are analysed on credit card transactions.

This paper presents the Naïve Bayes improved K-Nearest Neighbor method (NBKNN) for Fraud Detection of Credit Cards. Principal Component Analysis (PCA) transformation of input values has been done on the dataset which makes the dataset contain only numeric input values. The Naïve Bayes machine learning classifier is a supervised learning technique that tries to predict a class which is known as outcome class based on probabilities, and also conditional probabilities of its occurrence from the training data. KNN follows a process to learn in which it keeps focusing on storing the data until it is actually having the input data whose label or class is meant to be predicted. KNN classifier predicts that how close the unidentified tuple is to the K training set, and KNN does this by using some distance measure. Credit Card Fraud Detection for given data set was done using Naïve Bayes and KNN individually with the precision of approximately 95% and 90% respectively. Experimental results illustrate that both classifiers work differently for the same dataset.

This paper proposes countering the fraud activities through data mining and machine learning, which is one of the prominent approaches introduced by scholars intending to prevent the losses caused by these illegal acts. Data mining techniques were employed to study the patterns and characteristics of suspicious and non-suspicious transactions based on normalized and anomalies data. Machine learning (ML) techniques were employed to predict the

suspicious and non-suspicious transactions automatically by using classifiers. Therefore, the combination of machine learning and data mining techniques were able to identify the genuine and non-genuine transactions by learning the patterns of the data. This paper discusses the supervised based classification using Bayesian network classifiers namely K2, Tree Augmented Naïve Bayes (TAN), and Naïve Bayes, logistics and J48 classifiers.

This paper considers fraud detection problem as a sequence classification task and employ Long Short-term Memory networks to incorporate transaction consequences. It is a special type of Recurrent Neural network, whose structure s similar to that of a standard multilayer perceptron, with the addition that it allows connections among hidden units associated with discrete time steps. The time steps index the individual elements in a sequence of inputs. Through the connections across time steps the model can retain information about the past inputs, enabling it to discover temporal correlations. This is compared to the baseline Random Forest classifier, and the result showed that LSTM is best suited for offline transactions where the cardholder is physically present at the merchant.

In this paper, many supervised machine learning algorithms are applied to detect credit card fraudulent transactions using a real-world dataset. Further, these algorithms have been employed to implement a super classifier using ensemble learning methods and the most important variables that may lead to higher accuracy in credit card fraudulent transaction detection was identified. GBT is a collection of classification and regression models which produces a prediction model in the form of an ensemble of weak prediction models like decision trees. Boosting improves the tree accuracy. XGB (XG boost Classifier) is the most refined classifier that works with all type of dataset. SVM is a discriminative classifier defined by a separating hyperplane. The given labelled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples.

This paper proposes a combination of methods in which shopping behaviour and spending behaviour of cardholder's is analysed. In shopping behaviour, the fraud is detected based on the type of product customer buys and it is further matched with database created and then if the transaction is matched then transaction succeeds or else it is shown as fraud. In spending behaviour, the fraud is detected based on the maximum amount spent. If the amount matches with the amount stored in the database, then the transaction is legitimate, else the transaction is classified as fraud. Hidden Markov Model and Genetic Algorithm are used for credit card fraud detection. In Hidden Markov Model, profiles are maintained and statistics of a particular user and statistics of different fraud scenarios are clustered. Genetic Algorithm is used for calculation of threshold and accurate frauds. Finally, average is taken out by summing the result. This work explores different views of the same problem and see what can be learned from the application of each different technique.

This paper uses two advanced data mining approaches, Support Vector Machines (SVM) and Random Forests, together with the well-known Logistic Regression (LR) to detect and thus control credit card fraud. A real-life dataset on credit card transactions from the January 2006– January 2007 period was used in our evaluation. Random forests and SVM are two approaches that have gained prominence in recent years with noted superior performance across a range of applications.

 Copyright 2020.All rights reserved.

The performance of k-NN, Naive Bayes and Logistic Regression is investigated on highly skewed credit card fraud data. A hybrid technique of under sampling and over sampling is carried out to balance the data set. The performance is measured based on accuracy, sensitivity, specificity, precision, Mathews Correlation Coefficient and balanced classification rate. The results show of optimal accuracy for naïve bayes, k-nearest neighbour and logistic regression classifiers are 97.92%, 97.69% and 54.86% respectively. The comparative results show that k-nearest neighbour performs better than naïve bayes and logistic regression techniques.

Table 1: Related works comparison

Sl. No	Work	Methods	Accuracy
1	Navin Kasa, Andrew Dahbura, Charishma Ravoori, Stephen Adams	Random Forest XGBoost	90%
2	Changjun Jiang, Jiahui Song, Guanjun Liu, Lutao Zheng, and Wenjing Luan	AggRF RawLR AggRF+FB	75% 60% 80%
3	Shiyang Xuan , Guanjun Liu, Zhenchuan Li, Lutao Zheng, Shuo Wang, Changjun Jiang	Random-tree based RF CART-based RF	91.96% 96.77%
4	Sai Kiran,Jyothi Guru,Rishabh kumar,Naveen Kumar,Deepak Katariya,Maheshwar Sharma	Naive Bayes improved K-Nearest Neighbor method (NBKNN)	95%
5	Ong Shu Yee, Saravanan Sagadevan and Nurul Hashimah Ahamed Hassain Malim	K2 Naive Bayes TAN Logisitic Regression J48	95.8% 96.7% 99.7% 100% 100%
6	Johannes Jurgovsky, Michael Granitzer, Konstantin Ziegler, Sylvie Calabretto, Pierre-Edouard Portier, Liyun He-Guelton, Olivier Caelen	LSTM	
7	Sahil Dhankhad , Emad A.Mohammed, Behrouz Far	RF XGB LR MLP SVM Decision Tree NB KNN GB SC	94.5% 94.5% 93.9% 93.2% 93.2% 90.8% 90.5% 94.2% 93.5% 95.2%
8	Ayushi Agrawal, Shiv Kumar, Amit Kumar Mishra	HMM Genetic Algorithm	
9	Siddhartha Bhattacharyya Sanjeev Jha, Kurian Tharakunnel Christopher Westland	LR SVM RF	94.7% 93.8% 96.2%
10	John O Awoyemi, Adebayo O Adetunmbi, Samuel A Oluwadare	Naive Bayes KNN LR	97.92% 97.69% 54.86%

III. CONCLUSION

In the present-day scenario credit card frauds occur everywhere, many detection techniques are carried out to identify such frauds. Advanced machine learning and deep learning techniques are used in developing detection models. This paper attempts to summarize few techniques designed for credit card fraud detection. There is still a need for further research an enhancement of techniques in this regard to ensure that the developed systems can be deployed for use by banks and other financial institutions.

REFERENCES

[1] "Improving Credit Card Fraud Detection by Profiling and Clustering Accounts", Navin Kasa, Andrew Dahbura, Charishma Ravoori, Stephen Adams 2019 Systems and Information Engineering Design Symposium (SIEDS) IEEE, 2019.

[2] "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism" Changjun Jiang, Jiahui Song, Guanjun Liu, Lutao Zheng, and Wenjing Luan IEEE Internet of Things Journal 2018.

[3] "Random Forest for Credit Card Fraud Detection" Shiyang Xuan ; Guanjun Liu ; Zhenchuan Li, Lutao Zheng, Shuo Wang, Changjun Jiang, 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)

[4] "Credit card fraud detection using Naïve Bayes model based and KNN classifier"Sai Kiran,Jyothi Guru,Rishabh kumar,Naveen Kumar,Deepak Katariya,Maheshwar Sharma.International Journal of Advance Research,Ideas and Innovations in Technology 4.3 (2018).

[5] "Credit Card Fraud Detection Using Machine Learning As Data Mining Technique" Ong Shu Yee, Saravanan Sagadevan and Nurul Hashimah Ahamed Hassain Malim. Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 10.1-4 (2018)

[6] "Sequence Classification for Credit-Card Fraud Detection" Johannes Jurgovsky, Michael Granitzer, Konstantin Ziegler, Sylvie Calabretto, Pierre-Edouard Portier, Liyun He-Guelton, Olivier Caelen. Expert Systems with Applications 100 (2018): Elsevier.

[7] "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study", Sahil Dhankhad , Emad A.Mohammed, Behrouz Far. 2018 IEEE International Conference on Information Reuse and Integration for Data Science

[8] "Implementation of Novel Approach for Credit Card a Fraud Detection", Ayushi Agrawal, Shiv Kumar, Amit Kumar Mishra, 2nd International Conference on Computing for sustainable global development (INDIACom)- 2015

[9] "Data mining for credit card fraud: A comparative study", Siddhartha Bhattacharyya Sanjeev Jha, Kurian Tharakunnel Christopher Westland. Decision Support

[10] "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis", John O Awoyemi, Adebayo O Adetunmbi, Samuel A Oluwadare, (2017) IEEE.

 Copyright 2020.All rights reserved.