Credit Card Fraud Detection using SVM and Reduction of False Alarms

Nancy Demla
Research Scholar in CSE Department
H.C.T.M College, Kaithal(Haryana),India

Alankrita Aggarwal
Assistant Professor &Head
H.C.T.M College, Kaithal(Haryana),India

Abstract - In day to day life credit cards are used for purchasing goods and services with the help of virtual card for online transaction or 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 can lead 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. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. The only way to detect this kind of fraud is to analyze the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns.

Keywords-Credit Card ,Fraud Detection

I. INTRODUCTION

Credit is a method of selling goods or services without the buyer having cash in hand. A credit card is only an automatic way of offering credit to a consumer. Today, every credit card carries an identifying number that speeds shopping transactions .In the credit card business, fraud occurs when a lender is fooled by a borrower offering him/her purchases, believing that the borrower credit card account will provide payment for this purchase. Really, no payment will be made. If the payment is made, the credit card issuer will reclaim the amount paid. Today, with the expansion of e-commerce, it is on the internet that half of all credit card fraud is conducted. Fraudsters have usually connections with the affected business. In the credit card business, it can be an internal party but most likely an external party. As an external party, fraud is committed being a prospective/existing customer or a prospective/existing supplier[8].

1.1 Big Data

The concept of big data has been endemic within computer science since the earliest days of computing. "Big Data"[8] originally meant the volume of data that could not be processed by traditional database methods and tools. Each time a new storage medium was invented, the amount of data accessible exploded because it could be easily accessed. The original definition focused on structured data, but most researchers and practitioners have come to realize that most of the world's information resides in massive, unstructured information, largely in the form of text and imagery. The explosion of data has not been accompanied by a corresponding new storage medium .Today, we are thinking in tens to hundreds of terabytes. Thus, big data is a moving target. Put another way, it is that amount of data that is just beyond our immediate grasp, e.g., we have to work hard to store it, access it, manage it, and process it. The current growth rate in the amount of data collected is staggering. The major challenges is that this growth rate is fast exceeding our ability to both:

1. Design appropriate systems to handle the data effectively and appropriately
2. Analyze it to extract relevant meaning for decision making.

II. PROPOSED WORK

Though most of the fraud detection systems show good results in detecting fraudulent transactions, they also lead to the generation of too many false alarms. This assumes significance especially in the domain of credit card fraud

detection where a credit card company needs to minimize its losses but, at the same time, does not wish the cardholder to feel restricted too often. We need a novel credit card fraud detection system based on the integration support vector machines. We will use the SVM classification technique for the grouping of the similar instances and employ the incremental learning technique to reduce the misclassification rate based on the attributes transaction amount, type of items purchased and time of transaction.

III. OBJECTIVES

1.To Study and analyze various Credit Card Fraud Detection techniques
2. To Propose a new Credit Card Fraud Detection based on Data Mining using Support Vector Machines
3.To employ the incremental learning technique to reduce the misclassification rate and generation of false alarms
4.To Evaluate the proposed technique using various input and output parameters such as Classification errors, Accuracy and False Alarms.

IV. SUPPORT VECTOR MACHINE

The Support Vector Machine (SVM) is a powerful machine learning tool based on firm statistical and mathematical foundations concerning generalization and optimization theory. It offers a robust technique for many aspects of data mining including classification, regression, and outlier detection. SVM is based on Vapnik's statistical learning theory and falls at the intersection of kernel methods and maximum margin classifiers. Support vector machines have been successfully applied to many real-world problems such as face detection, intrusion detection, handwriting recognition, information extraction, and others.Support Vector Machine is an attractive method due to its high generalization capability and its ability to handle high-dimensional input data.[4].

4.1 Linearly Separable Case

In the linearly separable case, there exists one or more hyperplanes that may separate the two classes represented by the training data with 100% accuracy.

4.2 Non-Linearly Separable Case

In the non-linearly separable case, it is not possible to find a linear hyperplane that separates all positive and negative examples. To solve this case, the margin maximization technique may be relaxed by allowing some data points to fall on the wrong side of the margin, i.e., to allow a degree of error in the separation. Slack Variablesξiare introduced to represent the error degree for each input data point.

V. KERNAL TRICK

.The kernel "trick" allows the computation of the vector product $\Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$ in the lower dimension input space.

From Mercer's theorem, there is a class of mappings Φ such that $\Phi(\mathbf{x})^T \Phi(\mathbf{y}) = K(\mathbf{x},\mathbf{y})$, where K is a corresponding kernel function. Being able to compute the vector products in the lower dimension input space while solving the classification problem in the linearly separable feature space is a major advantage of SVMs using a kernel function. The dual problem then becomes to:

find α that maximizes $\sum_i \alpha_i - 1/2 \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$

subject to $\sum_{i=1}^{N} \alpha_i y_i = 0,$ $0 \le \alpha_i \le C, \quad \forall i$ and the resulting SVM takes the form:

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}_i) + b = \sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b$$

VI. PROPOSED SVM BASED ALGORITHM

Training set: a set of examples used for learning: to fit the parameters of the classifier In the SVM case, we would use the training set to find the "optimal" Support Vectors

Validation set: a set of examples used to tune the parameters of a classifier For SVM case, we would use the validation set to find the "optimal" number of support vectors or determine a stopping point for the algorithm

Test set: a set of examples used only to assess the performance of a fully-trained classifier In the SVM case, we would use the test to estimate the error rate, FP rate or TP rate after we have chosen the final model.

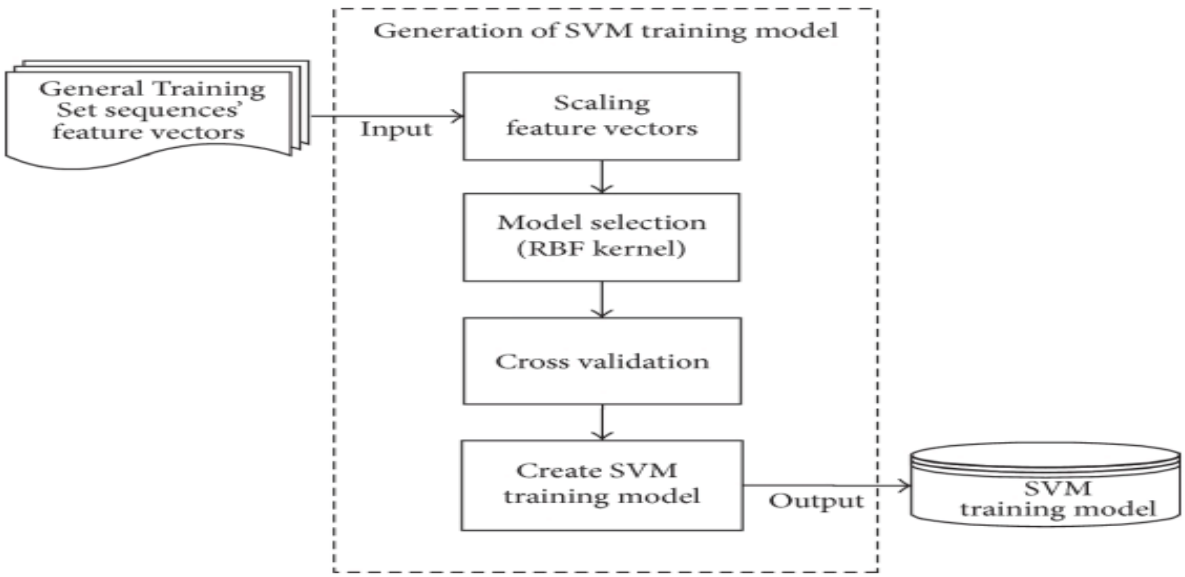


Fig1: Proposed Flow chart of the Algorithm

VII. RESULT &ANALYSIS-

We analyse German Credit Card Dataset

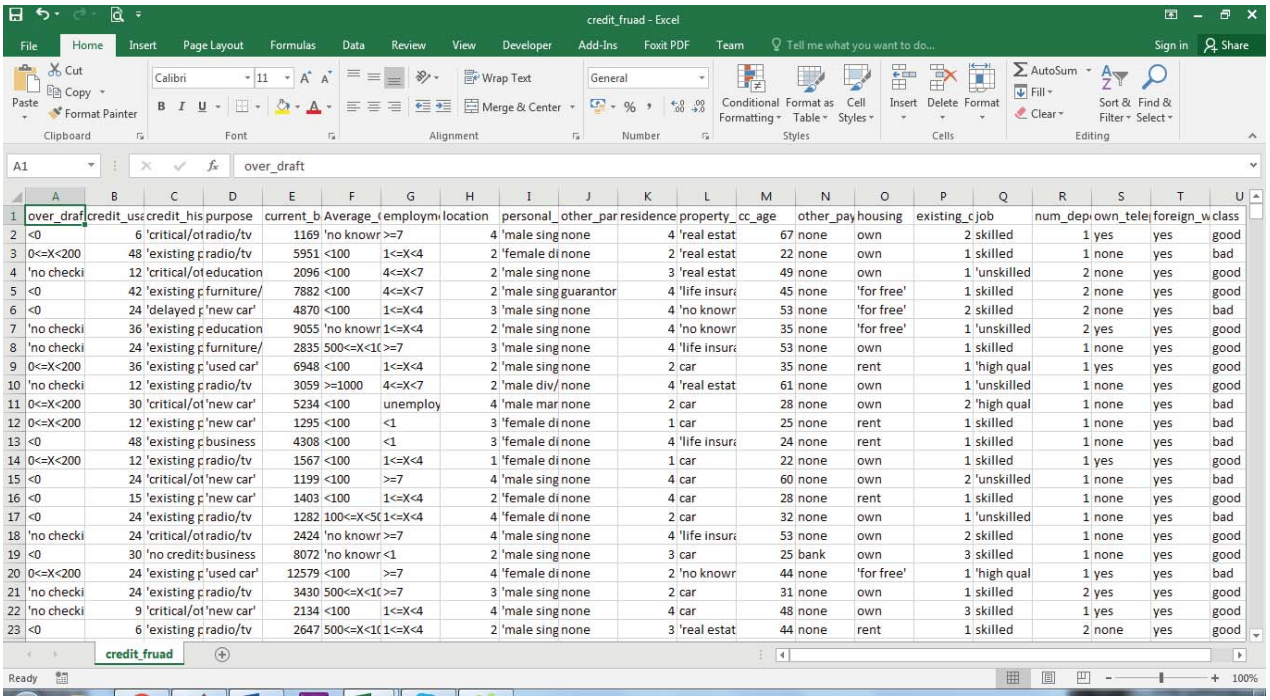
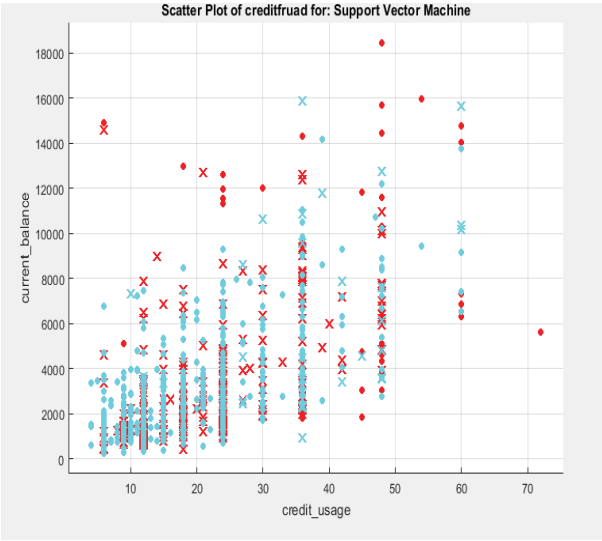


Fig2: German Credit Dataset viewed in Excel



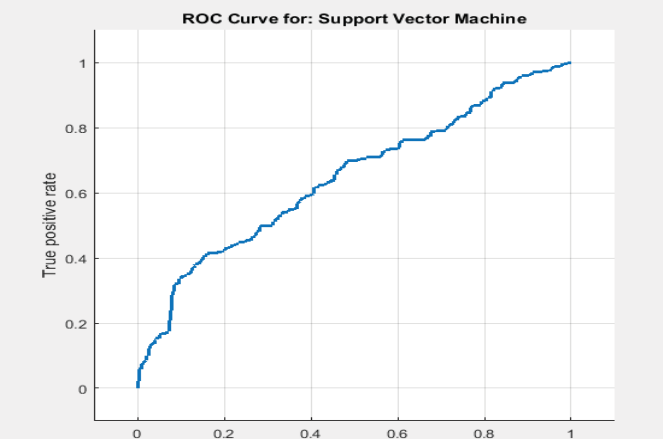


Fig5: Area Under the Curve for SVM showing good area for both detected frauand normal users
Fig3: Scatter Plot showing relationship between current balance and Credit usage of customer

VIII. ACCURACY MEASURES-

The confusion matrix is used as an indication of the properties of a classification (discriminant) rule. It contains the number of elements that have been correctly or incorrectly classified for each class. We can see on its main diagonal the number of observations that have been correctly classified for each class; the off-diagonal elements indicate the number of observations that have been incorrectly classified .one benefit of a confusion Matrix is that it is easy to see if the system is confusing two classes

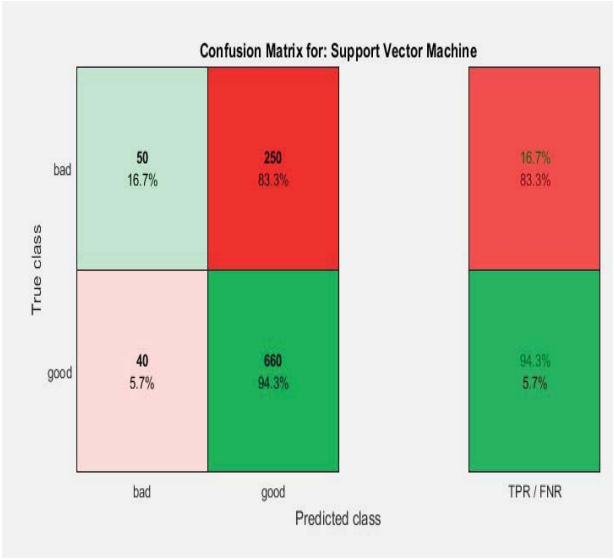


Fig4: Final Confusion Matrix of the Credit dataset with 94.3% overall accuracy

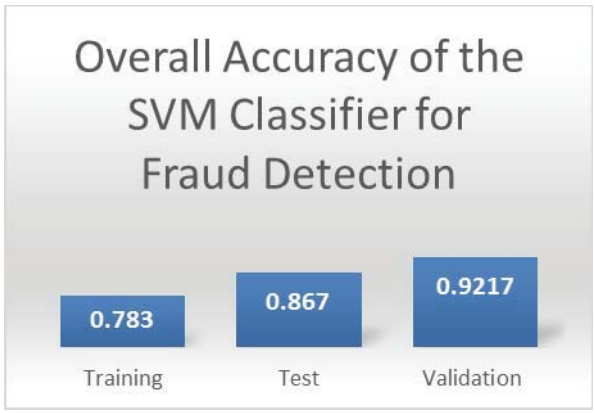


Fig6:Overall Accuracy of the SVM classifier While training, testing and validation

IX. CONCLUSION

This work examined the performance of advanced data mining techniques support vector machines, together with RBF kernel, for credit card fraud detection. For performance assessment, we use a test dataset with much lower fraud rate (0.5%) than in the training datasets with different levels of under sampling. This helps provide an indication of performance that may be expected when models are applied for fraud detection where the proportion of fraudulent transactions are typically low.SVM predicts 94.3% customers correctly; only 6.7% true bad customers are predicted as good customers; and 13.3% true good customers are predicted as bad ones. To compare single tree data mining method with ensemble methods, considering the two wrongly prediction situations the same bad, SVM, bagging, boosting, and random forest are also applied into this dataset. All methods tell us a customer's checking account existing status and duration time are important variables to predict his or her credit risk. Without exception, ensemble methods have lower misclassification rates than the single tree method SVM where bagging shows the best predicting result that 94.3.7% customers in the test sample are predicted correctly.

X. FUTURE SCOPE

Future research can explore possibilities for creating ingenious derived attributes to help classify more transactions more accurately. We created derived attributes based on past research, but future work can usefully undertake a broader study of attributes best suited for fraud modeling, including the issue of transaction aggregation. Another interesting issue for investigation is how the fraudulent behavior of a card with multiple fraudulent transactions is different from a card with few fraudulent transactions. As mentioned above, a limitation in our data was the non-availability of exact time stamp data beyond the date of credit card transactions. Future study may focus on the difference in sequence of fraudulent and legitimate transactions before a credit card is withdrawn. Future research may also examine differences in fraudulent behavior among different types of fraud, say the difference in behavior between stolen and counterfeit cards. Alternative means for dividing the data into training and test remains another issue for investigation. The random sampling of data into training and test as used in this study assumes that fraud patterns will remain essentially same over the anticipated time period of application of such patterns. Given the increasingly sophisticated mechanisms being applied by fraudsters and the potential for their varying such mechanisms over time to escape detection, such assumptions of stable patterns over time may not hold. Consideration of data drift issues can then become important. To better match how developed models may be used in real application, training and test data can be set up such that trained models are tested for their predictive ability in subsequent time periods. With availability of data covering a longer time period, it will be useful to examine the extent of concept drift and whether fraud pattern effect over time.

REFERENCES

[1] Delamaire,L., Abdou,H., and Pointon,J.,(2009) "Credit card fraud and detection techniques: a review." Banks and Bank systems ,4(2),pp. 57-68.

[2] Phua,C., Lee,V., Smith,K., and Gayler,R.,(2010) "A comprehensive survey of data mining-based fraud detection research." arXiv preprint arXiv:1009.6119 .
[3] Raj, S., and A. Annie Portia(2011). "Analysis on credit card fraud detection methods." In International Conference On Computer, Communication and Electrical Technology (ICCCET),IEEE, pp. 152-156.
[4] Sahin, Y., and Duman,E., (2011)"Detecting credit card fraud by decision trees and support vector machines," In International Multi conference of Engineers and computer scientists, (1).
[5] Ganji,V.R., and Mannem,S.N.P.,(2012) "Credit card fraud detection using anti-k nearest neighbor algorithm." International Journal on Computer Science and Engineering ,4(6),pp. 1035.
[6] Chaudhary, K.,Yadav,J., and Malik.B.,(2012) "A review of fraud detection techniques: Credit card," International Journal of Computer Applications ,(0975–8887) .
[7] Dhok, Shailesh S., and Bamnote,G.R.,(2012) "Credit card fraud detection using hidden markov model." International Journal of Advanced Research in Computer Science ,2(4).
[8] Richhariya,P.,Singh,P.K.,(2012),"A Survey on Financial Fraud Detection Methodologies."International Journal of Computer Applications, 45,pp.22.
[9] Singh, A., and Narayan,D.,(2012) "A survey on hidden markov model for credit card fraud detection." International Journal of Engineering and Advanced Technology (IJEAT),pp. 49-52.
[10] Tripathi,K.K.,and Pavaskar,M.A.,(2012), "Survey on credit card fraud detection methods." International Journal of Emerging Technology and Advanced Engineering ,2(11),pp.721-726.
[11] Akhilomen, J.,(2013) "Data mining application for cyber credit-card fraud detection system," In Advances in Data Mining. Applications and Theoretical Aspects, pp. 218-228.