International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072

CREDIT CARD FRAUD DETECTION USING HYBRID MODELS

ASWATHY M S, LIJI SAMEUL

[1]*ASWATHY M S M.Tech Computer Science & Engineering. Sree Buddha College of Engineering, Ayathil, Elavumthitta Pathanamthitta , Kerala, India.*
[2]*Ms. LIJI SAMEUL Assistant Professor Computer Science & Engineering. Sree Buddha College of Engineering, Ayathil, Elavumthitta, Pathanamthitta, Kerala, India*

Abstract - *Charge card extortion is a difficult issue in money related administrations. Billions of dollars are lost because of charge card misrepresentation consistently. There is an absence of research thinks about on examining genuine charge card information inferable from privacy issues. In this paper, AI calculations are utilized to recognize charge card extortion. Standard models are first utilized. At that point, half breed techniques which use AdaBoost and larger part casting a ballot strategies are connected. To assess the model adequacy, an openly accessible charge card informational index is utilized. At that point, a true Credit card informational collection from a monetary establishment is examined. What's more, clamor is added to the information tests to additionally survey the heartiness of the calculations. The exploratory outcomes decidedly demonstrate that the lion's share casting a ballot technique accomplishes great exactness rates in distinguishing extortion cases in Visas. Misrepresentation is an illegitimate or criminal trickiness meant to bring money related or individual increase. In evading misfortune from extortion, two instruments can be utilized: misrepresentation counteractive action and extortion location. Misrepresentation counteractive action is a proactive technique, where it prevents extortion from occurring in any case. Then again, misrepresentation identification is required when a deceitful exchange is endeavored by a fraudster.*

Key Words: CCF, FRAUD DETECTION

1. INTRODUCTION

Generally, fraud "is the act of deceiving to gain unfair, undeserved and/or illegal financial profit". Fraud detection is an important issue in many areas including credit loans, credit cards, long distance communications and insurance. Any attempt to detect fraud in these areas is called a fraud detection process. In banking, fraud happens in credit cards, online bank accounts, and call centers (telephone banking). The sooner the fraudulent transactions are detected, more damages can be prevented by stopping the transactions of counterfeit credit cards. There are two main and important types of frauds related to credit cards. The first one is counterfeit fraud, which is done by organized crime gangs. The second type of credit card fraud is the illegal use of a missing or stolen credit card.

Fraud detection is one of the best applications of data mining in the industry and the government. Statistical methods of fraud detection are divided into two broad categories, supervised and unsupervised. Traditional fraud detection is very costly due to expensive experts and broadness of the databases. Another deficiency is that not every human expert is able to detect the most recent patterns of fraud. Thus a data mining algorithm should analyze huge databases of transactions, and only then the expert will be able to do a further investigation about the diagnosed risky measures. Credit card fraud detection is an incredibly troublesome, yet in addition famous issue to illuminate. There comes just a restricted measure of information with the exchange being submitted. Additionally, there can be past exchanges made by fraudsters which likewise fit an example of typical conduct. Besides the issue has numerous limitations. As a matter of first importance, the profiles of typical and fake practices change always. Besides, the advancement of new extortion discovery strategies is made increasingly troublesome by the way that the trading of thoughts in misrepresentation location, particularly in Visa extortion recognition is seriously constrained because of security and protection concerns. Thirdly, informational indexes are not made accessible and the outcomes are frequently blue-penciled, making them hard to evaluate. Indeed, a portion of the investigations are finished utilizing artificially produced information. Fourthly, Visa extortion informational indexes are profoundly skewed sets. Finally, the informational collections are additionally continually advancing making the profiles of ordinary and deceitful practices continually evolving. In this way, charge card misrepresentation identification is as yet a famous testing and hard research point. Visa reports about charge card fakes in European nations express that about half of the entire Credit card misrepresentation misfortunes in 2008 are because of online fakes. Numerous papers announced immense measures of misfortunes in various nations. Along these lines new methodologies improving the classifier execution in this area have both money related ramifications and research commitments. Characterizing another cost-delicate methodology is a standout amongst the most ideal ways for such an improvement because of the attributes of the area. Misrepresentation discovery includes distinguishing rare extortion exercises among various genuine exchanges as fast as could be expected under the circumstances. Extortion recognition techniques are growing quickly so as to adjust with new approaching false methodologies over the world. Be that as it may, improvement of new misrepresentation recognition strategies turns out to be progressively

troublesome because of the extreme confinement of the thoughts trade in extortion location. Then again, extortion discovery is basically an uncommon occasion issue, which has been differently called exception investigation, peculiarity location, special case mining, mining uncommon classes, mining imbalanced information and so forth. The quantity of deceitful exchanges is normally an exceptionally low division of the complete exchanges. Consequently the undertaking of recognizing misrepresentation exchanges in an exact and proficient way is genuinely troublesome and challengeable. In this way, improvement of effective techniques which can recognize uncommon extortion exercises from billions of authentic exchange appears to be fundamental.

Fraud detection systems are prune to several difficulties and challenges enumerated bellow. An effective fraud detection technique should have abilities to address these difficulties in order to achieve best performance.

1.1 Objective

The objective of the proposed system is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications. Credit card fraud is concerned with the illegal use of credit card information for purchases. The Credit Card Fraud Detection Problem includes modelling past credit card transactions with the knowledge of the ones that turned out to be fraud. This model is then used to identify whether a new transaction is fraudulent or not. Machine Learning are used for detecting fraud. These algorithms can be used either stand alone or can be combined together.

2. METHODOLOGY

2.1 Existing System

Credit card fraud is worried about the illicit utilization of charge card data for buys. Credit card exchanges can be cultivated either physically or carefully. In physical exchanges, the charge card is included amid the exchanges. In computerized exchanges, this can occur via phone or the web. Cardholders regularly give the card number, expiry date, and card confirmation number through phone or site. With the ascent of internet business in the previous decade, the utilization of Credit cards has expanded drastically. Misfortune from Visa extortion influences the vendors, where they bear all costs, including card guarantor expenses, charges, and authoritative charges. Since the traders need to tolerate the misfortune, a few products are estimated higher, or limits and motivations are decreased. Along these lines, it is basic to lessen the misfortune, and a successful extortion discovery framework to decrease or take out misrepresentation cases is significant.

2.1.1 Disadvantages

- Imbalanced information: The Visa misrepresentation identification information has imbalanced nature. It implies that exceptionally little rates of all Visa exchanges are deceitful. This reason the location of misrepresentation exchanges extremely troublesome and uncertain.
- Different misclassification significance: In misrepresentation discovery task, diverse misclassification mistakes have distinctive significance. Misclassification of an ordinary exchange as misrepresentation isn't as destructive as identifying an extortion exchange as typical. Since in the main case the slip-up in characterization will be recognized in further examinations.
- Overlapping information: numerous exchanges might be viewed as deceitful, while really they are ordinary (false positive) and conversely, a fake exchange may likewise appear to be authentic (false negative). Henceforth getting low rate of false positive and false negative is a key test of extortion discovery frameworks.
- Lack of flexibility: characterization calculations are typically looked with the issue of distinguishing new sorts of ordinary or fake examples. The directed and unsupervised extortion location frameworks are wasteful in distinguishing new examples of ordinary and misrepresentation practices, individually.
- Fraud identification cost: The framework should consider both the expense of false conduct that is recognized and the expense of forestalling it. For instance, no income is acquired by ceasing a fake exchange of a couple of dollars.
- Lack of standard measurements: there is no standard assessment foundation for surveying and contrasting the aftereffects of extortion discovery frameworks.

2.2 Proposed System

A study of credit card fraud detection using machine learning algorithms have been proposed. Machine learning algorithms like Random Forest, Decision Tree, Bayesian Learning and Convolutional Neural Network is being used. Naïve Bayes (NB) uses the Bayes' theorem with strong or naïve independence assumptions for classification. Certain features of a class are assumed to be not correlated to others. It requires only a small training data set for estimating the means and variances is needed for classification. The presentation of data in form of a tree structure is useful for ease of interpretation by users. The Decision Tree (DT) is a collection of nodes that creates decision on features connected to certain classes. Every node represents a splitting rule for a feature. New nodes are established until the stopping criterion is met. The class label is determined based on the majority of samples that belong to a particular leaf. The Random Forest (RF) creates an ensemble of random trees. The user sets the number of trees. The resulting model employs voting of all created trees to determine the final classification outcome. The MLP network consists of at least three layers of nodes, i.e., input, hidden, and output. Each node uses a non-linear activation function, with the exception of the input nodes. It uses the supervised back propagation

algorithm for training. The version of MLP used in this study is able to adjust the learning rate and hidden layer size automatically during training. It uses an ensemble of networks trained in parallel with different rates and number of hidden units. These algorithms are being used as single models and as an enhancement these algorithms are also used in hybrid forms. Majority voting is frequently used in data classification, which involves a combined model with at least two algorithms. Each algorithm makes its own prediction for every test sample. The final output is for the one that receives the majority of the votes. Adaptive Boosting or AdaBoost is used in conjunction with different types of algorithms to improve their performance. These algorithms are evaluated with precision, recall and accuracy and their corresponding graphs are plotted.

3. SYSTEM DESIGN

Credit card fraud detection system can be divided into three parts:

- Master Data Manager
- Public
- Fraud Detection

3. 1 MASTER DATA MANAGER

The main part of the proposed system is the master data manger module or the admin module. The administrator controls the whole part of the system. It manages various credit card types, Credit Card Company, Vendor management and Data set management. Initially a user should register to the system and after that the particular user can request for credit card. Depending on their financial status the administrator can either accept or deny their request. Various credit card types like plain vanilla, etc are listed in this section and the customer can select the particular card with proper credit limit and interest limit. The credit limit and interest limit of different cards will be varying according to the standard of the card. Credit Card Company includes various banks that provides money and it also suggest their approved cards. In Vendor management various vendors can suggest their particular products and services. In this various shopping companies, water and electricity services are coded as vendors. In Data Set Management, the transactions occurred in the payment part are converted into data set. Apart from that a real data set is also uploaded for fraud detection. The dataset is initially converted into Arff file and after that the machine learning algorithms like Naïve Bayes, Decision Tree, Random Forest and Convolutional Neural Network are applied. A comparison of these algorithms are made based on precision, recall and accuracy values. In addition to improve the performance for detecting fraud Adaptive Boosting algorithm is also used. After this Majority Voting algorithm is also used as a combination of two algorithms where each classifier makes its own prediction. An application for credit card is processed and the company can determine whether a card should be provided for the concerned person depending on the background details like annual income. After checking back the details bank can either approve or reject the credit card application. Different bank make different scale of income for accepting the cards. So it is the responsibility of the user to check out which bank is suitable for them. The approved user will get a credit card number also. Mapping is also done in this section in which training data and testing data are mapped. The testing data is converted into data set during this section.

3. 2 PUBLIC

The public module is mainly for different users. It includes Credit Card Application, Credit card payment and prediction. The public can register and apply for various credit cards like plain vanilla, Balance transfer, rewards etc. as they wish. The credit cards are different based on cash limit and interest limit. The user can choose particular cards provided by the company. Another important step is payment part, Credit card is needed to pay bills either for shopping purpose or service bill payment. The payment is processed with proper authentication. At the time of payment an OTP is send to the registered users email. When the user enters the particular OTP in payment section then only the complete payment occurs. The user is also authenticated with a particular username and password. Another facility in public module is the Alert View. In this proper alerts will be send by Credit Card Company or admin in case of any problem. For example if the user hasn't pay back the amount withdrawed proper alerts will be provided and also the balance amount alerts will also be given to the user to the registered numbers. The output of fraud detection will be displayed here which helps to trace fraudulent credit card. If the occurred transaction turned out to be fraud then that message will be delivered to the particular credit card company. Then the company should take further steps to prevent this fraud. If the occurred transaction turned out to be normal then that action will also be reported to the particular credit card company.

3.3 FRAUD DETECTION

Fraud Detection module mainly includes Prediction using classifiers and Evaluation. In prediction part classifiers are applied. The classifiers like Random forest, Bayesian, Decision Trees and CNN are applied and their corresponding precision, recall and accuracy are calculated. In order to improve the features hybrid models like Majority voting and Adaboost are also applied. Evaluation part includes the prediction using classifiers. The performance of enhanced method with an existing method is also evaluated. The time complexity of each algorithm is also evaluated.

4. ALGORITHMS

4.1 NAIVE BAYES

Naive Bayes classifiers are an accumulation of order calculations dependent on Bayes' Theorem. It's anything but a solitary calculation however a group of calculations where every one of them share a typical standard, for example each pair of highlights being characterized is free of one another. Bayes' Theorem finds the likelihood of an occasion happening given the likelihood of another occasion that has just happened. Bayes' hypothesis is expressed numerically as the accompanying condition:

P(A|B) = (P(A)P(B|A))/(P(B)) where An and B are occasions and P(B) ? 0. Fundamentally, we are endeavoring to discover likelihood of occasion A, given the occasion B is valid. Occasion B is additionally named as proof. P(A) is the priori of A (the earlier likelihood, for example Likelihood of occasion before proof is seen). The proof is a characteristic estimation of an obscure example (here, it is occasion B). P(A|B) is a posteriori likelihood of B, for example likelihood of occasion after proof is seen. Innocent Bayes classifiers are very adaptable, requiring various parameters direct in the quantity of factors (highlights/indicators) in a learning issue. Greatest probability preparing should be possible by assessing a shut structure articulation, which takes direct time, as opposed to by costly iterative estimate as utilized for some different kinds of classifiers.

4.2 DECISION TREES

A decision tree is a choice help instrument that utilizes a tree-like model of choices and their potential results, including chance occasion results, asset expenses, and utility. It is one approach to show a calculation that just contains restrictive control explanations. Choice trees are regularly utilized in tasks look into, explicitly in choice examination, to help distinguish a methodology destined to achieve an objective, but on the other hand are a well known device in AI. A choice tree is a flowchart-like structure in which each inward hub speaks to a "test" on a property (for example regardless of whether a coin flip comes up heads or tails), each branch speaks to the result of the test, and each leaf hub speaks to a class mark (choice taken in the wake of registering all properties). The ways from root to leaf speak to grouping rules. In choice examination, a choice tree and the firmly related impact graph are utilized as a visual and scientific choice help apparatus, where the normal qualities (or anticipated utility) of contending options are determined. A decision tree comprises of three kinds of hubs:

- Decision hubs – normally spoken to by squares
- Chance hubs – normally spoken to by circles
- End hubs – normally spoken to by triangles

Decision trees are ordinarily utilized in tasks research and activities the executives. On the off chance that, practically speaking, choices must be taken online with no review under deficient learning, a choice tree ought to be paralleled by a likelihood model as a best decision model or online choice model calculation. Another utilization of choice trees is as a spellbinding methods for ascertaining contingent probabilities.

4.3 RANDOM FOREST

Random Forest or random Decision Forest are a gathering learning strategy for characterization, relapse and different undertakings that works by building a huge number of choice trees at preparing time and yielding the class that is the method of the classes (grouping) or mean forecast (relapse) of the individual trees. Arbitrary choice backwoods right for choice trees' propensity for over fitting to their preparation set. The preparation calculation for arbitrary backwoods applies the general system of bootstrap amassing, or packing, to tree students. This bootstrapping methodology prompts better model execution since it diminishes the difference of the model, without expanding the inclination. This implies while the expectations of a solitary tree are very delicate to clamor in its preparation set, the normal of numerous trees isn't, the length of the trees are not corresponded. Essentially preparing numerous trees on a solitary preparing set would give emphatically corresponded trees (or even a similar tree commonly, if the preparation calculation is deterministic); bootstrap testing is a method for de-relating the trees by appearing changed preparing sets. Arbitrary timberlands contrast in just a single path from this general plan: they utilize a changed tree learning calculation that chooses, at every competitor split in the learning procedure, an irregular subset of the highlights. This procedure is now and again called "highlight sacking".

4.4 CONVOLUTIONAL NEURAL NETWORK

CNNs are regularized variants of multilayer perceptron's. Multilayer perceptron's generally allude to completely associated systems, that is, every neuron in one layer is associated with all neurons in the following layer. The "completely connectedness" of these systems make them inclined to over fitting information. Run of the mill methods for regularization incorporates including some type of greatness estimation of loads to the misfortune work. Be that as it may, CNNs adopt an alternate strategy towards regularization: they exploit the various leveled design in information and gather increasingly complex examples utilizing littler and less complex examples. In this manner, on the size of connectedness and multifaceted nature, CNNs are on the lower extraordinary. Convolutional systems were roused by natural procedures in that the network design between neurons looks like the association of the creature visual cortex. Individual cortical neurons react to improvements just in a limited locale of the visual field known as the open field. The responsive fields of various

neurons halfway cover with the end goal that they spread the whole visual field. They have applications in picture and video acknowledgment, recommender frameworks, picture characterization, therapeutic picture examination, and normal language preparing. A convolutional neural system comprises of an info and a yield layer, just as different shrouded layers. The concealed layers of a CNN ordinarily comprise of convolutional layers, RELU layer for example actuation work, pooling layers, completely associated layers and standardization layers. Depiction of the procedure as a convolution in neural systems is by show. Numerically it is a cross-relationship instead of a convolution (albeit cross-connection is a related activity). This just has importance for the files in the framework, and along these lines which loads are set at which record. Each convolutional neuron forms information just for its open field. Albeit completely associated feed forward neural systems can be utilized to learn includes just as arrange information, it isn't reasonable to apply this engineering to pictures. A high number of neurons would be essential, even in a shallow (inverse of profound) engineering, because of the enormous info sizes related with pictures, where every pixel is an important variable. For example, a completely associated layer for a (little) picture of size 100 x 100 has 10000 loads for every neuron in the second layer. The convolution task conveys an answer for this issue as it lessens the quantity of free parameters, enabling the system to be more profound with less parameters.

4.5 MAJORITY VOTING

The Boyer–Moore majority vote algorithm is a calculation for finding most of an arrangement of components utilizing straight time and steady space. In its least complex structure, the calculation finds a dominant part component, if there is one: that is, a component that happens over and again for the greater part of the components of the information. In any case, if there is no dominant part, the calculation won't identify that reality, will in any case yield one of the components.. The calculation won't, as a rule, discover the method of a succession (a component that has the most reiterations) except if the quantity of redundancies is sufficiently huge for the mode to be a greater part. It isn't workable for a spilling calculation to locate the most regular component in under direct space, when the quantity of reiterations can be small.The calculation keeps up in its nearby factors a grouping component and a counter, with the counter at first zero. It at that point forms the components of the arrangement, each one in turn. When preparing a component x, if the counter is zero, the calculation stores x as its recollected succession component and sets the counter to one. Else, it looks at x to the put away component and either augments the counter (on the off chance that they are equivalent) or decrements the counter (generally). Toward the finish of this procedure, if the grouping has a lion's share, it will be the component put away by the calculation. This can be communicated in pseudo code as the accompanying advances:

- Initialize an element m and a counter i with i = 0
- For each element x of the input sequence:
  - If i = 0, then assign m = x and i = 1
  - else if m = x, then assign i = i + 1
  - else assign i = i − 1
- Return m

Notwithstanding when the information succession has no larger part, the calculation will report one of the grouping components as its outcome. In any case, it is conceivable to play out a second ignore a similar info grouping so as to tally the occasions the announced component happens and decide if it is really a greater part. This second pass is required, as it isn't workable for a sub straight space calculation to decide if there exists a dominant part component in a solitary go through the input.

4.6 ADAPTIVE BOOSTING

Adaptive Boosting or AdaBoost is utilized related to various kinds of calculations to improve their exhibition. AdaBoost is versatile as in ensuing feeble students are changed for those examples misclassified by past classifiers. AdaBoost is touchy to uproarious information and exceptions. In certain issues it tends to be less powerless to the over fitting issue than other learning calculations. The individual students can be feeble, however as long as the presentation of every one is marginally superior to irregular speculating, the last model can be demonstrated to combine to a solid student.

5. RESULT AND ANALYSIS

This section discusses the experimental results of the regularization of the classifier model and prediction model for credit card fraud detection. The system that uses the operating system for windows 10 and windows platforms here is c#.net. And the database created is a SQL server. The proposed system is using synthetic data for results assessment. Synthetic data is developed data. The synthetic data is created to attain specific needs or specific criteria that may not be establish in the original real data. Synthesizing data is very helpful for designing any type of system because this data can be used as a simulation. The proposed system is implemented using three modules and different sub-modules. The administrator controls the whole part of the system. It manages various credit card types, Credit Card Company, Vendor management and Data set management. The main module is the Fraud detection part. It includes Prediction using classifiers, Enhanced Neural Network and Evaluation. In prediction part classifiers are applied. The classifiers like Random forest, Bayesian and Decision Trees are applied and their corresponding precision, recall and accuracy are calculated. In addition CNN, majority voting and Adaboost is also used. After all these machine learning algorithms the system returns whether the transaction is fraudulent or not. The analysis of the proposed system performed are:

- Prediction Using Classifiers
- CNN and Adaboost CNN
- Time Complexity

Prediction using classifiers shows the precision, recall and accuracy values for Bayesian classifiers, Decision Trees, Random Forest, Convolutional Neural Network and Adaboost algorithm. The result is shown in Fig 5.1.

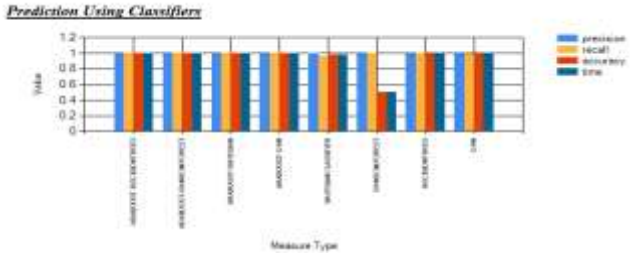

Fig 5.1 Prediction Using Classifiers

CNN and Adaboost CNN graph shows the variation in precision, recall and accuracy values. The result is shown in Fig 5.2.

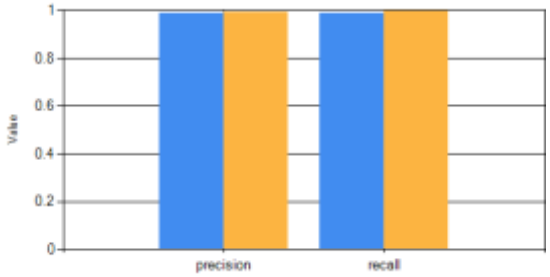

Fig 5.2: CNN- AdaBoost CNN

Time Complexity determines the time taken to perform the algorithm and the result is shown in Fig 5.3.

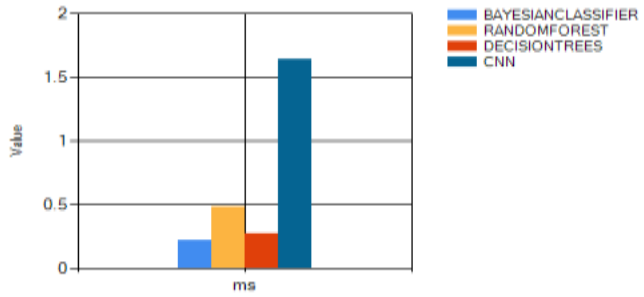

Fig 5.3 Time Complexity

CONCLUSIONS

The authors can acknowledge any person/authorities in this section. This is not mandatory. The proposed system is a credit card fraud detection system using hybrid models. In this machine learning algorithms are used to detect fraud. Algorithms like Naïve Bayes, Decision Tree, Random Forest and Convolutional Neural Network are used. These algorithms are used as single models. Then these are used as hybrid models using majority voting technique. In order to boost the performance of these classifiers adaptive boosting algorithm is also used. A publicly available credit card data set has been used for evaluation using individual (standard) models and hybrid models using AdaBoost and majority voting combination methods.

As a future work the work can be extended to online model. The use of online learning will enable rapid detection of fraud cases, potentially in real-time. This in turn will help detect and prevent fraudulent transactions before they take place, which will reduce the number of losses incurred every day in the financial sector.

REFERENCES

[1] Y. Sahin, S. Bulkan, and E. Duman, ``A cost-sensitive decision tree approach for fraud detection,'' Expert Syst. Appl., vol. 40, no. 15.

[2] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, ``Data mining for credit card fraud: A comparative study,'' Decision Support Syst., vol. 50, no. 3, 2011.

[3] A. Srivastava, A. Kundu, S. Sural, and A. Majumdar, ``Credit card fraud detection using hidden Markov model,'' IEEE Trans. Depend. Sec. Com- put., vol. 5, no. 1, Jan. 2008.

[4] J. T. Quah and M. Sriganesh, ``Real-time credit card fraud detection using computational intelligence,'' Expert Syst. Appl., vol. 35, no. 4, 2008.

[5] N. S. Halvaiee and M. K. Akbari, ``A novel model for credit card fraud detection using artificial immune systems,'' Appl. Soft Comput., vol. 24,2014.

[6] S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, ``Credit card fraud detection: A fusion approach using Dempster Shafer theory and Bayesian learning,'' Inf. Fusion, vol. 10, no. 4,2009.

[7] N. Mahmoudi and E. Duman, ``Detecting credit card fraud by modified fisher discriminant analysis,'' Expert Syst. Appl., vol. 42, no. 5, 2015.

[8] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, ``Association rules applied to credit card fraud detection,'' Expert Syst. Appl., vol. 36, no. 2, 2009.

[9] F. H. Glancy and S. B. Yadav, ``A computational model for financial reporting fraud detection,'' Decision Support Syst., vol. 50, no. 3, 2011.

[10] I. T. Christou, M. Bakopoulos, T. Dimitriou, E. Amolochitis, S. Tsekeridou, and C. Dimitriadis, ``Detecting fraud in online games of chance and lotteries,'' Expert Syst. Appl., vol. 38, no. 10, 2011.

[11] C.-F. Tsai, ``Combining cluster analysis with classifier ensembles to predict financial distress,'' Inf. Fusion, vol. 16, pp. 4658, Mar. 2014.

[12] F. H. Chen, D. J. Chi, and J. Y. Zhu, ``Application of random forest, rough set theory, decision tree and neural network to detect financial statement fraud-Taking corporate governance into consideration,'' in Proc. Int. Conf. Intell. Comput., 2014.

BIOGRAPHIES

Aswathy M S, She is currently pursuing her Masters degree in Computer Science and Engineering in Sree Buddha College Of Engineering, Kerala, India. Her area of research include Intelligence, Data Mining and Security

Liji Sameul, She is an Assistant Professor in the Department of Computer Science and Engineering, Sree Buddha College Of Engineering. Her main area of interest is Data Mining.