Detecting Credit Card Fraud by Decision Trees and Support Vector Machines

Y. Sahin and E. Duman

***Abstract*— With the developments in the Information Technology and improvements in the communication channels, fraud is spreading all over the world, resulting in huge financial losses. Though fraud prevention mechanisms such as CHIP&PIN are developed, these mechanisms do not prevent the most common fraud types such as fraudulent credit card usages over virtual POS terminals or mail orders. As a result, fraud detection is the essential tool and probably the best way to stop such fraud types. In this study, classification models based on decision trees and support vector machines (SVM) are developed and applied on credit card fraud detection problem. This study is one of the firsts to compare the performance of SVM and decision tree methods in credit card fraud detection with a real data set.**

***Index Terms*— Credit card fraud detection, decision trees, Support Vector Machines, classification**

I. Introduction

Fraud can be defined as wrongful or criminal deception intended to result in financial or personal gain [1], or to damage another individual without necessarily leading to direct legal consequences. The two main mechanisms to avoid frauds and losses due to fraudulent activities are fraud prevention and fraud detection systems. Fraud prevention is the proactive mechanism with the goal of disabling the occurrence of fraud. Fraud detection systems come into play when the fraudsters surpass the fraud prevention systems and start a fraudulent transaction. Nobody can understand whether a fraudulent transaction has passed the prevention mechanisms. Accordingly, the goal of the fraud detection systems is to check every transaction for the possibility of being fraudulent regardless of the prevention mechanisms, and to identify fraudulent ones as quickly as possible after the fraudster has begun to perpetrate a fraudulent transaction. A review of the fraud detection can be found in [2-5]. The most well-known fraud types are fraudulent transactions in credit card systems and e-commerce systems, money laundering in financial systems, intrusions to computer systems, fraudulent calls or service usages in telecommunication systems, and fraudulent claims in health and auto insurance systems.

With the developments in the Information Technology and improvements in the communication channels, fraud is spreading all over the world with results of huge financial losses. Though fraud can be perpetrated through many types of media, including mail, wire, phone and the Internet, online media such as Internet are the most popular ones. Because of the international availability of the web and ease with which users can hide their location and identity over Internet transactions, there is a rapid growth of committing fraudulent actions over this medium. Furthermore, with the improvements in the bandwidth of internetworking channels, fraudsters have the chance to form fraud networks among themselves through information change and collaboration all over the world. As a result, frauds committed over Internet such as online credit card frauds become the most popular ones because of their nature.

Credit card frauds can be made in many ways such as simple theft, application fraud, counterfeit cards, never received issue (NRI) and online fraud (where the card holder is not present). In online fraud, the transaction is made remotely and only the card's details are needed. A manual signature, a PIN or a card imprint are not required at the time of purchase. Though prevention mechanisms like CHIP&PIN decrease the fraudulent activities through simple theft, counterfeit cards and NRI; online frauds (Internet and mail order frauds) are still increasing in both amount and number of transactions. There has been a growing amount of financial losses due to credit card frauds as the usage of the credit cards become more and more common. Many papers reported huge amounts of losses in different countries [2, 6-7]. According to Visa reports about European countries, about 50% of the whole credit card fraud losses in 2008 are due to online frauds.

Credit card fraud detection is an extremely difficult, but also popular problem to solve. Firstly, there comes only a limited amount of data with the transaction being committed, such as transaction amount, date and time; address, Merchant Category Code (MCC) and acquirer number of the merchant. There are millions of possible places and e-commerce sites to use a credit card which makes it extremely difficult to match a pattern. Also, there can be past transactions made by fraudsters which also fit a pattern of normal (legitimate) behavior [8]. Also the problem has many constraints. First of all, the profile of normal and fraudulent behavior changes constantly. Secondly, the development of new fraud detection methods is made more difficult by the fact that the exchange of ideas

Manuscript received December 08, 2010; revised January 15, 2011. This work was supported in part by The Scientific and Research Council of Turkey (TUBITAK) under Grant 3100018 (TUBITAK TEYDEB 1501 Program).

Y. Sahin is with the Department of Electrical & Electronics Engineering, Marmara University, Kadikoy, 34722, Istanbul, Turkey (phone: +90-533-5566217; fax: +90-216-3480293; e-mail: ysahin@marmara.edu.tr).

E. Duman is with the Department of Industrial Engineering, Dogus University, Acibadem, Istanbul Turkey (phone: +90-216-5445555; fax: +90-216-3279631; e-mail: eduman@dogus.edu.tr).

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

in fraud detection, especially in credit card fraud detection is severely limited due to security and privacy concerns. Thirdly, data sets are not made available and the results are often censored, making them difficult to assess. Because of this problem, there is no chance of benchmarking for the models built. Even, some of the studies are done using synthetically generated data [9-10]. None of the previous studies with real data in the literature give details about the data and the variables used in classifier models. Fourthly, credit card fraud data sets are highly skewed sets with a ratio of about 10000 legitimate transactions to a fraudulent one. Lastly, the data sets are also constantly evolving making the profiles of normal and fraudulent behaviors always changing [2-4].

Techniques used in fraud detection can be divided into two: Supervised techniques where past known legitimate/fraud cases are used to build a model which will produce a suspicion score for the new transactions; and unsupervised ones where there are no prior sets in which the state of the transactions are known to be fraud or legitimate. A brief discussion of supervised and unsupervised techniques can be found in [2]. Many of these techniques like artificial neural networks can be used in both manners, supervised or unsupervised.

The most commonly used fraud detection methods are rule-induction techniques, decision trees, neural networks, Support Vector Machines (SVM), logistic regression, and meta-heuristics such as genetic algorithm, k-means clustering and nearest neighbor algorithms. These techniques can be used alone or in collaboration using ensemble or meta-learning techniques to build classifiers. From the well-known decision tree techniques, ID3, C4.5 and C&RT are used for credit card fraud detection in [11-14]. Also, SVM is used in [10, 15] for detecting credit card frauds.

Fraud detection systems evaluate the transactions and produce a suspicion score (generally a probability between 0 and 1) which shows the possibility of that transaction to be fraudulent. Computational procedures of these scores are relevant to the techniques used to build the model/models in the fraud detection systems. These scores are used with a predefined threshold value to differentiate the fraudulent transactions from the legitimate ones. However, most of the time, these scores are not directly used; but help the observer staff with domain expertise who examine and try to identify the frauds. Because the organizations have limited staff for this process, the ability of the detection systems to produce accurate suspicion scores helps these staff in many ways. Nevertheless, the success of the detection systems lies in distinguishing the fraudulent transactions from legitimate ones through producing suspicion scores with high precisions.

In this study, a credit card fraud detection system based on a number of decision tree and SVM methods is developed. In this system, each account is monitored separately using suitable descriptors, and the transactions are attempt to be identified and flagged as legitimate or normal. The identification will be based on the suspicion score produced by the classifier models developed. When a new transaction is going, the classifier can predict whether the transaction is normal or fraud. To the best of our knowledge, there is no prior work on the comparison of performance of decision tree and SVM based classifiers in the subject of credit card fraud detection with a real data set.

The rest of this paper is organized as follows: Section 2 gives some insights to the structure of credit card data. Section 3 is a summary of the classification methods used to develop the classifier models of the credit card fraud detection system given in this paper. Section 4 is about the details of our approach. Section 5 gives the results and a short discussion about the results. Section 6 concludes the study and shows directions for future work.

II. Structure of the Credit Card Data

The credit card data used in this study are taken from a national bank's credit card data warehouses with the required permissions. The past data in the credit card data warehouses are used to form a data mart representing the card usage profiles of the customers. Though some of the customers may have more than one credit card, each card is taken as a unique profile because customers with more than one card generally use each card for a different purpose. Every card profile consists of variables each of which discloses a behavioral characteristic of the card usage. These variables may show the spending habits of the customers with respect to geographical locations, days of the month, hours of the day or Merchant Category Codes (MCC) which show the type of the merchant where the transaction takes place. Later on, these variables are used to build a model to be used in the fraud detection systems to distinguish fraudulent activities which show significant deviations from the card usage profile stored in the data-mart.

The number of transactions for each card differs from one to other; however, each transaction record is of the same fixed length and includes the same fields. Hand and Blunt gave a detailed description of the characteristics of credit card data [16]. These fields range from the date and hour of the transaction to the amount, transaction type, MCC code, address of the merchant where the transaction is done and etc. The date and hour of the transaction record shows when the transaction is made. Transaction type shows whether this transaction is a purchase or a cash-advance transaction. MCC code shows the type of the merchant store where the transaction takes place. These are fixed codes given by the members of the VISA International Service Association. However; however, many of these codes form natural groups. So, instead of working with hundreds of codes, we grouped them into 25 groups according to their nature and the risk of availability to commit a fraud. The goods or services bought from merchant stores in some MCC codes can be easily converted to cash. As a result, transactions belonging to these MCC codes are more open to fraud and more risky from the transactions belonging to others. The grouping of the MCC codes are done according to both the number of the fraudulent transactions made belonging to each MCC code and the interviews done with the personnel of the data supplier bank with domain expertise about the subject.

The distribution of the data with respect to being normal

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

or fraudulent is highly imbalanced with a ratio of about 20000 normal transaction records to one fraudulent transaction record. So, to enable the models to learn both types of profiles, some under sampling or oversampling techniques should be used. Instead of oversampling the fraudulent records by making multiple copies or etc., we use stratified sampling to under sample the legitimate records to a meaningful number.

III. Classification Methods

There are a lot of studies done on credit card fraud detection. The general background of the credit card systems and non-technical knowledge about this type of fraud can be learned from [17] and [18]. Most of the credit card fraud detection systems are using supervised algorithms such as neural networks and SVMs [7, 9, 11, 15, 19-24]. In this study, the performance of classifier models built by using the well-known decision tree methods C5.0, C&RT and CHAID and a number of different SVM methods (SVM with polynomial, sigmoid, linear and RBF kernel functions) are compared.

A decision tree is a tree structure which attempts to separate the given records into mutually exclusive subgroups. To do this, starting from the root node, each node is split into child nodes in a binary or a multi split fashion related to the method used based on the value of the attribute (input variable) which separates the given records best. Records in a node is recursively separated into child nodes until there is no split that makes statistical difference on the distribution of the records in the node or the number of records in a node is too small. Each decision tree method uses its own splitting algorithms and splitting metrics. From the well-known decision tree algorithms, ID3, C5.0 and C&RT methods use impurity measures to choose the splitting attribute and the split value/s. ID3 [25] uses information gain while the successor, C5.0 uses gain ratio, and C&RT [26] uses Gini coefficient for impurity measurements. Unlikely, CHAID uses chi-square or F statistic to choose the splitting variable [26].

As the tree is grown, the resultant tree may overfit the training data containing possible errors or noise or some of the branches of the resultant tree may contain anomalies. So, the resultant tree should be checked whether removal of some nodes, starting from the leaf ones, make a significant effect on the tree's classification performance. This operation is called as pruning.

After the tree is grown, a new observation or record is classified by tracing the route on the tree up to a leaf node according to the values of the attributes of the record. This is done by recursively checking the values of the splitting attribute of the record at each node and following the required branch of the tree until a leaf node is reached. The label of the leaf node reached gives the class which the new observation or record is classified in.

Unlike the decision tree methods, SVM tries to find a hyperplane to separate the two classes while minimizing the classification error. SVM is a new and promising classification and regression technique proposed by Vapnik and his group at AT&T Bell Laboratories [27]. SVM learns a separating hyperplane which maximizes the margin and produces good generalization ability [28]. In prior literature, SVM has been successfully applied to many areas such as telecommunication fraud detection [29], pattern recognition [28], system intrusion detection [30] SVM's basic idea is to transform the attributes to a higher dimensional feature space and find the optimal hyperplane in that space that maximizes the margin between the classes. Briefly, SVM does this by using a polynomial, sigmoid, radial basis or a linear kernel function which satisfies the Mercer condition [31]. The kernel function reflects the geometric relationship between the input record and the support vector, or the similarities of the features of the classes. The details of the theory have been given in [32].

IV. Approach

First of all, the collected data is pre-processed before starting the modeling phase. As mentioned earlier, the distribution of data with respect to the classes is highly imbalanced. The time period that is used to build our sample included 978 fraudulent records and 22 million normal ones with a ratio of about 1:22500. So, stratified sampling is used to under sample the normal records so that the models have chance to learn the characteristics of both the normal and the fraudulent records' profile. To do this, first of all, the variables that are most successful in discriminating the fraudulent and legitimate transactions are founded. Then, these variables are used to form stratified samples of the legitimate records. Later on, these stratified samples of the legitimate records are combined with the fraudulent ones to form three samples with different fraudulent to normal record ratios. The first sample set has a ratio of one fraudulent record to one normal record; the second one has a ratio of one fraudulent record to four normal ones; and the last one has the ratio of one fraudulent to nine normal ones.

The variables which form the card usage profile and the methods used to build the model make the difference in the fraud detection systems. Our aim in defining the variables used to form the data-mart is to discriminate the profile of the fraudulent card usage by the fraudsters from the profile of legitimate (normal) card usage by the card holders. We will be content with mentioning about the type of variables used but regarding the privacy, confidentiality and security concerns, we are not allowed to talk on the full list of variables (a variable is a certain level of deviation from a personal average statistics). The variables are one of the five main types: all transactions statistics, regional statistics, sectorial statistics, daily amount statistics and daily number of transactions statistics.

The chosen methods to build classifier models are C5.0, C&RT and CHAID from decision tree methods and SVMs with kernels of polynomial, sigmoid, linear and radial basis functions. All these methods are used to develop models using the three data sets. These methods and the parameters used with these methods are given in Table 1.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
INPUT PARAMETERS FOR CLASSIFIER MODELS

Classifier Model		Parameters
Decision Tree Methods	C&RT	Impur. Measure: Gini Max. Surrogates:10, Tree_depth:6
	C5.0	Impur. Measure: Entropy Pruning Severity:75% 10-Fold Cross-validate, boosting=true Tree_depth:15
	CHAID	α for splitting:0.05 α for merging:0.05 Tree_depth:5
SVM Methods	SVM with RBF Kernel	C=10 RBF ɣ=0.1
	SVM with Polynomial Kernel	C=10 ɣ=1 Bias=0 Degree=3
	SVM with Sigmoid Kernel	C=10 ɣ=1 Bias=0
	SVM with Linear Kernel	C=10

C&RT method takes two parameters. The first one is the impurity measure used for deciding which split to use which is chosen as Gini coefficient, and the second one is the maximum number of surrogates. Surrogates are a method for dealing with missing values. For each split in the tree, decision tree method identifies the input fields that are most similar to the selected split field. Those fields are the surrogates for that split. When a record must be classified but has a missing value for a split field, its value on a surrogate field can be used to make the split. The maximum number of surrogates is chosen as 10. Unlike C&RT method, C5.0 uses gain ratio in the entropy as the impurity measure.

CHAID uses Chi-Square distribution for splitting and merging operations. Accordingly, it takes two probabilities, the first one showing the significance level for splitting the node, and the second one showing the significance level for merging the nodes.

SVM models built in this study use four different kernel functions: polynomial, sigmoid, radial basis and linear kernel functions. There are two parameters associated with RBF kernel: C and ɣ. Regularization parameter (C) controls the trade-off between maximizing the margin and minimizing the training error term. Increasing the value improves the classification accuracy (or reduces the regression error) for the training data, but this can also lead to overfitting. RBF gamma should normally take a value between 3/k and 6/k, where k is the number of input fields. For example, if there are 12 input fields, values between 0.25 and 0.5 would be worth trying. Increasing the value improves the classification accuracy (or reduces the regression error) for the training data, but this can also lead to overfitting. Gamma is used in Polynomial or Sigmoid kernels. Increasing the value improves the classification accuracy (or reduces the regression error) for the training data, but this can also lead to overfitting. Bias is enabled only if the kernel type is also set to Polynomial or Sigmoid. It sets the constant coefficient value in the kernel function. The default value 0 is suitable in most cases. Degree is enabled only if Kernel type is set to Polynomial. It controls the complexity (dimension) of the mapping space.

TABLE II
TRAINING AND TEST SET SIZES FOR THE SAMPLES

Samples		# of Records	
		Training Set	Test Set
Set-1F-To-1N	Normal	1362	583
	Fraud	657	316
Set-1F-To-4N	Normal	3404	1460
	Fraud	657	316
Set-1F-To-9N	Normal	6787	2942
	Fraud	657	316

As mentioned above, three different sets are formed to compare the performances of the methods under these different conditions. In the first set, there is one normal transaction for each fraudulent one. In the second set, there are four normal transactions for each fraudulent one and there are nine normal ones for each fraudulent one in the third set. For each sample, 70% of the data, both 70% of the normal transactions and 70% of the fraudulent transactions, are taken as the training set for the models; and 30% of the data are taken as the test set to evaluate the performance of the models developed. The training set and test set of each sample is designed to have the same fraudulent records as the sets of other samples. The training and test set sizes are given in Table 2. In each sample, the number of fraudulent transactions is fixed to 973. The number of normal transactions changes according to the given ratios. For the training sets, 70% of the normal and 70% of the fraudulent transactions are chosen. The rest 30% of the normal and fraudulent transactions are left as test data set. Every model for a sample uses the same training and test data set so that a comparison of the performance of the models for each sample becomes possible.

V. RESULTS AND DISCUSSIONS

In this study, seven alternative models based on decision tree algorithms and SVM were built using the relevant training data set for each sample. To evaluate these models, we used the remaining transactions in the relevant test sets for each model. The fraudulent transactions in the test sets of the samples are identical. Accuracy rates were used to describe the usefulness of the models. Accuracy is probably the most commonly used metric to measure the performance of targeting models in classification applications.

The performance among the predictive models is presented in Table 3. The left column shows the methods used to build the models, the columns named as “Train” show the prediction accuracy of the models on the training data set of the given samples, the columns named as “Test” show the prediction accuracy of the models on the testing

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE III
PERFORMANCE OF CLASSIFIERS W.R.T. ACCURACY OVER TRAINING & TEST SETS

Classifier Model		Set-1F-To-1N				Set-1F-To-4N				Set-1F-To-9N			
		Train	Test	Build Time	Frauds	Train	Test	Build Time	Frauds	Train	Test	Build Time	Frauds
DT Methods	C&RT	90,01%	86,79%	<3m	254	92,13%	92,53%	<3m	242	93,85%	94,69%	<3m	216
	C5.0	92,51%	91,08%	<1m	260	97,44%	92,81%	<1m	227	99,15%	94,52%	<1m	176
	CHAID	92,51%	89,37%	<1m	252	92,92%	92,53%	<1m	233	94,32%	94,76%	<1m	165
SVM Methods	SVM with RBF Kernel	99,78%	83,02%	<74m	228	98,00%	88,97%	<74m	184	98,75%	93,08%	<74m	158
	SVM with Polynomial Kern.	99,78%	83,02%	<2m	228	98,00%	88,97%	<2m	184	98,75%	93,08%	<2m	158
	SVM with Sigmoid Kernel	99,78%	83,02%	<2m	228	98,00%	88,97%	<2m	184	98,75%	93,08%	<2m	158
	SVM with Linear Kernel	99,78%	83,02%	<2m	228	98,00%	89,18%	<2m	191	98,75%	93,08%	<2m	158

data set of the given samples, the columns named as "Build Time" show the time elapsed for building the given model over the given sample, and columns "Frauds" show the number of fraudulent transactions in the test data sets assigned as fraud (True Positive) by the models over the given samples.

From Table 3, it is clear that decision tree based models outperform the SVM based models when their performances are compared over the test data sets. Though the result is reverse when the performances are compared over the training data sets, it just shows that SVM based models overfit the training data. In this problem, assigning as many fraudulent transactions as fraudulent is the most important success factor. However, accuracy shows the rate of true assignments regardless of whether it a true fraud assignment or a true normal assignment. When the performances of the models are compared w.r.t. accuracy, it is seen that as the number of the training data increases, this overfitting behavior becomes less remarkable and the performances of the SVM based models become comparable to decision tree based models. But the number of frauds caught by SVM models are still far less than the decision tree models, especially C&RT model. Nevertheless, the number of actual fraudulent transactions assigned as fraudulent by the models are not strongly related with the accuracy of the assignments done by the models. Accordingly, accuracy is not a well matched performance metric for this problem domain. Though C5.0 model is the champion over the other models w.r.t. accuracy for each sample among the seven models, C&RT model catches the largest number of frauds for each sample except the first one. This provides a key factor in choosing the models and we choose the C&RT and C5.0 models as the final methods to build the prediction model.

VI. CONCLUSIONS

As usage of credit cards become more and more common in every field of the daily life, credit card fraud has become much more rampant. To improve security of the financial transaction systems in an automatic and effective way, building an accurate and efficient credit card fraud detection system is one of the key tasks for the financial institutions. In this study, seven classification methods were used to build fraud detecting models. The work demonstrates the advantages of applying the data mining techniques including decision trees and SVMs to the credit card fraud detection problem for the purpose of reducing the bank's risk. The results show that the proposed classifiers of C&RT and other

decision tree approaches outperform SVM approaches in solving the problem under investigation. However, as the size of the training data sets become larger, the accuracy performance of SVM based models reach the performance of the decision tree based models, but the number of frauds caught by SVM models are still far less than the number of frauds caught by decision tree methods, especially C&RT model.

Under this framework, financial institutions can utilize credit card fraud detection models to compare the transaction information with the historical profile patterns to predict the probability of being fraudulent for a new transaction, and provide a scientific basis for the authorization mechanisms. Furthermore, resources of the institutions can be focused on more suspicious transactions to decrease the fraud levels.

As a future work, other data mining algorithms such as different versions of Artificial Neural Networks (ANN) and logistic regression will be used to build new classification models on the same real world dataset and the performance of the new models will be compared with the performance of the models given in this paper. Also, instead of making performance comparisons just over the prediction accuracy, these comparisons will be extended to include the comparisons over other performance metrics.

REFERENCES

[1] Hobson, A. 2004. The Oxford Dictionary of Difficult Words. The Oxford University Press. New York.
[2] Bolton, R. J. and Hand, D. J. 2002. Statistical fraud detection: A review. Statistical Science 28(3):235-255.
[3] Kou, Y. et. al. 2004. Survey of fraud detection techniques. In Proceedings of the 2004 IEEE International Conference on Networking, Sensing and Control, Taipei, Taiwan, March 21-23.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[4] Phua, C. et al. 2005. A comprehensive survey of data mining-based fraud detection research. Artificial Intelligence Review.
[5] Sahin, Y., Duman E. An overview of business domains where fraud can take place, and a survey of various fraud detection techniques. In Proceedings of the 1st International Symposium on Computing in Science and Engineering, Aydin, Turkey, June 2010.
[6] Leonard, K. J. 1993. Detecting credit card fraud using expert systems. Computers and Industrial Engineering, 25
[7] Ghosh, S. and Reilly D. L. 1994. Credit card fraud detection with a neural network. In Proceedings of the 27th Hawaii International Conference on system Sciences, 3.
[8] Mena, J. 2003. Investigate Data Mining for Security and Criminal Detection. Amsterdam, Butterworth-Heinemann.
[9] Aleskerov, E., Freisleben, B. and Rao, B. 1997. CARDWATCH: A neural network based data mining system for credit card fraud detection. In Computational Intelligence for Financial Engineering.
[10] Chen, R., Chiu, M., Huang, Y. and Chen, L. 2004. Detecting credit card fraud by using questionnaire-responded transaction model based on SVMs. In Proceedings of IDEAL2004.
[11] Brause, R., Langsdorf, T. and Hepp, M. 1999. Neural data mining for credit card fraud detection. In Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence.
[12] Stolfo, S. J., Fan, D. W., Lee, W., Prodromidis, A. L. and Chan, P. K. 1997. Credit card fraud detection using meta-learning: Issues and initial results. In AAAI Workshop on AI Approaches to Fraud Detection and Risk Management. AAAI Press, Menlo Park, CA.
[13] Stolfo, S., Fan, W., Lee, W., Prodromidis, A. L. and Chan, P. 1999. Cost-based modeling for fraud and intrusion detection: Results from the JAM Project. In Proceedings of the DARPA Information Survivability Conference and Exposition. IEEE Computer Press, New York.
[14] Prodromidis, A. L., Chan P. and Stolfo S. J. 2000. Meta-learning in distributed data mining systems: issues and approaches. Advances of Distributed Data Mining, Editors Kargupta H. and Chan, P. AAAI Press.
[15] Chen, R.-C., Luo, S.-T., Liang, X. and Lee, V. C. S. 2005. Personalized approach based on SVM and ANN for detecting credit card fraud. In Proceedings of the IEEE International Conference on Neural Networks and Brain, Beijing, China
[16] Hand, D. J. and Blunt G. 2001. Prospecting gems in credit card data. IMA Journal of Management Mathematics, 12.
[17] Dahl, J. 2006. Card Fraud. In Credit Union Magazine.
[18] Schindeler, S. 2006. Fighting Card Fraud in the USA. In Credit Control, House of Words Ltd.
[19] Dorronsoro, J. R., Ginel, F., Sanchez, C. and Cruz, C. S. 1997. Neural fraud detection in credit card operations. IEEE Transactions on Neural Networks, 8
[20] Hanagandi, V., Dhar, A. and Buescher, K. 1996. Density-Based Clustering and Radial Basis Function Modeling to Generate Credit Card Fraud Scores. In Proceedings of the IEEE/IAFE 1996 Conference.
[21] Juszczak, P., Adams, N. M., Hand, D. J., Whitrow, C., and Weston, D. J. 2008. Off-the-peg and bespoke classifiers for fraud detection. Computational Statistics & Data Analysis. 52(9).
[22] Quah, J. T. and Sriganesh, M. 2008. Real-time credit card fraud detection using computational intelligence. Expert Systems with Applications. 35(4).
[23] Shen, A., Tong, R., Deng, Y. 2007. Application of classification models on credit card fraud detection. In International Conference on Service Systems and Service Management, Chengdu, China, June 2007.
[24] Syeda, M., Zhang, Y. and Pan, Y. 2002. Parallel granular neural networks for fast credit card fraud detection. In Proceedings of the 2002 IEEE International Conference on Fuzzy Systems.
[25] Han, J., & Camber, M. (2000). Data mining concepts and techniques. San Diego, USA: Morgan Kaufman
[26] Koh, H. C., & Low, C. K. (2004). Going concern prediction using data mining techniques. Managerial Auditing Journal, 19(3), 462–476.
[27] Cortes, C., Vapnik, V. 1995. Support vector network. Machine Learning, vol:20 pg.273–297
[28] Burges, C.J.C. 1998. A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, Vol. 2, No.2, pp. 955-974.
[29] Kim, H.-C., Pang, S., Je, H.-M., Kim, D., Bang, S. Y. 2003. Constructing support vector machine ensemble. Pattern Recognition, Vol. 36, pg. 2757 – 2767
[30] Chen, R.C., Chen, J., Chen, T.S., Hsieh, C.H., Chen, T.Y. and K.Y. Wu. 2005. Building an Intrusion Detection System Based on Support Vector Machine and Genetic Algorithm. Lecture Notes in Computer Science (LNCS), Vol. 3498, pp. 409-414.
[31] Mercer, J. Function of positive and negative type and their connection with theory of integral equations, Philosophical Transactions of the Royal Society, London A 209 (1909) 415–446.
[32] Cristianini, N., Shawe-Taylor, J. An Introduction to Support Vector Machines and other Kernel-based Learning Methods, Cambridge University Press, Cambridge, UK, 2000.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)