International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) e-ISSN: 2395-5325
Volume 5, Issue 3 (September '2019)

CREDIT CARD FRAUD DETECTION USING P-SVM CLASSIFICATION TECHNIQUE

S.Narmatha,
M.Phil Research Scholar,
Department of Computer Science,
Thiruvalluvar Government Arts College,
Rasipuram,Tamilnadu,India.

Dr C.Kavitha,
Assistant professor,
Department of Computer Science,
Thiruvalluvar Government Arts College,
Rasipuram,Tamilnadu,India.

Abstract: Nowadays, the banking sector is a very necessary sector in our present day generation where almost each and every human has to deal with the bank either physically or online. In dealing with the banks, the clients and the banks face the possibilities of been trapped by fraudsters. Examples of fraud include insurance fraud, credit card fraud, accounting fraud, etc. The most frequent payment mode is credit card for both on-line and offline in today's world, it presents cashless purchasing at each store in all countries. It will be the most convenient way to do on-line shopping, paying bills etc. Hence Detection of fraudulent activity is necessary in credit card transactions. Credit card fraud can be efficiently detected by data mining techniques. Data mining techniques such as classification, clustering and association rules analyze the credit card transaction data and predict the patterns which lead to fraud. Because of increased credit card fraud, the main aim of this research is to develop a model that should efficiently analyze and predict the fraudulent data set. In this research, an efficient credit card fraud detection model P-SVM classification is proposed and compared with decision tree and Naïve Bayes algorithms. The proposed model uses credit card holder's expenditure behavior and location analysis for analyzing and predicting fraud in the credit card transaction data. Various evaluation measures are used for analyzing the performance of Naïve Bayes, Decision Tree and P-SVM classification technique and proved that P-SVM yields higher accuracy when compared with Naïve Bayes and Decision Tree.

Keywords: Fraud Detection, P-SVM classification, Classification Algorithm, Naïve Bayes , Profile, Decision Tree, Credit Card, Time Series.

I.INTRODUCTION

The use of credit cards has increased, due to the fact of a speedy development in the digital commerce technology. The credit card will become the most famous kind of payment for each on-line as properly as normal purchase, instances of credit card fraud also increasing. In Modern day the fraud is one of the principal outcomes of outstanding business losses, not only for merchants, the individual customers are also affected. Credit Card Fraud: Credit card fraud has been dividing into two types: Offline fraud and On-line fraud. The Offline fraud is devoted through the use of a stolen physical card at call center or any other place. The On-line fraud is devoted through phone, shopping, and web or in absence of card holder. **Offline Fraud:** Most offline fraud incidences appear as a end result of theft of mail, touchy information associated to clients bank or credit card accounts, stolen ATM/debit/credit cards, forged/ stolen cheque etc. client can defend from such instances through exercising warning while receiving, storing and disposing consumer account statements as properly as Cheque, ATM/Debit and Credit Cards. **Online Fraud:** Online fraud happens when anyone poses as a reliable organization (that may also or may not be in order to acquire sensitive non-public data and illegally conducts transactions on your current accounts. Typically, criminals send emails that appear like they're from authentic sources, but are not. The fake messages typically consist of a hyperlink to phony, or spoofed, websites, where victims are requested to supply sensitive non-public information. The information goes to criminals, instead than the reliable business or "spoofing" An on-line identification theft scam. Typically, criminals send emails that seem to be like they're from reliable sources, but are now not (phishing). Web sites and pop-up windows, or any aggregate of such methods. The major goal of each offline as properly as on-line fraud is to steal your 'identity'. This phenomenon is frequently recognized as "identity theft". Identity theft (A crook recreation where a thief appropriates fundamental records such as your name, delivery date, account number, or credit card quantity except your knowledge) happens when anyone illegally obtains your non-public information — such as your credit card number, bank account number, or different identification and makes use of it many times to open new accounts or to initiate transactions in your name. Identity theft can occur even to these who do not shop, communicate, or transact online. A majority of identification theft happen offline. Stealing wallets and purses, intercepting or rerouting your mail, and rummage through your trash are some of the frequent techniques that thieves can use to achieve non-public information.

Type of fraud in Banking Sector

The Banking used to be keeps data on frauds on the groundwork of location of operation below which the frauds have been perpetrated. According to such data pertaining, pinnacle classes below which frauds have been reported through banks are as follows,

1) Credit Cards
2) Deposits – Savings A/C
3) Internet Banking
4) Housing Loans
5) Term Loans
6) Cheque / Demand Drafts
7) Cash Transactions

8) Cash Credit A/c (Types of Overdraft A/C)
9) ATM / Debit Cards.

Credit Card Fraud Detection

Although the use of credit cards as a payment technique can be actually convenient for our every day transactions; people need to be aware of the risks that they impose themselves while using their credit cards. More precisely, the incremental usage of credit cards gave the possibility to fraudsters to take advantage of their vulnerabilities. Credit card fraud refers to any illegal and unauthorized activity on the use of credit cards which is undertaken through a fraudster. According to credit card fraud has been multiplied between 2010 and 2019.

The credit card fraud detection elements uses user behavior and region scanning to test for unusual patterns. These patterns consist of user characteristics such as user spending patterns as well as common user geographic locations to confirm his identity. If any unusual pattern is detected, the system requires revivification. The system analyses user credit card data for various characteristics. These characteristics consist of user country, regular spending procedures. Based upon previous data of that user the system acknowledges unusual patterns in the charge procedure. So now the system may also require the user to login once more or even block the user for greater than 3 invalid attempts.

Core Features:

- The system stores previous transaction patterns for each user.
- Based upon the user spending ability and even country, it calculates user's characteristics.
- More than 20 -30 %deviation of users transaction(spending history and operating country) is considered as an invalid attempt and system takes action.

Advantages

- Due to Behavior and location analysis approach, there is a drastic reduction in the number of False Positives transactions identified as malicious by an Fraud Detection System although they are actually genuine..
- The system stores previous transaction patterns for each user.
- Based upon previous data of that user the system recognizes unusual patterns in the payment procedure.
- The System will block the user for more than 3 invalid attempts.

Nowadays fraud detection is a hot topic in the context of electronic payments. This is mostly due to considerable financial losses incurred by payment card companies for fraudulent activities [1].

In this context the transactional profile can reveal the frauds. Many researches consider this kind offraudulent activitiesandconstructatransactionalprofile[7],[8],[9]and [10]. But more cautious fraudsters try to follow the normal behaviors of card holder or perform low value transactions in short time intervals. In this case the frequency or volume of transactions is a much better indicator of fraud compared to the characteristics of each individual transaction. For instance, in these frauds the total number or total amount spent on a credit card over a specific time windowincreases.

A few researches consider this type of frauds and construct an aggregated profile. The problem with this approach is the late detection because the system has to wait until the end of the aggregation period before it can make a decision. This problem seems more crucial when the aggregation period is considerable. Also some useful information like the order of data is lost during the aggregation. This order of data is anotheraspectofacardholderbehaviorwhichcanbeusedto detect some types offrauds.

In this research, we approach the credit card fraud detection problem with an improved aggregated profile. For this purpose the sequence of aggregated daily amounts spentonanindividualcardholderinatimewindowhasbeen considered.Thentheinherentpatternsinthesetimeseries havebeenextractedtoshortenthetimebetweenwhenafraud occurs and when it is finally detected. Indeed we have taken advantage of the order of data to timelier fraud detection.We demonstrate that the proposed approach leads to improved detection rate and timeliness while it decreases the cost involved in somecircumstances.

II.RELATED WORKS

B.Pushpalatha et al., [10] the look at Data mining methods like Bayesian networks, Bayes Minimum Risk, Genetic algorithm, Hidden markov model (HMM) and Ontology for enhance fraud detection in credit cards.. Moreover, assessment employed criterions in literature are accrued and discussed. The findings in this work highlight the fraud detection enchantment that a getting to know approach can supply when it is used in conjunction with a set up fraud detection system. The chosen data mining methods amongst the present structures and also show the discovering charge of fraud detection.

Y. Sahin et al., [11] the experiment of classification models primarily based on decision trees and support vector machines (SVM) are developed and utilized on credit card fraud detection problem. This evaluation is one of the firsts to evaluate the overall performance of SVM and decision tree techniques in credit card fraud detection with an actual data set. This framework, financial establishments can utilize credit card fraud detection models to evaluate the transaction data with the historic profile patterns to predict the likelihood of being fraudulent for a new transaction, and supply a scientific foundation for the authorization mechanisms.

Marwan Fahmi et al.,[12] this evaluation of the four fraud detection models based totally on data mining methods Support vector machine, K-nearest neighbors, Decision Trees, Naïve Bayes have been developed and their performances had been in contrast when utilized on a actual lifestyles data set of transaction. Four applicable metrics had been used in evaluating the overall performance of the classifiers which are True effective rate (TPR), False Positive Rate (FPR), Balanced Classification Rate (BCR) and Matthews Correlation Coefficient (MCC). There is no data mining technique that

is universally higher than others. Performance enchantment may want to be carried out via creating a fraud detection model the usage of an aggregate of exclusive data mining techniques (ensemble).

T. Kavipriya et al., [13] to developed and analyze, notice and understand the fraudulent transactions and the system is based totally on environment friendly clustering and classification techniques such as apriori and support vector machine respectively. The result proposed technique offers higher effects which help to achieve excessive fraud insurance blended with a low false alarm rate than the present Hidden Markov Model. The effects had been promising, nearly all the fraudulent transactions may want to be detected efficiently and the proposed technique have been in contrast with the present technique and the result suggests that the proposed method performs better than present methods. In this research fraudulent transactions have been detected and recognized which illustrates the robustness of the proposed system.

III. PROPOSED METHOD

CREDIT CARD FRAUD DETECTION USING P-SVM ALGPRITHM

Payments the use of credit cards have increased in current years. It may also be used in on-line or in everyday shopping. Now-a-days credit card payments are fundamental and convenient to use. Due to the increase of fraudulent transactions, there is a want to locate the efficient fraud detection model. Fraud is growing dramatically with the expansion of modern-day technology. The amount of information in the present day world is also explosive. The promising way to detect the fraud is to analyze the spending behavior of the cardholder. Detecting the fraud means identifying the suspicious one. If any abnormality arises in the spending behavior then it is considered as suspicious and taken for further consideration. In this research, credit card fraud is detected using three classification algorithms, such as, Decision Tree, Naïve Bayes and P-SVM.

Behavioral fraud, on the different hand, has 4 main types: stolen/lost card, mail theft, counterfeit card and card holder not existing fraud. Stolen/lost card fraud happens when fraudsters steal a credit card or get access to a lost card. Mail theft fraud happens when the fraudster get a credit card in mail or personal information from financial institution earlier than attaining to real cardholder. In each counterfeit and card holder not existing frauds, credit card details are received without the information of card holders.

The fraud is mainly detected with time, amount and location attributes in each transaction. If there is a large variation on time and amount then it will be considered as behavior change and it should be detected. Based on age and gender attributes, we can find how much amount taken by child, youngster and elders (spending behavior). There is any abnormality of amount in the particular age group (child, younger and elders) it considers as a fraud based on previous history. Finally if there is any change of location then it is considered as a fraud and it should be detected.

3.1. DECISION TREE

Decision tree algorithms are a technique for coming near discrete-valued goal functions, in which the discovered characteristic is denoted through a decision tree. These kinds of algorithms are famous in inductive learning and have been effectively utilized to range of tasks. Decision trees arrange instances through sorting them down the tree from the root to some leaf node, which provides the classification of the instance. Each node in the tree specifies a test of some attribute of the instance and every department descending from those node hyperlinks to one of the possible values for this attribute.

A decision tree is a flowchart-like tree structure, where every internal node represents a check on an attribute, every branch represents an effect of the test, and category label is represented through every leaf node (or terminal node). Given a tuple X, the attribute values of the tuple are examined against the decision tree. A direction is traced from the root to a leaf node which holds the classification prediction for the tuple. It is handy to convert decision trees into classification rules. Decision tree learning uses a decision tree as a predictive model which maps observations about an object to conclusions about the item's goal value. It is one of the predictive modeling strategies used in statistics, data mining and machine learning. Tree models where the goal variable can take a finite set of values are called classification trees, in this tree structure, leaves characterize classification labels and branches signify conjunctions of elements that lead to these classification labels. SQL statements can be developed from tree that can be used to access databases efficiently. Decision tree classifiers achieve comparable or higher accuracy when compared with different classification methods. A number of data mining techniques have already been executed on instructional data mining to enhance the overall performance of college students like Regression, Genetic algorithm, Bays classification, k-means clustering, companion rules, prediction etc.

Data mining techniques can be used in instructional area to decorate our perception of learning procedure to focus on identifying, extracting and evaluating variables associated to the learning technique of students. Classification is one of the most frequently. [29].

Decision Tree Algorithm

The decision tree algorithm is a top-down induction algorithm. The purpose of this algorithm is to construct a tree that has leaves that are homogeneous as possible. The main step of this algorithm is to proceed to divide leaves that are not homogeneous into leaves that are as homogeneous as possible. Steps of this algorithm are given below.

Input:

- Data partition, D, which is a set of training tuples and their associated class labels.
- Attribute list, the set of candidate attributes.
- Attribute_selection_method, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes.

Output: A decision tree.

Attribute selection
Attribute selection consists basically of two different types of classes:

- evaluator – determines the merit of single attributes or subsets of attributes • search algorithm – the search heuristic

Evaluator
The evaluator algorithm is responsible for determining merit of the current attribute selection
Super classes and interfaces.
The ancestor for all evaluators is the weka. Attribute Selection. AS Evaluation
Class.
Here are some interfaces that are commonly implemented by evaluators:

- Attribute Evaluator – evaluates only single attributes • Subset Evaluator – evaluates subsets of attributes

Attribute Transformer – evaluators that transform the input data.
Can be used for optional post-processing of the selected attributes, e.g., for ranking purposes.

3.2. NAIVE BAYES ALGORITHM

Given a set of objects, every of which belongs to a recognized class, and every of which has a recognized vector of variables, our goal is to assemble a rule which will enable us to assign future objects to a class, given only the vectors of variables describing the future objects. Problems of this kind, known as troubles of supervised classification, are ubiquitous, and many techniques for developing such guidelines have been developed. One very necessary one is the naive Bayes method—also known as idiot's Bayes, easy Bayes, and independence Bayes.

This technique is necessary for a number of reasons. It is very convenient to construct, not requiring any difficult iterative parameter estimation schemes. It may also be effortlessly utilized to large data sets. It is convenient to interpret, so customers unskilled in classifier technological know-how can recognize why it is making the classification it makes. And finally, it frequently does incredibly well: it may also not Probabilistic techniques to classification usually contain modeling the conditional likelihood distribution P(C|D), where C levels over lessons and D over descriptions, in some language, of objects to be classified.

Given a description d of a particular object, it assign the class argmaxcP(C = c|D = d). A Bayesian approach splits this posterior distribution into a prior distribution P(C) and a likelihood P (D|C): P(D = d|C = c)P(C = c) argmaxc P(C = c|D = d) =argmaxc (1) P(D = d) The denominator P(D = d) is a normalizing factor that can be ignored when determining the maximum a posteriori class, as it does not depend on the class. The key term in Equation (1) is P(D = d|C = c), the likelihood of the given description given the class (often abbreviated to P(d|c)). A Bayesian classifier estimates these likelihoods from training data, but this typically requires some additional simplifying assumptions.

For instance, in an attribute-value representation (also called propositional or single- table representation), the individual is described by a vector of values a1, . . . , and for a fixed set of attributes A1, . . . ,An. Determining P(D = d|C = c) here requires an estimate of the joint probability P(A1 = a1, . . . ,An = an|C = c), abbreviated to P(a1, . . . ,an|c). This joint probability distribution is problematic for two reasons: (1) its size is exponential in the number of attributes n, and (2) it requires a complete training set, with several examples for each possible description. These problems vanish if it can assume that all attributes are independent given the class:

The Naive Bayes algorithm is an intuitive technique that uses the conditional possibilities of every attribute belonging to every category to make a prediction. It uses Bayes' Theorem, a formulation that calculates a probability through counting the frequency of values and combinations of values in the historic data. Parameter estimation for naive Bayes models uses the technique of most likelihood. In spite over-simplified assumptions, it frequently performs higher in many complicated actual world situations. One of the main benefits of Naïve Bayes theorem is that it requires a small quantity of training data to estimate the parameters.
Algorithm:
INPUT

- Set of tuples = D
- Each Tuple is an 'n' dimensional attribute vector
- X : (x1,x2,x3,…. xn)

Let there be 'm' Classes: C1, C2, C3…Cm
Naïve Bayes classifier predicts X belongs to Class Ciiff

- P (Ci/X) > P(Cj/X) for 1<= j <= m , j <> i

Maximum Posteriori Hypothesis

- P(Ci/X) = P(X/Ci) P(Ci) / P(X) □ Maximize P(X/Ci) P(Ci) as P(X) is constant

With many attributes, it is computationally expensive to evaluate P(X/Ci). Naïve Assumption of "class conditional independence"
P (X/Ci) = ΠP (xk/Ci)=*l*
P(X/Ci) = P(x1/Ci) * P(x2/Ci) *…* P(xn/ Ci)
P(fraud) = yi / y = 7/30 = 0.23
P(legal) = yi / y =23/30 = 0.77
Bayesian networks ease many of the theoretical and computational difficulties of rule based totally structures through using graphical elements for representing and managing probabilistic knowledge. They also have the capacity to deal with incomplete data sets. Bayesian networks can predict the category label when only partial information about the attribute is available. Bayesian faith networks are very positive for modeling conditions where information about the previous and/or the modern-day state of affairs is vague, incomplete, conflicting, and uncertain, whereas rule-based models result in ineffective or inaccurate predictions when the data is unsure or unavailable.
BayesNet
The graphs that the BayesNet classifier (package weka.classifiers.bayes).
Generates can be displayed using the GraphVisualizer class (located in package.
weka.gui.graphvisualizer). The GraphVisualizer can display graphs that Are either in Graphic's DOT language or in XML BIF format. For displaying DOT format, one needs to use the method read DOT, and for the BIF format

the method read BIF.

3.3. P-SVM CLASSIFICATION

Principal component analysis (PCA) is a statistical technique that uses an orthogonal transformation to convert a set of observations of maybe correlated variables (entities every of which takes on a variety of numerical values) into a set of values of linearly uncorrelated variables known as principal components. Support vector machine (SVM) classification has been applied for the credit card fraud detection. However, if the index of the training data has a great deal noise and redundancy, the generalized overall performance of SVM will be weakened, so this can purpose some negative aspects of slow convergence speed and low classification accuracy. A SVM classification model based on principal component analysis (P-SVM) is in this research, the use of PCA (principal component analysis) is used to minimize the dimensionality of indexes, and then extract principal components to exchange the unique indexes. Applying this model P-SVM to credit card fraud detection, shows extra generalized overall performance and higher classification accuracy compared with the technique of SVM.

Support vector machine is a technique used in sample cognizance and classification. It is a classifier to predict or classify patterns into two categories, fraudulent or non fraudulent. It is properly suitable for binary classifications. As any synthetic talent tool, it has to be trained to achieve a discovered model. P-SVM has been used in many classification sample awareness troubles such as textual content categorization, bioinformatics and face detection [32]. P-SVM is correlated to and having the fundamentals of non-parametric utilized statistics, neural networks and machine learning. P-SVM is described in the following passage [33]. It's developed from the concept of Structural Risk Minimization.

$$f(x) = sgn(x, w) + \mathrm{b}$$

Where, x is the input vector which includes weight and b is a constant. It is used to discover the selection boundary between two classes. The parameter values of w and b have to be realized through the P-SVM on the training part and b are derived through maximizing the margin of separation between the two classes. The criterion used through P-SVM is based on the margin maximization between the two classes.

In statistical learning concept (SLT) the trouble of supervise getting to know is formulated as follows. There are given a set of one training data {(x1,y1)...(xl,yl)} in Rn × R sampled according to unknown likelihood distribution P(x,y), and a loss function V(y,f(x)) that measures the error carried out when, for a given x, f(x) is "predicted" rather of the real cost y. The problem consists in discovering a feature f that minimizes the expectation of the error on new data, that is, discover a feature f that minimizes the predicted error[34].

$$\int V(y, f(x))P(x, y)dx\,dy$$

Since P(x, y) is unknown, it wants to use some induction in order to infer from the one reachable training examples a feature that minimizes the predicted error. The principle used is Empirical Risk Minimization (ERM) over a set of feasible functions, known as speculation space. Formally this can be written as minimizing the empirical error.

$$\sum_{i-1}^{1} V(yi, f(xi))$$

A necessary query is how close the empirical error of the solution (minimize of the empirical error) is to the minimum of the predicted error that can be carried out with features from H. A central result of the concept states the prerequisites below which the two mistakes are close to every other, and presents probabilistic bounds on the distance between empirical and predicted mistakes. These bounds are given in terms of a measure of complexity of the speculation space H: the extra "complex" H is the large the distance between the empirical and predicted errors.

It can be confirmed that the real error on the future data is bounded through a sum of two terms. The first time period is the training error, and the second term if comparative to the rectangular root of the dimension. Thus, if it can reduce h, it can decrease the future error, as lengthy as it also reduce the training error, P-SVM can be without problems prolonged to function numerical calculations. Here it talks about two such extensions. Careful consideration goes into the preference of the error models; in guide vectorregression, or the error is described to be zero when the variations between real and expected values are inside an epsilon amount. Otherwise, the epsilon insensitive error will develop linearly. The aid vectors can then be discovered through the minimization.

A benefit of support vector regression is mentioned to be its insensitivity to outliers. One of the preliminary drawbacks of P-SVMis its computational inefficiency. However, this trouble is being solved with excellent success. One strategy is to damage a large optimization trouble into a collection of smaller problems, where every trouble only includes a couple of carefully chosen variables so that the optimization can be accomplished efficiently. The method iterates till all the decomposed optimization troubles are solved successfully [30]. Fig 3.1 shows the SVM Classification result.

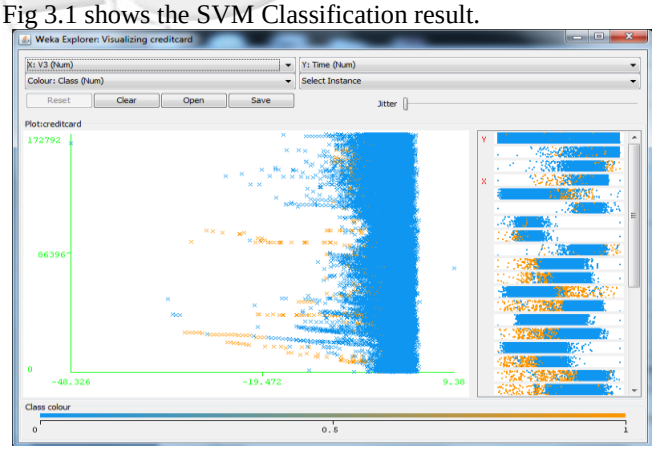

Fig. 3.1 P- SVM Result

IV. EXPERIMENTALRESULTS

4.1 CREDIT CARD FRAUD DATASETS

The datasets includes transactions made through credit cards in September 2018 through European cardholders. This dataset offers transactions that happened in two days, where it has 492 frauds out of 2,84,807 transactions. The dataset is incredibly unbalanced, the advantageous category (frauds)

account for 0.172% of all transactions.

It consists of only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, it can't provide the original features and more background information about the data set. Features V1, V2 ….. V28 (Table 4.1 shows the attributes name and description) are the principal components obtained with PCA, the only feature which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' includes the seconds elapsed between each transaction and the first transaction in the dataset.

The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise. Given the category imbalance ratio, it suggests measuring the accuracy of the usage of the Area under the Precision-Recall Curve (AUPRC). Confusion matrix accuracy is not significant for unbalanced classification.

The dataset has been accrued and analyzed throughout a lookup collaboration of World line and the Machine Learning Group (http://mlg.ulb.ac.be) of ULB (University Libre de Bruxelles) on large data mining and fraud detection. More details on modern-day and previous tasks on associated topics are accessible on http://mlg.ulb.ac.be/BruFence and http://mlg.ulb.ac.be/ARTML

Table: 4.1 Credit Card Transaction- Attributes name and description

Attributes name	Description
V1	Transaction ID
V2	Time
V3	Account number
V4	Card number
V5	Transaction type
V6	Entry mode
V7	Amount
V8	Mail Id
V9	Phone No2
V10	Country
V11	Country 2
V12	Type of card
V13	Gender
V14	Age
V15	Bank
V16	Transaction Amount
V17	Frequency of card usage
V18	Place
V19	Average amount of transactions per month
V20	Account holder name
V21	Bank code
V22	Bank address
V23	Duration in month
V24	Card secret no
V25	Credit history
V26	Purpose
V27	Telephone 1
V28	Savings account

4.3 RESULTS AND DISCUSSION

The performance of the model can be evaluated using various performance metrics. This measures performance of the algorithm using three performance metrics namely, accuracy, Precision Ratio (Specificity), Recall (Sensitivity). These metrics are calculated from the confusion matrix. The confusion matrix is a table that is used to predict the performance of a classification model on a sample set of data.

It is used for summarizing the result of a classifier. It is a matrix that shows the number of True Positives (TP), False Negative (FN), False Positives (FP), and True Negatives (TN). The format of the confusion matrix is shown in Table 4.2. Since classification need training and testing data, 70% of data is considered for training and 30% of data is considered for testing this is shown in Table 4.3.

Table 4.2 Confusion Matrix

Actual vs Predictive	**Correct Labels**	
	Positive	**Negative**
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

- True Positive: The fraud cases that the model predicted as 'fraud.'
- False Positive: The non-fraud cases that the model predicted as 'fraud.'
- True Negative: The non-fraud cases that the model predicted as 'non-fraud.'
- False Negative: The fraud cases that the model predicted as 'non-fraud.'

Threshold Cutoff Probability: Probability at which the true positive ratio and true negatives ratio are both highest. It can be noted that this probability is minimal, which is reasonable as the probability of frauds is low.

Accuracy: The measure of correct predictions made by the model – that is, the ratio of fraud transactions classified as fraud and non-fraud classified as non-fraud to the total transactions in the test data.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision Ratio: Precision is a ratio of correctly predicted fraud case to totally predicted fraud case.

$$\text{PositivePredictiveRatio} = \frac{TP}{TP + FP}$$

Recall: Recall is the ratio of correctly identified fraud case to totally fraud case.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Confusion matrix is calculated through the evaluation outcomes and with the help of confusion matrix precision, recall and f-measure is calculated. Analysis outcomes indicates True Positives(TP), False Positives(FP), True

Negatives(TN) and False Negatives(FN) from Table 4.2. Naive Bayes classifier work on the precept of possibilities and the Bayes rule given by: p (c/d) = (p(c)p(d/c))/p(d). Where P (c|d) is the likelihood of a given file (text) belongs to category c, which is the classification section which that are involved in. Below is the confusion matrix for the naive bayes classifier in this thesis. The classifier has received accuracy of 88.04 %. It can be seen that naïve bayes classifier has misclassified extra quantity of data factors as in contrast to P-SVM.

Features	Values in parentage %
Size of Training Set	70 %
Size of Testing Set	30 %
Classes	Fraud, Legal
Existing Methods	Decision Tree, Naïve Bayes

Table 4.3: The description of processing training and testing data set

The accuracy of this machanism comes out to be 94.37 % which is greater than exisiting algorithm. Naïve Bayes and Decision tree algorithm can deal with Noisy or Incomplete Data but not affected through outliers and error charge is excessive because of overfitting and becoming below when the pattern dimension is small. Decision Tree algorithm is excellent in contrast to others, data is categorized except many calculations. The accuracy of Decision Tree model is 89.57 %. The P-SVM algorithm is excellent for Larger Samples and Memory Efficient.It can be observed that scores of the above methods are relatively stable in ten experiments. In above three methods, the classification effects of Naive Bayesian are poor, and score is only around 88%.When using Decision treefor classification, score can reach around 89%. The proposed P-SVM classification for the given data, obtained a accuracy of 94.37% percentage and it outperforms the other two methods.

Fig.4.1 shows the represents the value comparison of Data Mining techniques such as Naïve Bayes, Decision tree and P-SVM with each other. These techniques are evaluated using the common measures such as precision, recall, and accuracy.

Table 4.4 Comparative table Decision Tree, Naïve Bayes and P-SVM

Data set	Classifier	F-Measure in Percentage	Recall in Percentage	Precision in Percentage	Accuracy in Percentage
Credit card	Naïve Bayes	85.93	86.07	84.82	88.04
	Decision Tree	87.41	88.56	86.22	89.57
	P-SVM	91.48	92.87	89.59	94.37

The X- axis in the chart diagram represents the classifier algorithms. The Y-axis represents the percentage of the common measures. Hence, False Positive (FP) and False Negative (FN) rate has been minimized; True Positive (TP) and True Negative (TN) rate has been maximized. To improve the accuracy of credit card fraud detection.

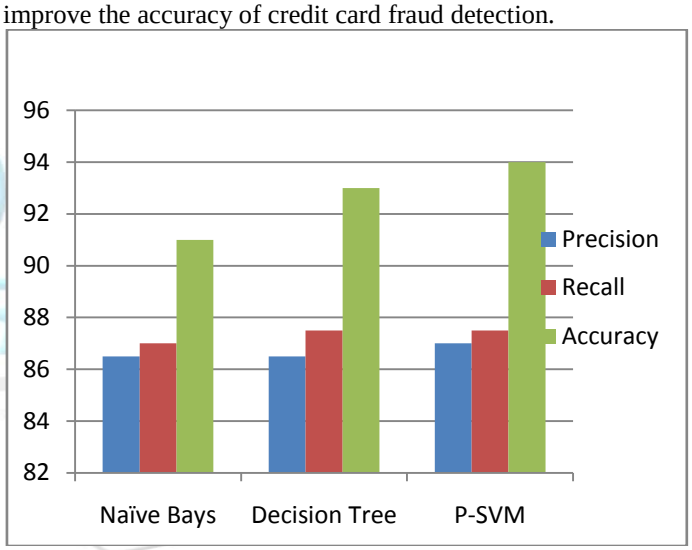

Fig. 4.1 Precision, Recall and Accuracy comparison

V.CONCLUSION

The credit card's expenditure behaviour and location analysis is the centre of attention in this research. An efficient classification model P-SVM is proposed to detect credit card fraud. The proposed model analyzed the credit card transaction data based on the expenditure behaviour, location on different age groups. The performance of the Decision Tree, Naïve bayes and P-SVM classification has been evaluated by using various measures such as Precision, Recall and Accuracy. The proposed P-SVM classification model outperforms the Naïve Bayes and Decision Tree algorithms.This model can be enhanced dynamically in training of classifiers, the usage of special SVM models like incremental SVM, detrimental SVM and evolutionary SVM etc.Rough Set theory and Dynamic feature selection method could be implemented in future. The system should be adaptable to new behavioral pattern in finding credit card fraud.

VI.REFERENCES

[1] CyberSource;"11th Annual Online Fraud Report"; 2010. http://forms.cybersource.com/forms/FraudReport2010NACYBSwwwQ109 last accessed on 2010/09/10.

[2] R. Brause, L. T., and M. Hepp, "Neural data mining for credit card fraud detection," *11th IEEE International Conference on Machine Learning and Cybernetics*, Vol7, 2008,pp.3630-3634.

[3] R. Chen, S. Luol, X. Liang, and V.C. Lee, "Personalized approach basedonSVMandANNfordetectingcreditcardfraud",*International Conference on Neural Networks and Brain*, 2005, pp.810-815.

[4] A. Shen, R. Tong, and Y. Deng, "Application of classification models on credit card fraud detection," *International Conference on Service Systems and Service Management*, June 2007, pp.1-4.

[5] M.F. Gadi, X. Wang, and A.P. Lago, "Comparison with parametric optimization in credit card fraud detection," *Seventh International Conference on Machine Learning and Applications*, 2008, pp. 279-285.

[6] C. Whitrow, D.J. Hand, P. Juszczak, D. Weston, and N.M. Adams, "Transaction aggregation as a strategy for credit card fraud detection," *Data Mining and Knowledge Discovery*, Vol.18, No. 1, 2009, pp.30-55.

[7] J.QuahandM.Sriganesh,"Real-timecreditcardfrauddetectionusing computational intelligence," *Expert Systems with Applications*, Vol. 35, No. 4, 2008, pp.1721-1732.

[8] S. Panigrahi, A. Kundu, S. Sural, and A. Majumdar, "Credit card frauddetection:AfusionapproachusingDempster–Shafertheoryand Bayesian learning," *Information Fusion*, Vol. 10, No. 4, 2009, p.9.

[9] J.Xu,A.H.Sung,andQ.Liu,"Behaviourminingforfrauddetection," *Journal of Research and Practice in Information Technology*, Vol.39, No. 1, 2007, pp.3-18.

[10] A. Srivastava, A. Kundu, S. Sural, and A. Majumdar, "Credit card fraud detection using Hidden Markov Model," *IEEE Transactions on Dependable and Secure Computing*, Vol. 5, No. 1, 2008, pp.37-48.

[11] A. Kundu, S. Sural, and A. Majumdar, "Two-stage credit card fraud detection using sequence alignment," *Information Systems Security*, Springer Berlin / Heidelberg, 2006, pp.260-275.

[12] D.J. Weston, D.J. Hand, N.M. Adams, C. Whitrow, and P. Juszczak, "Plastic card fraud detection using peer group analysis," *Advances in Data Analysis and Classification*, Vol. 2, No. 1, 2008, pp.45-62.

