Credit Card Fraud Detection Using Machine Learning Algorithm

Aparna Shukla[1], Kavya DN[2]
[1]Dept of Computer Science and Engineering
[2]Professor, Dept of Computer Science and Engineering
[1, 2] Atria Institute of Technology, Bangalore, Karnataka, India

Abstract- *In the fast moving world everything has become digital. From paying small merchants to huge financial arrangements, everything is accomplished by using credit/debit cards. These payment cards leads to illicit financial manoeuvres. The drawback of these payment cards is that not only the legalized owner can use it but also the defrauder. The motive behind these activities is to hack the account for various reasons. PCI DSS formally called as the payment card industry data security standard is the data security standard created to help business process cards payment securely and reduce card frauds. There are many methods to tackle the situation but the consequences are compromised due to some factors. The perspective of this paper focuses on adaboost and majority voting along with online learning methods. This perspective permits us to form a hybrid model using ML algorithms and spread awareness using online learning methods that will lead to detection of such activities in prior to the incident.*

Keywords- Credit card, PCI DSS, Adaboost, Majority Voting, Online learning methods, Detection.

I. INTRODUCTION

A credit card is a payment card issued to users (cardholders) to enable the cardholder to pay a merchant for goods and services based on the cardholder's promise to the card issuer to pay them for the amounts plus the other agreed charges. The card issuer (usually a bank) creates a revolving account and grants a line of credit to the cardholder, from which the cardholder can borrow money for payment to a merchant or as a cash advance.Fraud is intentional deceit to secure unfair or unlawful gain, or to deprive a victim of a legal right. Fraud can violate civil law (i.e., a fraud victim may sue the fraud perpetrator to avoid the fraud or recover monetary compensation), a criminal law (i.e., a fraud perpetrator may be prosecuted and imprisoned by governmental authorities), or it may cause no loss of money, property or legal right but still be an element of another civil or criminal wrong.

The purpose of fraud may be monetary gain or other benefits, for example by obtaining a passport, travel document, or driver's license, or mortgage fraud, where the perpetrator may attempt to qualify for a mortgage by way of false statements. Fraud can be handled in, two mechanism can be used. First, fraud prevention, it is an action which prevents the fraud from happening. Second, fraud detection, the action or process of detecting the presence of something concealed. Credit card fraud is when someone uses your credit card or credit account to make a purchase you didn't authorize. We can use credit cards either directly or we can use them to make online payments. In the direct payments we simply scan the card and insert the pins to make payments, whereas in online payments is usually done by calls or from websites where they ask the users to provide credentials such as card number, validity, and card verification value.

In the era of ecommerce, credit cards pays the most vital role. Payment using credit card is one of most common mode of electronic payment. When a customer purchases a product via credit card, credit card issuer bank pays on behalf of the customer and customer has a certain time period after which he/she can pay the credit card bill. It is required to have a bank account before getting a debit card from the bank. The major difference between a debit card and a credit card is that in case of payment through debit card, the amount gets deducted from the card's bank account immediately and there should be sufficient balance in the bank account for the transaction to get completed whereas in case of a credit card transaction, there is no such compulsion.

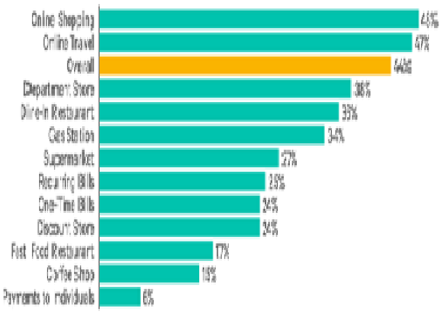


Fig.1:Percentage rate of consumer using credit card

Out of 176 million credit card users, most cardholders carry a balance month to month. An April 2018 report by the

American Bankers Association reveals that 44% of credit card users are revolvers, meaning they carry a balance to the next month at least once every quarter. Another 29.5% of credit card users are transactors, which means they don't carry a balance and have no financing charges. The remaining 26.5% of credit card users show no activity.

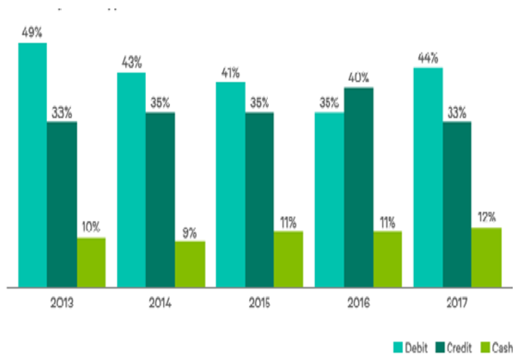


Fig.2:Payment types that consumer prefers

While some people may never experience credit card fraud, it happens every day. In 2015, global credit, debit and prepaid card fraud amounted to $21.84 billion in losses, according to the Nilson Report. By 2020, the same report estimates that card fraud worldwide will reach a whopping $31.67 billion.

In this paper, Ada-boost and majority voting as a hybrid model is used, various machine learning algorithms are also used and implementation of online learning methods for rapid detection of fraud cases. To further evaluate the validity and accuracy, noise is added to the real world data set.

II. LITERATURE SURVEY

[1] E. Rahimika, S. Mohammadi, t. Rahmani, and M. Ghazanfari, "Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran. It concentrates on the effectiveness of using a hybrid intelligent system that combines multi layer perceptron(MLP) neural network support vector machine (SVM) and logistics reasoning(LR) classification models with harmony search(HS) optimization algorithm to detect corporate tax evasion for the Iranian National Tax Administration(INTA). [2]N. Mahmoudi and E. Duman, "Detecting credit card fraud by modified fisher discrimination analysis," This study investigates a linear discriminant, called Fisher discriminant function for the first time in credit card fraud detection problem. Cost of false negatives is very high than false positives and is different for each transaction. A modified Fisher Discriminant Function is proposed in this study which makes the traditional function more sensitive to the important instances.[3] R. Sainand S. Carla, "Evaluating credit card transaction to the frequency domain for a proactive fraud detection approach". Approach presented in the paper takes advantage of novel evaluation criterion based on the analysis, in the frequency domain, of the spectral pattern of the data. This strategy allows us to obtain a more stable model for representing information ,with respect to canonical ones, reducing both the problems of imbalance and heterogeneity of data.[4] N. S. Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using artificial immune system," In this approach we address credit card fraud detection using Artificial Immune Systems (AIS) and introduce a new model called AIS- based fraud detection Model(AFDM). We will use an immune system inspired algorithm(AIRS) and improve it for fraud detection. The accuracy increases upto 25%, reduce the cost upto 85%,and decrease system response time upto 40% compared to base algorithm.

In this approach, the transactions are scored. The legitimacy and fraudulency of the transaction is decided with the help of these scores.The available limit of the card should be the misclassification cost. In the approach a combination of meta-heuristic approaches, genetic algorithm and scatter search is used.

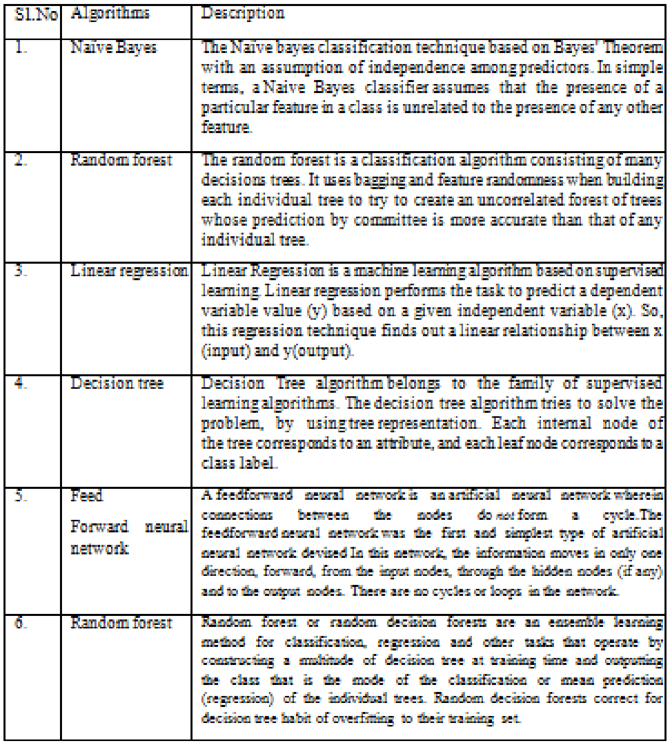
III. EXISTING SYSTEM

In this section we deal with the already existing system applied for credit card fraud detection.

[4]Fraud detection in Credit Card by Clustering ApproachIn this approach, K-means algorithm is used and data is generated for credit card. Then the K-means algorithm is used to detect if the transaction is fraud or not. Clusters are formed namely low, high, risky and high risky to detect fraud. [5]GMM based under sampling and its application for credit card fraud detection. This approach is based on Gaussian mixture under sampling. It consist of three steps. First the majority samples are fitted using Gaussian mixture model. Second, the maximum of probability density function of predicted minority samples are selected as cross edge between two classes. The majority samples near the cross edge are under sampled. The obvious disadvantage of assuming diagonally, however, is that the different features may in practice be strongly correlated. In this case, the model might be incapable of representing the feature distribution accurately.

IV. RELATED WORK

Many algorithms are used in the detection of credit card frauds.Some of them are listed here:



Sl.No	Algorithms	Description
1.	Naive Bayes	The Naive bayes classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
2.	Random forest	The random forest is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.
3.	Linear regression	Linear Regression is a machine learning algorithm based on supervised learning. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output).
4.	Decision tree	Decision Tree algorithm belongs to the family of supervised learning algorithms. The decision tree algorithm tries to solve the problem, by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label.
5.	Feed Forward neural network	A feedforward neural network is an artificial neural network wherein connections between the nodes do *not* form a cycle.The feedforward neural network was the first and simplest type of artificial neural network devised In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.
6.	Random forest	Random forest or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision tree at training time and outputting the class that is the mode of the classification or mean prediction (regression) of the individual trees. Random decision forests correct for decision tree habit of overfitting to their training set.

V. PROPOSED SYSTEM

Majority Voting:

The data classification, Majority voting, uses two algorithms combined model.Every model makes a prediction (votes) for each test instance and the final output prediction is the one that receives more than half of the votes. If none of the predictions get more than half of the votes, we may say that the ensemble method could not make a stable prediction for this instance. Although this is a widely used technique, you may try the most voted prediction (even if that is less than half of the votes) as the final prediction. In some articles, you may see this method being called “plurality voting”.

Ada-boost:

Ada-Boost short for Adaptive Boosting, is a machine learning meta-algorithm. Adaboost is combines with different algorithms to improve their performance. The output of the other learning algorithms ('weak learners') is combined into aweighted sum that represents the final output of the boosted classifier. AdaBoost is adaptive in the sense that subsequent weak learners are tweaked in favor of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. In some problems it can be less susceptible to the overfitting problem than other learning algorithms. The individual learners can be weak, but as long as the performance of each one is slightly better than random guessing, the final model can be proven to converge to a strong learner. Ada-Boost refers to a particular method of training a boosted classifier.

Hybrid model:

A combination of many algorithms becomes hybrid model. In this paper a combination of feed forward network, SVM and decision tree and many algorithms work together to form a hybrid model along with ada-boost and majority voting. Their hybrid models includes two basic submodels, namely, a knowledge-driven (KD) submodel, and a data-driven(DD) sub model. When KD can be given from the first principle or physically based functions, the DD can be neural networks Hybrid machine learning models combine strengths of both knowledge representation model types. Typical hybrid machine learning methods available in.

Weak environment are:

- Model Trees: LMT (Logistic model trees)
- Decision trees and Naive Bayes: NBTree
- Functional trees: FT [18]
- Stacking generalization: StackingC
- Hybrid Hoeding Trees: HT and variants for massive online analysis (MOA)

Online learning methods:

This combined hybrid effort will give the best accuracy in fraud detection. To spread awareness about these methods we take into account the online learning methods. These will help people understand the problem and be prepared for it in advance. By using online learning methods we can detect and stop the fraudulent behaviour well in advance. By using online learning methods and by spreading awareness we can reduce the occurrence of such frauds and the rate will drop eventually. Basically there are four different ways of online learning methods: Asynchronous, synchronous, hybrid and competence based.

Asynchronous learning is self-paced and time independent. Synchronous is time dependent and hybrid is a mixture of both. Hybrid and competence is best suited for the process of online learning methods. It is a personalized approach to learning, where specific skills are mastered to support a particular employment goal. A faculty mentor oversees the process and evaluates if and when competencies are reached.

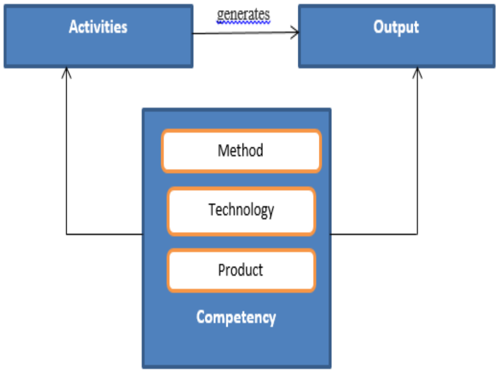


Fig.3:System architecture of competence learning

VI. CONCLUSIONS

This paper provides a brief explanation of credit card fraud detection using various machine learning algorithms. Related studies include Naïve Bayes, SVM and decision tree algorithm and many more. We started from benchmark to deep learning algorithms. These machine learning are separately used for fraud detection, along with majority voting and adaboost making the system a hybrid model. Then we used some datasets to perform the tasks and then we took some real life examples. The performance measure considered is MCC metric. We get 1 as performance measure while using ada-boost and majority voting. Then we keep adding noise to check the accuracy. All this process is unknown to people. An online learning method is needed for people who wants to learn and understand about the process and how to deal with it. There are many approaches to learn these methods involving asynchronous, synchronous, hybrid and competence methods. The paper focuses more on competence learning because it’s the most suited and comfortable method of learning.

VII. ACKNOWLEDGEMENT

I would like to express my thanks to Dr. Aishwarya P, Head of Department, Computer science and Engineering, Atria Institute of Technology, Bangalore for her consistent guidance that helped me to completing the dissertation successfully.

I would like to especially thank my Project guide Prof. Kavya DN, Department of Computer Science and Engineering, Atria Institute of technology, Bangalore for her constant guidance and valuable advice, support and constructive suggestions.

REFERENCES

[1] E.Rahimikia, S.Mohammadi, T.Rahmani, and M.Ghazanfari, ‘‘Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran,’’ Int. J. Account. Inf. Syst., vol. 25, pp. 1–17, May 2017.

[2] N. Mahmoudi and E. Duman, ‘‘Detecting credit card fraud by modified fisher discriminant analysis,’’ Expert Syst. Appl., vol. 42, no. 5, pp. 2510–2516, 2015

[3] R.SaiaandS.Carta,‘‘Evaluatingcreditcardtransactionsinthe frequency domain for a proactive fraud detection approach,’’ in Proc. 14th Int. Joint Conf. e-Bus. Telecommun., vol. 4. 2017, pp. 335–342.

[4] N. S. Halvaiee and M. K. Akbari, ‘‘A novel model for credit card fraud detection using artificial immune systems,’’ Appl. Soft Compute., vol. 24, pp. 40–49, Nov. 2014.

[5] Credit Card Fraud Detection. Accessed: Nov.3,2017 .[Online].Available:https://www.kaggle.com/dalpozz/creditcardfraud.