American Journal of Intelligent Systems 2015, 5(1): 27-33
DOI: 10.5923/j.ajis.20150501.03

An Antifraud System for Tourism SMEs in the Context of Electronic Operations with Credit Cards

Fidel Rebón[*], Iñigo Castander, Jon Argandoña, Jon Kepa Gerrikagoitia, Aurkene Alzua-Sorzabal

CI Ctour GUNE, Donostiako Parke Teknologikoa, Mikeletegi Pasealekua, Donostia, Spain

Abstract The new paradigm of marketing and communication, in which the "tourism marketer" conducts its strategies by adapting services to the tourists' interests has enabled an increase in the number of transactions made by tourists, before, during and after the trip. This act, in which both parties, buyer and seller, annul their differences are generally carried out on the internet by means of credit cards. Knowledge Discovery in Databases techniques bring up an optimal solution for discovering fraud conducted in internet transactions before it is executed. This study proposes a novel system of antifraud model in order to improve the analyzed techniques and help in the decision-making process of fraud with credit cards in electronic transactions for the tourism SMEs.

Keywords Credit Fraud, Credit Fraud Model Process, Knowledge Discovery in Databases

1. Introduction

Traditionally, sociocultural elements of the environment have been embraced by human beings and incorporated into their personality in order to adapt to society's needs. However, the real world is leading to an attractive digital and virtual realm that, in contrast, is able to rapidly adapt to these requirements. Society has started to respond to the globalization and digitalization phenomena, where Information and Communication Technologies (ICTs) represent one of the pillars in the new knowledge economy [1].

In the context of tourism, the new paradigm of marketing and communication, in which the "tourism marketer" conducts its strategies by adapting services to the tourists' interests [2], has enabled an increase in the number of transactions made by tourists before, during and after the trip [3], [4]. This act, in which both of the parties, buyer and seller annul their differences, is generally carried out through credit cards based on the Internet.

The statistics on electronic Commerce (e-commerce) state that between the years 2009 to 2012 the volume of online buyers in the United States has experienced a growth of over 22 million. Forecasts estimate that by 2015 these numbers will be up to around 38 million [5]. Not only the number of users will rise rapidly, but also the business volume will increase. For example, in the first quarter of 2013, the 25 million transactions originated in Spain and directed out of its borders represented an amount of 1.2 billion euros. The largest part of the purchases was directed towards the European Union, adding up to 89% of the total. The transactions directed to the United States stand for a 5%. The value of the foreign transaction directed to Spanish websites reached 463 million euros. The biggest share of the transactions had its origin in the European Union, counting for a total volume of business of around 72% [6].

The data provided shows an upward tendency in the number of daily trade transactions that will take place in the market. This fact inevitably leads to the increase of fraud in credit cards operations. This type of fraud has an estimated waste cost of billions of dollars for both consumers and the financial industry [7].

These financial and industrial entities are the ones that have put more means for preventing this type of fraud. Nevertheless, small and medium size enterprises (SMEs) have little access to resources, capital and knowledge needed to fight against fraud. Furthermore, a large share of enterprises in the tourism sector pertains to this structural segment [8]. Besides this, the incorporation of ICTs to the cooperative network culture is barely developed, despite it could positively influence in its level of competitiveness [9].

Actually, information technology and Business Intelligence provide a novel direction to support enterprise business in the new age [10]. At this moment a confluence of practices is taking place and technologies are leading into smarter computing capacities. This paradigm shift will enable organizations to overtake intelligent actions to address time-sensitive business processes and benefit from analysis [11]. This analysis is developed by applying Knowledge Discovery in Databases (KDD). These techniques bring up an optimal solution for discovering

* Corresponding author:
FidelRebon@tourgune.org (Fidel Rebón)
Published online at http://journal.sapub.org/ajis

fraud before it is executed in transactions conducted by means of the credit cards. This is a non-trivial process as discovering knowledge and useful information within the data contained in a repository of information is a tedious task [12]. There is a high natural imbalance within the data when significant and critical characteristics regarding fraudulent cases from the non-fraudulent are extracted.

In order to make the fraud detection in commercial transactions carried out with credit cards easier and also to be able to detect them before they are executed, the authors of this study will analyze the different existing techniques. After that, an improved solution to help in the detection of fraud with credit cards in electronic transactions for tourism SMEs is presented.

2. Related Work

A broad variety of fraud forms have been gathered in a taxonomy [13] (See Figure 1).

Until the emergence of information technologies and the internet Internal, Insurance and Customs were the prevailing fraud types. Internal fraud is the fraud perpetrated against an insurance company or its policyholders by insurance agents, managers, executives, or other insurance employees. Currently, frauds such as Web Network, Fraud Computer Intrusion, Telecommunication and Credit Card are more common than the Internal type one.

Fraud, by means of credit cards can be effectuated in two manners: a) (Credit) Application Fraud: it is produced when fraudsters obtain credit cards from emitterfraud; and b) Behavioral or Credit Card Transaction Fraud: it is happened when cards have been stolen or card's holder is no presented[14], [15] (See Figure 2).

Once the credit fraud types considered by this study have been explained, the techniques are introduced. Several techniques have been applied in recent years to detect fraud on credit cards a short summary of the most significant ones is provided next.

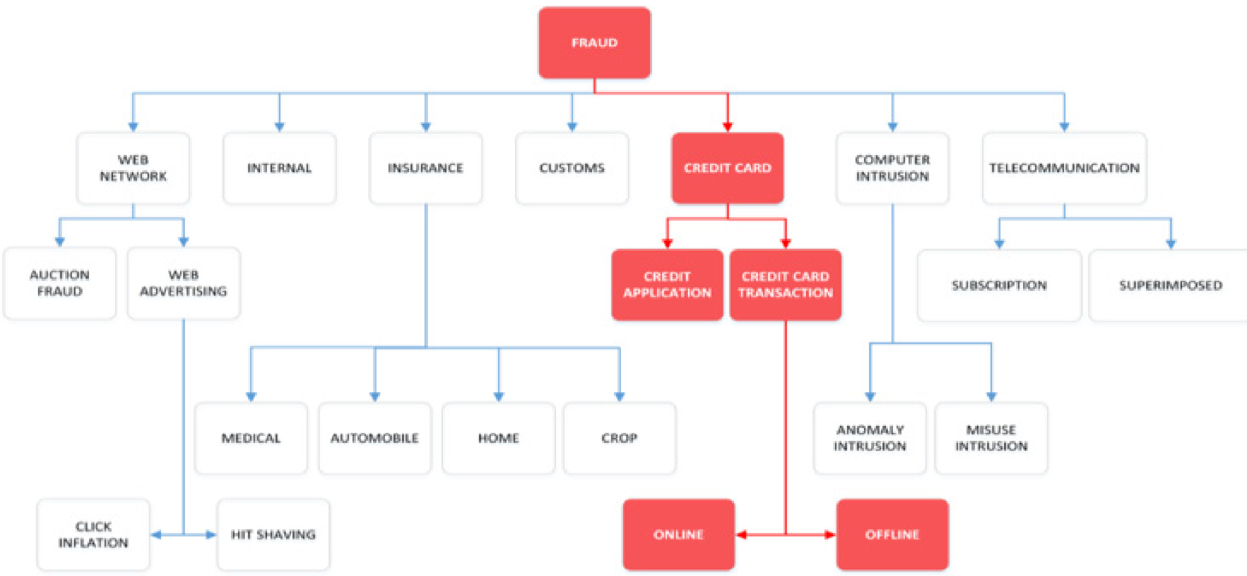

Figure 1. Taxonomy of fraud types [Laleh & Azgomi, 2009]

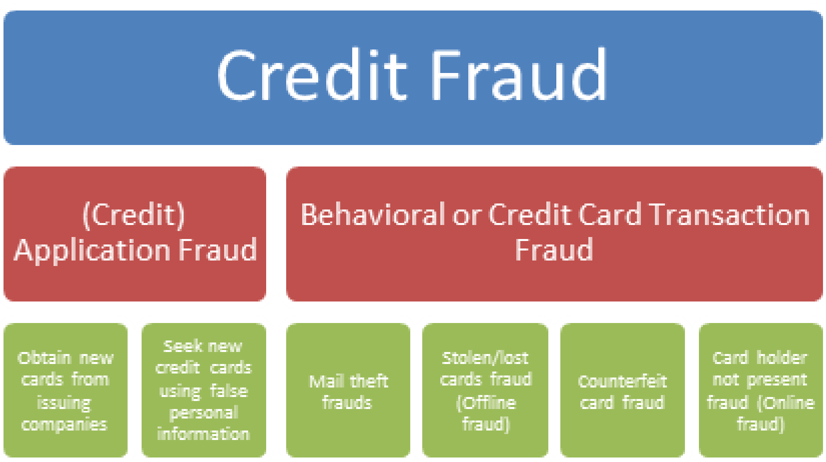

Figure 2. Credit fraud types

The Linear Discriminant Analysis (LDA) [16-21] and the Logistic Regressions (LR) [22-25] are well known statistical tools that have been used on classification tasks for a long time in sectors such as the financial one. The LDA performs a classification by means of a linear discriminant function that utilizes the centroids in order to separate the two groups considered. However, a better group discrimination is often achieved with the LR [26]. This happens because the LR exploits the reliance of the binomial response on a set of independent variables. Artificial Neural Networks (ANN) and Self-Organizing Maps (SOM) are recently created techniques that have achieved promising results on some classification tasks inside of credit fraud field.

ANN technique is a universal approach by means of functions that have been successfully used in a vast amount of research fields: air traffic control, voice and character recognition, medical diagnosis and research, weather forecasting, etc. But they have also been intensely applied in economics and finances [27]. This technique represents an architecture compounded by neurons receiving inbound variables; they are later processed in order to provide a solution in accordance to one or several variables. An implementation of this technique is called SOM [28]. Its main characteristics are: the encapsulation of its functionality and its enriched visualization.

Classification and Regression Tress (CART) and RF (Random Forest) based their architecture in a tree shape structure of which the purpose is to conduct a data classification or regression [29]. CART implies the application of three steps: 1) a tree named "*maximal tree*" is built; it deeply describes the training set and it grows by means of a method of binary separation; 2) the tree is pruned and derived towards a series of less complex trees and 3) through the Cross-Validation (CV) method, the tree with the optimal size is selected.

Although the greater applicability of CART has been in the medical sciences [30], in the social sciences [31] and in meteorology for weather forecasting [32], it has also been applied to models for credit granting validation [28]. In contrast to the previous method, in RF a group of trees is available for classification. The process improves whether each of the trees is different and whether this randomness among trees is provided through diverse independent training data samples and considering a subset of the data attributes randomly selected [33]. Besides, its application on the fraud detection area is rather new.

The purpose of AR is to create a vast number of objective and subjective rules and, then, the useful ones are chosen. In an objective way, it is done through the statistical analysis of the data. Definitely, the tendencies are located inside the data. Although this technique is mostly used in the "market-basket" analysis, it is also selected for the extraction of knowledge from the users' behavioral patterns [34]. This occurs mainly because the algorithm "*Min Has*" works properly in situations of low support and high confidence [35].

Support Vector Machine (SVM) is a powerful Machine Learning (ML) technique enabling data classification and regressions [36]. For this purpose, it generates a hyper-surface that empowers the classification of that data meeting the condition in one side, and on the other, those data not meeting the condition. The suggested data separation is fulfilled by obtaining the largest distance among the closest data meeting the condition and those data not reaching it [20]. Simplicity in the precise moment of applying lineal classifications, and its ability to use a vast field of variables, are those facts making it such an attractive inside of credit fraud field technique for fraud detection.

Inspired in the natural evolution, Genetic Algorithm (GA) have the goal of providing a better solution while the time goes, as the probability of surviving for those members pertaining to the strongest group of population is bigger than that of the weakest ones. Usually, new generations are produced by crossing the members, but mutations can also appear within the members enabling population diversity [34]. It must be highlighted that this technique easily admits complementarity in the evolutionary algorithm, thus enabling the process improvement.

3. Results and Methodology

The study has firstly performed an extensive review of the state of the art. Once the current capabilities and deficiencies have been identified, the authors propose a novel system which facilitates an early detection of fraudulent transactions.

A comparison among the methods described in the previous section is performed first. It should note that this comparison is extracted by means of the conclusions based on related work.

With this aim, a value between zero and one is assigned to each cell. There are two cases in which the value will be zero: a) the comparison is performed on the technique itself; b) the proposed technique on the row has not a behavior as good as the technique offered in the column. Otherwise, the assigned value will be one (See Table 1).

Table 1. Comparison among different techniques of KDD

	LDA & LR	ANN (SOM)	CART & RF	AR	SVM	GA
LDA & LR	0	0	0	0	0	-
ANN (&SOM)	1	0	0	0	1	0
CART&RF	1	1	0	-	1	-
AR	1	1	-	0	1	-
SVM	1	0	0	0	0	0
GA	-	1	-	-	1	0

The comparisons among the proposed techniques provide the following conclusions:

- ANN techniques are more accurate and reliable than LDA and RL [22], [28]
- In most of the scenarios RF presents a better behavior than LR and SVM [15]
- SVM technique are not as precise as ANN and GA when used for credit requests classification [30]
- CART technique offers better classification ratios than other techniques like LDA, LR, ANN and SVM. This does not occur in the context of fraudulent detection but in the area of credit granting [28]
- ANN has a higher overall accuracy than the LR [37]
- When working with vast volume of data that contains information gathered during several years, LR offers better results than AR [30]
- SVM present some improvements compared to LDA and LR [38]. This difference is not so significant with respect to the LR technique when the subject of interest is credit granting related data [34].

3.1. Credit Fraud Model Process

CFMP is an intelligent and expert system, capable of validating payment operations through a digital media before they have happen.

It is considered an intelligent and expert system because it initially learns from an expert and it continues improving through experience as it grows. All the decisions are stored and are re-evaluated in order to anticipate to the new fraud behaviors.

In order to be able to discard an operation, CFMP applies KDD techniques. These techniques look for valid, novel and potentially useful and comprehensible patterns through transaction history.

3.1.1. Internal Operation

The process is modelled in the system by a series of operations made on the initial commercial operation data (See Figure 3 and Figure 4).

The initial data are received by a Multivariable Adaptation Regression Splines (MARS) task; which will reduce the input variables to the most significant. This will facilitate working in a more efficient fashion. After this, the variables are converted in an intake and they are checked by an ANN.

These inputs are credit cards' data and/or the information about the cards' owner that the MARS consider more significant.

The evaluation will indicate if the operation is "*fraudulent*" or "*not fraudulent*". In case of a "*not fraudulent*" result the input will pass to a *Case Based Reasoning* (CBR). The CBR has the aim of filtering incorrectly detected fraud cases.

The mechanism of the CBR is simple; it looks up the number of previous inputs in the database that are close to the input and show a "*fraudulent*" result (SG) in front of the previous inputs that are close to the input and hold a "*not fraudulent*" result (SB). If SG is less or equal to the SB, the CBR will indicate that the operation as "*not fraudulent*", however, if the SG is greater than SB, the CBR will show it as "*fraudulent*".

Once the operation is discarded or accepted, the system will store the result with their respective inputs in order to be able to incorporate this new case to CBR.

The system will also evaluate the raid of the new case inside the training data of ANN, seriously considering to add it if the case has been considered "*fraudulent*".

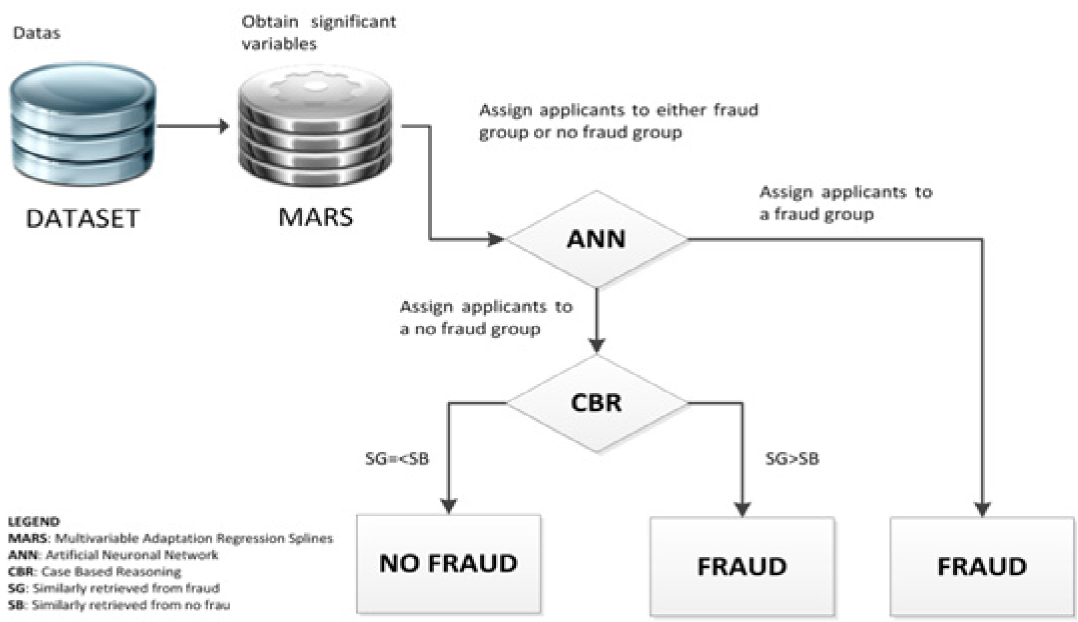

Figure 3. Internal Operation of CFMP (1/2)

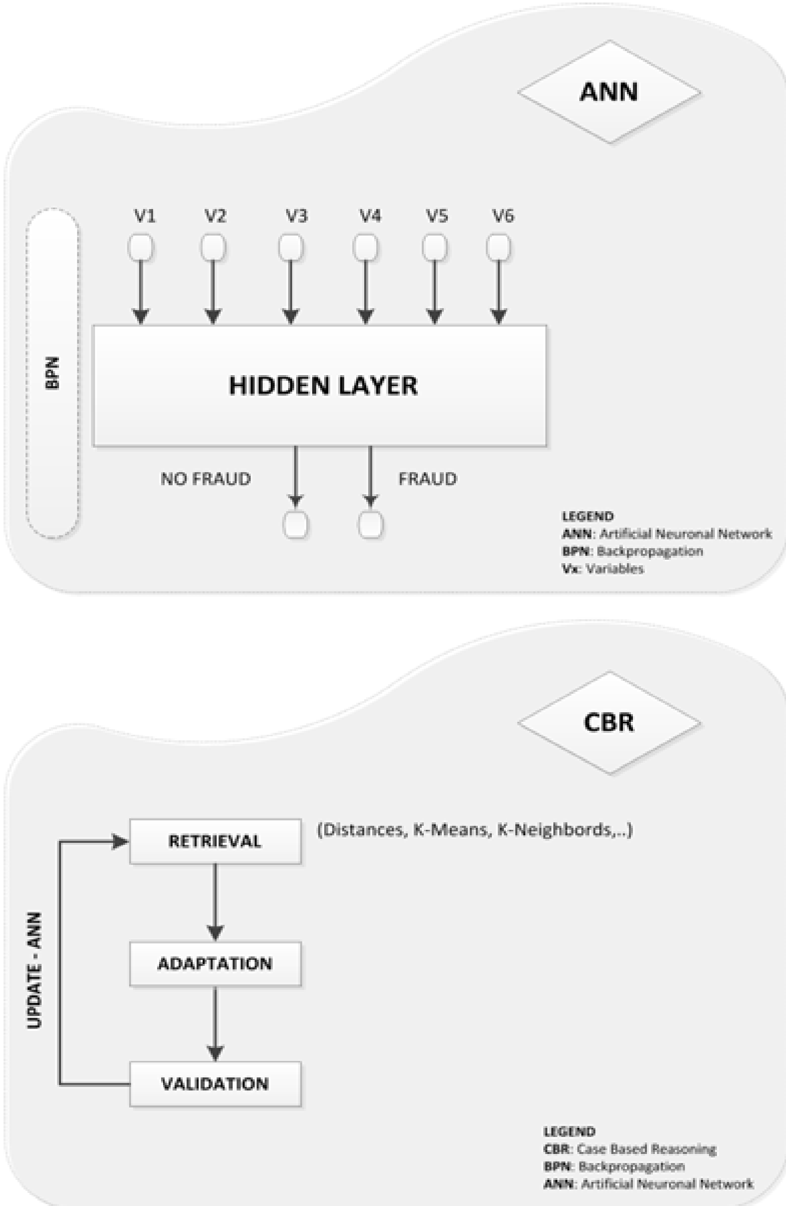

Figure 4. Internal Operation of CFMP (2/2)

3.1.2. Advantages of the CFMP over Other Tools

When considering fraud detection tools, it is visible that CFMP and, in particular, the CBR component is better than GA and ANN if the features (skills, functional facilities,…) are taken into consideration. However, it is not the case when comparing LR and ANN [37]. This happens because ANN has a higher overall accuracy than LR but it has a lower one than the AR when working with datasets containing several years of data. It is also highlighted, that considering the same sample, AR achieves a better precision than LR.

In addition, it should be noted that the system includes all MARS + ANN + CBR techniques and those have a more optimal ratio of accuracy, as a whole, than other techniques utilized in the current context of fraud credit: LDA, LR, CART Y ANNs [22].

Furthermore, the technical level of CBR entitles the usefulness of supplying a similar case based on prior knowledge. This allows reallocating a case that has already been classified. From a business perspective, the combination of the ANN + CBR provides a significant improvement to the decision makers [22].

4. Conclusions and Future Work

This study has presented a new novel system to prevent fraud based on commercial transactions.

CFMP is a system that can improve the current lack of preventive systems by means of KDD techniques.

This proposal should be elevated to another testing data environment, where the system and the other techniques will confront again; although the advantages of the CFMP about these KDD methods are clearly demonstrated.

In addition, this system must adapt to the users' behavior changes especially in Internet related purchasing activities performed from mobile apps. Having in account that the mobile commerce revenue is expected to reach 31 billion U.S. dollars in 2016 it will be crucial in the future.

ACKNOWLEDGMENTS

We would like to thank the participants and partners in European project: "e-COMMERCE Proficient Applications in Security and Sales for SMEs (SME E-COMPASS)".

REFERENCES

[1] M. Castells, La sociedad red: unavisión global. Alianza Editorial, 2006.

[2] O. Miralbell, "Portales de destinosturísticosen Internet: unareflexiónestratégica," Métodos Inf., vol. 8, no. 42, pp. 74–81, 2010.

[3] S. Park, D. Wang, and D. Fesenmaier, "Assessing structure in American online purchase of travel products," Anatolia, vol. 22, no. 3, pp. 401–417, 2011.

[4] L. Li and D. Buhalis, "E-Commerce in China: The case of travel," Int. J. Inf. Manage., vol. 26, no. 2, pp. 153–166, 2006.

[5] Stadista, "E-commerce and mail-order trade: statistics and market data," 2014. [Online]. Available: http://www.statista.com/markets/10/e-commerce-mail-order-trade/.

[6] Dirección de Estudios, Estadísticas Documentales, "Informesobre el comercioelectrónicoen España a través de entidades de medios de pago", 2013. [Online]. Available: http://www.cmt.es/c/document_library/get_file?uuid=e995b42a-bd79-4dc2-a7ea-b0e068033da9&groupId=10138

[7] R.-C. Chen, T.-S. Chen, and C.-C. Lin, "A new binary support vector system for increasing detection rate of credit card fraud," Int. J. Pattern Recognit. Artif.Intell., vol. 20, no. 02, pp. 227–239, 2006.

[8] B. Walder, "Sources and determinants of innovations-the role of market forces," B. Walder, K. Weiermaier, A. Sancho Innov. Prod. Dev. Tour. Berlin, Erich Schmidt Verlag, vol. 164, 2006.

[9] A.-M. Hjalager, "Repairing innovation defectiveness in tourism," Tour. Manag., vol. 23, no. 5, pp. 465–474, 2002.

[10] S. Chaudhuri, U. Dayal, and V. Narasayya, "An overview of business intelligence technology," Commun.ACM, vol. 54, no. 8, pp. 88–98, 2011.

[11] A. Alzua-Sorzabal, J. K. Gerrikagoitia, and E. Torres-Manzanera, "Opening and closing Internet booking channel for hotels: A first approximation," Tour. Manag. Perspect., vol. 5, no. 0, pp. 5–9, 2013.

[12] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Mag., vol. 17, no. 3, p. 37, 1996.

[13] N. Laleh and M. A. Azgomi, "A taxonomy of frauds and fraud detection techniques," in Information Systems, Technology and Management, Springer, 2009, pp. 256–267.

[14] S. Jha, M. Guillen, and J. Christopher Westland, "Employing transaction aggregation strategy to detect credit card fraud," Expert Syst. Appl., vol. 39, no. 16, pp. 12650–12657, 2012.

[15] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decis. Support Syst., vol. 50, no. 3, pp. 602–613, 2011.

[16] R. A. Fisher, "The use of multiple measurements in taxonomic problems," Ann. Eugen., vol. 7, no. 2, pp. 179–188, 1936.

[17] J.-C. Kim, D.-H. Kim, J.-J. Kim, J.-S. Ye, and H.-S. Lee, "Segmenting the Korean housing market using multiple discriminant analysis," Constr. Manag. Econ., vol. 18, no. 1, pp. 45–54, 2000.

[18] E. K. Laitinen, "Predicting a corporate credit analyst's risk estimate by logistic and linear models," Int. Rev. Financ. Anal., vol. 8, no. 2, pp. 97–121, 1999.

[19] L. J. Trevino and J. D. Daniels, "FDI theory and foreign direct investment in the United States: a comparison of investors and non-investors," Int. Bus. Rev., vol. 4, no. 2, pp. 177–194, 1995.

[20] E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," Expert Syst. Appl., vol. 38, no. 10, pp. 13057–13063, 2011.

[21] V. S. Desai, J. N. Crook, and G. A. Overstreet Jr, "A comparison of neural networks and linear scoring models in the credit union environment," Eur. J. Oper.Res., vol. 95, no. 1, pp. 24–37, 1996.

[22] C. Chuang and R. Lin, "Constructing a reassigning credit scoring model," Expert Syst. Appl., vol. 36, no. 2, pp. 1685–1694, 2009.

[23] W. Kay Olsen, A. Warde, and L. Martens, "Social differentiation and the market for eating out in the UK," Int. J. Hosp. Manag., vol. 19, no. 2, pp. 173–190, 2000.

[24] E. K. Laitinen and T. Laitinen, "Bankruptcy prediction: application of the Taylor's expansion in logistic regression," Int. Rev. Financ. Anal., vol. 9, no. 4, pp. 327–349, 2001.

[25] E. H. Suh, K. C. Noh, and C. K. Suh, "Customer list segmentation using the combined response model," Expert Syst. Appl., vol. 17, no. 2, pp. 89–97, 1999.

[26] F. E. Harrell and K. L. Lee, "A comparison of the discrimination of discriminant analysis and logistic regression under multivariate normality," Biostat. Stat. Biomed. Public Heal. Environ. Sci. North-Holland, New York, United States, pp. 333–343, 1985.

[27] R. Wehrens and L. M. C. Buydens, "Self-and super-organizing maps in R: the Kohonen package," J. Stat. Softw., vol. 21, no. 5, pp. 1–19, 2007.

[28] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[29] I. Kurt, M. Ture, and A. T. Kurum, "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease,"

Expert Syst. Appl., vol. 34, no. 1, pp. 366–374, 2008.

[30] S. Zhang, C. Tjortjis, X. Zeng, H. Qiao, I. Buchan, and J. Keane, “Comparing data mining methods with logistic regression in childhood obesity prediction,” Inf. Syst. Front., vol. 11, no. 4, pp. 449–460, 2009.

[31] R. E. Davis, K. Elder, D. Howlett, and E. Bouzaglou, “Relating storm and weather factors to dry slab avalanche activity at Alta, Utah, and Mammoth Mountain, California, using classification and regression trees,” Cold Reg. Sci. Technol., vol. 30, no. 1, pp. 79–89, 1999.

[32] T.-S. Lee, C.-C. Chiu, Y.-C. Chou, and C.-J. Lu, “Mining the customer credit using classification and regression tree and multivariate adaptive regression splines,” Comput. Stat. Data Anal., vol. 50, no. 4, pp. 1113–1130, 2006.

[33] 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, pp. 3630–3640, 2009.

[34] T. Bellotti and J. Crook, “Support vector machines for credit scoring and discovery of significant features,” Expert Syst. Appl., vol. 36, no. 2, pp. 3302–3308, 2009.

[35] V. Vapnik, “Statistical learning theory. 1998.” Wiley, New York, 1998.

[36] R.-C. Chen, M.-L. Chiu, Y.-L. Huang, and L.-T. Chen, “Detecting credit card fraud by using questionnaire-responded transaction model based on support vector machines,” in Intelligent Data Engineering and Automated Learning--IDEAL 2004, Springer, 2004, pp. 800–806.

[37] I. Bose and R. K. Mahapatra, “Business data mining - a machine learning perspective,” Inf. Manag., vol. 39, no. 3, pp. 211–225, 2001.

[38] D. Vidaurre, C. Bielza, and P. Larrañaga, “Classification of neural signals from sparse autoregressive features,” Neurocomputing, 2013.