Evaluation of Credit Card Fraud Detection Using SVM and ANN

[1]Saramath Shamshida, [2]Mustafa Basthikodi, [3]Fathimathul Zohara, [4]Thameeza, [5]Mumthaz M
Dept. of CSE, Bearys Institute of Technology, Mangalore

Abstract – A system for credit card fraud detection is essential for today's technology-driven market since majority of population uses the facility of credit card in the economic system. Here we use two algorithms in machine learning and compare the accuracy of system in credit card fraud detection. It uses full historical transactions of a person including normal transaction data or fraudulent to get normal/fraud transaction features and then these features are used to check whether a transaction is normal or not. The two machine learning supervised algorithms are Support Vector Machine (SVM) and Artificial Neural Network (ANN). The final result will be based on the most accurate algorithm used.

Keywords—Fraud Detection, Classification, Support Vector Machine, Artificial Neural Network.

I. INTRODUCTION

In today's world most of the people uses credit cards for transactions. Changes in mobile intelligent devices and e-commerce made the people to use the credit cards for easy transactions. Credit cards are very useful. It is very easy to carry. Card – not – present transactions or online transactions are very famous today.

Like any other technology, the credit cards have both advantages and disadvantages. The users of credit cards may normal or fraudulent. So it is important to detect to which category a user belongs to. There are many chances to perform fraudulent activities using a credit card.

Fraud detection is a process of monitoring the transaction behavior of a cardholder in order to detect whether an incoming transaction is done by the cardholder or others. It uses full historical transactions of a person including normal transaction data or fraudulent to get normal/fraud transaction features and then these features are used to check whether a transaction is normal or not.

This paper uses two machine learning algorithms Support Vector Machine (SVM) and Artificial Neural Network (ANN). These algorithms are supervised algorithms. Hence it has training and testing phases.

II. FRAUD DETECTION

A. *Fraud Detection Problem*:

Fraud is defined as criminal deception. The purpose of fraud may be to obtain goods without paying or to obtain unauthorized funds from an account. Fraud can either be prevented or detected. In prevention, precaution activities are made to reduce the fraud. If the fraud prevention fails, the problem of detection is taken for consideration. Fraud detection is identifying wrongful, suspect and illegitimate behavior. There are a lot of procedures used for fraud detection. The main target is defending the transactions from illegal use and maximizes the correct predictions.

The credit card fraud can be of two types: offline fraud and online fraud. The fraud begins either with the theft of the credit card or the compromise of data associated with the account, including the card account number or other information that would routinely and necessarily be available to a merchant during a legitimate transaction.

The offline fraud associated with the theft of the credit card. The physical card is stolen by the unauthorized person. Then purchases made by him by using the stolen card. He may use it to purchase until the usage of card is cancelled. If the user does not know about the theft, then it is a great loss to him and to the corresponding financial institutions.

In online fraud, the physical card is not needed but only the information about the card is enough to purchase. Thus here the fraudster simply needs some important card details. Stealing the information from the user is called Identity Theft. In this type fraud the transactions are done through phone or internet. Generally, the genuine cardholder is not responsive if someone else has seen or stolen his card information.

B. Challenging issues in Fraud Detection:

The datasets are extreme imbalance and highly skewed. The genuine transactions dominate than fraudulent transactions. The fraudulent events occur rarely. So it is difficult to find the fraudulent. If the fraudulent transaction is consider as legal then it will cause great loss.

The huge amount of datasets and the dimensionality is very high. It is not an easy process to handle the massive amount of data efficiently. The scalable machine learning system is needed to process the large amount of data.

The real data is not shared for the number of reasons such as to maintain the privacy of the user. Generally the misclassification cost is high for these detections. Efficient measure should take to reduce them is classification cost.

III. SUPPORT VECTOR MACHINE

SVM is a supervised algorithm associated with machine learning. Support vector machine is a method used in pattern recognition and classification. It is a classifier to predict or classify patterns into two categories; fraudulent or non- fraudulent. It is well suited for binary classifications. As any artificial intelligence tool, it has to be trained to obtain a learned model. SVM has been used in many classification pattern- recognition problems such as text categorization, bioinformatics and face detection. SVM is correlated to and having the basics of non-parametric applied statistics, neural networks and machine learning.

IV. EXPERIMENT AND RESULTS OF SVM

A. Data Pre Processing and Selection:

Firstly, the features used in the dataset are converting into numerical data. Feature selection is a very important stage in fraud detection. The features in the data efficiently portray the usage of behavior of an individual credit card account. In this model, the features which interpret the behavior of the customer only are selected for detection. Adding irreverent features make the classifier inefficient.

Transaction amount is the most important behavior it varies from person to person. Frequency of card usage is calculated from the Date and Time Attributes. Average amount of transactions are calculated from each transaction. The important features used in this model are shown in below Table.1.

Table 1: Selected Important Features

Feature No.	Feature Name
1	Transaction Amount
2	Date
3	Time
4	Frequency of card usage
5	Place
6	Customer ID
7	Average amount of transactions per month

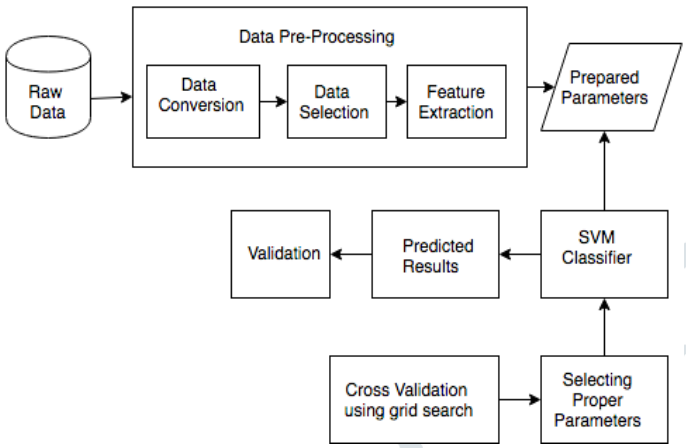

Figure 1: Workflow of the SVM Model

B. Feature Extraction :

Feature extraction is special form of dimensionality reduction. Here the input data is transformed into a reduced representation set of features. The features represent the relevant characteristics of the input data. Instead of using full size input one may use this reduced representation set. If it is properly chosen, then it will give successful task.

Principal component analysis (PCA) is a suitable tool for feature compression. The original feature space is reduced to low dimensional spaces but it will not affect the solution. The computational cost also less for training and testing the SVM because of using PCA.

C. SVM Training and Classification:

LIBSVM classifier is used for training and classification. Libsvm has lot of functions. The 591 samples selected including 576 positive samples and 15 negative samples. Usually SVM suffers from large number of features. To overcome this problem only selected features are used in this model. If the numbers of features are less and the instances are high, then one may have to use the kernel function. In this model RBF kernel function is used. The accuracy is obtained by optimizing the RBF kernel parameter ¥ and the penalty parameter C.

Even though the accuracy is important, the fraud catching rate and false alarm rate are the better metrics for the fraud detection domain . Here the confusion matrix is used for evaluating the fraud catching rate and false alarm rate. The standard confusion matrix format is shown in the following Table 2.

Table 2 : Confusion Matrix

		Predicted	
		Positive	**Negative**
Actual	Positive	TP	FN
	Negative	FP	TN

In Confusion Matrix, the column signifies the predicted class and the row signifies the actual class. TP is True Positive (Fraud catching rate) which shows the number of genuine transactions correctly identified as non fraudulent. FP is False Positive (False alarm rate) which gives the number of genuine transactions incorrectly identified as fraudulent. FN is False Negative mistakenly consider fraudulent transaction as genuine. TN is True Negative which shows the number of fraudulent transactions correctly identified as fraudulent. Achieving highest fraud catching rate and lowest false alarm rate is the important task of this model. The True Positive rate (TP) and False Positive rate (FP) are found by the following Eqs.(1) and (2),

$$TP_{rate}=\frac{TP}{TP+FN} \quad (1)$$

$$FP_{rate}=\frac{FP}{TN+FP} \quad (2)$$

TP_{rate} represents the ratio of positive class that was correctly identified. FP_{rate} represents the ratio of the negative cases that was incorrectly identified as positive.

Accuracy represents the ratio of the total number of transactions that were correctly identified. The accuracy of the classifier is calculated by the Eq.(3) and the error rate is calculated by the Eq.(4),

$$Accuracy=\frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

$$Error\ Rate=\frac{FP+FN}{TP+FP+FN+} \quad (4)$$

The accuracy and the error rate of the proposed work are shown in Figure 2.

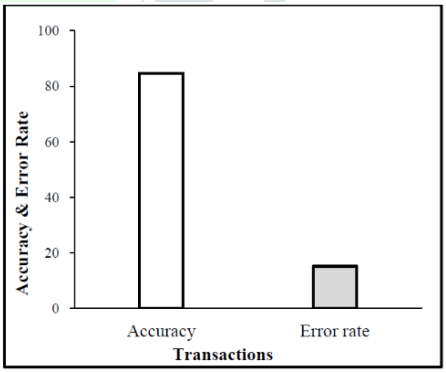

Figure 2: Accuracy and Error rate

In the Result, the TP (True Positive) value raises and FP (False Positive) value becomes low. The highest fraud catching rate and low false alarm rates are obtained by selecting the appropriate parameter values. The parameter values are found out by checking the behavior profiles of the cardholders. The parameter values of the proposed work are based on the average amount of transactions and the frequency of the card usage. This model achieves the accuracy more than 80 percent. Achieving high accuracy is a vital one and reducing the false alarms are also the important tasks in the credit card fraud detection. Too many false alarms restricted the customer from the use of credit card. In this Approach false alarm rates are reduced.

V. ARTIFICIAL NEURAL NETWORKS

Neural Network is a supervised machine learning technique. Artificial Neural Network works similar to the human mind. Human cerebrum comprise of number of neurons associated with each other. Similarly, ANN comprises of artificial neurons, called nodes in network, connected with each other.

In 1943, Warren S.McCulloch presented ANN as a data processing unit for prediction or classification problems. In 1997 Dorronsoro "et al." developed a system that can detect credit card fraud by using neural network. ANN is under the category of machine learning. The most important thing to note about ANN is that it can be used both as supervised or unsupervised method of learning.

ANN is extensively used in fraud detection because it has the ability to detect the hidden pattern in a large and complex data. A computer is capable to think due to neural network technology. It is same as human brain because human brain learns from past experience and is capable in making decision in everyday problem. The same technique is used for credit card fraud detection. When any consumer uses its credit card, a fixed pattern is used that is made by the way consumer uses the credit card.

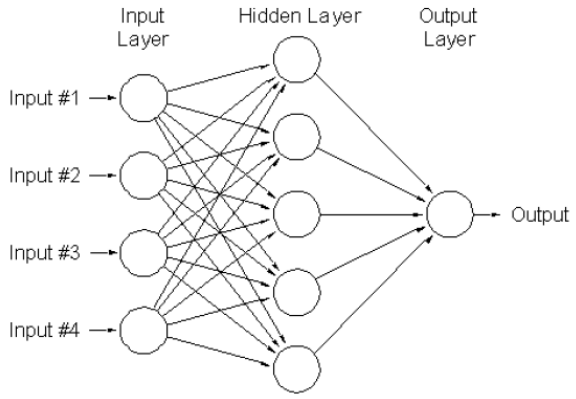

Figure 3: Artificial Neural Network

As shown in the above figure 3, Neural network is a type of train about the information of various things about the card holder such as income, information about large purchase, occupation of card holder, etc. Classification can be done using these patterns of credit card through neural network. One can classify weather a transaction is fraudulent or genuine. We had taken zero as genuine and one as fraudulent. When any unauthorized user uses the credit card, the neural network based fraud detection system checks for the pattern and compares it with the pattern of original card holder on which neural network has been trained. If both the patterns matches, ANN declares the transaction is ok means it is used by valid user.

VI. RESULTS ANALYSIS OF SVM AND ANN

The Accuracy rate of the two algorithms called Support Vector Machine (SVM) and Artificial Neural Network (ANN) can be measured with the help of below graph figure 4.

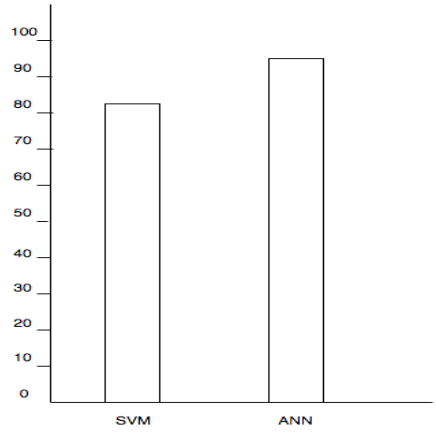

Figure 4: Accuracy graph of algorithms

Table 3: Comparison Table

The below Table 3 describes the results of comparison of SVM and ANN algorithms.

METHODS	SPEED OF DETECTION	ACCURACY	COST
SVM	Low	82%	Expensive
ANN	Fast	92.86%	Expensive

VII. CONCLUSION

The objective of our work is to detect the presence of fraud in credit card database with the help of two classification algorithms and then compared the result of these two algorithms. In our work first we used support vector machine and got the accuracy of 82%. Then we used ANN for the classification of fraud. It gave the accuracy of 92.86%. This is very good result. So when we compared the result of these two algorithms we can see that the neural network gives better result for the classification of fraud. So it indicates that neural network can give better result than other algorithms in the case of classification problem.

REFERENCES

[1] Duman, Ekrem, Ayse Buyukkaya, and Ilker Elikucuk. "A novel and successful credit card fraud detection system implemented in a turkish bank." Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on. IEEE, 2013.

[2] Gaikwad, Jyoti R., et al. "Credit Card Fraud Detection using Decision Tree Induction Algorithm." International Journal of Innovative Technology and Exploring Engineering (IJITEE) 4 (2014).

[3] Dal Pozzolo, Andrea, et al. "Credit card fraud detection and concept-drift adaptation with delayed supervised information." Neural Networks (IJCNN), 2015 International Joint Conference on. IEEE, 2015.

[4] Fathimeh Ghobadi, Mohen Rohani. "Cost Sensitive Modelling of Credit Card Fraud Using Neural Networks." IEEE 2016.

[5] John O.awyemi Adebayo O.Adetumbi, Samuel A.Oluwadare. "Credit Card Fraud Detection using Machine Learning

Techniques:" Department of Computer Science Federal University of Technology. IEEE 2017.

[6] Sahil Dhankhad ,Emad A.Mohammed, Behrouz ."Supervised Machine Learning Algorithms for Credit card Fraudulent Transaction Detection :A comparative Study" Electrical and Computer Engineering.IEEE 2018.

[7] Kuldeep Randhava, Manjeevan Seera. "Credit Card Fraud Detection Using AdaBoost and Majority Voting"IEEE 2018.

[8] Ayushi Agarwal,Shiv Kumar,amith Kumar Mishra. "Credit Card Fraud Detection: A Case Study. IEEE 2015.

[9] Emanuel Mineda Carneiro, Luiz Alberto Vieira Dias,Lineu Fernado Stege Mialaret. "Cluster Analysis and Artificial Neural Networks:A Case Study in Credit Card Fraud Detection ".IEEE 2015.