IJARIIT

INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

*(Volume2, Issue4)*

Available online at: www.Ijariit.com

# Credit Card Fraud Detection and False Alarms Reduction using Support Vector Machines

**Mehak Kamboj , Shankey Gupta**

*Student*[#1]*, Assistant Professor*[#2]

*Department Of Computer Science and Engineering*

*DVIET, Karnal, Haryana-132001*

***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. 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. 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. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds.*

***As manually processing credit card transactions is a time-consuming and resource-demanding task, credit card issuers search for high-performing and efficient algorithms that automatically look for anomalies in the set of incoming transactions. Data mining is a well-known and often suitable solution to big data problems involving risk such as credit risk modeling, churn prediction and survival analysis. Nevertheless, fraud detection in general is an atypical prediction task which requires a tailored approach to address and predict future fraud. 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. In this work, we propose a novel credit card fraud detection system based on the integration support vector machines.***

## I. Introduction

**Fraud:** Fraud can be defined as criminal trick aimed to result in financial or personal gain. Fraud prevention and fraud detection systems are two main mechanisms to avoid frauds and losses due to fraudulent activities. Fraud prevention is the upbeat mechanism with the goal of disabling the happening of fraud. Fraud detection systems come into play when the fraudsters go beyond to the fraud prevention systems and start a fraudulent transaction.

**Types of Fraud:** . Credit cards are one of the most famous targets of fraud but not the only one; fraud can occur with any type of credit products, such as personal loans, home loans, and retail. Furthermore, the face of fraud has changed dramatically during the last few decades as technologies have changed and developed. A critical task to help businesses and financial institutions including banks is to take steps to prevent fraud and to deal with it efficiently and effectively, when it does happen.

Anderson has identified and explained the different types of fraud

**Bankruptcy fraud** This section focuses on bankruptcy fraud and advises the use of credit report from credit bureaux as a source of information regarding the applicants' public records as well as a possible implementation of a bankruptcy model. Bankruptcy fraud is one of the most difficult types of fraud to predict. However, some methods or techniques may help in its prevention. Bankruptcy fraud means using a credit card while being insolvent. In other words, purchasers use credit cards knowing that they are not able to pay for their purchases. The bank will send them an order to pay. However, the customers will be recognized as being in a state of personal bankruptcy and not able to recover their debts. The bank will have to cover the losses itself. Usually, this type of fraud loss is not included in the calculation of the fraud loss provision as it is considered a charge-off loss. The only way to prevent this bankruptcy fraud is by doing a pre-check with credit bureaux in order to be informed about the banking history of the customers.

**Theft fraud** This section focuses on theft fraud and counterfeit fraud, which are related to each other. Theft fraud means using a card that is not yours. The perpetrator will steal the card of someone else and use it as many times as possible before the card is blocked. The sooner the owner will react and contact the bank, the faster the bank will take measures to stop the thief. Similarly, counterfeit fraud occurs when the credit card is used remotely; only the credit card details are needed. At one point, one will copy your card number and codes and use it via certain web-sites, where no signature or physical cards are required. On-line merchants are at risk because they have to offer their clients payment by credit card. In cases where fraudsters use stolen or manipulated credit card data the merchant loses money because of so-called "charge-backs". Note that charge-backs are generated if credit card holders object to items on their monthly credit card statements because they were not responsible for the purchase transactions.

**Application fraud** Application fraud is when someone applies for a credit card with false information. To detect application fraud, the solution is to implement a fraud system that allows identifying suspicious applications. To detect application fraud, two different situations have to be distinguished: when applications come from a same individual with the same details, the so-called duplicates, and when applications come from different individuals with similar details, the so called identity fraudsters.

In most banks, to be eligible for a credit card, applicants need to complete an application form. This application form is mandatory except for social fields. The information required includes identification information, location information, contact information, confidential information and additional information. Recurrent information available would be for identification purposes, such as the full name and the date of birth. The applicant would inform the bank about his/her location details: the address, the postal code, the city and the country. The bank would also ask for contact details, such as e-mail address, land-line and mobile phone numbers. Confidential information will be the password. In addition, the gender will be given. All those characteristics may be used while searching for duplicates.

To identify the so-called duplicates, cross-matching techniques are in common use. Rather than using statistical techniques, another method easy to implement is cross-matching. For instance, simple queries that give fast results are to cross-identify information with location details. Examples would be "last name and date of birth and postal code and address" or "last name and address and e-mail and gender". By those queries, individuals with more than one card are identified. Those are quite simplistic queries but will remove most duplicates from the system. Note that duplicates may usually be genuine. Customers can reapply filling in a new address or spelling differently in one of the fields. By contrast, identity crime, as it is named, is perpetrated by real criminals filling wrong application data consciously.

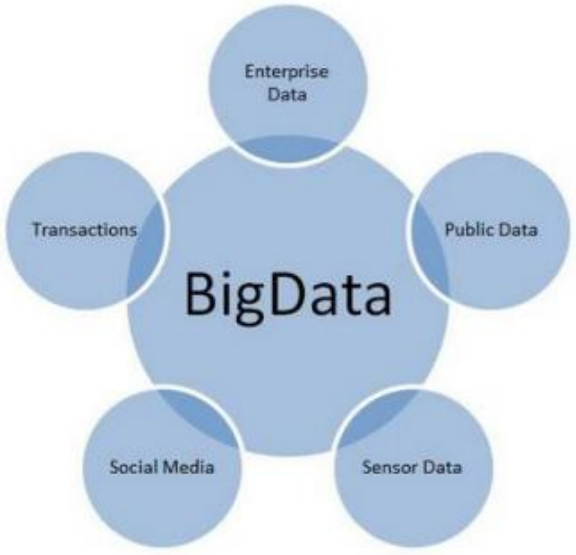
**Behavioral fraud** Behavioral fraud occurs when details of legitimate cards have been obtained fraudulently and sales are made on a 'cardholder present' basis. These sales include telephone sales and e-commerce transactions, where only the card details are required. Behavioral fraud can be detected by implementing a fraud scorecard predicting which customers are likely to default. Traditional credit scorecards are used to detect customers who are likely to default, and the reasons for this may include fraud. Regarding the process, using scoring for fraud prevention is similar to any other use, including profit, default, and collection. The score reflects experience of past cases, and the result is a binary outcome: a genuine customer or a fraudster.

The key difference is that professional fraudsters will make their application look very genuine. Therefore, some scoring developments for fraud prevention have not proved worthwhile because they are unable to differentiate between genuine applications and fraudulent applications. On the other hand, if one uses scoring as a fraud check in addition to using a different scoring model as a credit risk check, any improvement will add value. However, the value of this additional check relies on it not presenting too many false-positive cases. To detect fraudulent applications is possible once they have gone through the system and have been bank customers for a certain time. To build a scorecard, it is important to define what the profile of a fraudulent customer is, and especially the cardholder level profiles encapsulating normal transaction patterns, such as frequency of use, typical value range, types of goods purchased, transaction types, retailer profiles, cash usage, balance and payment histories, overseas spending patterns and daily, weekly, monthly and seasonal patterns.

**Big data:** The concept of big data has been endemic within computer science since the earliest days of computing. "Big Data"[1] 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. We define "Big Data" as the amount of data just beyond technology's capability to store, manage and process efficiently. These imitations are only discovered by a robust analysis of the data itself, explicit processing needs, and the capabilities of the tools used to analyze it. As with any new problem, the conclusion of how to proceed may lead to a recommendation that new tools need to be forged to perform the new tasks. As little as 5 years ago, we were only thinking of tens to hundreds of gigabytes of storage for our personal computers [2].

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 challenge is that this growth rate is fast exceeding our ability to both:

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



**Reach of Big Data in to Multitude of Areas**

**Big Data Analytics:** Big data analytics is the process of examining large data sets containing a variety of data types i.e., big data **--** to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. The analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over rival organizations and other business benefits. The primary goal of big data analytics is to help companies make more informed business decisions by enabling data scientists, predictive modelers and other analytics professionals to analyze large volumes of transaction data, as well as other forms of data that may be untapped by conventional business intelligence (BI) programs[3].

That could include Web server logs and Internet click stream data, social media content and social network activity reports, text from customer emails and survey responses, mobile-phone call detail records and machine data captured by sensors connected to the Internet of Things.

**Credit Card Transactions in big data :**As manually processing credit card transactions is a time-consuming and resource-demanding task, credit card issuers search for high-performing and efficient algorithms that automatically look for anomalies in the set of incoming transactions. Data mining is a well-known and often suitable solution to big data problems involving risk such as credit risk modelling, churn prediction and survival analysis. Nevertheless, fraud detection in general is an a typical prediction task which requires a tailored approach to address and predict future fraud. We say that fraud is an uncommon, well-considered, imperceptibly concealed, time-evolving and often carefully organized crime which appears in many types and forms.

**Detection techniques**

**Decision tree:** The idea of a similarity tree using decision tree logic has been developed. A similarity tree is defined recursively: nodes are labelled with attribute names, edges are labelled with values of attributes that satisfy some condition and ‘leaves’ that contain an intensity factor which is defined as the ratio of the number of transactions that satisfy these condition(s) over the total number of legitimate transaction in the behavior.

**Genetic algorithms and other algorithms:** Algorithms are often recommended as predictive methods as a means of detecting fraud. One algorithm that has been suggested by is based on genetic programming in order to establish logic rules capable of classifying credit card transactions into suspicious and non-suspicious classes.

**Clustering techniques:** The peer group analysis is a system that allows identifying accounts that are behaving differently from others at one moment in time whereas they were behaving the same previously. Those accounts are then flagged as suspicious. Fraud analysts have then to investigate those cases. The hypothesis of the peer group analysis is that if accounts behave the same for a certain period of time and then one account is behaving significantly differently, this account has to be notified. Breakpoint analysis uses a different approach. The hypothesis is that if a change of card usage is notified on an individual basis, the account has to be investigated. In other words, based on the transactions of a single card, the break-point analysis can identify suspicious behavior. Signals of suspicious behavior are a sudden transaction for a high amount, and a high frequency of usage.

**Neural networks:** Neural networks are also often recommended for fraud detection. However, the main constraint is that data need to be clustered by type of account. Similar concepts are: Card watch Back-propagation of error signals FDS SOM improving detection efficiency “mis-detections”. Data mining tools, such as ‘Clementine’ allow the use of neural network technologies, which have been used in credit card fraud.

**Bayesian networks** are also one technique to detect fraud, and have been applied to detect fraud in the telecommunications industry and also in the credit card industry. Results from this technique are optimistic. However, the time constraint is one main disadvantage of such a technique, especially compared with neural networks. Furthermore, expert systems have also been used in credit card fraud using a rule-based expert system.

## Data Mining

Data mining is used to mine useful data from a large amount of data in the same way as extraction of minerals from mine fields.
Clustering is the classification of similar objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset share some common trait.

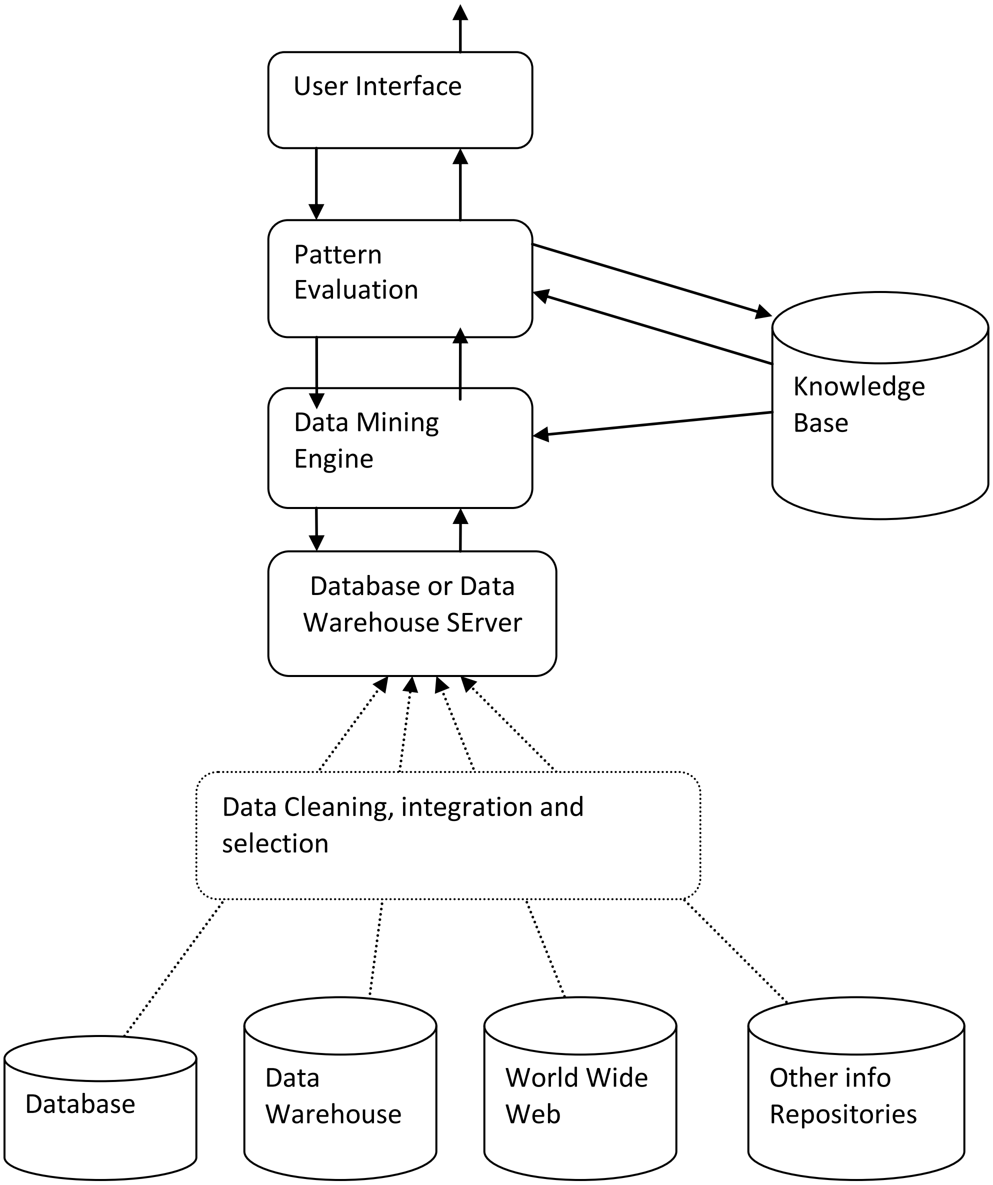
Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. Data mining is more than just conventional data analysis. It uses traditional analysis tools like statistics and graphics plus those associated with artificial intelligence such as rule induction and neural nets. It is all of these, but different. It is a distinctive approach or attitude to data analysis. The emphasis is not so much on extracting facts, but on generating hypotheses.

### Knowledge Discovery in Data

Knowledge Discovery in Data (KDD) is employed to describe the whole process of extraction of knowledge from data.

### 1.1.1 Architecture of Data Mining

The architecture of typical data mining system may have following major components as shown in fig 1.1 below



**Support Vector Machines** 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. Compared to neural networks or decision trees, SVM does not suffer from the local minima problem, it has fewer learning parameters to select, and it produces stable and reproducible results. If two SVMs are trained on the same data with the same learning parameters, they produce the same results independent of the optimization algorithm they use. However, SVMs suffer from slow training especially with non-linear kernels and with large input data size. Support vector machines are primarily binary classifiers. Extensions to multi-class problems are most often done by combining several binary machines in order to produce the final multi-classification results.

The more difficult problem of training one SVM to classify all classes uses much more complex optimization algorithms and are much slower to train than binary classifiers.

In the following sections, we present the SVM mathematical foundation for the binary classification case, then discuss the different approaches applied for multi-classification.

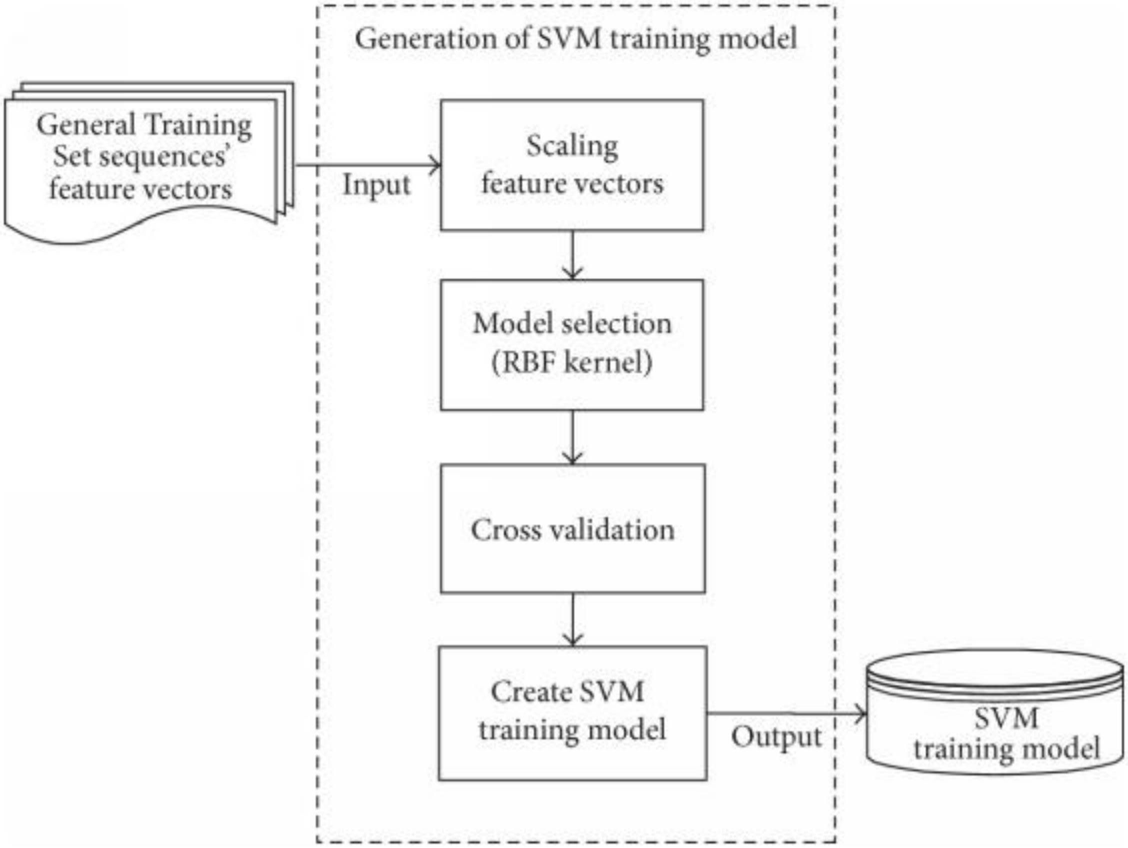
**Proposed SVM based Algorithm**

**Step 1**: In first Step dataset a training set is implemented to build up a model, while a test (or validation) set is to validate the model built. Data points in the training set are excluded from the test (validation) set. Usually a dataset is divided into a training set, a validation set in each iteration, or divided into a training set, a validation set and a test set in each iteration.

**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.


Generation of SVM training model
General Training Set sequences' feature vectors
Input
Scaling feature vectors
Model selection (RBF kernel)
Cross validation
Create SVM training model
Output
SVM training model

Proposed Flow chart of the Algorithm

**Step 2: Scaling Feature Vectors:** In this process the feature vectors are scaled or transformed into numeric vectors as the dataset contains various strings like **0<x<100** conversion or scaling of these features is required for SVM as the SVM can only work with numeric inputs

**Step 3: RBF Kernel (or Kernel Selection):** The Gaussian RBF kernel is very popular and makes a good default kernel especially in absence of expert knowledge about data and domain because it kind of subsumes polynomial and linear kernel as well. Linear Kernels and Polynomial Kernels are a special case of Gaussian RBF kernel. Gaussian RBF kernels are non-parametric model which essentially means that the complexity of the model is potentially infinite because the number of analytic functions are infinite. If you see it from the point of view of polynomial kernel, it essentially is infinite polynomial kernel. As Gaussian kernels including **RBF** are universal kernels i.e. their use with appropriate regularization guarantees a globally optimal predictor which minimizes both the estimation and approximation errors of a classifier. Here, approximation error refers to the error incurred by limiting the space of classification models over which search is performed, and estimation error refers to error in estimation of the model parameters. **Also the credit dataset has the property of Exponential Gaussian curves which make the decision of choosing RBF function, as seen below,**

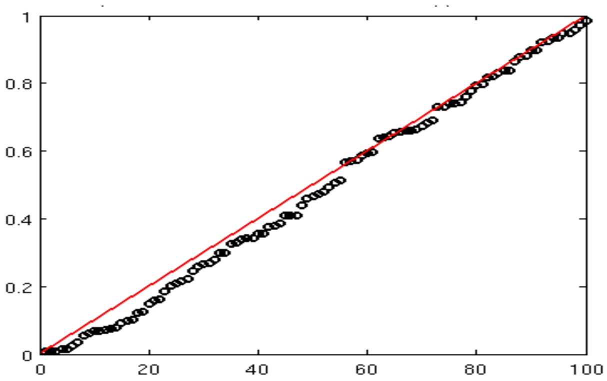


Fig: Linear Distribution best used with Linear SVM

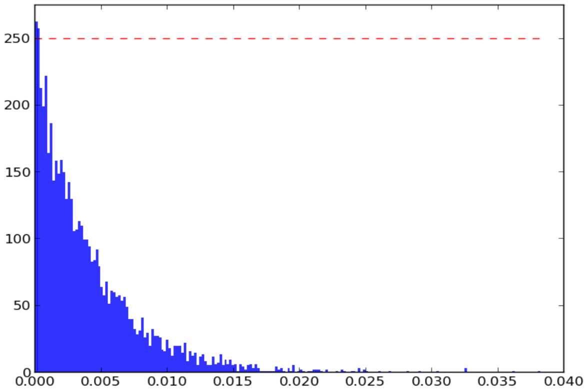
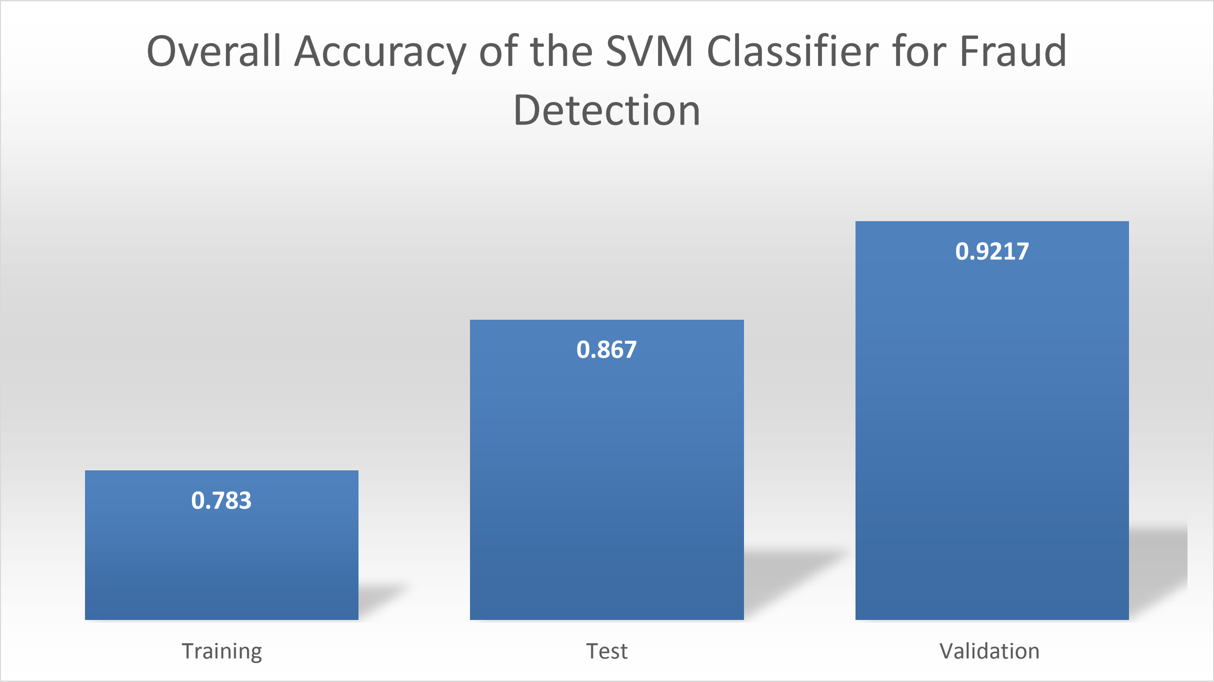


Fig: Exponential Distribution used with RBF kernel

**Step 4: Cross Validation:** Cross-validation sometimes is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. In a prediction problem, a model is usually given a dataset of known data on which training is run (training dataset), and a dataset of unknown data (or first seen data) against which the model is tested (testing dataset). The goal of cross validation is to define a dataset to "test" the model in the training phase (i.e., the validation dataset), in order to limit problems like over fitting, give an insight on how the model will generalize to an independent dataset (i.e., an unknown dataset, for instance from a real problem), etc. **KFOLD analysis with 10 folds in RBF SVM was used for German credit dataset.**

**Step 5: SVM Training Model:** The input feature data is fed into SVM Model for training and testing, after training the SVM Model is achieved which is used for prediction from the data. Matrix of training data, where each row corresponds to an observation or replicate, and each column corresponds to a feature or variable.

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.



**Overall Accuracy of the SVM classifier while training, testing and validation**

**Conclusion** This work examined the performance of advanced data mining techniques support vector machines, together with RBF kernel, for credit card fraud detection. A German credit card transaction form that was used in our evaluation. Till date, their use for credit card fraud prediction has been limited. With the typically very low fraud cases in the data compared to legitimate transactions, some form of sampling is necessary to obtain a training dataset carrying an adequate proportion of fraud to non-fraud cases. The work provides a performance of SVM considering various traditional measures of classification performance and certain measures related to the implementation of such models in practice. 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.

**Future Scope** Future research can explore possibilities for creating ingenious derived attributes to help classify 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 patterns remain in effect over time.

## References

[1]. West, Jarrod, and Maumita Bhattacharya. "Some Experimental Issues in Financial Fraud Detection: An Investigation." arXiv preprint arXiv:1601.01228(2016).

[2]. Carneiro, Emanuel Mineda, Luiz Alberto Vieira Dias, Adilson Marques da Cunha, and Lineu Fernando Stege Mialaret. "Cluster Analysis and Artificial Neural Networks: A Case Study in Credit Card Fraud Detection." In 2015 12th International Conference on Information Technology-New Generationxs (ITNG), pp. 122-126. IEEE, 2015.

[3]. Shimpi, Priya Ravindra, and Vijayalaxmi Kadroli. "Survey on Credit Card Fraud Detection Techniques."

[4]. Sonepat, H. C. E., and Mitali Bansal. "Survey Paper on Credit Card Fraud Detection."

[5]. Kumari, Nitu, S. Kannan, and A. Muthukumaravel. "Credit Card Fraud Detection Using Hidden Markov Model-A Survey." Middle-East Journal of Scientific Research 19, no. 6 (2014): 821-825.

[6]. Richhariya, Pankaj, Prashant K. Singh, Endu Duneja, Bhopal BITS, and I. S. C. Softwares. "A Survey on Financial Fraud Detection Methodologies."International Journal of Computer Applications 45, no. 22 (2012).

[7]. Dhok, Shailesh S., and G. R. Bamnote. "Credit card fraud detection using hidden markov model." International Journal of Advanced Research in Computer Science 3, no. 3 (2012).

[8]. Raj, S., and A. Annie Portia. "Analysis on credit card fraud detection methods." In Computer, Communication and Electrical Technology (ICCCET), 2011 International Conference on, pp. 152-156. IEEE, 2011.

[9]. Sahin, Y., and E. Duman. "Detecting credit card fraud by decision trees and support vector machines." In International Multiconference of Engineers and computer scientists, vol. 1. 2011.

[10]. Phua, Clifton, Vincent Lee, Kate Smith, and Ross Gayler. "A comprehensive survey of data mining-based fraud detection research." arXiv preprint arXiv:1009.6119 (2010).