ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

An Approach to Identify Credit Card Frauds through Support Vector Machine using Kernel Trick

Er. Ritika Wadhwa, Er. Ritika Mehra

Dept. of CSE, RPIIT, Karnal, Kurukshetra University, India

ABSTRACT: The number of Credit Card transactions has been astonishingly rising and due to this breathtaking problem, bunch of users are trapped in Credit Card Fraud. So, up gradation of existing Credit Card Fraud Detection techniques has become a major priority for the banks to minimize their daily losses. In this paper, the researcher has used the concept of Support Vector Machine methodology along with the usage of Specialized Kernels to detect the frauds. The Kernel Trick is a class of algorithms used for pattern analysis, whose best known member is the support vector machine (SVM). The objective of pattern analysis is to find and study general types of relations (for example clusters, correlations, classifications) in datasets. This methodology has been proved to be one of the finest and magnificent approaches in fraud identification because the accuracy of the system has been tested against Linear, RBF and Polynomial Kernel. As a result, polynomial kernel shows the maximum accuracy in the output.

KEYWORDS: Credit Card Discovery; Credit Card Fraud Detection; Kernel Trick; Linear Kernel; RBF Kernel; Polynomial Kernel; Support Vector Machine

I. INTRODUCTION

Nowadays, the E-Commerce world is dynamically ruling our global business. Due to time and work constraints, people tends to buy stuff online through one of the major money weapon known as Credit Card. This weapon has undoubtedly lessened the pressure of buying things physically. But, this is not the end; in fact accessing credit card at every point has tremendously increased the graph of "CREDIT CARD FRAUDS". We cannot deny the fact that prevention measures has not been taken against it but enhancement is always a necessity.

In this paper, the researcher has identified one of the phenomenal approaches i.e. Support Vector Machine along with the usage of Specialized Kernels. The unique point in this approach is that the system is tested for ACCURACY and the researcher has opted for Polynomial Kernel in which it is displaying the maximum accuracy as compare to Linear and RBF Kernel. The **"German Credit Card Data Set** has been used for analysing and identifying Frauds because it is a publically available data set and generally any algorithm works on a good formatted and standardized data without missing values. We ourselves cannot use predefined dataset as we need to take special permissions from the bank authorities.

The polynomial kernel is a non-stationery kernel and works well on the normalized data set just like a German Credit Card Data Set which is purely normalized. Polynomial kernel captures and separates non – linear data set patterns by mapping data to higher dimensional data sets such as 3D. Before moving towards the concept of the Polynomial Kernel under Support Vector Machine in detail, the researcher performed various surveys and studied the existing methodologies for Credit Card Fraud Detection as mentioned in the next section.

II. LITERATURE SURVEY

In [1], the researchers have used the concept of Hidden Markov Model (HMM) to identify the fraudulent transactions in credit card. The fraudulent transactions will only be detected after the transaction is done. This model aims to obtain a high fraud transaction coverage combined with low false alarm rate, thus providing a better and convenient way to detect frauds. Using hidden markov model, customer's pattern is analysed and any deviation from

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

the regular pattern is considered to be a fraudulent transaction. The spending habits of the customers are recorded and the Fraud Detection System (FDS) based on HMM creates the clusters of training sets and identify the expenditure profile of the card holder. The main focus of FDS is on **"Amount of Item Purchased"** and is used for further processing. Different amount of transactions are stored in the form of clusters depending on the transaction amount which will be either 'low', 'medium' or 'high' value series. This paper also includes the additional security features such IP Address detection, MAC Address detection and shipping details detection.

According to [2], the researchers have proposed three phases to detect the credit card frauds. In first phase the Hidden Markov Model is activated which checks the authorization by screening the spending profile of the user. In case the spending profile doesn't match according to the database then access is denied. HMM also traces the IP address of the organization from where the user tries to gain access and if the transaction is found to be fraudulent and it immediately raises the alarm to Admin Department. In second phase, K-Clustering algorithm is used to determine the clusters based on the similarity in the attribute values. At last in the third phase, One Time Password (OTP) authentication takes place in which the user is provided with a password on his / her hand held device which is valid for only one login session or transaction.

In this technique [3], the researchers have used Baum-Welch algorithm to train HMM for each cardholder. All the transactions have been modelled using HMM in a sequential manner. First of all to detect fraud, they have considered three behaviours of the card holder i.e. Low Spending Behaviour, Medium Spending Behaviour and High Spending Behaviour and recorded these behaviours in the form of observation symbols on the basis of K – Clustering Algorithm. The whole system is basically divided into two parts-one is **generating observation symbol** and training and other is **detection**. Training part is performed offline, whereas detection part is an online process. It has been also explained that how an HMM can detect whether the incoming transaction is fraudulent or not. At last they have calculated the performance of system using TP (True Positive) and FP (False Positive) parameters and it is observed that accuracy of system is near to 75%.

According to [4], credit card fraud detection can be done with the help of neural networks i.e. human brain working principle. They have trained the system (neural networks) on the basis of the attributes of a credit card holder. They say that first of all there is very limited time span in which the acceptance or rejection decision has to be made. And, the second one is the huge amount of credit card transactions that have to be processed in a particular time frame. As we know that our brain learn maximum from our past experiences, similarly neural networks learns. They are trained on the basis of some certain protocols. For example: in case of credit card fraud detection, there is a fixed protocol for the customers to access the credit card, which further decides whether the credit card user is fraudulent or not. The neural networks are trained on the basis of certain information from various categories i.e. cardholder's occupation, income and frequency of large amount purchases, location where purchases take place etc.

A hidden Markov model is one in which certain number of emissions are observed but we do not know the sequence of states the model went through to generate the emissions. And, the hidden Markov model helps to analyse the sequence of states from the observed data. For example: predicting the weather for tomorrow or day after tomorrow after observing today's or yesterday's weather. In case of Credit Card Fraud Detection, this methodology has been used by [5] in which the fraudulent or suspicious transactions have been detected on the basis of user's spending profile history. The author said that the fraud detection system will only be activated after 10 transactions done by the user. The probability of occurrences of the huge number of amounts in the transaction history of the user will generate a kind of suspicion for the user.

According to [6], the researchers have proposed a behaviour based classification approach using Support Vector Machines. The main objective of this approach is to achieve the maximum accuracy and fraud catching rate with low false alarm by checking any kind of discrepancies that may occur in the transaction pattern. To accomplish this, they have used only one feature from the dataset i.e. behaviour of the customer. The behaviour feature includes some spending patterns like transaction amount, date, time, place, frequency of purchase and billing Address. Based on these details, the model is trained and then used to classify between the legitimate and fraudulent customers. The proposed method gives higher accuracy of detection and is also scalable for handling large volumes of transactions.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

III. PROPOSED METHODOLOGY AND DISCUSSION

Before proposing the methodology for the problem, it is important to go through the problem statement of the scenario. The Problem Statement is that **"The ever increasing credit card frauds are troubling our worldwide financial institutions"**. It means the frauds have become a threat for our electronic society. So, to overcome this major threat, we have proposed an approach of Support Vector Machine using Polynomial Kernel.

A. Support Vector Machine:

It is a supervised machine learning algorithm that can be employed for both classification and regression purposes. SVMs are more commonly used in classification problems like here we need to classify between the fraudulent and legitimate users of the credit card.

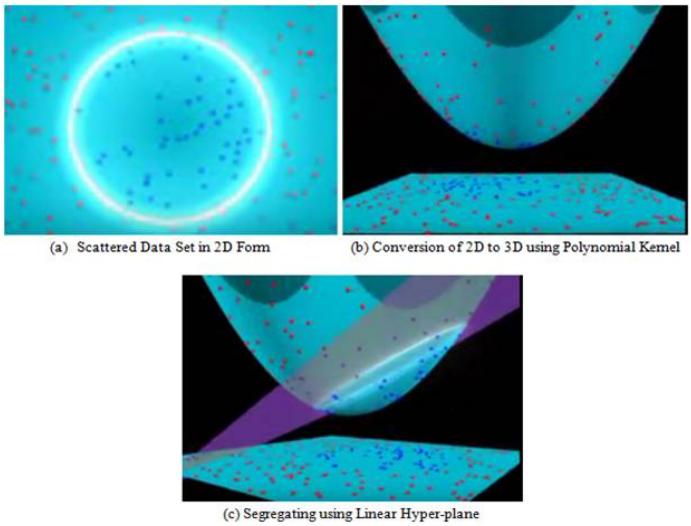
B. Methodology:

Following are the steps which need to be accomplished to train and test the data:

- **Data Set Pre-processing:** If the dataset is available in a numerical form, then it is required to convert it in the parse format or libsvm format because libsvm package cannot work on the format other than sparse.
- **Kernel Function:** There are various kernel in the libsvm package which can be chosen by changing the value of '–t' parameter which is denoted as the type of kernel used. SVM models have been using the linear kernel function, polynomial kernel function and radial basis function kernel function
- **Parameters Selection:** Polynomial kernel has been used to find the best accuracy, training time and prediction time for the Credit Card German Dataset.
- **SVM Classifier Training:** At last the classifier has been trained and validation accuracy has been computed in case of credit card fraud detection.

C. Polynomial Kernel Visualization:

With the help of kernel trick, the researcher has been able to implement the non-linearly separable data sets. As the given problem of Credit Cards Frauds also is in the form of a non-linear data i.e. cannot be segregated with help of a linear (straight) hyper-plane. So, to separate the fraudulent users from the genuine users, polynomial kernel has been used to convert the 2 Dimensional input spaces into 3 Dimensional Feature Space. For example: see the plotting of the graph of the data set in the below Fig.1, which cannot be separated merely by a linear hyper plane, so here comes the polynomial kernel which will be trained and tested in such a way that the ring will bulge out in the 3-D form.



(a) Scattered Data Set in 2D Form (b) Conversion of 2D to 3D using Polynomial Kernel

(c) Segregating using Linear Hyper-plane

Figure 1: Polynomial Kernel Visualization

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

D. Proposed Algorithm:

Step 1: Read the given data.
Step 2: Re-categorize the data in five groups as transaction month, date, day, amount of transaction & difference between successive transaction amounts.
Step 3: Make each transaction in the form of data as vector of five fields.
Step 4: Then make two separate groups of data named True & False transaction group (if false transaction data is not available add randomly generate data in this group).
Step 5: Select polynomial Kernel.
Step 6: Train SVM.
Step 7: Save the classifier.
Step 8: Read the current Transaction.
Step 9: Restart the process from step1 to step3 for current transaction data only.
Step 10: Replaced the saved classifier & currently generated vector in classifier.
Step 11: Admit the generated decision from the classifier.

E. Training Classifier:

Step 1: The training data to be imported should be in the matrix or tabular form.
Step 2: The training data is then trained using the classifier.
Step 3: After training the classifier, perform cross validation.
Step 4: Compute Validation Accuracy. It is a double containing the accuracy in percent.
Step 5: Compute Validation Prediction and Scores.

IV. RESULTS AND ANALYSIS

A. Plotting of Existing Credits and Credits Usage

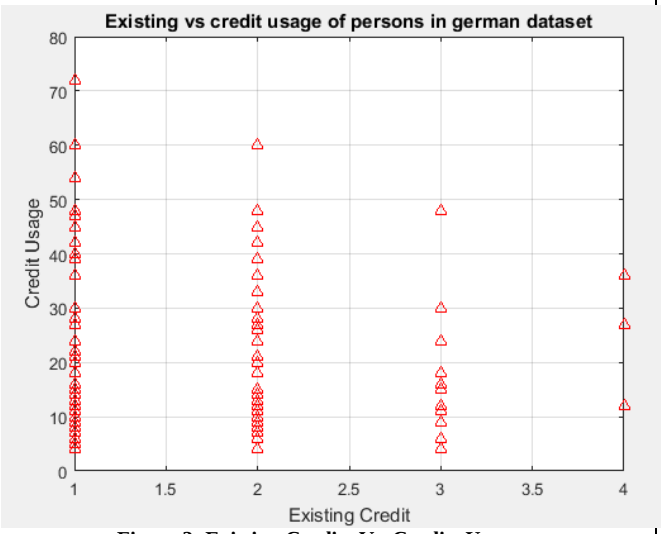


Figure 2: Existing Credits Vs. Credits Usage

B. Plotting of User Age and Credit Usage

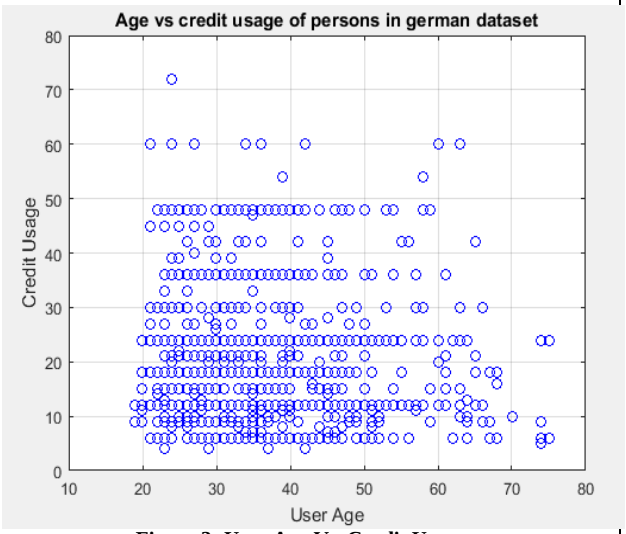


Figure 3: User Age Vs. Credit Usage

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

C. Plotting of Credit Usage and Current Balance

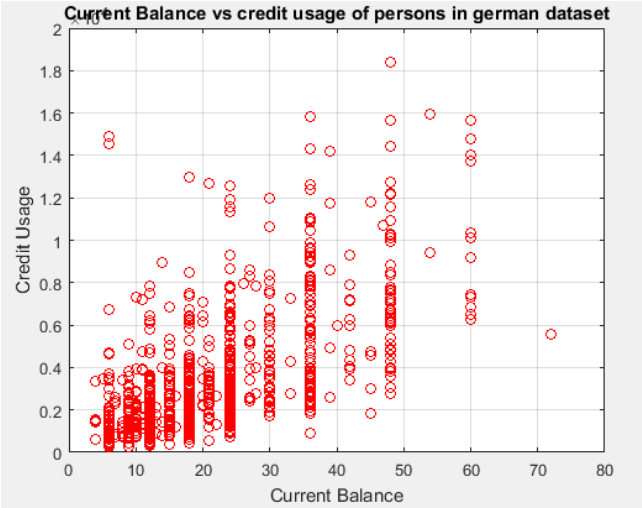


Figure 4: Credit Usage Vs. Current Balance

D. Scatter Plot of Credit Fraud Data

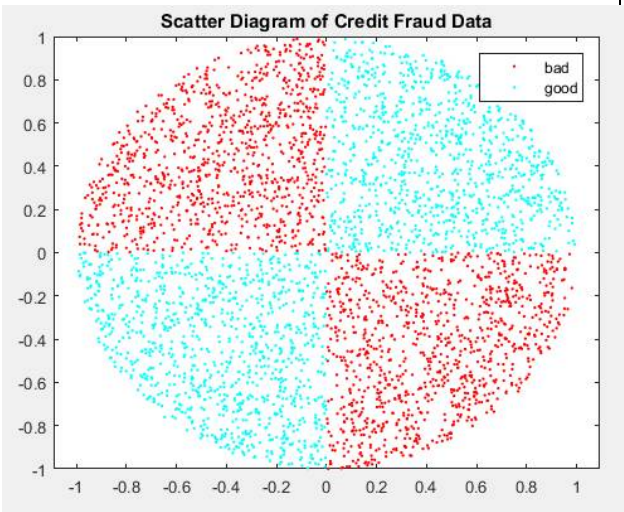


Figure 5: Scatter Plot of Credit Fraud Data

After plotting of certain graphs, "Validation Accuracy" parameter has been calculated for all three kernels (i.e. Linear, RBF and Polynomial) and the value of accuracy along with the elapsed time has been shown below in the command window:

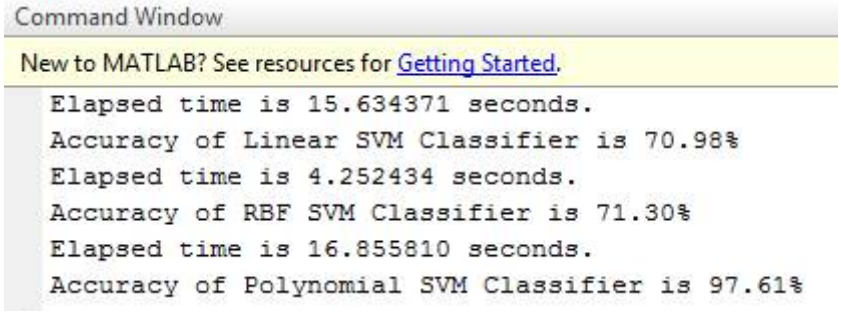


Figure 6: Credit Card Fraud Detection System using SVM

Here, we can observe that the accuracy for Polynomial SVM Classifier is maximum i.e. **(97.61%)** as compare to Linear and RBF SVM Classifiers. At last, the support vector with decision boundary has been displayed as an output which classifies bad and good users.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

Output for the Support Vector with the Decision Boundary:

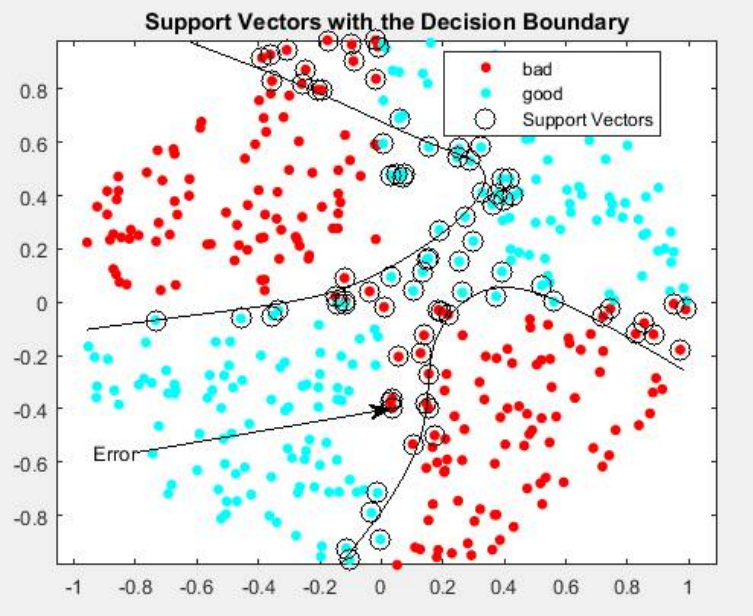


Figure 7: Support Vector with Decision Boundary

V. CONCLUSION

After researching a lot about Credit Card Frauds and their optimal detection methodologies, the researcher has concluded that Support Vector Machine is one of the finest and optimal techniques to differentiate among the genuine and fraudulent users. The major plus point of Support Vector Machine is that it allows the usage of kernel functions and due to which it become little easier to reach the target. The main aim of Kernel function is to take input vector in the original space and then to return the dot product of the two vectors in a feature space. To choose the best kernel for the Credit Card Fraud Detection problem, comparison among three different kernels (Linear, RBF and Polynomial) have been done. And, on the basis of "accuracy" parameter, the classifiers have been trained and then executed for all these kernels. After executing the algorithms (as stated in implementation section), it has been concluded that polynomial kernel gives the maximum accuracy i.e. 97.61% in the output box.

REFERENCES

[1] Ashish Thakur, Bushra Shaikh, Vinita Jain and A.M Magar 'Credit Card Fraud Detection Using Hidden Markov Model and Enhanced Security Features', International Journal of Engineering Sciences and Research Technology (IJERT), pp. 72-77, 2015.
[2] MohdAvesh Zubair Khan, Jabir Daud Pathan and Ali Haider Ekbal Ahmed, 'Credit Card Fraud Detection System Using Hidden Markov Model and K-Clustering', International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 2, 2014.
[3] Avinash Ingole and Dr. R.C Thool 'Credit Card Fraud Detection Using Hidden Markov Model and its Performance', International Journal of Advance Research in Computer Science and Software Engineering, Vol. 3, Issue 6, pp. 626-632, 2013.
[4] Raghavendra Patidar, Lokesh Sharma, 'Credit Card Fraud Detection Using Neural Network', International Journal of Soft Computing and Engineering (IJSCE), Vol. 1, Issue-NCAI2011, 2011.
[5] Gaurav Mhatre , Oshan Almeida , Dhiraj Mhatre and Poonam Joshi, 'Credit Card Fraud Detection Using Hidden Markov Model', International Journal of Computer Science and Information Technologies, Vol. 5, Issue 2, pp. 2053-2055, 2014.
[6] V. Dheepa and R. Dhanapal, 'Behavior based credit card fraud detection using support vector machines', ICTACT journal on soft computing, Vol. 2, Issue 4, 2012